Shipment Analysis

Data Analytics Project Part- 1

---

1. Exploratory Data Analysis (EDA)

A. Initial Data Cleaning:

- Handle missing values, data types, and duplicates
- Convert 'Date of Shipment' to datetime format
- Derive new columns like 'Shipping Duration'

B. Univariate Analysis:

- Use histograms and boxplots to explore Profit, Cost, Delay, etc.
- Value counts for categorical variables like Company, Mode, Priority

C. Bivariate & Multivariate Analysis:

- Analyze relationships using scatter plots and boxplots
- Correlation matrix for numeric columns

2. Feature Engineering

A. Create New Features:

- Cost per km = Shipment Cost / Distance
- Cost per kg = Shipment Cost / Weight
- Delay Category (Low, Medium, High)
- Is Delayed = Delay > 0
- Is Fragile & High Value = Complex interaction feature

B. Outlier Treatment:

- Use IQR or Z-Score to detect and flag/remove outliers
- Apply capping or transformation where needed

C. Encoding & Scaling:

- One-hot encode categorical fields: Mode, Priority, Company, etc.
- Standardize/normalize numerical fields for ML readiness

3. Business Analysis

A. Business Questions to Explore:

- Which company is most cost-efficient per km?
- Which shipment modes are most delayed?
- Which regions have the highest logistic costs?
- Which product categories result in the most returns or delays?