

Applied Data Science with Python

Course-End Project Problem Statement

Course-End Project: Sales Analysis

Project Statement:

AAL is a household name in Australia, known for its clothing business since 2000. It caters to all groups— kids, women, men, and seniors. It has opened branches in many states in Australia, in metropolises and tier-1 and tier-2 cities.

The business is booming, and the company is in an expansion mode. It wants to get an in-depth understanding of its sales so that it can make investment decisions. The CEO has tasked the Head of Sales and Marketing (S&M) of AAL to:

- Determine the states that are generating the highest revenues and
- Formulate sales programs for states with lower revenues. The Head of S&M has approached you for the same.

Analyze the company's sales data for the fourth quarter across Australia, and state by state and help the company make data-driven decisions for the coming year.

Perform the following steps:

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

- Data Wrangling
- Data Analysis
- Data Visualization and
- Report Generation

- Data Wrangling
- Ensure that the data is clean and that there is no missing or incorrect data.
 - Inspect the data manually for missing/incorrect data using the functions `isna()`, and `notna()`.
- Based on your knowledge of Data Analytics, include your recommendations for treating missing data and incorrect data. (*dropping the null values or filling them*).
- Select an appropriate Data Wrangling approach — data standardization or data normalization. Perform the standardization or normalization and present the data. (*Normalization is the preferred approach for this problem.*)
- Share your recommendation on the usage of the `groupby()` function for data chunking or merging.

- Data Analysis
- Perform descriptive statistical analysis on the data (Sales and Unit columns) (*Techniques such as mean, median, mode and standard deviation can be used.*).
- Determine which group is generating the highest sales, and which group is generating the lowest sales.

- Determine which state is generating the highest sales, and which state is generating the lowest sales.
- Generate weekly, monthly and quarterly reports for the analysis made.

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

- Data Visualization
- Use appropriate data visualization libraries to build a dashboard for the Head of S&M that includes for the key parameters like
 - State-wise sales analysis for different groups (kids, women, men, and seniors)
 - Group-wise sales analysis (kids, women, men, and seniors) across different states.
 - Time-of-the-day analysis: during which time of the day are sales the highest, and during which time are sales the lowest? [This helps S&M teams design programs for increasing sales such as hyper-personalization and Next Best Offers].

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.)

- Include your recommendation, and indicate why you are choosing the recommended visualization package.
- Report Generation
- Use JupyterLab Notebook for report generation (wrangling, analysis and visualization) *Please note that JupyterLab allows you to mix code with graphs and plots etc.*
- Use Markdown in suitable places, while presenting your report.
- The report should contain suitable graphs, plots and analysis reports, and recommendations. *Please note that different aspects of analysis demand different graphs/plots.*
 - Use box plot for descriptive statistics
 - Use Seaborn distribution plot for any other statistical plotting