

Proyecto final Análisis de Datos.

TECNOLOGIA SUPERIOR EN DESARROLLO DE SOFTWARE

INTEGRANTES

Adrian Cadena

Joahn Cardenas

Kevin Almeida

INSTRUCTOR

Ing. Lorena Chulde

FECHA DE ENTREGA

13 de agosto de 2024

2024

Contenido

1. DEFINICION DEL ESTUDIO DE CASO.....	3
2. EQUIPO DE TRABAJO.	3
3. CRONOGRAMA.	3
4. RECURSOS Y HERRAMIENTAS UTILIZADAS.	3
5. ARQUITECTURA DE LA SOLUCIÓN.....	3
6. EXTRACCIÓN DE DATOS.	3
7. ANÁLISIS DE INFORMACIÓN.	4
8. VISUALIZACION DE LA	4
9. INFORMACIÓN.....	4
10. RESULTADOS OBTENIDOS.....	8
11. CONCLUSIONES.....	8
12. RECOMENDACIONES.....	8

1. DEFINICION DEL ESTUDIO DE CASO.

Objetivo: Utilizar PowerBI para aplicar los conocimientos adquiridos sobre el análisis de datos, que incluyen la extracción, limpieza y visualización de datos.

Descripción: Crear dashboards explicativos y una arquitectura de datos que incluya al menos diez fuentes de datos de diversas temáticas, utilizando tanto bases de datos relacionales como NoSQL.

Objetivo general: Realizar una solución completa de análisis de datos que posibilite la toma de decisiones basadas en la información recopilada de varias fuentes de datos.

Objetivos específicos:

- Crear dashboards en Power BI que permitan visualizar la información de manera efectiva.
- Diseñar una arquitectura de datos integrando bases de datos relacionales y NoSQL.
- Analizar y presentar los resultados obtenidos a partir de los dashboards.

2. EQUIPO DE TRABAJO.

Adrian Cadena
Joahn Cardenas
Kevin Almeida

3. CRONOGRAMA.

Semana 1: Recolección de fuentes de datos (data sets) para su respectivo análisis con diferentes dashboards explicativos.

Semana 2: Consultas dentro de los data sets con la ayuda de la aplicación de Power BI para su posterior análisis.

4. RECURSOS Y HERRAMIENTAS UTILIZADAS.

Power BI: Permite la toma de decisiones rápidas, puesto que, es un servicio analítico basado en la nube permitiendo el análisis y visualización de datos de forma rápida.

MongoDb: Es una base de datos que maneja datos NoSQL, proporcionando un sistema de almacenamiento de datos flexible permitiendo a los usuarios gestionar datos multivariados.

SQL: Es un sistema de administración de bases de datos que maneja datos relacionales, es decir maneja un modelo de datos específico por medio de tablas.

5. ARQUITECTURA DE LA SOLUCIÓN.



6. EXTRACCIÓN DE DATOS.

Conversión de datos por medio de SQLite

En primer lugar, se realiza la importación de a MongoDB

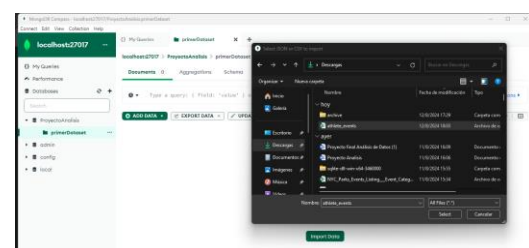


Ilustración 1 –Importación de datos

Realizamos una consulta por medio de Jupyter para comprobar



Ilustración 2 – Consulta

Se exporta el dataset pero esta vez en formato JSON.

Ilustración 3 – Exportación Completa

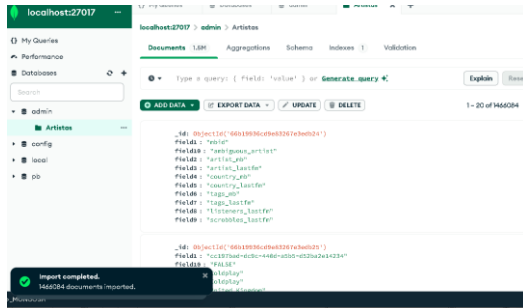


Ilustración 4 - Importar a MongoDB

Una vez se haya completado la exportación del dataset se lo puede importar dentro de la plataforma de MongoDB.

Conversión de datos por medio de PostgreSQL

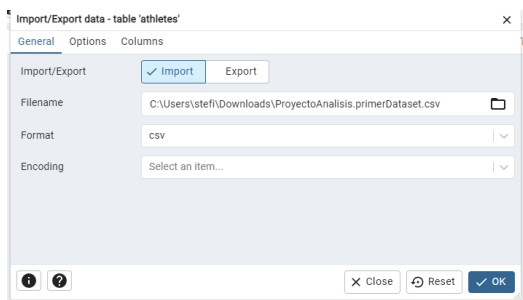


Ilustración 5 – Importación de dato.

En primer lugar, se realiza la importación de datos a la plataforma en formato csv.

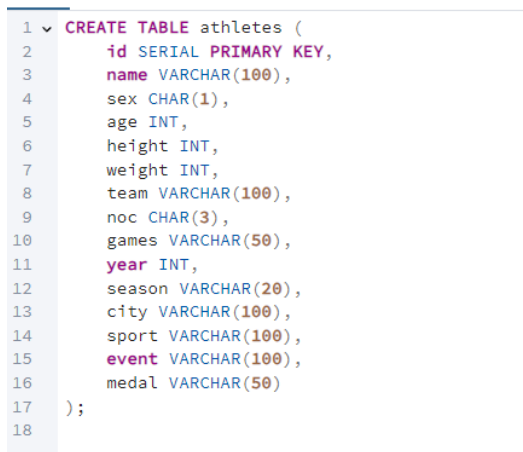


Ilustración 6 – Creación de tablas.

Para importar a PostgreSQL se necesita tener creada antes la tabla con los campos, además de activar la opción de que se cuente la primera línea como cabecera.

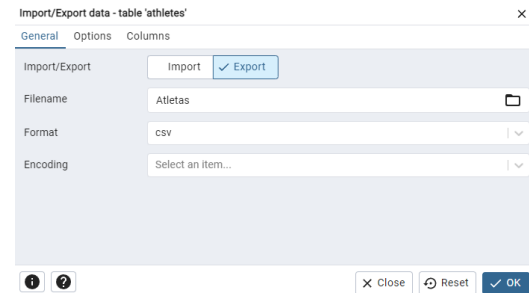


Ilustración 7 - Exportación de datos

Una vez se realizó la importación correctamente y se insertó el Dataset en la tabla se puede realizar la exportación en formato JSON.

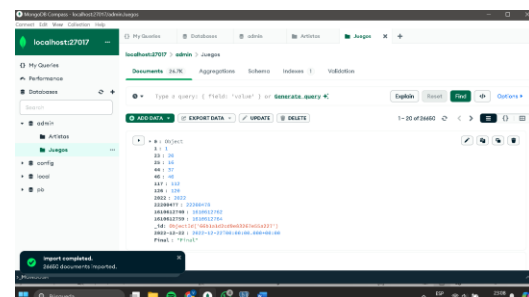


Ilustración 8 – Importar datos en MongoDB

Finalmente, ya se puede insertar los datos dentro de la base de datos en MongoDB.

7. ANÁLISIS DE INFORMACIÓN.

Usamos PowerBi para hacer un análisis en los datos y llegar a algunas conclusiones.

8. VISUALIZACION DE LA INFORMACIÓN.

DATASET 1: Atletas

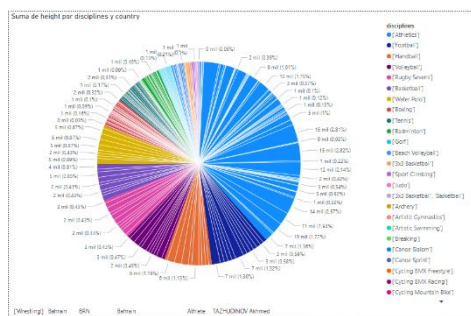


Ilustración 9 – Consulta 1

Los deportes que se realizaron cuentan con una variedad de deportistas de diferentes pesos.

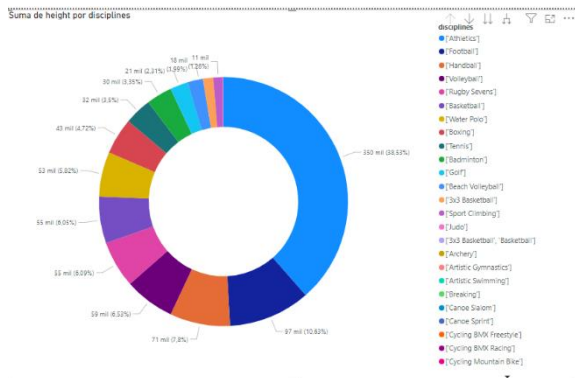


Ilustración 10 – Consulta 2

En la siguiente consulta, se la sumatoria del peso de los deportistas por disciplina.

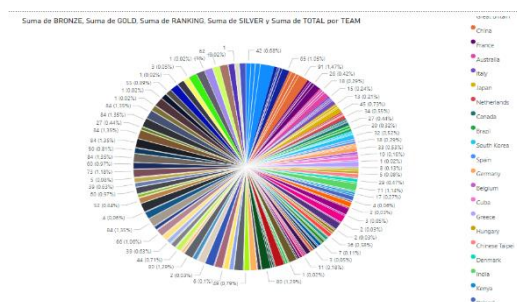


Ilustración 11 – Medallas

Se muestra las medallas ganadas por equipos o mejor dicho países que participaron.

Suma de BRONZE	Suma de GOLD	Suma de SILVER
384	328	327

Ilustración 12 – Medallas totales

Se presenta en la imagen el numero total de medallas entregadas.

TEAM	Suma de SILVER	Suma de GOLD	Suma de BRONZE	Suma de RANKING
Chinese Taipei	3	0	1	66
Colombia	0	0	1	84
Côte d'Ivoire	2	2	3	30
Croatia	1	2	6	32
Cuba	1	0	0	74
Cyprus	0	3	2	28
Czechia	2	2	5	29
Denmark	0	1	0	62
Dominica	0	1	2	59
Dominican Republic	2	1	2	49
Ecuador	1	1	1	52
Egypt	3	1	0	47
Ethiopia	1	0	0	74
Fiji	26	16	22	5
France	3	3	1	24
Georgia	13	12	8	10
Germany	327	328	384	4123

Ilustración 13 – Intento de Medallero

En esta consulta se intento hacer un

DATASET 2: Artistas

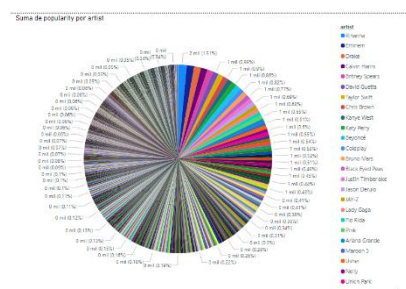


Ilustración 15– Popularidad de Artistas.

Se presenta a os artistas y su popularidad alrededor del mundo

artist	Suma de popularity
Rihanna	1803
Britney	1519
Drake	1424
Calvin Harris	1356
Britney Spears	1198
David Guetta	1185
Taylor Swift	1125
Chris Brown	1080
Kanye West	1049
Katy Perry	1033
Beyonce	1030
Goldplay	979
Bruno Mars	938
Black Eyed Peas	917
Justin Timberlake	847
Jason Derulo	824
JAY-Z	813
Lady Gaga	809
Flo Rida	788
Rick	784
Ariana Grande	759
Maroon 5	725
Usher	719
Nelly	716
Linkin Park	678
Ed Sheeran	655
Pitbull	649
Destiny's Child	646
Post Malone	645
50 Cent	640
Avril Lavigne	627
One Direction	623
Total	119745

Ilustración 16– Popularidad en tabla

Se muestra de mayor a menor la popularidad de cada artista

artist	genre	song	Suma de popularity	Suma de d
*NSYNC	pop	Bye Bye Bye	65	614
*NSYNC	pop	Grease	50	746
*NSYNC	pop	Gone	45	689
*NSYNC	pop	It's Gonna Be Me	60	644
112	hip hop, pop, R&B	It's Over Now	87	66
112	hip hop, pop, R&B	Reaches & Cream	63	677
2 Chanz	hip hop, pop	It's A Vibe	71	622
3 Chanz	hip hop, pop	We Own It (Fast & Furious)	68	543
21 Savage	hip hop	a lot	78	637
21 Savage	hip hop	Bank Account	74	854
2pac	hip hop	Ghetto Gospel	68	794
2pac	hip hop	Until The End Of Time	37	737
3 Doors Down	pop, rock, metal	Here Without You	74	157
3 Doors Down	pop, rock, metal	It's Not My Time	56	528
3 Doors Down	pop, rock, metal	Kryptonite	78	545
3 Doors Down	pop, rock, metal	Let Me Go	52	478
3 Doors Down	pop, rock, metal	When I'm Gone	67	53
3 Doors Down	pop, rock, metal	My First Kiss (feat. Kellie)	70	791
3 Doors Down	pop, rock, metal	Starstrukk	32	605
3 Doors Down	pop, rock, metal	East	1	505
5 Seconds of Summer	pop	She Looks So Perfect	74	494
5 Seconds of Summer	pop	Youngblood	69	596
50 Cent	hip hop, pop	21 Questions	72	645
50 Cent	hip hop, pop	Ayo Technology	67	63
50 Cent	hip hop, pop	Candy Shop	79	614
50 Cent	hip hop, pop	Disco Inferno	72	925
50 Cent	hip hop, pop	In Da Club	81	899
50 Cent	hip hop, pop	Just A Lil Bit	76	489
50 Cent	hip hop, pop	Outro (feat. Remy)	63	772
50 Cent	hip hop, pop	P.I.M.P.	76	712
Total			119745	1212223

Ilustración 147 – Canciones mas famosas

Muestra los artistas y sus canciones mas escuchadas hasta el momento

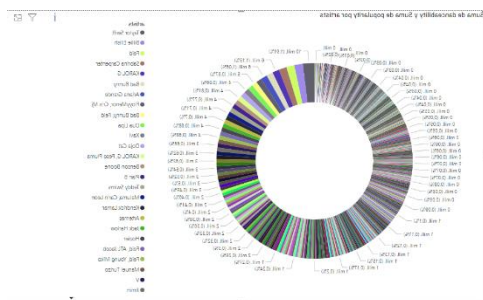


Ilustración 18 – “Bailabilidad”

La capacidad de que un artista ponga a bailar a su oyente.

DATASET 3: Películas

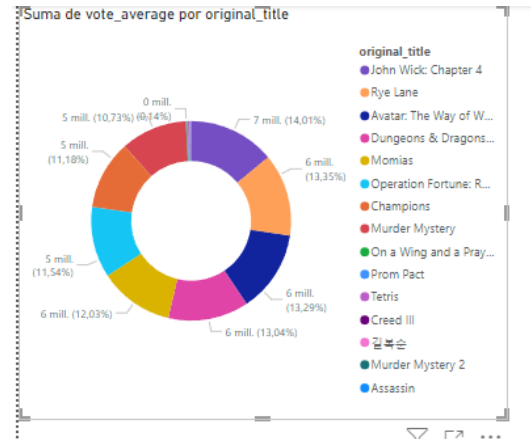


Ilustración 19– Aprobación a películas

Se muestran las películas con mayor aprobación por medio de los fans.

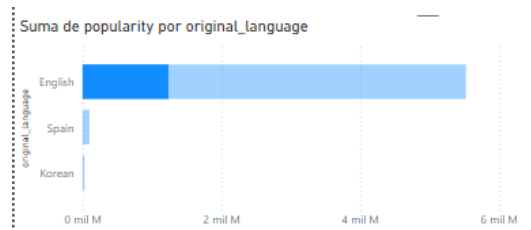


Ilustración 20 – Idioma original

Se muestra los 3 idiomas en la cual originalmente se hacen las películas dominando por completo del inglés.

original_title	English	Korean	Spain	Total
Assassin	8844			8844
Avatar: The Way of Water	5093340			5093340
Champions	34572			34572
Creed III	451044			451044
Dungeons & Dragons: Honor Among Thieves	102912			102912
John Wick: Chapter 4	510540			510540
Momias			51456	51456
Murder Mystery	2928168			2928168
Murder Mystery 2	150348			150348
On a Wing and a Prayer	3216			3216
Operation Fortune: Ruse de Guerre	238788			238788
Prom Pact	8040			8040
Rye Lane	14472			14472
Tetris	80400			80400
김복순		31356		31356
Murder Mystery 2			51456	51456
Assassin				8844
Total	9624684	31356	51456	9707496

Ilustración 21 – Dominio de votos

La mayoría de gente que ha votado sobre lo buena o mala que haya sido una película son de habla inglesa.

DATASET 4: Vehiculos Electricos

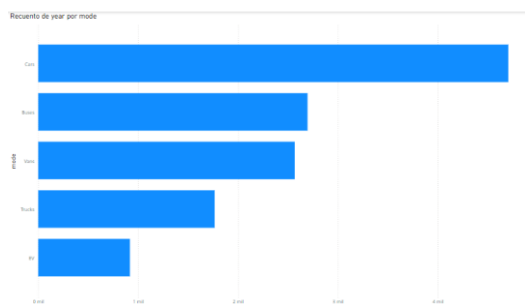


Ilustración 22 – Cantidad de carros generados

Carros generados por año por año.

region	unit	mode	Recuento de year
Australia	charging points	EV	12
Australia	percent	Cars	26
Australia	Vehicles	Cars	56
Austria	charging points	EV	26
Austria	percent	Cars	25
Austria	Vehicles	Cars	65
Belgium	charging points	EV	22
Belgium	percent	Buses	22
Belgium	percent	Cars	28
Belgium	percent	Trucks	22
Belgium	percent	Vans	28
Belgium	Vehicles	Buses	41
Belgium	Vehicles	Cars	79
Belgium	Vehicles	Trucks	24
Belgium	Vehicles	Vans	44
Brazil	charging points	EV	12
Brazil	percent	Buses	17
Brazil	percent	Cars	24
Brazil	percent	Trucks	12
Brazil	percent	Vans	17
Brazil	Vehicles	Buses	23
Brazil	Vehicles	Cars	49
Brazil	Vehicles	Trucks	12
Brazil	Vehicles	Vans	17
Bulgaria	percent	Cars	6
Bulgaria	Vehicles	Cars	17
Canada	charging points	EV	24
Canada	percent	Buses	14
Canada	percent	Cars	26
Canada	percent	Trucks	23
Canada	percent	Vans	10
Canada	Vehicles	Buses	14
Canada	Vehicles	Cars	71
Total			12654

Ilustración 23 – Modelos por año

Se presentan la cantidad de veces que se hizo un modelo en un país por año.

DATASET 5: Restaurantes Ranking

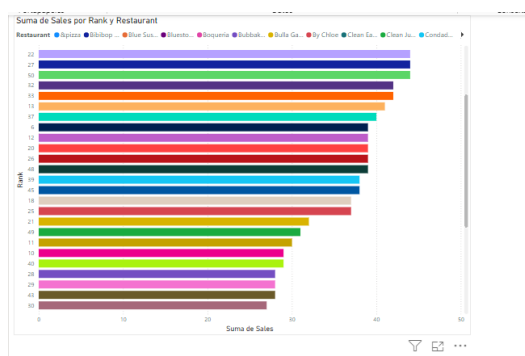


Ilustración 24 – Ventas

Muestra de mayor a menor el numero de venta que tiene cada Restaurante

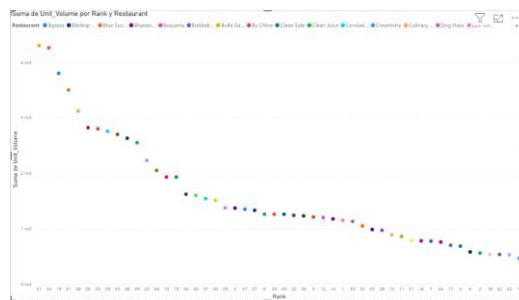


Ilustración 25 – Tamaño de restaurant.

Se muestran el restaurante con mayor unidad de volumen



Ilustración 26 – Máximo de ventas

Muestra el número máximo de ventas anuales que recibió un restaurante

9. RESULTADOS OBTENIDOS.

10. CONCLUSIONES.

11. DESAFIOS Y COMPLICACIONES.

Cadena: Aquí Me gustaría expresar despecho con la forma en que algunos miembros del grupo han manejado sus responsabilidades durante este proyecto. Mientras uno al menos cumplió con su parte al realizar el video y algunas consultas, el otro no aportó absolutamente nada más que excusas. Debido a la falta de compromiso y responsabilidad del Sr Almeida, confiamos en él para completar que pasara los datos de mongo a sqllite y solo dijo que no valía o no le salía, y esto no solo causó que nos demoremos más, sino que también nos obligó a hacer todo el trabajo un día antes de la entrega. Y todo el informe se paso a mi y son la 1:37 de la mañana del día 13 de agosto y sigue sin aportar nada.

Aparte de eso la mayoría de los datos no valían de los datasets descordinaban todas la graficas y el paso de una tras otro tipo de base de datos corrompio 2 datasets.

12. RECOMENDACIONES.

Hacerlo todo tú mismo.

13. BIBLIOGRAFIA

<https://www.kaggle.com/datasets/mohdshahnawazaadil/restaurant-dataset>

<https://www.kaggle.com/datasets/rajku marpandey02/electric-vehicle-population-data>

<https://datasetsearch.research.google.com/>