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
import seaborn as sns
```

Step 2. Import the dataset from this address

```
url = "https://raw.githubusercontent.com/thieu1995/csv-files/main/data/pandas/titanic_train.csv"
```

Step 3. Assign it to a variable titanic

```
titanic = pd.read_csv(url)

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
				Futrelle, Mrs. Jacques Heath								

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

Step 4. Set PassengerId as the index

```
titanic = titanic.set_index("PassengerId")

titanic.head()
```

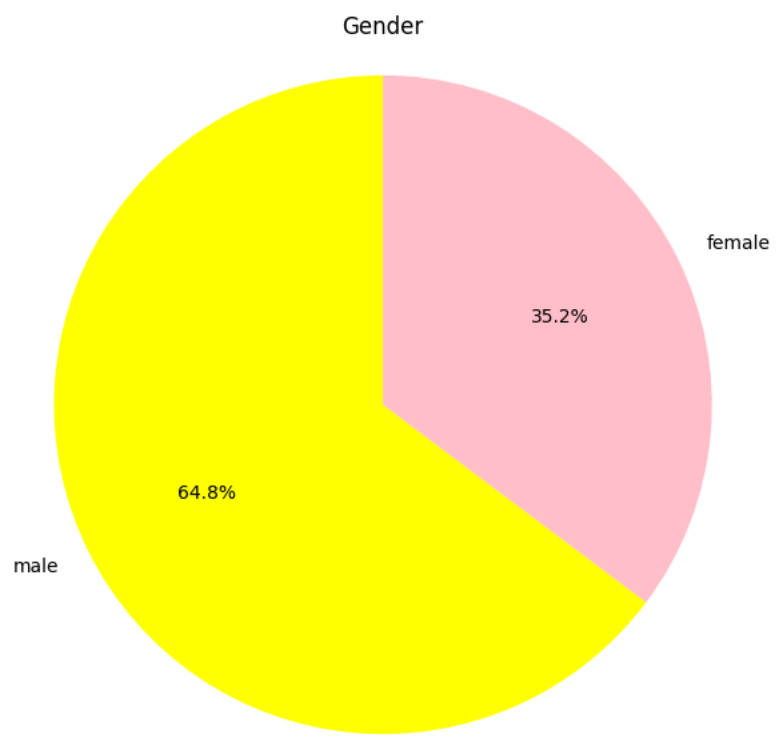
PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
			Futrelle, Mrs. Jacques Heath (Lily								

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

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

```
gender_count = titanic['Sex'].value_counts()

plt.figure(figsize=(7,7))
plt.pie(gender_count, labels=gender_count.index, autopct='%1.1f%%', startangle=90, colors=['yellow', 'pink'])
plt.title('Gender')
plt.axis('equal')
plt.show()
```

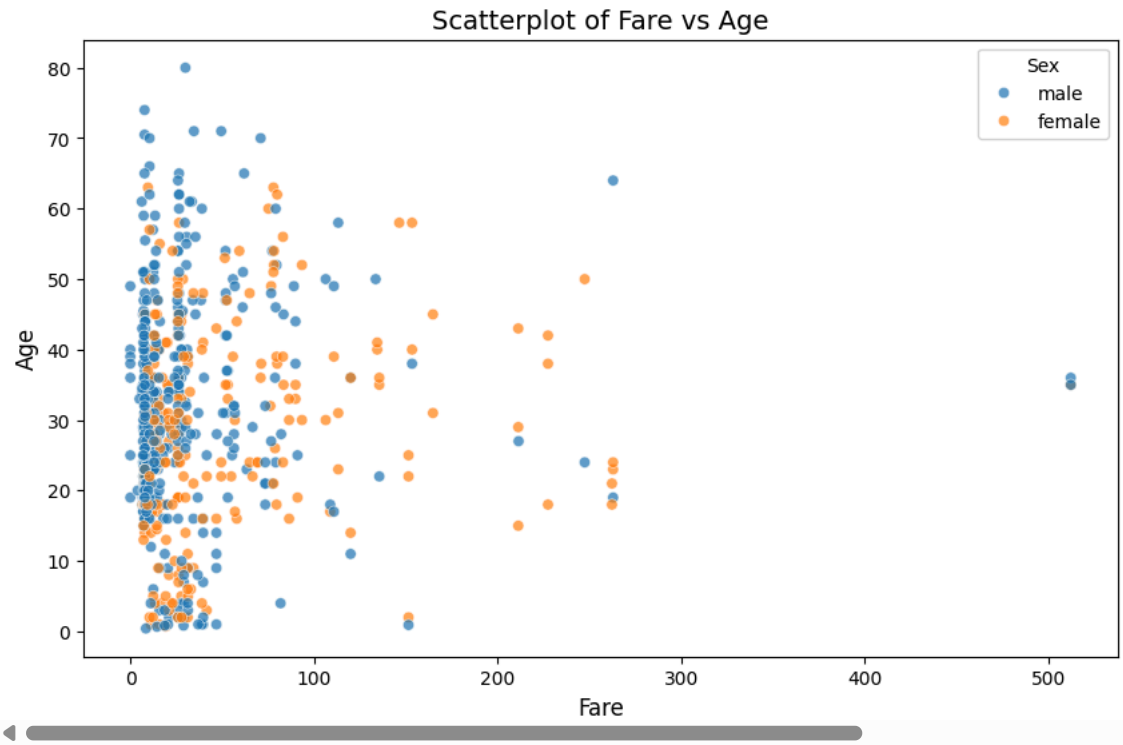


Step 6. Create a scatterplot with the Fare payed and the Age, differ the plot color by gender

```
plt.figure(figsize=(10, 6))
sns.scatterplot(data=titanic, x='Fare', y='Age', hue='Sex', alpha=0.7)

plt.title('Scatterplot of Fare vs Age', fontsize=14)
plt.xlabel('Fare', fontsize=12)
plt.ylabel('Age', fontsize=12)

plt.show()
```



Step 7. How many people survived?

```
survived = titanic['Survived'].sum()

print(survived)
```

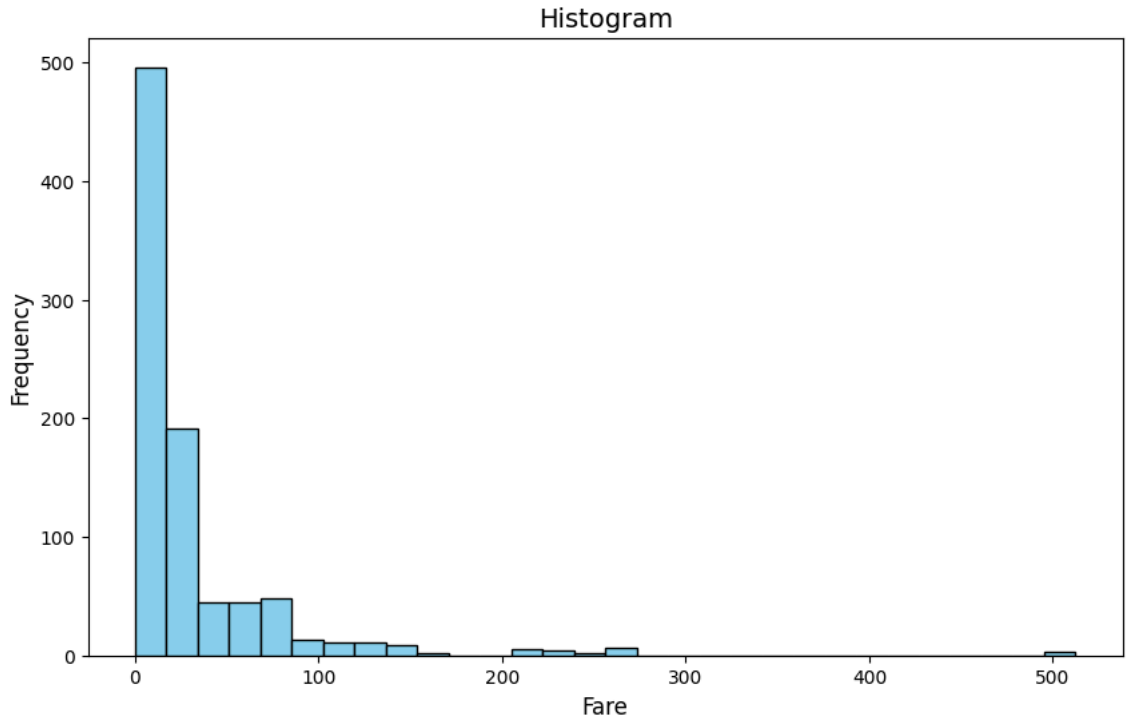
```
342
```

Step 8. Create a histogram with the Fare payed

```
plt.figure(figsize=(10, 6))
plt.hist(titanic['Fare'], bins=30, edgecolor='black', color='skyblue')

plt.title('Histogram', fontsize=14)
plt.xlabel('Fare', fontsize=12)
plt.ylabel('Frequency', fontsize=12)
```

```
Text(0, 0.5, 'Frequency')
```



BONUS: Create your own question and answer it.

How can I find the average age of passengers who survived and those who didn't on the Titanic?

```
average_age_by_survival = titanic.groupby('Survived')['Age'].mean()

print(average_age_by_survival)
```

```
Survived
0    30.626179
1    28.343690
Name: Age, dtype: float64
```