Exploring Fast Food

For this challenge we will be using the fast food dataset (Fast_Food.csv) located at:

`https://www.kaggle.com/datasets/rakkesharv/fast-food-joint-nutrition-values-dataset`

The variables in the dataset are as follows:

- **Company**: The name of the fast food chain.
- **Category**: Category of item in the fast food chain.
- **Product**: The name of the menu item.
- **Serving Size**: The serving size of the menu item in grams.
- **Calories**: The number of calories in the menu item.
- **Carbohydrates**: The amount of carbohydrates in grams in the menu item.
- **Protein**: The amount of protein in grams in the menu item.
- **Fiber**: The amount of fiber in grams in the menu item.
- **Sugar**: The amount of sugars in grams in the menu item.
- **Total Fat**: The amount of fat in grams in the menu item.
- **Saturated Fat**: The amount of saturated fat in grams in the menu item.
- **Trans Fat**: The amount of trans fat in grams in the menu item.
- **Cholesterol**: The amount of cholesterol in miligrams in the menu item.
- **Sodium**: The amount of sodium in milligrams in the menu item.

Complete the following Questions regarding the dataset:

1. Visualize and describe in full the distribution for each of the following variables:
    - Serving Size
    - Calories
    - Fat
    - Trans Fat
    - Sodium
    - Carbohydrate
    - Sugars
    - Protein

2. For each of the following variables list the observations that are the top 5:

    - Fat
    - Calories
    - Protein
    - Serving size
    - Fat per gram serving size
    - Calories per gram serving size

3. Summarize each of the following:

    - Average serving size for each company
    - Average fat content for each company
    - Average protein content for each company
    - Average calories for each company
    - Average fat per gram for each company

4. Use visualizations/statistics to explain association between the following variables:

    - Brand and calories
    - Brand and fat
    - Brand and protein
    - Protein and energy (calories)

5. A body builder knows that they eat 200 g of protein a day. Build a linear regression model to predict the average number of calories the body builder expects to eat.

6. Are the variables 'company' and 'calories' *independent*? Explain heuristically what independence between these two variables would mean.

7. Suppose you order a *random* item from the dataset. Determine each of the following probabilities:

    - $P(\text{Company is Macdonalds})$
    - $P(\text{Fat Content} > 15\text{g})$
    - $P(\text{Fat Content} > 30 \cap \text{Protein Content} > 20)$
    - $P(\text{Fat Content} > 20 \,|\, \text{Macdonalds})$