

Summary Statistics

1. General Dataset Overview:

- Total records: 176,248
- Features include both numerical and categorical data related to delivery time and store properties.

2. Key Features:

- **time_taken_mins:**
 - Average: 47.76 minutes
 - Standard Deviation: 27.65 minutes
 - Minimum: ~1.68 minutes
 - Maximum: 6,231.31 minutes (possibly an outlier).
- **subtotal:**
 - Average: ₹2,696.50
 - Range: ₹0 to ₹26,800.
- **total_items:**
 - Average: 3.2 items per order.
 - Range: 1 to 411 items (some extreme values noted).

3. Anomalies in Data:

- Negative values in features such as min_item_price and total_onshift_partners might need further investigation or correction.

Correlation Analysis

1. Strongest Relationships:

- **subtotal and total_items:** Moderate positive correlation (0.555) indicates that larger orders generally have higher subtotals.
- **num_distinct_items and total_items:** High correlation (0.758) suggests a strong relationship between the variety and quantity of items in an order.

2. Weak Relationships:

- Most features, such as `store_primary_category_encoded`, have weak correlations with `time_taken_mins`.
- **time_taken_mins and subtotal:** Low positive correlation (0.144), implying that the subtotal has a minimal effect on delivery time.

3. Negative Correlations:

- **hours and total_onshift_partners:** Moderate negative correlation (-0.375), indicating fewer partners on-shift during off-peak hours.
-

Recommendations for Next Steps

1. Data Cleaning:

- Investigate and handle negative values in features like `min_item_price` and `total_onshift_partners`.
- Address potential outliers in `time_taken_mins` and `total_items`.

2. Feature Importance:

- Perform detailed feature engineering, focusing on correlated features (`subtotal`, `num_distinct_items`) for model training.

3. Visualization Insights:

- Delivery time peaks at certain hours (e.g., mornings or evenings) and days (e.g., weekends vs. weekdays).
- Use heatmaps and scatterplots to further validate patterns in the data.

4. Modeling Strategy:

- Focus on enhancing the feature set by incorporating external factors like traffic or weather.
- Perform hyperparameter tuning for Random Forest and explore feature selection techniques.