Training Day 13 Report

Date: 9 July 2025

Topic: Data Visualization with Seaborn (Titanic Dataset)

Overview

On this day, you worked with **Seaborn** (a statistical data visualization library built on top of Matplotlib). The focus was on analyzing the **Titanic dataset** using different plots to understand passenger demographics, survival, fare distribution, and other insights.

Key Code and Explanations

1. Importing Libraries & Dataset

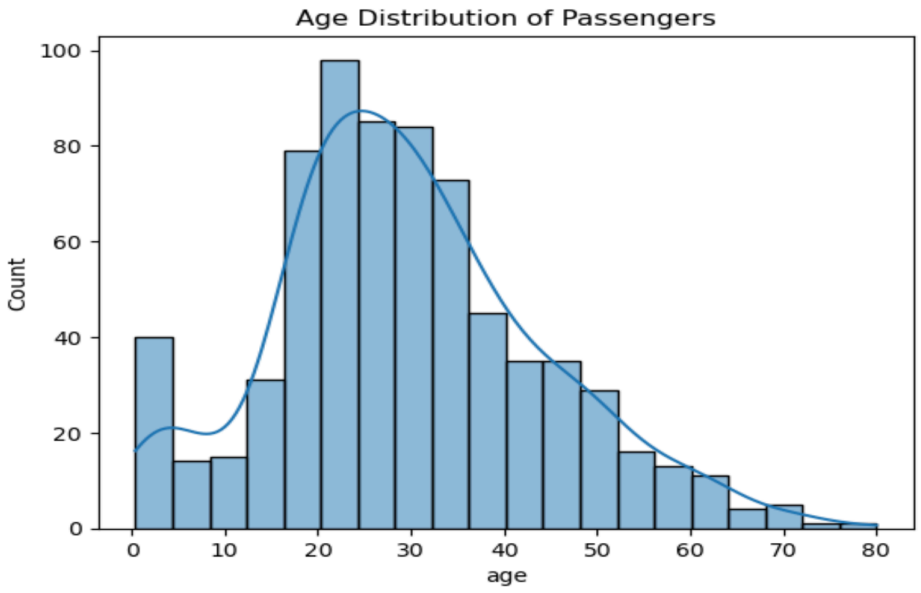
```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd

titanic = sns.load_dataset("titanic")
```

- **Seaborn dataset** titanic contains passenger details like age, sex, class, fare, and survival info.

2. Age Distribution (Histogram)

```
sns.histplot(data=df, x='age', kde=True)
plt.title("Age Distribution of Passengers")
plt.show()
```

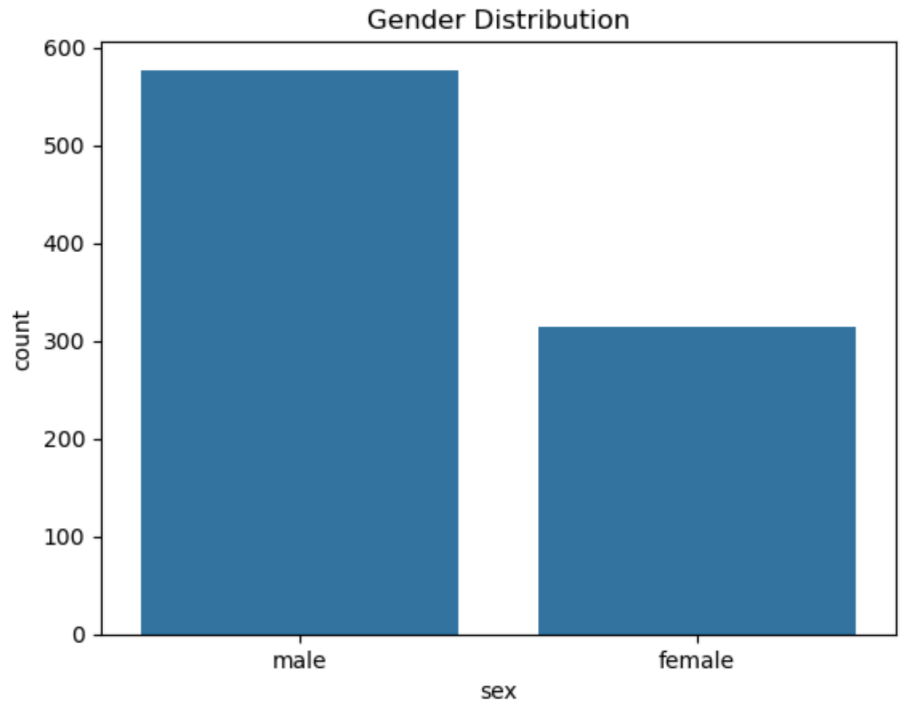


- Shows distribution of passenger ages.
- Most passengers are between **20–40 years**.

3. Gender Distribution (Bar Chart)

sns.countplot(x='sex', data=titanic)

```
sns.countplot(data=df, x='sex')
plt.title("Gender Distribution")
plt.show()
```

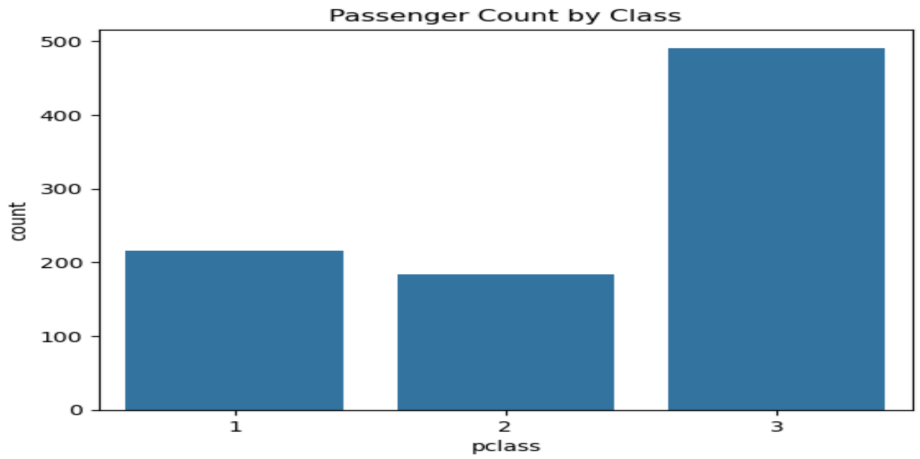


- More **males than females** on Titanic.

4. Passenger Class Distribution

sns.countplot(x='class', data=titanic)

```
sns.countplot(data=df, x='pclass')
plt.title("Passenger Count by Class")
plt.show()
```

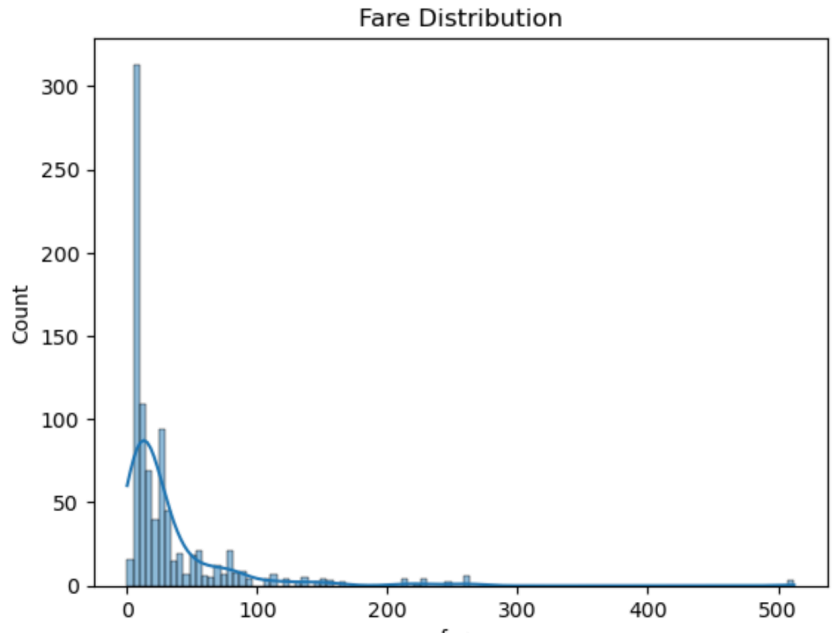


- Most passengers belonged to **3rd class**.

5. Fare Distribution (Histogram + Boxplot)

sns.histplot(titanic['fare'], bins=50, kde=True)
sns.boxplot(x=titanic['fare'])

```
sns.histplot(data=df, x='fare', kde=True)
plt.title("Fare Distribution")
plt.show()
```

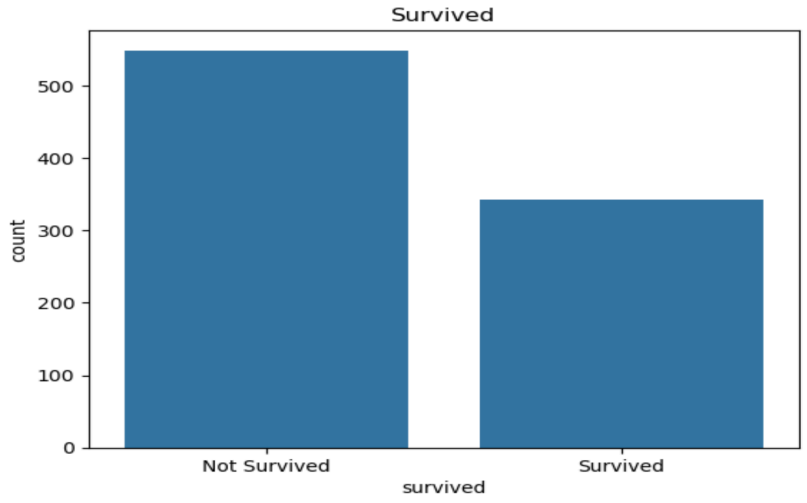


- **Histogram:** Shows most fares between 0–100.
- **Boxplot:** Detects outliers (very high fares).

6. Survival Count (Bar Chart)

sns.countplot(x='survived', data=titanic)

```
sns.countplot(data=df,x='survived')
plt.title("Survived")
plt.xticks([0, 1], ['Not Survived', 'Survived'])
plt.show()
```

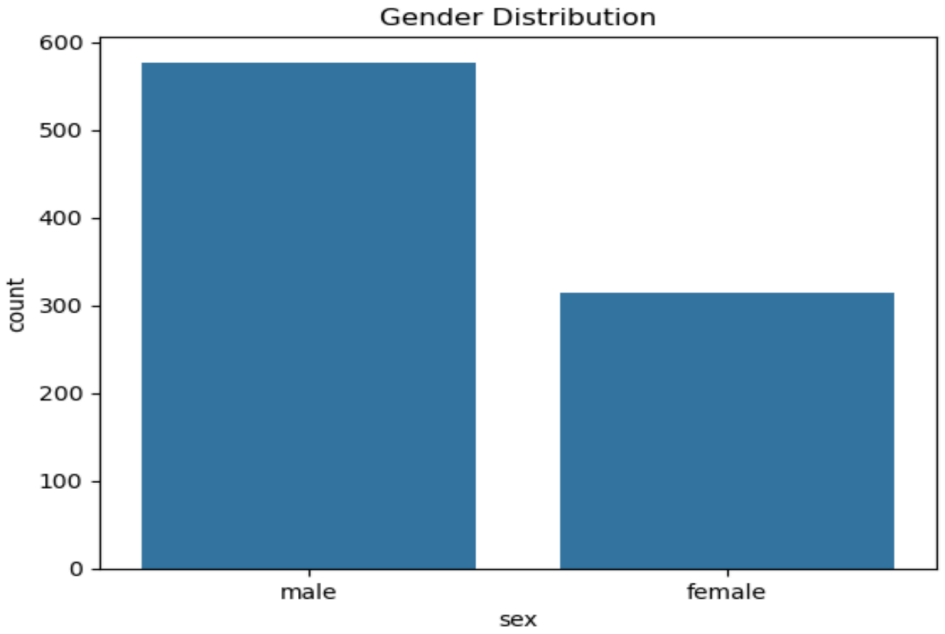


- Shows how many survived vs. not.
- More passengers **did not survive**.

7. Survival by Gender

sns.countplot(x='sex', hue='survived', data=titanic)

```
sns.countplot(data=df, x='sex')
plt.title("Gender Distribution")
plt.show()
```



- **Females survived more** than males.
- Reflects the "Women and children first" policy.

Learning Outcomes

By the end of this session, you learned:
✔ How to use **Seaborn** for advanced statistical plotting.
✔ Different plot types: histogram, bar chart, boxplot, countplot.
✔ How to analyze **categorical vs. numerical data**.
✔ Real insights from Titanic dataset:

- More males than females.
- Higher-class passengers had better survival chances.
- Fare had many outliers (rich passengers).
- Southampton was the main boarding point.