Assignment 4

Due on July 11th, 2023
Introduction to Programming with Python
Columbia University Pre-College Program−Summer Session I
Instructor: Daniel Kadyrov

In this assignment, you will apply your python skills including `pandas` and the graphing libraries `matplotlib` or `plotly` to analyze the cereal dataset. The dataset is available at the following link: `https://www.kaggle.com/crawford/80-cereals`. The dataset contains information about 80 cereals. The dataset is also available on the course repository.

You can choose any graphing package to generate the graphs including but not limited to matplotlib, seaborn, or plotly. Make sure each plot is visually appealing, has a title, and has labeled axes. Make sure that the axis ticks make sense. Make each graph have a consistent theme and color scheme across all graphs.

Problem 1
Load the dataset into a `pandas` dataframe.

Problem 2
Print the first 5 rows of the dataset.

Problem 3
Use pandas to generate a description of the dataset.

Problem 4
Print the correlation matrix of the dataset. Which two columns have the highest correlation? Which two columns have the lowest correlation? Generate a correlation matrix plot using pandas.

Problem 5
Use pandas to plot a scatter matrix of the dataset.

Problem 6
Create a bar chart of the number of cereals in each manufacturer using the graphing library of your choice.

Problem 7
Generate a scatter plot of the calories vs. the rating of the cereal using the graphing library of your choice.

Problem 8
Select a few interesting looking graphs from the scatter matrix generated in problem 6 and make them visually appealing using the graphing library of your choice.

Problem 9
Create a histogram of the ratings of the cereals using the graphing library of your choice.

Problem 10
Describe any conclusions you can make about the dataset.