



IMDB Movie Analysis

-Ranapratap Ghosh

-Project Description-

In these days the analytical insights helps us to identify many hidden aspect within the bigger datasets.

This projects helps us to understand the insights about the Cinematic world and their effects on the people's mind and pocket.

In this project we will identify Top grossing and profitable movies, best directors, best lead actors as well as which actor is the critic and fan favorite. This analysis will help us understand why some movies and some actor are better than others.

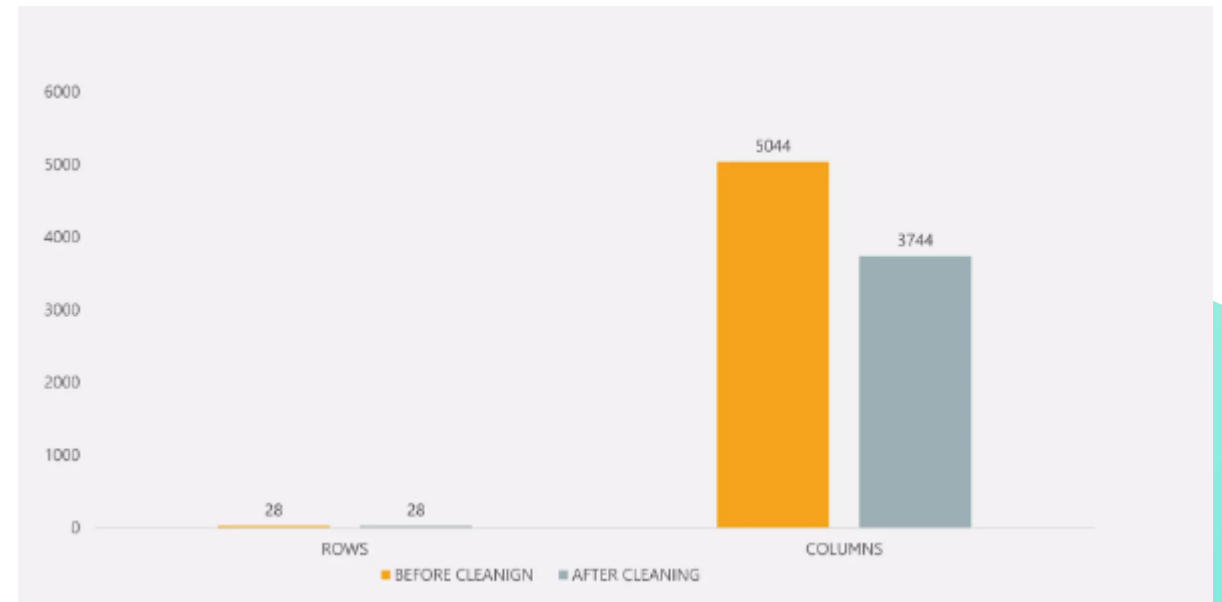
Approach & Tech Stack:-

- Cleaning the data and get rid of the existing null and blank values in the dataset.
- Brainstorming for finding possible answers from the dataset.
- Representing the result values in charts for better understanding.
- Using Excel, PowerBi and PowerPoint to complete the analysis.

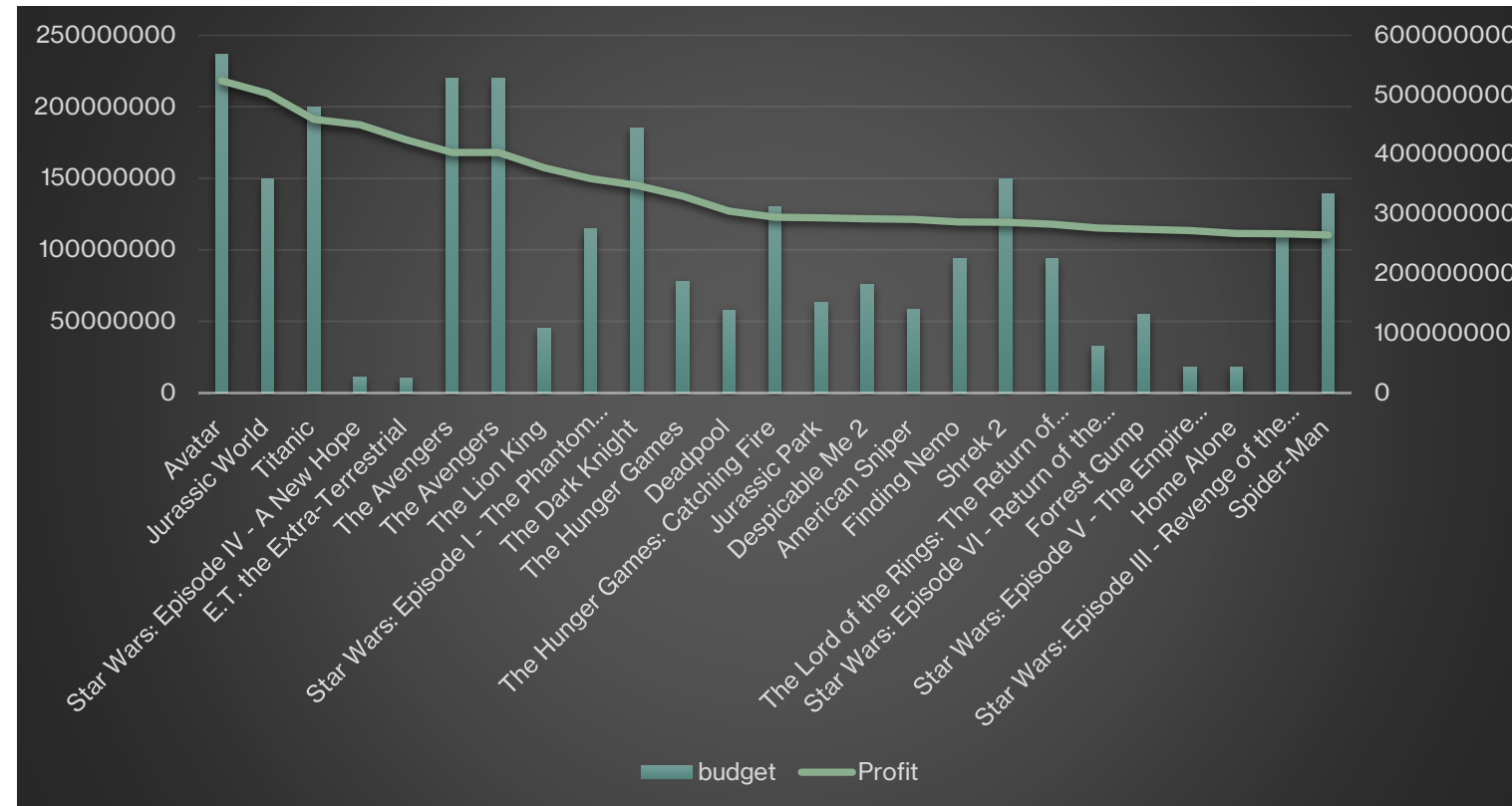


1. Clean The Data

- At first, The raw dataset was containing a total no of 5044 Rows and after cleaning the null and blank values we get 3744 Rows of data.
- We mainly used the excel and PowerBi to clean the data.
- The chart, Shows how the data is looking before and after disposing the garbage value.

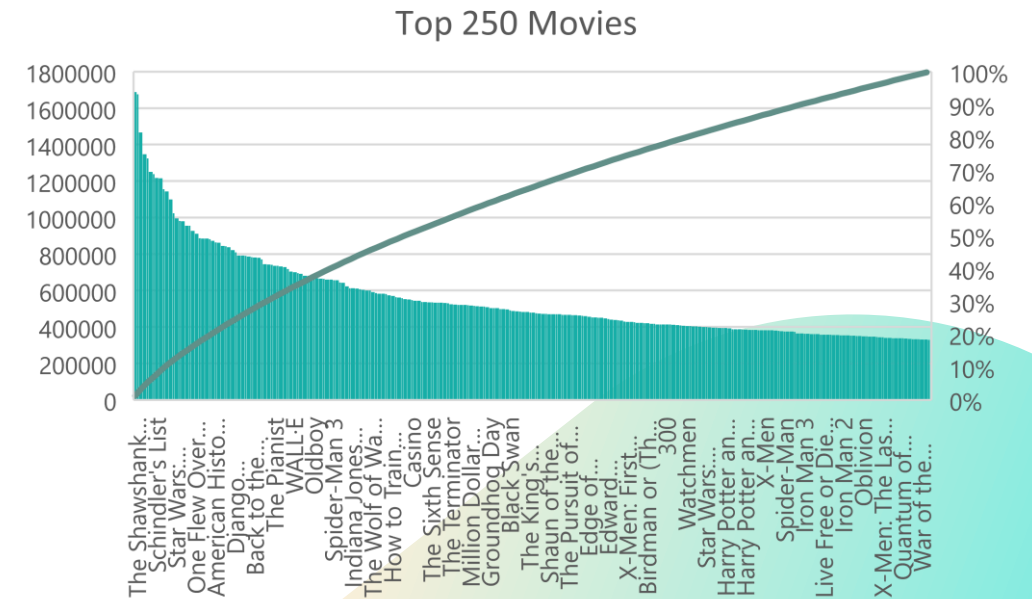


2. Movies with highest profit:



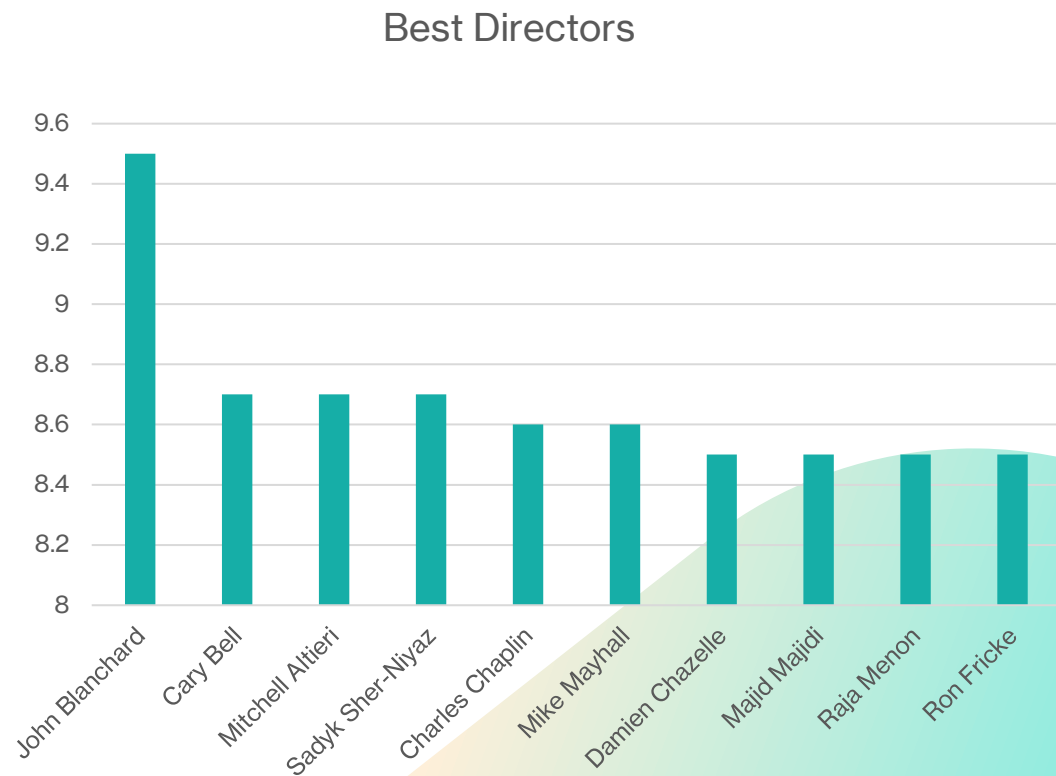
3. Find IMDB Top 250

1	movie_title	Movie_rated_by_above_25000	imdb_score
2	The Shawshank Redemption	1689764	9.3
3	The Godfather	1676169	9.2
4	The Dark Knight	1468200	9
5	The Godfather: Part II	1347461	9
6	Pulp Fiction	1324680	8.9
7	The Lord of the Rings: The Return of the King	1251222	8.9
8	Schindler's List	1238746	8.9
9	The Good, the Bad and the Ugly	1217752	8.9
10	Inception	1215718	8.8
11	Fight Club	1155770	8.8
12	Forrest Gump	1144337	8.8
13	The Lord of the Rings: The Fellowship of the Ring	1100446	8.8
14	Star Wars: Episode V - The Empire Strikes Back	1023511	8.8
15	The Matrix	995415	8.7
16	The Lord of the Rings: The Two Towers	982637	8.7
17	Star Wars: Episode IV - A New Hope	980946	8.7
18	Goodfellas	955174	8.7
19	One Flew Over the Cuckoo's Nest	928227	8.7
20	City of God	911097	8.7
21	Se7en	887467	8.6
22	Interstellar	886204	8.6
23	The Silence of the Lambs	885175	8.6



4. The Best Directors

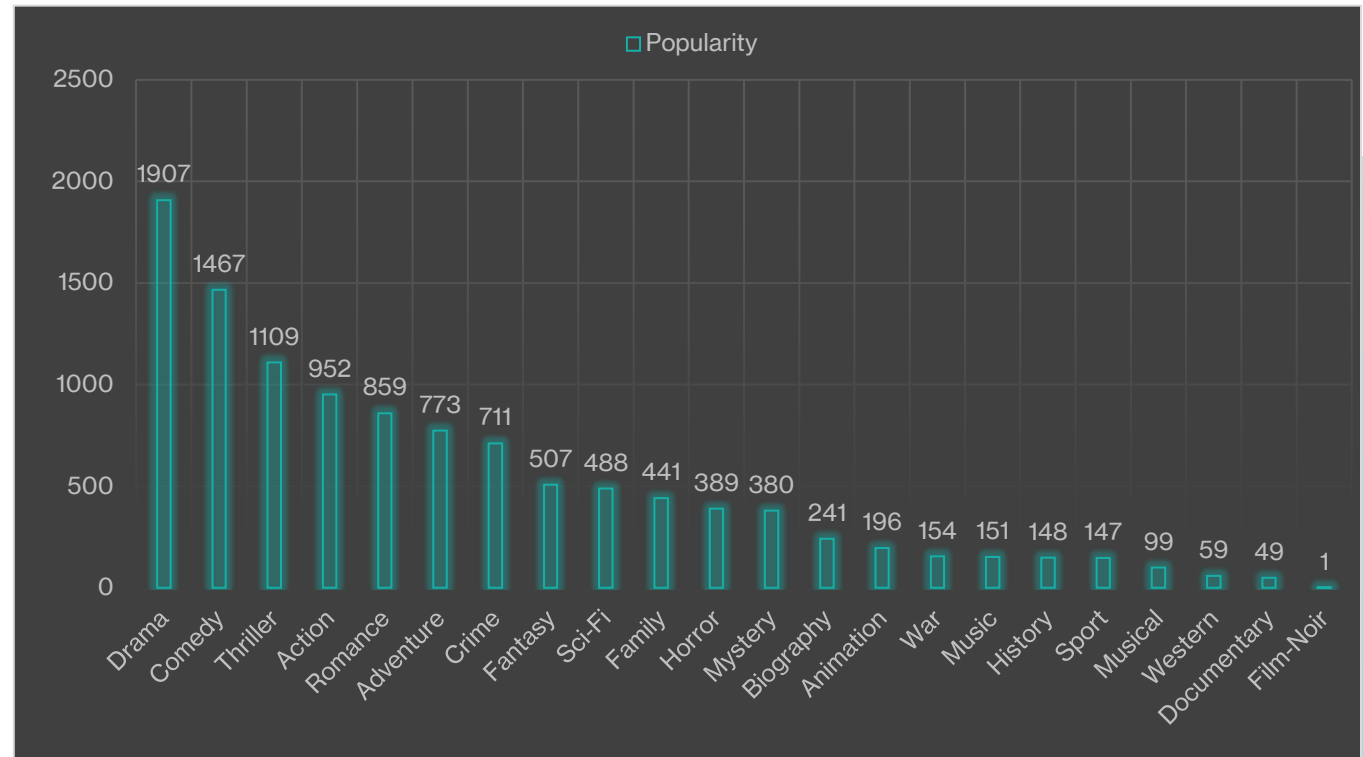
Director name	Mean of imdb_score
John Blanchard	9.5
Cary Bell	8.7
Mitchell Altieri	8.7
Sadyk Sher-Niyaz	8.7
Charles Chaplin	8.6
Mike Mayhall	8.6
Damien Chazelle	8.5
Majid Majidi	8.5
Raja Menon	8.5
Ron Fricke	8.5



5. Find Popular Genres

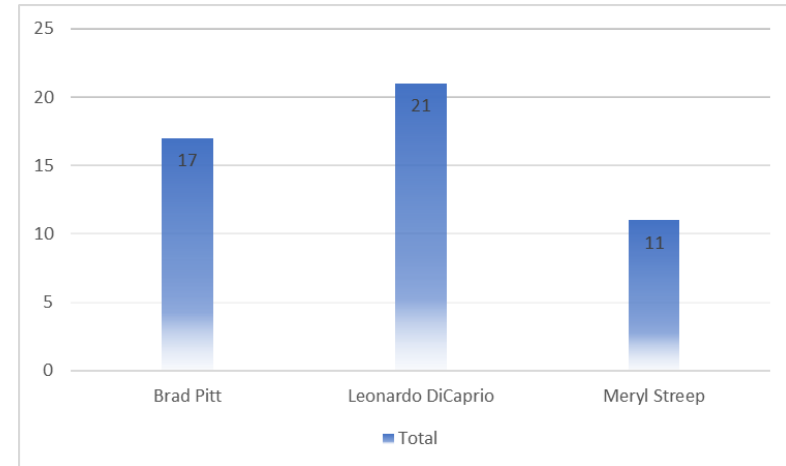
Genre	Popularity
Drama	1907
Comedy	1467
Thriller	1109
Action	952
Romance	859
Adventure	773
Crime	711
Fantasy	507
Sci-Fi	488
Family	441
Horror	389
Mystery	380
Biography	241
Animation	196
War	154
Music	151
History	148
Sport	147
Musical	99
Western	59
Documentary	49
Film-Noir	1

First, we used the delimiter to identify specific genre then we used the “countifs” function to calculate the Total number of movies from each genre.



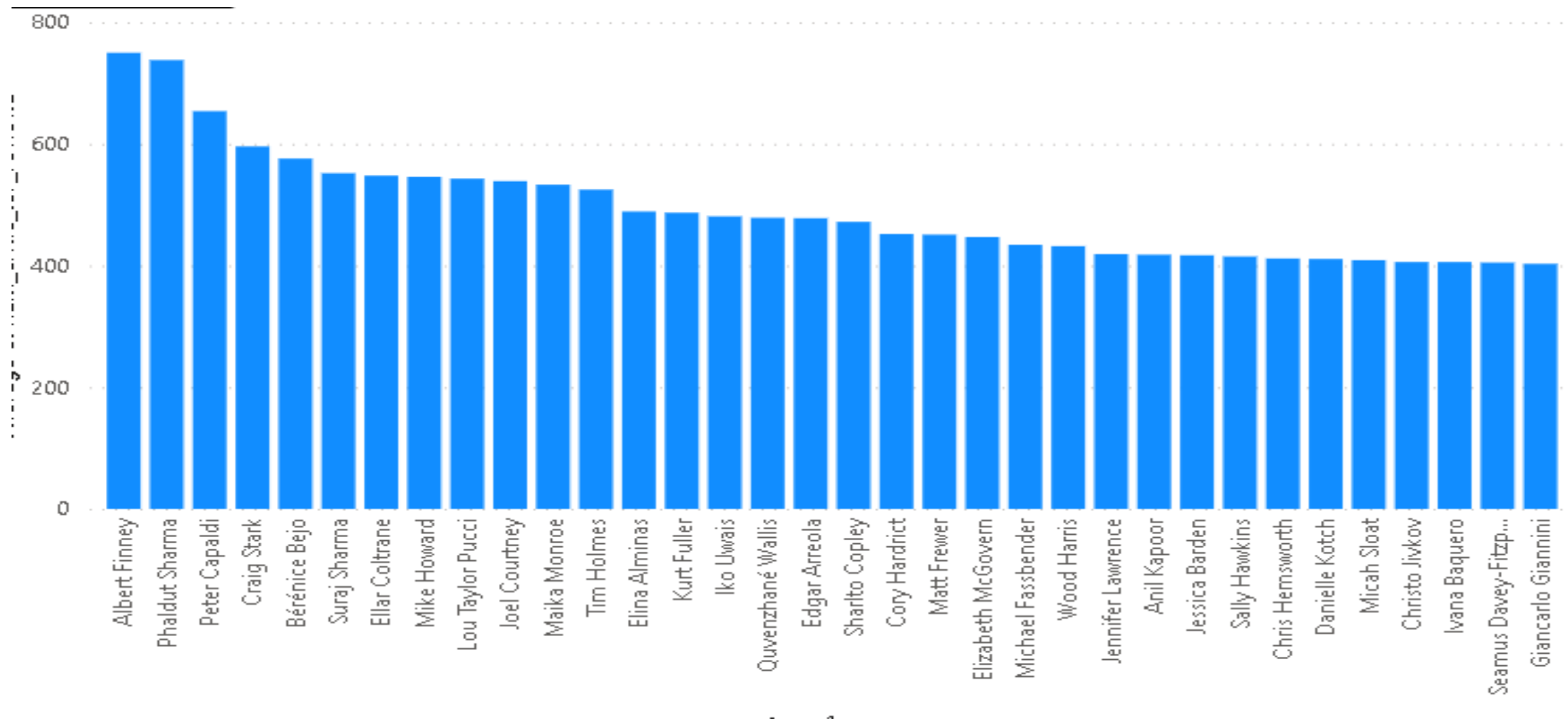
6.Find the critic-favorite and audience-favorite actors

- First, we calculated the total count of these three individual actors. Then Plotted this Graph based on the result.
- Then We Calculated the mean of num_of_critic and num_users for review and compared them with the actors mean to identify the actor with the highest mean.

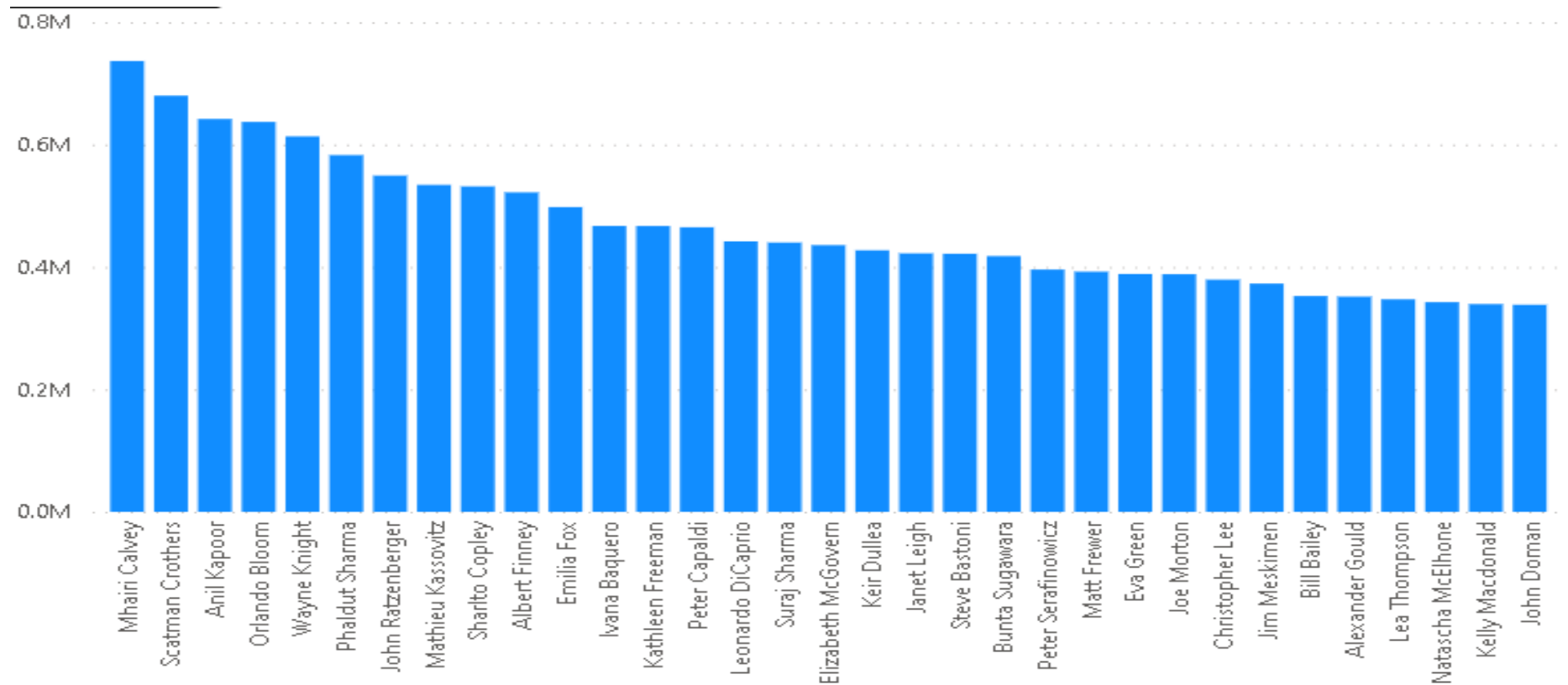


Row Labels	Average of num_critic_for_reviews	Average of num user for reviews
Brad Pitt	231.9444444	702.4444444
Leonardo DiCaprio	330.1904762	914.4761905
Meryl Streep	163.1538462	257.3076923
Grand Total	254.4230769	676.7884615

Then with the help of PowerBi, We plotted this graph to create a column chart on the base of our result from the previous slide. We put Actor_1_name on X-axis and mean of num_of_critic review on Y-axis and got the following chart. Which Shows us the critic favorite actors.



Again, with the help of PowerBi, We plotted this graph to create a column chart on the base of our result from the previous slide. We put Actor_1_name on X-axis and mean of num_of_voted_user on Y-axis and got the following chart. Which Shows us the Audience Favorite actor list tier.



Conclusion

- After going through the analysis, we can identify many key aspect hidden within the dataset. We Identify which movies achieved the most profit and which director we can call the best. And also, which actors are fan favorite and which actor are critic favorite.
- The whole dataset was more diverse, and the key aspect of data cleaning can be learned by going through this project.

THANK YOU

