2022

First Year Project

COVID-19 AND THE WEATHER

VIGGO GASCOU, EMMA JENSEN, BJØRN SØVAD, KATARINA KRALJEVIC, MIRKA KATUSCÁKOVÁ

1 Introduction

Since COVID-19 became a pandemic the confirmed cases of infected people have fluctuated during the year. During autumn and winter months the cases increased and during spring and summer the cases again fell. This pattern continued to occur during 2020 and 2021. More cases can lead to a rise in the number of hospitalizations and societal costs, so understanding what drives the fluctuations can help shape policy and deploy preventive measures.
This project seeks to analyze the relationship between weather seasonality and covid-cases by finding the meteorological variables that strongest correlate to the confirmed cases. Any correlated variables are held up against the same weather parameter from another country to verify the relationship.
Our research question is therefore formulated as such:

"How are corona and weather conditions correlated in Germany?"

This research question will help provide insight to the German authorities to assess whether there is a basis for leveraging weather conditions to make, change or improve measures for preventing unnecessary cases. If there is a strong relationship to a weather condition the stringency and implementation of restrictions can be capped to the period where a certain weather condition is not as prevalent.

2 Data

We were provided with four datasets/tables to use within this project. The first dataset was issued by our lecturer. This dataset describes the number of new COVID-19 infections per day and the number of new casualties. These were filtered by day and region in Germany for each day in the period from 2020-01-02 to 2021-02-21. The second dataset was provided to us by our lecturer and is from The Weather Company, IBM. It contains information about several weather conditions for each region in Germany, Denmark, Sweden, and Netherlands for each day in the period from 2020-02-13 to 2021-02-21. The third data table was provided to us by our lecturer and describes metadata information about Germany, namely the population of each region. The fourth and final data table that we were provided was a geojson that describes the borders of Germany along with its regional borders. For more detailed descriptions on the two major datasets namely the corona dataset and the weather dataset look in the "*fyp2022p01g06_Notebook.ipynb*" under "Task 0: Data filtering and cleaning". Before beginning to work with our two datasets we had to check for empty values or NaN values and duplicate rows. Luckily neither of our two datasets contained empty values or duplicate rows. We then filtered the weather dataset to only contain values belonging to Germany. We did this by masking the dataset for the ISO code that started with "DE". When doing our initial data inspection, we noticed that the temperature values seemed abnormal, and we would need to change those values. We noticed that it was measured in Kelvin. We dealt with this by converting the temperature to Celsius by subtracting each of the temperature readings by 273.15. We also created two new columns in the corona dataset that contained the number of new COVID-19 cases and casualties per capita i.e., by dividing the number of new cases and casualties by the population of that region. We then merged these two now filtered and cleaned datasets. This resulted in us losing all the COVID-19 data that occurred before 2020-02-13 since we only have weather data from 2020-02-13 to 2021-02-21. It also meant that we lost some weather reports. This could be due to the fact that on the days that the two datasets match one of the datasets did not have a report on that same day in the same region as the other dataset, thus resulting in more lost data.

Finally, we added an external data source from Our World in Data (Our World in Data, 2022) to get the “Stringency Index” for Germany. This index is a measure of how many preventative measures a country has deployed during the period. Even though this index can be found per region the data from OWID is an aggregate value for the entire country of Germany.

3 Results and discussion

3.1 Single variable analysis

With our processed data we wanted to visualize how the different weather variables changed over time. We did this by grouping the data by weeks as the visualization by day was not readable.

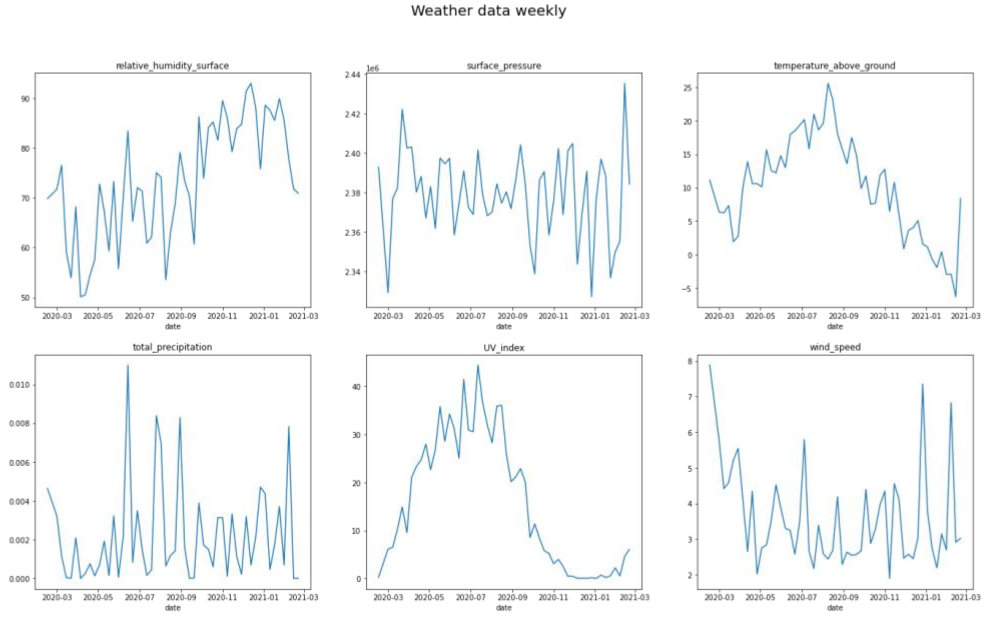

Figure 1: Line-plots of the different weather variables by week

We chose to focus on relative humidity, temperature above ground and UV index, as we can observe on the graphs that the *temperature above ground* and UV index follow each other, whereas relative humidity shows the opposite tendency.

For *Relative Humidity [on] Surface* we can see that there is an increase in the values throughout the year, they reach the highest values in the colder months, and then decrease drastically during March. This decrease is more difficult to notice in the monthly weather data plot, compared to the daily plot where the difference is showcased more clearly.

The graph for *Temperature Above Ground* shows what typical temperatures are for each season. During the warmer months the values keep increasing, reaching their peak in August before decreasing during the colder months to reach the lowest values in February.

UV Index values can be compared to the Temperature Above Ground values in the way that both are higher during warm months and lower during cold ones. The difference would be that UV Index has a steeper line and therefore declines more rapidly to the low values, for example, after the peak that it reaches in August.

3.2 Associations

For multiple variable analysis we use R^2 and Pearson-correlation to evaluate their significance to covid cases. We have a significance threshold of 0.05 as per convention, which is used to determine whether a variable has a significant association with our dependent variable "confirmed_addition". If the variables have a P-value less than this (0.05) we cannot deny that there could be a correlation. Out of 7 independent variables, *"temperature_above_ground"* is the only one that can be excluded, because its P-value is 0.133.

Even though wind speed has the most significant value we chose to focus on UV-index instead, the reason being that we could not see the development on the plot with wind speed over time. Proceeding with the UV Index as our leading variable we visualized the UV-index compared to total confirmed cases.

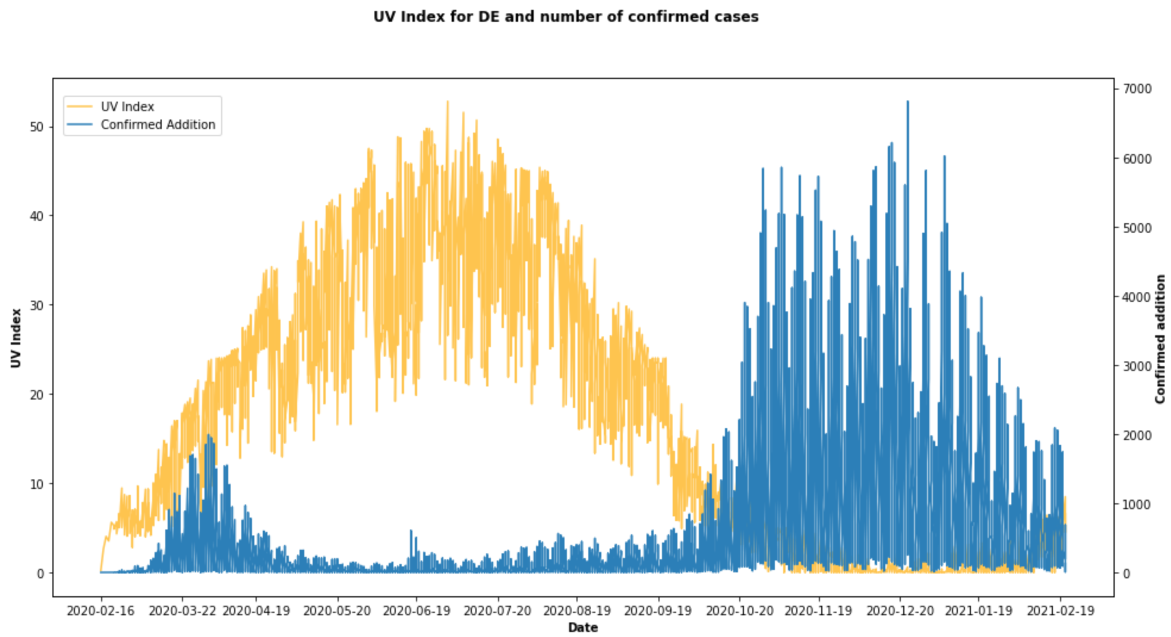

Figure 2: Line-plots for UV Index and Confirmed Cases in Germany by data

Here we can see how the cases are clearly negatively correlated with the UV-index which further suggests that when the UV index rises the cases reversely fall.

One cause for the decrease in cases due to the rise in UV Index has already been addressed by researchers from University of Edinburgh (Cherrie et al., 2021):

"One explanation for the lower number of deaths, which the researchers are following up, is that sunlight exposure causes the skin to release nitric oxide. This may reduce the ability of SARS Coronavirus2 – the cause of Covid-19 – to replicate, as has been found in some lab studies"

Another study suggests D-vitamin as a result of sun exposure. (Mukherje et al., 2021)

However, we cannot say for sure if there is not a confounding variable or indirect cause in play that causes our variables to change like they do. But comparing correlation with another country, Sweden, we can see a similar pattern occurs, where UV index is correlated.

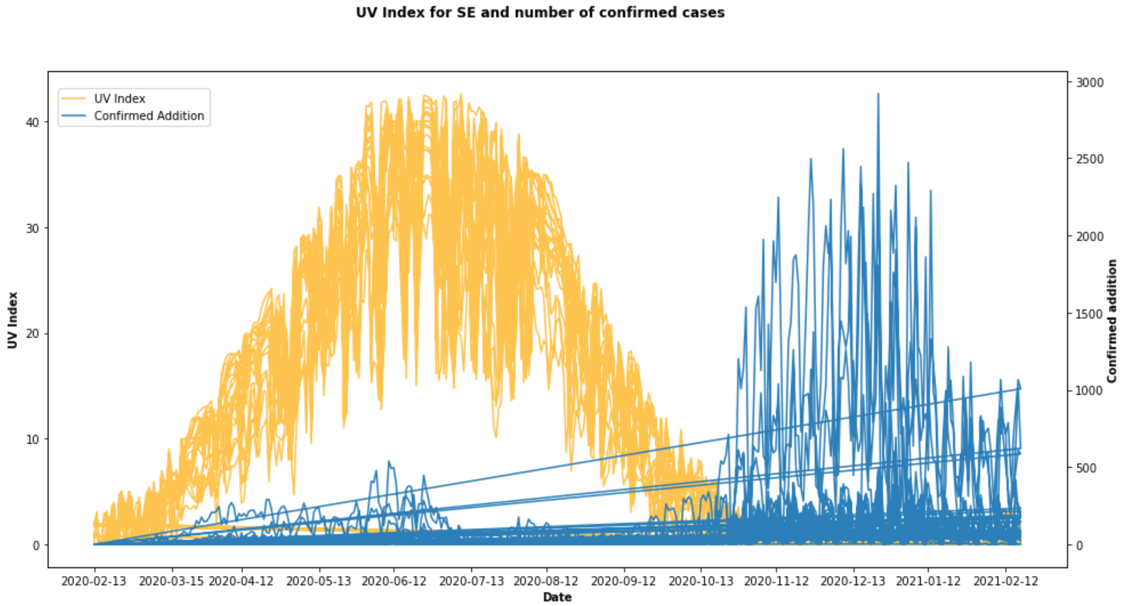

Figure 3: Line-plots for UV Index and Confirmed Cases in Sweden by date

4 Limitations

During this study we discovered some limitations. Firstly, is the fact that we had to work with new Python libraries such as pandas, folium and seaborn. This meant that we had to spend some time to understand and get used to working with these libraries. Another factor was that everyone in our group had little to no experience with GitHub, which also took time to familiarize ourselves with. Another limitation is the time "restriction" that comes from the fact that we had to spend time to familiarize ourselves with the different new Python libraries and GitHub.
With our maps we also wanted to visualize the UV index along with the number of confirmed cases, however after spending multiple hours coding, we did not succeed.

5 Concluding remarks and future works:

5.1 Concluding remarks

In conclusion from our work in both single variable analysis as well as multivariable analysis we can determine that there could be correlation between the UV index and the amount of COVID-19 cases in Germany. Specifically when the UV index increases then the number of COVID-19 cases decreases.

In terms of recommendations to German authorities for implementing restrictions our work shows that the winter months are the ones that need the strictest measures.

5.2 Future work - Cases per capita

For future work it would be an aid for the German authorities in terms of restrictions to also look at population density.

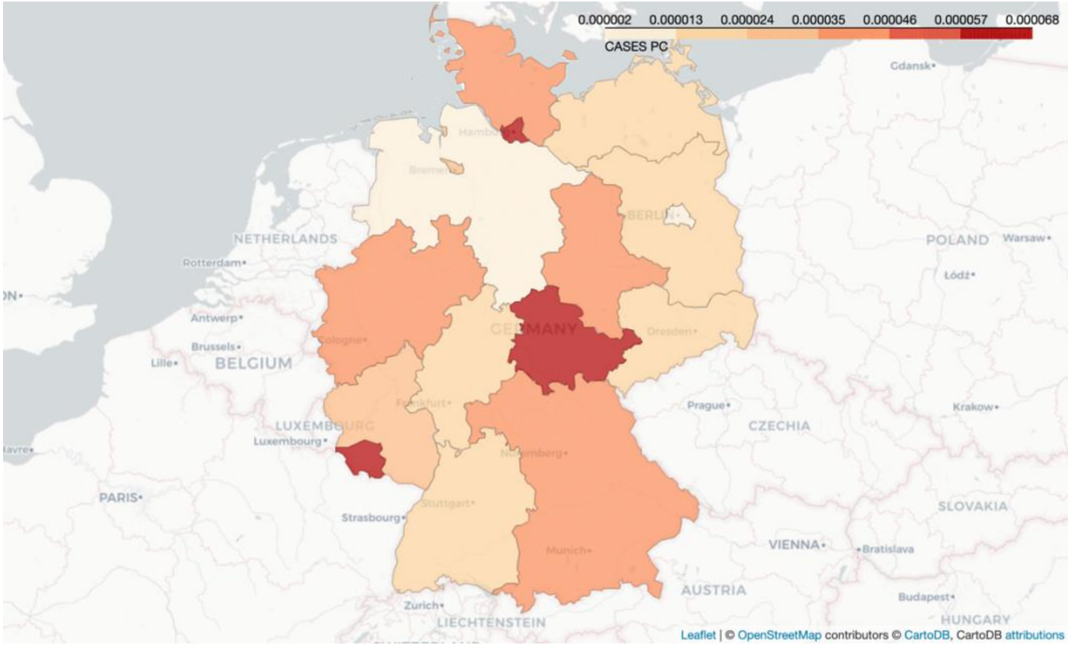

Figure 4: Choropleth map of cases per capita in Germany

On this map we can see the different regions' levels of confirmed cases per capita. The darkest red regions are not the most populated in Germany, and therefore it seems that there's other factors, such as population density and cases pr. capita, that affects the amount of confirmed Covid-19 cases.

6 Disclosure statement

In the gitlog.txt it looks like one person did most of the work/commits. The gitlog.txt does not reflect the fact that we worked together on completing all the tasks in the *"fyp2022p01g06_Report.ipynb"*, this meant that most of the commits were done from one computer.

References

Cherrie, M., & Weller, R. (2021, April 13). *Sunlight linked with lower Covid-19 deaths*. The University of Edinburgh. Retrieved March 3, 2022, from https://www.ed.ac.uk/news/2021/sunlight-linked-with-lower-covid-19-deaths

Gascou, V., Jensen, E., Søvad, B., Kraljevic, K., & Katuscáková, M. (n.d.). fyp2022p01g06_Report [Python code for data analysis]. In *Jupyter Notebook*.

Mukherje, S. B., Gorohovsk, A., Merzo, E., Lev, E., Mukherjee, S., & Frenkel-Morgenstern, M. (2021). Seasonal UV exposure and vitamin D: association with the dynamics of COVID-19 transmission in Europe. *FEBS PRESS*, (Seasonal UV exposure and vitamin D: association with the dynamics of COVID-19 transmission in Europe), 12. https://febs.onlinelibrary.wiley.com/doi/pdfdirect/10.1002/2211-5463.13309

Our World in Data. (2022). *Covid-19 Stringency Index*. Covid-19 Stringency Index. Retrieved February 2022, from https://ourworldindata.org/grapher/covid-stringency-index?tab=chart&country=~DEU