BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai – 400058-India
Department of Computer Engineering

Name	Sujal Sandeep Dingankar
UID no.	2024301005
Experiment No.	5

AIM:	**Learning Data Visualization Libraries**
Program 1	
PROBLEM STATEMENT :	In the '**Marks.csv**' file, you can find the scores obtained by 200 students in 4 subjects of a standardised test. The different columns - Score A, Score B, Score C and Score D indicate the score obtained by a particular student in the respective subjects A, B, C and D.Load the dataset to your notebook and answer the following questions 1. Load the dataset and plot a histogram for the Score A column by keeping the number of bins to 6. Which bin range among the following has the highest frequency? 2. Plot a box plot for the column Score C and choose the correct option. In the **superstore.csv** file, you have the details of orders purchased in an American online retail store. Load the dataset, observe and analyse the different columns and answer the following questions. 1. Plot a pie-chart to find the Ship Mode through which most of the orders are being delivered. 2. Plot a bar chart comparing the average Discount across all 3. the Regions and report back the Region getting the highest average discount

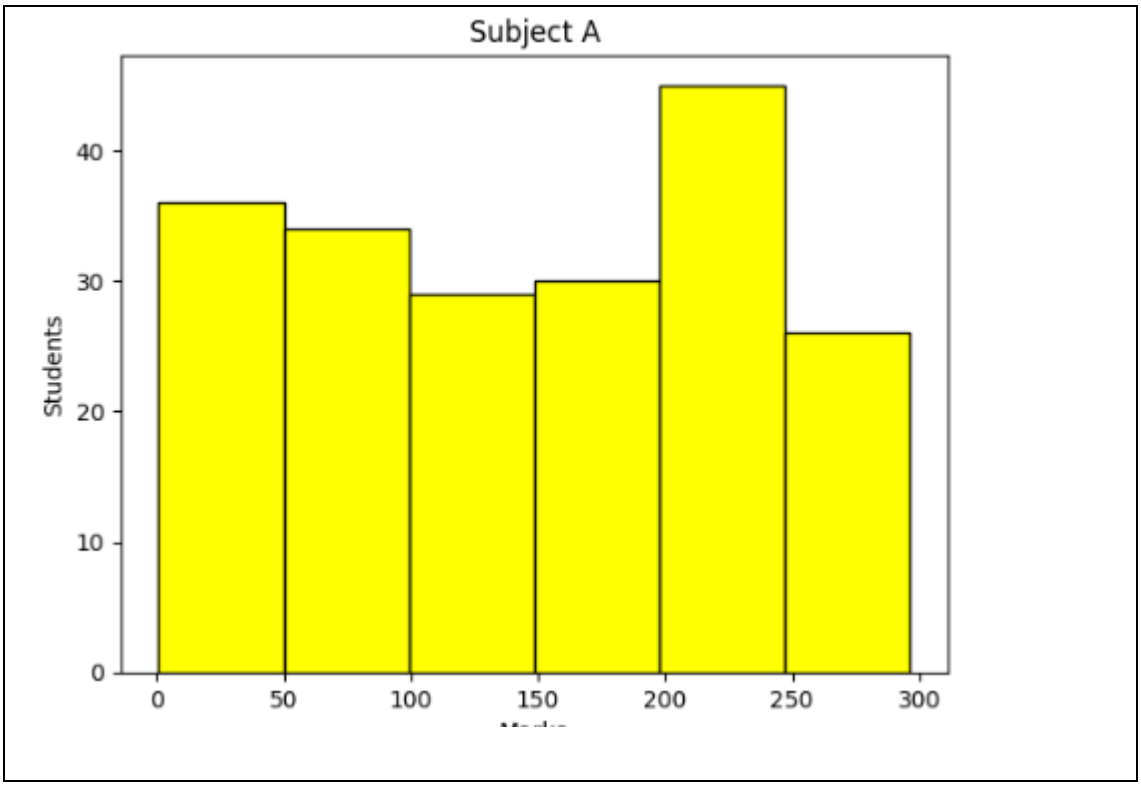
PROGRAM:

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

# Reading data from CSV files
df1 = pd.read_csv('Marks.csv')  # Marks data
df2 = pd.read_csv('superstore.csv')  # Superstore data

# Creating a Histogram for 'Score A' column
plt.hist(df1['Score A'], bins=6, color='yellow', edgecolor='black')
plt.title('Subject A')
plt.xlabel('Marks')
plt.ylabel('Students')
plt.show()

# Creating a boxplot for 'Score C' column
sns.boxplot(x=df1['Score C'])
plt.title('Box Plot for Subject C column')
plt.xlabel('Score')
plt.ylabel('Frequency')
plt.show()

# Creating a Pie chart for 'Ship Mode'
mylabels = ["Standard Class", "First Class", "First Class", "Same Day"]
y = df2['Ship Mode'].value_counts()
myexplode = [0.2, 0, 0, 0]
plt.pie(y, labels=mylabels, explode=myexplode)
plt.legend()
plt.show()

# Converting 'Discount' column to float and calculating average discount by
region
df2['Discount'] = df2['Discount'].astype(str).str.rstrip('%').astype('float') /
100
average_discount_by_region = df2.groupby('Region')['Discount'].mean()

# Creating a bar chart for Average Discount by Region
```

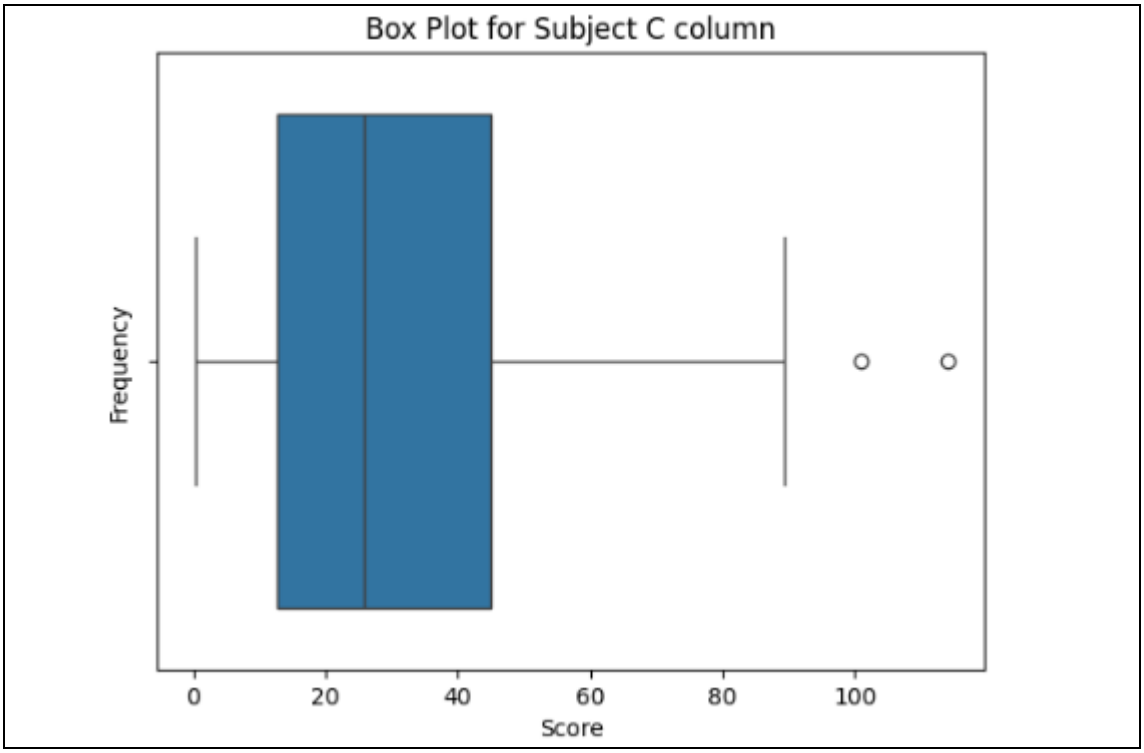
```
plt.figure(figsize=(8, 8))
plt.bar(average_discount_by_region.index,
average_discount_by_region.values)
plt.xlabel('Region')
plt.ylabel('Average Discount')
plt.title('Average Discount by Region')
plt.show()

# Finding the region with the highest average discount
region_with_highest_discount = average_discount_by_region.idxmax()
print(f"The region with the highest average discount is:
{region_with_highest_discount}")

# Extra 3 graphs

# Pie Chart for 'Segment' column
mylabels = ["Corporate", "Consumer", "Home Office"]
y = df2['Segment'].value_counts()
myexplode = [0.1, 0.1, 0]
colors = ['green', 'gray', 'purple']
plt.pie(y, labels=mylabels, explode=myexplode, colors=colors)
plt.legend()
plt.show()

# Bar Chart for Average Marks per Subject
avg_marks = df1[['Score A', 'Score B', 'Score C', 'Score D']].mean()
plt.bar(avg_marks.index, avg_marks.values, color='orange')
plt.xlabel('Subject')
plt.ylabel('Average Marks')
plt.title('Average Marks in Each Subject')
plt.show()

# Bar Chart for Region-wise Profit
profit_by_region = df2.groupby('Region')['Profit'].sum()
plt.bar(profit_by_region.index, profit_by_region.values, color='#C02323')
plt.xlabel('Region')
plt.ylabel('Profit')
plt.title('Profit per Region')
```

	plt.show()

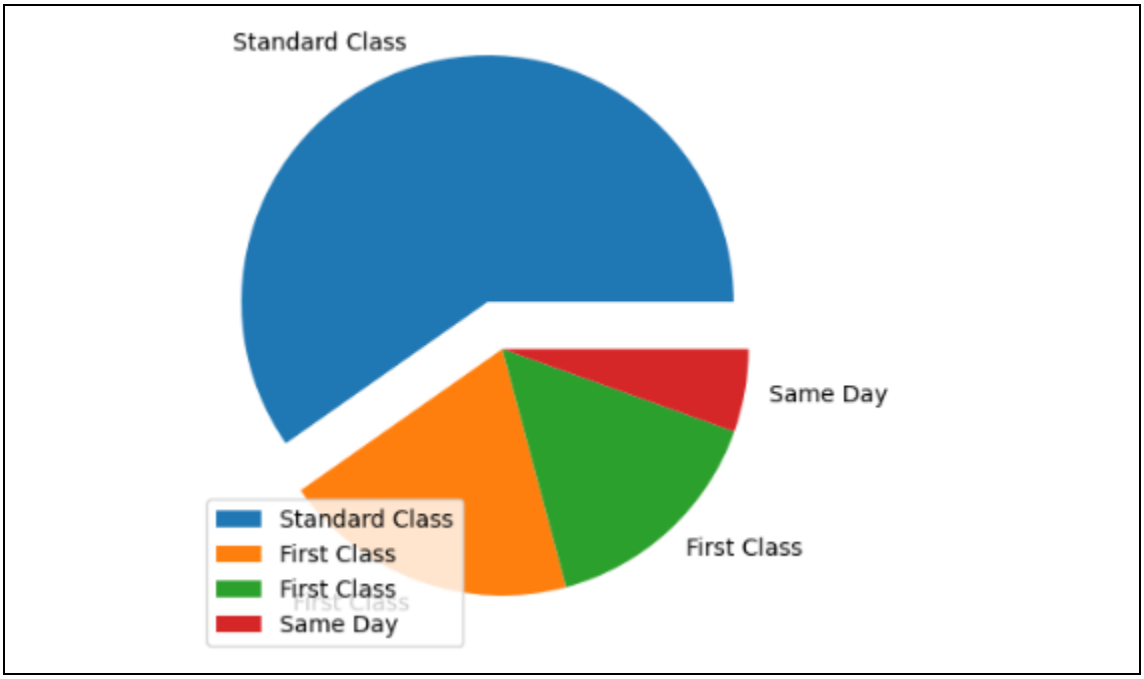
RESULT:



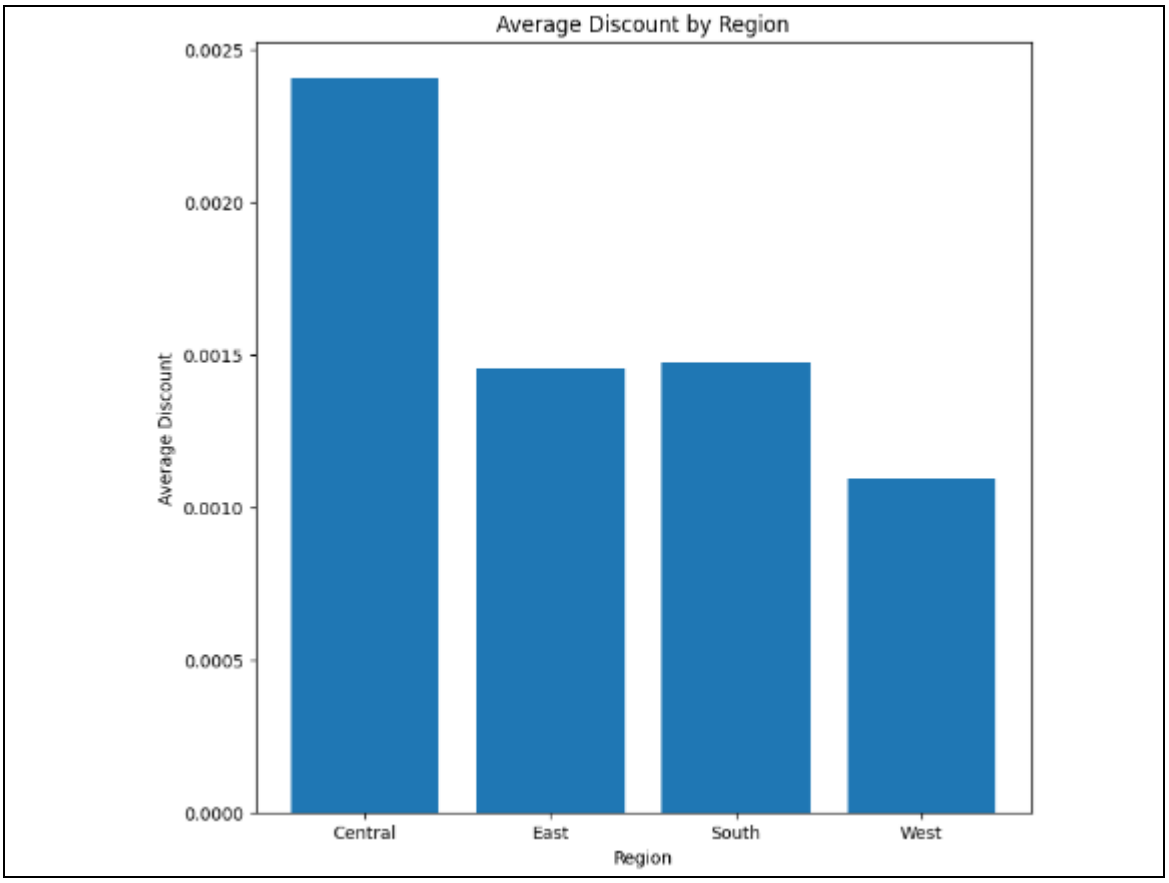
Bin range 200-250 has the highest frequency



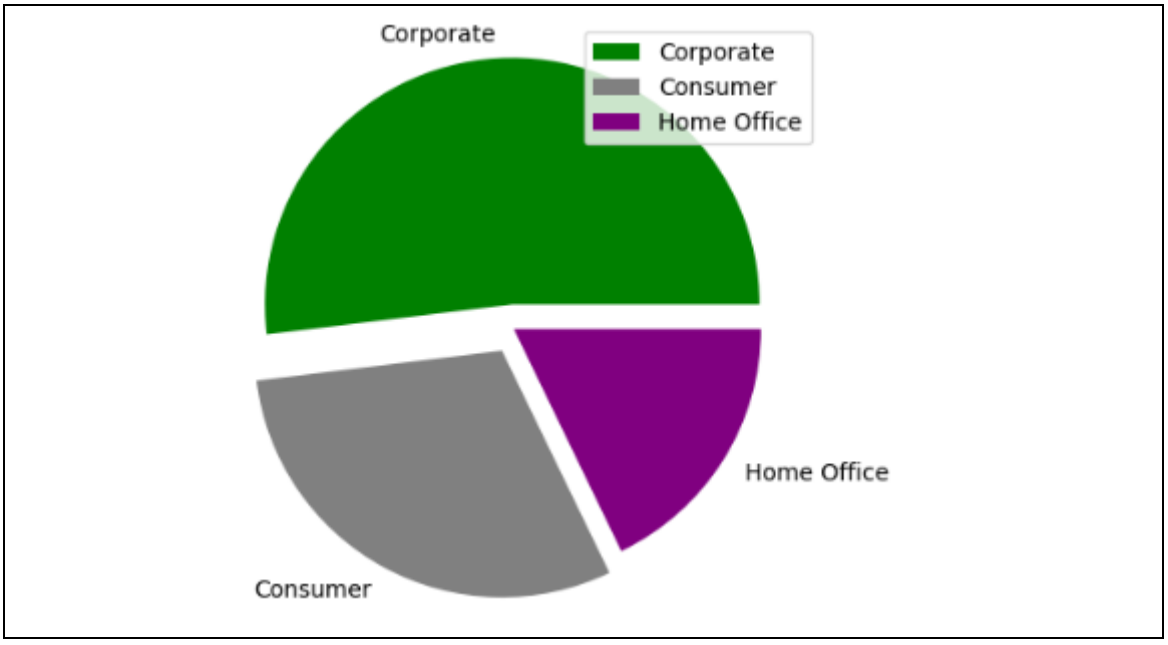
The 75th percentile lies between 40 and 60 while the 25th percentile lies between 0 and 20



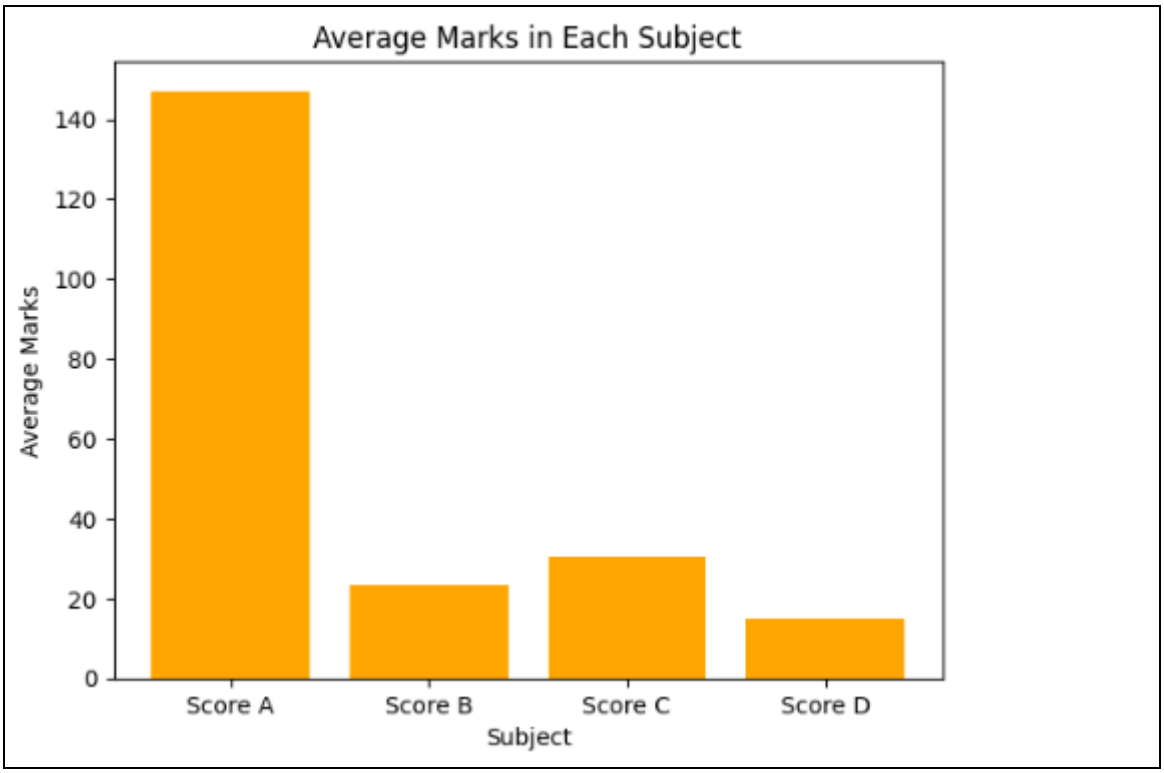
The Ship Mode through which most of the orders are being delivered is Standard Class.



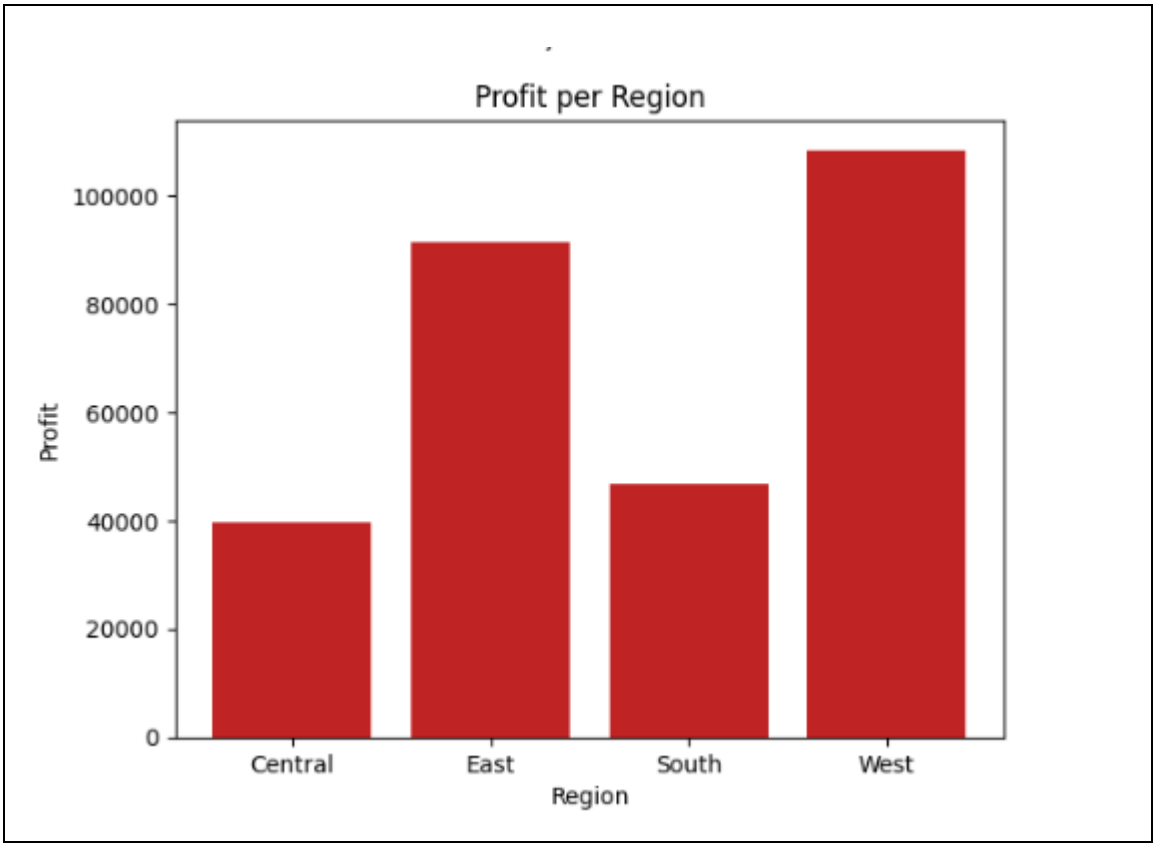
The region with the highest average discount is: Central



Majority of the orders are given by the Corporate segment while Home Office gives least orders.



Subject A has the highest average marks while Subject D has least average marks.



The graph shows the Profit per region. It is evident that West Region is the most profitable Followed by East, South and Central regions.

CONCLUSION:	In this experiment, I was able to understand the Matplotlib and seaborn libraries in Python and how to create histrograms, bar plots, boxplots and other graphs with help of it.