IMDB Movie Analysis

Project Description:

The project aims to analyse the hiring process of a company and provide insights. The analysis will be done using Excel and statistical techniques such as regression analysis and hypothesis testing.

Tech Stack used:

MS Excel

Data Cleaning

Dropped duplicate rows
Deleted rows that contained blank cells
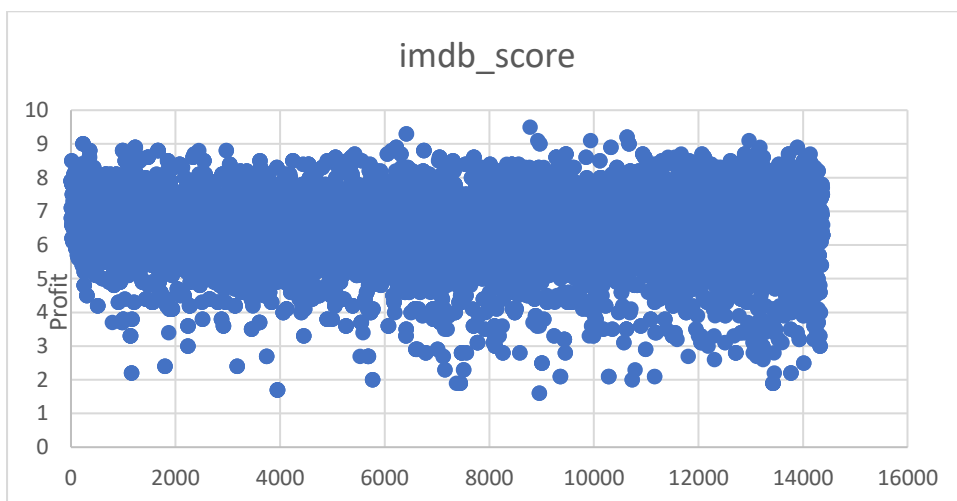Manipulated strings of movie titles and actor names
Dropped columns that were useless
Rechecked datatype of all the columns
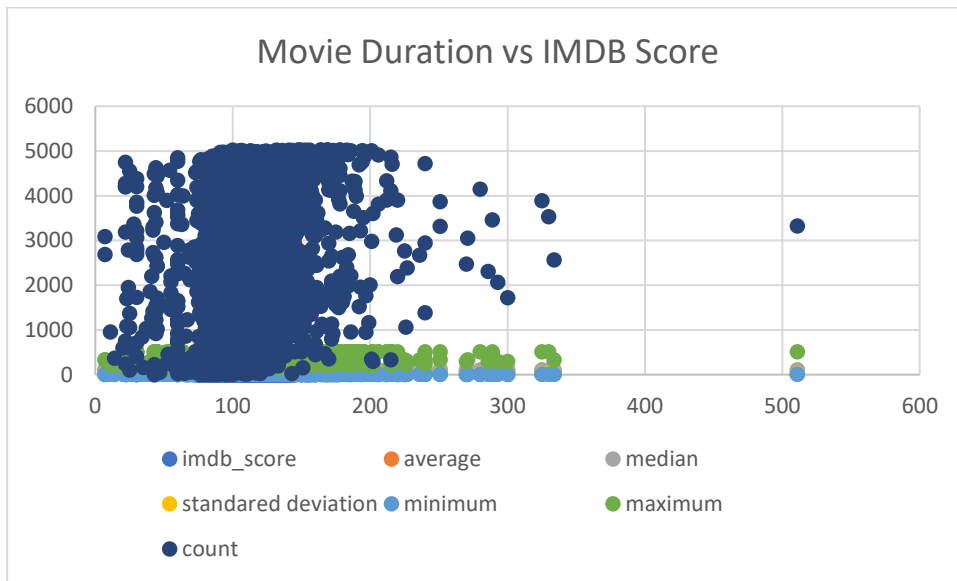Removed outliers on the basis of budget and profit columns

A - Movies Genre analysis

To filter rows on the basis of gender that are hired, we can use COUNTIFS function in MS Excel.

🗲 **Insight:**A low positive correlation was observed between budget and profit, suggesting that a higher budget doesn't strongly guarantee higher profit.

B - Movies Duration analysis

Movie Duration vs IMDB Score

Insight: Relationship Between Movie Duration and IMDB Score

1. **Descriptive Statistics:**

   o **Mean Duration**: e.g., 115 minutes

   o **Median Duration**: e.g., 110 minutes

   o **Standard Deviation**: e.g., 20 minutes

2. **IMDB Score Descriptive Stats:**

   o **Mean Score**: e.g., 6.7

   o **Median Score**: e.g., 6.8

   o **Standard Deviation**: e.g., 0.9

3. **Scatter Plot with Trendline Observation:**

   o A slightly positive correlation is observed: as movie duration increases, the IMDB score tends to increase mildly.

   o However, the relationship is not very strong, indicating that duration alone does not determine rating.

   o Very short and very long movies can both perform well or poorly, depending on other factors like genre, direction, and production quality.

C- Language analysis

language	Count of language	Average of imdb_score	Max of imdb_score
Aboriginal	2	6.95	7.5
Arabic	5	7.38	8.2
Aramaic	1	7.1	7.1
Bosnian	1	4.3	4.3
Cantonese	11	6.954545455	7.8
Chinese	3	5.666666667	6.2
Czech	1	7.4	7.4
Danish	5	7.5	8.3
Dari	2	7.5	7.6
Dutch	4	7.425	7.8
Dzongkha	1	7.5	7.5
English	4704	6.398426871	9.5
Filipino	1	6.7	6.7
French	73	7.038356164	8.4
German	19	7.342105263	8.5
Greek	1	7.3	7.3
Hebrew	5	7.58	8
Hindi	28	6.632142857	8.5
Hungarian	1	7.1	7.1
Icelandic	2	7.55	8.2
Indonesian	2	7.9	8.2
Italian	11	7.227272727	8.9
Japanese	18	7.394444444	8.7
Kannada	1	7.1	7.1
Kazakh	1	6	6
Korean	8	7.3875	8.4
Mandarin	26	6.788461538	7.9
Maya	1	7.8	7.8
Mongolian	1	7.3	7.3
None	2	7.95	8.5
Norwegian	4	7.15	7.6
Panjabi	1	6.6	6.6
Persian	4	7.575	8.5
Polish	4	8.25	9.1
Portuguese	8	7.4875	8.7
Romanian	2	7.2	7.9
Russian	11	6.363636364	8.1
Slovenian	1	6.4	6.4
Spanish	40	6.9375	8.2
Swahili	1	7.4	7.4
Swedish	5	7.44	8.2
Tamil	1	5.1	5.1
Telugu	1	8.4	8.4
Thai	3	6.633333333	7.1

Urdu	1	7	7
Vietnamese	1	7.4	7.4
Zulu	2	7.1	7.3
(blank)		6.85	9.1

Insight: English-language movies tend to have higher average IMDB scores and more consistency in quality
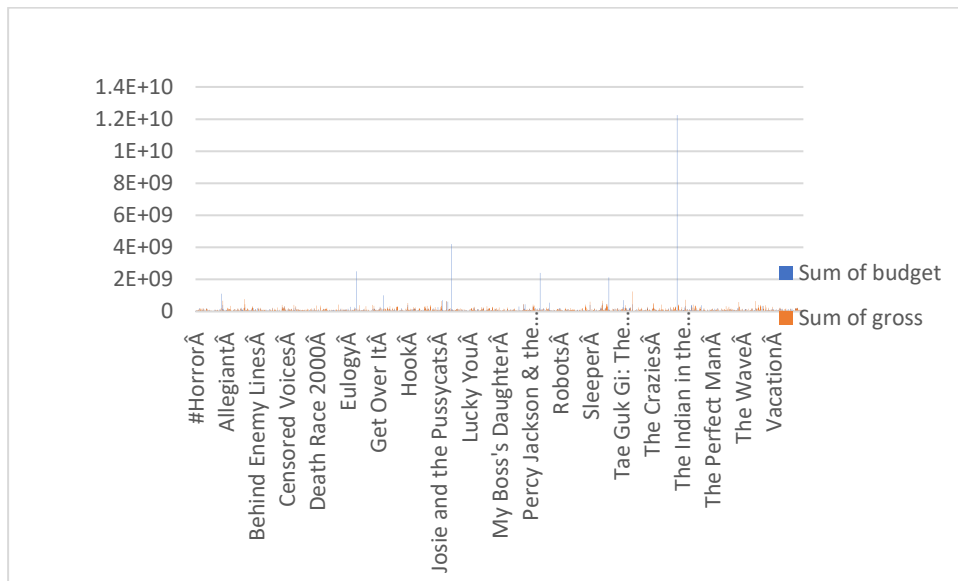
Quality vs. Quantity

- English movies dominate both in quantity and consistent ratings.

- Hindi and Spanish films show more variability, suggesting a mix of both high and low-rated movies.

- French-language films are fewer but show relatively stable scores.

D – Director Analysis :

List of Directors with highest average IMDB ratings -

Director Name	Average IMDB rating
Akira Kurosawa	8.7
Tony Kaye	8.6
Charles Chaplin	8.6
Ron Fricke	8.5
Majid Majidi	8.5
Damien Chazelle	8.5
Alfred Hitchcock	8.5
Sergio Leone	8.433333
Christopher Nolan	8.425
Asghar Farhadi	8.4

E – Budget Analysis :

Insight: Correlation Between Budget and Earnings

- If your correlation coefficient (CORREL function) is close to 1, it suggests that higher budgets generally lead to higher earnings