
# Data Analytics

## Bicycles use in Paris
## State of the art and future

Damian Borsiak

December, 2022

## Table of content

# Introduction

As big metropolises around the world are becoming more active in the field of sustainability and are becoming more ecologically oriented, Paris as one of the most influential capitals is trying to lead the way. Following an initiative announced by the Mayor of the city in 2021, **Paris plans to be completely cyclable by 2026 (source)** - and aims to add another 130 kilometers of bike-safe pathways between 2021-2026. Together with the announced 130 '000 parking spots created by planned removal of 50% of the city car spaces, it seems nothing can stop this change.

Following recent policy changes in Paris, aimed to encourage conversion from cars to alternative forms of transport, I have decided to create an analysis of historic data between 2018 and 2021, and try to predict future use of bicycles.

What is especially interesting to me is that the time frame falls partially into Covid-19 global pandemic, which had a large impact on every possible sphere of our lives, therefore I feel it's appropriate to expect to see some interesting details (taking into consideration lockdowns and government restrictions).

The City of Paris has been deploying permanent **bicycle counters** for the last several years to assess the development of cycling. I will be using them to:

- Provide insight into bicycles traffic in Paris
- Visualize map of counters / data collection points
- Decide which Paris District has the highest bicycle traffic
- See the influence of Covid-19 pandemy on the bicycle use
- Forecast future traffic

**IMPORTANT !**

- A counting site can be equipped with a counter in the case of a one-way cycle path or two counters in the case of a two-way cycle path.
- Counting sites and counters are located on cycle paths and in certain bus lanes open to bicycles. Other vehicles (e.g. scooters, etc.) are not counted.
- The number of counters changes over time and some meters may be deactivated for work or occasionally suffer a breakdown.

## Data and data sources

**Main data source:**

https://opendata.paris.fr/

Website was created by an initiative of the City of Paris. You can find there datasets published by the City's services and its partners under the ODbL licence .

**Data source for my Project:**
https://opendata.paris.fr/explore/dataset/comptage-velo-historique-donnees-compteurs/information/

**Dataset ID:**
comptage-velo-historique-donnees-compteurs

**Subjects:**
Mobility and Public Space

**Key words**:
bike, bike count, bike path, DVD, IoT, bike lane

**Licence:**
Open Database License (ODbL)

**Producer:** Department of Roads and Transport - City of Paris
**Territory:** Paris
**Time zone:** Europe/Paris
**Language:** French

**Last processing:**
April 25, 2022 12:28 (metadata)
April 30, 2020 12:23 (data created and posted)

## Data collection

Data were downloaded separately for each year: 2018,2019,2020,2021
**Format:** compressed (zip) .csv files

**2018**
https://opendata.paris.fr/api/datasets/1.0/comptage-velo-historique-donnees-compteurs/attachments/2018_comptage_velo_donnees_compteurs_csv_zip/

**2019**
https://opendata.paris.fr/api/datasets/1.0/comptage-velo-historique-donnees-compteurs/attachments/2019_comptage_velo_donnees_compteurs_csv_zip/

**2020**
https://opendata.paris.fr/api/datasets/1.0/comptage-velo-historique-donnees-compteurs/attachments/2020_comptage_velo_donnees_compteurs_csv_zip/

**2021**
https://opendata.paris.fr/api/datasets/1.0/comptage-velo-historique-donnees-compteurs/attachments/2021_comptage_velo_donnees_compteurs_csv_zip/

| | 2018 | 2019 | 2020 | 2021 | |
|---|---|---|---|---|---|
| **Number of Rows:** | 157,825 | 436,729 | 2,314,738 | 5,851,680 | **Total Rows** |
| **Number of Columns** | 9 | 9 | 7 | 6 | 8,760,972 |

---

**Starting Features (max columns):**

*'Identifiant du compteur'* - ID / Counter identification number
*'Nom du compteur'* - Name of the counter
*'Identifiant du site de comptage'* - Another ID / Counter identification number
*'Nom du site de comptage'* - Counting site name
*'Comptage horaire'* - Number of bike rides counter
*'Date et heure de comptage'* - Date and time of measurement
*'Date d'installation du site de comptage'* - Counter installation date
*'Lien vers photo du site de comptage'* - Link to photo of counting site
*'Coordonnées géographiques'* - Geographical coordinates

## Data cleaning

**To achieve my goals, I will be using below Python Libraries:**

| Python Libraries | |
|---|---|
| **#Basic calculations and dataframe creation** | **#Plots** |
| import numpy as np | import matplotlib.pyplot as plt |
| import pandas as pd | import seaborn as sns |
| **#Operations on Date** | %matplotlib inline |
| from datetime import datetime | **#Forecasting** |
| from datetime import timezone | from prophet import Prophet |
| **#Visualisation / Maps** | from prophet.plot import plot_plotly |
| import folium | import plotly.offline as py |
| from folium import plugins | py.init_notebook_mode() |
| import seaborn as sns | from prophet.plot import add_changepoints_to_plot |
| **#Clustering** | |
| from sklearn.cluster import KMeans, DBSCAN | |

Data cleaning was performed using Python in Visual Studio Code

1. Import of 4 files with data into Python

```
#We import all datasets (for years 2018,2019,2020,2021)

data2018 = pd.read_csv(r'C:\Users\borys\IronHack\Final Project\Data Cycles in Paris\2018_comptage-velo-donnees-
data2019 = pd.read_csv(r'C:\Users\borys\IronHack\Final Project\Data Cycles in Paris\2019_comptage-velo-donnees-
data2020 = pd.read_csv(r'C:\Users\borys\IronHack\Final Project\Data Cycles in Paris\2020_comptage-velo-donnees-
data2021 = pd.read_csv(r'C:\Users\borys\IronHack\Final Project\Data Cycles in Paris\2021-comptage-velo-donnees-
```

2. Identify data Shapes - number of rows and columns

```
#We might want to combine all datasets into 1 dataframe
#Lets check if they are of the same shape and what are their sizes?

print("Shape of DataFrames:\n")
print("Number of rows and columns for year 2018:",data2018.shape)
print("Number of rows and columns for year 2019:",data2019.shape)
print("Number of rows and columns for year 2020:",data2020.shape)
print("Number of rows and columns for year 2021:",data2021.shape)

print("\nColumns on each DataFrames:\n")
print("List of columns for 2018:\n\n",data2018.columns)
print("\nList of columns for 2019:\n\n",data2019.columns)
print("\nList of columns for 2020:\n\n",data2020.columns)
print("\nList of columns for 2021:\n\n",data2021.columns)
```

3. Print data Headers to identify similarities between datasets
4. List names of columns

5. Drop columns that prevent us from successful merge of data sets

```
# Drop of columns we can't combine between all datasets
data2018.drop(['Identifiant du compteur',
               'Nom du site de comptage',
               "Date d'installation du site de comptage",
               "Lien vers photo du site de comptage"], axis=1, inplace=True)
data2019.drop(['Identifiant du compteur',
               'Nom du site de comptage',
               "Date d'installation du site de comptage",
               "Lien vers photo du site de comptage"], axis=1, inplace=True)
data2020.drop(["Date d'installation du point de comptage",
               "Lien vers photo du point de comptage"], axis=1, inplace=True)
data2021.drop(["Lien vers photo du point de comptage"], axis=1, inplace=True)
```

6. Switch order of columns where needed
7. Unification of columns names
8. Check if any data are missing / check for Null

```
#We check if any data are missing

print("\nList of blanks for 2018:\n\n",data2018.isnull().sum())
print("\nList of blanks for 2019:\n\n",data2019.isnull().sum())
print("\nList of blanks for 2020:\n\n",data2020.isnull().sum())
print("\nList of blanks for 2021:\n\n",data2021.isnull().sum())
```

```
List of blanks for 2020:


 Identifiant du site de comptage        0
Nom du compteur                     35132
Comptage horaire                        0
Date et heure de comptage               0
```

9. Drop identified 35’132 rows = 1.5% of data from dataset for 2020 (Null)

```
data2020 = data2020.dropna()
```

10. Addition of extra column with separate Year number for each dataset

```
data2018['Year'] = '2018'
data2019['Year'] = '2019'
data2020['Year'] = '2020'
data2021['Year'] = '2021'
```

11. Separation of Lat and Lon from column “Coordonnées géographiques”

```
data2018[['Lat', 'Lon']] = data2018['Coordonnées géographiques'].str.split(',', expand=True)
data2019[['Lat', 'Lon']] = data2019['Coordonnées géographiques'].str.split(',', expand=True)
data2020[['Lat', 'Lon']] = data2020['Coordonnées géographiques'].str.split(',', expand=True)
data2021[['Lat', 'Lon']] = data2021['Coordonnées géographiques'].str.split(',', expand=True)
```

12. Drop of “Coordonnées géographiques” column
13. Extraction of Hour,WeekDay,Month from column "Date et heure de comptage"

```
data2018["Date et heure de comptage"]=pd.to_datetime(data2018["Date et heure de comptage"], utc=True )
data2018['hour']=data2018["Date et heure de comptage"].dt.hour
data2018['weekday']=data2018["Date et heure de comptage"].dt.day_name()
data2018['month']=data2018["Date et heure de comptage"].dt.month
```

14. Drop of "Date et heure de comptage" column
15. Check of the unique values count in each dataset - no inconsistency detected
16. Final change of column names to make them shorter / translated into English
17. Merge of all datasets into 1

```
data = data2018.append([data2019,data2020,data2021])
```

**Comment for Data Cleaning -** data was relatively clean. As I worked with 4 datasets collected by the same organization, they had largely maintained consistency. I had to identify columns with the same information and prepare them for further merge. Biggest challenge was the size of the data (almost 9 '000' 000 rows), any reload or mistake was penalized by long loading time that extended my work.

```
data
```

✓ 0.5s

| | ID | Counter_Name | Bikes_Recorded | Year | Lat | Lon | hour | weekday | month |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 100047547 | 6 rue Julia Bartet NE-SO | 4.0 | 2018 | 48.82648 | 2.303149 | 0 | Thursday | 11 |
| 1 | 100047547 | 6 rue Julia Bartet NE-SO | 30.0 | 2018 | 48.82648 | 2.303149 | 21 | Thursday | 11 |
| 2 | 100047547 | 6 rue Julia Bartet NE-SO | 116.0 | 2018 | 48.82648 | 2.303149 | 16 | Friday | 11 |
| 3 | 100047547 | 6 rue Julia Bartet NE-SO | 0.0 | 2018 | 48.82648 | 2.303149 | 0 | Monday | 12 |
| 4 | 100047547 | 6 rue Julia Bartet NE-SO | 18.0 | 2018 | 48.82648 | 2.303149 | 10 | Monday | 12 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 5851675 | 100063175 | 20 Avenue de Clichy | 1.0 | 2021 | 48.88529 | 2.32666 | 22 | Friday | 12 |
| 5851676 | 100063175 | 20 Avenue de Clichy | 0.0 | 2021 | 48.88529 | 2.32666 | 23 | Friday | 12 |
| 5851677 | 100063175 | 20 Avenue de Clichy | 0.0 | 2021 | 48.88529 | 2.32666 | 23 | Friday | 12 |
| 5851678 | 100063175 | 20 Avenue de Clichy | 0.0 | 2021 | 48.88529 | 2.32666 | 23 | Friday | 12 |
| 5851679 | 100063175 | 20 Avenue de Clichy | 0.0 | 2021 | 48.88529 | 2.32666 | 23 | Friday | 12 |

8725840 rows × 9 columns

## Exploratory data analysis

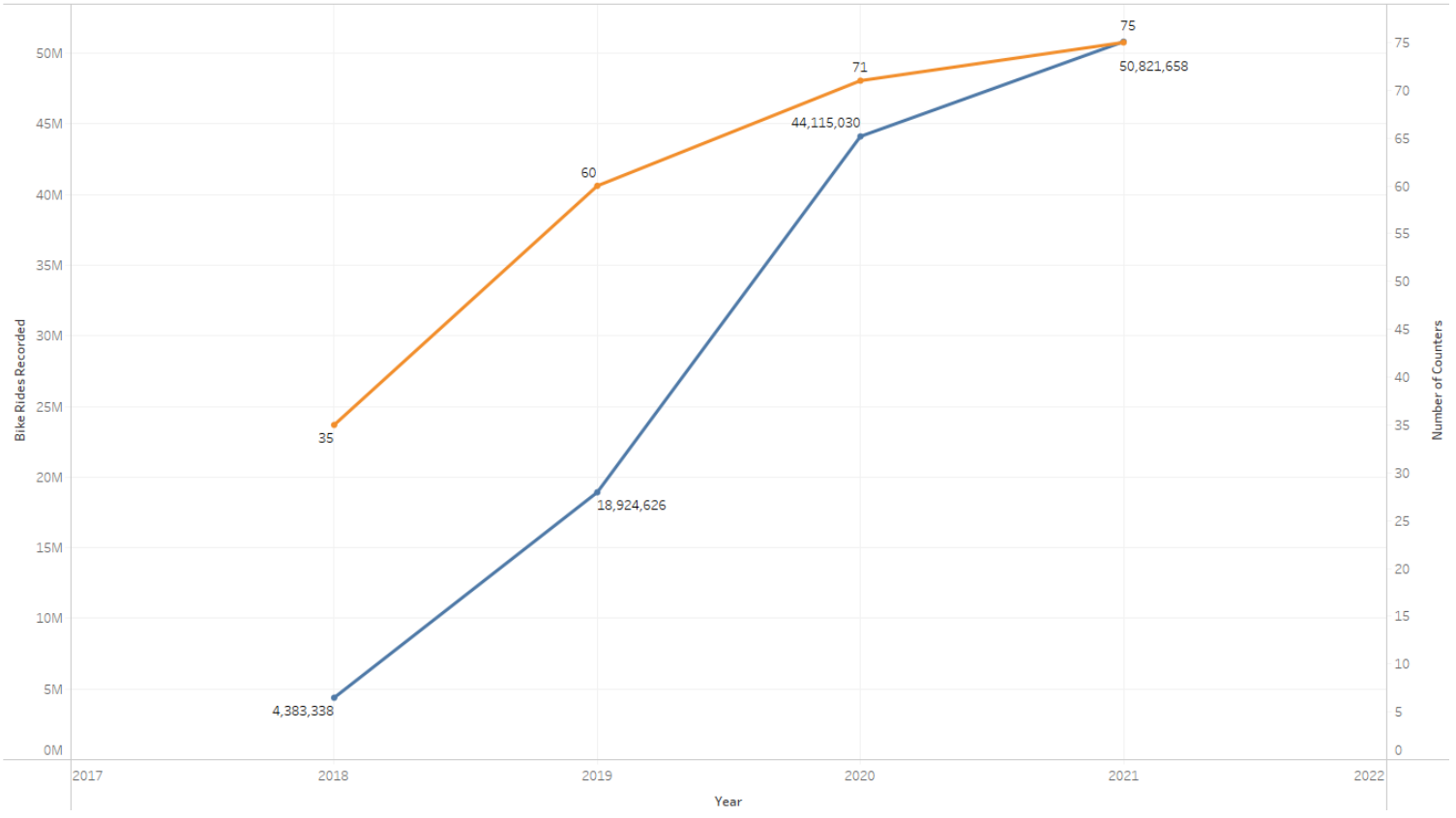
### Bike rides recorder per Year vs Number of Counters

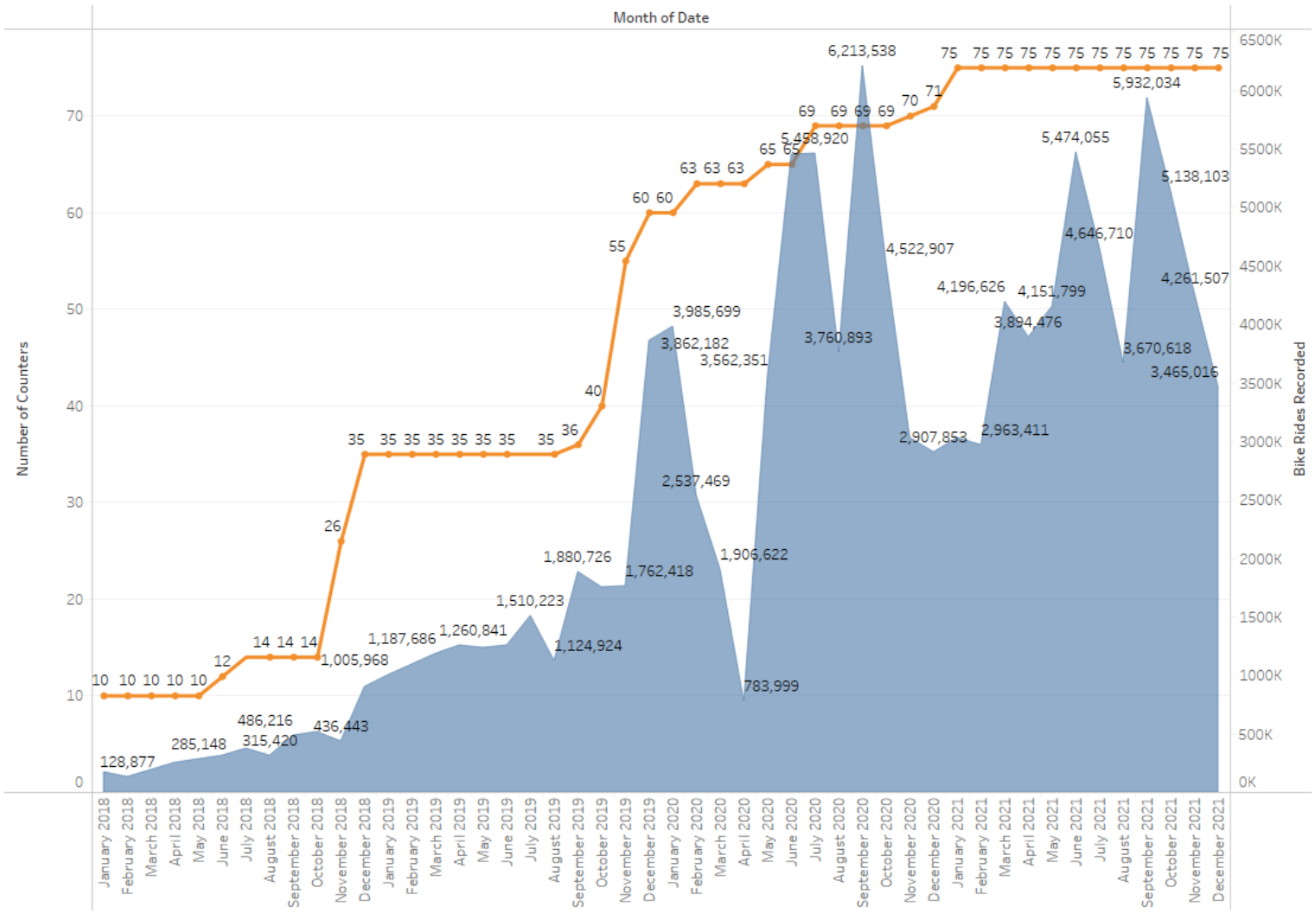
| Year of Date | Quarter of Date | Bikes Recorded | Number of Counters |
|---|---|---|---|
| 2018 | Q1 | 496,442 | 10 |
| | Q2 | 851,444 | 12 |
| | Q3 | 1,178,255 | 14 |
| | Q4 | 1,857,197 | 35 |
| 2019 | Q1 | 3,286,317 | 35 |
| | Q2 | 3,744,407 | 35 |
| | Q3 | 4,515,873 | 36 |
| | Q4 | 7,378,029 | 60 |
| 2020 | Q1 | 8,429,790 | 63 |
| | Q2 | 9,793,269 | 65 |
| | Q3 | 15,433,351 | 69 |
| | Q4 | 10,458,620 | 71 |
| 2021 | Q1 | 10,187,340 | 75 |
| | Q2 | 13,520,330 | 75 |
| | Q3 | 14,249,362 | 75 |
| | Q4 | 12,864,626 | 75 |
| Grand Total | | 118,244,652 | 75 |

Paris has been steadily expanding its number of Counters since 2018, starting with 10 machines in Q1 of 2018, to achieve a maximum number of 75 machines in the last quarter of 2021.

Higher number of machines means increased number of rides recorded. Starting with almost 500'000 trips in Q1 of 2018, Paris has experienced a total of 118 million bicycle rides by the end of 2021.

Bike rides recored per Year vs Number of Counters






Above graph presents a monthly change in the number of operating Counters and their measured bicycle rides. We see that the number of machines was increasing first by large quantities at one time (increasing at the end of 2018 from 14 to 35 and at the end of 2019 from 35 to 60 machines). After this period additional machines were gradually added to reach their peak at 75.

We can clearly observe a correlation between the number of Counters and the volume of rides.

We can also notice an unusual drop recorded in March / April 2020 - as we remember on 16 March 2020, President Macron has announced the first Covid-19 lockdown in France.

Again in October 2020 we have experienced decreased volume of bicycle users due to Overnight curfew (21:00 to 6:00) in Paris and suburbs (later in December 2020 extended from 18:00 to 6:00).

## Bike rides recorded per Hour vs Year / Quarter

| | Date | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2019 | | | | 2020 | | | | 2021 | | | |
| Hour | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 |
| 0 | 23,479 | 21,409 | 29,424 | 55,523 | 66,409 | 56,510 | 100,039 | 28,729 | 28,669 | 113,043 | 271,036 | 196,594 |
| 1 | 16,361 | 11,537 | 15,491 | 36,982 | 44,914 | 31,340 | 51,780 | 18,059 | 17,606 | 63,621 | 148,900 | 112,540 |
| 2 | 10,308 | 9,282 | 12,091 | 20,043 | 24,854 | 22,583 | 36,567 | 13,839 | 11,137 | 38,406 | 100,173 | 85,012 |
| 3 | 8,995 | 13,578 | 17,822 | 17,952 | 21,230 | 34,108 | 55,656 | 19,577 | 12,376 | 24,774 | 54,940 | 48,195 |
| 4 | 13,206 | 29,708 | 38,550 | 34,759 | 35,993 | 86,880 | 128,994 | 49,745 | 10,780 | 20,150 | 40,493 | 37,538 |
| 5 | 25,299 | 101,211 | 126,657 | 96,124 | 79,890 | 266,654 | 439,942 | 158,862 | 27,477 | 38,875 | 55,638 | 48,741 |
| 6 | 90,447 | 308,837 | 375,983 | 313,017 | 261,058 | 671,831 | 1,285,287 | 531,835 | 81,521 | 107,687 | 115,569 | 96,014 |
| 7 | 294,940 | 319,854 | 395,561 | 630,371 | 773,738 | 645,423 | 1,260,213 | 889,910 | 307,730 | 371,789 | 372,632 | 345,224 |
| 8 | 300,247 | 166,666 | 191,849 | 588,254 | 808,914 | 410,094 | 622,473 | 687,644 | 901,006 | 1,050,522 | 1,110,281 | 1,104,914 |
| 9 | 147,758 | 156,585 | 175,317 | 327,863 | 379,470 | 407,373 | 572,345 | 472,481 | 847,859 | 1,018,184 | 1,078,995 | 1,059,996 |
| 10 | 138,941 | 185,064 | 207,724 | 303,240 | 333,144 | 485,011 | 708,680 | 528,771 | 484,071 | 611,206 | 574,956 | 542,450 |
| 11 | 167,500 | 180,627 | 198,210 | 353,422 | 406,628 | 487,332 | 679,050 | 610,390 | 507,251 | 626,682 | 558,324 | 519,488 |
| 12 | 165,273 | 174,255 | 198,516 | 357,173 | 408,495 | 502,531 | 683,028 | 606,252 | 674,809 | 794,293 | 696,057 | 674,214 |
| 13 | 163,563 | 179,640 | 208,230 | 358,138 | 395,481 | 567,618 | 737,294 | 603,435 | 665,930 | 755,130 | 673,330 | 663,751 |
| 14 | 166,756 | 211,029 | 243,595 | 390,713 | 405,508 | 667,406 | 857,100 | 645,924 | 643,104 | 733,350 | 674,805 | 636,774 |
| 15 | 193,311 | 286,239 | 342,614 | 485,436 | 476,712 | 809,888 | 1,191,838 | 786,062 | 677,777 | 781,719 | 716,387 | 655,055 |
| 16 | 257,190 | 374,815 | 476,789 | 634,102 | 647,746 | 990,441 | 1,651,018 | 993,187 | 776,067 | 896,084 | 821,418 | 746,780 |
| 17 | 333,039 | 352,479 | 435,240 | 720,589 | 857,027 | 979,533 | 1,556,231 | 1,031,458 | 1,048,458 | 1,128,908 | 1,090,857 | 972,880 |
| 18 | 296,750 | 226,788 | 275,166 | 608,109 | 775,008 | 611,324 | 941,542 | 833,679 | 981,812 | 1,439,630 | 1,481,352 | 1,297,841 |
| 19 | 177,807 | 137,299 | 169,798 | 377,189 | 460,354 | 337,119 | 546,773 | 444,202 | 747,509 | 1,217,371 | 1,391,654 | 1,213,667 |
| 20 | 104,367 | 109,897 | 136,649 | 229,465 | 262,415 | 252,982 | 463,047 | 239,744 | 384,381 | 728,362 | 863,950 | 733,846 |
| 21 | 83,308 | 92,820 | 118,040 | 185,547 | 212,126 | 214,415 | 423,819 | 140,158 | 208,203 | 470,118 | 528,925 | 422,441 |
| 22 | 66,228 | 61,244 | 81,590 | 154,778 | 178,429 | 158,194 | 284,314 | 77,837 | 95,436 | 273,809 | 436,109 | 347,594 |
| 23 | 41,244 | 33,544 | 44,967 | 99,240 | 114,247 | 96,679 | 156,321 | 46,840 | 46,371 | 216,617 | 392,581 | 303,077 |

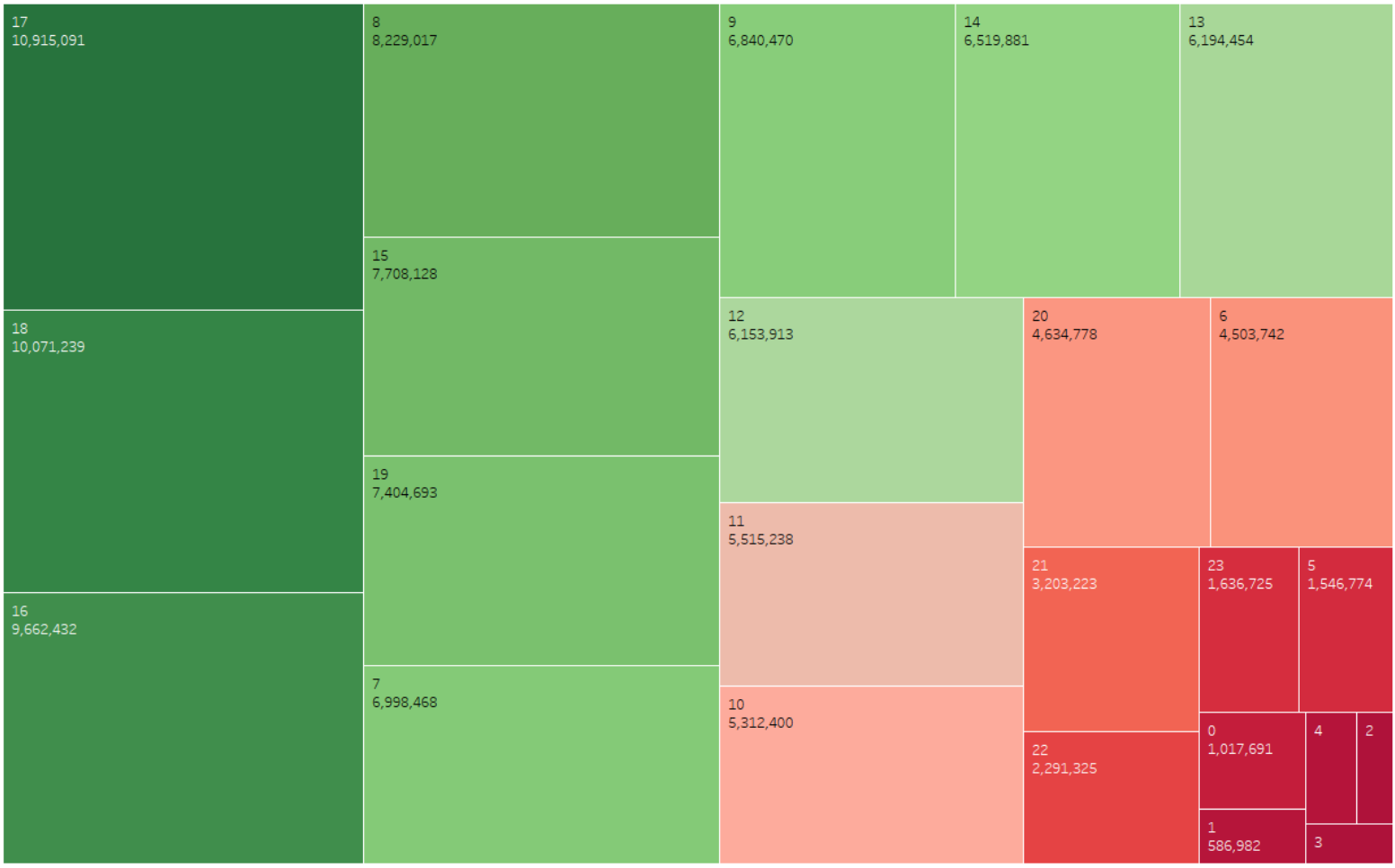
Following the above visualization, we can observe that the highest bicycle traffic occurs usually between 7:00 and 9:00 when people travel in the morning to work and between 16:00 and 19:00 when they are returning to their homes.

In 2021 we can notice a high number of green fields as the year benefited from the highest active number of measuring instruments across the dataset.

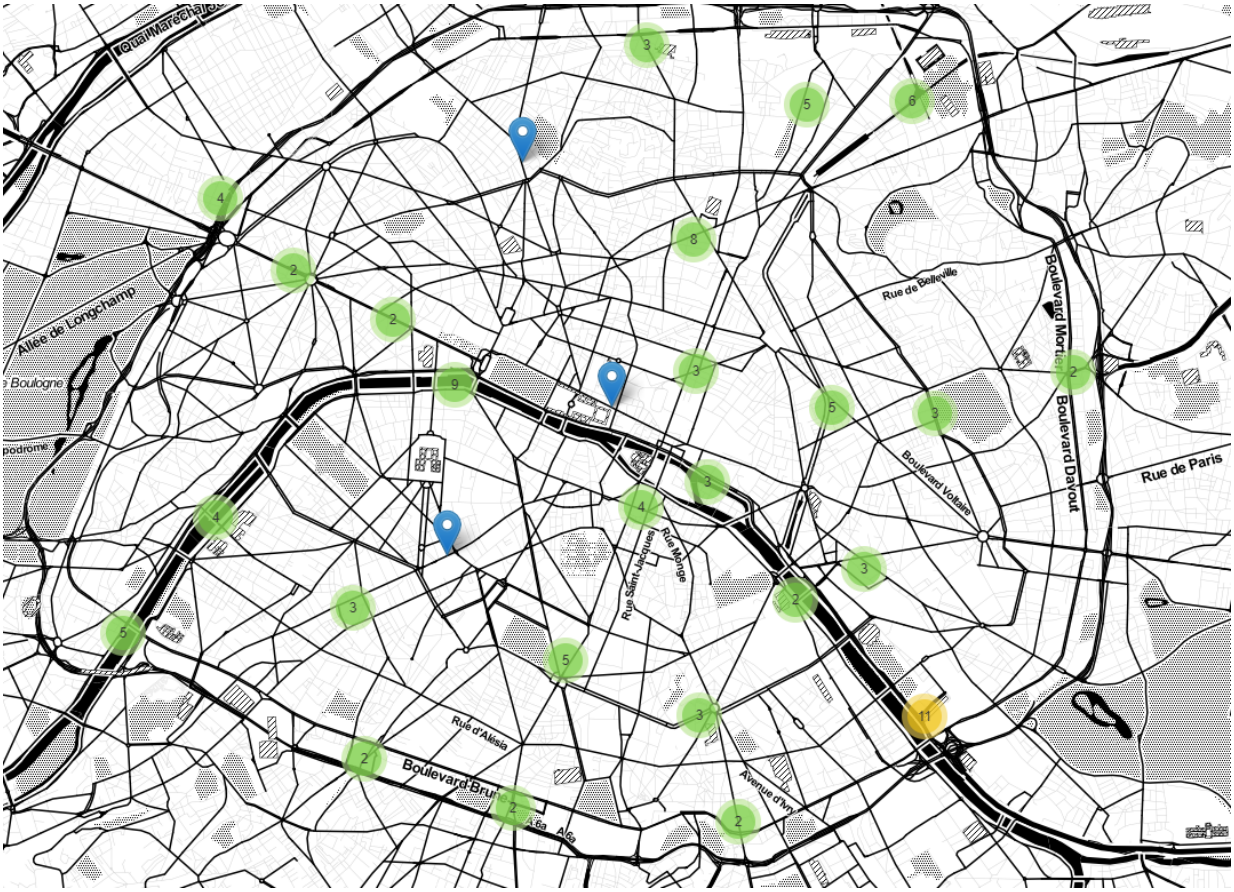
2020 has high levels of traffic - most probably due to the lockdown being lifted at the end of July of that year.

Data for 2018 are hidden as their volume of trips was very small, due to little number of counters.

Bikes traffic per Hour across all Years




Above TreeMap visualization confirms our earlier comment about rush hours in Paris. Number of rides in the night hours is minimal, during the day outside of rush hours, the number of travelers is somewhere in between.




Visualization of Counters location by point and heat map in the City of Paris

## % of total bicycles traffic in Paris by District

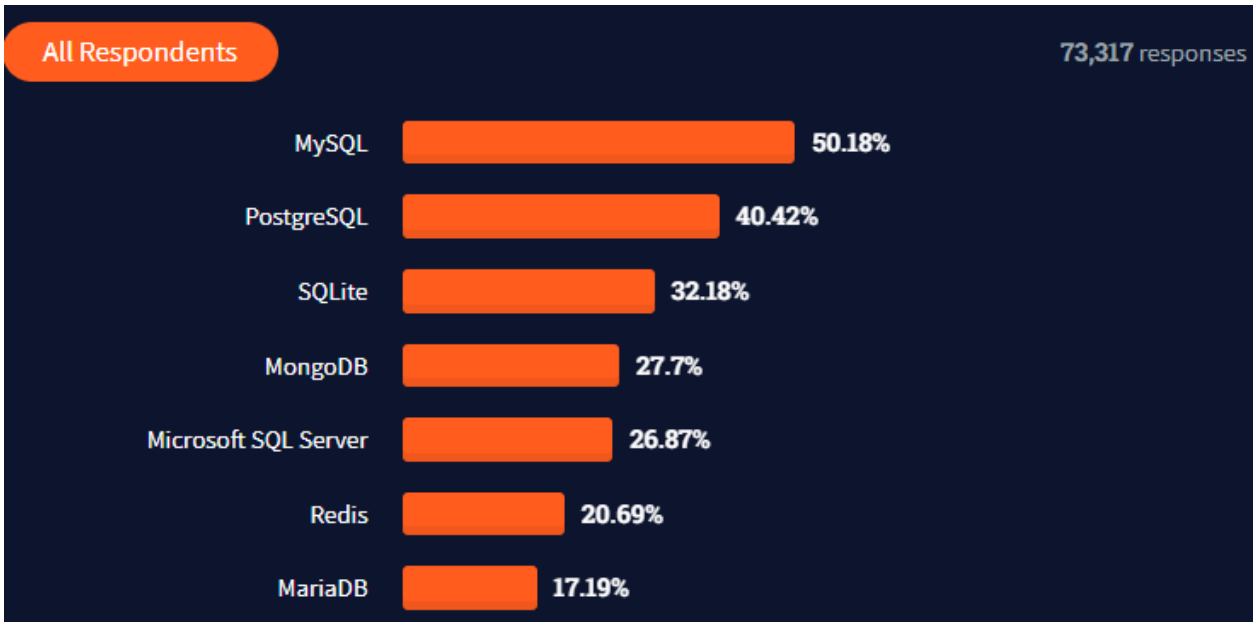
| District Full | |
|---|---|
| 11 - Popincourt | 14.08% |
| 10 - Entrepôt | 10.02% |
| 13 - Gobelins | 9.62% |
| 01 - Louvre | 7.64% |
| 15 - Vaugirard | 7.51% |
| 12 - Reuilly | 7.11% |
| 19 - Buttes-Chaumont | 6.83% |
| 07 - Palais-Bourbon | 6.71% |
| 14 - Observatoire | 5.63% |
| 02 - Bourse | 5.27% |
| 05 - Panthéon | 4.45% |
| 08 - Élysée | 3.99% |
| 03 - Temple | 2.40% |
| 18 - Butte-Montmartre | 2.23% |
| 09 - Opéra | 2.15% |
| 04 - Hôtel-de-Ville | 1.70% |
| 17 - Batignolles-Monceau | 1.07% |
| 20 - Ménilmontant | 0.96% |
| 16 - Passy | 0.65% |

Traffic per District of Paris


Map and Table presents Parisian districts with color codes based on their participation in the bicycle Traffic.

Highest traffic recorded was visible in 11, 10 and 13 Arrondissements of Paris.

## Database Comparison




Most popular database systems. Source: 2021 Developer Survey by StackOverflow

**I have selected MySQL Database** as its one of the most popular relational database systems in the industry and offer perfect capabilities for my project scale and complexity. As it is widely used and I am already familiar with it, it was the best choice.

**MySQL benefits**:

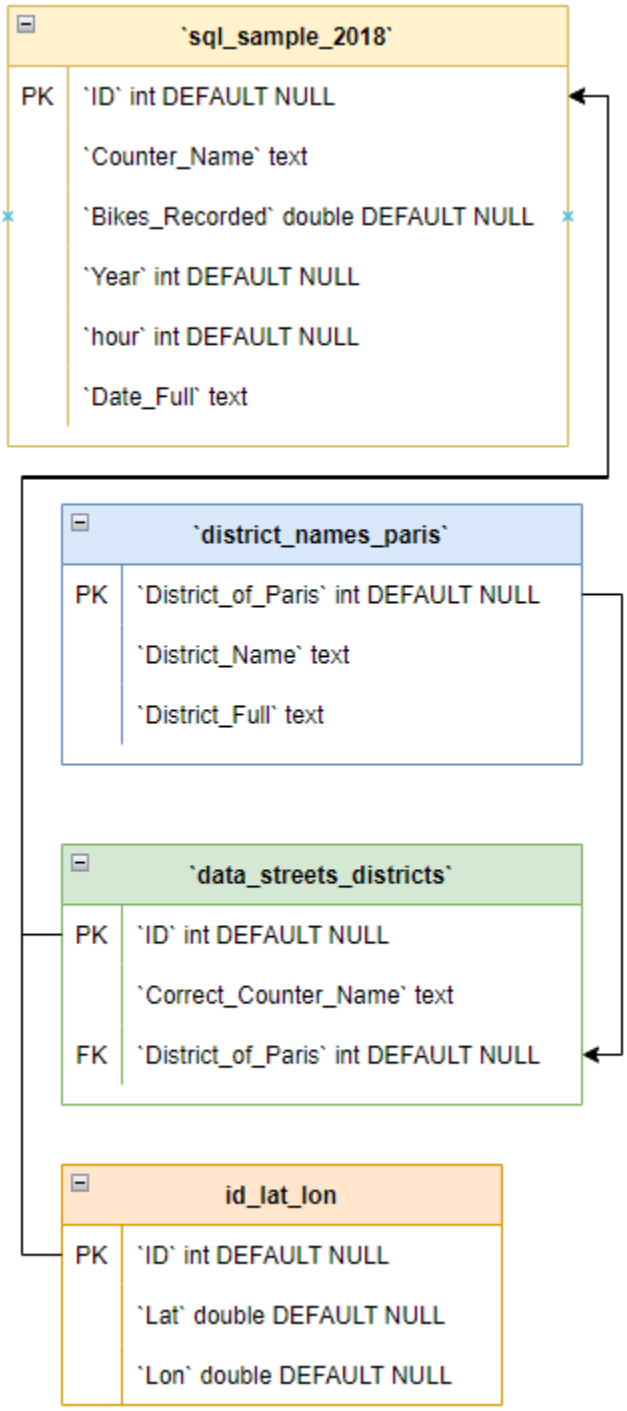
- Free installation
- Simple syntax and mild complexity
- Cloud compatibility

I was also considering PostgreSQL as a potential Database, but decided that it could be less tailored to my needs, simply by offering other - not needed in this project functionalities.

**PostgreSQL benefits:**

- Great scalability
- Support for custom data types
- Easily-integrated third-party tools
- Open-source and community-driven support

## Entities. ERD




```sql
create database if not exists Final_Project;
use Final_Project;

create table sql_sample_2018 (
ID int,
Counter_Name varchar(255),
Bikes_Recorded float,
Year int,
hour int,
Date_Full varchar(255)
);

create table data_streets_districts (
ID int,
Correct_Counter_Name varchar(255),
District_of_Paris int
);

create table data_names_parisA (
District_of_Paris int,
District_Name varchar(255),
district_Full varchar(255)
);

create table id_lat_lonA (
ID int,
Lat float,
Lon float
);

-- We are importing the data into tables
-- using Table Data Import Wizard
```

## Queries aimed to check if content of tables was correctly imported

```
select * from sql_sample_2018;
```

| ID | Counter_Name | Bikes_Recorded | Year | hour | Date_Full |
|---|---|---|---|---|---|
| 100042374 | Voie Georges Pompidou SO-NE | 10 | 2018 | 10 | 2018-01-01 |
| 100003097 | 105 Rue La Fayette E-O | 4 | 2018 | 1 | 2018-01-01 |
| 100041488 | 27 Boulevard Diderot E-O | 0 | 2018 | 4 | 2018-01-01 |

```
select * from data_streets_districts;
```

| ID | Correct_Counter_Name | District_of_Paris |
|---|---|---|
| 100003096 | 97 Avenue Denfert Rochereau SO-NE | 75014 |
| 100003097 | 105 Rue La Fayette E-O | 75010 |
| 100003098 | 106 Avenue Denfert Rochereau NE-SO | 75014 |

```
select * from district_names_paris;
```

| District_of_Paris | District_Name | District_Full |
|---|---|---|
| 75001 | Louvre | 01 - Louvre |
| 75002 | Bourse | 02 - Bourse |
| 75003 | Temple | 03 - Temple |

```
select * from id_lat_lon;
```

| ID | Lat | Lon |
|---|---|---|
| 100047547 | 48.82648 | 2.303149 |
| 100047546 | 48.8295233 | 2.38699 |
| 100047544 | 48.860852 | 2.372279 |

## Query Joining 2 tables
## - addition of geographical dimensions (Lat/Lon) to the main table

```
-- Left Join to add geographical dimensions to the main table
select sql_sample_2018.ID, sql_sample_2018.Counter_Name, sql_sample_2018.Bikes_Recorded,
        sql_sample_2018.Year, sql_sample_2018.hour, sql_sample_2018.Date_Full,
        id_lat_lon.Lat, id_lat_lon.Lon
from sql_sample_2018
left join id_lat_lon on sql_sample_2018.ID = id_lat_lon.ID;
```

| ID | Counter_Name | Bikes_Recorded | Year | hour | Date_Full | Lat | Lon |
|---|---|---|---|---|---|---|---|
| 100042374 | Voie Georges Pompidou SO-NE | 10 | 2018 | 10 | 2018-01-01 | 48.848399 | 2.275932 |
| 100003097 | 105 Rue La Fayette E-O | 4 | 2018 | 1 | 2018-01-01 | 48.877667 | 2.350556 |
| 100041488 | 27 Boulevard Diderot E-O | 0 | 2018 | 4 | 2018-01-01 | 48.846099 | 2.375456 |

**Query Joining 2 tables**
**- addition of corrected name of the street with Counter and District number**

```
select sql_sample_2018.ID, sql_sample_2018.Bikes_Recorded,
        sql_sample_2018.Date_Full, data_streets_districts.Correct_Counter_Name,
        data_streets_districts.District_of_Paris
from sql_sample_2018
left join data_streets_districts on sql_sample_2018.ID = data_streets_districts.ID;
```

| ID | Bikes_Recorded | Date_Full | Correct_Counter_Name | District_of_Paris |
|---|---|---|---|---|
| 100036719 | 33 | 2018-05-01 | 18 quai de l'hotel de ville NO-SE | 75004 |
| 100044495 | 1 | 2018-10-01 | 7 Avenue de la Grande Armee NO-SE | 75016 |
| 100007049 | 17 | 2018-06-01 | 28 Boulevard Diderot E-O | 75012 |

**Query Group By**

- **Grouping to see number of bicycles per Date**

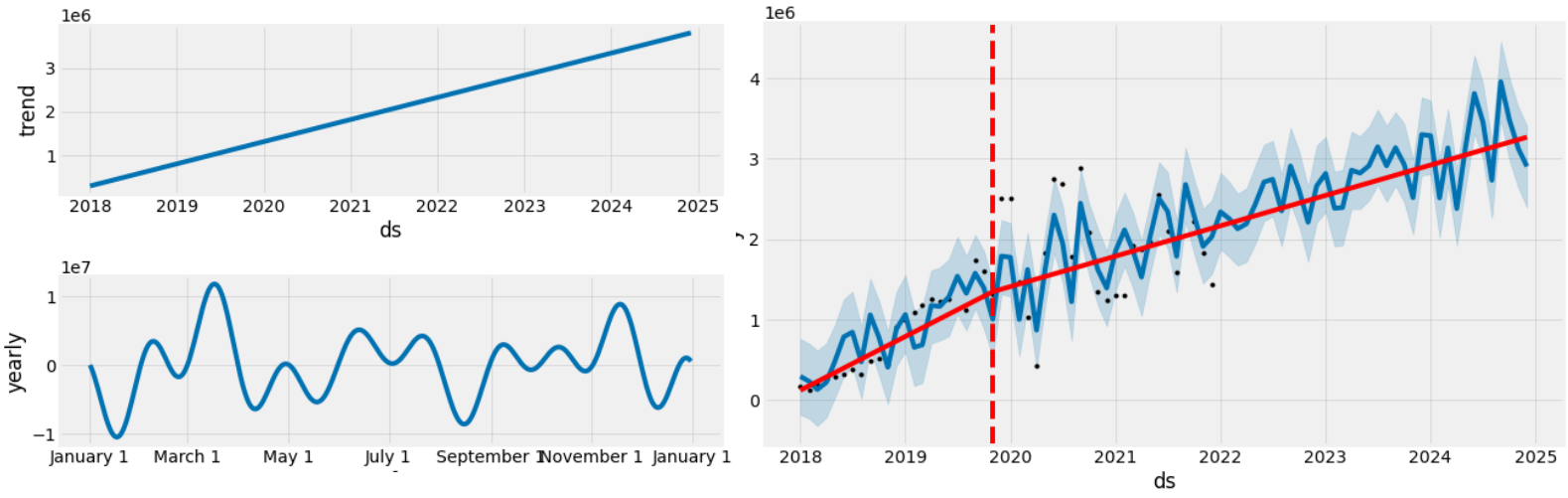
```
SELECT Date_Full, Bikes_recorded
FROM sql_sample_2018
GROUP BY Date_Full
ORDER BY Bikes_Recorded Desc;
```

| Date_Full | Bikes_recorded |
|---|---|
| 2018-11-01 | 176 |
| 2018-09-01 | 97 |
| 2018-05-01 | 33 |

## Forecasting - Time Series Forecasting with Prophet

**The Prophet library** is an open-source library designed for making forecasts for univariate time series datasets. It is easy to use and designed to automatically find a good set of hyperparameters for the model in an effort to make skillful forecasts for data with trends and seasonal structure by default.

- The input to Prophet is always a dataframe with two columns: ds and y.
- The ds (datestamp) column should be of a format expected by Pandas, ideally YYYY-MM-DD for a date or YYYY-MM-DD HH:MM:SS for a timestamp.
- The y column must be numeric, and represents the measurement we wish to forecast.




Using the Prophet Library in Python I was able to visualize trends and forecasts for upcoming years. As the number of Counters has changed across the years, I have decided to base my analysis on the set of machines installed in 2018 - I hope this will add stability to my analysis.
Following the above models, we can expect a **strong increase in the number of bicycle rides in Paris** - despite temporarily lower numbers during the Covid-19 pandemy. Trend is clearly positive and together with earlier information about plans to make Paris completely cyclable by 2026 - **we can be optimistic about their future.**

## Conclusions

Based on detailed analysis of received data, we can conclude that Bicycles have a bright future in the city of Paris.

Number of bicycle rides is expected to increase significantly in coming years, despite Covid-19 Pandemic. More - it's possible that a pandemic indeed increased support for this type of transport, as citizens are less reluctant to use public transport, at the same time owning a car in Paris will become more expensive and difficult.

While working on the data, I had to consider an uneven number of the counting machines that were first installed in small numbers in 2018 - factor that could negatively affect my analysis and the forecast.

Top 3 Districts with highest number of travelers are:

| District Full | |
|---|---|
| 11 - Popincourt | 16,643,303 |
| 10 - Entrepôt | 11,849,945 |
| 13 - Gobelins | 11,369,603 |

| District Full | |
|---|---|
| 11 - Popincourt | 14.08% |
| 10 - Entrepôt | 10.02% |
| 13 - Gobelins | 9.62% |

## Links

Github:
https://github.com/Borysdb/Final_Project_Bootcamp.git

Trello project planning:
https://trello.com/invite/b/ZCjQ2NCa/ATTI35a3736b31a64c53698c38770e84aacbC4EAD14A/final-project-ironhack-2022-bicycle-counters-in-paris