AMRUTVAHINI COLLEGE OF ENGINEERING, SANGAMNER

BE(E&C)	Data Science & Visualization Lab	
Experiment No.: 04	**Importing dataset (CSV file) and Python program to demonstrate the various plots using Matplotlib library on dataset.**	**Page:** /9

Aim of Experiment:

Importing dataset (CSV file) and Python program to demonstrate the various plots using Matplotlib library on dataset.

Software Used: Python 3.12, Jupyter Notebook.

Learning Objective

To learn how to plot various graphs/charts using Matplotlib library on the given dataset.

Learning Outcomes:

After performing the experiment students will be able to-

1. Understand Matplotlib library for data visualization
2. Write Python scripts to perform data visualization

Theory:

Matplotlib-

Matplotlib is Python package used for data visualization. It provides an object-oriented API that helps in embedding plots in applications using Python. Matplotlib library is used for creating static, animated, and interactive 2D- plots or figures in Python. It can be installed using the following pip command from the command prompt:

pip install matplotlib

For plotting using Matplotlib, we need to import its Pyplot module using the following command:

import matplotlib.pyplot as plt

Here, plt is an alias or an alternative name for matplotlib.pyplot. We can use any other alias also.

1. Pandas Read CSV in Python

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
df = pd.read_csv("E:\Data Science and Visualization\Lab Manual ISO\DSV
    Practicals\diabetes.csv")
df
```

2. Drop the Columns and Rows

```
df= df.drop(columns=['Pregnancies', 'SkinThickness',
'DiabetesPedigreeFunction', 'Insulin', 'Outcome'])
df
```

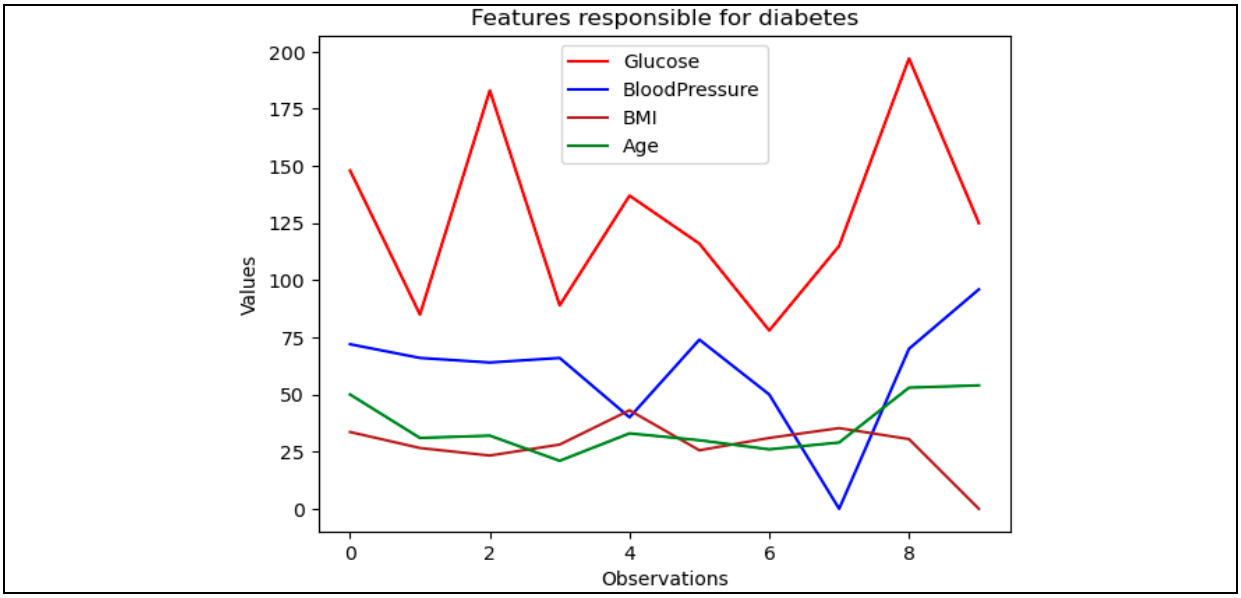
```
df = df.drop(df.index[10:])
df
```

Line chart

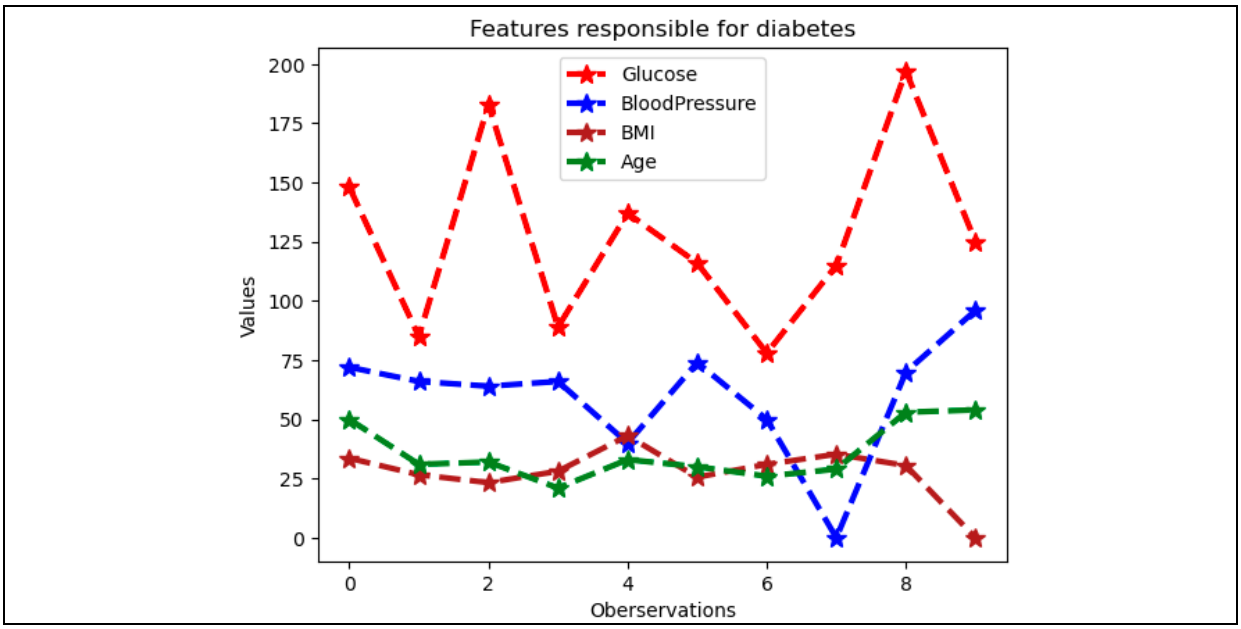
A line plot is a graph that shows the frequency of data along a number line. It is used to show continuous dataset. A line chart or line graph is a type of chart which displays information as a series of data points called ‘markers’ connected by straight line segments. A line plot is used to visualise growth or decline in data over a time interval. A line chart can be created using the Matplotlib plot() function.

```
#create a line plot of different color for each feature
df.plot(kind='line', color=['red','blue','brown','green'])
#Set title to "Features responsible for diabetes"
plt.title('Features responsible for diabetes')
#Label x axis as "Observations"
plt.xlabel('Observations')
#Label y axis as "Values"
plt.ylabel('Values')
#Display the figure
plt.show()
```

When the code is executed, the following graph is displayed



```
#create a line plot of different color for each feature
df.plot(kind='line', color=['red','blue','brown','green'],
marker='*',markersize=10,linewidth=3,linestyle="--")
# Set title to "Features responsible for diabetes"
plt.title('Features responsible for diabetes')
# Label x axis as "Oberservations"
plt.xlabel('Oberservations')
# Label y axis as "Values"
plt.ylabel('Values')
#Display the figure
plt.show()
```

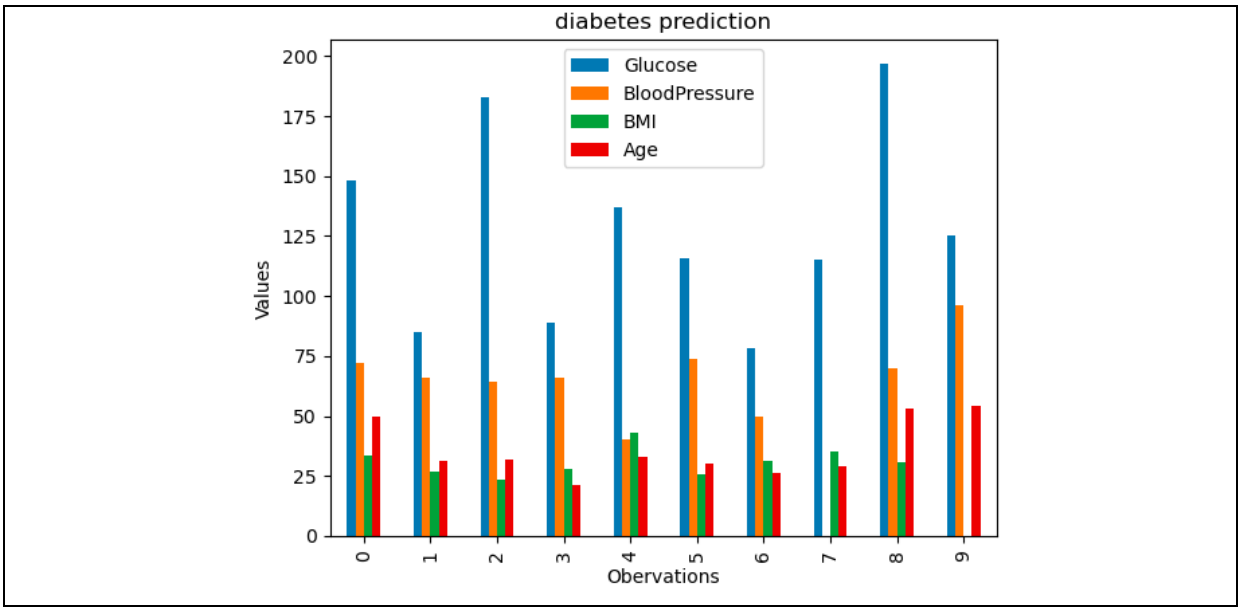


Bar Chart

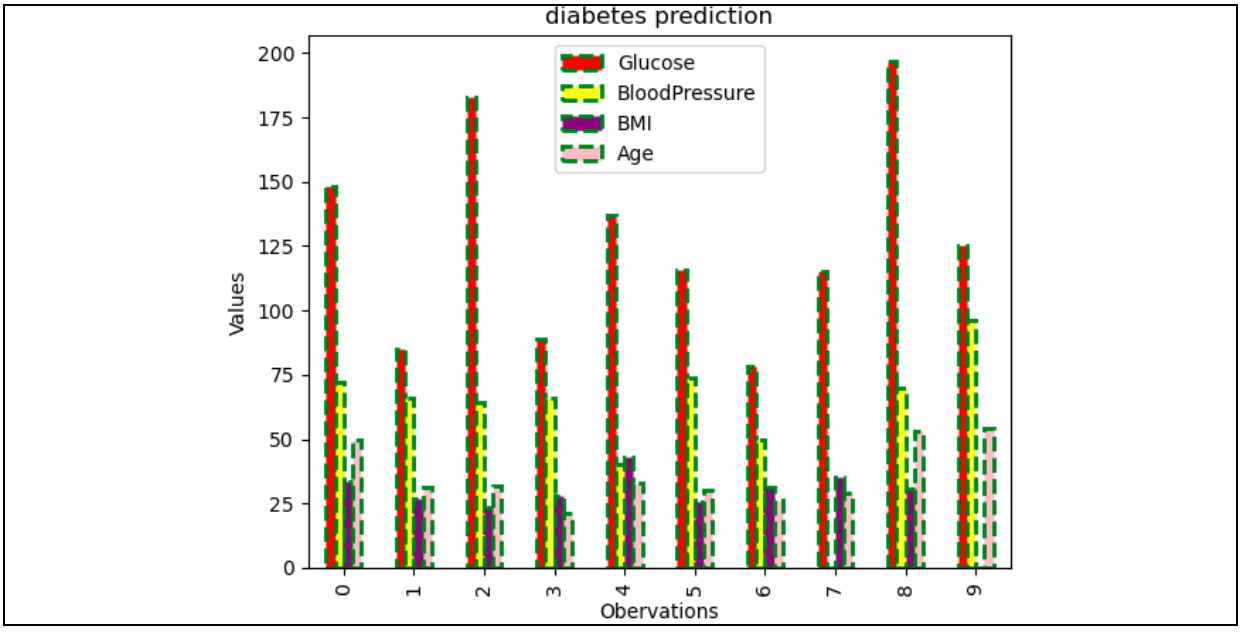
A bar chart or bar graph is a chart or graph that presents categorical data with rectangular bars with heights proportional to the values that they represent. Matplotlib API provides the bar() function that can be used to plot bar chart.

```
#plots a bar chart
df.plot(kind='bar', title='diabetes prediction')
#set title, set xlabel and ylabel
plt.xlabel('Obervations')
plt.ylabel('Values')
plt.show()
```

When the code is executed, the following graph is displayed




```
#Customising Bar Chart
df.plot(kind ='bar',title ='diabetes prediction',
color=['red','yellow','purple','pink'],edgecolor='Green', linewidth=2,
linestyle='--')
#set title, set xlabel and ylabel
plt.xlabel('Obervations')
plt.ylabel('Values')
plt.show()
```

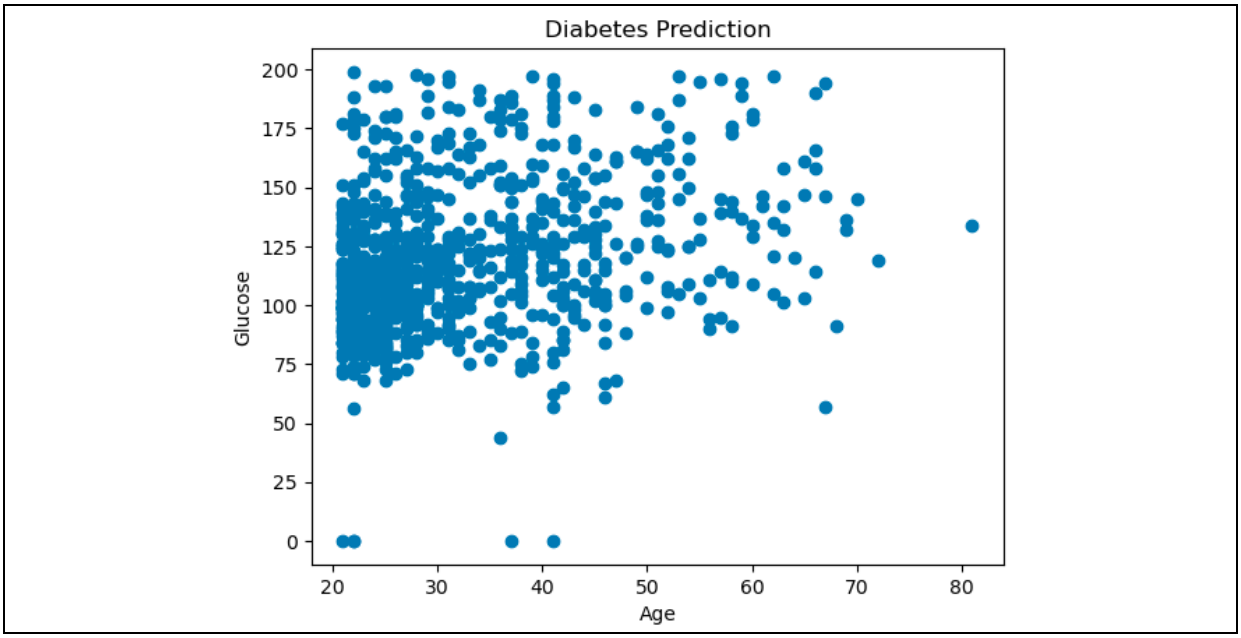


Scatter plot

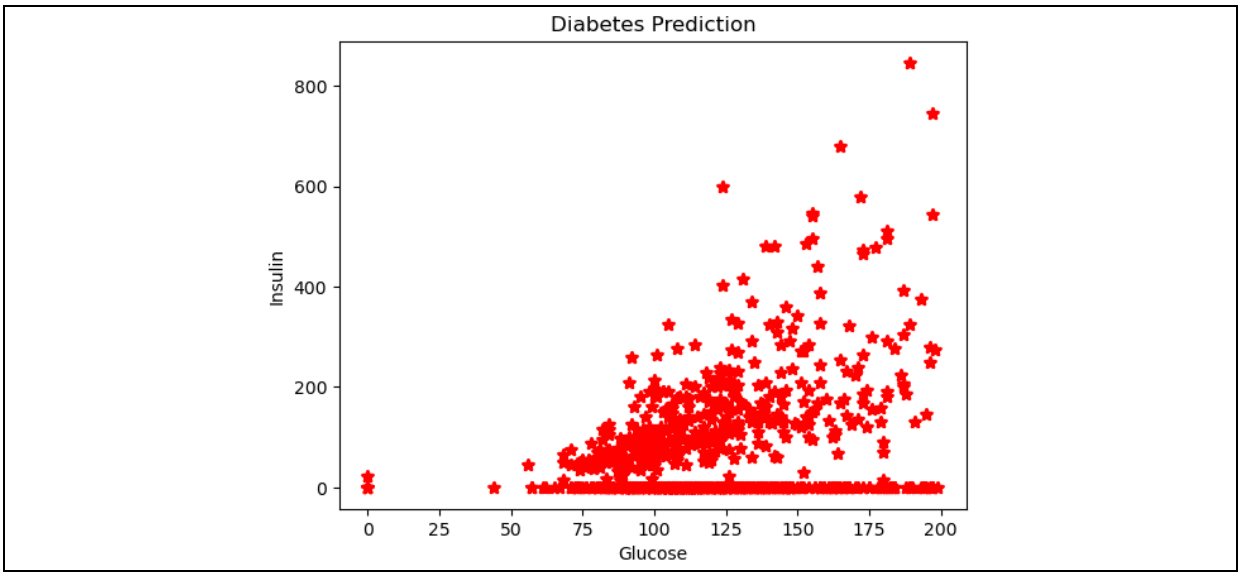
A scatter chart is a two-dimensional data visualisation method that uses dots to represent the values obtained for two different variables —one plotted along the x-axis and the other plotted along the y-axis. The scatter() method in the matplotlib library is used to draw a scatter plot.

```
x= df['Age']
y= df['Glucose']
plt.scatter(x,y)
plt.title('Diabetes Prediction')
plt.xlabel('Age')
plt.ylabel('Glucose')
plt.show()
```

When the code is executed, the following graph is displayed




```
#Customising Scatter Plot
x= df['Glucose']
y= df['Insulin']
plt.scatter(x,y,color='red',linewidth=2,marker='*',edgecolor='red')
plt.title('Diabetes Prediction')
plt.xlabel('Glucose')
plt.ylabel('Insulin')
plt.show()
```

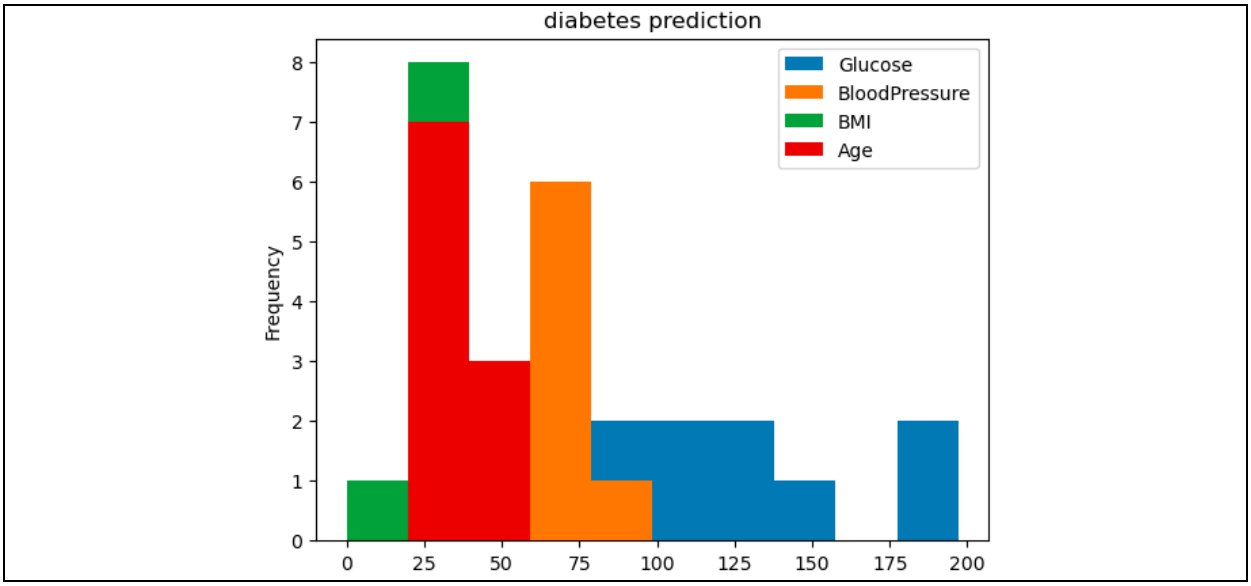


Histogram

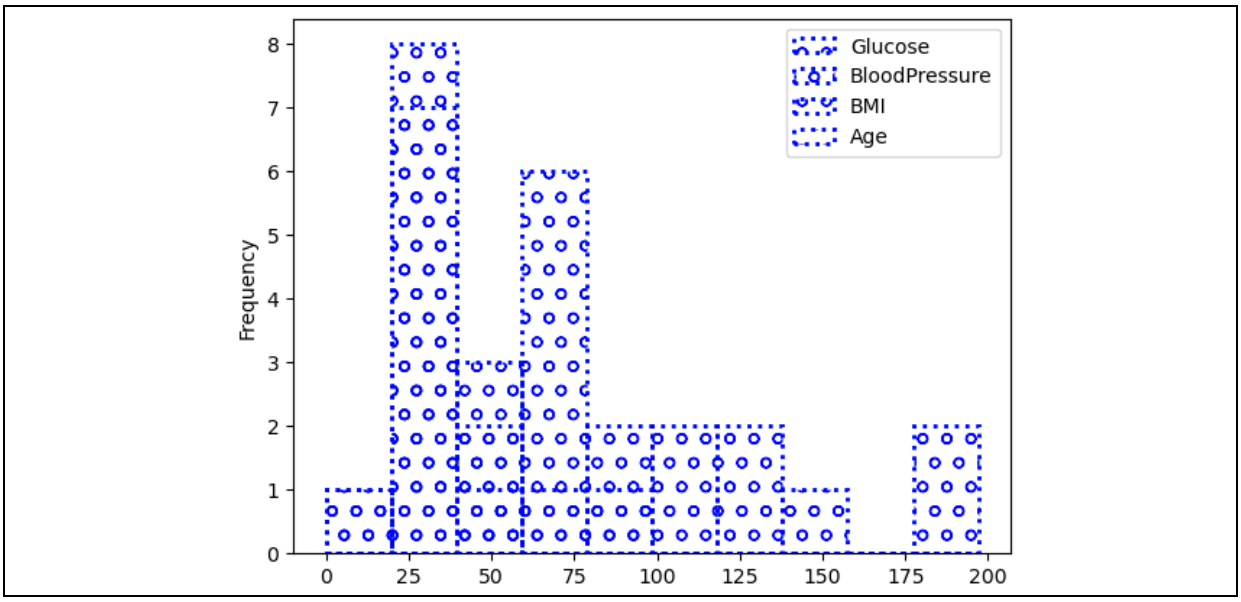
Histograms are column-charts, where each column represents a range of values, and the height of a column corresponds to how many values are in that range. To make a histogram, the data is sorted into "bins" and the number of data points in each bin is counted. The height of each column in the histogram is then proportional to the number of data points its bin contains. A histogram is a graph showing frequency distributions.

```
df.plot(kind='hist',title='diabetes prediction')
plt.show()
```

When the code is executed, the following graph is displayed



```
#Customising Histogram
df.plot(kind='hist',edgecolor='Blue',linewidth=2,linestyle=':',fill=False,h
atch='o')
plt.show()
```

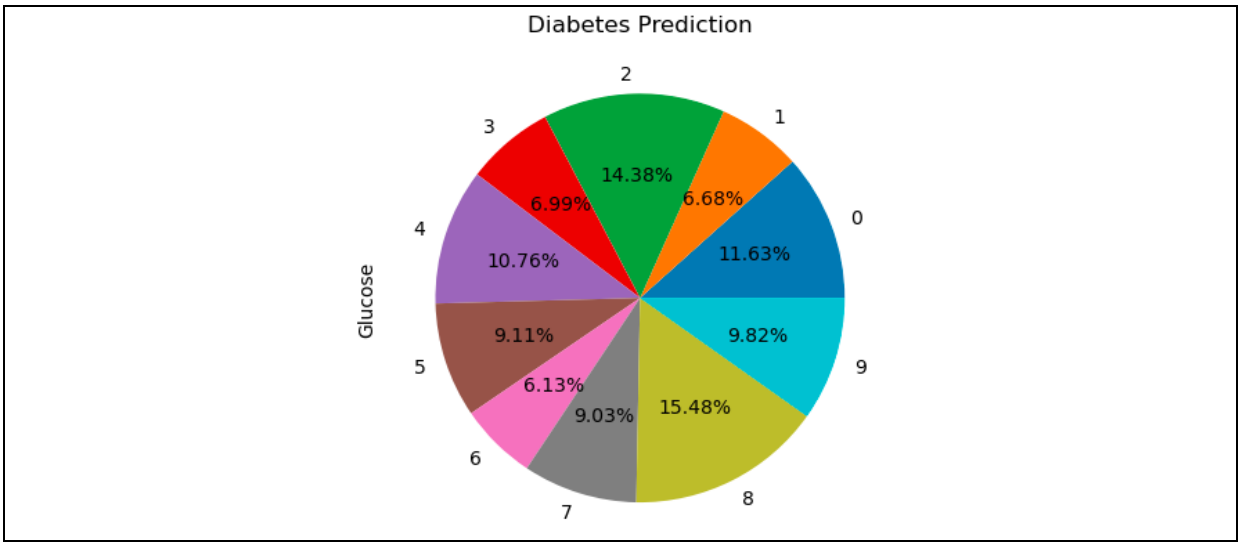


Pie Chart

Pie is a type of graph in which a circle is divided into different sectors and each sector represents a part of the whole. A pie plot is used to represent numerical data proportionally. To plot a pie chart, either column label y or 'subplots=True' should be set while using df.plot(kind='pie') . If no column reference is passed and subplots=True, a 'pie' plot is drawn for each numerical column independently.

```
df.plot(kind='pie',y='Glucose', title='Diabetes Prediction', legend=False,
autopct="%.2f%%")
plt.show()
```

BE(E&C)	Data Science & Visualization Lab	
Experiment No.: 04	Importing dataset (CSV file) and Python program to demonstrate the various plots using Matplotlib library on dataset.	Page: /9



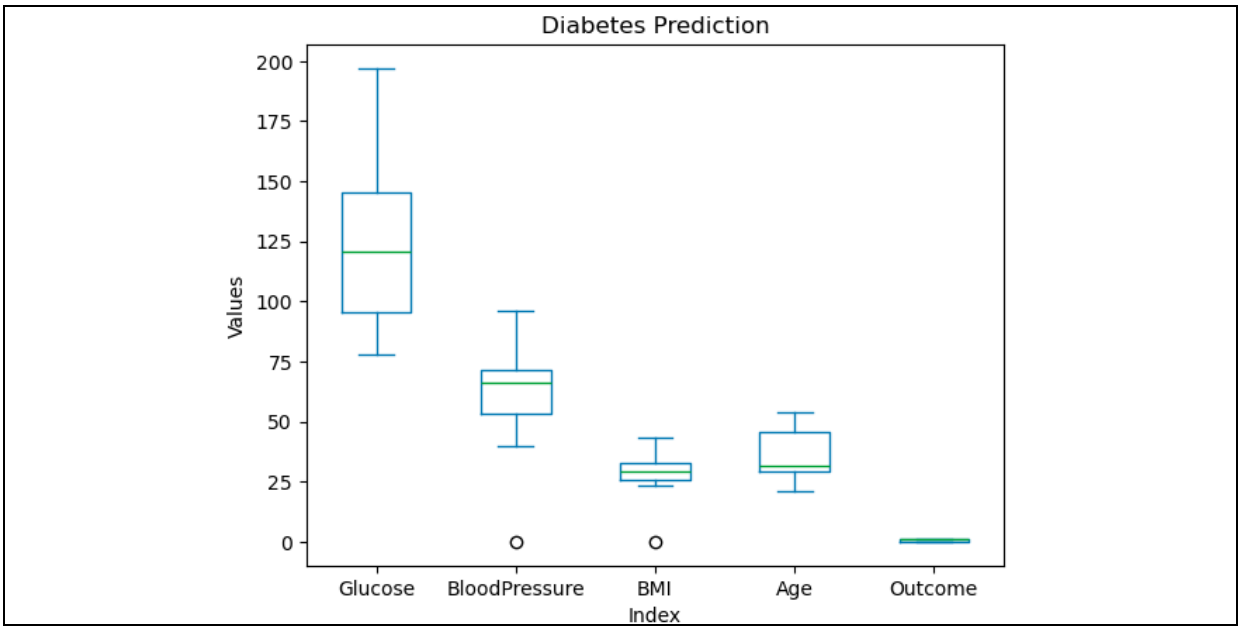
Quartiles and Box plot

A Box Plot is the visual representation of the statistical summary of a given data set. The summary includes Minimum value, Quartile 1, Quartile 2, Median, Quartile 4 and Maximum value. The whiskers are the two lines outside the box that extend to the highest and lowest values. It also helps in identifying the outliers.

```
df.plot(kind='box')
#set title,xlabel,ylabel
plt.title('Diabetes Prediction')
plt.xlabel('Index')
plt.ylabel('Values')
plt.show()
```

AMRUTVAHINI COLLEGE OF ENGINEERING, SANGAMNER

BE(E&C)	Data Science & Visualization Lab	
Experiment No.: 04	Importing dataset (CSV file) and Python program to demonstrate the various plots using Matplotlib library on dataset.	Page: /9



Questions:

1. What is the purpose of the Matplotlib library?
2. What are some of the major components of any graphs or plot?
3. Write short notes on different customisation options available with any plot.
4. Define Pandas visualisation.
5. Describe the function matplotlib.pyplot.scatter in detail.

Conclusion:
