

Bikes sales and repairs shops in Paris using geographical data

1. Description of the problem and background

1.1. Background

In 2015, a constatation has been made that Paris was among the european capitals where citizen use bike the least. Only 2 % of distances were made by bike which drastically contrast with cities like Copenhagen, Amsterdam or Berlin with 31%, 22% and 13% of distance made by bike [\[ref\]](#).

Therefore the Paris's city council decided to launch the program "Paris, world capital of bike" aiming at reversing the situation and make Paris the top capital of the world where bike is used the most. Between 2015 - 2020 many realisations have been made including constructing new bikeways, the reduction speed limit in certain avenues and roads in Paris to insure safety cycling. In 2020 some statistic revealed that the usage of bike in Paris has increased to above 15 % [\[ref, ref\]](#).

1.2. Business Problem

As investor in the sector of bike selling or reparation, such an increase of bike usage represents a great business opportunity.

The objective of this project is to find an optimal location for a bike shops either for selling, reparation or both. The location of interest should preferentially have no many bike shops nearby and should be dense of bike users.

A second objective of the project is to know borough or neighborhood more efficiently covered by bike facilities including bikeways and sharing shops. This can be interesting to citizens willing to make things easier by using a bike instead of a car giving the jump in the traffic and difficulties finding a parking which in fact may be expensive.

2. Data description

The data were obtained from two main sources. The Foursquare API and [Paris open data](#) which is a website of open source data concerning the city of Paris.

For the project, I have requested four main datasets from these two databases which are :

- **Geographical and administrative information regarding Borough in Paris**
- **Data of Paris bikeway network**
- **Bike counting data in Paris**
- **Paris bike shops obtained from foursquare**

To make the data ready, I have gone through and important data preprocessing steps including, data cleaning, data selection and feature engineering. Then, I merged them all into a single dataset to be used in the machine learning model.

In the following sections, I describe the four dataset and all the transformation made.

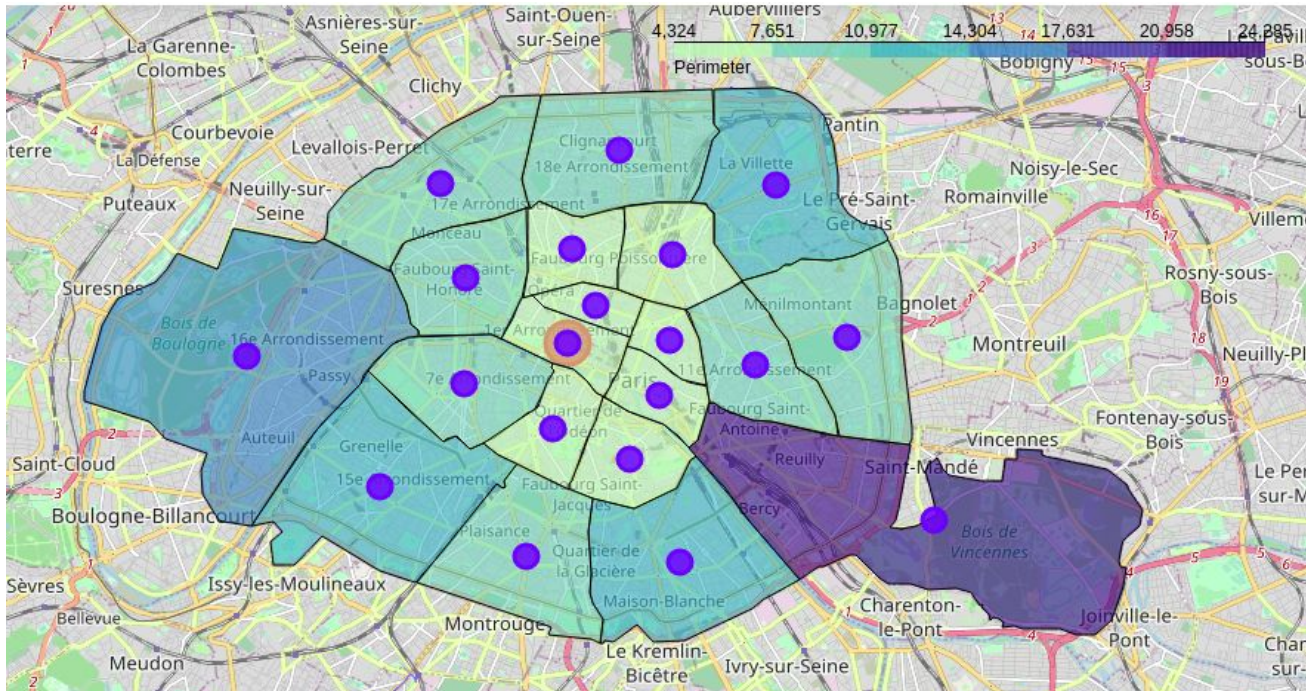
2.1. Geographical and administrative information regarding Borough in Paris

This dataset is identified as 'Arrondissements' in [Paris open data](#) website. Note that the word 'Arrondissements' is the french translation of 'borough'. In other word this dataset provides geographical and administrative informations regarding the 20 boroughs of Paris. This includes common and official names, the perimeter and the area, the geographical coordinates and other metadata. Here is the head of the DataFrame constituted (restricted to the relevant data) from the retrieve Json file.

	official_name	common_name	lat_borough	long_borough	perimeter	approx_radius
0	Reuilly	12ème Ardt	48.834974	2.421325	24089.666298	3835.934124
1	Gobelins	13ème Ardt	48.828388	2.362272	11546.546526	1838.622058
2	Popincourt	11ème Ardt	48.859059	2.380058	8282.011886	1318.791702
3	Vaugirard	15ème Ardt	48.840085	2.292826	13678.798315	2178.152598
4	Opéra	9ème Ardt	48.877164	2.337458	6471.588290	1030.507690

The columns '**approx_radius**' refers to the approximated radius of the any borough assuming that they have circular shapes while in reality they are polygons. **This feature is to be used as the radius when retrieving the bike shops from the Foursquare.**

I used python folium library to visualize the geographic chart of Paris and its boroughs superimposed. I created the using the longitude and latitude coordinates of borough limits and centers. In the previous DataFrame, I conceived only the center borough (represented with blue circles in the chart) coordinates as they will be necessary for retrieving data from Foursquare. The color of borough are scale on their perimeters, one can see that bigger the borough is the darker it color is in the map.



2.2. Data of Paris bikeway network

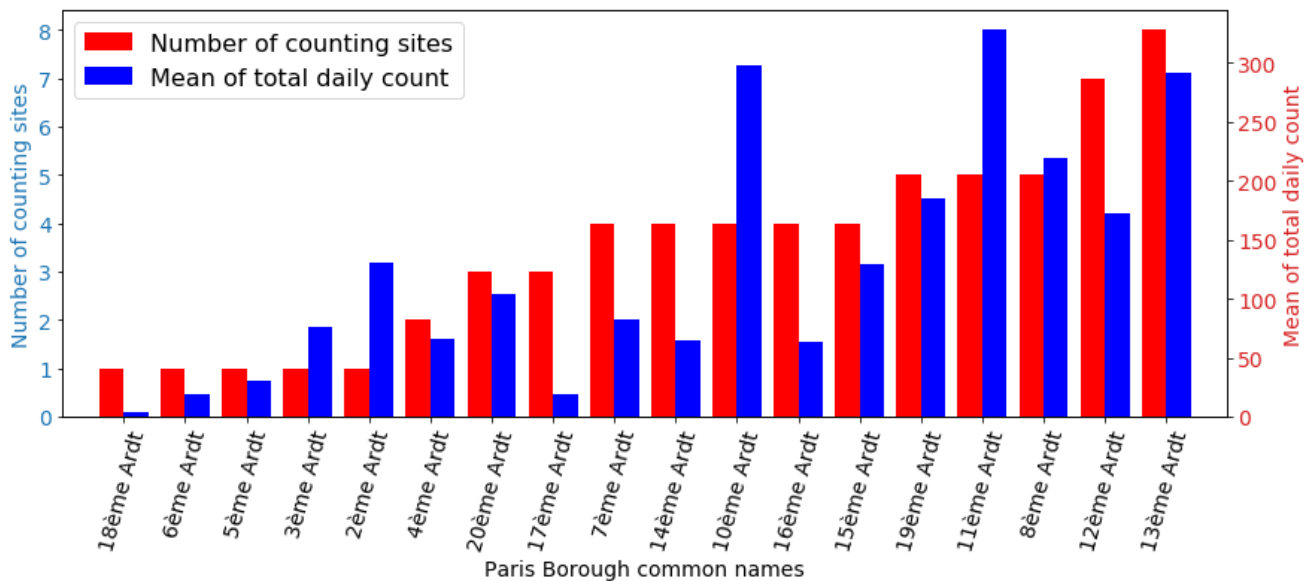
This dataset contains information regarding the Paris bikeway network such as the length, the longitude and latitude coordinate, their location, the date of construction, the direction of circulation (double or unique) and many other features. In total, this data set represents 11954 bikeways with 19 descriptive features¹. However, for the use case of concern here, not all the descriptive features are relevant. I cleaned the dataset and reduced it to the below table (transposed for representation).

Borough common name	8ème Ardt	13ème Ardt	9ème Ardt	1er Ardt	4ème Ardt	17ème Ardt	12ème Ardt	14ème Ardt	15ème Ardt	16ème Ardt	6ème Ardt	7ème Ardt	5ème Ardt	10ème Ardt	3ème Ardt	2ème Ardt	11ème Ardt	19ème Ardt	20ème Ardt
Total borough bikeway distance (km)	37.35	86.73	27.98	39.12	33.88	48.98	112.09	55.35	76.88	72.05	23.16	35.14	31.12	39.85	16.8	24.72	56.33	76.43	80.17

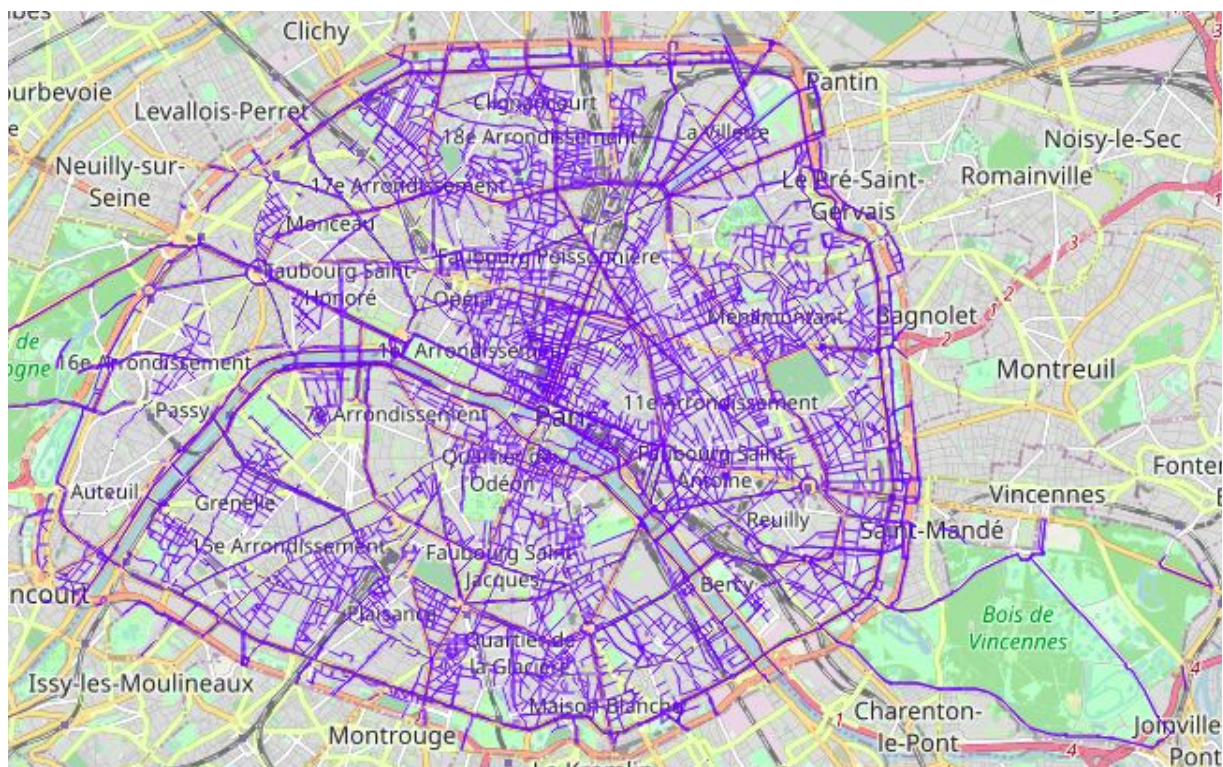
This table presents the total bikeway distance in km in each of the 20 boroughs of Paris.

The bar plot below provides a more clearer description of such reduced data. One observes a clear correlation between the total bikeway distance constructed and the size (perimeter) of the borough. Though this is clearly normal, caution has to be taken given that some borough host large parks. This is the case for example of Luxembourg (6ème Ardt) and Reuilly (12ème Ardt) boroughs respectively hosting the two most famous parks in Paris namely the '**vincennes park**' and '**boulogne park**'.

¹ df_paris_bikeway.shape = (11954, 19)



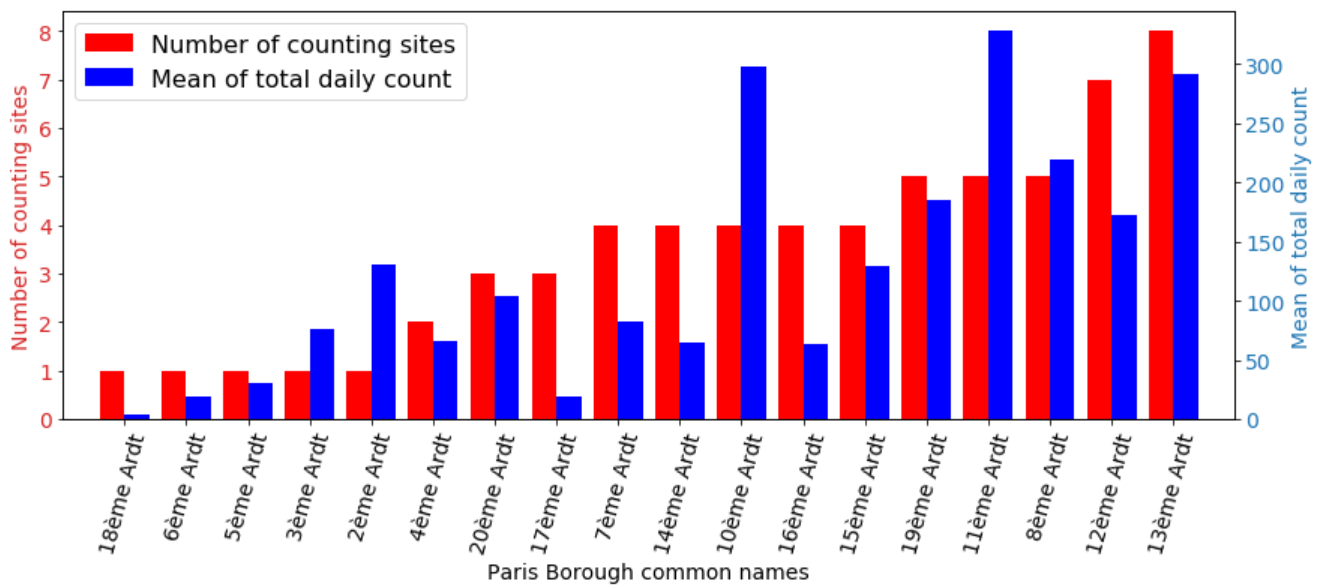
Following, I provided a folium map chart with all the bikeways superposed onto. It gives a geographical illustration of the distribution bikeway network in different borough of Paris.



2.3. Bike counting data in Paris

This is a hourly bike counting dataset in Paris since 2012. The counters devices are randomly installed in certain avenues and bikeway of Paris. This means that, there are not the equal number of counters in each borough. Therefore, this data provides simplement an indication on the daily number of cyclists using a bikeway in each neighborhood of Paris. Thus, it give a tendency on the the

population of cyclists in Paris neighborhoods which is an important information for an investors in bike repairation and/or selling sectors.



I cleaned the dataset and reduced it to the average daily bike count for each counting site. The bar plot below above the average daily bike count per counting site (blue bar, right axis) and the number of counting site per borough (red bar, left axis). One see a clear disproportion in the distribution of counting sites per borough remaining the caution to be taken when referring on this data during the interpretation of the model.

2.4. Paris bike shops : Foursquare data

I used Foursquare API to retrieve bike shops of given Borough. To do this, I deduced the approximated radius parameter from the perimeter of the borough approximating them to a circle.

Here is the 'search query' I used. I multiplied the keywords considering also the name of bike shops I found on google map.

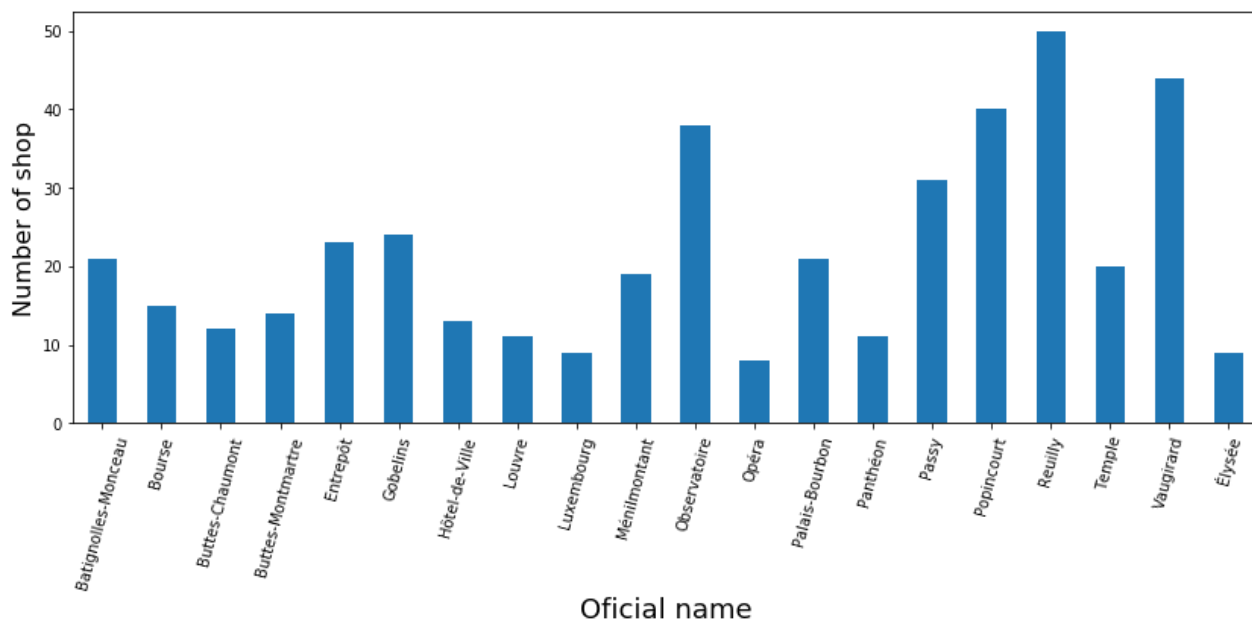
```
search_query_list = ['Bike', 'Byke', 'Vélo', 'bicyclette', 'Cyclable',
                    'Bicycle', 'CYCLES', 'cyclistes', 'Bicloune', 'Buzibi',
                    'Bicycland', 'Bicycl'Art', 'Velotority', 'Cycloparis15',
                    'Cyclo-store', 'Velobecane', 'vintelo', 'Cyclable']
```

I got in total² 2752 which after cleaning items with categories not referring anyhow to a bike shops it was reduced to 433 shops folding in a category related to bike shops. The head below illustrates the clean dataframe of the Foursquare data.

² df_bikeshops_paris.shape = (2752, 6)

	bike shop name	shop lat_long coord	Official name	categories	common name
0	Vélo Electro	[48.87780147075355, 2.33739685142654]	Opéra	Bike Shop	9ème Ardt
1	Vélib' [15-53]	[48.8393974466911, 2.2914993051433905]	Vaugirard	Bike	15ème Ardt
2	Vélib' [01-23]	[48.863602052393816, 2.335514831205776]	Louvre	Bike	1er Ardt
3	Vélib' [09-15]	[48.876719598116196, 2.339334127013303]	Opéra	Bike	9ème Ardt
4	Cycles Sport Urbain	[48.858334, 2.381948]	Popincourt	Bike Shop	11ème Ardt

The bar chart below shows the distribution of the bike shops retrieved from Foursquare API in Paris Boroughs.



The final step in data preprocessing consisted of merging all dataframe to a single dataframe.