

```
import pandas as pd
```

```
titanic = pd.read_csv("train.csv")
```

```
titanic.shape
```

(891, 20)

```
test = pd.read_csv("test.csv")
```

```
test.shape
```

(418, 11)

```
titanic.head
```

<bound method NDFrame.head of

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

	Parch	Ticket	Fare	Cabin	Embarked_C	Embarked_Q	\
0	0	A/5 21171	7.2500	NaN	0	0	
1	0	PC 17599	71.2833	C85	1	0	
2	0	STON/O2. 3101282	7.9250	NaN	0	0	
3	0	113803	53.1000	C123	0	0	
4	0	373450	8.0500	NaN	0	0	
..	
886	0	211536	13.0000	NaN	0	0	
887	0	112053	30.0000	B42	0	0	
888	2	W./C. 6607	23.4500	NaN	0	0	
889	0	111369	30.0000	C148	1	0	
890	0	370376	7.7500	NaN	0	1	

	Embarked_S
0	1
1	0
2	1
3	1
4	1
..	...
886	1
887	1
888	1
889	0
890	0

[891 rows x 14 columns]>

```
titanic.tail
```

<bound method NDFrame.tail of

	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_C      Embarked_Q      \
0      0      A/5 21171      7.2500      NaN      0      0
1      0      PC 17599      71.2833      C85      1      0
2      0      STON/O2. 3101282      7.9250      NaN      0      0
3      0      113803      53.1000      C123      0      0
4      0      373450      8.0500      NaN      0      0
..      ...      ...      ...      ...      ...
886      0      211536      13.0000      NaN      0      0
887      0      112053      30.0000      B42      0      0
888      2      W./C. 6607      23.4500      NaN      0      0
889      0      111369      30.0000      C148      1      0
890      0      370376      7.7500      NaN      0      1

      Embarked_S      Embarked_C_port      Embarked_Q_port      Embarked_S_port      \
0      1      0      0      1
1      0      1      0      0
2      1      0      0      1
3      1      0      0      1
4      1      0      0      1
..      ...      ...      ...      ...
886      1      0      0      1
887      1      0      0      1
888      1      0      0      1
889      0      1      0      0
890      0      0      1      0

      Embarked_C_port      Embarked_Q_port      Embarked_S_port
0      0      0      1
1      1      0      0
2      0      0      1
3      0      0      1
4      0      0      1

titanic.info

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_C      Embarked_Q      \
0      0      A/5 21171      7.2500      NaN      0      0
1      0      PC 17599      71.2833      C85      1      0
2      0      STON/O2. 3101282      7.9250      NaN      0      0
3      0      113803      53.1000      C123      0      0
4      0      373450      8.0500      NaN      0      0
..      ...      ...      ...      ...
886      0      211536      13.0000      NaN      0      0
887      0      112053      30.0000      B42      0      0
888      2      W./C. 6607      23.4500      NaN      0      0
889      0      111369      30.0000      C148      1      0
890      0      370376      7.7500      NaN      0      1



      Embarked_S      Embarked_C_port      Embarked_Q_port      Embarked_S_port      \
```

```
888      1      0      0      1
889      0      1      0      0
890      0      0      1      0

      Embarked_C_port Embarked_Q_port Embarked_S_port
0              0              0              1
1              1              0              0
2              0              0              1
3              0              0              1
4              0              0              1
..          ...          ...          ...
886              0              0              1
887              0              0              1
888              0              0              1
889              1              0              0
890              0              1              0

[891 rows x 20 columns]>

ports = pd.get_dummies(titanic.Embarked, prefix = 'Embarked')
ports.head()
```

	Embarked_C	Embarked_Q	Embarked_S	
0	0	0	1	
1	1	0	0	
2	0	0	1	
3	0	0	1	
4	0	0	1	

```
titanic = titanic.join(ports, rsuffix='_port')
titanic.drop(['Embarked'], axis=1, inplace=True)

titanic.Sex=titanic.Sex.map({"male":0,"female":1})

y= titanic.Survived.copy()
x = titanic.drop(["Survived"],axis=1)

x.drop(['Cabin', 'Ticket', 'Name', "PassengerId"],axis = 1,inplace = True)

x.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 9 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Pclass      891 non-null    int64
1   Sex         891 non-null    int64
2   Age        714 non-null    float64
3   SibSp       891 non-null    int64
4   Parch       891 non-null    int64
5   Fare        891 non-null    float64
6   Embarked_C  891 non-null    uint8
7   Embarked_Q  891 non-null    uint8
8   Embarked_S  891 non-null    uint8
dtypes: float64(2), int64(4), uint8(3)
memory usage: 44.5 KB

x.isnull().values.any()

True

x[pd.isnull(x).any(axis=1)]
```

	Pclass	Sex	Age	SibSp	Parch	Fare	Embarked_C	Embarked_Q	Embarked_S	
5	3	0	NaN	0	0	8.4583	0	1	0	
17	2	0	NaN	0	0	13.0000	0	0	1	
19	3	1	NaN	0	0	7.2250	1	0	0	
26	3	0	NaN	0	0	7.2250	1	0	0	
28	3	1	NaN	0	0	7.8792	0	1	0	
...	

```
x.Age.fillna(x.Age.mean(),inplace=True)
x.isnull().values.any()

False
888      3      1  NaN      1      2 23.4500      0      0      1

from sklearn.model_selection import train_test_split
x_train,x_valid,y_train,y_valid = train_test_split(x,y,test_size=0.2,random_state = 7)
```

```
from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
```

```
model.fit(x_train,y_train)
```

/usr/local/lib/python3.10/dist-packages/sklearn/linear_model/_logistic.py:458: ConvergenceWarning: lbfgs failed to
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
<https://scikit-learn.org/stable/modules/preprocessing.html>
Please also refer to the documentation for alternative solver options:
https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
n_iter_i = _check_optimize_result(

```
LogisticRegression()
LogisticRegression()
```

```
model.score(x_train,y_train)
```

0.8103932584269663

```
model.score(x_valid,y_valid)
```

0.7541899441340782