

Proyecto Final de Análisis de Datos

*CHANCUSIG CACHIMIL FABIAN DARWIN, VITERI MENA ALISSON MISHELL,
COBACANGO ACERO JOHN WALTER*

Desarrollo de Software - Escuela Politécnica Nacional

darwin.cachimil@epn.edu.ec
alisson.viteri01@epn.edu.ec
John.cobacango@epn.edu.ec

Resumen – El proyecto se enfoca en la integración y análisis de datos obtenidos de diversas fuentes, cubriendo cinco temáticas principales: eventos deportivos, actividades y hobbies, restaurantes y sitios de esparcimiento, tránsito, y eventos o noticias mundiales. Para asegurar la calidad de los datos, se empleó pandas en todas las etapas del proceso: limpieza, importación y exportación de datasets. Los datos depurados fueron importados a sistemas de gestión de bases de datos como MongoDB Atlas, MySQL, SQLite y MongoDB Compass. Posteriormente, los datos fueron cargados en Microsoft SQL Server en la nube. Finalmente, se crearon dashboards en Microsoft Power BI para facilitar la toma de decisiones mediante un análisis visual de la información.

Índice de Términos – Datasets, Pandas, Microsoft Power BI, Mongo Atlas, Microsoft SQL Server, MySQL, SQLite, Mongo DB Compass, SQL Server, Sistema Gestor de Base de Datos (SGBD), dashboards.

I. DEFINICIÓN DE LOS CASOS DE ESTUDIO

Caso de Estudio Eventos deportivos a nivel mundial.

Este caso de estudio examina los Juegos Olímpicos utilizando dos conjuntos de datos. El primero ofrece información general sobre cada edición de los juegos, incluyendo medallas, comités, disciplinas, eventos, y ciudades anfitrionas. El segundo dataset detalla a los atletas, incluyendo nombre, sexo, edad, altura, peso, país, medalla ganada, y eventos. La integración de estos datos proporciona una visión completa del desempeño y la organización de los eventos deportivos globales.

Caso de Estudio Actividades y hobbies.

El caso de estudio se encuentra como una actividad y hobbies a la danza en el cual se analizan: tipos de danza, estilos, origen, periodo, características, instrumentos, eventos y festivales, adaptaciones modernas, géneros, entre otros aspectos.

Caso de Estudio Tránsito en el Ecuador

Este caso de estudio analiza los incidentes de tránsito en el Ecuador, considerando factores como provincia, cantón, mes, día, zona, hora, clase de accidente, causa y número de fallecidos, víctimas y lesionados.

Caso de Estudio Restaurantes y sitios de esparcimiento

Este caso de estudio analiza el sector de restaurantes y sitios de esparcimiento, utilizando dos conjuntos de datos. El primer dataset recolecta información de la aplicación Zomato que proporciona detalles de los lugares, países, servicios y ranking de los restaurantes a nivel mundial. El segundo dataset proporciona información sobre el tipo de restaurante, ranking, costos y dirección del local.

Para el apartado sitios de esparcimiento, tenemos un dataset que proporciona información sobre los principales sitios turísticos visitados a nivel mundial, la cantidad de turistas y las aerolíneas usadas.

Caso de Estudio Eventos o noticias mundiales

Este caso de estudio se centra en el análisis de eventos y noticias internacionales utilizando dos conjuntos de datos. El primer dataset hace referencia a al conflicto bélico Ucrania/Rusia, donde se recolectan datos sobre sitios estratégicos, fallecidos, ubicaciones, etc. El segundo dataset se refiere a la pandemia de covid-19 en Estados Unidos, principalmente en los hospitales que atendieron la emergencia, la cantidad de camas de camas disponible y fallecimientos, etc.

II. OBJETIVOS DEL PROYECTO

Caso de Estudio Eventos deportivos a nivel mundial

Objetivo General

Analizar los eventos deportivos a nivel mundial, enfocándose en los Juegos Olímpicos, mediante la integración y análisis de datos relacionados con las ediciones de los juegos y los atletas participantes, con el fin de comprender mejor su impacto y relevancia en el ámbito deportivo global.

Objetivos Específicos

- Obtener información sobre las diferentes ediciones de los Juegos Olímpicos, incluyendo detalles como la cantidad de medallas, los eventos realizados, los atletas participantes y las ciudades anfitrionas, utilizando múltiples fuentes de datos.
- Preparar y depurar los datos recopilados para asegurar su calidad y coherencia, corrigiendo errores y completando cualquier información faltante para mejorar la precisión del análisis.
- Elaborar visualizaciones en Power BI que ilustren la distribución de medallas entre países y disciplinas, proporcionando una visión clara del rendimiento global a través de diferentes ediciones.

Caso de Estudio Actividades y hobbies.

Objetivo General

Explorar y comprender la diversidad y evolución de la danza como actividad y hobby, utilizando la recolección de datos sobre tipos de danza, estilos, origen, periodos históricos, características, instrumentos, eventos y festivales, adaptaciones modernas y géneros, para captar su impacto y relevancia en la cultura actual.

Objetivos Específicos

- Estructurar y organizar los datos relacionados con eventos, festivales y géneros de danza en una base de datos adecuada, facilitando la extracción y el acceso eficiente a la información.
- Aplicar técnicas de limpieza de datos con Pandas, incluyendo la eliminación de duplicados, el manejo de valores faltantes y la corrección de errores en el DataFrame, para garantizar la integridad y precisión de la información antes de su análisis y visualización en Power BI.

- Crear visualizaciones y dashboards en Power BI que representen la evolución de los estilos de danza, eventos relevantes y adaptaciones modernas, proporcionando una visión clara y accesible de las tendencias actuales.

Caso de Estudio Tránsito en el Ecuador

Objetivo General

Explorar el dataset de incidentes de tránsito en Ecuador para identificar patrones y tendencias relacionados con factores como provincia, cantón, mes, día, zona, hora, clase de accidente, causa, y el número de fallecidos, víctimas y lesionados, con el fin de proporcionar recomendaciones para mejorar la seguridad vial y la gestión del tránsito.

Objetivos Específicos

- Integrar el dataset de incidentes de tránsito en Ecuador a un gestor de base de datos, asegurando que la información se almacene correctamente para su análisis
- Establecer métodos de depuración y ajuste de datos para corregir errores, manejar valores ausentes y mejorar la calidad de la información antes de su exportación para análisis avanzado.
- Configurar la conexión entre SQL Server en la nube y el dataset, y elaborar un dashboard en Power BI que represente visualmente los incidentes de tránsito, facilitando la identificación de patrones y áreas de mejora en la seguridad vial.

Caso de Estudio Restaurantes y sitios de esparcimiento

Objetivo General

Evaluar el sector de restaurantes y sitios de esparcimiento a nivel mundial mediante el análisis de los conjuntos de datos de la aplicación Zomato y sobre sitios turísticos, con el propósito de identificar tendencias en la clasificación, costos, servicios, y popularidad de restaurantes y lugares de esparcimiento, así como en el flujo de visitantes y el uso de aerolíneas, para ofrecer recomendaciones que optimicen la experiencia y la gestión en estos sectores.

Objetivos Específicos

- Analizar las tendencias de clasificación y popularidad de restaurantes y sitios de esparcimiento a nivel mundial mediante los datos de Zomato y sitios turísticos.

- Detectar patrones en la clasificación y popularidad de restaurantes y sitios de esparcimiento globalmente mediante el análisis de datos en dashboards de Power BI, para captar tendencias actuales

- Exportar datos procesados y analizados desde gestores de bases de datos a dashboards en Power BI para facilitar la visualización y el análisis de tendencias en la clasificación, costos y popularidad de restaurantes y sitios de esparcimiento.

Caso de Estudio Eventos o noticias mundiales

Objetivo General

Analizar el impacto y la gestión de dos eventos internacionales críticos utilizando conjuntos de datos específicos: el conflicto bélico entre Ucrania y Rusia y la pandemia de COVID-19 en Estados Unidos. Este análisis busca identificar patrones en la respuesta a estas crisis y evaluar cómo la información sobre sitios estratégicos y capacidad hospitalaria influyó en la gestión de cada situación, con el fin de proporcionar recomendaciones para mejorar la respuesta ante futuras emergencias globales.

Objetivos Específicos

- Integrar los conjuntos de datos sobre el conflicto bélico y la pandemia en gestores de bases de datos como MongoDB, MySQL, SQLite y Azure para un análisis eficiente.
- Eliminar errores y preparar los datos para asegurar su calidad, consistencia y adecuación para análisis y visualización en Power BI.
- Exportar los datos procesados y crear informes y visualizaciones en MongoDB Compass, Azure y Power BI para resumir y presentar los dashboards adecuados.

III. DESCRIPCIÓN DEL EQUIPO Y ACTIVIDADES

Para llevar a cabo el proyecto, el equipo acordó asignar tareas específicas a cada miembro, basándose en sus habilidades y conocimientos. Esto garantiza que cada tarea sea realizada por quien tiene la experiencia necesaria, contribuyendo a un desarrollo más fluido y organizado del proyecto.

Actividades realizadas por cada integrante

Las actividades asignadas a cada miembro del equipo son las siguientes:

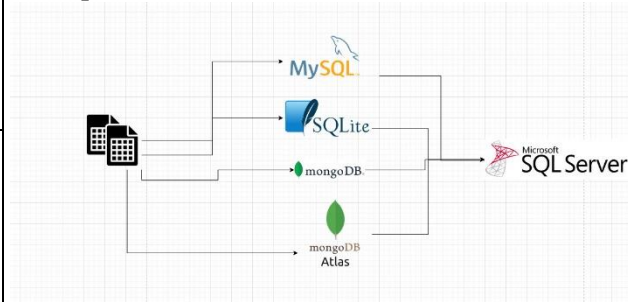
Integrante	Actividades
Walter Cobacango	<ul style="list-style-type: none"> - Investigación y recopilación de los dataset para el caso de estudio 4 y 6. - Limpieza de datos con Pandas. - Importación y exportación de datos en Mongo DB, Atlas, MySQL, SQLite y Azure.

	<ul style="list-style-type: none"> - Conexión de Azure con Power BI. - Dashboard de los respectivos casos en Power BI. - Informe - Video de la conexión - Creación del repositorio en github. - Presentación
Darwin Cachimil	<ul style="list-style-type: none"> - Investigación y recopilación de los dataset para el caso de estudio 7 y 8. - Limpieza de datos con Pandas. - Importación y exportación de datos en Mongo DB, Atlas, MYSQL, SQLite. - Dashboard de los respectivos casos en Power BI. - Informe - Video de las conclusiones del dashboard. - Presentación
Alisson Viteri	<ul style="list-style-type: none"> - Investigación y recopilación de los dataset para el caso 2. - Limpieza de datos con Pandas. - Importación y exportación de datos en Mongo DB, Atlas, MYSQL, SQLite. - Dashboard de los respectivos casos en Power BI. - Informe - Presentación

MongoClient, pandas, sqlite3, sqlalchemy, create_engine, mysql.connector, IPython.display, entre muchas otras más)

VI. ARQUITECTURA DE LA SOLUCIÓN

Arquitectura



Recolección e Importación de Datos

Como paso inicial se realizó la búsqueda de las bases de datos necesarias para el proyecto; se escogió los data sets con la ayuda de diferentes sitios web y herramientas como: Kaggle, Google Dataset, data.world, webscraping entre otros. A continuación, se realizó la importación de los distintos datasets, transformados al tipo de archivo permitidos por las bases de datos las cuales fueron escogidas en una reunión de equipo las cuales son: MySQL, Microsoft SQL Server, SQLite, MongoDB, Mongo Atlas.

```
[1]: import pandas as pd
    from pymongo import MongoClient

[2]: # Conectar a la base de datos de MongoDB Atlas
    url = "mongodb+srv://Nalter:WalyFox22@cluster0.p2y1kw.mongodb.net/ANALISIS?retryWrites=true&majority=asynchronous"
    client = MongoClient(url)

    # Seleccionar la base de datos
    db = client['ANALISIS']

    # Seleccionar la colección
    collection = db['Transito']

    # Leer los documentos de la colección
    data = list(collection.find())

[3]: # Cargar los datos en un DataFrame de Pandas
    df = pd.DataFrame(data)

[6]: # Mostrar el DataFrame
    df
```

Gráfico 2 Conversión del Dataset al tipo de archivo permitido por el gestor de base de datos para su importación y extracción del SGBD

Extracción de Datos

Para este proceso se debe realizar la conexión a la base de datos respectiva utilizando el lenguaje de programación Python e importando las librerías necesarias con el fin de realizar la extracción de datos, los cuales están contenidos en tablas por cada dataset, es decir, un dataset pertenece a una tabla independiente de la base de datos final que es el SQL SERVER en la nube de Azure. Este gestor nos sirve como repositorio para posteriormente la conexión con PowerBI.

IV. CRONOGRAMA DE ACTIVIDADES

Actividades	Duración de la Planificación											
	Julio				Agosto							
	29	30	31	6	7	8	9	10	11	12		
Preparación de herramientas para la recolección de datos												
Selección de datasets												
Limpieza de datos												
Importación de datos a los gestores de BD												
Exportación de los datos de los gestores de BD												
Conexión Azure y Power BI												
Importación de datos a Azure												
Elaboración de visualizaciones con su respectivo análisis												
Elaboración de la documentación, presentación y videos												
Subir el proyecto al repositorio de github												

Gráfico 1 Cronograma de Actividades

V. RECURSOS O HERRAMIENTAS UTILIZADAS

Las Herramientas y Recursos utilizados para la creación de los dashboards fueron MongoDB Compass, MongoDB Atlas, MySQL, Microsoft SQL Server, Azure de SQL Server, SQLite Browser, Jupyter notebook, Librerías de Python (pymongo,

Introduction					
Specify Input File					
Preview Data					
This operation analyzed the input file structure to generate the preview below for up to the first 50 rows.					
Rank	Restaurant	Content	Sales	YOY_Sales	Units
1	McDonald's		40412	4.9%	13846
2	Starbucks		21380	8.6%	15049
3	Chick-fil-A	While Popeyes...	11320	13.0%	2470
4	Taco Bell		11293	9.0%	6766
5	Burger King		10204	2.7%	7346
6	Subway		10200	-2.0%	23801
7	Wendy's		9762	4.2%	5852
8	Dunkin'	Dunkin' droppe...	9228	5.0%	9630
9	Dominos		7044	6.9%	6126
10	Panera Bread		5890	4.0%	2160
11	Pizza Hut	The chain mov...	5558	0.6%	7306
12	Chipotle Mexic...		5509	14.8%	2584
13	Sonic Drive-In	The drive-in ch...	4687	4.6%	3526
14	KFC		4546	2.5%	4065
15	Olive Garden	Olive Garden b...	4287	5.0%	866
16	Applebee's		4085	-3.0%	1665

Gráfico 9 Identificación de los campos vacíos y nulos del dataset

Conexión y envío de la Información al SQL Server en Azure

Se inicia la conexión con Jupyter notebook, que en este contexto actuará como el punto central para exportar los datos depurados de los diversos temas seleccionados para este proyecto. La información primero se lee, el cual está organizado según los temas y se separa para cada uno, por medios de librerías se importa a la Sql server ingresando las credenciales que posee la base.

```

1: import pandas as pd
from sqlalchemy import create_engine

2: import pyodbc
from azure.identity import DefaultAzureCredential

# Configuración de la conexión
server = 'projectoanalysis.database.windows.net'
database = 'Projecto'
username = 'john.cobacang@epn.edu.ec' # Tu correo electrónico
driver = 'ODBC Driver 17 for SQL Server'

# Obtener token de acceso
credential = DefaultAzureCredential()
token = credential.get_token("https://database.windows.net/.default").token

# Crear la conexión usando el token de acceso
conn = pyodbc.connect(
    f'DRIVER={driver};SERVER={server};PORT=1433;DATABASE={database};UID={username};Authentication=ActiveDirectoryAccessToken;',
    attrs=({'AccessToken': token})
)

cursor = conn.cursor()

```

Gráfico 10 Conexión y envío de los Datasets al SQL Server

```

[ ]: import json

# Ruta al archivo JSON
file_path = r'C:\Temp\independence_rankings.json'

# Leer el archivo JSON
with open(file_path, 'r') as file:
    data = json.load(file)

# Crear la tabla si no existe
create_table_query = '''
CREATE TABLE independence_rankings (
    Rank VARCHAR(100) PRIMARY KEY,
    Restaurant VARCHAR(400),
    Sales INT,
    Average_Check INT,
    City VARCHAR(100),
    State VARCHAR(50),
    Meals_Served FLOAT
)
'''

cursor.execute(create_table_query)
conn.commit()

```

Gráfico 11 Creación de tablas para el almacenar en la base de datos

```

# Insertar datos en la tabla
for item in data:
    cursor.execute('''
INSERT INTO independence_rankings (Rank, Restaurant, Sales, Average_Check, City, State, Meals_Served)
VALUES (?, ?, ?, ?, ?, ?, ?)
''')
    item['Rank'], item['Restaurant'], item['Sales'], item['Average_Check'], item['City'], item['State'], item['Meals_Served'])

conn.commit()

# Consultar la tabla para verificar los datos
cursor.execute('SELECT * FROM independence_rankings')
rows = cursor.fetchall()

for row in rows:
    print(row)

```

Gráfico 12 Inserción de los datasets a la SQL server

Rank	Restaurant	Content	Sales	YOY_Sales	Units
1	McDonald's		40412	4.9%	13846
2	Starbucks		21380	8.6%	15049
3	Chick-fil-A	While Popeyes...	11320	13.0%	2470
4	Taco Bell		11293	9.0%	6766
5	Burger King		10204	2.7%	7346
6	Subway		10200	-2.0%	23801
7	Wendy's		9762	4.2%	5852
8	Dunkin'	Dunkin' droppe...	9228	5.0%	9630
9	Dominos		7044	6.9%	6126
10	Panera Bread		5890	4.0%	2160
11	Pizza Hut	The chain mov...	5558	0.6%	7306
12	Chipotle Mexic...		5509	14.8%	2584
13	Sonic Drive-In	The drive-in ch...	4687	4.6%	3526
14	KFC		4546	2.5%	4065
15	Olive Garden	Olive Garden b...	4287	5.0%	866
16	Applebee's		4085	-3.0%	1665

Gráfico 13 Inserción de los datasets a la SQL server

Nota Importante

La conexión de la base de datos SQL Server no se logró tras varios intentos ya que como el proyecto se lo llevo a cabo en Azure for Students, el ingreso de las credenciales eran rechazadas constantemente y no se permitía la conexión y por consiguiente el subir los datasets verificados. En las imágenes colocadas se represente el cómo debió de ser la conexión y exportación de datos SQL Server.

Visualización de Información

Para empezar con la visualización de los datos vamos a iniciar Power BI Desktop en nuestro equipo y procedemos a logearnos; una vez hecho eso empezamos un informe en blanco y en la barra de opciones vamos hacia la opción “Obtener Datos”, posteriormente se nos abre una ventana en la cual buscamos “Azure sql”

Obtener datos

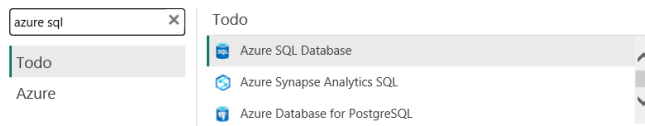


Gráfico 14 Obtener datos

Una vez escogida la fuente de los datos procedemos a conectarnos y llenar la información requerida en la ventana pertinente, en esta ventana colocamos el enlace y el nombre de nuestra base de datos sql que contiene todos los archivos requeridos. El enlace como tal sirve como puente entre power BI y Azure para la correcta compartición de datos.

Base de datos SQL Server

Servidor

Base de datos (opcional)

Modo Conectividad de datos ☐ Importar ☒ DirectQuery

☐ Opciones avanzadas

Aceptar

Cancelar

Gráfico 15 Ventana de conexión

Llenamos todos los datos que garantizan la conexión hacia nuestros archivos y finalmente procedemos a cargarlos, para lo cual se nos despliega una ventana que permite ver los datasets almacenados en la base de datos, esta ventana temporal nos deja escoger y limpiar los datos, si estamos conformes con los mismo lo cargamos hacia producción.

De esta manera tenemos una correcta integración entre estos dos ecosistemas “Azure” y “Power BI” permitiéndonos una correcta visualización de nuestros datos y posterior manejo de estos mismos.

En el siguiente apartado de imágenes se puede observar el detalle de los datasets importados hacia Power Bi.

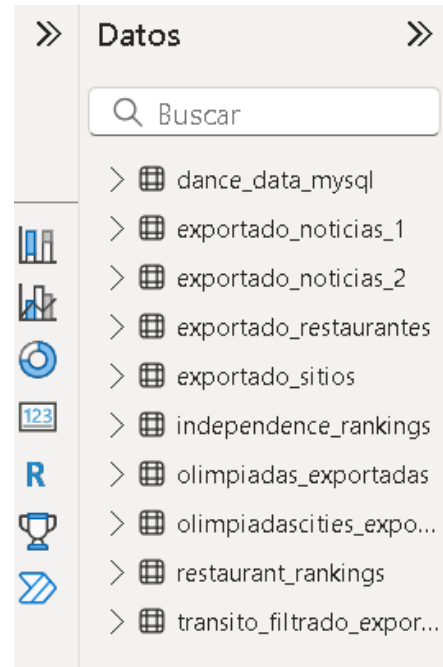
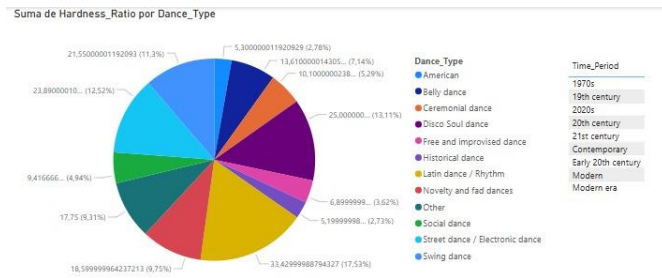


Gráfico 16 Visualización de datos

EXPORT DATA											Details
EXPORT ID	EXPORT DATE	YEAR	IMPORTED TYPE	EXPORT TYPE	SUB EXPORT TYPE	ACTOR1	ACTOR2	ACTOR3	ACTOR4	ACTOR5	
1902180	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902181	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902182	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902183	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902184	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902185	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902186	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902187	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902188	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902189	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902190	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902191	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902192	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902193	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902194	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902195	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902196	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902197	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902198	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902199	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902200	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902201	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902202	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902203	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902204	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902205	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902206	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902207	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902208	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902209	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902210	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902211	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902212	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902213	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902214	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902215	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902216	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902217	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902218	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902219	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902220	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902221	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902222	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902223	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902224	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902225	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902226	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902227	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902228	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902229	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902230	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902231	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902232	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902233	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902234	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902235	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902236	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902237	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902238	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902239	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902240	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902241	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902242	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902243	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902244	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902245	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902246	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902247	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902248	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902249	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902250	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902251	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902252	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902253	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902254	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902255	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902256	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902257	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902258	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902259	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902260	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902261	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902262	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902263	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902264	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902265	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902266	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902267	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902268	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902269	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902270	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902271	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902272	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902273	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902274	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902275	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902276	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902277	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902278	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902279	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902280	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902281	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902282	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902283	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902284	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902285	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows	Full disk volume	Full disk volume	Full disk volume	Full disk volume	
1902286	2020-12-29 00:00:00	2020	Full disk volume	Full disk volume	Full disk volume	Exported from: Windows					

Resultados Obtenidos

Caso de Estudio Tipos de Baile a Nivel Mundial



Caso de estudio 1. En la siguiente gráfica se presenta la relación y avance que han tenido los diferentes tipos de baile a lo largo del mundo por su grado de popularidad en los diferentes tipos.

Gráfico 18 Gráficos de resultados obtenidos del caso de estudio Actividades

En este caso se usó un diagrama pastel para representar la relación y avance que han tenido los diferentes tipos de baile a lo largo del mundo medidos por su popularidad.

Caso de Estudio Conflicto Ucrania/Rusia

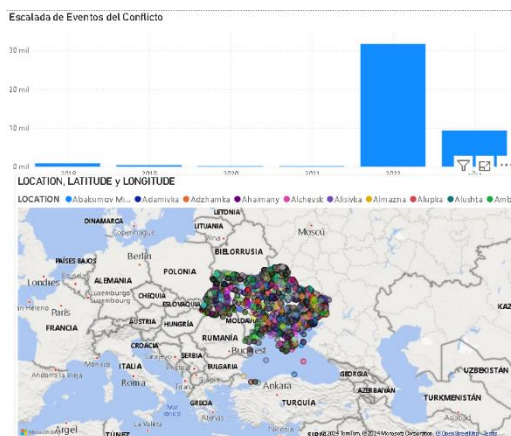


Gráfico 19 Gráficos de resultados obtenidos del caso de estudio Noticias

En este caso examinamos la escalada de eventos conflictivos entre Rusia y Ucrania, teniendo como mayor afectado del mismo a Ucrania. Se puede ver de forma gráfica los principales sitios inmersos en el conflicto, que han sido los principales escenarios de batallas o algún tipo de atentado. También observamos los años en que este tipo de eventos se desencadenaron con mayor frecuencia.

Caso de Estudio Covid-19 en Estados Unidos

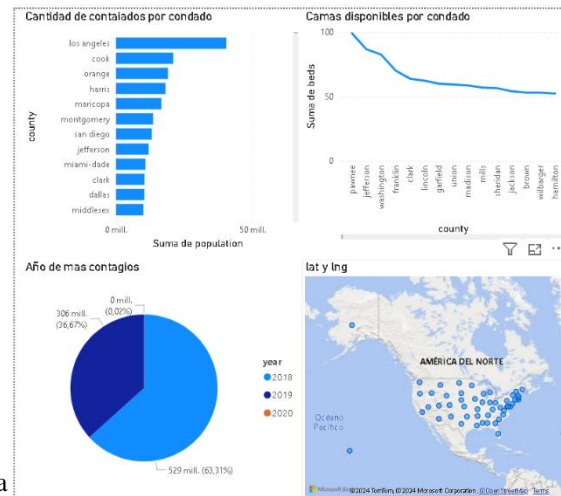


Gráfico 20 Gráficos de resultados obtenidos del caso de estudio Noticias

Estos datos nos permiten cotejar el impacto que tuvo el covid-19 en Estados Unidos. Los sitios con más contagios, la cantidad de camas disponibles en cada condado para atender la emergencia, los años donde se produjo el mayor número de contagios, y un mapa que permite visualizar los principales sitios de contagios.

Caso de Estudio análisis de restaurantes según Zomato

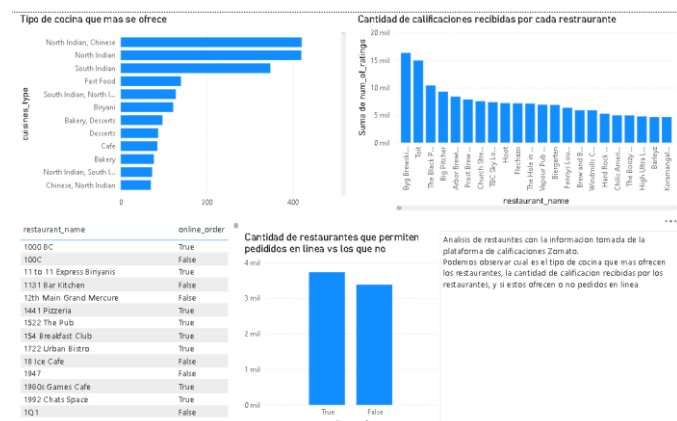


Gráfico 21 Gráficos de resultados obtenidos del caso de estudio Restaurantes

Análisis de restaurantes con la información tomada de la plataforma de calificaciones Zomato. Podemos observar cual es el tipo de cocina que más ofrecen los restaurantes, la cantidad de calificación recibidas por los restaurantes, y si estos ofrecen o no pedidos en línea

Caso de Estudio Análisis Principales destinos turísticos a nivel mundial.



Gráfico 22 Gráficos de resultados obtenidos del caso de estudio sitios esparcimiento

Observamos los sitios turísticos más visitados por los turistas a nivel mundial, las principales agencias contratadas, las preferencias de los turistas entre agencias de viajes o vuelos directos, y la edad de los turistas que más viajan.

Caso de Estudio Top restaurantes en América del Norte

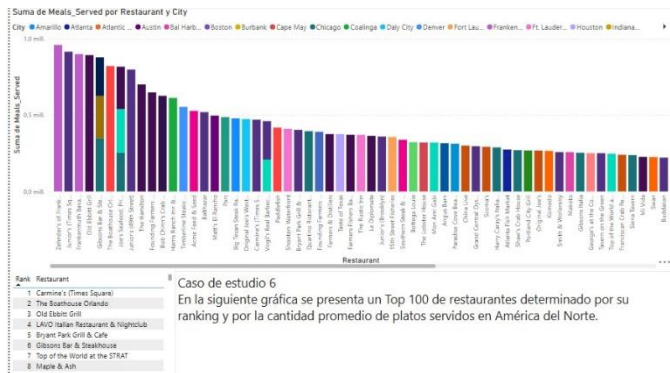


Gráfico 23 Gráficos de resultados obtenidos del caso de estudio Restaurantes

En este caso de estudio se un gráfico de tipo barras perpendiculares para denotar de mejor manera el top 100 de restaurantes cuyos valores han sido determinados por su ranking y por la cantidad promedio de sus platos servidos en América del Norte.

Caso de Estudio Eventos terroristas

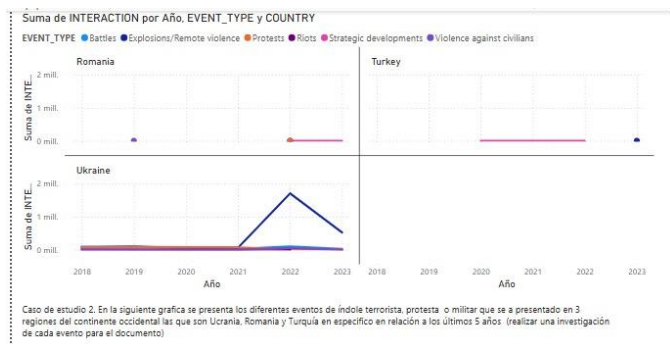
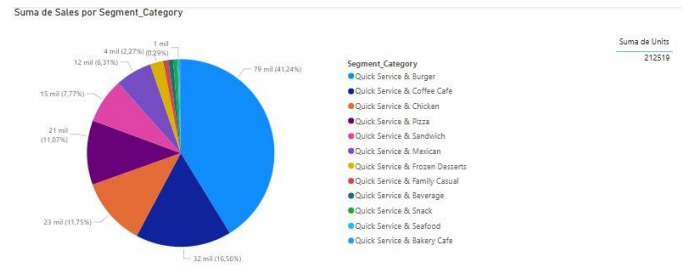


Gráfico 24 Gráficos de resultados obtenidos del caso de estudio sitios Noticias

En la siguiente grafica se presentan los diferentes eventos de índole terrorista, protesta o militar que se han presentado en 3

regiones del continente occidental como son Ucrania, Romania y Turquía en estos últimos 5 años.

Caso de Estudio Top restaurantes Quick service



Caso de estudio 9

En la siguiente tabla se visualiza toda los restaurantes que pertenecen a la categoría de Quick Service en Norte América tomando en cuenta la suma de sus ventas. Además se visualiza la suma de las unidades disponibles.

Gráfico 24 Gráficos de resultados obtenidos del caso de estudio sitios Restaurantes

En el grafico pastel se visualiza todos los restaurantes que pertenecen a la categoría de Quick Service en Norte América tomando en cuenta la suma de sus ventas. También se puede observar la suma de las unidades disponible.

Caso de Estudio análisis de medallas obtenidas



Gráfico 25 Gráficos de resultados obtenidos del caso de estudio Deportes

El presente caso de estudio es un análisis de las medallas obtenidas en las Olimpiadas. El mapa muestra la distribución de medallas por país, incluyendo oro, plata y bronce. Se filtro, además, los datos de los países ganadores de medallas de oro. Así podemos identificar los países con mayor éxito en las competiciones olímpicas.

Caso de Estudio países que más participaron en las Olimpiadas



Gráfico 26 Gráficos de resultados obtenidos del caso de estudio Deportes

La grafica usada muestra los cinco países que más participaron en deportes durante las Olimpiadas, reflejando así su capacidad para este tipo de eventos importantes.

Caso de Estudio análisis país que compite en más disciplinas.



Gráfico 27 Gráficos de resultados obtenidos del caso de estudio Deportes

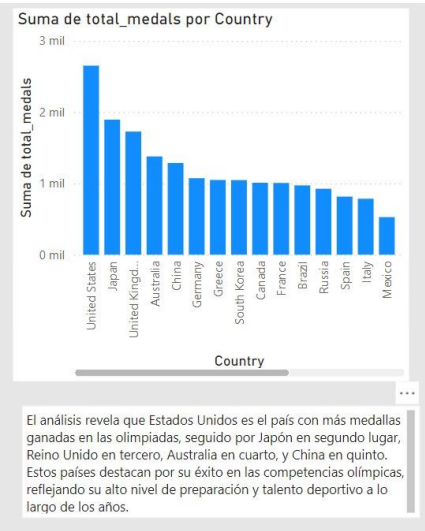


Gráfico 28 Gráficos de resultados obtenidos del caso de estudio Deportes

El análisis muestra que Estados Unidos es el país que más ha competido en más disciplinas en las Olimpiadas, con 143 disciplinas en total, seguido por Japón. De igual manera se puede observar en el gráfico de barras perpendiculares que Estados Unidos es el país con más medallas ganadas.

VII. CONCLUSIONES

-El proyecto nos permitió demostrar que se puede gestionar grandes volúmenes de datos provenientes de diversas fuentes y su consolidación en un repositorio, así de esta manera se evidencio la importancia de una arquitectura planificada para

manejar datos.

- El proyecto fue un esfuerzo de todo el equipo en las distintas etapas del proceso, desde la recolección y limpieza de datos hasta la visualización final. Este enfoque mejoró la eficiencia del trabajo, habilidades y conocimientos de cada uno.

- El uso de la librería Pandas en Python para la limpieza y transformación de datos fue crucial para garantizar la calidad de los datasets, de esta manera se minimizo errores y se mejoró la precisión de los resultados.

VIII. DESAFIOS Y PROBLEMAS ENCONTRADOS

- Al crear dashboards en Power BI, algunos datos resultaron difíciles de representar de manera clara y comprensible, especialmente cuando se trataba de relaciones complejas.
- La conexión y sincronización entre diferentes sistemas de bases de datos y la unión de todos los datasets en SQL server fueron complicadas por la diferencia en los esquemas de datos y las capacidades y limitaciones de cada base de datos.
- La calidad de los datos era variable, con datos incompletos, duplicados o inconsistentes en varios conjuntos de datos.

IX. LINKS DEL PROYECTO

Link de las Visualizaciones PowerBI:

<https://app.powerbi.com/groups/me/reports/d297a714-9f40-408e-b202-a337e3395714/af2a7357bafb4c890ea6?experience=power-bi>

Link de Github:

https://github.com/WalterFox22/Proyecto_Final_Analisis.git

Link de Video del Proyecto:

- <https://www.youtube.com/watch?v=WKTbBF3o2F4>
- <https://youtu.be/OW7xXfpJBUE>