Experiment No : 04

Date: 06.08.2025

EDA - Data Inspection and Analysis using Pandas

Aim: To inspect and analyze data using Pandas through DataFrame viewing, filtering, and calculating descriptive statistics.

Code:

```
# Import necessary libraries

import pandas as pd import

numpy as np

from scipy import stats  # For mode


# Sample DataFrame data

= {

   'Name': ['Alice', 'Bob', 'Charlie', 'David', 'Eve'],

   'Age': [24, 27, 22, 32, 29],

   'Score': [88, 92, 85, 70, 95]

}


df = pd.DataFrame(data)


# --------------------------------

# 1. Viewing and Inspecting DataFrame
```

```
# --------------------------------
print("Full DataFrame:\n", df)
print("\nDataFrame Info:")
print(df.info()) print("\nFirst 3
Rows:") print(df.head(3))
print("\nColumn Names:")
print(df.columns)

# --------------------------------
# 2. Filtering and Subsetting Data
# --------------------------------
# Filter rows where Score > 85 high_scores =
df[df['Score'] > 85] print("\nStudents with Score >
85:\n", high_scores)

# Filter rows where Age is between 25 and 30 age_range =
df[(df['Age'] >= 25) & (df['Age'] <= 30)] print("\nStudents
aged between 25 and 30:\n", age_range)

# --------------------------------
# 3. Descriptive Statistics # ------
--------------------------
print("\nDescriptive Statistics:")
print(df.describe())
```

```
# Central Tendency
mean_score = df['Score'].mean() median_score =
df['Score'].median() mode_score =
stats.mode(df['Score'], keepdims=False)

# Measures of Dispersion
range_score = df['Score'].max() - df['Score'].min()
variance_score = df['Score'].var() std_dev_score =
df['Score'].std()

print(f"\nMean Score: {mean_score}") print(f"Median
Score: {median_score}") print(f"Mode Score:
{mode_score}") print(f"Range of Scores:
{range_score}") print(f"Variance of Scores:
{variance_score}") print(f"Standard Deviation of
Scores: {std_dev_score}")
```

Output:

```
Full DataFrame:
      Name  Age  Score
0    Alice   24     88
1      Bob   27     92
2  Charlie   22     85
3    David   32     70
4      Eve   29     95

DataFrame Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5 entries, 0 to 4
Data columns (total 3 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Name    5 non-null      object
 1   Age     5 non-null      int64
 2   Score   5 non-null      int64
dtypes: int64(2), object(1)
memory usage: 248.0+ bytes
None

First 3 Rows:
      Name  Age  Score
0    Alice   24     88
1      Bob   27     92
2  Charlie   22     85

Column Names:
Index(['Name', 'Age', 'Score'], dtype='object')

Students with Score > 85:
     Name  Age  Score
0  Alice   24     88
1    Bob   27     92
4    Eve   29     95

Students aged between 25 and 30:
   Name  Age  Score
1  Bob   27     92
4  Eve   29     95



Descriptive Statistics:
             Age      Score
count   5.000000   5.000000
mean   26.800000  86.000000
std     3.962323   9.721111
min    22.000000  70.000000
25%    24.000000  85.000000
50%    27.000000  88.000000
75%    29.000000  92.000000
max    32.000000  95.000000

Mean Score: 86.0
Median Score: 88.0
Mode Score: ModeResult(mode=70, count=1)
Range of Scores: 25
Variance of Scores: 94.5
Standard Deviation of Scores: 9.72111104761179
```

Result: Successfully inspected, filtered, and analyzed the dataset using Pandas and computed key descriptive statistics.