**ADTA 5230**

**Lab 1 - Python Coding Help File**

First, to set up you need to have Python installed and upload the datafile from Canvas.

**Question 1:**

Show the appropriate software output of the summary statistics including missingness, min,  
max, mean, and standard deviation for all the numerical variables and number of levels for all  
the categorical variables.

**You should see output like this:**

A screenshot of a computer

Description automatically generated

**Question 2:**

How does the weekly sales compare between these stores? Specifically, which store has the  
highest weekly sale value overall, which has the highest median weekly sales, and which has the  
highest average weekly sales? Provide a box graph and a brief discussion to explain the graph.  
Remember to include your first and last name in the footnote of your graph.

**You should see output like this:**A screenshot of a computer

Description automatically generated

**Question 3:**

How does the weekly sales compare for the different weeks of the year? Specifically, which  
week has the highest weekly sale value overall, the highest median weekly sales, and the highest  
average weekly sales. Provide a box graph and a brief discussion to explain the graph.  
Remember to include your first and last name in the footnote of your graph.

**You should see output like this:**

A screenshot of a computer

Description automatically generated

**Question 4: (10 points)**

Make some comments on distribution of the weekly sales? What about its center and  
variability? Is it symmetric, right-, or left- skewed? Provide a histogram and descriptive statistics  
to support your claim. Remember to include your first and last name in the footnote of your  
graph.

|  |  |
| --- | --- |
| **A screenshot of a computer  Description automatically generated** | **A screenshot of a computer program  Description automatically generated** |

**Question 5: (10 points)**

Look at the pairwise scatter plot below. Generate the same pairwise  
scatter plot of all selected variables with different colors indicating the holidays. Then find two  
variables that show clear correlation with each other. **There are multiple correct answers!!**

**You should see output like this:**

A screenshot of a computer screen

Description automatically generated