# Report

Optimizing Bolt Pricing

Objective

Improve upfront pricing precision.

Observations

- The data consisted of 4943 entries from 2nd February 2020 to 13th March 2020 out of which 2770 were from EU and the remaining 2173 were from other continents/countries where BOLT is operational. Hence, for this task I have only analyzed EU data as for the other chunk further information regarding countries would be needed as relationship of price with distance and duration would vary.

total upfront price per prediction type

- 20 entries in metered price and 243 entries in upfront price had missing values – *this was not linked to the rider's or driver's app version.*
- However, upfront price was NULL whenever metered price was NULL and it was also NULL for all other prediction_price_type except for category 'upfront', *hence the updated upfront price when destination was changed couldn't be analyzed.*
- Since predicted distance and predicted duration are strongly correlated therefore the model used for price prediction should not be affected by multicollinearity issue, otherwise dropping one of these variables might be beneficial.

- The strong correlation of predicted duration and predicted distance with upfront price, as well as the strong correlation of the duration and distance with metered_price indicates that predicted duration and predicted distance should be optimized to get the right upfront_price (which in turn would give right metered price)
- The actual duration is not as strongly correlated to predicted duration as compared to actual distance to predicted distance which means that we need to optimize predicted duration or add other relevant variables.
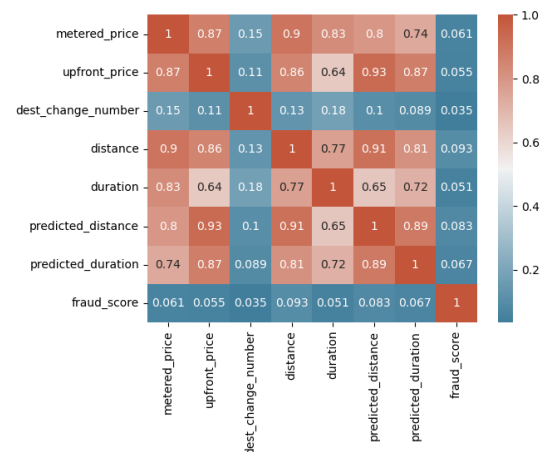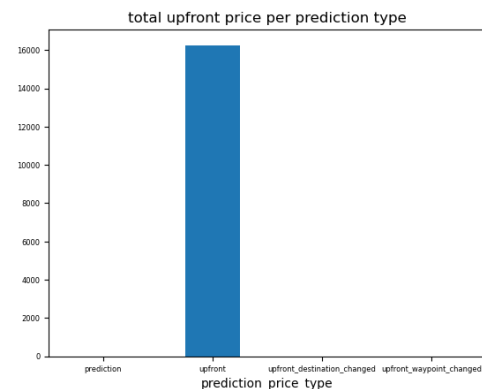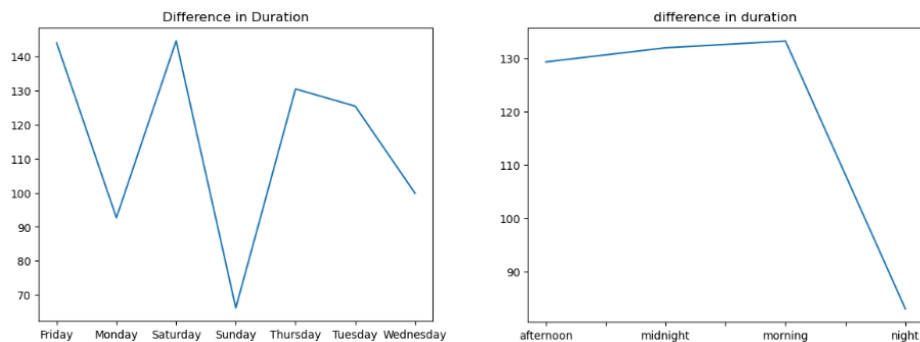
- The predicted price, distance and duration follow the same trend as of the actual price, distance, and duration, however, predicted price and duration seem to be underestimated as compared to predicted distance.
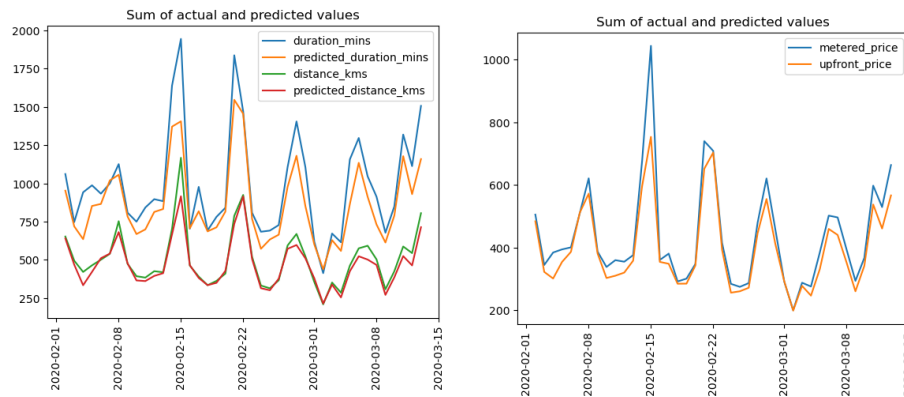
- What percentage of rides result in 20% greater (price, distance, and duration) as compared to the predictions.

Metered price > Upfront Price	34%
Duration > Predicted Duration	33%
Distance > Predicted Distance	28%

- The graphs below show that time of the day and day of the week might be important features to consider as the difference in duration varies at different times and days.

- The Mean Square Error (MSE) between existing upfront price and metered price is 42. After applying a simple decision tree for regression using predicted_duration and predicted_distance as features and metered_price as output gave MSE 48.
- In another decision tree by using predicted duration and day of the week as features the MSE was reduced to 43.

Results

The upfront price precision can possibly be improved by following two techniques:

1. By improving the predicted duration time
2. By adding day of the week and time of the day as features to determine peak hours.