Covid Analysis of Germany

Introduction

Coronavirus (Covid-19) disrupted the modern world like no other event in recent history. There are many unanswered questions – which also includes the origin of the disease and its overall impact both to society and economy across the globe

Corona Virus Data

Several sources who have been updating their websites daily with new numbers of infected cases, where they were found, and the number of deaths, based on different demographics and sections of society, or based on countries and regions. In this report, we have assessed the impact using number of cases across all the states in Germany and total death toll stored in a csv file ranging from Mar 02, 2020 to Feb 08, 2021.

Two data files:

- *deaths-rki-by-state.csv*
- *cases-rki-by-state.csv*

Column-wise we have all the states with the respective daily cases and daily death count, the rightmost columns are a mathematical summation of count across all states on each particular day.

Pre-processing

Step1: The dataset is in csv format, I used pandas' library to convert the data into a pandas' data-frame. So, once I imported pandas' I could read the csv into pandas and further analyse the data and extract patterns from it. Also, libraries such as Matplotlib, DateFormatter and NumPy are also imported to help with further analysis.

- ***Code Snippet***

```
import pandas as pd
import time
from pandas import to_datetime
import matplotlib.pyplot as plt
import matplotlib.dates as mdates
from matplotlib.dates import DateFormatter
import numpy as np
df_cases = pd.read_csv('Data/cases-rki-by-state.csv')
```

- ***Image:*** *Sample screenshot from the pandas' dataframe*

	time_iso8601	DE-BB	DE-BE	DE-BW	DE-BY	DE-HB	DE-HE	DE-HH	DE-MV	DE-NI	DE-NW	DE-RP	DE-SH	DE-SL
0	2020-03-02T17:00:00+0000	2	0	30	34	2	9	2	0	4	116	2	4	3
1	2020-03-03T17:00:00+0000	2	6	54	40	4	13	5	2	10	145	3	5	3
2	2020-03-04T17:00:00+0000	2	9	92	50	4	15	6	4	14	235	7	5	3
3	2020-03-05T17:00:00+0000	4	19	128	69	4	21	11	5	20	332	8	6	3

- Step2: Extract date from the given Datetime from csv file
- ***Code Snippet***

```
df_cases[['Date_xx', 'Time_xx']] = df_cases["time_iso8601"].str.split("T", n = 1, expand = True)
df_cases.drop(['Time_xx', 'time_iso8601'], axis=1, inplace =True) #extract dates from the given information
df_cases['Date_xx'].dtypes #check data type for further analysis
df_cases['Date_xx'] = to_datetime(df_cases['Date_xx']) #convert datetype and create index
df_cases.index= (df_cases['Date_xx'])
```

- Step3: Filtering the data to extract daily cases and death by subtracting each row with its previous value.
- ***Code Snippet***

```
df_cases = df_cases.loc['2020-03-02':'2021-02-01']  #Find daily cases, subtract given row from previous row
df_cases_daily = df_cases.diff()
df_cases_daily.Date_xx = df_cases_daily.index #convertion of date format
df_cases_daily.Date_xx = pd.to_datetime(df_cases_daily.Date_xx)
```

- Step4: Familiarizing with datatypes
- **Code Snippet**

```
df_cases_monthly.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 12 entries, 03 to 02
Data columns (total 17 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   DE-BB      12 non-null     float64
 1   DE-BE      12 non-null     float64
 2   DE-BW      12 non-null     float64
 3   DE-BY      12 non-null     float64
 4   DE-HB      12 non-null     float64
 5   DE-HE      12 non-null     float64
 6   DE-HH      12 non-null     float64
 7   DE-MV      12 non-null     float64
 8   DE-NI      12 non-null     float64
 9   DE-NW      12 non-null     float64
 10  DE-RP      12 non-null     float64
 11  DE-SH      12 non-null     float64
 12  DE-SL      12 non-null     float64
 13  DE-SN      12 non-null     float64
 14  DE-ST      12 non-null     float64
 15  DE-TH      12 non-null     float64
 16  sum_cases  12 non-null     float64
dtypes: float64(17)
memory usage: 1.7+ KB
```

Similar pre-processing steps were performed for the csv file with death count for 16 states for Germany.

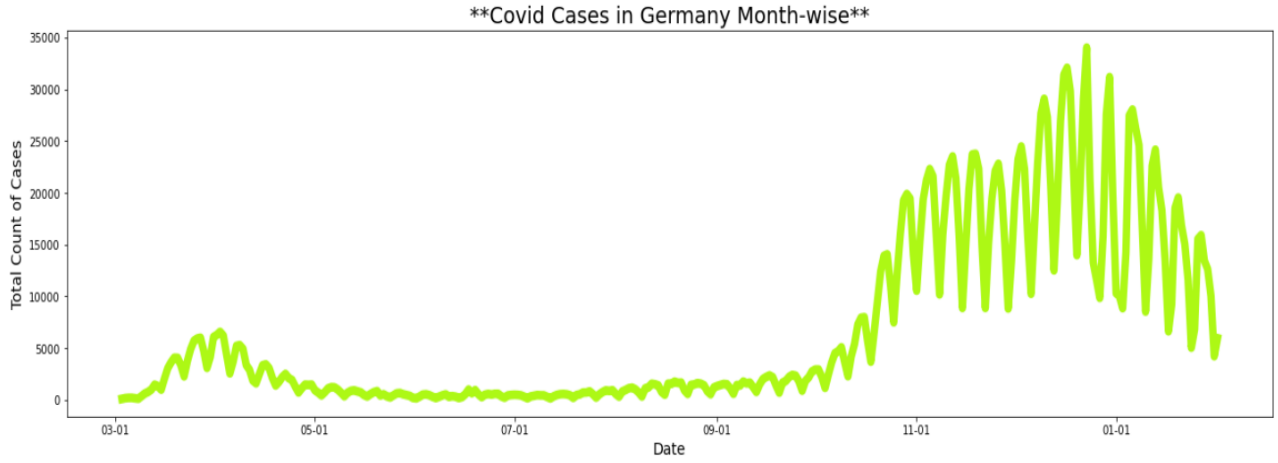
- **How has the number of confirmed cases changed for Germany from begin 2020 - January 2021?**

Discussion There has been a constant increase in cases during the late 2020 and early 2021 in Germany. Through consistent prevention and early case identification, transmission and subsequent cases can bereduced.

Code Snippet

```
fig, ax = plt.subplots(figsize=(22, 6))
ax.plot(df_cases_daily.Date_xx,df_cases_daily['sum_cases'],linewidth=7.5,color='#adf815')
ax.set_title('v = 1',fontweight="bold", size=20) # Title
ax.set_xlabel("Date", fontsize = 14.0)
ax.set_ylabel("Cases", fontsize = 14.0)
ax.set_title("**Covid Cases in Germany Month-wise**", fontsize = 20.0)
date_form = DateFormatter("%m-%d")
ax.xaxis.set_major_formatter(date_form)
```

- ***Image:*** *Screenshot depicting cases count in Germany from beginning of 2020 to January 2020*

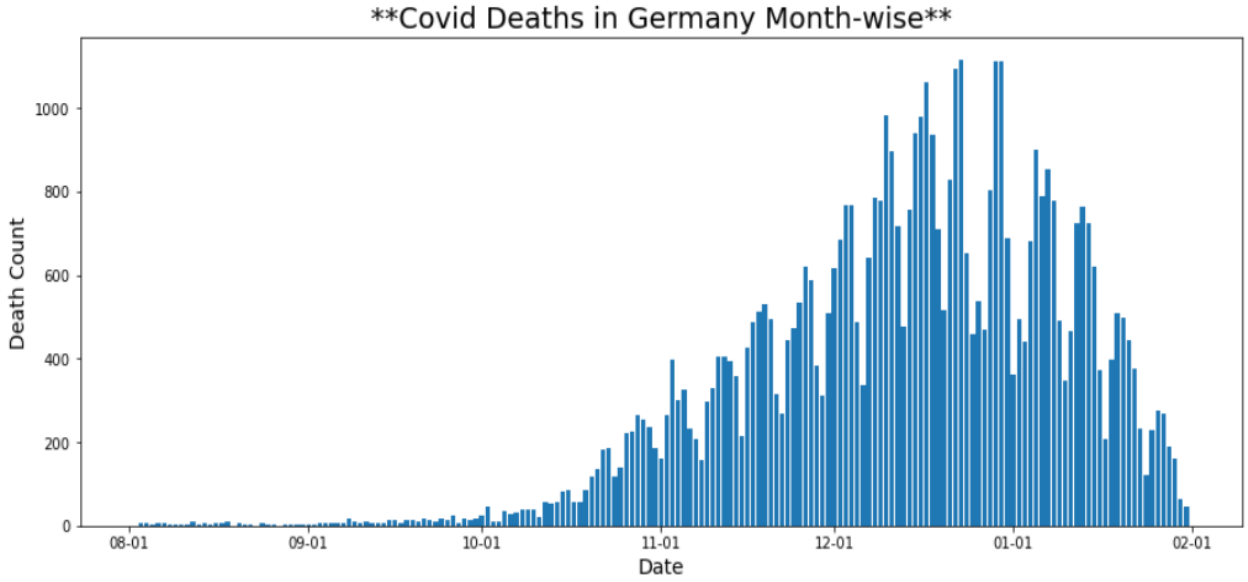


- **How has the number of deaths has increased or decreased in the last six months (from August 2020 to January 2021?**

Discussion Germany has recorded more than 10,000 coronavirus deaths in December. Mathematically speaking the numbers have doubled and have been increasing in an exponential pattern. Almost 50000 death has been recorded from Oct 2020 – Feb 2021. It can inferred from the data that the virus has been spreading in waves and the datat for Oct 2020 – Feb 2021 have proved to be very impactful taking away too many lives.

- **Code Snippet**

```
#Truncate the date range as per requirement
df_death = df_death.loc['2020-08-01':'2021-01-31']
df_death_daily = df_death.diff()
df_death_daily.Date_ = df_death_daily.index
fig, ax = plt.subplots(figsize=(22, 6))
ax.plot(df_death_daily.Date_, df_death_daily['sum_deaths'], linewidth=7.5, color='#adf815', markerfacecolor='blue', markersize=12)
ax.set_title('v = 1',fontweight="bold", size=20) # Title
ax.set_xlabel("Date", fontsize = 14.0)
ax.set_ylabel("Death Count", fontsize = 14.0)
ax.set_title("**Covid Deaths in Germany Month-wise**", fontsize = 20.0)
date_form = DateFormatter("%m-%d")
ax.xaxis.set_major_formatter(date_form)
```



- **How is the situation (in terms of new cases and death) in Sachsen-Anhalt compare to that in some of the hotspots in Germany? You may want to compare with a certain state in a certain time frame.**

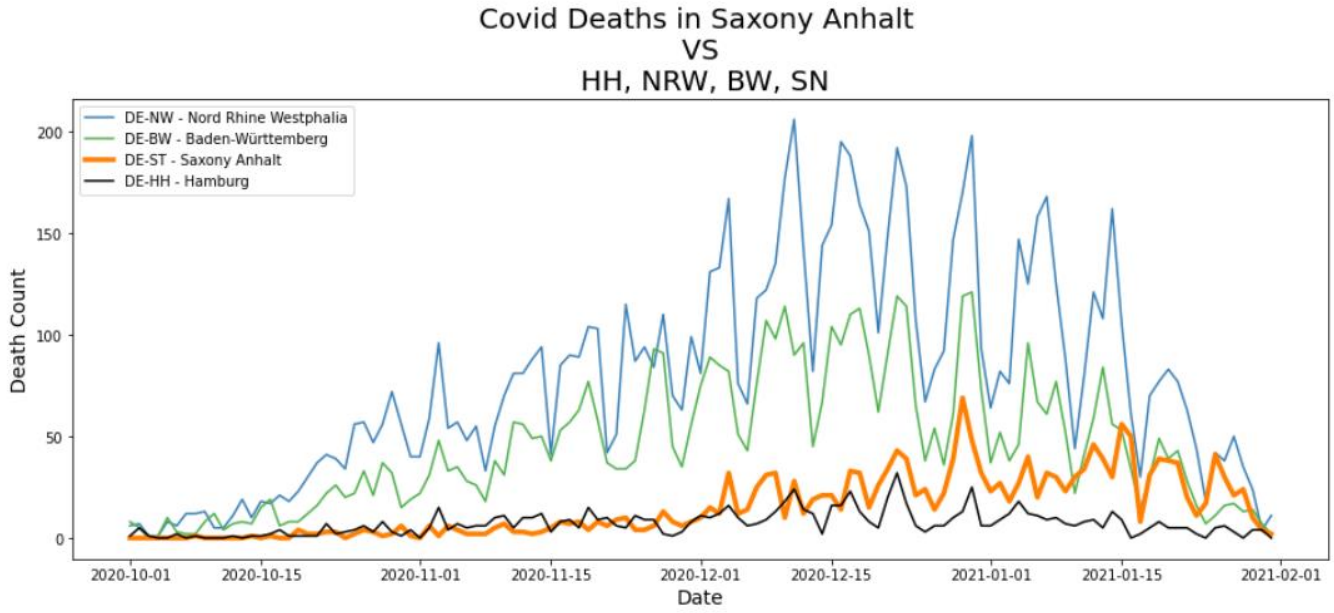
Discussion I have tried to compare the death count and covid cases between Saxony Anhalt and NRW, Baden Wurttemberg, and Hamburg. ***The idea was to include states from four sides of Germany and then compare with Saxony Anhalt from* Oct 2020 to 31 Jan 2021**. The death toll in Nord Rhine Westphalia has been significantly higher compared to other states. **Saxony Anhalt is highlighted in bold orange color which is relatively low compared to other states during this time period.**

A similar pattern can be observed for number of covid cases between Saxony Anhalt and other states.

- **Code Snippet**

```
df_Q3_Death = df_death_daily[['DE-HH', 'DE-NW', 'DE-BW', 'DE-SN', 'DE-ST', 'Date_']]
df_Q3_Death = df_Q3_Death.loc['2020-10-01':'2021-01-31']
fig, ax = plt.subplots(figsize=(16, 6))
ax.plot(df_Q3_Death.Date_,df_Q3_Death['DE-NW'], linewidth=1.5, color='#377eb8', label='DE-NW - Nord Rhine Westphalia')
ax.plot(df_Q3_Death.Date_,df_Q3_Death['DE-BW'],linewidth=1.5,color='#4daf4a',label='DE-BW Baden-Württemberg')
ax.plot(df_Q3_Death.Date_, df_Q3_Death['DE-ST'], linewidth=3.5, color='#ff7f00', label='DE-ST - Saxony Anhalt')
ax.plot(df_Q3_Death.Date_,df_Q3_Death['DE-HH'],linewidth=1.5,color='black', label='DE-HH - Hamburg')
ax.legend(loc='upper left')
```

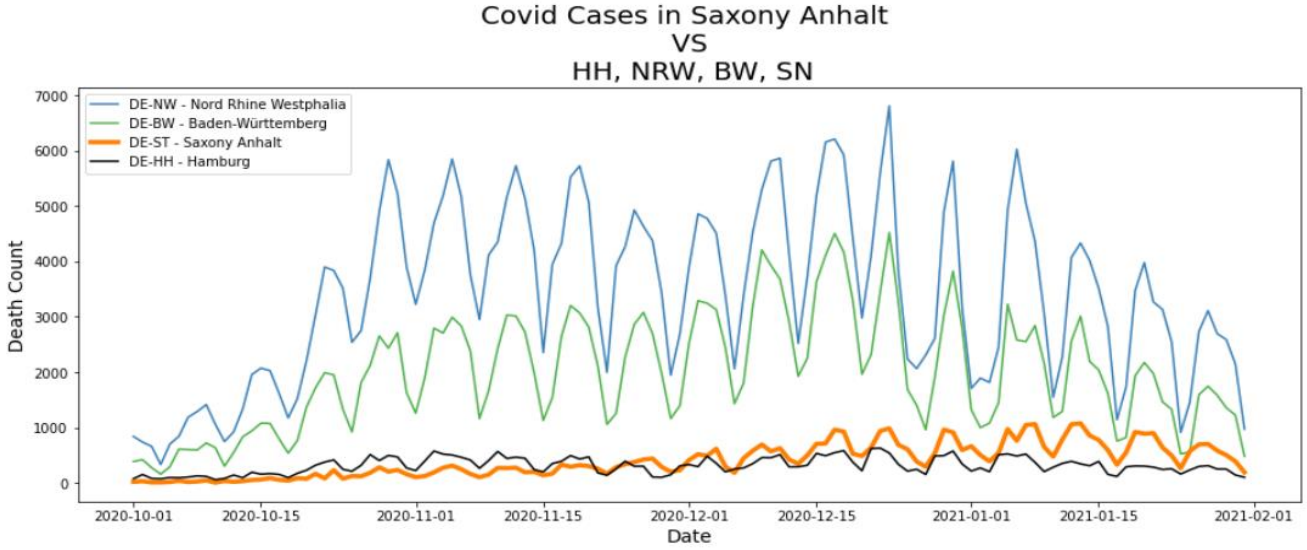
```
ax.set_title('v = 1',fontweight="bold", size=20) # Title
ax.set_xlabel("Date", fontsize = 14.0)
ax.set_ylabel("Death Count", fontsize = 14.0)
ax.set_title("Covid Deaths in Saxony Anhalt \n VS \n HH, NRW, BW, SN", fontsize = 20.0)
```



- **Code Snippet**

```
df_Q3_Cases = df_cases_daily[['DE-HH', 'DE-NW', 'DE-BW', 'DE-SN', 'DE-ST', 'Date_xx']]
df_Q3_Cases = df_Q3_Cases.loc['2020-10-01':'2021-01-31']
```

Similar code is used to plot Covid cases in Saxony Anhalt vs HH, NRW, BW, SN.



- **Could you relate the sudden spike or decline in the data to certain events such as a festival celebration, the introduction of mask wearing, et cetera?**

Situation 01: Sep21st 2020; New Test Strategy implemented with improved test quality and increased number of tests per day

Discussion While the absolute figures for cases have been increasing exponentially. I have attempted to highlight the fact that increased number of tests actually help us track the numbers better, the red color of the bar graph depicts an exponential increase of rise in cases. Most cluster of cases reported were in areas in the context of cultural gathering, celebrations with family and friends which further contributed to the transmission.

I have attempted to highlight the necessary steps taken by the government to reduce the impact. Sep21, 2020 was a significant date where daily number of tests were increased and a better testing strategy was delployed. We can observe the impact in the graph below. **The testing strategy helped us to understand the real situation of cases which could have gone unnoticed otherwise.**

- **Code Snippet**

```
df_Q4_situation01_cases = df_cases_daily.loc['2020-04-02':'2021-03-31']
```

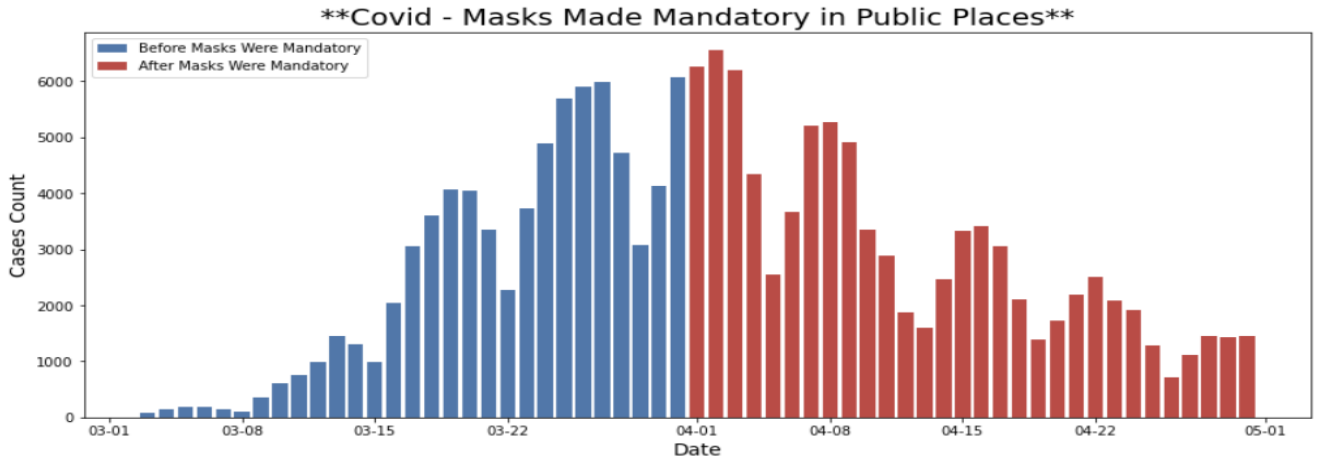
```
df_Q4_situation01_cases_a = df_Q4_situation01_cases.loc['2020-07-01':'2020-09-20']
df_Q4_situation01_cases_b = df_Q4_situation01_cases.loc['2020-09-21':'2020-10-31']
fig, ax = plt.subplots(figsize=(15, 6))
ax.bar(df_Q4_situation01_cases_a.Date_xx, df_Q4_situation01_cases_a['sum_cases'], color = '#5177a9')
ax.bar(df_Q4_situation01_cases_b.Date_xx, df_Q4_situation01_cases_b['sum_cases'], color = '#b94c46')
ax.set_title('v = 1',fontweight="bold", size=20) # Title
ax.set_xlabel("Date", fontsize = 14.0)
ax.set_ylabel("Cases Count", fontsize = 14.0)
ax.set_title("Covid - Testing Strategy Changed - Improved and Increased Testing", fontsize = 20.0)
date_form = DateFormatter("%m-%d")
ax.xaxis.set_major_formatter(date_form)
```

Situation 02: Apr1st 2020-Apr10 2020; New Test Strategy implemented with improved test quality and increased number of tests per day

Discussion Here, we have attempted to analyse the change in number of cases after wearing masks were mandatory by the government officials in Germnay. The blue bars highlight the patterns before masks were introduced. The red bar in the graph highlights the drop in daily cases in Germany. A **decrease of transmission within the population can be noticed**, the rule took approximately 10 days to be implemented and to make general public aware of the benefits of covering their mouth and nose in public places.

- **Code-Snippet**

```
df_Q4_situation02_cases_a = df_cases_daily.loc['2020-03-02':'2020-03-31']
df_Q4_situation02_cases_b = df_cases_daily.loc['2020-04-01':'2020-04-30']
fig, ax = plt.subplots(figsize=(15, 6))
ax.bar(df_Q4_situation02_cases_a.Date_xx, df_Q4_situation02_cases_a['sum_cases'], color = '#5177a9')
ax.bar(df_Q4_situation02_cases_b.Date_xx, df_Q4_situation02_cases_b['sum_cases'], color = '#b94c46')
ax.set_title('v = 1',fontweight="bold", size=20) # Title
ax.set_xlabel("Date", fontsize = 14.0)
ax.set_ylabel("Cases Count", fontsize = 14.0)
ax.set_title("**Covid - Masks Made Mandatory in Public Places**", fontsize = 20.0)
date_form = DateFormatter("%m-%d")
ax.xaxis.set_major_formatter(date_form)
```



- **Perhaps you can discuss any other interesting observation from these data with a plot or a table**.

Discussion As the number of infected persons is high in Germany, this means that the daily number of newly infected persons was high during the timeperiod Mar 2020- Feb 2021 hence the average color of heatmap is > 100000 in each state. For the state of Thuringia Dec 2020-Jan 2021 were very grave where number of cases is highest.

- **Code Snippet**

```
df_cases_monthly=df_cases_daily.groupby(df_cases_daily
['Date_nameChanged'].dt.strftime('%m'), sort=False).sum()
#Group the data to extract
statewise cumulative data
df_cases_monthlyArray = df_cases_monthly.to_numpy()
#Conversion to numpy array
df_cases_monthlyarrayTranspose =
df_cases_monthlyArray.transpose() #transpose the array
fig, ax = plt.subplots(figsize = (5, 6))
im = ax.imshow(df_cases_monthlyarrayTranspose, cmap =
"Blues")
plt.colorbar(im)
ax.set_xticks(np.arange(len(monYear)))
ax.set_yticks(np.arange(len(columns)))
ax.set_xticklabels(monYear)
ax.set_yticklabels(columns)
plt.setp(ax.get_xticklabels(), rotation = 45, ha = "right",
rotation_mode ="anchor")
fig.tight_layout()
```

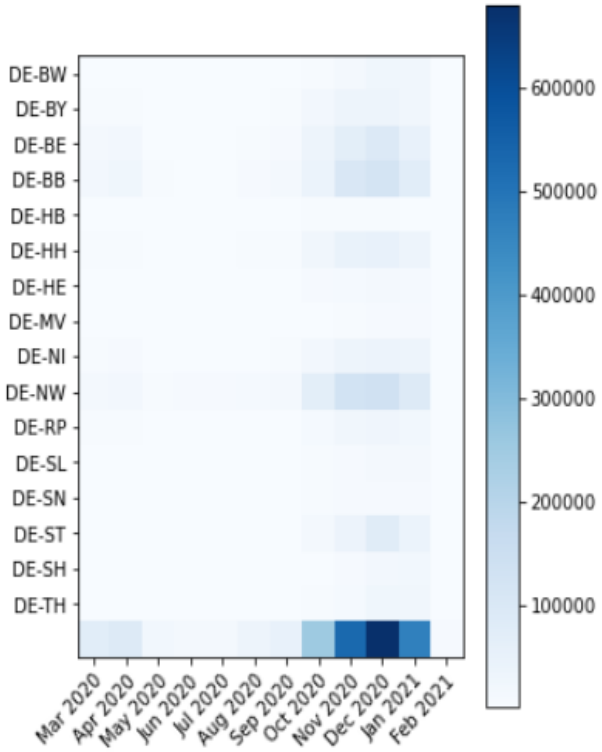


Image: *Heatmap of total cases across all 16 states in Germany during Mar2020-Feb 2021*

- **Conclusion** It remains essential that people follow the rules of physical distancing and hygiene and avoid crowds, and that anyone who develops symptoms compatible with COVID-19 be tested immediately.

- **References**

https://wiki.unece.org/display/CTRBSBC/Germany
https://www.dw.com/en/germany-monthly-covid-death-toll-set-to-double-in-december/a-56000364
https://www.dw.com/en/coronavirus-germanys-new-face-mask-regulations-explained/a-53260732
https://www.worldometers.info/coronavirus/