4.EDA-Data Inspection and Analysis

AIM:

- Viewing and inspecting DataFrames
- Filtering and subsetting data using conditions
- Descriptive statistics: measures of central tendency (mean, median, mode) and measures of dispersion (range, variance, standard deviation)

PROGRAM:

```
import pandas as pd

import seaborn as sns


# Load iris dataset from seaborn

df = sns.load_dataset('iris')


# Step 1: View the first few rows

print("First 5 rows:")

print(df.head())


# Step 2: Inspect DataFrame info

print("\nDataFrame info:")

print(df.info())


# Step 3: Summary statistics

print("\nSummary statistics:")

print(df.describe())


# Step 4: Filtering examples
```

```
# Filter flowers with sepal_length > 5.5

filtered_sepal_length = df[df['sepal_length'] > 5.5]

print("\nFlowers with sepal_length > 5.5:")

print(filtered_sepal_length)


# Subset flowers of species 'setosa'

setosa_flowers = df[df['species'] == 'setosa']

print("\nSetosa species flowers:")

print(setosa_flowers)


# Step 5: Descriptive statistics on sepal_length

print("\nDescriptive statistics for sepal_length:")

print("Mean:", df['sepal_length'].mean())SN

print("Median:", df['sepal_length'].median())

print("Mode:", df['sepal_length'].mode()[0])

print("Range:", df['sepal_length'].max() - df['sepal_length'].min())

print("Variance:", df['sepal_length'].var())

print("Standard Deviation:", df['sepal_length'].std())
```

OUTPUT:

```
First 5 rows:
   sepal_length  sepal_width  petal_length  petal_width species
0           5.1          3.5           1.4          0.2  setosa
1           4.9          3.0           1.4          0.2  setosa
2           4.7          3.2           1.3          0.2  setosa
3           4.6          3.1           1.5          0.2  setosa
4           5.0          3.6           1.4          0.2  setosa

DataFrame info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   sepal_length  150 non-null    float64
 1   sepal_width   150 non-null    float64
 2   petal_length  150 non-null    float64
 3   petal_width   150 non-null    float64
 4   species       150 non-null    object
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
None

Summary statistics:
       sepal_length  sepal_width  petal_length  petal_width
count    150.000000   150.000000    150.000000   150.000000
mean       5.843333     3.057333      3.758000     1.199333
std        0.828066     0.435866      1.765298     0.762238
min        4.300000     2.000000      1.000000     0.100000
25%        5.100000     2.800000      1.600000     0.300000
50%        5.800000     3.000000      4.350000     1.300000
75%        6.400000     3.300000      5.100000     1.800000
max        7.900000     4.400000      6.900000     2.500000

Flowers with sepal_length > 5.5:
     sepal_length  sepal_width  petal_length  petal_width     species
14            5.8          4.0           1.2          0.2      setosa
15            5.7          4.4           1.5          0.4      setosa
18            5.7          3.8           1.7          0.3      setosa
50            7.0          3.2           4.7          1.4  versicolor
51            6.4          3.2           4.5          1.5  versicolor
..            ...          ...           ...          ...         ...
145           6.7          3.0           5.2          2.3   virginica
146           6.3          2.5           5.0          1.9   virginica
147           6.5          3.0           5.2          2.0   virginica
148           6.2          3.4           5.4          2.3   virginica
149           5.9          3.0           5.1          1.8   virginica

[91 rows x 5 columns]
```

```
Setosa species flowers:
    sepal_length  sepal_width  petal_length  petal_width species
0            5.1          3.5           1.4          0.2  setosa
1            4.9          3.0           1.4          0.2  setosa
2            4.7          3.2           1.3          0.2  setosa
3            4.6          3.1           1.5          0.2  setosa
4            5.0          3.6           1.4          0.2  setosa
5            5.4          3.9           1.7          0.4  setosa
6            4.6          3.4           1.4          0.3  setosa
7            5.0          3.4           1.5          0.2  setosa
8            4.4          2.9           1.4          0.2  setosa
9            4.9          3.1           1.5          0.1  setosa
10           5.4          3.7           1.5          0.2  setosa
11           4.8          3.4           1.6          0.2  setosa
12           4.8          3.0           1.4          0.1  setosa
13           4.3          3.0           1.1          0.1  setosa
14           5.8          4.0           1.2          0.2  setosa
15           5.7          4.4           1.5          0.4  setosa
16           5.4          3.9           1.3          0.4  setosa
17           5.1          3.5           1.4          0.3  setosa
18           5.7          3.8           1.7          0.3  setosa
19           5.1          3.8           1.5          0.3  setosa
20           5.4          3.4           1.7          0.2  setosa
21           5.1          3.7           1.5          0.4  setosa
22           4.6          3.6           1.0          0.2  setosa
23           5.1          3.3           1.7          0.5  setosa
24           4.8          3.4           1.9          0.2  setosa
25           5.0          3.0           1.6          0.2  setosa
26           5.0          3.4           1.6          0.4  setosa
27           5.2          3.5           1.5          0.2  setosa
28           5.2          3.4           1.4          0.2  setosa
29           4.7          3.2           1.6          0.2  setosa
30           4.8          3.1           1.6          0.2  setosa
```

```
Descriptive statistics for sepal_length:
Mean: 5.843333333333334
Median: 5.8
Mode: 5.0
Range: 3.6000000000000005
Variance: 0.6856935123042505
Standard Deviation: 0.8280661279778629
```

RESULT:

Thus, the given program was written and executed successfully.