Predicting Bikesharing Demand

Group Nyon

2022-05-21

1 Introduction

This project focuses on estimating demand at each of the 13 bike-sharing stations in the city of Nyon, Switzerland. The stations are Changins, Débarcadère, Gare Sud, Piscine du Cossy, Hôpital, Petit Perdtemps, Hostel, Place de Savoie, La Plage, Stade de Colovray, Triangle de l'Etraz, Gare Nord, and Château.

This map shows how the stations are spread out across Nyon. Apart from the two bike stations around Nyon Gare, most stations are spread out.

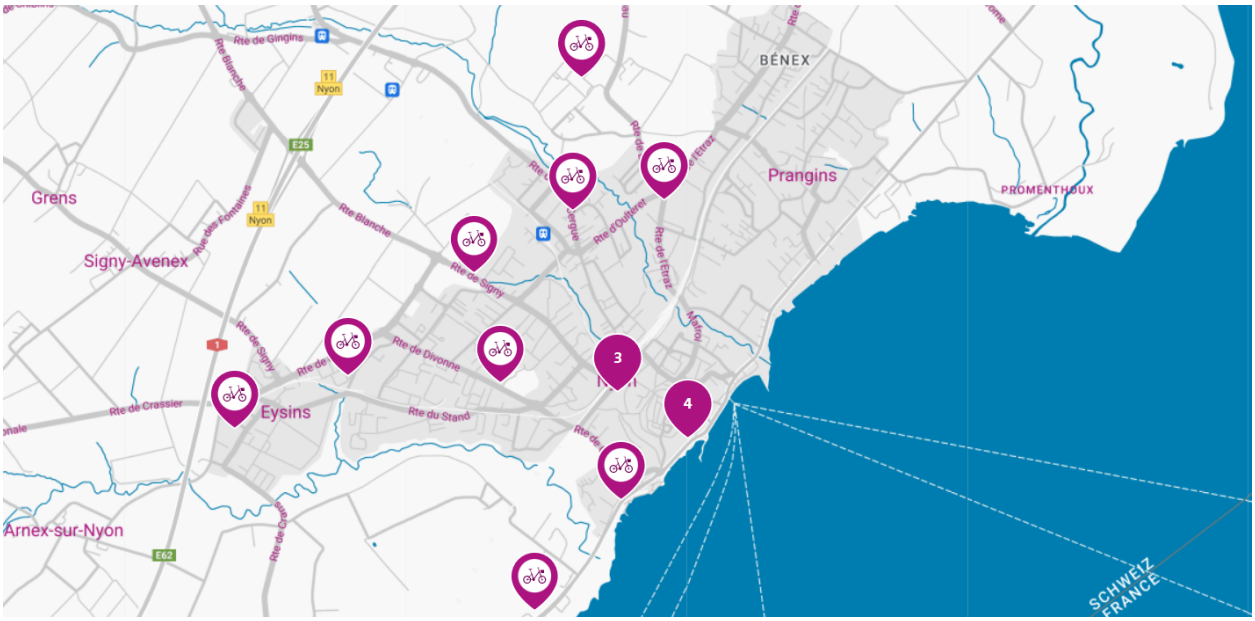


Figure 1: map

2 Exploratory Data Analysis and Data Wrangling / Cleaning

2.1 Data Wrangling

The dataset had several missing values. While most of the data was at intervals of 10 minutes, there were several cases where data was at half-hourly intervals. In order to deal with the missing data, graphs were created that showed how much the data varied across hours. The analysis revealed that the number of bikes and e-bikes at stations does not differ that much from hour to hour, therefore it is unlikely that it differs within 30 minute intervals either. The gaps were hence filled with the last previous available value. Capacity was also removed to save space as it is a constant value and does not affect the forecast. Dates were converted

into R's POSIXct format for ease of use. `TotalBikes` was created as a sum of `Bike` and `E-Bike`. After this, the dataset was divided into training and test sets to test accuracy. Data prior to May 11th was used as the training set and the remaining two weeks data was used as a test set.

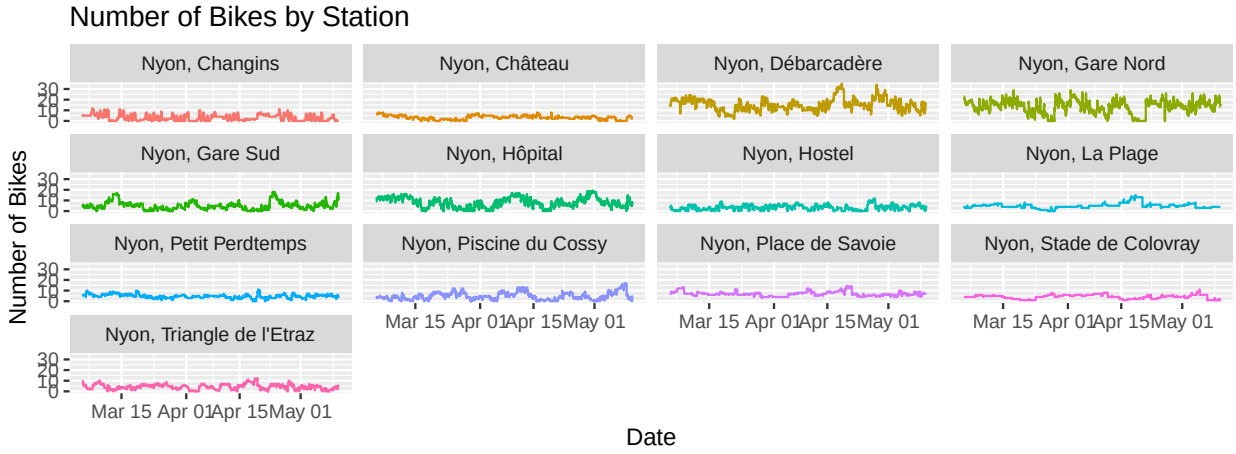


Figure 2: Variable demand across each station

Figure 2 with all its plots shows the number of bikes at each station in Nyon. There is no clear correlation between the graphs and their locations are far enough that the group decided against them. As each graph exhibits variable seasonality in terms of time of day and week. Further decomposition of seasonality to motivate this choice can be found in Fig 1.1 in the Appendix.

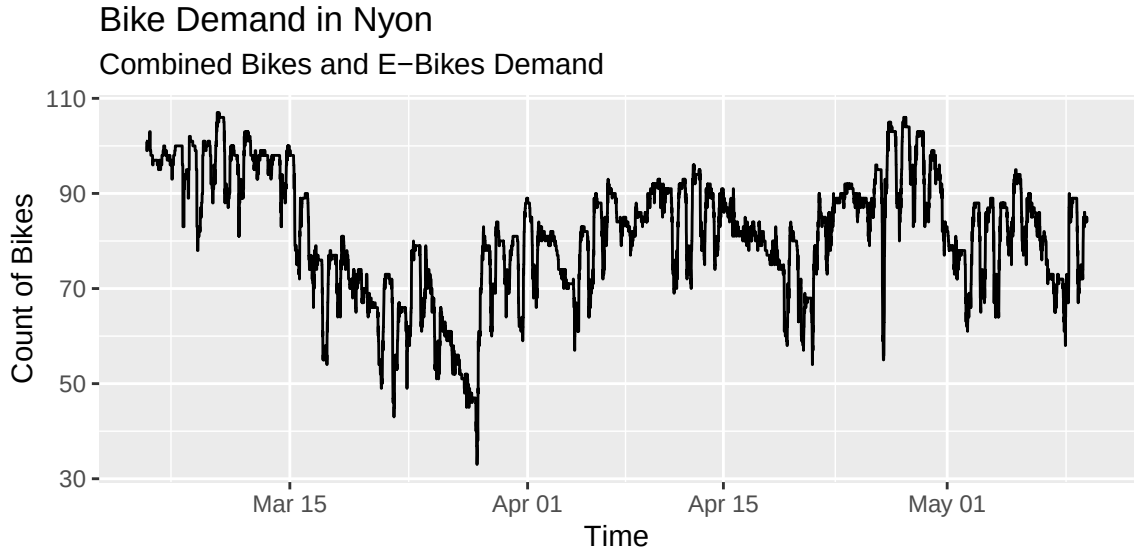
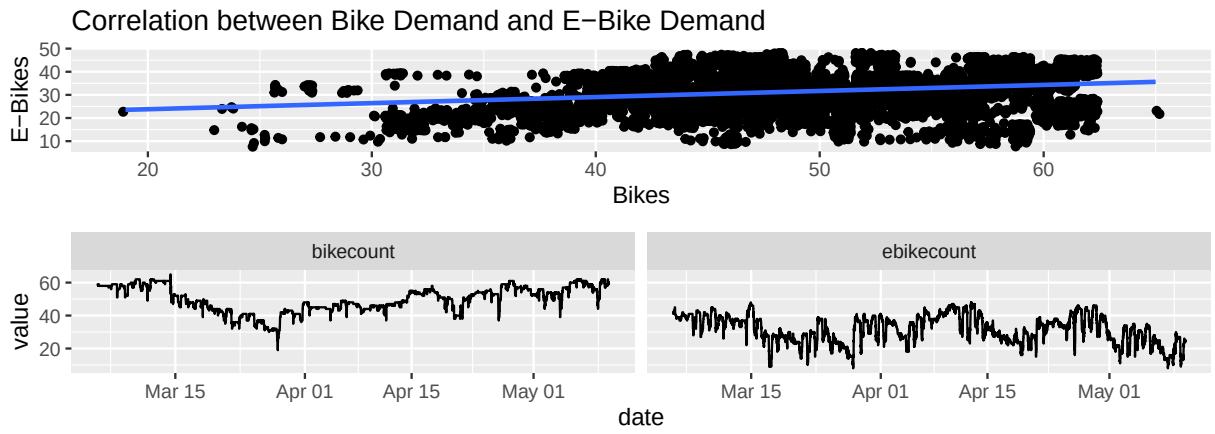


Figure 3: Bike and E-Bike Demand

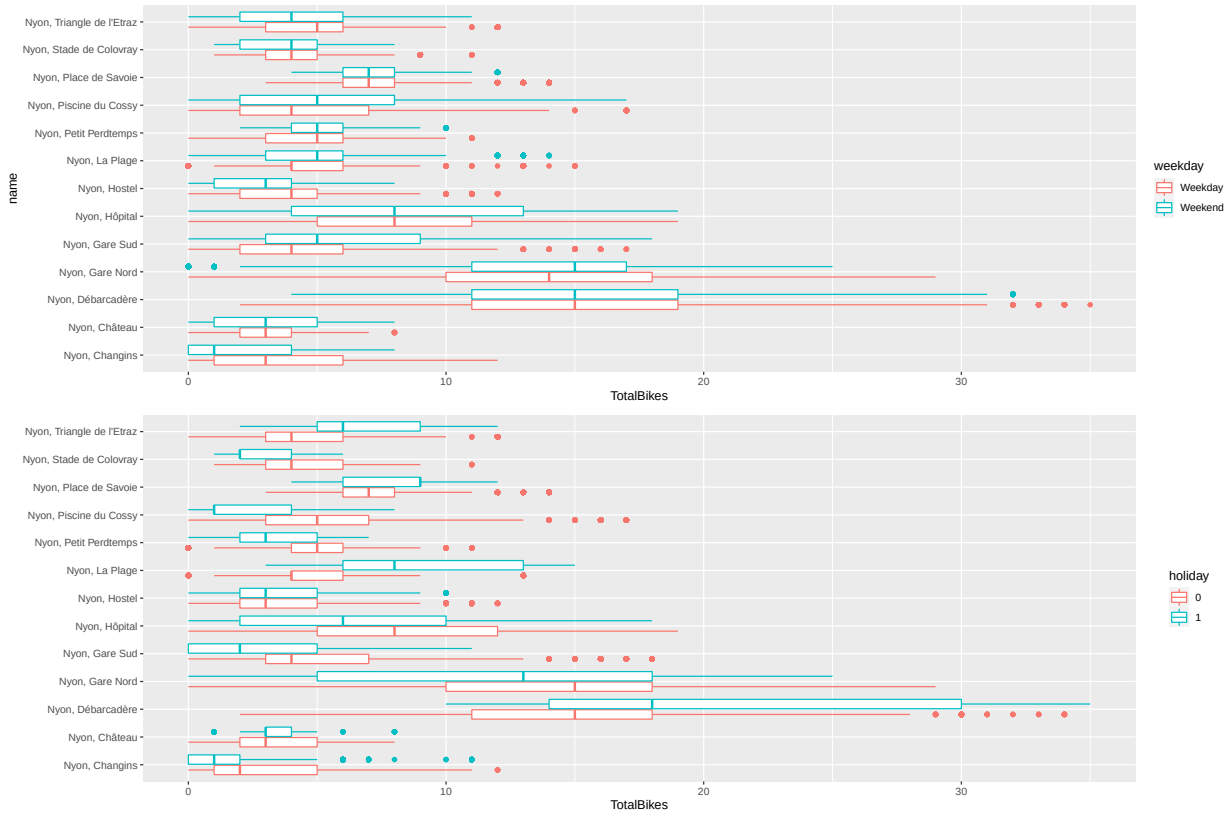
Figure 3 shows that the total bike demand is erratic. There is a clear drop in the last week of March, possibly owing due to the snowy conditions in Western Switzerland.

As part of our data understanding process, we wanted to see whether there was a relationship or even a difference in demand between the normal bikes and the e-bikes. Therefore, we have calculated the correlation between them.



After studying the correlation, we could see a slight correlation between bikes and e-bikes but there wasn't a huge difference between the demand on both so we have decided to combine them and wirk on "Total Bikes" instead. Intuitively, bike demand depends on a number of factors such as time of day, the day of the week, public holidays and the weather. The dataset already contains information that can be used to create new features such as `timeofday`, `weekday` and `holidays`. Data regarding holidays was extracted from the commune's website.

The boxplots show how `TotalBikes` differs at each station depending on whether it's a weekend or a weekday and whether the day in question is a public holiday (0 for No and 1 for Yes). In this time frame, the most relevant public holidays were: Easter and Labour Day.



There doesn't seem to be much variation in the total number of bikes available when it comes to weekdays vs weekends. However, there is a clear difference between holidays and non-holidays. Areas like La Plage have

no demand on holidays such as Easter Monday and the number of bikes available at the station is very high.

Next, we add the weather data and see if it has an impact on the available bikes at each station. Weather data was accessed using `WorldWeatherOnline` API. WWO provides a number of metrics such as temperature, precipitation, visibility among many others. The assumption is that better weather leads to more bikes in demand.

Let's look at that assumption in figure 4 using a temperature colored graph for one station `Nyon, Gare Nord`:

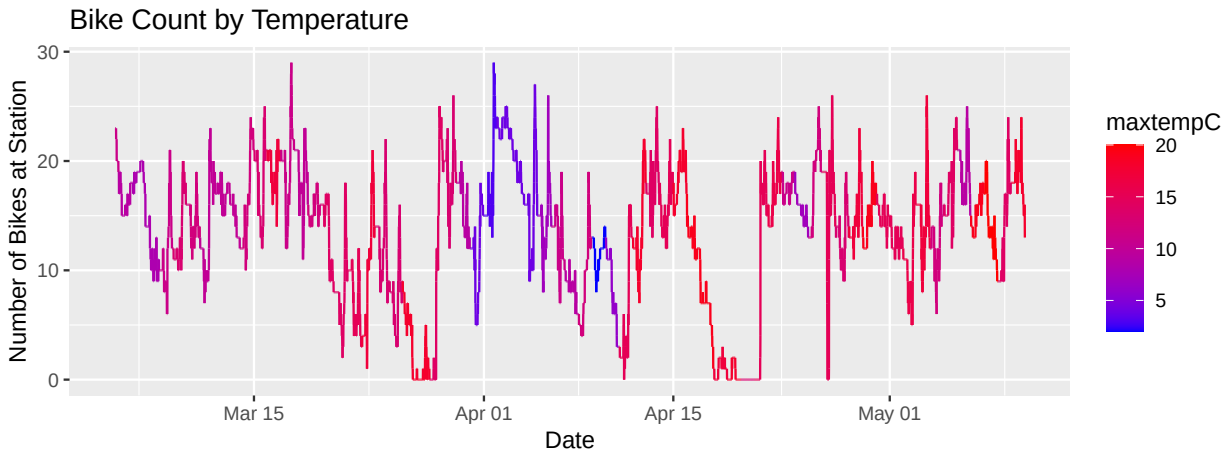


Figure 4: Correlation between temperature and number of bikes at station

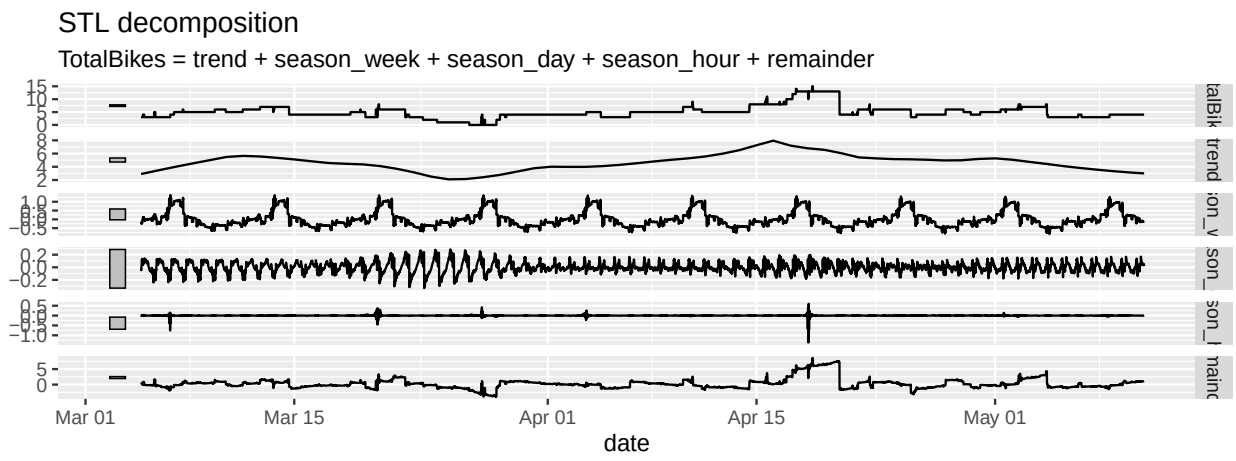
We can see that there is some correlation between weather and the number of bikes. As the temperature is higher, there are fewer bikes available at a station, as the temperature is lower. There are more bikes available. Therefore, we will be accounting for the temperature in our forecast.

3 Forecast

For our forecast, we have tried several forecasting models to get a closest forecast. After fitting several models, we have chosen TSLM (Time Series Linear Model) to fit the forecast as it provided best results along time.

The table below shows the first 6 rows of the forecast that we have done for May 25, 2022. You can refer to the .csv file for all the forecasts.

4 Appendix


STL decomposition
TotalBikes = trend + season_week + season_day + season_hour + remainder
Mar 01
Mar 15
Apr 01
Apr 15
May 01
date