Context:

This DIY project is to work on the world's most sensitive situation that has occurred and analysis of data using Data Mining approaches to develop a model is absolutely necessary. This is based on the Corona Virus pandemic and its threat.

You are provided with 2 real life data-sets in the format of CSV files, called : covid_19_data.csv and covid19_line_list_data_modified.

The fields of covid_19_data.csv are as follows:

1. SNo 2. ObservationDate 3.Province/State 4.Country/Region 5. Last Update 6. Confirmed 7. Deaths 8.Recovered

[All the fields are self explanatory from the name]

The fields of covid19_line_list_data_modified are as follows:

1. id 2.case_in_country 3. reporting date 4. summary 5. location 6. country 7. gender 8. age

9. symptom_onset 10.If_onset_approximated 11.hosp_visit_date 12.exposure_start 13.exposure_end

14. visiting Wuhan 15.from Wuhan 16. death 17.recovered 18.symptom

[All the fields are self explanatory from the name]

Activities To Perform:

Perform the following activities on the dataset using Python Programming Language.
You may use any Python libraries as may be needed to complete the operations.

1) Clean, filter and Load data as necessary for analysis.

2) Develop appropriate models using Clustering techniques.

3) Use Data Analysis and mining techniques to develop solutions to queries :

   a. Which is the highest affected area and what is the number. Group from the model, the second highest affected area along with number.

   b. What is the mortality Vs. recovery ratio.

   c. Is there any general tendency towards particular age, gender or random?

   d. What is the mortality rate among different age groups?

4) Develop a simple User Interface including all the queries and processes above to make it a functional system.