IMDB Analysis

Name - Samiksha Zagade

Project Description

The dataset is related to IMDB Movies. A potential problem to investigate could be: "What factors influence the success of a movie on IMDB?" Here, success can be defined by high IMDB ratings or IMDB score that is assigned to each movie. The impact of this problem is significant for movie producers, directors, and investors who want to understand what makes a movie successful to make informed decisions in their future projects.

Approach and Tech Used

The approach here to first clean the data that includes handling the null values, handling outliers and understanding the data. After the data is cleaned and understood I have tried answering the questions to provide insights that can drive decision-making. The tech used here is excel. Excel is a versatile tool for data analysis due to its user-friendly interface and widespread accessibility. Its array of functions and features allows for quick data manipulation, facilitating easy organization and visualization. For straightforward analyses or initial data exploration, Excel's pivot tables, charts, and formulas provide a solid foundation. While it may have limitations for handling large datasets or complex statistical analyses compared to specialized software, Excel remains an efficient and practical tool for basic data analysis and quick insights.

Results

- The data has 3752 rows and 28 columns.
- **Movie Genre Analysis –**
    - I have analysed the distribution of movie genres and their impact on the IMDB score.

- The most common genres of movies in the dataset are Comedy, Action, Drama, Adventure and Crime with count to be 988, 957, 668, 368 and 255 respectively.
- Comedy has average of 6.2 imdb rating.
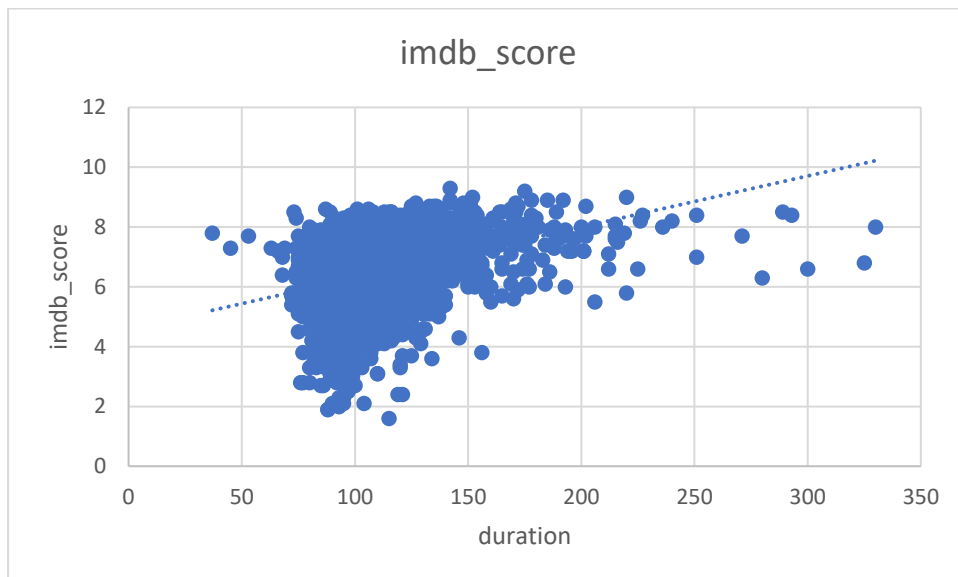- Then, for the first three mot common genre, I have calculated descriptive statistics as shown below.

Comedy			Action			DRAMA	
Mean	6.16842		Mean	6.29112		Mean	6.83159
Standard Error	0.03286		Standard Error	0.0336		Standard Error	0.03502
Median	6.3		Median	6.4		Median	6.9
Mode	6.4		Mode	6.6		Mode	6.7
Standard Deviation	1.03275		Standard Deviation	1.03954		Standard Deviation	0.90519
Sample Variance	1.06658		Sample Variance	1.08064		Sample Variance	0.81938
Kurtosis	1.78242		Kurtosis	0.84862		Kurtosis	1.75238
Skewness	-0.93076		Skewness	-0.49203		Skewness	-0.94173
Range	6.9		Range	6.9		Range	6.7
Minimum	1.9		Minimum	2.1		Minimum	2.1
Maximum	8.8		Maximum	9		Maximum	8.8
Sum	6094.4		Sum	6020.6		Sum	4563.5
Count	988		Count	957		Count	668

- **Duration Analysis –**
  - Analysed the distribution of movie durations and identified the relationship between movie duration and IMDB score.
  - The descriptive statistics for the duration is shown below where the mean is 110.2 , median is 106 and mode as 101.

Duration	
Mean	110.2579872
Standard Error	0.369523832
Median	106
Mode	101
Standard Deviation	22.64671656
Sample Variance	512.8737711
Kurtosis	12.62739704
Skewness	2.402551754
Range	293
Minimum	37
Maximum	330
Sum	414129
Count	3756

o The relationship between imdb score and duration is that the imdb score is saturated when duration is between 75 and 150 approx.

o Towards the start and the end very few ratings are seen but those ratings are pretty high around 8.

- **Language Analysis-**
  o I determine the most common languages used in movies and analysed their impact on the IMDB score using descriptive statistics.

English			French			Mandarian	
Mean	6.427043		Mean	7.355882		Mean	7.08
Standard Error	0.017514		Standard E	0.089082		Standard E	0.199332
Median	6.5		Median	7.3		Median	7.4
Mode	6.7		Mode	7.2		Mode	7.9
Standard Deviation	1.050538		Standard D	0.519435		Standard D	0.77201
Sample Variance	1.10363		Sample Va	0.269813		Sample Va	0.596
Kurtosis	1.186555		Kurtosis	1.160239		Kurtosis	-0.95414
Skewness	-0.72834		Skewness	-0.49292		Skewness	-0.65778
Range	7.7		Range	2.6		Range	2.3
Minimum	1.6		Minimum	5.8		Minimum	5.6
Maximum	9.3		Maximum	8.4		Maximum	7.9
Sum	23124.5		Sum	250.1		Sum	106.2
Count	3598		Count	34		Count	15

- The most common Language preffered is English with highest count of 3598.

- Compared to English no other language has such a large audience.

- French has count 34 and mandarian as almost half the audience of French.

- The standard deviation of English language is 1.051. The standard deviation shows the variation and spread of the data. Here the standard deviation is less which shows that it is less deviated.

- **Director Analysis –**
  - I have then Identified the top directors based on their average IMDB score and analysed their contribution to the success of movies using percentile calculations.

Row Labels	Average of imdb_score				
Akira Kurosawa	8.7				
Alfred Hitchcock	8.5		Max avg	8.7	
Asghar Farhadi	8.4		Percent rank	1	
Billy Wilder	8.3		Percentile	8.7	
Charles Chaplin	8.6				
Christopher Nolan	8.425				
Damien Chazelle	8.5				
Fritz Lang	8.3				
Lee Unkrich	8.3				
Lenny Abrahamson	8.3				
Majid Majidi	8.5				
Richard Marquand	8.4				
Ron Fricke	8.5				
Sergio Leone	8.433333333				
Tony Kaye	8.6				

o   The percentile shows that success rate is 100%.

o   The top three directors are Akira Kurosawa, Alfred Hitchcock and Asghar Farhadi.

o   The maximum average imdb score is 8.7 director is Akira Kurosawa.

- **Budget Analysis-**
    o   I have then analysed the correlation between movie budgets and gross earnings, and identified the movies with the highest profit margin.

    o   The correlation between the budget and gross is 0.091which shows that the relationship between the two is quite weak though it is positively correlated.

    o   That tells us that the budget and gross earnings have a huge gap and no correlation. The gross earning is either maximum compared to budget in some cases and in some cases the earning might be very less compared to budget.

Insights

- Comedy, Action and Drama are in more demand and common genres.

- Directors can prefer making a movie of Comedy genre as this genre has highest demand by the audience.

- The duration between 75 and 150 has maximum count of ratings ranging from 8 to 2 approx. Highest variation is seen between this duration range.

- The highest viewed movies is of English language. The movies with English language will have a big audience which will increase the reach of the movie to maximum people.

- So making a movie of English language will be very beneficial.

- The directors like Akira Kurosawa, Alfred Hitchcock and Asghar Farhadi have the making average rating for their movies. Thus working with these top three directors has high percent of movie success.

- If these directors make a movie in English then maximum profit will be gained.

Link to the Excel sheet-

[LINK](#)