

TMDB Movie Dataset Analysis







Overview and Business Understanding

- This project involves exploratory data analysis on two box office company movie datasets ,that is, movie revenues and TMDB movie datasets.
- The aim is to generate actionable insights that Microsoft could use as a guide when producing their own movies.



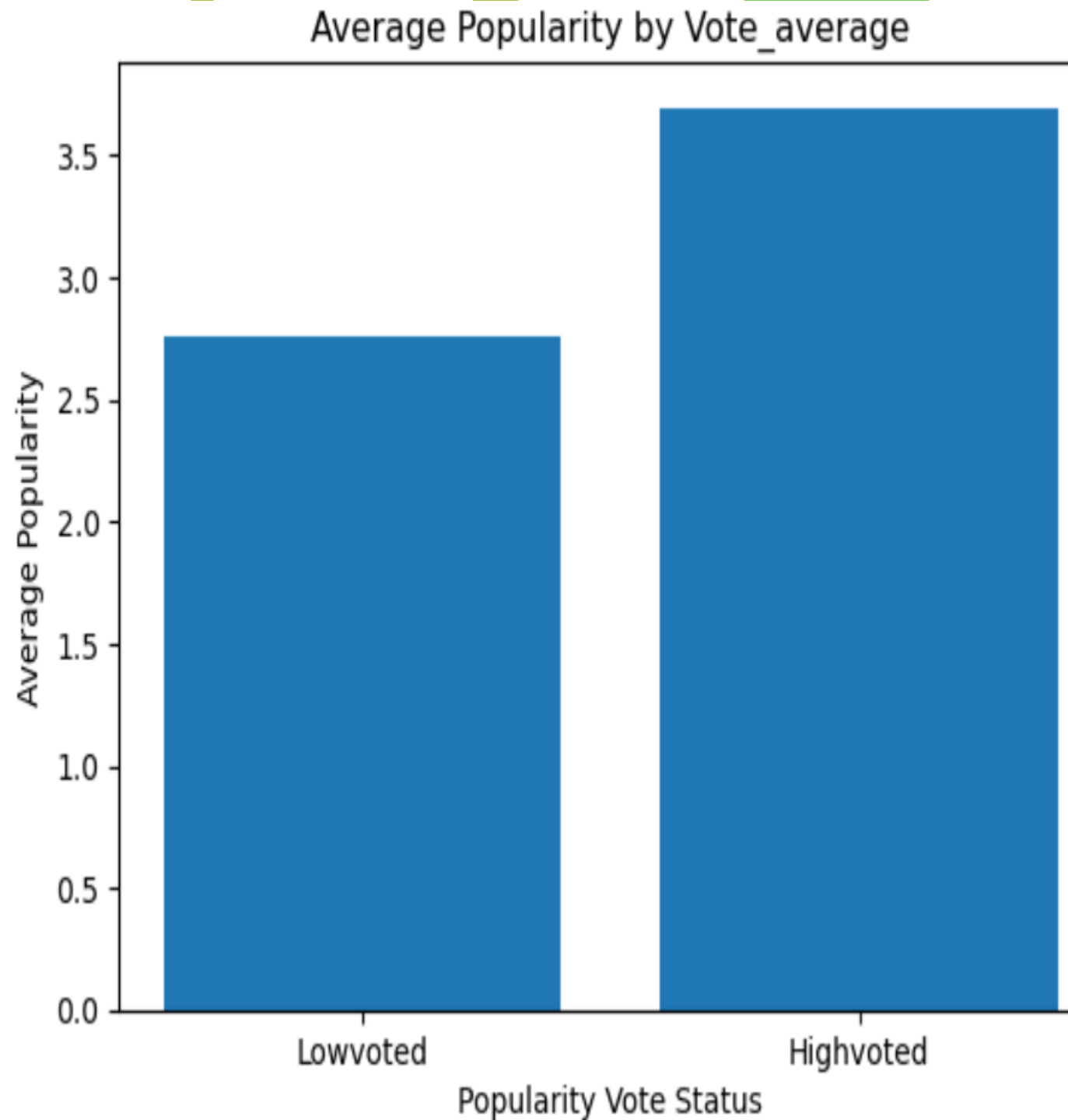


Data Understanding and Analysis

- The budgets dataset contains movies financials such as production budget, worldwide gross and domestic gross.
 - The TMDb dataset contains each movies metadata such as movie title, genre, release date, popularity, vote count to say the least.
 - Combining the two datasets into one generates a dataset with 26855 rows and 15 columns
 - The merged dataset had its own shortcomings such as null values in some rows, incorrect datatypes in the financial columns and unnecessary columns that were not useful for the exploratory analysis
 - The data cleaning and preparation process involved dropping the unnecessary columns, filling the financial nulls with the mean as well as correcting the financials columns datatypes
 - After data cleaning, the dataset had 26855 rows and 13 columns.
- 
- 
- 
- 

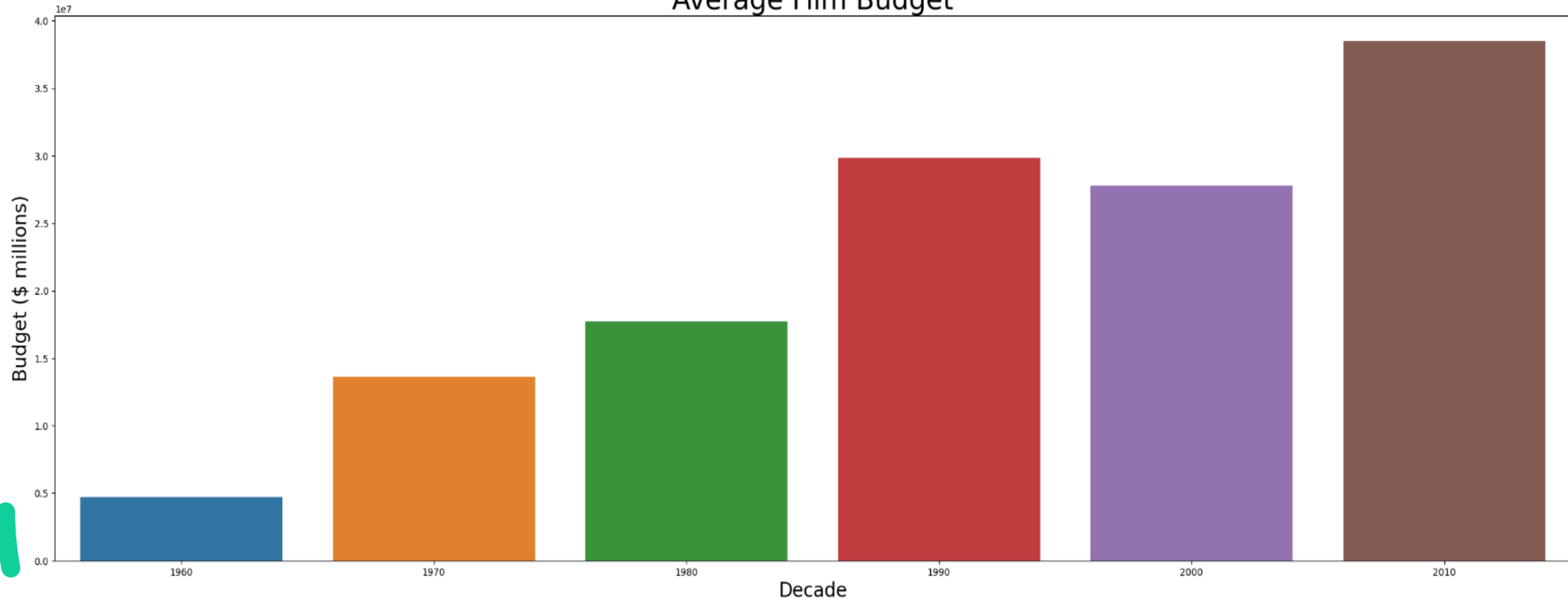
Recommendations

- From the analysis, it was clear that movies that have high vote counts were the most popular movies.
- Highly voted movies refer to movies whose vote_average was above 6.0
- Low voted movies had an average popularity of about 2.7 while high voted movies had an average popularity of about 3.7
- With these findings, it would be appropriate for Microsoft to produce movie genres that were highly voted as the profit would also be high.



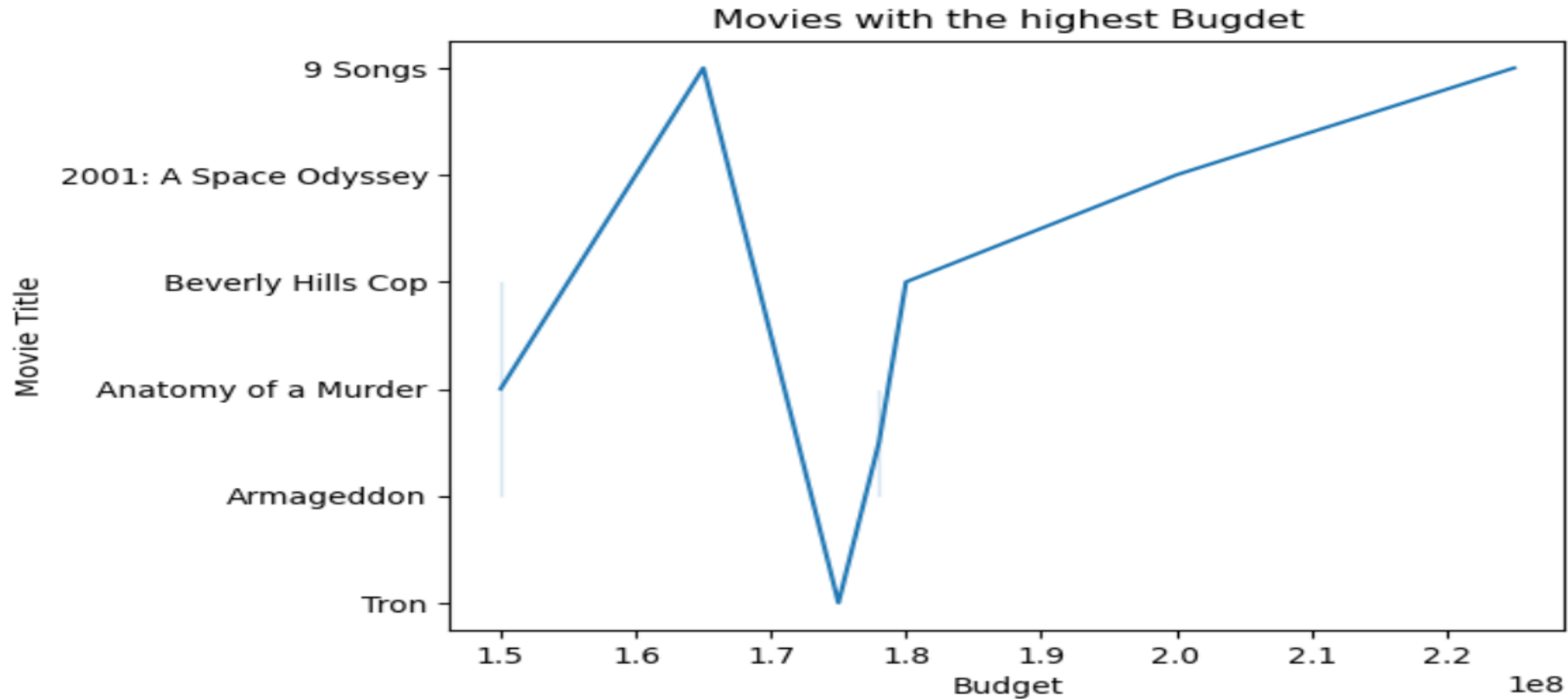
Budget change from one decade to another

Average Film Budget



- 2010s saw the highest film budget on average while 1960s was the lowest.
- There is an increase in film production from 1960s to 1990s.
- From the trend analysis, the budget is bound to increase over the decades. This means that Microsoft would have to have enough budget to be able to produce movies within each 10 years.

Movies with high budgets



- The the top three movies with the highest budget are 9 Songs, Beverly Hills Cop and Anatomy of a Murder.
- These movies were also the most popular movies.
- It seems that budget and popularity have a positive correlation in that the higher the budget the more popular the movie is.

Summary Findings

- Movies that have high vote counts were the most popular movies.
- The budget needed to produce movies is bound to increase every decade
- The top three movies with the highest budget are 9 Songs, Beverly Hills Cop and Anatomy of a Murder. These movies are mainly belonged to action and drama genres



Thank You

