• Include a short description of the Project Title, Motivation, Problem Statement, Dataset used, Methodology, Tools. You could create a brief PPT and use it while recording the video. • Provide a walkthrough of your project implementation, providing insights on what you have built and emphasizing on the approach taken, final results obtained and their evaluation.

• Delivering the data mining results in an attractive and understandable way is important. You can provide visualizations and use-cases to show how the results are interesting.

• You can mention how this project can be further improved and extended.

My project title is COVID-19 forecaster. First let me introduce the video outline. There are three parts. The first part is the brief introduction of my project idea and methodology. The second part is to walk through my implementation and talk about user cases. This final one is the improvement and conclusion of my project.

Let’s go to the first part.

Due to the spread of COVID-19, many countries have locked down, and people have quarantined for a long time. Recently, many people start back to workplace and students start their summer break so they have some traveling plan. However, the daily confirmed cases of covid-19 keep increasing. For people who want to travel, they may want to know which cities they should choose. For people who back to work, want to know the future trend of virus, so they can take some preventive action to protect themselves.

Many factors can affect the growth of covid-19, in this project, I analyze how weather factor affects Covid-19 and predict the growth of it. This project will be useful for people who have traveling plan and back to work already. Since it can provide some suggestions to them about the future trend of the virus and help them make the decision of choosing traveling cities. For those back to workplace already, they can take some preventive action before the next explode of covid-19.

There are three datasets used in this project. One of them stores covid-19 situation of most of cities, including the accumulative confirmed, deaths and recovered cases from Jan to Aug. The second table stores weather station information and the last one contains weather information like temperature, wind speed and etc. I combined all of them to make a new dataset and used it to do the analysis and prediction later.

I use Kaggle notebook as my development platform. In this project, I applied roll-up and drill down operations from OLAP to select data based on the users’ queries. And use a statistic technique to find out the relation between weather and the spread of covid-19 in this country. Finally, using a time series forecasting methods called Autoregressive integrated moving average to predict the future trend of covid-19 in this country.

Implementation:

Next let’s go to the implementation part.

There are two parts of my implementation. Firstly, I will go through my code and then provides user cases to illustrate the usefulness of my project. Let look at the code file that retrieves and integrates data and this file is called createData. The table that contains covid-19 data will be imported at the beginning. This is how the dataset looks like. And then use API to get the location of each city. This is the updated dataset. Use Bigquery to retrieve station table and weather information. Finally, I will combine all of them based on their common columns. This new dataset is sorted to the Kaggle output folder. This is the integrated dataset. It will be imported to the next part of my code.

I gonna move to the second part.

The integrated dataset from the previous code file will be imported firstly in this analysis\_ prediction file. In this part, I will use three user cases to explain my implementation. The first scenario is that a user wants to know the relation between weather and the growth of COVID-19 in a state. She enters ‘1’. She feels interested about Victoria state in Australia. After entering state names, there are two figures shown to the user. One describes the weather information of Victoria from Jan to Aug, including mean temperature, sea level pressure, wind speed and etc. Another figure shows the daily confirmed, deaths and recovered cases. Also, the data is recorded from Jan to Aug. Looking at figures, she may find there are some relation between them. Since the spread of covid-19 increase as the temperature decreases. After program analyzing the relation, the most positive and negative correlation will be shown to her. She finds the number of daily confirmed cases will be increased in colder weather based on the figure.

Move to the second scenario. A user has a traveling plan to China so he wants to know the spread of covid-19 for the next week. He enters ‘2’ and then type Mainland China. A relation between confirmed, deaths and recovered cases and weather factors are shown to him. He can find there is a negative correlation between the spread of Covid-19 and temperature. He also can find temperature is the most important factor to affect the growth of Covid-19 in China compared to other weather factors. Four more straightforward figures are given to user. All of them show when temperature is high, more confirmed and deaths cases in China. After analyzing, program start to search the optimal parameter to make the prediction. Th median absolute error decreases by 33% and it becomes 10 after using the optimal parameter. The model becomes more accurate. Finally, the user gets the predicted result. It seems like there is an increasing trend of covid-19 in China. He has decided to choose another city.

The last scenario is that a user wants to know the covid-19 situation and how weather affect covid-19 worldwide. She enters ‘3’. And then she can see 2 interactive map about confirmed and death cases. Both show the situation of COVID-19 as time goes. After showing the basic situation, the relation of daily confirmed and deaths cases and temperature are shown to the user.

Finally, move to the conclusion part.

Some improvements can be made to my project. When more data about covid-19 are provided, we should use a seasonal time series forecasting technique to predict covid-19. Since seasonal factors have influence to the spread of covid-19. Seasonal forecasting technique can improve the accuracy. Also, we can implement a GUI instead of a simple command-line interface.

In the conclusion, there is a relation between weather and the growth of covid-19 in most of countries. The spread of COVID-19 tends to increase in colder weather.