## US Homelessness EDA

This project is an in-depth exploration data analysis on Homelessness data from US Housing and Urban Development to identify the variables that can predict homelessness and run correlation analysis on multiples features using R programing to measure the relationship of multiple features to homelessness population in united states.

1. US Population by State

2. Unemployment by State

3. Poverty by State

4. Percent of Adult with Depression by State

5. Violence and Crime Rate

6. Available Public support like number of bed available to shelter homelessness is effective in decreasing

Homeless

7. Weather

8. Housing Rental Rate

## Housing Sales Price Prediction

This assignment performed a single and multiple regression models to predict housing sales price in Washington State using R programming. I looked at multiple features that influence the house price and analyzed the Coefficient of Determination, Confidence interval, Anova Test result and validated multicollinearity.

## MLB Attendance and Recommendation

This project looks into the MLB details for an entire MLB season played at Dodger Stadium. This aims to identify the different factors that could impact the high attendance in Dodger Stadium and provide actionable insight to management to improve game attendance. The Exploration data analysis uses Python programming.

## Yelp Dataset and Wikipedia Restaurant Information Data Wrangling

In this project, I have demonstrated the process of data wrangling of Yelp business dataset and Wikipedia restaurant information which involved working on data sources, including CVS, JSON, Website data scraping and API call. The data was cleaned, transformed, organized and stored in a database using Python code.

## Movie Recommendation System

This assignment uses movies dataset and using title, genres, user ratings and tags, I have performed exploration data analysis and built recommendation system using collaborative filtering approach. In this, the system looks for users who share the same rating patterns as the user whom the prediction is for using K-Nearest Neighbors in Python Programming.

## ALS Data Clustering (Unsupervised Learning)

In this project, I used the historical data for Amyotrophic lateral sclerosis (ALS) and performed a k-means clustering model to uncover groups of similar patterns within ALS dataset using Python programming.

## Prosumer Electric Consumption Prediction (Timeseries)

This project built a machine learning model to predict energy consumption and production for a prosumer for a specific period of time and understand the different features and variables that have a substantial weight in predicting energy targets using Python programming.

## Gym Member Workout Calories Burn Prediction

This project built a machine learning regression modal that predict calories burned by considering critical metrics with substantial weight in estimating fitness progression.

## DSC 680 Project Milestone 2 (TBD)

## DSC 680 Project Milestone 3 (TBD)