Variables

|  |  |
| --- | --- |
| diabetes\_all | Raw data with all values |
| diabetes\_male | Raw data for gender “Male” |
| Diabetes\_female | Raw data for gender “Female” |
| age\_boxplot | Boxplot for age variable for all values |
| male\_age\_boxplot | Boxplot for age variable for Male |
| female\_age\_boxplot | Boxplot for age variable for Female |
| corr\_all | Correlation matrix for all values |
| corr\_male | Correlation matrix for Male |
| corr\_female | Correlation matrix for Female |
| corr\_all\_plot | Plot of correlation matrix for all values |
| hist\_age | Histogram of variable age for all values |
| hist\_age\_male | Histogram of variable age for male |
| hist\_age\_female | Histogram of variable age for Female |
| ageGrp | New variable for age for all values grouping age into 10 year increments |
| dia\_all | All values with the new ageGrp variable added and the age variable removed |
| dia\_male | Male values with the new ageGrp variable added and the age variable removed |
| dia\_female | Female values with the new ageGrp variable added and the age variable removed |
| diabetes\_input\_all | All values with ageGrp but class and age variables removed |
| diabetes\_input\_male | Male values with ageGrp but class and age variables removed |
| diabetes\_input\_female | Female values with ageGrp but class and age variables removed |
| dia\_all\_one\_hot | All values with ageGrp included, class/age removed, with one hot encoding applied |
| dia\_male\_one\_hot | Male values with ageGrp included, class/age removed, with one hot encoding applied |
| dia\_female\_one\_hot | Female values with ageGrp included, class/age removed, with one hot encoding applied |
| kmeans\_all | K means classification for all values with a K value of 2 |
| kmeans\_male | K means classification for all malewith a K value of 2 |
| kmeans\_female | K means classification for all female with a K value of 2 |
| kmeans\_all\_table | Table showing results of kmeans for all values |
| kmeans\_male\_table | Table showing results of kmeans for all male |
| kmeans\_female\_table | Table showing results of kmeans for all female |
| shuffle\_index\_all | Index to shuffle all values to ensure consistent class variable distribution in train/test |
| shuffle\_index\_male | Index to shuffle all values to ensure consistent class variable distribution in train/test |
| shuffle\_index\_female | Index to shuffle all values to ensure consistent class variable distribution in train/test |
| dia\_all\_dt | Shuffled dataset of all values for decision tree analysis |
| index | To hold the 70/30 train/test split |
| train\_all | 70% of all values |
| train\_male | 70% of all values male |
| train\_female | 70% of all values female |
| test\_all | 30% of all values |
| fit\_all | Decision tree analysis model for all values |
| fit\_male | Decision tree analysis model for male values |
| fit\_female | Decision tree analysis model for female values |
| predict\_all | Predicted value from decision tree analysis all values |
| predict\_male | Predicted value from decision tree analysis male values |
| predict\_female | Predicted value from decision tree analysis female values |
| predTable\_all | Table showing results of decision tree for all values |
| predTable\_male | Table showing results of decision tree for male values |
| predTable\_female | Table showing results of decision tree for female values |
| acc\_dt\_all/male/female | Accuracy for decision tree for all/male/female values |
| pre\_dt\_all/male/female | Precision for decision tree for all/male/female values |
| sen\_dt\_all/male/female | Sensitivity for decision tree for all/male/female values |
| nb\_model\_all | Naives Bayes model for all values |
| nb\_model\_male | Naives Bayes model for male values |
| nb\_model\_female | Naives Bayes model for female values |
| nb\_all\_predict | Predicted values for class from Naives Bayes for all values |
| Nb\_male\_predict | Predicted values for class from Naives Bayes for male values |
| nb\_female\_predict | Predicted values for class from Naives Bayes for female values |
| acc\_nb\_all/male/female | Accuracy for Naives Bayes for all/male/female values |
| pre\_nb\_all/male/female | Precision for Naives Bayes for all/male/female values |
| sen\_nb\_all/male/female | Sensitivity for Naives Bayes for all/male/female values |
| train\_all\_hot | Data table of train\_all |
| train\_male\_hot | Data table of train\_male |
| train\_female\_hot | Data table of train\_female |
| train\_all\_log | Training dataset for all data one hot encoded |
| train\_male\_log | Training dataset for male data one hot encoded |
| train\_female\_log | Training dataset for female data one hot encoded |
| linreg\_all | Linear regression model for all values |
| linreg\_male | Linear regression model for male values |
| linreg\_female | Linear regression model for female values |
| test\_all\_hot | Data table of test\_all |
| test\_male\_hot | Data table of test\_male |
| test\_female\_hot | Data table of test\_female |
| test\_all\_log | Testing dataset for all data one hot encoded |
| test\_male\_log | Testing dataset for male data one hot encoded |
| test\_female\_log | Testing dataset for female data one hot encoded |
| pred\_lm\_all | Predicted values for linear regression for all values |
| pred\_lm\_male | Predicted values for linear regression for male values |
| pred\_lm\_female | Predicted values for linear regression for female values |
| errors\_all/male/female | Fit errors from linear regression for all/male/female values |
| rmse\_all/male/female | Root mean square errors for all/male/female values for linear regression |
| pred25\_all/male/female | Predicted values from linear regression for all/male/female values that have less than 25% error |