

Project Report

Analysis of Personal Netflix
Streaming Data

İrem Şahin

NETFLIX

About the Project

Introduction

The purpose of this project was to explore my personal streaming data on Netflix, focusing on the correlation between the genre of what I watched, time of day, and device type I used for streaming. My hypothesis was: "During afternoon hours, I am more likely to watch romantic content on larger screens like TVs, while evenings I watch comedies on PC devices." The purpose of this study is to get understanding about my own watching habits and preferences.

Data Source

For my main data, I requested my streaming history data from Netflix, and got 4 years of streaming data. But, to test my hypotheses, I need genres and types of what I watched. I provide these values and add them into my Data.



Data Preprocessing

The Netflix streaming data was initially loaded into a Pandas DataFrame for manipulation and analysis. Main preprocessing steps included:

1. Separating 'Start Time' into distinct date and time columns.
2. Simplifying 'Title' by removing season and episode number
3. Formatting "Duration" in terms of a minute.
4. Categorizing 'Device Type' into broader classes (Mobile, Tablet, PC, TV).
5. Dropping rows where 'Latest Bookmark' was 'Not latest view' since they are duplicates.
6. Creating a 'Part of Day' feature to categorize viewing times into morning, afternoon, evening, and night.

Exploratory Data Analysis (EDA)

To investigate the data, several visualizations were made, including:

- The distribution of genres demonstrated the variety and frequency of genres watched.
- The distribution of device types revealed preferences for various streaming devices.
- Time of day analysis revealed viewing patterns across different times.
- Additional analyses included genre popularity trends over time, and a correlation heatmap to understand relationships between different variables.

In-Depth Analysis

I tested my hypothesis using statistical methods and visualizations:

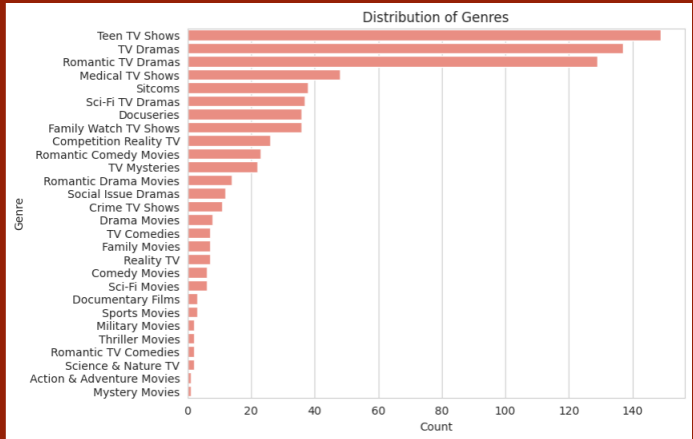
- Chi-Square tests were applied to inspect the association between time of day and genre, and between time of day and device type.
- An analysis of the interaction effect shed light on how genre, device type, and time of day all affect watching habits.

Findings

The analysis revealed several key insights:

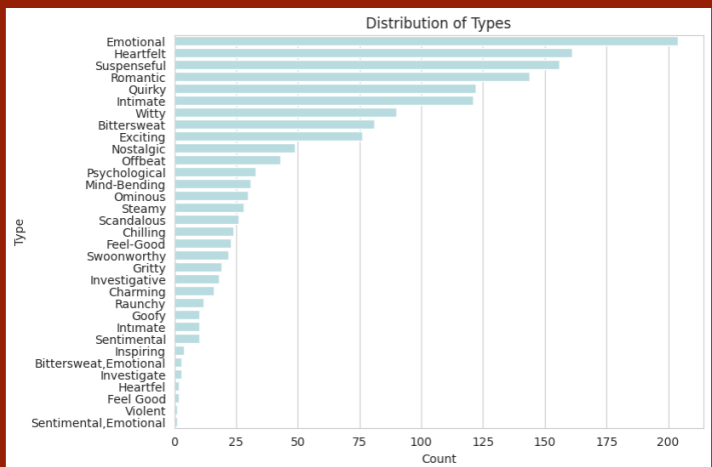
- Top 3 watched genres were:

- o Teen TV Shows : 149 times
- o TV Dramas: 137 times
- o Romantic TV Dramas: 129 times



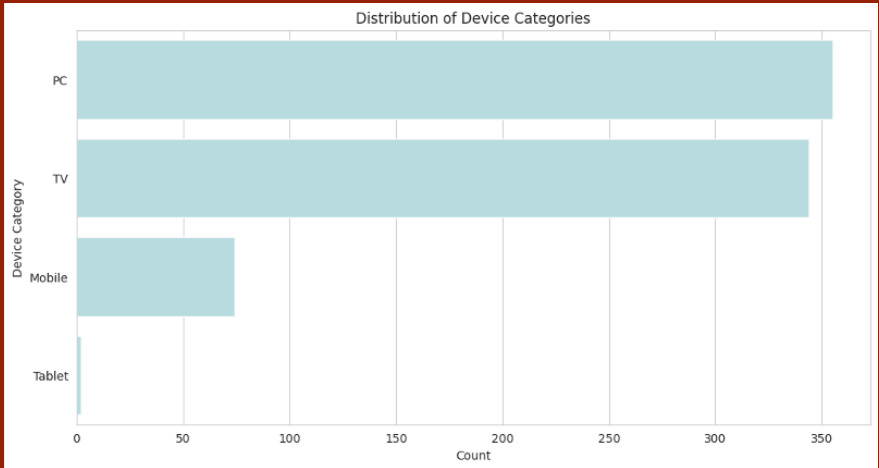
- Top 3 watched types were:

- o Emotional: 204 times
- o Heartfelt: 161 times
- o Suspenseful: 156 times

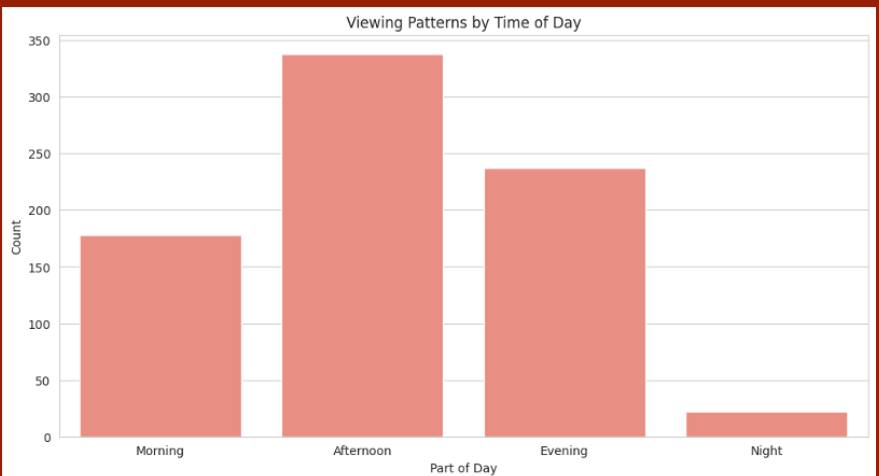


Findings

- Mostly, TV was preferred to watch movies and TV series.

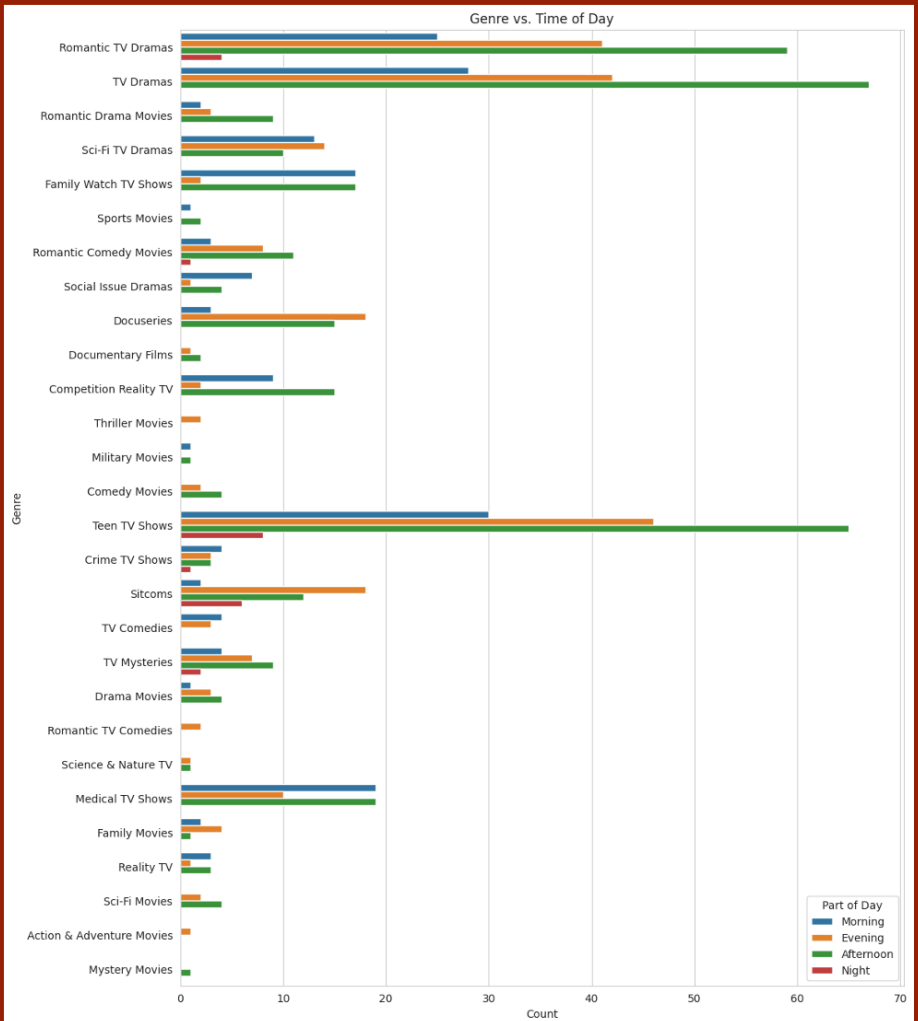


- Looking at the viewing time, more viewing was done in the afternoons.



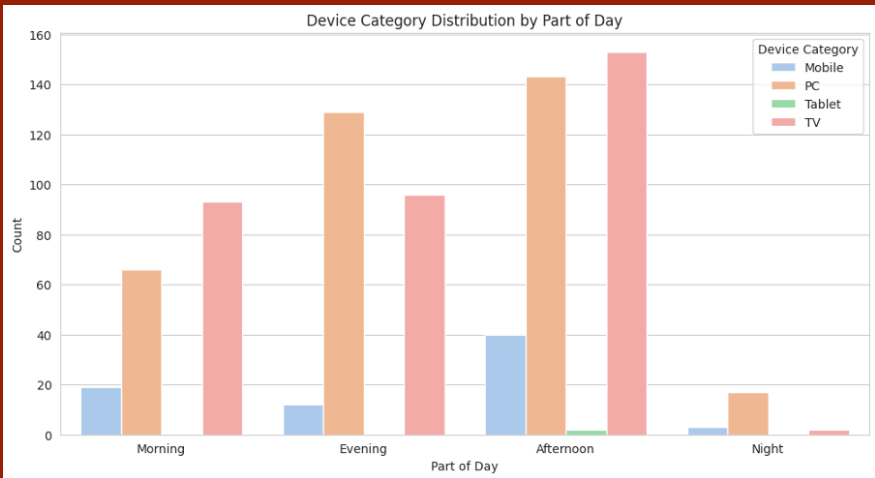
Findings

- There was a significant association between the time of day and the genre of content watched. Romantic content was more frequently watched during afternoon hours, aligning with the hypothesis.

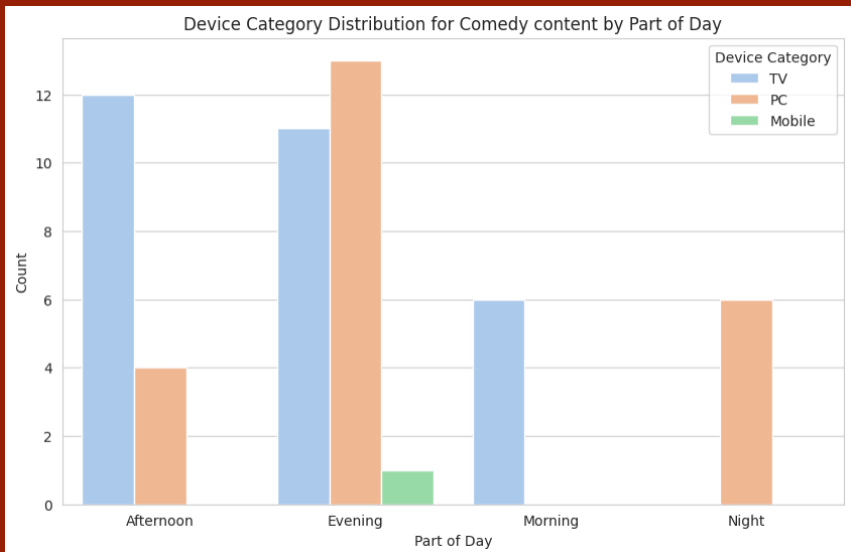


Findings

- The choice of device also varied with the time of day. Larger screens like TVs were preferred during the afternoon.

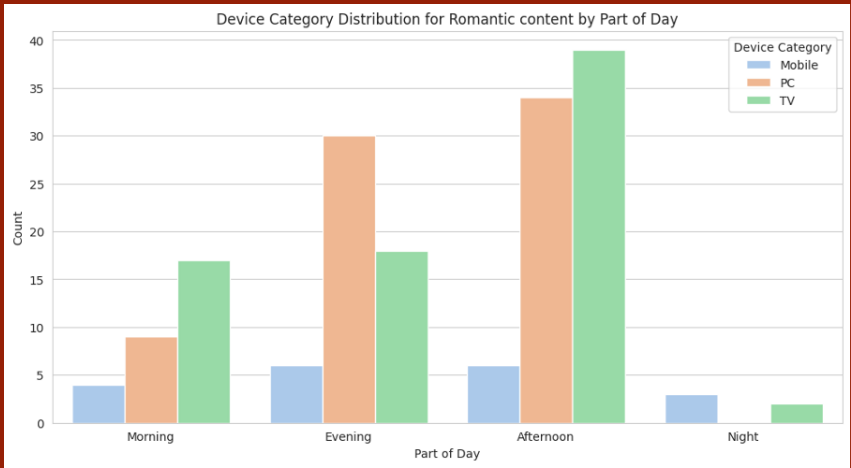


- Consumption of comedy content increases in evenings, predominantly on PC devices, supporting the hypothesis.

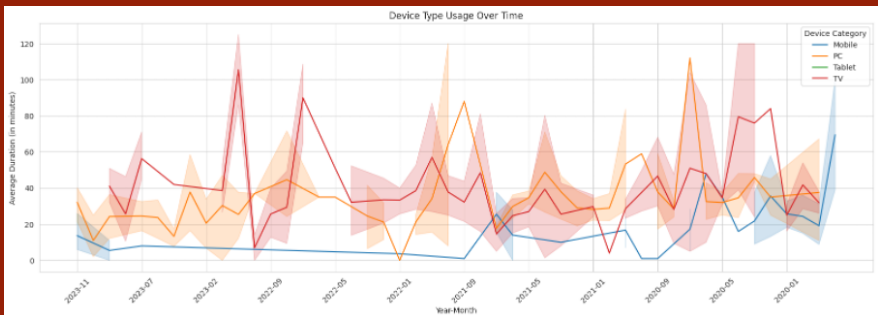


Findings

- Consumption of romantic content increases in afternoons, predominantly on TV devices, supporting the hypothesis.

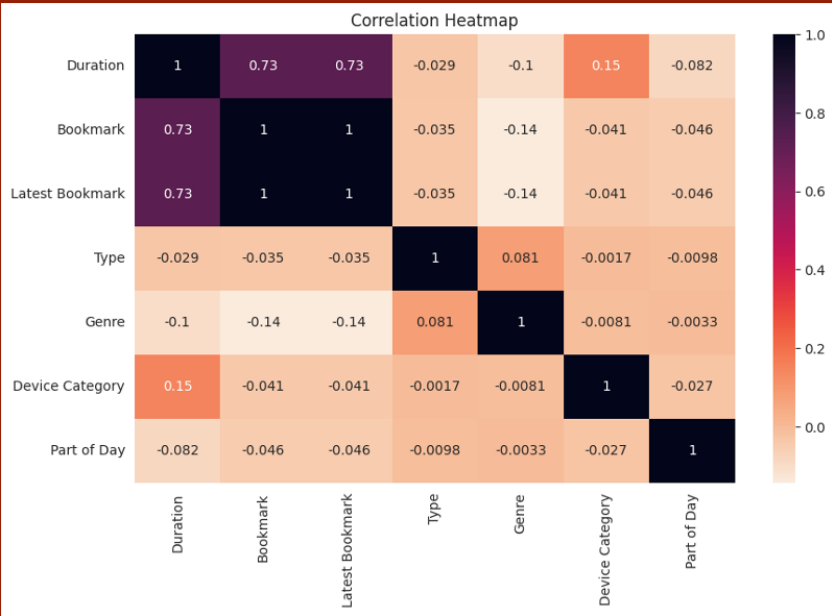


- Device type usage/popularity is not change considerably over the time. It tends to maintain similar popularity rates.



Findings

- Watching duration and watched device has the highest correlation compared to other categories.



Conclusion

The idea was validated by the data analysis, which showed clear trends in streaming behavior depending on the device type, content genre, and time of day. These results offer insightful information about individual viewing preferences and can guide future selections of devices and content for the best possible streaming experience.

All these data shows us that my hypothesis is useful and during afternoon hours, I am more likely to watch romantic content on larger screens like TVs, while evenings I watch comedies on PC devices.