Exploratory Analysis

1. Provide a set of recommendations on how to improve our business or product based on the attached dataset. This is intended to be fairly open-ended - there's no right or wrong answer! We're more concerned with your approach and the insights you uncover.

Total delivery time

In the first step, i analyzed the total delivery time from the time the order was placed to the time the order was delivered to the customer. It's the most important metric, since on-demand food service imply timely delivery. The customer satisfaction also depends a lot on the time of delivery as people order the food towards the meal time.

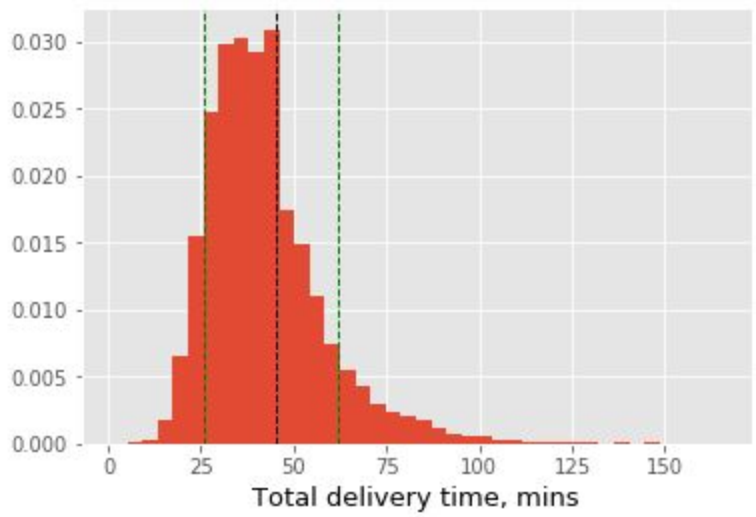


Fig. 1 shows total delivery time histogram along with lines of the 10th, 50th and 90th percentile, which corresponds to 25, 39 and 62 minutes accordingly.

count	17996.000000
mean	45.168426
std	321.377510
min	1.000000
25%	31.000000
50%	39.000000
75%	50.000000
max	42990.000000

Around 10% of total delivery time is more than 62 minutes long, the Company needs to work on reducing the time of delivery for the 10% of the orders.

What segment is slowing down the delivery process

Let's look only at the 10% of all orders, the ones that take more than 62 mins (Fig. 2). I looked at all segments that comprise their total delivery time: from placing the order, to driver picking up the food, and to bringing the food to the customer. Based on the data on Fig. 2 it takes long time to find the driver and the prepare food. One recommendation would be to incentivize people to order ahead or provide the restaurant with the estimate of the orders for the specific day & time. It would be interesting to split the data (Time between "Placed order with restaurant datetime" and "Driver at restaurant datetime") into 2 components: how long the restaurant was preparing the order versus how long did it take the Company to find a driver. I assume that the most time

consuming and the most under-Company-control factor is finding the drivers. Thus, dispatching more drivers or incentivizing them to be available could be helpful.

Placing the order with restaurant should take less than 5 mins. Delivery time should be reduced to be less than 1 hour.

	Time_Cust_to_Rest	Time_Rest_to_Dr	Time_Dr_to_Cust
mean	157.394563	51.345964	33.534596
std	603.372342	78.231065	17.506142
min	1.000000	3.000000	5.000000
25%	3.000000	30.000000	20.000000
50%	12.000000	46.000000	29.000000
75%	41.500000	58.000000	44.000000
max	5869.000000	1786.000000	168.000000

Figure 2. Statistics of the segments of the delivery time of the orders that take >62 mins

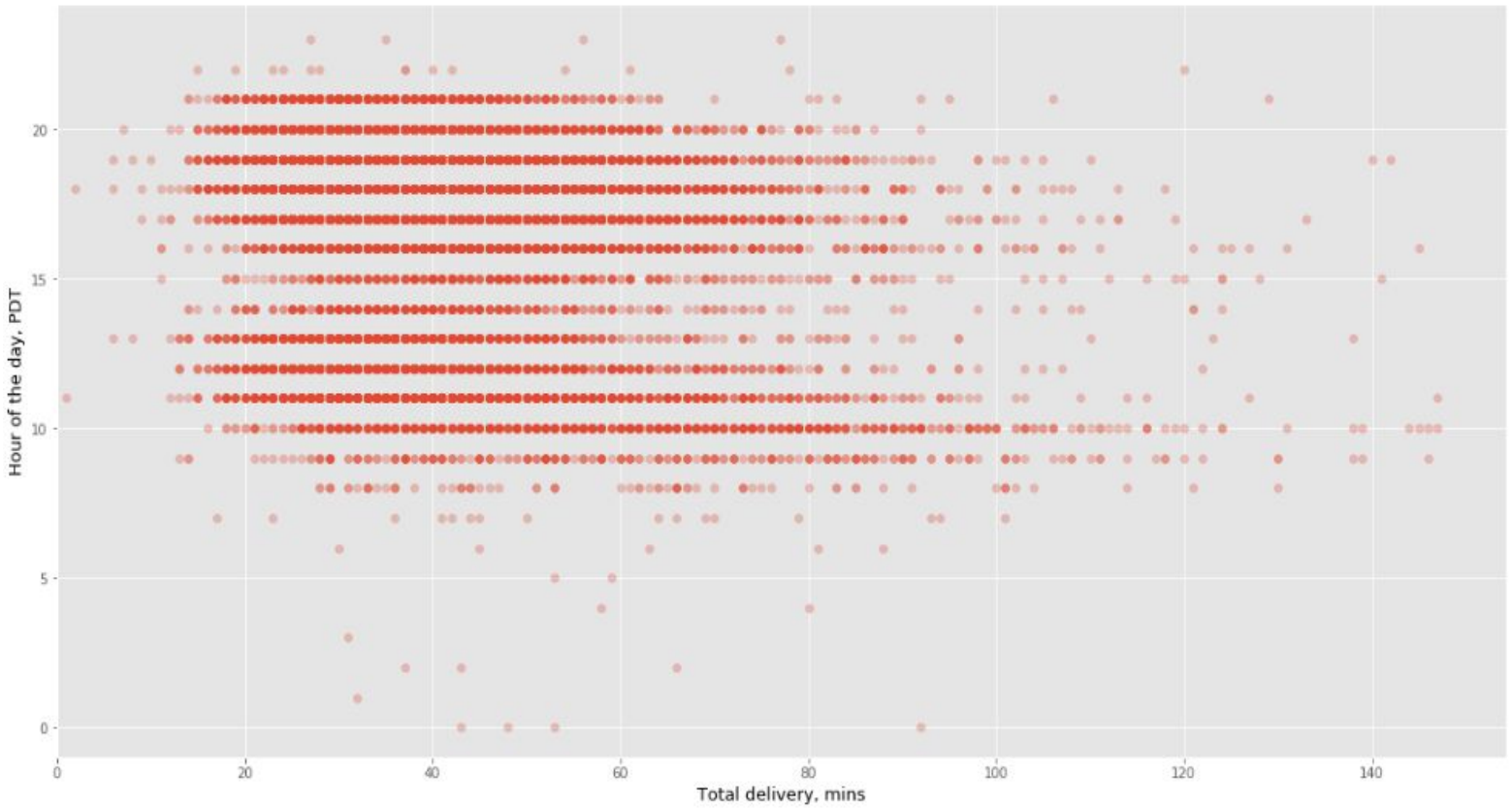
How does the time of delivery depend on the hour of the day?

Based on Fig.3 the jump in order numbers is around lunch and dinner times. It looks like during morning time, between 8am and 11am, around 50% of the deliveries take longer than 60 mins. 10% of the customers ordering at 9 am wait more than 100 mins. The Company needs to incentivize drivers to be available during morning hours by, for example, offering a better pay at that time. Many people won't consider the Company as a breakfast options if the time wait is so long. Today the Company don't have many orders during breakfast, but we could receive more breakfast orders in the future if we make the service more reliable, so that the customers can trust us.

Next step would be to find out if these early orders are on weekend or weekdays to find out the reason for longer time. If it's week day, these orders probably represent catering to companies, the orders are probably bigger and thus takes longer to prepare. If it's weekends, probably many drivers/restaurants are busy.

New customers

3512 out of 18078 of customers are new customers, or around 20%. It's quite a big number and probably is good for a new player in the market who is aiming to grab the large market share. However, the Company has been around for a few years, so the work needs to be done to retain the existing customers and incentivize them to become loyal and frequent users. One example could be a loyalty program which would allow the users to earn rewards (e.g. free delivery, free meal) for placing frequent orders.



	Total_time			
	mean	count	p90	fraction_above1hr
Hour_ordered				
0	59.000000	4	80.3	0.250000
1	32.000000	1	32.0	0.000000
2	48.666667	3	61.4	0.333333
3	31.000000	1	31.0	0.000000
4	69.000000	2	77.8	0.500000
5	56.000000	2	58.4	0.000000
6	61.400000	5	85.2	0.600000
7	58.375000	16	93.5	0.500000
8	60.863014	73	100.8	0.520548
9	67.353414	249	101.0	0.497992
10	60.436275	816	90.5	0.420343
11	44.144876	1415	66.0	0.136396
12	74.797162	1198	58.0	0.080968
13	37.586441	885	53.0	0.054237
14	49.495413	327	67.4	0.134557
15	50.784916	358	76.3	0.220670
16	48.189455	1119	68.0	0.159964
17	42.488364	2879	60.0	0.094825
18	41.747774	3370	58.0	0.080415
19	39.807882	2842	57.0	0.067910
20	37.688068	1760	54.0	0.052273
21	35.737730	652	49.0	0.024540
22	43.214286	14	72.9	0.214286
23	74.800000	5	138.2	0.400000

Figure 3: Total delivery time & hour of the day and statistics of the delivery time by hour

Order totals

There a $2 difference between average order total with and without discounts/refunds. It's quite a lot and thus the Company needs to work on reducing the refunds and discounts. Optimizing the the delivery process (especially time) will reduce the refunds due to a long wait time and will lead to happier customers. Average discount can be reduced by targeting discounts as specific customers more efficiently.

```
df['Order total'].describe()
```

```
count    17996.000000
mean        50.219467
std         43.340137
min          0.000000
25%         26.660000
50%         38.610000
75%         57.550000
max        884.180000
Name: Order total, dtype: float64
```

```
df['order_minus_discounts'].describe()
```

```
count    17996.000000
mean        48.256690
std         43.277057
min        -85.920000
25%         25.300000
50%         37.270000
75%         55.970000
max        884.180000
```

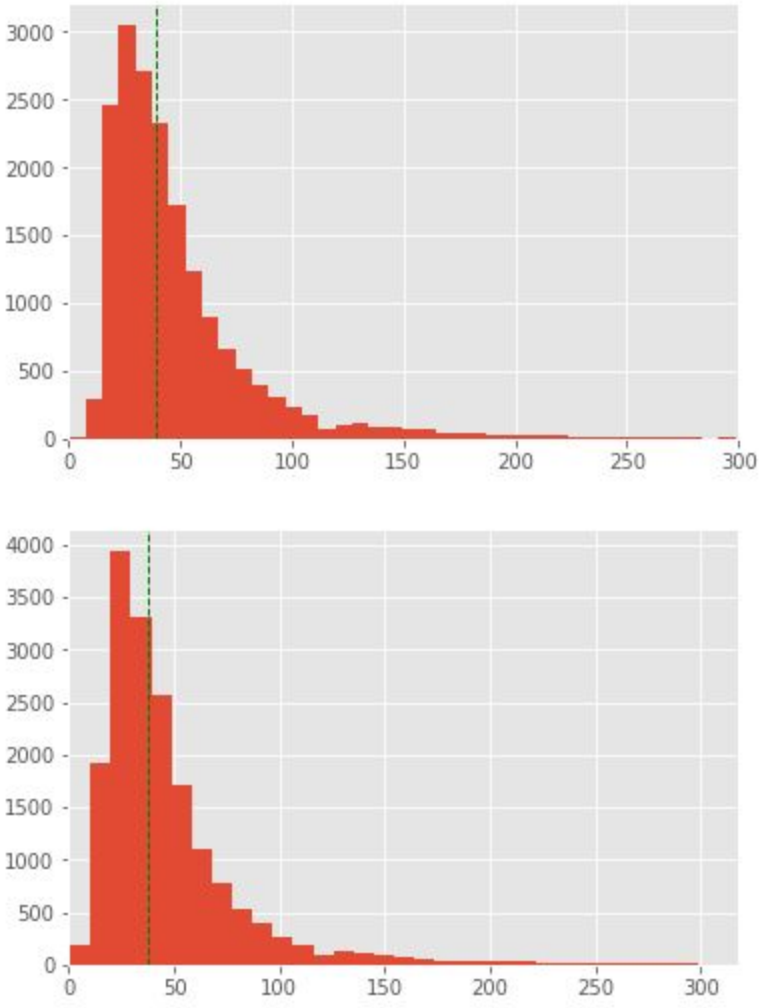
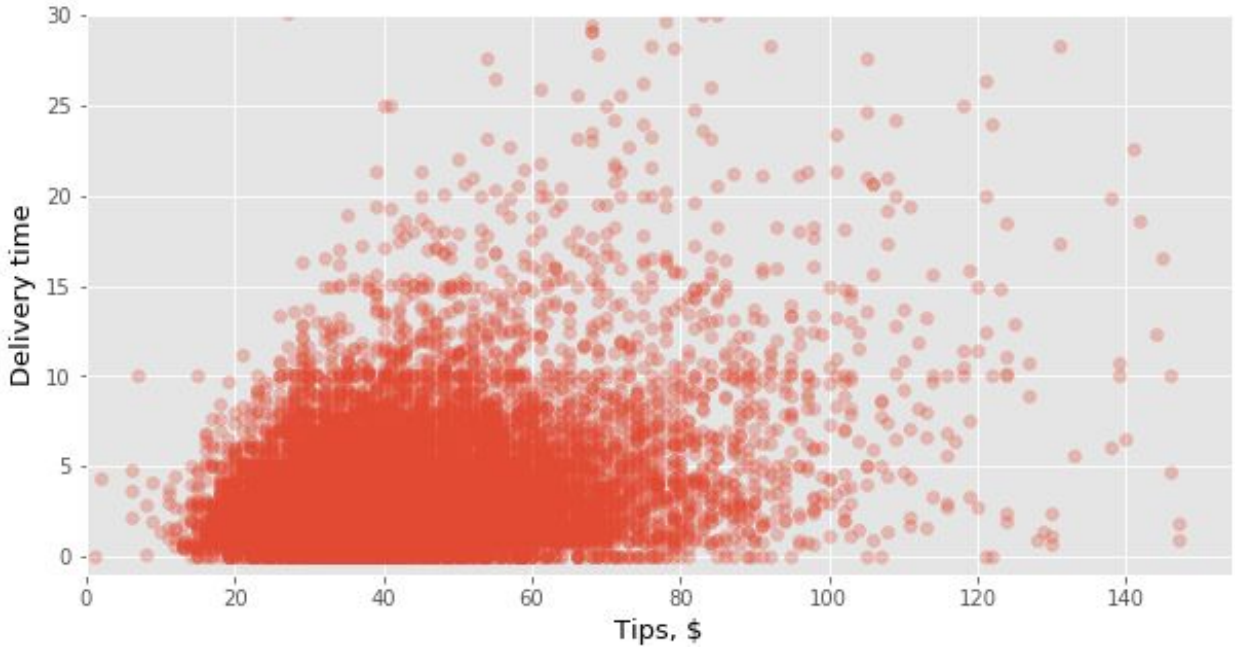


Figure 4: Histogram of the order prices (upper), order price minus discounts and refund (lower)

Tips and delivery time

On average people leave $3.5 as tips (Fig. 5). Again, optimizing the delivery process will make customers happy and will lead to bigger tips. Easiness of leaving tips such as in-app tip option or showing the recommended tip amount will help too.




```
df['Amount of tip'].describe()
```

```
count    17996.000000
mean         3.474255
std          3.613268
min          0.000000
25%          1.400000
50%          2.550000
75%          4.450000
max        120.300000
Name: Amount of tip, dtype: float64
```

Figure 5: Scatter Plot of the tips versus delivery time and statistics of the tip amounts in Dollars.

Experiment

2. Choose *one* of the recommendations/insights you uncovered (in #1) and outline an experiment you would like to run to test your suggested product/business recommendation. Please state your hypothesis, describe how you would structure your experiment, list your success metrics and describe the implementation.

I would propose running the experiment for the reduction of the morning delivery time. As said earlier, there are many orders before 11 am. However, 50% of the morning orders are delivered in more than 60 mins. Hypothesis: reduction of the morning delivery time by adding more drivers will lead to more orders in the morning in future. The way to increase the number of drivers is to offer an increased pay for delivery time in morning hours with bonuses for delivering faster. I would run a switchback experiment during which I would alternate treatment and control days. On control days the drivers will be paid regular fares. On treatment days the fares during the morning (8am to 11am) will be increased by 20% (or we can test different amounts on different days). We'll need to make sure to clearly communicate the 20% fare increase upfront to the drivers (e.g. via email and text messages ahead of time, as well as on the morning of the treatment days) to make sure that they respond properly.
Hypothesis: increased driver pay during the morning hours will reduce delivery times, as well as customer complaints and refunds due to long delivery times.

After 2-4 weeks, the 2 groups of mornings will be compared. The metrics to check:

1. Average, median and p90 delivery time.
2. Average driver payment per order
3. Average profit per order
4. Average % of orders with support tickets
5. Average refund amount per order

Implementation and risks

3. Let's assume that the experiment you ran (in #2) proved your hypothesis was true. How would you suggest implementing the change on a larger scale? What are some operational challenges you might encounter and how would you mitigate their risk?

If the experiment went well, we would need to decide whether we continue to pay the drivers the bonus to work in the morning. Also, we can experiment with different bonus amounts to determine how much exactly we need to offer to get just enough drivers on the road.

We'll need to come up with a communications strategy to notify the drivers of the bonus and explain how it works.

Before the experiment the drivers were not incentivized to work in the morning due to low number of orders. However, if the number of orders eventually increases and the drivers see it, they might shift the preference of working in the mornings instead of evenings even without a bonus. In the future we might want to remove the bonus after our market reaches the new

equilibrium in which breakfast volume is higher and drivers are motivated to drive during the breakfast time even without the bonus.

Other things to look at and take into the account: how rush hour affects morning delivery and how offering availability of the closer breakfast options would decrease the delivery time. Rush hour might be a limiting factor in reducing the morning delivery time. The option of adding closer to the customer restaurant options might just slightly change the situation because the driver still has to drive through traffic.