In [40]:
```
import pandas as pd
import numpy as np
```

In [42]:
```
# Read the Data with Pandas
csv_data = pd.read_csv("train.csv")
csv_data
```

Out[42]:

	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	7
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8
...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7

891 rows × 12 columns

In [44]:
```
# show only first five rows
csv_data.head()
```

Out[44]:

	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
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.28
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.92
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.10
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.05

In [46]:

```
# show only first n rows
csv_data.head(10)
```

Out[46]:

	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
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.28
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.92
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.10
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
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.13
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.07

In [48]:

```
# show only last five rows
csv_data.tail()
```

Out[48]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.45
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75

In [50]:
```
# show only last n rows
csv_data.tail(8)
```

Out[50]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
883	884	0	2	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068	1
884	885	0	3	Sutehall, Mr. Henry Jr	male	25.0	0	0	SOTON/OQ 392076	
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	2
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	1
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	3
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	2
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	3
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	

In [52]:
```
#For negative values of n, this function returns all rows except the first |n| rows
csv_data.tail(-4)
```

Out[52]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	F
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1
...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7

887 rows × 12 columns

In [54]:
```
#columns to access the column of the data source
csv_data.columns
```

Out[54]:
```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

In [56]:
```
csv_data.index
```

Out[56]:
```
RangeIndex(start=0, stop=891, step=1)
```

In [58]:
```
#Return the dtypes in the DataFrame
csv_data.dtypes
```

Out[58]:
```
PassengerId      int64
Survived         int64
Pclass           int64
Name            object
Sex             object
Age            float64
SibSp            int64
Parch            int64
Ticket          object
Fare           float64
Cabin           object
Embarked        object
dtype: object
```

In [59]:
```
csv_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 [61]:
```
csv_data.values
```

Out[61]:
```
array([[1, 0, 3, ..., 7.25, nan, 'S'],
       [2, 1, 1, ..., 71.2833, 'C85', 'C'],
       [3, 1, 3, ..., 7.925, nan, 'S'],
       ...,
       [889, 0, 3, ..., 23.45, nan, 'S'],
       [890, 1, 1, ..., 30.0, 'C148', 'C'],
       [891, 0, 3, ..., 7.75, nan, 'Q']], shape=(891, 12), dtype=object)
```

In [63]:
```
csv_data.value_counts()
```

Out[63]:
```
PassengerId  Survived  Pclass  Name
Sex     Age   SibSp  Parch  Ticket    Fare     Cabin        Embarked
2            1         1       Cumings, Mrs. John Bradley (Florence Briggs Thayer)
female  38.0  1      0      PC 17599  71.2833  C85          C           1
4            1         1       Futrelle, Mrs. Jacques Heath (Lily May Peel)
female  35.0  1      0      113803    53.1000  C123         S           1
7            0         1       McCarthy, Mr. Timothy J
male    54.0  0      0      17463     51.8625  E46          S           1
11           1         3       Sandstrom, Miss. Marguerite Rut
female  4.0   1      1      PP 9549   16.7000  G6           S           1
12           1         1       Bonnell, Miss. Elizabeth
female  58.0  0      0      113783    26.5500  C103         S           1

..
872          1         1       Beckwith, Mrs. Richard Leonard (Sallie Monypeny)
female  47.0  1      1      11751     52.5542  D35          S           1
873          0         1       Carlsson, Mr. Frans Olof
male    33.0  0      0      695       5.0000   B51 B53 B55  S           1
880          1         1       Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)
female  56.0  0      1      11767     83.1583  C50          C           1
888          1         1       Graham, Miss. Margaret Edith
female  19.0  0      0      112053    30.0000  B42          S           1
890          1         1       Behr, Mr. Karl Howell
male    26.0  0      0      111369    30.0000  C148         C           1
Name: count, Length: 183, dtype: int64
```

In [65]:
```
#Return an int representing the number of axes / array dimensions
csv_data.ndim
```

Out[65]:
```
2
```

In [67]:
```
#Return an int representing the number of elements in this object
csv_data.size
```

Out[67]:
```
10692
```

In [69]:
```
#Return a tuple representing the dimensionality of the DataFrame
csv_data.shape
```

Out[69]:
```
(891, 12)
```

In [71]:
```
# Indicator whether Series/DataFrame is empty
csv_data.empty
```

Out[71]:
```
False
```

In [73]:
```
#access column using label
csv_data["PassengerId"]
```

Out[73]:
```
0        1
1        2
2        3
3        4
4        5
      ...
886    887
887    888
888    889
889    890
890    891
Name: PassengerId, Length: 891, dtype: int64
```

In [74]:
```
# Access a group of rows and columns by label(s)
csv_data.loc[5,'PassengerId']
```

Out[74]:
```
np.int64(6)
```

In [76]:
```
csv_data.loc[5]
```

Out[76]:
```
PassengerId                  6
Survived                     0
Pclass                       3
Name           Moran, Mr. James
Sex                        male
Age                         NaN
SibSp                         0
Parch                         0
Ticket                   330877
Fare                     8.4583
Cabin                       NaN
Embarked                      Q
Name: 5, dtype: object
```

In [78]:
```
csv_data.loc[2:5]
```

Out[78]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fa
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.92
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1(
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

In [80]:
```
#Access row
csv_data.iloc[5]
```

Out[80]:
```
PassengerId                  6
Survived                     0
Pclass                       3
Name           Moran, Mr. James
Sex                        male
Age                         NaN
SibSp                         0
Parch                         0
Ticket                   330877
Fare                     8.4583
Cabin                       NaN
Embarked                      Q
Name: 5, dtype: object
```

In [82]:
```
csv_data["Name"].iloc[1]
```

Out[82]:
```
'Cumings, Mrs. John Bradley (Florence Briggs Thayer)'
```

In [84]:
```
csv_data.iloc[5:10]
```

Out[84]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Far
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.458
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.862
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.075
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.133
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.070

In [86]:

```
# Data Filtering
csv_data[csv_data['Sex'] == 'female']
```

Out[86]:

	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	71
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30
...	...	...	...	...	...	...	...	...	...	
880	881	1	2	Shelley, Mrs. William (Imanita Parrish Hall)	female	25.0	0	1	230433	26
882	883	0	3	Dahlberg, Miss. Gerda Ulrika	female	22.0	0	0	7552	10
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23

314 rows × 12 columns

In [88]:
```
csv_data[(csv_data.Sex=='female') & (csv_data.Age>60)]
```

Out[88]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	F
275	276	1	1	Andrews, Miss. Kornelia Theodosia	female	63.0	1	0	13502	77.9
483	484	1	3	Turkula, Mrs. (Hedwig)	female	63.0	0	0	4134	9.5
829	830	1	1	Stone, Mrs. George Nelson (Martha Evelyn)	female	62.0	0	0	113572	80.0

In [89]:
```
# str accessor to filter rows based on strings.
csv_data[csv_data.Sex.str.startswith('m')]
```

Out[89]:

	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	7
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21
...	...	...	...	...	...	...	...	...	...	
883	884	0	2	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068	10
884	885	0	3	Sutehall, Mr. Henry Jr	male	25.0	0	0	SOTON/OQ 392076	7
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7

577 rows × 12 columns

In [91]:
```
# isin method to filter the names that exist in a given list
names = ['m','f','male']
csv_data[csv_data.Sex.isin(names)]
```

Out[91]:

	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	7
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21
...	...	...	...	...	...	...	...	...	...	
883	884	0	2	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068	10
884	885	0	3	Sutehall, Mr. Henry Jr	male	25.0	0	0	SOTON/OQ 392076	7
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7

577 rows × 12 columns

In [93]:
```
# query function can pass the conditions as a string
csv_data.query('Sex == "female" and Age > 60')
```

Out[93]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	F
275	276	1	1	Andrews, Miss. Kornelia Theodosia	female	63.0	1	0	13502	77.9
483	484	1	3	Turkula, Mrs. (Hedwig)	female	63.0	0	0	4134	9.5
829	830	1	1	Stone, Mrs. George Nelson (Martha Evelyn)	female	62.0	0	0	113572	80.0

In [95]:
```
# Descriptive statistics include those that summarize the central tendency, dispersio
csv_data.describe()
```

Out[95]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.00
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.20
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.69
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.00
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.91
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.45
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.00
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.32

In [97]:
```
# For numeric data, the result's index will include count, mean, std, min, max as wel
# For object data (e.g. strings or timestamps), the result's index will include count
# The top is the most common value. The freq is the most common value's frequency. Ti
csv_data.describe(include="all")
```

Out[97]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	
count	891.000000	891.000000	891.000000	891	891	714.000000	891.000000	89
unique	NaN	NaN	NaN	891	2	NaN	NaN	
top	NaN	NaN	NaN	Braund, Mr. Owen Harris	male	NaN	NaN	
freq	NaN	NaN	NaN	1	577	NaN	NaN	
mean	446.000000	0.383838	2.308642	NaN	NaN	29.699118	0.523008	
std	257.353842	0.486592	0.836071	NaN	NaN	14.526497	1.102743	
min	1.000000	0.000000	1.000000	NaN	NaN	0.420000	0.000000	
25%	223.500000	0.000000	2.000000	NaN	NaN	20.125000	0.000000	
50%	446.000000	0.000000	3.000000	NaN	NaN	28.000000	0.000000	
75%	668.500000	1.000000	3.000000	NaN	NaN	38.000000	1.000000	
max	891.000000	1.000000	3.000000	NaN	NaN	80.000000	8.000000	

In [98]:
```
# Including only numeric columns
csv_data.describe(include=[np.number])
```

Out[98]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.00
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.20
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.69
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.00
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.91
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.45
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.00
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.32

In [100…
```
# ignore non-numeric data for processing
csv_data.describe(exclude=["O"])
```

Out[100…

	PassengerId	Survived	Pclass	Age	SibSp	Parch	
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.00
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.20
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.69
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.00
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.91
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.45
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.00
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.32

In [102…
```
csv_data.isnull()
```

Out[102…

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cal
0	False	False	False	False	False	False	False	False	False	False	T
1	False	False	False	False	False	False	False	False	False	False	Fa
2	False	False	False	False	False	False	False	False	False	False	T
3	False	False	False	False	False	False	False	False	False	False	Fa
4	False	False	False	False	False	False	False	False	False	False	T
...	...	...	...	...	...	...	...	...	...	...	
886	False	False	False	False	False	False	False	False	False	False	T
887	False	False	False	False	False	False	False	False	False	False	Fa
888	False	False	False	False	False	True	False	False	False	False	T
889	False	False	False	False	False	False	False	False	False	False	Fa
890	False	False	False	False	False	False	False	False	False	False	T

891 rows × 12 columns

In [104…
```
csv_data.isna()
```

Out[104…

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cal
0	False	False	False	False	False	False	False	False	False	False	T
1	False	False	False	False	False	False	False	False	False	False	Fa
2	False	False	False	False	False	False	False	False	False	False	T
3	False	False	False	False	False	False	False	False	False	False	Fa
4	False	False	False	False	False	False	False	False	False	False	T
...	...	...	...	...	...	...	...	...	...	...	
886	False	False	False	False	False	False	False	False	False	False	T
887	False	False	False	False	False	False	False	False	False	False	Fa
888	False	False	False	False	False	True	False	False	False	False	T
889	False	False	False	False	False	False	False	False	False	False	Fa
890	False	False	False	False	False	False	False	False	False	False	T

891 rows × 12 columns

In [105…
```
# find missing value
csv_data.isnull().values.any()
```

Out[105… np.True_

In [108…
```
csv_data.isna().values.any()
```

Out[108… np.True_

In [113…
```
csv_data.duplicated().values.any()
```

Out[113… np.False_

In [115…
```
data_missing = pd.read_csv("train.csv")
print(data_missing.isnull().values.any())
```

True

In [119…
```
data_missing.dropna()
```

Out[119…

	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	71.
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.
...	...	...	...	...	...	...	...	...	...	
871	872	1	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0	1	1	11751	52.
872	873	0	1	Carlsson, Mr. Frans Olof	male	33.0	0	0	695	5.
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0	1	11767	83.
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.

183 rows × 12 columns

In [121…
```
data_missing.isnull().values.any()
```

Out[121… np.True_

In [125…
```
data_missing=data_missing.dropna()
data_missing.isnull().values.any()
```

Out[125… np.False_

In [127…
```
csv_data= csv_data.drop("Sex", axis=1)
csv_data
```

Out[127…

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

891 rows × 11 columns

In [133…
```
csv_data['Survived'].mean()
```

Out[133…
```
np.float64(0.3838383838383838)
```

In [137…
```
csv_data['Survived'].median()
```

Out[137…
```
0.0
```

In [141…
```
data_filling=data_missing.fillna(csv_data['Survived'].mean())
data_filling.head(10)
```

Out[141…

	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	71.
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.
21	22	1	2	Beesley, Mr. Lawrence	male	34.0	0	0	248698	13.
23	24	1	1	Sloper, Mr. William Thompson	male	28.0	0	0	113788	35.
27	28	0	1	Fortune, Mr. Charles Alexander	male	19.0	3	2	19950	263.
52	53	1	1	Harper, Mrs. Henry Sleeper (Myna Haxtun)	female	49.0	1	0	PC 17572	76.
54	55	0	1	Ostby, Mr. Engelhart Cornelius	male	65.0	0	1	113509	61.

In [143…

```
data_filling=data_missing.fillna(csv_data['Survived'].median())
data_filling.head(10)
```

Out[143…

	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	71.
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.
21	22	1	2	Beesley, Mr. Lawrence	male	34.0	0	0	248698	13.
23	24	1	1	Sloper, Mr. William Thompson	male	28.0	0	0	113788	35.
27	28	0	1	Fortune, Mr. Charles Alexander	male	19.0	3	2	19950	263.
52	53	1	1	Harper, Mrs. Henry Sleeper (Myna Haxtun)	female	49.0	1	0	PC 17572	76.
54	55	0	1	Ostby, Mr. Engelhart Cornelius	male	65.0	0	1	113509	61.

In [155…

```
#In order to convert data types in pandas, there are three basic options:
#Use astype() to force an appropriate dtype
#Create a custom function to convert the data
#Use pandas functions such as to_numeric() or to_datetime()
csv_data['Age'] = pd.to_numeric(csv_data['Age'], errors='coerce').round().astype('Int
csv_data['Age']
```

Out[155…

```
0        22
1        38
2        26
3        35
4        35
       ...
886      27
887      19
888    <NA>
889      26
890      32
Name: Age, Length: 891, dtype: Int64
```

In [157…

```
csv_data.dtypes
```

Out[157…

```
PassengerId      int64
Survived         int64
Pclass           int64
Name            object
Age              Int64
SibSp            int64
Parch            int64
Ticket          object
Fare           float64
Cabin           object
Embarked        object
dtype: object
```