

IMDB MOVIE ANALYSIS

Career Fair Capstone

Problem Statement:

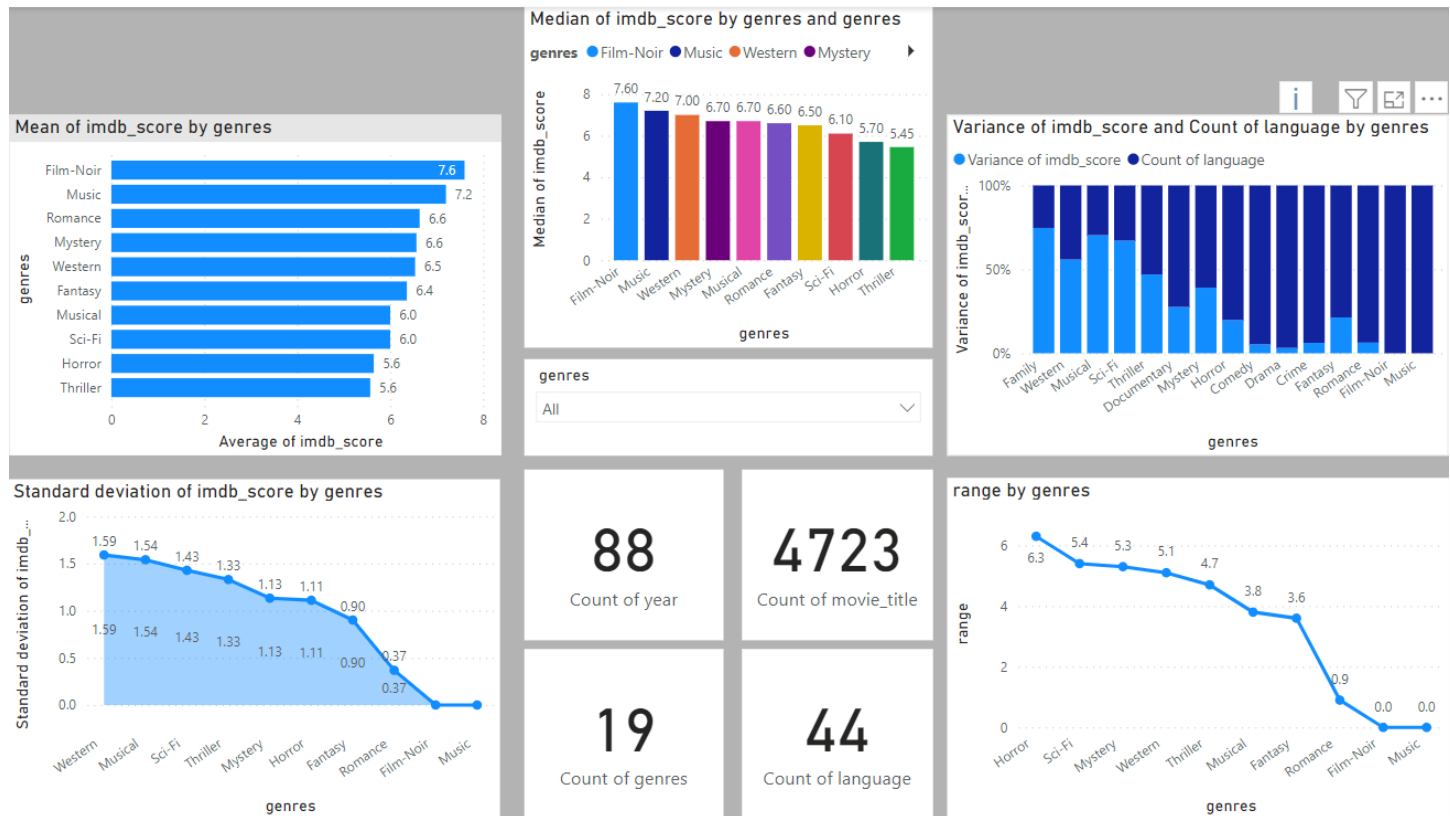
The dataset provided is related to IMDB Movies. A potential problem to investigate could be: **"What factors influence the success of a movie on IMDB?"** Here, success can be defined by high IMDB ratings. The impact of this problem is significant for movie producers, directors, and investors who want to understand what makes a movie successful to make informed decisions in their future projects. Consider this as an open-ended question and come up with more analysis points.

Data Cleaning:

- This step involves preprocessing the data to make it suitable for analysis. It includes handling missing values, removing duplicates, converting data types if necessary, and possibly feature engineering.
- The NULL imputation was used to fill the NULL values.
- The column “genres” contains many categories separated by “|”. It was splitted using string split method and splitted categories were exploded into separate rows.
- The mean, median, mode, standard deviation of the IMDB score was calculated.
- The consolidated data frame was converted into excel or csv file and uploaded in Power-BI for further visualization.

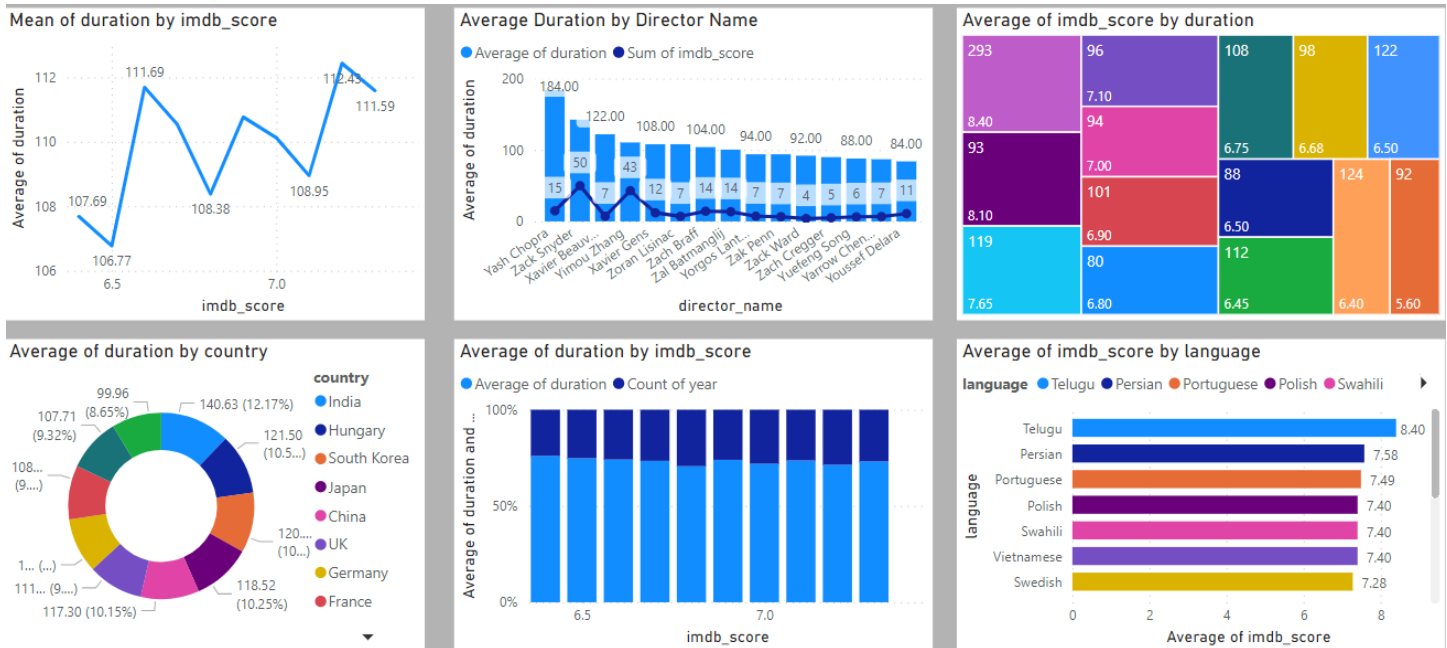
Data Analysis:

A. Movie Genre Analysis:



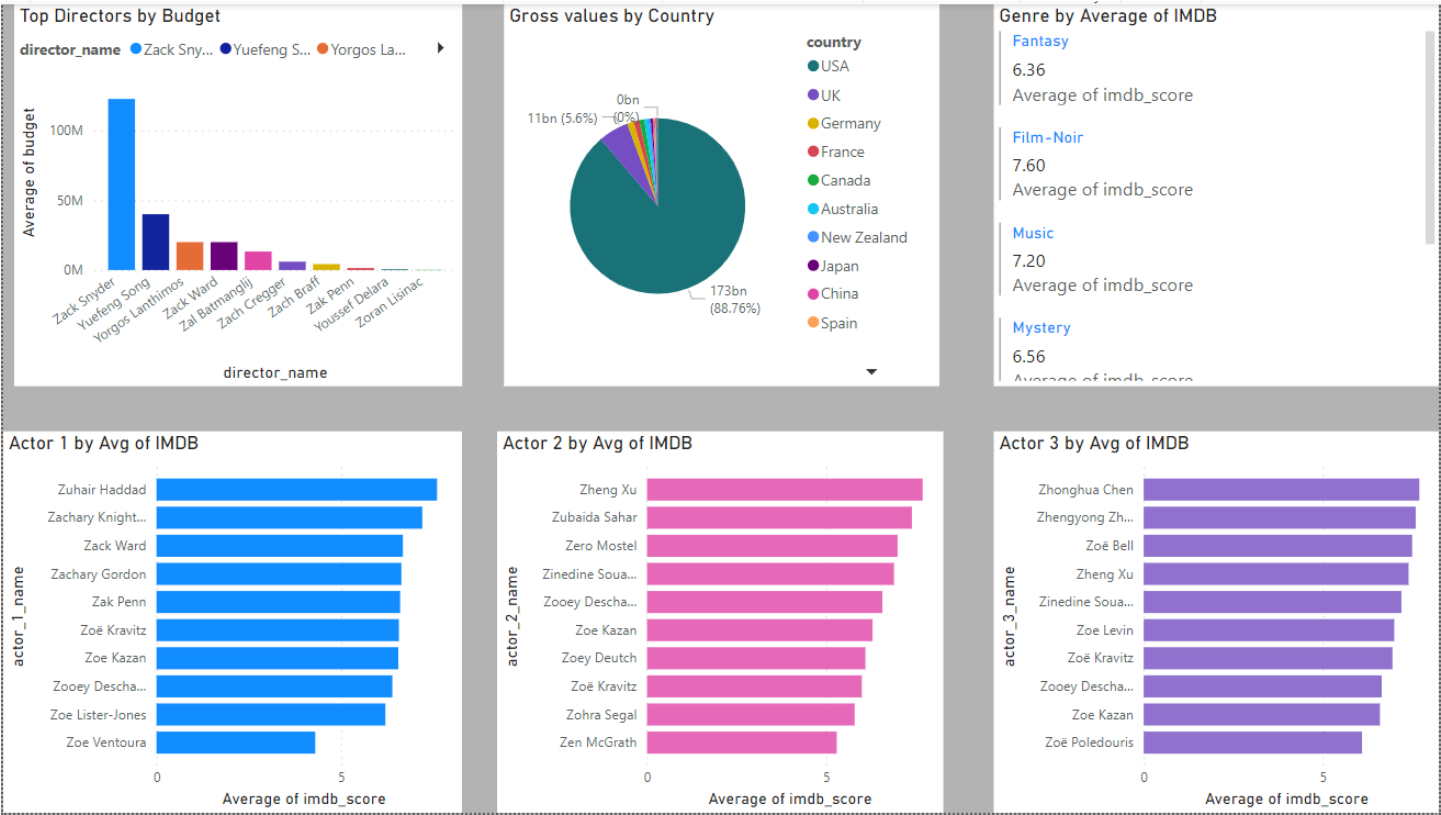
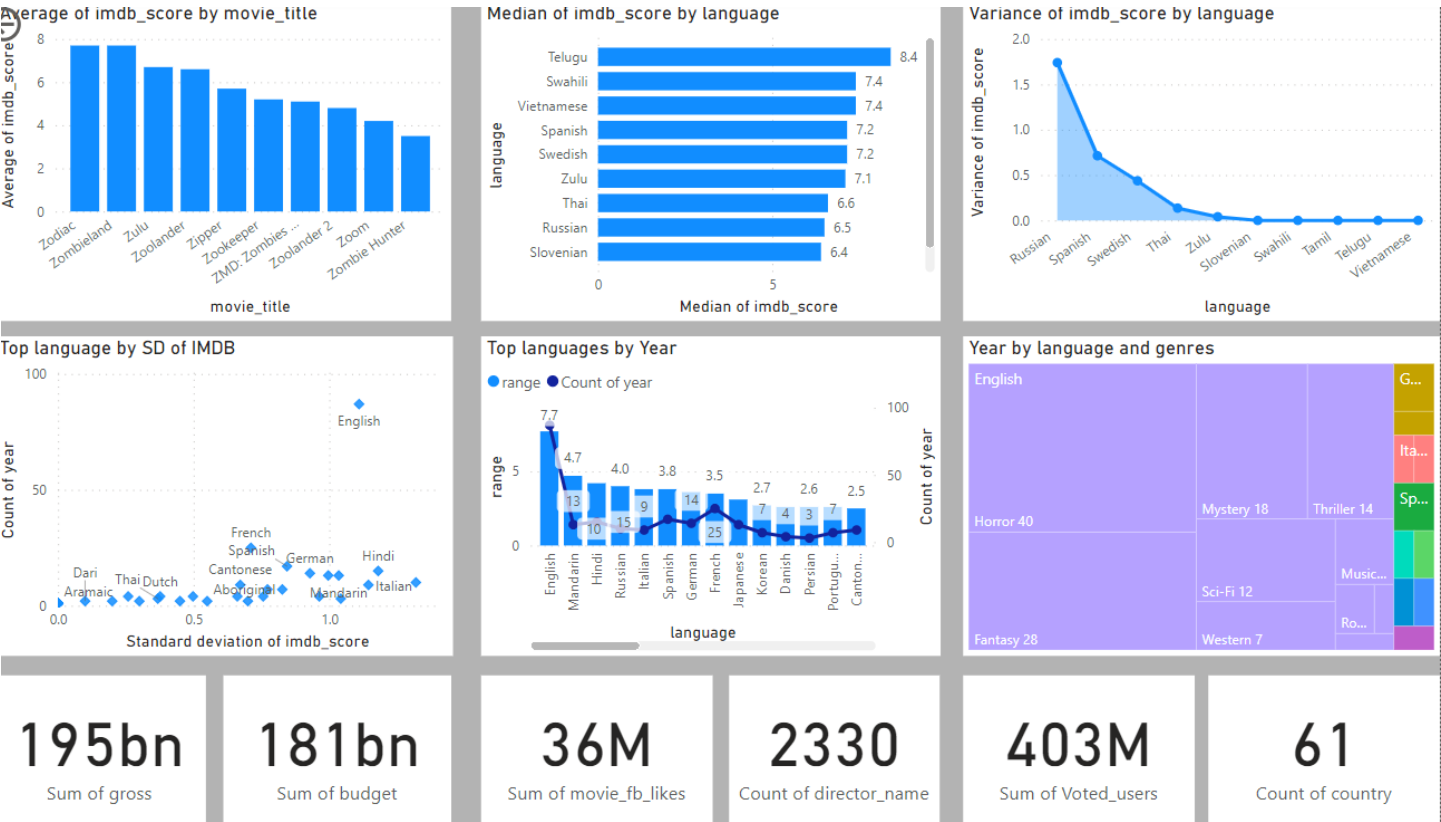
- In this analysis the Statistical analysis Mean, Median, Variance, Standard Deviation, Range of the IMDB score impacted with genres was analyzed.
- By Mean of IMDB score with genres the Film-Noir is the top one with the average IMDB score of 7.6.
- The Median of IMDB score also Film-Noir is the top in one genres.
- But the variance analysis the Family genres is the top one.
- And the overall year is 88, Overall movie count was 4723, count of language is 44 and genres is 19.

B. Movie Duration Analysis:



- The next task is analyzing the durations.
- The mean of IMDB score above 7 have an average duration is 112.43.
- The director Yash Chopra was the top one with the average duration value is 184.0.
- By country wise the India is the first one have an high average of duration 140.63.
- But the language wise the Telugu is the top one with high average IMDB score by year wise is 8.40.

C,D,E : Language Analysis, Director Analysis & Budget Analysis:



- The balance analysis is based on Language, Director and Budget analysis.
- By language wise Telugu is the first one with high average IMDB score 8.4.
- In director wise the Zack Snyder was the top one with the budget values.
- By Gross values the USA is the top most one using high Gross values 173 bn.
- Zuhair Haddad, Zheng Xu, and Zhongua Chen these are the 3 actors is the top one with the average of IMDB scores.
- By the average of IMDB the Fantasy Genres is the top one with The 6.36 IMDB score.
- The sum of Budget is 181 bn.
- Overall directors count is 2330.