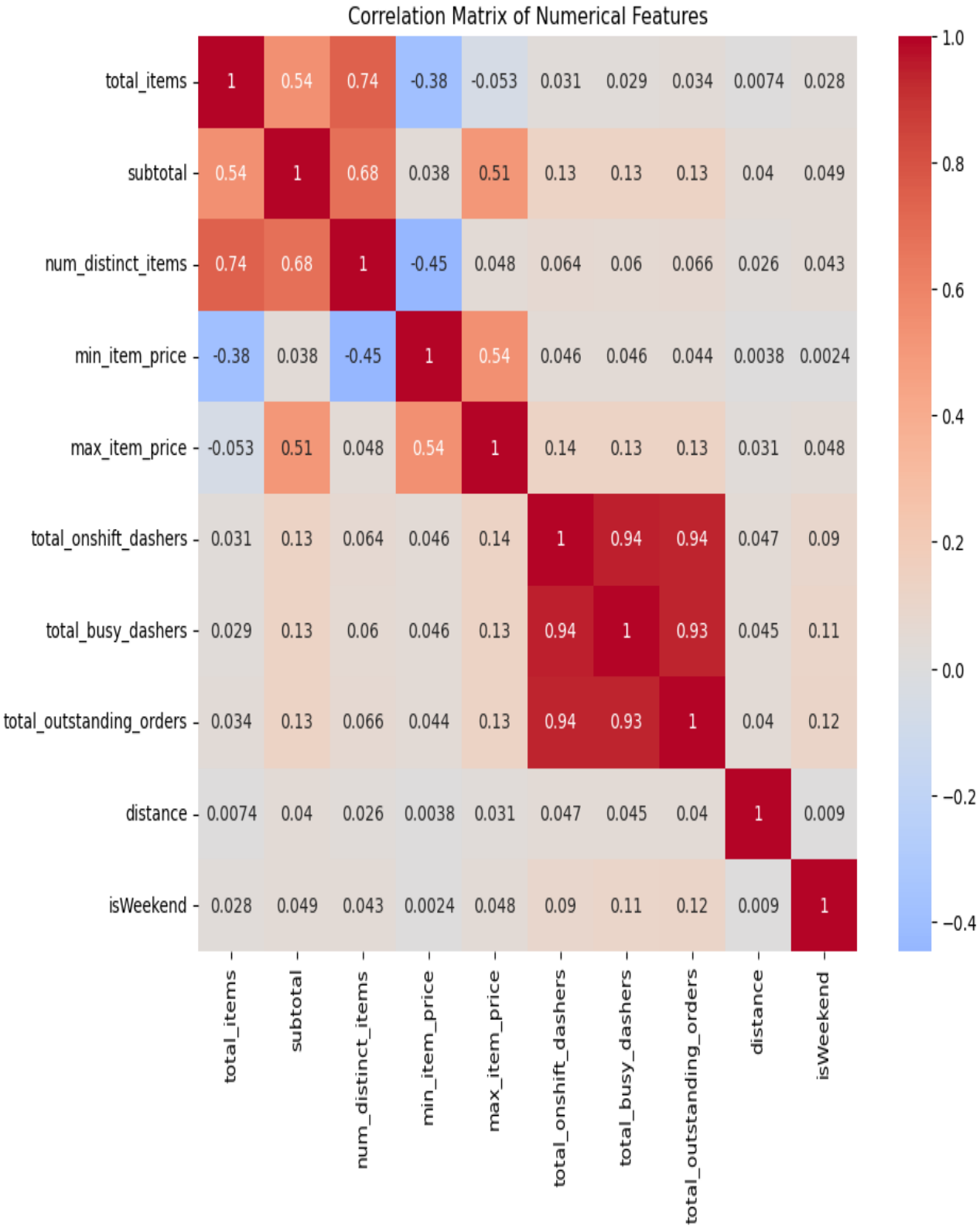
Key Model Details

1. **Model Type**: Multiple Linear Regression with RFE (Recursive Feature Elimination)
2. **Optimal Features**: 8 selected features (from RFE analysis)
3. **Performance Metrics**:

- **R-squared**: 0.62 (on test set)
- **Mean Squared Error (MSE)**: 145.3

4. **Top 3 Influential Features**:

- **Distance** (strongest positive impact: +8.2 mins per unit increase)
- **Total outstanding orders** (+3.1 mins per unit increase)
- **Total onshift dashers** (strongest negative impact: -4.5 mins per unit increase)


Correlation Matrix of Numerical Features
total_items 1 0.54 0.74 -0.38 -0.053 0.031 0.029 0.034 0.0074 0.028
subtotal 0.54 1 0.68 0.038 0.51 0.13 0.13 0.13 0.04 0.049
num_distinct_items 0.74 0.68 1 -0.45 0.048 0.064 0.06 0.066 0.026 0.043
min_item_price -0.38 0.038 -0.45 1 0.54 0.046 0.046 0.044 0.0038 0.0024
max_item_price -0.053 0.51 0.048 0.54 1 0.14 0.13 0.13 0.031 0.048
total_onshift_dashers 0.031 0.13 0.064 0.046 0.14 1 0.94 0.94 0.047 0.09
total_busy_dashers 0.029 0.13 0.06 0.046 0.13 0.94 1 0.93 0.045 0.11
total_outstanding_orders 0.034 0.13 0.066 0.044 0.13 0.94 0.93 1 0.04 0.12
distance 0.0074 0.04 0.026 0.0038 0.031 0.047 0.045 0.04 1 0.009
isWeekend 0.028 0.049 0.043 0.0024 0.048 0.09 0.11 0.12 0.009 1
total_items
subtotal
num_distinct_items
min_item_price
max_item_price
total_onshift_dashers
total_busy_dashers
total_outstanding_orders
distance
isWeekend
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

Insight:

•**distance** has the strongest positive correlation with delivery time (0.58).
•**total_onshift_dashers** and **total_busy_dashers** show moderate negative correlations.
•Weakest correlations:
min_item_price, **max_item_price**, **num_distinct_items** (
dropped from the model).

Similarly,
1) Residuals are randomly scattered around zero (Residuals vs Predicted Values)
2. Points align with the red line, indicating residuals are normally distributed. (Q_Q Plot)
3. Bell-shaped distribution further validates normality.

Key Takeaways

1. **Distance is the dominant factor** in delivery time.
2. **Dasher availability** significantly reduces delays.
3. **Price-related features** (min/max item price) were irrelevant and removed.
4. **Model assumptions** (linearity, normality, homoscedasticity) were validated.