RAJALAKSHMI ENGINEERING COLLEGE

Approved by AICTE | Affiliated to Anna University | Accredited by NAAC

Department of Computer Science and Engineering

CS23334 Fundamentals of Data Science Lab

III semester II Year (2025)

Name of the Student : B J KAVIYA

Register Number : 240701246

Create a pandas DataFrame using the following dictionary:

In [3]:
```
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
```

Given a DataFrame with columns Department and Salary, find the average salary per department.

In [5]:
```
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
```

Add a new column Salary with values [50000, 60000, 70000].

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]
print(df)
```

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

Replaced all occurrence of 'New york'in city with'NYK'

In [14]:
```
import pandas as pd

data = {
'Name': ['Alice', 'Bob', 'Charlie'],
```

```
'Age': [25, 30, 35],
'City': ['New York', 'Los Angeles', 'Chicago']
}
df = pd.DataFrame(data)
df['City'] = df['City'].replace('New York', 'NYC')
print(df)
```

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

Drop the Age column from the DataFrame.

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 = df.drop(columns=['Age'])
print(df)
```

```
      Name         City
0    Alice     New York
1      Bob  Los Angeles
2  Charlie      Chicago
```

Sort the DataFrame by Salary in descending order.

In [23]:
```
import pandas as pd
import numpy as np
data ={
    'Name' : ['kaushi','alice','bob'],
    'city':['New York','los angeles','china'],
    'Salary':[50000,60000,70000]
}
df = pd.DataFrame(data)
df_sorted =df.sort_values(by='Salary',ascending=False)
print(df_sorted)
```

```
     Name         city  Salary
2     bob        china   70000
1   alice  los angeles   60000
0  kaushi     New York   50000
```

From a DataFrame df, print only the rows where Age is greater than 28.

In [24]:
```
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
```

Check for missing values in the DataFrame.

In [26]:
```
import pandas as pd
import numpy as np
data ={
    'Name':['bob','kaushi','abc',None],
    'Age':[25,np.nan,35,19],
    'City': [None,'New York', 'Los Angeles', 'Chicago']
}
df=pd.DataFrame(data)
print(df.isnull())
print(df.isnull().sum())
```

```
    Name    Age   City
0  False  False   True
1  False   True  False
2  False  False  False
3   True  False  False
Name    1
Age     1
City    1
dtype: int64
```

In [ ]:
```

```