**Project Description**

**The dataset under analysis contains information on IMDB movies, presenting an opportunity to uncover factors influencing a movie's success, particularly in terms of high IMDB ratings.**

**Objective: Investigate and understand the impact of various features such as genre, duration, language, director, and budget on IMDB scores.**

**Tech Stack Used**

**Microsoft Excel**

**Data Cleaning:**

**Removed Duplicates**

**Removed rows with missing values**

**Formatted all columns with numbers into “Number” type from “General”**

**Data Analytics Tasks:**

**A. Movie Genre Analysis:**

**Analyse the distribution of movie genres and their impact on the IMDB score.**

**Task:**

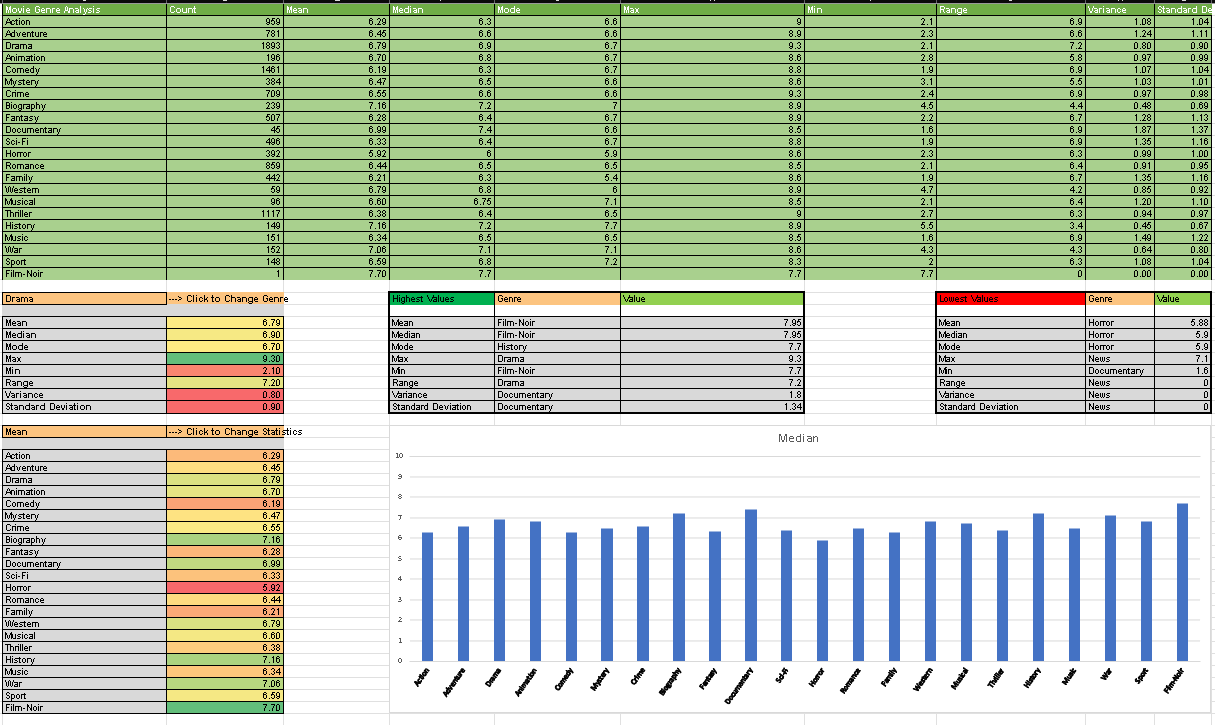
**Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.**

* Used Text to Column to separate the genres on the delimiter “|” from the genre column into separate columns towards the end of the data table.
* Found out the individual genres, made a table for its individual count.
* Found out the descriptive statistics using excel formulas

**Insights:**

* **Film Noir is rated higher on average, but on further inspection only 1 movie comes under the genre, so the data is not representative of the genre itself.**
* **Biography, History, and War genres is rated high on average and has enough data to back up the insight.**
* **Drama has the greatest number of movies.**
* **The movie with the highest rating is a Crime movie.**
* **The movie with the lowest rating is under Documentary and Music genre.**
* **Documentary has the highest standard deviation indicating a huge variation in the Imdb score of movies with the Documentary genre.**

**Results:**

****

**B. Movie Duration Analysis:**

**Analyze the distribution of movie durations and its impact on the IMDB score.**

**Task:**

**Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.**

* Calculated Mean Median and Standard Deviation of the Duration of movies.
* Scatter Plot between IMDB Score and Duration is made.
* A linear trendline is applied in the chart along with its equation.

**Insights:**

* **The average movie lasts 1 hour and 40 mins.**
* **Approximately 68% of movies have durations between 110−23 and 110+23 minutes.**

**Results:**

**A blue and white graph

Description automatically generated**

**C. Language Analysis:**

**Examine the distribution of movies based on their language.**

**Task:**

**Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.**

* Calculated Mean Median Standard deviation and total count of movies for each language.
* Visualised the results of Average IMDB Score for each language in a bar chart.
* Found Top 5 and Bottom 5 Language for each statistic.

**Insights:**

* Most of the movies are released in English.
* Persian movies are rated highest on average, but the sample size is 3 making it too low for a reliable insight.
* Bosnian movies are rated lowest on average but the sample size is relatively small being 14, but enough to get an round about idea of its performance.

**Results:**

**A screenshot of a spreadsheet

Description automatically generated**

**D. Director Analysis:**

**Influence of directors on movie ratings.**

**Task:**

**Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.**

* Calculated Average IMDB Score for each Director
* Calculated the global average (Average IMDB score of all directors) and 95th Percentile IMDB Score.
* Applied a filter to only show the directors in the 95th Percentile.
* Found out the top 5 Directors based on their Average IMDB Score and Made a bar chart comparing their Average Imdb Score to the Global Average.

**Insights:**

* **The top 95th Percentile of Directors have an Average Imdb score above 7.7 for their movies, and the Average being 6.31.**
* **All the top 5 directors have 1 movie released granting them the top spot.**

**Results:**

**A screenshot of a spreadsheet

Description automatically generated**

**E. Budget Analysis:**

**Explore the relationship between movie budgets and their financial success.**

**Task:**

**Analyze the correlation between movie budgets and gross earnings and identify the movies with the highest profit margin.**

* Calculated the profit for each Movie.
* Found out the Correlation coefficient between Budget and Gross revenue.
* Found the top 10 earning movies.

**Insights:**

* **The Correlation** coefficient between Budget and Gross revenue is found out to be 0.0995.
* The Avatar is the most earning movie of all time.

**Results:**

**A screenshot of a spreadsheet

Description automatically generated**

**GOOGLE DRIVE LINK:**

**Video Link** [**https://drive.google.com/file/d/1n8gc3YwsOLKD1QZ\_oimp56lBv5ANkFNw/view?usp=sharing**](https://drive.google.com/file/d/1n8gc3YwsOLKD1QZ_oimp56lBv5ANkFNw/view?usp=sharing)

**Excel Workbook Link**

[**https://docs.google.com/spreadsheets/d/1w8EJ2qF5zbBa7CIdYitn1xSxUPoLhPJL/edit?usp=sharing&ouid=106630094537104676920&rtpof=true&sd=true**](https://docs.google.com/spreadsheets/d/1w8EJ2qF5zbBa7CIdYitn1xSxUPoLhPJL/edit?usp=sharing&ouid=106630094537104676920&rtpof=true&sd=true)