Homework 3

1. Read the data from the CSV file into a DataFrame.

2. Display type, memory consumption, and null count information using the info() method.

3. Display the number of unique values in each column.

4. Display all the rows of data that JupyterLab displays by default.

5. Display the first and last five rows of data and the first and last four columns of data.

6. Choose any three columns, access them with bracket notation, and display the first five rows of this data.

7. Select one column and access it with dot notation.

8. Multiply the Total Volume and AveragePrice columns, and store the result in a new column called EstimatedRevenue. Then, display the first five rows of this data to confirm that the column was added and has the correct values.

9. Create a DataFrame that’s grouped by region and type and that includes the average price for the grouped columns. Then, reset the index and display the first five rows.

10. Create a bar plot that shows the mean, median, and standard deviation of the Total Volume column by year.

11. Create a new DataFrame that contains the total of the Small Bags, Large Bags, and XLarge Bags columns grouped by type, and then display the DataFrame.

12. Use the grouped data to create a bar plot that shows the number of small, large, and extra-large bags for both types of avocado.

13. Use the original data to create a scatter plot for the Total Volume and AveragePrice

columns.

Requirements and Submission:

1. Create a GitHub repository named CIS410\_Homework3.
2. Use at least 1 branch (in addition to the default main branch).
3. Merge all branches into the main branch prior to submission but do not delete any branches.
4. Include a .gitignore and README.md file in the repository.
5. Write a brief description of the results and include that in the body of the README.md file.
6. Submit a link to the repository in Blackboard.