**TRAINITY ASSIGNMENT**

**IMDB MOVIE ANALYSIS**

**Problem Statement:** The dataset provided 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. 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.

**Data cleaning:**

* To check blanks in each column separately and remove them by using filter.
* Rename the columns(if necessary).
* To remove duplicates.
* To create another sheet with clean data where we keep the numerical columns towards the end.

**TASKS:**

**A. Movie Genre Analysis:** Analyze the distribution of movie genres and their impact on the IMDB score.

* **Task:** Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.

| **Genre** | **Count** |
| --- | --- |
| Drama | 2562 |
| Comedy | 1860 |
| Thriller | 1393 |
| Action | 1139 |
| Romance | 1097 |

**Analysis:** With a mean IMDb score of 6.7, a median of 6.6, and a mode of 6.7, we can infer that the majority of movies fall within the 6-7 range, indicating that audiences generally expect a decent level of quality. However, the differences in range, variance, and standard deviation across genres suggest that audience expectations and reception vary depending on the genre.

**Insights** - Genre specific recommendation.

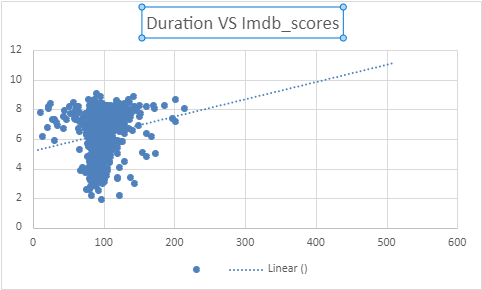
Identify target audience based on the previous history

Since the IMDB score is evenly spread there is an opportunity for every kind.

**B. Movie Duration Analysis:** Analyze the distribution of movie durations and its impact on the IMDB score.

* **Task:** Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.

| Mean | Median | Standarddeviation |
| --- | --- | --- |
| 107.2133253 | 103 | 25.24877521 |



**Analysis:** From the graph we can see a positive correlation, longer movies might generally have higher IMDb scores,

**C. Language Analysis:** Situation: Examine the distribution of movies based on their language.

* **Task:** Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.

The most common languages are English,French,Spanish,Hindi and Mandarin.

**Analysis:** Although there are more movies made in the popular languages , it is evident from descriptive measure that imdb is more for the unpopular language movies. This says that no matter if the movie is made in popular language or not , the movie does well if the content is good.

**D. Director Analysis:** Influence of directors on movie ratings.

* **Task:** Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.

| **director\_name** | **imdb\_score** | **Average\_imdb\_score** |
| --- | --- | --- |
| John Blanchard | 9.5 | 9.5 |
| Mitchell Altieri | 8.7 | 8.7 |
| Sadyk Sher-Niyaz | 8.7 | 8.7 |
| Cary Bell | 8.7 | 8.7 |
| Mike Mayhall | 8.6 | 8.6 |
| Charles Chaplin | 8.6 | 8.6 |

**Analysis:** From the average score distribution we can see that there is only one extreme score i.e for tower inferno movie from the director John Blanchard but the rest most of the movies fall around the score 8 which means most of all are average.

**E. Budget Analysis:** Explore the relationship between movie budgets and their financial success.

* **Task:** Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

| **TOP\_5\_MOVIES** |
| --- |
| Avatar |
| Jurassic World |
| Titanic |
| Star Wars: Episode IV - A New Hope |
| E.T. the Extra-Terrestrial |

| **Correlation** | **Max\_profit** |
| --- | --- |
| 0.101033478 | 523505847 |

**Analysis:** From this we can see that the animated movies with average budget has earned the best profits. It tells that using animation to represent the stories can have a significant effect on the profit.

**EXCEL SHEET (MICROSOFT 365 ONLINE):**

https://1drv.ms/x/c/349a68758f9026f4/EcLlg1lJHXpPsa-JoMcP3Z0BNSfDNHb7OKW3YPMukgK1eA?e=6YTaeR