BRINGING BACK COVID-19 INTO THE SPOTLIGHT

USING MACHINE LEARNING IN FORECASTING COVID-19 CONFIRMED CASES IN THE PHILIPPINES FOR CY 2022

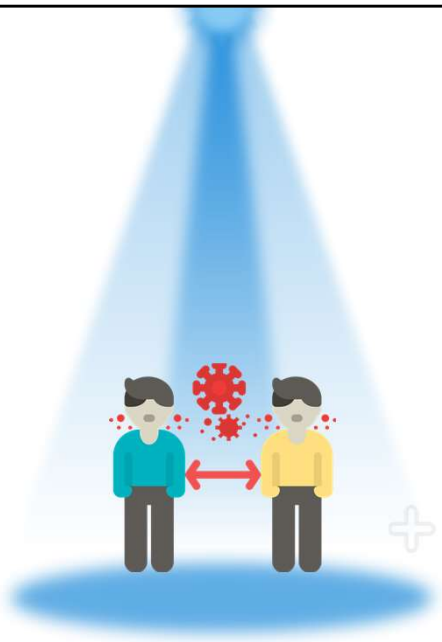
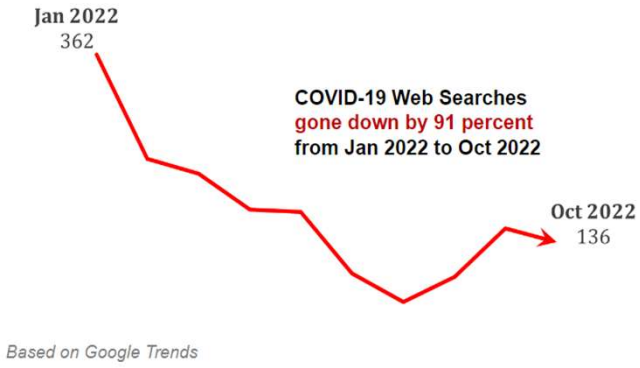


Table of Contents

Overview

- Covid-19 is a global epidemic that greatly threatened the whole world, as it claimed many lives, but is now slowly being ignored.
- Google Trend shows that from January 2022 to October 2022 google searches about Covid-19 in the Philippines had been reduced by 91 percent
- Using the Department of Health (DOH) 2022 Covid-19 data, we will try to forecast the Covid-19 confirmed cases trend in the Philippines using Machine Learning Algorithm in order to intelligently analyze and understand the Covid-19 Situation in our country these days.



Based on Google Trends

Problem Statement

- How can we use Machine Learning Algorithms to explore and predict the number of Covid-19 Confirmed Cases in the Philippines?

Figure 1. Covid-19 Time Series Model Diagram

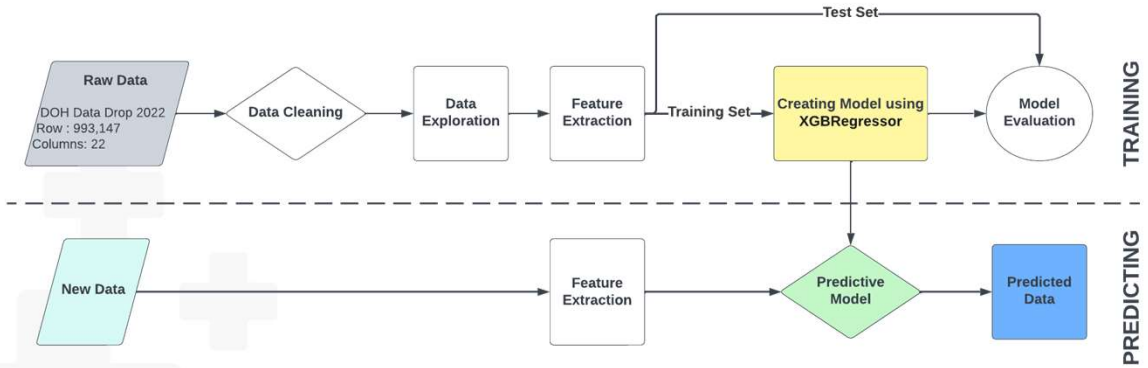
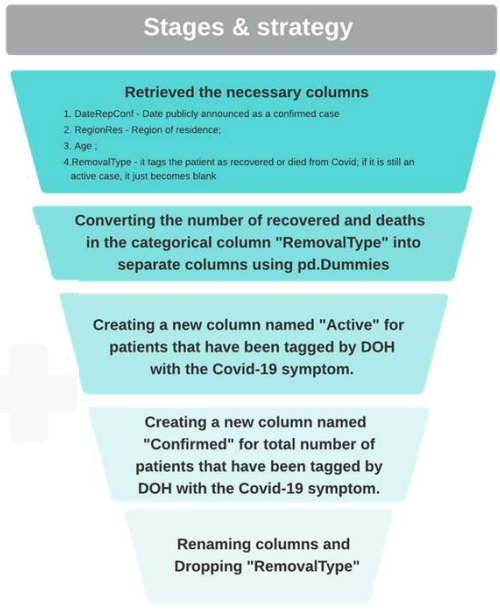


Figure 2. Data Cleaning



ORIGINAL DATA SET

Columns: 22
Rows: 993,147

Total Data Points: 18,200,926

PROCESSED DATA SET

Columns: 7 (Date, Age, Region, Died, Recovered, Active, Confirmed)
Rows: 993,147

Total Data Points: 6,945,781

Data Exploration

Table 1: Top 10 Regions in the Philippines with Highest Confirmed cases of Covid-19 in 2022

Region	Confirmed	Died	Recovered	Recovery Rate	Mortality Rate
NCR	323461	952.0	315176.0	97.438640	0.294317
Region IV-A: CALABARZON	186043	515.0	181412.0	97.510791	0.276818
Region III: Central Luzon	98128	941.0	94430.0	96.231453	0.958952
Region VI: Western Visayas	57482	572.0	55086.0	95.831739	0.995094
Region VII: Central Visayas	49517	625.0	47725.0	96.381041	1.262193
Region XI: Davao Region	43798	412.0	41902.0	95.671035	0.940682
Region I: Ilocos Region	36280	363.0	35138.0	96.852260	1.000551
CAR	32818	306.0	32125.0	97.888354	0.932415
Region II: Cagayan Valley	31187	413.0	30203.0	96.844839	1.324270
Region X: Northern Mindanao	25053	81.0	24385.0	97.333653	0.323315

As we process the DOH dataset, we will try to explore the present situation of Covid-19 in the Philippines. Table 1 shows the Top 10 regions with the highest number of confirmed cases of Covid-19 in the Philippines for 2022. NCR has the highest number of confirmed cases followed by Region IV-A, and Region III. Fortunately, the mortality rate of all regions did not reach 2%, with the highest mortality rate of 1.32 % from Region II, followed by Region VII with 1.26%, and Region I with a 1% mortality rate

Take note that these data only cover the reported cases of Covid-19 in the Philippines for 2022 and did not represent all the cases of the disease from previous years.

Figure 3: Top 10 Regions with Highest Active Covid-19 Cases

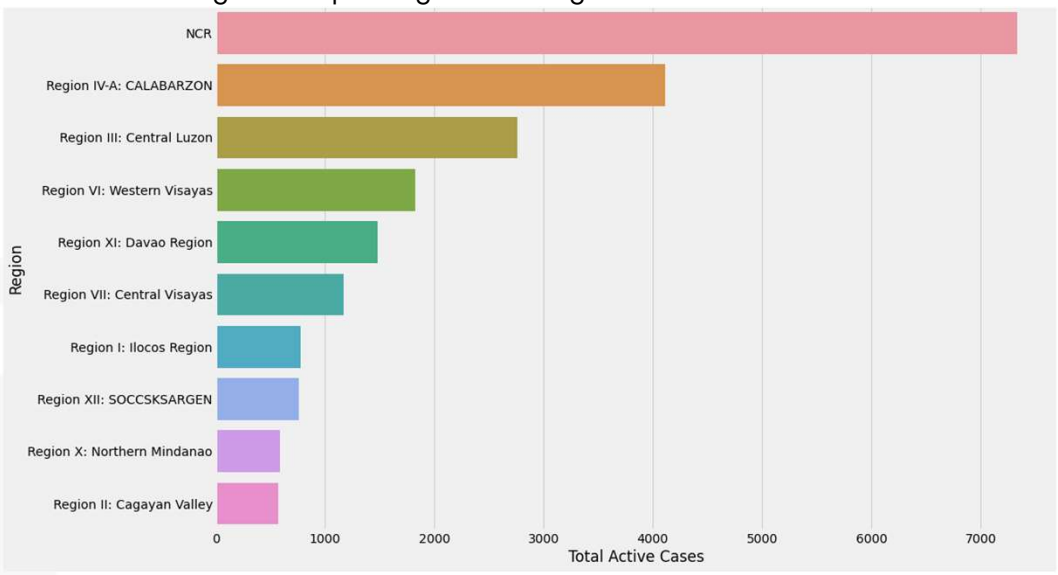


Figure 3 shows that the highest number of active Covid-19 Cases in the Philippines is from the NCR, followed by Region IV-A, and Region 3. The least number of cases are from Region II, followed by Region X, and Region XII

Figure 4: Daily new confirmed cases of COVID-19

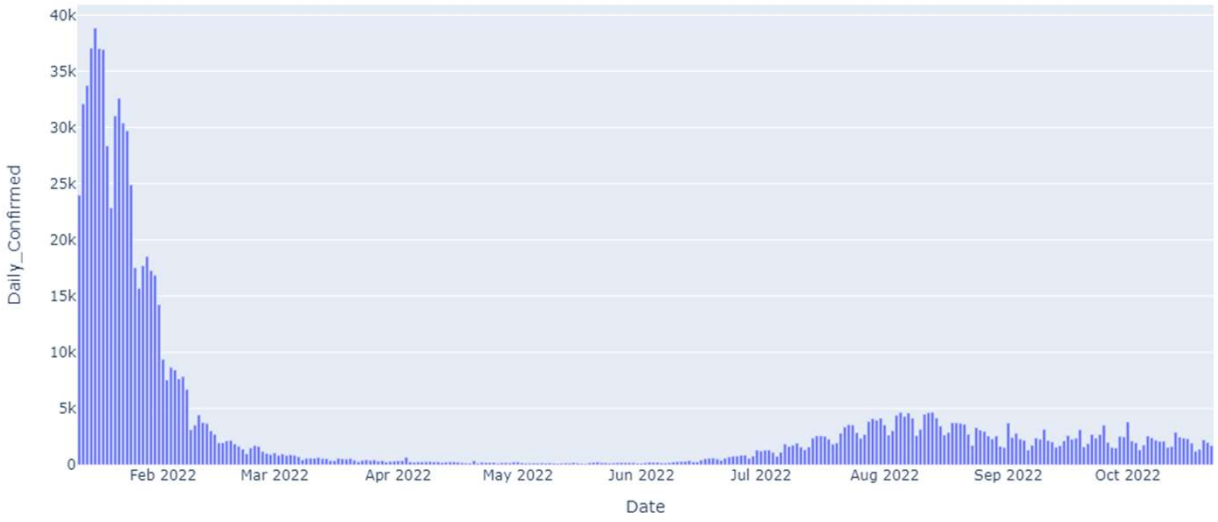
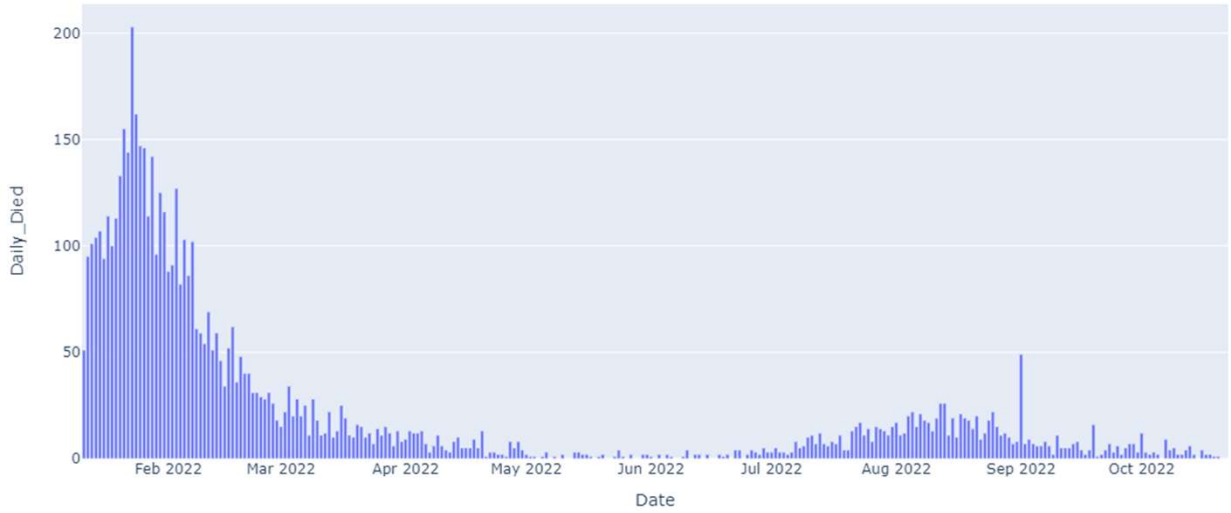


Figure 4 shows that the peak of Covid-19 daily new confirmed cases in the Philippines in 2022 was the month of January, but it drastically goes down from March to June and has starting to rise from July to August, and begun to consolidate from September to October.

This figure was supported by the journal article titled "Lineage BA., 2 dominated the Omicron SARS-CoV-2 epidemic wave in the Philippines," which stated that both the Omicron BA.1 and BA.2 lineages occurred in the Philippines in December 2021, before the number of patients spiked in January 2022 (Li et al., 2022).
Further, the increase in cases starting from the month of July can be explained by the relaxation of compliance with safety protocols against Covid-19.

Figure 5: Daily new confirmed COVID-19 deaths



Similarly, figure 5 also shows a peak number of Covid-19 confirmed deaths in the Philippines was in January in 2022 and likewise drastically go down starting March to June and has starting to rise from July to August and begun to consolidate from September to October.

Figure 6: Daily new confirmed COVID-19 Recovered cases

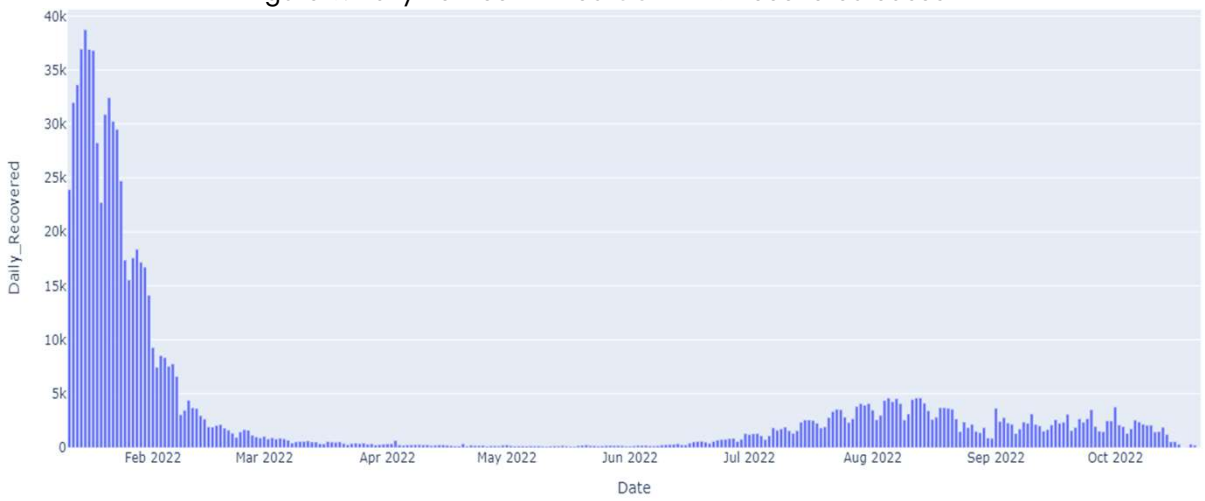


Figure 6 supports the previous charts which also show that the peak recovered the number of Covid-19 Cases in the Philippines in 2022 was in the month of January and likewise drastically go down starting March to June and has starting to rise from July to August and begun to consolidate from September to October.

Feature Extraction

01

We created a new dataset only containing the Date and Number of Confirmed Covid Cases

02

Using the Date Column we created additional Datetime columns as features such as:

1. Day of Week
2. Month
3. Day of Month

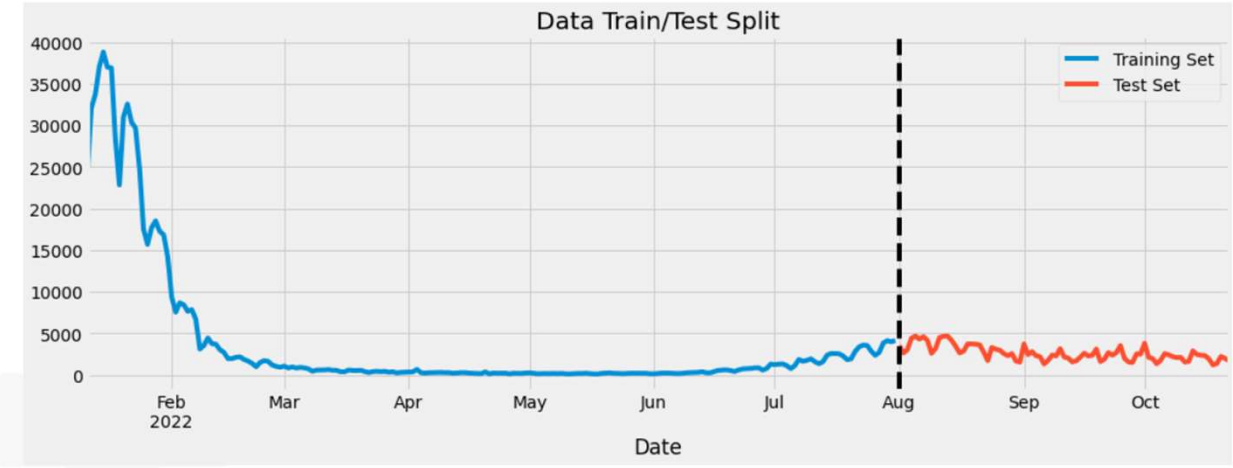
03

We also created Lag features to our model including:

1. 30 Day
2. 60 Day
3. 90 Day

Figure 7. Train / Test Split

Using a 70/30 Split for the Training Set, we will try to forecast the trend of Covid-19 Confirmed cases in the Philippines from August to October 2022 based on the trend of confirmed cases from January to August 2022.



Creating the Model

We used the XGBRegressor as our model with the following parameters:

- base_score=0.5
- n_estimators=900
- early_stopping_rounds=50
- learning_rate=0.01
- objective='reg:linear'
- max_depth=3

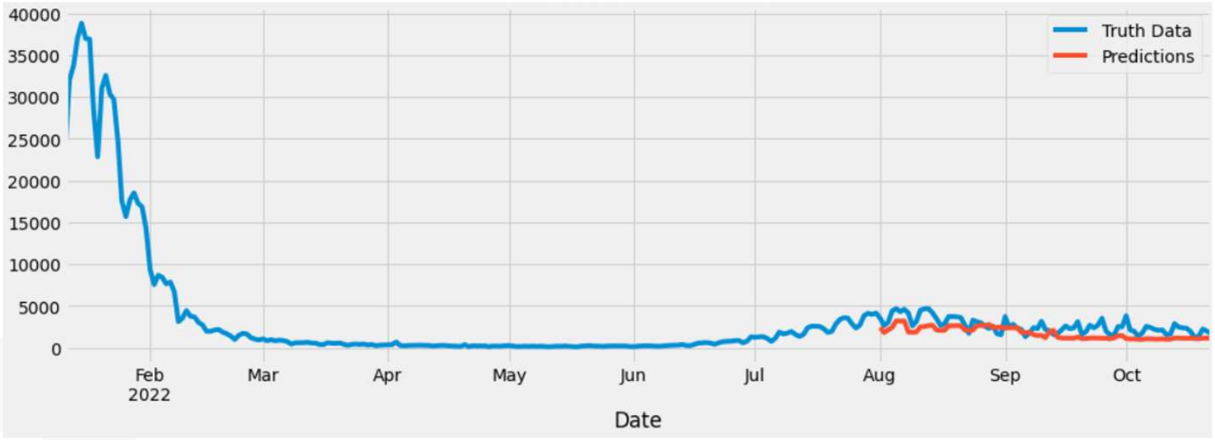
Different Iteration of the Model

```
[08:26:02] WARNING: /workspace/src/objective/regression_obj.cu:152: reg:linear is now deprecated in favor of reg:squarederror.
[0]     validation_0-rmse:9071.98       validation_1-rmse:2773.23
[100]   validation_0-rmse:3798.73       validation_1-rmse:1879.03
[200]   validation_0-rmse:1738.83       validation_1-rmse:1437.41
[300]   validation_0-rmse:899.838       validation_1-rmse:1278.73
[400]   validation_0-rmse:558.906       validation_1-rmse:1191.22
[500]   validation_0-rmse:417.029       validation_1-rmse:1157.89
[600]   validation_0-rmse:347.423       validation_1-rmse:1154.62
[700]   validation_0-rmse:299.223       validation_1-rmse:1149.29
[800]   validation_0-rmse:275.399       validation_1-rmse:1144.16
[899]   validation_0-rmse:262.326       validation_1-rmse:1141.49
XGBRegressor(early_stopping_rounds=50, learning_rate=0.01, n_estimators=900)
```

Model Validation

To validate the model, I used Root Mean Squared Error and got an **RMSE score of 1141.49**

Figure 8. Original Data vs. Predicted Data



Model Application

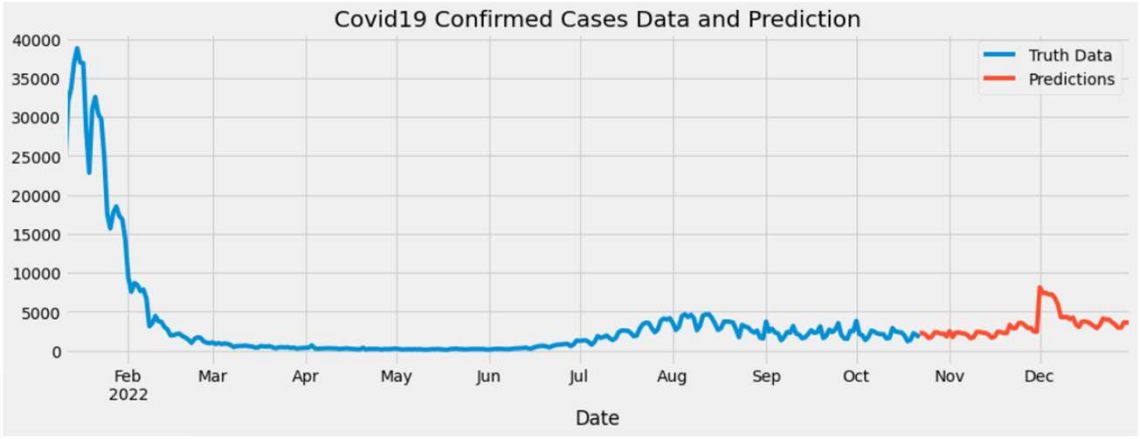
Using our model, we have successfully predicted the trend of daily confirmed cases of Covid-19 in the Philippines for the month of November to December 2022

Figure 9: Forecasted Results

Figure 9, forecasted that the daily confirmed cases of Covid-19 in the Philippines will reach its peak by December 1st before it will start to decline and consolidate with higher volume than the months of October and November 2022.

This trend can be Justified by the policy of allowing the voluntary wearing of face masks in indoor and outdoor spaces pursuant to Executive Order No. 7, series of 2022, issued by President Ferdinand Marcos, Jr. (Exec. Order No. 07, 2022) and the resumption of face-to-face classes in the country, as well as the celebration of Christmas this December.

Figure 10. Data Trend and Predictions



Conclusion and Recommendation

- Using data analytics, this project has successfully provided a forecast of the trend of Covid-19 confirmed cases in the Philippines for the months of November to December 2022 using the DOH data set from January to October 2022.
- This forecast forewarns the policy-makers and the Filipino people, that the threat of Covid-19 in the Philippines is still present. Therefore, it is highly recommended that our policymakers to review our policy of allowing the voluntary use of facemasks in indoor and outdoor spaces. Likewise, we should remind the public to still stay vigilant and follow the safety protocols in order to prevent the recurrence of major spikes in the number of confirmed covid-19 cases in our country.
- Proven to have produced a useful predictive model, using Machine Learning Algorithm, I encourage other data scientists to use this script to try to explore other factors that can provide further analysis of the Covid-19 situation in our country and even in the whole world to prevent its recurrence.

References

- Department of Health. (2022). DOH COVID Data Drop – Case Information batch 3. [Data file]. Retrieved from https://drive.google.com/drive/folders/1oy9gCSR1SC4WIkoEAj2bn8-2SPhMIOIg
- Office of the President, Allowing Voluntary Wearing of Facemasks in Indoor and Outdoor Settings, Reiterating the Continued Implementation of Minimum Public Health Standards During the State of Public Health Emergency Relative to the Covid-19 Pandemic, Exec. Ord. No. 07, § 2022 (Oct. 28, 2022) (Phil.), https://www.officialgazette.gov.ph/downloads/2022/10oct/20221028-EO-7-FRM.pdf.
- Li, Y., Polotan, G. F. M., Sotelo, I. G. S., Alpino, P. A. A., Dolor, Y. A. M., Tujan, A. M. A., Gomez, R. M. R., Onza, J. O. T., Chang, K. A. T., Bautista, C. T., Carandang, J. C., Yangzon, S. M. L., Pangilinan, A. E. R., Mantaring, R. J., Telles, J. A. E., Egana, M. J. C., Endozo, J. J. S., Cruz, P. R. S., Tablizo, F. A., . . . Brunker, K. (2022, August 19). Lineage BA.2 dominated the Omicron SARS-CoV-2 epidemic wave in the Philippines. OUP Academic. https://academic.oup.com/ve/article/8/2/veac078/6672639