Bishal Sharma Roy
UB ID: 50571143

Project Proposal: Understanding Content Trends Across Streaming Platforms

Problem Statement:
With the rise of multiple streaming platforms such as Netflix, Amazon Prime, Disney+, HBO Max, and Apple TV+, there is increasing competition in content offerings. This project aims to analyze how these streaming platforms differentiate themselves through their content, learning their key trends in genres, runtime of content, IMDb scores, and platform-specific strengths.

Execution Plan

To conduct this analysis, multiple datasets from **Kaggle** will be used, consisting around 20,000 content titles, including:

- **Amazon Prime TV Shows and Movies Dataset**
- **Apple TV+ Movies and TV Shows Dataset**
- **Disney+ TV Shows and Movies Dataset**
- **HBO Max TV Shows and Movies Dataset**
- **Netflix TV Shows and Movies Dataset**
- **IMDB Movie Reviews Dataset**
- **The Movies Dataset (In general)**

These datasets will be aggregated into a single CSV file, ensuring that all relevant data is structured for effective visualization. The data cleaning process will involve handling missing values wherever possible, ensuring consistency in formats, and applying appropriate transformations to unify the datasets.

For the analysis, **Power BI** will be used to create insightful visualizations that uncover content distribution patterns, genre popularity, and platform-specific trends. If data

inconsistencies or missing values pose a challenge, filtering techniques will be applied within Power BI to ensure the quality and accuracy of the insights generated.

Potential Barriers and Mitigation Strategies

1. **Data Quality and Missing Values:** Some datasets may have missing or inconsistent data. This will be addressed by data cleaning techniques in **Jupyter Notebook** before importing the final dataset into Power BI.
2. **Data Aggregation Complexity:** Merging multiple datasets into a single structured format may be challenging. Proper data mapping and standardization techniques will be used to maintain consistency.
3. **Visualization Challenges:** Some trends may not be immediately evident in Power BI. Interactive filtering and multiple visual representations will be utilized to extract meaningful insights.
4. **Time Constraints:** To ensure timely completion, the project will follow a structured timeline, focusing on **data preparation, visualization, and interpretation** in distinct phases.
5. **Technical Roadblocks:** If technical difficulties arise, I'll consult with **Professor Dominic Sellitto** or leverage additional resources to address roadblocks efficiently.

By systematically following this plan, this project will provide valuable insights into how streaming platforms curate their content and compete for audience attention.

Note: ChatGPT was used to improve the grammar and punctuation of the lines I've written for this project proposal.