CSE 564 Assignment 1 Report

The dataset that I have chosen for this homework assignment is the Movies Dataset available in Kaggle (https://www.kaggle.com/rounakbanik/the-movies-dataset) as movies_metadata.csv.

This dataset contains various interesting features pertaining to movie information such as:

1. Adult - defines if a movie is an adult movie or not (binary, categorical)
2. Budget - defines the budget of a specified movie (numeric)
3. Genres - defines the genres of the various movies (categorical)
4. IsCollection - defines if a movie is part of a collection or not (binary, categorical)
5. Original Language - defines the original language the movie was made in (categorical)
6. Popularity - defines the popularity score of the movie (numeric)
7. Production Companies - defines the various companies which produced the movies (categorical)
8. Release Date - defines the release date of the various movies (numeric)
9. production countries
10. Revenue - defines the amount of money earned by the movie (numeric)
11. Runtime - defines the duration of the movies (numeric)
12. Available Languages- defines the various language the movie was released in (categorical)
13. Status - defines if the movie was released or not (binary,categorical)
14. Vote average/Ratings - defines the ratings received by the audience (numeric)
15. Vote Count - defines the number of people who voted for the movies (numeric)

The dataset also consists of other features which were discarded for the purposes of this assignment as they not useful/ were beyond the scope of this assignment such as:

1. Homepage
2. Id
3. Imdb Id
4. Original Title
5. Video
6. Overview
7. Poster path
8. Tagline
9. Title

The reason I chose this dataset is because it can be useful to identify patterns between the various features (by means of exploratory data analysis and visualization) as well as be used in building different types of recommender systems which are used by various companies such as Netflix, Amazon Prime, etc

The code is divided as follows:

- Index.html : consists of the main page
- final datas : consists of csv data files obtained after data cleaning in python
- MovieDataCleaning.py : consists of code which was used to clean data in python
- src/app.js : consists of the various functions such as data reading and calling plotting functions
- src/barchart.js: consists of code for plotting barcharts
- src/histogram.js : consists of code for plotting histograms
- src/scatterplot.js : consists of code for plotting scatter plot

Link to video :
https://drive.google.com/file/d/1DeyEuDyrfCi7r9RFUgcZWWQifOU_Ud5B/view?usp=sharing