NATIONAL INSTITUTE OF BUSINESS MANAGEMENT

In collaboration with **COVENTRY UNIVERSITY**

Advanced Diploma in Data Science 21.1P
July – September 2021

DATA VISUALIZATION WITH PYTHON

Final Assignment 1 (10 %; Due 12th October 2021)

Submission Points:

1. Your final sub missions should be a html file and a Jupiter notebook file.
2. While creating the Final HTML file, please use below instructions:
    - Your HTML file will be used to evaluate the final marks and it is mandatory that the final HTML file should not have any python codes.
    - To create a HTML file by hiding codes:
        - Save your final Jupiter notebook in any directory.
        - Open the anaconda prompt
        - Check whether nbconvert installed.
            - If not **install it**
        - Run below line

- jupyter nbconvert --to html --TemplateExporter.exclude_input=True **"file path of Jupiter notebook"**
    - An HTML file will be saved as final document in the same directory of your Jupiter notebook.
    - Use semi colons at the end of the each code cell to avoid displaying **out[ ]** in Jupiter notebook which will give unintended output in your final HTML file.+

Notes:

3. You can select dataset of your own but with below features:
    - It should have at least 3 numerical variables, 3 categorical variables. (Categorical variables can be nominal or ordinal)
    - It should also have a datetime column
    - Number of rows is arbitrary, but number of columns should be between 7 – 12.
4. Marks will be given for selecting suitable graphs for the datatypes in the questions and adding extra information inside the graphs.
5. **Specially give importance to title, axis names, fig size, color maps, legends and adding extra dimension by introducing another visual variable. (Lecture 4 >> color, size, shape etc)**
6. plagiarism, copying, giving, or receiving aid are strictly not permitted and considered as exam violation. (Jupiter notebook codes will be examined to validate your assignment) but of course you can discuss with your friends and get their help.
7. Selection of graphs is completely arbitrary, but it should explain the question properly.
8. You are expected to use matplotlib but use seaborn if you feel it is necessary.

Questions:

The overall objective is to explain the data to your audience with the help of matplotlib package. But for the marking purpose below guidelines should be followed.

1. Interpret the data and the variables. (25 to 50 words) [10 marks]
2. Interpret why you selected that dataset and what is the message you are planning to give to your audience. (50 - 100 words) [10 marks]
3. Using scatter plots to explain relationship. [7.5 marks]
4. Using line plots to explain a time series or relationship. [7.5 marks]
5. Using Pie Plot to explain a categorical variable with a numerical variable. [7.5 marks]
6. Using Bar Plot to explain a categorical variable with a numerical variable. [7.5 marks]
7. Using insets or subplots of matplotlib at least one time in any of your charts. [10 marks]
8. Annotate any of the above graph using text and patch. (Minimum one patch and minimum one Text) [10 marks]
9. Use at least one 3D chart either **line scatter** or **bar** plot. [10 marks]
10. Overall Idea conveyed to audience using the charts presented earlier with respect to objective you defined in Question 2. [10 marks]
11. Overall presentation of the HTML file. (Using markdown documents, and styling) [10 marks]

Note: You can use charts that are not requested in these questions also if you feel necessary. But including these charts are important.