

Proyecto Final de Análisis de Datos

Eduardo Caza, Mateo Garzon, Luis Guaygua, Lenin Proaño, Yuverly Verdezoto

Escuela Politécnica Nacional

Eduardo.caza@epn.edu.ec

Mateo.garzon@epn.edu.ec

Luis.guaygua@epn.edu.ec

Lenin.proaño@epn.edu.ec

Yuverly.verdezoto@epn.edu.ec

I. Definición del Caso de estudio

Análisis de Ganancias: Jugadores, Equipos y Videojuegos a Nivel Global

Evaluación del Valor: Clasificación de Eventos Deportivos a Escala Mundial

Estudio de Mortalidad: Tasas de Enfermedades Prevalentes a Nivel Internacional

Impacto del Consumo de Comida Rápida en la Salud: Un Análisis Profundo

Exploración de los World Tours de Taylor Swift: Un Estudio Exhaustivo

II. Objetivos Generales y Específicos

Análisis de Ganancias: Jugadores, Equipos y Videojuegos a Nivel Global

Objetivo General

Realizar un análisis exhaustivo de las ganancias generadas por jugadores, equipos y videojuegos a nivel global, con el fin de identificar patrones, tendencias y oportunidades en el mercado de los videojuegos.

Objetivos Específicos

- Analizar las ganancias de los principales videojuegos a lo largo del tiempo, identificando aquellos que han tenido un mayor impacto en el mercado.

- Investigar las preferencias de los jugadores en cuanto a géneros de videojuegos, lo que, para comprender mejor el comportamiento del mercado.
- Identificar las tendencias emergentes en la industria de los videojuegos, como la realidad virtual,
- Proponer estrategias y recomendaciones para maximizar las ganancias de los jugadores, equipos y empresas de videojuegos

Exploración de los World Tours de Taylor Swift: Un Estudio Exhaustivo

Objetivo General

Analizar de manera integral los datos de las giras de Taylor Swift con el fin de obtener una vista de sus ventas en diferentes ciudades del mundo significativas de las giras, mediante gráficos implementados en Looker con un dataset elegido

Objetivos Específicos

Analizar la evolución de la asistencia y los ingresos a lo largo de las diferentes giras de Taylor Swift.

Evaluar la relación entre la asistencia y la ubicación geográfica de los conciertos.

Identificar los lugares que han demostrado ser más exitosos en términos de asistencia e ingresos.

Identificar oportunidades para mejorar la experiencia del espectador, ya sea a través de la selección de canciones, colaboraciones especiales u otros aspectos relevantes.

Impacto del Consumo de Comida Rápida en la Salud: Un Análisis Profundo

Objetivo General

Analizar el impacto en la salud derivado del consumo habitual de comida rápida, identificar específicamente aquellos establecimientos que ofrecen alimentos con alto contenido calórico, y utilizar estos hallazgos para fomentar la adopción de hábitos alimenticios más saludables en la población.

Objetivos Específicos

Identificar los principales establecimientos de comida rápida en la localidad y analizar su menú en términos de contenido calórico y nutricional.

Investigar la frecuencia de consumo de comida rápida en el establecimiento con mayor índice para conocer los alimentos con mayor demanda.

Evaluar las políticas actuales de los establecimientos de comida rápida en relación a la divulgación de información nutricional a sus clientes.

Analizar el impacto de las regulaciones gubernamentales en la oferta y demanda de comida rápida.

Evaluación del Valor: Clasificación de Eventos Deportivos a Escala Mundial

Objetivo General

Realizar un caso de estudio analítico que explore y compare la dificultad por habilidad en diferentes deportes, así como la valoración

económica de los equipos deportivos, con el fin de obtener insights significativos que contribuyan a la comprensión del panorama deportivo mundial

Objetivos Específicos

Analizar el dataset relacionado con la dificultad de los deportes por habilidad, considerando variables como el nivel de competencia, la complejidad técnica y la demanda física.

Identificar los deportes que se consideran más difíciles y aquellos que presentan diferencias significativas en términos de desafío para los participantes.

Explorar el dataset de valoración económica de equipos deportivos, teniendo en cuenta factores como ingresos, patrocinios, valor de mercado de jugadores y activos del equipo.

Clasificar los equipos deportivos según su valor económico y analizar las tendencias en la valoración, identificando posibles correlaciones con el rendimiento en competiciones y la popularidad global.

Estudio de Mortalidad: Tasas de Enfermedades Prevalentes a Nivel Internacional

Objetivo General

Analizar las tasas de mortalidad asociadas a enfermedades prevalentes a nivel internacional para comprender la distribución geográfica y las tendencias temporales, con el fin de identificar patrones y áreas de intervención prioritaria en salud pública.

Objetivos Específicos

Evaluar las variaciones geográficas en las tasas de mortalidad por enfermedades prevalentes, incluyendo diferencias regionales y nacionales.

Analizar las tendencias temporales en las tasas de mortalidad de las enfermedades estudiadas a lo largo de un periodo específico, identificando posibles cambios significativos.

Identificar los factores socioeconómicos y de salud pública que pueden influir en las tasas de mortalidad por enfermedades prevalentes, con el objetivo de proponer estrategias de prevención y control adecuadas en diferentes contextos.

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

Integrante	Actividades
Eduardo Caza	Impacto del Consumo de Comida Rápida en la Salud: Un Análisis Profundo
Mateo Garzon	Análisis de Ganancias: Jugadores y Videojuegos a Nivel Global
Luis Guaygua	Exploración de los World Tours de Taylor Swift: Un Estudio Exhaustivo
Lenin Proaño	Evaluación del Valor: Clasificación de Eventos Deportivos a Escala Mundial
Yuverly Verdezoto	Estudio de Mortalidad: Tasas de Enfermedades Prevalentes a Nivel Internacional

IV. Cronograma de actividades.

Semana	Fecha	Actividades
1	10/02/24	Selección de los casos de Estudio Definición del caso de estudio Descripción del equipo de trabajo Cronograma de Actividades

2	17/02/24	Arquitectura de la solución Extracción de datos Análisis de información
3	24/02/24	Visualización de los datos Resultados Obtenidos Conclusiones y Recomendaciones
4	02/03/24	Desafíos y problemas encontrados Subida del proyecto a GitHub.

V. Recursos y herramientas utilizadas.

Durante el desarrollo del proyecto, se emplearon una variedad de recursos y herramientas para llevar a cabo el análisis de los datos, estos fueron:

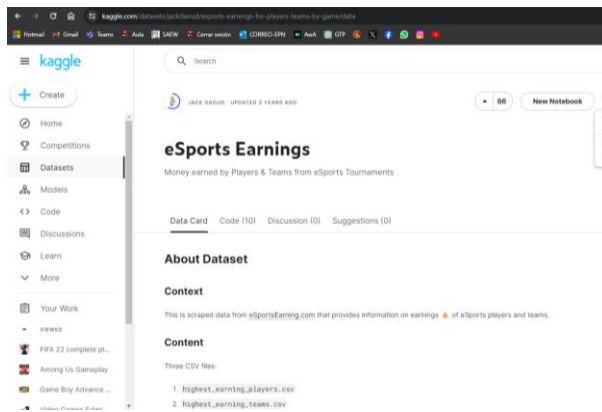
- Python
- Jupyter notebook
- Google Looker
- Power BI
- Mongoddb
- Canva
- Mysql
- Google database search
- kaggle

VI. Arquitectura de la solución.



VII. Extracción de datos.

Análisis de Ganancias: Jugadores, Equipos y Videojuegos a Nivel Global



Importamos los conjuntos de datos propuestos para el análisis.

```
[29]: import pandas as pd
ruta_archivo = "GM ciudad-y-continentes-codigos-lista.csv"
ruta_archivo2 = "GM jugadores-con-mayores-ganancias.csv"

[30]: dataframeciudades = pd.read_csv(ruta_archivo)
dataframejugadores = pd.read_csv(ruta_archivo2)
```

Cargamos los archivos.

```
] : #Limpiar datos
dataframe = dataframeciudades.dropna()
dataframe2 = dataframejugadores.dropna()

] : #Limpiar dataframe
dataframelimpio = len(dataframe)
dataframelimpio2 = len(dataframe2)

] : #Verificar Limpido
print("El largo del DataFrame es:", dataframelimpio)

El largo del DataFrame es: 214

] : print("El largo del dataframe es:", dataframelimpio2)

El largo del dataframe es: 1000
```

Limpiamos los datos duplicados y nulos

```
#Verificar Limpido
print("El largo del DataFrame es:", dataframelimpio)
El largo del DataFrame es: 214

print("El largo del dataframe es:", dataframelimpio2)
El largo del dataframe es: 1000

dataframe['Ciudad'] = dataframe['Ciudad'].str.split(',').str[0]
#LIMPIAR DATAFRAME CIUDAD

# Convertir los códigos de ciudad en dataframe2 a mayúsculas
dataframe2['codigociudad'] = dataframe2['codigociudad'].str.upper()

#Renombrar columnas para que coincidan
dataframe.rename(columns={'Ciudad-Code': 'codigociudad'}, inplace=True)

# Fusionar los DataFrames basados en el código de ciudad en mayúsculas
dataframefinal = pd.merge(dataframe2, dataframe[['codigociudad', 'Ciudad']], on='codigociudad', how='left')

dataframefinal = dataframefinal.dropna()
# Ahora dataframefinal no contendrá filas con valores nulos
print(dataframefinal.head())

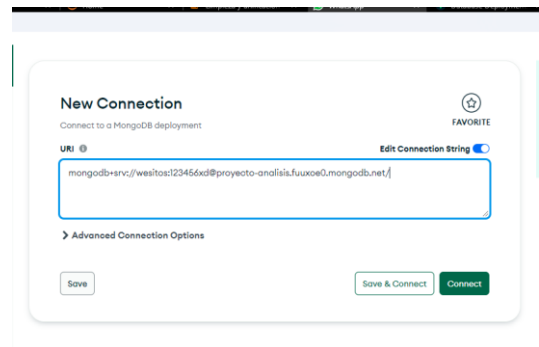
nombre_archivo_csv = "GM-Jugadores-Codigos.csv"
dataframefinal.to_csv(nombre_archivo_csv, index=False)
print("DataFrame guardado como CSV en la carpeta actual con el nombre:", nombre_archivo_csv)

DataFrame guardado como CSV en la carpeta actual con el nombre: GM-Jugadores-Codigos.csv
```

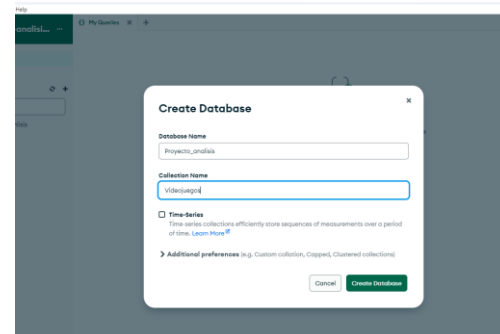
Combinamos los dos archivos CSV para realizar un análisis más preciso y, posteriormente, aplicamos la función dropna para eliminar columnas duplicadas.

Metodología

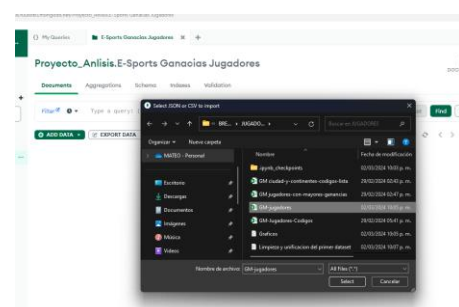
- Elección del csv a analizar: Herramienta usada Kaggle
- Limpieza de Datos: Eliminar datos incorrectos o incompletos. Con ayuda de pandas



Importamos los datos a MongoDB Atlas para unificar la información.



Creamos la base y la colección



Subimos el CSV de los jugadores

Import

To collection Proyecto_Analisis.E-Sports Ganacias Jugadores

Import file: GM-jugadores.csv

Options

Select delimiter: Comma

☒ Ignore empty strings

☐ Stop on errors

Specify Fields and Types

	idjugador	nombre	apellido	nickname	codigociudad
1	3883	Peter	Rasmussen	dupreeh	DK
2	3679	Andreas	Højsteth	Xyp9x	DK
3	3885	Nicolai	Reedtz	device	DK
4	3672	Lukas	Rosander	glove	DK
5	17800	Emil	Relf	Magisk	DK
6	12183	Epiðacio	de Melo	TACO	BR

Cancel Import

Verificamos que los datos que mantengan su integridad

Documents Aggregations Schema Indexes Validation

Filter Type a query: { field: 'value' } or Generate query

ADD DATA EXPORT DATA

```

_id: ObjectId('65e3ea7d6cb39e89292b3884')
Idjugador: 3883
nombre: "Peter"
apellido: "Rasmussen"
nickname: "dupreeh"
codigociudad: "DK"
total-ganancias: 1822989.41
juego: "Counter-Strike: Global Offensive"
genero: "First-Person Shooter"
Ciudad: "Denmark"

```

```

_id: ObjectId('65e3ea7d6cb39e89292b3885')
Idjugador: 3679
nombre: "Andreas"
apellido: "Højsteth"
nickname: "Xyp9x"
codigociudad: "DK"
total-ganancias: 1780000.07

```

Subimos junto con las demás colecciones

Proyecto_Analisis

Conciertos_Taylor

E-Sports Ganacias J...

deportes mundial 1

deportes mundial 2

ADD DATA

Evaluación del Valor: Clasificación de Eventos Deportivos a Escala Mundial

Extraemos los datos de los dos conjuntos de datos, empezando por el conjunto de datos 1 y

luego por el conjunto de datos 2.

Kaggle

Toughest Sports by Skill

Top 60 Sports Ranked by Skill

About Dataset

Context

Many people want to know what sports are easy and which ones are hard. This dataset will tell you exactly that.

Content

In this dataset you'll find 60 sports ranked from top to bottom based on their skill level by ESPN.

Acknowledgements

This data comes from <http://www.espn.com/espn/page2/sportData>.

Usability

8.24

License

CC0: Public Domain

Expected update frequency

Never

Tags

Arts and Entertainment

Health

Sports

Kaggle

Most Valuable Sports Team

Most Valuable Clubs, Most Valuable Franchise

About Dataset

This is Dataset about the "World's Most Valuable Sports Teams 2021".

I have Scraped this Dataset Using Selenium From Forbes website.

It contains 50 most valuable Sports team of World.

Explore this Dataset and you will find interesting things.

One thing that I noticed is Real Madrid or Barcelona are not richest club and there are no IPL team there :)

Usability

8.75

License

CC0: Public Domain

Expected update frequency

Annually

Tags

Sports

Beginner

Data Visualization

Maps/Geo

end

A continuación, cargamos ambos conjuntos de datos en phpMyAdmin para conectarlos a MySQL.

Data set uno:

phpMyAdmin

Importar

Archivo a importar

El archivo puede ser comprimido (gzip, bz2) o descomprimido. Un archivo comprimido tiene que terminar en .(formato)comprimido. Por ejemplo: xyz.zip

Buscar en su ordenador: phpmyadmin

Seleccionar archivo: toughestsports.csv

También puede seleccionar un archivo en cualquier página.

Conjunto de caracteres del archivo: utf-8

Importación parcial

☒ Permitir la interrupción de una importación en caso que el script detecte que se ha accedido al límite de tiempo PHP. Esto puede ser útil cuando se importan muchos archivos grandes de archivo, para evitar los timeouts.

Only this cantidad de consultas (en SQL) desde la primera: 1

Otras opciones

☒ Habilita la revisión de los claves binarias

Formato: CSV

Nota: si el archivo contiene múltiples tablas, van a ser combinadas en una.

phpMyAdmin

Opciones específicas al formato

☒ Actualizar datos cuando los datos importados están duplicados (agregar ON DUPLICATE KEY UPDATE)

Columnas dependientes por:

Columnas encadenadas entre:

Carácter de escape de columnas:

Líneas terminadas en: auto

Nombre de la tabla nueva (opcional):

Nombre de la base de datos nueva (opcional):

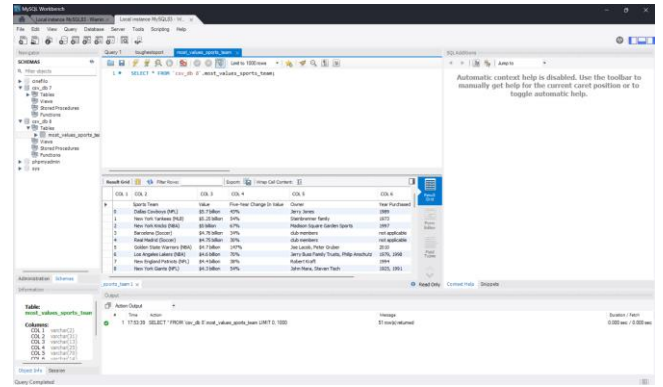
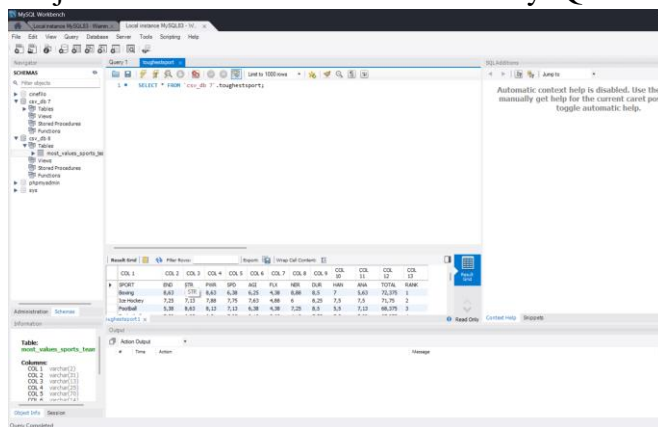
Importe este gran número de filas (opcional):

☒ La primera fila del archivo contiene los nombres de columna de la tabla (si no está activado la primera fila será parte de los datos)

☒ No añadir la columna si existe con el mismo nombre

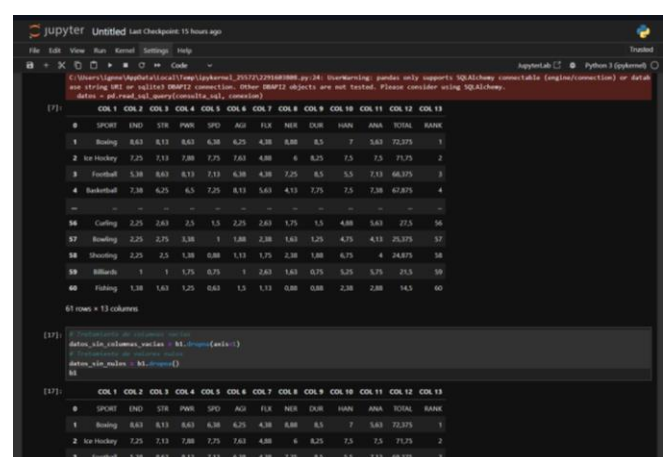
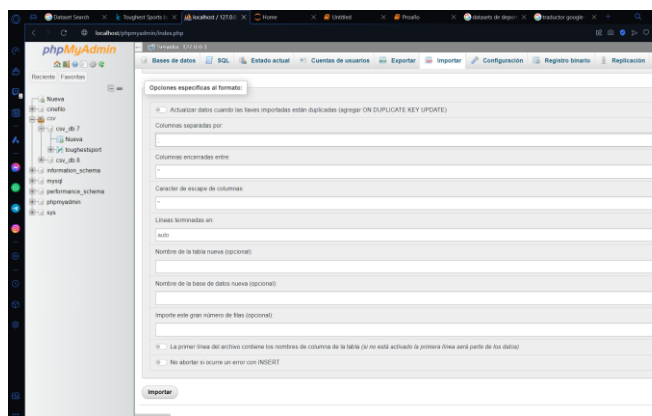
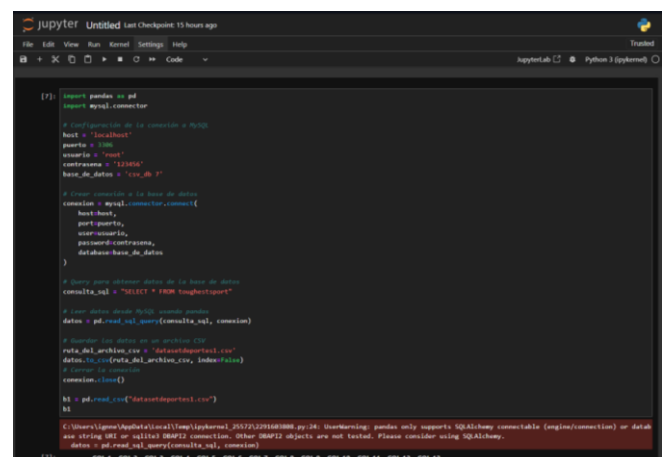
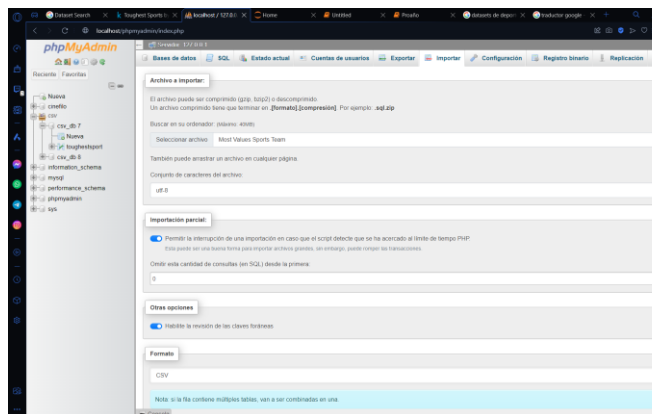
Importar

Confirmación de que se ha subido el primer conjunto de datos a MySQL.



Lo siguiente que haremos es la extracción y limpieza de los datos con jupyter notebook utilizando las bibliotecas de numpy y pandas

Ahora lo hacemos otra vez con la data set 2



Confirmación de que el data set dos se subio exitosamente al Mysql


```
[12]: import pandas as pd
import mysql.connector

# Configuración de la conexión a MySQL
host = "localhost"
puerto = 3306
usuario = "root"
contraseña = "123456"
base_de_datos = "causas_muertes"

# Crear conexión a la base de datos
conexion = mysql.connector.connect(
    host=host,
    port=puerto,
    user=usuario,
    password=contraseña,
    database=base_de_datos,
)

# Query para obtener datos de la base de datos
consulta_sql = "SELECT * FROM causas_muertes_datos"

# Leer datos desde MySQL usando pandas
datos = pd.read_sql_query(consulta_sql, conexion)

# Mostrar los datos en un archivo CSV
ruta_archivo_csv = "datos_causas_muertes.csv"
datos.to_csv(ruta_archivo_csv, index=False)

# Guardar la conexión
conexion.close()

# Crear un objeto de conexión a MongoDB
mongo = pymongo.MongoClient("mongodb://localhost:27020/")
db = mongo["causas_muertes"]
collection = db["datos_causas_muertes"]

# Insertar los datos en MongoDB
collection.insert_many(datos.to_dict("records"))
```

	COL 1	COL 2	COL 3	COL 4	COL 5	COL 6	COL 7
0	NaN	Sports Team	Value	Flow Year Change In Value	Owner	Year Purchased	Price Paid
1	0.0	Dallas Cowboys (NFL)	\$5.7 billion	43%	Jerry Jones	1989	\$150 million
2	1.0	New York Yankees (MLB)	\$5.25 billion	54%	Steinbrenner family	1973	\$8.8 million
3	2.0	New York Knicks (NBA)	\$5 billion	67%	Madison Square Garden Sports	1997	\$300 million
4	3.0	Barcelona (Soccer)	\$4.76 billion	34%	club members	not applicable	not applicable
5	4.0	Real Madrid (Soccer)	\$4.75 billion	30%	club members	not applicable	not applicable
6	5.0	Golden State Warriors (NBA)	\$4.7 billion	147%	Joe Lacob, Peter Guber	2010	\$450 million
7	6.0	Los Angeles Lakers (NBA)	\$4.6 billion	70%	Jerry Buss Family Trusts, Philip Anschutz	1976, 1996	200million, 260 million
8	7.0	New England Patriots (NFL)	\$4.4 billion	38%	Robert Kraft	1994	\$172 million
9	8.0	New York Giants (NFL)	\$4.3 billion	54%	John Mara, Steven Tisch	1925, 1991	\$80,150 million
10	9.0	Bayern Munich (Soccer)	\$4.21 billion	57%	club members	not applicable	not applicable
11	10.0	Manchester United (Soccer)	\$4.2 billion	27%	Glazer family	2005	\$1.4 billion
12	11.0	Liverpool (Soccer)	\$4.1 billion	160%	John Henry, Tom Werner	2010	\$476 million
13	12.0	Los Angeles Rams (NFL)	\$4.1 billion	176%	Stanley Kronzberg	2010	\$750 million
14	13.0	Manchester City (Soccer)	\$4 billion	108%	Khalid Mansour bin Zayed Al Nahyan	2008	\$385 million
15	14.0	San Francisco 49ers (NFL)	\$3.8 billion	41%	Debrah DeLauro York, John York	1977	\$13 million
16	15.0	Los Angeles Dodgers (MLB)	\$3.57 billion	43%	Coughlin Baseball Management	2012	\$2 billion
17	16.0	New York Jets (NFL)	\$3.55 billion	37%	Johnson family	2000	\$635 million
18	17.0	Chicago Bears (NFL)	\$3.53 billion	44%	McCawley family	1920	\$100 million
19	18.0	Washington Football Team (NFL)	\$3.5 billion	22%	Daniel Snyder	1999	\$750 million
20	19.0	Boston Red Sox (MLB)	\$3.47 billion	51%	John Henry, Tom Werner	2002	\$380 million
21	20.0	Philadelphia Eagles (NFL)	\$3.4 billion	42%	Jeffrey Lurie	1994	\$185 million

```
[13]: # Transformación de columnas para MongoDB
datos_sin_columnas_varias = datos.drop(['COL 1', 'COL 2', 'COL 3', 'COL 4', 'COL 5', 'COL 6', 'COL 7'], axis=1)
datos_sin_columnas_varias.to_dict('records')

[14]: # Transformación de columnas para MongoDB
datos_sin_columnas_varias = datos.drop(['COL 1', 'COL 2', 'COL 3', 'COL 4', 'COL 5', 'COL 6', 'COL 7'], axis=1)
datos_sin_columnas_varias.to_dict('records')
```

Estudio de Mortalidad: Tasas de Enfermedades Prevalentes a Nivel Internacional

Para comenzar, accedemos a Kaggle, una plataforma especializada en la recopilación de conjuntos de datos para análisis. Una vez en la plataforma, localizamos y descargamos el

archivo que vamos a trabajar.



Una vez que hemos descargado el archivo, abrimos Jupyter Notebook para llevar a cabo el proceso de limpieza necesario. En primer lugar, verificamos si los datos se leen correctamente.

```
# Leer el archivo CSV
import pandas as pd

# Leer el archivo CSV
datos = pd.read_csv('ruta_archivo')
print(datos)

# Verificar la estructura de los datos
print(datos.info())
```

Realizamos la limpieza de datos utilizando las siguientes líneas de código. Esto implica eliminar cualquier valor nulo y reemplazarlo por

```
# Limpieza del CSV
import pandas as pd

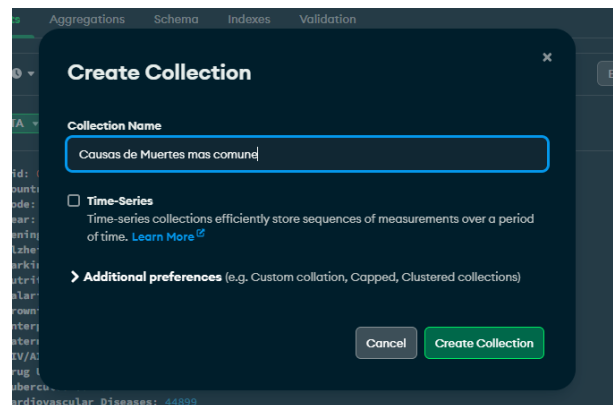
nombre_archivo = 'cause_of_deaths.csv'
datos = pd.read_csv(nombre_archivo)

datos_limpios = datos.fillna(0)

# Guardar el archivo limpio
nombre_archivo_limpio = 'cause_of_deaths_limpio.csv'
datos_limpios.to_csv(nombre_archivo_limpio, index=False)

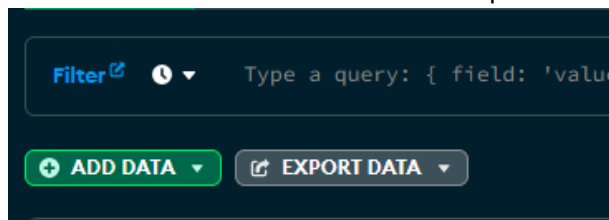
print("Limpieza de datos completa. El archivo limpio se ha guardado como:", nombre_archivo_limpio)
```

Añadiremos una nueva conexión al repositorio de MongoDB para su posterior análisis. Le asignaremos un nombre identificativo para reconocerlo fácilmente dentro de la base de datos.

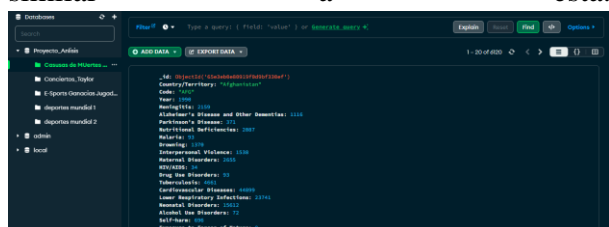


Para incorporar el archivo CSV al proyecto general de MongoDB y así acceder a toda la información, simplemente haremos clic en el botón 'Agregar

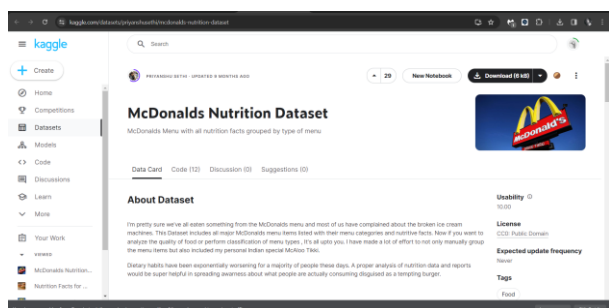
datos al repositorio'.



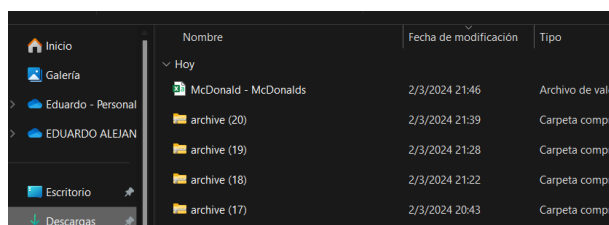
Localizamos la ruta donde se encuentran los datos y los cargamos en MongoDB. Una vez completada la carga, veremos una pantalla similar a esta.



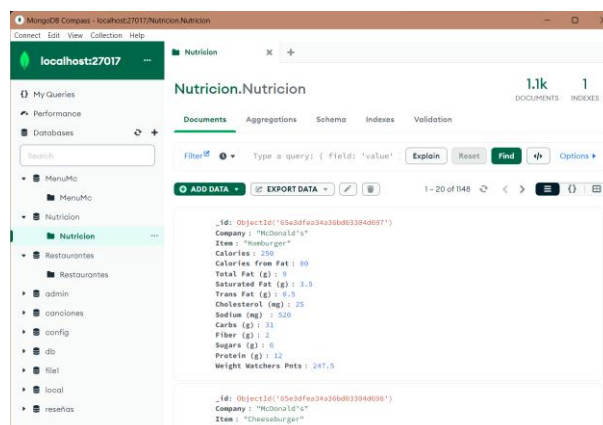
Buscamos e importamos conjuntos de datos de fuentes de información gratuitas, como en este caso, que proviene de Kaggle.



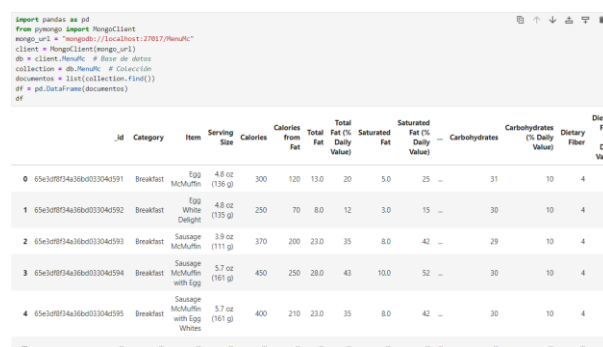
Una vez identificadas las bases de datos que consideramos atractivas para el análisis de sus datos, procedemos a descargarlas.



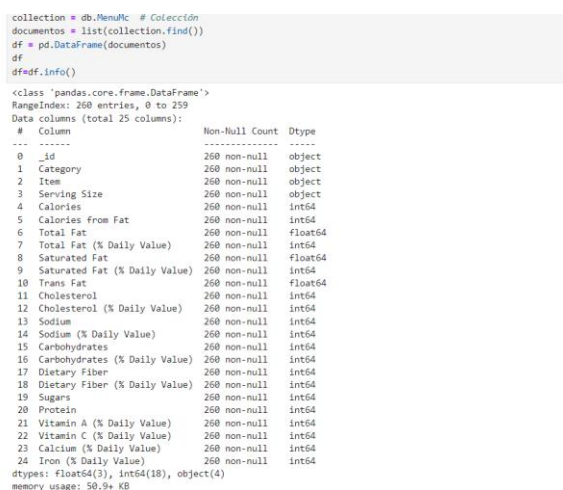
Una vez completado este paso, importamos las bases de datos a MongoDB para poder visualizar los datos de manera más efectiva.



Después de haber importado los datos, conectamos MongoDB con Jupyter Notebook para llevar a cabo la limpieza de datos correspondiente.



Para examinar más detalladamente los datos registrados, utilizamos las funciones de la biblioteca Pandas para realizar modificaciones en la base de datos que se ha guardado en un DataFrame.



También verificamos si la base de datos contiene valores nulos para poder reemplazarlos y asegurarnos de que no se

muestren en la base de datos, evitando que afecten nuestro análisis posterior.

```
[34]: import pandas as pd
from pymongo import MongoClient
mongo_url = "mongodb://localhost:27017/MenúMc"
client = MongoClient(mongo_url)
db = client.MenúMc # Base de datos
collection = db.MenúMc # Colección
documentos = list(collection.find())
df = pd.DataFrame(documentos)
df = df.dropna()
print(df.isnull().any())
```

```
_id      False
Category False
Item      False
Serving Size False
Calories  False
Calories from Fat False
Total Fat (% Daily Value) False
Saturated Fat (% Daily Value) False
Trans Fat (% Daily Value) False
Cholesterol (% Daily Value) False
Sodium (% Daily Value) False
Carbohydrates (% Daily Value) False
Dietary Fiber (% Daily Value) False
Sugars (% Daily Value) False
Protein  False
Vitamin A (% Daily Value) False
Vitamin C (% Daily Value) False
Calcium (% Daily Value) False
Iron (% Daily Value) False
dtype: bool
```

En caso de encontrar valores nulos, los reemplazamos con valores predeterminados. Además, procedemos a eliminar columnas que no sean necesarias.

```
db = client.MenúMc # Base de datos
collection = db.MenúMc # Colección
documentos = list(collection.find())
df = pd.DataFrame(documentos)
df = df.drop(columns=['_id'])
df = df.fillna(valor_predeterminado)
df = df.dropna()
df
```

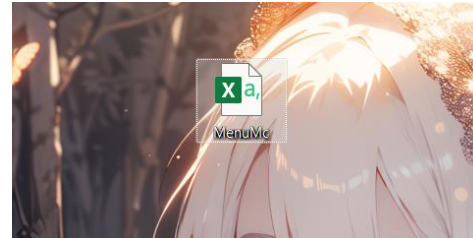
	Category	Item	Serving Size	Calories	Calories from Fat	Total Fat (% Daily Value)	Saturated Fat (% Daily Value)	Trans Fat (% Daily Value)	Carbohydrates (% Daily Value)	Dietary Fiber (% Daily Value)	Sugars (% Daily Value)	Protein
0	Breakfast	Egg McMuffin	4.8 oz (136 g)	300	120	13.0	20	5.0	25	0.0	...	17
1	Breakfast	Egg White Delight	4.8 oz (133 g)	250	70	8.0	12	3.0	15	0.0	...	18
2	Breakfast	Sausage McMuffin	3.9 oz (111 g)	370	200	23.0	35	8.0	42	0.0	...	14
3	Breakfast	Sausage McMuffin	5.7 oz (161 g)	450	250	28.0	43	10.0	52	0.0	...	21

También eliminamos las filas que contienen valores nulos por si acaso se encontrara algún dato que no se leyó correctamente.

```
df = df.fillna(valor_predeterminado)
df = df.drop_duplicates()
df
```

	Category	Item	Serving Size	Calories	Calories from Fat	Total Fat (% Daily Value)	Saturated Fat (% Daily Value)	Trans Fat (% Daily Value)	Carbohydrates (% Daily Value)	Dietary Fiber (% Daily Value)	Sugars (% Daily Value)	Protein
0	Breakfast	Egg McMuffin	4.8 oz (136 g)	300	120	13.0	20	5.0	25	0.0	...	17
1	Breakfast	Egg White Delight	4.8 oz (133 g)	250	70	8.0	12	3.0	15	0.0	...	18

Una vez que la base de datos esté limpia, procedemos a descargar el archivo CSV corregido para poder analizarlo en los siguientes pasos. Este proceso también lo realizamos con las demás bases de datos con las que llevaremos a cabo los análisis respectivos.



```
import pandas as pd
from pymongo import MongoClient
mongo_url = "mongodb://localhost:27017/Nutricion"
client = MongoClient(mongo_url)
db = client.Nutricion
collection = db.Nutricion
documentos = list(collection.find())
df = pd.DataFrame(documentos)
df = df.drop(columns=['_id'])
valor_predeterminado=0
df = df.fillna(valor_predeterminado)
df = df.dropna()
df = df.drop_duplicates()
df
df.to_csv('Nutricion.csv', index=False)
```

```
import pandas as pd
from pymongo import MongoClient
mongo_url = "mongodb://localhost:27017/Restaurantes"
client = MongoClient(mongo_url)
db = client.Restaurantes
collection = db.Restaurantes
documentos = list(collection.find())
df = pd.DataFrame(documentos)
df = df.drop(columns=['_id'])
valor_predeterminado=0
df = df.fillna(valor_predeterminado)
df = df.dropna()
df = df.drop_duplicates()
df
df.to_csv('Restaurantes.csv', index=False)
```

Una vez que las bases de datos estén limpias, procedemos a establecer la conexión con MongoDB Atlas, la cual contendrá varias colecciones que serán de los demás integrantes del grupo.

New Connection

Connect to a MongoDB deployment

URI

Edit Connection String

mongodb+srv://wesitos:*****@proyecto-analisis.fuwxe0.mongodb.net/

Advanced Connection Options

Save

Save & Connect

Connect

Una vez conectados creamos nuestra colección para la base

Proyecto_Análisis

123 MB

Conciertos_Taylor

Storage size: 56.84 kB

deportes mundiales

Storage size: 24.58 kB

Create Collection

Collection Name

Restaurantes

☐ Time-Series

Additional preferences (e.g. Custom collation, Capped, Clustered collections)

Importamos nuestro documento csv

Exploración de los World Tours de Taylor Swift: Un Estudio Exhaustivo

Importar Pandas, Mysql conector, validar credenciales y crear el dataframe.

```
import pandas as pd
import mysql.connector

# Configuración de la conexión a la base de datos
host = 'localhost'
user = 'root'
password = '123456'
database = 'conciertos'

# Crear la conexión a la base de datos
conn = mysql.connector.connect(host=host, user=user, password=password, database=database)

# Consulta SQL para obtener datos de la tabla (reemplaza 'tu_tabla' con el nombre real de tu tabla)
query = 'SELECT * FROM conciertos_taylor'

# Utilizar pandas para ejecutar la consulta y cargar los resultados en un DataFrame
df = pd.read_sql(query, conn)

# Cerrar la conexión a la base de datos
conn.close()

# Visualizar el DataFrame (puedes realizar otras operaciones con los datos aquí)
print(df)

import pandas as pd
# Cargar el dataset
df = pd.read_csv('Conciertos_Taylor.csv')
```

Mostramos la información del dataframe para asegurarnos que los elementos debemos de limpiar.

```
# Mostrar las primeras filas del dataset
print(df.head())

# Obtener información sobre las columnas
print(df.info())

# Estadísticas descriptivas del dataset
print(df.describe())
```

Eliminamos las filas con valores nulos.

```
# Eliminar filas con valores nulos
df = df.dropna()
```

A los valores que sean nulos los reemplazamos con el valor más repetido de dichas tablas

```
# Reemplazar los valores nulos con el valor más repetido en esa columna
valor_mas_repetido = df['City'].mode()[0]
df['City'].fillna(valor_mas_repetido, inplace=True)

[26]:

valor_mas_repetido = df['Country'].mode()[0]
df['Country'].fillna(valor_mas_repetido, inplace=True)

[27]:

valor_mas_repetido = df['Venue'].mode()[0]
df['Venue'].fillna(valor_mas_repetido, inplace=True)

[28]:

valor_mas_repetido = df['Opening act(s)'].mode()[0]
df['Opening act(s)'].fillna(valor_mas_repetido, inplace=True)

[29]:

valor_mas_repetido = df['Attendance (tickets sold / available)'].mode()[0]
df['Attendance (tickets sold / available)'].fillna(valor_mas_repetido, in
```

Eliminamos duplicados y modificamos la tabla Revenue para utilizar los valores como enteros eliminando caracteres especiales y comas.

```
# Eliminar duplicados
df = df.drop_duplicates()

[49]:

# columna que contiene valores con el símbolo de dólar
df['Revenue'] = df['Revenue'].str.replace('$', '')

[53]:

# columna que contiene valores con comas
df['Revenue'] = df['Revenue'].str.replace(',', '')

[54]:

# columna que contiene valores con caracteres no deseados
df['Revenue'] = df['Revenue'].replace('[^\d.]', '', regex=True)
```

Transformamos la columna a valores enteros

```
df['Revenue'] = df['Revenue'].astype(int)
```

Reemplazamos los valores de cero con el más repetido

```
df['Revenue'].replace(0, pd.NA, inplace=True)

# Encontrar el valor más repetido
valor_mas_repetido = df['Revenue'].mode()[0]

# Reemplazar ceros con el valor más repetido y NaN con 0
df['Revenue'] = df['Revenue'].replace({0: valor_mas_repetido, pd.NA: 0})

# Imprimir el DataFrame después de la sustitución
print(df)
```

Comprobamos que los datos han sido cambiados con

```
df.head()
```

Creamos una copia con la limpieza avanzada para evitar posibles errores

```
# Crear una copia del DataFrame original
df_copy = df.copy()
```

Para nuestro analisis utilizamos una tabla llamada “Attendance (tickets sold)/Attendance (available)” por lo que al haber dos datos en la misma tabla, la separamos en dos y continuamos con la limpieza realizada antes

```
# Crear nuevas columnas 'Attendance (tickets sold)' y 'Attendance (available)'
df_copy[['Attendance (tickets sold)', 'Attendance (available)']] = df_copy['Attendance (tickets sold / available)']

# Eliminar la columna original 'Attendance (tickets sold / available)'
df_copy.drop(columns=['Attendance (tickets sold / available)'], inplace=True)

# Eliminar los caracteres ", " de las nuevas columnas
df_copy['Attendance (tickets sold)'] = df_copy['Attendance (tickets sold)'].str.replace(", ", "")
df_copy['Attendance (available)'] = df_copy['Attendance (available)'].str.replace(", ", "")

# Reemplazar valores nulos, vacíos, "0" y None con cero en las nuevas columnas
df_copy['Attendance (tickets sold)'] = df_copy['Attendance (tickets sold)'].fillna(0)
df_copy['Attendance (available)'] = df_copy['Attendance (available)'].fillna(0)

# Imprimir el DataFrame copia después de la transformación
print(df_copy)
```

Luego de la limpieza convertimos los valores de string a enteros de las dos columnas, además de reemplazar los valores de nulo con el mas repetido

```
# Convertir las columnas a tipo de datos entero
df_copy['Attendance (tickets sold)'] = df_copy['Attendance (tickets sold)'].astype(int)
df_copy['Attendance (available)'] = df_copy['Attendance (available)'].astype(int)

[113]:

# Encontrar el valor más repetido excluyendo ceros en 'Attendance (tickets sold)'
valor_tickets_repetido = df_copy['Attendance (tickets sold)'].mode()[0]

# Encontrar el valor más repetido excluyendo ceros en 'Attendance (available)'
valor_available_repetido = df_copy['Attendance (available)'].mode()[0]

# Reemplazar ceros con el valor más repetido excluyendo ceros
df_copy['Attendance (tickets sold)'] = df_copy['Attendance (tickets sold)'].fillna(valor_tickets_repetido)
df_copy['Attendance (available)'] = df_copy['Attendance (available)'].fillna(valor_available_repetido)
```

Comprobamos que las dos columnas han sido creadas con ayuda de df.head(20)

Attendance (tickets sold)	Attendance (available)
7463	7463
7822	7822
13764	13764
11208	11208
8751	8751
11072	11072

Por ultimo una vez comprobada la limpieza guardamos el dataset en un csv nuevo

```
#Guardamos el df en un nuevo csv
df_copy.to_csv('Concerts_Taylor_Clean.csv', index=False)
```

VIII. Análisis de información

Análisis de Ganancias: Jugadores, Equipos y Videojuegos a Nivel Global

1-Grafica

La conclusión principal La diferencia significativa en las ganancias entre China y los otros países mencionados indica el dominio chino en este campo. Esto podría atribuirse a varios factores, como la gran población de jugadores el apoyo gubernamental, la infraestructura de e-sports desarrollada y una cultura arraigada de juegos en China Corea del Sur, históricamente conocida por su dominio en los e-sports, sigue siendo un competidor fuerte, con una impresionante cantidad de ganancias y una gran cantidad de jugadores

los e-sports en Corea del Sur ha sido impulsada por la popularidad de los juegos en línea y el apoyo de la industria del entretenimiento en resumen, los e-sports continúan siendo un campo emocionante y lucrativo para los jugadores talentosos en todo el mundo, con China liderando en términos de ganancias, seguido de cerca por Corea del Sur y Suecia

2-Grafica

E-sports es diverso y global, con varios países destacando en diferentes juegos y regiones emergentes mostrando un crecimiento prometedor en la industria. China y Corea del Sur son líderes reconocidos, con una gran presencia en juegos como League of Legends, Dota 2, StarCraft, Overwatch y PUBG. Tailandia, Taiwán y Rusia también tienen comunidades activas de esports, con juegos populares y jugadores talentosos. Además, países europeos como Francia, Suecia y Alemania tienen una presencia significativa en varios juegos, mientras que Estados Unidos y Dinamarca cuentan con jugadores destacados en una variedad de títulos. En resumen, los esports son un fenómeno global en constante crecimiento, con una amplia participación internacional y un enorme potencial tanto en términos de entretenimiento como de oportunidades económicas.

3-Grafica

La conclusión principal que se puede extraer es que los géneros de videojuegos más populares varían en función de la preferencia de los jugadores y la disponibilidad de diferentes plataformas y juegos. Los juegos en línea multijugador y los de estrategia son los más populares, lo que sugiere que los jugadores disfrutan tanto de la competencia directa como de la planificación estratégica. Los juegos de Battle Royal, FPS y los de cartas

coleccionables también tienen una base considerable de seguidores, lo que demuestra la diversidad de gustos en la comunidad de jugadores. En general, esta variedad en la preferencia de géneros muestra la amplitud y la profundidad de la industria de los videojuegos, que ofrece opciones para todos los tipos de jugadores.

4-Grafica

La conclusión principal que se puede sacar es que Dota 2 ha sido extremadamente lucrativo para algunos jugadores prominentes. Los cinco jugadores más exitosos en términos de ganancias individuales provienen todos de este juego. Esto refleja la importancia de Dota 2 en el panorama de los esports y sugiere que este juego ofrece algunas de las mayores oportunidades de ganancias para los jugadores profesionales.

El hecho de que estos jugadores hayan obtenido el 100% de sus ingresos de Dota 2 resalta la especialización y dedicación al juego por parte de estos jugadores. Dota 2 es un juego altamente competitivo y complejo que requiere un alto nivel de habilidad y compromiso para alcanzar el éxito a nivel profesional. La consistencia en el rendimiento y el éxito en este juego ha sido clave para generar estas impresionantes ganancias.

Además, la presencia dominante de jugadores de Dota 2 en la lista sugiere que el juego ha sido particularmente exitoso en términos de premios monetarios en comparación con otros esports.

5.Grafica

Muestra una jerarquía clara en términos de ganancias para los jugadores individuales en diferentes juegos de esports. Dota 2 se destaca como el juego más lucrativo, con ganancias superiores a los \$7 millones, seguido de cerca por CS: GO y LoL, con

alrededor de \$4 millones y \$3 millones respectivamente.

Evaluación del Valor: Clasificación de Eventos Deportivos a Escala Mundial

Grafica 1:

Es un gráfico circular que analiza la popularidad o el seguimiento de varios equipos deportivos:

Equipos Deportivos: Los equipos mencionados son New York Knicks (NBA), Real Madrid (Fútbol), Golden State Warriors (NBA), Los Angeles Lakers (NBA), New York Giants (NFL), Bayern Munich (Fútbol), Manchester United (Fútbol), New England Patriots (NFL) y Barcelona (Fútbol).

Aquí cada equipo tiene una sección en el gráfico circular que representa su porcentaje de seguimiento en comparación con los demás. La mayoría del gráfico (82%) pertenece a la categoría “otros”, lo que sugiere que hay una gran diversidad de equipos seguidos además de los mencionados específicamente.

Grafico 2:

Este es un diagrama de barras horizontal que analiza el conteo de registros (Record Count) para diferentes equipos deportivos. Aquí están los datos que analiza:

Equipos Deportivos: Los equipos mencionados son New York K., Real Madrid, Golden Statz, Los Angeles..., New York G., Bayern Munz, New England..., Barcelona L., New York Y., y Dallas Cow...

Conteo de Registros: Cada equipo tiene una barra en el gráfico que representa su conteo de registros. Todos los equipos mostrados tienen un conteo de registros de 1 o equivalente a 0,8 en la escala del gráfico.

Grafica 3:
Este un diagrama de barras horizontal que analiza la cantidad de calorías quemadas por

hora en diferentes deportes y actividades físicas. Aquí están los datos que analiza:

Los deportes y actividades físicas mencionados incluyen boxeo, fútbol, alpinismo, lacrosse, salto con pértiga, salto de altura, salto de longitud, béisbol/softbol, baloncesto, lucha, rugby, patinaje de velocidad, ciclismo, carreras de autos, buceo, orientación, escalada de rocas/montañas, deportes de rodeo, esquí acuático, tenis de mesa, artes marciales/kick boxing, gimnasia, hockey sobre césped, patinaje artístico, racquetball/squash, esgrima, fútbol, hockey sobre patines, voleibol de playa, tenis, golf, equitación, tiro con arco, y pesca.

Cada deporte o actividad física tiene una cantidad asociada de calorías quemadas por hora. Por ejemplo, el boxeo quema 800 cal/hr, mientras que el yoga quema 183 cal/hr.

Grafico 4:
Es un diagrama de barras horizontal que analiza la cantidad de registros (Record Count) asociados con varios deportes. Aquí están los datos que analiza:

Los deportes mencionados en el gráfico son Fútbol, Béisbol, Rugby, Speed Skating, Team Handball, Badminton, Diving, Rowing, Track & Field, Table Tennis, Golf, Archery y Billiards Gymnastics.

Cantidad de Registros: Cada deporte tiene una barra en el gráfico que representa su conteo de registros. Los deportes con barras más largas tienen más registros.

Grafica 5:
Es un diagrama de barras horizontal que analiza la riqueza neta de diferentes individuos. Aquí están los datos que analiza:

Los individuos mencionados en el gráfico son Mark Zuckerberg, Jeff Bezos, Bernard Arnault, Queen Elizabeth II, John F. Kennedy Jr., Joe Biden y Emmanuel Macron.

Cada individuo tiene una barra en el gráfico que representa su riqueza neta en miles de millones

de dólares. Por ejemplo, Mark Zuckerberg tiene una riqueza neta de \$24.5 mil millones, mientras que Emmanuel Macron tiene una riqueza neta de \$4.7 mil millones.

Estudio de Mortalidad: Tasas de Enfermedades Prevalentes a Nivel Internacional

Grafica 1

En el gráfico se presenta la distribución de muertes debido a la hepatitis aguda a nivel mundial a lo largo del tiempo, desde el descubrimiento de la enfermedad hasta el año 2021. Se destaca que la India registra el 44.2% de las muertes a nivel global, seguida por China con el 8.2%. Además, otros países como Pakistán, Indonesia, Bangladesh, Nigeria, Afganistán, Sudán y Etiopía muestran porcentajes de mortalidad que oscilan entre el 2% y el 3.4% aproximadamente.

Grafica 2:

El gráfico muestra las muertes causadas por la meningitis a nivel mundial. Se destaca que la India registra la cifra más alta, con 2 millones de muertes, seguida por Nigeria con 1.5 millones. Es importante resaltar la notable brecha con el siguiente país en la lista, Pakistán, que cuenta con tan solo 500 mil muertes, mientras que los demás países se sitúan por debajo de esta cifra.

Grafica 3:

Esta gráfica representa a escala global el número de muertes atribuidas al trastorno por consumo de alcohol en el mundo, alcanzando un total de 4.8 millones. Estas muertes pueden deberse a una variedad de causas, que incluyen lesiones como las ocasionadas por accidentes de tránsito, autolesiones y violencia interpersonal, trastornos digestivos, enfermedades cardiovasculares, así como también a enfermedades infecciosas, cánceres, trastornos mentales y otras afecciones.

Grafica 4:

La gráfica representa las muertes por trastornos respiratorios crónicos desde 1990 hasta el año 2015 y más allá. En los años 90, se observa un inicio con 3.1 millones de muertes, pero este número crece de manera exponencial. En tan solo 5 años, aumenta a 3.3 millones, y aproximadamente para el año 2003 alcanza los 3.5 millones, manteniéndose así hasta 2010. Sin embargo, en 2015 se produce un notable incremento, llegando a 3.7 millones de muertes, y esta tendencia continúa en aumento.

Grafica 5:

En el mapa mundial, se destacan los países con un mayor número de muertes debido al Alzheimer y otras enfermedades demenciales. Estados Unidos y Canadá sobresalen, registrando entre 3.3 y 3.5 millones de muertes por estas enfermedades. Sin embargo, por encima de ellos se encuentra la India, con un registro de 5.3 millones de muertes.

Impacto del Consumo de Comida Rápida en la Salud: Un Análisis Profundo

Grafico 1:

El gráfico 1 revela que McDonald's lidera en la suma de calorías, azúcares y sodio entre las empresas analizadas, seguido por Burger King, Taco Bell, Wendy's, KFC y Pizza Hut. Esto sugiere que los productos de McDonald's tienden a tener un contenido más elevado de estos componentes en comparación con las demás cadenas. Además, se destaca que la cantidad de calorías es la más predominante en todas las empresas, seguida por el sodio y luego los azúcares, indicando una tendencia general de altos niveles de calorías y sodio en los productos de comida rápida. Finalmente, se observa que Pizza Hut acumula la mayor cantidad de estos tres elementos, lo que implica que sus productos podrían tener el mayor contenido de calorías, azúcares y sodio entre las cadenas mencionadas.

Grafico 2:

El siguiente mapa ilustra la distribución geográfica de restaurantes que ofrecen alimentos con alto contenido calórico. Se destaca una densa concentración de estos establecimientos en América del Norte, con una presencia notable también en Europa, así como en ciertas regiones de Asia y América del Sur. Este patrón podría indicar un acceso más amplio a alimentos ricos en calorías en estas áreas. Sin embargo, es esencial considerar otros aspectos, como la disponibilidad de opciones de alimentos saludables y los hábitos dietéticos de los habitantes locales.

Grafico 3:

El gráfico, un diagrama de dispersión, ilustra la combinación de calorías y calcio presentes en varios productos alimenticios ofrecidos por McDonald's, Burger King, Taco Bell, Wendy's, KFC y Pizza Hut. Cada punto en el gráfico representa un producto específico, y se distinguen por colores. Se aprecia una diversidad de valores tanto en calorías como en calcio entre los diferentes productos de cada compañía. Esto sugiere una amplia gama de opciones alimenticias en términos de su contenido calórico y de calcio dentro de estas cadenas de restaurantes.

Grafico 4:

El diagrama de dispersión representa la composición de proteínas, grasa saturada y colesterol en una selección de alimentos, todos pertenecientes al menú de McDonald's. Cada punto azul representa un artículo específico del menú, proporcionando una visualización clara de su contenido en estos tres componentes. Se observa una concentración mayor de puntos en la esquina inferior izquierda, lo que indica que muchos de los alimentos de McDonald's tienen niveles más bajos de estas sustancias. Sin embargo, algunos puntos dispersos hacia la esquina superior derecha sugieren que también existen opciones en el menú con cantidades más altas de proteínas, grasa saturada y colesterol. Este análisis resalta la variedad de opciones disponibles en el menú de

McDonald's en términos de contenido nutricional, ofreciendo alternativas tanto para aquellos que buscan opciones más saludables como para quienes prefieren indulgencias ocasionales.

Grafico 5:

El gráfico muestra la distribución de calorías por categoría de alimentos a través de un diagrama de barras. Destaca que la categoría "Breakfast" lidera en cantidad de calorías, seguida por "Chicken & Beef". Por otro lado, las categorías "Desserts", "Salads" y "Sides" exhiben las cifras más bajas en calorías. Este análisis sugiere que los alimentos típicos del desayuno y aquellos que incluyen pollo y carne de res tienden a ser más calóricos, mientras que los postres, ensaladas y acompañamientos presentan niveles más bajos de calorías.

Exploración de los World Tours de Taylor Swift: Un Estudio Exhaustivo

Grafico 1:

Como se muestra en la gráfica se puede ver que la ciudad con más ingresos fue Foxborough así como, la ciudad con menos ingresos es Winchester

Grafico 2:

El grafico nos muestra un mapa donde se concentran el mayor número de tickets vendidos por ciudad, vemos que estados unidos es el país con más ventas a comparación de Japón que es el que menos ventas tiene, aunque este grafico nos sirve también para ver en donde se concentran más los conciertos realizados por Taylor Swift

Grafico 3:

Este grafico nos enseña la cantidad de eventos por tour realizados Siendo el Fearless_Tour el más grande con 29,7% y Reputation_Stadium_Tour el que menos tiene con 10,91%

Grafico 4:

El grafico nos muestra la suma de ingresos totales que ha tenido los tours mundiales de Taylor Swift con un total de 932 millones en un medidor que va de 0 a 2 mil millones, se puede decir que es casi la mitad de esta medida

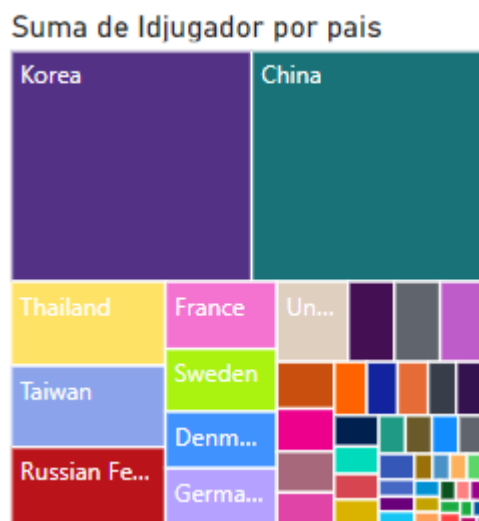
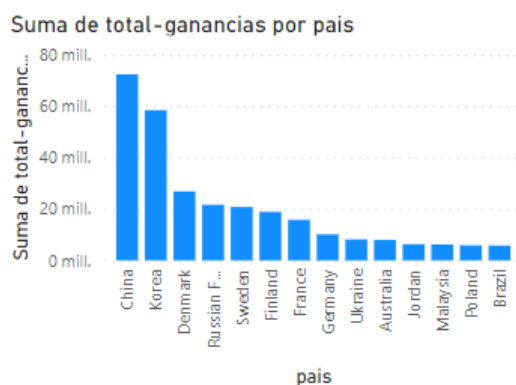
Grafico 5:

Este grafico nos muestra la cantidad de músicos de apertura por tour, teniendo un mayor número Fearless_Tour en comparación con el más bajo, siendo este el Reputation_Stadium_Tour

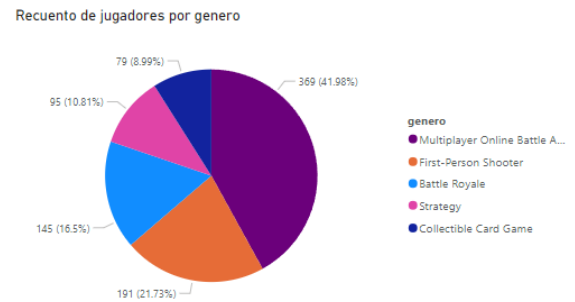
IX. Visualización de información.

Análisis de Ganancias: Jugadores, Equipos y Videojuegos a Nivel Global

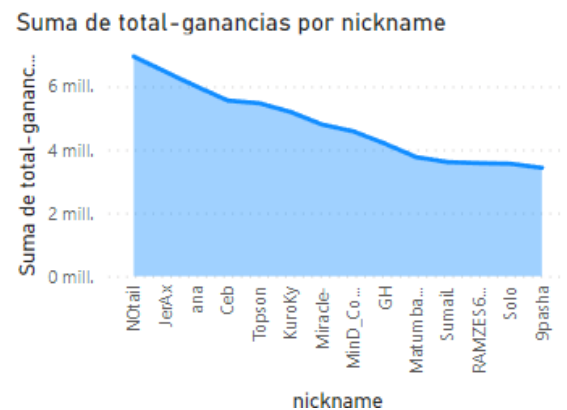
1.Grafica



3.Grafica



4.Grafica

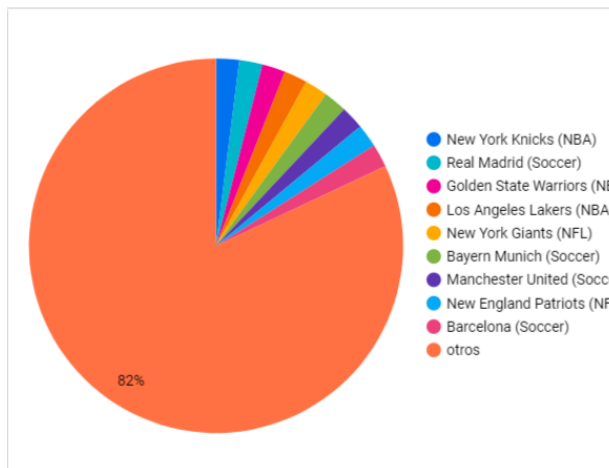


5.Grafica

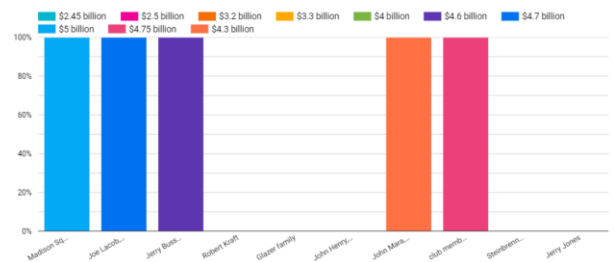


Evaluación del Valor: Clasificación de Eventos Deportivos a Escala Mundial

Grafica 1

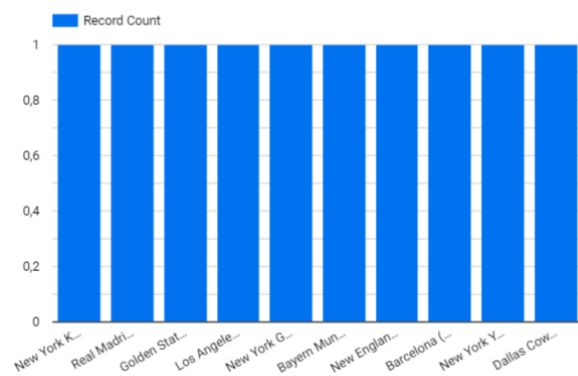


Grafica 2

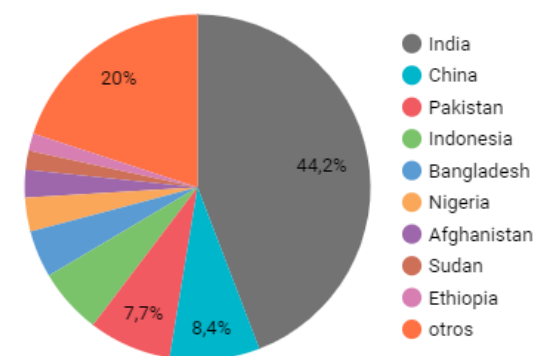


Estudio de Mortalidad: Tasas de Enfermedades Prevalentes a Nivel Internacional

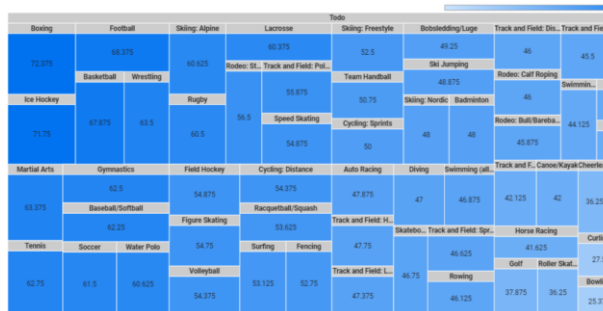
Grafica 1



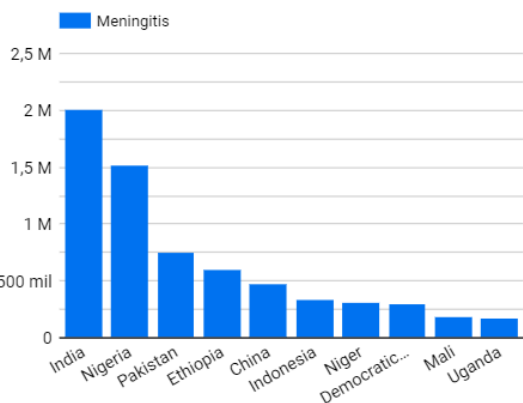
Grafica 3



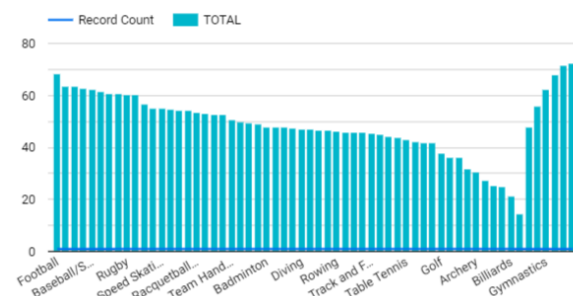
Grafica 2



Grafica 4



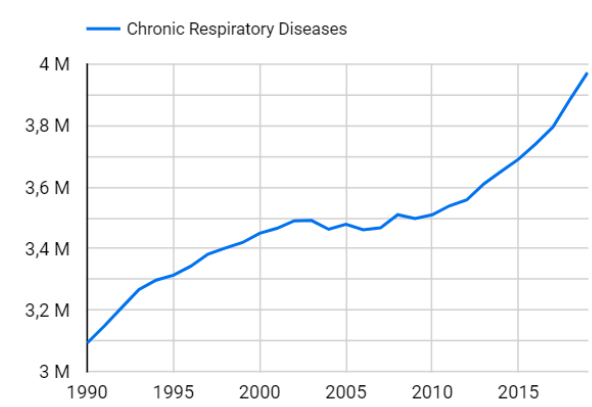
Grafica 3



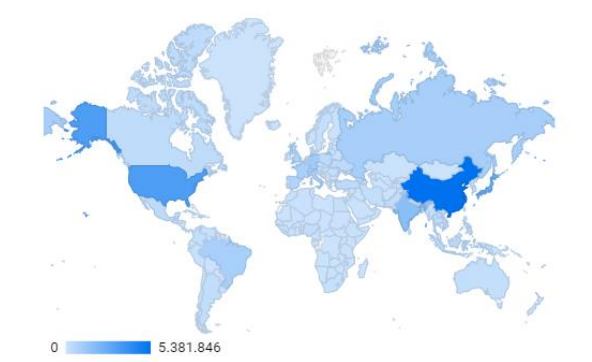
Grafica 5



Grafica 4



Grafica 5



Impacto del Consumo de Comida Rápida en la Salud: Un Análisis Profundo

Grafico 1:

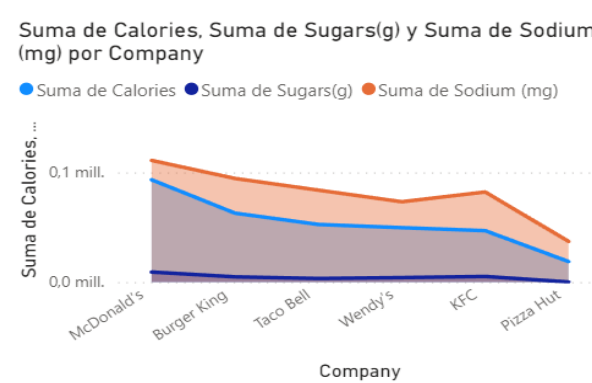


Grafico 2:

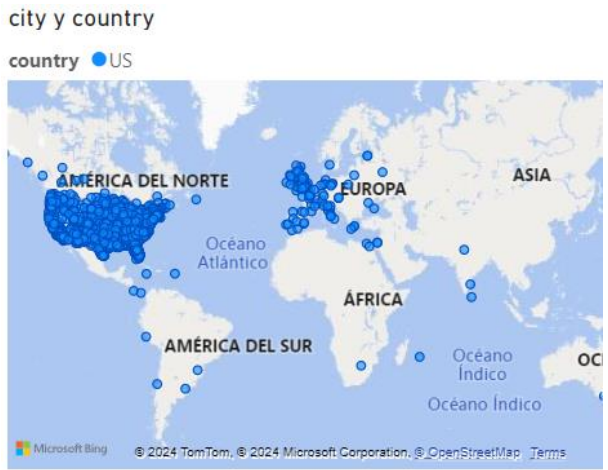


Grafico 3:

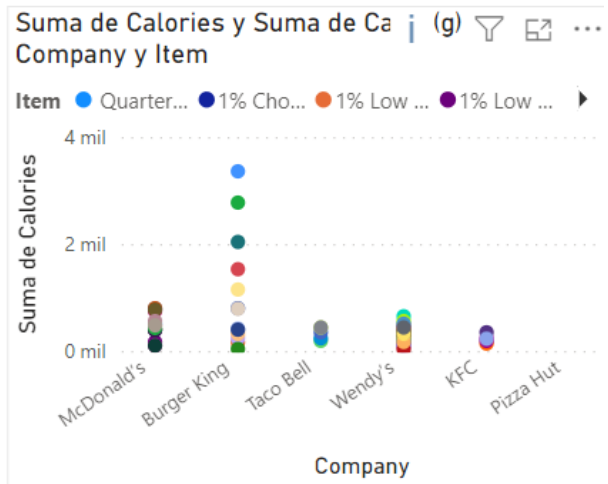


Grafico 4:

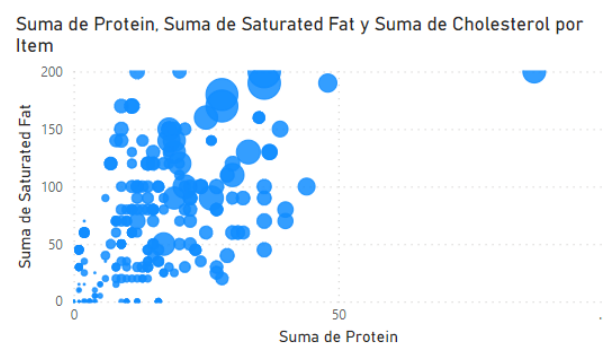
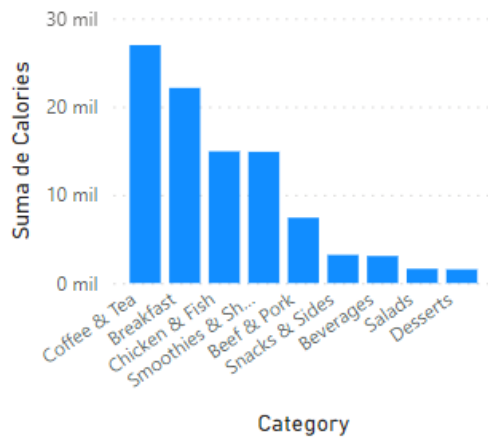


Grafico 5:

Suma de Calories por Category



Exploración de los World Tours de Taylor Swift: Un Estudio Exhaustivo

Grafico 1:

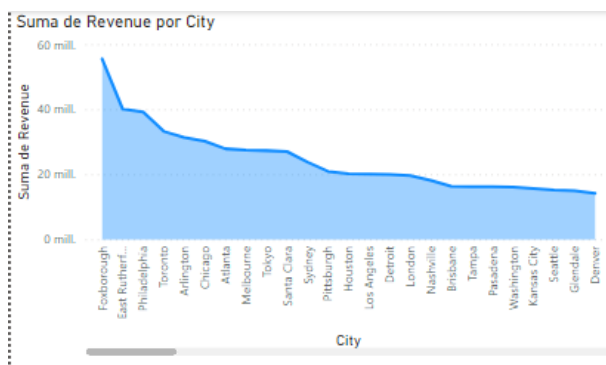


Grafico 2:



Grafico 3:

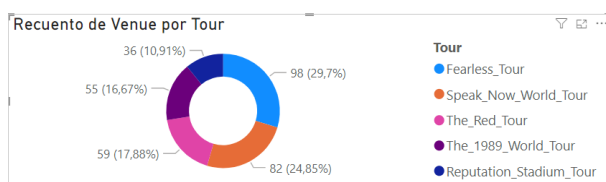
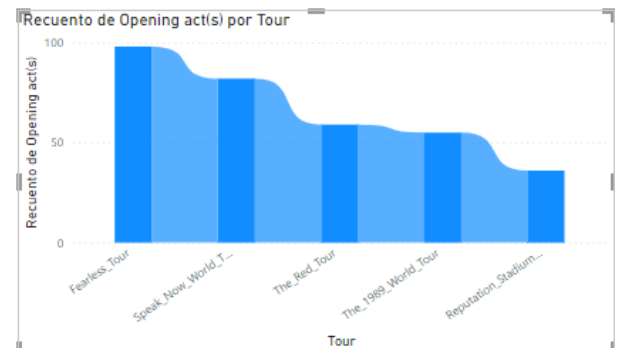


Grafico 4:

Suma de Revenue



Grafico 5:



X. Resultados obtenidos.

Análisis de Ganancias: Jugadores, Equipos y Videojuegos a Nivel Global

1-Grafica

Dominio Chino en Esports: China lidera en términos de ganancias en esports, seguido de cerca por Corea del Sur y Suecia. Los esports siguen siendo un campo emocionante y lucrativo para jugadores talentosos en todo el mundo.

2-Grafica

Diversidad Global en Esports: Los esports son diversos y globales, con países como China, Corea del Sur, Tailandia, Taiwán, Rusia, Francia, Suecia, Alemania, Estados Unidos y Dinamarca destacando en diferentes juegos. Los esports muestran un crecimiento prometedor con una amplia participación internacional.

3-Grafica

Preferencias de Géneros de Videojuegos: Los juegos en línea multijugador y de estrategia son los más populares, seguidos de Battle Royale, FPS y juegos de cartas coleccionables. La diversidad de gustos en la comunidad de

jugadores muestra la amplitud y profundidad de la industria de los videojuegos.

4-Grafica

Lucratividad de Dota 2: Dota 2 es extremadamente lucrativo para jugadores prominentes, con los cinco jugadores más exitosos en términos de ganancias individuales provenientes de este juego. Dota 2 ofrece algunas de las mayores oportunidades de ganancias en esports.

5-Grafica

Jerarquía en Ganancias por Juego: Dota 2 lidera como el juego más lucrativo para jugadores individuales, seguido de CS: GO y LoL. Los juegos mencionados muestran una jerarquía clara en términos de ganancias para los jugadores individuales en esports

Evaluación del Valor: Clasificación de Eventos Deportivos a Escala Mundial

Gráfico 1:

El gráfico muestra que la mayoría de los seguidores no se concentran en un solo equipo, sino que están distribuidos entre una variedad de equipos, con una gran proporción que sigue a equipos no especificados en el gráfico (“otros”).

Grafico 2:

El gráfico muestra que todos los equipos tienen el mismo conteo de registros. Esto podría indicar que los datos se recogieron de una manera que cada equipo sólo podría tener un registro, o que todos los equipos tuvieron el mismo número de eventos registrados en el conjunto de datos analizado.

Grafico 3:

El gráfico muestra que diferentes deportes y actividades físicas queman diferentes cantidades de calorías por hora. Esto puede ser útil para las personas que buscan maximizar la quema de calorías a través del ejercicio.

Grafico 4:

El gráfico muestra que el fútbol tiene la mayor cantidad de registros, mientras que la gimnasia tiene la menor cantidad de registros. Esto podría indicar que el fútbol es el deporte más popular o más frecuentemente registrado en este conjunto de datos, mientras que la gimnasia es el menos popular o menos frecuentemente registrado. Sin embargo, sin más detalles sobre el conjunto de datos, es difícil hacer afirmaciones definitivas.

Grafico5:

El gráfico muestra que Bernard Arnault tiene la mayor riqueza neta de los individuos mencionados, mientras que John F. Kennedy Jr. tiene la menor. Esto podría indicar que Bernard Arnault es el individuo más rico entre los mencionados en este gráfico, mientras que John F. Kennedy Jr. es el menos rico. Sin embargo, sin más detalles sobre el conjunto de datos, es difícil hacer afirmaciones definitivas.

Estudio de Mortalidad: Tasas de Enfermedades Prevalentes a Nivel Internacional

Los resultados obtenidos de los textos muestran una distribución de muertes debido a diferentes enfermedades a nivel mundial. La hepatitis aguda revela una alta incidencia en la India, representando el 44.2% de las muertes a nivel global, seguida por China con el 8.2%. Por otro lado, la meningitis causa un mayor impacto en la India y Nigeria, con 2 millones y 1.5 millones de muertes respectivamente. En cuanto al trastorno por consumo de alcohol, se registran 4.8 millones de muertes a nivel global, con causas que van desde lesiones por accidentes de tránsito hasta enfermedades cardiovasculares. Los trastornos respiratorios crónicos muestran un aumento progresivo en las muertes desde 1990, alcanzando 3.7 millones en 2015. Finalmente, el Alzheimer y otras enfermedades demenciales registran un alto número de muertes en la India, con 5.3 millones, seguida

por Estados Unidos y Canadá con entre 3.3 y 3.5 millones de muertes.

Impacto del Consumo de Comida Rápida en la Salud: Un Análisis Profundo

Los análisis revelan que McDonald's lidera en calorías, azúcares y sodio entre las principales cadenas de comida rápida, seguido por Burger King, Taco Bell, Wendy's, KFC y Pizza Hut. Además, se destaca una concentración densa de restaurantes de alto contenido calórico en América del Norte, Europa y ciertas regiones de Asia y América del Sur. La diversidad en los valores de calorías y calcio entre los productos de McDonald's, Burger King, Taco Bell, Wendy's, KFC y Pizza Hut sugiere una amplia variedad de opciones nutricionales dentro de estas cadenas. En cuanto al menú de McDonald's, se observa una predominancia de opciones con niveles más bajos de proteínas, grasa saturada y colesterol, aunque también hay productos con cantidades más altas. Este panorama indica una oferta diversa en términos de contenido nutricional, con preferencias hacia alimentos altos en calorías en ciertas categorías, como desayunos y platos de carne, mientras que las opciones más bajas en calorías se encuentran en postres, ensaladas y acompañamientos.

Exploración de los World Tours de Taylor Swift: Un Estudio Exhaustivo

En conclusión los conciertos realizados en los diferentes tour de Taylor Swift han dejado una gran cantidad de ingresos 932 millones para ser exactos, además que hubo un recibimiento de casi el 98% por la cantidad de tickets comprados con los disponibles, También se puede visualizar que existe una mayor concentración sobre todo en ciudades de estados unidos también vemos que no existe una gran diferencia entre los eventos realizados por tour lo que nos dé a interpretar que existe equivalencia entre los diferentes eventos.

XI. Conclusiones y Recomendaciones.

Impacto del Consumo de Comida Rápida en la Salud: La alta concentración de calorías, azúcares y sodio en los productos de cadenas de comida rápida como McDonald's, Burger King y Taco Bell plantea graves preocupaciones sobre la salud pública, especialmente en regiones como América del Norte, Europa y ciertas partes de Asia y América del Sur.

Exploración de los World Tours de Taylor Swift: Los conciertos de los World Tours de Taylor Swift han demostrado ser altamente lucrativos y con una alta tasa de asistencia, sin embargo, la concentración predominante de estos eventos en ciudades de Estados Unidos podría limitar el acceso global a su música.

Estudio de Mortalidad: Tasas de Enfermedades Prevalentes a Nivel Internacional: Las tasas de enfermedades prevalentes como hepatitis aguda, meningitis, trastornos por consumo de alcohol, trastornos respiratorios crónicos y enfermedades demenciales muestran disparidades significativas en todo el mundo, destacando la necesidad de enfoques específicos para abordar estas enfermedades en diferentes regiones.

Análisis de Ganancias: Jugadores, Equipos y Videojuegos a Nivel Global: China lidera en términos de ganancias en esports, lo que subraya la creciente importancia económica de los videojuegos competitivos a nivel mundial y la necesidad de una mayor atención a esta industria en términos de regulación y desarrollo.

Evaluación del Valor: Clasificación de Eventos Deportivos a Escala Mundial: La distribución de seguidores entre diferentes equipos, la igualdad en el conteo de registros por equipo y la variabilidad en la quema de calorías por deporte resaltan la diversidad y complejidad de los eventos deportivos a nivel mundial, lo que destaca la necesidad de un enfoque integral en su gestión y promoción.

XII. Desafíos y problemas encontrados

A la hora de buscar y descargar archivos CSV para iniciar el análisis, nos enfrentamos a dificultades significativas. La mayoría de los archivos disponibles contenían información poco relevante para nuestra investigación, lo que complicaba el proceso de selección. Además, nos encontramos con la frustración de que algunos archivos que contenían información valiosa estaban disponibles solo mediante pago, lo que nos obligaba a descartarlos como opción. Esta situación dificultó considerablemente nuestro acceso a datos pertinentes y significativos para nuestro análisis.

Una vez que habíamos preparado los archivos y los habíamos limpiado, nos encontramos con dificultades para cargarlos en MongoDB Atlas debido a incompatibilidades del sistema. Pasamos bastante tiempo intentando resolver este problema y finalmente llegamos a la conclusión de que debíamos instalar la versión más reciente de MongoDB Atlas para poder cargar los archivos con éxito.

Una vez que los sistemas estuvieron cargados, intentamos conectar dos visualizadores de datos: Power BI y Google Looker. Power BI resultó ser compatible con MongoDB Atlas, pero estaba en una fase temprana de demostración, lo que generaba cierta desconfianza debido a problemas de rendimiento durante el análisis de datos. Además, requería una extensión externa para funcionar correctamente. Existían dos extensiones disponibles: una especializada para MongoDB Atlas y otra para la conexión con data lakes convencionales. Decidimos probar primero la especializada para MongoDB Atlas, pero al intentar instalarla, nos encontramos con un error de compatibilidad inesperado, lo cual resultó extraño dado que la extensión era distribuida por MongoDB. Al probar con la otra extensión, nos encontramos con el mismo problema de compatibilidad. La solución fue instalar ambas extensiones para que funcionara correctamente. Durante nuestra búsqueda de

soluciones, también intentamos utilizar Google Looker, pero descubrimos que aún no tenía compatibilidad con este tipo de programas, por lo que descartamos su uso en ese momento.

XIII. Link de GitHub del proyecto.

<https://github.com/YuverlyHidokun/Proyecto-Final-Analisis-de-Datos/tree/master>