**Practical Labs: Python Pandas + Matplotlib/Seaborn**

**Exercise 1: Getting Started with Pandas**

**Objective:** Load and explore dataset.

import pandas as pd

# Load Titanic dataset

df = pd.read\_csv("titanic.csv")

# Preview data

print(df.head())

print(df.info())

print(df.describe())

# Check missing values

print(df.isnull().sum())

***Task*: Count how many passengers survived vs. not survived.**

**Exercise 2: Data Cleaning & Transformation**

**Objective:** Handle missing values and create new columns.

# Fill missing Age with mean

df['Age'].fillna(df['Age'].mean(), inplace=True)

# Fill Embarked with mode

df['Embarked'].fillna(df['Embarked'].mode()[0], inplace=True)

# Create a new column "FamilySize"

df['FamilySize'] = df['SibSp'] + df['Parch'] + 1

print(df[['Name','FamilySize']].head())

***Task: Find the passenger with the largest family size.***

**Exercise 3: Grouping & Aggregations**

**Objective:** Explore survival rate by groups.

# Survival rate by Gender

print(df.groupby('Sex')['Survived'].mean())

# Survival rate by Passenger Class

print(df.groupby('Pclass')['Survived'].mean())

# Combine multiple groupings

print(df.groupby(['Pclass', 'Sex'])['Survived'].mean())

***Task: Which class and gender had the highest survival rate?***

**Exercise 4: Visualization with Matplotlib**

**Objective:** Create basic plots.

import matplotlib.pyplot as plt

# Plot Age distribution

plt.hist(df['Age'], bins=30, color='skyblue', edgecolor='black')

plt.title("Age Distribution of Titanic Passengers")

plt.xlabel("Age")

plt.ylabel("Count")

plt.show()

# Survival by Class (bar chart)

df['Pclass'].value\_counts().sort\_index().plot(kind='bar', color='orange')

plt.title("Passenger Count by Class")

plt.xlabel("Class")

plt.ylabel("Count")

plt.show()

***Task: Plot survival count per gender as a bar chart.***

**Exercise 5: Visualization with Seaborn**

**Objective:** Use Seaborn for advanced plots.

import seaborn as sns

# Survival countplot by Gender

sns.countplot(data=df, x='Sex', hue='Survived')

plt.title("Survival Count by Gender")

plt.show()

# Boxplot of Age vs. Passenger Class

sns.boxplot(data=df, x='Pclass', y='Age')

plt.title("Age Distribution per Class")

plt.show()

# Heatmap of correlations

sns.heatmap(df.corr(), annot=True, cmap='coolwarm')

plt.title("Correlation Heatmap")

plt.show()

***Task: Plot survival rate by Embarked location.***

**Exercise 6: Sales Data Analysis**

**Dataset:** sales.csv with columns: Date, Region, Product, Sales, Profit

sales = pd.read\_csv("sales.csv")

# Convert Date column to datetime

sales['Date'] = pd.to\_datetime(sales['Date'])

# Monthly Sales

monthly\_sales = sales.groupby(sales['Date'].dt.to\_period("M"))['Sales'].sum()

monthly\_sales.plot(kind='line', marker='o')

plt.title("Monthly Sales Trend")

plt.xlabel("Month")

plt.ylabel("Total Sales")

plt.show()

# Sales by Region (Seaborn barplot)

sns.barplot(data=sales, x='Region', y='Sales', estimator=sum)

plt.title("Total Sales by Region")

plt.show()

***Task: Plot top 5 products by total profit.***

**Exercise 7: Combined Insights**

**Mini Project:**

1. Titanic → Find top 3 features that correlate with survival.
2. Sales → Find best-performing region and month.
3. Create at least 3 charts (bar, line, heatmap).

By completing these labs, you’ll practice:

* Data wrangling with **Pandas**
* Visualization with **Matplotlib & Seaborn**
* Real-world dataset analysis (Titanic + Sales)