Netflix Viewing History Analyzer

A Mini Project Report

Jaimin Hadvani
Dhirubhai Ambani University / Information and Communication Technology

2023-27

Contents

Executive Summary

This project presents a comprehensive data analysis tool designed to analyze Netflix viewing history data. The application processes exported viewing history CSV files and generates detailed insights through statistical analysis and data visualization. The tool provides users with information about viewing patterns, including top-watched content, viewing trends, binge-watching sessions, and temporal viewing behaviors.

Chapter 1

Introduction

1.1 Project Overview

The Netflix Viewing History Analyzer is a Python-based data analysis application that transforms raw Netflix viewing history data into meaningful insights and visualizations.

1.2 Objectives

- Develop an automated tool to analyze Netflix viewing history.
- Identify viewing patterns and trends.
- Detect binge-watching sessions.
- Visualize habits across days, weeks, and months.

1.3 Scope

The project analyzes:

- Total viewing sessions and watch time
- Frequently watched titles
- Viewing patterns by weekday and hour
- Monthly viewing trends
- Binge-watching detection

Chapter 2

Methodology

2.1 Data Source

The project uses Netflix viewing history exported as a CSV file, containing:

- Title
- Date
- Duration (optional)

2.2 Data Processing Pipeline

2.2.1 Data Loading

The tool automatically detects CSV structure and performs error handling.

2.2.2 Data Cleaning

- Date parsing
- Duration extraction (numeric and time string formats)
- Handling missing values

2.2.3 Feature Engineering

- Weekday
- Hour
- Month period

2.2.4 Statistical Analysis

- Total watch time
- Top titles
- Grouping by time periods
- Binge session detection

Chapter 3

Features and Functionality

3.1 Summary Statistics

- Total viewing sessions
- Total watch time
- Most watched title
- Title with highest time investment

3.2 Top Content Analysis

Top 10 titles by:

- Views
- Watch time

3.3 Temporal Analysis

- Weekday-hour heatmap
- Monthly trends
- Last 30 days trend

3.4 Binge Detection

A binge is defined as watching 3+ episodes of the same title in one day.

3.5 Visualizations

- Horizontal bar charts
- Line graph
- Heatmap
- Summary dashboard

Chapter 4

Technical Implementation

4.1 Technology Stack

- Python 3.x
- pandas
- numpy
- matplotlib
- seaborn

4.2 Code Architecture

Core functions include:

- `load_netflix_csv()`
- `detect_datetime_column()`
- `detect_duration_column()`
- `try_extract_minutes_from_str()`
- `main()`

4.3 Output Structure

All results are saved in `netflix_output/`:

- summary.csv

- top_by_views.csv
- top_by_time.csv
- binge_sessions.csv
- All visualizations (.png)

Chapter 5

Results and Analysis

5.1 Sample Results

- Total Sessions: 733
- Total Watch Time: 733 minutes (approx. 12 hours)

5.2 Key Insights

- Viewing frequency trends
- High time-investment titles
- Preferred viewing hours
- Monthly viewing habits

Chapter 6

Challenges and Solutions

6.1 Data Format Variability

Solved using flexible column detection.

6.2 Duration Parsing

Implemented regex-based robust extraction.

6.3 Missing Data

Default values and fallback logic were used.

Chapter 7

Future Enhancements

- Genre analysis
- Recommendation engine
- GUI/Web interface
- Interactive dashboards
- Advanced statistics

Chapter 8

Conclusion

The Netflix Viewing History Analyzer successfully applies data science techniques to real-world personal viewing data. The project demonstrates skills in data processing, visualization, and Python automation.

Chapter 9

References

- Pandas Documentation
- Matplotlib Documentation
- Seaborn Documentation
- Netflix Viewing Activity Export