



# PROYECTO FINAL

Chang Anthon  
Morales Eddy  
Cisneros Paulo

# OBJETIVO

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

## 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.

# FUENTES DE DATOS

Los lugares de donde obtuvimos los datos fueron:

The logo for Kaggle, featuring the word "kaggle" in a lowercase, rounded, sans-serif font. The letters are a light blue color.The logo for Rotten Tomatoes, featuring the words "Rotten" and "Tomatoes" stacked vertically in a bold, sans-serif font. The letter "o" in "Rotten" is replaced by a tomato icon, and the letter "o" in "Tomatoes" is replaced by a tomato icon. The logo is white.

# HERRAMIENTAS UTILIZADAS

Herramientas para código:  python™  jupyter

Bases de datos:    MongoDB®  CouchDB  
relax

Herramientas para visualizaciones:



# LIMPIEZA DE DATOS

La limpieza de los archivos se dio por medio de Jupyter notebook

Hobbies:

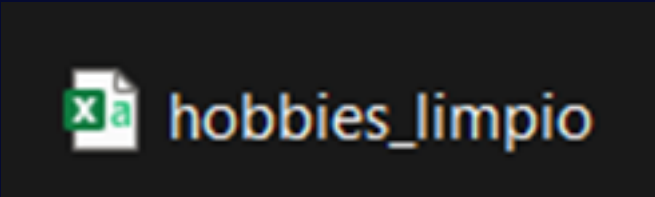
Limpieza de datos archivo Hobbies

```
#Importacion de librerias
import pandas as pd
import numpy as np
import json
```

[1] Python

```
#Cargar el archivo
df_csv = pd.read_csv("C:/Users/Personal/OneDrive/Documentos/Archivos de programación/Jupyter/datos/Student Attitude and Behavior.csv")
```

[6] Python



#Obtener informacion de los datos  
df\_csv.info()

Python

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 235 entries, 0 to 234
Data columns (total 19 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Certification Course                 235 non-null   object
1   Gender                               235 non-null   object
2   Department                           235 non-null   object
3   Height(CM)                           235 non-null   float64
4   Weight(KG)                           235 non-null   float64
5   10th Mark                            235 non-null   float64
6   12th Mark                            235 non-null   float64
7   college mark                         235 non-null   float64
8   hobbies                             235 non-null   object
9   daily studing time                   235 non-null   object
10  prefer to study in                   235 non-null   object
11  salary expectation                    235 non-null   int64
12  Do you like your degree?              235 non-null   object
13  willingness to pursue a career based on their degree  235 non-null   object
14  social medai & video                  235 non-null   object
15  Travelling Time                       235 non-null   object
16  Stress level                          235 non-null   object
17  Financial Status                      235 non-null   object
18  part-time job                        235 non-null   object
dtypes: float64(5), int64(1), object(13)
memory usage: 35.0+ KB
```

Videojuegos:

Limpieza de datos archivo Videojuegos

```
#cargar el archivo
df2_csv = pd.read_csv("C:/Users/Personal/OneDrive/Documentos/Archivos de programación/Jupyter/datos/metacritic_games.csv")
```

[1] ✓ 0.0s Python

```
#Revisar primeras filas
df2_csv.head()
```

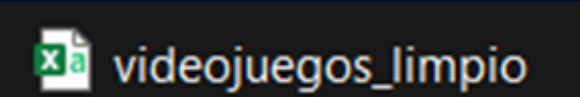
[4] ✓ 0.0s Python

	game	platform	developer	genre	number_players	rating	release_date	positive_critics	neutral_critics	negative_critics	positive_users	neutral_users
0	Portal 2	PC	Valve Software	Action	NaN	E10+	Apr 18, 2011	51	1	0	1700	107
1	The Elder Scrolls V: Skyrim	PC	Bethesda Game Studios	Role-Playing	No Online Multiplayer	M	Nov 10, 2011	32	0	0	1616	322
2	The Legend of Zelda: Ocarina of Time 3D	3DS	GREZZO	Miscellaneous	No Online Multiplayer	E10+	Jun 19, 2011	84	1	0	283	20
3	Batman: Arkham City	PC	Rocksteady Studios	Action Adventure	NaN	T	Nov 21, 2011	27	0	0	240	34
...	Super Mario	...	...	...	No Online	...	Nov 11	...	...	...	...	...

#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                     5689 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
```



# LIMPIEZA DE DATOS

## Web-Scrapping

```
Reseña 1: ...over two films, Joker looks rather different from the edgy-origin story this was taken to be o
he messaging but that corrosive message about the dangeous direction society was heading was here all alo
---
Reseña 2: It had me at the beginning and the last half hour or more is superb, but Phillips earnest, deli
erse and heavy-handed stab at making it germane to socioeconomic issues hurt my suspension of disbelief.
---
Reseña 3: Joker, directed by Todd Phillips, is the worst kind of bad film. It's a technically proficient
ects responsibility as it strives to understand and empathize with the stereotypical white mass murderer.
---
Reseña 4: The environment and its atmosphere are a central aspect narratively, that is why Joker begins w
ork from Taxi Driver, is represented as a putrid place with social violence and rising class resentment.
---
Reseña 5: Joker is what so many controversial films turn out to be: it's fine – neither masterpiece nor t
y thought out in others.
---
Reseña 6: What the film wants to say – about mental illness or class divisions in society – is not as int
eness.
---
Reseña 7: a clever gem of a movie that approaches its complex themes with extreme simplicity.
---
Reseña 8: The joke in both cases is that the director both gets to disavow white male murderousness and c
---
Reseña 9: It's not a film about the Joker. It's a very realistic portrayal of someone (anyone!) who can b
illiant!
---
Reseña 10: "What do you get when you cross a mentally ill loner with a society that abandons him and treat
the titular character of Joker, a film that luxuriates in giving us the answer
```

```
import requests

from bs4 import BeautifulSoup

import pandas as pd
```

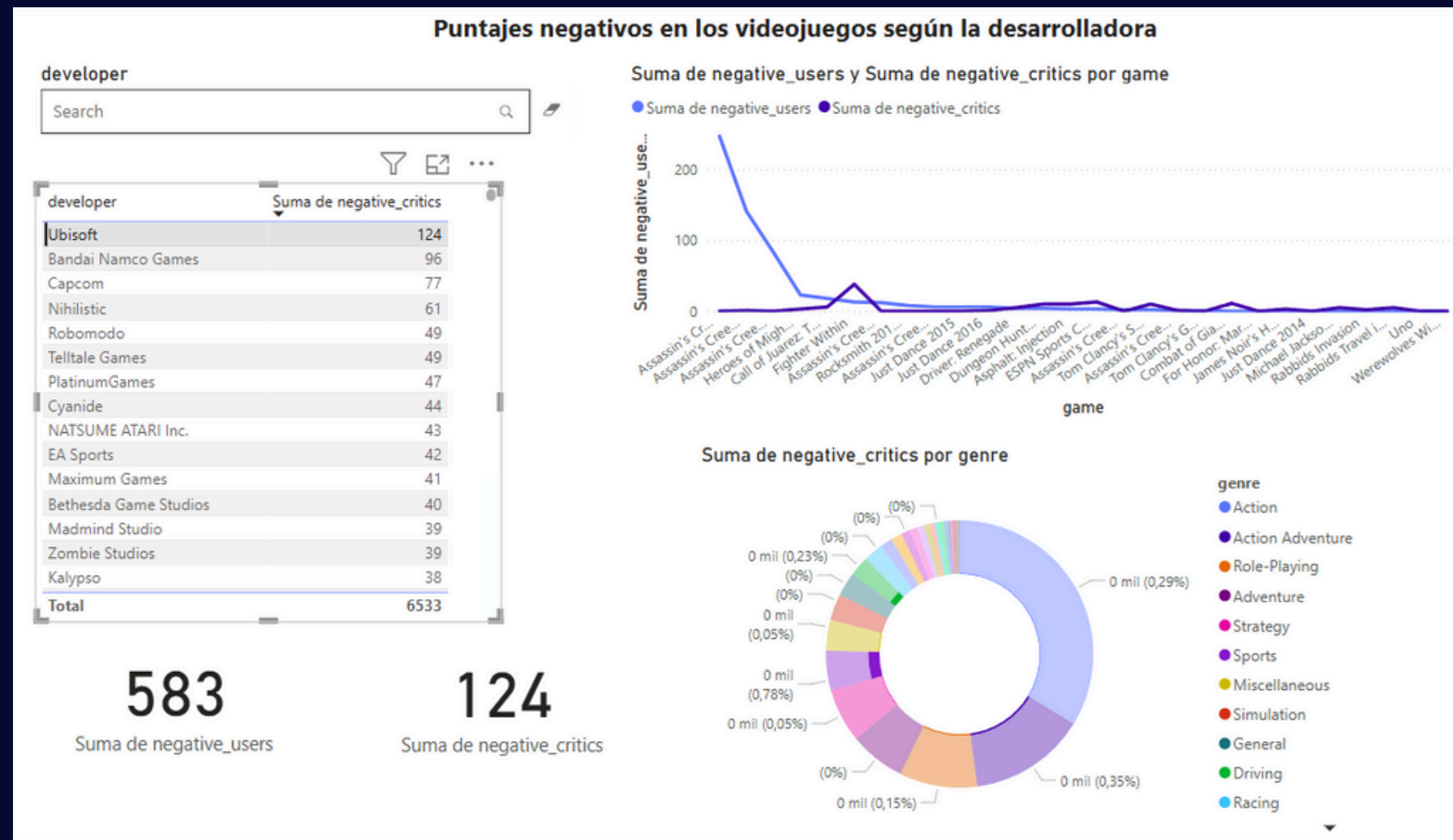
```
# Encabezados para simular una solicitud desde un navegador
headers = {
    'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.124 Safari/537.36'
}

# Realizar la solicitud HTTP con encabezados
response = requests.get(url, headers=headers)
```

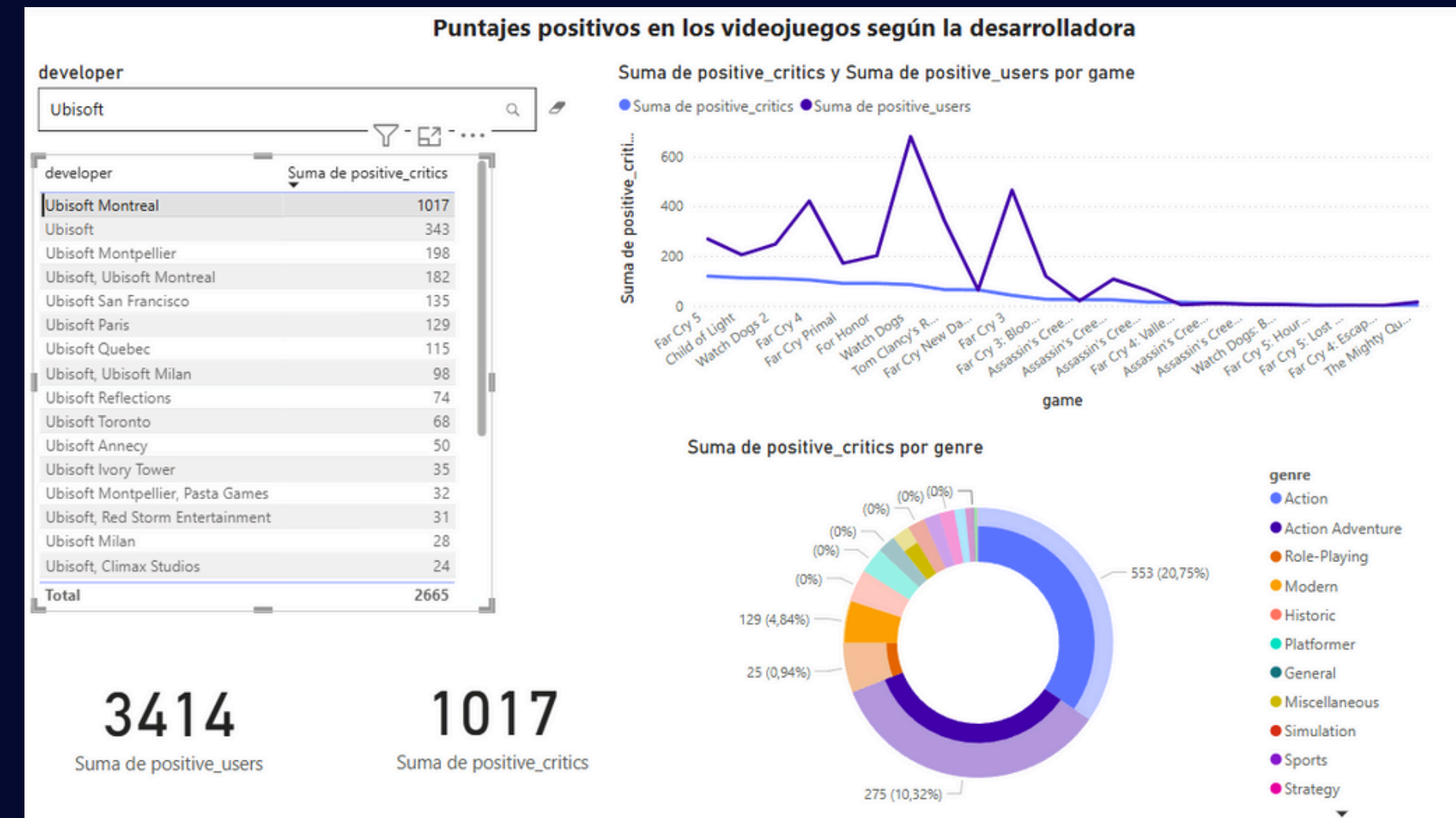


# CASOS PRUEBA

¿CUÁL ES LA DESARROLLADORA QUE MÁS CRÍTICAS NEGATIVAS TUVO Y CUÁL ES SU JUEGO MÁS CRITICADO?



¿CUÁL ES LA DESARROLLADORA QUE MÁS CRÍTICAS POSITIVAS TUVO MÁS GUSTADO?

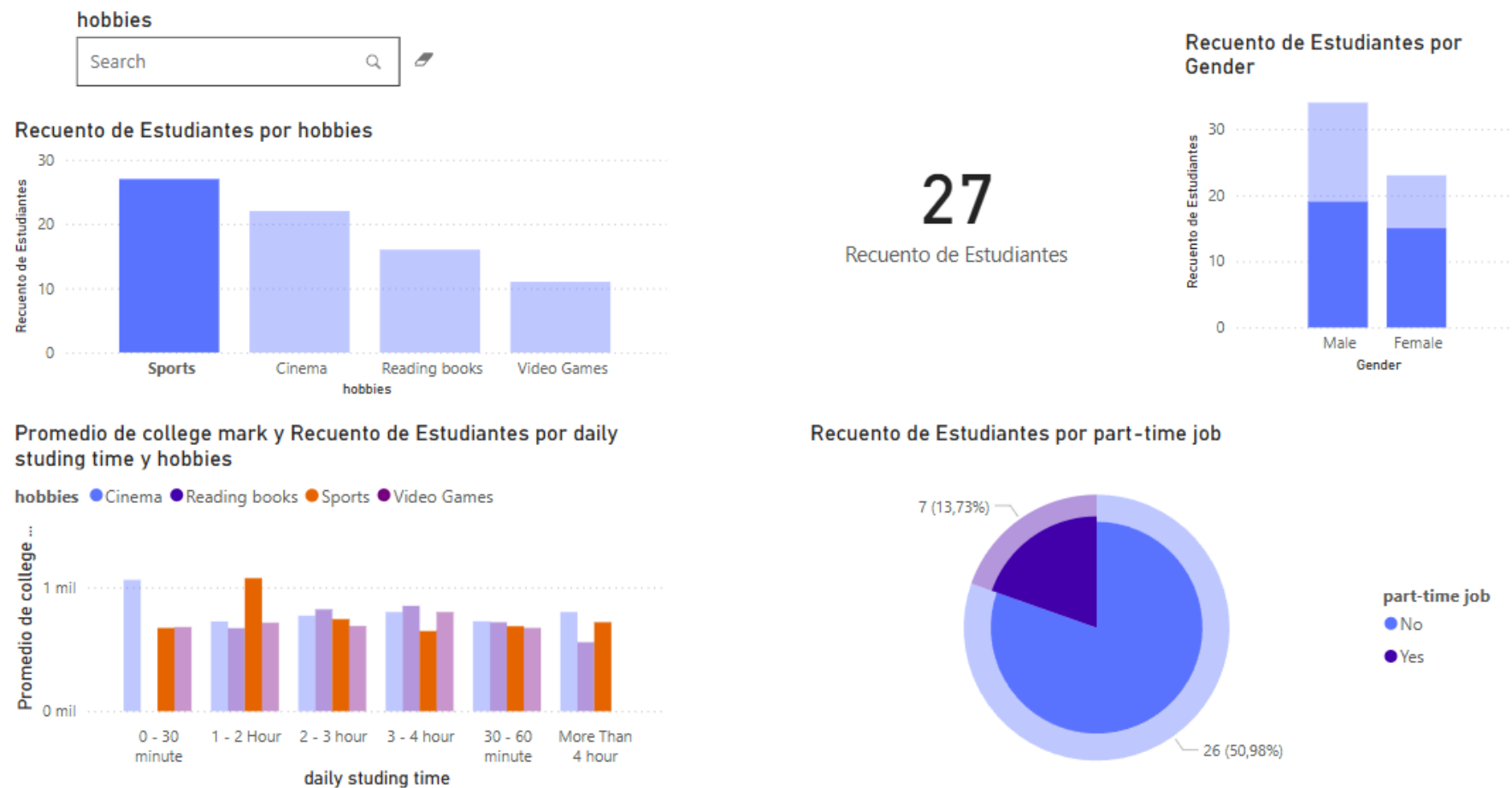


# CASOS PRUEBA

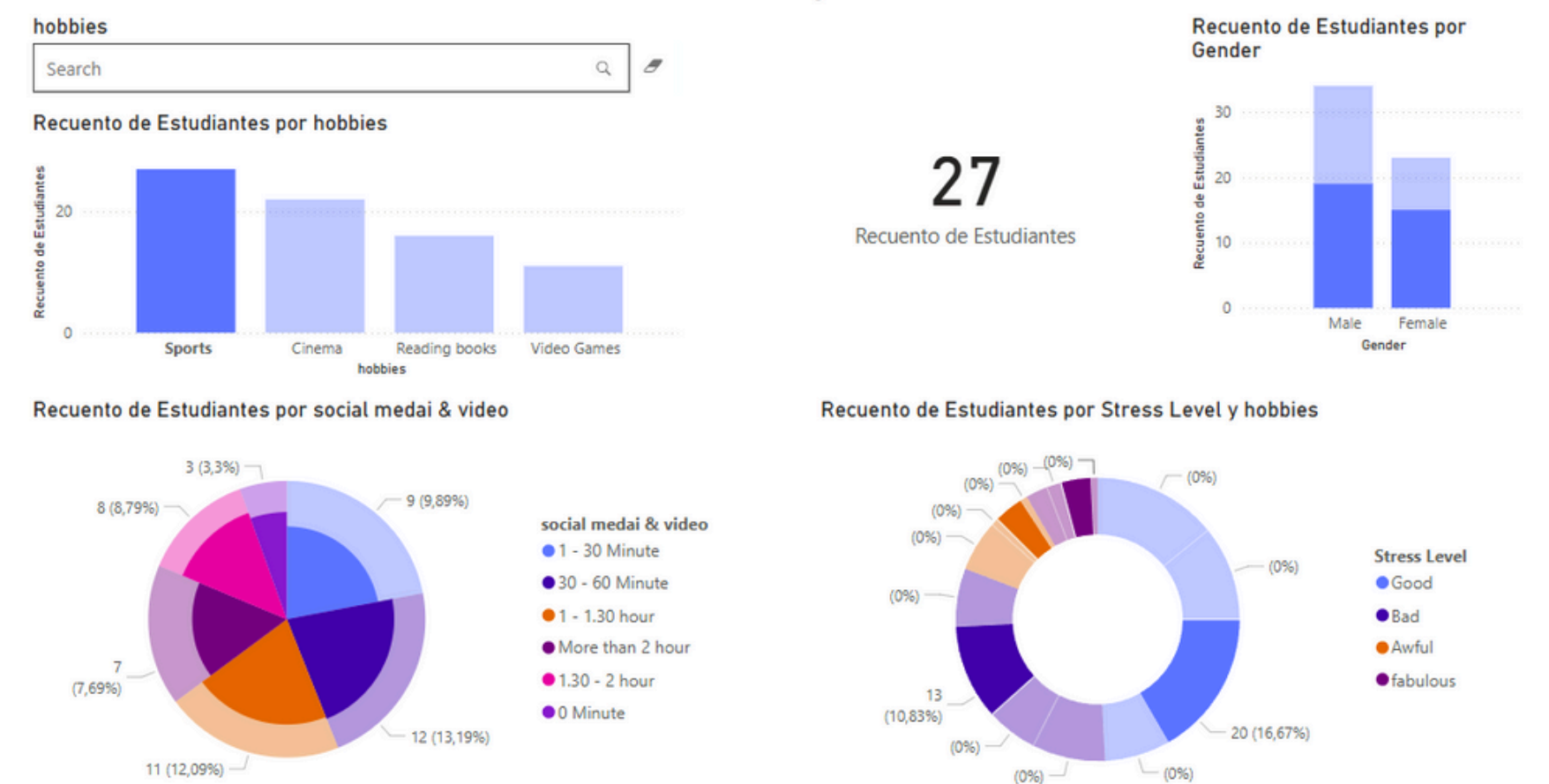
¿CUÁL ES UN HOBBY QUE SE PUEDE PRACTICAR Y QUE PERMITE TENER TIEMPO PARA ESTUDIAR Y TRABAJAR A MEDIO TIEMPO?

¿CÓMO INFLUYEN LOS HOBBIES QUE PRACTICAMOS EN LA REDUCCIÓN DEL ESTRÉS Y CÓMO PUEDEN SERVIR COMO INDICADORES DE BIENESTAR?

Influencia de los hobbies en el tiempo de estudio y posibilidad de tener un trabajo a medio tiempo



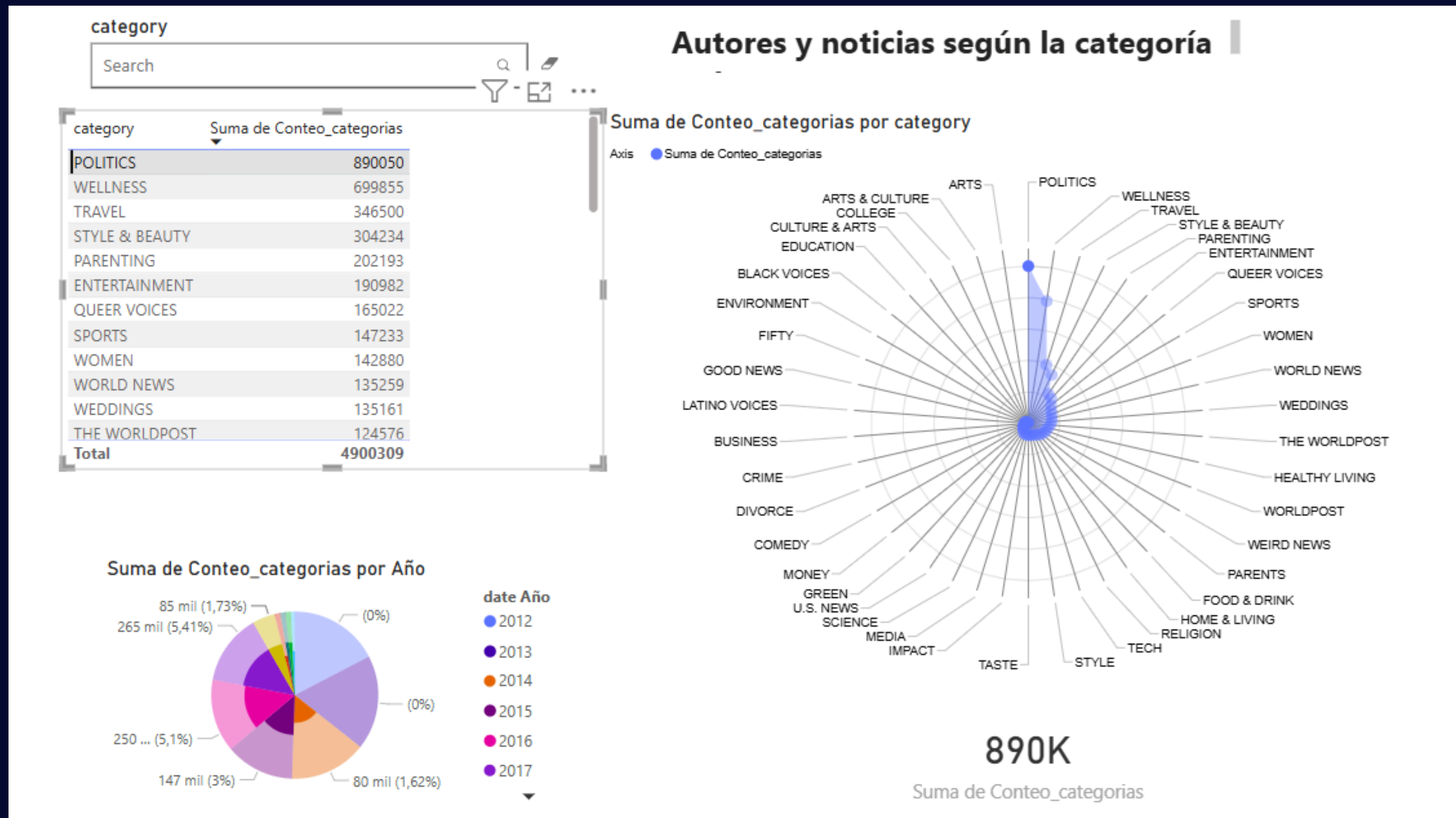
Influencia de los hobbies en el tiempo de redes sociales e indicador de estrés





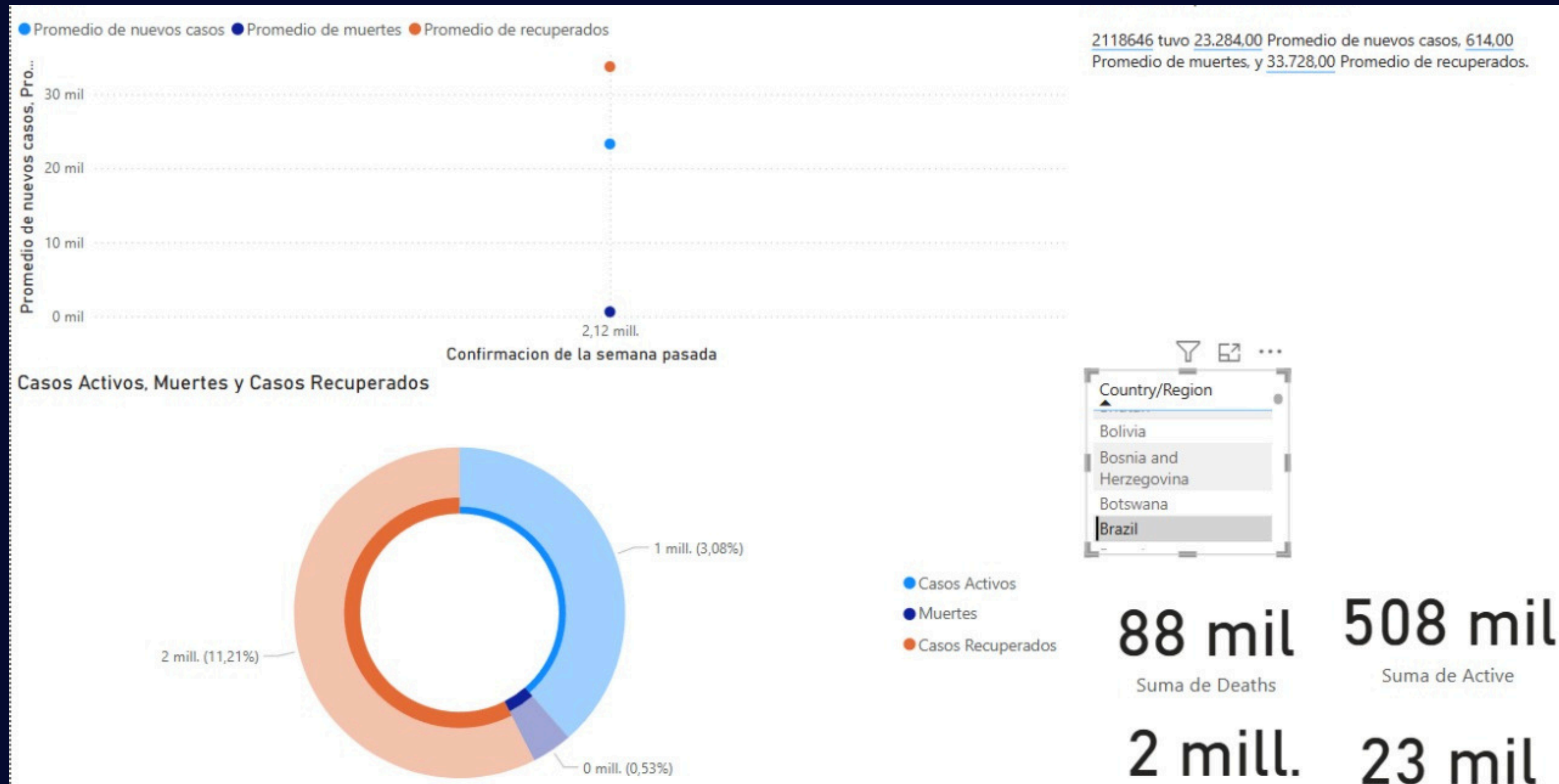
# CASOS PRUEBA

¿CUÁLES SON LAS TENDENCIAS MUNDIALES EN LAS NOTICIAS?



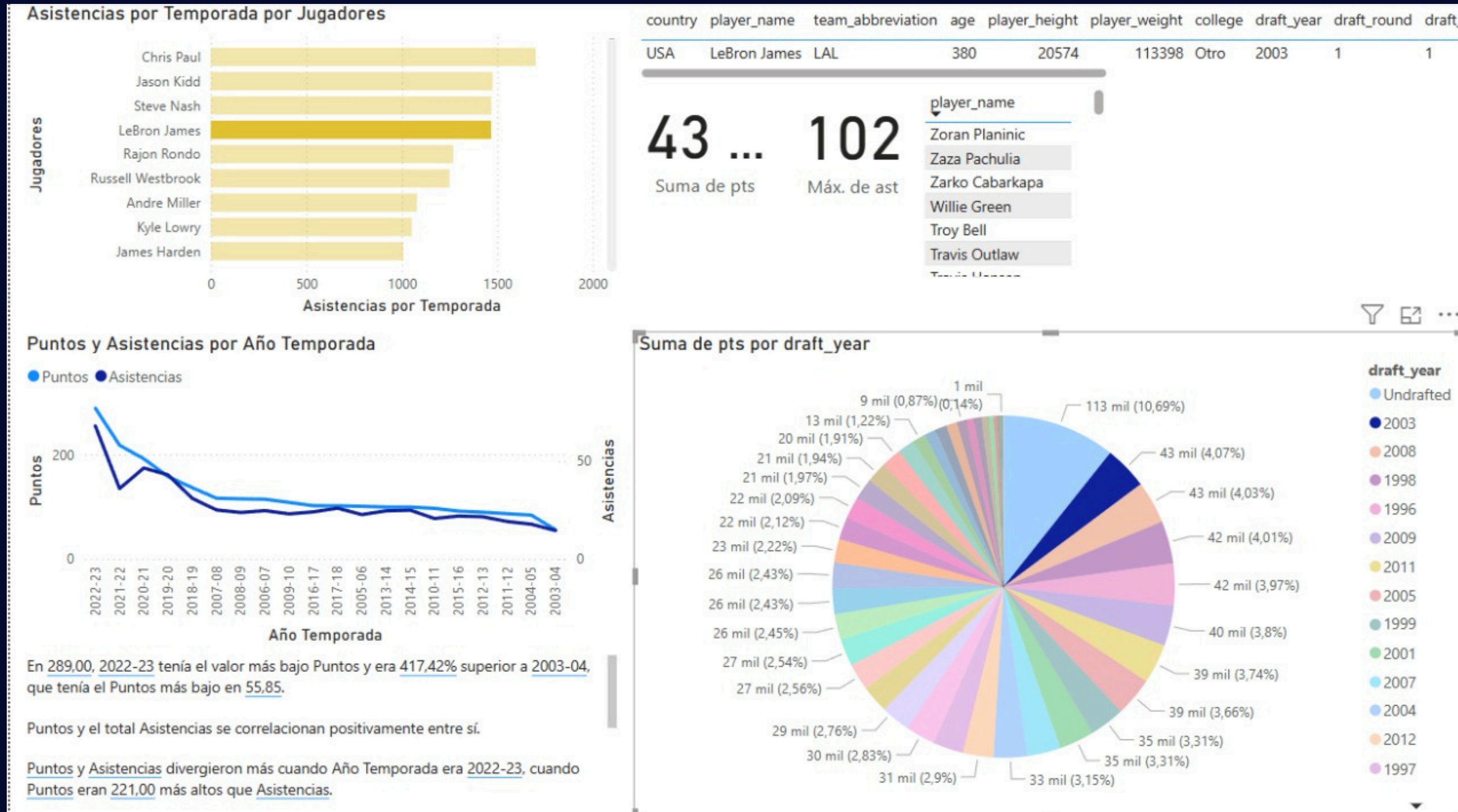
# CASOS PRUEBA

¿CUÁL ES LA CANTIDAD DE CASOS CONFIRMADOS, MUERTES Y NUEVOS CASOS DE COVID-19 EN BRASIL?



# CASOS PRUEBA

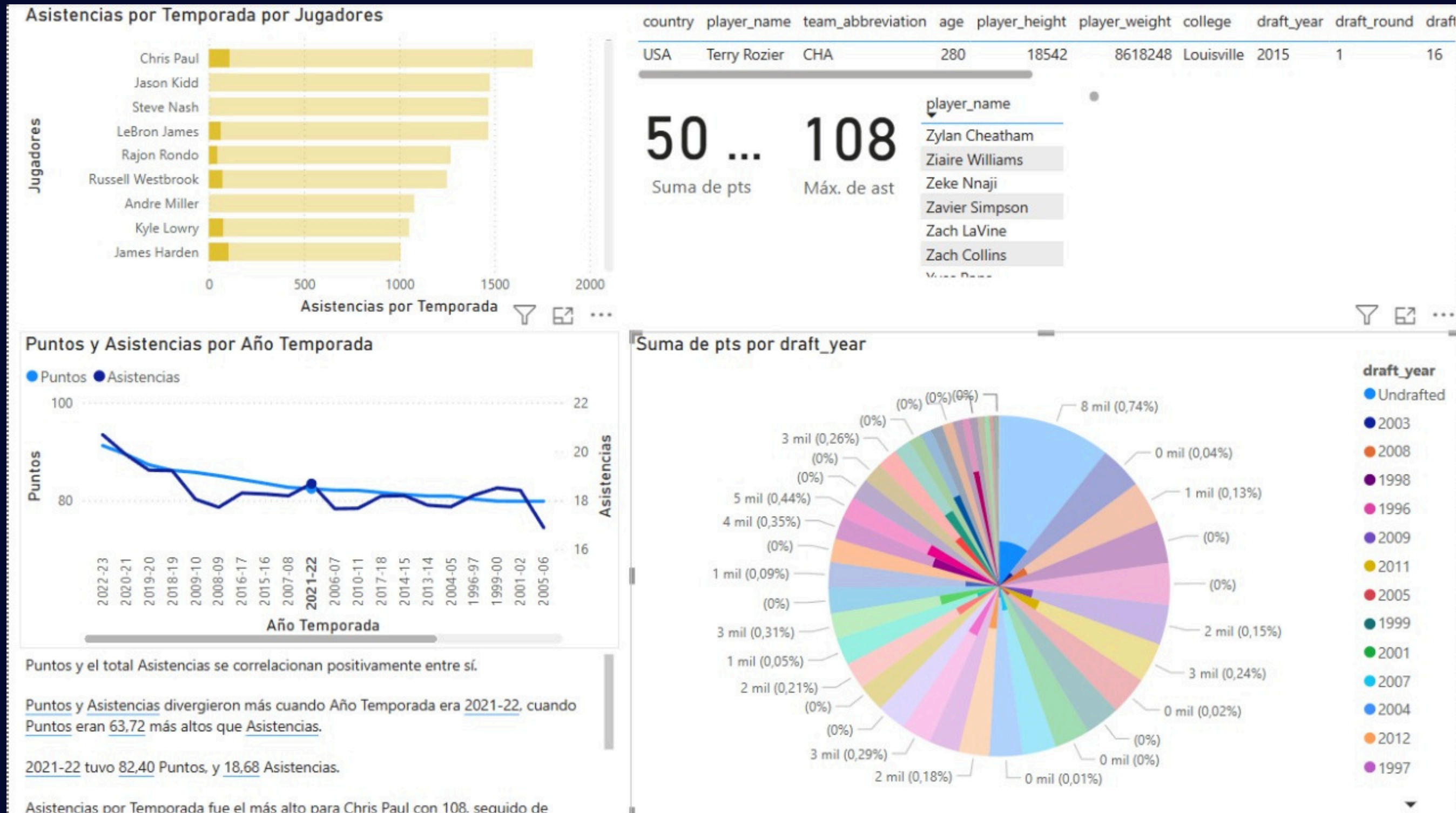
¿CUÁL HA SIDO EL MEJOR AÑO DE LEBRÓN JAMES EN TÉRMINOS DE PUNTOS Y ASISTENCIAS?





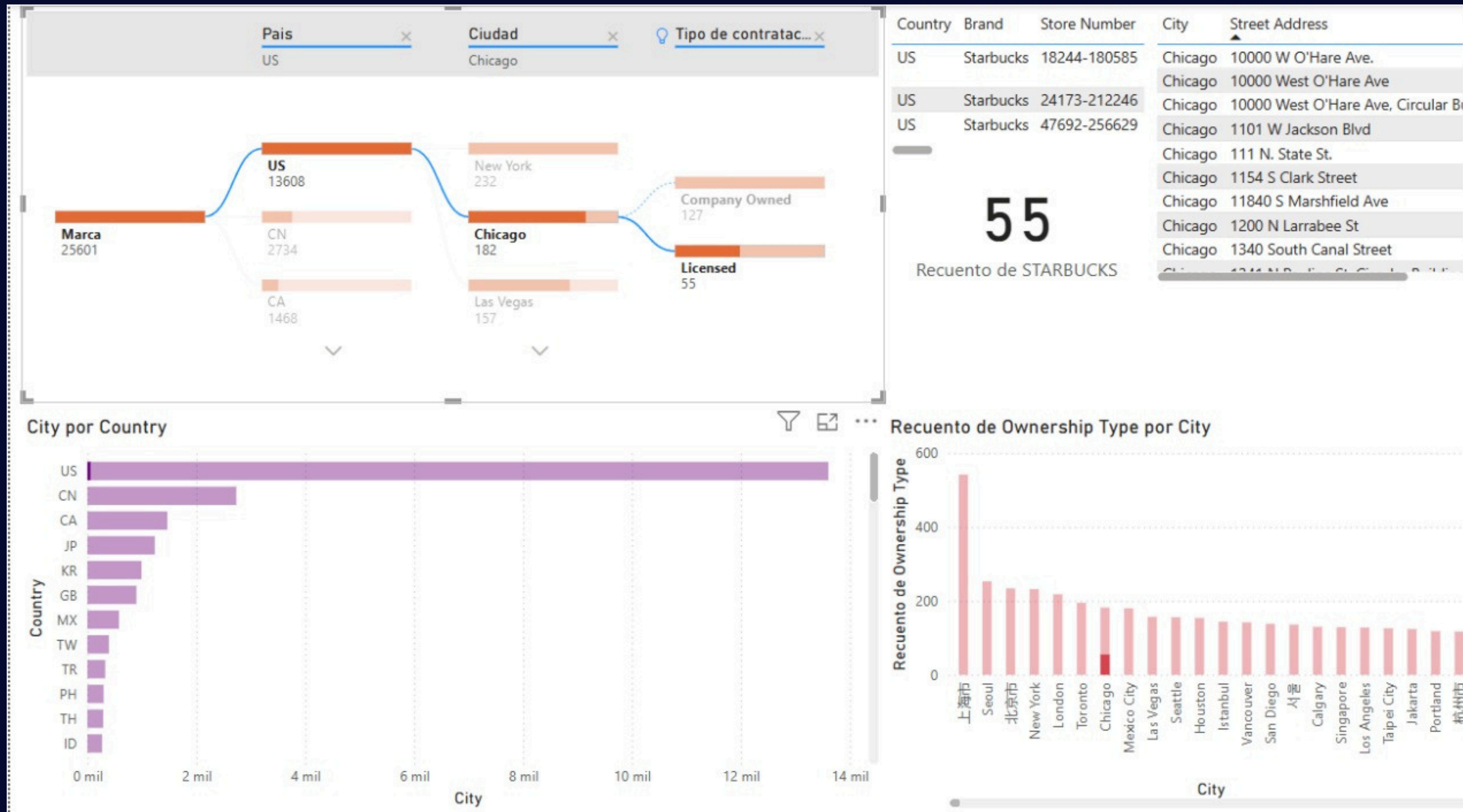
# CASOS PRUEBA

¿CUÁL FUE LA SUMA TOTAL DE PUNTOS DE LOS JUGADORES EN LA TEMPORADA 2021?

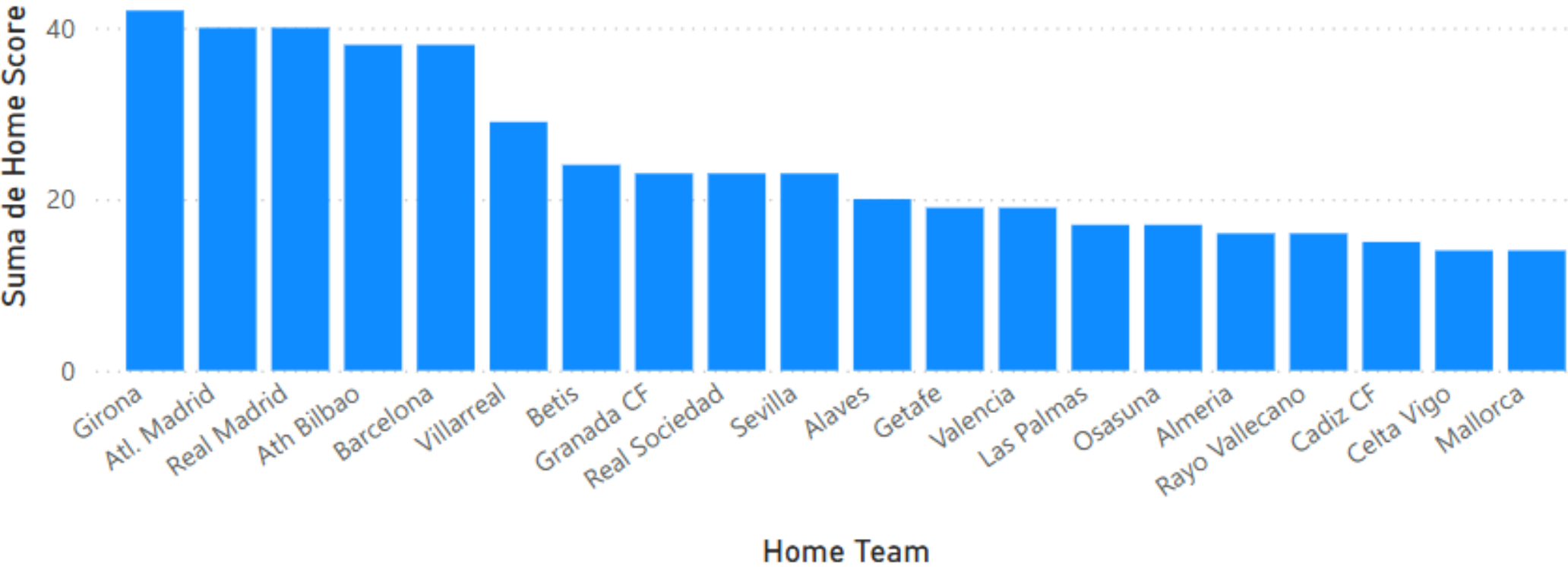


# CASOS PRUEBA

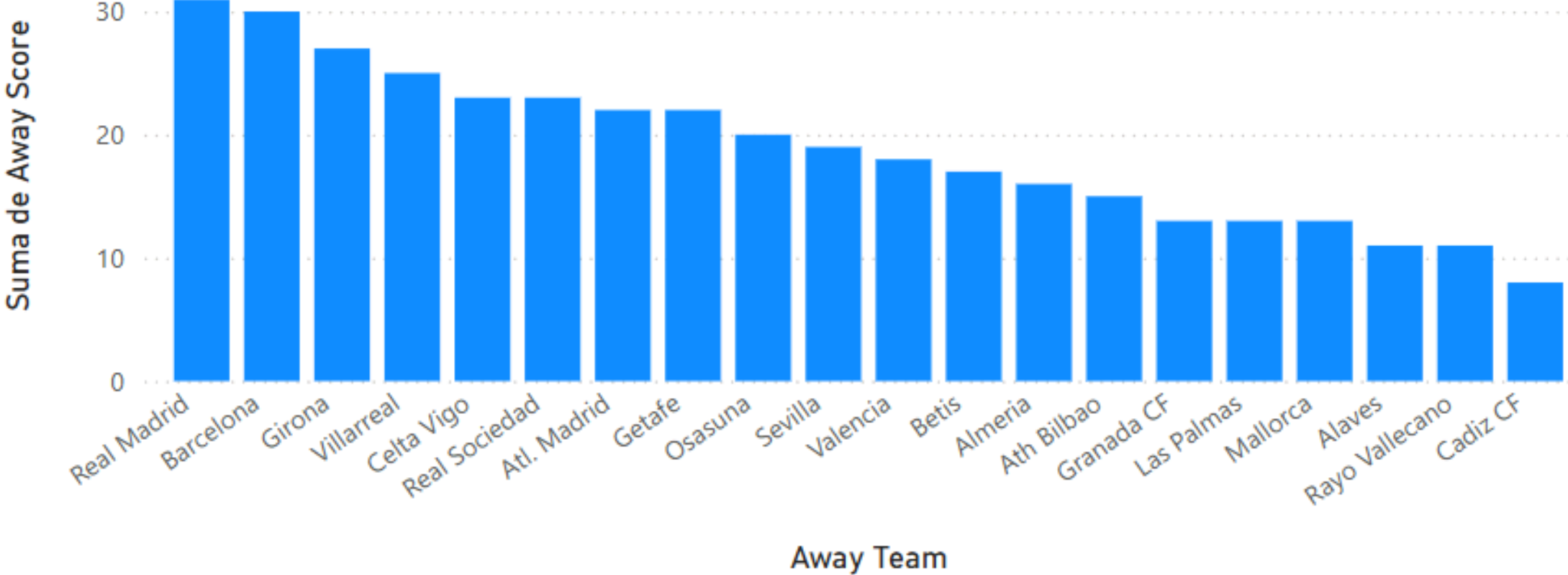
¿CUÁL ES EL NÚMERO TOTAL DE LOCALES DE STARBUCKS CON LICENCIA EN CHICAGO?



Total de puntos realizados en Casa



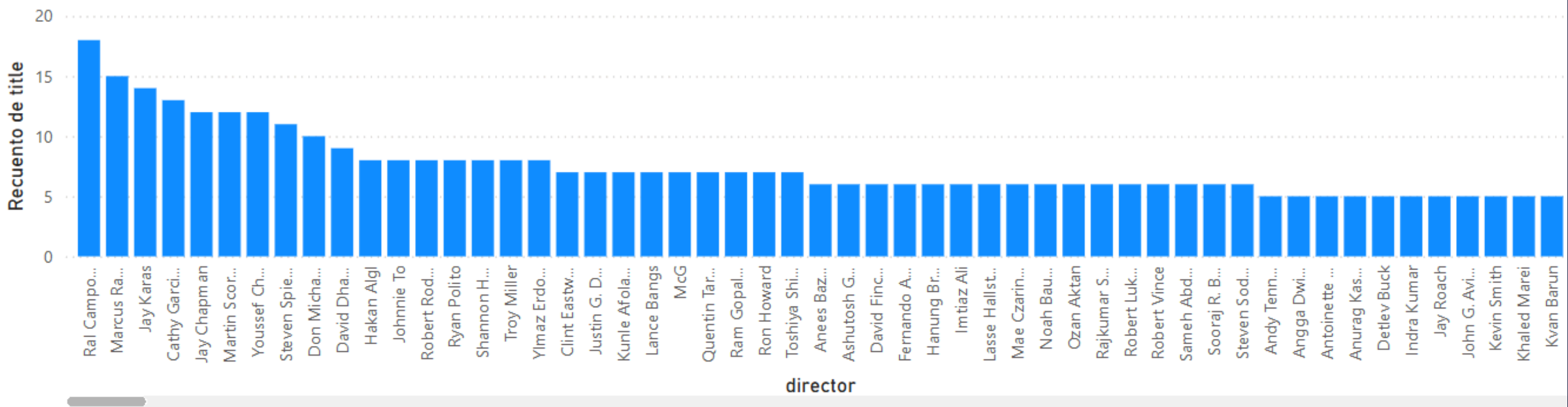
Total de puntos realizados como Visitante



¿CUÁL ES EL TOTAL DE PUNTOS REALIZADOS POR EQUIPO TANTO EN CASA COMO DE VISITANTE?



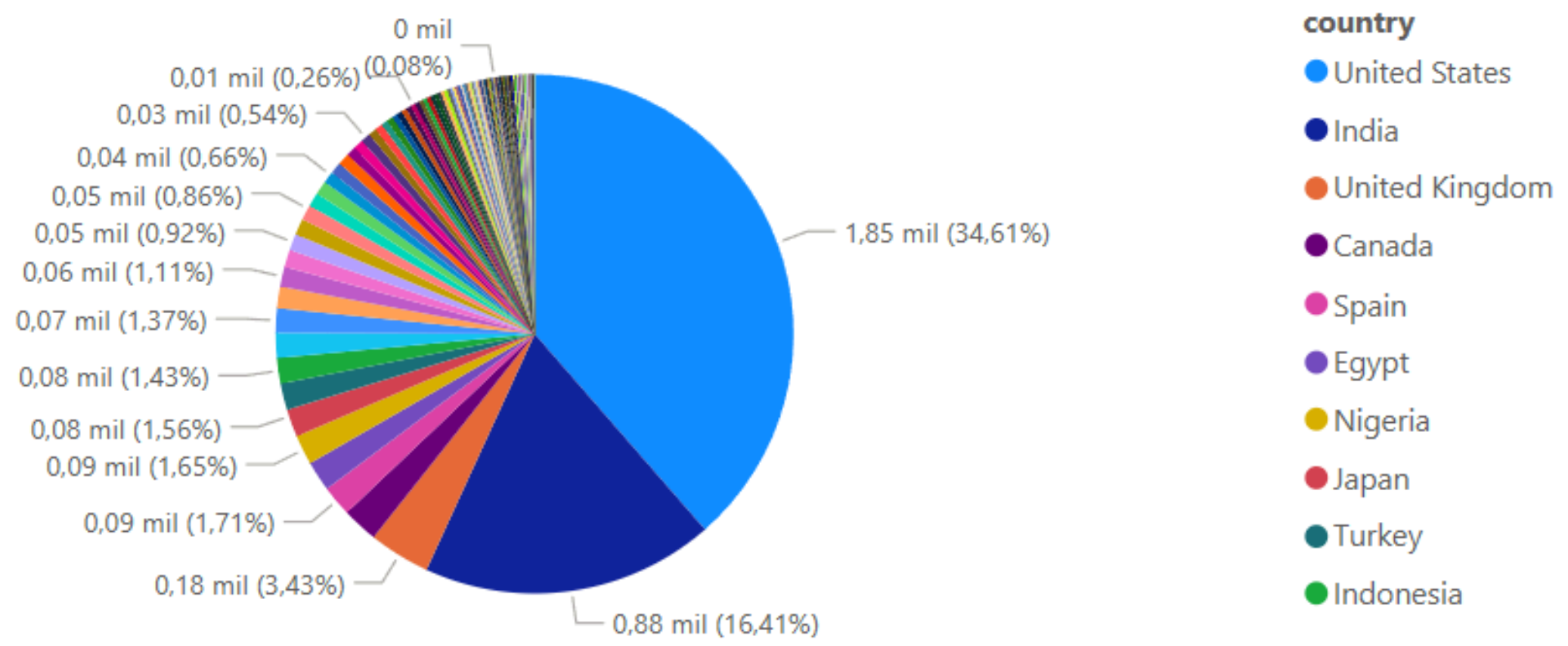
Cantidad de películas por director



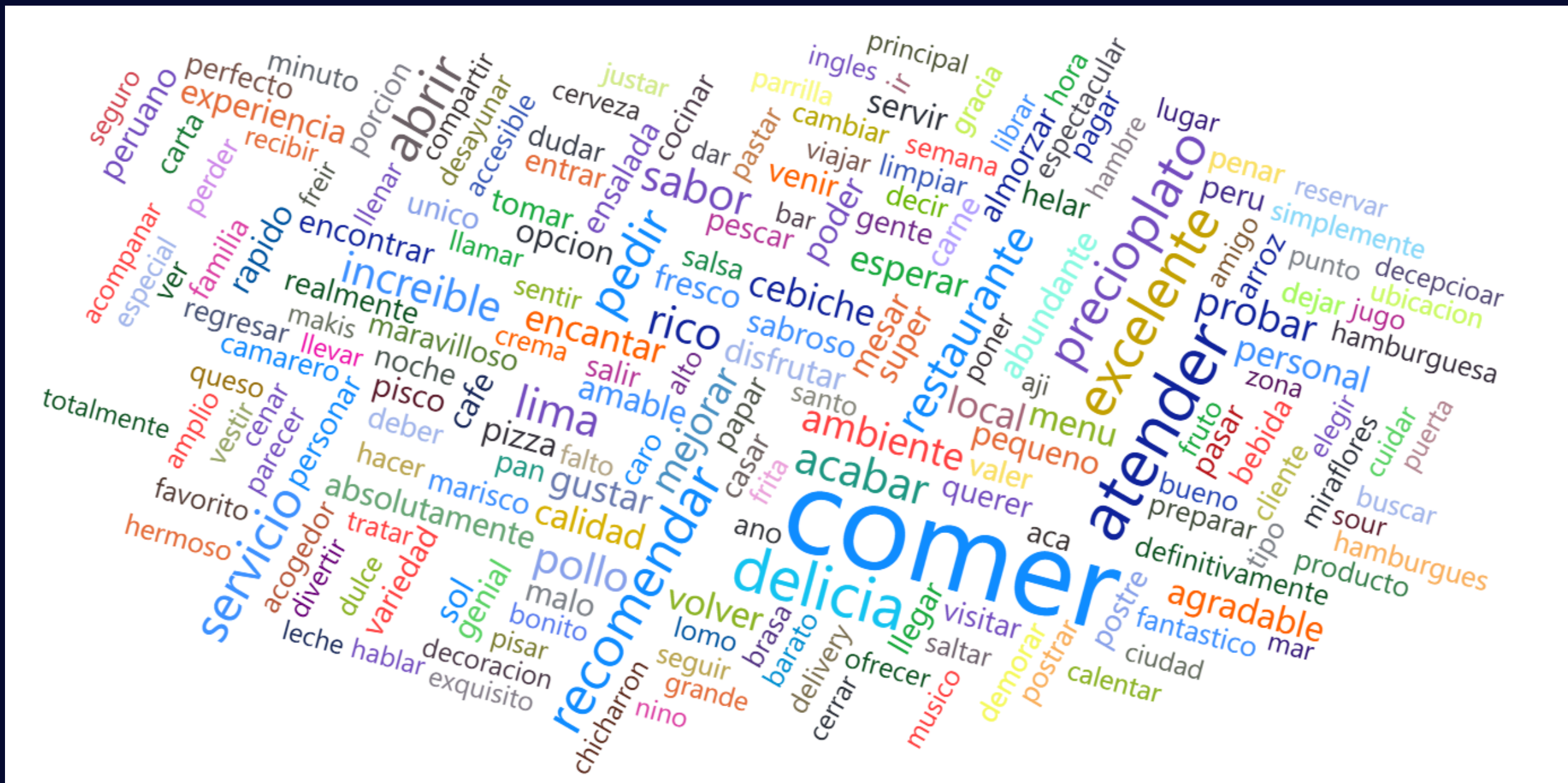
¿CUÁL ES LA CANTIDAD DE PELICULAS HECHAS POR DIRECTOR?

¿CUÁL ES LA CANTIDAD DE PELICULAS HECHAS POR PAÍS?

Cantidad de películas por País



¿CUÁL SON LAS PALABRAS MAS FRECUENTES EN LA RESEÑA?





# CONCLUSIÓN

El proyecto de recopilación y análisis de datos ha demostrado la crucial importancia de la limpieza de datos para garantizar su calidad y precisión.

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.



# RECOMENDACION

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.



GRACIAS