IMDB MOVIE ANALYSIS

Name: Kaila Sai Pranava Karthik

PROJECT DESCRIPTION

This project involved a comprehensive analysis of data extracted from IMDb, focusing on identifying trends and patterns among movies, actors, and directors. Data manipulation and visualization were primarily conducted using Excel, enabling the effective presentation of insights.

APPROACH

I have used the below approach to complete this project.

1. Gathering the information from the discerption to complete the tasks.
2. Excel is used.
3. Clean the datasets
4. Process the data to answer the asked questions
5. Use formulas, filters and other functions for finding insights
6. Create charts and graphs for easy and meaningful data representation
7. Made pdf file for insights.

TECH STACK

Excel is used for doing EDA as well as for graphs.

Microsoft word to make report and publish to leadership.

INSIGHTS

Cleaning the data:

This is one of the most important steps to perform before moving forward with the analysis.

The following steps has been taken to clean data

- Format values in the column for better view.
- Delete Duplicate Data.
- Drop Null values.
- Delete columns which are not related.
- Removed special character from movie_title

Movies with highest profit:

To find the movies with the highest profit. I used the gross and budget columns to calculate the profit for each movie, then I arranged the profit in descending order and extracted the top ten highest profit movies.

movie_title	gross	budget	Profit
Avatar	760505847	237000000	523505847
Jurassic World	652177271	150000000	502177271
Titanic	658672302	200000000	458672302
Star Wars: Episode IV - A New Hope	460935665	11000000	449935665
E.T. the Extra-Terrestrial	434949459	10500000	424449459
The Avengers	623279547	220000000	403279547
The Lion King	422783777	45000000	377783777
Star Wars: Episode I - The Phantom Menace	474544677	115000000	359544677
The Dark Knight	533316061	185000000	348316061
The Hunger Games	407999255	78000000	329999255
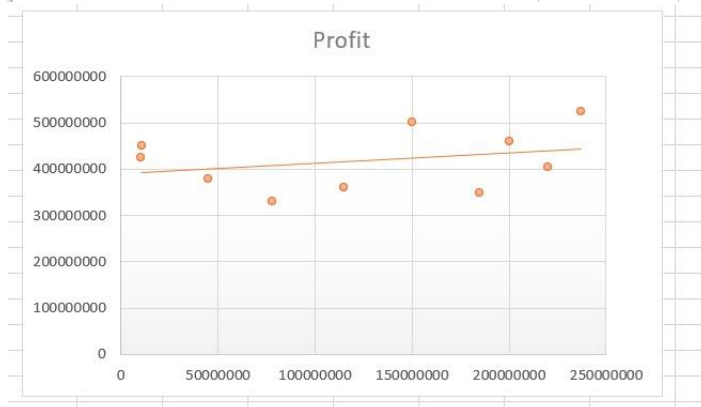
Top 250:

In this section, I attempted to find the IMDB top 250 movies based on IMDB ratings with more than 25000 votes. For this task, I filtered the No. of voted users greater than 25000 and then arranged the IMDB ratings in descending order and extracted the top 250 movies regardless of language.

In the following section, we have filtered the foreign language films that were in the top 250 on IMDB.

Results has been added in the Sheet C1 and C2.

Best Directors:

In this section, we have ranked the top ten directors based on their IMDB ratings. I ranked the directors in ascending order and then performed a subtotal on the groups to get their average ratings for all of their films, and then ranked the average ratings in descending order to get the top ten directors.

director_name	imdb_score
Akira Kurosawa Average	8.7
Charles Chaplin Average	8.6
Tony Kaye Average	8.6
Alfred Hitchcock Average	8.5
Damien Chazelle Average	8.5
Majid Majidi Average	8.5
Ron Fricke Average	8.5
Sergio Leone Average	8.433333333
Christopher Nolan Average	8.425
Asghar Farhadi Average	8.4

Results has been added in the Sheet D1 and D2.

Popular Genres:

To locate the most popular genres I assumed that the higher the rating, the more popular the genre, and with the help of the subtotal function, we calculated the total for genre from each movie and discovered the three most popular genera as shown below.

genres	num_voted_users
Action Total	175308173
Comedy Total	84226515
Drama Total	41280002

Charts

We calculated the average of one 's reviews to find the most popular lead actor among Brad Pitt, Leonardo DiCaprio, and Meryl Streep. Leonardo Di Capri is the winner here.

actor_1_name	num_critic_for_reviews
Brad Pitt Average	245
Leonardo DiCaprio Average	330.1904762
Meryl Streep Average	181.4545455
Grand Average	267.244898

actor_1_name	num_user_for_reviews
Brad Pitt Average	742.3529412
Leonardo DiCaprio Average	914.4761905
Meryl Streep Average	297.1818182

From Both the perspective Leonardo DiCaprio is the popular lead actor.

Detail information has been added into the sheet F.

RESULT

We've thoroughly cleaned and analysed the data we gathered, diving deep into different layers to uncover the answers we needed. This project taught me a lot about how to sift through data effectively, craft the right queries, and execute them to get the results we were looking for.