Applied Data Science Capstone Project

Electrical Vehicle Charging Stations in Paris

Table of contents

I. Introduction

An electric vehicle charging station, also called EV charging station is a machine that supplies electric energy for the recharging of plug-in electric vehicles.

Some electric vehicles have converters on board that plug into a standard electrical outlet or a high-capacity appliance outlet. Others use a charging station that provides electrical conversion, monitoring, or safety functionality. These stations can support faster charging at higher voltages and currents.

Public charging stations are typically on-street facilities provided by electric utility companies or located at retail shopping centers, restaurants, and parking places. They can be operated by various private companies.

1. Business understanding

Our client is a private company which wants to invest on charging stations for electrical vehicles. The company choose Paris for their investments. The stakeholders want to investigate the city to choose the right places where they can set their new charging stations.

So, the questions that we are intended to answer are:

Are there sufficient customers that have electrical vehicles in Paris ?

Can we predict the number of electrical vehicles in the future years ?

What are the places in the city that have not sufficient charging stations ?

II. Data

We will use the following data to answer the questions:

- https://www.automobile-propre.com/dossiers/chiffres-vente-immatriculations-france/
  This site is well known in France for the expertise in the automotive field for ecology. We will use Web Scraping to extract tables of the number of electrical vehicles sold in the last years.

- We will use the Foursquare location data to identify the venues which contain the category of "EV Charging Station" as mentioned in their official website:
  https://developer.foursquare.com/docs/build-with-foursquare/categories/
  we can see an extract of the category in the following image:

- https://www.data.gouv.fr/fr/datasets/bornes-de-recharge-pour-vehicules-electriques-2/
  This site, which is an official French government site, contains dataset of charging stations of the entire country of France with the coordinates XLongitude and YLatitude. We will use the csv file to complete information in our data frame.

III. Methodology

1. Analytic approach

We will start answering the two first questions (Are there sufficient customers that have electrical vehicles in Paris ? and Can we predict the number of electrical vehicles in the future years ?) by exploring the tables of the web sites described in the Data section and try to set a model to predict the number of electrical cars for the future years. After we will use the foursquare API to cluster the neighborhoods in function of the existing Electrical charging stations in the venues of each neighborhood to answer the last question (What are the places in the city that have not sufficient charging stations ?).

2. Data wrangling

We used Web Scraping to extract the table of the number of electrical cars and trucks in function of the last 10 years. We obtained the following data frame after cleaning and converting data:

	Year	Electrical_cars	Electrical_trucks
0	2010	180	796
1	2011	2630	1683
2	2012	5663	3651
3	2013	8779	5175
4	2014	10560	4485
5	2015	17266	4949
6	2016	21751	5556
7	2017	24910	6011
8	2018	31055	8103
9	2019	42763	7958
10	2020	110916	8792

We used the "read_csv" method to read the csv file for the coordinates of the neighborhoods of Paris. After wrangling data, we obtained the following data frame:

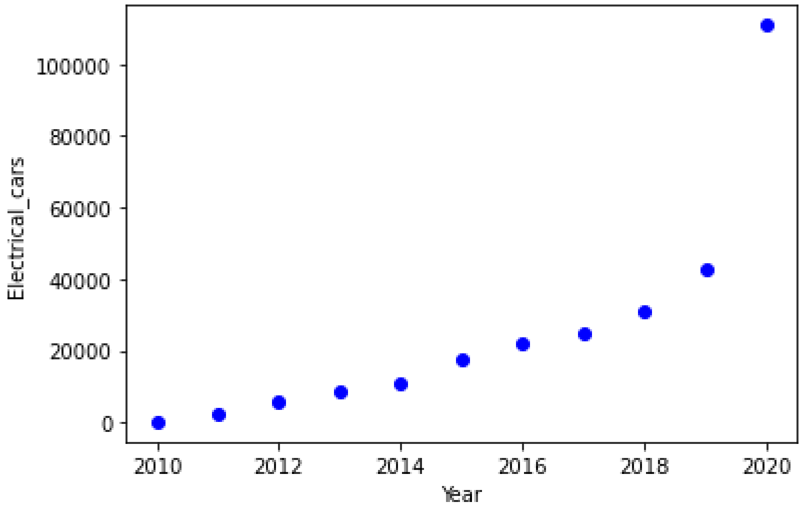
	Neighborhood	Latitude	Longitude
0	Monnaie	48.8543844036	2.34003537113
1	Odéon	48.8478006293	2.33633882759
2	Champs-Elysées	48.8670744922	2.30865168468
3	Maison-Blanche	48.8231278057	2.35243314954
4	Croulebarbe	48.8337336761	2.34767304607

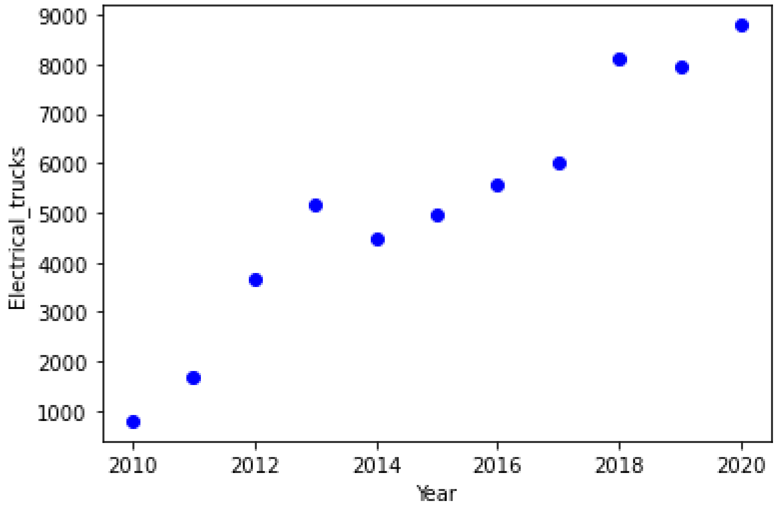
```
paris_info.shape
```

```
(80, 3)
```

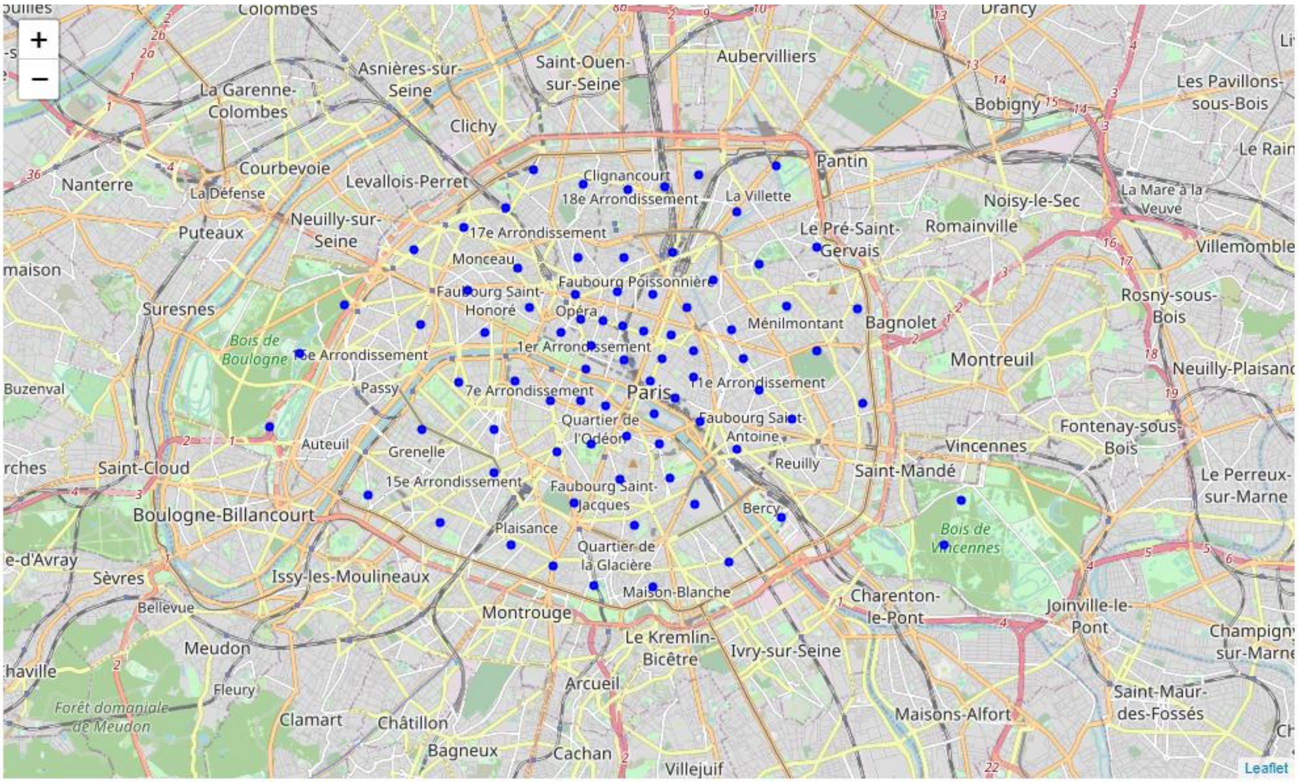
3. Data analysis

We wanted to see if there is any correlation between the 10 years and the number of electrical cars and trucks. We found that there is a positive correlation between them:





We created the map of Neighborhoods of Paris using Folium library and the data frame containing the latitude and longitude:



We can see that we have all the 80 neighborhood that we will analyses their venues with the Foursquare API by grouping these neighborhoods by venue's categories:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Amérique	12	12	12	12	12	12
Archives	100	100	100	100	100	100
Arsenal	68	68	68	68	68	68
Arts-et-Métiers	100	100	100	100	100	100
Auteuil	15	15	15	15	15	15
...	...	...	...	...	...	...
Sorbonne	100	100	100	100	100	100
Ternes	66	66	66	66	66	66
Val-de-Grâce	45	45	45	45	45	45
Villette	58	58	58	58	58	58
Vivienne	91	91	91	91	91	91

80 rows × 6 columns

	Neighborhood	Accessories Store	Afghan Restaurant	African Restaurant	Alsatian Restaurant	American Restaurant	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	...	Vegetarian / Vegan Restaurant	Venezuelan Restaurant	Video Game Store
0	Amérique	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.00	...	0.000000	0.0	0.0
1	Archives	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.05	...	0.000000	0.0	0.0
2	Arsenal	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.00	...	0.029412	0.0	0.0
3	Arts-et-Métiers	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.01	0.02	...	0.020000	0.0	0.0
4	Auteuil	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.00	...	0.000000	0.0	0.0

5 rows × 301 columns

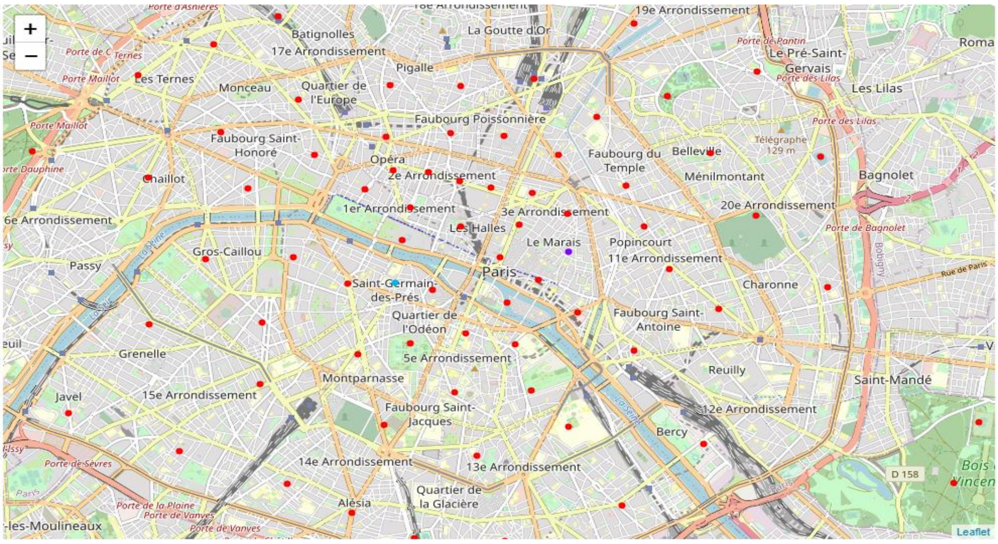
4. Modeling

Since we have continuous values in our features and we want to predict a numerical value, we have a supervised learning and we will use a simple linear regression to predict the number of electrical cars and trucks for the year 2021. The target is "Electrical_trucks", and the feature are "Year". We won't predict the Electrical cars because we can already be sure that the number of those cars is increasing with a very high rate, we will only predict the Electrical trucks:

```
# Let's import the libraries
import numpy as np
from sklearn import linear_model
from sklearn.model_selection import train_test_split


X=EV[['Year']]
y=EV['Electrical_trucks']

# We will split data into train and test split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.1)

regr = linear_model.LinearRegression()
regr.fit (X_train, y_train)

# The coefficients
print ('Coefficients: ', regr.coef_)
print ('Intercept: ',regr.intercept_)
```

```
Coefficients:  [737.15441176]
Intercept:  -1480251.9264705887
```

Concerning the neighborhoods of Paris, we used clustering to group the similar neighborhoods based on the existence of Electrical Charging Stations in each neighborhood’s venue. We used K-means clustering:



5. Model evaluation

We used the R-squared metric which is a popular metric for accuracy of our model. It represents how close the data are to the fitted regression line. The higher the R-squared, the better the model fits our data. Best possible score is 1.0 and it can be negative (because the model can be arbitrarily worse).

```
# Let's import the libraries
from sklearn.metrics import r2_score

print("R2-score: %.2f" % r2_score(y_test , _y_test) )
```

```
R2-score: 0.71
```

we have 0.71 as R-squared metric, it means that our model is well fitted to predict the number of Electrical trucks.

6. Model deployment

Let's predict the number of Electrical trucks for the years 2021 and 2022:

```
d = {'Year': [2021, 2022]}
X_predict=pd.DataFrame(data=d)
y_result=regr.predict(X_predict)
y_result
```

```
array([ 9537.13970588, 10274.29411765])
```

we will have year_2021=9537 Electrical trucks and year_2022=10274 Electrical trucks. It means that the number of Electrical trucks will increase the coming years.

IV. Results

The coming years, the number of electrical cars and trucks will increase with a very high pace. Our model predicted ten thousands electrical vehicles for the coming two years.

The result of the clustering is that there is not enough Electrical Charging Stations at the city of Paris, except two Neighborhoods "Le Marais" and "Saint-Germain-des-Prés". The stakeholders can try to set their Electrical Charging stations at any Neighborhood they want except the two identified "Le Marais" and "Saint-Germain-des-Prés" neighborhoods to avoid competition with other companies.

V. Conclusion

The project of installing Electrical Charging Stations will be a success for the Energy Companies for two main reasons: Profitability and respect to the environment using green and clean energy. The city of Paris is a good choice for the project, because the city is very crowded and needs to adapt to the coming challenges concerning the ecology and health care for the residents who can't keep harming them selves with dirty energy and CO2 emissions.

We discussed in the results section that nearly all neighborhoods are well suited to settle the stations except two neighborhoods. However, which of the neighborhoods are cheaper or expansive for the rent ? We can propose to stakeholders this question to discuss and answer by using a new Data Science Project for the pricing of renting fields.