
TASK 02
Perform data cleaning and exploratory data analysis (EDA) on a dataset of your choice, such as the Titanic dataset from Kaggle. Explore the relationships between variables and identify patterns and trends in the data.
Titanic Dataset
SkillCraft Technology

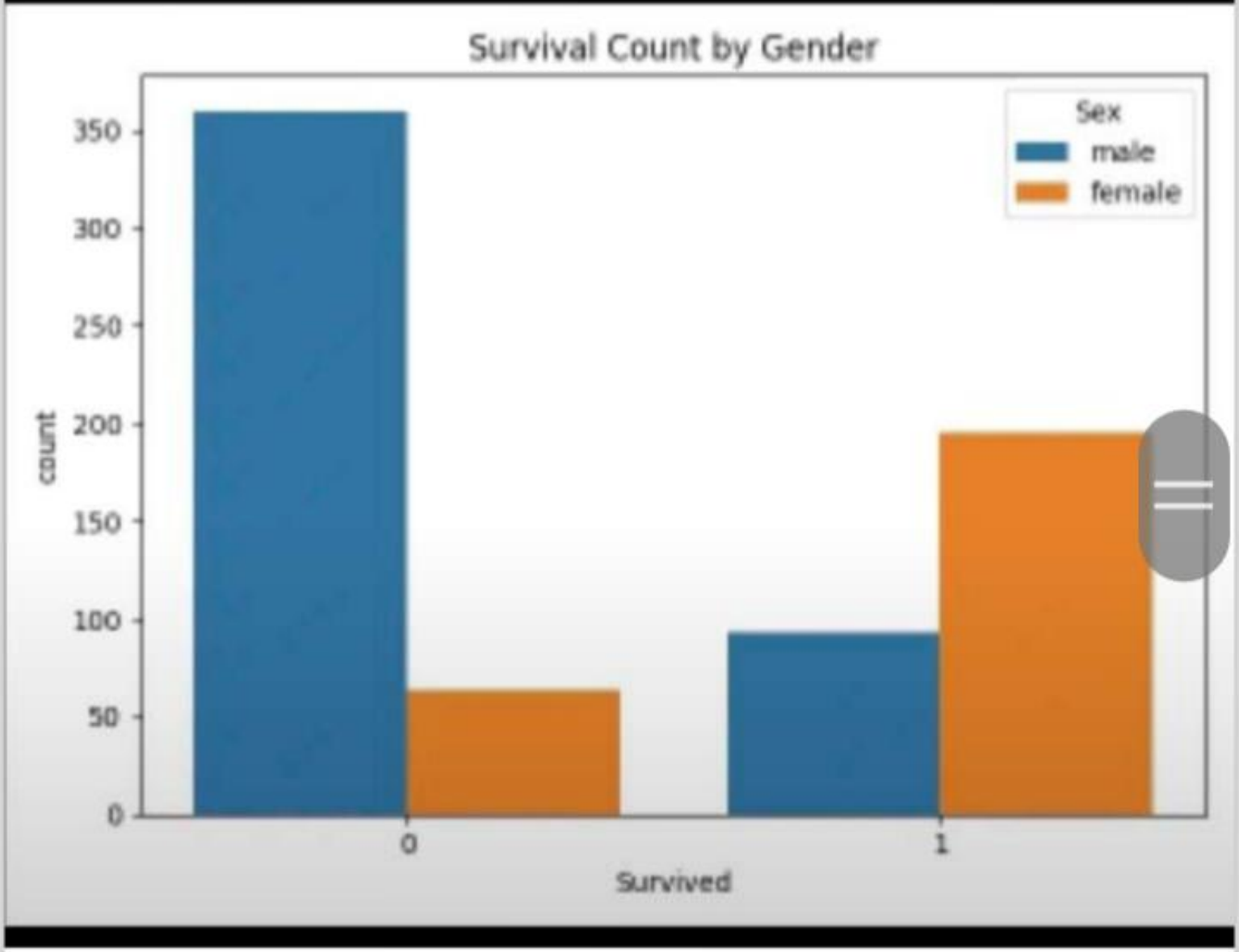
Programiz Python Online Compiler

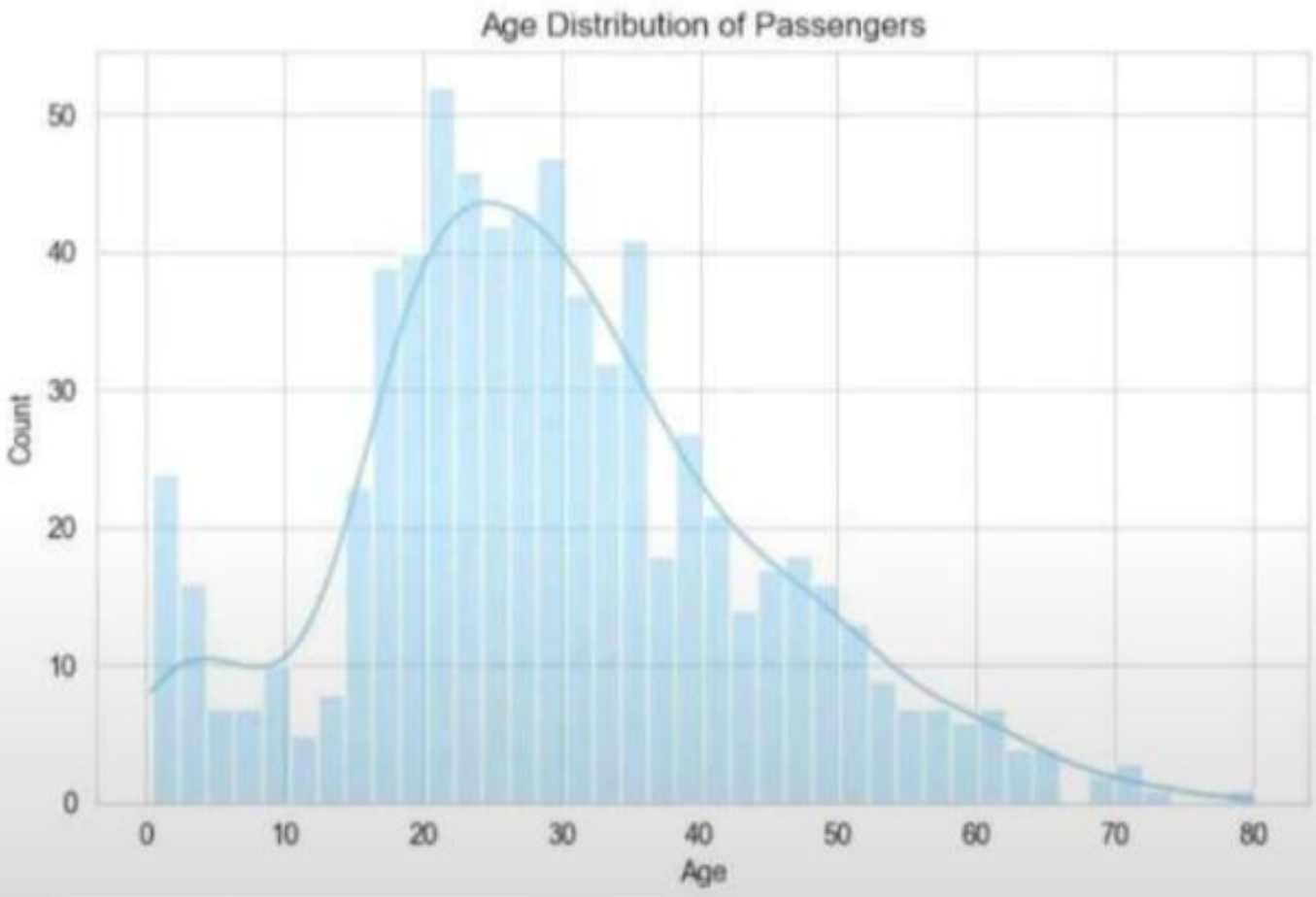
main.py

Share Run

```python
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Load dataset
df= pd.read_csv("https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv")

# Clean missing values
df.dropna(subset=['Age', "Embarked"], inplace=True)

# written summary
total_passengers = len(df)
survived=df[" Survived"].sum()
survival_rate = round((survived/total_passengers)*100, 2)

print("Titanic Dataset EDA")
print(f"Total Passengers (after cleaning): {total_passengers}")
print(f"Survivors: {survived} ({survival_rate}%)")
print("\nGender Breakdown:")
print(df['Sex'].value_counts())
print("\nEsburkation Ports:")
print(df['Tabarked'].value_counts())

# Plot survival by gender
sns.countplot(x='Survived', hue='Sex', data=df)
plt.title("Survival Count by Gender")
plt.tight_layout()
plt.show()

```

```python
# Plot age distribution
sns.histplot(df['Age'], bins=30, kde=True)
plt.title("Age Distribution of Passengers")
plt.tight_layout()
plt.show()
```


Survival Count by Gender
Sex
male
female
count
350
300
250
200
150
100
50
0
0
1
Survived


Age Distribution of Passengers
Count
0
10
20
30
40
50
0
10
20
30
40
50
60
70
80
Age