

IMDB Movie Analysis

9.30

9090 Brief:

This project is about the analysis of Movies rated through IMDB. Here as a data analyst, I have to find some metrics regarding the movies and actors according to rating and people and critics likings. The job data set is provided through excel sheet where movie name, actor_1, 2,3 names, number of people voted, facebook likes of actor_1,2,3, language of movie, gross and budget of the movies, imdb overall ratings, number of critics rated, genres and country of origin.

The project is classified into different objectives:

- To find out the top movies with highest profit and scatter plot of gross and budget of movies
- To find out the top 250 movies according to imdb rating and foreign language films among them
- To find out the top 10 directors according to average imdb ratings received
- To find out the top 10 genres of the movies according to the average imdb ratings received
- To find out the top actors among the Brad Pitt, Leonardo di Caprio and Meryl Streep according to average of number of users and critic reviews
- To plot the decade with respect to number of users voted in each year

Approach:

To derive the insights and results mentioned above, dataset is cleaned where unwanted columns are removed and then blanks from the respective usable columns for the insights are removed using the sort function. After sorting, calculation is done using the excel function and then graph is plotted. For questions d,e,f – pivot table is used where average field settings are used.

Details of the excel is given here through this [link](#)

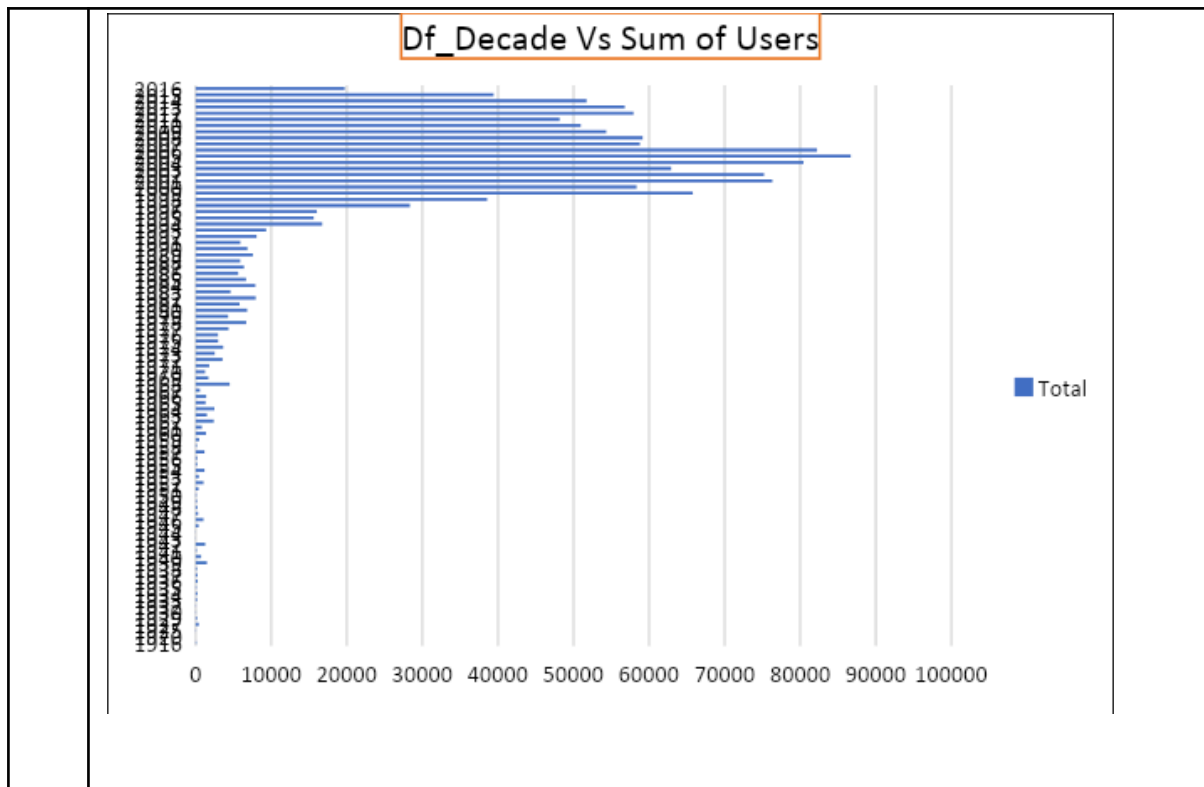
Tech-Stack Used:

Excel is Used to perform the analysis.

Q- A	Here the q-1 is about basic formatting of the data based on the question requirement											
	<table><tr><th>movie_title</th><th>gross</th><th>budget</th><th>profit</th></tr><tr><td>Avatar</td><td>7.61E+08</td><td>2.37E+08</td><td>523505847</td></tr></table>				movie_title	gross	budget	profit	Avatar	7.61E+08	2.37E+08	523505847
movie_title	gross	budget	profit									
Avatar	7.61E+08	2.37E+08	523505847									

Q- B	Jurassic World	6.52E+08	1.5E+08	502177271																						
	Titanic	6.59E+08	2E+08	458672302																						
	Star Wars: Episode IV - A New Hope	4.61E+08	11000000	449935665																						
	E.T. the Extra-Terrestrial	4.35E+08	10500000	424449459																						
	The Avengers	6.23E+08	2.2E+08	403279547																						
	The Avengers	6.23E+08	2.2E+08	403279547																						
	The Lion King	4.23E+08	45000000	377783777																						
	Star Wars: Episode I - The Phantom Menace	4.75E+08	1.15E+08	359544677																						
	The Dark Knight	5.33E+08	1.85E+08	348316061																						
	Note: Top 10 movies by Profit is shown																									
Q-B (2 nd Part)	Here Graph is plotted based on the budget vs gross which is shown in excel attached																									
Q-C	- Here movie, number of users voted, imdb score and language is extracted and then number of users whether greater than 25000 is checked using Or function. All the true results are then ranked from 1 -250 and placed . - From the Ranked results foreign language column is created using the IF function to check whether they are English or foreign language																									
Q- D	Director's name and imdb score is extracted. Then pivot table is created to extract the average IMDB score and then Top 10 is formed. <table><tr><th>Top_10 Directors</th><th>Average of imdb_score</th></tr><tr><td>Charles Chaplin</td><td>8.6</td></tr><tr><td>Tony Kaye</td><td>8.6</td></tr><tr><td>Alfred Hitchcock</td><td>8.5</td></tr><tr><td>Damien Chazelle</td><td>8.5</td></tr><tr><td>Majid Majidi</td><td>8.5</td></tr><tr><td>Ron Fricke</td><td>8.5</td></tr><tr><td>Sergio Leone</td><td>8.433333333</td></tr><tr><td>Asghar Farhadi</td><td>8.4</td></tr><tr><td>Christopher Nolan</td><td>8.4</td></tr><tr><td>Marius A. Markevicius</td><td>8.4</td></tr></table>				Top_10 Directors	Average of imdb_score	Charles Chaplin	8.6	Tony Kaye	8.6	Alfred Hitchcock	8.5	Damien Chazelle	8.5	Majid Majidi	8.5	Ron Fricke	8.5	Sergio Leone	8.433333333	Asghar Farhadi	8.4	Christopher Nolan	8.4	Marius A. Markevicius	8.4
Top_10 Directors	Average of imdb_score																									
Charles Chaplin	8.6																									
Tony Kaye	8.6																									
Alfred Hitchcock	8.5																									
Damien Chazelle	8.5																									
Majid Majidi	8.5																									
Ron Fricke	8.5																									
Sergio Leone	8.433333333																									
Asghar Farhadi	8.4																									
Christopher Nolan	8.4																									
Marius A. Markevicius	8.4																									

	<table> <tr> <td>Richard Marquand</td><td>8.4</td></tr> <tr> <td>S.S. Rajamouli</td><td>8.4</td></tr> </table>	Richard Marquand	8.4	S.S. Rajamouli	8.4																		
Richard Marquand	8.4																						
S.S. Rajamouli	8.4																						
Q- E	Genre's name and imdb score is extracted. Then pivot table is created to extract the average IMDB score and then Top 10 is formed. <table> <tr> <th>Top_10 Genres</th><th>Average of imdb_score</th></tr> <tr> <td>Adventure Animation Drama Family Musical</td><td>8.5</td></tr> <tr> <td>Crime Drama Fantasy Mystery</td><td>8.5</td></tr> <tr> <td>Action Adventure Drama Fantasy War</td><td>8.4</td></tr> <tr> <td>Adventure Animation Fantasy</td><td>8.4</td></tr> <tr> <td>Adventure Drama Thriller War</td><td>8.4</td></tr> <tr> <td>Adventure Animation Comedy Drama Family Fantasy</td><td>8.3</td></tr> <tr> <td>Biography Drama History Music</td><td>8.3</td></tr> <tr> <td>Documentary Drama Sport</td><td>8.3</td></tr> <tr> <td>Documentary War</td><td>8.3</td></tr> <tr> <td>Adventure Drama War</td><td>8.25</td></tr> </table>	Top_10 Genres	Average of imdb_score	Adventure Animation Drama Family Musical	8.5	Crime Drama Fantasy Mystery	8.5	Action Adventure Drama Fantasy War	8.4	Adventure Animation Fantasy	8.4	Adventure Drama Thriller War	8.4	Adventure Animation Comedy Drama Family Fantasy	8.3	Biography Drama History Music	8.3	Documentary Drama Sport	8.3	Documentary War	8.3	Adventure Drama War	8.25
Top_10 Genres	Average of imdb_score																						
Adventure Animation Drama Family Musical	8.5																						
Crime Drama Fantasy Mystery	8.5																						
Action Adventure Drama Fantasy War	8.4																						
Adventure Animation Fantasy	8.4																						
Adventure Drama Thriller War	8.4																						
Adventure Animation Comedy Drama Family Fantasy	8.3																						
Biography Drama History Music	8.3																						
Documentary Drama Sport	8.3																						
Documentary War	8.3																						
Adventure Drama War	8.25																						
Q- F	Actor_1 name, number of users, number of critics for review and title year is extracted in r. Then pivot table is created to extract the average number of users, average number of critics in respect of three actors – Brad Pitt, Leonardo, Dicaprio and Merl Streep and title year as df_decade in new column is created. Then using the average number of users, average number of critics Top 10 is formed. From the df_decade vs number if users bar graph is plotted. <table> <tr> <th>Actor_1 Name</th><th>Average of num_user_for_reviews</th></tr> <tr> <td>Leonardo DiCaprio</td><td>914.4761905</td></tr> <tr> <td>Brad Pitt</td><td>702.4444444</td></tr> <tr> <td>Meryl Streep</td><td>257.3076923</td></tr> </table> <table> <tr> <th>Actor_1 Name</th><th>Average of num_critic_for_reviews</th></tr> <tr> <td>Leonardo DiCaprio</td><td>330.1904762</td></tr> <tr> <td>Brad Pitt</td><td>231.9444444</td></tr> <tr> <td>Meryl Streep</td><td>163.1538462</td></tr> </table>	Actor_1 Name	Average of num_user_for_reviews	Leonardo DiCaprio	914.4761905	Brad Pitt	702.4444444	Meryl Streep	257.3076923	Actor_1 Name	Average of num_critic_for_reviews	Leonardo DiCaprio	330.1904762	Brad Pitt	231.9444444	Meryl Streep	163.1538462						
Actor_1 Name	Average of num_user_for_reviews																						
Leonardo DiCaprio	914.4761905																						
Brad Pitt	702.4444444																						
Meryl Streep	257.3076923																						
Actor_1 Name	Average of num_critic_for_reviews																						
Leonardo DiCaprio	330.1904762																						
Brad Pitt	231.9444444																						
Meryl Streep	163.1538462																						



Insights:

From this project I have derived certain insights that helps in better understanding:

- From the analysis, I have obtained top 10 movies with highest profit, The profit is obtained by subtracting the gross and budgeted columns. Then for all the movies according to profits are plotted in scatter plots with outliers.
- Top 250 movies according to imdb ratings are extracted using the if function where true options are taken where number of users greater than 25000. Then they are ranked from 1-250. After that a new column (Foreign language) according to language where English and Foreign language are segregated using the if function.
- Top 10 directors according average imdb ratings is extracted using the pivot function and sort function.
- Top 10 genres of the films according average imdb ratings is extracted using the pivot function and sort function.
- Top actor among Brad Pitt, Leonardo, Dicaprio and Merl Streep according to the average critic reviews and number of users voted is obtained using the pivot the sort function.
- The number of users is plotted against the decade where the title year is considered in a bar plot.

Result:

Through this project, it is found the different metrics of the movie market analysis and it effects the investment in a movie.

Through this project, I have learnt how database is used to generate business usable insights. The Business analyst views have developed and I can leverage that to use in other live projects