# Parcel Delivery Time Prediction - Linear Regression

## 1. Problem Statement

The objective is to build a linear regression model that predicts the time it will take to deliver a parcel based on factors such as total busy dashers, distance, and other operational metrics.

## 2. Business Objective

- Improve accuracy in estimated delivery times (ETAs).
- Optimize resource allocation by understanding delivery delays.
- Provide better customer experience through accurate prediction models.

## 3. Data Preprocessing

- Timestamps were converted to datetime format.
- Delivery duration was calculated in minutes.
- Missing values were handled using forward fill.
- Features selected: total_onshift_dashers, total_busy_dashers, total_outstanding_orders, distance.

## 4. Model Building & Evaluation

Linear Regression model was trained on standardized feature values. Recursive Feature Elimination (RFE) was applied to identify the most significant features.

Selected Features: ['total_busy_dashers', 'total_outstanding_orders', 'distance']

Evaluation Metrics:
- $R^2$ Score: ~0.74
- RMSE: ~9.5 minutes

## 5. Insights

- Distance and number of busy dashers significantly affect delivery time.

# Parcel Delivery Time Prediction - Linear Regression

- Residuals are mostly normally distributed, suggesting the model fits well.
- There may still be nonlinear effects or categorical drivers not captured by this model.

## 6. Conclusion

The linear regression model provides a good starting point for predicting delivery times. Future iterations can explore advanced models (e.g., Random Forest, XGBoost) and include more features such as weather, traffic, and restaurant preparation time.