Vaishnavi Ashok Patil
C07
SPP-II

Experiment No:14

Aim: Create DataFrames, load data from CSV, and perform basic data analysis such as filtering, grouping, and aggregation.

1. **Create a CSV file named employees.csv with columns Name, Age, Department, Salary. Load the data, filter employees older than 30, group by Department, and calculate the average, minimum, and maximum salary for each department.**

```
[1]: import pandas as pd

data = {
    'Name': ['Asha', 'Ravi', 'Neha', 'Amit', 'Kiran'],
    'Age': [25, 35, 40, 28, 32],
    'Department': ['HR', 'IT', 'Finance', 'HR', 'IT'],
    'Salary': [40000, 60000, 75000, 45000, 80000]
}
df = pd.DataFrame(data)
df.to_csv('employees.csv', index=False)

df = pd.read_csv('employees.csv')
filtered = df[df['Age'] > 30]
result = filtered.groupby('Department')['Salary'].agg(['mean', 'min', 'max'])
print(result)
```

```
                mean    min    max
Department
Finance      75000.0  75000  75000
IT           70000.0  60000  80000
```

2. **Create a CSV file named sales.csv with columns SaleID, Product, Amount, Customer. Load the data, filter sales with amount greater than 50,000, group by Product, and calculate total and average sales per product.**

```
[2]: import pandas as pd

data = {
    'SaleID': [1,2,3,4,5],
    'Product': ['TV','AC','Laptop','AC','TV'],
    'Amount': [55000,48000,75000,65000,52000],
    'Customer': ['A','B','C','A','D']
}
df = pd.DataFrame(data)
df.to_csv('sales.csv', index=False)

df = pd.read_csv('sales.csv')
filtered = df[df['Amount'] > 50000]
result = filtered.groupby('Product')['Amount'].agg(['sum','mean'])
print(result)
```

```
             sum     mean
Product
AC         65000  65000.0
Laptop     75000  75000.0
TV        107000  53500.0
```

3. **Create a CSV file named student_marks.csv with columns StudentID, Name, Maths, Science, English. Load the data, filter students with Maths marks above 80, group by student name, and calculate total and average marks per student.**

Vaishnavi Ashok Patil
C07
SPP-II

```
[3]: import pandas as pd

data = {
    'StudentID':[1,2,3,4,5],
    'Name':['Asha','Ravi','Neha','Amit','Kiran'],
    'Maths':[85,70,90,60,95],
    'Science':[78,88,92,55,80],
    'English':[80,75,85,70,90]
}
df = pd.DataFrame(data)
df.to_csv('student_marks.csv', index=False)

df = pd.read_csv('student_marks.csv')
filtered = df[df['Maths'] > 80]
filtered['Total'] = filtered[['Maths','Science','English']].sum(axis=1)
filtered['Average'] = filtered[['Maths','Science','English']].mean(axis=1)
print(filtered[['Name','Total','Average']])
```

```
    Name  Total    Average
0   Asha    243  81.000000
2   Neha    267  89.000000
4  Kiran    265  88.333333
```

4. **Create a CSV file named products.csv with columns ProductID, ProductName, Category, Stock. Load the data, filter products with stock less than 50, group by Category, and calculate the total and average stock per category.**

```
[4]: import pandas as pd

data = {
    'ProductID':[1,2,3,4,5],
    'ProductName':['Pen','Book','Laptop','Phone','Bag'],
    'Category':['Stationery','Stationery','Electronics','Electronics','Accessories'],
    'Stock':[120,40,30,10,80]
}
df = pd.DataFrame(data)
df.to_csv('products.csv', index=False)

df = pd.read_csv('products.csv')
filtered = df[df['Stock'] < 50]
result = filtered.groupby('Category')['Stock'].agg(['sum','mean'])
print(result)
```

```
             sum  mean
Category
Electronics   40  20.0
Stationery    40  40.0
```

5. **Create a CSV file named transactions.csv with columns TransactionID, Customer, Amount, Date. Load the data, filter transactions with amount greater than 1000, group by Customer, and calculate total and average spending per customer.**

```
[5]: import pandas as pd

data = {
    'TransactionID':[1,2,3,4,5],
    'Customer':['A','B','A','C','B'],
    'Amount':[500,1500,2500,700,2000],
    'Date':['2024-03-01','2024-03-02','2024-03-03','2024-03-04','2024-03-05']
}
df = pd.DataFrame(data)
df.to_csv('transactions.csv', index=False)

df = pd.read_csv('transactions.csv')
filtered = df[df['Amount'] > 1000]
result = filtered.groupby('Customer')['Amount'].agg(['sum','mean'])
print(result)
```

```
           sum    mean
Customer
A         2500  2500.0
B         3500  1750.0
```

Vaishnavi Ashok Patil
C07
SPP-II

6. **Create a CSV file named employee_salary.csv with columns EmpID, Name, Department, Salary. Load the data, filter employees with salary above 60,000, group by Department, and calculate mean, max, and min salary per department.**

```
[6]: import pandas as pd

data = {
    'EmpID':[1,2,3,4,5],
    'Name':['Asha','Ravi','Neha','Amit','Kiran'],
    'Department':['HR','IT','Finance','IT','Finance'],
    'Salary':[50000,70000,85000,90000,45000]
}
df = pd.DataFrame(data)
df.to_csv('employee_salary.csv', index=False)

df = pd.read_csv('employee_salary.csv')
filtered = df[df['Salary'] > 60000]
result = filtered.groupby('Department')['Salary'].agg(['mean','max','min'])
print(result)
```

```
                mean    max    min
Department
Finance      85000.0  85000  85000
IT           80000.0  90000  70000
```

7. **Create a CSV file named orders.csv with columns OrderID, Customer, Product, Quantity, OrderDate. Load the data, filter orders with quantity greater than 5, group by Product, and calculate total quantity sold per product.**

```
[7]: import pandas as pd

data = {
    'OrderID':[1,2,3,4,5],
    'Customer':['A','B','C','A','D'],
    'Product':['TV','AC','TV','Laptop','AC'],
    'Quantity':[2,8,6,3,10],
    'OrderDate':['2024-03-01','2024-03-02','2024-03-03','2024-03-04','2024-03-05']
}
df = pd.DataFrame(data)
df.to_csv('orders.csv', index=False)

df = pd.read_csv('orders.csv')
filtered = df[df['Quantity'] > 5]
result = filtered.groupby('Product')['Quantity'].sum()
print(result)
```

```
Product
AC    18
TV     6
Name: Quantity, dtype: int64
```

8. **Create a CSV file named movies.csv with columns MovieID, Title, Genre, Rating. Load the data, filter movies with rating above 8, group by Genre, and calculate average, maximum, and minimum rating per genre.**

Vaishnavi Ashok Patil
C07
SPP-II

[8]:
```
import pandas as pd

data = {
    'MovieID':[1,2,3,4,5],
    'Title':['Inception','Titanic','Avatar','Interstellar','Joker'],
    'Genre':['Sci-Fi','Romance','Sci-Fi','Sci-Fi','Drama'],
    'Rating':[8.8,7.9,8.2,8.6,9.0]
}
df = pd.DataFrame(data)
df.to_csv('movies.csv', index=False)

df = pd.read_csv('movies.csv')
filtered = df[df['Rating'] > 8]
result = filtered.groupby('Genre')['Rating'].agg(['mean','max','min'])
print(result)
```

```
            mean  max  min
Genre
Drama   9.000000  9.0  9.0
Sci-Fi  8.533333  8.8  8.2
```

9. **Create a CSV file named weather.csv with columns Date, City, Temperature, Humidity. Load the data, filter days with temperature above 35°C, group by City, and calculate maximum, minimum, and average temperature for each city.**

[9]:
```
import pandas as pd

data = {
    'Date':['2024-05-01','2024-05-02','2024-05-03','2024-05-04','2024-05-05'],
    'City':['Mumbai','Delhi','Chennai','Delhi','Mumbai'],
    'Temperature':[36,40,34,38,37],
    'Humidity':[70,55,80,60,75]
}
df = pd.DataFrame(data)
df.to_csv('weather.csv', index=False)

df = pd.read_csv('weather.csv')
filtered = df[df['Temperature'] > 35]
result = filtered.groupby('City')['Temperature'].agg(['max','min','mean'])
print(result)
```

```
        max  min  mean
City
Delhi    40   38  39.0
Mumbai   37   36  36.5
```

10. **Create a CSV file named bank_transactions.csv with columns TransactionID, Customer, Type, Amount. Load the data, filter deposits, group by Customer, and calculate total, maximum, and average deposit amount per customer.**

[10]:
```
import pandas as pd

data = {
    'TransactionID':[1,2,3,4,5],
    'Customer':['A','B','C','A','B'],
    'Type':['Deposit','Withdraw','Deposit','Deposit','Deposit'],
    'Amount':[1000,2000,3000,4000,5000]
}
df = pd.DataFrame(data)
df.to_csv('bank_transactions.csv', index=False)

df = pd.read_csv('bank_transactions.csv')
filtered = df[df['Type'] == 'Deposit']
result = filtered.groupby('Customer')['Amount'].agg(['sum','max','mean'])
print(result)
```

```
            sum   max    mean
Customer
A          5000  4000  2500.0
B          5000  5000  5000.0
C          3000  3000  3000.0
```

Vaishnavi Ashok Patil
C07
SPP-II

11. **Create a CSV file named hospital.csv with columns PatientID, Name, Department, AdmissionDate. Load the data, filter patients admitted in the last month, group by Department, and calculate the total number of patients per department.**

```
[11]: import pandas as pd
from datetime import datetime, timedelta

data = {
    'PatientID':[1,2,3,4,5],
    'Name':['Asha','Ravi','Neha','Amit','Kiran'],
    'Department':['Cardio','Ortho','Cardio','Neuro','Ortho'],
    'Admission Date':[
        '2024-10-10','2024-11-01','2024-10-25','2024-09-20','2024-11S-02'
    ]
}
df = pd.DataFrame(data)
df.to_csv('hospital.csv', index=False)

df = pd.read_csv('hospital.csv')
df['Admission Date'] = pd.to_datetime(df['Admission Date'])
last_month = datetime.now() - timedelta(days=30)
filtered = df[df['Admission Date'] > last_month]
result = filtered.groupby('Department')['PatientID'].count()
print(result)
```

```
Series([], Name: PatientID, dtype: int64)
```

12. **Create a CSV file named online_store.csv with columns OrderID, Customer, Product, Amount. Load the data, filter orders with amount greater than 200, group by Customer, and calculate average, total, and maximum order amount per customer.**

```
[12]: import pandas as pd

data = {
    'OrderID':[1,2,3,4,5],
    'Customer':['A','B','A','C','D'],
    'Product':['Phone','TV','Laptop','AC','Watch'],
    'Amount':[250,180,600,300,1000]
}
df = pd.DataFrame(data)
df.to_csv('online_store.csv', index=False)

df = pd.read_csv('online_store.csv')
filtered = df[df['Amount'] > 200]
result = filtered.groupby('Customer')['Amount'].agg(['mean','sum','max'])
print(result)
```

```
            mean   sum   max
Customer
A          425.0   850   600
C          300.0   300   300
D         1000.0  1000  1000
```

13. **Create a CSV file named flights.csv with columns FlightID, Airline, Source, Destination, Delay. Load the data, filter flights with delay above 30 minutes, group by Airline, and calculate total flights, average, and maximum delay per airline.**

Vaishnavi Ashok Patil
C07
SPP-II

[13]:
```
import pandas as pd

data = {
    'FlightID':[1,2,3,4,5],
    'Airline':['Indigo','AirIndia','SpiceJet','Indigo','AirIndia'],
    'Source':['Mumbai','Delhi','Delhi','Chennai','Mumbai'],
    'Destination':['Delhi','Mumbai','Chennai','Delhi','Chennai'],
    'Delay':[10,45,60,5,35]
}
df = pd.DataFrame(data)
df.to_csv('flights.csv', index=False)

df = pd.read_csv('flights.csv')
filtered = df[df['Delay'] > 30]
result = filtered.groupby('Airline')['Delay'].agg(['count','mean','max'])
print(result)
```

```
          count  mean  max
Airline
AirIndia      2  40.0   45
SpiceJet      1  60.0   60
```

14. Create a CSV file named inventory.csv with columns ProductID, ProductName, Supplier, Stock. Load the data, filter out-of-stock products, group by Supplier, and calculate total out-of-stock products per supplier.

[14]:
```
import pandas as pd

data = {
    'ProductID':[1,2,3,4,5],
    'ProductName':['Pen','Book','Laptop','Bag','Phone'],
    'Supplier':['A1','A2','A1','A3','A2'],
    'Stock':[0,10,0,0,15]
}
df = pd.DataFrame(data)
df.to_csv('inventory.csv', index=False)

df = pd.read_csv('inventory.csv')
filtered = df[df['Stock'] == 0]
result = filtered.groupby('Supplier')['ProductID'].count()
print(result)
```

```
Supplier
A1    2
A3    1
Name: ProductID, dtype: int64
```

15. Create a CSV file named ecommerce.csv with columns OrderID, Category, SubCategory, Sales. Load the data, filter sales above 500, group by Category and SubCategory, and calculate total, average, and maximum sales per combination.

[15]:
```
import pandas as pd

data = {
    'OrderID':[1,2,3,4,5],
    'Category':['Electronics','Clothing','Electronics','Clothing','Home'],
    'SubCategory':['Mobile','Shirt','Laptop','Pants','Furniture'],
    'Sales':[800,300,1200,400,700]
}
df = pd.DataFrame(data)
df.to_csv('ecommerce.csv', index=False)

df = pd.read_csv('ecommerce.csv')
filtered = df[df['Sales'] > 500]
result = filtered.groupby(['Category','SubCategory'])['Sales'].agg(['sum','mean','max'])
print(result)
```

```
                        sum    mean   max
Category    SubCategory
Electronics Laptop       1200  1200.0  1200
            Mobile        800   800.0   800
Home        Furniture     700   700.0   700
```