

```
1 C:\Users\kavan\PyCharmMiscProject\venv\Scripts\python.exe "C:/Users/kavan\  
PyCharmMiscProject\AIML INTERNSHIP\task1.py"  
2 Dataset Loaded Successfully!  
3  
4 First 5 rows of the dataset:  
5 PassengerId Survived Pclass Sex Age SibSp Parch Fare Embarked  
6 0 1 0 male 22.0 1 0 7.25 S  
7 1 2 1 female 38.0 1 0 71.28 S  
8 2 3 1 female 26.0 0 0 7.92 S  
9 3 4 1 female 35.0 1 0 53.10 S  
10 4 5 0 male 35.0 0 0 8.05 S  
11  
12 Last 5 rows of the dataset:  
13 PassengerId Survived Pclass Sex Age SibSp Parch Fare Embarked  
14 15 16 1 2 female 40.0 0 1 27.72 S  
15 16 17 0 3 male 21.0 0 0 8.66 S  
16 17 18 1 1 female 30.0 1 0 151.55 C  
17 18 19 0 3 male 18.0 0 0 7.80 S  
18 19 20 1 2 female 32.0 0 0 13.00 S  
19  
20 Dataset Information:  
21 <class 'pandas.core.frame.DataFrame'>  
22 RangeIndex: 20 entries, 0 to 19  
23 Data columns (total 9 columns):  
24 # Column Non-Null Count Dtype  
25 --- --- --- --- --- --- --- --- ---  
26 0 PassengerId 20 non-null int64
```

File - task1

```
27   1   Survived      20 non-null    int64
28   2   Pclass        20 non-null    int64
29   3   Sex           20 non-null    object
30   4   Age          18 non-null    float64
31   5   SibSp         20 non-null    int64
32   6   Parch         20 non-null    int64
33   7   Fare          20 non-null    float64
34   8   Embarked      20 non-null    object
35   dtypes: float64(2), int64(5), object(2)
36   memory usage: 1.5+ KB
37   None
```

38

39 Statistical Summary:

	PassengerId	Survived	Pclass	SibSp	Parch	Fare
40	20.00000	20.00000	20.00000	20.00000	20.00000	20.00000
41 count	20.00000	20.00000	20.00000	20.00000	20.00000	20.00000
42 mean	10.50000	0.50000	2.15000	0.50000	0.30000	27.494500
43 std	5.91608	0.512989	0.875094	0.760886	0.656947	34.394082
44 min	1.00000	0.00000	1.00000	0.00000	0.00000	7.220000
45 25%	5.75000	0.00000	1.00000	0.00000	0.00000	8.050000
46 50%	10.50000	0.50000	2.00000	0.00000	0.00000	13.000000
47 75%	15.25000	1.00000	3.00000	1.00000	0.00000	28.307500
48 max	20.00000	1.00000	3.00000	3.00000	2.00000	151.550000

49 [8 rows x 7 columns]

50

51 Feature Classification:

52 Numerical Features: ['PassengerId', 'Age', 'SibSp', 'Parch', 'Fare']

```
54 Categorical Features: ['Sex', 'Embarked']
55 Ordinal Features: ['Pclass']
56 Binary Features: ['Survived']
57
58 Unique Values in Categorical Columns:
59
60 Sex:
61 Sex
62 male    10
63 female   10
64 Name: count, dtype: int64
65
66 Embarked:
67 Embarked
68 S      14
69 C      4
70 Q      2
71 Name: count, dtype: int64
72
73 Target Variable: Survived
74 Input Features: ['PassengerId', 'Pclass', 'Sex', 'Age', 'SibSp', 'Parch', 'Fare', 'Embarked']
75
76 Dataset Size: 20 rows and 9 columns
77 Dataset is suitable for basic machine learning tasks.
78
79 Missing Values in Dataset:
```

```
80 PassengerId 0
81 Survived 0
82 Pclass 0
83 Sex 0
84 Age 2
85 SibSp 0
86 Parch 0
87 Fare 0
88 Embarked 0
89 dtype: int64
90
91 Data Quality Observations:
92 - Age column contains missing values.
93 - Dataset has both numerical and categorical data.
94 - Target variable is binary (Survived).
95 - Categorical features need encoding before ML modeling.
96 - Dataset may have class imbalance.
97
98 Task 1 Completed Successfully!
99
100 Process finished with exit code 0
101
```