Covid-19 France Analysis

Nassim Ben Yaagoub
Lourdvic Alcindor

03 Mai 2020

—

Visualisation of Massive Data

—

Nicolas Lehir

Introduction

Covid-19 pandemic is the subject of those days, so we wanted to know more about his spread over France.

In this analysis we will answer two main question:

Does the age really influence the contamination of the virus?

Which France's department are the more affected ?

Do the spread will continue to grow, maintained or reduce in the next 2 weeks?

In order to answered those question we make a data analysis of this two datasets:

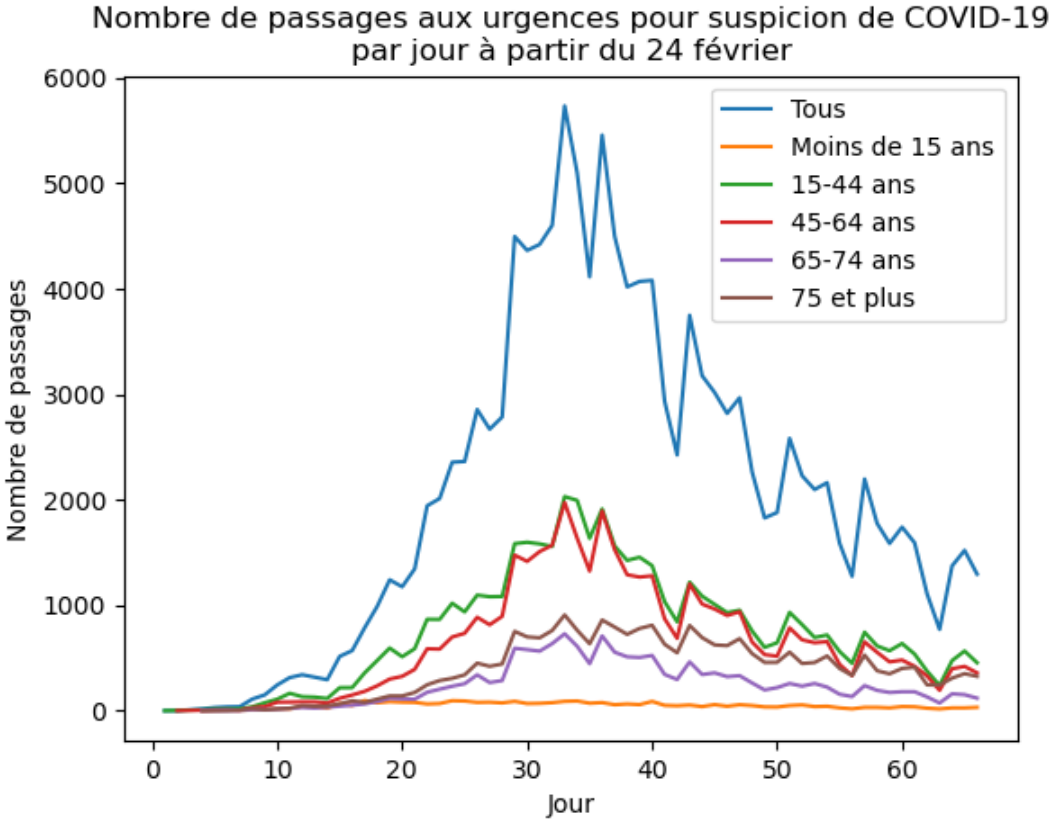
- *Données des urgences hospitalières et de SOS médecins relatives à l'épidémie de COVID-19*
- *COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University*

DATA ANALISYS

AGE VISUALIZATION – SCATTER PLOT

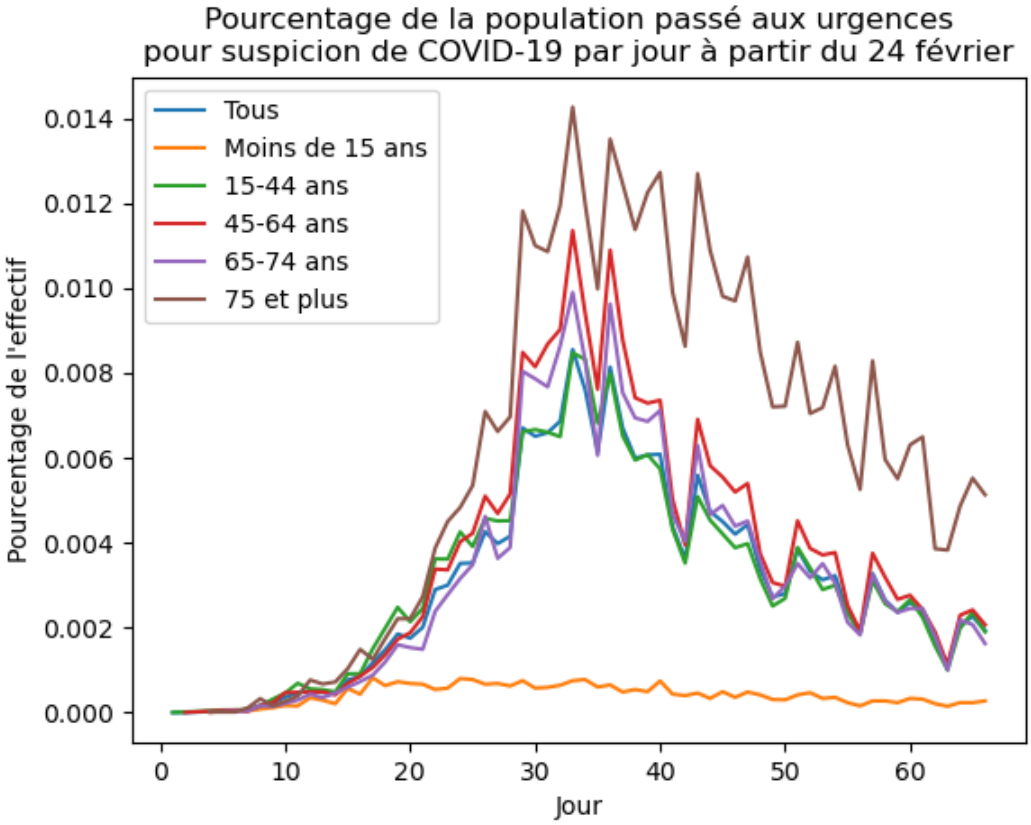
To answer the first problematic, we made an age visualization by doing a scatter plot of the evolution of the emergency room visits for COVID-19 suspicion according to the age range.

(0 : Tous, A: Moins de 15 ans, B: 15-44 ans, C: 45-64 ans, D: 65-74 ans, E:75 et plus)



Surprisingly, we see that it's the "15-44 ans" with the "45-64 ans" who come more often to the emergency room.
Maybe it's because they are more often outside as a difference to the "65-74 ans" & the "75 et plus".

So to be sure let's redo a scatter plot but this time let's analyses by percentage because it's possible that there are more B & C than D & E. Depending on the number of the people, the results can be slightly different.



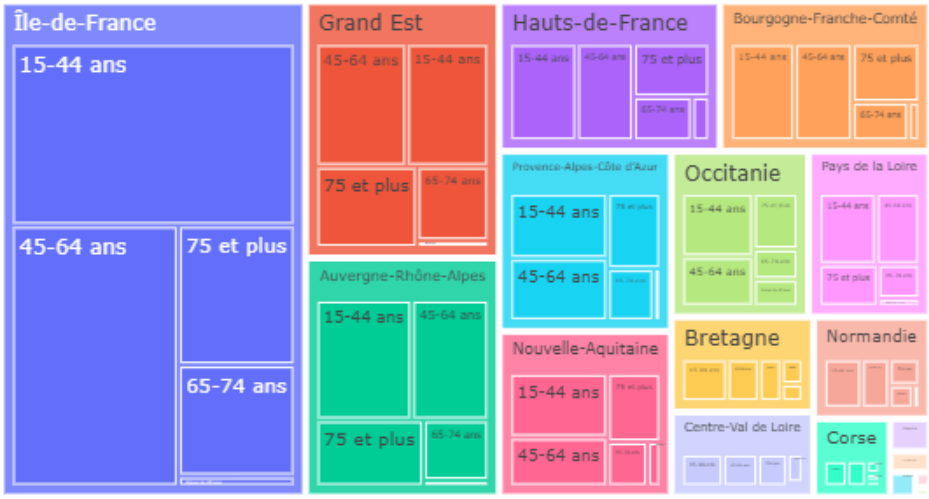
So, there we see that the "75 ans et plus" are coming way more often than the other, then the C & D. The C's are coming more often can be explain that they are more often outside than the D's.
We also see that the "Moins de 15 ans" are not affected that much.

Thanks to this scatter plot by percentage we can conclude that the age influences a lot in the contraction of the virus. The old people are enormously affected at the opposite of de young who are way less affected by the COVID-19 pandemic.

AGE VISUALIZATION – TREE MAP

Now let us see in which region the people are coming the more to the emergency for a COVID-19 suspicion.

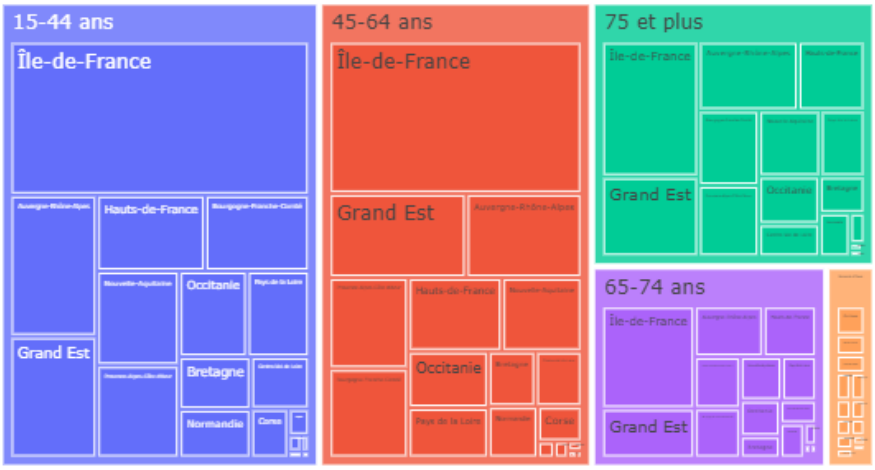
To see this clearly, we going to plot a tree map of the regional breakdown.



We can see that the "Ile-de-France" is more a way more affected by the COVID-19, the fact that IDF it is the most populated region And the third smallest region of the country can explain this result.

The fact that the "Grand Est" and "Auvergne-Rhône-Alpes" are 2nd & 3rd most populated region of France explain why they coming after IDF in the people coming at emergency for a suspicion of the COVID-19.

What about the ages of the people coming to the emergency? Let' us analyses it with a tree map to see how it is.



The tree map above shows that the "15-44ans" are the most coming to the emergency closely followed by the "45-64ans", we can conclude this by the fact that there is more young and active people in France than older people, moreover we can notice that the older people are more used to be lock-up than the younger one because of their vulnerability also because the most are in retirement.

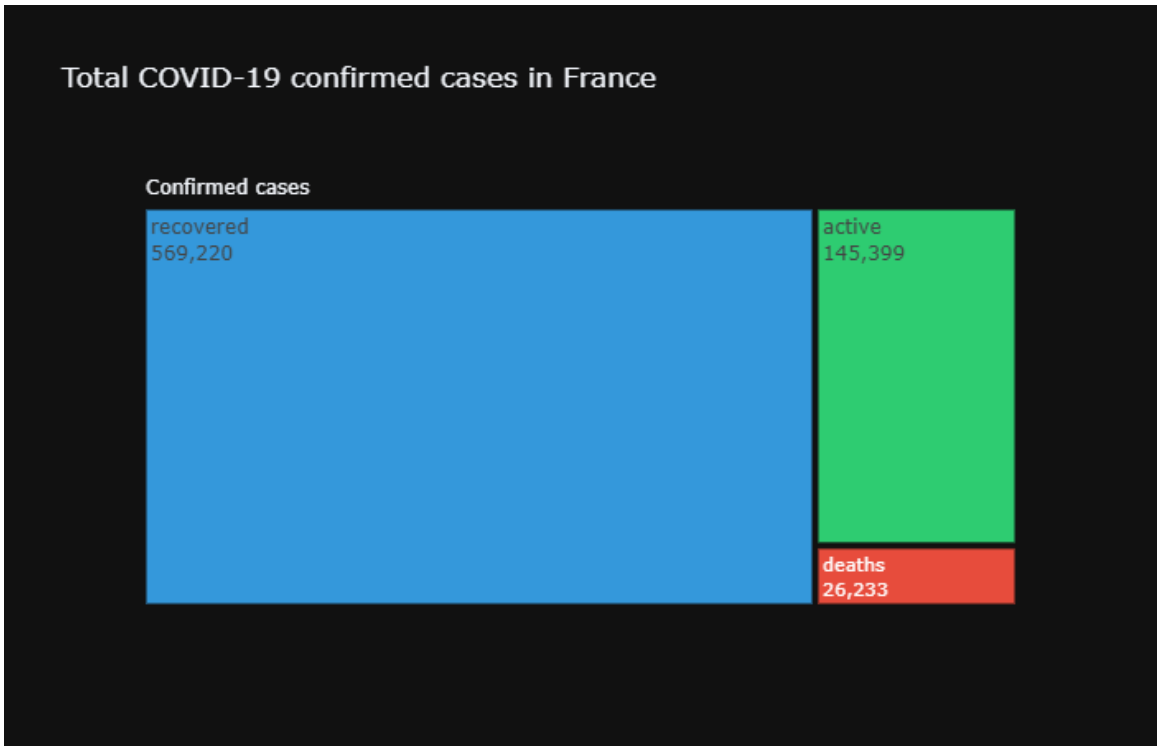
COVID-19 SPREAD ANALISYS – TREE MAP

We going to study the spread of the confirmed cases of the corona virus in France.

We use a daily-updated dataset that get the confirmed/deaths/recovered cases and we add and active columns by doing this simple calculation:

Active = Confirmed – Death – Recovered

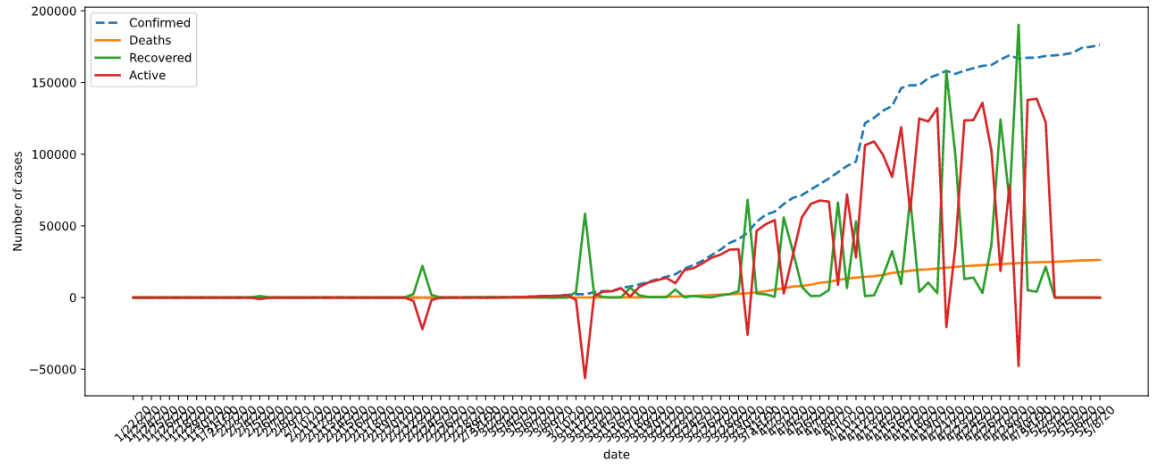
First, let us plot a tree map to see clearly if the virus kills more people than hospitals recovered.

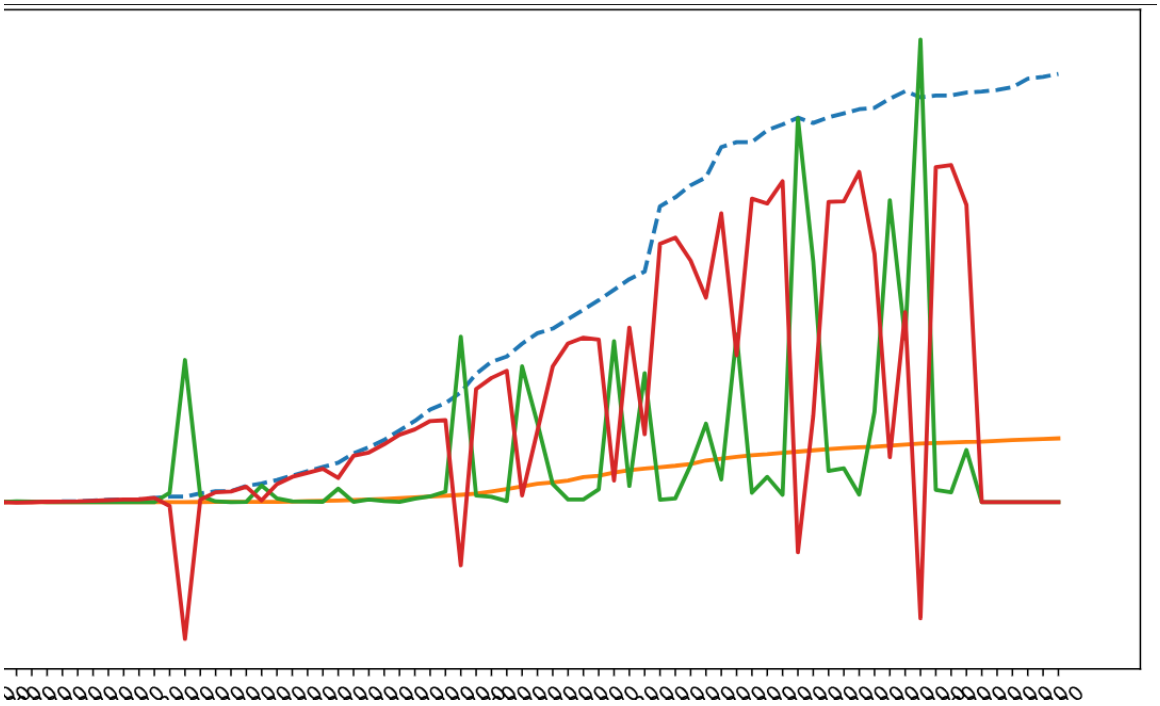


Thanks to this tree map we see clearly that there are more people who was recovered than dead. Also, we see that there is still a lot of people active but it is quite a few comparing all the people recovered. We can predict that we will arrive at the end of this crisis.

To be more precise let us plot a scatter plot to see the evolution of the spread since the beginning.

s



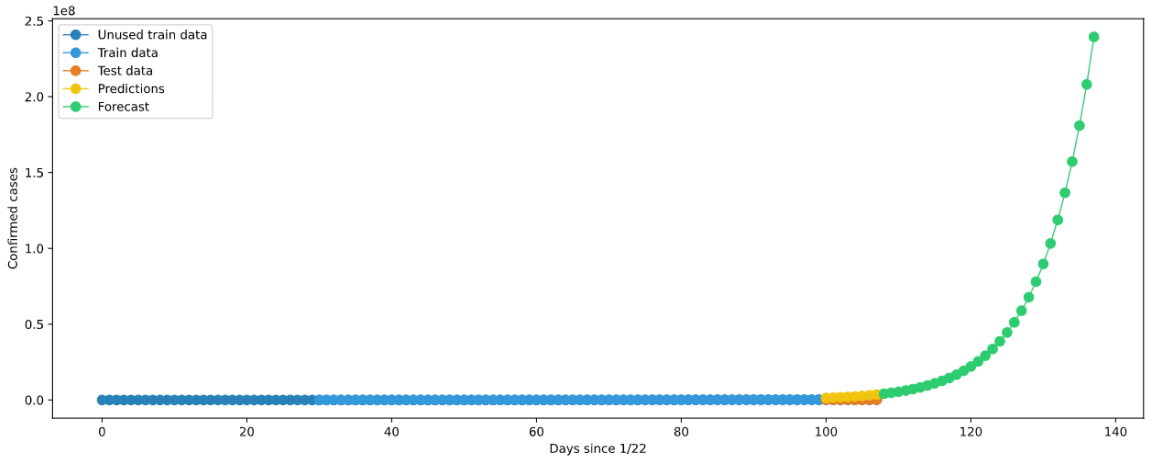


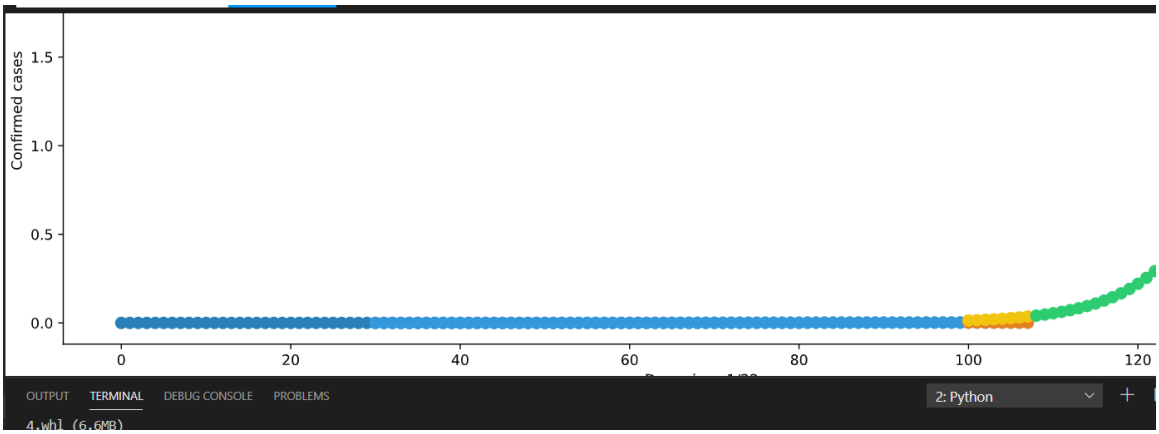
This scatter plot confirms that the recovered people increased and the people who were active reduce. That can explain why the government choose to open the lockdown.

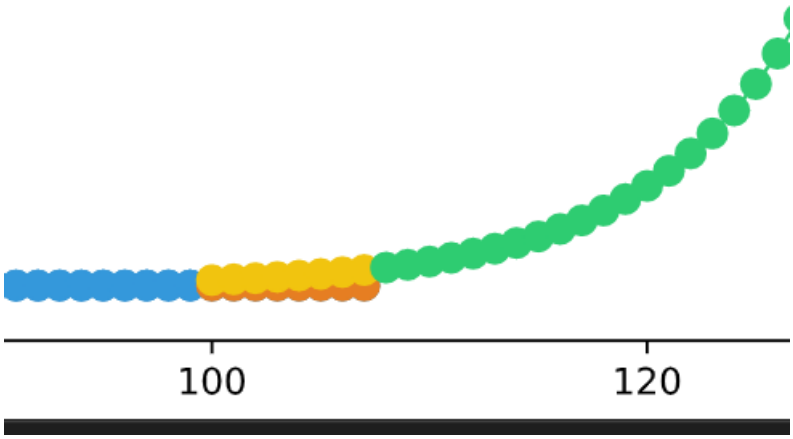
COVID-19 PREDICTION – LINEAR REGRESSION

In order to predict the forecast of the COVID-19 we are going to make linear regression model by using the sklearn library.

After split the dataset, for training & testing we evaluate we obtain a mean absolute error of `2.5447523775153593` , then we use it to predict the behavior of the spread of the next 30 days and here is the result that we get :



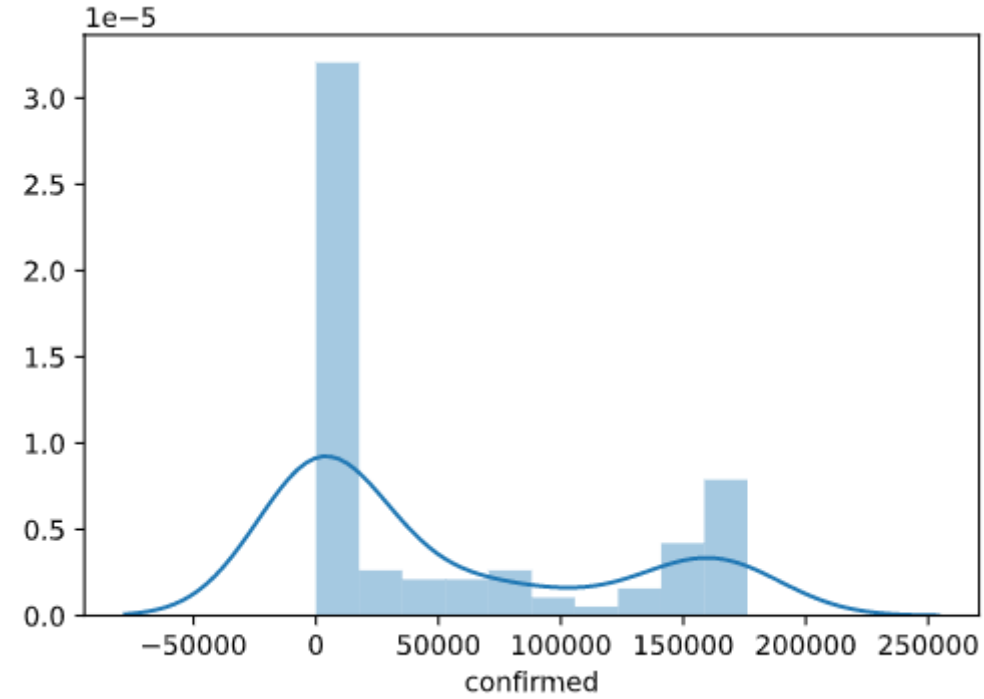


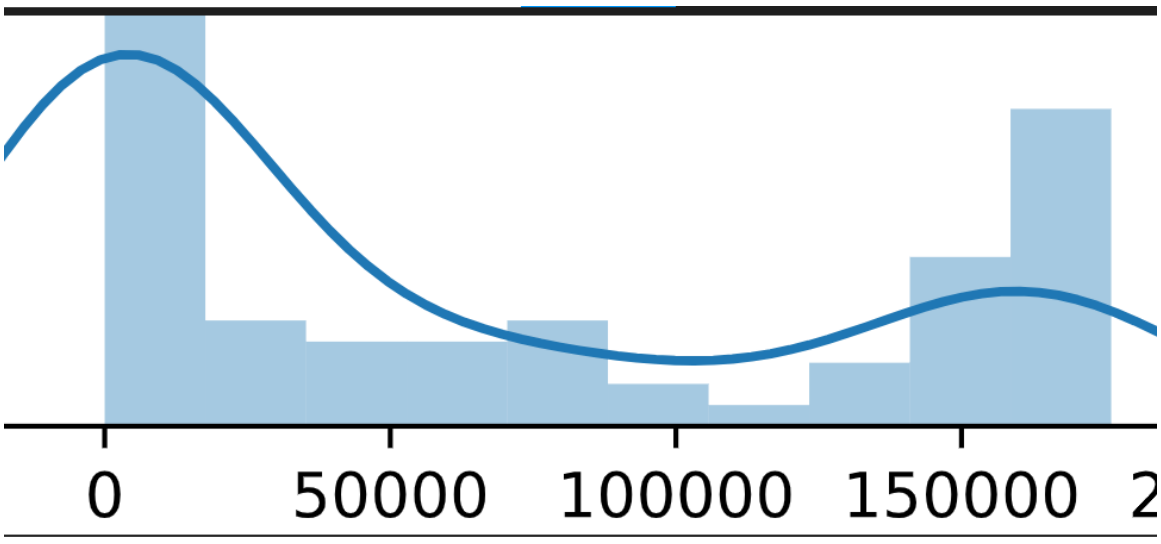


Through the prediction we realize that the spread will continue to grow and that we are in a danger of another wave of contamination of the virus.
So even if it is just a prediction we can not be sure 100% that it will goes like that and we hope not.
Nevertheless, even we about to be out of the lockdown, we still need to be cautious to avoid any return of this deadly virus.

COVID-19 FIT DATA TO DISTRIBUTION

We use the seaborn library to fit the data to a histogram distribution.





Thanks to this distribution we can notice that the apogee of this pandemic was from the first 20 000 confirmed cases after that the spread confirmed case spread diminished.

We also see that the spread increase between 125K and 170K confirmed cases but quickly reduce, maybe we were not that

surprise and the hospitals and the people where more caution and prepare to face this wave.

Sources:

Dataset 1 : https://www.data.gouv.fr/fr/datasets/donnees-des-urgences-hospitalieres-et-de-sos-medecins-relatives-a-lepidemie-de-covid-19/
Dataset 2 : https://github.com/CSSEGISandData/COVID-19