Air quality in Madrid

Ines Vilar

Final Project - Data Science - AllWomen

April 2020

kaggle

Search

Home
Compete
Data
Notebooks
Discuss
Courses
More

Data Tasks Kernels (29) Discussion (5) Activity Metadata

Download (570 MB) New Notebook

Usability 7.9 **License** Other (specified in description) **Tags** natural and physical sciences, time series, green living and environmental issues, environment, pollution

Context

In recent years, high levels of pollution in Madrid have forced the authorities to take measures to monitor and address this issue. One of the most effective and controversial measures was putting restrictions on the use of polluting vehicles in the city center.

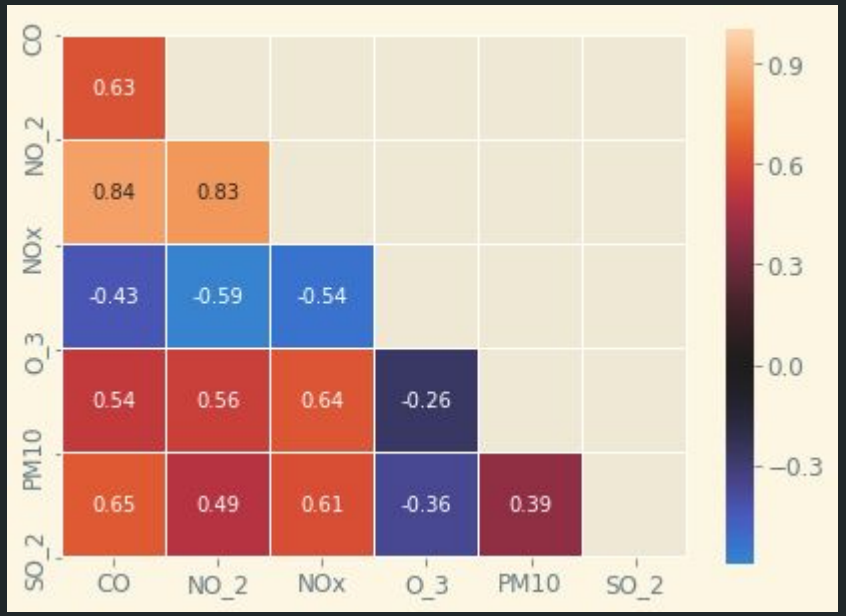
To monitor the success of these measures Madrid's City Council Open Data website, makes publicly available daily and hourly historical data of the levels of contamination from different toxic particles, for a number of active stations across the whole city, from 2001 to real time.

NO2: nitrogen dioxide level measured in µg/m^3.
Long-term exposure is a cause of chronic lung diseases, and are harmful for the vegetation.

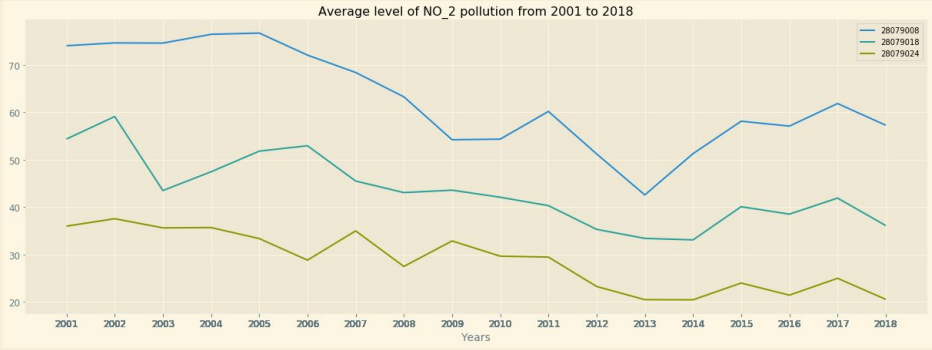
O3: ozone level measured in µg/m^3.
High levels can produce asthma, bronchitis or other chronic pulmonary diseases in sensitive groups or outdoor workers.

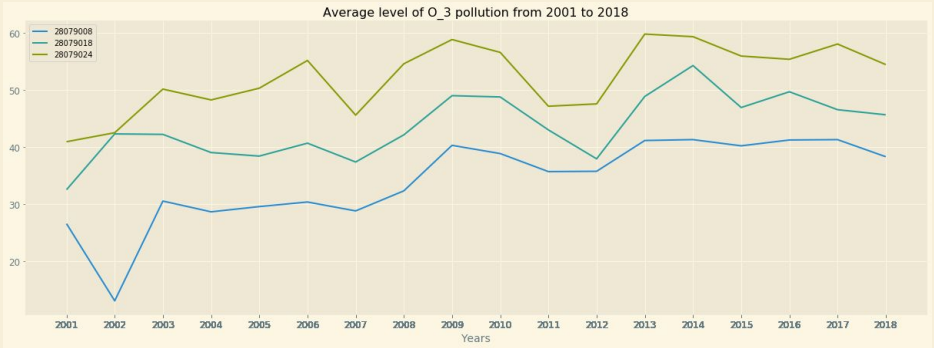
PM10: particles smaller than 10 µm measured in µg/m^3.
Even though the cannot penetrate the alveolus, they can still penetrate through the lungs and affect other organs. Long term exposure can result in lung cancer and cardiovascular complications.

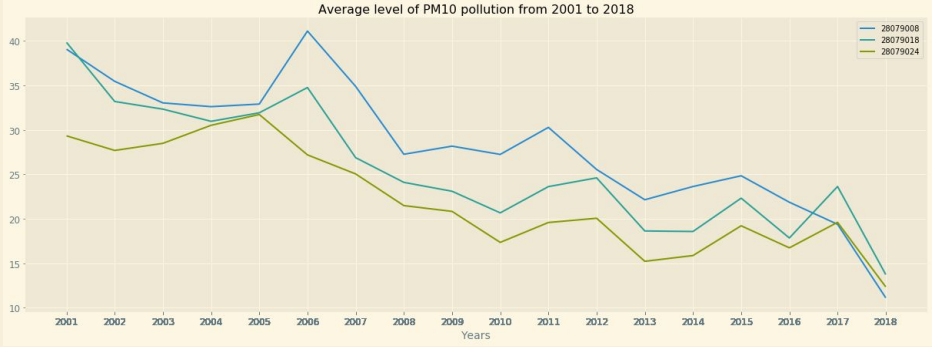
SO2: sulphur dioxide level measured in µg/m^3.
High levels of sulphur dioxide can produce irritation in the skin and membranes, and worsen asthma or heart diseases in sensitive groups.

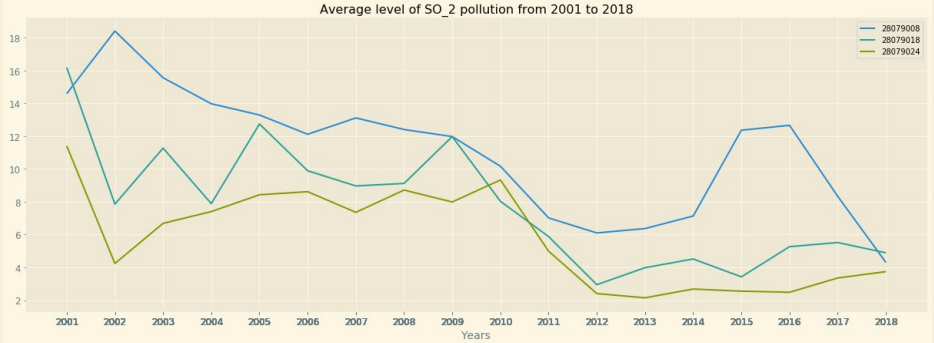


Let’s start by looking at the 3 main components in any time series analysis: trend, seasonality and noise...

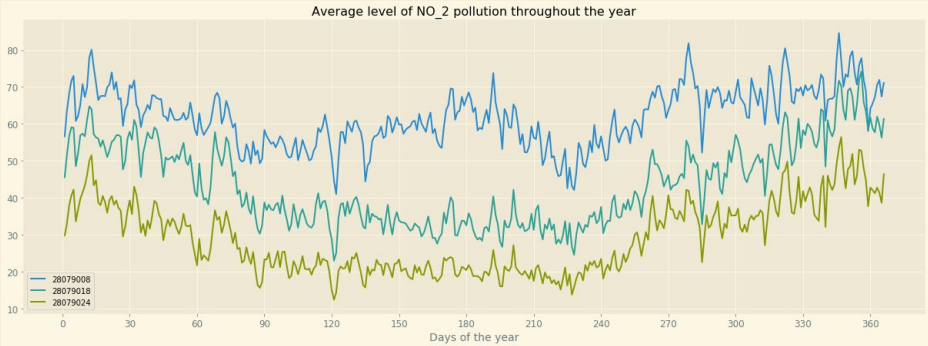


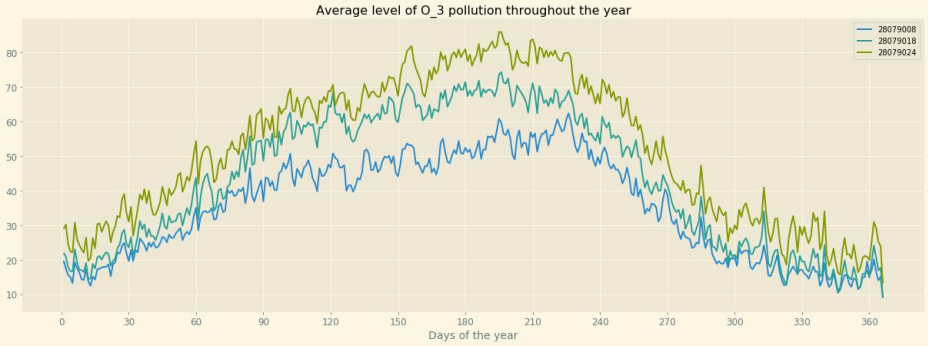


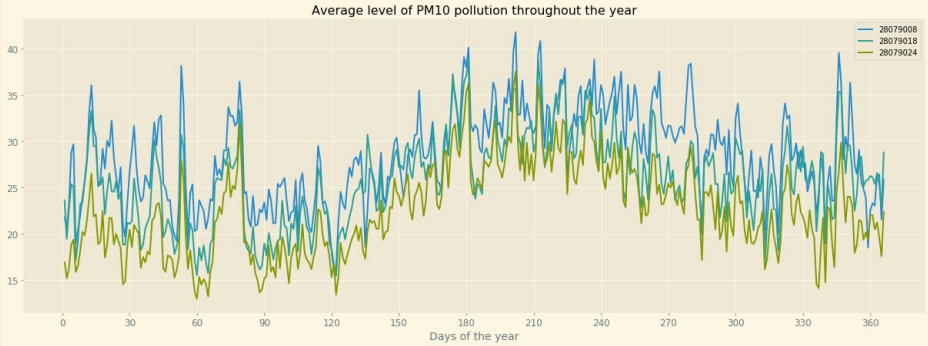


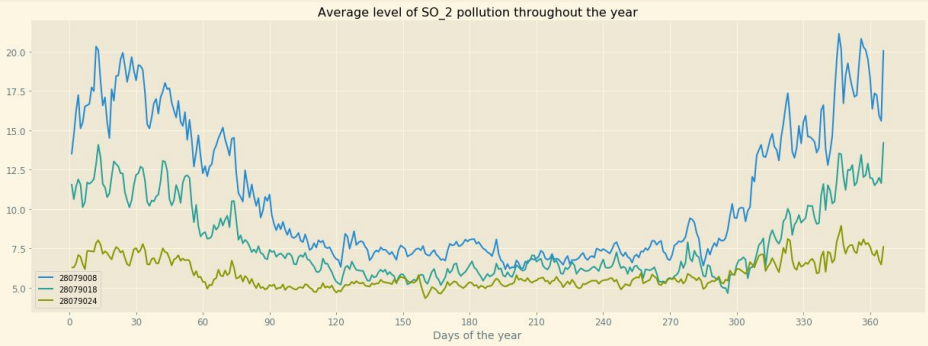


Do you see any significant patterns in the data throughout the year? What differences can you spot?

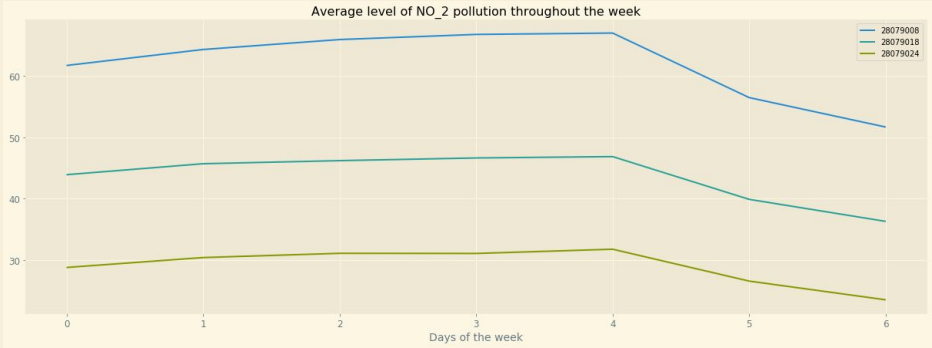


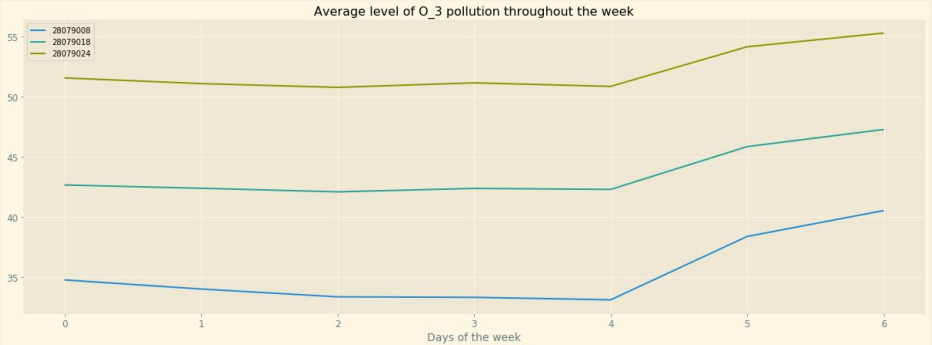


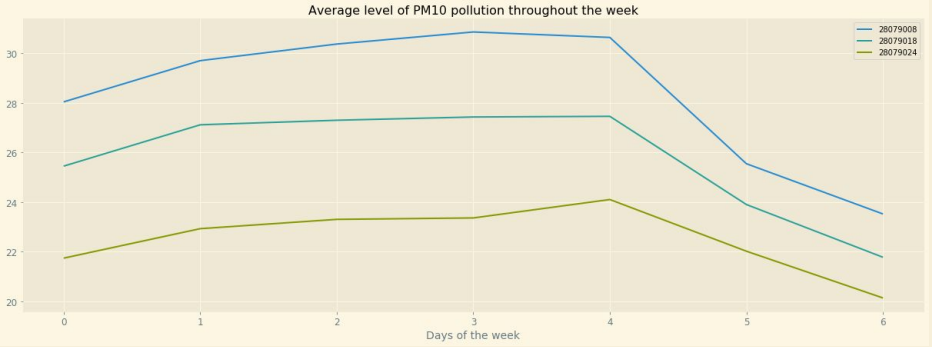


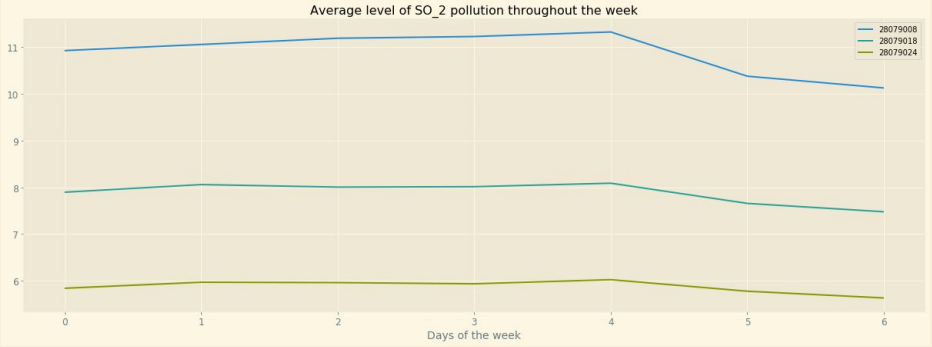


What about throughout the week? Can you see the difference between weekdays and weekends?

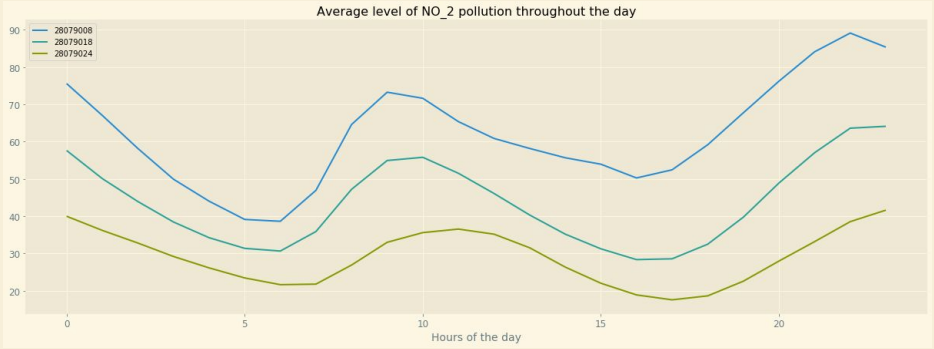


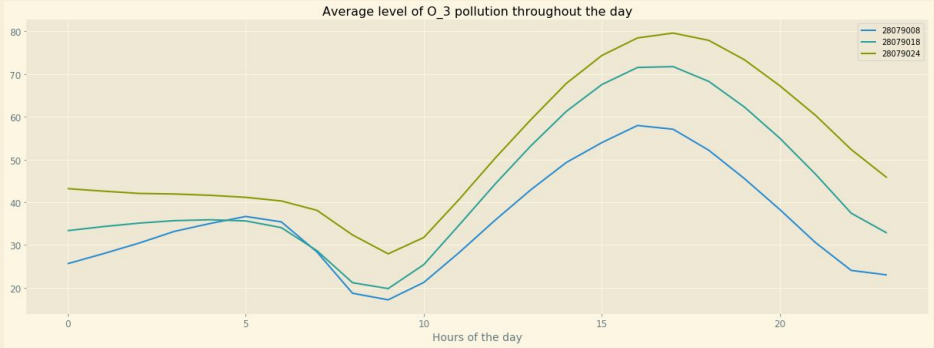


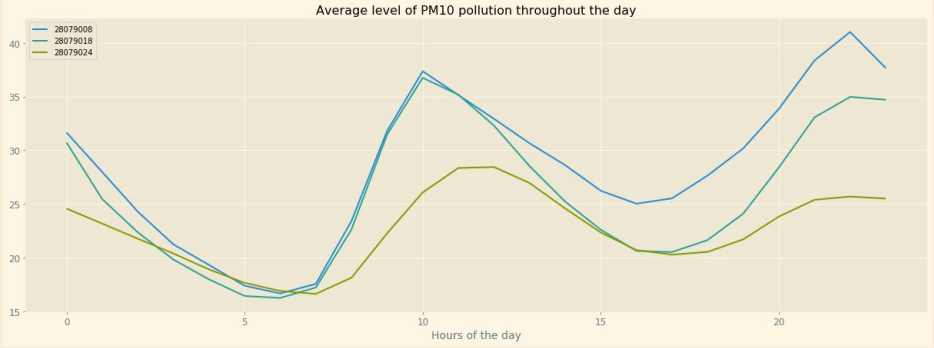


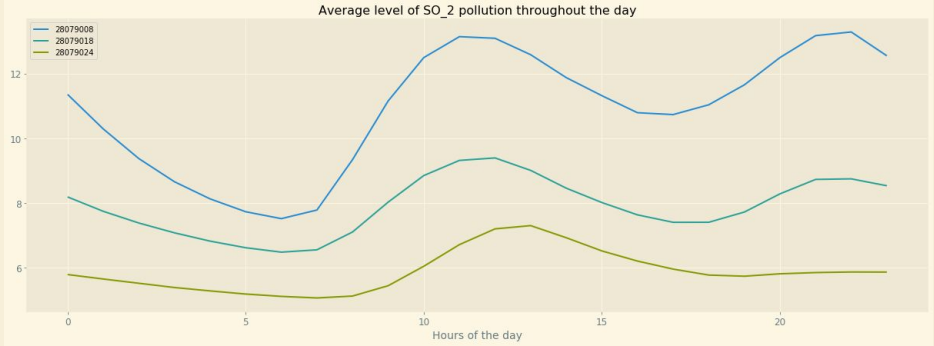


Can you spot the intra day patterns? What are the peak hours for pollution in the day?









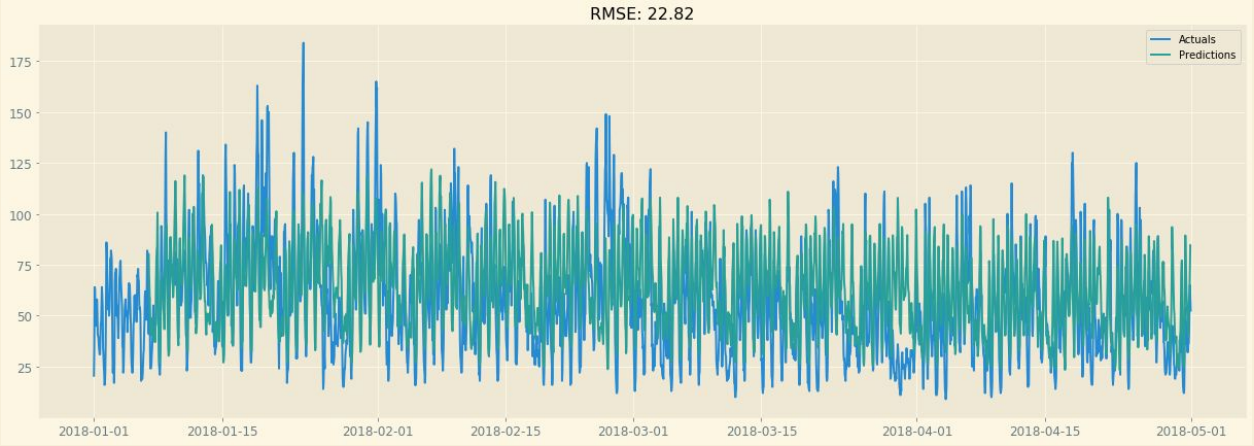
PREDICTIONS

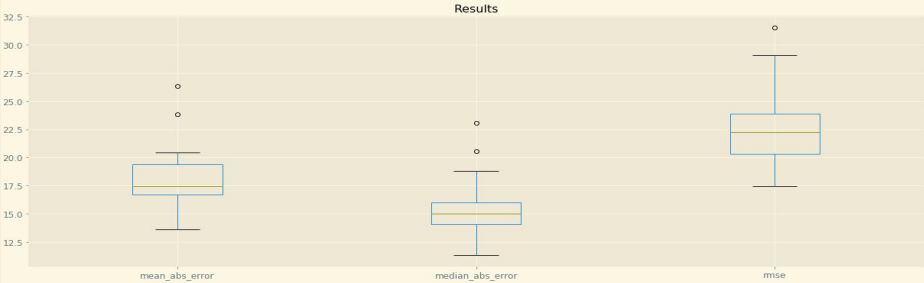
Let’s look at the rolling weekly predictions for one particle and one station, using an ARIMA model.

The particle is:
Nitrogen Dioxide (NO2)

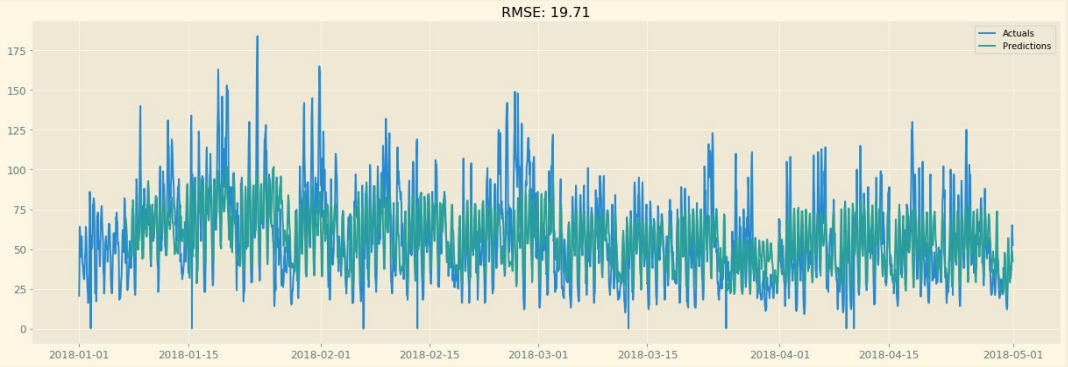
The location is:
Station ID: 28079008
Retiro/ Escuelas Aguirre

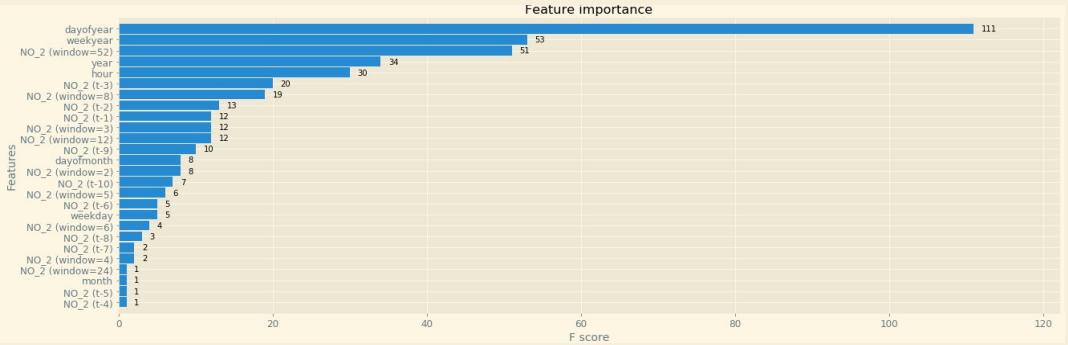
The model is:
Auto ARIMA

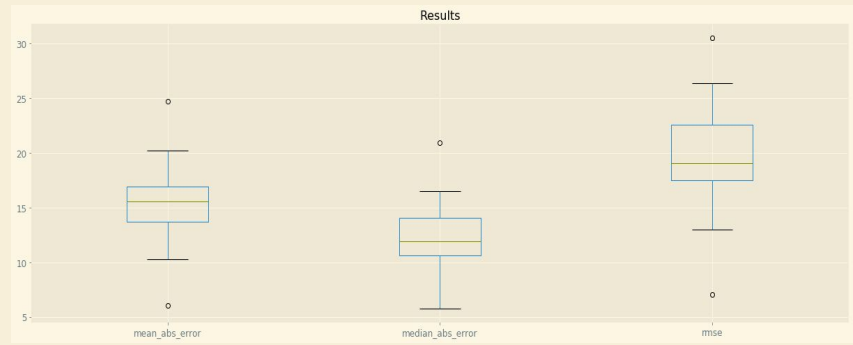




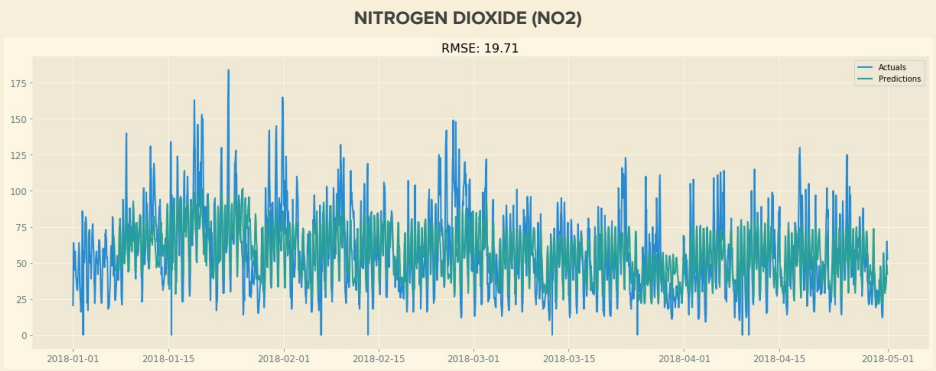
Now look at the rolling weekly predictions for the same particle and station, using an XGBoost model.

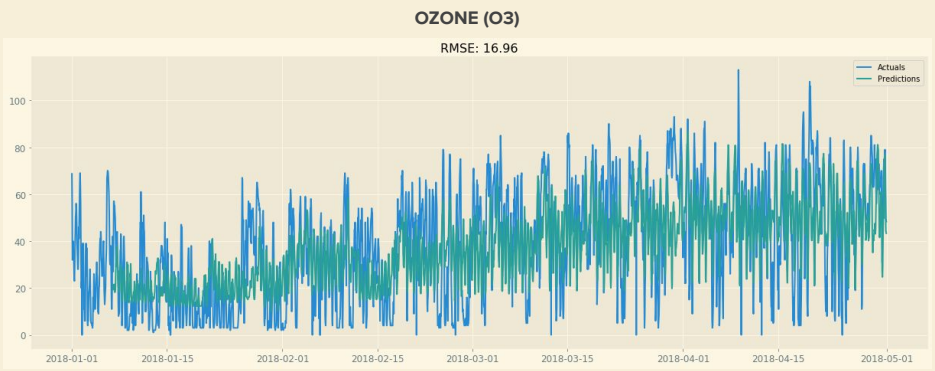


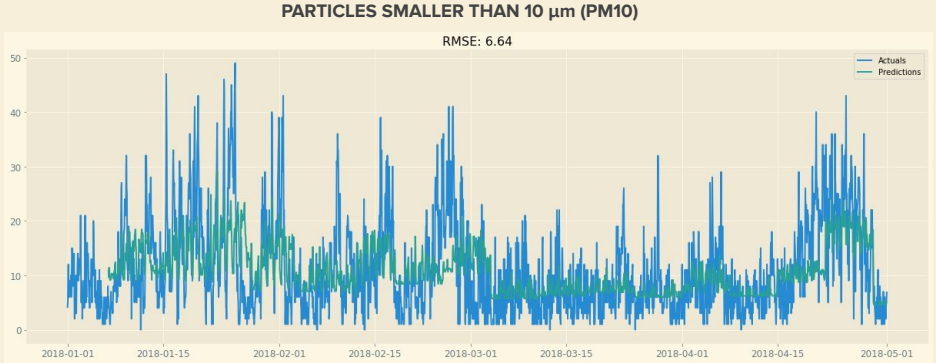


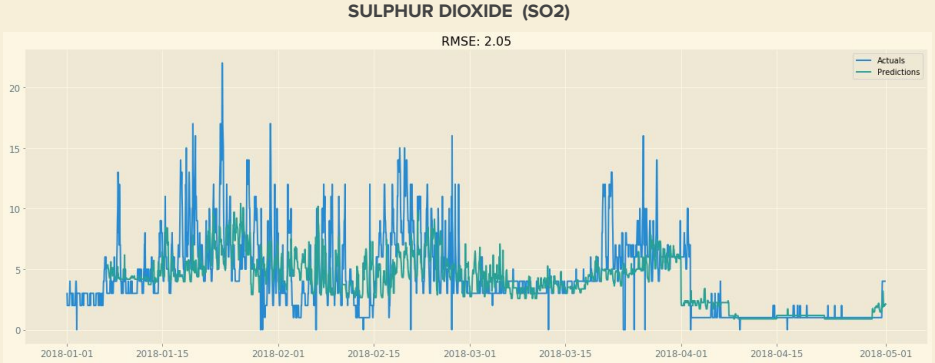


Since the XGBoost model gave us the best results for NO2, let's see how it performed for other particles.

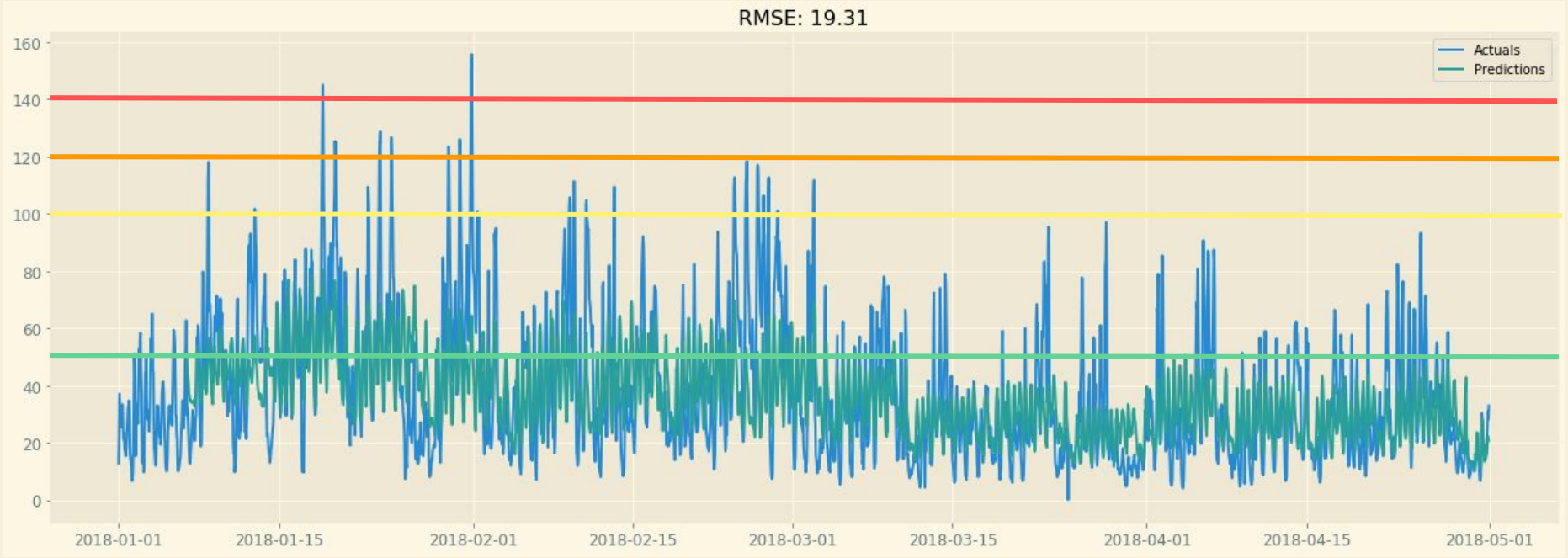








Finally, let’s compare the average of the 3 stations with the WHO threasholds for NO2 contamination.



3 key take aways:

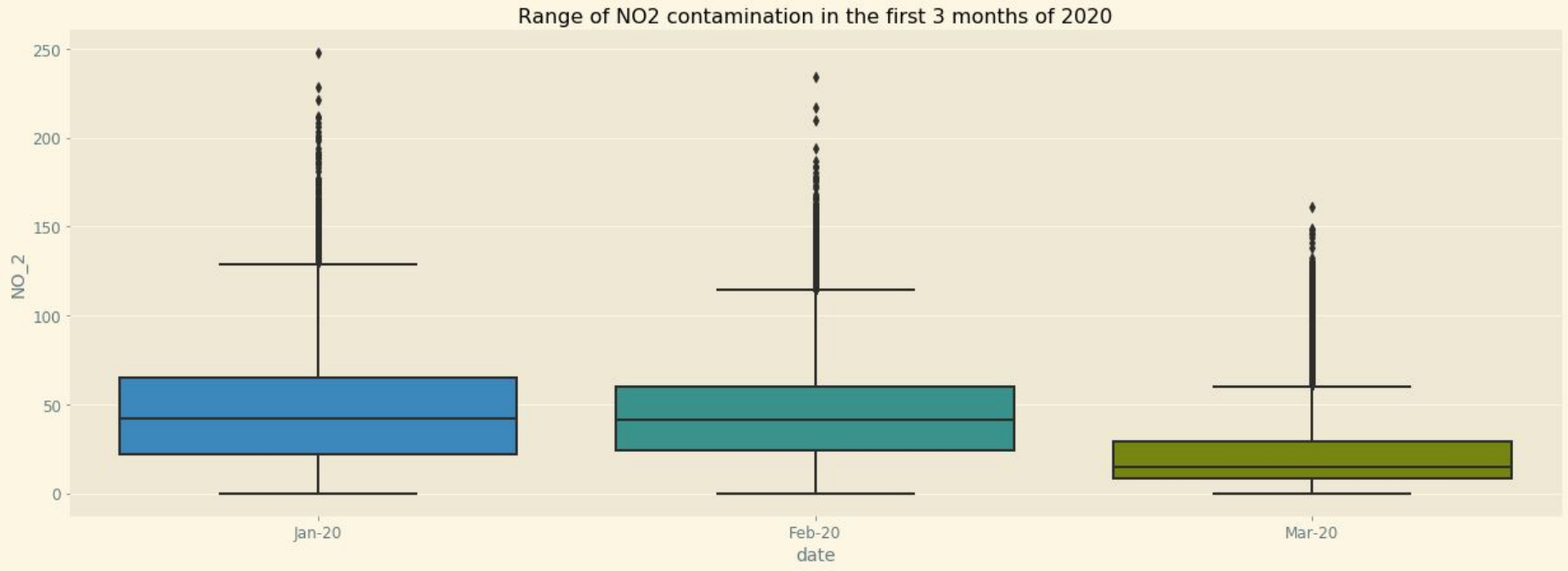
1. Most data is time series in its nature. Whether we want to predict trends in financial markets, sales of a product or levels of pollution, time is an important factor and should be factored into our models.

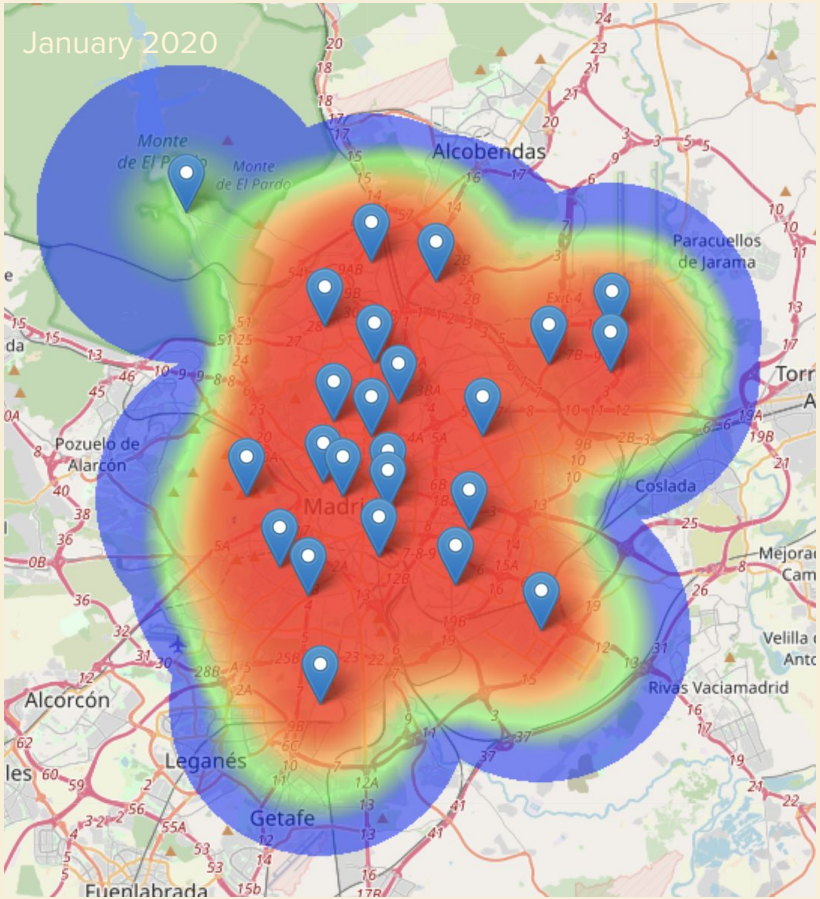
2. Both models perform better on stationary data. It's important to analyse any trends and patterns and transform the time series data if needed, so that it becomes stationary. Using non-stationary time series data produces unreliable results and leads to poor forecasting.

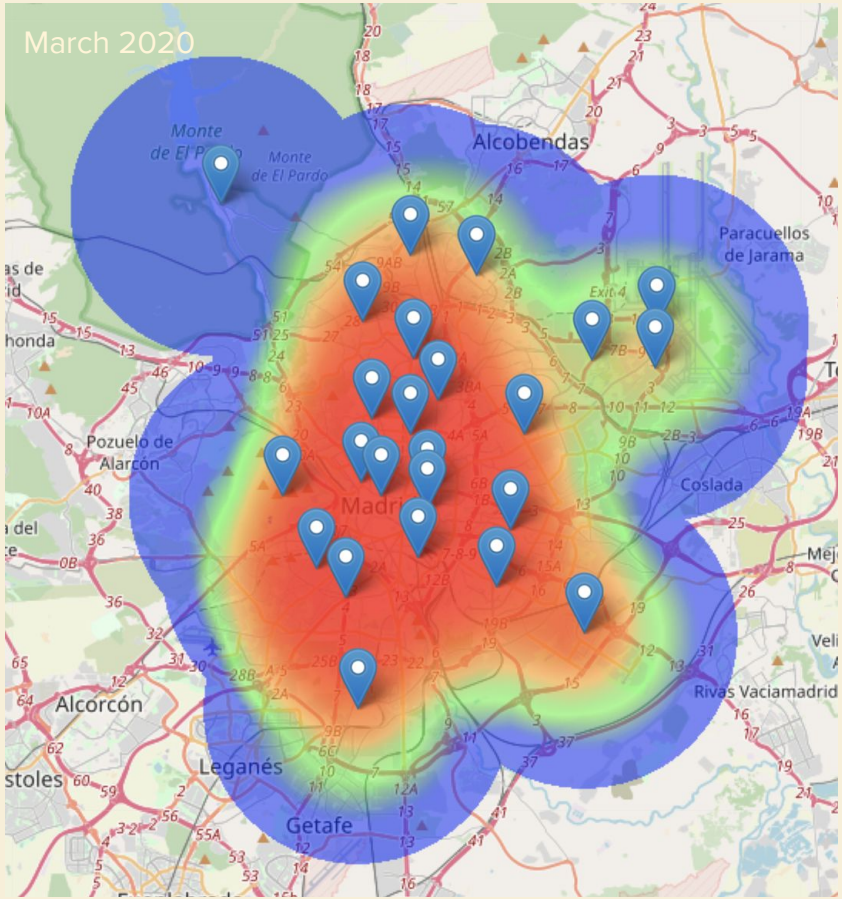
3. Any model has pros and cons. In this case, the XGB model had better overall performance while the ARIMA performed better at predicting dangerous levels of the particle. We can optimise our model by using an evaluation metric that penalises these bad predictions.


BONUS
74
30

Since the start of the Covid-19 crisis, levels of NO2 contamination have dropped significantly! 🌳🌍




January 2020
Alcobendas
Paracuellos de Jarama
Pozuelo de Alarcón
Coslada
Alcorcón
Leganés
Getafe
Rivas Vaciamadrid

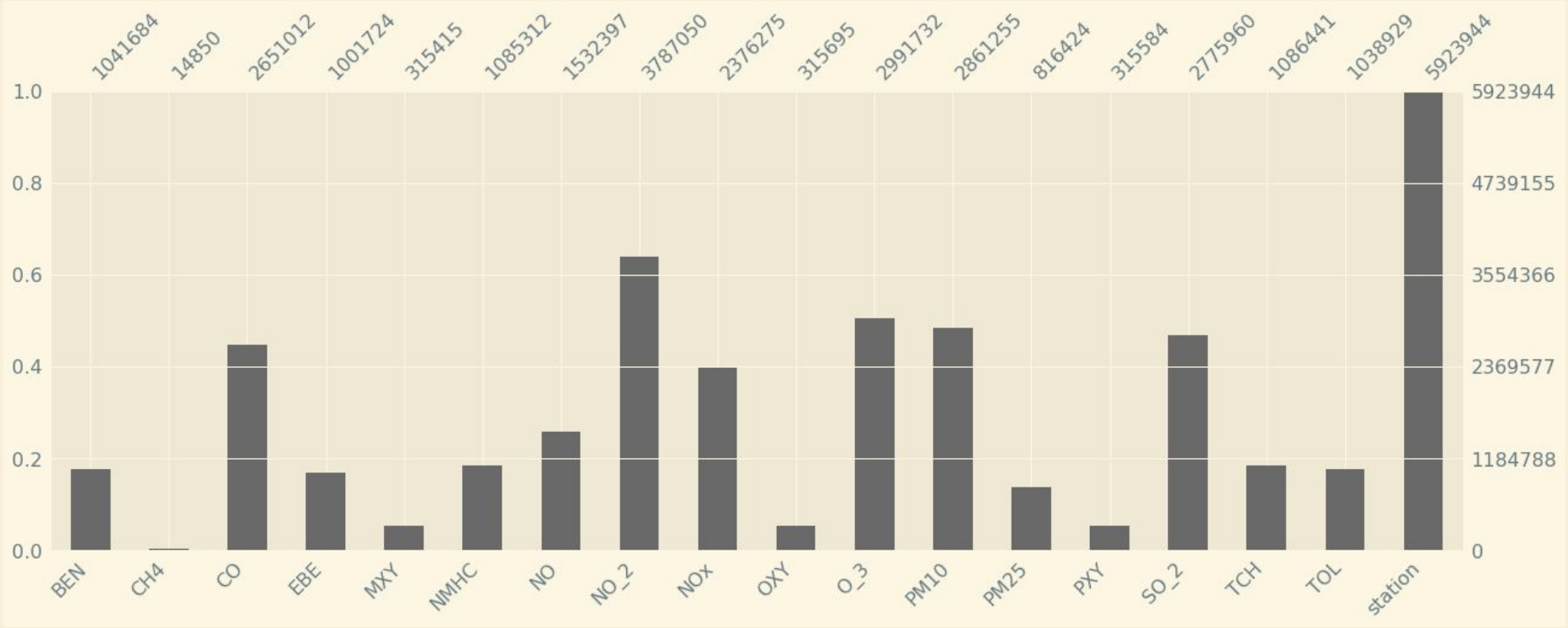

March 2020
Alcobendas
Paracuellos de Jarama
Pozuelo de Alarcón
Coslada
Alcorcón
Leganés
Getafe
Rivas Vaciamadrid

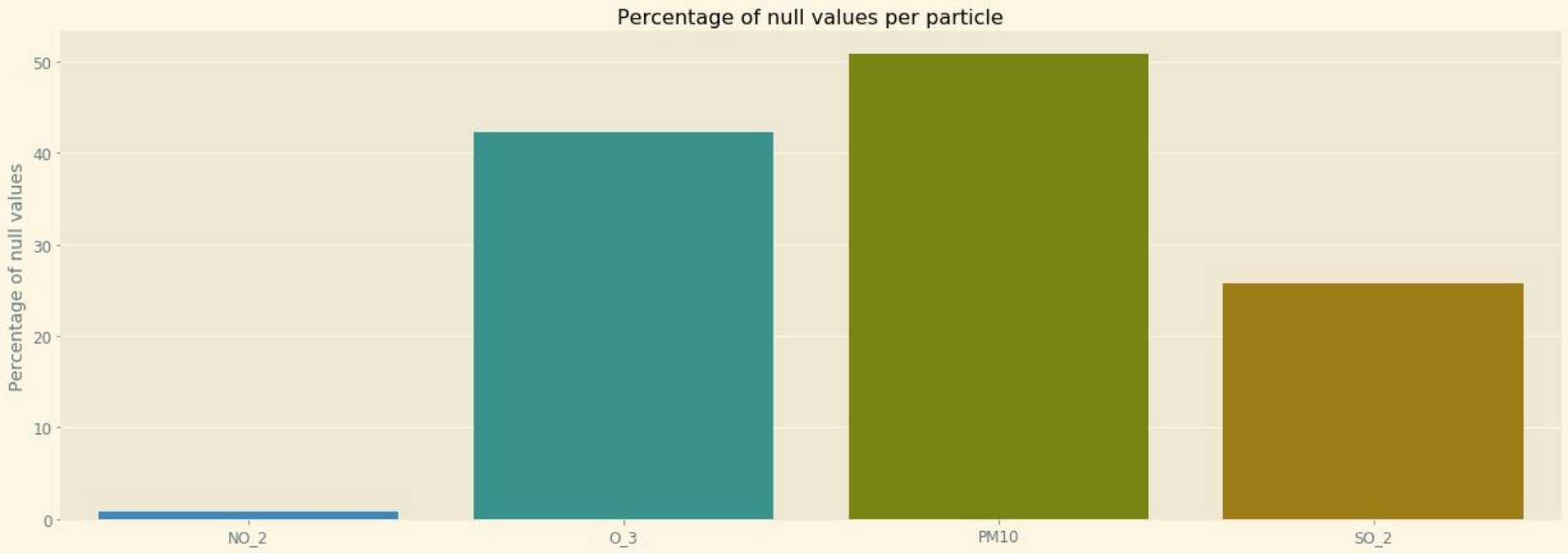

January 2020
Chamartín
Cerro de Los Locos o de Las Balas
M-500
Carretera de Castilla
Chamberí
Cerro de Garabitas 675 m
Parque del Oeste
Río Manzanares
Salamanca
Cerro de las Ánimas 660 m
626 m
590 m
610 m
608 m
Madrid
Retiro
Autovía del Suroeste
Latina
Arganzuela
Cerro de la Mica 663 m
Aluche
Usera
Bypass Sur

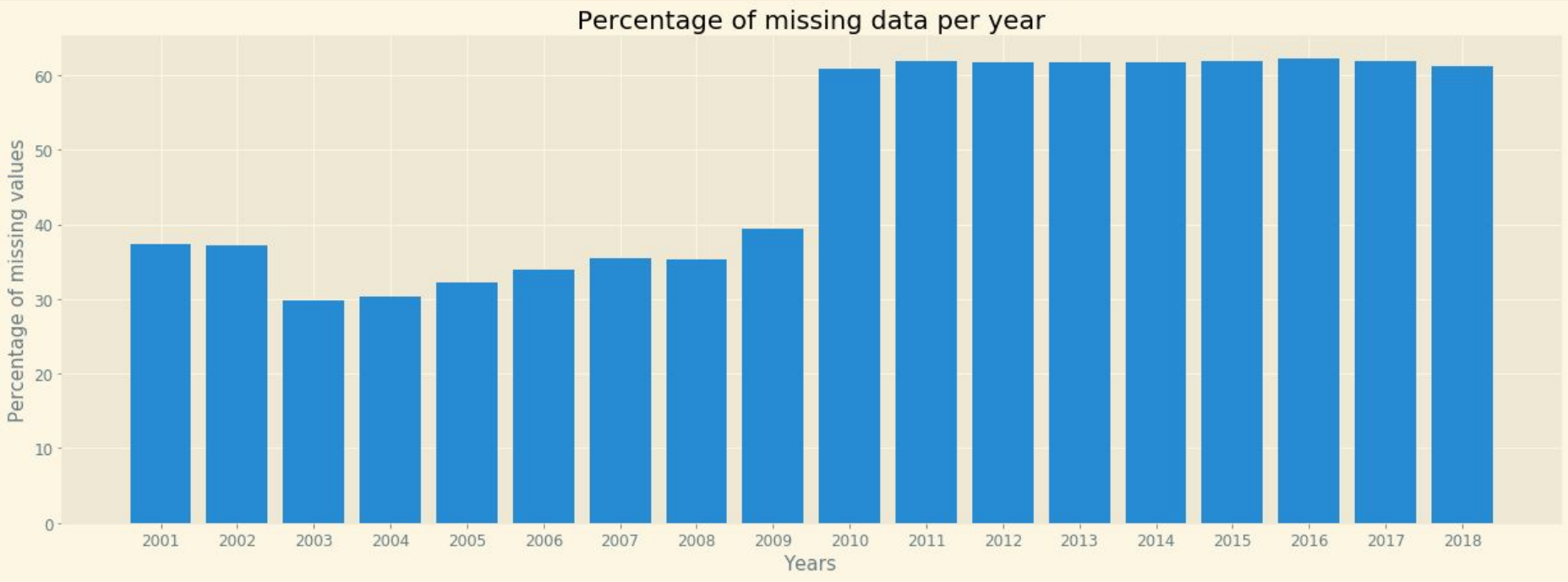

March 2020
Chamartín
Cerro de Los Locos o de Las Balas
M-500
Carretera de Castilla
Chamberí
Cerro de Garabitas 675 m
Parque del Oeste
Río Manzanares
Salamanca
Cerro de las Ánimas 660 m
626 m
590 m
610 m
608 m
Madrid
Retiro
Autovía del Suroeste
Latina
Arganzuela
Cerro de la Mica 663 m
Aluche
Usera
Bypass Sur

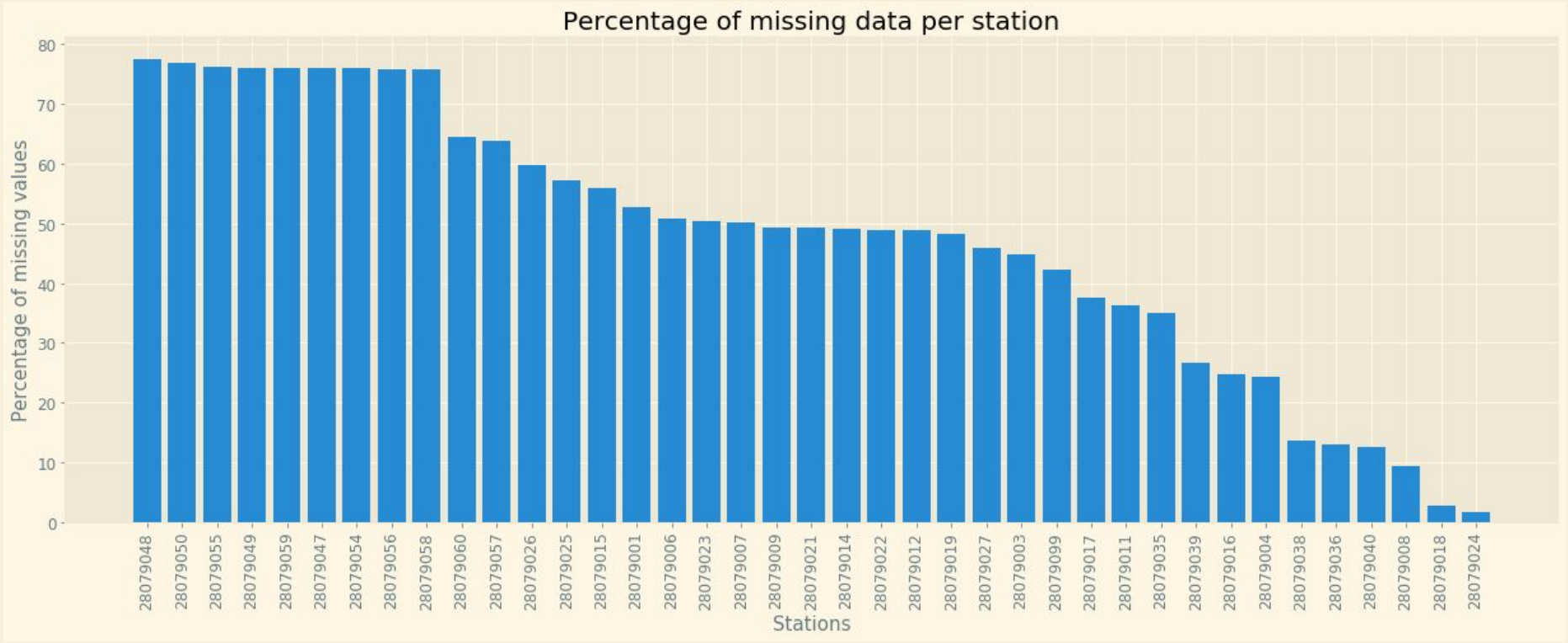

THANK YOU

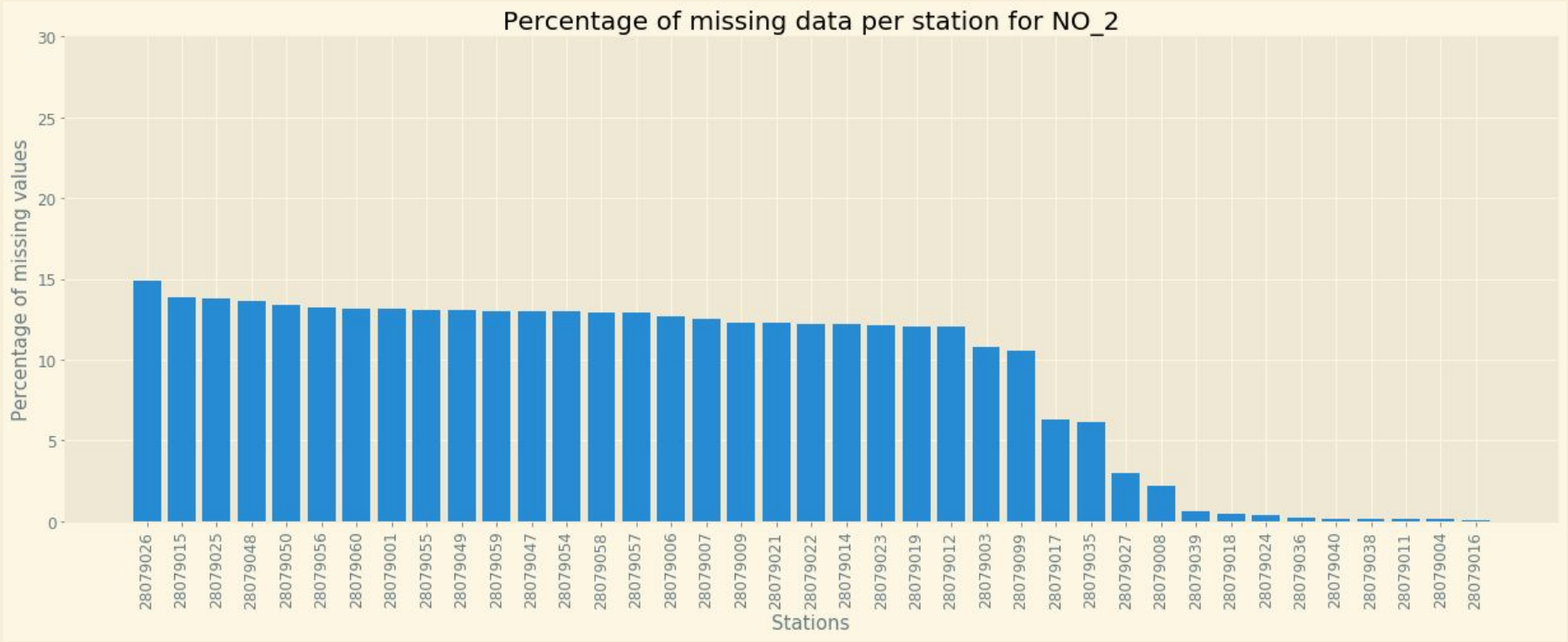
Missing Values

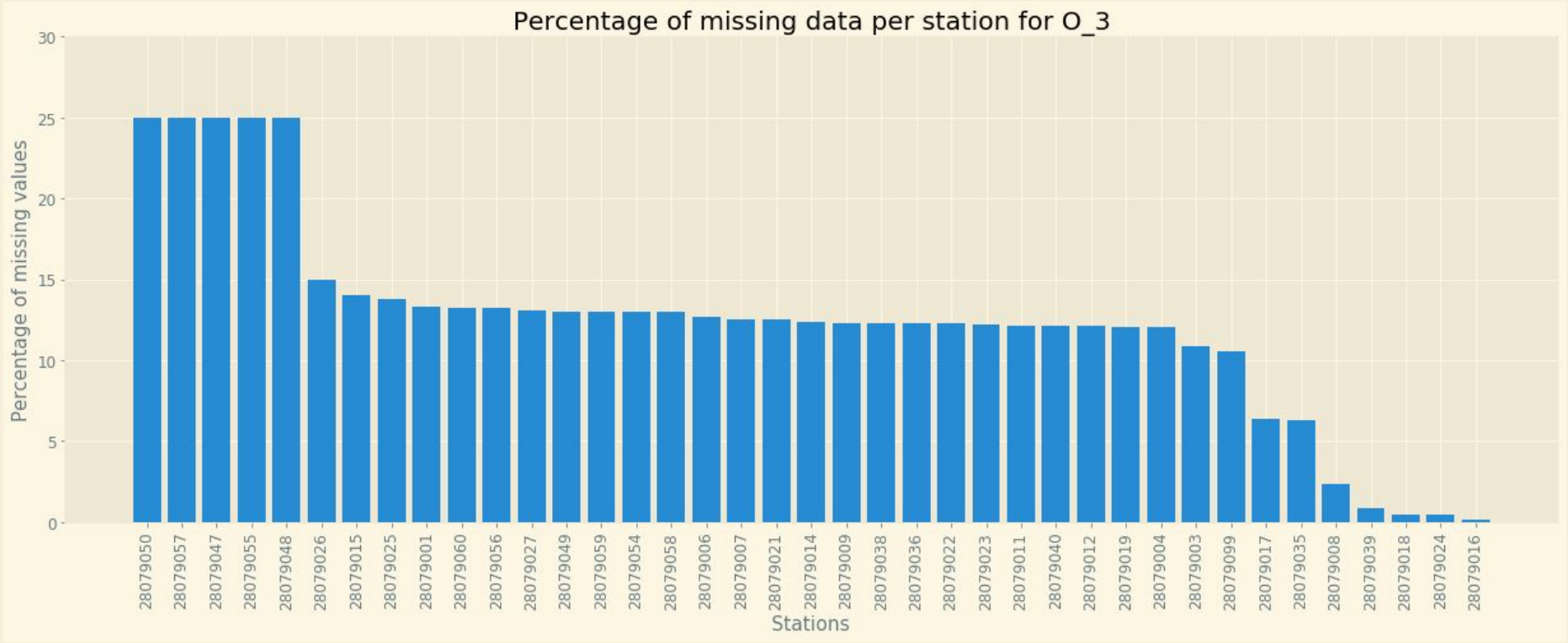

1041684
14850
2651012
1001724
315415
1085312
1532397
3787050
2376275
315695
2991732
2861255
816424
315584
2775960
1086441
1038929
5923944
1.0
0.8
0.6
0.4
0.2
0.0
5923944
4739155
3554366
2369577
1184788
0
BEN
CH4
CO
EBE
MXY
NMHC
NO
NO_2
NOx
OXY
O_3
PM10
PM25
PXY
SO_2
TCH
TOL
station

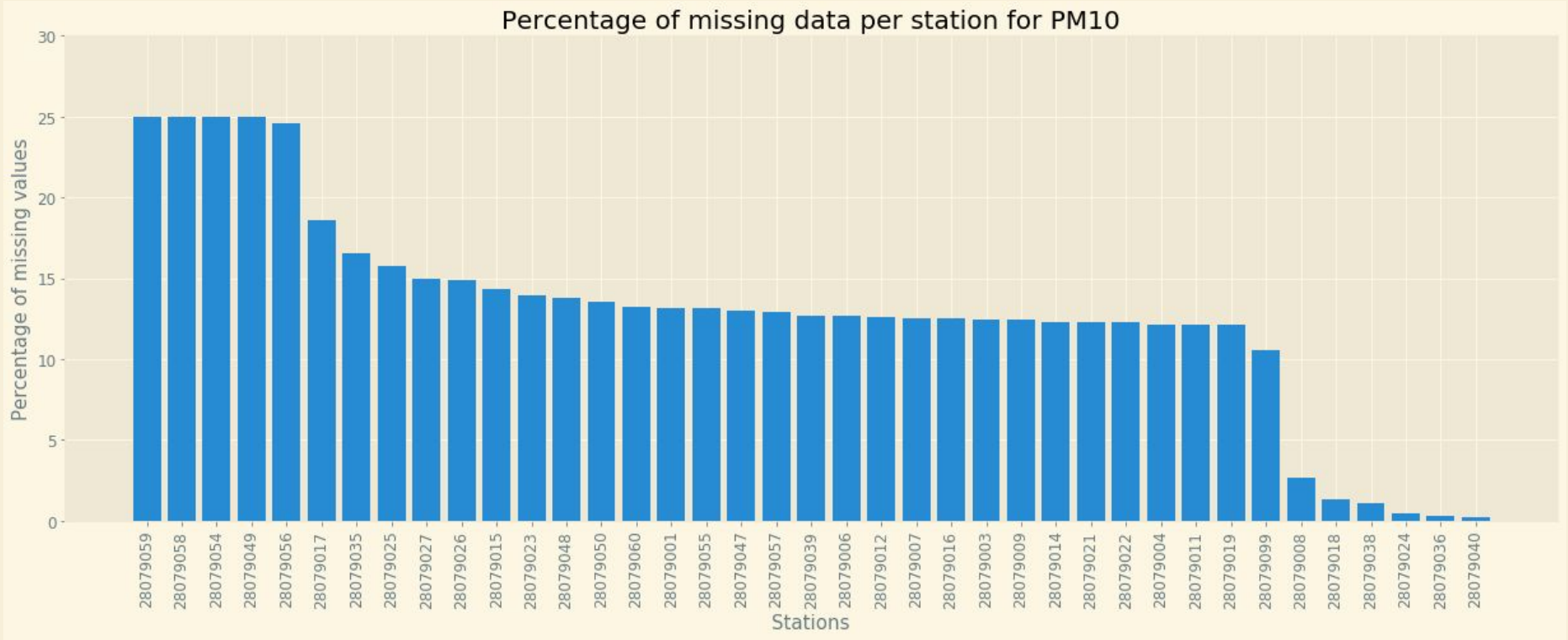

Percentage of null values per particle
Percentage of null values
0
10
20
30
40
50
NO_2
O_3
PM10
SO_2

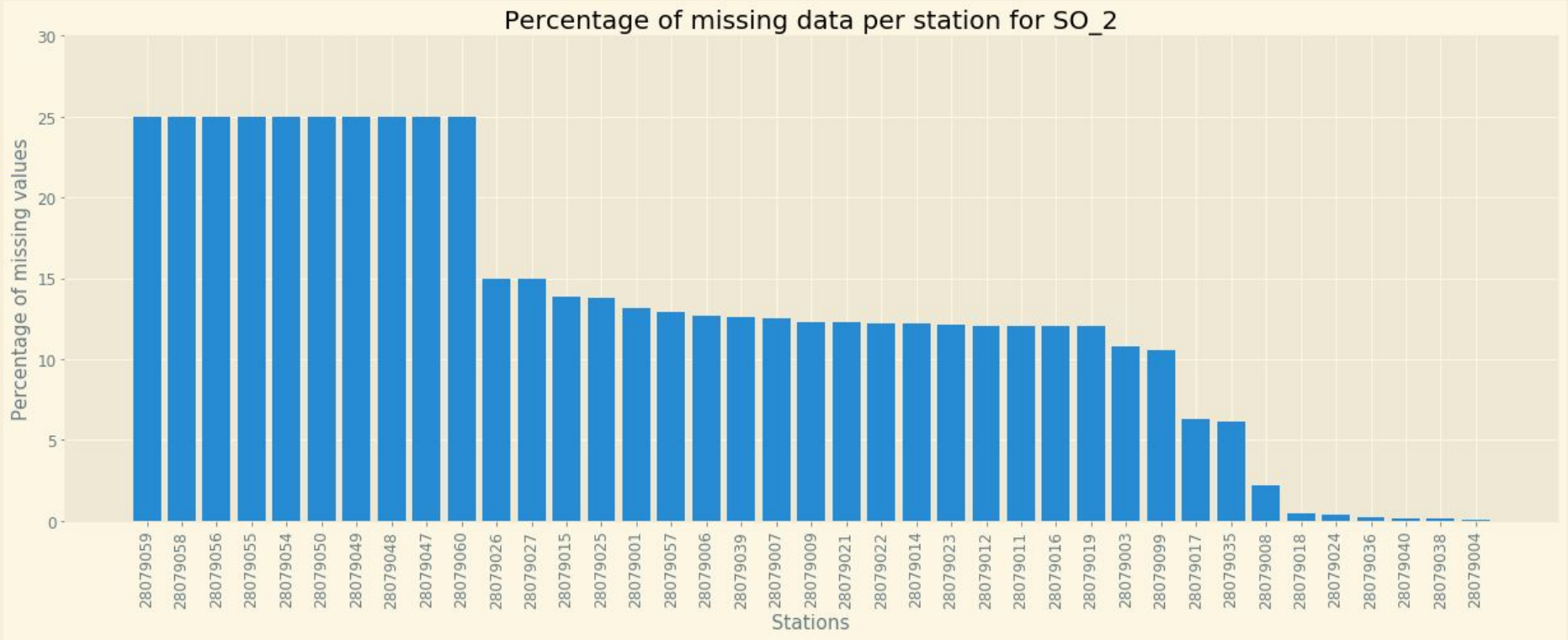

Percentage of missing data per year
Percentage of missing values
0
10
20
30
40
50
60
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years


Percentage of missing data per station
Percentage of missing values
0
10
20
30
40
50
60
70
80
28079048
28079050
28079055
28079049
28079059
28079047
28079054
28079056
28079058
28079060
28079057
28079026
28079025
28079015
28079001
28079006
28079023
28079007
28079009
28079021
28079014
28079022
28079012
28079019
28079027
28079003
28079099
28079017
28079011
28079035
28079039
28079016
28079004
28079038
28079036
28079040
28079008
28079018
28079024
Stations

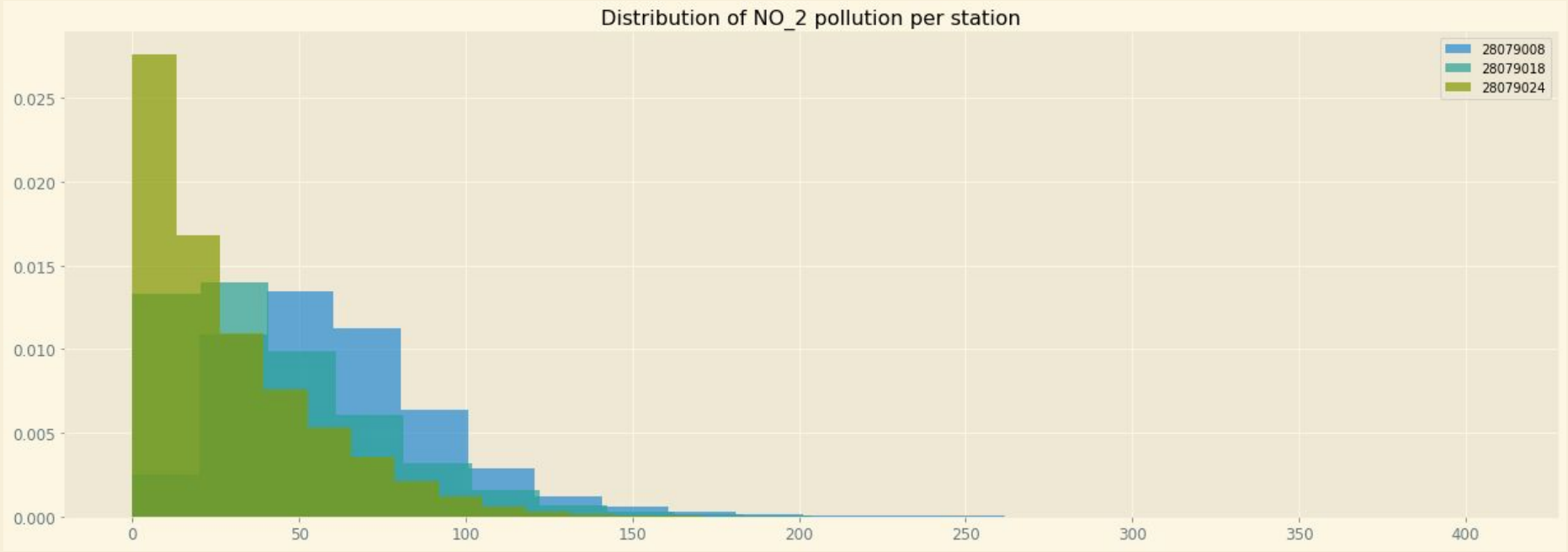

Percentage of missing data per station for NO_2
Percentage of missing values
0
5
10
15
20
25
30
28079026
28079015
28079025
28079048
28079050
28079056
28079060
28079001
28079055
28079049
28079059
28079047
28079054
28079058
28079057
28079006
28079007
28079009
28079021
28079022
28079014
28079023
28079019
28079012
28079003
28079099
28079017
28079035
28079027
28079008
28079039
28079018
28079024
28079036
28079040
28079038
28079011
28079004
28079016
Stations

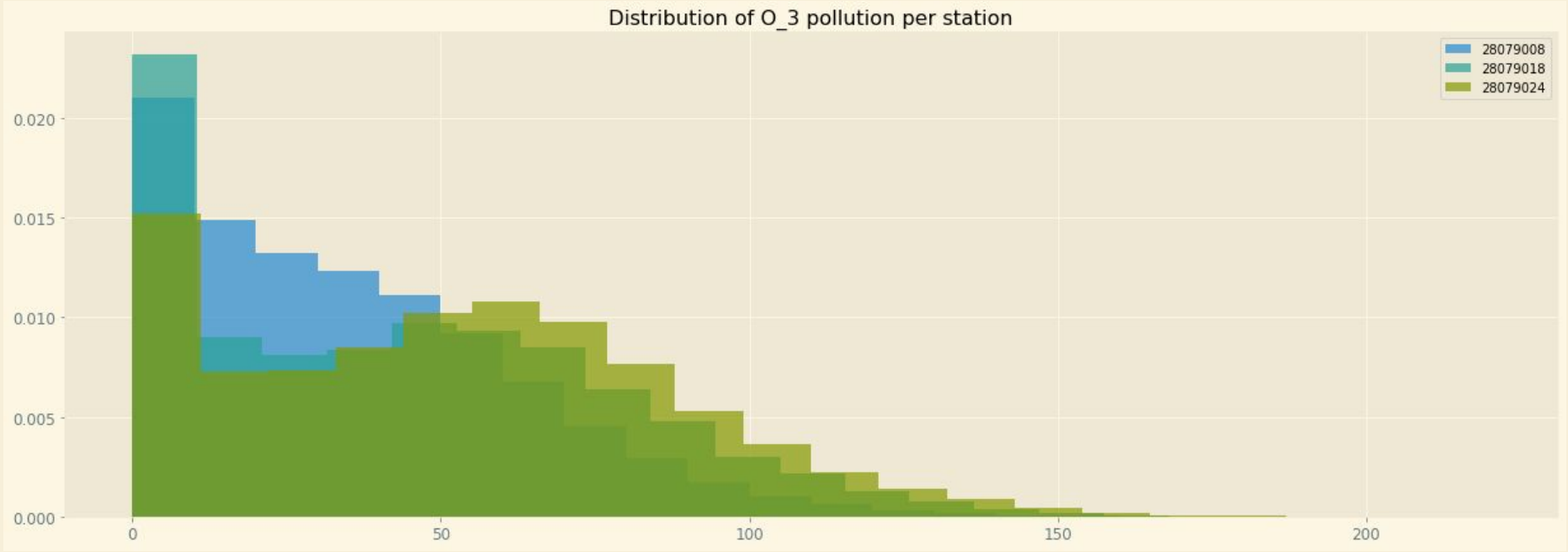

Percentage of missing data per station for O_3
Percentage of missing values
30
25
20
15
10
5
0
28079050
28079057
28079047
28079055
28079048
28079026
28079015
28079025
28079001
28079060
28079056
28079027
28079049
28079059
28079054
28079058
28079006
28079007
28079021
28079014
28079009
28079038
28079036
28079022
28079023
28079011
28079040
28079012
28079019
28079004
28079003
28079099
28079017
28079035
28079008
28079039
28079018
28079024
28079016
Stations

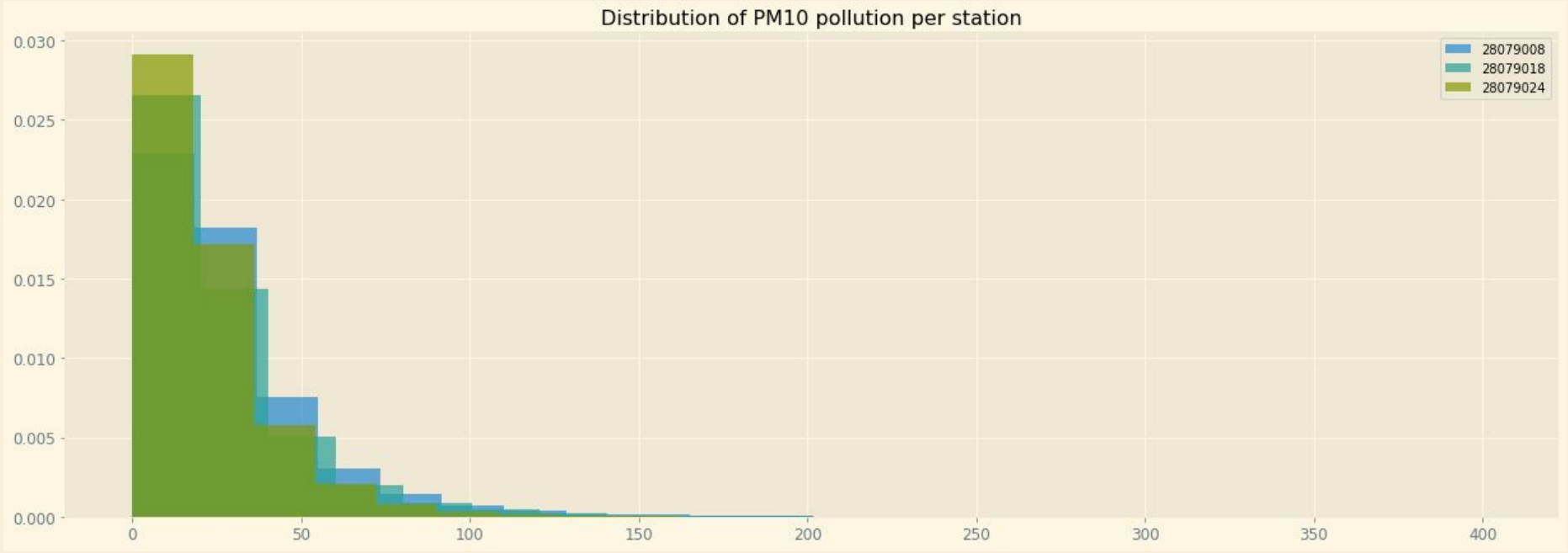

Percentage of missing data per station for PM10
Percentage of missing values
30
25
20
15
10
5
0
28079059
28079058
28079054
28079049
28079056
28079017
28079035
28079025
28079027
28079026
28079015
28079023
28079048
28079050
28079060
28079001
28079055
28079047
28079057
28079039
28079006
28079012
28079007
28079016
28079003
28079009
28079014
28079021
28079022
28079004
28079011
28079019
28079099
28079008
28079018
28079038
28079024
28079036
28079040
Stations

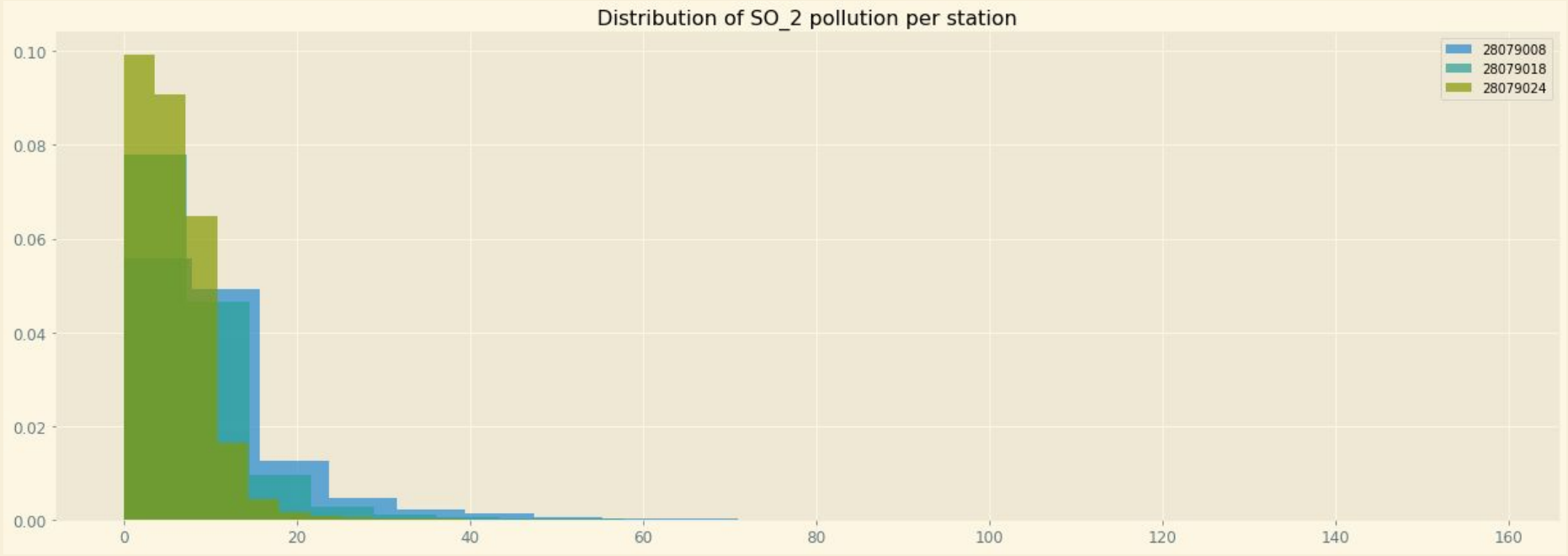

Percentage of missing data per station for SO_2
Percentage of missing values
30
25
20
15
10
5
0
28079059
28079058
28079056
28079055
28079054
28079050
28079049
28079048
28079047
28079060
28079026
28079027
28079015
28079025
28079001
28079057
28079006
28079039
28079007
28079009
28079021
28079022
28079014
28079023
28079012
28079011
28079016
28079019
28079003
28079099
28079017
28079035
28079008
28079018
28079024
28079036
28079040
28079038
28079004
Stations

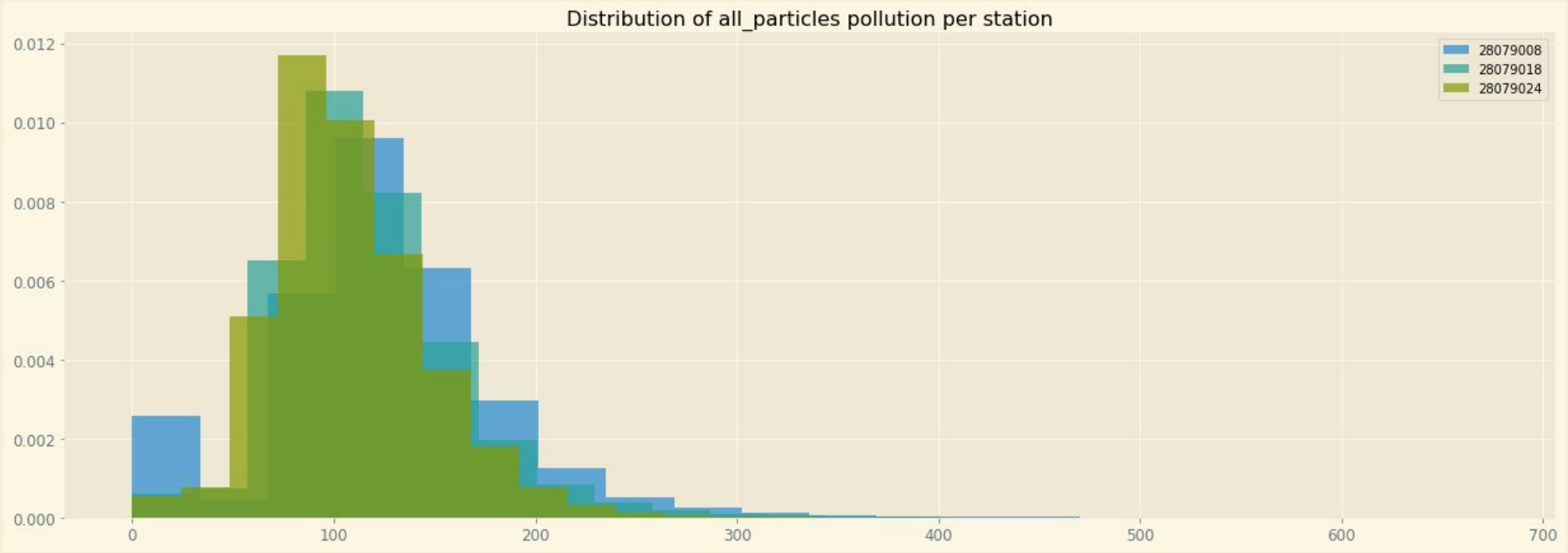
Distribution of the levels of pollution for each particle


Distribution of NO_2 pollution per station
28079008
28079018
28079024
0.025
0.020
0.015
0.010
0.005
0.000
0
50
100
150
200
250
300
350
400

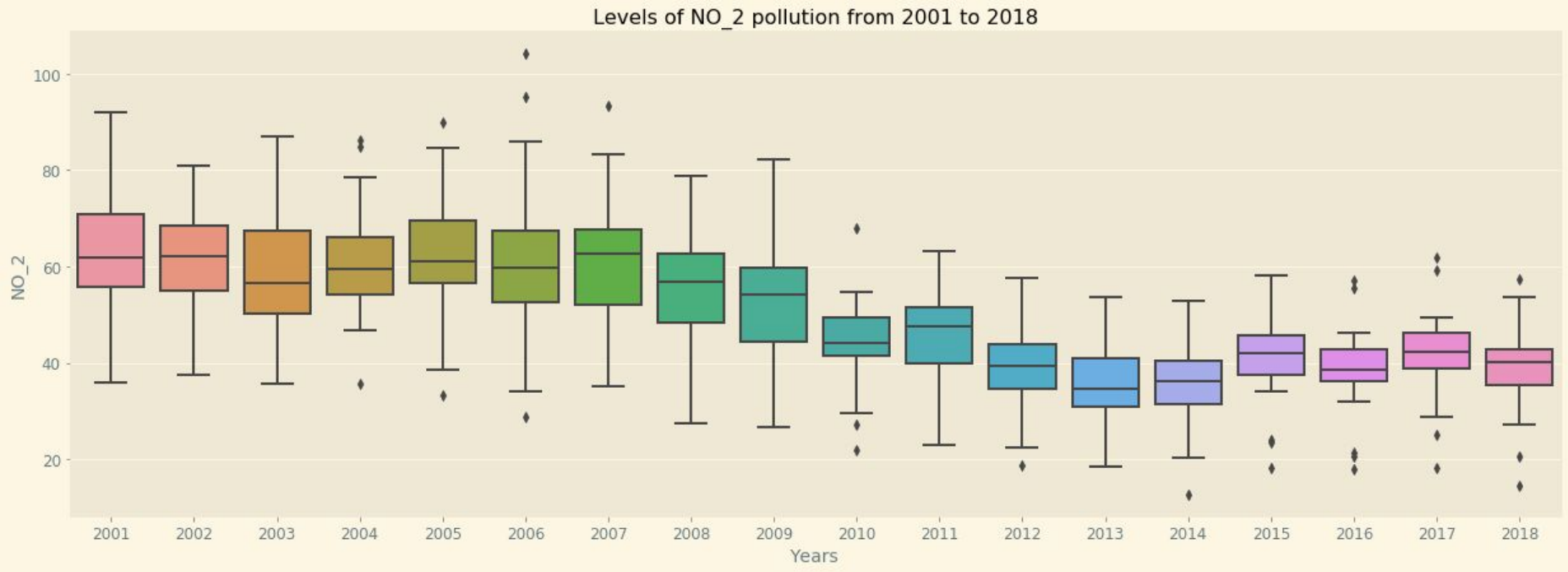

Distribution of O_3 pollution per station
28079008
28079018
28079024
0.020
0.015
0.010
0.005
0.000
0
50
100
150
200

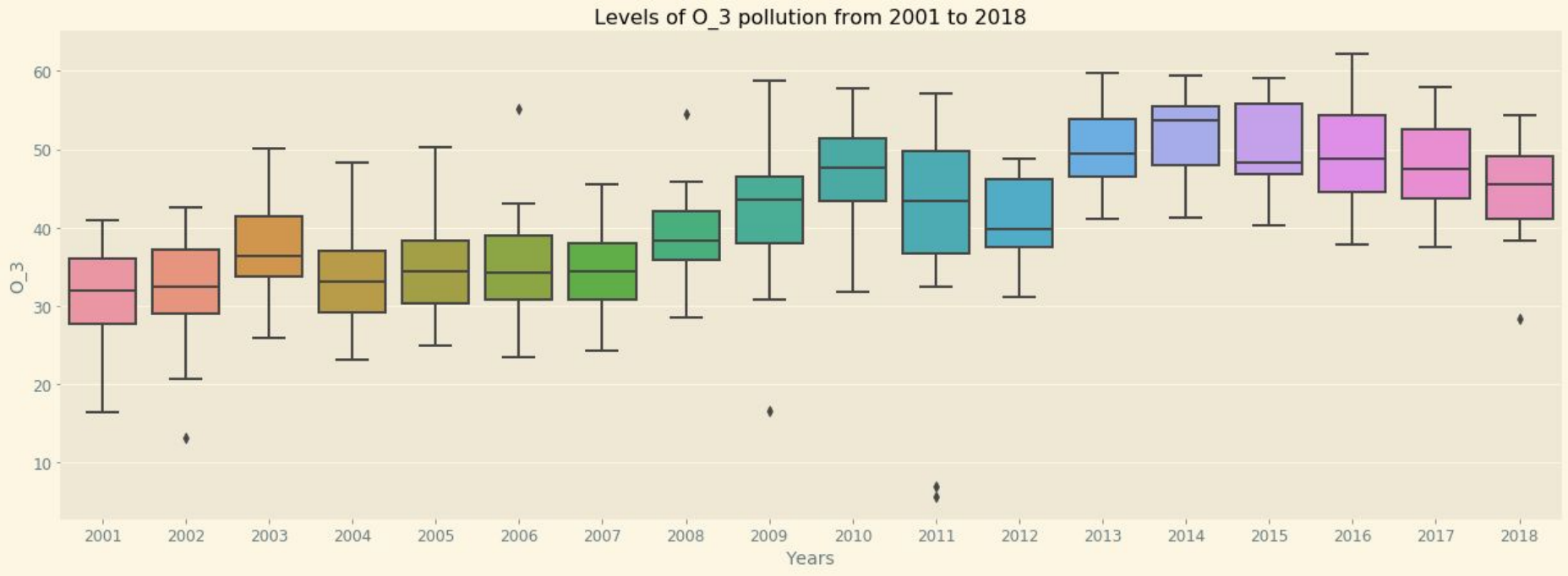

Distribution of PM10 pollution per station
28079008
28079018
28079024
0.030
0.025
0.020
0.015
0.010
0.005
0.000
0
50
100
150
200
250
300
350
400

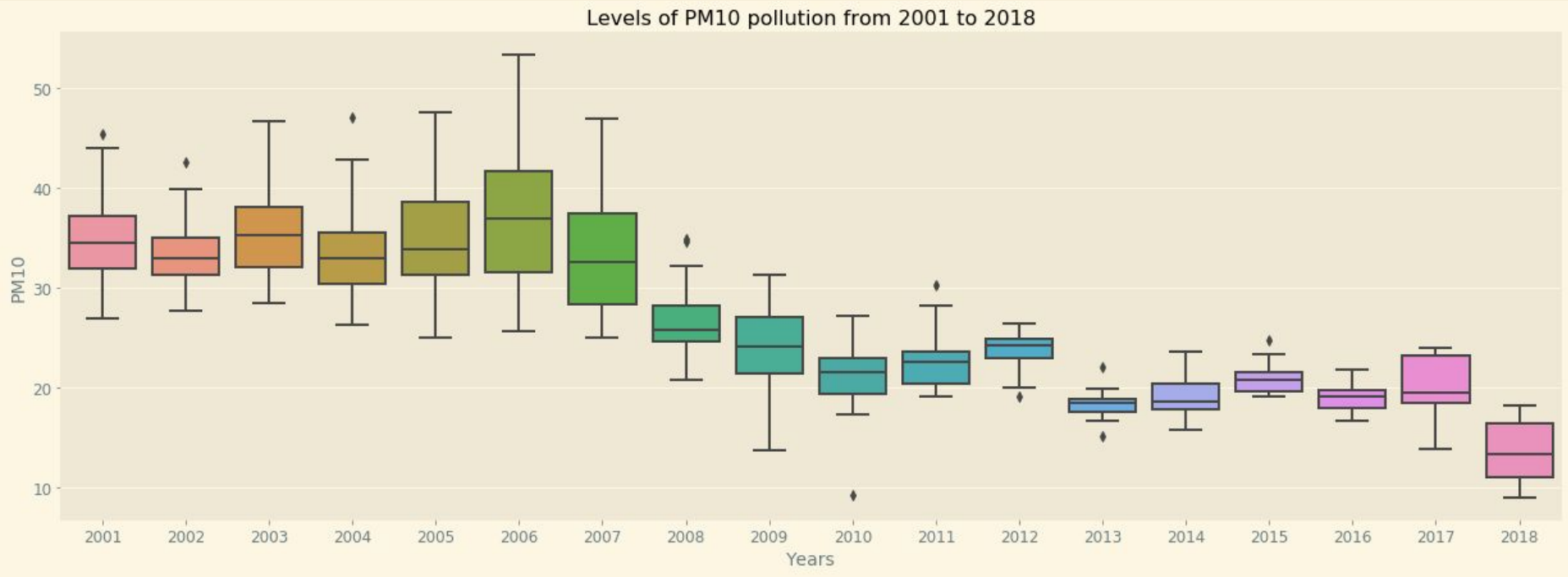

Distribution of SO_2 pollution per station
28079008
28079018
28079024
0.10
0.08
0.06
0.04
0.02
0.00
0
20
40
60
80
100
120
140
160

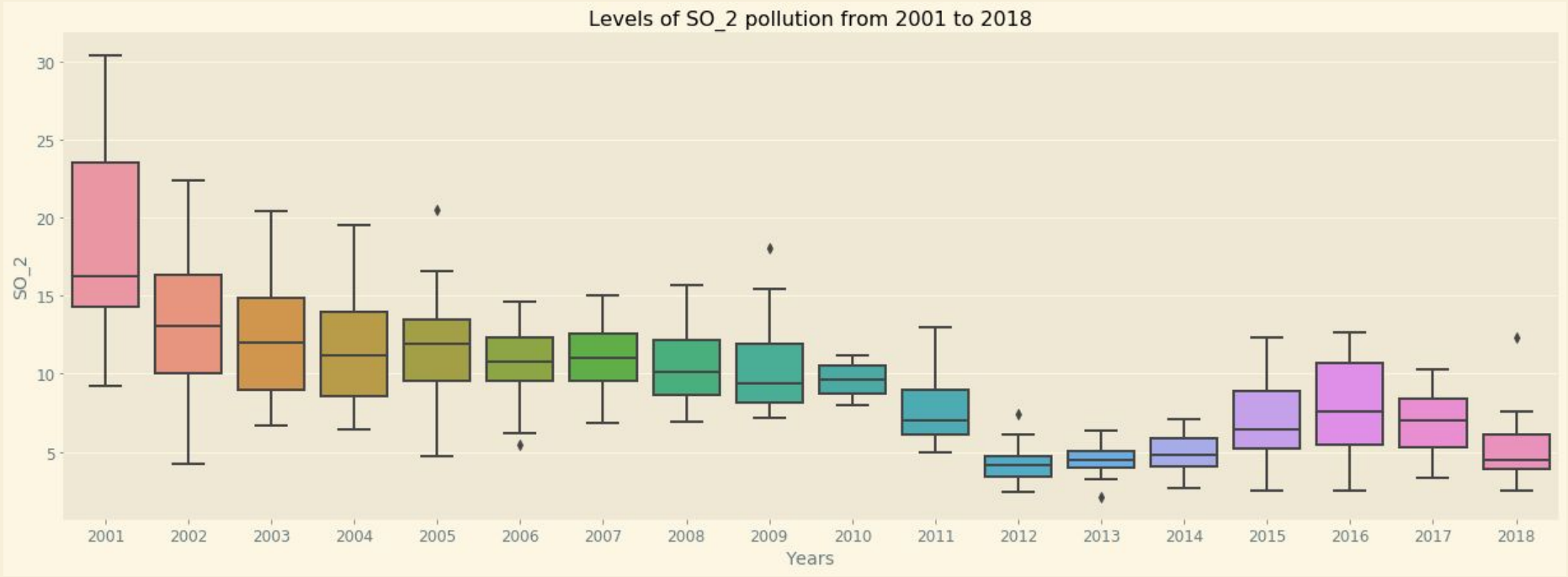

Distribution of all_particles pollution per station
28079008
28079018
28079024
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
100
200
300
400
500
600
700

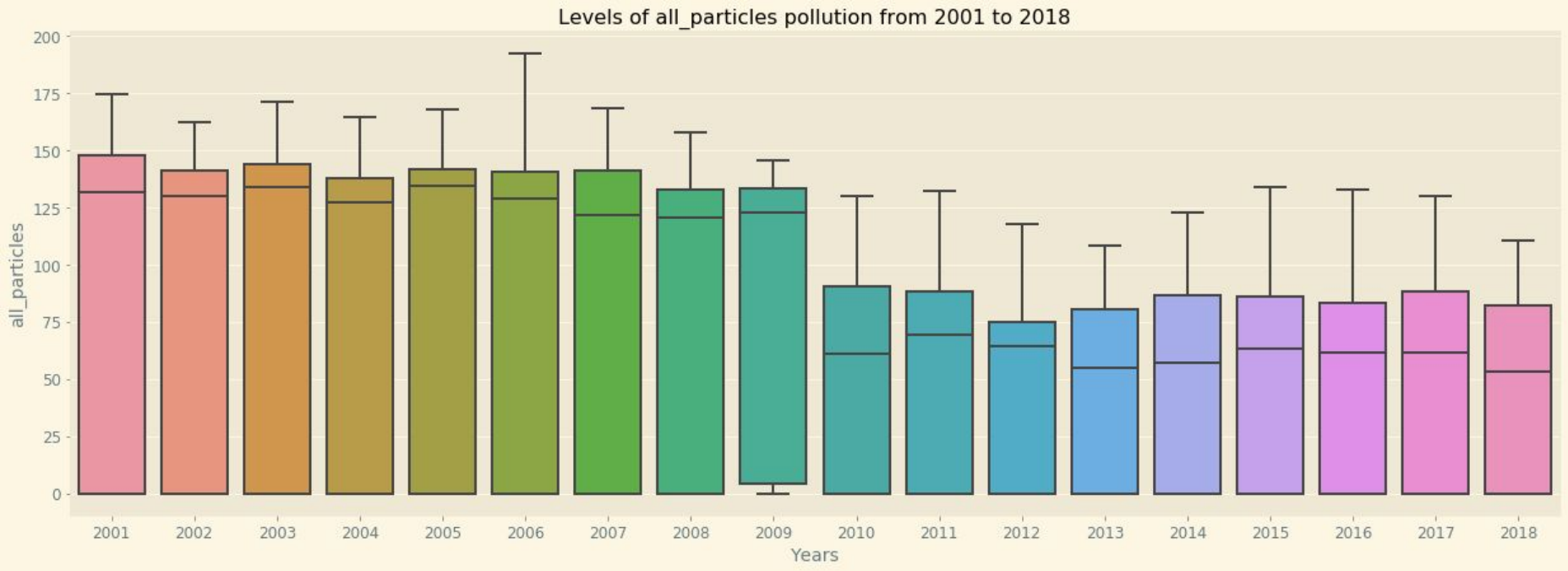
Evolution of the levels of pollution throughout the years


Levels of NO_2 pollution from 2001 to 2018
NO_2
100
80
60
40
20
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

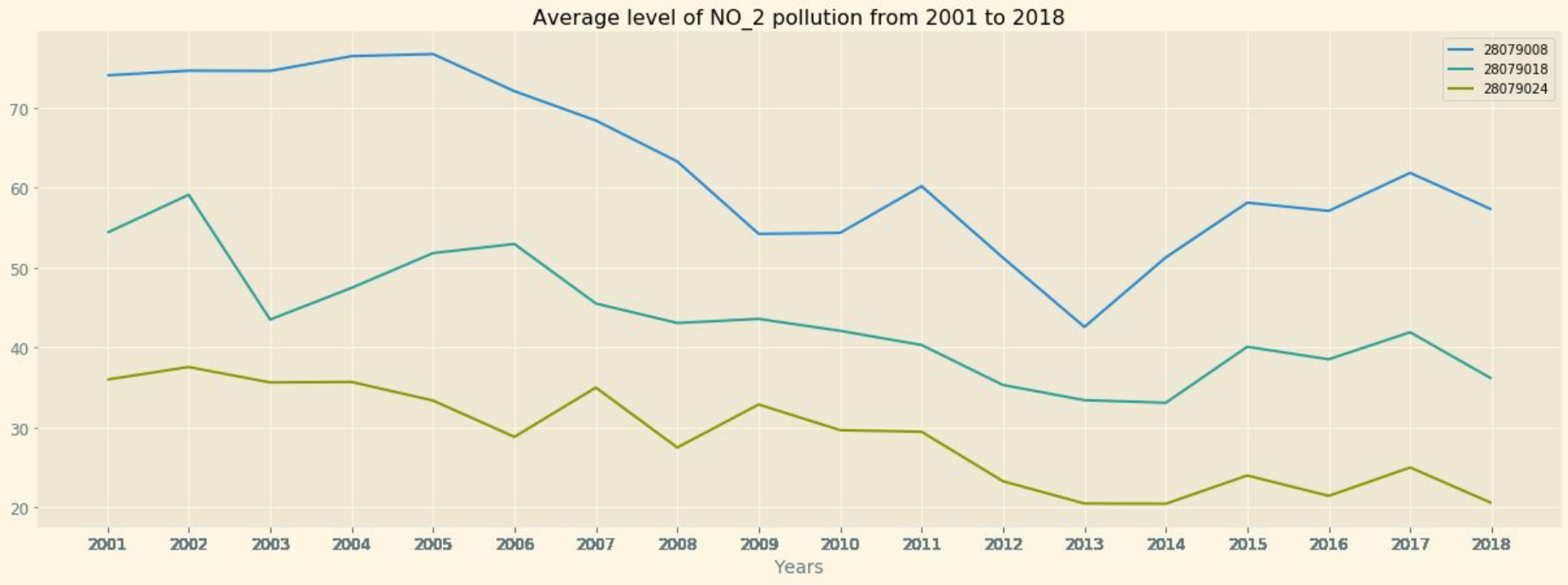

Levels of O_3 pollution from 2001 to 2018
O_3
60
50
40
30
20
10
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

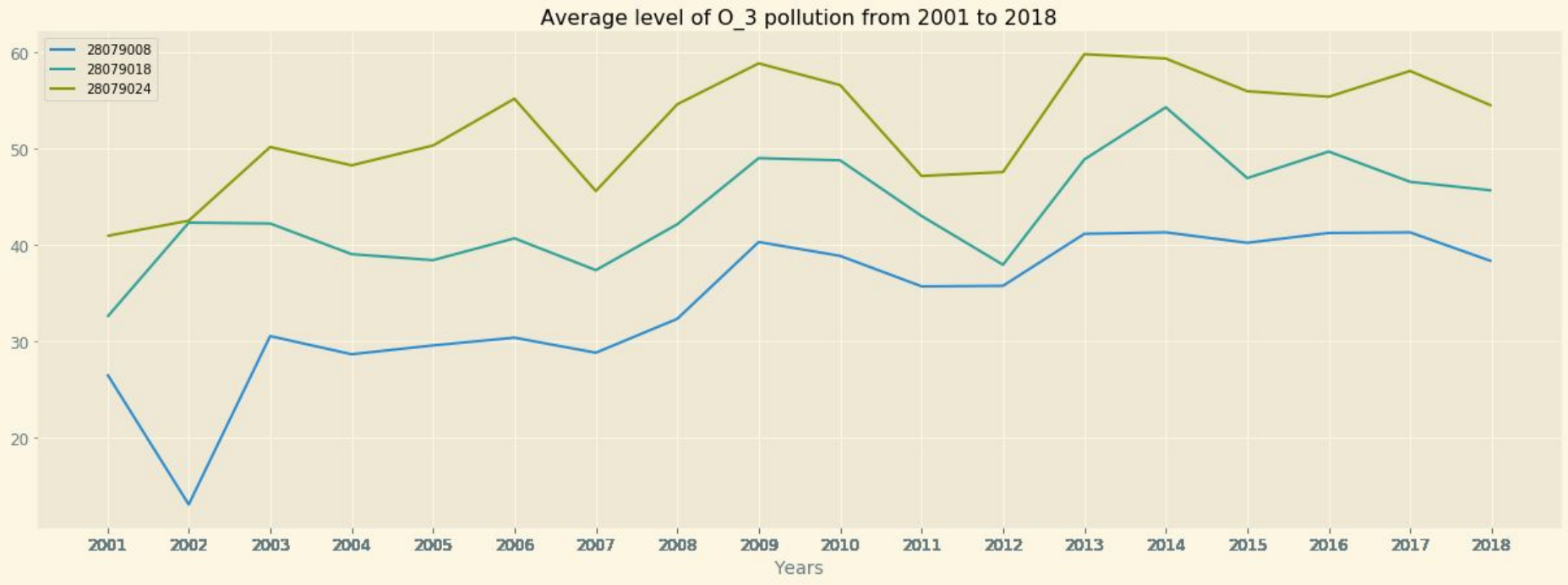

Levels of PM10 pollution from 2001 to 2018
PM10
50
40
30
20
10
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

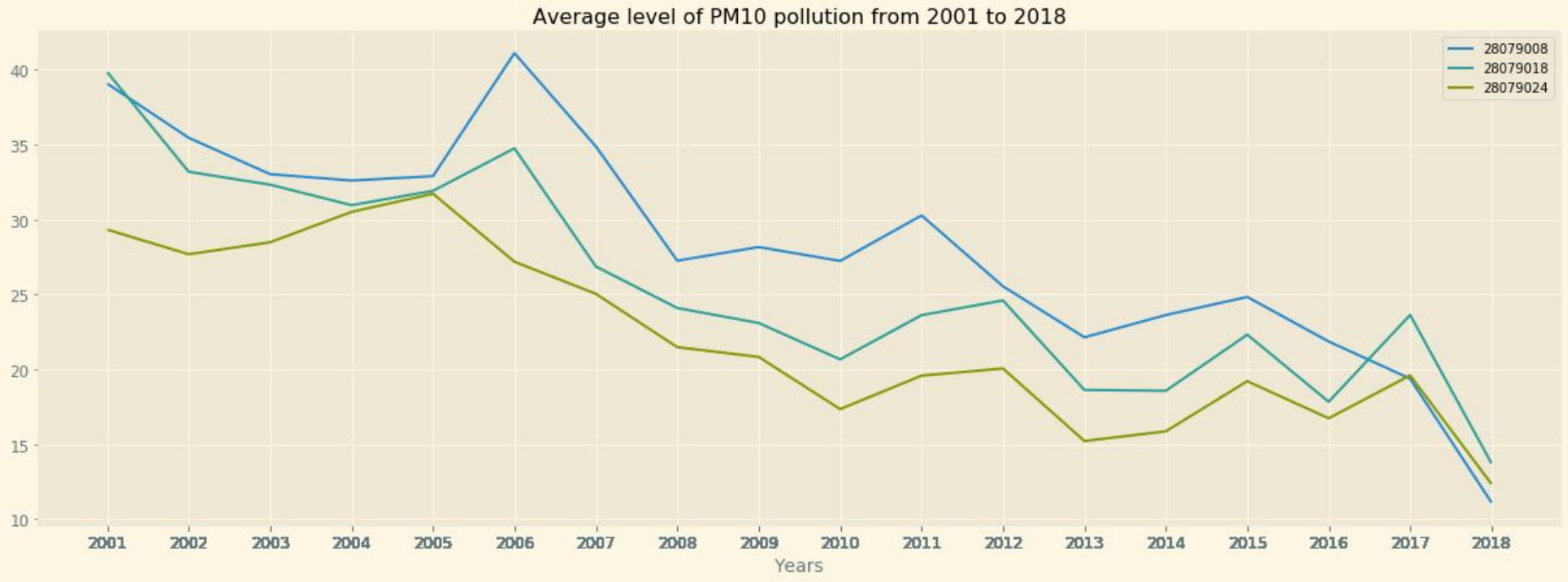

Levels of SO_2 pollution from 2001 to 2018
SO_2
30
25
20
15
10
5
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

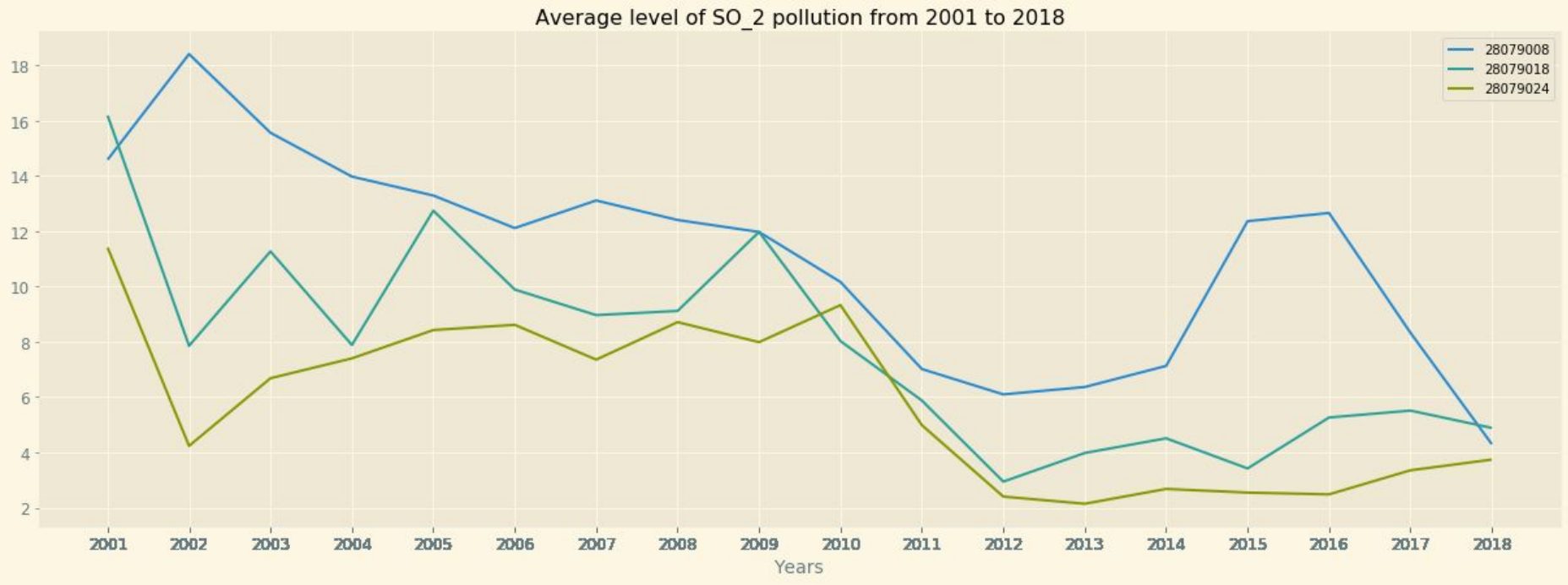

Levels of all_particles pollution from 2001 to 2018
all_particles
200
175
150
125
100
75
50
25
0
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

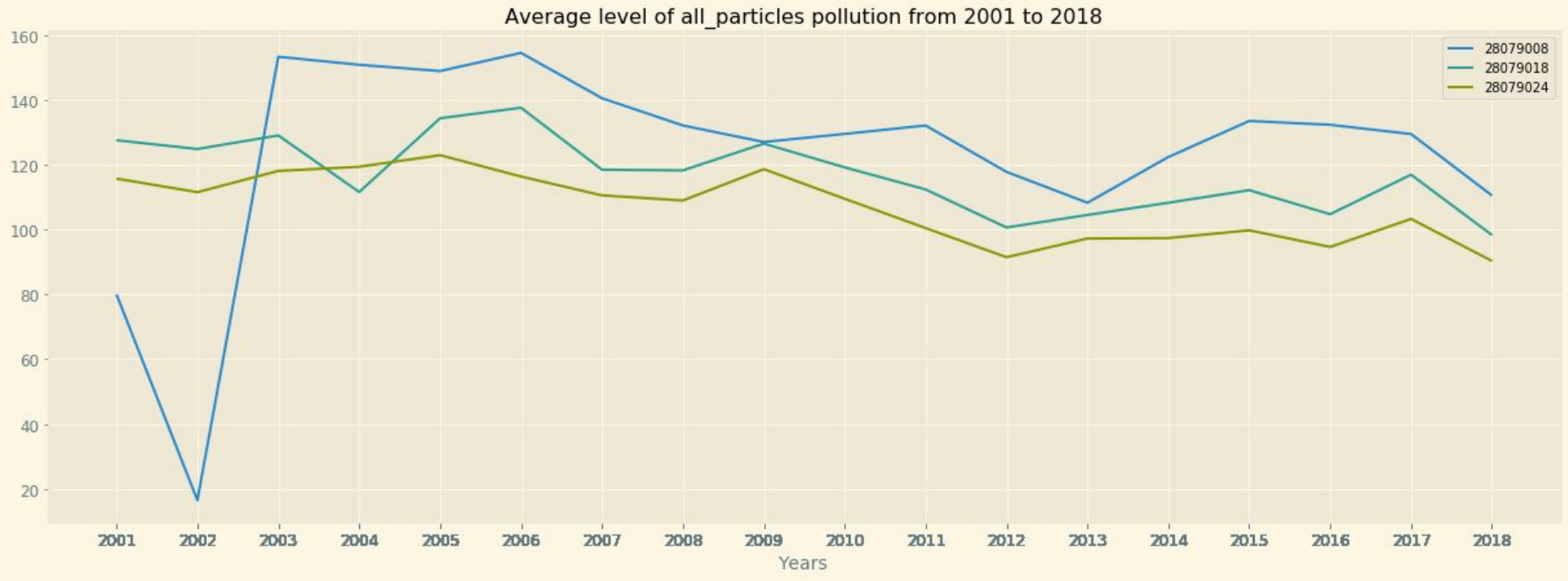
Average levels of pollution throughout the years


Average level of NO_2 pollution from 2001 to 2018
28079008
28079018
28079024
70
60
50
40
30
20
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

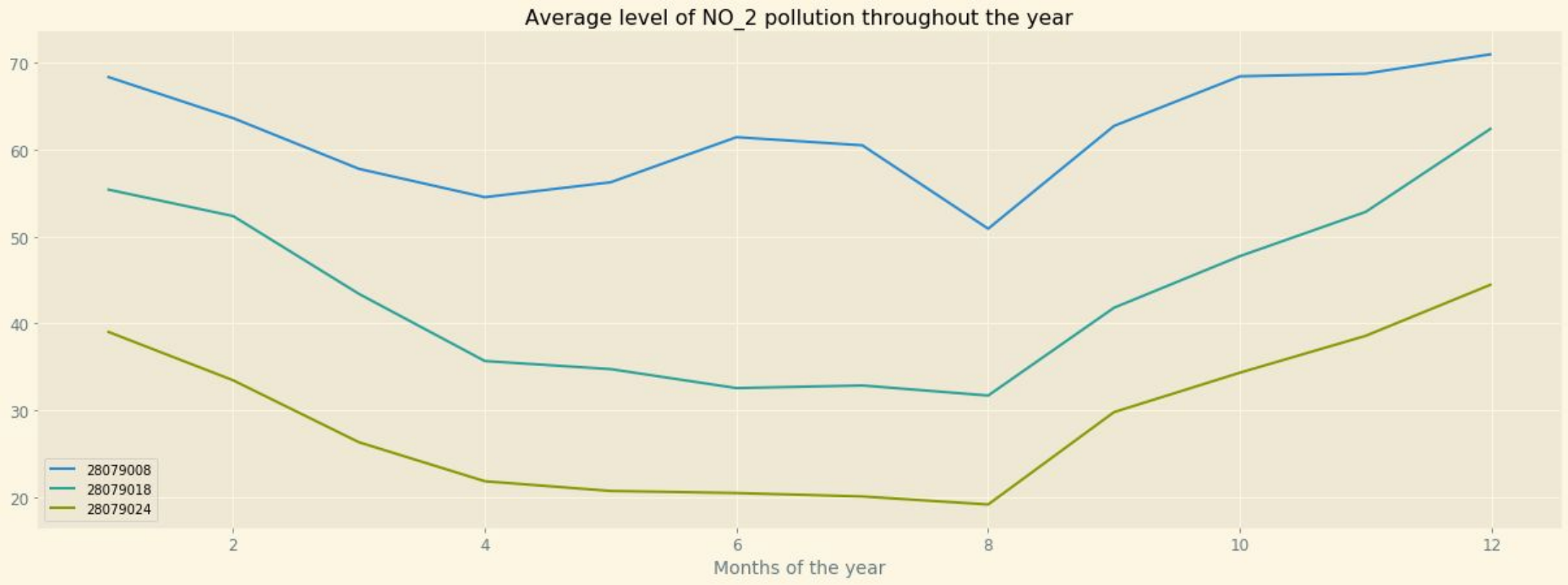

Average level of O_3 pollution from 2001 to 2018
28079008
28079018
28079024
60
50
40
30
20
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

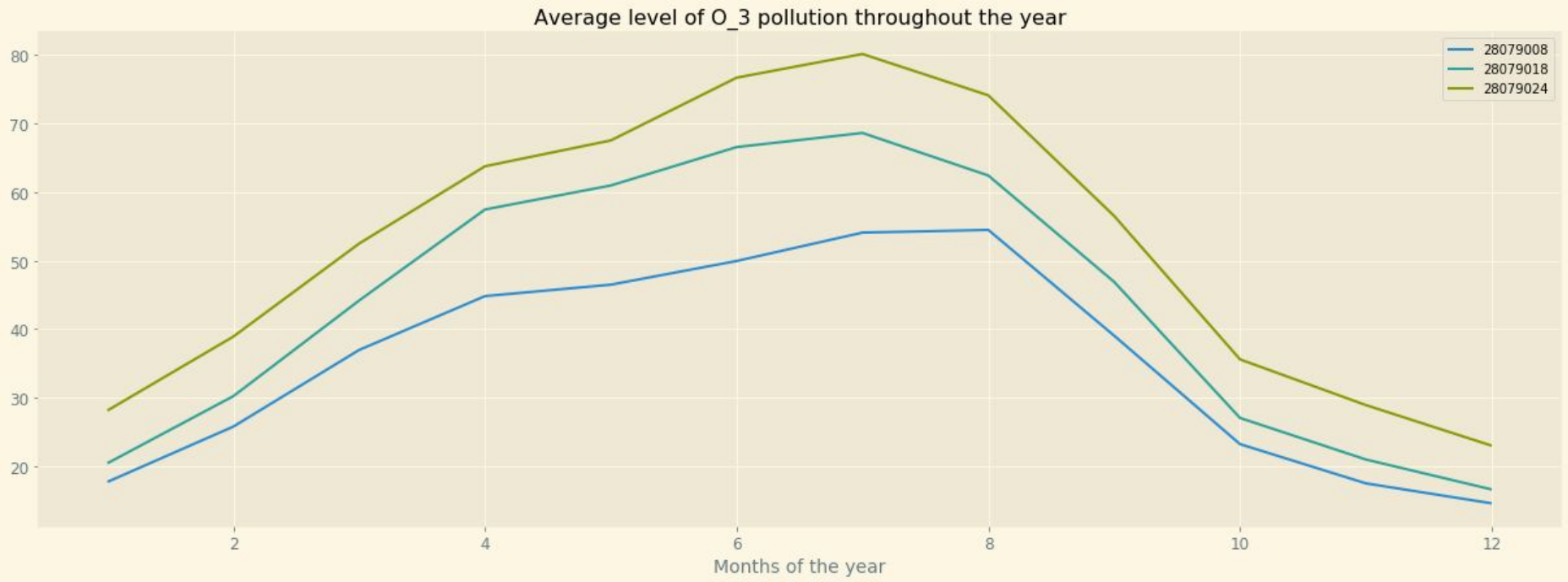

Average level of PM10 pollution from 2001 to 2018
28079008
28079018
28079024
40
35
30
25
20
15
10
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

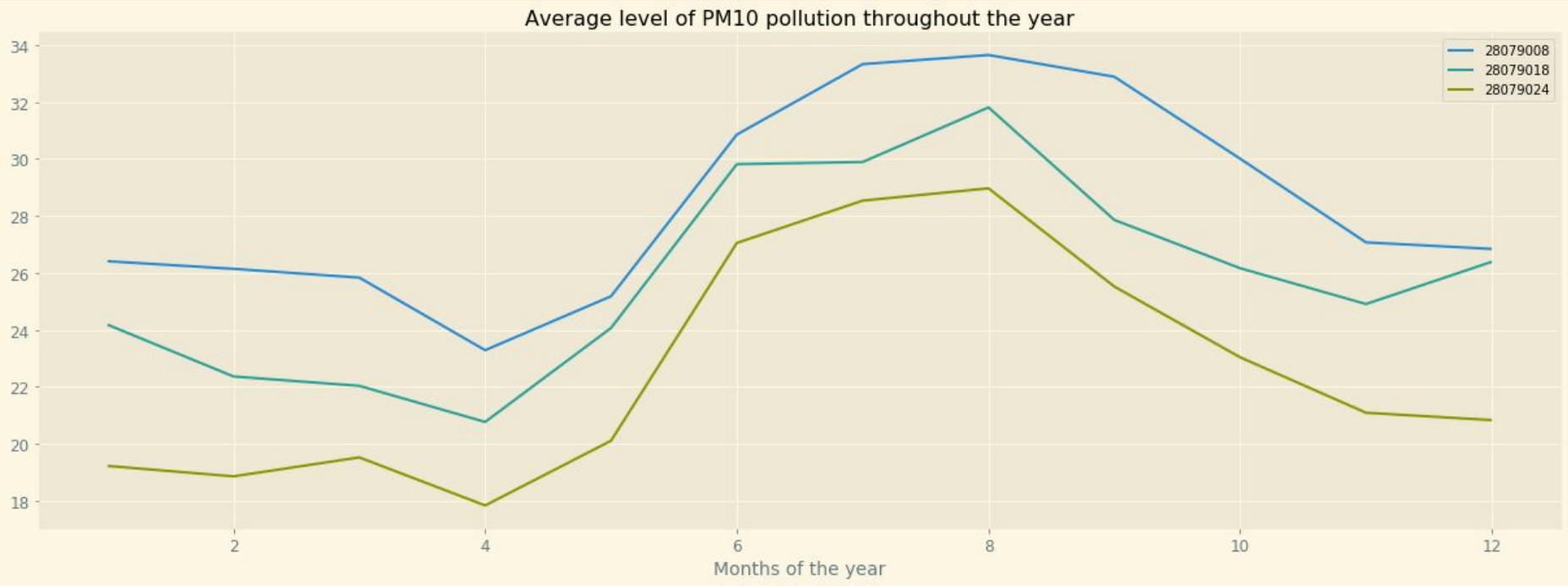

Average level of SO_2 pollution from 2001 to 2018
28079008
28079018
28079024
18
16
14
12
10
8
6
4
2
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years

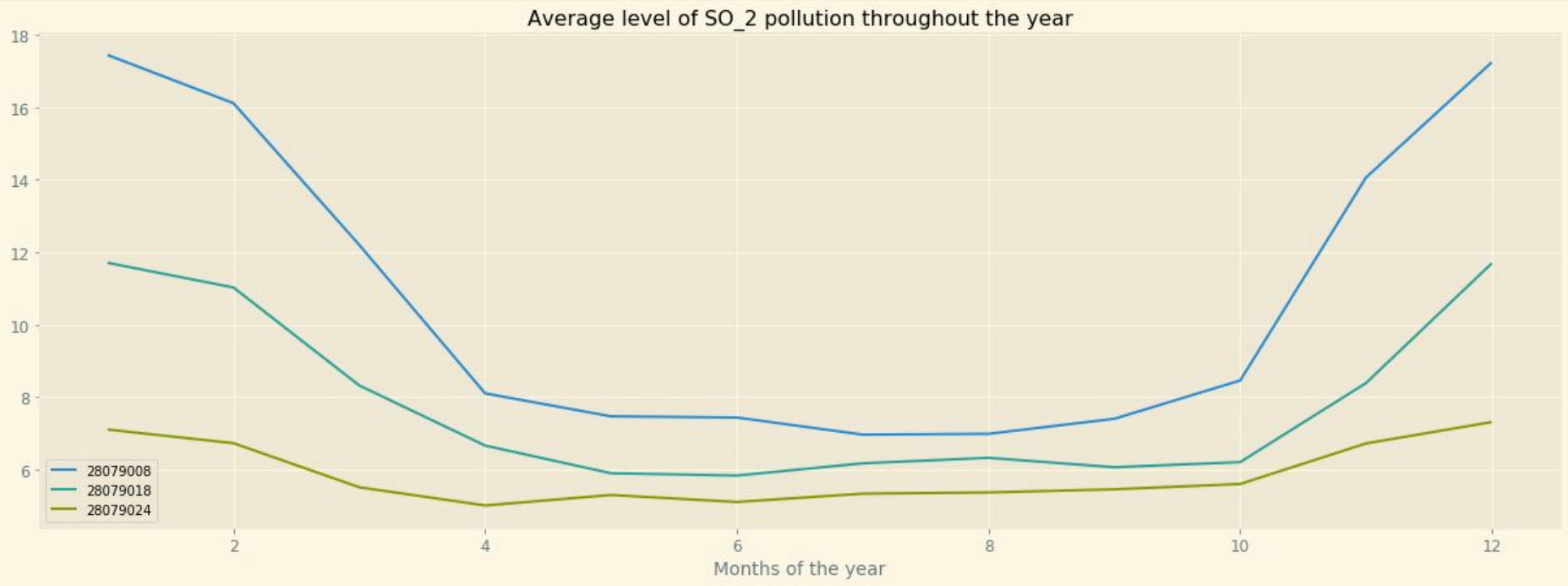

Average level of all_particles pollution from 2001 to 2018
160
140
120
100
80
60
40
20
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
Years
28079008
28079018
28079024

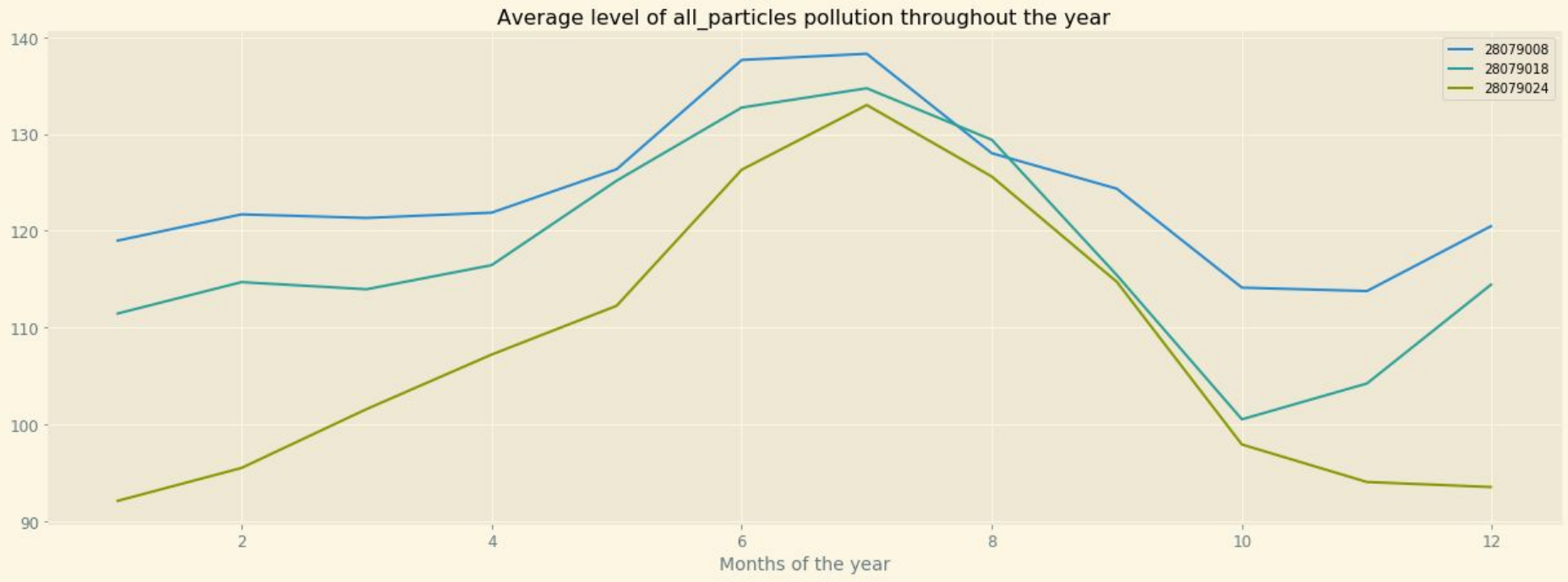
Average levels of pollution throughout the months of the year


Average level of NO_2 pollution throughout the year
70
60
50
40
30
20
2
4
6
8
10
12
Months of the year
28079008
28079018
28079024

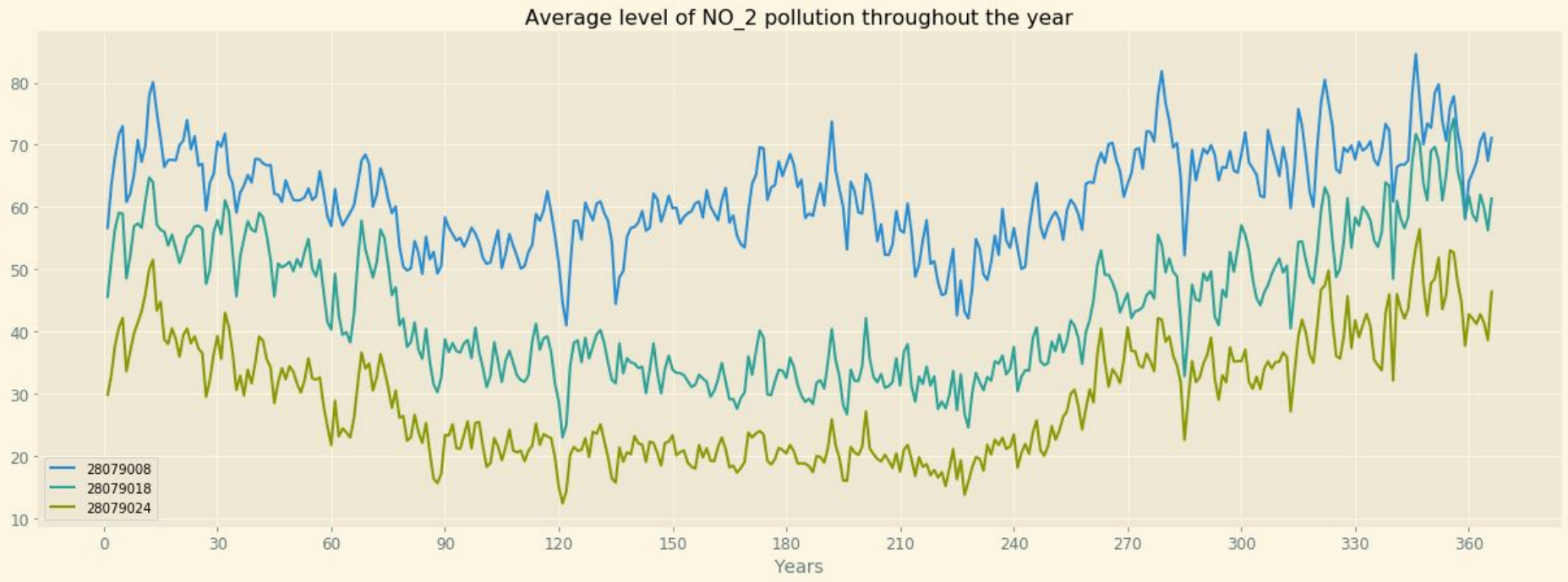

Average level of O_3 pollution throughout the year
28079008
28079018
28079024
80
70
60
50
40
30
20
2
4
6
8
10
12
Months of the year

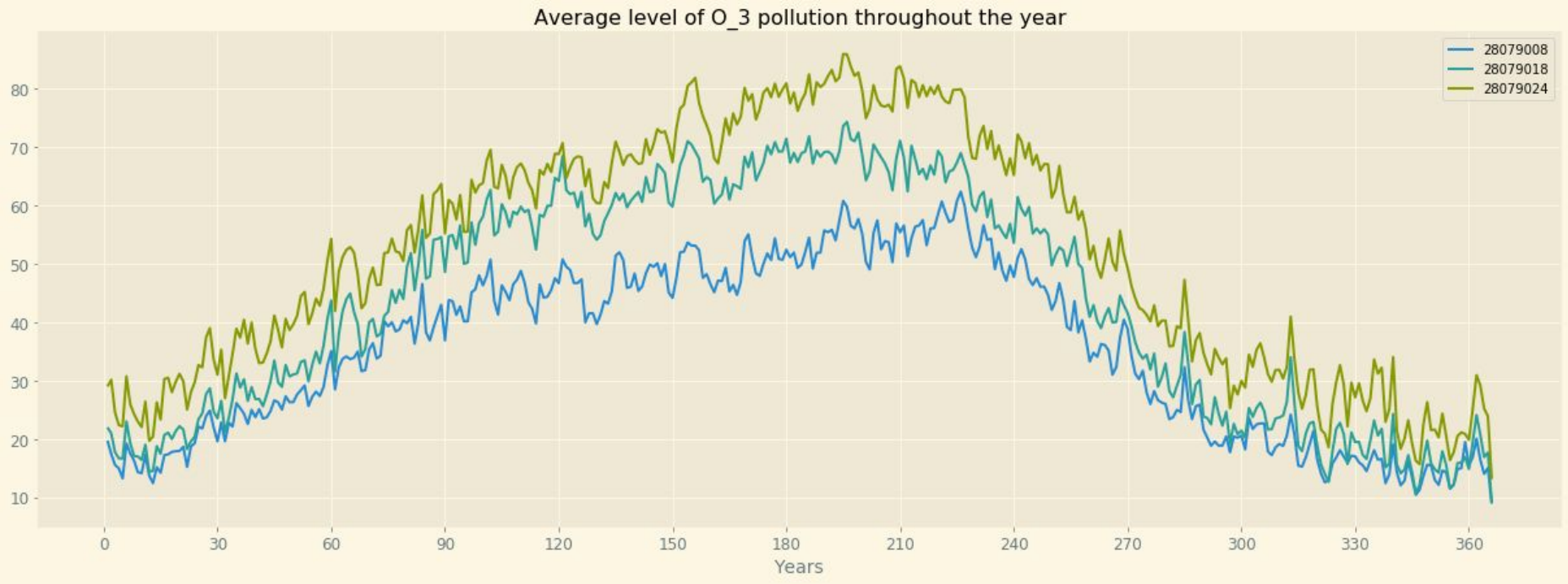

Average level of PM10 pollution throughout the year
28079008
28079018
28079024
34
32
30
28
26
24
22
20
18
2
4
6
8
10
12
Months of the year

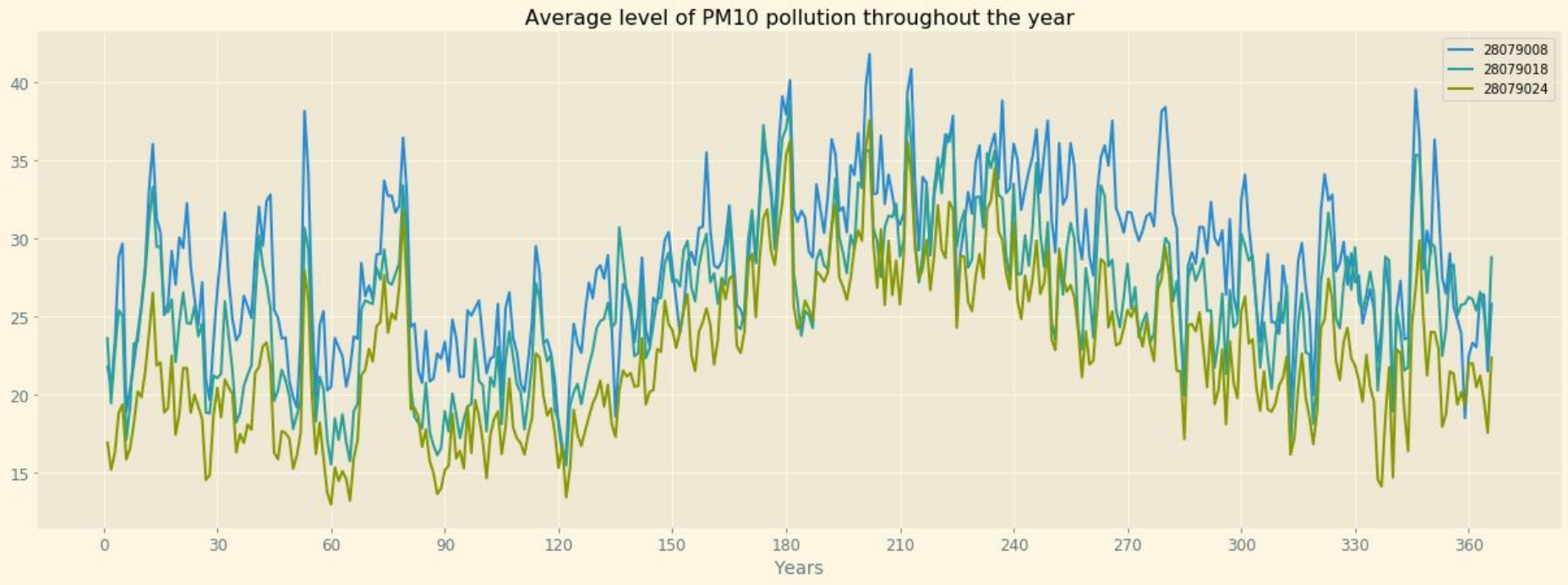

Average level of SO_2 pollution throughout the year
18
16
14
12
10
8
6
2
4
6
8
10
12
Months of the year
28079008
28079018
28079024

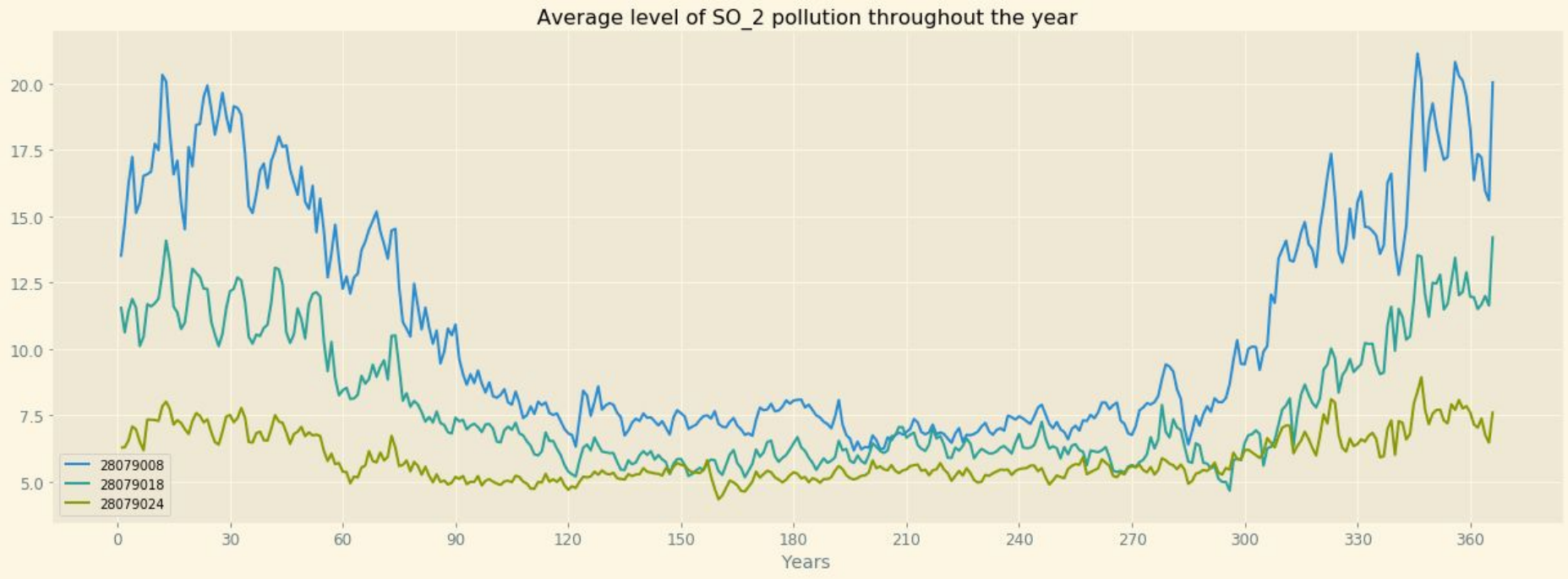

Average level of all_particles pollution throughout the year
140
130
120
110
100
90
2
4
6
8
10
12
Months of the year
28079008
28079018
28079024

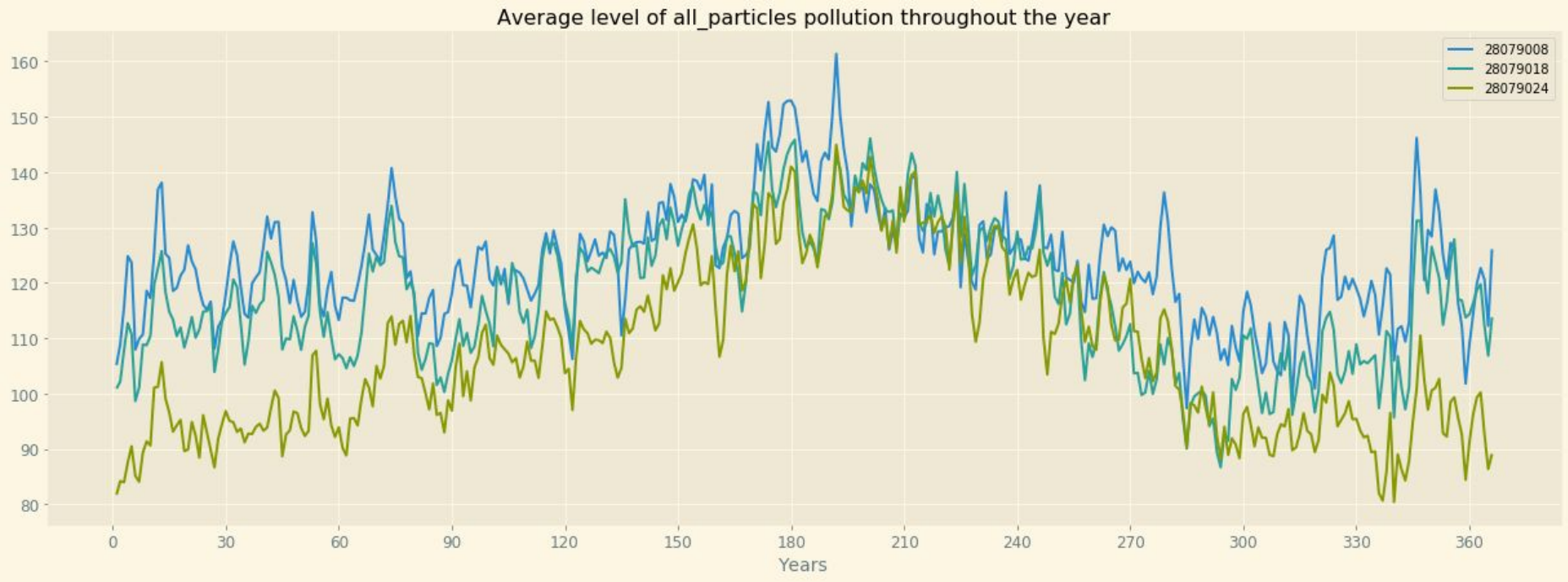
Average levels of pollution throughout the days of the year


Average level of NO_2 pollution throughout the year
80
70
60
50
40
30
20
10
0
30
60
90
120
150
180
210
240
270
300
330
360
Years
28079008
28079018
28079024

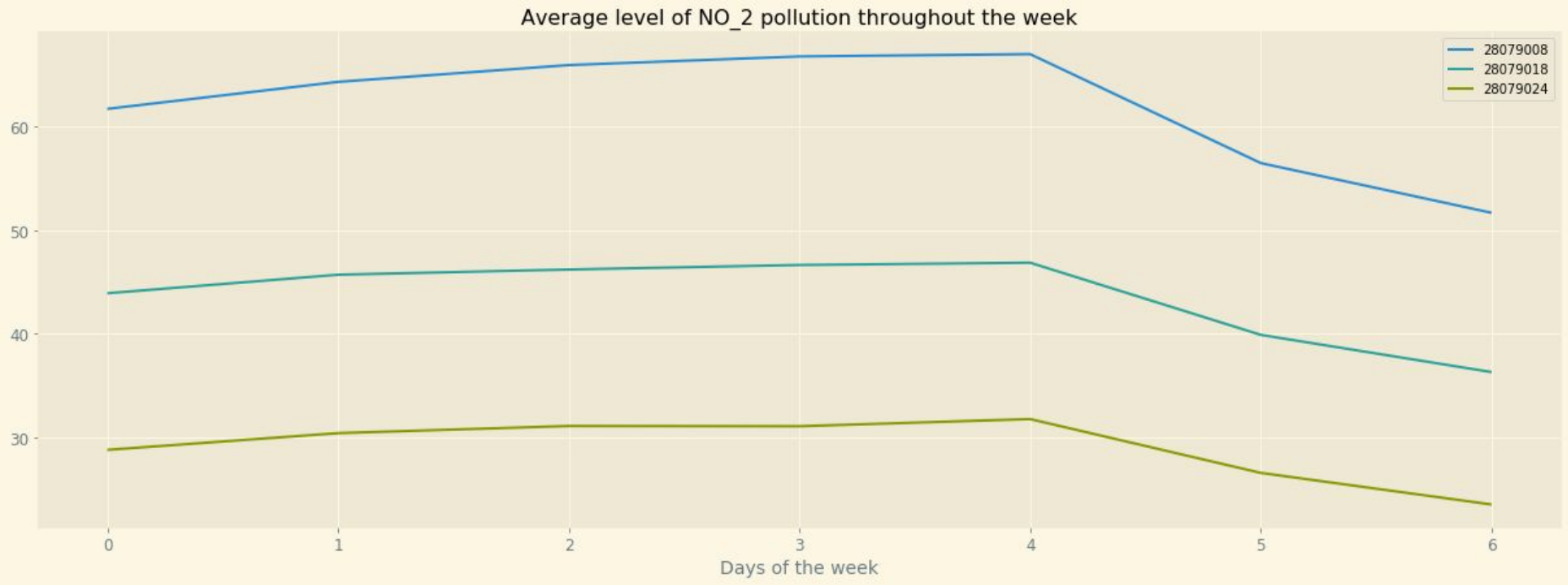

Average level of O_3 pollution throughout the year
28079008
28079018
28079024
80
70
60
50
40
30
20
10
0
30
60
90
120
150
180
210
240
270
300
330
360
Years

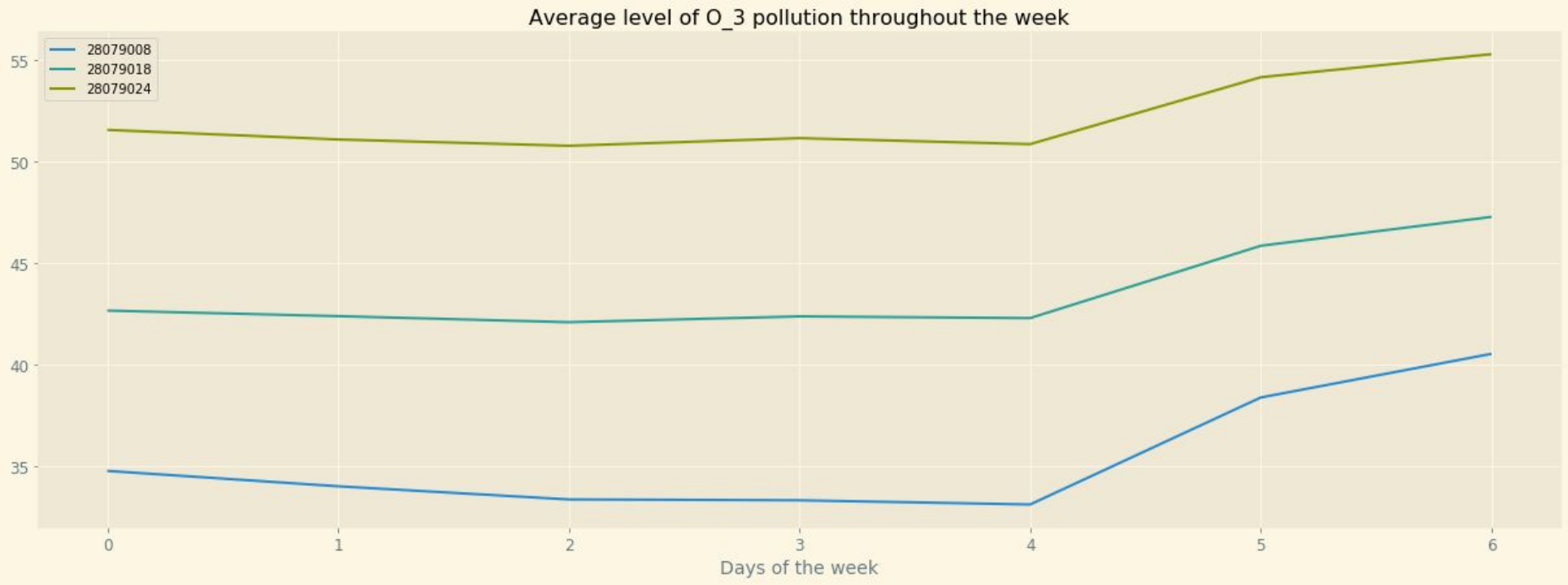

Average level of PM10 pollution throughout the year
28079008
28079018
28079024
40
35
30
25
20
15
0
30
60
90
120
150
180
210
240
270
300
330
360
Years

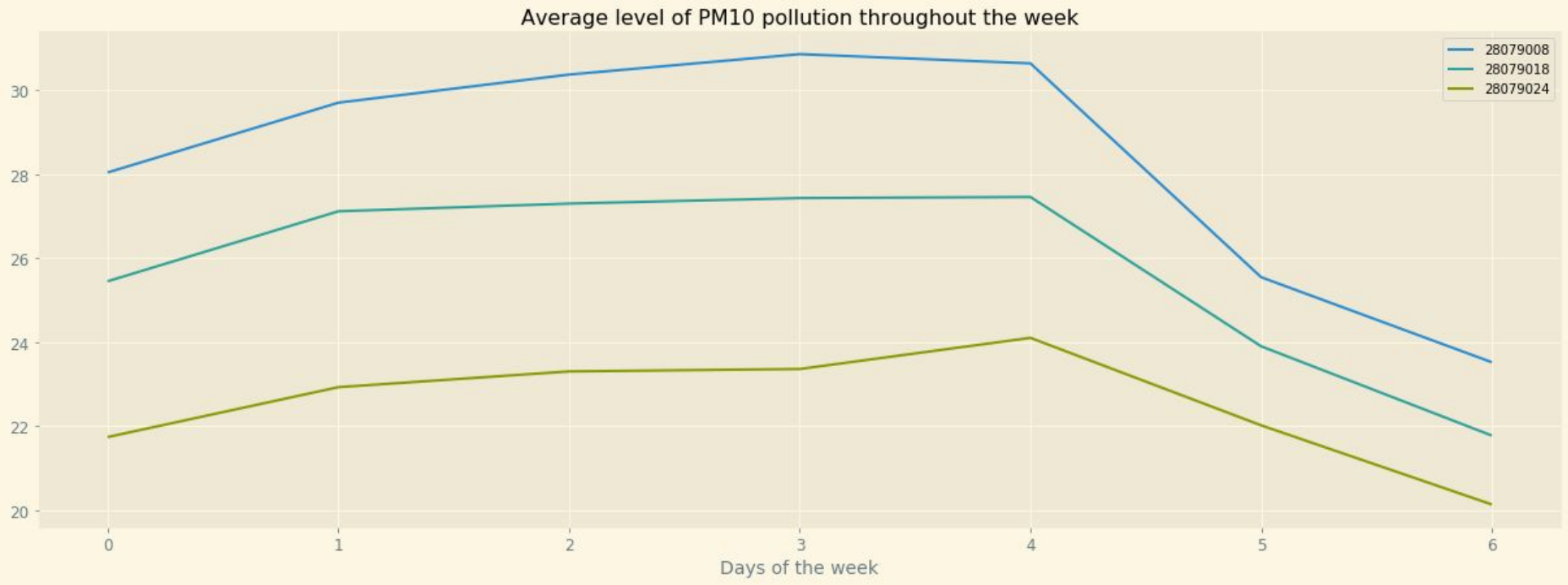

Average level of SO_2 pollution throughout the year
20.0
17.5
15.0
12.5
10.0
7.5
5.0
0
30
60
90
120
150
180
210
240
270
300
330
360
Years
28079008
28079018
28079024

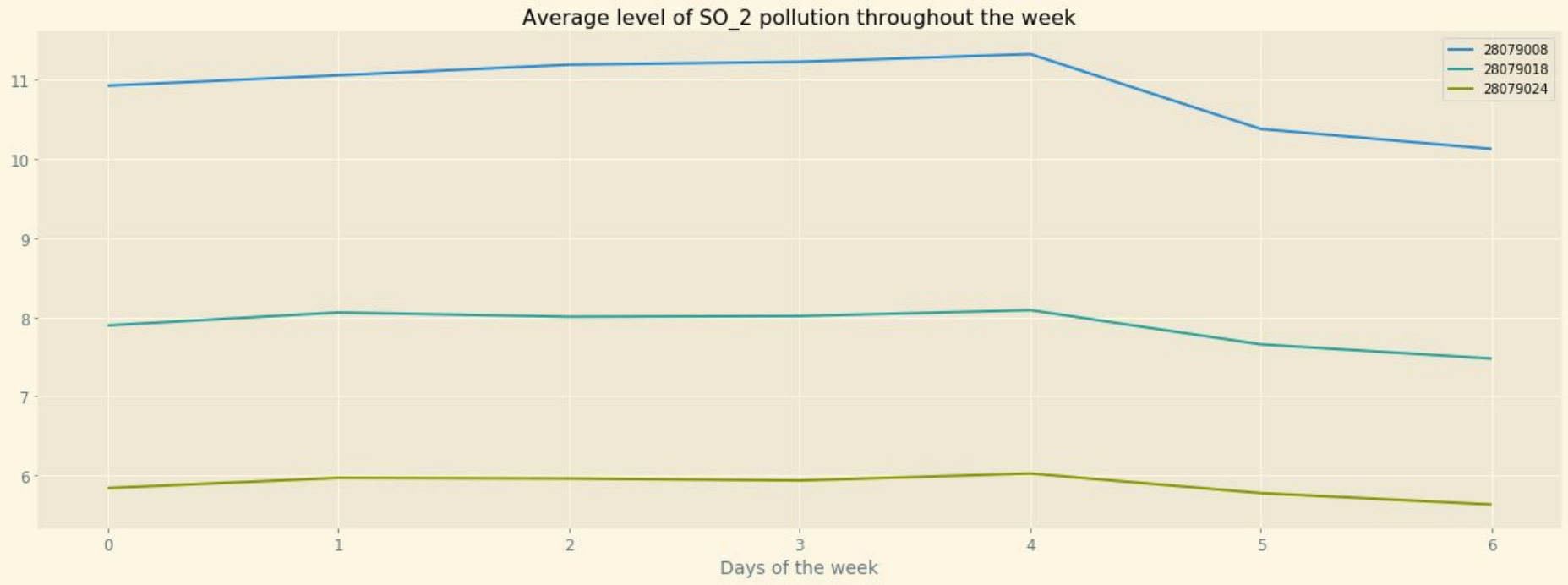

Average level of all_particles pollution throughout the year
28079008
28079018
28079024
160
150
140
130
120
110
100
90
80
0
30
60
90
120
150
180
210
240
270
300
330
360
Years

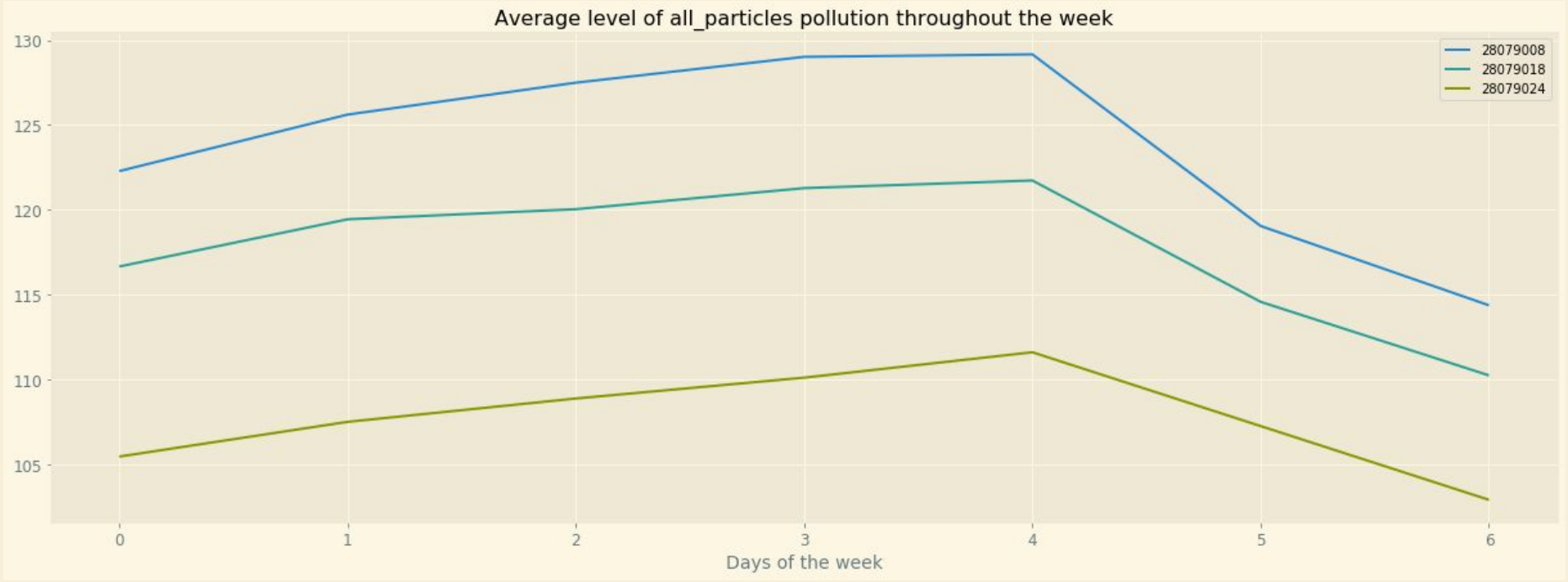
Average levels of pollution throughout the week


Average level of NO_2 pollution throughout the week
28079008
28079018
28079024
60
50
40
30
0
1
2
3
4
5
6
Days of the week

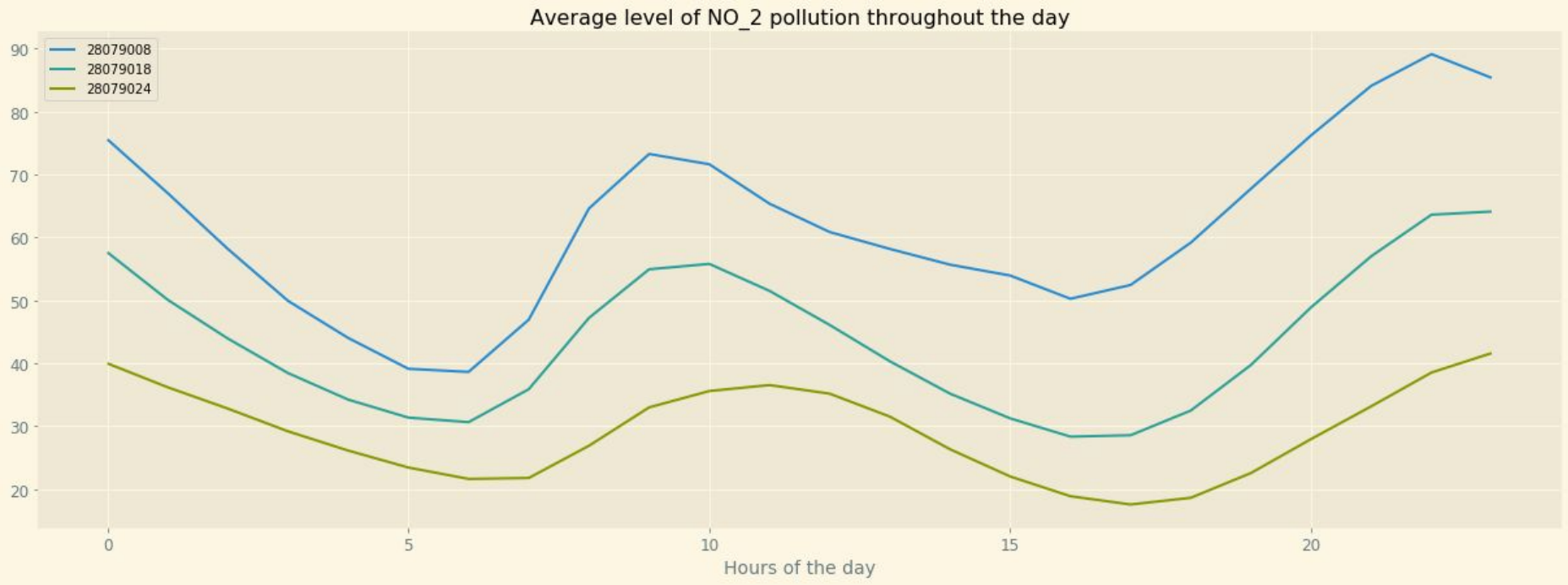

Average level of O_3 pollution throughout the week
28079008
28079018
28079024
55
50
45
40
35
0
1
2
3
4
5
6
Days of the week

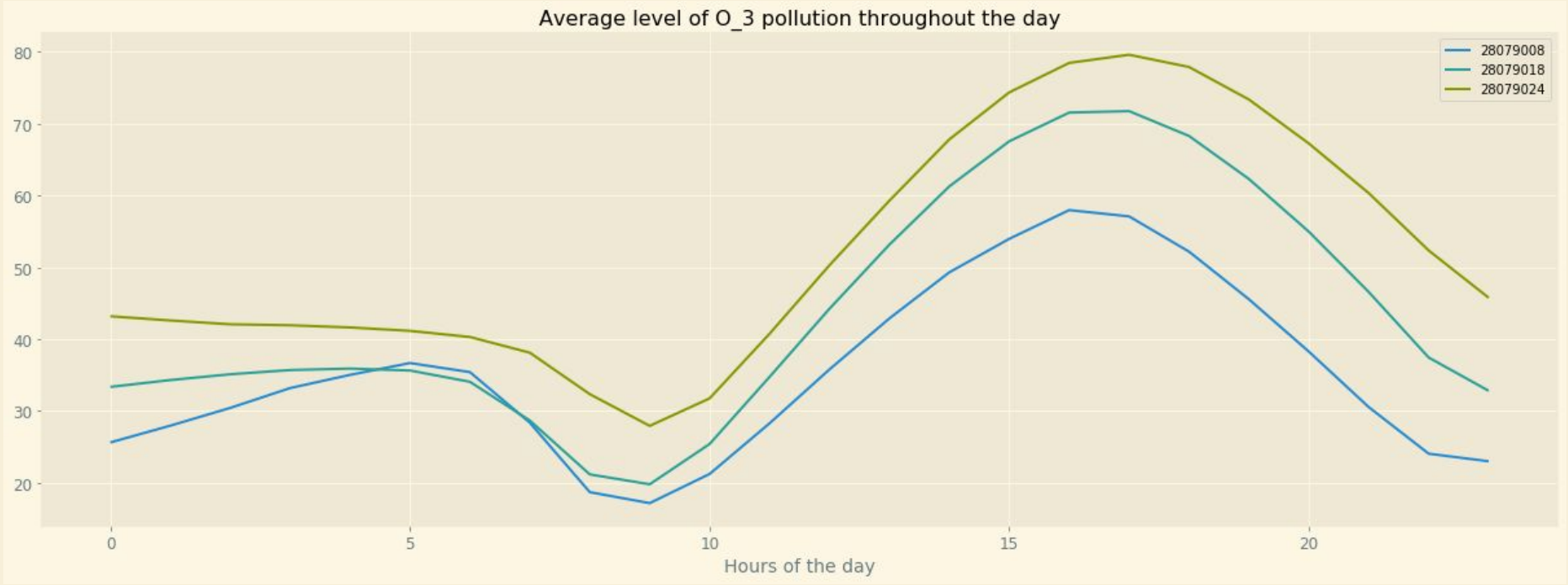

Average level of PM10 pollution throughout the week
28079008
28079018
28079024
30
28
26
24
22
20
0
1
2
3
4
5
6
Days of the week

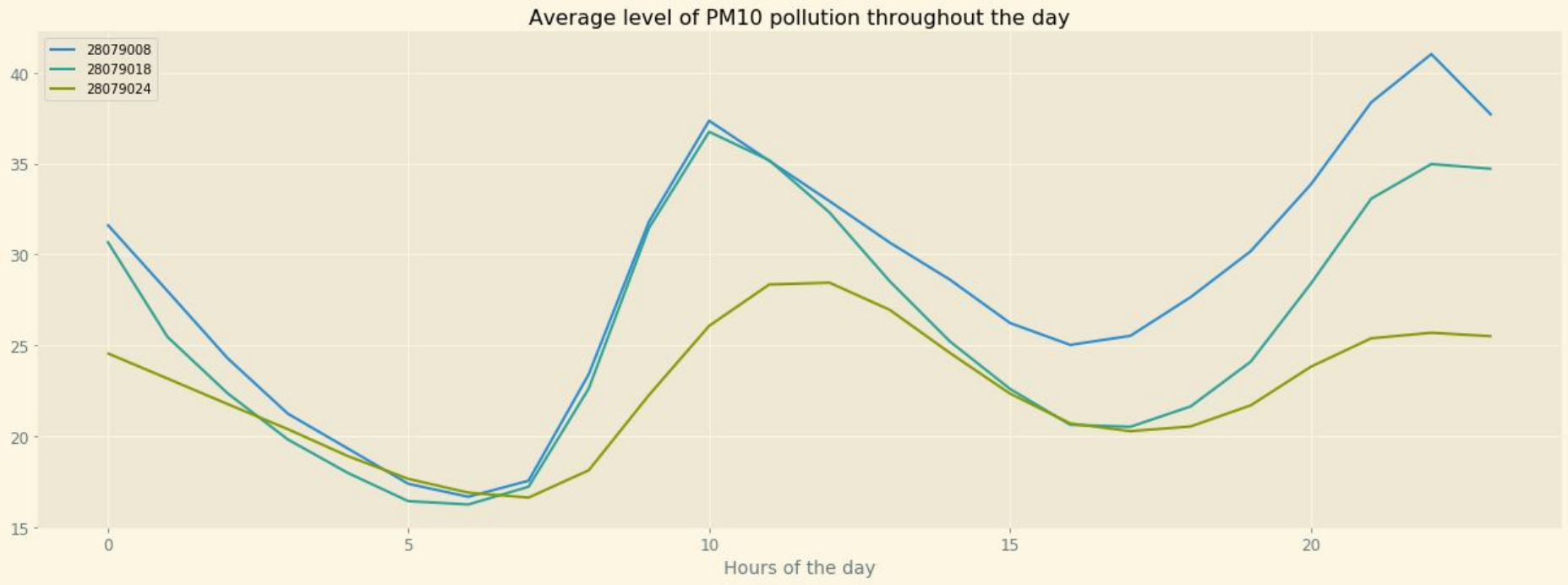

Average level of SO_2 pollution throughout the week
28079008
28079018
28079024
11
10
9
8
7
6
0
1
2
3
4
5
6
Days of the week

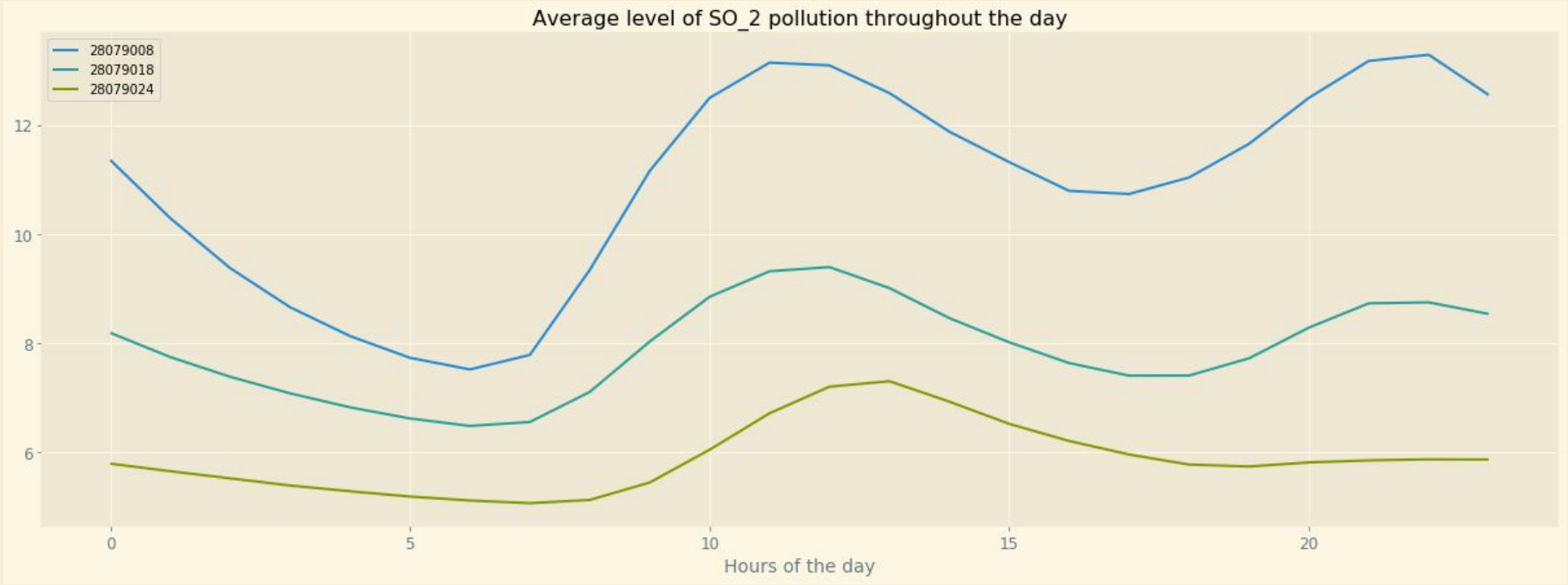

Average level of all_particles pollution throughout the week
130
125
120
115
110
105
0
1
2
3
4
5
6
Days of the week
28079008
28079018
28079024

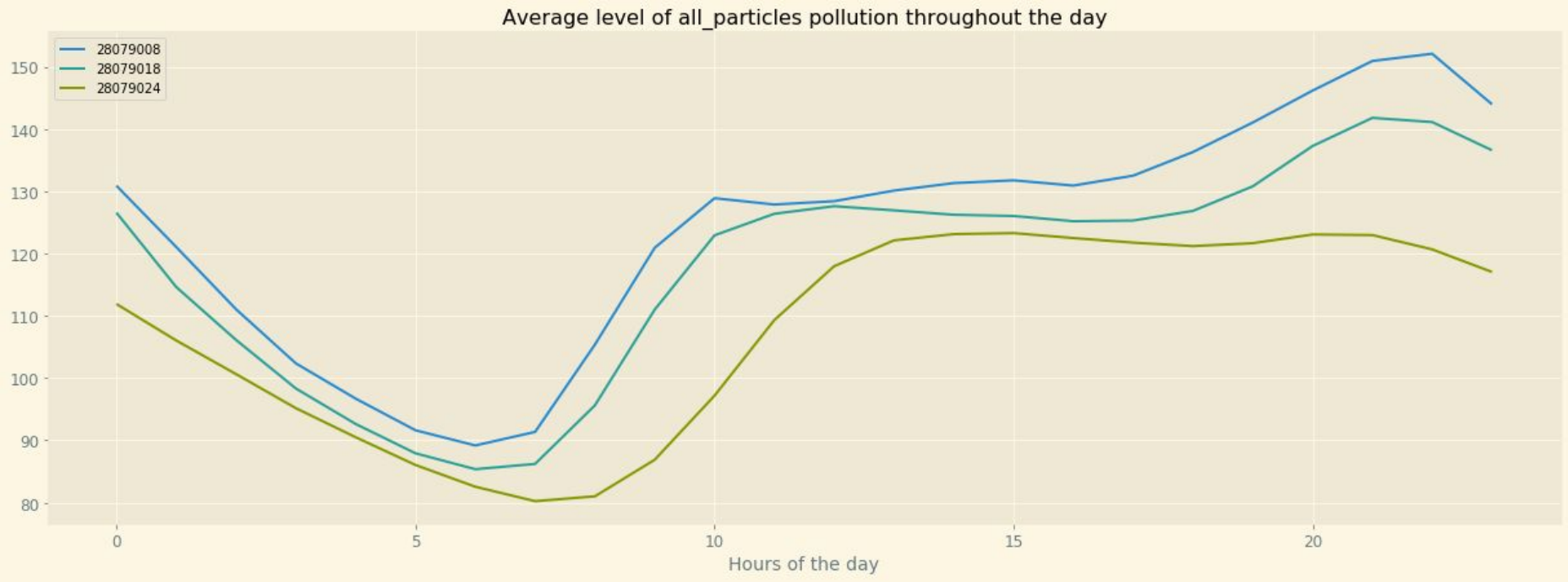
Average levels of pollution throughout the day


Average level of NO_2 pollution throughout the day
28079008
28079018
28079024
90
80
70
60
50
40
30
20
0
5
10
15
20
Hours of the day

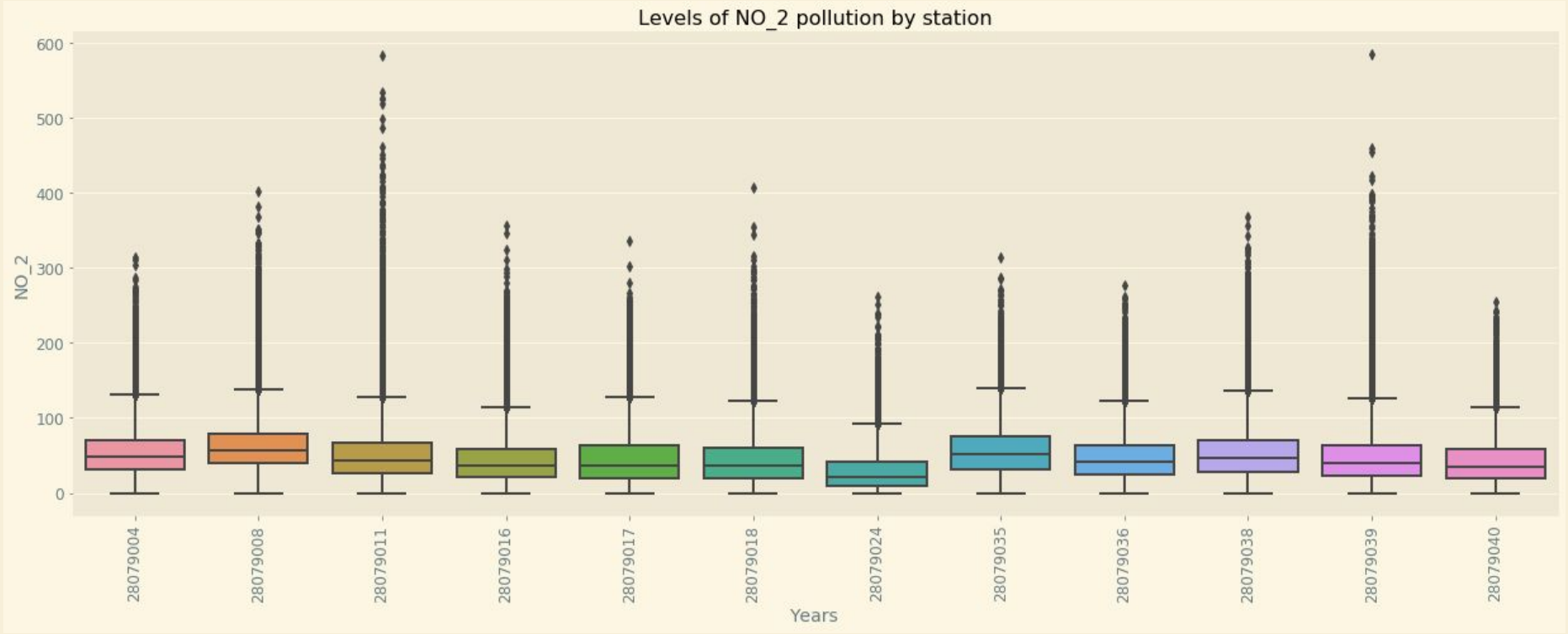

Average level of O_3 pollution throughout the day
80
70
60
50
40
30
20
0
5
10
15
20
Hours of the day
28079008
28079018
28079024

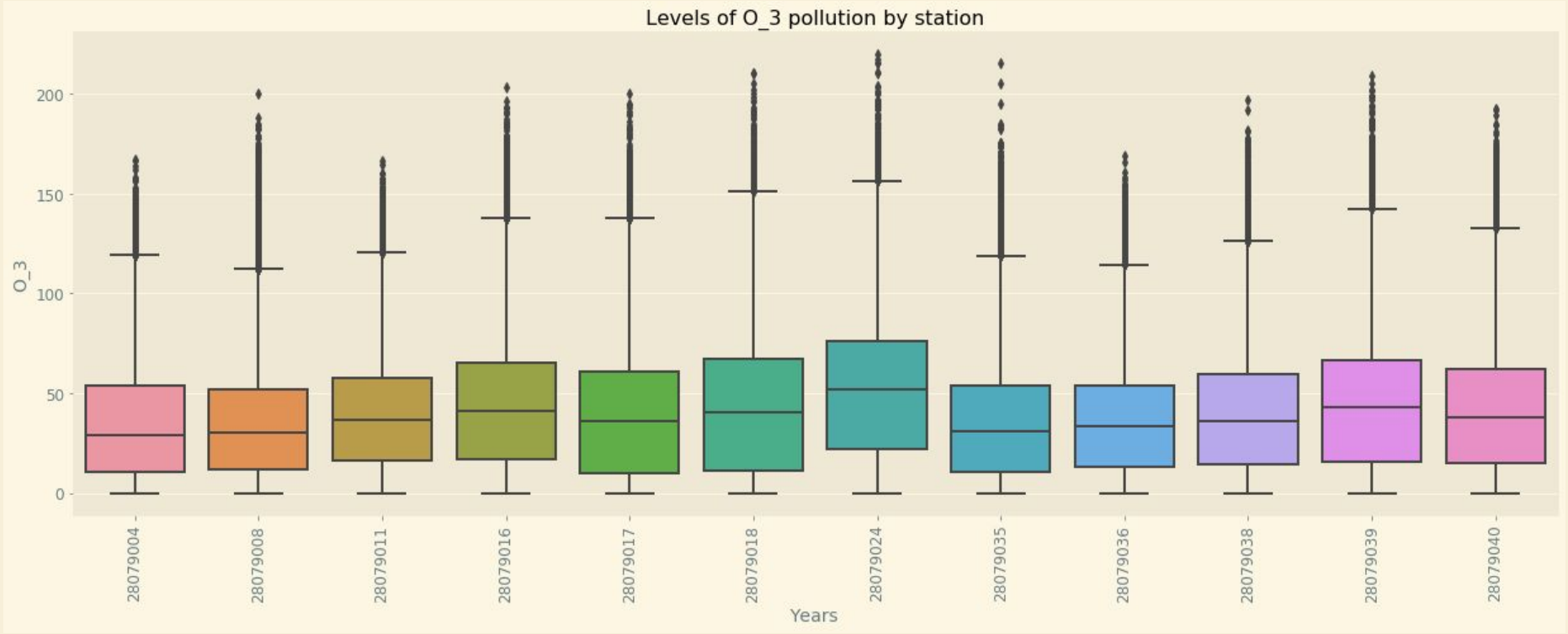

Average level of PM10 pollution throughout the day
28079008
28079018
28079024
40
35
30
25
20
15
0
5
10
15
20
Hours of the day

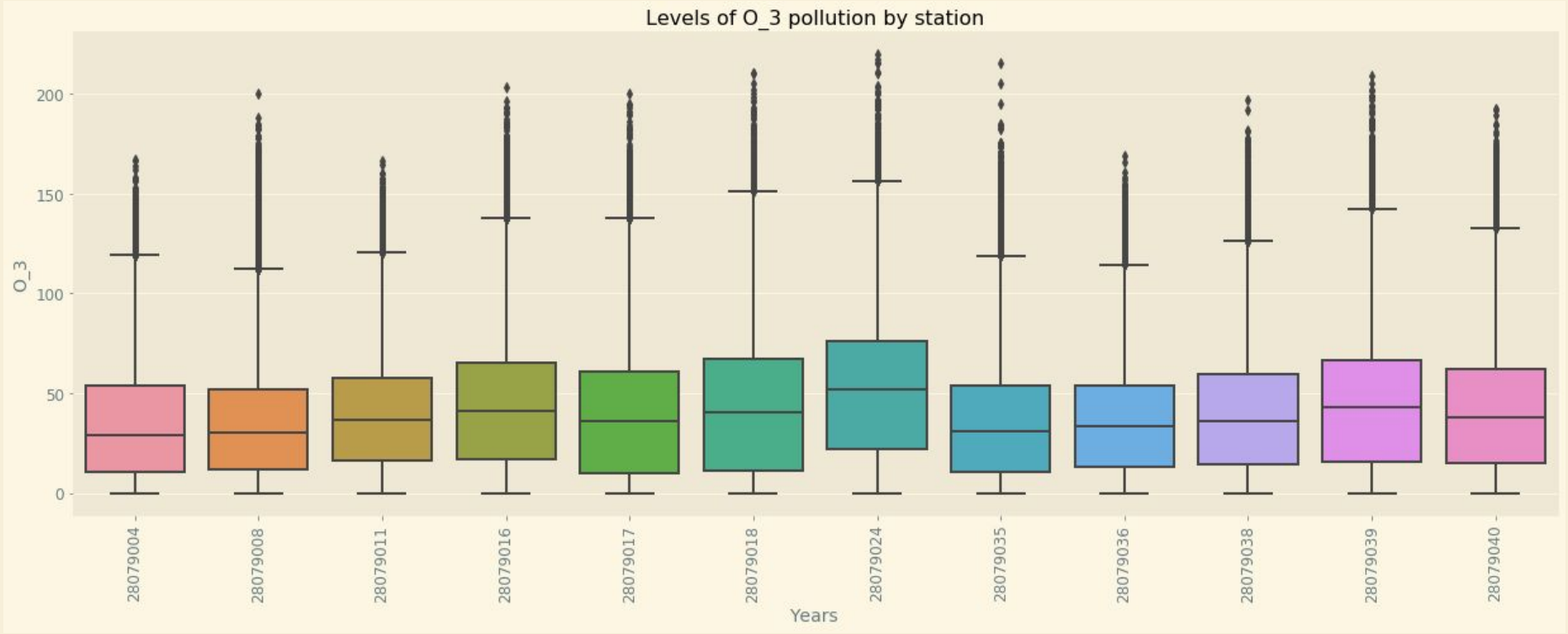

Average level of SO_2 pollution throughout the day
28079008
28079018
28079024
12
10
8
6
0
5
10
15
20
Hours of the day

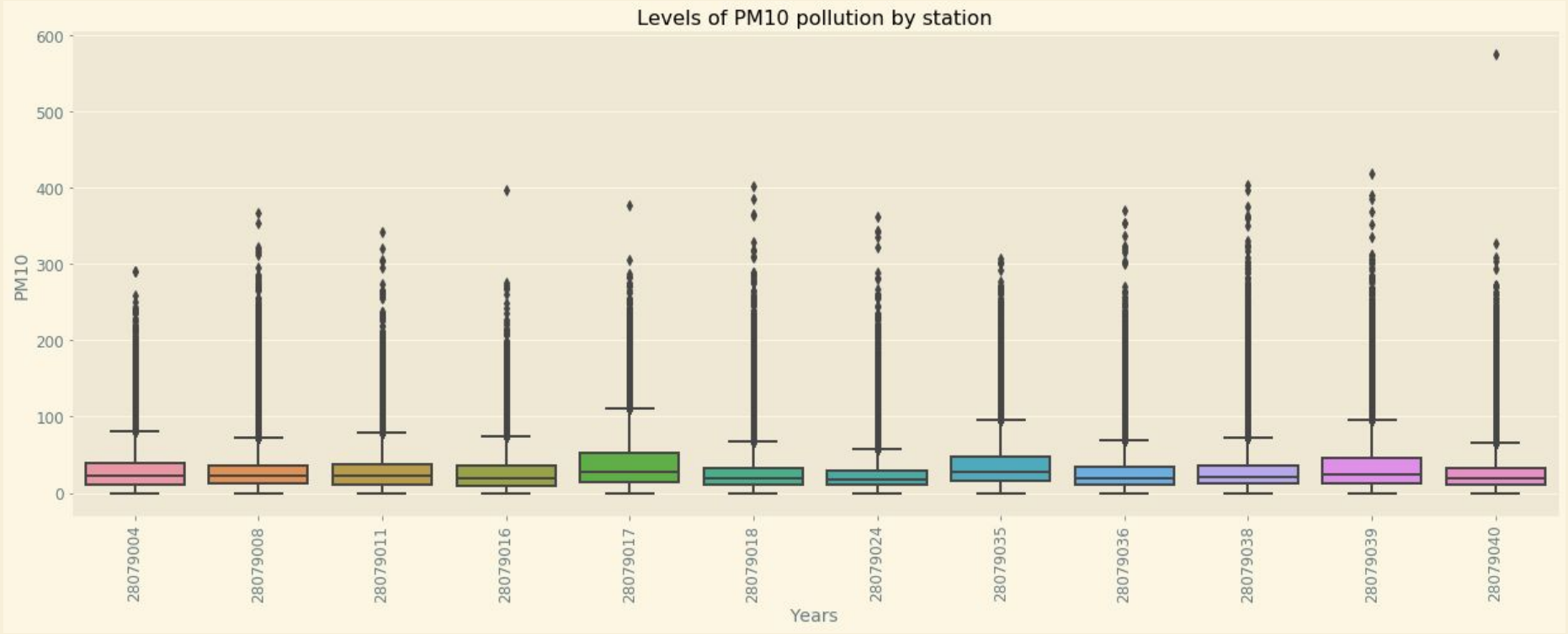

Average level of all_particles pollution throughout the day
28079008
28079018
28079024
150
140
130
120
110
100
90
80
0
5
10
15
20
Hours of the day

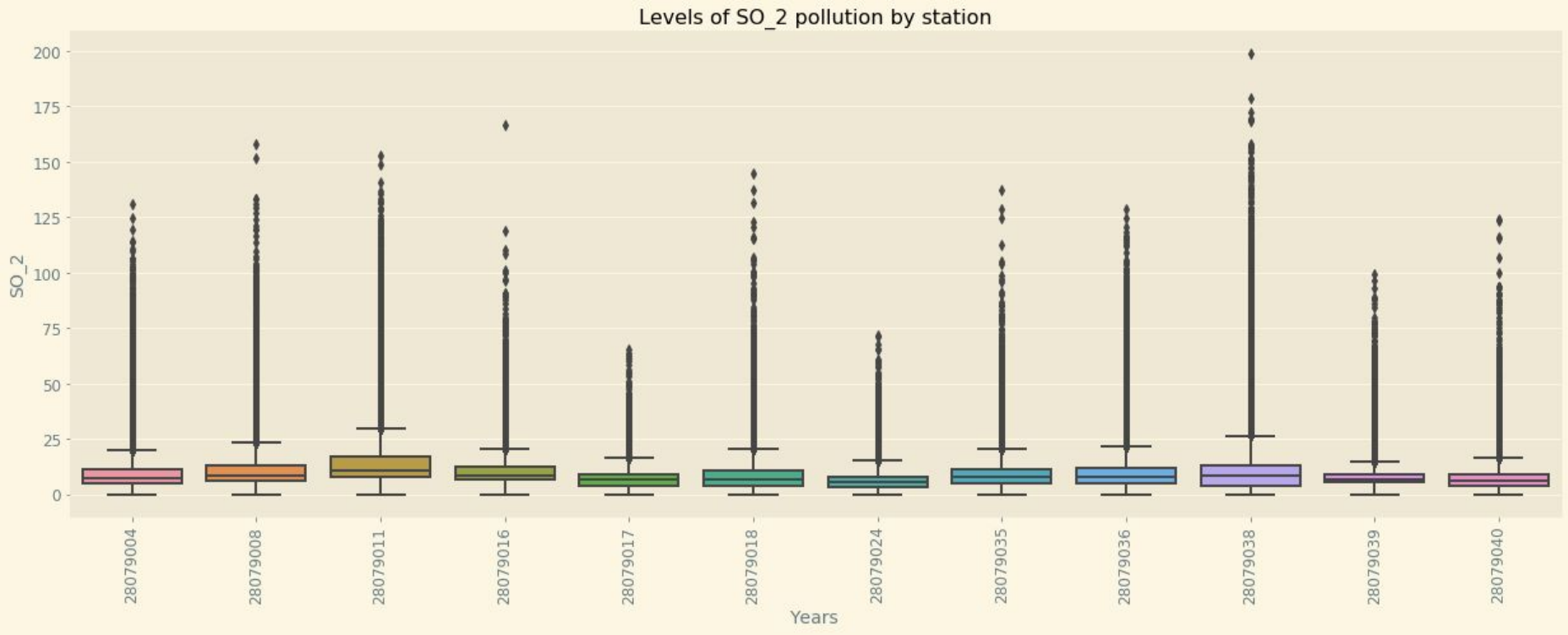
Levels of pollution across different stations


Levels of NO_2 pollution by station
NO_2
600
500
400
300
200
100
0
28079004
28079008
28079011
28079016
28079017
28079018
28079024
28079035
28079036
28079038
28079039
28079040
Years


Levels of O_3 pollution by station
O_3
200
150
100
50
0
28079004
28079008
28079011
28079016
28079017
28079018
28079024
28079035
28079036
28079038
28079039
28079040
Years


Levels of O_3 pollution by station
O_3
200
150
100
50
0
28079004
28079008
28079011
28079016
28079017
28079018
28079024
28079035
28079036
28079038
28079039
28079040
Years

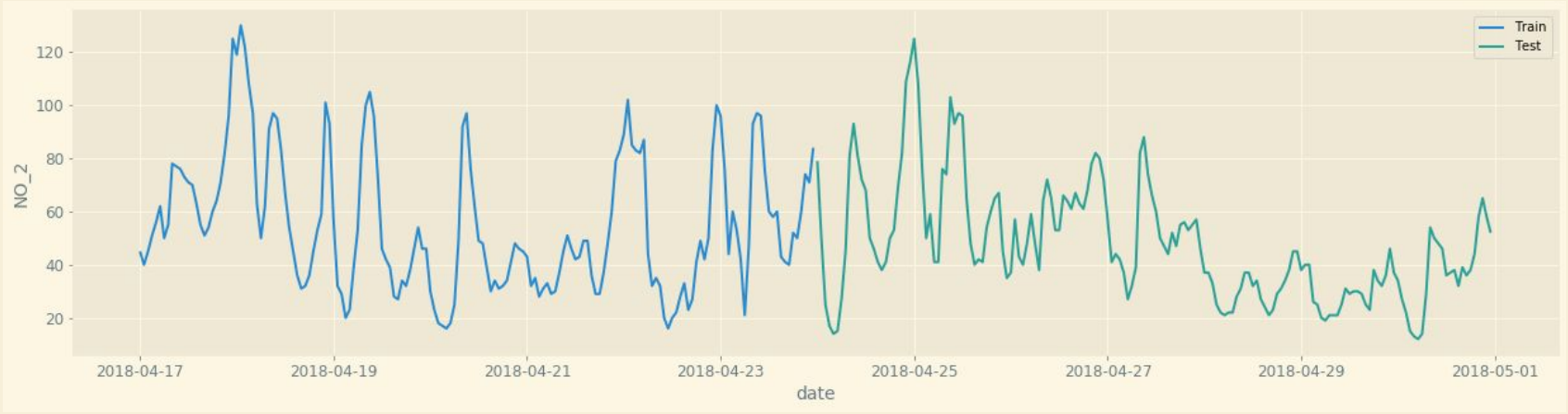

Levels of PM10 pollution by station
PM10
600
500
400
300
200
100
0
28079004
28079008
28079011
28079016
28079017
28079018
28079024
28079035
28079036
28079038
28079039
28079040
Years

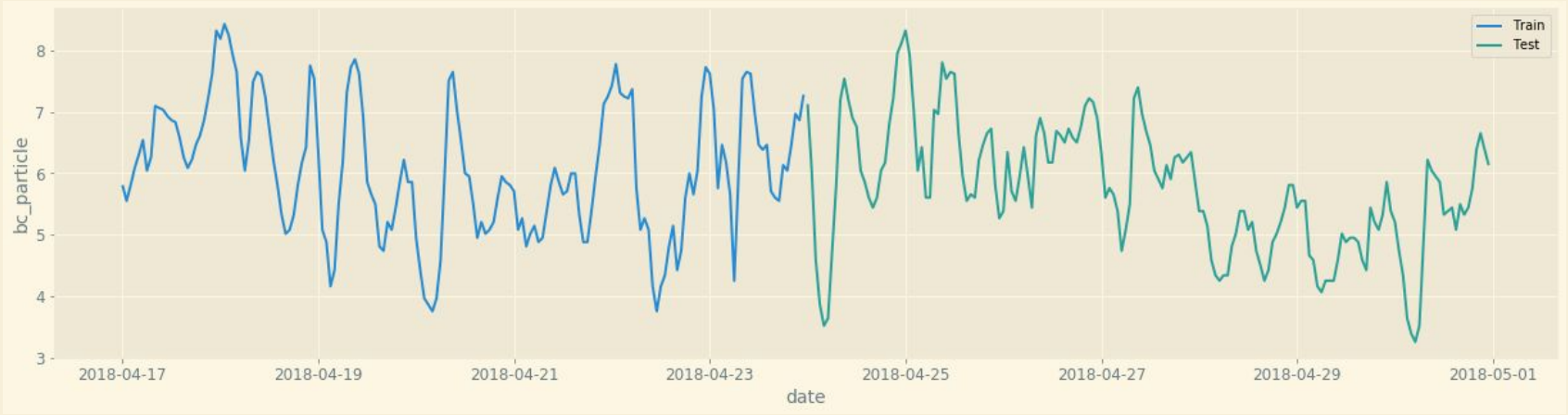

Levels of SO_2 pollution by station
SO_2
200
175
150
125
100
75
50
25
0
28079004
28079008
28079011
28079016
28079017
28079018
28079024
28079035
28079036
28079038
28079039
28079040
Years

Maps

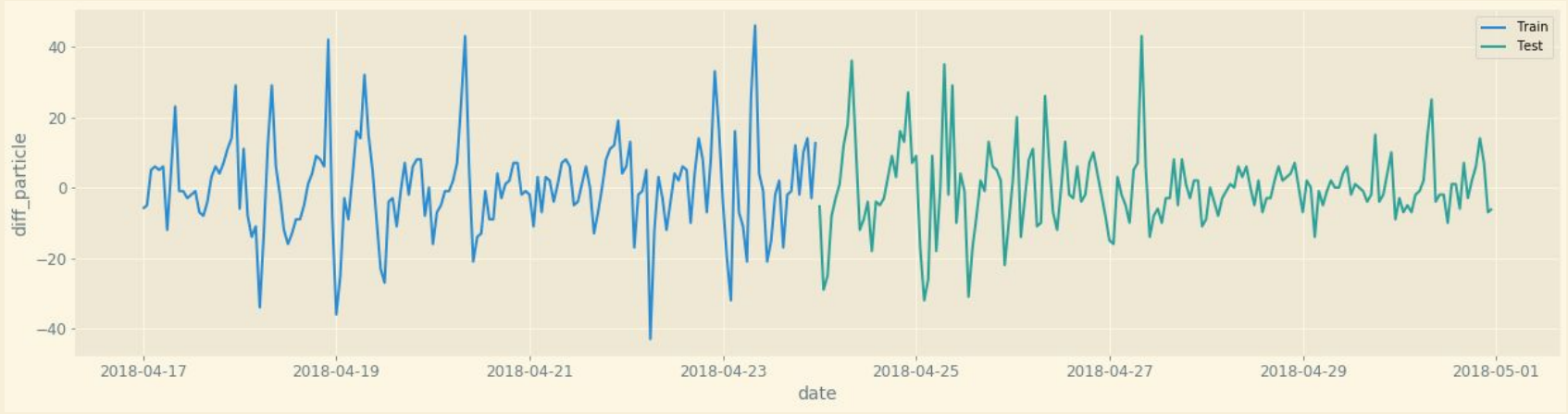
Target transformations

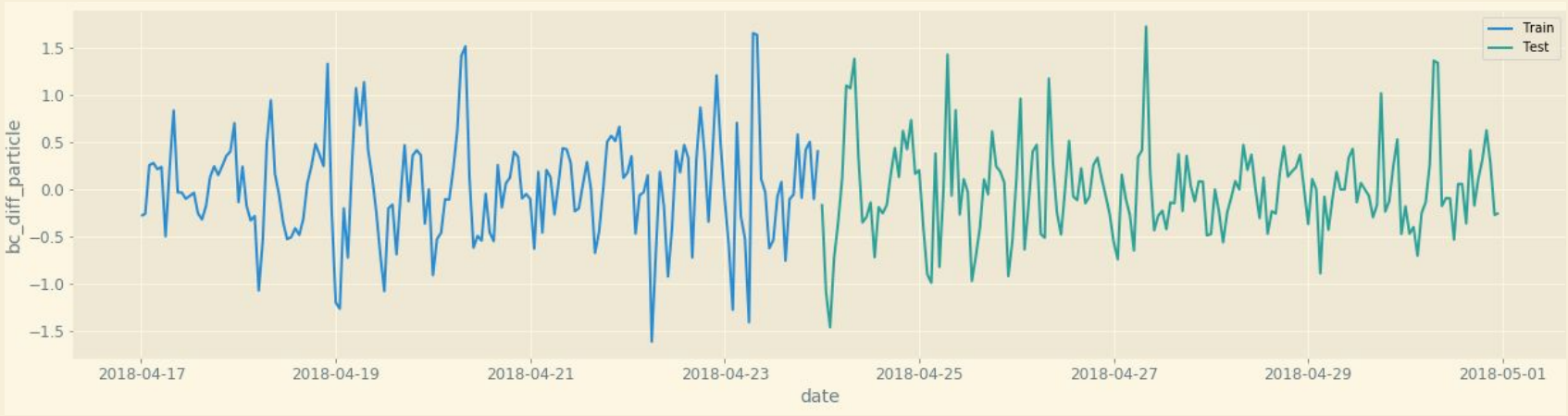
Box Cox Transformation





Difference & Box Cox Transformation

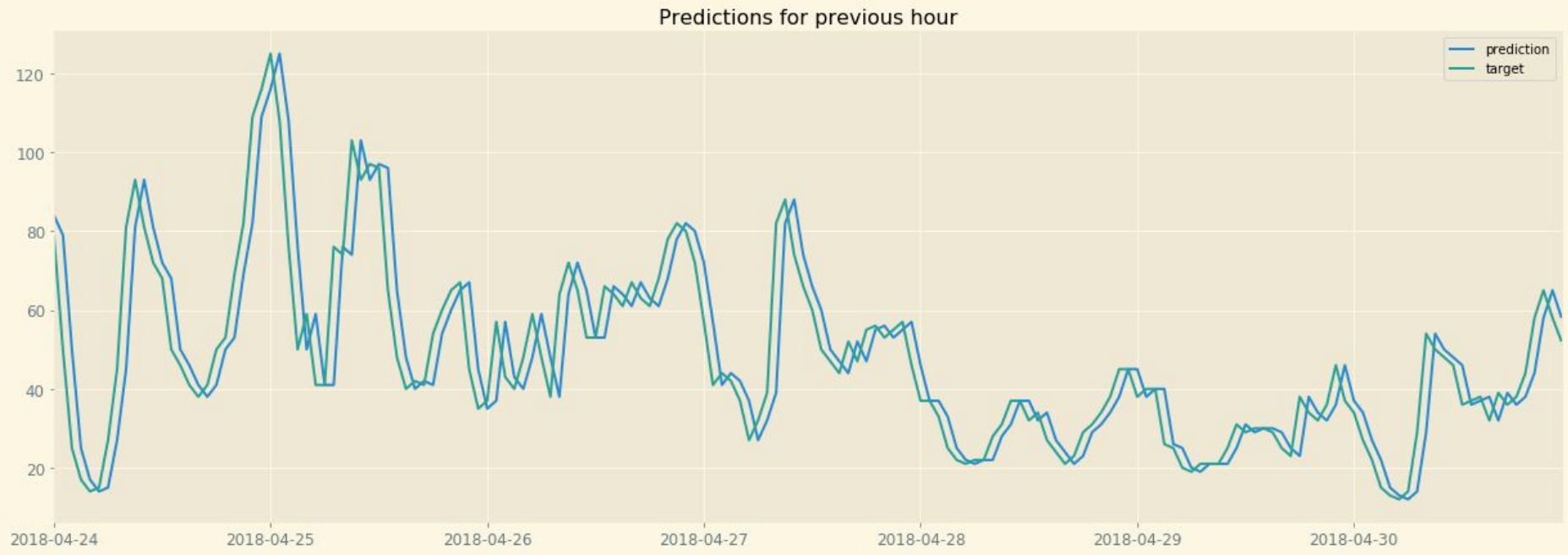




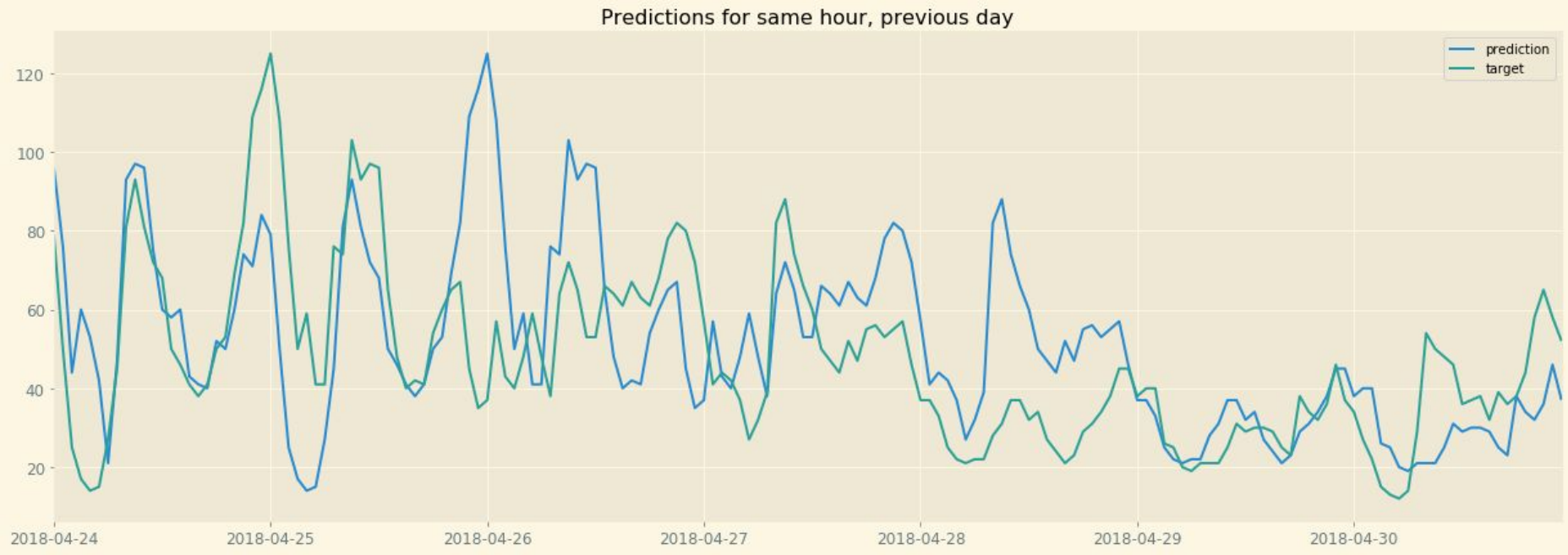
Basic Models

models	r2_score	mean_abs_error	median_abs_error	mse	rmse
previous hour	0.746614	7.773810	5.500000	123.452381	11.110913
same hour, previous day	-0.073556	16.880952	13.000000	523.047619	22.870234
same hour, previous week	-0.285305	19.583333	14.000000	626.214286	25.024274
rolling average	0.606051	9.854167	6.500000	191.936012	13.854097
exponential rolling average	0.584022	10.318985	7.176439	202.668927	14.236184

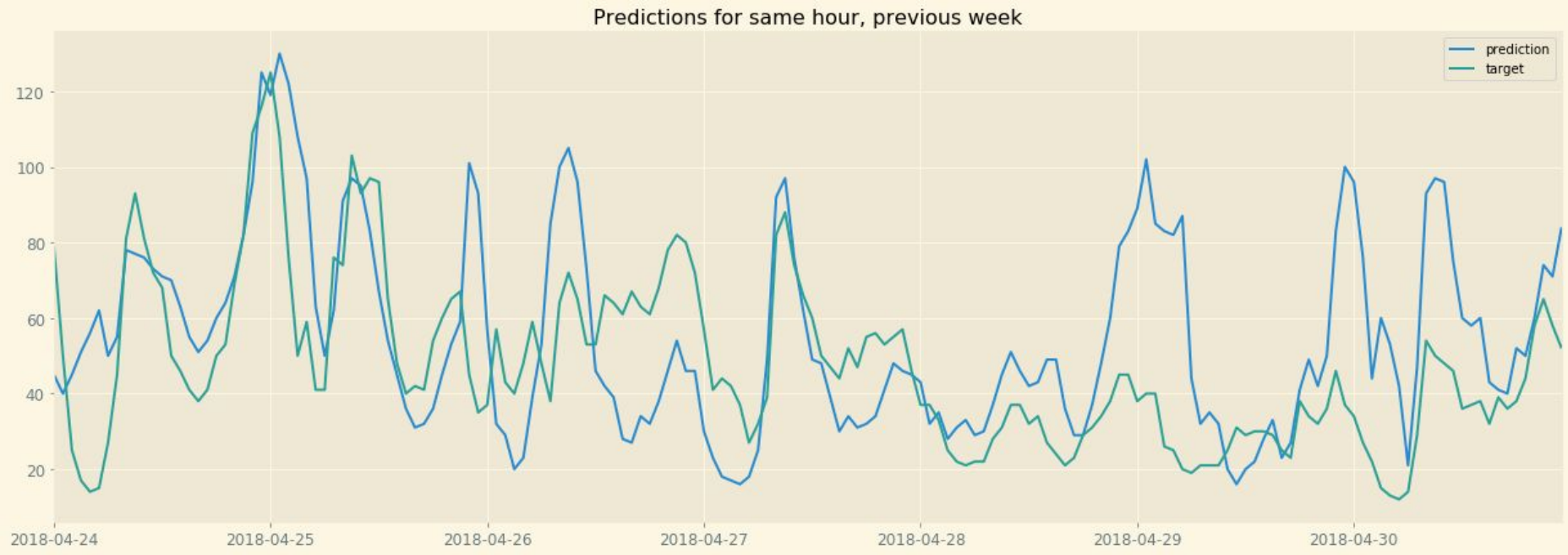
One hour shift



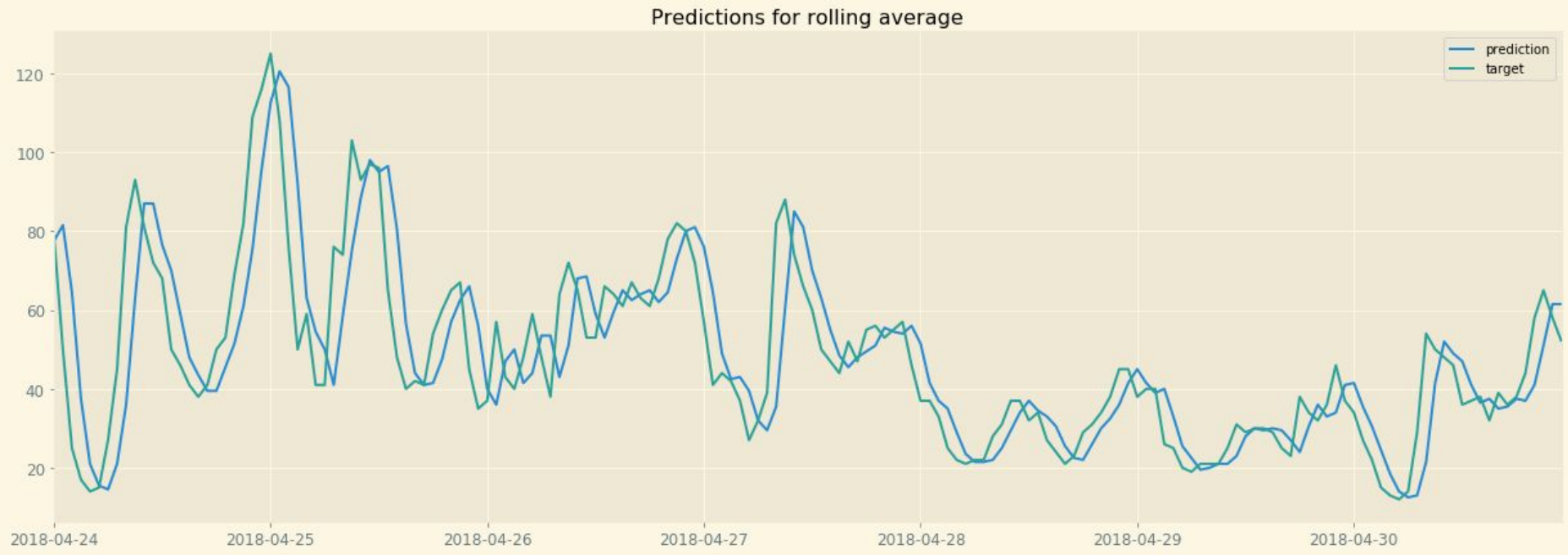
24 hour shift



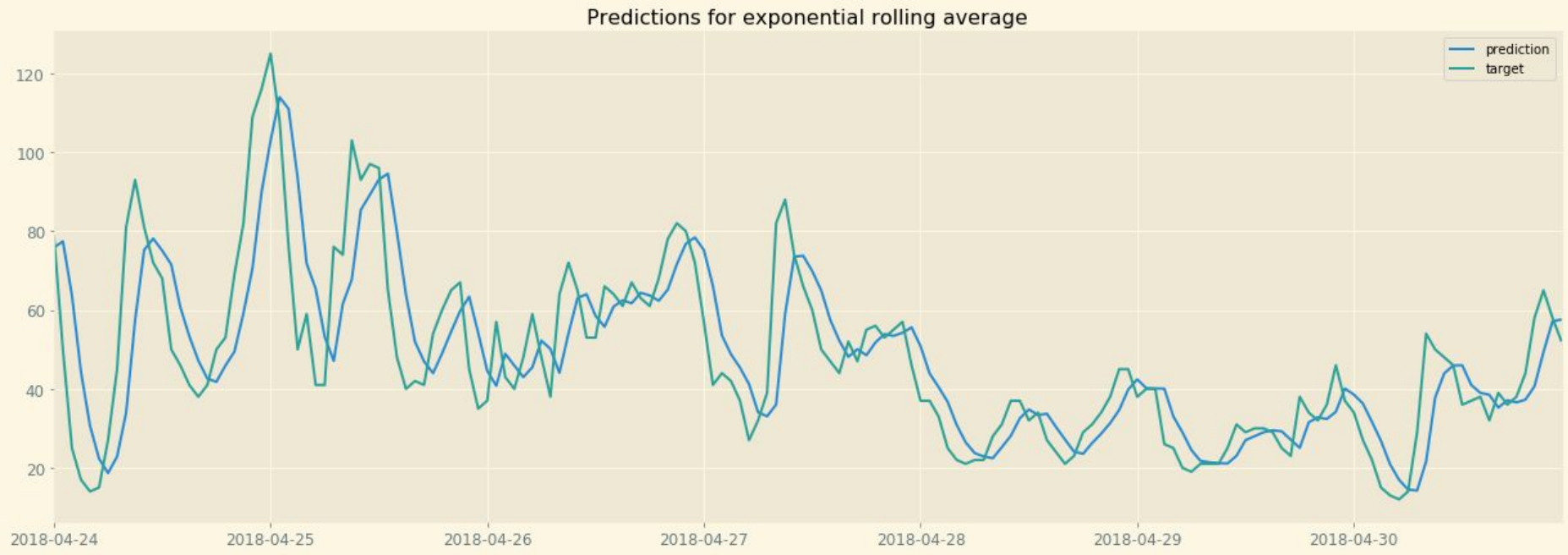
One week shift



Rolling Average, window=2



Exponential Rolling Average, halflife=1

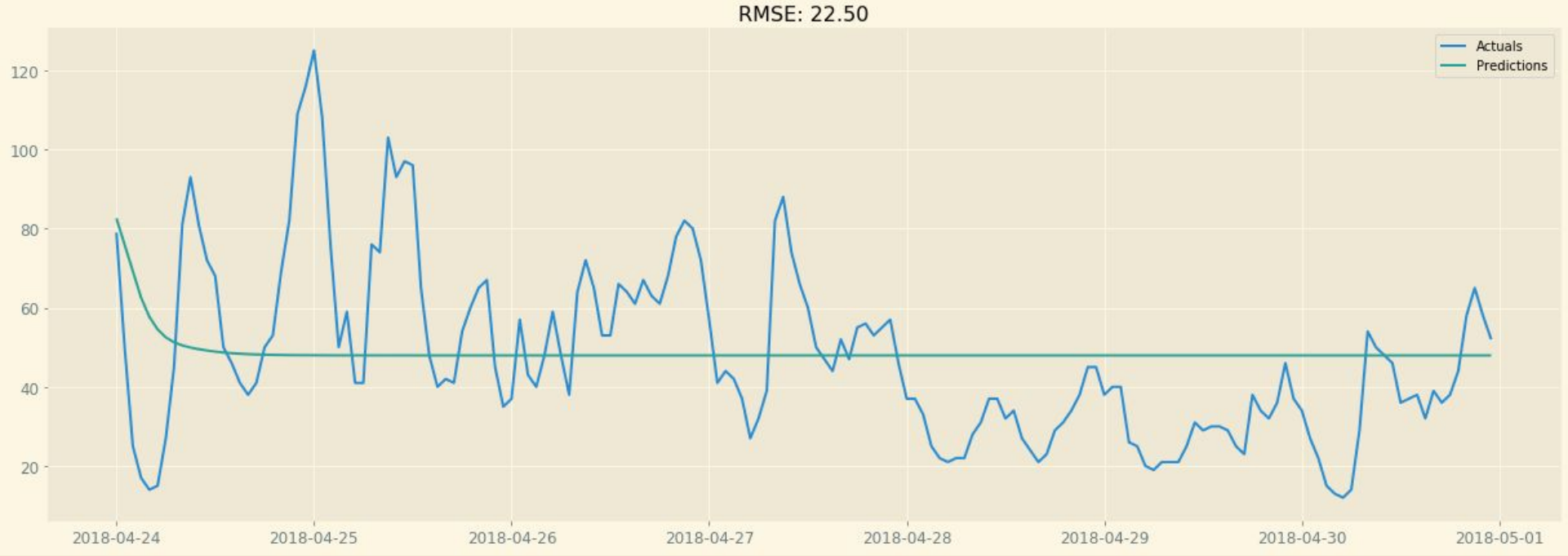


Arima

	models	r2_score	mean_abs_error	median_abs_error	mse	rmse
	ARIMA, order=(5, 0, 0), target: bc_particle	-0.039023	17.699984	15.000000	506.222909	22.499398
	ARIMA, order=(5, 1, 0), target: bc_particle	-1.761423	32.776925	34.945345	1345.394317	36.679617
	ARIMA, order=(5, 0, 0), target: bc_diff_particle	-2.836248	39.318706	41.298895	1869.060587	43.232633
	ARIMA, order=(0, 1, 5), target: bc_particle	-0.195759	20.287273	19.104218	582.586512	24.136829
	ARIMA, order=(0, 0, 5), target: bc_diff_particle	-2.769662	38.932398	40.999978	1836.619312	42.855797
	ARIMA, order=(5, 0, 5), target: bc_particle	0.000398	17.431171	14.866814	487.016664	22.068454
	ARIMA, order=(5, 0, 5), target: bc_diff_particle	-2.791674	39.062975	41.013558	1847.343691	42.980736

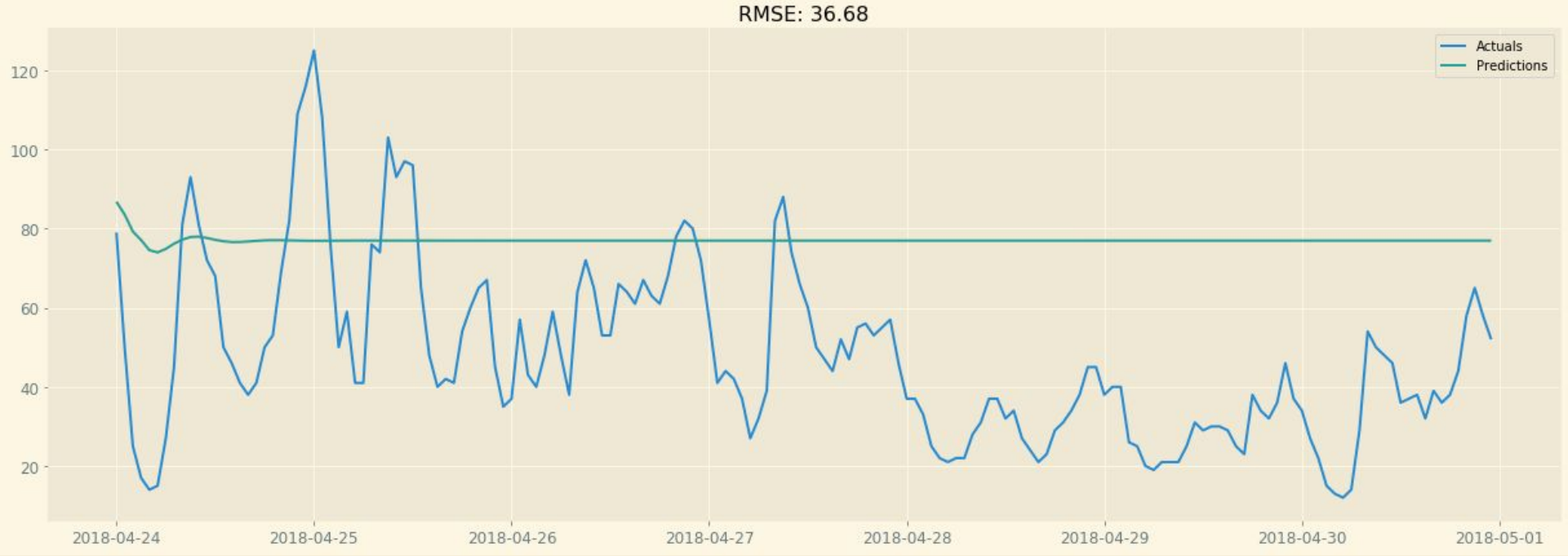
Target: NO2 log | Station: 28079008 - Parque del Retiro/ Escuelas Aguirre | Model: AR

Order = (5, 0, 0)



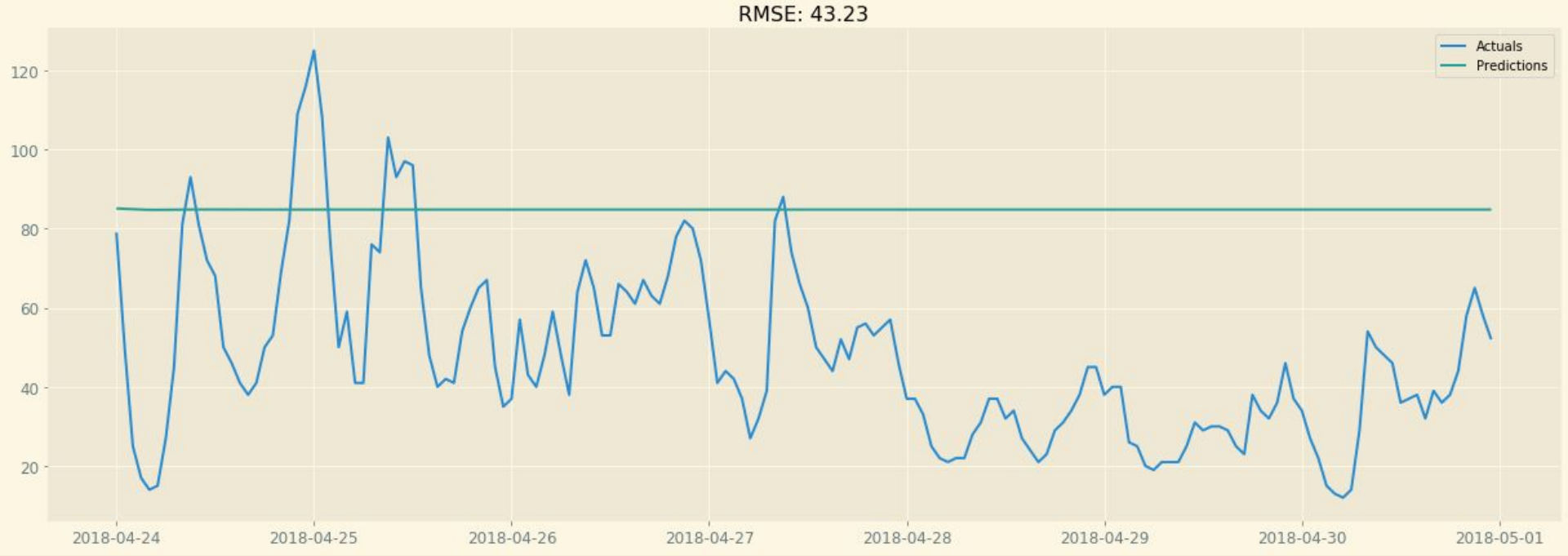
Target: NO2 log | Station: 28079008 - Parque del Retiro/ Escuelas Aguirre | Model: AR

Order = (5, 1, 0)



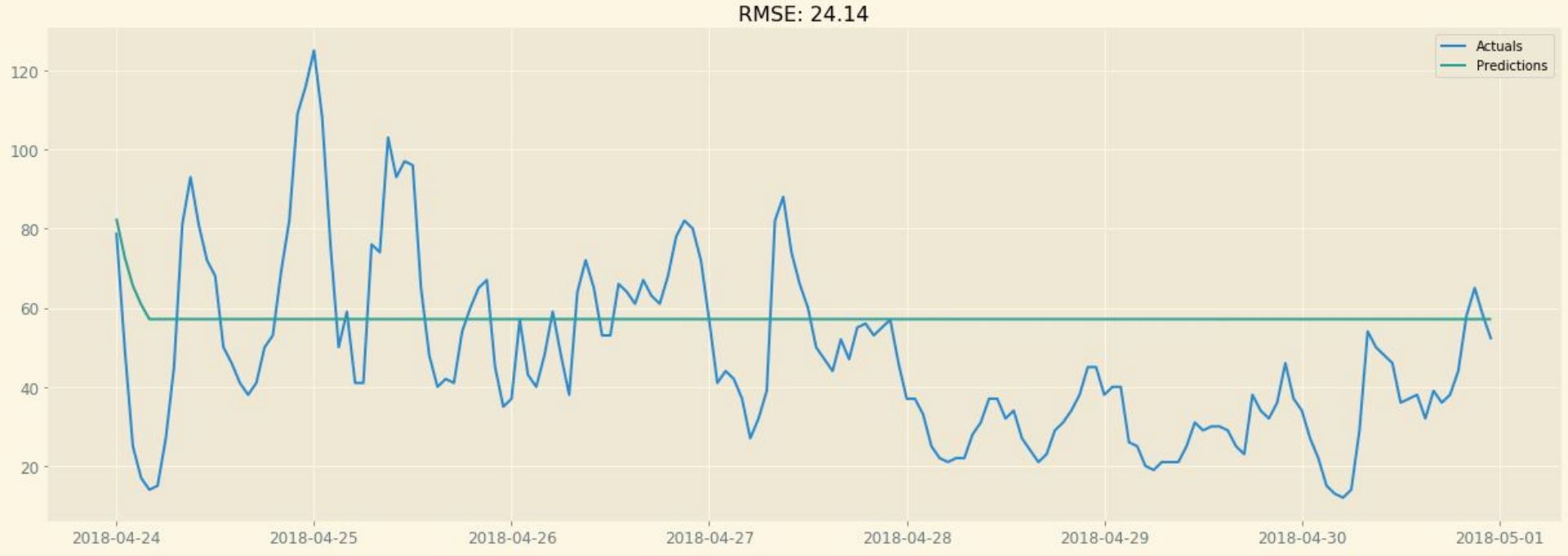
Target: NO2 log diff | Station: 28079008 - Parque del Retiro/ Escuelas Aguirre | Model: AR

Order = (5, 0, 0)



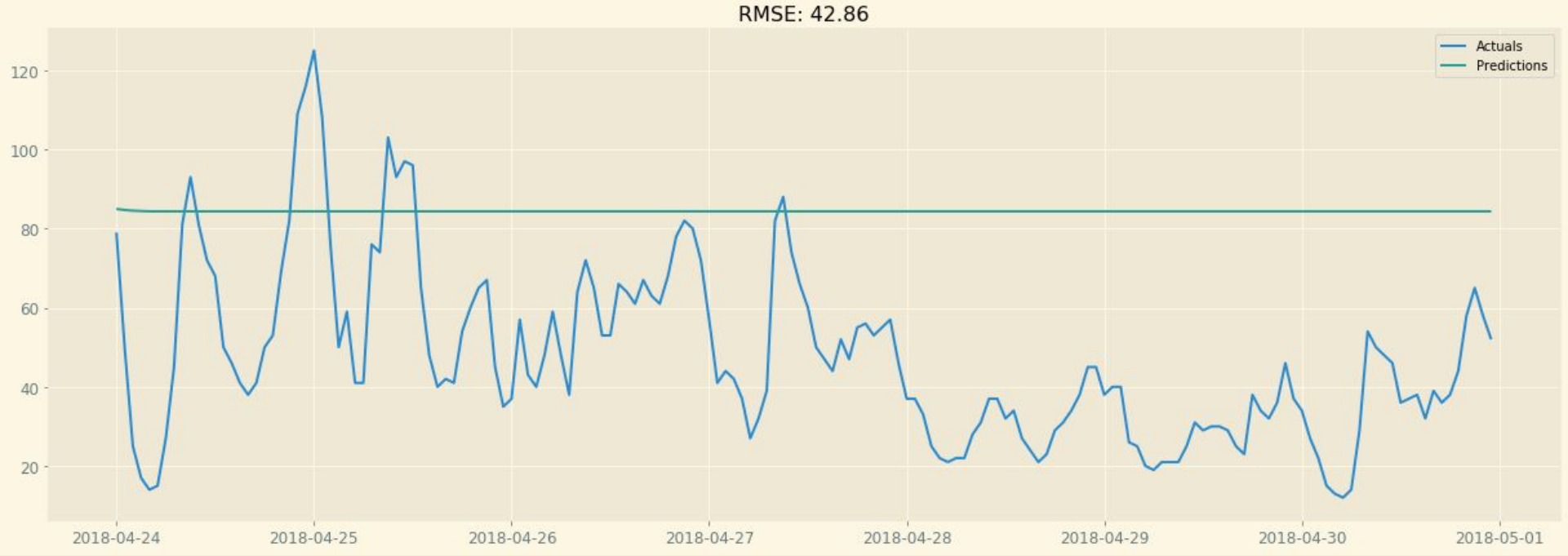
Target: NO2 log | Station: 28079008 - Parque del Retiro/ Escuelas Aguirre | Model: MA

Order = (0, 1, 5)



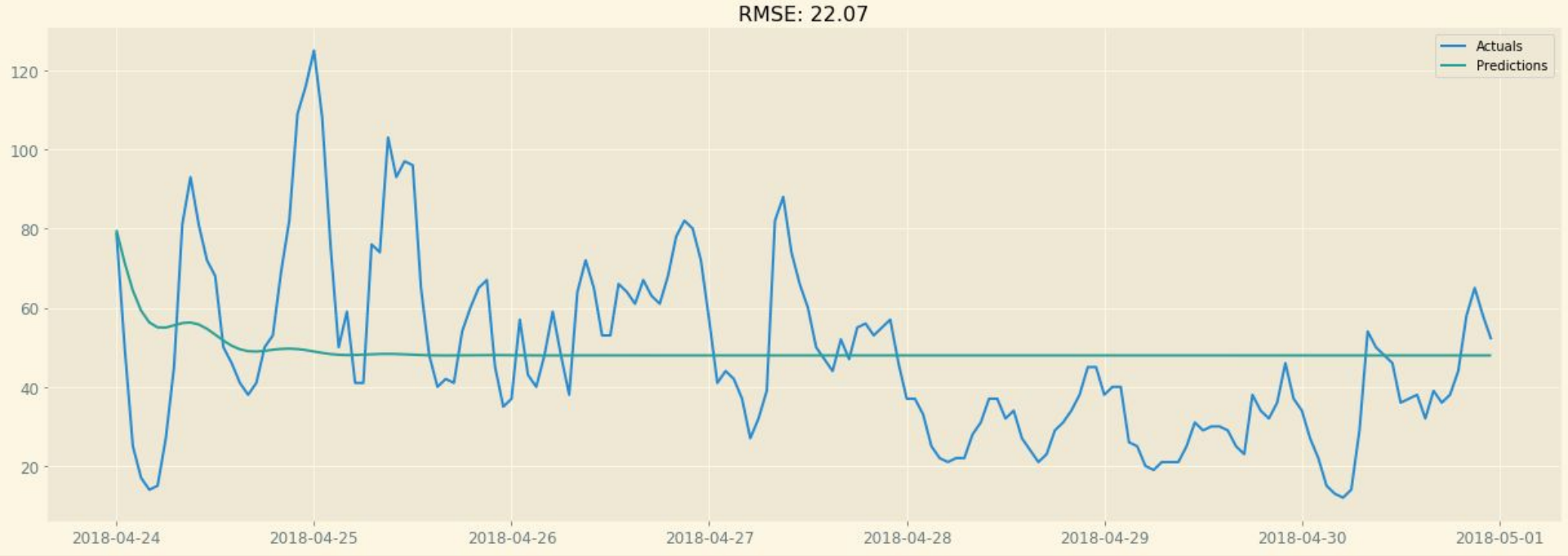
Target: NO2 log diff | Station: 28079008 - Parque del Retiro/ Escuelas Aguirre | Model: MA

Order = (0, 0, 5)



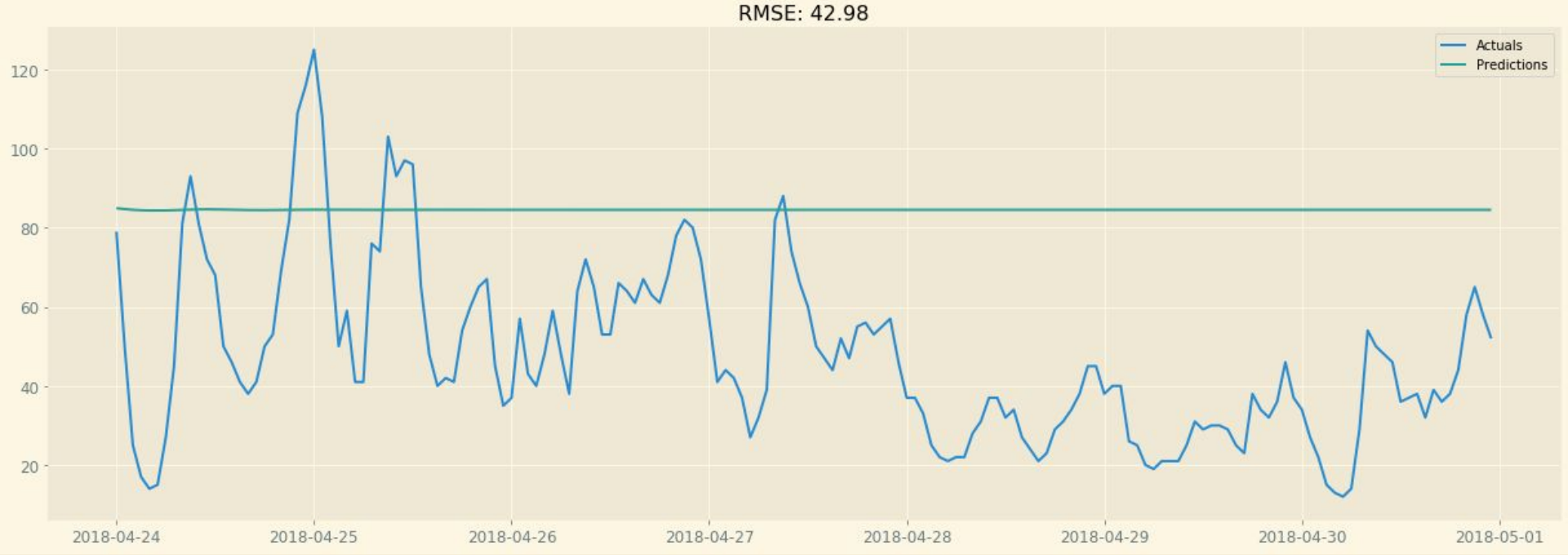
Target: NO2 log | Station: 28079008 - Parque del Retiro/ Escuelas Aguirre | Model: ARMA

Order = (5, 0, 5)



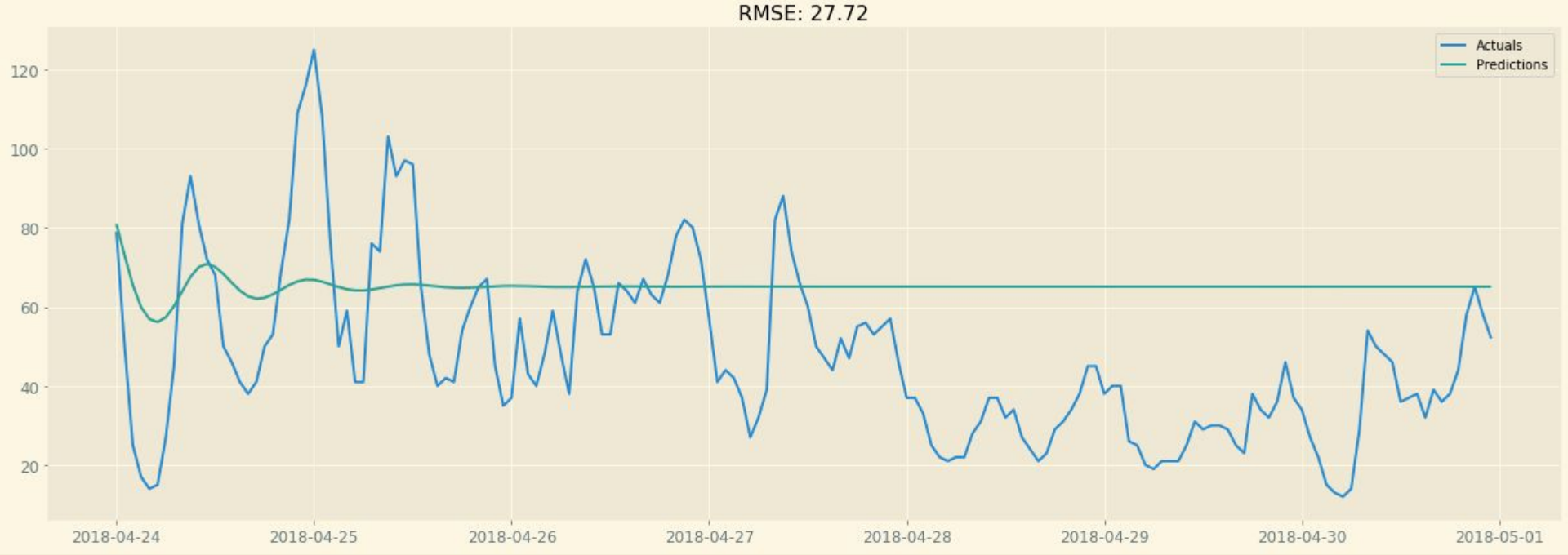
Target: NO2 log diff | Station: 28079008 - Parque del Retiro/ Escuelas Aguirre | Model: ARMA

Order = (5, 0, 5)



Target: NO2 log | Station: 28079008 - Parque del Retiro/ Escuelas Aguirre | Model: ARIMA

Order = (5, 1, 5)

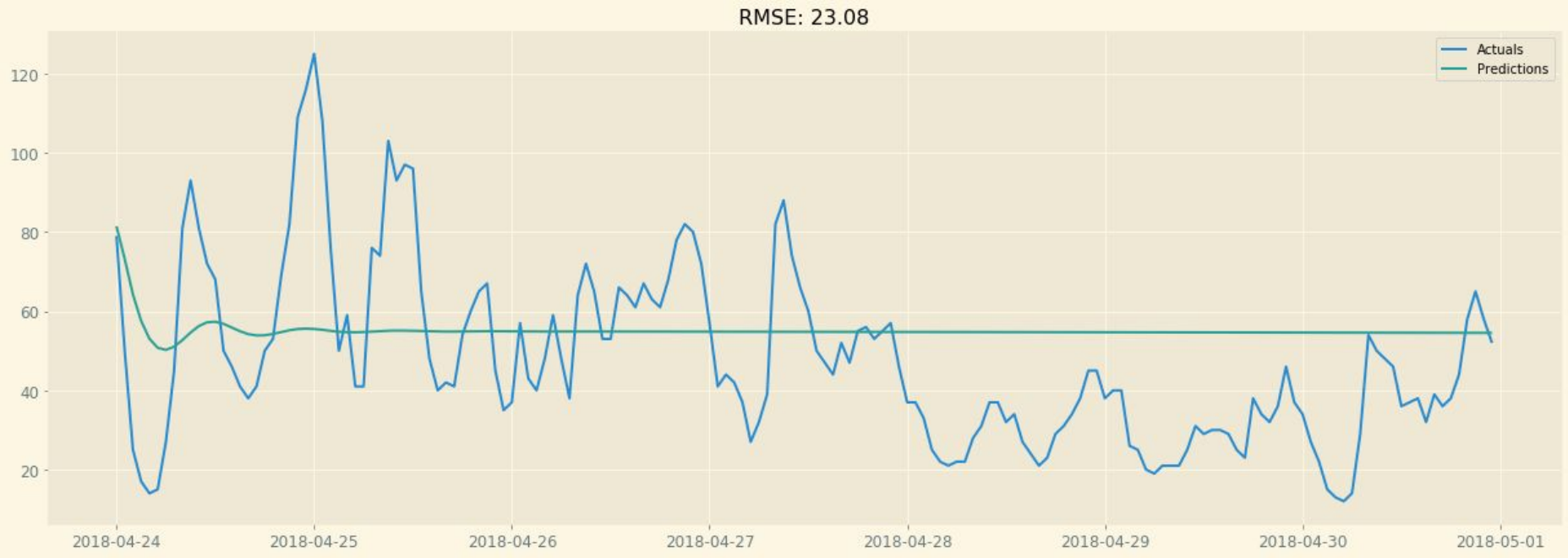


Model Tuning

models	r2_score	mean_abs_error	median_abs_error	mse	rmse
Auto ARIMA: bc_particle	-0.093783	19.169096	17.359020	532.902866	23.084689
Auto ARIMA: bc_diff_particle	-2.769928	38.934043	40.996413	1836.748712	42.857306
Auto SARIMA: bc_particle	-1.923727	31.099324	31.353438	1424.470578	37.742159

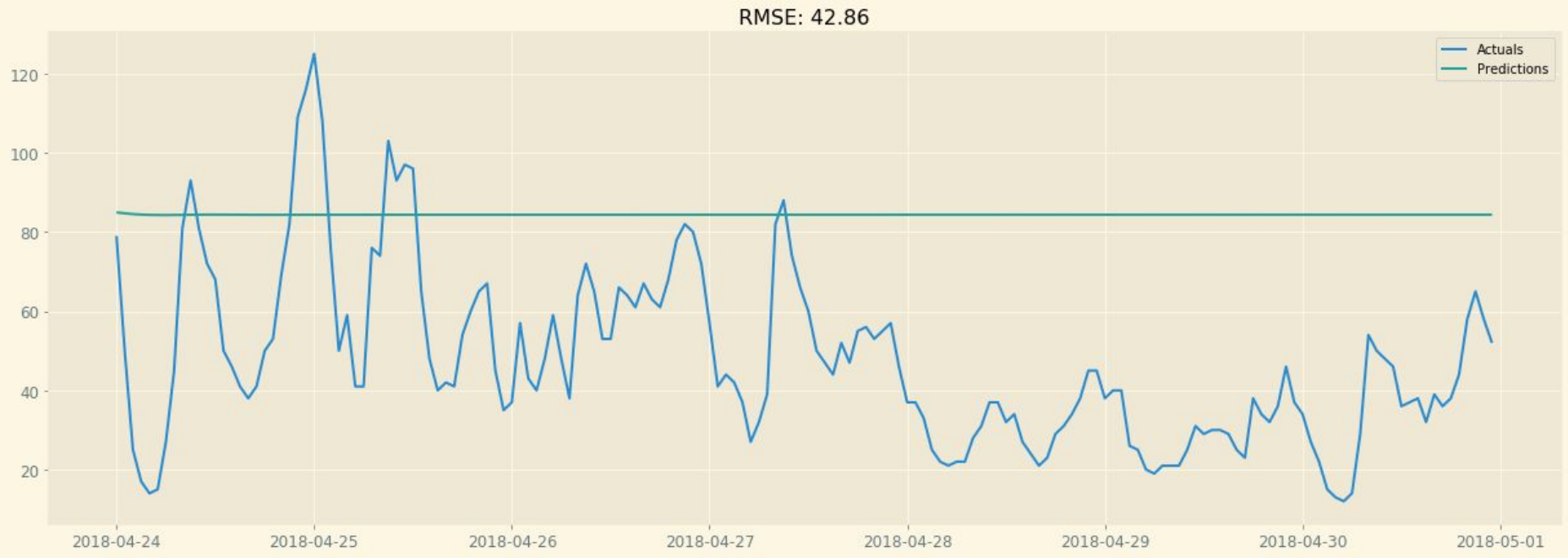
Target: NO2 log | Station: 28079008 - Parque del Retiro/ Escuelas Aguirre | Model: Auto ARIMA

Order = (5, 1, 5)



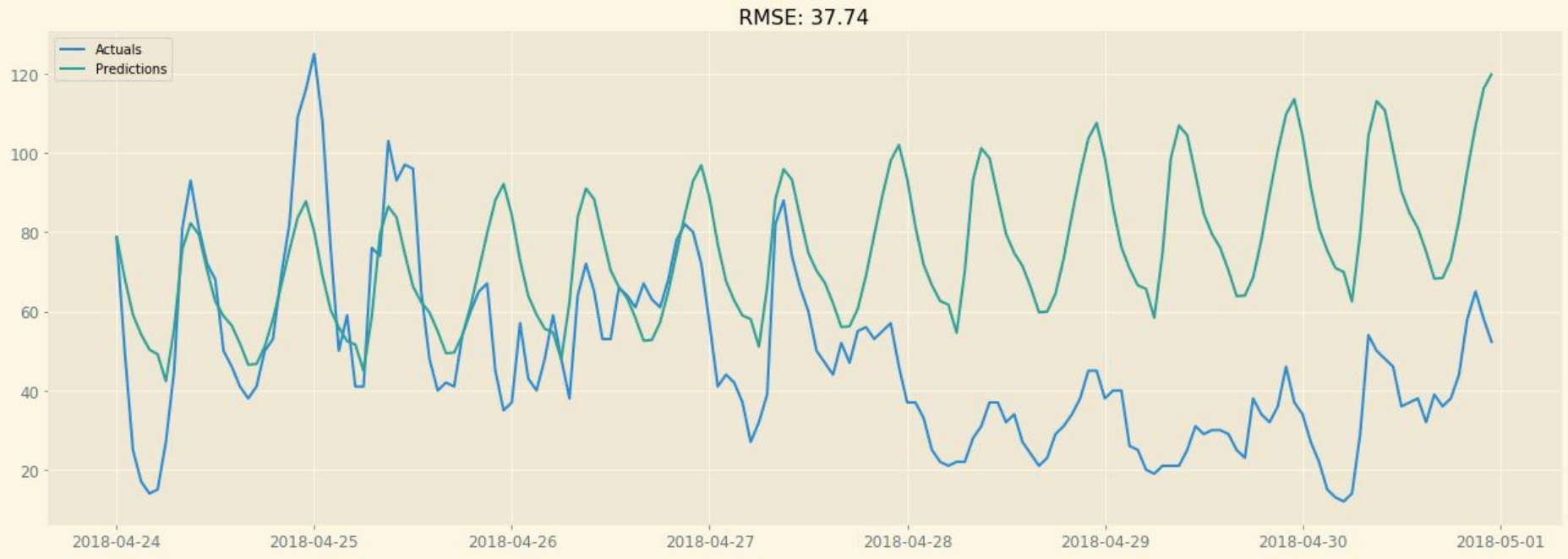
Target: NO2 log diff | Station: 28079008 - Parque del Retiro/ Escuelas Aguirre | Model: Auto ARIMA

Order = (5, 0, 5)



Target: NO2 log | Station: 28079008 - Parque del Retiro/ Escuelas Aguirre | Model: Auto SARIMA

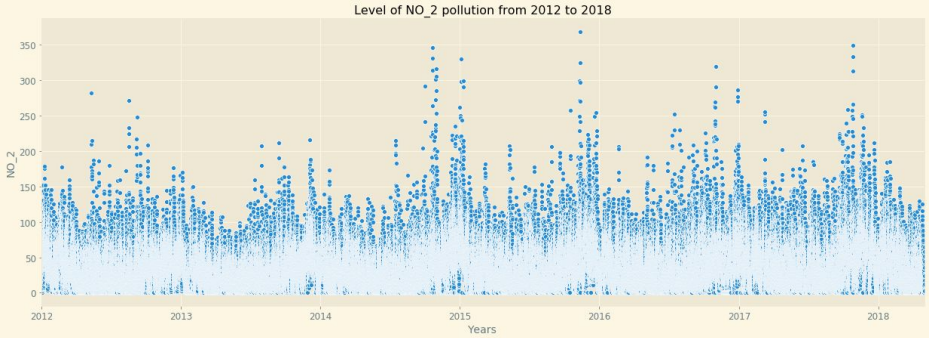
Order = (2, 1, 2) | Seasonal order = (2, 0, 2, 12)

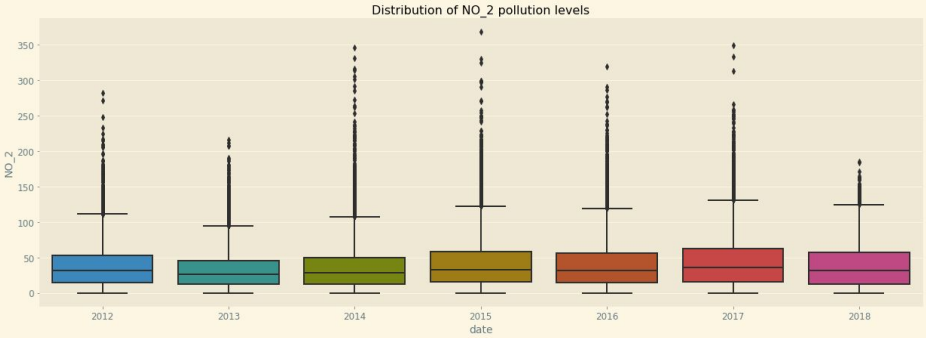


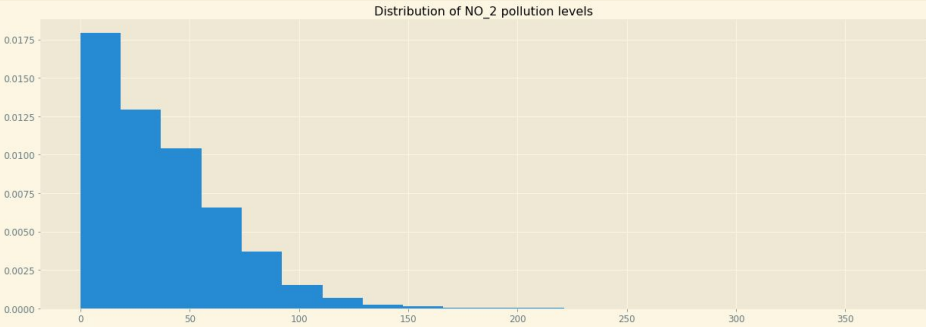
Rolling predictions with ARIMA

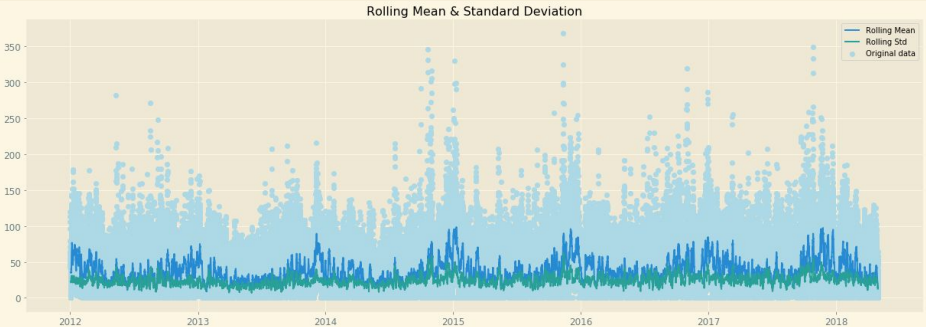
NO2

Visualising the target variable









Testing Stationarity

Results of Dickey-Fuller Test:

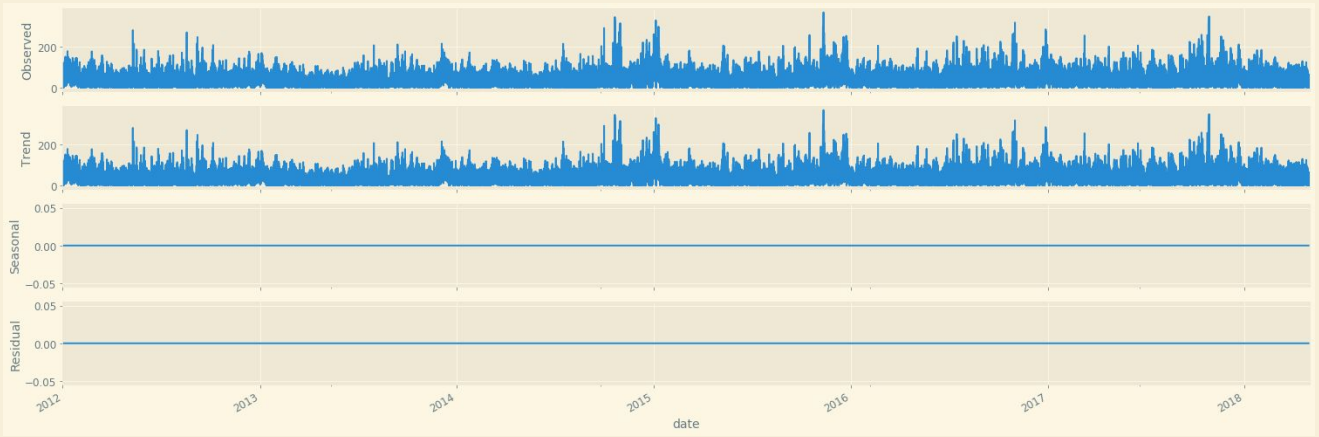
Null Hypothesis: Unit Root Present
Test Statistic < Critical Value => Reject Null
P-Value =< Alpha(.05) => Reject Null

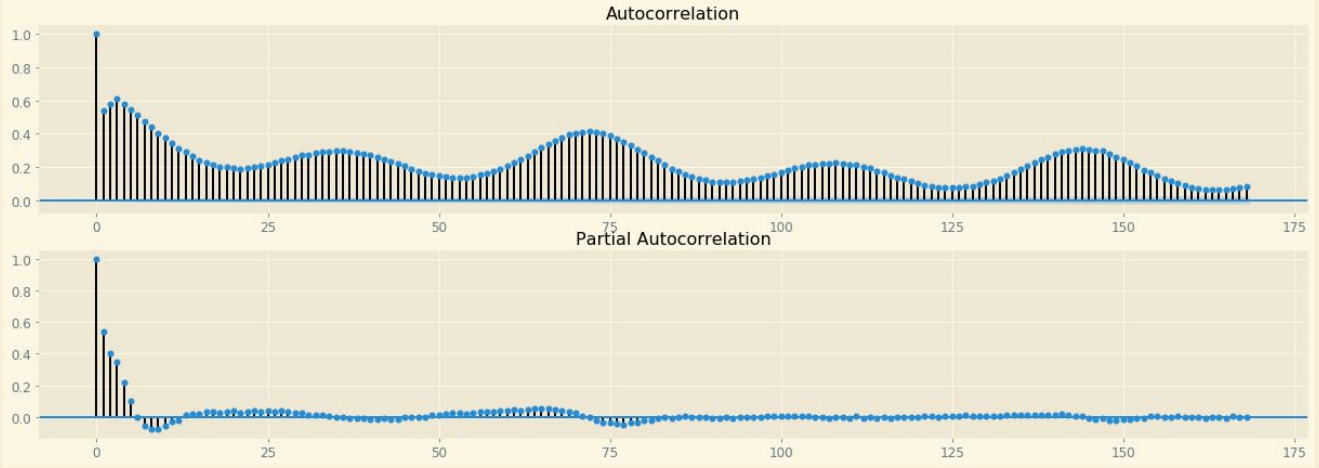
Test Statistic	-19.065
p-value	0.000
#Lags Used	77.000
Nr of Obs Used	166386
Critical Value 1%	3.430
Critical Value 5%	-2.862
Critical Value 10%	-2.567

Results of KPSS Test:

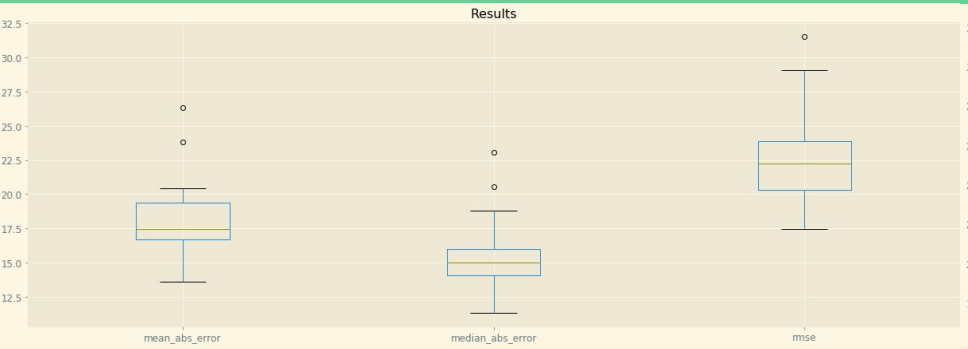
Null Hypothesis: Data is Stationary
Test Statistic > Critical Value => Reject Null
P-Value =< Alpha(.05) => Reject Null

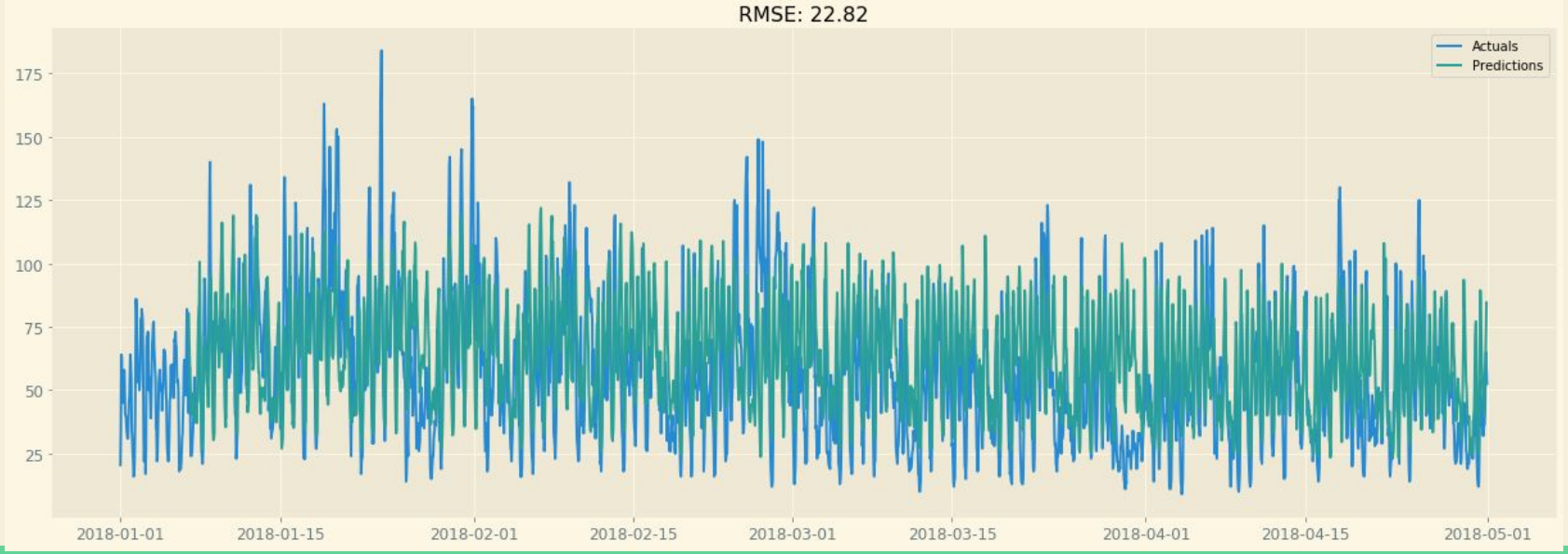
Test Statistic	8.344
p-value	0.010
Lags Used	77.000
Critical Value 10%	0.347
Critical Value 5%	0.463
Critical Value 2.5%	0.574
Critical Value 1%	0.739



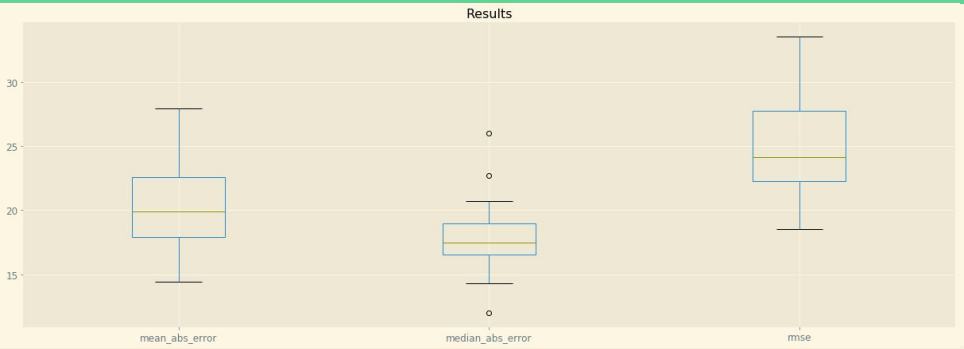


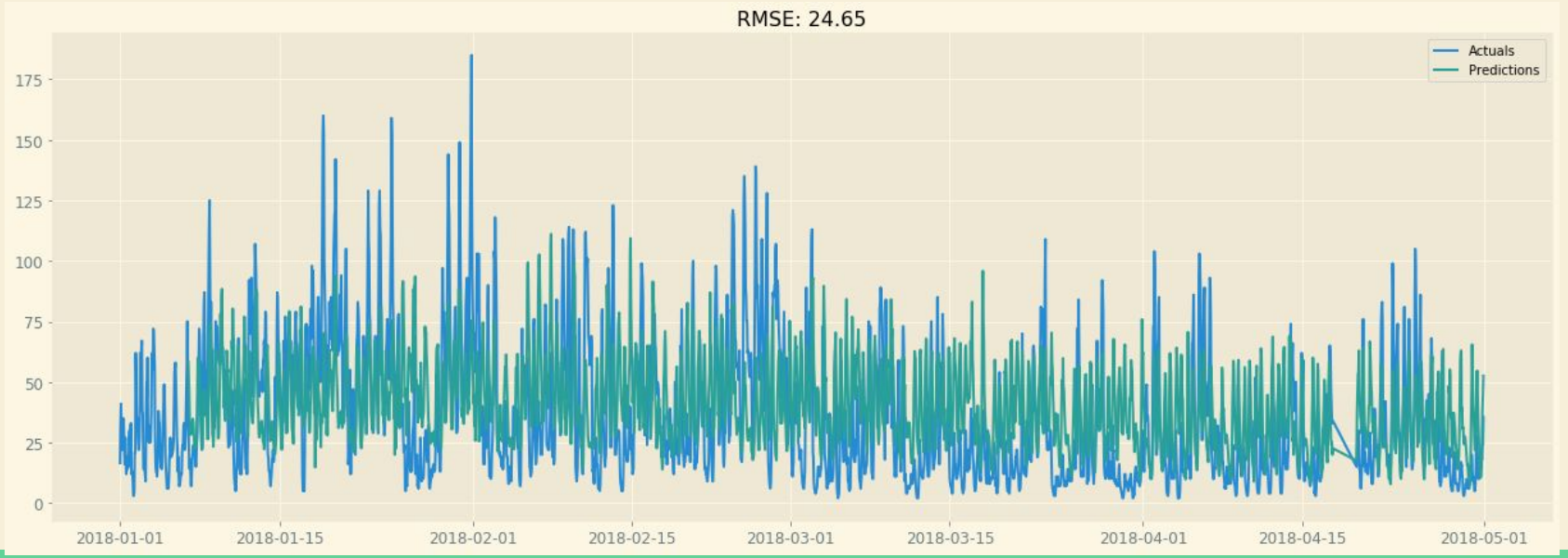
Station nr 28079008 - Parque del Retiro/ Escuelas Aguirre



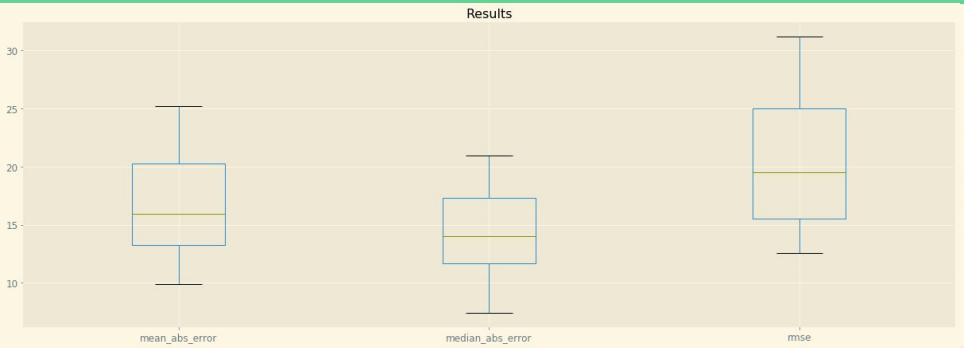


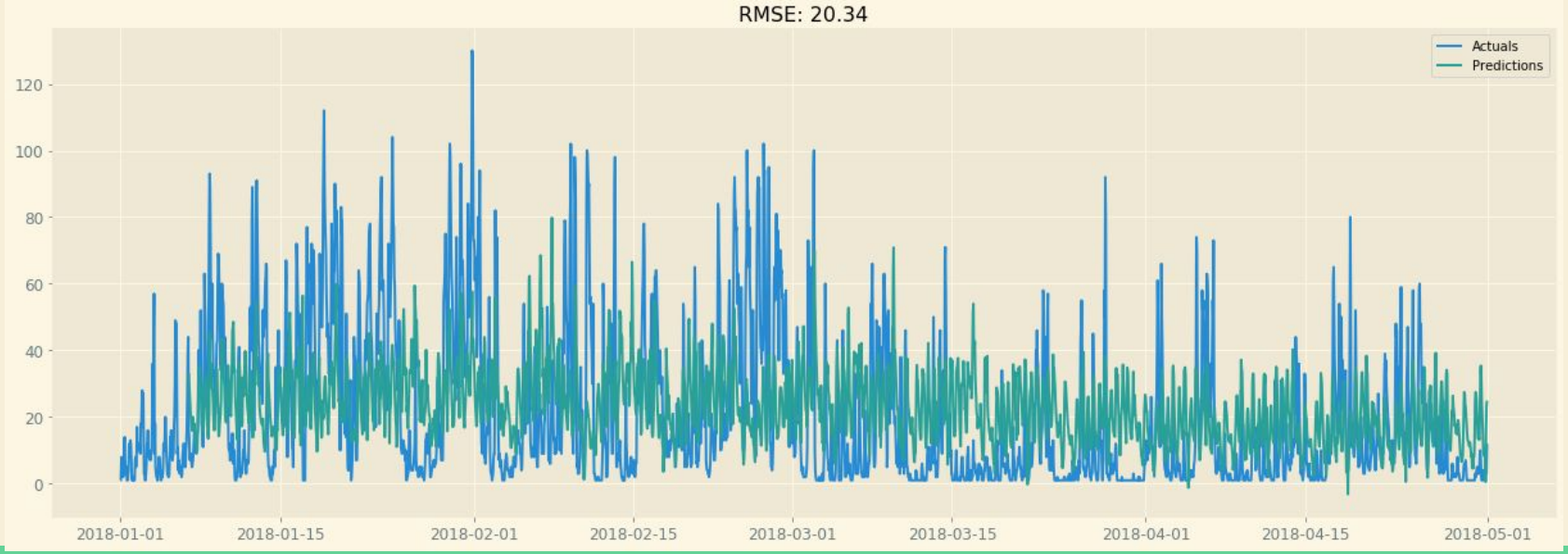
Station nr 28079018 - San Isidro/Calle Farolillo





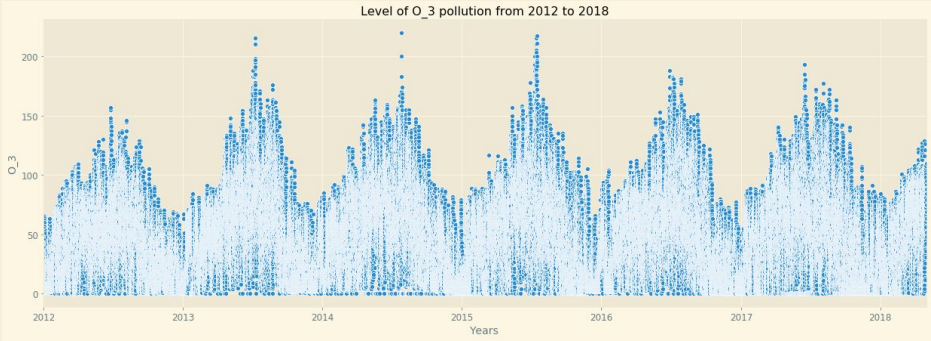
Station nr 28079024 - Casa de Campo

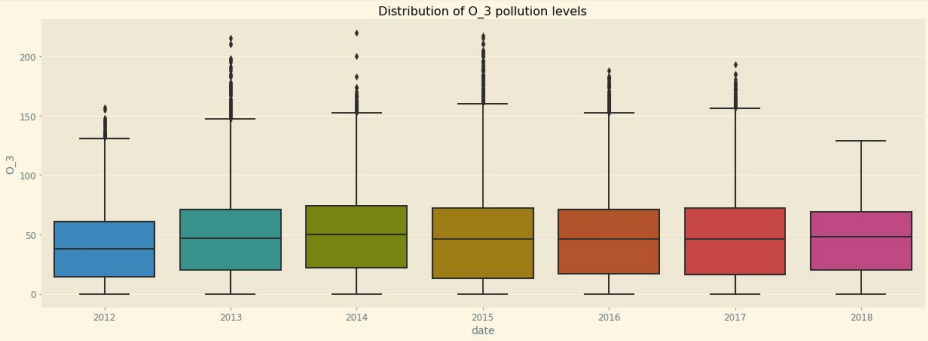


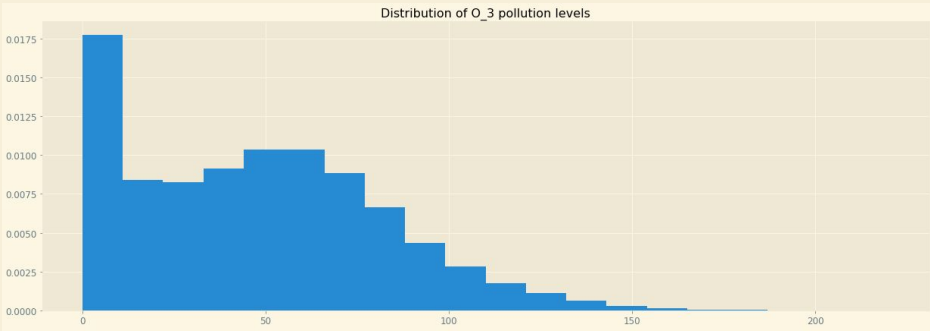


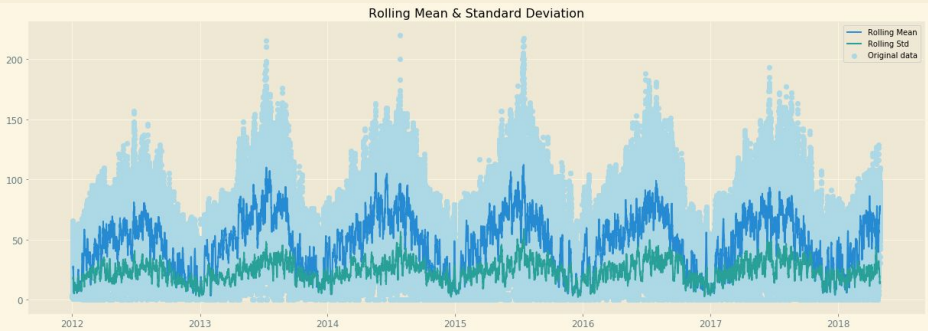
03

Visualising the target variable









Testing Stationarity

Results of Dickey-Fuller Test:

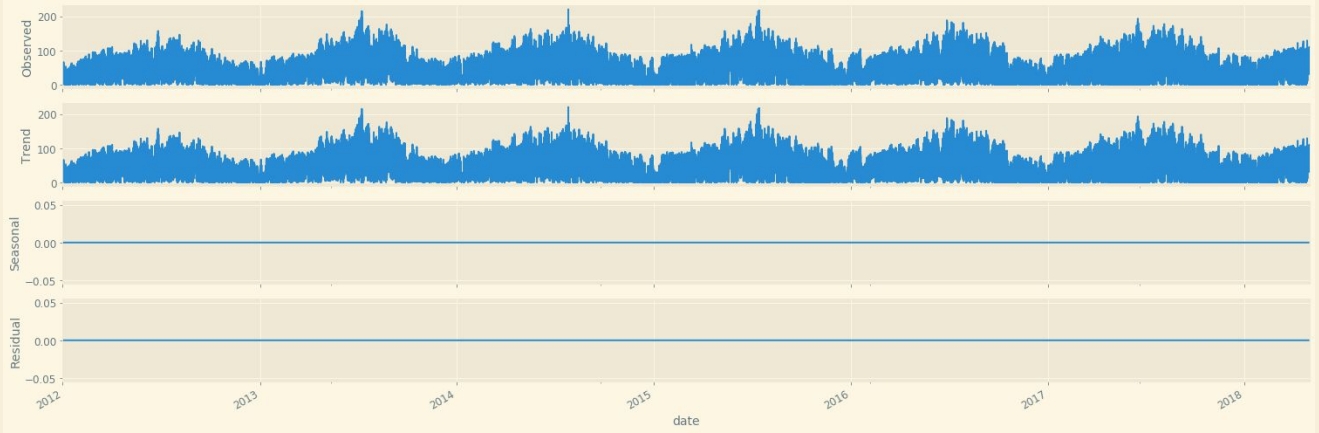
Null Hypothesis: Unit Root Present
Test Statistic < Critical Value => Reject Null
P-Value =< Alpha(.05) => Reject Null

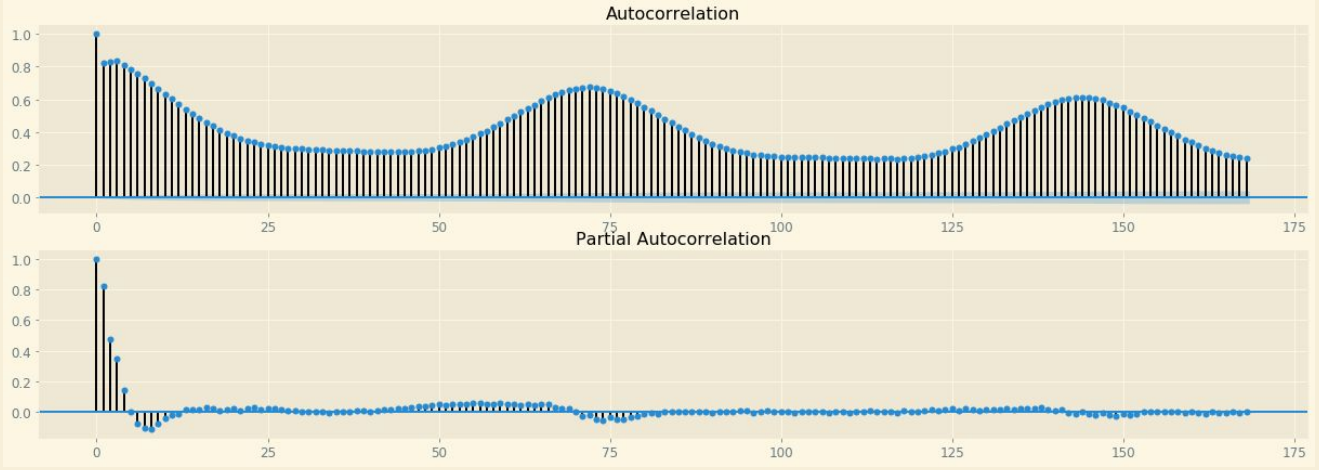
Test Statistic	-1.483e+01
p-value	1.891e-27
#Lags Used	7.700e+01
Nr of Obs Used	1.664e+05
Critical Value 1%	-3.430
Critical Value 5%	-2.862
Critical Value 10%	-2.567

Results of KPSS Test:

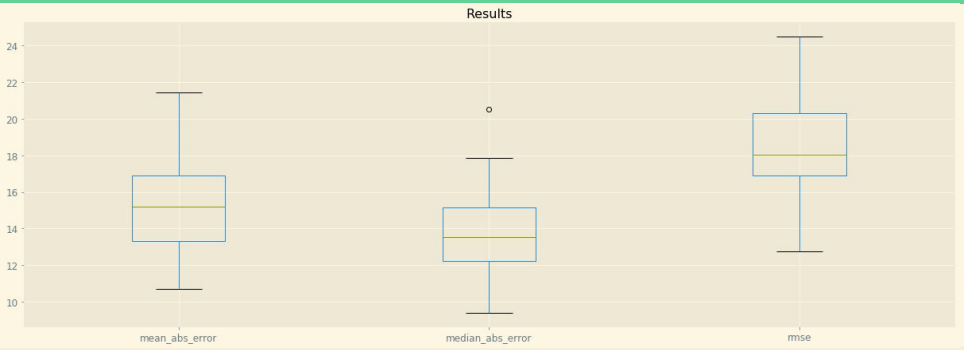
Null Hypothesis: Data is Stationary
Test Statistic > Critical Value => Reject Null
P-Value =< Alpha(.05) => Reject Null

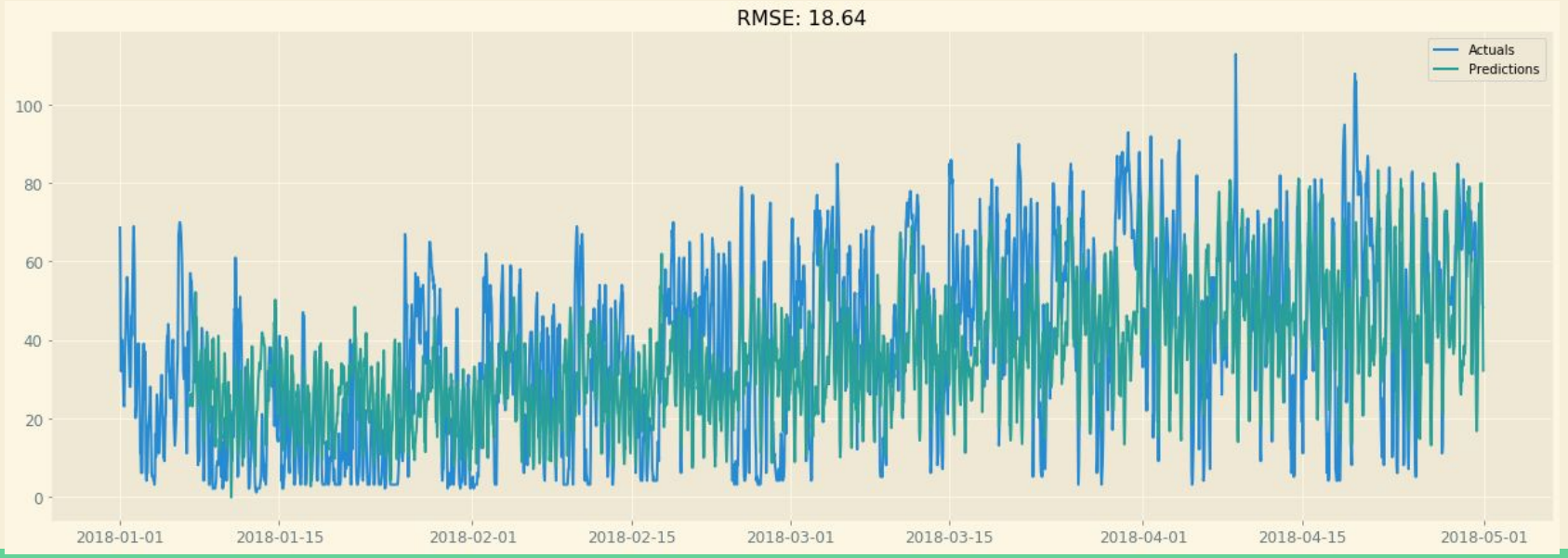
Test Statistic	1.562
p-value	0.010
Lags Used	77.000
Critical Value 10%	0.347
Critical Value 5%	0.463
Critical Value 2.5%	0.574
Critical Value 1%	0.739



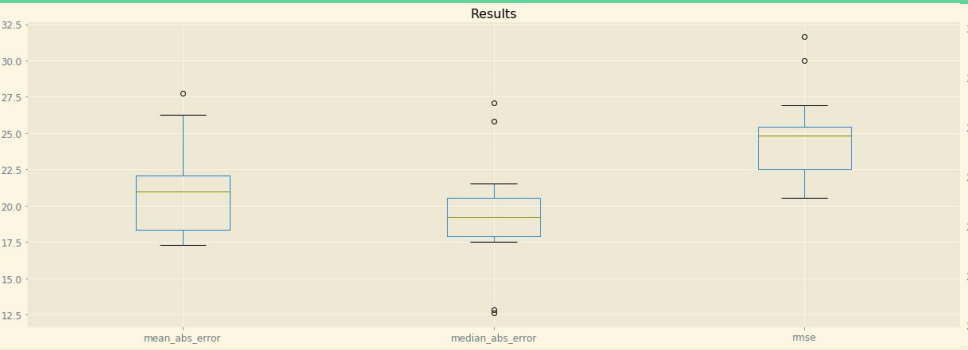


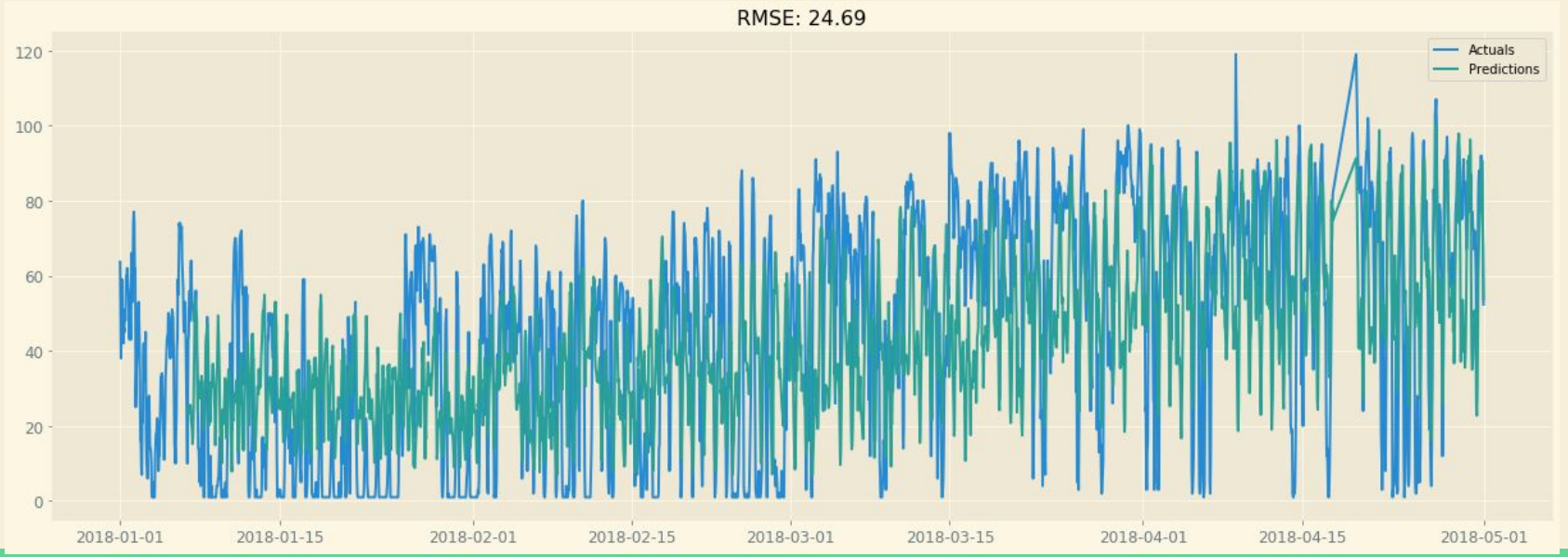
Station nr 28079008 - Parque del Retiro/ Escuelas Aguirre



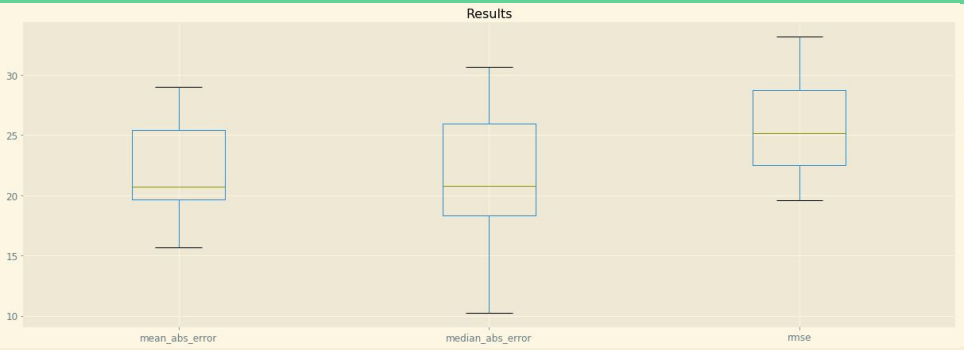


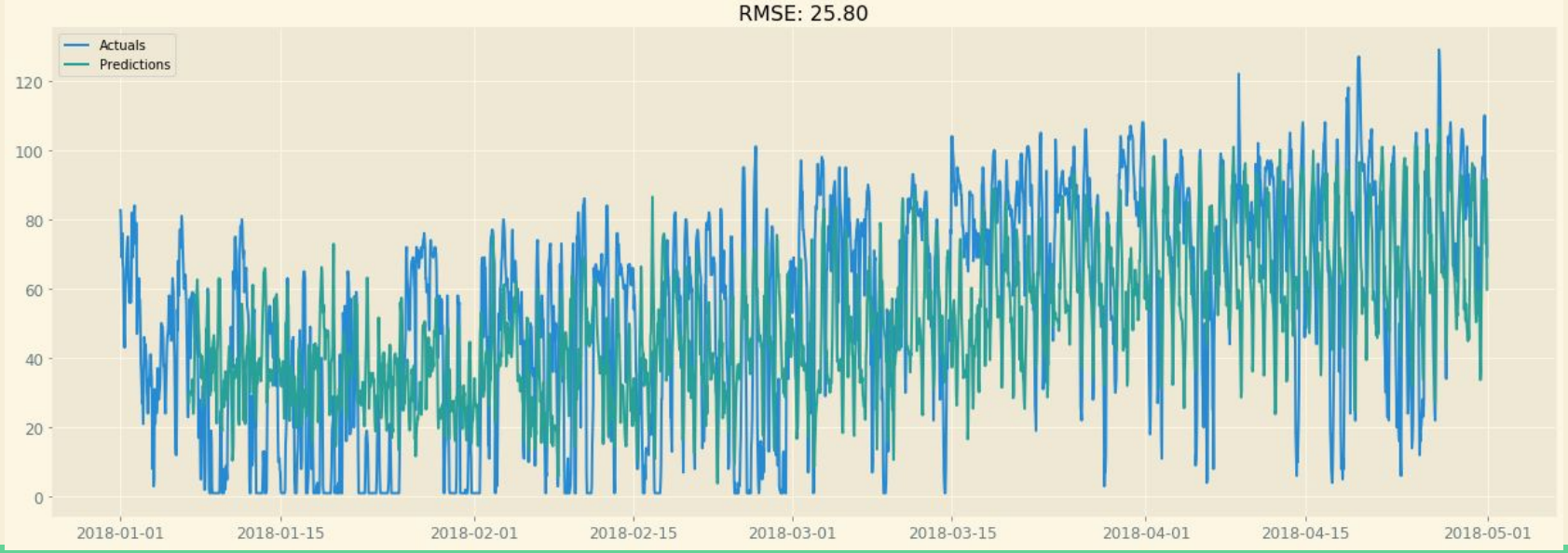
Station nr 28079018 - San Isidro/Calle Farolillo





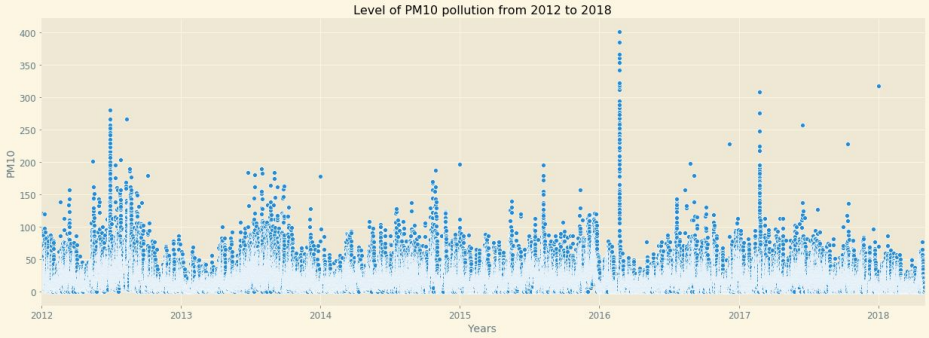
Station nr 28079018 - San Isidro/Calle Farolillo

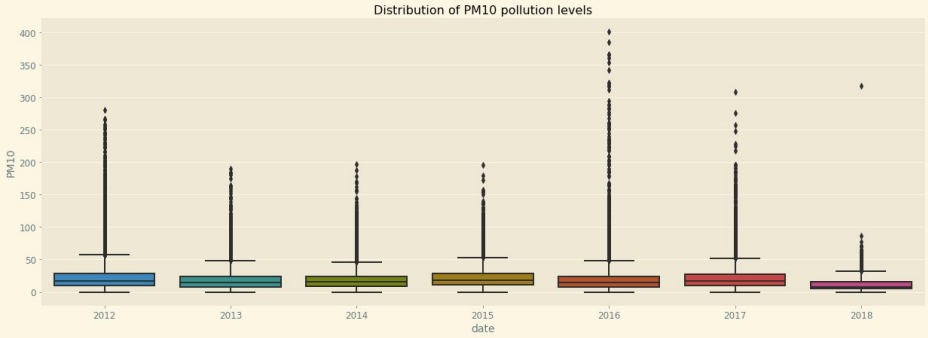


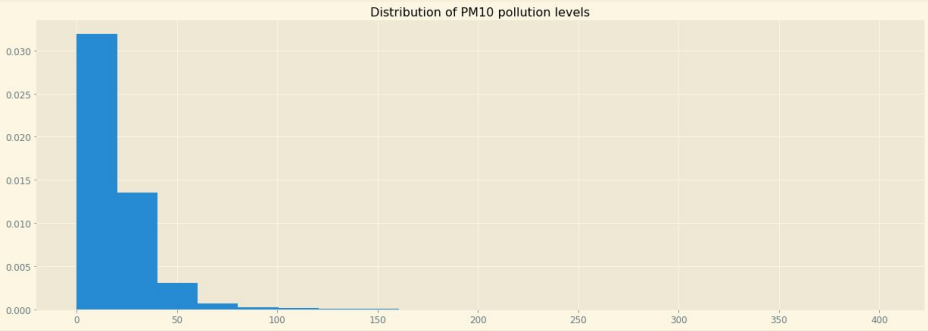


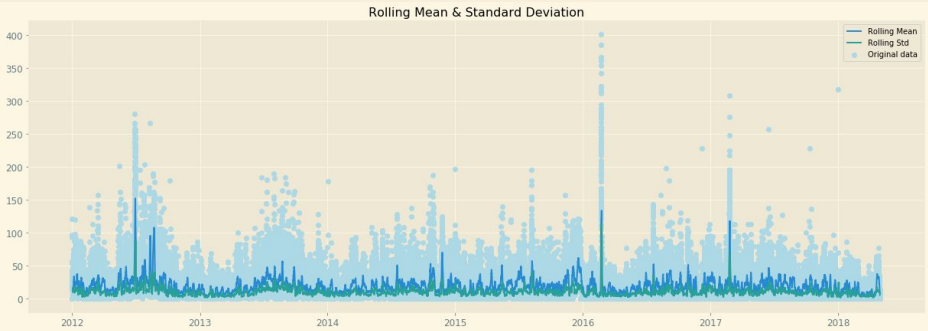
PM10

Visualising the target variable









Testing Stationarity

Results of Dickey-Fuller Test:

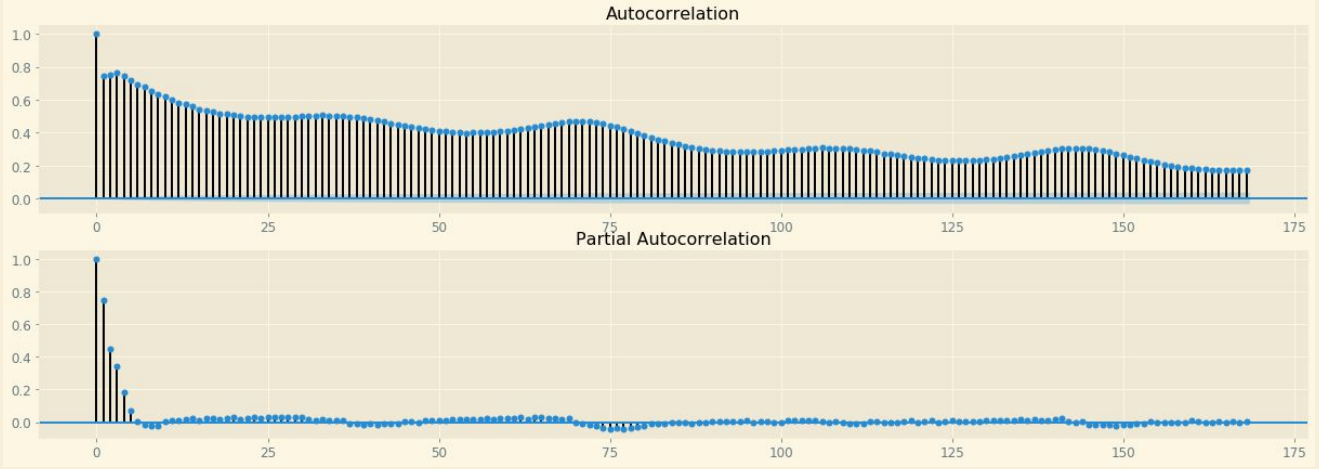
Null Hypothesis: Unit Root Present
Test Statistic < Critical Value => Reject Null
P-Value =< Alpha(.05) => Reject Null

Test Statistic	-21.758
p-value	0.000
#Lags Used	77.000
Nr of Obs Used	166386
Critical Value 1%	-3.430
Critical Value 5%	-2.862
Critical Value 10%	-2.567

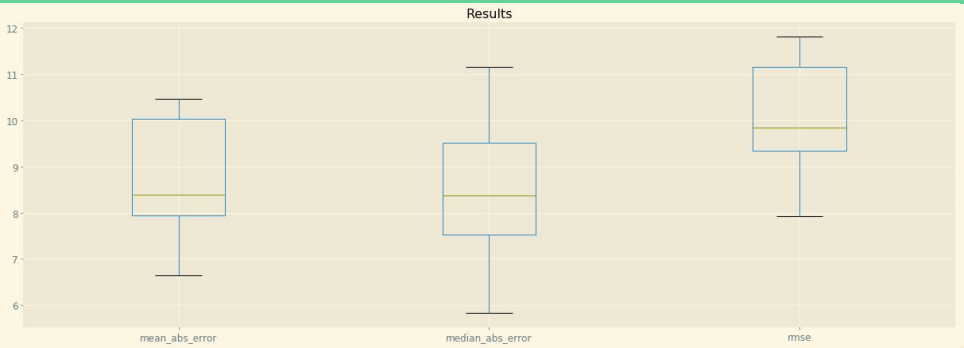
Results of KPSS Test:

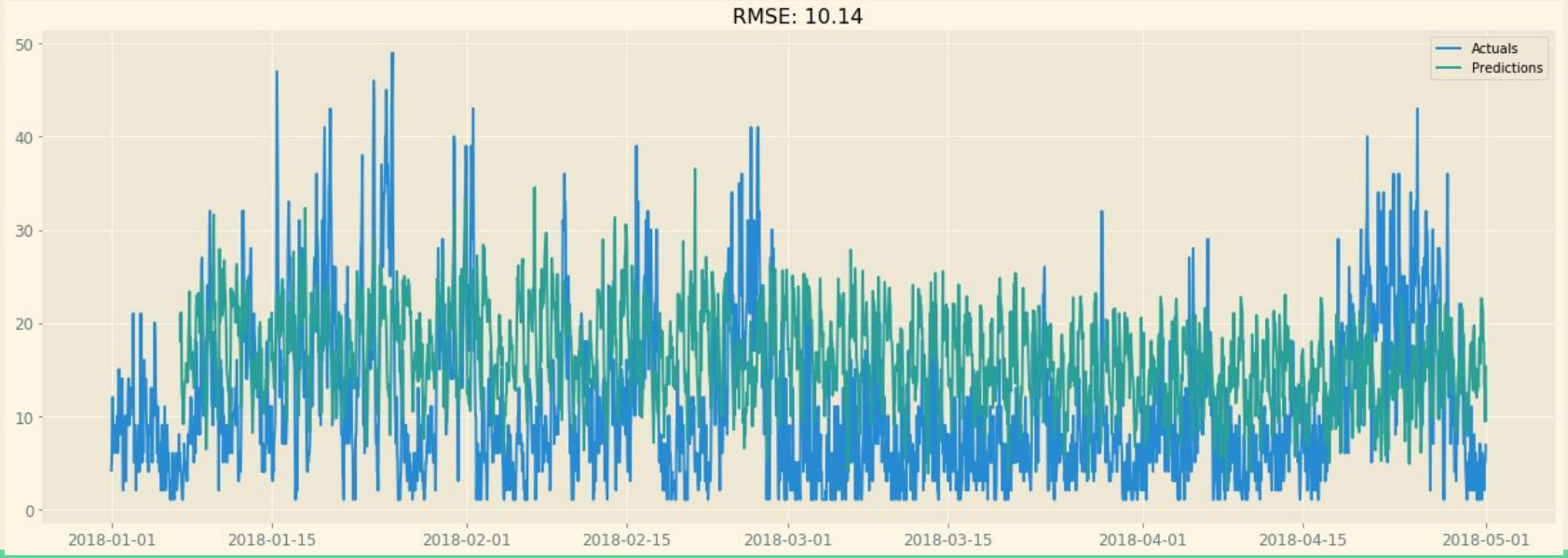
Null Hypothesis: Data is Stationary
Test Statistic > Critical Value => Reject Null
P-Value =< Alpha(.05) => Reject Null

Test Statistic	1.330
p-value	0.010
Lags Used	77.000
Critical Value 10%	0.347
Critical Value 5%	0.463
Critical Value 2.5%	0.574
Critical Value 1%	0.739

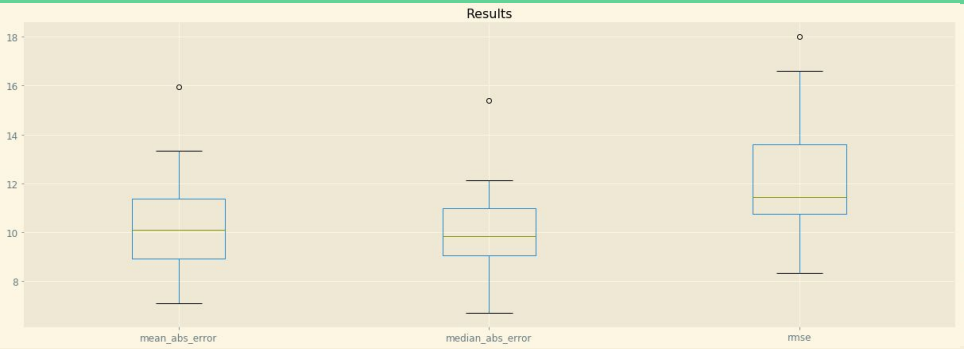


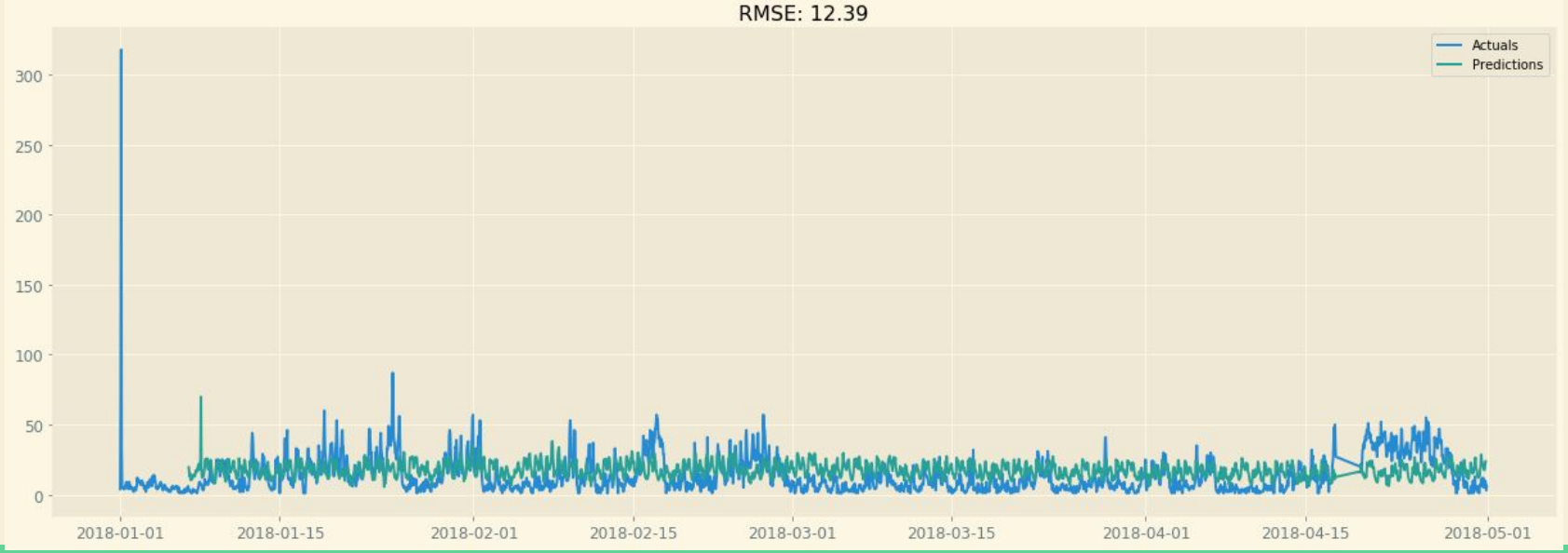
Station nr 28079008 - Parque del Retiro/ Escuelas Aguirre



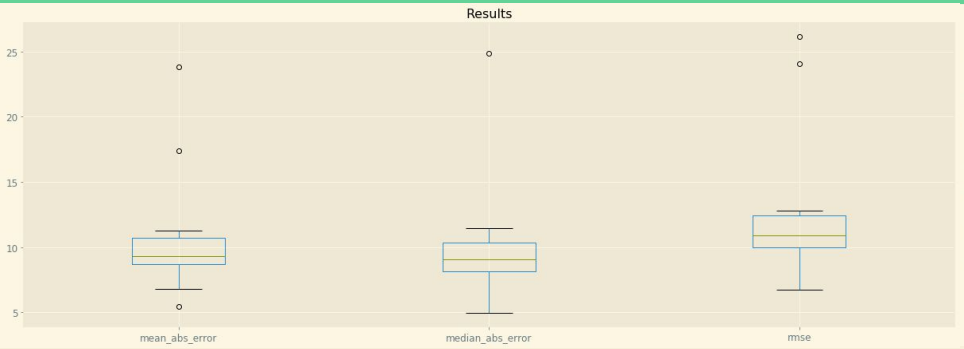


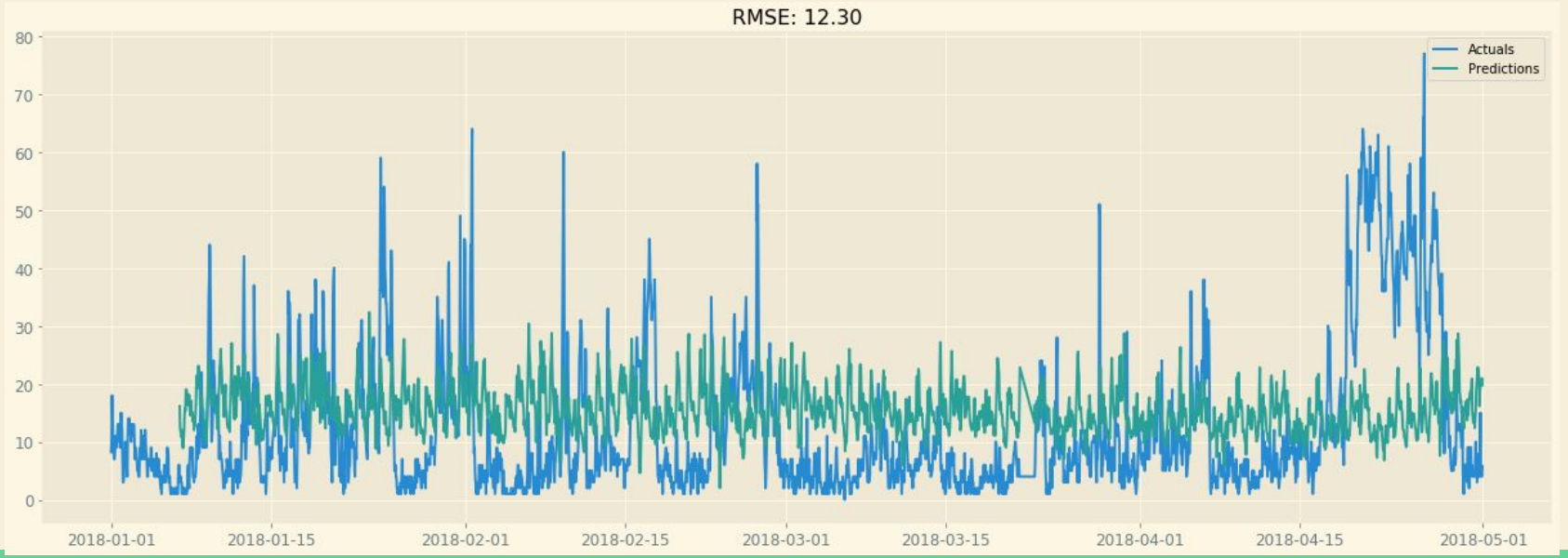
Station nr 28079018 - San Isidro/Calle Farolillo





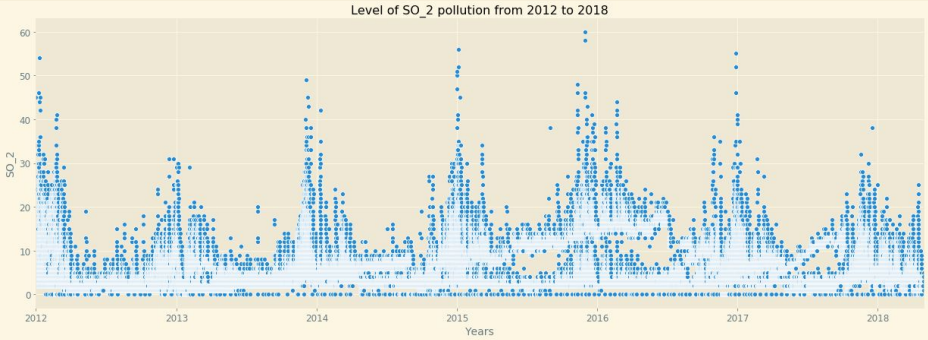
Station nr 28079018 - San Isidro/Calle Farolillo

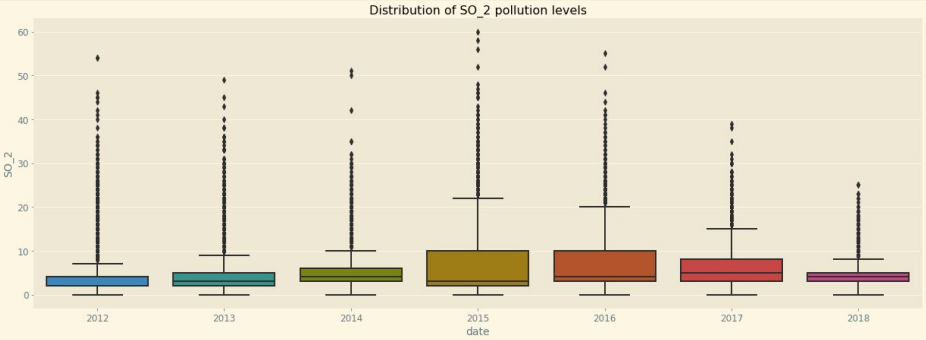


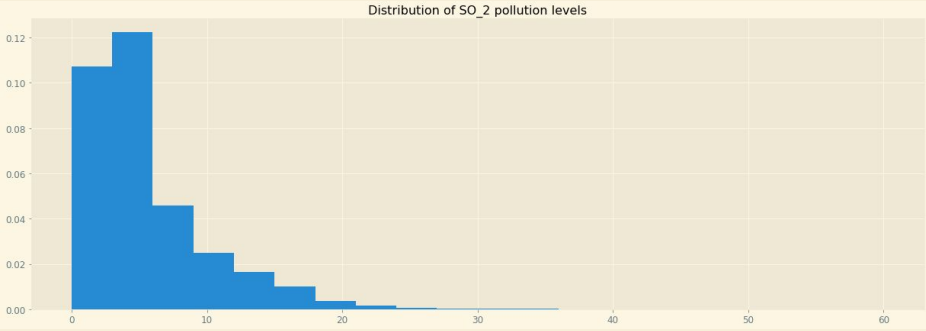


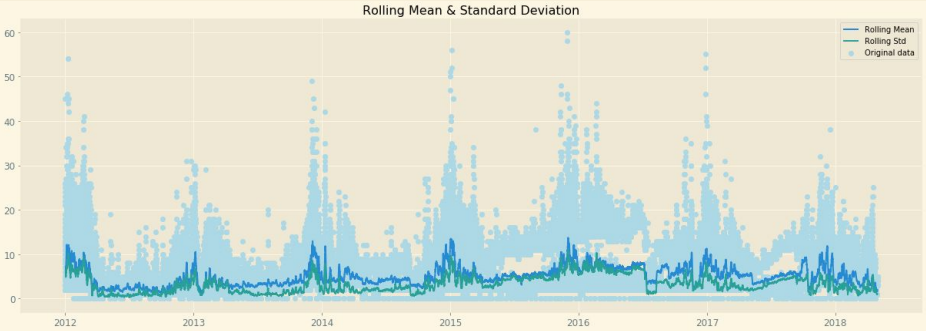
SO2

Visualising the target variable









Testing Stationarity

Results of Dickey-Fuller Test:

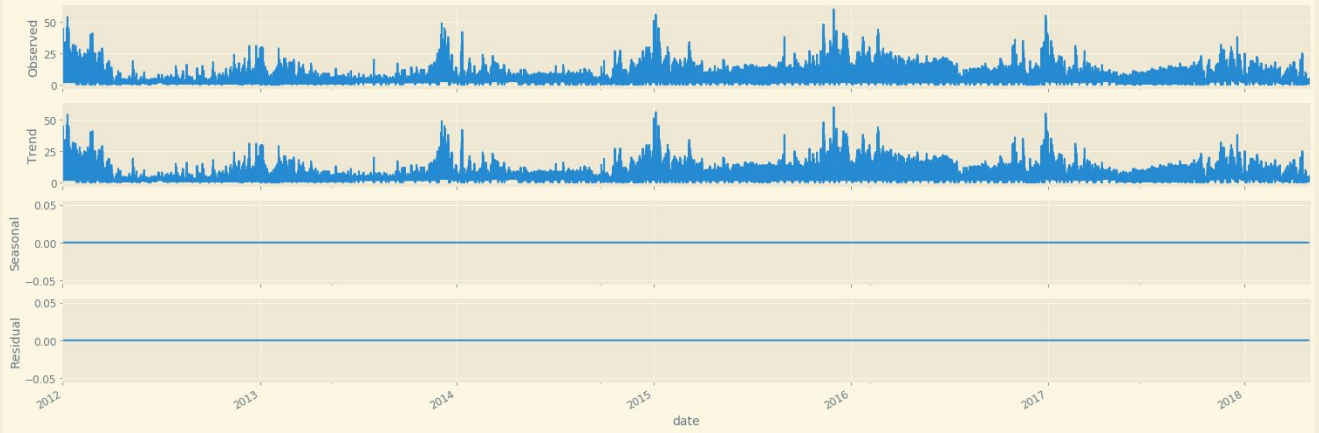
Null Hypothesis: Unit Root Present
Test Statistic < Critical Value => Reject Null
P-Value =< Alpha(.05) => Reject Null

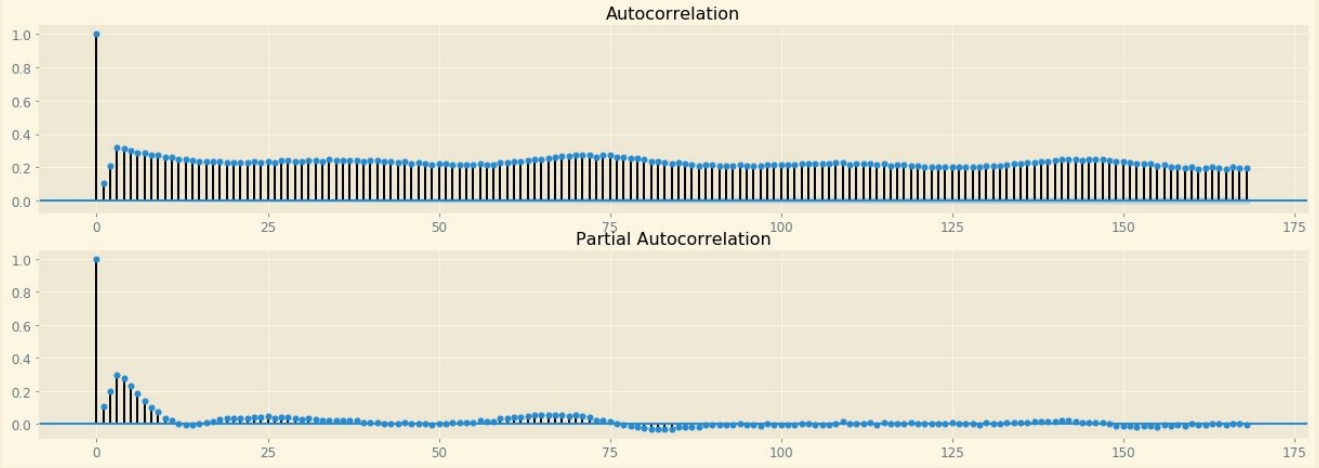
Test Statistic	-8.382
p-value	2.490e-13
#Lags Used	7.700e+01
Nr of Obs Used	1.664e+05
Critical Value 1%	-3.430
Critical Value 5%	-2.862
Critical Value 10%	-2.567

Results of KPSS Test:

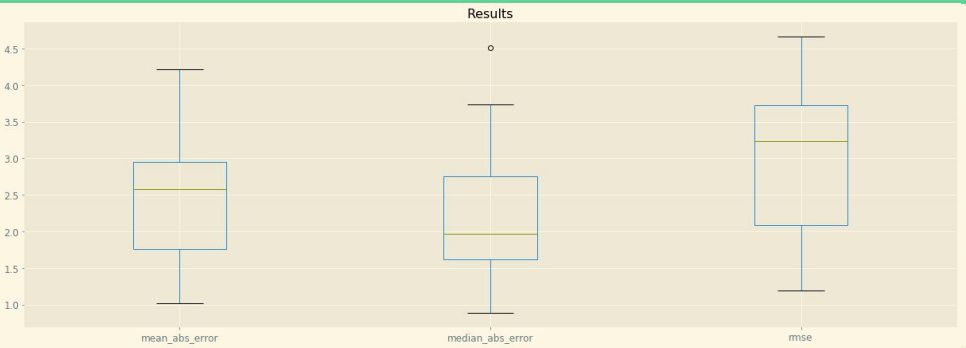
Null Hypothesis: Data is Stationary
Test Statistic > Critical Value => Reject Null
P-Value =< Alpha(.05) => Reject Null

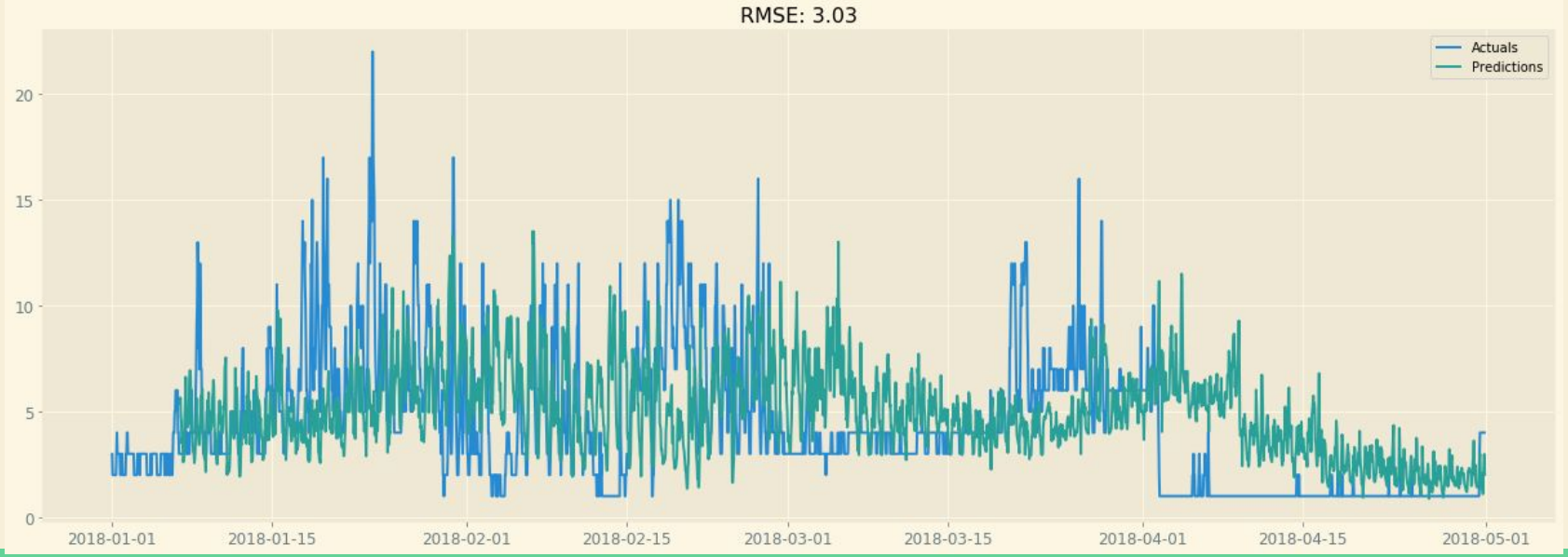
Test Statistic	32.792
p-value	0.010
Lags Used	77.000
Critical Value 10%	0.347
Critical Value 5%	0.463
Critical Value 2.5%	0.574
Critical Value 1%	0.739



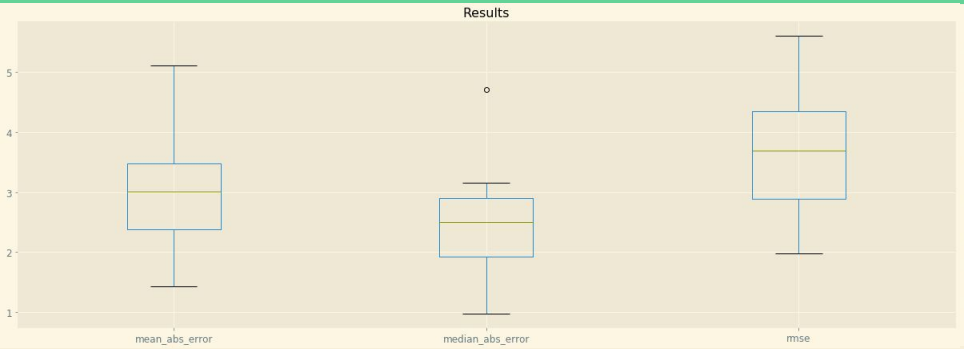


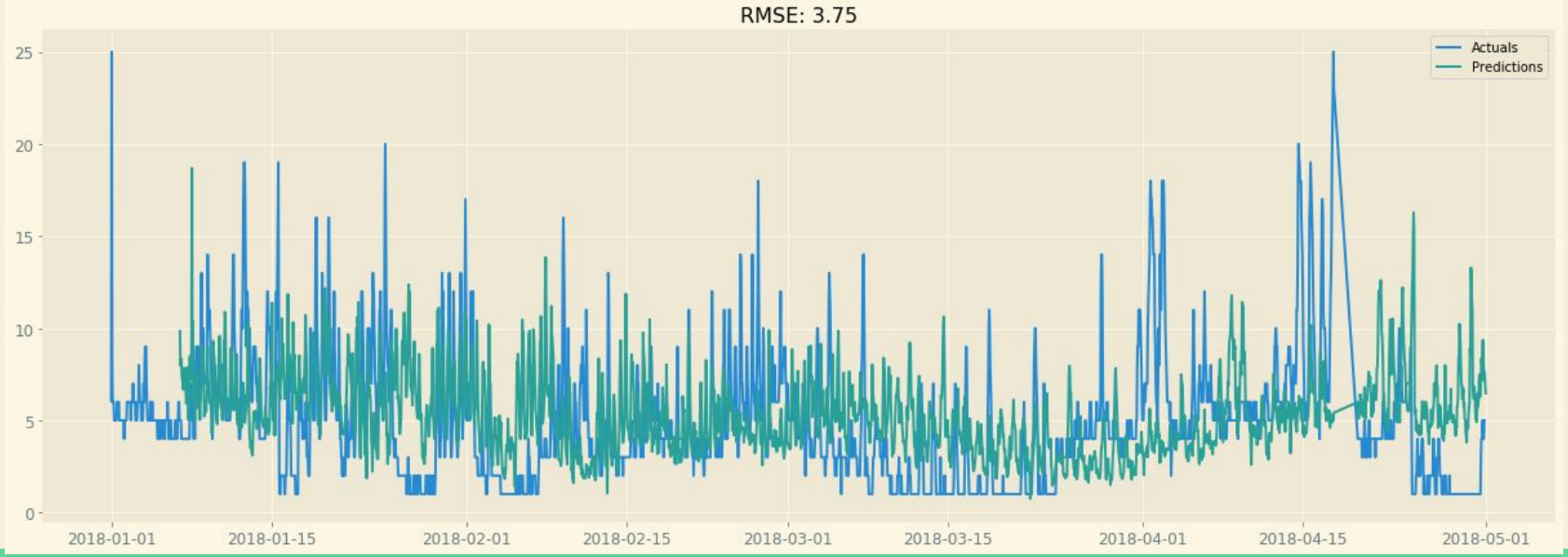
Station nr 28079008 - Parque del Retiro/ Escuelas Aguirre



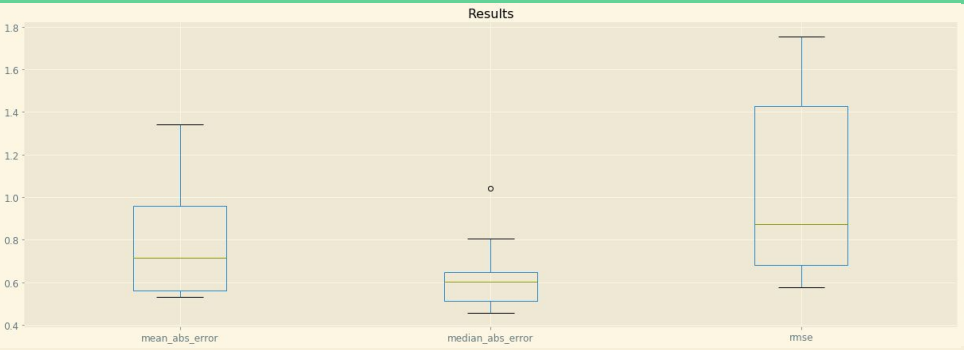


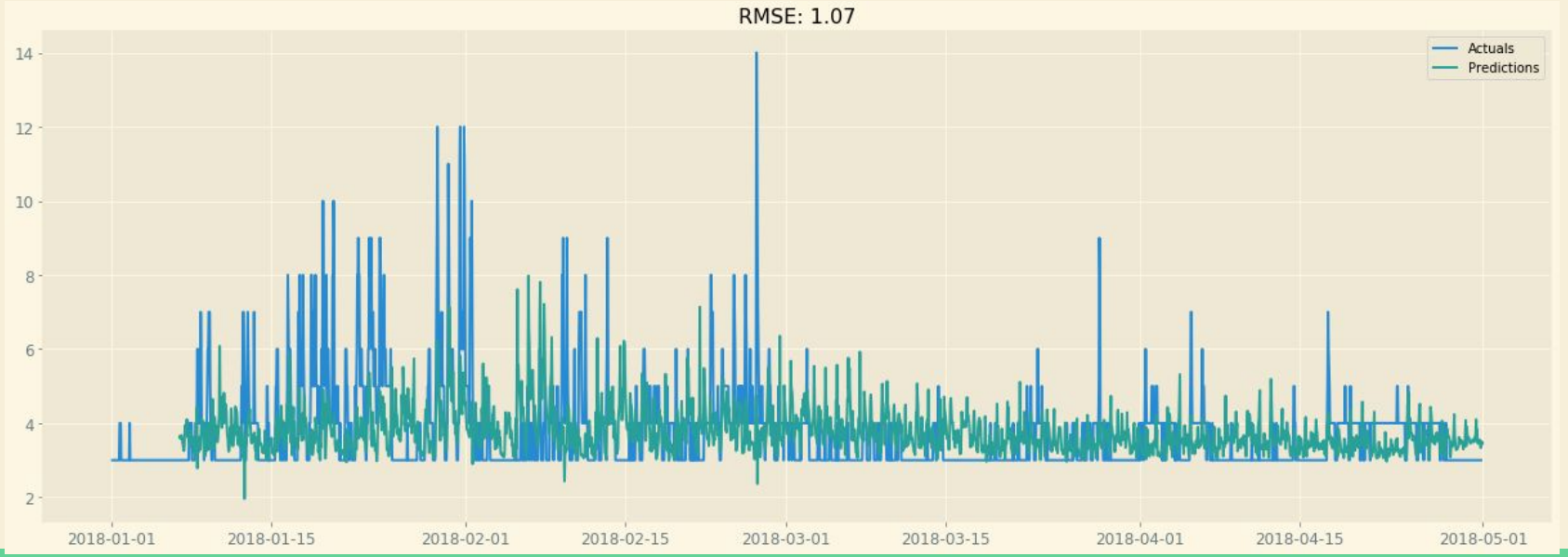
Station nr 28079018 - San Isidro/Calle Farolillo





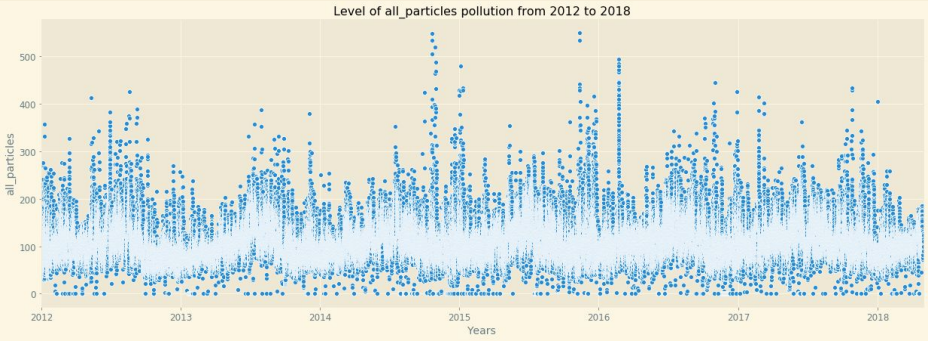
Station nr 28079018 - San Isidro/Calle Farolillo

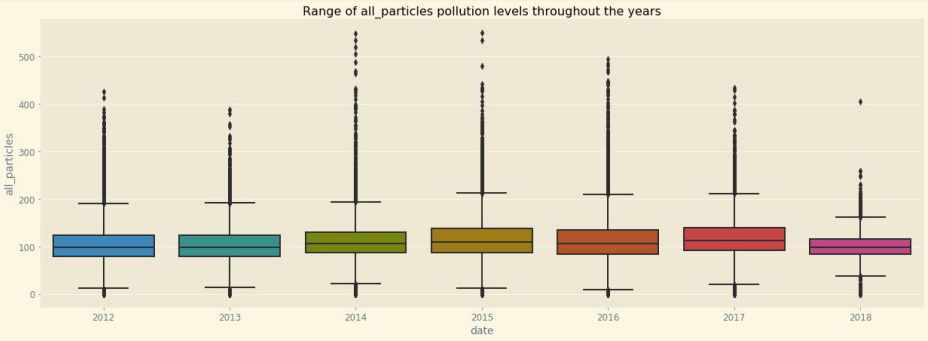


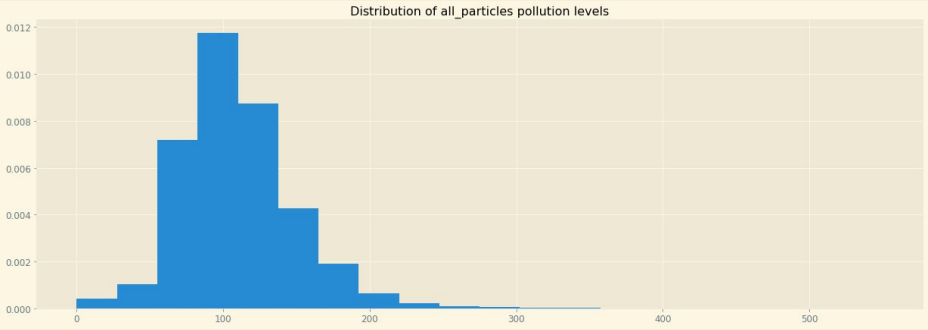


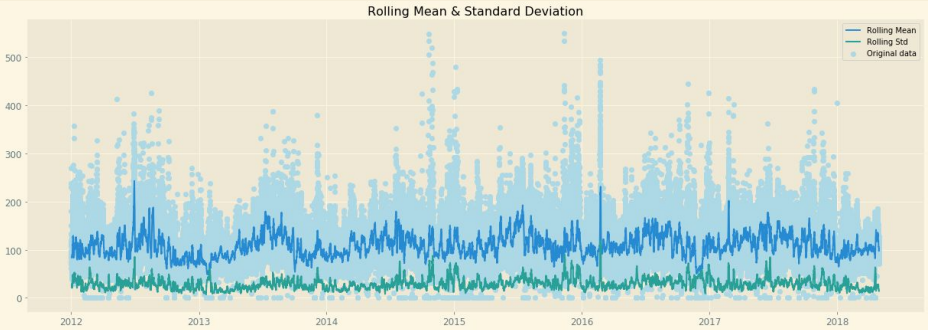
All Particles

Visualising the target variable









Testing Stationarity

Results of Dickey-Fuller Test:

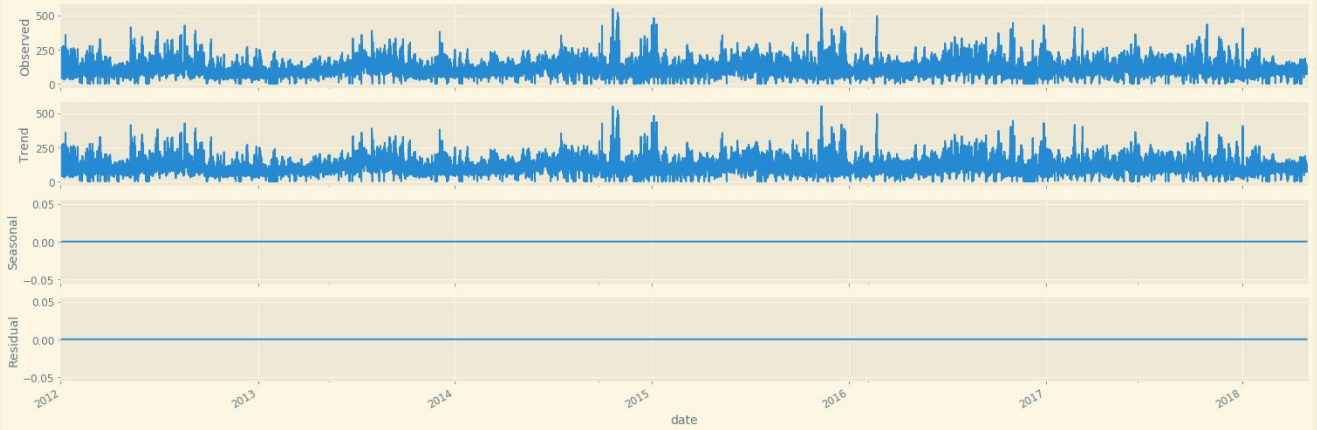
Null Hypothesis: Unit Root Present
Test Statistic < Critical Value => Reject Null
P-Value =< Alpha(.05) => Reject Null

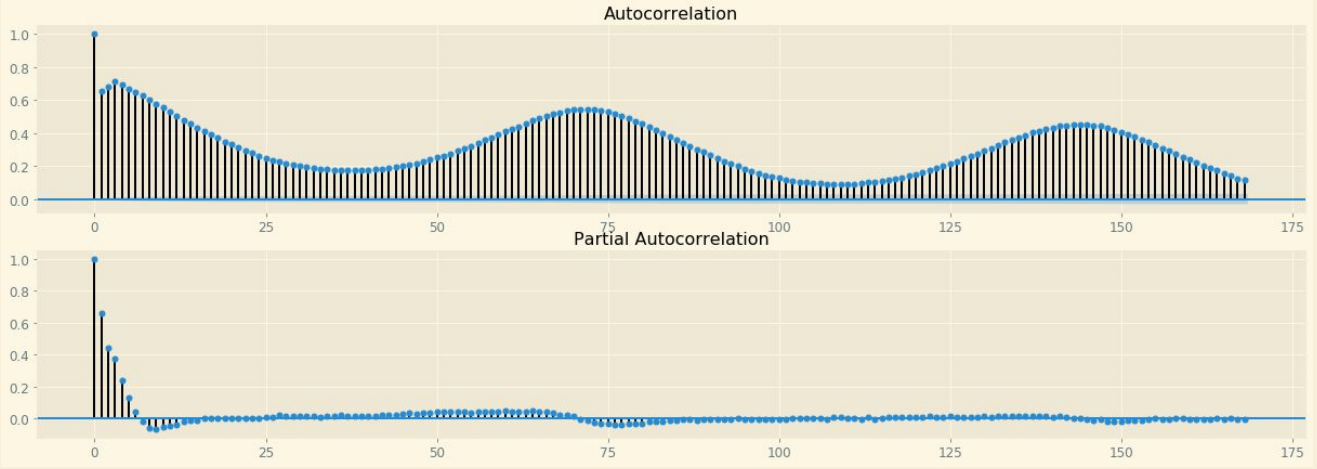
Test Statistic	-1.568 e+01
p-value	1.464 e-28
#Lags Used	7.700 e+01
Nr of Obs Used	1.664 e+05
Critical Value 1%	-3.430
Critical Value 5%	-2.862
Critical Value 10%	-2.567

Results of KPSS Test:

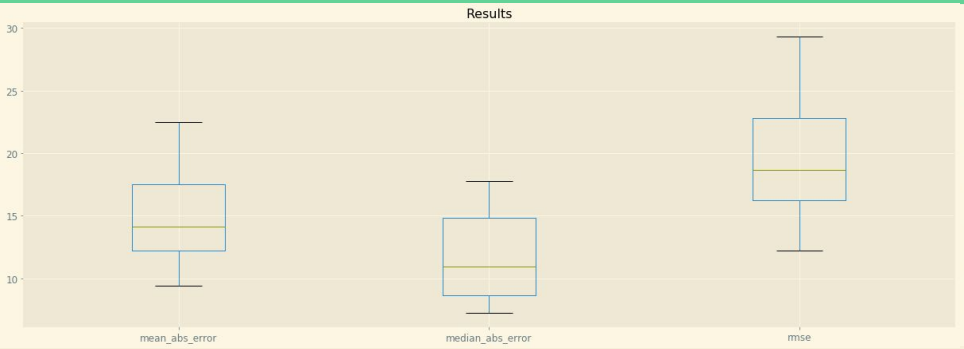
Null Hypothesis: Data is Stationary
Test Statistic > Critical Value => Reject Null
P-Value =< Alpha(.05) => Reject Null

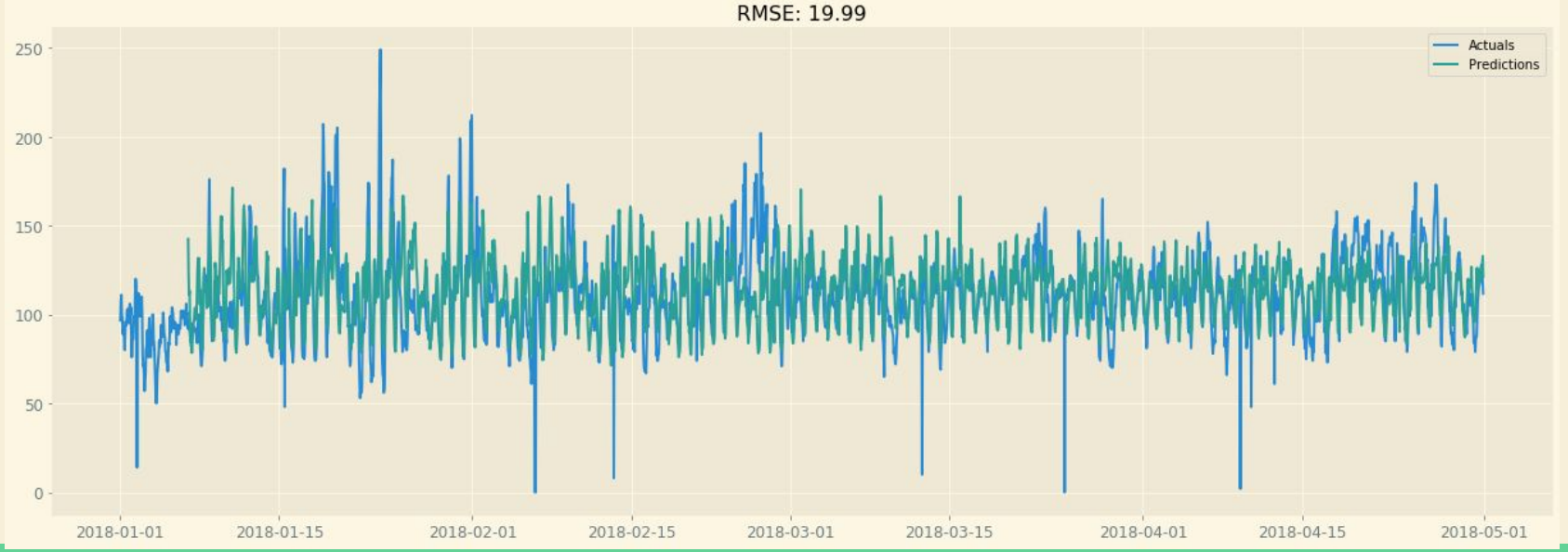
Test Statistic	5.715
p-value	0.010
Lags Used	77.000
Critical Value 10%	0.347
Critical Value 5%	0.463
Critical Value 2.5%	0.574
Critical Value 1%	0.739



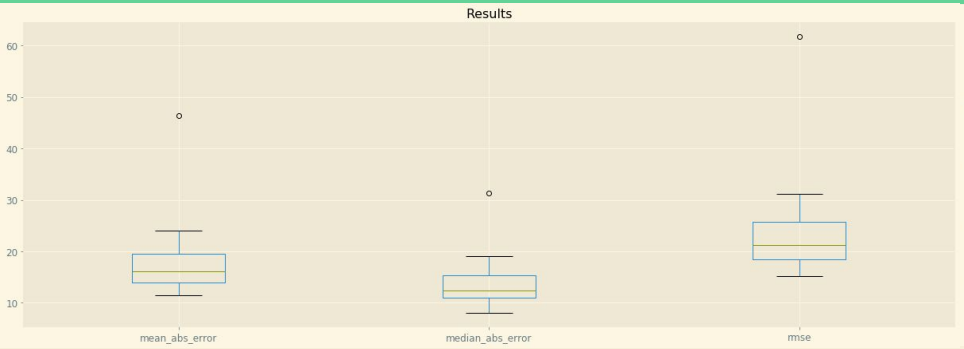


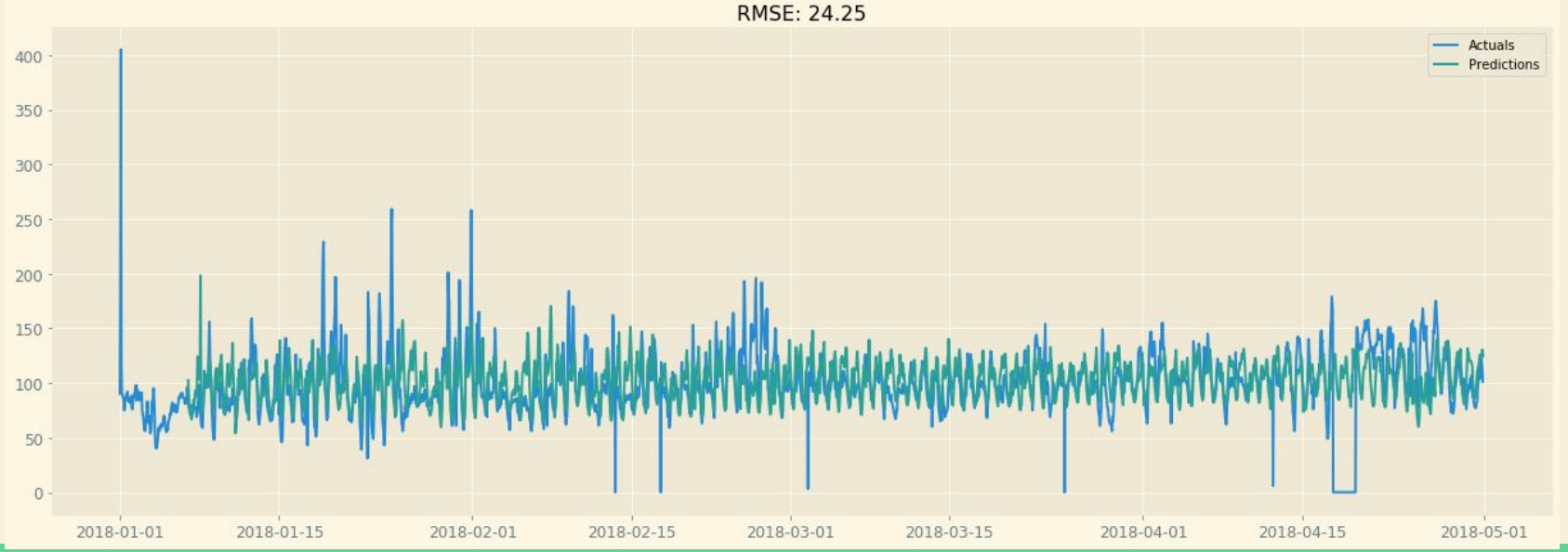
Station nr 28079008 - Parque del Retiro/ Escuelas Aguirre



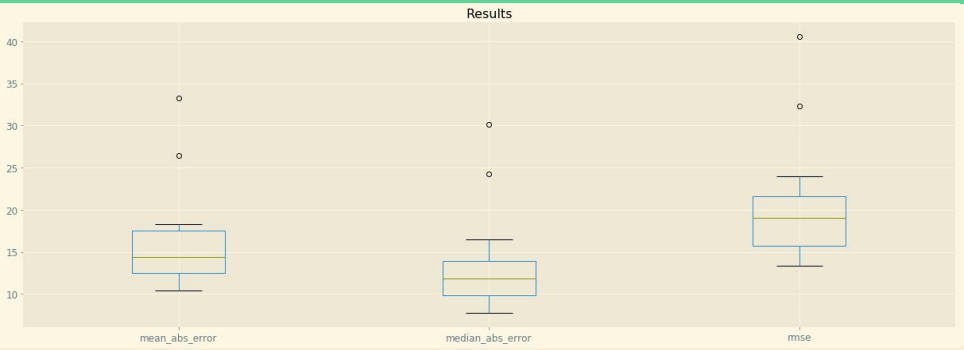


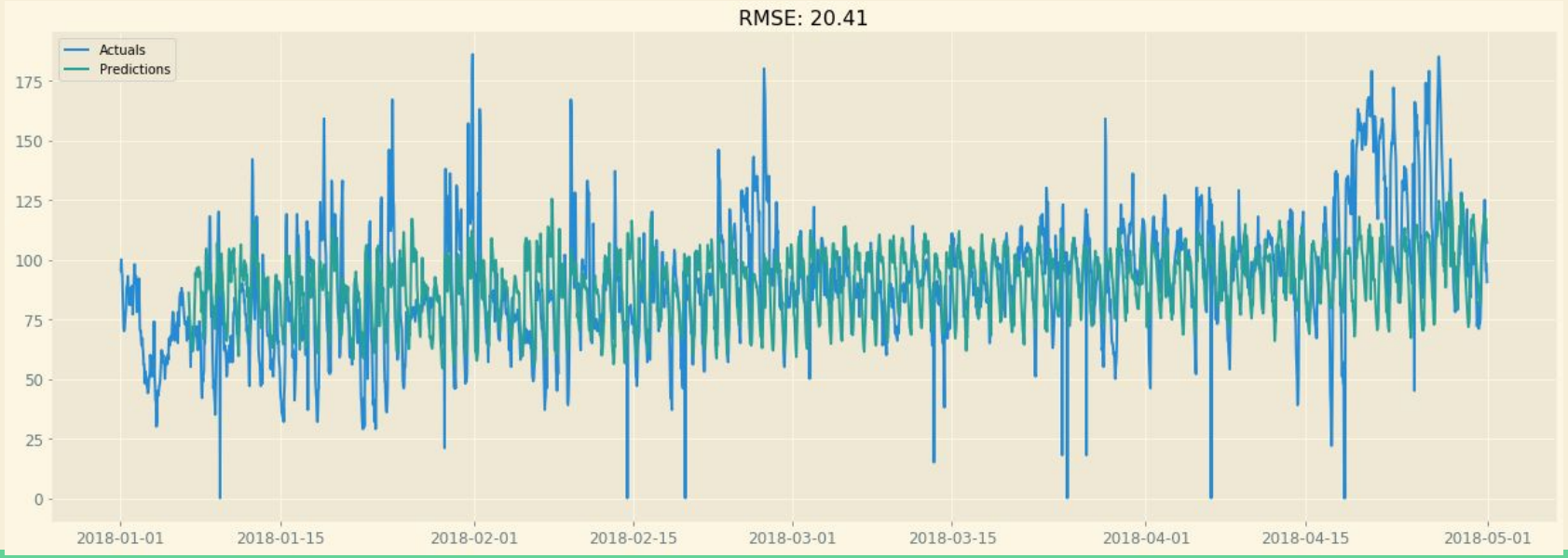
Station nr 28079018 - San Isidro/Calle Farolillo





Station nr 28079024 - Casa de Campo

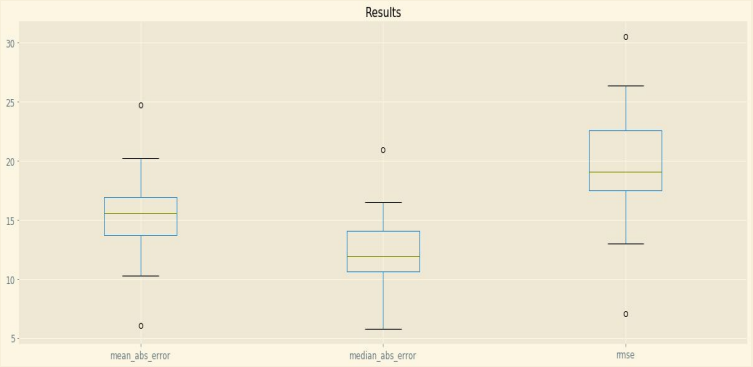


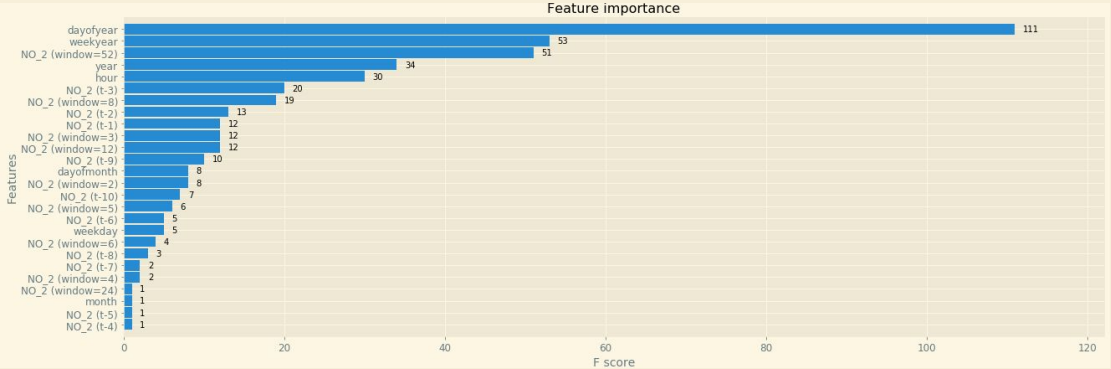


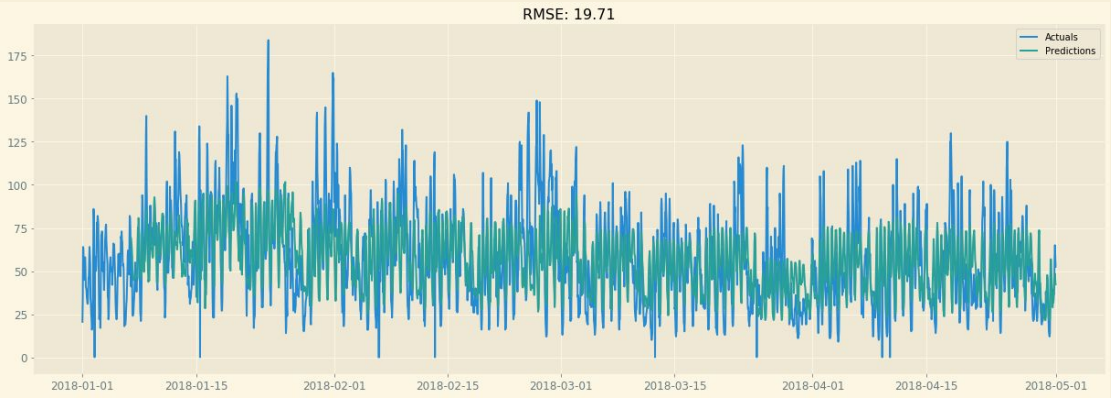
Rolling predictions with XGB

NO2

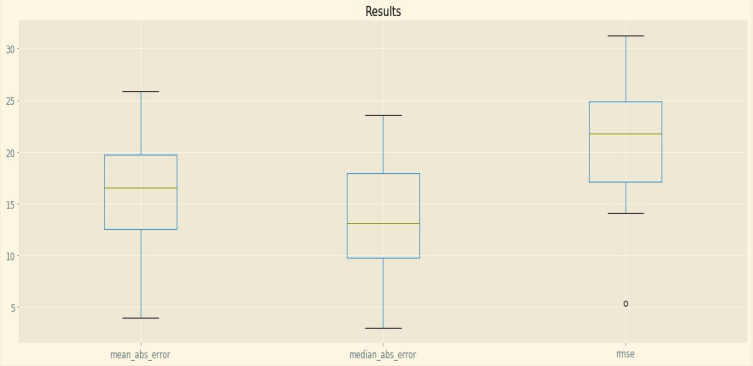
Station nr 28079008 - Parque del Retiro/ Escuelas Aguirre

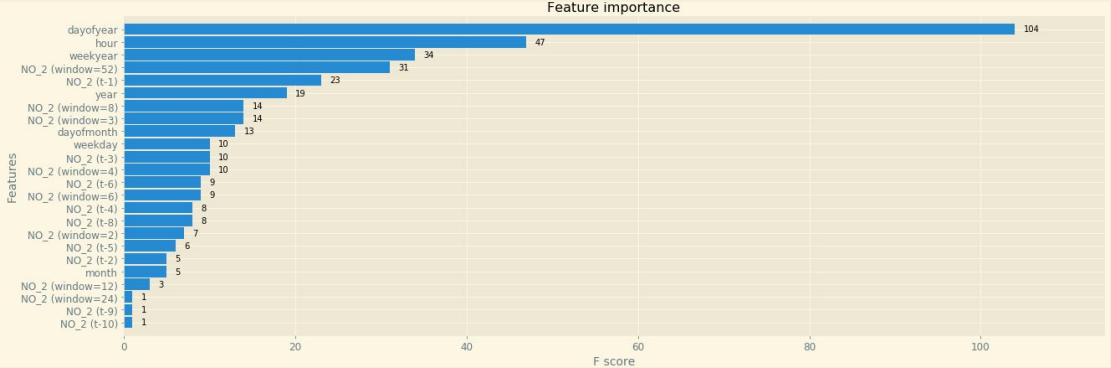


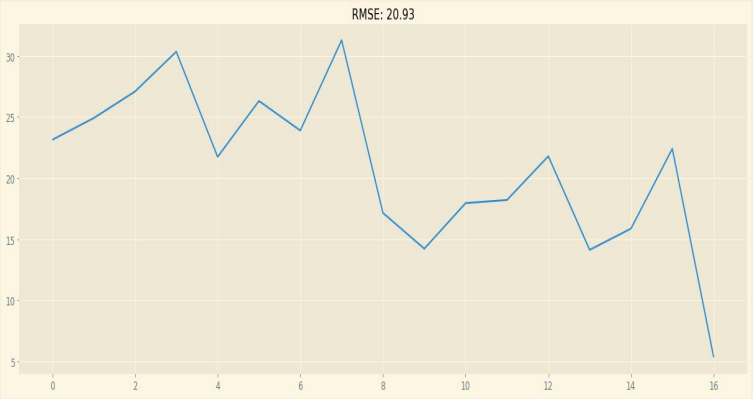


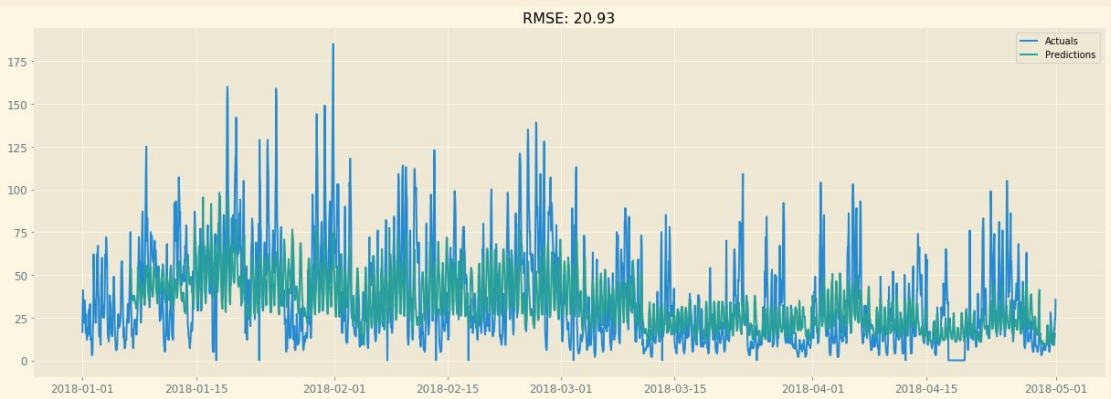


Station nr 28079018 - San Isidro/Calle Farolillo

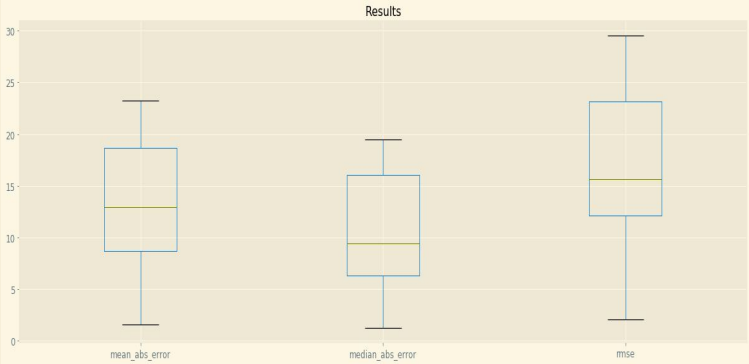


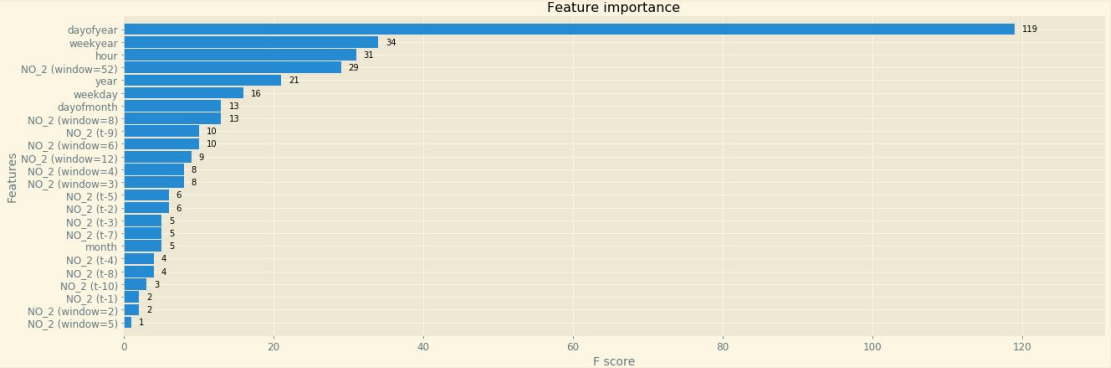


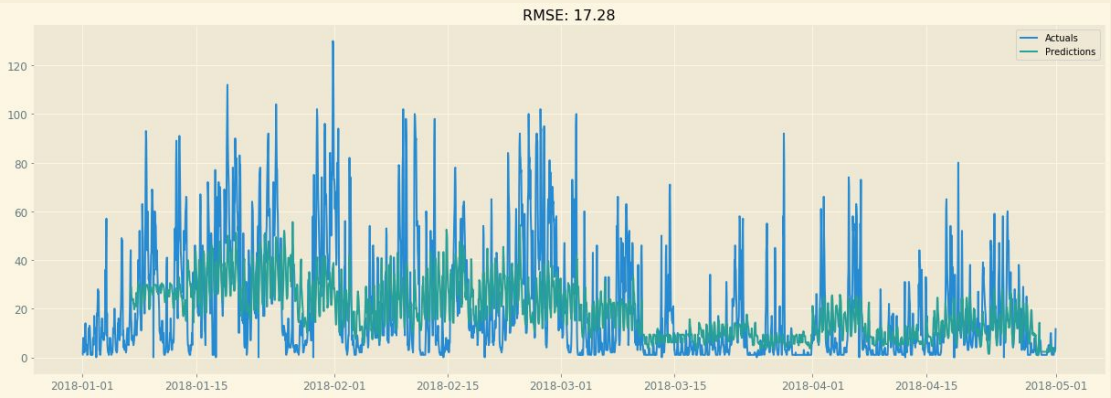




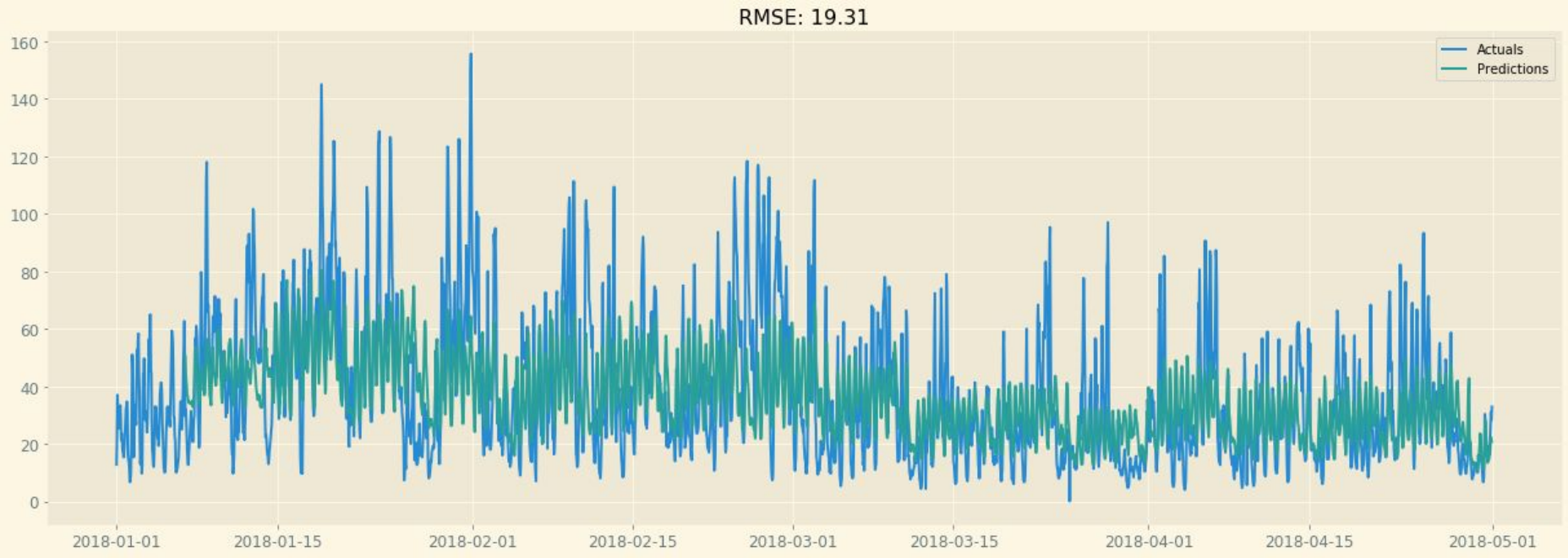
Station nr 28079024 - Casa de Campo





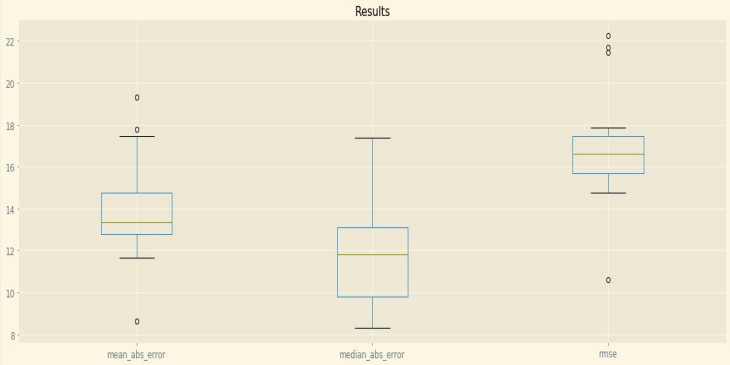


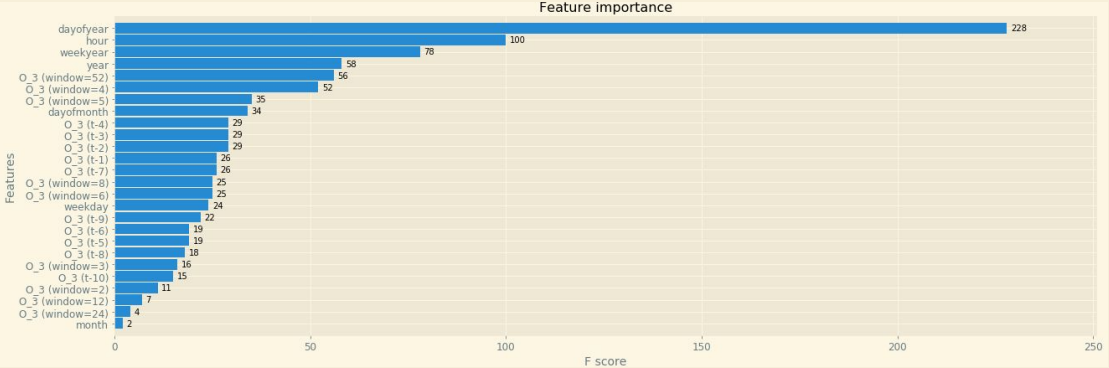
Average of the 3 stations

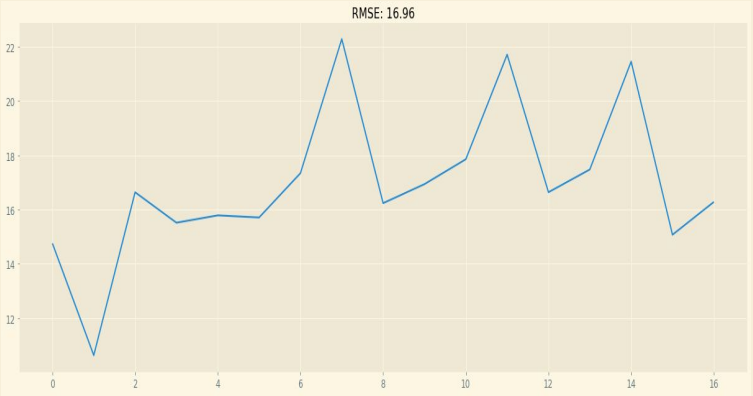


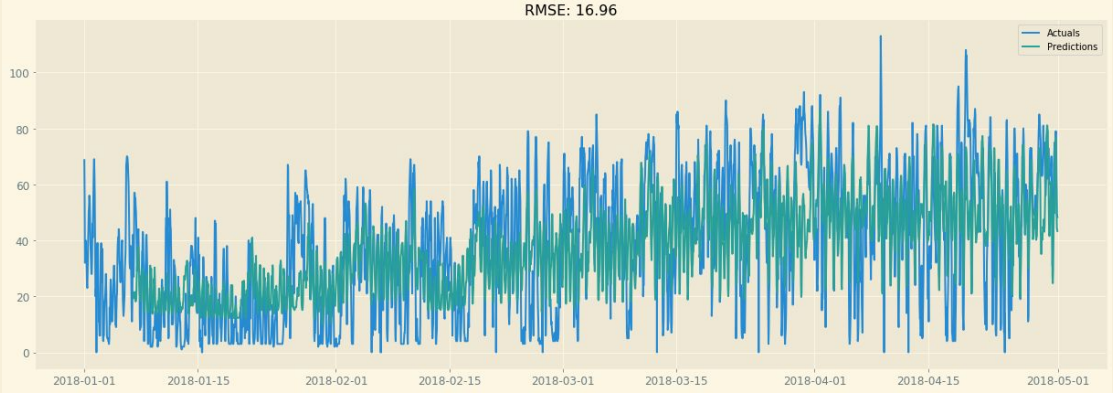
03

Station nr 28079008 - Parque del Retiro/ Escuelas Aguirre

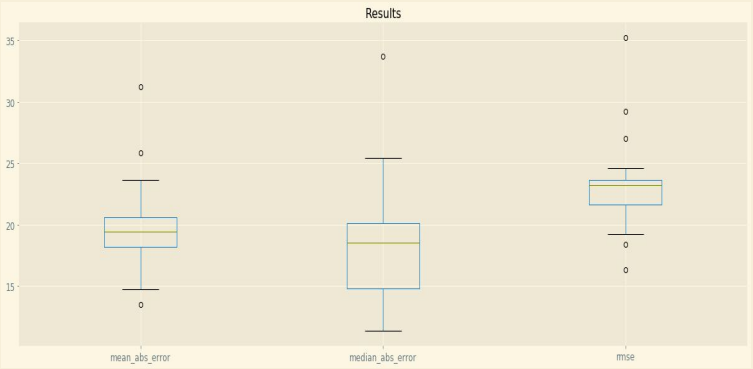


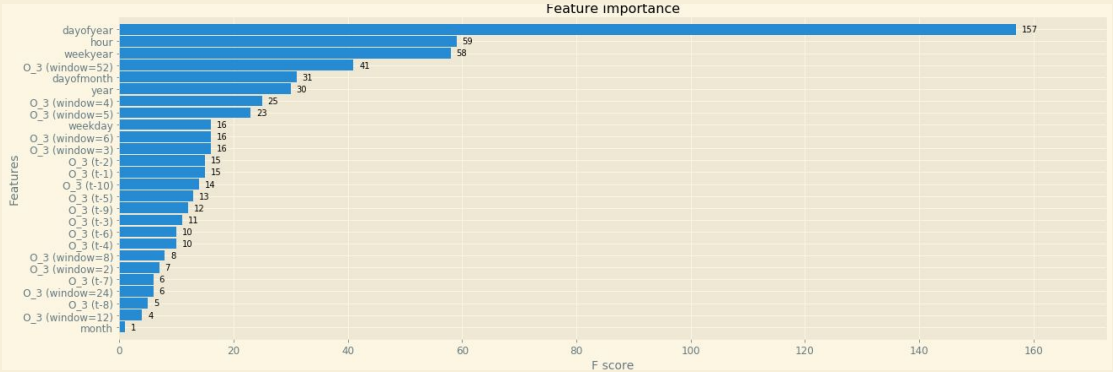


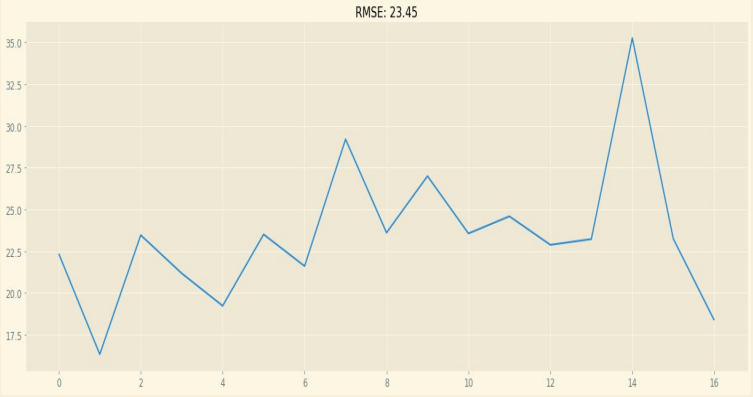


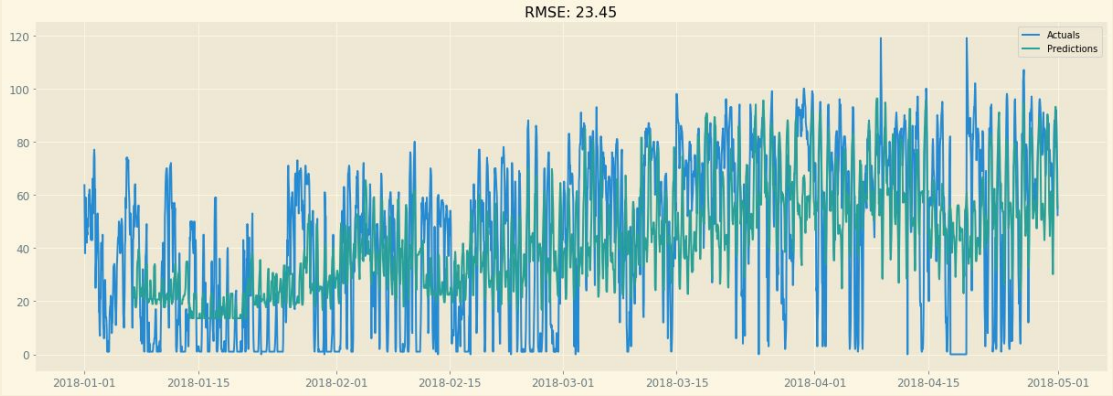


Station nr 28079018 - San Isidro/Calle Farolillo

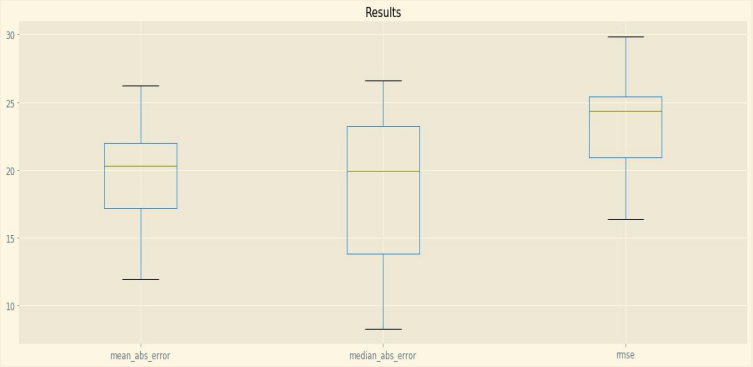


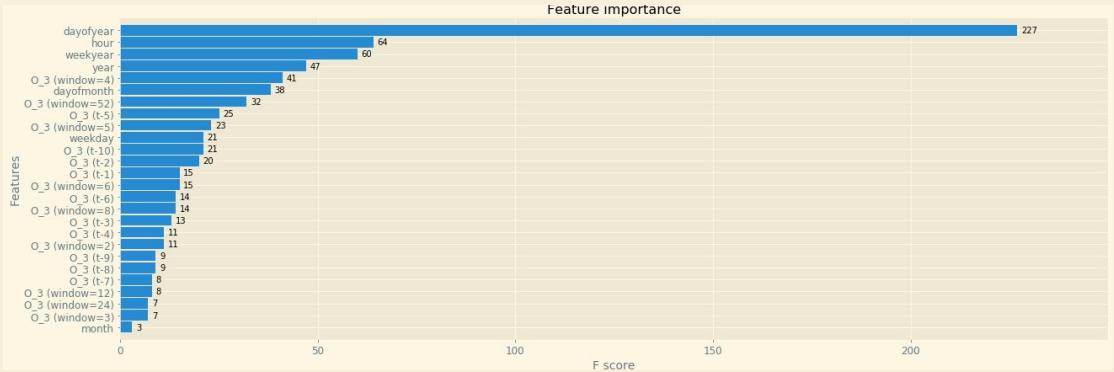


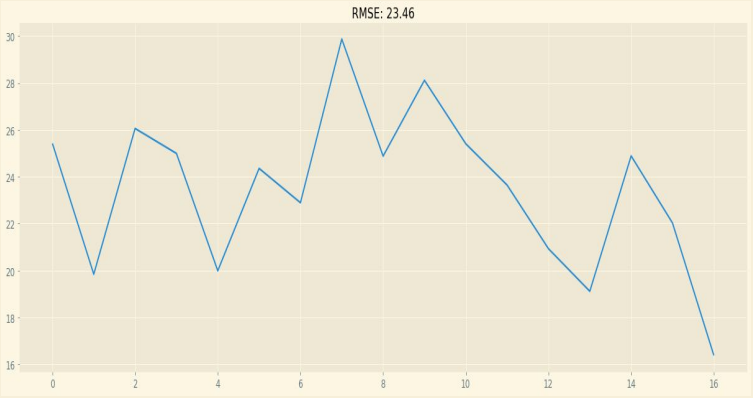


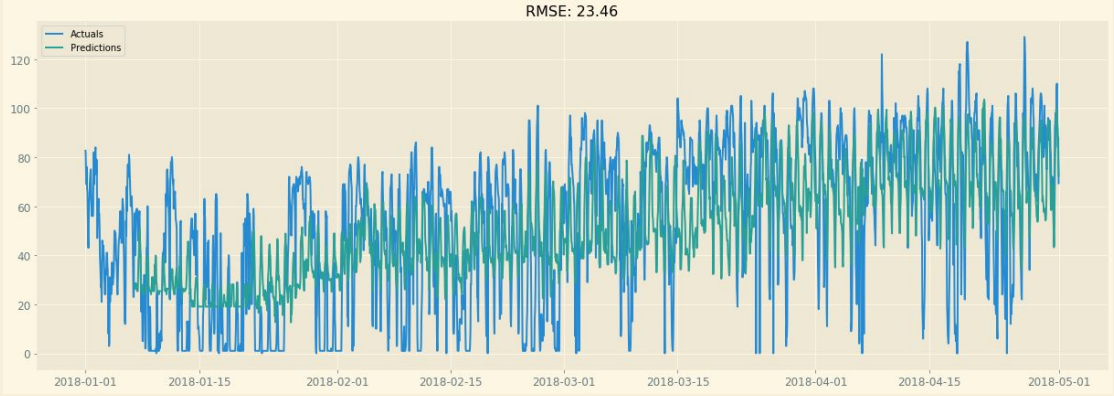


Station nr 28079024 - Casa de Campo

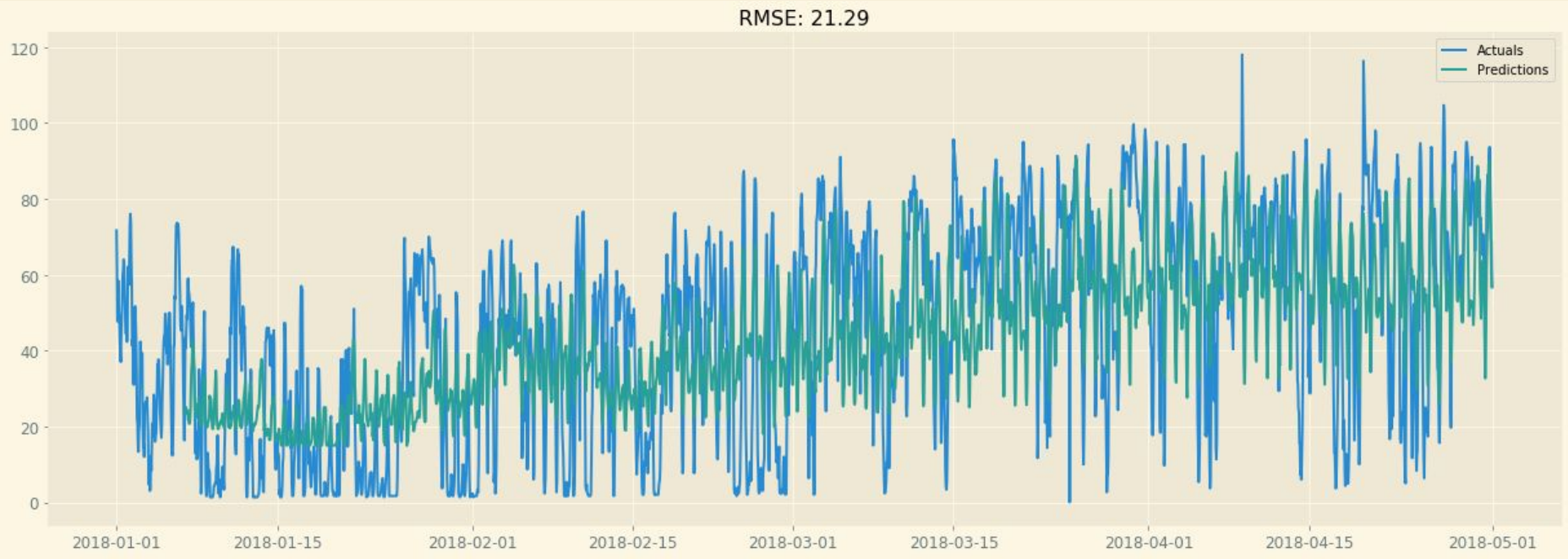






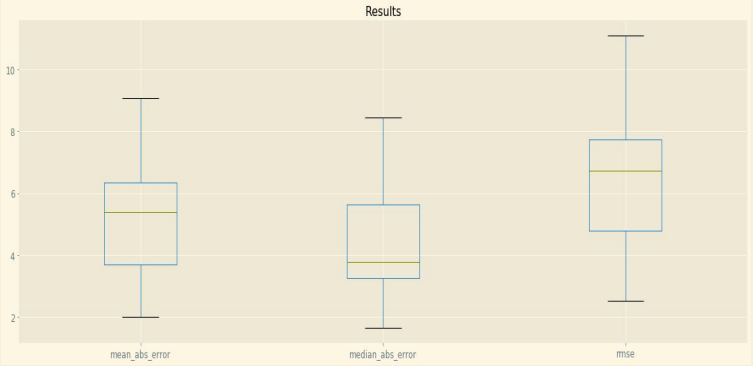


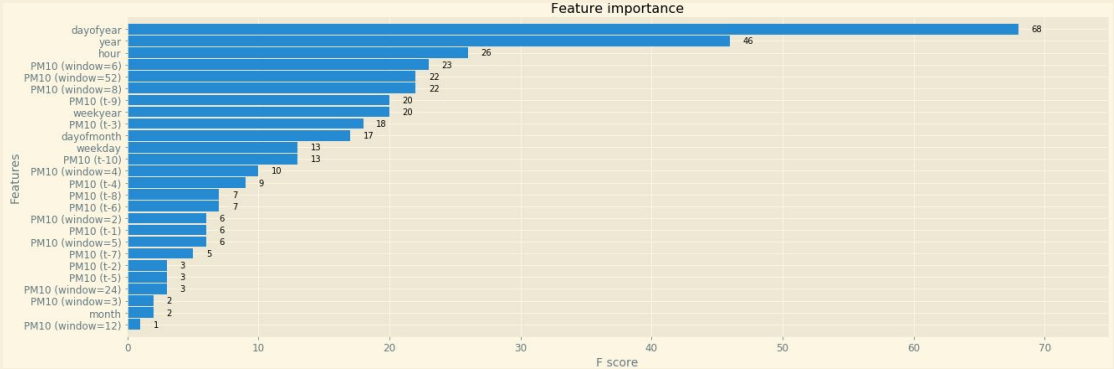
Average of the 3 stations

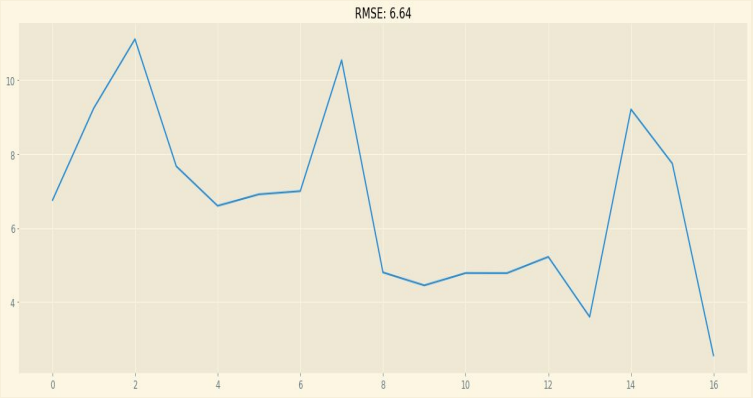


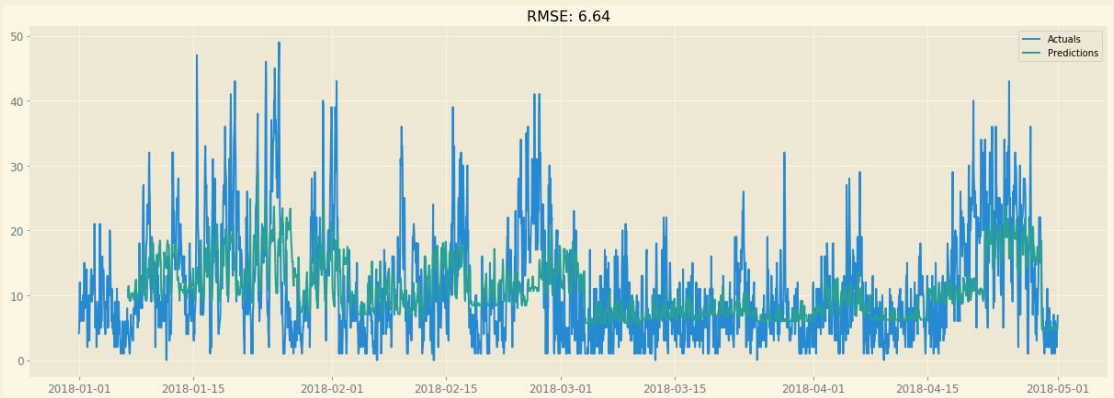
PM10

Station nr 28079008 - Parque del Retiro/ Escuelas Aguirre

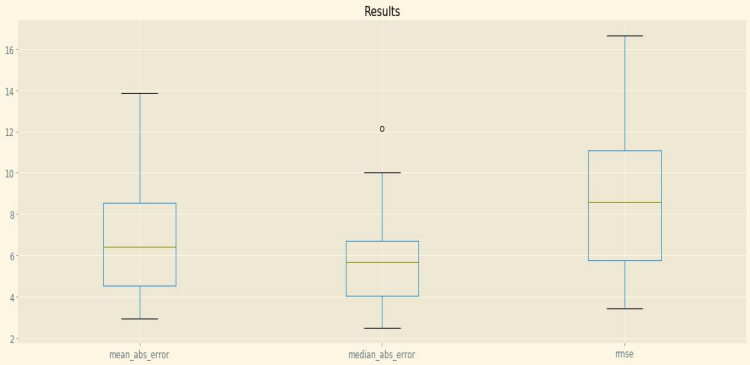


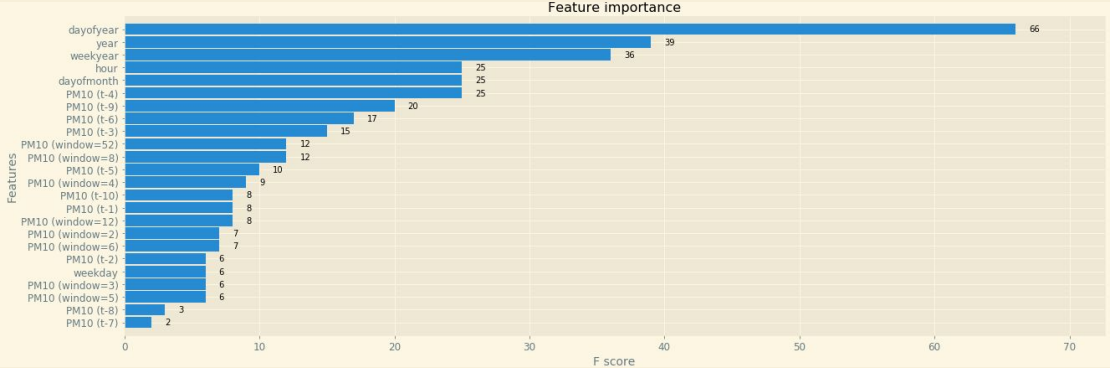


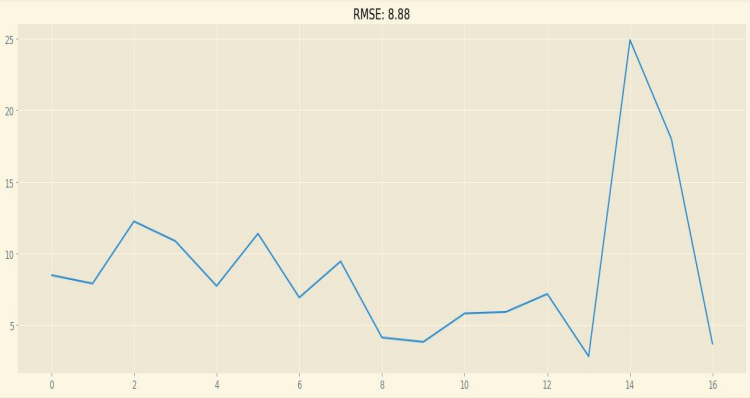


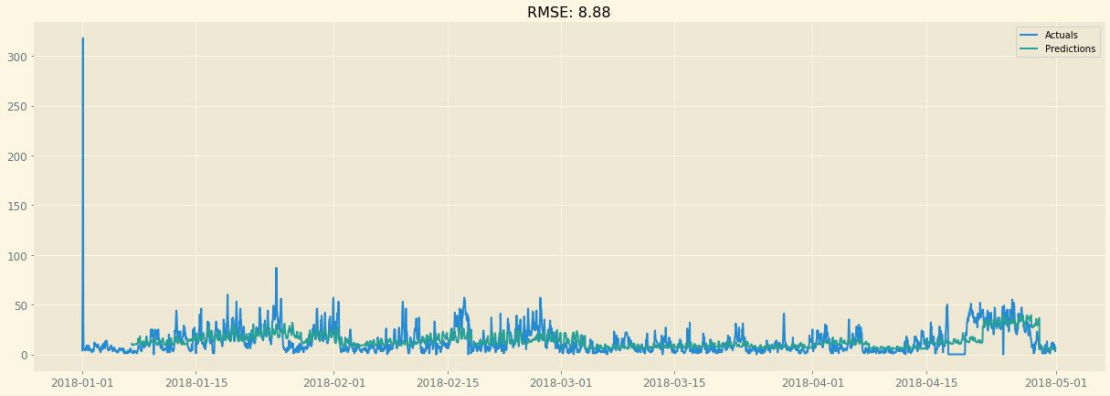


Station nr 28079018 - San Isidro/Calle Farolillo

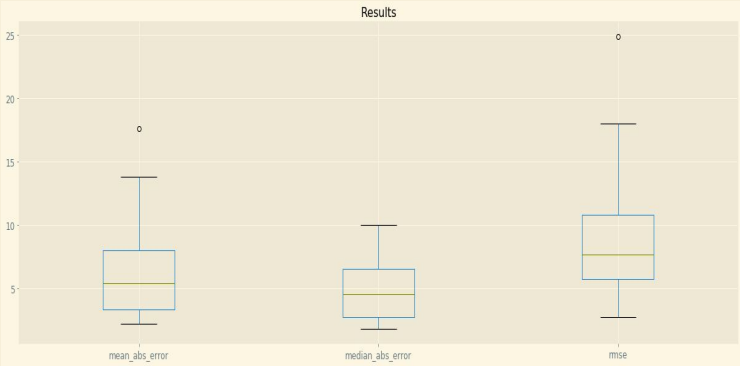


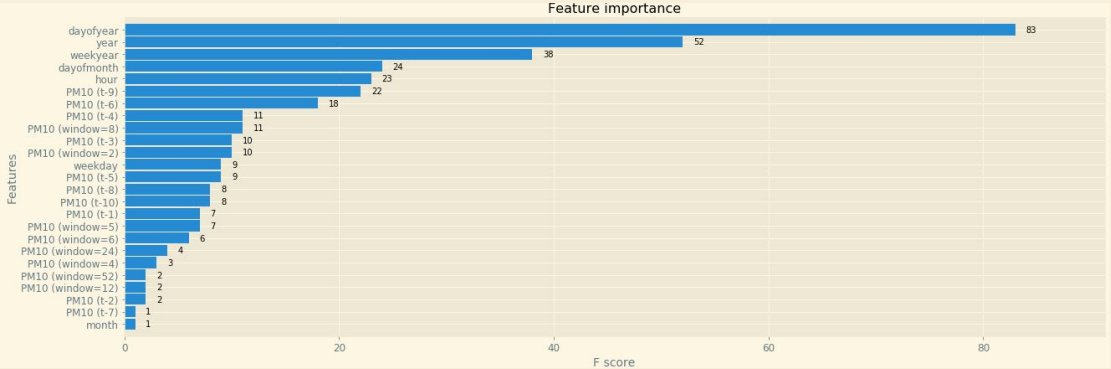


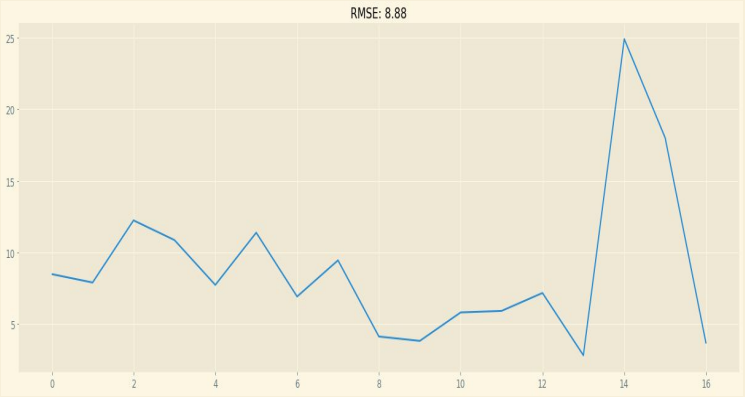


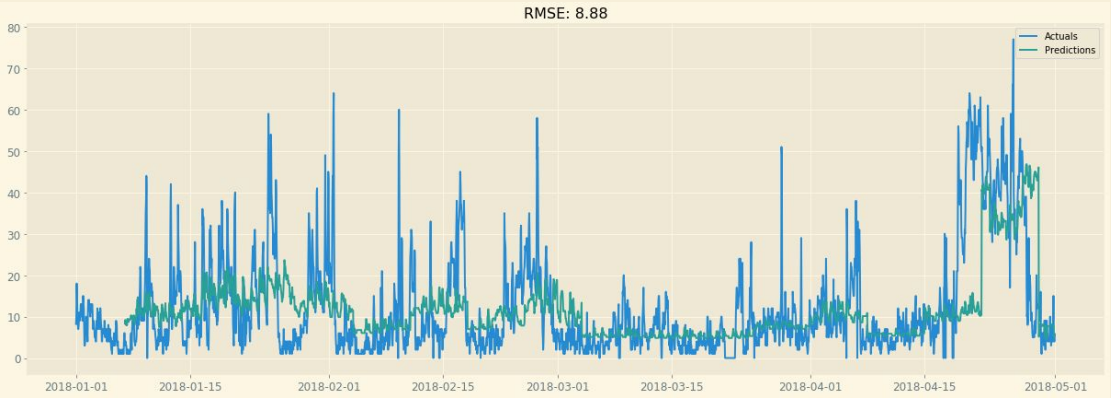


Station nr 28079024 - Casa de Campo

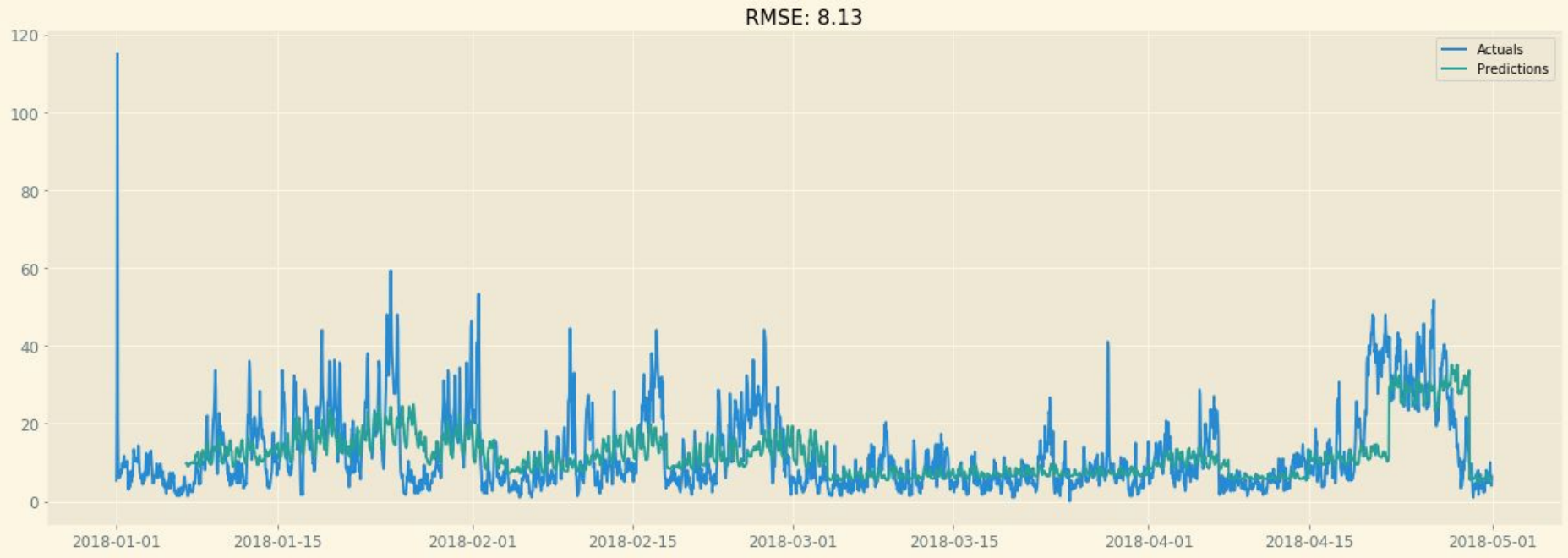






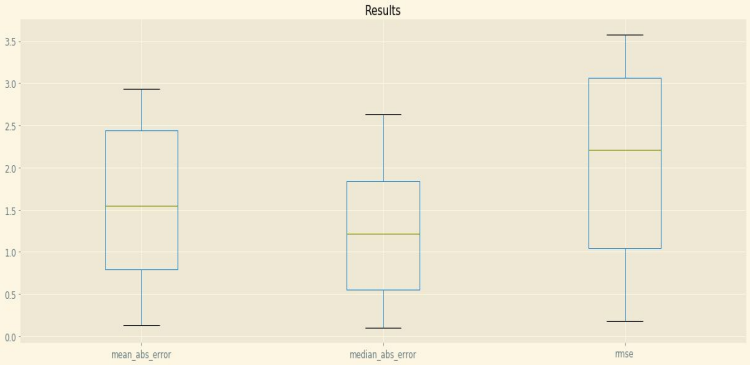


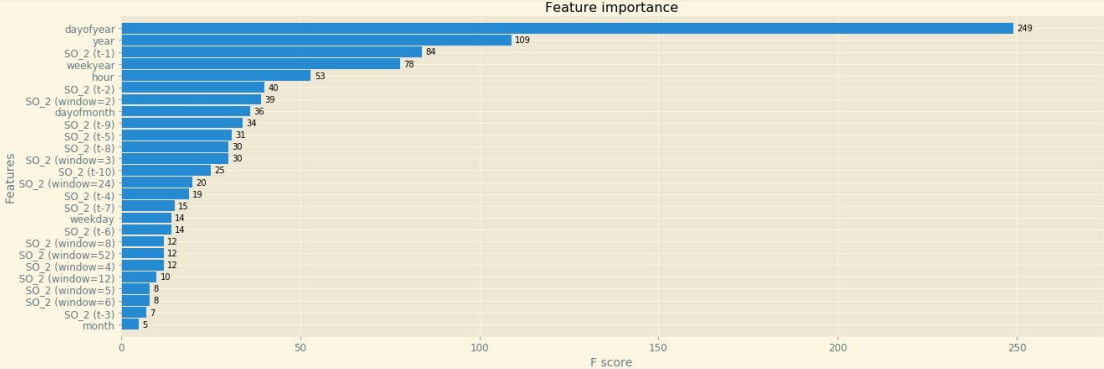
Average of the 3 stations

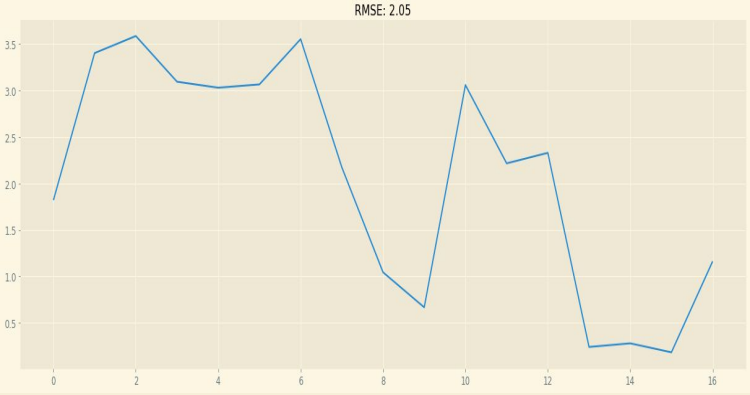


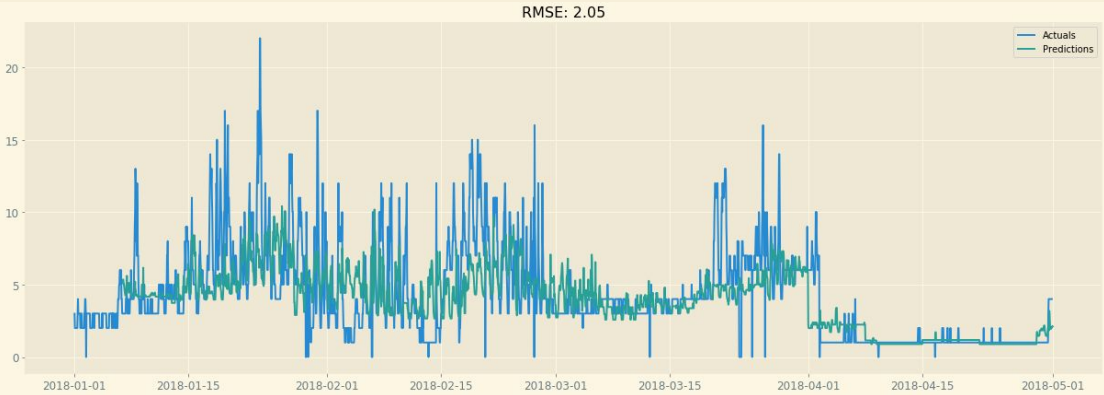
SO2

Station nr 28079008 - Parque del Retiro/ Escuelas Aguirre

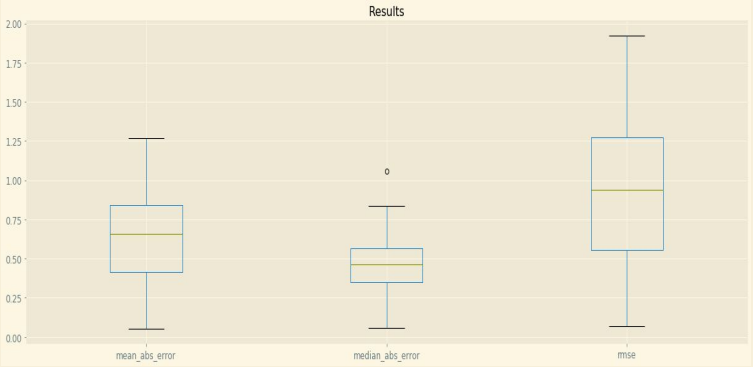


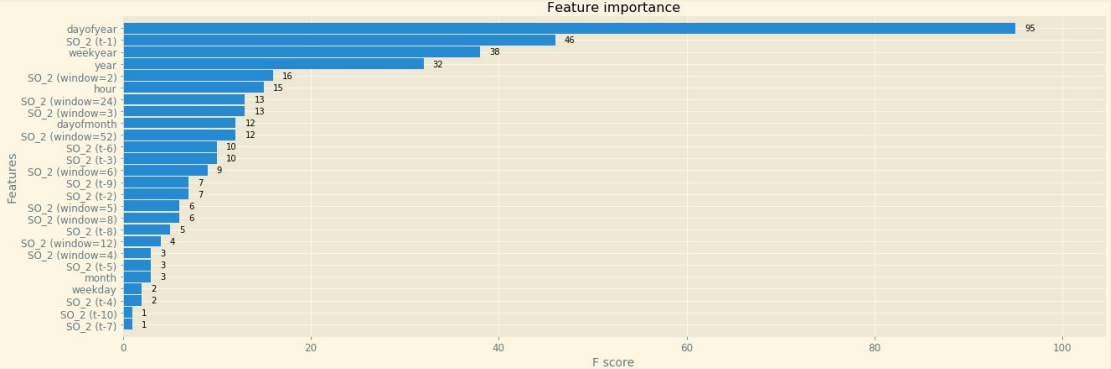


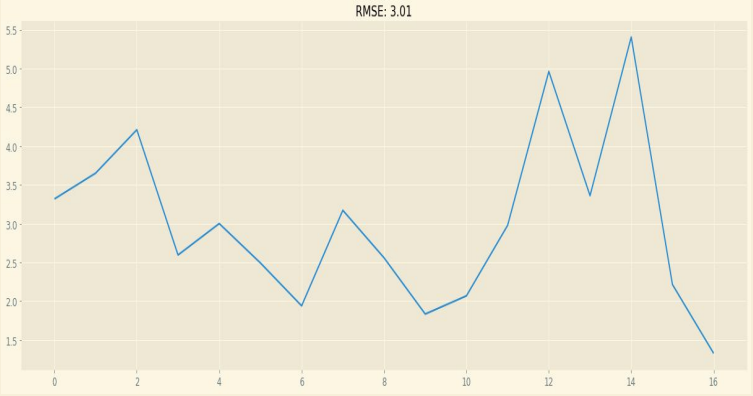


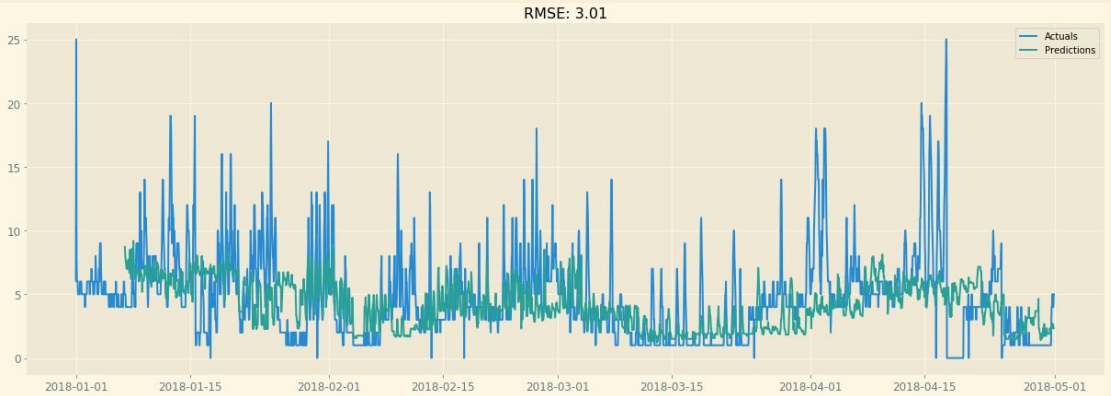


Station nr 28079018 - San Isidro/Calle Farolillo

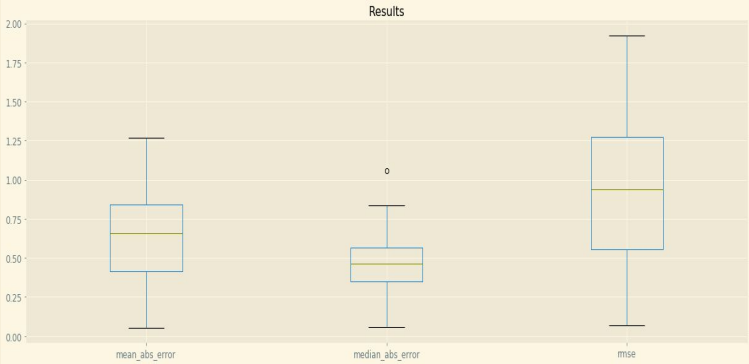


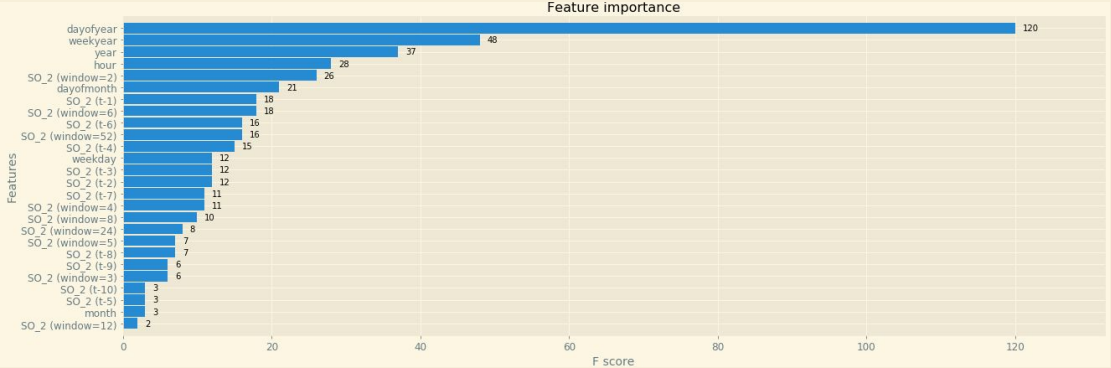


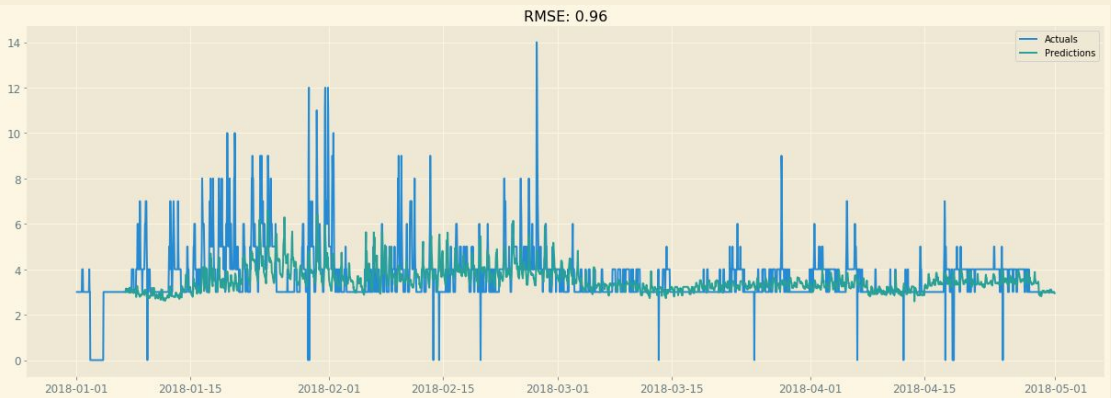




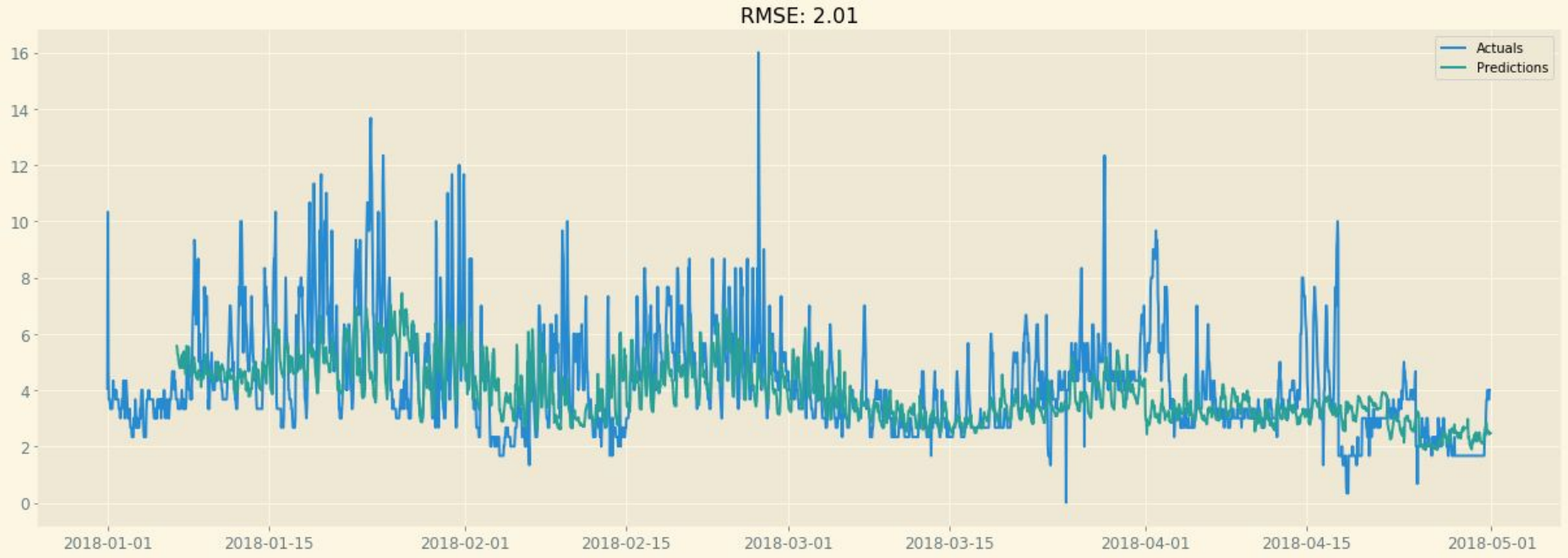
Station nr 28079024 - Casa de Campo





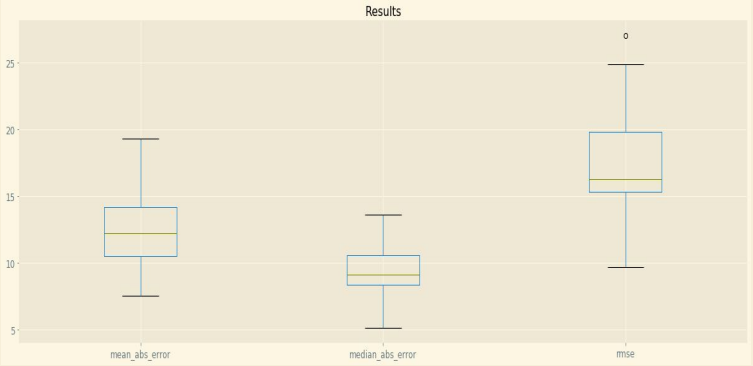


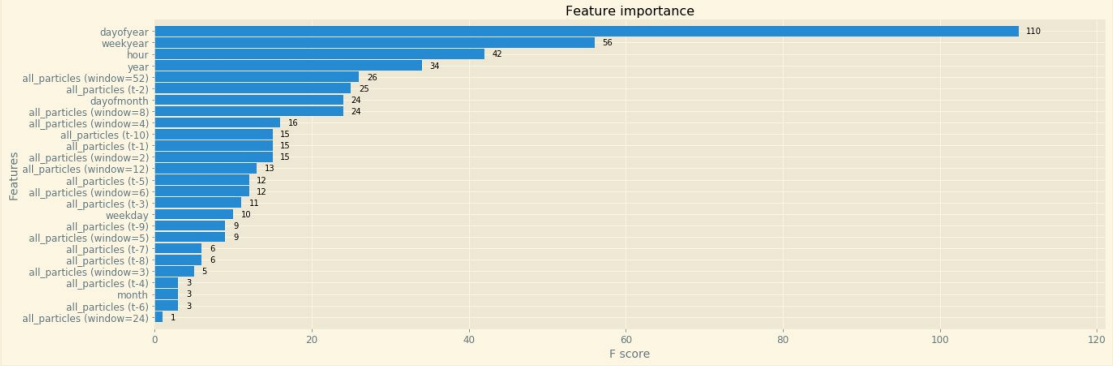
Average of the 3 stations

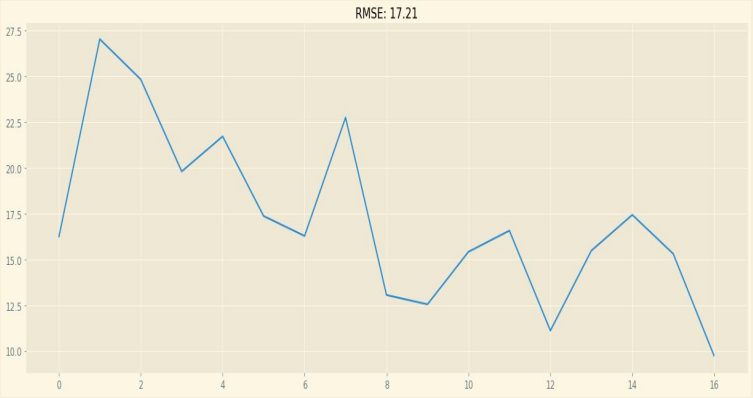


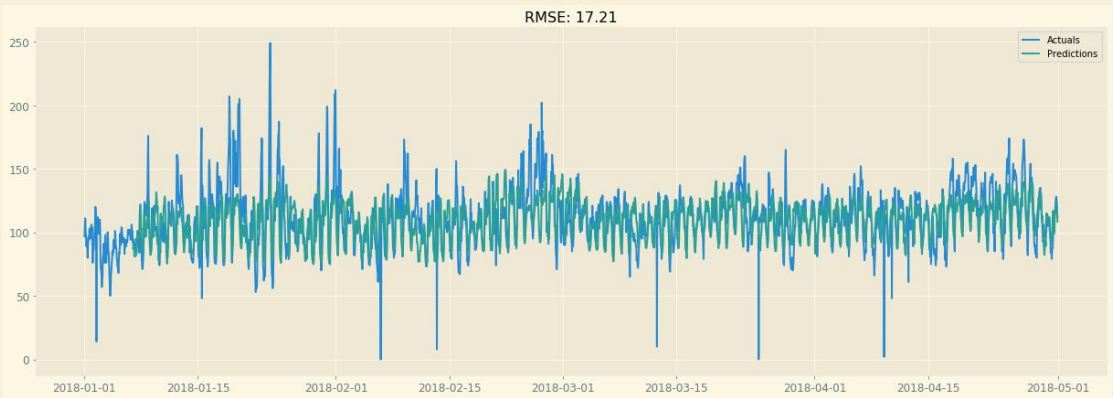
All Particles

Station nr 28079008 - Parque del Retiro/ Escuelas Aguirre

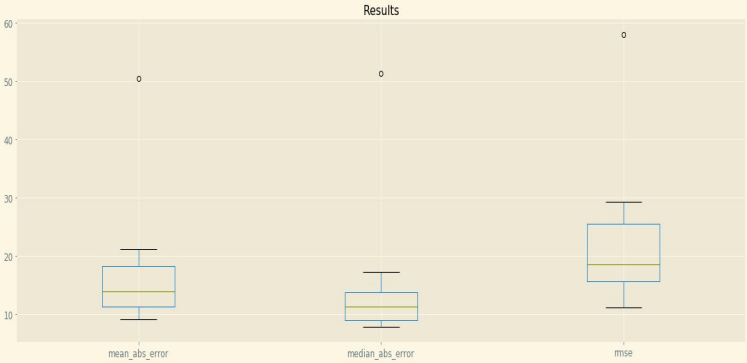


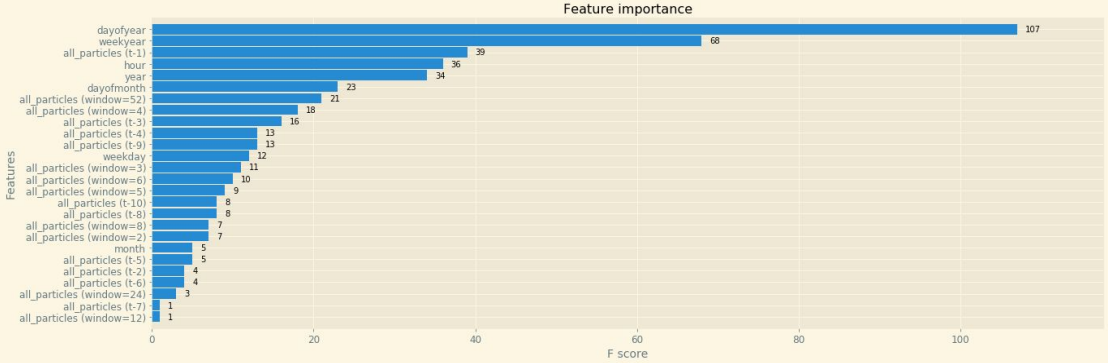


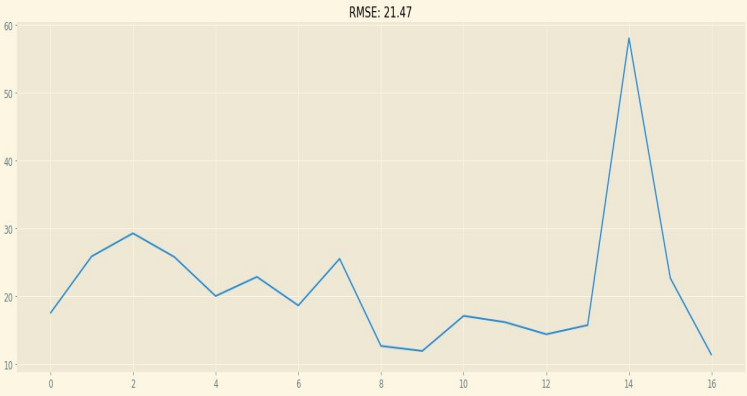


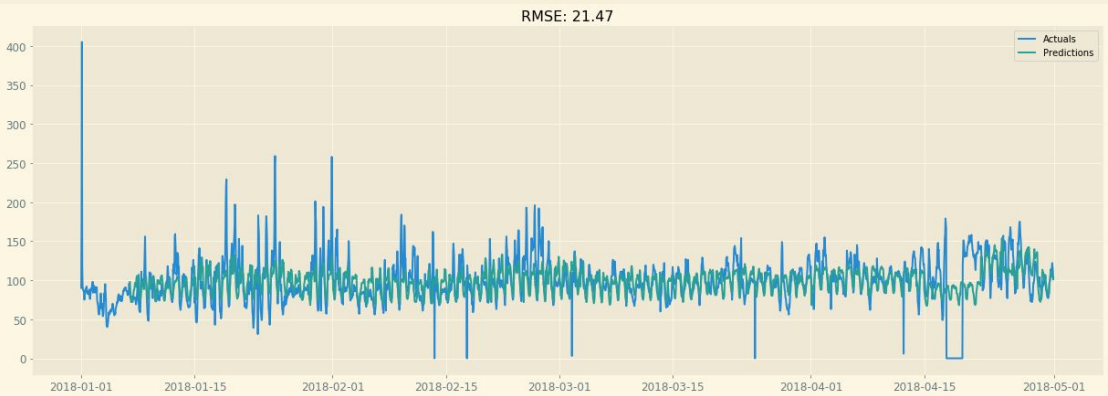


Station nr 28079018 - San Isidro/Calle Farolillo

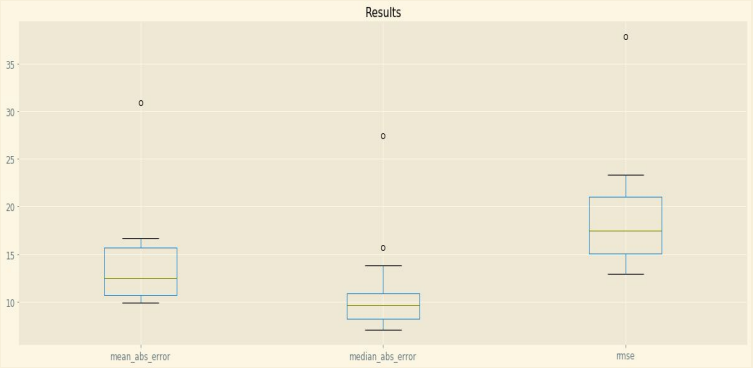


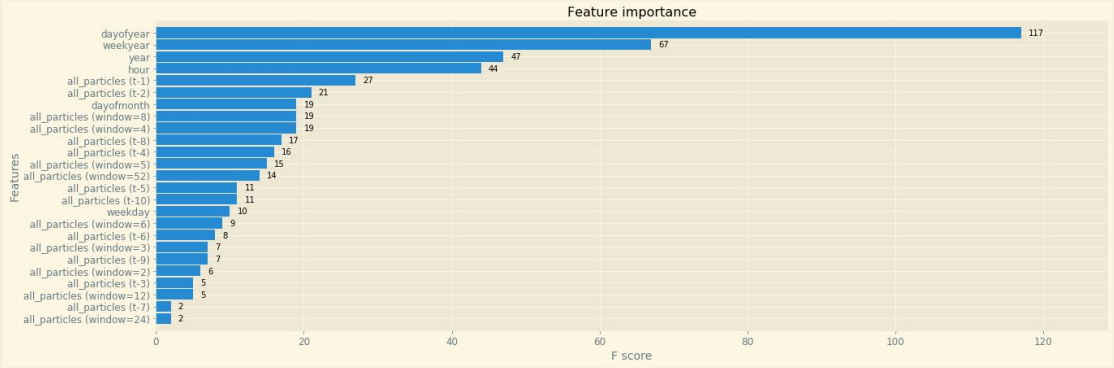


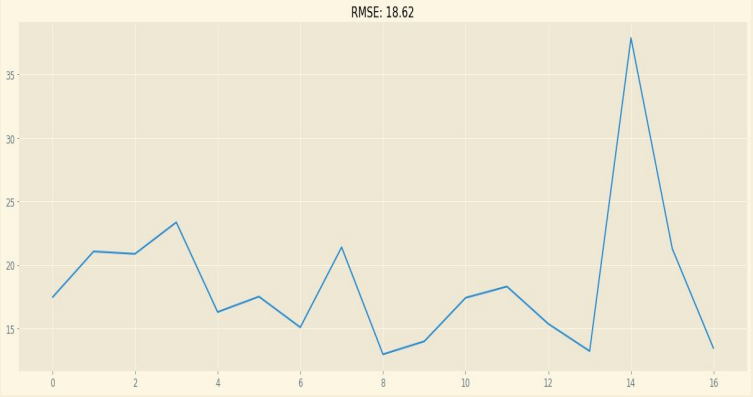


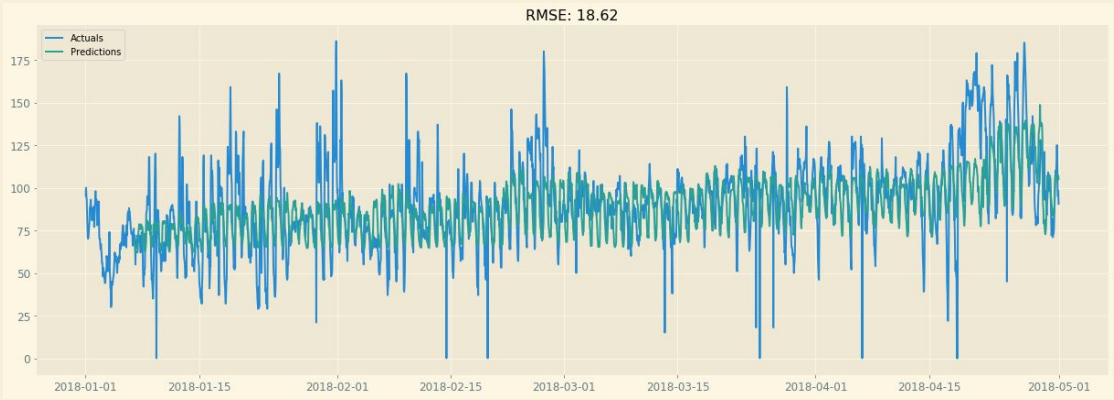


Station nr 28079024 - Casa de Campo









Average of the 3 stations