EXPERMINT: 05

● Aim: Implementation of Data Discretization (any one) & Visualization (any one).

● Theory:

Data discretization:

Data discretization refers to a method of converting a huge number of data values into smaller ones so that the evaluation and management of data become easy. In other words, data discretization is a method of converting attributes values of continuous data into a finite set of intervals with minimum data loss. There are two forms of data discretization first is supervised discretization, and the second is unsupervised discretization.

Some Famous techniques of data discretization:

- Histogram analysis
- Binning
- Cluster Analysis
- Data discretization using decision tree analysis

Why is Discretization important?

As we know, an infinite of degrees of freedom mathematical problem poses with the continuous data. For many purposes, data scientists need the implementation of discretization. It is also used to improve signal noise ratio.

Data Visualization: -

Data visualization is a graphical representation of quantitative information and data by using visual elements like graphs, charts, and maps. Data visualization convert large and small data sets into visuals, which is easy to understand and process for humans.Data visualization tools provide accessible ways to understand outliers, patterns, and trends in the data.

In the world of Big Data, the data visualization tools and technologies are required to analyze vast amounts of information. Data visualizations are common in your everyday life, but they always appear in the form of graphs and charts. The combination of multiple visualizations and bits of information are still referred to as Info graphics.

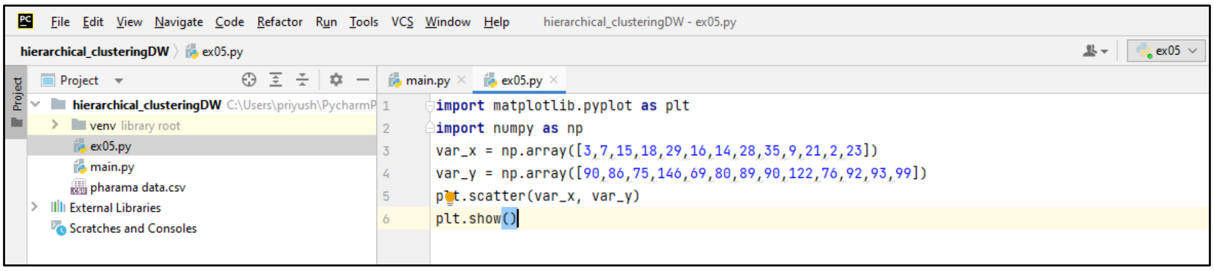
Scatter Plot:- A scatter plot (also called a scatterplot, scatter graph, scatter chart, scattergram, or scatter diagram) is a type of plot or mathematical diagram using Cartesian coordinates to display values for typically two variables for a set of data. If the points are coded (color/shape/size), one additional variable can be displayed.

Bar chart: - A bar chart or bar graph is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar chart is sometimes called a column chart

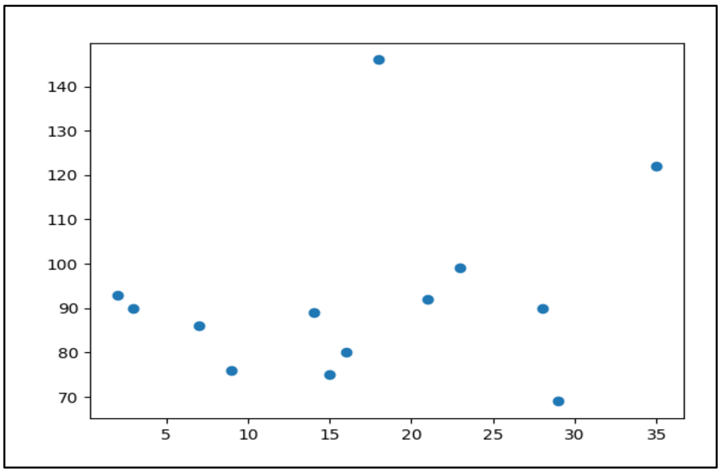
Pie chart: - A pie chart (or a circle chart) is a circular statistical graphic, which is divided into slices to illustrate numerical proportion. In a pie chart, the arc length of each slice (and consequently its central angle and area), is proportional to the quantity it represents.

Histogram: - A histogram is a graphical representation that organizes a group of data points into user-specified ranges. Similar in appearance to a bar graph, the histogram condenses a data series into an easily interpreted visual by taking many data points and grouping them into logical ranges or bins.

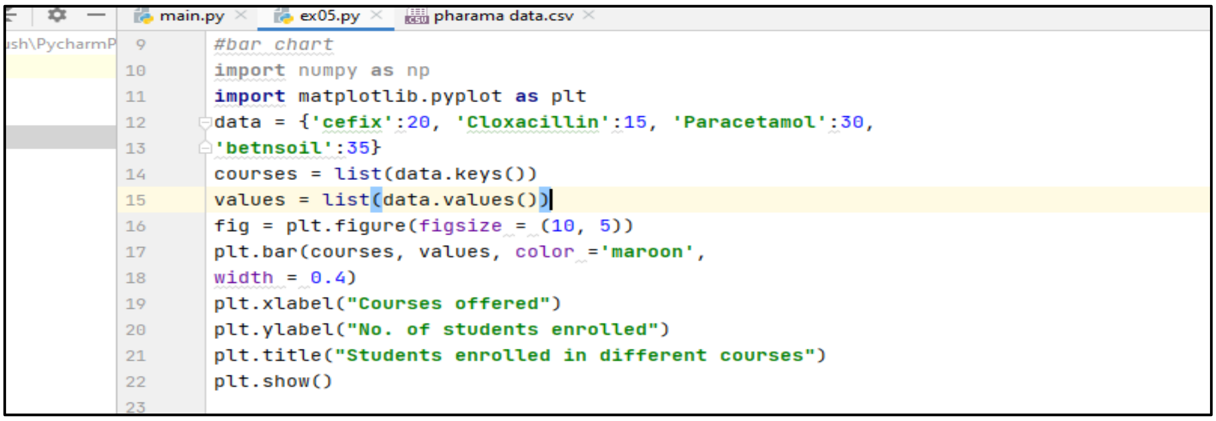
Scatter Plot:


File Edit View Navigate Code Refactor Run Tools VCS Window Help
hierarchical_clusteringDW - ex05.py
hierarchical_clusteringDW 〉 ex05.py
ex05
Project
main.py
ex05.py
hierarchical_clusteringDW C:\Users\priyush\PycharmP
venv library root
ex05.py
main.py
pharama data.csv
External Libraries
Scratches and Consoles

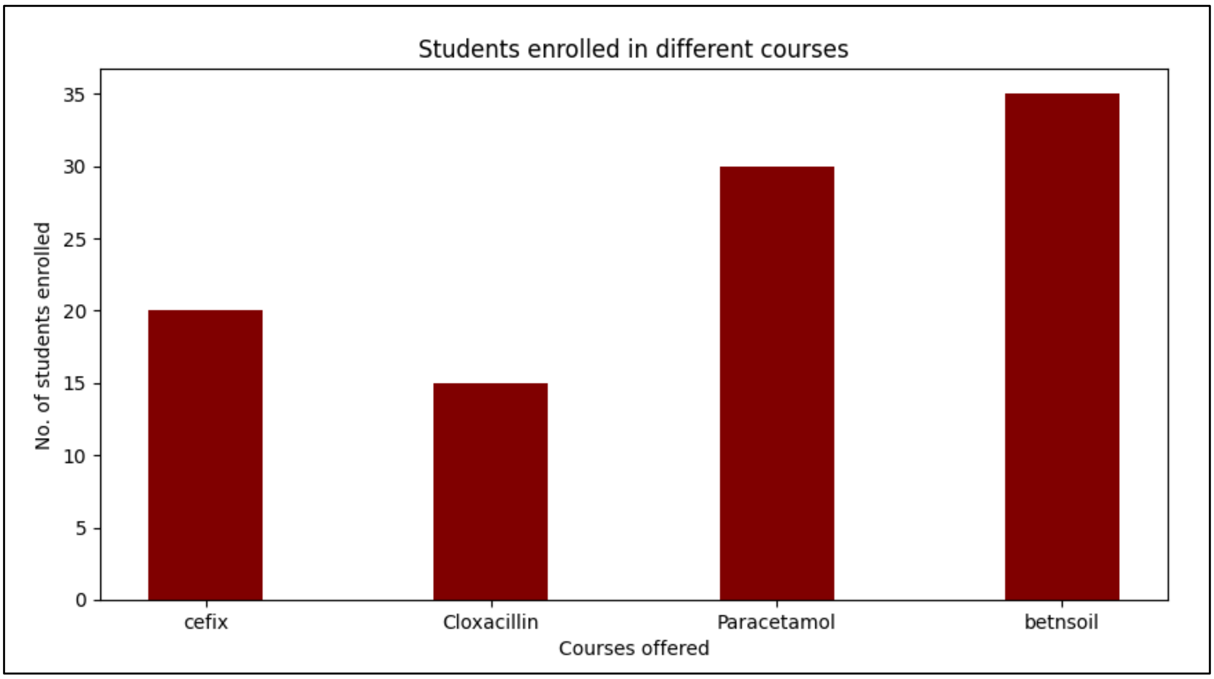
```python
import matplotlib.pyplot as plt
import numpy as np
var_x = np.array([3,7,15,18,29,16,14,28,35,9,21,2,23])
var_y = np.array([90,86,75,146,69,80,89,90,122,76,92,93,99])
plt.scatter(var_x, var_y)
plt.show()
```



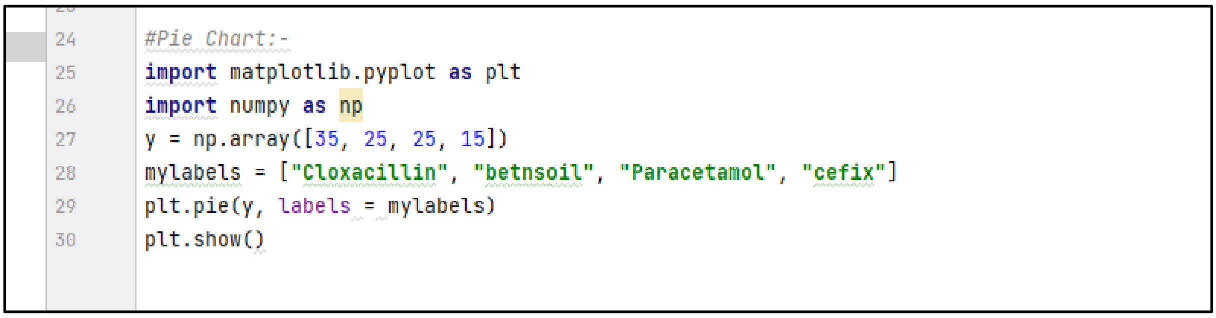
Bar chart



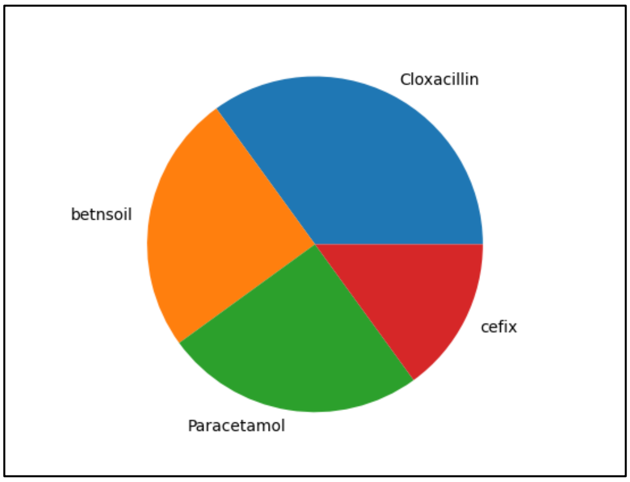
```python
#bar chart
import numpy as np
import matplotlib.pyplot as plt
data = {'cefix':20, 'Cloxacillin':15, 'Paracetamol':30,
'betnsoil':35}
courses = list(data.keys())
values = list(data.values())
fig = plt.figure(figsize = (10, 5))
plt.bar(courses, values, color ='maroon',
width = 0.4)
plt.xlabel("Courses offered")
plt.ylabel("No. of students enrolled")
plt.title("Students enrolled in different courses")
plt.show()
```



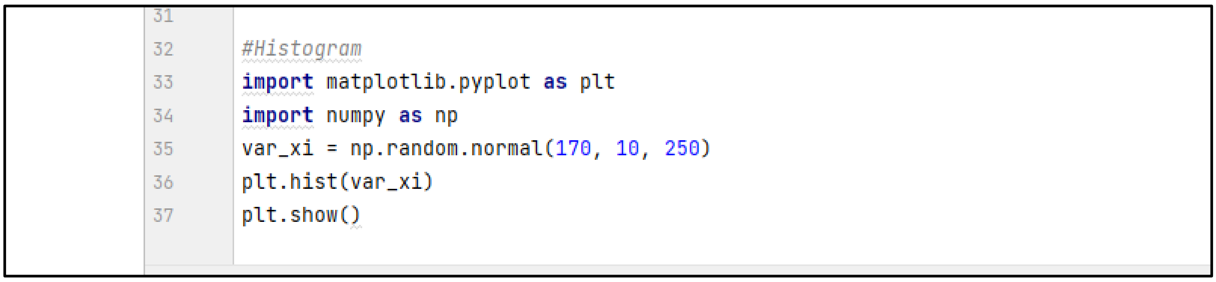
Pie chart



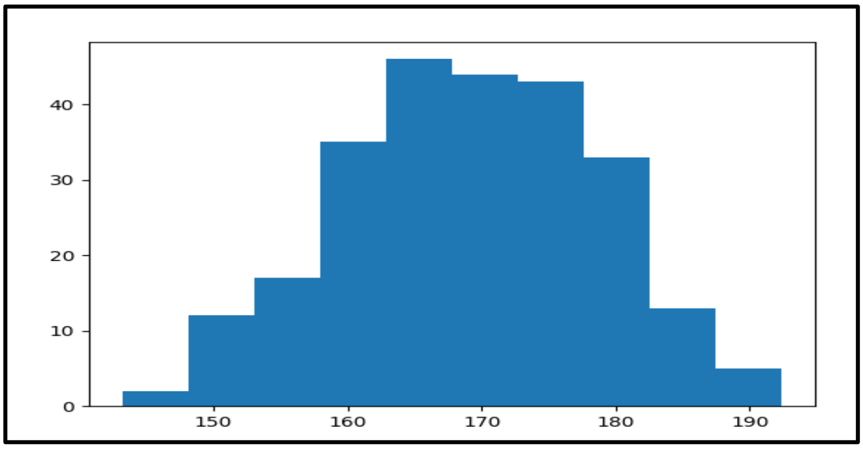
```
#Pie Chart:-
import matplotlib.pyplot as plt
import numpy as np
y = np.array([35, 25, 25, 15])
mylabels = ["Cloxacillin", "betnsoil", "Paracetamol", "cefix"]
plt.pie(y, labels = mylabels)
plt.show()
```



Histogram:



```
#Histogram
import matplotlib.pyplot as plt
import numpy as np
var_xi = np.random.normal(170, 10, 250)
plt.hist(var_xi)
plt.show()
```



● <u>**Conclusion:**</u>

We implemented binning as a method of **data discretization** and various **visualization** method such as scatter plot, bar chart, pie chart and histogram as a method of data visualization.