campusx-official / 100-days-of-machine-learning Public

Code Issues 3 Pull requests Actions Projects Security Insights

main

100-days-of-machine-learning / day20-univariate-analysis / day20.ipynb

campusx-official Add files via upload History

1 contributor

501 lines (501 sloc) | 52.3 KB

In [4]:
```
import pandas as pd
import seaborn as sns
```

In [2]:
```
df = pd.read_csv('train.csv')
```

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

Out[3]:

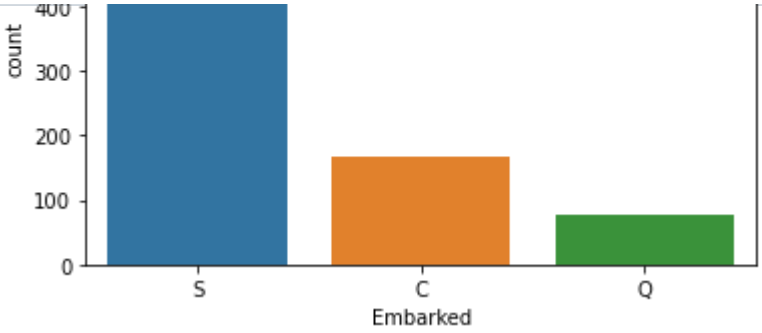
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250
3	4	1	1	Futrelle,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

1. Categorical Data

a. Countplot

In [12]:
```
sns.countplot(df['Embarked'])
#df['Survived'].value_counts().plot(kind='bar')
```

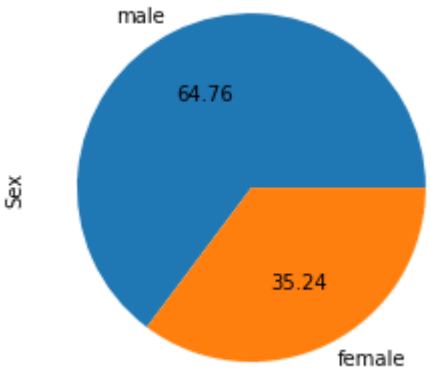
Out[12]:



b. PieChart

In [16]:
```
df['Sex'].value_counts().plot(kind='pie',autopct='%.2f')
```

Out[16]:



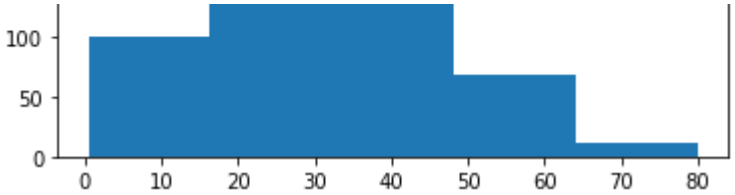
---

2. Numerical Data

a. Histogram

In [25]:
```
import matplotlib.pyplot as plt
plt.hist(df['Age'],bins=5)
```

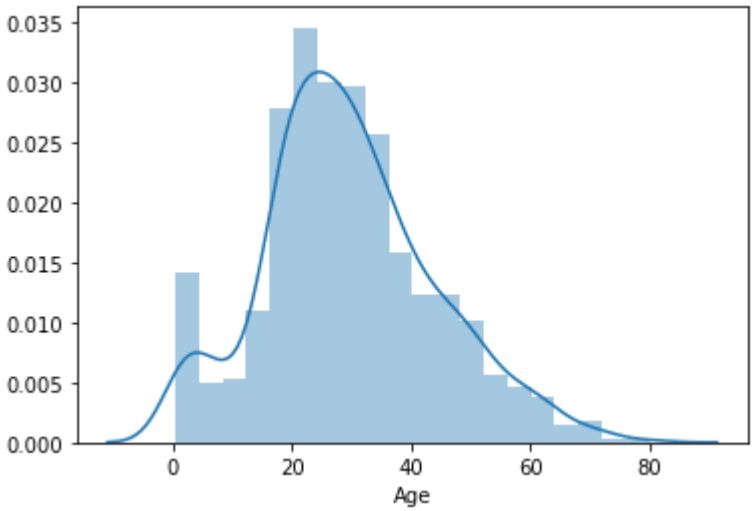
Out[25]:
```
(array([100., 346., 188.,  69.,  11.]),
 array([ 0.42 , 16.336, 32.252, 48.168, 64.084, 80.   ]),
 )
```



b. Distplot

In [26]:
```
sns.distplot(df['Age'])
```

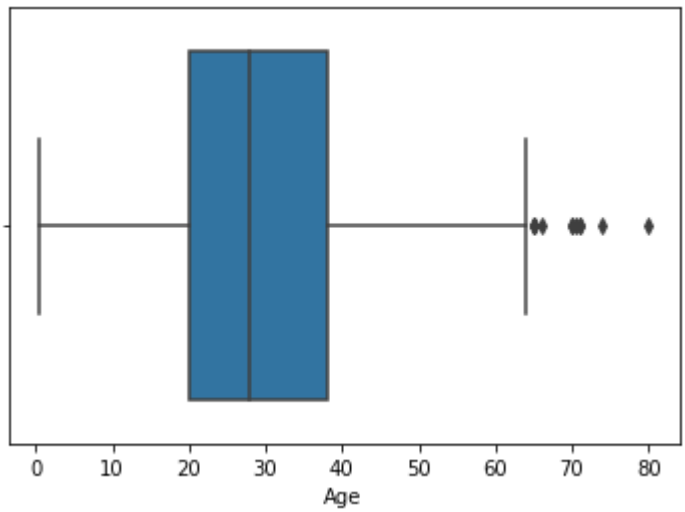
Out[26]:



c. Boxplot

In [28]:
```
sns.boxplot(df['Age'])
```

Out[28]:



In [29]:
```
df['Age'].min()
```

Out[29]:
```
0.42
```

In [30]:
```
df['Age'].max()
```