In [4]:
```
import pandas as pd
import numpy as np
import seaborn as sns
```

In [29]:
```
data = pd.read_csv('titanic.csv')
```

In [30]:
```
data.head(10)
```

Out[30]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarke
0	1.0	1.0	Allen, Miss. Elisabeth Walton	female	29.0000	0.0	0.0	24160	211.3375	B5	
1	1.0	1.0	Allison, Master. Hudson Trevor	male	0.9167	1.0	2.0	113781	151.5500	C22 C26	
2	1.0	0.0	Allison, Miss. Helen Loraine	female	2.0000	1.0	2.0	113781	151.5500	C22 C26	
3	1.0	0.0	Allison, Mr. Hudson Joshua Creighton	male	30.0000	1.0	2.0	113781	151.5500	C22 C26	
4	1.0	0.0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.0000	1.0	2.0	113781	151.5500	C22 C26	
5	1.0	1.0	Anderson, Mr. Harry	male	48.0000	0.0	0.0	19952	26.5500	E12	
6	1.0	1.0	Andrews, Miss. Kornelia Theodosia	female	63.0000	1.0	0.0	13502	77.9583	D7	
7	1.0	0.0	Andrews, Mr. Thomas Jr	male	39.0000	0.0	0.0	112050	0.0000	A36	
8	1.0	1.0	Appleton, Mrs. Edward Dale (Charlotte Lamson)	female	53.0000	2.0	0.0	11769	51.4792	C101	
9	1.0	0.0	Artagaveytia, Mr. Ramon	male	71.0000	0.0	0.0	PC 17609	49.5042	NaN	

In [32]:
```
data.tail(10)
```

Out[32]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarke
1300	3.0	1.0	Yasbeck, Mrs. Antoni (Selini Alexander)	female	15.0	1.0	0.0	2659	14.4542	NaN	C
1301	3.0	0.0	Youseff, Mr. Gerious	male	45.5	0.0	0.0	2628	7.2250	NaN	C
1302	3.0	0.0	Yousif, Mr. Wazli	male	NaN	0.0	0.0	2647	7.2250	NaN	C
1303	3.0	0.0	Yousseff, Mr. Gerious	male	NaN	0.0	0.0	2627	14.4583	NaN	C
1304	3.0	0.0	Zabour, Miss. Hileni	female	14.5	1.0	0.0	2665	14.4542	NaN	C
1305	3.0	0.0	Zabour, Miss. Thamine	female	NaN	1.0	0.0	2665	14.4542	NaN	C
1306	3.0	0.0	Zakarian, Mr. Mapriededer	male	26.5	0.0	0.0	2656	7.2250	NaN	C
1307	3.0	0.0	Zakarian, Mr. Ortin	male	27.0	0.0	0.0	2670	7.2250	NaN	C
1308	3.0	0.0	Zimmerman, Mr. Leo	male	29.0	0.0	0.0	315082	7.8750	NaN	S
1309	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In [33]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1310 entries, 0 to 1309
Data columns (total 14 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   pclass     1309 non-null   float64
 1   survived   1309 non-null   float64
 2   name       1309 non-null   object 
 3   sex        1309 non-null   object 
 4   age        1046 non-null   float64
 5   sibsp      1309 non-null   float64
 6   parch      1309 non-null   float64
 7   ticket     1309 non-null   object 
 8   fare       1308 non-null   float64
 9   cabin      295 non-null    object 
 10  embarked   1307 non-null   object 
 11  boat       486 non-null    object 
 12  body       121 non-null    float64
 13  home.dest  745 non-null    object 
dtypes: float64(7), object(7)
memory usage: 143.4+ KB
```

In [34]:
```
data.describe()
```

Out[34]:

	pclass	survived	age	sibsp	parch	fare	body
count	1309.000000	1309.000000	1046.000000	1309.000000	1309.000000	1308.000000	121.000000
mean	2.294882	0.381971	29.881135	0.498854	0.385027	33.295479	160.809917
std	0.837836	0.486055	14.413500	1.041658	0.865560	51.758668	97.696922
min	1.000000	0.000000	0.166700	0.000000	0.000000	0.000000	1.000000
25%	2.000000	0.000000	21.000000	0.000000	0.000000	7.895800	72.000000
50%	3.000000	0.000000	28.000000	0.000000	0.000000	14.454200	155.000000
75%	3.000000	1.000000	39.000000	1.000000	0.000000	31.275000	256.000000
max	3.000000	1.000000	80.000000	8.000000	9.000000	512.329200	328.000000

In [35]:
```
#Cleaning The Data
```

In [36]:
```
print(data.isnull().sum())
```

```
pclass          1
survived        1
name            1
sex             1
age           264
sibsp           1
parch           1
ticket          1
fare            2
cabin        1015
embarked        3
boat          824
body         1189
home.dest     565
dtype: int64
```

In [39]:
```
#Fill missing values for 'Age' with the mean age
data['age'].fillna(data['age'].mean(), inplace=True)

#Fill missing values for 'Embarked' with the mode
data['embarked'].fillna(data['embarked'].mode()[0], inplace=True)

#Fill missing values for 'Cabin' with the mode
data['cabin'].fillna(data['cabin'].mode()[0], inplace=True)
```

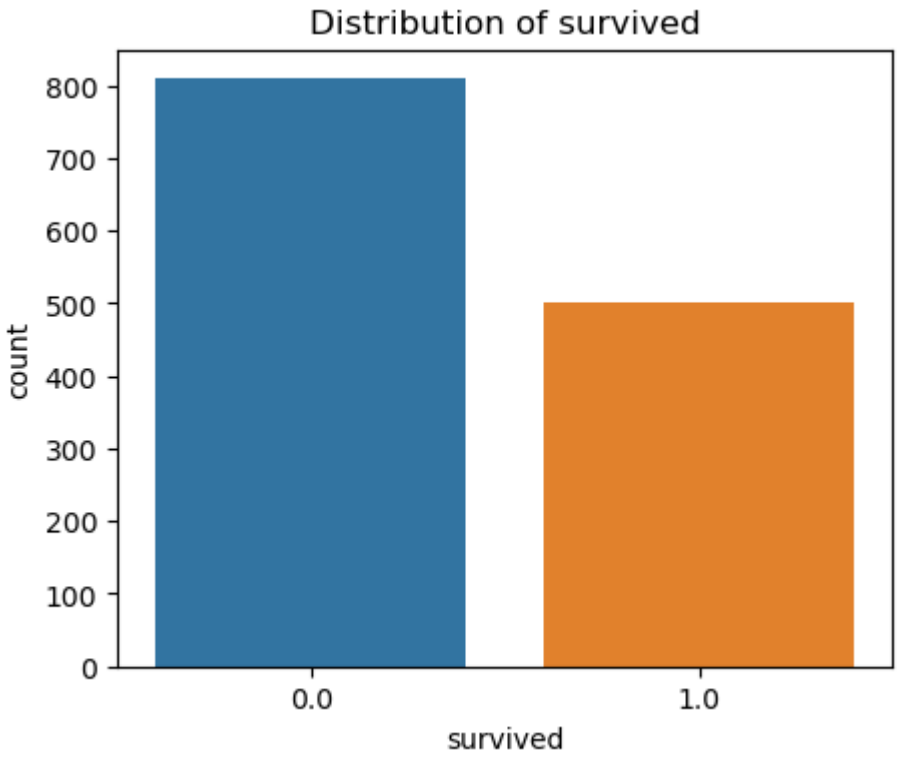
In [42]:
```
# Dropping the Unnecessary data
data.drop(['pclass', 'name', 'ticket', 'fare'], axis=1, inplace=True)
```

In [43]:
```
print(data.isnull().sum())
```

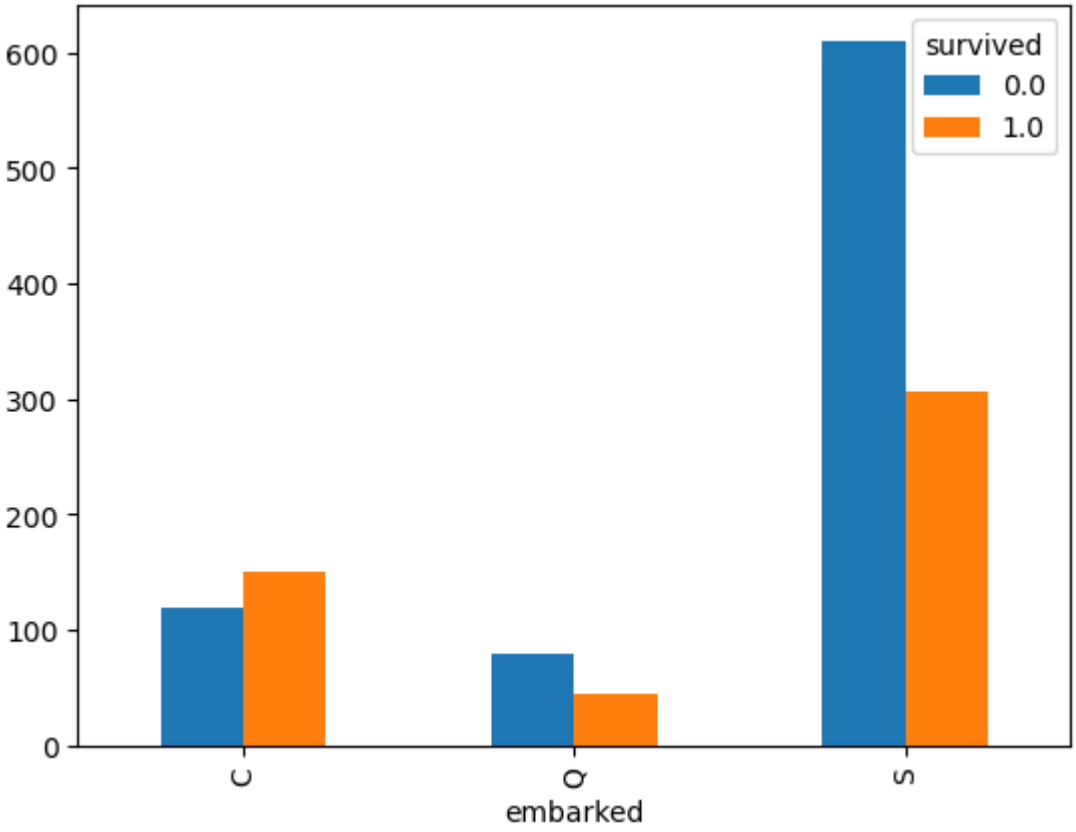
```
survived        1
sex             1
age             0
sibsp           1
parch           1
cabin           0
embarked        0
boat          824
body         1189
home.dest     565
dtype: int64
```

In [44]:
```
#Analyzing The Data
```

In [45]:
```
# Plot distribution of Survived
plt.figure(figsize=(5, 4))
sns.countplot(x='survived', data=data)
plt.title('Distribution of survived')
plt.show()
```



In [46]:
```
# Plot the distribution of 'Embarked' with respect to 'Survived'
counts = data.groupby(['embarked', 'survived']).size().unstack()
counts.plot(kind="bar")
plt.show()
```



In [ ]:
```
# END
```