Screening Task: Python & Data Science Evaluation

Predicting Delivery Delays

RITESH BANKAR

Problem Statement & Objective

- **Problem:**

1. **Timely delivery is crucial to customer satisfaction and business effectiveness.**
2. **Delays in outbound deliveries can result in missed sales, higher costs, and lower consumer trust. Currently, delivery performance is influenced by factors such as invoice processing time, outbound processing time, and logistics restrictions.**

- **Objective:**

1. **Analyze past delivery data to discover trends.**
2. **Predict potential delays to improve supply chain efficiency.**

Dataset & EDA (Exploratory Data Analysis)

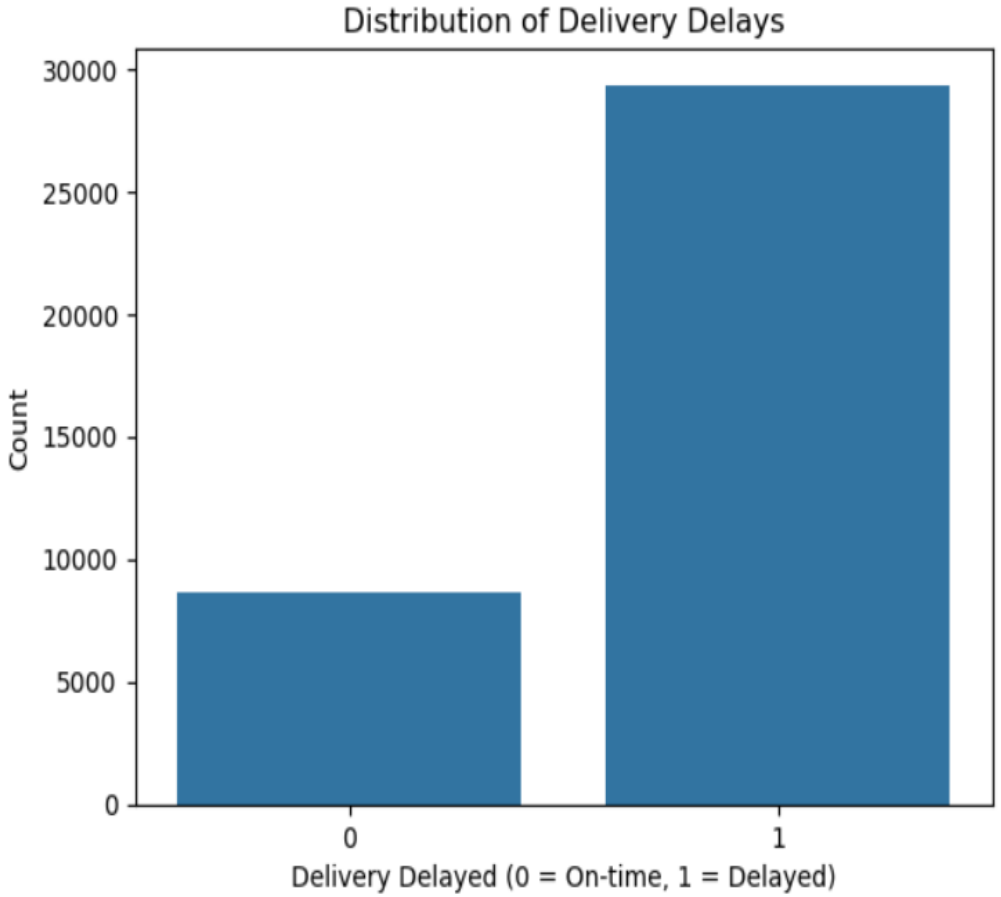
- **Dataset Overview:**

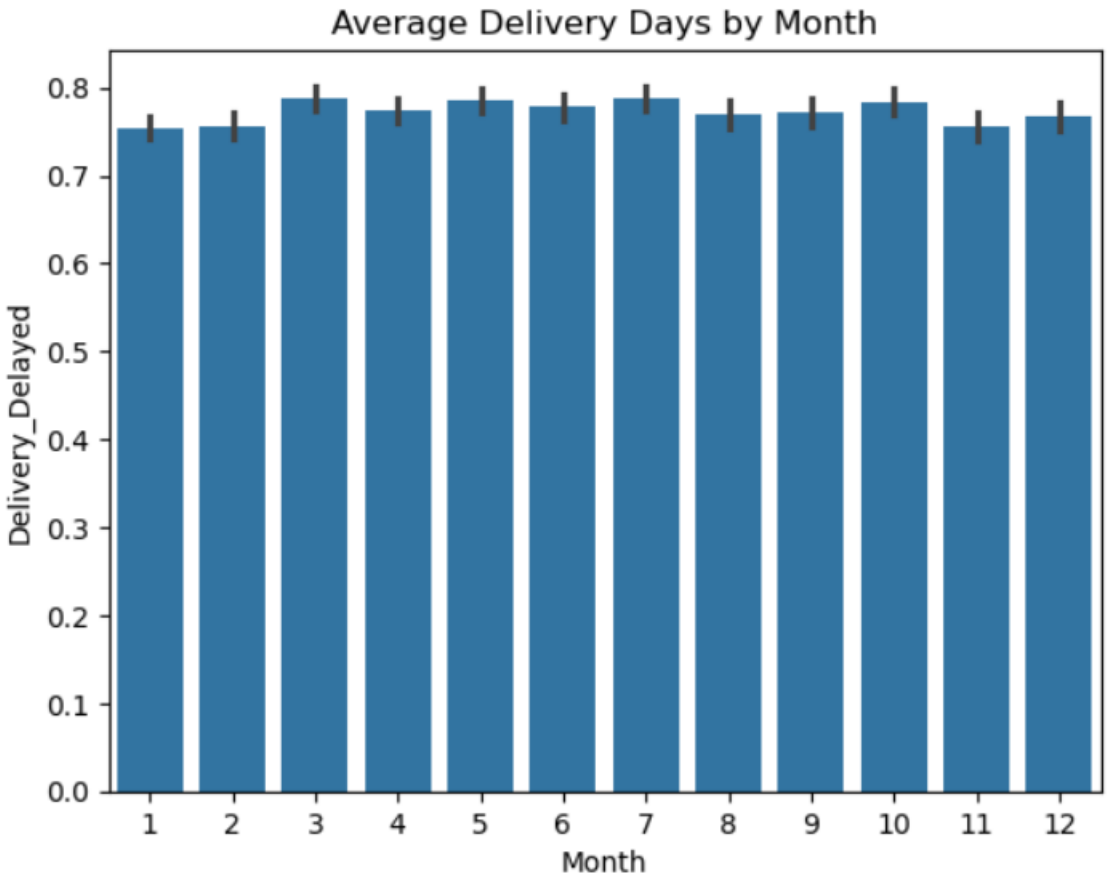
1. Includes order dates, delivery times, and product details.
2. Identified missing values and handled them appropriately.

- **EDA Key Insights:**

1. Correlation between order dates and delays.
2. Distribution of delivery delays using countplot.
3. The dataset contained outliers, particularly in Billing Amount and processing times.

Important Data Visuals





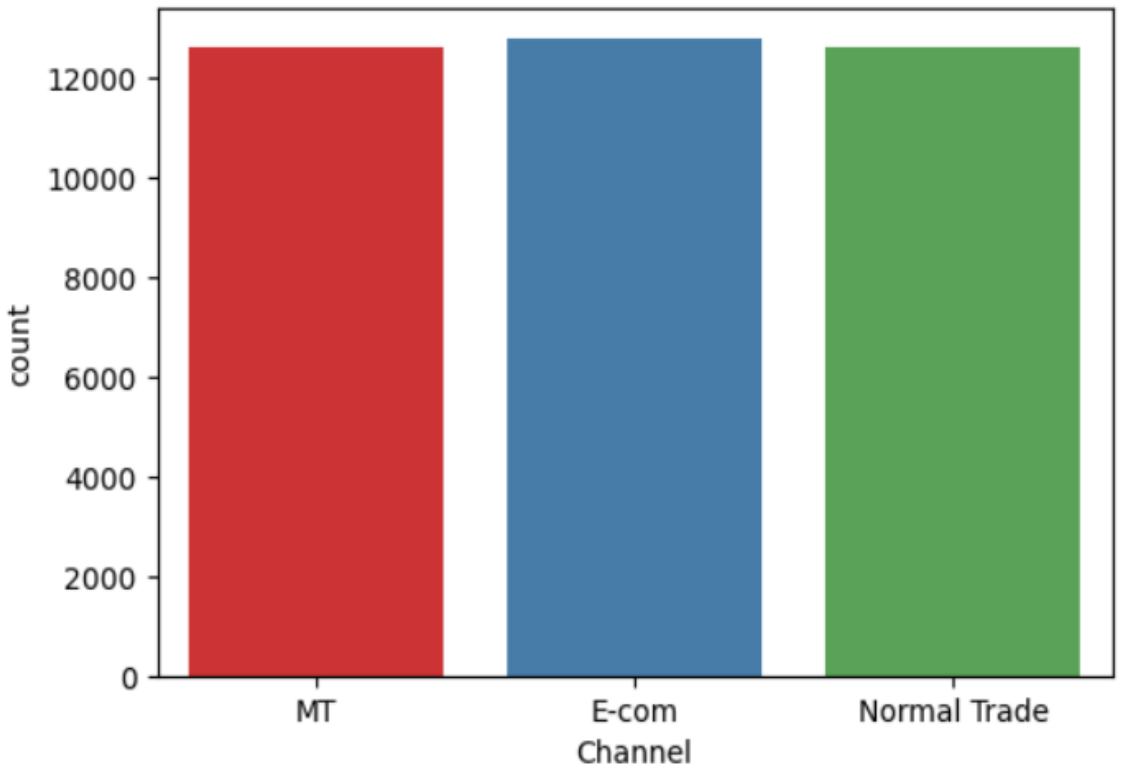
Correlation Analysis & Model Selection

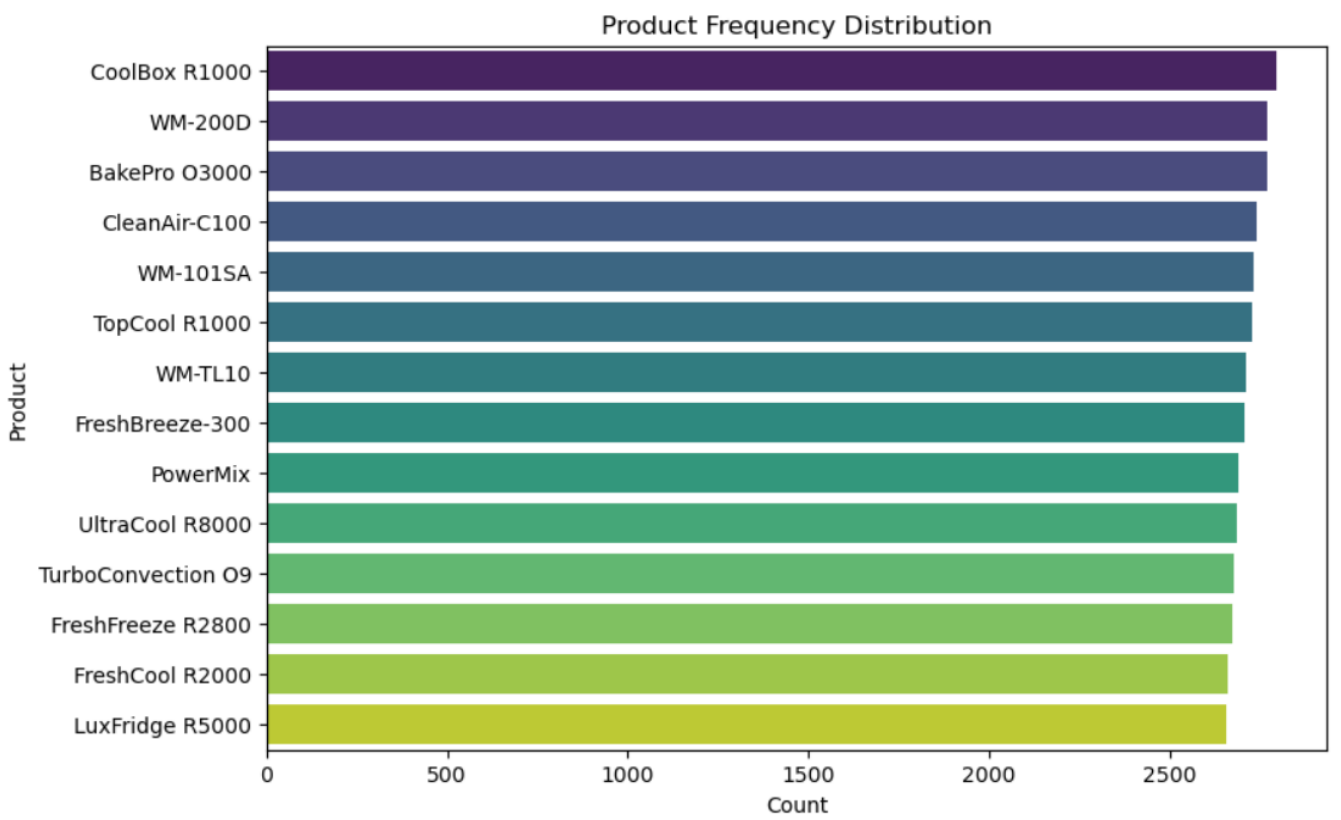
- **Correlation Analysis:**

1. Some variables showed strong correlations with delivery delays.
2. Understanding these relationships helps in feature selection for the model.
3. Predictive Modeling Approach:

- **Model Selection:**
- Various machine learning techniques were likely used (e.g., logistic regression, decision trees and random forests).
- Model performance was evaluated using metrics like accuracy, precision, recall, or F1-score.

Important Data Visuals





Results & Insights

- **Model Performance & Predictions:**

Model	Training Accuracy	Testing Accuracy	Observations
Decision Tree	98.21%	94.56%	Slight Overfitting
Random Forest	98%	95%	Slight Overfitting

Conclusion:

- **Conclusion:**

1. A data-driven approach empowers targeted interventions in the delivery process, reducing delays and optimizing the supply chain.

- **Improved Delivery Delay Predictions:**

1. The predictive model accurately identifies delayed shipments, enabling proactive interventions.
2. Random Forest performed best, with high recall to ensure most delays are detected.

Future of the Model:

- **Model Enhancement:**
- Incorporate Real-Time Data: Integrate live tracking data, weather conditions, and traffic updates for better delay predictions.
- **Advanced Machine Learning Techniques:**
- Deep Learning: Use Neural Networks (LSTMs or GRUs) for time-series forecasting of delivery delays.
- **Deployment & Business Integration**
- Dashboard & Reporting: Build interactive BI dashboards for logistics managers to track and act on high-risk deliveries.