



LAST MILE DELIVERY

Optimizing Customer Delivery Time

How might we optimize customer delivery time
by predicting future orders from historical data ?

Problem Space

1

Supply Chain

Last Mile Delivery

Final step of the supply chain process where the order arrives at the customer's doorsteps

2

E-commerce

Growth of e-commerce

In 2020, e-retail sales surpassed 4.2 trillion U.S. dollars worldwide (COVID-19)

3

Customer Delivery

Delivery Time

The last mile is arguably the biggest problem in e-commerce as it addresses the question, “When will I receive my order?”.

4

Industry

Development on CDT

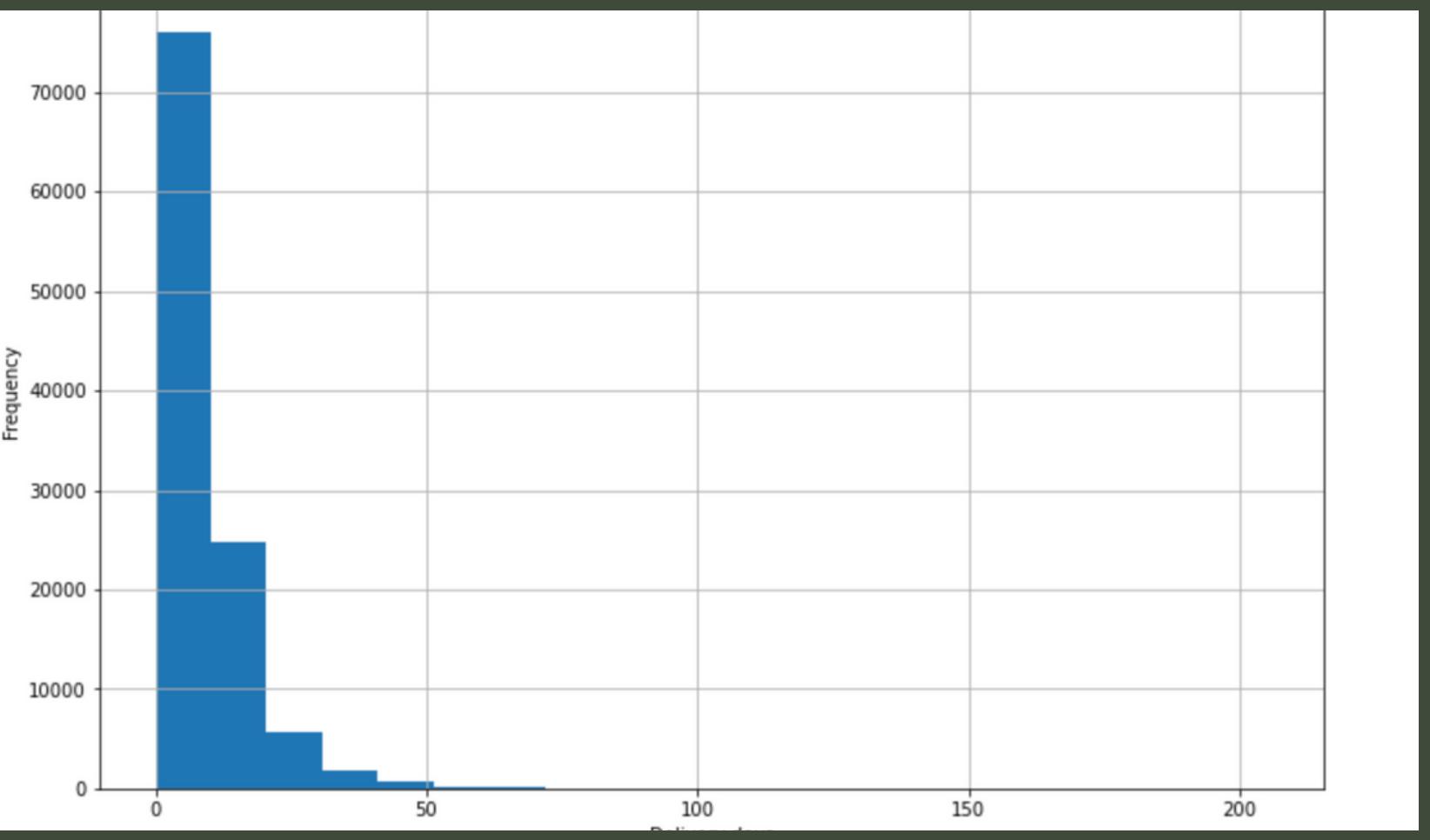
Industry leaders, such as Shopify, Amazon, and Uber have focus on improving CDT which impact on sales and overall service delivery.

The Data

Olist e-commerce data set

- The data has ~110k orders from Oct. 2016 to Oct. 2018.
- The data has 8 different tables
- The target is to predict the delivery time and compare it with previous predictions done by the Olist team.

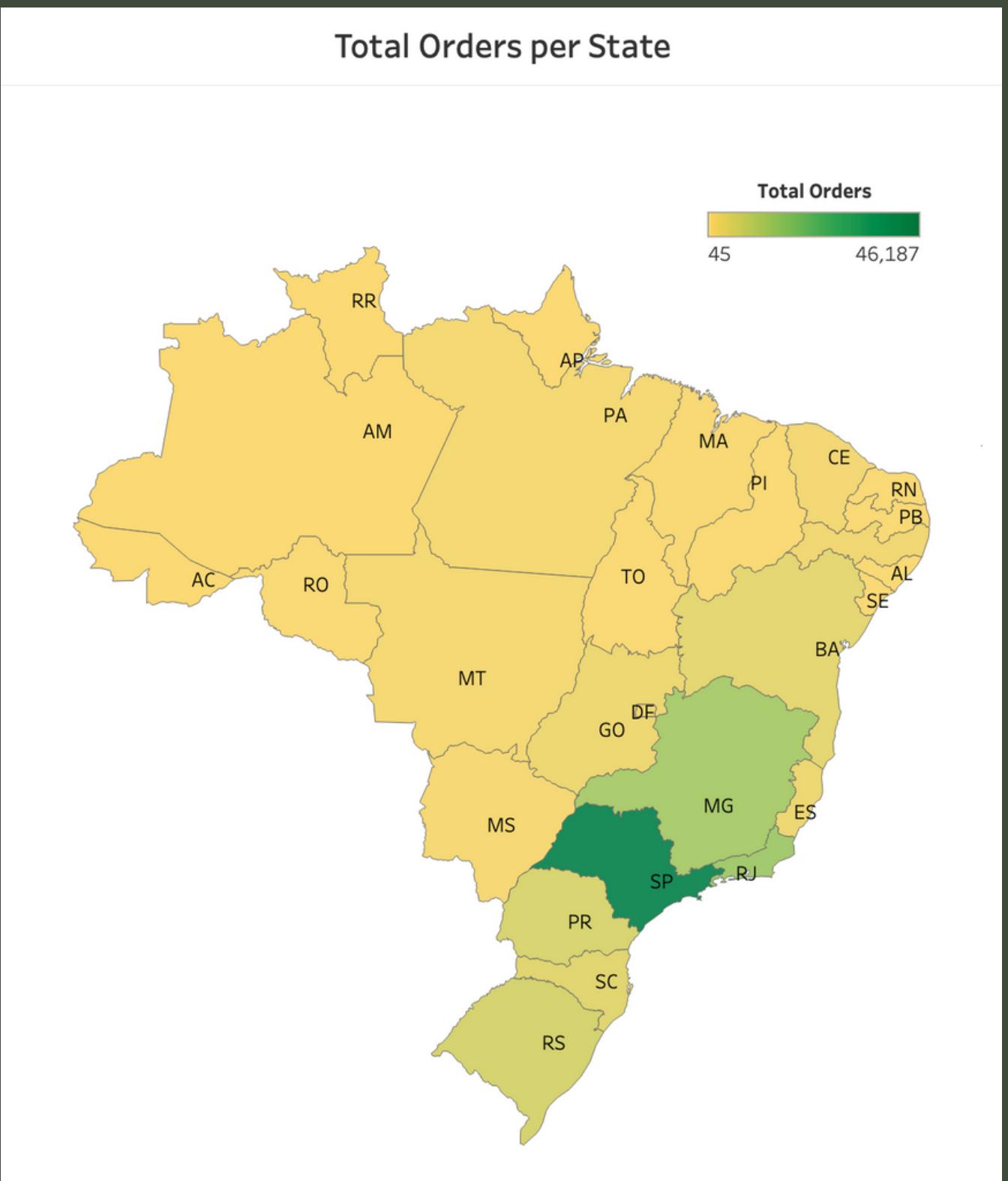
Delivery time in Days



EDA

Data Exploration - Geolocation

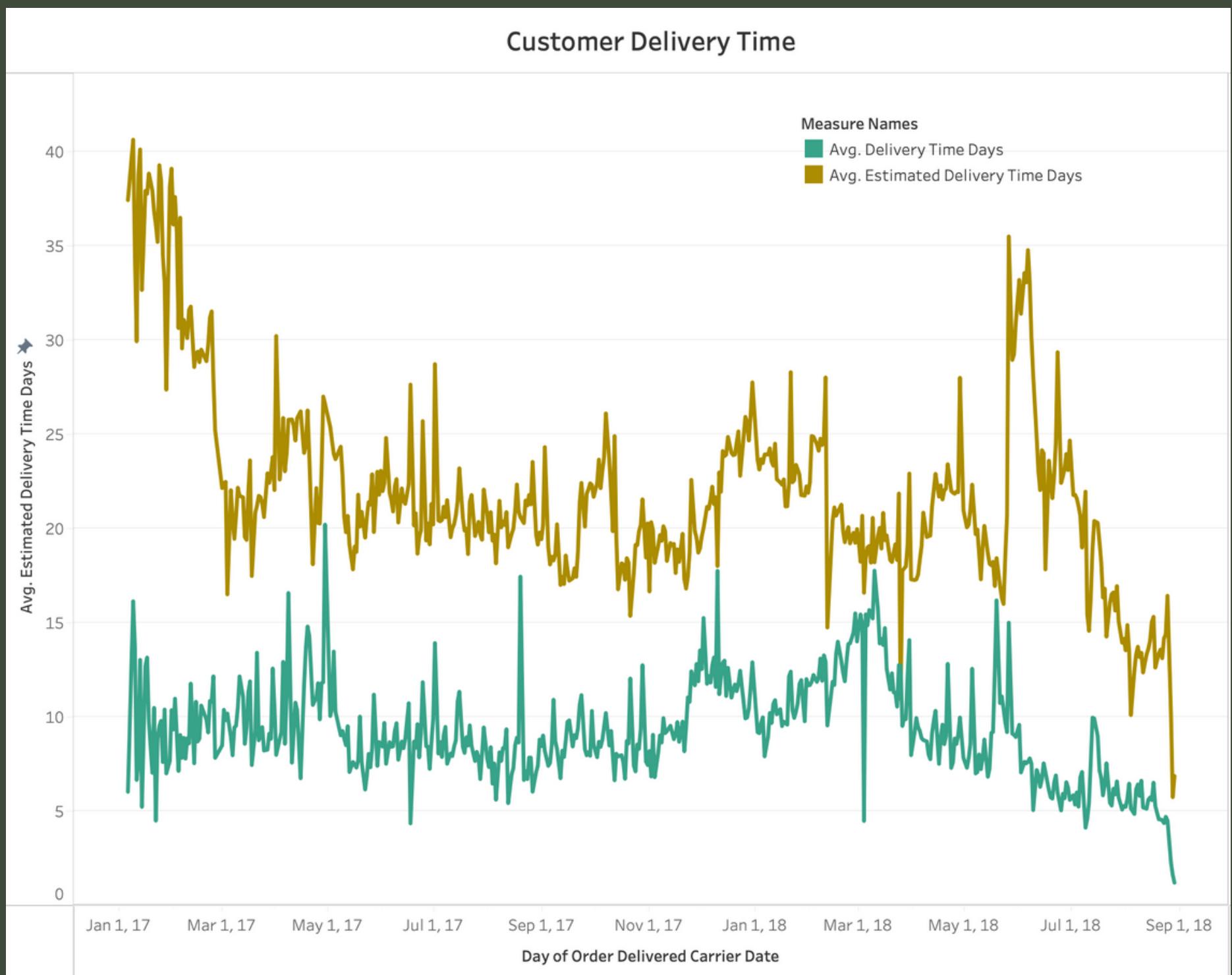
- Most of the orders originate from the states of São Paulo (~41% of all orders) and Rio de Janeiro (~13% of all orders), both are also coincidentally cities too.
- Most of the orders are coming from coastal cities



EDA

Data Exploration - Orders

- Olist overestimates delivery times by about 11 days on average where the average delivery time is 9days
- Actual delivery from EDA shows delivery can take up to 200days after logistics partner pickup.
- The target was divided into 3 categories- **3-days, 3-7days, and over 7 days**, due to its heavy skewness



Feature Selection

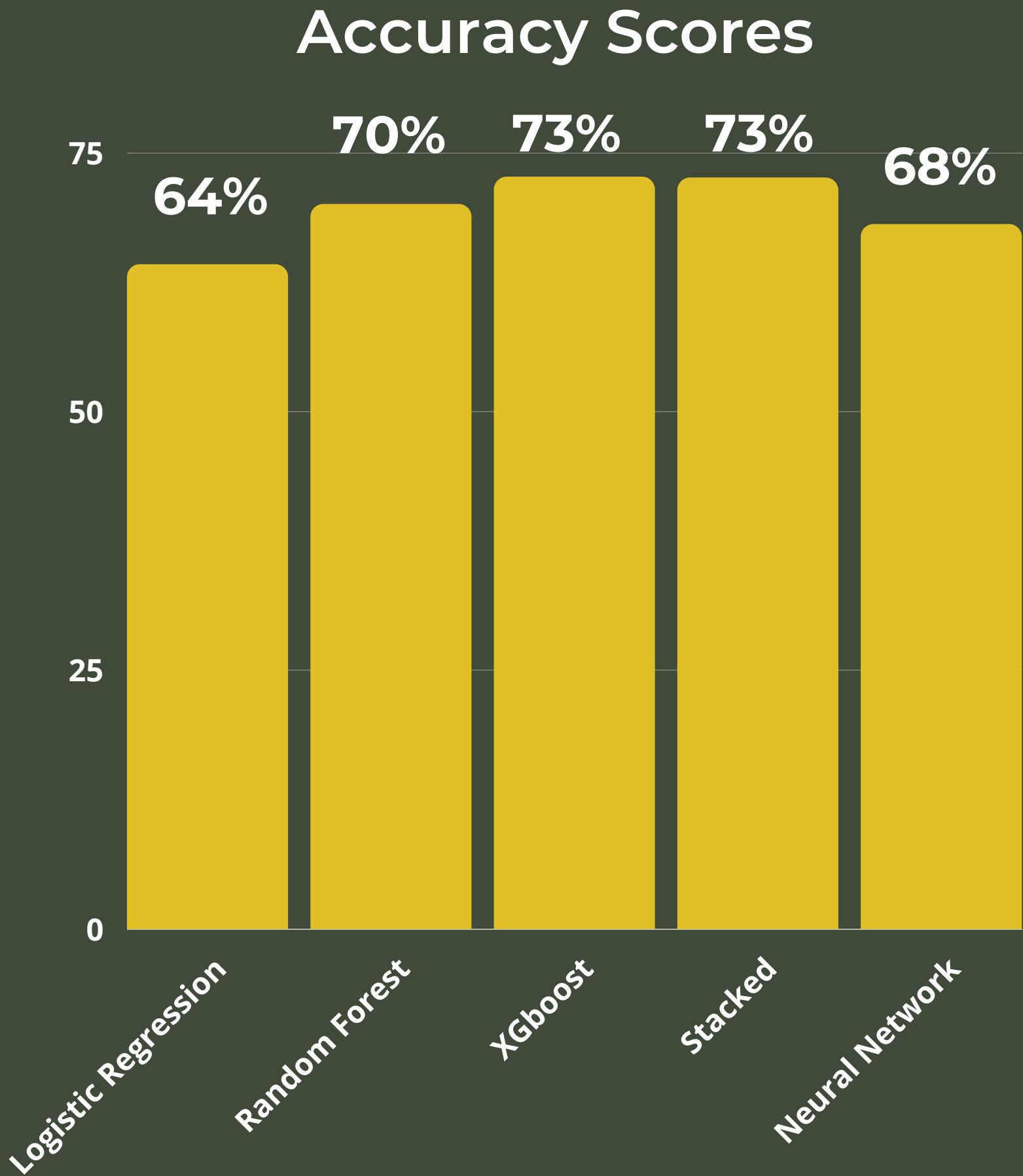
Features are divided into three:
Customer, Orders and External factors

- **Distance**
- **Holidays** - scraped from internet
- Time based features (day of week, week of year)
- State category- This was due to imbalance in our state data
- **Same state delivery**
- **Product weight and size - these were log transformed**



Modeling & Findings

- 5 different models were used for the modeling
- XGBoost is the best model with accuracy of ~73%
- Predicting deliveries 3-7 days appears to be very difficult and there is no underlining pattern with these orders
- XGBoost did a good job at predicting the delivery window with a certain amount of confidence and this can help reduce the cost associated with deliveries.



Next Steps

- A more robust dataset is required that would give a good level of detail on the logistics partner operations, fulfillment centers and type of customer.
- Explore and research more on how other types of neural networks(RNN and CNN) can help in improving accuracy of our model.
- Investigate further on why the accuracy for the 3-7 days delivery window, performing badly.



Thank You for Listening



<https://github.com/Oluwafemi-Jegede>



<https://www.linkedin.com/in/oluwafemijegede>



jegede45@gmail.com



Tools

