


```
from google.colab import files
uploaded = files.upload()
```

 Choose Files

No file chosen


Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

Saving gender_submission.csv to gender_submission.csv

Saving test.csv to test.csv


Saving train.csv to train.csv

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv('train.csv')
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 = df.dropna()
df.info()
```



<class 'pandas.core.frame.DataFrame'>
Index: 712 entries, 0 to 890
Data columns (total 9 columns):
Column Non-Null Count Dtype
--- -
0 PassengerId 712 non-null int64
1 Survived 712 non-null int64
2 Pclass 712 non-null int64
3 Sex 712 non-null object
4 Age 712 non-null float64
5 SibSp 712 non-null int64
6 Parch 712 non-null int64
7 Fare 712 non-null float64
8 Embarked 712 non-null object
dtypes: float64(2), int64(5), object(2)
memory usage: 55.6+ KB

```
import pandas as pd
dummies = []
cols = ['Pclass', 'Sex', 'Embarked']
for col in cols:
    dummies.append(pd.get_dummies(df[col], prefix=col))
df = pd.concat([df] + dummies, axis=1)
df = df.drop(cols, axis=1)
print(df)
```



	PassengerId	Survived	Age	SibSp	Parch	Fare	Pclass_1	Pclass_2	\
0	1	0	22.0	1	0	7.2500	False	False	
1	2	1	38.0	1	0	71.2833	True	False	
2	3	1	26.0	0	0	7.9250	False	False	
3	4	1	35.0	1	0	53.1000	True	False	
4	5	0	35.0	0	0	8.0500	False	False	
..	
885	886	0	39.0	0	5	29.1250	False	False	
886	887	0	27.0	0	0	13.0000	False	True	
887	888	1	19.0	0	0	30.0000	True	False	
889	890	1	26.0	0	0	30.0000	True	False	
890	891	0	32.0	0	0	7.7500	False	False	
	Pclass_3	Sex_female	Sex_male	Embarked_C	Embarked_Q	Embarked_S			
0	True	False	True	False	False	True			
1	False	True	False	True	False	False			
2	True	True	False	False	False	True			
3	False	True	False	False	False	True			
4	True	False	True	False	False	True			

```

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

```

[712 rows x 14 columns]

```

df['Age'] = df['Age'].interpolate()
df.info()

```

```

<class 'pandas.core.frame.DataFrame'>
Index: 712 entries, 0 to 890
Data columns (total 14 columns):
#   Column          Non-Null Count  Dtype
---  -
0   PassengerId      712 non-null    int64
1   Survived         712 non-null    int64
2   Age              712 non-null    float64
3   SibSp            712 non-null    int64
4   Parch            712 non-null    int64
5   Fare             712 non-null    float64
6   Pclass_1         712 non-null    bool
7   Pclass_2         712 non-null    bool
8   Pclass_3         712 non-null    bool
9   Sex_female       712 non-null    bool
10  Sex_male         712 non-null    bool
11  Embarked_C       712 non-null    bool
12  Embarked_Q       712 non-null    bool
13  Embarked_S       712 non-null    bool
dtypes: bool(8), float64(2), int64(4)
memory usage: 44.5 KB

```

```

#Converting the dataframe to numpy
X = df.values
y = df['Survived'].values
X = np.delete(X, 1, axis=1)
print(X)

```

```

[[1 3 'Braund, Mr. Owen Harris' ... 7.25 nan 'S']
 [2 1 'Cumings, Mrs. John Bradley (Florence Briggs Thayer)' ... 71.2833
  'C85' 'C']
 [3 3 'Heikkinen, Miss. Laina' ... 7.925 nan 'S']
 ...
 [889 3 'Johnston, Miss. Catherine Helen "Carrie"' ... 23.45 nan 'S']
 [890 1 'Behr, Mr. Karl Howell' ... 30.0 'C148' 'C']
 [891 3 'Dooley, Mr. Patrick' ... 7.75 nan 'Q']]

```

```

#Dividing data set into training set and test set
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0)
#print(X_train, X_test, y_train, y_test)

```

print(X_train)

```

[[858 1 'Daly, Mr. Peter Denis ' ... 26.55 'E17' 'S']
 [53 1 'Harper, Mrs. Henry Sleeper (Myna Haxtun)' ... 76.7292 'D33' 'C']
 [387 3 'Goodwin, Master. Sidney Leonard' ... 46.9 nan 'S']
 ...
 [630 3 'O'Connell, Mr. Patrick D' ... 7.7333 nan 'Q']
 [560 3 'de Messemaeker, Mrs. Guillaume Joseph (Emma)' ... 17.4 nan 'S']
 [685 2 'Brown, Mr. Thomas William Solomon' ... 39.0 nan 'S']]

```

print(X_test)

```

[[496 3 'Yousseff, Mr. Gerious' ... 14.4583 nan 'C']
 [649 3 'Willey, Mr. Edward' ... 7.55 nan 'S']
 [279 3 'Rice, Master. Eric' ... 29.125 nan 'Q']
 ...
 [621 3 'Yasbeck, Mr. Antoni' ... 14.4542 nan 'C']
 [787 3 'Sjoblom, Miss. Anna Sofia' ... 7.4958 nan 'S']
 [65 1 'Stewart, Mr. Albert A' ... 27.7208 nan 'C']]

```

print(y_train)

```

[[1 1 0 0 0 1 0 0 0 1 1 0 0 1 0 1 0 0 0 0 0 0 1 0 1 1 1 0 0 0 1 0 1 0 0 1 1
  1 0 0 1 0 1 0 0 0 0 1 0 1 0 1 0 1 1 1 0 0 0 0 0 0 1 0 1 0 0 0 1 0 1 0 0 0
  1 0 1 1 1 0 0 0 1 1 0 0 1 0 1 0 0 0 0 1 1 1 0 1 0 1 0 1 0 0 1 0 0 0 0 0 1
  0 1 0 1 0 0 0 0 1 0 0 1 0 0 0 1 0 0 1 0 0 1 0 1 0 0 0 0 1 0 0 0 1 0 0 0

```

```
0 0 0 0 0 0 0 1 1 0 0 0 0 1 0 0 0 0 0 1 1 0 1 0 1 0 0 0 0 1 1 0 0 0 0 0 0 0 1 0 0 1
1 0 0 0 1 0 1 0 0 1 0 0 1 1 0 0 0 1 1 1 0 1 0 0 1 0 0 0 0 0 0 1 0 0 0 1 0 0 1
1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 1 0 1 0 0 1 0 1 0 1 0 0 0 0 1 1 1 1 0 0 0
0 0 0 0 1 1 1 0 0 0 1 0 1 1 1 0 0 0 0 0 1 1 0 0 1 0 1 1 0 0 0 0 1 1 0 1 0 1
0 1 0 0 0 1 0 0 0 1 0 0 0 0 0 0 1 1 0 0 0 0 0 1 1 0 0 0 1 0 0 0 1 1 0 0 1 0 0
0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 1 1 0 0 1 1 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 1 1
0 0 0 1 0 1 0 1 0 1 0 1 1 1 0 0 1 0 1 0 0 1 1 1 0 0 0 1 0 0 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 1 1 0 0 0 1 1 0 0 1 1 1 0 0 0 1 0 0 1 0 1 0 1 0 0 1 0 0 0
0 0 1 0 0 0 0 1 1 1 0 0 1 1 1 0 1 1 0 0 0 0 1 0 0 1 1 0 0 0 1 1 0 1 0 1 1
1 0 1 1 1 0 0 0 1 0 0 0 1 0 0 1 0 0 0 1 0 0 1 1 1 0 1 1 0 1 0 1 1 0 0 0 0
0 0 1 1 0 0 0 0 0 1 0 0 0 1 1 1 0 0 1 0 0 0 0 0 1 1 1 0 1 0 0 1 0 0 0 0 0
0 0 0 0 1 1 1 1 1 1 0 1 0 1 0 0 1 0 0 1 0 1 0 0 0 1 1 0 0 0 1 1 0 1 1 0
1 0 0 1 1 1 1 0 0 0 1 0 0 0 1 1 0 1 1 1 0 0 1 1 1 1 1 0 1 0 1 0]
```

```
print(y_test)
```

```
→ [0 0 0 1 1 1 1 1 1 1 0 1 1 0 0 0 0 0 1 0 1 0 0 0 1 0 1 1 0 0 1 0 1 0 1 0
0 0 0 1 0 0 0 1 0 0 1 0 0 1 1 1 0 1 0 0 0 0 1 0 0 1 0 1 0 1 0 1 1 1 1 0 0
0 1 0 0 0 0 0 1 0 0 0 1 1 1 1 0 0 0 1 1 0 0 1 0 0 1 0 0 0 0 0 1 1 0 0 1 0
1 1 0 1 1 1 1 0 1 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 1
1 0 0 1 0 0 1 0 0 1 0 1 0 1 1 1 0 0 0 0 0 0 0 0 1 0 0 1 0 1 0 0 0 0 0 0 0
0 1 0 0 1 0 0 1 1 0 0 0 1 1 0 1 0 0 1 1 0 0 0 1 0 0 1 0 0 0 0 0 1 0 1 0 1
1 0 1 0 0 1 1 0 0 1 1 0 0 0 1 1 1 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 1 1 1
0 0 0 0 0 0 0 1 0]
```