In [13]:
```
import pandas as pd
```

In [14]:
```
import numpy as np
```

In [18]:
```
autos = pd.read_csv('titanic.csv', encoding = "Latin-1")
```

In [19]:
```
autos.head()
```

Out[19]:

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

1. How many columns are there in the data set?

In [21]:
```
autos.shape
```

Out[21]:
```
(891, 12)
```

In [26]:
```
list(autos.columns)
```

Out[26]:
```
['PassengerId',
 'Survived',
 'Pclass',
 'Name',
 'Sex',
 'Age',
```

```
'SibSp',
'Parch',
'Ticket',
'Fare',
'Cabin',
'Embarked']
```

There are 12 columns in the data set.

1. How many passengers in the data set?

In [24]:
```
autos.count()
```

Out[24]:
```
PassengerId    891
Survived       891
Pclass         891
Name           891
Sex            891
Age            714
SibSp          891
Parch          891
Ticket         891
Fare           891
Cabin          204
Embarked       889
dtype: int64
```

There were 891 passengers in the data set.

1. What percentage of the passengers are male?

In [28]:
```
sex_value = autos.groupby('Sex').size()
```

In [29]:
```
sex_value
```

Out[29]:
```
Sex
female    314
male      577
dtype: int64
```

In [30]:
```
autos.Sex.value_counts(normalize = True) * 100
```

Out[30]:
```
male      64.758698
female    35.241302
Name: Sex, dtype: float64
```

64.8% of the passengers were male.

1. How many of the passengers were female children (<18)

In [36]:
```
Children = autos[(autos.Sex=='Female') & (autos.Age<18)]
```

In [47]:
```
autos[(autos.Sex=='Female') & (autos.Age<18)].shape
```

Out[47]:
```
(0, 12)
```

In [45]:
```
autos.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
```

**Missing info from excel sheet

1. What is the overall survival rate?

In [50]:
```
survived = autos.groupby('Survived').size()
```

In [51]:
```
survived
```

Out[51]:
```
Survived
0    549
1    342
dtype: int64
```

In [54]:
```
(342/891) * 100
```

Out[54]:
```
38.38383838383838
```

In [57]:
```
autos.Survived.value_counts(normalize = True) * 100
```

Out[57]:
```
0    61.616162
1    38.383838
Name: Survived, dtype: float64
```

38.4% of passengers survived.

1. Is there a difference in survival rates between males and females?

In [59]:
```
survival_gender = autos.groupby(['Sex','Survived'])[['Survived']].agg(count_survived=('
```

```
survival_gender
```

Out[59]:

	Sex	Survived	count_survived
0	female	0	81
1	female	1	233
2	male	0	468
3	male	1	109

In [60]:
```
233/(233+81) *100
```

Out[60]:
```
74.20382165605095
```

In [61]:
```
109/(109+468) * 100
```

Out[61]:
```
18.890814558058924
```

74.2 % of the females on board survived while only 18.9% of males on board survived.

1. What are the min and max ages of the passengers? Who are they? Did they survive? Why?

In [67]:
```
autos.loc[autos['Age'].idxmax()]
```

Out[67]:
```
PassengerId                                631
Survived                                     1
Pclass                                       1
Name           Barkworth, Mr. Algernon Henry Wilson
Sex                                       male
Age                                       80.0
SibSp                                        0
Parch                                        0
Ticket                                   27042
Fare                                      30.0
Cabin                                      A23
Embarked                                     S
Name: 630, dtype: object
```

In [68]:
```
autos.loc[autos['Age'].idxmin()]
```

Out[68]:
```
PassengerId                                804
Survived                                     1
Pclass                                       3
Name           Thomas, Master. Assad Alexander
Sex                                       male
Age                                       0.42
SibSp                                        0
Parch                                        1
Ticket                                    2625
Fare                                    8.5167
Cabin                                      NaN
Embarked                                     C
Name: 803, dtype: object
```

Mr. Barkworth was 80 and likely survived because of being a first class passenger and possibly given preferential treatment over other male passenegers. Assadwas under the age of 1 and also survived likely becausewomen and children were placed on life boats first.

In [ ]: