Maria Knapczyk

Delivery Time Hypothesis

As it was seen in the first part, proposed solution to split data by sectors is a great potential way to upgrade the algorithm. I would start by calculating model metrics of the current model on a testing set such as: RMSE, MAE, R-squared. After calculation on the first model, I would group the data and for each sector make a prediction the same way as before. Then I would calculate the same metrics and compare them with the previous version. If the error rates are lower, then it means that the new solution is a more efficient way. The expected outcome is reduced prediction error rate especially in the first sector.

As a more advanced approach, I could propose using machine learning models, for a solid start, instead of calculating a simple average, a linear regression model could be used. For more advanced algorithms, with access to route data, a genetic algorithm for finding the most optimal route between delivery addresses could be made.

Some deliveries may take more time, some potential reasons I see are:
- Modern architecture, especially big estates, have a lot of different staircases for the same building, navigating between these big flats can be hard, also some of them have security guards who need to contact flat owner for approval
- People can also put on incomplete address or provide incorrect contact information which can lead to missing parcels
- Some customers may expect the courier to wait for them to open the package, in case of a missing item or a fraud
- Delivering to points such as shops may be more complicated due to crowd or needing to wait for free staff members
- Traffic on the streets, especially during peak hours, weather conditions

To improve future analysis and make it more insightful I would definitely add detailed address info to track, if packages with 0 time are processed during another operation and verify data accuracy or to see if higher floor numbers increase delivery time significantly. Many parcels delivered to the nearby addresses may take more time to process. For example there could me multiple orders for a bigger place like żabka, where delivery guy has to leave multiple packages for different customers and with different order id. This could also help explaining why some sectors take longer to proceed. This brings me to adding more details about drivers such as experience level or average pace, combined with geo, weather or traffic data, it could give a solid understanding of underlaying problems. The last thing I would propose is customer availability eg. whether they were waiting at the door.

The risk of underestimating the delivery times delayed or even missed deliveries, which can lead to unsatisfied customers and overloaded employees. I think that most customers will be disappointment after waiting the whole day at home just for the courier not to come. Also couriers will have to do deliveries the next day which can lead to further delays or even dangerous situations.
The risk of overestimating the delivery times is decreased efficiency which generates extra costs such as more employees and equipment than required.