```
import pandas as pd
import numpy as np
import matplotlib as plt
import seaborn as sns
```

```
df=pd.read_csv('titanic_test.csv')
```

```
df.head()
```

	passenger_id	pclass	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
0	295	1	Thayer, Mr. John Borland Jr	male	17.0	0	2	17421	110.8833	C70	C	B	NaN	Haverford, PA
1	1150	3	Risien, Mr. Samuel Beard	male	NaN	0	0	364498	14.5000	NaN	S	NaN	NaN	NaN
2	89	1	Davidson, Mr. Thornton	male	31.0	1	0	F.C. 12750	52.0000	B71	S	NaN	NaN	Montreal, PQ
			Nirva, Mr.											

+ Code + Text

```
df.describe()
```

	passenger_id	pclass	age	sibsp	parch	fare	body
count	459.000000	459.000000	370.000000	459.000000	459.000000	459.000000	48.000000
mean	637.673203	2.248366	30.541216	0.455338	0.389978	31.968854	153.187500
std	372.769720	0.835467	14.133361	0.896176	0.840038	47.977896	96.104034
min	0.000000	1.000000	0.333300	0.000000	0.000000	0.000000	1.000000
25%	317.000000	1.000000	21.250000	0.000000	0.000000	7.925000	69.250000
50%	612.000000	3.000000	29.000000	0.000000	0.000000	14.500000	142.500000
75%	960.000000	3.000000	40.000000	1.000000	0.000000	31.387500	250.500000
max	1308.000000	3.000000	71.000000	8.000000	6.000000	512.329200	322.000000

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 459 entries, 0 to 458
Data columns (total 14 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   passenger_id  459 non-null    int64
 1   pclass        459 non-null    int64
 2   name          459 non-null    object
 3   sex           459 non-null    object
 4   age           370 non-null    float64
 5   sibsp         459 non-null    int64
 6   parch         459 non-null    int64
 7   ticket        459 non-null    object
 8   fare          459 non-null    float64
 9   cabin         104 non-null    object
 10  embarked      458 non-null    object
 11  boat          178 non-null    object
 12  body          48 non-null     float64
 13  home.dest     281 non-null    object
dtypes: float64(3), int64(4), object(7)
memory usage: 50.3+ KB
```

```
df.isnull()
```

	passenger_id	pclass	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
0	False	False	False	False	False	False	False	False	False	False	False	False	True	False
1	False	False	False	False	True	False	False	False	False	True	False	True	True	True
2	False	False	False	False	False	False	False	False	False	False	False	True	True	False
3	False	False	False	False	False	False	False	False	False	True	False	True	True	False
4	False	False	False	False	False	False	False	False	False	True	False	True	True	True
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
454	False	False	False	False	True	False	False	False	False	True	False	True	True	True
455	False	False	False	False	False	False	False	False	False	True	False	True	True	False
456	False	False	False	False	True	False	False	False	False	True	False	False	True	True
457	False	False	False	False	False	False	False	False	False	True	False	True	True	False
458	False	False	False	False	False	False	False	False	False	True	False	True	True	True

459 rows × 14 columns

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

```
passenger_id      0
pclass            0
name              0
sex               0
age              89
sibsp             0
parch             0
ticket            0
fare              0
cabin           355
embarked          1
boat            281
body            411
home.dest       178
dtype: int64
```

```
df["cabin"].fillna('Unknown',inplace=True)
```

```
df.dropna(subset=["embarked"],inplace=True)
```

```
df["age"].fillna(df["age"].mean(),inplace=True)
```

```
df["boat"].fillna('Unknown',inplace=True)
```

```
df["home.dest"].fillna('Unknown values',inplace=True)
```

```
df["body"].fillna('Unknown values',inplace=True)
```

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

```
passenger_id    0
pclass          0
name            0
sex             0
age             0
sibsp           0
parch           0
ticket          0
fare            0
cabin           0
embarked        0
boat            0
body            0
home.dest       0
dtype: int64
```

```
df.duplicated().sum()
```

```
0
```