**Project – 5**

**IMDB Movie Analysis**

**Project Description**

This project aims to investigate "What factors influence the success of a movie on IMDB?" Here in this project we will analyze movie genre, duration, language, director and budget analysis defined by high IMDB ratings. This analysis is crucial for movie producers, directors, and investors to make informed decisions regarding future projects.

**Approach**

* Dataset was downloaded and imported into Microsoft excel
* Ensured dataset was clean and free of missing values
* Then we started with analysis using pivot tables and its function.
* For genre analysis we calculated no\_of\_movies, mean\_imdb, max\_imdb, min\_imdb, StdDev\_imdb and var\_imdb from pivot table function and we derived the median\_imdb, mode\_imdb and range\_imdb using function in excel sheet.
* Then we generated graph of no\_of\_movies v/s genre to understand the distribution and a line graph of imdb\_score v/s genre to see the trend between there descriptive statistical function.
* For movie duration analysis we analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score using the scatter plot and by observing the trend line.
* For language analysis we have determined the most common languages used in movies and analyzed their impact on the IMDB score using descriptive statistics.
* For director analysis we have calculated the average IMDB score for each director. Used excel's percentile function to identify the directors with the highest scores. Then we compared the scores of these directors to the overall distribution of scores.
* For budget analysis we calculated the correlation coefficient between movie budgets and gross earnings using excel's correl function. Then we calculated the profit margin for each movie and identified the movies with the highest profit margin using excel's max function.

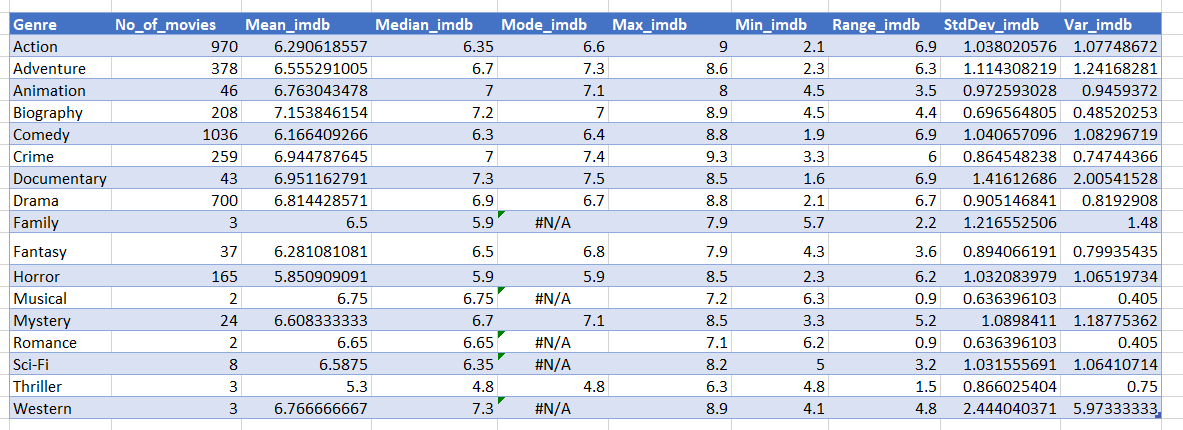
**Tech Stack Used**

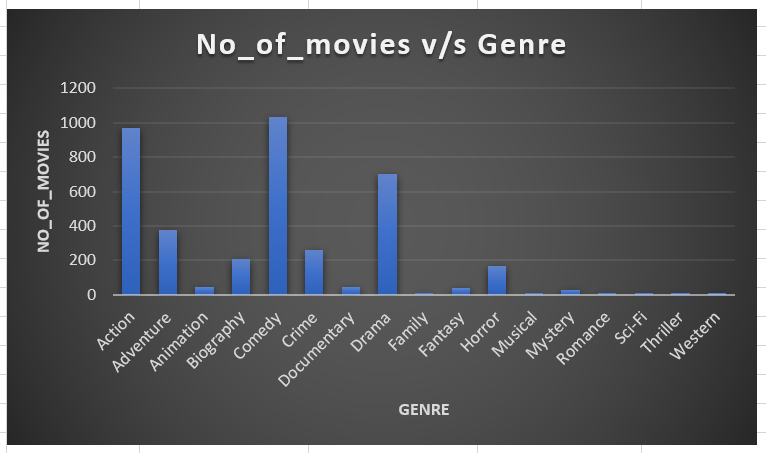
Microsoft® Word 2019 MSO (Version 2501 Build 16.0.18429.20132) 64-bit for data analysis, pivot table creation and chart generation.

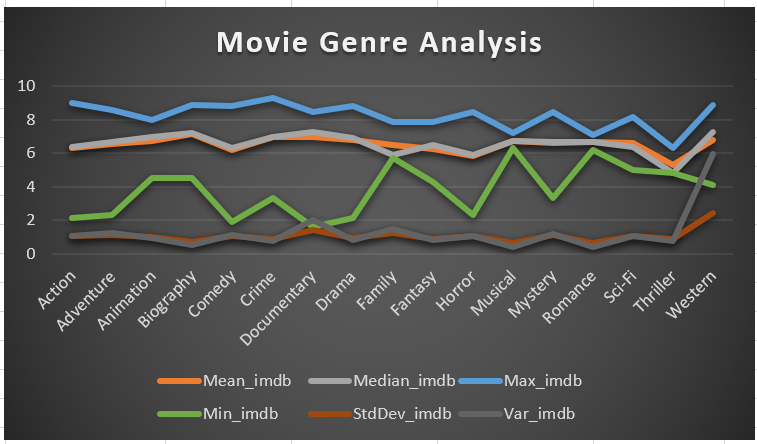
**Insights**

1. **Movie Genre Analysis:**

* Most common movie genre includes comedy with 1036 no. of movies and then action, drama and adventure with 9770, 700 and 378 no. of movies respectively.
* Least common genres include musical, romance, thriller and western.
* Highest mean imdb score is of biography, documentary and crime.
* #N/A appears when no single rating appers more than once in that genre.
* Documentary is having most rating of 7.5 which is highest and most repeated.
* The highest rated movie is having crime genre with 9.3 imdb score and lowest rated movie is genre having action and horror with 2.1 imdb score.
* Higher standard deviation means more unpredictable in audience reception.
* Western movies have most inconsistent ratings while animation and musical movies tend to have stable ratings.

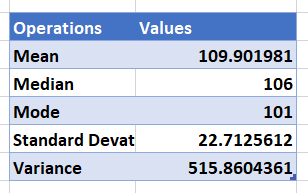
****

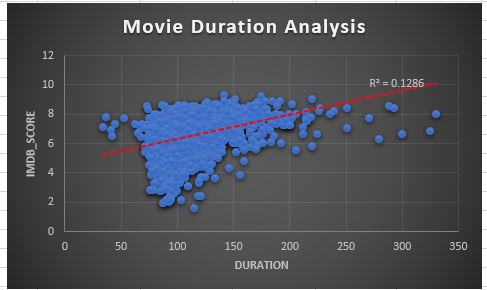
****

****

1. **Movie Duration Analysis:**

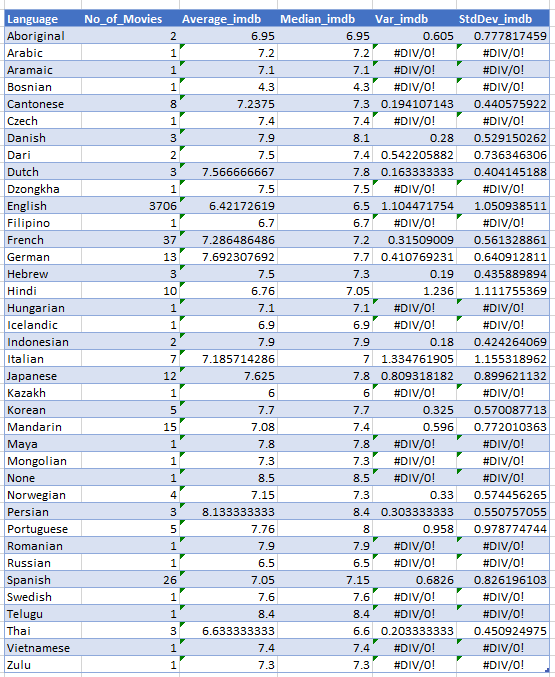
* The scatter plot shows positive correlation between movie duration and IMDB score.
* The R2 value shows that the impact of duration of a movie has less impact on IMDB score and it is not a strong predictor.
* Most movies have duration around 100-110 minutes.
* Longer movies might tend to have slightly higher ratings, but duration only cannot determine IMDB score.

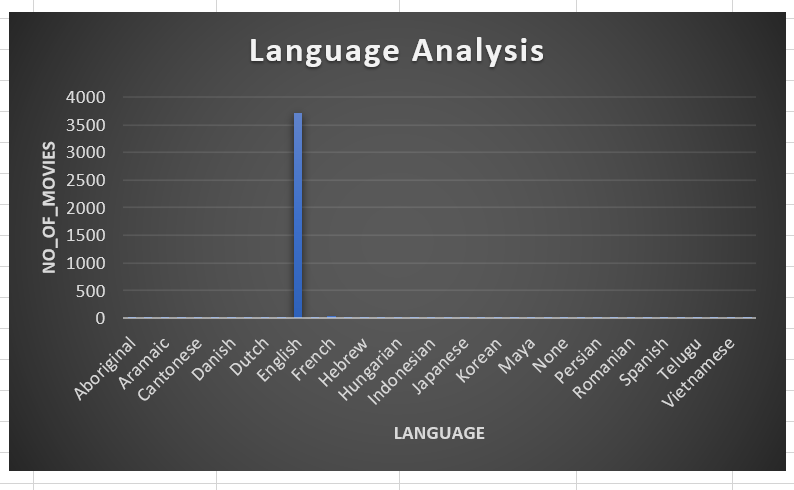
****

****

1. **Language Analysis:**

* With 3706 movies English is the dominating language.
* English movies have mean of 6.42 of imdb score which is lower it might be due larger dataset that contains both high and low rated movies.
* In conclusion we can observed that smaller language dataset might be biased towards high quality films, while English movies would cover wider spectrum including lower rated films.

****

****

1. **Director Analysis:**

* The average IMDB score of highest rated directors is 7.5 percentile.
* By filtering the data we can identify top directors and which help to understand contributors to higher movie ratings.

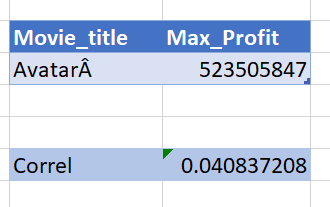
****

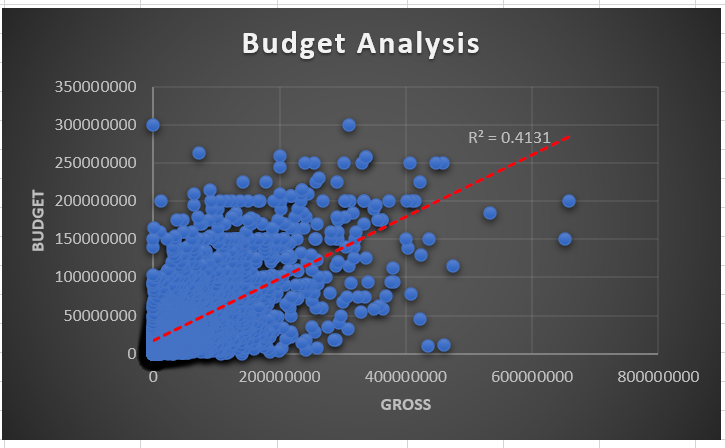


1. **Budget Analysis:**

* Avatar is having the highest progit margin of $523,505,847 which was calculated by gross earnings – budget.
* The correlation coefficient is 0.0408 indicating very weak relationship between budget and gross earnings. This means higher budget does not guarantee higher revenue always.
* We have plot scatter plot by filtering some outliers to observe the values visually and R2 suggest that 42% of variance in gross earnings can be explained by budget.
* In short lower budget movies may achieve exceptional financial success, while some expensive production may sometime under perform.

****

****

****

**Result**

In this project we were able to derive key trends and patterns in factors that influence the success of a movie on IMDB. This insights help movie producers, directors, and investors who want to understand what makes a movie successful to make informed decisions in their future projects. Hence project helps to provide comprehensive understanding of factors that influence success of a movie.

**Drive link of Excel Sheet**

[**https://docs.google.com/spreadsheets/d/1nG343gLUPXV4S1TqFZwNE9gRRVbkvg3j/edit?usp=sharing&ouid=109001208060904860088&rtpof=true&sd=true**](https://docs.google.com/spreadsheets/d/1nG343gLUPXV4S1TqFZwNE9gRRVbkvg3j/edit?usp=sharing&ouid=109001208060904860088&rtpof=true&sd=true)

**Loom video link**

[**https://www.loom.com/share/234580d50bcc4375a2d3c936e650908b?sid=ef1238f6-d24e-4e0a-8233-ca887d15aba2**](https://www.loom.com/share/234580d50bcc4375a2d3c936e650908b?sid=ef1238f6-d24e-4e0a-8233-ca887d15aba2)