



MOVIE RECOMMENDATION SYSTEM

PROD.

COMP.



BY:

Developed and Presented by Varshini Chellapilla for Springboard



OVERVIEW

The goal of this third capstone project is to create a recommendation system based on collaborative filtering using data from Letterboxd and TMDB (The Movie Database) to provide Letterboxd users with movie recommendations based on past ratings and other calculated trends.



DATA ACQUISITION AND WRANGLING

Letterboxd

Letterboxd is a global social network for grass-roots film discussion and discovery. Users can leave numeric and custom text reviews, create “playlists” of films, select their top 4 favorite films to be featured on their profile, and follow other users to keep up to date with their activity.

Dataset

Kaggle and Github datasets uploaded by Sam Learner. Three datasets: ***Ratings***, ***Users***, ***Movies***.

A black clapperboard with a white and black striped top bar. The text is in white. The top bar has three small white circles on the left side.

*SIZE OF
COMBINED
DATASET:*

~10,000,000

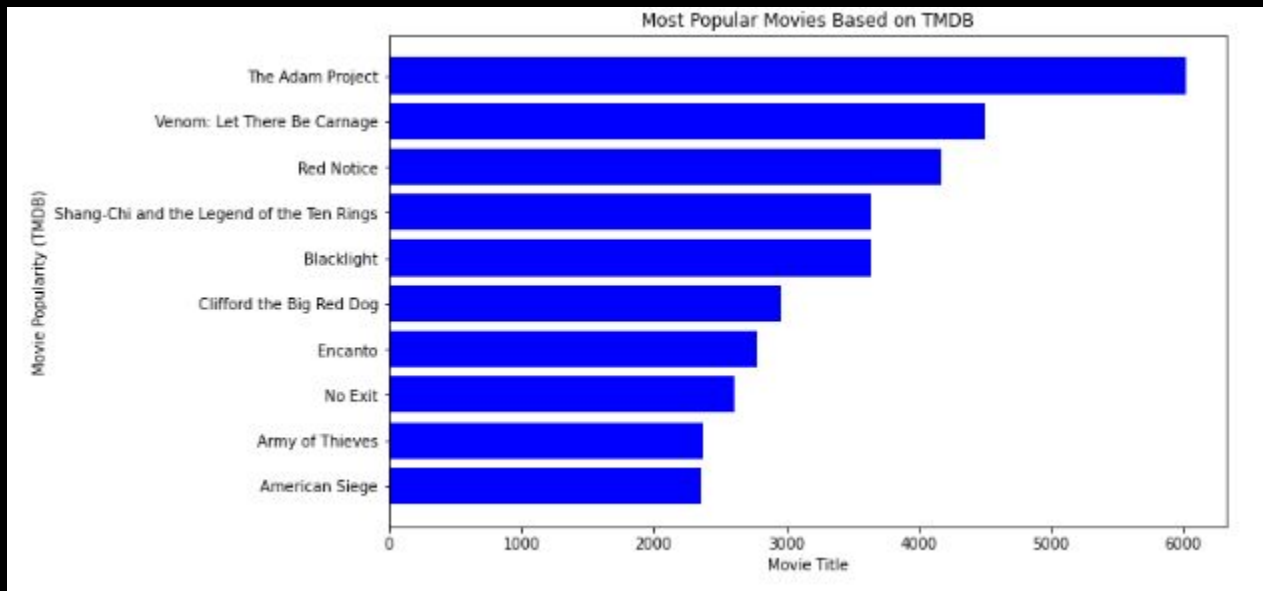


EXPLORATORY DATA ANALYSIS

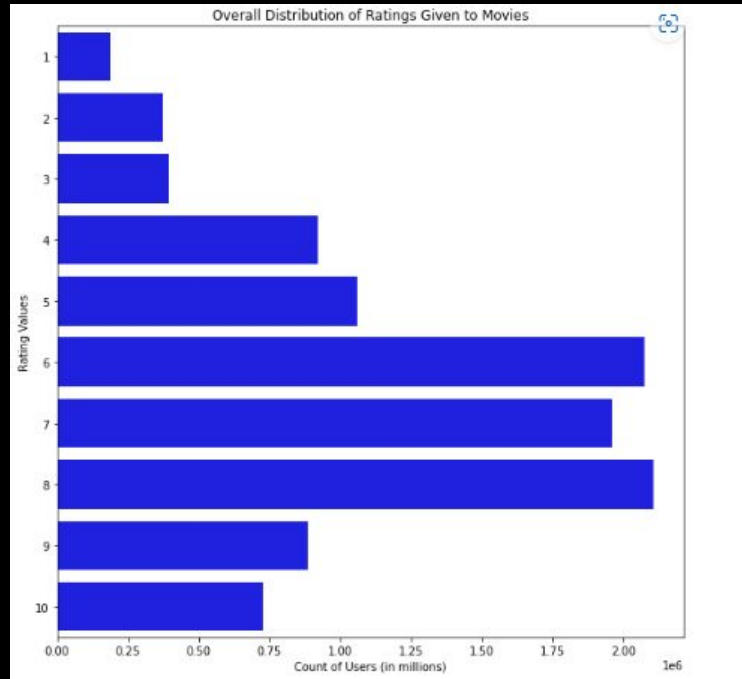
At the beginning of the Exploratory Data Analysis, the following questions were laid out:

- The distribution of ratings & number of reviews among movies, users, years, and genres
- The most and least popular movies and genres
- Changes in popularity of genres and movies over time

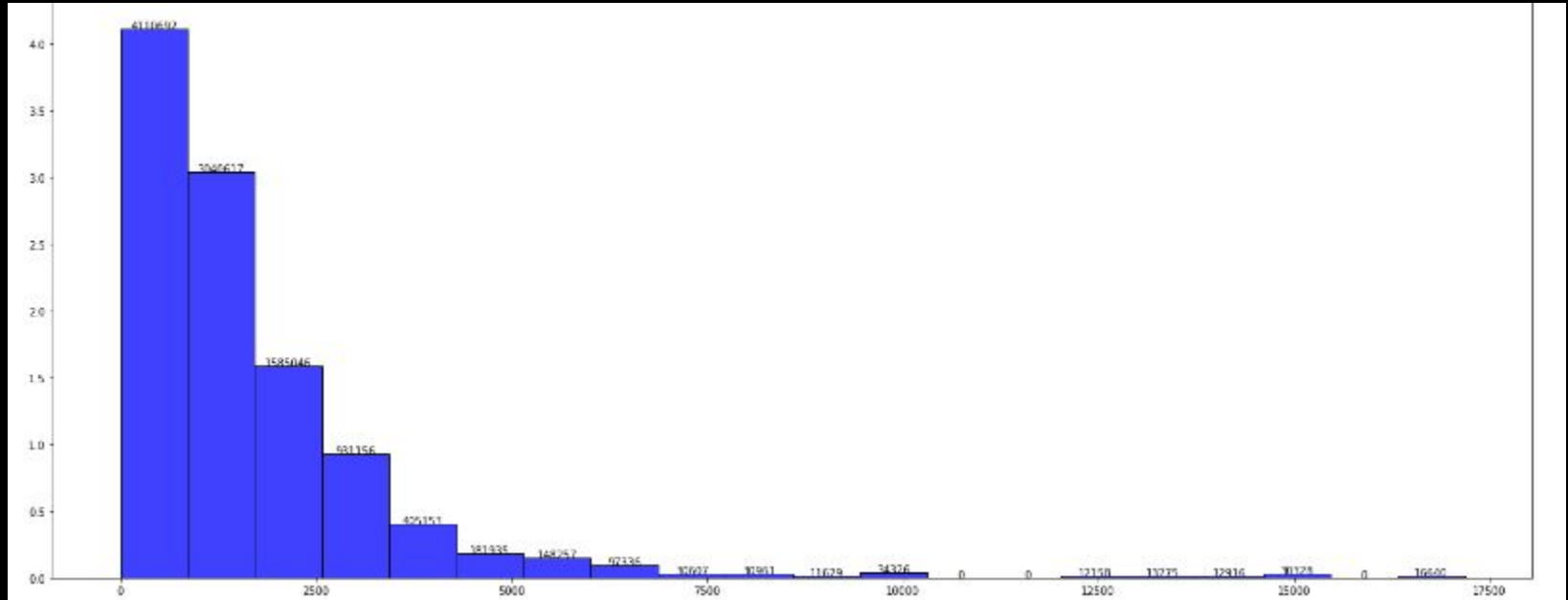
EDA: MOST POPULAR MOVIES



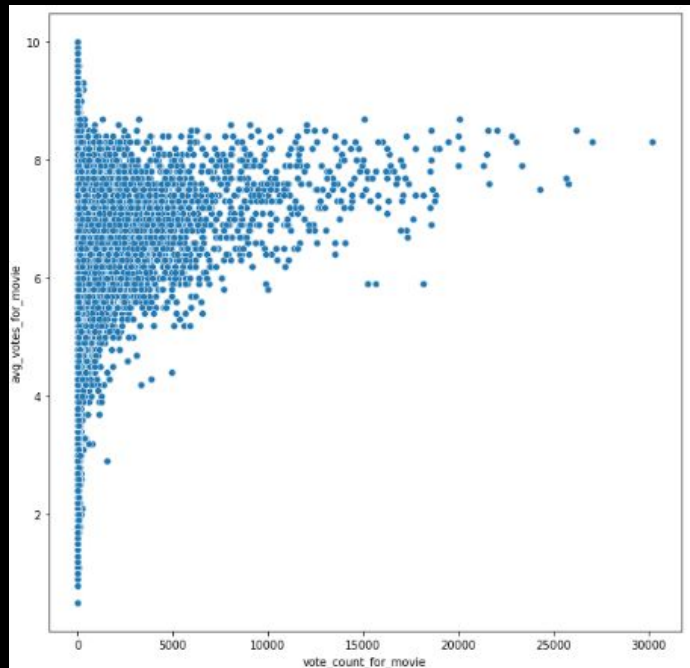
EDA: DISTRIBUTION OF RATING VALUES



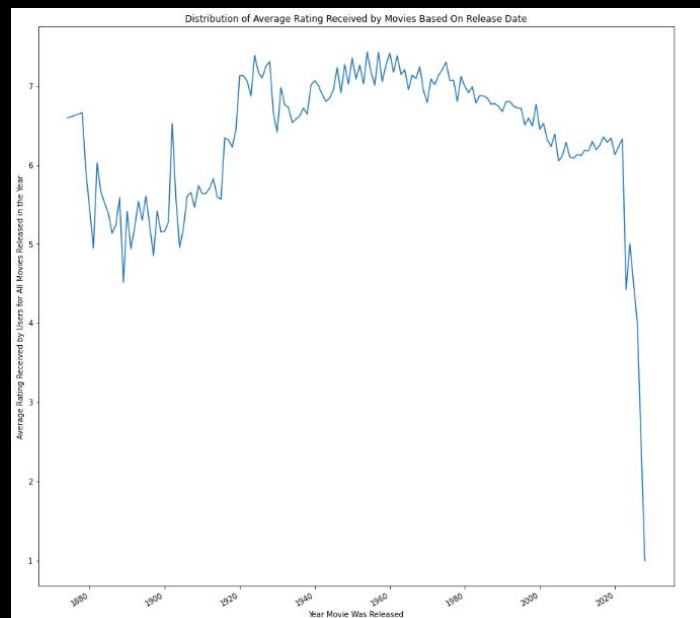
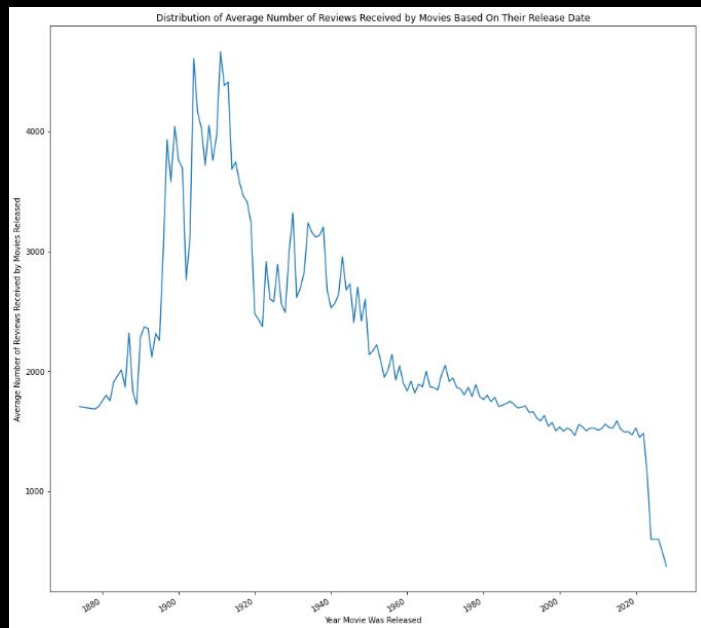
EDA: DISTRIBUTION OF NUMBER OF REVIEWS



EDA: NUMBER OF VOTES PER MOVIE AND AVG. RATING RECEIVED

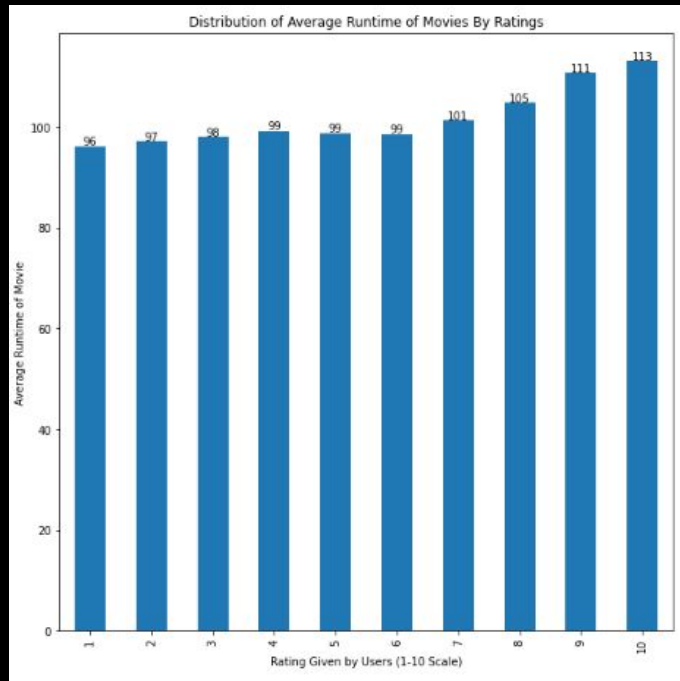


EDA: RATINGS AND REVIEWS BASED ON RELEASE DATE

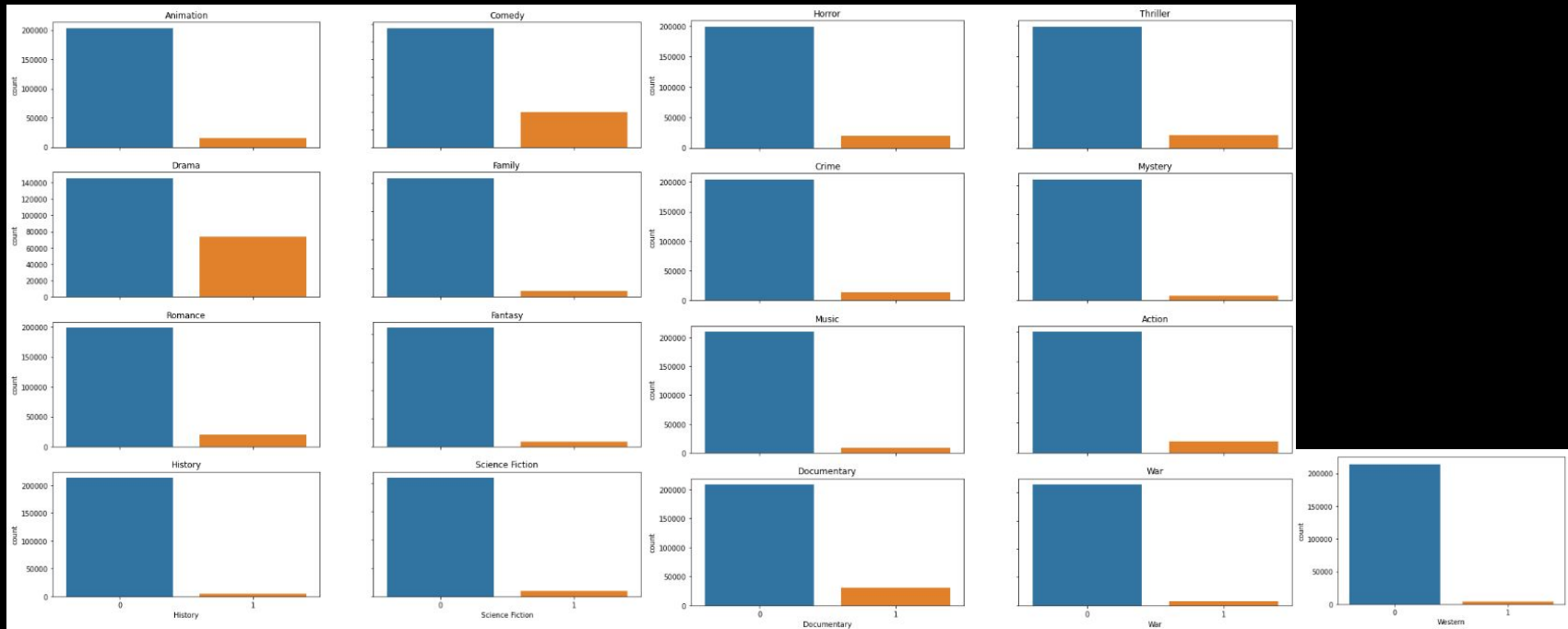


EDA:

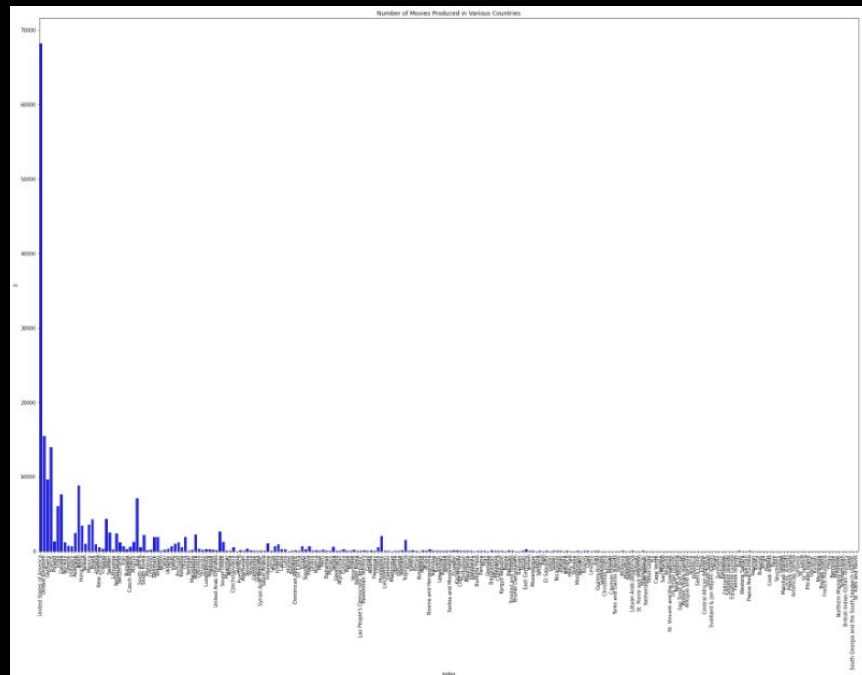
DISTRIBUTION OF RUNTIME BY RATINGS



EDA: UNIQUE MOVIES IN EACH GENRE




























EDA: MOVIES BY PRODUCTION COUNTRY





DATA MODELING

**USER-BASED,
MODEL-BASED,
COLLABORATIVE FILTERING,
RECOMMENDATION SYSTEM**

					
A					
B					
C					
D					
E					

Via [Recommendation Systems : User-based Collaborative Filtering using N Nearest Neighbors](#) | by Ashay Pathak | SFU Professional Computer Science | Medium



MODELS PURSUED

STEP 1:

Models



SVD



SVD++



KNNWithMeans



NMF



CoClustering

STEP 2:

Hyperparameter Tuning

After selecting a model, tuning some of the hyperparameters to produce better results and performance.

STEP 3:

Custom Function

A custom function that takes the trained, tested and tuned model and dataset as input and output recommendations.



THANKS!

TAKE:

CREDITS: This presentation template was created by **Slidesgo**,
including icons by **Flaticon** and infographics & images by **Freepik**

SCENE:
