Project Description

The dataset provided is related to IMDB Movies. We have to investigate 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.

Our analysis aims to provide actionable insights that can help stakeholders make informed decisions.

Approach

Data Cleaning: This step involves preprocessing the data to make it suitable for analysis.

I have placed average duration i.e. 107 mins, average gross earning i.e. 48468407.4 , average budget i.e. 39752620.4 at place of missing values in their respective columns.

Data Analysis: explore the data to understand the relationships between different variables. Look at the correlation between movie ratings and other factors like genre, director, budget, etc.

Report and Data Story: Create a report that tells a story with data, include initial problem, findings, and the insights gained.

Tech-Stack Used

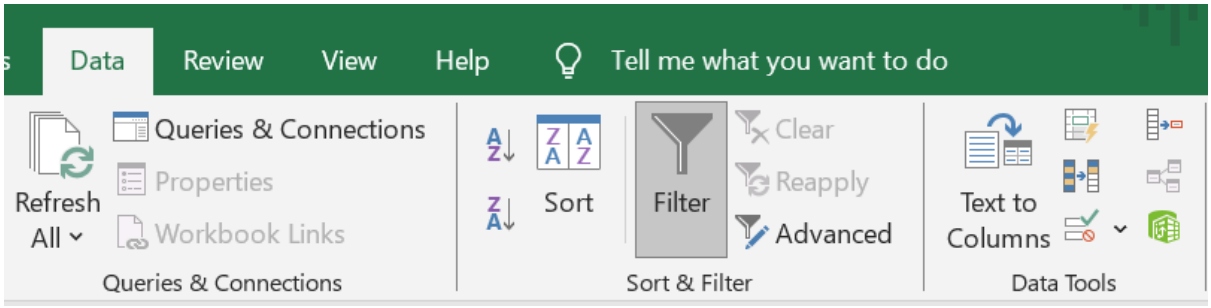
I have used Microsoft Excel 2019.

Insights

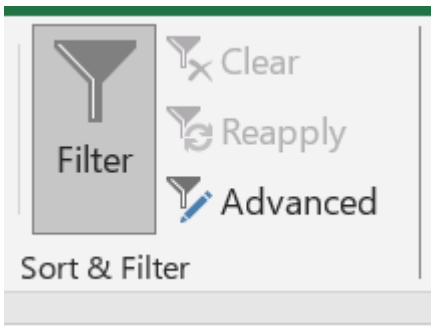
Movie Genre Analysis: Analyze the distribution of movie genres and their impact on the IMDB score.

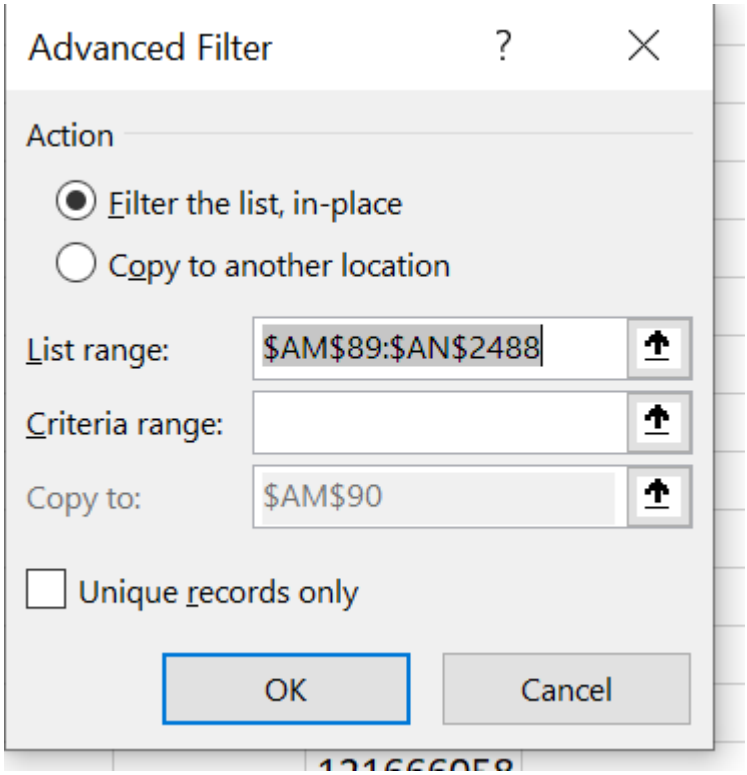
Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.

.Here first I removed multiple genres of a movie into separate columns by using text to column option in Data tab.



Then there will be a dialog box where you can select delimiter to separate out the genre. Then I filtered all unique genres by clicking on advance filter option then in dialog box select the genre range and select unique records only option and select where you want to copy the unique records then click OK.





To calculate the no. of movies

```
=COUNTIF($AC$2:$AJ$5044,AM12)
```

To calculate the mean

```
=AVERAGEIF($AC$2:$AJ$5044,AM12,$Z$2:$Z$5044)
```

To calculate the median

```
{=MEDIAN(IF($AC$2:$AJ$5044=AM12,$Z$2:$Z$5044))}
```

To calculate the mode

```
{=MODE.SNGL(IF($AC$2:$AJ$5044=AM12,$Z$2:$Z$5044))}
```

To calculate the variance

```
{=VAR.P(IF($AC$2:$AJ$5044=AM12,$Z$2:$Z$5044))}
```

To calculate the standard deviation

```
{=STDEV.P(IF($AC$2:$AJ$5044=AM12,$Z$2:$Z$5044))}
```

To calculate the max

```
{=MAX(IF($AC$2:$AJ$5044=AM12,$Z$2:$Z$5044))}
```

To calculate the min

```
{=MIN(IF($AC$2:$AJ$5044=AM12,$Z$2:$Z$5044))}
```

To calculate the range I subtracted min from max.

Movie Duration Analysis: Analyze the distribution of movie durations and its impact on the IMDB score.

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

Average duration	107.200476	Median duration	103	Std dev duration	25.15744

Language Analysis: Examine the distribution of movies based on their language.

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

Get all unique values of language column using the advance filter option as described above.

To calculate the no. of movies

```
=COUNTIF($T$2:$T$5044,AM41)
```

To calculate the mean

```
=AVERAGEIF($T$2:$T$5044,AM41,$Z$2:$Z$5044)
```

To calculate median

```
{=MEDIAN(IF($T$2:$T$5044=AM41,$Z$2:$Z$5044))}
```

To calculate the standard deviation

```
{=STDEV.P(IF($T$2:$T$5044=AM41,$Z$2:$Z$5044))}
```

Director Analysis: Influence of directors on movie ratings. Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.

Find all the unique values in director column using the advanced filter as shown above then find the average rating for each director using following formula:

```
=AVERAGEIF($B$2:$B$5044,AM90,$Z$2:$Z$5044)
```

Then find 90% percentile using following formula:

```
=PERCENTILE.INC(AN90:AN2488,90%)
```

Budget Analysis: Explore the relationship between movie budgets and their financial success.

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

Create a column Profit and calculate the profit by subtracting budget from gross earning using formula:

```
=(I2-W2)
```

To calculate correlation between budget and gross earning use the following formula:

```
=CORREL(I2:I5044,W2:W5044)
```

To calculate highest profit use:

```
=MAX(AK2:AK5044)
```

Max Profit	523505847	Movie Name	AvatarÂ

Result

We found that Comedy was most common genre followed by Thriller and Action.

English is the most common language for movies to be made in. There are 4704 movies made in English.

The correlation between movie budget and gross earning of movie is 0.101.