Visualizing the Titanic Disaster

Introduction:

This exercise is based on the titanic Disaster dataset avaiable at Kaggle.
To know more about the variables check here

Step 1. Import the necessary libraries

```
import pandas as pd
import matplotlib.pyplot as plt
```

Step 2. Import the dataset from this address

Step 3. Assign it to a variable titanic

```
titanic = pd.read_csv('https://raw.githubusercontent.com/thieu1995/csv-files/main/data/pandas/titanic_train.csv')
titanic.head()
```

	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

Next steps: Generate code with titanic | View recommended plots | New interactive sheet

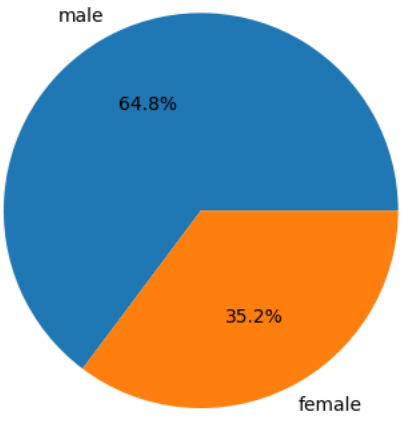
Step 4. Set PassengerId as the index

```
titanic.set_index('PassengerId', inplace=True)
```

Step 5. Create a pie chart presenting the male/female proportion

```
gender_counts = titanic['Sex'].value_counts()
plt.pie(gender_counts, labels=gender_counts.index, autopct='%1.1f%%')

plt.show()
```

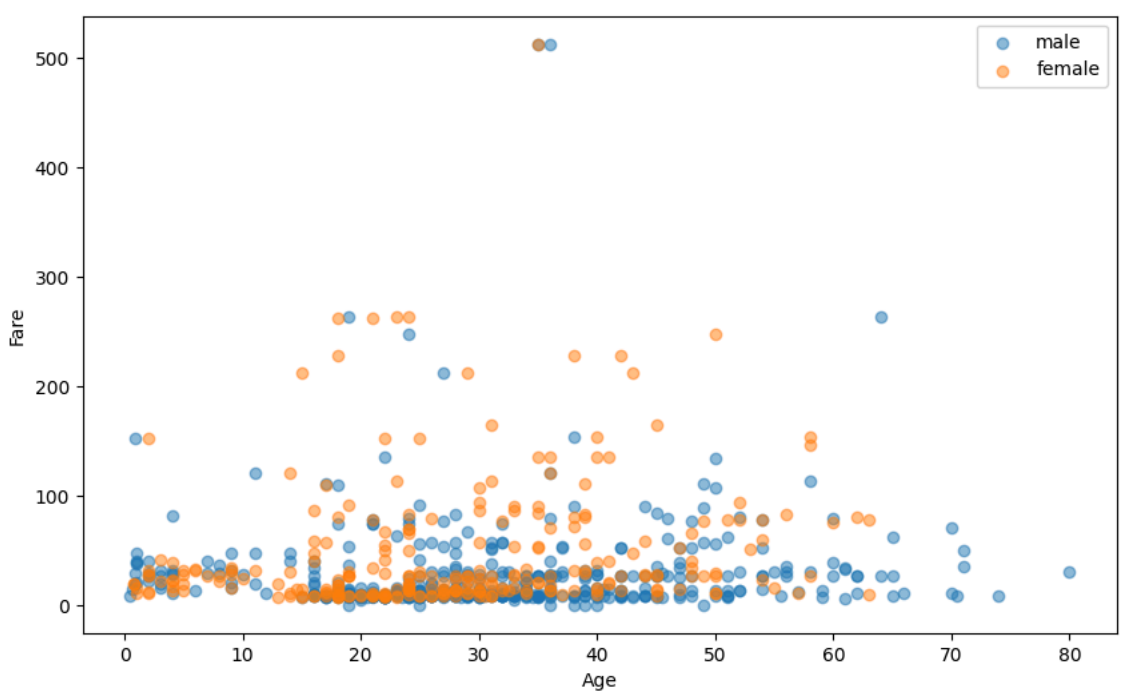


Step 6. Create a scatterplot with the Fare payed and the Age, differ the plot color by gender

```
plt.figure(figsize=(10, 6))
for sex in titanic['Sex'].unique():
    subset = titanic[titanic['Sex'] == sex]
    plt.scatter(subset['Age'], subset['Fare'], label=sex, alpha=0.5)

plt.xlabel('Age')
plt.ylabel('Fare')

plt.legend()
plt.show()
```



Step 7. How many people survived?

```
int(titanic['Survived'].sum())
```

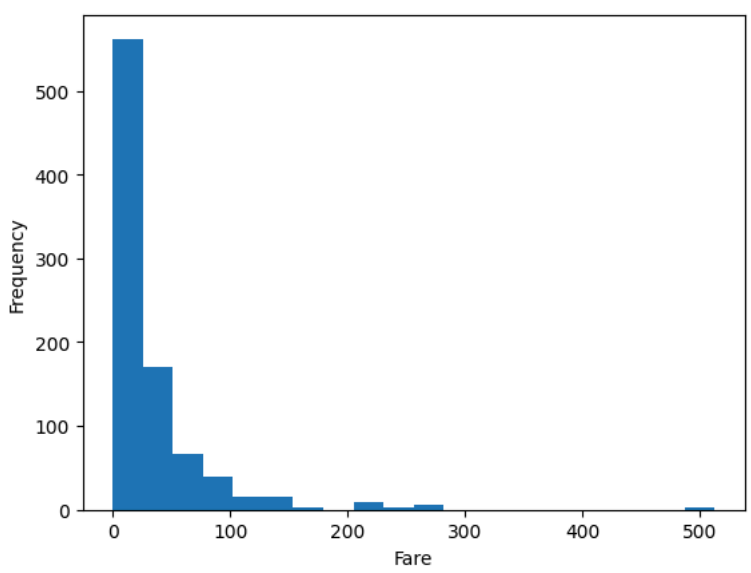
```
342
```

Step 8. Create a histogram with the Fare payed

```
plt.hist(titanic['Fare'], bins=20)

plt.xlabel('Fare')
plt.ylabel('Frequency')

plt.show()
```



BONUS: Create your own question and answer it.

```
# q: avg age of passengers each class?

# a:
age_by_class = titanic.groupby('Pclass')['Age'].mean()

age_by_class
```

Pclass	Age
1	38.233441
2	29.877630
3	25.140620

dtype: float64