README FILES

1) Predicting Type 2 diabetes to identify potential high-risk patients.

   A Kaggle dataset that consisted of over 200,000 different patients that have type 1 diabetes, type 2 diabetes and non-diagnosed were analyzed to see if there were any key indicators (demographics, and lab results) that shows correlation to predict diabetes. Exploratory data analysis was done on the dataset and then prepped for modeling. The data was split into a 75/25 train and test. Linear regression, Lasso, Naive Bayes, Random Forest Classifier, and other models were all done to see if it can achieve the highest accuracy of over 90% to predict the next generation of diabetic patients.

2) Crime Rates: Are selected crimes more dominant in different states compared to others?

   Crime rates from multiple sources such as website data, flat file and API that measure if selected crimes occur more likely in certain states than others. This analysis was done utilizing correlation and creating dummy variables for categorical data. It shows how to obtain different data that is available and how to unify it to one usable dataset.