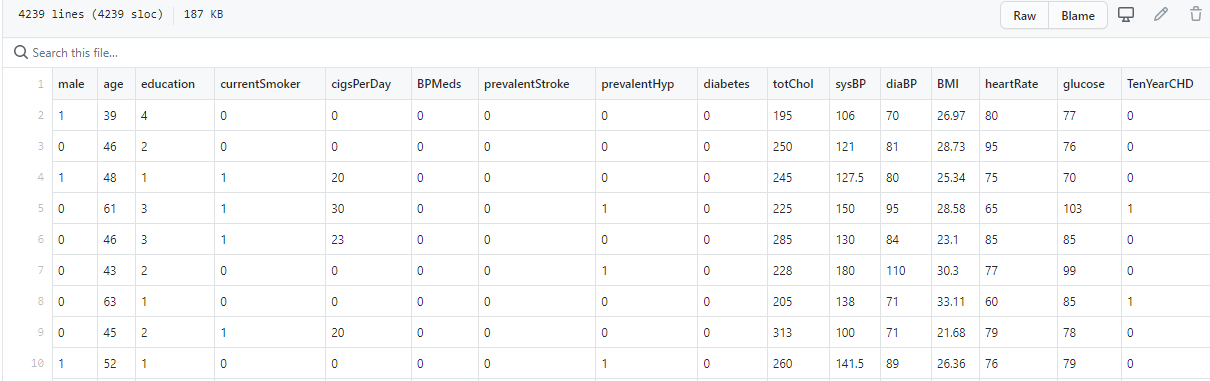
**DATA SETS**

Topic: Female Heart Disease

Resources: Framingham.csv (Kaggle) ~ 2,419 rows 16 columns



ETL – used Jupyter notebook

* Checked columns
* Checked datatypes
* Drop null values

1. ETL clean data set ~ heart\_disease\_females.csv ~ 2,034 rows 16 columns

Grouped the dataset to the following

* Demographics.csv (age, education)
* Habits.csv (currentSmoker,cigsPerDay)
* Pre-existing\_conditions.csv (BP Meds, prevStroke, prevalHyp, diabetes)
* Parameters.csv (totChol, sysBP,diaBP, BMI, heartrate, glucose)

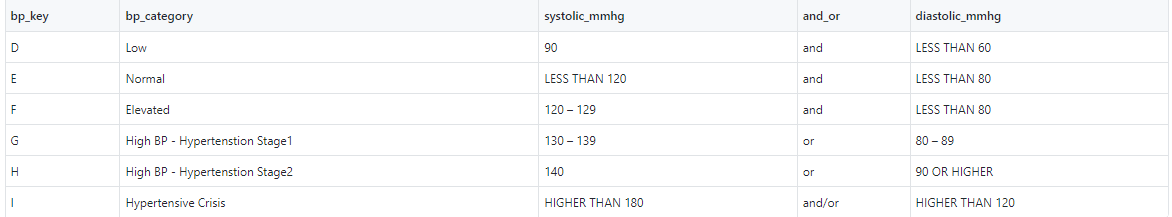
ETL:

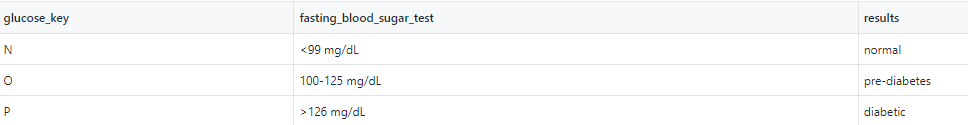
Additional dataset added ~ join to create a new data set with predictor

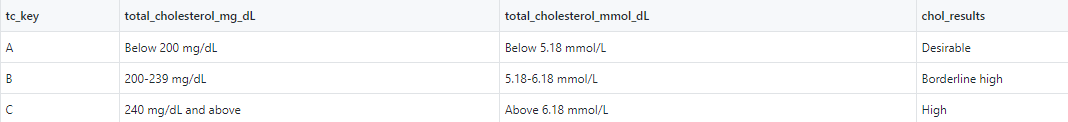
* Bmi\_pks.csv



* Bp\_pks.csv



* Glucose\_pks.csv
* Totlchol\_pks.csv



1. Clean\_FHD\_predictor\_dataset.csv ~ 2,034 rows 20 columns