CEN445 Introduction to Data Visualization - Project Report

Project Title: Netflix Data Visualization Dashboard

Course: CEN445 - Introduction to Data Visualization

Team Members: Emre Kahraman, Asusena Ela Öztürk, Orhan Karaman

Date: November 23, 2025

Dataset Context and Source

This project uses the "Netflix Movies and TV Shows" dataset from Kaggle, containing about 8,800 entries as of mid-2021. It includes 12 core attributes covering metadata like type, title, cast, country, release year, rating, duration, and genre. The dataset was thoroughly cleaned by converting dates and durations into usable formats, handling missing country and director values, and standardizing genre labels for consistent categorization.

Analysis Goals

The goal of this dashboard is to uncover Netflix's content strategy by analyzing how its library has evolved over time, how geographically diverse its productions are, and how content attributes-such as duration, maturity ratings, and genres-interrelate, using a range of visual analytics techniques.

We implemented **9 distinct visualization techniques** (3 Basic, 6 Advanced) using **Python (Streamlit & Plotly)**. Some of them:

1. **Treemap:** Genre Composition - A hierarchical view separating Movies vs. TV Shows to show which genres dominate each format.

2. **Parallel Coordinates Plot:** *Length Trends* - An advanced multivariate plot connecting Release Year to Duration, allowing users to spot trends in movie runtimes over the last decade.

3. **Sankey Diagram:** *Maturity Flow* - Visualizes the complex flow from Content Type to Maturity Rating to Genre, showing which genres tend to be rated for mature audiences.

4. **Heatmap:** *Rating Intensity* - A matrix view comparing Rating vs. Type to highlight the frequency of specific ratings (e.g., TV-MA) across different formats.

Key Insights and Findings

The analysis shows that Netflix's content library grew steadily until 2015 before accelerating rapidly with global expansion. Movie runtimes cluster around 90-100 minutes, with some longer durations emerging in critically favored genres. Rating patterns highlight a strong shift toward adult-oriented content, dominated by TV-MA and R categories. Regionally, the U.S. produces the broadest mix, while countries like South Korea and Japan show more specialized outputs, particularly in dramas and anime.

The complete source code, dataset, and documentation are available at:

GitHub Link: https://github.com/EmreKahraman9797/CEN445-Introduction-to-Data-Visualization.git