

Titanic Data Analysis

December 5, 2021

```
[2]: import pandas as pd
df=pd.read_csv("https://raw.githubusercontent.com/catamaican/AED/main/titanic.
↪csv")
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
```

```
[3]: df.head(100)
```

```
[3]:   PassengerId  Survived  Pclass  \
0             1         0       3
1             2         1       1
2             3         1       3
3             4         1       1
4             5         0       3
..          ...         ...     ...
95           96         0       3
96           97         0       1
97           98         1       1
98           99         1       2
```

```
99          100          0          2
```

	Name	Sex	Age	SibSp	\
0	Braund, Mr. Owen Harris	male	22.0	1	
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	
2	Heikkinen, Miss. Laina	female	26.0	0	
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	
4	Allen, Mr. William Henry	male	35.0	0	
..		
95	Shorney, Mr. Charles Joseph	male	NaN	0	
96	Goldschmidt, Mr. George B	male	71.0	0	
97	Greenfield, Mr. William Bertram	male	23.0	0	
98	Doling, Mrs. John T (Ada Julia Bone)	female	34.0	0	
99	Kantor, Mr. Sinai	male	34.0	1	

	Parch	Ticket	Fare	Cabin	Embarked
0	0	A/5 21171	7.2500	NaN	S
1	0	PC 17599	71.2833	C85	C
2	0	STON/O2. 3101282	7.9250	NaN	S
3	0	113803	53.1000	C123	S
4	0	373450	8.0500	NaN	S
..	
95	0	374910	8.0500	NaN	S
96	0	PC 17754	34.6542	A5	C
97	1	PC 17759	63.3583	D10 D12	C
98	1	231919	23.0000	NaN	S
99	0	244367	26.0000	NaN	S

[100 rows x 12 columns]

```
[4]: print(df)
```

	PassengerId	Survived	Pclass	\
0	1	0	3	
1	2	1	1	
2	3	1	3	
3	4	1	1	
4	5	0	3	
..	
886	887	0	2	
887	888	1	1	
888	889	0	3	
889	890	1	1	
890	891	0	3	

	Name	Sex	Age	SibSp	\
0	Braund, Mr. Owen Harris	male	22.0	1	
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	

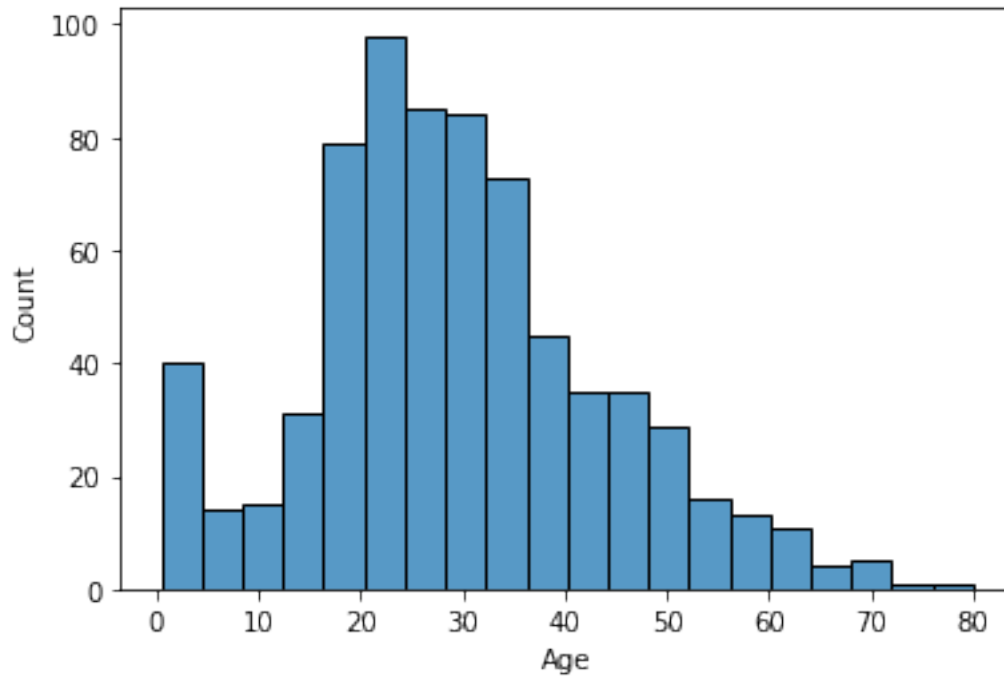
2		Heikkinen, Miss. Laina	female	26.0	0
3		Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1
4		Allen, Mr. William Henry	male	35.0	0
..	
886		Montvila, Rev. Juozas	male	27.0	0
887		Graham, Miss. Margaret Edith	female	19.0	0
888		Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1
889		Behr, Mr. Karl Howell	male	26.0	0
890		Dooley, Mr. Patrick	male	32.0	0

	Parch		Ticket	Fare	Cabin	Embarked
0	0		A/5 21171	7.2500	NaN	S
1	0		PC 17599	71.2833	C85	C
2	0	STON/02.	3101282	7.9250	NaN	S
3	0		113803	53.1000	C123	S
4	0		373450	8.0500	NaN	S
..
886	0		211536	13.0000	NaN	S
887	0		112053	30.0000	B42	S
888	2	W./C.	6607	23.4500	NaN	S
889	0		111369	30.0000	C148	C
890	0		370376	7.7500	NaN	Q

[891 rows x 12 columns]

```
[57]: import seaborn as sms
      varsta=df["Age"]
      sms.histplot(varsta)
```

```
[57]: <AxesSubplot:xlabel='Age', ylabel='Count'>
```



```
[6]: df.describe()
```

```
[6]:
```

	PassengerId	Survived	Pclass	Age	SibSp	\
count	891.000000	891.000000	891.000000	714.000000	891.000000	
mean	446.000000	0.383838	2.308642	29.699118	0.523008	
std	257.353842	0.486592	0.836071	14.526497	1.102743	
min	1.000000	0.000000	1.000000	0.420000	0.000000	
25%	223.500000	0.000000	2.000000	20.125000	0.000000	
50%	446.000000	0.000000	3.000000	28.000000	0.000000	
75%	668.500000	1.000000	3.000000	38.000000	1.000000	
max	891.000000	1.000000	3.000000	80.000000	8.000000	

	Parch	Fare
count	891.000000	891.000000
mean	0.381594	32.204208
std	0.806057	49.693429
min	0.000000	0.000000
25%	0.000000	7.910400
50%	0.000000	14.454200
75%	0.000000	31.000000
max	6.000000	512.329200

```
[7]: df.sort_values(by=['Survived'])
df.head(100)
```

```
[7]: PassengerId  Survived  Pclass  \
0           1         0         3
1           2         1         1
2           3         1         3
3           4         1         1
4           5         0         3
..          ...         ...         ...
95          96         0         3
96          97         0         1
97          98         1         1
98          99         1         2
99         100         0         2
```

```

                                Name    Sex  Age  SibSp  \
0                        Braund, Mr. Owen Harris    male  22.0    1
1  Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0    1
2                        Heikkinen, Miss. Laina    female  26.0    0
3  Futrelle, Mrs. Jacques Heath (Lily May Peel)    female  35.0    1
4                        Allen, Mr. William Henry    male  35.0    0
..          ...         ...         ...         ...
95                        Shorney, Mr. Charles Joseph    male   NaN    0
96                        Goldschmidt, Mr. George B    male  71.0    0
97                        Greenfield, Mr. William Bertram    male  23.0    0
98      Doling, Mrs. John T (Ada Julia Bone)    female  34.0    0
99                        Kantor, Mr. Sinai    male  34.0    1
```

```

Parch    Ticket    Fare    Cabin Embarked
0      0    A/5 21171   7.2500   NaN      S
1      0    PC 17599  71.2833   C85      C
2      0  STON/O2. 3101282   7.9250   NaN      S
3      0    113803  53.1000  C123      S
4      0    373450   8.0500   NaN      S
..     ...         ...         ...         ...
95     0    374910   8.0500   NaN      S
96     0    PC 17754  34.6542   A5      C
97     1    PC 17759  63.3583  D10 D12    C
98     1    231919  23.0000   NaN      S
99     0    244367  26.0000   NaN      S
```

[100 rows x 12 columns]

```
[8]: df[['Pclass', 'Survived']].groupby("Pclass").count()
```

```
[8]:      Survived
Pclass
1         216
2         184
```

3 491

```
[28]: df[['Pclass', 'Embarked']].groupby("Embarked").count()
```

```
[28]:
```

	Pclass
Embarked	
C	168
Q	77
S	644

```
[29]: df[['Pclass', 'Embarked']].groupby("Pclass").count()
```

```
[29]:
```

	Embarked
Pclass	
1	214
2	184
3	491

```
[30]: df[['Age', 'Survived', 'Pclass']].groupby("Survived").max()
```

```
[30]:
```

	Age	Pclass
Survived		
0	74.0	3
1	80.0	3

```
[31]: df[['Age', 'Survived', 'Pclass']].groupby("Survived").min()
```

```
[31]:
```

	Age	Pclass
Survived		
0	1.00	1
1	0.42	1

```
[32]: df[['Sex', 'Survived']].groupby("Sex").mean()
```

```
[32]:
```

	Survived
Sex	
female	0.742038
male	0.188908

```
[33]: df[['Age', 'Pclass', 'SibSp']].groupby("Age").mean()
```

```
[33]:
```

	Pclass	SibSp
Age		
0.42	3.0	0.0
0.67	2.0	1.0
0.75	3.0	2.0
0.83	2.0	0.5
0.92	1.0	1.0

```

...      ...      ...
70.00      1.5      0.5
70.50      3.0      0.0
71.00      1.0      0.0
74.00      3.0      0.0
80.00      1.0      0.0

```

[88 rows x 2 columns]

```
[34]: df[['SibSp', 'Survived']].groupby("SibSp").mean()
```

```
[34]:      Survived
SibSp
0      0.345395
1      0.535885
2      0.464286
3      0.250000
4      0.166667
5      0.000000
8      0.000000

```

```
[35]: df[['Age', 'Parch', 'Survived']].groupby("Parch").mean()
```

```
[35]:      Age  Survived
Parch
0      32.178503  0.343658
1      24.422000  0.550847
2      17.216912  0.500000
3      33.200000  0.600000
4      44.500000  0.000000
5      39.200000  0.200000
6      43.000000  0.000000

```

```
[36]: corelatie=df.corr()
corelatie
```

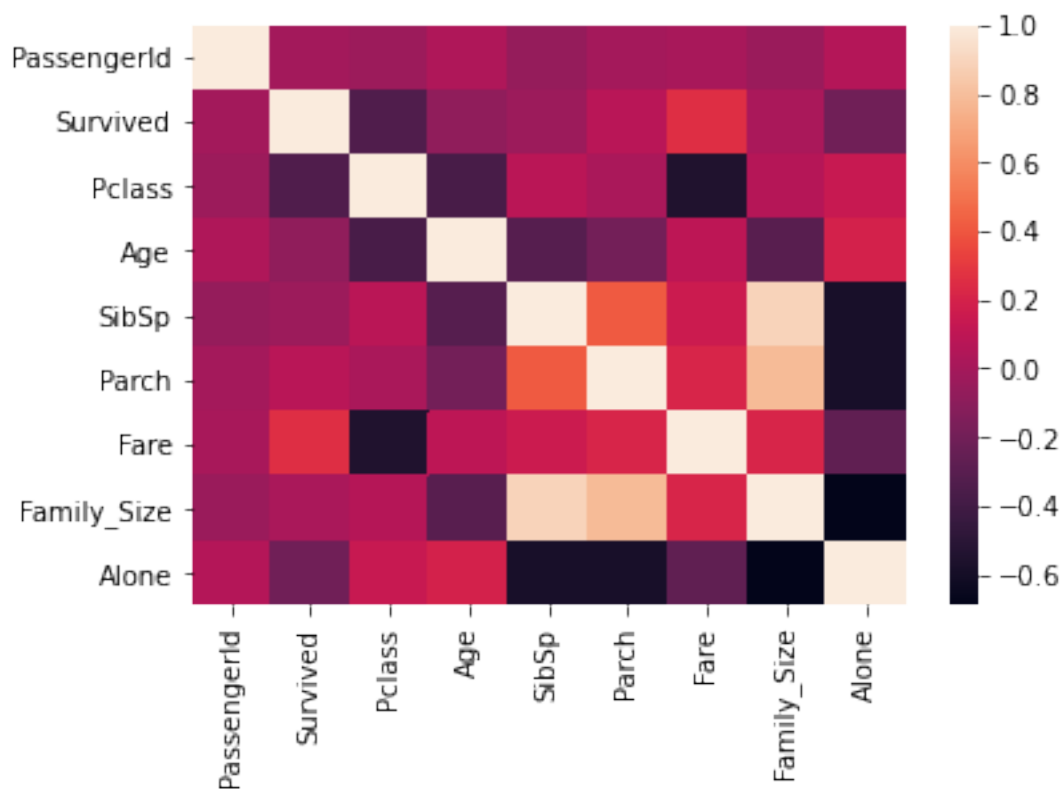
```
[36]:      PassengerId  Survived  Pclass     Age     SibSp  Parch  \
PassengerId      1.000000 -0.005007 -0.035144  0.036847 -0.057527 -0.001652
Survived          -0.005007  1.000000 -0.338481 -0.077221 -0.035322  0.081629
Pclass            -0.035144 -0.338481  1.000000 -0.369226  0.083081  0.018443
Age                0.036847 -0.077221 -0.369226  1.000000 -0.308247 -0.189119
SibSp             -0.057527 -0.035322  0.083081 -0.308247  1.000000  0.414838
Parch             -0.001652  0.081629  0.018443 -0.189119  0.414838  1.000000
Fare              0.012658  0.257307 -0.549500  0.096067  0.159651  0.216225
Family_Size      -0.040143  0.016639  0.065997 -0.301914  0.890712  0.783111
Alone             0.057462 -0.203367  0.135207  0.198270 -0.584471 -0.583398

```

	Fare	Family_Size	Alone
PassengerId	0.012658	-0.040143	0.057462
Survived	0.257307	0.016639	-0.203367
Pclass	-0.549500	0.065997	0.135207
Age	0.096067	-0.301914	0.198270
SibSp	0.159651	0.890712	-0.584471
Parch	0.216225	0.783111	-0.583398
Fare	1.000000	0.217138	-0.271832
Family_Size	0.217138	1.000000	-0.690922
Alone	-0.271832	-0.690922	1.000000

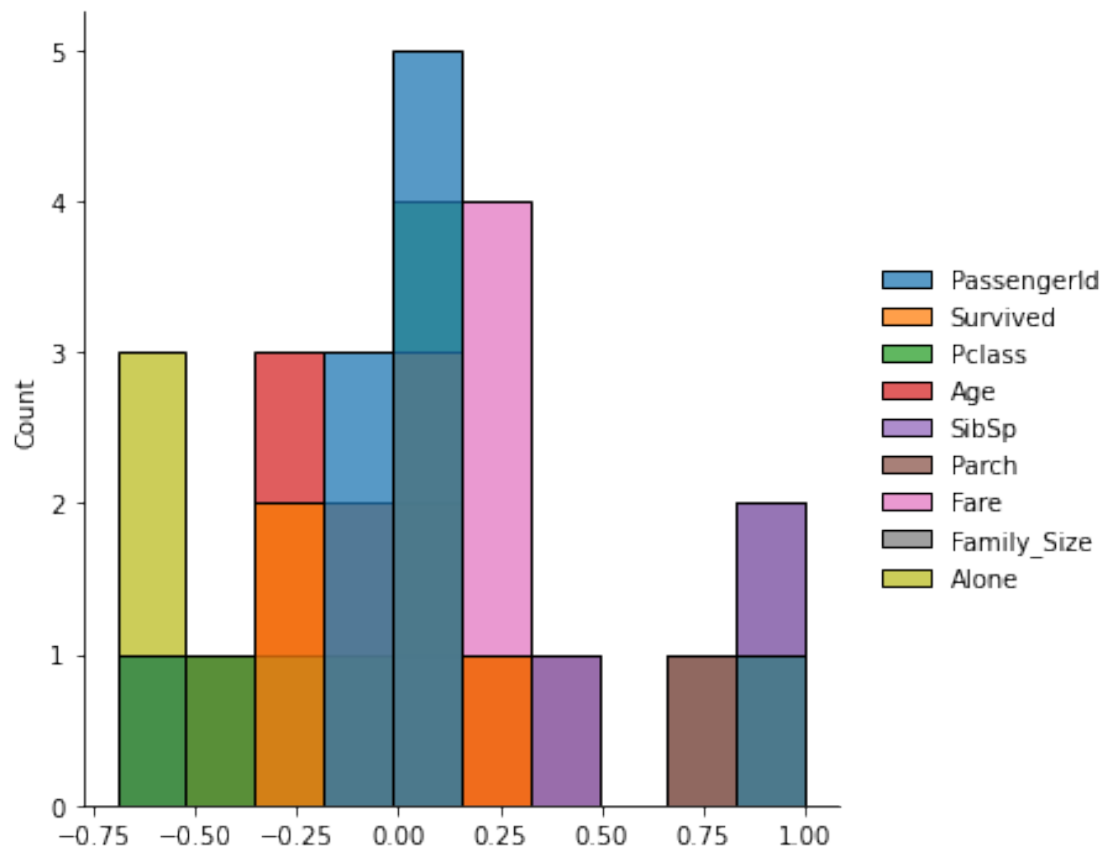
```
[37]: sms.heatmap(corelatie)
```

```
[37]: <AxesSubplot:>
```



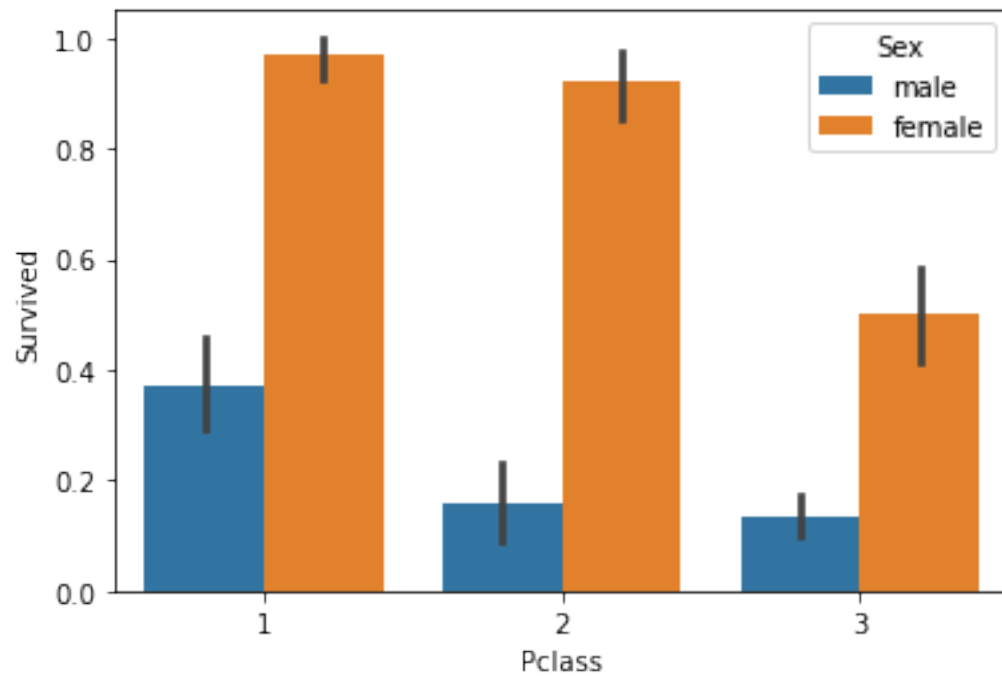
```
[38]: sms.displot(corelatie)
```

```
[38]: <seaborn.axisgrid.FacetGrid at 0x2c81bb47280>
```

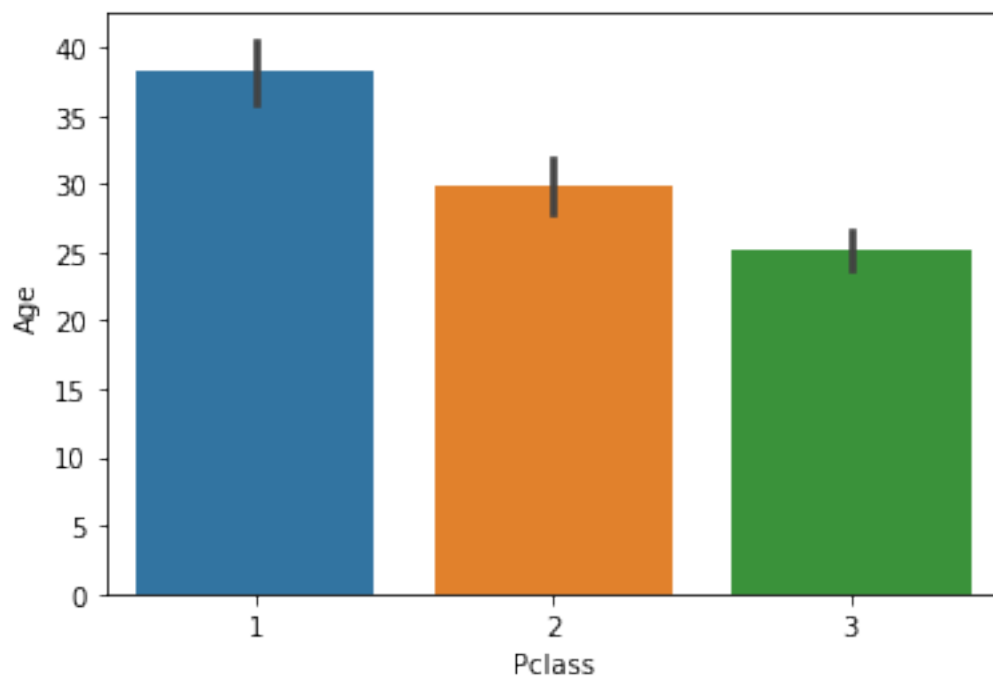
```
[39]: sms.barplot(x=df["Pclass"],y=df["Survived"],hue=df["Sex"])
```

```
[39]: <AxesSubplot:xlabel='Pclass', ylabel='Survived'>
```



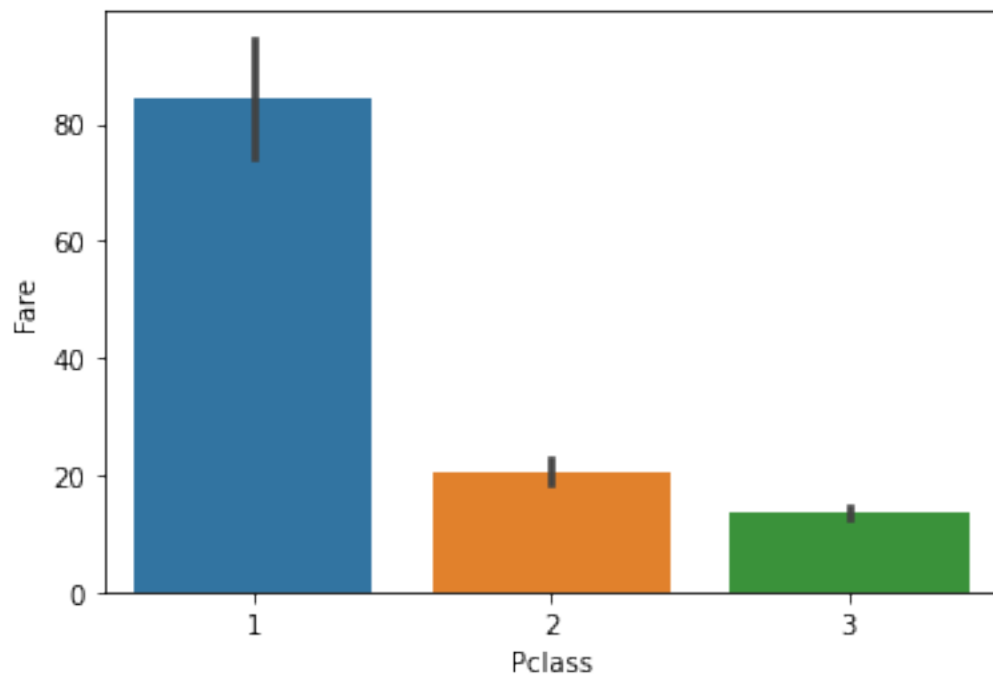
```
[40]: sms.barplot(x=df["Pclass"],y=df["Age"])
```

```
[40]: <AxesSubplot:xlabel='Pclass', ylabel='Age'>
```



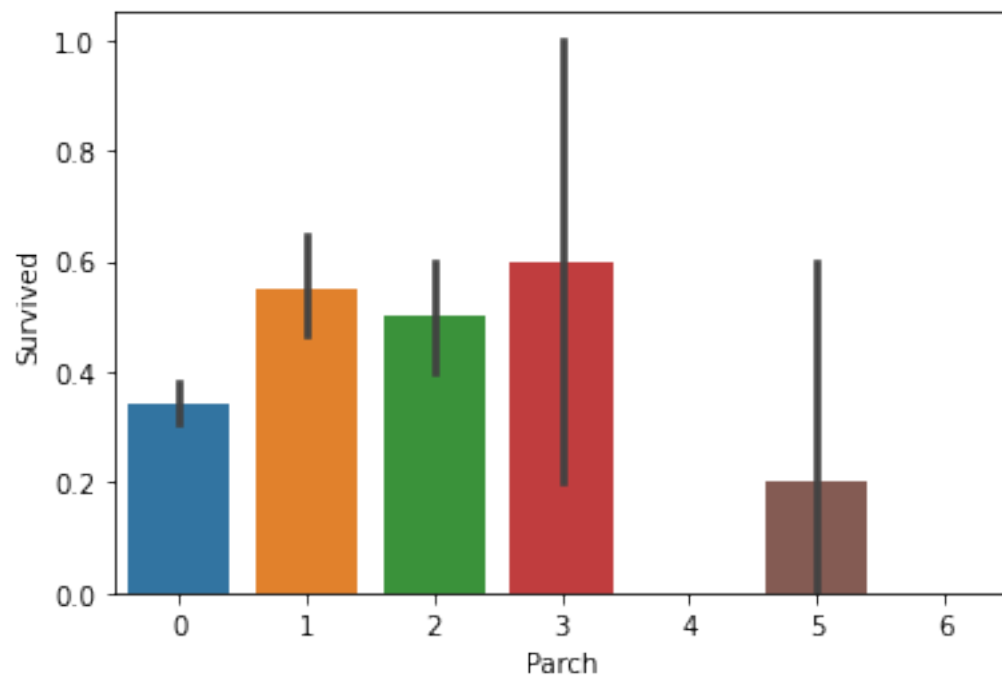
```
[41]: sms.barplot(x=df["Pclass"],y=df["Fare"])
```

```
[41]: <AxesSubplot:xlabel='Pclass', ylabel='Fare'>
```



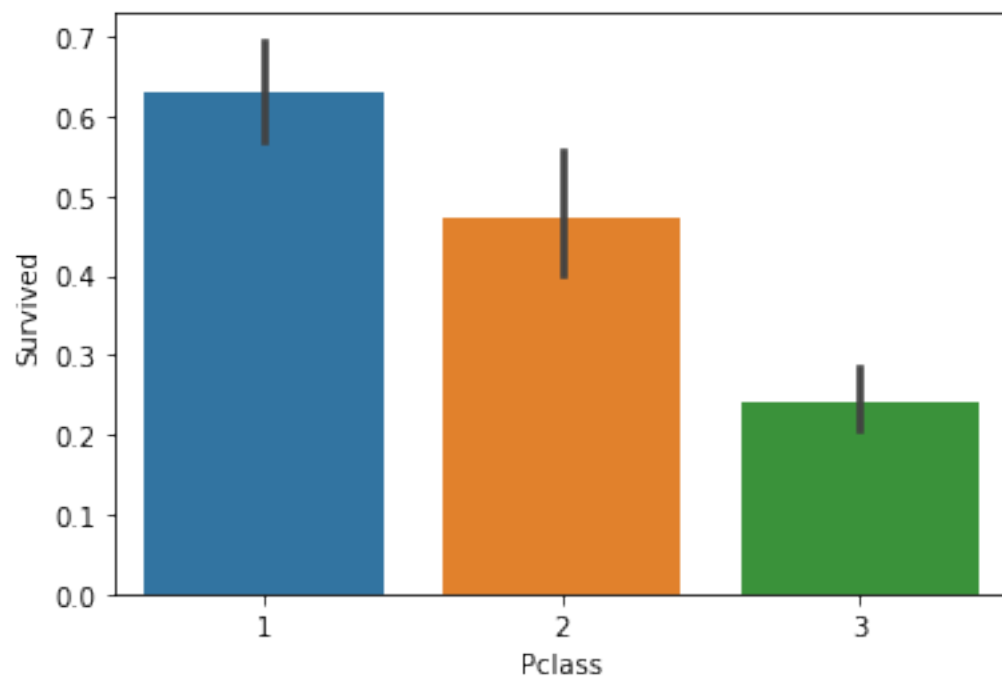
```
[42]: sms.barplot(x=df["Parch"],y=df["Survived"])
```

```
[42]: <AxesSubplot:xlabel='Parch', ylabel='Survived'>
```



```
[43]: sms.barplot(x=df["Pclass"], y=df["Survived"])
```

```
[43]: <AxesSubplot:xlabel='Pclass', ylabel='Survived'>
```

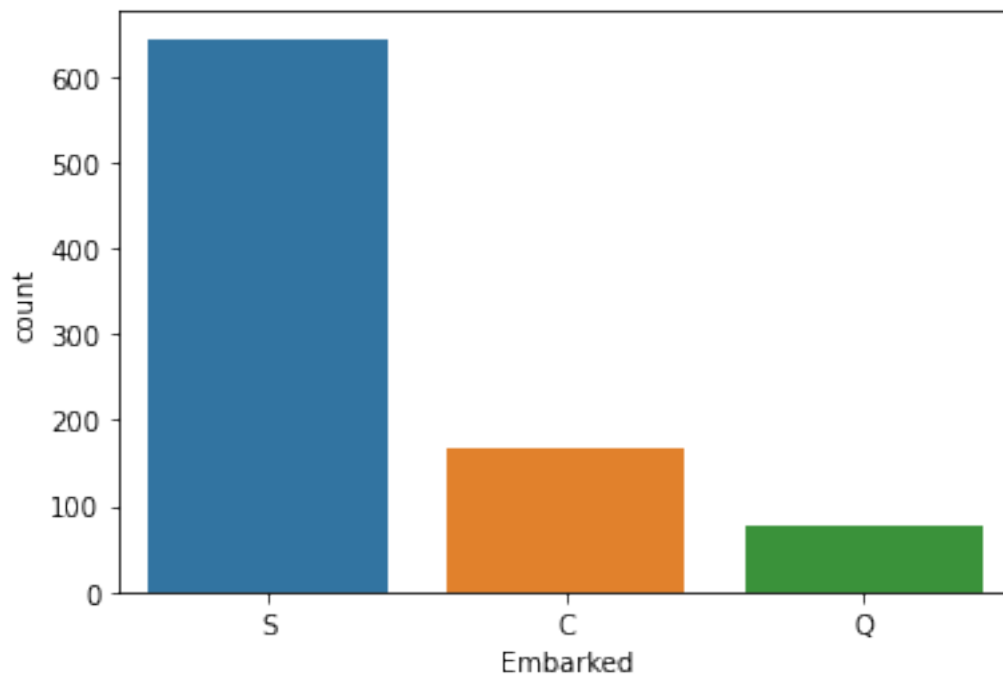


```
[44]: dfSurvived_Calc=df["Survived"]  
dfSurvived_Calc
```

```
[44]: 0      0  
      1      1  
      2      1  
      3      1  
      4      0  
      ..  
     886      0  
     887      1  
     888      0  
     889      1  
     890      0  
      Name: Survived, Length: 891, dtype: int64
```

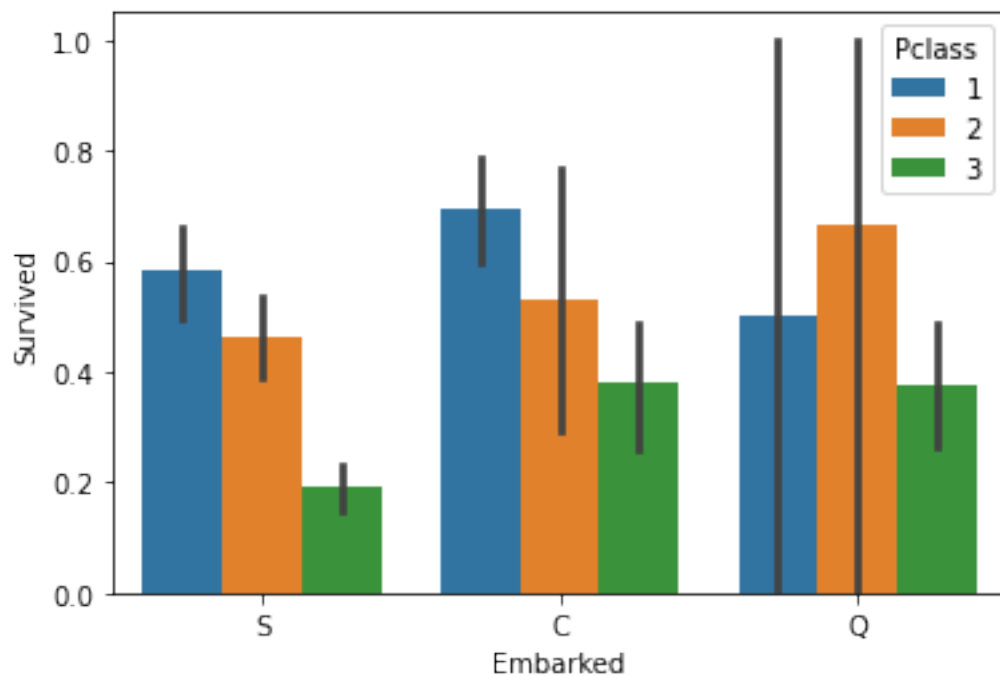
```
[45]: sms.countplot(x=df["Embarked"])
```

```
[45]: <AxesSubplot:xlabel='Embarked', ylabel='count'>
```



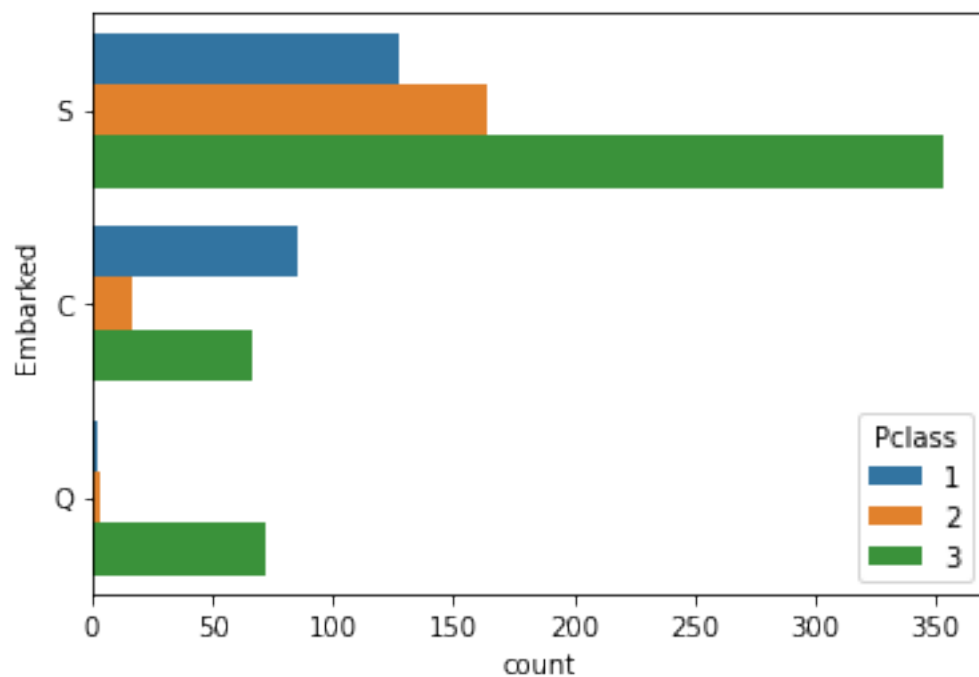
```
[46]: sms.barplot(x=df["Embarked"],y=df["Survived"],hue=df["Pclass"])
```

```
[46]: <AxesSubplot:xlabel='Embarked', ylabel='Survived'>
```



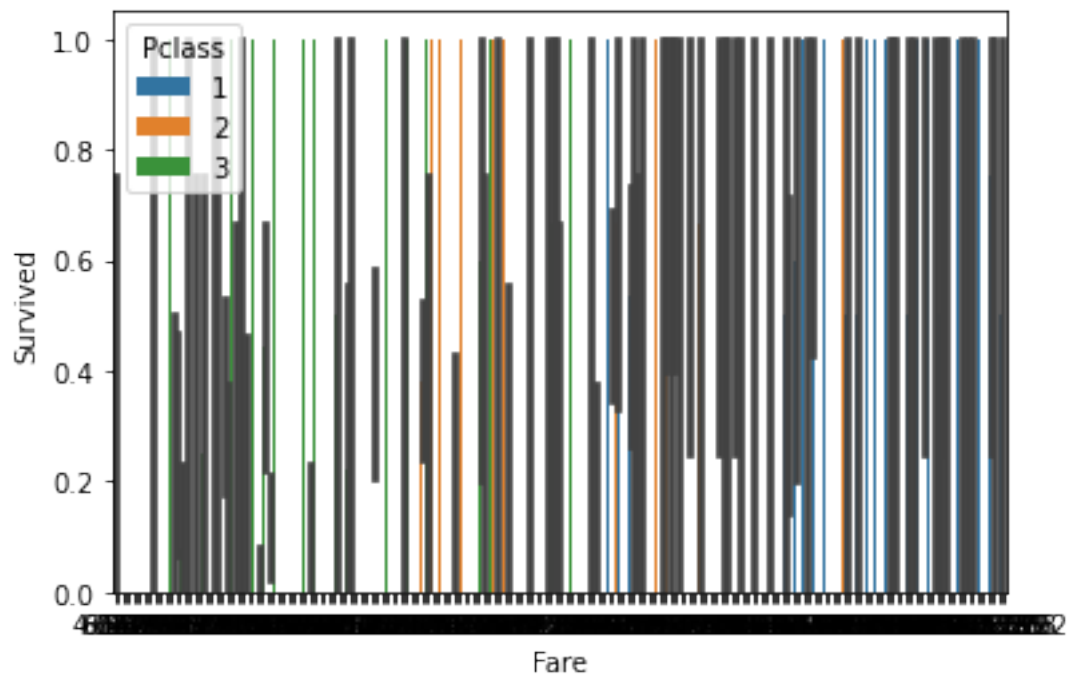
```
[47]: sms.countplot(y=df["Embarked"],hue=df["Pclass"])
```

```
[47]: <AxesSubplot:xlabel='count', ylabel='Embarked'>
```



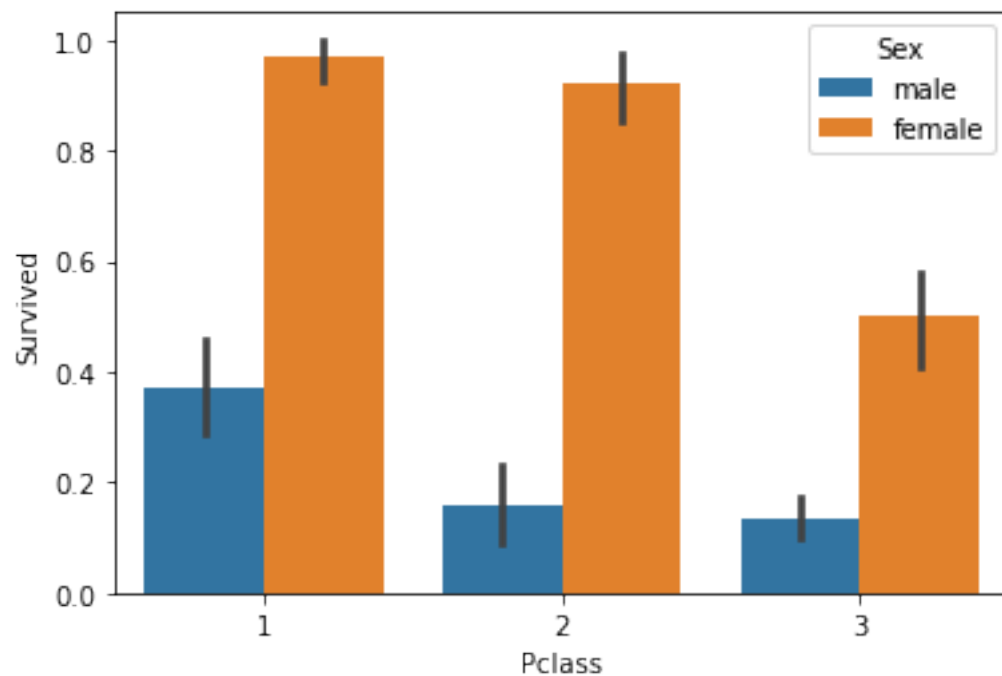
```
[48]: sms.barplot(x=df["Fare"],y=df["Survived"],hue=df["Pclass"])
```

```
[48]: <AxesSubplot:xlabel='Fare', ylabel='Survived'>
```



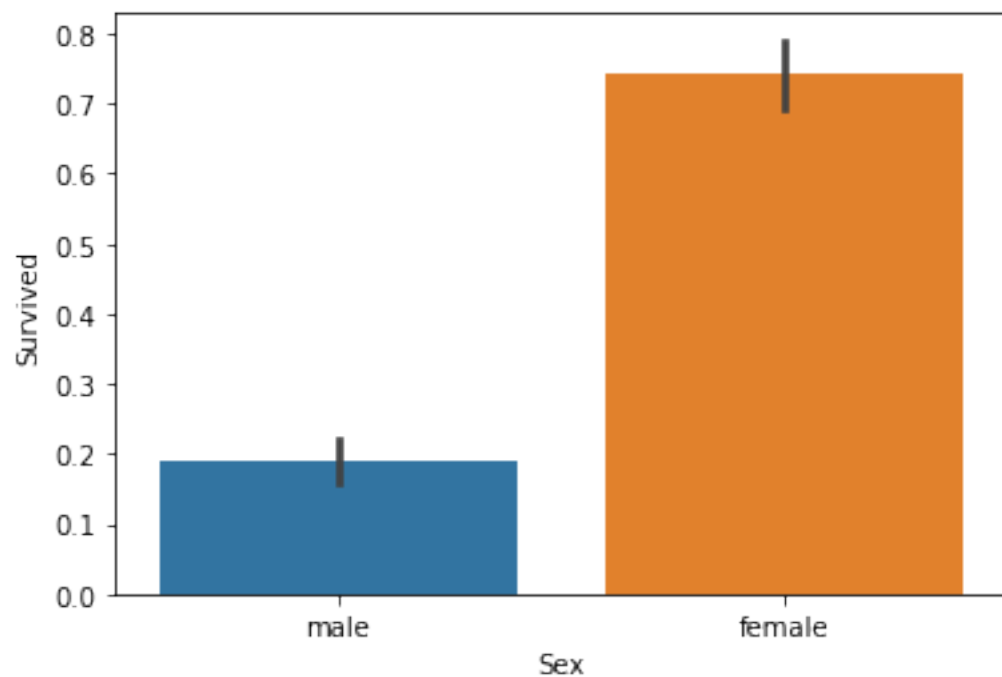
```
[49]: sms.barplot(x=df["Pclass"],y=df["Survived"],hue=df["Sex"])
```

```
[49]: <AxesSubplot:xlabel='Pclass', ylabel='Survived'>
```



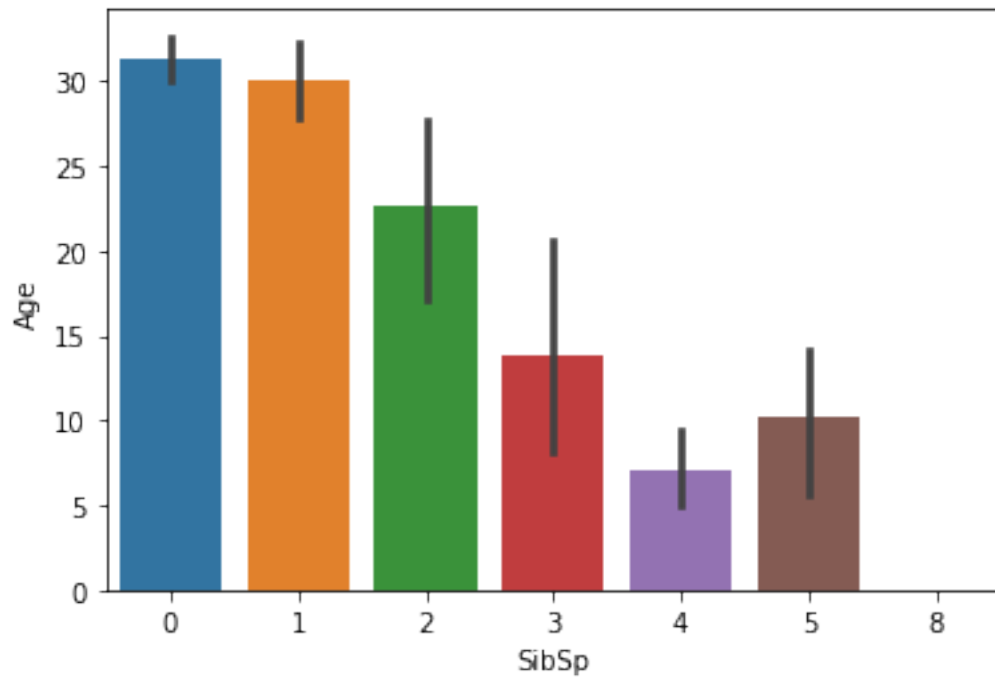
```
[50]: sms.barplot(x=df["Sex"],y=df["Survived"])
```

```
[50]: <AxesSubplot:xlabel='Sex', ylabel='Survived'>
```




```
[51]: sms.barplot(x=df["SibSp"],y=df["Age"])
```

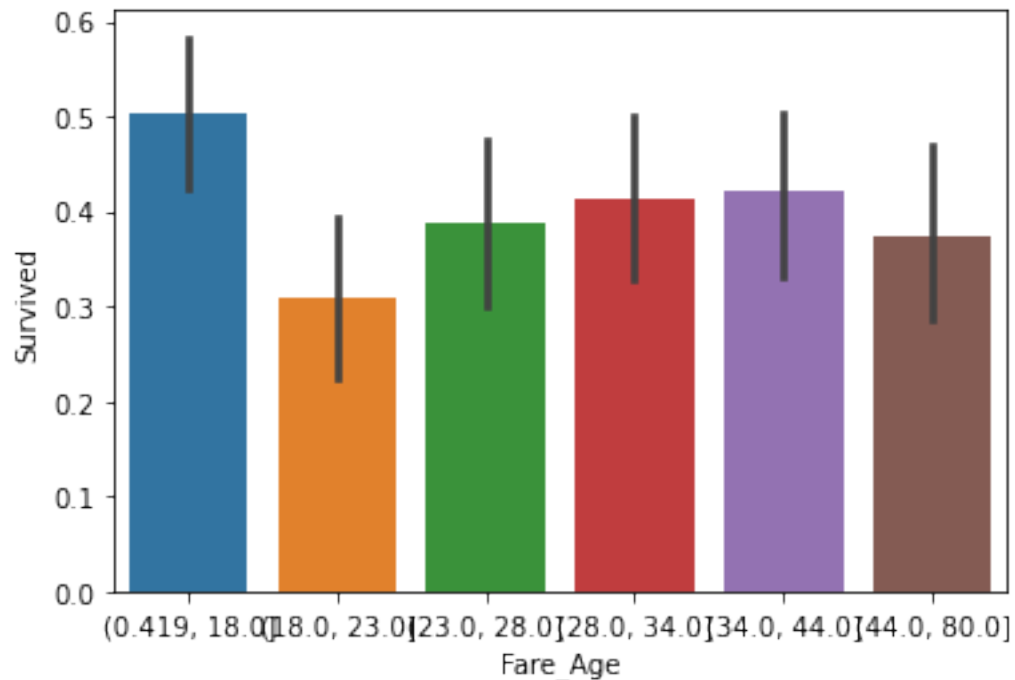
```
[51]: <AxesSubplot:xlabel='SibSp', ylabel='Age'>
```



```
[52]: df['Fare_Age'] = pd.qcut(df['Age'], 6)

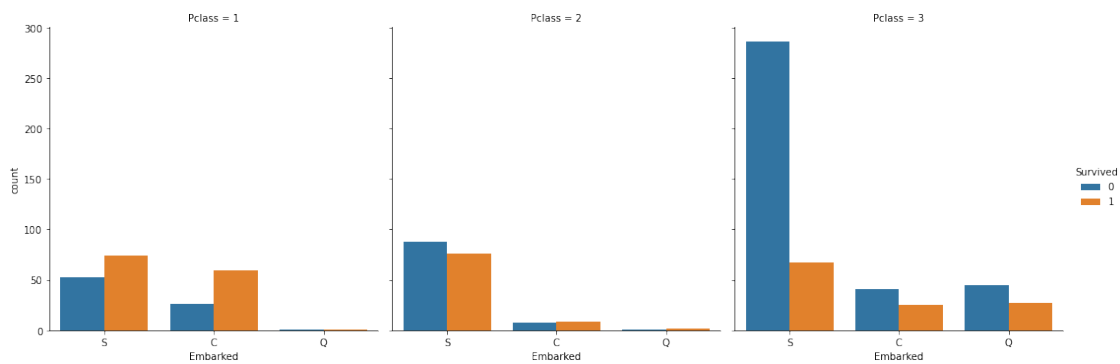
sms.barplot(x='Fare_Age', y='Survived',data = df)
```

```
[52]: <AxesSubplot:xlabel='Fare_Age', ylabel='Survived'>
```



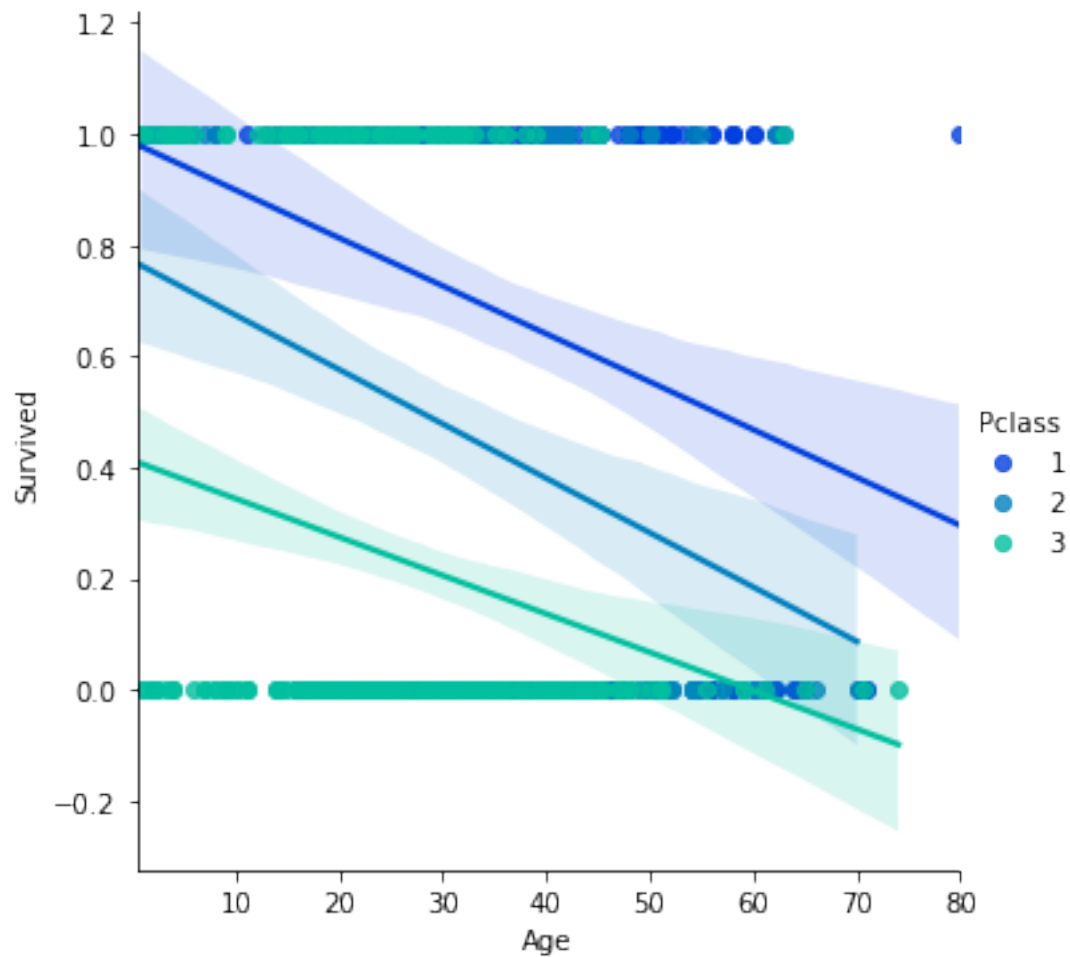
```
[53]: sms.catplot(x='Embarked', hue='Survived', kind='count', col='Pclass', data=df)
```

```
[53]: <seaborn.axisgrid.FacetGrid at 0x2c81e686550>
```



```
[66]: sms.lmplot(x='Age', y='Survived', hue='Pclass', palette='winter', data=df)
```

```
[66]: <seaborn.axisgrid.FacetGrid at 0x2c81fdbca00>
```

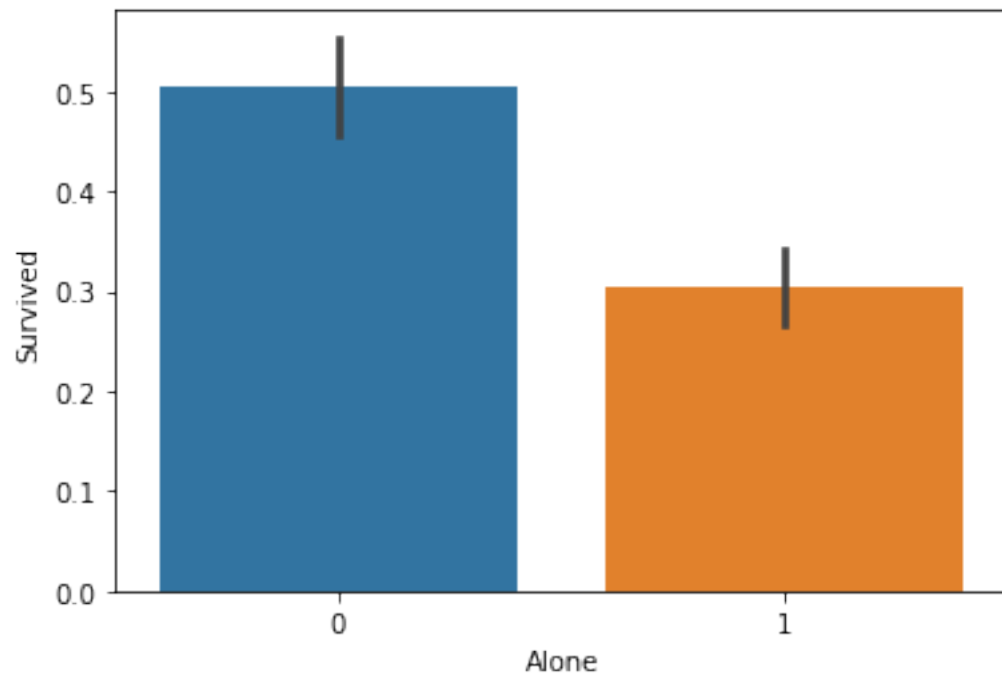


```
[61]: df['Family_Size'] = 0
      df['Family_Size'] = df['Parch']+df['SibSp']

      df['Alone'] = 0
      df.loc[df.Family_Size == 0, 'Alone'] = 1

      sms.barplot(x='Alone', y='Survived', data = df)
```

```
[61]: <AxesSubplot:xlabel='Alone', ylabel='Survived'>
```



[]: