

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

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

```
df
```

	PassengerId	Survived	Pclass	\
0	1	0	3	
1	2	1	1	
2	3	1	3	
3	4	1	1	
4	5	0	3	
..	...	...	...	
886	887	0	2	
887	888	1	1	
888	889	0	3	
889	890	1	1	
890	891	0	3	

	Name	Sex	Age
SibSp \			
0	Braund, Mr. Owen Harris	male	22.0
1			
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0
1			
2	Heikkinen, Miss. Laina	female	26.0
0			
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0
1			
4	Allen, Mr. William Henry	male	35.0
0			
..	...	...	...
...			
886	Montvila, Rev. Juozas	male	27.0
0			
887	Graham, Miss. Margaret Edith	female	19.0
0			
888	Johnston, Miss. Catherine Helen "Carrie"	female	NaN
1			
889	Behr, Mr. Karl Howell	male	26.0
0			
890	Dooley, Mr. Patrick	male	32.0
0			

	Parch	Ticket	Fare	Cabin	Embarked
0	0	A/5 21171	7.2500	NaN	S
1	0	PC 17599	71.2833	C85	C

2	0	STON/02.	3101282	7.9250	NaN	S
3	0		113803	53.1000	C123	S
4	0		373450	8.0500	NaN	S
...	...		...	...	...	...
886	0		211536	13.0000	NaN	S
887	0		112053	30.0000	B42	S
888	2	W./C.	6607	23.4500	NaN	S
889	0		111369	30.0000	C148	C
890	0		370376	7.7500	NaN	Q

[891 rows x 12 columns]

df.head()

	PassengerId	Survived	Pclass	\
0	1	0	3	
1	2	1	1	
2	3	1	3	
3	4	1	1	
4	5	0	3	

		Name	Sex	Age
SibSp	\			
0		Braund, Mr. Owen Harris	male	22.0
1				
1		Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0
1				
2		Heikkinen, Miss. Laina	female	26.0
0				
3		Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0
1				
4		Allen, Mr. William Henry	male	35.0
0				

	Parch		Ticket	Fare	Cabin	Embarked
0	0		A/5 21171	7.2500	NaN	S
1	0		PC 17599	71.2833	C85	C
2	0	STON/02.	3101282	7.9250	NaN	S
3	0		113803	53.1000	C123	S
4	0		373450	8.0500	NaN	S

df.tail()

	PassengerId	Survived	Pclass	
Name	\			
886	887	0	2	Montvila, Rev. Juozas
887	888	1	1	Graham, Miss. Margaret Edith
888	889	0	3	Johnston, Miss. Catherine Helen

"Carrie"					
889	890	1	1		Behr, Mr. Karl
Howell					
890	891	0	3		Dooley, Mr.
Patrick					

	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
886	male	27.0	0	0	211536	13.00	NaN	S
887	female	19.0	0	0	112053	30.00	B42	S
888	female	NaN	1	2	W./C. 6607	23.45	NaN	S
889	male	26.0	0	0	111369	30.00	C148	C
890	male	32.0	0	0	370376	7.75	NaN	Q

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

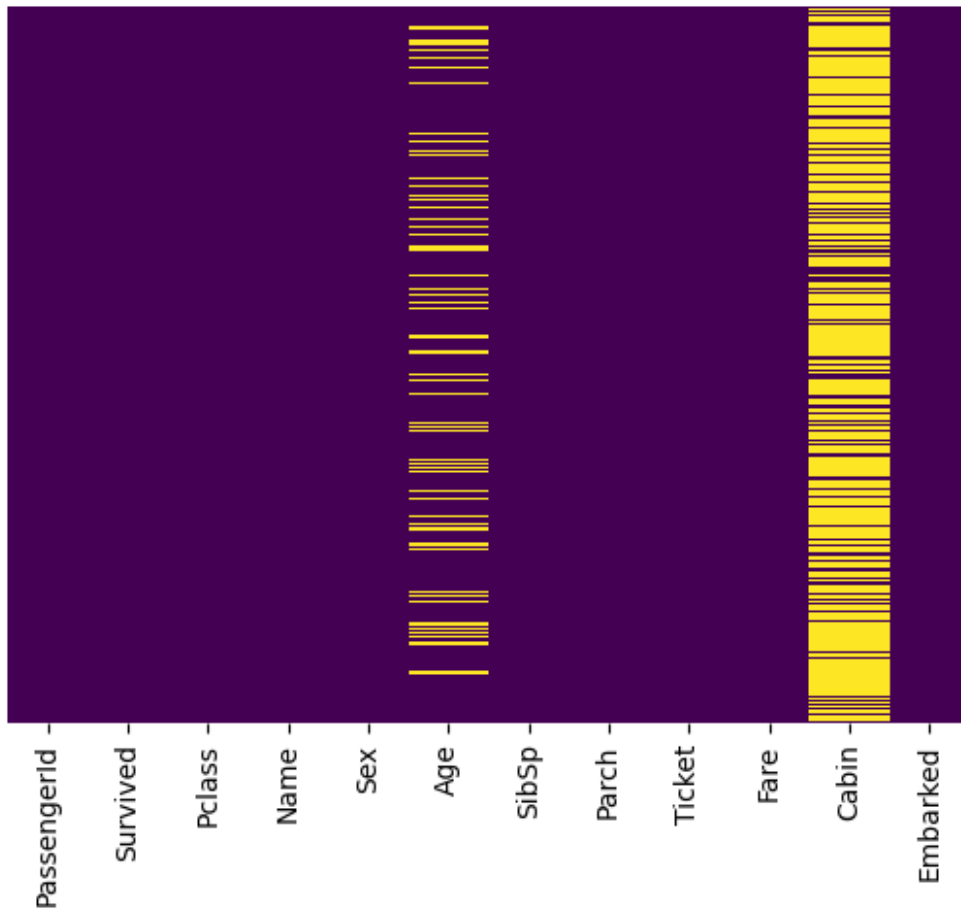
```

PassengerId      0
Survived          0
Pclass           0
Name             0
Sex              0
Age             177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin           687
Embarked         2
dtype: int64

```

```
sns.heatmap(df.isnull(), yticklabels=False, cbar=False,
cmap='viridis')
```

```
<Axes: >
```



```
df['Age'].isnull().sum()
```

```
177
```

```
df['Cabin'].isnull().sum()
```

```
687
```

```
df.describe()
```

	PassengerId	Survived	Pclass	Age	SibSp	\
count	891.000000	891.000000	891.000000	714.000000	891.000000	
mean	446.000000	0.383838	2.308642	29.699118	0.523008	
std	257.353842	0.486592	0.836071	14.526497	1.102743	
min	1.000000	0.000000	1.000000	0.420000	0.000000	
25%	223.500000	0.000000	2.000000	20.125000	0.000000	
50%	446.000000	0.000000	3.000000	28.000000	0.000000	
75%	668.500000	1.000000	3.000000	38.000000	1.000000	
max	891.000000	1.000000	3.000000	80.000000	8.000000	

	Parch	Fare
count	891.000000	891.000000

```

mean      0.381594    32.204208
std       0.806057    49.693429
min       0.000000    0.000000
25%      0.000000    7.910400
50%      0.000000    14.454200
75%      0.000000    31.000000
max       6.000000   512.329200

```

```
df.describe()
```

	PassengerId	Survived	Pclass	Age	SibSp \
count	891.000000	891.000000	891.000000	714.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008
std	257.353842	0.486592	0.836071	14.526497	1.102743
min	1.000000	0.000000	1.000000	0.420000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000
50%	446.000000	0.000000	3.000000	28.000000	0.000000
75%	668.500000	1.000000	3.000000	38.000000	1.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000

	Parch	Fare
count	891.000000	891.000000
mean	0.381594	32.204208
std	0.806057	49.693429
min	0.000000	0.000000
25%	0.000000	7.910400
50%	0.000000	14.454200
75%	0.000000	31.000000
max	6.000000	512.329200

```
df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   PassengerId     891 non-null   int64
1   Survived        891 non-null   int64
2   Pclass          891 non-null   int64
3   Name            891 non-null   object
4   Sex             891 non-null   object
5   Age             714 non-null   float64
6   SibSp           891 non-null   int64
7   Parch           891 non-null   int64
8   Ticket          891 non-null   object
9   Fare            891 non-null   float64
10  Cabin           204 non-null   object
11  Embarked        889 non-null   object

```

```
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

```
df.dtypes
```

```
PassengerId    int64
Survived        int64
Pclass          int64
Name            object
Sex             object
Age            float64
SibSp           int64
Parch           int64
Ticket          object
Fare            float64
Cabin           object
Embarked        object
dtype: object
```

```
df.shape
```

```
(891, 12)
```

```
df.drop('Cabin',axis=1 ,inplace=True)
```

```
df.dropna(inplace=True)
```

```
df.head()
```

	PassengerId	Survived	Pclass	\
0	1	0	3	
1	2	1	1	
2	3	1	3	
3	4	1	1	
4	5	0	3	

		Name	Sex	Age
SibSp	\			
0		Braund, Mr. Owen Harris	male	22.0
1				
1		Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0
1				
2		Heikkinen, Miss. Laina	female	26.0
0				
3		Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0
1				
4		Allen, Mr. William Henry	male	35.0
0				

	Parch	Ticket	Fare	Embarked
0	0	A/5 21171	7.2500	S

1	0	PC	17599	71.2833	C
2	0	STON/02.	3101282	7.9250	S
3	0		113803	53.1000	S
4	0		373450	8.0500	S

```
df.tail()
```

	PassengerId	Survived	Pclass	
Name \				
885	886	0	3	Rice, Mrs. William (Margaret Norton)
886	887	0	2	Montvila, Rev. Juozas
887	888	1	1	Graham, Miss. Margaret Edith
889	890	1	1	Behr, Mr. Karl Howell
890	891	0	3	Dooley, Mr. Patrick

	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
885	female	39.0	0	5	382652	29.125	Q
886	male	27.0	0	0	211536	13.000	S
887	female	19.0	0	0	112053	30.000	S
889	male	26.0	0	0	111369	30.000	C
890	male	32.0	0	0	370376	7.750	Q

```
df.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

```

```
df['Age']=df['Age'].round(0).astype('int')
```

```
df.dtypes
```

```

PassengerId    int64
Survived        int64
Pclass          int64
Name            object
Sex             object

```

```
Age          int32
SibSp        int64
Parch        int64
Ticket       object
Fare         float64
Embarked     object
dtype: object
```

```
df.head()
```

	PassengerId	Survived	Pclass	\
0	1	0	3	
1	2	1	1	
2	3	1	3	
3	4	1	1	
4	5	0	3	

		Name	Sex	Age
SibSp	\			
0		Braund, Mr. Owen Harris	male	22
1				
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38	
1				
2	Heikkinen, Miss. Laina	female	26	
0				
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	
1				
4	Allen, Mr. William Henry	male	35	
0				

	Parch	Ticket	Fare	Embarked
0	0	A/5 21171	7.2500	S
1	0	PC 17599	71.2833	C
2	0	STON/O2. 3101282	7.9250	S
3	0	113803	53.1000	S
4	0	373450	8.0500	S

```
cat=pd.get_dummies(df, columns=['Sex'])
```

```
cat.head()
```

	PassengerId	Survived	Pclass	\
0	1	0	3	
1	2	1	1	
2	3	1	3	
3	4	1	1	
4	5	0	3	

		Name	Age	SibSp
Parch	\			
0		Braund, Mr. Owen Harris	22	1



```

0
1 Cumings, Mrs. John Bradley (Florence Briggs Th... 38 1
0
2 Heikkinen, Miss. Laina 26 0
0
3 Futrelle, Mrs. Jacques Heath (Lily May Peel) 35 1
0
4 Allen, Mr. William Henry 35 0
0

```

```

      Ticket      Fare Embarked Sex_female Sex_male
0      A/5 21171    7.2500      S      False      True
1      PC 17599   71.2833      C      True      False
2 STON/O2. 3101282    7.9250      S      True      False
3      113803   53.1000      S      True      False
4      373450    8.0500      S      False      True

```

```
cat['Sex_female']
```

```

0      False
1       True
2       True
3       True
4      False

```

```

...
885      True
886     False
887      True
889     False
890     False

```

```
Name: Sex_female, Length: 712, dtype: bool
```

```
cat['Sex_male']
```

```

0       True
1      False
2      False
3      False
4       True

```

```

...
885     False
886      True
887     False
889      True
890      True

```

```
Name: Sex_male, Length: 712, dtype: bool
```