```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sn

read=pd.read_csv(r"D:\titanic\train.csv")
read
```

```
     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/O2. 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]

read.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

read.describe()

        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
```

```
read.isnull().sum()
```

```
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
```

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

```
29.69911764705882
```

```
df=read['Age'].fillna(value=29)
df
```

```
0        22.0
1        38.0
2        26.0
3        35.0
4        35.0
         ...
886      27.0
887      19.0
888      29.0
889      26.0
890      32.0
Name: Age, Length: 891, dtype: float64
```

```
read['Age']=df
df
```

```
0        22.0
1        38.0
2        26.0
3        35.0
4        35.0
         ...
886      27.0
887      19.0
888      29.0
```

```
889    26.0
890    32.0
Name: Age, Length: 891, dtype: float64
```

```
df.isnull().sum()
```

```
0
```

```
read.isnull().sum()
```

```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age              0
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

```
dd=read.drop('Cabin',axis=1)
dd
```

```
     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  29.0
1
889                             Behr, Mr. Karl Howell    male  26.0
0
890                                 Dooley, Mr. Patrick   male  32.0
0

     Parch            Ticket     Fare Embarked
0        0         A/5 21171   7.2500        S
1        0          PC 17599  71.2833        C
2        0  STON/O2. 3101282   7.9250        S
3        0            113803  53.1000        S
4        0            373450   8.0500        S
..     ...               ...      ...      ...
886      0            211536  13.0000        S
887      0            112053  30.0000        S
888      2        W./C. 6607  23.4500        S
889      0            111369  30.0000        C
890      0            370376   7.7500        Q

[891 rows x 11 columns]

dd.isnull().sum()

PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Embarked       2
dtype: int64

dd

     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  29.0
1
889                               Behr, Mr. Karl Howell    male  26.0
0
890                                 Dooley, Mr. Patrick    male  32.0
0

     Parch            Ticket       Fare Embarked
0        0         A/5 21171     7.2500        S
1        0          PC 17599    71.2833        C
2        0   STON/O2. 3101282     7.9250        S
3        0            113803    53.1000        S
4        0            373450     8.0500        S
..     ...               ...        ...      ...
886      0            211536    13.0000        S
887      0            112053    30.0000        S
888      2         W./C. 6607    23.4500        S
889      0            111369    30.0000        C
890      0            370376     7.7500        Q

[891 rows x 11 columns]
```

```python
pd.Categorical(dd['Sex'])
```

```
['male', 'female', 'female', 'female', 'male', ..., 'male', 'female',
'female', 'male', 'male']
Length: 891
Categories (2, object): ['female', 'male']

dd['Sex'].replace(['female','male'],[0,1],inplace=True)
dd
```

```
     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    1  22.0
1
1      Cumings, Mrs. John Bradley (Florence Briggs Th...    0  38.0
1
2                              Heikkinen, Miss. Laina    0  26.0
0
3        Futrelle, Mrs. Jacques Heath (Lily May Peel)    0  35.0
1
4                            Allen, Mr. William Henry    1  35.0
0
..                                               ...  ...   ...    .
..
886                            Montvila, Rev. Juozas    1  27.0
0
887                     Graham, Miss. Margaret Edith    0  19.0
0
888         Johnston, Miss. Catherine Helen "Carrie"    0  29.0
1
889                            Behr, Mr. Karl Howell    1  26.0
0
890                                Dooley, Mr. Patrick    1  32.0
0

     Parch            Ticket     Fare Embarked
0        0         A/5 21171   7.2500        S
1        0          PC 17599  71.2833        C
2        0  STON/O2. 3101282   7.9250        S
```

```
3          0          113803  53.1000          S
4          0          373450   8.0500          S
..       ...               ...      ...        ...
886        0          211536  13.0000          S
887        0          112053  30.0000          S
888        2      W./C. 6607  23.4500          S
889        0          111369  30.0000          C
890        0          370376   7.7500          Q

[891 rows x 11 columns]
```

```
ss=dd.drop(['Name','Ticket'],axis=1)
ss
```

```
     PassengerId  Survived  Pclass  Sex   Age  SibSp  Parch      Fare
Embarked
0              1         0       3    1  22.0      1      0    7.2500
S
1              2         1       1    0  38.0      1      0   71.2833
C
2              3         1       3    0  26.0      0      0    7.9250
S
3              4         1       1    0  35.0      1      0   53.1000
S
4              5         0       3    1  35.0      0      0    8.0500
S
..           ...       ...     ...  ...   ...    ...    ...       ...
...
886          887         0       2    1  27.0      0      0   13.0000
S
887          888         1       1    0  19.0      0      0   30.0000
S
888          889         0       3    0  29.0      1      2   23.4500
S
889          890         1       1    1  26.0      0      0   30.0000
C
890          891         0       3    1  32.0      0      0    7.7500
Q

[891 rows x 9 columns]
```

```
ss
```

```
     PassengerId  Survived  Pclass  Sex   Age  SibSp  Parch      Fare
Embarked
0              1         0       3    1  22.0      1      0    7.2500
S
1              2         1       1    0  38.0      1      0   71.2833
C
2              3         1       3    0  26.0      0      0    7.9250
S
```

```
3                4          1     1    0  35.0        1        0  53.1000
S
4                5          0     3    1  35.0        0        0   8.0500
S
..             ...        ...   ...  ...   ...      ...      ...      ...
...
886            887          0     2    1  27.0        0        0  13.0000
S
887            888          1     1    0  19.0        0        0  30.0000
S
888            889          0     3    0  29.0        1        2  23.4500
S
889            890          1     1    1  26.0        0        0  30.0000
C
890            891          0     3    1  32.0        0        0   7.7500
Q

[891 rows x 9 columns]


ss.isnull().sum()

PassengerId    0
Survived       0
Pclass         0
Sex            0
Age            0
SibSp          0
Parch          0
Fare           0
Embarked       2
dtype: int64

ss['Embarked'].value_counts()

S    644
C    168
Q     77
Name: Embarked, dtype: int64

ss['Embarked']=ss['Embarked'].fillna(value='S')
ss['Embarked']

0      S
1      C
2      S
3      S
4      S
      ..
886    S
887    S
```

```
888    S
889    C
890    Q
Name: Embarked, Length: 891, dtype: object
```

```
ss['Embarked'].isnull().sum()
```

```
0
```

```
ss
```

```
        PassengerId  Survived  Pclass  Sex    Age  SibSp  Parch      Fare
Embarked
0                 1         0       3    1   22.0      1      0    7.2500
S
1                 2         1       1    0   38.0      1      0   71.2833
C
2                 3         1       3    0   26.0      0      0    7.9250
S
3                 4         1       1    0   35.0      1      0   53.1000
S
4                 5         0       3    1   35.0      0      0    8.0500
S
..              ...       ...     ...  ...    ...    ...    ...       ...
...
886             887         0       2    1   27.0      0      0   13.0000
S
887             888         1       1    0   19.0      0      0   30.0000
S
888             889         0       3    0   29.0      1      2   23.4500
S
889             890         1       1    1   26.0      0      0   30.0000
C
890             891         0       3    1   32.0      0      0    7.7500
Q
```
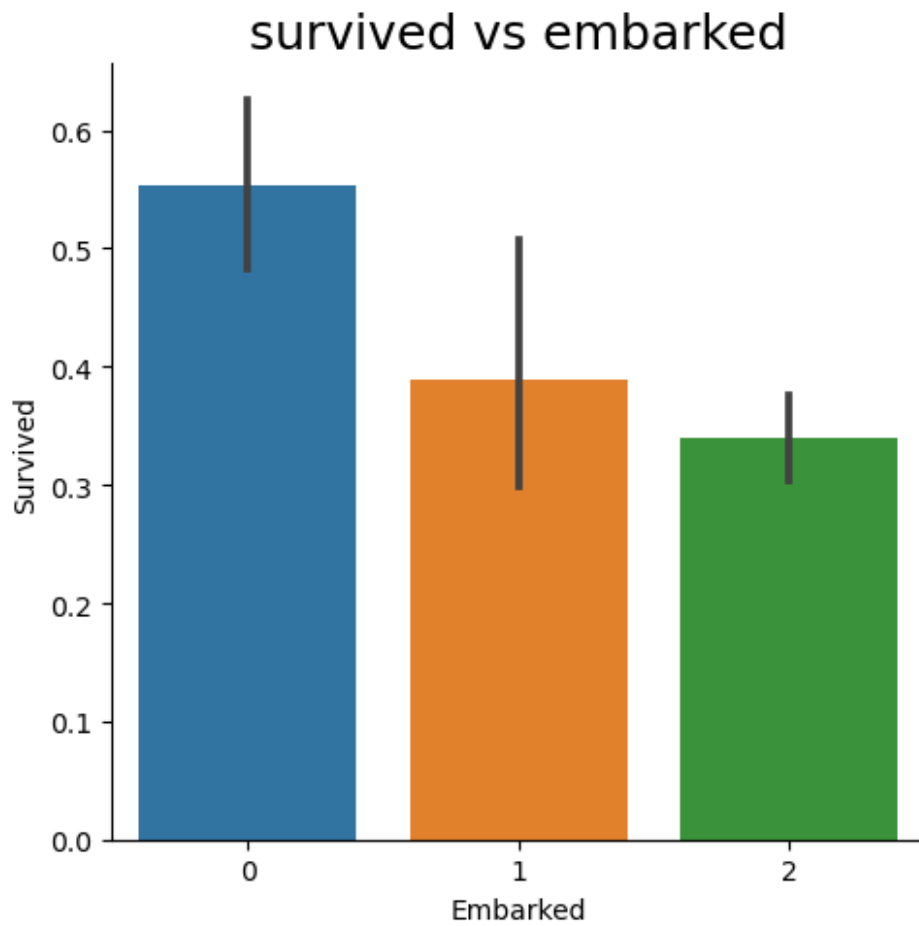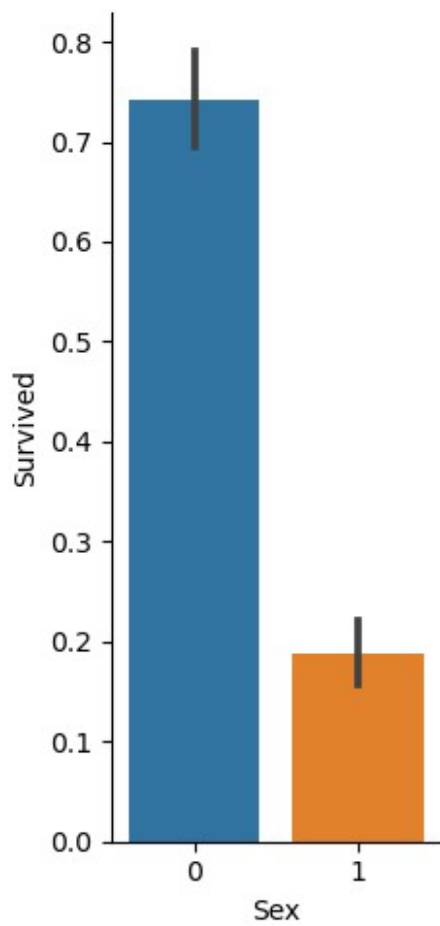
```
[891 rows x 9 columns]
```

```
from sklearn import preprocessing
```

```
label_encoder=preprocessing.LabelEncoder()
```

```
ss['Embarked']=label_encoder.fit_transform(ss['Embarked'])
ss['Embarked'].value_counts()
```

```
2    646
0    168
1     77
Name: Embarked, dtype: int64
```

```
ss
```

```
       PassengerId  Survived  Pclass  Sex   Age  SibSp  Parch      Fare
Embarked
0                1         0       3    1  22.0      1      0    7.2500
2
1                2         1       1    0  38.0      1      0   71.2833
0
2                3         1       3    0  26.0      0      0    7.9250
2
3                4         1       1    0  35.0      1      0   53.1000
2
4                5         0       3    1  35.0      0      0    8.0500
2
..             ...       ...     ...  ...   ...    ...    ...       ...
...
886            887         0       2    1  27.0      0      0   13.0000
2
887            888         1       1    0  19.0      0      0   30.0000
2
888            889         0       3    0  29.0      1      2   23.4500
2
889            890         1       1    1  26.0      0      0   30.0000
0
890            891         0       3    1  32.0      0      0    7.7500
1

[891 rows x 9 columns]

sn.catplot(x="Embarked",
           y="Survived",
           kind="bar",
           data=ss,
           height=5,
           aspect=1.0)           #Embarked C=0,Q=1,S=2
plt.xlabel("Embarked")
plt.ylabel("Survived")
#plt.title("survived vs embarked", size=18)
plt.tight_layout()
```

survived vs embarked


```
sn.catplot(x="Sex",
           y="Survived",
           kind="bar",
           data=ss,
           height=5,
           aspect=0.5)
plt.xlabel("Sex")
plt.ylabel("Survived")
#plt.title("survived vs embarked", size=18)
plt.tight_layout()
```
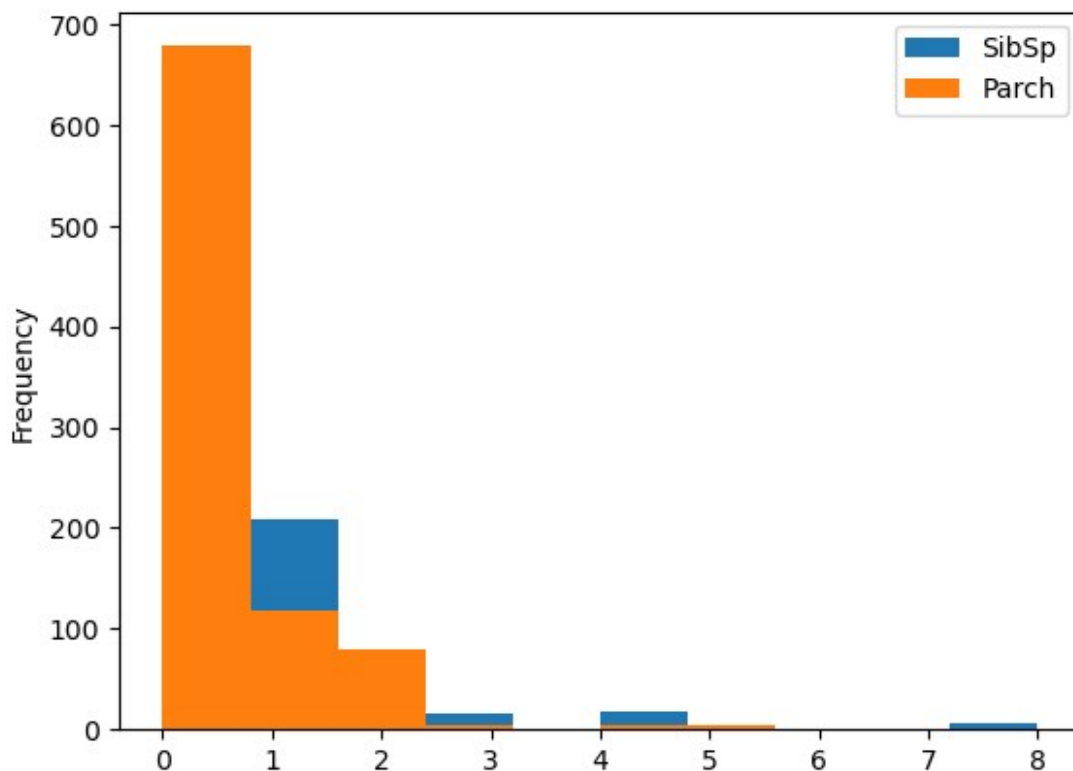
```
ss.plot(x='Embarked',y=['SibSp','Parch'],kind='hist')
```
```
<Axes: ylabel='Frequency'>
```

ss.corr()

```
              PassengerId  Survived    Pclass       Sex       Age
SibSp  \
PassengerId      1.000000 -0.005007 -0.035144  0.042939  0.033632 -
0.057527
Survived        -0.005007  1.000000 -0.338481 -0.543351 -0.067814 -
0.035322
Pclass          -0.035144 -0.338481  1.000000  0.131900 -0.334974
0.083081
Sex              0.042939 -0.543351  0.131900  1.000000  0.082949 -
0.114631
Age              0.033632 -0.067814 -0.334974  0.082949  1.000000 -
0.232978
SibSp           -0.057527 -0.035322  0.083081 -0.114631 -0.232978
1.000000
Parch           -0.001652  0.081629  0.018443 -0.245489 -0.176486
0.414838
Fare             0.012658  0.257307 -0.549500 -0.182333  0.093706
0.159651
Embarked         0.013128 -0.167675  0.162098  0.108262 -0.023464
0.068230


                Parch      Fare  Embarked
PassengerId -0.001652  0.012658  0.013128
Survived     0.081629  0.257307 -0.167675
```

```
Pclass         0.018443 -0.549500  0.162098
Sex           -0.245489 -0.182333  0.108262
Age           -0.176486  0.093706 -0.023464
SibSp          0.414838  0.159651  0.068230
Parch          1.000000  0.216225  0.039798
Fare           0.216225  1.000000 -0.224719
Embarked       0.039798 -0.224719  1.000000
```

```
sn.heatmap(ss.corr())
```

```
<Axes: >
```