Applied Data Science with Python

Course-End Project Problem Statement

Sales Analysis

Project statement:

AAL, established in 2000, is a well-known brand in Australia, particularly recognized for its clothing business. It has opened branches in various states, metropolises, and tier-1 and tier-2 cities across the country.

The brand caters to all age groups, from kids to the elderly.

Currently experiencing a surge in business, AAL is actively pursuing expansion opportunities. To facilitate informed investment decisions, the CEO has assigned the responsibility to the head of AAL's sales and marketing (S&M) department. The specific tasks include:

1) Identify the states that are generating the highest revenues.
2) Develop sales programs for states with lower revenues. The head of sales and marketing has requested your assistance with this task.

Analyze the sales data of the company for the fourth quarter in Australia, examining it on a state-by-state basis. Provide insights to assist the company in making data-driven decisions for the upcoming year.

*Enclosed is the CSV (AusApparalSales4thQrt2020.csv) file that covers the said data.

Perform the following steps:

As a data scientist, you must perform the following steps on the enclosed data:

1. Data wrangling
2. Data analysis
3. Data visualization
4. Report generation

1. Data wrangling
   a. Ensure that the data is clean and free from any missing or incorrect entries.
      o Inspect the data manually to identify missing or incorrect information using the functions isna() and notna().

b. Based on your knowledge of data analytics, include your recommendations for treating missing and incorrect data *(dropping the null values or filling them)*.

c. Choose a suitable data wrangling technique—either data standardization or normalization. Execute the preferred normalization method and present the resulting data. *(Normalization is the preferred approach for this problem.)*

d. Share your insights regarding the application of the GroupBy() function for either data chunking or merging, and offer a recommendation based on your analysis.

2. Data analysis

a. Perform descriptive statistical analysis on the data in the Sales and Unit columns. Utilize techniques such as mean, median, mode, and standard deviation for this analysis.

b. Identify the group with the highest sales and the group with the lowest sales based on the data provided.

c. Identify the group with the highest and lowest sales based on the data provided.

d. Generate weekly, monthly, and quarterly reports to document and present the results of the analysis conducted.

(Use suitable libraries such as NumPy, Pandas, and SciPy for performing the analysis.)

3. Data visualization

a. Use suitable data visualization libraries to construct a dashboard for the head of sales and marketing. The dashboard should encompass key parameters:

   o State-wise sales analysis for different demographic groups (kids, women, men, and seniors).

   o Group-wise sales analysis (Kids, Women, Men, and Seniors) across various states.

   o Time-of-the-day analysis: Identify peak and off-peak sales periods to facilitate strategic planning for S&M teams. This information aids in designing programs like *hyper-personalization* and *Next Best Offers* to enhance sales.

b. Ensure the visualization is clear and accessible for effective decision-making by the head of sales and marketing (S&M).

The dashboard must contain daily, weekly, monthly, and quarterly charts.

(Any visualization library can be used for this purpose. However, since statistical analysis is being done, Seaborn is preferred.)

c. Include your recommendation and indicate why you are choosing the recommended visualization package.

4. Report generation
   a) Use JupyterLab Notebook for generating reports, which includes tasks such as data wrangling, analysis, and visualization. Please note that JupyterLab enables you to integrate code seamlessly with graphs and plots.
   b) Use Markdown in suitable places while presenting your report.
   c) Use suitable graphs, plots, and analysis reports in the report, along with recommendations. Note that various aspects of analysis require different graphs and plots.
      - Use a box plot for descriptive statistics.
      - Use the Seaborn distribution plot for any other statistical plotting.