

## **Part 1-GitHub**

### **Aggregation and Visualization of Data for Disease Outbreak Surveillance**

**Introduction** -In the modern world, health informatics is a hotspot area in medical sciences. The treatment criteria for several diseases are still subpar and challenging as well, full of complexity and very high cost, which is not affordable in many people in the world. Furthermore, traditional treatment criteria are time-consuming, and the success ratio is very low especially in lethal or advanced stages of diseases.

**Significance**- Health informatics is an emerging branch of medical science which may not only improve treatment efficiency but also quick, convenient and affordable. With the help of AI, ML and deep learning save billions of dollars every year in the world as well as interoperability system much easier to interpret, diagnosed and efficient for finding potential outcomes for diseases. Synthea and google looker studio of informatics technology helps to reduce the spread of diseases, but also reduce the cost in healthcare setting both for patients and paying organization.

**Problems to be solved** -With the help of health informatics technology, such as Synthea and google looker studio , I am able to find the out most affected region of covid-19 so that precautionary treatment can be measured as earliest and stop the spread of diseases. Resultantly it not only reduced the cost in healthcare setting but also reduced the workload in hospital and waiting time. Many works have already been established like Remote patient monitoring, telemedicine. which can access and handle the patient remotely. Resultantly, it is not only reducing the treatment cost but also reduces the workload as well provides quality and efficient treatment as compared to traditional treatment. And after visualization the data , we can take precautionary measurements.

**Specialization** - I am a health informatics and medical professional with strong knowledge of E.H.R, E.M.R, python, MySQL, and Linux/Ubuntu with a background of pharmacology and ophthalmology. My clinical background along with molecular biology and animal studies

make me a pioneer in the healthcare sector after gaining knowledge of health informatics. With the help of health informatics knowledge such AI, ML, and DL will not help to make me pioneer of medical fields but also beneficial for patients as well I can predict the progression of diseases and avert the diseases in very early stage.

**Personal contribution** – Being a clinicians (optometrist & pharmacologist ), with the help of this health informatics technology, I can advice to concern physician how and where the stop the progression of covid 19 as earliest which would be also very helpful to investigate to different mutant types of COVID-19..

### 1. Python Script:

- ❖ A completed Python script has been provided, which illustrates the output of commands based on the aggregation of data from the student's work. It generates insights using the combined data (attached).

### 2. Comments in the Script:

- ❖ Significant comments have been added within the script for more complicated processes and data points to assist in understanding and learning.
- ❖ The original code provided was not working properly. We created a different script, which involved converting JSON to Python code. Due to the large file size (1481 patients from the Synthea dataset, approximately 9GB), the process took considerable time, and the system hanged frequently.
- ❖ The file was first downloaded from VM software to a personal laptop, and after multiple modifications, the code worked successfully.

### 3. Insights from Results:

- ❖ Two interesting insights found from the script's results include:
  - Identifying COVID-19 cases in relevant cities.
  - Analyzing the extent of COVID-19 outbreaks (high or low infection rates) in those cities.

### 4. Google Looker Studio Visualization:

- ❖ A link to the publicly shared **Google Looker Studio COVID Disease Outbreak Surveillance Page** is included below. This demonstrates how the data visualizations were configured: Link to Looker Studio Dashboard Permission was provided on April 7th.

[https://lookerstudio.google.com/s/pQ\\_eoC8BB9U](https://lookerstudio.google.com/s/pQ_eoC8BB9U)

❖ **Title of the Dashboard:** COVID-19 Monitored Dashboard

❖ **Visualizations Included:**

- ✓ **Plot Map:** Displays COVID-19 cases based on geographic regions - It provides spatial representation of the outbreak, highlighting areas with high or low infection rates. Screenshots have been attached for MGH.
- ✓ **Bar Chart:** Plots the number of COVID-19 cases per city for easy comparison and identification of significant variations. Screenshots have been attached.
- **Dashboard Refresh Interval:** Set to refresh every 12 hours automatically, as it cannot refresh every 15 minute.