-I viewed “.describe and .info the data to see what needs to be done in the dataset.

-Also the .unique values were to see if there’re unexpected values in the dataset

-the preprocessing like dropping duplicates and filling the nulls and encoding the categorical columns were necessary for cleaning the data.

-isnull is for knowing the counts for the nulls in all columns

-I transformed the datatypes from objects to perform operations on them like correlations

- strip is for deleting the spaces and the percentages of the columns

-caffeine needed extra cleaning as it included ‘varies’ word and needed to be turned into a value by the mode function.

- the histograms were needed for checking the distribution of every feature

-Features selection: the correlational matrix and the drop of the cholesterol feature was because of the high correlation between it and the calories and sugar features.

-for the visualization I used groupby to group the sugar and the calories columns based on their beverage as required by taking the average and collecting the repeated values of the common one.

How to run the solution:

-download the dataset and the solution at the same place on your computer

-open the solution and run it.