

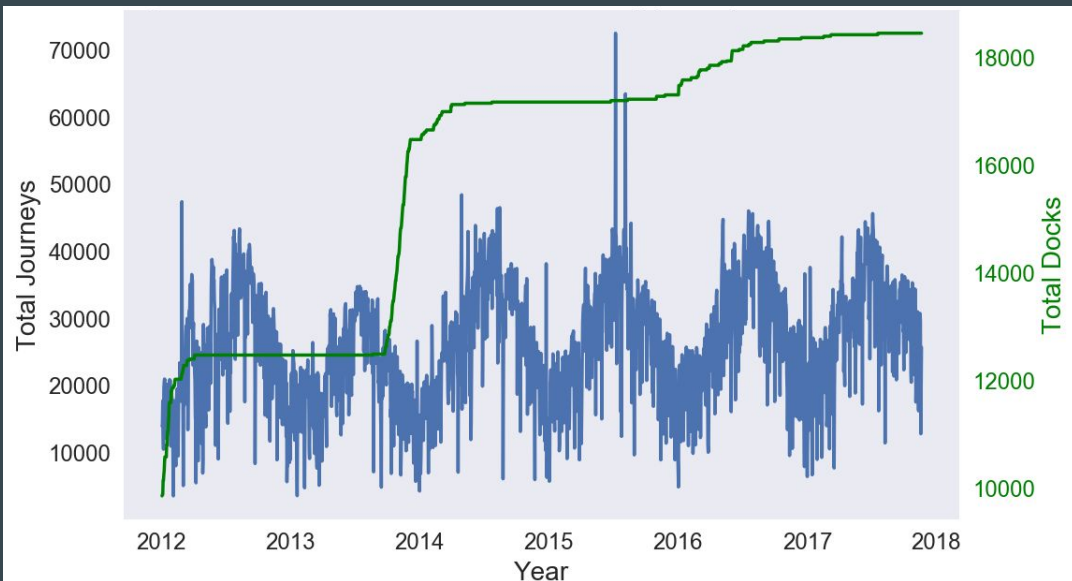
Sharing is Caring Predicting Daily Cycle Hires

...

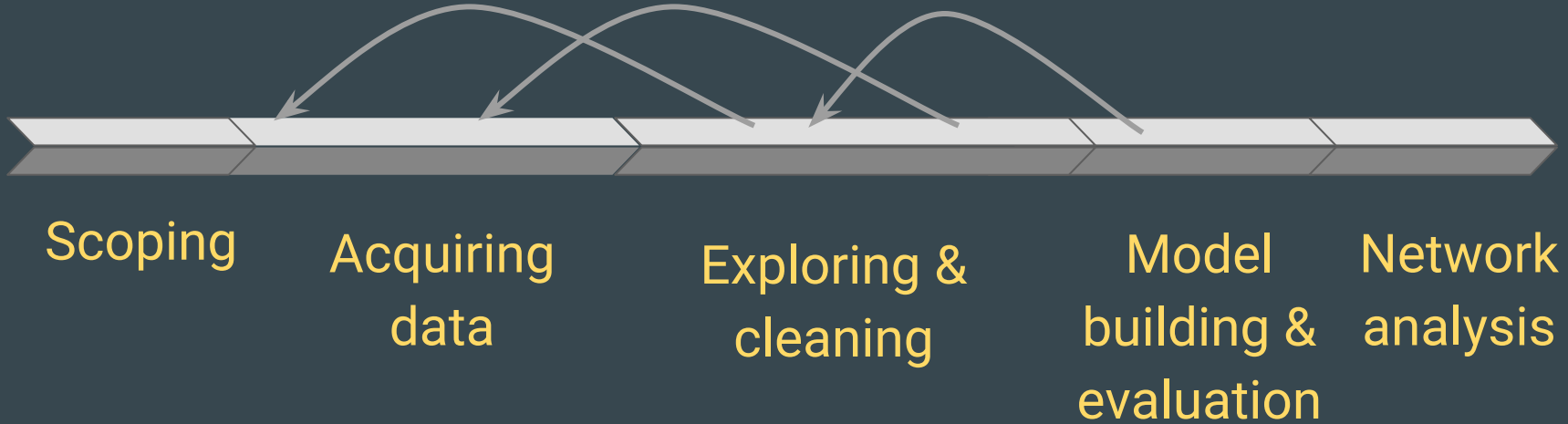
Blae Quayle
General Assembly DSI4

Santander Cycles

780 docking stations
2nd largest scheme in Europe



Project Framework



Making a dataset

API



 Transport for London Unified API

CSV

 [06JourneyDataExtract18May2016-24May2016.csv](#)

 [07JourneyDataExtract25May2016-31May2016.csv](#)

 [08JourneyDataExtract01Jun2016-07Jun2016.csv](#)

DIY

London faces biggest tube strike in more than 10 years over night shift pay

Exploring the data

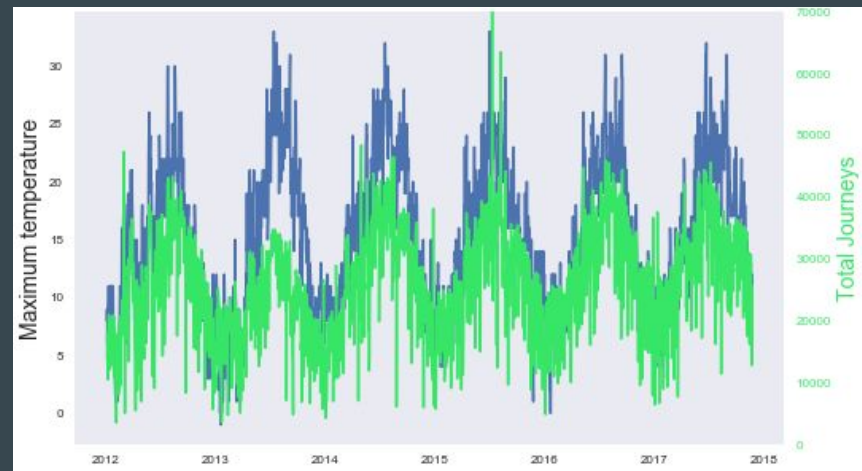
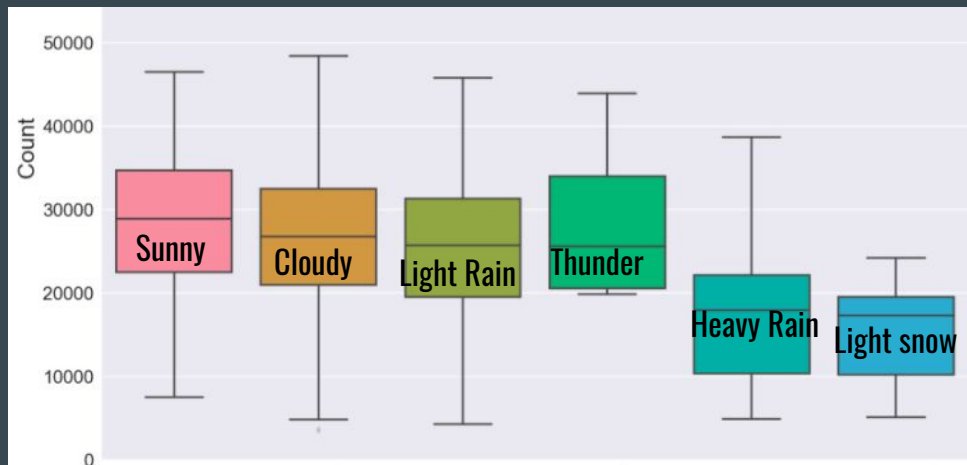
72504 max
daily hires

Sunday
quietest day

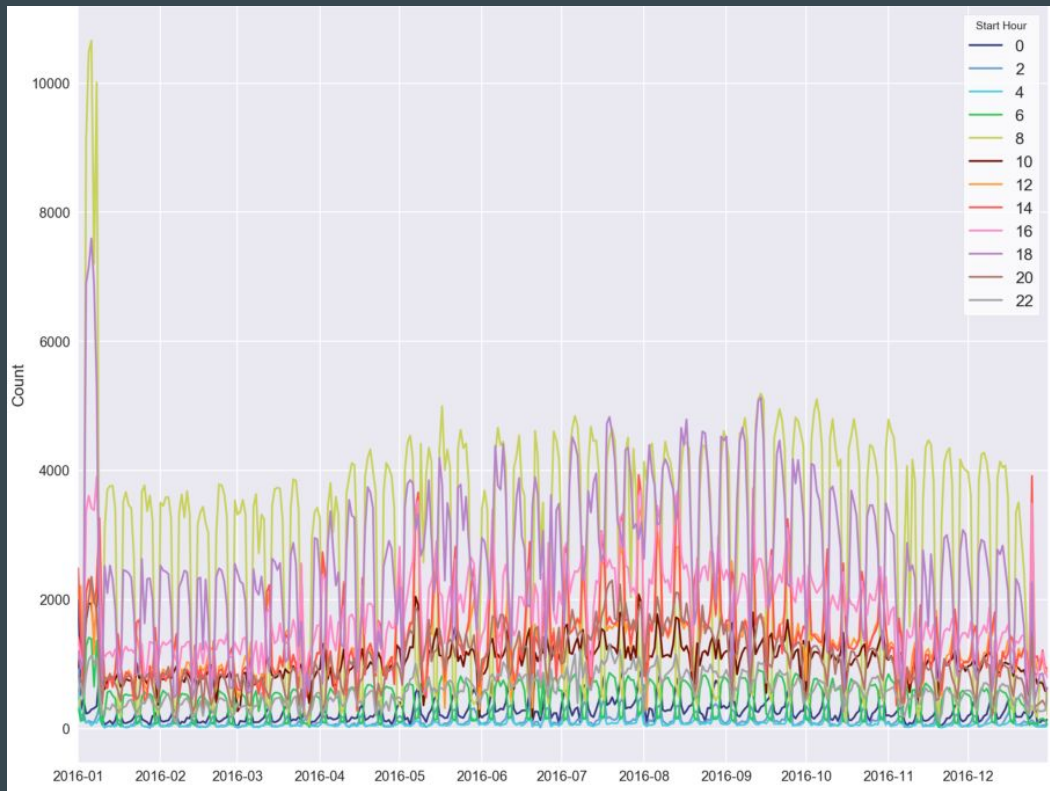
2016
743 hours
longest hire

2016
13 minutes
median hire

2016
2 hires per
bike/day

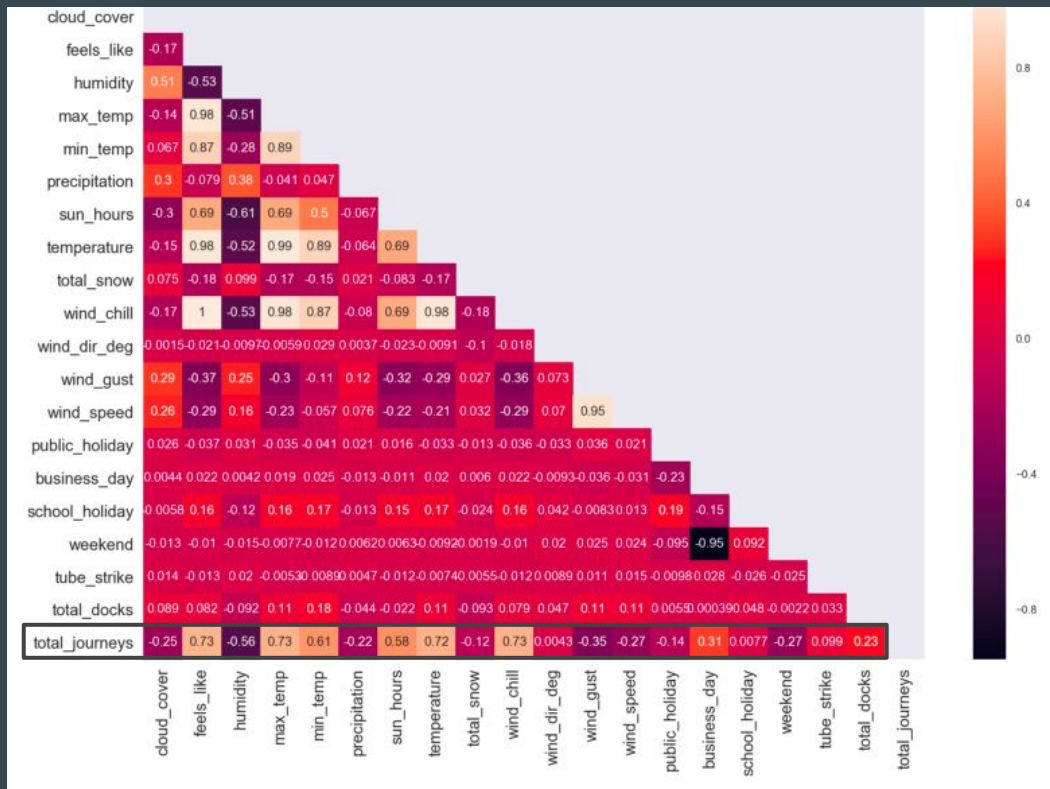


Down the rabbit hole...



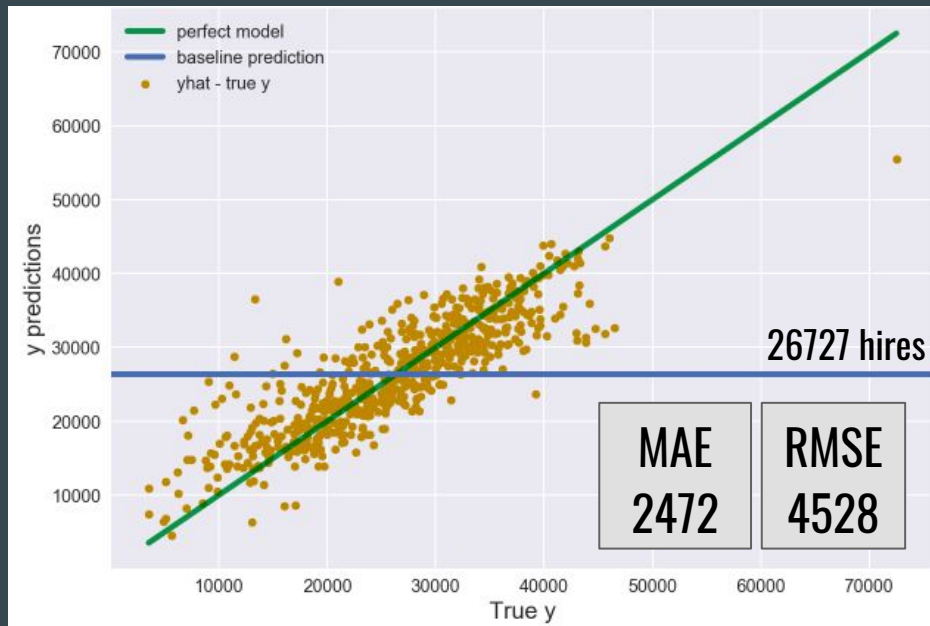
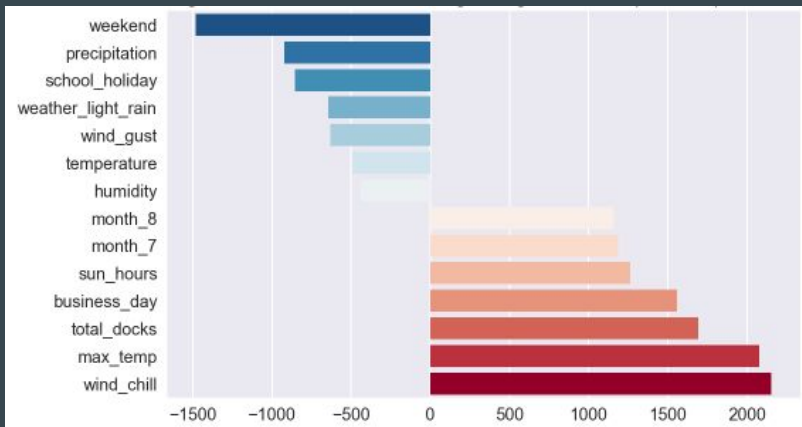
Commuters failing
hard at their New
Year's resolution

Heat Map



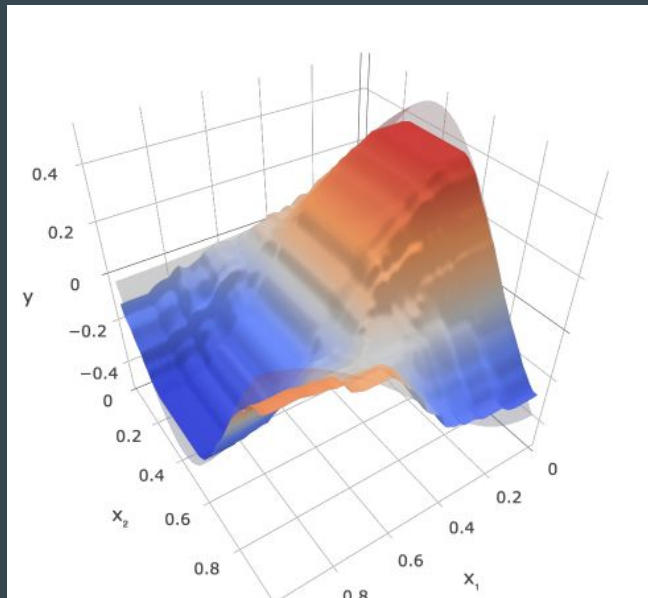
Predicting daily hires

- Drop min_temp & feels_like
- Dummify categorical variables
- 41 predictors
- Train/test split 70:30
- Standardise training set

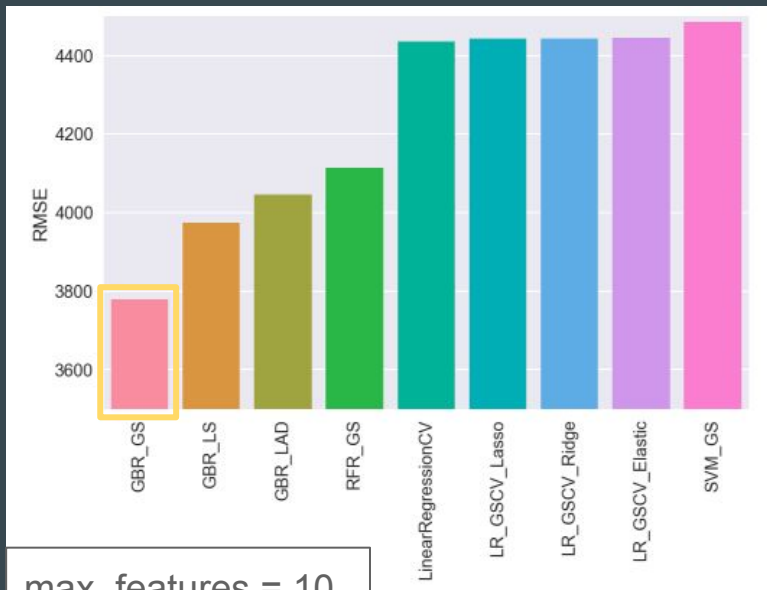


Gradient Boosting

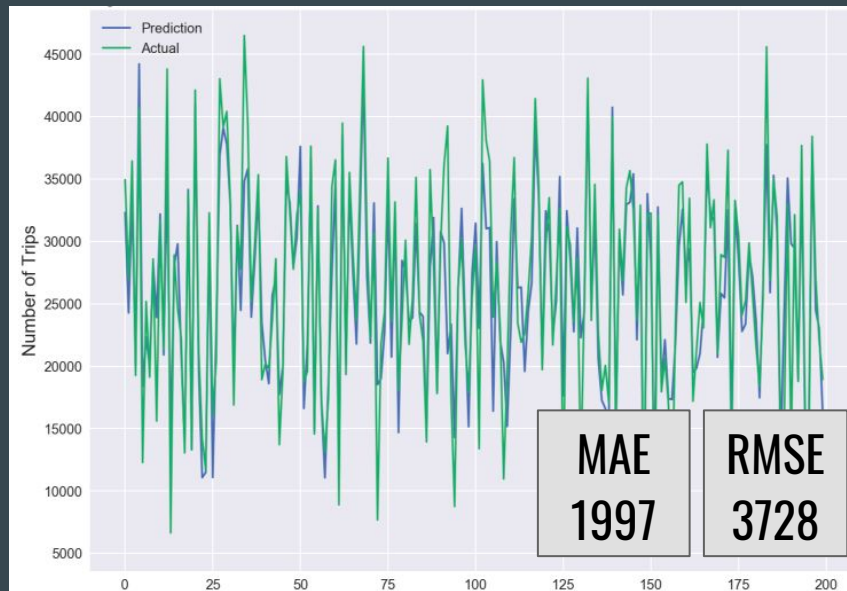
- Many weak learners -> overlapping regions of feature space
- Internal regression model trained iteratively on the residuals
- Single strong learner -> optimised for bias and variance



Final Model



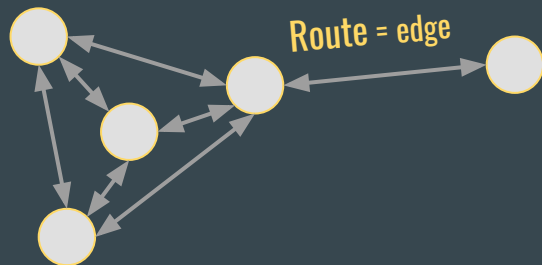
max_features = 10
learning_rate = 0.1
n_estimators = 400
max_depth = 3



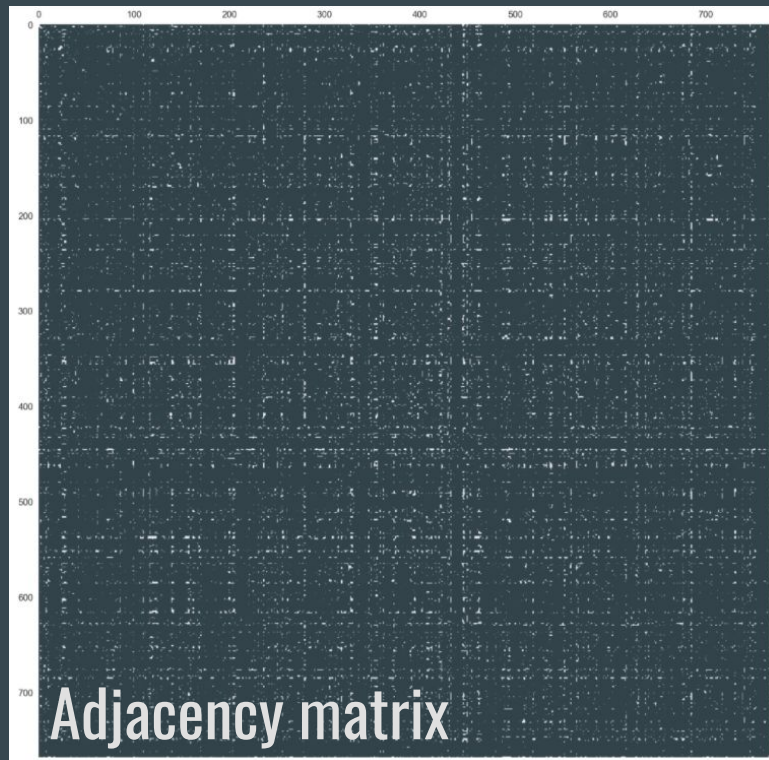
Important features: total docks, wind dir deg, humidity

Network Analysis

Dock = node

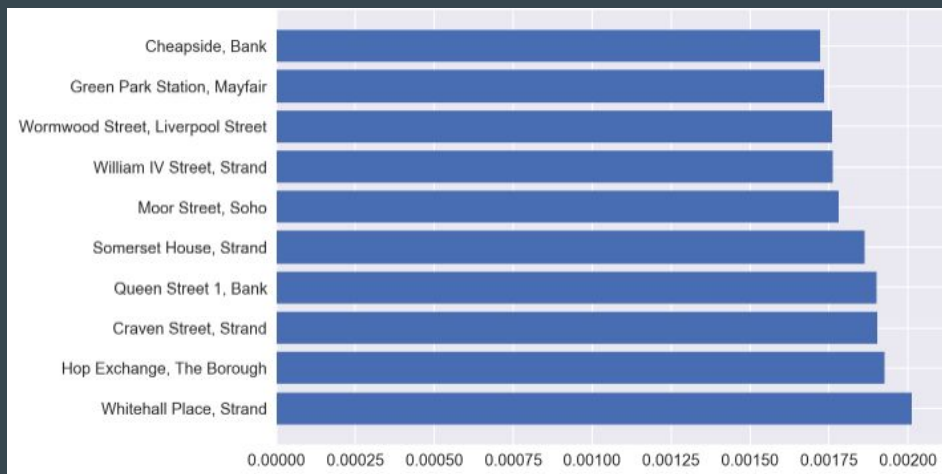
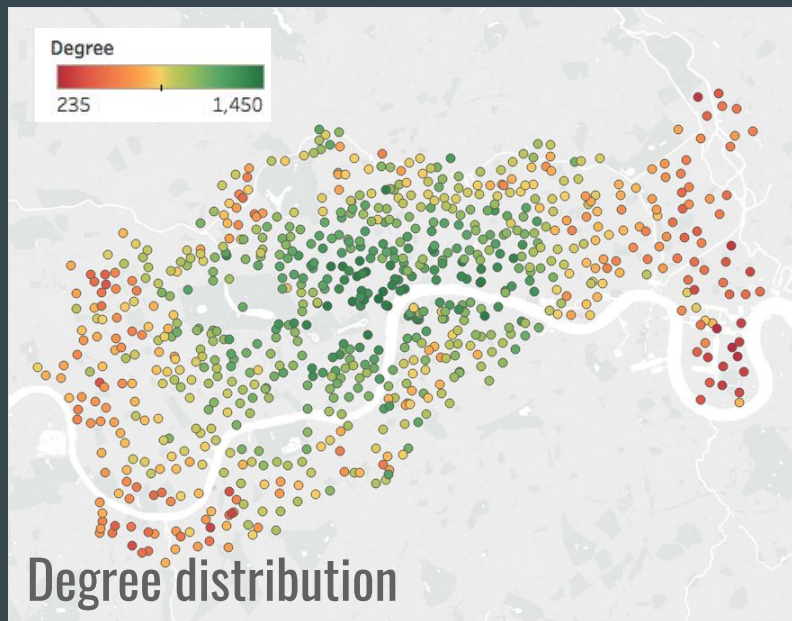


- Directed graph
- In-Degree
- Out-Degree
- Weighted by number of journeys



Degree & centrality

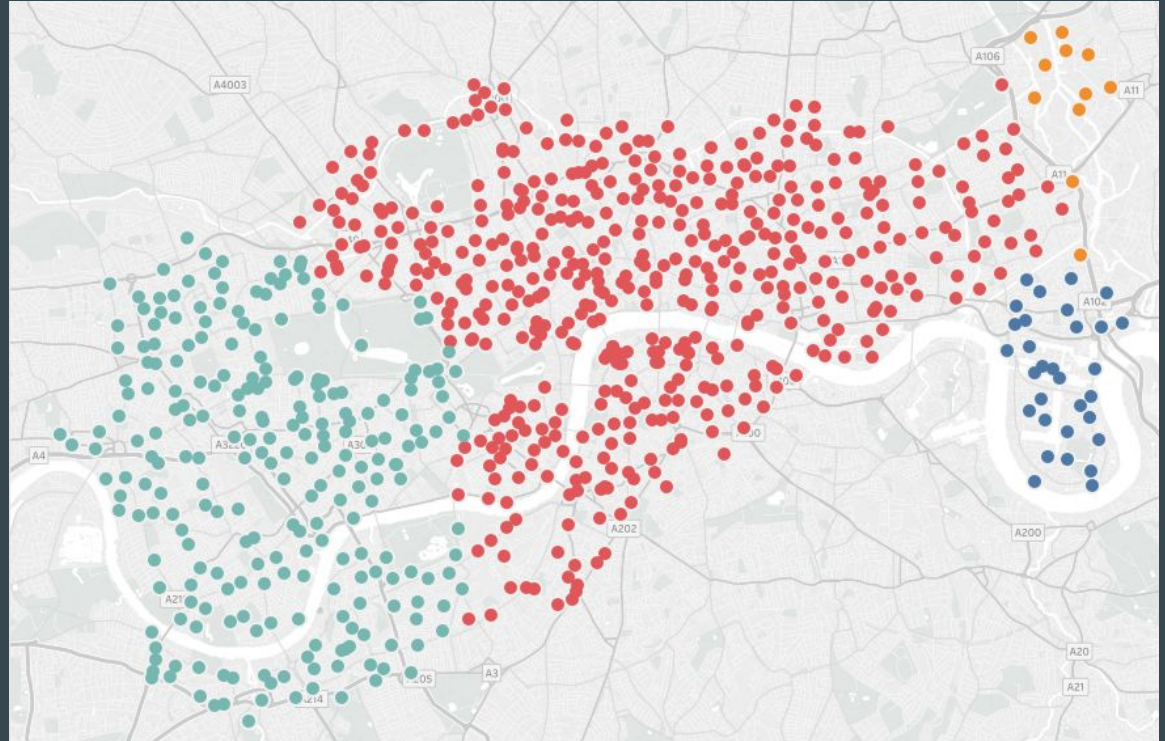
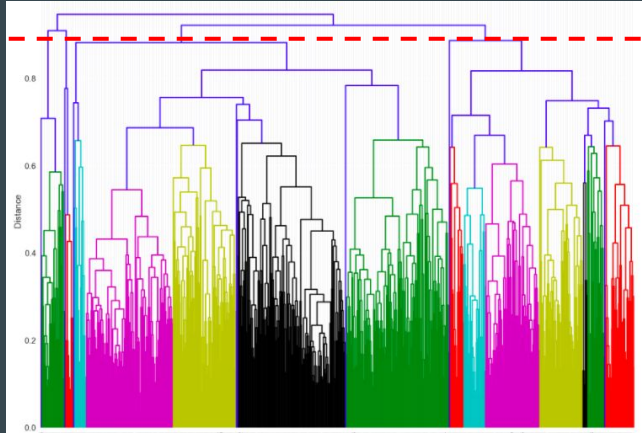
→ Importance of nodes within network



Betweenness centrality is number of shortest paths from all nodes to all others that pass through that node

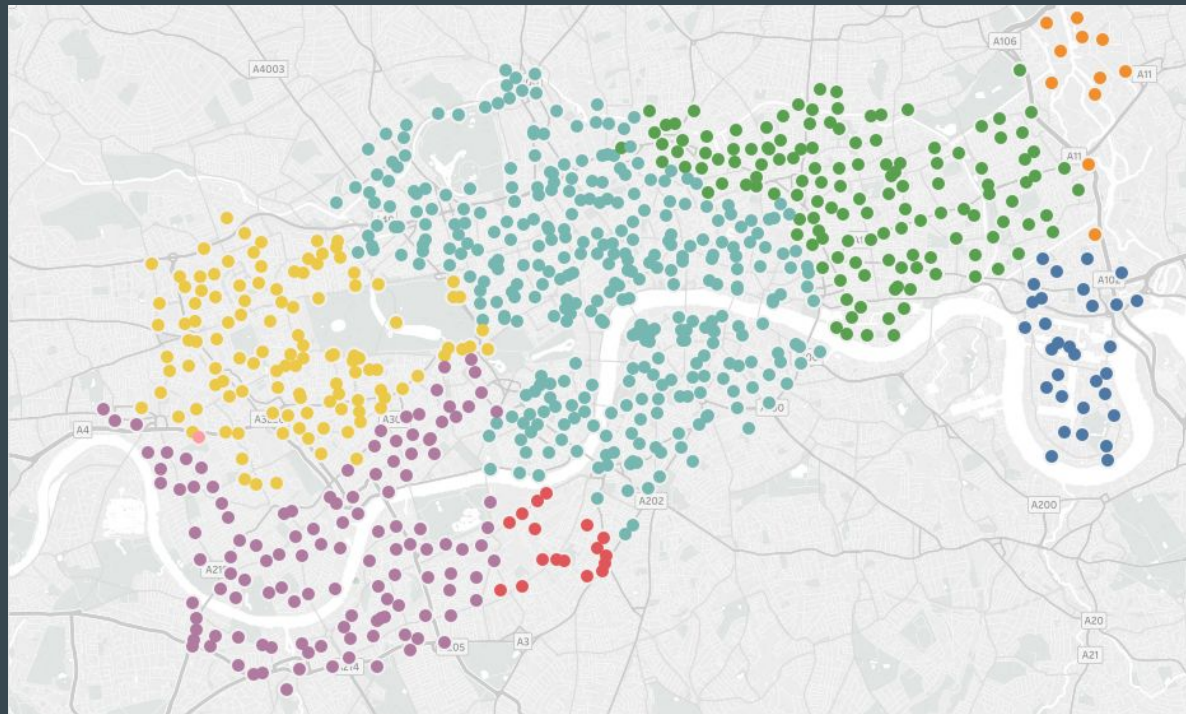
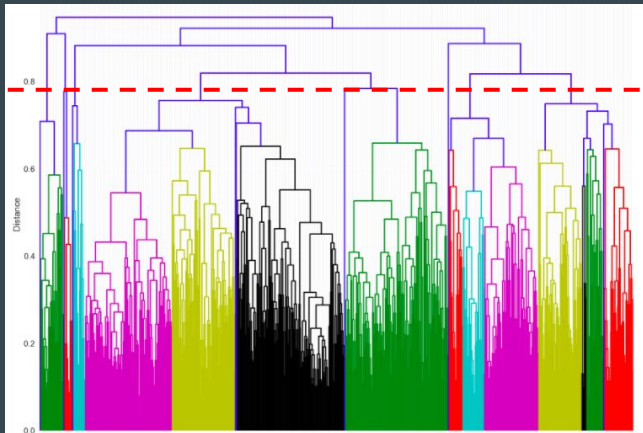
Hierarchical clustering

0.9 - 4 clusters



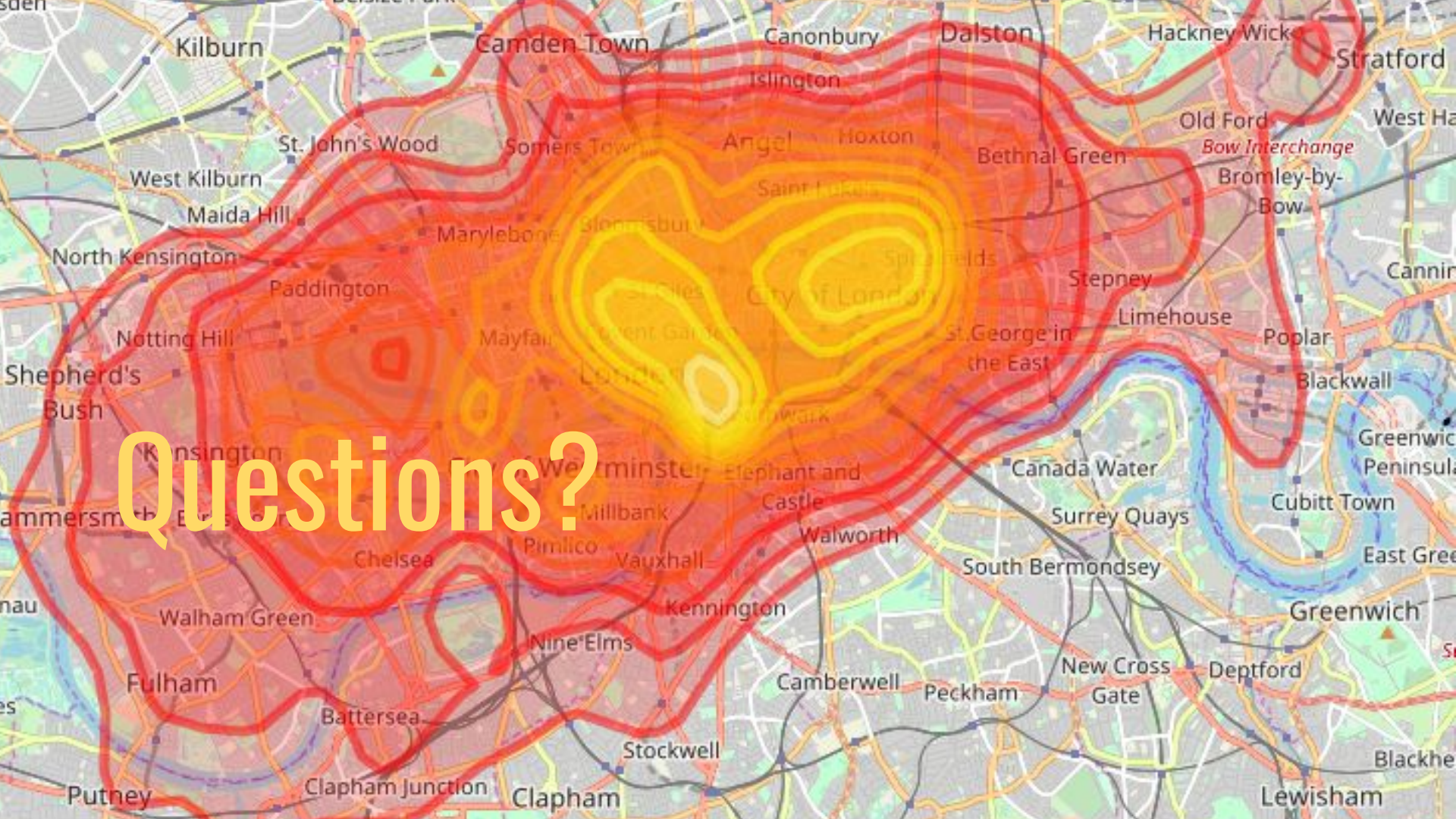
Hierarchical clustering

0.8 - 8 clusters



Things I ran out of time to do...

- Bike redistribution data
- Classifier to predict availability at a docking station: 'Spaces' or 'Full'
- Calculate nearest dock and repeat, until one with 'Spaces' is found
- Develop a Flask app to host the outputs



Questions?