

```

import pandas as pd

dataset = pd.read_csv('tested.csv')

dataset.head()

{"summary":{"\n  \"name\": \"dataset\",\n  \"rows\": 418,\n  \"fields\": [\n    {\n      \"column\": \"PassengerId\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 120,\n        \"min\": 892,\n        \"max\": 1309,\n        \"num_unique_values\": 418,\n        \"samples\": [\n          1213,\n          1216,\n          1280\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"Survived\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 0,\n        \"min\": 0,\n        \"max\": 1,\n        \"num_unique_values\": 2,\n        \"samples\": [\n          1,\n          0\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"Pclass\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 0,\n        \"min\": 1,\n        \"max\": 3,\n        \"num_unique_values\": 3,\n        \"samples\": [\n          3,\n          2\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"Name\",\n      \"properties\": {\n        \"dtype\": \"string\",\n        \"num_unique_values\": 418,\n        \"samples\": [\n          \"Krekorian, Mr. Neshan\",\n          \"Kreuchen, Miss. Emilie\"\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"Sex\",\n      \"properties\": {\n        \"dtype\": \"category\",\n        \"num_unique_values\": 2,\n        \"samples\": [\n          \"female\",\n          \"male\"\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"Age\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 14.181209235624422,\n        \"min\": 0.17,\n        \"max\": 76.0,\n        \"num_unique_values\": 79,\n        \"samples\": [\n          10.0,\n          34.5\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"SibSp\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 0,\n        \"min\": 0,\n        \"max\": 8,\n        \"num_unique_values\": 7,\n        \"samples\": [\n          0,\n          1\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"Parch\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 0,\n        \"min\": 0,\n        \"max\": 9,\n        \"num_unique_values\": 8,\n        \"samples\": [\n          1,\n          6\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"Ticket\",\n      \"properties\": {\n        \"dtype\": \"string\",\n        \"num_unique_values\": 363,\n        \"samples\": [\n
```

```

\"2673\", \n          \"W./C. 6607\" \n          ], \n
\"semantic_type\": \"\", \n          \"description\": \"\" \n          } \n
n      }, \n      { \n          \"column\": \"Fare\", \n          \"properties\": { \n
\"dtype\": \"number\", \n          \"std\": 55.907576179973844, \n
\"min\": 0.0, \n          \"max\": 512.3292, \n
\"num_unique_values\": 169, \n          \"samples\": [ \n
41.5792, \n          57.75 \n          ], \n          \"semantic_type\":
\"\", \n          \"description\": \"\" \n          } \n      }, \n      { \n
\"column\": \"Cabin\", \n          \"properties\": { \n          \"dtype\":
\"category\", \n          \"num_unique_values\": 76, \n
\"samples\": [ \n          \"A21\", \n          \"E45\" \n          ], \n
\"semantic_type\": \"\", \n          \"description\": \"\" \n          } \n
n      }, \n      { \n          \"column\": \"Embarked\", \n          \"properties\":
{ \n          \"dtype\": \"category\", \n          \"num_unique_values\":
3, \n          \"samples\": [ \n          \"Q\", \n          \"S\" \n          ], \n
\"semantic_type\": \"\", \n          \"description\": \"\" \n          } \n
      } \n      ] \n      }, \"type\": \"dataframe\", \"variable_name\": \"dataset\"}

```

```
dataset.tail()
```

```

{ \"summary\": { \n      \"name\": \"dataset\", \n      \"rows\": 5, \n      \"fields\":
[ \n          { \n          \"column\": \"PassengerId\", \n          \"properties\": { \n
\"dtype\": \"number\", \n          \"std\": 1, \n          \"min\":
1305, \n          \"max\": 1309, \n          \"num_unique_values\": 5, \n
\"samples\": [ \n          1306, \n          1309, \n          1307 \n          ], \n
          \"semantic_type\": \"\", \n          \"description\": \"\" \n          } \n
      }, \n      { \n          \"column\": \"Survived\", \n          \"properties\": { \n
          \"dtype\": \"number\", \n          \"std\":
0, \n          \"min\": 0, \n          \"max\": 1, \n
          \"num_unique_values\": 2, \n          \"samples\": [ \n          1, \n
0 \n          ], \n          \"semantic_type\": \"\", \n          \"description\":
\"\" \n          } \n      }, \n      { \n          \"column\":
\"Pclass\", \n          \"properties\": { \n          \"dtype\": \"number\", \n
          \"std\": 0, \n          \"min\": 1, \n          \"max\": 3, \n
          \"num_unique_values\": 2, \n          \"samples\": [ \n          1, \n
3 \n          ], \n          \"semantic_type\": \"\", \n          \"description\":
\"\" \n          } \n      }, \n      { \n          \"column\":
\"Name\", \n          \"properties\": { \n          \"dtype\": \"string\", \n
          \"num_unique_values\": 5, \n          \"samples\": [ \n          \"Oliva y
Ocana, Dona. Fermina\", \n          \"Peter, Master. Michael J\" \n          ], \n
          \"semantic_type\": \"\", \n          \"description\": \"\" \n          } \n
      }, \n      { \n          \"column\": \"Sex\", \n          \"properties\": { \n
          \"dtype\": \"category\", \n          \"num_unique_values\": 2, \n
          \"samples\": [ \n          \"female\", \n          \"male\" \n          ], \n
          \"semantic_type\": \"\", \n          \"description\": \"\" \n          } \n
      }, \n      { \n          \"column\": \"Age\", \n          \"properties\": { \n
          \"dtype\": \"number\", \n          \"std\": 0.3535533905932738, \n
          \"min\": 38.5, \n          \"max\": 39.0, \n
          \"num_unique_values\": 2, \n          \"samples\": [ \n          38.5, \n
39.0 \n          ], \n          \"semantic_type\": \"\", \n          \"description\":
\"\" \n          } \n      } \n      ] \n      } \n      } \n      } \n      } \n

```



```
9   Fare          417 non-null   float64
10  Cabin          91 non-null    object
11  Embarked       418 non-null    object
dtypes: float64(2), int64(5), object(5)
memory usage: 39.3+ KB
```

```
dataset.columns
```

```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age',
      'SibSp',
      'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

```
dataset.isnull().sum()
```

```
PassengerId    0
Survived        0
Pclass          0
Name            0
Sex             0
Age            86
SibSp           0
Parch           0
Ticket          0
Fare            1
Cabin          327
Embarked        0
dtype: int64
```

```
dataset.fillna(dataset['Age'].mean(), inplace=True)
```

```
dataset.fillna(dataset['Fare'].mean(), inplace=True)
```

```
dataset.drop(columns=['Cabin'], inplace=True)
```

```
dataset.isnull().sum()
```

```
PassengerId    0
Survived        0
Pclass          0
Name            0
Sex             0
Age            0
SibSp           0
Parch           0
Ticket          0
Fare            0
Embarked        0
dtype: int64
```

```
dataset["Embarked"].value_counts()
```

```
Embarked
S    270
C    102
Q     46
Name: count, dtype: int64
```

```
dataset.describe()
```

```
{ "summary": "{\n  \"name\": \"dataset\",\n  \"rows\": 8,\n  \"fields\": [\n    {\n      \"column\": \"PassengerId\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 412.1232851470217,\n        \"min\": 120.81045760473994,\n        \"max\": 1309.0,\n        \"num_unique_values\": 7,\n        \"samples\": [\n          418.0,\n          1100.5,\n          1204.75\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"Survived\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 147.6421942886486,\n        \"min\": 0.0,\n        \"max\": 418.0,\n        \"num_unique_values\": 5,\n        \"samples\": [\n          0.36363636363636365,\n          1.0,\n          0.4816221409322309\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"Pclass\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 147.0758997861715,\n        \"min\": 0.8418375519640503,\n        \"max\": 418.0,\n        \"num_unique_values\": 5,\n        \"samples\": [\n          2.2655502392344498,\n          3.0,\n          0.8418375519640503\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"Age\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 139.0269633070591,\n        \"min\": 0.17,\n        \"max\": 418.0,\n        \"num_unique_values\": 7,\n        \"samples\": [\n          418.0,\n          30.272590361445783,\n          35.75\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"SibSp\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 147.28745840271156,\n        \"min\": 0.0,\n        \"max\": 418.0,\n        \"num_unique_values\": 6,\n        \"samples\": [\n          418.0,\n          0.4473684210526316,\n          8.0\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"Parch\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 147.29373273558997,\n        \"min\": 0.0,\n        \"max\": 418.0,\n        \"num_unique_values\": 5,\n        \"samples\": [\n          0.3923444976076555,\n          9.0,\n          0.9814288785371691\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"Fare\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 206.410611408246,\n        \"min\": 0.0,\n        \"max\": 512.3292,\n        \"num_unique_values\": 8,\n        \"samples\": [\n          35.614378445840785,\n          14.4542,\n          418.0\n        ],\n
```

```

{"semantic_type": "\\n", "description": "\\n"}\n
n ]\n}", "type": "dataframe"}

print(dataset['Survived'].value_counts())

# Value counts for 'Pclass'
print(dataset['Pclass'].value_counts())

# Value counts for 'Sex'
print(dataset['Sex'].value_counts())

# Value counts for 'Embarked'
print(dataset['Embarked'].value_counts())

Survived
0      266
1      152
Name: count, dtype: int64
Pclass
3      218
1      107
2       93
Name: count, dtype: int64
Sex
male      266
female    152
Name: count, dtype: int64
Embarked
S      270
C      102
Q       46
Name: count, dtype: int64

dataset.to_csv('cleaned_tested.csv', index=False)

```