Creado por:

Isabel Maniega

In [1]:
```
import pandas as pd
```

Ejercicio 1

1) Lee con pandas el archivo train.csv correspondiente al titanic dataset

In [2]:
```
df = pd.read_csv("train.csv")
df.head()
```

Out[2]:

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

2) Hacer un bucle for para automatizar las gráficas de pd.crosstab

Se pide relacionar la columna Survived con Pclass, Sex y Embarked

Nota:

Se pide que dentro del bucle for se encuentre la gráfica requerida.

Entonces, en una sola celda, tenemos 3 gráficas mostradas y todo automatizado.

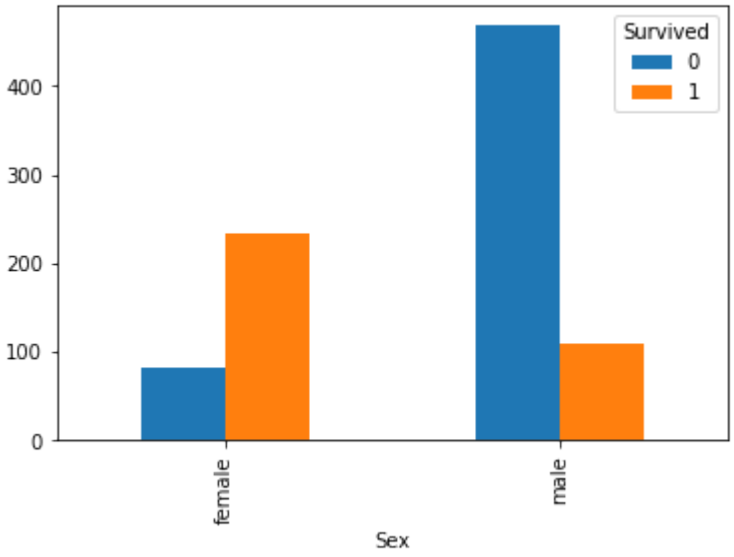
In [3]:
```
pd.crosstab(df.Sex, df.Survived)
```

Out[3]:

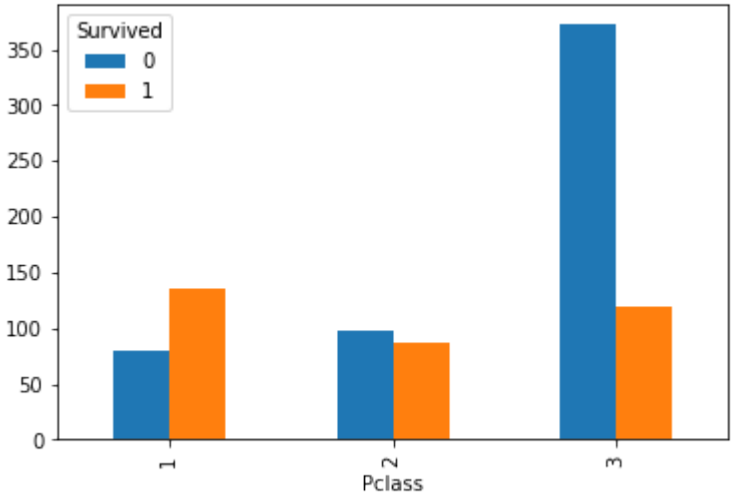
Survived	0	1
Sex		
female	81	233
male	468	109

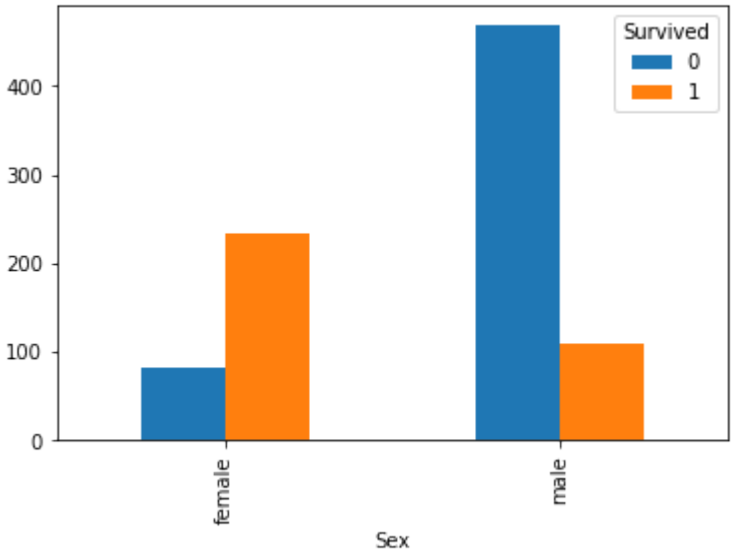
In [5]:
```
pd.crosstab(df.Sex, df.Survived).plot(kind='bar')
```

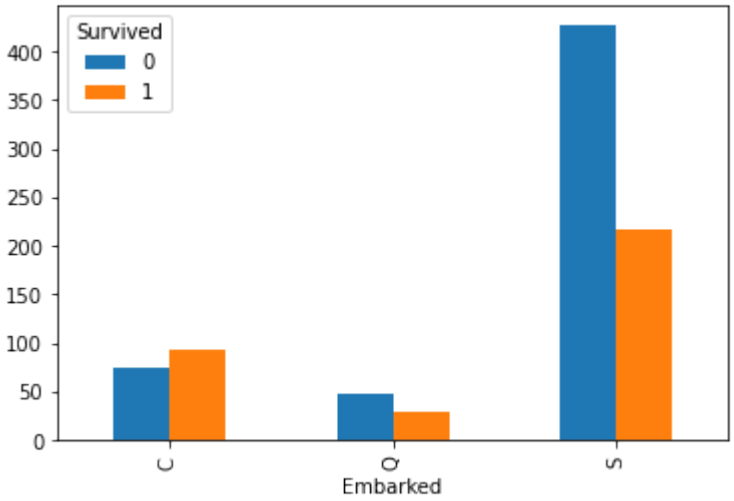
Out[5]: <AxesSubplot:xlabel='Sex'>



In [7]:
```
features = ["Pclass", "Sex", "Embarked"]
for feature in features:
    pd.crosstab(df[feature], df["Survived"]).plot(kind="bar")
```







3) Hacer una función para automatizar las gráficas de pd.crosstab

Se pide relacionar la columna Survived con Pclass, Sex y Embarked

NOTA:

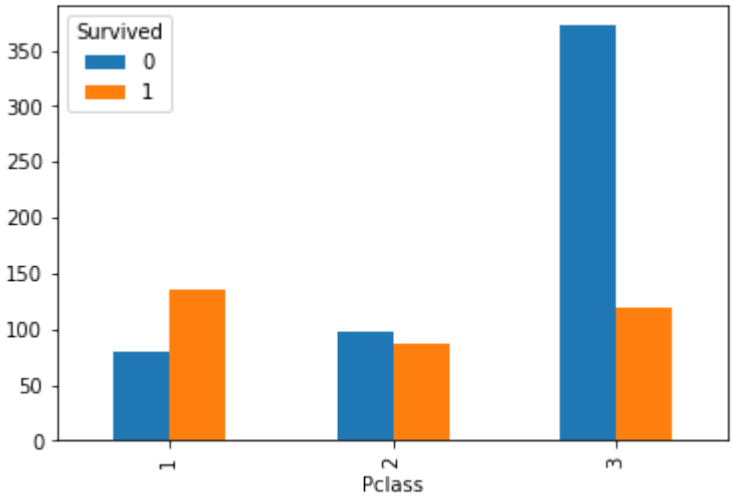
Se pide definir una función (1 vez por ello)

y hacer llamadas a la función (3 en este caso, para: Pclass, Sex, Embarked)

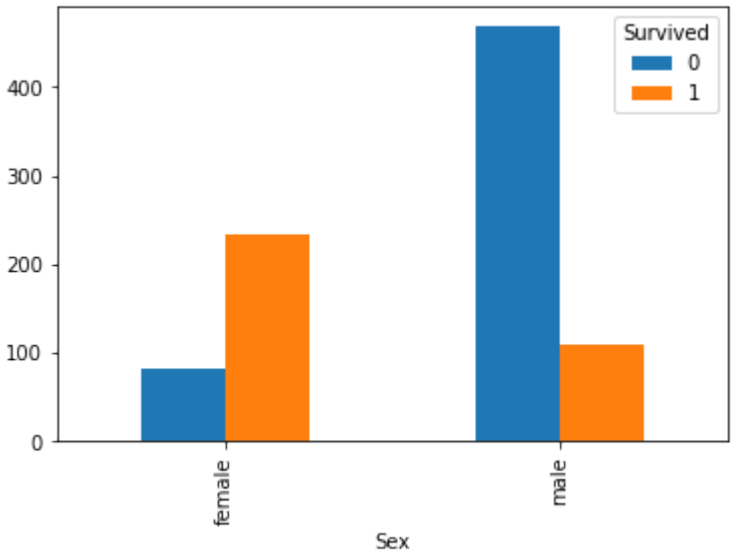
In [10]:
```
# defino la función

def funcion_crosstab(feature):
    pd.crosstab(df[feature], df.Survived).plot(kind="bar")
```

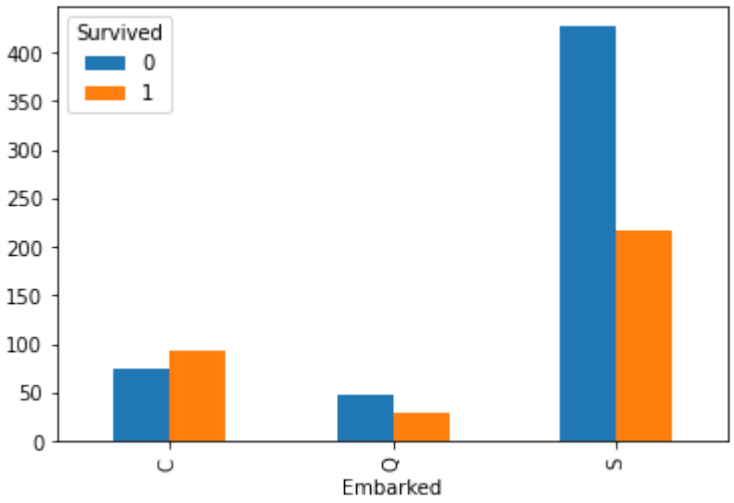
In [11]:
```
# llamo a la función (Llamada primera)
funcion_crosstab("Pclass")
```



In [12]:
```
funcion_crosstab("Sex")
```



In [13]:
```
funcion_crosstab("Embarked")
```



Ejercicio 2

Ejercicio de obtener los valores que muestra el pd.crosstab de Sex y Pclass sin usar el propio pd.crosstab

1) Imprime nuevamente los primeros 5 valores

In [14]:
```
df.head()
```

Out[14]:

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

2) Usando value_counts() observa cuantos hombres y mujeres hay

(No hace falta plotear, simplemente mostrar los números de cada)

In [15]:
```
# NombreDataFrame.NombreColumna.value_counts()
df.Sex.value_counts()
```

Out[15]:
```
male      577
female    314
Name: Sex, dtype: int64
```

3) Sin usar value_counts() observa cuantos hombres y mujeres hay

(con un algoritmo)

In [16]:
```
hombres=0  # contador inicializado en 0
mujeres=0  # contador inicializado en 0

for persona in df.Sex:
    if persona=="male":
        hombres+=1
    else:
        mujeres+=1
```

```
hombres, mujeres
```

Out[16]: (577, 314)

In [23]:
```
# si observaste el mismo número que con value_counts, es que está bien
```

4) Ahora haz lo mismo de otra forma

En esta ocasión se pide que:

crees un dataframe con el formato del original,

bajo la permisa que sea un dataframe con todo hombres (primeramente)

y con todo mujeres (a continuación)

(2 DataFrames por tanto)

Y observes si el número de filas de ambos nuevos DataFrames coincide con los valores anteriores

In [17]:
```
df_hombres = df[df["Sex"]=="male"]
df_hombres.head()
```

Out[17]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fa
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.25
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.05
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.45
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.86
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.07

In [18]:
```
len(df_hombres)
```

Out[18]: 577

In [19]:
```
df_hombres.Sex.value_counts()
```

Out[19]:
```
male    577
Name: Sex, dtype: int64
```

In [20]:
```
df_mujeres = df[df["Sex"]=="female"]
```

```
df_mujeres.head()
```

Out[20]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736

In [21]:
```
len(df_mujeres)
```

Out[21]:
```
314
```

In [22]:
```
df_mujeres.Sex.value_counts()
```

Out[22]:
```
female    314
Name: Sex, dtype: int64
```

In [24]:
```
# si nuevamente observas que los valores son los mismos es que está bien.

# y has hecho lo mismo de 3 formas diferentes.
```

Creado por:

Isabel Maniega