IT UNIVERSITY OF CPH

Covid-19, Weather and Government Response in Denmark

Investigation of which variables correlate with case numbers

First Year Project

Course code: BSFIYEP1KU

March 4th. 2022

Christian Alexander Hetling, *chrhe@itu.dk* Malthe Have Musaeus, *mhmu@itu.dk*
Laurids Robert Holme Pedersen, *lrpe@itu.dk* Christian Weidemann, *weid@itu.dk*
Hubert Wójcik, *huwo@itu.dk*

1 Introduction

Since march of 2020, when the Danish Government recognized the Corona Virus (Covid-19), various lockdowns and other less extreme measures have been enforced. These measures were an attempt to lower cases and lower the stress on hospitals, from hospitalizations due to Covid-19. In this report, we will look at what factors have played a role in the number of cases and hospitalizations of people in Denmark. This analysis will cover danish weather data and later the stringency index. We hope that our findings will provide value to the country by helping them understand the implications weather might bring and if they can do anything to help their citizens.

This leads to us formulating two research questions;

Research Questions

What parts of the weather have a significant correlation with Covid-19 cases in Denmark?

Does the stringency index correlate with Covid-19 cases in Denmark?

2 Data

We used four datasets in this project: Weather data by danish region (The Weather Company, IBM), daily new hospitalizations by region (Statens Serum Institut, 2020), Stringency Index data (Hale, T. et al, 2021) and Covid-19 case data (CSSE JHU, 2020). Figure 1 shows examples of columns in hospitalization, weather and stringency index data. To more accurately model the effect of the variables on Covid-19 hospitalizations we shift the cases by a time delta of 7 and 12 days. In (BCGCHS, 2020) they found the mean time between hospitalization and symptom onset to be around 12 days. Other studies such as (Hersh, E., 2021) pinned the incubation time to be between 5 and 12 days. As a result, we decided to account for the time lag by shifting Covid-19 case data before analysis.

Column	Type
date	string
region_code	string
hospitalized_addition	int

Column	Type
date	string
iso3166-2	string
RelativeHumiditySurface	float
⋮	⋮
WindSpeed	float

Column	Type
CountryName	string
CountryCode	string
Date	int
C1_School closing	float
⋮	⋮
StringencyIndex	float

Fig. 1: Data frame Layouts

Our data contained no missing values. When working with the hospitalization dataset, the region code column was changed to mirror the ISO3166-2 naming convention used in the hospitalization dataset. Because dates had different formats between datasets, they were all formatted to pandas datetime objects, before merging into a dataframe containing all variables.

3 Results and Discussion

3.1 Task 1 - Single variable analysis

We analyze nine distinct weather variables, and see which of these variables correlate with the new confirmed cases of COVID-19 on a daily basis. To initially explore the data and get an understanding we perform a Pearson, Spearman and log-Pearson correlation for each variable. The results are presented in depth in section 3.2.

Five out of the initial nine weather variables were significant in correlating with new confirmed COVID-19 cases. The variables: Surface Pressure, Temperature Above Ground, Total precipitation, UVIndex, and Wind Speed were all significant enough to warrant further attention. To see the distribution of values in the weather variables. We plotted them by date. We found that the variables Total Precipitation and UVIndex had zero values in certain intervals. See figure 2 for a plot visualizing the variables Total Precipitation and UV-Index.

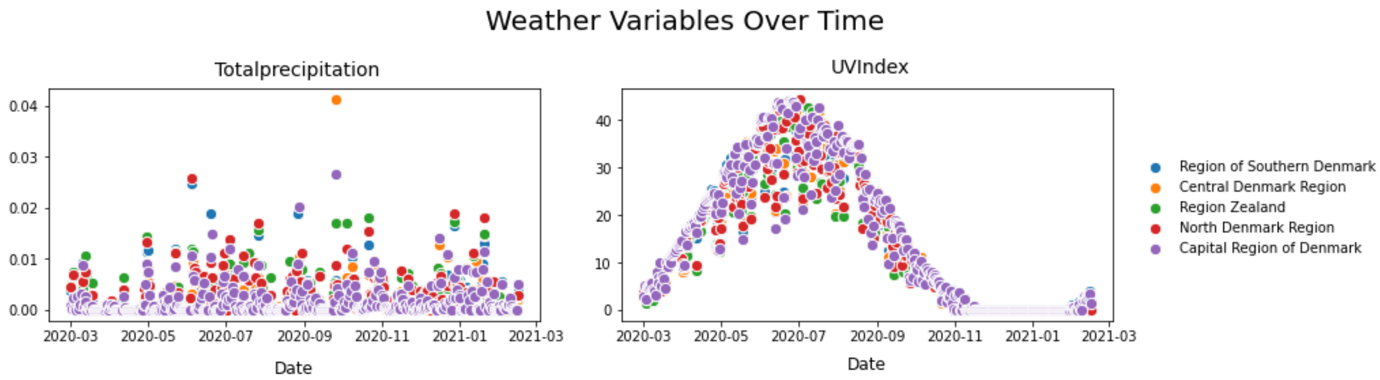

Fig. 2: Totalprecipitation and UVIndex in each region

3.2 Task 2 - Associations

To find associations between certain weather variables and Covid-19 cases in Denmark, we used Pearson, Spearman rank and Log-Pearson correlation. Then, Holm-Bonferroni corrections were implemented.

Each of the tests demonstrated that WindSpeed does not correlate with the number of cases $(p > 0.005)$. When it comes to TotalPrecipitation, only Spearman rank r marked it as 'significant'. TemperatureAboveGround, UVIndex and SurfacePressure all correlate with the number of cases in Denmark. P values for these variables are extremely low for every method. See figure 3 for results of Holm-Bonferroni test.

-	TemperatureAboveGround	UVIndex	Surfacepressure	Totalprecipitation	WindSpeed
Pearson	True	True	True	False	False
Spearman rank r	True	True	False	True	False
Log Pearson r	True	True	True	False	False

Fig. 3: Holm-Bonferroni result table

3.3 Task 3 - Map visualization

Unfortunately, we do not possess data with Denmark's Covid-19 cases per region.
That is why for this task we used the data from (Statens Serum Institut, 2020) with hospitalized cases per region. We have decided to plot three interactive maps using the folium library. The first choropleth shows how many people were hospitalized in every region. The second one is a map of population per region, which helped us with understanding the density of certain parts of Denmark. The last choropleth displays hospitalized people per capita.
As we can observe from the choropleths, Denmark's population is not evenly located. For instance, the capital region has a significantly higher population density as well as hospitalization per capita than the other regions in Denmark. This was an important piece of information for us when deciding to cluster the standard errors by region.

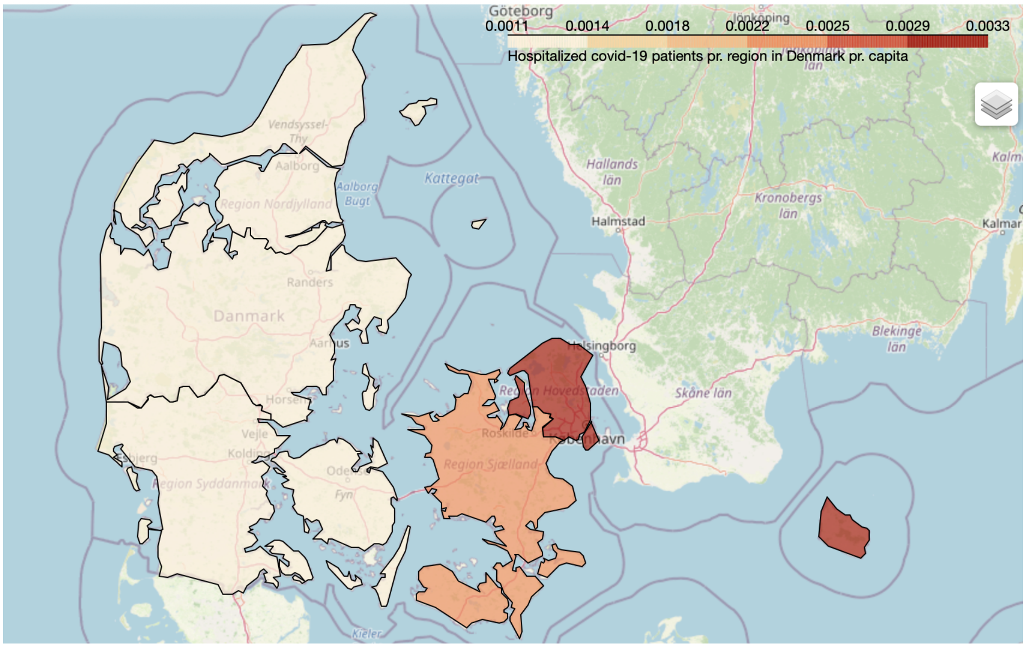

Fig. 4: Hospitalized Covid-19 patients pr. capita; pr. region

3.4 Task 4 - Open question

Weather data, which we have used to research the correlation to COVID-19 cases, is a variable controlled by nature only. Given this fact, we introduced a set of new variables containing data about the danish government interventions under the COVID-19 pandemic. This data is denoted as the Stringency Index [Hale, T. et al. (2021)]. As opposed to the weather data, the stringency index is a factor controlled by human beings, as it reflects the interventions made by the government, in the form of social restrictions.

We computed a new multivariate regression model, with logged COVID-19 cases as the dependent variable and weather and stringent data the as independent variables. Figure 5 displays a list of independent variables each with their respective coefficient, standard error and p-value, derived from the multivariate regression model.

Variable	coefficient	std err	P>|t|
TemperatureAboveGround	0.0474	0.004	<0.001
Totalprecipitation	-21.6088	3.232	< 0.001
UVIndex	-0.0352	0.002	< 0.001
C1_School closing	-0.3636	0.025	< 0.001
C2_Workplace closing	-0.3774	0.054	< 0.001
C3_Cancel public events	1.3269	0.056	< 0.001
C4_Restrictions on gatherings	0.4019	0.024	< 0.001
C5_Close public transport	-0.3375	0.065	< 0.001
C6_Stay at home requirements	0.5860	0.056	< 0.001
C7_Restrictions on internal movement	0.2345	0.046	< 0.001
C8_International travel controls	0.0978	0.033	0.003
H2_Testing policy	-0.4173	0.030	< 0.001
H3_Contact tracing	-12.9176	1.602	< 0.001
H4_Emergency investment in healthcare	7,66E-06	1.75e-09	< 0.001
H6_Facial Coverings	0.9138	0.028	< 0.001
H7_Vaccination policy	-0.2367	0.024	< 0.001
H8_Protection of elderly people	0.3457	0.027	< 0.001

Fig. 5: Significant variables with coefficients, standard errors and p-values.
Variables color coded by sign of coefficient.

4 Limitations

The methods used in this project were purely linear regression. It might be that our data is non-linear which our model simply cannot account for.

The data used in this report is only a one year span and only for one country, so the findings might be hyper specific to Denmark and could also be a result of unusual weather conditions.

In our findings we only see correlations between variables and not causation. Furthermore we cannot tell which variable indicates the other.

5 Concluding Remarks and Future Work

We believe this report could help the danish government with making decisions for the future. It is always beneficial to try to predict the future situation and plan necessary measures to implement.

The three most intriguing variables from our main linear regression model are temperature, total precipitation and UV-index (See fig 5.). We can see that both UV-index and precipitation have a negative correlation with cases and that temperature has a positive correlation with cases.

Other than the weather, we see correlations between government interventions measured by Stringency Index and Covid-19 cases. This could be that government intervention works, but could also be that the government is too slow to react and act according to Covid-19 cases.

Future work on this subject could take advantage of a longer time frame. Our data only spanned one year, a longer time frame could control for an extreme summer or winter. Most of our data is also only of the first variant. An interesting approach could see if the weather correlations changed as different variants emerge.

Another approach would be to compare different countries with Denmark. It could be that the weather is actually not the real reason behind a spike or valley in cases but that different holidays coincide with the weather and result in a fake correlation. Australia would be a prime example for this as their seasons are flipped compared to Denmark.

On the topic of other countries, comparing Denmark's stringency index to that of other countries might also lead to an interesting analysis.

6 Disclosure Statement

Code was written collaboratively using Deepnote, and all group members contributed to the project. Pushes to our repository were made by one person at the end of a day of collaboration, so our gitlog.txt is not representative of the contribution of each group member.

7 References

The Weather Company, IBM. Advancing weather science with accurate forecasting products & technology for businesses everywhere. https://www.ibm.com/weather

Statens Serum Institut. (2020). Arkiv med overvågningsdata for COVID-19. Archived March 3rd 2022 at: https://web.archive.org/web/20201101060302/https://www.ssi.dk/sygdomme-beredskab-og-forskning/sygdomsoverv c/covid19-overvaagning/arkiv-med-overvaagningsdata-for-covid19

Center for Systems Science and Engineering at Johns Hopkins University (CSSE JHU). (2020). COVID-19 Data Repository. https://github.com/CSSEGISandData/COVID-19. Accessed March 3rd 2022 at: https://github.com/owid/covid-19-data/tree/master/public/data

Hale, T. et al. (2021). A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). Nature Human Behaviour. https://doi.org/10.1038/s41562-021-01079-8. Accessed March 3rd 2022 at: https://github.com/OxCGRT/covid-policy-tracker

Hersh, E and Seladi-Schulman, J. (2021). After Exposure to the Coronavirus, How Long Before Symptoms Appear?. Healthline Media. Accessed March 3rd 2022 at: https://www.healthline.com/health/coronavirus-incubation-perio coronavirus-incubation

Belgian Collaborative Group on COVID-19 Hospital Surveillance (BCGCHS). (2020). Time between Symptom Onset, Hospitalisation and Recovery or Death: Statistical Analysis of Belgian COVID-19 Patients. Int J Environ Res Public Health. doi:10.3390/ijerph17207560.