Page No.
Date

Assignment 2

Q.1 a) Explain how watson studio will provide the environment to solve bussiness problems.

→ (1) Watson studio provides you with the environment and tools to solve your business problems by collaboratively working with data.

(2) You can choose the tools you need to analyze and visualize data, to cleanse and shape data, to ingest streaming data, or to create and train machine learning models.

(3) Visualizing information in graphical way can give you insights into your data.

(4) By enabling you to look at and explore data from different perspectives, visualization can help you identify patterns, connections, and relationships within that data as well as understand large amounts of information very quickly.

(5) Watson studio is an IBM cloud-based platform designed to help businesses and data professionals through data science and machine learning.

(6) It provides an environment that facilitates problem solving in several ways like data preparation, model development, scalability and ex..

Page No.
Date

Q.1 b) What is data types ? List out the types of data types with example.

→ Data type describe the characteristic of a variable. It defines what type of data we are going to store in variable.

(1) number :→

There are mainly three types which includes,

a=5 → integer number.

b=2.5 → float number.

c=6+2j → complex number.

(2) String :→

String is an ordered sequence of character. we can use single or double quotes to represent string. Example: Str = "Welcome"

(3) List :→

A list can contain a series of values. It declared using [] . List is mutable , which means we can modify the list.

Example: list = [2, 3, 4, 5.5, 'Hi']

(4) Tuple :→

A tuple is a sequence of python objects separated by commas. Tuples are immutable , which

means tuples once created cannot be modified.
They defined using parenthesis ().
Example: tuple = (50, 25.6, "python")

(5) Set: →
A set is an unordered collection of items.
Set is defined by values separated by a commas inside braces { }.
Example: set = {5, 1, 2.6, "python"}

(6) Dictionary: →
Dictionary items are stored and fetched by using key. They are used to store huge amount of data. It contain key: value pair.
Example: dict = {1: 'Hi', 2: 7.5, 3: "Hello"}

Q.2a) What is dictionary? Explain the methods available in dictionary.
→ (1) Dictionary are used to store data values in key-value pairs.
(2) A dictionary is a collection which is ordered, changeable and do not allow duplicates.
(3) Creating dictionary: →
With many key-value pairs surrounded in curly brackets and a colon separating each key
Syntax: Dict = { "Name": "Sahil", "Roll No.": 140 }

Page No.
Date

(4) Accessing the dictionary Values :→
The key of dict can be used to obtain the values because they are unique from one another.
Ex. Emp = {"Name": "Dev", "Age": 20}
print ("Name : %s" %. Emp["Name"])
print ("Age : %d" %. Emp["Age"])

(5) Deleting element using pop() method is one of the way to get rid of elements from dict.
Ex. Dict = {1: "Hello", 2: "Key", 3: "Hii"}
pop - key : Dict.pop(2)
print (Dict)

(6) Iterating dict. can be iterated using 'for' loop
Ex. Emp = {"Name": "Sahil", "Age": 21}
for x in Emp:
print (x)

(7) Delete keyword can be used to in-place delete key that is present in dict in python.
dict = {"Arushi": 22, "mani": 21}
del dict ["mani"]

Page No.
Date

Q. 2b) Differentiate betn numpy arrays and lists.

→

	Numpy Array	List
(1)	It is the core library of python which is used for scientific computing	The core library of python provides list.
(2)	It contain similar datatypes	It contain different datatypes
(3)	It is homogeneous	It is both homogeneous as well as heterogeneous
(4)	It is faster as compared to list and also have some optimism function	It does not have optimism function and also slow as compared to Numpy Array.
(5)	It required smaller memory consumption as compared to python list	It required more memory as compared to Numpy array.
(6)	In this element wise operation is available	Element wise list is not possible on the list.

Page No.
Date

Q. 26) Differentiate betn numpy arrays and lists.

→

Numpy Array	List
(1) It is the core library of python which is used for scientific computing	The core library of python provides list.
(2) It contain similar datatypes	It contain different datatypes
(3) It is homogeneous	It is both homogeneous as well as heterogeneous
(4) It is faster as compared to list and also have some optimism function	It does not have optimism function and also slow as compared to Numpy Array.
(5) It required smaller memory consumption as compared to python list	It required more memory as compared to Numpy array.
(6) In this element wise operation is available	Element wise list is not possible in the list.

Page No.
Date

Q.3 a) Explain the following term:

i) Raw code :→

① In data visualisation, raw code typically refers to the direct programming code or scripts used to create visualizations from raw data.

② This code can be written in many programming languages like python, R and etc.., depending on the visualization library or tool being used.

③ Raw code in data visualization provides flexibility and control over the design and presentation of visualizations.

ii) Waffle chart :→

① A waffle chart is an interesting visualization that is normally created to display progress towards goals.

② It is a commonly an effective option when you are trying to add interesting visualization features to a visual that consists mainly of cells, such as excel dashboard.

Page No.
Date

Q.3b) How will you create the normalized weight?

→ (1) Identify the Range :- Determine the range of values in your dataset. For example: find minimum & maximum values.

(2) Choose a Normalized method : There are various methods to normalize data, depending on your specific needs. Two common methods are:

a) min-max scaling :

$$\text{Normalized value} = \frac{\text{Value} - \text{minimum}}{\text{maximum} - \text{Minimum}}$$

b) z-score standardization :

$$\text{Normalized value} = \frac{\text{Value} - \text{mean}}{\text{std. deviation}}$$

(3) Apply the chosen method :- Use the chosen normalized method to scale your data. This will give you normalized values that represent the relative importance or weight of each data point within the chosen range.

(4) Visualization : These normalized values will make it easier to compare and interpret the imp. of different data points or categories.

Page No.
Date

Q.4 a) Explain pie chart specialized visualization tool using matplotlib.

→ (1) Import matplotlib: Being by importing matplotlib in your python script or Jupyter Notebook.
import matplotlib.pyplot as plt.

(2) Prepare your data: You need to have the data you want to visualize in the form of numerical values. labels = ['category A', 'category B', 'category C', 'category D']
sizes = [25, 30, 15, 30]

(3) Create the piechart: Use the plt.pie function to create the pie chart.
plt.pie(sizes, labels=labels, autopct="%1.1f%%", startangle=90, shadow=True)
'autopct': Specifies the format for displaying the percentages on each slice.
'startangle': Controls the starting angle for the first slice.
'shadow': Adds a shadow effect to the chart.

(4) Display the pie chart: You can use the 'plt.show()' to display the piechart.
plt.show().

Page No.
Date

Q.4 b) Explain Bubble plots specialized visualisation tool using matplotlib.

→ (1) Import matplotlib: import matplotlib.pyplot as plt.

(2) prepare your data: you'll need data with three variables x-values, y-values and third variable that represent the size of the bubbles.

(3) create the bubble plot: Example

```
x = [1, 2, 3, 4, 5]
y = [10, 15, 13, 17, 20]
bubble_size = [100, 200, 150, 250, 300]
plt.scatter(x, y, s=bubble_size, alpha=0.5, c='blue',
            label='Bubble plot')
plt.xlabel('x-axis label')
plt.ylabel('y-axis label')
plt.title("Bubble plot")
plt.legend()
plt.show()
```

(4) In this example: "scatter" is used to create the bubble plot. The 's' parameter specified the size of each bubble based on the 'bubble-size" data. 'alpha' control the transparency of the bubbles. 'c' specifies the color of the bubbles. 'label' sets a label for legend.

Page No.
Date

Q.6 a) Explain the following term:

i) Regression plots :→

- The regression plots in seaborn are primarily intended to add a visual guid that help to emphasize pattern in a dataset during EDA.
- Regression plot as the name suggest creates a regression line between 2 parameters and help to visualise their linear relationship.
- Regression plots in seaborn can be easily implemented with the help of lmplot() function

ii) matrix plot :→

- Matrix plots, also known as heatmap matrices, are a type of data visualization techniques used to display the relationship between multiple variables in a dataset.
- They are particularly useful for exploring correlations or patterns in multivariate data.
- They can be used in various fields, including data analysis, biology and more to uncover hidden patterns

Page No.
Date

1.6 (b) Explain the following terms.

i) Distribution plots :→

- Distribution plot visually asses the distribution of sample data by comparing the empirical distribution of the data with the theoretical values expected from a specified distribution.
- They are graphical representations that help you understand the distribution or spread of a dataset.

ii) Categorical plots :→

- categorical plots are a type of data visualization used to display the distribution of data across different categories or groups.
- We have two different kinds of categorical distribution plots, box plot and violin plots.
- These kinds of plots allow us to to choose a numerical variable, like age, and plot the distribution of age for each category in selected categorical variables.

Page No.
Date

Q.5 b) Explain spatial visualizations and analysis in python with folium.

→ (1) Folium is a powerful data visualisation library in python that was built primarily to help people visualize geospatial data.

(2) With folium, you can create a map of any location in the world if you know its latitude and longitude value.

(3) You can also create a map and superimpose makers as well as clusters of makers on top of the map for cool and very interesting visualizations.

(4) You can also create maps of different styles such as street level map, stamen map.

(5) You can create a basic map by calling the 'function folium.map() and specifying the initial center and zoom level.

```
m = folium.map(location = [Latitude, longitude],
    zoom_start = zoom_level).
```

Q.5a) An e-commerce company want to get into logistics "Delivery". It wants to know the pattern for maximum pickup call from different areas of the city throughout the day. This will result in:

i) Build optimum no. of stations where its pickup calls from different areas delivery personnel will be located.

ii) Ensure pickup personnel reaches the pickup location at the earliest possible time.

→ Find the highest density of probable pickup location in the future.

→ ① pre-requisites: python, Jupyter Notebook, pandas.

② Dataset: dataset contains two separate data files - train_del.csv and test_del.csv.

③ Importing libraries like pandas and numpy and drop unnecessary attributes and combine 2 different files as one dataframe.

④ maps are defined as folium.map Object. We will need to add other Object on top of this before rendering.

(5) Visualise the rides data using a class method called Heatmap().

(6) There is high demand for cabs in areas marked by the heat map which is central Delhi must probably and other surrounding areas.

(7) We can also animate our heat maps to dynamically change the data on timely basis based on a certain dimension of time.

(8) Conclusion:

Throughout the city, pickups are more probable from central areas so better to set lot of pickup stops at these location.

therefore, by using maps we can highlight trends and uncover pattern and derive insights from the data.