

ANALYSIS FOR

JUMPMAN23

By Hua-lai, Tseng (Justin)

INTRO

- **Background:**

Jumpman23 is an on-demand delivery platform connecting “Jumpmen” and customers purchasing a variety of goods. **Jumpman23** will send Jumpmen to merchants to purchase and pickup any items requested by the customer. Whenever possible, **Jumpman23** will order the requested items ahead to save the Jumpmen time. Each time a **Jumpman23** delivery is completed, a record is saved to the **Jumpman23** database that contains information about that delivery. **Jumpman23** is growing fast and has just launched in its newest market -- New York City.

- **Challenge:**

The CEO of **Jumpman23** has just asked you how are things going in New York ? The CEO is considering pouring more resources into NYC and wants you to come up with a plan to grow the market by 20% in two months.

They have also heard of data integrity issues in the data from NYC. Please think through the CEO’s ask and present to both the CEO and the CTO, in any format you choose, an analysis of the market. In addition, dive into the reports on data integrity issues and if they indeed exist, outline where they may be and how they may impact the analysis. The CEO is a visual learner, the CTO loves to see code and technical work, and both are obsessed with maps.

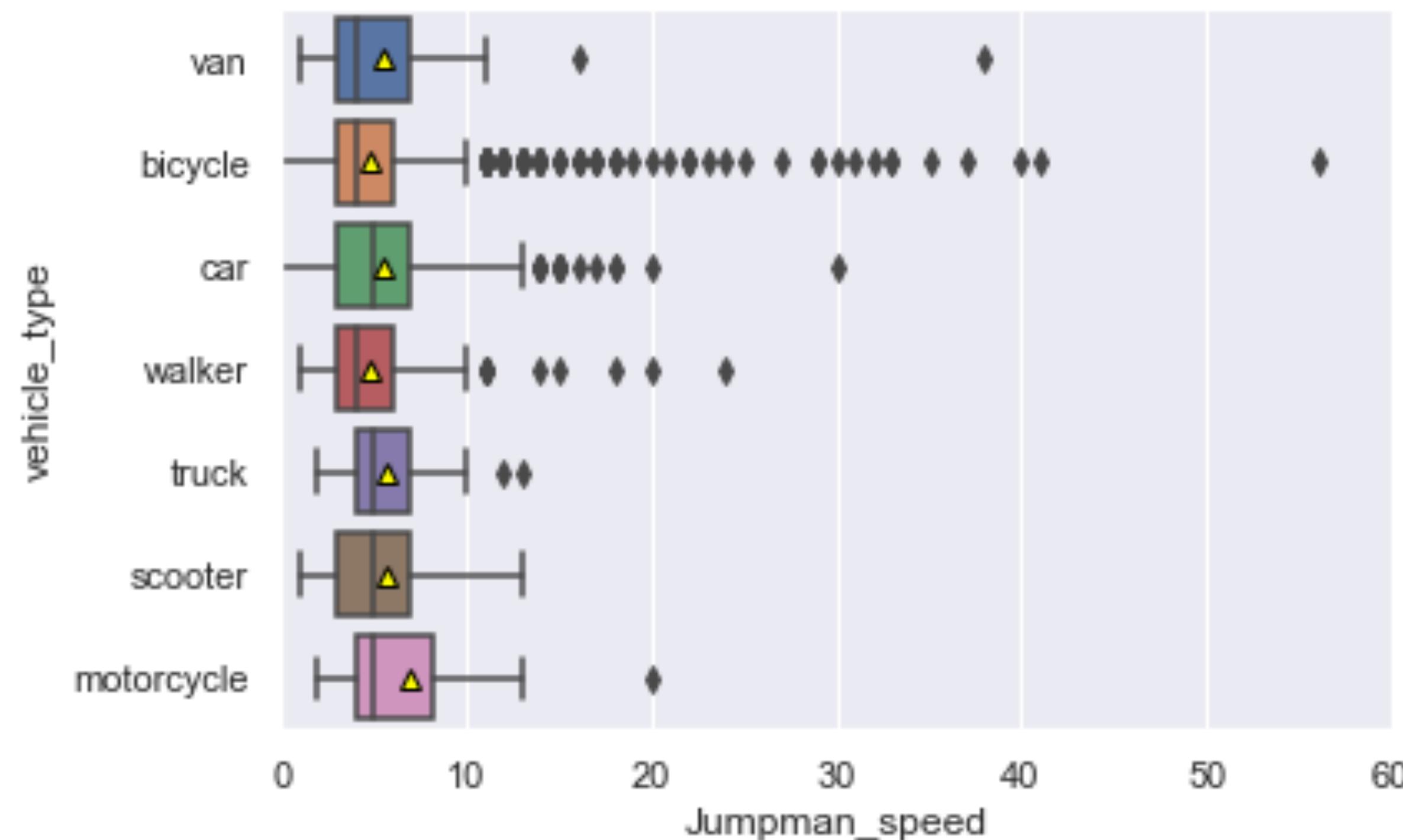
PROBLEM BREAKDOWN

- 1. Data integrity issues?
 - Suspicious Speed
 - Negative time interval
 - Summary
- 2. How are things going in NYC?
 - About Orders
 - About Customers
 - About Jumpmen
 - About Geo
- 3. Come up with a plan to grow the market by 20% in two months.
 - Strategies from analysis
 - Strategies from ML
 - Conclusion & future Studies

DATA INTEGRITY ISSUES

SUSPICIOUS SPEED

— SUSPICIOUS SPEED



— SOLUTION

Typical walking speed is 3.4mph, running: 6.67mph, cycling: 12mph, car 29-67mph, from the box plot on the left we notice some abnormal speed (ex: bicycle at 57mph).

Going to remove the outliers for biker and walker. For speed that is around 0, might be due to traffic or drivers have problem finding destinations, not going to remove them.

NEGATIVE TIME INTERVAL

■ NEGATIVE TIME INTERVAL

```
jump_1['extra_food_prep_time'].sort_values(ascending=True).head()
```

```
1478    -1 days +23:58:55.682249  
623     -1 days +23:59:35.885382
```

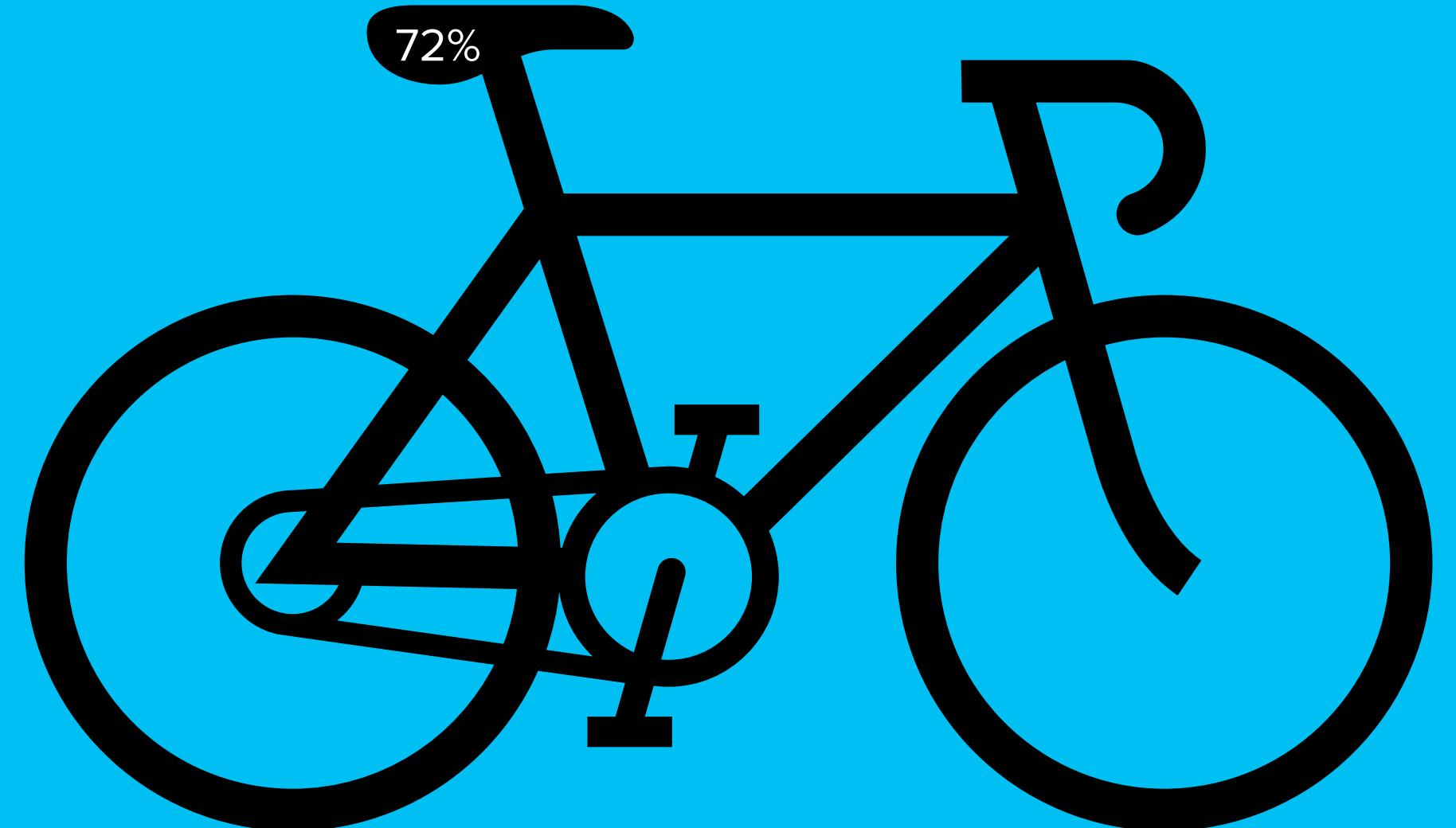
■ SOLUTION:

Extra_food_prep_time column is created by (when_the_Jumpman_left_pickup - when_the_delivery_started), so negative values in this column indicates that the Jumpman left the pickup before the delivery started, this is not reasonable. All negative values in this column are dropped.

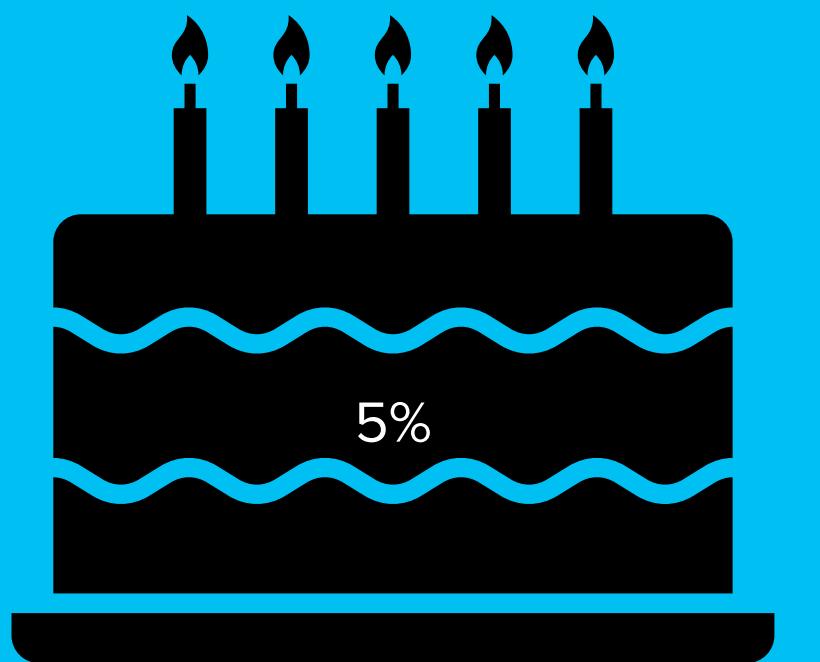
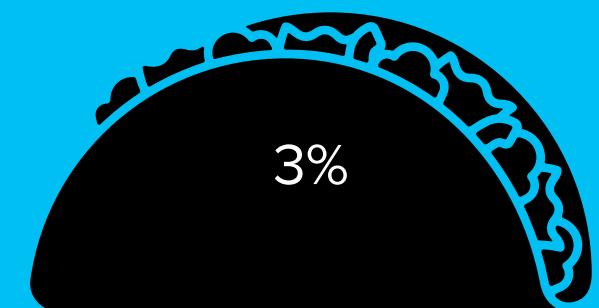
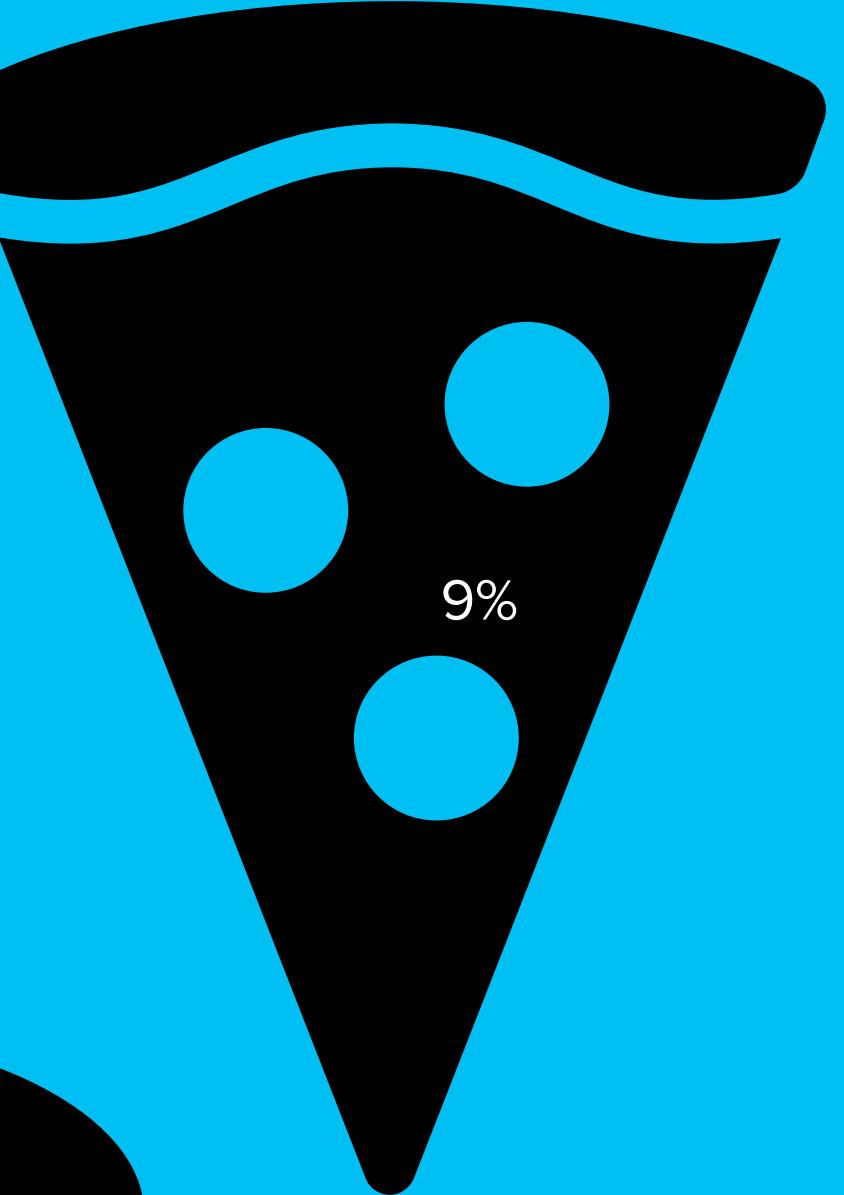
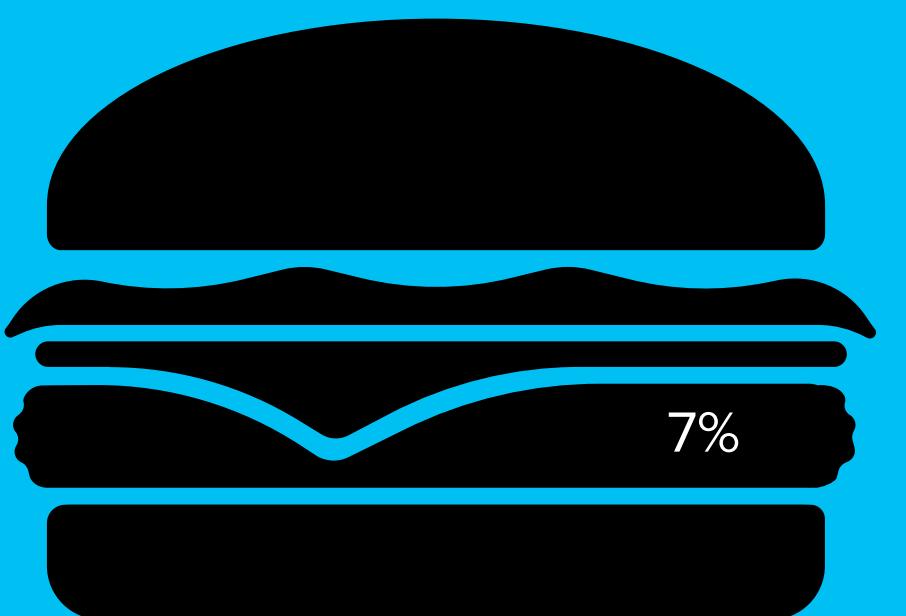
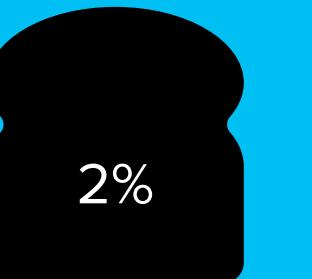
SUMMARY

- After inspecting every column, this data set does not have serious integrity issues. For categorical columns, there do not exist serious bias (eg: A: 99% of the data, B: 1% of the data). For continuous columns, though there are some outliers, the majority of the data are normal.
- Other minor issues & solution
 - Item columns: Items related columns is the reason why there exist duplicate id, by sort with item_quantity & delivery_id, going to drop duplicates.
 - Place_category: Replace null values with others.
 - For time columns: 'when_the_Jumpman_arrived_at_pickup' & 'when_the_Jumpman_left_pickup', both of these columns have very detail data, it is not ideal to replace them with average since it will affect analysis for particular date and time. Going to drop around 500 rows here.
 - how_long_it_took_to_order: Almost half of the values in this column are missing, after analysis, going use average to replace missing value.
 - Why not remove negative values in time_for_Jumpman_to_arrive_at_pickup column?

Full report & code: <https://github.com/hakunamatata9981/Jumpman23/blob/master/Jumpman23%20-%20Data%20Wrangling.ipynb>



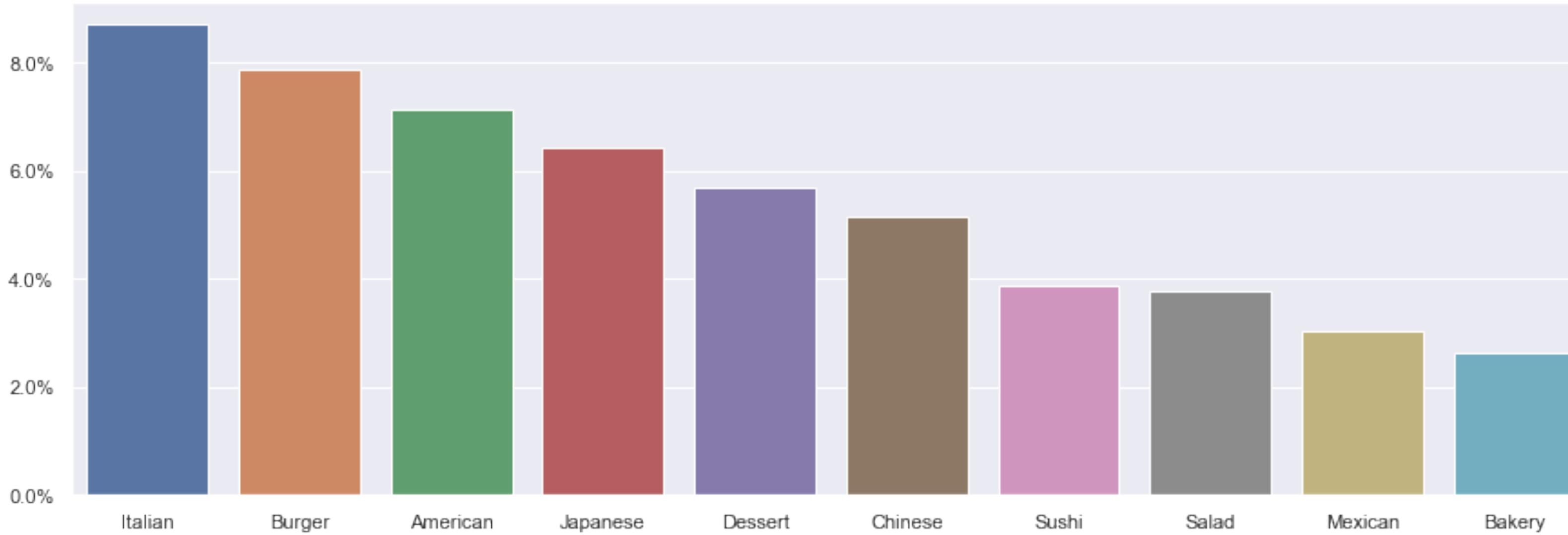
HOW ARE THINGS GOING IN NYC



ABOUT ORDERS

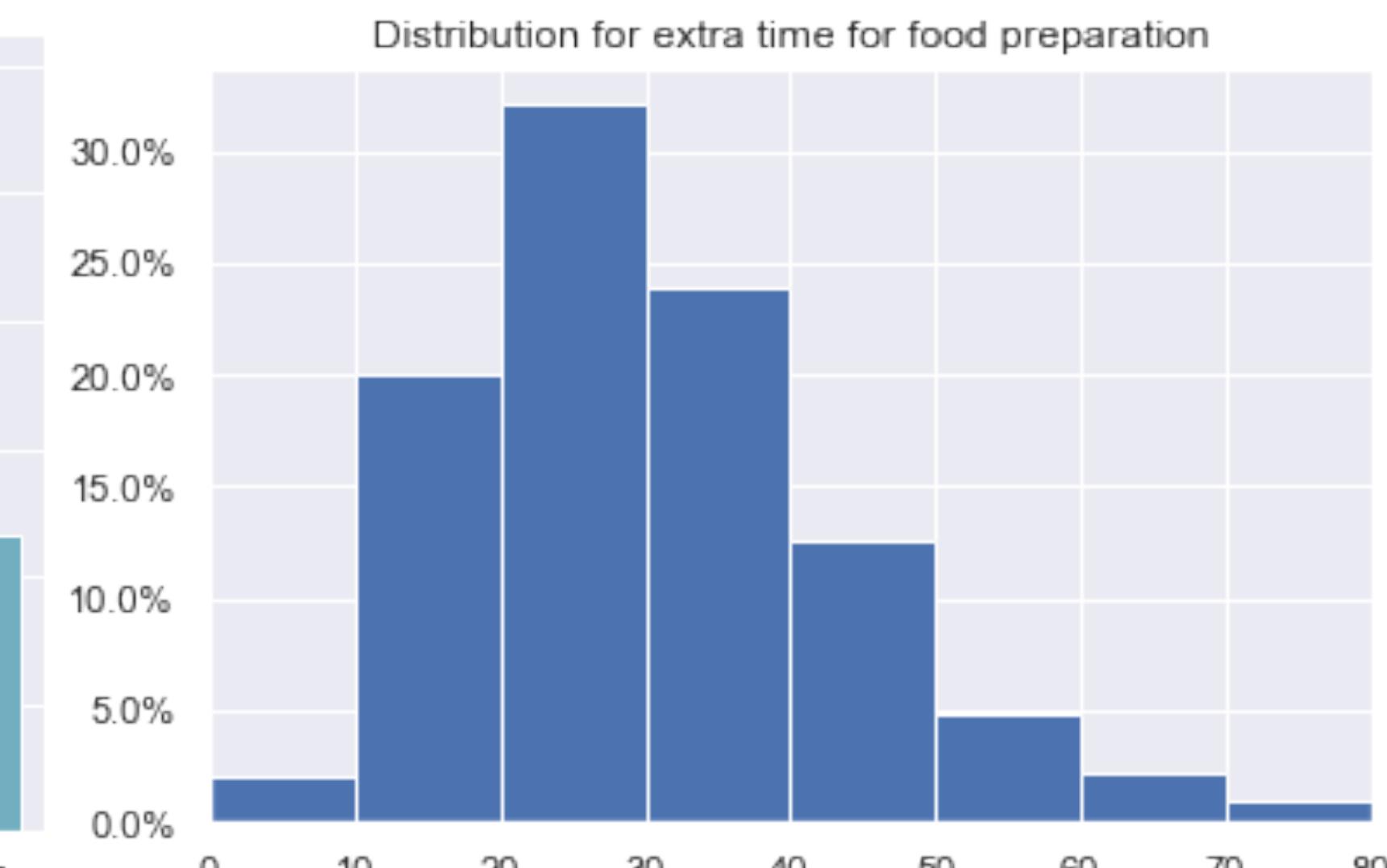
- When the first delivery start: 2014-10-01 00:26:32
- When the last delivery complete: 2014-10-30 23:29:45
- After filtering out suspicious data and orders with 1+ item, there are: 4646 orders in the first month
- Orders in the first month are from 828 different place

Top 10 places customers order from in the first month



ABOUT ORDERS

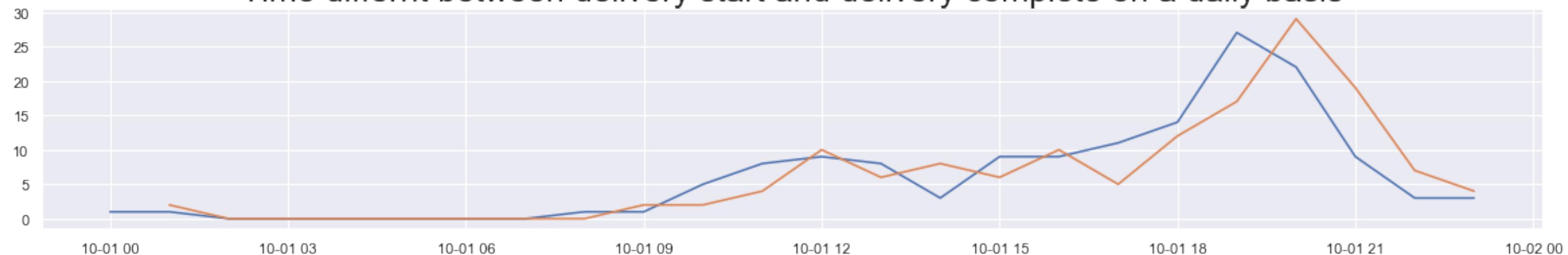
- Average quantity of items in a single order 1 item
- Average extra time for food preparation: 31 minutes



"Naked ball" is an Italian dish from The Meatball Shop.

ABOUT ORDERS

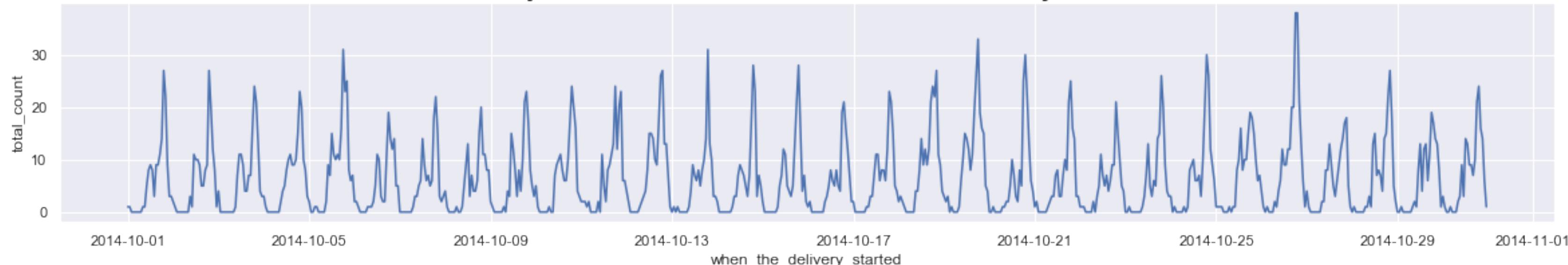
Time differnt between delivery start and delivery complete on a daily basis



Hourly trend for week 1 when delivery started



Hourly trend for first month when delivery started

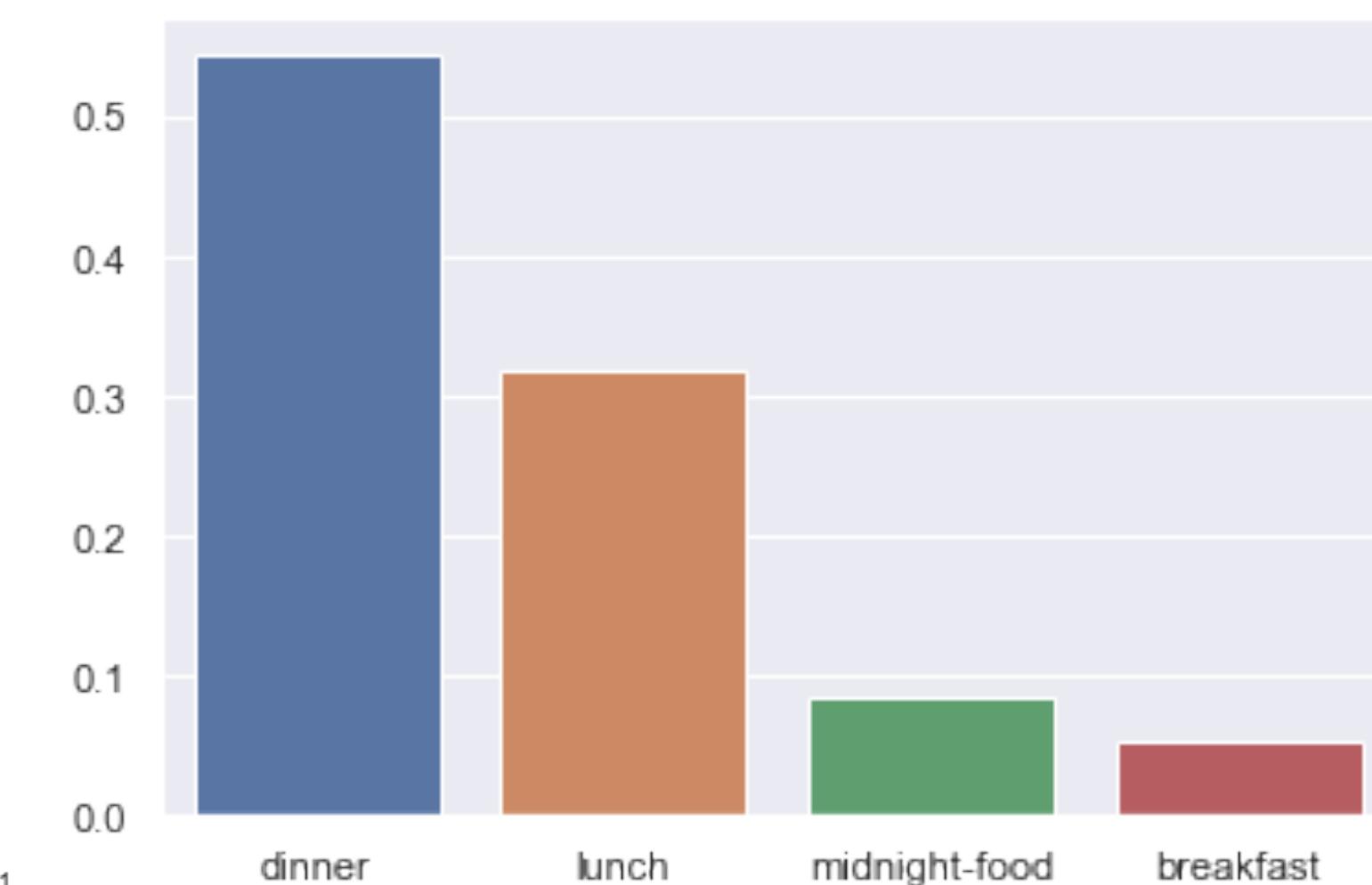


total_count

when_the_delivery_started

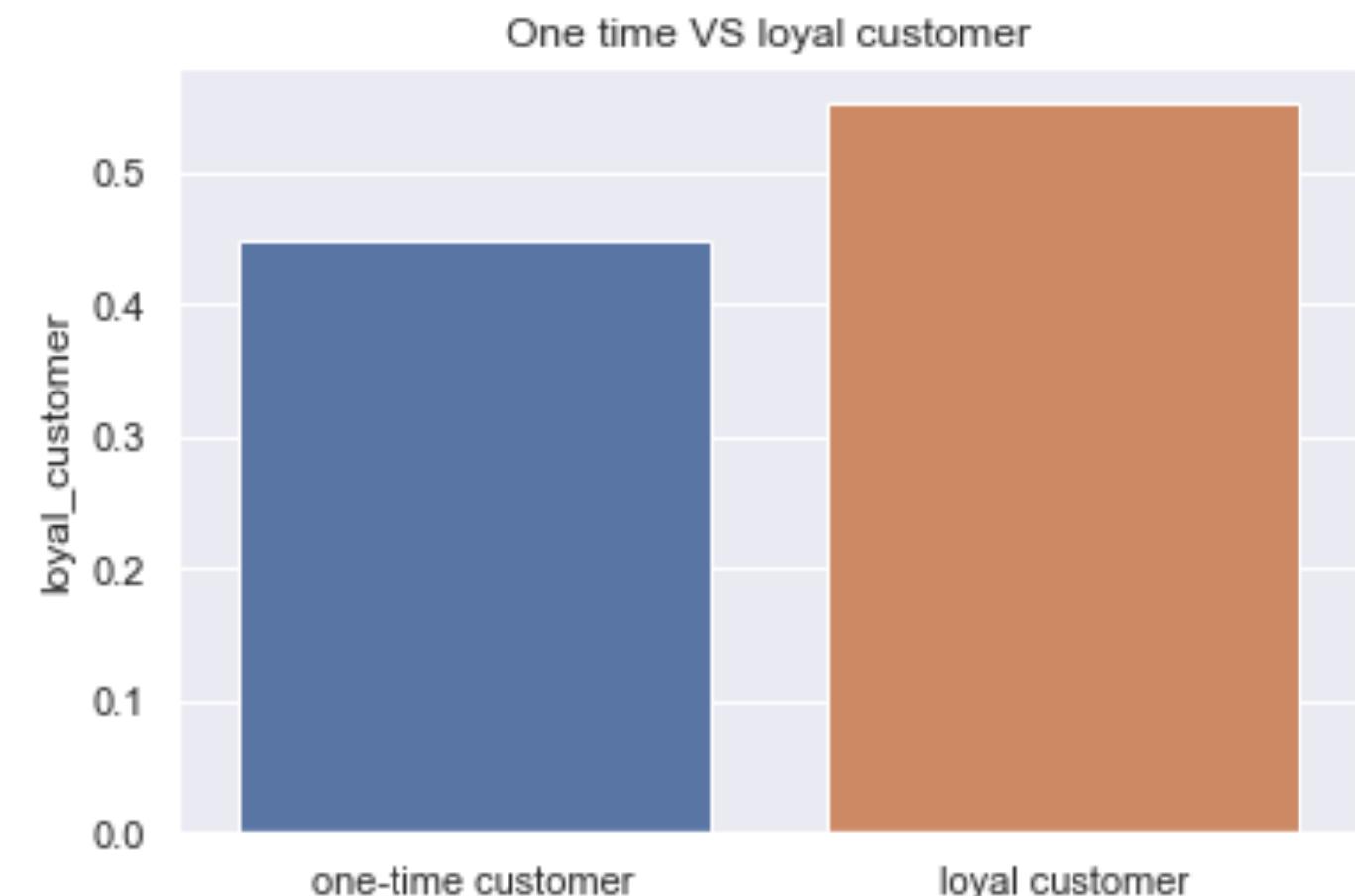
when_the_delivery_started	total_count
2014-10-07	1029
2014-10-14	1061
2014-10-21	1127
2014-10-28	1119

Distribution of meal ordered

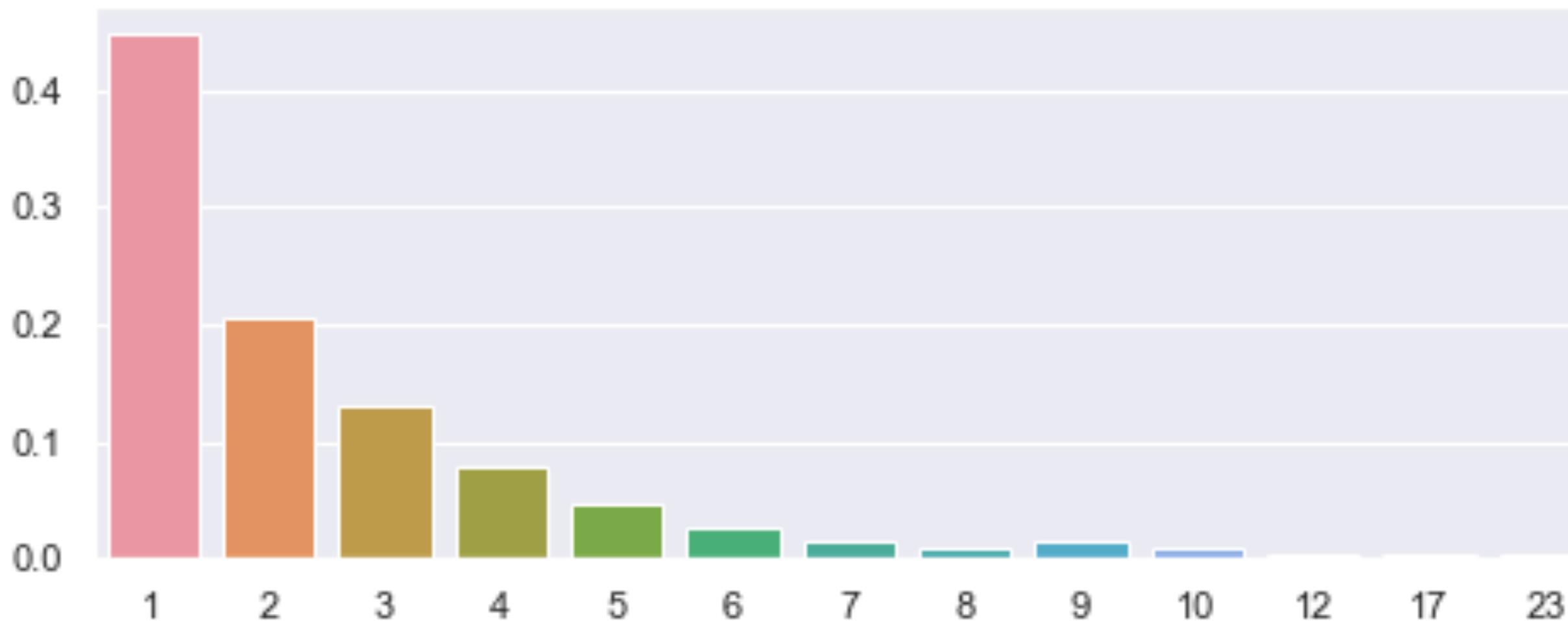


ABOUT CUSTOMER

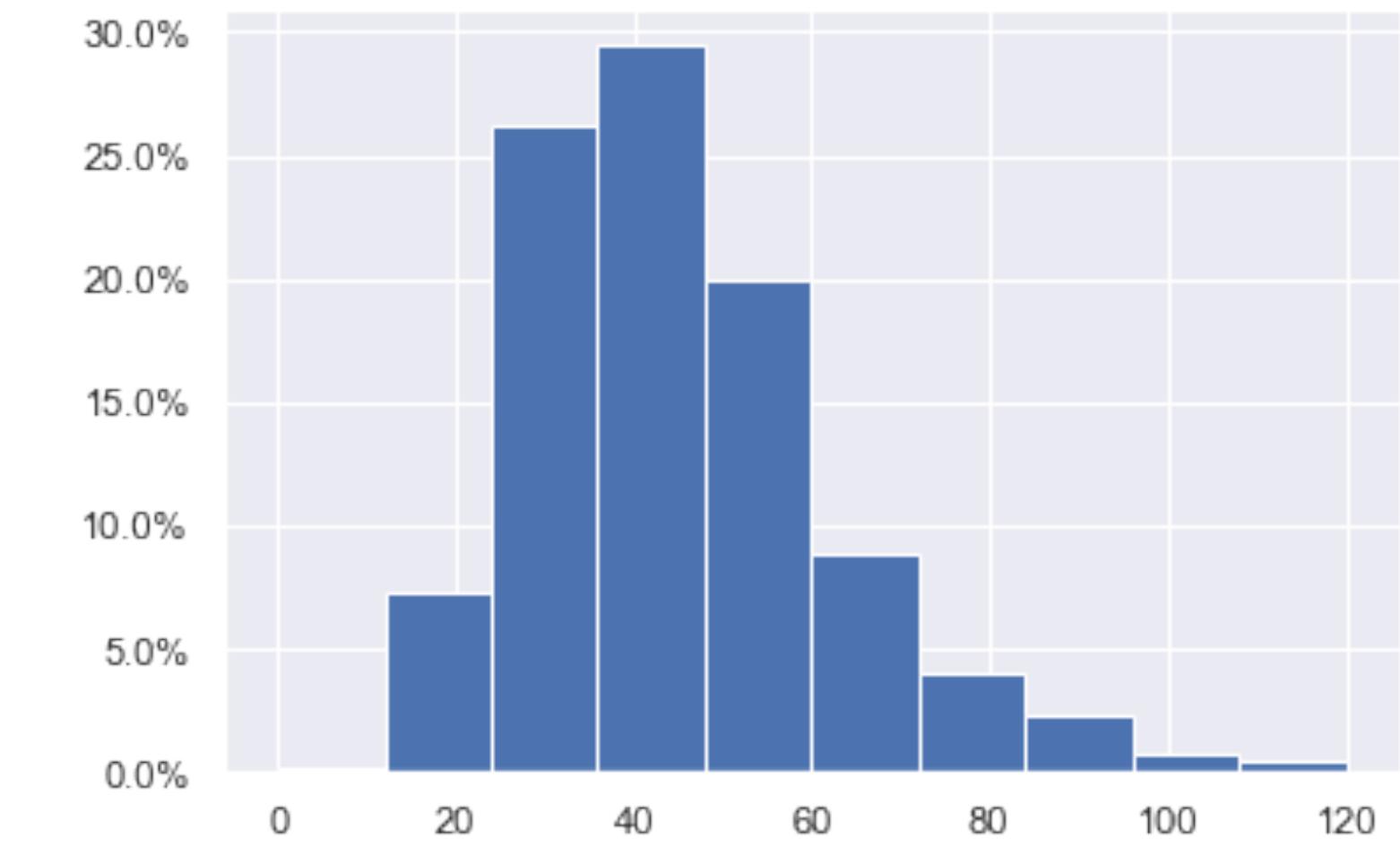
- Top loyal customer is: ID 369272, with 23 orders in a month
- Average customer order frequency: 3 times
- Average time it takes for a customer to execute an order: 8 minutes
- Average waiting time for customer since the delivery starts: 45 minutes



Barplot of times customer order

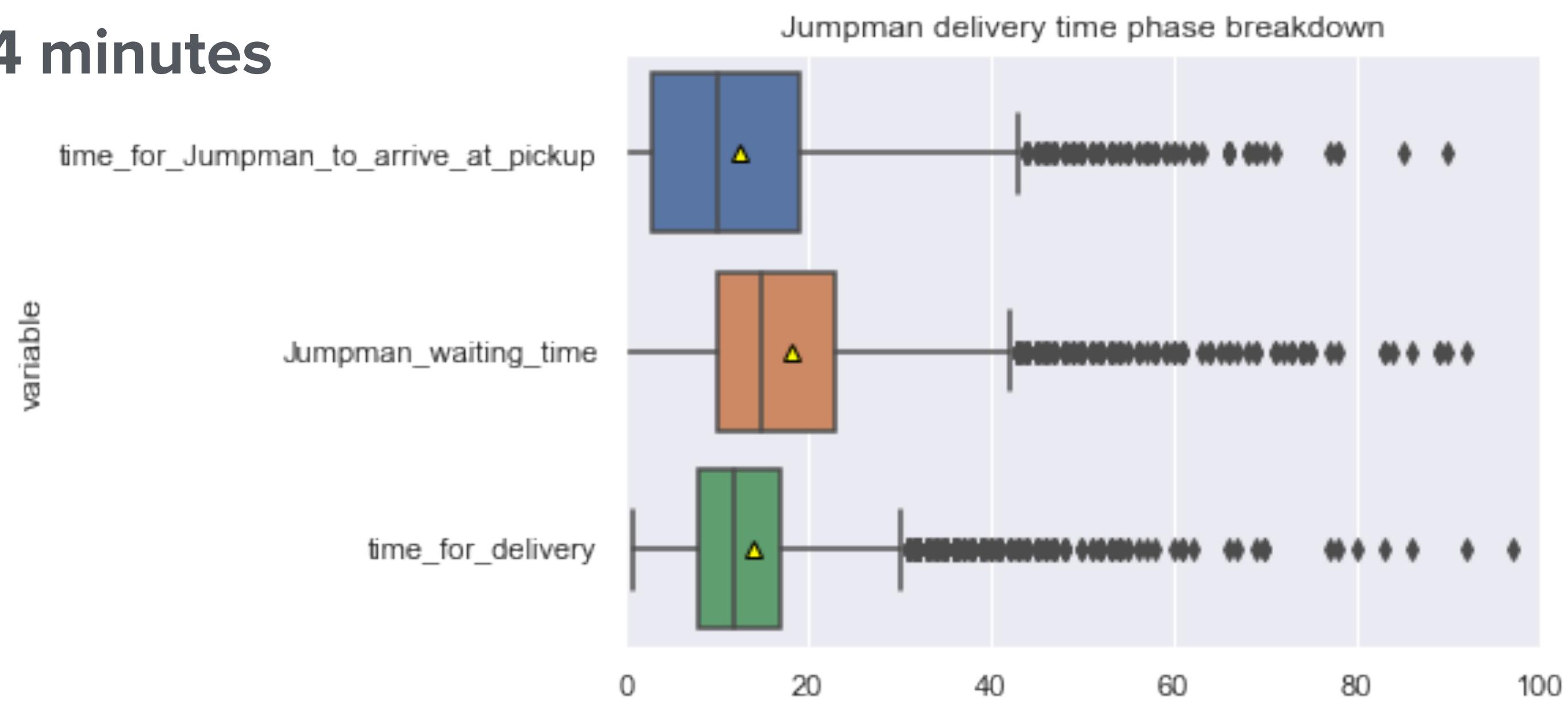
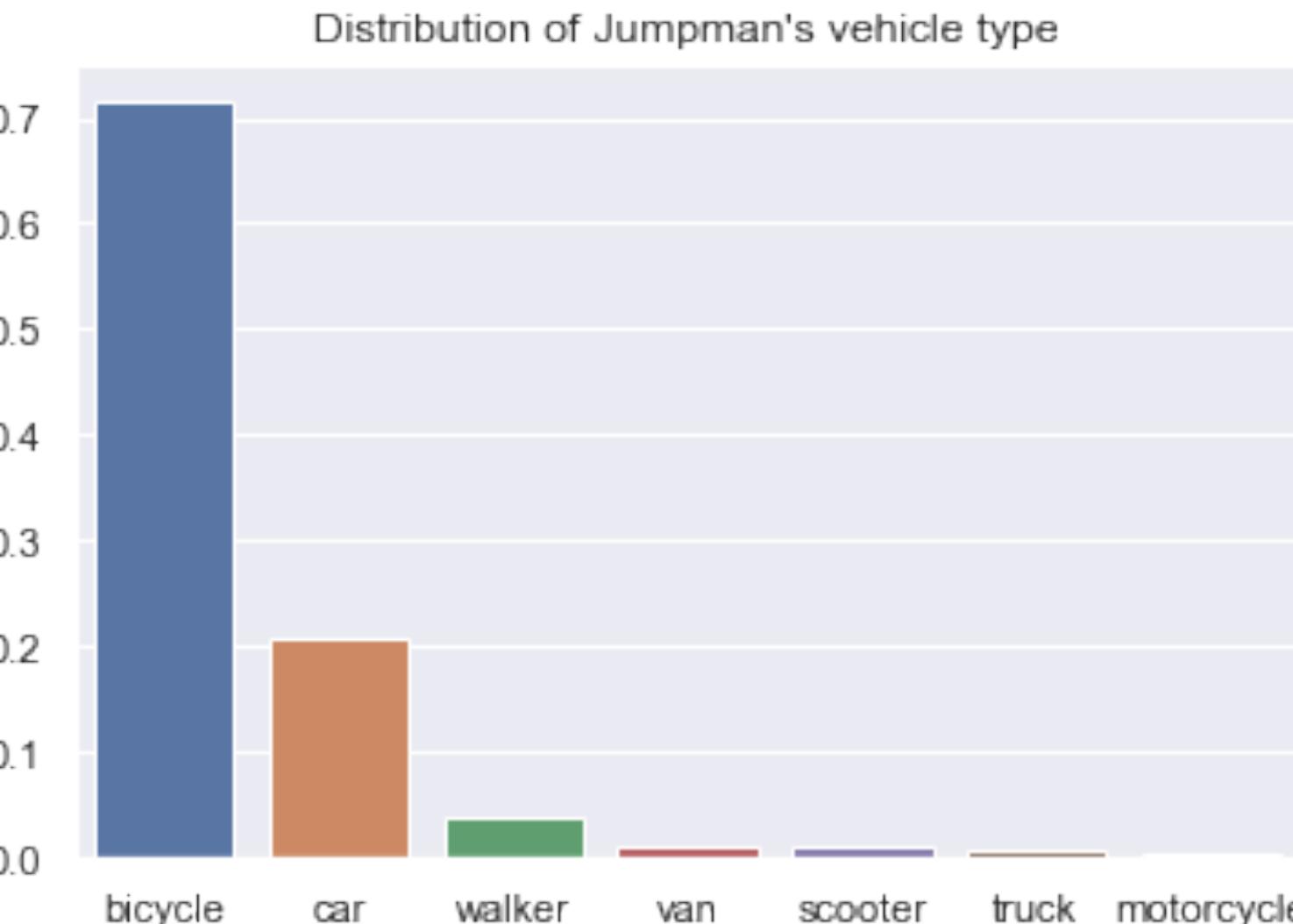


Distribution for customer waiting time since delivery start



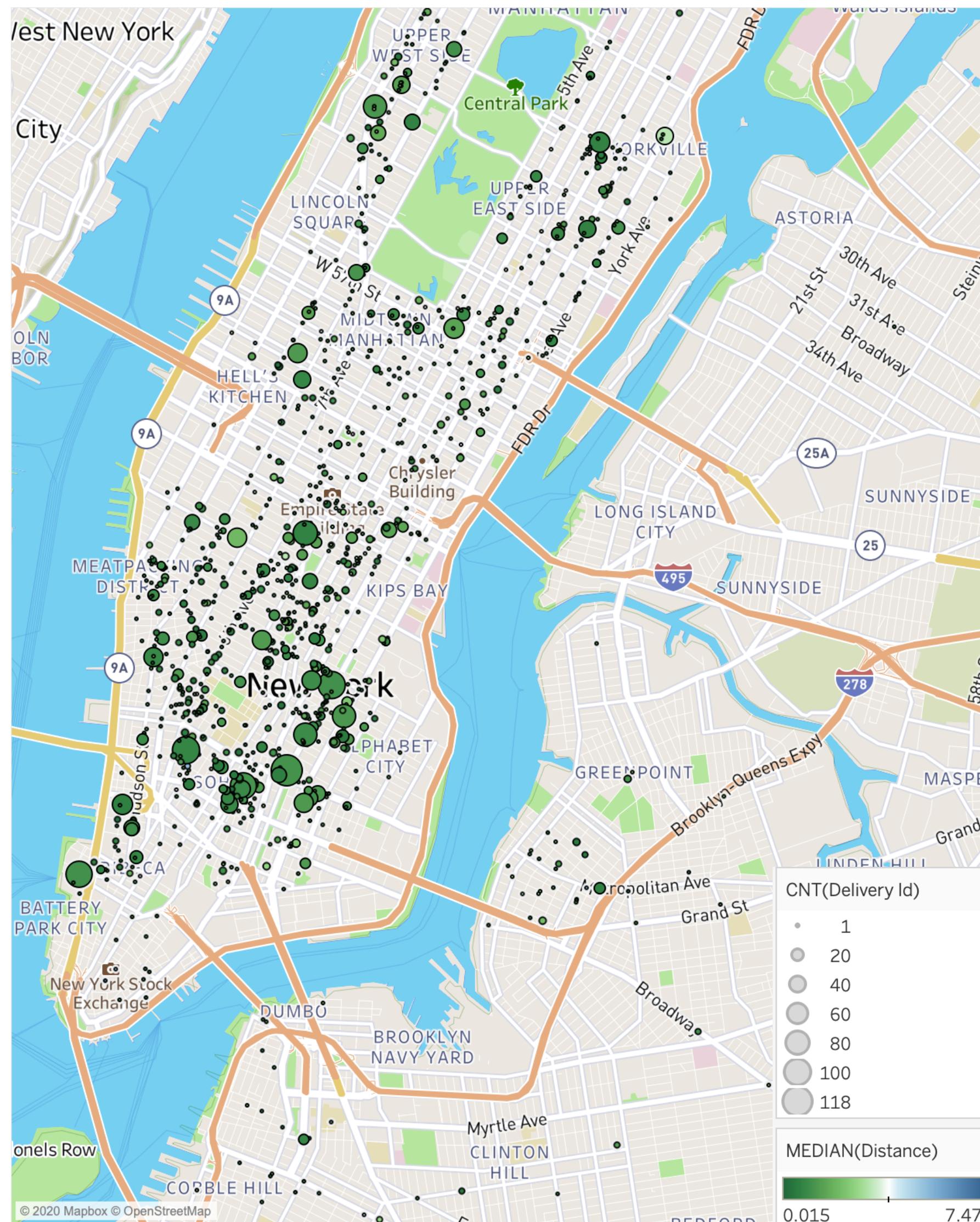
ABOUT JUMPMAN

- Top hard working Jumpman is: ID 99219 with 59 orders delivered in a month
- Average times of delivery: 8 times
- Average Jumpman delivery speed: 4.7 miles/hr
- Average waiting time for Jumpman to arrive at pickup: 13 minutes
- Average waiting time for Jumpman at pickup: 18 minutes
- Average Jumpman delivery time: 14 minutes

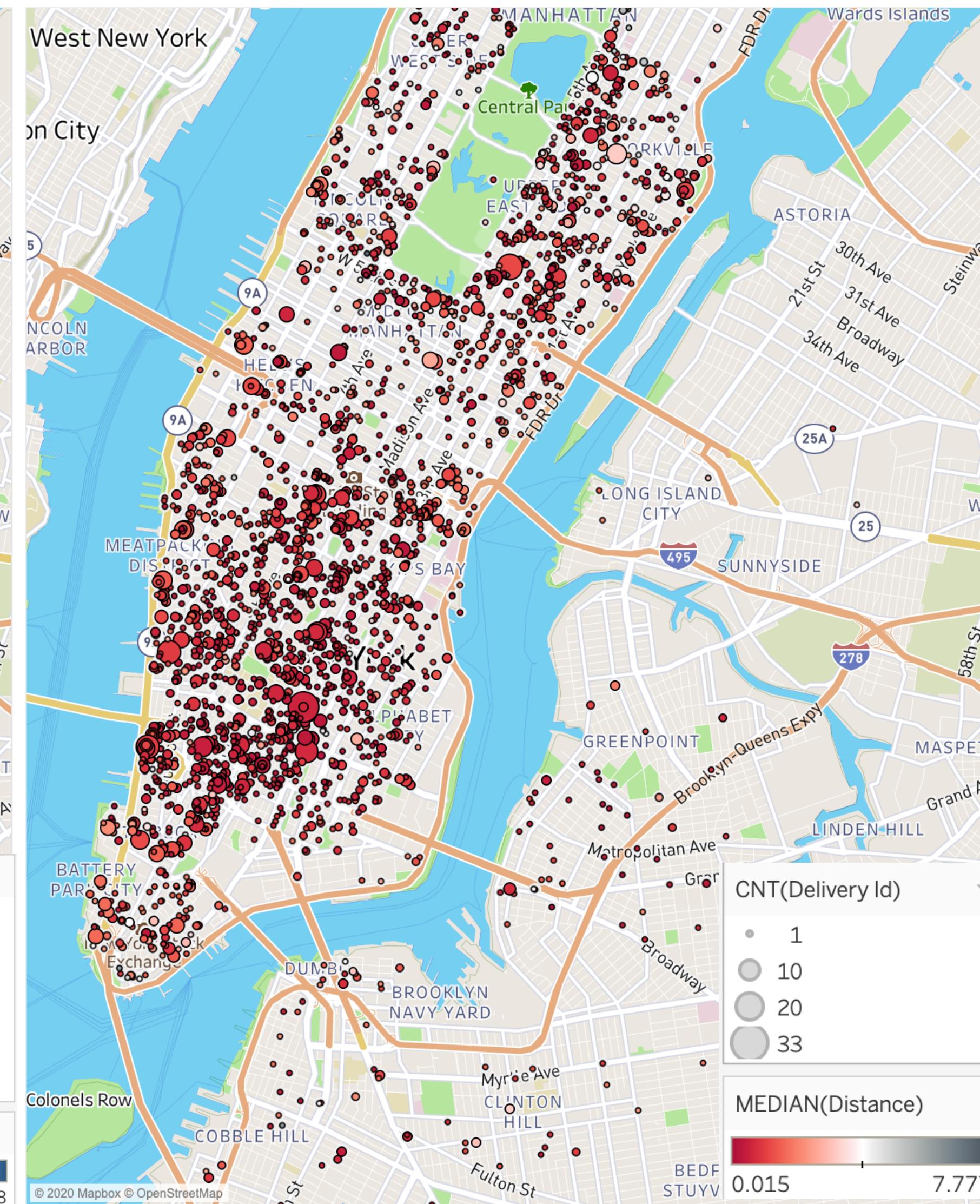


ABOUT GEO

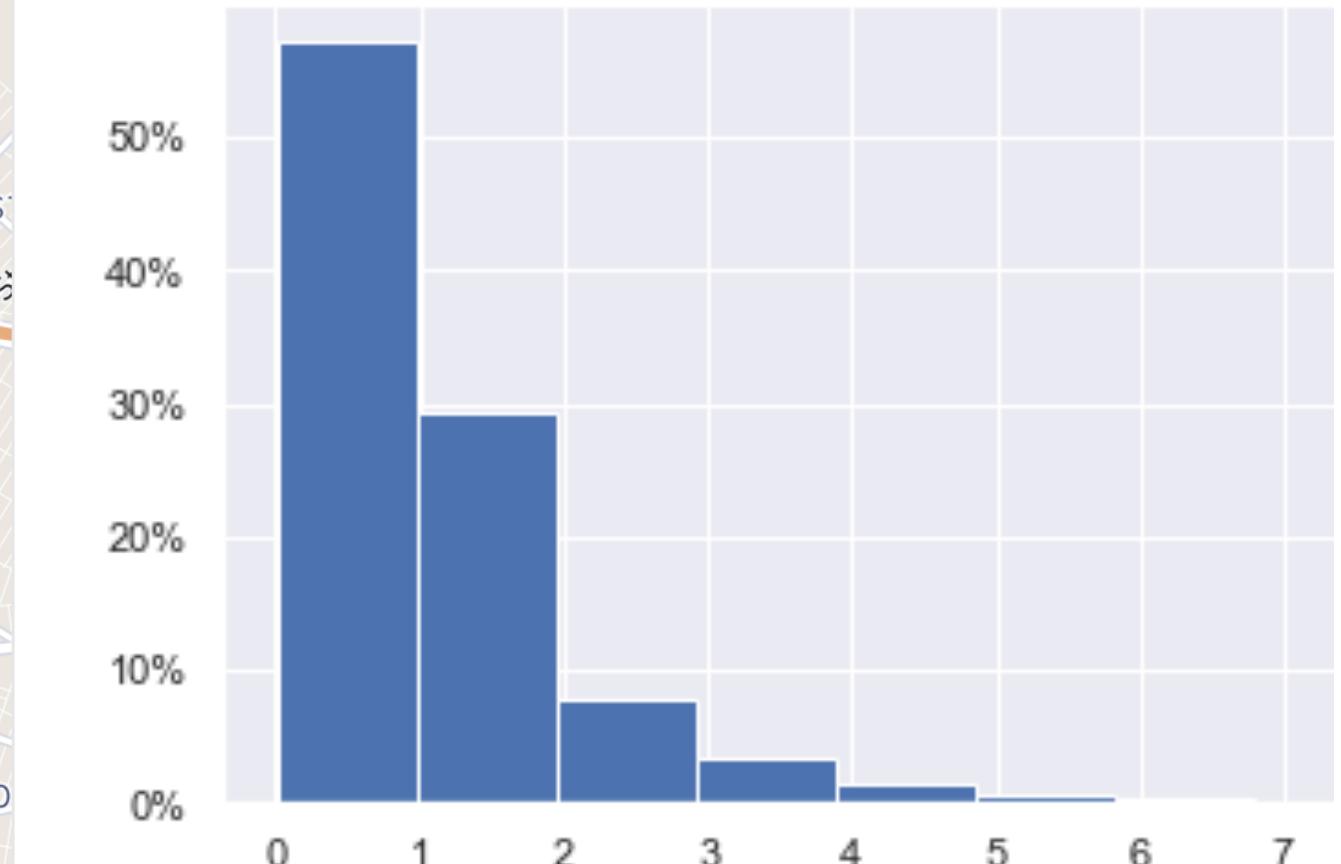
Pickup Amount & Distance



Dropoff Amount & Distance



Distribution for delivery distance



- Longest delivery distance: 7.78 miles
- Average of distance: 1.13 miles
- Shortest of distance: 0.02 miles

PLANS TO GROW THE MARKET BY 20% IN 2 MONTHS

STRATEGIES FROM ANALYSIS

■ Research findings

From research <https://www.usfoods.com/our-services/business-trends/2019-food-delivery-statistics.html> & <https://aytm.com/blog/food-delivery-survey/>, we can see that the no.1 things people care about for food delivery services is -- speed.

According to survey results from USfood, only 25% of customers will accept a delivery which takes longer than an hour. And on average people can only accept delivery that takes no more than 40 minutes. For Jumpmen, according to USfood survey, one of their biggest complaints is that food is not ready when they arrived at the pick_up place.

■ Full analysis

What is on slides are a short version of analysis

Full Report & code: <https://github.com/hakunamatata9981/Jumpman23/blob/master/Jumpman23%20-%20Data%20Analysis.ipynb>

STRATEGIES FROM ANALYSIS

Strategy 1 (Time, Time & Time):

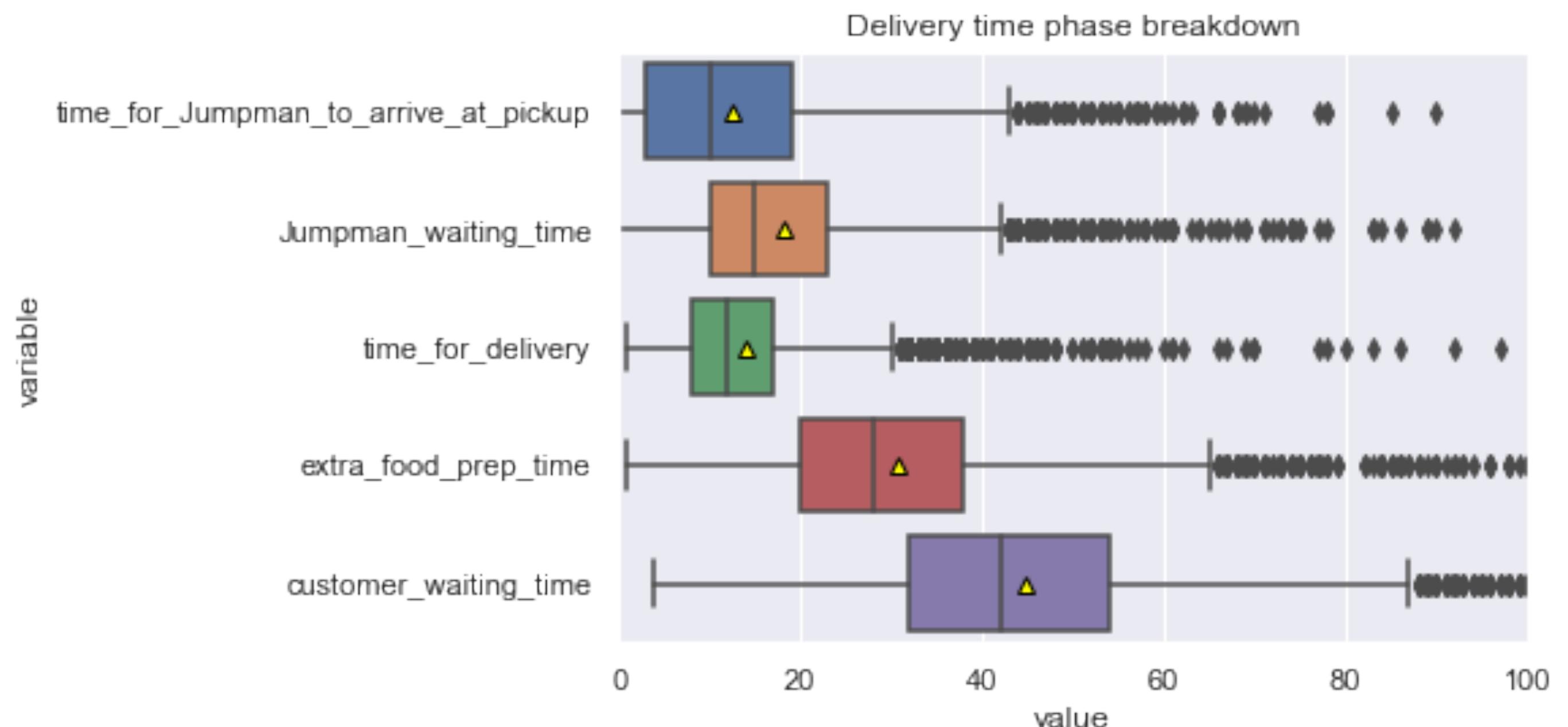
To expand the market, what we can do is shorten the time for delivery. From the boxplot below, we can see that there is, on average, an extra 30 minutes of food preparation time.

To shorten this time, we can first encourage the cooperative restaurant to prepare a menu for delivery, which takes a shorter time to prepare. Or give our customer options to pick between a speedy delivery menu or an extra delivery time menu.

Secondly, we can promote the pre-order function for customers, this will give the restaurant more time to sort through their orders.

By shortening the food preparation time, we can have fewer complaints from both Jumpmen and customers, it is the best of both worlds. There will be more Jumpman who are willing to join our network since the waiting time at pickup is significantly reduced, which can also help reduce delivery time.

As a result, more customers can be attracted by Jumpman23's speedy delivery then we can achieve the goal for market expansion.

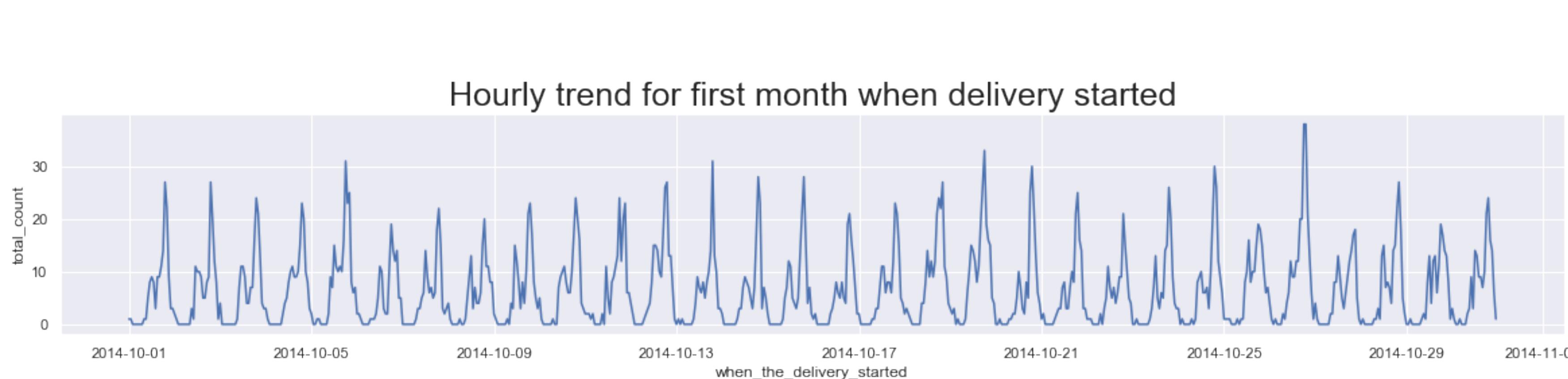


STRATEGIES FROM ANALYSIS

— Strategy 2 (Trend Analysis and market expansion):

We can see from the hourly orders count plot below that there exists a significant peak around dinner time. What we can do to increase the market share is to increase the breakfast, lunch, and midnight-food market.

For those who never order from Jumpman23 for breakfast, lunch, and midnight-food, we can offer a discount or promotion to motivate them. I speculate that the other factor for people to not execute a delivery for breakfast and lunch is the time uncertainty. To solve this, we can promote the pre-order function in the app in order to help to expand the breakfast and lunch market share.



STRATEGIES FROM ML

- Machine learning goal: Target loyal customers
- Machine learning method: Turn data into label encoded data and one hot encoded method. Apply logistic regression, random forest classification, gradient boosting classifier, catboost classifier & XGBoost classifier(half) to both sets of data.

Full Report & Code: <https://github.com/hakunamatata9981/Jumpman23/blob/master/Jumpman23%20-%20Machine%20Learning%20%26%20Predictions.ipynb>

- Winner & results: Gradient Boost + LE + GRID,

Training\Testing accuracy: 0.56, 0.60

==== Confusion Matrix on test====

```
[[ 63 196]
 [ 37 285]]
```

==== Mean AUC Score on CV====

Mean AUC Score: 0.5370141751950885

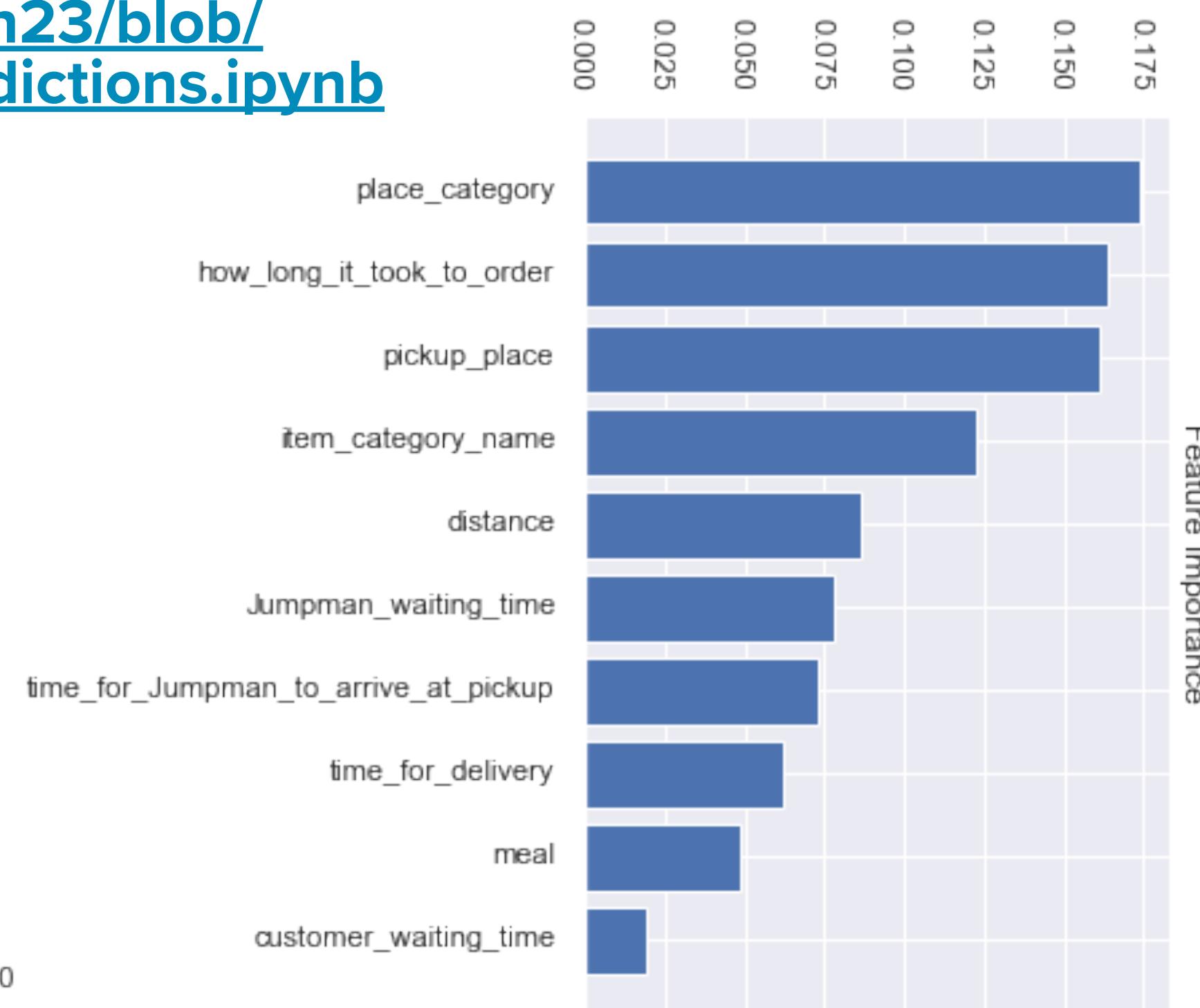
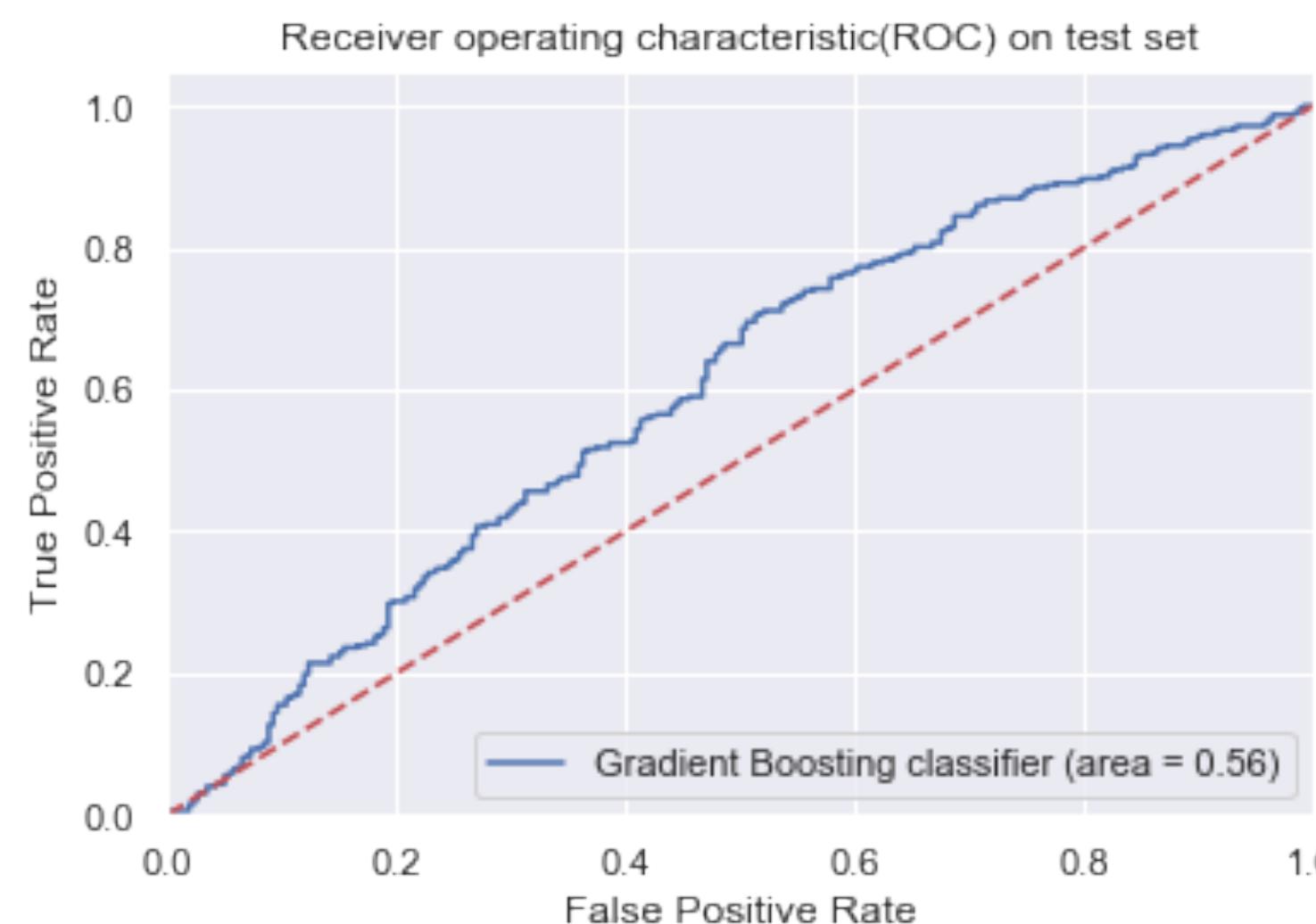
==== Mean f1 Score on CV====

Mean f1 Score: 0.6855547898176784

==== Training and tesing accuracy====

Training accuracy: 0.5601519049541361

Testing accuracy: 0.5989672977624785

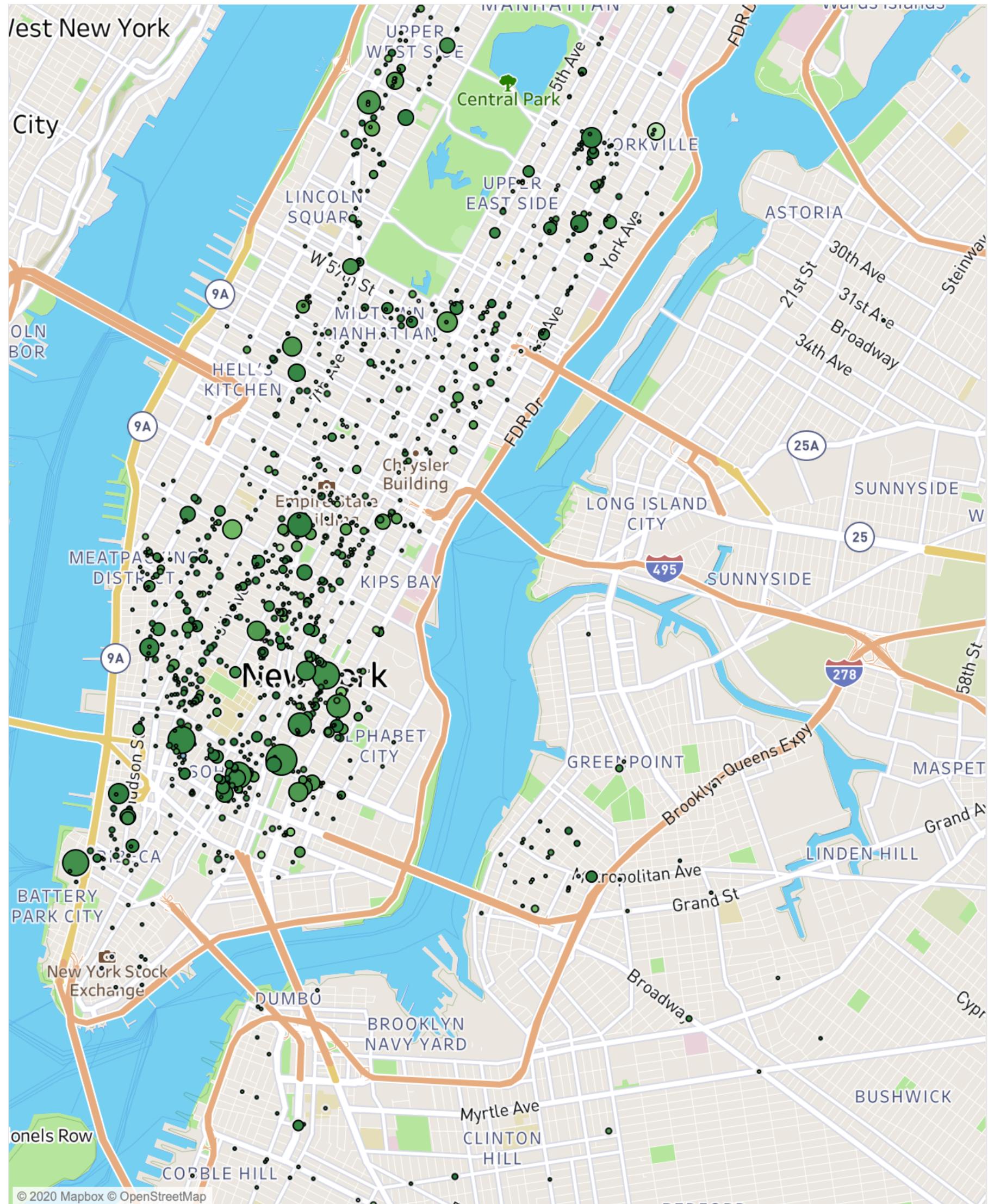


STRATEGIES FROM ML

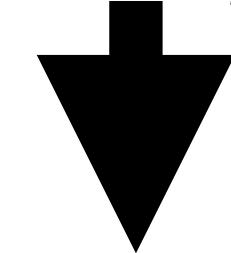
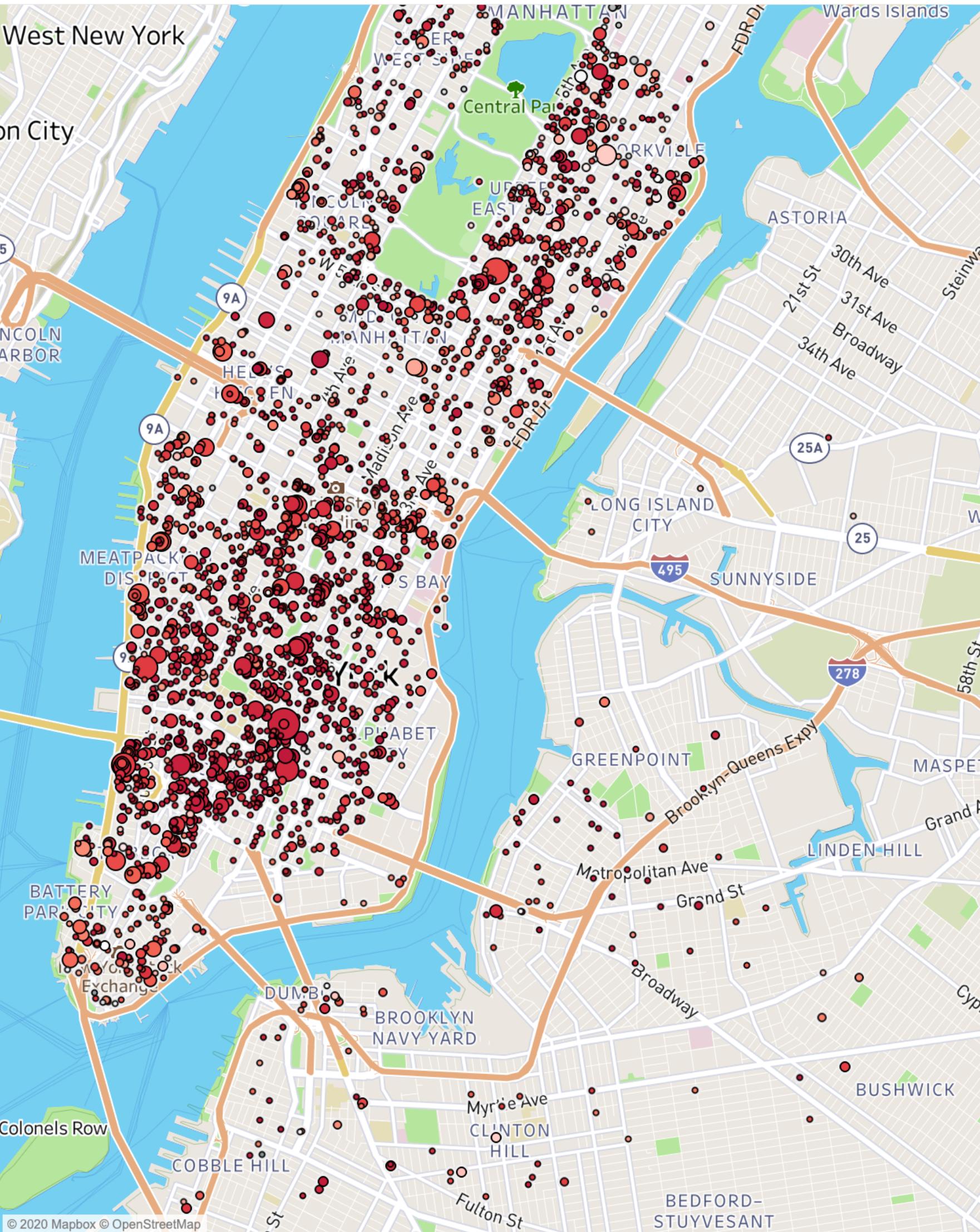
ML Feature importance

From ML Feature importance plot, we can see that distance and food categories are important features to detect differentiate if the customer is loyal or not

Pickup Amount & Distance



Dropoff Amount & Distance



Distance & Market:

We can see from the map that on two sides of the central park, the supply is comparably smaller than demand.

Having more cooperative merchandise on the upper east and west side can decrease the distance. Jumpmen have to travel which shortens the delivery time and make it more economically efficient.

By cooperating with more restaurants in these two areas, it will shorten distance and help expand the market.

CONCLUSION & FUTURE STUDIES

- The machine learning accuracy score can be improved, but more data are required (Ex: age, gender, money spent, tips spent), highly recommended to record more customer profile data if possible, this can also be helpful for target marketing.
- Can create more variables such as day of the week, mark weekend, etc, to see if it help improve the machine learning accuracy score.

THE END