

PROJECT: INVESTIGATING NETFLIX MOVIES



```
##Azaneth Gonzalez##  
#Investigating Netflix Movies  
#26 agosto de 2025
```

Netflix! What started in 1997 as a DVD rental service has since exploded into one of the largest entertainment and media companies.

Given the large number of movies and series available on the platform, it is a perfect opportunity to flex your exploratory data analysis skills and dive into the entertainment industry.

You work for a production company that specializes in nostalgic styles. You want to do some research on movies released in the 1990's. You'll delve into Netflix data and perform exploratory data analysis to better understand this awesome movie decade!

You have been supplied with the dataset `netflix_data.csv`, along with the following table detailing the column names and descriptions. Feel free to experiment further after submitting!

The data

netflix_data.csv

Column	Description
<code>show_id</code>	The ID of the show
<code>type</code>	Type of show
<code>title</code>	Title of the show
<code>director</code>	Director of the show
<code>cast</code>	Cast of the show
<code>country</code>	Country of origin
<code>date_added</code>	Date added to Netflix
<code>release_year</code>	Year of Netflix release
<code>duration</code>	Duration of the show in minutes
<code>description</code>	Description of the show
<code>genre</code>	Show genre

```
# Importing pandas and matplotlib
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
# Read in the Netflix CSV as a DataFrame
netflix_df = pd.read_csv("netflix_data.csv")
```

#Primero recomiendo que se imprima el dataset para ver con qué tipos de datos se va a trabajar

```
print(netflix_df)
```

	show_id	...	genre
0	s2	...	Dramas
1	s3	...	Horror Movies
2	s4	...	Action
3	s5	...	Dramas
4	s6	...	International TV
...
4807	s7779	...	Comedies
4808	s7781	...	Dramas
4809	s7782	...	Children
4810	s7783	...	Dramas
4811	s7784	...	Dramas

```
[4812 rows x 11 columns]
```

#1 Filtramos los datos: Solo interesan las películas que se estrenaron en 1990s para posteriormente encontrar la duración de las películas más frecuentes.

```
movies_90s = netflix_df[(netflix_df['release_year'] >= 1990) &
                        (netflix_df['release_year'] <= 1999) & (netflix_df['type'] == 'Movie')]

duration = movies_90s['duration'].mode()[0] #Aquí solo seleccionamos la duración

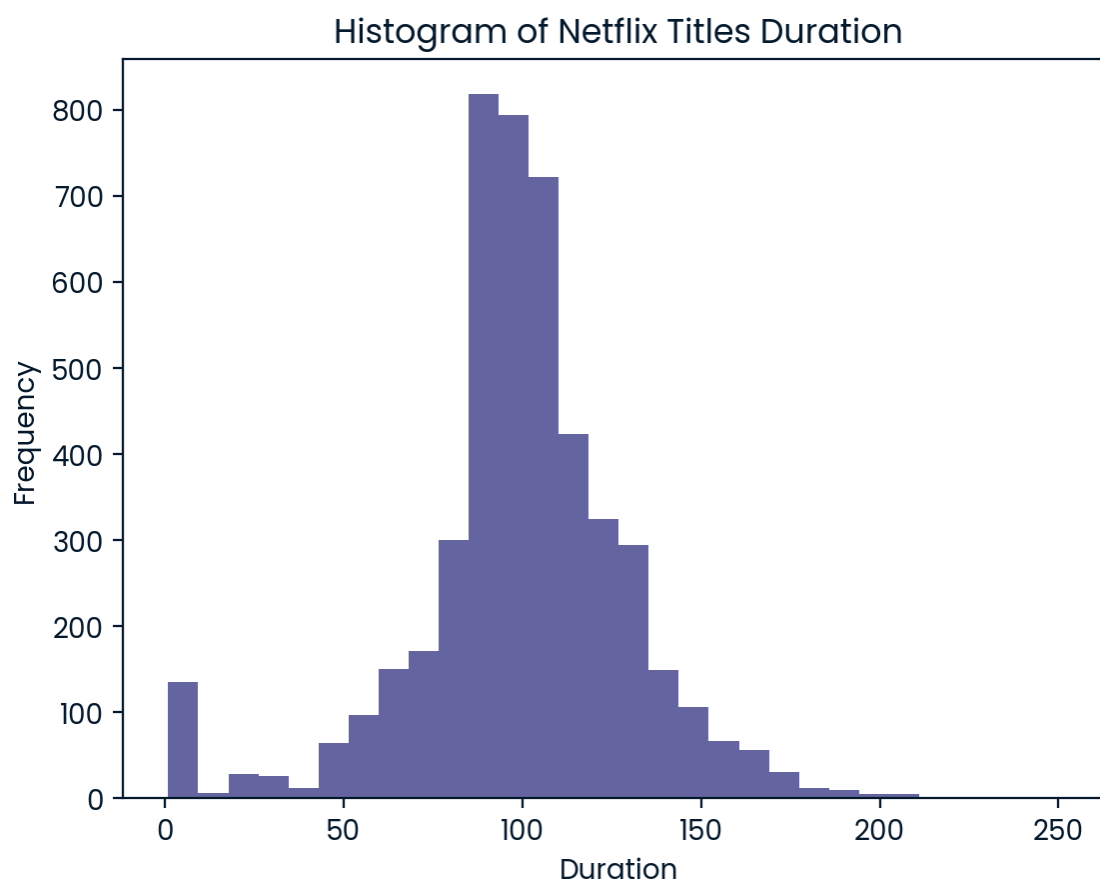
print(f"La duración más frecuente de las películas en 1990 fue de {duration} minutos.")
```

La duración más frecuente de las películas en 1990 fue de 94 minutos.

#Con la información anterior, podemos crear un historgrama para ver cuál es la duración de las películas más frecuente en el periodo de tiempo dado de una forma más visual

#Como duration es una columna de cadenas, primero lo paramos a un array
duration = np.array(netflix_df["duration"])

```
# Graficamos
plt.hist(duration, bins=30)
plt.xlabel('Duration')
plt.ylabel('Frequency')
plt.title('Histogram of Netflix Titles Duration')
plt.show()
```



#Ahora nos piden contar el número de películas de accion cortas (<90 min) de la década de 1990. Hacemos algo similar al inciso anterior.

```
movies2 = netflix_df[(netflix_df['release_year'] >= 1990) &
(netflix_df['release_year'] <= 1999) & (netflix_df['type'] == 'Movie') &
(netflix_df['genre'] == 'Action') & (netflix_df['duration'] < 90)]

print("El número de películas de acción cortas de la década de 1990 es de",
movies2['show_id'].count())
```

El número de películas de acción cortas de la década de 1990 es de 7

```
#Graficamos
```

```
movie_durations = movies2['duration']
```

```
print("En promedio, las películas de acción de la década de los 90s  
duraban:",movie_durations.mean())  
plt.hist(movie_durations)  
plt.title('Duración de películas de acción cortas de los 90')  
plt.xlabel('Duración (minutos)')  
plt.ylabel('Número de películas')  
plt.show()
```

En promedio, las películas de acción de la década de los 90s duraban:
84.14285714285714

