In [4]:

```
import numpy as np
import pandas as pd

temp = pd.read_csv("C://Users/ARPIT/Desktop/titanic/titanic_train.csv")
temp
```

Out[4]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

In [5]:

```
temp.head()
```

Out[5]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

In [6]:

```
temp.describe()
```

Out[6]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

In [7]:

```
temp2 = temp.copy()
temp2.head()
```

Out[7]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

In [12]:

```
del temp2["Name"]
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
~\anaconda3\lib\site-packages\pandas\core\indexes\base.py in get_loc(self, key, method, tolerance)
   3360             try:
-> 3361                 return self._engine.get_loc(casted_key)
   3362             except KeyError as err:

~\anaconda3\lib\site-packages\pandas\_libs\index.pyx in pandas._libs.index.IndexEngine.get_loc()

~\anaconda3\lib\site-packages\pandas\_libs\index.pyx in pandas._libs.index.IndexEngine.get_loc()

pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.get_item()

pandas\_libs\hashtable_class_helper.pxi in pandas._libs.hashtable.PyObjectHashTable.get_item()

KeyError: 'Name'

The above exception was the direct cause of the following exception:

KeyError                                  Traceback (most recent call last)
~\AppData\Local\Temp/ipykernel_17372/43614661.py in <module>
----> 1 del temp2["Name"]

~\anaconda3\lib\site-packages\pandas\core\generic.py in __delitem__(self, key)
   3961             # there was no match, this call should raise the appropriate
   3962             # exception:
-> 3963             loc = self.axes[-1].get_loc(key)
   3964             self._mgr = self._mgr.idelete(loc)
   3965

~\anaconda3\lib\site-packages\pandas\core\indexes\base.py in get_loc(self, key, method, tolerance)
   3361                 return self._engine.get_loc(casted_key)
   3362             except KeyError as err:
-> 3363                 raise KeyError(key) from err
   3364
   3365         if is_scalar(key) and isna(key) and not self.hasnans:

KeyError: 'Name'
```

In [13]:

```
temp2.head()
```

Out[13]:

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	male	35.0	0	0	373450	8.0500	NaN	S

In [14]:

```
del temp2["Ticket"]

temp2.head()
```

Out[14]:

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Fare	Cabin	Embarked
0	1	0	3	male	22.0	1	0	7.2500	NaN	S
1	2	1	1	female	38.0	1	0	71.2833	C85	C
2	3	1	3	female	26.0	0	0	7.9250	NaN	S
3	4	1	1	female	35.0	1	0	53.1000	C123	S
4	5	0	3	male	35.0	0	0	8.0500	NaN	S

In [16]:

```
def a(n):
    if n == "male":
        return 1
    else:
        return 0


temp2["gender"] = temp2.Sex.apply(a)

temp2.head(9)
```

Out[16]:

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Fare	Cabin	Embarked	gender
0	1	0	3	male	22.0	1	0	7.2500	NaN	S	1
1	2	1	1	female	38.0	1	0	71.2833	C85	C	0
2	3	1	3	female	26.0	0	0	7.9250	NaN	S	0
3	4	1	1	female	35.0	1	0	53.1000	C123	S	0
4	5	0	3	male	35.0	0	0	8.0500	NaN	S	1
5	6	0	3	male	NaN	0	0	8.4583	NaN	Q	1
6	7	0	1	male	54.0	0	0	51.8625	E46	S	1
7	8	0	3	male	2.0	3	1	21.0750	NaN	S	1
8	9	1	3	female	27.0	0	2	11.1333	NaN	S	0

In [17]:

```
del temp2["Sex"]
temp2.head()
```

Out[17]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	Cabin	Embarked	gender
0	1	0	3	22.0	1	0	7.2500	NaN	S	1
1	2	1	1	38.0	1	0	71.2833	C85	C	0
2	3	1	3	26.0	0	0	7.9250	NaN	S	0
3	4	1	1	35.0	1	0	53.1000	C123	S	0
4	5	0	3	35.0	0	0	8.0500	NaN	S	1

In [18]:

```
temp2.isnull().sum()
```

Out[18]:

```
PassengerId      0
Survived         0
Pclass           0
Age            177
SibSp            0
Parch            0
Fare             0
Cabin          687
Embarked         2
gender           0
dtype: int64
```

In [19]:

```
temp2.Cabin.fillna(0,inplace = True)
temp2.head()
```

Out[19]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	Cabin	Embarked	gender
0	1	0	3	22.0	1	0	7.2500	0	S	1
1	2	1	1	38.0	1	0	71.2833	C85	C	0
2	3	1	3	26.0	0	0	7.9250	0	S	0
3	4	1	1	35.0	1	0	53.1000	C123	S	0
4	5	0	3	35.0	0	0	8.0500	0	S	1

In [23]:

```
a = temp2[temp2.Survived==1]
a
```

Out[23]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	Cabin	Embarked	gender
1	2	1	1	38.0	1	0	71.2833	C85	C	0
2	3	1	3	26.0	0	0	7.9250	0	S	0
3	4	1	1	35.0	1	0	53.1000	C123	S	0
8	9	1	3	27.0	0	2	11.1333	0	S	0
9	10	1	2	14.0	1	0	30.0708	0	C	0
...	...	...	...	...	...	...	...	...	...	...
875	876	1	3	15.0	0	0	7.2250	0	C	0
879	880	1	1	56.0	0	1	83.1583	C50	C	0
880	881	1	2	25.0	0	1	26.0000	0	S	0
887	888	1	1	19.0	0	0	30.0000	B42	S	0
889	890	1	1	26.0	0	0	30.0000	C148	C	1

342 rows × 10 columns

In [27]:

```
temp2.Age.fillna(a.Age.mean(),inplace = True)
temp2
```

Out[27]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	Cabin	Embarked	gender
0	1	0	3	22.00000	1	0	7.2500	0	S	1
1	2	1	1	38.00000	1	0	71.2833	C85	C	0
2	3	1	3	26.00000	0	0	7.9250	0	S	0
3	4	1	1	35.00000	1	0	53.1000	C123	S	0
4	5	0	3	35.00000	0	0	8.0500	0	S	1
...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	27.00000	0	0	13.0000	0	S	1
887	888	1	1	19.00000	0	0	30.0000	B42	S	0
888	889	0	3	28.34369	1	2	23.4500	0	S	0
889	890	1	1	26.00000	0	0	30.0000	C148	C	1
890	891	0	3	32.00000	0	0	7.7500	0	Q	1

891 rows × 10 columns

In [28]:

```
b = temp2[temp2.Survived==0]

temp2.Age.fillna(b.Age.mean(),inplace = True)

temp2
```

Out[28]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	Cabin	Embarked	gender
0	1	0	3	22.00000	1	0	7.2500	0	S	1
1	2	1	1	38.00000	1	0	71.2833	C85	C	0
2	3	1	3	26.00000	0	0	7.9250	0	S	0
3	4	1	1	35.00000	1	0	53.1000	C123	S	0
4	5	0	3	35.00000	0	0	8.0500	0	S	1
...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	27.00000	0	0	13.0000	0	S	1
887	888	1	1	19.00000	0	0	30.0000	B42	S	0
888	889	0	3	28.34369	1	2	23.4500	0	S	0
889	890	1	1	26.00000	0	0	30.0000	C148	C	1
890	891	0	3	32.00000	0	0	7.7500	0	Q	1

891 rows × 10 columns

In [29]:

```
temp2.isnull().sum()
```

Out[29]:

```
PassengerId    0
Survived       0
Pclass         0
Age            0
SibSp          0
Parch          0
Fare           0
Cabin          0
Embarked       2
gender         0
dtype: int64
```

temp2.Embarkment

In [31]:

```
temp2.Embarked
```

Out[31]:

```
0      S
1      C
2      S
3      S
4      S
      ..
886    S
887    S
888    S
889    C
890    Q
Name: Embarked, Length: 891, dtype: object
```

In [32]:

```
def e(n):
    if n == 'S':
        return 1
    else:
        return 0


temp2["southhampton"] = temp2.Embarked.apply(e)

temp2.head()
```

Out[32]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	Cabin	Embarked	gender	southhampton
0	1	0	3	22.0	1	0	7.2500	0	S	1	1
1	2	1	1	38.0	1	0	71.2833	C85	C	0	0
2	3	1	3	26.0	0	0	7.9250	0	S	0	1
3	4	1	1	35.0	1	0	53.1000	C123	S	0	1
4	5	0	3	35.0	0	0	8.0500	0	S	1	1

In [33]:

```
def c(n):
    if n == 'C':
        return 1
    else:
        return 0


temp2["cherbourgh"] = temp2.Embarked.apply(c)

temp2.head()
```

Out[33]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	Cabin	Embarked	gender	southhampton	cherbourgh
0	1	0	3	22.0	1	0	7.2500	0	S	1	1	0
1	2	1	1	38.0	1	0	71.2833	C85	C	0	0	1
2	3	1	3	26.0	0	0	7.9250	0	S	0	1	0
3	4	1	1	35.0	1	0	53.1000	C123	S	0	1	0
4	5	0	3	35.0	0	0	8.0500	0	S	1	1	0

In [34]:

```
def q(n):
    if n == 'Q':
        return 1
    else:
        return 0


temp2["queenstown"] = temp2.Embarked.apply(q)

temp2.head()
```

Out[34]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	Cabin	Embarked	gender	southhampton	cherbourgh	queenstown
0	1	0	3	22.0	1	0	7.2500	0	S	1	1	0	0
1	2	1	1	38.0	1	0	71.2833	C85	C	0	0	1	0
2	3	1	3	26.0	0	0	7.9250	0	S	0	1	0	0
3	4	1	1	35.0	1	0	53.1000	C123	S	0	1	0	0
4	5	0	3	35.0	0	0	8.0500	0	S	1	1	0	0

In [35]:

```
del temp2["Embarked"]
temp2.head()
```

Out[35]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	Cabin	gender	southhampton	cherbourgh	queenstown
0	1	0	3	22.0	1	0	7.2500	0	1	1	0	0
1	2	1	1	38.0	1	0	71.2833	C85	0	0	1	0
2	3	1	3	26.0	0	0	7.9250	0	0	1	0	0
3	4	1	1	35.0	1	0	53.1000	C123	0	1	0	0
4	5	0	3	35.0	0	0	8.0500	0	1	1	0	0

In [36]:

```
temp2.isnull().sum()
```

Out[36]:

```
PassengerId     0
Survived        0
Pclass          0
Age             0
SibSp           0
Parch           0
Fare            0
Cabin           0
gender          0
southhampton    0
cherbourgh      0
queenstown      0
dtype: int64
```

In [37]:

```
temp2.describe()
```

Out[37]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	gender	southhampton	cherbourgh	queenstown
count	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.429858	0.523008	0.381594	32.204208	0.647587	0.722783	0.188552	0.086420
std	257.353842	0.486592	0.836071	13.013270	1.102743	0.806057	49.693429	0.477990	0.447876	0.391372	0.281141
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	22.000000	0.000000	0.000000	7.910400	0.000000	0.000000	0.000000	0.000000
50%	446.000000	0.000000	3.000000	28.343690	0.000000	0.000000	14.454200	1.000000	1.000000	0.000000	0.000000
75%	668.500000	1.000000	3.000000	35.000000	1.000000	0.000000	31.000000	1.000000	1.000000	0.000000	0.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200	1.000000	1.000000	1.000000	1.000000

In [38]:

```
temp2.to_csv("C://Users/ARPIT/Desktop/titanic/titanic_clean.csv")
```

In [ ]: