Introduction to Data Analysis (DATA 1200)
Assignment #1 – Descriptive Analysis (15% of Total Grade)
Professor: Ritwick Dutta

Mr. John Hughes has been reviewing the **IQAnalysis.csv** dataset.

The dataset has 40 observations and 4 variables:

Independent Variables:

Brain – circumference (cm.)
Height – in inches (in.)
Weight – in pounds (lbs.)

Dependent Variable:

PIQ – intelligence quotient (IQ)

The Ask:

Part A: Create a Python Script using Jupyter Notebook (then convert to .html) ***– 3%***

Using Python create the **Key Statistics (i.e., using .describe())** and **Standard Deviation Spread Graphs) for the three (3) independent variables** for the **IQAnalysis.csv** dataset

Part B: PowerPoint (.ppt or pptx) ***- 12%***

Create a PowerPoint (PPT) presentation that includes the following:

a) Cover Page (Title, Name (1st and last) and Student Number)
b) Complete the following analysis from the Standard Deviation Graphs. Remember that the below needs to be completed for each of the **three (3) independent variables**:
   i. Identify and Justify the type of distribution (i.e. Normal or Skewed) by reviewing the Central Tendency Metrics (i.e. Mean, Median and Mode)
   ii. Identify and Justify which Central Tendency Metric (i.e., Mean, Median or Mode) should be used to best describe the independent variable.
   iii. Present the Standard Deviation Spread Graphs and explain the results (i.e., how spread out are the datapoints).
c) Using the results from **Key Statistics (i.e., using .describe())**, Identify and explain **three (3) key generalized insights** that can be used to explain the entire dataset.

Hint: Leverage the Wk2c-Tutorial-DescStats

Please post your PowerPoint Document (.ppt or .pptx) and Jupyter Notebook in HTML (.html) format via assignments under Assignment #1 by

Tuesday, Oct 3rd, 2023 @ 11:59 p.m.

Rubric				
Question	**Exemplary (14-15)**	**Proficient (10-13)**	**Incomplete (7-9)**	**Needs Improvement (0-6)**
Part A ***(3%)***	Python HTML file is complete	Python HTML file is mostly complete. Missing headings or structure.	Python HTML file is incomplete. Incorrect use of heading or code.	Python HTML file is missing or incorrect.
Part B ***(12%)***	Three (3) independent variables with properly identified distribution, associated Central Tendency Metric, and STD Spread graphs with detailed explanations Three (3) generalized insights with detailed explanations	Three (3) independent variables with properly identified distribution, associated Central Tendency Metric, and STD Spread graphs with high-level explanations Three (3) generalized insights with high level explanations	Less than Three (3) independent variables with properly identified distribution, associated Central Tendency Metric, and STD Spread graphs with missing or limited explanations Less than Three (3) generalized insights with limited or missing explanations	Three (3) variable insights missing or incorrect Three (3) generalized insights missing or incorrect