

Milad Toufani

**Data science for life science
Programming 01
2021-2022**

Research Question

How much vaccination has impact for population of the Netherlands people catching COVID-19 and decreases number of death and hospitalization?

First of all for finding answers for my research question I needed data for the Netherlands about people with positive Cov-19, Hospitalized and death amount of people, and also amount of vaccination for first, second and third doze of COV-19. I found some dataset which some of them contained data for all of the world and used them to find out my answers.

I found datasets from <https://ourworldindata.org/covid-vaccinations> website. I used three different dataset which I should merge those data and combine them together to investigate my research question.

First dataset (full_data.csv):

Information about people which diseased by COVID-19 and death cause of this problem for all countries in the world for every day from start of this disease.

Second dataset (covid-hospitalizations.csv):

This dataset shows the data about people hospitalized in all of the world caused by COV-19 for every day.

Third dataset (Netherlands vaccination.csv):

This dataset shows data about vaccination progress in the Netherlands country. It contains first, second and booster data which people got during this problem.

Steps To reach my answers from datasets:

First of all, I added all libraries which I used in my project.

I used Jupiter notebook .ipynb format.

In second step, I opened datasets (.csv files) to use them as a data frame. For this step I made a function which opens dataset and save them into different data frames.

After that I made a function to clean, rename, filter, and delete extra columns and prepare them for the next steps and merging all of them to use it for our research question.

There were some extra data which it was not related to my research question, and I dropped them out. The most important column which I made it as my index and use that column to merge my data frame for the next step.

For the next step, I made one function to merge my separate data frames together and use that to my source data frame.

Also, in this step I changed columns datatypes to use them in my plots.

I had some columns which they filled out with weekly data. Because of that, I used interpolating (linear method) to fill Nan data. And for the rest of Nan values which they were zero, I changed them to zero value.

For visualizing my data, first of all I used line bar to showing tabular data for positive cases, hospitalized and death during the period of time. The plot use hover tool to show exact date and number of cases.

For the next plot, I showed amount of first, second and booster doze of vaccination for the Netherlands people during time.

It shows from first month of 2021 vaccination process started and increased significantly until first month of 2022 which reached around 30 million of shots totally.

For the next plot I showed first, second and booster shot separately in a graph with different colors.

For the next step I made two graphs which shows number Covid patient and also number of death people cause of Cov-19 for separated months and years.

For showing the result I made a new data frame which calculate the percentage of covid patient and percentage of death caused by Cov-19.

Result:

In the last two graph, there are lots of fluctuation for the number of people diseased by Covid 19, which can be caused by other reasons such as lockdown, using face mask and limitations during the time.

But the last graph shows number of deaths people which they had diseased decreased significantly which one the reasons are caused by vaccination of the Netherlands people during the time.