Assignment 4 - Soumya Gite - SIRSS2276

August 27, 2021

[1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
```

[2]:
```
data = pd.read_csv("C:/Users/SUDHAKAR/Downloads/EDA1-master/EDA1-master/
↪titanic_train.csv")
```

[3]:
```
#1 Understanding the data
```

[4]:
```
data.head()
```

[4]:

	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

[5]:
```
data.tail()
```

[5]:

	PassengerId	Survived	Pclass	Name
886	887	0	2	Montvila, Rev. Juozas
887	888	1	1	Graham, Miss. Margaret Edith
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"

```
889          890         1       1                                Behr, Mr. Karl Howell
890          891         0       3                                  Dooley, Mr. Patrick

        Sex   Age  SibSp  Parch      Ticket   Fare Cabin Embarked
886    male  27.0      0      0      211536  13.00   NaN        S
887  female  19.0      0      0      112053  30.00   B42        S
888  female   NaN      1      2  W./C. 6607  23.45   NaN        S
889    male  26.0      0      0      111369  30.00  C148        C
890    male  32.0      0      0      370376   7.75   NaN        Q
```

[6]:
```
data.shape
```

[6]:
```
(891, 12)
```

[7]:
```
data.describe()
```

[7]:
```
       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
```

[8]:
```
data.columns
```

[8]:
```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

[ ]:
```
# cleaning the data
```

[19]:
```
data.isnull().sum()
```

[19]:
```
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
```

[20]:
```
titanic_train = data.drop(['Survived','SibSp'],axis = 1)
```

[21]:
```
titanic_train.head()
```

[21]:

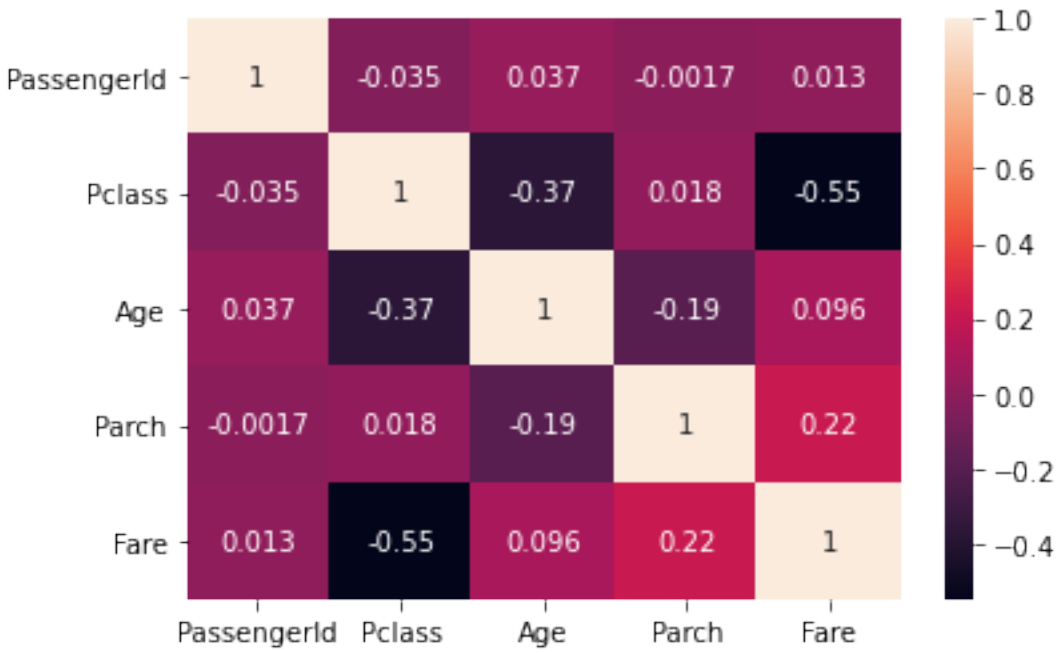
	PassengerId	Pclass	Name	Sex	Age	Parch	Ticket	Fare	Cabin	Embarked
0	1	3	Braund, Mr. Owen Harris	male	22.0	0	A/5 21171	7.2500	NaN	S
1	2	1	Cumings, Mrs. John Bradley (Florence Briggs Th…	female	38.0	0	PC 17599	71.2833	C85	C
2	3	3	Heikkinen, Miss. Laina	female	26.0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	Futrelle,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

[ ]:
```
# relationship analysis
```

[22]:
```
corelation = titanic_train.corr()
```

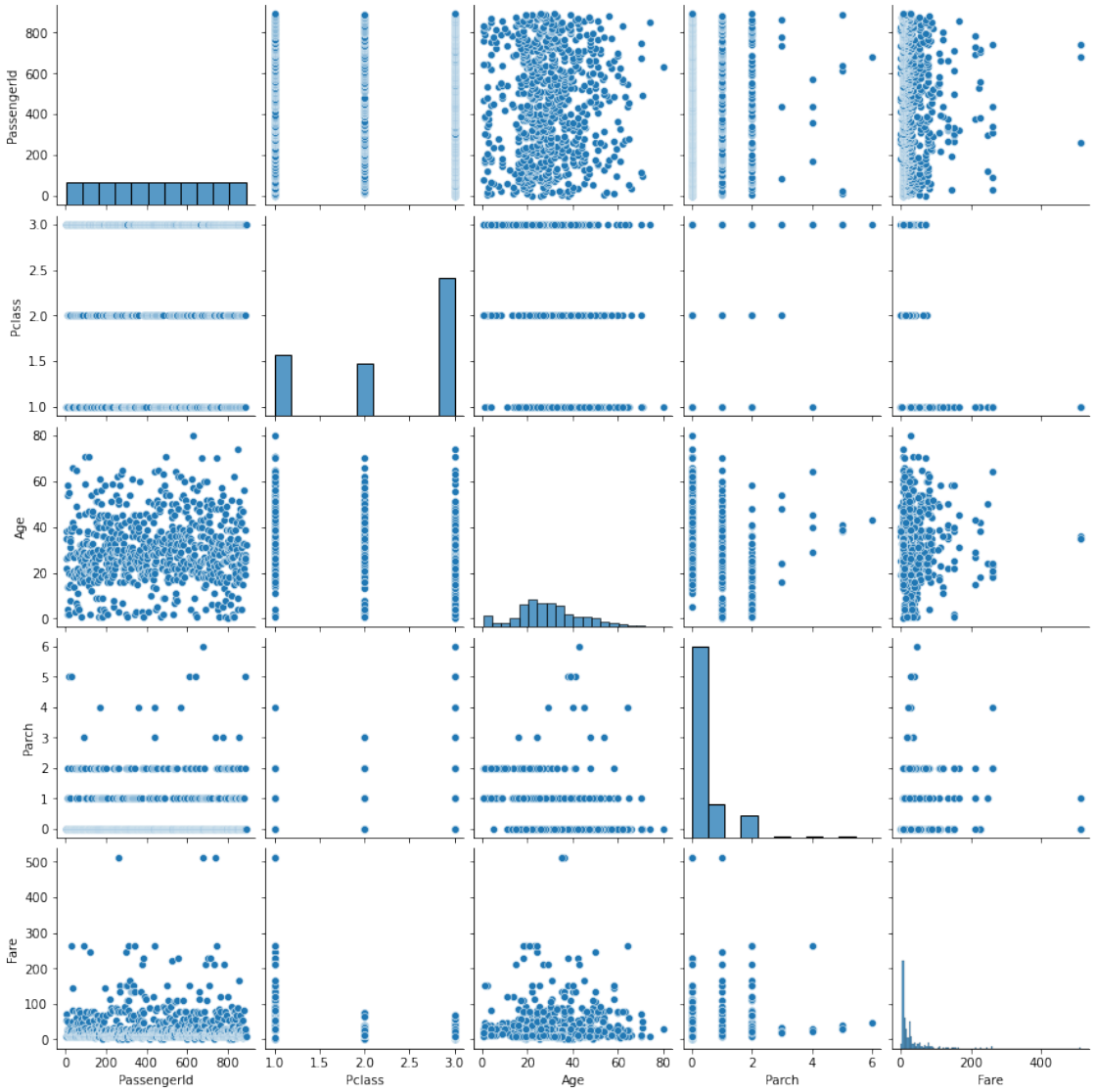
[23]:
```
sns.heatmap(corelation,xticklabels = corelation.columns, yticklabels =␣
↪corelation.columns,annot = True)
```

[23]:
```
<AxesSubplot:>
```



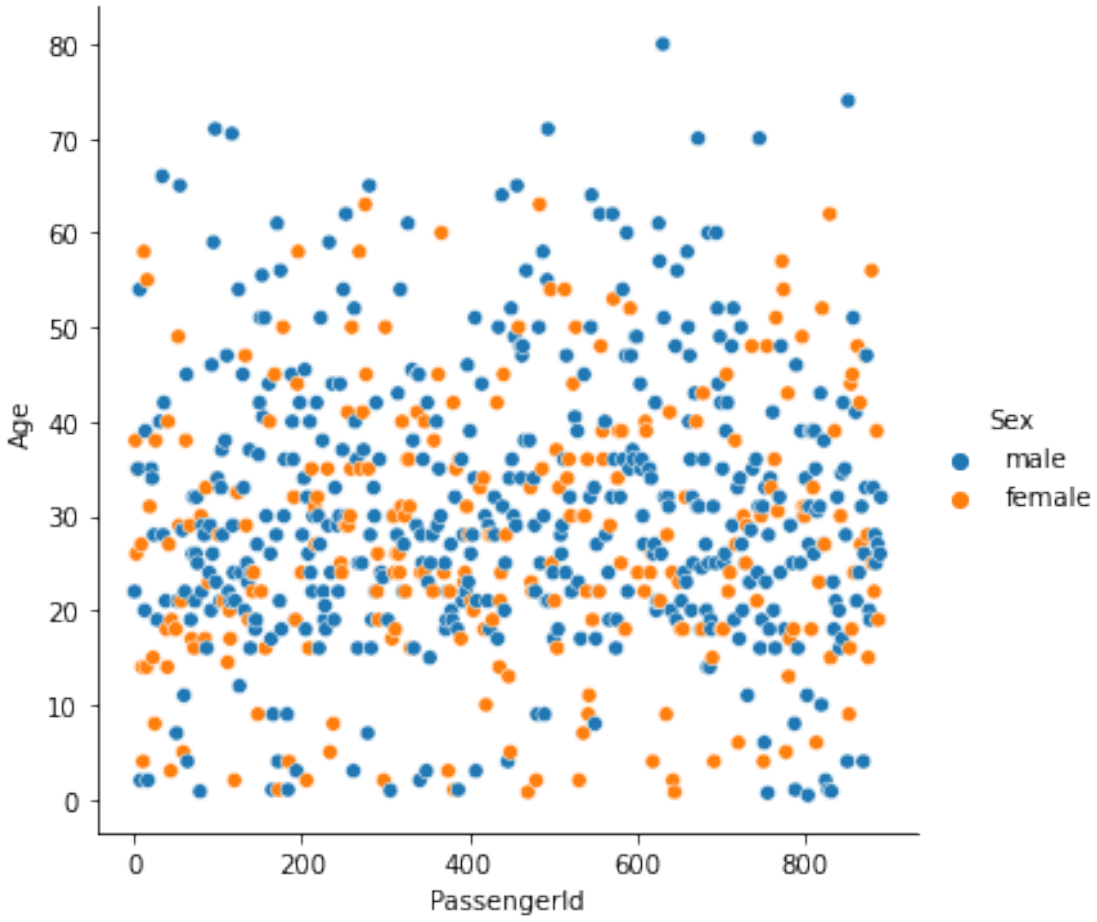
[24]:
```
sns.pairplot(titanic_train)
```

[24]: <seaborn.axisgrid.PairGrid at 0x2369acea760>



[30]:
```
sns.relplot(x = 'PassengerId',y = 'Age',hue = 'Sex',data = titanic_train)
```

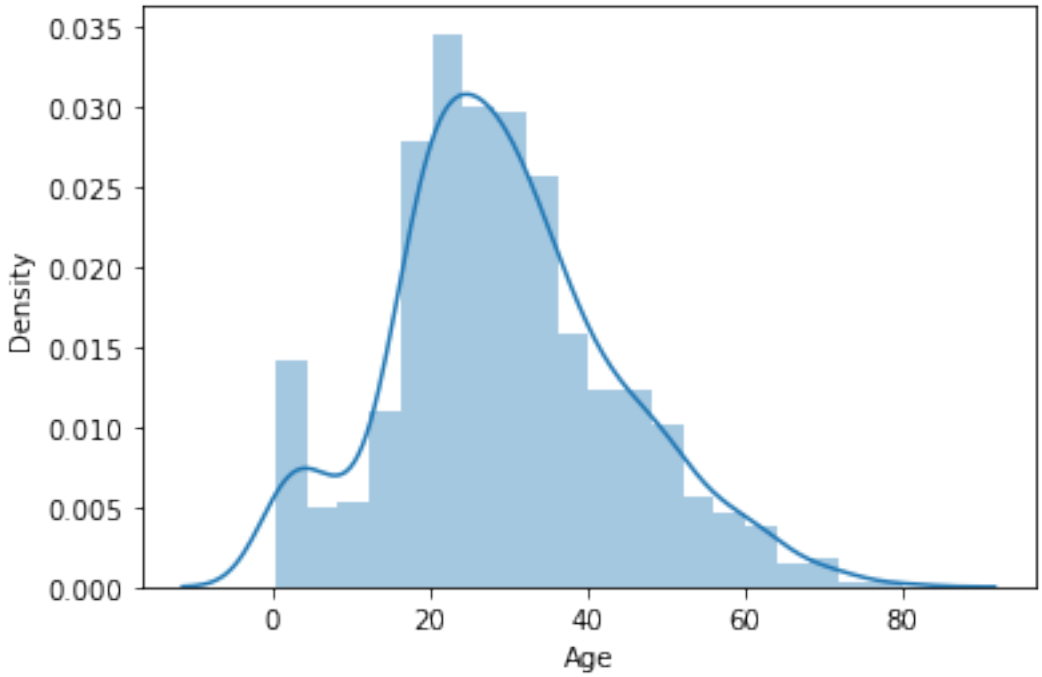
[30]: <seaborn.axisgrid.FacetGrid at 0x2369c79a100>



[34]:
```
sns.distplot(titanic_train['Age'])
```

```
C:\Users\SUDHAKAR\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `histplot` (an axes-level function for
histograms).
  warnings.warn(msg, FutureWarning)
```

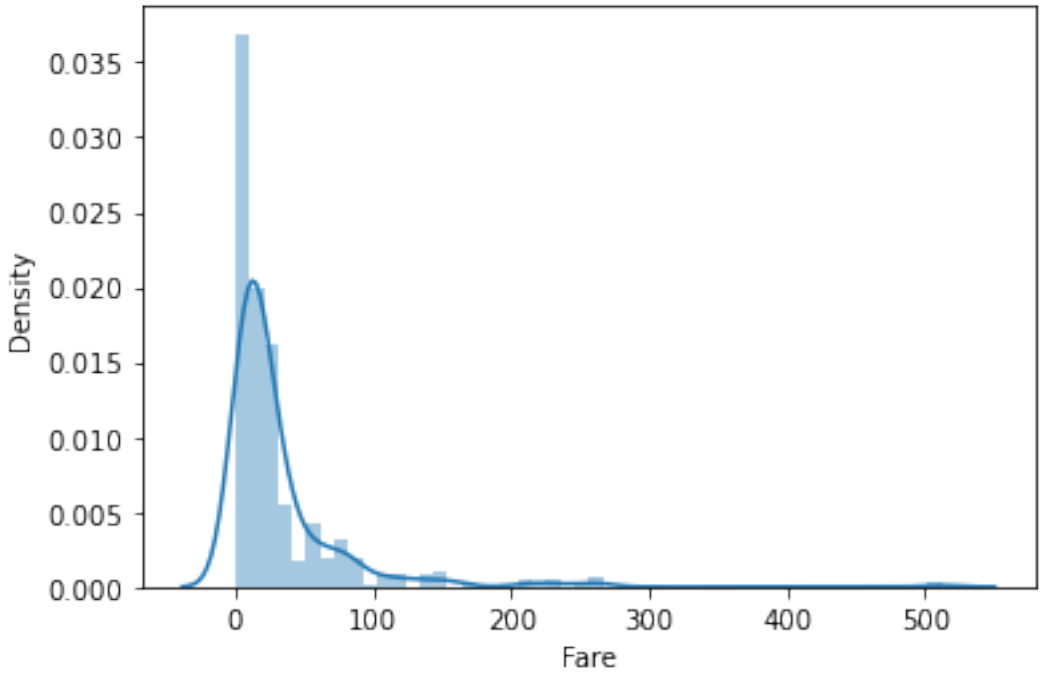
[34]:
```
<AxesSubplot:xlabel='Age', ylabel='Density'>
```



[35]:
```
sns.distplot(titanic_train['Fare'])
```

```
C:\Users\SUDHAKAR\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning: `distplot` is a deprecated function and will be removed in a
future version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `histplot` (an axes-level function for
histograms).
  warnings.warn(msg, FutureWarning)
```

[35]:
```
<AxesSubplot:xlabel='Fare', ylabel='Density'>
```


0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
Density
0
100
200
300
400
500
Fare

[ ]: