

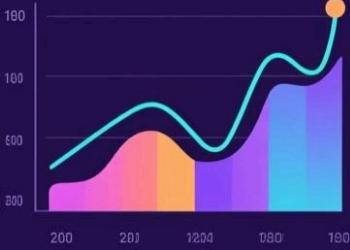
Predicting On-Time Deliveries

An analytical approach to optimize logistics and customer satisfaction.



Made with GAMMA

Delivery times



Delivery times



Delivery times



Logistics



Exploratory Data Analysis



Weight Impact

Heavier items are more prone to late deliveries.



Discount Strategy

High-discount products often arrive on time, suggesting a campaign opportunity.



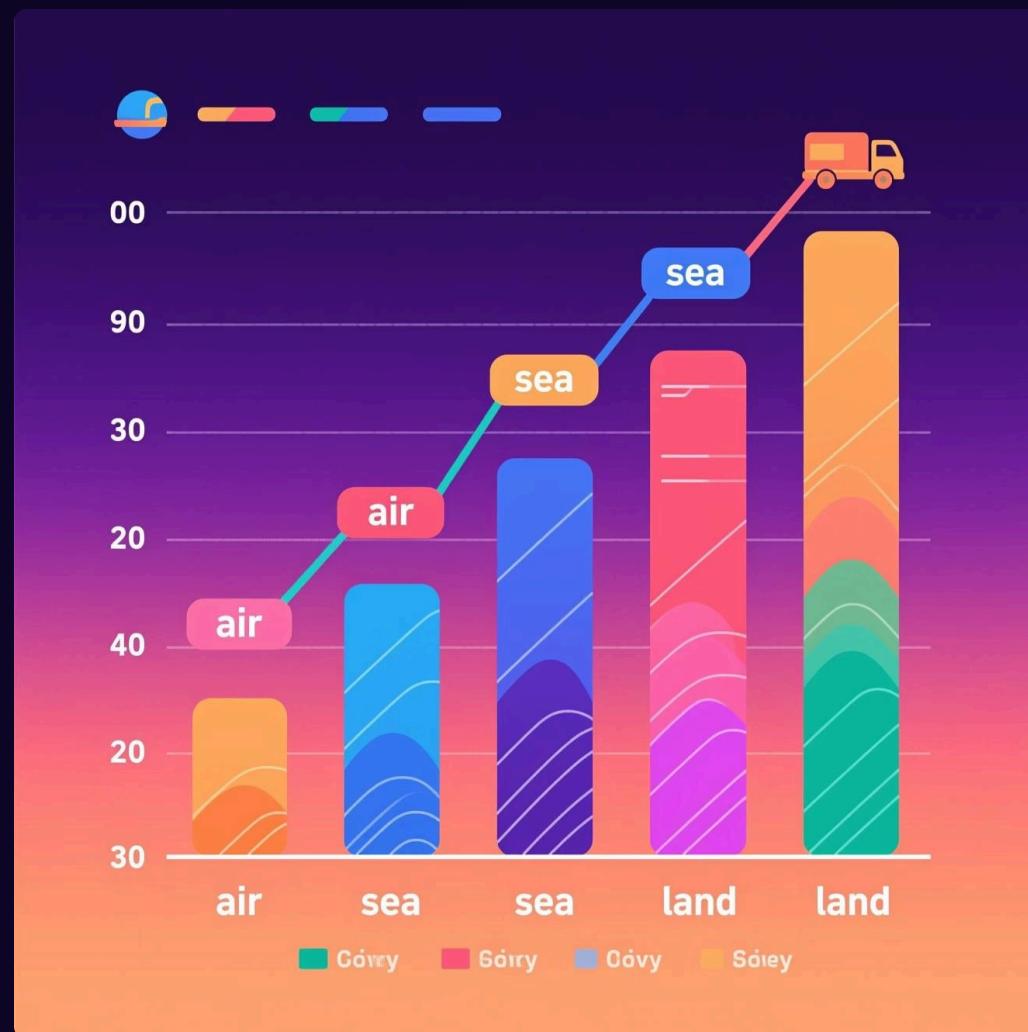
Cost vs. Delivery

Analyzing product cost against delivery performance.

Shipment & Product Insights

Mode of Shipment

Distribution of on-time deliveries across different shipping methods.



Product Importance

How product priority influences delivery punctuality.





Data Preprocessing: Encoding Categorical Features

1

Identify Unique Values

Extracting unique categories for
'Warehouse_block',
'Mode_of_Shipment',
'Product_importance', and 'Gender'.

2

Label Encoding

Transforming categorical text data
into numerical representations for
model compatibility.

3

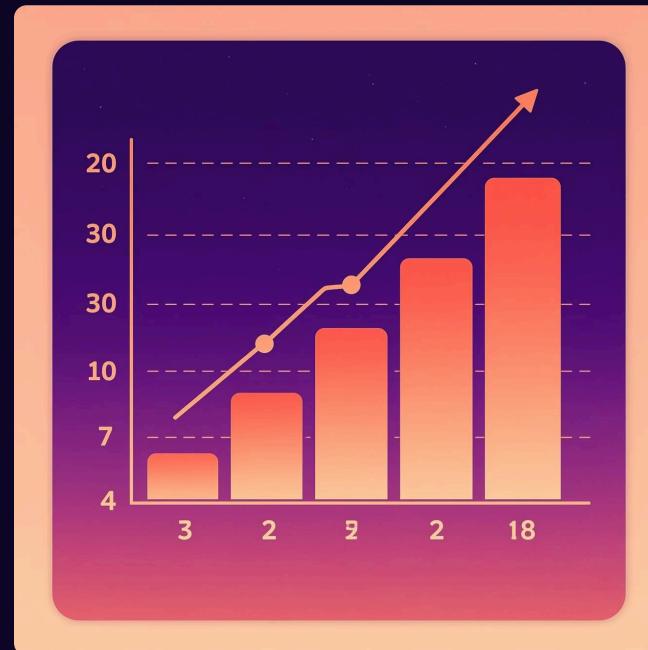
Convert to Numeric

Ensuring all encoded columns are of
a numeric data type for analysis.

Outlier Detection

Visualizing data distribution to identify extreme values that could skew model accuracy.

Cost of Product



Discount Offered



Weight in Grams





Outlier Removal with Z-Score

Applying the Z-score method to clean the dataset by removing statistical outliers, enhancing data quality for modeling.

"Clean data is the foundation of reliable machine learning models."



Balancing Class Values

01

Class Imbalance

Initial count reveals an imbalance between on-time (1) and late (0) deliveries.

02

Upsampling Minority

The minority class (late deliveries) is upsampled to match the majority class size (6282 samples).

03

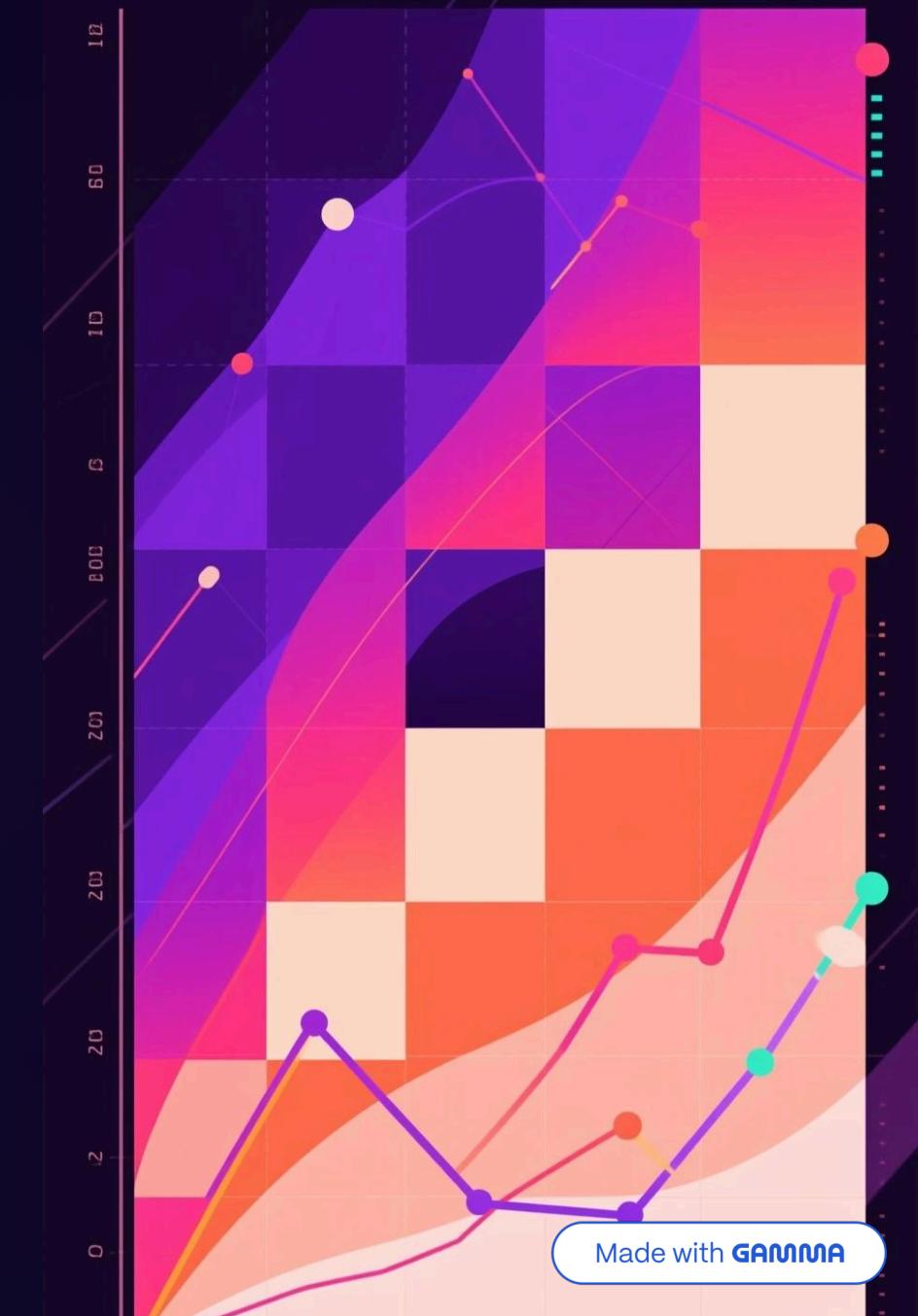
Balanced Dataset

Combining upsampled minority with majority class creates a balanced dataset for training.

Feature Correlation Analysis

A heatmap visualizes the correlation between different features in the cleaned and balanced dataset, revealing interdependencies.

- Understanding feature relationships is crucial for model interpretability and selection.





Machine Learning Model Building

Data Split

Dataset divided into 80% training and 20% testing sets.

Decision Tree

Accuracy Score: 96.94%

F-1 Score: 0.97, Precision: 0.96,
Recall: 0.98

Random Forest

Accuracy Score: 98.41%

F-1 Score: 0.98, Precision: 0.98,
Recall: 0.99

Model Performance: Logistic Regression

Logistic Regression

Accuracy Score: 61.54%

F-1 Score: 0.62

Precision Score: 0.61

Recall Score: 0.63



Random Forest Classifier shows superior performance for predicting on-time deliveries.