In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:

```
df = pd.read_csv('train.csv')
df.head()
```

Out[2]:

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

In [3]:

```
1 df.tail(5)
```

Out[3]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cab
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00	Na
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00	B
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.45	Na
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00	C1
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75	Na

In [4]:

```
1 df.sample(5)
```

Out[4]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
737	738	1	1	Lesurer, Mr. Gustave J	male	35.0	0	0	PC 17755	512.3292
766	767	0	1	Brewe, Dr. Arthur Jackson	male	NaN	0	0	112379	39.6000
741	742	0	1	Cavendish, Mr. Tyrell William	male	36.0	1	0	19877	78.8500
360	361	0	3	Skoog, Mr. Wilhelm	male	40.0	1	4	347088	27.9000
173	174	0	3	Sivola, Mr. Antti Wilhelm	male	21.0	0	0	STON/O 2. 3101280	7.9250

In [5]:

```
df.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 [ ]:

```
# There are 891 entries in the data set
# We have 12 Columns in train.csv
# Memory usage is 83.7+ KB
# Datatypes used in the data set are int64, float64, object
# Column Count
# Index ranges from 0 to 891
```

In [ ]:

```
df.columns
```

In [ ]:

```
df.describe()
```

In [ ]:

```
df.corr()
```

In [ ]:

```
import seaborn as sns
sns.heatmap(df.corr())
```

In [ ]:

```
df.size
```

In [ ]:

```
df.ndim
```

In [ ]:

```
df.shape
```

In [ ]:

```
df.shape[0]
```

In [ ]:

```
df.shape[1]
```

In [ ]:

```
count = df.sum()/int(df.shape[0])
count
```

In [ ]:

```
df.notnull()
```

In [ ]:

```
df.isnull()
```

In [ ]:

```
df.mean()['Age']
```

In [ ]:

```
df.median()['Age']
```

In [ ]:

```
df['Sex'].nunique()
```

In [ ]:

```
df['Sex'].value_counts()
```

In [ ]:

```
df.nlargest(5,'Sex')
```

In [ ]:

```
sns.countplot(df['Sex'])
```

In [ ]:

```
1 sns.barplot(df['Age'])
```

In [ ]:

```
1 sns.lineplot(x = df['Age'], y = df['Sex'])
```

In [ ]:

```
1
```