Project Description:

The dataset has 28 columns, 5044 rows of different IMDB movies.

Define a problem, clean the data as necessary remove null and duplicate values. Explore the data set and derive insights. Use root cause analysis of why approach. Do the required analysis and create reports for
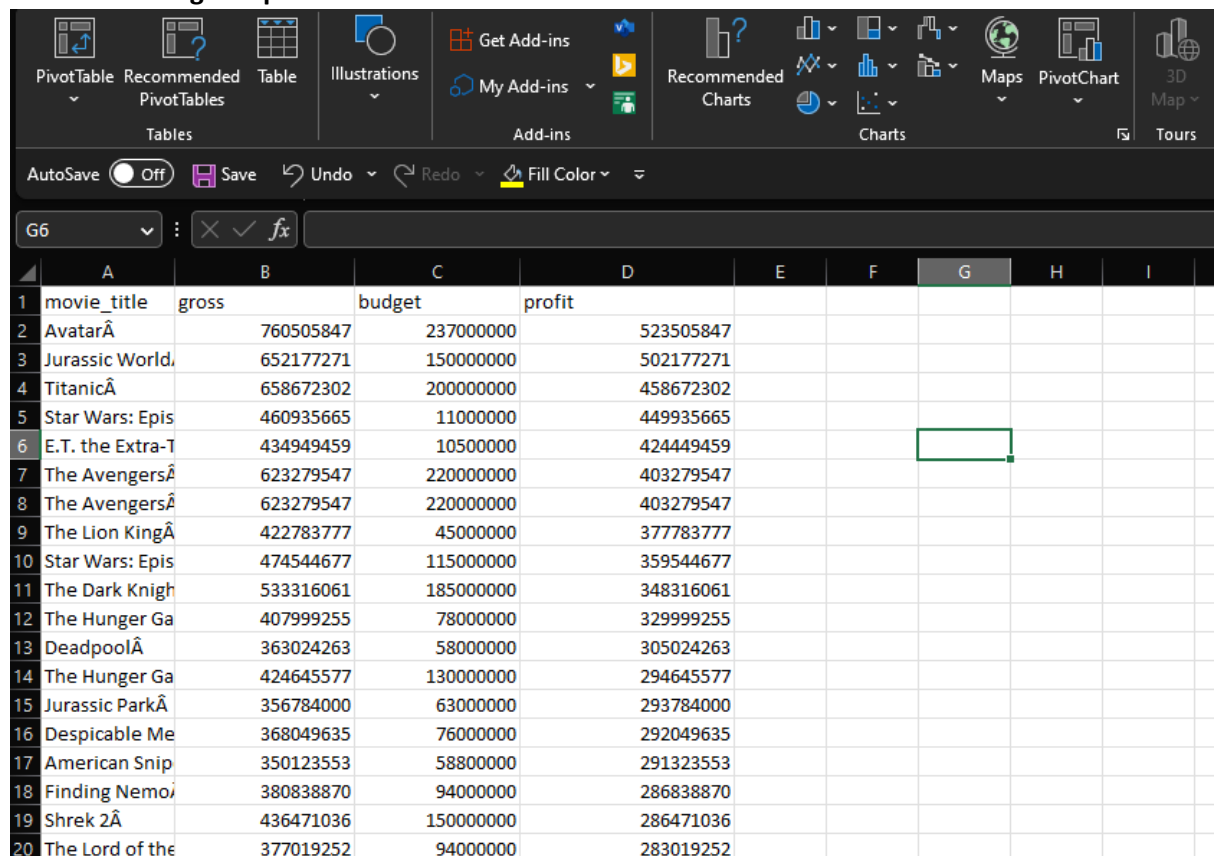
Approach:

First, I will do the data cleaning. Then I will perform analysis on the data using Microsoft excel in built function. Use other functions to give answer to asked questions. Create charts to for easy and meaningful description. Create report in ppt format and submit the project.
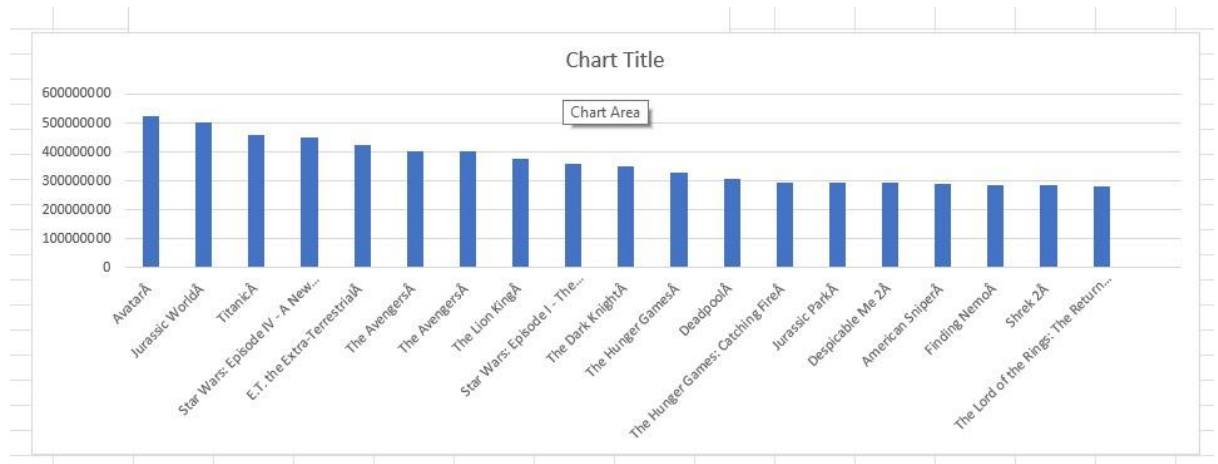
Tech Stack used: Microsoft Excel

Insights:

1) **Movies with highest profit**

Chart Title

2) Top 250 movies with highest mean IMDB rating

| Movie Name | Voted users>25000 | Language | Rating | RANK |
|---|---|---|---|---|
| T+A2:D841he Shawshank RedemptionÂ | 1689764 | English | 9.3 | 1 |
| The GodfatherÂ | 1155770 | English | 9.2 | 2 |
| The Dark KnightÂ | 1676169 | English | 9 | 3 |
| The Godfather: Part IIÂ | 790926 | English | 9 | 4 |
| FargoÂ | 170055 | English | 9 | 5 |
| The Lord of the Rings: The Return of the Ki | 1215718 | English | 8.9 | 6 |
| Schindler's ListÂ | 865020 | English | 8.9 | 7 |
| Pulp FictionÂ | 1324680 | English | 8.9 | 8 |
| The Good, the Bad and the UglyÂ | 503509 | Italian | 8.9 | 9 |
| 12 Angry MenÂ | 447785 | English | 8.9 | 10 |
| InceptionÂ | 1468200 | English | 8.8 | 11 |
| The Lord of the Rings: The Fellowship of th | 1238746 | English | 8.8 | 12 |
| DaredevilÂ | 213483 | English | 8.8 | 13 |
| Fight ClubÂ | 1347461 | English | 8.8 | 14 |
| Forrest GumpÂ | 1251222 | English | 8.8 | 15 |
| It's Always Sunny in PhiladelphiaÂ | 133415 | English | 8.8 | 16 |
| Star Wars: Episode V - The Empire Strikes E | 837759 | English | 8.8 | 17 |
| The Lord of the Rings: The Two TowersÂ | 1100446 | English | 8.7 | 18 |
| The MatrixÂ | 1217752 | English | 8.7 | 19 |

Top 250 movies with highest mean IMDB rating other than English

| Title | Votes | Language | IMDb | Rank |
|---|---|---|---|---|
| The Good, the Bad and the Ugly | 503509 | Italian | 8.9 | 9 |
| City of GodÂ | 533200 | Portuguese | 8.7 | 24 |
| Seven SamuraiÂ | 229012 | Japanese | 8.7 | 25 |
| Spirited AwayÂ | 417971 | Japanese | 8.6 | 32 |
| AirliftÂ | 30977 | Hindi | 8.5 | 57 |
| The Lives of OthersÂ | 259379 | German | 8.5 | 59 |
| Children of HeavenÂ | 27882 | Persian | 8.5 | 61 |
| AmÃ©lieÂ | 534262 | French | 8.4 | 65 |
| Baahubali: The BeginningÂ | 62756 | Telugu | 8.4 | 66 |
| Princess MononokeÂ | 221552 | Japanese | 8.4 | 69 |
| Das BootÂ | 168203 | German | 8.4 | 74 |
| Rang De BasantiÂ | 70233 | Hindi | 8.4 | 77 |
| OldboyÂ | 356181 | Korean | 8.4 | 82 |
| A SeparationÂ | 151812 | Persian | 8.4 | 85 |
| MetropolisÂ | 111841 | German | 8.3 | 99 |
| DownfallÂ | 248354 | German | 8.3 | 101 |
| The HuntÂ | 170155 | Danish | 8.3 | 109 |
| Howl's Moving CastleÂ | 214091 | Japanese | 8.2 | 129 |

3) Top 10 Best Directors with highest mean:

| Actors | Mean 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 |
| Damien Chazelle | 8.5 |
| Majid Majidi | 8.5 |
| Raja Menon | 8.5 |
| Ron Fricke | 8.5 |

4) Popular genres:

| Row Labels | Count of genres |
|---|---|
| Drama | 236 |
| Comedy | 209 |
| Comedy|Drama | 191 |
| Comedy|Drama|Romance | 187 |
| Comedy|Romance | 158 |
| Drama|Romance | 152 |
| Crime|Drama|Thriller | 101 |
| Horror | 71 |
| Action|Crime|Drama|Thriller | 68 |
| Action|Crime|Thriller | 65 |
| Grand Total | 1438 |

Drama is the most popular genres.

5) Charts:

| Row Labels | Sum of num_voted_users |
|---|---|
| 1916-1925 | 15572 |
| 1926-1935 | 221320 |
| 1936-1945 | 1701477 |
| 1946-1955 | 1201344 |
| 1956-1965 | 3866611 |
| 1966-1975 | 7906554 |
| 1976-1985 | 19394407 |
| 1986-1995 | 39999052 |
| 1996-2005 | 137679063 |
| 2006-2016 | 206821367 |
| Grand Total | 418806767 |



Num of users/ num of critics

| Row Labels | Average of num_critic_for_reviews | Average of num_user_for_reviews |
|---|---|---|
| Brad Pitt | 224.5 | 574.64 |
| Leonardo DiCaprio | 80 | 292.00 |
| Meryl Streep | 210.09 | 424.73 |
| (blank) | | |
| Grand Total | 212.8461538 | 500.3461538 |
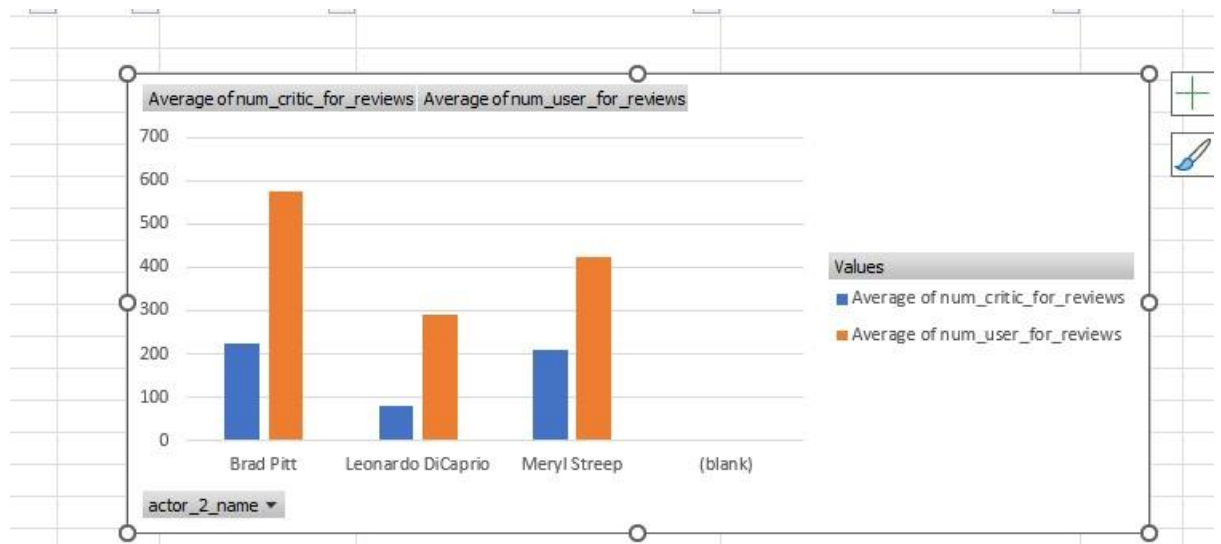
Brad Pitt is the actor with the highest num critic and num users for review.

Result: This project has helped to understand basis steps involved in exploratory data analysis. EDA is require to make data free from error and making it fit for making models. The whole process also helped me to get real hands experience on data set and how to clean, manipulate data, visualize data from insights.