Universidad Nacional
del Nordeste

Universidad Nacional
de Misiones

Universidad Nacional del Nordeste

Facultad de Ciencias Exactas y Naturales y Agrimensura

Trabajo Final de Maestría en Tecnologías de la Información

PROCESAMIENTO Y ANÁLISIS DE DATOS DEL SISTEMA EDUCATIVO DE UNA JURISDICCION ESPECÍFICA EMPLEANDO BASES DE DATOS NOSQL

Autor: Víctor Hugo Ruchinsky

Directora: Dra. María Fernanda Golobisky

Codirector: Dr. Aldo Vecchietti

Año 2024

Dedicatoria

A mis hijos: Alejandra Mariel, Andrea Verónica, Valeria Nazarena y Hugo Nicolás. A Graciela, mi Señora. A mis familiares, a los que están y a los que ya partieron como Marisa. ¡¡¡A Dios y a la Virgen María!!!

Resumen

Los datos educativos abiertos existentes en grandes conjuntos de datos públicos, se potencian y mejoran su calidad con la utilización de una base de datos NoSQL («Not Only SQL»). El presente Trabajo Final de Maestría consiste en extraer datos educativos abiertos de la jurisdicción de Formosa, gestionarlos en modo local utilizando la plataforma MongoDB Compass, procesarlos, analizarlos y presentarlos, utilizando MongoDB Atlas, un servicio de la base de datos MongoDB en la nube. La potencia de estas tecnologías se pone en valor, a la hora de organizar la información en una estructura flexible, en el manejo y modificación de datos autocontenidos, por medio de la utilización de objetos JSON. Esta potencia permite, además, tener en un solo lugar, datos referidos a distintos sistemas de relevamiento de datos educativos, recolectados por más de una década, como ser, Relevamientos Anuales de 2010 a 2023 de la educación común y sus modalidades; y operativos destinados a la evaluación de estudiantes del Sistema Educativo Argentino, tales como ONE 2013 y operativos Aprender. El almacenamiento en una base de datos NoSQL, garantizó la disponibilidad y actualización de los datos, reduciendo la imprecisión y la inconsistencia. Además, se logró auditar los resultados obtenidos con los anuarios estadísticos. Asimismo, se realizó el análisis y discusión de los indicadores de la educación secundaria, incorporando nuevas categorías de promovidos.

Este trabajo resalta la importancia de mejorar la calidad y precisión de los datos educativos para fortalecer el análisis y la toma de decisiones en el ámbito educativo.

Palabras claves: *Base de Datos, Ciencia de Datos, Datos Abiertos, Relevamientos Anuales, Operativos ONE, Operativos Aprender.*

Abstract

Existing open educational data in large public datasets are enhanced and improved in quality by using a NoSQL database ("Not Only SQL"). The present Master's Thesis consists of extracting open educational data from the jurisdiction of Formosa, managing them locally using the MongoDB Compass platform, processing, analyzing and presenting them using MongoDB Atlas, a MongoDB database service in the cloud. The power of these technologies is valued when organizing information in a flexible structure, in the management and modification of self-contained data, through the use of JSON objects. This power also allows

having in one place, data referring to different educational data collection systems, collected for more than a decade, such as, Annual Surveys from 2010 to 2023 of common education and its modalities; and operations aimed at the evaluation of students of the Argentine Educational System, such as ONE 2013 and Aprender operations. The storage in a NoSQL database ensured the availability and updating of data, reducing inaccuracy and inconsistency. In addition, it was possible to audit the results obtained with the statistical yearbooks. Furthermore, the analysis and discussion of secondary education indicators was carried out, incorporating new categories of promoted students.

This work highlights the importance of improving the quality and accuracy of educational data to strengthen analysis and decision-making in education.

Keywords: *Database, Data Science, Open Data, Annual Surveys, ONE Operations, Learn Operations.*

Agradecimientos

A María Fernanda Golobisky y Aldo Vecchietti, que son mis Directores de Tesis.

A Melina Vidoni, por todo su acompañamiento.

A las Autoridades de la Maestría, en las personas de Laly Dapozo, Horacio Kuna, David La Red Martinez y Emmanuel Irrazabal.

A los Profesores de la Maestría.

Al Personal No Docente de la Maestría, en la persona de Silvana Romero.

Índice de Contenido

Capítulo 1 ..... *19*
Introducción *19*
1.1 Introducción *20*
1.2 Estado del arte. *22*
1.3 Objetivos *23*
Capítulo 2 ..... *25*
Marco teórico *25*
2.1 Caracterización del Sistema Educativo Argentino *26*
2.2 Caracterización de datos abiertos del Sistema Educativo Argentino *27*
2.3 Necesidad de una base de datos NoSQL *28*
2.3.1 La utilización de MongoDB *29*
2.3.2 ¿Qué es JSON: JavaScript Object Notation? *29*
2.4 Metodología del trabajo *30*
Capítulo 3 ..... *31*
Construcción de la base de datos en MongoDB *31*
3.1 Recolección y limpieza de los datos *32*
3.1.1 Transformaciones iniciales *33*
3.1.2 Eliminación de filas sin información *37*
3.1.3 Importación de datos con el MongoDB Compass *37*
3.2 Primera aproximación a la estructura de la base de datos *41*
3.2.1 Composición de las colecciones *41*
3.2.2 Comprobación de totales generales con datos externos *42*
3.2.3 Ejemplos de relación CUE_anexo-ID *42*
3.2.4 Límite en el análisis de la relación CUE-ID *47*
Capítulo 4 ..... *49*
Procesamiento en modo local y completamiento de datos *49*
4.1 Necesidad de procesamiento de los datos en modo local *50*
4.1.1 Agrupamiento por año, sector y ámbito *50*
4.1.2 Combinación de colecciones por año, sector y ámbito *57*
4.2 Completamiento de datos e información *58*

4.2.1 Completamiento de la variable "Unidades Educativas" de los 4 niveles de la educación común *58*
4.2.2 Completamiento de las modalidades *59*
4.2.3 Completamiento de información de Operativos Aprender *61*
4.2.3.1 Completamiento de datos de Aprender de Primario 2021 ..... *61*
4.2.3.2 Completamiento de datos de Aprender 2022 de Secundario *62*
4.2.3.3 Completamiento de datos de Aprender 2023 de Primaria – Operativo censal *63*
4.2.4 Completamiento de datos geoespaciales *64*
4.2.4.1 Utilización de herramientas de visualización y carga de datos *64*
4.2.4.2 Transformación de latitud y longitud *67*
Capítulo 5 ..... *69*
Análisis y presentación de resultados de las principales variables en la nube *69*
5.1 Relevamientos Anuales 2010-2022 y padrón de establecimientos *70*
5.1.1 Exportación en formato JSON e importación en la nube *70*
5.1.2 Geolocalización y mapas *72*
5.1.2.1 Consulta de proximidad de establecimientos educativos cercanos a otro establecimiento educativo geolocalizado *72*
5.1.2.2 Mapas en la nube utilizando MongoDB Atlas *73*
5.1.3 Cuadros estadísticos *75*
5.1.3.1 Cuadros de educación común de 4 niveles educativos y modalidades especial y adulto *75*
5.1.3.2 Cuadros de la educación común - nivel inicial *80*
5.1.3.3 Cuadros de la educación común - nivel primario *81*
5.1.3.4 Cuadros de la educación común - nivel secundario *82*
5.1.3.5 Evolución de las principales variables de la educación común - nivel superior no universitario *84*
5.1.3.6 Modalidad especial *84*
5.1.3.7 Modalidad de jóvenes y adultos *85*
5.1.4 Indicadores educativos *86*
5.1.4.1 Descripción de los indicadores educativos *86*
5.1.4.2 Codificación de estos 3 indicadores en MongoDB Atlas *89*
5.1.4.3 Indicadores educativos de primaria. *92*
5.1.4.4 Indicadores educativos de secundario. *93*
5.1.5 Gráficos estadísticos en Atlas *95*
5.2 Operativos ONE y Aprender *97*

5.2.1 Aprender primaria 2021-censal *97*

5.2.1.1 Aprender 2021. Matemática y Lengua. Resultados de estudiantes de Desempeños Avanzado y Satisfactorio *97*

5.2.1.2 Aprender 2021. Matemática. Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2021 y 2018 ..... *98*

5.2.1.3 Aprender 2021. Lengua. Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2021 y 2018. ..... *100*

5.2.1.4 Aprender 2021. Matemática. Diferencia de los desempeños avanzado y satisfactorio del sector estatal y del sector privado de Aprender 2021 y 2018. .... *100*

5.2.1.5 Aprender 2021. Matemática. Diferencia de los desempeños avanzado y satisfactorio del ámbito urbano y el ámbito rural de Aprender 2021 y 2018. ..... *101*

5.2.1.6 Aprender 2021. Matemática. Diferencia de los desempeños avanzado y satisfactorio del género varones y del género mujeres de Aprender 2021 y 2018. *102*

5.2.1.7 Aprender 2021. Matemática. Diferencia de los desempeños avanzado y satisfactorio del NSE bajo y NSE alto de Aprender 2021. ..... *103*

5.2.1.8 Aprender 2021. Matemática. NSE alto en 2018 y 2021, y cuánto cae del 2018 al 2021 y NSE bajo en 2018 y 2021, y cuánto cae del 2018 al 2021. ..... *104*

5.2.1.9 Aprender 2021. Matemática y Lengua. Promedio de desempeños avanzado y satisfactorio de relacionado el cociente de las sobreedades de sexto grado de 2021 y primero de 2016, con sus respectivas matriculas. *105*

5.2.2 Aprender secundaria 2022 ..... *107*

5.2.2.1 Aprender 2022. Matemática. Distribución de estudiantes por niveles de desempeño y suma de los niveles Avanzado y Satisfactorio *107*

5.2.2.1 Aprender 2022. Lengua. Distribución de estudiantes por niveles de desempeño y suma de los niveles Avanzado y Satisfactorio *107*

5.2.3 Aprender primaria 2023-censal *108*

5.2.3.1 Aprender 2023. Matemática. Distribución de estudiantes por nivel de desempeño y suma de los niveles Avanzado y Satisfactorio *108*

5.2.3.2 Aprender 2023. Matemática. Menor diferencia de la suma de desempeños avanzado y satisfactorio entre los NSE bajo y NSE alto. *109*

5.2.3.3 Aprender 2023. Matemática. Mayor crecimiento en la diferencia de los niveles de desempeño avanzado y satisfactorio entre las pruebas ONE 2013 y Aprender 2023. ..... *110*

5.3 Embeber el gráfico de Atlas en cualquier página web *111*

Capítulo 6 ..... *113*

Conclusiones y futuras investigaciones *113*

6.1 Conclusiones *114*

6.2 Futuras investigaciones *117*

Bibliografía..... *119*

Índice de Figuras

Figura 1. La plataforma de big data educativa que cubre el flujo de datos educativos desde la producción de datos hasta el análisis de datos, con un ciclo de retroalimentación. Fuente: Munshi et al. *22*

Figura 2. Extracto de datos provistos por argentina.gob.ar sobre el RA 2018. Fuente: Red Federal de Información Educativa (RedFIE). *27*

Figura 3. Sitio del Ministerio de Capital Humano en donde se encuentran las bases de datos. Fuente: Argentina.gob.ar. *32*

Figura 4. Bases de datos del RA 2023. Fuente: Argentina.gob.ar. *33*

Figura 5. Esquema de transformación de datos. Fuente: Ruchinsky et al. [21]. *34*

Figura 6. Transformaciones iniciales: campos pertenecientes a subgrupos. Fuente propia. *34*

Figura 7. Cabeceras de un sólo reglón, sin formar subgrupos. Fuente propia. *35*

Figura 8. Prueba de los nombres de cabecera. Fuente propia. *36*

Figura 9. Importación del padrón de establecimientos educativos con mongoimport. Fuente propia. *36*

Figura 10. Eliminación de registros totalmente nulos. Fuente propia. *37*

Figura 11. Los pasos necesarios para poder importar con MongoDB Compass. Fuente propia. *37*

Figura 12. El MongoDB Compass, actualmente detecta automáticamente el tipo de datos. Fuente propia. *38*

Figura 13. Se consignaba una "X" para indicar que una institución ofrecía una oferta determinada Fuente propia. *38*

Figura 14. Procesamiento de transformaciones en NodeJS. Fuente propia. *40*

Figura 15. Diferencias en 2018, entre la cantidad de documentos del padrón y la colección de matrícula y secciones. Fuente Ruchinsky et al. [22]. *40*

Figura 16. Control de totales con datos externos. Fuente propia y Anuario Estadístico Educativo 2018. ..... *42*

Figura 17. Esquema que relaciona las principales colecciones. Fuente propia. *43*

Figura 18. Listado de establecimientos y matrículas que tienen las 4 ofertas educativas. Fuente propia. *44*

Figura 19. Asignación de un CUE-anexo específico. Fuente propia. *45*

Figura 20. Consulta de una institución en cada una de las colecciones. Fuente propia. *46*

Figura 21. Consulta de establecimientos combinando colecciones. Fuente propia. *47*

Figura 22. Documentos de matrículas y secciones correspondientes a 13 años (18.3 k documentos). Fuente propia. *50*

Figura 23. Agrupamiento de documentos por año, sector y ámbitos de trayectoria (11.8k documentos). Fuente propia. *51*

Figura 24. Subdocumentos embebidos y vectores para cada variable, sexo y nivel. Fuente propia. *51*

Figura 25. Proyección y grabado en forma local. Fuente propia. *52*

Figura 26. Distintas agregaciones en modo local, tomando como clave el año anterior en la trayectoria. Fuente propia. *53*

Figura 27. Realización de distintos ajustes. Fuente propia. *54*

Figura 28. Control de cantidad de documentos en cada colección. Fuente propia. *54*

Figura 29. Consulta de claves (año, y sector) de tres colecciones resúmenes. Fuente propia. *55*

Figura 30. Matrícula total del sector privado 2018-Consulta en modo local. Fuente: DIIE y fuente propia. *56*

Figura 31. Control de matrícula de primer año de secundaria de 2018. Fuente propia. .. *57*

Figura 32. Combinación de colecciones por medio del mismo _id. Fuente propia. *57*

Figura 33. Relación entre localizaciones y unidades de servicios educativos del anuario 2022, con documentos bajados en bases de datos NoSQL y resultados procesados localmente. Fuente propia y anuario. *59*

Figura 34. Colección de datos de la modalidad especial por año y sector. Fuente: anuarios y fuente propia. *60*

Figura 35. Recorte de la información publicada con los sectores en un mismo cuadro. Fuente: Anuario 2017. ..... *61*

Figura 36. Estructura de documento MongoDB para Aprender 2021. Fuente propia. *62*

Figura 37. Ejemplo de un subdocumento de lengua por nivel socioeconómico. Fuente: Aprender 2021. ..... *62*

Figura 38. Estructura de un subdocumento de matemática, y dentro de este, la variable correspondiente a "resultados generales" de Aprender 2023. ..... *63*

Figura 39. API del Instituto Geográfico Nacional. Fuente: IGN. *65*

Figura 40. Consulta de latitud y longitud. Fuente: leaflet.js y propia. *65*

Figura 41. Utilización de íconos para referenciar y diferenciar establecimientos. Fuente propia. *66*

Figura 42. Carga de niveles diferenciada por colores en Leafletjs.com. Fuente propia. .. *66*

Figura 43. Ofertas educativas en las localidades del interior. Fuente propia. *67*

Figura 44. Al mover los datos, la latitud y longitud se vuelven alfanuméricas. Fuente propia. *67*

Figura 45. Exportación de datos preprocesados en modo local e importación en la nube: cloud.mongodb.com. Fuente propia. *70*

Figura 46. Estructura de matrícula, trayectoria, cargos y horas. Fuente propia. 71

Figura 47. Estructura de la variable promovidos, en subdocumentos en distintos niveles y vectores. Fuente propia. 71

Figura 48. Consulta de proximidad de establecimientos educativos cercanos a otro establecimiento geolocalizado. Fuente propia. 72

Figura 49. Mapa de establecimientos educativos de la provincia de Formosa. Fuente propia. 73

Figura 50. Mapa de establecimientos educativos filtrando el atributo localidad del documento. Fuente propia. 73

Figura 51. Mapa de establecimientos especificando un nivel específico de los documentos. Fuente propia. 74

Figura 52. Mapa de calor utilizando un criterio determinado. Fuente propia 74

Figura 53. Evolución de principales variables. Educación común y modalidades. Fuente: 76

Figura 54. Evolución de las principales variables de la educación común por año. Fuente: RA2011 a RA2023 y propia. 76

Figura 55. Principales variables del 2022 por departamento. Fuente: Anuario 2022..... 77

Figura 56. Diferencia de la sumatoria de unidades educativas por departamento en el resumen jurisdiccional del anuario 2022. Fuente: Anuario 2022. 78

Figura 57. Extracto de la evolución de las principales variables de la educación común por año y sector de gestión. Fuente: R2011 a RA2023 y propia..... 78

Figura 58. Facilidad de MongoDB Atlas para agregar un grupo al diseño del cuadro. Fuente propia. 79

Figura 59. Evolución de las principales variables de la educación común por año y ámbito. Fuente: RA2011 a 2023 y propia. 79

Figura 60. Evolución de principales variables del nivel inicial. Fuente: RA2011 a RA2023 y propia. 80

Figura 61. Nivel inicial, salas de 3, 4 y 5 años, y matrículas, por sector y año. Fuente: RA2011 a RA2023 y propia. 80

Figura 62. Evolución de las principales variables del nivel primario, por año. Fuente: RA2011 a RA2023 y propia. 81

Figura 63. Extracto de la evolución de las principales variables del nivel primario, por año y ámbito. Fuente: RA2011 a RA2023 y propia. 81

Figura 64. Análisis de evolución de matrículas y cantidad de repitentes por grado y año. Fuente: RA2011 a RA2023 y propia. 82

Figura 65. Evolución de las principales variables del nivel secundario de la educación común. Fuente: RA2011 a RA2023 y propia. 82

Figura 66. Evolución de matrícula y repitentes del nivel secundario de la educación común, por año de estudio y año lectivo. Fuente: RA2011 a RA2023 y propia. 83

Figura 67. Evolución de egresados del nivel secundario por orientación y año de egreso de la educación común. Fuente: RA2011 a RA2023 y propia. *83*

Figura 68. Evolución de las principales variables del nivel superior no universitario de la educación común, por año. Fuente: RA2011 a RA2023 y propia. *84*

Figura 69. Evolución de las principales variables de la modalidad especial por año. Fuente: RA2011 a RA2023 y propia. *85*

Figura 70. Evolución de las principales variables de la modalidad de jóvenes y adultos por año. Fuente: RA2011 a RA2023 y propia. *85*

Figura 71. Indicadores educativos del nivel primario: tasa de promoción efectiva, tasa de repitencia y tasa de abandono interanual por año. TPE por grado. Fuente: Indicadores educativos de Argentina.gob.ar y propia. *92*

Figura 72. Indicadores educativos del nivel secundario: tasa de promoción efectiva, tasa de repitencia y tasa de abandono interanual por año. TPE por año de estudio. Fuente: Indicadores educativos de Argentina.gob.ar y propia. *93*

Figura 73. Ajuste necesario en el cálculo de la tasa de promoción efectiva después de la pandemia COVID-19. Fuente propia. *94*

Figura 74. Tasa de promoción efectiva ajustada, considerando todos los promovidos. Fuente: Indicadores educativos de Argentina.gob.ar hasta 2021, y propia. *94*

Figura 75. Gráfico de matrícula por departamento y nivel del año 2022, de la educación común. Fuente: Anuario estadístico 2022 y propia. *95*

Figura 76. Matrícula de la educación común por nivel en 2022. Fuente: Anuario 2022 y propia. *96*

Figura 77. Composición de matrícula de la educación común por nivel y departamento en 2022. Fuente: Anuario estadístico 2022 y propia. *96*

Figura 78. Aprender Matemática y Lengua. Resultados de estudiantes de Desempeños Avanzado y Satisfactorio de Aprender 2021. Fuente: Aprender 2021 y propia. *98*

Figura 79. Aprender-Matemática. Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2021 y 2018: Distribución de estudiantes por nivel de desempeño: Avanzado y Satisfactorio. Ranking de jurisdicciones que mejoraron y que cayeron en el desempeño. Fuente: Aprender 2021 y propia. *99*

Figura 80. Aprender-Lengua. Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2021 y 2018. Distribución de estudiantes por nivel de desempeño: Avanzado y Satisfactorio. Ranking de jurisdicciones en la caída de desempeños entre ambos periodos. Fuente: Aprender 2021 y propia. *100*

Figura 81. Aprender-Matemática Sector de gestión. Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2021 y 2018: Distribución de estudiantes por nivel de desempeño según SECTOR DE GESTIÓN: Avanzado y Satisfactorio. Se resaltan jurisdicciones que mejoran en el sector estatal. Fuente Aprender 2021 y propia. *101*

Figura 82. Matemática - Ámbito. Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2021 y 2018: Distribución de

estudiantes por niveles de desempeños Avanzado y Satisfactorio según ÁMBITO. Ranking alfabético (en CABA y y Tierra del Fuego se consignó 0 en el ámbito rural). Fuente: Aprender 2021 y propia. *102*

Figura 83. Matemática - Género. Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2021 y 2018: Distribución de estudiantes por niveles de desempeños Avanzado y Satisfactorio según GÉNERO. Ranking alfabético. Fuente: Aprender 2021 y propia. *103*

Figura 84. Aprender-Matemática Niveles Socioeconómicos 2021. Diferencias entre porcentajes de estudiantes con desempeño Av. y Sat. de NSE bajo y alto de 2021: Distribución de estudiantes por niveles de desempeños Avanzado y Satisfactorio según NSE. Ranking diferencia entre NSE bajo menos NSE alto de 2021. Fuente: Aprender 2021 y propia. *104*

Figura 85. Aprender-Matemática_NSE alto y bajo por año. Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2021 y 2018: Aprender-Matemática_NSE alto y bajo por año. Diferencias entre porcentajes de estudiantes con desempeño avanzado y satisfactorio de 2021 y 2018. Fuente: Aprender 2021 y propia. *105*

Figura 86. Aprender 2021-Matemática y Lengua y la relación de matrículas de edades teóricas de 6to. RA2021 y de 1ro. RA2016. Distribución de estudiantes por niveles de desempeños Avanzado y Satisfactorio. Ranking según multiplicación relación de edades teóricas del RA y desempeño Aprender. Fuente: Aprender 2021, Relevamientos anuales y propia. *106*

Figura 87. Aprender 2022 Secundaria-Matemática. Distribución de estudiantes por nivel de desempeño y suma de los niveles Avanzado y Satisfactorio. Fuente: Aprender 2022 y propia. *107*

Figura 88. Aprender 2022 Secundaria-Lengua. Distribución de estudiantes por nivel de desempeño y suma de los niveles Avanzado y Satisfactorio. Fuente: Aprender 2022 y propia. *108*

Figura 89. Aprender 2023 Primaria-Matemática. Distribución de estudiantes por nivel de desempeño y suma de los niveles Avanzado y Satisfactorio. Ranking por nivel de desempeño. Fuente: Aprender 2023 y propia. *109*

Figura 90. Aprender 2023 Primaria-Matemática. Distribución de estudiantes por suma de los niveles de desempeño Avanzado y Satisfactorio de los Niveles socioeconómicos: bajo, medio y alto. Ranking de las jurisdicciones por menor diferencia entre los niveles socioeconómicos bajo y alto. Fuente: Aprender 2023 y propia. *110*

Figura 91. Aprender 2023 Primaria-Matemática. Distribución de estudiantes por suma de los niveles de desempeño Avanzado y Satisfactorio de los operativos ONE2013, Aprender 2016, 2018, 2021, 2022 y 2023. Ranking de las jurisdicciones que más crecieron entre las pruebas ONE2013 y Aprender 2023.Fuente: ONE13-Apreder16,18,21,22 y 23 y propia. *111*

Figura 92. Embeber un gráfico de MongoDB Atlas en una página web. Fuente propia. *111*

Índice de Scripts

Script 1. Ejemplo de un subdocumento en formato JSON. Fuente propia. *30*

Script 2. Prueba de importación de datos. Fuente propia. *35*

Script 3. Transformación de un campo alfabético a un tipo de dato numérico. Fuente propia. *39*

Script 4. Transformación de un campo alfabético a numérico, en NodeJS. Fuente propia. *39*

Script 5. Actualización de un CUE_anexo en la colección de matrículas y secciones. Fuente propia. *44*

Script 6. Asignación de CUE_anexo a un documento de la colección de matrícula y secciones. Fuente propia *45*

Script 7. Combinación de las colecciones padrón y bu2mys2018. Fuente propia *46*

Script 8. Ejemplo de un subdocumento JSON de lengua de un operativo Aprender 2022. Fuente propia. *63*

Script 9. Transformación de latitud a un formato de doble precisión. Fuente propia. *68*

Script 10. Transformación de longitud a un formato de doble precisión. Fuente propia. ... *68*

Script 11. Creación del campo de coordenadas. Fuente propia. *68*

Script 12. Consulta de establecimientos cercanos a un punto. Fuente propia *72*

Script 13. Creación de campos calculados en el gráfico para totalizar las distintas variables. Fuente propia. *75*

Script 14. Agregación para el cálculo de los indicadores. Fuente propia. *92*

Script 15. Campo calculado: obtener desempeño avanzado en Matemática de 2018. Fuente propia. *97*

Script 16. Campo calculado: obtener desempeño avanzado y satisfactorios en Matemática de 2018. Fuente propia. *97*

Script 17. Campo calculado: diferencia en Matemática entre Aprender 2021 y Aprender 2018. Fuente propia. *97*

Script 18. Campo calculado de la diferencia en matemática del sector estatal, entre los años 2021 y 2018. Fuente propia. *101*

Script 19. Campo calculado de la diferencia del desempeño en matemática del ámbito urbano entre 2021 y 2018. Fuente propia. *101*

Script 20. Campo calculado de la diferencia desempeño de avanzado y satisfactorio del género mujeres de matemática entre 2021 y 2018. ..... *102*

Script 21. Campo calculado de la diferencia en matemática de los NSE alto y bajo de 2021. Fuente propia. *103*

Script 22. Campo calculado de la diferencia entre matemática del NSE bajo de 2021 y NSE bajo de 2018. Fuente propia. 104

Script 23. Campo calculado que relaciona el promedio de desempeño de lengua y matemática de Aprender 2021, relacionándolos con la sobreedad y matrícula del RA2021 y RA2016. Fuente propia..... 105

Índice de Ecuaciones

Ecuación 1. Tasa de promoción efectiva por grado/año. 86
Ecuación 2. Tasa de promoción efectiva de cada nivel. 87
Ecuación 3. Tasa de repitencia de un grado/año. 87
Ecuación 4. Tasa de repitencia de un nivel. 88
Ecuación 5. Tasa de abandono interanual de un grado/año. 88
Ecuación 6. Tasa de abandono interanual de un nivel. 89

Capítulo 1

Introducción

1.1 Introducción

El sistema educativo argentino es complejo y tratar de analizarlo y entenderlo [1], nos ayudará a implementar mejores políticas públicas en educación, de tal forma que nos den un norte y que nos indiquen cómo optimizar los recursos [2] y en donde poner más énfasis, en un marco de calidad educativa. En la actualidad existen 2 grandes operativos, que indagan sobre la realidad educativa argentina [3]: a) los denominados *Relevamientos Anuales* (RA), o simplemente *operativos RA*, que indagan sobre datos estadísticos de los 4 niveles y sus modalidades, de todas las jurisdicciones provinciales, y que forman una base de datos nacional, y b) los operativos que forman parte del dispositivo nacional de evaluación de los aprendizajes de los estudiantes y de sistematización de información acerca de algunas condiciones en las que este proceso se desarrolla, tales como el operativo ONE (Operativo Nacional de Evaluación) 2013 y los operativos Aprender.

Los relevamientos anuales (RA), constituyen un operativo educativo de carácter censal, cuya unidad de relevamiento y análisis son las unidades educativas. Dichos operativos recogen al 30 de abril de cada año la información consolidada a nivel nacional sobre las principales variables del sistema educativo. Están conformados por dos sistemas, uno denominado "Padrón" de establecimientos educativos, con datos que cambian poco a lo largo del tiempo, y el otro, denominado "Relevamiento Anual" que indaga sobre distintas variables de cada año lectivo, en cada escuela. En la actualidad, se culminó con el Relevamiento Anual de información estadística del año 2023 [4], aclarando que el RA 2024 está en proceso de carga y verificación de la información. Estos operativos RA se desarrollan a nivel federal, implementados a nivel jurisdiccional (en cada provincia) y conformados por unos 10 cuadernillos en formato de revista para escuelas que no tienen Internet en el país, o en formato de formularios en páginas web, para aquellas que sí tienen:

- Los datos se ingresan por cada escuela por medio de páginas web, y para aquellas escuelas que no tienen Internet en el país, se rellenan los cuadernillos en forma manual, y luego se cargan por un organismo central de cada provincia. En cualquiera de los casos, cada escuela tiene que proporcionar datos sobre 50, o más, cuadros estadísticos.
- Los sistemas antes mencionados, Padrón y RA, almacenan los datos en bases de datos PostgreSQL (bases de datos relacionales), y generalmente, la implementación en cada jurisdicción consta de entre 100 a 400 tablas, y a veces más.

- El Ministerio de Capital Humano-Secretaría de Educación (ex Ministerio de Educación Nacional) comparte estos datos y publica las bases de datos a nivel de establecimiento educativo en forma completa, aunque con uno a dos años de atraso [5].
- La información publicada está contenida sólo en 8 tablas, en grandes archivos planos tipo planillas de cálculo Excel, produciendo una gran cantidad de datos redundantes y datos nulos, por lo que, para organizar y almacenar los datos abiertos del sistema educativo argentino se necesita hacer una limpieza de los mismos. La redundancia se produce ya que existen instituciones con los 4 niveles educativos (inicial, primario, secundario y superior), y muchas instituciones tienen un solo nivel.
- Los datos compartidos por el Ministerio de Capital Humano están en forma desagregada y compleja:
    - No todas las jurisdicciones tienen el mismo esquema educativo en los niveles primario y secundario, ya que algunas tienen 7 años de primaria y 5 de secundaria, y otras, 6 años de primaria y 6 años de secundaria. En Formosa, un ciclo educativo completo, se integra con 6 años de primaria y 6 de secundaria.
    - Aparentemente tienen el mismo formato, pero conviene recordar que son más de 60.000 unidades educativas en el país, y los datos son cargados en distintos equipos con distintos sistemas operativos en las diferentes jurisdicciones. De una base de datos PostgreSQL constituida por cientos de tablas, el Ministerio de Capital Humano lo pasa a una planilla de cálculo. Esta herramienta utilizada para la publicación, que es de fácil manejo, muy conocida por todos y que en la actualidad tiene mucha potencia para filtrar información, no tiene tanta potencia para manejar todos los sistemas operativos, razón por la cual, es posible que un valor numérico tenga un formato alfanumérico.
    - Con respecto al formato empleado para indicar las horas, algunas tablas tienen porciones de horas enteras, y otras, horas decimales.
    - El identificador (id) también cambia de longitud a lo largo de los 12 años: empieza con 10 dígitos en 2011, y en 2022 tiene 15 dígitos.

Con respecto a los operativos ONE 2013 y Aprender, correspondientes a distintos períodos, el Ministerio de Capital Humano compartió todas las bases de datos, faltando a la fecha del presente trabajo, las bases de datos de ONE 2013 y las bases de datos de Aprender 2022 y

2023, razón por la que se cargaron en formato JSON (JavaScript Object Notation) los informes finales de las distintas jurisdicciones y luego se procesaron.

1.2 Estado del arte.

Se están produciendo enormes cantidades de datos educativos y un desafío común al que se enfrentan muchas organizaciones educativas es encontrar un método eficaz para aprovechar y analizar estos datos para ofrecer una educación mejorada de forma continua. Hoy en día, los datos educativos están evolucionando y se han vuelto grandes en volumen, amplios en variedad y de alta velocidad. Estos datos producidos deben manejarse de manera eficiente para extraer valor y tomar decisiones informadas. Un trabajo publicado por Munshi et al., en 2021 [6], toma dichos datos como un desafío de big data y presenta una plataforma integral diseñada para realizar aplicaciones analíticas de big data educativas. Se destacan las etapas de implementación de la plataforma de big data educativa en una plataforma de computación en la nube y la organización de los datos educativos en una arquitectura de lago de datos.

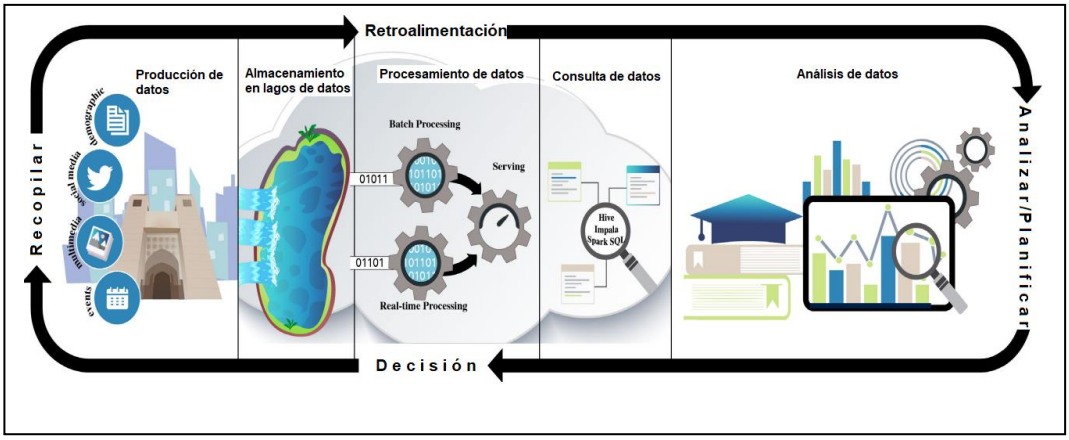


Figura 1. La plataforma de big data educativa que cubre el flujo de datos educativos desde la producción de datos hasta el análisis de datos, con un ciclo de retroalimentación. Fuente: Munshi et al.

Los conjuntos de datos educativos pueden considerarse desde dos aspectos: a) conjuntos de datos directamente relacionados con información educativa que contienen recursos educativos, datos institucionales e indicadores educativos; y b) conjuntos de datos de diferentes dominios que pueden utilizarse en entornos educativos. Una de las plataformas ocupadas para procesar la analítica y el big data educativo, es Hadoop, que consiste en un conjunto de tecnologías, en las que se destacan Hbase, que es un sistema de base de datos que proporciona un mecanismo para el almacenamiento y recuperación de datos con menos restricciones que las bases de datos SQL tradicionales [7].

En un sistema que procesa el rendimiento académico de alumnos de cursos MOOC (Massive

Online Open Courses), en Países Bajos [8], a través de paneles de análisis de aprendizaje. Este sistema calcula un conjunto de indicadores que se muestran por medio de pequeñas aplicaciones en la pantalla (widget) y posteriormente se exportan datos de investigación de Coursera en más de 100 tablas con información como contenido del curso, datos demográficos de los estudiantes, interacción de los estudiantes con cada parte del contenido del curso y datos del foro. Coursera procesa las exportaciones de investigación al final de cada día y están listas para su descarga solo después de un retraso. El procesamiento de datos comienza con una solicitud automatizada diaria de la exportación de datos del día anterior. Una vez descargada, la exportación de datos se procesa con Python para calcular los indicadores de cada alumno que se almacenan en una instancia de MongoDB.

1.3 Objetivos

El objetivo del presente trabajo consiste en extraer datos educativos abiertos que se encuentran en grandes conjuntos de datos (datasets), gestionarlos y presentarlos utilizando bases de datos NoSQL, con el fin de facilitar su posterior procesamiento, visualización y toma de decisiones, haciendo foco en la jurisdicción de la provincia de Formosa durante el último ciclo educativo completo que se tiene información (2011-2023), para poner a disposición de personas u Organismos interesados.

Como objetivos específicos se persigue:

- Buscar, extraer y limpiar datos disponibles en conjuntos de datos oficiales, que están en distintos formatos, incluso con datos nulos y redundantes.
- Proponer técnicas de análisis de datos para evaluar la información disponible, buscando correlacionar con la situación educativa actual a fin de profundizar el entendimiento de diversas situaciones del ámbito educativo.
- Correlacionar la información en diversos grupos, asociándose por categorías tales como departamentos, geolocalización de grandes instituciones educativas, niveles del sistema educativo (inicial, primario, secundario, superior), género de los estudiantes, entre otros.
- Facilitar el acceso a los datos, propiciando favorecer la disponibilidad y socialización de estos.

Capítulo 2

Marco teórico

2.1 Caracterización del Sistema Educativo Argentino

En Argentina, en las últimas décadas, se evidencia el crecimiento de la escolarización y una prolongación del período escolar obligatorio. Como resultado de esto, los estudiantes han logrado obtener mayores niveles educativos que en períodos anteriores. Asimismo, la cobertura del sistema educativo también experimentó una expansión, la cual tuvo que ver con distintos factores intervinientes, entre otros, la mencionada ampliación de la obligatoriedad, la expansión de la oferta educativa y distintas acciones orientadas a fomentar la inclusión educativa y a estimular el incremento de la escolarización [9]. Sin embargo, el desempeño académico se ve afectado por la existencia de muchos factores y desigualdades que deben ser considerados en el marco de la calidad educativa. Disponer de esta información es fundamental para poder orientar políticas educativas y superar los déficits.

La Ley de Educación Nacional 26.206 [10], establece que el Sistema Educativo Nacional está formado por un conjunto organizado de servicios y acciones regulados por el Estado, que posibilitan el ejercicio del derecho a la educación.

Este sistema se estructura en cuatro *niveles*: educación inicial (de 45 días a cinco años), primaria (a partir de los 6 años), secundaria (aquellos que completaron la primaria) y superior. La obligatoriedad escolar en todo el país se extiende desde la edad de cinco (5) años hasta la finalización de la educación secundaria. En particular, la educación superior comprende universidades e institutos universitarios, e institutos de educación superior, tanto en el estamento nacional, como en los estamentos provinciales.
Además, existen *modalidades.* Éstas son opciones organizativas o curriculares, que responden a necesidades específicas de formación, tales como educación técnico profesional, artística, intercultural bilingüe, especial, permanente para jóvenes y adultos, rural, domiciliaria, hospitalaria, y en contextos de privación de libertad. Si bien la educación es pública, puede ser impartida por un determinado *sector* de gestión: estatal o privado, y en un determinado *ámbito*, que en forma básica la podemos calificar en urbano o rural.

La educación se imparte en *establecimientos* educativos. Éstos pueden tener uno o más niveles u ofertas educativas, y pueden enfocarse en una modalidad o incluir varias. Sin embargo, siempre corresponde a un sólo sector de gestión y ámbito educativo. Cada establecimiento se identifica por el Código Único de Establecimiento (CUE). Éste consta de siete dígitos, donde los dos primeros corresponden al código de la provincia y el resto a un número adjudicado correlativamente a cada establecimiento de la jurisdicción, a medida que

se van creando las instituciones [11]. Adicionalmente, se agrega un código de 2 dígitos, correspondiente al anexo: generalmente la sede del establecimiento termina con 00, y a partir de ahí, empiezan los anexos (01, 02, etc.).

El funcionamiento y la estructura del sistema educativo se caracterizan siguiendo diversos parámetros, los cuales pueden agruparse en:

- Estructurales: instituciones por jurisdicción, sus niveles, modalidades, sector de gestión y ámbito.
- Datos Estadísticos de Alumnos: la *matrícula* (cantidad de alumnos inscriptos y aceptados), cantidad por género, repitentes, número de promovidos (aquellos que aprueban un año completo), egresados (los que completan un nivel) y la matrícula de alumnos con sobreedad.
- Datos Estadísticos de Docentes: la cantidad de docentes, la cantidad de cargos cubiertos y vacantes, la cantidad de horas cubiertas y las secciones.
- Otros Indicadores Estadísticos: como ser tasa de promoción efectiva, tasa de repitencia, tasa de abandono interanual, entre otros.

2.2 Caracterización de datos abiertos del Sistema Educativo Argentino

Los datos educativos abiertos (DEA) son conjuntos de datos sobre el sistema educativo que se publican en formatos electrónicos accesibles y gratuitos para cualquier persona. Estos datos pueden incluir información sobre unidades educativas, alumnos, cargos, horas, etc. Los datos abiertos del sistema educativo argentino presentan un problema adicional: poseen un formato inconsistente, que no puede ser empleado de forma directa para su posterior análisis. Esto se debe a que los datos provistos por “argentina.gob.ar” acerca de los niveles y modalidades de las escuelas se encuentran presentados como columnas individuales, donde una celda con una cruz indica la pertenencia del establecimiento.

Fuente: Relevamiento Anual 2018. DiNIEE. Ministerio de Educación , Cultural, Ciencia y Tecnología de la Nación.

cue	Provincia	Sector	Ámbito	Oferta						
				Inicial		Primaria		Secundaria		
				Maternal	Infantes	rimara de	rimara de	Ciclo básico Req	Ciclo básico Req	Completa Req6
420000707	Buenos Aires	Estatal	Urbano		X					
420007007	Buenos Aires	Privado	Urbano			X				
420700007	Buenos Aires	Privado	Urbano		X					
427000007	Buenos Aires	Estatal	Urbano			X				
480001608	Buenos Aires	Estatal	Urbano			X				
480016008	Buenos Aires	Estatal	Urbano			X				
480088008	Buenos Aires	Estatal	Urbano							X

Figura 2. Extracto de datos provistos por argentina.gob.ar sobre el RA 2018. Fuente: Red Federal de Información Educativa (RedFIE).

En la Figura 2 se ejemplifica este formato. Debido a la extensión del archivo original se presenta un extracto del mismo.

Cada uno de los niveles de educación se subdivide en distintos tipos. Así, se obtiene: *inicial* (maternal o infantes), *primaria* (con seis grados o con siete), *secundaria* (ciclo básico de seis o siete grados, completo de seis o siete, o polimodal), junto con otros tipos no agrupados. Si bien en las hojas de cálculo esto no es problema, mantener esta representación consume memoria adicional, y genera dificultades de operación.

2.3 Necesidad de una base de datos NoSQL

Al tener las distintas jurisdicciones distintas estructuras y tipologías de datos, surge la necesidad de contar con una base de datos más potente en el manejo de los datos, y que soporte en un mismo campo autorreferenciado, distintos tipos de datos (aunque todos los datos parezcan ser del mismo tipo). Por eso, una alternativa positiva son las bases de datos NoSQL («Not Only SQL»), que pueden manejar grandes cantidades de datos desestructurados y cambiantes. Una de las ventajas más importantes de las bases de datos NoSQL es la de lograr una mayor flexibilidad al soportar datos semiestructurados y autodescriptivos [12], característica fundamental para su elección para este trabajo de tesis, ya que en las distintas jurisdicciones manejan conceptos similares, pero varían sus tipos de datos y sus representaciones de datos. Los sistemas NoSQL basados en documentos u orientados a documentos, suelen almacenar datos como colecciones de documentos similares. Estos tipos de sistemas también se conocen como almacenes de documentos. La tecnología NoSQL constituye una base de datos moderna que está diseñada para proporcionar escalabilidad y para admitir datos voluminosos, lo que llevó al surgimiento de NoSQL como la solución más viable de bases de datos [13].

Como resultado, las bases de datos NoSQL son ampliamente aceptadas en la ciencia de datos debido a su escalabilidad, facilidad de trabajo en la nube, y aptitud para el desarrollo ágil de aplicaciones [14] [15] . En particular, la ciencia de datos es un campo interdisciplinario que usa métodos científicos, procesos, algoritmos y sistemas para extraer conocimiento e inferencias de datos en distintos formatos, ya sea estructurados o no [16]. Es un concepto que unifica estadística, análisis de datos, aprendizaje automático y métodos relacionados para comprender un problema real mediante los datos disponibles. Esto se hace a través de técnicas y teorías extraídas de múltiples campos, como ser las matemáticas, estadística, sistemas de información y ciencias de la computación. Las organizaciones modernas están

inundadas de datos; hay una proliferación de dispositivos que pueden recopilar y almacenar información de manera automática [17]. Ciertos autores consideran a la ciencia de datos como un nuevo paradigma de la ciencia, afirmando que este cambio se debe a la rápida generación y procesamiento eficiente de datos y al impacto de la tecnología de la información.

2.3.1 La utilización de MongoDB

MongoDB es una base de datos NoSQL que almacena información en documentos. En lugar de almacenar los datos en filas y columnas fijas, las bases de datos de documentos utilizan documentos flexibles. Los documentos almacenan datos en pares campo-valor. Los valores pueden ser diversos tipos de estructuras, incluidas cadenas, números, fechas, matrices u objetos. Los documentos se pueden almacenar en formatos JSON, BSON y XML. Un grupo de documentos conforma una colección, donde cada documento puede tener contenido similar. No es necesario que todos los documentos de una colección tengan los mismos campos, debido a que las bases de datos de documentos tienen un esquema flexible.

Tres factores claves diferencian las bases de datos de documentos de las bases de datos relacionales:

a) La **intuición del modelo de datos**, ya que no es necesario descomponer los datos en tablas, ni ejecutar uniones costosas. Los datos a los que se accede en conjunto se almacenan juntos, por lo que tienen menos código al escribir.

b) La **ubicuidad de los documentos JSON**: los documentos JSON son livianos, independientes del lenguaje y legible por humanos.

c) La **flexibilidad del esquema**: el esquema de un documento es dinámico y se autodescribe, por lo que no se necesita predefinir previamente la estructura de la base de datos. La estructura se puede cambiar, evitando migraciones disruptivas de esquemas.

2.3.2 ¿Qué es JSON: JavaScript Object Notation?

JSON [18], o notación de objetos JavaScript, es un formato de intercambio de datos legible por humanos, especificado a principios de la década de 2000. Aunque se basa en un subconjunto del estándar del lenguaje de programación JavaScript, es completamente independiente del lenguaje. Los objetos JSON son contenedores asociativos, en los que una clave de cadena se asigna a un valor. Los objetos JSON son fáciles de entender para los humanos y para que las computadoras los analicen y generen. Ejemplo de subdocumento JSON:

```
"matematica": {"resultados_grales_anios": {
     "anio_2013": [1.3, 15.6, 25, 58.1],
     "anio_2016": [1.6, 15,28.6,54.8],
     "anio_2017": [1.8,17.9,28.4,51.9],
     "anio_2019": [ 0.6,15.7, 27.5, 56.2],
     "anio_2022": [ 0,10.2,26.8,63]},
  "sector": {"estatal": {   "anio_2013": [ 0.5,11.5,24.6,63.4],
       "anio_2016": [ 0.8,
```

Script 1. Ejemplo de un subdocumento en formato JSON. Fuente propia.

En el Script 1, se observa un subdocumento “matematica” de una jurisdicción de un operativo Aprender, que está compuesto por lo menos, por dos subdocumentos: uno denominado “resultados_generales_anios” y otro “sector”. El primer subdocumento “resultados_generales_anios”, está compuesto por 5 vectores: anio_2013, anio_2016, anio_2017, anio_2019 y anio_2022. El segundo subdocumento que se observa, denominado “sector”, y se observa el inicio del subdocumento correspondiente al sector “estatal”. Este último, está conformado por vectores.

2.4 Metodología del trabajo

Para el desarrollo de este Trabajo Final de Maestría se empleó una metodología conformada por 4 fases, donde cada una de ellas se realizó de manera iterativa, hasta obtener la información necesaria. Las fases de la metodología son las siguientes:

1) Estudio y caracterización de los elementos, parámetros y agrupaciones que conforman el sistema educativo argentino. El capítulo 2 se corresponde con esta fase.
2) Generación del modelo de datos NoSQL. Esta fase contempló la limpieza [19] y organización de los datos de la fase 1, a fin de corregir datos erróneos o fallas en los mismos. Se buscó la estandarización [20] de datos, la no duplicación de éstos y la eliminación de registros nulos. Se generó el esquema de BD NoSQL y la posterior migración de los datos, desde las diferentes fuentes hacia la base de datos. El capítulo 3 se corresponde con esta fase.
3) Validación de los datos cargados. Esta fase contempló el completamiento de los datos de la BD NoSQL, su preprocesamiento en modo local a fin de validar los resultados obtenidos con los datos de origen y su exportación hacia Atlas MongoDB para su procesamiento en la nube. El capítulo 4 se corresponde con esta fase.
4) Análisis de los datos y discusión de resultados, para un mejor enfoque del sistema educativo de la jurisdicción de Formosa. El capítulo 5 se corresponde con esta fase.

Capítulo 3

Construcción de la base de datos en MongoDB

3.1 Recolección y limpieza de los datos

Tanto el padrón de establecimientos educativos, como las bases de datos de los operativos RA, se descargan de:
https://www.argentina.gob.ar/educacion/evaluacion-informacion-educativa/bdd, cómo se observa en la Figura 3.

Figura 3. Sitio del Ministerio de Capital Humano en donde se encuentran las bases de datos. Fuente: Argentina.gob.ar.

Si bien, al momento de finalizar la revisión final de este trabajo de tesis (agosto de 2024) se terminó de recabar información del año 2023, los datos publicados corresponden al año 2023, como se observa en la Figura 4.

Con respecto al Padrón de Establecimientos Educativos, éste se puede descargar desde distintos sitios, como ser:
https://www.argentina.gob.ar/educacion/evaluacion-e-informacion-educativa/padron-oficial-de-establecimientos-educativos, actualizado al 15/07/2024 y también desde https://datos.gob.ar/dataset/educacion-padron-oficial-establecimientos-educativos, con datos a Setiembre de 2019

Como ya fue mencionado, estas bases de datos públicas tienen un formato de planilla de cálculo, tipo Excel, con mucha información redundante, con muchos casilleros nulos y con muchos campos que poseen distintos tipos de datos.

A continuación, se observa en la Figura 4, el conjunto de datos del Relevamiento Anual 2023.

Figura 4. Bases de datos del RA 2023. Fuente: Argentina.gob.ar.

Resulta necesario aclarar que todas las colecciones (tablas) en cada año, se relacionan por un campo ID, es decir, figura el mismo ID para cada institución en todas las colecciones del mismo año, o por conjunto de años donde gobierna el mismo partido político en la Argentina. Cada cierta cantidad de años, el ID cambia su longitud y composición.

En el padrón de establecimientos educativos, los establecimientos educativos de todo el país y sus anexos, de gestión estatal y de gestión privada, se identifican a través de una “Clave única de establecimientos y anexos” (CUE-anexo) generada por el Ministerio de Educación de cada jurisdicción. Lo que no es público, es la relación del CUE-anexo del padrón de establecimientos educativos con esos ID.

3.1.1 Transformaciones iniciales

Como parte del proceso de unificación, se convirtieron los niveles de educación, y sus tipos, a variables *categóricas*. Éstas toman un valor dentro de un conjunto finito determinado, representando categorías nominales. Con esto, es posible crear una variable categórica por cada nivel, facilitando así su posterior análisis. No obstante, esta conversión requiere considerar dos características clave:

- Un registro puede tener múltiples categorías para el mismo nivel.

- Dado que un registro puede no tener oferta en un nivel (por ejemplo, ser sólo una escuela primaria), continúan existiendo valores nulos que deben manejarse.

El esquema de transformación de datos se visualiza en la Figura 5.

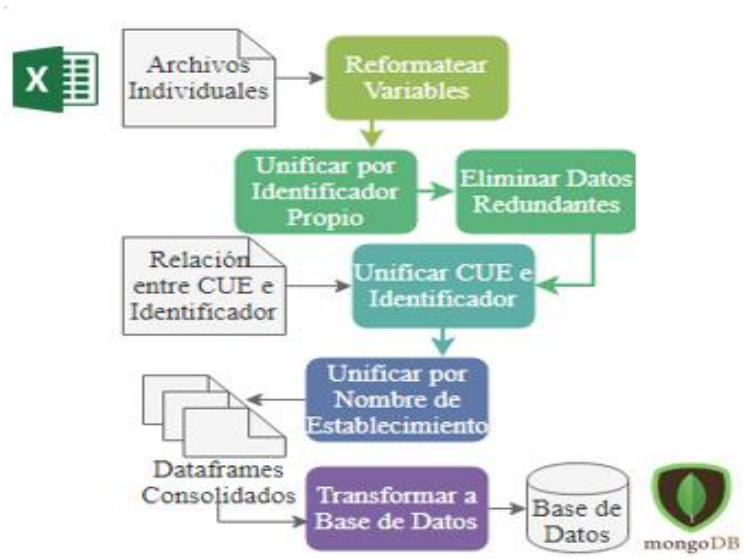


Figura 5. Esquema de transformación de datos. Fuente: Ruchinsky et al. **[21]**.

Además, en Excel, para convertir a tablas, se debe eliminar información innecesaria como los títulos, y homogeneizar los nombres de las columnas, en aquellos casos en que el nombre está detallado en varios renglones conformando subgrupos, como se observa en la Figura 6,y acotarlo en un sólo renglón.

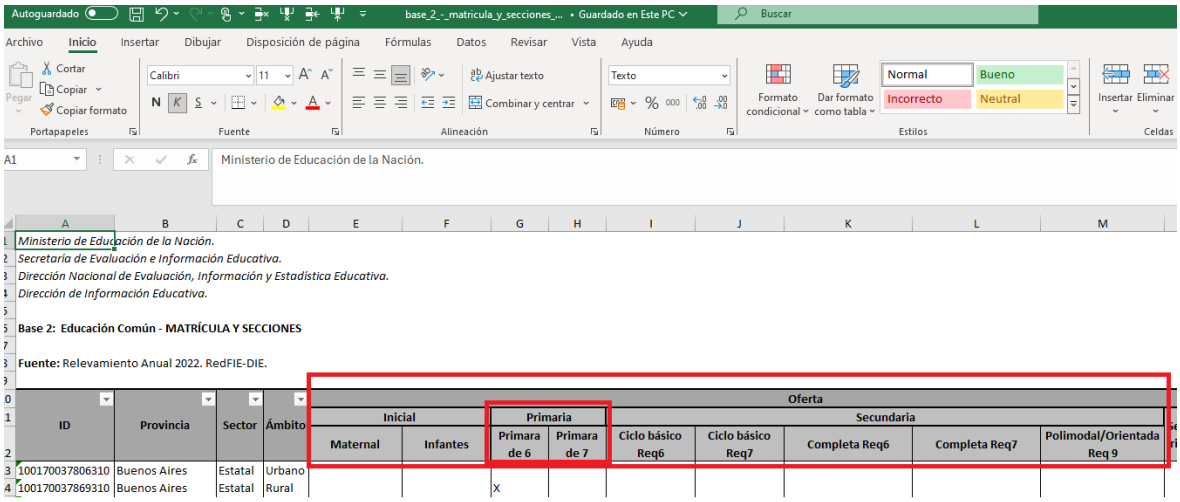


Figura 6. Transformaciones iniciales: campos pertenecientes a subgrupos. Fuente propia.

En la Figura 7 se observa la misma tabla, pero con cabeceras de un sólo renglón.

A	B	C	D	E	F	G	H	I	J	K	L
id	provincia	sector	ambito	inicial_mate	inicial_infan	primaria6	primaria7	secundaria_cb6	secundaria_cb7	secundaria_completa6	secundaria_completa7
1,00E+14	Buenos Aires	Estatal	Urbano								
1,00E+14	Buenos Aires	Estatal	Rural			X					
1,00E+14	Buenos Aires	Estatal	Urbano	X	X						
1,00E+14	Buenos Aires	Estatal	Urbano			X					
1,00E+14	Buenos Aires	Estatal	Rural	X	X						
1,00E+14	Buenos Aires	Privado	Urbano			X					

Figura 7. Cabeceras de un sólo reglón, sin formar subgrupos. Fuente propia.

Posteriormente, se graba la tabla en formato de texto, con extensión CSV, para que MongoDB la pueda importar.

En MongoDB se debe probar que los nombres de cabecera coincidan con la estructura de datos que se pretende importar, como se observa en el Script 2. Para eso, todos los campos se definieron de tipo string:

```
mongoimport --db dea --collection padron --type tsv --columnsHaveTypes
--
fields="prov.string(),_id.string(),nombre.string(),sector.string(),amb
ito.string(),domicilio.string(),cp.string(),tel_area.string(),tel_nro.
string(),cod_loca.string(),localidad.string(),departamento.string(),ma
il.string(),ed_comun.string(),ed_especial.string(),ed_adultos.string()
,ed_artistica.string(),ed_hospitalaria.string(),ed_eib.string(),ed_con
texto_e.string(),ec_jar_mat.string(),ec_ini.string(),ec_pri.string(),e
c_sec.string(),ec_sec_inet.string(),ec_snu.string(),ec_snu_inet.string
(),ea_sec.string(),ea_snu.string(),ea_snu_cursos.string(),ea_cyt.strin
g(),ee_edu_tem.string(),ee_ini.string(),ee_pri.string(),ee_sec.string(
),ee_int.string(),eja_pri.string(),eja_egb3.string(),eja_sec.string(),
eja_alf.string(),eja_fp.string(),eja_fp_inet.string(),ehd_ini.string()
,ehd_pri.string(),ehd_sec.string(),serv_comp.string()"          --file
c:/data/db/dea/cabecera_p_20190916_b.txt -ignoreBlanks
```

Script 2. Prueba de importación de datos. Fuente propia.

Y debe haber una coincidencia con los nombres de campos y la cabecera que se está importando, como se observa en la Figura 8.

```
Símbolo del sistema - mongo
> show collections
padron
> db.padron.find().pretty()
{
        "_id" : "CUE_anexo",
        "prov" : "jurisdicci⊡n",
        "cue" : "CUE",
        "nombre" : "nombre",
        "latitud" : "latitud",
        "longitud" : "longitud",
        "sector" : "Sector",
        "ambito" : "ambito",
        "departamento" : "departamento",
        "cod_depar" : "cod_depar",
        "localidad" : "localidad",
        "cod_loca" : "cod_loca",
        "domicilio" : "domicilio",
        "cod_postal" : "cod_postal",
        "tel_cod_area" : "cod_area",
        "tel_nro" : "telefono",
        "email" : "email",
        "ed_comun" : "ed_comun",
        "ed_especial" : "ed_especial",
        "ed_adultos" : "ed_adultos",
        "ed_artistica" : "ed_artistica",
        "ed_hospitalaria" : "ed_hospitalaria",
        "ed_eib" : "EIB",
        "ed_contexto_e" : "ed_contexto_e",
        "ec_jar_mat" : "ec_jardin_mat",
        "ec_ini" : "ec_inicial",
```

Figura 8. Prueba de los nombres de cabecera. Fuente propia.

Una vez comprobada la coincidencia entre la cabecera y la estructura que se quiere importar, se elimina la cabecera (db.padron.drop()), se cambian los campos a sus tipos de datos reales, y se procede a importar los datos reales. El parámetro ignoreBlanks, impide importar los datos que no están definidos, pero sí importa aquellos valores definidos como espacios “ ” y los que tienen un valor cero, importando los 63.583 documentos pertenecientes a establecimientos educativos del país a Octubre de 2020, como se observa en la Figura 9.

```
C:\Program Files\MongoDB\Server\4.4\bin>
C:\Program Files\MongoDB\Server\4.4\bin>mongoimport --db dea --collection padron --type tsv --columnsHaveTypes --fields="_id.string(),prov.string(),cue.string(),nombre.
string(),latitud.string(),longitud.string(),sector.string(),ambito.string(),departamento.string(),cod_depar.string(),localidad.string(),cod_loca.string(),domicilio.stri
ng(),cod_postal.string(),tel_cod_area.string(),tel_nro.string(),email.string(),ed_comun.string(),ed_especial.string(),ed_adultos.string(),ed_artistica.string(),ed_hospi
talaria.string(),ed_eib.string(),ed_contexto_e.string(),ec_jar_mat.string(),ec_ini.string(),ec_pri.string(),ec_sec.string(),ec_sec_inet.string(),ec_snu.string(),ec_snu_
inet.string(),ee_edu_tem.string(),ee_ini.string(),ee_pri.string(),ee_sec.string(),ee_int.string(),eja_pri.string(),eja_egb3.string(),eja_sec.string(),eja_alf.string(),e
ja_fp.string(),eja_fp_inet.string(),ea_sec.string(),ea_snu.string(),ea_cyt.string(),ehd_ini.string(),ehd_pri.string(),ehd_sec.string(),serv_comp.string(),latiuno.int32(
),latidos.int32(),latitres.int32(),laticuatro.double()"  --file c:/data/db/0padron/padron20200630c.txt --ignoreBlanks
2020-10-03T08:13:08.430-0300    connected to: mongodb://localhost/
2020-10-03T08:13:11.453-0300    [........................] dea.padron   128KB/14.1MB (0.9%)
2020-10-03T08:13:14.430-0300    [........................] dea.padron   256KB/14.1MB (1.8%)
2020-10-03T08:13:17.436-0300    [........................] dea.padron   256KB/14.1MB (1.8%)
2020-10-03T08:13:20.431-0300    [........................] dea.padron   256KB/14.1MB (1.8%)
2020-10-03T08:13:23.440-0300    [........................] dea.padron   256KB/14.1MB (1.8%)
2020-10-03T08:13:26.435-0300    [........................] dea.padron   256KB/14.1MB (1.8%)
2020-10-03T08:13:29.432-0300    [........................] dea.padron   256KB/14.1MB (1.8%)
2020-10-03T08:13:32.440-0300    [###.....................] dea.padron   1.91MB/14.1MB (13.6%)
2020-10-03T08:13:35.432-0300    [##############..........] dea.padron   8.36MB/14.1MB (59.4%)
2020-10-03T08:13:38.431-0300    [####################....] dea.padron   11.8MB/14.1MB (84.0%)
2020-10-03T08:13:41.431-0300    [#######################.] dea.padron   13.9MB/14.1MB (99.1%)
2020-10-03T08:13:41.576-0300    [########################] dea.padron   14.1MB/14.1MB (100.0%)
2020-10-03T08:13:41.834-0300    63583 document(s) imported successfully. 0 document(s) failed to import.

C:\Program Files\MongoDB\Server\4.4\bin>_
```

Figura 9. Importación del padrón de establecimientos educativos con mongoimport. Fuente propia.

3.1.2 Eliminación de filas sin información

Existen registros totalmente en blanco (nulos) en la base de datos pública, como se observa en la Figura 10 . Se deben eliminar dichos registros, con su clave.

id	sector	ambito	inicial_mate	inicial_infan	primaria6	secundaria_(	secundaria_(	sec_inet	snu	snu_inet
105600014960881	Estatal	Urbano			X					
108450015300901	Estatal	Rural			X					
108570011220991	Estatal	Rural								
109210022780671	Estatal	Rural			X					
109720088410410	Estatal	Rural			X					
110520061210811	Estatal	Rural			X					
110521861210811	Estatal	Rural			X					
111520023120681	Estatal	Rural			X					
111520023127481	Estatal	Urbano			X					

Figura 10. Eliminación de registros totalmente nulos. Fuente propia.

Este grupo de operaciones se debe seguir con todas las planillas, y funciona bien, siempre y cuando todas tengan el mismo tipo de datos en toda la columna (en todo el país). Sin embargo, esto no siempre es así, por lo que resulta necesario definir a esas columnas como string, y posteriormente, transformarlas en MongoDB, a su verdadero tipo de datos.

3.1.3 Importación de datos con el MongoDB Compass

Actualmente, en 2024, con el MongoDB Compass se pueden importar datos en formato JSON o CSV. El MongoDB Compass es una interfaz gráfica de usuario, que permite realizar un conjunto de operaciones sobre las bases de datos MongoDB. En la Figura 11, se observan los pasos de la importación: transformación de datos la planilla a un formato CSV, posteriormente se pasa a texto con un cotejamiento de caracteres UTF8, y finalmente se importa en MongoDB Compass.

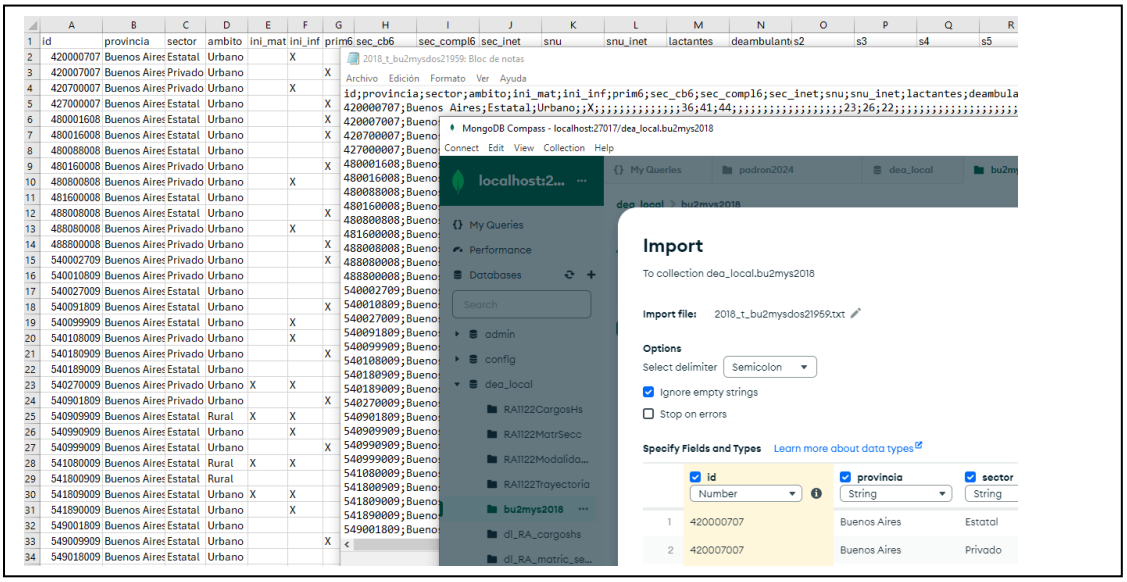


Figura 11. Los pasos necesarios para poder importar con MongoDB Compass. Fuente propia.

El MongoDB Compass, detecta automáticamente los tipos de datos correspondientes a los primeros documentos de la colección, como se observa en la Figura 12.

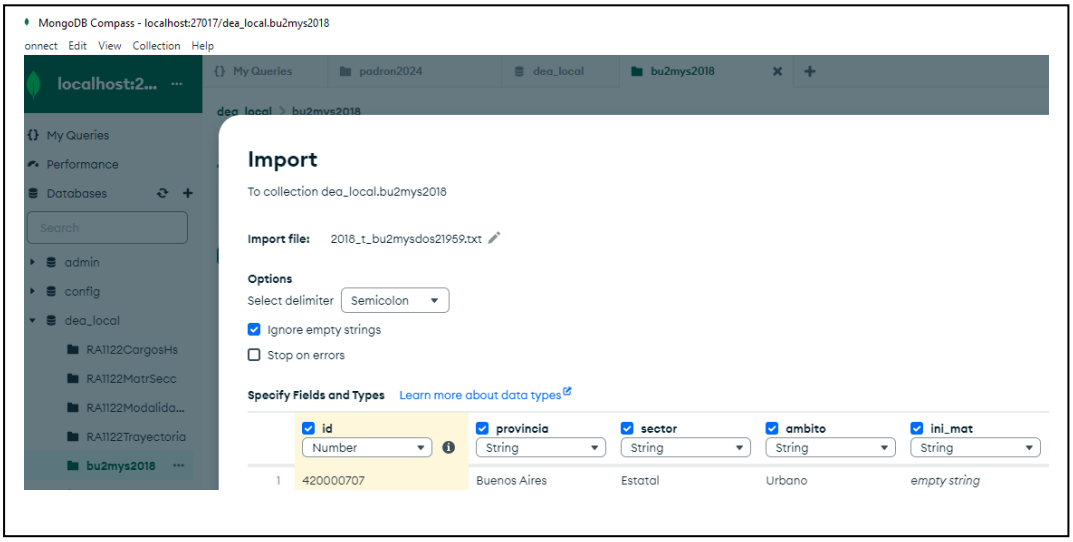


Figura 12. El MongoDB Compass, actualmente detecta automáticamente el tipo de datos. Fuente propia.

No obstante, siempre debemos estar atentos a que, si el tipo de dato en el campo específico es numérico, debe serlo para todos los documentos de la colección. Si así no fuera, debemos transformar a un valor numérico.

Tanto la colección padrón, que se fue actualizando a medida que fue pasando el tiempo, como en las otras colecciones, resulta necesario cambiar el tipo de dato alfabético a numérico, para poder operarlos mejor.

Obsérvese, que en muchas tablas desde 2011, se consignaba una “X” para indicar que una institución ofrecía una oferta determinada, como se observa en Figura 13.

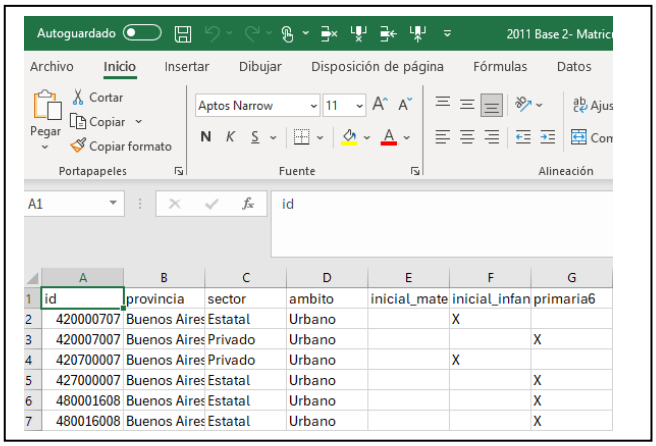


Figura 13. Se consignaba una "X" para indicar que una institución ofrecía una oferta determinada Fuente propia.

Una vez que los datos están importados, resulta necesario transformar la “X” a un entero, en este caso el 1, para poder operar y contar más fácilmente.

Por ejemplo, en la colección de matrículas y secciones del 2011, es necesario cambiar los campos que tienen una “X” indicando que esa institución tiene una oferta educativa de inicial maternal, a un valor numérico, con dos sentencias para cada campo. Como usuario MongoDB (en el mongosh), las sentencias se observan en el Script 3:

```
db.bu2mys2011.updateMany({{$and:[{prov:/Formosa/},{ini_mat:"X"}]},{$s
et:{ini_mat:"1"}},multi:true})
db. bu2mys2011.updateMany(
{ini_mat:”1”},
[{$set: {ini_mat:{ $convert: { input: “$ini_mat”, to: ”int” } } } }] )
```

Script 3. Transformación de un campo alfabético a un tipo de dato numérico. Fuente propia.

El script anterior, hace la conversión en un par de sentencias: la primera cambia la “X” por un “1” del campo ini_mat, y la segunda, transforma el tipo de dato del campo ini_mat a numérico entero.

Debido a que son 34 campos en esta colección para este año, conviene ocupar NodeJS. Para el mismo campo indicado en el Script 3, se muestra una parte del código equivalente, perteneciente al módulo transformo02.js, como se observa en el Script 4 .

```
var url = “mongodb://localhost:27017/”;
MongoClient.connect(url,function(err,db){
if (err) throw err;
var dbo = db.db(“dea_local”); //base de datos
console.log(“paso 003”);
dbo.collection(“bu2mys2011”).updateMany(
{ini_mat:{$regex: ”X” }},
[{$set: {ini_mat:
{$replaceOne:{input:” $ini_mat”,find:” X”, replacement:”1” }} }
}]);
dbo.collection(“bu2mys2011”).updateMany(
{ini_mat:”1”},
[{$set: {ini_mat:{ $convert: { input: “$ini_mat”, to: ”int” } } } }]
);
```

Script 4. Transformación de un campo alfabético a numérico, en NodeJS. Fuente propia.

Posteriormente, se ejecutan distintas transformaciones como se observa en la Figura 14. Cada una de estas transformaciones, tienen más de 200 líneas de código.

```
Node.js command prompt
C:\Users\vhruc>node "transformo01b.js"
->paso 001
->paso 002
(node:12708) [MONGODB DRIVER] Warning: Current Server Discovery and Monitoring engine is deprecated, and will be removed
 in a future version. To use the new Server Discover and Monitoring engine, pass option { useUnifiedTopology: true } to
the MongoClient constructor.
(Use `node --trace-warnings ...` to show where the warning was created)
->paso 003

C:\Users\vhruc>node "transformo02.js"
->paso 001
->paso 002
(node:10104) [MONGODB DRIVER] Warning: Current Server Discovery and Monitoring engine is deprecated, and will be removed
 in a future version. To use the new Server Discover and Monitoring engine, pass option { useUnifiedTopology: true } to
the MongoClient constructor.
(Use `node --trace-warnings ...` to show where the warning was created)
->paso 003
->termino

C:\Users\vhruc>node "transformo03.js"
->paso 001
->paso 002
(node:4420) [MONGODB DRIVER] Warning: Current Server Discovery and Monitoring engine is deprecated, and will be removed
in a future version. To use the new Server Discover and Monitoring engine, pass option { useUnifiedTopology: true } to t
he MongoClient constructor.
(Use `node --trace-warnings ...` to show where the warning was created)
->paso 003
->termino

C:\Users\vhruc>
```

Figura 14. Procesamiento de transformaciones en NodeJS. Fuente propia.

En esta instancia se tienen todos los datos en MongoDB, sin espacios ni ceros, y sin relación entre los establecimientos y los valores estadísticos. Sin embargo, los datos están incompletos, porque las bases de datos no tienen datos referidos a las modalidades de la educación especial ni la educación de adultos.

Si se cuenta la cantidad de documentos la colección de “matrícula y secciones” del año 2.018 correspondientes a la provincia de Formosa, y un “padrón” de establecimientos educativos, también de Formosa, de principios de 2019, existe una diferencia de 320 documentos, como se observa en la Figura 15.

```
localhost:27017 > dea_local > padron20190916
Documents 63.2K   Aggregations   Schema   Indexes 1   Validation
{"jurisdiccion":/Formosa/}   Generate query   Explain   Reset   Find   Opt
ADD DATA   EXPORT DATA   UPDATE   DELETE   201 - 220 of 1736
    ec_pri : "X"
1   _id: ObjectId('663406af36dc4
2   jurisdiccion : "Formosa"
3   CUE_anexo : 340018700
4   Nombre : "EPEP N 308 PATRICI
5   Sector : "Estatal"
6   ambito : "Rural"
7   Domicilio : "   COLONIA AB.
mongosh mongodb://127.0.0.1:27017/?directConnection=true&serverSelectionTimeoutMS=2000
Enterprise dea_local> db.bu2mys2018.find({"provincia":/Formosa/}).count()
1416
Enterprise dea_local> db.padron20190916.find({"jurisdiccion":/Formosa/}).count()
1736
Enterprise dea_local>
```

Figura 15. Diferencias en 2018, entre la cantidad de documentos del padrón y la colección de matrícula y secciones. Fuente Ruchinsky et al. **[22]**.

Esta diferencia se debe a que sólo se informa la matrícula de los cuatro niveles (inicial, primario, secundario y superior no universitario), a diferencia del padrón que muestra todas las unidades educativas, inclusive las correspondientes a la modalidad de la educación permanente de jóvenes y adultos (adulto primario, adulto secundario y formación profesional), y educación especial. Sería deseable que el Ministerio de Capital Humano informara las bases de datos de las modalidades.

3.2 Primera aproximación a la estructura de la base de datos

3.2.1 Composición de las colecciones

Todas las colecciones tienen un ID, la jurisdicción (provincia) a la que pertenecen, el ámbito al que pertenecen, el sector de gestión, y el nivel de cada oferta educativa que prestan. Se distinguen en:

a) El padrón tiene el nombre del establecimiento, el domicilio, el departamento y la localidad. El ID del padrón se denomina CUE-anexo.

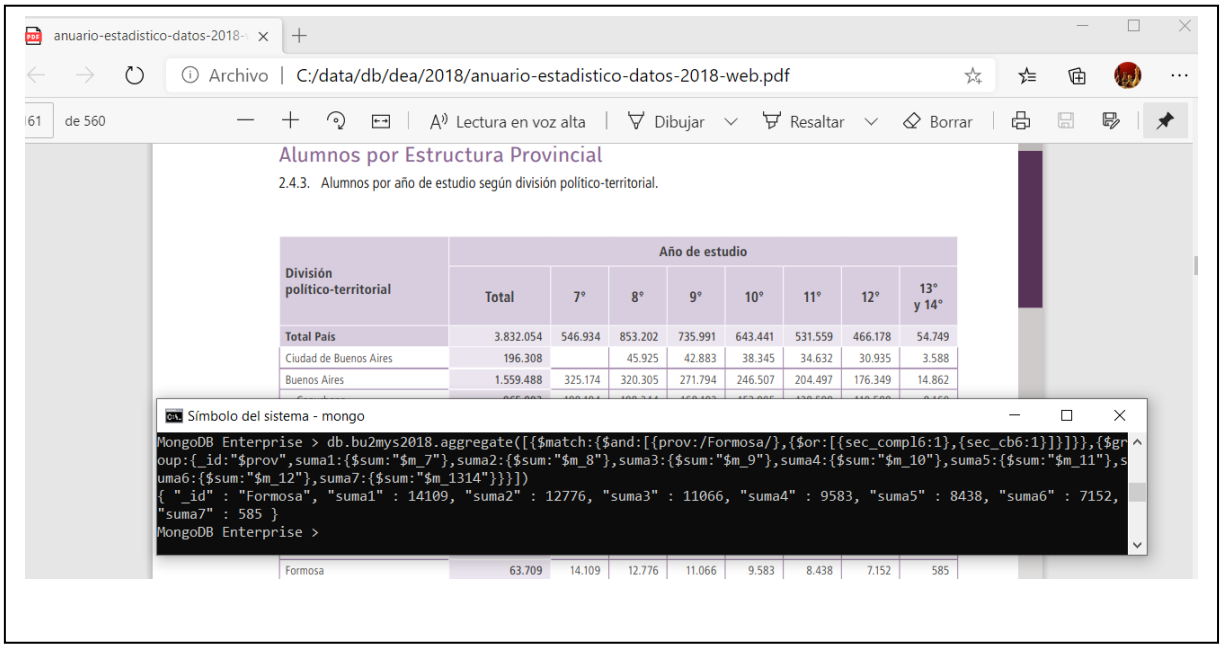
b) Las otras colecciones tienen todas el mismo ID cuando hacen referencia al mismo establecimiento educativo, asignados aleatoriamente por el Ministerio de Capital Humano para publicar los datos.

Además de estos datos, cada colección tiene:

- Datos de matrículas y secciones, como ser la bu2mys2018: que significa, bases usuarias de matrículas y secciones 2018. Esta colección tiene datos de matrícula por salas de 2, 3, 4 y 5 años cuando la institución es del nivel inicial; de cada uno de los grados de la primaria, si tiene esa oferta; de los años de la secundaria, hasta el séptimo para las instituciones que son de escuelas de educación técnica conocidas como INET; el total de la matrícula del nivel superior no universitario. También tiene la matrícula de varones por sala/grado/año, los repitentes por grado/año y los repitentes varones por grado/año. Más adelante veremos en este trabajo, que los resultados de Aprender, se informan las mujeres. Además, se indica la cantidad de secciones independientes y múltiples, si existen. Una sección es múltiple cuando en un mismo espacio físico, hay alumnos de distintos trayectos educativos o de distintos grados/años.
- Colecciones de trayectoria que hacen referencia a las matrículas del año anterior.
- Colecciones de cargos y horas.

3.2.2 Comprobación de totales generales con datos externos

Para realizar la comprobación de matrícula a nivel jurisdicción, por cada año en el nivel secundario, por ejemplo, se compara la colección de matrículas y secciones 2018, con el Anuario Estadístico 2018 [23], en formato PDF, como se observa en la Figura 16.



División político-territorial	Año de estudio							
	Total	7°	8°	9°	10°	11°	12°	13° y 14°
Total País	3.832.054	546.934	853.202	735.991	643.441	531.559	466.178	54.749
Ciudad de Buenos Aires	196.308		45.925	42.883	38.345	34.632	30.935	3.588
Buenos Aires	1.559.488	325.174	320.305	271.794	246.507	204.497	176.349	14.862
Formosa	63.709	14.109	12.776	11.066	9.583	8.438	7.152	585

Figura 16. Control de totales con datos externos. Fuente propia y Anuario Estadístico Educativo 2018.

Obsérvese que, en el caso de Formosa, en el anuario estadístico de 2018, en primer año de secundaria, que aparece como séptimo año en Formosa, coinciden los totales en 14.109, tanto en el cuadro para la fila Formosa, como el campo “suma1” de la consulta de MongoDB; y así en todos los campos. Lo mismo sucede en los otros niveles.

Cabe aclarar que se trabajó con datos del 2018 porque se hace foco en los últimos 12 años, correspondientes a un ciclo educativo completo.

3.2.3 Ejemplos de relación CUE_anexo-ID

Es conveniente separar las distintas unidades educativas agrupándolas en colecciones que hacen referencia a las distintas variables, y a cada una de estas colecciones, identificarlas con su CUE-anexo, como se observa en la Figura 17. En la misma se observa a la colección establecimiento, relacionada con las distintas variables: matrícula, horas cubiertas, cargos cubiertos, cargos sin cubrir, etc.

Para identificar cada CUE-anexo, se deben distinguir los establecimientos por las ofertas educativas que ofrecen, relacionando la información de las distintas colecciones de las bases

de datos publicadas por el Ministerio de Capital Humano entre sí, y datos públicos existentes en la web, iniciando el análisis por aquellas instituciones más grandes, y que tienen las 4 ofertas educativas, ya que estas instituciones van a tener los cuatro campos de educación común marcados en el padrón, y van a tener la matrícula en los 4 niveles educativos. El esquema que relaciona las colecciones principales se puede observar en la Figura 17.

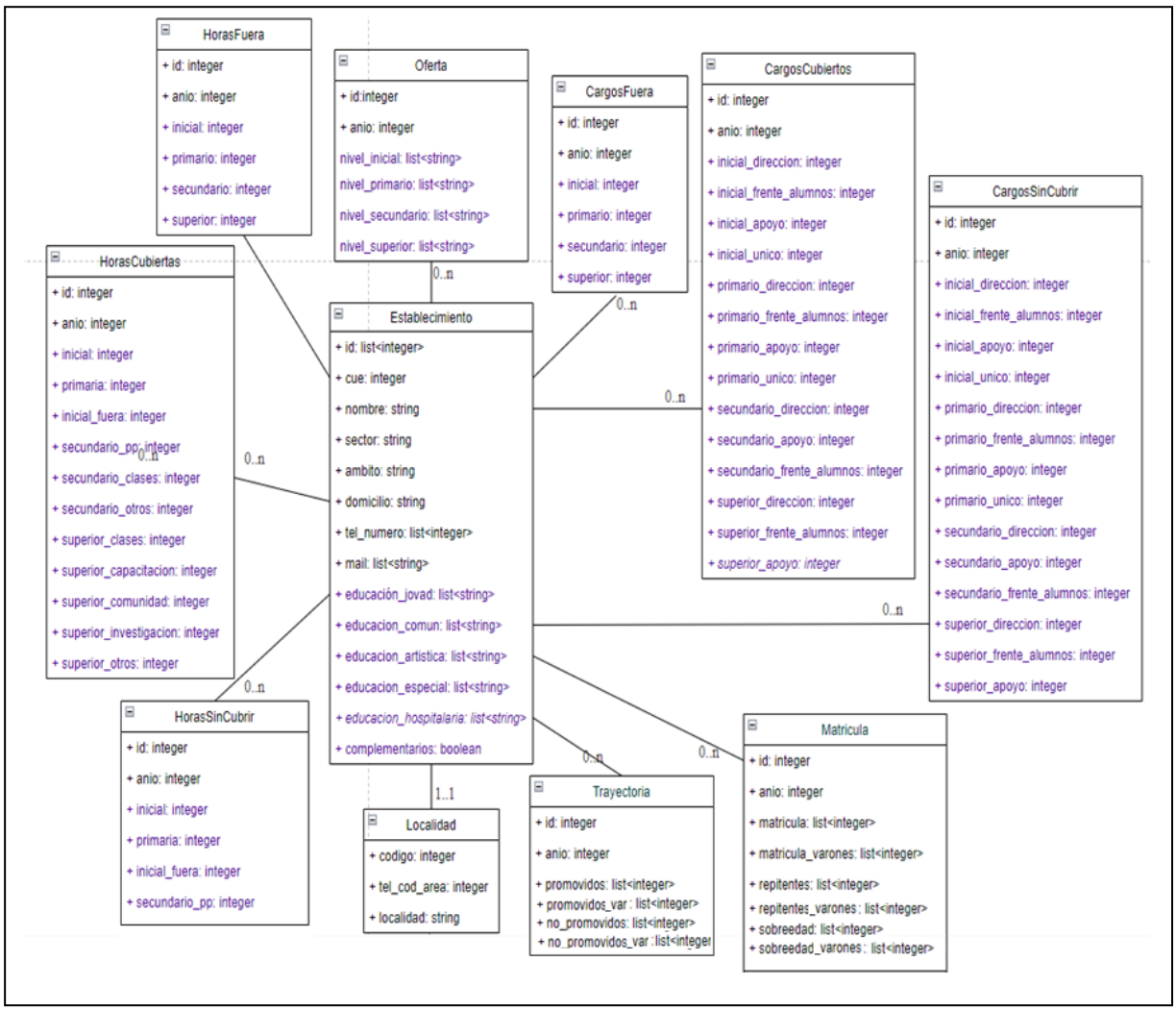


Figura 17. Esquema que relaciona las principales colecciones. Fuente propia.

En primer lugar, podríamos asignar los CUE-anexo a las instituciones que tienen ofertas en los 4 niveles educativos. Para ello, se consulta en MongoDB la colección “padrón”, la colección de “matrícula y las secciones”, como se observa en la Figura 18.

```
MongoDB Enterprise > db.padron20190916f.find({$and:[{prov:/Formosa/},{ec_ini:1},{ec_pri:1},{ec_sec:1},{ec_snu:1}]},{_id:1,nombre:1,departamento:1,loc
alidad:1})
{ "_id" : "340049900", "nombre" : "ESCUELA NORMAL SUPERIOR REPUBLICA DEL PARAGUAY", "localidad" : "FORMOSA", "departamento" : "FORMOSA" }
{ "_id" : "340051100", "nombre" : "INSTITUTO SANTA ISABEL", "localidad" : "FORMOSA", "departamento" : "FORMOSA" }
{ "_id" : "340054700", "nombre" : "ISFDCYT FELIX ATILIO CABRERA", "localidad" : "FORMOSA", "departamento" : "FORMOSA" }
{ "_id" : "340070300", "nombre" : "INSTITUTO TECNICO SAN JOSE □-15", "localidad" : "CLORINDA", "departamento" : "PILCOMAYO" }
{ "_id" : "340070500", "nombre" : "SANTA CATALINA LABOURE (PR-04)", "localidad" : "CLORINDA", "departamento" : "PILCOMAYO" }
{ "_id" : "340071400", "nombre" : "INSTITUTO SAN JOSE □12", "localidad" : "COMANDANTE FONTANA", "departamento" : "PATI□O" }
MongoDB Enterprise > db.bu2mys2018.aggregate([{$match:{$and:[{prov:/Formosa/},{ini_inf:1},{prim6:1},{sec_compl6:1},{snu:1}]}},{$group:{_id:"$_id",mat
r_inicial:{$sum:{$sum:["$ms3","$ms4","$ms5"]}},matr_primaria:{$sum:{$sum:["$m_1","$m_2","$m_3","$m_4","m_5","$m_6"]}},matr_prim_varon:{$sum:{$sum:["$
mv_1","$mv_2","$mv_3","$mv_4","$mv_5","$mv_6"]}},matr_secundaria:{$sum:{$sum:["$m_7","$m_8","$m_9","$m_10","$m_11","$m_12","$m_1314"]}},matr_sec_varo
n:{$sum:{$sum:["$mv_7","$mv_8","$mv_9","$mv_10","$mv_11","$mv_12","$mv_1314"]}},matr_inet:{$sum:"$m_1314"},matr_superior_no_univ:{$sum:"$m_snu"},matr
_TOTAL:{$sum:{$sum:["$ms3","$ms4","$ms5","$m_1","$m_2","$m_3","$m_4","m_5","$m_6","$m_7","$m_8","$m_9","$m_10","$m_11","$m_12","$m_1314","$m_snu"]}}}
}])
{ "_id" : "9861447129", "matr_inicial" : 208, "matr_primaria" : 322, "matr_prim_varon" : 202, "matr_secundaria" : 695, "matr_sec_varon" : 377, "matr_
inet" : 0, "matr_superior_no_univ" : 693, "matr_TOTAL" : 1918 }
{ "_id" : "4760715414", "matr_inicial" : 120, "matr_primaria" : 317, "matr_prim_varon" : 0, "matr_secundaria" : 393, "matr_sec_varon" : 0, "matr_inet
" : 0, "matr_superior_no_univ" : 17, "matr_TOTAL" : 847 }
{ "_id" : "7821258123", "matr_inicial" : 95, "matr_primaria" : 388, "matr_prim_varon" : 231, "matr_secundaria" : 566, "matr_sec_varon" : 255, "matr_i
net" : 0, "matr_superior_no_univ" : 780, "matr_TOTAL" : 1829 }
{ "_id" : "6461339519", "matr_inicial" : 129, "matr_primaria" : 356, "matr_prim_varon" : 188, "matr_secundaria" : 528, "matr_sec_varon" : 228, "matr_
inet" : 0, "matr_superior_no_univ" : 266, "matr_TOTAL" : 1279 }
{ "_id" : "6461356619", "matr_inicial" : 25, "matr_primaria" : 136, "matr_prim_varon" : 79, "matr_secundaria" : 236, "matr_sec_varon" : 98, "matr_ine
t" : 0, "matr_superior_no_univ" : 161, "matr_TOTAL" : 558 }
{ "_id" : "5781195117", "matr_inicial" : 152, "matr_primaria" : 340, "matr_prim_varon" : 222, "matr_secundaria" : 693, "matr_sec_varon" : 348, "matr_
inet" : 76, "matr_superior_no_univ" : 151, "matr_TOTAL" : 1336 }
```

Figura 18. Listado de establecimientos y matrículas que tienen las 4 ofertas educativas. Fuente propia.

De las 6 instituciones resultantes, puede observarse que 3 son de Formosa y 3 son del interior de la provincia. De las 3 instituciones de Formosa, 1 es el Instituto Santa Isabel, que se la distingue con claridad, ya que en sus niveles inicial, primario y secundario, sólo admite mujeres y obsérvese la etiqueta “matr.prim.varon”:0, y “matr_sec.varon”. La actualización del CUE_anexo se observa en el Script 5.

```
dbo.collection("bu2mys2018").updateOne({$and:[{prov:/Formosa/},
{cueanexo:{$not:/34/}},{ini_inf:1},{prim6:1},{sec_compl6:1},{sn
u:1},{mv_1:0},{mv_2:0},{mv_3:0},{mv_4:0},{mv_5:0},{mv_6:0},{mv_
7:0},{mv_8:0},{mv_9:0},{mv_10:0},{mv_11:0},{mv_12:0}]},        [{
$set: { cueanexo:"340051100" }  }]  );
```

Script 5. Actualización de un CUE_anexo en la colección de matrículas y secciones. Fuente propia.

Obsérvese, además, que existe una institución que tiene matrícula del INET, tiene matrícula en el 7º año de secundaria técnica (76 alumnos), y en la colección padrón, existe una sola institución que resalta en su nombre la palabra técnico: es el Instituto Técnico San José, CUE: 340070300 de Clorinda.

Quedan 4 instituciones, con una matrícula de 1918, 1829, 1279 y 558 alumnos, respectivamente, y falta asignar a 2 instituciones que pertenecen al departamento Formosa de la localidad Formosa, a una institución del departamento Pilcomayo de la localidad Clorinda y a una del departamento Patiño de la localidad Fontana. Según el INDEC [24], en el último censo de población de 2010, el departamento Formosa cuenta con 234.354 habitantes, el departamento Pilcomayo con 85.024 y el departamento Patiño con 68.581

habitantes. Las unidades educativas, debería pasar por el mismo prisma de la población general, ya que éstas son establecimientos educativos que tienen los 4 niveles educativos y son los más representativos de la población general, razón por la cual, las 2 instituciones de mayor matrícula se podrían asignar a Formosa (al ISFDCYT Félix Atilio Cabrera - FAC y a la Escuela Normal Superior República del Paraguay), la siguiente a Pilcomayo (Santa Catalina Labouré), y la siguiente a Patiño (Instituto San José).

El problema es distinguir a cada una de las 2 instituciones de Formosa. De estas dos instituciones, una tiene 780 alumnos en el nivel superior no universitario, y la otra 693; una de ellas tuvo como director por muchos años al primer Rector de la Universidad Nacional de Formosa [25] (el Instituto Félix Atilio Cabrera [26]), entonces podríamos asignar la mayor matrícula del nivel superior al Instituto Félix Atilio Cabrera, CUE: 340054700, como se observa en la Figura 19; la segunda institución es la Escuela Normal República del Paraguay, CUE: 340049900.

Ejecutando el módulo: asigno0_bu2mys2018.js, que tiene entre otras, la asignación específica del CUE_anexo, como se observa en Script 6:

```
dbo.collection("bu2mys2018").updateOne(
{$and:[{prov:/Formosa/},{ini_inf:1},{prim6:1},{sec_compl6:1},{s
nu:1}      ,{m_snu:780}]},    [{ $set: { cueanexo:"340054700"   }
}] );
```

Script 6. Asignación de CUE_anexo a un documento de la colección de matrícula y secciones. Fuente propia

y realizando una nueva consulta, se ve que ya tiene el atributo CUE_anexo, como se observa en la Figura 19.

```
MongoDB Enterprise > db.bu2mys2018.find({$and:[{prov:/Formosa/},{ini_inf:1},{prim6:1},{sec_compl6:1},{snu:1},{m_snu:780}]})
{ "_id" : "7821258123", "prov" : "Formosa", "sector" : "Estatal", "ambito" : "Urbano", "ini_inf" : 1, "prim6" : 1, "sec_compl6" : 1, "snu" : 1, "ms
 : 26, "ms5" : 69, "m_1" : 78, "m_2" : 80, "m_3" : 77, "m_4" : 80, "m_5" : 78, "m_6" : 73, "m_7" : 108, "m_8" : 120, "m_9" : 98, "m_10" : 82, "m_11
: 95, "m_12" : 63, "m_snu" : 780, "mvs4" : 12, "mvs5" : 31, "mv_1" : 37, "mv_2" : 38, "mv_3" : 33, "mv_4" : 43, "mv_5" : 46, "mv_6" : 34, "mv_7" :
, "mv_8" : 50, "mv_9" : 48, "mv_10" : 35, "mv_11" : 40, "mv_12" : 28, "mv_snu" : 159, "r_1" : 0, "r_2" : 1, "r_3" : 2, "r_4" : 0, "r_5" : 1, "r_6"
2, "r_7" : 5, "r_8" : 14, "r_9" : 12, "r_10" : 12, "r_11" : 7, "r_12" : 2, "so_2" : 1, "so_3" : 2, "so_5" : 1, "so_6" : 8, "so_7" : 8, "so_8" : 18,
so_9" : 16, "so_10" : 12, "so_11" : 7, "so_12" : 3, "si_lac" : 0, "si_deam" : 0, "si_s2" : 0, "si_s3" : 0, "si_s4" : 1, "si_s5" : 3, "si_1" : 3, "s
2" : 3, "si_3" : 3, "si_4" : 3, "si_5" : 4, "si_6" : 3, "si_7" : 4, "si_8" : 4, "si_9" : 4, "si_10" : 3, "si_11" : 3, "si_12" : 2, "si_1314" : 0, "
_ini" : 0, "sm_pri" : 0, "sm_multi" : 0, "cueanexo" : "340054700" }
```

Figura 19. Asignación de un CUE-anexo específico. Fuente propia.

Aquí se ve como ya se pueden relacionar las dos colecciones mediante el campo CUE_anexo. En la Figura 20 se visualizan las colecciones en forma separada.

```
MongoDB Enterprise > db.padron20190916f.find({$and:[{prov:/Formosa/},{nombre:/ATILIO CABRERA/}]}).pretty()
{
        "_id" : "340054700",
        "prov" : "Formosa",
        "nombre" : "ISFDCYT FELIX ATILIO CABRERA",
        "sector" : "Estatal",
        "ambito" : "Urbano",
        "domicilio" : "CORDOBA 1125 DON BOSCO Bo. DON BOSCO -CORDOBA 1125",
        "cp" : "P3600",
        "tel_area" : "370",
        "tel_nro" : "4422545",
        "cod_loca" : "34014051",
        "localidad" : "FORMOSA",
        "departamento" : "FORMOSA",
        "mail" : "mcye.isfdcytatiliocabrera@formosa.gob.ar",
        "ed_comun" : 1,
        "ec_ini" : 1,
        "ec_pri" : 1,
        "ec_sec" : 1,
        "ec_snu" : 1,
        "vlati" : "-58.1766",
        "vlongi" : "-26.1905"
}
MongoDB Enterprise > db.bu2mys2018.find({$and:[{prov:/Formosa/},{ini_inf:1},{prim6:1},{sec_compl6:1},{snu:1},{m_snu:780}]})
{ "_id" : "7821258123", "prov" : "Formosa", "sector" : "Estatal", "ambito" : "Urbano", "ini_inf" : 1, "prim6" : 1, "sec_compl6" : 1, "snu" : 1, "ms4
 : 26, "ms5" : 69, "m_1" : 78, "m_2" : 80, "m_3" : 77, "m_4" : 80, "m_5" : 78, "m_6" : 73, "m_7" : 108, "m_8" : 120, "m_9" : 98, "m_10" : 82, "m_11"
: 95, "m_12" : 63, "m_snu" : 780, "mvs4" : 12, "mvs5" : 31, "mv_1" : 37, "mv_2" : 38, "mv_3" : 33, "mv_4" : 43, "mv_5" : 46, "mv_6" : 34, "mv_7" : 5
, "mv_8" : 50, "mv_9" : 48, "mv_10" : 35, "mv_11" : 40, "mv_12" : 28, "mv_snu" : 159, "r_1" : 0, "r_2" : 1, "r_3" : 2, "r_4" : 0, "r_5" : 1, "r_6" :
2, "r_7" : 5, "r_8" : 14, "r_9" : 12, "r_10" : 12, "r_11" : 7, "r_12" : 2, "so_2" : 1, "so_3" : 2, "so_5" : 1, "so_6" : 8, "so_7" : 8, "so_8" : 18,
so_9" : 16, "so_10" : 12, "so_11" : 7, "so_12" : 3, "si_lac" : 0, "si_deam" : 0, "si_s2" : 0, "si_s3" : 0, "si_s4" : 1, "si_s5" : 3, "si_1" : 3, "si
2" : 3, "si_3" : 3, "si_4" : 3, "si_5" : 4, "si_6" : 3, "si_7" : 4, "si_8" : 4, "si_9" : 4, "si_10" : 3, "si_11" : 3, "si_12" : 2, "si_1314" : 0, "
```

Figura 20. Consulta de una institución en cada una de las colecciones. Fuente propia.

Haciendo la combinación de ambas colecciones, en Node JS, como se observa en el Script 7.

```
dbo.collection("bu2mys2018").aggregate([{ $lookup: { from:
"padron"                                ,localField:"cueanexo",
foreignField:"_id",as:"cpadron"}          },{          $match:
{$and:[{prov:/Formosa/},{cueanexo:/34/}   ]   }},{   $group:
{_id:
{vpadronid:"$cueanexo",vnom:"$cpadron.nombre",vdepartamen
to:"$cpadron.departamento",vlocalidad:"$cpadron.localidad
"},matr_inicial:{$sum:{$sum:["$ms3","$ms4","$ms5"]}},matr
_primaria:{$sum:{$sum:["$m_1","$m_2","$m_3","$m_4","$m_5"
,"$m_6"]}},matr_prim_varon:{$sum:{$sum:["$mv_1","$mv_2","
$mv_3","$mv_4","$mv_5","$mv_6"]}},matr_secundaria:{$sum:{
$sum:["$m_7","$m_8","$m_9","$m_10","$m_11","$m_12","$m_13
14"]}},matr_sec_varon:{$sum:{$sum:["$mv_7","$mv_8","$mv_9
","$mv_10","$mv_11","$mv_12","$mv_1314"]}},matr_inet:{$su
m:"$m_1314"},matr_superior_no_univ:{$sum:"$m_snu"},matr_t
ot:{$sum:{$sum:["$ms3","$ms4","$ms5","$m_1","$m_2","$m_3"
,"$m_4","$m_5","$m_6","$m_7","$m_8","$m_9","$m_10","$m_11
","$m_12","$m_1314","$m_snu"]}}         }         }         ]
).forEach(function(r){vii=vii+1;console.log(vii+"*CUE_ane
xo:"+r._id.vpadronid+"--
>"+r._id.vnom+"Departamento:"+r._id.vdepartamento+"Locali
dad:"+r._id.vlocalidad+                                  "--
>Matriculas:Inicial:"+r.matr_inicial+"          Primaria:"
+r.matr_primaria+"    Secundaria:"    +r.matr_secundaria+"
INET(7):"             +r.matr_inet+"            Superior:"
+r.matr_superior_no_univ        +"    TOTAL:"+r.matr_tot);
console.log("    ")});
```

Script 7. Combinación de las colecciones padrón y bu2mys2018. Fuente propia

Se obtiene la consulta que se observa en la Figura 21.

```
Node.js command prompt
1*CUE_anexo:340070500-->SANTA CATALINA LABOURE (PR-04) Departamento:PILCOMAYO Localidad:CLORINDA-->Matriculas:Inicial:129 Primaria:435 Secundaria:5
 INET(7):0 Superior:266 TOTAL:1358

2*CUE_anexo:340051100-->INSTITUTO SANTA ISABEL Departamento:FORMOSA Localidad:FORMOSA-->Matriculas:Inicial:120 Primaria:383 Secundaria:393 INET(7):
Superior:17 TOTAL:913

3*CUE_anexo:340071400-->INSTITUTO SAN JOSE 12 Departamento:PATIO Localidad:COMANDANTE FONTANA-->Matriculas:Inicial:25 Primaria:169 Secundaria:236
NET(7):0 Superior:161 TOTAL:591

4*CUE_anexo:340054700-->ISFDCYT FELIX ATILIO CABRERA Departamento:FORMOSA Localidad:FORMOSA-->Matriculas:Inicial:95 Primaria:466 Secundaria:566 INE
7):0 Superior:780 TOTAL:1907

5*CUE_anexo:340070300-->INSTITUTO TECNICO SAN JOSE -15 Departamento:PILCOMAYO Localidad:CLORINDA-->Matriculas:Inicial:152 Primaria:409 Secundaria:
3 INET(7):76 Superior:151 TOTAL:1405

6*CUE_anexo:340049900-->ESCUELA NORMAL SUPERIOR REPUBLICA DEL PARAGUAY Departamento:FORMOSA Localidad:FORMOSA-->Matriculas:Inicial:208 Primaria:392
ecundaria:695 INET(7):0 Superior:693 TOTAL:1988


C:\Users\vhruc>
```

Figura 21. Consulta de establecimientos combinando colecciones. Fuente propia.

En segundo lugar, podríamos asignar a las instituciones que tienen ofertas en 3 niveles educativos.

3.2.4 Límite en el análisis de la relación CUE-ID

El trabajo con el ID tiene un par de limitaciones:

a) Trabajando sobre las matrículas que más se alejan del promedio, con información pública y abierta, se podría continuar con el análisis de la relación entre CUE - ID, pero un límite es el secreto estadístico que impera en los datos educativos. En este sentido, la Ley de Secreto Estadístico Nº 17.622 del" año 1968 afecta la posibilidad de difundir datos que permitan identificar a las entidades que declararon esos datos, así como de suministrarlos a cualquier persona externa al área de información o estadística, incluyendo a las altas autoridades de un Ministerio. En este sentido, según el artículo 10 de la mencionada ley, las informaciones que se suministren a los organismos que integran el Sistema Estadístico Nacional, serán estrictamente secretas y sólo se utilizarán con fines estadísticos. Además de esta ley, el secreto estadístico se enmascara en el art. 97 de la Ley 26206 de Educación Nacional y la ley 25326 de Protección de Datos Personales. Por lo que resulta conveniente no avanzar en su determinación, procesamiento, análisis y publicación.

b) Además, el ID, que sirve como campo clave para todas las tablas de las bases de datos que informa el Ministerio de Capital Humano, cambia cada vez que cambia el presidente de la Nación. En 2011 el ID constaba de 10 dígitos, y en 2022, de 14 dígitos, impidiendo realizar el cálculo continuo y completo de las tasas.

No obstante, sin determinar la relación CUE_anexo-ID, trabajaremos con clave enmascarada, sin indicar en forma precisa la unidad educativa.

Capítulo 4

Procesamiento en modo local y completamiento de datos

4.1 Necesidad de procesamiento de los datos en modo local

MongoDB Atlas [27] permitió subir datos de bases de datos a la nube, sin límites, hasta mediados del año 2022. Como durante todo el desarrollo del presente trabajo se utilizó una versión gratuita del MongoDB Atlas, hacia mediados de 2022 el clúster utilizado no permitió importar como se necesitaba. Agrupando los datos de matrículas y secciones correspondientes a Formosa para los 12 años de análisis, se totalizan 16.900 documentos. Para la importación de los datos de los 16.900 documentos, el proceso de importación llevó más de 24 hs de duración, y sólo se permitió subir 400 documentos, solamente aquellos que tenían la menor cantidad de campos, es decir, se lograron subir los documentos más chicos. Por la razón expuesta, los resultados debieron preprocesarse en forma local, y posteriormente exportarse en formato JSON y subirse al MongoDB Atlas, para terminar su procesamiento y presentar los resultados. Resulta importante aclarar que se utilizó una versión gratuita de Atlas. Para la última revisión de este trabajo, son 13 años de registro, y totalizan 18.300 documentos como se observa en la Figura 22.

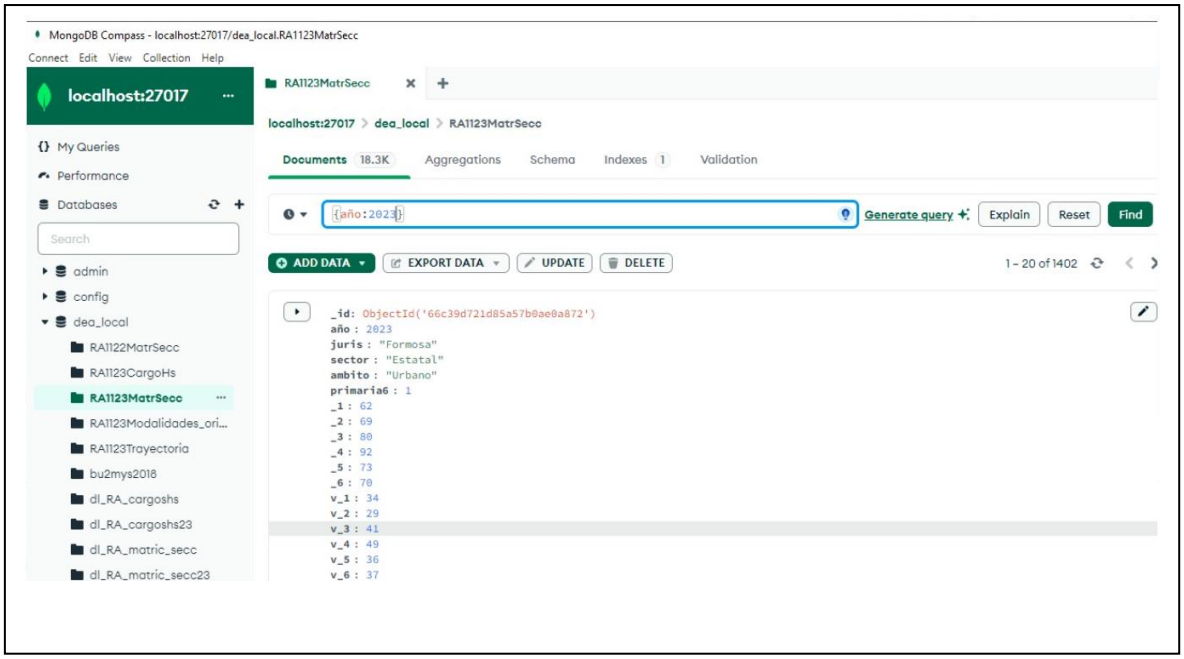


Figura 22. Documentos de matrículas y secciones correspondientes a 13 años (18.3 k documentos). Fuente propia.

4.1.1 Agrupamiento por año, sector y ámbito

Para poder resumir la información la nueva clave de agrupamiento está conformada por el año, el sector y el ámbito. En este agrupamiento se suman los distintos valores para cada uno de los campos. Al final de la Figura 23 se observa cómo se suman cada uno de los 103 campos distintos de trayectoria: `"..., mga_1: {$sum:"$ A_1"},mga_2:{$sum:"$`

`A_2"}, mga_3:{$sum:"$ A_3"}, mga_4:{$sum:"$ A_4"},mga_5:{$sum:"$ A_5"}…”,` y posteriormente se agrupa.

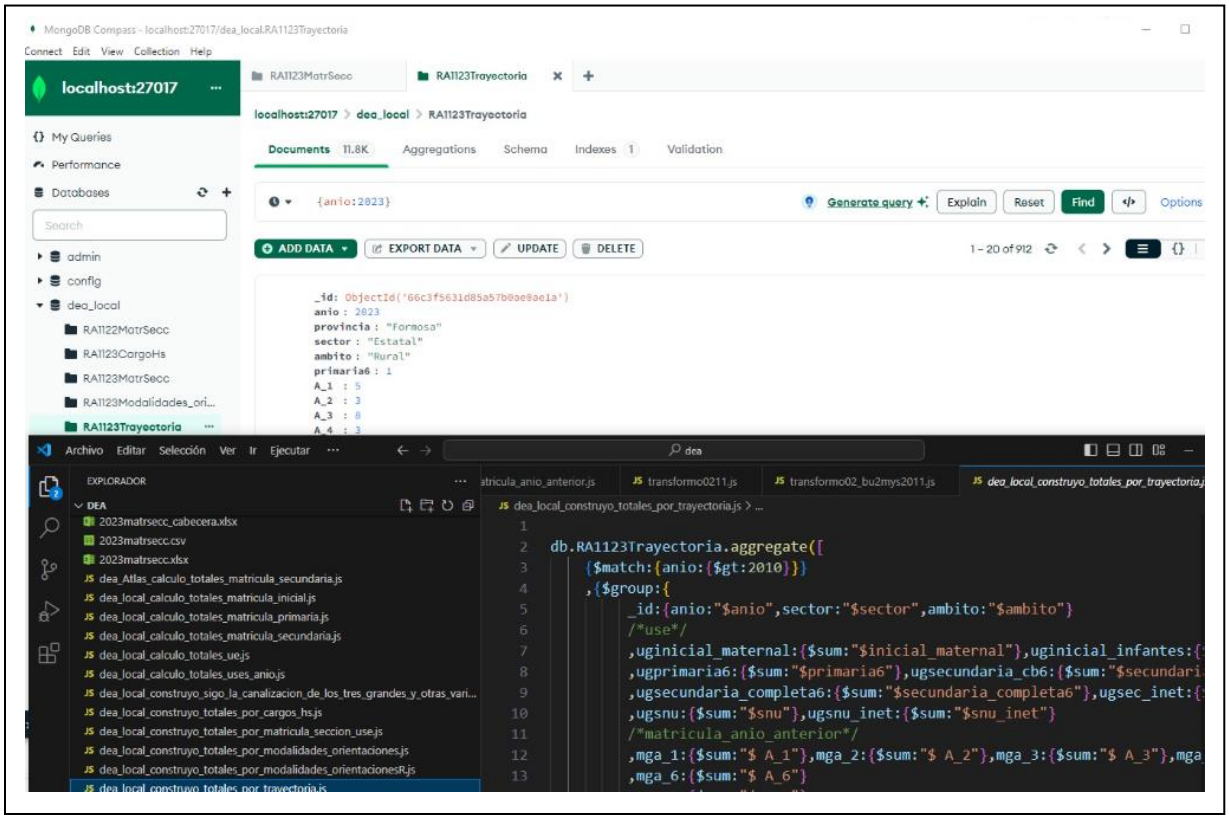


Figura 23. Agrupamiento de documentos por año, sector y ámbitos de trayectoria (11.8k documentos). Fuente propia.

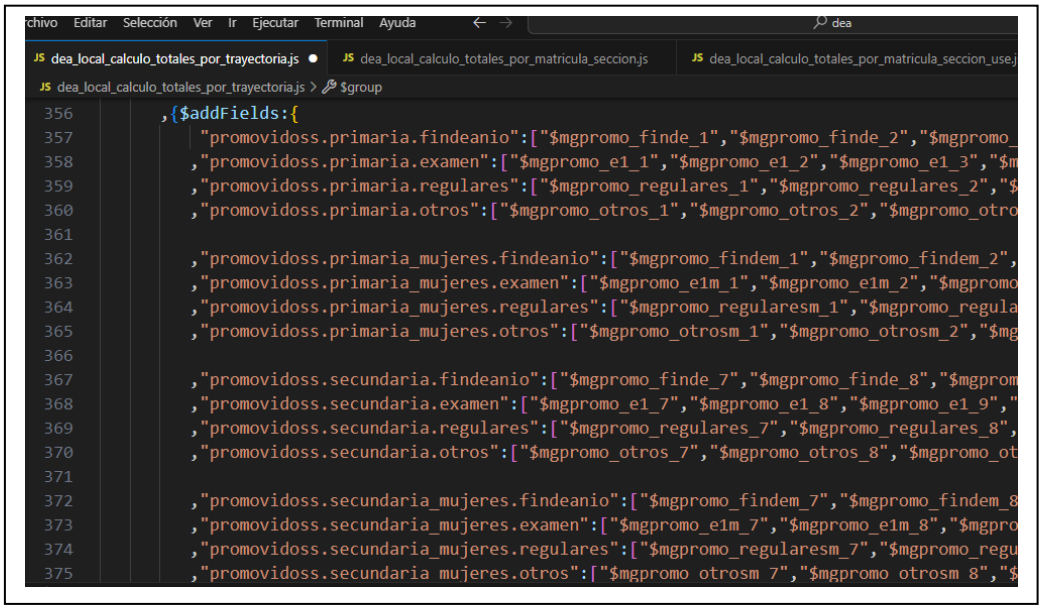


Figura 24. Subdocumentos embebidos y vectores para cada variable, sexo y nivel. Fuente propia.

En la Figura 24 se puede visualizar cómo se forman los distintos subdocumentos y vectores que componen cada variable en estudio, por ejemplo, los alumnos promovidos del año anterior. En la misma se observa cómo se construye la variable cantidad de alumnos

promovidos del año anterior, constituida por subdocumentos de “promovidos.primaria”, “promovidos.primaria_mujeres”, “promovidos.secundaria” y “promovidos. secundaria_mujeres”, y dentro de cada uno de estos subdocumentos, encontramos los subdocumentos de promovidos a fin de año, promovidos con examen, promovidos regulares y otros promovidos. A su vez, cada uno de estos documentos se componen de vectores para el manejo de los valores correspondientes a cada año de estudio. Esta construcción permite el reemplazo de más de 100 campos de la información publicada por el Ministerio de Capital Humano

En la Figura 25 se visualiza cómo se proyectan los distintos subdocumentos y se graban en modo local, en este caso en la colección dl_RA_trayectoria (significa dea_local de la colección RA trayectoria):

```
458      ,{$project:{
459          _id:1
460          ,anio:"$_id.anio"
461          ,sector:"$_id.sector"
462          ,ambito:"$_id.ambito"
463          ,use:"$uses"
464          ,matricula_anio_ant:"$matriculas_anio_ant"
465          ,matricula_anio_ant_mujeres:"$matriculas_anio_ant_mujeres"
466          ,matricula_entrados:"$matriculas_entrados"
467          ,matricula_entrados_mujeres:"$matriculas_entrados_mujeres"
468          ,matricula_salidos_con_pase:"$matriculas_salidos_con_pase"
469          ,matricula_salidos_con_pase_mujeres:"$matriculas_salidos_con_pase_mujeres"
470          ,matricula_salidos_sin_pase:"$matriculas_salidos_sin_pase"
471          ,matricula_salidos_sin_pase_mujeres:"$matriculas_salidos_sin_pase_mujeres"
472          ,matricula_ultimo_dia:"$matriculas_ultimo_dia"
473          ,matricula_ultimo_dia_mujeres:"$matriculas_ultimo_dia_mujeres"
474          ,promovidos:"$promovidoss"
475          ,egresados:"$egresadoss"  }    }
476      ,{$out: {  db:"dea_local",coll: "dl_RA_trayectoria" }    }
477
478  ]);
```

Figura 25. Proyección y grabado en forma local. Fuente propia.

Posteriormente, se ejecutan las principales agregaciones y combinaciones, como se observa en la

Figura 26. Para poder calcular las tasas, necesitamos operar con las matrículas del año anterior y del año actual. En cada año, los establecimientos informan las dos matrículas: la matrícula del año actual se informa en la colección dl_RA_matr_secc, y la matrícula del año anterior se informa en la colección dl_RA_trayectoria. Lo informado en trayectoria de este año, debe coincidir con la matrícula del año anterior, pero dado que esto no sucede, se lo debe calcular a partir de lo informado el año anterior. En la

Figura 26, se observa la combinación de ambas colecciones, tomando como campo común para unirlas, a una clave combinada compuesta por el año anterior, el sector y ámbito. Si el año es menor a 2012, es decir, es 2011, la matrícula del año anterior es lo que figura en la trayectoria de la colección dl_RA_trayectoria, ya que no se tienen bases de 2010. En cambio,

```
418  db.dl_RA_trayectoria.aggregate([
419   {$match:{anio:{$gt:2010}}}
420   ,{$lookup: { from: "dl_RA_matric_secc"
421     ,let: { traye_anio: {$subtract:["$anio",1]}
422           , traye_sector: "$sector"
423           , traye_ambito: "$ambito" }
424          /*de la coleccion dl_RA_matric_secc*/
425       ,pipeline: [
426           { $match: { $expr:
427                 { $and:
428                   [{ $eq: [ "$anio",  "$$traye_anio" ] },
429                    { $eq: [ "$sector",  "$$traye_sector" ] },
430                    { $eq: [ "$ambito",  "$$traye_ambito" ] } ]}}}
431       ]
432       ,as: "vtrayectoria_matrsecc"  }  }
433   ,{$addFields:{
434     "matriculas_anio_ant.primaria":{$cond:{if:{$lt:["$_id.anio",2012]
435       ,then:[
436         {$arrayElemAt: [ "$matricula_anio_ant.primaria", 0 ] }
```

si el año es posterior a 2011, la matrícula del año anterior se calcula por el dato informado en dl_RA_matr_secc, pero del año anterior.

Figura 26. Distintas agregaciones en modo local, tomando como clave el año anterior en la trayectoria. Fuente propia.

Se realizan algunos ajustes, como por ejemplo: si el año es 2011, la matrícula de mujeres del año anterior se toma de la colección dl_RA_trayectoria, y si es posterior, se calcula la matrícula del año anterior de mujeres, por medio de la diferencia entre dicha matrícula y la matrícula de varones de la colección dl_RA_matric_secc, pero del documento del año anterior: `let: { traye_anio: {$subtract:["$anio",1]} , traye_sector: "$sector" , traye_ambito: "$ambito" }`, como se observa en la Figura 27:

```
JS dea_local_construyo_totales_por_trayectoria.js > $addFields > "matriculas_anio_ant_mujeres.secundaria" > $cond > else > $subtract > $arrayElemAt
481   ,{$addFields:{
519           }
520
521         ,"matriculas_anio_ant_mujeres.secundaria":
522           {$cond:
523              {if:{$lt:["$_id.anio",2012]}
524              ,then:[
525                      {$arrayElemAt: [ "$matricula_anio_ant_mujeres.secundaria", 0 ] }
526                     ,{$arrayElemAt: [ "$matricula_anio_ant_mujeres.secundaria", 1 ] }
527                     ,{$arrayElemAt: [ "$matricula_anio_ant_mujeres.secundaria", 2 ] }
528                     ,{$arrayElemAt: [ "$matricula_anio_ant_mujeres.secundaria", 3 ] }
529                     ,{$arrayElemAt: [ "$matricula_anio_ant_mujeres.secundaria", 4 ] }
530                     ,{$arrayElemAt: [ "$matricula_anio_ant_mujeres.secundaria", 5 ] }
531                     ,{$arrayElemAt: [ "$matricula_anio_ant_mujeres.secundaria", 6 ] }
532
533                    ]
534              ,else:
535                 [{$subtract:[
536                    {$arrayElemAt:[{$arrayElemAt:["$vtrayectoria_matrsecc.matricula.secundaria",0]},0]},
537                    {$arrayElemAt:[{$arrayElemAt:["$vtrayectoria_matrsecc.matricula_varones.secundaria",0]},0]}
538                    ]}
539                   ,{$subtract:[
540                     {$arrayElemAt:[{$arrayElemAt:["$vtrayectoria_matrsecc.matricula.secundaria",0]},1]},
541                     {$arrayElemAt:[{$arrayElemAt:["$vtrayectoria_matrsecc.matricula_varones.secundaria",0]},1]}
542                       ]}
```

Figura 27. Realización de distintos ajustes. Fuente propia.

En la Figura 28 se controlan los totales de documentos resultantes (las colecciones que comienzan con dl_RA_... deberían tener 52 documentos, ya que son 4 por año, por un total de 13 años).

```
dea_local> db.dl_RA_matric_secc23.find().count()
48
dea_local> db.dl_RA_cargoshs23.find().count()
52
dea_local> db.dl_RA_trayectoria23.find().count()
52
dea_local> db.dl_vRA_UE.find().count()
52
dea_local>
```

Figura 28. Control de cantidad de documentos en cada colección. Fuente propia.

En este caso, se deben revisar las claves por año, sector y ámbito de la colección dl_RA_matric_secc23.

```
dea_local> db.dl_RA_matric_secc.aggregate([{$group:{_id:{anio:"$anio",sector:"$sector"}}}])
[    dl_RA_matric_secc
  { _id: { anio: 2021, sector: 'Estatal' } },
  { _id: { anio: 2018, sector: 'Social/cooperativa'
  { _id: { anio: 2017, sector: 'Privado' } },
  { _id: { anio: 2016, sector: 'Privado' } },
  { _id: { anio: 2012, sector: 'Social/cooperativa'
  { _id: { anio: 2011, sector: 'Privado' } },
  { _id: { anio: 2013, sector: 'Privado' } },
  { _id: { anio: 2014, sector: 'Estatal' } },
  { _id: { anio: 2018, sector: 'Privado' } },
  { _id: { anio: 2016, sector: 'Estatal' } },
  { _id: { anio: 2020, sector: 'Privado' } },
  { _id: { anio: 2022, sector: 'Privado' } },
  { _id: { anio: 2014, sector: 'Privado' } },
  { _id: { anio: 2019, sector: 'Privado' } },
  { _id: { anio: 2021, sector: 'Privado' } },
  { _id: { anio: 2012, sector: 'Estatal' } },
  { _id: { anio: 2018, sector: 'Estatal' } },
  { _id: { anio: 2015, sector: 'Social/cooperativa'
  { _id: { anio: 2016, sector: 'Social/cooperativa'
  { _id: { anio: 2012, sector: 'Privado' } }
]
Type "it" for more
dea_local> it
[
  { _id: { anio: 2015, sector: 'Privado' } },
  { _id: { anio: 2013, sector: 'Social/cooperativa'
  { _id: { anio: 2017, sector: 'Estatal' } },
  { _id: { anio: 2015, sector: 'Estatal' } },
  { _id: { anio: 2020, sector: 'Estatal' } },
  { _id: { anio: 2022, sector: 'Estatal' } },
  { _id: { anio: 2011, sector: 'Estatal' } },
  { _id: { anio: 2013, sector: 'Estatal' } },
  { _id: { anio: 2014, sector: 'Social/cooperativa'
  { _id: { anio: 2019, sector: 'Estatal' } }
```

```
dl_RA_trayectoria
dea_local> db.dl_RA_trayectoria.aggregat
[
  { _id: { anio: 2022, sector: 'Estatal'
  { _id: { anio: 2015, sector: 'Estatal'
  { _id: { anio: 2011, sector: 'Privado'
  { _id: { anio: 2021, sector: 'Privado'
  { _id: { anio: 2018, sector: 'Privado'
  { _id: { anio: 2016, sector: 'Estatal'
  { _id: { anio: 2014, sector: 'Estatal'
  { _id: { anio: 2013, sector: 'Privado'
  { _id: { anio: 2017, sector: 'Privado'
  { _id: { anio: 2015, sector: 'Privado'
  { _id: { anio: 2012, sector: 'Privado'
  { _id: { anio: 2016, sector: 'Privado'
  { _id: { anio: 2013, sector: 'Estatal'
  { _id: { anio: 2011, sector: 'Estatal'
  { _id: { anio: 2017, sector: 'Estatal'
  { _id: { anio: 2019, sector: 'Estatal'
  { _id: { anio: 2018, sector: 'Estatal'
  { _id: { anio: 2021, sector: 'Estatal'
  { _id: { anio: 2012, sector: 'Estatal'
  { _id: { anio: 2022, sector: 'Privado'
]
Type "it" for more
dea_local> it
[
  { _id: { anio: 2020, sector: 'Privado'
  { _id: { anio: 2019, sector: 'Privado'
  { _id: { anio: 2014, sector: 'Privado'
  { _id: { anio: 2020, sector: 'Estatal'
]
dea_local>
```

```
dl_RA_cargoshs
dea_local> db.dl_RA_cargoshs.aggregate([{$group:{
[
  { _id: { anio: 2022, sector: 'Estatal' } },
  { _id: { anio: 2011, sector: 'Privado' } },
  { _id: { anio: 2015, sector: 'Estatal' } },
  { _id: { anio: 2013, sector: 'Privado' } },
  { _id: { anio: 2018, sector: 'Privado' } },
  { _id: { anio: 2016, sector: 'Estatal' } },
  { _id: { anio: 2014, sector: 'Estatal' } },
  { _id: { anio: 2021, sector: 'Privado' } },
  { _id: { anio: 2017, sector: 'Privado' } },
  { _id: { anio: 2015, sector: 'Privado' } },
  { _id: { anio: 2012, sector: 'Privado' } },
  { _id: { anio: 2016, sector: 'Privado' } },
  { _id: { anio: 2013, sector: 'Estatal' } },
  { _id: { anio: 2011, sector: 'Estatal' } },
  { _id: { anio: 2021, sector: 'Estatal' } },
  { _id: { anio: 2019, sector: 'Estatal' } },
  { _id: { anio: 2018, sector: 'Estatal' } },
  { _id: { anio: 2012, sector: 'Estatal' } },
  { _id: { anio: 2017, sector: 'Estatal' } },
  { _id: { anio: 2022, sector: 'Privado' } }
]
Type "it" for more
dea_local>
[
  { _id: { anio: 2020, sector: 'Privado' } },
  { _id: { anio: 2019, sector: 'Privado' } },
  { _id: { anio: 2014, sector: 'Privado' } },
  { _id: { anio: 2020, sector: 'Estatal' } }
]
```

Figura 29. Consulta de claves (año, y sector) de tres colecciones resúmenes. Fuente propia.

En la Figura 29 se ve que solamente en la colección dl_RA_matric_secc, aparece el sector “Social/cooperativa” -en los años 2012, 2013, 2014, 2015, 2016 y 2018-. Dado que este sector no aparece en las otras colecciones, significa que sólo tiene matrículas y secciones, pero no tiene cargos u horas, ni tampoco trayectoria. Consultando un poco más, se toma conocimiento de que todas las matrículas corresponden al nivel inicial: lactantes, deambuladores, sala de 2, sala de 3 y sala de 4 años. Además, en los anuarios estadísticos [28], en el año 2018, se consideran 2 sectores solamente (Estatal y Privado), y la matrícula correspondiente al sector “Social/cooperativa” de la

Figura **30** forma parte del total del sector “Privado” en el anuario 2018, razón por la cual en este trabajo de tesis también vamos a considerar esas matrículas del nivel inicial para ese sector.

Año 2018: Educación Común. Nivel Inicial
2.2.3.2.Alumnos del sector de gestión privado por ciclo y año de estudio según división político-territorial.

División político-territorial	Ciclo				
	Total	Jardín Maternal	Jardín de Infantes		
			Sala de 3 años	Sala de 4 años	Sala de 5 años
Total País	605.417				
Ciudad de Buenos Aires	69.962				
Buenos Aires	295.705				
Conurbano	200.298				
Buenos Aires Resto	95.407				
Catamarca	3.739				
Córdoba	37.514				
Corrientes	8.789				
Chaco	7.999				
Chubut	4.498				
Entre Ríos	16.381				
Formosa	5.837				
Jujuy	8.167				
La Pampa	3.139				
La Rioja	2.423				
Mendoza	31.375				
Misiones	12.286				
Neuquén	4.264				
Río Negro	6.158				
Salta	8.399				
San Juan	8.220				
San Luis	3.581				
Santa Cruz	2.711				
Santa Fe	39.211				
Santiago del Estero	6.459				
Tucumán	17.010				
Tierra del Fuego	1.590				

Fuente: Relevamiento Anual 2018, DIEE, MECCyT.

dea_local_calculo_totales_matricula_inicial.js × | dea_local_calculo_totales_por_trayectoria.js | dea_local_calculo_totales_por_matricula_seccion_us

dea_local_calculo_totales_matricula_inicial.js > ...

```
1  db.dl_RA_matric_secc.aggregate([
2      {$match:{anio:2018}},
3      {$group:{
4          _id:{anio:"$anio",sector:"$sector"}
5          ,matr_tot_inicial:{$sum:{$sum:["$matricula.inicial"]}}
6         }
7      }
8      ,{$project:{_id:1,matr_tot_inicial:1 } }
9  ]
10 ).forEach(function(r){console.log(' id',r._id,' Total_matr_inicial:',r.matr_tot_inicial
```

mongosh mongodb://127.0.0.1:27017/?directConnection=true&serverSelectionTimeoutMS=2000

```
11
12
13 dea_local> load("c:/users/vhruc/dea/dea_local_calculo_totales_matricula_ini
14 cial.js")
15  id { anio: 2018, sector: 'Privado' }  Total_matr_inicial: 3986
16  id { anio: 2018, sector: 'Social/cooperativa' }  Total_matr_inicial: 1851
17  id { anio: 2018, sector: 'Estatal' }  Total_matr_inicial: 16770
18 true
19 dea_local>
```

Figura 30. Matrícula total del sector privado 2018-Consulta en modo local. Fuente: DIIE y fuente propia.

Controlamos nuevamente la matrícula total de primer año de secundaria, ya no de la colección individual de matrículas y secciones de 2018 como en la subsección 3.2.2, sino de la colección resultante del proceso local, como se observa en la Figura 31.

```
db.dl_RA_matric_secc23.aggregate([
    {$match:{anio:2018}},
    {$group:{
        _id:{anio:"$anio"}
        ,matr_tot_secundaria:{$sum:{$sum:[{$slice:["$matricula.secundaria",
      } }
    ,{$project:{_id:1,matr_tot_secundaria:1 } }
 ]).forEach(function(r){console.log(' id',r._id,' Total_matr_1er.año de secu
```

```
mongosh mongodb://127.0.0.1:27017/?directConnection=true&serverSelectionTimeoutMS=2000

dea_local> load ("C:/Users/vhruc/dea/dea_local_calculo_totales_matricula_secundaria.js")
 id { anio: 2018 }  Total_matr_1er.año de secundaria: 14109
true
dea_local>
```

Figura 31. Control de matrícula de primer año de secundaria de 2018. Fuente propia.

4.1.2 Combinación de colecciones por año, sector y ámbito

Para combinar las colecciones, continuamos la canalización, utilizando $loockup en la agregación, ya que todas las colecciones tienen el mismo _id, como se observa en la Figura 32:

```
db.dl_RA_matric_secc23.aggregate([
    {$lookup:
        {from:"dl_RA_trayectoria23",localField:"_id",foreignField:"_id",as:"vmatrsecc_trayectoria" } }
    ,{$unwind:"$vmatrsecc_trayectoria"}
    ,{$lookup:
        {from:"dl_RA_cargoshs23",localField:"_id",foreignField:"_id",as:"vmatrsecc_cargoshs" } }
    ,{$unwind:"$vmatrsecc_cargoshs"}
    ,{$lookup:
      {from:"dl_vRA_UE",localField:"_id",foreignField:"_id",as:"vmatrsecc_vune" } }
    ,{$unwind:"$vmatrsecc_vune"}
    ,{$lookup:
      {from:"dl_RA_modalidades_espadutec23",localField:"_id",foreignField:"_id",as:"vmodal" } }
    ,{$unwind:"$vmodal"}
    ,{$project:{
        _id:1
        ,anio:"$_id.anio"
        ,sector:"$_id.sector"
        ,ambito:"$_id.ambito"
        ,use:"$use"
        ,matricula:"$matricula"
        ,matricula_varones:"$matricula_varones"
        ,repitientes:"$repitientes"
        ,sobreedad:"$sobreedad"
```

Figura 32. Combinación de colecciones por medio del mismo _id. Fuente propia.

4.2 Completamiento de datos e información

4.2.1 Completamiento de la variable "Unidades Educativas" de los 4 niveles de la educación común

Al no poder procesar la clave enmascarada, se pierde el concepto de unidades educativas, ya que no se puede determinar la relación CUE-anexo con la clave enmascarada. Las unidades educativas son unidades con alumnos y docentes y que tienen un director; dichas unidades, en 2022 en Formosa, fueron 908 en los 4 niveles que se tienen en las bases de datos. Éste, resulta un valor muy distinto a las 1.448 unidades de servicio educativo que tienen ofertas educativas concretadas, con estructuras organizativas, alumnos, profesores e infraestructuras (edificios), tengan o no tengan director, y que, además, se pueden diferenciar en las bases de datos por tener documentos distintos (registros distintos en una base de datos SQL).

Las unidades educativas se cuentan por los primeros 7 caracteres del CUE-anexo, ya que los dos últimos, hacen referencia a los anexos. Para resumir, las unidades educativas son estructuras organizacionales que tienen un director. Para conseguir la cantidad de unidades educativas, se procesaron localmente los datos de los anuarios. Posteriormente, se continuó con la canalización, usando como clave: año, sector y ámbito, para completar los objetos JSON también localmente procesados. Se procedió a controlar los resultados publicados, utilizando distintas herramientas y ayudas de MongoDB [29] [30] [31].

En la Figura 33 se muestra la relación entre datos del anuario 2022, indicando las localizaciones (documentos almacenados en la base de datos NoSQL) y unidades de servicios educativos correspondientes a resultados procesados localmente.

Año 2022: Educación Común. Resumen

2.1.2. Localizaciones y unidades de servicio por nivel / ciclo de enseñanza según división político-territorial.

División político-territorial	Localizaciones	Unidades de Servicio por Nivel / Ciclo de enseñanza									
		Total	Inicial			Primaria		Secundaria			Superior no Universitario
			Sólo Jardín maternal	Sólo Jardín de infantes	Ambos ciclos	6 años	7 años	Sólo Ciclo básico	Sólo Ciclo orientado	C. básico y C. orientado	
Total País	50.277	60.413	487	17.694	2.979	13.198	9.675	1.152	127	12.385	2.716
Ciudad de Buenos Aires	1.782	2.552	15	474	411	2	891	3	3	494	259
Buenos Aires	16.314	16.722	152	3.587	1.800	5.881		155	38	4.477	632
Conurbano	7.579	7.840	58	1.966	566	2.493		71	20	2.415	251
Buenos Aires Resto	8.735	8.882	94	1.621	1.234	3.388		84	18	2.062	381
Catamarca	997	1.043	2	323	1	449		39	4	191	34
Córdoba	4.542	5.292		1.897	26	2.064		22	1	1.028	254
Corrientes	1.699	2.184		778	58	905		7	1	332	103
Chaco	2.211	2.548		710	66		1.179	8	2	405	178
Chubut	680	760	22	199	64	243		12		171	49
Entre Ríos	2.095	3.115	2	1.088	87	1.266		15	2	555	100
Formosa	1.397	1.448		456	25	581		63	1	273	49

```
mongosh mongodb://127.0.0.1:27017/?directConnect...
dea_local> db.RA1122MatrSecc.find({año:2
022,$or:[{inicial_infantes:1},{primaria6
:1},{secundaria_cb6:1},{secundaria_compl
eta6:1},{snu:1}]}).count()
1397
dea_local> load("c:/users/vhruc/dea/dea_
local_calculo_totales_uses_anio.js")
 id { anio: 2022 }  Total_unidades_de_se
rvicios_educativos: 1448
```

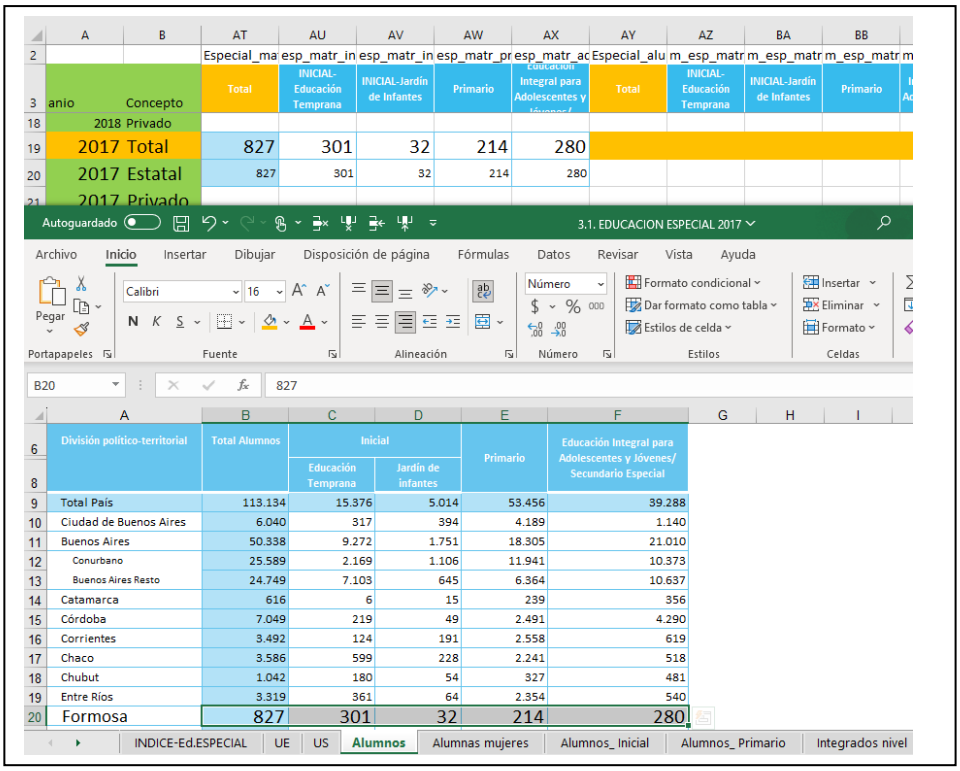
```
Selección  Ver  Ir  ...   dea
_calculo_totales_uses_anio.js ●   dea_local_calculo_totales_por_trayectoria.js   dea_local_sigo_la_canalizacion.js   dea_local
al_calculo_totales_uses_anio.js > ...
db.dl_RA_matric_secc.aggregate([{$match:{anio:2022}},{$group:{_id:{anio:"$anio"}
  ,tot_uses_anio:{$sum:{$sum:["$use.inicial_ambos_mat_e_inf","$use.inicial_infantes"
  ,"$use.primaria6","$use.secundaria_cb6","$use.secundaria_completa6","$use.snu"]}}}
  ,{$project:{_id:1,tot_uses_anio:1 } }
 ]).forEach(function(r){console.log(' id',r._id,
 ' Total_unidades_de_servicios_educativos:' ,r.tot_uses_anio ); });
Lín. 7, col. 41
```

Figura 33. Relación entre localizaciones y unidades de servicios educativos del anuario 2022, con documentos bajados en bases de datos NoSQL y resultados procesados localmente. Fuente propia y anuario.

4.2.2 Completamiento de las modalidades

El Ministerio de Capital Humano no publica las bases de datos referidas a cada uno de los establecimientos de educación especial y a educación permanente de jóvenes y adultos, razón por la cual se tomaron datos de los distintos anuarios publicados, pero que están incompletos. Por ejemplo, en adultos, la oferta “formación profesional” representa la mitad de la matrícula (14327 alumnos en 2017), y se informa una modalidad denominada rural, en donde no aparecen los cargos ni las horas referidas a formación profesional, razón por la que se puede coleccionar información únicamente por año y sector, y no por ámbito rural.

El Ministerio de Capital Humano publica la información en tres cuadros distintos para cada variable: un cuadro por el total, y un cuadro por cada sector, y no trata el ámbito. En la Figura 34 se observa la colección de información de matrícula de educación especial del año 2017. En matrícula existen tres grupos de columnas (inicial, primario y adolescentes y jóvenes); sin embargo, se publican solamente 2 solapas (Alumnos_inicial y Alumnos_primario, y después se pasa a alumnos integrados); son resúmenes incompletos. A medida que uno retrocede en el tiempo, se observan faltantes de información más acentuados.



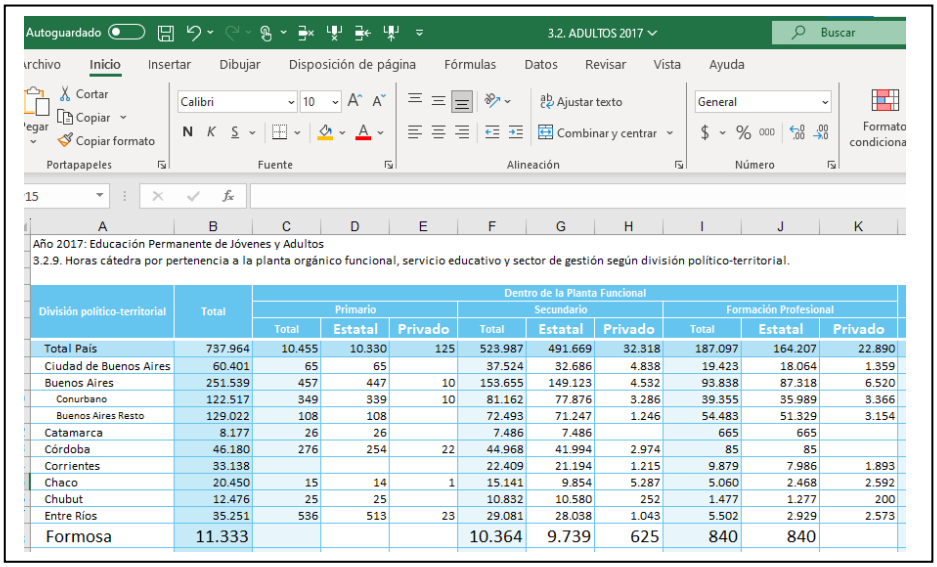
anio	Concepto	Total	INICIAL-Educación Temprana	INICIAL-Jardín de Infantes	Primario	Educación Integral para Adolescentes y Jóvenes/	Total	INICIAL-Educación Temprana	INICIAL-Jardín de Infantes	Primario
2018	Privado									
2017	Total	827	301	32	214	280				
2017	Estatal	827	301	32	214	280				
2017	Privado									

División político-territorial	Total Alumnos	Inicial		Primario	Educación Integral para Adolescentes y Jóvenes/ Secundario Especial
		Educación Temprana	Jardín de infantes		
Total País	113.134	15.376	5.014	53.456	39.288
Ciudad de Buenos Aires	6.040	317	394	4.189	1.140
Buenos Aires	50.338	9.272	1.751	18.305	21.010
Conurbano	25.589	2.169	1.106	11.941	10.373
Buenos Aires Resto	24.749	7.103	645	6.364	10.637
Catamarca	616	6	15	239	356
Córdoba	7.049	219	49	2.491	4.290
Corrientes	3.492	124	191	2.558	619
Chaco	3.586	599	228	2.241	518
Chubut	1.042	180	54	327	481
Entre Ríos	3.319	361	64	2.354	540
Formosa	827	301	32	214	280

Figura 34. Colección de datos de la modalidad especial por año y sector. Fuente: anuarios y fuente propia.

En la modalidad adultos, en el resumen de información de la variable horas, el Ministerio de Capital Humano mezcla los sectores (estatal y privado), como se observa en la Figura 35, esta situación imposibilita tomar la información de manera automática, debiendo acomodar en la planilla de Excel los valores, en las filas correspondientes.

Sería deseable que el Ministerio de Capital Humano publique los datos de las modalidades, en bases de datos, como lo hace con la educación común.



Año 2017: Educación Permanente de Jóvenes y Adultos

3.2.9. Horas cátedra por pertenencia a la planta orgánico funcional, servicio educativo y sector de gestión según división político-territorial.

División político-territorial	Total	Dentro de la Planta Funcional								
		Primario			Secundario			Formación Profesional		
		Total	Estatal	Privado	Total	Estatal	Privado	Total	Estatal	Privado
Total País	737.964	10.455	10.330	125	523.987	491.669	32.318	187.097	164.207	22.890
Ciudad de Buenos Aires	60.401	65	65		37.524	32.686	4.838	19.423	18.064	1.359
Buenos Aires	251.539	457	447	10	153.655	149.123	4.532	93.838	87.318	6.520
Conurbano	122.517	349	339	10	81.162	77.876	3.286	39.355	35.989	3.366
Buenos Aires Resto	129.022	108	108		72.493	71.247	1.246	54.483	51.329	3.154
Catamarca	8.177	26	26		7.486	7.486		665	665	
Córdoba	46.180	276	254	22	44.968	41.994	2.974	85	85	
Corrientes	33.138				22.409	21.194	1.215	9.879	7.986	1.893
Chaco	20.450	15	14	1	15.141	9.854	5.287	5.060	2.468	2.592
Chubut	12.476	25	25		10.832	10.580	252	1.477	1.277	200
Entre Ríos	35.251	536	513	23	29.081	28.038	1.043	5.502	2.929	2.573
Formosa	11.333				10.364	9.739	625	840	840	

Figura 35. Recorte de la información publicada con los sectores en un mismo cuadro. Fuente: Anuario 2017.

4.2.3 Completamiento de información de Operativos Aprender

4.2.3.1 Completamiento de datos de Aprender de Primario 2021

Recientemente se difundieron resultados de Aprender 2021, que tuvo un carácter censal, y que incluyó datos de alumnos del nivel primario de pruebas Aprender 2016, 2018 y 2021. Si bien existen datos a nivel de alumno de Aprender 2021, para este trabajo no se avanzó en el proceso y análisis de dicha base de datos ya que el factor de escalamiento y las preguntas de anclaje entre las distintas pruebas para aplicar las técnicas de escalamiento y poder comparar pruebas de distintos años, no son conocidas. Como los alumnos pasan de año, tampoco tenemos grupo de anclaje entre las pruebas. Para el completamiento de la información y su posterior análisis, se tomaron los 24 informes jurisdiccionales proporcionados por el Ministerio de Capital Humano [32]. Se cargó la información en formato JSON, para que posteriormente pueda ser procesado en Atlas MongoDB.

La estructura de cada documento en MongoDB es la siguiente: cada documento se asocia con una jurisdicción, en donde se cargan datos de la población de estudiantes que realizaron la prueba: la edad, el género, los estudiantes por sector, ámbito y género. A los resultados de Matemática y Lengua se los trata como documentos embebidos. A su vez, dentro de cada asignatura, las distintas variables constituyen subdocumentos y vectores. El informe de cada jurisdicción tiene cerca de 70 páginas, y se cargaron las 24 jurisdicciones para su procesamiento en Atlas de MongoDB, como se observa en la Figura 36.

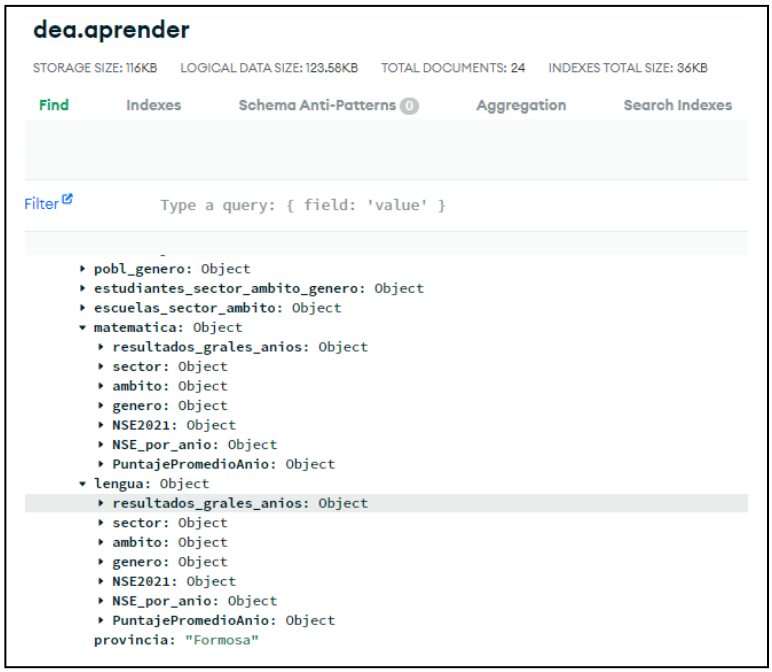


Figura 36. Estructura de documento MongoDB para Aprender 2021. Fuente propia.

Por ejemplo, los resultados de Lengua por niveles socioeconómicos y año de la prueba Aprender 2021 de una jurisdicción, se observan en la Figura 37.

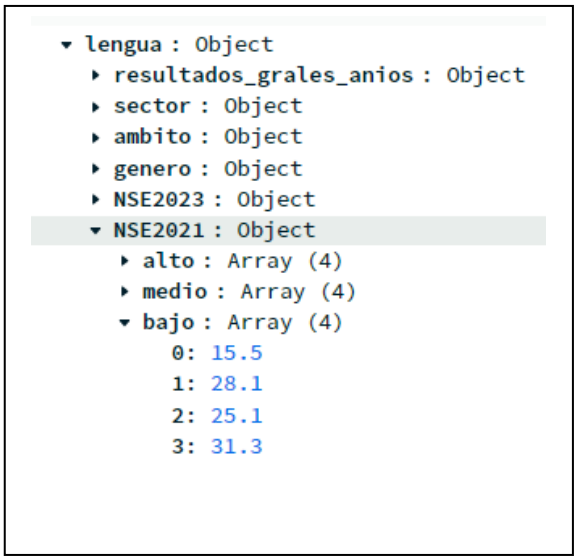


Figura 37. Ejemplo de un subdocumento de lengua por nivel socioeconómico. Fuente: Aprender 2021.

4.2.3.2 Completamiento de datos de Aprender 2022 de Secundario

También se cargaron los objetos JSON de las 24 jurisdicciones de 65 páginas, y para diferenciar los documentos de primaria, los _id de los documentos de secundaria, comienzan con APR22S.

Los resultados del 2022 fueron publicados junto con los resultados de las pruebas ONE 2013, y Aprender 2016, 2017, 2019 y 2022. El inicio de un subdocumento JSON de Lengua, de la Provincia de Formosa se observa en el Script 8.

"lengua":{"resultados_grales_anios":{
"anio_2013":[{"$numberDouble":"4"},{"$numberDouble":"29"},{"$numberDouble":"25.4"},{"$numberDouble":"41.6"}]
,"anio_2016":[{"$numberDouble":"4.5"},{"$numberDouble":"37.5"},{"$numberDouble":"27.8"},{"$numberDouble":"30.2"}]
,"anio_2017":[{"$numberDouble":"7.2"},{"$numberDouble":"37.7"},{"$numberDouble":"25.1"},{"$numberDouble":"30"}]
,"anio_2019":[{"$numberDouble":"6.7"},{"$numberDouble":"36.2"},{"$numberDouble":"24.8"},{"$numberDouble":"32.3"}]
,"anio_2022":[{"$numberDouble":"8.4"},{"$numberDouble":"36.1"},{"$numberDouble":"22.9"},{"$numberDouble":"32.6"}] }

,"sector": { "estatal":{ …

Script 8. Ejemplo de un subdocumento JSON de lengua de un operativo Aprender 2022. Fuente propia.

4.2.3.3 Completamiento de datos de Aprender 2023 de Primaria – Operativo censal

Los resultados se publicaron el 07/12/2023, juntamente con pruebas ONE 2013, Aprender 2016, 2018, 2021, 2022 (que fue muestral), 2023. En este caso, sólo se modificaron los objetos JSON de Aprender primaria 2021, como se observa en la Figura 38.

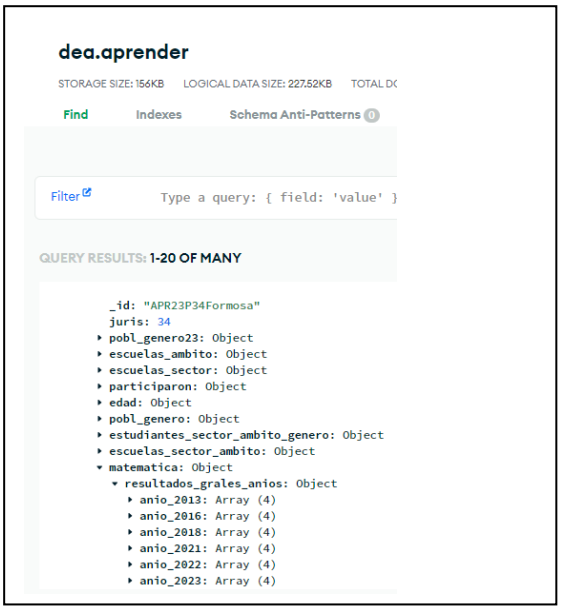


Figura 38. Estructura de un subdocumento de matemática, y dentro de este, la variable correspondiente a “resultados generales” de Aprender 2023.

En este caso, sólo se agregaron resultados de los operativos ONE 2013 y Aprender 2022 muestral, y Aprender 2023 censal, para modificar el objeto JSON, en cada subdocumento y para cada una de las variables.

4.2.4 Completamiento de datos geoespaciales

A los datos públicos del padrón de establecimientos educativos publicado en Excel, le faltan los datos geoespaciales para poder realizar mapas.

En un principio se tomaron los datos del mapa interactivo del Ministerio Capital Humano , que consiste en una implementación en donde se pueden consultar algunas ubicaciones de las escuelas en http://mapa.educacion.gob.ar/mapa-interactivo y en https://mapa.educacion.gob.ar/maps/cobertura-alumnos-nominalizados; pero aún faltaban actualizar algunos datos, y además, a partir de la tercera consulta o acercamiento, el servidor se volvía muy lento, razón por la cual para este trabajo de tesis se implementó una API (Interfaz de Programación de Aplicaciones). Los datos de latitud y longitud de los establecimientos educativos fueron cargados en forma manual al padrón, consultando distintas infraestructuras de datos espaciales. Cuando la carga manual de las localizaciones de las instituciones educativas estaba en más de un 70%, se encontraron otras implementaciones de mucha actualidad, como idef.formosa.gov.ar [33], pero dado que la carga estaba avanzada, se prefirió no utilizarlas y continuar con el trabajo manual.

4.2.4.1 Utilización de herramientas de visualización y carga de datos

Si bien se puede utilizar el Google Maps para poder ubicar las instituciones educativas en el mapa, para facilitar la tarea en un principio se utilizó una API (por sus siglas en inglés Application Programming Interface) o interfaz de programación de aplicaciones, que consiste en conjunto de reglas o protocolos para permitir que diferentes programas de software se comuniquen entre sí, y que fuera desarrollada en este caso, por el Instituto Geográfico Nacional, como se observa en la Figura 39.

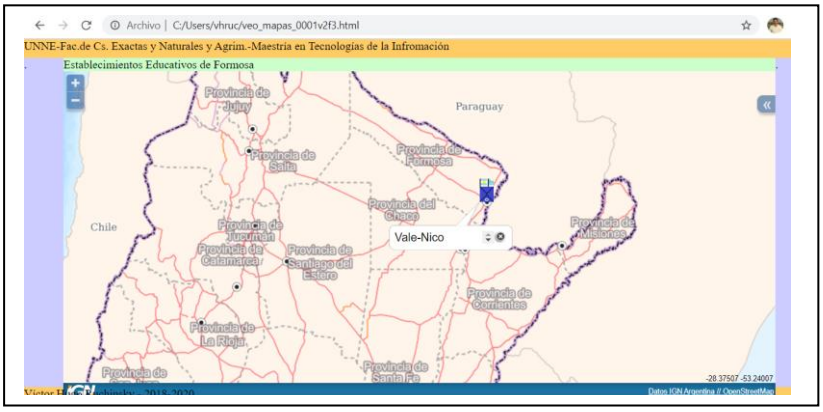


Figura 39. API del Instituto Geográfico Nacional. Fuente: IGN.

Existe también una biblioteca, Leaflet.js [34], que tiene una información más completa que la anterior, como ser el nombre de los barrios y con una apariencia más realista; además, para no tapar la información del mapa detrás de la etiqueta, se pueden trabajar con las etiquetas con transparencia [35].

Para confirmar las ubicaciones de las escuelas de Formosa capital, aprovechando la pandemia por COVID-19, se recorrieron todas las instituciones educativas de Formosa Capital, enviando su posición vía WhatsApp. Además, es posible consultar en Google Maps el color de los techos (color azul) de la gran mayoría de las escuelas del interior, pero no de todas, por falta de actualización del Google Maps. Con ayuda del visualizador, se desarrolló una consulta para obtener latitud y longitud de las escuelas, como se observa en la Figura 40.

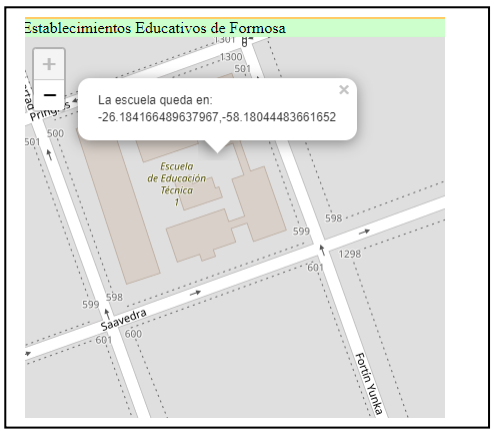


Figura 40. Consulta de latitud y longitud. Fuente: leaflet.js y propia.

También, para referenciar y diferenciar los establecimientos, y facilitar la carga, se crearon íconos de diferentes, como se observa en la Figura 41:

a) Por el color:

- lila claro, para el nivel inicial
- celeste, para el nivel primario

- azul, para el nivel secundario
- lila fuerte, para el nivel superior

b) Por un orificio en los íconos de los establecimientos educativos:

- el ícono del nivel inicial tiene el orificio en la parte superior
- el nivel primario lo tiene a la derecha
- el nivel secundario lo tiene en la parte inferior
- el nivel superior lo tiene a la izquierda

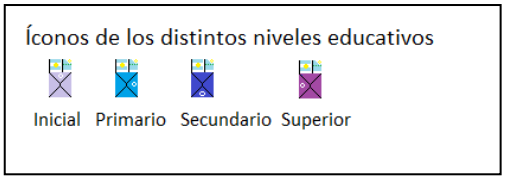


Figura 41. Utilización de íconos para referenciar y diferenciar establecimientos. Fuente propia.

Los distintos colores de los establecimientos, en los diferentes niveles, facilitan la carga y su control, como se observa en la Figura 42.

Figura 42. Carga de niveles diferenciada por colores en Leafletjs.com. Fuente propia.

Es posible que en un par de manzanas existan muchas ofertas educativas, sobre todo en las localidades del interior, como se observa en la Figura 43.

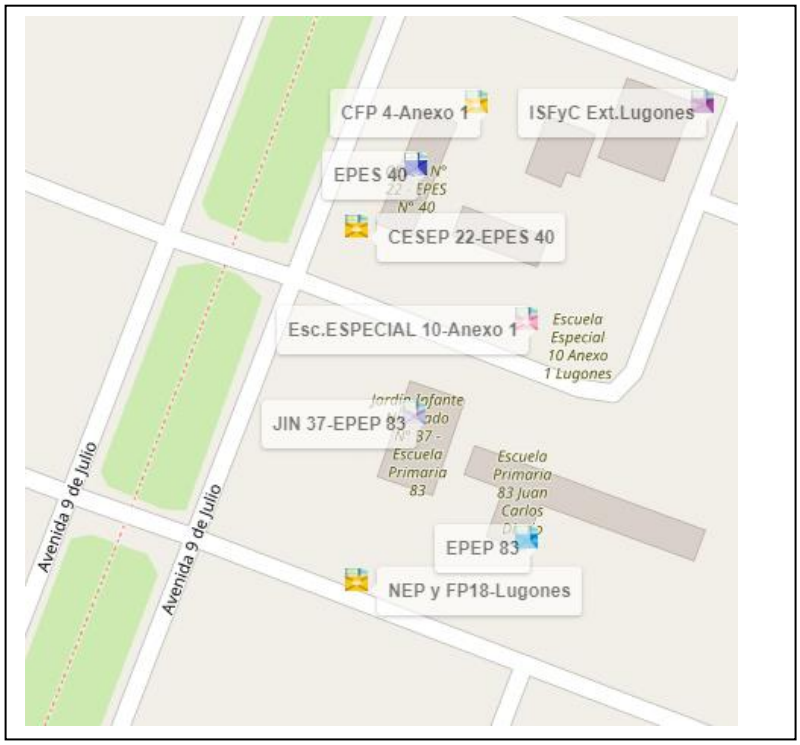


Figura 43. Ofertas educativas en las localidades del interior. Fuente propia.

4.2.4.2 Transformación de latitud y longitud

MongoDB toma la latitud y longitud como cadenas de caracteres: `lati1:"25.2144", y longi1:"-58.5135"`, como se observa en la Figura 44 y eso constituye un problema dado que se necesita que sean datos de tipo numéricos, de doble precisión.

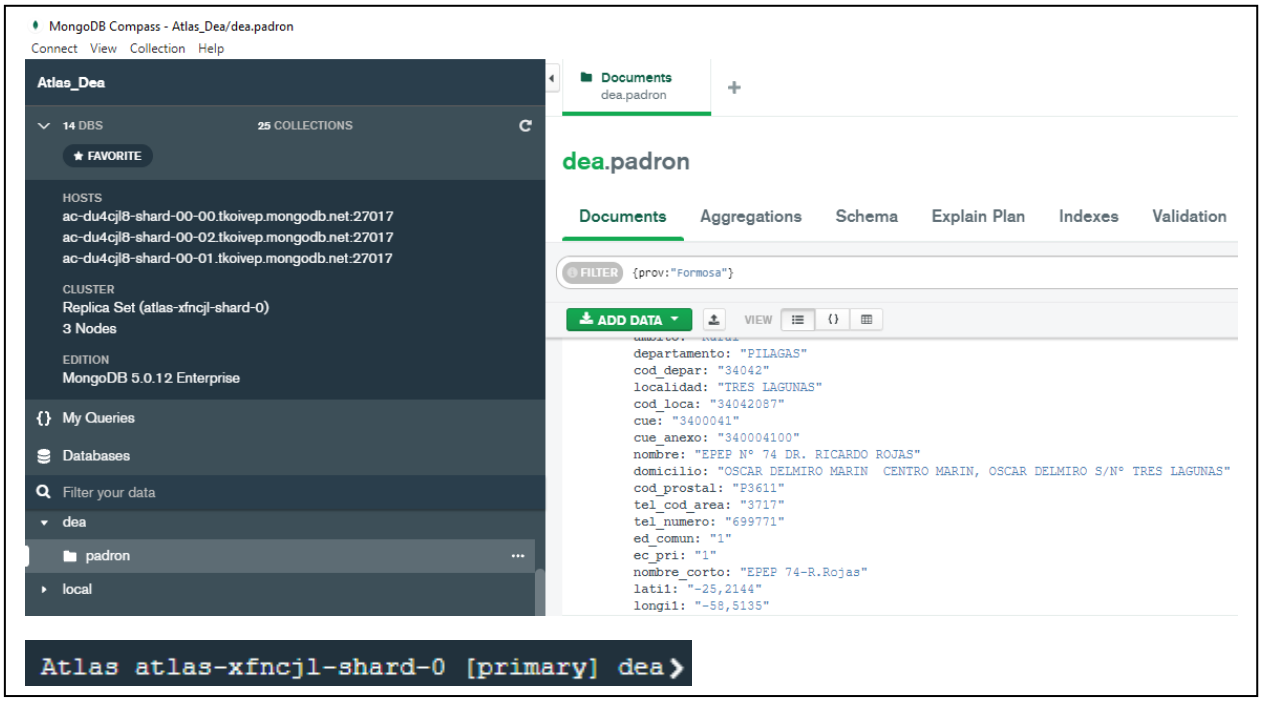


Figura 44. Al mover los datos, la latitud y longitud se vuelven alfanuméricas. Fuente propia.

Para intentar aproximar a un primer mapa, convenía ocupar en primera instancia la herramienta gráfica del Atlas, que sólo funciona en la nube. En este caso, la herramienta del MongoDB Compass, conectada a un servidor de la nube, tiene incorporada una herramienta de comando llamada Mongosh. Se ejecutaron los siguientes comandos en Mongosh, cuyo resultado se observa en el Script 9.

```
db.padron.updateMany(   {lati1:{$regex:/,/}},
   [{   $set: { lati1:
{$replaceOne:{input:"$lati1",find:",", replacement:"."}}
}    }]   );
db.padron.updateMany(  {prov:"Formosa"},   [{       $set: {
lati1:             { $convert: { input: "$lati1", to:
"double" } }            }    }] );
```

Script 9. Transformación de latitud a un formato de doble precisión. Fuente propia.

Obsérvese cómo cambió el formato de la latitud.

Se hizo lo propio, para la longitud, como se observa en el Script 10.

```
 db.padron.updateMany(   {longi1:{$regex:/,/}},    [{
      $set: {longi1:
{$replaceOne:{input:"$longi1",find:",", replacement:"."}}
}    }]      );
db.padron.updateMany(    {prov:"Formosa"},    [{       $set:
{ longi1:             { $convert: { input: "$longi1", to:
"double" } }             }    }] );
```

Script 10. Transformación de longitud a un formato de doble precisión. Fuente propia.

Posteriormente, se creó un nuevo campo al que se llamó coordenadas, y que es un vector de 2 posiciones (latitud y longitud), como se observa en el Script 11.

```
db.padron.updateMany(
      {prov:/Formosa/},
         [{ $set: {coordenadas:["$longi1","$lati1"]} }]  );
```

Script 11. Creación del campo de coordenadas. Fuente propia.

Capítulo 5

Análisis y presentación de resultados de las principales variables en la nube

5.1 Relevamientos Anuales 2010-2022 y padrón de establecimientos

5.1.1 Exportación en formato JSON e importación en la nube

En la Figura 45 se observa la exportación de resultados preprocesados en el localhost:27017, el formato JSON de los datos, e importación en la nube, en el Atlas: cloud.mongodb.com.

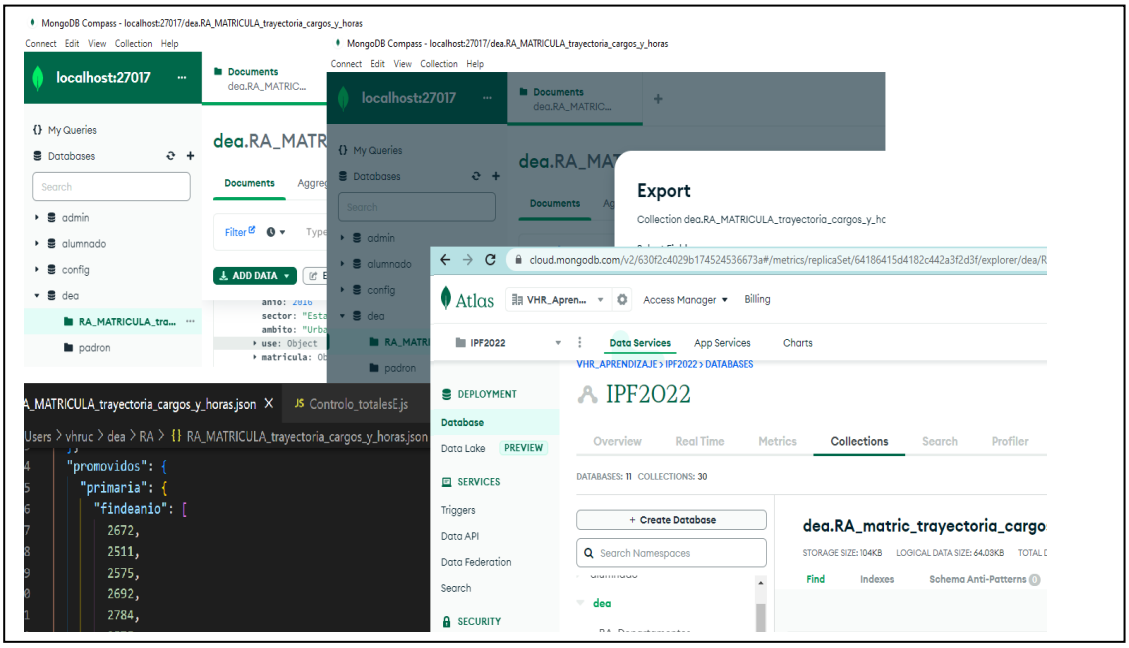


Figura 45. Exportación de datos preprocesados en modo local e importación en la nube: cloud.mongodb.com. Fuente propia.

Es importante aclarar que existe información publicada en bases de datos sobre el ámbito, pero no se realiza el análisis de la dimensión ámbito, en los Anuarios de Relevamientos Anuales del Ministerio de Capital Humano, y que sí se muestra en este trabajo. En los anuarios, se toma al ámbito rural como una modalidad. A diferencia de los análisis de los informes Aprender, en los anuarios de relevamientos anuales, solamente se hace un análisis del total de cada jurisdicción, y se discrimina por sector de gestión, en estatal y privado. Conviene recordar que sólo dos jurisdicciones no tienen ámbito rural, y 22 provincias si lo tienen.

La estructura final de resultados intermedios de los relevamientos anuales quedó conformada por documentos, subdocumentos, vectores y campos individuales, por ejemplo: se crearon documentos por cada año, gestión y ámbito; cada documento tiene campos individuales y subdocumentos como ser: unidades de servicio educativo (use), matrícula, matrícula_varones, secciones, matrícula_anio_anterior, repitentes, sobreedad, secciones, promovidos, cargos, horas, como se observa en la Figura 46. A su vez, para la variable promovidos, se tienen 4 subdocumentos (primaria, primaria_mujeres, secundaria y

secundaria_mujeres) y a su vez, por ejemplo, secundaria tiene 4 vectores: promovidos a fin de año, promovidos con exámenes, promovidos regulares y otros promovidos, como se observa en la Figura 47.

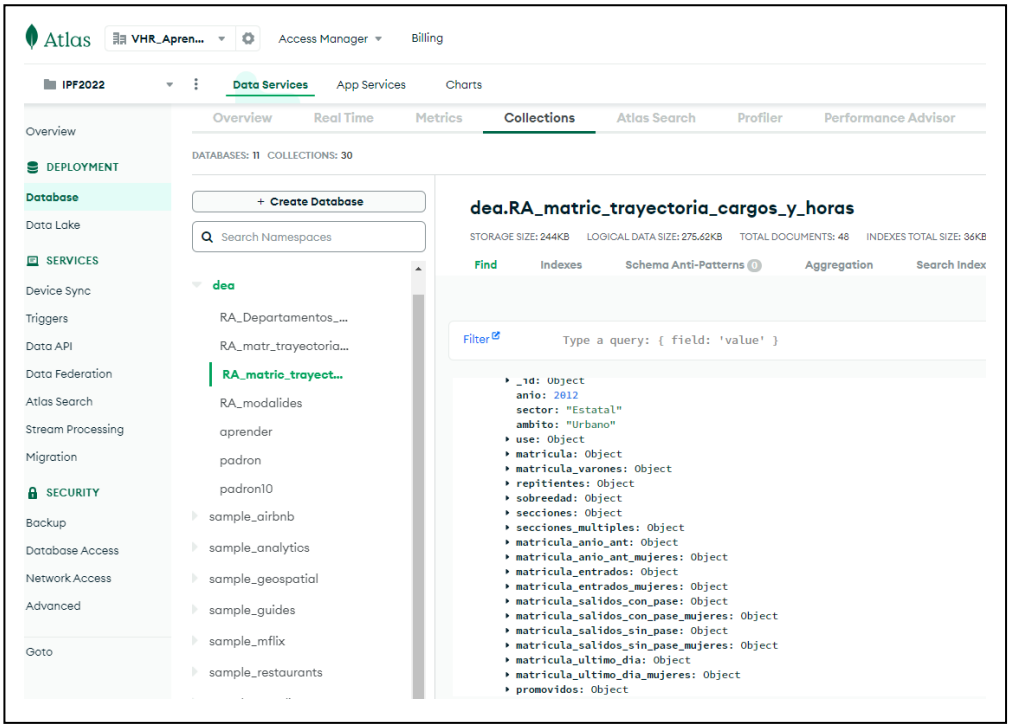


Figura 46. Estructura de matrícula, trayectoria, cargos y horas. Fuente propia.

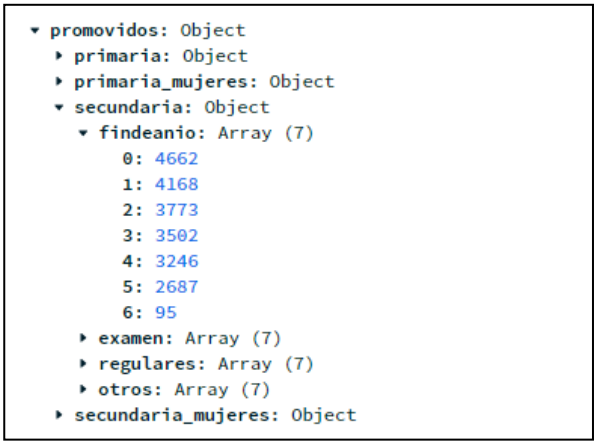


Figura 47. Estructura de la variable promovidos, en subdocumentos en distintos niveles y vectores. Fuente propia.

5.1.2 Geolocalización y mapas

5.1.2.1 Consulta de proximidad de establecimientos educativos cercanos a otro establecimiento educativo geolocalizado

MongoDB Compass permite realizar consultas de establecimientos cercanos a establecimiento educativo geolocalizado, estableciendo una conexión a la base de datos en la nube. Con la consulta del Script 12, se traen todos los establecimientos cercanos, de 0 a 2000 metros a la redonda, de un establecimiento educativo geolocalizad (EPES 20) de Villa Escolar, como se observa en la Figura 48. Cabe aclarar, que no se puede hacer esta consulta de proximidad en forma conjunta en un mapa del MongoDB Atlas, al menos en la versión gratuita utilizada en este trabajo.

```
db.padron.find({ubicaciones:{ $near :{$geometry: { type:
"Point",  coordinates: [ -58.67234, -26.62399 ] },
$minDistance: 0,$maxDistance:
2000}}},{_id:0,cue_anexo:1,nombre:1})
```

Script 12. Consulta de establecimientos cercanos a un punto. Fuente propia

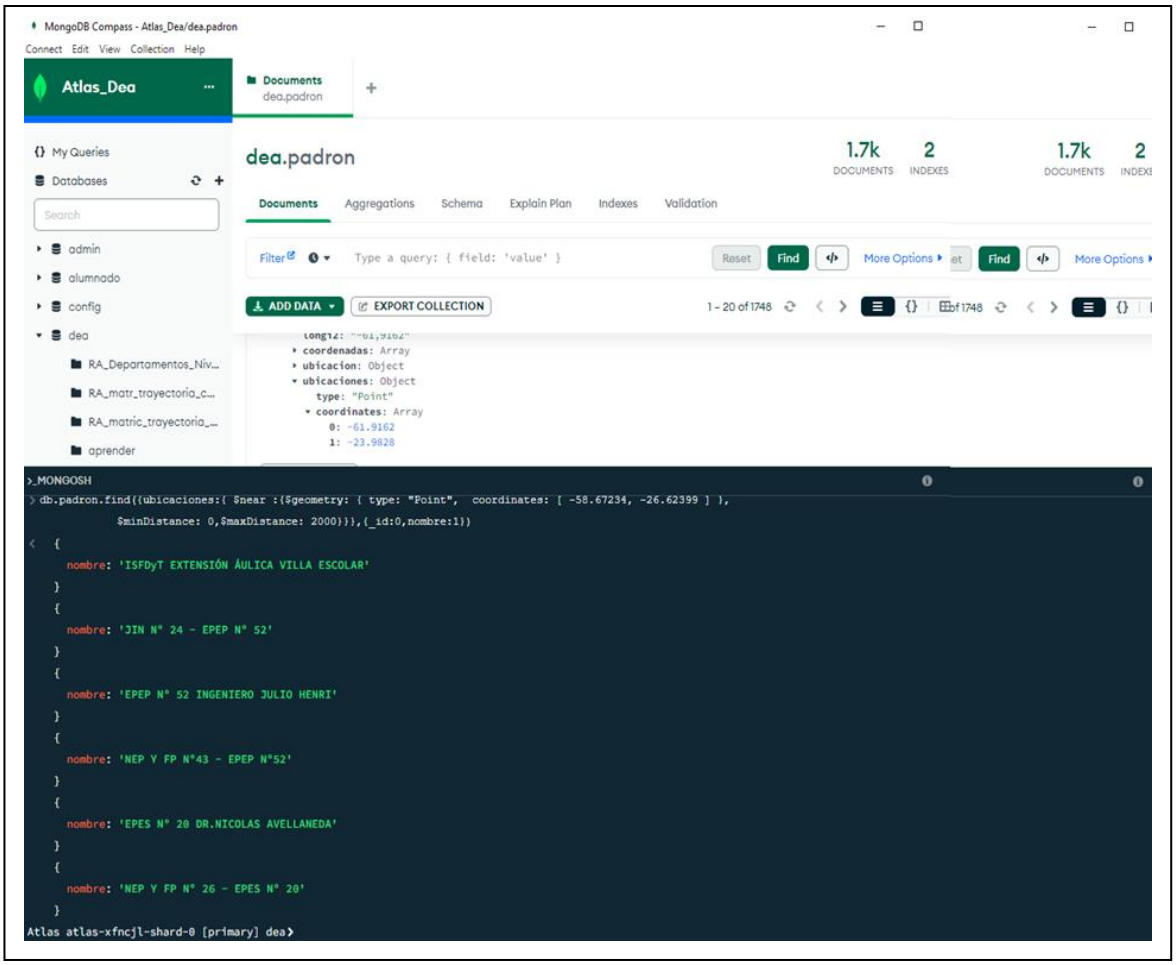


Figura 48. Consulta de proximidad de establecimientos educativos cercanos a otro establecimiento geolocalizado. Fuente propia.

5.1.2.2 Mapas en la nube utilizando MongoDB Atlas

MongoDB Atlas permite hacer mapas en la nube, como se observa en la

Figura 49, en donde se indican los establecimientos educativos de Formosa. Los puntos verdes, corresponden a establecimientos educativos de la educación común del nivel primario.

Figura 49. Mapa de establecimientos educativos de la provincia de Formosa. Fuente propia.

El problema que tiene la versión gratuita del MongoDB Atlas utilizada es que solamente se pueden visualizar 1.000 geolocalizaciones. Y esta cantidad supera la cantidad de establecimientos educativos de la provincia de Formosa.

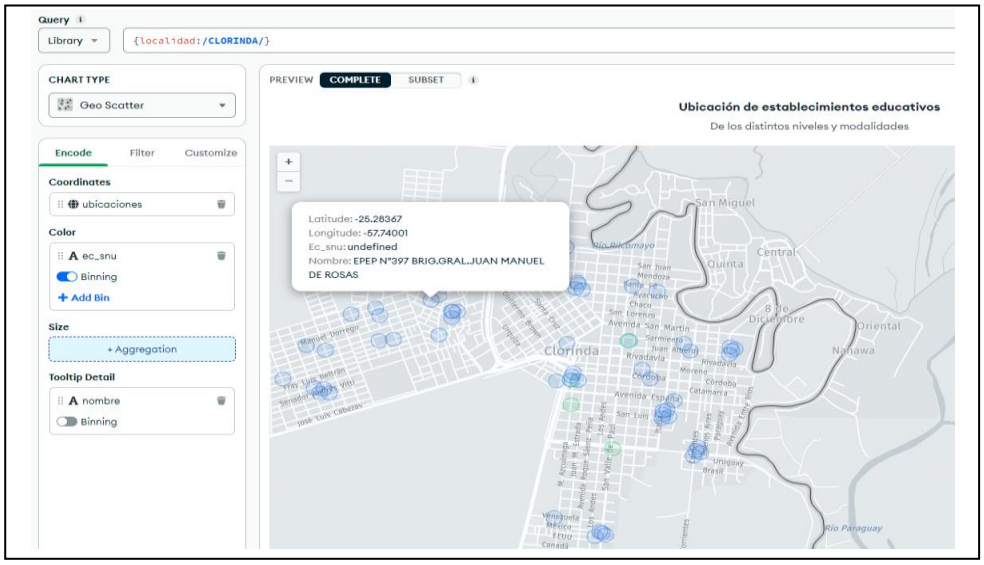


Figura 50. Mapa de establecimientos educativos filtrando el atributo localidad del documento. Fuente propia.

Es posible consultar en el mapa por cualquier atributo del documento, por ejemplo, visualizar establecimientos educativos de la localidad de Clorinda, e identificando los establecimientos educativos del nivel superior no universitario con el color verde, como se observa en la Figura 50.

También se puede hacer una consulta de un mapa, de un nivel específico, como se observa en la Figura 51.

Figura 51. Mapa de establecimientos especificando un nivel específico de los documentos. Fuente propia.

Asimismo, conforme a la cantidad de establecimientos secundarios, se puede obtener un mapa de calor como el de la Figura 52, en donde se puede observar que existe concentración de instituciones educativas secundarias en la ciudad de Formosa.

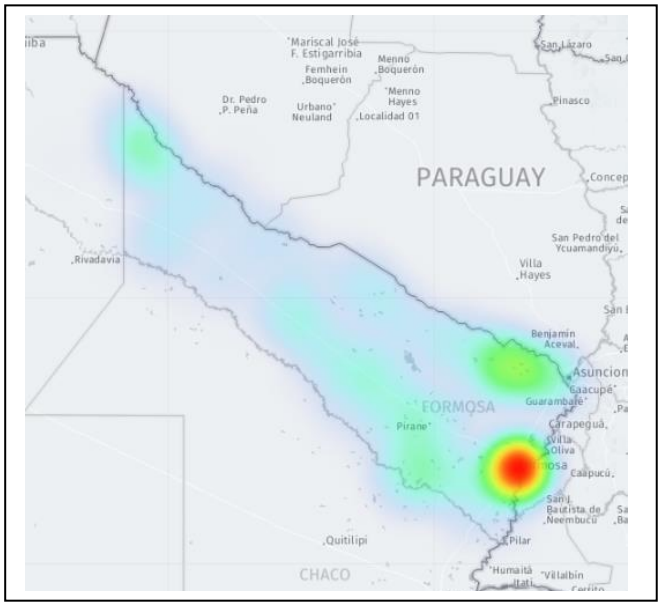


Figura 52. Mapa de calor utilizando un criterio determinado. Fuente propia

También se observa que las instituciones educativas secundarias se encuentran repartidas en toda la geografía provincial, ya que existen muchas instituciones en el ámbito rural.

5.1.3 Cuadros estadísticos

5.1.3.1 Cuadros de educación común de 4 niveles educativos y modalidades especial y adulto

Se construyen campos calculados de los distintos subdocumentos, como se observa en el Script 13.

```
TG_UNE= {$sum: [{$sum:[ "$une.inicial"]},{$sum:[
"$une.primaria"]},{$sum:[ "$une.secundaria"]},{$sum:[
"$une.superior"]},{$sum:[ "$mod_une.esp"]},{$sum:[
"$mod_une.adu"]}]}
TG_USE= {$sum:
["$use.inicial_ambos_mat_e_inf","$use.inicial_infantes","$us
e.primaria6","$use.secundaria_cb6","$use.secundaria_completa
6","$use.snu",{$sum:[ "$mod_use.esp"]},{$sum:[
"$mod_use.adu"]}]}
TG_Matricula= {$sum: [{$sum:[ "$matricula.inicial"]},{$sum:[
"$matricula.primaria"]},{$sum:[
"$matricula.secundaria"]},{$sum:[ "$matricula.snu"]},{$sum:[
"$esp_matricula_sin_integracion"]},{$sum:[
"$adu_matricula"]}]}
TG_cargos= {$sum: [{$sum:[ "$cargos.inicial"]},{$sum:[
"$cargos.primaria"]},{$sum:[ "$cargos.secundaria"]},{$sum:[
"$cargos.snu"]},{$sum:[ "$mod_cargos.esp"]},{$sum:[
"$mod_cargos.adu"]}]}
TG_horas= {$sum: [{$sum:[ "$horas.inicial"]},{$sum:[
"$horas.primaria"]},{$sum:[ "$horas.secundaria"]},{$sum:[
"$horas.snu"]},{$sum:[ "$mod_horas.esp"]},{$sum:[
"$mod_horas.adu"]}]}
```

Script 13. Creación de campos calculados en el gráfico para totalizar las distintas variables. Fuente propia.

En la Figura 53 se visualiza un cuadro que indica la evolución de las principales variables de la educación común (niveles inicial, primario, secundario y superior) y modalidades especial y adultos.

Evolución de las Principales Variables del Sistema Educativo - EDUCACIÓN COMÚN Y MODALIDADES especial y adultos

Principales variables de la EDUCACIÓN COMÚN de los niveles Inicial, Primario, Secundario y Superior No Universitario y MODALIDADES especial y adultos por AÑO

Año	Unid.Educ.	Unid.Serv.Educativo	Matrícula	Cargos	Horas
2.011	1.090	1.967	209.211	13.235	108.309
2.012	1.124	1.990	205.978	14.223	107.644
2.013	1.081	2.004	204.748	13.684	110.620
2.014	1.128	2.003	204.809	14.374	124.392
2.015	1.133	1.989	205.510	14.137	117.058
2.016	1.137	1.961	205.485	14.181	120.526
2.017	1.165	1.988	205.603	14.865	119.252
2.018	1.171	1.986	205.030	14.962	124.773
2.019	1.138	1.965	204.614	14.848	116.355
2.020	1.158	1.953	200.439	14.357	118.087
2.021	1.129	1.920	198.112	14.239	119.531
2.022	1.148	1.932	197.511	14.466	123.855
2.023	1.153	1.931	196.215	14.634	122.179

Figura 53. Evolución de principales variables. Educación común y modalidades. Fuente: Relevamientos anuales (RA2011 a RA2023), anuarios y propia.

En la Figura 54 se puede observar el cuadro de la evolución de las principales variables de la educación común, consistente en los niveles inicial, primario, secundario y superior no universitario, por año.

Evolución Principales Variables del Sistema Educativo - Educación común por año

Principales variables de la EDUCACIÓN COMÚN de los niveles Inicial, Primario, Secundario y Superior No Universitario por AÑO

Año	Unid.Educ.	Unid.Serv.Educativo	Matrícula	Cargos	Horas
2.011	892	1.458	182.474	11.523	101.157
2.012	904	1.469	177.159	12.467	101.398
2.013	910	1.478	176.183	11.816	100.296
2.014	909	1.477	176.496	12.311	114.097
2.015	913	1.477	176.798	12.134	106.539
2.016	915	1.465	176.661	12.227	109.662
2.017	916	1.461	176.293	12.911	107.919
2.018	920	1.468	174.634	12.963	113.241
2.019	913	1.463	173.415	12.764	104.851
2.020	913	1.460	172.256	12.344	106.566
2.021	888	1.437	170.721	12.296	108.211
2.022	908	1.448	170.228	12.584	112.565
2.023	913	1.453	167.547	12.728	110.279

Figura 54. Evolución de las principales variables de la educación común por año. Fuente: RA2011 a RA2023 y propia.

En la Figura 55 se muestra el cuadro de las variables principales del año 2022, por departamento.

Relevamiento Anual - Año:2022

Principales Variables de la Educación Común de los Niveles Inicial -Primario - Secundario y Superior No Universitario por Departamento

Departamento	Unidades Educa...	Unidades de Servicio Ed...	Matricula	Porc.Estatal	Cargos	Horas
Bermejo	72	129	6.342	94,9%	640	3.904
Formosa	223	270	68.862	79,6%	4.782	40.462
Laishí	40	66	5.639	87,4%	457	5.124
Matacos	25	53	5.819	100,0%	357	3.091
Patiño	201	319	24.515	89,3%	1.925	17.081
Pilagás	50	82	5.236	100,0%	465	4.092
Pilcomayo	117	180	26.294	80,1%	1.658	18.687
Pirané	141	240	19.872	90,4%	1.718	15.680
Ramón Lista	57	109	7.649	100,0%	582	4.444
Total	**926**	**1.448**	**170.228**		**12.584**	**112.565**

Activar Windows
Ve a Configuración para activar Windows.

Figura 55. Principales variables del 2022 por departamento. Fuente: Anuario 2022.

Obsérvese que no coinciden los valores de la Unidades Educativas por Departamento, que deberían ser 908 unidades educativas (estructuras organizativas que tienen director), tal como se observa en la Figura 54. Al no poder procesar las claves enmascaradas e identificar cada una de las unidades educativas con su correspondiente departamento, se completaron datos de los objetos JSON procesados en modo local con el valor de esta variable. Sin embargo, existen diferencias entre los cuadros de la Figura 54 y Figura 55.

Indagando el origen de la diferencia, la misma está dada por la mala consignación de los totales de unidades educativas por departamento, del Resumen Jurisdiccional obtenido del sitio https://www.argentina.gob.ar/educacion/evaluacion-informacion-educativa/anuarios, existiendo una diferencia de 18 en el total y en la suma parcial de cada nivel, como se observa en la Figura 56. En el resumen jurisdiccional por departamentos proporcionados por los anuarios de 2022, la variable unidades educativas es la única que no coincide con valores de bases de datos, como se observa en la diferencia entre la Figura 55 y la Figura 56.

la educación nuestra bandera

Secretaría de Evaluación e Información Educativa

Ministerio de Educación Argentina

5.2.9. FORMOSA. Educación común. Unidades educativas, unidades de servicio, alumnos, cargos y horas por departamento. Resumen jurisdiccional

Departamento	Unidades educativas	Unidades de servicio	Alumnos	Porcentaje estatal	Cargos	Horas	UE	Diferencia
TOTAL	908	1.448	170.228	88,1	12.584	112.565	926	18
Bermejo	72	129	6.342	93,1	640	3.904	72	
Formosa	223	270	68.862	83,9	4.782	40.462	223	
Laishi	40	66	5.639	87,5	457	5.124	40	
Matacos	25	53	5.819	100,0	357	3.091	25	
Patiño	201	319	24.515	90,6	1.925	17.081	201	
Pilagás	50	82	5.236	100,0	465	4.092	50	
Pilcomayo	117	180	26.294	84,9	1.658	18.687	117	
Pirané	141	240	19.872	91,1	1.718	15.680	141	
Ramón Lista	57	109	7.649	100,0	582	4.444	57	

Figura 56. Diferencia de la sumatoria de unidades educativas por departamento en el resumen jurisdiccional del anuario 2022. Fuente: Anuario 2022.

Debido a la extensión del informe, en la Figura 57 se muestra un extracto (años 2017 a 2023) del cuadro de la evolución de las principales variables de la educación común por sector de gestión y año.

Evolución Principales Variables del Sistema Educativo - Educación común por año y sector de gestión

Principales variables de la EDUCACIÓN COMÚN de los niveles Inicial, Primario, Secundario y Superior No Universitario por AÑO y sector

Año	Sector	Unid.Educ.	Unid.Serv.Educativo	Matrícula	Cargos	Horas
2.017	Privado	132	137	21.305	1.653	15.220
2.018	Estatal	788	1331	153.131	11.228	98.087
2.018	Privado	132	137	21.503	1.735	15.154
2.019	Estatal	780	1325	152.202	11.253	92.021
2.019	Privado	133	138	21.213	1.511	12.830
2.020	Estatal	782	1324	151.701	10.951	93.258
2.020	Privado	131	136	20.555	1.393	13.308
2.021	Estatal	782	1326	152.375	11.240	94.019
2.021	Privado	106	111	18.346	1.056	14.192
2.022	Estatal	781	1316	149.945	11.134	98.508
2.022	Privado	127	132	20.283	1.450	14.057
2.023	Estatal	787	1323	147.145	11.321	96.434
2.023	Privado	126	130	20.402	1.407	13.845

Figura 57. Extracto de la evolución de las principales variables de la educación común por año y sector de gestión. Fuente: R2011 a RA2023 y propia.

En MongoDB Atlas, sencillamente se agrega un grupo más en el diseño del cuadro, como se observa en la Figura 58.

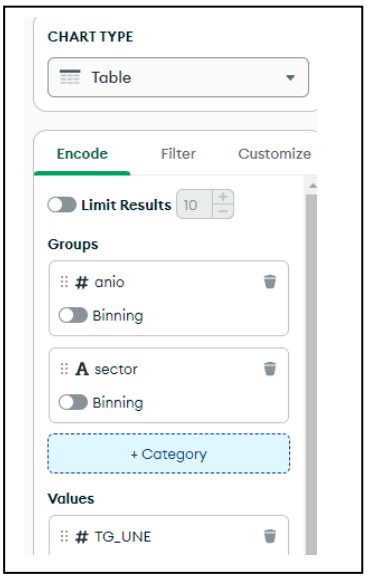


Figura 58. Facilidad de MongoDB Atlas para agregar un grupo al diseño del cuadro. Fuente propia.

El cuadro de la evolución de las principales variables de la educación común, por año y ámbito, se observa en la Figura 59.

Evolución Principales Variables del Sistema Educativo - Educación común por año y ámbito

Principales variables de la EDUCACIÓN COMÚN de los niveles Inicial, Primario, Secundario y Superior No Universitario por AÑO y ámbito

Año	Ámbito	Unid.Educ.	Unid.Serv.Educativo	Matrícula	Cargos	Horas
2.017	Rural	505	879	32.322	3.746	19.098
2.017	Urbano	411	582	143.971	9.165	88.821
2.018	Rural	506	882	31.892	3.728	23.882
2.018	Urbano	414	586	142.742	9.235	89.359
2.019	Rural	493	878	31.362	3.710	19.041
2.019	Urbano	420	585	142.053	9.054	85.810
2.020	Rural	491	875	31.133	3.554	18.449
2.020	Urbano	422	585	141.123	8.790	88.117
2.021	Rural	491	876	31.349	3.838	18.979
2.021	Urbano	397	561	139.372	8.458	89.232
2.022	Rural	488	868	30.782	3.871	22.578
2.022	Urbano	420	580	139.446	8.713	89.987
2.023	Rural	487	868	30.435	3.878	22.694
2.023	Urbano	426	585	137.112	8.850	87.585

Figura 59. Evolución de las principales variables de la educación común por año y ámbito. Fuente: RA2011 a 2023 y propia.

Obsérvese que, en el ámbito rural de la provincia de Formosa existen más unidades educativas y unidades de servicio educativo, a pesar de que sus respectivas matrículas son

menores. MongoDB Atlas, proporciona importantes herramientas de análisis, en este caso a través del formato condicional.

5.1.3.2 Cuadros de la educación común - nivel inicial

A continuación, se muestran los distintos parciales que conforman la Figura 53, y se visualizarán la evolución de los 4 niveles de la educación común y las modalidades. En la Figura 60, se observa las principales variables del nivel inicial. En la Figura 61, la descomposición de salas de 3, 4 y 5 años, y salas múltiples, indicando la matrícula de maternal e inicial.

Nivel inicial-Evolución Principales Variables del Sistema Educativo por año

Principales variables del Nivel Inicial por AÑO

Año	Unid.Educ.	Unid.Serv.Educativo	Matrícula	Cargos	Horas
2.011	115	484	17.851	1.323	17
2.012	129	497	19.635	1.545	4
2.013	136	501	20.273	1.676	51
2.014	137	504	20.899	1.767	10
2.015	137	504	21.599	1.629	217
2.016	137	500	22.330	1.620	10
2.017	138	495	22.301	2.105	41
2.018	139	496	22.607	2.096	112
2.019	140	494	22.708	1.926	108
2.020	140	492	21.416	1.843	114
2.021	114	468	18.565	1.513	42
2.022	136	481	20.281	1.883	8
2.023	143	487	20.841	1.958	120

Figura 60. Evolución de principales variables del nivel inicial. Fuente: RA2011 a RA2023 y propia.

Educación Inicial-Unid.Serv.Educ., secciones y matrícula por año y sector

Principales variables de la EDUCACIÓN COMÚN Nivel INICIAL por año y sector

Año	sector	USE_Ini	USE Ambas	S_de_3	S_de_4	S_de_5	Secc.Múlt.	Matr_MAT	Matr_IN
2.017	Estatal	422	1	1	243	399	273	10	16.55
2.017	Privado	40	32	73	94	59	53	1184	4.55
2.018	Estatal	423	1	1	246	411	267	6	16.76
2.018	Privado	40	32	69	91	62	43	1153	4.68
2.019	Estatal	420	1	0	250	445	231	2	16.88
2.019	Privado	41	32	81	105	60	29	1098	4.72
2.020	Estatal	419	1	2	217	483	233	3	16.18
2.020	Privado	41	31	74	100	61	16	881	4.34
2.021	Estatal	421	1	0	232	448	255	4	15.35
2.021	Privado	40	6	21	57	65	8	86	3.11
2.022	Estatal	414	0	0	244	426	264	0	15.45
2.022	Privado	42	25	63	85	61	14	863	3.96
2.023	Estatal	419	0	1	252	455	250	0	15.74
2.023	Privado	40	28	75	91	65	14	1028	4.07

Figura 61. Nivel inicial, salas de 3, 4 y 5 años, y matrículas, por sector y año. Fuente: RA2011 a RA2023 y propia.

5.1.3.3 Cuadros de la educación común - nivel primario

Se observa en la Figura 62, la evolución de las principales variables del nivel primario por año. En la Figura 63, se desagregan las principales variables, por año y ámbito.

Primaria_Evolución Principales Variables del Sistema Educativo por año

Principales variables de la EDUCACIÓN COMÚN del nivel Primario por año

Año	Unid.Educ.	Unid.Serv.Educativo	Matrícula	Cargos	Horas
2.011	513	598	85.809	6.896	128
2.012	513	597	82.512	6.897	143
2.013	512	597	80.013	6.892	162
2.014	512	597	78.732	7.034	172
2.015	510	593	77.588	6.989	919
2.016	511	593	75.890	7.075	625
2.017	510	592	73.627	7.031	54
2.018	510	592	73.290	7.132	270
2.019	509	588	72.367	7.045	87
2.020	507	583	72.390	6.965	189
2.021	507	583	72.515	6.901	483
2.022	506	581	71.956	6.912	293
2.023	506	580	70.777	6.920	377

Figura 62. Evolución de las principales variables del nivel primario, por año. Fuente: RA2011 a RA2023 y propia.

Primaria_Evolución Principales Variables del Sistema Educativo por año y ámbito

Principales variables de la EDUCACIÓN COMÚN del nivel Primario por año y ámbito

Año	Ámbito	Unid.Educ.	Unid.Serv.Educativo	Matrícula	Cargos	Horas
2.017	Rural	337	401	16.256	2.199	0
2.017	Urbano	173	191	57.371	4.832	54
2.018	Rural	337	401	15.925	2.223	74
2.018	Urbano	173	191	57.365	4.909	196
2.019	Rural	335	398	15.311	2.223	40
2.019	Urbano	174	190	57.056	4.822	47
2.020	Rural	331	393	14.941	2.207	22
2.020	Urbano	176	190	57.449	4.758	167
2.021	Rural	331	393	14.934	2.175	10
2.021	Urbano	176	190	57.581	4.726	473
2.022	Rural	329	391	14.496	2.179	10
2.022	Urbano	177	190	57.460	4.733	283
2.023	Rural	329	390	13.891	2.196	20
2.023	Urbano	177	190	56.886	4.724	357

Figura 63. Extracto de la evolución de las principales variables del nivel primario, por año y ámbito. Fuente: RA2011 a RA2023 y propia.

En el cuadro de la Figura 64 se observa la evolución de las matrículas y cantidad de repitentes por grado y año. La matrícula fue descendiendo a lo largo del tiempo, al igual que la cantidad de repitentes. MongoDB Atlas asigna automáticamente distintos colores, según los valores de la matrícula y de cantidad de repitentes.

Nivel Primario-Evolución de matrícula y repitentes por año

Matríoula y Repitentes por año

Año	Matrícula	Matr_1°g	M2	M3	M4	M5	M6	Repitentes	R1	R2	R3	R4	R5	R6
2.011	85.809	15.021	14.431	14.598	14.353	13.827	13.579	6775	2006	1279	1194	1055	789	452
2.012	82.512	13.467	13.985	14.182	14.200	13.649	13.029	5047	1480	993	872	759	609	334
2.013	80.013	12.600	12.955	13.924	13.936	13.658	12.940	2552	480	441	556	485	365	225
2.014	78.732	12.262	12.555	13.294	13.747	13.652	13.222	1793	130	193	526	391	345	208
2.015	77.588	13.058	12.203	12.833	12.944	13.412	13.138	1953	118	164	524	418	387	342
2.016	75.890	12.315	12.879	12.646	12.479	12.618	12.953	2142	131	193	595	432	426	365
2.017	73.627	12.577	12.002	13.132	12.072	11.897	11.947	2738	108	272	709	606	518	525
2.018	73.290	12.194	12.364	12.570	12.721	11.869	11.572	3007	185	327	894	543	520	538
2.021	72.515	12.399	12.057	12.227	11.922	12.207	11.703	1459	36	131	333	323	281	355
2.020	72.390	12.121	11.960	12.184	12.280	11.758	12.087	2281	98	279	616	375	456	457
2.019	72.367	12.015	11.892	12.783	11.864	12.257	11.556	2413	119	281	730	515	386	382
2.022	71.956	11.620	12.265	12.080	12.063	11.801	12.127	1389	22	116	206	289	318	438
2.023	70.777	11.152	11.596	12.312	12.044	11.930	11.743	698	7	60	103	166	168	194

Figura 64. Análisis de evolución de matrículas y cantidad de repitentes por grado y año. Fuente: RA2011 a RA2023 y propia.

5.1.3.4 Cuadros de la educación común - nivel secundario

Se observa la evolución de las principales variables del nivel secundario en la Figura 65.

Secundaria-Evolución Principales Variables del Sistema Educativo por año

Principales variables de la EDUCACIÓN COMÚN del nivel Secundario por AÑO

Año_del_Relevamiento	Unid.Educ.	Unid.Serv.Educativo	Matrícula	Cargos	Horas	Egresados_año_anterior
2.011	226	337	68.111	2.985	81.247	5.509
2.012	224	333	64.799	3.126	84.029	5.705
2.013	224	335	64.233	2.933	83.501	5.787
2.014	221	328	64.138	3.175	83.672	5.773
2.015	227	334	64.379	3.130	85.335	5.819
2.016	228	327	64.225	3.171	85.887	5.495
2.017	228	329	64.171	3.392	86.816	5.429
2.018	230	332	63.709	3.333	88.160	5.585
2.019	223	333	63.220	3.363	86.242	5.486
2.020	225	337	62.480	3.075	85.702	5.768
2.021	225	337	64.537	3.462	87.832	5.858
2.022	224	337	62.873	3.439	91.568	6.145
2.023	223	338	62.560	3.474	92.475	5.658

Figura 65. Evolución de las principales variables del nivel secundario de la educación común. Fuente: RA2011 a RA2023 y propia.

.

En la Figura 66, se observa la evolución de las matrículas del año que se informa y cantidad de repitentes del año anterior, y tanto en matrícula como en repitentes, se discriminan los distintos años de estudio, todos pertenecientes al nivel secundario de la educación común.

Ec_Nivel Secundario-Evolución de matrícula y repitentes por año

Matrícula y Repitentes por año

Año	Matrícula	1°A	M2	M3	M4	M5	M6	M7	Rep.	R1	R2	R3	R4	R5	R6	R7
2.011	68.111	15.841	14.003	12.144	10.367	8.476	7.073	207	6.350	2212	1842	1101	708	372	115	0
2.012	64.799	15.003	12.699	10.867	9.841	8.706	7.475	208	5.950	1827	1654	1174	718	464	104	9
2.013	64.233	14.676	13.011	10.766	9.543	8.499	7.415	323	6.156	2001	1627	1096	798	530	104	0
2.014	64.138	14.351	12.944	11.148	9.616	8.272	7.458	349	5.365	1745	1473	866	725	409	144	3
2.015	64.379	14.620	12.830	11.093	9.812	8.333	7.299	392	5.643	1833	1584	990	679	431	126	0
2.016	64.225	14.724	12.833	10.964	9.676	8.384	7.141	503	6.751	2147	1786	1274	871	555	118	0
2.017	64.171	14.701	12.901	10.901	9.709	8.250	7.206	503	6.837	2122	1732	1256	935	579	209	4
2.018	63.709	14.109	12.776	11.066	9.583	8.438	7.152	585	6.734	2300	1746	1177	754	548	203	6
2.019	63.220	13.445	12.394	11.117	9.917	8.365	7.350	632	5.858	1871	1593	1072	697	441	173	11
2.020	62.480	13.248	11.810	10.543	9.991	8.692	7.590	606	5.163	1570	1299	922	733	438	195	6
2.021	64.537	12.309	12.381	11.261	10.205	9.351	8.297	733	2.549	669	553	457	359	301	206	4
2.022	62.873	12.946	12.281	10.643	9.741	8.830	7.629	803	5.554	1234	1466	1141	879	638	182	14
2.023	62.560	13.995	12.114	10.777	9.492	8.230	7.179	773	6.506	1591	1550	1196	1127	776	258	8

Figura 66. Evolución de matrícula y repitentes del nivel secundario de la educación común, por año de estudio y año lectivo. Fuente: RA2011 a RA2023 y propia.

Se evidencia un crecimiento de la matrícula del séptimo año del secundario, correspondiente a la educación técnica.

En la Figura 67, se pueden observar las distintas evoluciones de las orientaciones de egresados. Recuérdese, que los egresados de un año, se informan en el relevamiento anual del año siguiente.

Ec_Nivel Secundario- Evolución de egresados por orientación por año

Egresados del nivel secundario por ORIENTACIÓN por AÑO DE EGRESO

Año_de_egreso	Total	Cs.Salud	Cs.Básicas	Cs.Aplicadas a Tecnología	Cs.Sociales	Cs.Humanas
2010	5509	5	971	1112	3313	108
2011	5705	7	992	1356	3161	189
2012	5787	0	1023	1431	3295	38
2013	5773	0	1120	1470	3108	75
2014	5819	0	1070	1641	2991	117
2015	5495	0	1182	1478	2727	108
2016	5429	8	1089	1629	2564	139
2017	5585	0	1166	1557	2674	188
2018	5486	16	1118	1526	2698	128
2019	5768	21	997	1741	2886	123
2020	5858	0	1074	1066	2657	1061
2021	6145	28	1049	959	2602	1507
2022	5658	0	969	1177	2589	923

Figura 67. Evolución de egresados del nivel secundario por orientación y año de egreso de la educación común. Fuente: RA2011 a RA2023 y propia.

5.1.3.5 Evolución de las principales variables de la educación común - nivel superior no universitario

Se observa la evolución de las principales variables del nivel superior no universitario en la Figura 68.

Ec_Superior No Universitario-Evolución Principales Variables por año

Principales variables de la EDUCACIÓN COMÚN del nivel Superior No Universitario por AÑO

Año	Unid.Educ.	Unid.Serv.Educativo	Matrícula	Cargos	Horas
2.011	38	39	10.703	319	19.765
2.012	38	42	10.213	899	17.222
2.013	38	45	11.664	315	16.582
2.014	39	48	12.727	335	30.243
2.015	39	46	13.232	386	20.068
2.016	39	45	14.216	361	23.140
2.017	40	45	16.194	383	21.008
2.018	41	48	15.028	402	24.699
2.019	41	48	15.120	430	18.414
2.020	41	48	15.970	461	20.561
2.021	42	49	15.104	420	19.854
2.022	42	49	15.118	350	20.696
2.023	41	48	13.369	376	17.307

Figura 68. Evolución de las principales variables del nivel superior no universitario de la educación común, por año. Fuente: RA2011 a RA2023 y propia.

5.1.3.6 Modalidad especial

Se observa la evolución de las principales variables de la modalidad especial en la Figura 69. La matrícula hace referencia únicamente a los alumnos que no están integrados (sin integración), ya que, a los alumnos integrados, se los consignan en los otros niveles educativos.

Modalidad Especial-Evolución de las Principales Variables

Principales variables de la modalidad especial por AÑO-En la variable matrícula, se consgina ALUMNOS NO INTEGRADOS

Año	Unid.Educ.	Unid.Serv.Educativo	Matrícula	Cargos	Horas
2011	65	78	691	611	0
2012	64	78	582	622	0
2013	67	81	598	607	0
2014	65	79	488	623	0
2015	62	75	892	638	40
2016	60	74	790	630	0
2017	83	103	827	643	0
2018	85	103	943	654	0
2019	59	100	863	660	0
2020	81	96	781	653	807
2021	78	94	748	631	0
2022	77	94	762	631	0
2023	77	94	797	634	0

Figura 69. Evolución de las principales variables de la modalidad especial por año. Fuente: RA2011 a RA2023 y propia.

5.1.3.7 Modalidad de jóvenes y adultos

Se observa en la Figura 70, la evolución de las principales variables de la modalidad de jóvenes y adultos.

Modalidad de jóvenes y adultos-Evolución de las Principales Variables por año

Principales variables de la modalidad adultos por año

Año	Unid.Educ.	Unid.Serv.Educativo	Matrícula	Cargos	Horas
2011	133	431	26046	1101	7152
2012	156	443	28237	1134	6246
2013	104	445	27967	1261	10324
2014	154	447	27825	1440	10295
2015	158	437	27820	1365	10479
2016	162	422	28034	1324	10864
2017	166	424	28483	1311	11333
2018	166	415	29453	1345	11532
2019	166	402	30336	1424	11504
2020	164	397	27402	1360	10714
2021	163	389	26643	1312	11320
2022	163	390	26521	1251	11290
2023	163	384	27871	1272	11900

Figura 70. Evolución de las principales variables de la modalidad de jóvenes y adultos por año. Fuente: RA2011 a RA2023 y propia.

5.1.4 Indicadores educativos

En este trabajo, solamente se consideran 3 indicadores educativos, a los fines ilustrativos.

5.1.4.1 Descripción de los indicadores educativos

La **tasa de promoción efectiva** [36] refiere al porcentaje de alumnos matriculados en un grado/año de estudio dado de un nivel de enseñanza en particular, que se matriculan como alumnos nuevos en el grado /año de estudio inmediato superior de ese nivel, en el año lectivo siguiente.

El propósito es mostrar en términos relativos, la eficiencia del sistema educativo en retener a los alumnos de un grado/año de estudio dado, como alumnos nuevos al año siguiente en el grado/año de estudio inmediatamente superior y bajo el supuesto de que los alumnos nuevos promovieron dicho grado /año de estudio. Además, este indicador integra el conjunto de las tasas de transición que se utilizan en el análisis de una cohorte educativa. Básicamente, mide la eficiencia interna del sistema.

El método de cálculo es el cociente entre los alumnos nuevos en un grado/año de estudio en un año lectivo dado y los alumnos en el grado/año de estudio anterior en el año lectivo próximo pasado, por cien. Su fórmula es:

$$TPE_{i;n}^{t} = \frac{AN_{i+1;n}^{t+1}}{A_{i;n}^{t}} * 100 \quad ; i = 1, \ldots, k-1$$

Ecuación 1. Tasa de promoción efectiva por grado/año.

$$TPE_{k;n}^{t} = \frac{Prom_{k;n}^{t}}{A_{i;n}^{t}} * 100$$

En donde:

$TPE_{i;n}^{t}$ = Tasa de promoción efectiva del grado/año de estudio *i* del nivel n en el año *t*.

$AN_{i+1;n}^{t+1}$ = Alumnos nuevos del grado/año de estudio *i*+1 en el nivel n en el año *t*+1.

$A_{i;n}^{t}$= Alumnos del grado/año de estudio *i* en el nivel *n* en el año *t*.

$Prom_{k;n}^{t}$= Promovidos del grado/año de estudio *k* del nivel *n* en el año *t*.

i = Grado/año de estudio.

k = Último grado/año de estudio del nivel *n*.

n = Nivel de la enseñanza.

Este indicador para el total del nivel de enseñanza se construye como un promedio ponderado de las tasas de promoción efectiva de los grados/años de estudio que componen dicho nivel. Por lo tanto, se tiene,

$$TPE_n^t = \frac{\sum_{i=1}^{k-1} AN_{i+1;n}^{t+1} + Prom_{k;n}^t}{\sum_{i=1}^{k} A_{i;n}^t} * 100$$

Ecuación 2. Tasa de promoción efectiva de cada nivel.

Los alumnos nuevos derivan de restar a los alumnos matriculados en el grado/año de estudio *i*, los alumnos repitentes en ese grado/año en ese año lectivo.

La **tasa de repitencia** es el porcentaje de alumnos matriculados en un grado/año de estudio dado de un nivel de enseñanza en particular, que se matriculan como alumnos repitentes en el mismo grado /año de estudio de ese nivel, en el año lectivo siguiente. Su propósito es mostrar en términos relativos la eficiencia del sistema educativo en función de los alumnos que al no promover el grado/año de estudio en el cual estaban matriculados, vuelven a inscribirse como repitentes en el mismo grado/año al año lectivo siguiente. También, este indicador integra el conjunto de las tasas de transición que se utilizan en el análisis de una cohorte educativa. La inclusión de este indicador es de suma relevancia para medir la ineficiencia del sistema educacional, por cuanto es uno de los antecedentes, para algunos especialistas el más importante, de la deserción escolar. También mide la eficiencia del sistema. Su método de cálculo es el cociente entre los alumnos repitentes de un grado/año de estudio en un año lectivo dado y los alumnos en el mismo grado/año de estudio en el año lectivo anterior, por cien.

$$TR_{i;n}^t = \frac{R_{i;n}^{t+1}}{A_{i;n}^t} * 100 \; ; \; i=1, \ldots, k$$

Ecuación 3. Tasa de repitencia de un grado/año.

En donde:

$TR_{i;n}^t$ = Tasa de repitencia del grado/año de estudio *i* del nivel *n* en el año *t*.

$R_{i;n}^{t+1}$ = Repitentes del grado/año de estudio *i* en el nivel *n* en el año *t*+1.

$A_{i;n}^t$= Alumnos del grado/año de estudio *i* en el nivel *n* en el año *t*.

i = Grado/año de estudio.

k = Último grado/año de estudio del nivel *n*.

n = Nivel de la enseñanza.

Este indicador para el total del nivel de enseñanza se construye como un promedio ponderado de las tasas de repitencia de los grados/años de estudio que componen dicho nivel. Por lo tanto, se tiene:

$$TR_n^t = \frac{\sum_{i=1}^{k} R_{i;n}^{t+1}}{\sum_{i=1}^{k} A_{i;n}^{t}} * 100$$

Ecuación 4. Tasa de repitencia de un nivel.

La **<u>tasa de abandono interanual</u>** es el porcentaje de alumnos matriculados en un grado/año de estudio dado de un nivel de enseñanza en particular, que no se vuelve a matricular al año lectivo siguiente como alumno nuevo o repitente. El propósito es mostrar, en términos relativos, el volumen de alumnos que abandona el sistema educativo durante el año lectivo y no vuelve a matricularse al año lectivo siguiente, sumado a los alumnos que abandonan el sistema educativo entre dicho año lectivo y el siguiente. Este indicador completa el conjunto de las tasas de transición que se utilizan en el análisis de una cohorte educativa, describiendo entre todos, la posible trayectoria escolar de un alumno dentro del sistema educativo entre dos años lectivos consecutivos.

Las tasas de transición describen exhaustivamente la trayectoria escolar de los alumnos matriculados en el sistema educativo en un año lectivo dado, lo cual significa, relativo al volumen de dicha matrícula, que entre todos suman el 100%. Es decir que el 100% de los alumnos matriculados en un grado/año de estudio, se distribuyen en términos relativos, entre cuatro situaciones posibles, promoción efectiva, repitencia, reinscripción y abandono interanual. Es por ello, que la tasa de abandono interanual suele construirse como diferencia entre dicho 100% y la sumatoria del resto de las tasas. Su fórmula es:

$$TAI_{i;n}^t = 100 - TPE_{i;n}^t - TR_{i;n}^t \;\; ; i = 1,\ldots,k$$

Ecuación 5. Tasa de abandono interanual de un grado/año.

En donde:

$TAI_{i;n}^t$ = Tasa de abandono interanual del grado/año de estudio *i* del nivel n en el año *t*.

$TPE_{i;n}^t$ = Tasa de promoción efectiva del grado/año de estudio *i* del nivel *n* en el año *t*.

$TR_{i;n}^t$ = Tasa de repitencia del grado/año de estudio *i* del nivel *n* en el año *t*.

i = Grado/año de estudio.

k = Último grado/año de estudio del nivel *n*.

n = Nivel de la enseñanza.

La fórmula de la tasa de abandono interanual de un nivel es:

$$TAI_n^t = 100 - TPE_n^t - TR_n^t$$

Ecuación 6. Tasa de abandono interanual de un nivel.

5.1.4.2 Codificación de estos 3 indicadores en MongoDB Atlas

En MongoDB Atlas, en la parte superior de un gráfico o cuadro, se puede colocar una consulta o una agregación para preprocesar los datos de una o varias colecciones. En el

Script **14**, se observa la sentencia de agregación que se coloca en la parte superior de un cuadro, y en este caso, se utiliza para calcular los 3 indicadores explicados más arriba:

```
[ {  $group: {
        _id: "$anio",   "suma_de_matr_por_doc_2_a_6": {$push: $sum:
[{$slice: ["$matricula.secundaria",1,5]}]  }}
,"suma_de_alumnos_nuevos_2_a_6_por_doc":
       {$push:                 {$subtract:                 [{$sum:[
{$sum:[{$slice:["$matricula.secundaria",1,5]}] }
,               {$sum:               [{$sum:               [{$slice:
["$promovidos.secundaria.findeanio",5,1]}]   },   {$sum:   [{$slice:
["$promovidos.secundaria.examen",5,1]}]}
, $sum:[{$slice:["$promovidos.secundaria.regulares",5,1]}] }
, {$sum:[{$slice:["$promovidos.secundaria.otros",5,1]}] }      ]} ]
}, {$sum: [{$slice:["$repitentes.secundaria",1,5]}] }  ]} }
,"suma_alumnos_nuevos_1":
       {$push: {$subtract:[
                              {$sum:
[{$slice:["$matricula.secundaria",1,1]}] }
, {$sum: [{$slice:["$repitentes.secundaria",1,1]}] }  ]} }
,"suma_alumnos_nuevos_2":
       {                    $push:                    {$subtract:[
{$sum:[{$slice:["$matricula.secundaria",2,1]}] }
, {$sum:[{$slice:["$repitentes.secundaria",2,1]}] }   ]} }
,"suma_alumnos_nuevos_3":
       {                    $push:                    {$subtract:[
{$sum:[{$slice:["$matricula.secundaria",3,1]}] }
, {$sum:[{$slice:["$repitentes.secundaria",3,1]}] }    ]} }
,"suma_alumnos_nuevos_4":
       {                    $push:                    {$subtract:[
{$sum:[{$slice:["$matricula.secundaria",4,1]}] }
, {$sum:[{$slice:["$repitentes.secundaria",4,1]}] }   ]} }
,"suma_alumnos_nuevos_5":
       { $push: {$subtract:[
{$sum:[{$slice:["$matricula.secundaria",5,1]}] }
,{$sum:[{$slice:["$repitentes.secundaria",5,1]}] }   ]}   }
,"suma_alumnos_nuevos_6":
```

```
        {                        $push:                        {$sum:[
{$sum:[{$slice:["$promovidos.secundaria.findeanio",5,1]}] }
,{$sum:[{$slice:["$promovidos.secundaria.examen",5,1]}] }
,{$sum:[{$slice:["$promovidos.secundaria.regulares",5,1]}] }
,{$sum:[{$slice:["$promovidos.secundaria.otros",5,1]}] }     ]  } }
,"suma_de_alumnos_repitentes_2_a_6_por_doc":        {        $push:
{$sum:[{$slice:["$repitentes.secundaria",1,5]}] }   }
,"suma_de_alumnos_repitentes_1_a_6_por_doc":        {        $push:
{$sum:[{$slice:["$repitentes.secundaria",0,6]}] }   }
,"suma_promovidos_6_por_doc":               {               $push:
{$sum:[{$sum:[{$slice:["$promovidos.secundaria.findeanio",5,1]}] }
,$sum:[{$slice:["$promovidos.secundaria.examen",5,1]}] }
,$sum:[{$slice:["$promovidos.secundaria.regulares",5,1]}] }
,{$sum:[{$slice:["$promovidos.secundaria.otros",5,1]}] }     ]}  }
,"suma_de_alumnos_anio_anterior_1_a_6_por_doc":      {       $push:
{$sum:[{$slice:["$matricula_anio_ant.secundaria",0,6]}] }   }
,"suma_alumnos_anio_anterior_1":             {              $push:
{$sum:[{$slice:["$matricula_anio_ant.secundaria",0,1]}] }   }
,"suma_alumnos_anio_anterior_2":             {              $push:
{$sum:[{$slice:["$matricula_anio_ant.secundaria",1,1]}] }   }
,"suma_alumnos_anio_anterior_3":             {              $push:
{$sum:[{$slice:["$matricula_anio_ant.secundaria",2,1]}] }   }
,"suma_alumnos_anio_anterior_4":             {              $push:
{$sum:[{$slice:["$matricula_anio_ant.secundaria",3,1]}] }   }
,"suma_alumnos_anio_anterior_5":             {              $push:
{$sum:[{$slice:["$matricula_anio_ant.secundaria",4,1]}] }   }
,"suma_alumnos_anio_anterior_6":             {              $push:
{$sum:[{$slice:["$matricula_anio_ant.secundaria",5,1]}]   }       }
}     },
    {$project: {
        "description": 1,
        "Matr_Secu_2a6": {
          $reduce: {
            input: "$suma_de_matr_por_doc_2_a_6",
            initialValue: 0,
            in: { $sum: [ "$$value", "$$this" ] } }  }
        ,"Alu_Repit": {
          $reduce: {
            input: "$suma_de_alumnos_repitentes_2_a_6_por_doc",
            initialValue: 0,
            in: { $sum: [ "$$value", "$$this" ] } } }
        ,"Alu_Repi_1_a_6": {
          $reduce: {
            input: "$suma_de_alumnos_repitentes_1_a_6_por_doc",
            initialValue: 0,
            in: { $sum: [ "$$value", "$$this" ] } } }
         ,"Alu_nuevos": {
          $reduce: {
            input: "$suma_de_alumnos_nuevos_2_a_6_por_doc",
            initialValue: 0,
            in: { $sum: [ "$$value", "$$this" ] } } }
       ,"Alu_anio_anterior": {
```

```
          $reduce: {
            input: "$suma_de_alumnos_anio_anterior_1_a_6_por_doc",
            initialValue: 0,
            in: { $sum: [ "$$value", "$$this" ] } } }
           ,"Alu_Pormovidos_6": {
          $reduce: {
            input: "$suma_promovidos_6_por_doc",
            initialValue: 0,
            in: { $sum: [ "$$value", "$$this" ] } } }
           ,"GA_1_Alu_Nuevos": {
          $reduce: {
            input: "$suma_alumnos_nuevos_1",
            initialValue: 0,
            in: { $sum: [ "$$value", "$$this" ] } } }
     ,"GA_2_Alu_Nuevos": {
          $reduce: {
            input: "$suma_alumnos_nuevos_2",
            initialValue: 0,
            in: { $sum: [ "$$value", "$$this" ] } }  }
     ,"GA_3_Alu_Nuevos": {
          $reduce: {
            input: "$suma_alumnos_nuevos_3",
            initialValue: 0,
            in: { $sum: [ "$$value", "$$this" ] } }  }
     ,"GA_4_Alu_Nuevos": {
          $reduce: {
            input: "$suma_alumnos_nuevos_4",
            initialValue: 0,
            in: { $sum: [ "$$value", "$$this" ] } }  }
     ,"GA_5_Alu_Nuevos": {
          $reduce: {
            input: "$suma_alumnos_nuevos_5",
            initialValue: 0,
            in: { $sum: [ "$$value", "$$this" ] } } }
       ,"GA_1_Anterior": {
          $reduce: {
            input: "$suma_alumnos_anio_anterior_1",
            initialValue: 0,
            in: { $sum: [ "$$value", "$$this" ] } } }
       ,"GA_2_Anterior": {
          $reduce: {
            input: "$suma_alumnos_anio_anterior_2",
            initialValue: 0,
            in: { $sum: [ "$$value", "$$this" ] } } }
       ,"GA_3_Anterior": {
          $reduce: {
            input: "$suma_alumnos_anio_anterior_3",
            initialValue: 0,
            in: { $sum: [ "$$value", "$$this" ] } }  }
       ,"GA_4_Anterior": {
          $reduce: {
            input: "$suma_alumnos_anio_anterior_4",
```

```
        initialValue: 0,
        in: { $sum: [ "$$value", "$$this" ] }  } }
,"GA_5_Anterior": {
    $reduce: {
        input: "$suma_alumnos_anio_anterior_5",
        initialValue: 0,
        in: { $sum: [ "$$value", "$$this" ] } } }
,"GA_6_Anterior": {
    $reduce: {
        input: "$suma_alumnos_anio_anterior_6",
        initialValue: 0,
        in: { $sum: [ "$$value", "$$this" ] } } }}}]
```

Script 14. Agregación para el cálculo de los indicadores. Fuente propia.

Aunque dicha sentencia parezca complicada, resulta necesario recordar que se están relacionando datos de 13 años.

5.1.4.3 Indicadores educativos de primaria.

En la Figura 71, se observan la evolución de los indicadores educativos del nivel primario: la tasa de promoción efectiva del nivel, la tasa de repitencia del nivel, la tasa de abandono interanual del nivel, y la tasa de promoción efectiva por grado. A la fecha de publicación de este trabajo, el Ministerio de Capital Humano, sólo tenía indicadores hasta el año 2021.

INDICADORES EDUCATIVOS-PRIMARIA-Educación Común

Principales tasas de la EDUCACIÓN COMÚN Primaria por AÑO-Tasa de Promoción efectiva por grado y año

Año	T_Promoción_Efectiva	T_Repitencia	T_Abandono_Intera...	TPrEf_1_grado	TPE_2	TPE...	TPE_4	TPE_5	TPE_6
2,010	89.71%	7.75%	2.54%	84.2%	90.8%	91.0%	89.9%	90.2%	92.9%
2,011	91.27%	5.86%	2.85%	86.5%	92.2%	92.1%	90.9%	91.8%	94.6%
2,012	94.30%	3.09%	2.60%	92.9%	95.6%	94.8%	93.6%	93.2%	95.7%
2,013	96.65%	2.24%	1.11%	98.1%	98.6%	95.9%	95.5%	95.3%	96.8%
2,014	95.97%	2.48%	1.55%	98.2%	98.0%	94.2%	94.7%	93.7%	97.3%
2,015	95.71%	2.76%	1.52%	97.2%	98.8%	93.9%	94.2%	93.9%	96.7%
2,016	93.35%	3.61%	3.05%	95.2%	96.5%	90.7%	91.2%	90.5%	95.9%
2,017	94.72%	4.08%	1.20%	95.7%	97.3%	92.7%	94.0%	92.7%	96.0%
2,018	94.49%	3.29%	2.21%	95.2%	97.5%	90.3%	93.3%	94.1%	96.8%
2,019	95.70%	3.15%	1.15%	97.2%	97.3%	93.1%	95.3%	94.9%	96.7%
2,020	97.39%	2.02%	0.60%	98.4%	99.4%	95.2%	97.1%	96.5%	97.7%
2,021	97.22%	1.92%	0.86%	98.0%	98.5%	96.3%	96.3%	95.8%	98.5%
2,022	98.57%	0.97%	0.46%	99.3%	99.5%	98.3%	97.5%	97.9%	98.9%

Figura 71. Indicadores educativos del nivel primario: tasa de promoción efectiva, tasa de repitencia y tasa de abandono interanual por año. TPE por grado. Fuente: Indicadores educativos de Argentina.gob.ar y propia.

Obsérvese que la provincia de Formosa mejoró este indicador en la última década. Se resalta nuevamente el poder de análisis del MongoDB Atlas, para analizar, los distintos años y grados, colocando distinta gama de colores.

5.1.4.4 Indicadores educativos de secundario.

En la Figura 72 se observan la evolución de los indicadores educativos del nivel secundario: la tasa de promoción efectiva del nivel, la tasa de repitencia del nivel, la tasa de abandono interanual del nivel, y la tasa de promoción efectiva por año, años 2010 a 2021.

INDICADORES EDUCATIVOS-SECUNDARIA-Educación Común

Principales tasas de la EDUCACIÓN COMÚN Secundaria por AÑO LECTIVO-Tasa de Promoción efectiva por año grado y año

Año	T_Promoción_Efectiva	T_Repitencia	T_Abandono_Intera...	TPrEf_1_Año	TPE_2	TPE_3	TPE_4	TPE_5	TPE_6
2.010	78,88%	9,41%	11,71%	75,2%	75,3%	80,9%	80,1%	88,5%	78,7%
2.011	75,26%	8,75%	15,99%	69,7%	69,2%	75,1%	79,5%	87,0%	79,6%
2.012	78,60%	9,53%	11,87%	75,9%	76,1%	80,5%	81,0%	84,0%	76,1%
2.013	80,58%	8,39%	11,03%	78,2%	79,0%	82,6%	82,4%	86,1%	76,5%
2.014	80,44%	8,85%	10,72%	78,4%	78,1%	81,9%	82,2%	86,7%	77,1%
2.015	77,86%	10,55%	11,59%	75,6%	75,5%	79,4%	79,8%	84,3%	74,3%
2.016	77,88%	10,72%	11,39%	75,9%	75,2%	80,0%	79,3%	83,5%	75,2%
2.017	78,61%	10,57%	10,82%	75,0%	76,7%	81,0%	81,3%	84,2%	75,8%
2.018	80,42%	9,26%	10,32%	76,6%	78,6%	83,3%	82,7%	85,1%	78,2%
2.019	81,15%	8,24%	10,61%	78,2%	77,6%	83,3%	83,2%	88,4%	78,3%
2.020	89,62%	4,11%	6,27%	89,3%	91,5%	93,4%	90,6%	93,1%	76,8%
2.021	79,92%	8,68%	11,40%	87,9%	76,7%	78,7%	80,3%	79,6%	74,4%

Figura 72. Indicadores educativos del nivel secundario: tasa de promoción efectiva, tasa de repitencia y tasa de abandono interanual por año. TPE por año de estudio. Fuente: Indicadores educativos de Argentina.gob.ar y propia.

Obsérvese, por ejemplo, que la tasa de promoción efectiva del año 2021/2022 del nivel secundario es 79.92% y coincide con el valor publicado por Nación. Este valor se consigue, considerando los promovidos de fin de año, y los promovidos por exámenes. Sin embargo, con la pandemia de COVID-19, se implementó en el país la promoción acompañada, que permite promocionar a un estudiante de un grado/año al subsiguiente, siempre y cuando en el transcurso del mismo pueda garantizársele el logro de los aprendizajes no acreditados en el año anterior, y los distintos operadores/directores que registraban estos datos, colocaron promovidos en otros campos, como ser **promovidos regulares**, y otros **promovidos**. Por esta razón, considero necesario sumar todos los promovidos, desde la pandemia COVID-19, como se observa en la agregación de la Figura 73.

```
[
  {
    $group: {
      _id: "$anio",
      "suma_de_matr_por_doc_2_a_6": { $push: {$sum:[{$slice:["$matricula.secundaria",1,5]}] }}
      ,"suma_de_alumnos_nuevos_2_a_6_por_doc":
      { $push: {$subtract:[
                {    $sum:[      {$sum:[{$slice:["$matricula.secundaria",1,5]}] }
                                                                  , {$sum:[{$sum:[{$slice:["$promovidos.secundaria.findeanio",5,1]}] }
                                                                                {$sum:[{$slice:["$promovidos.secundaria.examen",5,1]}] }
                                                                                ,{$sum:[{$slice:["$promovidos.secundaria.regulares",5,1]}] }
                                                                                ,{$sum:[{$slice:["$promovidos.secundaria.otros",5,1]}] }
                                                                                ]}
                                                                                
                                                                                ]
                                                                              }
                                                                  , {$sum:[{$slice:["$repitientes.secundaria",1,5]}] }
                                                                  ]}
      }
      ,"suma_promovidos_6_por_doc": { $push:  {$sum:[{$sum:[{$slice:["$promovidos.secundaria.findeanio",5,1]}] }
                                                                                {$sum:[{$slice:["$promovidos.secundaria.examen",5,1]}] }
                                                                                ,{$sum:[{$slice:["$promovidos.secundaria.regulares",5,1]}] }
                                                                                ,{$sum:[{$slice:["$promovidos.secundaria.otros",5,1]}] }
                                                                                ]}  }
```

Figura 73. Ajuste necesario en el cálculo de la tasa de promoción efectiva después de la pandemia COVID-19. Fuente propia.

INDICADORES EDUCATIVOS-SECUNDARIA-Educación Común

Principales tasas de la EDUCACIÓN COMÚN Secundaria por AÑO LECTIVO-Tasa de Promoción efectiva por año grado y año

Año	T_Promoción_Efectiva	T_Repitencia	T_Abandono_Intera...	TPrEf_1_Año	TPE_2	TPE_3	TPE_4	TPE_5	TPE_6
2.010	79.09%	9.41%	11.50%	75.2%	75.3%	80.9%	80.1%	88.5%	80.8%
2.011	75.52%	8.75%	15.73%	69.7%	69.2%	75.1%	79.5%	87.0%	82.1%
2.012	78.99%	9.53%	11.48%	75.9%	76.1%	80.5%	81.0%	84.0%	79.5%
2.013	80.87%	8.39%	10.74%	78.2%	79.0%	82.6%	82.4%	86.1%	79.1%
2.014	80.73%	8.85%	10.42%	78.4%	78.1%	81.9%	82.2%	86.7%	79.7%
2.015	78.29%	10.55%	11.16%	75.6%	75.5%	79.4%	79.8%	84.3%	78.1%
2.016	78.09%	10.72%	11.18%	75.9%	75.2%	80.0%	79.3%	83.5%	77.1%
2.017	78.92%	10.57%	10.51%	75.0%	76.7%	81.0%	81.3%	84.2%	78.6%
2.018	80.57%	9.26%	10.17%	76.6%	78.6%	83.3%	82.7%	85.1%	79.6%
2.019	81.45%	8.24%	10.31%	78.2%	77.6%	83.3%	83.2%	88.4%	80.8%
2.020	89.74%	4.11%	6.15%	89.3%	91.5%	93.4%	90.6%	93.1%	77.8%
2.021	80.21%	8.68%	11.11%	87.9%	76.7%	78.7%	80.3%	79.6%	76.6%
2.022	78.48%	10.47%	11.06%	81.6%	78.0%	78.6%	76.5%	78.4%	76.4%

Figura 74. Tasa de promoción efectiva ajustada, considerando todos los promovidos. Fuente: Indicadores educativos de Argentina.gob.ar hasta 2021, y propia.

El resultado que surge, y que se observa en la Figura 74 cambia el valor de la tasa de promoción efectiva (TPE) de 2021 de 79,92% a 80,21%. De modo similar, también sube la TPE de 2021 de sexto año de 74,4% a 76,6%, y baja la tasa de abandono interanual del mismo año de 11,40 a 11,11%, que surge por diferencia. La tasa de repitencia de 2021 no tiene por qué moverse, y queda en 8,68%. Se observa también en esta, que ya están calculadas las tasas de promoción efectiva, de repitencia y abandono de 2022, y el Ministerio de Capital Humano aún no lo hace, a la fecha de la última revisión de este trabajo (31/08/2024) [37].

5.1.5 Gráficos estadísticos en Atlas

Los gráficos se obtienen de manera similar, ocupando campos calculados, como ser:

```
{$sum: [{$arrayElemAt: [ "$secundario.alumnos", 5 ] }]}
```

Sólo se muestran 3 gráficos a modo ilustrativo.

En la Figura 75 se observa la matrícula por departamento y nivel del año 2022, indicando la matrícula total de la educación común en cada departamento.

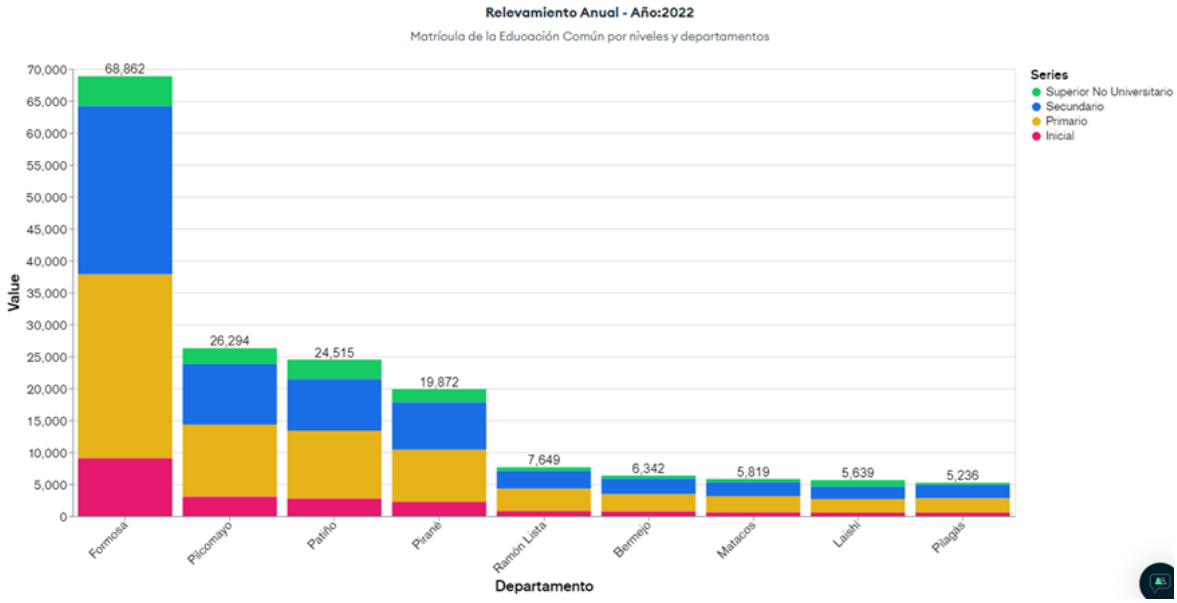


Figura 75. Gráfico de matrícula por departamento y nivel del año 2022, de la educación común. Fuente: Anuario estadístico 2022 y propia.

En la Figura 76 se observa la matrícula de la educación común por nivel en 2022.

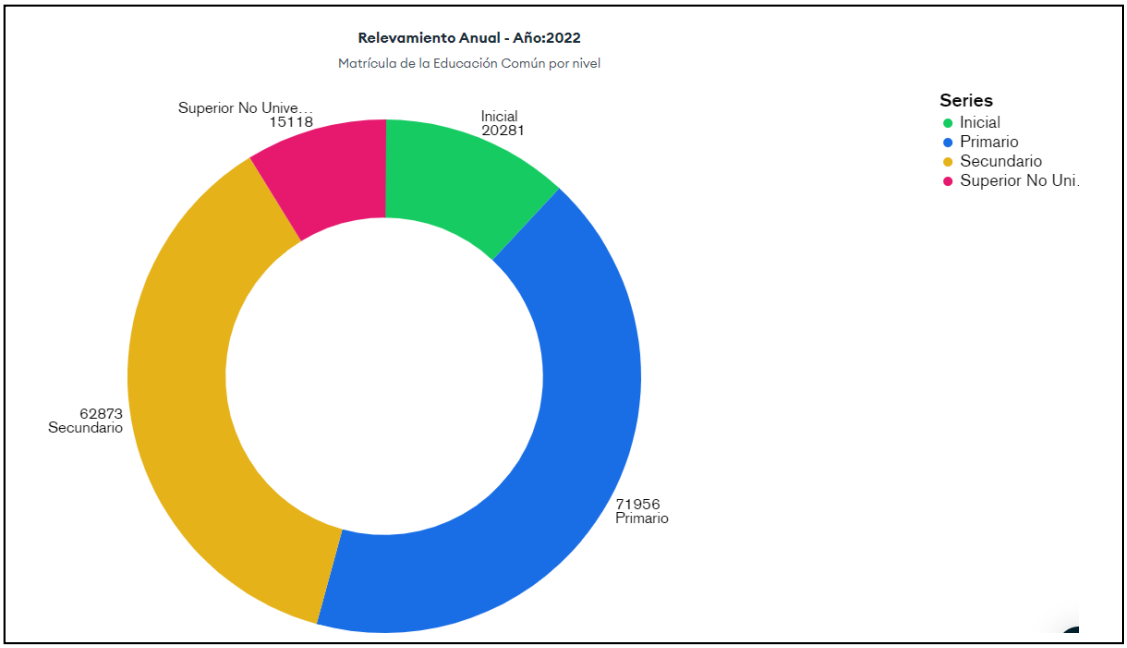


Figura 76. Matrícula de la educación común por nivel en 2022. Fuente: Anuario 2022 y propia.

En la Figura 77 se observa la matrícula de la educación común por departamento y nivel en 2022, indicando la matrícula en cada nivel y en cada departamento.

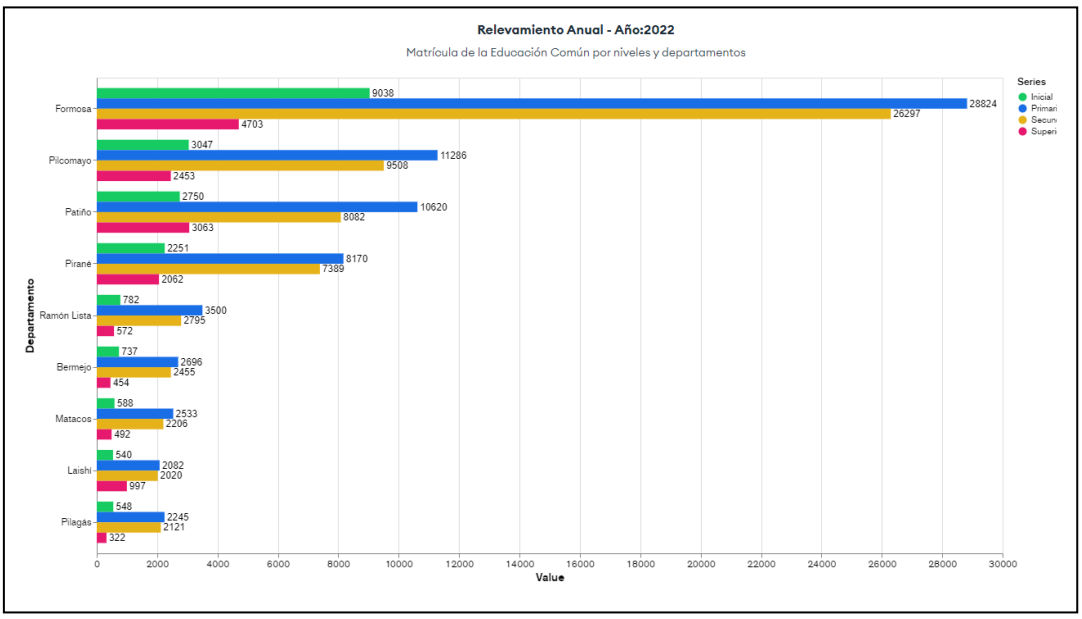


Figura 77. Composición de matrícula de la educación común por nivel y departamento en 2022. Fuente: Anuario estadístico 2022 y propia.

5.2 Operativos ONE y Aprender

5.2.1 Aprender primaria 2021-censal

5.2.1.1 Aprender 2021. Matemática y Lengua. Resultados de estudiantes de Desempeños Avanzado y Satisfactorio

Para poder acceder a los valores que tienen las jurisdicciones se ocupan campos calculados. Por ejemplo, para obtener el valor del desempeño "Avanzado" de 2018 en Matemática, se lo debe referenciar de la siguiente forma (
Script **15**):

```
{$arrayElemAt:
[ "$matematica.resultados_grales_anios.anio_2018", 0 ] }
```

Script 15. Campo calculado: obtener desempeño avanzado en Matemática de 2018. Fuente propia.

Para calcular la suma de los desempeños Avanzado y Satisfactorio en Matemática de 2018, se hace (
Script **16**):

```
{$sum: [{$arrayElemAt:
[ "$matematica.resultados_grales_anios.anio_2018", 0 ]}
,{$arrayElemAt:
[ "$matematica.resultados_grales_anios.anio_2018", 1 ] }]}
```

Script 16. Campo calculado: obtener desempeño avanzado y satisfactorios en Matemática de 2018. Fuente propia.

Para calcular la diferencia de los valores de Matemática, de desempeños "Avanzado" y "Satisfactorio" de Aprender 2021 menos iguales desempeños de Aprender 2018, se utiliza el siguiente campo calculado (
Script **17**):

```
Dif_Apr_Mat_21_18= {$subtract:[{$sum:[{$arrayElemAt: [
"$matematica.resultados_grales_anios.anio_2021", 0 ]
},{$arrayElemAt: [
"$matematica.resultados_grales_anios.anio_2021", 1 ]
}]},{$sum:[{$arrayElemAt: [
"$matematica.resultados_grales_anios.anio_2018", 0 ]
},{$arrayElemAt: [
"$matematica.resultados_grales_anios.anio_2018", 1 ] }]}]}
```

Script 17. Campo calculado: diferencia en Matemática entre Aprender 2021 y Aprender 2018. Fuente propia.

Los resultados de Aprender 2021, indicando los desempeños avanzado y satisfactorio, tanto en Lengua como en Matemática, se observan en la Figura 78. Los resultados obtenidos coinciden con el resultado informado por los medios de comunicación.

Aprender-Matemática y Lengua- Resultados de estudiantes de Desempeños Avanzado y Satisfactorio de Aprender 2021

Distribución de estudiantes por niveles de desempeños Avanzado y Satisfactorio -Ranking según desempeño en matemática

Provincia	Ap21_L_Resultados_grales_AvSa	Ap21_M_Resultados_grales_AvSa
CABA	73,6%	72,3%
Córdoba	66,3%	68,2%
La Pampa	61,0%	60,5%
Mendoza	57,1%	59,3%
Rio Negro	57,4%	57,5%
Tierra del Fuego	63,7%	55,2%
Formosa	53,6%	55,0%
Santa Fé	55,4%	54,8%
Jujuy	54,0%	54,3%
Buenos Aires	56,8%	54,0%
Tucumán	52,5%	53,8%
Salta	51,4%	53,8%
Chubut	58,3%	53,4%
Neuquén	54,0%	52,0%
San Luis	52,4%	50,7%
Santiago del Estero	45,0%	49,9%

Figura 78. Aprender Matemática y Lengua. Resultados de estudiantes de Desempeños Avanzado y Satisfactorio de Aprender 2021. Fuente: Aprender 2021 y propia.

Se observa que Formosa se posiciona entre las 7 primeras jurisdicciones en Matemática, después de Ciudad de Buenos Aires (72.2%), Córdoba (68.2%), La Pampa (60.5%), Mendoza (59.3%), Río Negro (57.5%), Tierra del Fuego (55.2%) y Formosa (55%). En Lengua, Formosa cae a la posición 13, con un 53.6% en la suma de los desempeños avanzado y satisfactorio.

5.2.1.2 Aprender 2021. Matemática. Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2021 y 2018

Ahora, si se analizan las mejoras y caídas de desempeños avanzado y satisfactorio, entre las pruebas Aprender 2018 y 2021, cambia la situación, como se observa en la Figura 79.

Aprender-Matemática-Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2021 y 2018

Distribución de estudiantes por nivel de desempeño: Avanzado y Satisfactorio-Ranking de jurisdicciones que mejoraron y que cayeron en el desempeños en Matemática entre ambos períodos

Provincia	Apr2018_Mat_Av...	Apr2018_Mat_Satis	Apr2018_Mat_AvSa	Apr2021_Mat_Av	Apr2021_Mat_Sat	Apr2021_Mat_AvSa	Dif_Apr_Mat_21_18_...
Tucumán	16,40%	34,70%	51,10%	15,50%	38,30%	53,80%	2,70%
Mendoza	17,20%	39,60%	56,80%	17,90%	41,40%	59,30%	2,50%
Santiago del Estero	16,90%	32,30%	49,20%	15,00%	34,90%	49,90%	0,70%
Formosa	18,10%	36,30%	54,40%	17,00%	38,00%	55,00%	0,60%
Jujuy	15,00%	38,70%	53,70%	13,30%	41,00%	54,30%	0,60%
Córdoba	29,00%	39,20%	68,20%	26,10%	42,10%	68,20%	0,00%
Santa Cruz	11,90%	37,50%	49,40%	9,10%	39,60%	48,70%	-0,70%
Tierra del Fuego	16,50%	40,40%	56,90%	12,40%	42,80%	55,20%	-1,70%
San Juan	14,90%	36,40%	51,30%	11,20%	38,30%	49,50%	-1,80%
Buenos Aires	18,30%	38,00%	56,30%	14,40%	39,60%	54,00%	-2,30%
Catamarca	9,30%	33,20%	42,50%	6,70%	33,20%	39,90%	-2,60%
CABA	34,10%	41,20%	75,30%	27,40%	44,90%	72,30%	-3,00%
Santa Fé	20,70%	37,50%	58,20%	15,50%	39,30%	54,80%	-3,40%
Rio Negro	21,20%	41,00%	62,20%	15,00%	42,50%	57,50%	-4,70%
Chubut	17,70%	40,60%	58,30%	12,60%	40,80%	53,40%	-4,90%
Neuquén	16,50%	40,50%	57,00%	13,00%	39,00%	52,00%	-5,00%
Corrientes	15,90%	34,90%	50,80%	11,70%	33,70%	45,40%	-5,40%

Figura 79. Aprender-Matemática. Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2021 y 2018: Distribución de estudiantes por nivel de desempeño: Avanzado y Satisfactorio. Ranking de jurisdicciones que mejoraron y que cayeron en el desempeño. Fuente: Aprender 2021 y propia.

En este caso, la provincia de Formosa mejoró su nivel de desempeño a pesar de la pandemia. Los resultados obtenidos son similares a los de los medios: *"En matemática ocurrió una de las sorpresas. Contrario a lo esperado, hubo cinco provincias que, pese al cierre de escuelas, mejoraron sus resultados respecto a 2018. Fueron Mendoza (+2,5%), Tucumán (+2,4%), Jujuy (+0,7%), Santiago del Estero (+0,7%) y Formosa (+0,6%)* [38]*",* aunque el ranking procesado en este trabajo, no coincide plenamente, ya que Formosa no aparece quinta, sino cuarta.

5.2.1.3 Aprender 2021. Lengua. Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2021 y 2018.

Aprender-Lengua-Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2021 y 2018

Distribución de estudiantes por nivel de desempeño: Avanzado y Satisfactorio-Ranking de jurisdicciones en la caída de desempeños en Lengua entre ambos períodos

Provincia	Apr2018_Leng_A...	Apr2018_Leng_S...	Apr2018_Leng_A...	Apr2021_Leng...	Apr2021_Leng_Sat	Apr2021_Leng_AvSa	Dif_Apr_Leng_21_18...
CABA	53,8%	34,1%	87,9%	38,0%	35,6%	73,6%	-14,3%
Formosa	29,3%	40,1%	69,4%	20,8%	32,8%	53,6%	-15,8%
Córdoba	45,5%	37,2%	82,7%	30,4%	35,9%	66,3%	-16,4%
Tucumán	31,2%	39,0%	70,2%	19,5%	33,0%	52,5%	-17,7%
Buenos Aires	35,5%	40,0%	75,5%	22,7%	34,1%	56,8%	-18,7%
Mendoza	33,5%	42,4%	75,9%	22,0%	35,1%	57,1%	-18,8%
Santiago del Estero	25,5%	38,5%	64,0%	15,7%	29,3%	45,0%	-19,0%
Santa Fé	34,9%	39,8%	74,7%	21,7%	33,7%	55,4%	-19,3%
Jujuy	30,6%	43,1%	73,7%	18,2%	35,8%	54,0%	-19,7%
Tierra del Fuego	42,0%	41,6%	83,6%	25,5%	38,2%	63,7%	-19,9%
Santa Cruz	32,6%	44,6%	77,2%	20,2%	36,9%	57,1%	-20,1%
La Rioja	30,6%	40,9%	71,5%	18,4%	32,9%	51,3%	-20,2%
Chaco	21,8%	39,0%	60,8%	13,3%	26,5%	39,8%	-21,0%
San Juan	28,9%	41,8%	70,7%	16,6%	32,9%	49,5%	-21,2%
Río Negro	36,3%	42,5%	78,8%	22,4%	35,0%	57,4%	-21,4%
Catamarca	23,7%	41,9%	65,6%	13,7%	30,2%	43,9%	-21,7%

Figura 80. Aprender-Lengua. Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2021 y 2018. Distribución de estudiantes por nivel de desempeño: Avanzado y Satisfactorio. Ranking de jurisdicciones en la caída de desempeños entre ambos periodos. Fuente: Aprender 2021 y propia.

En el cuadro de la Figura 80, se observa que las jurisdicciones que menos sufrieron en Lengua el cierre escolar por la pandemia, fueron Ciudad de Buenos Aires que cayó 14.3%, Formosa 15,8% y Córdoba 16,4%.

5.2.1.4 Aprender 2021. Matemática. Diferencia de los desempeños avanzado y satisfactorio del sector estatal y del sector privado de Aprender 2021 y 2018.

A partir de la consulta de la Figura 81, poco hablan los medios:

La consulta de la diferencia de los desempeños avanzado y satisfactorio del **sector estatal** en Matemática del 2021 menos el de 2018 (Script 18):

```
{$subtract:[{$sum:[{$arrayElemAt: [
"$matematica.sector.estatal.anio_2021", 0 ] },{$arrayElemAt: [
"$matematica.sector.estatal.anio_2021", 1 ] }]},{$sum:[{
$arrayElemAt:[ "$matematica.sector.estatal.anio_2018", 0 ]},
{$arrayElemAt: [ "$matematica.sector.estatal.anio_2018",1]}]}]}
```

Script 18. Campo calculado de la diferencia en matemática del sector estatal, entre los años 2021 y 2018. Fuente propia.

Y muy similar, en el sector privado.

Aprender-Matemática_Sector de gestión-Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2021 y 2018

Distribución de estudiantes por nivel de desempeño según SECTOR DE GESTIÓN: Avanzado y Satisfactorio-Ranking alfabético-Formato condicional para jurisdicciones que mejoraron en el S.ESTATAL

Provincia	Ap18_M_SPrivado_AvSa	Ap21_M_SPrivado_AvSa	Dif_A21_18_M_SPrivado	Ap18_M_SEstatal_AvSa	Ap21_M_SEstatal_AvSa	Dif_A21_18_M_SEstatal
Buenos Aires	74,4%	71,6%	-2,8%	45,8%	45,0%	-0,8%
CABA	85,2%	80,1%	-5,1%	66,1%	65,2%	-0,9%
Catamarca	63,9%	57,1%	-6,8%	38,0%	35,6%	-2,4%
Chaco	66,1%	62,8%	-3,3%	43,2%	36,7%	-6,5%
Chubut	82,9%	68,7%	-14,2%	55,7%	51,1%	-4,6%
Corrientes	70,3%	68,2%	-2,1%	48,1%	42,3%	-5,8%
Córdoba	82,0%	78,2%	-3,8%	63,4%	64,6%	1,2%
Entre Rios	67,6%	61,4%	-6,2%	49,6%	41,6%	-8,0%
Formosa	67,0%	66,0%	-1,0%	53,0%	53,8%	0,8%
Jujuy	73,2%	68,1%	-5,1%	51,1%	52,5%	1,4%
La Pampa	82,1%	77,6%	-4,5%	68,9%	58,5%	-10,4%
La Rioja	75,4%	65,8%	-9,6%	45,9%	40,1%	-5,8%
Mendoza	78,8%	76,9%	-1,9%	51,7%	55,6%	3,9%
Misiones	66,7%	55,9%	-10,8%	51,3%	43,6%	-7,7%
Neuquén	75,4%	68,6%	-6,8%	54,3%	49,4%	-4,9%
Rio Negro	78,5%	73,6%	-4,9%	58,8%	54,1%	-4,7%
Salta	80,1%	73,4%	-6,7%	56,2%	50,4%	-5,8%

Figura 81. Aprender-Matemática Sector de gestión. Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2021 y 2018: Distribución de estudiantes por nivel de desempeño según SECTOR DE GESTIÓN: Avanzado y Satisfactorio. Se resaltan jurisdicciones que mejoran en el sector estatal. Fuente Aprender 2021 y propia.

Específicamente, Formosa cae 1% en el sector privado, y mejora 0.8% en el sector estatal.

5.2.1.5 Aprender 2021. Matemática. Diferencia de los desempeños avanzado y satisfactorio del ámbito urbano y el ámbito rural de Aprender 2021 y 2018.

Para el procesamiento de los resultados de niveles de desempeño avanzado y satisfactorio del **ámbito urbano de matemática**, la consulta para la diferencia de los períodos 2021 y 2018 (Script 19):

```
{$subtract:[{$sum:[{$arrayElemAt: [
"$matematica.ambito.urbano.anio_2021", 0 ] },{$arrayElemAt: [
"$matematica.ambito.urbano.anio_2021", 1 ]
}]},{$sum:[{$arrayElemAt:
["$matematica.ambito.urbano.anio_2018", 0 ] },{$arrayElemAt: [
"$matematica.ambito.urbano.anio_2018", 1 ] }]}]}
```

Script 19. Campo calculado de la diferencia del desempeño en matemática del ámbito urbano entre 2021 y 2018. Fuente propia.

Aprender-Matemática_Ámbito-Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2021 y 2018

Distribución de estudiantes por niveles de desempeños Avanzado y Satisfactorio según ÁMBITO-Ranking alfabético (en CABA y Tierra del Fuego se consignó 0 en A.Rural)

Provincia	Ap18_M_AUrbano_AvSa	Ap21_M_AUrbano_AvSa	Dif_A21_18_M_AUrbano	Ap18_M_ARural_AvSa	Ap21_M_ARural_AvSa	Dif_A21_18_M_ARural
Buenos Aires	56,4%	53,9%	-2,5%	55,9%	57,2%	1,3%
CABA	75,3%	72,3%	-3,0%	0,0%	0,0%	0,0%
Catamarca	44,9%	38,2%	-6,7%	38,5%	46,3%	7,8%
Chaco	41,7%	37,9%	-3,8%	63,6%	49,5%	-14,1%
Chubut	58,1%	53,3%	-4,8%	60,3%	54,9%	-5,4%
Corrientes	51,3%	44,5%	-6,8%	49,5%	49,5%	0,0%
Córdoba	68,0%	67,6%	-0,4%	71,2%	74,6%	3,4%
Entre Rios	53,4%	45,7%	-7,7%	61,4%	54,7%	-6,7%
Formosa	50,0%	52,7%	2,7%	67,7%	63,4%	-4,3%
Jujuy	52,9%	53,8%	0,9%	58,4%	58,5%	0,1%
La Pampa	69,4%	59,7%	-9,7%	75,4%	66,2%	-9,2%
La Rioja	49,7%	41,5%	-8,2%	48,3%	53,7%	5,4%
Mendoza	59,3%	59,7%	0,4%	49,7%	58,0%	8,3%
Misiones	52,5%	44,9%	-7,6%	56,4%	48,6%	-7,8%
Neuquén	56,5%	51,8%	-4,7%	66,5%	54,6%	-11,9%
Rio Negro	62,7%	57,4%	-5,3%	59,2%	58,9%	-0,3%
Salta	59,7%	53,7%	-6,0%	61,0%	54,0%	-7,0%

Figura 82. Matemática - Ámbito. Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2021 y 2018: Distribución de estudiantes por niveles de desempeños Avanzado y Satisfactorio según ÁMBITO. Ranking alfabético (en CABA y y Tierra del Fuego se consignó 0 en el ámbito rural). Fuente: Aprender 2021 y propia.

Como se observa en la Figura 84, Formosa mejora en el ámbito urbano en 2.7%, y cae 3.7% en el ámbito rural. Se da una situación particular en el ámbito rural de la provincia de Corrientes, ya que la diferencia de ambas pruebas da 0%, pero cambian el nivel de desempeño avanzado 18.8% y satisfactorio 30.7% en 2018 y 17.2% y 32.3% en 2021, respectivamente. En CABA y Tierra del Fuego, se consignó 0 en el ámbito rural, ya que no tienen dicho ámbito.

5.2.1.6 Aprender 2021. Matemática. Diferencia de los desempeños avanzado y satisfactorio del género varones y del género mujeres de Aprender 2021 y 2018.

Para el procesamiento de los resultados de niveles de desempeño avanzado y satisfactorio del **<u>género mujeres de matemática</u>**, la consulta para la diferencia de los períodos 2021 y 2018 (Script 20):

```
{$subtract:[{$sum:[{$arrayElemAt: [
"$matematica.genero.mujeres.anio_2021", 0 ] },{$arrayElemAt: [
"$matematica.genero.mujeres.anio_2021", 1 ]
}]},{$sum:[{$arrayElemAt: [
"$matematica.genero.mujeres.anio_2018", 0 ] },{$arrayElemAt: [
"$matematica.genero.mujeres.anio_2018", 1 ] }]}]}
```

Script 20. Campo calculado de la diferencia desempeño de avanzado y satisfactorio del género mujeres de matemática entre 2021 y 2018.

Aprender-Matemática_Género-Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2021 y 2018

Distribución de estudiantes por niveles de desempeños Avanzado y Satisfactorio según GÉNERO-Ranking alfabético

Provincia	Ap18_M_GMujeres_Av...	Ap21_M_GMujeres_Av...	Dif_A21_18_M_GMujeres	Ap18_M_GVarones_A...	Ap21_M_GVarones_Av...	Dif_A21_18_M_GVarones
Buenos Aires	55,7%	53,2%	-2,5%	57,8%	59,1%	1,3%
CABA	75,6%	70,7%	-4,9%	76,7%	75,9%	-0,8%
Catamarca	43,0%	38,9%	-4,1%	42,4%	43,7%	1,3%
Chaco	44,0%	39,6%	-4,4%	47,2%	45,9%	-1,3%
Chubut	57,5%	52,5%	-5,0%	59,7%	57,8%	-1,9%
Corrientes	49,4%	44,4%	-5,0%	52,7%	49,0%	-3,7%
Córdoba	67,4%	66,9%	-0,5%	69,3%	71,7%	2,4%
Entre Rios	53,8%	47,1%	-6,7%	55,7%	49,9%	-5,8%
Formosa	53,9%	55,4%	1,5%	56,0%	57,5%	1,5%
Jujuy	52,8%	53,2%	0,4%	54,8%	58,3%	3,5%
La Pampa	71,3%	59,2%	-12,1%	69,5%	64,3%	-5,2%
La Rioja	48,9%	42,5%	-6,4%	50,3%	45,3%	-5,0%
Mendoza	56,7%	59,0%	2,3%	57,2%	63,0%	5,8%
Misiones	56,7%	59,0%	2,3%	55,3%	49,0%	-6,3%
Neuquén	56,0%	55,0%	-1,0%	58,5%	53,7%	-4,8%
Rio Negro	61,8%	57,7%	-4,1%	63,7%	61,2%	-2,5%
Salta	59,3%	53,3%	-6,0%	60,7%	58,2%	-2,5%

Figura 83. Matemática - Género. Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2021 y 2018: Distribución de estudiantes por niveles de desempeños Avanzado y Satisfactorio según GÉNERO. Ranking alfabético. Fuente: Aprender 2021 y propia.

Como se observa en la Figura 83, en ambos géneros informados, en Formosa aumentaron en una proporción similar, aunque la combinación de niveles de desempeño de avanzado y satisfactorio, cambia en ambos períodos.

5.2.1.7 Aprender 2021. Matemática. Diferencia de los desempeños avanzado y satisfactorio del NSE bajo y NSE alto de Aprender 2021.

Para el procesamiento de los resultados de niveles de desempeño avanzado y satisfactorio de los **niveles socioeconómicos del año 2021**, la consulta para la diferencia de los NSE bajo y alto (Script 21):

```
{$subtract:[{$sum:[{$arrayElemAt: [ "$matematica.NSE2021.bajo",
0 ] },{$arrayElemAt: [ "$matematica.NSE2021.bajo", 1 ]
}]},{$sum:[{$arrayElemAt: [ "$matematica.NSE2021.alto", 0 ]
},{$arrayElemAt: [ "$matematica.NSE2021.alto", 1 ] }]}]}
```

Script 21. Campo calculado de la diferencia en matemática de los NSE alto y bajo de 2021. Fuente propia.

Primaria_Aprender21-Matemática_Niveles Socioeconómicos-Diferencias entre porcentajes de estudiantes con desempeños Av. y Sat. de NSE bajo y alto

Distribución de estudiantes por niveles de desempeños Avanzado y Satisfactorio según NSE-Ranking diferencia entre NSE bajo menos NSE alto de 2021

Provincia	Ap21_M_NSBajo_AvSa	Ap21_M_NSMedio_AVSa	Ap21_M_NSAlto_AvSa	AP21_M_Dif_NSE_Ba_Al
Santiago del Estero	44.0%	52.8%	65.3%	-21.3%
Misiones	36.3%	48.8%	60.7%	-24.4%
Formosa	45.5%	60.2%	70.4%	-24.9%
Jujuy	35.5%	60.4%	67.5%	-32.0%
Córdoba	46.2%	69.2%	80.5%	-34.3%
Chaco	29.8%	43.6%	64.1%	-34.3%
La Rioja	24.6%	45.6%	59.7%	-35.1%
San Luis	30.7%	52.0%	66.7%	-36.0%
Catamarca	22.3%	43.4%	58.3%	-36.0%
Corrientes	29.1%	50.3%	65.4%	-36.3%
Salta	36.7%	59.4%	73.5%	-36.8%
Mendoza	37.0%	61.4%	75.7%	-38.7%
Tucumán	35.6%	58.4%	74.7%	-39.1%
Santa Cruz	18.9%	51.8%	60.2%	-41.3%

Figura 84. Aprender-Matemática Niveles Socioeconómicos 2021. Diferencias entre porcentajes de estudiantes con desempeño Av. y Sat. de NSE bajo y alto de 2021: Distribución de estudiantes por niveles de desempeños Avanzado y Satisfactorio según NSE. Ranking diferencia entre NSE bajo menos NSE alto de 2021. Fuente: Aprender 2021 y propia.

Como se observa en la Figura 84, el ranking de la diferencia de nivel socioeconómico bajo y alto de 2021, en los niveles de desempeño avanzado y satisfactorio de matemática, la encabeza Santiago del Estero con una diferencia 21.3%, continúa la provincia de Misiones con 24.4%, y Formosa con 24.9%, Chaco (34.3%) y Corrientes (36.3%). Estas provincias, están entre los 10 primeros y sus sistemas educativos tienen menos diferencias sociales en el aprendizaje que el resto del país.

5.2.1.8 Aprender 2021. Matemática. NSE alto en 2018 y 2021, y cuánto cae del 2018 al 2021 y NSE bajo en 2018 y 2021, y cuánto cae del 2018 al 2021.

Ahora bien, podemos analizar lo que sucede en el NSE alto en 2018 y 2021, y cuánto cae del 2018 al 2021 y NSE bajo en 2018 y 2021, y cuánto cae del 2018 al 2021 en Matemática. La diferencia del NSE bajó entre 2018 y 2021 en niveles de desempeño avanzado y satisfactorio de Matemática (Script 22):

```
{$subtract:[{$sum:[{$arrayElemAt: [
"$matematica.NSE_por_anio.bajo.anio_2021", 0 ] },{$arrayElemAt: [
"$matematica.NSE_por_anio.bajo.anio_2021", 1 ]
}]},{$sum:[{$arrayElemAt: [
"$matematica.NSE_por_anio.bajo.anio_2018", 0 ] },{$arrayElemAt: [
"$matematica.NSE_por_anio.bajo.anio_2018", 1 ] }]}]}
```

Script 22. Campo calculado de la diferencia entre matemática del NSE bajo de 2021 y NSE bajo de 2018. Fuente propia.

Aprender-Matemática_NSE alto y bajo por año-Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2021 y 2018

Distribución de estudiantes por niveles de desempeños Avanzado y Satisfactorio según NSE alto y bajo-Ranking diferencia NSE Bajo de los años 2018 a 2021

Provincia	Ap18_M_NSEaLTO_Av...	Ap21_M_NSEAlto_AvSa	Dif_A21_18_M_NSEAlto	Ap18_M_NSEBajo_Av...	Ap21_M_NSEBajo_AvSa	Dif_A21_18_M_NSEBajo
Tucumán	76,6%	74,7%	-1,9%	37,7%	35,6%	-2,1%
Mendoza	77,5%	75,7%	-1,8%	39,8%	37,0%	-2,8%
San Luis	72,3%	66,7%	-5,6%	36,6%	30,7%	-5,9%
CABA	87,8%	83,1%	-4,7%	47,2%	40,2%	-7,0%
Santiago del Estero	68,3%	65,3%	-3,0%	51,5%	44,0%	-7,5%
San Juan	77,9%	71,0%	-6,9%	34,4%	26,2%	-8,2%
Córdoba	84,3%	80,5%	-3,8%	55,4%	46,2%	-9,2%
Formosa	68,8%	70,5%	1,7%	54,7%	45,5%	-9,2%
Buenos Aires	77,8%	74,5%	-3,3%	36,1%	25,9%	-10,2%
Santa Fé	80,8%	74,5%	-6,3%	39,2%	28,9%	-10,3%
Santa Cruz	64,4%	60,2%	-4,2%	29,2%	18,8%	-10,4%
Jujuy	67,8%	67,5%	-0,3%	46,3%	35,5%	-10,8%
Chubut	77,7%	71,9%	-5,8%	41,3%	30,1%	-11,2%
Tierra del Fuego	71,3%	67,2%	-4,1%	35,5%	24,3%	-11,2%
Catamarca	61,2%	58,3%	-2,9%	34,0%	22,3%	-11,7%
Chaco	64,5%	64,1%	-0,4%	42,2%	29,8%	-12,4%
Misiones	70,6%	60,7%	-9,9%	51,6%	36,3%	-15,3%

Figura 85. Aprender-Matemática_NSE alto y bajo por año. Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2021 y 2018: Aprender-Matemática_NSE alto y bajo por año. Diferencias entre porcentajes de estudiantes con desempeño avanzado y satisfactorio de 2021 y 2018. Fuente: Aprender 2021 y propia.

Se observa en la Figura 85, que Tucumán es la que menos cayó en el NSE bajo, seguidas de Mendoza, San Luis, CABA, Santiago del Estero, San Juan, Córdoba y Formosa con 9.2% de caída (está entre las 10 primeras).

5.2.1.9 Aprender 2021. Matemática y Lengua. Promedio de desempeños avanzado y satisfactorio de relacionado el cociente de las sobreedades de sexto grado de 2021 y primero de 2016, con sus respectivas matriculas.

```
{$multiply:[{$divide:[{$sum:[{$arrayElemAt: [
"$lengua.resultados_grales_anios.anio_2021", 0 ] },{$arrayElemAt:
[ "$lengua.resultados_grales_anios.anio_2021", 1 ]
},{$arrayElemAt: [
"$matematica.resultados_grales_anios.anio_2021", 0 ]
},{$arrayElemAt: [
"$matematica.resultados_grales_anios.anio_2021", 1 ]
}]},2]},{$divide:[{$arrayElemAt: [
"$ra2016_ra2021_sobreedad.anio_2021_M6_S6_en_edad_teorica", 0 ]
},{$arrayElemAt: [
"$ra2016_ra2021_sobreedad.anio_2016_M1_S1_en_edad_teorica", 0 ]
}]}]}
```

Script 23. Campo calculado que relaciona el promedio de desempeño de lengua y matemática de Aprender 2021, relacionándolos con la sobreedad y matrícula del RA2021 y RA2016. Fuente propia.

Al procesar Aprender 2021 Matemática y Lengua y las matrículas de sexto grado en edad teórica del Relevamiento Anual 2021 y la matrícula de primer grado en edad teórica del Relevamiento Anual 2016, se ve que Formosa no queda relegada al séptimo lugar al final de la tabla, como muestra un medio del 09 de marzo de 2023 [39].

La consulta final de la multiplicación de Aprender 2021 con el RA2021-RA2016 (Script 23):

Aprender 2021-Matemática y Lengua y la relación de matrículas de edades teóricas de 6to.RA2021 y de 1ro.RA2016

Distribución de estudiantes por niveles de desempeños Avanzado y Satisfactorio -Ranking según multiplicación relación de edades teóricas del RA y desempeño APRENDER

Provi...	Ap21_L_Resultado...	Ap21_M_Result...	Ap21_Prom_L_M_A...	RA2016_Matricula_1_Edad_te...	RA2021_Matricula_6_edad_te...	Sobreedad_6_2021_...	Ap21_LM_...
CABA	73,6%	72,3%	73,0%	41.474	38.570	0,93	67,842%
Córdoba	66,3%	68,2%	67,3%	59.074	58.296	0,99	66,364%
La Pampa	61,0%	60,5%	60,8%	5.614	5.441	0,97	58,878%
Río Negro	57,4%	57,5%	57,5%	12.319	12.155	0,99	56,685%
Tierra del Fuego	63,7%	55,2%	59,5%	2.985	2.783	0,93	55,427%
Chubut	58,3%	53,4%	55,8%	10.163	10.002	0,98	54,965%
Mendoza	57,1%	59,3%	58,2%	32.907	30.470	0,93	53,890%
Jujuy	54,0%	54,3%	54,2%	13.257	13.168	0,99	53,786%
Neuquén	54,0%	52,0%	53,0%	11.754	11.664	0,99	52,594%
Santa Fé	55,4%	54,8%	55,1%	52.341	49.202	0,94	51,796%
Buenos Aires	56,8%	54,0%	55,4%	285.389	264.316	0,93	51,309%
Tucumán	52,5%	53,8%	53,2%	28.562	27.346	0,96	50,887%
Santa Cruz	57,1%	48,7%	52,9%	6.288	5.928	0,94	49,871%
Salta	51,4%	53,8%	52,6%	26.358	23.827	0,90	47,549%
San Luis	52,4%	50,7%	51,6%	8.476	7.552	0,89	45,930%
Formosa	53,6%	55,0%	54,3%	11.663	9.773	0,84	45,501%
Entre Ríos	49,5%	46,7%	48,1%	22.604	19.956	0,88	42,465%

Figura 86. Aprender 2021-Matemática y Lengua y la relación de matrículas de edades teóricas de 6to. RA2021 y de 1ro. RA2016. Distribución de estudiantes por niveles de desempeños Avanzado y Satisfactorio. Ranking según multiplicación relación de edades teóricas del RA y desempeño Aprender. Fuente: Aprender 2021, Relevamientos anuales y propia.

Como se observa en la Figura 86, en el promedio de los desempeños Avanzado y Satisfactorio de Lengua y Matemática de Aprender 2021, Formosa queda en el décimo lugar con el 54.3%, lo que constituye un buen resultado. Al considerar las diferencias de los niveles socioeconómicos, y del NSE bajo, este resultado se convierte en muy bueno. Sin embargo, al multiplicar por la relación de matrículas en edades teóricas, Formosa queda en el lugar décimo sexto con el 45,50% de los alumnos en edad teórica con niveles avanzado y satisfactorio de aprendizaje, y no con el 36% como señala un artículo periodístico de marzo de 2023.

5.2.2 Aprender secundaria 2022

5.2.2.1 Aprender 2022. Matemática. Distribución de estudiantes por niveles de desempeño y suma de los niveles Avanzado y Satisfactorio

Solo se van a visualizar los resultados generales de Aprender Matemática, como se observa en la Figura 87.

SECUNDARIA-Matemática-Aprender 2022- Distribución de estudiantes por nivel de desempeño y suma de los niveles por Jurisdicción

Distribución de estudiantes por nivel de desempeño: Avanzado y Satisfactorio-Ranking alfabético

Provincia	Apr22_M_Av	Apr22_M_Sa	Apr22_M_Básí...	Apr22_M_deb...	Apr22_M_AvSa
Buenos Aires	0.1%	17.0%	30.2%	52.7%	17.1%
CABA	0.5%	36.7%	32.4%	30.4%	37.2%
Catamarca	0.0%	8.3%	26.2%	65.5%	8.3%
Chaco	0.0%	7.3%	21.3%	71.4%	7.3%
Chubut	0.2%	17.9%	31.1%	50.8%	18.1%
Corrientes	0.0%	10.2%	26.8%	63.0%	10.2%
Córdoba	0.1%	25.8%	33.9%	40.2%	25.9%
Entre Rios	0.1%	16.9%	33.1%	49.9%	17.0%
Formosa	0.0%	9.2%	24.9%	65.9%	9.2%
Jujuy	0.0%	13.9%	30.7%	55.4%	13.9%
La Pampa	0.0%	20.4%	33.4%	46.2%	20.4%
La Rioja	0.0%	8.9%	26.0%	65.1%	8.9%
Mendoza	0.1%	16.8%	32.2%	50.9%	16.9%
Misiones	0.0%	9.5%	26.9%	63.6%	9.5%
Neuquén	0.0%	22.3%	35.6%	42.1%	22.3%
Rio Negro	0.2%	19.3%	30.9%	49.6%	19.5%

Figura 87. Aprender 2022 Secundaria-Matemática. Distribución de estudiantes por nivel de desempeño y suma de los niveles Avanzado y Satisfactorio. Fuente: Aprender 2022 y propia.

5.2.2.1 Aprender 2022. Lengua. Distribución de estudiantes por niveles de desempeño y suma de los niveles Avanzado y Satisfactorio

Los resultados generales de Aprender Lengua, por niveles de desempeño, se pueden observar en la Figura 88.

SECUNDARIA-Lengua-Aprender 2022- Distribución de estudiantes por nivel de desempeño y suma de los niveles por Jurisdicción

Distribución de estudiantes por nivel de desempeño: Avanzado y Satisfactorio-Ranking alfabético

Provincia	Apr22_L_Av	Apr22_L_Sa	Apr22_L_Básico	Apr22_L_deba...	Apr22_L_AvSa
Buenos Aires	15.4%	41.7%	19.2%	23.7%	57.1%
CABA	26.1%	44.7%	13.9%	15.3%	70.8%
Catamarca	10.0%	39.1%	22.6%	28.3%	49.1%
Chaco	8.4%	33.5%	22.5%	35.6%	41.9%
Chubut	15.1%	46.0%	19.9%	19.0%	61.1%
Corrientes	10.6%	38.9%	22.2%	28.3%	49.5%
Córdoba	18.8%	46.1%	18.0%	17.1%	64.9%
Entre Rios	14.3%	43.6%	20.4%	21.7%	57.9%
Formosa	8.4%	36.1%	22.9%	32.6%	44.5%
Jujuy	10.9%	42.1%	21.7%	25.3%	53.0%
La Pampa	17.2%	46.3%	19.8%	16.7%	63.5%
La Rioja	10.8%	38.3%	22.6%	28.3%	49.1%
Mendoza	13.3%	43.2%	20.5%	23.0%	56.5%
Misiones	10.6%	40.0%	22.2%	27.2%	50.6%
Neuquén	16.3%	46.8%	20.1%	16.8%	63.1%
Rio Negro	14.8%	44.7%	20.1%	20.4%	59.5%

Figura 88. Aprender 2022 Secundaria-Lengua. Distribución de estudiantes por nivel de desempeño y suma de los niveles Avanzado y Satisfactorio. Fuente: Aprender 2022 y propia.

5.2.3 Aprender primaria 2023-censal

Solo se van a mostrar los resultados generales de Matemática, con dos características en las que se destaca la provincia de Formosa.

5.2.3.1 Aprender 2023. Matemática. Distribución de estudiantes por nivel de desempeño y suma de los niveles Avanzado y Satisfactorio

Los resultados generales de Aprender Matemática, por niveles de desempeño, se pueden observar en Figura 89. Se observa que Formosa está segunda.

PRIMARIA-Matemática-Aprender 2023- Distribución de estudiantes por nivel de desempeño y suma de los niveles Avanzado y Satisfactorio por Jurisdicción

Distribución de estudiantes por nivel de desempeño: Avanzado y Satisfactorio-Ranking de jurisdicciones

Provincia	Apr23_Mat_Avanzado	Apr2023_Mat_Satisfactorio	Apr2023_Mat_Básico	Apr2023_Mat_debajo_del_básico	Apr2023_Mat_AvSa
CABA	23.1%	47.3%	17.4%	12.20%	70.4%
Formosa	16.6%	40.1%	20.6%	22.70%	56.7%
Tierra del Fuego	8.6%	44.6%	24.8%	22.00%	53.2%
La Pampa	10.7%	41.3%	23.2%	24.80%	52.0%
Neuquén	9.8%	42.1%	24.6%	23.50%	51.9%
Rio Negro	10.7%	41.0%	24.3%	24.00%	51.7%
Chubut	9.9%	41.7%	25.7%	22.70%	51.6%
Buenos Aires	10.9%	40.7%	24.3%	24.10%	51.6%
Jujuy	9.6%	41.6%	25.9%	22.90%	51.2%
Mendoza	9.8%	40.4%	25.1%	24.70%	50.2%
Santa Fé	11.0%	39.0%	24.3%	25.70%	50.0%
Salta	10.3%	39.2%	25.4%	25.10%	49.5%
San Luis	9.3%	39.5%	26.1%	25.10%	48.8%
Entre Rios	9.8%	38.3%	25.0%	26.90%	48.1%
Corrientes	9.6%	37.7%	26.4%	26.30%	47.3%

Figura 89. Aprender 2023 Primaria-Matemática. Distribución de estudiantes por nivel de desempeño y suma de los niveles Avanzado y Satisfactorio. Ranking por nivel de desempeño. Fuente: Aprender 2023 y propia.

5.2.3.2 Aprender 2023. Matemática. Menor diferencia de la suma de desempeños avanzado y satisfactorio entre los NSE bajo y NSE alto.

En la Figura 90, se observa que Formosa se destaca por tener una menor diferencia entre los niveles socioeconómicos bajo y alto[1].

[1] Formosa.gob.ar, “Apertura de las Sesiones Ordinarias de la Honorable Cámara de Diputados -1 Marzo 2024”, en https://www.formosa.gob.ar/accionesdegobierno; [Ultima Consulta:15/07/2024]

Primaria_Aprender23-Matemática_Niveles Socioeconómicos-Diferencias entre porcentajes de estudiantes con desempeños Av. y Sat. de NSE bajo y alto

Distribución de estudiantes por niveles de desempeños Avanzado y Satisfactorio según NSE-Ranking diferencia entre NSE bajo menos NSE alto de 2023 (Córdoba:NO INFORMADO

Provincia	Ap23_M_NSBajo_AvSa	Ap23_M_NSMedio_AVSa	Ap23_M_NSAlto_AvSa	Ap23P_M_Dif_NS_Ba_Al
Formosa	62.1%	59.2%	66.8%	-4.7%
Santiago del Estero	48.5%	49.1%	63.8%	-15.3%
Misiones	45.5%	50.2%	64.5%	-19.0%
Jujuy	45.6%	54.3%	65.8%	-20.2%
Tierra del Fuego	43.7%	54.0%	64.3%	-20.6%
Catamarca	33.2%	43.9%	56.2%	-23.0%
Chaco	36.6%	42.8%	60.7%	-24.1%
Salta	42.6%	54.2%	68.5%	-25.9%
La Rioja	32.4%	43.4%	58.7%	-26.3%
San Luis	37.0%	50.5%	63.7%	-26.7%
Corrientes	42.5%	51.9%	69.5%	-27.0%
Entre Rios	36.7%	50.7%	66.6%	-29.9%
Chubut	37.2%	53.0%	67.6%	-30.4%
Santa Cruz	30.1%	45.1%	61.0%	-30.9%
Rio Negro	39.3%	53.1%	70.5%	-31.2%

Figura 90. Aprender 2023 Primaria-Matemática. Distribución de estudiantes por suma de los niveles de desempeño Avanzado y Satisfactorio de los Niveles socioeconómicos: bajo, medio y alto. Ranking de las jurisdicciones por menor diferencia entre los niveles socioeconómicos bajo y alto. Fuente: Aprender 2023 y propia.

5.2.3.3 Aprender 2023. Matemática. Mayor crecimiento en la diferencia de los niveles de desempeño avanzado y satisfactorio entre las pruebas ONE 2013 y Aprender 2023.

En la Figura 91, se ve que Formosa se destaca por ser la jurisdicción que más creció en la diferencia de los niveles de desempeño Alto y Satisfactorio a nivel país, entre las pruebas ONE 2013 y Aprender 2023.

PRIMARIA-Matemática-ONE13-Aprender 16-18-21-22-23-Diferencias entre porcentajes de estudiantes con desempeño Avanzado y Satisfactorio de Aprender 2023 y ONE2013

Distribución de estudiantes por nivel de desempeño: Avanzado y Satisfactorio-Ranking de jurisdicciones que mejoraron desempeños en Matemática entre ambos períodos

Provincia	ONE13_AvSa	Apr16_AvSa	Apr18_AvSa	Apr21_AvSa	Apr22_Av...	Apr23_Av...	Dif_Apr_Mat_21_18_Av...	Dif_A23_O13
Formosa	44.0%	58.7%	54.4%	55.0%	66.8%	56.7%	0.60%	12.7%
Corrientes	35.2%	53.6%	50.8%	45.4%	48.2%	47.3%	-5.40%	12.1%
Santiago del Estero	34.5%	49.9%	49.2%	50.4%	53.0%	46.2%	1.20%	11.7%
Jujuy	46.8%	52.9%	53.7%	54.3%	56.9%	51.2%	0.60%	4.4%
Tierra del Fuego	49.6%	56.9%	56.9%	55.2%	58.8%	53.2%	-1.70%	3.6%
Misiones	43.7%	46.7%	53.8%	45.7%	45.4%	47.0%	-8.10%	3.3%
La Rioja	38.1%	46.7%	49.4%	33.3%	47.7%	39.9%	-16.10%	1.8%
Entre Ríos	46.5%	55.3%	54.3%	46.7%	51.2%	48.1%	-7.60%	1.6%
Catamarca	38.1%	41.2%	42.5%	39.9%	46.0%	39.6%	-2.60%	1.5%
Córdoba	0.0%	67.6%	68.2%	68.2%	0.0%	0.0%	0.00%	0.0%
Chubut	52.4%	55.0%	58.3%	53.4%	58.2%	51.6%	-4.90%	-0.8%
San Luis	49.6%	60.4%	58.4%	50.7%	53.8%	48.8%	-7.70%	-0.8%
Santa Cruz	46.8%	49.9%	49.4%	48.3%	52.9%	45.3%	-1.10%	-1.5%

Figura 91. Aprender 2023 Primaria-Matemática. Distribución de estudiantes por suma de los niveles de desempeño Avanzado y Satisfactorio de los operativos ONE2013, Aprender 2016, 2018, 2021, 2022 y 2023. Ranking de las jurisdicciones que más crecieron entre las pruebas ONE2013 y Aprender 2023.Fuente: ONE13-Apreder16,18,21,22 y 23 y propia.

5.3 Embeber el gráfico de Atlas en cualquier página web

MongoDB Atlas brinda la posibilidad de embeber el gráfico, en cualquier página web, como se observa en las opciones de la Figura 92.

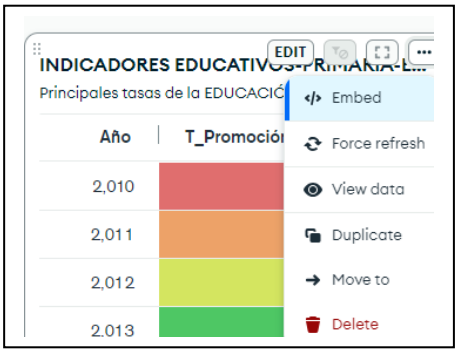


Figura 92. Embeber un gráfico de MongoDB Atlas en una página web. Fuente propia.

Capítulo 6

Conclusiones y futuras investigaciones

6.1 Conclusiones

En este trabajo se estudiaron y caracterizaron los elementos, parámetros y agrupaciones que conforman el Sistema Educativo Argentino (en adelante, SEA), logrando como resultado, extraer y organizar la información correspondiente a los datos abiertos con el objetivo de permitir una mejor visualización de éstos.

Como parte del proceso de organización de los datos, se crearon variables categóricas para cada uno de los niveles y modalidades de las escuelas; se eliminó información innecesaria; se homogeneizaron los nombres de las columnas. De esta manera, el mantenimiento de la información que se almacene se realizará con menor dificultad.

La extracción y organización de los datos del SEA permitió realizar un análisis exhaustivo, utilizando la herramienta Atlas para generar visualizaciones gráficas, facilitando la interpretación de los datos y apoyando a la toma de decisiones.

El almacenamiento en una base de datos NoSQL, de los datos abiertos del SEA, garantizó la disponibilidad y actualización de éstos, reduciendo la información imprecisa e inconsistente. Además, la aplicación desarrollada, permitió comparar los resultados del procesamiento de los anuarios almacenados y realizar un trabajo de auditoría de los datos. Se pudo realizar también un análisis y discusión de los indicadores de la educación secundaria, al incorporar categorías de promovidos por examen, y otros promovidos, mejorándolo ya que, por ejemplo, en la pandemia COVID-19 se implementó en el país la promoción acompañada, y los operadores del RA, colocaron los datos de “promovidos”, en las columnas “promovidos regulares” y “otros promovidos”.

La utilización de una base de datos NoSQL y particularmente MongoDB, proporcionó una serie de características que fueron apropiadas para la implementación de las estadísticas educativas que se realizaron:

- La característica no estructurada del modelo de datos NoSQL resultó adecuada para almacenar y gestionar los datos educativos, ya que las publicaciones que realiza el Ministerio de Capital Humano contienen mucha redundancia de datos. Con la estructura flexible de los documentos de MongoDB sólo se ocuparon los campos de datos estrictamente necesarios, eliminando duplicados. Por otra parte, un mismo campo se pudo representar con distintos tipos de datos, facilitando su incorporación de la base de datos, dada la heterogeneidad de las fuentes.

- En este trabajo se gestionó información de distintos tipos y calidades, que presentó las siguientes particularidades:
    - La información de los Relevamientos Anuales se carga en miles de estaciones de trabajo en el país, con distintos sistemas operativos, aparentemente con el mismo formato en las planillas Excel, pero al intercambiar la información, de Excel a MongoDB empleando diferentes herramientas (Mongo Shell, Compass, Atlas) es común que aparezcan datos erróneos, o en doble precisión, que no serían adecuados para el estudio que se desarrolla en esta tesis. Se aprovechó la facilidad de importación de MongoDB para el procesamiento de la información y la detección de errores.
    - Aún con estas facilidades debieron realizarse algunas transformaciones posteriores de los datos como, por ejemplo, algunos atributos con un valor "X" para indicar que una institución presenta una oferta determinada, en lugar de informar tal condición con un 1, para facilitar la operación.
    - Ante la imposibilidad de trabajar con la clave de cada institución, ya sea por el secreto estadístico o porque dicha clave fue cambiando de 10 a 14 dígitos a lo largo de la docena de años, las capacidades de agregación de MongoDB, permitieron resumir la información por medio de una nueva clave de agrupamiento conformada por el año, el sector y el ámbito, y realizar la construcción de documentos embebidos en distintos niveles y el empleo de vectores para cada una de las variables. Por ejemplo, la variable cantidad de alumnos promovidos en un año, se representa por medio de un conjunto de documentos de promovidos primaria, promovidos primaria_mujeres, promovidos secundaria y promovidos secundaria_mujeres, y dentro de cada uno de estos documentos, se incluyeron otros documentos de promovidos a fin de año, promovidos con examen, promovidos regulares y otros promovidos. A su vez, estos documentos se almacenaron en vectores para el manejo de los valores correspondientes para cada año de estudio. Esta construcción permitió el reemplazo de más de 100 campos sólo para esta variable, de la información publicada por el Ministerio de Capital Humano, y redundó en la facilidad para hacer los distintos agrupamientos, agregaciones, resúmenes y cálculos en los informes de cuadros y gráficos estadísticos. Se debe destacar que estos formatos no se pueden emplear en una base de datos SQL, por la restricción de atomicidad de sus campos.

    - La construcción mencionada más arriba, permitió el cálculo de las tasas de promoción efectiva, de repitencia y de abandono interanual, tanto a nivel provincial como para cada grado o año de estudio, con una cantidad cercana a 200 líneas de código para hacer la consulta principal, aunque la misma demandó más de 20 días de trabajo para su codificación. Es importante destacar que, para otro tipo de agrupamiento como, por ejemplo, el cálculo de las mismas tasas por otro criterio, por ejemplo, por ámbito se consigue simplemente arrastrando el campo de ámbito a la parte de grupos, en la edición del gráfico. Lo mismo sucede si necesitamos las tasas por sector de gestión.
    - Se calcularon las tasas de eficiencia interna 2022: la tasa de promoción efectiva, tasa de repitencia y de abandono interanual 2022/2023, que a la fecha de última revisión de este trabajo (agosto de 2024), aún no publica el Ministerio de Capital Humano.
    - Se realizaron controles de totales en todo momento: a la hora de importar y mover los datos, y también, cuando se calcularon los distintos totales y tasas con los Anuarios Estadísticos publicados por el Ministerio de Capital Humano. La gran mayoría de los totales y tasas coincidieron, aunque no sucedió lo mismo, con la suma de las cantidades parciales de unidades educativas por departamento.
- Con utilización de objetos JSON autocontenidos se pudo realizar:
    - el manejo sencillo de la información en todo momento
    - el completamiento de la información no informada en los relevamientos anuales
    - el registro de información de los operativos ONE y Aprender
    - una fácil actualización de documentos.
- Con el empleo de MongoDB, se facilitó la realización de cuadros, gráficos y mapas modernos en la nube, para una mejor visualización de la información. El agregado, modificación y eliminación de grupos de datos estadísticos se pudo realizar sin afectar los programas de la aplicación. La utilización de colores en formatos condicionales fue de gran utilidad para el análisis ya que permitió visualizar las características educativas relevantes de la provincia de Formosa como, por ejemplo, a) mayor cantidad de unidades educativas y unidades de servicios educativos en el ámbito rural, a pesar de tener menor matrícula; b) menor diferencia entre los niveles

socioeconómicos de matemática primaria de APRENDER 2023; c) mayor crecimiento entre las pruebas ONE 2013 y APRENDER 2023 en matemática de primaria.

Con la realización del presente trabajo, se detectó que la información publicada en las bases de datos por parte del Ministerio de Capital Humano, presenta varios problemas:

- Falta de la apertura de egresados por orientación en la base de datos de trayectoria, ya que la orientación se obtuvo de los anuarios.
- Al estar la clave enmascarada y ser todos los registros unidades de servicio educativo, falta un campo de unidad educativa en la base de datos, que asigne la calidad de unidad educativa, para poder contar con precisión la cantidad precisa de unidades educativas.
- Falta de registros o documentos de modalidad adultos, especial y otras modalidades.
- Falta de la dimensión ámbito en gran parte de los análisis.
- Diferencias en el cálculo de total de promovidos del año anterior, sobre todo en los años de la pandemia COVID-2019.
- Falta de la actualización de la publicación de tasas de eficiencia interna 2022, a la fecha de la última revisión del presente trabajo (Agosto de 2024).

6.2 Futuras investigaciones

Existen otros aspectos que necesitan ser investigados y trabajados, los cuales surgen a partir de lo realizado en esta tesis:

- El factor de escalamiento que relaciona las distintas pruebas y calidad de cada una de las pruebas Aprender, y que pueda ser un proceso público y abierto, como el cálculo de cualquiera de las variables de este trabajo. En el contexto de relacionar dos pruebas de aprendizaje, el factor de escalamiento se refiere a un valor numérico que se utiliza para ajustar la puntuación obtenida en una prueba para que sea comparable a la puntuación obtenida en otra prueba. Dicho factor es necesario porque las dos pruebas pueden tener diferentes escalas, es decir, diferentes rangos de puntuaciones posibles.
- Estudiar una relación coherente entre distintas herramientas de medición de calidad educativa para un ciclo de 12 años, ya que al multiplicar las distintas tasas de promoción efectiva de 0.9 * 0.9 …, 12 veces, por el resultado de una prueba

Aprender, da un número muy pequeño, cercano o muy inferior al 25%, que constituye un número irracional para la calidad.

- Desarrollar una plataforma web regional o nacional, de calidad, que permita automatizar las tareas de extracción de los datos alojados en la web, su organización y unificación de datos estadísticos y geoespaciales en un formato común y estandarizado, se puedan realizar todo tipo de análisis con mayor apertura, cobertura, velocidad y precisión. Como sostiene Guillermina Tiramonti en un artículo de una revista publicada por la Universidad de Belgrano correspondiente a Julio-Agosto de este año, al tratar el tema de la transformación de la escuela media, que uno de los temas importantes a tratar para el mejoramiento de la calidad de toda la educación es: *"2 – La generación de un sistema integral de información digital educativa (SINIDE) que hace más de 10 años que estamos en camino de formarlo y todavía no lo hemos logrado"* [40].

Bibliografía

[1] Piovani J., Salvia A. , «Trayectorias y capitales socioeducativos», de *La Argentina en el siglo XXI: cómo somos, vivimos y convivimos en una sociedad desigual*, Ciudad Autónoma de Buenos Aires, Siglo XXI Editores, 2018, pp. 221-236.

[2] UNESCO-IESALC, «Los futuros de la Educación Superior en América Latina y el Caribe», *Educación Superior y Sociedad,* vol. 35, nº 1, p. 40, 2023.

[3] Argentina.gob.ar, «Ministerio de Capital Humano - Educacíon - Información y Evaluación Educativa», [En línea]. Available: https://www.argentina.gob.ar/educacion/evaluacion-e-informacion-educativa. [Último acceso: 26 02 2024].

[4] Argentina.gob.ar, «Capital Humano Educación Evaluación e Información Educativa-Cuadernillos de Relevamiento Anual», [En línea]. Available: https://www.argentina.gob.ar/educacion/evaluacion-informacion-educativa/cuadernillos. [Último acceso: 15 08 2024].

[5] Argentina.gob.ar, «Capital Humano Educación Evaluación e Información Educativa-Base de Datos - Información Estadística», [En línea]. Available: https://www.argentina.gob.ar/educacion/evaluacion-informacion-educativa/bdd. [Último acceso: 27 02 2024].

[6] Munshi A. & Alhindi A., «Big Data Plattfomr for Educational Analytics», *IEEE Acces,* vol. 9, nº 2021, p. 52884, 2021.

[7] Li-Minn Ang K., Lu Ge F. & Seng K., «Big Educational Data & Analytics: Survey, Architecture and Challenges», *IEEE Access,* vol. 8, nº 2020, p. 166399, 2020.

[8] Jivet J., Wong J., Scheffel M., Torre M., «Quantum of Choice: How learners' feedback monitoring decisions, goals and self-regulated learning skills are related», *11th International Learning Analytics an Knowledge Conference,* pp. 416-427, 2021.

[9] Otero, A.E. & Córica, A., «La performance del sistema educativo argentino en las últimas décadas», *Análisis,* vol. 49, nº 90, pp. 17-37. https://doi.org/10.15332/s0120-8454.2017.0090.01, 2017.

[10] Argentina.gob.ar, «Ley 20206 De Educación Nacional», [En línea]. Available: https://www.argentina.gob.ar/sites/default/files/ley-de-educ-nac-58ac89392ea4c.pdf. [Último acceso: 27 02 2020].

[11] M. Herrera Gómez, «¿La localización de la escuela importa?. Condicionantes espacio-contextuales de la tasa de repitencia en un panel de datos georreferenciados», *Munich Personal RePEc Archive ,* Agosto 2016.

[12] Elmasri R., Navathe S., «Fundamentals of Database Systems - Seventh Edition», Pearson, 2016.

[13] Ramzan S., Bajwa I., Ramzan B., Anwar W., «Intelligent Data Engineering for Migration to NoSQL Based Secure Environments», *IEEE Access - Special section on Advanced Software and Data Engineering for Secure Societies,* vol. 7, pp. 69042-69056, 2019.

[14] Schram, A.; Anderson, K.M., «MySQL to NoSQL: data modeling challenges in supporting scalability», de *3rd Annual Conference on Systems, Programming, and Applications: Software for Humanity*, Tucson, Arizona, USA, 2012.

[15] J. Pokorny, «NoSQL databases: a step to database scalability in web environment», *International Journal of Web Information Systems,* vol. 9, nº 1, pp. 69-82, 2013.

[16] V. Dhar, «Data Science and Prediction», *Communications of the ACM,* vol. 56, nº 12, pp. 64-73, 2013.

[17] Amazon.com, «¿Qué es ciencia de datos?», [En línea]. Available: https://aws.amazon.com/es/what-is/data-science/. [Último acceso: 14 01 2024].

[18] MongoDB , «JSON y BSON,» [En línea]. Available: https://www-mongodb-com.translate.goog/json-and-bson?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=rq#:~:text=BSON's%20binary%20structure%20encodes%20type,been%20missing%20some%20valuable%20support. [Último acceso: 15 12 2021].

[19] Batini C., Cappiello C., Francalanci C., y Maurino A., «Methodologies for data quality assessment and improvement», *ACM Comput. Surv.,doi: 10.1145/1541880.1541883,* vol. 41, nº 3, pp. 1-52, 2009.

[20] Ehrlinger, L. & Wöß, W., «A Survey of Data Quality Measurement and Monitoring Tools», *Frontiers in Big Data, https://doi.org/10.3389/fdata.2022.850611,* vol. 2022, nº 5, p. 850611, 2022.

[21] Ruchinsky V., Vidoni M., Golobisky M., «Desafíos para datos educativos abiertos en Argentina», de *CONAIISI 2018-6to.Congreso Nacional de Ingeniería Informática-Sistema de Información* , Mar del Plata, 2018.

[22] V. Ruchinsky, «Limpieza de datos educativos abiertos en bases de datos NoSQL MongoDb», de *XXIII Jornadas de Ciencia y Tecnología - Universidad Nacional de Formosa*, Formosa, 2020.

[23] Argentina.gob.ar, «Anuario Estadístico Educativo 2018», [En línea]. Available: https://www.argentina.gob.ar/sites/default/files/anuario-estadistico-datos-2018-web.pdf. [Último acceso: 18 05 2022].

[24] INDEC, «Censo Nacional de Población, Hogares y Viviendas 2010: Cuadro P1-P. provincia de Formosa. Población total y variación intercensal absoluta y relativa por departamento. Años 2001-2010», [En línea]. Available: https://www.indec.gob.ar/indec/web/Nivel4-Censoprovincia-3-999-34-00. [Último acceso: 30 05 2022].

[25] UNaF, «UNaf-Universidad Nacional de Formosa-Historia», [En línea]. Available: http://www.unf.edu.ar/historia/. [Último acceso: 30 005 2022].

[26] I.S.F.D.C.y T.F.A.C., «I.S.F.D.C.y.T. FELIZ ATILIO CABRERA – Autoridades», [En línea]. Available: https://isfdcytcabrera-for.infd.edu.ar/sitio/autoridades/. [Último acceso: 30 05 2022].

[27] MongoDB Atlas, «MongoDB. La plataforma de datos para aplicaciones multicloud», [En línea]. Available: https://www.mongodb.com/es/atlas. [Último acceso: 15 12 2023].

[28] Argentina.gob.ar, «Anuarios Estadísticos», Ministerio de Capital Humano. Educación. Información y Evaluación Educativa, [En línea]. Available: https://www.argentina.gob.ar/educacion/evaluacion-informacion-educativa/anuarios. [Último acceso: 15 08 2024].

[29] MongoDB, «Documentación de MongoDB-Manual de Referencia de MongoDB», [En línea]. Available: https://www.mongodb.com/docs/manual/reference. [Último acceso: 10 12 2022].

[30] MongoDB, «Controlador MongoDB NodeJS», [En línea]. Available: https://mongodb.github.io/node-mongodb-native/. [Último acceso: 10 12 2022].

[31] MongoDB, «MongoDB Developer - JavaScript», [En línea]. Available: https://www.mongodb.com/developer/languages/javascript/. [Último acceso: 10 12 2022].

[32] Argentina.gob.ar, «Resultados Aprender 2021», [En línea]. Available: https://www.argentina.gob.ar/educacion/evaluacion-informacion-educativa/aprender/aprender-2021/informes-jurisdiccionales. [Último acceso: 12 07 2022].

[33] formosa.gov.ar, «Infraestructura de datos espaciales de Formosa», [En línea]. Available: idef.formosa.gov.ar. [Último acceso: 18 12 2022].

[34] V. Agafonkin, «Leaflet», [En línea]. Available: https://leafletjs.com/ . [Último acceso: 19 09 2023].

[35] w3schools, «Información sobre herramientas CSS», [En línea]. Available: https://www.w3schools.com/css/css_tooltip.asp. [Último acceso: 10 12 2021].

[36] Argentina.gob.ar, «Documentos metodológicos. Estadística educativa. Indicadores educativos», Capital Humano. Educación. Evaluación e Información educativa, [En línea]. Available: https://www.argentina.gob.ar/educacion/evaluacion-informacion-educativa/metodologicos. [Último acceso: 15 03 2024].

[37] Argentina.gob.ar, «Indicadores Educativos», Ministerio de Capital Humano. Educación. Información y Evaluación Educativa., [En línea]. Available:

https://www.argentina.gob.ar/educacion/evaluacion-e-informacion-educativa/indicadores. [Último acceso: 31 08 2024].

[38] Infobae, «Pruebas Aprender: cuáles fueron las provincias que más sufrieron el cierre de escuelas y las sorpresas», [En línea]. Available: https://www.infobae.com/educacion/2022/07/06/pruebas-aprender-cuales-fueron-las-provincias-que-mas-sufrieron-el-cierre-de-escuelas-y-las-sorpresa. [Último acceso: 15 12 2022].

[39] Infobae, «Menos de la mitad de los alumnos llega a sexto grado a tiempo y con los aprendizajes necesarios», [En línea]. Available: https://www.infobae.com/educacion/2023/03/09/menos-de-la-mitad-de-los-alumnos-llega-a-sexto-grado-a-tiempo-y-con-los-aprendizajes-necesarios/. [Último acceso: 115 12 2023].

[40] T. Guillermina, «La transsformación de la escuela media», *CEAUB-Centro de Estudios de la Educación Argentina de la Universidad de Belgrano,* vol. Año 13, nº 128, 2024.