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.	4

AIM:	**Learning Pandas Library**
Program 1	
PROBLEM STATEMENT :	**Analyse Sales Data** 1. Write a Pandas program to capitalize all the string values of specified columns of a given DataFrame. 2. Write a Pandas program to create a Pivot table and find the total sale amount region wise, manager wise. (Salesdata.xlsx). 3. Create a DataFrame from a dictionary of lists and perform following functions: a. Select the first 3 rows of a DataFrame. b. Add a new column to an existing DataFrame. c. Group a DataFrame by a column and calculate the mean of each group. d. Concatenate two DataFrames. 4. Write a Pandas program to create a line plot of the historical stock prices of Alphabet Inc. between two specific dates. Extract data from alphabet_stock_data.csv file. 5. Write a Pandas program to count the number of missing values in each column of a given DataFrame.

PROGRAM:

```
import pandas as pd

# Reading Excel file
df = pd.read_excel('SaleData.xlsx')

# Write a Pandas program to create a Pivot table and find the total sale
# amount region wise, manager wise.
print("Pivot Table:")
pivot_table = pd.pivot_table(df, values='Sale_amt', index=['Region',
'Manager'], aggfunc='sum')
print(pivot_table)

# Write a Pandas program to capitalize all the string values of the specified
column of a given DataFrame.
print("\nCapitalized column 'Manager':")
df['Manager'] = df['Manager'].str.upper()
print(df)

# Create a DataFrame from a dictionary of lists and perform different
functions:
data1 = {
   'Name': ['Rahul', 'Ramesh', 'Raj'],
   'Age': [12, 33, 21],
   'City': ['Mumbai', 'Delhi', 'Pune']
}

df1 = pd.DataFrame(data1)
print("\nFirst 3 rows of the DataFrame:\n", df1.head(3))

# Add a new column to an existing DataFrame.
df1['Profession'] = ['Student', 'Teacher', 'Manager']
print("\nAfter adding a new column:\n", df1)

# Group by 'City' and calculate the mean age for each city.
Mean = df1.groupby('City')['Age'].mean()
print("\nMean Age by City:\n", Mean)

# Create another DataFrame and concatenate them.
```

```
data2 = {
   'Name': ['Soham', 'Sahil', 'Harsh'],
   'Age': [22, 13, 19],
   'City': ['Mumbai', 'Delhi', 'Pune'],
   'Profession': ['Student', 'Web Developer', 'Data Scientist']
}

df2 = pd.DataFrame(data2)

# Concatenate two DataFrames.
concat = pd.concat([df1, df2], ignore_index=True)
print("\nConcatenated DataFrame:\n", concat)

# Reading CSV for Alphabet Stock Data
adf = pd.read_csv('alphabet_stock_data.csv')

# Write a Pandas program to create a line plot of the historical stock prices
of Alphabet Inc. between two specific dates.
adf['Date'] = pd.to_datetime(adf['Date'])
start_date = '2020-05-22'
end_date = '2020-08-31'

# Filter the DataFrame based on the date range.
filtered_df = adf[(adf['Date'] >= start_date) & (adf['Date'] <= end_date)]

# Create a line plot of the 'Close' prices.
filtered_df.plot(x='Date', y='Close', kind='line', figsize=(10, 7), marker='o',
title='Alphabet Stock Prices')

# Write a Pandas program to count the number of missing values in each
column of a given DataFrame.
missing_values = df.isnull().sum()
print("\nFor Alphabet Stock Data:\nMissing Values:\n", missing_values)
```

RESULT:

Pivot Table:

Region	Manager	Sale_amt
Central	Douglas	124016.0
	Hermann	365108.5
	Martha	199690.0
	Timothy	140955.0
East	Douglas	48204.0
	Martha	272803.0
West	Douglas	66836.0
	Timothy	88063.0

Capitalized column 'Manager':

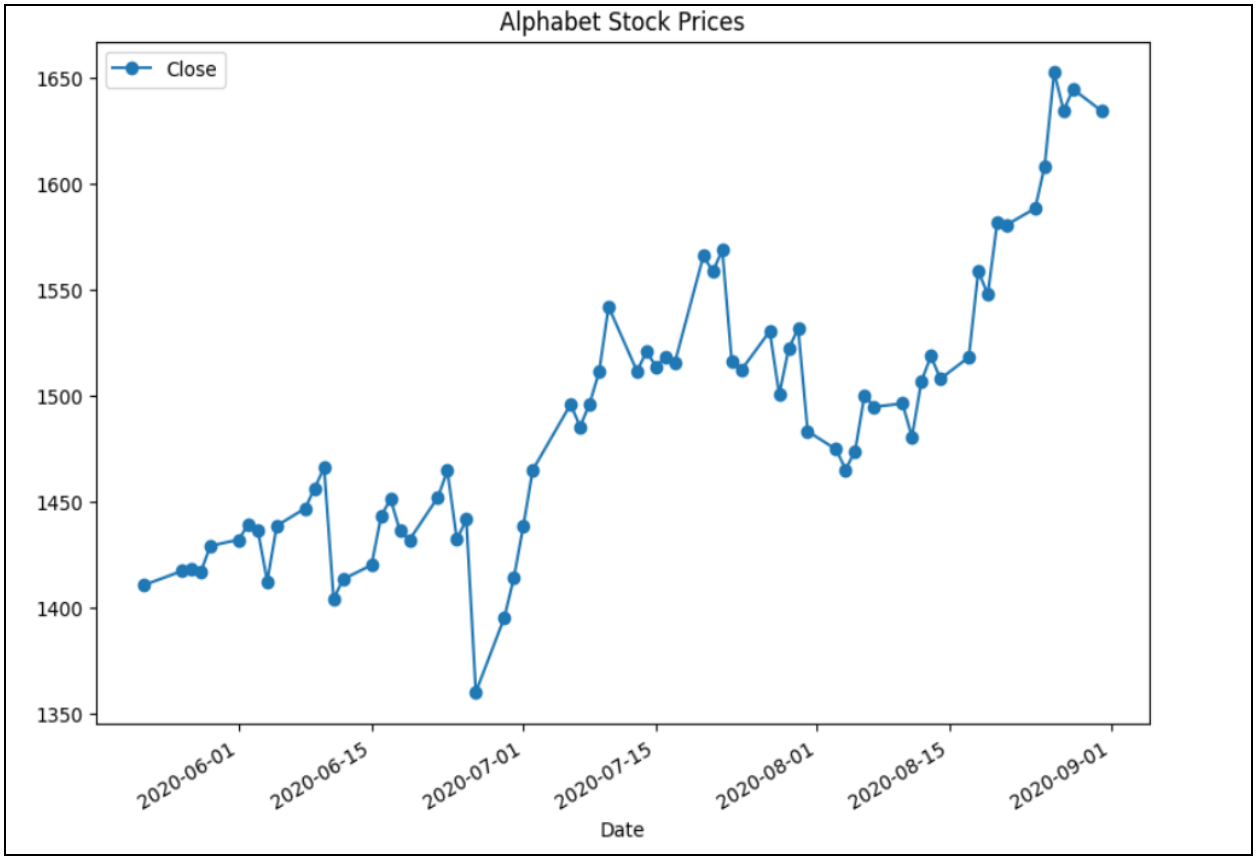
	OrderDate	Region	Manager	SalesMan	Item	Units	Unit_price
0	2018-01-06	East	MARTHA	Alexander	Television	95.00	1198.000
1	2018-01-23	Central	HERMANN	Shelli	Home Theater	50.00	500.000
2	2018-02-09	Central	HERMANN	Luis	Television	36.00	1198.000
3	2018-02-26	Central	TIMOTHY	David	Cell Phone	27.00	225.000
4	2018-03-15	West	TIMOTHY	Stephen	Television	56.00	1198.000
5	2018-04-01	East	MARTHA	Alexander	Home Theater	60.00	500.000
6	2018-04-18	Central	MARTHA	Steven	Television	75.00	1198.000
7	2018-05-05	Central	HERMANN	Luis	Television	90.00	1198.000
8	2018-05-22	West	DOUGLAS	Michael	Television	32.00	1198.000
9	2018-06-08	East	MARTHA	Alexander	Home Theater	60.00	500.000
10	2018-06-25	Central	HERMANN	Sigal	Television	90.00	1198.000
11	2018-07-12	East	MARTHA	Diana	Home Theater	29.00	500.000
12	2018-07-29	East	DOUGLAS	Karen	Home Theater	81.00	500.000
13	2018-08-15	East	MARTHA	Alexander	Television	35.00	1198.000
14	2018-09-01	Central	DOUGLAS	John	Desk	2.00	125.000
15	2018-09-18	East	MARTHA	Alexander	Video Games	16.00	58.500
16	2018-10-05	Central	HERMANN	Sigal	Home Theater	28.00	500.000
17	2018-10-22	East	MARTHA	Alexander	Cell Phone	64.00	225.000
18	2018-11-08	East	DOUGLAS	Karen	Cell Phone	15.00	225.000
19	2018-11-25	Central	HERMANN	Shelli	Video Games	96.00	58.500
20	2018-12-12	Central	DOUGLAS	John	Television	67.00	1198.000
21	2018-12-29	East	DOUGLAS	Karen	Video Games	74.00	58.500
22	2019-01-15	Central	TIMOTHY	David	Home Theater	46.00	500.000
23	2019-02-01	Central	DOUGLAS	John	Home Theater	87.00	500.000
24	2019-02-18	East	MARTHA	Alexander	Home Theater	4.00	500.000
25	2019-03-07	West	TIMOTHY	Stephen	Home Theater	7.00	500.000

26	2019-03-24	Central	HERMANN	Luis	Video Games	50.00	58.500
27	2019-04-10	Central	MARTHA	Steven	Television	66.00	1198.000
28	2019-04-27	East	MARTHA	Diana	Cell Phone	96.00	225.000
29	2019-05-14	Central	TIMOTHY	David	Television	53.00	1198.000
30	2019-05-31	Central	TIMOTHY	David	Home Theater	80.00	500.000
31	2019-06-17	Central	HERMANN	Shelli	Desk	5.00	125.000
32	2019-07-04	East	MARTHA	Alexander	Video Games	62.00	58.500
33	2019-07-21	Central	HERMANN	Sigal	Video Games	55.00	58.500
34	2019-08-07	Central	HERMANN	Shelli	Video Games	42.00	58.500
35	2019-08-24	West	TIMOTHY	Stephen	Desk	3.00	125.000
36	2019-09-10	Central	TIMOTHY	David	Television	7.00	1198.000
37	2019-09-27	West	TIMOTHY	Stephen	Cell Phone	76.00	225.000
38	2019-10-14	West	DOUGLAS	Michael	Home Theater	57.00	500.000
39	2019-10-31	Central	MARTHA	Steven	Television	14.00	1198.000
40	2019-11-17	Central	HERMANN	Luis	Home Theater	11.00	500.000
41	2019-12-04	Central	HERMANN	Luis	Home Theater	94.00	500.000
42	2019-12-21	Central	MARTHA	Steven	Home Theater	28.00	500.000
43	NaT	NaN	NaN	NaN	NaN	278.00	1125.000
44	NaT	NaN	NaN	NaN	NaN	34.75	140.625

	Sale_amt
0	113810.00
1	25000.00
2	43128.00
3	6075.00
4	67088.00
5	30000.00
6	89850.00
7	107820.00
8	38336.00
9	30000.00
10	107820.00
11	14500.00
12	40500.00
13	41930.00
14	250.00
15	936.00
16	14000.00
17	14400.00

```
18    3375.00
19    5616.00
20   80266.00
21    4329.00
22   23000.00
23   43500.00
24    2000.00
25    3500.00
26    2925.00
27   79068.00
28   21600.00
29   63494.00
30   40000.00
31     625.00
32    3627.00
33    3217.50
34    2457.00
35     375.00
36    8386.00
37   17100.00
38   28500.00
39   16772.00
40    5500.00
41   47000.00
42   14000.00
43   62550.00
44    7818.75

First 3 rows of the DataFrame:
      Name  Age    City
0   Rahul   12  Mumbai
1  Ramesh   33   Delhi
2     Raj   21    Pune

After adding a new column:
      Name  Age    City Profession
0   Rahul   12  Mumbai    Student
1  Ramesh   33   Delhi    Teacher
2     Raj   21    Pune    Manager
```

```
Mean Age by City:
 City
Delhi     33.0
Mumbai    12.0
Pune      21.0
Name: Age, dtype: float64

Concatenated DataFrame:
      Name  Age    City       Profession
0   Rahul   12  Mumbai          Student
1  Ramesh   33   Delhi          Teacher
2     Raj   21    Pune          Manager
3   Soham   22  Mumbai          Student
4   Sahil   13   Delhi   Web Developer
5   Harsh   19    Pune  Data Scientist

For Alphabet Stock Data:
Missing Values:
 OrderDate     2
Region         2
Manager        2
SalesMan       2
Item           2
Units          0
Unit_price     0
Sale_amt       0
dtype: int64
```



CONCLUSION:	Through this experiment, I got well versed with the extensive library of Pandas in python and learned many functions we can use with the help of it. I also understood about the different applications of pandas in data science, financing and machine learning.