Sprint 2:
Project #20 Implementation Plan

Haosong Ma
hma81@gatech.edu

Abstract—Project #20 is an individual project focusing on the employment of health analytics tool kits. Students working on this project are expected to get familiar with various health analytics and visualization tools and provide tutorials of these tools to help the health informatics community. In my personal implementation, I would like to (because I am not assigned any) build a python tool to explore efficacy and side effects of the three CDC-authorized COVID-19 vaccines in the United States (Pfizer, Moderna, and Johnson & Johnson). If the user could provide the vaccine brand (company) and some other demographic parameters (race, sex, age, etc…), the program should be able to return a visualization of the efficacy (#infected after vaccinated / #total vaccinated), side effects (if any), and the immunization rate of one vaccine for a certain demographic group (if the datasets are available). Moreover, if the program is in "prediction" mode, the script could be able to predict the daily positive cases after today and predict the date that the pandemic will settle down (when the daily cases are < 1000 the pandemic would be considered "settled down") using machine learning. After the code is implemented I will provide a detailed walk-through about how to use the script to explore the data.

1 Project design

This section will have 4 sub-threads: Tools and technology, data sources, diagrams and screen mockups. I will walk through some top level design ideas of the project by text interpretation and image visualization.

1.1 Tools and Technology

The program will be written in python or MATLAB and it will be tested in Jupyter Notebook or MATLAB application. The main() function of the program will take in user inputs as parameters and output corresponding visualizations based on these inputs. The following python libraries are essential in the script (MATLAB is not my first choice because I am more familiar with python so I am not listing MATLAB functions here, but if I find MATLAB is easier to use I might switch to MATLAB):

- Numpy
- Pandas
- Matplotlib
- scikit-learn

1.2 Data Sources

I will import datasets from CDC (https://covid.cdc.gov/covid-data-tracker/#vaccination) and from Our World in Data(https://github.com/owid/covid-19-data/blob/master/public/data/vaccinations/us_state_vaccinations.csv). The dataset from Our World in Data is mainly on the immunization rate of the US states and the CDC Covid-19 tracker has detailed information of vaccination based on demographics, vaccine brand and positive cases after receiving the vaccine.

1.3 Diagrams

The entire program consists of 5 major modules, and each module will complete a specific task based on the user command. According to figure 1, the main() function accepts user inputs, and based on the user inputs the execution flow will be directed to different modules.

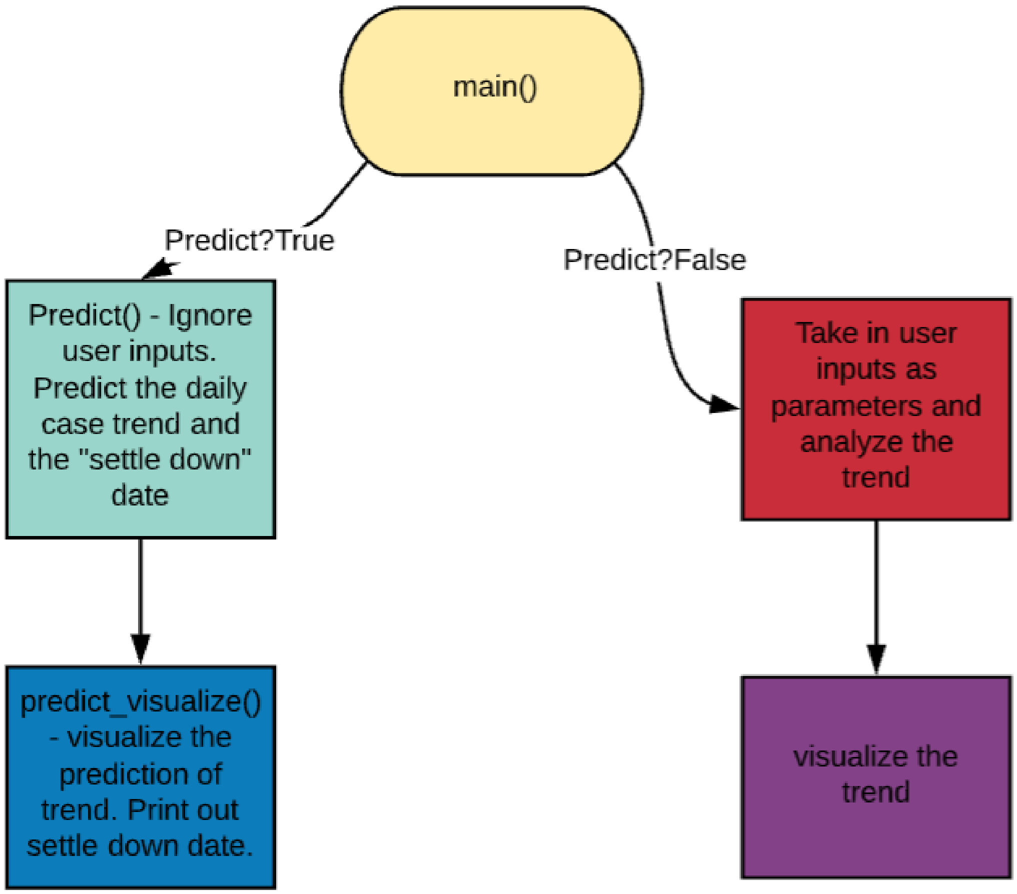
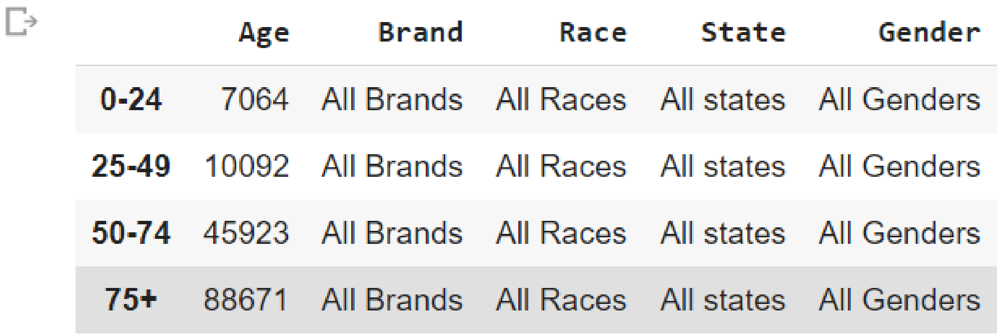

main()
Predict?True
Predict?False
Predict() - Ignore user inputs. Predict the daily case trend and the "settle down" date
Take in user inputs as parameters and analyze the trend
predict_visualize() - visualize the prediction of trend. Print out settle down date.
visualize the trend

Figure 1—The temporary flowchart of the program. Source: LucidChart.

1.4 Screen Mock-ups

It is not realistic to present real demos of how the program will finally work, but I made some dummy data to help illustrate how the script will be like when runned. Figure 2 shows the vaccination trend with regard to age groups, and figure 3 shows the visualization of the prediction of following daily positive cases and the possible “settle down” date.



	Age	Brand	Race	State	Gender
0-24	7064	All Brands	All Races	All states	All Genders
25-49	10092	All Brands	All Races	All states	All Genders
50-74	45923	All Brands	All Races	All states	All Genders
75+	88671	All Brands	All Races	All states	All Genders

Figure 2—The dummy vaccination trend based on age groups.

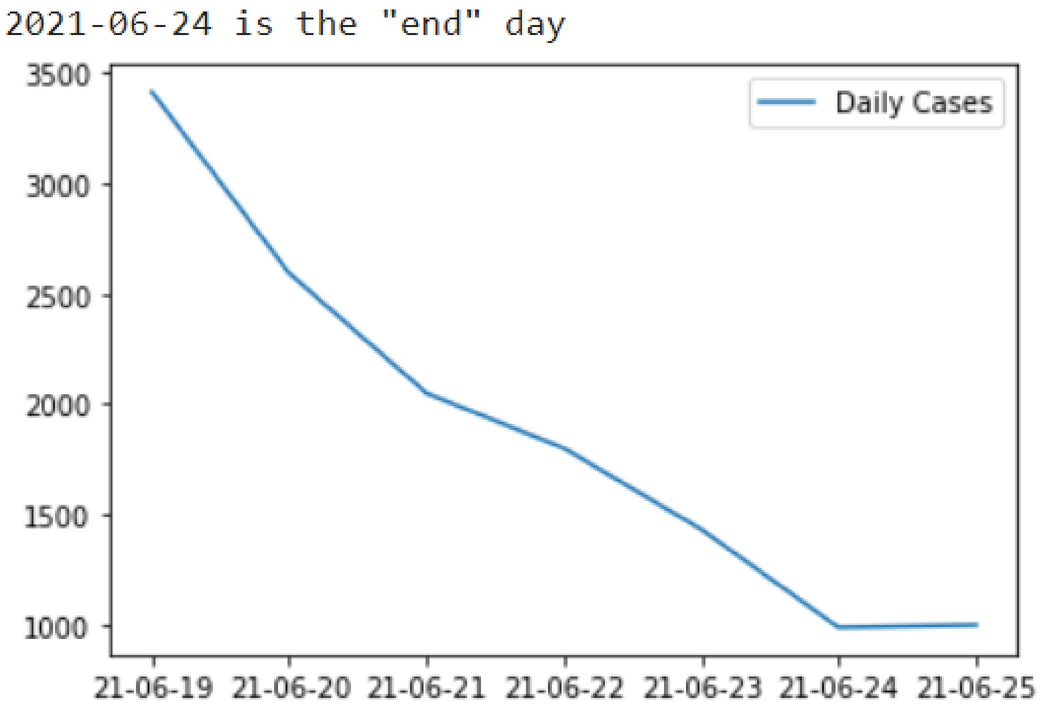


Figure 3—Visualization of daily cases prediction.

2 Implementation Plan

In this section I will list the high-level tasks to be completed in each week and the schedule for them.

2.1 Project Tasks

Different tasks will be assigned to each week and the things to do are listed in the subtitles. The division of the tasks should be subject to change, but the entire implementation process should follow it as the baseline.

2.1.1 *Week 11*

Gather and clean the datasets from CDC and Our World in Data. Make sure all datasets are ready for use and nicely cleaned. Implement the main function.

2.1.2 *Week 12*

Implement the trend analyze module. Test different user inputs and make sure the program is able to handle a bundle of user parameters including edge cases. Start to study how to implement the prediction module using machine learning.

2.1.3 *Week 13*

Start implementing the prediction module with machine learning.

2.1.4 *Week 14*

Finish up the coding of the entire project. Write test cases and test the program thoroughly. Clean up the code and comments and upload to web repositories.

2.1.5 *Week 15 & Week 16*

If coding is still in progress finish up coding in week 15. Wrap up the scripts and datasets and prepare for the presentation and final report.

2.2 Project Timeline

The project schedule will be visualized as a gantt chart in figure 4. Tasks expected to be long and time-consuming are stretched in 2 weeks. Short tasks can ideally be finished in a week.

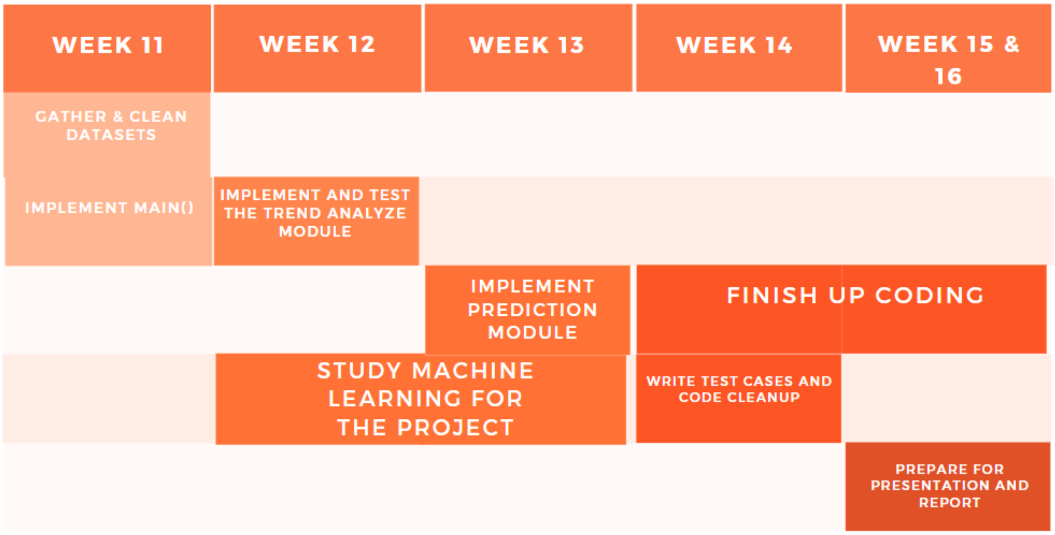


Figure 4—The weekly schedule.

2.3 Needs/ Risks

For the preprocessing (data gathering and cleaning) and coding tasks (implementation, testing and commenting), I need to make sure the program does not have apparent and fatal bugs. Also, it should be easier for the community to use because not all community members are tech-savvy. Studying machine learning algorithms would be a tough task because I need to research a suitable algorithm by myself. This part is risky because there would be a case I cannot figure out a proper algorithm to make predictions.

3 REFERENCES

1. CDC. (2020, March 28). COVID Data Tracker. Centers for Disease Control and Prevention. https://covid.cdc.gov/covid-data-tracker/#vaccinations.
2. owid/covid-19-data.(n.d.).GitHub.https://github.com/owid/covid-19-data.