Subission Date: 30 December 2022 11.59 PM

1. Business Context

A new coronavirus designated 2019-nCoV was first identified in Wuhan, the capital of China's Hubei province. People developed pneumonia without a clear cause and for which existing vaccines or treatments were not effective The virus has shown evidence of human-to-human transmission. Transmission rate (rate of infection) appeared to escalate in mid-January 2020. As of 30 January 2020, approximately 8,243 cases had been confirmed. As the number of these cases was increasing at alarming rate, the WHO department decided to analyze the COVID data to find out the frequent patterns present in the entire world.

2. Business Problem Understanding

Every country has been affected by the virus's spread, which has caused chaos. Therefore, governments from each nation and WHO have determined that raising people's awareness of the virus' spread would be a first step toward giving them the fortitude to fight pandemics that might arise due to the rise in macho activities.

Data on the death rate, the number of people affected by the disease were gathered in order to retain knowledge of COVID in each nation. For each person to have access to historical statistics, this data needs to be transformed into a form that is useful. Using the data available, trends in the illness distribution attributable to each nation can be discovered.

Your team has been appointed to take a closer look at the records of COVID dataset and analyze the effects caused due to the pandemic.

1) Identifying the spread among the countries and comparing them.

2) Trying to know the reasons for the vast spread among few countries and the consequences caused because of the same.

3. Data Understanding

For this analysis, the department is expecting your team to explore the usage of MongoDB for the storage and querying the COVID_19 data. The data is available in Kaggle. Click here for accessing the dataset.

Below are the datasets which can be used to solve the respective questions / queries which are sub datasets of the main dataset.

Q1) Country, Q2) World, Q3) Day, Q4) Country Q5) Country, Q6) Country, Q7) Country, Q8) Day, Q9) Covid19, Q10) FullGrouped, Q11) Covid 9, Q12) FullGrouped.

4. Data preparation and Exploratory Data Analysis

You are supposed to utilize appropriate data pre-processing techniques on the given data set. If required, make appropriate assumptions and make it explicitly known while using them in the query. Make appropriate selection of the attributes with sound justification for the same. The data set allows for several new combinations of attributes and attributes exclusions, or the modification of the attribute type (categorical, integer, or real) depending on the purpose of the analysis.

5. Expected Outcomes

You are expected to find out the answers to following questions.

1) The number of new cases, new deaths and new recovered

2) The number of death cases in each country of continent Asia and also the corresponding WHO regions

3) The number of deaths that occurred on 12-02-2020

4) The number of active new cases (new cases-(new death+new recovered)) in a reverse sorted order based on the country name

5) The names of the countries with more than 9000 active cases and more than 800 deaths

6) The country with the highest number of active cases and also with second highest death rate

7) The total number of deaths all around the world

8) The number of death cases and active cases between 28-01-2020 and 21-02-2020

9) The latitude and longitude of countries ending with "**ia"** and the number of countries

10) The countries with active cases on 30/03/2020

11) The latitude and longitude of those countries which are having active cases greater than 100

12) The countries and respective dates in which maximum increase of active cases occurred.

The submission should consist of 2 files:

a) **A PDF file containing answer to the thirteen questions based on the analysis that you have carried out earlier along with the supporting MongoDB queries that you have written to extract the answers.**

- Name the PDF file in format like "Grp_<your_group_number>.doc" only. Don't add anything into the file names. Add the group member names in the PDF.

- Make sure that you upload the file well ahead of deadline. At last moments, we have seen several groups have faced issues while doing the submissions.

- Note - Since it is a group assignment, only one submission is expected from each group. Unnecessarily don't upload the solution on individual basis. If it's observed, then a penalty (25% reduction) will be applicable on it.

b) **Every group should record a mp4 video which should contain the executions with queries/answers.**

- Name the video file in format like "Grp_<your_group_number>.mp4" only. Don't add anything into the file names.

6. **MongoDB instance: You can use any instance in lab, local, on Cloud. Some example pointers given below.**

   a) https://www.mongodb.com/try/download/community

   b) https://www.mongodb.com/cloud/atlas/register

7. **References**

   - Covid Data Set
   - MongoDB documentation
   - Groups information