EXPERIMENT-8

Student Name: Garv Khurana
UID: 21BCS6615
Branch: BE CSE
Section/Group: 21AML-9/A
Semester: 3
Date of Performance: 01 -11-22
Subject: Python for Machine Learning
Subject Code: 21CSH-238

1. Aim/Overview of the practical:

- Create a dataset of customers orders including Customer_id, customer_name, Customer_age, Product_name, Cost_of_product, quantity.

  Create a Scatter plot by taking price and quantity columns.
  Also, explain the relationship among them in two-three lines.

- Create a dataset of student registration and implement some data formatting and cleaning techniques on your own. Ex.
  Duplicacy removal, removing typo errors, etc.

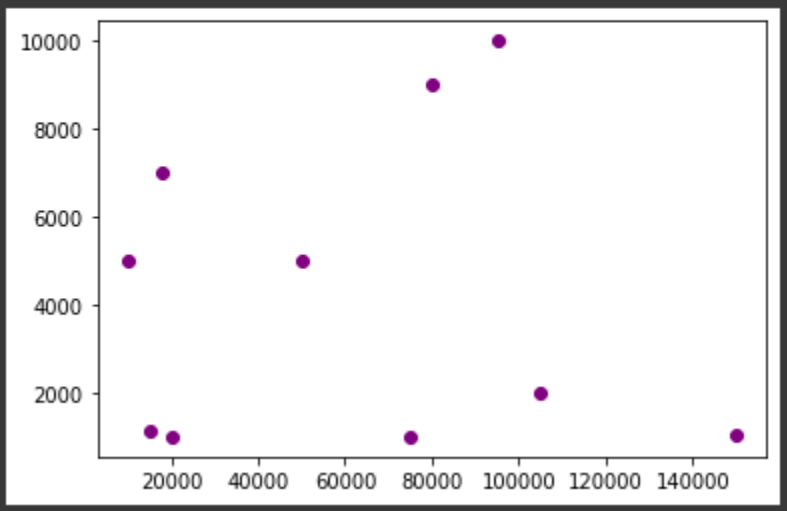
2. Code and Output:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

cust_id=[101,102,103,104,105,106,107,108,109,110]
cust_nm=['Garv Khurana','Naruto Uzumaki','Itadori Yuji','Eren Yeager','Izuku Midoria',
         'Light Yagami','Tanjiro Kamado','Ichigo Kurosaki','Denji','Saitama']
cust_age=[19,30,17,19,16,24,16,19,20,28]
pdt_name=['Basket of chocolates','Makeup products','Watches','Iphone','Kurta','Ice cream','Books'
          ,'Hoodies','Glasses','Cold coffee']
cst=[50000,20000,150000,105000,80000,75000,15000,18000,10000,95000]
qty=[5000,1000,1050,2000,9000,1000,1150,7008,5006,10000]
ci=pd.Series(cust_id)
cn=pd.Series(cust_nm)
ca=pd.Series(cust_age)
pn=pd.Series(pdt_name)
cost=pd.Series(cst)
quant=pd.Series(qty)
data={"Customer ID":ci,"Customer Name":cn,"Customer Age":ca,"Product Name":pn,"Cost":cost,"Quantity":quant}
res=pd.DataFrame(data)
print(res)
```

	Customer ID	Customer Name	Customer Age	Product Name	Cost	Quantity
0	101	Garv Khurana	19	Basket of chocolates	50000	5000
1	102	Naruto Uzumaki	30	Makeup products	20000	1000
2	103	Itadori Yuji	17	Watches	150000	1050
3	104	Eren Yeager	19	Iphone	105000	2000
4	105	Izuku Midoria	16	Kurta	80000	9000
5	106	Light Yagami	24	Ice cream	75000	1000
6	107	Tanjiro Kamado	16	Books	15000	1150
7	108	Ichigo Kurosaki	19	Hoodies	18000	7008
8	109	Denji	20	Glasses	10000	5006
9	110	Saitama	28	Cold coffee	95000	10000

```
plt.scatter(cst,qty,c="purple")
plt.show()
```



```
#create dataset of student registration and implement some data formatting
#and cleaning techniques on your own.
#Ex. duplicacy removal,removing typo errors etc
import pandas as pd
import numpy as np
reg=[1210789,1256891,2359874,5689145,7853219,5689145,9587632,1258746,5697100,3267842]
nm=['Dimple Ahuja','Rishi Singh Shekhawat','Namrata Winsariya',
    'Celina','Krish','Celina','Anmol','Dev','Garv','knowhere']
age=[19,20,21,19,21,19,20,19,19]
crs=['App Developement','Animation','Web Developement',
     'App Developement','App Developement','App Developement',
    'App Developement','App Developement','App Developement']
r=pd.Series(reg)
n=pd.Series(nm)
a=pd.Series(age)
c=pd.Series(crs)
data={"Registration no":r,"Student Name":n,"Student Age":a,"Course Name":c}
result=pd.DataFrame(data)
print(result)
```

```
   Registration no           Student Name  Student Age       Course Name
0          1210789           Dimple Ahuja         19.0  App Developement
1          1256891  Rishi Singh Shekhawat         20.0         Animation
2          2359874     Namrata Winsariya         21.0  Web Developement
3          5689145                Celina         19.0  App Developement
4          7853219                 Krish         21.0  App Developement
5          5689145                Celina         19.0  App Developement
6          9587632                 Anmol         20.0  App Developement
7          1258746                   Dev         19.0  App Developement
8          5697100                  Garv         19.0  App Developement
9          3267842              knowhere          NaN               NaN
```

```
result.values
```

```
array([[1210789, 'Dimple Ahuja', 19.0, 'App Developement'],
       [1256891, 'Rishi Singh Shekhawat', 20.0, 'Animation'],
       [2359874, 'Namrata Winsariya', 21.0, 'Web Developement'],
       [5689145, 'Celina', 19.0, 'App Developement'],
       [7853219, 'Krish', 21.0, 'App Developement'],
       [5689145, 'Celina', 19.0, 'App Developement'],
       [9587632, 'Anmol', 20.0, 'App Developement'],
       [1258746, 'Dev', 19.0, 'App Developement'],
       [5697100, 'Garv', 19.0, 'App Developement'],
       [3267842, 'knowhere', nan, nan]], dtype=object)
```

```
[5]  result.size
```

```
40
```

```
[6]  result.head()
```

	Registration no	Student Name	Student Age	Course Name
0	1210789	Dimple Ahuja	19.0	App Developement
1	1256891	Rishi Singh Shekhawat	20.0	Animation
2	2359874	Namrata Winsariya	21.0	Web Developement
3	5689145	Celina	19.0	App Developement
4	7853219	Krish	21.0	App Developement

```
result.loc[1:3]
```

	Registration no	Student Name	Student Age	Course Name
1	1256891	Rishi Singh Shekhawat	20.0	Animation
2	2359874	Namrata Winsariya	21.0	Web Developement
3	5689145	Celina	19.0	App Developement

```
[8]  result.tail()
```

	Registration no	Student Name	Student Age	Course Name
5	5689145	Celina	19.0	App Developement
6	9587632	Anmol	20.0	App Developement
7	1258746	Dev	19.0	App Developement
8	5697100	Garv	19.0	App Developement
9	3267842	knowhere	NaN	NaN

```
[9]  result.describe()
```

	Registration no	Student Age
count	1.000000e+01	9.000000
mean	4.387038e+06	19.666667
std	2.966524e+06	0.866025
min	1.210789e+06	19.000000
25%	1.534028e+06	19.000000
50%	4.478494e+06	19.000000
75%	5.695111e+06	20.000000
max	9.587632e+06	21.000000

```
#Removing duplicate values
result.drop_duplicates(inplace=True)
display(result)
```

	Registration no	Student Name	Student Age	Course Name
0	1210789	Dimple Ahuja	19.0	App Developement
1	1256891	Rishi Singh Shekhawat	20.0	Animation
2	2359874	Namrata Winsariya	21.0	Web Developement
3	5689145	Celina	19.0	App Developement
4	7853219	Krish	21.0	App Developement
6	9587632	Anmol	20.0	App Developement
7	1258746	Dev	19.0	App Developement
8	5697100	Garv	19.0	App Developement
9	3267842	knowhere	NaN	NaN

```
#Changing name of columns
new_data = result.rename(columns = {"Student Name": "Name of Students"})
display(new_data)
```

	Registration no	Name of Students	Student Age	Course Name
0	1210789	Dimple Ahuja	19.0	App Developement
1	1256891	Rishi Singh Shekhawat	20.0	Animation
2	2359874	Namrata Winsariya	21.0	Web Developement
3	5689145	Celina	19.0	App Developement
4	7853219	Krish	21.0	App Developement
6	9587632	Anmol	20.0	App Developement
7	1258746	Dev	19.0	App Developement
8	5697100	Garv	19.0	App Developement
9	3267842	knowhere	NaN	NaN

```
[12] #Removing row containing null value
     result.dropna(inplace=True)
     display(result)
```

	Registration no	Student Name	Student Age	Course Name
0	1210789	Dimple Ahuja	19.0	App Developement
1	1256891	Rishi Singh Shekhawat	20.0	Animation
2	2359874	Namrata Winsariya	21.0	Web Developement
3	5689145	Celina	19.0	App Developement
4	7853219	Krish	21.0	App Developement
6	9587632	Anmol	20.0	App Developement
7	1258746	Dev	19.0	App Developement
8	5697100	Garv	19.0	App Developement

```
[13] #Removing typo error
     result.loc[4, 'Student Name'] = "Sande"
     display(result)
```

	Registration no	Student Name	Student Age	Course Name
0	1210789	Dimple Ahuja	19.0	App Developement
1	1256891	Rishi Singh Shekhawat	20.0	Animation
2	2359874	Namrata Winsariya	21.0	Web Developement
3	5689145	Celina	19.0	App Developement
4	7853219	Sande	21.0	App Developement
6	9587632	Anmol	20.0	App Developement
7	1258746	Dev	19.0	App Developement
8	5697100	Garv	19.0	App Developement

Learning outcomes (What I have learned):

- I have learned about the python programming language.
- I have learned about cleaning techniques using pandas.
- I have learned to plot scatter charts.
- I have learned about different libraries and packages.

Evaluation Grid (To be created as per the SOP and Assessment guidelines by the faculty):

Sr. No.	Parameters	Marks Obtained	Maximum Marks
1.	PERFORMANCE		12
2.	WORKSHEET		08
3.	VIVA		10