

Determining the preparedness of the German Government and Medical Authorities in handling COVID-19 Crisis

Dr. D Narayana,

Professor, AIML, Great Learning
darapaneni@gmail.com,

Aranya Das

Student, PGP-AIML, Great Learning
aranyads2018@gmail.com

Varun Sivasubramanian

Student, PGP-AIML, Great Learning
varunsiva95@gmail.com

Debajyoti Das

Student, PGP-AIML, Great Learning
debajyoti.das.bookworm@gmail.com

Abstract

Purpose: In this paper, we try and analyse the trends in the spread of COVID-19 in the German population. We look into the statistical significance of the lockdown enforced by the German authorities in curbing the spread of the virus across the German population. We also try to develop a forecasting model to predict the number of infected patients & fatalities arising out of the infection. **Materials & Methods:** As part of this study, we obtained the dataset from 2 sources, namely Kaggle & Institute of Health Metrics & Evaluation (IHME). The datasets had data like number of cases, age-group of patient, number of available beds, number of tests done on a given day across the 2 datasets & the datasets were merged based different criterion like State, County, Date of Infection, etc. The dataset had data from 24-Jan-2020 up till 3-June-2020, and the same was obtained from across the 16 German states & their individual counties. **Results:** Statistical methods like t-Test & ANOVA resulted in extremely low p-values, based on which we can say with 99% confidence level that the different measures undertaken by German authorities, especially the imposition of a nationwide lockdown, had a statistically significant impact on controlling the spread of the SARS-CoV-2 virus in Germany. Visually we saw that the SARS-CoV-2 virus spread rose significantly in the pre-lockdown period (which we have assumed to be till 07-April-2020) and then started slowly tapering off. This also helped avoid severe stress on the medical institutions & we once again saw visually that the hospital admission rate was always below the rate at which new hospital beds were being added daily. This was also vindicated in our time-series forecasting model using Prophet, an open-source forecasting library from Facebook. **Conclusion:** Based on the above, we could statistically prove that the timely measures and steps taken by German authorities, like imposition of the lockdown, helped in both controlling the spread of COVID-19 across Germany, as well as kept the rate of fatalities due to the same at a relatively low rate.

1. Introduction:

The COVID-19 pandemic, also known as the Coronavirus Pandemic, is an ongoing pandemic of coronavirus disease 2019 (COVID-19), caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) ^[1]. The outbreak was first detected in Wuhan city, which is the capital of Hubei province in China, in December 2019 ^[2]. The World Health Organization (WHO) declared the outbreak a Public Health Emergency of International Concern on 30-January-2020, and a pandemic on 11-March-2020 ^{[3][4]}. As of 14-June-2020, more than 7.76 million cases of COVID-19 have been reported globally across 188 countries and territories, resulting in more than 429,000 deaths. At the same time, more than 3.68 million people have also recovered ^[5].

Authorities across the globe have responded by implementing one or more of the below measures, in an effort to curb the spread of the virus via, travel restrictions, lockdowns, workplace hazard controls and facility closures. Many places have also worked to increase testing capacity and contact tracing of infected persons.

Germany reported its first case on 27-January-2020, near Munich, Bavaria. Germany's disease and epidemic control is advised by the Robert Koch Institute (RKI) according to a national pandemic plan, which describes the responsibilities and measures of the healthcare system actors in case of a huge epidemic.

The outbreaks were first managed in a Containment stage ^[6], under the aforesaid plan, which targeted minimizing the expansion of clusters. The German government did not initially implement special measures to stockpile medical supplies or limit public freedom. However, since 13-March-2020, the pandemic has been managed in the Protection stage ^[6] as per the above RKI plan, with states mandating school and kindergarten closures, postponing academic semesters and prohibiting visits to nursing homes to protect the elderly. Two

days later, on 15-March-2020; borders to Austria, Denmark, France, Luxembourg and Switzerland were closed. By 22-March-2020, curfews were imposed in six German states while other states prohibited physical contact with more than one person who resides outside of a household. Complete restriction of movement started from 17-March-2020 & by 27-March-2020 it was imposed across all the states

We look at various statistical methods like t-Test [7] & ANOVA [8], to determine if the different measures undertaken by German authorities had had any statistically significant impact on the control of spread of SARS-CoV-2. Specifically, we look at the Student's t-Test [8], which is an inferential statistical technique useful for comparison of means of two groups. We have separated our dataset in 2 sets, viz-a-vis one set containing the *Infection Rates* (which is the *ratio of the total number of confirmed cases in a given day & the total number of tests conducted in the same day*) before initiation of lockdown and another set containing *Infection Rates* post lockdown. Here, using the t-Test analysis we will ascertain whether the lockdown was statistically significant in having an impact on curtailing the spread of the infection. We have also used the ANOVA test, to verify the same.

Furthermore, looking at the distribution of infection, recovery, deaths, availability of beds, admissions to hospital, availability of ICUs & ventilators etc. over time, we can give visual projections, that can help in taking informed decisions.

We also build a forecasting model using Facebook's open-sourced Prophet library [9], using which we will try to predict the number of infected cases in a day as knowing this number beforehand can enable a government to be adequately prepared with necessary resources to tackle the pandemic. Prophet looks to solve the problem of "forecasting at scale" & builds a decomposable time-series prediction model which can handle non-linear trends by fitting yearly, weekly, and daily seasonality, with additional holiday effects [10][11]. We used Prophet for its robust handling of missing data, shifts in trend & typically excellent handling of outliers. At its core Prophet uses an additive regression model [12] which has 4 main components: 1. A piecewise linear logistic growth curve trend 2. A yearly seasonal component modelled using Fourier series. 3. A weekly seasonal component using dummy variables, and 4. A user-provided list of important holidays. [11]

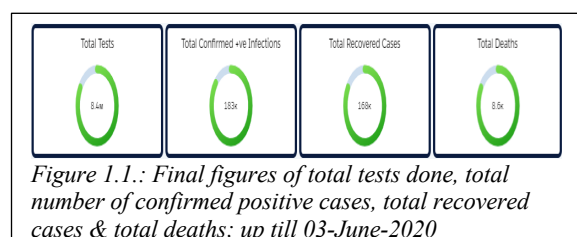
2. Materials & Methods:

The COVID-19 data for Germany is being collected by the Robert Koch Institute [13] and can be

downloaded through the National Platform for Geographic Data (which also hosts an interactive dashboard) [14]. The earliest recorded cases are from 24-January-2020. For our analysis, we obtained one part of our research dataset from *Kaggle*, the link to which is given in the references section [15]. This dataset, *covid_de.csv*, comes with an associated blog, which explains the data extraction & beautification methodologies [16]. The 2nd part of the dataset came from *Institute of Health Metrics & Evaluation (IHME)* [17]. All the data have been extracted from the above 2 open data sources and we are grateful to all of the acknowledged.

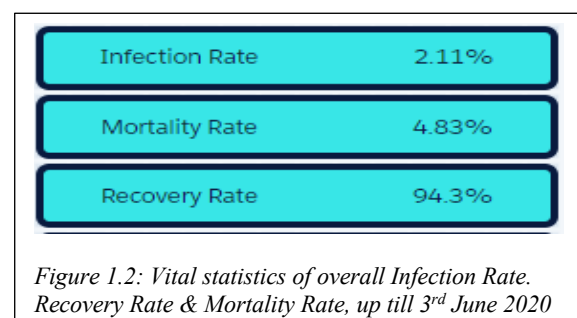
These 2 datasets, contained information about: State, County, Age-Group, Gender, Date, Cases, Deaths, Recovered, Number of available beds on a given date, Number of available ICUs on a given date, Number of available Ventilators on a given date, Total Number of tests done till date, etc. This information was spread across 2 data files and had to be merged based on the State, County & Date. Thus we had with us cumulative number of cases being reported in a county, in a state, on a given date. For the same date, we also had the cumulative number of recovered patients, the cumulative number of expired patients, the cumulative number of tests being done; again for a given county, in a given state. This data was segregated based on the gender & age of the patients.

In Germany, from 28-January-2020 to 12:09pm CEST, 30-July-2020, there have been 206,926 confirmed cases of COVID-19 with 9,128 deaths [18]. *Figure 1.1* shows the different important figures in Germany's battle against COVID-19 for our observed period, i.e. up to 03-



June-2020.

We can also draw some vital statistics from the above data, which is highlighted in *Figure 1.2*.

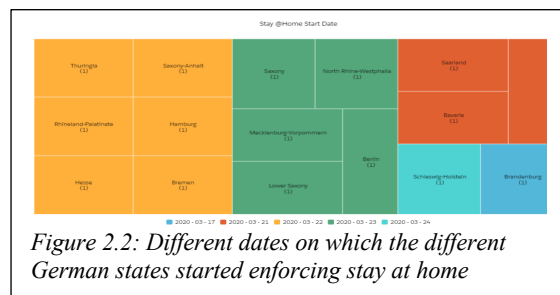
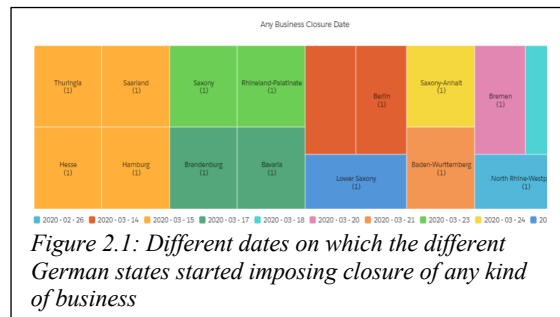


Germany with a COVID-19 fatality rate at roughly 5% has one of the lowest mortality rates amidst some of the other First-World Countries like UK (~15%), France (~13.5%), Spain (~10%), and USA (~3%). It has a record equivalent to that of China (~5%), where the virus was first reported [5].

Although we had data available from 28-January-2020, on close inspection we observed that the total number of infected people were only 54 by 03-March-2020. We thus decided to restrict our analysis to a 3-month time period of 03-March-2020 to 03-June-2020 for our analysis.

As we have already discussed, there was staggered imposition of restriction of movement across different German states & counties, wherein since 13-March-2020, the pandemic was being managed in the PROTECTION STAGE as per the RKI plan [6], which is national pandemic plan prepared by the Robert Koch Institute (RKI). Two days later, on 15-March-2020; borders to Austria, Denmark, France, Luxembourg and Switzerland were closed. By 22-March-2020, curfews were imposed in six German states while other states prohibited physical contact with more than one person who resides outside of a household.

Below we look at the different dates on which the various restrictions started coming in place across the different German districts.



The above figures, 2.1 & 2.2, are presented in chronological order & hence depict the staggered fashion in which the various restrictions were imposed across the different German states. Thus, as we can see, North Rhine-Westphalia was the first to start any kind of restrictions, with it imposing restrictions on any kind of businesses on 26-

February-2020. Slowly other states also started following suite, with Lower Saxony being the last to impose these restrictions on 27-March-2020. Gradually this progressed towards other forms of restrictions also like *Closure of Educational Facilities*, *Closure of Non-Essential Business*, *Gatherings* & finally *Complete Lockdown*. Complete restriction of movement started from 17-March-2020 & by 27-March-2020 it was imposed across all the states. To arrive at a pre & post-lockdown period, we added 11 days (the maximum incubation period - for up to 97.5% of patients - for COVID-19 is about 11.5 days [21]) to the final lockdown date (27-March-2020) and arrived a cut-off date of 07-April-2020. We will test our hypothesis around this date.

A pre-cursor to hypothesis testing is to understand the distribution of the dataset. Statistical tests like t-Test [19], expect the dataset to have a normal distribution. When such a distribution exists, we can identify the probability of a particular outcome. This is especially true for t-Test, which aims to statistically check whether the means of 2 distributions are different. Another statistical test that we will look at is *One-Way Analysis of Variance* (ANOVA) [21], which comes into picture if the independent variable has multiple levels. In our experiment, the independent variable is the *date* feature, which we have segregated into 2 levels (pre & post-lockdown). The dependent variable is the *Infection Rate*. We thus perform t-Test to understand if there is any statistically significant difference between the *means of Infection Rate* before and after 07-April-2020. We set our acceptable confidence level at 99%, i.e. we will reject the Null Hypothesis - "*Mean Infection Rate Remained the Same Before & After Lockdown*" - if the p-value of the statistical test is greater than 0.01.

The datasets upon visual analysis seemed to be *right-skewed*, which was expected since the entire dataset in itself was right-skewed. Nonetheless, we utilised in-built scikit-learn libraries like *skew* to bring out this skewness in numbers & our visual analysis was confirmed statistically. Thus, the datasets had to be normalized before we proceeded with our t-Test hypothesis testing. As the datasets were *right-skewed*, we transformed the datasets using the lower powers in the *ladder of powers* [20], i.e. powers like square-root, cube-root, logarithmic, etc.

3. Results

3.1. Exploratory Data Analysis

From the dataset, we could visualize the different insights which we have illustrated in the below figures. In Figure 3.1 below, we see the distribution of the 3 components *Daily Sum of Confirmed Infections*, *Daily Sum of Recovered Patients* &

Daily Sum of Deceased Patients; across the timeframe from 03-March-2020 to 23-May-2020. As can be seen, the number of daily confirmed cases reaches a peak around 02-April-2020 & then onwards slowly starts to taper-off. A similar trend is observed in the daily number of recovered cases also, with the number of recoveries always closely following the number of confirmed infections. This makes us realise that different measures taken by the medical authorities across Germany always kept the number of infections within a controllable & treatable range.

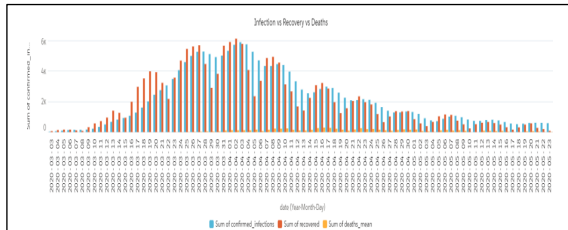


Figure 3.1. : The above figure shows the distribution of 3 components, Daily Sum of Confirmed Infections, Daily Sum of Recovered Patients & Daily Sum of Deceased Patients; across the time-frame from 3rd

This is further confirmed in *Figure 3. 2* wherein we see that while the German medical authorities were slow to start adding dedicated facilities for COVID positive patients, they ramped up the effort pretty quickly, and as the beds were being continuously added, the number of daily cases was progressively falling.

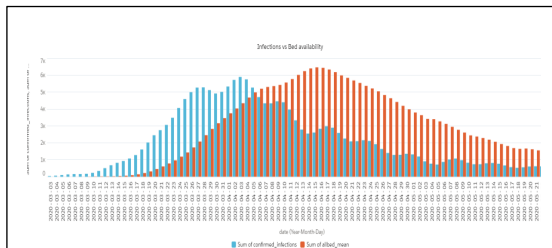


Figure 3.2: Distribution of daily number of confirmed infections & daily number of dedicated beds being added

In *Figure 3.3* we see that although the number of reported positive cases progressively rose & then started falling, the number of hospital admissions

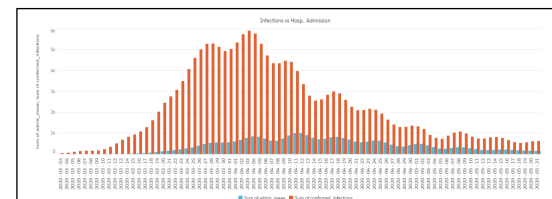


Figure 3.3: Distribution of cumulative number of daily hospital admissions due to SARS-CoV-2 infection & daily number of confirmed COVID positive patients.

was always lower & within manageable limits. The number of cases started to fall after 04-April-2020, around a week after the country went into a strict domestic confinement across all states.

In the next figure, *Figure 3.4*, we can see that the authorities always had more than sufficient number of hospital beds (both general & ICU) as well as ventilators; than the actual number of admissions on any given date. As this number was always lower than the available number of ICUs, the medical facilities were never over-burdened. Additionally, daily new beds being added also accounted for any unforeseen surge in the daily number of admissions. This allowed the hospitals to provide better care for the admitted patients, probably with more severe symptoms, and thus in turn kept the mortality rate at a relatively low rate & count.

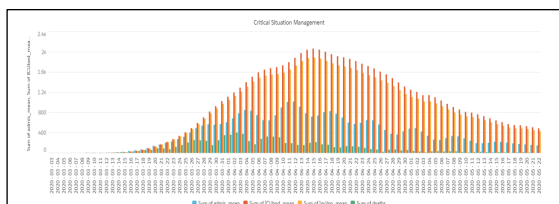


Figure 3.4: Distribution of daily number of hospital admissions, available ICU beds, Ventilators & Number of Daily Deaths

3.2. Hypothesis Testing

The *Infection Rates* centred around 07-April-2020 showed a skewness of 1.22 & 3.25 for the pre & post-lockdown datasets respectively. *Figure 4.1* shows this distribution visually.

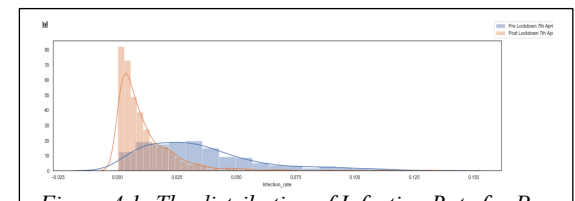


Figure 4.1: The distribution of Infection Rate for Pre & Post-Lockdown periods, before applying any transformation

A cube-root transformation brought these to -0.1 & 0.53 respectively. Similar attempts were made with square-root, which resulted in 0.28, 1.07 respectively & 4th root which resulted in -0.31, 0.28 respectively. We observed that cube-root transformation resulted in best transformation, with the overall skewness across both the datasets being removed equally well both the datasets centred better. This is also confirmed in *Figure 4.2* below.

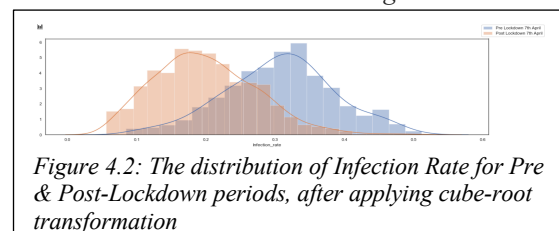


Figure 4.2: The distribution of Infection Rate for Pre & Post-Lockdown periods, after applying cube-root transformation

The test statistic of the t-Test is a *t-value*, which is conceptually an extension of *z-scores*, wherein *t-values* represent how many standard units the mean of the 2 groups are apart. Our experiment showed the same to be 25.89. The corresponding p-value was 2.85e-121.

Similar experimentation with One-way ANOVA produced ANOVA statistics of 670.45. The corresponding p-value was 2.85e-121.

3.3. Forecasting

In this section, we estimate the magnitude of cases that were reduced due to the lockdown enforced on 27-March-2020. By forecasting cases using the pre-lockdown trend, we can extrapolate the number of cases that could have occurred, had that same trend continued. We must also keep in mind that forecasting using the pre-lockdown trend is the best-case scenario, since it is possible that the rate of spread of infections could have possibly increased without the lockdown.

To model this, we consider the cumulative sum of cases on each date from 03-March-2020 till 07-April-2020 as the input to the forecasting model. The reason for considering 07-April-2020, though the lockdown was enforced on 27-March-2020, is due to the fact that the maximum incubation period (for up to 97.5% of patients) for COVID-19 is about 11.5 days^[22]. The model is then able to forecast the cumulative sum of cases for the next 57 days (3-June-2020 - which is the end of our analysis period - is 57 days away from 07-April-2020) using a 99% confidence interval.

We then plot the curves for forecasted cases and actual cases to visualize the difference in slopes, post lockdown. As shown in *Figure 5.1*, we can see that the curve for actual cases starts to flatten around the 07-April-2020 and the cumulative sum of actual cases on the 03-June-2020 was roughly 182,000. Whereas, the curve for the forecasted cases continues at the same rate as the pre-lockdown period and reaches a cumulative case count of roughly 400,000 on 03-June-2020.

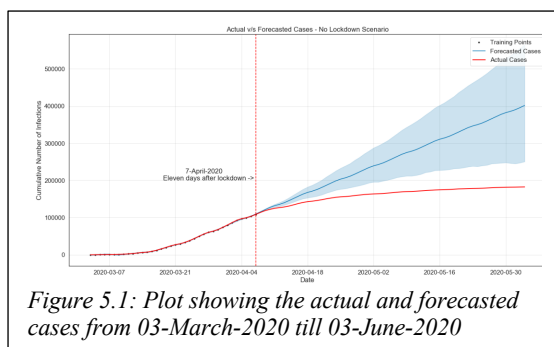


Figure 5.1: Plot showing the actual and forecasted cases from 03-March-2020 till 03-June-2020

Similarly, from *Figure 5.2*, we can see that while the actual number of fatalities due to COVID-19 was around 8,600 as on 03-June-2020, the projected number of deaths without the lockdown would have been 22,359 for the same date.

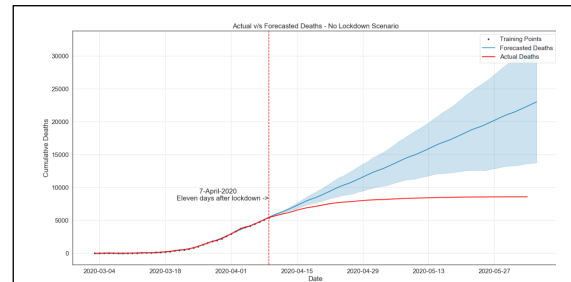


Figure 5.2: Plot showing the actual and forecasted deaths from 03-March-2020 till 03-June-2020

4. Conclusion

As is evident from the low p-values, we *rejected* the *Null Hypothesis* that “*Mean Infection Rate Remained the Same Before & After Lockdown*”. Thus, we have statistically proved that the lockdown was successful in reducing the *Infection Rate* and was indeed an effective measure in controlling the spread of COVID-19 across the German population. This was further corroborated by the high t-stats, which showed that the mean of the distributions, centred around 07-April-2020, varied significantly for the periods before & after this date.

Finally, the forecast on 03-June-2020 tells us that without the lockdown an additional 218,000 people would have caught the disease. The fatality count would have increased by 13,750. We can thus conclude that there would’ve been a 120% increase in the number of cases, and a corresponding 260% increase in fatalities, without the lockdown. The lockdown has thus saved the lives of at least 13,750 individuals.

5. References

1. Naming the coronavirus disease (COVID-19) and the virus that causes it. Retrieved 30 July 2020, from Who.int website: [https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-\(covid-2019\)-and-the-virus-that-causes-it](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it)
2. Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., ... Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*, 395(10223), 497–506.
3. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). <https://www.who.int/news-room/detail/30-01-2020-nCoV>

- [2020-statement-on-the-second-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-\(2019-ncov\)](#). Retrieved 30 July 2020.
4. WHO Director-General's opening remarks at the media briefing on COVID-19—11 March 2020.
<https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>. Retrieved 30 July 2020.
 5. COVID-19 Dashboard by the Centre for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Johns Hopkins University.
<https://coronavirus.jhu.edu/map.html>. Retrieved 30 July 2020.
 6. Ergänzung zum Nationalen Pandemieplan – COVID-19 – neuartige Coronaviruserkrankung (PDF). Robert Koch Institute.
https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Ergaenzung_Pandemieplan_Covid.pdf;jsessionid=AEF9B67828049CE1FBC0B3B7742350FC.internet092?_blob=publicationFile. Retrieved 4 March 2020.
 7. de Winter, J.C.F. (2013) "Using the Student's t-test with extremely small sample sizes," Practical Assessment, Research, and Evaluation: Vol. 18, Article 10.
 8. Hypothesis Testing - Analysis of Variance (ANOVA): Lisa Sullivan, PhD. Professor of Biostatistics. Boston University School of Public Health
 9. Taylor, S. J., & Letham, B. (2017). Forecasting at scale. doi:10.7287/peerj.preprints.3190v2. Retrieved 30 July 2020.
 10. <https://facebook.github.io/prophet/>. Retrieved 30 July 2020.
 11. Letham, B. (2017, February 23). Prophet: forecasting at scale - Facebook Research. Retrieved August 26, 2020, from Research.fb.com website:
<https://research.fb.com/blog/2017/02/prophet-forecasting-at-scale>. Retrieved 30 July 2020.
 12. Linton, O. (1996). Estimation of additive regression models with known links. *Biometrika*, 83(3), 529–540.
 13. RKI - Homepage. (n.d.). Retrieved August 26, 2020, from Rki.de website:
https://www.rki.de/EN/Home/homepage_node.html. Retrieved 30 July 2020.
 14. RKI COVID-19 Germany. Retrieved July 30, 2020, from Arcgis.com website: <https://npgeo-corona-npgeo-de.hub.arcgis.com/app/478220a4c454480e823b17327b2bffd4>.
 15. *COVID-19 Tracking Germany* [Data set]. <https://www.kaggle.com/headsortails/covid19-tracking-germany>. Retrieved 30 July 2020.
 16. Animations in the time of Coronavirus - head spin - the Heads or Tails blog. (n.d.). Retrieved July 30, 2020, from Github.io website: <https://headsortails.github.io/2020/04/30/animate-map-covid>.
 17. United States COVID-19 Hospital Needs & Death Projections-Institute for Health Metrics and Evaluation (IHME).
<http://ghdx.healthdata.org/record/ihme-data/united-states-covid-19-hospital-needs-and-death-projections>. Retrieved 30 July 2020.
 18. Germany: WHO Coronavirus disease (COVID-19) dashboard. (n.d.). Retrieved July 30, 2020, from Who.int website: <https://covid19.who.int/region/euro/country/de>
 19. Siegle, D. (2015, May 22). t Test | Educational Research Basics by Del Siegle. Retrieved August 26, 2020, from Uconn.edu website: <https://researchbasics.education.uconn.edu/t-test/>. Retrieved 30 July 2020.
 20. Radford, P. J., Velleman, P. F., & Hoaglin, D. C. (1983). Applications, basics, and computing of exploratory data analysis. *Biometrics*, 39(3), 815.
 21. Hypothesis testing - analysis of variance (ANOVA). (n.d.). Retrieved August 26, 2020, from Bumc.bu.edu website:
https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_HypothesisTesting-ANOVA/BS704_HypothesisTesting-Anova_print.html. Retrieved 30 July 2020.
 22. Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., ... Lessler, J. (2020). The incubation period of Coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine*, 172(9), 577–582