In [1]:
```
pip install openpyxl
```

```
Requirement already satisfied: openpyxl in /srv/conda/envs/notebook/lib/python3.7/site-packages (3.0.8)
Requirement already satisfied: et-xmlfile in /srv/conda/envs/notebook/lib/python3.7/site-packages (from openpyxl) (1.1.0)
Note: you may need to restart the kernel to use updated packages.
```

In [2]:
```
import pandas as pd
import numpy as np
from sklearn.metrics import accuracy_score
```

In [3]:
```
data=pd.read_excel('train.xlsx')
```

In [4]:
```
df=pd.DataFrame(data)
print(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/O2. 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]
```

In [5]:
```
data.shape
```

Out[5]:
```
(891, 12)
```

In [6]:
```
data.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
```

In [7]:
```
data.describe()
```

Out[7]:

	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 [8]:
```
data.head()
```

Out[8]:

	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 [9]:
```
data.isnull().sum()
```

Out[9]:
```
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
```

In [10]:
```
from sklearn.utils import shuffle
```

In [11]:
```
data=shuffle(data,random_state=42)
```

In [12]:
```
data
```

Out[12]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
709	710	1	3	Moubarek, Master. Halim Gonios ("William George")	male	NaN	1	1	2661	15.2458	NaN	C
439	440	0	2	Kvillner, Mr. Johan Henrik Johannesson	male	31.0	0	0	C.A. 18723	10.5000	NaN	S
840	841	0	3	Alhomaki, Mr. Ilmari Rudolf	male	20.0	0	0	SOTON/O2 3101287	7.9250	NaN	S
720	721	1	2	Harper, Miss. Annie Jessie "Nina"	female	6.0	0	1	248727	33.0000	NaN	S
39	40	1	3	Nicola-Yarred, Miss. Jamila	female	14.0	1	0	2651	11.2417	NaN	C
...	...	...	...	...	...	...	...	...	...	...	...	...
106	107	1	3	Salkjelsvik, Miss. Anna Kristine	female	21.0	0	0	343120	7.6500	NaN	S
270	271	0	1	Cairns, Mr. Alexander	male	NaN	0	0	113798	31.0000	NaN	S
860	861	0	3	Hansen, Mr. Claus Peter	male	41.0	2	0	350026	14.1083	NaN	S
435	436	1	1	Carter, Miss. Lucile Polk	female	14.0	1	2	113760	120.0000	B96 B98	S
102	103	0	1	White, Mr. Richard Frasar	male	21.0	0	1	35281	77.2875	D26	S

891 rows × 12 columns

In [13]:
```
div=int(data.shape[0]/4)
```

In [14]:
```
div
```

Out[14]:
```
222
```

In [15]:
```
train=data.loc[:3*div+1,:]
test=data.loc[3*div+1:]
```

In [16]:
```
train
```

Out[16]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
709	710	1	3	Moubarek, Master. Halim Gonios ("William George")	male	NaN	1	1	2661	15.2458	NaN	C
439	440	0	2	Kvillner, Mr. Johan Henrik Johannesson	male	31.0	0	0	C.A. 18723	10.5000	NaN	S
840	841	0	3	Alhomaki, Mr. Ilmari Rudolf	male	20.0	0	0	SOTON/O2 3101287	7.9250	NaN	S
720	721	1	2	Harper, Miss. Annie Jessie "Nina"	female	6.0	0	1	248727	33.0000	NaN	S
39	40	1	3	Nicola-Yarred, Miss. Jamila	female	14.0	1	0	2651	11.2417	NaN	C
...	...	...	...	...	...	...	...	...	...	...	...	...
36	37	1	3	Mamee, Mr. Hanna	male	NaN	0	0	2677	7.2292	NaN	C
452	453	0	1	Foreman, Mr. Benjamin Laventall	male	30.0	0	0	113051	27.7500	C111	C
253	254	0	3	Lobb, Mr. William Arthur	male	30.0	1	0	A/5. 3336	16.1000	NaN	S
303	304	1	2	Keane, Miss. Nora A	female	NaN	0	0	226593	12.3500	E101	Q
667	668	0	3	Rommetvedt, Mr. Knud Paust	male	NaN	0	0	312993	7.7750	NaN	S

621 rows × 12 columns

In [17]:
```
test
```

Out[17]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
667	668	0	3	Rommetvedt, Mr. Knud Paust	male	NaN	0	0	312993	7.7750	NaN	S
571	572	1	1	Appleton, Mrs. Edward Dale (Charlotte Lamson)	female	53.0	2	0	11769	51.4792	C101	S
636	637	0	3	Leinonen, Mr. Antti Gustaf	male	32.0	0	0	STON/O 2. 3101292	7.9250	NaN	S
714	715	0	2	Greenberg, Mr. Samuel	male	52.0	0	0	250647	13.0000	NaN	S
262	263	0	1	Taussig, Mr. Emil	male	52.0	1	1	110413	79.6500	E67	S
...	...	...	...	...	...	...	...	...	...	...	...	...
106	107	1	3	Salkjelsvik, Miss. Anna Kristine	female	21.0	0	0	343120	7.6500	NaN	S
270	271	0	1	Cairns, Mr. Alexander	male	NaN	0	0	113798	31.0000	NaN	S
860	861	0	3	Hansen, Mr. Claus Peter	male	41.0	2	0	350026	14.1083	NaN	S
435	436	1	1	Carter, Miss. Lucile Polk	female	14.0	1	2	113760	120.0000	B96 B98	S
102	103	0	1	White, Mr. Richard Frasar	male	21.0	0	1	35281	77.2875	D26	S

271 rows × 12 columns

In [18]:
```
test['Simple_mode']=train['Survived'].mode()[0]
```

```
/srv/conda/envs/notebook/lib/python3.7/site-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  """Entry point for launching an IPython kernel.
```

In [19]:
```
test['Simple_mode']
```

Out[19]:
```
667    0
571    0
636    0
714    0
262    0
      ..
106    0
270    0
860    0
435    0
102    0
Name: Simple_mode, Length: 271, dtype: int64
```

In [20]:
```
simple_mode_accuracy=accuracy_score(test['Survived'],test['Simple_mode'])
simple_mode_accuracy
```

Out[20]:
```
0.6346863468634686
```

In [21]:
```
gender_mode=pd.crosstab(train['Survived'],train['Sex'])
gender_mode
```

Out[21]:

Survived	Sex	female	male
	0	57	321
	1	167	76

In [22]:
```
pd.pivot_table(df,index=['Age','Sex'],aggfunc='sum')
```

Out[22]:

		Fare	Parch	PassengerId	Pclass	SibSp	Survived
Age	**Sex**						
0.42	**male**	8.5167	1	804	3	0	1
0.67	**male**	14.5000	1	756	2	1	1
0.75	**female**	38.5166	2	1115	6	4	2
0.83	**male**	47.7500	3	911	4	1	2
0.92	**male**	151.5500	2	306	1	1	1
...	**...**	...	...	...	...	...	...
70.00	**male**	81.5000	1	1419	3	1	0
70.50	**male**	7.7500	0	117	3	0	0
71.00	**male**	84.1584	0	591	2	0	0
74.00	**male**	7.7750	0	852	3	0	0
80.00	**male**	30.0000	0	631	1	0	1

145 rows × 6 columns

In [23]:
```
pd.pivot_table(df,index='Sex').aggfunc='sum')
```

Out[23]:

	Age	Fare	Parch	PassengerId	Pclass	SibSp	Survived
Sex							
female	7286.00	13966.6628	204	135343	678	218	233
male	13919.17	14727.2865	136	262043	1379	248	109

In [29]:
```
test['gender_mode']=test['Survived']
for i in test['Sex'].unique():
    test['gender_mode'][(test['Sex']==str(i))==train['Survived'][train['Sex']==str(i)].mode()[0]
```

```
/srv/conda/envs/notebook/lib/python3.7/site-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  """Entry point for launching an IPython kernel.
```

In [30]:
```
gender_accuracy=accuracy_score(test['Survived'],test['gender_mode'])
gender_accuracy
```

Out[30]:
```
1.0
```

In [26]:
```
class_mode=pd.crosstab(train['Survived'],train['Pclass'])
class_mode
```

Out[26]:

Pclass	1	2	3
Survived			
0	50	66	262
1	97	65	81

In [33]:
```
test['class_mode']=test['Survived']
for i in test['Pclass'].unique():
    test['class_mode'][(test['Pclass']==str(i))==train['Survived'][train['Pclass']==str(i)].mode()
```

```
/srv/conda/envs/notebook/lib/python3.7/site-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  """Entry point for launching an IPython kernel.
```

In [34]:
```
class_accuracy=accuracy_score(test['Survived'],test['class_mode'])
class_accuracy
```

Out[34]:
```
1.0
```

In [ ]:
```

```