

**TASK**

**Exploratory Data Analysis on the Movies Data Set**

[](https://www.hyperiondev.com/)

**Introduction**

This is a dataset about different movies, it contains information about movies, including details such as title, budget, revenue, release date, popularity, vote average, and more. The goal is to analyse this data to uncover insights and trends in the movie industry.

**DATA CLEANING**

* Removing Unnecessary Columns: Columns that were not relevant to the analysis, such as ‘homepage’, ‘keywords’, ‘original language’, ‘original title’, ‘overview’, ‘production companies’, ‘status’, and ‘tagline’, were removed.
* Removing Duplicates: Duplicate rows were removed to ensure the dataset’s integrity.
* Handling Zero Values: Entries with zero budget or zero revenue were discarded as they do not provide meaningful insights.
* Converting Data Types: The ‘release date’ column was converted to DateTime format, and the budget and revenue columns were converted to integer data types.

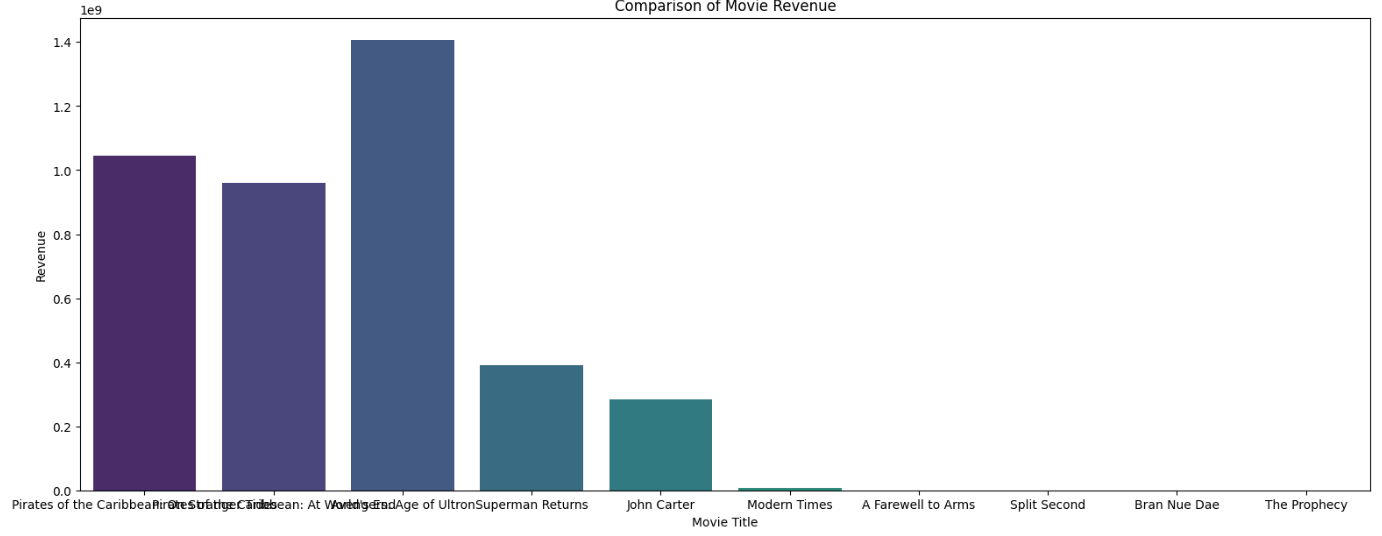
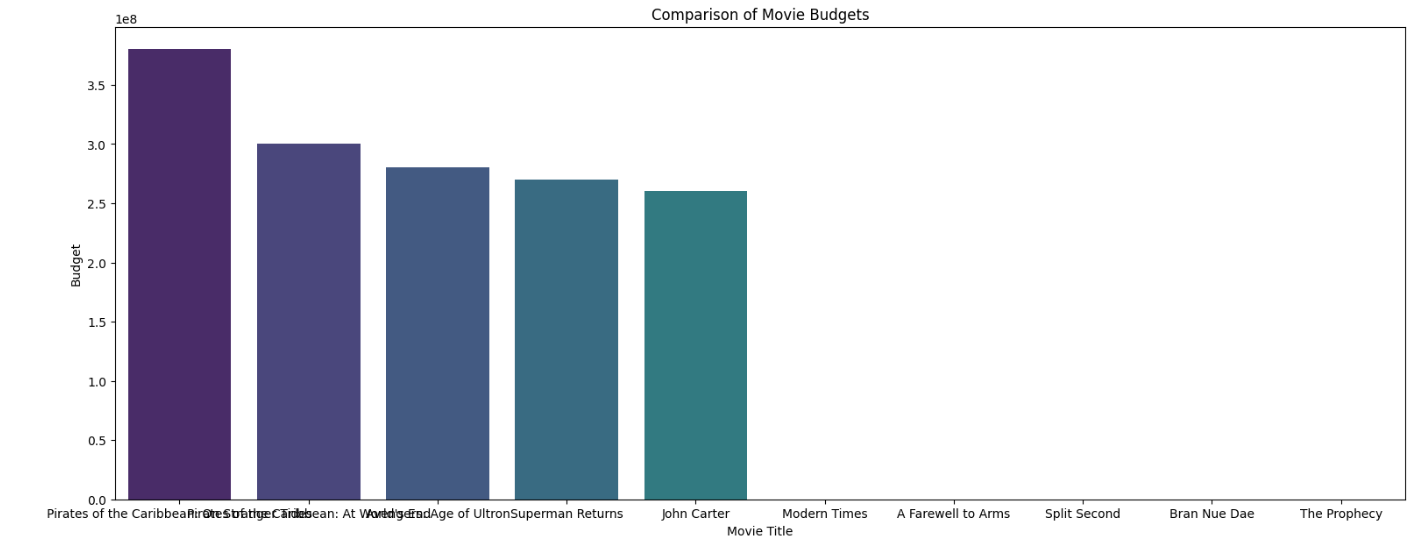
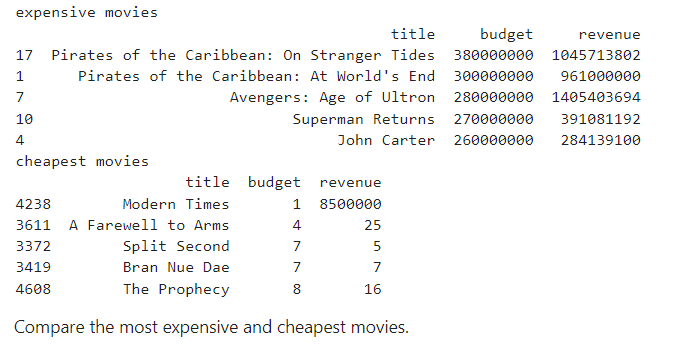
**MISSING DATA**

Some movies in the database had zero budget or zero revenue. Such rows were discarded.

**DATA STORIES AND VISUALISATIONS**

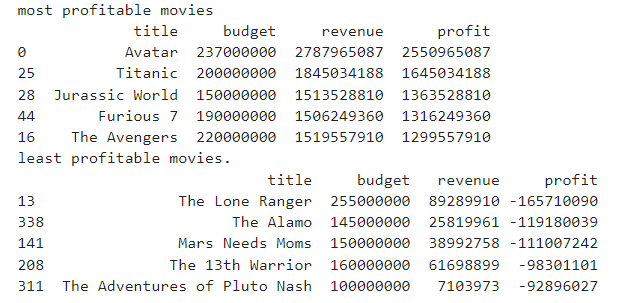
Which are the 5 most expensive movies?

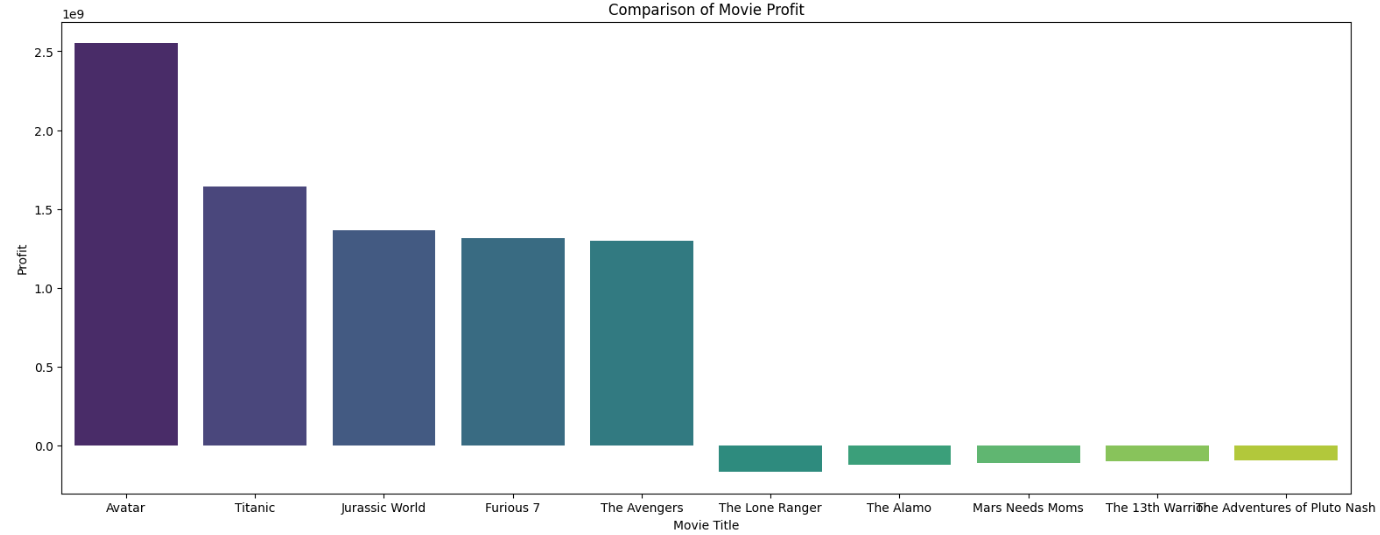
How do the most expensive and cheapest movies compare? Exploring the most expensive movies highlights if some movies are worth the money spent on them based on their performance and revenue generated.

  
  
Pirates of the Caribbean: On Stranger Tides and Avengers: Age of Ultron, had high budgets and they also generated very high revenue. Modern Times had a very low budget but was able to generate high revenue. Superman Returns and John by generated a revenue below the budget

What are the top 5 most profitable movies?

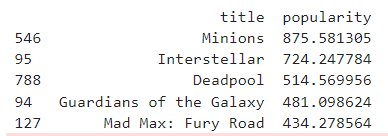
Compare the min and max profits. The comparison helps us identify the different approaches which failed and succeeded. Subtracting the budget from the revenue generated, will return the profit earned.

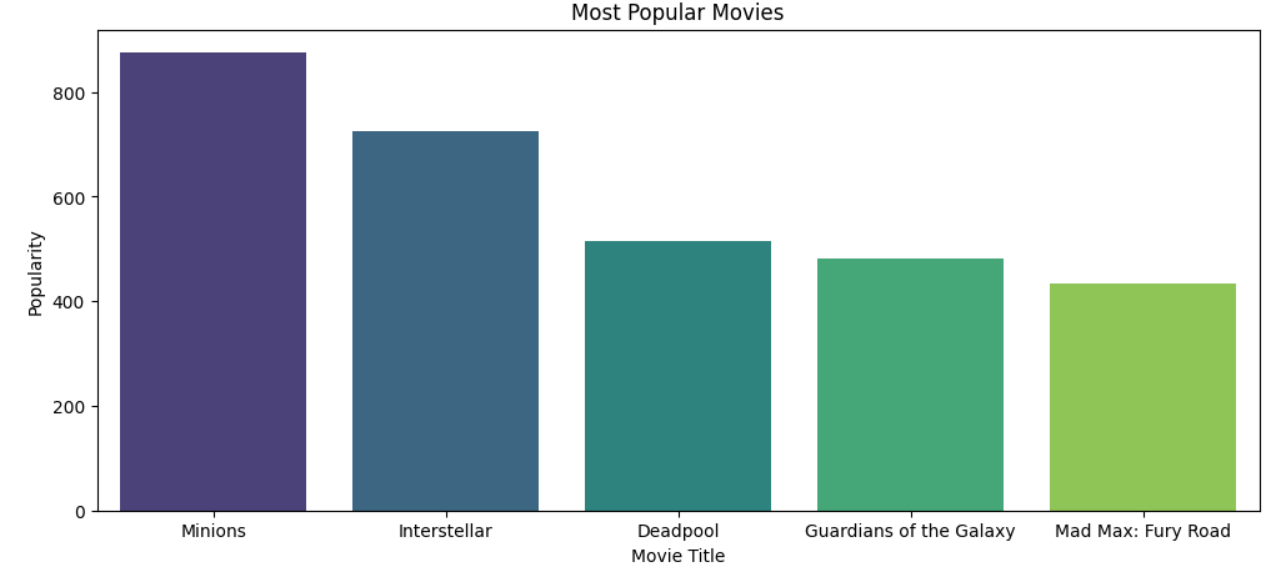




High budget does not always garantes high profit. The Lone Ranger was one of the movies with a very high budget but they also had a very high loss.

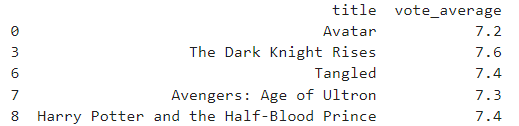
Find the most popular movies.

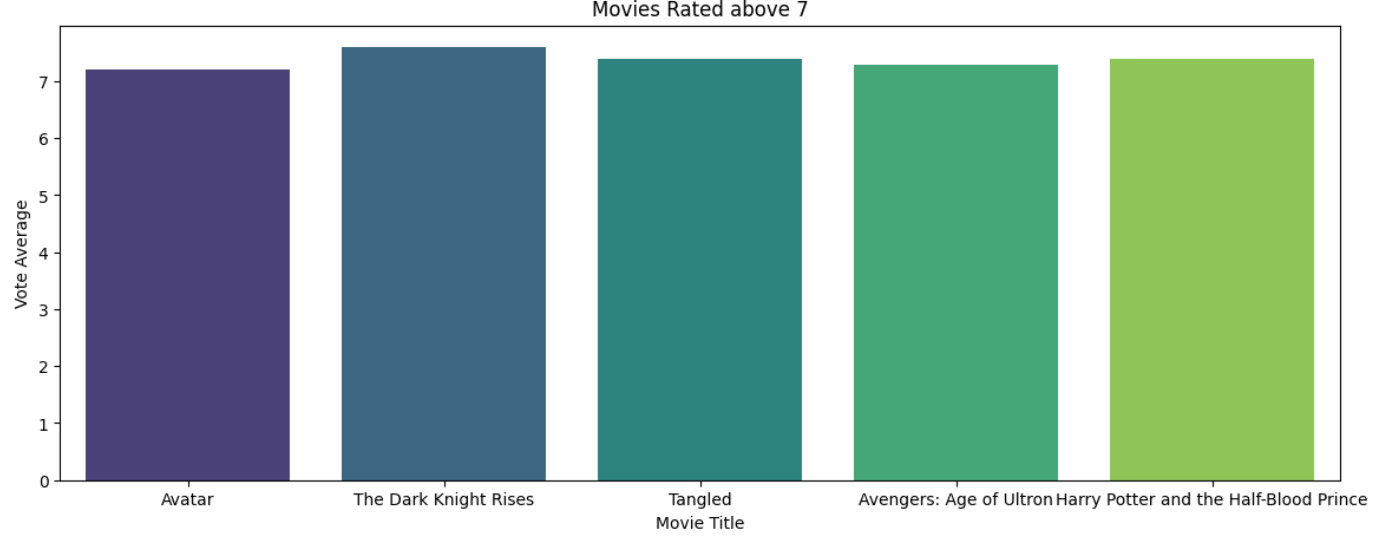




Minions had a high popularity above 800, followed by Interstellar with the popularity around 725. Mad Max: Fury Road had a popularity just above 400 but was also ranked together with most popular movies.

Movies which are rated above 7

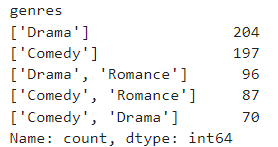


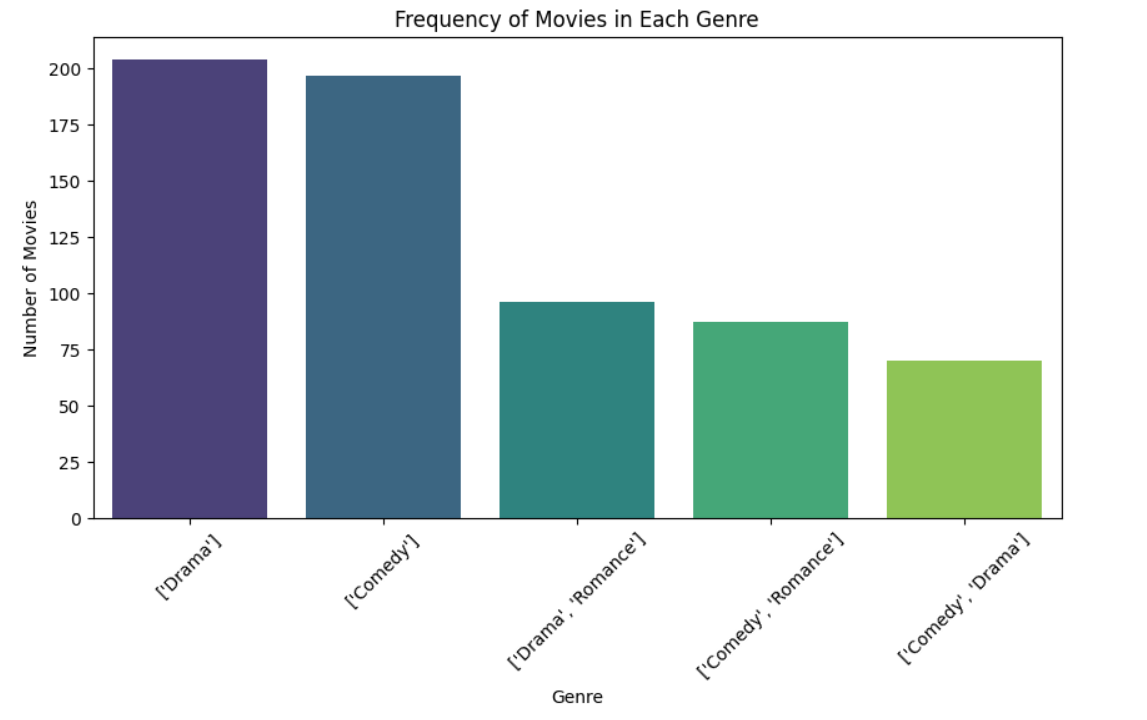


The above graph shows the top 5 movies rated above 7, with The Dark Night Rises being the highest rated movie at 7.6 rating.

Most successful genres

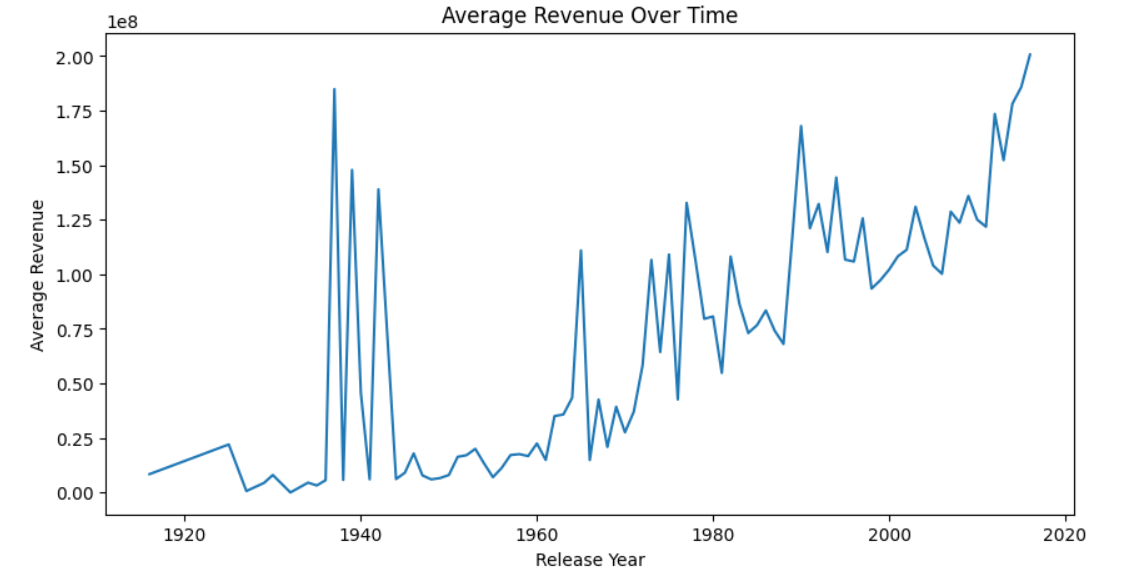
Create a bar plot explaining the frequency of movies in each genre.





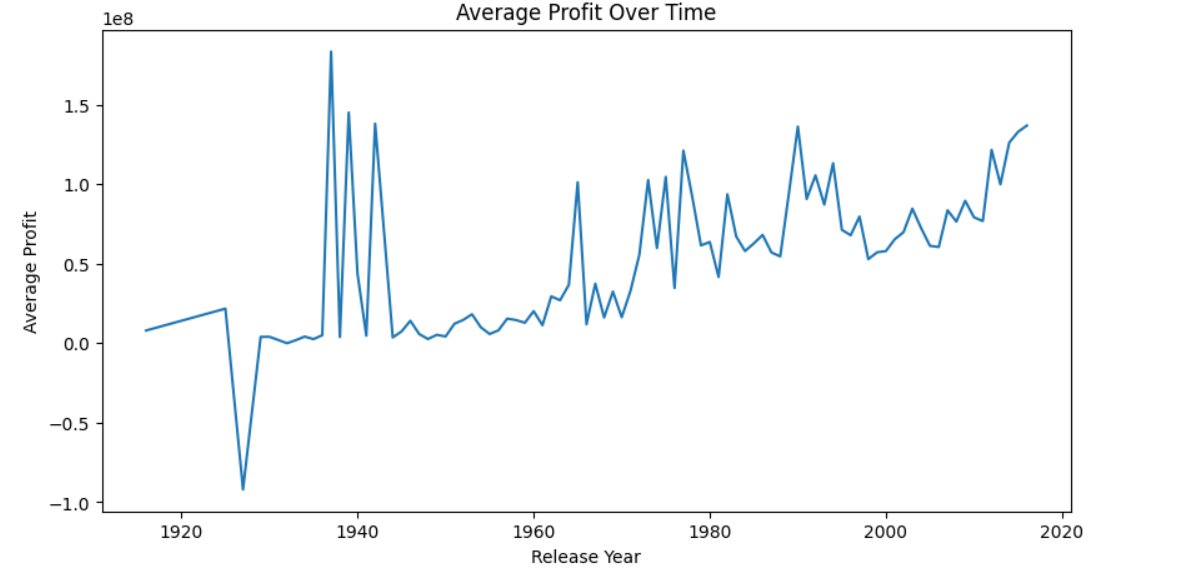
The genre with most frequency is drama with 204, followed closely by comedy with the frequency of 197.

Average revenue for movies over the years.



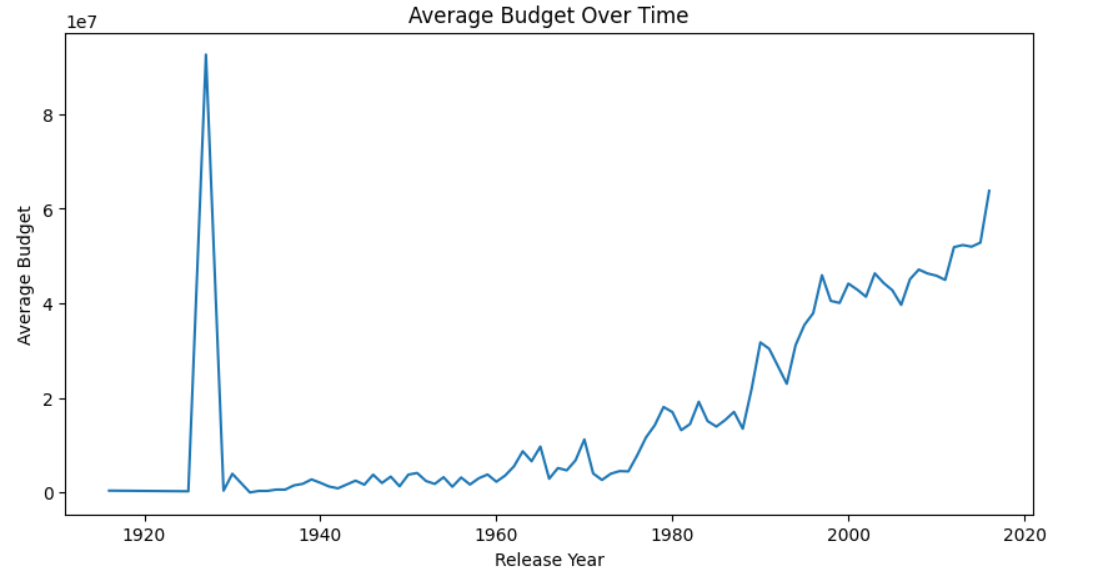
The above line plot shows the average revenue for the movies over the years, it makes it easy to understand trends in movie earnings. From the plot we can tell the revenue was high around 1940.

Average profit for movies over the years.



The above line plot shows the average profit for movies over the years, helping us understand treads in movie profitability. From the plot we can tell the profit have been on the increase from after 1940.

Average profit for movies over the years.



This line plot shows the average budget for the movies over the years, it helps us understand trends in movie production costs. Production costs started to rise around 1980.

**THIS REPORT WAS WRITTEN BY : SPHELELE KHUMALO**

