Kevin Chiang
Romil Shah
Leo Dai
Aryan Barik
Krutik Doshi
Salman Alsabah

WatchTower Data

Introduction:

In today's data-driven world, businesses have access to vast amounts of data that can provide valuable insights into consumer behavior and preferences. However, working with large and complex datasets can be a daunting task, requiring specialized skills and tools. In this context, we were tasked with exploring three massive datasets from a streaming app and extracting relevant information from them. Our objective was to take a sample from the datasets and identify the most popular shows among them, as well as any commonalities between them. To achieve this, we used the powerful data analysis tools R and Tableau, which enabled us to conduct exploratory data analysis and derive meaningful insights from the data. The datasets we worked with were collected by NBCUniversal through their old iOS streaming app and contain information about user viewing habits across different brands. However, before we could analyze the data, we had to clean it first, a crucial step in any data analysis process.

Methodology:

The methodology we followed to extract insights from the three large datasets provided involved several steps. Firstly, we selected the first 100,000 rows from each dataset to create a manageable sample size for analysis. We then removed any missing values and duplicates from the data to ensure data quality. Next, we extracted the relevant information from the "value" column, which included the show name and watch duration. To further simplify the data, we summed the watch duration for each show across. This allowed us to determine which shows were the most popular in each sample. Finally, we utilized Tableau, a powerful data visualization tool, to create clear and easy-to-understand visualizations of the data. By presenting the results in an accessible manner, we were able to communicate our findings effectively and facilitate better decision-making.

Results summary:

Across all data sets, we found that the shows The Real Housewives of Beverly Hills, Vanderpump Rules, The Real Housewives of Atlanta, and Top Chef were the most popular shows, from most popular to fourth most popular respectively. The Expanse and Mr. Robot consistently appeared at the bottom of our graphs.
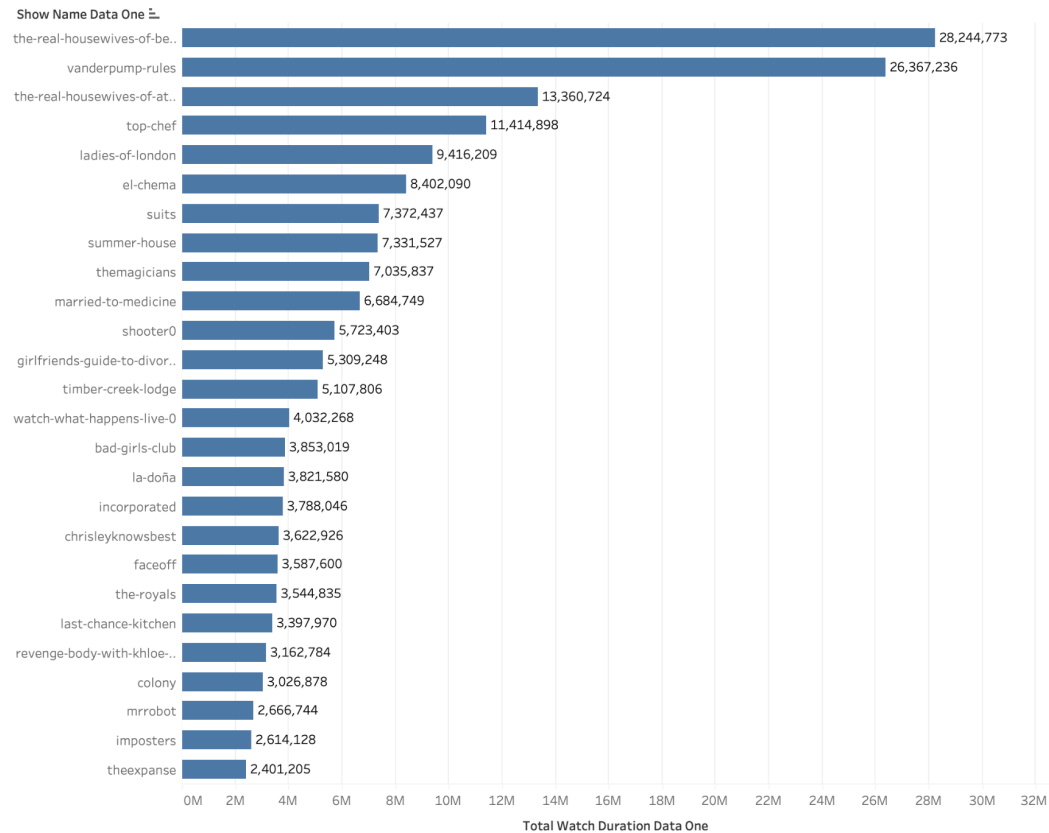

Figure 1: A bar chart of total watch durations (in million minutes) of various Netflix shows for Data Set 1
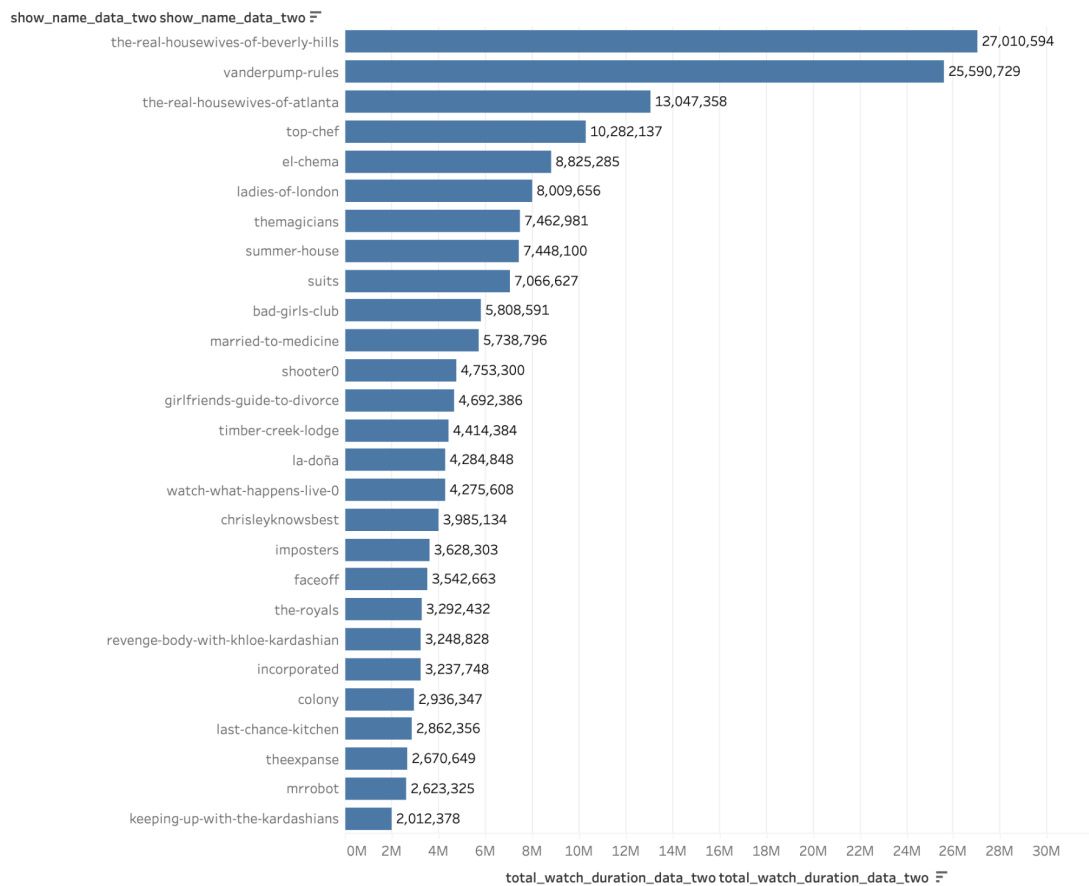
Figure 2: A bar chart of total watch durations (in million minutes) of various Netflix shows for Data Set 2
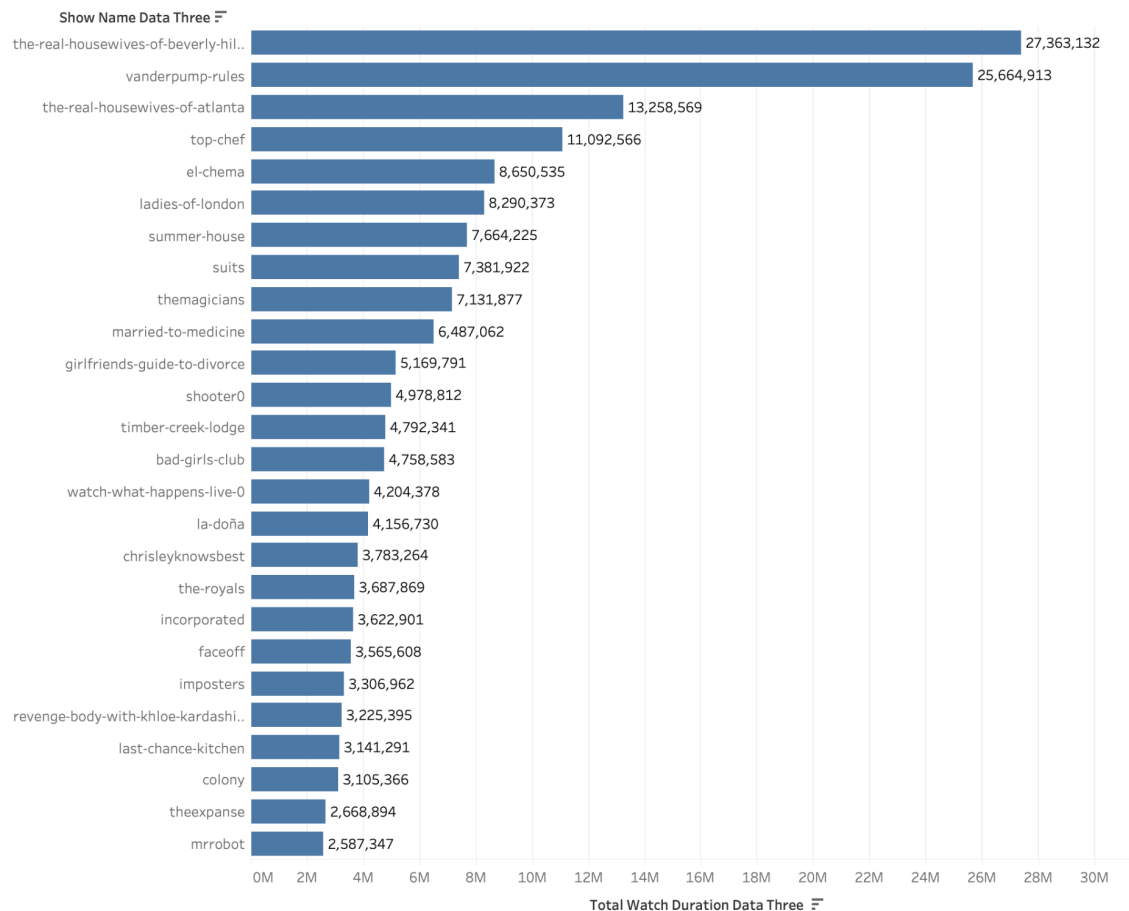
Figure 3: A bar chart of total watch durations (in million minutes) of various Netflix shows for Data Set 3

Conclusion:

The top four shows, Real Housewives of Beverly Hills, Vanderpump Rules, The Real Housewives of Atlanta, and Top Chef are all reality television shows, showing that the genre is extremely popular and has widespread appeal. We can conclude that the audience really enjoys Real Housewives types of shows, however, the housewives who live in Beverly Hills have more appeal. From the low viewing numbers of Mr. Robot and The Expanse, it can be concluded that science fiction shows have low appeal.

Recommendation:

We recommend NBCUniversal to create more reality television shows and create fewer science fiction shows to maximize viewers. Vanderpump Rules is a spin-off of The Real Housewives of Beverly Hills, which shows that the Real Housewives brand is extremely strong. We believe that creating more Real Housewives television shows or spin-offs of the series would be extremely successful. We would also suggest creating fewer science fiction shows as they seem to gain fewer viewers than other genres.