COMP1891- Week 4 Tutorial/Lab

Task sheet

The aim of this tutorial is to perform data visualisation with the Python libraries including Pandas, Matplotlib and Seaborn. All tasks should be performed using Notebook.

Task 1:
Using the examples shown in the lecture, perform data visualisation with different types of charts and graphs listed below. Create the **student_df** dataframe for charts shown with student dictionary examples and the **searches_df** dataframe for internet searches examples.

    a) Comparison charts
    b) Distribution plots
    c) Relationship plots

Comment on the results of your illustrative diagrams where applicable by creating a markup cell underneath the task.

Task 2:
Use a bar chart to compare movie ratings of your favourite 5 movies. Use movie rating sites for movie scores (for example a score of 6 out of 10 (or 60%) given to a movie).
Compare the rating scores among the five movies. You may consider the following:

1. Use pandas to read the data/create dataframes.
2. Use Pandas/Matplotlib or Seaborn to create a visually appealing bar plot comparing the scores for all five movies.
3. Use the movie titles as labels for the x-axis. Add a legend and a suitable title to your plot.

Task 3:
Using the IQ scores data (**IQ.csv**) visualise the IQ test performance of adults by plotting histogram and box-plot as follows:
    a) Determine the mean and standard deviation of IQ scores.
    b) Plot a histogram with 10 bins for the given IQ scores. Add labels and a title. What can you learn about the data from the output?
    c) Plot a histogram with KDE using seaborn. Comment on the outputs of your plots in (b) and (c) if there are any noticeable differences.
    d) Create a box plot to visualize the same IQ scores. Add labels and a title. Comment on the distribution of data with regards to 5 key components of the box plot. Are there any outliers?

Task 4:

Extend task 3 by adding IQ scores for three different groups of adults in the data file. Visualise and compare the performance across the groups via box plots, i.e.,

    a) Create a box plot for each of the IQ scores of the different test groups. Add labels and a title. Comment on your findings.

    b) Are there any outliers in the dataset. Follow the code sample shown in the lecture to identify and remove the outliers.

Task 5:

During the first 20 days of the COVID-19 pandemic, the number of newly infected patients per day in a country were recorded as below:

       174, 335, 278, 214, 422, 513, 737, 672, 489, 412, 1301,
       1105, 1123,1376, 1502, 894, 665, 1704, 1656, 1342

Display a bar chart showing the number of newly infected patients for each day. Include title and labels.

Task 6:

Using the housing property data (**Property.csv**) determine the relationships among variables by performing the following visualisations:

    a) A scatter plot between sqrt_living and price

    b) A pairplot to show relationships among the pairs of variables in the dataset

    c) A heatmap to calculate and display the correlation coefficient values among the pairs of variables in the dataset.

    d) Comment on your findings (i.e. strengths of relationships) in (a), (b) and (c).

Task 7:

By using any suitable data create a visualisation example of a composition chart (e.g., pie chart) of your choice.