Task 5 – Exploratory Data Analysis (Titanic Data

This notebook performs Exploratory Data Analysis (EDA) on the Titanic dataset.

We will:

- Explore the dataset structure
- Check for missing values
- Visualize important features
- Identify patterns and insights

```
In [ ]: import pandas as pd
        import seaborn as sns
        import matplotlib.pyplot as plt

        sns.set(style="whitegrid")
```

```
In [ ]: # Load Titanic dataset (place train.csv in the same directory)
        df = pd.read_csv("train.csv")
        df.head()
```

```
In [ ]: df.info()
```

```
In [ ]: df.describe()
```

```
In [ ]: df.isnull().sum()
```

```
In [ ]: df['Survived'].value_counts()
```

Visualizations

```
In [ ]: plt.figure(figsize=(8,5))
        sns.histplot(df['Age'], kde=True, bins=30, color='skyblue')
        plt.title("Age Distribution")
        plt.show()
```

```
In [ ]: plt.figure(figsize=(6,4))
        sns.countplot(x='Survived', hue='Sex', data=df, palette='Set2')
        plt.title("Survival Count by Gender")
        plt.show()
```

```
In [ ]: plt.figure(figsize=(8,5))
        sns.boxplot(x='Pclass', y='Age', data=df, palette='Pastel1')
        plt.title("Age by Passenger Class")
        plt.show()
```

```
In [ ]: plt.figure(figsize=(10,6))
        sns.heatmap(df.corr(), annot=True, cmap='coolwarm', fmt=".2f")
        plt.title("Correlation Heatmap")
        plt.show()
```

```
In [ ]: plt.figure(figsize=(8,5))
        sns.scatterplot(x='Age', y='Fare', hue='Survived', data=df, palette='coolwarm')
```

```
plt.title("Fare vs Age (Colored by Survival)")
plt.show()
```

Observations

- Survival rate is higher among females than males.
- Younger passengers had better survival rates.
- First-class passengers were more likely to survive.
- Higher fare passengers had higher chances of survival.