

Project 05

IMDB Movie Analysis

Final Project-1

NAME :- NIRAJ INGOLE

INTERNSHIP PROJECT

USING : Microsoft Excel

Description:

For your Final Project, we are providing you with dataset having various columns of different IMDB Movies. You are required to Frame the problem. For this task, you will need to define a problem you want to shed some light on.

Once you have defined a problem, clean the data as necessary, and use your Data Analysis skills to explore the data set and derive insights.

How to do this Project?

Download the dataset: You are supposed to download the dataset and perform the analysis using Excel or Google sheets.

Perform Analysis: Use Excel or Google Sheets to perform your entire analysis answering the questions asked above.

Submit a Report: Make a report (PDF/PPT) to be presented to the leadership team.

You are required to provide a detailed report for the below data record mentioning the answers of the questions that follows:

1. **Cleaning the data**:: PThis is one of the most important step to perform before moving forward with the analysis. Use your knowledge learned till now to do this. (Dropping columns, removing null values, etc.)

Your Task: Clean the data

- 1.Dropping unnecessary columns
2. Remove blank cell/na value
3. Remove duplicate
- 4.Show top 10 rating
- 5.handle missing value
- 6.convert data
- 7.merge data sets

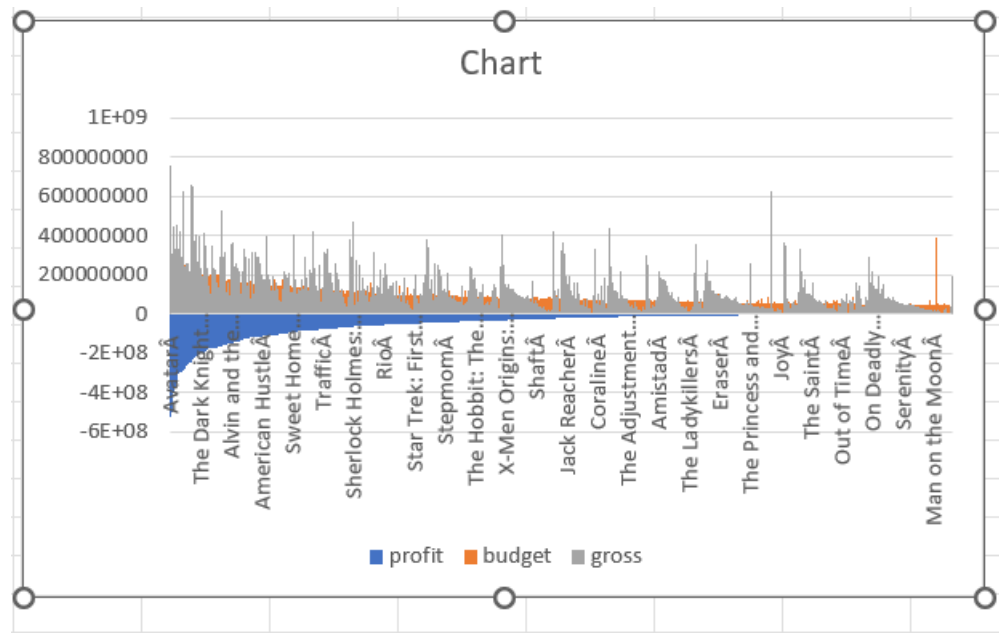
Cleaned data:-

<https://drive.google.com/file/d/1N8OKg6xP3N6xMu88V4zdyKyW46DkoFVQ/view?usp=drivesdk>

2. **Movies with highest profit**: Create a new column called profit which contains the difference of the two columns: gross and budget. Sort the column using the profit column as reference. Plot profit (y-axis) vs budget (x- axis) and observe the outliers using the appropriate chart type.

Your Task: Find the movies with the highest profit?

1. Create a new column called profit
2. . Sort the column using the profit column as reference
3. Plot profit (y-axis) vs budget (x- axis)



Cleaned data:-

<https://drive.google.com/file/d/1N8evpKrVjVHYOacaYbNa-mmT8WPTNe7o/view?usp=drivesdk>

- 3. Top 250:** Create a new column IMDb_Top_250 and store the top 250 movies with the highest IMDb Rating (corresponding to the column: imdb_score). Also make sure that for all of these movies, the num_voted_users is greater than 25,000. Also add a Rank column containing the values 1 to 250 indicating the ranks of the corresponding films. Extract all the movies in the IMDb_Top_250 column which are not in the English language and store them in a new column named Top_Foreign_Lang_Film. You can use your own imagination also!

Your task: Find IMDB Top 250

1. Create column IMDb_Top_250
2. Sort data using imdb_score in descending order
3. Use 1st 250 entry for analysis
4. Then rank them
5. Find out without English top moves

Cleaned data:-

Create a new column IMDb_Top_250

=<https://drive.google.com/file/d/1NDgUVx0jzLfA7LiiU99x4jmZBjmlxRDv/view?usp=drivesdk>

not in the English language =https://drive.google.com/file/d/1NHe62uhcGQMAYIhMYEMiXk6m-EwRszR_/view?usp=drivesdk

4. Best Directors: TGroup the column using the director_name column.

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.

Your task : Find the best directors

Cleaned data:-

<https://drive.google.com/file/d/1NI7Syquno5AXoJobBGS3b-8c5s6qOFEm/view?usp=drivesdk>

5. Popular Genres: Perform this step using the knowledge gained while performing previous steps.

Your task: Find popular genres

Cleaned data:- <https://drive.google.com/file/d/1NI91-yYNY1L-0P6l4aWD9VHTzZc7bREX/view?usp=drivesdk>

6. Charts: Create three new columns namely, Meryl_Streep, Leo_Caprio, and Brad_Pitt which contain the movies in which the actors: 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' are the lead actors. Use only the actor_1_name column for extraction. Also, make sure that you use the names 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' for the said extraction. Append the rows of all these columns and store them in a new column named Combined. Group the combined column using the actor_1_name column. Find the mean of the num_critic_for_reviews and num_users_for_review and identify the actors which have the highest mean. Observe the change in number of voted users over decades using a bar chart. Create a column called decade which represents the decade to which every movie belongs to. For example, the title_year year 1923, 1925 should be stored as 1920s. Sort the column based on the column decade, group it by decade and find the sum of users voted in each decade. Store this in a new data frame called df_by_decade.

Your task: Find the critic-favorite and audience-favorite actors

Cleaned data:-

mean of 3 actore

https://drive.google.com/file/d/1NLByjmg8NSAieTLj1adSMae99DZC744_/view?usp=drivesdk

no.of voted user

<https://drive.google.com/file/d/1NMGzBNUjGJs9A3EjleRGremGPqW0Vssl/view?usp=drivesdk>

RESULT:

In this project of IMBD Movie Analysis, I have gained various Logical, Statistics and Technical Skill for get desired answers from the dataset. Average, Frequency Table, and Discovering Outliers are statistics concepts that help me better connect with data, offer me a thorough understanding of it, and aid in the analytics of supplied data .My ability to apply statistics and Microsoft Excel's technical capabilities to analyse data. It speeds up data analytics tasks considerably. Simplifies difficult and time-consuming calculations. I also get a sense of how visual presentation of data makes it very simple to understand through its data visualisation functionality. I gain knowledge on when to utilise each visualisation graph or chart based on the data and desired results.