**2023 Data Derby**

**Minnesota**

**DATAGON Rules**

**Background**

In keeping with this year's DataDerby theme, food. Everybody nowadays is mindful of what food they eat. Counting calories and watching nutrition facts is the number one advice given by all dieticians and nutritionists. The calories in food come from nutrition such as carbohydrates, protein, fat, and saturate fat. Accordingly, you will consider type of food nutrition/food category in predicting healthiest food. To make such analysis valid, you are provided data sets representing the nutritional facts and calories for most common food.

**Data description**

The dataset contains a csv file with 299 foods each with the nutrient including Calories, Fats, Proteins, Saturated Fats, Carbohydrates, and Fibers. Also, the foods are also categorized into various groups like Desserts, Vegetables, Fruits etc.  
The 'Food' column denotes name of the item and 'Measure Cup' denotes quantity of the food item used to find the other nutrients.

Note: "t" indicates that only a trace amount is available(miniscule). Please treat those values as 0. Your challenge is to analyze food nutrition and do Exploratory Data Analysis (EDA) of given dataset.

If you provide any analysis or answers to the questions not for your division, your team will be classified at a higher division.

**Tasks**

**High School Division**

If you participate in the high school division, you will find and visualize category wise per nutrient and predict the healthiest foods (High protein low carbs).

1. Exploratory analysis of the data and visualizations. Your ultimate goal is to find food category wise visualization per nutrient (referring to the dependent variables of Calories, Protein, Fat, Saturated Fat, Fiber and Carb). Generate appropriate graphs to help with the analysis and prediction. A minimum of six graphs are required, with one graph for each dependent variable.
2. Your second task is to find the healthiest foods (High protein low carbs). Generate appropriate graph to find the result. Explain and perform all necessary analysis to support your results.

**Novice Division**

If you participate in the novice division, you will visualize distribution of features with one another and predict most protein rich food in the category of vegetables and grains food that has the most calories with proper visualizations.

1. Exploratory analysis of the data using visualizations. Your ultimate goal is to start the analysis by plotting the features (Calories, Protein, Fat, Saturated fat, Fiber and Carbs) of each type of food. Generate appropriate graphs to help with the analysis and predictions. A minimum of six graphs are required, with each graph for one dependent variable.
2. Your second task is to compare the average protein level between the category of Fish, Seafood and the category of Meat, Poultry. You need to generate appropriate graphs based on your analysis. Provide complete and appropriate analysis to explain if there is a significant difference.
3. Your last task is to predict food that has the most calories. You need to generate appropriate graphs as evidence. Provide complete and appropriate analysis and rationale to explain your results.

**Advanced Division**

If you participate in the advanced division, you will find and visualize category wise visualization per nutrient, start the analysis by plotting the features with one another and predict foods that have the largest combination of calories and fats and analyze categories. Finally, you have to find the food group with the most calorie content.

1. Exploratory analysis of the data using visualizations. Your ultimate goal is to find food category wise visualization per nutrient (referring to the dependent variables of Calories, Protein, Fat, Saturated Fat, Fiber and Carb). Generate appropriate graphs to help with the analysis and prediction. A minimum of six graphs are required, with one graph for each dependent variable.
2. A certain amount of fat is required for a healthy gut. Let's look at some fatty foods. Your second task is to predict top 10 foods that have the largest combination of calories and fats. You need to utilize all necessary and appropriate predictor variables given in Datagon dataset for your analysis. Provide complete and appropriate analysis to explain your results.

## Your need to analyze the Drinks, Alcohol, Beverages and Desserts, use all predictor variables (Calories, Protein, Fat, Saturated fat, Fiber and Carbs) in Datagon dataset for your analysis. Provide complete and appropriate analysis and graphs to explain your results.

1. Find the top ten protein rich food in the dataset. Provide analysis and graphs to support your results.

## You need to analyze meat, poultry, seafood. You need to find meat/seafood with high fat content. Then you have to scatter 3D plot of Fatty Foods (% Daily Value) and carbohydrate rich foods. Provide complete and appropriate analysis and graphs to explain your results.

**What to turn in**

1. Your final report in a word file. Your report should include the following:

* A title of your report
* Your Team **number** (Please do NOT include any information about your team member names, your institution/school name, your advisor name, or your team name; points will be deducted if any of that information is disclosed to the judges)
* Methodologies used
* Answers to address all the questions in your division, in a format of a report
* Your final conclusion
* If necessary, please copy and paste your code to your report.

Please limit your report to 5-6 pages, including the graphs. If you include the code in your report, that does not count into the page limit.