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

In [2]:
```
pd.read_csv("https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv")
```

Out[2]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

In [4]:
```
df=pd.read_csv("/home/harshit/Desktop/Titanic.csv")
```

In [6]:
```
#names of columns


df.columns
```

Out[6]:
```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

In [ ]:
```

```

In [7]:
```
#see number of rows as number of columns


df.shape
```

Out[7]:
```
(891, 12)
```

""" rows----->records

columns------>features/attributes of the data """

In [8]:
```
#random samples (this is a function)



df.sample(5)
```

Out[8]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
354	355	0	3	Yousif, Mr. Wazli	male	NaN	0	0	2647	7.2250	NaN	C
786	787	1	3	Sjoblom, Miss. Anna Sofia	female	18.0	0	0	3101265	7.4958	NaN	S
271	272	1	3	Tornquist, Mr. William Henry	male	25.0	0	0	LINE	0.0000	NaN	S
417	418	1	2	Silven, Miss. Lyyli Karoliina	female	18.0	0	2	250652	13.0000	NaN	S
762	763	1	3	Barah, Mr. Hanna Assi	male	20.0	0	0	2663	7.2292	NaN	C

In [9]:
```
#first N rows from the data


df.head(4) #first 4 rows from the given data set
```

Out[9]:

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

In [10]:
```
#how to locate rows?

#find a single row by its number

#765

df.loc[  [765] ,       ]
```

Out[10]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
765	766	1	1	Hogeboom, Mrs. John C (Anna Andrews)	female	51.0	1	0	13502	77.9583	D11	S

In [11]:
```
#765th row, Name and Fare column

df.loc[      [765]  ,  ['Name', 'Fare']      ]
```

Out[11]:

	Name	Fare
765	Hogeboom, Mrs. John C (Anna Andrews)	77.9583

In [12]:
```
#765th row, Name and Fare column

df.loc[      765      ] #series data type
```

Out[12]:
```
PassengerId                                766
Survived                                     1
Pclass                                       1
Name           Hogeboom, Mrs. John C (Anna Andrews)
Sex                                       female
Age                                           51
SibSp                                          1
Parch                                          0
Ticket                                     13502
Fare                                     77.9583
Cabin                                        D11
Embarked                                       S
Name: 765, dtype: object
```

In [15]:
```
#multiple rows

#765 and 560

df.loc[      [ 765, 560 ]  ,  ['Name', 'Fare']      ]
```

Out[15]:

	Name	Fare
765	Hogeboom, Mrs. John C (Anna Andrews)	77.9583
560	Morrow, Mr. Thomas Rowan	7.7500

In [16]:
```
df.loc[      [ 765, 560,19,86,167,190 ]  ,  ['Name', 'Fare']      ]
```

Out[16]:

	Name	Fare
765	Hogeboom, Mrs. John C (Anna Andrews)	77.9583
560	Morrow, Mr. Thomas Rowan	7.7500
19	Masselmani, Mrs. Fatima	7.2250
86	Ford, Mr. William Neal	34.3750
167	Skoog, Mrs. William (Anna Bernhardina Karlsson)	27.9000
190	Pinsky, Mrs. (Rosa)	13.0000

In [21]:
```
# df.loc[      :   , ['Name', 'Fare']      ]
```

In [23]:
```
#sequence of rows

#all rows between 1 and 10(both inclusive)

df.loc[      1:10  ,  ['Name', 'Fare']      ]
```

Out[23]:

	Name	Fare
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	71.2833
2	Heikkinen, Miss. Laina	7.9250
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	53.1000
4	Allen, Mr. William Henry	8.0500
5	Moran, Mr. James	8.4583
6	McCarthy, Mr. Timothy J	51.8625
7	Palsson, Master. Gosta Leonard	21.0750
8	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	11.1333
9	Nasser, Mrs. Nicholas (Adele Achem)	30.0708
10	Sandstrom, Miss. Marguerite Rut	16.7000

In [29]:
```
#show columns of my choice


df[     [  'Name','Age','Fare'  ]     ]
```

Out[29]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

In [25]:
```
#find out the number of missing values in each column


df.isna().sum()
```

Out[25]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

In [30]:
```
#chaining of commands

#show name and age of first 5 rows
```

```
df.head(5)[ ['Name','Age'] ]
```

Out[30]:

	Name	Age
0	Braund, Mr. Owen Harris	22.0
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	38.0
2	Heikkinen, Miss. Laina	26.0
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	35.0
4	Allen, Mr. William Henry	35.0

In [32]:
```
#show Names of 8 random records
df.sample(8)[ ['Name'] ]
```

Out[32]:

	Name
468	Scanlan, Mr. James
883	Banfield, Mr. Frederick James
347	Davison, Mrs. Thomas Henry (Mary E Finck)
88	Fortune, Miss. Mabel Helen
413	Cunningham, Mr. Alfred Fleming
510	Daly, Mr. Eugene Patrick
479	Hirvonen, Miss. Hildur E
153	van Billiard, Mr. Austin Blyler

In [33]:
```
#show Names of 8 random records
df.sample(8)[ 'Name' ]
```

Out[33]:
```
46                          Lennon, Mr. Denis
580               Christy, Miss. Julie Rachel
369              Aubart, Mme. Leontine Pauline
789                   Guggenheim, Mr. Benjamin
787                  Rice, Master. George Hugh
100                    Petranec, Miss. Matilda
579                        Jussila, Mr. Eiriik
830    Yasbeck, Mrs. Antoni (Selini Alexander)
Name: Name, dtype: object
```

In [34]:
```
#show names and age of n number of rows and print the missing value count of each

df.sample(8)[[ "Name",'Age' ]].isna().sum()
```

Out[34]:
```
Name    0
Age     2
dtype: int64
```

In [35]:
```
df
```

Out[35]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

In [37]:
```
df[['Fare']].sum() #sum of all the values
```

Out[37]:
```
Fare    28693.9493
dtype: float64
```

In [38]:
```
df[['Fare']].mean() #average of all the values
```

Out[38]:
```
Fare    32.204208
dtype: float64
```

In [39]:
```
df[['Fare']].min() #minimum of all the values
```

Out[39]:
```
Fare    0.0
dtype: float64
```

In [40]:
```
#average Age and average Fare in one command?

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

Out[40]:
```
Age     29.699118
Fare    32.204208
dtype: float64
```

In [42]:
```
#aggregation operations
"""

apply multiple aggregations on one or more than one column at the same time

"""

df[['Age','Fare']].agg(  ['mean','sum']  )
```

Out[42]:

	Age	Fare
mean	29.699118	32.204208
sum	21205.170000	28693.949300

In [45]:
```
df[['Age','Fare']].agg(  ['mean','sum','min','max']  )
```

Out[45]:

	Age	Fare
mean	29.699118	32.204208
sum	21205.170000	28693.949300
min	0.420000	0.000000
max	80.000000	512.329200

In [46]:
```
df.describe()
```

Out[46]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

In [49]:
```
df[['Age']].quantile(0.9) #90 percentile
```

Out[49]:
```
Age    50.0
Name: 0.9, dtype: float64
```

In [53]:
```
#apply different aggregates on different columns


#sum of fare and average age in the same command

df[['Age','Fare']].agg( {  'Age':'mean' , 'Fare':'sum'  }  )
```

Out[53]:
```
Age        29.699118
Fare    28693.949300
dtype: float64
```

In [54]:
```
#sum of fare and min, max and average age in the same command
```

```
df[['Age','Fare']].agg( {  'Age':['mean','min','max'] , 'Fare':'sum'  }  )
```

Out[54]:

	Age	Fare
max	80.000000	NaN
mean	29.699118	NaN
min	0.420000	NaN
sum	NaN	28693.9493

In [ ]:
```
#calculate the average age of female passengers in the dataset(filtering!)



#average fare collected from Pclass 1 and Pclass 2 passengers(filtering)
```

In [55]:
```
df
```

Out[55]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

In [57]:
```
condition = ( df ['Sex'] == 'female'  )


df[   condition  ]
```

Out[57]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	C
...	...	...	...	...	...	...	...	...	...	...	...	...
880	881	1	2	Shelley, Mrs. William (Imanita Parrish Hall)	female	25.0	0	1	230433	26.0000	NaN	S
882	883	0	3	Dahlberg, Miss. Gerda Ulrika	female	22.0	0	0	7552	10.5167	NaN	S
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.1250	NaN	Q
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S

888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S

314 rows × 12 columns

In [58]:
```
#column data type information

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 [60]:
```
#pclass 1 OR pclass 2


condition = ( df ['Pclass'] == 1  ) | (df ['Pclass'] == 2)
df[condition]
```

Out[60]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	C
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500	C103	S
...	...	...	...	...	...	...	...	...	...	...	...	...
880	881	1	2	Shelley, Mrs. William (Imanita Parrish Hall)	female	25.0	0	1	230433	26.0000	NaN	S
883	884	0	2	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068	10.5000	NaN	S
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C

400 rows × 12 columns

In [61]:
```
#age > 25 and gender is female

condition=(df['Age'] > 25 )  & (df['Sex']=='female')
df[condition]
```

Out[61]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
8	9	1	3	Johnson, Mrs. Oscar W								

8	9	1	3	(Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500	C103	S
...	...	...	...	...	...	...	...	...	...	...	...	...
866	867	1	2	Duran y More, Miss. Asuncion	female	27.0	1	0	SC/PARIS 2149	13.8583	NaN	C
871	872	1	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0	1	1	11751	52.5542	D35	S
874	875	1	2	Abelson, Mrs. Samuel (Hannah Wizosky)	female	28.0	1	0	P/PP 3381	24.0000	NaN	C
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0	1	11767	83.1583	C50	C
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.1250	NaN	Q

139 rows × 12 columns

In [63]:
```
#female passengers above the age of 20 from Pclass 3

condition = ( df ['Age'] > 25 ) & (df ['Sex'] == 'female') & (df ['Pclass'] == 3)
df[condition]
```

Out[63]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S
18	19	0	3	Vander Planke, Mrs. Julius (Emelia Maria Vande...	female	31.0	1	0	345763	18.0000	NaN	S
25	26	1	3	Asplund, Mrs. Carl Oscar (Selma Augusta Emilia...	female	38.0	1	5	347077	31.3875	NaN	S
40	41	0	3	Ahlin, Mrs. Johan (Johanna Persdotter Larsson)	female	40.0	1	0	7546	9.4750	NaN	S
79	80	1	3	Dowdell, Miss. Elizabeth	female	30.0	0	0	364516	12.4750	NaN	S
85	86	1	3	Backstrom, Mrs. Karl Alfred (Maria Mathilda Gu...	female	33.0	3	0	3101278	15.8500	NaN	S
100	101	0	3	Petranec, Miss. Matilda	female	28.0	0	0	349245	7.8958	NaN	S
132	133	0	3	Robins, Mrs. Alexander A (Grace Charity Laury)	female	47.0	1	0	A/5. 3337	14.5000	NaN	S
167	168	0	3	Skoog, Mrs. William (Anna Bernhardina Karlsson)	female	45.0	1	4	347088	27.9000	NaN	S
216	217	1	3	Honkanen, Miss. Eliina	female	27.0	0	0	STON/O2. 3101283	7.9250	NaN	S
251	252	0	3	Strom, Mrs. Wilhelm (Elna Matilda Persson)	female	29.0	1	1	347054	10.4625	G6	S
254	255	0	3	Rosblom, Mrs. Viktor (Helena Wilhelmina)	female	41.0	0	2	370129	20.2125	NaN	S
255	256	1	3	Touma, Mrs. Darwis (Hanne Youssef Razi)	female	29.0	0	2	2650	15.2458	NaN	C
276	277	0	3	Lindblom, Miss. Augusta Charlotta	female	45.0	0	0	347073	7.7500	NaN	S
279	280	1	3	Abbott, Mrs. Stanton (Rosa Hunt)	female	35.0	1	1	C.A. 2673	20.2500	NaN	S
315	316	1	3	Nilsson, Miss. Helmina Josefina	female	26.0	0	0	347470	7.8542	NaN	S
328	329	1	3	Goldsmith, Mrs. Frank John (Emily Alice Brown)	female	31.0	1	1	363291	20.5250	NaN	S
362	363	0	3	Barbara, Mrs. (Catherine David)	female	45.0	0	1	2691	14.4542	NaN	C
396	397	0	3	Olsson, Miss. Elina	female	31.0	0	0	350407	7.8542	NaN	S
423	424	0	3	Danbom, Mrs. Ernst Gilbert (Anna Sigrid	female	28.0	1	1	347080	14.4000	NaN	S

				Maria ...								
483	484	1	3	Turkula, Mrs. (Hedwig)	female	63.0	0	0	4134	9.5875	NaN	S
503	504	0	3	Laitinen, Miss. Kristina Sofia	female	37.0	0	0	4135	9.5875	NaN	S
534	535	0	3	Cacic, Miss. Marija	female	30.0	0	0	315084	8.6625	NaN	S
559	560	1	3	de Messemaeker, Mrs. Guillaume Joseph (Emma)	female	36.0	1	0	345572	17.4000	NaN	S
567	568	0	3	Palsson, Mrs. Nils (Alma Cornelia Berglund)	female	29.0	0	4	349909	21.0750	NaN	S
610	611	0	3	Andersson, Mrs. Anders Johan (Alfrida Konstant...	female	39.0	1	5	347082	31.2750	NaN	S
617	618	0	3	Lobb, Mrs. William Arthur (Cordelia K Stanlick)	female	26.0	1	0	A/5. 3336	16.1000	NaN	S
638	639	0	3	Panula, Mrs. Juha (Maria Emilia Ojala)	female	41.0	0	5	3101295	39.6875	NaN	S
657	658	0	3	Bourke, Mrs. John (Catherine)	female	32.0	1	1	364849	15.5000	NaN	Q
678	679	0	3	Goodwin, Mrs. Frederick (Augusta Tyler)	female	43.0	1	6	CA 2144	46.9000	NaN	S
736	737	0	3	Ford, Mrs. Edward (Margaret Ann Watson)	female	48.0	1	3	W./C. 6608	34.3750	NaN	S
767	768	0	3	Mangan, Miss. Mary	female	30.5	0	0	364850	7.7500	NaN	Q
797	798	1	3	Osman, Mrs. Mara	female	31.0	0	0	349244	8.6833	NaN	S
799	800	0	3	Van Impe, Mrs. Jean Baptiste (Rosalie Paula Go...	female	30.0	1	1	345773	24.1500	NaN	S
823	824	1	3	Moor, Mrs. (Beila)	female	27.0	0	1	392096	12.4750	E121	S
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.1250	NaN	Q

In [64]:
```
#female passengers between the age of 20 to 25(both included) from Pclass 3
condition = ( df ['Age'].between(20,25) ) & (df ['Sex'] == 'female') & (df ['Pclass'] == 3)
df[condition]
```

Out[64]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
106	107	1	3	Salkjelsvik, Miss. Anna Kristine	female	21.0	0	0	343120	7.6500	NaN	S
113	114	0	3	Jussila, Miss. Katriina	female	20.0	1	0	4136	9.8250	NaN	S
141	142	1	3	Nysten, Miss. Anna Sofia	female	22.0	0	0	347081	7.7500	NaN	S
142	143	1	3	Hakkarainen, Mrs. Pekka Pietari (Elin Matilda ...	female	24.0	1	0	STON/O2. 3101279	15.8500	NaN	S
246	247	0	3	Lindahl, Miss. Agda Thorilda Viktoria	female	25.0	0	0	347071	7.7750	NaN	S
289	290	1	3	Connolly, Miss. Kate	female	22.0	0	0	370373	7.7500	NaN	Q
293	294	0	3	Haas, Miss. Aloisia	female	24.0	0	0	349236	8.8500	NaN	S
376	377	1	3	Landergren, Miss. Aurora Adelia	female	22.0	0	0	C 7077	7.2500	NaN	S
394	395	1	3	Sandstrom, Mrs. Hjalmar (Agnes Charlotta Bengt...	female	24.0	0	2	PP 9549	16.7000	G6	S
402	403	0	3	Jussila, Miss. Mari Aina	female	21.0	1	0	4137	9.8250	NaN	S
404	405	0	3	Oreskovic, Miss. Marija	female	20.0	0	0	315096	8.6625	NaN	S
436	437	0	3	Ford, Miss. Doolina Margaret "Daisy"	female	21.0	2	2	W./C. 6608	34.3750	NaN	S
474	475	0	3	Strandberg, Miss. Ida Sofia	female	22.0	0	0	7553	9.8375	NaN	S
501	502	0	3	Canavan, Miss. Mary	female	21.0	0	0	364846	7.7500	NaN	Q
554	555	1	3	Ohman, Miss. Velin	female	22.0	0	0	347085	7.7750	NaN	S
649	650	1	3	Stanley, Miss. Amy Zillah Elsie	female	23.0	0	0	CA. 2314	7.5500	NaN	S
				Ilmakangas, Miss. Pieta					STON/O2.			

729	730	0	3	Sofia	female	25.0	1	0	3101271	7.9250	NaN	S
816	817	0	3	Heininen, Miss. Wendla Maria	female	23.0	0	0	STON/O2. 3101290	7.9250	NaN	S
858	859	1	3	Baclini, Mrs. Solomon (Latifa Qurban)	female	24.0	0	3	2666	19.2583	NaN	C
882	883	0	3	Dahlberg, Miss. Gerda Ulrika	female	22.0	0	0	7552	10.5167	NaN	S

In [65]:
```
#show records of passengers who paid fare between 100 to 200

condition=(df["Fare"].between(100,200))
df[condition]
```

Out[65]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
31	32	1	1	Spencer, Mrs. William Augustus (Marie Eugenie)	female	NaN	1	0	PC 17569	146.5208	B78	C
195	196	1	1	Lurette, Miss. Elise	female	58.00	0	0	PC 17569	146.5208	B80	C
215	216	1	1	Newell, Miss. Madeleine	female	31.00	1	0	35273	113.2750	D36	C
268	269	1	1	Graham, Mrs. William Thompson (Edith Junkins)	female	58.00	0	1	PC 17582	153.4625	C125	S
269	270	1	1	Bissette, Miss. Amelia	female	35.00	0	0	PC 17760	135.6333	C99	S
297	298	0	1	Allison, Miss. Helen Loraine	female	2.00	1	2	113781	151.5500	C22 C26	S
305	306	1	1	Allison, Master. Hudson Trevor	male	0.92	1	2	113781	151.5500	C22 C26	S
306	307	1	1	Fleming, Miss. Margaret	female	NaN	0	0	17421	110.8833	NaN	C
307	308	1	1	Penasco y Castellana, Mrs. Victor de Satode (M...	female	17.00	1	0	PC 17758	108.9000	C65	C
318	319	1	1	Wick, Miss. Mary Natalie	female	31.00	0	2	36928	164.8667	C7	S
319	320	1	1	Spedden, Mrs. Frederic Oakley (Margaretta Corn...	female	40.00	1	1	16966	134.5000	E34	C
325	326	1	1	Young, Miss. Marie Grice	female	36.00	0	0	PC 17760	135.6333	C32	C
332	333	0	1	Graham, Mr. George Edward	male	38.00	0	1	PC 17582	153.4625	C91	S
334	335	1	1	Frauenthal, Mrs. Henry William (Clara Heinshei...	female	NaN	1	0	PC 17611	133.6500	NaN	S
337	338	1	1	Burns, Miss. Elizabeth Margaret	female	41.00	0	0	16966	134.5000	E40	C
373	374	0	1	Ringhini, Mr. Sante	male	22.00	0	0	PC 17760	135.6333	NaN	C
390	391	1	1	Carter, Mr. William Ernest	male	36.00	1	2	113760	120.0000	B96 B98	S
393	394	1	1	Newell, Miss. Marjorie	female	23.00	1	0	35273	113.2750	D36	C
435	436	1	1	Carter, Miss. Lucile Polk	female	14.00	1	2	113760	120.0000	B96 B98	S
498	499	0	1	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.00	1	2	113781	151.5500	C22 C26	S
505	506	0	1	Penasco y Castellana, Mr. Victor de Satode	male	18.00	1	0	PC 17758	108.9000	C65	C
537	538	1	1	LeRoy, Miss. Bertha	female	30.00	0	0	PC 17761	106.4250	NaN	C
544	545	0	1	Douglas, Mr. Walter Donald	male	50.00	1	0	PC 17761	106.4250	C86	C
550	551	1	1	Thayer, Mr. John Borland Jr	male	17.00	0	2	17421	110.8833	C70	C
581	582	1	1	Thayer, Mrs. John Borland (Marian Longstreth M...	female	39.00	1	1	17421	110.8833	C68	C
609	610	1	1	Shutes, Miss. Elizabeth	female	40.00	0	0	PC	153.4625	C125	S

609	610	1	1	W	female	40.00	0	0	17582	153.4625	C125	S
659	660	0	1	Newell, Mr. Arthur Webster	male	58.00	0	2	35273	113.2750	D48	C
660	661	1	1	Frauenthal, Dr. Henry William	male	50.00	2	0	PC 17611	133.6500	NaN	S
698	699	0	1	Thayer, Mr. John Borland	male	49.00	1	1	17421	110.8833	C68	C
708	709	1	1	Cleaver, Miss. Alice	female	22.00	0	0	113781	151.5500	NaN	S
763	764	1	1	Carter, Mrs. William Ernest (Lucile Polk)	female	36.00	1	2	113760	120.0000	B96 B98	S
802	803	1	1	Carter, Master. William Thornton II	male	11.00	1	2	113760	120.0000	B96 B98	S
856	857	1	1	Wick, Mrs. George Dennick (Mary Hitchcock)	female	45.00	1	1	36928	164.8667	NaN	S

In [ ]:
```
#records of passengers who age is 20, 28 , 34 or 18

#lengthy and verbose way
condition=(df["Age"] == 20 ) | (df["Age"] == 28 ) | (df["Age"] == 34 ) | (df["Age"] == 18 )
```

In [66]:
```
condition=(df['Age'].isin(  [20,28,34,18]  )   ) #simple and elegant
df[condition]
```

Out[66]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
12	13	0	3	Saundercock, Mr. William Henry	male	20.0	0	0	A/5. 2151	8.0500	NaN	S
21	22	1	2	Beesley, Mr. Lawrence	male	34.0	0	0	248698	13.0000	D56	S
23	24	1	1	Sloper, Mr. William Thompson	male	28.0	0	0	113788	35.5000	A6	S
34	35	0	1	Meyer, Mr. Edgar Joseph	male	28.0	1	0	PC 17604	82.1708	NaN	C
38	39	0	3	Vander Planke, Miss. Augusta Maria	female	18.0	2	0	345764	18.0000	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
848	849	0	2	Harper, Rev. John	male	28.0	0	1	248727	33.0000	NaN	S
855	856	1	3	Aks, Mrs. Sam (Leah Rosen)	female	18.0	0	1	392091	9.3500	NaN	S
874	875	1	2	Abelson, Mrs. Samuel (Hannah Wizosky)	female	28.0	1	0	P/PP 3381	24.0000	NaN	C
876	877	0	3	Gustafsson, Mr. Alfred Ossian	male	20.0	0	0	7534	9.8458	NaN	S
883	884	0	2	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068	10.5000	NaN	S

81 rows × 12 columns

In [ ]:
```
#locate the rows 17 , 89 and 20 and show names of passengers in those rows
```

In [89]:
```
#find out the maximum fare paid by passenger.
#Show record of these passengers(multiple passengers have paid max fare)


condition=( df['Fare'] == df['Fare'].max() )

df[condition][['Name of Passenger']]
```

Out[89]:

	Name of Passenger
258	Ward, Miss. Anna
679	Cardeza, Mr. Thomas Drake Martinez
737	Lesurer, Mr. Gustave J

In [ ]:
```
#find the records of passengers in the age group of 20-50 who paid a fare of over 100 and travelled pclass 2 or 3
```

In [ ]:
```
#find the total fare collected from male passengers in Pclass 3
```

In [ ]:
```
#find the total number of passengers in Pclass 2
```

In [67]:
```
df
```

Out[67]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

In [69]:
```
df.rename(  columns={'Name':  'Name of Passenger'},inplace=True  )
```

In [72]:
```
df.rename(  columns={'Name of Passenger':  'XYZ'}  )
```

Out[72]:

	PassengerId	Survived	Pclass	XYZ	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

In [73]:
```
df
```

Out[73]:

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

886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

In [ ]:
```
#if number of unique categories are less than or equal to 5----> categorical

#more than 5 variations--->continuous
```

In [74]:
```
#how many unique values do you have???

df['Sex'].unique() #categorical
```

Out[74]:
```
array(['male', 'female'], dtype=object)
```

In [76]:
```
df['Pclass'].unique()
```

Out[76]:
```
array([3, 1, 2])
```

In [77]:
```
df['Fare'].unique()
```

Out[77]:
```
array([  7.25  ,  71.2833,   7.925 ,  53.1   ,   8.05  ,   8.4583,
        51.8625,  21.075 ,  11.1333,  30.0708,  16.7   ,  26.55  ,
        31.275 ,   7.8542,  16.    ,  29.125 ,  13.    ,  18.    ,
         7.225 ,  26.    ,   8.0292,  35.5   ,  31.3875, 263.    ,
         7.8792,   7.8958,  27.7208, 146.5208,   7.75  ,  10.5   ,
        82.1708,  52.    ,   7.2292,  11.2417,   9.475 ,  21.    ,
        41.5792,  15.5   ,  21.6792,  17.8   ,  39.6875,   7.8   ,
        76.7292,  61.9792,  27.75  ,  46.9   ,  80.    ,  83.475 ,
        27.9   ,  15.2458,   8.1583,   8.6625,  73.5   ,  14.4542,
        56.4958,   7.65  ,  29.    ,  12.475 ,   9.    ,   9.5   ,
         7.7875,  47.1   ,  15.85  ,  34.375 ,  61.175 ,  20.575 ,
        34.6542,  63.3583,  23.    ,  77.2875,   8.6542,   7.775 ,
        24.15  ,   9.825 ,  14.4583, 247.5208,   7.1417,  22.3583,
         6.975 ,   7.05  ,  14.5   ,  15.0458,  26.2833,   9.2167,
        79.2   ,   6.75  ,  11.5   ,  36.75  ,   7.7958,  12.525 ,
        66.6   ,   7.3125,  61.3792,   7.7333,  69.55  ,  16.1   ,
        15.75  ,  20.525 ,  55.    ,  25.925 ,  33.5   ,  30.6958,
        25.4667,  28.7125,   0.    ,  15.05  ,  39.    ,  22.025 ,
        50.    ,   8.4042,   6.4958,  10.4625,  18.7875,  31.    ,
       113.275 ,  27.    ,  76.2917,  90.    ,   9.35  ,  13.5   ,
         7.55  ,  26.25  ,  12.275 ,   7.125 ,  52.5542,  20.2125,
        86.5   , 512.3292,  79.65  , 153.4625, 135.6333,  19.5   ,
        29.7   ,  77.9583,  20.25  ,  78.85  ,  91.0792,  12.875 ,
         8.85  , 151.55  ,  30.5   ,  23.25  ,  12.35  , 110.8833,
       108.9   ,  24.    ,  56.9292,  83.1583, 262.375 ,  14.    ,
       164.8667, 134.5   ,   6.2375,  57.9792,  28.5   , 133.65  ,
        15.9   ,   9.225 ,  35.    ,  75.25  ,  69.3   ,  55.4417,
       211.5   ,   4.0125, 227.525 ,  15.7417,   7.7292,  12.    ,
       120.    ,  12.65  ,  18.75  ,   6.8583,  32.5   ,   7.875 ,
        14.4   ,  55.9   ,   8.1125,  81.8583,  19.2583,  19.9667,
        89.1042,  38.5   ,   7.725 ,  13.7917,   9.8375,   7.0458,
         7.5208,  12.2875,   9.5875,  49.5042,  78.2667,  15.1   ,
         7.6292,  22.525 ,  26.2875,  59.4   ,   7.4958,  34.0208,
        93.5   , 221.7792, 106.425 ,  49.5   ,  71.    ,  13.8625,
         7.8292,  39.6   ,  17.4   ,  51.4792,  26.3875,  30.    ,
        40.125 ,   8.7125,  15.    ,  33.    ,  42.4   ,  15.55  ,
        65.    ,  32.3208,   7.0542,   8.4333,  25.5875,   9.8417,
         8.1375,  10.1708, 211.3375,  57.    ,  13.4167,   7.7417,
         9.4833,   7.7375,   8.3625,  23.45  ,  25.9292,   8.6833,
         8.5167,   7.8875,  37.0042,   6.45  ,   6.95  ,   8.3   ,
         6.4375,  39.4   ,  14.1083,  13.8583,  50.4958,   5.    ,
         9.8458,  10.5167])
```

In [75]:
```
df['Age'].unique() #continuous
```

Out[75]:
```
array([22.  , 38.  , 26.  , 35.  ,   nan, 54.  ,  2.  , 27.  , 14.  ,
        4.  , 58.  , 20.  , 39.  , 55.  , 31.  , 34.  , 15.  , 28.  ,
        8.  , 19.  , 40.  , 66.  , 42.  , 21.  , 18.  ,  3.  ,  7.  ,
       49.  , 29.  , 65.  , 28.5 ,  5.  , 11.  , 45.  , 17.  , 32.  ,
       16.  , 25.  ,  0.83, 30.  , 33.  , 23.  , 24.  , 46.  , 59.  ,
       71.  , 37.  , 47.  , 14.5 , 70.5 , 32.5 , 12.  ,  9.  , 36.5 ,
```

```
       51.  , 55.5 , 40.5 , 44.  ,  1.  , 61.  , 56.  , 50.  , 36.  ,
       45.5 , 20.5 , 62.  , 41.  , 52.  , 63.  , 23.5 ,  0.92, 43.  ,
       60.  , 10.  , 64.  , 13.  , 48.  ,  0.75, 53.  , 57.  , 80.  ,
       70.  , 24.5 ,  6.  ,  0.67, 30.5 ,  0.42, 34.5 , 74.  ])
```

In [80]:
```
#how many passengers are present in Pclass 1 , 2 and 3?


#categorical attributes

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

Out[80]:
```
3    491
1    216
2    184
Name: Pclass, dtype: int64
```

In [82]:
```
df['Sex'].value_counts()
```

Out[82]:
```
male      577
female    314
Name: Sex, dtype: int64
```

In [86]:
```
df['Sex'].value_counts( normalize=True  )
```

Out[86]:
```
male      0.647587
female    0.352413
Name: Sex, dtype: float64
```

In [91]:
```
#grouping


#data of pclass 1 , 2 and 3, average fare
condition1=(df['Pclass']==1)
condition2=(df['Pclass']==2)
condition3=(df['Pclass']==3)


df[condition1]
```

Out[91]:

	PassengerId	Survived	Pclass	Name of Passenger	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500	C103	S
23	24	1	1	Sloper, Mr. William Thompson	male	28.0	0	0	113788	35.5000	A6	S
...	...	...	...	...	...	...	...	...	...	...	...	...
871	872	1	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0	1	1	11751	52.5542	D35	S
872	873	0	1	Carlsson, Mr. Frans Olof	male	33.0	0	0	695	5.0000	B51 B53 B55	S
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0	1	11767	83.1583	C50	C
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C

216 rows × 12 columns

In [92]:
```
df[condition2]
```

Out[92]:

	PassengerId	Survived	Pclass	Name of Passenger	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	C
15	16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female	55.0	0	0	248706	16.0000	NaN	S
17	18	1	2	Williams, Mr. Charles Eugene	male	NaN	0	0	244373	13.0000	NaN	S
20	21	0	2	Fynney, Mr. Joseph J	male	35.0	0	0	239865	26.0000	NaN	S
21	22	1	2	Beesley, Mr.	male	34.0	0	0	248698	13.0000	D56	S

				Lawrence								
...	...	...	...	...	...	...	...	...	...	...	...	...
866	867	1	2	Duran y More, Miss. Asuncion	female	27.0	1	0	SC/PARIS 2149	13.8583	NaN	C
874	875	1	2	Abelson, Mrs. Samuel (Hannah Wizosky)	female	28.0	1	0	P/PP 3381	24.0000	NaN	C
880	881	1	2	Shelley, Mrs. William (Imanita Parrish Hall)	female	25.0	0	1	230433	26.0000	NaN	S
883	884	0	2	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068	10.5000	NaN	S
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S

184 rows × 12 columns

In [93]:
```
#average fare in each Pclass

df.groupby( ['Pclass']  )[['Fare']].mean()
```

Out[93]:

Pclass	Fare
1	84.154687
2	20.662183
3	13.675550

In [97]:
```
df[['Fare']].mean()
```

Out[97]:
```
Fare    32.204208
dtype: float64
```

In [98]:
```
df['Fare'].mean()
```

Out[98]:
```
32.204207968574636
```

In [94]:
```
#total fare according to gender

df.groupby( ['Sex']  )[['Fare']].sum()
```

Out[94]:

Sex	Fare
female	13966.6628
male	14727.2865

In [101…
```
#reset_index
df.groupby( ['Sex','Pclass']  )[['Fare']].sum().reset_index()
```

Out[101…

	Sex	Pclass	Fare
0	female	1	9975.8250
1	female	2	1669.7292
2	female	3	2321.1086
3	male	1	8201.5875
4	male	2	2132.1125
5	male	3	4393.5865

In [103…
```
temp = df.groupby( ['Sex','Pclass']  ) #multiple filter


temp.groups.keys() #which groups were formed?
```

Out[103…
```
dict_keys([('female', 1), ('female', 2), ('female', 3), ('male', 1), ('male', 2), ('male', 3)])
```

In [110…
```
#see records from a particular group???


temp.get_group(('female', 3)) #equivalent to filtering for female and pclass 2
```

Out[110…

	PassengerId	Survived	Name of Passenger	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
2	3	1	Heikkinen, Miss. Laina	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
			Johnson, Mrs. Oscar W (Elisabeth							

8	9	1	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	27.0	0	2	347742	11.1333	NaN	S
10	11	1	Sandstrom, Miss. Marguerite Rut	4.0	1	1	PP 9549	16.7000	G6	S
14	15	0	Vestrom, Miss. Hulda Amanda Adolfina	14.0	0	0	350406	7.8542	NaN	S
18	19	0	Vander Planke, Mrs. Julius (Emelia Maria Vande...	31.0	1	0	345763	18.0000	NaN	S
...	...	...	...	...	...	...	...	...	...	...
863	864	0	Sage, Miss. Dorothy Edith "Dolly"	NaN	8	2	CA. 2343	69.5500	NaN	S
875	876	1	Najib, Miss. Adele Kiamie "Jane"	15.0	0	0	2667	7.2250	NaN	C
882	883	0	Dahlberg, Miss. Gerda Ulrika	22.0	0	0	7552	10.5167	NaN	S
885	886	0	Rice, Mrs. William (Margaret Norton)	39.0	0	5	382652	29.1250	NaN	Q
888	889	0	Johnston, Miss. Catherine Helen "Carrie"	NaN	1	2	W./C. 6607	23.4500	NaN	S

144 rows × 10 columns

In [106…
```
#number of groups

temp.ngroups
```

Out[106… 6

In [119…
```
temp[['Age','Fare']].agg(  ['min','max']  )
```

Out[119…

		Age		Fare	
		min	**max**	**min**	**max**
Sex	**Pclass**				
female	**1**	2.00	63.0	25.9292	512.3292
	2	2.00	57.0	10.5000	65.0000
	3	0.75	63.0	6.7500	69.5500
male	**1**	0.92	80.0	0.0000	512.3292
	2	0.67	70.0	0.0000	73.5000
	3	0.42	74.0	0.0000	69.5500

In [118…
```
#min fare of females in Pclass 2
temp[['Age','Fare']].agg(  ['min','max']  ).loc[  [('female',2)],   [('Fare','min')]   ]
```

Out[118…

		Fare
		min
Sex	**Pclass**	
female	**2**	10.5

In [120…
```
#min fare of females in Pclass 2

temp.get_group(   ('female', 2)  )[['Fare']].min()
```

Out[120… Fare 10.5
dtype: float64

In [111…
```
#6.75000--------> Minimum Fare paid by Female passengers in Pclass 3


#NOT RECOMMENDED way of obtaining the result

condition=(df['Pclass']==3 ) & (df['Sex']=='female')

df[  condition  ][['Fare']].agg(['min'])
```

Out[111…

	Fare
min	6.75

"""

group data first

extract the relevant column

apply an operation

"""

In [121…
```
#group data by Pclass and Embarked. Find average age in each group

#1 line syntax
df.groupby(  ['Pclass','Embarked']  )[['Age']].mean()
```

Out[121…

		Age
Pclass	**Embarked**	
1	**C**	38.027027
	Q	38.500000
	S	38.152037
2	**C**	22.766667
	Q	43.500000
	S	30.386731
3	**C**	20.741951
	Q	25.937500
	S	25.696552

In [122…
```
temp=df.groupby(  ['Pclass','Embarked']  )


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

Out[122…

		Age
Pclass	**Embarked**	
1	**C**	38.027027
	Q	38.500000
	S	38.152037
2	**C**	22.766667
	Q	43.500000
	S	30.386731
3	**C**	20.741951
	Q	25.937500
	S	25.696552

In [ ]:
```
#find out the minimum, maximum and average fare by each category of Pclass
```

In [123…
```
#find the total fare paid by passengers grouped according to their gender, Pclass and embarked location


df.groupby(['Pclass','Sex','Embarked'])[['Fare']].sum()
```

Out[123…

			Fare
Pclass	**Sex**	**Embarked**	
1	**female**	**C**	4972.5333
		Q	90.0000
		S	4753.2917
	male	**C**	3928.5417
		Q	90.0000
		S	4183.0458
2	**female**	**C**	176.8792
		Q	24.7000
		S	1468.1500

	male	**C**	254.2125
		Q	12.3500
		S	1865.5500
3	**female**	**C**	337.9833
		Q	340.1585
		S	1642.9668
	male	**C**	402.1462
		Q	465.0458
		S	3526.3945

In [133…

```
# 0 to 80


#average fare paid in every age bracket(  0 -10 , 10-20, 20-30,30-40 .......)
import numpy as np #numpy


                                    #start,end,interval

bins=pd.cut(  df['Age'],  np.arange( 0,90,10  ) )

df.groupby(  bins  )[['Fare']].mean()
```

Out[133…

Age	**Fare**
(0, 10]	30.434439
(10, 20]	29.529531
(20, 30]	28.306719
(30, 40]	42.496100
(40, 50]	41.163181
(50, 60]	44.774802
(60, 70]	45.910782
(70, 80]	25.936680

In [134…

```
df
```

Out[134…

	PassengerId	Survived	Pclass	Name of Passenger	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

In [135…

```
#sort this data in ascending order of age


df.sort_values(by=['Age'],ascending=True)
```

Out[135…

	PassengerId	Survived	Pclass	Name of Passenger	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked

803	804	1	3	Thomas, Master. Assad Alexander	male	0.42	0	1	2625	8.5167	NaN	C
755	756	1	2	Hamalainen, Master. Viljo	male	0.67	1	1	250649	14.5000	NaN	S
644	645	1	3	Baclini, Miss. Eugenie	female	0.75	2	1	2666	19.2583	NaN	C
469	470	1	3	Baclini, Miss. Helene Barbara	female	0.75	2	1	2666	19.2583	NaN	C
78	79	1	2	Caldwell, Master. Alden Gates	male	0.83	0	2	248738	29.0000	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
859	860	0	3	Razi, Mr. Raihed	male	NaN	0	0	2629	7.2292	NaN	C
863	864	0	3	Sage, Miss. Dorothy Edith "Dolly"	female	NaN	8	2	CA. 2343	69.5500	NaN	S
868	869	0	3	van Melkebeke, Mr. Philemon	male	NaN	0	0	345777	9.5000	NaN	S
878	879	0	3	Laleff, Mr. Kristo	male	NaN	0	0	349217	7.8958	NaN	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S

891 rows × 12 columns

In [138…
```
#sort this data in descending order of age


#select * from df order by Age Desc
df.sort_values(by=['Age'],ascending=False)
```

Out[138…

	PassengerId	Survived	Pclass	Name of Passenger	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
630	631	1	1	Barkworth, Mr. Algernon Henry Wilson	male	80.0	0	0	27042	30.0000	A23	S
851	852	0	3	Svensson, Mr. Johan	male	74.0	0	0	347060	7.7750	NaN	S
493	494	0	1	Artagaveytia, Mr. Ramon	male	71.0	0	0	PC 17609	49.5042	NaN	C
96	97	0	1	Goldschmidt, Mr. George B	male	71.0	0	0	PC 17754	34.6542	A5	C
116	117	0	3	Connors, Mr. Patrick	male	70.5	0	0	370369	7.7500	NaN	Q
...	...	...	...	...	...	...	...	...	...	...	...	...
859	860	0	3	Razi, Mr. Raihed	male	NaN	0	0	2629	7.2292	NaN	C
863	864	0	3	Sage, Miss. Dorothy Edith "Dolly"	female	NaN	8	2	CA. 2343	69.5500	NaN	S
868	869	0	3	van Melkebeke, Mr. Philemon	male	NaN	0	0	345777	9.5000	NaN	S
878	879	0	3	Laleff, Mr. Kristo	male	NaN	0	0	349217	7.8958	NaN	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S

891 rows × 12 columns

In [145…
```
d1={

    "Name":[ 'Arun',"Ajay","Dhavan", "Leonard", "Jai" ],
    "Age" :[  20, 19, 24, 39, 39  ],
    "Salary": [  18000, 21000, 21000, 17000, 34000  ]
}


df1=pd.DataFrame(  d1 )
```

In [147…
```
df1
```

Out[147…

	Name	Age	Salary
0	Arun	20	18000
1	Ajay	19	21000
2	Dhavan	24	21000
3	Leonard	39	17000
4	Jai	39	34000

In [146…
```
df1.sort_values(by='Age' ) #ascending order
```

Out[146…

	Name	Age	Salary
1	Ajay	19	21000
0	Arun	20	18000
2	Dhavan	24	21000
3	Leonard	39	17000
4	Jai	39	34000

In [148…
```
#based on Names

df1.sort_values(by='Name' ) #alphabetical order
```

Out[148…

	Name	Age	Salary
1	Ajay	19	21000
0	Arun	20	18000
2	Dhavan	24	21000
4	Jai	39	34000
3	Leonard	39	17000

In [149…
```
#based on Age or Salary

df1.sort_values(by=['Age','Salary'] ) #ascending order
```

Out[149…

	Name	Age	Salary
1	Ajay	19	21000
0	Arun	20	18000
2	Dhavan	24	21000
3	Leonard	39	17000
4	Jai	39	34000

In [150…
```
#select * from df1 order by Age Asc, salary DESC;

df1.sort_values(by=['Age','Salary'], ascending=[ True , False ] ) #ascending order of Age and descending order of
```

Out[150…

	Name	Age	Salary
1	Ajay	19	21000
0	Arun	20	18000
2	Dhavan	24	21000
4	Jai	39	34000
3	Leonard	39	17000

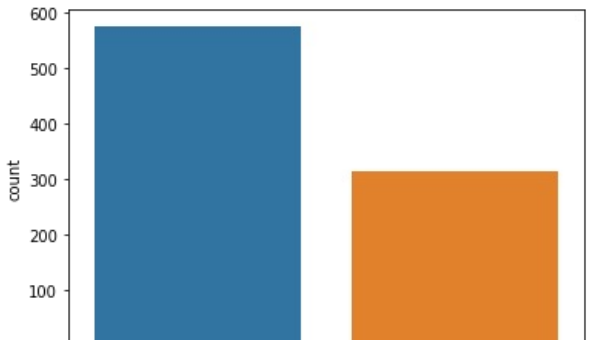
In [151…
```
import seaborn as sns #visual charts , matplotlib
```

In [152…
```
df['Sex'].value_counts()
```

Out[152…
```
male      577
female    314
Name: Sex, dtype: int64
```

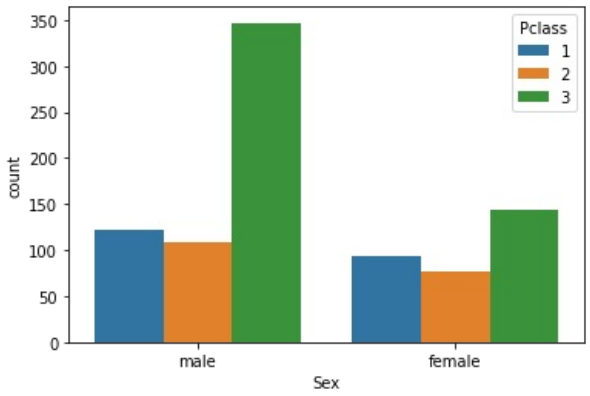
In [153…
```
sns.countplot( x='Sex',data=df  )
```

Out[153…
```
<AxesSubplot:xlabel='Sex', ylabel='count'>
```



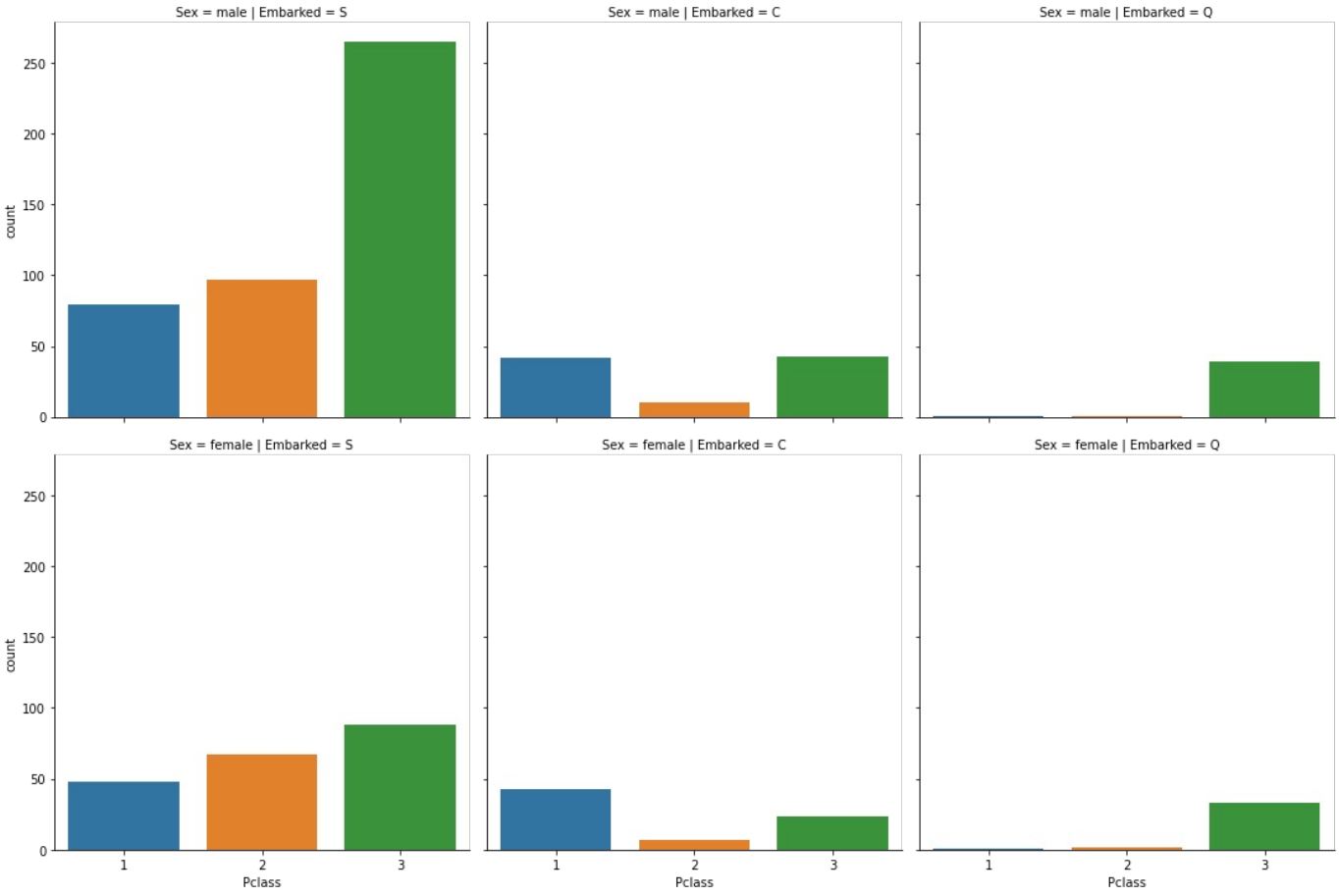
In [154…
```
sns.countplot( x='Sex',data=df,hue='Pclass'  ) #2 parameters
```

Out[154…
```
<AxesSubplot:xlabel='Sex', ylabel='count'>
```



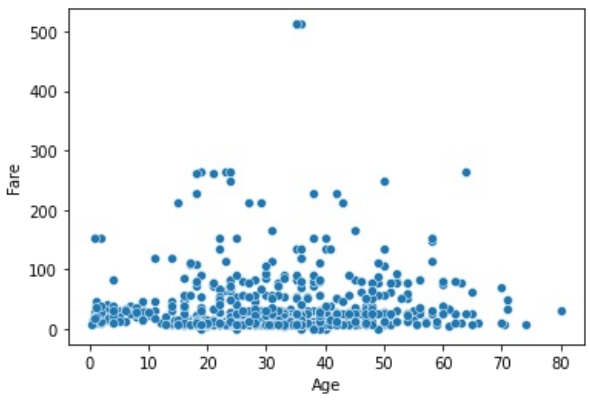
In [157…
```
sns.catplot( row='Sex', col='Embarked',data=df ,kind='count', x='Pclass') #2 parameters
```

Out[157…
```
<seaborn.axisgrid.FacetGrid at 0x7fbaa9e0f760>
```



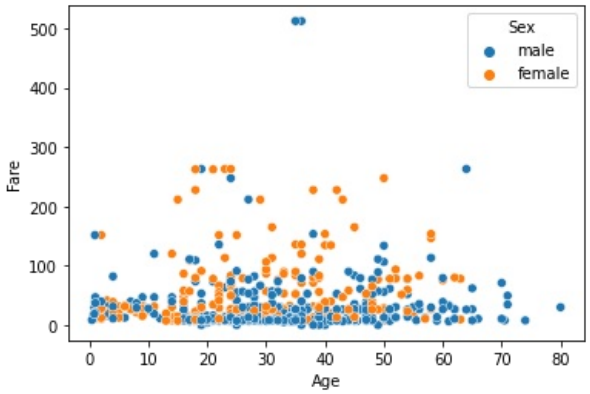
In [160…
```
#relationship between 2 continuous attributes

sns.scatterplot(   x='Age',y='Fare',data=df  )
```

Out[160…
```
<AxesSubplot:xlabel='Age', ylabel='Fare'>
```



In [161…
```
#relationship between 2 continuous attributes

sns.scatterplot(   x='Age',y='Fare',data=df, hue='Sex'  )
```

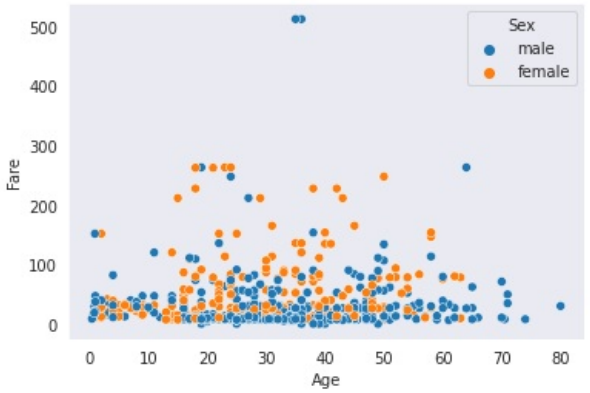
Out[161…
```
<AxesSubplot:xlabel='Age', ylabel='Fare'>
```



In [162…
```
#seaborn global settings

sns.set_style('dark') #darkgrid, dark, white, whitegrid
```

In [163…
```
sns.scatterplot(   x='Age',y='Fare',data=df, hue='Sex'  )
```

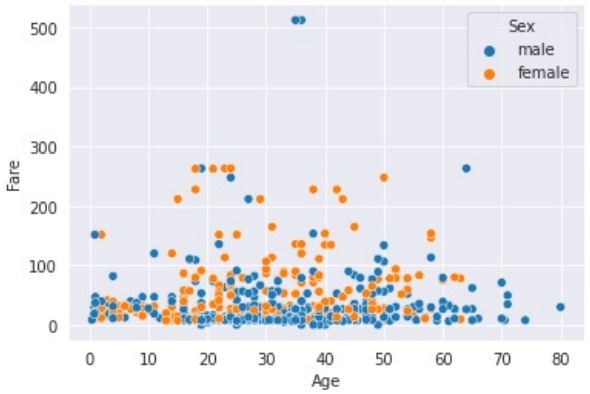
Out[163…
```
<AxesSubplot:xlabel='Age', ylabel='Fare'>
```



In [164…
```
sns.set_style('darkgrid') #darkgrid, dark, white, whitegrid
```

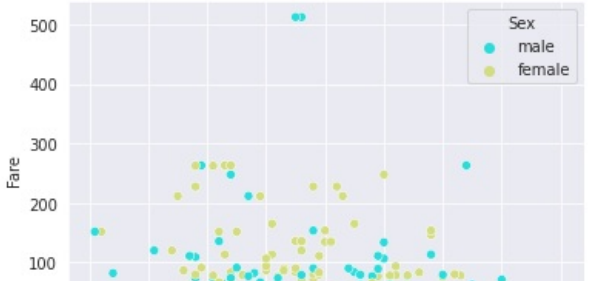
In [165…
```
sns.scatterplot(   x='Age',y='Fare',data=df, hue='Sex'  )
```

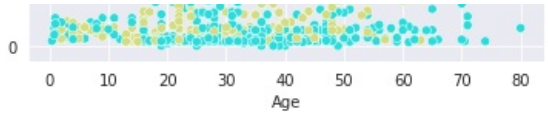
Out[165…
```
<AxesSubplot:xlabel='Age', ylabel='Fare'>
```



In [175…
```
sns.scatterplot(   x='Age',y='Fare',data=df, hue='Sex' ,palette='rainbow' )
```

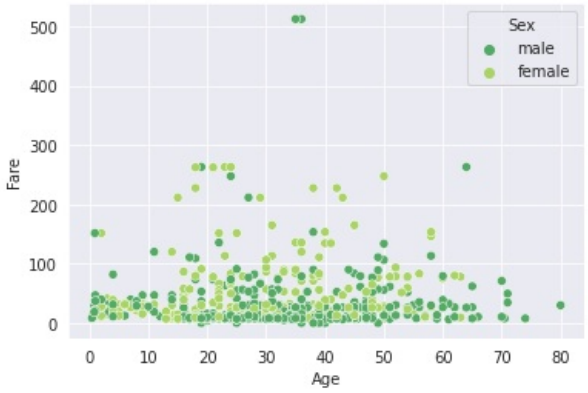
Out[175…
```
<AxesSubplot:xlabel='Age', ylabel='Fare'>
```





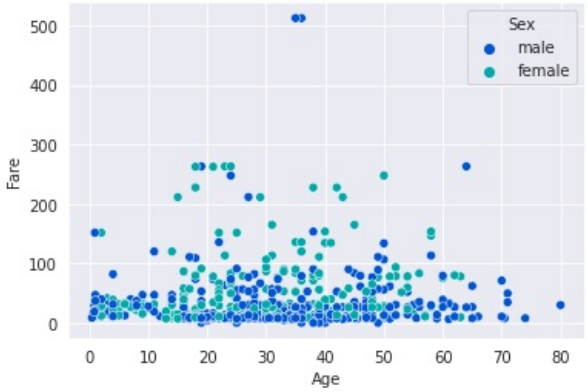
In [168…
```
sns.scatterplot(   x='Age',y='Fare',data=df, hue='Sex' ,palette='summer' )
```

Out[168… <AxesSubplot:xlabel='Age', ylabel='Fare'>



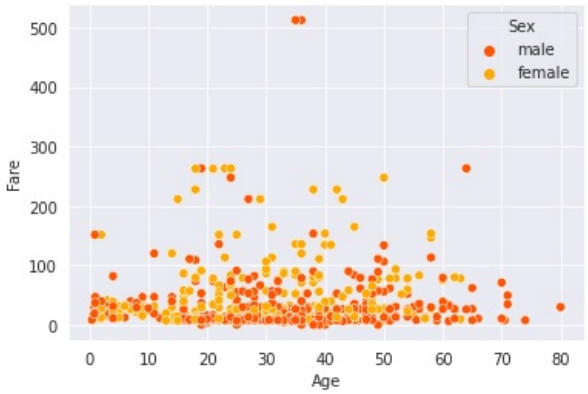
In [169…
```
sns.scatterplot(   x='Age',y='Fare',data=df, hue='Sex' ,palette='winter' )
```

Out[169… <AxesSubplot:xlabel='Age', ylabel='Fare'>



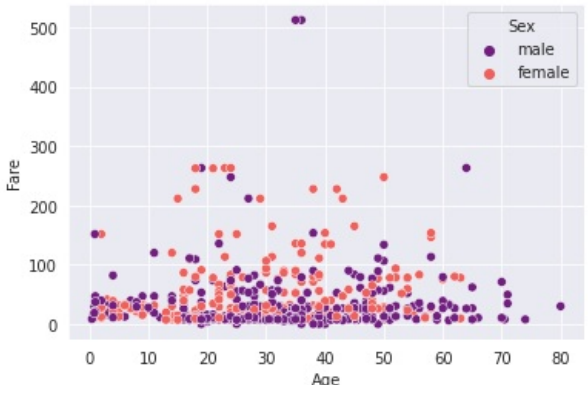
In [170…
```
sns.scatterplot(   x='Age',y='Fare',data=df, hue='Sex' ,palette='autumn' )
```

Out[170… <AxesSubplot:xlabel='Age', ylabel='Fare'>



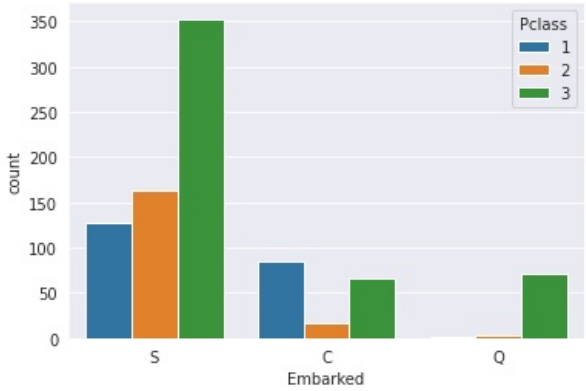
In [171…
```
sns.scatterplot(   x='Age',y='Fare',data=df, hue='Sex' ,palette='magma' )
```

Out[171… <AxesSubplot:xlabel='Age', ylabel='Fare'>



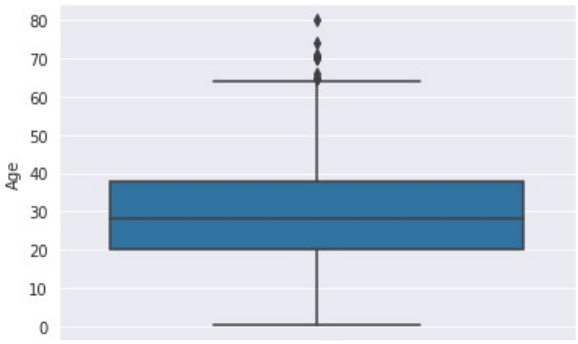
In [176…
```
#please generate a graph

#create a graph of count of passengers categorized by Embarked location distributed by Pclass!
sns.countplot(x='Embarked',data=df,hue='Pclass')
```

Out[176… <AxesSubplot:xlabel='Embarked', ylabel='count'>



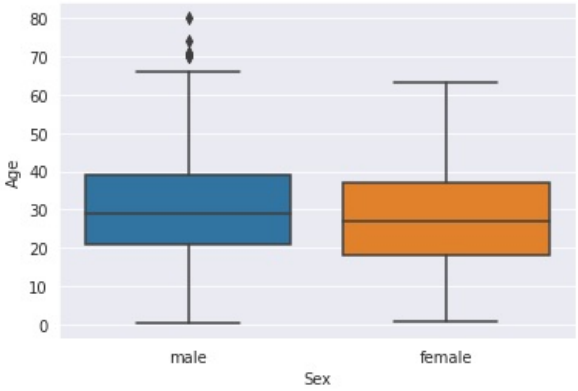
In [179…
```
#distribution of data in terms of percentiles or quantiles


#boxplot! (box or whisker plot)

#continuous
sns.boxplot( y='Age',data=df ) #IQR???
```

Out[179… <AxesSubplot:ylabel='Age'>



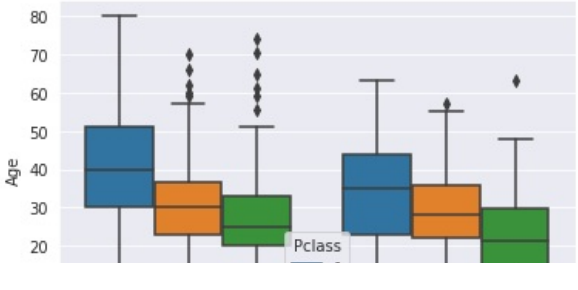
In [180…
```
#one continuous & one categorical attribute
sns.boxplot( y='Age',x='Sex',data=df )
```

Out[180… <AxesSubplot:xlabel='Sex', ylabel='Age'>



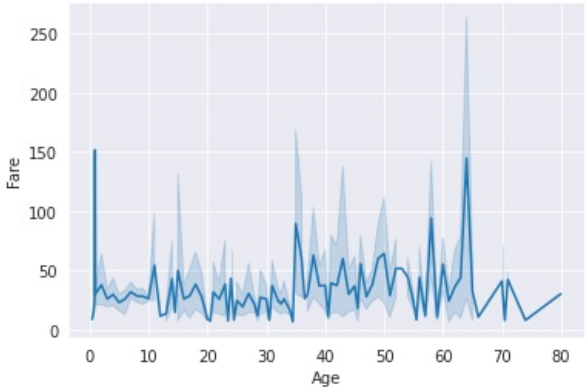
In [181…
```
#one continuous & one categorical attribute
sns.boxplot( y='Age',x='Sex',data=df,hue='Pclass' )
```

Out[181… <AxesSubplot:xlabel='Sex', ylabel='Age'>



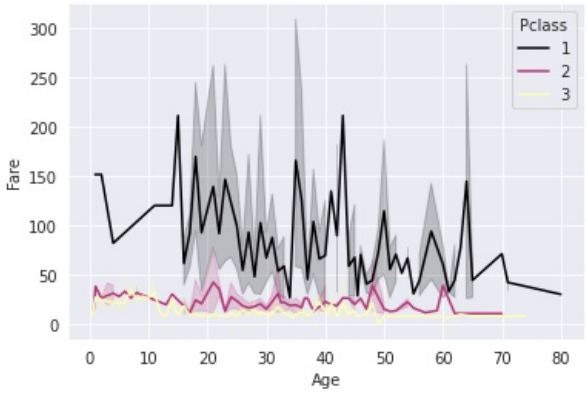
In [183…
```
sns.lineplot(x='Age',y='Fare',data=df) #timeseries data(financial domain)
```

Out[183…
```
<AxesSubplot:xlabel='Age', ylabel='Fare'>
```



In [185…
```
sns.lineplot(x='Age',y='Fare',data=df,hue='Pclass',palette='magma') #timeseries data(financial domain)
```

Out[185…
```
<AxesSubplot:xlabel='Age', ylabel='Fare'>
```



In [187…
```
temp=pd.read_csv('/home/harshit/DataSets/YESBANK.csv')
temp
```

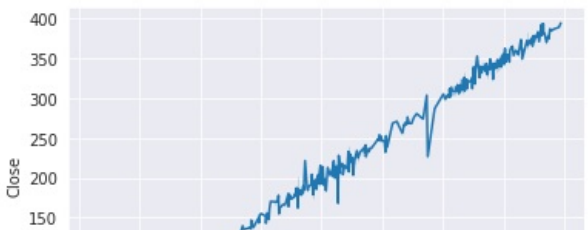
Out[187…

	Date	Open	High	Low	Close	Adj Close	Volume
0	2017-12-11	313.500000	315.799988	310.600006	311.600006	300.880615	4416465.0
1	2017-12-12	312.000000	312.000000	305.899994	306.799988	296.245758	5457103.0
2	2017-12-13	306.350006	307.350006	301.049988	301.899994	291.514282	6911856.0
3	2017-12-14	303.899994	304.649994	301.750000	303.899994	293.445526	4904177.0
4	2017-12-15	307.000000	317.450012	307.000000	315.899994	305.032715	20571225.0
...	...	...	...	...	...	...	...
733	2020-12-02	15.700000	15.900000	14.850000	15.450000	15.450000	311349886.0
734	2020-12-03	15.650000	15.800000	15.250000	15.450000	15.450000	152445535.0
735	2020-12-04	15.600000	15.600000	15.050000	15.350000	15.350000	149691622.0
736	2020-12-07	15.650000	15.850000	15.500000	15.750000	15.750000	193242183.0
737	2020-12-08	16.000000	17.299999	16.000000	17.299999	17.299999	562741066.0

738 rows × 7 columns

In [188…
```
sns.lineplot(x='Open',y='Close',data=temp)
```

Out[188…
```
<AxesSubplot:xlabel='Open', ylabel='Close'>
```




100
50
0
0
50
100
150
200
250
300
350
400
Open