Project Overview

This project involves analyzing a my Netflix viewing history. The analysis is conducted using a dataset obtained from the user's Netflix account and processed through a Jupyter notebook.

Data Collection

- **Source**: The data was collected from my personal Netflix account.
- **Method**: Scraped the data from Netflix's own website using Selenium & Beautifulsoup libraries and converted it into CSV file. No any other 3rd party website/app is used in this process.

Data Cleaning and Preprocessing

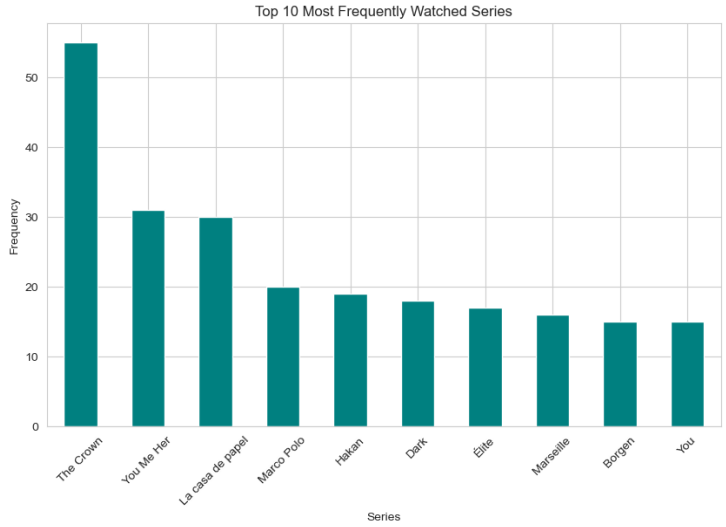
- **Libraries Used**: `pandas` for data manipulation.
- **Methods**:
    1. **Handling Missing Values**: Checked and handled missing values in the dataset.
    2. **Parsing Dates**: Converted the 'Date' column into a datetime format for easier analysis.
    3. **Splitting 'Title' Column**: Extracted the series name, season, and episode from the 'Title' column.
    4. **Data Type Conversion**: Converted data to appropriate types, e.g., 'Season' to categorical data type.
    5. **Removing Duplicates**: Identified and removed duplicate entries.

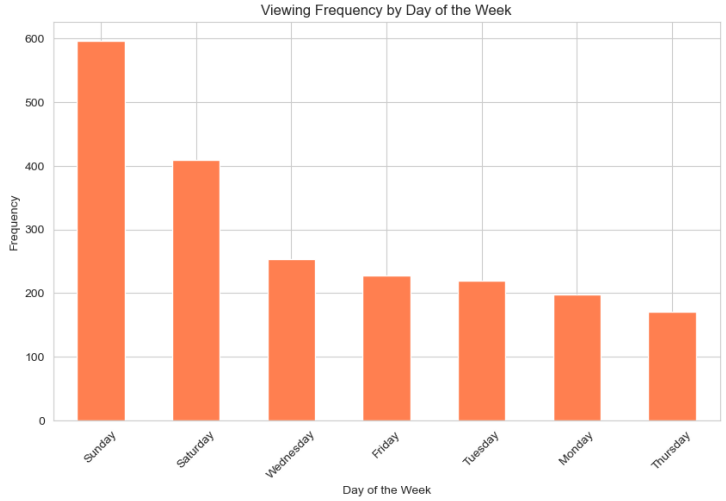
Exploratory Data Analysis (EDA)

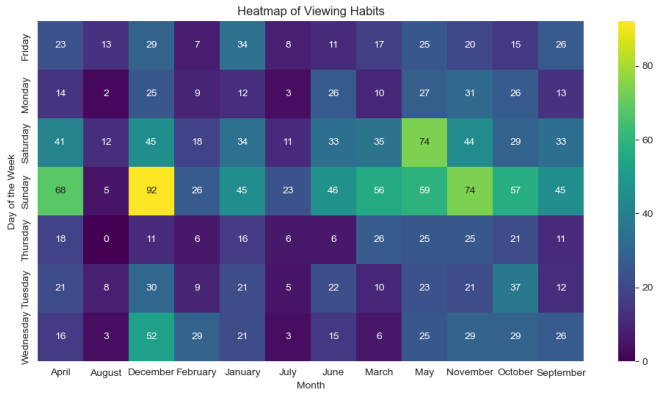
- **Libraries Used**:
    - `pandas` for data handling.
    - `matplotlib` and `seaborn` for data visualization.
- **Methods**:
    - **Frequency Analysis**: Identified the most frequently watched shows or movies.
    - **Trend Analysis**: Analyzed viewing habits over time, including daily, weekly, and monthly trends.
    - **Binge-Watching and Series on Repeat**: Determined patterns of binge-watching and identified series watched multiple times.

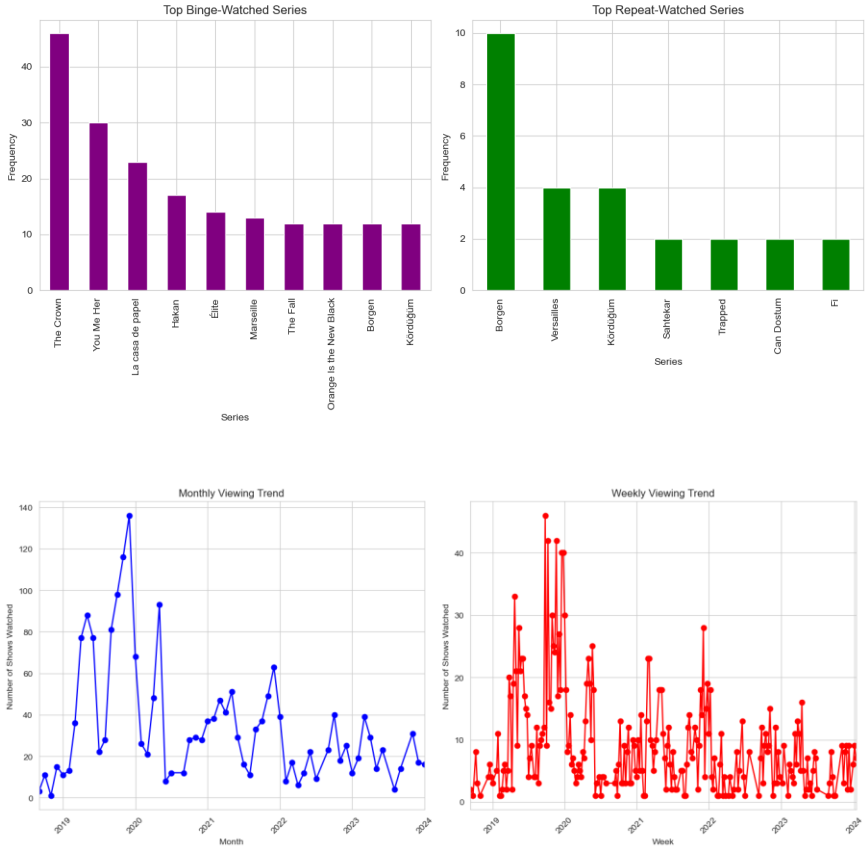
- **Visualization**: Created bar charts, line graphs, and other visual representations to illustrate findings.

Some of the important Visuals









Machine Learning

Implemented a machine learning algorithm using scikit-learn for clustering, one-hot encoding, labelling ‘Interest Level’, ‘Engagement’, ‘Numeric Interest’ based on training and predictions. Successfully completed the model with a 0.89 accuracy score.

Conclusion

The project utilized data analysis techniques and visualizations to reveal interesting patterns in my Netflix viewing habits. It serves as a practical application of data science methods to personal entertainment data.