Homework Assignment 1 – Exploratory Data Analysis

Dr. Gail Gilboa-Freedman
Dr. Naveh Eskinazi

Submission: 7/4/2022

GENERAL INSTRUCTIONS

The purpose of this exercise is to align all students with importing Python libraries and modules and perform basic exploratory data analysis.

The work will be based on a CSV named **"superstore_data.csv"** located in the course's Moodle site,

SUBMISSION:

Through assignment box within the course Moodle, submit a **Jupyter Notebook file named HWA1_<student name>.ipynb** (e.g. HWA1.karin_tenne.ipynb)
Should include all the relevant code needed to perform the assignment's tasks along with code's output.
(Recommendation: Add headers and sub-headers using the Markdown option)

Good Luck!

PART1: PREQUISITES

TASK 1: SETTING THE FOLDER

1. Create a new and blank Jupyter Notebook named **HWA1_<student name>.ipynb.**
2. Download from the CSV file named **"superstore_data.csv"** from Moodle.
3. Upload the CSV file to Jupyter (Note: make sure the file is placed places in the same location as your Jupyter Notebook)

TASK 2: IMPORT LIBRARIES & MODULES

4. Import the following libraries and modules within your notebook: **scipy, numpy, matplotlib, pandas, os, math, seaborn, matplotlib.pyplot, mpl_toolkits, and mpl_toolkits.mplot3d**

PART 2: EXPLORATORY DATA ANALYSIS

To complete the following tasks, use what you learned in the lecture and tutorial and rely on the **superstore_data** dataset.

TASK 3: DESCRIBE STATISTICS

Use Python commands (e.g., shape, describe bars, iloc) to plot and provide answers to the following questions:

5. How many rows and columns are in the data?
6. What are the categories of products sold in transactions **2 till 5**?
7. How many types of **Ship Mode** there are?
8. How many transactions are categorized as **Technology**?
9. What was the **order date** of the **latest** transaction made?
10. What was the **ship date** of **earliest** transaction made?
11. What was the **standard deviation** of **sales amount**?
12. What was the **lowest quantity** sold?
13. What was the **highest discount** given?
14. What was the **mean profit** made?
15. What was the **total profit** made?
16. What was the **number of transactions** made in the sub-category **Chairs**?
17. What was the **most** popular sub-category?
18. What was the **least** popular sub-category?

TASK 4: VISUALIZE STATISTICS

19. Using a Box & Whisker plot on the **Quantity** data, what represents the vertical line below the box, the vertical line above the box and the vertical line which goes through the box?
20. the line between the values 2 and 4 (shown on the Y-axis), and what's its value's?
21. Using a Histogram:
    a. What is the **least popular ship mode**?
    b. What is the **most popular ship mode**?
22. Using a Bar chart, what is the **name of the region** in which the sales were **the lowest**?

Good Luck!