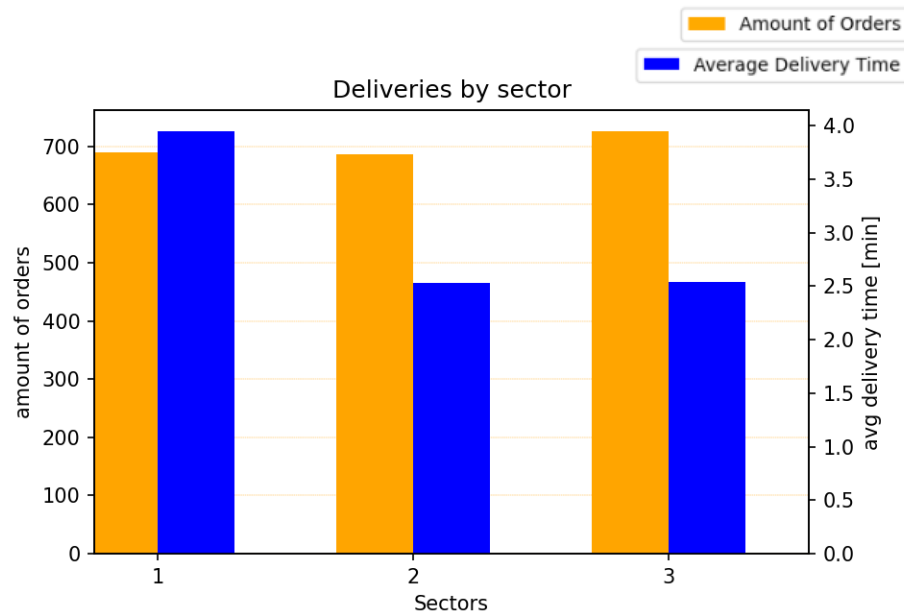


Report

Predicting delivery time per sector

Considering dividing average delivery duration on sectors I would calculate delivery duration for every segment/order (type STOP with assigned *order_id*) taking place in that sector. I would examine the data and designate data frame for analysis by disregarding outliers. In this case it comes down to dismissing any record that has time duration less than or equal to 0 seconds and those lasting for more than hour – considering them an error. This gives me 2101 records, which I would group and order by *sector_id*.

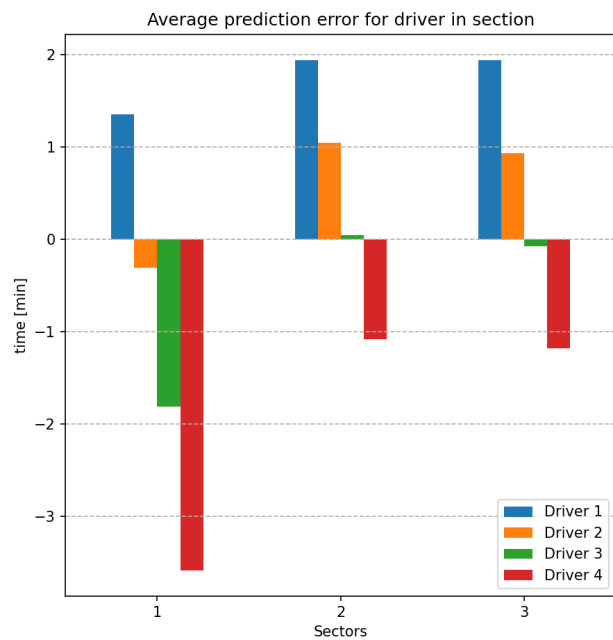
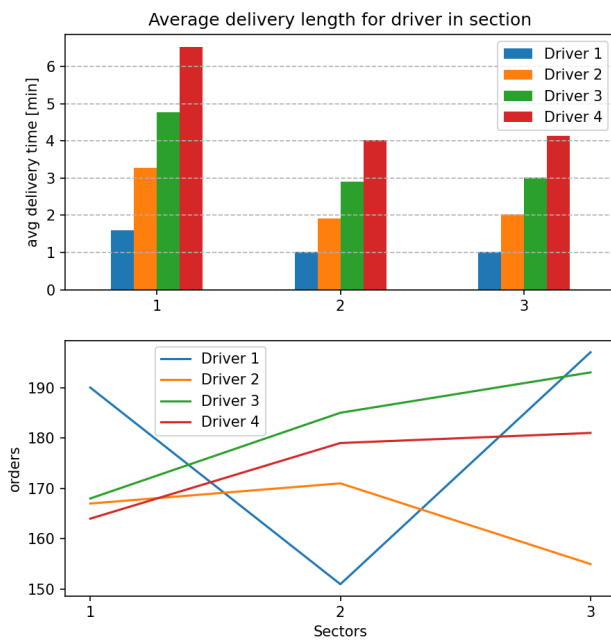


Because not in every sector an average delivery time is the same, the idea of separating calculation of *planned delivery duration* into specific sectors seems like a good **first** formula improvement.

Other solutions

I would take previous idea a little bit further and because of each sector is being handled by multiple drivers I would suggest predicting time per sector **and per driver**.

The data preparation is very similar to one described before. Only change is to group and order by sector_id and driver_id.

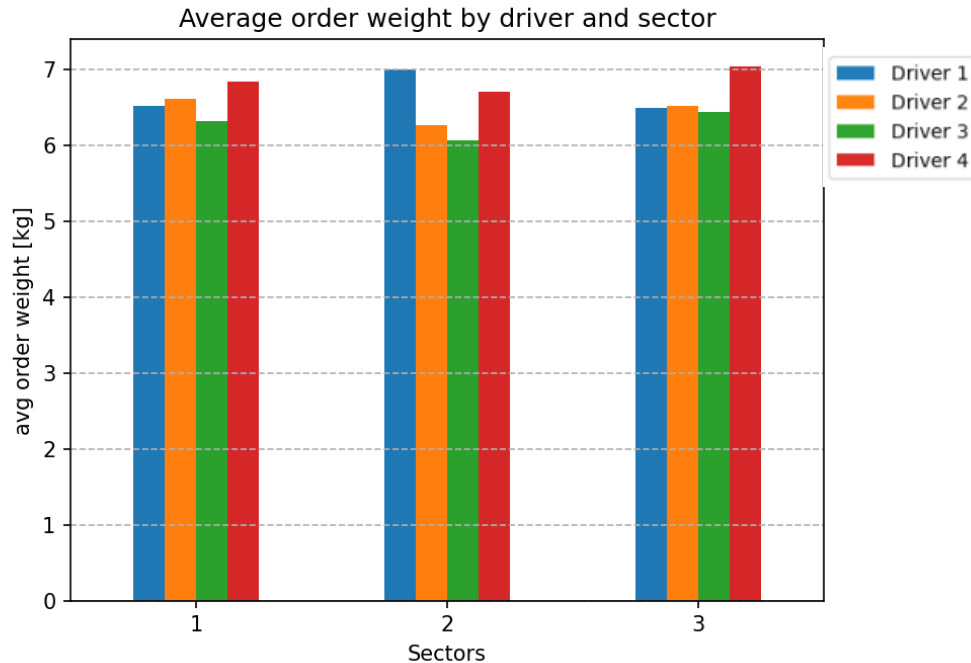


Here we see that dividing calculation only by sector is not enough, even though it will decrease the over all prediction error. Now the planned delivery duration is about 3 minutes for each order. So for sector 2 and 3 it would drop to about 2 minutes per order. But in sector 1 it would still be 3 minutes, so in this case first proposed improvement doesn't change anything. The prediction error for this sector would stay the same, which is the biggest one of all.

If only we could assign the driver to the order in advance we could reduce prediction error to close to 0.

Why some deliveries take longer?

My first thought was that maybe some orders are a lot heavier which makes it harder to deliver fast. I calculated weight of each order by summing up the weight of product multiplied by its quantity in order and grouped data by *order_id*. Then I grouped orders by *sector_id* and *driver_id* as before (chart below). Unfortunately it turned out that in average every driver delivers orders weighting between 6 and 7kg. Nothing deviating from an average.



Assuming that order weight is not a problem, there has to be another reason for longer deliveries. This might be an issue with buildings, like lack of (access to) elevator, especially in high buildings, difficulty in getting inside: broken intercom or front door, angry dog barking at the entrance, being unable to contact client when he's not home.

The problem might be with a driver. Because we can see the difference in how faster driver 1 is delivering packages than driver 4, regardless of sector. Perhaps he's not as strong and this makes him move slower with heavy packages, especially on the stairs. Maybe he is a smoker and adds time by smoking before moving on to the next delivery – so it would be a data error, because the actual delivery took less time.

What additional data could be collected

Type of building is a good information to have. Should be provided by a client while placing an order. More than type of building the important information is to which floor the order needs to be delivered. Every additional floor makes the delivery a bit longer and harder.

Does the building have a reception/porter. If so, the order might be left there and the apartment building is then equal to a house.

Delivery failure and success status. Very short time of delivery might occur when delivery attempt failed – no one was at home so driver moved on to the next order. Or it could go the other way – no one was home so driver waited for them and in result delivery took much more time than intended.

Risks of over- and under-estimating delivery time

Overestimating delivery time might cause customer dissatisfaction because they are presented with longer time than they think is needed for the order to be delivered.

Customers might start looking for the faster alternative.

The delivery schedule might fit less deliveries per day.

Underestimating delivery time can cause customer frustration that they are consistently waiting longer for the delivery than it was promised. These can lead to negative reviews, complaints, loss of clients.

It might cause undesirable rush of deliveries that drivers cannot handle.

It might result in creating additional costs, like compensating clients, paying employees overtime, covering expedited shipping costs.

ADDITIONAL INFO

If you want to see full SQL queries used for data extracting see file: queriesVisualization.ipynb