**Assignment Report**

**Task 1**

**Dataset Name:** Exploring Global AI, ML, Data Science Jobs Overview for 2023

**1. Dataset Selection:**

I chose the dataset from [ai-jobs. Net](https://ai-jobs.net/salaries/download) provides information about AI, ML, and Data Science salaries. This dataset is valuable for understanding trends, job titles, and experience levels in the industry for the year 2023.<https://www.kaggle.com/datasets/lorenzovzquez/data-jobs-salaries/versions/3/data>

**2. Initial Setup:**

The code begins with the necessary imports, including libraries like NumPy, Pandas, and Matplotlib. Additionally, it checks the file directory to ensure the correct file path.

**3. Data Loading and Inspection:**

The dataset is loaded using Pandas, and its structure is inspected using the `shape` attribute to get the number of rows and columns. The code then checks for missing values and duplicated rows, removing the latter to ensure data cleanliness.

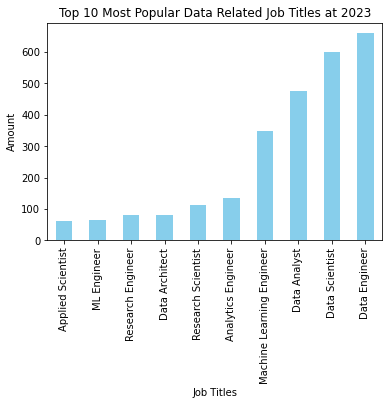
**4. Exploratory Data Analysis (EDA):**

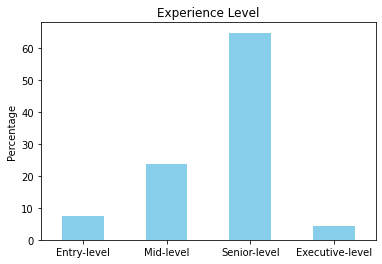
- The code counts the number of unique job titles in the dataset.

- It focuses on 2023, extracting and visualising the top 10 most popular data-related job titles using a bar chart.

**5. Visualizing Experience Levels:**

The code analyses and visualises the distribution of experience levels in the dataset, showing the percentage of each group. The bar chart is labelled with relevant information.





**Task 2  
Dataset Name:** Exploratory Data Analysis of Global Food Price Estimates

**1. Purpose:**  - The code aims to perform Exploratory Data Analysis (EDA) on a global food price estimates and inflation information dataset. The dataset includes monthly data by product, market, and country.

**2. Dataset Information:**

- Source[https://www.kaggle.com/datasets/anshtanwar/monthly-food-price-estimates/data]

- Files: The dataset consists of three CSV files:

- Monthly Food Price Inflation Estimates By Country

- Monthly food price inflation estimates aggregated for all food products

- Monthly food price estimates by product and market

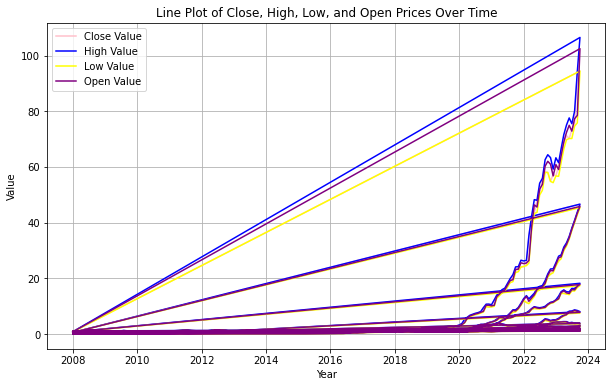
**3. Data Cleaning:**

- Handling Missing Values:

- The code identifies and addresses missing values in the 'Open', 'Close', 'High', and 'Low' columns by filling them with the respective mode values.

- Remaining missing values are dropped.

**4. Time Series Visualization:**

- Plotted a line chart showing the trends of 'Close', 'High', 'Low', and 'Open' prices over time, with the x-axis representing years.  


**Task 3**

**1. Dataset Selection:**

The dataset used in this project is the "Used Car Prices in the UK" dataset. It is sourced from the popular automotive marketplace website autotrader.co.uk. The dataset comprises 3,685 data points, each representing a unique vehicle listing.

[https://www.kaggle.com/datasets/muhammadawaistayyab/used-cars-prices-in-uk]

**2. Initial Setup:**

The initial setup involves importing necessary libraries and defining a function (`plot\_box`) for generating box plots. The libraries include:

- `pandas` for data manipulation

- `matplotlib.pyplot` for plotting

- `seaborn` for enhanced data visualisation

- `sklearn.preprocessing.LabelEncoder` for label encoding categorical variables

**3. Data Loading and Inspection:**

The dataset is loaded into a Pandas DataFrame using the `pd.read\_csv` function. The unnecessary column, "Unnamed: 0," is dropped from the DataFrame. This step ensures a clean and structured dataset for further analysis.

**4. Exploratory Data Analysis (EDA):**

Data Preprocessing and Label Encoding:

Categorical values in the "Emission Class" and "Fuel Type" columns are mapped to numerical values. Additionally, non-ordinal flat features, such as "Engine" and "Body Type," are label-encoded using `LabelEncoder`.

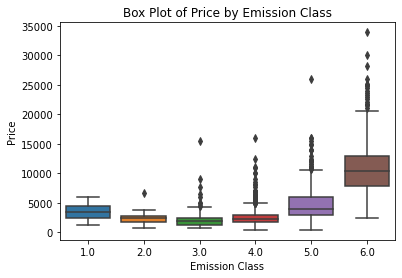
Correlation Analysis:

A correlation matrix is computed for relevant numerical columns, providing insights into the relationships between variables. This aids in understanding how different features are correlated with the target variable, "Price."

**5. Visualizing Experience Levels:**

**Box Plot for Price by Emission Class:**

A box plot is generated to visualise the distribution of car prices based on emission classes. This visualisation helps identify potential patterns or trends related to emission classes and their impact on car prices.



**Box Plot for Price by Mileage Group:**

The "Mileage(miles)" column is converted to an integer type, and values are grouped into specific ranges. A box plot is then created to explore the relationship between car prices and mileage groups. This visualisation provides insights into how mileage may influence used car prices.

