##The code is analyzing the ***Movie Data Set***

Cleaning data

Dropping irrelevant columns: I dropped the columns 'id', 'imdb_id', 'homepage', 'tagline', and 'overview' as they are not relevant for the analysis.

Converting release_date column to datetime format:I converted the 'release_date' column to datetime format using the pd.to_datetime() function.

Extracting the primary genre: I defined a function extract_primary_genre() to extract the primary genre from the 'genres' column. Then applied this function to create a new column called 'primary_genre' that contains only the primary genre for each movie.

Categorizing movie duration: I defined a function categorize_duration() to categorize the movie duration into 'Short', 'Medium', or 'Long' based on predefined thresholds. Then I applied this function to create a new column called 'duration_category' that categorizes the duration accordingly.

Remove zeros and negative numbers:remove the negative numbers and zeros from revenue and popularity as it is impossible to exist by condition greater than zero.

Dropping rows with missing values: I dropped rows with missing values in the 'revenue', 'genres', and 'runtime' columns using the dropna() function.

Converting budget columns to integers: I converted the 'budget', 'revenue', 'budget_adj', and 'revenue_adj' columns to integers using the astype() function.

The Statement of Questions:

1.What is the total profit by decade?

- To calculate the total profit by decade, subtract the budget from the revenue for each movie. then group the data by decade and calculate the sum of profits. Finally, visualize the results using a line plot.
- The total profit fluctuates over decades. There was a significant increase in profit from the 1970s to the 1990s, followed by a slight decline in the 2000s. However, the profit has been steadily increasing since the 2010s.

2.What is the distribution of runtime in movies?

- To analyze the distribution of runtime, I created a histogram to visualize the frequency of different runtime values.
- The majority of movies have a runtime between 80 and 140 minutes, with a peak around 100 minutes. There are fewer movies with extremely short or long runtimes.

3.How many movies are released each month?

- I calculated the number of movies released each month and visualize the results using a bar chart.
- Movie releases are relatively consistent throughout the year, with a slight increase in the summer months. December has the highest number of movie releases, possibly due to the holiday season.

4.How many movies are there for each duration category?

- I calculated the number of movies for each duration category and visualize the results using a bar chart.
- Most movies fall into the "Medium" duration category, with durations between 60 and 120 minutes. There are fewer movies categorized as "Short" (less than 60 minutes) or "Long" (more than 120 minutes).

5.Is there a correlation between popularity and revenue?

- I calculated the correlation coefficient between popularity and revenue to measure the strength and direction of the relationship. We can also visualize the relationship using a scatter plot.
- There is a positive correlation between popularity and revenue, indicating that movies with higher popularity tend to generate higher revenues. However, it's important to note that correlation does not imply causation, and other factors may contribute to a movie's success.