|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Experiment No. 11 |  |  |  |  |
| Program to perform using Numpy | Exploratory  And Pandas | Data | Analysis |  |
| Date of Performance:18/3/2024 |  |  |  |  |
| Date of Submission:25/3/2024 |  |  |  |  |

**Experiment No. 11**

**Title:** Program to demonstrate data frame creation and Manipulation using Pandas

**Aim:** To study and implement data frame creation and Manipulation using Pandas

**Objective:** To introduce Pandas package for python **Theory:**

**Pandas** is an open-source library that is built on top of NumPy library. It is a Python package that offers various data structures and operations for manipulating numerical data and time series. It is mainly popular for importing and analyzing data much easier. Pandas is fast and it has high-performance & productivity for users.

**Code:**

import pandas as pd data = {'Name': ['ABC', 'ADC', 'PQR', 'XYZ'],

'Age': [25, 30, 35, 40],

'Salary': [50000, 60000, 70000, 80000]}

df = pd.DataFrame(data) print("Original DataFrame:") print(df)

df['Department'] = ['HR', 'Finance', 'IT', 'Marketing'] print("\nDataFrame after adding a new column:")

print(df)

print("\nAccessing specific columns:") print(df[['Name', 'Salary']])

print("\nFiltering rows based on a condition:") print(df[df['Age'] > 30])

df.loc[0, 'Salary'] = 55000 print("\nDataFrame after updating values:") print(df)

del df['Age'] print("\nDataFrame after deleting a column:") print(df)

df\_sorted = df.sort\_values(by='Salary', ascending=False) print("\nDataFrame sorted by Salary:") print(df\_sorted)

**Output:**

Original DataFrame:

Name Age Salary

1. ABC 25 50000
2. ADC 30 60000
3. PQR 35 70000
4. XYZ 40 80000

DataFrame after adding a new column:

Name Age Salary Department

1. ABC 25 50000 HR
2. ADC 30 60000 Finance
3. PQR 35 70000 IT
4. XYZ 40 80000 Marketing

Accessing specific columns:

Name Salary

1. ABC 50000
2. ADC 60000
3. PQR 70000
4. XYZ 80000

Filtering rows based on a condition:

Name Age Salary Department

1. PQR 35 70000 IT
2. XYZ 40 80000 Marketing DataFrame after updating values:

Name Age Salary Department

1. ABC 25 55000 HR
2. ADC 30 60000 Finance
3. PQR 35 70000 IT
4. XYZ 40 80000 Marketing

DataFrame after deleting a column:

Name Salary Department

1. ABC 55000 HR
2. ADC 60000 Finance
3. PQR 70000 IT
4. XYZ 80000 Marketing

DataFrame sorted by Salary:

Name Salary Department

3 XYZ 80000 Marketing

2 PQR 70000 IT

1 ADC 60000 Finance

0 ABC 55000 HR

**Conclusion:** In this Python program, data frames were created and manipulated using the Pandas library. Initially, a data frame was created from a dictionary, showcasing various operations such as adding a new column, accessing specific columns, filtering rows based on conditions, updating values, deleting a column, and sorting the data frame by a column. Pandas proved to be an efficient tool for data manipulation, providing intuitive methods for handling tabular data effectively.