Phase two

Data analyst

intro

Anaconda ➔ is a distribution of the Python and R programming languages for scientific computing, that aims to simplify package management and deployment. It comes with a large collection of pre-installed packages used in data science, machine learning, and scientific computing.

Jupyter ➔ is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text. It supports various programming languages, including Python, R, Julia, and others.

Jupyter Notebooks are the primary interface for working with Jupyter. These notebooks allow you to write and execute code in individual cells, view the results, and add formatted text, equations, and visualizations. Jupyter Notebooks have become very popular in data science and scientific computing due to their ability to create interactive and reproducible computational narratives.

1-NumPy

NumPy ➔ is a Python library that provides support for mathematical and scientific operations with large multidimensional arrays and matrices.

Usage of NumPy ➔ First import the numpy library

1.Creating NumPy Arrays using np.array() function.

2. Some of Attributes

```
print(arr2d.shape)   # Shape of the array
print(arr2d.size)    # Number of elements in the array
print(arr2d.dtype)   # Datatype of the array
```

3. Array Operations ➔ + , * , - , multiplication using np.dot(arr1,arr2)

5. Array indexing ➔

```
arr = np.array([[1, 2, 3], [4, 5, 6], [7, 8, 9]])
print(arr[0, 0])      # Access element at row 0, column 0
print(arr[1])         # Access entire second row
print(arr[:, 0])      # Access entire first column
```

End