**Project: Investigate a Dataset (TMDb Movie Dataset)**

**Questions used for Analysis:**

1. Which movies are with highest and lowest Budgets?
2. Which movies are with highest and lowest Revenue?
3. Is there a relationship between The Revenue & The Budget?
4. Which movies are with the highest and lowest Profit?
5. Which movies are with highest and lowest Votes?
6. Does Voting Affect the Revenue?
7. Classification of Movies Based on Votes
8. Which Genre Has the Highest Release of Movies?
9. Correlations

**Used data wrangling, data cleaning, and summary statistics for the whole data.**

For the data wrangling:

* Found missing values in the dataset and filled it with zero in the original data.

For the data cleaning:

* Removed duplicate rows from the dataset.
* Removed unused columns for the data analysis.
* Change the date and time format.
* Adding a new column to the dataset which is “profit”.
* Creating a histogram to display the distribution of the whole data.

**Exploring the dataset:**

1. Find which movie has the highest budget and which one has the lowest budget in a bar plot.

& Display the top 10 movies based on their budget in ascending order in a bar plot.

1. Find which movie has the highest revenue and which one has the lowest revenue in a bar plot.

& Display the top 10 movies based on their revenue in ascending order in a bar plot.

1. Find if there’s a relationship between the revenue and the budget in a scatterplot graph and display the correlation between them.
2. Find which movie has the highest profit and which one has the lowest profit in a bar plot.

& Display the top 10 movies based on their profit in ascending order in a bar plot.

1. Find which movie has the highest votes and which one has the lowest votes in a bar plot.

& Display the top 10 movies based on their votes in ascending order in a bar plot.

1. Find if voting affects revenue in a scatterplot.
2. Classifying movies based on their votes to into classifications: excellent, average, poor.
3. Find which genre has the highest release of movies in a pie chart.
4. Correlations between revenue, budget, popularity, and vote\_average.

**Final results:**

1. Found that “The Warrior's way” is the highest budgetary movie.

& Displayed the top 10 movies ascendingly.

1. Found that “Avatar” is the highest earned revenue movie.

& “Wild Card” is the lowest earned revenue movie.

& Displayed the top 10 movies ascendingly.

1. Found that there’s a positive relationship between Revenues & Budget by 0.73, which means there's a high chance that movies with a higher budget make high revenues.
2. Found that "The Warrior's way" is the lowest profitable movie, although it cost a very high budget.

& "Avatar" is the highest profitable movie, although it cost less budget.

& Displayed the top 10 movies ascendingly.

1. "The Story of Film: An Odyssey" is the most-voted movie,

& "Transmorphers" & "Manos: The Hands of Fate" are the least-voted movies.

& Displayed the top 10 movies ascendingly.

1. Found that voting affection on Revenues is weak, so whenever average votes are high, it’s not necessarily that revenue is high. Votes were classified into excellent, average, and poor.
2. Observed that the Drama genre has the highest release of movies followed by Comedy and Thriller.
3. Correlation showed that there’s a positive correlation between revenue, budget, and popularity. Besides, there’s a weak positive correlation between vote\_average and revenue, budget, and popularity.