IMDB MOVIE ANALYSIS

Project Description:

This project involves analyzing an IMDB dataset provided by the trainity team. The dataset consists of movies from 1920 to 2010 and contains various information such as movie titles, ratings, actors, directors, budget, and collection. The goal of the project is to clean the dataset using Microsoft Excel and extract valuable insights by answering specific questions. Once the dataset is cleaned, various analysis can be performed using Excel's various functions and tools. For example, we can calculate average ratings, identify the highest-grossing movies, determine the most popular actors or directors, and examine the financial aspects of the movies.

Overall, this project aims to utilize Microsoft Excel as a tool to clean and analyze the IMDB dataset, providing valuable insights into movie ratings, director and actor profiles, and financial aspects of movies from 1920 to 2010.

Approach:

For this project, first we'll have to understand the given data. After that we'll have to clean the data as per our requirements like removing duplicate and null values, delete unnecessary column which will never be used etc.

When our data will clean we'll use pivot table, some mathematical and statistical function and create charts for easy representation.

Tech-Stack Used:

MS Excel – I used this tool because this tool is used to create graphical representation of the result and understand the result set better.

It also allows us to analyze large amount of data quickly and easily with less efforts than other tools.

Insights:

A. Clean the data

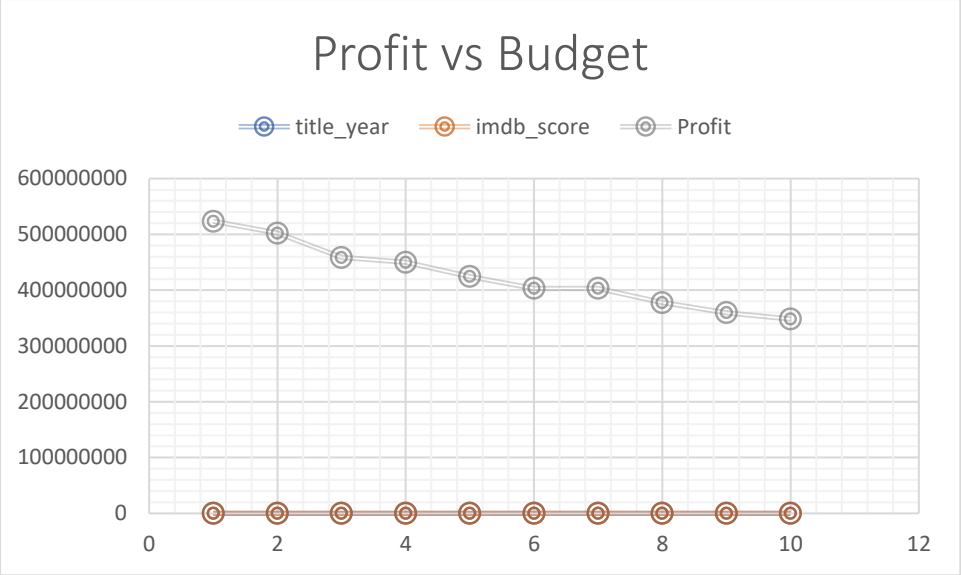
1. Remove unnecessary columns
   Like:
   (Color, director_facebook_likes, actor_3_facebook_likes,
   actor_2_name, actor_1_facebook_likes,
   cast_total_facebook_likes, actor_3_name, facenumber_in_poster,
   plot_keywords, movie_imdb_link, content_rating,
   actor_2_facebook_likes, aspect_ratio, movie_facebook_likes.)

2. Remove blank cell/ Null value
3. Remove Duplicate
   Here the link is provided for the clean data here:
   clean_data.xlsv

B. Find the movies with the highest profit?

Top 10 highest profit movie

director_name	actor_1_name	movie_title	title_year	imdb_score	Profit
James Cameron	CCH Pounder	Avatar	2009	7.9	523505847
Colin Trevorrow	Bryce Dallas Howard	Jurassic World	2015	7	502177271
James Cameron	Leonardo DiCaprio	Titanic	1997	7.7	458672302
George Lucas	Harrison Ford	Star Wars: Episode IV - A New Hope	1977	8.7	449935665
Steven Spielberg	Henry Thomas	E.T. the Extra-Terrestrial	1982	7.9	424449459
Joss Whedon	Chris Hemsworth	The Avengers	2012	8.1	403279547
Joss Whedon	Chris Hemsworth	The Avengers	2012	8.1	403279547
Roger Allers	Matthew Broderick	The Lion King	1994	8.5	377783777
George Lucas	Natalie Portman	Star Wars: Episode I - The Phantom Menace	1999	6.5	359544677
Christopher Nolan	Christian Bale	The Dark Knight	2008	9	348316061



C. Find IMDB Top 250.

1. First filter the data of num_voted_users > 25,000.
2. Sort the data of imdb_score in descending order.
3. Select 250 entry columns.
4. Give Rank of the IMDB using =row() function

Link is provided here for top 250 movies:

top_250_movie.xlsv

5. Filter the language by unselecting the English language and the data will show having the foreign language movies in Top 250s list

Link for Top Foreign Language Movie:

top_foreign_lang.xlsv

D. Find the best directors.

top_10_directors	Average of imdb_score
Charles Chaplin	8.6
Tony Kaye	8.6
Alfred Hitchcock	8.5
Damien Chazelle	8.5
Majid Majidi	8.5
Ron Fricke	8.5
Sergio Leone	8.433333333
Christopher Nolan	8.425

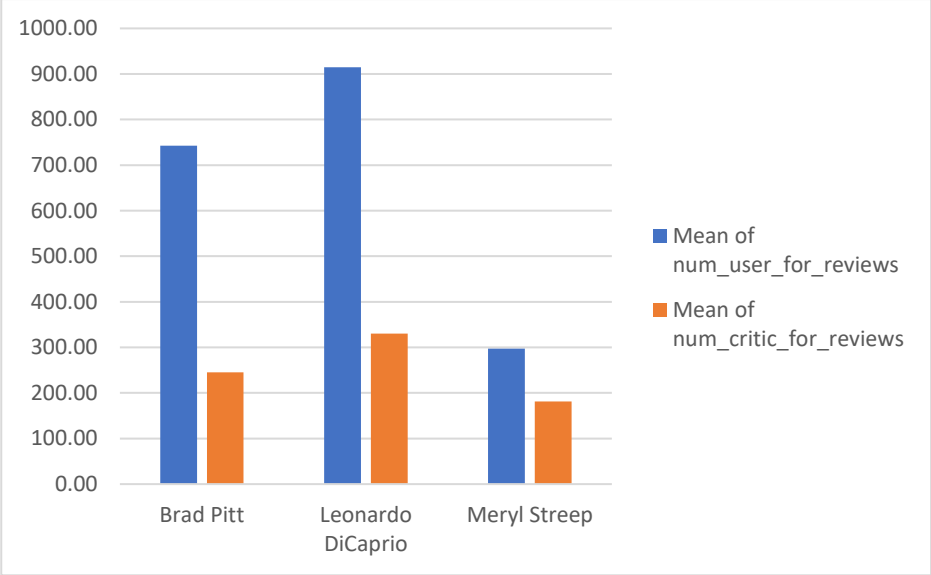
Richard Marquand	8.4
S.S. Rajamouli	8.4

E. Find the popular genres.

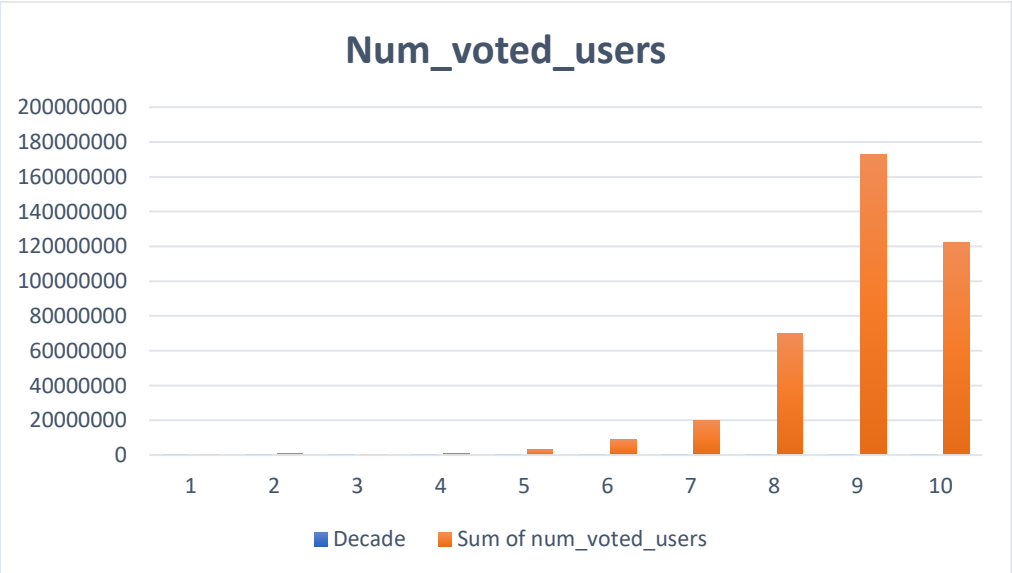
top_10_genres	Count of genres
Drama	154
Comedy|Drama|Romance	151
Comedy|Drama	148
Comedy	147
Comedy|Romance	136
Drama|Romance	120
Crime|Drama|Thriller	83
Action|Crime|Thriller	56
Action|Crime|Drama|Thriller	50
Action|Adventure|Sci-Fi	48

F. Find the critic-favorite and audience-favorite

actor 1 name	Mean of num_user_for_reviews	Mean of num_critic_for_reviews
Brad Pitt	742.35	245.00
Leonardo DiCaprio	914.48	330.19
Meryl Streep	297.18	181.45
Grand Total	**716.18**	**267.24**



Decade	Sum of num_voted_users
1920	116392
1930	804839
1940	230838
1950	678336
1960	2985581
1970	8704723
1980	20101705
1990	70090204
2000	173033966
2010	122492496



Results

Through this IMDB Movie Analysis project, I have developed a diverse range of logical, statistical, and technical skills to extract the desired insights from the dataset. Utilizing statistical techniques and leveraging the capabilities of MS Excel, I have significantly expedited data analytics tasks and streamlined complex calculations. This has enabled me to transform the data and derive meaningful conclusions more efficiently.

I also get how visual representation of data make it very simple to understandable. I gain knowledge on when to utilize each visualizations graph or chart based on the data and desired results.