DSF-PT08: Group1

By: Gilbert Cheruiyot, Millicent Cheptoi,
Edwin George, Mercy Ayub, James Muthee,
Daniel Muigai, Angela Mwanzia, Reagan Adajo

Agenda

- Business Problem Definition
- Business Understanding
- Data Understanding
- Data Analysis
- Recommendations
- Next Steps
- Q&A (Thank You)

Business Problem

- Our company wants to invest in a new movie studio, but they don't know anything about creating movies. A research is to be done to explore what types of films are currently doing the best at the box office. And translate the findings into actionable insights to help decide what type of films to create.

Key questions to answer

- What are the trends in movie release dates and what impact do they have on revenue?
- Is there a relationship between production budget and profitability and ROI of a movie? i.e. Does a higher production budget automatically result to a higher profitability and vice versa.

Objectives

1. **Analyze Box Office Data**: Identify characteristics of high-performing movies, such as genre, budget, runtime, release date, or lead actors/directors and look for trends in movie popularity and revenue growth over recent years.
2. **Understand Market Demand**: Examine the types of movies audiences prefer (e.g., genre) and Identify seasonal trends (e.g., holiday releases, summer releases) to determine the best time for movie releases.
3. **Identify Key Success Factors**: Explore factors that will indicate success such as production budget and evaluate their impact on box office revenue.
4. **Translate Findings into Recommendations**:
    - Provide specific, actionable insights for the company to make data-informed decisions on the types of movies to create.
    - Recommend strategies around genre, budget, target demographics, and release timing.

Business Understanding

- Data
  - **movie_gross**: Contains data on movie gross earnings
  - **movie_budget**: Contains information about movie budgets
  - **reviews**: Contains movie reviews, loaded with specific encoding
  - **movie_info**: Additional movie metadata
  - **im.db**: SQL database containing additional tables for analysis.

Insights from Data Cleaning

- Initial exploration focused on understanding data spread and identifying potential inconsistencies or missing values.
- Key statistics helped determine the completeness and reliability of each dataset for further analysis.

Data Analysis & Key Findings

- Descriptive Statistics of
  - movie_info Dataset: 1,560 movie records, 12 columns, Mean ID Value: 1,007.30 and Standard Deviation: 579.16, indicating variability around the mean.
  - Data Types: The majority of columns contain text data (11 text columns), with only the ID column as an integer.

- Data Structure of movie_

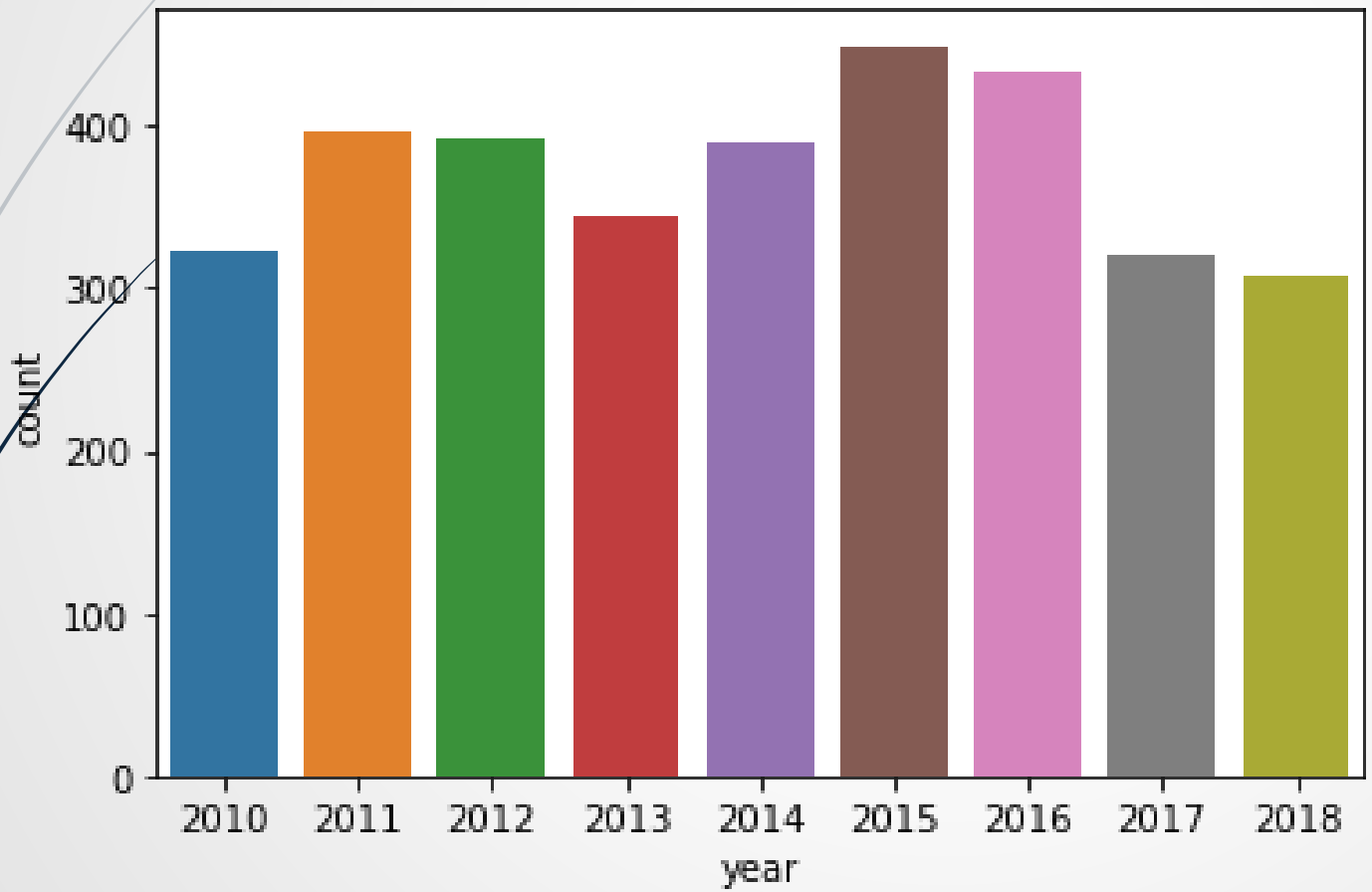
Checking and Handling Missing Values

- Missing Values Analysis:
  - Key columns with high missing values: Studio, box_office, and currency columns had over 50% missing data.
- Data Cleaning:
  - Columns with more than 50% missing values were removed to maintain data quality.
  - Remaining missing values in categorical columns were filled with the mode (most frequent value) for consistency.

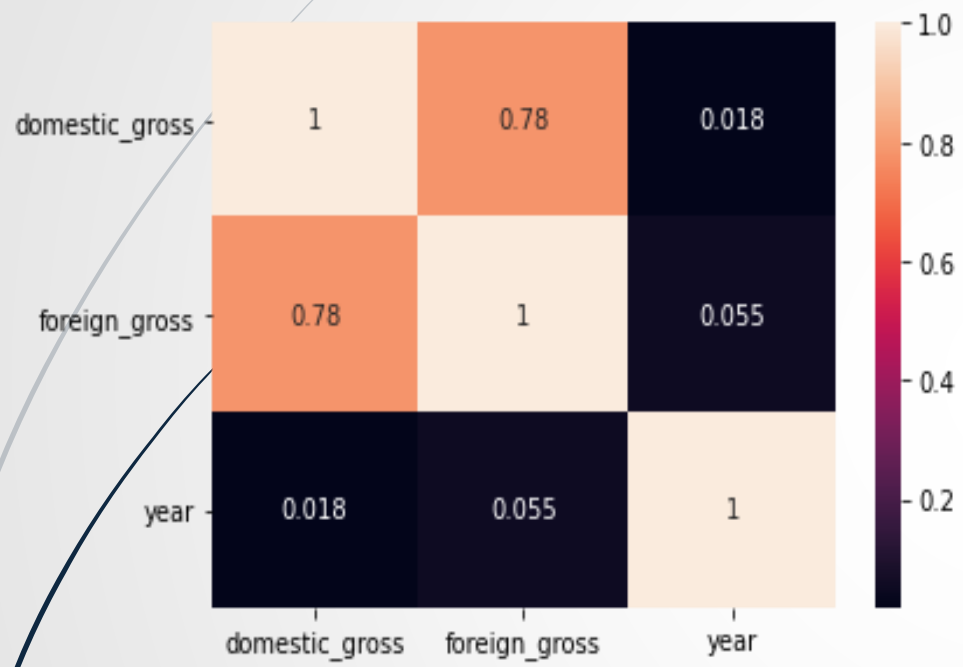
Data Analysis

- Univariate Analysis
  - Check the production rate.....
- Bivariate Analysis
  - Relationship between Production_budget vs domestic_gross vs worldwide_gross.
  - Trends on production(start_year)
- Multi-variate Analysis

Production Numbers per Year

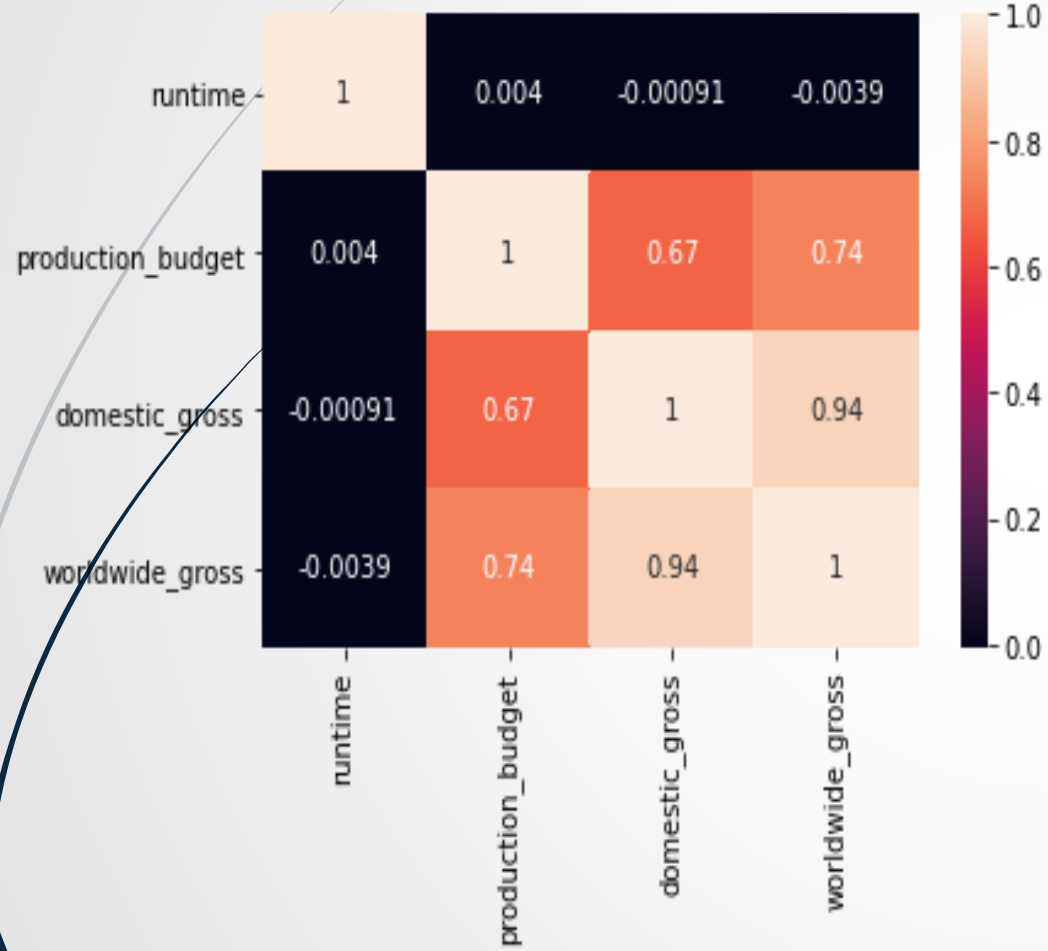


Statistical Analysis – Movie gross



Finding

- **Relationship Between Domestic and Foreign Earnings**
  - Movies that earn more domestically also tend to do well internationally. In other words, if a movie makes a lot of money in its home country, it's likely to earn a good amount abroad as well.
  - However, the year a movie was released doesn't seem to have a strong effect on its earnings, either domestically or internationally.
- **Conclusion**
  - For predicting movie earnings, the strong relationship between domestic and international earnings is key. Understanding this connection can help us make better predictions about a movie's potential success worldwide.



Key Insights

- **Production Budget and Earnings**: Movies with higher production budgets tend to earn more both in the U.S. and worldwide. The link is especially strong for worldwide earnings, meaning a larger budget often results in higher global revenue.
- **Runtime and Earnings**: The length of a movie (runtime) doesn't significantly impact its budget or earnings. In fact, there's a very slight negative relationship, suggesting that longer or shorter movies don't necessarily earn more or less.

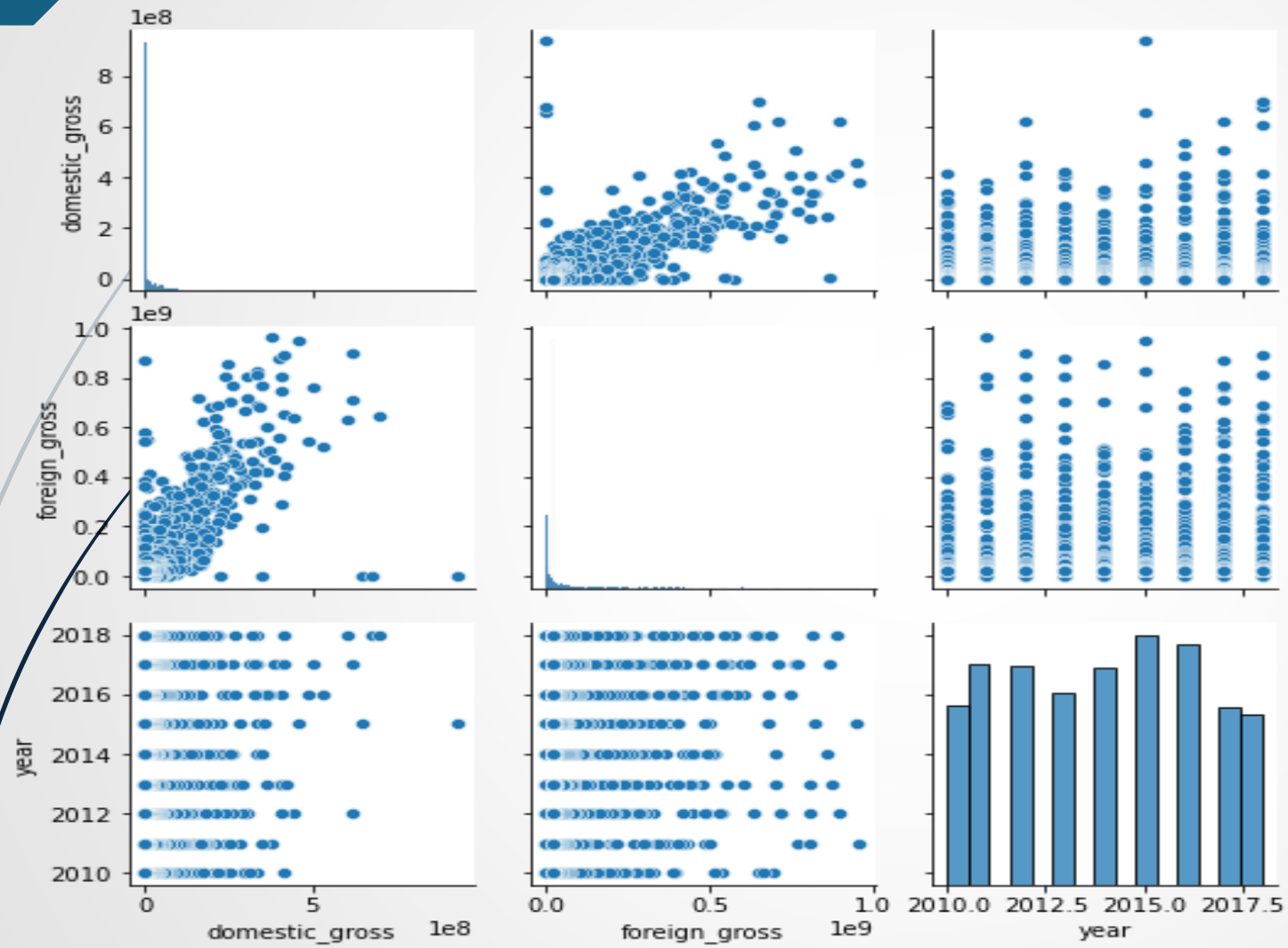
Conclusion

- **Budget Matters**: Investing more in a movie's production generally leads to higher earnings both domestically and worldwide. This insight could guide budget decisions for maximizing potential revenue.

Data Modelling – Simple Linear Regression

- The Ordinary Least Squares(OLS) model was used to determine the relationships between different variables: domestic_gross, production_budget and worldwide_gross, foreign_gross.
- The model was used to predict various variables on the data.

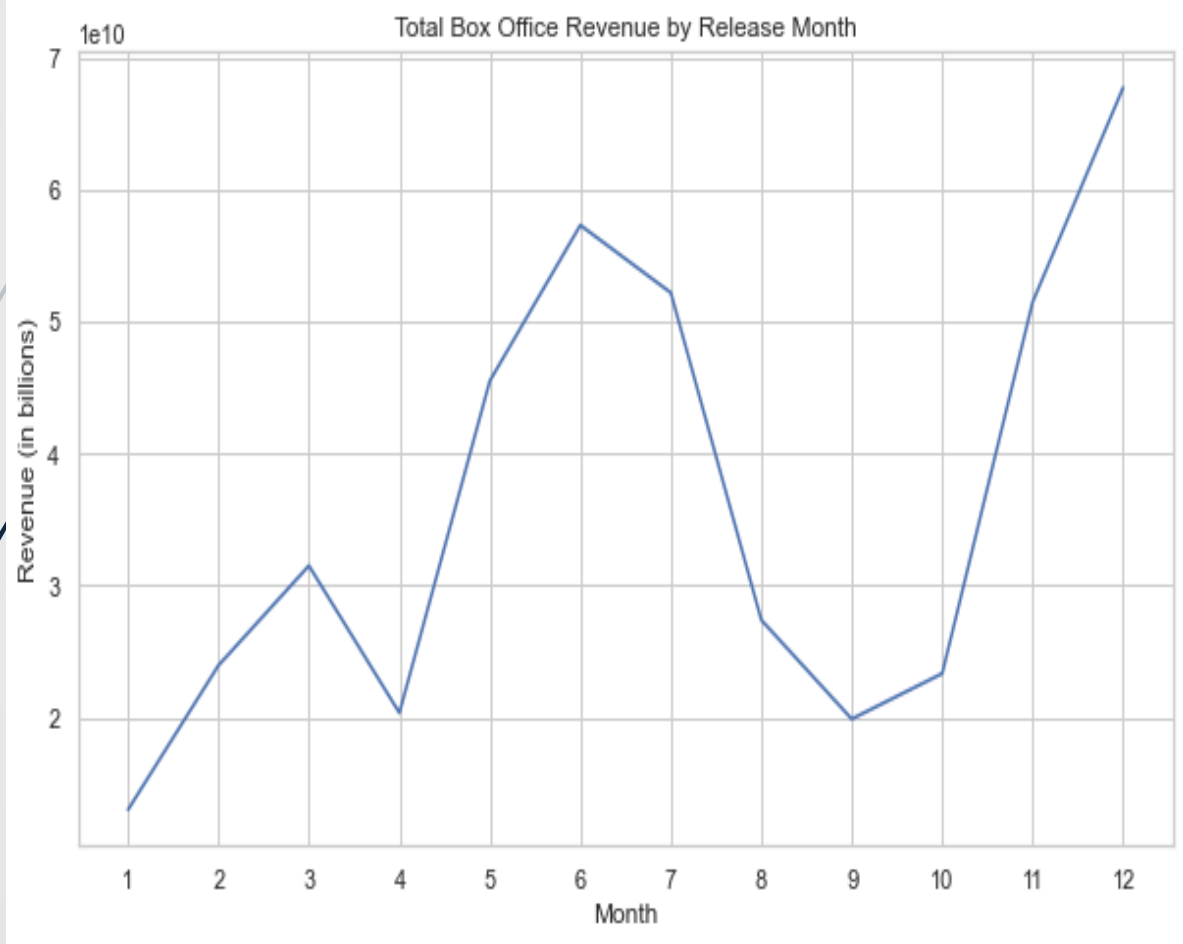
Findings:



- There is a positive linear relationship between domestic gross and foreign gross. Most of the values lie between 0-2e8.

id
runtime
production_budget
domestic_gross
worldwide_gross
1e8
1e9

Revenue per Release Month



- The revenue is higher in the months of June and December, More movies are released during these months

Recommendations

- To maximize revenue, the movie studio should focus on releasing their films in the months with the highest revenue, such as December, June, and July.
- They should prioritize releasing high-budget films with a high return on investment (ROI) to ensure their production budgets are not wasted.
- They should invest in marketing and promotional efforts to reach a wider audience and generate more revenue.
- They should prioritize releasing their films in the top performing genres, such as Musical, Performing Arts, Horror, Science Fiction, and Fantasy. This will help maximize their box office revenue and attract more viewers.

Next Steps

- Thank You ☺