

EDA ANALYSIS FOR BOX OFFICE DATA. (by Alan. O)

ANALYSIS OF MOVIE DATA FROM:

1. BOX OFFICE MOJO.
- 2.THE MOVIE DATABASE.
3. THE NUMBERS DATABASE.

BACKGROUND INFO.

- **Business Problem.**
- Microsoft sees all the big companies creating original video content and they want to get in on the fun. They have decided to create a new movie studio, but they don't know anything about creating movies. You are charged with exploring what types of films are currently doing the best at the box office. You must then translate those findings into actionable insights that the head of Microsoft's new movie studio can use to help decide what type of films to create.

Research Objectives.

- 1. Identify the types of films that are currently doing the best at the box office.
- 2. Analyze the characteristics of these successful films, including their genre, budget, release date, target audience, and critical reception.

Methodology.

- To achieve our research objectives, we will use exploratory data analysis to analyze data on films that have been released in recent years, including their box office performance, genre, budget, release date, target audience, and critical reception. We will use statistical analysis and data visualization techniques to identify patterns and trends in the data.

Expected Outcomes.

Recording the Experimental Design.

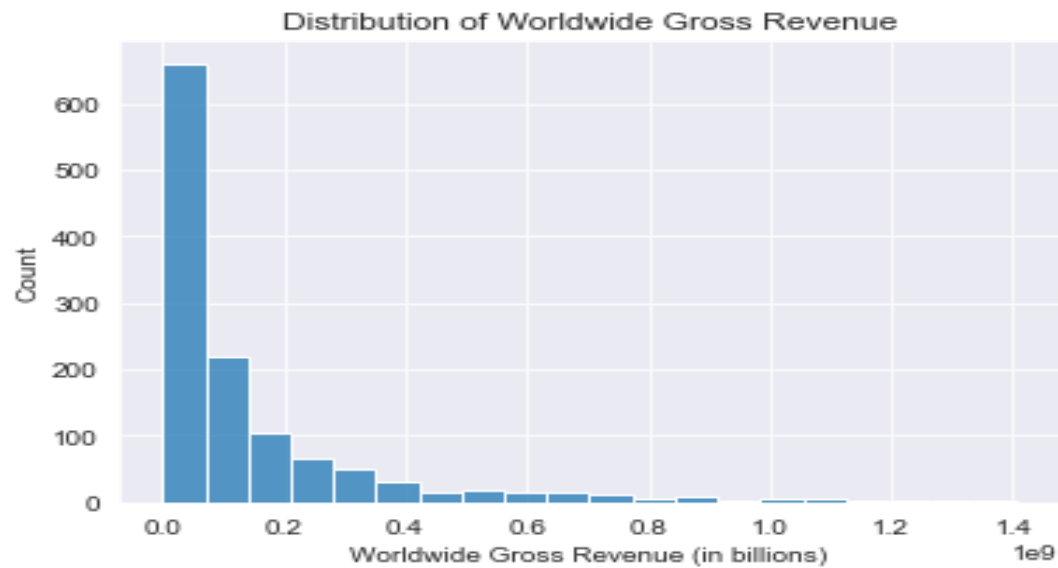
- **Expected Outcomes.**
- Based on our research, we expect to identify the types of films that are currently successful at the box office, as well as the characteristics that contribute to their success. We will provide actionable insights for Microsoft's new movie studio to use in creating films that are more likely to resonate with audiences and generate revenue.
- **Recording the Experimental Design.**
- .upload and read our csv files
- .clean our dataset
- .perform EDA
- .Conclusion
- .Recommendation

Data Loading and Pre-processing.

- **Data Loading and Pre-processing.**
- To start with the analysis, we first need to import the necessary libraries and load the dataset into our Jupyter Notebook. Here's how you can do that:
- **Importing Libraries.**
- We will be using the following libraries for this project:
 - 1.numpy
 - 2.pandas
 - 3.matplotlib.

Exploratory data analysis.

- 1. Univariate analysis.



The histogram above shows that most of the movies in the dataset have worldwide gross revenue less than 500million. *There are some movies that have grossed over 500 billion worldwide*, but they are few in number. The distribution is right-skewed, indicating that there are a few movies that have made a lot of money while most have made less.

CONTINUATION OF EDA.

BIVARIATE ANALYSIS.

- we can perform bivariate analysis to explore the relationship between two variables. One example is to explore the relationship between the production budget and the worldwide gross revenue.



Feature engineering.

Feature engineering is the process of selecting, extracting, transforming, and refining relevant data features from raw data to create an effective input dataset for a machine learning model. It is one of the most critical aspects of the data science workflow because it directly impacts the quality and accuracy of the model's predictions.



CONCLUSION.

- The movie industry is a profitable and growing industry, but success is heavily reliant on brand recognition and established franchises.
- Timing of movie releases, particularly during the summer months, can impact revenue significantly.

RECOMMENDATION.

- Studios should continue to invest in established franchises to ensure continued success.
- Strategic timing of movie releases, particularly during the summer months, should be considered to maximize revenue potential.
- Studios should also factor in reputable movie directors as this has been seen to impact movie success.