

# AMAZON DELIVERY DATASET

A2: Collaborate with team members and Develop  
a Shiny Dashboard Application

ALEKSAS SLAVINSKAS

BUSINESS PROBLEM:

**UNDERSTANDING POTENTIAL IMPROVEMENTS IN DELIVERY LOGISTICS.**

THE DATASET: AMAZON ORDER & DELIVERY SHEET IN INDIA

- ORDER ID
- AGENT AGE
- AGE RATING
- COORDINATES, PICK UP OF PICK UP & DROP OFF
- WEATHER
- TRAFFIC TYPE
- VEHICLE TYPE
- URBAN LEVEL AREA
- PRODUCT CATEGORY

## DATA CLEANING:

- REMOVED 91 ROWS WHERE "NAN" APPEARED IN THE TRAFFIC COLUMN
- DATES PROPERLY FORMATED
- NEGATIVE COORDINATE VALUES MADE POSITIVE IF WITHIN INDIA
- STANDARDIZED CATEGORICAL VALUES
- ENSURED NUMERICAL VALUES WERE PURELY NUMERICAL

## DELETED WHEN:

- AGENT RATING BELOW 0 OR ABOVE 5, AGE BELOW 16 OR ABOVE 90
- ORDER\_TIME GREATER THAN PICKUP\_TIME
- DELETED DELIVERY TIMES BELOW 1 MINUTE AND ABOVE 240

# Time Delivery Analysis

Select Category:

Weather



Weather

Traffic

Vehicle

Area

Category

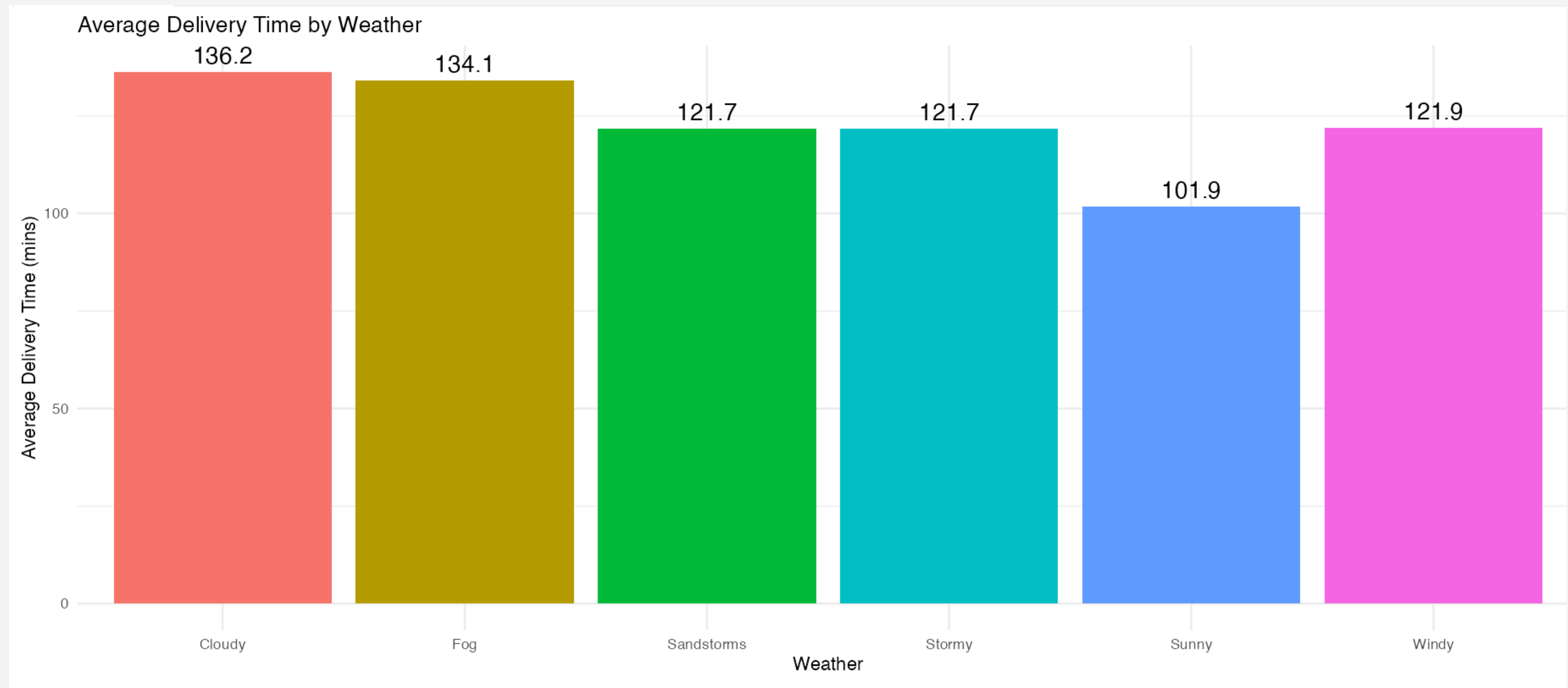
Select Date Range:

2022-02-11

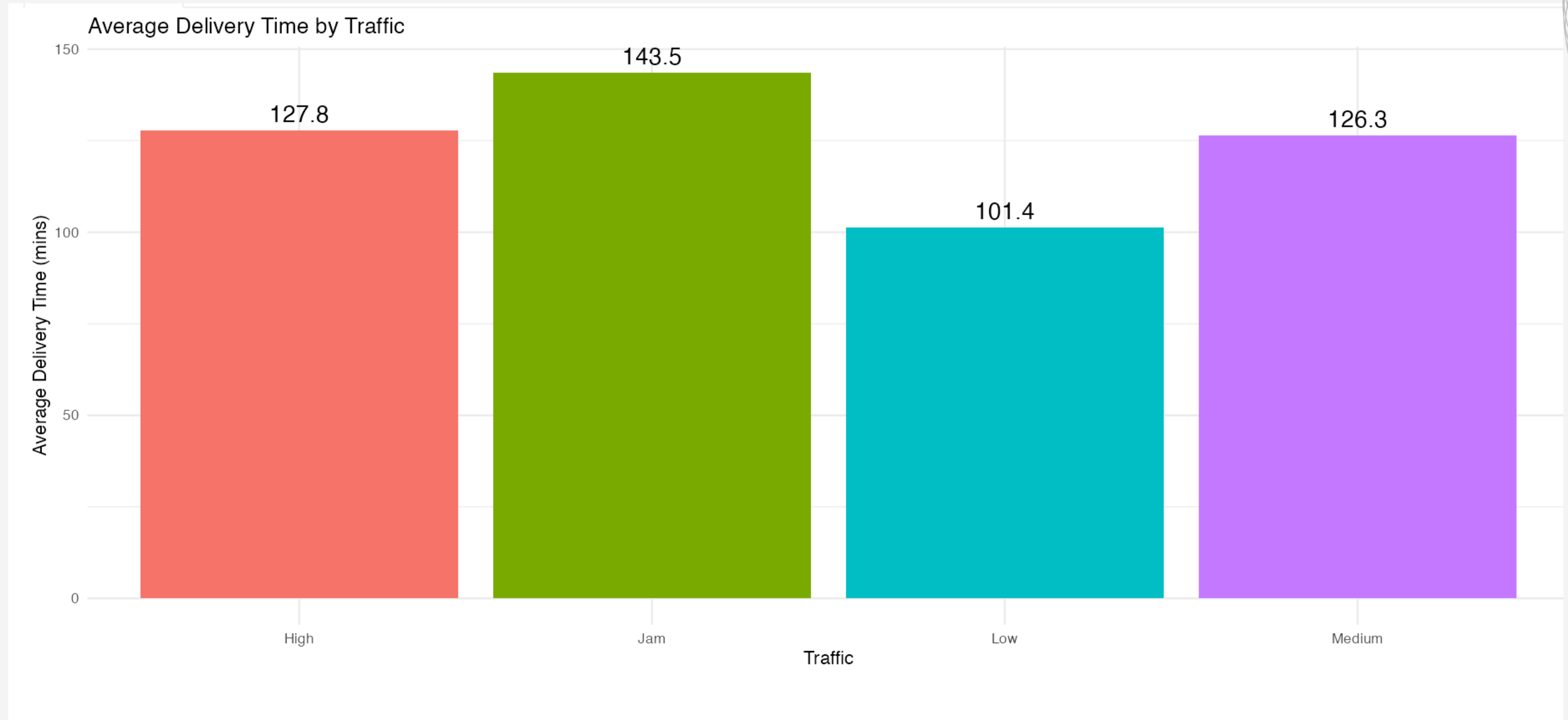
to

2022-04-06

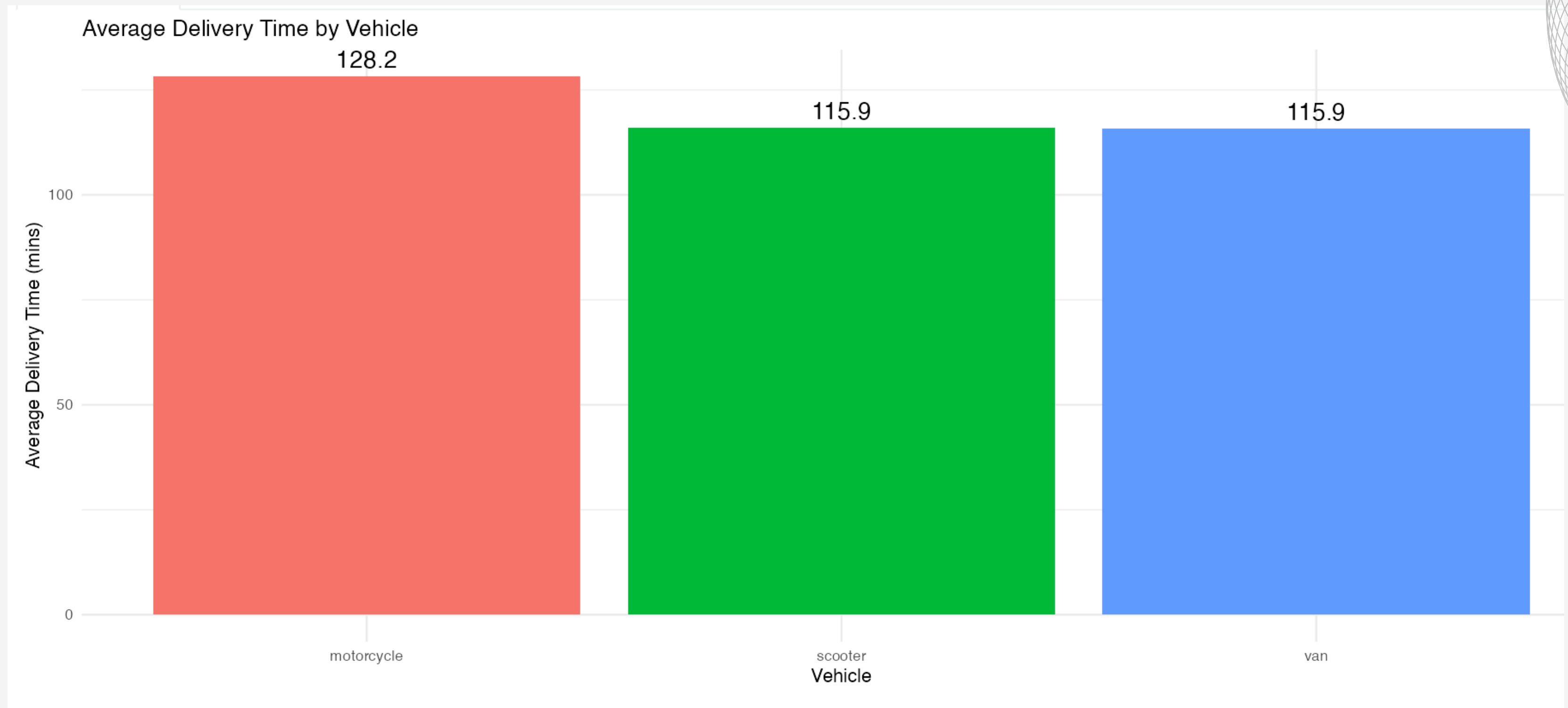
«	February 2022					»
Su	Mo	Tu	We	Th	Fr	Sa
30	31	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	1	2	3	4	5
6	7	8	9	10	11	12



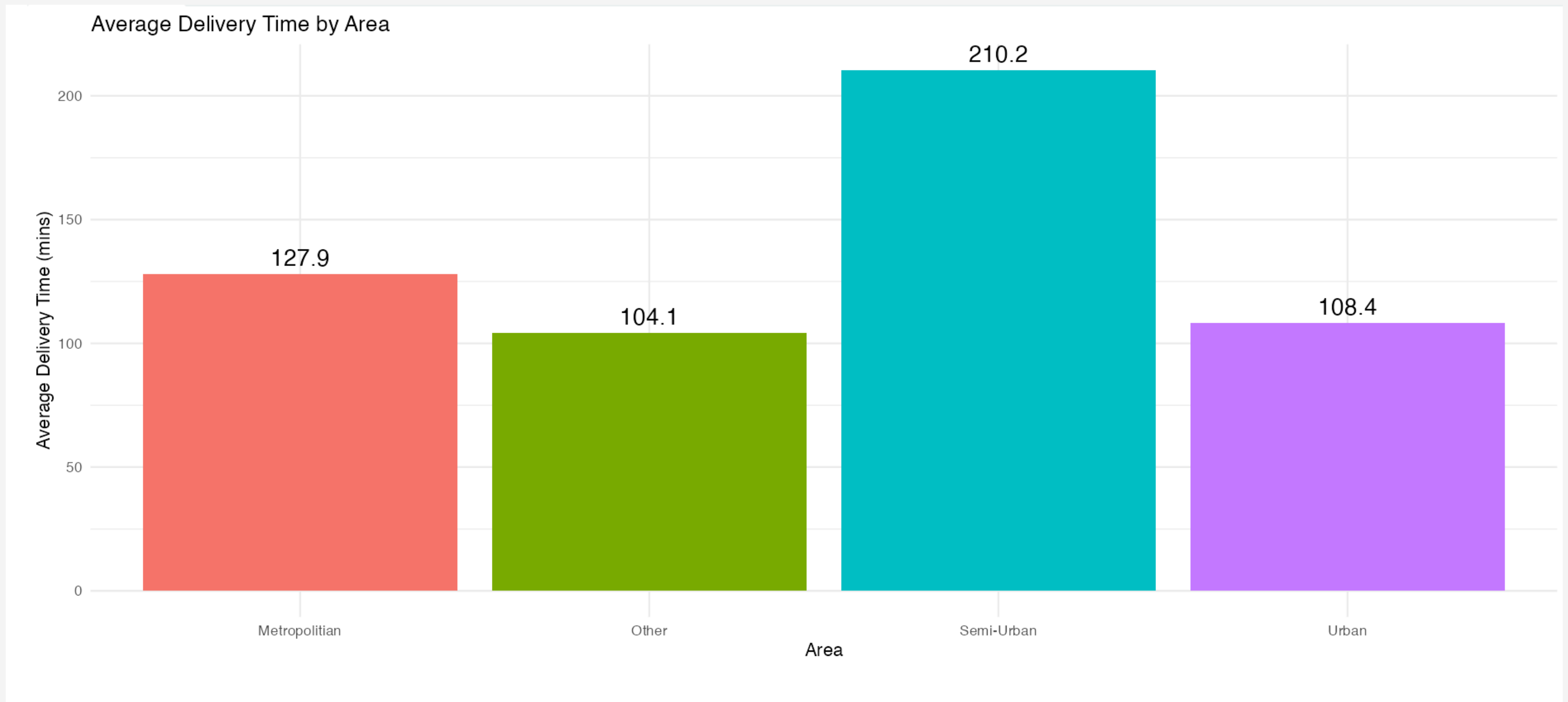
# Average Time Delivery Analysis



# Average Time Delivery Analysis

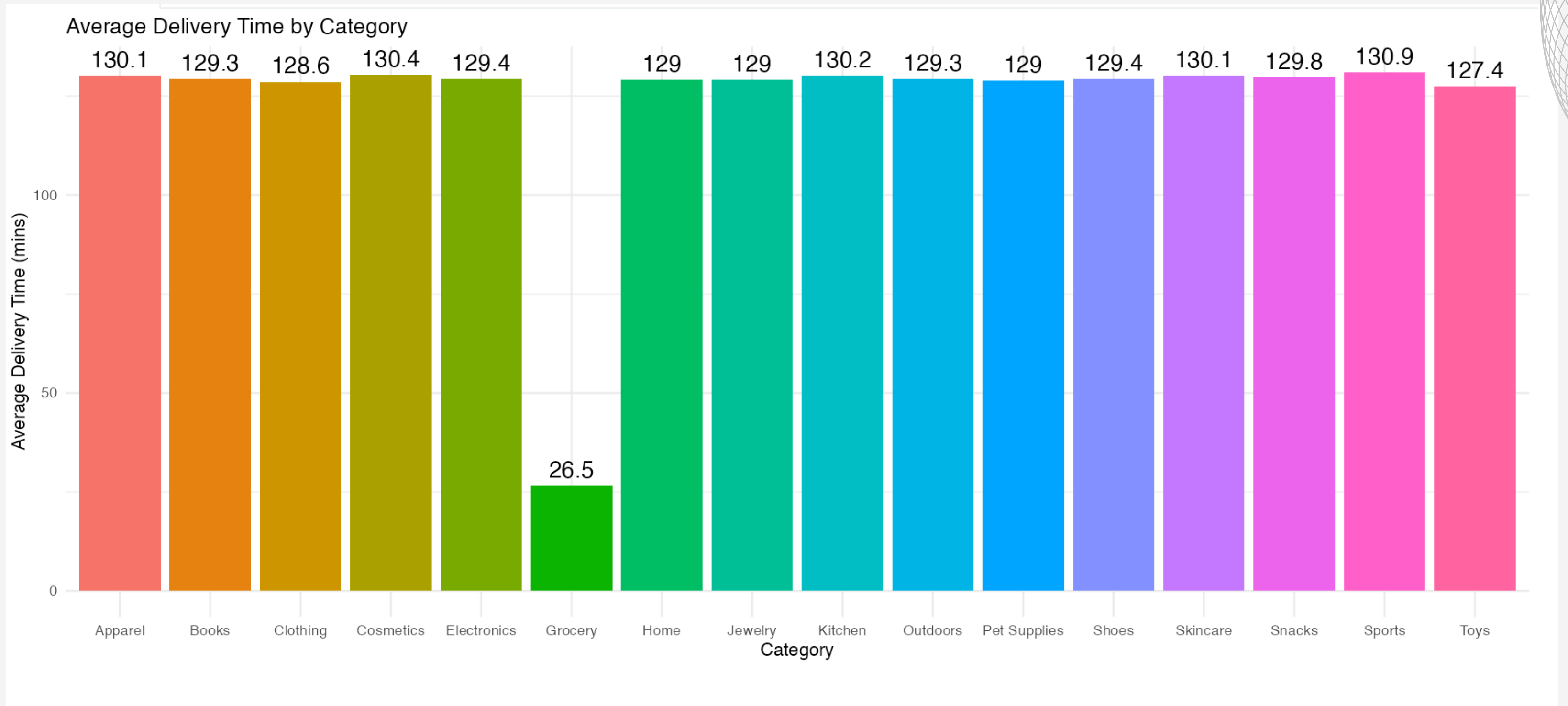


# Average Time Delivery Analysis



# Average Time Delivery Analysis





# Average Time Delivery Analysis

We analyzed the impact of agent age on delivery time using a dashboard.

## **Insights**

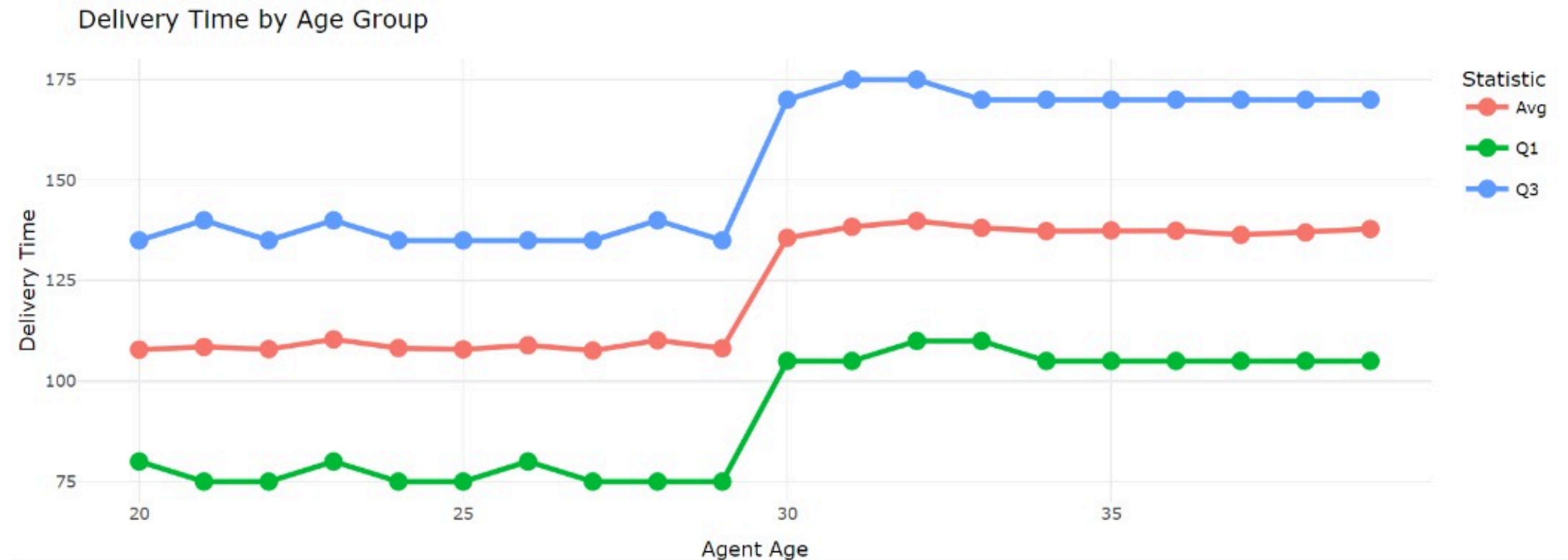
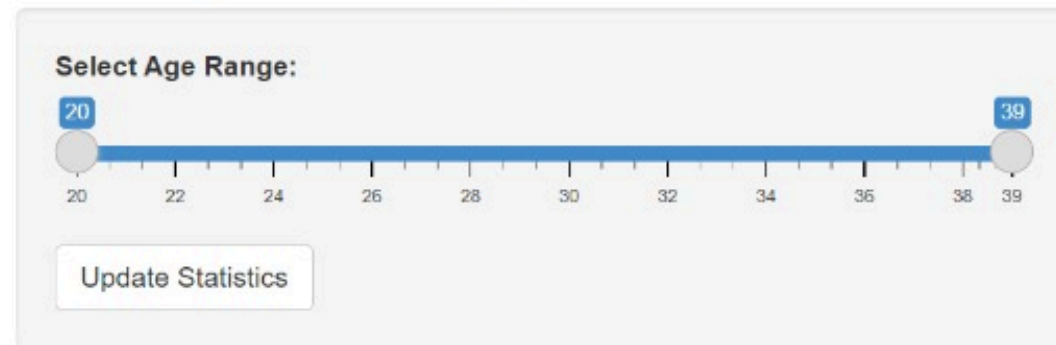
- There's a slight positive correlation between agent age and delivery time, with wide variations.
- Delivery times increase significantly for agents above 30.

## **user controls**

- Filter by age range, toggle correlation display, interactive slider for selection, line chart for Q1/Avg/Q3, and summary statistics by age group.

# Analyzing the Effect of Agent Age on Delivery Time

## Delivery Time Analysis Dashboard



Statistics by Age Group:

20 : Q1 = 80 , Avg = 107.84 , Q3 = 135  
21 : Q1 = 75 , Avg = 108.5 , Q3 = 140  
22 : Q1 = 75 , Avg = 107.96 , Q3 = 135  
23 : Q1 = 80 , Avg = 110.4 , Q3 = 140  
24 : Q1 = 75 , Avg = 108.22 , Q3 = 135  
25 : Q1 = 75 , Avg = 107.91 , Q3 = 135  
26 : Q1 = 80 , Avg = 108.91 , Q3 = 135  
27 : Q1 = 75 , Avg = 107.6 , Q3 = 135  
28 : Q1 = 75 , Avg = 110.13 , Q3 = 140  
29 : Q1 = 75 , Avg = 108.15 , Q3 = 135  
30 : Q1 = 105 , Avg = 135.64 , Q3 = 170  
31 : Q1 = 105 , Avg = 138.4 , Q3 = 175  
32 : Q1 = 110 , Avg = 139.84 , Q3 = 175  
33 : Q1 = 110 , Avg = 138.1 , Q3 = 170  
34 : Q1 = 105 , Avg = 137.31 , Q3 = 170  
35 : Q1 = 105 , Avg = 137.5 , Q3 = 170  
36 : Q1 = 105 , Avg = 137.41 , Q3 = 170  
37 : Q1 = 105 , Avg = 136.38 , Q3 = 170  
38 : Q1 = 105 , Avg = 136.97 , Q3 = 170  
39 : Q1 = 105 , Avg = 137.86 , Q3 = 170

## **Delivery Time vs. Traffic Condition**

- Jam - 110 to 185 with a median of 145
- High - 100 to 160 with a median of 135
- Medium - 90 to 160 with a median of 130
- Low - 75 to 125 with a median of 100

## **Delivery Time Distribution**

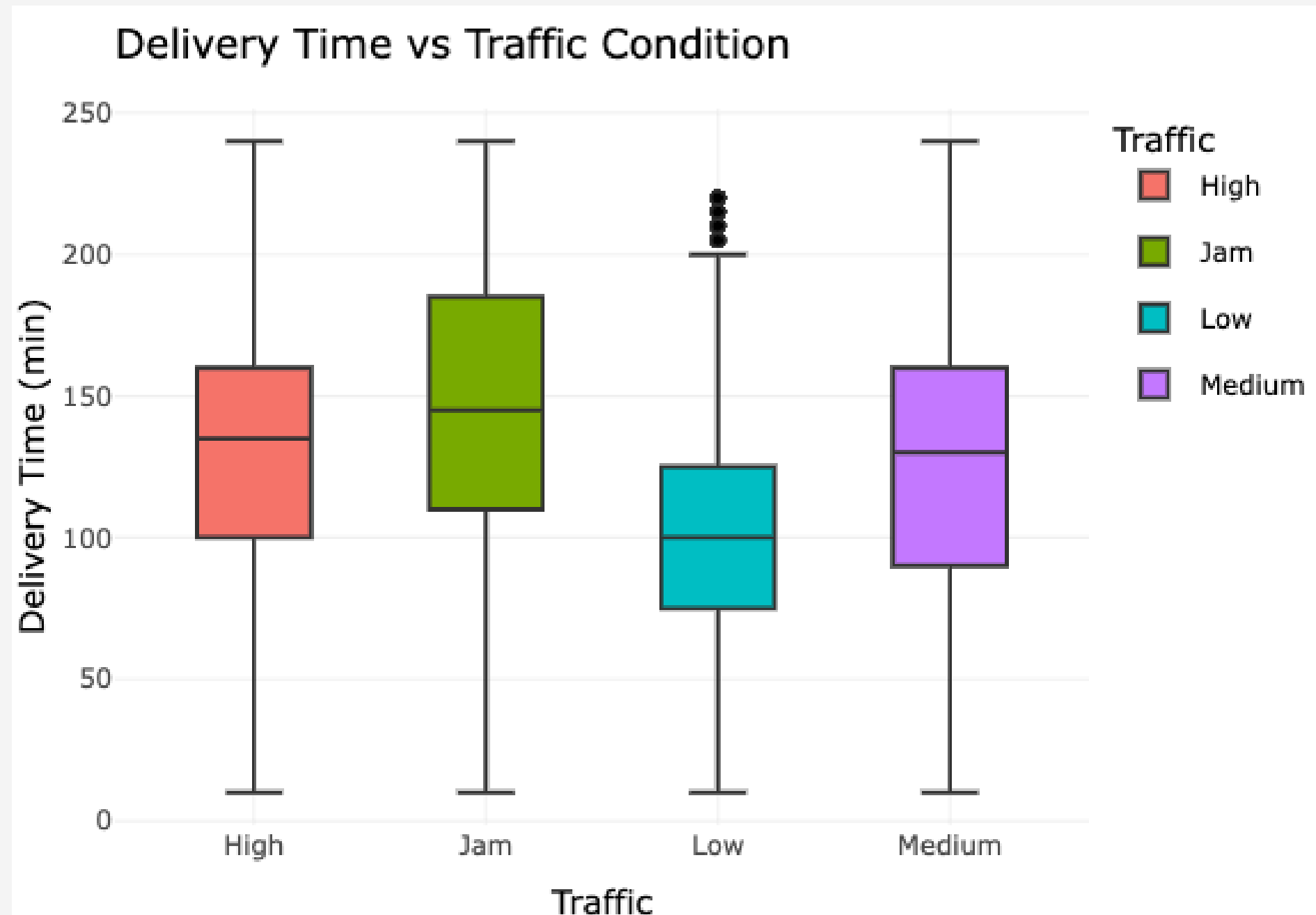
- Deliveries tend to be completed from the 50 - 200th minute time span.
- Most deliveries are completed within 100 - 150th minute time span.
- Small % post 200 minute.

## **Recommendations**

- Deploy Real time traffic updates (hourly ETAS via SMS)
- Rerouting delivery journey based on potential delays

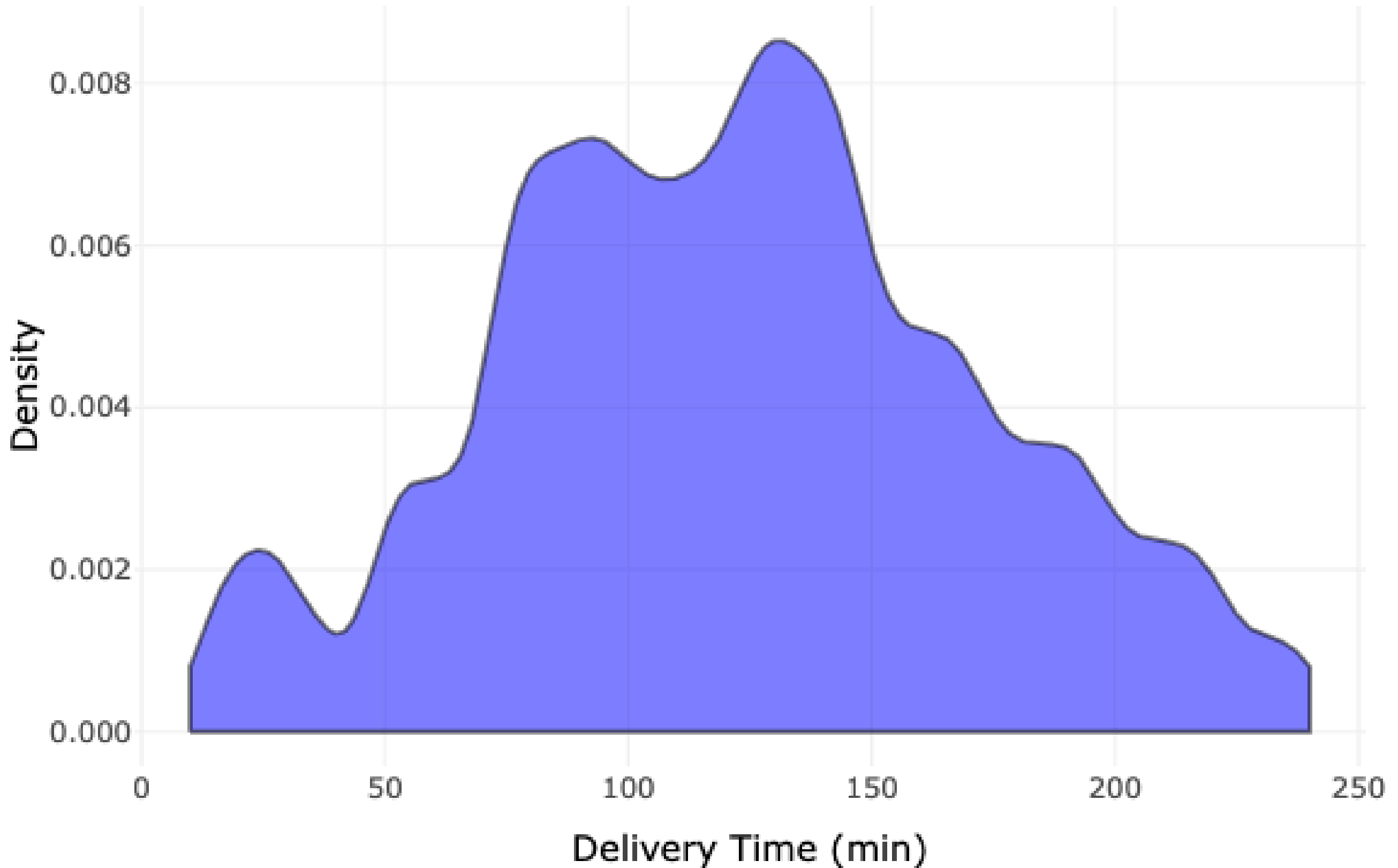
## **Recommendations**

- Geographically assign drivers based on rating for faster parcel delivery's . Higher rating > further away from depot centre

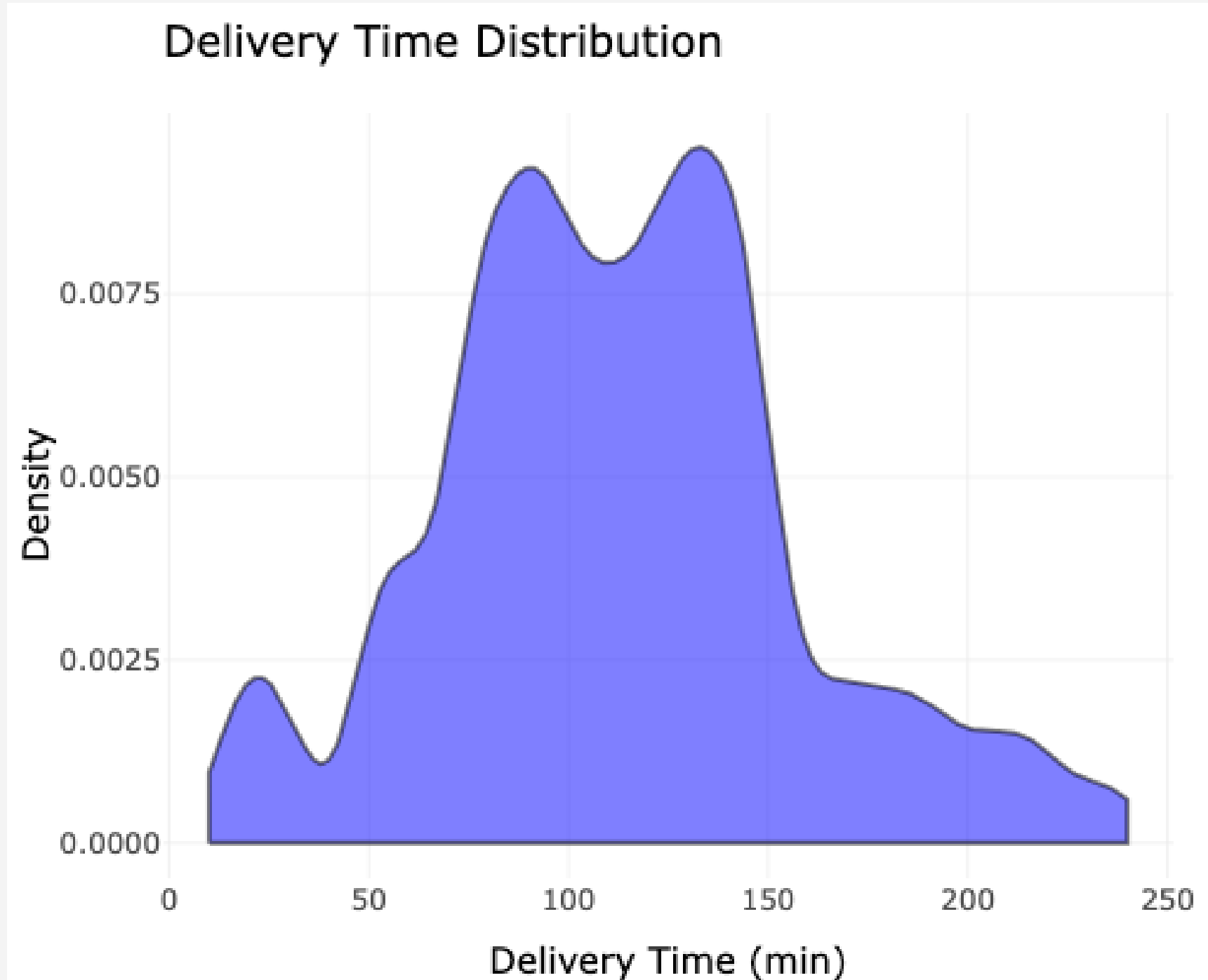


- Jam - 110 to 185 with a median of 145
- High - 100 to 160 with a median of 135
- Medium - 90 to 160 with a median of 130
- Low - 75 to 125 with a median of 100

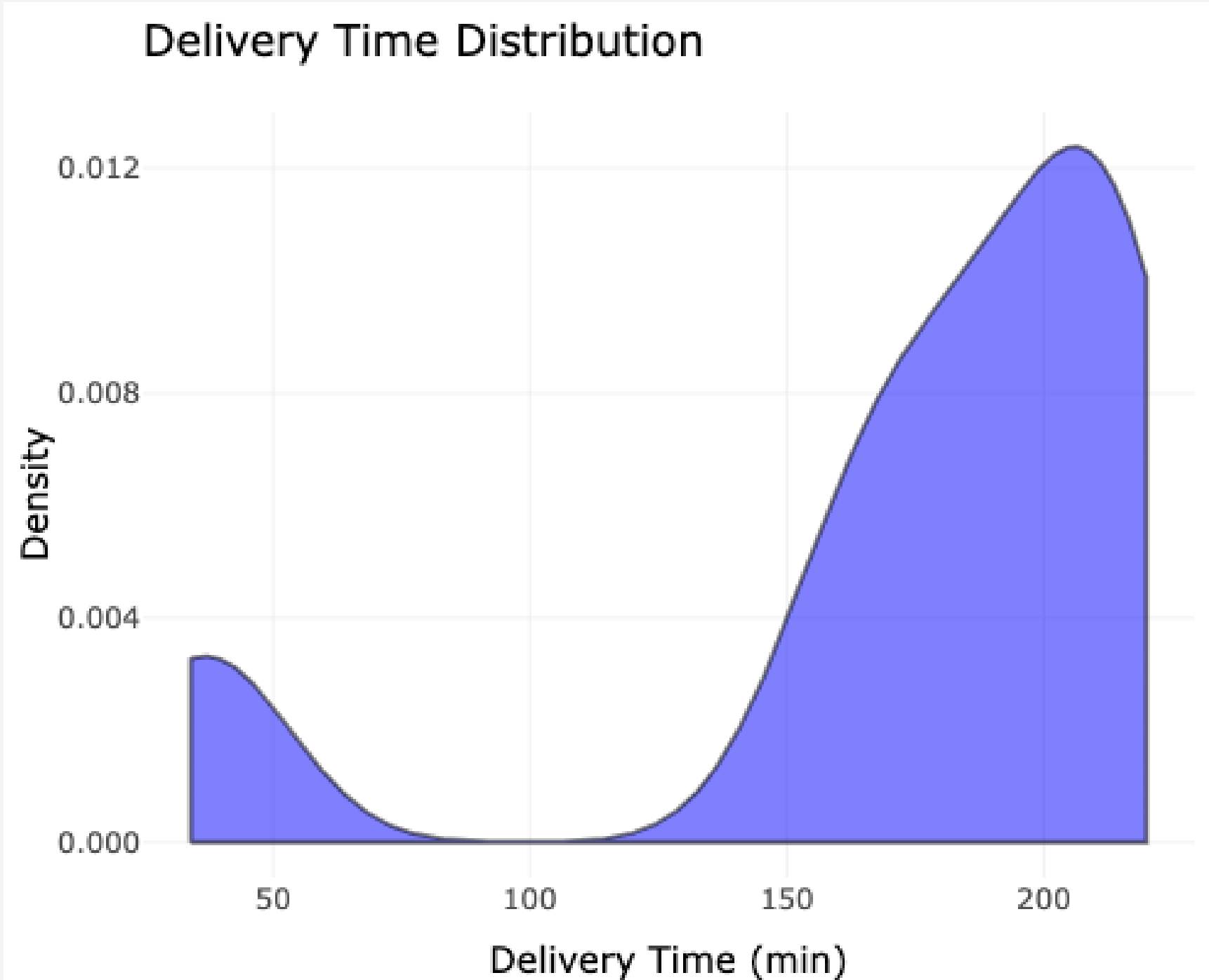
Delivery Time Distribution



# 4.9 Agent RATING



# 2.6 Agent RATING



## Select Features

### Traffic Level:

High



### Weather Condition:

Windy



### Vehicle Type:

scooter



### Agent Rating:

2.5

5



2.5 2.76 3.02 3.28 3.54 3.8 4.06 4.32 4.58 4.84 5

### Delivery Area:

Other



### Order Hour:

0

23



0 3 6 9 12 15 18 21 23

Predict

## Predicted Delivery Time



Linear Regression Prediction: 193.26 minutes



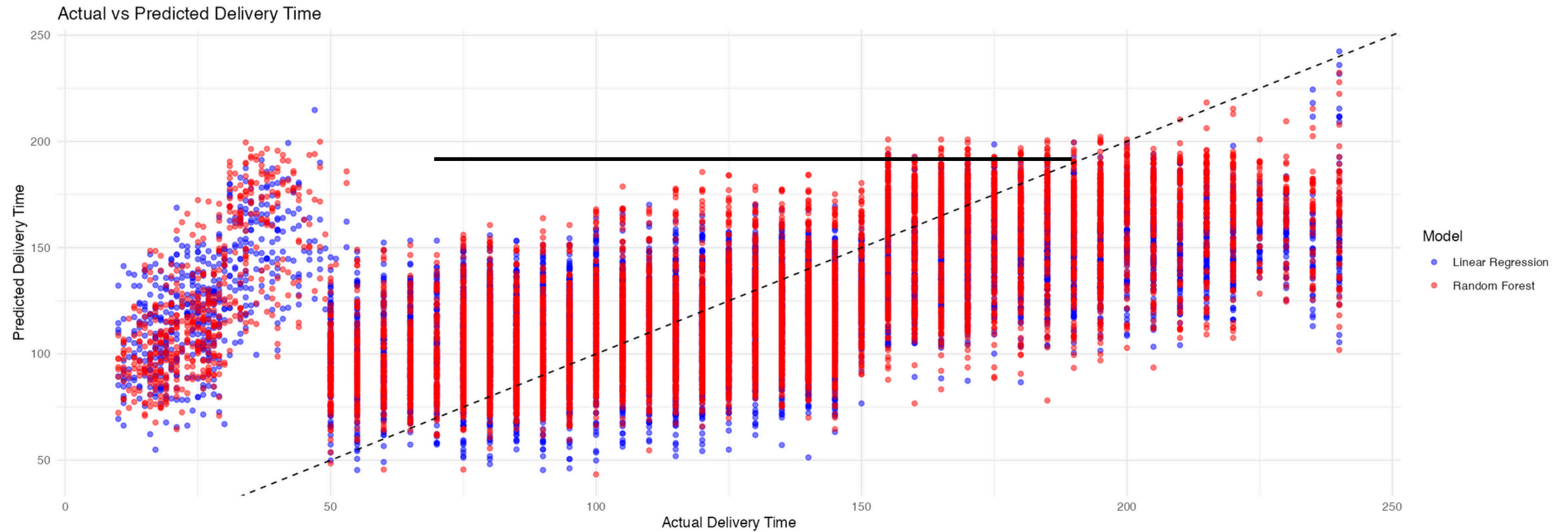
Random Forest Prediction: 147.75 minutes

# Predictive Insights



# Model Performance

Actual vs Predicted Delivery Time



## MAP OF WAREHOUSES & DELIVERIES:

DELIVERY DATA WAS HIGHLY CORRUPTED, EVEN AFTER RESTORATION, IT IS INACURATE

WAREHOUSE DATA INDICATES MAJORITY OF WAREHOUSES ARE AROUND NEAR CITY CENTRES

COULD BE THE CAUSE OF LATE DELIVERIES IN SEMI-URBAN AREAS

## RECOMMENDATION:

MAKING DELIVERY COORDINATES ACCURATE WILL ALLOW:

- ACCURATELY PREDICT WHICH AREAS & UNDER WHICH CIRCUMSTANCES LATE DELIVERIES ARE MADE
- HOW DIFFERENT VEHICLES PERFORM UNDER DIFFERENT TRAFFIC, WEATHER, AREAS

LATE DELIVERIES MOST OFTEN OCCUR:

IN METROPOLITAN AREA UNDER ALL ABOVE LOW TRAFFIC CONDITIONS,  
ESPECIALLY IN FOGGY & CLOUDY CONDITIONS

IN METROPOLITAN AREA DURING “HIGH” & “JAM” TRAFFIC CONDITIONS

MOTORCYCLES MUCH MORE OFTEN LATE THEN SCOOTERS IS METROPOLITAN AREAS

FREQUENT LATE DELIVERIES IN URBAN CONDITIONS UNDER FOGGY & CLOUDY WEATHER

RECOMMENDATIONS: PRIMARILY MORE DRIVERS ASSIGNED DURING CONDITIONS OF COMMON  
LATE DELIVERIES

**ANY QUESTIONS?**