STA326-Assignment2

12111408 GuoYingpu

April 2024

1 Q1

group by the **order status** and **cancel type**, we have following results.

Figure 1: reasons for failure

The most cases is the order is cancelled by client before a driver is assigned, totally 4496 cases. The least case is the order is cancelled by the system after a driver is cancelled, only 3 cases.

For those cancelled orders, I have following findings.

- If a driver is assigned to an order, then almost every order is cancelled by the client(2011 cases). Nearly none is cancelled by the system(3 cases)
- If a driver is not assigned to an order, these orders may either cancelled by the client(4496 cases) or by the system(3406 cases).

2 Q2

For this task, I draw two figures. First is a barplot grouped by cancellation type for each hour. Second is a bar plot grouped by hours shows different color means different cancellation type.

First: each line is the cancellation proportion of different reasons.

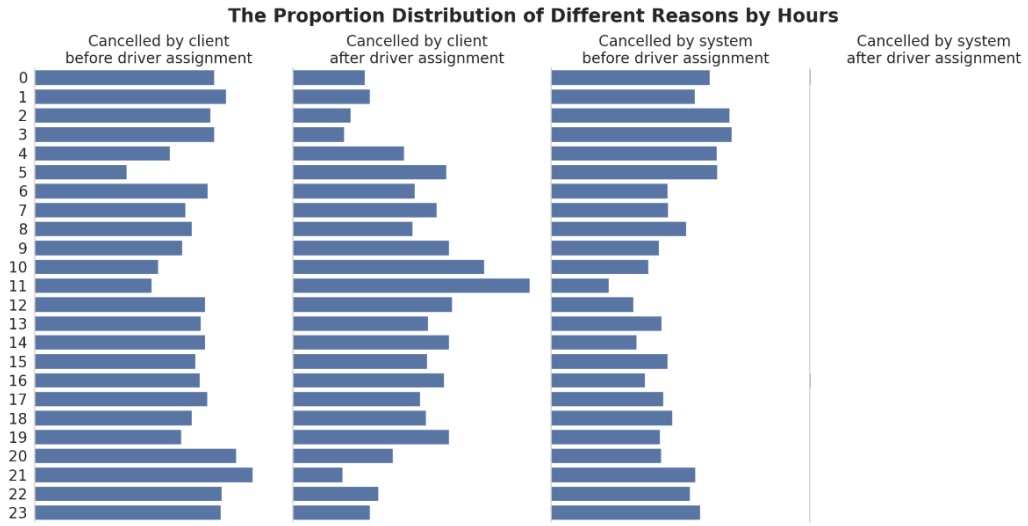

Figure 2: order cancellation by hours

Second: each bar is the cancellation orders of different reasons(in different colors) in each hour.

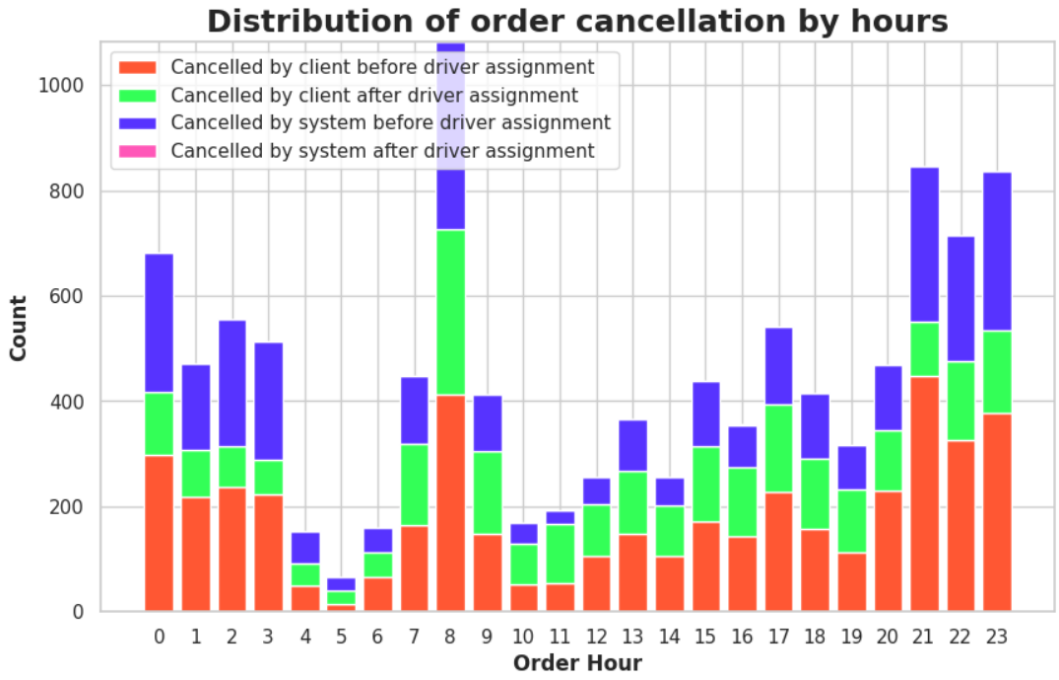

Figure 3: order cancellation by hours

From the above two figures, we conclude that

- hour 8, 21 has the most cancellation orders of type **cancelled by client before driver assignment**. The reason might be that during those times, the drivers are busy so client might cancel the orders.
- hour 8 has the most cancellation orders of type **cancelled by client after driver assignment**. The reason might be that during rush hour(8 a.m), a client is busy and the driver might take a long time before pick up the client, so client cancel these orders.
- during night(9 p.m to 3 a.m) and early rush hour(8 a.m), **cancelled by system before driver assignment** type tend to be very much. That's because during night time, drivers are less than usual, and in early rush hour, drivers are busy too.

3 Q3

After removing the outliers(data points that fall below Q1 - 1.5 * IQR or above Q3 + 1.5 * IQR.), we gain the data and plot the following line chart.

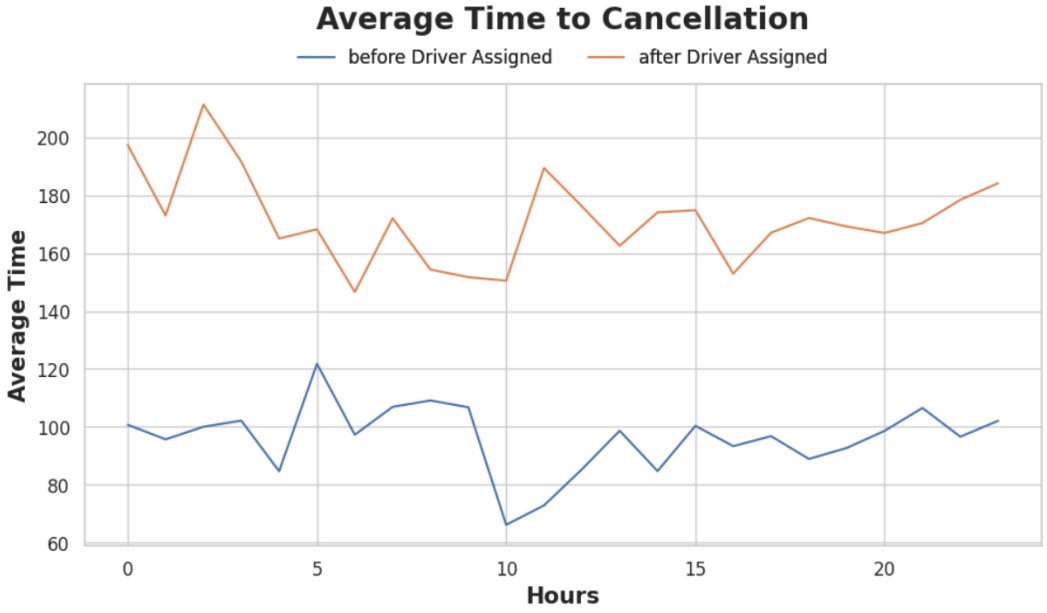

Figure 4: average time to cancellation by the hour

- In general, the cancellation time for an order after driver is assigned is more than an order before the driver is assigned. That's because when a driver is assigned, the client tend to wait for more time.
- During night time(20 p.m-3 a.m), the average time for cancellation time is longer than usual after driver is assigned. That's because the time is late, the clients are reluctant to cancel so they wait for a long time.
- During early rush hours, the average time for cancellation time before driver assignment is more than usual. That's because people have no idea for other public transportation but tend to wait for a taxi.

4 Q4

ETA: time before order arrival.

For this task, I draw a bar plot and shows the top 3 hours which have the longest time in dark blue.

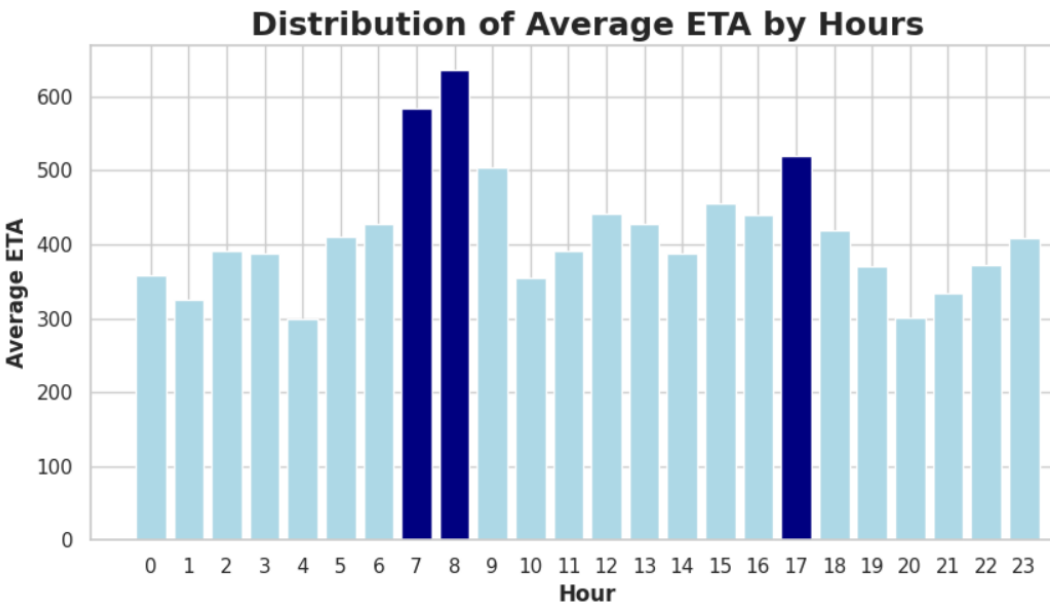

Figure 5: average ETA by the hour

We can see that during morning and afternoon rush hours(7 8a.m, 5 p.m), the average ETA is the longest. The traffic is busy during these hours, so it takes more time to pick up the client.

5 Q5

Totally 24 sizes contain 80% of all orders in London.

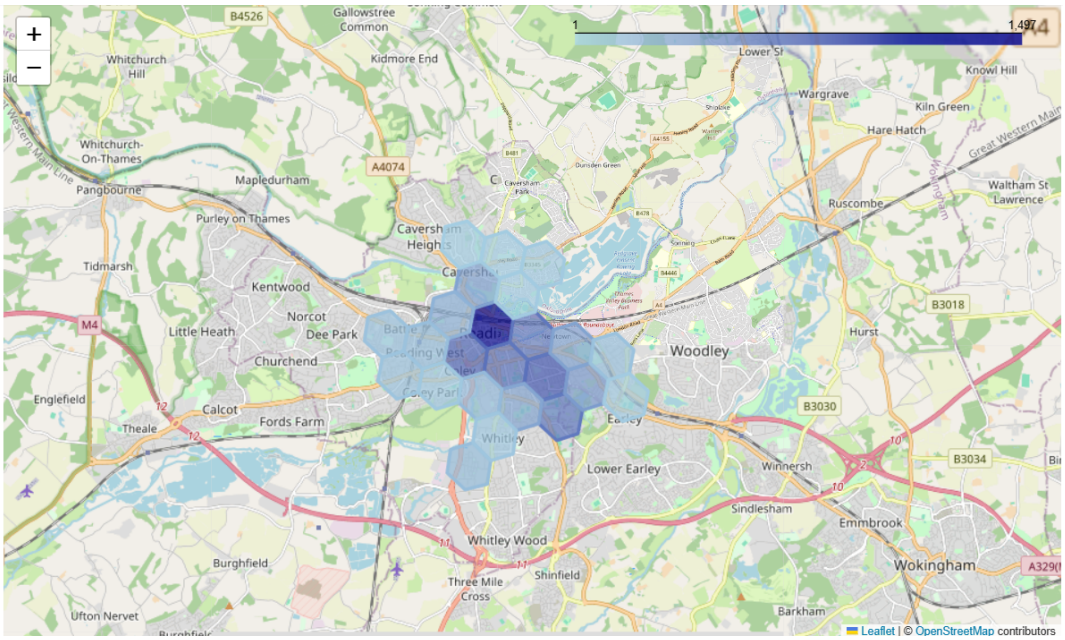

Figure 6: hexagon