

Proyecto de Analisis TalentoTECH

RETO OBJETIVO

En este reto, trabajarán con un dataset que contiene información sobre las mejores películas del 2019.

El objetivo es crear un dashboard interactivo en Looker Studio que permita responder preguntas clave relacionadas con el rendimiento, popularidad y características principales de estas películas.

El proceso implicará identificar las columnas necesarias para el análisis, justificar las decisiones tomadas y diseñar visualizaciones claras y funcionales. Finalmente, interpretarán los resultados para responder preguntas clave.

Preguntas Claves

- ¿Cuáles son las películas con mejores calificaciones en IMDb?
- ¿Qué género fue el más representativo en el Top de películas del 2019?
- ¿Existe alguna relación entre la duración y la calificación de IMDb?
- ¿Qué director recibió más votos en sus películas?

Herramientas a usar

- Vscode
- Looker Studio

MANIPULACION DE DATOS

Para este Analisis descartare las columnas (Position, Const, Description y URL) ya que no contiene informacion que sea relevante para este reto en especifico

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Position              100 non-null   int64
1   Const                 100 non-null   object
2   Created               100 non-null   object
3   Modified              100 non-null   object
4   Description            100 non-null   object
5   Title                 100 non-null   object
6   Original Title        100 non-null   object
7   URL                   100 non-null   object
8   Title Type            100 non-null   object
9   IMDb Rating           100 non-null   float64
10  Runtime (mins)        100 non-null   int64
11  Year                  100 non-null   int64
12  Genres                 100 non-null   object
13  Num Votes             100 non-null   object
14  Release Date          100 non-null   object
15  Directors              46 non-null    object
dtypes: float64(1), int64(3), object(12)
memory usage: 12.6+ KB

```

Adicionalmente Normalizare la columna de "Genres" y "Title Type" capitalizando sus valores.

Se convierten columnas donde tienen datos de tipo fecha a "Datetime"

CODIGO USADO PARA MANIPULACION

```

import pandas as pd

# %%
ruta = r'c:\Users\ACER-PC\Downloads\Data Set top peliculas 20
df = pd.read_csv(ruta, encoding='ansi')
df.head()

# %%
df.info()

# %%
# Borra Columnas no utiles para el analisis
df.drop(columns=['Position', 'Const', 'Description', 'URL', 'Ti

# %%

```

```
# Capitalizamos los Generos y corrige errores de escritura
df['Genres'] = df['Genres'].str.capitalize()
df['Title Type'] = df['Title Type'].str.capitalize().str.replace(' ', '-')

# %%
df.head(1)

# %%
columna_fechas = ['Created', 'Modified', 'Release Date']

for columna in columna_fechas:
    df[columna] = pd.to_datetime(df[columna], errors='coerce')

df.select_dtypes(include='datetime').isna().sum()

# %%
df.to_csv(r'c:\Users\ACER-PC\Downloads\Data Set top peliculas')
```

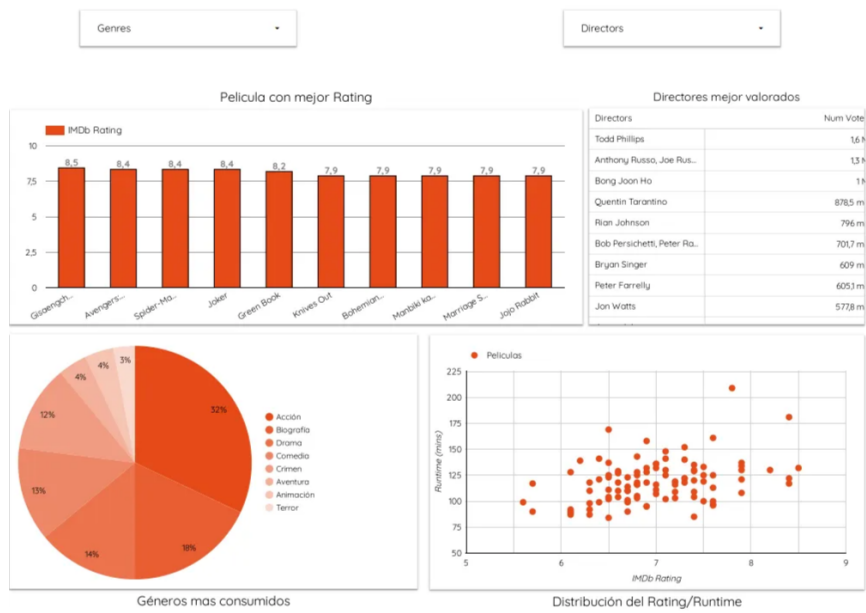
CONCLUSIONES DEL ANALISIS

Luego de hacer la presentacion grafica podemos concluir que:

- La pelicula con mejor rating en el 2019 fue Parasitos con una puntuacion de 8,5 IMDb
- Todd Phillips encabeza la lista con una cantidad mayor a 1,6 millones de votos siendo el director que mejor reputacion tiene
- Los Generos que mas se consumieron en el año 2019 fue Accion, Biografia y Drama

Con relacion a la distribucion o relacion que existe entre el rating y su duracion se puede concluir que:

- No se evidencia una fuerte correlacion positiva entre estas 2 variables. los valores de estas estan dispersas por lo que se recomendaria hacer un analisis mas exhaustivo entre estas mismas variables y generos o incluso por director.



Enlace al Dashboard:

<https://lookerstudio.google.com/s/tzAKX2gPtVU>