



ASIGNATURA:

Análisis de datos

PROFESOR:

Ing. Juan Pablo Zaldumbide Proaño

PROYECTO FINAL

Grupo 3

INTEGRANTES:

- Carapaz Mendoza Miguel Angel
- Panchi Melo Jose Rafael
- Flores Quimbia Dilan Alexander

FECHA DE ENTREGA:

06/03/2023



INDICE DE CONTENIDOS

1.	Definición del caso de estudio	5
2.	Objetivos Generales y específicos	5
3.	Descripción del equipo de trabajo y actividades realizadas por cada uno	5
4.	Desarrollo.....	5
5.	Recurso y herramientas utilizadas.....	11
6.	Arquitectura de la solución	11
7.	Conclusiones	11
8.	Recomendaciones.....	12
9.	Desafíos y problemas encontrados.....	12
10.	Bibliografía	12



INDICE DE ILUSTRACIONES

Ilustración 1 Base de datos CouchDB	5
Ilustración 2 Base de datos MongoDB	6
Ilustración 3 Base de datos MySQL Peru	6
Ilustración 4 Colombia.....	7
Ilustración 5 Colombia.....	7
Ilustración 6 Ecuador	7
Ilustración 7 Ecuador.....	8
Ilustración 8 Ecuador	8
Ilustración 9 Ecuador	8
Ilustración 10 Nivel mundial	8
Ilustración 11 Ecuador.....	9
Ilustración 12 Ecuador	9
Ilustración 13 Ecuador.....	9
Ilustración 14 Ecuador	9
Ilustración 15 Ecuador	10
Ilustración 16 Ecuador	10
Ilustración 17 Ecuador.....	10
Ilustración 18 Perú	10
Ilustración 19 Perú	11
Ilustración 20 Arquitectura de la solución.....	11



INDICE DE TABLAS

Tabla 1 Recursos utilizados en el proyecto	11
--	----

1. Definición del caso de estudio

Identificar la población general de estudiantes que pudieron acceder a la educación antes y durante la pandemia COVID-19 en los países de Ecuador, Colombia, Perú y a nivel mundial.

2. Objetivos Generales y específicos

- Identificar cual fue el país más afectado a nivel estudiantil por la pandemia, respecto a la cantidad de estudiantes que dejó de estudiar durante la pandemia COVID-19.
- Identificar cual fue el factor que más influyó en cada país para que los estudiantes no pudieran acceder a la educación.

3. Descripción del equipo de trabajo y actividades realizadas por cada uno

A continuación, se describe el trabajo realizado por cada uno de los integrantes:

Carapaz Migue: Recolección de datos de diferentes fuentes, extracción de datos, limpieza de datos, visualización de datos, análisis de datos, dashboard, realización del informe.

Panchi Jose: Recolección de datos de diferentes fuentes, extracción de datos, limpieza de datos, visualización de datos, análisis de datos, dashboard, realización del informe.

Flores Dilan: Recolección de datos de diferentes fuentes, extracción de datos, limpieza de datos, visualización de datos, análisis de datos, dashboard, realización del informe.

4. Desarrollo

Para poder realizar el análisis de datos, primero se realizó la búsqueda de información de diferentes fuentes de datos con la finalidad de recabar una gran cantidad de información que facilite la obtención de resultados en base a nuestros objetivos planteados anteriormente.

Luego de recabar la información de diferentes fuentes se procedió a repartir la data en diferentes bases de datos NoSQL y bases de datos

relacionales para recopilar total o parcialmente los datos mediante el lenguaje de programación Python, mismo que se usó para la limpieza de datos como se puede ver a continuación en las siguientes ilustraciones:



Name	Size	# of Docs	Partitioned
canasta	1.3 KB	2	No
canasta2	0 bytes	0	No
colombia	3.3 MB	3365	No
intento	0 bytes	0	No
primaria	265.0 KB	532	No
shopping_cart	2.6 KB	11	No

Ilustración 1 Base de datos CouchDB

Como se puede apreciar en la ilustración1, se cuenta con una base de datos NoSQL llamada “Colombia” y “primaria” donde se encuentra almacenada la información extraída de las fuentes: “GOV.CO” y “Banco Mundial de Datos”. Por otra parte, luego de haber extraído los datos y almacenado en una base de datos se procedió a obtener los datos mediante Python y a realizar la limpieza de datos de esta, eliminando datos nulos y aquellos campos que no sean relevantes para nuestro análisis. El código con el que se realizó la limpieza de datos se encuentra en el archivo ipynb que se encuentra en la tabla de Recurso y herramientas utilizadas.

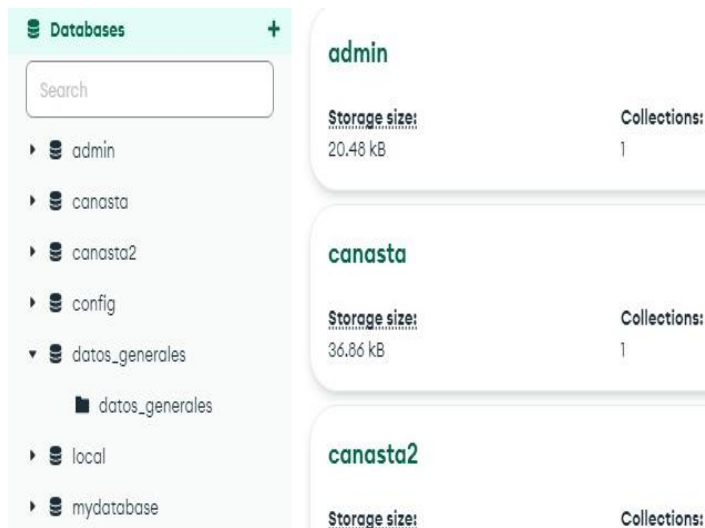


Ilustración 2 Base de datos MongoDB

De misma forma en la ilustración2, se puede apreciar una base de datos NoSQL donde se encuentra almacenada la base de datos “datos_generales” que contiene la información extraída de la fuente: “Gobierno del Ecuador”. Misma que al igual que con CouchDB se realizó la extracción de los datos almacenados en la base de datos y que mediante Python se realizó la limpieza de datos eliminando datos nulos y aquellos campos que no sean relevantes para el caso de estudio. El código con el que se realizó la limpieza de datos se encuentra en el archivo ipynb

que se encuentra en la tabla de Recurso y herramientas utilizadas.

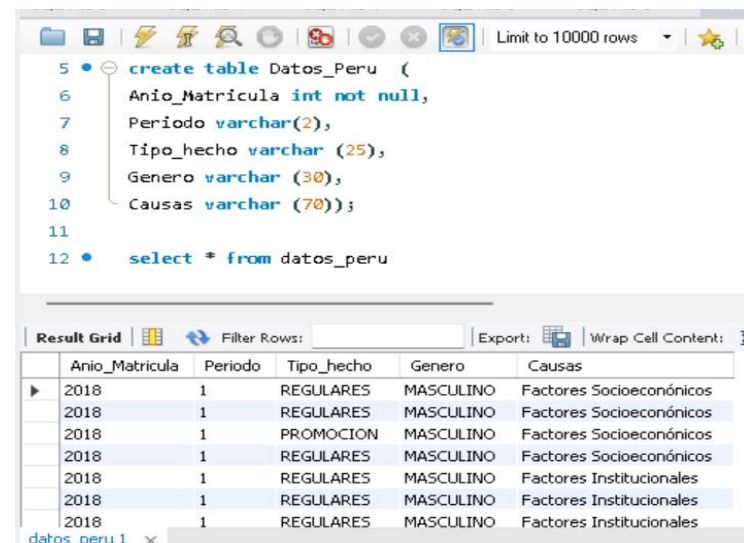


Ilustración 3 Base de datos MySQL Perú

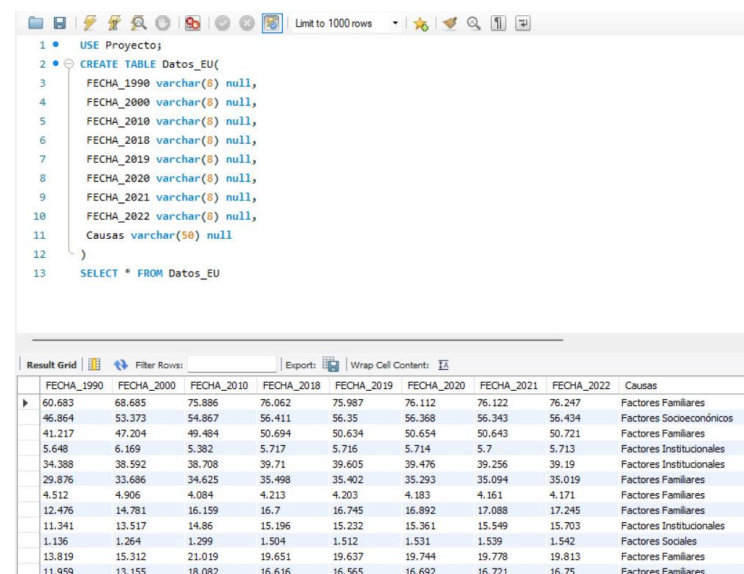


Ilustración 4 Base de datos MySQL Perú

En la ilustración3, se puede apreciar una base de datos relacional donde se encuentra almacenada la base de datos “Datos_Peru” y “Datos_EU”, misma que contiene la información extraída de las fuentes: “IES” y “Gob.pe”. Se realizó la extracción de los datos almacenados en la base de datos y mediante Python se procedió a hacer la limpieza de datos, eliminando los datos nulos y aquellos campos que no sean relevantes para el caso de estudio. El código con el que se realizó la limpieza

de datos se encuentra en el archivo ipynb que se encuentra en la tabla de Recurso y herramientas utilizadas.

Luego de obtener la data de todas las fuentes limpia y con aquellos datos que se utilizó para el análisis se procedió a guardar cada una de la data en diferentes archivos csv, con la finalidad de importar todos los 5 archivos csv, en una base de datos relacional en este caso

SQL Server, misma base de datos que se utilizó para importarla data en Power BI para el:

- Análisis de información.
- Visualización de información.
- Dashboard.

Gracias a la herramienta de Power BI se obtuvo la siguiente información.

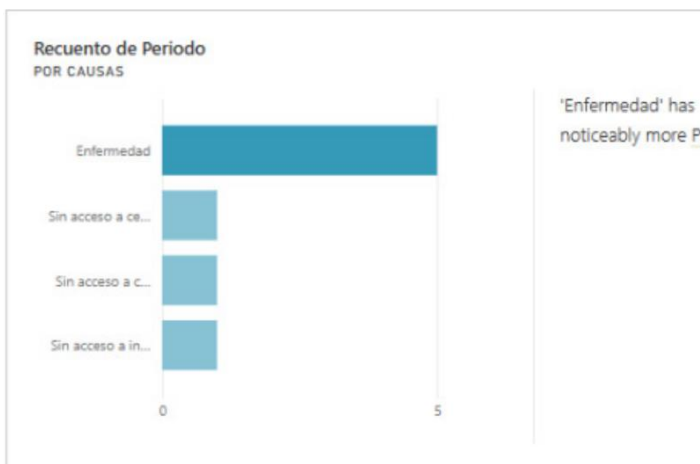


Ilustración 5 Colombia

La mayor cantidad de estudiantes se retiraron en los anteriores 4 años fue por temas de enfermedad, podemos tomar como pequeña referencia que son por el tema del covid en los años 2020 y 2021.

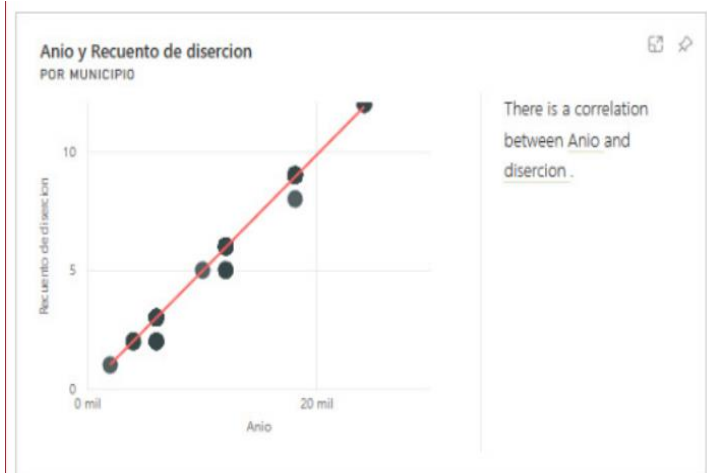


Ilustración 6 Colombia

Se puede observar como en los años de la pandemia que son del 2020 a inicios del 2022 se tiene una mayor deserción de los estudiantes cómo podemos observar de los tres puntos que se encuentran fuera de la línea roja.

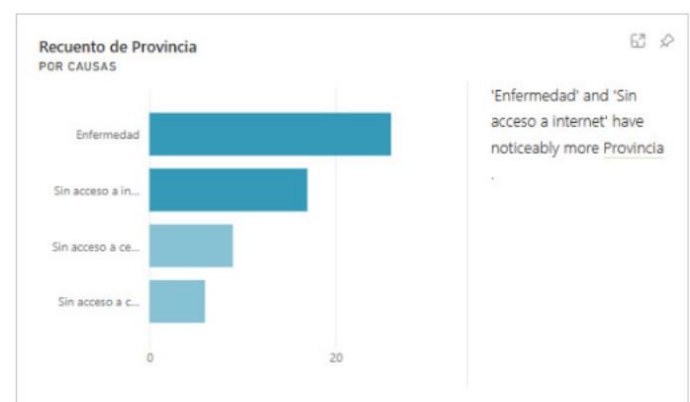


Ilustración 7 Ecuador

Se puede observar que todas las provincias del Ecuador tuvieron casos de enfermedad del covid 19 en los últimos años seguido por estudiantes que no poseen acceso al internet.

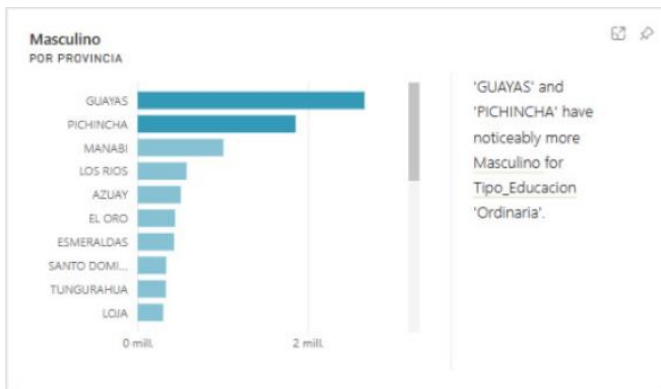


Ilustración 8 Ecuador

Podemos observar que Guayas tiene la mayor abundancia de estudiantes Masculinos en los últimos 5 años con más de 2 millones de estudiantes Masculinos seguido directamente por Pichincha con una cantidad cercana de 2 millones de estudiantes Masculinos.

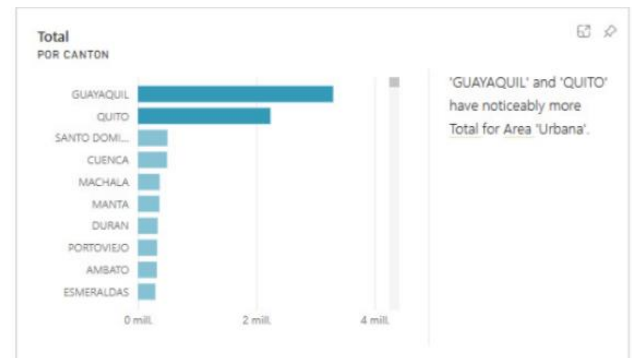


Ilustración 10 Ecuador

La cantidad de estudiantes pertenecientes a la zona urbana se encuentra reflejada directamente en el cantón de Guayaquil con un aproximado a los 3.5 millones de estudiantes de primaria, secundaria y bachillerato seguido directamente por el cantón Quito con un aproximado de 2.3 millones de estudiantes.

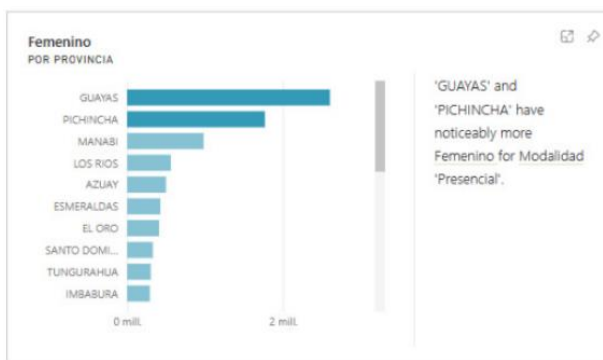


Ilustración 9 Ecuador

Podemos observar que Guayas tiene la mayor abundancia de estudiantes Femeninos en los últimos 5 años con más de 2 millones de estudiantes Femeninos seguido directamente por Pichincha con una cantidad cercana de 2 millones de estudiantes Femeninos.

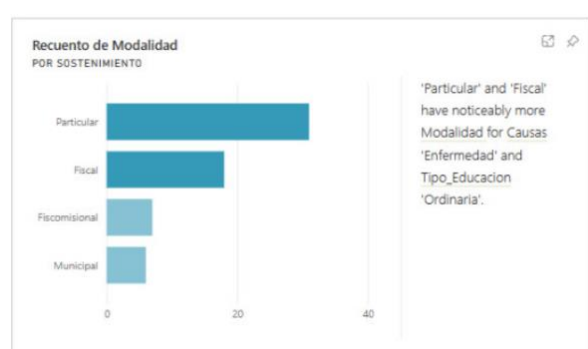


Ilustración 11 Nivel mundial

Se observa que las escuelas de carácter privado han sufrido una reducción de un 30% aproximada en los últimos 5 años, seguido por las escuelas fiscales con un aproximado del 19.5%.

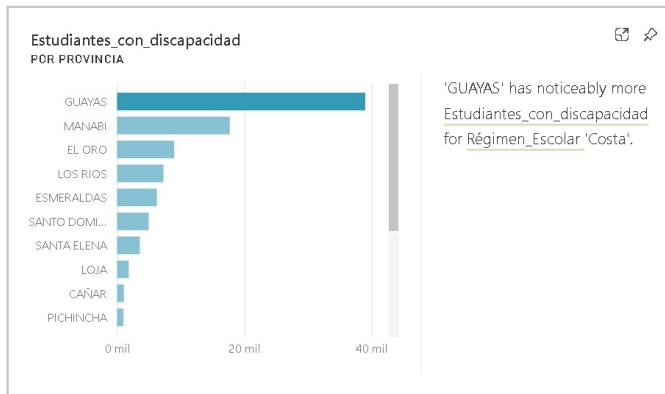


Ilustración 12 Ecuador

Observamos que Guayas cuenta con la mayor cantidad de estudiantes que poseen algún tipo de discapacidad con una cantidad aproximada de 40 mil, seguido directamente por Manabí que pose alrededor de 20 mil estudiantes con algún tipo de discapacidad.

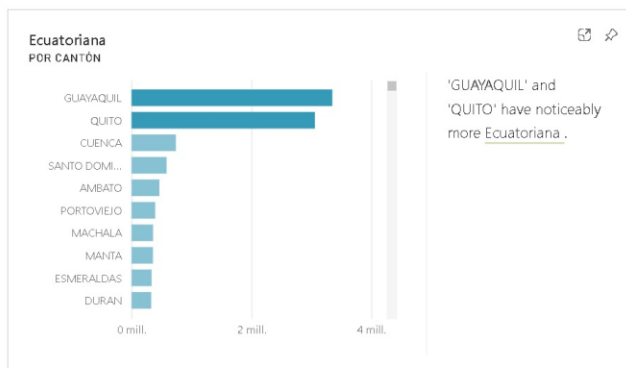


Ilustración 13 Ecuador

La mayor cantidad de población estudiantil con nacionalidad ecuatoriana se encuentra actualmente en Guayaquil seguido directamente por Quito ambos estarían rondando entre unos 3.5 y 3.3 millones respectivamente.

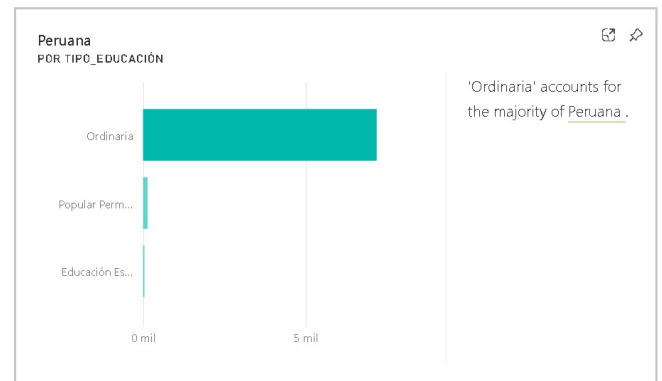


Ilustración 14 Ecuador

Podemos ver que más del 90% de los estudiantes de nacionalidad peruana se encuentran dentro del periodo de estudio llamado Ordinario.

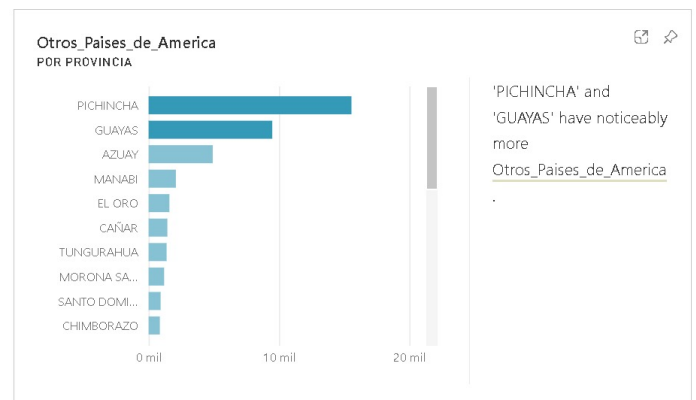


Ilustración 15 Ecuador

Se observa que Pichincha cuenta con la mayor cantidad de estudiantes Extranjero que no pertenecen a los países de Perú, Colombia o Venezuela con una cantidad que ronda entre 15 y 16 mil estudiantes seguido por Guayaquil que tiene una cantidad aproximada a los 10 mil estudiantes.

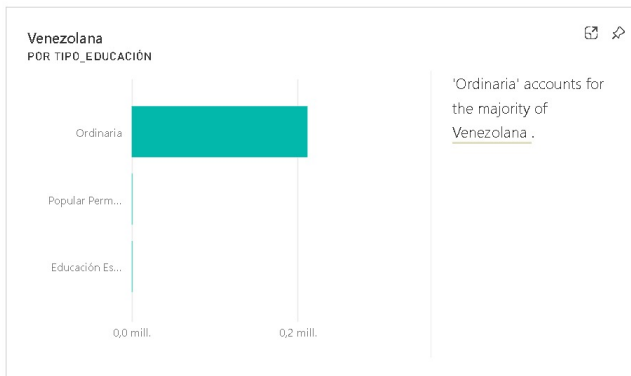


Ilustración 16 Ecuador

Podemos ver que más del 90% de los estudiantes de nacionalidad venezolana se encuentran dentro del periodo de estudio llamado Ordinario .

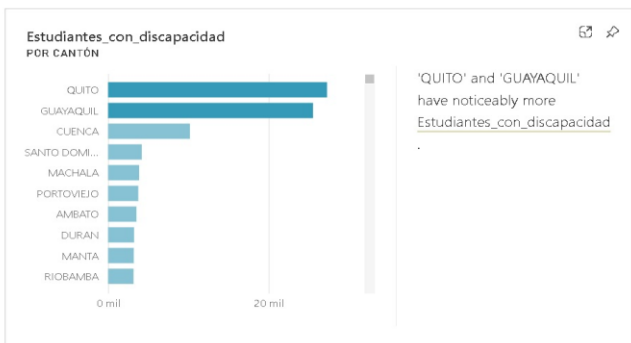


Ilustración 17 Ecuador

Observamos que Quito cuenta con la mayor cantidad de estudiantes que poseen algún tipo de discapacidad con una cantidad aproximada de 25 a 30 mil, seguido directamente por Guayaquil que posee alrededor de 23 a 26 mil estudiantes con algún tipo de discapacidad.

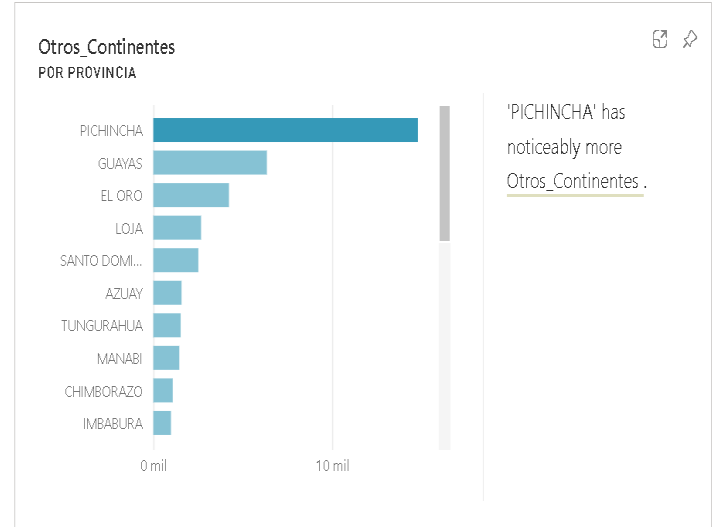


Ilustración 18 Ecuador

Pichincha cuenta con la mayor cantidad de estudiantes los cuales proceden desde otros continentes diferentes a América del Sur cerca a los 20 mil estudiantes entre hombres y mujeres seguido por la provincia de Guayas el cual tiene un aproximado a los 8 mil estudiantes.

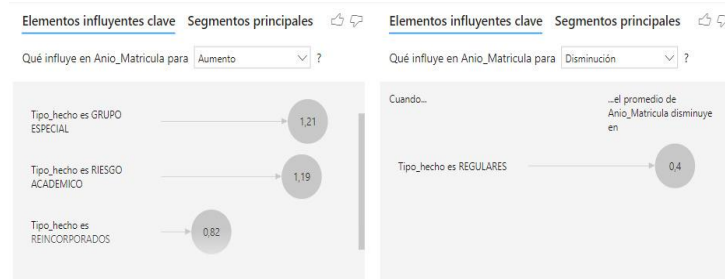


Ilustración 19 Perú

Podemos ver que el 1.21% de estudiantes se retiran es un por un pertenecer al grupo especial de estudiantes, mientras que el 1.19% son por temas de riesgo académico.

Suma de Anio_Matricula por Causas

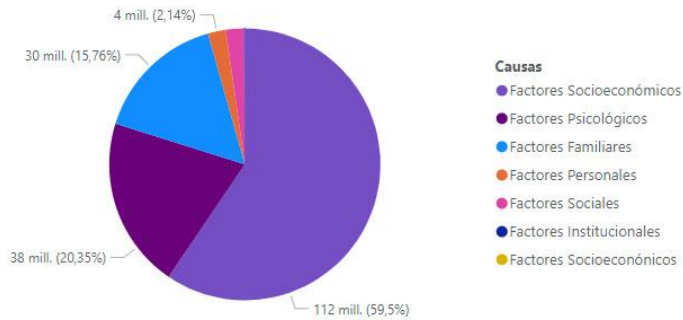


Ilustración 20 Perú

Se puede observar que en Perú gran parte de la población estudiantil posee un como factor principal del riesgo Socioeconómico con un 59.5% del total solo por delante de los Factores Psicológicos con un 20.35% del total y Factores familiares con el 15.76% del total.

5. Recurso y herramientas utilizadas

Tabla 1 Recursos utilizados en el proyecto

Recursos y herramientas utilizadas	
Git Hub	https://github.com/MiguelCarapaz/Proyecto_Analisis_de_Datos.git
Power BI	https://app.powerbi.com/groups/me/dashboards/5084dc7d-9dc6-4c78-aeb6-b972f4c6a9cd

6. Arquitectura de la solución

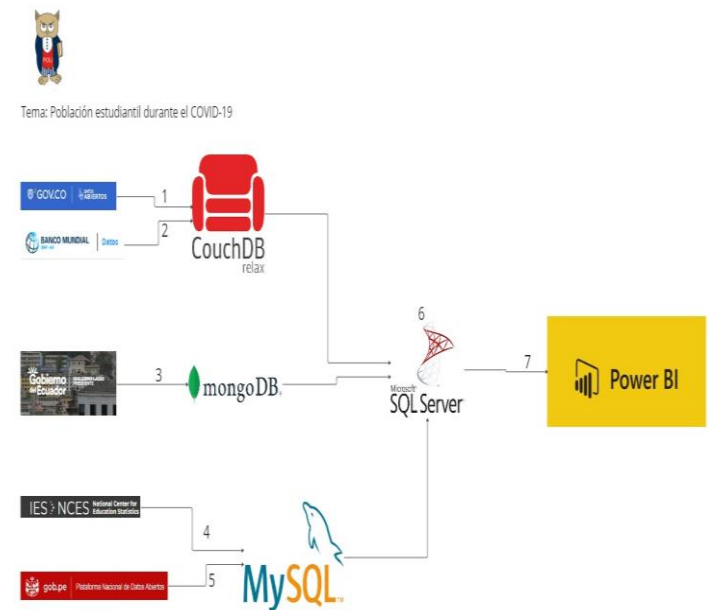


Ilustración 20 21 Arquitectura de la solución

7. Conclusiones

- En conclusión, la pandemia COVID-19 ha tenido un impacto significativo en el acceso a la educación en todo el mundo, especialmente en los países que se está analizando en el Data Lake y a nivel mundial. Los datos muestran que muchos estudiantes han perdido el acceso a la educación debido a la pandemia, y que aquellos que tienen acceso a la educación en línea a menudo enfrentan desafíos como la brecha digital y la falta de recursos adecuados. Para mejorar el acceso a la educación, se deben implementar políticas públicas y programas de inversión en infraestructura educativa y apoyo económico para los estudiantes. Además, se

deben tomar medidas para cerrar la brecha digital y garantizar que todos los estudiantes tengan acceso a los recursos necesarios para el aprendizaje en línea. En general, el acceso a la educación es fundamental para el desarrollo de la sociedad y es necesario abordar estos desafíos para garantizar un futuro mejor para todos los estudiantes.

8. Recomendaciones

- En el caso de que no se pueda encontrar la información que se quiere de diversas fuentes, se recomienda buscar la información por partes, es decir, no buscar el tema en general, sino, buscar por puntos específicos del tema general, esto ayudara a que puedas encontrar mayor cantidad de información y más específica.
- Se recomienda que al momento de descargar la data verificar si la data está bien estructurada es decir que el archivo no se dañó al momento de la descarga o extracción de datos, ya que, esto puede causar confusiones al momento de importar la data.

9. Desafíos y problemas encontrados

- Encontrar la información fue uno de los desafíos y problemas en el que se invirtió la mayoría del tiempo.

- Al momento de realizar los gráficos en Power BI, se tuvo el problema de que los gráficos no salían como se esperaba.

10. Bibliografía

- [1] «Datos Abiertos del Ministerio de Educación del Ecuador – Ministerio de Educación», Gob.ec. [En línea]. Disponible en: <https://educacion.gob.ec/datos-abiertos/>. [Accedido: 05-mar-2023].
- [2] «Indicators», Bancomundial.org. [En línea]. Disponible en: <https://datos.bancomundial.org/indicador>. [Accedido: 05-mar-2023].
- [3] «Alumnos matriculados», Gob.pe. [En línea]. Disponible en: <https://www.datosabiertos.gob.pe/dataset/alumnos-matriculados>. [Accedido: 05-mar-2023].
- [4] Education Department, Digest of education statistics, 2020. Lanham, MD: Bernan Press, 2022.
- [5] «MEN_ESTADISTICAS_EN_EDUCACION_EN_P REESCOLAR, BÁSICA Y MEDIA_POR_MUNICIPIO», Gov.co. [En línea]. Disponible en: <https://acortar.link/gGNpyT>. [Accedido: 05-mar-2023].