In [ ]:
```
GURU PRASAD V 240701160
```

```
1.Create a pandas DataFrame using the following dictionary
```

In [1]:
```
import pandas as pd
data = {
'Name': ['Alice', 'Bob', 'Charlie'],
'Age': [25, 30, 35],
'City': ['New York', 'Los Angeles', 'Chicago']
}
df = pd.DataFrame(data)
print(df)
```

```
      Name  Age         City
0    Alice   25     New York
1      Bob   30  Los Angeles
2  Charlie   35      Chicago
```

```
2.Read a CSV file named data.csv into a DataFrame and print the first 5 rows.
```

In [2]:
```
import pandas as pd
df=pd.read_csv('data.csv')
print(df.head())
```

```
   SNO         SUBJECT  CAT 1 CAT 2  SEM
0    1             MAT     50    60   80
1    2            POAI     40    40   41
2    3  DATA STRUCTURE     54    64   60
3    4           TAMIL     40  None   70
4    5          PYTHON     65    60   80
```

```
3. Filter rows based on a condition
```

In [4]:
```
import pandas as pd
data = {
'Name': ['Alice', 'Bob', 'Charlie'],
'Age': [25, 30, 35],
'City': ['New York', 'Los Angeles', 'Chicago']
}
df = pd.DataFrame(data)
filtered_df = df[df['Age'] > 28]
print(filtered_df)
```

```
      Name  Age         City
1      Bob   30  Los Angeles
2  Charlie   35      Chicago
```

```
4.Add a new column
```

In [5]:
```
import pandas as pd
data = {
'Name': ['Alice', 'Bob', 'Charlie'],
'Age': [25, 30, 35],
'City': ['New York', 'Los Angeles', 'Chicago']
}
df = pd.DataFrame(data)
df['Salary'] = [50000, 60000, 70000]
print(df)
```

```
      Name  Age         City  Salary
0    Alice   25     New York   50000
1      Bob   30  Los Angeles   60000
2  Charlie   35      Chicago   70000
```

5.Group by a column and find the mean

In [6]:
```
import pandas as pd
data = {
'Department': ['HR', 'IT', 'HR', 'IT', 'Finance'],
'Salary': [40000, 60000, 42000, 62000, 50000]

}
df = pd.DataFrame(data)
mean_salary = df.groupby('Department')['Salary'].mean()
print(mean_salary)
```

```
Department
Finance    50000.0
HR         41000.0
IT         61000.0
Name: Salary, dtype: float64
```

6.Replace all occurrences of 'New York' in City with 'NYC'.

In [7]:
```
import pandas as pd
data = {
'Name': ['Alice', 'Bob', 'Charlie'],
'Age': [25, 30, 35],
'City': ['New York', 'Los Angeles', 'Chicago']
}
df = pd.DataFrame(data)
df['Salary'] = [50000, 60000, 70000]
df['City'] = df['City'].replace('New York', 'NYC')
print(df)
```

```
      Name  Age         City  Salary
0    Alice   25          NYC   50000
1      Bob   30  Los Angeles   60000
2  Charlie   35      Chicago   70000
```

7.Drop a column

In [8]:
```
import pandas as pd
data = {
'Name': ['Alice', 'Bob', 'Charlie'],
'Age': [25, 30, 35],
'City': ['New York', 'Los Angeles', 'Chicago']
}
df = pd.DataFrame(data)
df['Salary'] = [50000, 60000, 70000]
df['City'] = df['City'].replace('New York', 'NYC')
df = df.drop(columns=['Age'])
print(df)
```

```
      Name         City  Salary
0    Alice          NYC   50000
1      Bob  Los Angeles   60000
2  Charlie      Chicago   70000
```

8.Sort the DataFrame by a column

In [9]:
```
import pandas as pd
data = {
'Name': ['Alice', 'Bob', 'Charlie'],
'Age': [25, 30, 35],
'City': ['New York', 'Los Angeles', 'Chicago']
}
df = pd.DataFrame(data)
df['Salary'] = [50000, 60000, 70000]
df['City'] = df['City'].replace('New York', 'NYC')
df_sorted = df.sort_values(by='Salary', ascending=False)
print(df)
```

```
      Name  Age         City  Salary
0    Alice   25          NYC   50000
1      Bob   30  Los Angeles   60000
2  Charlie   35      Chicago   70000
```

9.Find missing values

In [12]:
```
import pandas as pd
data = {
'Name': ['Alice', 'Bob', 'Charlie'],
'Age': [25, 30, 35],
'City': ['New York', 'Los Angeles', 'Chicago']
}
df = pd.DataFrame(data)
df['Salary'] = [50000, 60000, 70000]
df['City'] = df['City'].replace('New York', 'NYC')
print(df.isnull(),"\n")
print(df.isnull().sum())
```

```
    Name    Age   City  Salary
0  False  False  False   False
1  False  False  False   False
2  False  False  False   False

Name      0
Age       0
City      0
Salary    0
dtype: int64
```

10.Save DataFrame to a CSV file

In [15]:
```
import pandas as pd
data = {
'Name': ['Alice', 'Bob', 'Charlie'],
'Age': [25, 30, 35],
'City': ['New York', 'Los Angeles', 'Chicago']
}
df = pd.DataFrame(data)
df['Salary'] = [50000, 60000, 70000]
df['City'] = df['City'].replace('New York', 'NYC')
df.to_csv('output.csv', index=False)
print(df)
```

```
      Name  Age         City  Salary
0    Alice   25          NYC   50000
1      Bob   30  Los Angeles   60000
2  Charlie   35      Chicago   70000
```