Programming for Artificial Intelligence (ATI803)
ACETEL, NOUN

Instructions: Attempt ___all___ questions
Upload your Python Notebook (Colab/Jupyter) with the your report
Submission due by 22nd March 2025

Q1. Using the `london_weather.csv` and the `london_energy.csv` datasets on the Moodle page, answer the following questions by drafting a report:

a. Load the `london_weather.csv` into a dataframe and clean up the data into a useable format. Based on the entries in `date` column, use the .groupby() to group the date into these categorization on a Year_Month form (e.g 1979_Jan, 1979_Feb, 1079_Mar…groups), then:

  i. Find the _highest_ temperature for each of the grouping based on `max_temp`. Do the same computation for `mean_temp` and `min_temp` separately. **[5marks]**

b. Load the `london_energy.csv` into a dataframe and clean up the data into a useable format. Then join this dataset with the `london_weather.csv` datasets on the `date` column. Identify and handle missing values by filling them with the most reoccurring value for that column. After some preprocessing, and perform the following data analytics tasks and compile your observations/results into a report:

  i. What are the top 10 homes which had most energy consumption measurement in kWh for the month of December 2011? Do the same separately for December 2012 and December 2013. **[5marks]**

Q2. Using the `gp-prescribing-july-20242.csv` dataset on the Moodle page, explore the GP prescription dataset for the month of July 2024, and harnessing various Matplotlib and Plotly visualizations to uncover important insights which should be drafted into your report:

  i. Calculate and visualise in a bar chart the top 10 prescriptions made within the month of July 2024. Secondly, create another bar chart to calculate and visualize the 10 least prescriptions made within the month of July 2024. Thirdly, create another bar chart to calculate and visualize the 10 average prescriptions made within the month of July 2024. Use a Matplotlib `fig, ax = plt.subplots()` to layout these bar charts in a single row three columns arrangement. **[5marks]**

  ii. Using a line plot on a single chart, visualize the `Gross Cost (£)` trend for each GP (`Practice`) for the month of July 2024. Limit your selection to only the top 15 GP based on summation of `Gross Cost (£)`. Each GP categorization should be delineated by a distinct line with evident markers. Incorporate hover functionality to present the exact `Gross Cost (£)` figures for each prescription. **[5marks]**

iii. Using a bar plot, compare the summation of `Actual Cost (£)` for prescriptions made by top 15 GP (`Practice`) within July 2024. Assign a distinct hue to each day to delineate daily variations. Incorporate hover details to display the exact sales numbers. **[5marks]**

iv. Set against each other `Total Items` and `Total Quantity` for each prescription across the distinct GPs represented for the month of July 2024 to create a scatter plot for each GP (`Practice`). Randomly select only 20 GP. And use a Matplotlib `fig, ax = plt.subplots()` to arrange all scatter plots into rows and columns. **[5marks]**

Compile these graphs into your report and provide detailed discussion on the patterns represented in the graphs. Ascertain that every graph has fitting x and y-axis denominations. Afford each visualization an elucidative title. Ensure every graph plotted has legend added to it. Refine the color scheme to both captivate and ensure uniformity across all presentations.