Prediction of Movie Profit

Arufa Khanom, Zachary Formes, Ana Ramirez

1 - State the problem you are solving

- The Movie industry generates over $100 billion in revenue every year.
- However, over 80% of films produced do not make a profit.

⇩

Understanding which factors lead to more profitable movies could help lead to more profitable movies that people enjoy more

2 - Explain what data sources you are going to use

We plan to use multiple Kaggle datasets together to allow our model to choose from many different predictors. We also may scrape data from the internet to develop new features.

Initial Dataset: *IMDB movies dataset*

3. Specify your outcome variable and the predictors used to predict this outcome.

Outcome Variable

Movie profit (revenue - budget/expenses)

Predictors

- 1. Genre
- 2. Original language
- 3. Popularity (TMDB rating)
- 4. Production companies
- 5. Budget
- Rating (data must be scraped from Rotten Tomatoes)
    - 6. Include audience review
    - 7. Critics review
- 8. Award nominations and wins (data must be scraped)
- 9. First trailer views (data must be scraped from YouTube)
- 10. Release date

4. Description of the dataset

The dataset has 20 columns and more than 7 million rows (which we will reduce through data cleaning as some values are set as null or 0).

Columns:

- Movie ID
- Movie Title
- Genre
- Original Language
- Overview
- Popularity
- Production_Company
- Release Date
- Budget
- Runtime
- Status (Released/Not)
- Tagline
- Vote_Average
- Vote_Count
- Credits
- Keywords
- Recommendations
- Revenue

Dropped Columns:

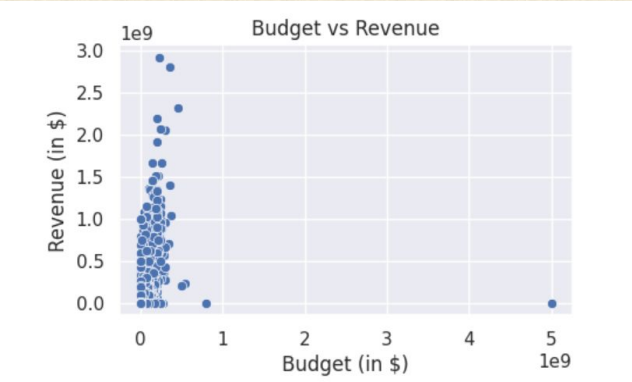
- Poster_Path
- Backdrop_Path

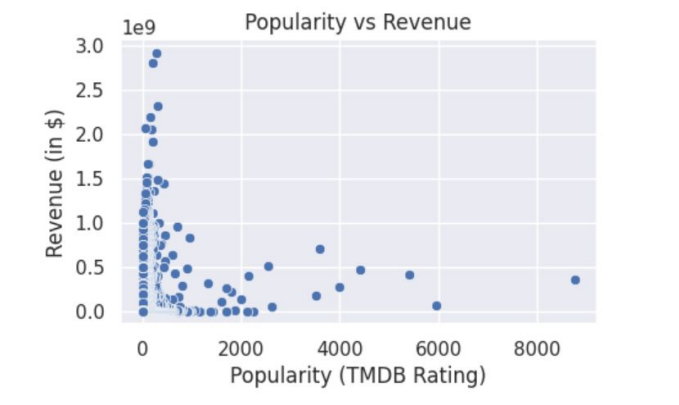
Revenue Statistics:

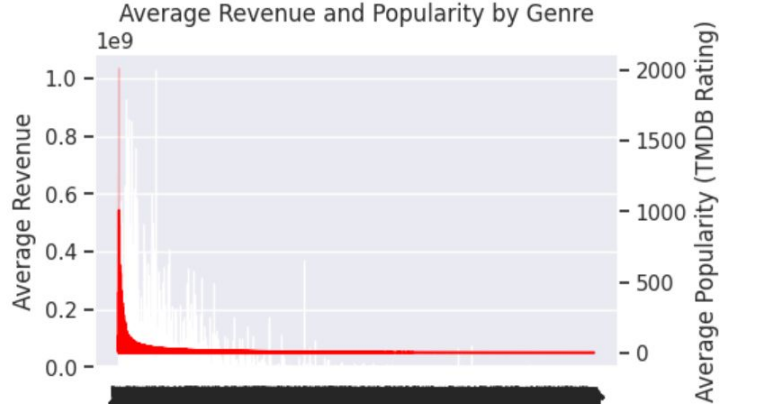
Mean: $993,672
Lowest: $0
Highest: $2,920,357,254
(We will remove all the data with $0 revenue)

We are also going to use other movie based database to add more related columns (e.g. Lead Actor, Theatres, Total Gross, Distributor)

5. Visualize a couple of interesting relationships between some of your predictors and the outcome.







THANK YOU