Exploration

Produce interesting statistics and graphs about the dataset. Show the most important features and explain what you see.

Reply: Plotting the features bar plot and scatter plot, it can find that orders are generating more at the weekend. Again, in a day, orders are generating more from morning 08:00 AM to evening 06:00 PM. One interesting fact is that orders are more in higher temperature comparatively to the lower temperature. On time delivery has not so significant impact.

Modeling

Why did you choose the approach, what kind of benefits do you see in solving it? What kind of metrics can you use to evaluate how good the solution is?

Based on the approach you choose, produce a model suitable for the task. You should include the preparation work, feature engineering and your thought process in your answer.

Reply: Co-ordinates of restaurant and client location are in Helsinki center. It has no other impact then the possible delivery time. So, it can be dropped. Again, considering the delivery is on time or not customer may have a complain. From 940 value count for on time delivery is 918. And as it is discrete for user it shouldn't impact on total delivery count.

I have chosen RandomForestRegressor and K-fold cross validation to find the best.

Evaluation

Are you happy with the results? What kind of results would you expect to see, if this was deployed to production?

Reply: Well, to be very honest not so happy. I have learned data-science by self-learning. But for the first time have analyzed a real-life data. I have learned a lot to find the best outcome. I am understanding what is going-on with the data, but lack of practice and knowledge couldn't implement/find solution from the resources according to understanding.

Further development

Make slight modifications to the model or take a completely different method to solve it. Compare your two solutions. Strengths, weaknesses? What should you consider when you compare different models? If you had more time and resources, what kind of development could be done to make the solution better?

Reply: I would like to try more using Linear regression and arranging the feature according to their importance.

Well Time series analysis ARIMA can also be experimented.