Sean McLean

Catherine Smereena Dommaty

Susheel Reddy Nelvoy

ALY 6040 – Module 1 Final Project

Our final project will be on a dataset from Kaggle that provides all the television shows and movies that are available for streaming on Netflix. There are 8,807 entries in the dataset with 12 variables that describe the movie or television shows on the streaming service. A few of the attributes that are the most intriguing are the director of the movie or television show, its release year, its rating, the runtime, what it is listed under in terms of genre, and its description. Other than the release year, which is an integer, the data type for the rest of the variables are strings.

What interested us with this dataset is the amount of variety it contains and how many different approaches or angles we can take when looking at the data. The size of the dataset provides the opportunity to understand user behavior and what kind of content they choose to watch. We hope to identify trends in how people use the streaming platform and how Netflix uses this data to build recommendation systems for each subscriber. We would also be curious how this impacts future decision-making regarding bringing in new television shows and movies and attracting new subscribers.