**IMDB Movie Analysis**

**Project Description-**

This project is about IMDB Movie Analysis where we must analyse the data and take decision on basis of this dataset.

**Required to provide a detailed report answering the questions below:**

1. **Task:** Clean the data.
2. Plot profit (y-axis) vs budget (x-axis) and observe the outliers using the appropriate chart type.  
   **Task:** Find the movies with the highest profit.
3. **Task:**Find IMDB Top 250
4. **Best Directors:** Find out the top 10 directors for whom the mean of imdb\_score is the highest and store them in a new column top10director. In case of a tie in IMDb score between two directors, sort them alphabetically.  
   **Task:**Find the best directors.
5. **Task:**Find popular genres.
6. **Charts:** Find the mean of the number of critics review and number of user review and identify the actors which have the highest mean. Observe the change in the number of voted users over decades using a bar chart. find the sum of users who voted in each decade.   
   **Task:**Find the critic-favourite and audience-favourite actors.

**Tech-Stack Used:**

Software used - Microsoft excel

Used for –

1) Cleaning Data.

2)use excel to perform entire analysis to answer the questions asked above.

**Approach:**

After downloading the dataset, I cleaned it and developed a hypothesis to work upon.

Post cleaning, I analysed different areas and drew insights on my developed problem statement to get a wider view of the problem at hand.

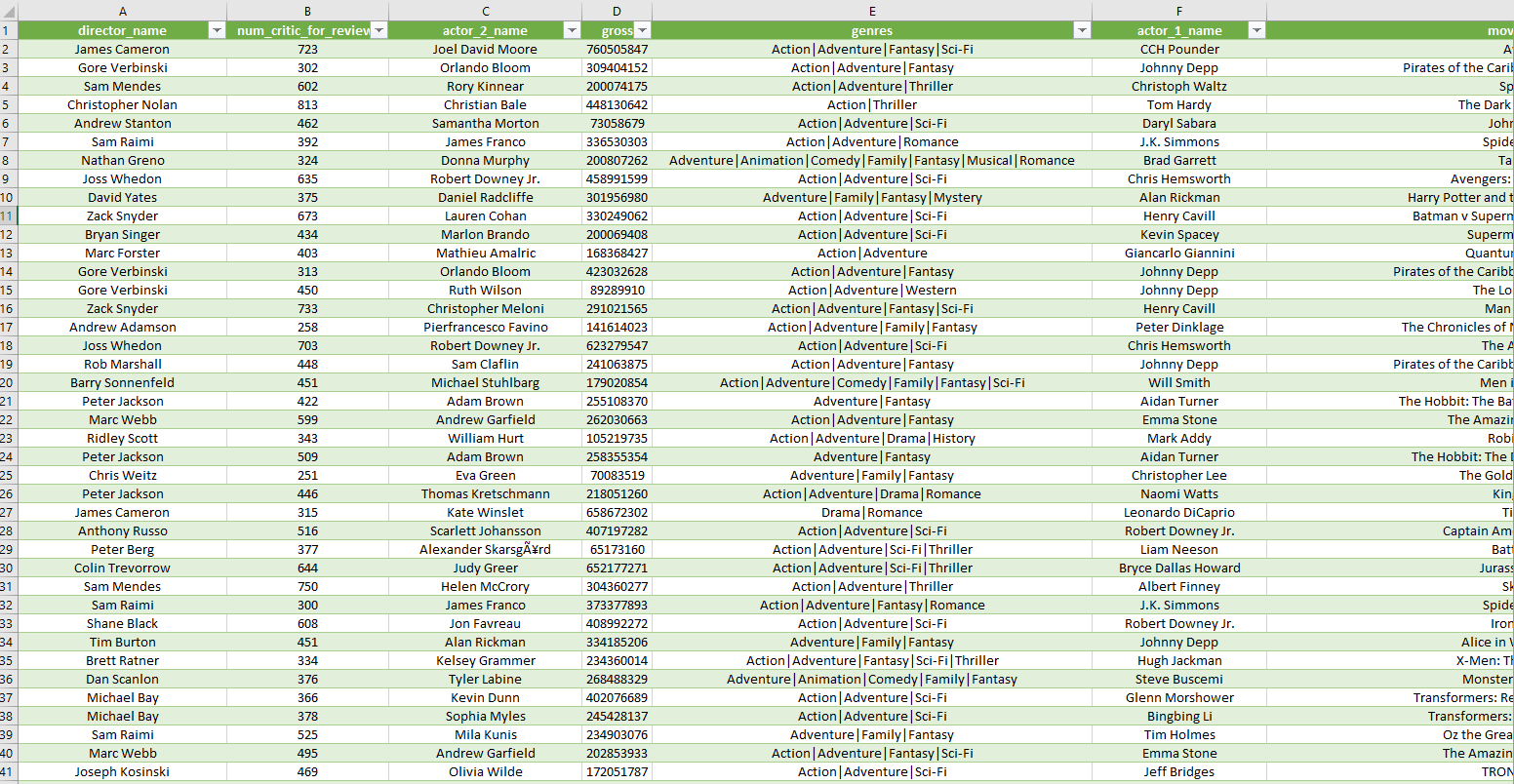
**Exploring the dataset**

The IMDb movies dataset had initially 5025 rows and 28 columns.

There are 5025 rows and 28 column which includes duplicates, blank and unknown values. Then I removed the duplicate rows, irrelevant column and blank data.

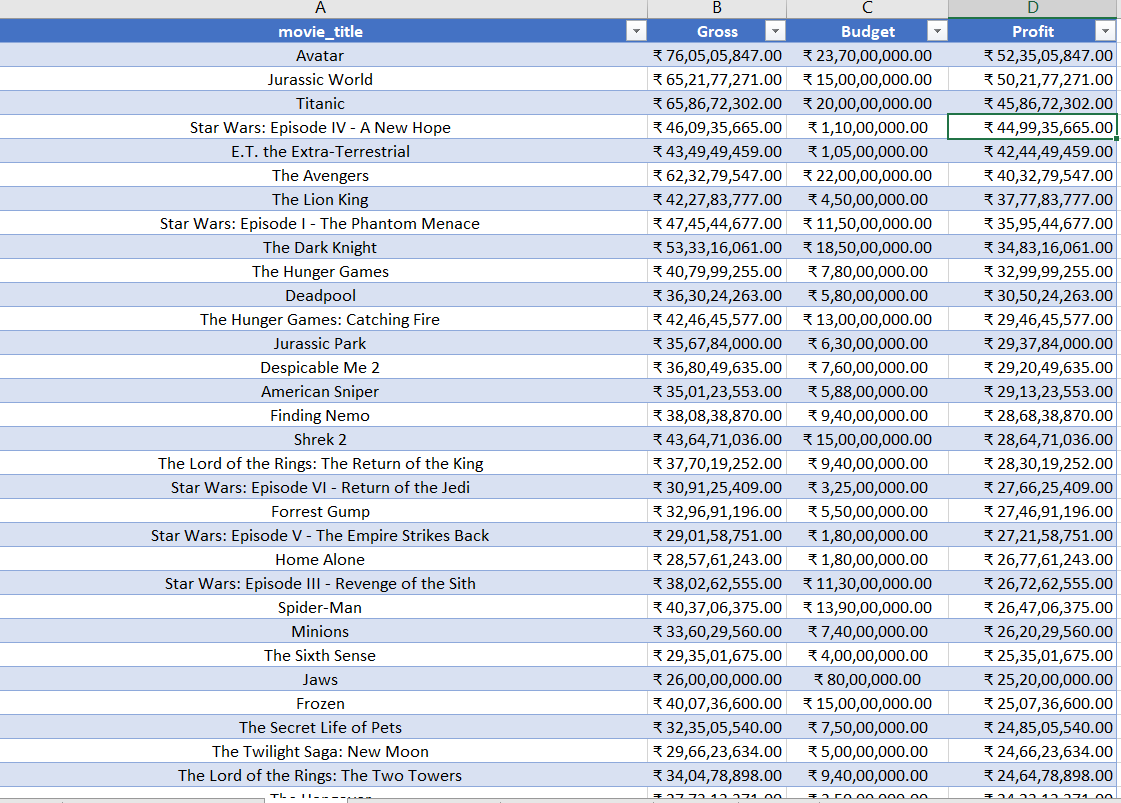
After cleaning, the total number of retains rows were 3724 and 17 columns from which I performed my further analysis.

Task A:



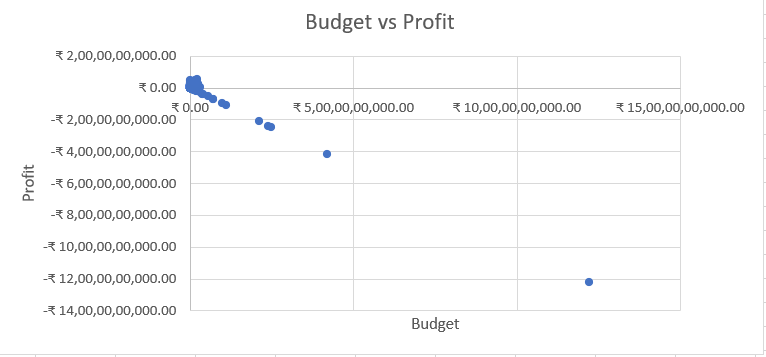
Clean dataset by deleting blank and irrelevant column.

Task B:



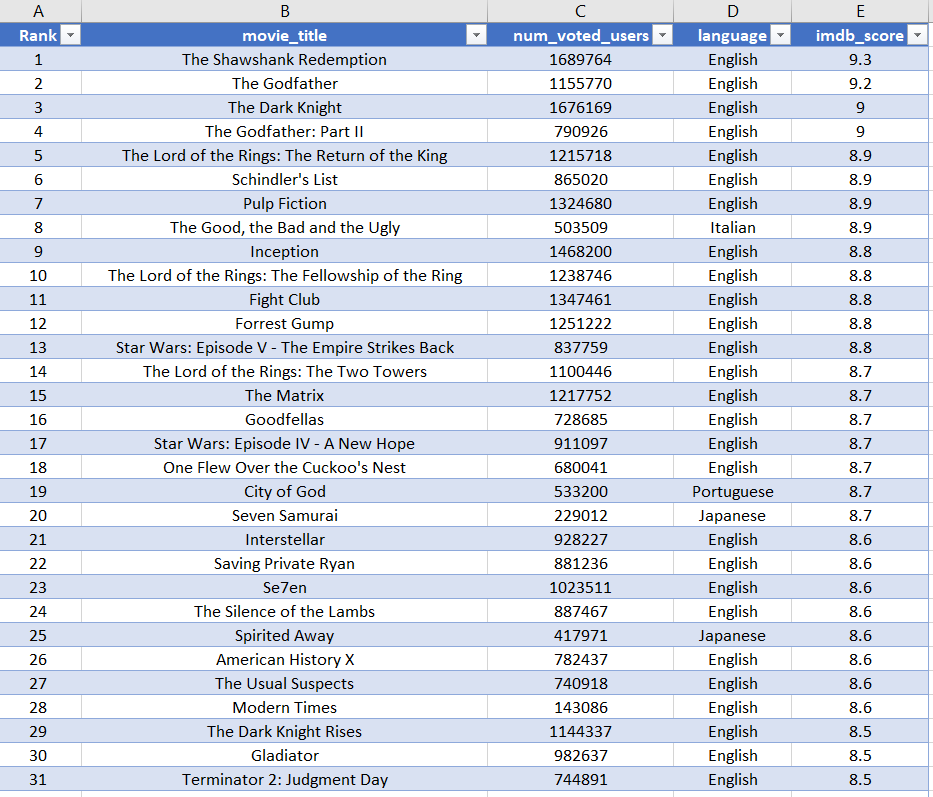
Profit= Gross-Budget

There are the top 10 most profitable movies out of which avatar is the most profitable one.



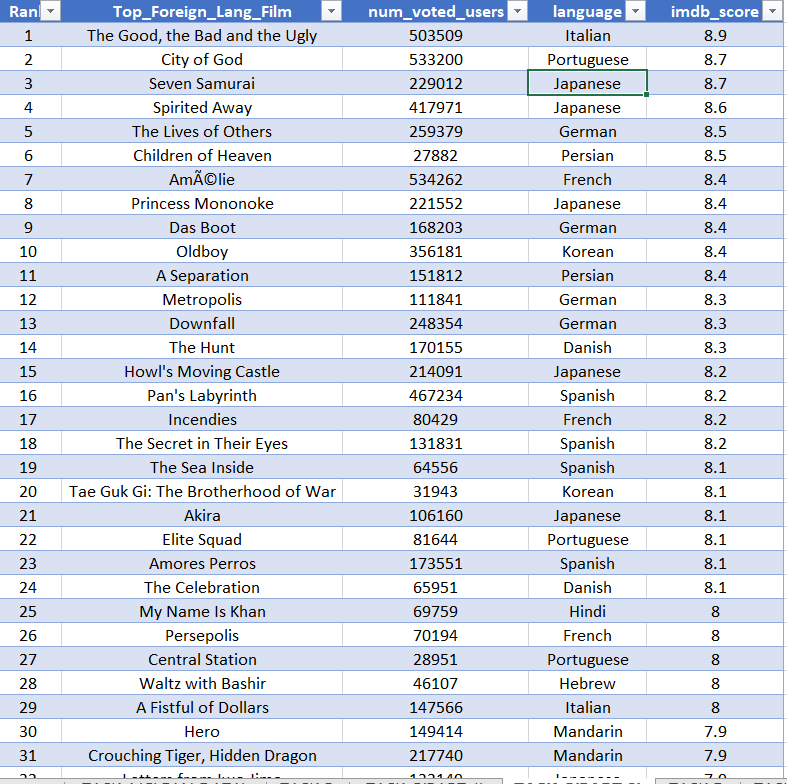
**Plotting profit (y-axis) vs budget (x- axis) and observing the outliers using the appropriate X-Y scattered chart type.**

Task C:

Part 1:

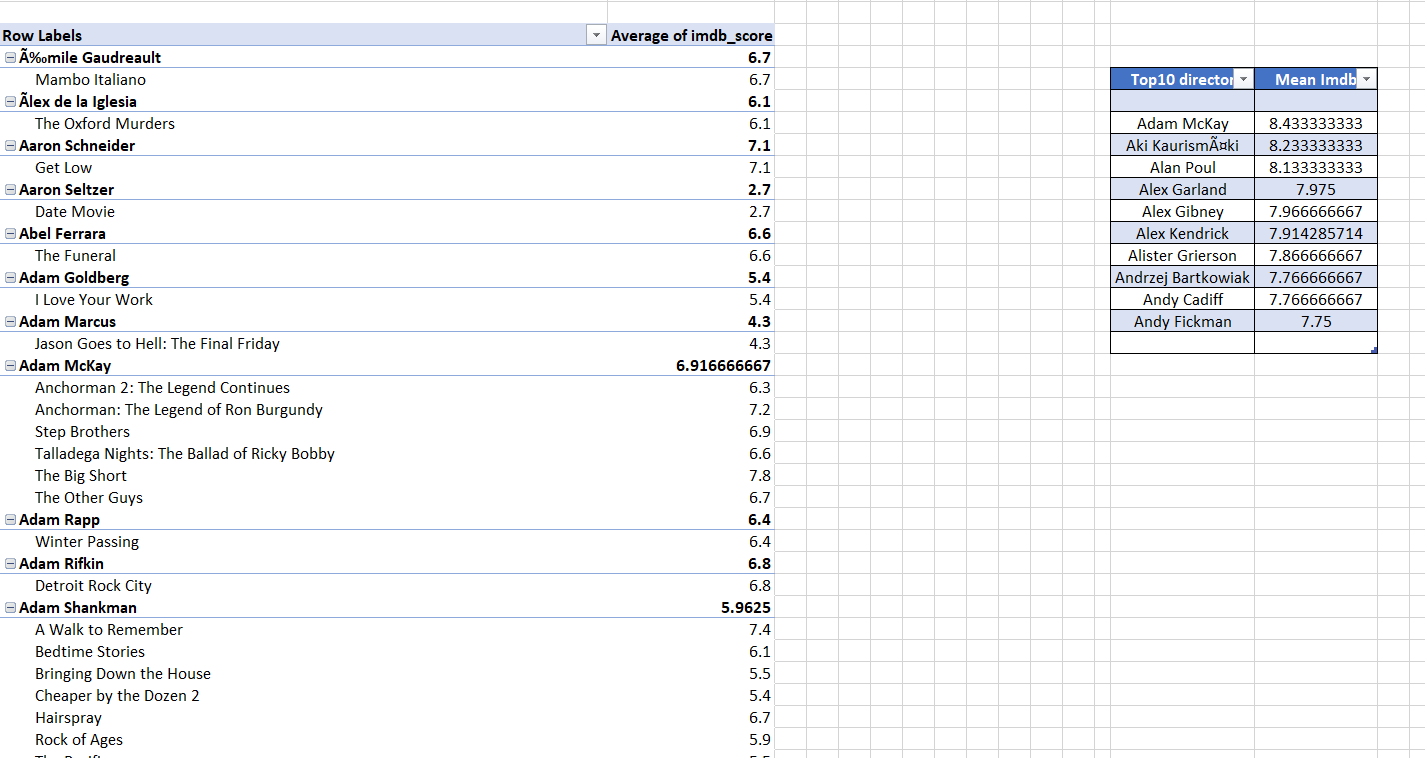
These are the top 250 movies where the number of votes given by users exceeds 25000. I have made a pivot table here with the necessary filters.

Part 2:



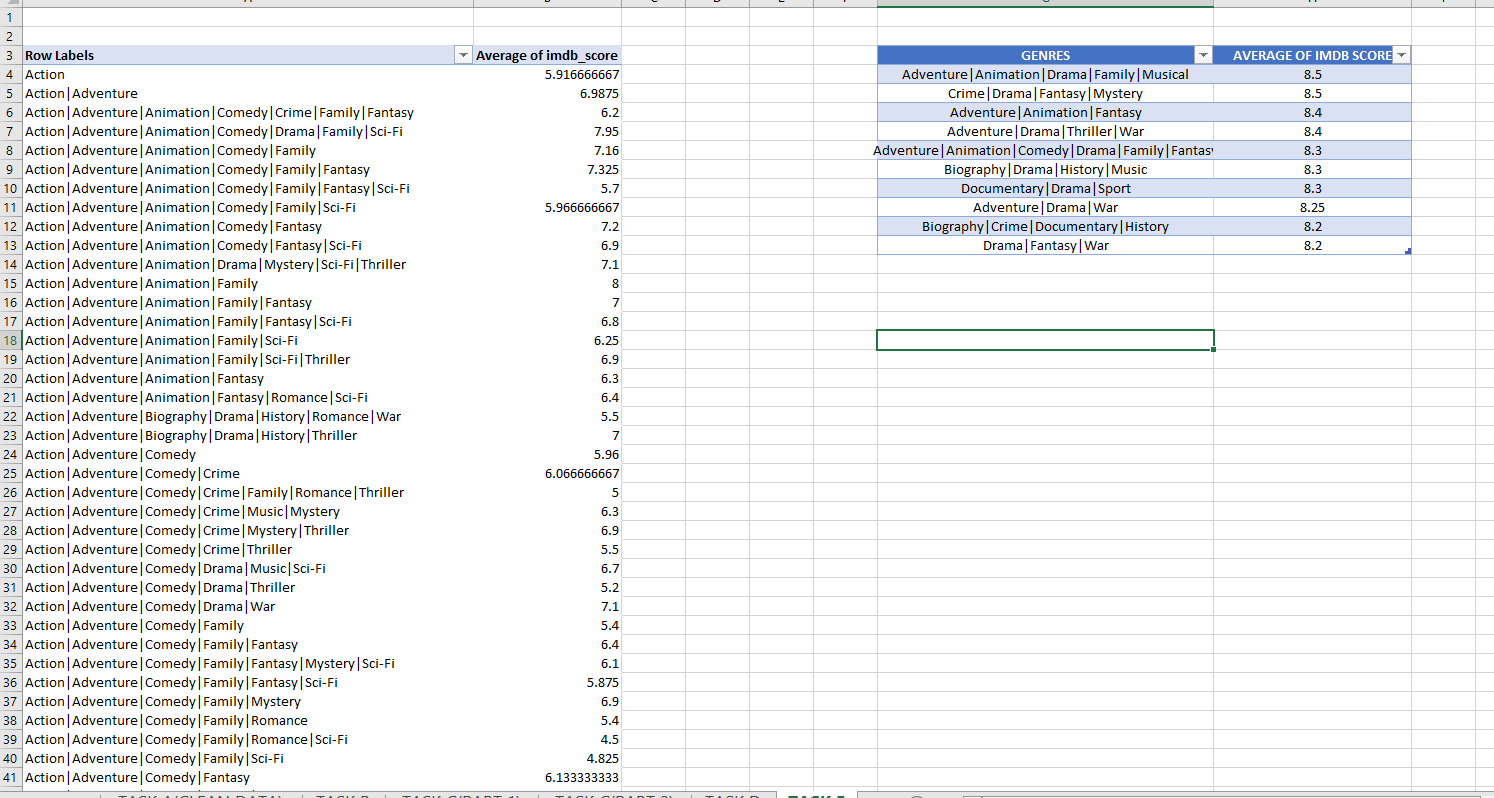
These are the values which are different language from top 250 movies list.

Task D:



These are the top 10 directors who have directed some of the greatest movies. I get these insights with the help of pivot table.

Task E:



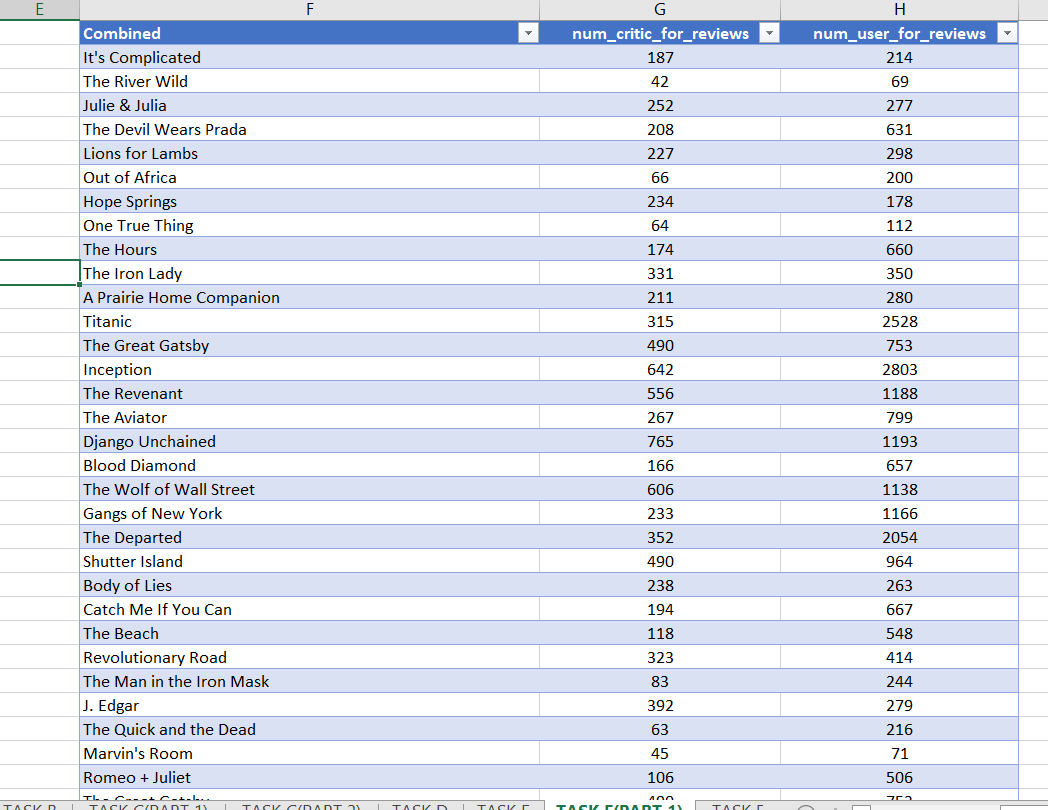
These are the top 10 genres in movie industry which have the highest imdb rating.

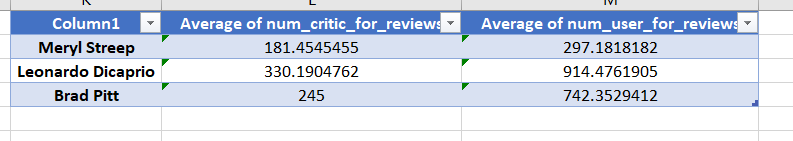
Task F:

Part 1:



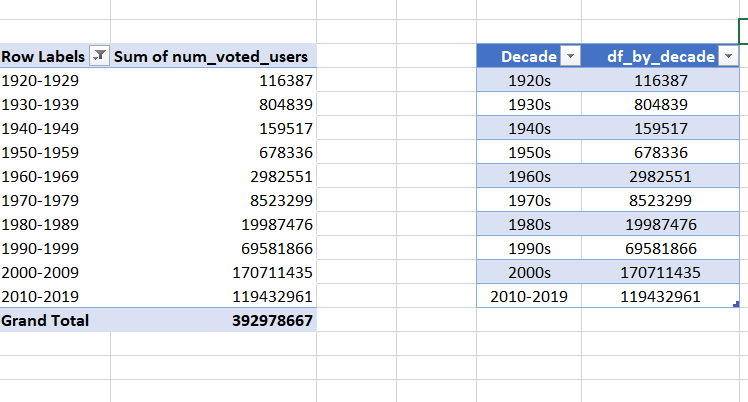
Separately finding the list of movies of the actors given above with the help of Data filtering.

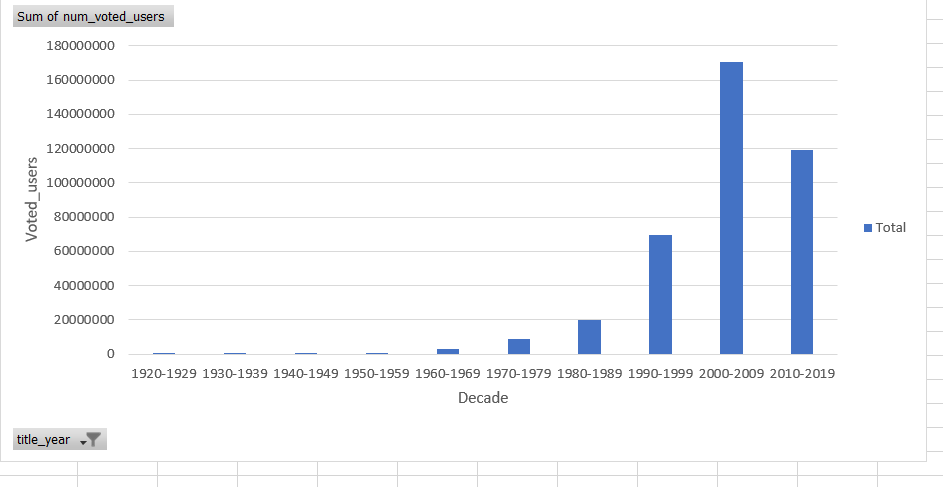




After combining, finding separately the average number critic for reviews and number user for reviews for Meryl Streep, Leonardo DiCaprio and Brad Pitt.

Part 2:

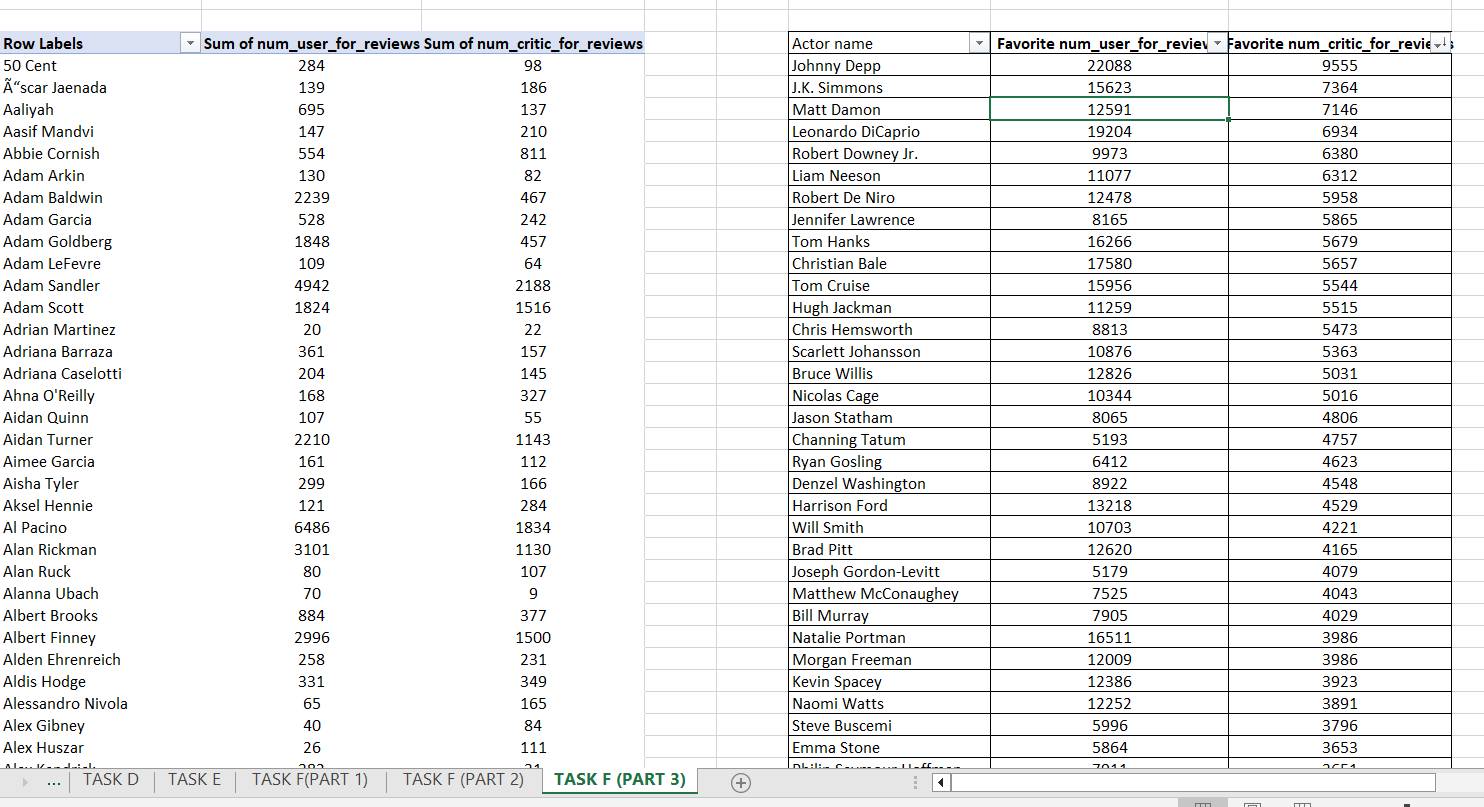




Creating a column name decade which represents the decade to which every movie belongs to. Then grouping it into decade and finding the sum of users voted in each decade.

Part 3:

Finding critic favourite & Audience favourite actor.



Link for excel: <https://docs.google.com/spreadsheets/d/19Y_TzTua9SXSHOT_9x6gW_3lDvlxpWvr/edit?usp=share_link&ouid=105895701067378695065&rtpof=true&sd=true>

Link for Video: <https://www.loom.com/share/eab9494415ac49ce84df45202a612ac2>

https://drive.google.com/file/d/1L8ICTg8zTtS\_TSMyT6hrG11Zuz0QKDr8/view?usp=share\_link