

Practical 12

```
import pandas as pd
```

```
# Load the CSV file into a DataFrame
```

```
df = pd.read_csv("placement.csv")
```

```
# 1. df.head()
```

```
print("First 5 rows of the DataFrame:")
```

```
print(df.head())
```

```
# 2. df.tail()
```

```
print("\nLast 5 rows of the DataFrame:")
```

```
print(df.tail())
```

```
# 3. df.info()
```

```
print("\nDataFrame info:")
```

```
print(df.info())
```

```
# 4. df.shape
```

```
print("\nShape of the DataFrame (rows, columns):", df.shape)
```

```
# 5. df.size
```

```
print("\nSize of the DataFrame (number of elements):", df.size)
```

```
# 6. df.ndim
```

```
print("\nDimension of the DataFrame:", df.ndim)
```

```
# 7. df.describe()
```

```
print("\nStatistical summary for numerical columns:")
```

```
print(df.describe())
```

```
# 8. df.sample()
```

```
print("\nRandom sample from the DataFrame:")
```

```
print(df.sample())
```

```
# 9. df.isnull().sum()
```

```
print("\nNumber of missing values in each column:")
```

```
print(df.isnull().sum())
```

```
# 10. df.nunique()
```

```
print("\nNumber of unique elements in each column:")
```

```
print(df.nunique())
```

```
# 11. df.dropna()
```

```
# For demonstration, assuming we want to drop rows with missing values
```

```
df_dropna = df.dropna()
```

```
print("\nDataFrame after dropping rows with missing values:")
```

```
print(df_dropna)
```

```
# 12. df.duplicated()
```

```
print("\nDuplicate rows in the DataFrame:")
```

```
print(df.duplicated())
```

```
# 13. value_counts()
```

```
# For demonstration, assuming we want to count unique values in a specific column
```

```
# Replace "Column_Name" with the actual column name you want to perform value_counts() on
```

```
value_counts_column = df['placement'].value_counts()
```

```
print("\nValue counts for the 'placement' column:")
```

```
print(value_counts_column)
```

Output :

First 5 rows of the DataFrame:

	Unnamed: 0	pointer	iq	placement
0	0	NaN	123.0	1.0
1	1	5.9	106.0	NaN
2	2	5.3	NaN	0.0
3	3	7.4	132.0	1.0
4	4	5.8	142.0	0.0

Last 5 rows of the DataFrame:

	Unnamed: 0	pointer	iq	placement
95	95	4.3	200.0	0.0
96	96	4.4	42.0	0.0
97	97	6.7	182.0	NaN
98	98	6.3	NaN	1.0
99	99	NaN	113.0	1.0

DataFrame info:

```
<class 'pandas.core.frame.DataFrame'>
```

RangeIndex: 100 entries, 0 to 99

Data columns (total 4 columns):

#	Column	Non-Null Count	Dtype
0	Unnamed: 0	100 non-null	int64
1	pointer	98 non-null	float64
2	iq	98 non-null	float64
3	placement	98 non-null	float64

dtypes: float64(3), int64(1)

memory usage: 3.3 KB

None

Shape of the DataFrame (rows, columns): (100, 4)

Size of the DataFrame (number of elements): 400

Dimension of the DataFrame: 2

Statistical summary for numerical columns:

	Unnamed: 0	pointer	iq	placement
count	100.000000	98.000000	98.000000	98.000000
mean	49.500000	5.980612	123.816327	0.500000
std	29.011492	1.152197	40.298205	0.502571
min	0.000000	3.300000	37.000000	0.000000
25%	24.750000	5.025000	100.500000	0.000000
50%	49.500000	6.000000	128.000000	0.500000
75%	74.250000	6.900000	149.000000	1.000000
max	99.000000	8.500000	233.000000	1.000000

Random sample from the DataFrame:

	Unnamed: 0	pointer	iq	placement
47	47	5.2	161.0	0.0

Number of missing values in each column:

```
Unnamed: 0    0
pointer       2
iq            2
placement     2
dtype: int64
```

Number of unique elements in each column:

```
Unnamed: 0    100
pointer      38
iq          71
placement     2
dtype: int64
```

DataFrame after dropping rows with missing values:

	Unnamed: 0	pointer	iq	placement
3	3	7.4	132.0	1.0
4	4	5.8	142.0	0.0
5	5	7.1	48.0	1.0
6	6	5.7	143.0	0.0
7	7	5.0	63.0	0.0
..
92	92	5.2	110.0	0.0
93	93	6.8	112.0	1.0
94	94	4.7	52.0	0.0
95	95	4.3	200.0	0.0
96	96	4.4	42.0	0.0

[94 rows x 4 columns]

Duplicate rows in the DataFrame:

```
0    False
1    False
2    False
3    False
4    False
...
95   False
96   False
97   False
98   False
99   False
```

Length: 100, dtype: bool

Value counts for the 'placement' column:

```
placement
1.0    49
0.0    49
```

Name: count, dtype: int64