Documentation:

1. Importing Dataset from current folder of py
2. EDA (Exploratory Data Analysis) and Data Preparation
    1. Check dataset columns information with Menu.info()
    2. Rename some columns to standard naming features
    3. Check Values of each column to see categorical, numerical, similar, and missing values
    4. Convert Categorical of numerical format to numerical values with column name indicating its value
    5. Filling missing value in Caffeine with 0 indicating not existence means 0 value in drink
    6. Replace undetermined value with Median value integerated with Mode to insure confirmation of the value in Caffeine
    7. Remove Nutrition values duplicates from Dataset: in "Tazo® Tea"
    8. Drop unnecessary with no addition or minimal information gain: "Beverage_category"

3. Data Visualisations
    1. Plot Beaverges (Drinks names) with Calories value to extract Drink with highest value of Calories: to determine that: Grande 2% Milk White Chocolate Mocha (Without Whipped Cream) is the highest Calories drink
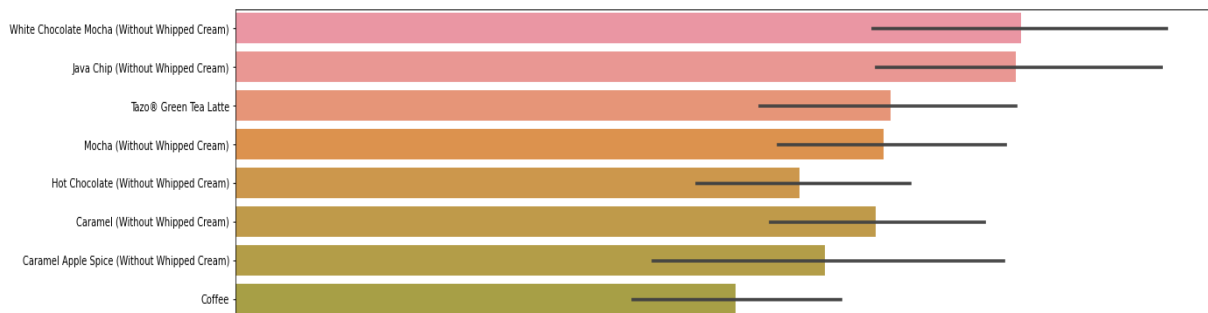

Figure 1: Cropped barplot of Calories vs Drink name

Figure 2: Cropped graph of Calories vs prep type of White Chocolate Mocha (Without Whipped Cream)

2. Plot Beaverges (Drinks names) with Sugars value to extract Drink with highest value of Sugars: to determine that: Venti Nonfat Milk Java Chip (Without Whipped Cream) and Venti Whole Milk Java Chip (Without Whipped Cream) are the highest Sugars Drinks
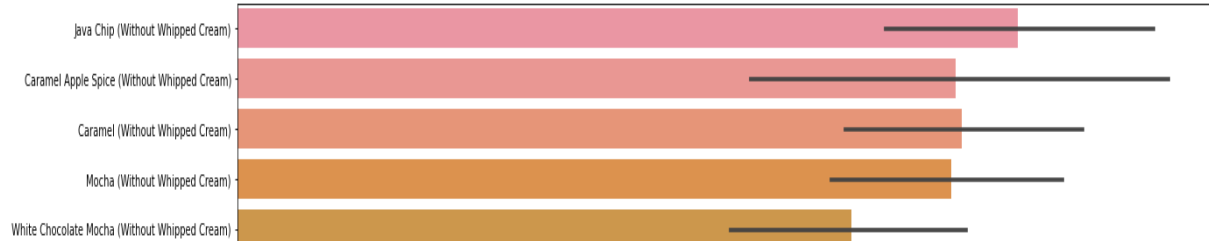

Figure 3: Cropped barplot of Sugars vs Drink name

Figure 4: Cropped graph of Sugars vs prep type of Java Chip (Without Whipped Cream)

Instructions:

1. Copy the Git Repo address into any service to run Notebook
2. Add in the same folder the CSV file of "drinkMenu.csv"
3. Copy path of "drinkMenu.csv" and replace in pd.read_csv: (Only if necessary)
4. Run the Cells of Notebook and read Markdown for further information