Application User Manual

CS 6440 Project #20 - US COVID-19 PREDICTOR

Haosong Ma

Table of Contents

Environment Setup

The environment setup is very easy for this project. You only need a working computer and a modern web browser like Chrome or Edge, and that is it.

You need to be able to use Google Colab on your browser. A Getting Started with Google Collab tutorial can be found here:

https://colab.research.google.com/github/jckantor/CBE30338/blob/master/docs/01.01-Getting-Started-with-Python-and-Jupyter-Notebooks.ipynb.

Although the predictor is very easy-to-use and does not require understanding of Python, if you are completely new to Python and Jupyter Notebook this is a good place to start.

Also make sure you have a stable Internet. Google Collab cannot run locally and it needs good Internet connection to interpret and execute the scripts.
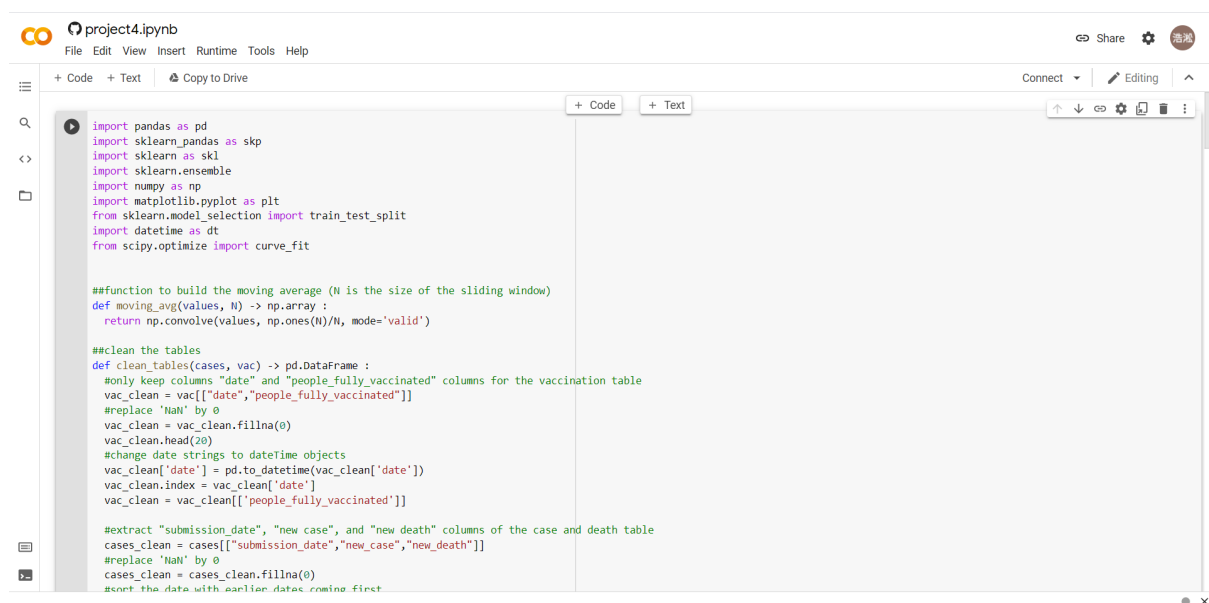
Usage Tutorial

This section will be a step-by-step walkthrough of the program usage.

To begin with, direct to the Google Collab Github page:

https://github.com/haosong830/CS6440/blob/main/finalProject.ipynb. You should be

able to see a blue "Open in Collab" button:
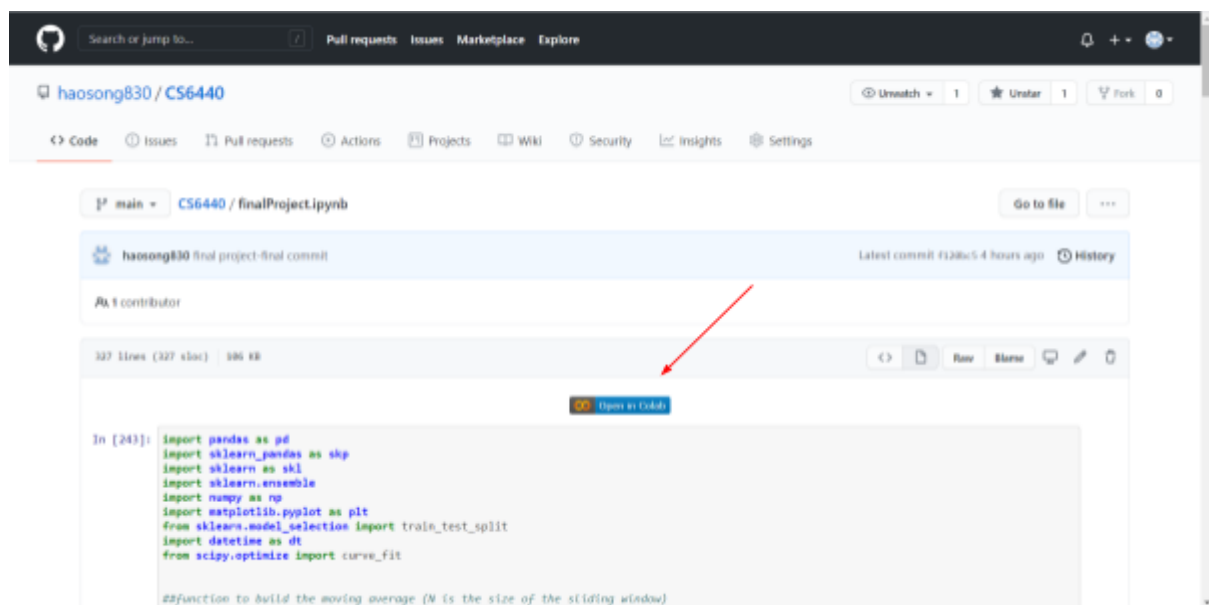
Click the button, it will bring you to the Google Collab Notebook page:
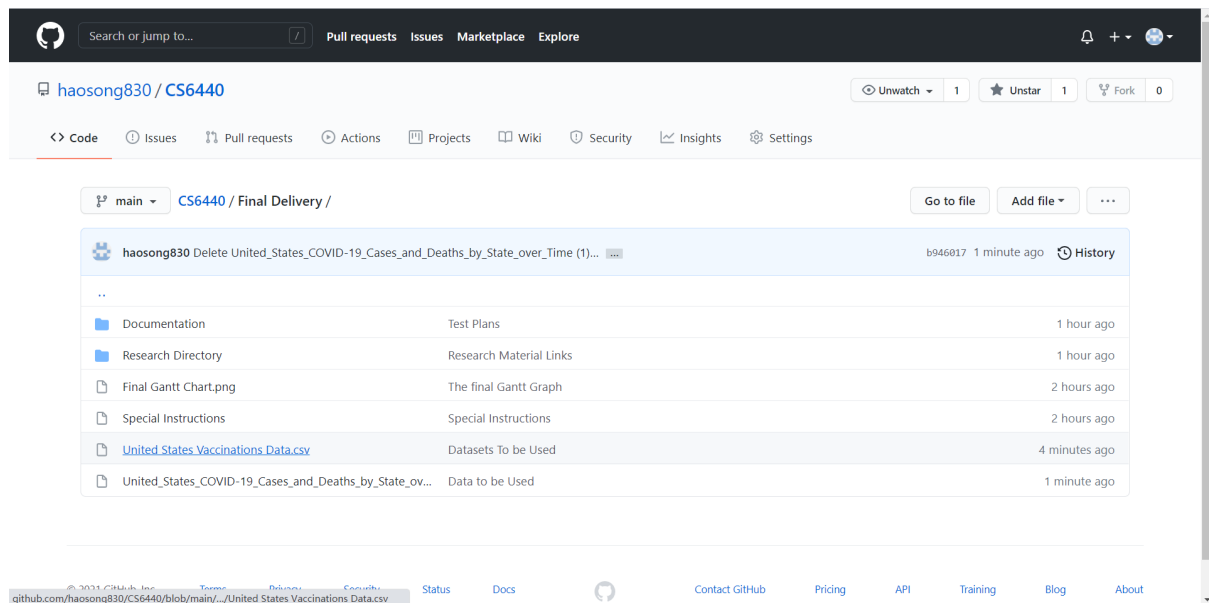

Click the folder icon to open the File Explorer.

In the repository root directory, go to the Final Delivery folder, and download "United States Vaccinations Data.csv" and "United_States_COVID-19_Cases_and_Deaths_by_State_over_Time.csv". These are the data tables we are going to use.
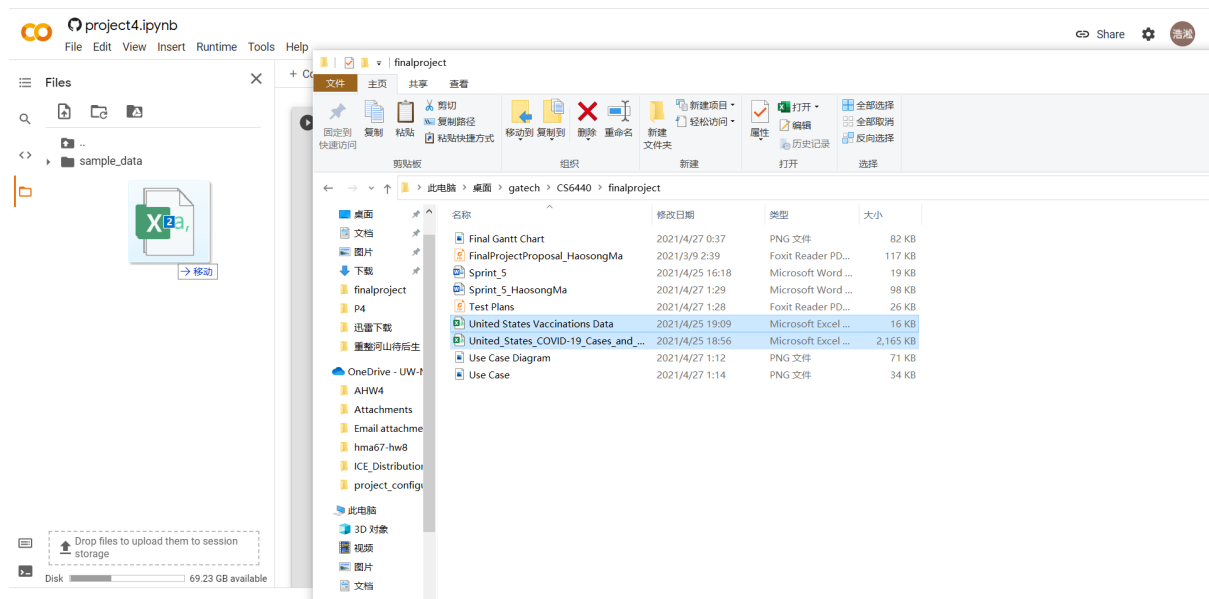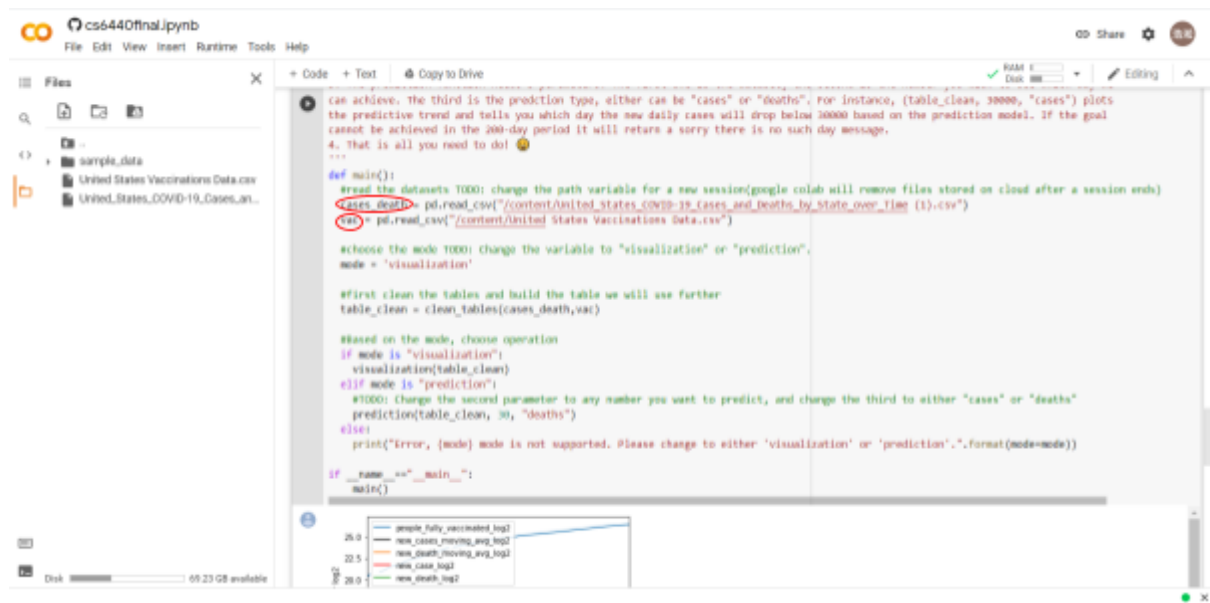
Go back to the Google Collab page, copy and paste these 2 csv files to the Google Collab file explorer(you can do so by dragging and releasing them onto the window pane).


Once the files are shown in the explorer, you first need to change the read_csv() function inputs because they are reading files based on file paths.

The "cases_death" variable corresponds to the "United_States_COVID-19_Cases_and_Deaths_by_State_over_Time" file. The "vac" variable relates to the "United States Vaccinations Data" file. Copy and paste the file paths by right clicking on the files and paste them to the corresponding variables.
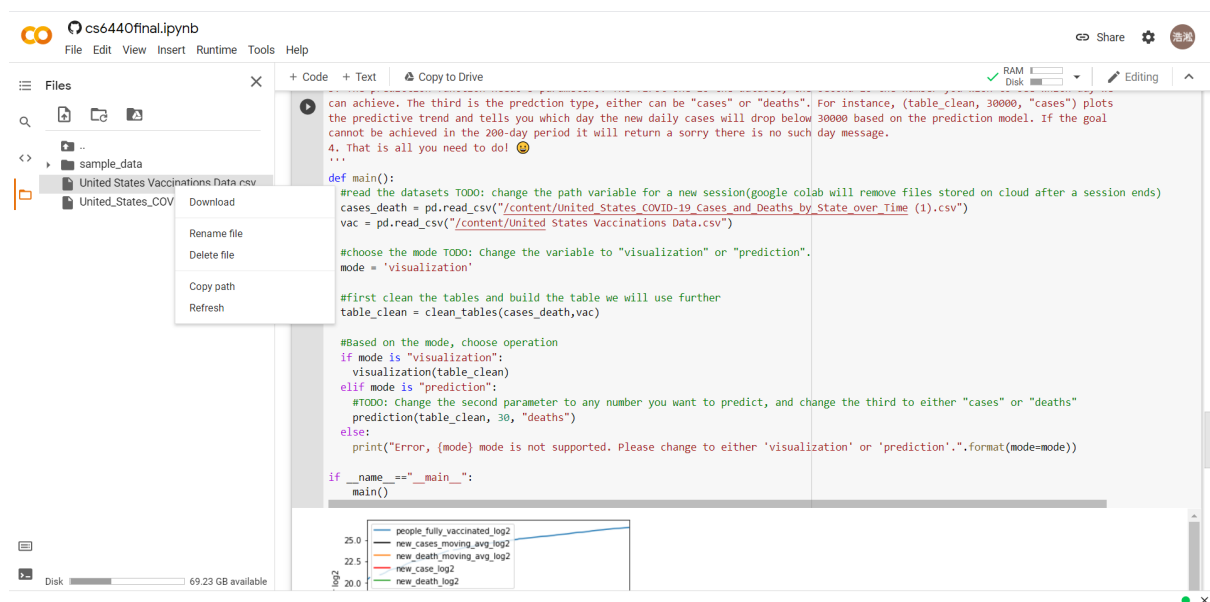
Now you need to change the "mode" variable. The mode variable is "visualization" by default so if you just need data visualization you can leave it there. We will go through visualization first. Click the play button on the top left corner and you may see an alert box popped out. Click "RUN ANYWAY".
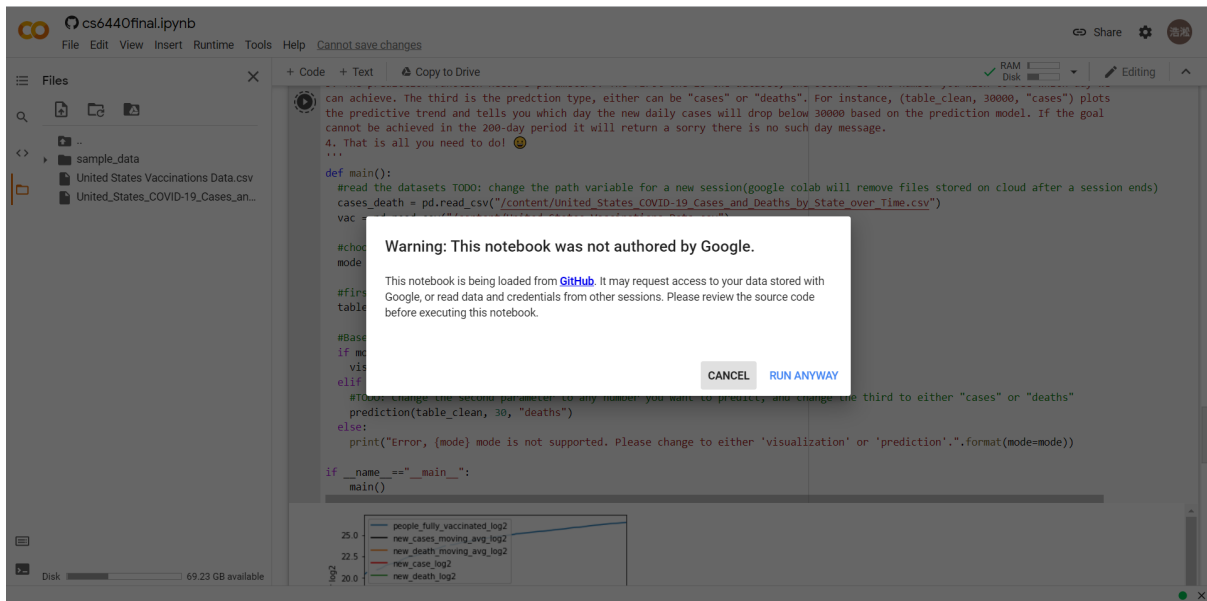
After a few seconds, you should be able to see the graphs plotted out.
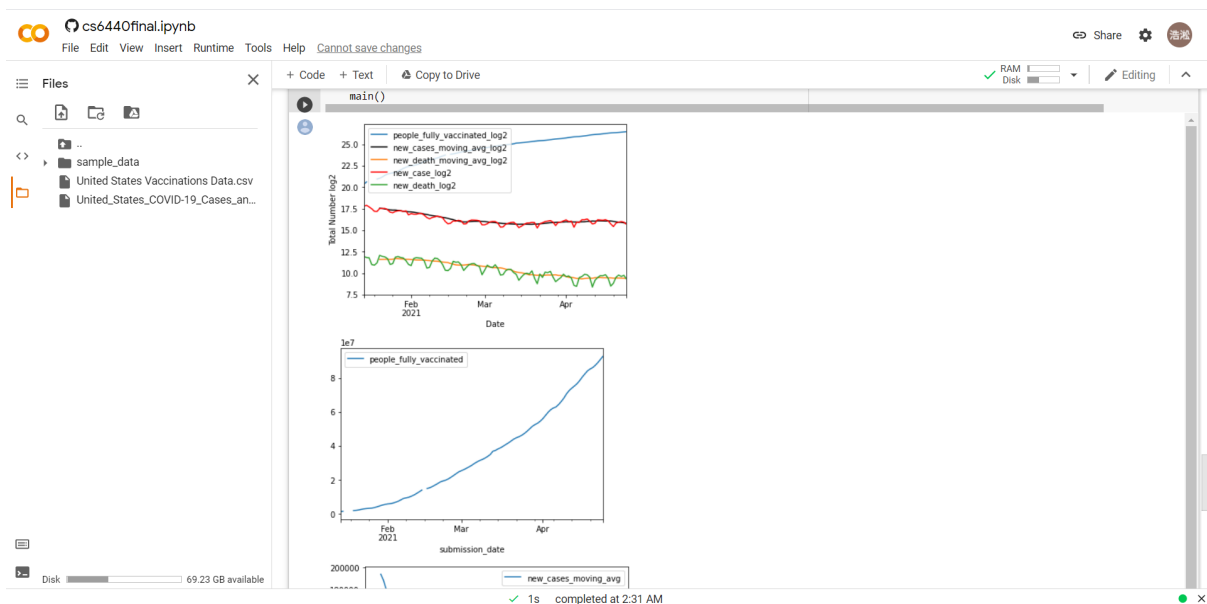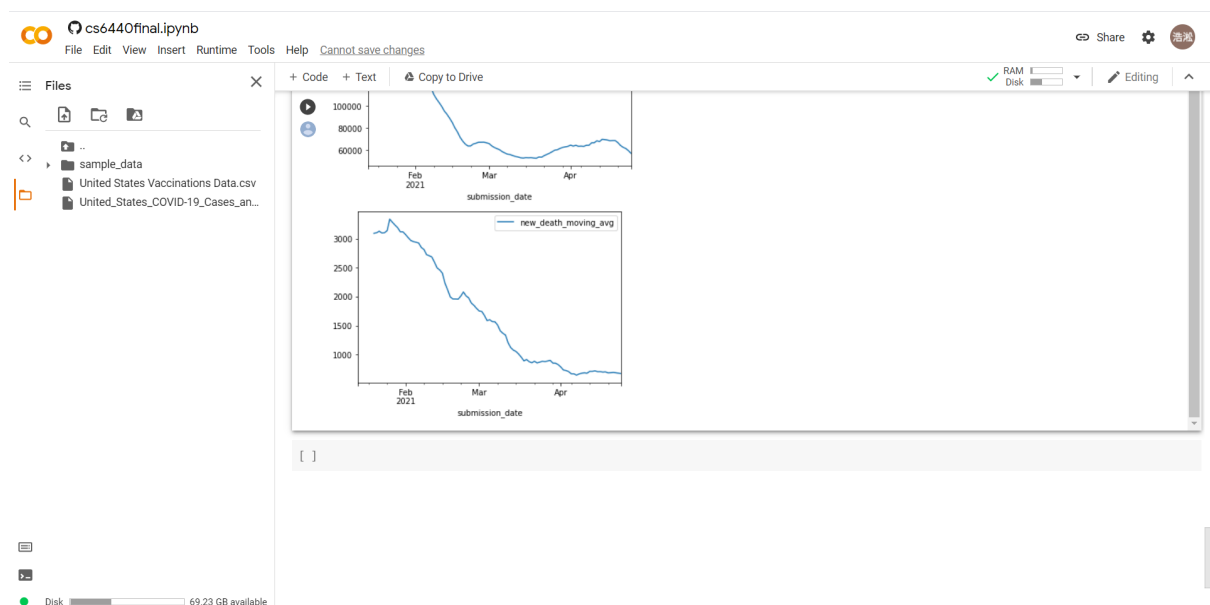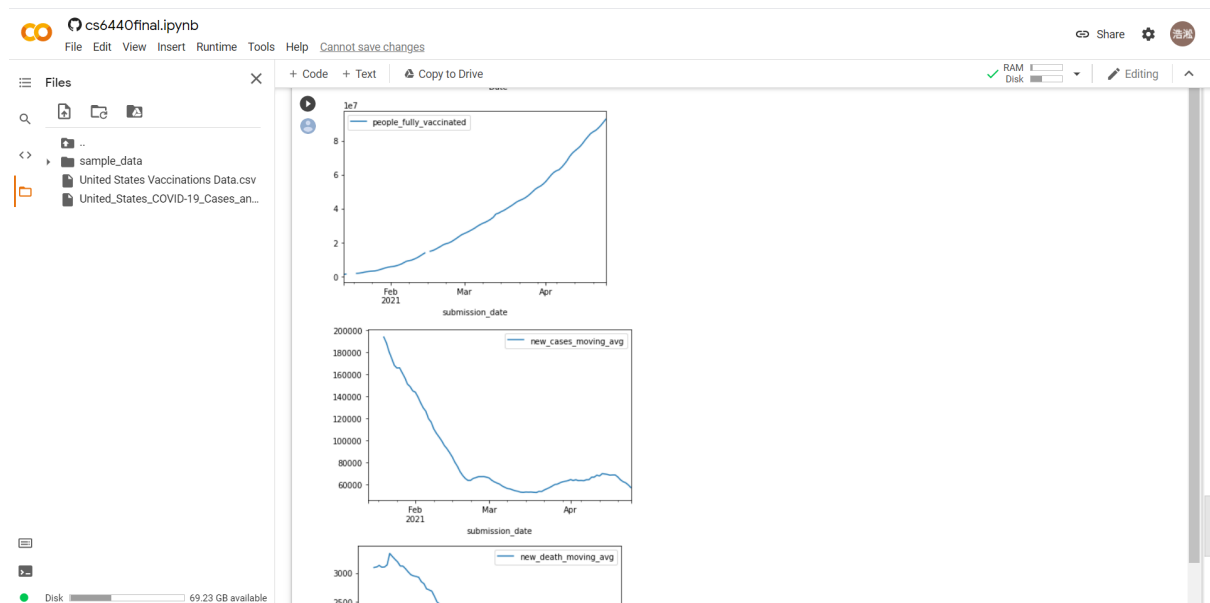
Now let's go through the prediction function. First change the "mode" variable to "prediction"

```
1. Copy the file paths of your TextGrid file and PitchTier file from Google Colab File Explorer (the left window pane)
   and paste them to the corresponding variables.
2. Change the variable "mode" to either "visualization" or "prediction". Visualization plots the new daily cases and new
   daily deaths with the effect of vaccinations of the past. Prediction predicts and plot the trend of new daily cases and new daily
   deaths up to 200 days later.
3. The prediction function needs 3 parameters. The first one is the dataset, the second is the number you wish to see which day we
   can achieve. The third is the prediction type, either can be "cases" or "deaths". For instance, (table_clean, 30000, "cases") plots
   the predictive trend and tells you which day the new daily cases will drop below 30000 based on the prediction model. If the goal
   cannot be achieved in the 200-day period it will return a sorry there is no such day message.
4. That is all you need to do! 😊
'''
def main():
    #read the datasets TODO: change the path variable for a new session(google colab will remove files stored on cloud after a session ends)
    cases_death = pd.read_csv("/content/United_States_COVID-19_Cases_and_Deaths_by_State_over_Time.csv")
    vac = pd.read_csv("/content/United States Vaccinations Data.csv")

    #choose the mode TODO: Change the variable to "visualization" or "prediction".
    mode = 'prediction'

    #first clean the tables and build the table we will use further
    table_clean = clean_tables(cases_death,vac)

    #Based on the mode, choose operation
    if mode is "visualization":
        visualization(table_clean)
    elif mode is "prediction":
        #TODO: Change the second parameter to any number you want to predict, and change the third to either "cases" or "deaths"
        prediction(table_clean, 30, "deaths")
    else:
        print("Error, {mode} mode is not supported. Please change to either 'visualization' or 'prediction'.".format(mode=mode))

if __name__=="__main__":
    main()
```

Scroll down to line where prediction() is called. You see 3 parameters are passed into the function. The first is the data table, and you SHOULD NOT MAKE ANY CHANGES TO IT. The second is a number N that you may wish to predict if there exists days of the 200-day period when new daily cases <= N or new daily deaths <= N. The third parameter is the prediction type, and you should pass either "deaths" or "cases". In the screenshot below prediction(table_clean, 30, "deaths") shows the regression and prediction lines and output strings if new daily deaths can drop below 30 in the future 200-day period. The model has an optimistic and pessimistic version so if the condition is satisfied at least 1 date will be outputted.

Click the run button, you can see the prediction lines and days on the conditions.

You can also play around with other valid parameters and see the results. The plots will always be the same, but not the texts.

```
#choose the mode TODO: Change the variable to "visualization" or "prediction".
mode = 'prediction'

#first clean the tables and build the table we will use further
table_clean = clean_tables(cases_death,vac)

#Based on the mode, choose operation
if mode is "visualization":
    visualization(table_clean)
elif mode is "prediction":
    #TODO: Change the second parameter to any number you want to predict, and change the third to either "cases" or "deaths"
    prediction(table_clean, 30000, "cases")
else:
    print("Error, {mode} mode is not supported. Please change to either 'visualization' or 'prediction'.".format(mode=mode))

if __name__=="__main__":
    main()
```

You want to know the first day that the new cases will drop below 30000
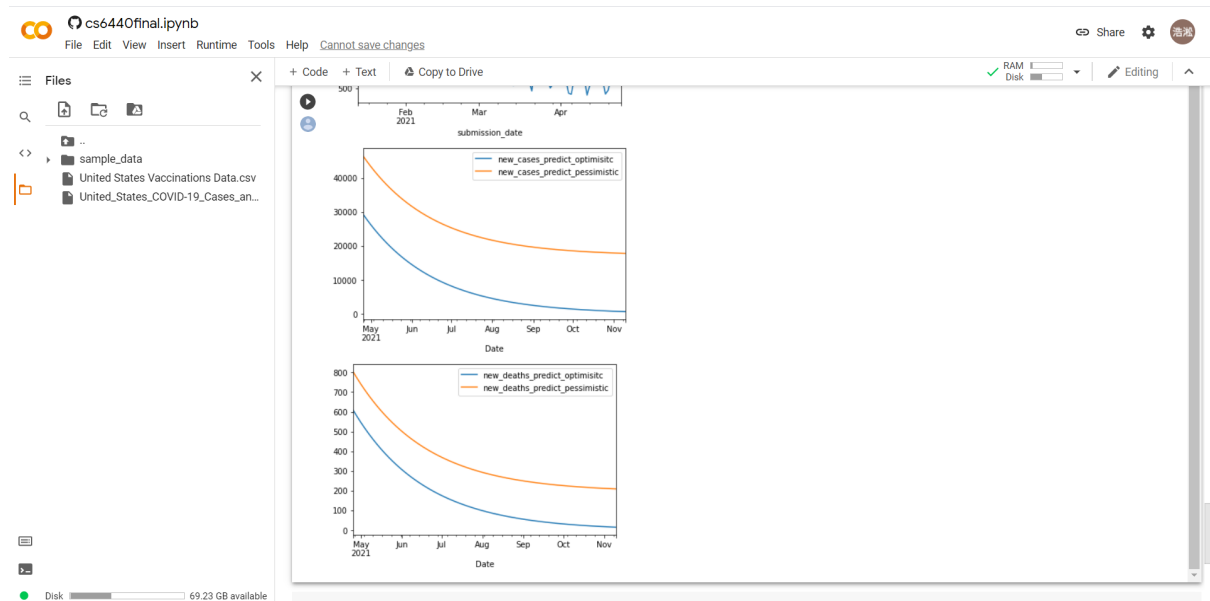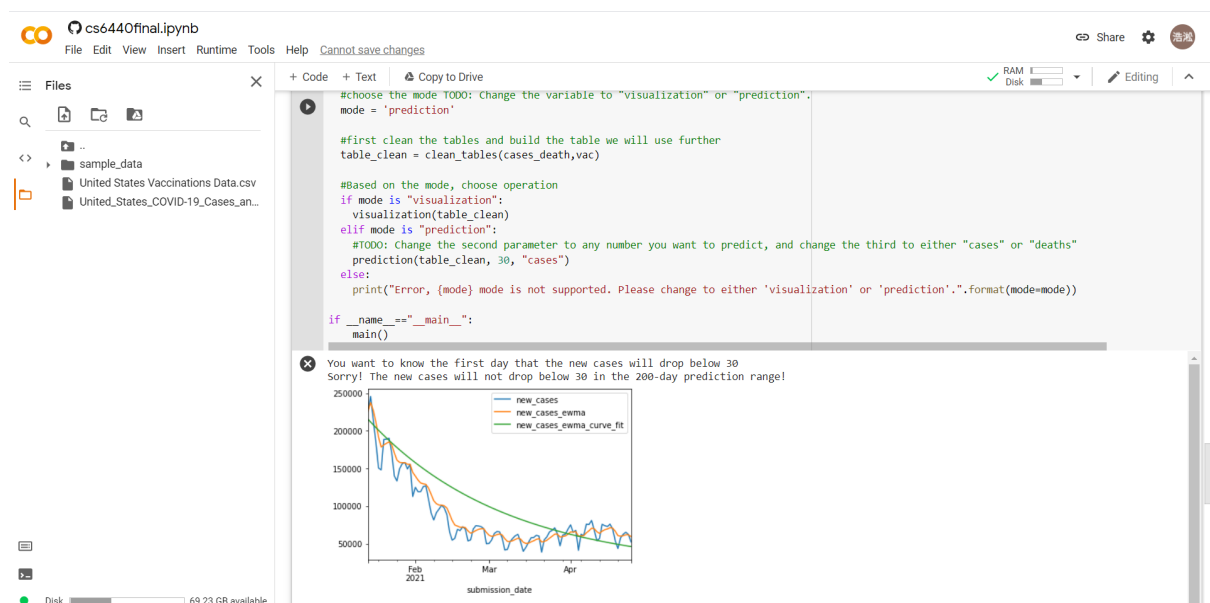The first day that new cases will drop below 30000 based on the optimistic prediction is 2021-04-25 00:00:00
The first day that new cases will drop below 30000 based on the pessimistic prediction is 2021-06-08 00:00:00

Play around with different numbers and prediction types, and see when the US COVID-19 pandemic will possibly settle down!

FAQs

Can I make the files permanently stay on the Collab?

Unless you mount your cloud drive and copy and paste the files to your personal

drive, you need to copy and paste each time you reopen it.

I copied the file paths to the correct variables, why does it still show "file not found"?

Do not use relative paths and always use absolute paths.