Assignment 2 on Data Visualisation (25%)

Due date: 15/11/2022

Q1. Create an animated time series plot from the given dataset, *time_series_19-covid-Confirmed_archived_0325.csv*, in Moodle to show the evolution of confirmed COVID cases for India from Jan 2020 until March 2020. (Marks: 50)

Requirements:

1. Import the required libraries in Python.
2. Select the Confirmed Cases of Country = “India”/row = 15.
3. Write a function in Python to get the required data and parameters for plotting.
4. Convert the date column to day/month format.
5. Plot the confirmed cases on Y-axis with the dates on X-axis.
6. Animate the plot using matplotlib animation library (Play with the fps and bitrate param values to obtain a smooth animation)
7. Save and display the animated video.

Q2. Apply Principal Component Analysis (PCA) to the **standardised input parameters** of Droplet Data.csv provided in Moodle under Week 8. This dataset has five input parameters, which include area, count, maximum Feret diameter, minimum Feret diameter and perimeter. It also includes the class/category column, which is the response.

Apply K-Means clustering to the most significant PCs (cumulatively explaining over 95% of variance) and plot the most significant PCs for the predicted response as well as the original response. Compare and present your inference.

Apply GMM to the same reduced dataset as in K-Means and present your inference.

(Marks: 50)

All the above should be done in Jupyter Notebook and the notebook file should be shared/uploaded via Github link in Moodle.