In this assignment, you are required to design and implement an exploratory data analysis (EDA) project.

**Datasets**

You can choose one of the two datasets below for your analysis:

1. **Birmingham Parking Dataset**

This dataset is from Birmingham City Council, and provides the usage of car parks operated by National Car Parks (NCP) in Birmingham. It includes half-hourly occupancy of the car parks for three months (October-December) in 2016. It is a widened version of the original dataset which can be found on the [UCI data repository](https://archive.ics.uci.edu/ml/datasets/Parking+Birmingham). The specific dataset version to be used for this project can be found in the attached ***"BhamParking.xlsx"*** file and the variable definitions are included in the second Excel worksheet (variable definitions).

1. **UK Vaccinations Dataset**

This data contains information about the numbers of people who took first, second and third doses of the COVID-19 vaccination in different regions of the UK. The data is compiled from the UK government’s [Coronavirus Vaccinations page](https://coronavirus.data.gov.uk/details/vaccinations) and covers a period of time from earlier 2021 to mid 2022.  The attached "UK\_VaccinationsData.xlsx" file contains both the data (first worksheet named ‘data’) and descriptions of the data variables (second worksheet named ‘variables' descriptions’)

**Format of submission**

You are free to choose a programming language to use for the assignment- either Python or R. You can also use Excel, however, do note that many types of the expected analysis are not directly available in Excel and you will need to do quite a lot of manual manipulation of the data if you do use Excel. You can prepare your project in the form of a Jupyter notebook (Python), RMarkdown notebook (R) or an Excel file.  With any of the three options, the final submission should be converted to a PDF file (if using Excel you may consider inserting relevant formulae, tables and plots into a MS Word document before converting it to PDF).

The pages in the file should be in the portrait orientation.

The word limit is 1500 words +/-10%, excluding the code and any code output. To count the number of words in Markdown cells in a Jupyter notebook, you may find it useful to use this code:  [Word count in Jupyter.ipynb](https://vle.aston.ac.uk/bbcswebdav/pid-3624823-dt-content-rid-25875867_1/xid-25875867_1) . To count the number of words in Markdown in an Rmd notebook, please see this code:  [Word count in Rmd.Rmd](https://vle.aston.ac.uk/bbcswebdav/pid-3624823-dt-content-rid-25875868_1/xid-25875868_1) .

**Steps of the EDA**

The EDA project should include the following analysis steps in their sequence:

1. Generate descriptive statistics for the dataset, and comment on the main trends.

2. Check any records with missing values and handle the missing data as appropriate.

3. Build graphs visualizing the following and comment on the obtained visual insights

1. the distribution of one or more individual continuous variables
2. the relationship of a pair of continuous variables.
3. the association b/w a categorical variable and a continuous one.
4. The relationship between more than two variables, e.g., using semantic mappings.

4. Display unique values of a categorical variable and their frequencies.

5. Build a contingency table of two potentially related categorical variables. Conduct a statistical test of the independence between them and interpret the results.

6. Retrieve one or more subset of rows based on two or more criteria and present descriptive statistics on the subset(s).

7. Conduct a statistical test of the significance of the difference between the means of two subsets of the data and interpret the results.

8. Create one or more tables that group the data by a certain categorical variable and display summarized information for each group (e.g., the mean or sum within the group).

9. Implement a linear regression model and interpret its output including its accuracy

**Note**: The EDA report should follow the above sequence of steps in their order with clear headers. Each step of the analysis has to be documented with comments that: describe what the step is meant to achieve, justify the implementation choices, and interpret the result of the step. Besides, the project report should start with an introduction section motivating the project (i.e., the issues to explore and their practical implications), and should conclude with a conclusion section summarizing key findings and learning.

And before you start to work on this assignment, please familiarise yourself with the detailed evaluation criteria for this assignment by studying the Assessment Brief (attached).