

# EDA Example on Titanic Dataset

## Exploratory Data Analysis (EDA) - Titanic Dataset

### 1. Import Required Libraries:

```
-----  
import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns  
import plotly.express as px
```

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

### 2. Load Titanic Dataset:

```
-----  
df = sns.load_dataset('titanic') # or use pd.read_csv('titanic.csv')  
df.head()
```

### 3. Basic Data Exploration:

```
-----  
df.shape  
df.info()  
df.describe()  
df.isnull().sum()
```

### 4. Univariate Analysis:

```
-----  
# Histogram  
df['age'].hist(bins=20)  
plt.title("Age Distribution")  
plt.xlabel("Age")  
plt.ylabel("Frequency")  
plt.show()
```

```
# Boxplot  
sns.boxplot(x='age', data=df)  
plt.title("Boxplot of Age")  
plt.show()
```

### 5. Categorical Features:

```
-----  
sns.countplot(x='sex', data=df)  
plt.title("Count of Passengers by Sex")  
plt.show()  
  
sns.countplot(x='class', data=df, hue='survived')  
plt.title("Survival by Passenger Class")  
plt.show()
```

## EDA Example on Titanic Dataset

### 6. Correlation Matrix:

```
-----  
plt.figure(figsize=(10,6))  
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='Blues')  
plt.title("Correlation Matrix")  
plt.show()
```

### 7. Pairplot:

```
-----  
sns.pairplot(df.dropna(subset=['age']), hue='survived', vars=['age', 'fare', 'pclass'])  
plt.suptitle("Pairwise Relationships", y=1.02)  
plt.show()
```

### 8. Plotly Interactive Visualization:

```
-----  
fig = px.scatter(df, x='age', y='fare', color='survived', hover_data=['sex', 'class'])  
fig.update_layout(title="Age vs Fare (Colored by Survival)")  
fig.show()
```

### 9. Inferences:

- ```
-----  
- Females had a higher survival rate.  
- Survival rate was higher in 1st class than 3rd class.  
- Younger passengers survived more frequently.  
- High fare -> More likely to be 1st class -> Higher survival.  
- Missing values in 'age', 'embark_town', 'deck' -> consider imputation or exclusion.
```