**Prog6280b**

**(Starbucks)**

**Program Description:** This lab program will guide you through using Pandas and Matplotlib in Python (or equivalent dataframe/plotting libraries in R, Julia, etc.) to analyze and visualize the “[Starbucks](https://www.kaggle.com/datasets/henryshan/starbucks)” dataset. The goal is to help you learn data wrangling and visualization techniques while uncovering interesting insights from the dataset.

1. Data Wrangling Exercises

* Remove leading/trailing spaces in column names.
  + df.columns = df.columns.str.strip()
  + Optionally, replace spaces in columns with underscores:

df.columns = df.columns.str.replace(' ', '\_')

* Handling missing values (if any).
* Converting data types (e.g., converting numbers to the correct format).
* Normalizing data (if applicable).

2. Data Visualization Tasks

Each task should guide you to tell a “story” with the data using various types of plots:

**Task 1: Caloric Content Across Beverage Categories**

* Plot Type: Bar plot (horizontal; use ‘barh’ in Pandas).
* Objective: Compare the average number of calories in each beverage category.
* Columns Used: Beverage\_category, Calories.

**Task 2: Fat Content in Different Beverages**

* Plot Type: Scatter plot.
* Objective: Explore the relationship between total fat and saturated fat in beverages.
* Columns Used: Total Fat (g), Saturated Fat (g).

**Task 3: Caffeine Content Variation**

* Plot Type: Histogram.
* Objective: Understand the distribution of caffeine content across all beverages.
* Columns Used: Caffeine (mg).
* **Normalize: coerce any non-numeric values to NaN and sort (ascending).**

# Replace ‘varies’ and ‘Varies’ with NaN

sorted\_df = df.sort\_values(by='Caffeine (mg)', ascending=True)

sorted\_df['Caffeine (mg)'] = pd.to\_numeric(sorted\_df['Caffeine (mg)'], errors='coerce')

**Task 4: Nutritional Content Comparison**

* Plot Type: Box plot.
* Objective: Analyze the spread of protein, sugars, and dietary fiber across beverage categories.
* Columns Used: Beverage\_category, Protein (g), Sugars (g), Dietary Fibre (g).

**Task 5: Vitamin and Mineral Contribution**

* Plot Type: Stacked bar plot (horizontal).
* Objective: Show the contribution of Vitamin A, Vitamin C, Calcium, and Iron in different beverage preparations.
* Columns Used: Beverage\_prep, Vitamin A (% DV), Vitamin C (% DV), Calcium (% DV), Iron (% DV).

3. Analysis and Interpretation

* Interpret the plots (one sentence each).
* Discuss possible health concerns based on the caloric/fat content, etc.

**Data Location:** starbucks.csv

**Sample Output** **(not including interpretations):**

Task 1: Caloric Content Across Beverage Categories

A graph of a bar chart

Description automatically generated with medium confidence

Task 2: Fat Content in Different Beverages

A graph of fats and beverages

Description automatically generated

Task 3: Caffeine Content Variation

data:image/png;base64,

Task 4: Nutritional Content Comparison

A group of colorful boxes

Description automatically generated with medium confidence

Task 5: Vitamin and Mineral Contribution

data:image/png;base64,