Chapter - 4
Plotting Data using Matplotlib

Que 1. What is the purpose of the Matplotlib library?

Ans. The matplotlib is a data visualisation library of Python — a high quality plotting library of python. It provides both a very quick way to visualize data from Python and publication-quality figures in many formats. The matplotlib library offers many different named collections of methods; pyPlot is one such interface.

One important feature of matplotlib library is that it is crossplatform, i.e., it can work with most operating systems and graphic backends.

Que 2. What are some of the major components of any graphs or plot?

Ans. Major components of charts are:-

(a) Figure , (b) Axes , (c) Axis Label, (d) Limits , (e) Ticks, (f) Chart Title, and (g) Legend

Que 3. Name the function which is used to save the plot.

Ans. pyplot.savefig()

Que 4. Write short notes on different customisation options available with any plot.

Ans. We can specify own title, axes labels, legends, color of markers, lines and bars.

Que 5. What is the purpose of a legend?

Ans. In a chart/graph, there may be multiple datasets plotted. To distinguish among various datasets plotted in the same chart, legends are used. Legends can be different colors/patterns assigned to different specific datasets. The legends are shown in a corner of a chart/graph.

Que 6. Define Pandas visualisation.

Ans. The graphical or visual representation of information and data using visual elements like charts, graphs, etc, called Visualisation.

Que 7. What is open data? Name any two websites from which we can download open data.

Ans. Data freely available for downloading, for research and analysis purposes, is known as open data. There are many websites, which make available open data, such as data.gov.in, which is Open Government Data (OGD) platform of Indian government.

Que 8. Give an example of data comparison where we can use the scatter plot.

Ans. To show the performance of a class in terms of marks scored, we may use scatter plot.

Que 9. Name the plot which displays the statistical summary.

Note: Give appropriate title, set xlabel and ylabel while attempting the following questions.

Ans. Boxplot

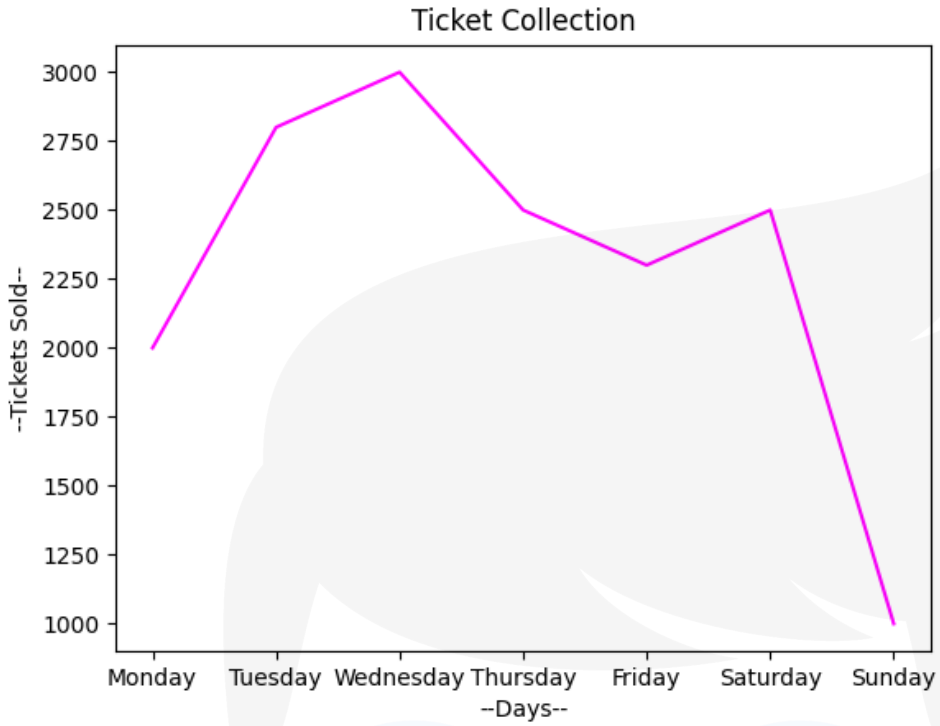
Que 10. Plot the following data using a line plot:

Day	1	2	3	4	5	6	7
Tickets sold	2000	2800	3000	2500	2300	2500	1000

- **Before displaying the plot display "Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday" in place of Day 1, 2, 3, 4, 5, 6, 7**
- **Change the color of the line to 'Magenta'.**

Ans.

```
import matplotlib.pyplot as plt

Day = list(range(1,8))
TicketsSold = [2000, 2800, 3000, 2500, 2300,
               2500, 1000]

plt.title("Ticket Collection")
plt.plot(Day, TicketsSold, color = 'magenta')

DayName = ['Monday', 'Tuesday', 'Wednesday',
           'Thursday', 'Friday', 'Saturday',
           'Sunday']
plt.xticks(Day, DayName)

plt.xlabel("--Days--")
plt.ylabel("--Tickets Sold--")

plt.show()
```



Que 11. Collect data about colleges in Delhi University or any other university of your choice and number of courses they run for Science, Commerce and Humanities, store it in a CSV file and present it using a bar plot.

Ans.

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

cdf = pd.read_csv("d:\\MyPythonProg\\courses.csv",
                  names =["Streams", "Courses"])

cdf.plot(x = "Streams", y = "Courses",
         ylabel = "No. of Courses", kind="bar")
plt.title("University Stream vs Courses")
plt.show()
```

Que 12. Collect and store data related to the screen time of students in your class separately for boys and girls and present it using a boxplot.

Ans.

```
import matplotlib.pyplot as plt
girls = [10.52, 5.3, 6.8, 12, 2, 6, 4, 4]
boys = [15, 12, 11, 5, 8.5, 8.8, 6.3, 6.3]

plt.boxplot([girls, boys], labels = ['Girls', 'Boys'])

plt.ylabel("Screen Time")
plt.title("Student's Screen Time")

plt.show()
```

Que 13. Explain the findings of the boxplot of Figure 4.17 by filling the following blanks:

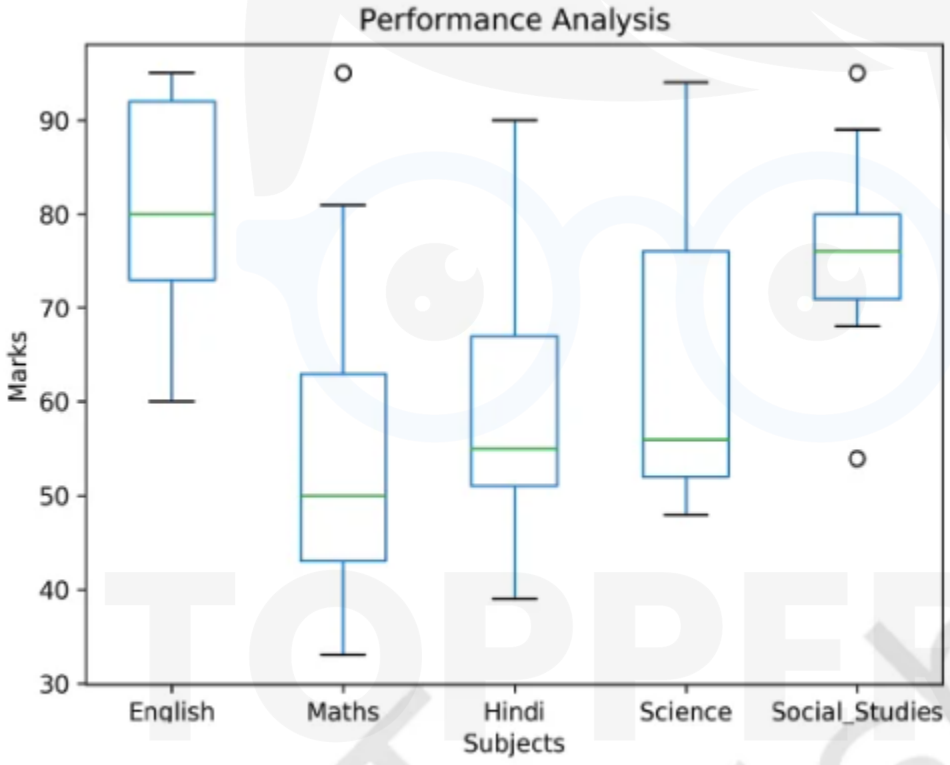


Figure 4.17: A boxplot of "Marks.csv"

a) The median for the five subjects is ____ , _____, _____, _____, _____

b) The highest value for the five subjects is : ____ , _____, _____, _____, _____

c) The lowest value for the five subjects is : ____ , _____, _____, _____, _____

d) _____ subject has two outliers with the value _____ and _____

e) _____ subject shows minimum variation

Ans. a) The median for the five subjects is **80, 50, 55, 56, 76 .**
b) The highest value for the five subjects is : **95, 95, 90, 94, 95 .**
c) The lowest value for the five subjects is : **60, 33, 39, 48, 54** .
d) **Soical_Studies** subject has two outliers with the value **95** and **54** .
e) **Social_Studies** subject shows minimum variation

Que 14. Collect the minimum and maximum temperature of your city for a month and present it using a histogram plot.

Ans. Do your self

Que 15. Conduct a class census by preparing a questionnaire. The questionnaire should contain a minimum of five questions. Questions should relate to students, their family members, their class performance, their health etc. Each student is required to fill up the questionnaire. Compile the information in numerical terms (in terms of percentage). Present the information through a bar, scatter–diagram. (NCERT Geography class IX, Page 60)

Ans. Do your self

Que 16. Visit data.gov.in , search for the following in "catalogs" option of the website:

- **Final population Totals, India and states**
- **State Wise literacy rate**

Download them and create a CSV file containing population data and literacy rate of the respective state. Also add a column Region to the CSV file that should contain the values East, West, North and South. Plot a scatter plot for each region where X axis should be population and Y axis should be Literacy rate. Change the marker to a diamond and size as the square root of the literacy rate.

Group the data on the column region and display a bar chart depicting average literacy rate for each region.

Ans. Do your self