Summary of what I learned and had to do to complete the project

Soukhna Wade

Bellevue University

Omaha, NE

**Abstract**

The purpose of this paper is to summarize what I learned and had to do to complete the project. COVID-19 has infected more than 9000 people in South Korea. KCDC (Korea Centers for Disease Control & Prevention) announces the information of COVID-19 . The following are the structured datasets based on the report materials of KCDC and local governments:

1. **Dataset1-** PatientInfo.csv: is the epidemiological data of COVID-19 patients in South Korea. Showing 8 to 33 of 2,771 entries(observations), and 18 total columns(variables)
2. **Dataset2-** Region.csv: location and statistical data of the regions in South Korea. This dataset has 244 observations(rows), and 12 variables(columns).
3. **Dataset3-** Weather.csv: data on the weather in the regions of South Korea. This dataset has 24,799 observations(rows), and 10 variables(columns).

In my previous millstones, I analyzed and visualized the above data using various data mining or visualization techniques. But the following lines show what will be done to complete the project.

**What will be done to accomplish all five milestones**

1. Importing the cleaned data from the web

countries = ["korea-south", "united-states", "india", "nicaragua"]

baseURLa = "https://api.covid19api.com/total/country/"

baseURLb = "/status/confirmed?from=2020-01-01T00:00:00Z& to=2020-20T00:00:00Z"

1. Construct questions that lead to a deeper analysis of the following four countries such as korea-south, united-states, india, Nicaragua, which one is the most impacted.

**Merging the Data and Storing in a Database/Visualizing Data**

1. I have downloaded the data from the web by using the following

link: <https://documenter.getpostman.com/view/10808728/SzS8rjbc?version=latest#6fbc46d6-0ddf-400b-a743-a149e9bba381>

1. Get JSON file from the following link by using Postman that I have installed on my computer: <https://api.covid19api.com/total/country/korea-south/status/confirmed?from=2020-03-01T00:00:00Z&to=2020-04-01T00:00:00Z>

"Country": "Korea (South)",

    "CountryCode": "",

    "Province": "",

    "City": "",

    "CityCode": "",

    "Lat": "0",

    "Lon": "0",

    "Confirmed": 10653,

    "Deaths": 232,

    "Recovered": 7937,

    "Active": 2484,

    "Date": "2020-04-18T00:00:00Z"

1. I downloaded and installed my MySQL on my computer and created the database (covid\_data), and the table( new table) where columns are id, country, province, cases, and date.
2. The data is merged in my database
3. Create query = ("INSERT INTO covid\_data.new\_table (country, cases, province, date) VALUES (%s, %s, %s, %s)")
4. Load libraries and used the connect function to connect to the database
5. Get all cases from the database by creating queries
6. Print the results
7. Create the visualizations that demonstrate the data.
8. Plotting the data
9. Show the dataset plot
10. Show the legend (where cases have risen in the USA)