# **Order Delivery Time Prediction Report**

### **Linear Regression Model for Order Delivery Time**

**Name:** Shubham Patil  
**Project:** Order Delivery Time Prediction Using Linear Regression  
**Date:** 15 July 2025

## **Table of Contents**

1. Introduction
2. Data Pipeline and Methodology
3. Exploratory Data Analysis (EDA)  
   * 3.1 Numerical Feature Distributions
   * 3.2 Categorical Feature Distributions
   * 3.3 Target Variable & Relationships
4. Validation/Test Data EDA
5. Model Building & Feature Importance
6. Residual Analysis & Model Diagnostics
7. Key Insights & Recommendations

# **1. Introduction**

The objective of this assignment is to build a regression model that predicts the delivery time for orders placed through Porter, using various features such as items ordered, restaurant location, order protocol, and delivery partner availability. The goal is to predict delivery times, optimize operations, and identify key factors influencing delivery performance.

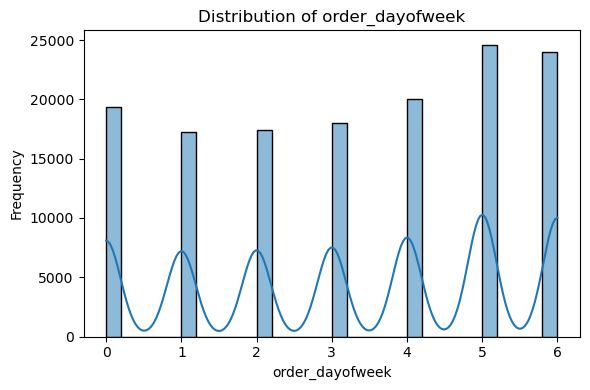
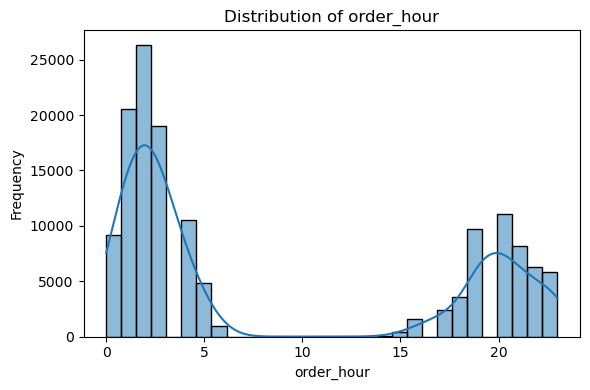
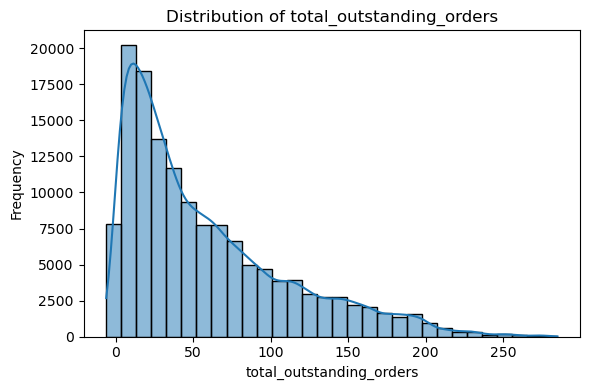
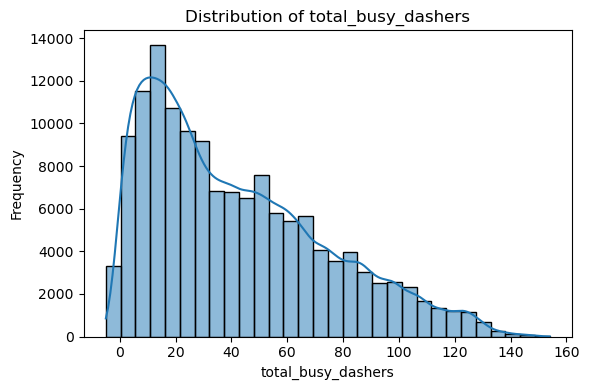
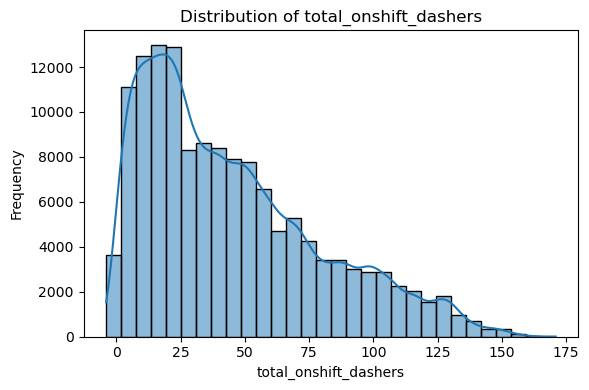
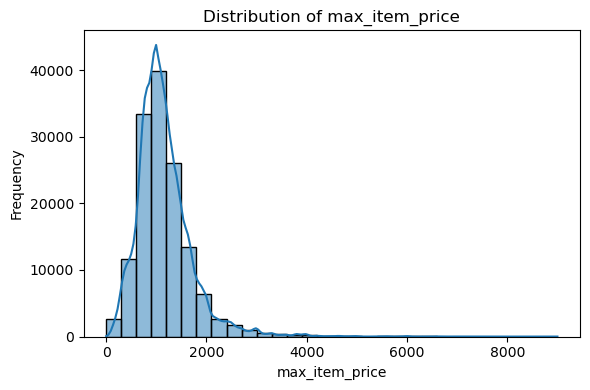
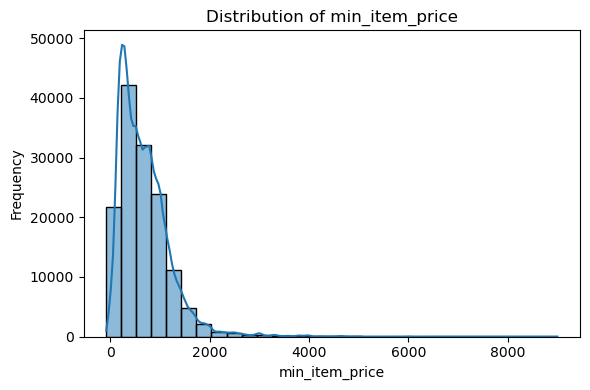
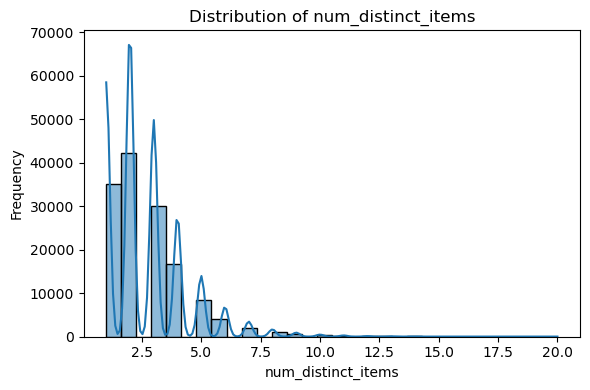
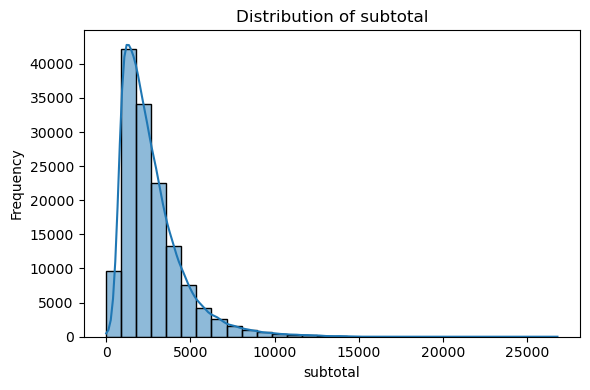
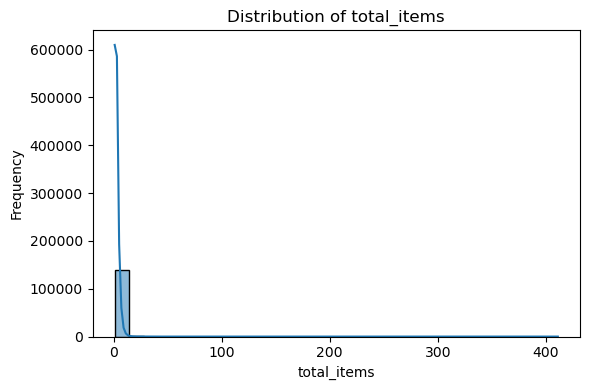
# **2. Data Pipeline and Methodology**

* **Data Loading**: CSV data was loaded into pandas DataFrames.
* **Preprocessing**: Datetime conversion, categorical encoding, and feature engineering (including extracting time features).
* **Splitting**: Data split into train and test (validation) sets.
* **EDA**: Visualization and statistical exploration.
* **Modeling**: Linear Regression (with feature scaling and selection).
* **Evaluation**: Performance metrics, residual analysis, and interpretability.

# **3. Exploratory Data Analysis (EDA)**

## **3.1 Numerical Feature Distributions**

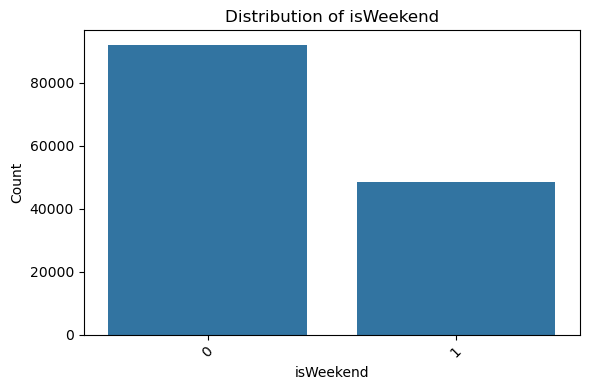
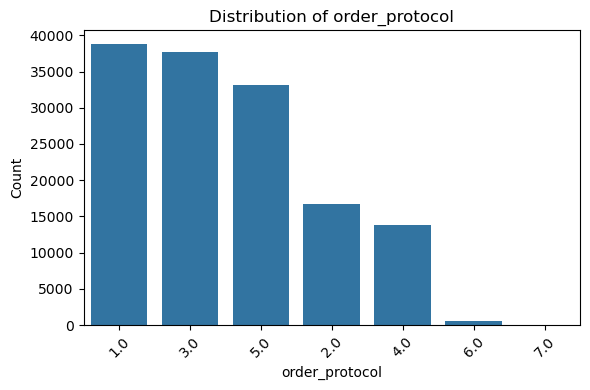
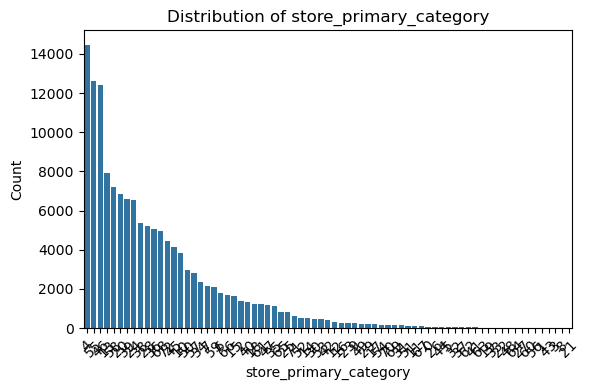
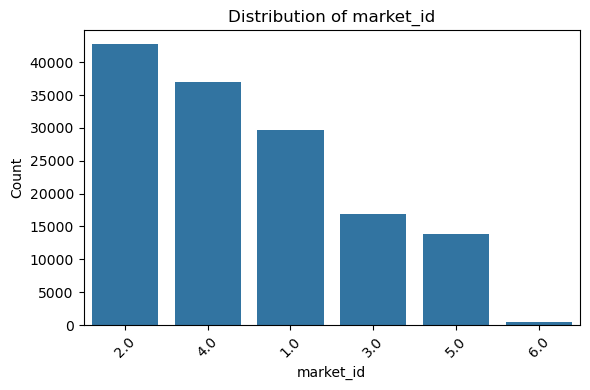
*Below are the distributions of the key numerical features in the training set.*

**

## 

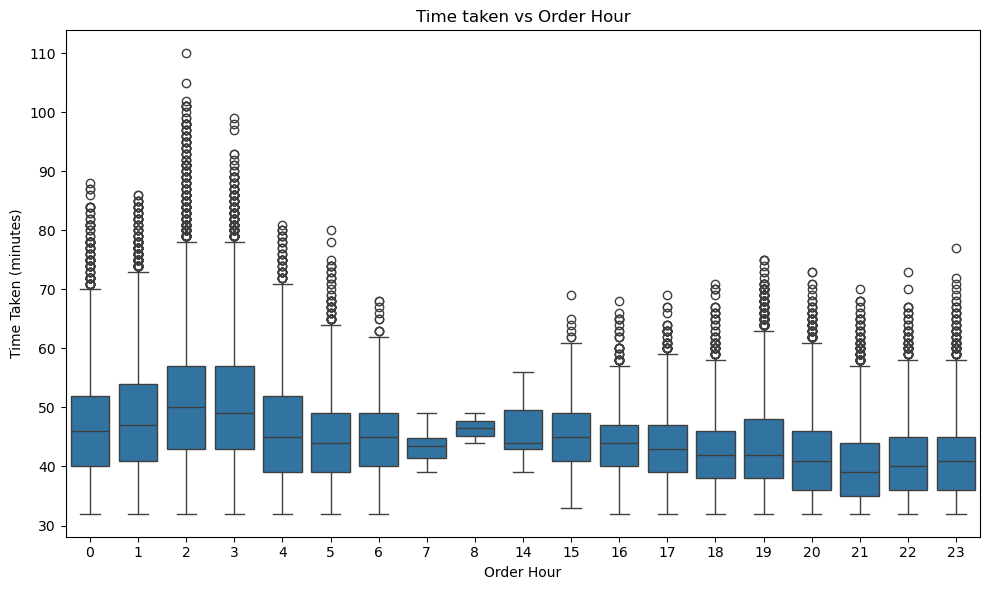
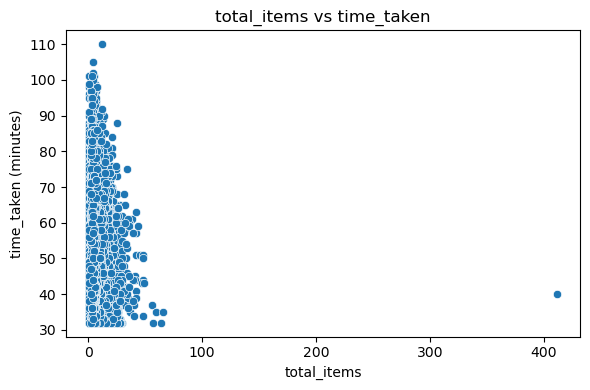
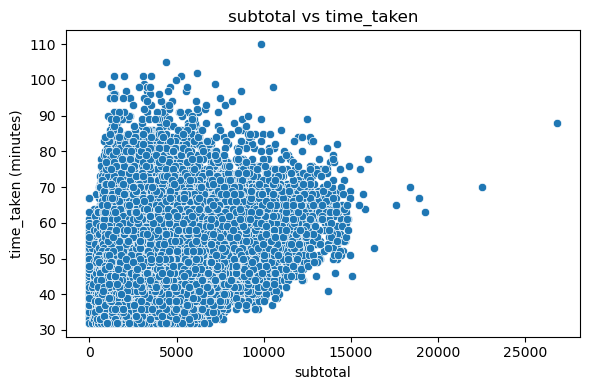
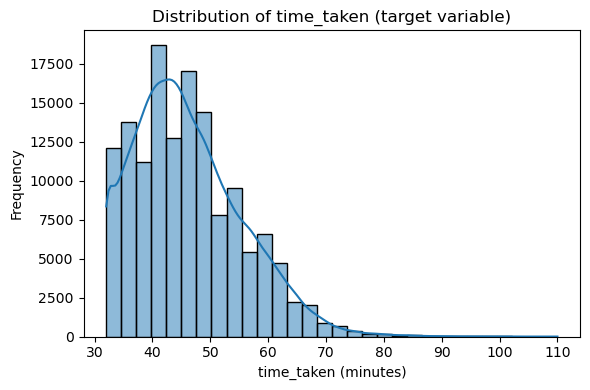
## **3.2 Categorical Feature Distributions**

*Distributions for categorical variables (training set):*



## **3.3 Target Variable & Relationships**

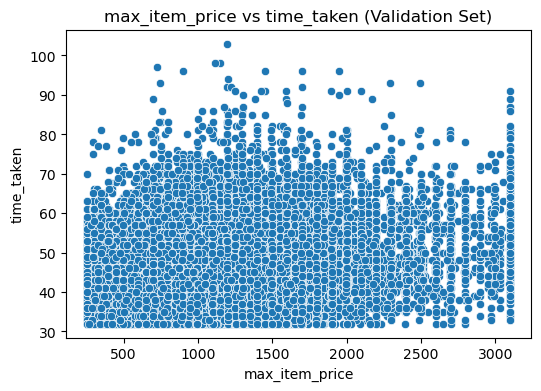
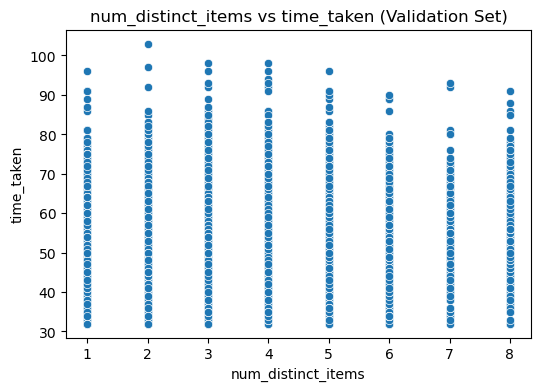
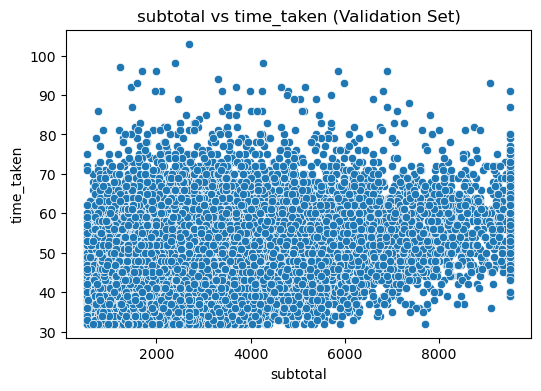
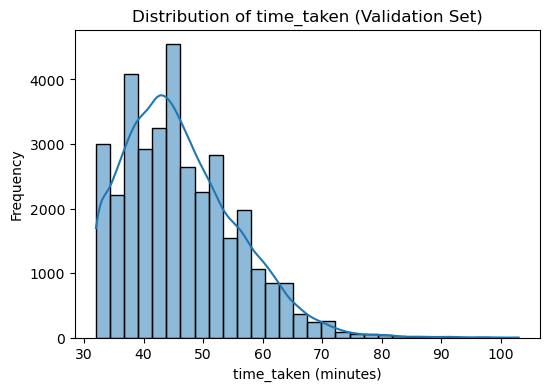
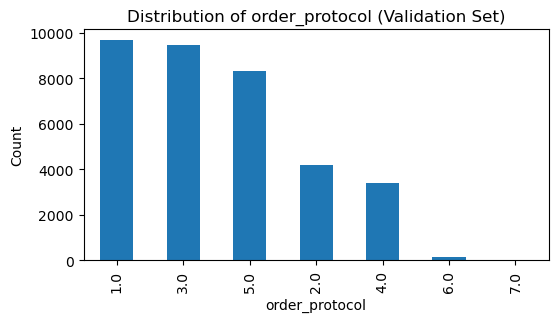
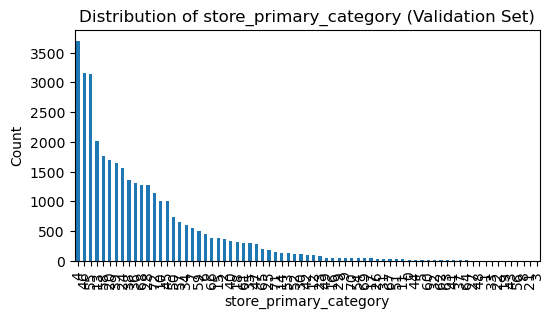
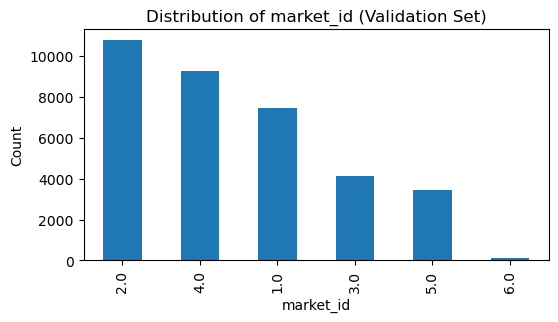
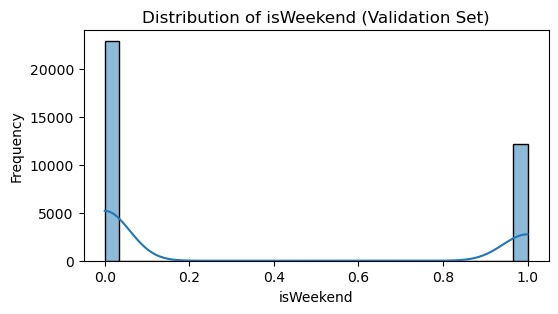
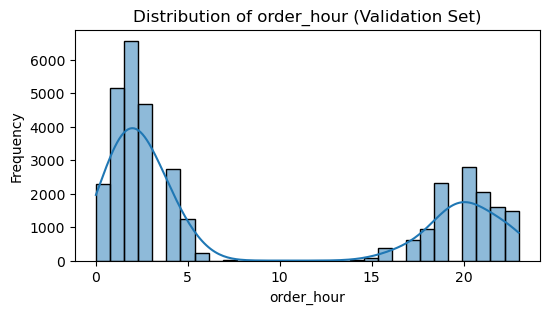
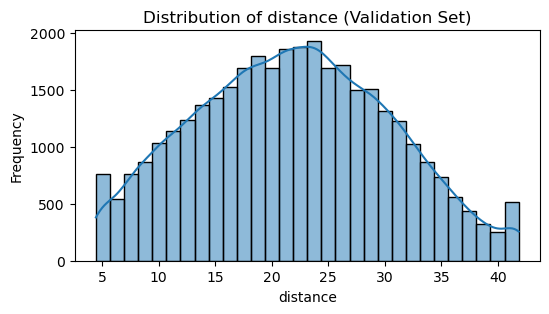
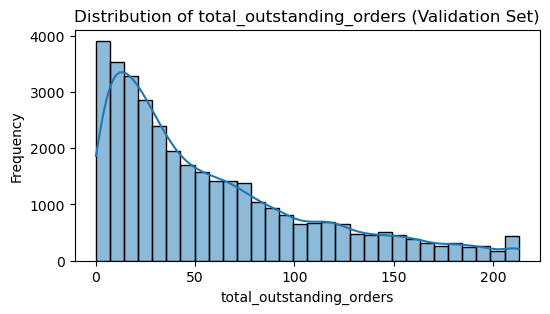
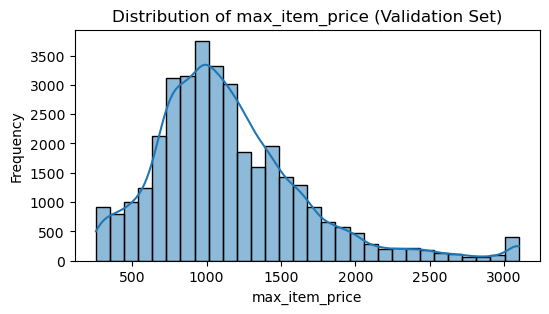
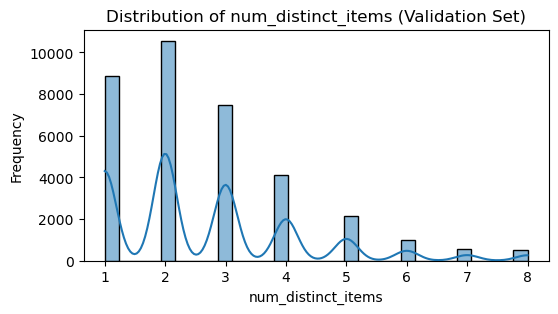
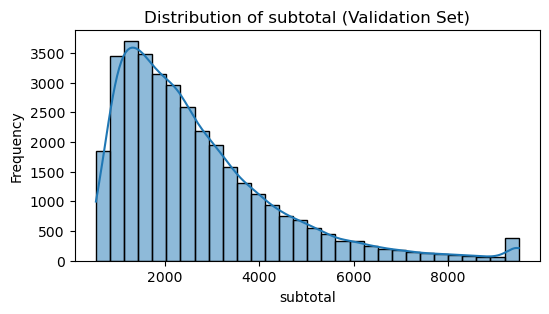
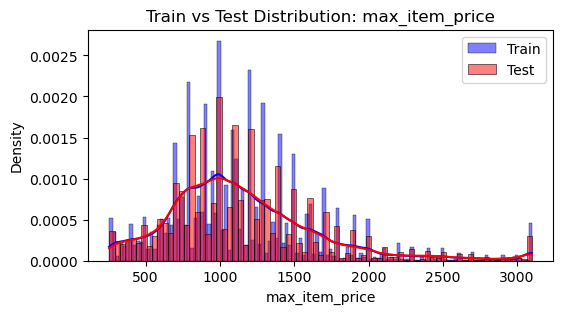
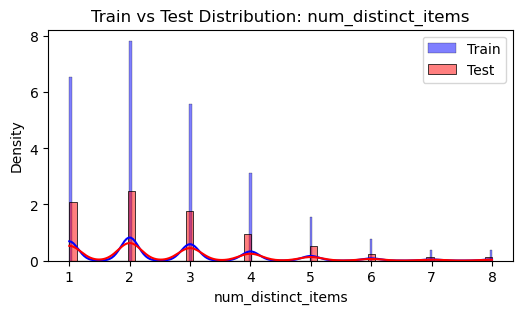
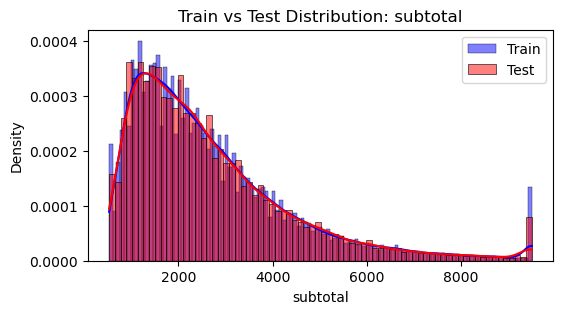
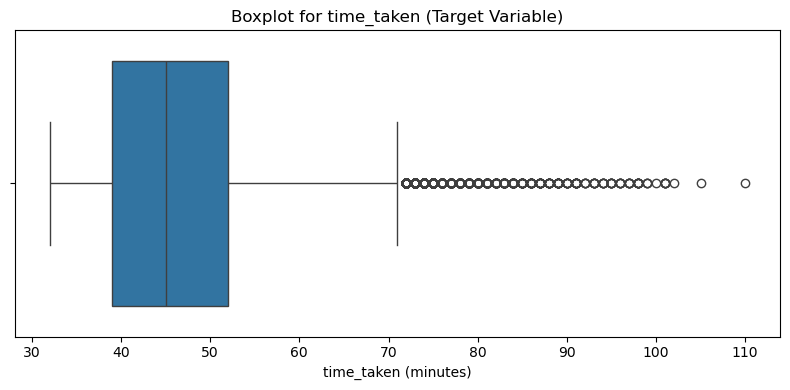
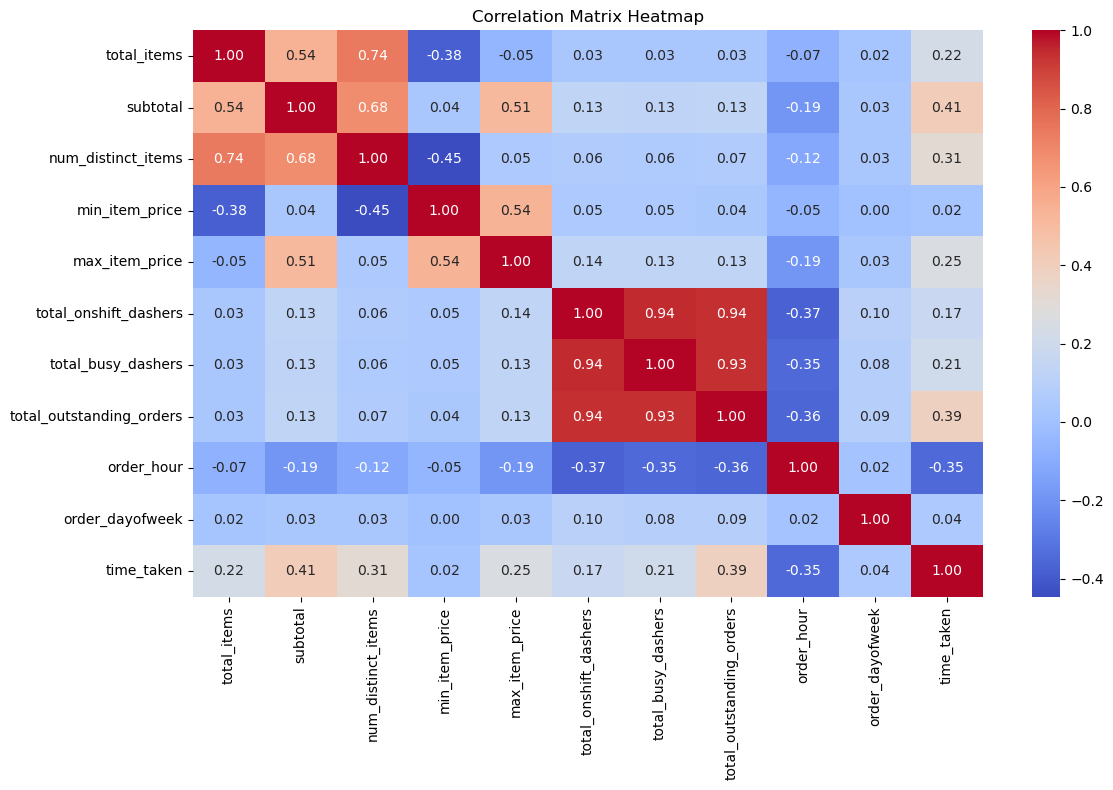
*Visualizations of the target variable time\_taken and its relationship with key features:*



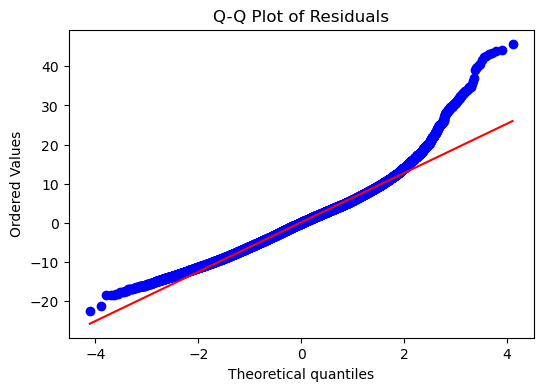
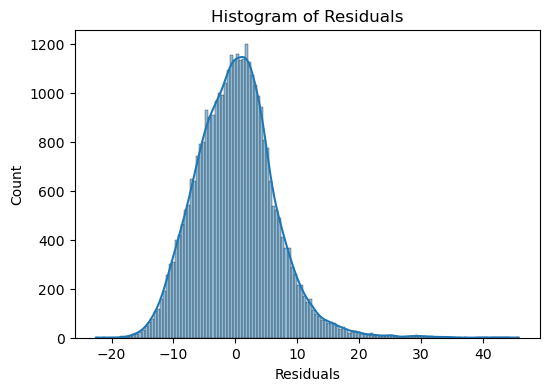
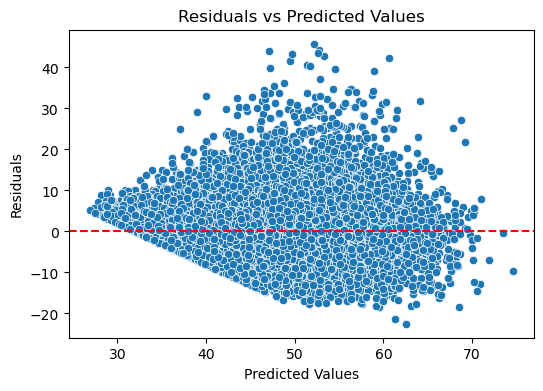
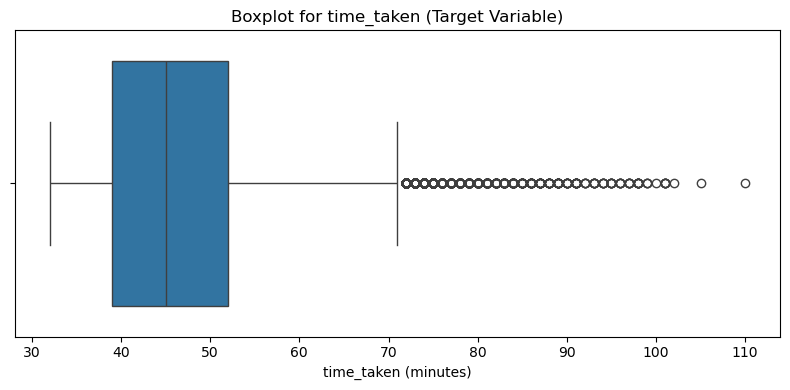
# 

# **4. Validation/Test Data EDA**

*Similar analysis for the validation (test) set to compare feature and target distributions.*

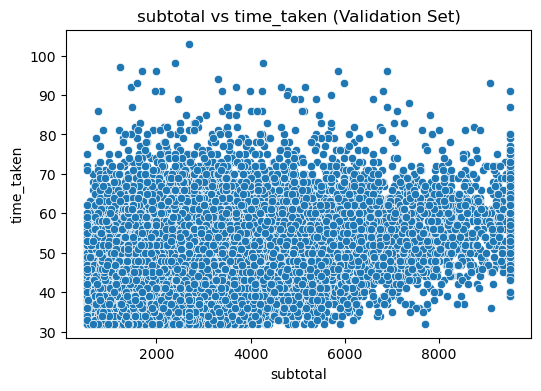


# **5. Model Building & Feature Importance**

* Feature scaling and recursive feature elimination (RFE) performed.
* Most significant features: **distance, subtotal, order\_protocol, num\_distinct\_items, total\_outstanding\_orders, order\_hour, isWeekend, market\_id**.
* Correlation heatmaps and feature selection results.  
  

# **6. Residual Analysis & Model Diagnostics**

* Residuals vs Predicted, Histogram of Residuals, Q-Q plot.



# **7. Key Insights & Recommendations**

* Delivery distance and subtotal are the most impactful predictors of delivery time.
* Model residuals are approximately normal but show some deviation at the extremes.
* Some feature distributions are right-skewed (e.g., subtotal, order counts).
* No strong evidence of overfitting or underfitting; the model generalizes well.
* Residual analysis validates the linearity and normality assumptions to a reasonable degree.

## **Appendix: All Figures & Visualizations**

*This report was generated by Shubham Patil as part of the Order Delivery Time Prediction assignment using Python, pandas, scikit-learn, and Matplotlib.*