

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
# load the dataset
db = pd.read_csv("/content/drive/MyDrive/Colab Notebooks/Titanic train_and_test2.csv")
db.head()
db.info()
db.describe()
db.isnull().sum()
```

```

↗ <class 'pandas.core.frame.DataFrame'>
RangeIndex: 1309 entries, 0 to 1308
Data columns (total 28 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Passengerid           1309 non-null   int64
1   Age                   1309 non-null   float64
2   Fare                  1309 non-null   float64
3   Sex                   1309 non-null   int64
4   sibsp                 1309 non-null   int64
5   zero                  1309 non-null   int64
6   zero.1                1309 non-null   int64
7   zero.2                1309 non-null   int64
8   zero.3                1309 non-null   int64
9   zero.4                1309 non-null   int64
10  zero.5                1309 non-null   int64
11  zero.6                1309 non-null   int64
12  Parch                 1309 non-null   int64
13  zero.7                1309 non-null   int64
14  zero.8                1309 non-null   int64
15  zero.9                1309 non-null   int64
16  zero.10               1309 non-null   int64
17  zero.11               1309 non-null   int64
18  zero.12               1309 non-null   int64
19  zero.13               1309 non-null   int64
20  zero.14               1309 non-null   int64
21  Pclass                1309 non-null   int64
22  zero.15               1309 non-null   int64
23  zero.16               1309 non-null   int64
24  Embarked              1307 non-null   float64
25  zero.17               1309 non-null   int64
26  zero.18               1309 non-null   int64
27  Survived              1309 non-null   int64
dtypes: float64(3), int64(25)
memory usage: 286.5 KB

```

```

      0
Passengerid  0
      Age    0
      Fare    0
      Sex     0
      sibsp   0
      zero    0
      zero.1  0
      zero.2  0
      zero.3  0
      zero.4  0
      zero.5  0
      zero.6  0
      Parch   0
      zero.7  0
      zero.8  0
      zero.9  0
      zero.10 0
      zero.11 0

```

```

zero.12    0
zero.13    0
zero.14    0
Pclass     0
zero.15    0
zero.16    0
Embarked   2
zero.17    0
zero.18    0
2urvived   0

```

dtype: int64

db.info()

```

↩ <class 'pandas.core.frame.DataFrame'>
RangeIndex: 1309 entries, 0 to 1308
Data columns (total 28 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Passengerid     1309 non-null   int64
1   Age             1309 non-null   float64
2   Fare            1309 non-null   float64
3   Sex             1309 non-null   int64
4   sibsp           1309 non-null   int64
5   zero            1309 non-null   int64
6   zero.1          1309 non-null   int64
7   zero.2          1309 non-null   int64
8   zero.3          1309 non-null   int64
9   zero.4          1309 non-null   int64
10  zero.5          1309 non-null   int64
11  zero.6          1309 non-null   int64
12  Parch           1309 non-null   int64
13  zero.7          1309 non-null   int64
14  zero.8          1309 non-null   int64
15  zero.9          1309 non-null   int64
16  zero.10         1309 non-null   int64
17  zero.11         1309 non-null   int64
18  zero.12         1309 non-null   int64
19  zero.13         1309 non-null   int64
20  zero.14         1309 non-null   int64
21  Pclass          1309 non-null   int64
22  zero.15         1309 non-null   int64
23  zero.16         1309 non-null   int64
24  Embarked        1307 non-null   float64
25  zero.17         1309 non-null   int64
26  zero.18         1309 non-null   int64
27  2urvived        1309 non-null   int64
dtypes: float64(3), int64(25)
memory usage: 286.5 KB

```

db.isnull().sum()



	0
Passengerid	0
Age	0
Fare	0
Sex	0
sibsp	0
zero	0
zero.1	0
zero.2	0
zero.3	0
zero.4	0
zero.5	0
zero.6	0
Parch	0
zero.7	0
zero.8	0
zero.9	0
zero.10	0
zero.11	0
zero.12	0
zero.13	0
zero.14	0
Pclass	0
zero.15	0
zero.16	0
Embarked	0
zero.17	0
zero.18	0
2urvived	0

dtype: int64

```
db.describe()
```