

Aplicación de análisis de datos

Autores:

- Eddy Morales
- Anthon Chang
- Paulo Cisneros

Fecha: lunes 3 de febrero del 2025

1. Definición del Caso de Estudio

Este caso de estudio analiza diversos datos provenientes de múltiples fuentes con el objetivo de obtener una visión integral de los diferentes factores que impactan en el área de interés. A través del cruce de información, se pretende identificar patrones, correlaciones y áreas de mejora, proporcionando una base sólida para la toma de decisiones y optimización de procesos.

2. Objetivos

2.1 Objetivo General

Analizar los datos provenientes de diversas fuentes para identificar patrones, correlaciones y áreas de mejora que contribuyan a optimizar procesos y tomar decisiones informadas.

2.2 Objetivos Específicos

1. Identificar las variables clave que afectan los procesos analizados.
2. Evaluar cómo las distintas fuentes de datos interactúan y contribuyen al análisis general.
3. Proponer acciones y estrategias basadas en los hallazgos obtenidos a partir de los datos analizados.

3. Descripción del Equipo de Trabajo y Actividades Realizadas

Equipo de Trabajo:

- Eddy Morales
- Anthon Chang
- Paulo Cisneros

Actividades:

1. Eddy Morales: Agreguen lo que hicieron
2. Anthon Chang: Limpieza de datos, la conexión al MySQL y al MongoDB y la creación de los dashboards.
3. Paulo Cisneros: Búsqueda de fuentes datos, limpieza de datos, Importación y exportación de datos en las bases de datos, creación de dashboards.

4. Cronograma de Actividades

Diagrama de Gantt:

Actividad	Lun 27	Mié 29	Jue 30	Vie 31	Sáb 1	Dom 2
Planificación del Proyecto						
Extraer Datasets						
Depuración de los Datos						
Exportar e Importar Datos						
Análisis de Sentimientos						
Realizar Dashboard						
Realización del Informe						
Presentación y Vídeos Expositivos						

5. Recursos y Herramientas Utilizadas

- **Recursos Humanos:**
 - Integrantes del grupo.
- **Herramientas Tecnológicas:**
 - Python para procesamiento de datos (Jupyter Notebook).
 - MongoDB como base de datos.
 - MySQL Workbench como base de datos.
 - SQL Server como base de datos.
 - CouchDB como base de datos.
 - Power BI para visualizaciones.

6. Arquitectura de la Solución

La arquitectura está hecha con las siguientes bases de datos SQL (MySQL Workbench, SQL Server) y NoSQL (MongoDB y CouchDB), el procesamiento de datos se hará en Python y visualización en Power BI.

Diagrama de Arquitectura

7. Extracción de Datos

Los datos fueron extraídos de la página web de “Kaggle”, que contiene data sets (.csv o .json) de la mayoría de los datos que encontramos. Y el tema de “Películas y series” fue extraído de una página web (https://www.rottentomatoes.com/m/joker_2019/reviews).

Luego se les paso a las bases de datos:

Hobbies

Datos en mysql:

Gender	Department	HeightCM	WeightKG	BMI	IQ	College mark	Hobbies	daily studying time	prefer to study in	salary expectation
Male	BCA	180	80	25	120	80	Video Games	0 - 30 minute	Evening	45000
Female	BCA	160	45	18	80	70	Cinema	30 - 60 minute	Morning	15000
Male	BCA	170	70	24	110	100	Cinema	1 - 2 hour	Anytime	12000
Female	BCA	140	30	16	90	80	Reading books	1 - 2 hour	Anytime	150000
Male	BCA	170	60	21	100	90	Video Games	30 - 60 minute	Evening	30000
Female	BCA	150	30	13	90	70	Cinema	30 - 60 minute	Night	20000
Male	BCA	180	80	25	120	80	Reading books	1 - 2 hour	Evening	10000
Male	BCA	180	40	16.7	110	70	Sports	1 - 2 hour	Evening	20000

Datos en couchdb:

Gender	hobbies	daily studying	Department	part-time job
<input type="checkbox"/> Male	Video Games	0 - 30 minute	BCA	No
<input type="checkbox"/> Female	Cinema	30 - 60 minute	BCA	No
<input type="checkbox"/> Male	Cinema	1 - 2 Hour	BCA	No
<input type="checkbox"/> Female	Reading books	1 - 2 Hour	BCA	No
<input type="checkbox"/> Male	Video Games	30 - 60 minute	BCA	No
<input type="checkbox"/> Female	Cinema	30 - 60 minute	BCA	No
<input type="checkbox"/> Male	Reading books	1 - 2 Hour	BCA	No
<input type="checkbox"/> Male	Sports	1 - 2 Hour	BCA	No
<input type="checkbox"/> Male	Video Games	0 - 30 minute	BCA	No
<input type="checkbox"/> Male	Video Games	0 - 30 minute	BCA	No
<input type="checkbox"/> Male	Video Games	1 - 2 Hour	BCA	No
<input type="checkbox"/> Female	Cinema	2 - 3 Hour	BCA	No
<input type="checkbox"/> Female	Reading books	30 - 60 minute	BCA	No
<input type="checkbox"/> Male	Cinema	0 - 30 minute	BCA	No

Videojuegos:

En mysql

game	platform	developer	genre	number_players	rating	release_date	positive_critics	negative_critics
Final 2	PC	Video Software	Action	1	8.5	Mar 10, 2011	15	0
The Elder Scrolls V: Skyrim	PC	Bethesda Game Studios	Role Playing	No Online Multiplayer	M	Nov 10, 2011	32	0
The Legend of Zelda: Breath of the Wild	Wii U	Nintendo	Adventure	No Online Multiplayer	E10+	Mar 10, 2011	84	1
Assassin's Creed City	PC	Ubisoft Montreal	Action Adventure	No Online Multiplayer	T	Nov 21, 2011	27	0
Super Mario 3D World	Wii U	Nintendo	Action	No Online Multiplayer	E	Nov 10, 2011	81	1
Overwatch	PC	Blizzard Entertainment	Action	No Online Multiplayer	M	Aug 23, 2011	52	0
Call of Duty: Warzone	PC	Infinity Ward	First Person Shooter	No Online Multiplayer	E	Mar 10, 2011	71	0
Fortnite	PC	Epic Games	Strategy	Online Multiplayer	T	Mar 10, 2011	42	0
Call of Duty: Warzone	PC	Epic Games	Strategy	Online Multiplayer	E	Mar 10, 2011	14	0

En couch db:

Table

Metadata

{ } JSON

	game	developer	release_date	genre	metascore
<input type="checkbox"/>	Portal 2	Valve Software	Apr 18, 2011	Action	95
<input type="checkbox"/>	The Elder Scrolls V: Skyrim	Bethesda Game Studios	Nov 10, 2011	Role-Playing	94
<input type="checkbox"/>	The Legend of Zelda: Ocarina of Time	GREZZO	Jun 19, 2011	Miscellaneous	94
<input type="checkbox"/>	Batman: Arkham City	Rocksteady Studios	Nov 21, 2011	Action Adventure	91
<input type="checkbox"/>	Super Mario 3D Land	Nintendo	Nov 13, 2011	Action	90
<input type="checkbox"/>	Deus Ex: Human Revolution	Nixxes Software	Aug 23, 2011	Action	90
<input type="checkbox"/>	Pushmo	Intelligent Systems	Dec 8, 2011	Miscellaneous	90
<input type="checkbox"/>	Total War: Shogun 2	Creative Assembly	Mar 15, 2011	Strategy	90
<input type="checkbox"/>	FIFA Soccer 12	Electronic Arts	Sep 27, 2011	Sports	89
<input type="checkbox"/>	Battlefield 3	EA DICE	Oct 25, 2011	Action	89
<input type="checkbox"/>	LIMBO	PLAYDEAD	Aug 2, 2011	Action	88
<input type="checkbox"/>	Assassin's Creed: Brotherhood	Ubisoft Montreal	Mar 22, 2011	Action Adventure	88
<input type="checkbox"/>	The Witcher 2: Assassins of Kings	CD Projekt Red Studio	May 17, 2011	Role-Playing	88
<input type="checkbox"/>	Dead Space 2	Visceral Games	Jan 25, 2011	Action	87

Noticias:

Table Metadata JSON					
_id	category	headline	date	short_descrt.	
<input type="checkbox"/>	ENTERTAINMENT	Justin Bieber And Lionel Richie Perform 'I Wanna Dance With Somebody' At The 2016 VMAs	2016-09-13	Hello? Goodbye.	
<input type="checkbox"/>	THE WORLDPOST	Airstrike On Yemen School Kills 10, Wounds 20	2016-09-13	Aid group Doctors Without Borders says it's 'a tragedy'	
<input type="checkbox"/>	WOMEN	20 Reasons We're Terrible At Being Single	2016-09-13	Tonight, my husband and I...	
<input type="checkbox"/>	HEALTHY LIVING	Diet, Doubts, And Dough: Why We're Not Losing Weight	2016-09-13	No, we are not, absolutely...	
<input type="checkbox"/>	QUEER VOICES	Two Months After Pulse A...	2016-09-13	It's sad to think that it coul...	
<input type="checkbox"/>	SPORTS	The Olympics: More Role...	2016-09-13	I have been obsessed wit...	
<input type="checkbox"/>	IMPACT	Foster Care vs Adoption: ...	2016-09-13	The Queensland Governm...	
<input type="checkbox"/>	LATINO VOICES	From NPR's Latino USA: ...	2016-09-13	Elizabeth Ramirez and he...	
<input type="checkbox"/>	PARENTS	Why Ignoring My Son Wa...	2016-09-13	Today, like many of our da...	
<input type="checkbox"/>	POLITICS	Why Donald Trump Is Sca...	2016-09-13	US presidential candidate...	
<input type="checkbox"/>	SPORTS	Sole Russian Track And F...	2016-09-13	Darya Kizhina is appea...	
<input type="checkbox"/>	SPORTS	Irish Rowers Deserve Gol...	2016-09-13	Brothers Paul and Gary O...	
<input type="checkbox"/>	PARENTS	Saying Goodbye: A Mexica...	2016-09-13	This has been a very emo...	
<input type="checkbox"/>	SPORTS	Cowboys! Deceit Proves R...	2016-09-13	Roger Goodell, commissio...	

Mongodb :

Reseñas:

Type a query (e.g. 'value') or generate query					
id	review	pre_review			
16	ObjectID('4796773a8c39628bc2544')	921245	pre_review: "presentacion servicio exageradamente costoso infuener amita carta llo ..."		
16	ObjectID('4796773a8c39628bc2544')	921245	pre_review: "presentacion plato calidad alimento vino experiencia recomendo"		
16	ObjectID('4796773a8c39628bc2544')	921247	pre_review: "frecuentame algo querido peru visitar statera emocionar probar fabul..."		
16	ObjectID('4796773a8c39628bc2544')	921245	pre_review: "excelente experiencia comer maravilloso excelente insume perfecto mas..."		
16	ObjectID('4796773a8c39628bc2544')	921245	pre_review: "poslon viajar disfrutar grandezza pais sobretodo aprecio gastronomia pa..."		
16	ObjectID('4796773a8c39628bc2544')	921238	pre_review: "vacacione fantastico disfrutar familia pareja reunion negocio perfecta..."		

Liga española:

Type a query (e.g. 'value') or generate query					
id	review	pre_review			
16	ObjectID('4796773a8c39628bc2544')	921245	pre_review: "presentacion servicio exageradamente costoso infuener amita carta llo ..."		
16	ObjectID('4796773a8c39628bc2544')	921245	pre_review: "presentacion plato calidad alimento vino experiencia recomendo"		
16	ObjectID('4796773a8c39628bc2544')	921247	pre_review: "frecuentame algo querido peru visitar statera emocionar probar fabul..."		
16	ObjectID('4796773a8c39628bc2544')	921245	pre_review: "excelente experiencia comer maravilloso excelente insume perfecto mas..."		
16	ObjectID('4796773a8c39628bc2544')	921245	pre_review: "poslon viajar disfrutar grandezza pais sobretodo aprecio gastronomia pa..."		
16	ObjectID('4796773a8c39628bc2544')	921238	pre_review: "vacacione fantastico disfrutar familia pareja reunion negocio perfecta..."		

Mysql: Netflix

Mysql COVID:

```
1 SELECT * FROM covid_19_data;
```

Mongo COVID:

```
_id: ObjectId('679da147f5bd746e@fb1c947')
index: 0
Country/Region: "Afghanistan"
Confirmed: 36263
Deaths: 1269
Recovered: 25198
Active: 19796
New cases: 106
New deaths: 10
New recovered: 18
Deaths / 100 Cases: 3.5
Recovered / 100 Cases: 69.49
Deaths / 100 Recovered: "5.04"
Confirmed last week: 35526
1 week change: 737
1 week % increase: 2.07
WHO Region: "Eastern Mediterranean"
```

```
_id: ObjectId('679da147f5bd746e@fb1c948')
index: 1
Country/Region: "Albania"
Confirmed: 4880
Deaths: 144
Recovered: 2745
Active: 1991
New cases: 117
New deaths: 6
New recovered: 63
```

Mysql NBA:

```
1 SELECT * FROM nba_players;
```

Mongo NBA:

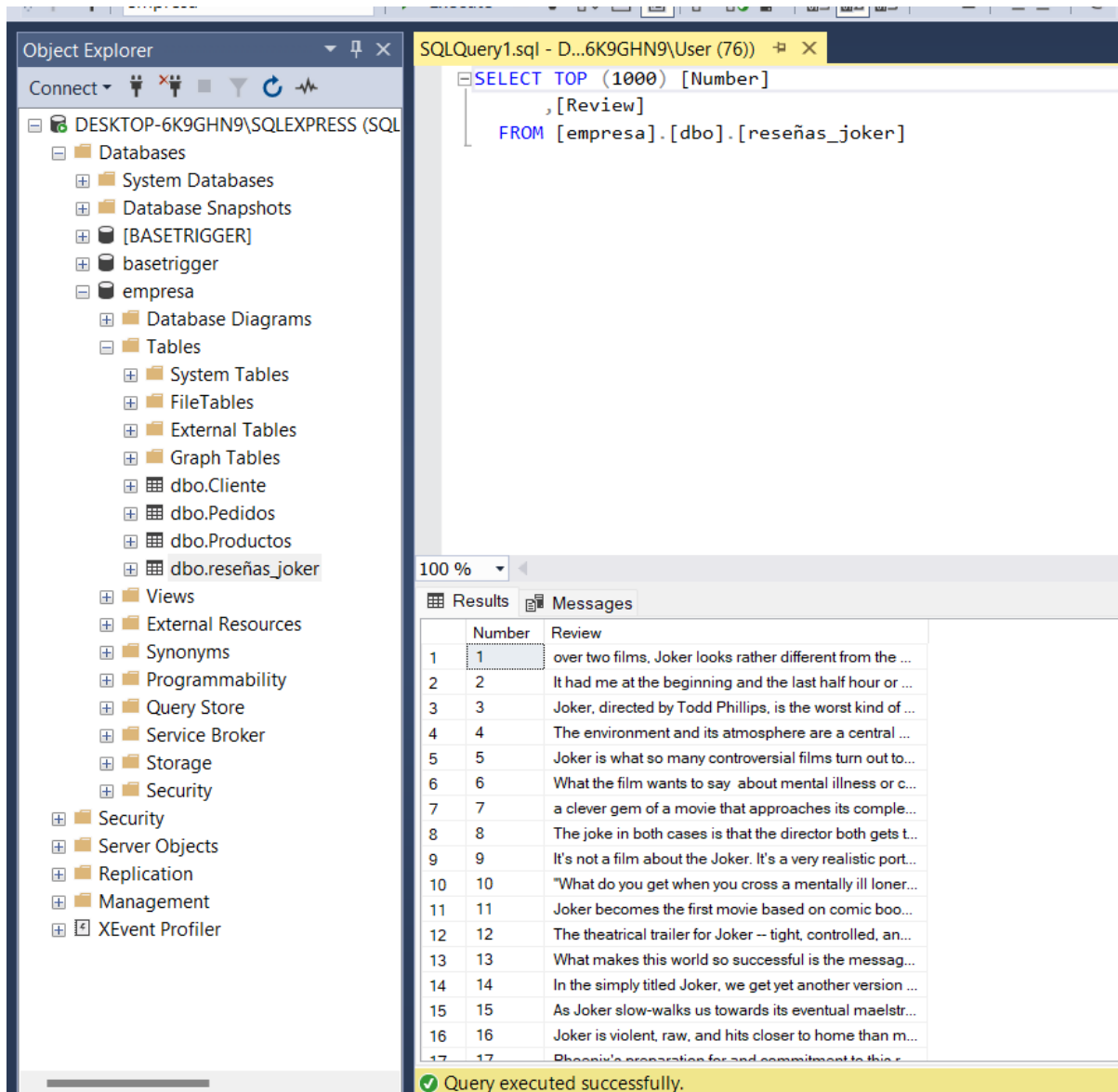
```
_id: ObjectId('679db2a7f5bd746e0fb1c104')
Unnamed: 0: 1
player_name: "Gaylon Nickerson"
team_abbreviation: "WAS"
```

```
1 • SELECT * FROM starbucks, t1order;
```

Mongo STARSBUCKS:


```
id: ObjectID("679db9c9f5bd746eefb1f331")
Brand: "Starbucks"
Store Number: "22331-212325"
Store Name: "Ajman Drive Thru"
Ownership Type: "Licensed"
Street Address: "11 Street 69, Al Jarf"
City: "Ajman"
State/Province: "AJ"
Country: "AE"
Postcode: "000000"
Phone Number: "No disponible"
Timezone: "GMT+04:00 Asia/Dubai"
Longitude: "85.47"
```

Sql server:



```
SELECT TOP (1000) [Number]
, [Review]
FROM [empresa].[dbo].[resenñas_joker]
```

	Number	Review
1	1	over two films, Joker looks rather different from the ...
2	2	It had me at the beginning and the last half hour or ...
3	3	Joker, directed by Todd Phillips, is the worst kind of ...
4	4	The environment and its atmosphere are a central ...
5	5	Joker is what so many controversial films turn out to...
6	6	What the film wants to say about mental illness or c...
7	7	a clever gem of a movie that approaches its comple...
8	8	The joke in both cases is that the director both gets t...
9	9	It's not a film about the Joker. It's a very realistic port...
10	10	"What do you get when you cross a mentally ill loner...
11	11	Joker becomes the first movie based on comic boo...
12	12	The theatrical trailer for Joker -- tight, controlled, an...
13	13	What makes this world so successful is the messag...
14	14	In the simply titled Joker, we get yet another version ...
15	15	As Joker slow-walks us towards its eventual maelstr...
16	16	Joker is violent, raw, and hits closer to home than m...
17	17	Phoenix's preparation for and commitment to this...

Query executed successfully.

8. Análisis de Información

Se aplicaron técnicas de limpieza de datos para eliminar duplicados y valores nulos en cada uno de los dataframes.

A continuación, se adjuntan algunas capturas de pantalla de cómo se limpiaron cada uno de los dataframes.

Hobbies:



```
Limpieza de datos archivo Hobbies

#Importación de librerías
import pandas as pd
import numpy as np
import json

df_csv = pd.read_csv('C:/Users/Personal/OneDrive/Documents/archivos de programación/hoyter/datos/Student Attitude and Behavior.csv')
```

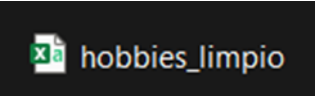
Predictor variables: <code>df_csv.head()</code>													Python			
Certification	Course	Gender	Department	Height(CM)	Weight(KG)	10th Mark	12th Mark	college mark	hobbies	daily study time	prefer study in	salary expected	Do you like your degree?	chances to pursue a career based on their degree	social media	Tr
0	No	Male	BCA	100.0	58.0	79.0	64.0	80.0	Video Games	8-10 minute	Morning	40000	No	50%	1-1.5 hour	
1	No	Female	BCA	90.0	40.0	70.0	80.0	70.0	Cinema	30-40 minute	Morning	15000	Yes	75%	1-1.5 hour	
2	Yes	Male	BCA	159.0	78.0	71.0	69.0	55.0	Cinema	1-2 Hour	Anytime	13000	Yes	50%	More than 2 hour	
3	Yes	Female	BCA	147.0	20.0	70.0	58.0	58.0	Reading books	1-2 Hour	Anytime	1500000	No	50%	1-3.0-2 hour	
4	No	Male	BCA	170.0	54.0	40.0	65.0	30.0	Video Games	10-15 minute	Morning	50000	Yes	25%	1-3.0-2 hour	

```

# Retrieve information of the data
df_csv.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 216 entries, 0 to 214
Data columns (total 19 columns):
 #   Column                Non-Null Count  Dtype  
---  -
 0   Certification Course   216 non-null    object 
 1   Gender                 216 non-null    object 
 2   Department             216 non-null    object 
 3   Height(cm)            216 non-null    float64 
 4   Weight(kg)            216 non-null    float64 
 5   Intro Mark            216 non-null    float64 
 6   12th Mark             216 non-null    float64 
 7   College Mark          216 non-null    float64 
 8   Hobbies               216 non-null    object 
 9   Daily Studing time    216 non-null    object 
10   prefer to study in    216 non-null    object 
11   salary expectation     216 non-null    int64  
12   Do you like your degree? 216 non-null    object 
13   willingness to pursue a career based on their degree 216 non-null    object 
14   social media & video   216 non-null    object 
15   travelling time       216 non-null    object 
16   stress level          216 non-null    object 
17   financial status      216 non-null    object 
18   part-time job         216 non-null    object 
dtypes: float64(5), int64(1), object(11)
memory usage: 35.4b KB

```



Archivo limpio:

Videojuegos:

[illegible]


```
#Obtener informacion de los datos
df2_csv.info()

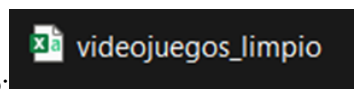
✓ 0.0s

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5699 entries, 0 to 5698
Data columns (total 15 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   game                   5699 non-null   object
1   platform               5699 non-null   object
2   developer              5685 non-null   object
3   genre                  5694 non-null   object
4   number_players         3059 non-null   object
5   rating                 4433 non-null   object
6   release_date           5699 non-null   object
7   positive_critics       5699 non-null   int64
8   neutral_critics        5699 non-null   int64
9   negative_critics       5699 non-null   int64
10  positive_users         5699 non-null   int64
11  neutral_users          5699 non-null   int64
12  negative_users         5699 non-null   int64
13  metascore              5699 non-null   int64
14  user_score             5699 non-null   int64
dtypes: int64(8), object(7)
memory usage: 668.0+ KB
```

```
#Eliminar filas duplicadas
df2_csv = df2_csv.drop_duplicates()

✓ 0.0s

#Guardar en un archivo csv
df2_csv.to_csv('C:/Users/Personal/OneDrive/Documents/Archivos de programación/Jupyter/datos/videojuegos_limpio.csv', index=False)
```



Este es el archivo que quedo:

Covid:

```
covid.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 187 entries, 0 to 186
Data columns (total 15 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   Country/Region        187 non-null   object
1   Confirmed              187 non-null   int64
2   Deaths                187 non-null   int64
3   Recovered              187 non-null   int64
4   Active                187 non-null   int64
5   New cases              187 non-null   int64
6   New deaths             187 non-null   int64
7   New recovered          187 non-null   int64
8   Deaths / 100 Cases    187 non-null   float64
9   Recovered / 100 Cases 187 non-null   float64
10  Deaths / 100 Recovered 187 non-null   float64
11  Confirmed last week    187 non-null   int64
12  1 week change          187 non-null   int64
13  1 week % increase      187 non-null   float64
14  WHO Region            187 non-null   object
dtypes: float64(4), int64(9), object(2)
memory usage: 22.0+ KB
```

```
covid.isnull().sum()
```

```
Country/Region      0
Confirmed            0
Deaths              0
Recovered           0
Active              0
New cases            0
New deaths           0
New recovered        0
Deaths / 100 Cases  0
Recovered / 100 Cases 0
Deaths / 100 Recovered 0
Confirmed last week  0
1 week change        0
1 week % increase    0
WHO Region           0
dtype: int64
```

Archivo limpio:



COVID.csv

Nba:

```
nba_csv['college'].fillna('Otro', inplace=True)
```

```
nba_csv.to_csv('NBA.csv', index=False)
```

```
nba_csv.isnull().sum()
```

```
Unnamed: 0          0
player_name         0
team_abbreviation   0
age                 0
player_height       0
player_weight       0
college             0
country             0
draft_year          0
draft_round         0
draft_number        0
gp                  0
pts                 0
reb                 0
ast                 0
net_rating          0
oreb_pct            0
dreb_pct            0
usg_pct             0
ts_pct              0
ast_pct             0
season              0
dtype: int64
```

Archivo limpio:



NBA.csv

Starbucks:

```
starbucks_csv['Street Address'].fillna('Desconocido', inplace=True)
```

```
starbucks_csv['City'].fillna('Ciudad Desconocido', inplace=True)
```

```
starbucks_csv['Postcode'].fillna('00000', inplace=True)
```

```
starbucks_csv['Phone Number'].fillna('No disponible', inplace=True)
```

```
starbucks_csv['Longitude'].fillna('0.0', inplace=True)
```

```
starbucks_csv['Latitude'].fillna('0.0', inplace=True)
```

```
starbucks_csv.to_csv('STARBUCKS.csv', index=False)
```

Archivo limpio:



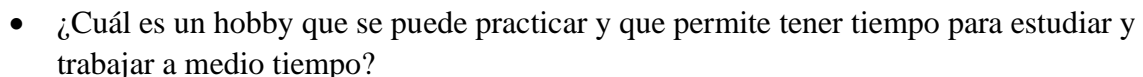
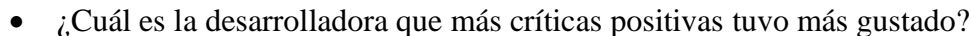
STARBUCKS.csv

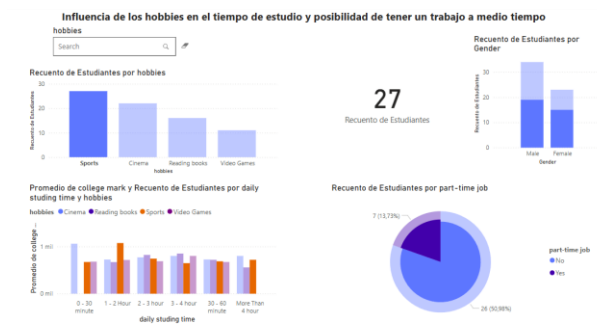
```

10 import requests
11 from bs4 import BeautifulSoup
12
13 url = "https://www.rottentomatoes.com/actor/joker_2019/reviews/"
14
15 # Encuentra una página con reviews de un navegador
16 headers = {}
17
18 """User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/78.0.4092.125 Safari/537.36"""
19
20 # Realiza la solicitud HTTP con encabezados
21 response = requests.get(url, headers=headers)
22
23 # Verifica si la solicitud fue exitosa
24 if response.status_code == 200:
25     html_content = response.content
26
27     # print("Error al acceder a la página: {}".format(response.status_code))
28     exit()
29
30 # Analiza el contenido HTML con BeautifulSoup
31 soup = BeautifulSoup(html_content, "html.parser")
32
33 # Extrae las reseñas usando el selector correcto
34 reviews = soup.find_all("p", class_="review-text", attrset={"data-src": "review-media"})
35
36 # Iterar las reseñas
37 for review in reviews:
38     # print("Reseña: {}".format(review.get("review-text")))
39     """User-Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/78.0.4092.125 Safari/537.36"""
40
41     # print("Se encontraron reseñas.")

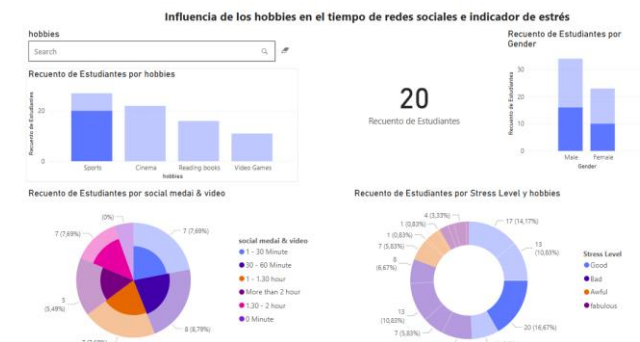
```

- ¿Cuál es la desarrolladora que más críticas negativas tuvo y cuál es su juego más criticado?



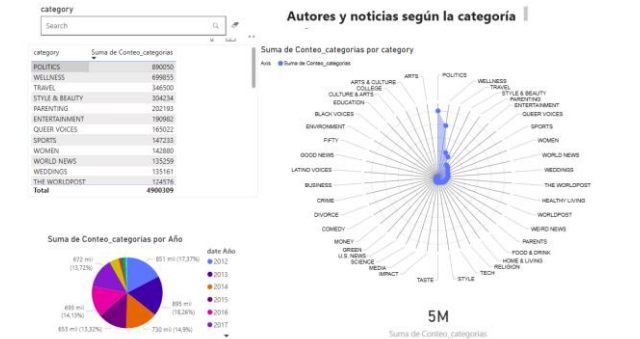


- ¿Cómo influyen los hobbies que practicamos en la reducción del estrés y cómo pueden servir como indicadores de bienestar?



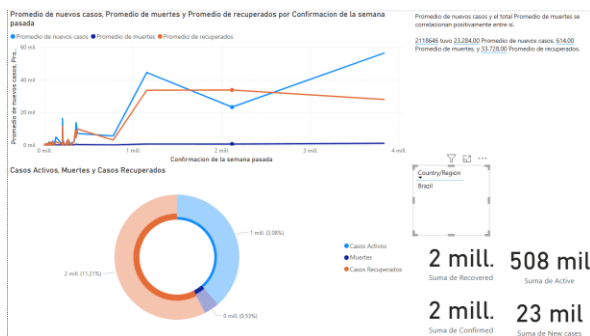
Noticias mundiales

- ¿Cuáles son las tendencias mundiales en las noticias?



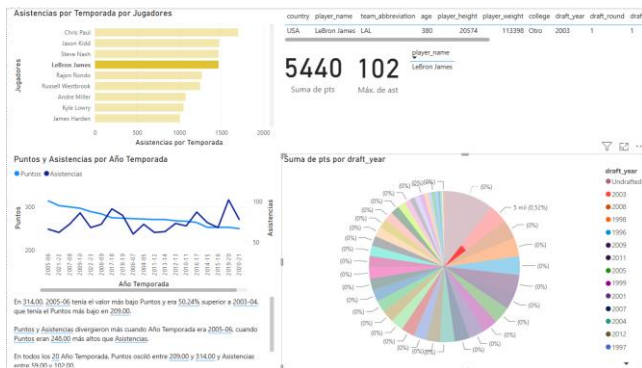
Covid

- ¿Cuál es la cantidad de casos confirmados, muertes y nuevos casos de COVID-19 en Brasil?

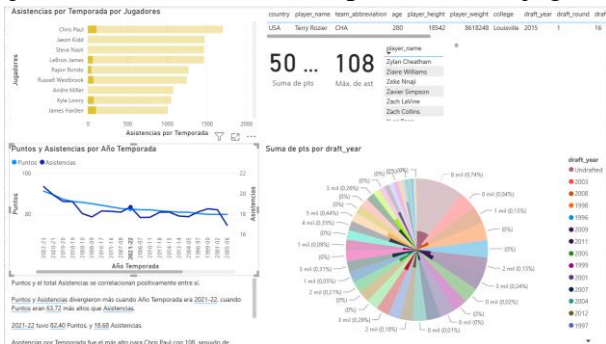


NBA

- ¿Cuál ha sido el mejor año de Lebrón James en términos de puntos y asistencias?

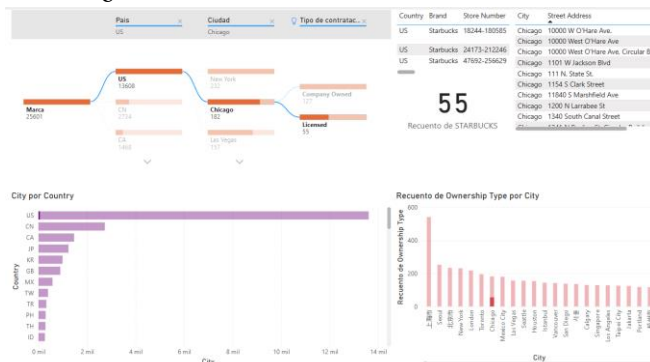


- ¿Cuál fue la suma total de puntos de los jugadores en la temporada 2021?



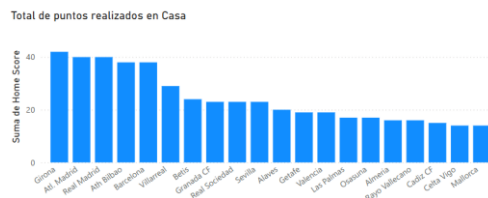
Starbucks

- ¿Cuál es el número total de locales de Starbucks con licencia en Chicago?



Liga española:

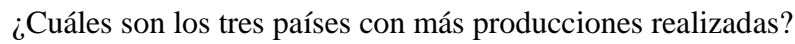
- ¿Cuáles son los puntos obtenidos por los equipos en Casa?



- ¿Cuáles son los puntos obtenidos por los equipos de visitante?



¿Cuáles son los tres directores con mayor cantidad de producciones realizadas?



¿Cuáles son las 5 palabras más frecuentes?:



¿Cuáles son las palabras más frecuentes?:



Videojuegos

- ¿Cuál es la desarrolladora que más críticas negativas tuvo y cuál es su juego más criticado?

Con los resultados que obtuvimos la empresa con mayores críticas negativas es “Ubisoft” con un total de 124 y usuarios negativos son 583, el juego de esta desarrolladora que obtuvo la mayoría de las críticas negativas fue “Figther within”.

- ¿Cuál es la desarrolladora que más críticas positivas tuvo más gustado?

Según el análisis podemos saber que una extensión de Ubisoft logro tener las mejores críticas positivas esta es “Ubisoft Montreal” con un total de 1017 de críticas positivas y 3414 de usuarios positivos, el juego más gustado según la información es el “Far Cry 5”.

Hobbies y actividades

- ¿Cuál es un hobby que se puede practicar y que permite tener tiempo para estudiar y trabajar a medio tiempo?

Basándonos en la información obtenida podemos decir que el deporte es la afición que les permite tener tiempo para otras cosas por ejemplo al estudiar los que practican deportes toman de 1 hora a 4 horas o más y el porcentaje de estos estudiantes que tiene trabajo a medio tiempo es de 50,98%.

- ¿Cómo influyen los hobbies que practicamos en la reducción del estrés y cómo pueden servir como indicadores de bienestar?

Según vemos en el gráfico podemos notar que la afición que tiene más porcentaje de nivel de estrés en “bueno” es deportes con 16,67 % a si mismo esta posee un el mayor porcentaje en “fabuloso” con un 3,33 % lo que nos ayuda a concluir que si queremos tener muy bajo el estrés el deporte es la mejor manera de hacerlo.

Noticias mundiales

- ¿Cuáles son las tendencias mundiales en las noticias?

El mayor tema del que se habla en las noticias es la “Política” con un total de 890 mil de este tema, siguiéndole el “Bienestar” con 699 mil y en tercer lugar los “Viajes” con 346 mil.

En comparación con estas las menos tendencias es “Arte” con un de 1509.

Covid:

- ¿Cuál es la cantidad de casos confirmados, muertes y nuevos casos de COVID-19 en Brasil?

En Brasil se han registrado aproximadamente 2 millones de casos confirmados de COVID-19. Además, el número de muertes asciende a 88 mil, mientras que los casos activos son 508 mil.

NBA:

- ¿Cuál ha sido el mejor año de Lebrón James en términos de puntos y asistencias?

En 2003, durante su temporada de debut en la NBA, LeBron James logró un máximo de 43 puntos en un partido, destacándose por su capacidad anotadora desde el inicio. En total, acumuló 102 asistencias a lo largo de esa temporada. Estos datos reflejan su capacidad para anotar y asistir, aunque fueron solo un indicio del rendimiento más destacado que alcanzaría en sus años posteriores.

- ¿Cuál fue la suma total de puntos de los jugadores en la temporada 2021?

En la temporada 2021 de la NBA, la suma total de puntos anotados por todos los jugadores en conjunto fue de aproximadamente 50,000 puntos. Este dato refleja la magnitud del rendimiento ofensivo colectivo durante la temporada, destacando la contribución individual de cada jugador al total, así como el alto ritmo de juego característico de la liga en ese periodo.

STARBUCKS:

- ¿Cuál es el número total de locales de Starbucks con licencia en Chicago?

En Chicago, Starbucks cuenta con un total de 55 locales operados bajo el modelo de licencia. Este número refleja la presencia significativa de la cadena en la ciudad, destacando su expansión a través de asociaciones con licenciatarios locales que gestionan las tiendas bajo los estándares de la marca.

Liga española:

- ¿Cuáles son los puntos obtenidos por los equipos en Casa?
En este caso el equipo con más puntos obtenidos en casa es Girona y el equipo con menos puntos realizados en casa es el Mallorca.

- ¿Cuáles son los puntos obtenidos por los equipos de visitante?

En este caso el equipo con más puntos obtenidos de visitante es el Real Madrid y el equipo con menos puntos realizados de visitante es el Cadiz FC.

Netflix:

- ¿Cuáles son los tres directores con mayor cantidad de producciones realizadas?
 - Raul C.
 - Marcus R.
 - Jay K.
- ¿Cuáles son los tres países con más producciones realizadas?
 - United States
 - India
 - United Kingdom

Reseña Comida Peruana:

- ¿Cuáles son las 5 palabras más frecuentes?

Las 5 palabras más frecuentes: comer, atender, recomendar, delicia y servicio.

11. Conclusiones y Recomendaciones

- **Conclusiones:**

El proyecto de recopilación y análisis de datos ha demostrado la crucial importancia de la limpieza de datos para garantizar su calidad y precisión. Utilizando Jupyter Notebook, se facilitó un análisis interactivo y eficiente, mientras que la integración de múltiples bases de datos (MySQL Server, SQL Server, MongoDB) permitió aprovechar las ventajas específicas de cada una. La visualización a través de Power BI reveló patrones y tendencias significativas, mejorando la comunicación de hallazgos. Además, el proceso fomentó el aprendizaje continuo en habilidades técnicas y destacó la relevancia del análisis exploratorio para abrir nuevas preguntas e ideas en futuras investigaciones. En conjunto, estas etapas resaltan la necesidad de un enfoque integral en el manejo de datos para una toma de decisiones informada.

- **Recomendaciones:**

Antes de crear una visualización de datos, es fundamental analizar y comprender la información contenida en los data sets. Este proceso implica identificar patrones, tendencias y posibles inconsistencias que puedan afectar la interpretación de los resultados. Al realizar un examen detallado de los datos,

se pueden detectar valores atípicos, datos faltantes o relaciones clave entre variables, lo que permite una representación gráfica más precisa y significativa. De esta manera, la visualización no solo será atractiva, sino que también facilitará una mejor toma de decisiones basada en información confiable y bien estructurada.

12. Desafíos y Problemas Encontrados

- Surgieron problemas con datos vacíos y errores de formato en especial en los .json, lo que requirió llenar esos datos nulos, para el formato se utilizó código en Python para solucionarlo.
- Al momento de general alguna visualización en PowerBi el programa no detectaba métricas en datos, se solucionó haciendo un recuento de cada una en una nueva columna.
- Al exportar el data sets de Covid no aceptaba la conexión para hacerlo, esto se debió a que el tipo de dato de la columna deaths/100 en algunos datos había incompatibilidad por lo que tuve que cambiar esos datos.

Referencias

[1] “Kaggle: your home for data science.” <https://www.kaggle.com/>

[2] “Joker - movie reviews,” *Rotten Tomatoes*.
https://www.rottentomatoes.com/m/joker_2019/reviews

Anexos

Repositorio de Github:

https://github.com/Anthon-Chang/analisis_datos.git

Enlaces de los videos:

- Recopilación de datos: <https://youtu.be/trl150zOnhs>
- Conclusiones de los dashboards: https://youtu.be/D_miZNeMvzs

Enlaces a los dashboards en PowerBI:

Liga Española:

<https://app.powerbi.com/view?r=eyJrIjoiMDczZTJhMjUtMmJjMi00MDlkLWJkYjYtZThjYzdmNmVINDMyIiwidCI6IjY4MmE0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMjY2ZDE4YWEzNyIsImMiOiR9>

Netflix:

<https://app.powerbi.com/view?r=eyJrIjoizTA5YjJlZDctNmFlZC00MDk2LTkxNzltZTk4YzgzMTM2JmNGIwIiwidCI6IjY4MmE0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMjY2ZDE4YWEzNyIsImMiOjR9>

Reseñas comida peruana:

<https://app.powerbi.com/view?r=eyJrIjojMGJwZTE5YjU0MWE3NC00ZjRlLWI2ZmMtNjhlYzljYTg2NDI2IiwidCI6IjY4MmE0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMjY2ZDE4YWEzNyIsImMiOjR9>

Videojuegos, Hobbies y Noticias:

<https://app.powerbi.com/view?r=eyJrIjojYzNiZjM2ODUtZjBmYi00OWQ5LTkzYWUtYmU3Yzk2YzJiMTUxIiwidCI6IjY4MmE0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMjY2ZDE4YWEzNyIsImMiOjR9>

Covid, NBA, Starbucks:

<https://app.powerbi.com/view?r=eyJrIjojZDM4MWQ2MDAtNmQ5Mi00MDNhLTlkMjYtMjgzM2I4MGMxNzdlIiwidCI6IjY4MmE0ZTZhLWE3N2YtNDk1OC1hM2FjLTllMjY2ZDE4YWEzNyIsImMiOjR9>