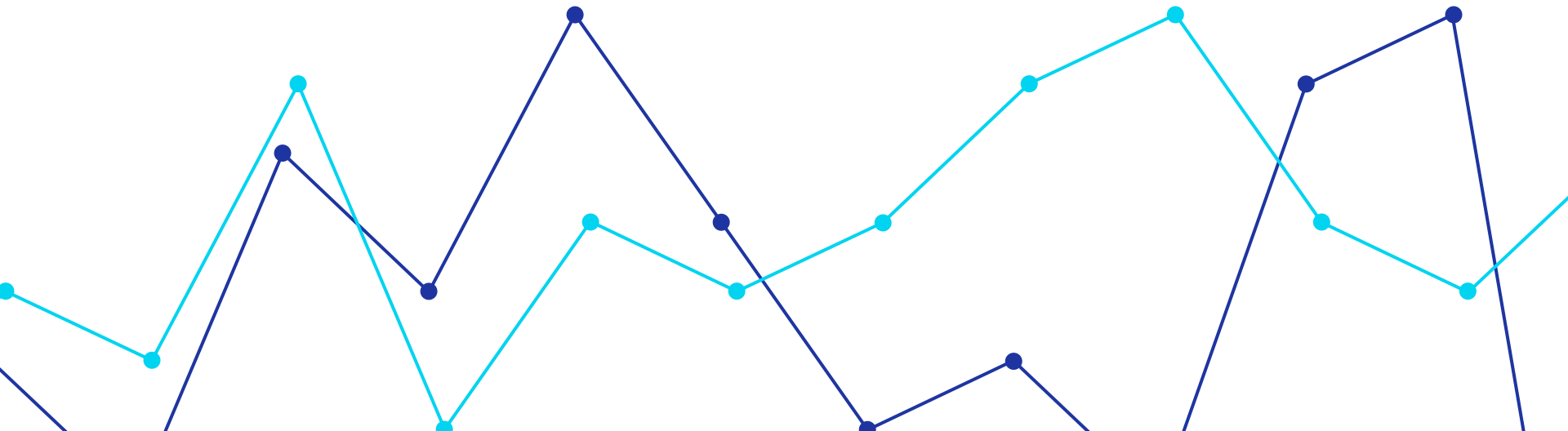


# Helsinki City Bikes

Mid Bootcamp Project

Heloisa Bal



# Helsinki City Bikes

## What?


Helsinki City Bikes is a public bicycle system in Helsinki.

## Why?

High interest in biking and micro mobility in general.

## Data sets

 Kaggle Dataset  
2018 - 2020 (3 years)

 Meteorological data  
2018 - 2020 (3years)

## City bikes



The new city bike season has started. Registration and season passes are now available!

### Buy a pass

The city bike season starts on 1 April and ends on 31 October.

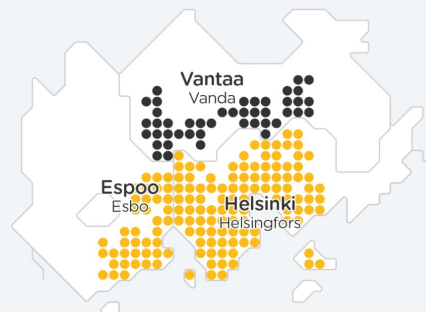
Please note that the bike system in place in Helsinki and Espoo is different from the system used in Vantaa and you cannot mix bikes from the two systems.

The pass for the whole season for Helsinki and Espoo costs EUR 35, for Vantaa EUR 30.

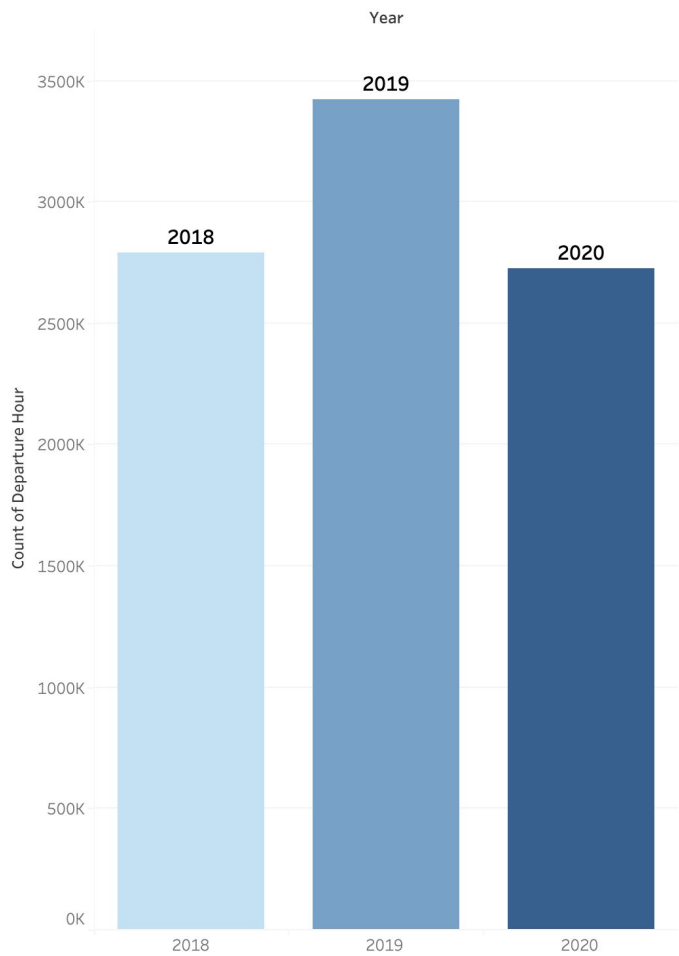
Start by selecting area

Helsinki and Espoo

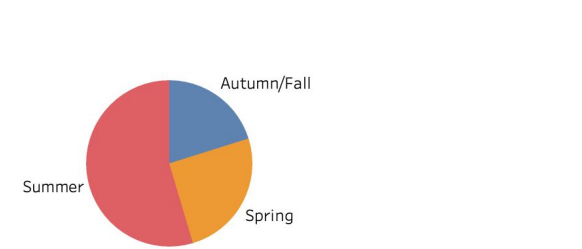
Vantaa



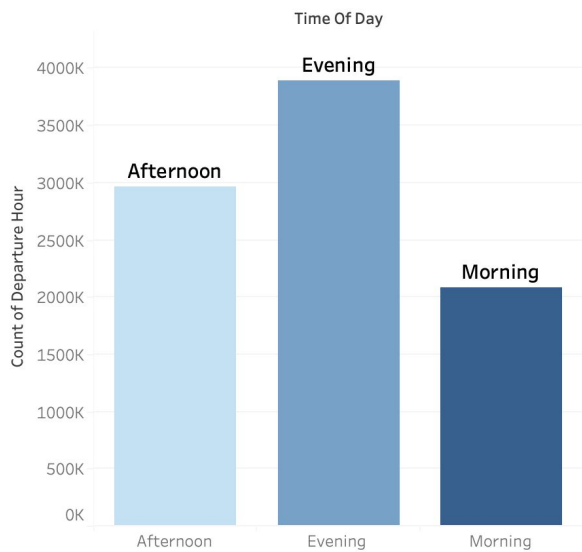
Bike Trips per Year



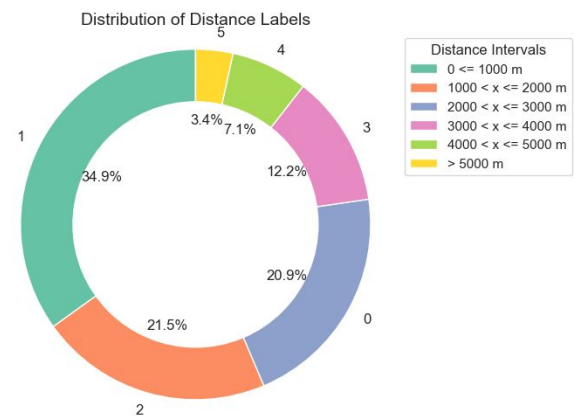
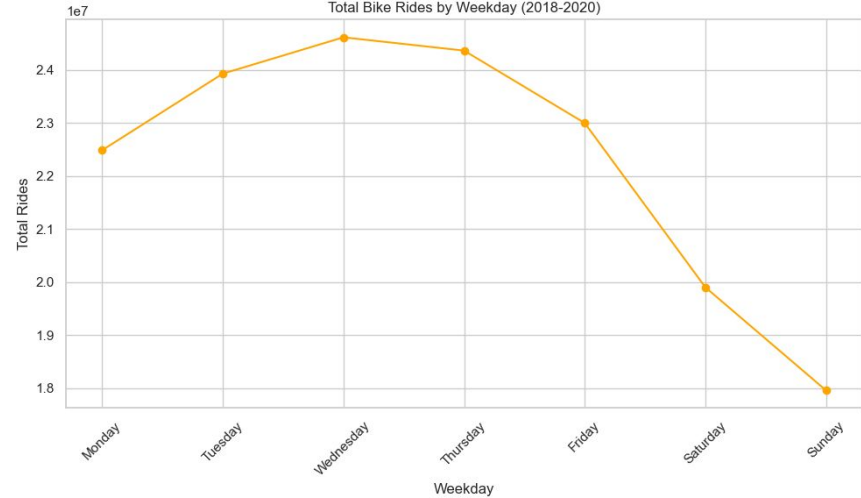
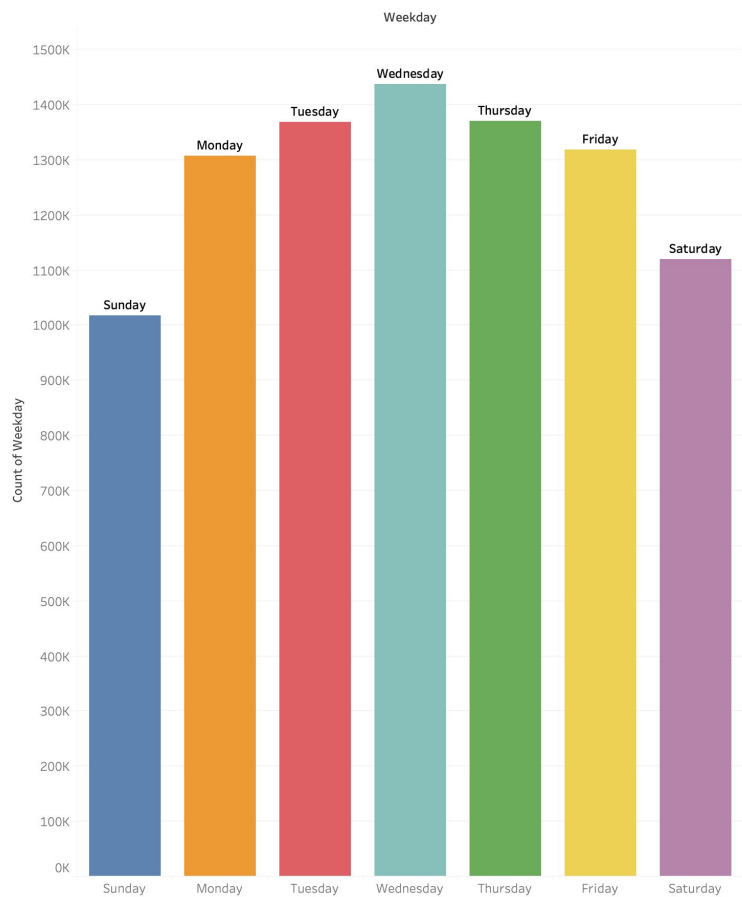
Trips by Season



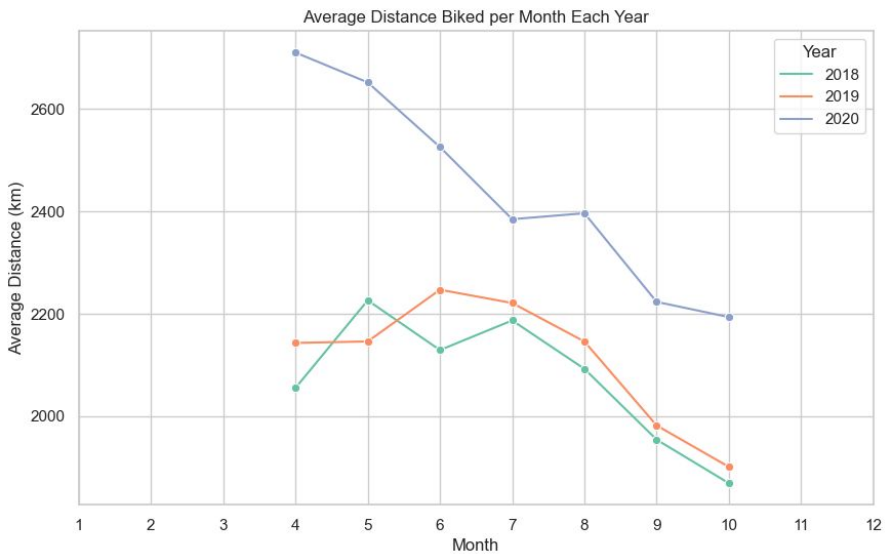
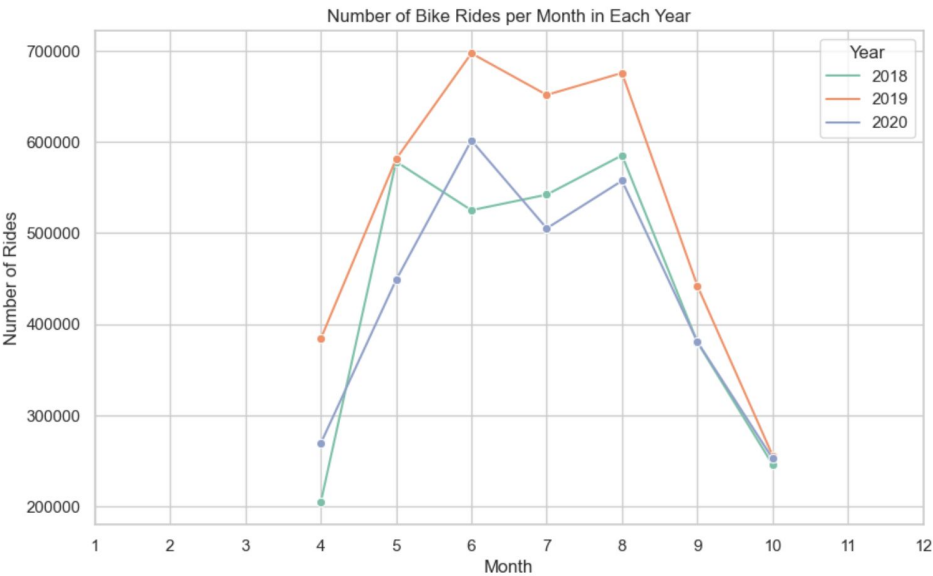
Trip Counts by Time of Day



Rides per Weekday

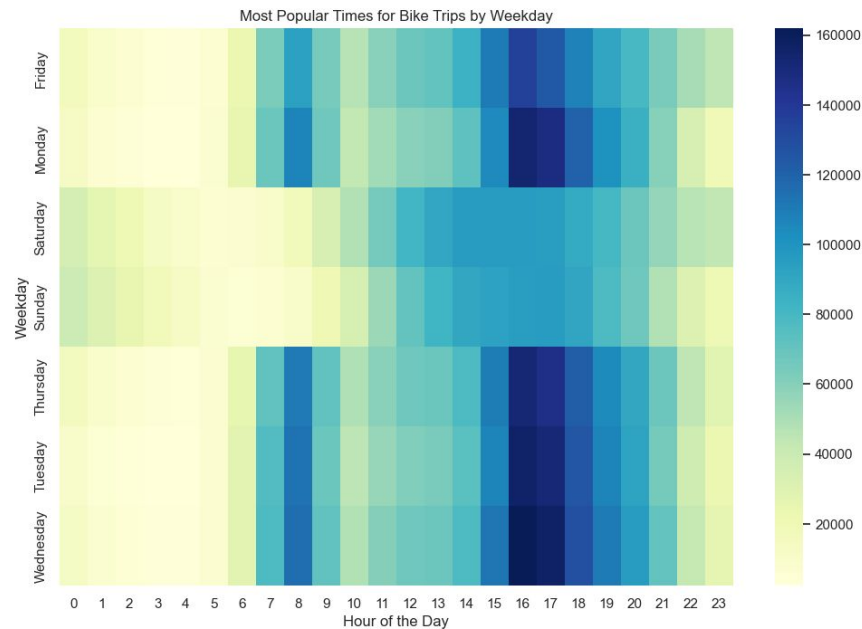
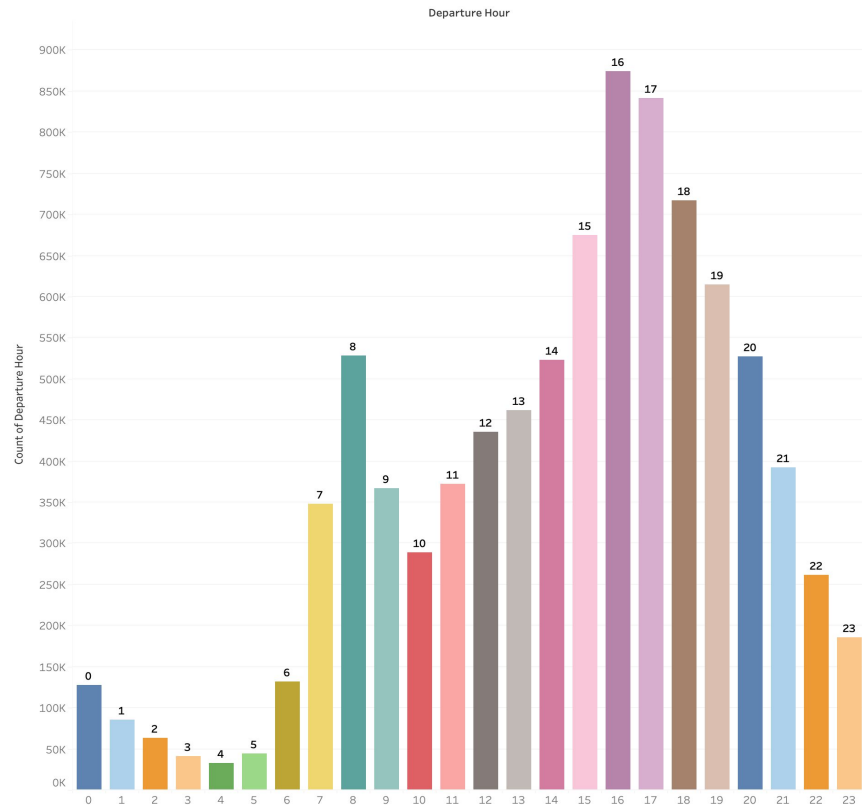


# Monthly bike rides & Monthly average bike distance



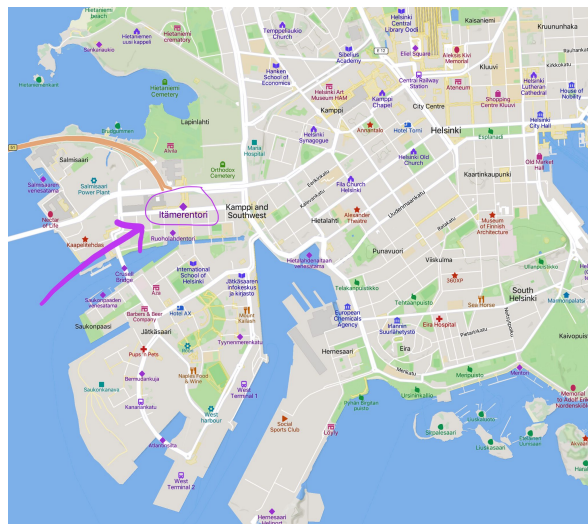
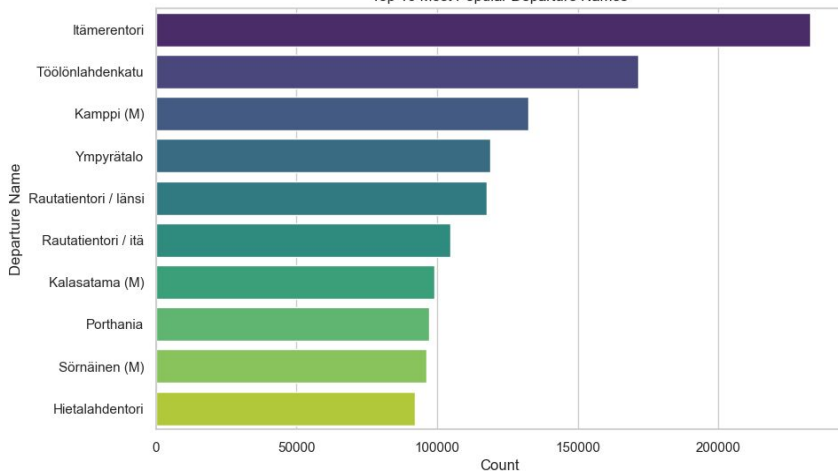
# Most popular biking hours

Most Popular Biking Hours

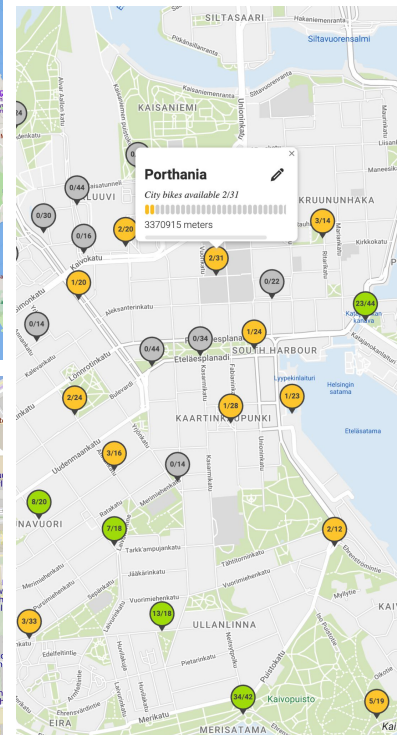
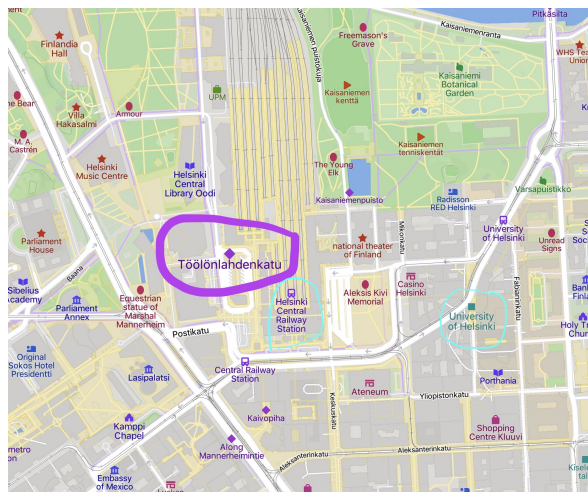
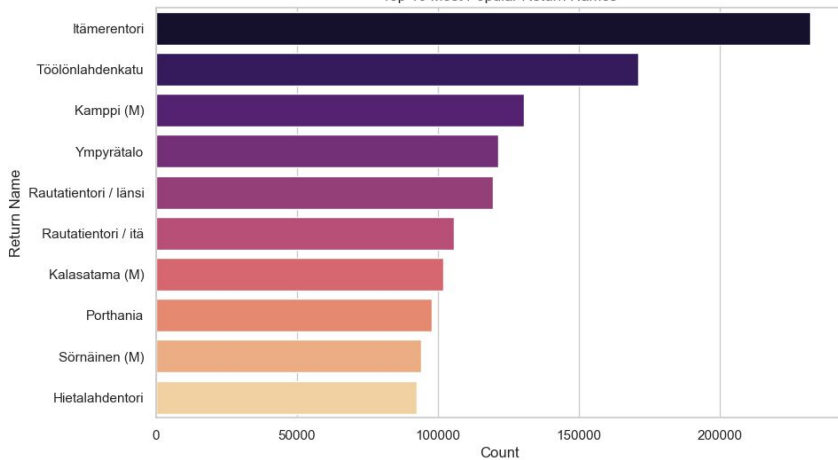


# Popular Docks

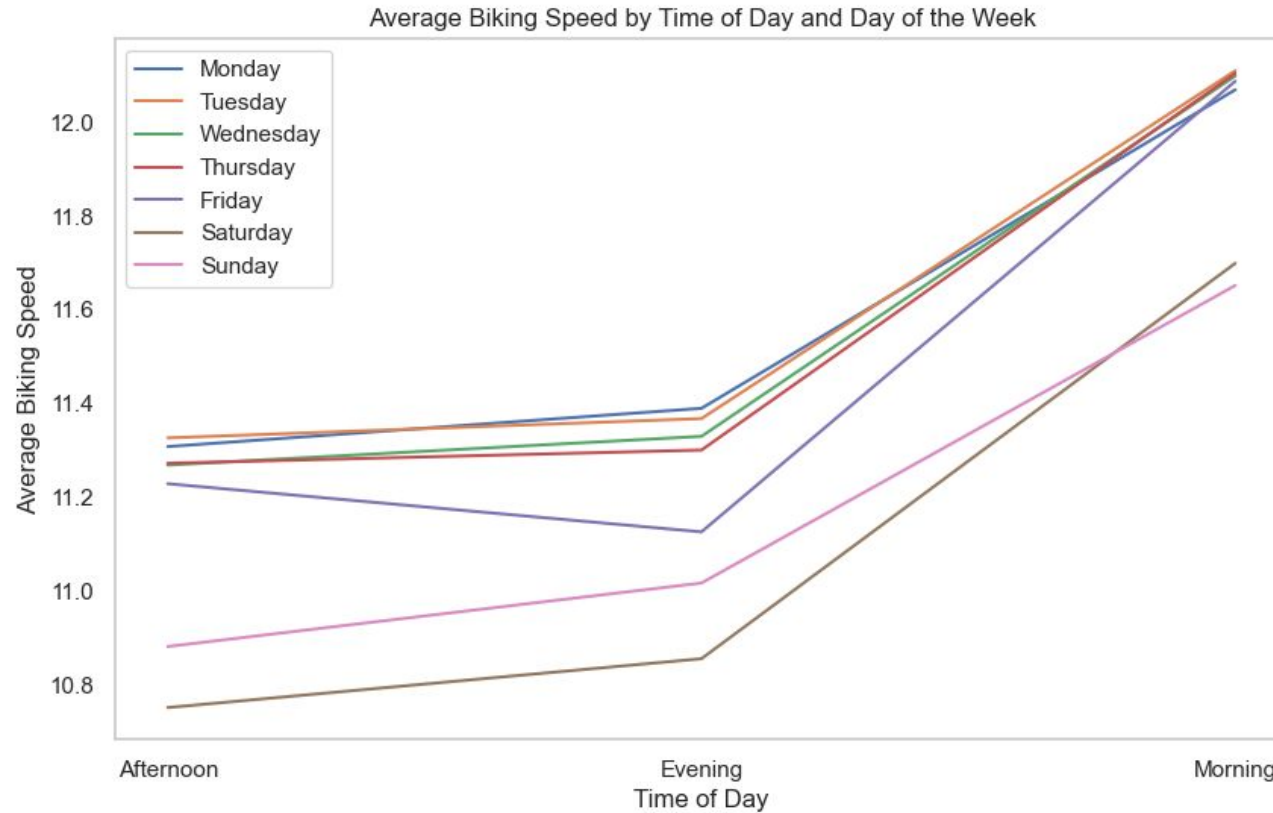
Top 10 Most Popular Departure Names



Top 10 Most Popular Return Names

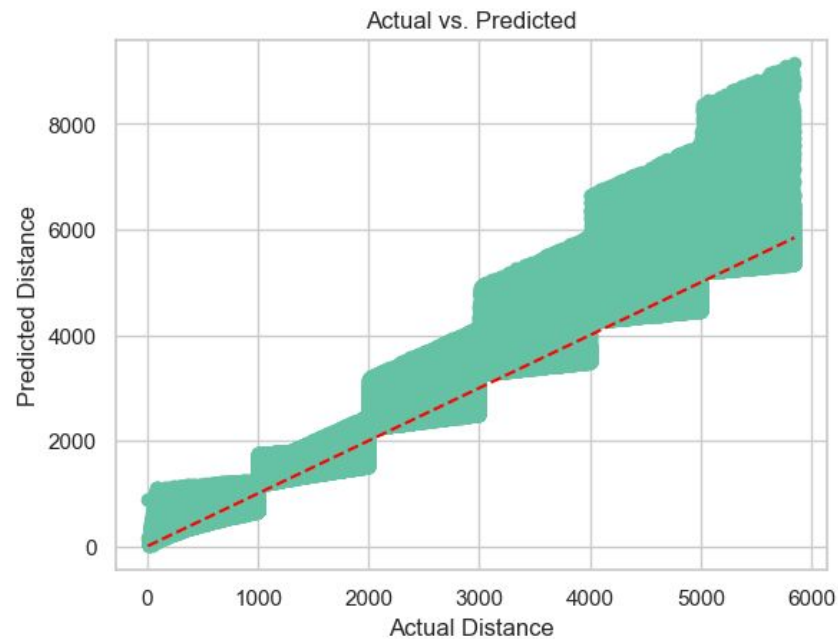


## Most interesting insight!





# Linear regression - Predicting distance (m)



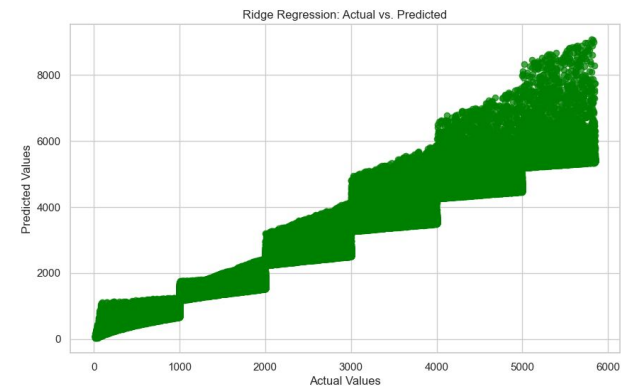
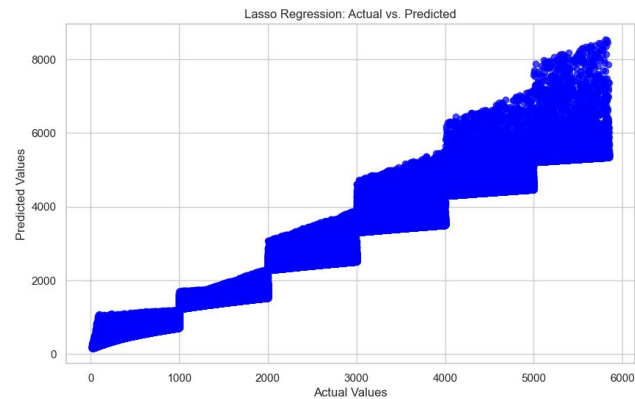
R-squared: 0.9687

Mean Squared Error: 51149.9066

Root Mean Squared Error: 226.1635

```
1 y_test[:5]
4590461 871.0
9638049 679.0
1277544 1823.0
3821326 1917.0
2089904 2664.0
Name: distance_m, dtype: float64

1 predictions_test[:5]
array([ 876.6875 , 640.30371094, 1512.50244141, 1679.890625 ,
        2528.74658203])
```

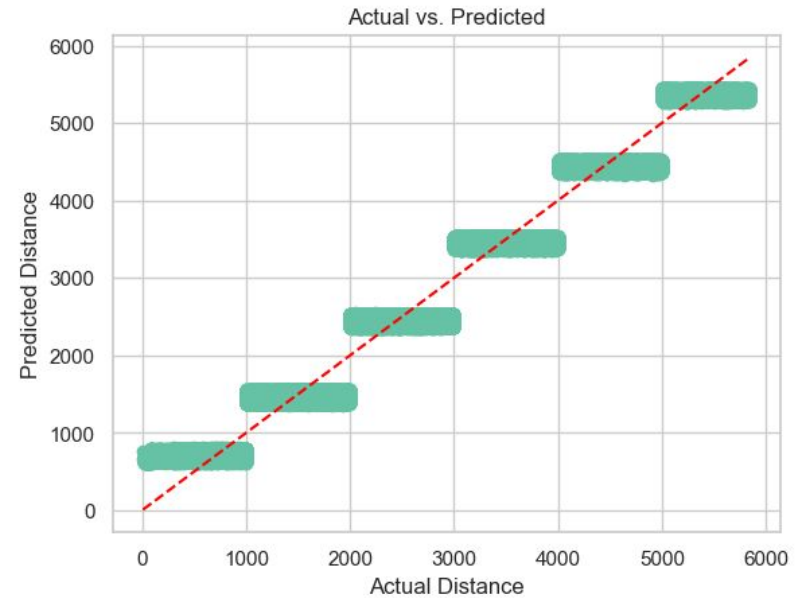
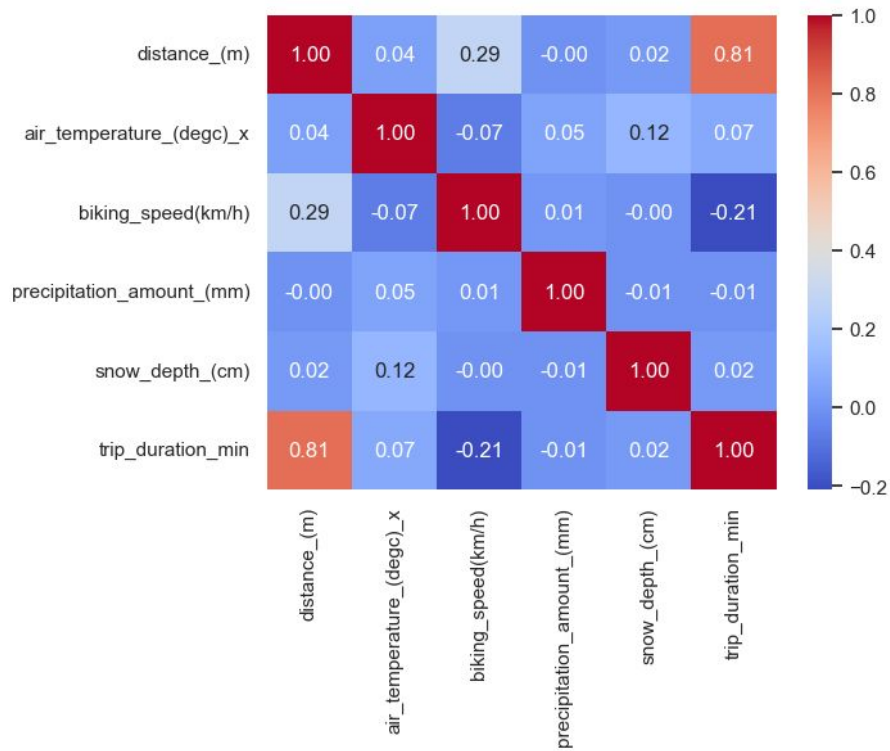


Model: lasso, train R2: 0.9682949228117823 -- test R2: 0.9682393998914398

Model: ridge, train R2: 0.9686873425868942 -- test R2: 0.9686433169378844

Model: elastic, train R2: 0.21608410810857315 -- test R2: 0.21614102981897487

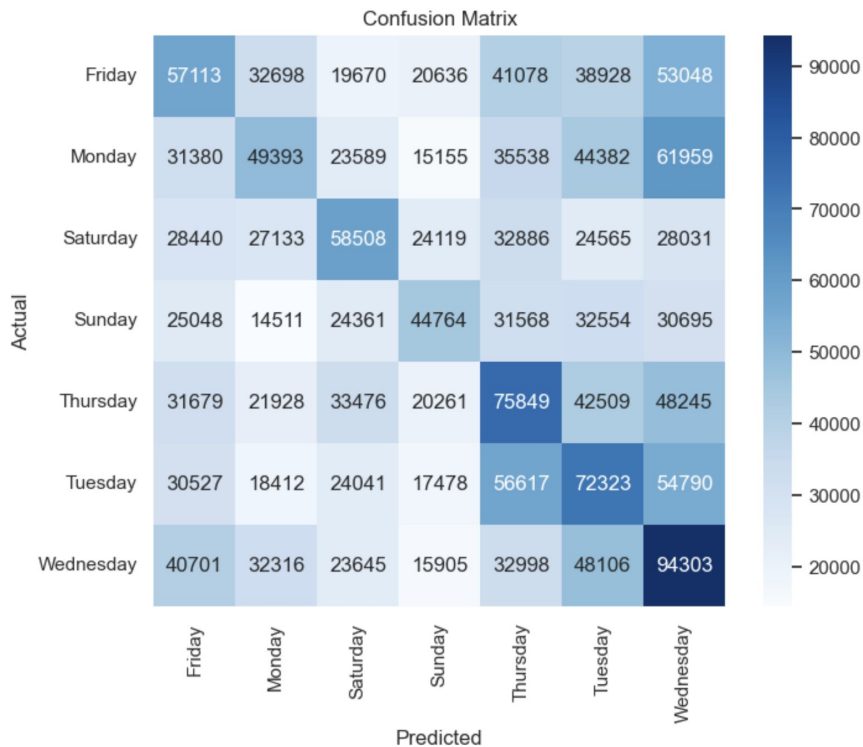
## New try based on multicollinearity



R-squared: 0.9563  
Mean Squared Error: 71338.1096  
Root Mean Squared Error: 267.0919

# Multinomial Logistic Regression - Predict day of the week

- Not a good model fit for predicting day of the week



```
LogisticRegression  
LogisticRegression(max_iter=1000, multi_class='multinomial')
```

Accuracy: 0.25295786748283844

Classification Report:

	precision	recall	f1-score	support
Friday	0.23	0.22	0.22	263171
Monday	0.25	0.19	0.22	261396
Saturday	0.28	0.26	0.27	223682
Sunday	0.28	0.22	0.25	203501
Thursday	0.25	0.28	0.26	273947
Tuesday	0.24	0.26	0.25	274188
Wednesday	0.25	0.33	0.29	287974
accuracy			0.25	1787859
macro avg	0.26	0.25	0.25	1787859
weighted avg	0.25	0.25	0.25	1787859

# Thank you!

Questions?

