CP2403: Project – Part 1 – 10%
Data Exploration, Management & Visualization

In Project Part 1, you will required to apply appropriate data management and data visualization techniques for a given scenario to create charts. The techniques for Project Part 1 are covered in Module 1 – Module 4 of the subject. You will have to explain what conclusions you draw from the charts.

Scenario
The California Cooperative Oceanic Fisheries Investigations (CalCOFI) was formed in 1949 to study the ecological aspects of the sardine population collapse off California. CalCOFI conducts quarterly cruises off southern & central California, collecting a suite of hydrographic and biological data on station and underway. The CalCOFI data set represents the longest (1949-present) and most complete (more than 50,000 sampling stations) time series of oceanographic in the world.

The physical, chemical, and biological data collected at regular time and space intervals quickly became valuable for documenting climatic cycles in the California Current and a range of biological responses to them. Data collected at depths down to 500 m include: temperature, salinity, oxygen, phosphate, silicate, nitrate and nitrite, chlorophyll, transmissometer, PAR and C14 primary productivity.

You are provided with the following:-
1. bottle.csv
2. CalCOFI Database Tables Description - Bottle Table.pdf
   (You can also access it via https://new.data.calcofi.org/index.php/database/calcofi-database/bottle-field-descriptions)

Using the dataset and codebook provided, apply appropriate data management techniques.
For this assessment, complete the following four tasks.

1. Select a categorical variable and quantitative variable from the dataset to draw a box plot. What is conclusion can you draw from the box plot?

2. Select a quantitative variable from the dataset to draw a histogram. What is conclusion can you draw from the histogram?

3. Select a quantitative variable from the dataset to draw a line chart. What is conclusion can you draw from the line chart?

4. Select three quantitative variables from the dataset to draw a bubble chart. What is conclusion can you draw from the bubble chart?

5. Go to the link below. Go through the different charts and the corresponding code provided there.

   Top 50 matplotlib Visualizations – The Master Plots (with full python code)
   https://www.machinelearningplus.com/plots/top-50-matplotlib-visualizations-the-master-plots-python/

   Then select one chart/plot from the 50 available and appropriate variable(s) from the dataset provided (bottle.cvs). Then create the selected chart using/modifying the corresponding code provided by the website for the variable data you selected from the dataset (bottle.cvs). What is conclusion can you draw from the chart you have created?
   (Note: you are required to select variable(s) which are different from what you select for previous tasks (1-4))

Hint: Refer to Modules 2, 3 and 4 and Practicals 2, 3 and 4 for help on data management and data visualisation

Ensure you complete, zip and submit both the '**CP2403 - Project - Part 1 - FirstNameLastName.docx**' and '**CP2403 - Project - Part 1 - FirstNameLastName.ipynb**' files to LearnJCU. Ensure you add your FirstName and LastName inside the files and to the file names.

Project – Part 1 (10%) Rubric – Total Raw Marks: 100

	Exemplary (10-9)	Good (8-7)	Satisfactory (6-5)	Limited (4-3)	Very Limited (2-0)
Formulate Investigative Question. 20%	All the questions are well-formed question and can be answered using the available dataset.	Most of the questions are well-formed question and can be answered using the available dataset.	Some of the questions are well-formed question and can be answered using the available dataset.	A few of the questions are well-formed question and can be answered using the available dataset.	Investigative question is meaningless.
Define Categorical and Quantitative Variables. 10%	All the selected explanatory and response variables are valid and connect well to the question.	Most of the selected explanatory and response variables are valid and connect well to the question.	Some of the selected explanatory and response variables are valid and connect well to the question.	A few of the selected explanatory and response variables are valid and connect well to the question.	Variables selected do not relate well to the question.
Create a Good sub-sample of variables for the investigation. 20%	All the sub samples are well represented from different groups.	Most of the sub samples are well represented from different groups.	Some of the sub samples are well represented from different groups.	A few of the sub samples are well represented from different groups.	Poor sampling such as considering only one type of group.
Recoding, Handle missing and NA samples. 10%	Data management techniques are used such as recoding labels and converting quantitative to categorical. Removed missing or NA samples.	Data management techniques are used such as recoding labels and converting quantitative to categorical in most charts. Removed missing or NA samples in most charts.	Data management techniques are used such as recoding labels and converting quantitative to categorical in some charts. Removed missing or NA samples in most charts.	Data management techniques are used such as recoding labels and converting quantitative to categorical in a few charts. Removed missing or NA samples in most charts.	Did not remove missing and NA samples.
Appropriate visualization of charts and legends. 20%	All the charts are easy to read. The title, legends and labels are provided. Correct use of Python commands and arguments.	Most of the charts are easy to read. The title, legends and labels are provided. Correct use of Python commands and arguments.	Some of the charts are easy to read. The title, legends and labels are provided. Correct use of Python commands and arguments.	A few of the charts are easy to read. The title, legends and labels are provided. Correct use of Python commands and arguments.	Charts are difficult to read. Title and legends are missing.

Interpretation of charts and conclusions. 20%	All the charts are accompanied with logical interpretation. The investigative question has been answered.	Most of the charts are accompanied with logical interpretation. The investigative question has been answered.	Some of the charts are accompanied with logical interpretation. The investigative question has been answered.	A few of the charts are accompanied with logical interpretation. The investigative question has been answered.	Limited or missing interpretation of charts. Investigative question is not answered.