

# 1 Introduction

At the end of 2019, Covid-19 has outbreak around the world. Many people's work and lives have been affected by Covid-19, and it is also a huge challenge for the Health Care System in each country. According to the World Health Organization Coronavirus Dashboard, until November 17, 2022, it has 633,263,617 cumulative cases, and 6,594,491 cumulative deaths around the world ( ). Covid-19 continues to impact healthcare systems in various countries, to avoid medical crash, it is important to do the prediction research for Covid-19 inpatient cases. That can help healthcare system to prepare enough medical resources for patients. Also, the prediction research can show the impact of epidemic prevention policy on the number of inpatient people.

It has been almost three years since the Covid-19 outbreak, but the rate of growth in the number of confirmed Covid19 cases has not decreased. on the one hand, as the economic needs for each government and the need for people to get back to normal life have grown stronger and stronger, more and more countries opened the border to allow people to enter. On the other hand, the coronavirus is keeping mutating, and its contagiousness has been getting stronger. Due to those two reasons, doing the prediction for the Covid-19 inpatient cases is very important especially in the period that most countries and regions choose to reopen the border.

In this paper, I use the Logistic model to do the prediction of inpatient people base on the data from the Google Covid-19 data set. I do research for inpatient cases in the United States and the Hong Kong SAR. Form the existing data, we can get the prediction data, and from this data, we can learn the impact of epidemic prevention policies on confirmed cases in different regions.

## 2 Data Description

The source for this project is form the Google Covid-19 Open Data ( ). Google Covid-19 Open data is a great source for the research. It includes many aspects such as search trends, vaccination, and hospitalization for the Covid-19 data around the world, and it is wildly use. In this project, I choose the hospitalization data from the Google Covid-19 Open Data as the primary data.

The hospitalization data contains 11 categories. The first category is the date. This category records the date of the daily change in the number of hospitalizations. The second category is the location key, this category records the location of the inpatient cases. The next category is the new hospitalized patients, this category records the number of new hospitalized patients every day. The next category is the cumulative hospitalized patients, this category records the number of total hospitalized patients every day. The next category is the current hospitalized patients, this category records the number of current inpatients cases every day. The next category is the new intensive care patients, this category records the number of new hospitalized patients in ICU every day. The next category is the cumulative intensive hospitalized patients, this category records the number of total hospitalized patients in ICU every day. The next category is the current hospitalized patients, this category records the number of current inpatients cases in ICU every day. The last three category is recording the new, current, and total inpatient cases that use the ventilators in hospital.

This database collects 34 countries and regions inpatients data around the world, this includes Argentina, Australia, and Belgium. For all the countries and regions on the list, I decide to choose United State and Hong Kong SAR. as the location to study in this project. According to Travel States Gov ( ), after June 12, 2022, at 12:01AM ET, entering US border do not need to provide negative Covid-19 test. That means there is no Covid check for entering the US border. Also, Hong Kong released the "0+3" entry policy for the Covid-19 on September 26, 2022 ( ). Although there is no quarantine required to enter Hong Kong, but the government in Hong Kong SAR still monitoring the Covid cases on the border. Comparing with United State, Hong Kong government has stricter control over the Covid outbreak, and this is the reason why choosing these two locations as the research data. For the inpatient cases data, I choose the category of cumulative hospitalized patients, because this category records all the inpatients cases that we need to use, and this data is the most complete of all categories.

For the implementation, I use Python to optimize the database. The original database is an csv form. I use the function `"file = file[file['location_key'].isin(['US'])]"` and `"file2 = file2[file2['location_key'].isin(['HK'])]"` to find all US and HK inpatient cases data form the original data. Then, I use the `file.drop` function to drop all other categories beside the category of cumulative hospitalized patients. After that, I created two csv form to store the completed processing data separately. The completed processing data only

contains date, location key, and cumulative hospitalized patients. The complete processing form is US\_hospitalizations.csv and HK\_hospitalizations.csv.

### 3 Math Model Description

For the math model description, I choose logistic model. According to the Wikipedia ( ), logistic regression is used in various fields, including machine learning, most medical fields, and social sciences. For example, the Trauma and Injury Severity Score , which is widely used to predict mortality in injured patients. Many other medical scales used to assess severity of a patient have been developed using logistic regression. Logistic regression may be used to predict the risk of developing a given disease, based on observed characteristics of the patient. So, Using the logistic model to predict the Covid inpatient cases is appropriately. The formula shows below is commonly used for this model:

$$y = \frac{K}{1 + be^{-ax}}, a > 0, b > 0$$

For this model, use this formula to predict cases and we can get:

$$y = K/(1 + b * \exp(-at)), a > 0, b > 0$$

In the above formula, y as cumulative hospitalized patients for each day, t as the date, K as the maximum number of cumulative inpatient cases. The a and b are the parameters controlling the speed of infection.

For the implementation, I use python to do the prediction. The code shows below is the implementation for the formula shows above. For the parameters, I set t as time, t0 as initial time, P0 is initial value, K as capacity, and r is increase rate.

### 4 Model Prediction