Task 2: Exploratory Data Analysis (EDA)

Tools: pandas, matplotlib, seaborn, plotly

1. Import Libraries & Load Dataset

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

df = sns.load_dataset('titanic')
df.head()
```

2. Generate Summary Statistics

```
df.describe()
df.describe(include=['object'])
df.info()
df.isnull().sum()
```

3. Histograms & Boxplots for Numeric Features

```
plt.figure(figsize=(8,5))
sns.histplot(df['age'].dropna(), kde=True)
plt.title('Age Distribution')
plt.show()

plt.figure(figsize=(8,5))
sns.boxplot(x=df['fare'])
plt.title('Fare Distribution')
```

```
plt.show()
```

4. Pairplot & Correlation Matrix

```
plt.figure(figsize=(10,8))
sns.heatmap(df.corr(), annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
plt.show()

sns.pairplot(df[['age', 'fare', 'pclass', 'survived']].dropna(), hue='survived')
plt.show()
```

5. Identify Patterns, Trends, Anomalies

```
sns.countplot(x='sex', hue='survived', data=df)
plt.title('Survival Count by Gender')
plt.show()

sns.countplot(x='pclass', hue='survived', data=df)
plt.title('Survival Count by Passenger Class')
plt.show()

fig = px.box(df, x='survived', y='fare', color='survived', title="Fare Distribution by Survival")
fig.show()
```

6. Basic Inferences

- Women had higher survival rates than men.
- Passengers in 1st class had higher survival rates.
- There are outliers in fare prices.
- Most passengers were young adults.
- Positive correlation between fare and survival probability.