SHIPMENT SURE

DOCUMENTATION

Week 1

24th September 2025 (Wednesday)

- Browsed the internet to identify relevant datasets for shipment and delivery prediction.
- Selected and downloaded a dataset from Kaggle containing 10,000+ rows with multiple shipment-related features.
- Briefly reviewed the dataset structure (rows, columns, feature types).

25th September 2025 (Thursday)

- Began data cleaning and preprocessing using Pandas in Google Colab.
- Identified missing values across the dataset.
- Replaced missing values systematically:
- Numerical columns → replaced with mean values.
- Categorical columns → replaced with mode values.
- Ensured the dataset was consistent and ready for analysis.

26th September 2025 (Friday)

- Generated descriptive statistics of the dataset (using .describe() and other summary functions).
- Explored measures like mean, median, min, max, standard deviation for numerical features.
- Reviewed distribution and summary of categorical features.
- Gained initial understanding of the dataset's structure and feature behavior.

SHIPMENT SURE

DOCUMENTATION

Week 2

29th September 2025 (Monday)

- Conducted Univariate Analysis:
- Studied each feature individually.
- Checked distributions, unique values, and frequency counts.
- Conducted Bivariate Analysis:
- Analyzed relationships between independent features and the target variable (reached_on_time).
- Checked correlations between features to detect patterns and dependencies.

30th September 2025 (Tuesday)

- Performed Data Visualization for deeper insights.
- Used libraries like Matplotlib and Seaborn to create plots:
- Histograms, boxplots, and bar charts for univariate analysis.
- Scatter plots, heatmaps, and pair plots for bivariate analysis.
- Improved overall understanding of data patterns, outliers, and feature importance.

1st October 2025 (Wednesday)

- Addressed class imbalance problem in the target variable (reached_on_time).
- Observed that the dataset had unequal distribution between "On Time" vs. "Not On Time" classes.
- Applied balancing techniques (e.g., oversampling) to create a more balanced dataset.
- This step ensures that future ML models won’t be biased towards the majority class.

SHIPMENT SURE

DOCUMENTATION

Progress Summary So Far:

- Dataset collected and cleaned.
- Descriptive statistics and exploratory analysis completed.
- Data visualization provided insights.
- Target variable imbalance resolved.