Midterm Project

Peng Liu

**Personal Statement:**

After graduation, I want to get a job in data analysis for an Internet company. In my opinion, this project will be a good practice and learning process for my future career. As a pioneer in the streaming service industry, Netflix is an excellent dataset for me. On this basis, I chose the "Latest Netflix data with 26+ joined attributes" data set for my project. This is a very comprehensive dataset, which includes sources from Netflix, Rotten Tomatoes, IMBD, posters, box office information, trailers on YouTube, and more using a variety of APIs. But I will focus on IMBD ratings to analyze movies on Netflix

**Question:**

* What are the factors that affect the IMDB of the film?
* What type of film has a higher IMDB?
* What factors have affected Netflix's original programs?

**The data sources:**

This dataset has more than 10,000 unique values, and I will use nearly five years of data for the project.

“Latest Netflix data with 26+ joined attributes” https://www.kaggle.com/ashishgup/netflix-rotten-tomatoes-metacritic-imdb

**Proposed timeline of work:**

* EDA: 11/5 - 11/10
* Data Processing: 11/10 - 11/15
* Modeling and Validation: 11/15 – 11/20
* Write up: 11/20 – 12/1