Project

Deliverable: Jupyter Notebook containing the code and the outputs. Make sure to present the Jupyter Notebook as a report with headings and text to describe what you are doing. Your Jupyter Notebook should be well organized.

Data preparation (75 pts)

- Import employees and orders Excel files into two data frames (**5 pts**)
- Merge the two data frames using the EmployeeID (**10 pts**)
    - After merging the data frames, reset the index
- Ensure that the data type of each column is correct. None of the columns should have a data type **Object**. For example, if the column contains dates and times, then you should make sure the data type of that column is date/time (**10 pts**).
- Check for missing values (**5 pts**). If there are missing values, suggest a way to deal with them.
- Remove duplicate rows if there are any (**5 pts**).
- Calculate the number of days to ship an order (**5 pts**)
    - The formula is as follow: Number of days to ship = Ship date – Order date
- Provide descriptive (e.g., mean, min, max, standard deviation, count, etc.) of all numeric columns (i.e., Sales, Profits, Quantity, and Discount) of the data (**10 pts**)
- Provide the correlation matrix (**10 pts**)
- Assume that an outlier is a value that is above (mean + 3*standard deviation). Find all the outliers
    - Based on Quantity (**5 pts**)
    - Based on Sales (**5 pts**)
    - Based on Profits (**5 pts**)

Data exploration (120 pts)

- Return the average and the total sales per state (**10 pts**)
    - Which state has the highest average sales?
    - Which state has the highest total sales?
- Return the average and the total profits per state (**10 pts**)
    - Which state has the highest average profits?
    - Which state has the highest total profits?
- Return the average and the total sales per sub-category (**10 pts**)
    - Which category has the highest average sales?
    - Which category has the highest total sales?
- Return the average and the total profits per sub-category (**10 pts**)
    - Which category has the highest average profits?
    - Which category has the highest total profits?
- Return the number of Order per Employee (**10 pts**)
    - What is the name of the employee with the highest number of orders?

- Return the average number of days to ship on order per state (**10 pts**)
    - Which state has the lowest number of days to ship an order?
- Return the average and the total sales per order date (in years) (**10 pts**)
    - Which year has the highest average sales?
    - Which year has the highest total sales?
- Return the average and the total profits per order date (in years) (**10 pts**)
    - Which year has the highest average profits?
    - Which year has the highest total profits?
- Answer the following questions:
    - Which state is the most profitable in 2017? (**10 pts**)
    - Which sub-category is the most profitable in the State of California in 2015? (**10 pts**)
    - What are the top 5 customers in terms of sales? (**10 pts**)
    - What are the top 5 products in terms of sales? (**10 pts**)

Data Visualization (60 pts)

- Create a bar plot that shows total sales per sub-category (**10 pts**)
- Create a bar plot that shows total sales per state (**10 pts**)
- Create a histogram of Quantity (**10 pts**)
- Plot how total sales change over time (**10 pts**)
    - Use year-month (e.g., Jan 2017, Feb 2017, March 2017, etc.) of Order Date as time.
- Plot how total sales change over time for the state of California (**10 pts**)
    - Use year-month (e.g., Jan 2017, Feb 2017, March 2017, etc.) of Order Date as time.
- Plot how total sales change over time for the state of Florida (**10 pts**)
    - Use year-month (e.g., Jan 2017, Feb 2017, March 2017, etc.) of Order Date as time.

Readability of the Jupyter Notebook (15 pts)

- Use markdown in Jupyter to create headings and text.
- Write sentences to describe what you are doing. It should be clear what question you are answering.
- Describe the results as much as you can. For example, if you create a bar plot of sales per state, you can describe the plot by indicating the states with the highest and lowest sales.
- Your Jupyter Notebook should be well organized and easy to read.