Click this Link to Get the Excel Sheet

IMDB MOVIE ANALYSIS

Project Description: The objective of this project is to conduct comprehensive analysis of the IMDB MOVIE dataset. This analysis focus on serval key aspect of the dataset like Genre, language, directors, Duration, Story and Budget etc. our main goal is to find a relation between all these aspect and how movie rating and success define on these aspect so that in future other movie directors can use this analysis for movie making and make a good movies and get their good rating and success.

Approach: To complete the task I follow the structured approach

Data preparation : I began by downloading the dataset providing for this project. This dataset contain information about Genre, language, directors, duration, Story, Budget, total gross, imbd rating. First I clean the data then I start working on this project.

Data Analysis: I use excel to performed various analysis on the dataset to answer the question outlined in the project. This involve manipulation the data to extract relevant information, calculating statistics and visualization relationship between variables.

Final report : Once I complete all the task like find relationship between all the variable then I try to convert the insight into visual representation using graph, chart, report etc. and create the detail report of my finding.

Tech-Stack Used : For this project I use excel for extract, transform, clean, statistics approach, data visualization and summarizing the results. And use google drive to save and share the final report.

Task A:

Movie Genre Analysis: Analyze 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.**

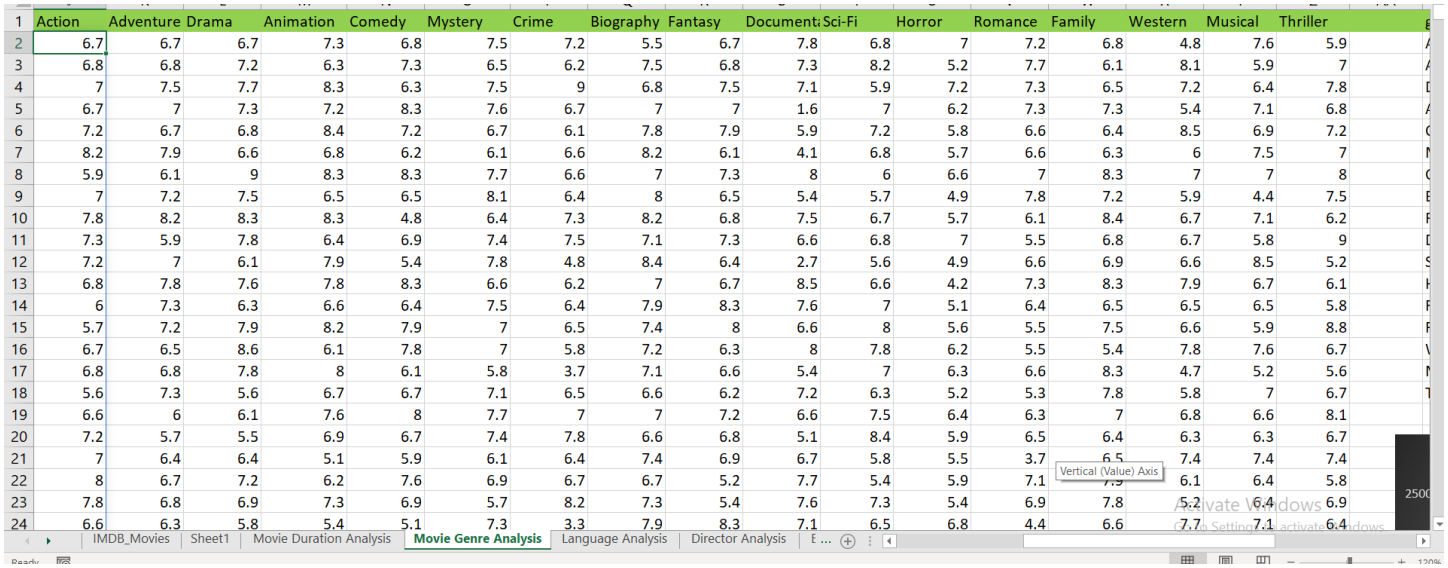
Solution: To analyze the distribution the movie genres and their impact on the IMBD Score

First we bring the Genre, IMBD in the different Sheet because only these two attribute we need for the analyze Movie Genre.

- We use the delimeter to split the column because genre column has multiple genre data in the single row .
- Once we split the column we use remove duplicate to get the total number of genre in the list and after remove the duplicate we find the there are 22 different genre in the list.
- Now we find the IMBD score for each movie using excel formula.

```
=FILTER($I$2:$I$3816,ISNUMBER(SEARCH(J1,$H$2:$H$3816)))
```

by using this formula we generate IMBD Score for each Genre and movies.

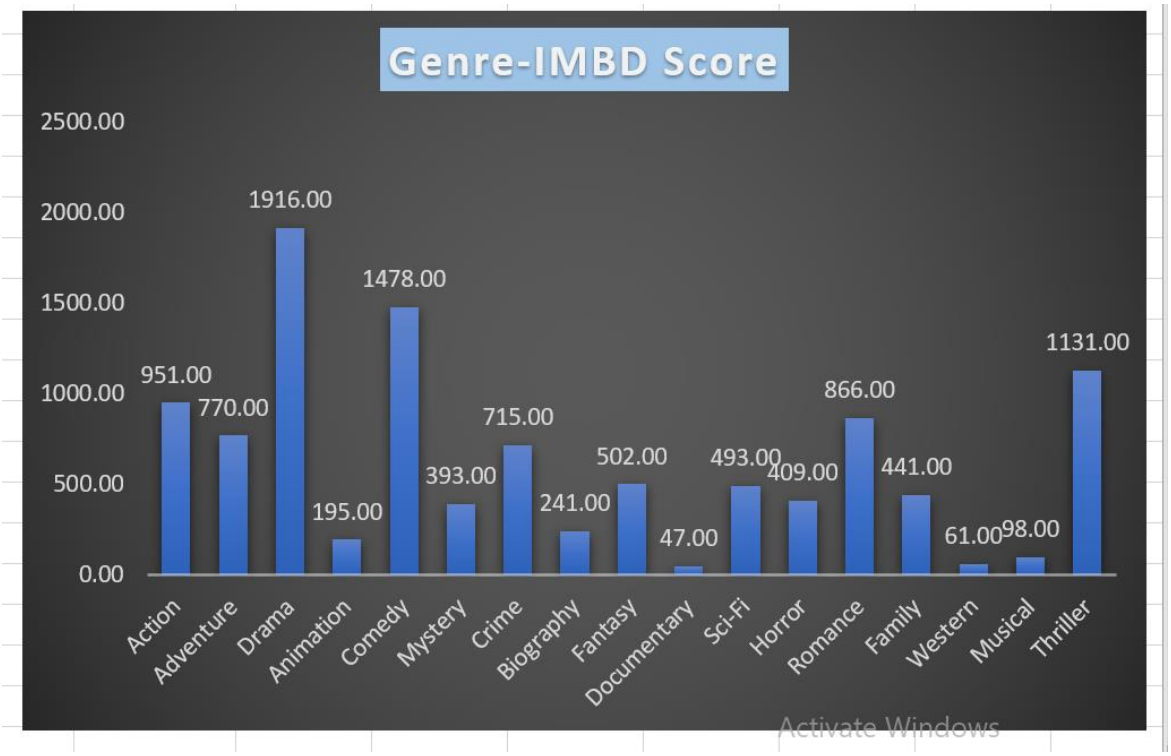


Action	Adventure	Drama	Animation	Comedy	Mystery	Crime	Biography	Fantasy	Documentary	Sci-Fi	Horror	Romance	Family	Western	Musical	Thriller
6.7	6.7	6.7	7.3	6.8	7.5	7.2	5.5	6.7	7.8	6.8	7	7.2	6.8	4.8	7.6	5.9
6.8	6.8	7.2	6.3	7.3	6.5	6.2	7.5	6.8	7.3	8.2	5.2	7.7	6.1	8.1	5.9	7
7	7.5	7.7	8.3	6.3	7.5	9	6.8	7.5	7.1	5.9	7.2	7.3	6.5	7.2	6.4	7.8
6.7	7	7.3	7.2	8.3	7.6	6.7	7	7	1.6	7	6.2	7.3	7.3	5.4	7.1	6.8
7.2	6.7	6.8	8.4	7.2	6.7	6.1	7.8	7.9	5.9	7.2	5.8	6.6	6.4	8.5	6.9	7.2
8.2	7.9	6.6	6.8	6.2	6.1	6.6	8.2	6.1	4.1	6.8	5.7	6.6	6.3	6	7.5	7
5.9	6.1	9	8.3	8.3	7.7	6.6	7	7.3	8	6	6.6	7	8.3	7	7	8
7	7.2	7.5	6.5	6.5	8.1	6.4	8	6.5	5.4	5.7	4.9	7.8	7.2	5.9	4.4	7.5
7.8	8.2	8.3	8.3	4.8	6.4	7.3	8.2	6.8	7.5	6.7	5.7	6.1	8.4	6.7	7.1	6.2
7.3	5.9	7.8	6.4	6.9	7.4	7.5	7.1	7.3	6.6	6.8	7	5.5	6.8	6.7	5.8	9
7.2	7	6.1	7.9	5.4	7.8	4.8	8.4	6.4	2.7	5.6	4.9	6.6	6.9	6.6	8.5	5.2
6.8	7.8	7.6	7.8	8.3	6.6	6.2	7	6.7	8.5	6.6	4.2	7.3	8.3	7.9	6.7	6.1
6	7.3	6.3	6.6	6.4	7.5	6.4	7.9	8.3	7.6	7	5.1	6.4	6.5	6.5	6.5	5.8
5.7	7.2	7.9	8.2	7.9	7	6.5	7.4	8	6.6	8	5.6	5.5	7.5	6.6	5.9	8.8
6.7	6.5	8.6	6.1	7.8	7	5.8	7.2	6.3	8	7.8	6.2	5.5	5.4	7.8	7.6	6.7
6.8	6.8	7.8	8	6.1	5.8	3.7	7.1	6.6	5.4	7	6.3	6.6	8.3	4.7	5.2	5.6
5.6	7.3	5.6	6.7	6.7	7.1	6.5	6.6	6.2	7.2	6.3	5.2	5.3	7.8	5.8	7	6.7
6.6	6	6.1	7.6	8	7.7	7	7	7.2	6.6	7.5	6.4	6.3	7	6.8	6.6	8.1
7.2	5.7	5.5	6.9	6.7	7.4	7.8	6.6	6.8	5.1	8.4	5.9	6.5	6.4	6.3	6.3	6.7
7	6.4	6.4	5.1	5.9	6.1	6.4	7.4	6.9	6.7	5.8	5.5	3.7	6.5	7.4	7.4	7.4
8	6.7	7.2	6.2	7.6	6.9	6.7	6.7	5.2	7.7	5.4	5.9	7.1	7.5	6.1	6.4	5.8
7.8	6.8	6.9	7.3	6.9	5.7	8.2	7.3	5.4	7.6	7.3	5.4	6.9	7.8	5.2	6.4	6.9
6.6	6.3	5.8	5.4	5.1	7.3	3.3	7.9	8.3	7.1	6.5	6.8	4.4	6.6	7.7	7.1	6.4

- After all this we calculate the descriptive statistics for each Genre

genres	count	Mean	Median	Mode	Variance	STV.DEV
Action	951.00	6.26	6.30	6.10	1.12	1.06
Adventure	770.00	6.43	6.50	6.70	1.25	1.12
Drama	1916.00	6.79	6.90	6.70	0.81	0.90
Animation	195.00	6.70	6.80	6.70	0.98	0.99
Comedy	1478.00	6.18	6.30	6.70	1.10	1.05
Mystery	393.00	6.45	6.50	6.60	1.12	1.06
Crime	715.00	6.53	6.60	6.60	1.00	1.00
Biography	241.00	7.16	7.20	7.00	0.48	0.69
Fantasy	502.00	6.27	6.35	6.70	1.28	1.13
Documentar	47.00	7.00	7.40	7.60	1.84	1.36
Sci-Fi	493.00	6.31	6.40	6.70	1.35	1.16
Horror	409.00	5.89	5.90	5.90	1.08	1.04
Romance	866.00	6.44	6.50	6.50	0.92	0.96
Family	441.00	6.21	6.30	5.40	1.35	1.16
Western	61.00	6.69	6.70	6.00	1.10	1.05
Musical	98.00	6.59	6.75	7.10	1.19	1.09
Thriller	1131.00	6.36	6.40	6.50	1.00	1.00

- After all this we showed the relation between Movie and Genre count



- Through this analysis we found Drama genre has a most number of movie and Comedy, Action, Thriller are second best after the drama.

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

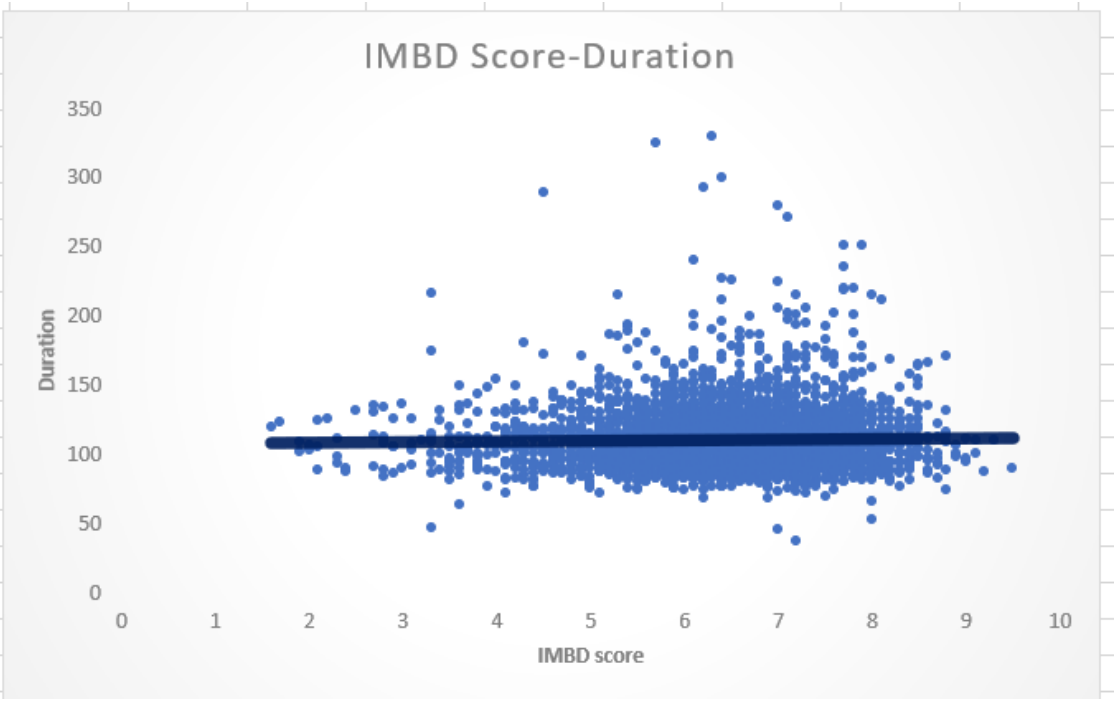
- **Task: Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score**.

Solution: To find the distribution of movie durations and its impact on the IMBD score first we bring the Duration and IMBD attribute on the different sheet.

- First we calculate the Descriptive Statistics for movie duration

Mean	110.047
Median	106
Mode	101
Stv.Dev	22.6432
variance	512.715

- Now we find the relation between IMBD Score and Movie Duration.



- Through the analysis we find out average movie duration time is 110 min.
- Most of the movie who has average IMBD Score there movie duration is lies between 90 min to 150 min.

C. Language Analysis: **Situation: 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.**

Solution: To Determine the most common language used in movies and analyze their impact on the IMBD first we bring the IMBD Score and Language attribute the spread sheet

Because we have lot of unnecessary column in the main data sheet and we have to work on only this two attribute.

- After that insert pivot table to analyze the relation between these two .

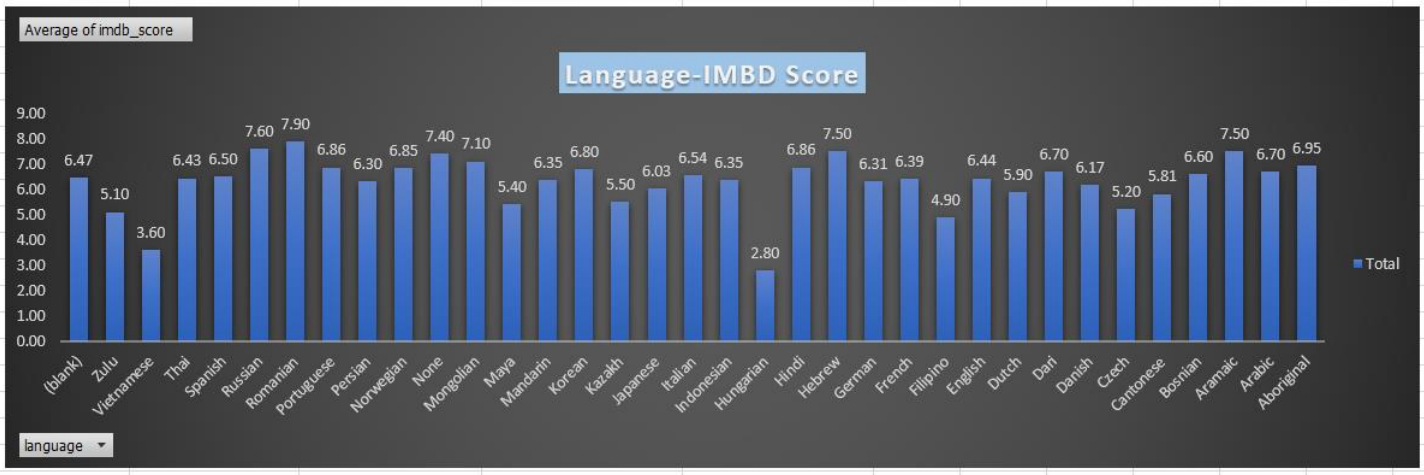
- Then we find out English is the most common language used in movies. Its number is 3657 it is whey more higher than the other language which mean people mostly watching English movie.

Language	Count
English	3657
French	34
Spanish	23
Mandarin	15
Japanese	10
German	10
Italian	7
Cantonese	7

- After that we find the average for each language on the basis of IMBD Score.

Row Labels	Average of imdb_score
(blank)	6.47
Zulu	5.10
Vietnamese	3.60
Thai	6.43
Spanish	6.50
Russian	7.60
Romanian	7.90
Portuguese	6.86
Persian	6.30
Norwegian	6.85
None	7.40
Mongolian	7.10
Maya	5.40
Mandarin	6.35
Korean	6.80
Kazakh	5.50
Japanese	6.03
Italian	6.54
Indonesian	6.35
Hungarian	2.80
Hindi	6.86
Hebrew	7.50
German	6.31
French	6.39
Filipino	4.90
English	6.44
Dutch	5.90
Dari	6.70
Danish	6.17
Czech	5.20

- Then we find the relation between Language and IMBD Score using column charts.



- After looking at this graph we can say there not any huge impact of language on the IMBD rating.

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.**

Solution: To find the relation between Director and IMBD score and analyze their contribution to the success first we fetch the director and IMBD score attribute to the different sheet.

- Now, we insert the pivot table on these attributes.

- After insert an pivot table we put Director in row attributes and IMBD score in values Then we convert it into Average that's how we find Average of IMBD score for all Director.

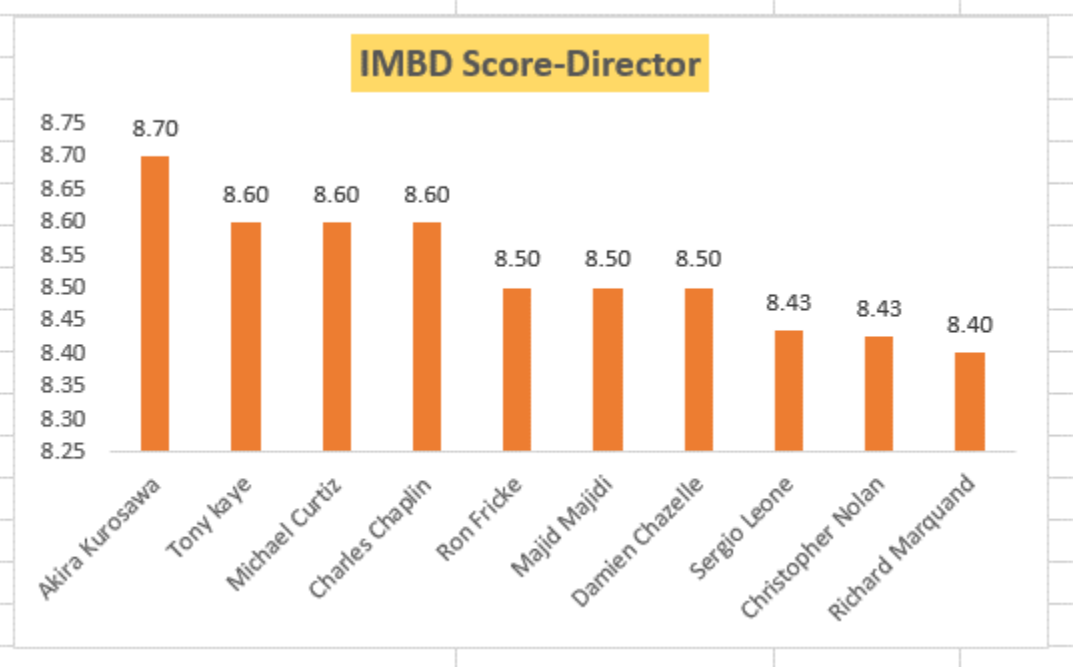
Row Labels	Average of imdb_score
Akira Kurosawa	8.7
Tony Kaye	8.6
Michael Curtiz	8.6
Charles Chaplin	8.6
Ron Fricke	8.5
Majid Majidi	8.5
Damien Chazelle	8.5
Sergio Leone	8.433333333
Christopher Nolan	8.425
Richard Marquand	8.4
Asghar Farhadi	8.4
Alfred Hitchcock	8.35
Lee Unkrich	8.3
Lenny Abrahamson	8.3
Fritz Lang	8.3
Billy Wilder	8.3
Pete Docter	8.233333333
Hayao Miyazaki	8.225
Quentin Tarantino	8.2
Joshua Oppenheimer	8.2
Juan JosÃ© Campanella	8.2
George Roy Hill	8.2
Elia Kazan	8.2
Milos Forman	8.133333333
Tim Miller	8.1
William Wyler	8.1
Sharon Greytak	8.1
Terry George	8.1
Michael Wadleigh	8.1
Je-kyu Kang	8.1
David Sington	8.1
Vincent Paronnaud	8

- These are the average of all Directors.

- Now, we Separate the top director of all time from this list.

Director Name	IMBD Score	Percentile
Akira Kurosawa	8.70	99.40%
Tony kaye	8.60	99.20%
Michael Curtiz	8.60	99.20%
Charles Chaplin	8.60	99.20%
Ron Fricke	8.50	98.70%
Majid Majidi	8.50	98.70%
Damien Chazelle	8.50	98.70%
Sergio Leone	8.43	98.70%
Christopher Nolan	8.43	98.70%
Richard Marquand	8.40	98.30%

- These are top Director with IMBD score and Percentile.
- Now, we show the relation between IMBD score and Director using column charts.



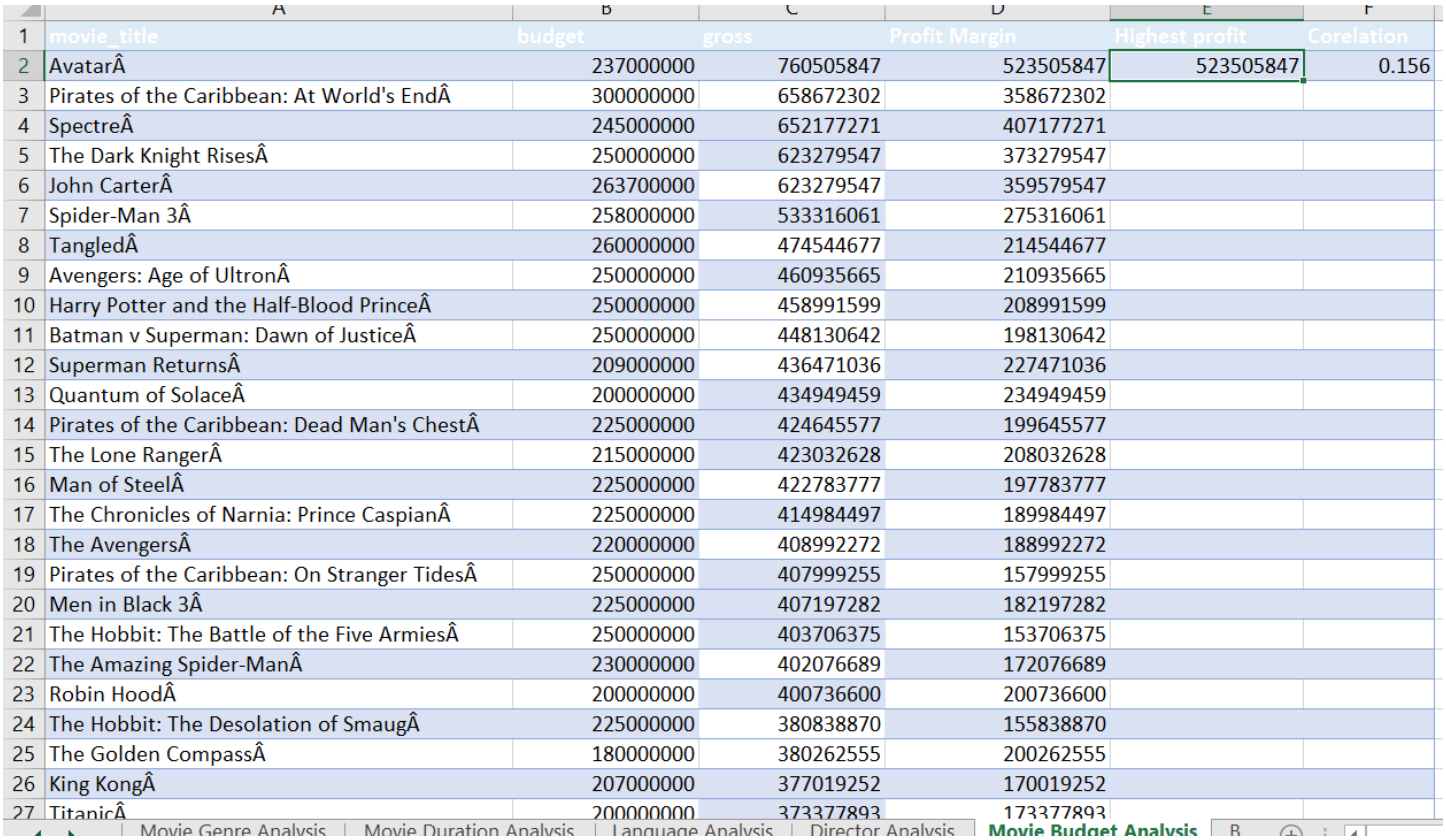
Notice: There is very mild impact on the basis of director.

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**.

Solution: To Analyze the correlation between movie budgets and gross earning first we bring the Movie Title , Budget, Gross earnings to the different sheet.

- Now, to find the Total profit we differentiate between Gross earning with Budgets. We get Total profit by doing this.



	A	B	C	D	E	F
1	movie_title	budget	gross	Profit Margin	Highest profit	Corelation
2	AvatarÂ	237000000	760505847	523505847	523505847	0.156
3	Pirates of the Caribbean: At World's EndÂ	300000000	658672302	358672302		
4	SpectreÂ	245000000	652177271	407177271		
5	The Dark Knight RisesÂ	250000000	623279547	373279547		
6	John CarterÂ	263700000	623279547	359579547		
7	Spider-Man 3Â	258000000	533316061	275316061		
8	TangledÂ	260000000	474544677	214544677		
9	Avengers: Age of UltronÂ	250000000	460935665	210935665		
10	Harry Potter and the Half-Blood PrinceÂ	250000000	458991599	208991599		
11	Batman v Superman: Dawn of JusticeÂ	250000000	448130642	198130642		
12	Superman ReturnsÂ	209000000	436471036	227471036		
13	Quantum of SolaceÂ	200000000	434949459	234949459		
14	Pirates of the Caribbean: Dead Man's ChestÂ	225000000	424645577	199645577		
15	The Lone RangerÂ	215000000	423032628	208032628		
16	Man of SteelÂ	225000000	422783777	197783777		
17	The Chronicles of Narnia: Prince CaspianÂ	225000000	414984497	189984497		
18	The AvengersÂ	220000000	408992272	188992272		
19	Pirates of the Caribbean: On Stranger TidesÂ	250000000	407999255	157999255		
20	Men in Black 3Â	225000000	407197282	182197282		
21	The Hobbit: The Battle of the Five ArmiesÂ	250000000	403706375	153706375		
22	The Amazing Spider-ManÂ	230000000	402076689	172076689		
23	Robin HoodÂ	200000000	400736600	200736600		
24	The Hobbit: The Desolation of SmaugÂ	225000000	380838870	155838870		
25	The Golden CompassÂ	180000000	380262555	200262555		
26	King KongÂ	207000000	377019252	170019252		
27	TitanicÂ	200000000	373377893	173377893		

Movie Genre Analysis | Movie Duration Analysis | Language Analysis | Director Analysis | Movie Budget Analysis | B

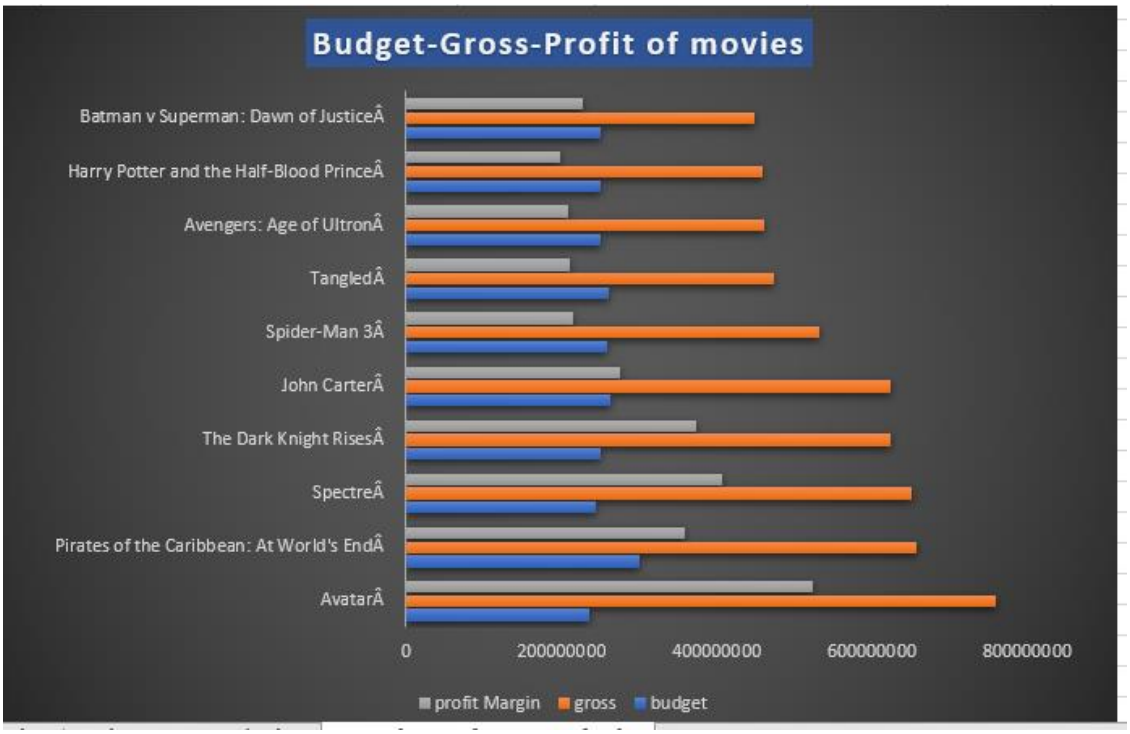
- Then we use Max function to get the highest profit movie from total profit.

```
=MAX(D2:D3319)
```

- Now we do analysis on the basis of Budget , Gross earning, Total profit.

movie_title	budget	gross	profit Margin
AvatarÂ	237000000	760505847	523505487
Pirates of the Caribbean: At World's EndÂ	300000000	658672302	358672302
SpectreÂ	245000000	652177271	407177271
The Dark Knight RisesÂ	250000000	623279547	373279547
John CarterÂ	263700000	623279547	275316061
Spider-Man 3Â	258000000	533316061	214544677
TangledÂ	260000000	474544677	210935665
Avengers: Age of UltronÂ	250000000	460935665	208991599
Harry Potter and the Half-Blood PrinceÂ	250000000	458991599	198130642
Batman v Superman: Dawn of JusticeÂ	250000000	448130642	227471036

- Now, we show the relation with using Bar graph.



- Avatar Movie has highest profit among all the movies
- Correlation between Budget and Gross earning is shown using correl function which is 0.156.

Result:

- Average Movie Duration is 110 min.
- Drama Genre has most number of movies which is 1916.
- English language has most number of movie no language is near to English which is 3657.
- Avatar movie is the most profitable movie and has big budget.
- Director has barely impact on imbd score.