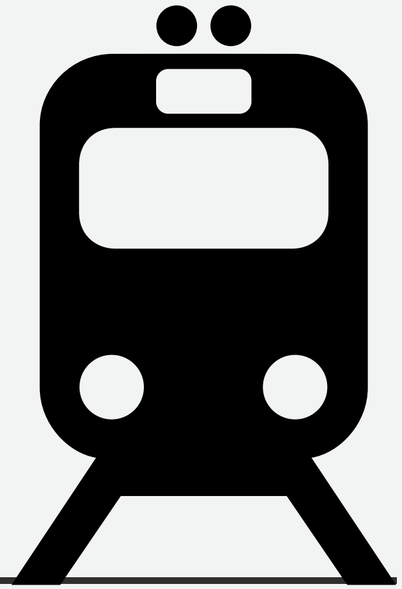
ANDREEA STROIA
HRITIKA KATHURIA
HALA ALBAHLOUL

# AIR QUALITY

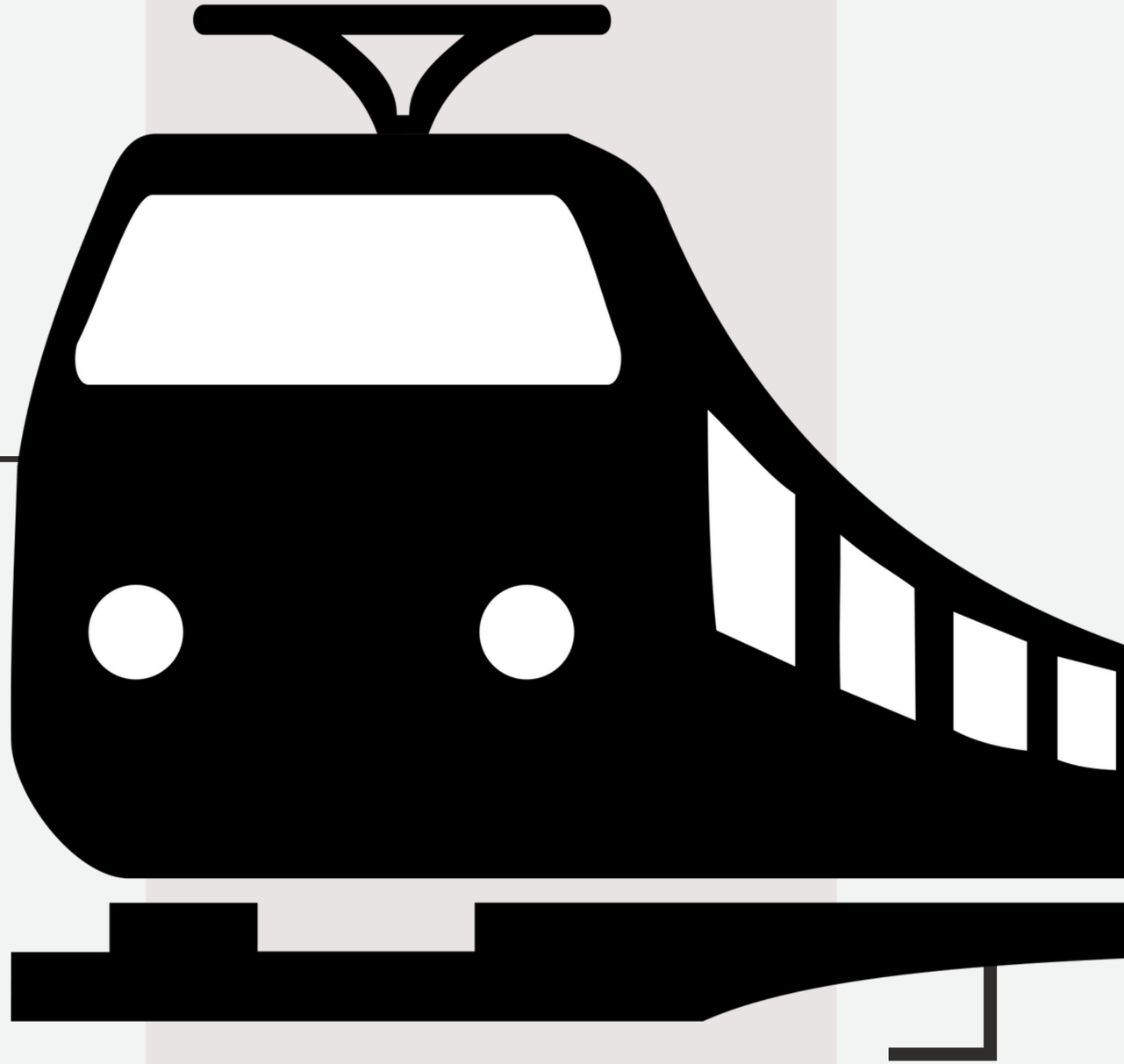


Metro stations : Chatelet & Auber & Roosvelt

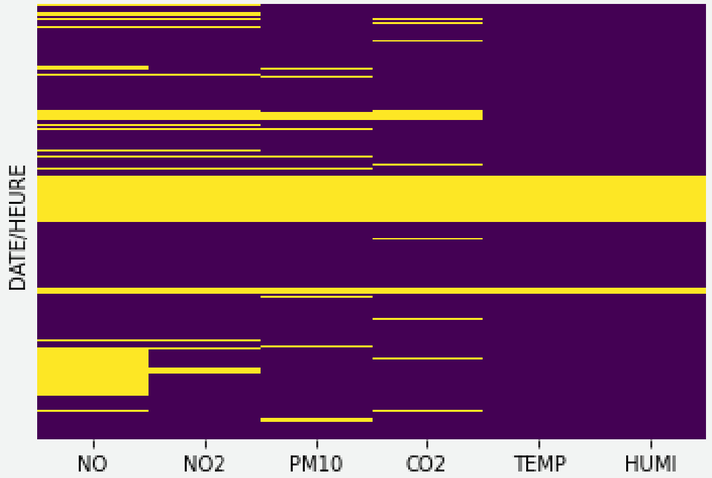
# Overview

WHAT WE DID;

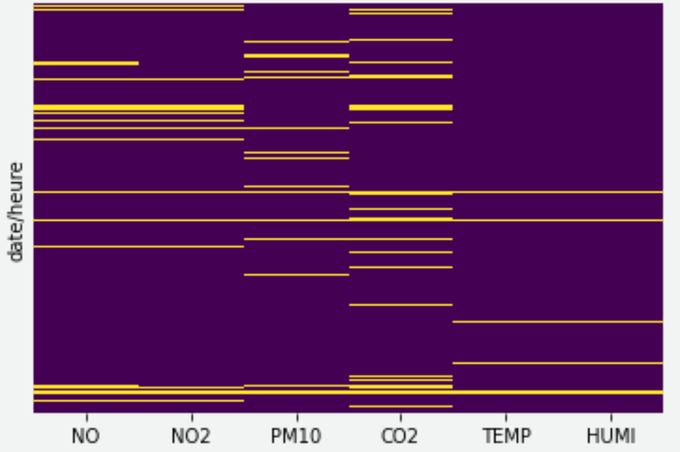
- Missing values
- Predicting $CO_2$ for Chatelet
- Predicting $NO_2$ for Chatelet
- Comparing models



# Missing values

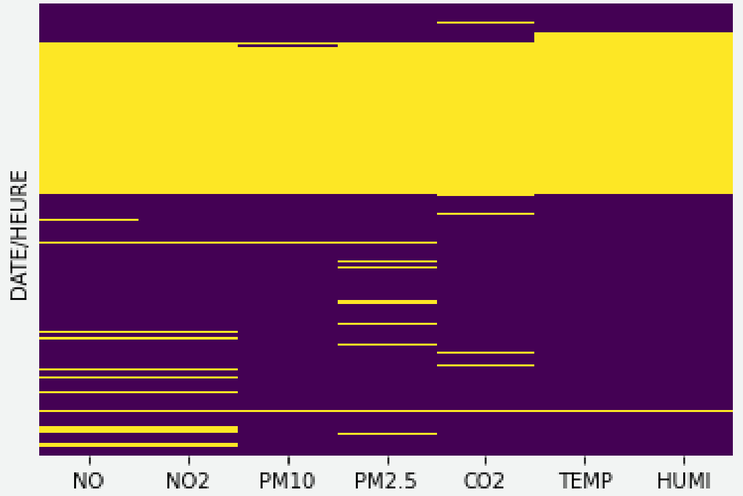



Chatelet




Roosevelt

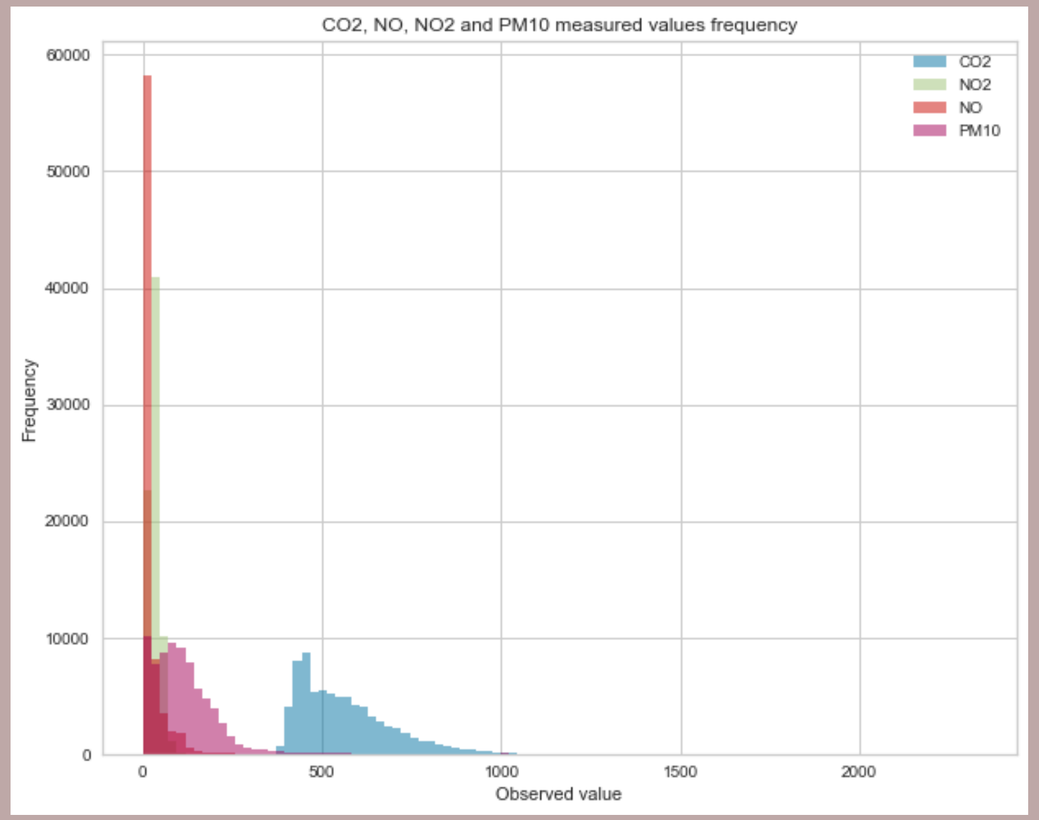
- Drop rows will all NAs
- Replace ND
- Replace , < and >
- BFill values
- Convert to float

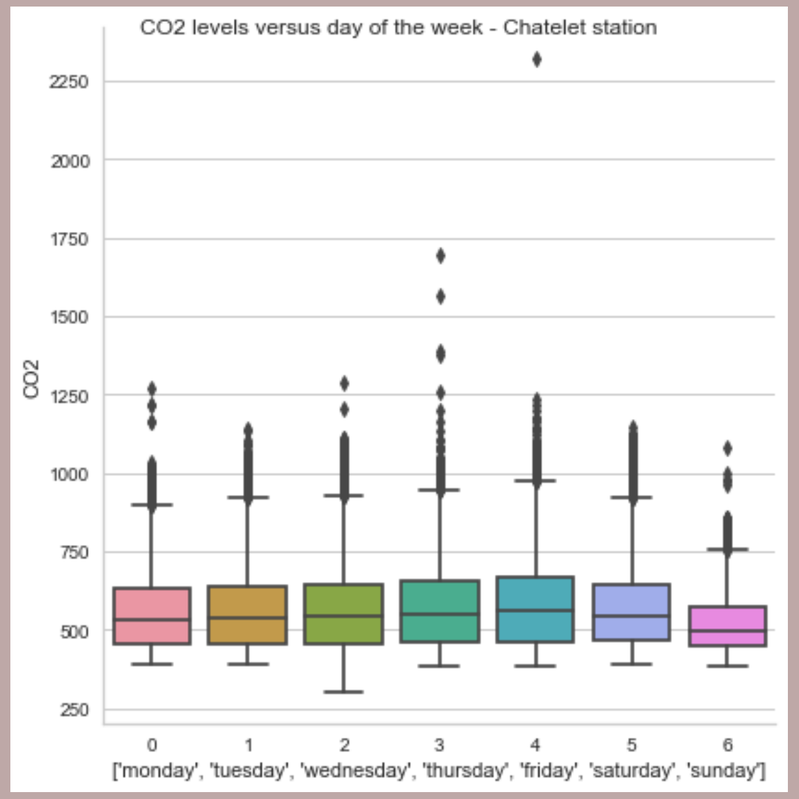



Auber

# Exploratory Data Analysis

Value range : 2013- 2016

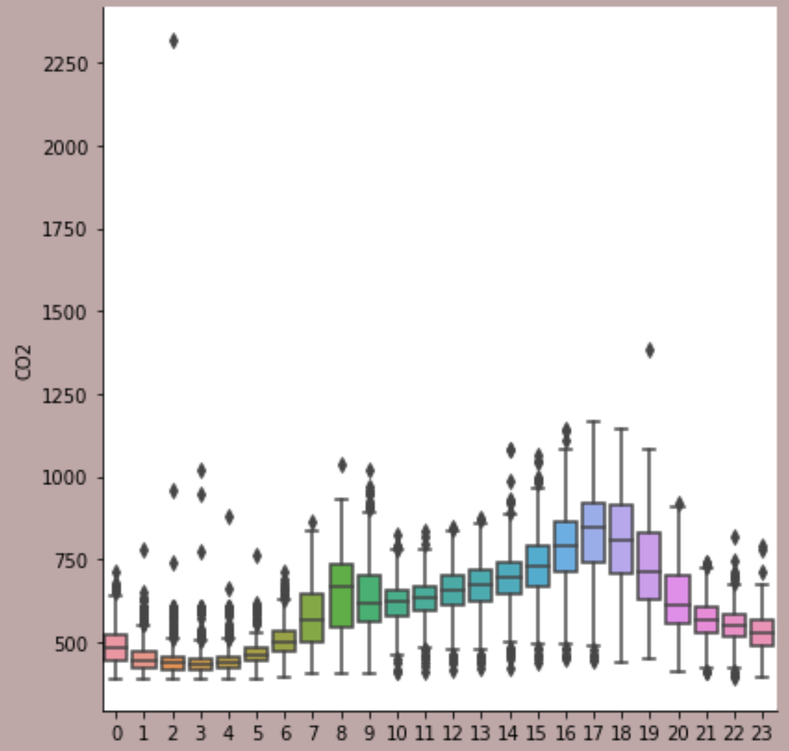


# Predicting CO2 from time of day, temperature and humidity

## -CHATELET-

Value range : 2013- 2016

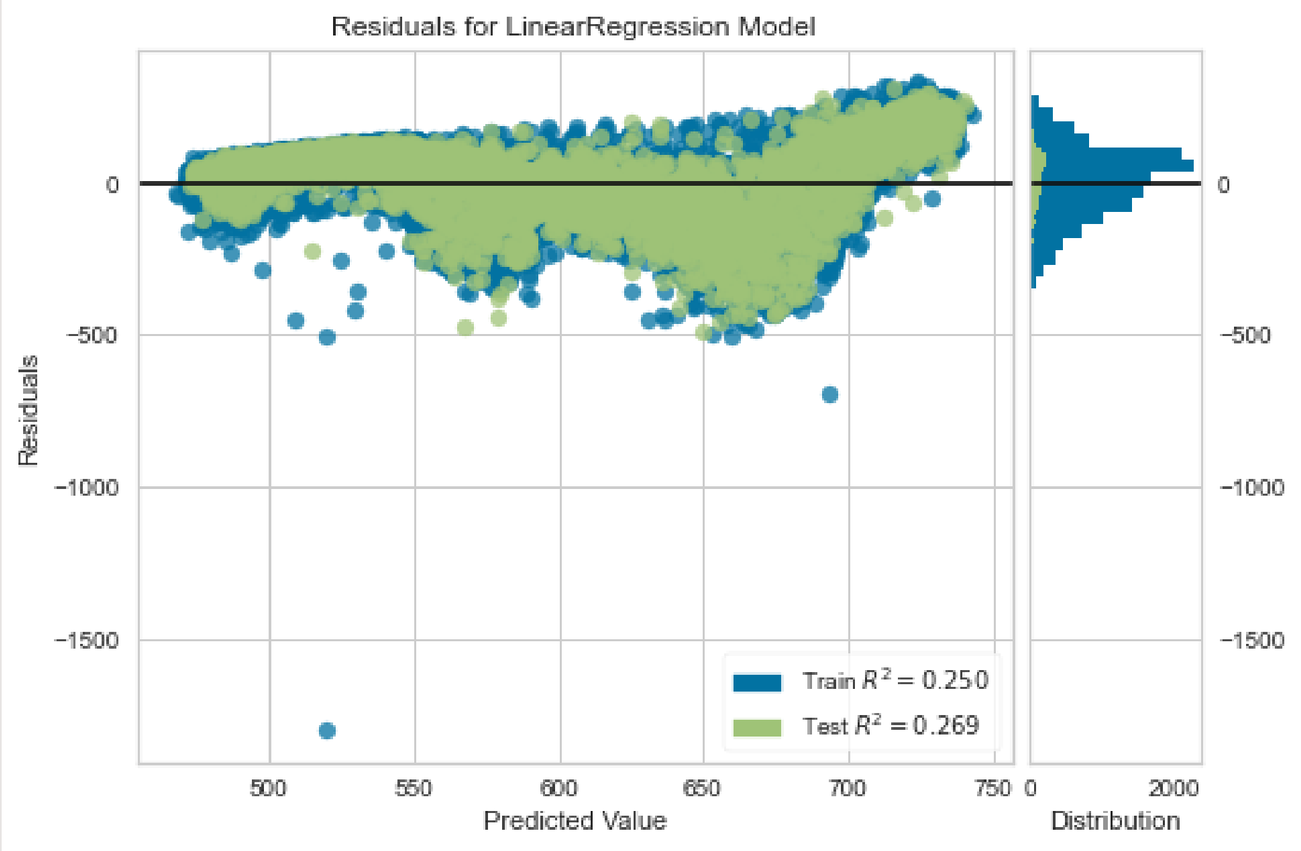

hours

# Predictions

Predictions for 2016-2022

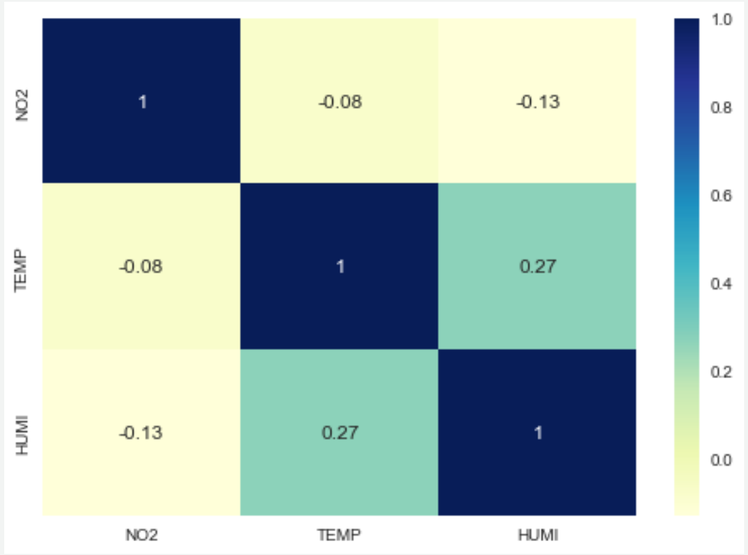
```
[[497.7235442 ],
 [740.12767956],
 [728.37994881],
 ...,
 [529.83469281],
 [519.00347581],
 [509.05179563]])
```

- The Mean Absolute Error: 97.74
- The Median Absolute Error: 81.36

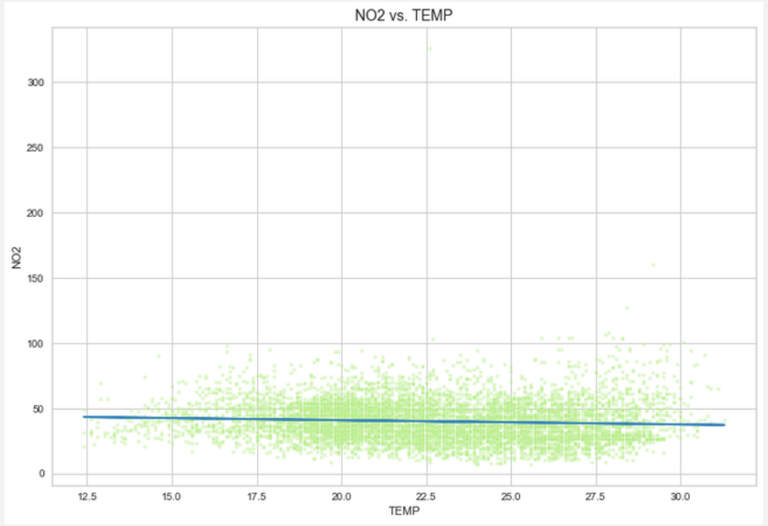


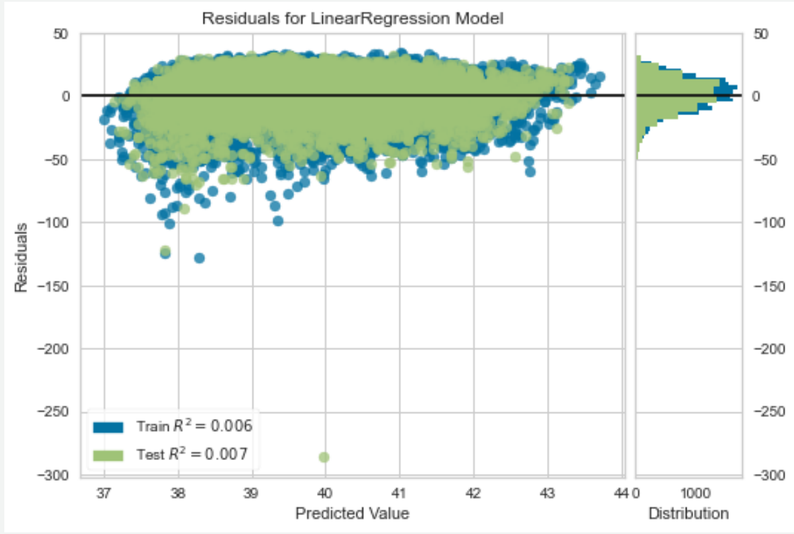
# Linear Regression

## NO2 vs TEMP




Heatmap

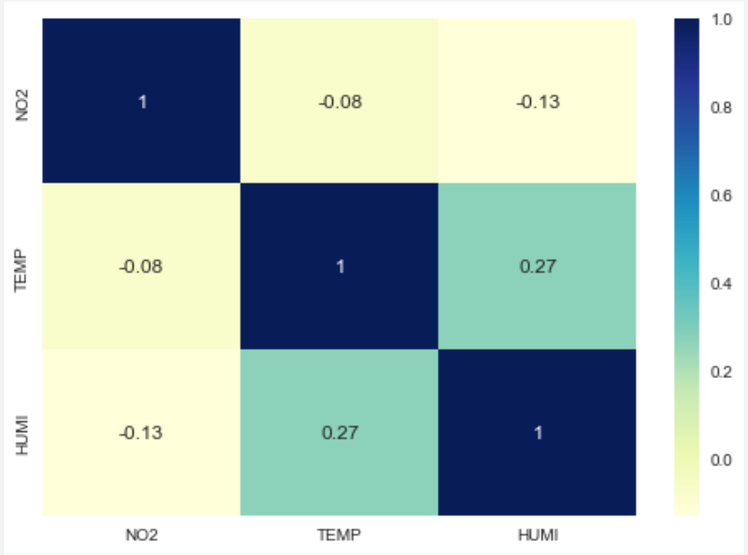


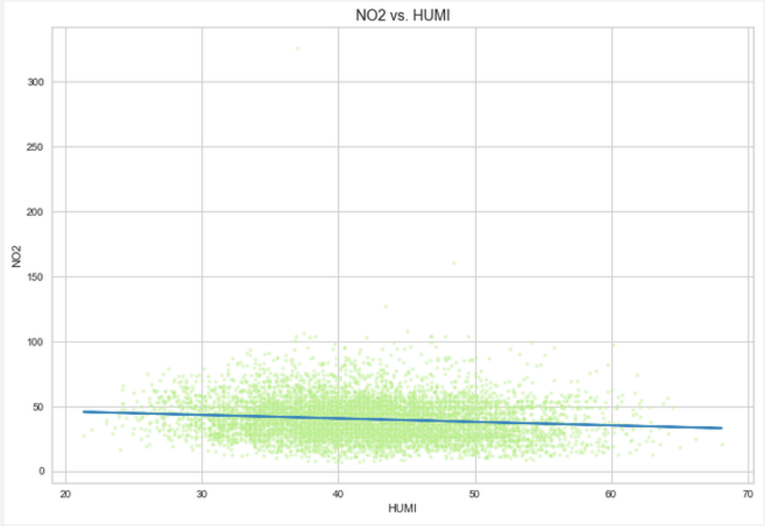

```
Slope: [[-0.32476106]]
Intercept: [47.30677617]
Mean squared error: 225.26
Root mean squared error: 15.01
R2 score:  0.007003440753828305
```

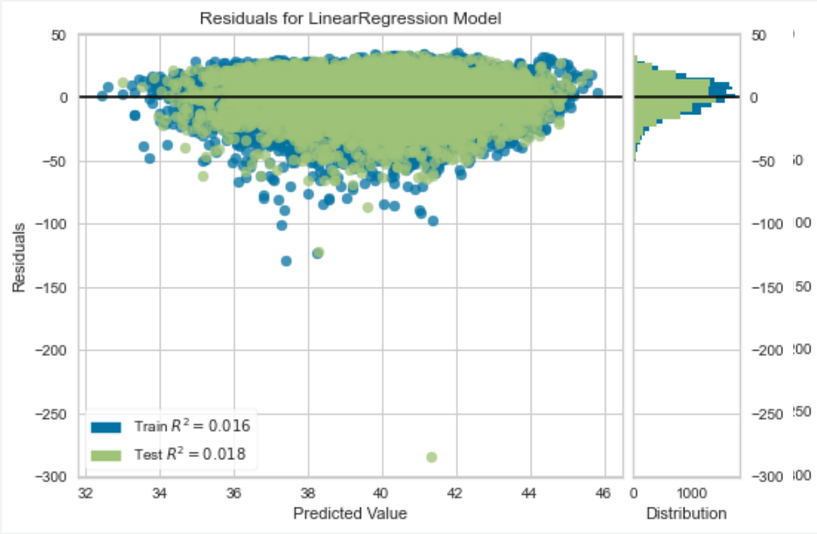
# Linear Regression

## NO2 vs HUMI



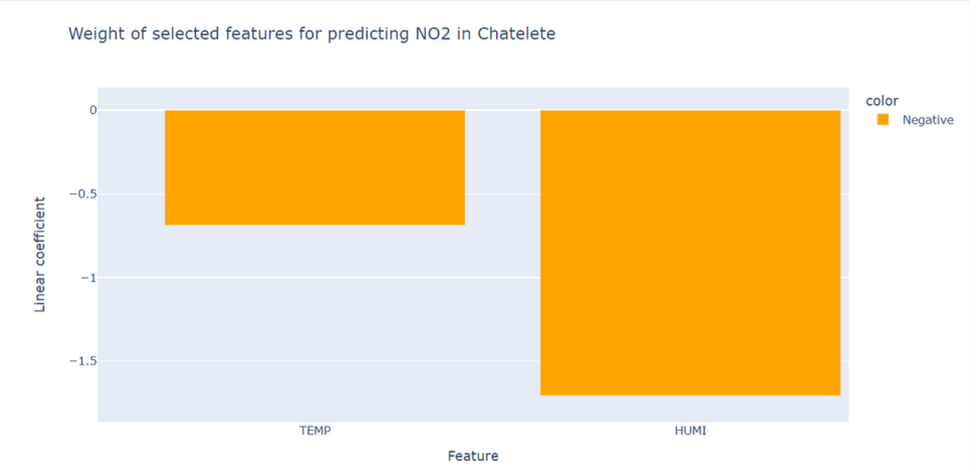

Heatmap

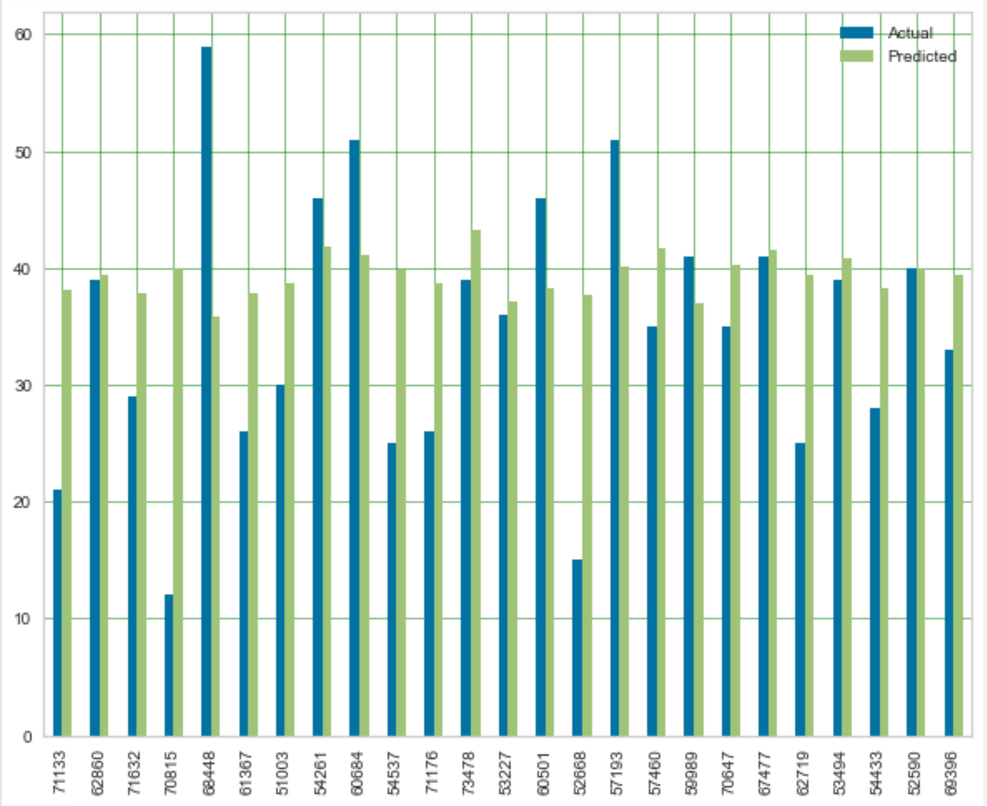



```
Slope: [[-0.26709828]]
Intercept: [51.2163539]
Mean squared error: 222.85
Root mean squared error: 14.93
R2 score:  0.017605094473889027
```

# Predict NO/NO2/PM10 levels in a selected station from past values, temperature and humidity.



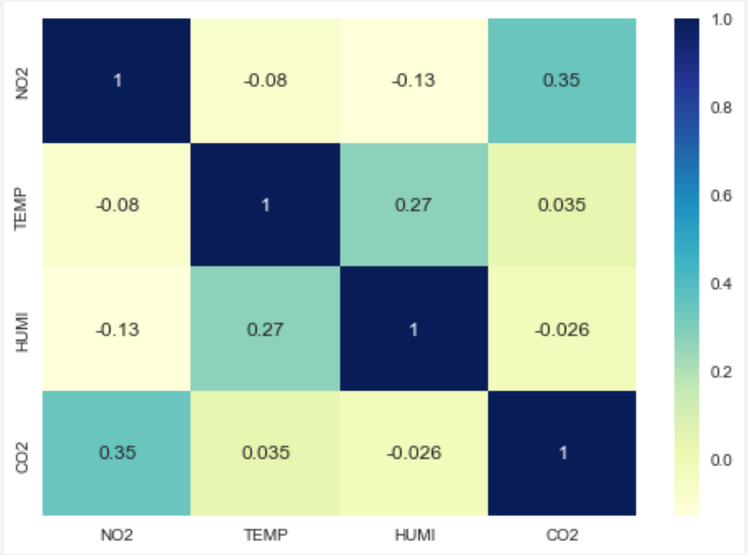




```
Mean squared error: 222.28
Root mean squared error: 14.91
Intercept: 39.97
R2 Score: 0.02
```

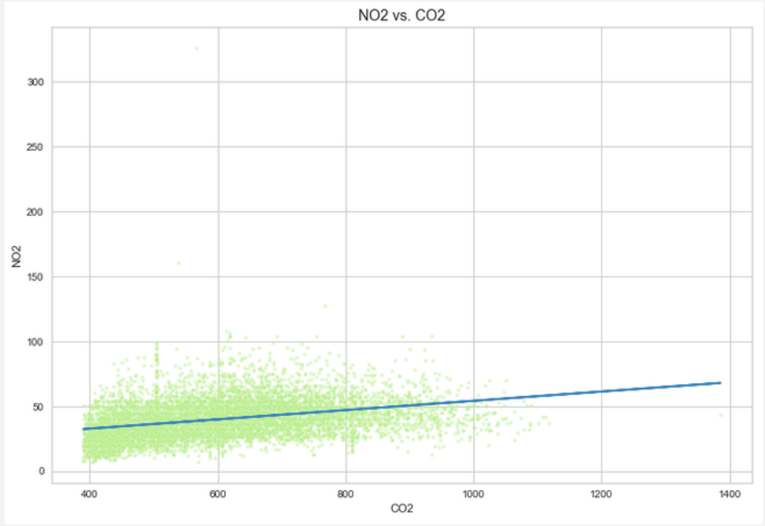
```
The Explained Variance: 0.02
The Mean Absolute Error: 11.37
The Median Absolute Error: 9.70
```

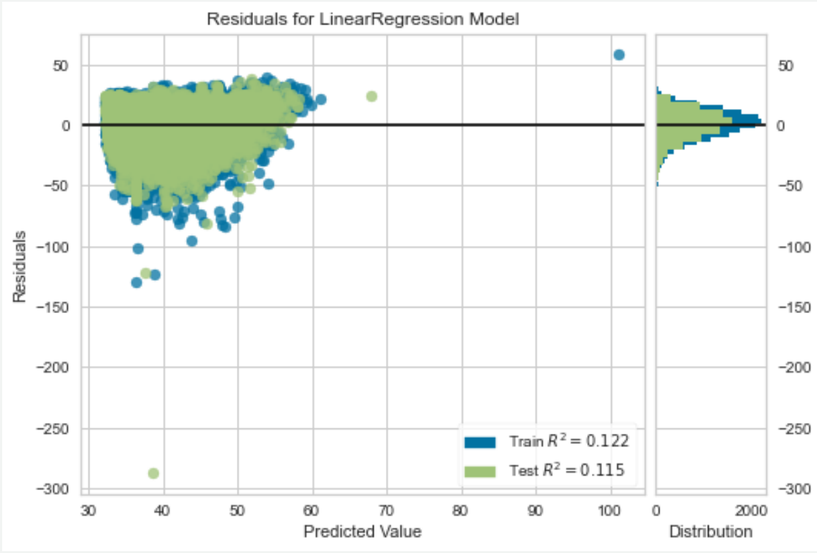
# Linear Regression

## Using CO2 now




Heatmap




```
Slope: [[0.03566195]]
Intercept: [18.40550976]
Mean squared error: 200.73
Root mean squared error: 14.17
R2 score:  0.11511227855230222
```

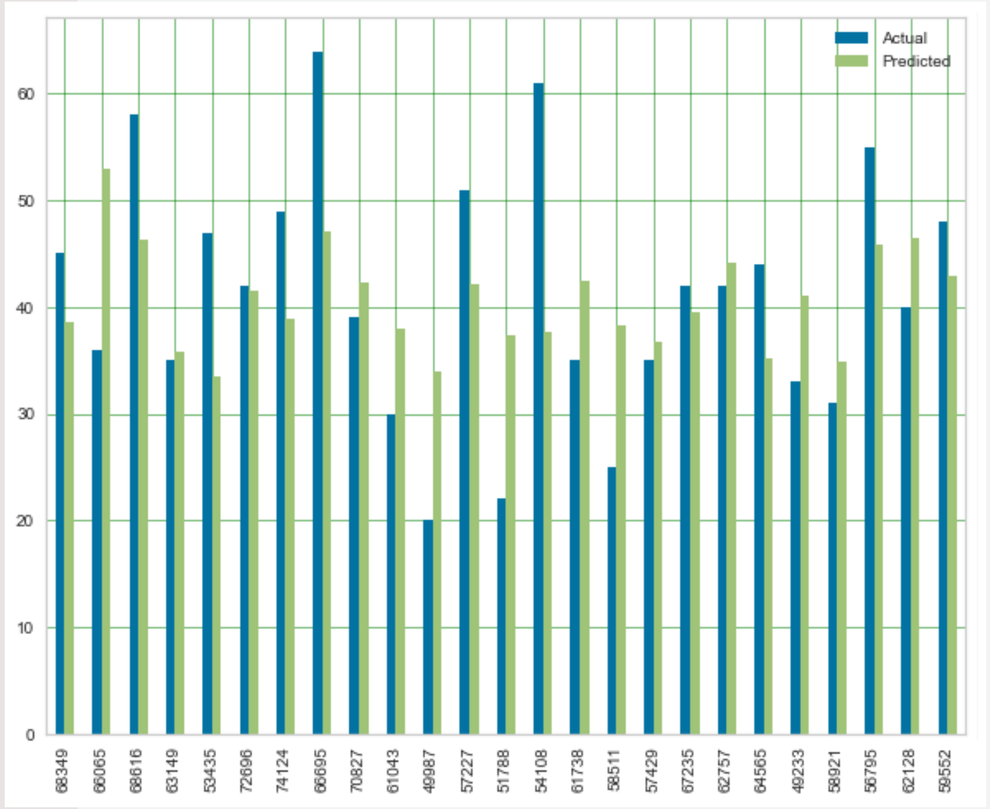
# Predict NO/NO2/PM10 levels in a selected station from past values, temperature, humidity, and CO2.

```
print(mse)
print(r2)
print(rmse)
```

```
192.60770055461018
0.1355865700941388
13.87831764136454
```

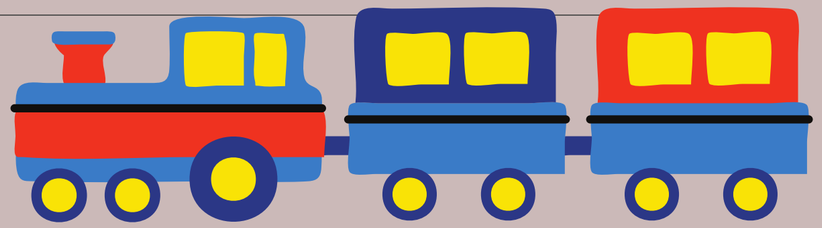
```
cross_validation(X_chatelet_no2_co2,Y_chatelet_no2_co2)
```

```
R_squared Mean Score: -0.02051225796606526
[-0.19839816 -0.02846629  0.07697044  0.0796446  -0.03231189]
```




```
The Explained Variance: 0.14
The Mean Absolute Error: 10.36
The Median Absolute Error: 8.44
```

# Comparing models



## Other models:

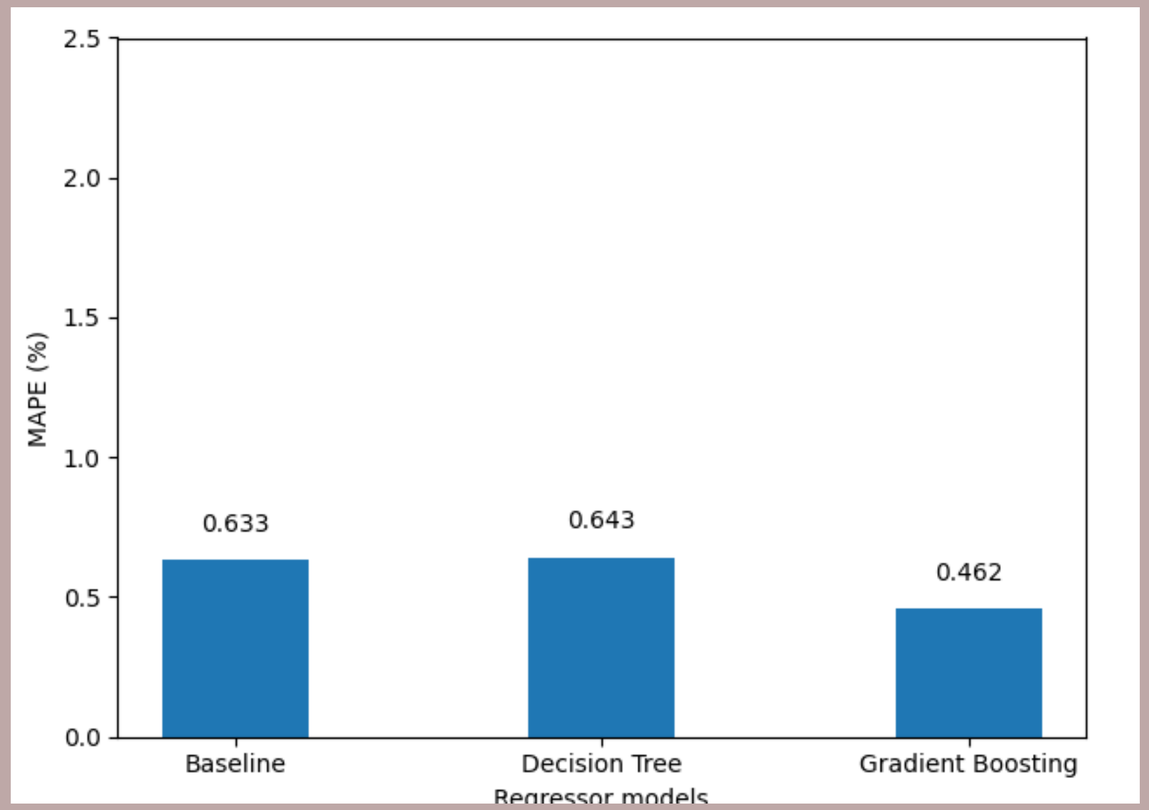
- Gradient boosting
- Decision Tree
- Baseline model

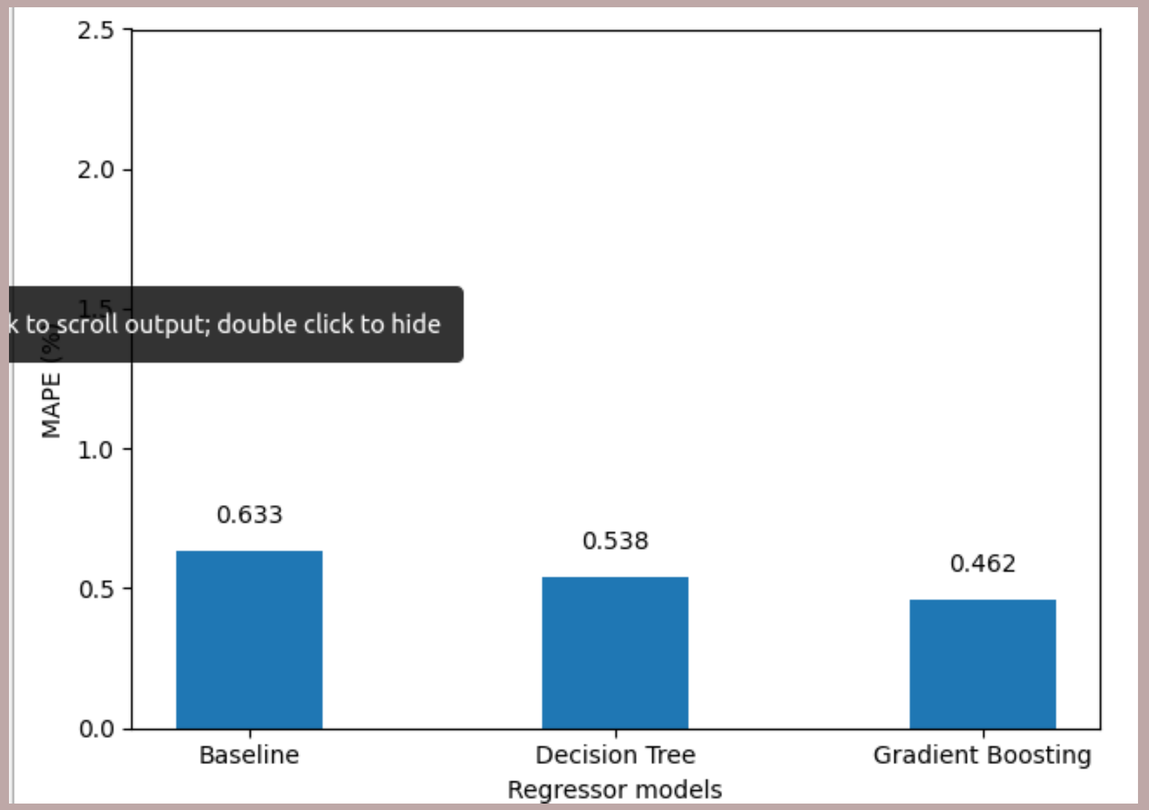
| CO2 | TEMP | HUMI | x_1 | x_2 | x_3 | x_4 | x_5 | x_6 | x_7 | x_8 | x_9 | y |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 610.166667 | 19.712500 | 39.187500 | 612.041667 | 597.291667 | 550.208333 | 568.875000 | 513.333333 | 558.125000 | 610.500000 | 644.791667 | 640.958333 | 645.739130 |
| 599.833333 | 19.587500 | 38.445833 | 610.166667 | 612.041667 | 597.291667 | 550.208333 | 568.875000 | 513.333333 | 558.125000 | 610.500000 | 644.791667 | 640.958333 |
| 500.000000 | 18.375000 | 37.145833 | 599.833333 | 610.166667 | 612.041667 | 597.291667 | 550.208333 | 568.875000 | 513.333333 | 558.125000 | 610.500000 | 644.791667 |
| 532.958333 | 17.087500 | 30.562500 | 500.000000 | 599.833333 | 610.166667 | 612.041667 | 597.291667 | 550.208333 | 568.875000 | 513.333333 | 558.125000 | 610.500000 |
| 517.666667 | 17.212500 | 35.891667 | 532.958333 | 500.000000 | 599.833333 | 610.166667 | 612.041667 | 597.291667 | 550.208333 | 568.875000 | 513.333333 | 558.125000 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 576.541667 | 17.104167 | 40.616667 | 547.791667 | 483.041667 | 532.666667 | 605.458333 | 605.958333 | 561.791667 | 569.000000 | 550.625000 | 549.375000 | 616.916667 |
| 589.041667 | 16.754167 | 38.708333 | 576.541667 | 547.791667 | 483.041667 | 532.666667 | 605.458333 | 605.958333 | 561.791667 | 569.000000 | 550.625000 | 549.375000 |
| 620.541667 | 16.075000 | 37.383333 | 589.041667 | 576.541667 | 547.791667 | 483.041667 | 532.666667 | 605.458333 | 605.958333 | 561.791667 | 569.000000 | 550.625000 |
| 671.916667 | 15.487500 | 37.483333 | 620.541667 | 589.041667 | 576.541667 | 547.791667 | 483.041667 | 532.666667 | 605.458333 | 605.958333 | 561.791667 | 569.000000 |
| 535.000000 | 14.900000 | 34.400000 | 671.916667 | 620.541667 | 589.041667 | 576.541667 | 547.791667 | 483.041667 | 532.666667 | 605.458333 | 605.958333 | 561.791667 |

# Comparing models

## -CHATELET-

mean absolute percentage error (MAPE)



# Thank you for your attention!