Bike Point

An analysis of London ´s bike sharing system

The future of urban mobility is shared.

Challenges

Urban areas are :

- Home to over 70 % of the EU’s population.
- Account for 85 % of the EU’s GDP.

Urban mobility accounts for:

- 40 % of all CO2 emissions from road transport,
- up to 70 % of other pollutants from transport.
- Congestion is in and around urban areas and estimated to cost nearly €130 billion per year — more than 1 % of the EU's GDP.

EIT Urban Mobility Strategic Agenda, 2022

Bike sharing services as part of the solution.

- Present today in most big cities
- Promotes **more liveable** spaces and a **healthier population**

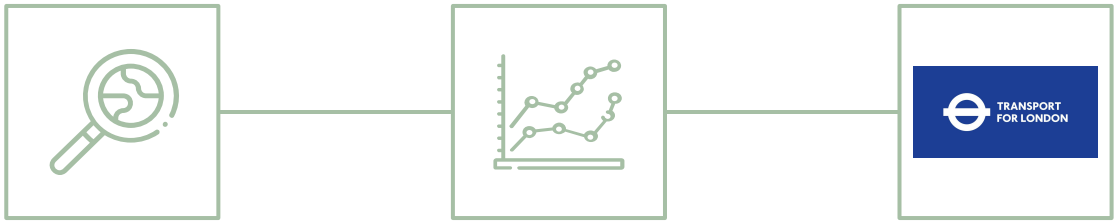
This project analyses data from **London City**:

- +/- 800 dock stations
- 12.000 bikes
- Data from 2015 to 2017

timestamp	total_count	real_temperature	felt_temperature	humidity	wind_speed	weather_code	is_holiday	is_weekend	season
2015-01-04 00:00:00	182	3.0	2.0	93.0	6.0	3.0	0.0	1.0	3.0
2015-01-04 01:00:00	138	3.0	2.5	93.0	5.0	1.0	0.0	1.0	3.0
2015-01-04 02:00:00	134	2.5	2.5	96.5	0.0	1.0	0.0	1.0	3.0
2015-01-04 03:00:00	72	2.0	2.0	100.0	0.0	1.0	0.0	1.0	3.0
2015-01-04 04:00:00	47	2.0	0.0	93.0	6.5	1.0	0.0	1.0	3.0
...	...	...	...	...	...	...	...	...	...
2017-01-03 19:00:00	1042	5.0	1.0	81.0	19.0	3.0	0.0	0.0	3.0
2017-01-03 20:00:00	541	5.0	1.0	81.0	21.0	4.0	0.0	0.0	3.0
2017-01-03 21:00:00	337	5.5	1.5	78.5	24.0	4.0	0.0	0.0	3.0
2017-01-03 22:00:00	224	5.5	1.5	76.0	23.0	4.0	0.0	0.0	3.0
2017-01-03 23:00:00	139	5.0	1.0	76.0	22.0	2.0	0.0	0.0	3.0

17414 rows × 9 columns

Project steps



Step 1

Insights from the data

Step 2

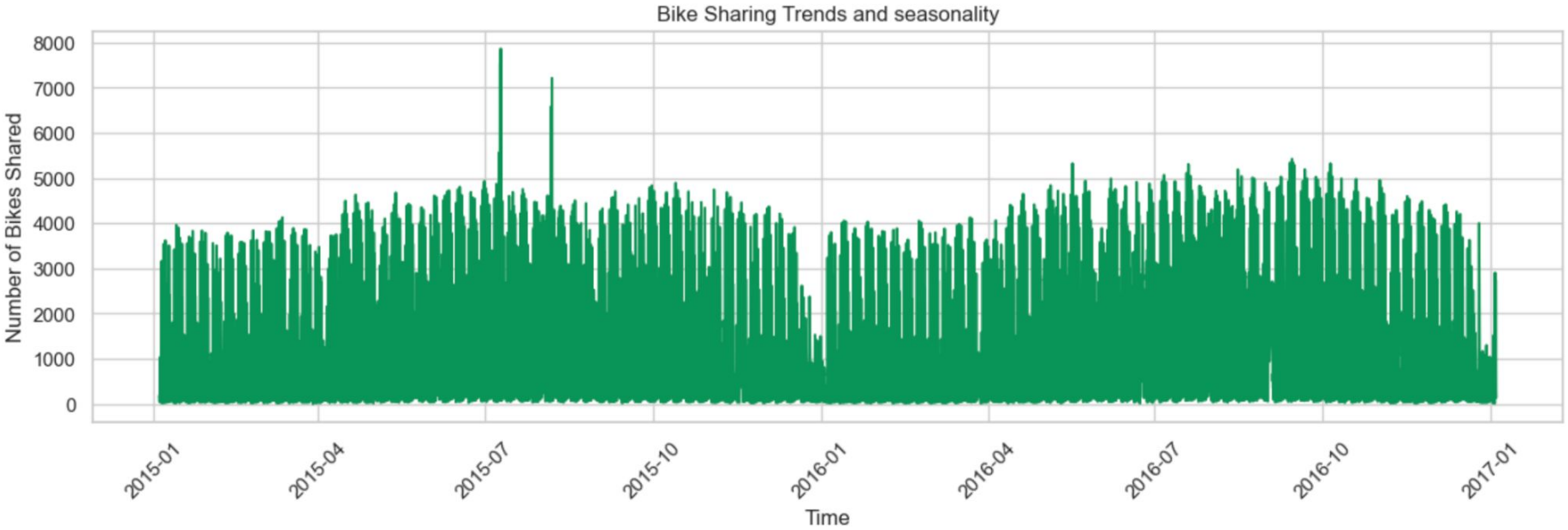
Demand forecast

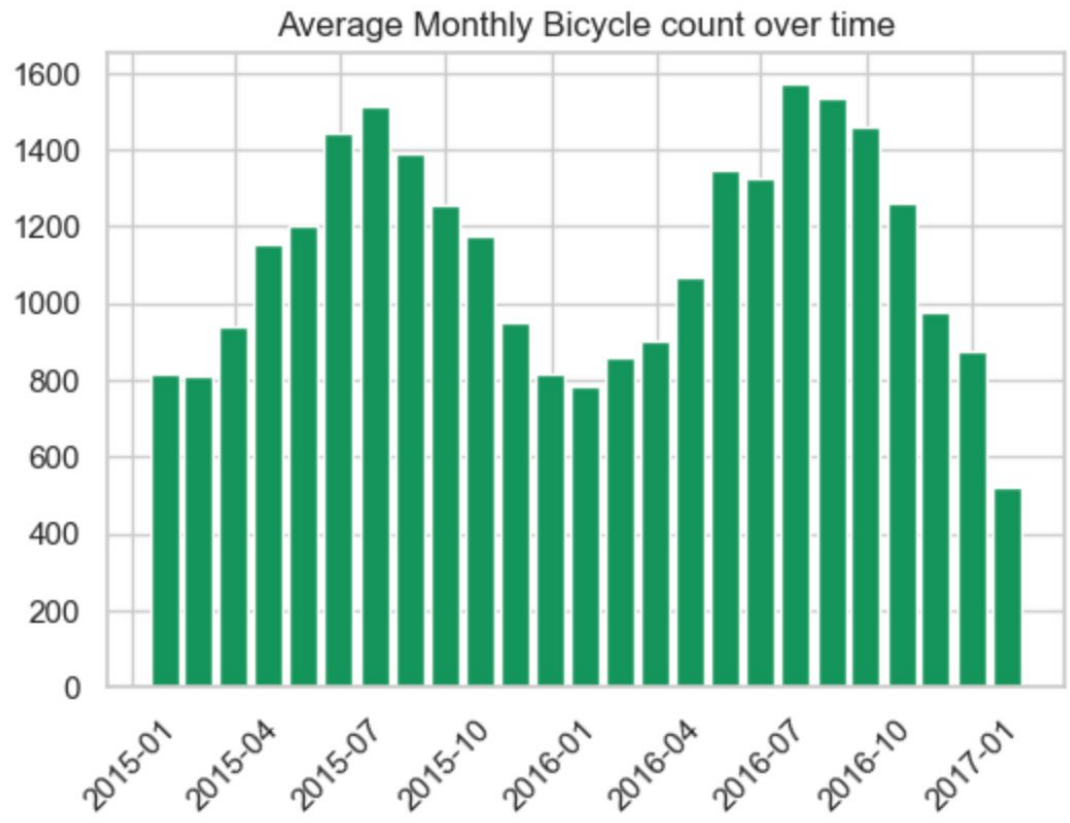
Step 3

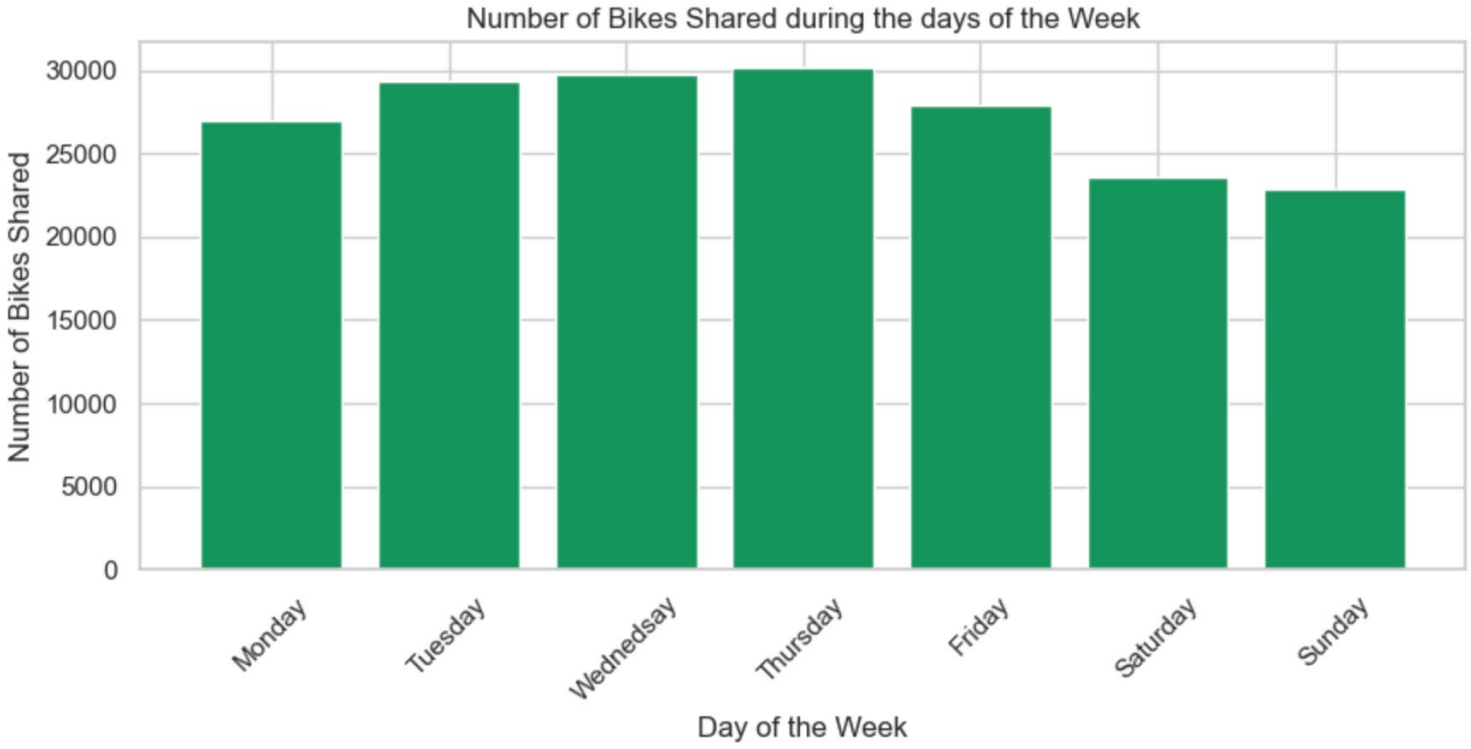
API data from TfL

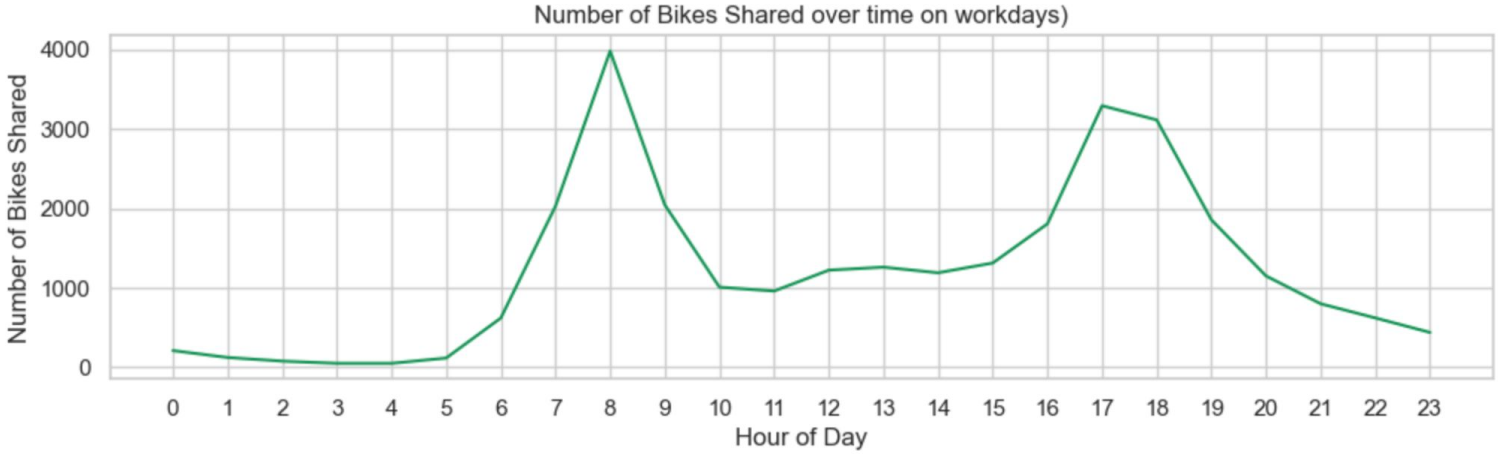
01

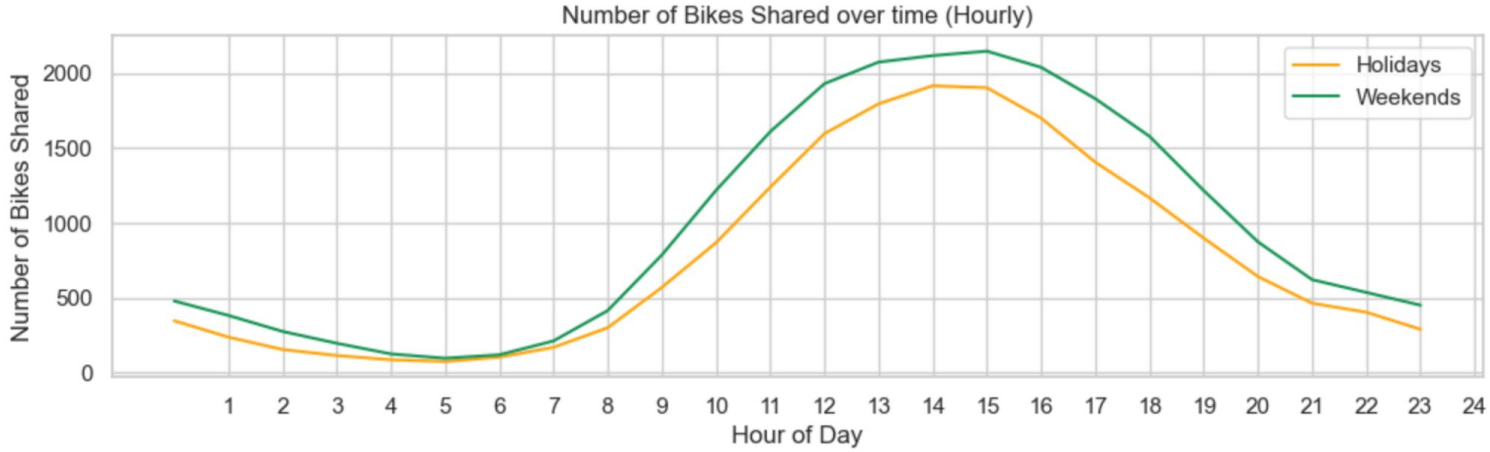
Insights

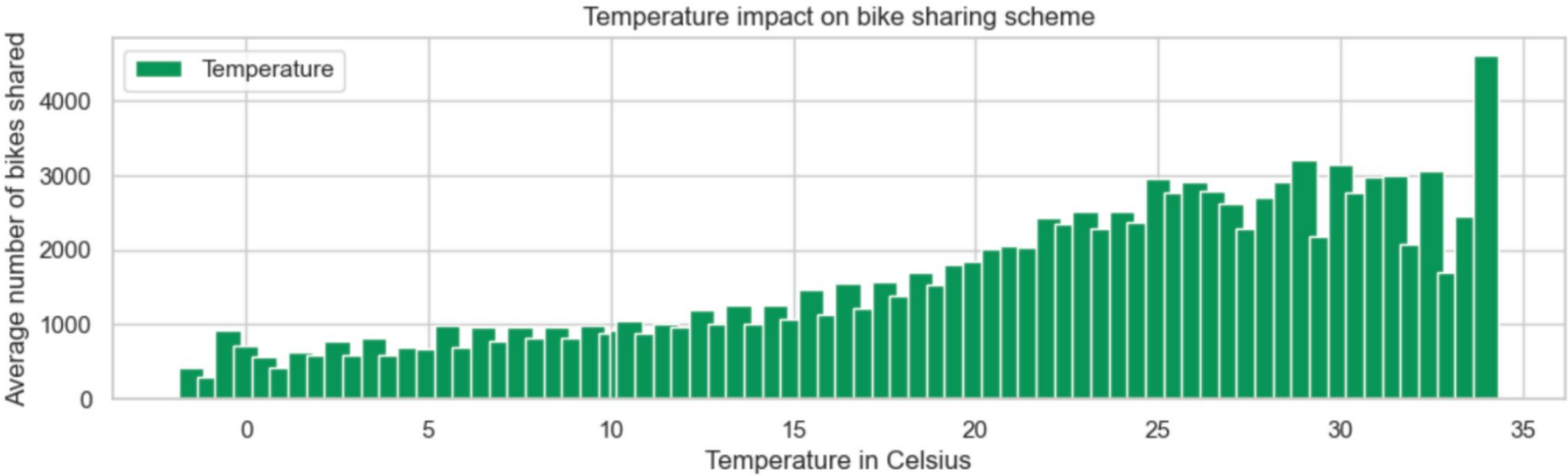

Bike Sharing Trends and seasonality
Number of Bikes Shared
8000
7000
6000
5000
4000
3000
2000
1000
0
2015-01
2015-04
2015-07
2015-10
2016-01
2016-04
2016-07
2016-10
2017-01
Time

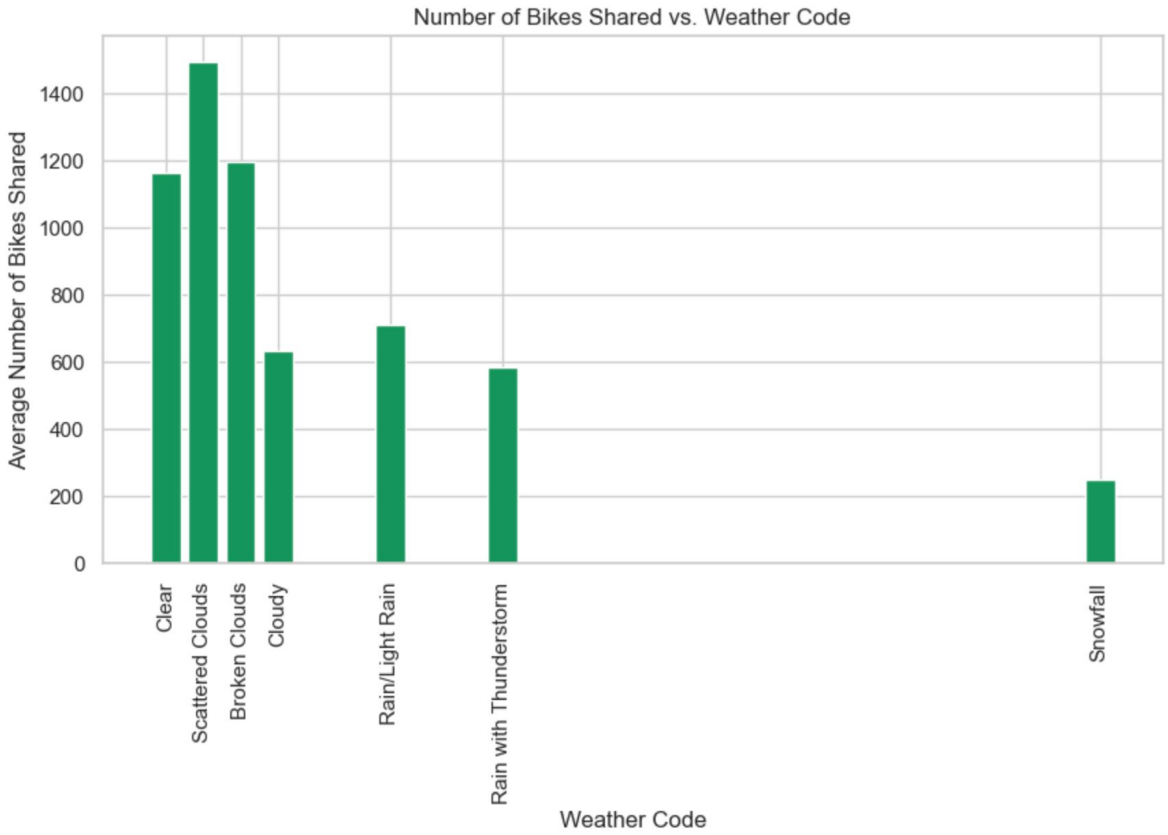

Average Monthly Bicycle count over time
1600
1400
1200
1000
800
600
400
200
0
2015-01
2015-04
2015-07
2015-10
2016-01
2016-04
2016-07
2016-10
2017-01


Number of Bikes Shared during the days of the Week
Number of Bikes Shared
30000
25000
20000
15000
10000
5000
0
Monday
Tuesday
Wednedsay
Thursday
Friday
Saturday
Sunday
Day of the Week


Number of Bikes Shared over time on workdays)
Number of Bikes Shared
4000
3000
2000
1000
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hour of Day


Number of Bikes Shared over time (Hourly)
Number of Bikes Shared
2000
1500
1000
500
0
Holidays
Weekends
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Hour of Day


Temperature impact on bike sharing scheme
Temperature
Average number of bikes shared
0
1000
2000
3000
4000
0
5
10
15
20
25
30
35
Temperature in Celsius


Number of Bikes Shared vs. Weather Code
Average Number of Bikes Shared
0
200
400
600
800
1000
1200
1400
Clear
Scattered Clouds
Broken Clouds
Cloudy
Rain/Light Rain
Rain with Thunderstorm
Snowfall
Weather Code

02

Forecast

Demand prediction model

1. Time series

Time series analysis is a specific way of analyzing a sequence of data points collected **over an interval of time**. It can show how variables change over time.

It requires a **large number of data points** to ensure:

- Consistency,
- Trends or patterns discovered are not outliers and can account for seasonality.

Models of time series analysis include classification, descriptive analysis, explanative analysis, exploratory analysis, forecasting...

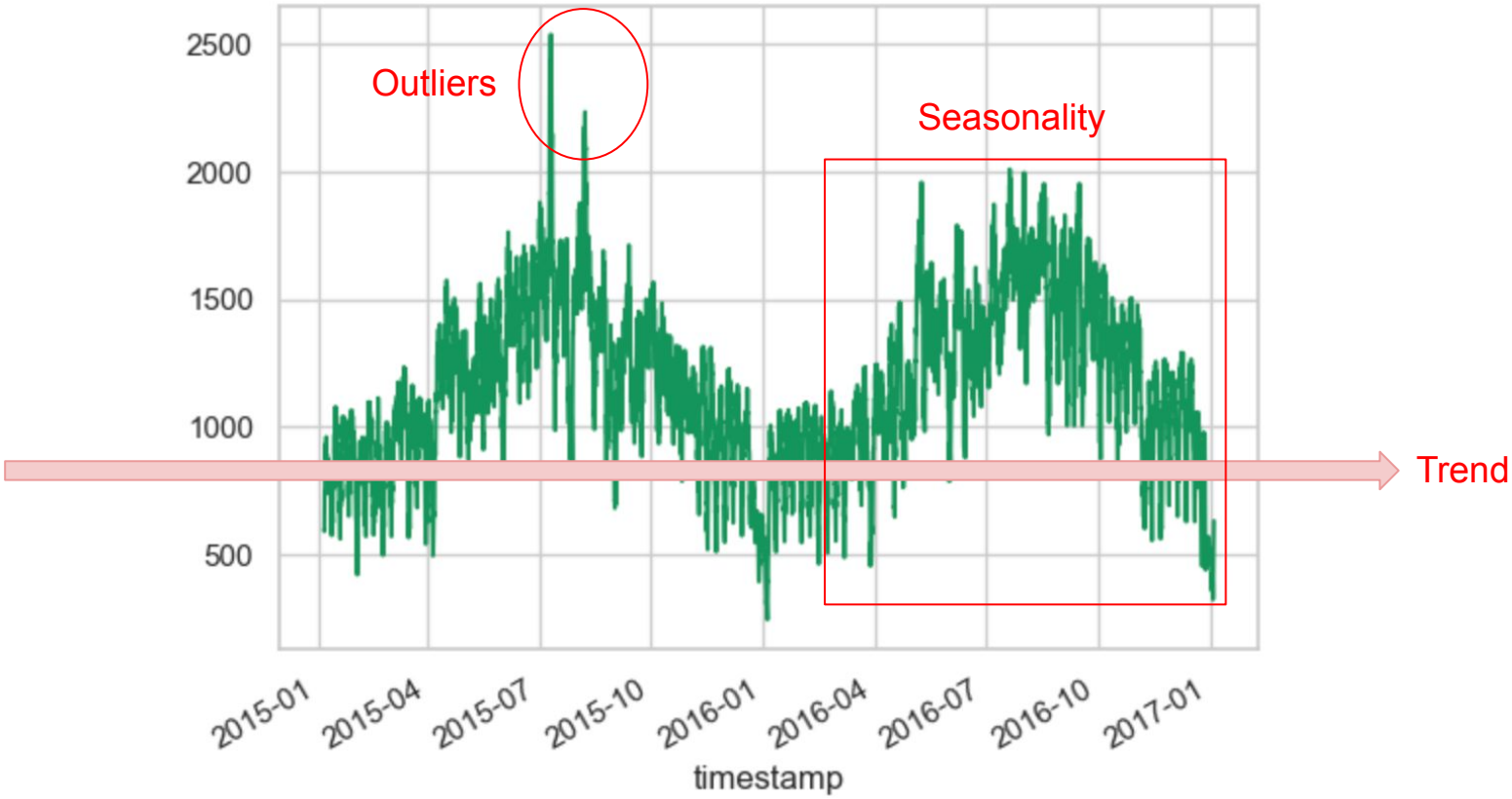
Main assumption: time series presents stationary data meaning that its statistical properties (mean, variance)remain constant across time.

2. Time series

In time series data, there are several types of patterns that can occur:

- Trend
- Seasonality
- Cycle
- Irregularity
- Autocorrelation
- Outliers
- Noise

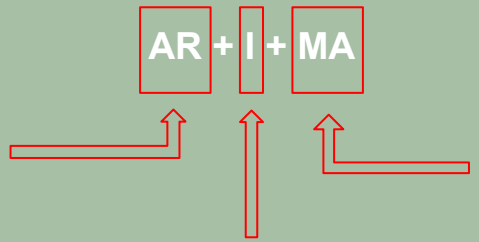
By identifying these patterns in time series data, analysts can better understand the underlying structure and make more accurate forecasts.


Outliers
Seasonality
Trend
2500
2000
1500
1000
500
2015-01
2015-04
2015-07
2015-10
2016-01
2016-04
2016-07
2016-10
2017-01
timestamp

3. Modelling with ARIMA

ARIMA is a general class of statistical models for time series analysis forecasting that uses time series past values or/and forecast errors to **predict future values.**

Three main components/parameters (non-negative integer values):



Auto-Regressive

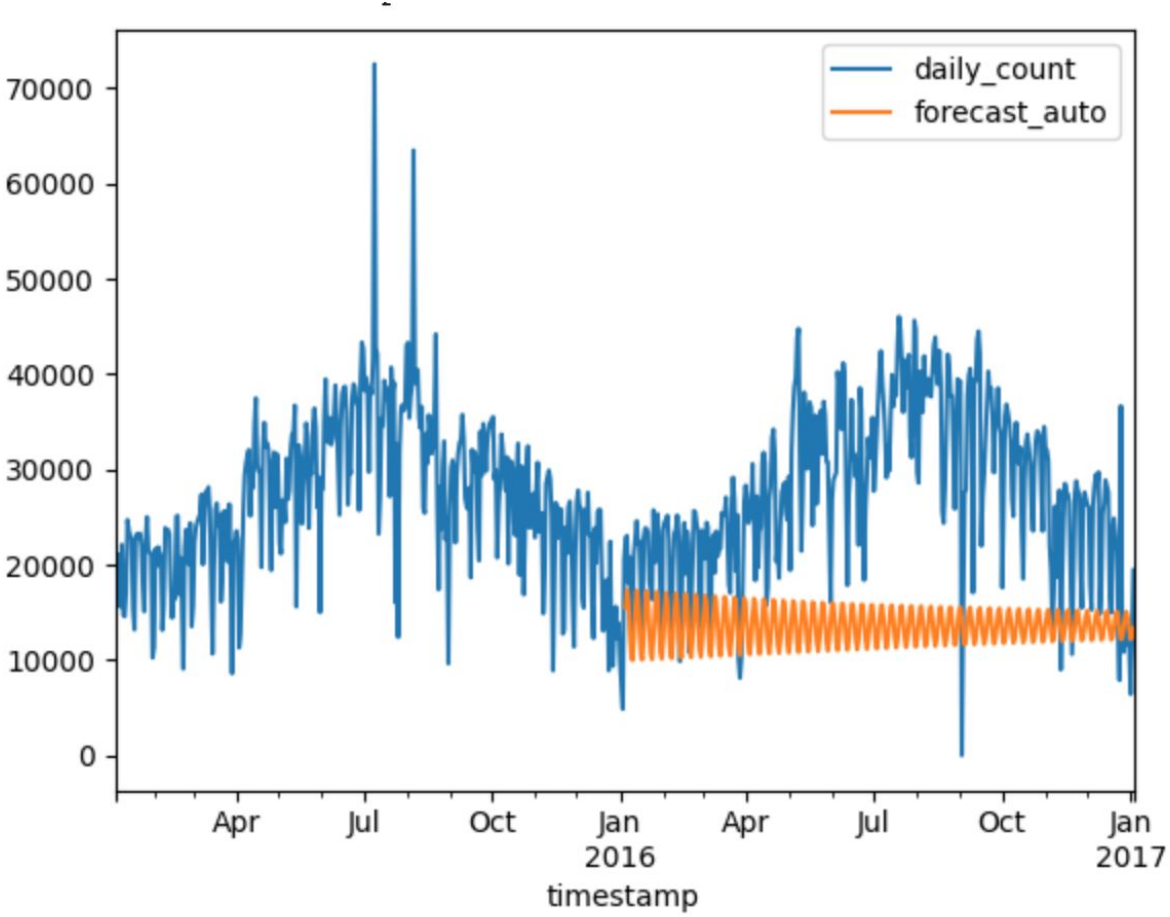
p = the number of past values included in the AR Model.

Integrated

d = the number of times the time series is differenced.

Moving Average

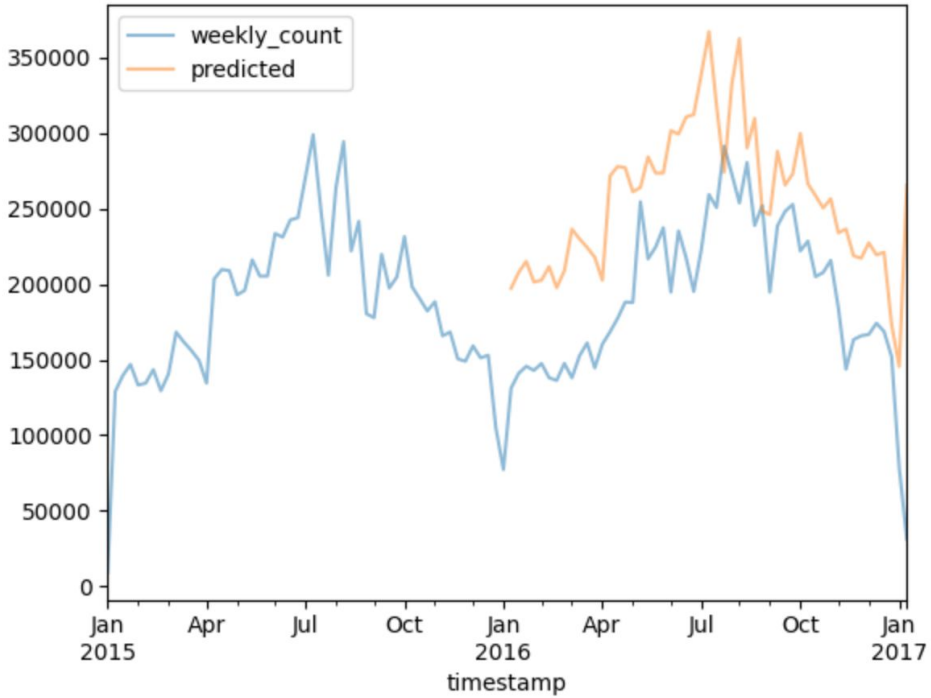
q = the number of past forecast errors included in the MA model.


daily_count
forecast_auto
70000
60000
50000
40000
30000
20000
10000
0
Apr
Jul
Oct
Jan
2016
Apr
Jul
Oct
Jan
2017
timestamp

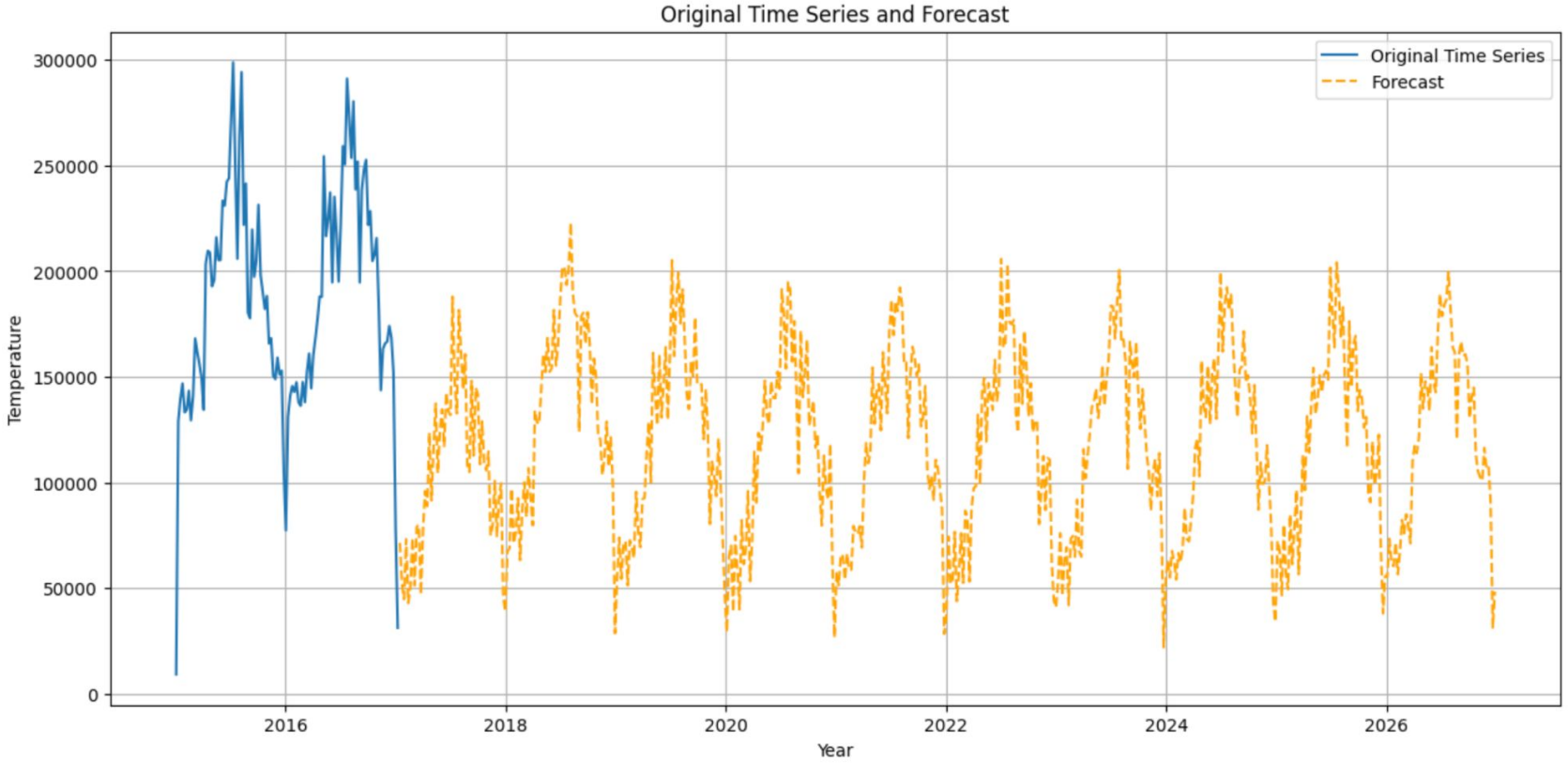
What am I missing?

What am I missing?

Seasonality.




```
model1 = SARIMAX(train["weekly_count"],
                 order = (3, 1, 3),
                 seasonal_order =(2, 1, 1, 52))
```


Original Time Series and Forecast
Original Time Series
Forecast
Temperature
Year
300000
250000
200000
150000
100000
50000
0
2016
2018
2020
2022
2024
2026

03

BikePoints

London public transport Open API

Improvements

- Analysis: add outlier data points, main flows (is the lack of bikes in certain location making users move differently);

- Demand prediction: extend data points for better predictions;

- API Connection : more frequent data updates.

Impacts of similar studies

Bike sharing has a long list of socio-economic benefits:

- Accessibility (many docking stations, and 24-7;
- Health (active mobility)
- Economic (increasing foot traffic - benefiting shops)
- Environmental

But for this it needs a vast and strong service that understands the needs of its users.

The future of urban mobility is shared.

Thank you.

Do you have any questions?

Esther Velasco
Data Analytics | Final project
December 2023