Project Description:

- This project is about conducting a comprehensive analysis of IMDb data to gain insights into the trends, patterns, and factors influencing movie ratings, profitability, and the performance of directors and genres.
- By leveraging statistical and data visualization techniques, we aim to provide a deeper understanding of the dynamics within the film industry.

Approach:

- We will clean the data by removing blanks, irregular patterns in any column and unwanted columns. Here we are handling missing and inconsistencies in data. Then use excel functions to gain insights from the data.

Tech-Stack Used:

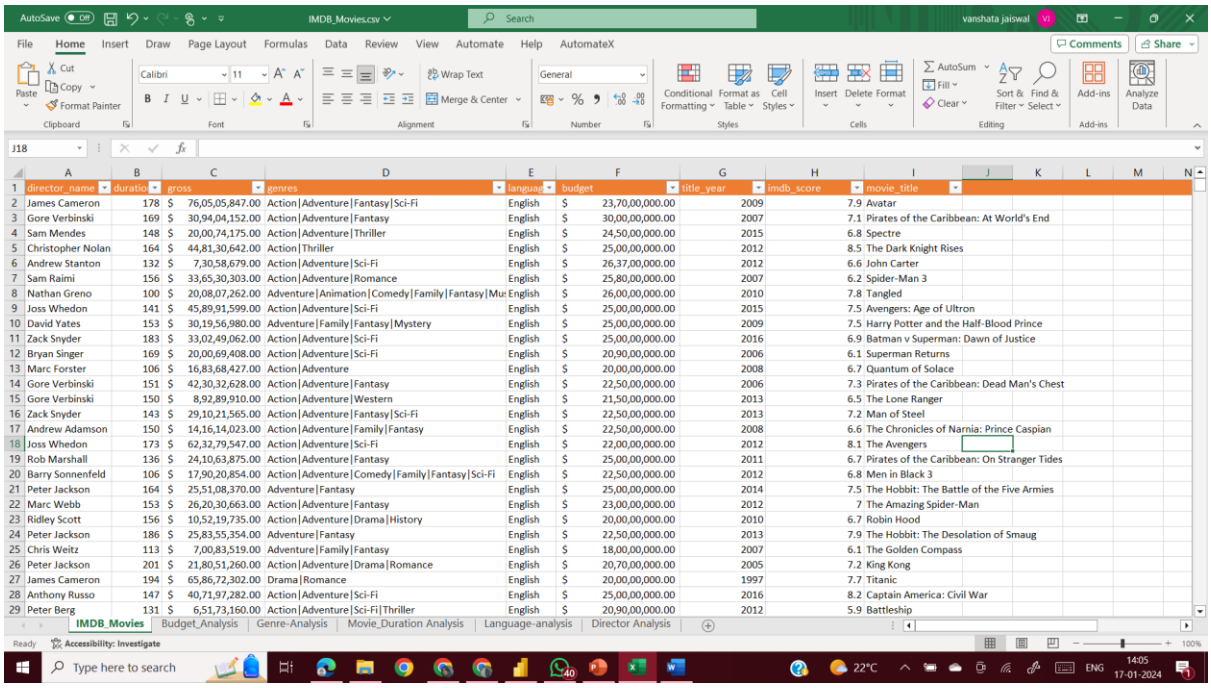
- MS Excel 2019 has been used for this task.

Insights:

There are 5 tasks provided:

- **Movie Genre Analysis**
- **Movie Duration Analysis**
- **Language Analysis**
- **Director Analysis**
- **Budget Analysis**

Before analysing the tasks, we need to clean the data. Below is the screenshot of good data.

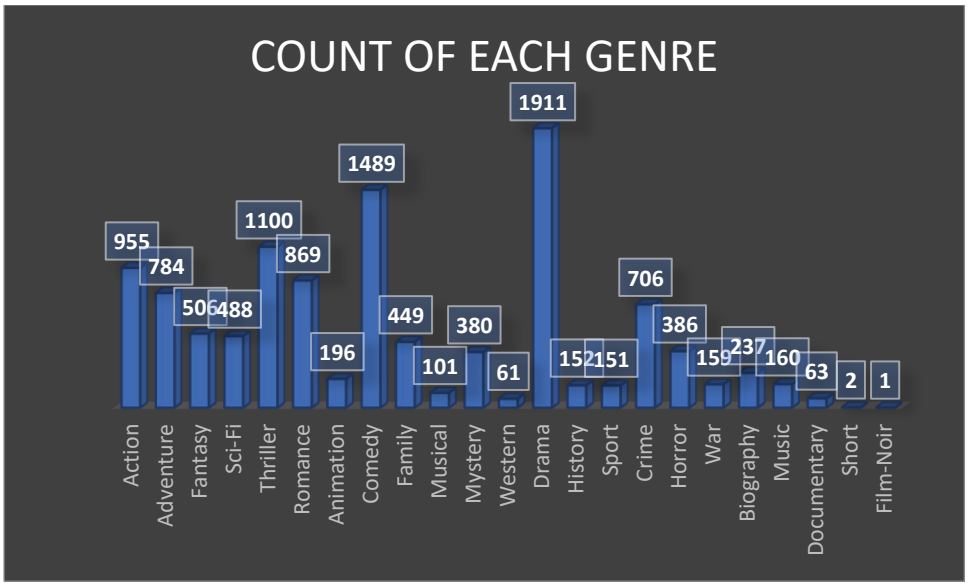


director_name	duration	gross	genres	language	budget	title_year	imdb_score	movie_title
James Cameron	178	$ 76,05,05,847.00	Action|Adventure|Fantasy|Sci-Fi	English	$ 23,70,00,000.00	2009	7.9	Avatar
Gore Verbinski	169	$ 30,94,04,152.00	Action|Adventure|Fantasy	English	$ 30,00,00,000.00	2007	7.1	Pirates of the Caribbean: At World's End
Sam Mendes	148	$ 20,00,74,175.00	Action|Adventure|Thriller	English	$ 24,50,00,000.00	2015	6.8	Spectre
Christopher Nolan	164	$ 44,81,30,642.00	Action|Thriller	English	$ 25,00,00,000.00	2012	8.5	The Dark Knight Rises
Andrew Stanton	132	$ 7,30,58,679.00	Action|Adventure|Sci-Fi	English	$ 26,37,00,000.00	2012	6.6	John Carter
Sam Raimi	156	$ 33,65,30,303.00	Action|Adventure|Romance	English	$ 25,80,00,000.00	2007	6.2	Spider-Man 3
Nathan Greno	100	$ 20,08,07,262.00	Adventure|Animation|Comedy|Family|Fantasy|Mu	English	$ 26,00,00,000.00	2010	7.8	Tangled
Joss Whedon	141	$ 45,89,91,599.00	Action|Adventure|Sci-Fi	English	$ 25,00,00,000.00	2015	7.5	Avengers: Age of Ultron
David Yates	153	$ 30,19,56,980.00	Adventure|Family|Fantasy|Mystery	English	$ 25,00,00,000.00	2009	7.5	Harry Potter and the Half-Blood Prince
Zack Snyder	183	$ 33,02,49,062.00	Action|Adventure|Sci-Fi	English	$ 25,00,00,000.00	2016	6.9	Batman v Superman: Dawn of Justice
Bryan Singer	169	$ 20,00,69,408.00	Action|Adventure|Sci-Fi	English	$ 20,90,00,000.00	2006	6.1	Superman Returns
Marc Forster	106	$ 16,83,68,427.00	Action|Adventure	English	$ 20,00,00,000.00	2008	6.7	Quantum of Solace
Gore Verbinski	151	$ 42,30,32,628.00	Action|Adventure|Fantasy	English	$ 22,50,00,000.00	2006	7.3	Pirates of the Caribbean: Dead Man's Chest
Gore Verbinski	150	$ 8,92,89,910.00	Action|Adventure|Western	English	$ 21,50,00,000.00	2013	6.5	The Lone Ranger
Zack Snyder	143	$ 29,10,21,565.00	Action|Adventure|Fantasy|Sci-Fi	English	$ 22,50,00,000.00	2013	7.2	Man of Steel
Andrew Adamson	150	$ 14,16,14,023.00	Action|Adventure|Family|Fantasy	English	$ 22,50,00,000.00	2008	6.6	The Chronicles of Narnia: Prince Caspian
Joss Whedon	173	$ 62,32,79,547.00	Action|Adventure|Sci-Fi	English	$ 22,00,00,000.00	2012	8.1	The Avengers
Rob Marshall	136	$ 24,10,63,875.00	Action|Adventure|Fantasy	English	$ 25,00,00,000.00	2011	6.7	Pirates of the Caribbean: On Stranger Tides
Barry Sonnenfeld	106	$ 17,90,20,854.00	Action|Adventure|Comedy|Family|Fantasy|Sci-Fi	English	$ 22,50,00,000.00	2012	6.8	Men in Black 3
Peter Jackson	164	$ 25,51,08,370.00	Adventure|Fantasy	English	$ 25,00,00,000.00	2014	7.5	The Hobbit: The Battle of the Five Armies
Marc Webb	153	$ 26,20,30,663.00	Action|Adventure|Fantasy	English	$ 23,00,00,000.00	2012	7	The Amazing Spider-Man
Ridley Scott	156	$ 10,52,19,735.00	Action|Adventure|Drama|History	English	$ 20,00,00,000.00	2010	6.7	Robin Hood
Peter Jackson	186	$ 25,83,55,354.00	Adventure|Fantasy	English	$ 22,50,00,000.00	2013	7.9	The Hobbit: The Desolation of Smaug
Chris Weitz	113	$ 7,00,83,519.00	Adventure|Family|Fantasy	English	$ 18,00,00,000.00	2007	6.1	The Golden Compass
Peter Jackson	201	$ 21,80,51,260.00	Action|Adventure|Drama|Romance	English	$ 20,70,00,000.00	2005	7.2	King Kong
James Cameron	194	$ 65,86,72,302.00	Drama|Romance	English	$ 20,00,00,000.00	1997	7.7	Titanic
Anthony Russo	147	$ 40,71,97,282.00	Action|Adventure|Sci-Fi	English	$ 25,00,00,000.00	2016	8.2	Captain America: Civil War
Peter Berg	131	$ 6,51,73,160.00	Action|Adventure|Sci-Fi|Thriller	English	$ 20,90,00,000.00	2012	5.9	Battleship

- Removed the blanks from columns and have retained the columns which are necessary to gain the insights and complete the tasks given.
- Movie_title contained unusual alphabets, so we remove that. Director names had special characters in the name, so we clear that as well.
- Removed unwanted columns like Facebook likes and actor names which were not required for this analysis or tasks.

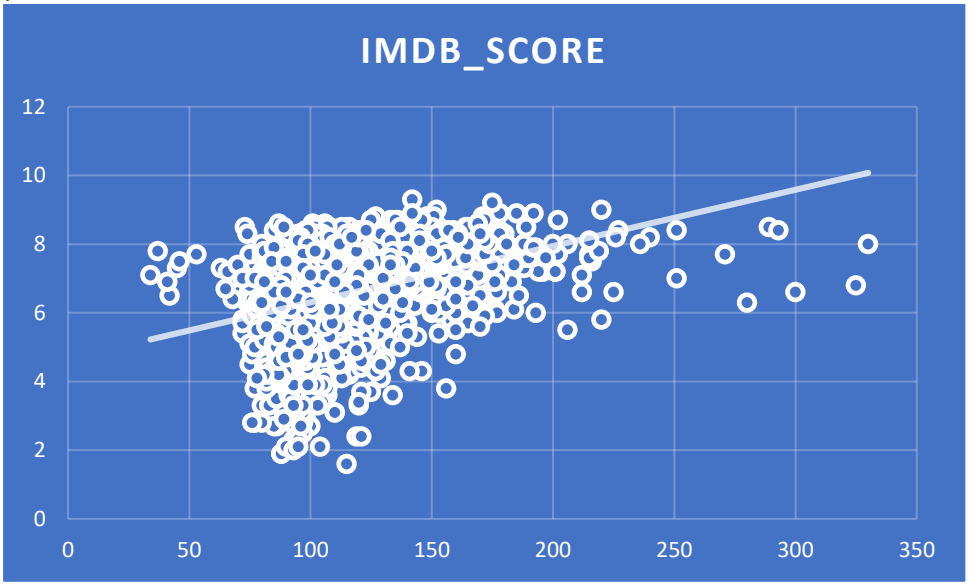
Task1: **Movie Genre Analysis----------**

- We analyse that the most popular genre is Drama.
- Used COUNTIF() to calculate the same. We use the AVERAGEIFS() formula to calculate the mean.
- AVERAGEIFS ($I$2:$I$3810,$A$2:$A$3810, "*"&J14&"*") - Here's a breakdown of the formula =AVERAGEIFS([IMDb Score Column], [Genre_column], ""&[Specific Genre]&"") according to your dataset
- This formula will calculate the genre's score according to the criteria (specific genre)
- if [Specific Genre] is "Action", the formula will average the IMDb scores for all rows where the [GenreColumn] contains the word "Action" anywhere in the string of genres.
- Similarly for standard deviation, median, mode, range, and variance.
- IF(ISNUMBER(SEARCH(""&[Specific Genre]&"", [Genre Column])), [IMDb Score Column]): This part checks each cell in the [Helper Column] for the presence of the [Specific Genre]. If the genre is found, it returns the corresponding IMDb score; otherwise, it returns FALSE. The entire calculation is in the excel sheet.



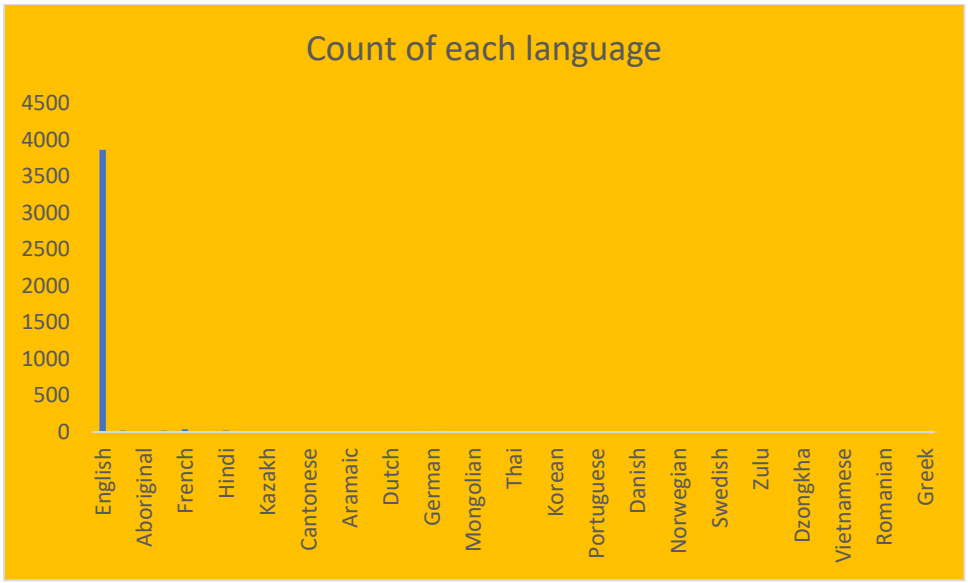
Task2: **Movie Duration Analysis ----------**

- Found the mean, standard deviation, minimum duration, maximum duration and median for the following task. The formula can be seen in the excel sheet.
- The analysis between the IMDB_SCORE and duration is seen below using the scatter plot with trendline.



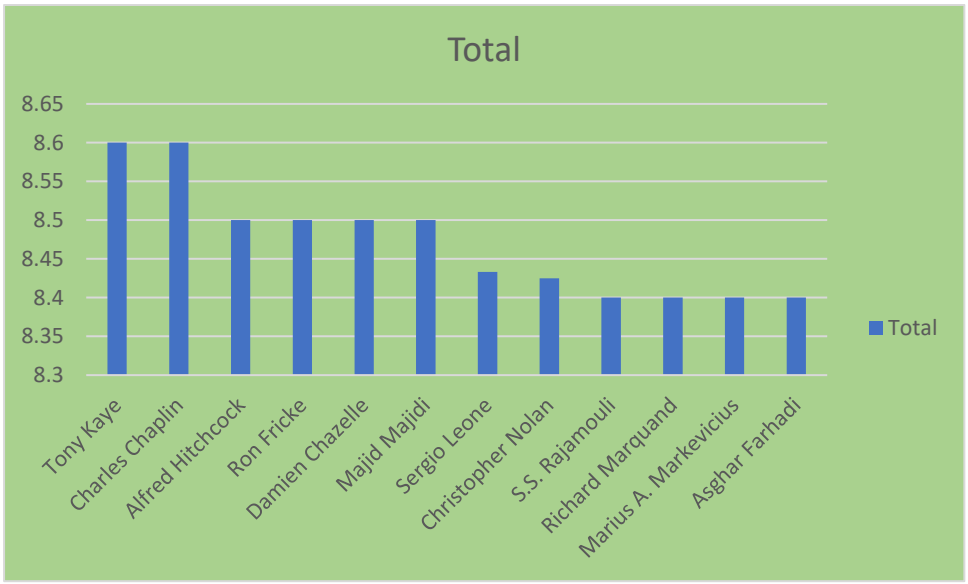
TASK3: **Language Analysis----------**

- English is the most popular language with 3860 as count using COUNTIF ()
- Calculated the median, mean and std. using the formulas averageif, median and std with IF condition. The details can be seen in excel sheet.
- Below is the graph for the languages count.



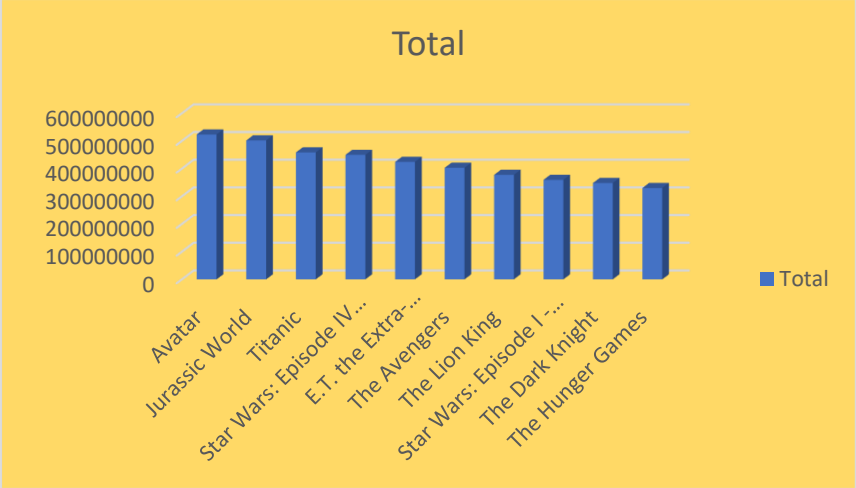
TASK4: **Director Analysis**----------

- Calculated the percentile- The top directors have been identified as those whose average IMDb score is above the 90th percentile of all IMDb scores.
- The percentile score is 7.7. Directors having average IMDB_SCORE more than this value will be rated as top directors.
- Directors like Charles Chaplin, Tony Kaye, and Alfred Hitchcock are among the top, with average scores ranging from 8.4 to 8.6. Created a pivot table and a chart to analyse the same. Refer excel sheet.



TASK5: **Budget Analysis**----------

- For this task, to find the profit, we subtract gross from budget columns.
- The find the maximum profit using max function from profit columns.
- The highest value found to be 523505847 which is the AVATAR.
- Avatar has the highest profit gain.
- Calculated the correlation coefficient between gross earnings and budget.
- Created a pivot table for the same. Please refer excel sheet.
-



RESULT:

- We learnt various excel functions through this project.
- Also, we now know how to analyse a dataset and create good insights so that the company knows the areas of improvement.
- LINK TO EXCEL SHEET---------
  https://docs.google.com/spreadsheets/d/1FzgAY2PpVNy3vQTqQm0qJr81nqPdmVMI/edit?usp=drive_link&ouid=100722541522733905001&rtpof=true&sd=true
- LOOM LINK -
  https://www.loom.com/share/3945d185ae5d4eef916feba5cc1c45ee?sid=d008da3b-98a8-4586-9522-bc413db88524