National University of Sciences & Technology
School of Electrical Engineering and Computer Science
Department of Computing

CS 423: Data Warehousing and Data Mining (3+1): BESE-10/BSCS-9 Fall 2022

Assignment #2	
CLO 3: Prepare data (i.e. pre-process data) for the application of data mining techniques	
Maximum Marks: 50	**Instructor:** Dr. Rabia Irfan
Announcement Date: 22nd Nov 2022 (Tuesday)	**Due Date:** 20th Dec 2022 (Tuesday)

Instructions:

- This assignment is to be done in a group of 3-4 students.
- Code file/notebook will be uploaded on LMS via the available link.
- Comments and clear code block will earn you more marks.

Questions:

Download the following dataset from John Hopkins University database for Covid-19:

a. Confirmed: https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv
b. Recovered: https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_recovered_global.csv
c. Death: https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_deaths_global.csv

1. Find the country-wise total number of death, confirmed, and recovered cases. [10 pts]
2. Calculate the country-wise active cases, mortality, and recovery rates. [10 pts]
3. Show the combined outcome of Q1 & Q2 via heatmap. [8 pts]
4. Present the country-wise number of confirmed cases in each month of the year present in the data. [12 pts]
5. Use some visualization techniques to demonstrate the trend of the outcome of Q4 for a period of last one year i.e. Nov 21-Nov 22. [10 pts]