Lab2 - Team Turquoise

Samantha Williams, Kate Kostelc, Mick Rejniak, Sean Baughman

3/20/2022

Contents

0.1 Proposal 1
0.2 Dataset 1
0.3 Core Questions/Investigations: 1
0.4 Plan of Action: 1

0.1 Proposal

We work for a start-up streaming service company and want to maximize viewer subscriptions to our new platform by offering the most watched films. We will conduct an analysis of IMDB movie data that clarifies how different movie attributes affect the meta score of a movie. The examination will include factors such as gross, IMDB rating, run time, etc. to find a selection of the top 1000 movies that will increase viewer subscriptions to our platform.

0.2 Dataset

Our project will utilize Kaggle's IMDB Movies data set (https://www.kaggle.com/datasets/harshitshankhdhar/imdb-dataset-of-top-1000-movies-and-tv-shows) that contains 1,000 records with 16 columns that describe the top 1,000 highest rated movies by IMDb, an online entertainment database from 1920 through 2020. The data set contains 13 columns but for our purposes we will be interested in the following:

- Series_Title = Name of the movie
- Released_Year - Year at which that movie released
- Runtime - Total run time of the movie
- Genre - Genre of the movie
- IMDB_Rating - Rating of the movie at IMDB site
- Meta_score - Score earned by the movie
- Director - Name of the Director
- Star1,Star2,Star3,Star4 - Name of the Stars
- No_of_Votes - Total number of votes
- Gross - Money earned by that movie

0.3 Core Questions/Investigations:

As a new streaming platform interested in increasing our paid subscribers, we want to know does a commercially successful (Gross) movie become a highly rated movie (Meta_Score)?

0.4 Plan of Action:

To help us dive into this question we will use the variable Meta_score as our output and Gross as our primary beta. We will use additional variables such as IMBD_Rating, No_of_Votes, and possibly Runtime to further refine our model. We may also want to consider the impact of a director, genre and actor has on a films Meta_score using dummy variables. We will meet on Sundays at 10AM (Pacific) to discuss our initial EDA, build our models and write our analysis. We will also stay in constant communicate through Slack, Zoom and email.