

IMDB Movie Analysis

Trinity-5

Description

In this project IMDB movie analysis has been performed.

This project will give insight and answers to the below Questions

- Data Cleaning
- Movie Genre Analysis
- Movie Duration Analysis
- Language Analysis
- Director Analysis
- Budget Analysis

Tech-Stack Used



MICROSOFT EXCEL

- This tool is used to draw insights & create graphical representation of the results and to understand the result set better.



Jupyter Notebook

- This tool is used to clean data and remove unnecessary columns and filling null values using Pandas library .

The Tableau Public logo, featuring the word 'tableau' in blue and 'public' in a lighter blue, with a small red plus sign between them.

Tableau Public

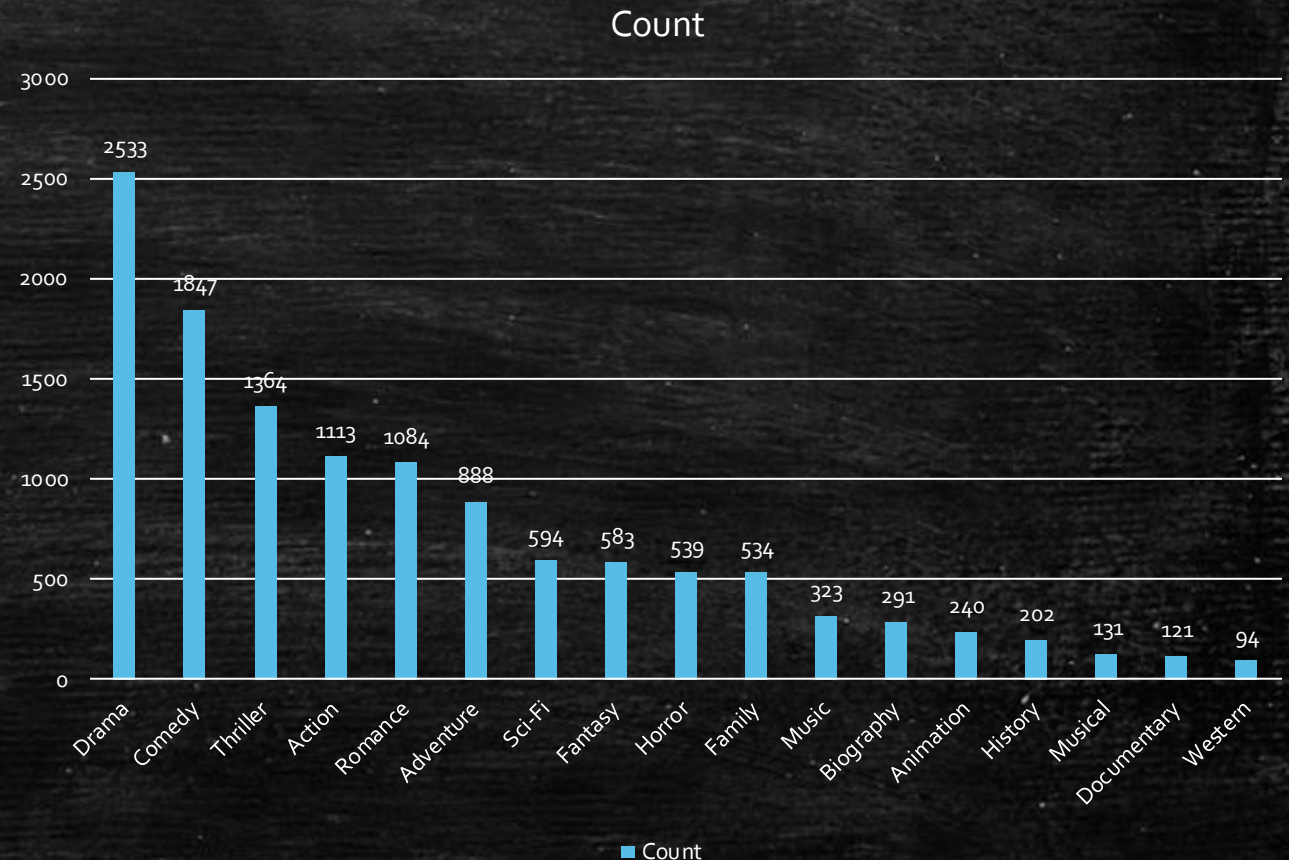
- This tool is used to create graphical representation of the results and to understand the result set better.

Data Cleaning

- In the dataset there are so many null values.
- First I removed all the unnecessary columns from the data set using pandas drop method.
- There are so many null so deleting these null values may cause disturbance in the data, so I replaced them with **Mean** and **Mode** accordingly.
- Imported the file into excel for further analysis.

Movie Genre Analysis

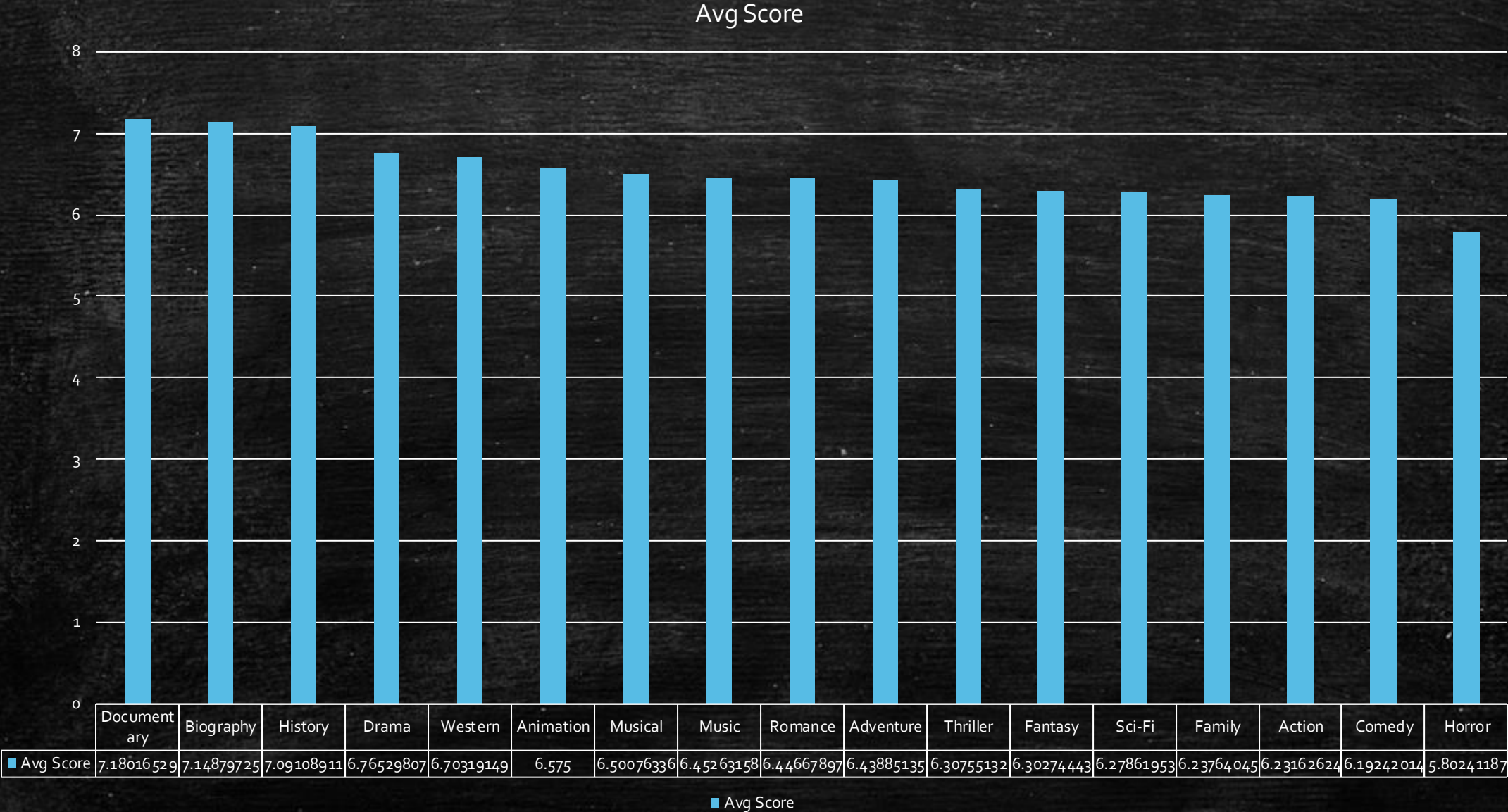
- **Task:** Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.
- Nearly half of the movies are part of **Drama** genre.



descriptive statistics

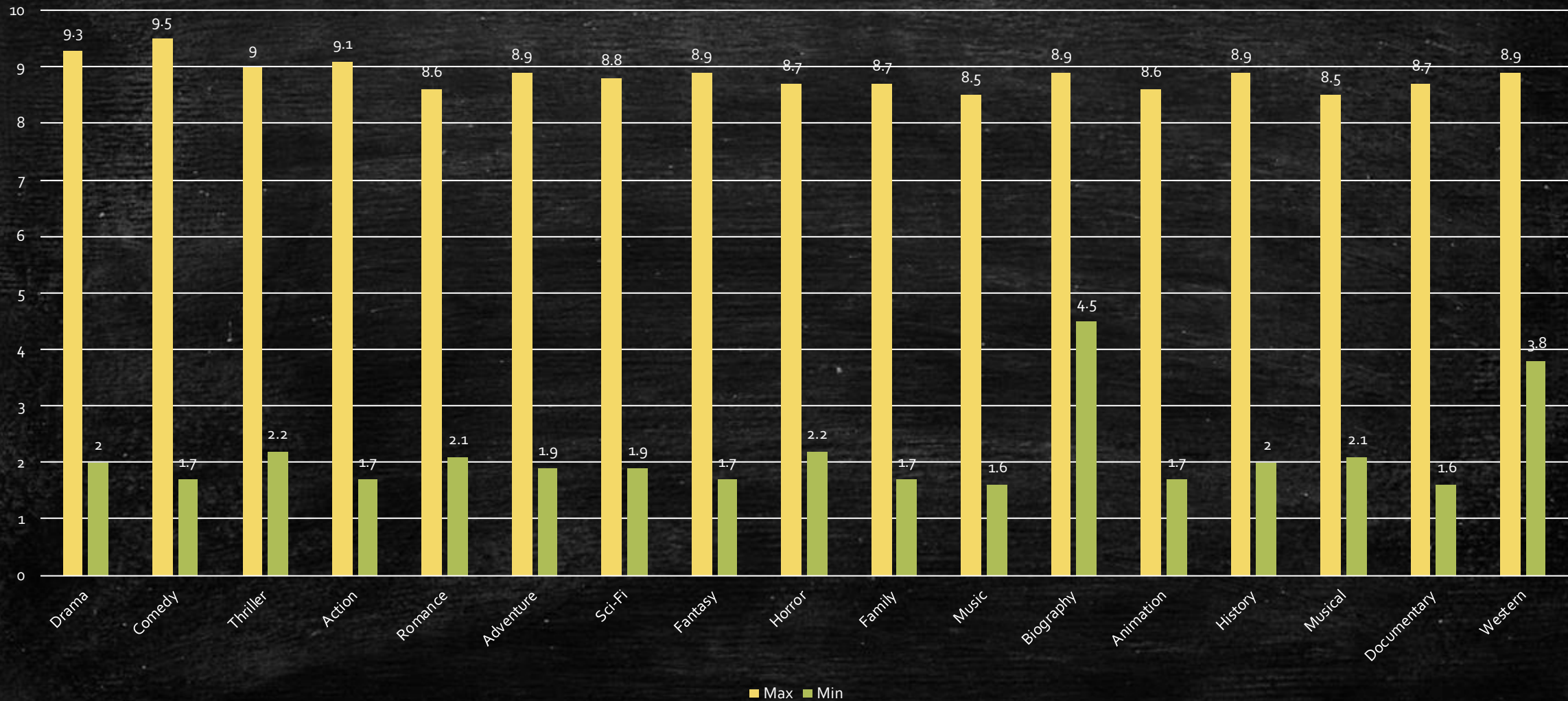
Genre	COUNT	Max	Min	MODE	AVG	MEDIAN	STDEV
Drama	2533	9.3	2	6.7	6.765298066	6.9	0.954005379
Comedy	1847	9.5	1.7	6.7	6.192420141	6.3	1.091405875
Thriller	1364	9	2.2	6.4	6.30755132	6.4	1.056222946
Action	1113	9.1	1.7	6.6	6.231626235	6.3	1.119226495
Romance	1084	8.6	2.1	6.5	6.446678967	6.5	0.998326413
Adventure	888	8.9	1.9	6.7	6.438851351	6.6	1.137388387
Sci-Fi	594	8.8	1.9	6.7	6.278619529	6.4	1.217369007
Fantasy	583	8.9	1.7	6.7	6.302744425	6.4	1.167070825
Horror	539	8.7	2.2	6.2	5.802411874	5.9	1.120248632
Family	534	8.7	1.7	6.7	6.237640449	6.4	1.210037311
Music	212	8.5	1.6	6.5	6.406132075	6.6	1.183139907
Biography	291	8.9	4.5	7	7.148797251	7.2	0.724704646
Animation	240	8.6	1.7	6.7	6.575	6.7	0.954005379
History	202	8.9	2	7.5	7.091089109	7.2	0.954005379
Musical	131	8.5	2.1	7	6.500763359	6.7	0.954005379
Documentary	121	8.7	1.6	7.5	7.180165289	7.4	0.954005379
Western	94	8.9	3.8	6.5	6.703191489	6.8	0.954005379

Average score by Genre



Min & Max by genre

Chart Title





Genre impact on IMDB Score

- Most common genre is Drama.
- comedy has highest score, Documentary has lowest score.
- Biography, History, Documentary has good mean, median and Mode.
- Biography has less standard deviation.

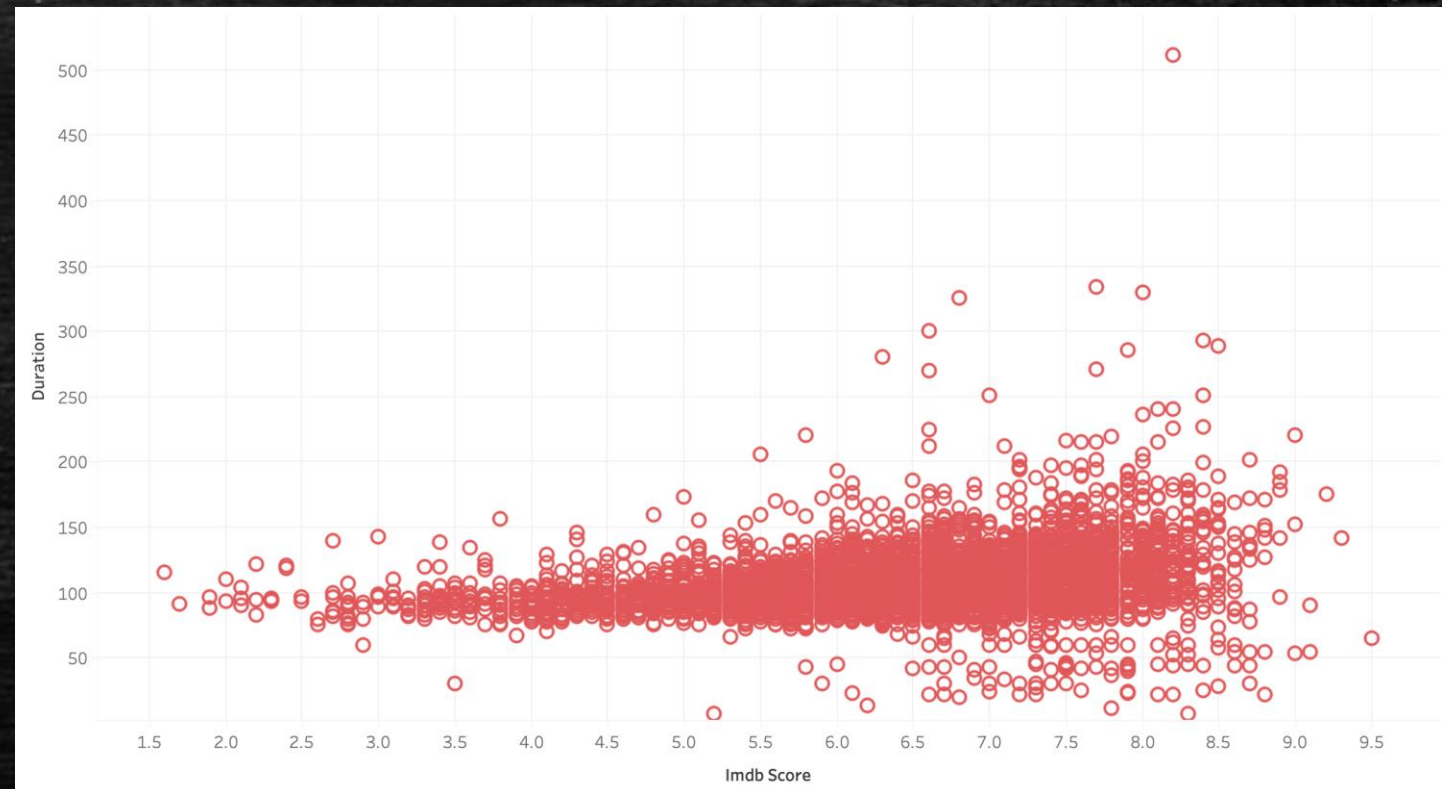
Movie Duration Analysis

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

Most of the movies made a cluster around average duration as move score increases the cluster is expanding in both sides.

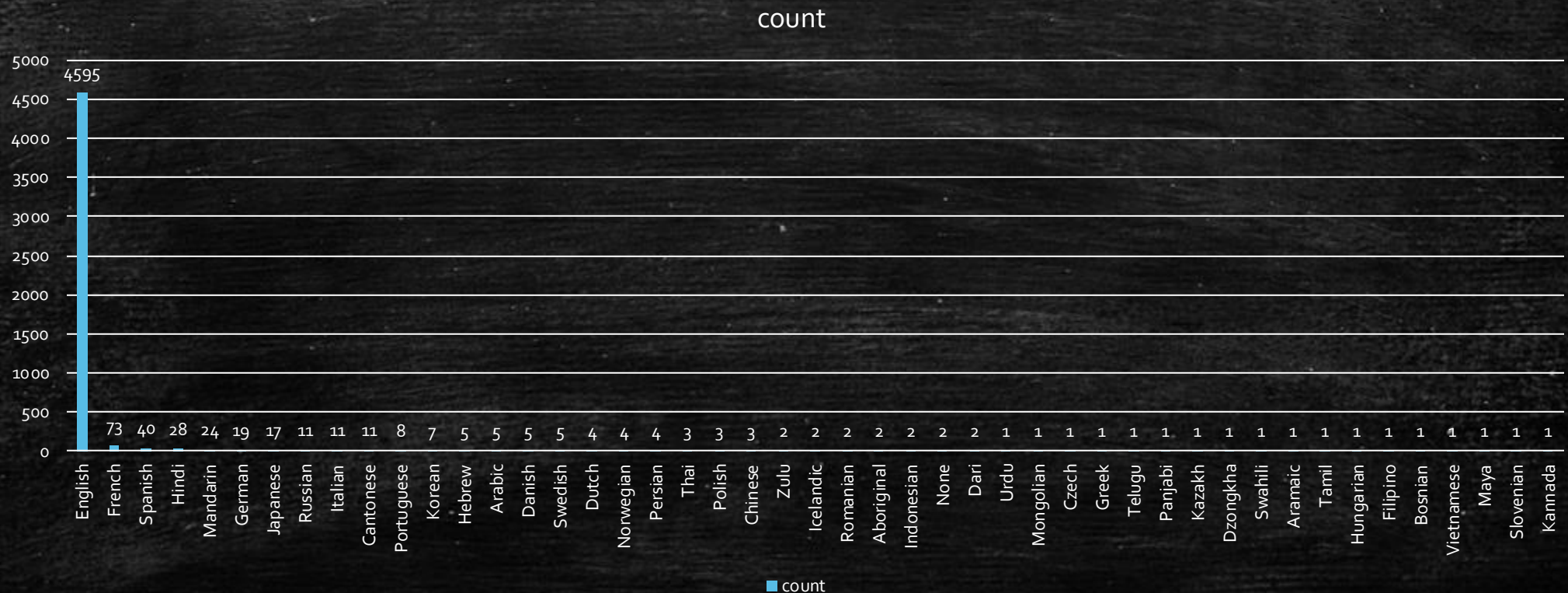
- Many high duration and low duration got good ratings.

	average	median	std
Duration	107.0899633	103	25.24490026



Language Analysis

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



Language	count	avg score	mode	Median	std
English	4595	6.394820457	6.7	6.5	1.125748116
French	73	7.038356164	7.2	7.2	0.726985812
Spanish	40	6.9375	7.2	7.15	0.855056603
Hindi	28	6.632142857	7.8	6.95	1.398955582
Mandarin	24	6.7875	7.9	7.05	1.036848276
German	19	7.342105263	7.8	7.6	0.954123093
Japanese	17	7.347058824	6.1	7.5	1.000073527
Russian	11	6.363636364	5.3	6.5	1.383671007
Italian	11	7.227272727	-	7.3	1.244259546
Cantonese	11	6.954545455	7.3	7.2	0.704788814
Portuguese	8	7.4875	-	7.7	0.883883476
Korean	7	7.442857143	7.7	7.7	0.87532307
Hebrew	5	7.58	-	7.6	0.334664011
Arabic	5	7.38	-	7.4	0.884307639
Danish	5	7.5	8.1	8.1	1.077032961
Swedish	5	7.44	-	7.6	0.756967635
Dutch	4	7.425	7.8	7.45	0.434932945
Norwegian	4	7.15	7.6	7.3	0.574456265
Persian	4	7.575	-	7.95	1.203813385
Thai	3	6.633333333	-	6.6	0.450924975
Polish	3	7.966666667	7.4	7.4	0.981495458
Chinese	3	5.666666667	-	5.7	0.550757055
Zulu	2	7.1	-	7.1	0.282842712
Icelandic	2	7.55	-	7.55	0.919238816
Romanian	2	7.2	-	7.2	0.989949494

Language	Count	Avg Score	Mode	Median.	Std
Aboriginal	2	6.95	-	95.6	0.777817459
Indonesian	2	7.9	-	7.9	0.424264069
None	2	7.95	-	7.95	0.777817459
Dari	2	7.5	-	7.5	0.141421356
Urdu	1	7	-	7	-
Mongolian	1	7.3	-	7.3	-
Czech	1	7.4	-	7.4	-
Greek	1	7.3	-	7.3	-
Telugu	1	8.4	-	8.4	-
Panjabi	1	6.6	-	6.6	-
Kazakh	1	6	-	6	-
Dzongkha	1	7.5	-	7.5	-
Swahili	1	7.4	-	7.4	-
Aramaic	1	7.1	-	7.1	-
Tamil	1	5.1	-	5.1	-
Hungarian	1	7.1	-	7.1	-
Filipino	1	6.7	-	6.7	-
Bosnian	1	4.3	-	4.3	-
Vietnamese	1	7.4	-	7.4	-
Maya	1	7.8	-	7.8	-
Slovenian	1	6.4	-	6.4	-
Kannada	1	7.1	-	7.1	-

Budget Analysis

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

Correlation=**0.219946752**

Director	gross	Movie title	budget	Imdb score	profit
James Cameron	760505847	Avatar	237000000	7.9	523505847
Colin Trevorrow	652177271	Jurassic World	150000000	7	502177271
James Cameron	658672302	Titanic	200000000	7.7	458672302
George Lucas	460935665	Star Wars: Episode IV - A New Hope	11000000	8.7	449935665
Steven Spielberg	434949459	E.T. the Extra-Terrestrial	10500000	7.9	424449459

Director Analysis

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

Percentile	Score	count
25th Percentile	5.6	1789
50th percentile	6.4	1168
75th percentile	7	590
90th percentile	7.5	220
95th percentile	7.785833333	120
99th percentile	8.3	22
99.5th percentile	8.400375	12

Director name	Average of IMDB score
John Blanchard	9.5
Cary Bell	8.7
Mitchell Altieri	8.7
Sadyk Sher-Niyaz	8.7
Charles Chaplin	8.6
Mike Mayhall	8.6
Raja Menon	8.5
Ron Fricke	8.5
Damien Chazelle	8.5
Majid Majidi	8.5
Sergio Leone	8.475
Christopher Nolan	8.425

Top 12 Directors with
Avg IMDB Score with
99.5 percentile

Links

- EXCEL – [Click me](#)
- Please go through the excel sheet where analysis has been done.
- Thank you.

Result

- I have no knowledge about Excel prior to this now I know little bit of Advanced excel.
- I also used Tableau for few visualizations
- I am waiting for upcoming projects I hope they are going to me more complex so I can test my limits.
- **THANK YOU.**