For this exercise, I selected a Netflix movies dataset that contains 16,000 records from 2010 to 2025. The dataset includes 16 descriptive attributes, such as movie titles, year of release, country, genre, revenue, budget, rating, popularity and vote count. Using this dataset, I aim to analyze trends in Netflix’s movie collection and viewer preferences.

To achieve this, I will use Matplotlib to create visual representation of key insights.

Research Questions

Using this dataset, I aim to answer the following questions:

1. Which genre is watched the most?
2. Which movie earned the highest revenue?
3. What country watches the most Netflix movies?
4. Which movie had the highest budget?
5. Which movie received the highest rating?
6. Which movie was the most popular?
7. Which movie had the highest vote count?

The Power of Visualization in Data Analysis

It is amazing how much insight can be gained through data visualization. By generating graphs, I have uncovered surprising facts that challenge common assumptions.

For example, I initially believed that action movies would be the most-watched genre on Netflix. However, the data revealed that drama is the most popular genre, with almost 7,000 different drama movies watched by customers. This demonstrates how personal beliefs can be misleading and how big data helps uncover objective truths.

Another common misconception I had was that Avatar: The Way of Water earned the highest revenue in movie history, including streaming revenue. Many people believe this to be true, but the data shows otherwise. In reality, Avengers: Endgame holds the record for the highest earnings at approximately $2.8 billion, compared to $2.3 billion for Avatar: The Way of Water.

It is not surprising that audiences in the United States watch the most Netflix movies, given that Netflix originated in the U.S. This aligns with expectations, as local audiences tend to drive the success of homegrown platforms.

One surprising fact was that Avatar: The Way of Water had the highest production budget at approximately $460 million. I always thought Titanic had the highest budget, but it did not even make it into the top 10 highest-budget movies.

I also discovered that 35 movies received the highest rating on Netflix. Some of these highly rated films include The Shepherd, The Williams, Queen of the Ring, Butterfly, and many more.

Surprisingly, The Gorge was the most popular movie on Netflix, with an engagement score of 3,876.

Finally, Inception had the highest vote count at 37,119 votes. This high number suggests that the rating for Inception is likely highly reliable, as a larger sample size generally leads to more accurate and representative feedback.

To summarize all the findings, I generated graphs to visualize the data. Using the Netflix dataset, I uncovered several insights that challenged common assumptions. Drama emerged as the most-watched genre, substantially surpassing action, which I initially thought would be the leader. This underscores how access to big data can correct personal biases.

I discovered that Avengers: Endgame earned the highest box office revenue at approximately $2.8 billion, contradicting to the common belief that Avatar: The Way of Water was the highest-grossing film. Additionally, while Avatar: The Way of Water had the highest production budget of around $460 million, Titanic – which many assume to be the most expensive film-did not even make the top 10 list.

The data also revealed that the United States has the largest Netflix audience, which was not surprising to me. This aligns with my expectations, given that Netflix originated in the U.S. The most popular movie in terms of engagement was The Gorge, with a popularity score of 3,876. Furthermore, Inception had the highest vote count, with 37,119 people rating the movie, making its rating highly reliable due to the large sample size.

Overall, this analysis demonstrates the power of data visualization in uncovering trends, correcting misconceptions, and providing objective insights. What we assume to be true is not always supported by data. This highlights the importance of data-driven decision-making in understanding consumer behavior and industry trends.