Data-science project Report :

ELABORETED BY:

BILEL JEDDA

TALEL JEDDA

NAJAH FREDJ

SAMAR LAAJILI

M2 SIA

2020-2021

1- Introduction :

Since Rshiny is one of the most popular tools to exploit and interact with data, we chose it in order to analyse information we have through our database.
In this project, we worked on "ma rue de la Mairie de Paris" data and we tried to apply the notion of ggplot and its main functions .
Throughout this report, we will try to explain the results .
Actually, we worked on 6 plots divided into 3 parts :
- Distribution of incidents by postal code and type .
- The rate of incidents according to years and postal code
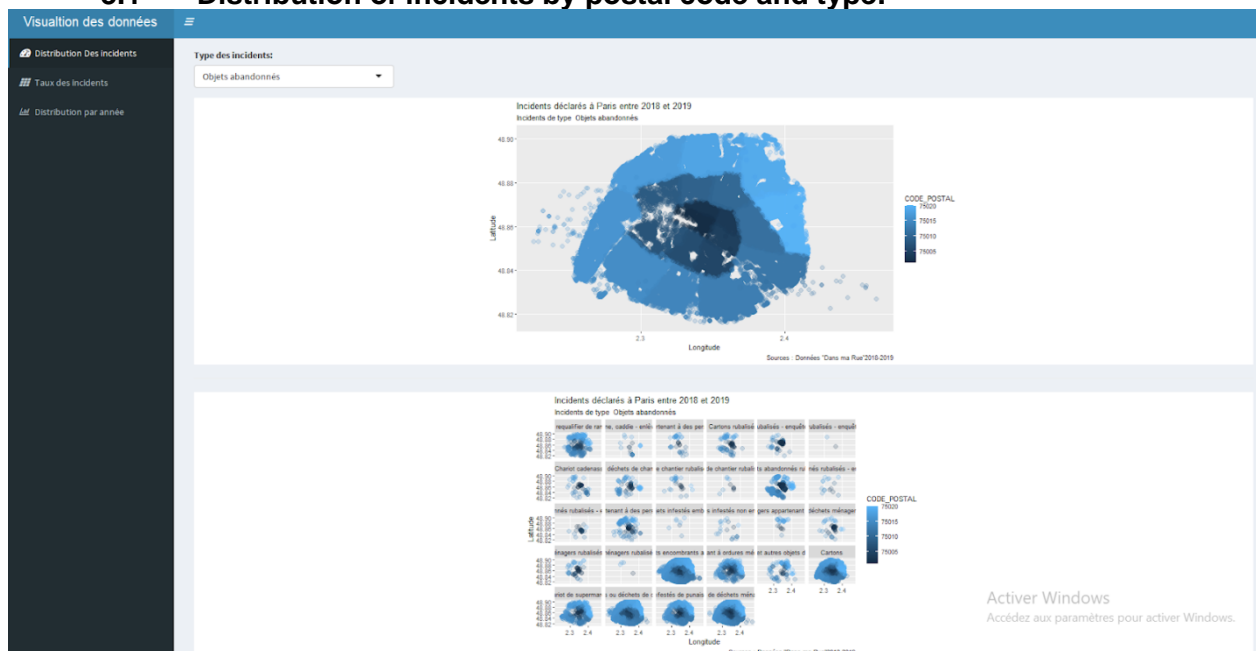- Distribution of incidents by type and months .

2- Data understanding :
We have 7 columns : TYPE , SOUSTYPE , CODE_POSTAL , DATEDECL , INTERVENANT , lat, long with a total of 538390.

3- Data analysis :

- We have to mention that the load of data and plotting will take some time since we are manipulating a huge amount of data .

3.1- Distribution of incidents by postal code and type:

In the first part , we have the choice to use the type of incident with a dropdown list. We figured that we have 11 possibilities . After picking an option, two plots will be displayed. Knowing that the colors represent the postal code,the first plot will show the distribution of the option chosen in Paris.. the density of the plot shows the number of incidents by location.
However, the second plot will show the distribution for all the subtypes of the option chosen. We can conclude that the density changes according to the subtypes .

3.2-   The rate of incidents according to years and postal code :
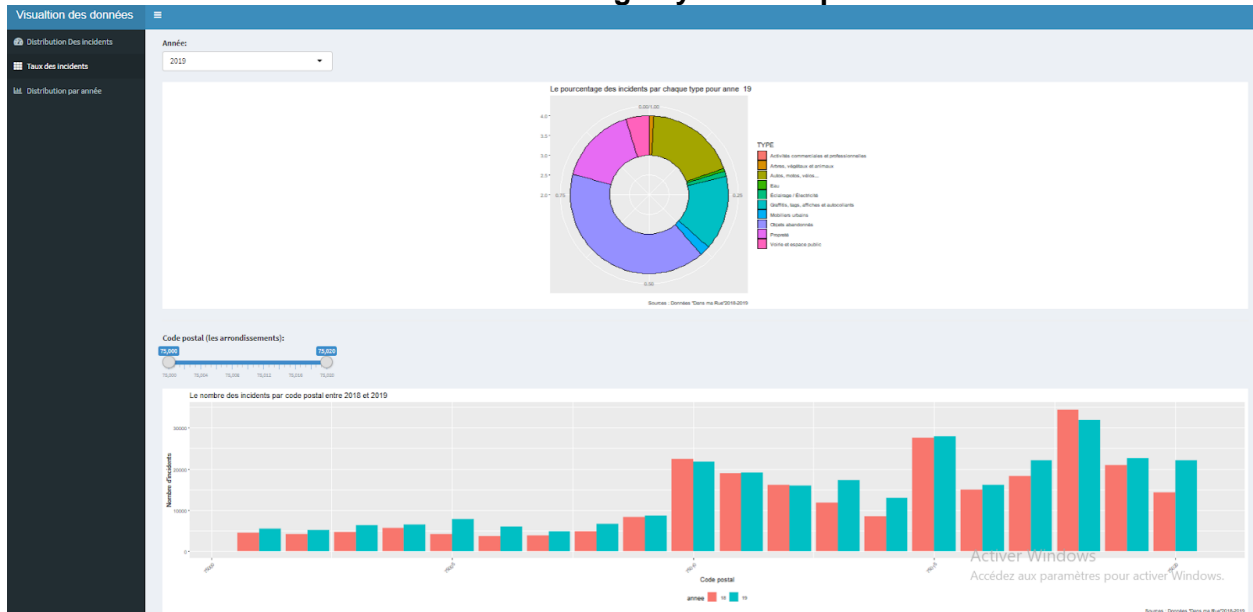
In the second part, we show firstly the percentage of each type of incident according to the year chosen in the dropdown list (in this case we have only two options : 2018 or 2019) . We also have a slider that gives us the opportunity to pick a postal code. As a result , we obtain a plot that displays the number of incidents per type and per year . According to this plot we can conclude that the numbers of incidents are moderately growing and centralized in some districts .

3.3-   Distribution of incidents by type and months :
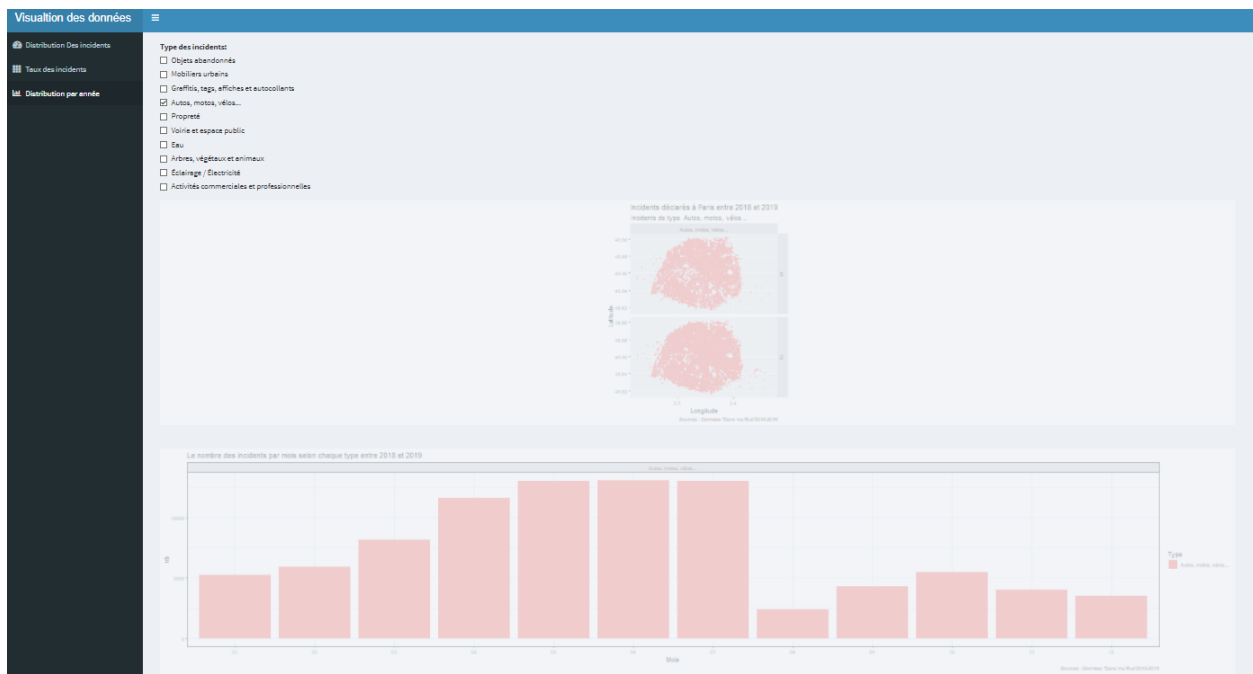
In the third part , we have the opportunity to pick many options since we had recourse to the check box . The purpose of the first plots is to show the density of the chosen incident according to latitude and longitude. Secondly , we can conclude the most periode that a specific incident can take place.

4- Conclusion :

During this project , we tried to explore our data using different types of plots and manipulate it using ggplot . it gives us the opportunity to know that the center of paris is the main source of incidents. It makes us think about the causes of this. It may be the number of the population . However we don't have the exact answer to this question since we don't have the data to explore the reasons .