

# REAL-TIME ANALYSIS OF TOP-PERFORMING BOX OFFICE MOVIES

JESSE NGUGI

DATA SCIENTIST, **LINC CLOUD CO.**

02/08/2025.

## OUTLINE.

- BUSINESS UNDERSTANDING.
- DATA UNDERSTANDING.
- DATA PREPARATION.
- DATA ANALYSIS.
- BUSINESS RECOMMENDATIONS.
- CONTACT INFORMATION.

## BUSINESS UNDERSTANDING

- The main business objective is to assist a new company navigate the film and production industry.
- Perform in-depth analysis on movie data to identify top-performing movie genres.
- Address stakeholder needs and expectations.
- Guide strategic and informed decision making.

## DATA UNDERSTANDING

- Data used for analysis included csv files and a system database.
- Variables that answered stakeholder expectations were used.
- Anomalies detected in the data were;
  - Missing values in several fields.
  - Data inconsistency such as incompatible data type formats.
  - Duplicates.
- The Data was cleaned prior to Analysis.

## DATA PREPARATION

In this step;

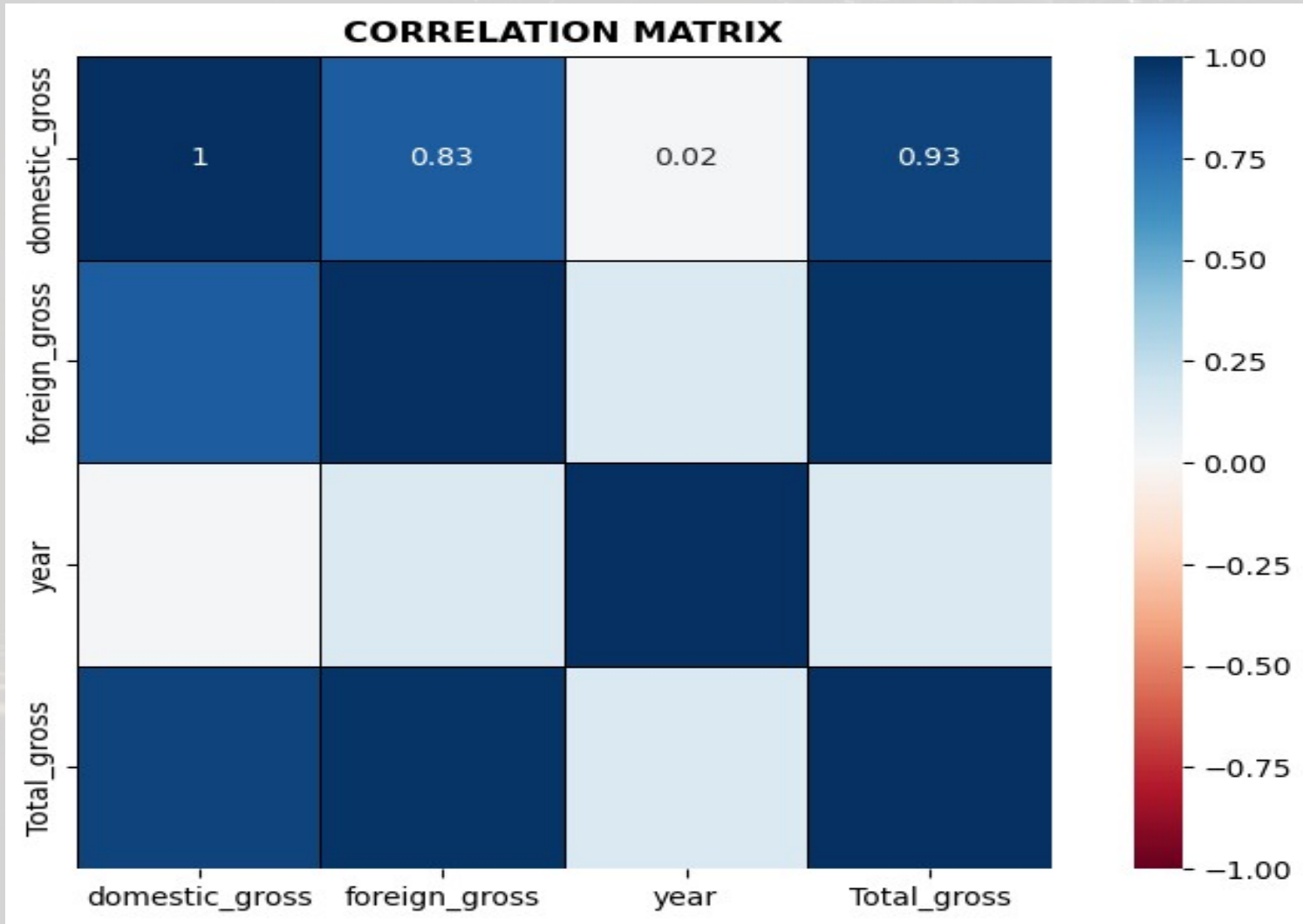
- Relevant datasets were loaded to the environment.
- Tables pulled from the database were combined to a single dataset.
- Missing data was dropped or imputed in various fields.
- Numeric and string columns were standardized.
- Duplicate records were dropped for accuracy.

## DATA ANALYSIS

- The analysis was done using python programming as the base tool within a jupyter environment.
- This step was done to uncover trends, patterns and relationships in the data to generate actionable insights.
- The following visuals were generated from key insights;

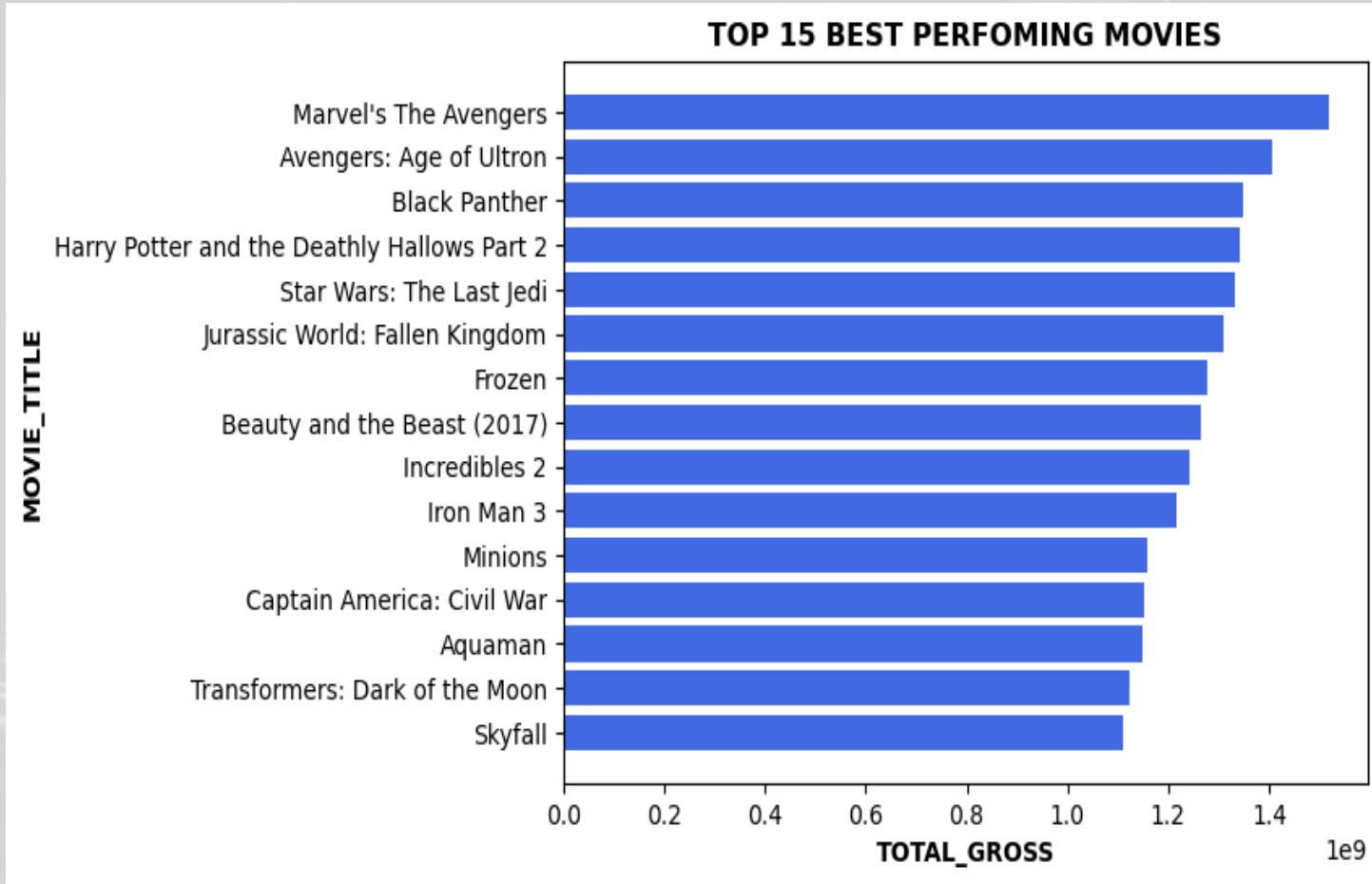


## Correlation matrix for numeric variables .(bom.movie\_gross.csv.gz)



- The strong positive correlation between the domestic gross and Total gross showed that movies that were produces locally also did well internationally.

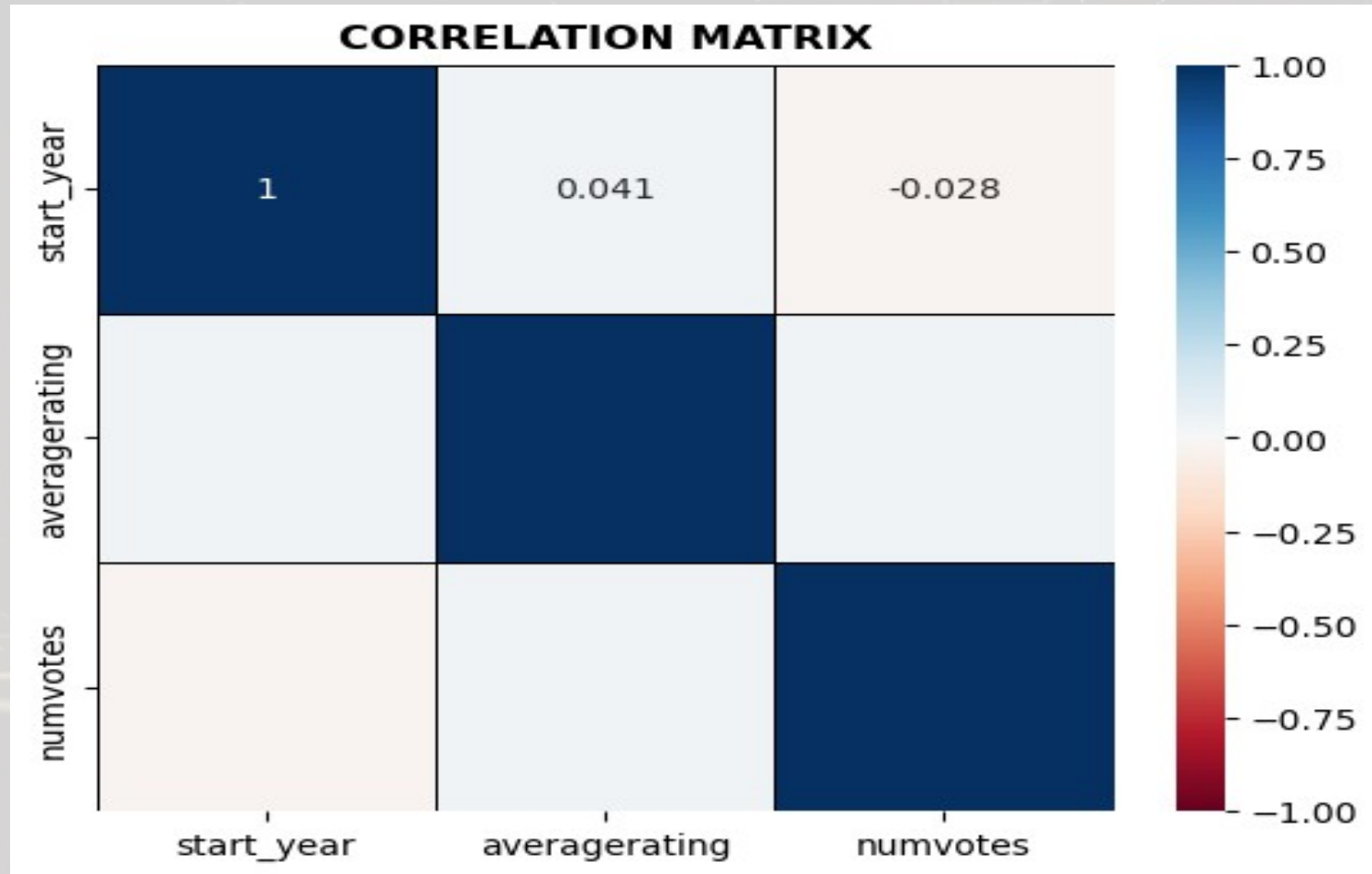
## TOP 15 BEST PERFORMING MOVIES. (*bom.movie\_gross.csv.gz*)



- This bar chart showed that Action and superhero movies generated the highest income.

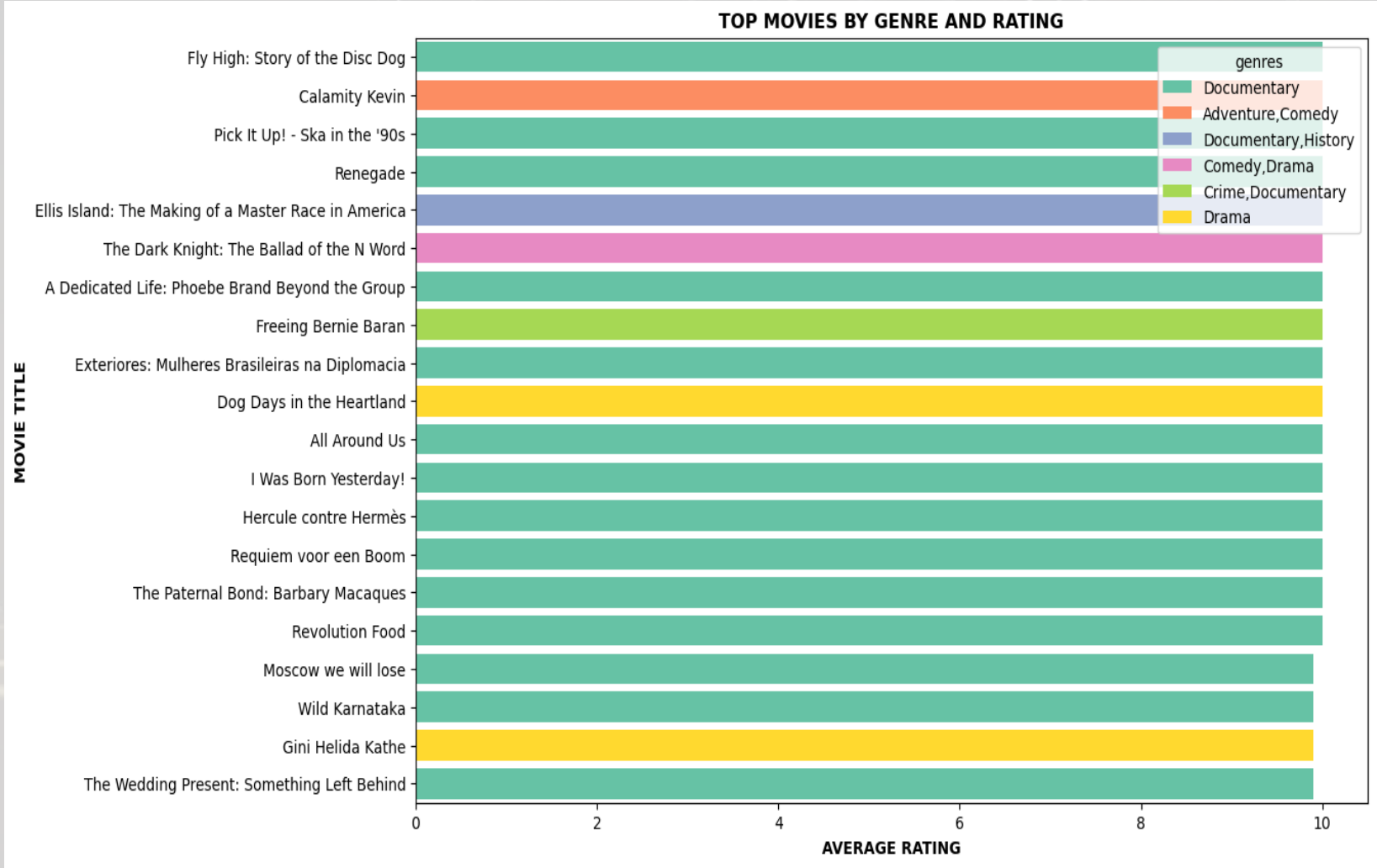


## Correlation matrix for numeric variables. (joined data frame)



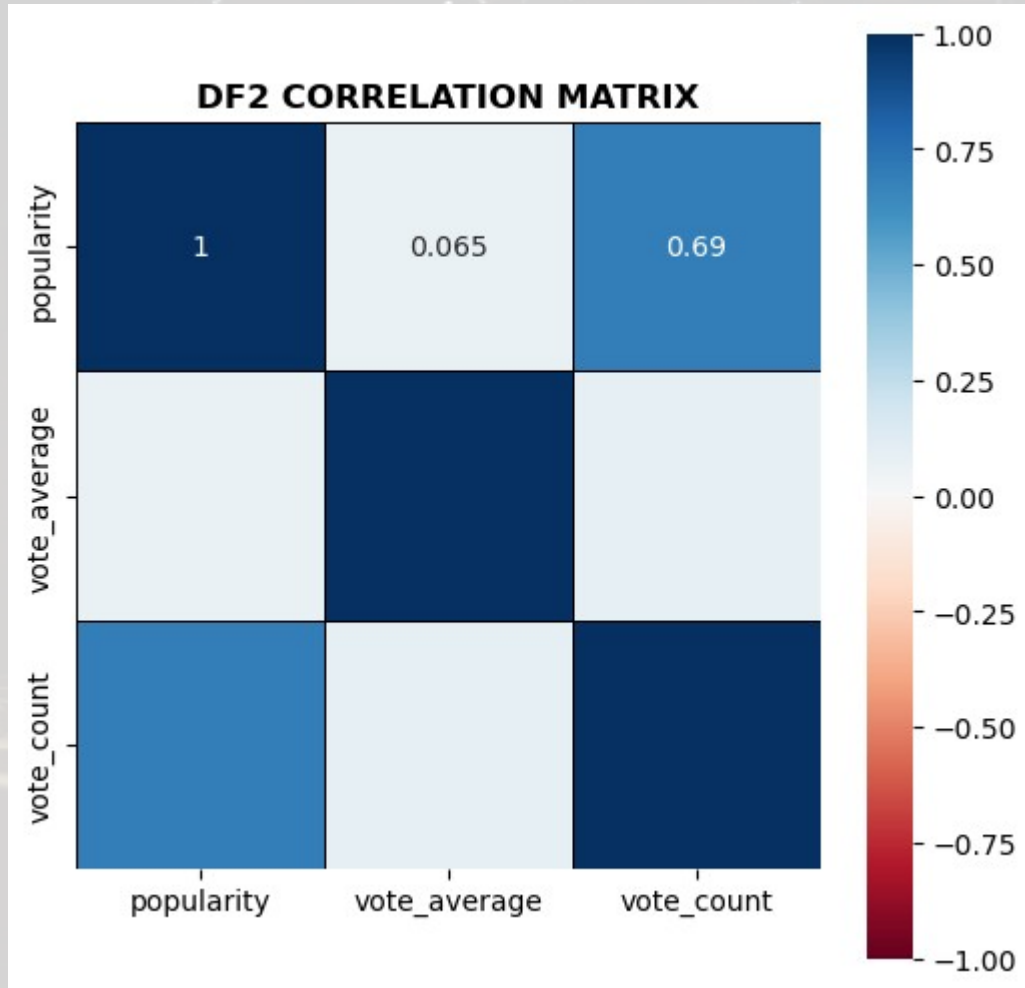
- The weak linear correlation displayed by the heatmap indicate that no linear relationship exist between the variables above and should therefore be treated as independent variables during analysis.

# TOP 20 MOVIES BY GENRE AND RATING. *(joined database tables)*



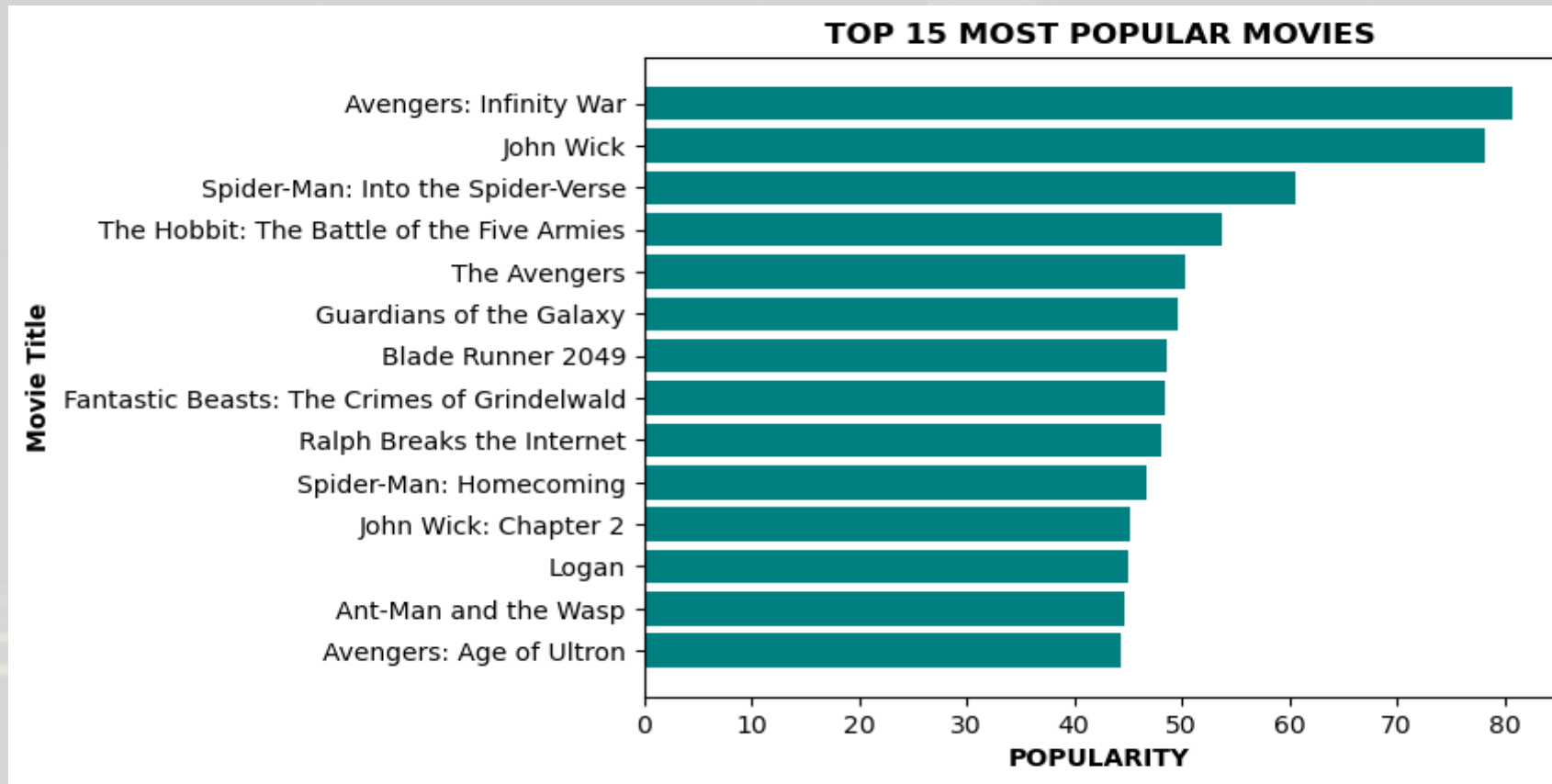
- This bar chart shows that documentaries also did well at the box office.
- Documentary based films appear more compared to other genres on the top list of best rated movies.
- Drama genre produced two movies that were most likely blockbusters(did well) based on their average rating.

## CORRELATION MATRIX. (tmdb.movies.csv.gz)



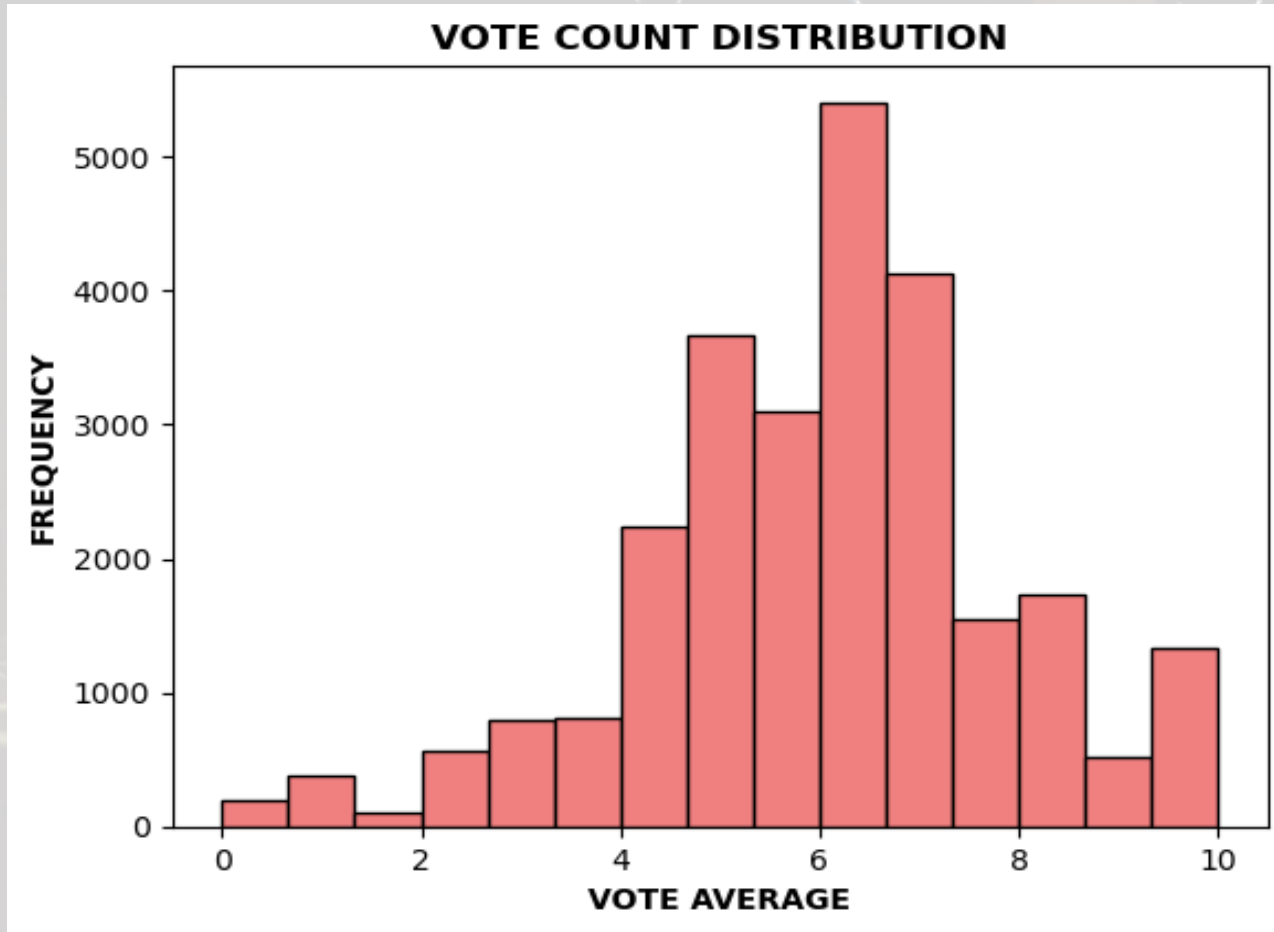
- The strong positive correlation between 'vote\_count' and 'popularity' indicated a strong linear relationship meaning movies with more vote counts were more popular.

## TOP 15 MOST POPULAR MOVIES. (tmdb.movies.csv.gz)



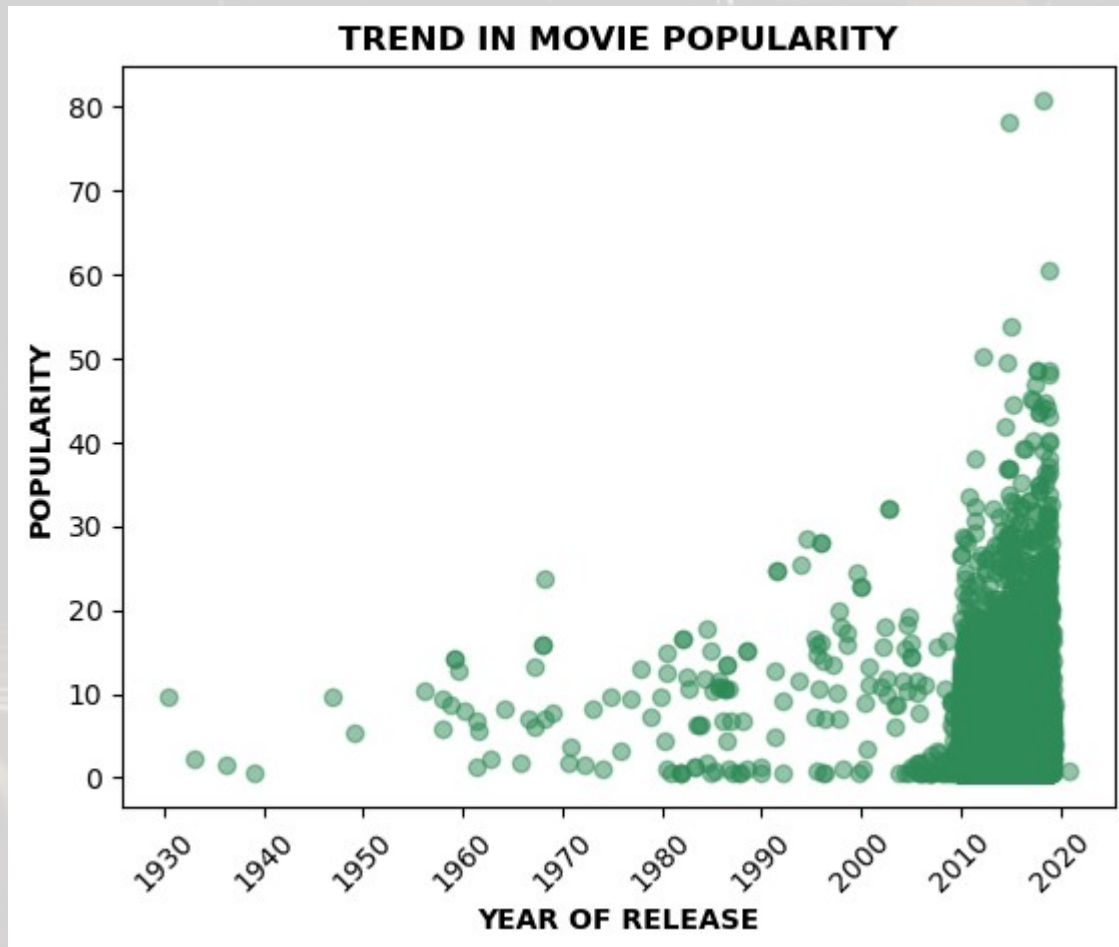
- This bar chart also proves that Action and superhero films performs the best at the box office based on their popularity.

## VOTE COUNT DISTRIBUTION. (tmdb.movies.csv.gz)



- Above 'vote count' histogram showed a normal distribution meaning most movies were fairly rated with a few of them having extremely good or bad rating.

## TREND IN MOVIE POPULARITY. (tmdb.movies.csv.gz)



- There was a sharp increase in movie production from 2005 to around 2022
- Extremely popular movies were produced in 2010 to 2020, it was highly likely they were blockbusters.



## **BUSINESS RECOMMENDATIONS**

1. The company should invest more in the following genres as they generate more income for a long period of time:
  - **Action and superhero films.**
  - **Documentaries.**
2. The company should focus more on films that are relatable internationally. This is evident as top performing movies do well internationally than domestically, in terms of net gross.
3. The company can invest in superhero merchandise as an extra income source mostly targeting children.

## **CONTACT INFORMATION.**

- NAME: JESSE NGUGI
- EMAIL ADDRESS: [jessengugi99@gmail.com](mailto:jessengugi99@gmail.com)
- LINKEDIN URL: [www.linkedin.com/in/jesse-ngugi-247a39203](http://www.linkedin.com/in/jesse-ngugi-247a39203)