IMDB Movie Analysis

Final Project-1

Project Description :

For our Final Project, we are provided with a dataset having various columns of different IMDB Movies. We are required to frame the problem. For this task, we need to define a problem we want to shed some light on.

1. Cleaning the data : This is one of the most important steps to perform before moving forward with the analysis. We have to clean the data.

2. Movies with highest profit: We have to find the movies with the highest profit.

3. Top 250: We have to find IMDB Top 250.

4. Best Directors: We have to find the best directors.

5. Popular Genres: We have to find popular genres.

6. Charts: We have to find the critic-favorite and audience-favorite actors.

Approach :

We plan to filter, summarize & visualize data with Excel on given database to create insights for the teams to make data driven decision. We got all the following answers to the questions.

1. We cleaned the data.

2. We found the movies with the highest profit.

3. We found IMDB Top 250.

4. We found the best directors.

5. We found popular genres.

6. We found the critic-favourite and audience-favourite actors

Tech-Stack Used :

Excel by Microsoft Corporation – For cleaning, extracting & manipulating the data

WordPad by Microsoft Corporation – For creating the project report

Insights :

1. Data cleaning :

a. Removed Duplicates .

b. Removed "Â" from the dataset .

c. Dropped irrelevant columns such as color, director_facebook_likes, actor_1_facebook_likes, actor_2_facebook_likes, actor_3_facebook_likes, actor_2_name, cast_total_facebook_likes, actor_3_name, duration, facenumber_in_poster, content_rating, country, movie_imdb_link, aspect_ratio, plot_keywords .

d. Found and replaced "Nan" values with "English" .

e. Removed all the Null values from specific columns.

2. Movies with the highest profit :

a. Created a column "Profit" and found profit value by subtracting Gross from Budget Sorted the movies with Sort and Filter Following are the movies with their profits:

Avatar - $523505847

Jurassic World - $502177271

Titanic - $458672302

Star Wars: Episode IV - A New Hope - $449935665

E.T. the Extra-Terrestrial - $424449459

3. IMDB top 250 :

We found IMDB Top 250

(Please refer sheet 2 of IMDB_Movies.csv file)

4. Best Directors :

We found the Best Directors by ratings as follows:

John Blanchard – 9.5

Frank Darabont – 9.3

Francis Ford Coppola – 9.2

John Stockwell – 9.1

Christopher Nolan - 9

Francis Ford Coppola - 9

Peter Jackson – 8.9

Sergio Leone – 8.9

Steven Spielberg – 8.9

Quentin Tarantino – 8.9

5. Popular genres :

We found popular Genres by ratings as follows:

Comedy - 9.5

Crime | Drama - 9.3

Crime | Drama - 9.2

Drama - 9.1

Drama - 9.1

Action - 9.1

Action | Crime | Drama | Thriller - 9

Crime | Drama - 9

Crime | Drama | Thriller - 9

Action | Adventure | Drama | Fantasy - 8.9

6. critic-favourite and audience-favourite actors :

We found Critic-favourite and Audience-favourite actor:

Meryl Streep Critic mean	163.1538
Meryl Streep user mean	257.3077
Leonardo DiCaprio Critic mean	330.1905
Leonardo DiCaprio user mean	914.4762
Brad Pitt Critic mean	231.9444
Brad Pitt user mean	702.4444

Based on Critics and audience Best Actor is Leonardo DiCaprio .

Result:

We have run all the above-mentioned Excel functions and got answers of the questions which will help us to take data driven decisions.
Also attached IMDB_Movies.csv file including all data analysis sheets for the reference.