Assignment No - 2

Q.1 a) Explain how walson studio will provide the environment to solve business problems.

Ans :- 1) Walson Studio provides the environment and tools for you to collaboratively work on data to solve your business problems.

2) You can choose the tools you need to analyze and visualize data. to chance cleange and shape data.

3) The architecture of walson studio is created around and analytic project.

Analytics projects		
Collaborators	Assets	Tools For These tasks
Admin Editor Viewer	Data assets Opeational assets	Prepare data Visualize data Schedule Jobs Script actions Build models Manage compute

Your data Source   Catalogs   Deployment other spaces   service

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

Ans :- Data type is a collection or grouping of data values, usually specified by a set of possible values a set of allowed operations on these values.

List of data types :-
1) Numeric
2) Sequence type
3) String
4) Dictionary
5) Set

1) **Numeric :-** Numeric data type is used to hold numeric value. It include int, Float, Complex

ex num = 5
print (num₁; is af type. type (num₁))
num = 2·0
print (num 2; is af type', type (num₂))

2) **List :-** List is an ordered collection of similar or different types of specified by common.
List = ['Name', 'Roll no']

3) **String :-** Sequence of characters represented by either single or double quotes.

ex

```
name = 'python'
Print (name)
```

output:-

Python

4) **Dictionary :-** Python dictionary is an ordered collection of items

```
dict = { 'Name' : 'Mayur', 'Roll no':
         '131' }
print (dict)
```

output:- { 'Name' : 'Mayur', 'Rollno':'131'}

5) **Set :-** Set is an unordered collection of unique items.

ex

```
student-id = { 12, 14, 16 }
    print ( student-id )
    print (type (student-id)
```

output :-

{ 12, 14, 16 }

< class 'set'>

Q. 2a) What is dictionary ? Explain the methods available in dictionary.

Ans :- Dictionary is mutable data structures that allows you to store key - value pairs. The dictionary can be created using the dict() The Methods available in dictionary are:

1) key() :- It use return list of all the available keys in the dictionary
ex :- dict = { 'Name' : 'Mayur',
                'Roll no' : '131' }
      print (dict. keys ())
Output :- dict - keys ('Name', 'Roll no')

2) values () :- It use returns list of dictionary value from the key value pairs.
ex : dict = {'Name' : 'Mayur', 'Roll no'
               : '131' }
      print (dict. values ())
Output :- dict - values ('Mayur', '131')

3) copy () :- This method returns a shallow copy of dictionary
ex
  dict = { 'Name' : 'Mayur', 'Roll no' :
            '131' }

```
dict-new = dict.copy()
print(dict-new)
```

Output :- { 'Name' : 'Mayur', 'Roll no'
         : '131' }

4) update() :- The update() inserts new item to the dictionary

example :-

```
dict = { 'Name' : 'Mayur', 'Roll no'
       : '131' }
dict.update({ 'age' : 22 })
print(dict)
```

Output :-
   { 'Name' : 'Mayur', 'Roll No' : '131'
         'Age' : 22 }

Q. 2> b> Difference between numpy arrays and lists.

Ans:

numpy arrays	Lists
1> It is the core library of python which used for scientific computing	1> The core library of python provides List.
2> It can contain similar data-types	2> It contain different types of datatypes
3> It is Homogeneous	3> It is both Homogeneous and heterogeneous
4> In this element wise operation is possible	4) In this not element wise operation is possible
5> It is faster as compared list	5> It is slow as compared arrays.
6> It also have some optimism function	6> It does not have some optimism function.
7> It store each item in sequential manner	7> It store item in random location of the memory.

Q.3 a) Explain the following term
i) Raw code   ii) Waffle chart

Ans :- i) Raw code :-

The Raw code feature is exactly how it sounds: Its a full block of code that contains all of the forms HTML, CSS and Javascript that can be embedded directly into your site.

1) In data visualization "raw of code" typically refers to the direct programming code or scripts used to create visulizations from raw data.

2) This code can be written in various programming languages like python, R etc.

ii) White chart :-

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

4) However, it may require good understanding of programming and data visualization concepts to effectively create and customize.

ii) Waffle chart :-

1) A waffle chart is a type of data visualization that is used to represent the distribution of a categorical variable.

2) It is similar to a 100% stacked bar chart but is displayed in a grid format that resembles a waffle or a checker board.

3) Waffle chart are effectively for showing the composition of data in an easily under-standable and visually engaging way.

Q.3 b> How will you create the normalized weight

Ans :- By Mathematically :- Simply divide the survey weight of each unit used in the analysis by the (unweighted) average of the survey weights of all the analyzed units.

To create normalised weights you typically follow three steps :-

1. Determine the weights :- First you need a set of weights that you want to normalize.

2. Calculate the Sum :- find the sum of all the weight.

$$Sum = W_1 + W_2 + W_3 + \cdots + W_n$$

3> Normalise Each weight :-
Divide each individual weight by the sum calculated in step 2.
The normalized weight $(w-nom)$ for a weight $w-i$ would be $w-nom-i = w-i / sum$.

Q.4 a) Explain pie chart specialised visualization Tools using Matplotlib.

Ans:- A Pie chart is a circular statistical plot that can display only one series of data.

<u>Creating pie chart :-</u>
Matplotlib API has pie() function in its pyplot module which create pie chart representing the data in an array.

<u>Syntax</u> :- Matplotlib.pyplot.pie(data, explode = None, labels = None, colors = None, autopct = None, shadow = False)

<u>Parameters :-</u>

data :- It represents the array of data values to be plotted.

labels :- It is a list of sequence of strings which sets the label of each wedge.

Color :- attribute is used to provide color to the wedges.

autopct :- It is a string used to label the wedge with their

numerical value.

Shadow :- is used to create
shadow of wedge.
Example :-
from matplotlib import pyplot as plt
import numpy as np
cars = { 'AUDI', 'BMW', 'FORD',
         'TESLA', 'JAGUAR' }
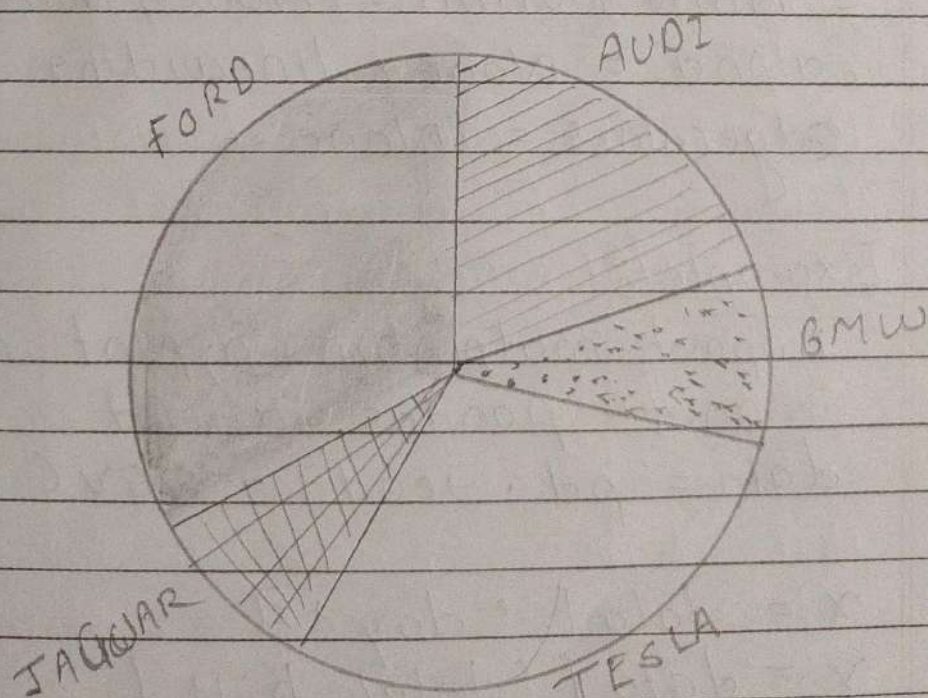
data = { 23, 17, 35, 29, 12 }
Fig = plt. figure(Figsize = (10, 7))
plt. pie(data .labels = cars)
plt. show ()

Output :-

Q.4 b) Explain Bubble plots specilised
Visualization Tools using matplotlib.
Ans:- A bubble plot is a Scatter
plot where the circle size is mapped
to the value of a third numeric
Variable.

The bubble plots are used
to observe relationship between
Variables. The bubble () method
in matplotlib library is used to
draw a bubble plot.

Syntax:-
matplotlib.pyplot.bubble(x - axis -
data, y-axis - data, S = None,
C = None, marker = None, cmap
= None, vmin = None, vmax = None
; alpha = None, linewidths - None,
edgecolors = None.

Example:-
import matplotlib.pyplot as plt
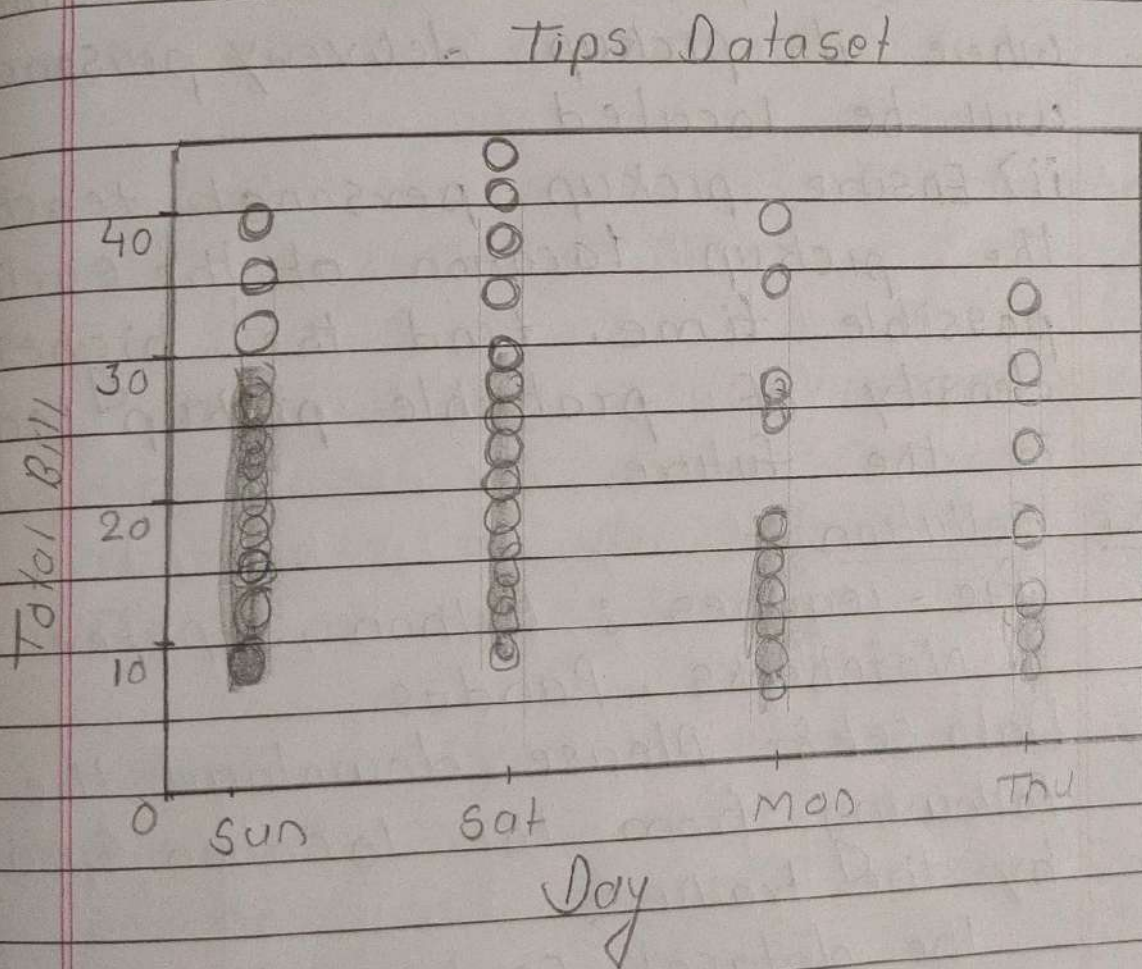import pandas as pd
data = pd.read - csv('tips.
csv')

x = data['day']
y = data['total_bill']

plt.bubble (x, y)
plt.title ("Tips Dataset")
plt.~~title~~ ylabel ('Total Bill')
plt.xlabel ('Day')
plt.show()

Output:-

Tips Dataset

Q.5 a) An e-commerce Campany wants to get into logistics "Deliver4U". It wants to know the pattern for maximum pickup calls from different areas of the city throughout the day. This will result in:

i) Build optimum number of stations where its pickup delivery personal will be located

ii) Ensure pickup personel reaches the pickup location at the earliest possible time. Find the highest density of probable pickup locations in the future.

Ans: Sollution :-

Pre-requires : Python, Jupyter Notebooks, Pandas.

Data set :- Please download the following from the location specified by the trainer.

The dataset contains two seperate data files - train-del.csv and test-del.csv. The difference is that train-del.csv contains additional column which is trip-duration which we will not be needed for our present analysis.

a) Import libraries - Pandas and folium. Drop the trip - duration column and combine the 2 different files as one dataframe.

```
import pandas as pd
import folium
```

```
df_train = pd.read_csv('train-
    del.csv').drop(column = ['
trip_duration', 'dropoff-datetime'])
df_train.head().
```

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

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

Q5 b) Explain spatial visualization and Analysis in python with folium.

Ans:- Visualizing spatial data with folium maps are defined as a folium maps object and we can add other folium objects on top of the folium.

Syntax :- folium.map(locations, files = " openstreetmap " zoom - start = 4)

Code :-
```
import folium
import Pandas as P
m = folium.Map( location = [40, -95]
                     , zoom-start = 4 )
m.save ( my map.html)
```

Output :-
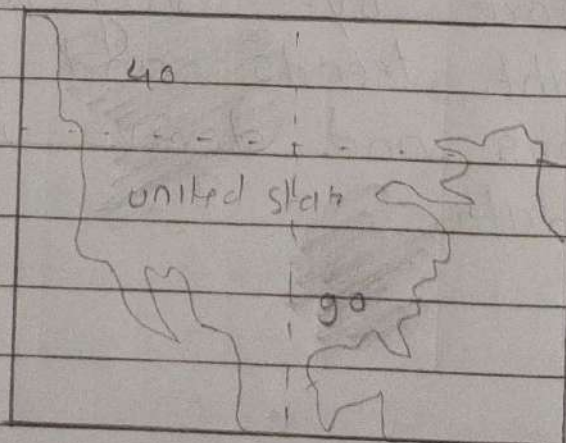
Q.6 a) Explain the following terms
i) regression plots ii) matrix plots.

Ans :- i) Regression plots :-

Regression plots as the son name suggests creates a regression line between 2 parameters and helps to visualize their linear relationships. Below are example of scalter plot and the same plot with a linear regression line added. The regression line is an attempt to find the best fit through the points in the scatter plot
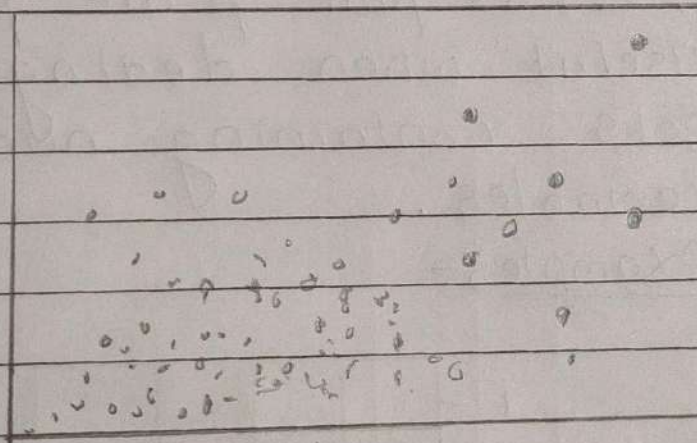

a)
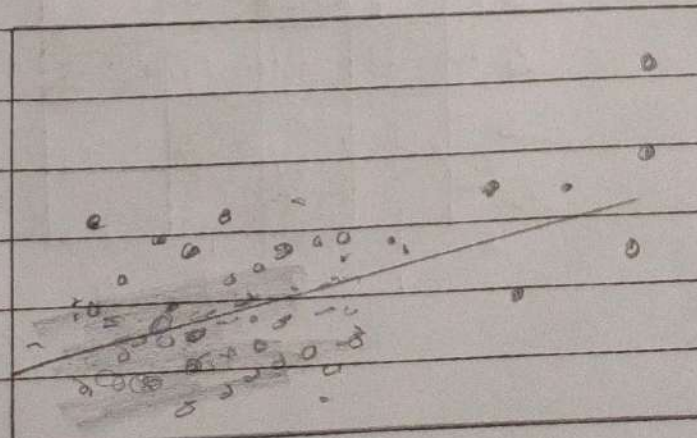
b)

ii> Matrix plot :

Ans:- 1) Matrix plots, also known as headmap matrices, arec a type of data visualization technique used to display the relationship between multiple vevriables in a dataset.

2) Theie are particularily useful for exploring correlations or patterns in multi-variate data.

3) Matrix plot p are particularly useful when dealing with data-sets containing numereous variables.

example :-
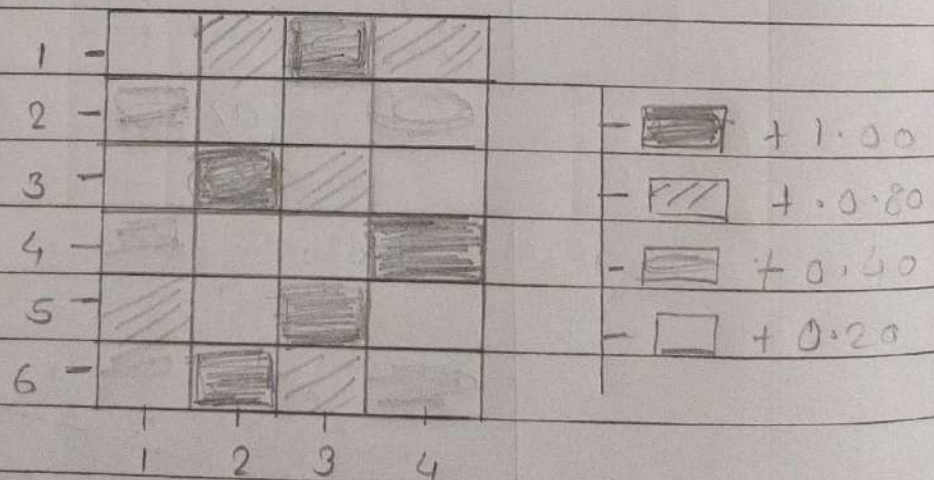
Q.6 b> Explain the following terms
i> Distribution plots ii> Categorical plots.

Ans:- 1) Distribution plots:-
1> Distribution plots visually asses that distribution of sample data by comparing the empirical distribution of the data with the theoratical values expected from a specified distribution.

2> The use of distribution plots is essential for exploratory data analysis

example
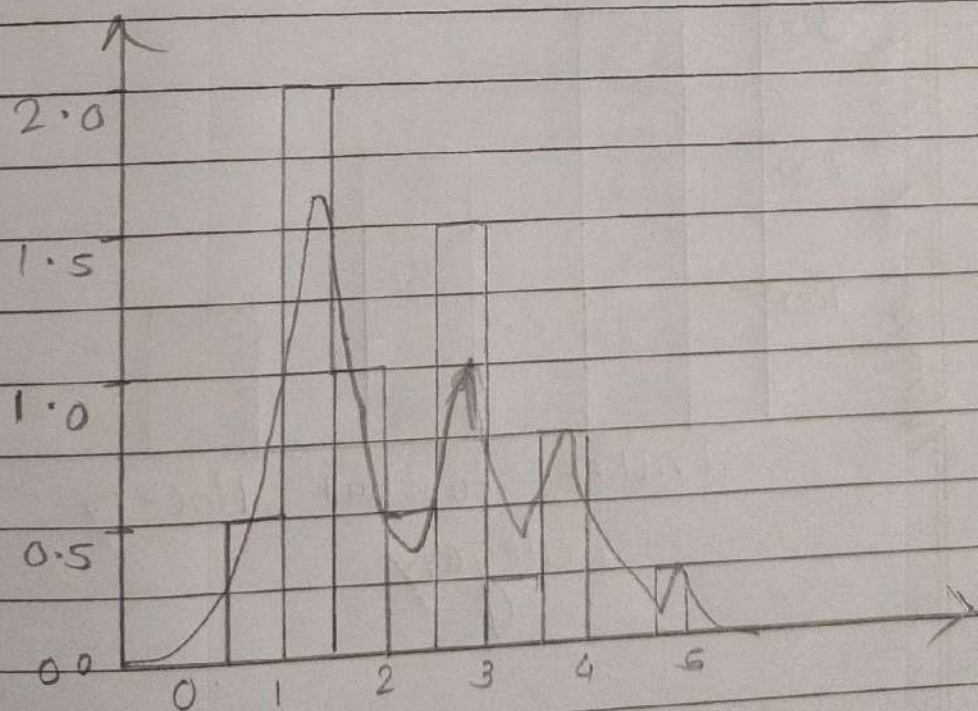
11) Categorical plot :-

There are two types of categorical plots i> box plot and ii> violin plots. These kinds of plot allow us to choose a numerical variable, like age and plot the distribution of age for each category in a selected categorical variable

ex :-

Chart of Number of Students