IMDB Movie Analysis

Final Project-1

susrithaj2102@gmail.com

TOPICS TO BE COVERED...

1. PROJECT DESCRIPTION
2. APPROACH
3. INSIGHTS
4. RESULTS
5. LINKS

Project Description

❖ IMBD is well known movie and series rating site for users and critics worldwide. In IMBD we can find movies or series' rating as well as its directors and actors compiled profile as well as financials of it. Here we're provided with IMBD's dataset for movies. Which contains information about movie, its actors, directors, budget, collection, etc. We're going to clean the dataset and get answers for required questions.

❖ This project aimed to find insights on IMDB movies including the
  o movie genre analysis.
  o movie duration analysis.
  o language analysis.
  o director analysis.
  o budget analysis.

Approach

✓ Download the dataset **(28 columns and 5027 rows)**and open in excel for analysis.
✓ Removed empty rows, special characters, and duplicate values and columns that were not relevant in finding insights.
✓ After cleaning dataset contains **12 columns and 3787 rows**.
✓ Process the data to answer the asked questions.
✓ Use formulas, filters, pivot tables and other functions for finding insights.
✓ Create charts and graphs for easy and meaningful data representation.
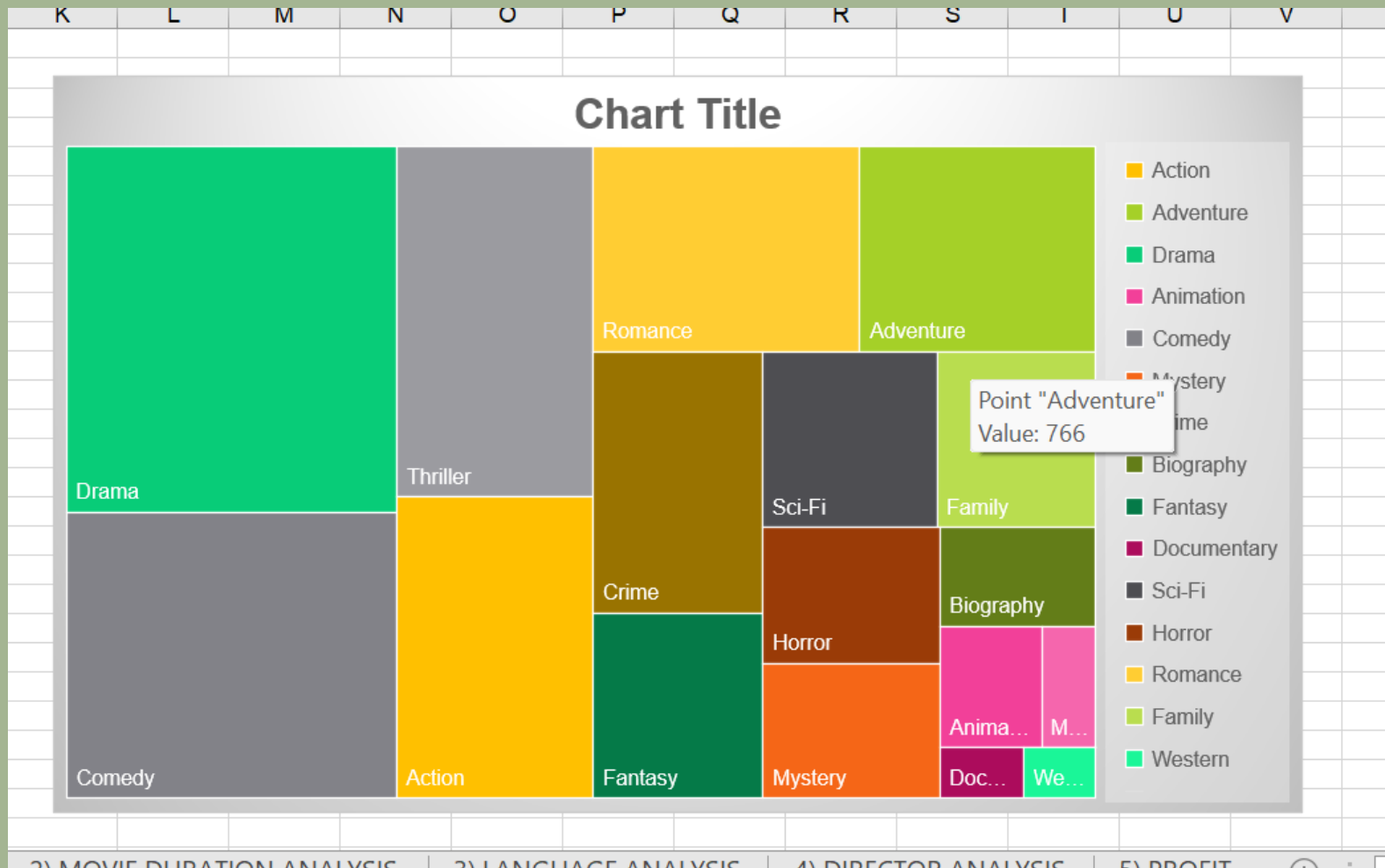✓ Create report and submit the project.

Tech Stacks Used

Insights

1) MOVIE GENRE ANALYSIS:-

GENRE	No_of_movies	Mean_imdb	Median_imdb	Mode_imdb	Max_imdb	Min_imdb	StdDev_imdb	Var_imdb
Action	935	6.2859893	6.3	6.6	9	2.1	1.038357736	1.07818679
Adventure	766	6.45496084	6.6	6.6	8.9	2.3	1.116926308	1.24752438
Drama	1911	6.78911565	6.9	6.7	9.3	2.1	0.891064898	0.79399665
Animation	197	6.70050761	6.8	7.3	8.6	2.8	0.993627525	0.98729566
Comedy	1492	6.18331099	6.3	6.3	8.8	1.9	1.039919012	1.08143155
Mystery	377	6.46949602	6.5	6.6	8.6	3.1	1.007391835	1.01483831
Crime	702	6.54814815	6.6	6.6	9.3	2.4	0.984105199	0.96846304
Biography	242	7.14008264	7.2	7	8.9	4.5	0.71009671	0.50423734
Fantasy	496	6.28508065	6.4	6.7	8.9	2.2	1.140414241	1.30054464
Documentary	67	7.0119403	7.2	6.6	8.5	1.6	1.199939694	1.43985527
Sci-Fi	484	6.32727273	6.4	7	8.8	1.9	1.16718415	1.36231884
Horror	379	5.90395778	5.9	6.2	8.6	2.3	0.991023285	0.98212715
Romance	866	6.42621247	6.5	6.5	8.5	2.1	0.968996249	0.93895373
Family	441	6.2	6.3	5.4	8.6	1.9	1.169576458	1.36790909
Western	58	6.76551724	6.8	6.8	8.9	4.1	0.998516746	0.99703569
Musical	102	6.55098039	6.7	7.1	8.5	2.1	1.143535	1.3076723
Thriller	1087	6.37230911	6.4	6.5	9	2.7	0.969078327	0.9391128

Chart Title

Legend:
- Action
- Adventure
- Drama
- Animation
- Comedy
- Mystery
- Crime
- Biography
- Fantasy
- Documentary
- Sci-Fi
- Horror
- Romance
- Family
- Western

Chart labels: Drama, Thriller, Comedy, Action, Romance, Crime, Fantasy, Sci-Fi, Horror, Mystery, Adventure, Family, Biography, Anima..., M..., Doc..., We...

Tooltip: Point "Adventure" Value: 766

Operations performed on duration of movies

2.Movie duration analysis

Operations	Values
Mean	109.808505
Median	105
Mode	101
Standard Devation	22.763201
Variance	518.16332

$R^2 = 0.4059$

- imdb_score

— Trendline for imdb_score
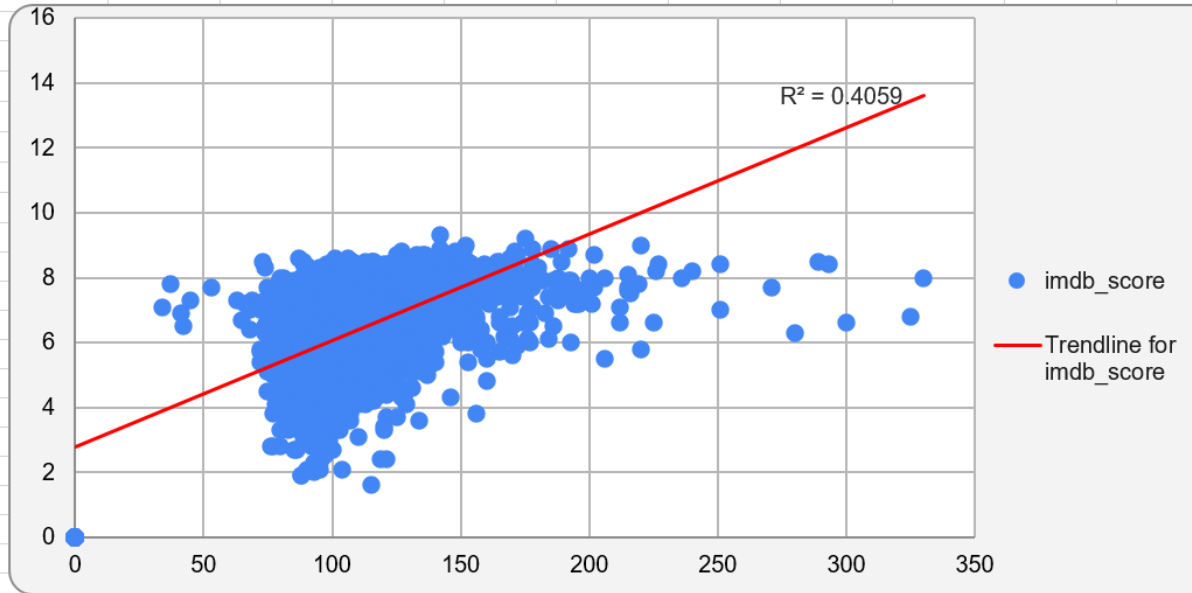
3) LANGUAGE ANALYSIS:-

Language	No_of_movies
English	3606
Mandarin	14
Aboriginal	2
Spanish	26
French	37
Filipino	1
Maya	1
Kazakh	1
Telugu	1
Cantonese	8
Japanese	12
Aramaic	1
Italian	7
Dutch	3
Dari	2
German	13
Mongolian	1
Thai	3
Bosnian	1
Korean	5

Most commonly used language in movies is English with count of 3606.

Entire language analysis data is provide in dataset attached to this report.

4) DIRECTOR ANALYSIS:-

Director	Average_imdb	percentile	Count_movies
James Cameron	7.914285714	0.976	7
Gore Verbinski	6.985714286	0.722	7
Sam Mendes	7.457142857	0.891	7
Christopher Nolan	8.425	0.995	8
Andrew Stanton	7.733333333	0.953	3
Sam Raimi	6.96	0.718	10
Nathan Greno	7.8	0.958	1
Joss Whedon	7.866666667	0.969	3
David Yates	7.2	0.812	3
Zack Snyder	7.142857143	0.804	7
Bryan Singer	7.2875	0.849	8
Marc Forster	7.228571429	0.841	7
Andrew Adamson	7.15	0.805	4
Rob Marshall	6.6	0.553	5
Barry Sonnenfeld	6.457142857	0.5	7
Peter Jackson	7.888888889	0.969	9
Marc Webb	7.133333333	0.801	3
Ridley Scott	7.13125	0.8	16
Chris Weitz	6.08	0.348	5
Anthony Russo	7	0.723	4
Peter Berg	6.666666667	0.592	6

IMDB Movies | 1) MOVIE GENRE ANALYSIS | 2) MOVIE DURATION ANALYSIS | 3) LANGUAGE

Top directors based on their average IMDB score and also calculated and analyzed their contribution to the success of movies using percentile calculations.

movies with the highest profit

5) PROFIT ANALYSIS:-

Movies	Profits in Millions
AvatarÂ	523505847
Jurassic WorldÂ	502177271
TitanicÂ	458672302
Star Wars: Episode IV - A New HopeÂ	449935665
E.T. the Extra-TerrestrialÂ	424449459
The Lion KingÂ	377783777
The Jungle BookÂ	375290282
Star Wars: Episode I - The Phantom MenaceÂ	359544677
The Dark KnightÂ	348316061
The Twilight Saga: Breaking Dawn - Part 2Â	344597846

Chart Title

AvatarÂ	523505847
Jurassic WorldÂ	502177271
TitanicÂ	458672302
Star Wars: Episode IV - A Ne...	449935665
E.T. the Extra-TerrestrialÂ	424449459
The Lion KingÂ	377783777
The Jungle BookÂ	375290282
Star Wars: Episode I - The...	359544677
The Dark KnightÂ	348316061
The Twilight Saga: Breaking...	344597846

Movie_Title	Profit
Avatar	523505847
Pirates of the Caribbean: At World's End	9404152
Spectre	-44925825
The Dark Knight Rises	198130642
John Carter	-190641321
Spider-Man 3	78530303
Tangled	-59192738
Avengers: Age of Ultron	208991599
Harry Potter and the Half-Blood Prince	51956980
Batman v Superman: Dawn of Justice	80249062
Superman Returns	-8930592
Quantum of Solace	-31631573
Pirates of the Caribbean: Dead Man's Chest	198032628
The Lone Ranger	-125710090
Man of Steel	66021565
The Chronicles of Narnia: Prince Caspian	-83385977
The Avengers	403279547
Pirates of the Caribbean: On Stranger Tides	-8936125
Men in Black 3	-45979146
The Hobbit: The Battle of the Five Armies	5108370

Results

1. Data Exploration and Cleansing:

- Scrutinize and clean the raw IMDB movie dataset to ensure accuracy and consistency.

2. Descriptive Statistics:

  - Utilize Excel's statistical functions to generate descriptive statistics for Language Analysis.

  - Employ pivot tables and charts to visually represent trends and variations within the dataset.

3. Genre Analysis:

  - Categorize movies based on genres and analyze the popularity ,performance of different genres over time.

4. Correlation Analysis:

  - Investigate relationships between variables, such as budget and gross earnings, using correlation analysis.

5. Top director Analysis:

  - Identify and analyze movies based on influence of directors using Excel's PERCENTILE function.

Contributions to Understanding IMDB Movie Analysis:

Through the diligent application of Excel's analytical tools and functions, this project has made substantial contributions to my understanding of IMDB movie data.

The knowledge gained include:

  - Data Analytics.

  - Advanced Excel Understanding.

  - Statistics Understanding

Link for Dataset

https://docs.google.com/spreadsheets/d/1SiuiDuAWiRJ_F6eY4jUrZp2ZBrV qZ_DJ/edit?usp=sharing&ouid=104873468157700933746&rtpof=true&sd =true

Thank you for your time 😊

Name: J.Susritha
Email: susrithaj2102@gmail.com