Delivery Duration Analysis Research

Zuzanna Jarlaczyńska

April 7, 2024

1 Assumptions and suggestions

On the created correlation plot we can't see any important correlations. However, we need to bear in mind that the dependency may not be linear and simple correlation won't show it. Instead of correlation plot, we now will use grouping to see how each feature impacts delivery time.

1.1 Product Id

From preformed data analysis we know that delivering some products takes much more time. The reason for that may be for example the size of the product - obviously delivering staff like furniture will be time consuming. Therefore, type of product delivered should be taken into account while making predictions.

1.2 Start/End hour

The deliveries that were started between 8 and 10 or ended around midnight or midday tend to be significantly longer than other ones. Moreover, also errors in predicted delivery times are the biggest in this hours. That leads to conclusion on this time of the day the deliveries are not only long, but also definitely longer than planned. The reason for that may be for example heavy traffic jam. This dependency should definitely be considered while planning delivery times. We can for example add a condition that if an order is supposed to start or end in mentioned hours, its duration should be calculated only basing on other orders started/ended in this interval.

1.3 Segment and Driver

Another thing worth mentioning is the fact that some routes may take much more time than the others. The reason for that might be the length of the route, traffic jams and type of the road. Also drivers have impact on delivery time.

After extracting long delivery time segments ids (over 30 minutes) and segments ids on which the error is usually bigger than 30 minutes, we got a useful set of segments. Therefore, we can easily identify those more time consuming segments and take that into account while predictions. That approach will significantly improve our delivery times accuracy.

```
Index([ 400,  656,  949, 1055, 1169, 1772, 1778, 1907, 1919, 1964, 2192, 2248,
        2370, 2686, 2775, 2786, 3114, 3207, 3348, 3883, 4139, 4191, 4797, 4936],
       dtype='int64', name='segment_id')
```

From the obtained data we can infer that not only routes but also drivers affect time of the delivery. Driver 4 seems to need much more time for his deliveries. Predictions for this driver are also frequently far from truth. Taking all of this into account, the the special time buffer should be provided for both mentioned segments and driver 4. Ideally, the predictions of delivery time should be done for each driver and segment individually.

1.4 Delivery times distribution

From predictions errors distribution we can conclude that for most of the predictions the predicted delivery time was a little bit longer that the actual time, which is an optimal solution. However, for about half of the predictions the expected delivery time was shorter that the actual time. That may lead to dangerous situation when the minor delays overlap and create a huge delay. Therefore, it would be a good idea to slightly extend each predicted delivery time.

Moreover, we can observe that the biggest errors happen for the longest deliveries. The reason for that is probably the fact, that the vast majority of deliveries is very short. As each prediction is given as a mean of previous delivery times, it can't be suitable for routes that are simply much longer and therefore more time consuming. Again, specifying those long segments ids is crucial for accurate predictions.

Finally, we can see that the delivery time is either between 0 and 10 minutes or over 240 minutes - we can't see any values between. It may be a good idea to associate special flag LONG with segments on which delivery took much more time in the past. Delivery time for LONG segments would be calculated only considering other LONG segments. Also an additional feature with total route length could be insert to the database.

1.5 Delivery time by sector

In the obtained plots the difference in delivery times is not very vivid - the difference is about 2 minutes. However, when we take into account the very long deliveries (over 3 hours), we can see that most of them takes place in sector 3. Also the longest delivery happened in sector 3. Drivers bad experiences with unexpectedly long routes may have lead to false impression, that in sector 3 the delivery times are significantly longer. Finally, we should rather consider the segment id than the sector itself.

2 Summary

To sum up, the most important step to improve the accuracy of delivery time prediction is considering the order route. The sector in itself doesn't seem to be crucial to predict delivery time. Further enhancements can be achieved by taking into account start and end hour of the delivery. Finally, we may introduce the factor depicting the relationship between the delivery time, specific product type and driver.