Proposal: Exploring Trends and Predicting Success in Netflix Daily Top 10 Rankings

Problem Identification

Problem Statement Formation: The project aims to analyze and predict factors influencing the daily top 10 rankings of movies and TV shows on Netflix in the United States. By exploring trends and patterns in viewership scores over time and identifying key attributes such as release date, exclusivity, and type (movie or TV show), the objective is to uncover insights that contribute to understanding audience preferences and streaming behavior.

Context: The landscape of streaming platforms like Netflix is highly competitive, with content popularity driving subscriber retention and acquisition. Understanding what makes a title trend on Netflix can inform content acquisition strategies and production decisions.

Criteria for Success

1. **Accurate Prediction:** Develop models that accurately predict which titles will enter or maintain their positions in the daily top 10 based on historical data.
2. **Insight Generation:** Extract meaningful insights into the factors (e.g., release timing, type of content) influencing viewership scores and rankings.
3. **Actionable Recommendations:** Provide actionable recommendations to content creators and distributors based on the analysis to optimize content strategy and increase viewership.

Scope of Solution Space

The project scope encompasses:

- Data preprocessing and cleaning to handle missing values and ensure data quality.
- Exploratory Data Analysis (EDA) to visualize trends and distributions of viewership scores, rank movements, and other relevant metrics.
- Statistical analysis to identify correlations between variables like release date, type, and viewership score.
- Predictive modeling using machine learning techniques to forecast daily top 10 rankings.
- Feature importance analysis to understand which attributes most significantly impact a title's ranking.

Constraints

- **Data Quality:** Ensuring the accuracy and completeness of the dataset, particularly in handling missing values and anomalies.
- **Temporal Dynamics:** Accounting for seasonality and trends in viewership patterns that may affect daily rankings.

Stakeholders

- **Netflix:** Interest in optimizing content placement and understanding audience preferences.
- **Content Creators and Distributors:** Benefit from insights into factors driving Netflix viewership and engagement.

Data Sources

The primary data source is the Netflix daily top 10 dataset covering movies and TV shows from 2020 to March 2022, including attributes such as title, release date, type, and viewership metrics.

Approach

1. **Data Cleaning and Preparation:**
    - Handle missing values and inconsistencies in the dataset.
    - Convert data types and ensure consistency for analysis.
2. **Exploratory Data Analysis (EDA):**
    - Visualize distributions and relationships between variables.
    - Explore trends in viewership scores and rank movements over time.
3. **Statistical Analysis:**
    - Conduct statistical tests to identify significant correlations between attributes and viewership scores.
    - Analyze the impact of factors like release date, type, and exclusivity on daily rankings.
4. **Predictive Modeling:**
    - Develop machine learning models (e.g., regression, classification) to predict daily top 10 rankings.
    - Evaluate model performance using appropriate metrics (e.g., accuracy, F1-score).
5. **Feature Importance Analysis:**
    - Determine which features (e.g., release date, type) are most influential in predicting top 10 rankings.

    - Provide insights into actionable recommendations based on the feature analysis.

Deliverables

- **GitHub Repository:**
    - Codebase for data preprocessing, EDA, modeling, and evaluation.
- **Project Report:**
    - Detailed documentation of methodology, findings, and conclusions.
- **Presentation Slides:**
    - Summary of project insights and recommendations for stakeholders.