

# Análisis y Visualización de Datos por Medio de Power BI

Bernal José; Porras Eduardo; Tacuri Mateo

Escuela Politécnica Nacional, Escuela de Formación de Tecnólogos, Desarrollo de Software, Quito, Ecuador

**Resumen:** Este proyecto se centra en el análisis y visualización de datos utilizando datasets que cubren temas como eventos deportivos, actividades y hobbies, conciertos y eventos públicos, ventas de consolas y videojuegos, restaurantes y sitios de esparcimiento, y eventos o noticias de relevancia global. Los datos fueron recolectados de diversas fuentes, como Google Dataset Search, Kaggle y UCI Machine Learning Repository, con un total aproximado de un millón de registros. El equipo de trabajo aplicó técnicas avanzadas de análisis de datos para crear dashboards interactivos en Power BI, demostrando la capacidad para manejar grandes volúmenes de información y presentarla de manera clara y efectiva. El proyecto es una aplicación práctica de los conocimientos adquiridos en la asignatura de Análisis de Datos de la carrera de Desarrollo de Software en la ESFOT.

**Palabras clave:** datasets; limpieza de datos; arquitectura de análisis; visualización interactiva; consolidación de datos; técnicas de análisis.

## Data Analysis and Visualization through Power BI

**Abstract:** This project focuses on data analysis and visualization using datasets covering topics such as sports events, activities and hobbies, concerts and public events, console and video game sales, restaurants and recreational sites, and global news or events. The data were collected from various sources, including Google Dataset Search, Kaggle, and UCI Machine Learning Repository, totaling approximately one million records. The team applied advanced data analysis techniques to create interactive dashboards in Power BI, demonstrating the ability to manage large volumes of information and present it clearly and effectively. The project is a practical application of the knowledge gained in the Data Analysis course in the Software Development program at ESFOT.

**Keywords:** datasets, data cleaning, analysis architecture, interactive visualization, data consolidation, analysis techniques.

### 1. DEFINICIÓN DEL CASO DE ESTUDIO

El presente proyecto se enmarca en la búsqueda y análisis de datasets sobre diversas temáticas, como eventos deportivos a nivel mundial, actividades y hobbies, conciertos y eventos públicos, ventas de consolas y videojuegos, restaurantes y sitios de esparcimiento, y eventos o noticias de relevancia global. La importancia de este caso de estudio radica en la aplicación práctica de los conocimientos adquiridos durante el semestre en la asignatura de Análisis de Datos, parte de la carrera de Desarrollo de Software en la ESFOT. Este proyecto tiene como objetivo principal demostrar la capacidad de los estudiantes para extraer, preparar, analizar y presentar datos provenientes de diferentes medios, utilizando herramientas y técnicas adecuadas.

Los problemas y desafíos abordados incluyen la identificación de datos relevantes, la diferenciación entre datos estructurados y no estructurados, la importancia de la consistencia de datos, y la necesidad de proteger y garantizar la fiabilidad de los datos en todo el proceso de análisis. Además, se enfrenta el desafío de consolidar y analizar alrededor de un millón de registros

provenientes de distintas fuentes, lo que requiere una arquitectura sólida y un enfoque colaborativo.

### 2. OBJETIVO GENERAL & ESPECÍFICO

**Objetivo General:** Aplicar los conceptos y herramientas de análisis de datos aprendidos durante la asignatura para extraer, preparar, analizar y presentar información relevante a partir de datasets sobre diversas temáticas, con el fin de demostrar la capacidad de análisis y la presentación efectiva de resultados en un entorno profesional.

#### Objetivos Específicos:

- Identificar y recolectar datasets relevantes** de distintas fuentes como Google Dataset Search, Kaggle, y UCI Machine Learning Repository, para cubrir las temáticas seleccionadas.
- Aplicar técnicas de limpieza y preparación de datos** para asegurar la consistencia y calidad de los datos antes del análisis.

- 3. **Desarrollar y aplicar una arquitectura de análisis de datos** que permita procesar eficientemente alrededor de un millón de registros.
- 4. **Crear dashboards interactivos en Power BI** que permitan visualizar y comunicar de manera clara los resultados del análisis.
- 5. **Documentar y explicar el proceso de análisis** y las conclusiones derivadas del mismo, destacando la importancia de la ética profesional en la protección y fiabilidad de los datos.

3. DESCRIPCIÓN DEL EQUIPO DE TRABAJO Y ACTIVIDADES REALIZADAS

Equipo de Trabajo:

- **José Bernal:** Responsable de la búsqueda y recolección de datasets sobre **eventos deportivos a nivel mundial y conciertos y eventos públicos**. Además, participó en la limpieza y preparación de estos datos, asegurando su consistencia para el análisis posterior.
- **Eduardo Porras:** Encargado de los datasets relacionados con **actividades y hobbies y ventas de consolas y videojuegos**. Eduardo también fue responsable de la implementación de la arquitectura de análisis de datos para estos temas.
- **Mateo Tacuri:** Se ocupó de la búsqueda y preparación de datos sobre **restaurantes y sitios de esparcimiento y eventos o noticias mundiales**. Mateo trabajó en la integración de los datos recolectados por todos los miembros en un solo archivo de Power BI, y fue clave en la creación de los dashboards.

Actividades Realizadas:

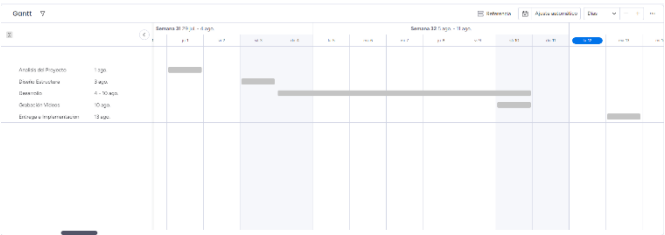
- **Recolección de Datasets:** Cada integrante se enfocó en dos temáticas específicas, buscando datasets relevantes que sumaran un total de alrededor de un millón de registros.
- **Preparación y Limpieza de Datos:** Todos los miembros participaron en la limpieza de los datos, aplicando técnicas para asegurar que los datasets fueran consistentes y estuvieran listos para el análisis.
- **Desarrollo de la Arquitectura de Análisis:** Juntos, el equipo diseñó y aplicó una arquitectura que permitiera manejar eficientemente la gran cantidad de datos recolectados.
- **Creación de Dashboards en Power BI:** Los datos procesados se integraron en un archivo de Power BI, donde se crearon visualizaciones interactivas que permiten una interpretación clara y concisa de la información.
- **Documentación del Proceso:** Finalmente, el equipo documentó todo el proceso, explicando las técnicas utilizadas y las conclusiones obtenidas, resaltando la importancia de la ética y la responsabilidad en el manejo de los datos.

4. CRONOGRAMA DE ACTIVIDADES

Mediante un diagrama de tablas se aprecia el cronograma de actividades que realizó el equipo de trabajo.

Actividad	Fecha Inicio	Fecha Fin	Responsables
Análisis del Proyecto	1/08/2024	1/08/2024	Mateo Tacuri Mateo Bernal  Eduardo Porras
Diseño Estructura	3/08/2024	3/08/2024	Mateo Tacuri Mateo Bernal  Eduardo Porras
Desarrollo	4/08/2024	10/08/2024	Mateo Tacuri Mateo Bernal  Eduardo Porras
Grabación Videos	10/08/2024	10/08/2024	Mateo Tacuri Mateo Bernal  Eduardo Porras
Entrega e Implementación	13/08/2024	13/08/2024	Mateo Tacuri Mateo Bernal  Eduardo Porras

Diagrama de Gantt



Recursos y Herramientas utilizadas

Recursos

- Mongo Atlas
- MongoDB
- MySQLWorkbench
- Excel
- SQLite
- Microsoft SQL Server
- Power BI
- Visual Studio Code

Arquitectura de la solución

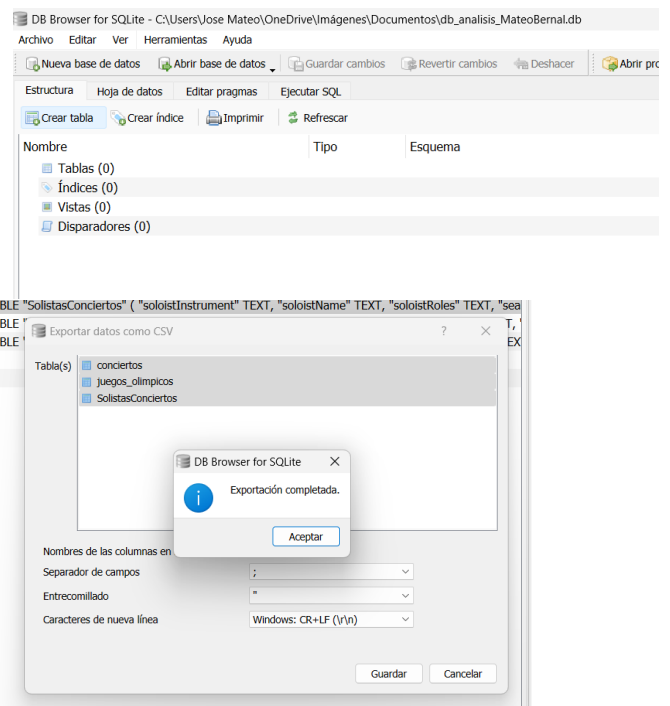
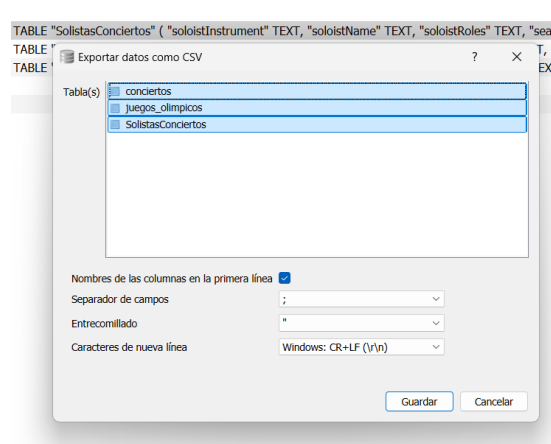
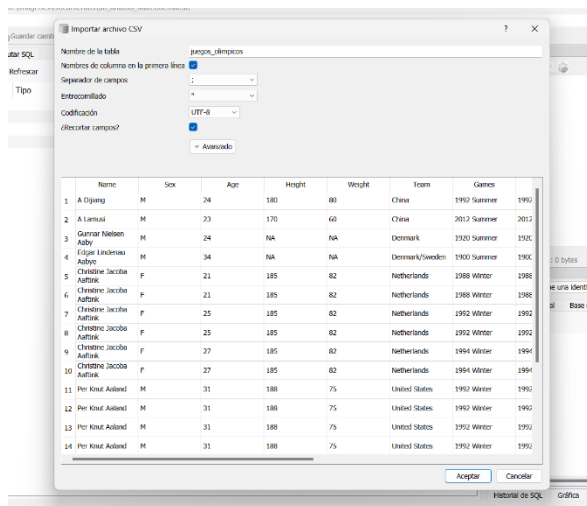


## 5. DESARROLLO

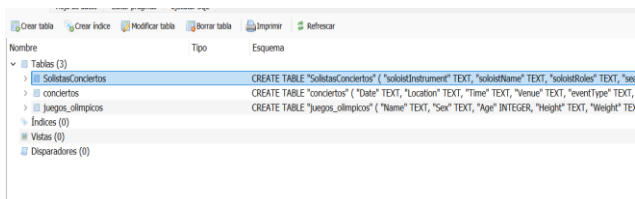
### Conexión a las distintas bases de datos

#### SQLite

- Creación de base de datos en 'DB browser'.
- Importación de los datos en formato CSV, los cuales serán transformados en tablas dentro de la base de datos.



- Se agregaron tres tablas a la base de datos de SQLite, las tablas tienen distintas temáticas (conciertos y eventos deportivos).



- Una vez guarda las tablas, se exportarán los datos en formato CSV para que estos puedan ser importados y usados en MySQL.

#### MySQL

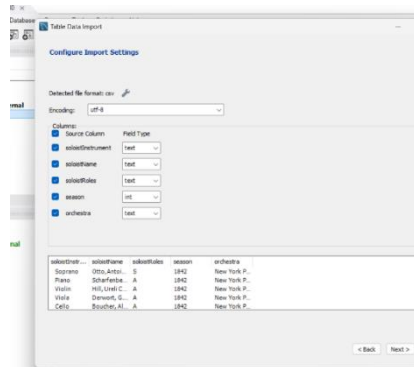
- Creación de una base de datos en MySQL con el objetivo de importar los archivos CSV provenientes de SQLite.

```

1
2 • create database db_analisis2_MateoBernal;
3 • use db_analisis2_MateoBernal;
4

```

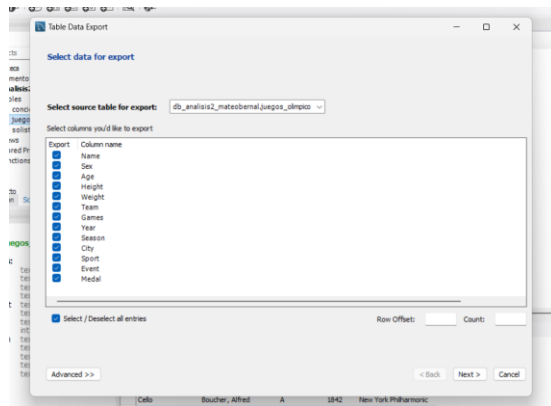
- Importación de archivos en formato CSV, para creación de las respectivas tablas.



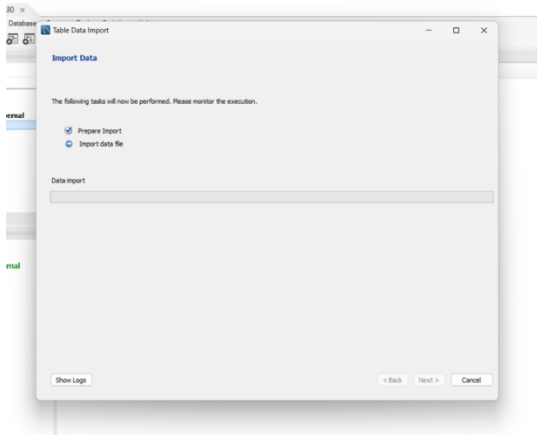
```
20 • select * from juegos_olimpicos;
```

Name	Sex	Age	Height	Weight	Team	Games	Year	Season	City	Sport	Event	Medal
A.Ojwang	M	24	180	80	China	1992 Summer	1992	Summer	Barcelona	Basketball	Basketball Men's Basketball	NA
A.Lamas	M	23	170	60	China	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Lightweight	NA
Gunnar Nielsen Aabye	M	24	NA	NA	Denmark	1920 Summer	1920	Summer	Antwerpen	Football	Football Men's Football	NA
Edgar Lindau Aabye	M	34	NA	NA	Denmark/Sweden	1900 Summer	1900	Summer	Paris	Tug-Of-War	Tug-Of-War Men's Tug-Of-War	Gold
Christine Jacobs Aafrika	F	21	185	82	Netherlands	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Women's 500 metres	NA
Christine Jacobs Aafrika	F	21	185	82	Netherlands	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Women's 1,000 metres	NA
Christine Jacobs Aafrika	F	25	185	82	Netherlands	1992 Winter	1992	Winter	Albertville	Speed Skating	Speed Skating Women's 500 metres	NA
Christine Jacobs Aafrika	F	25	185	82	Netherlands	1992 Winter	1992	Winter	Albertville	Speed Skating	Speed Skating Women's 1,000 metres	NA
Christine Jacobs Aafrika	F	27	185	82	Netherlands	1994 Winter	1994	Winter	Lillehammer	Speed Skating	Speed Skating Women's 500 metres	NA
Christine Jacobs Aafrika	F	27	185	82	Netherlands	1994 Winter	1994	Winter	Lillehammer	Speed Skating	Speed Skating Women's 1,000 metres	NA

- Una vez cargados los datos a la base de datos de MySQL, exportamos estos mismos datos en formato json para poder usarlos en la base de datos de mongo DB.



- La importación de datos, el tiempo puede variar dependiendo de la cantidad de datos que contenga el archivo CSV.



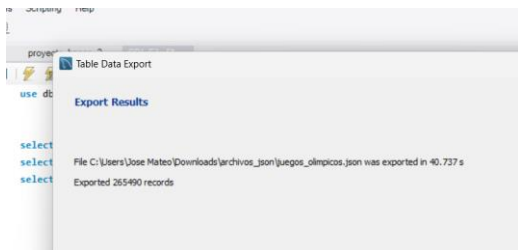
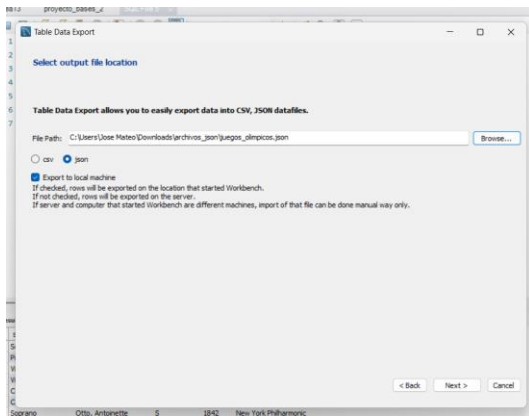
- Hacemos uso del comando 'select' para visualizar los datos agregados en la base de datos.

```
17
18 • select * from solistasconciertos;
19 • select * from conciertos;
```

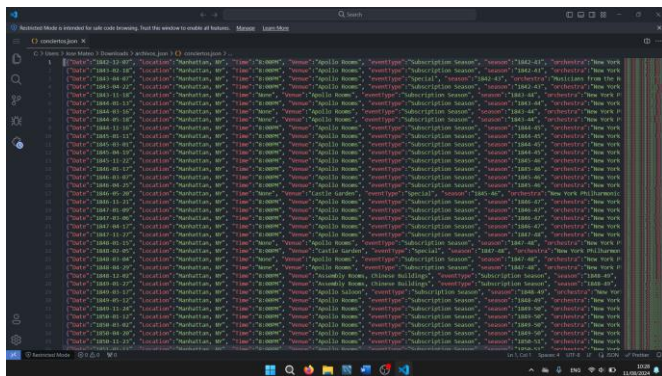
Date	Location	Time	Venue	eventType	season	orchestra
1842-12-07	Manhattan, NY	8:00PM	Apollo Rooms	Subscription Season	1842-43	New York Philharmonic
1843-02-18	Manhattan, NY	8:00PM	Apollo Rooms	Subscription Season	1842-43	New York Philharmonic
1843-04-07	Manhattan, NY	8:00PM	Apollo Rooms	Special	1842-43	Musicians from the New York Philharmonic
1843-04-22	Manhattan, NY	8:00PM	Apollo Rooms	Subscription Season	1842-43	New York Philharmonic
1843-11-18	Manhattan, NY	None	Apollo Rooms	Subscription Season	1843-44	New York Philharmonic
1844-01-13	Manhattan, NY	8:00PM	Apollo Rooms	Subscription Season	1843-44	New York Philharmonic
1844-03-16	Manhattan, NY	None	Apollo Rooms	Subscription Season	1843-44	New York Philharmonic
1844-05-18	Manhattan, NY	None	Apollo Rooms	Subscription Season	1843-44	New York Philharmonic
1844-11-16	Manhattan, NY	8:00PM	Apollo Rooms	Subscription Season	1844-45	New York Philharmonic
1845-01-11	Manhattan, NY	8:00PM	Apollo Rooms	Subscription Season	1844-45	New York Philharmonic

```
17
18 • select * from solistasconciertos;
19 • select * from conciertos;
```

soloistInstrument	soloistName	soloistRoles	season	orchestra
Soprano	Otto, Antoinette	S	1842	New York Philharmonic
Piano	Scharfenberg, William	A	1842	New York Philharmonic
Violin	Hill, Ureli Corelli	A	1842	New York Philharmonic
Viola	Derwort, G. H.	A	1842	New York Philharmonic
Cello	Boucher, Alfred	A	1842	New York Philharmonic
Contrabass	Rosier, F. W.	A	1842	New York Philharmonic
Soprano	Otto, Antoinette	S	1842	New York Philharmonic
Tenor	Horn, Charles Edward	S	1842	New York Philharmonic
Tenor	Horn, Charles Edward	S	1842	New York Philharmonic
Soprano	Otto, Antoinette	S	1842	New York Philharmonic

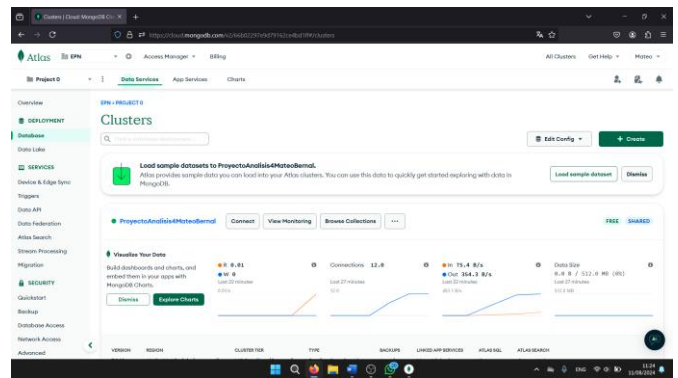


Nombre	Fecha de modificación	Tipo	Tamaño
hoy			
solistas_conciertos	11/08/2024 10:27	Archivo de origen ...	7.877 KB
juegos_olimpicos	11/08/2024 10:26	Archivo de origen ...	67.667 KB
conciertos	11/08/2024 10:22	Archivo de origen ...	3.951 KB

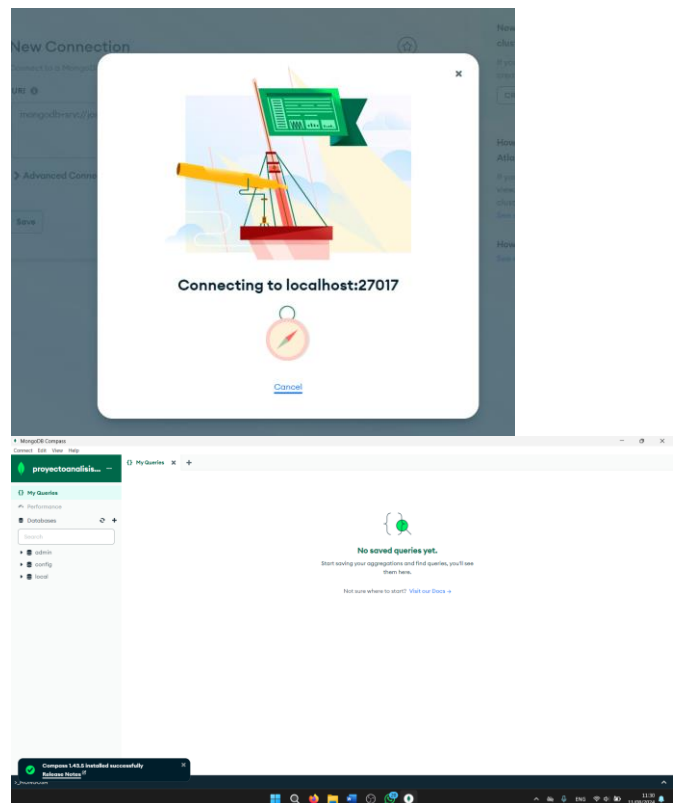
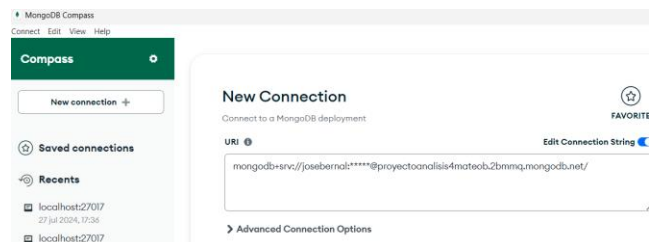


## Mongo Atlas

- Una vez creado un usuario y una conexión en mongo atlas vamos a realizar la conexión con nuestra base local de mongo compass.



- Se ingresa la url de la conexión de mongo atlas en mongo compass, para poder visualizar los datos de manera local y online.

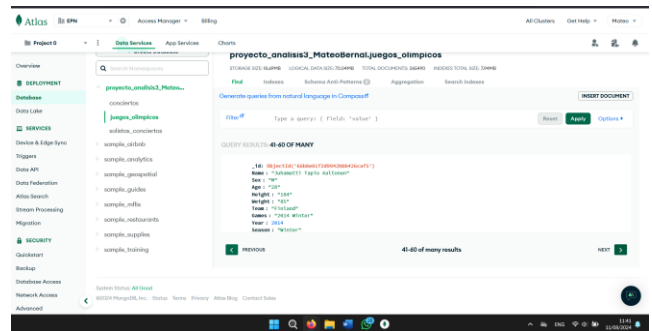
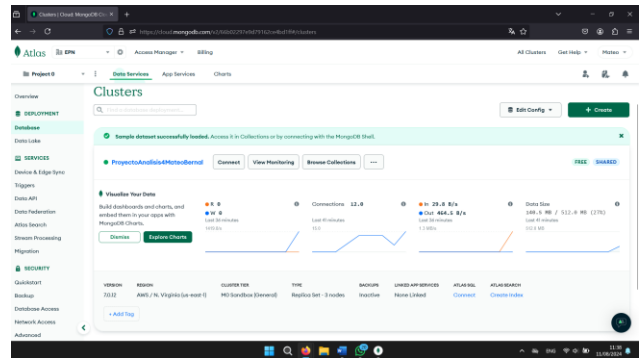
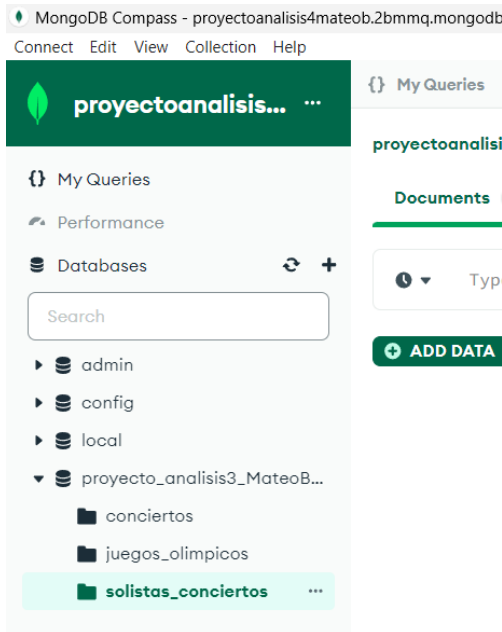


- Una vez conectada la base de datos, vamos a importar los archivos en formato json en mongo compass.

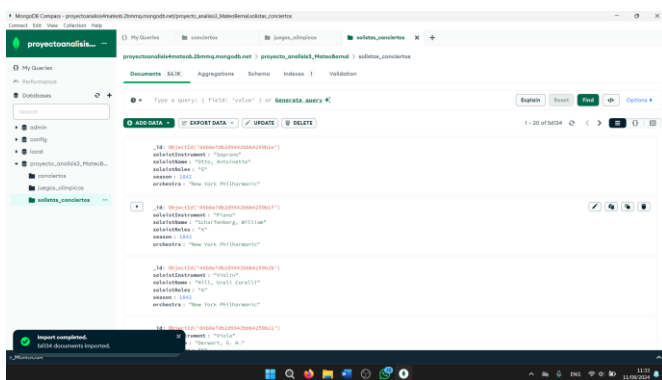
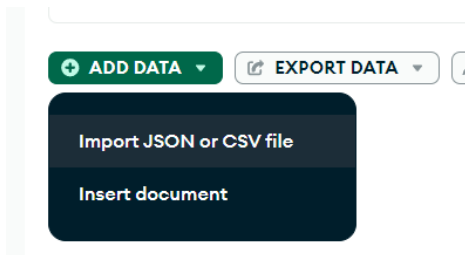
## Mongo Compass



- Creamos una base de datos en mongo compass, con diferentes apartados para mantener la integridad de los datos.

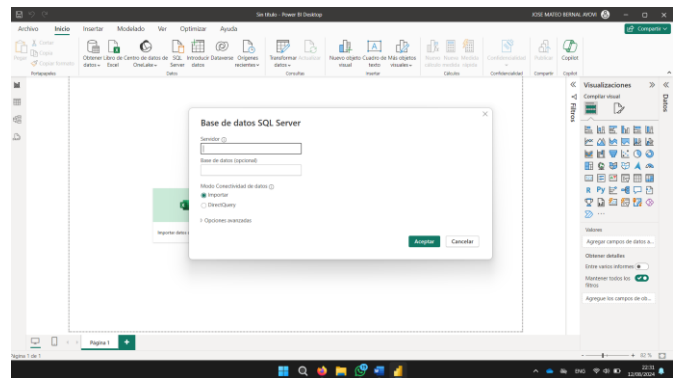


- Importamos los datos que están en formato json.

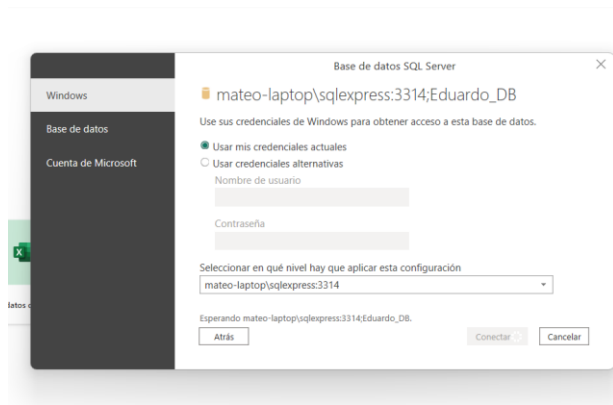


## Conexión Power BI

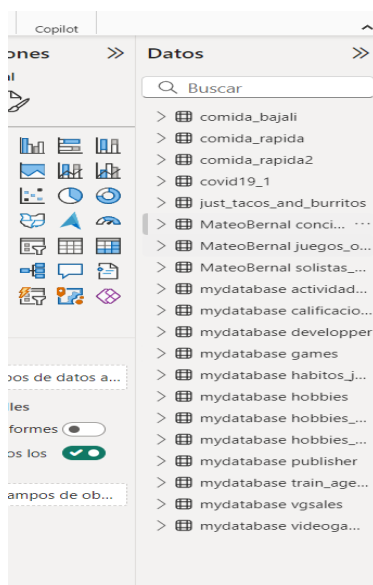
- Una vez en power BI, ingresamos las credenciales de nuestra base de datos de SQL server.



- Una vez importados los datos a mongo compass, los podemos visualizar desde mongo Atlas.



- Importamos las tablas de la base de datos, podremos visualizarlas desde power BI.



## 6. RESULTADOS Y DISCUSIÓN

En general, se recomienda que tanto el análisis como la discusión de resultados se presenten de manera integrada. Es necesario que los resultados queden clasificados dentro de arreglos significativos: tablas, cuadros, gráficos, diagramas, figuras, estadísticas, etc., que permitan reflejarla coherencia y la argumentación de estos. Cada representación de datos debe ir seguida de un texto que explique, interprete, contraste, amplíe... lo que los datos significan.

Los resultados cobran interés y valor cuando son interpretados a la luz de la teoría de base; presentada en la sección anterior. En sentido metafórico, los resultados, la teoría y la interpretación deben *dialogar* activamente en esta sección.

## 7. CONCLUSIONES Y RECOMENDACIONES

**Aplicación Integral de Conocimientos:** El proyecto permitió aplicar de manera integral los conocimientos adquiridos en la asignatura de Análisis de Datos. Desde la recolección y limpieza de datos hasta la creación de visualizaciones interactivas, se demostró la capacidad del equipo para manejar grandes volúmenes de información y presentarla de manera clara y efectiva.

**Importancia de la Consistencia de Datos:** A lo largo del proyecto, se destacó la importancia de la consistencia de los datos para obtener resultados precisos y confiables. El proceso de limpieza y preparación fue crucial para garantizar que los datasets fueran aptos para el análisis.

**Colaboración y Coordinación Efectiva:** La colaboración entre los miembros del equipo fue clave para el éxito del proyecto. La distribución clara de responsabilidades y la comunicación constante permitieron que cada miembro aportara significativamente al resultado final.

**Valor de la Visualización de Datos:** Las visualizaciones creadas en Power BI fueron fundamentales para interpretar los datos de manera intuitiva. Este enfoque facilitó la identificación de patrones y tendencias que podrían no ser evidentes en los datos en bruto.

**Relevancia del Análisis de Datos en Diversos Contextos:** El proyecto evidenció cómo el análisis de datos puede ser aplicado a diferentes áreas temáticas, desde eventos deportivos hasta ventas de videojuegos, subrayando su versatilidad y valor en diversos contextos.

**Mejorar la Recolección de Datos:** Para futuros proyectos, se recomienda ampliar las fuentes de datos y buscar datasets que no solo sean relevantes, sino también actualizados y de alta calidad, para enriquecer el análisis.

**Automatizar Procesos de Limpieza:** Implementar herramientas o scripts que automaticen la limpieza de datos puede mejorar la eficiencia y reducir errores humanos, facilitando un proceso más ágil y preciso.

**Explorar Herramientas Avanzadas de Análisis:** Se sugiere la incorporación de técnicas más avanzadas de análisis de datos, como el aprendizaje automático o el análisis predictivo, para obtener insights más profundos y valiosos.

**Optimizar la Arquitectura de Análisis:** Continuar mejorando la arquitectura de análisis utilizada, asegurando que sea escalable y eficiente, especialmente cuando se trabaja con grandes volúmenes de datos.

**Fomentar la Documentación Detallada:** Es crucial mantener una documentación detallada en todas las etapas del proyecto, lo que facilitará la revisión y comprensión del proceso, tanto para los miembros actuales como para futuros equipos.

**Incorporar Feedback Continuo:** Establecer mecanismos de retroalimentación continua durante el desarrollo del proyecto ayudará a identificar áreas de mejora y realizar ajustes a tiempo, asegurando un producto final de alta calidad.

