Exploration of COVID-19 Data

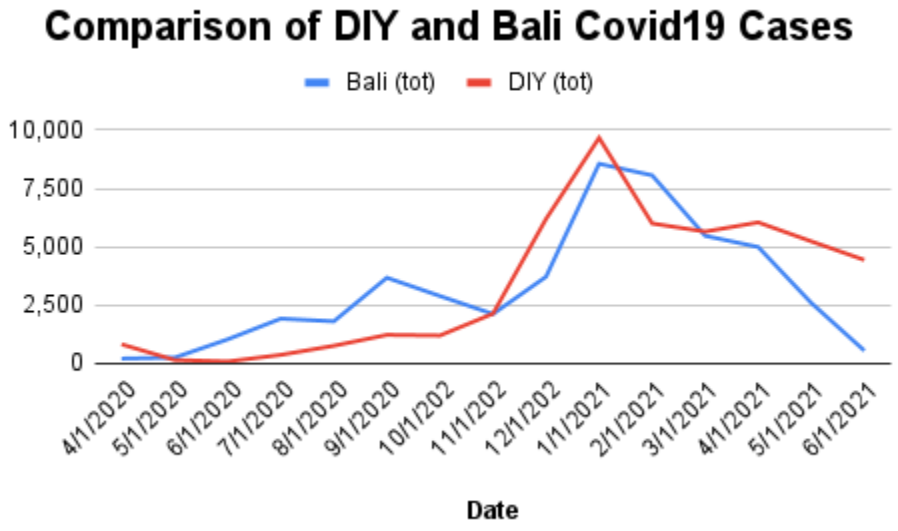
Introduction

I doubt anyone in 2019 could have prophesied how the world would change in just over a year's passing. Nevertheless, almost two million people have contracted the coronavirus in one form or another, and the numbers aren't getting smaller. More variants are causing more problems to more nations around the world. Our country seems to be handling the situation not as diligently as other nations, but that may be due to an erroneous amount of factors, most of which are out of our control.

This paper will report our data analysis project and cover multiple approaches to multiple solutions (as prescribed by the final project briefing). Among them are a descriptive, predictive, prescriptive, and cognitive analysis of cases, deaths, and approaches to solving the rising number of patients nationwide.

Descriptive

How do the total cases compare between Bali and DIY?



Bali and the Special Region of Yogyakarta compare in most other aspects and COVID-19 cases. Both are tourist destinations, with both long- and short-term visitors, along with a high concentration of ex-pats. Not sure how all of them interact to contribute, but there is something there.

The total cases for both provinces are very comparable, with DIY having a slight edge. We suspect this is due to DIY having more land routes to travel to neighboring regions while Bali is isolated. It is compensated by the number of tourists departing from Jakarta, though, the epicenter of this country's epidemic.

Look at the cumulative histogram, though. You can see that both provinces exhibit a normal distribution, with Bali having a more aggressive distribution, indicating it had its daily peaks earlier.

Despite this, the monthly averages for both provinces are pretty comparable, with only a 3 case difference between the two. So that indicates to us that even if the peaks are on different dates, the end outcome will be pretty similar.

Predictive

What will happen tomorrow?

Soon, we cannot deny that the number of COVID-19 cases will always rise from time to time. We can only depend on vaccines, immunization, and the government's health program because there are no definite cures. Due to this statement, we can easily assume that the total number of cases will increase each day until an unknown time.

Using the dataset provided, we can easily predict the output of the future by using predictive analysis. Our analysis uses the linear regression method to forecast the following months of the cumulative COVID-19 cases. There are a few steps to fulfill this analysis, which is :

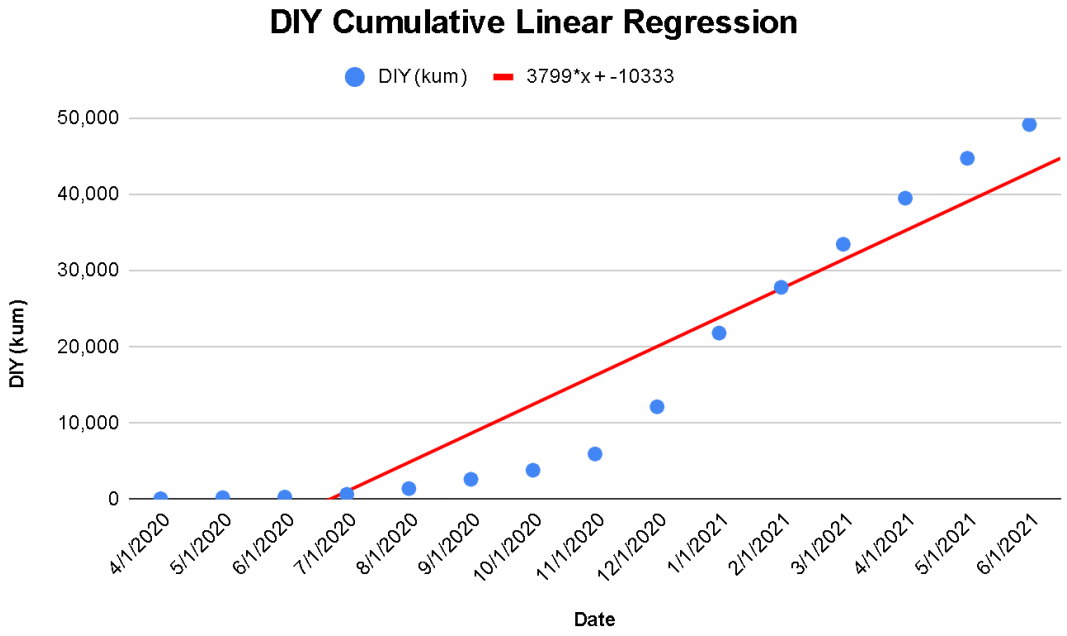
1. Data Exploration
   Data exploration is the initial step in data analysis, where users explore an extensive data set in an unstructured way to uncover initial patterns, characteristics, and points of interest. This process isn't meant to reveal every bit of information a dataset holds but rather to help create a broad picture of significant trends and significant points to study in greater detail. Before conducting the predictive analysis, we had already completed this step in the last part: a descriptive analysis. Therefore, we can continue our exploration to the next step.

2. Data Cleaning
   Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. For example, before conducting any exploration, we checked the "kumulatif prov" page and ensured that the data has no missing/null values. Therefore, the data is already cleaned.

3. Modeling

The process of creating a visual representation is called modeling. As said before, we're going to use linear regression as a model to represent our analysis.

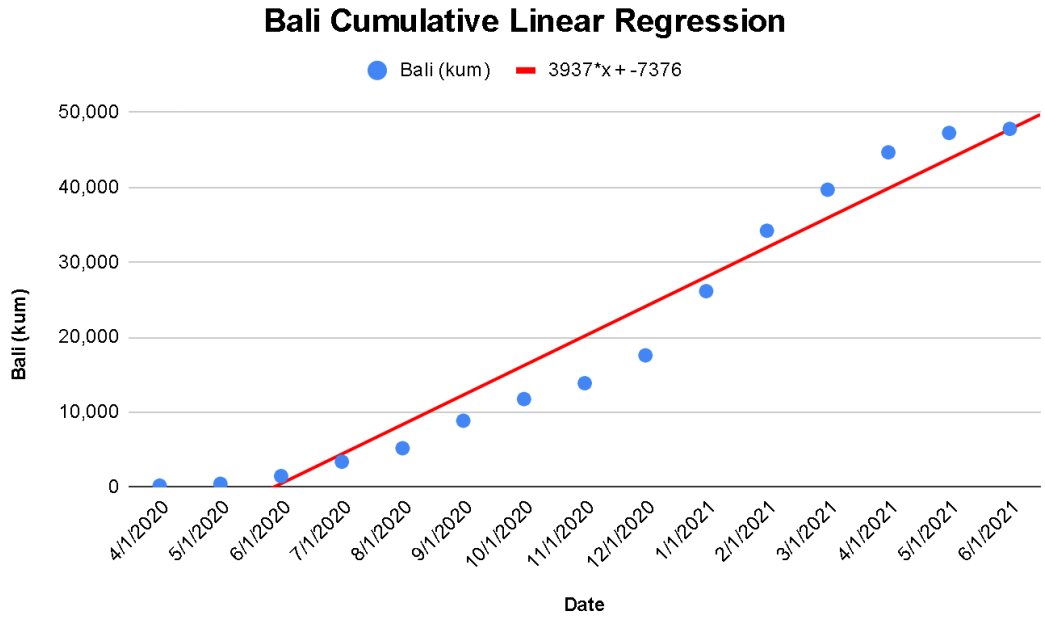
Yogyakarta



In this graph, the total cumulative cases from the early month of the analysis is relatively low. This is caused by the date of the data that started from the mid of March 2020, so we can only accumulate the information starting from April 2020.

According to the news, the COVID-19 cases in DIY (especially in the early month of 2020) is not as high as the other big Indonesian cities like Bali or Jakarta. But as time came by, the cases started to rise up and now is rivaling Bali. Therefore, we can conclude that the total amount of cases will still rise up until the society follows the government's health program more seriously

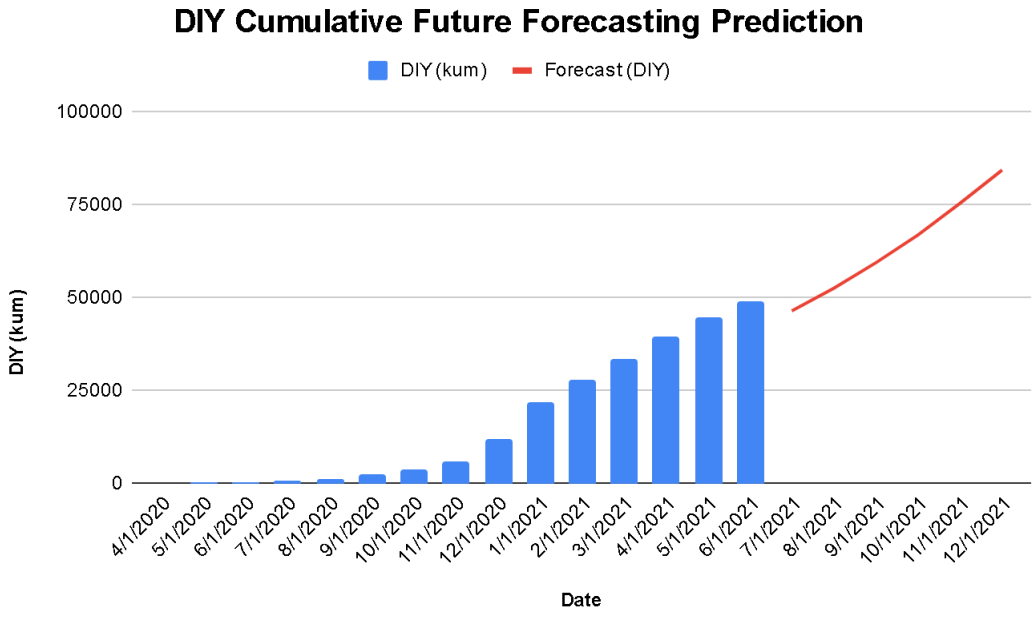
Bali



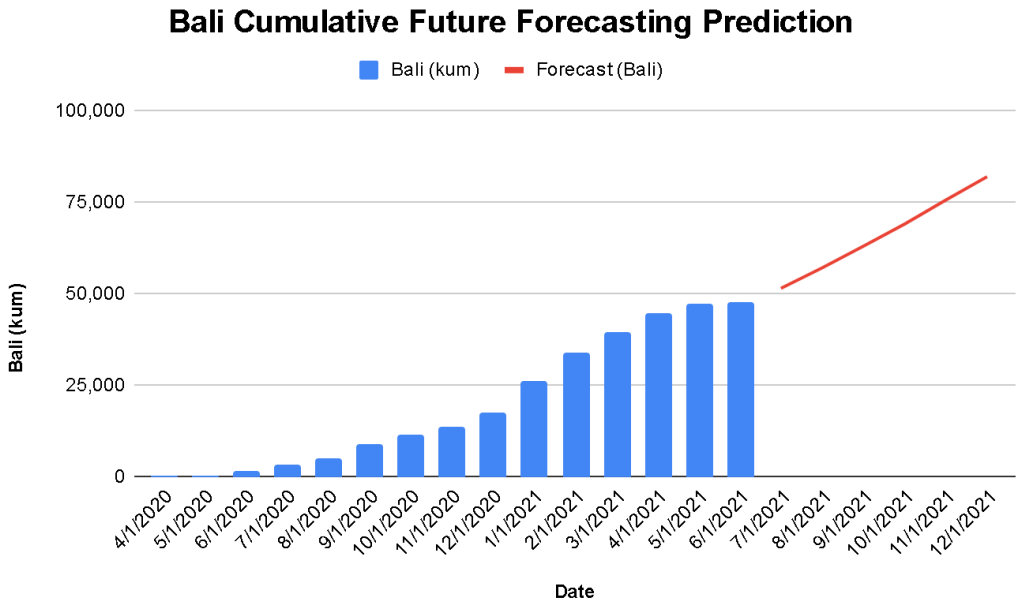
Same as DIY, the total cumulative cases will still rise up for the next following months. What makes this data different is that the total number of cases Bali had in the beginning of pandemic is quite high. Since then, Bali's government has tightened the regulations that causes the cases to reduce.

4. Analysis

Yogyakarta



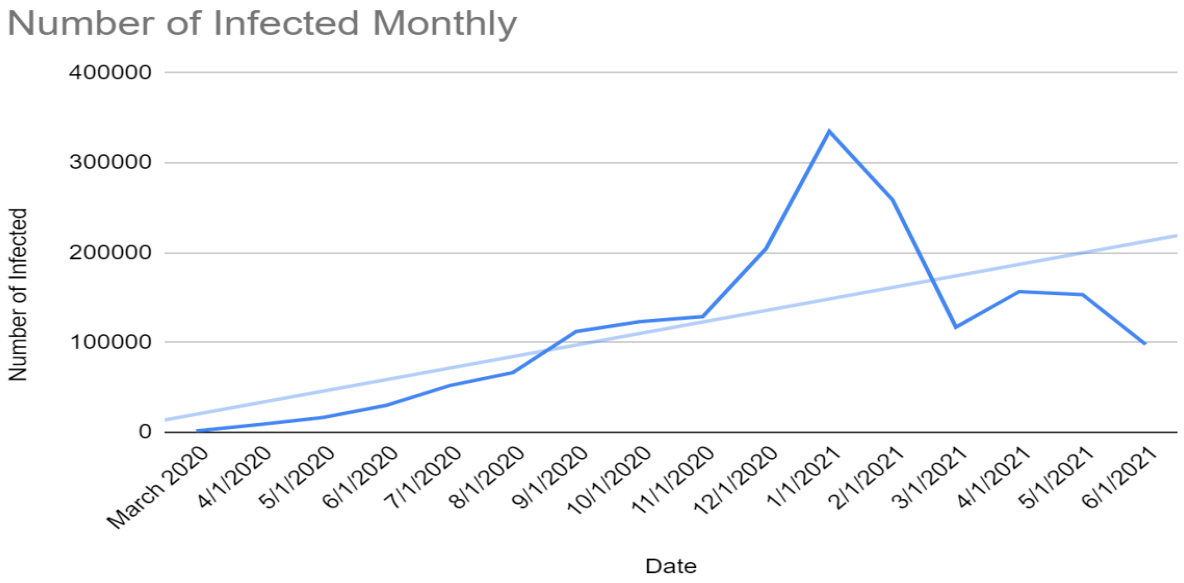
Bali



The prediction for the following months uses linear regression method as a prediction model to forecast the future outcome of the dataset. Because we use

cumulative data, the trend for the future will always rise up. Note that this prediction is only based on the historical data and can change anytime if future regulations are made.

Prescriptive

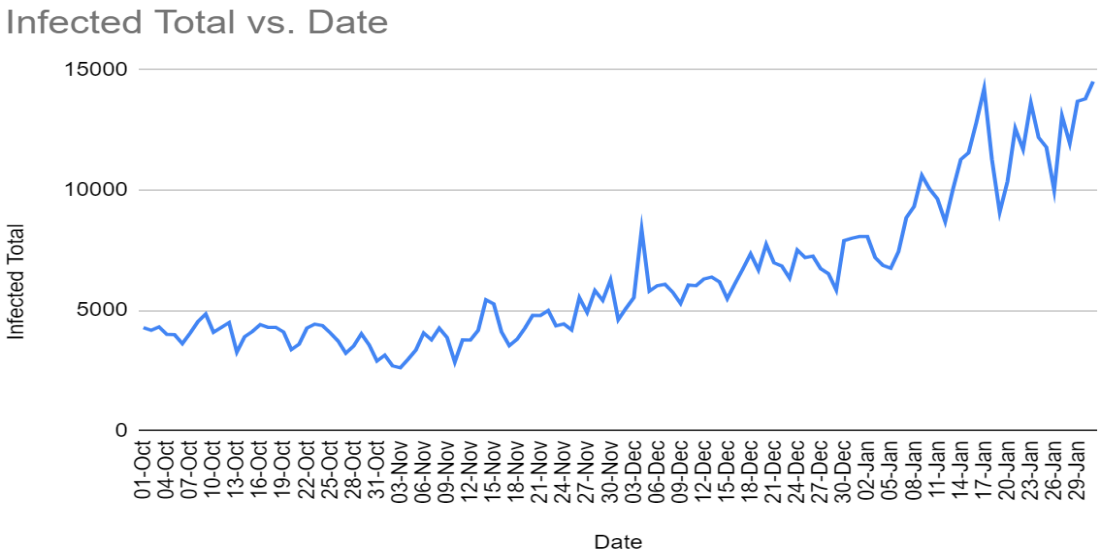


The number of infected individuals had seemed to rise recently and projected to keep increasing over the year. This fact is concerning, especially with the appearance of new Corona Viruses variants, increasing the infection rate. Despite vaccination efforts since January 2021, it still hasn't seemed to reduce infection rates.

There are many causes of high infection rates. One of the crucial factors is interactions between individuals. During the pandemic, news and reports of groups of people gather together despite the danger of the virus. To counter this, the government enacted large-scale social isolations multiple times during the pandemic. The goal is to limit the spread of the virus. With this, how crucial are large-scale social isolations?

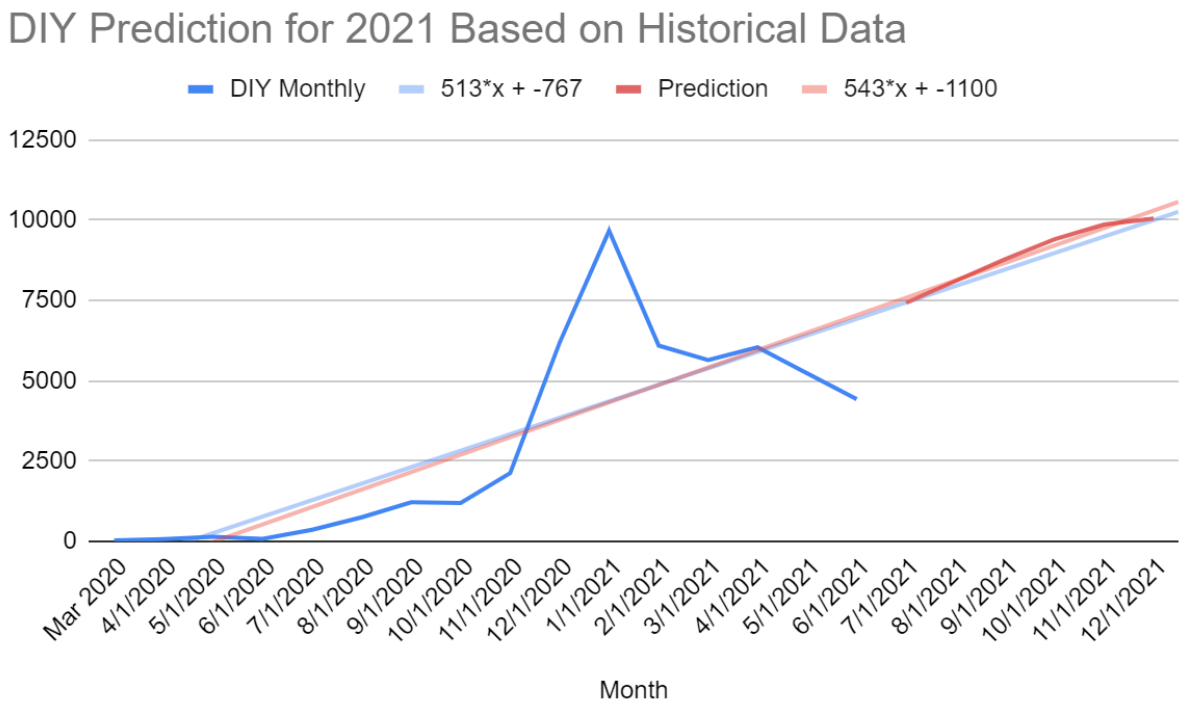
How Impactful Large Scale Isolations in Reducing Infection Rate?

The new year's eve of 2021 gave us a recent increase in the numbers of infected.

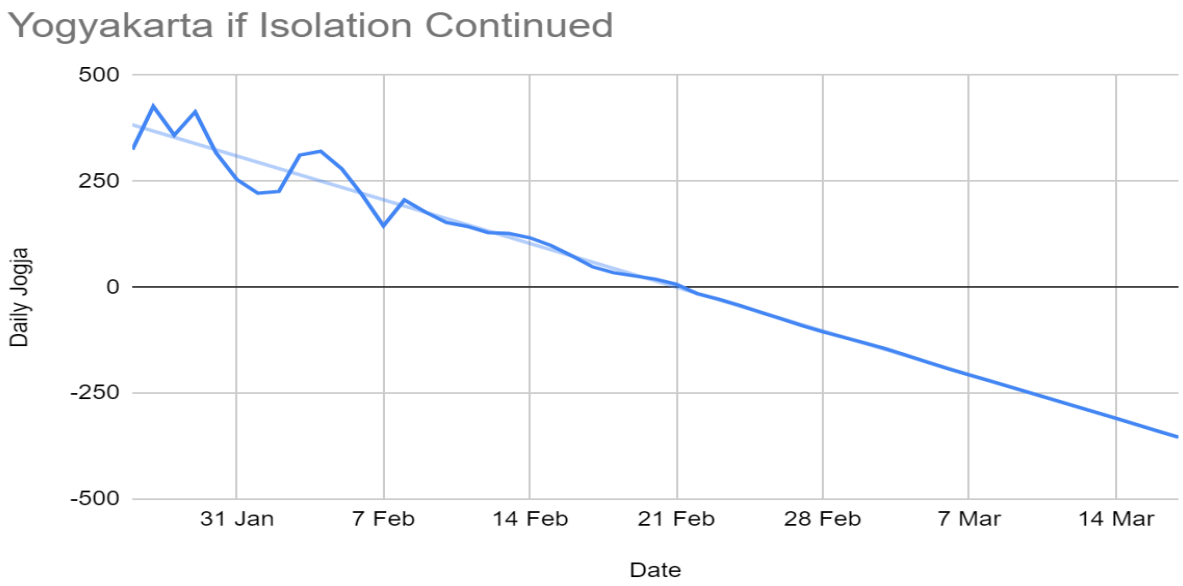


This increase, however, had been reduced by enacting large-scale isolation over early January until early February 2021. Taking a look deeper into the isolation period, we could see how the virus could be less of a threat when there is no one to infect.

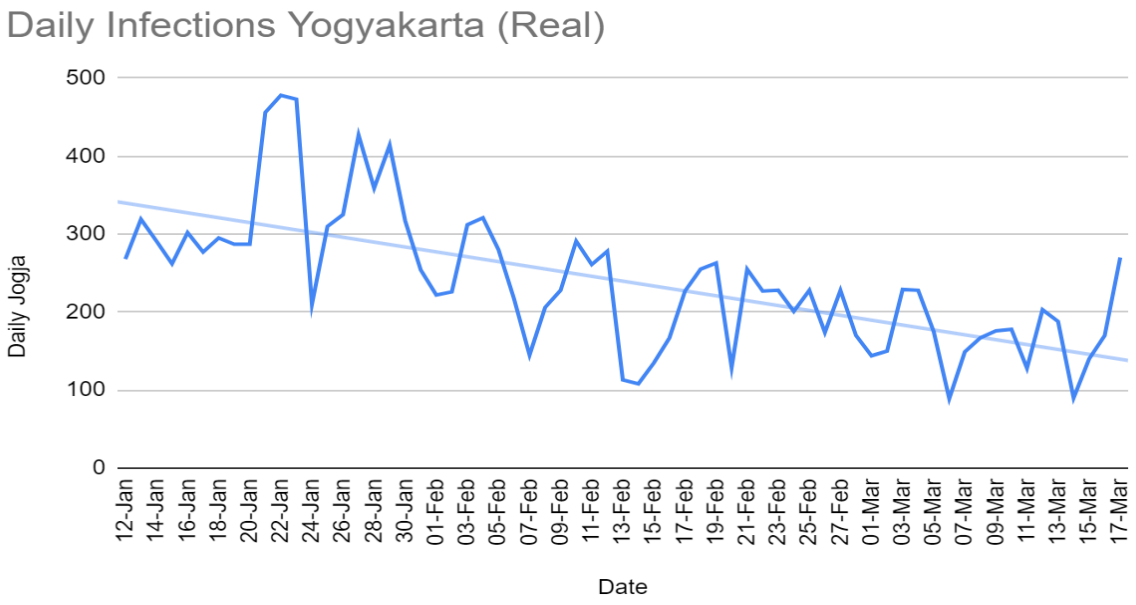
- Yogyakarta



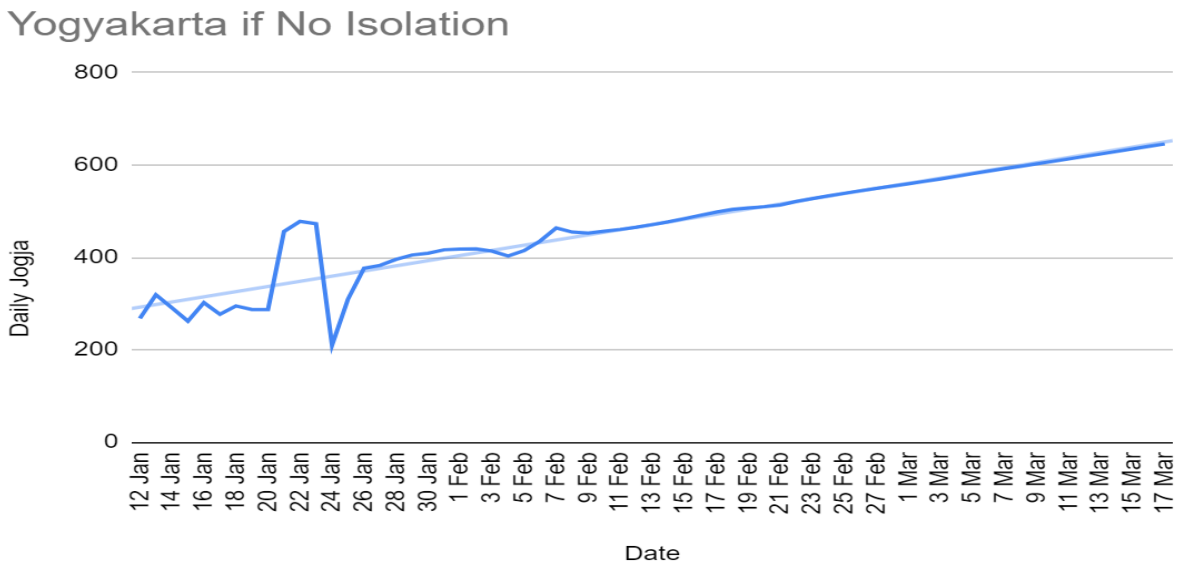
Yogyakarta had been projected to have an increase in the number of infected over 2021. But can any action be taken to prevent this? During the isolation of January 2021, the decrease in the infection rate is quite significant. Therefore, prolonging the isolation period could seem to help in reducing the infection rate further.



The prediction chart shows a surprising projection that by prolonging the isolation period, we could have no infection on February 21st. This is quite similar to the actual data where it did go down, but not 0.

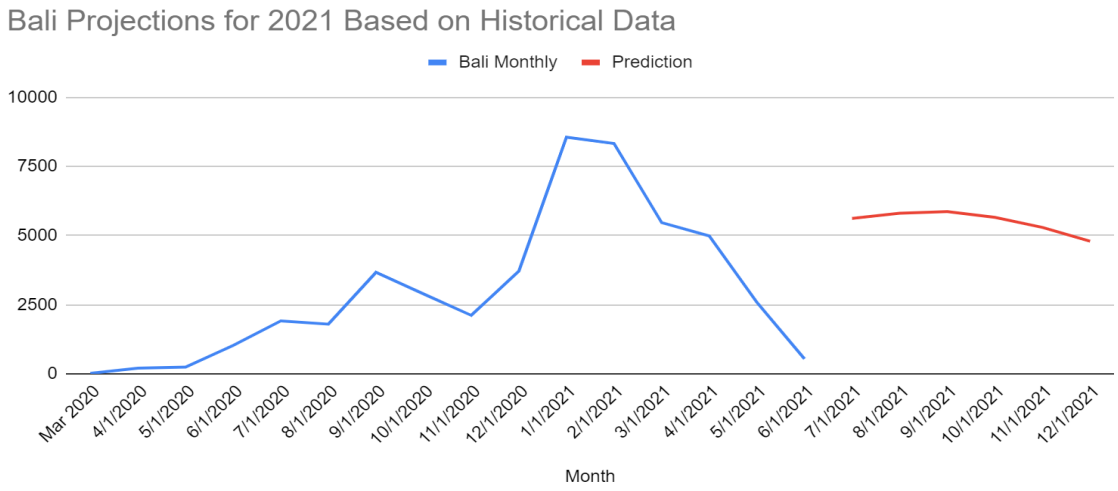


The large-scale isolation was right on time to be enacted since it would prevent infected people from infecting others if there weren't any.

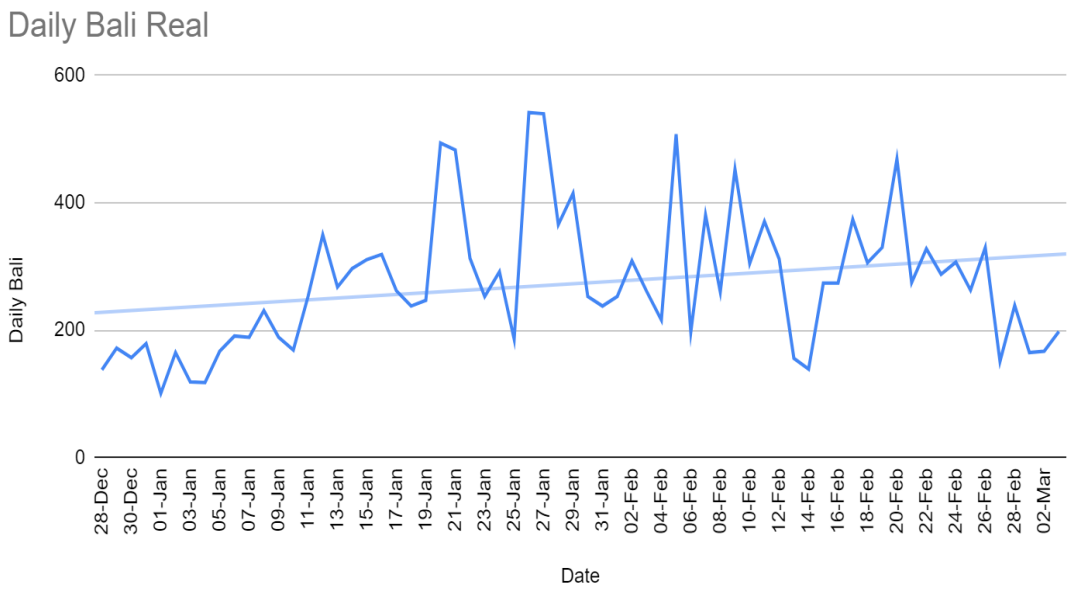


- Bali

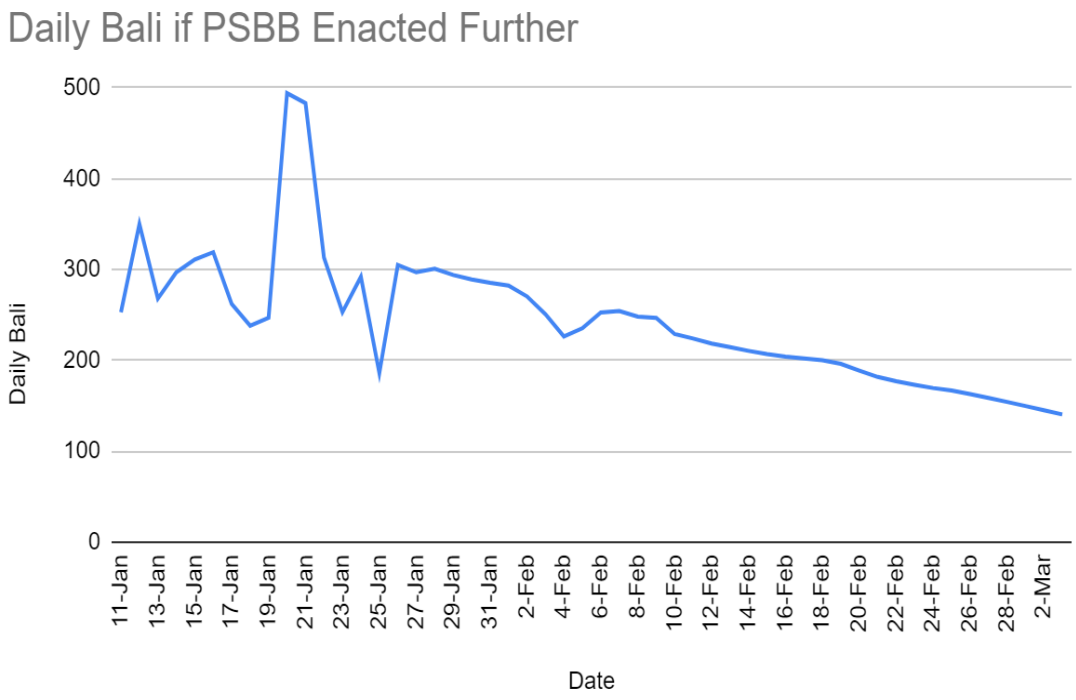
Bali also showed a similar trend during the New Year eve of 2021 and was subjected to large-scale isolation.



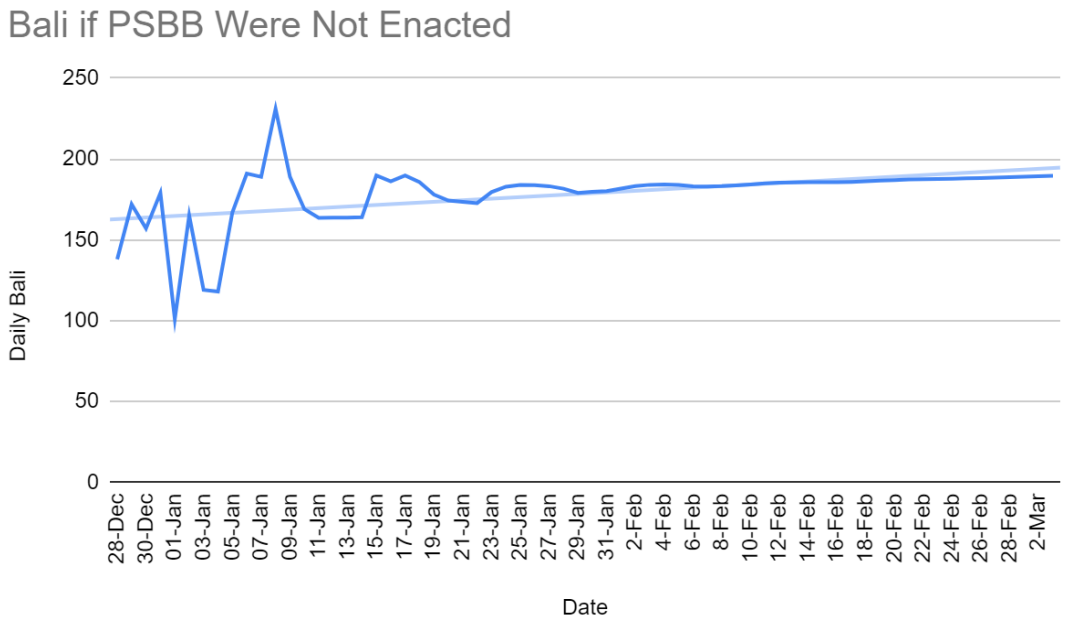
Like Yogyakarta, the isolation caused a drop from around 8000 infected in February 2021 to drop to about 6000 infected by March 2021. Despite the decrease, Bali is predicted to have a high case of around 5000 to 7000 people being infected over the last quarter of 2021. The large-scale isolation enacted was also in time to only stabilize the infection rate over the next couple of months.



If the period were to be prolonged, it would further decrease but not as significant as Yogyakarta projections.



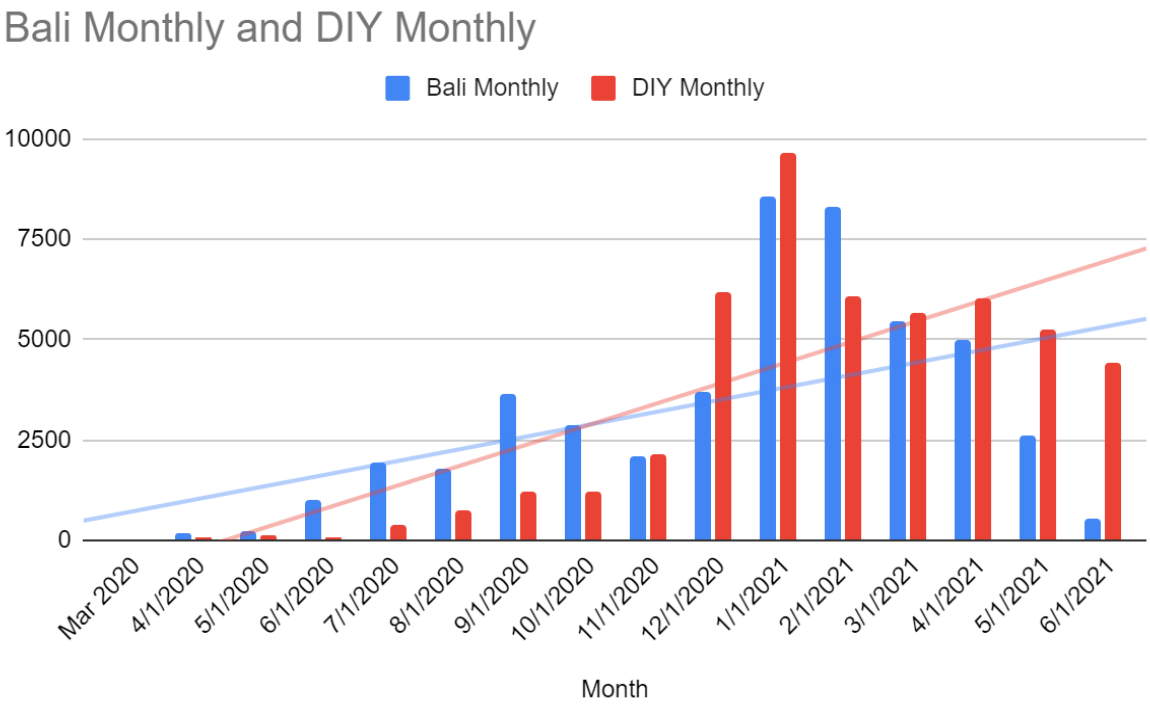
It could also have been worse for Bali if the large-scale isolation was not conducted.



Despite the hardships, Bali had seen a reduction of cases from March until May 2021.

- ## Why Compare Yogyakarta and Bali?

  Yogyakarta and Bali have a similar number in terms of cases. This shows the similar infection rates occurring in these different regions.



Another reason is a pairwise correlation that scored 0.987, which shows a relatively strong correlation between the two data.

- ## What Should We Do?

  Frequently prolonged isolation might be a good answer in reducing infection rates. This would help reduce interactions between the infected and the healthy. This helps minimize interaction with the fit and infected while also providing breathing room for medical infrastructures to treat and respond to the infections that occurred before the isolation was enacted. At the same time, the infection that occurs during the period does not come in a giant wave that could paralyze the medical infrastructure.

Conclusion

In conclusion, all of our analyses point toward a rise in cases without intervention. Our predictive analysis extrapolated the trend, and it registered an increase in patients in the future. Our prescriptive analysis compares two scenarios, both of which indicated a decrease in infection rates with the enactment of a lockdown. We should do our part in not contributing to the new positive cases list by staying at home, keeping our distance from others, and going out when it is necessary.

References

1. COVID-19 WRP (2021) Berita Terkini. In: covid19.go.id. https://covid19.go.id/p/berita/evaluasi-ppkm-dki-jakarta-dan-jawa-barat-dapat-menjadi-contoh. Accessed 23 Jun 2021