DATA CLEANING CODE

#Define Average Age for Calculating Age Buckets

import statistics as stats

avgage = stats.mean(data.age)

#Define Age Standard Deviation

stdage = stats.stdev(data.age)

#Set Up Age Buckets and New Column, Based on Average and Standard Deviation

data["age\_bucket"] = "NaN"

data.loc[(data["age"]< avgage - stdage), "age\_bucket"] = "1"

data.loc[(data["age"] > avgage - stdage) & (data["age"]< avgage), "age\_bucket"] = "2"

data.loc[(data["age"] > avgage) & (data["age"]< avgage + stdage), "age\_bucket"] = "3"

data.loc[(data["age"] > avgage + stdage), "age\_bucket"] = "4"

data.head()

#Impute Missing Values for Cholesterol

data.fillna(data.mean(), inplace=True)

data.head()