Predictive analytics

In our predictive analytics part, we start by importing the data, and collecting the rides within each hour of each day, e.g. 13 rides between midnight and 1 o’clock on the 1st of January. The weather data is then imported and paired with our existing data. After this, we add dummy variables for holidays or weekends, so we know not to expect the two daily peak from commuters.

After this, we then begin with our prediction models where we spilt our data into test and training data. The first three quarters of the year is used as training data, and the last quarter is used as test data. In our ADF test, we find the data to be stationary, however we improve on this with decomposing. From here we use ACF and PACF to estimate the parameters of the ARIMA model.

ARIMA model

We encountered problems during coding, which made us unable to get a MAE. Fortunately, we managed to find results for the other models

Random forests

We use 1000 estimators to train our model, which results in Mean Absolute Error of 27,37. This is the lowest MAE we find, however it is possible this could be further improved by pruning the trees. Due to the nature of the model, there is a risk of overfitting, which we can only comment on if we had multiple data set to test our model on.

Regression

In our simple polynomial regression, we find that 2 degrees is best, resulting in MAE of 90,57. In the Ridge regression we find an alpha of 1e-10 to be best, resulting in MAE of 68,28 bikes. Finally in the Lasso regression, we find the same alpha, giving us a MAE of 67,25.