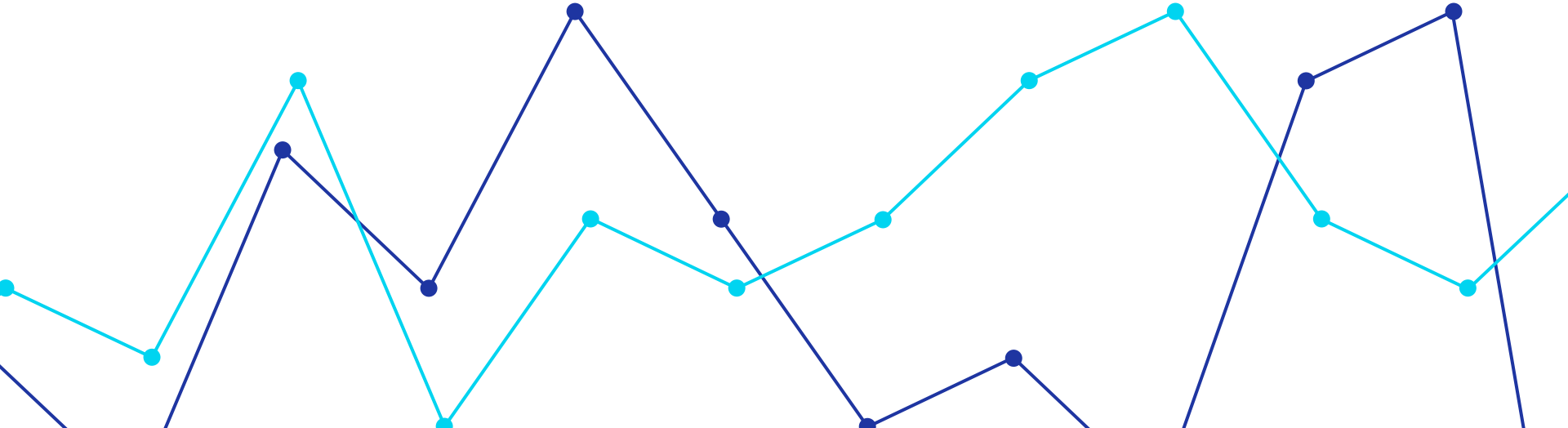
Helsinki City Bikes

Mid Bootcamp Project

Heloisa Bal



Helsinki City Bikes

What?

Helsinki City Bikes is a public bicycle system in Helsinki.

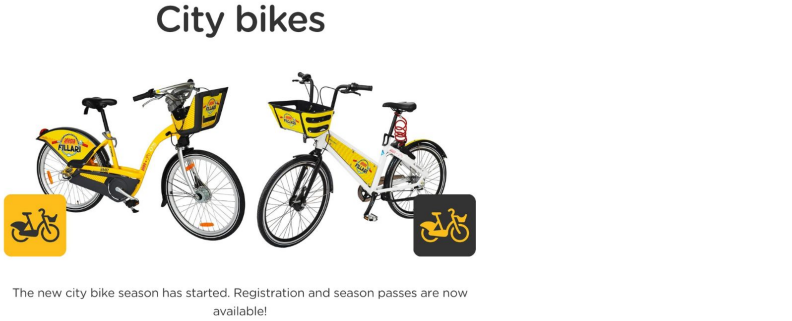
Why?

High interest in biking and micro mobility in general.

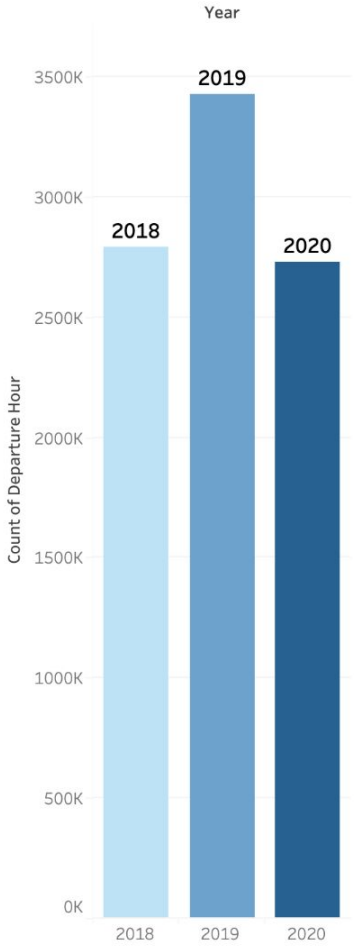
Data sets

Kaggle Dataset
2018 - 2020 (3 years)

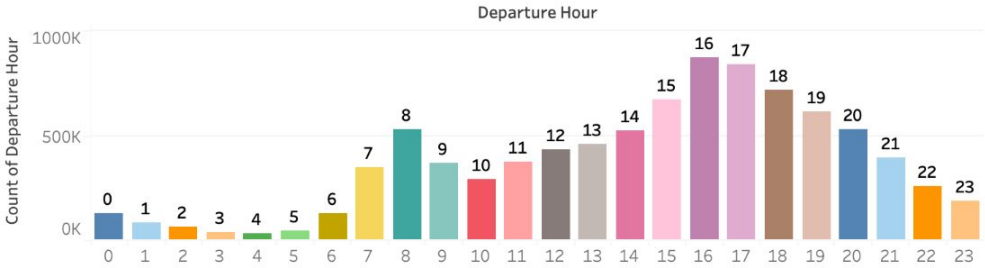
Meteorological data
2018 - 2020 (3years)



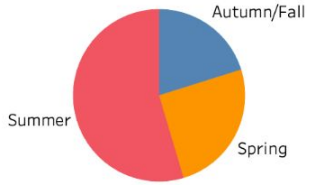
Bike Trips per Year



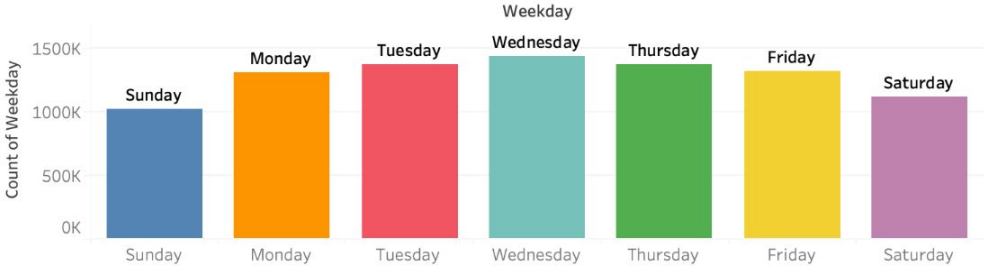
Most Popular Biking Hours



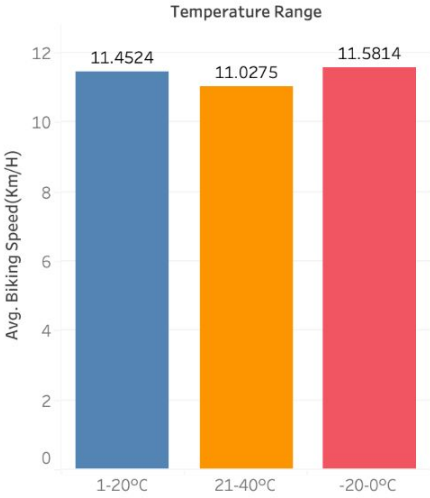
Trips by Season



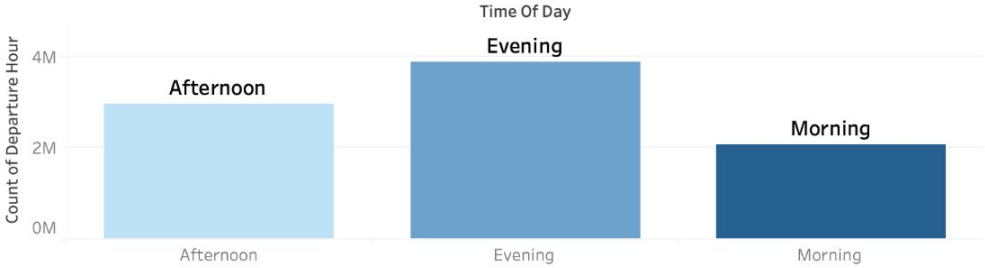
Rides per Weekday

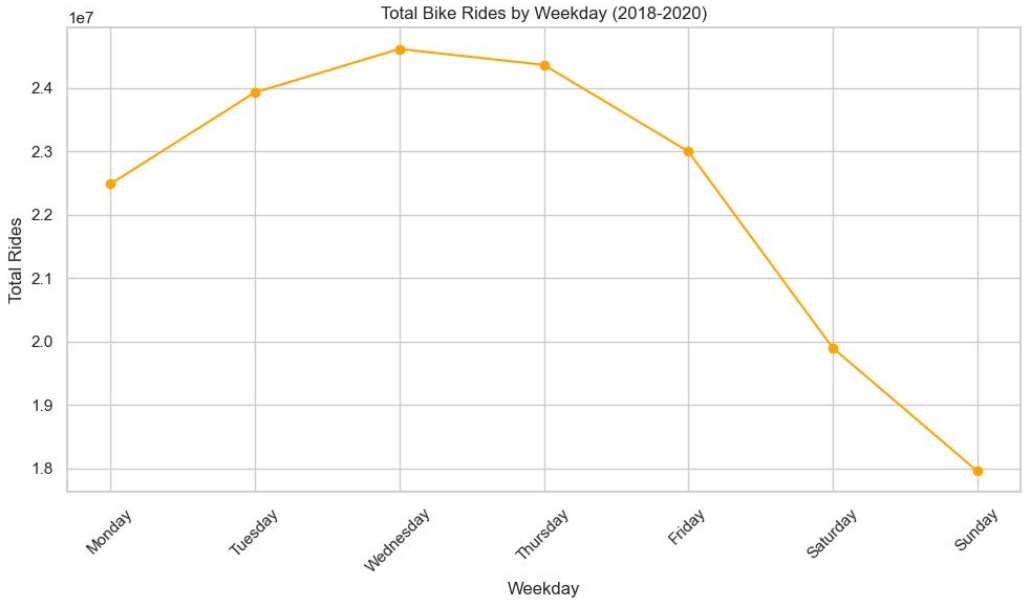


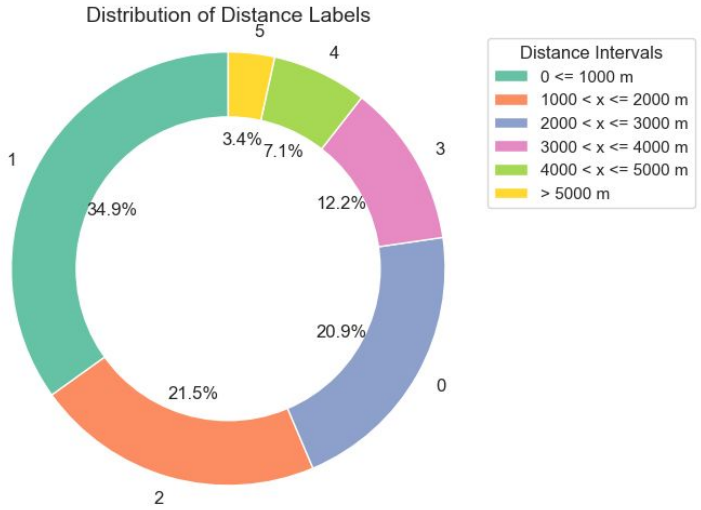
Biking Speed per Temp range



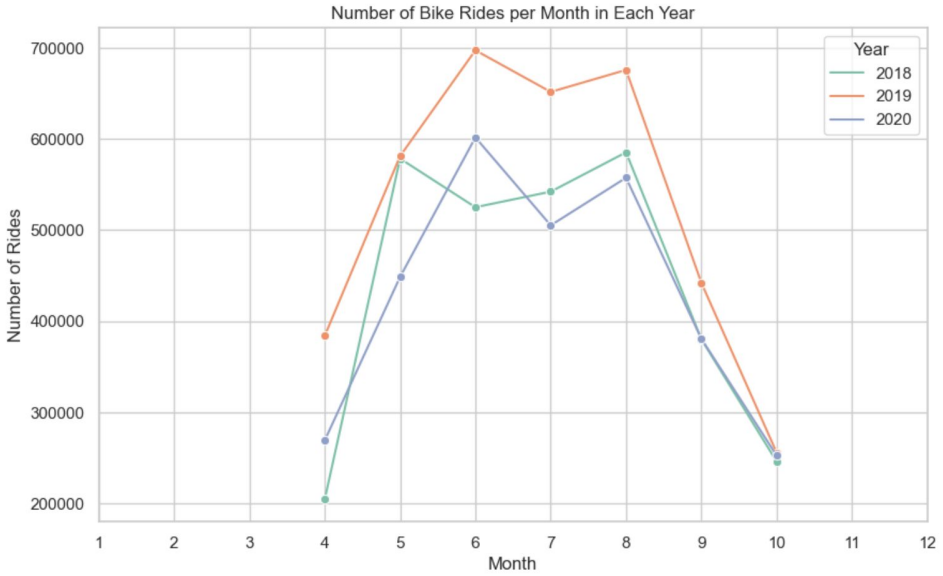
Trip Counts by Time of Day

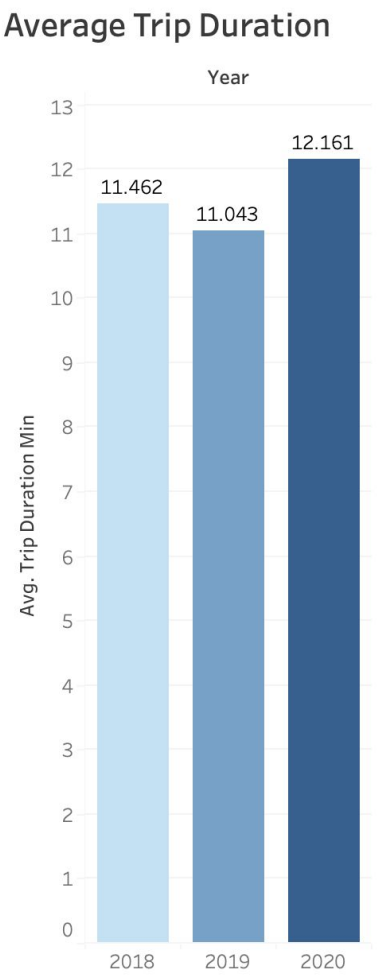



Total Bike Rides by Weekday (2018-2020)
1e7
2.4
2.3
2.2
2.1
2.0
1.9
1.8
Total Rides
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Weekday

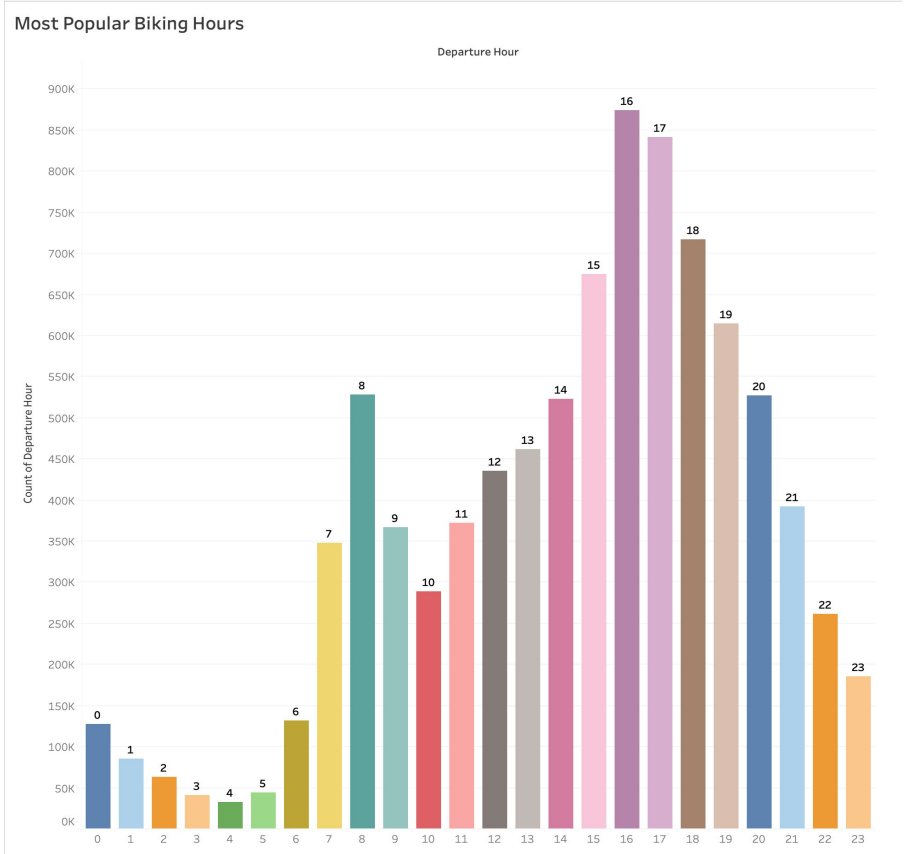

Distribution of Distance Labels
5
4
3
1
0
2
3.4%
7.1%
12.2%
34.9%
20.9%
21.5%
Distance Intervals
0 <= 1000 m
1000 < x <= 2000 m
2000 < x <= 3000 m
3000 < x <= 4000 m
4000 < x <= 5000 m
> 5000 m

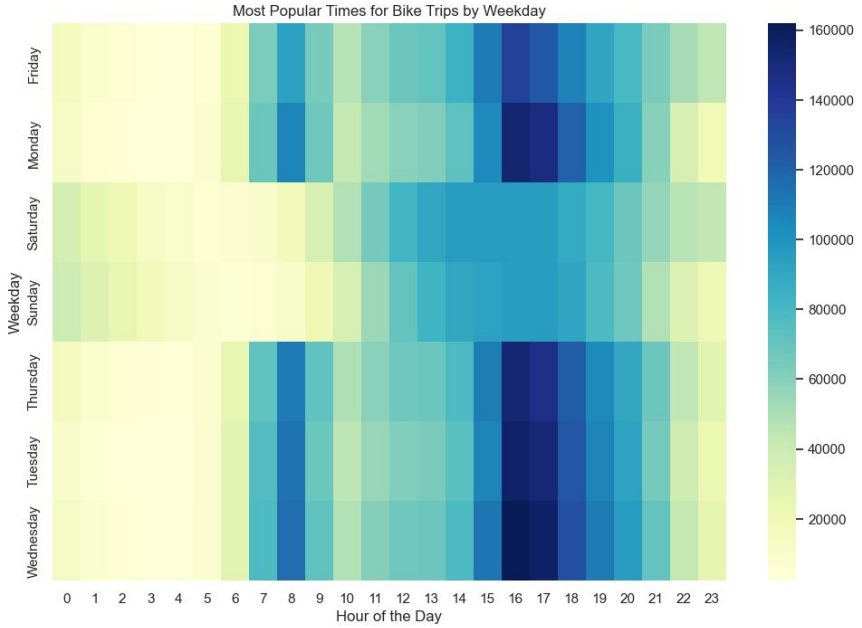
Monthly bike rides & Trip Duration



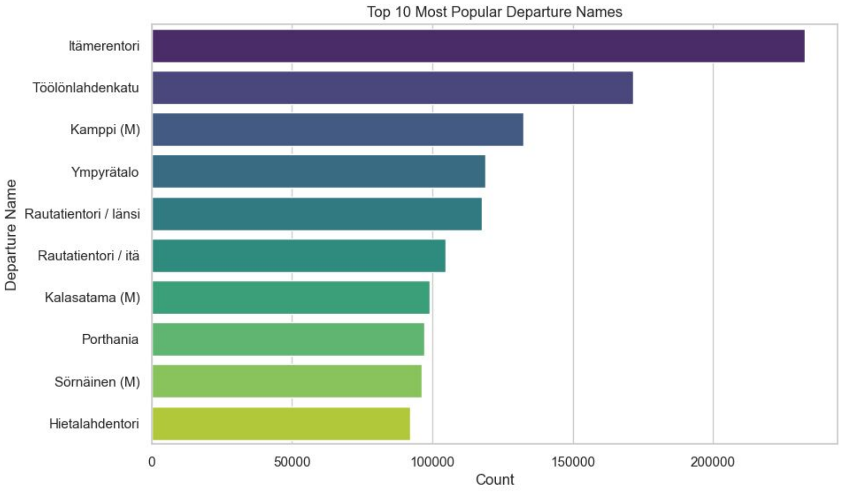


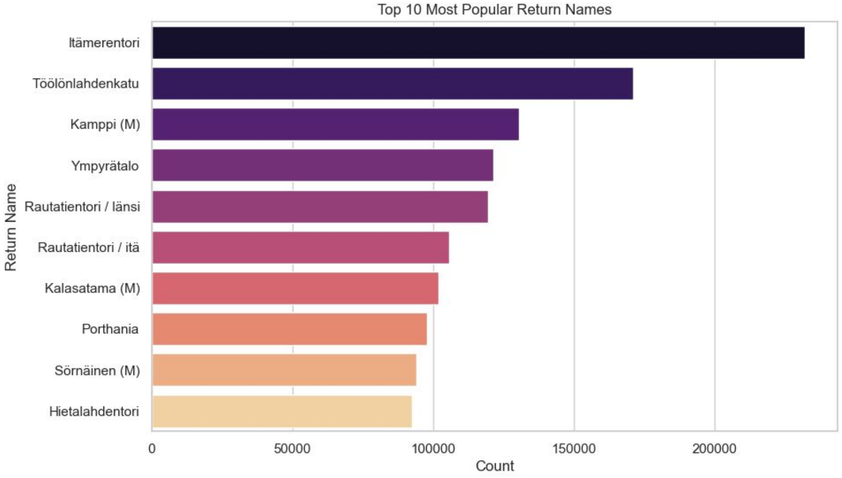
Most popular biking hours

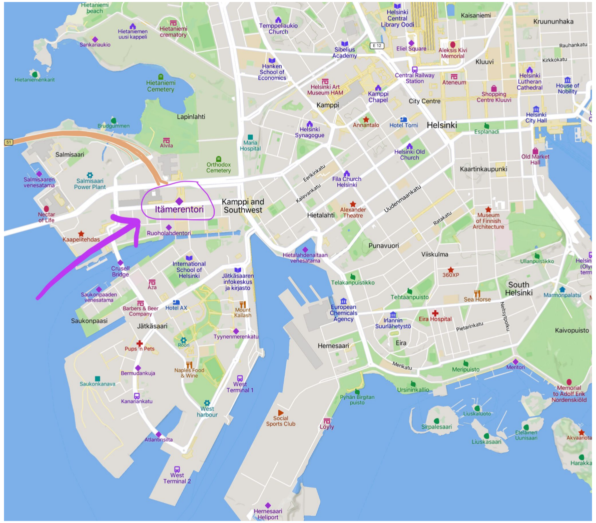


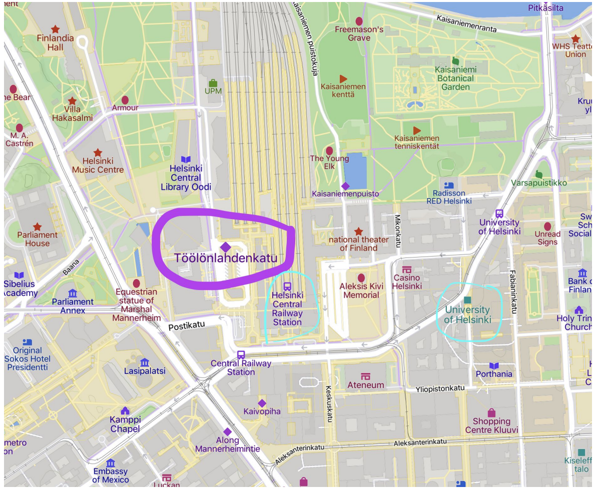


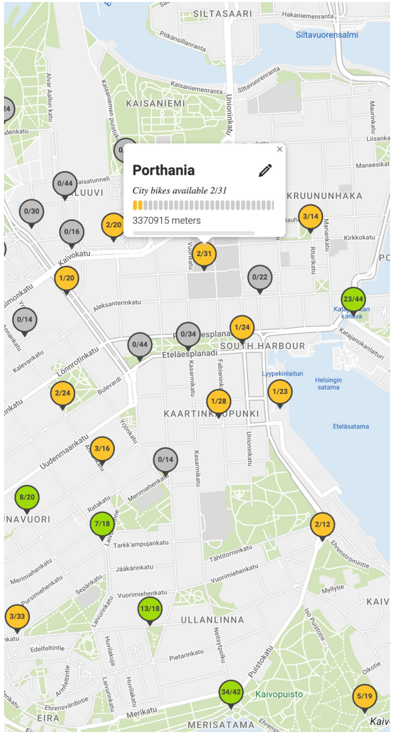
Popular Docks

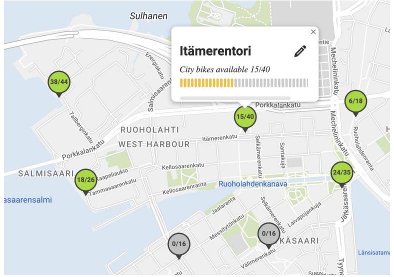




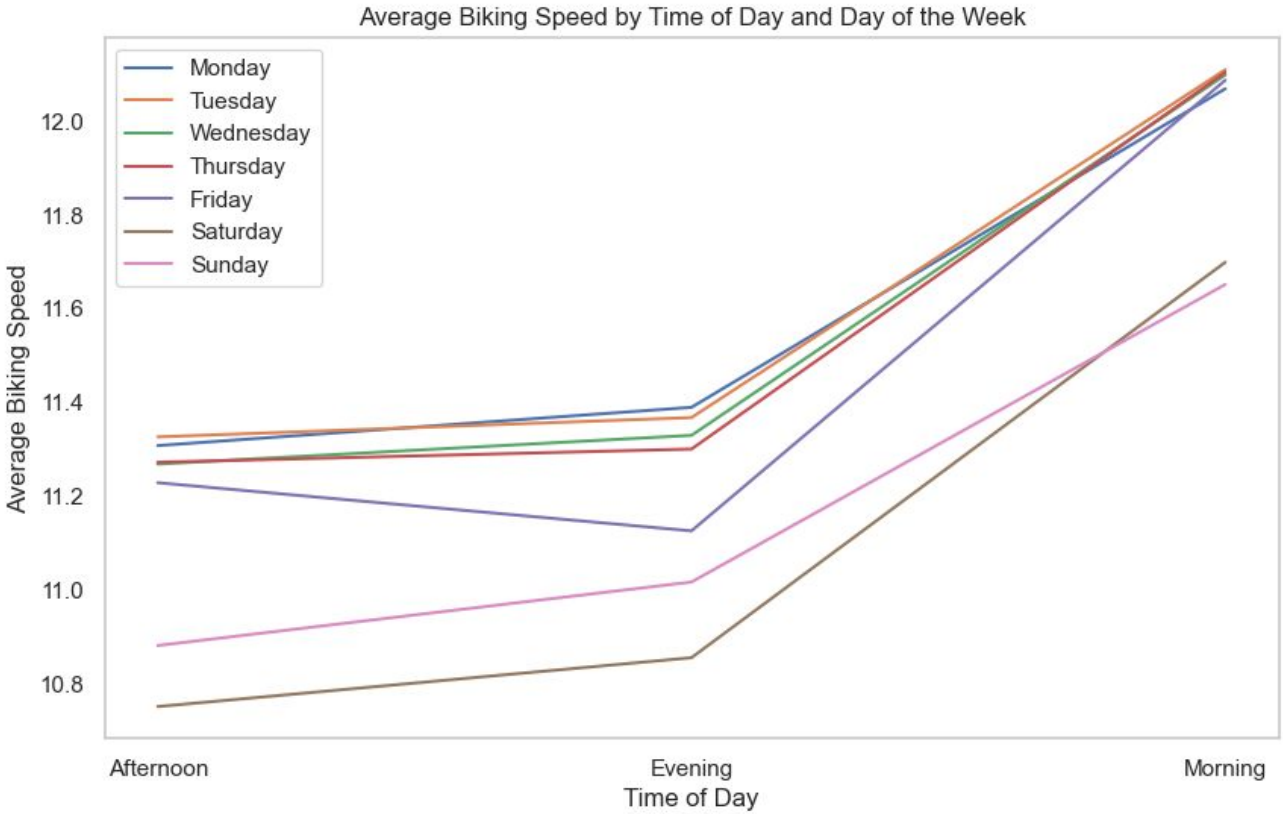






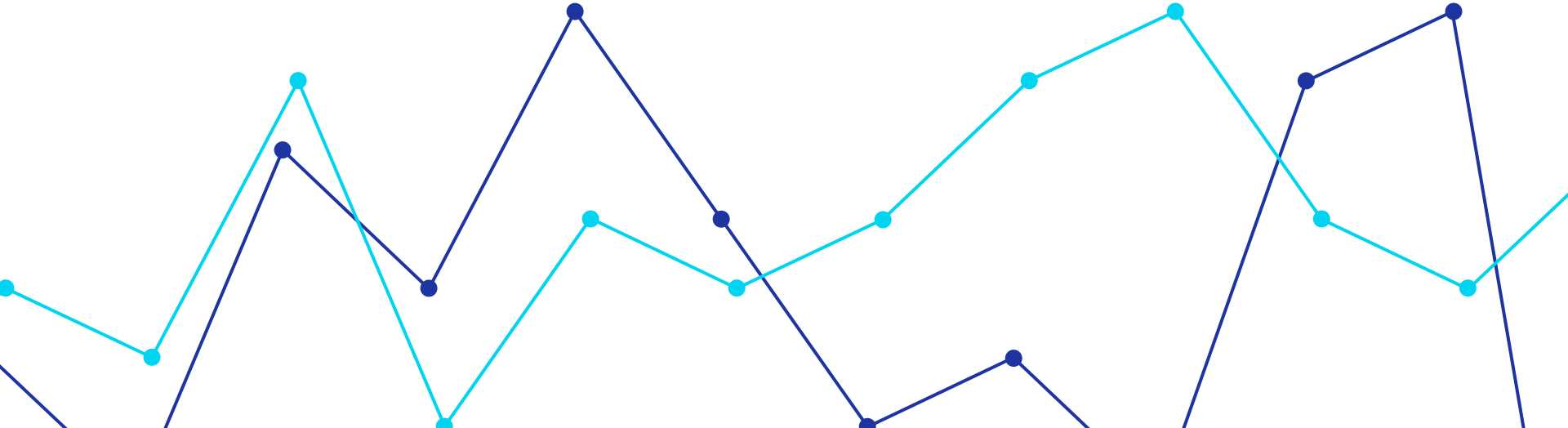


Biking speed per time of day & weekday

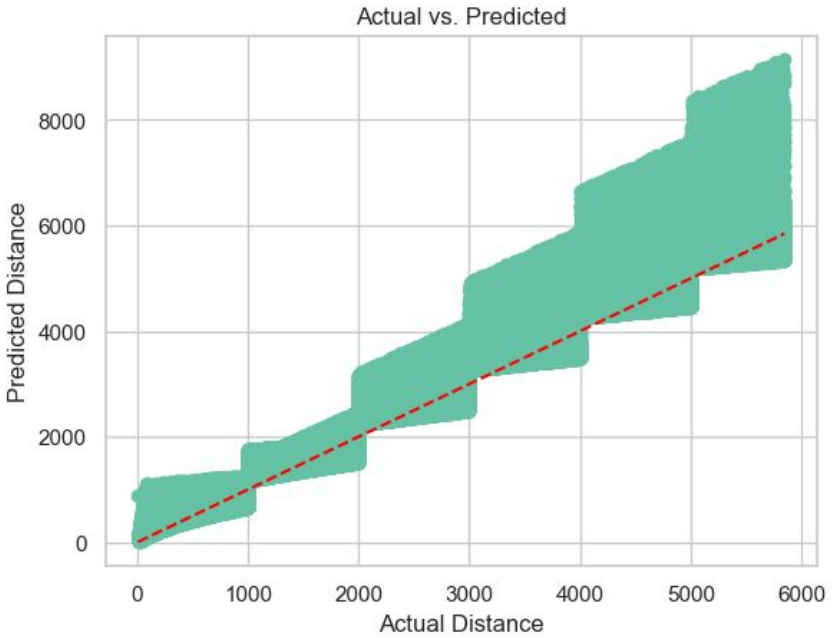


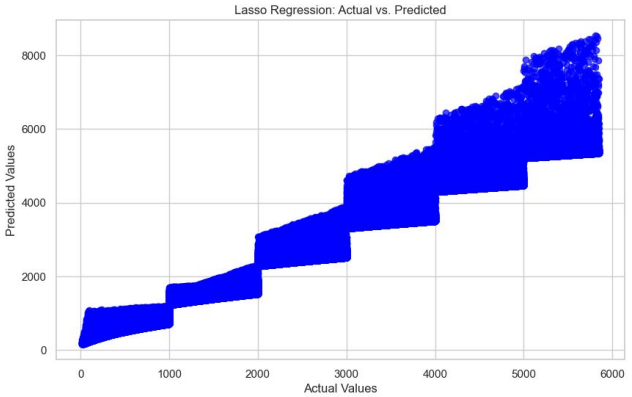
Thank you!

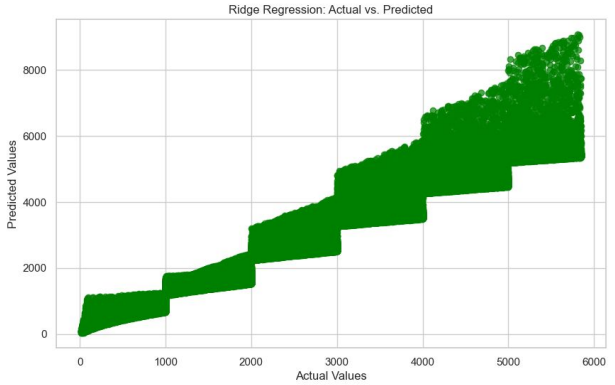
Questions?



Linear regression - Predicting distance (m)







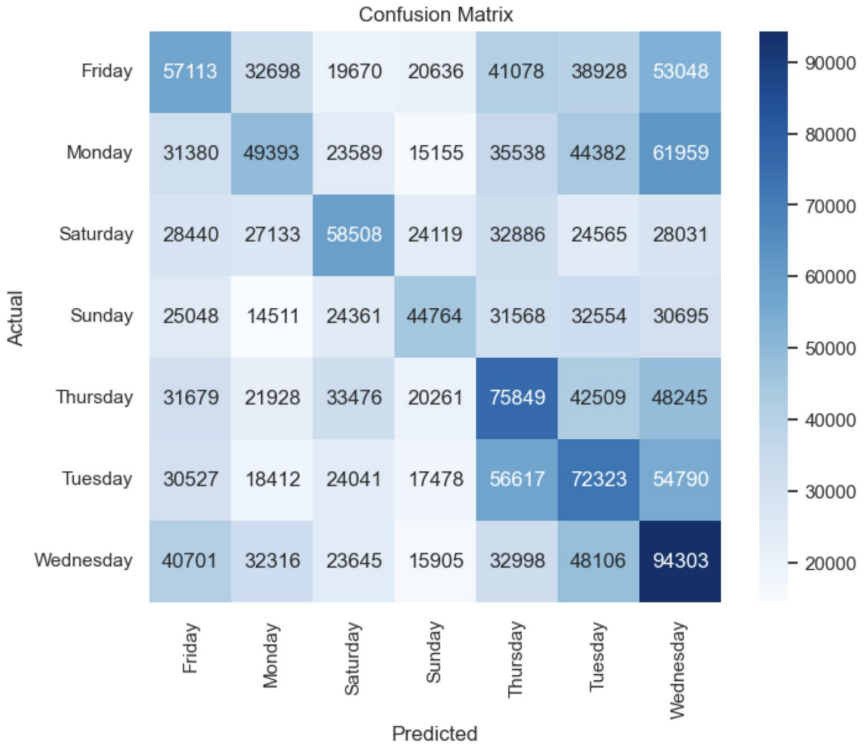

```
R-squared: 0.9687
Mean Squared Error: 51149.9066
Root Mean Squared Error: 226.1635
```

```
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
```

Multinomial Logistic Regression - Predict day of the week

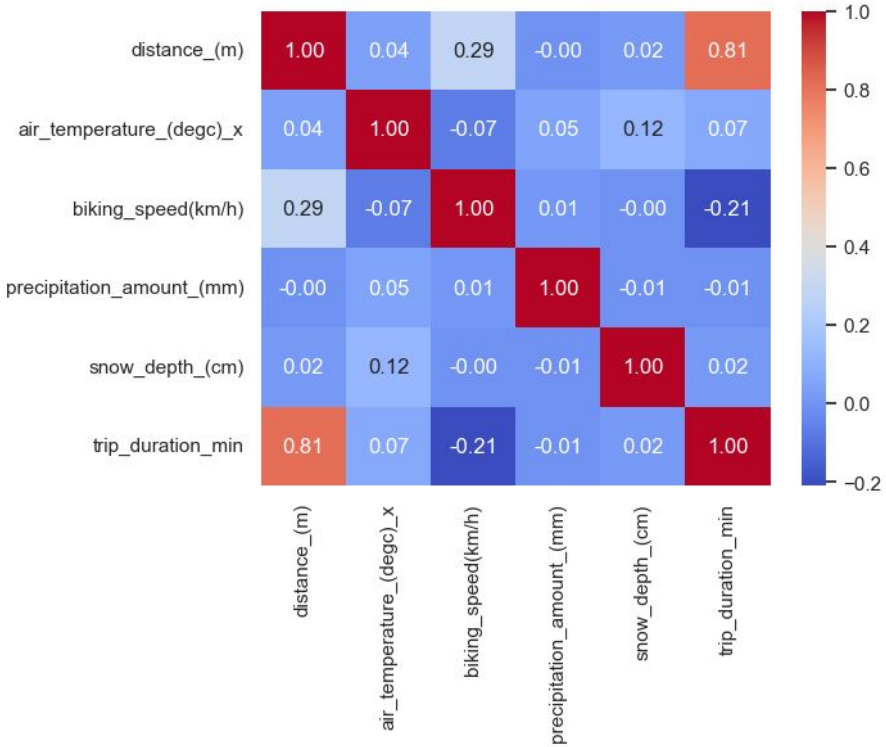
- 25% accurate for predicting day of the week

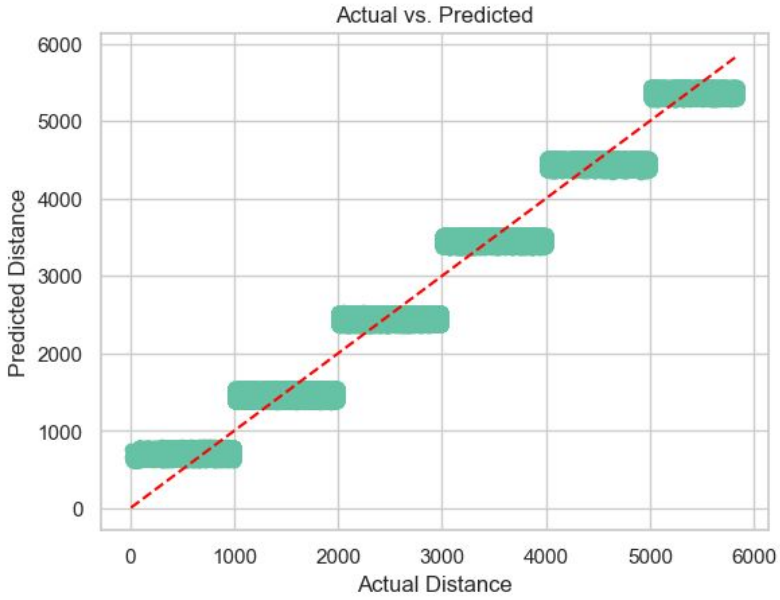



```
LogisticRegression
LogisticRegression(max_iter=1000, multi_class='multinomial')
```

```
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
```

New try based on multicollinearity






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