BOARD INFINITY

FOUNDATION COURSE IN DATA SCIENCE

A

PROJECT REPORT

ON

“IMDb MOVIES”

NAME : ASHIM MAITY

PROGRAM : B.TECH.CSE(DATA SCIENCE)

INSTITUTE : LOVELY PROFESSIONAL UNIVERSITY

SESSION : 2024-25

A PROJECT WORK OF DATA SCIENCE USING

TOOL : MICROSOFT EXCEL 365

IMDB MOVIE ANALYSIS

PROJECT DESCRIPTION :

The IMDb movie Analysis project aims to explore and analyze a comprehensive dataset of movies available by the board infinity platform. The dataset contains essential information about movies , movies titles, runtime ,genre , budget, gross earnings ,IMDb rating , and more. Through in-depth data analysis using Excel, Data visualization and statistics techniques this project seeks to extract valuable insights and trends that contribute to a movie's success and year wise requirements of common peoples and their involvement in that areas to have a specific analysis.

In this project , I was required to provide a detailed report for the below data record mentioning the answers of the questions that follows :

- **Movie Details :** Fetching the data about any specific movies listed in the IMDb movie dataset by the title name of the movie.
  - **Task :** Create a drop down menu with the title name of the movies and bring all the details about the movie from the dataset.
- **Movie Genre Analysis :** Analyze the distribution of movie genre and their demand.
  - **Task :** Determine the most common genres of the movie dataset. Then for each genre ,calculate the count of genre used in movies.
- **IMDb Rating Analysis :** Analyze the IMDb rating of the movies given in the dataset.
  - **Task :** Determine the top 10 IMDb rating movies and analyze the year wise count of IMDb rating of movies with slicers.
- **MetaCritic Rating Analysis :** Analyze the MetaCritic rating of the movies given in the dataset.
  - **Task :** Determine the top 10 MetaCritic rating movies and analyze the average metacritic rating of the movies year wise with slicers.
- **Lead actors and facebook likes :** Analyze the lead actors and their gained facebook likes from the given dataset.
  - **Task :** Determine the top 10 most facebook likes gained actors and top 10 highest average facebook likes gained actors with their graphical representation.
- **Country wise Analysis :** Analyze the country wise movie production and their graphical representation.
  - **Task :** Determine the countries with highest involvement in the production of movies.
- **Gross-Profit Analysis :** Analyze the gross profit earned by movies year wise and individually.
  - **Task :** Determine the Gross-Profit earned by top 10 countries and year wise total gross profit earning of movies.

- **C votes Analysis :** Analyze the C Votes rating that how many peoples are polling to which c votes rating.
    - **Task :** Determine the number of C Votes to each specifications year wise with slicers and their graphical representation.
- **Content_Rating Analysis :** Explore the different types of Content_Rating and analyze them.
    - **Task :** Determine the types of content rating and analyze them with year as well as country slicers.
- **Runtime Analysis :** Analyze the runtime of the movies and extract useful information from them.
    - **Task :** Determine the average runtime of movies year wise with country slicers and make graphical representation of them.
- **Average number of C_Votes :** Analyze the average number of different types of c votes involve in the polling about the rating of movies.
    - **Task :** Determine the average number of peoples involved in the C vote polling, extract their data and create graphical representation of them.
- **Average rating of C_Votes :** Analyze the average rating of different types of c votes involve in the polling about the movie's success.
    - **Task :** Determine the average rating given by the peoples of different measurement ,extract this data and create graphical representation of them.

MY APPROACH :

I have gone through the given dataset as IMDB movie dataset and understood all the given columns. Then I have observed that there are total 62 columns and 101 rows. I have organized all the data and after that extract given information from them with all my excel skills and knowledge.

TECH STACK :

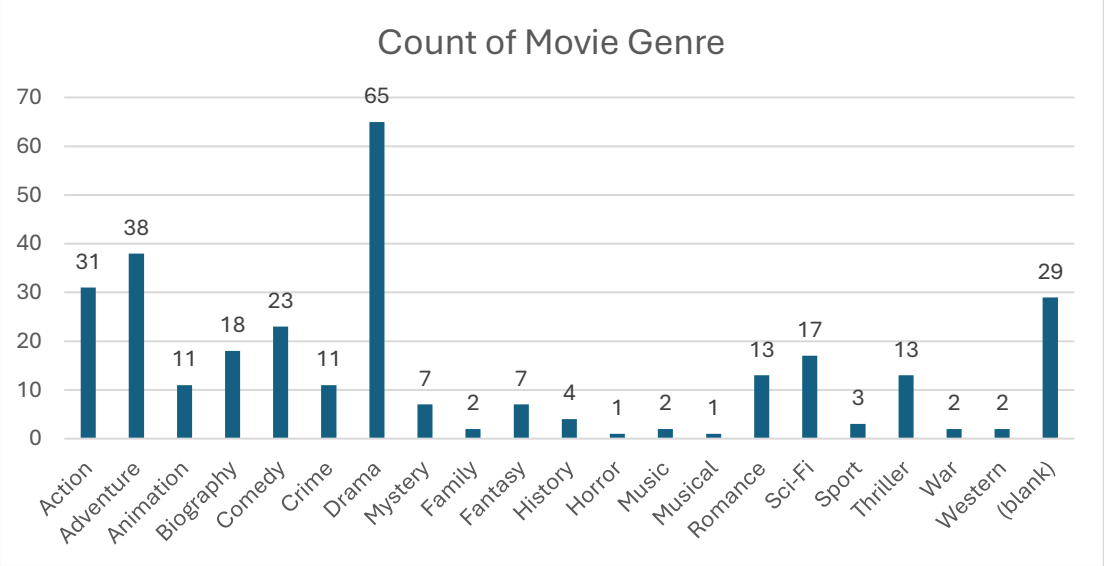
For this project I have used Microsoft Excel 365 to run the functions and get answers from the above questions. I also used this for the graphical representation of the data.

INSIGHTS :

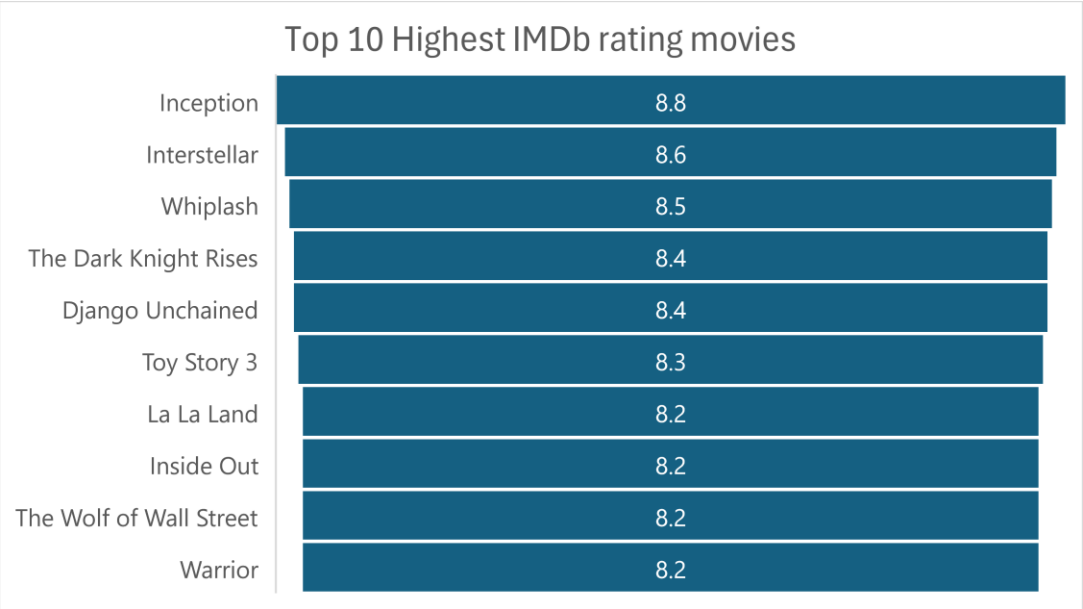
1. Movie Details :

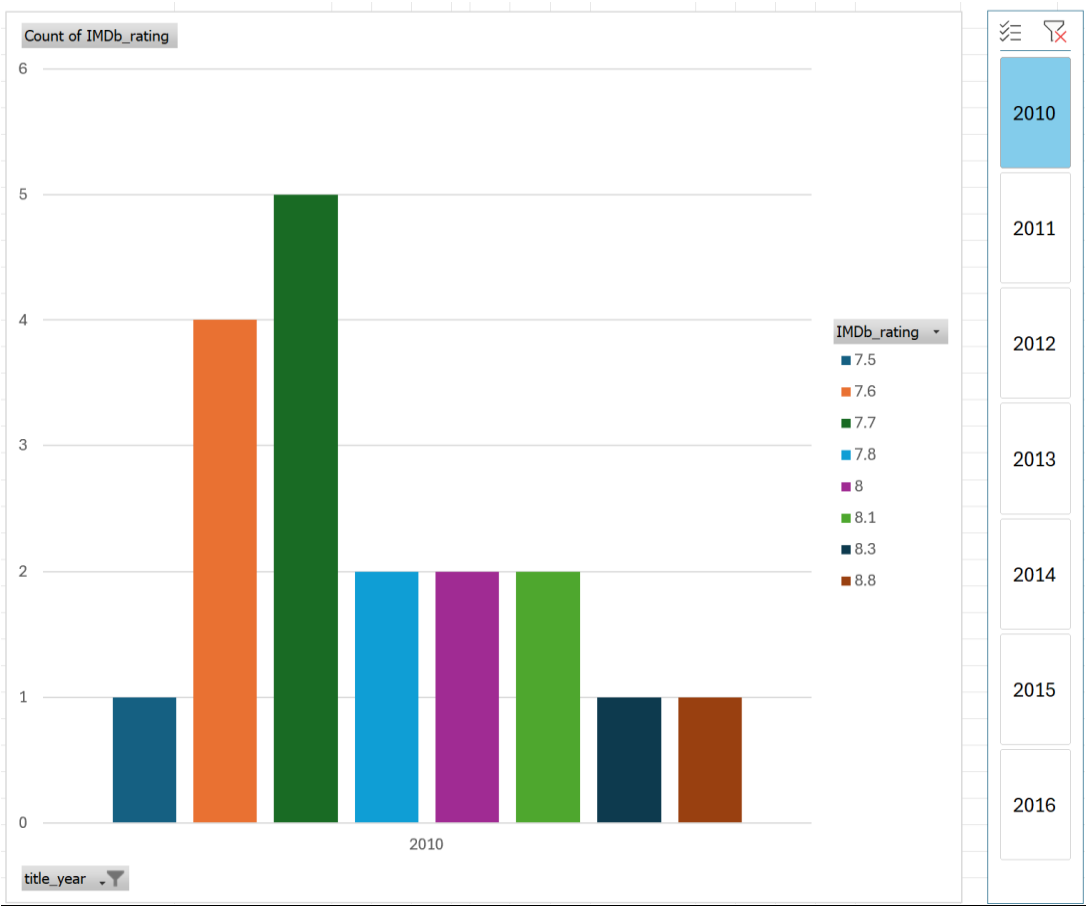
Column Name	Description	Movie Details
Title	Title of the movie	the avengers
budget	Budget of the movie in $	220000000
Gross	Gross of the movie in $	623279547
actor_1_name	Name of the lead actor in the movie	Chris Hemsworth
actor_2_name	Name of the second lead	Robert Downey Jr.
actor_3_name	Name of the third lead	Scarlett Johansson
actor_1_facebook_likes	Facebook likes of the lead actor in the movie	26000
actor_2_facebook_likes	Facebook Likes of the second lead actor in the movie	21000
actor_3_facebook_likes	Facebook Likes of the third lead actor in the movie	19000
IMDb_rating	IMDb rating of the movie	8.1
genre_1	Primary genre of the movie	Action
genre_2	Secondary genre of the movie	Sci-Fi
genre_3	Tertiary genre of the movie	0
MetaCritic	Average Rating on Metacritic (critic review website) on a scale of 100	69
Runtime	Movie length in minutes	143
CVotes10	Number of people who have voted 10/10 for the movie	260257
CVotes09	Number of people who have voted 9/10 for the movie	234203
CVotes08	Number of people who have voted 8/10 for the movie	264290
CVotes07	Number of people who have voted 7/10 for the movie	162604
CVotes06	Number of people who have voted 6/10 for the movie	67579
CVotes05	Number of people who have voted 5/10 for the movie	27957
CVotes04	Number of people who have voted 4/10 for the movie	12176
CVotes03	Number of people who have voted 3/10 for the movie	7201
Cvotes02	Number of people who have voted 2/10 for the movie	4996
CVotes01	Number of people who have voted 1/10 for the movie	15528
CVotesMale	Total number of Votes from the males	691783
CVotesFemale	Total number of Votes from the females	151617
CVotesU18	Number of votes from the age group under 18	4953
CVotesU18M	Number of votes from the male of age group under 18	3767
CVotesU18F	Number of votes from the females of the age group under 18	1150
CVotes1829	Number of votes from the age group 18 to 29	432999
CVotes1829M	Number of votes from the males of age group 18 to 29	343012
CVotes1829F	Number of votes from the females of age group 18 to 29	85465
CVotes3044	Number of votes from the age group 30 to 44	295318

2. Movie Genre Analysis :

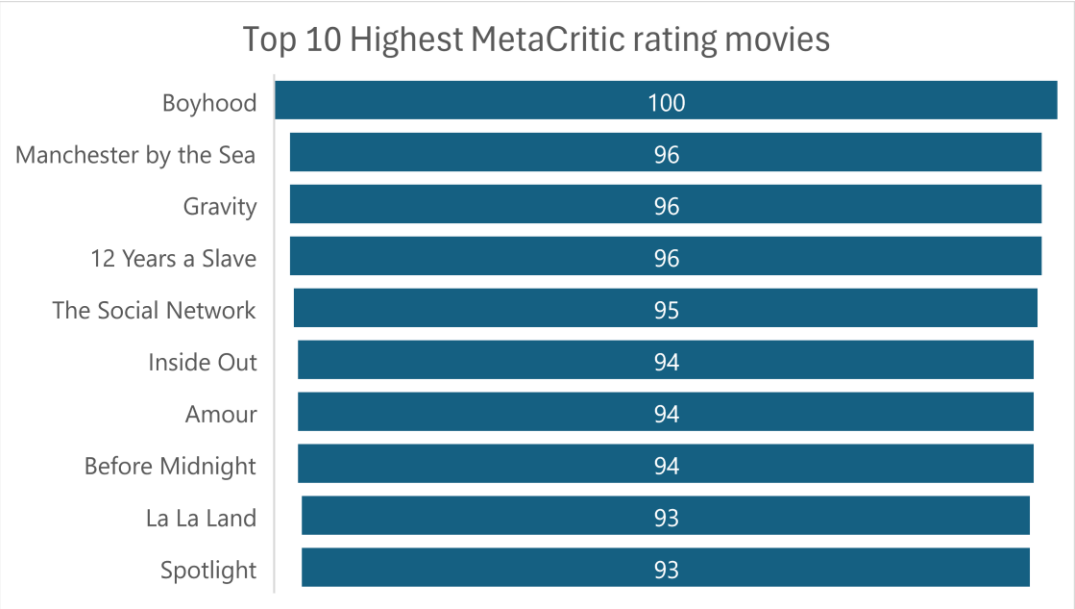


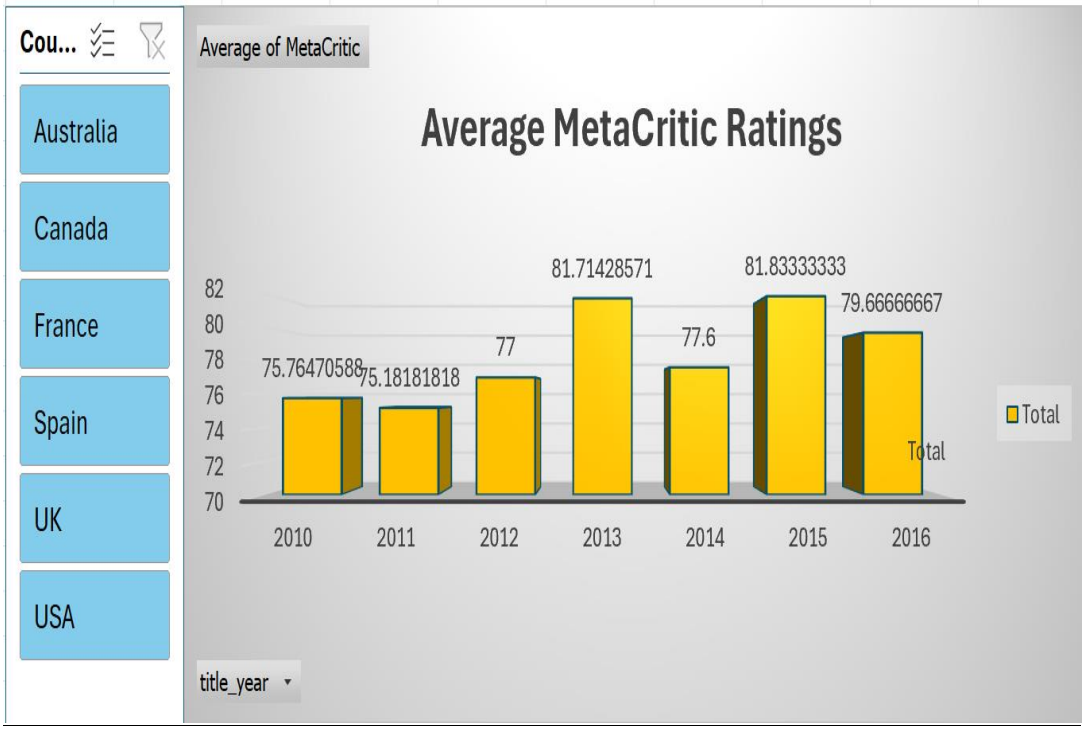
3. IMDb Rating Analysis :



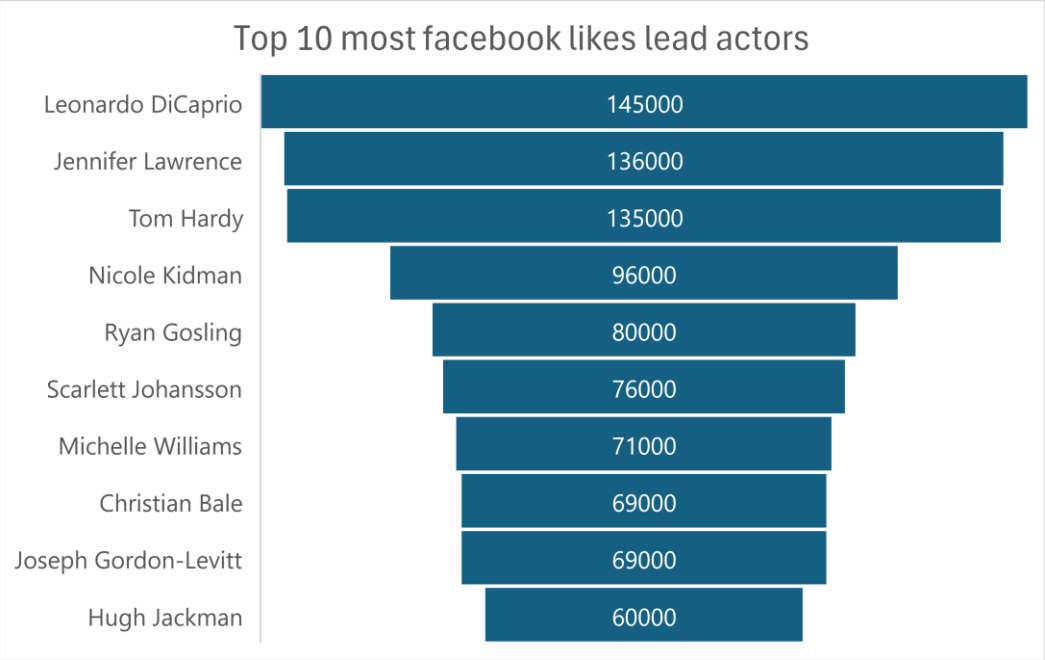


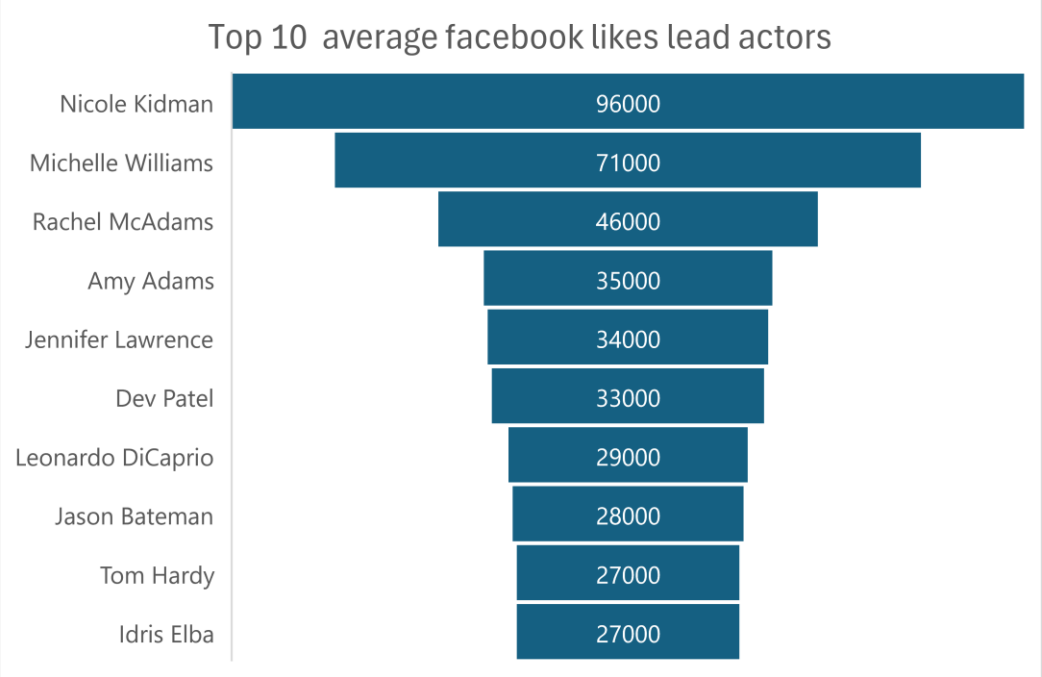
4. MetaCritic Rating Analysis :



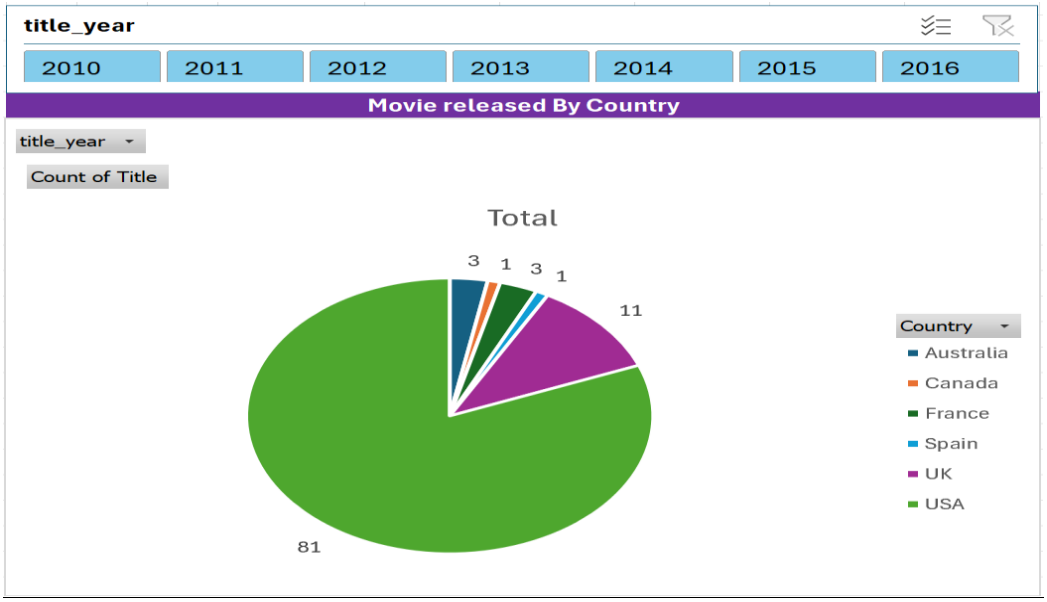


5. Lead actors and facebook rating :

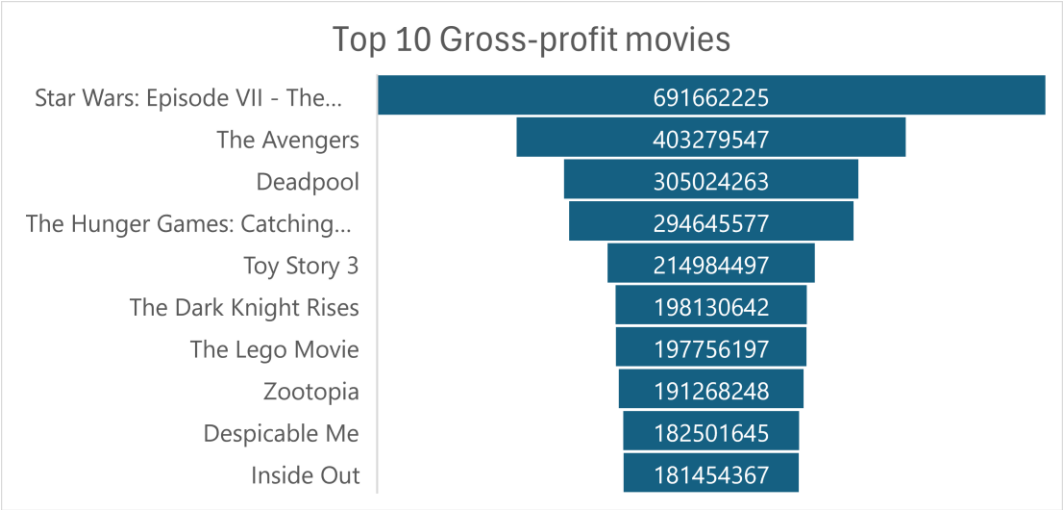


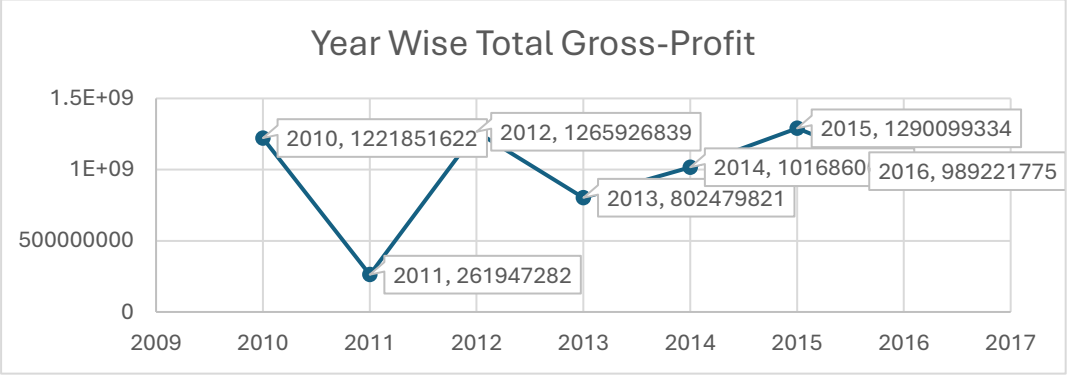


6. Country wise Analysis :

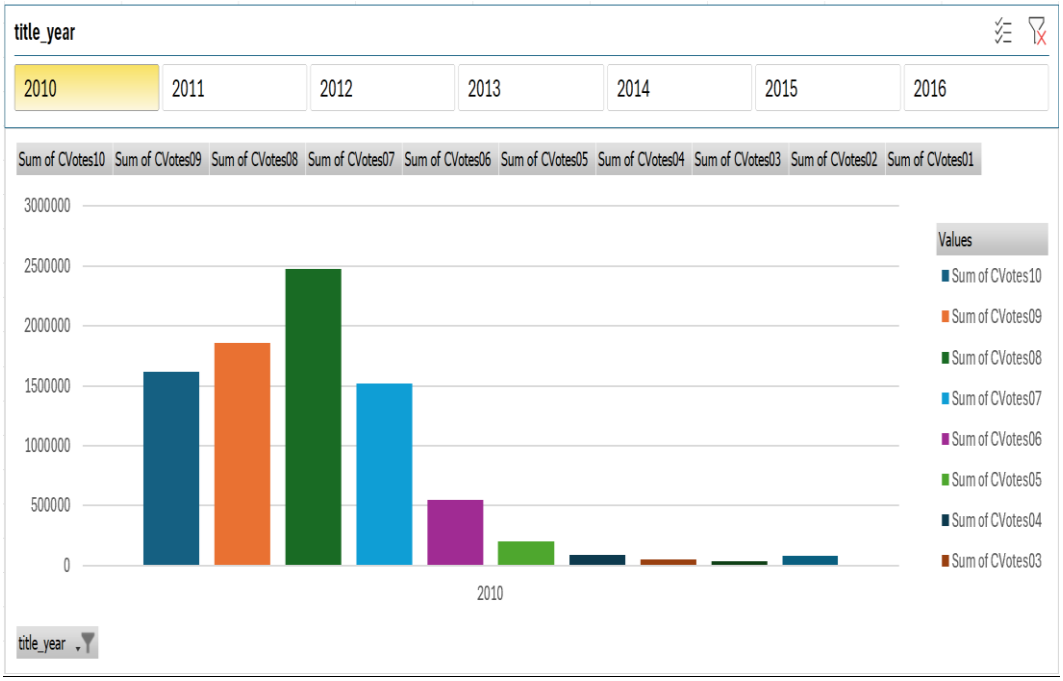


7. Gross-Profit Analysis :

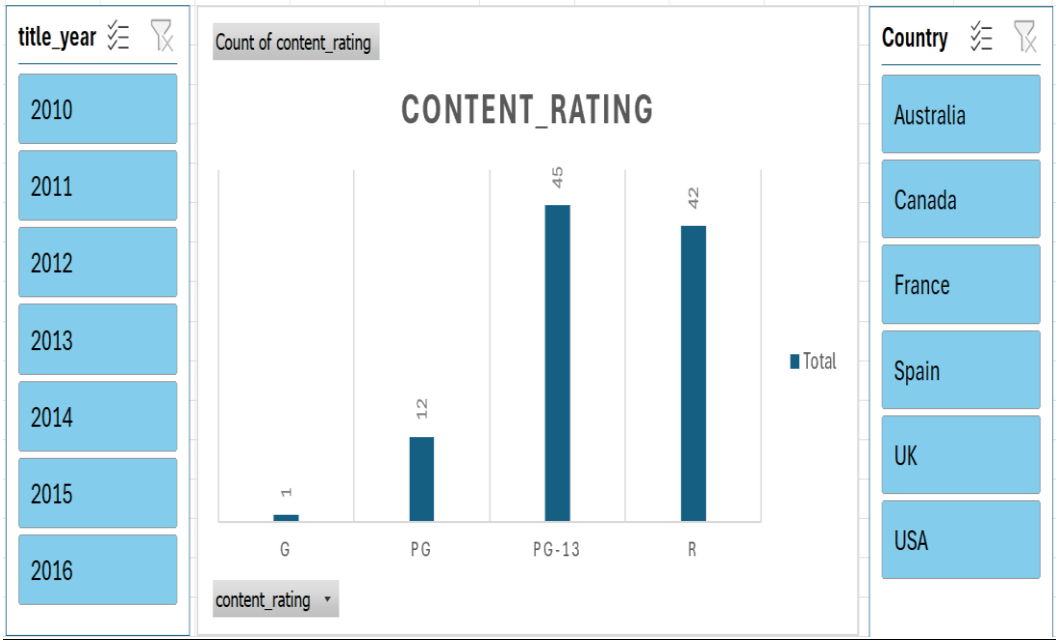




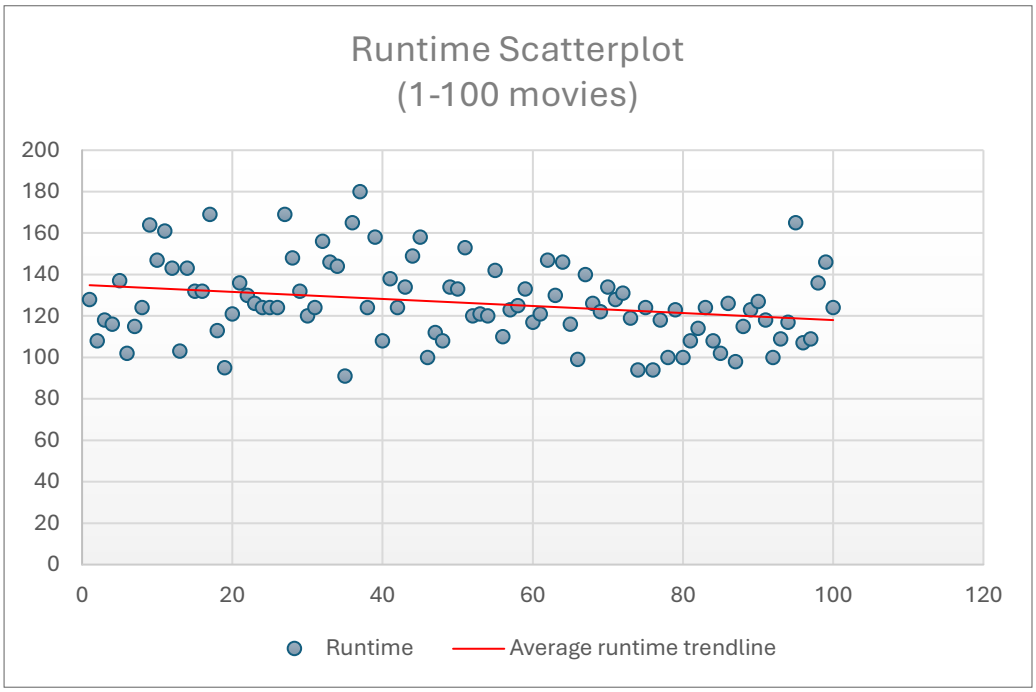
8. C Votes Analysis :

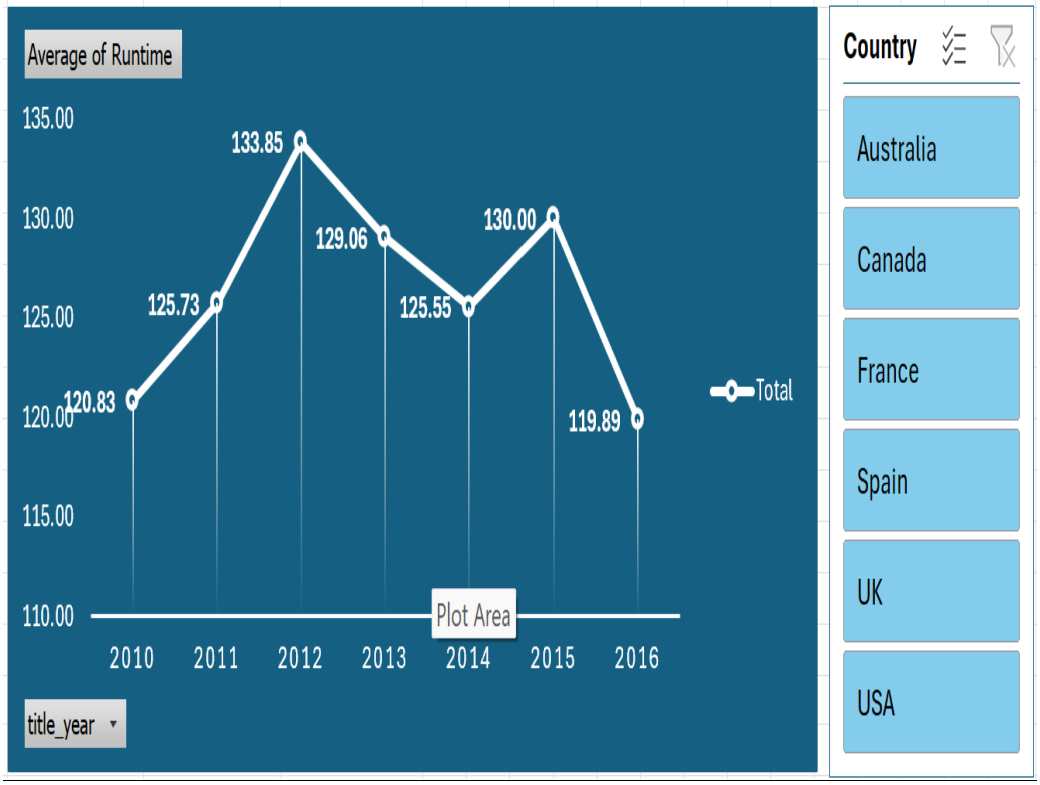


9. Content Rating Analysis :



10. Runtime analysis :





11. Average numbers of C Votes :

Average number of votes from different slots by year wise analyis with country slicers

Values	2010	2011	2012	2013	2014	2015	2016	Grand Total
Average of CVotesUS	73953	58784	***79540***	45782	54490	41600	37973	57113
Average of CVotesnUS	240005	185336	***255066***	170804	203859	147733	132028	195674
Average of CVotesMale	325484	257229	***381314***	246686	309388	224861	201345	284973
Average of CVotesFemale	85610	60363	***94621***	59319	68746	42835	44302	67117
Average of CVotesU18	1713	917	2145	1428	***2584***	1931	2387	1903
Average of CVotesU18M	1241	697	1580	1029	***1906***	1553	1826	1420
Average of CVotesU18F	458	210	547	387	***658***	362	544	467
Average of CVotes1829	208880	150434	***239898***	150272	190381	125576	119401	174416
Average of CVotes1829M	157301	116997	***184445***	116522	149822	102064	94051	135421
Average of CVotes1000	***706***	675	704	603	626	560	515	633
Average of CVotes45AF	4672	4320	***5745***	4051	4095	3223	2715	4193
Average of CVotes45AM	21856	19220	***25901***	18492	21138	17796	14364	20189
Average of CVotes45A	26964	23953	***32207***	22940	25704	21412	17404	24817
Average of CVotes3044F	26048	20596	***27609***	17715	19335	13278	12410	20067
Average of CVotes3044M	128698	108089	***141436***	92414	111710	85701	72189	108069
Average of CVotes3044	156719	130448	***171516***	111709	132987	100491	85872	129959
Average of CVotes1829F	49650	31972	***53059***	32257	38668	22173	24076	37270

Country: Australia, Canada, France, Spain, UK, USA

*** maximum average (slot wise) are conditional formatting with green color & **bold** fonts

12. Average rating of C Votes :

Average ratings from different slots by year wise analysis with country slicers

Values	2010	2011	2012	2013	2014	2015	2016	Grand Total
Average of VotesM	7.9	7.8	7.9	7.8	**7.9**	7.9	7.9	7.9
Average of VotesF	7.9	7.8	**8.0**	7.9	8.0	7.9	7.9	7.9
Average of VotesU18	8.1	8.1	8.2	8.2	8.4	8.2	**8.4**	8.2
Average of VotesU18M	8.1	8.1	8.1	8.2	8.3	8.2	**8.4**	8.2
Average of VotesU18F	8.1	8.1	8.2	8.2	**8.4**	8.1	8.4	8.2
Average of Votes1829	8.0	7.9	8.1	8.0	**8.1**	8.0	8.0	8.0
Average of Votes1829M	8.0	7.9	8.1	8.0	**8.1**	8.0	8.0	8.0
Average of Votes1829F	8.0	7.9	**8.1**	8.0	8.0	7.9	8.0	8.0
Average of Votes3044	7.8	7.7	7.8	7.7	**7.8**	7.8	7.7	7.7
Average of Votes3044M	7.8	7.7	7.8	7.7	**7.8**	7.8	7.7	7.7
Average of Votes3044F	7.8	7.7	**7.8**	7.8	7.8	7.7	7.8	7.8
Average of Votes45A	7.6	7.6	7.7	7.6	7.7	**7.7**	7.6	7.6
Average of Votes45AM	7.6	7.6	7.6	7.6	7.7	**7.7**	7.6	7.6
Average of Votes45AF	7.7	7.7	7.8	7.8	7.7	**7.9**	7.8	7.8
Average of Votes1000	7.3	7.2	7.2	7.2	**7.4**	7.4	7.3	7.3
Average of VotesUS	8.0	7.9	8.0	7.9	**8.0**	8.0	8.0	8.0
Average of VotesnUS	7.8	7.7	7.8	7.8	**7.9**	7.8	7.8	7.8

Country: Australia, Canada, France, Spain, UK, USA

*** maximum average (slot wise) are conditional formatting with Orange color & **bold** fonts

Results :

With the help of this project, I have gained valuable experience for data analysis using statistical knowledge and excel's data visualization . Through this, I have learnt to apply my data analysis skills and knowledge in solving real life problems.

THANK YOU