Project Deliverable 3: Business and Question

Team: Adetutu Daranijo, Krystal Branker, Paul Smith

CSV:

Credits.csv

Title.csv

MetaData:

Titles.csv:

- Title: The name of the title.
- ID: The title ID on Justwatch
- Show Type: TV show or movie.
- Description: A brief description.
- Release Year: The release year.
- Age Certification: The age certification.
- Runtime: The length of the episode (SHOW) or movie.
- Genres: A list of genres.
- Production Countries: A list of countries that produced the title.
- Seasons: Number of seasons if it's a SHOW.
- IMDB ID: The title ID on IMDB.
- IMDB Score: Score on IMDB.
- IMDB Votes: Votes on IMDB.
- TMDB Popularity: Popularity on TMDB.
- TMDB Score: Score on TMDB.

Credits CSV:

- person_ID: The person ID on JustWatch.

- id: The title ID on JustWatch.
- name: The actor or director's name.
- character_name: The character name.
- role: ACTOR or DIRECTOR.

Project Deliverable 2: Business and Question

Team: Adetutu Daranijo, Krystal Branker, Paul Smith

We are choosing to do the final project on HBO MAX, rebranded now as MAX, an entertainment streaming service. We plan on answering the following question: ***How has the popularity of HBO Max's content evolved over time, so we can identify any patterns or trends that would lead to growth and predict how well shows or movies will do ?*** By using Kaggle.com, we can use the titles.csv dataset provided by user Victor Soeiro. It has details for id (The title ID on JustWatch) , title (The name of the title) , show_type (TV show or movie), description (A brief description) , release_year (The release year) , age_certification (The age certification) , run_time (The length of the episode (SHOW) or movie) , genres (A list of genres), production_countries (A list of countries that produced the title), seasons (Number of seasons if it's a SHOW), imdb_id (The title ID on IMDB), imdb_score (Score on IMDB), imdb_votes (Votes on IMDB), tmdb_popularity (Popularity on TMDB), and tmdb_score (Score on TMDB). In this section, we will gather all the data from the dataset provided by kaggle. We will perform an EDA exploratory data analysis to understand the data by looking for outliers, distribution with box plots, missing values by using isna function , describe function to understand mean, max , min, standard deviation of each feature in the dataset. From here, we will start cleaning our dataset making sure columns are spelled correctly, dropping empty rows, and filling in data where necessary. Next, we will create visualizations using plotly , seaborn and matplotlib library from python to plot a interactive graphs such as line graphs,bar graphs , map plots, etc to

show how trends of the genre are performing the best, which genre has a higher rating, and the genre with highest amount of stars but not enough shows. In addition, we are going to perform linear regression to predict how well the show or movie will perform using imdb_score and indicate which features are significant in predicting that outcome. Last, provide recommendations to increase subscriptions based on our analysis.

https://developer.themoviedb.org/docs/popularity-and-trending

https://www.kaggle.com/datasets/victorsoeiro/hbo-max-tv-shows-and-movies/data?select=titles.csv