

Task 2: Exploratory Data Analysis (EDA)

Objective: Understand data using statistics and visualizations.


Tools: Pandas, Matplotlib, Seaborn, Plotly

CODE:


```
# Step 1: Import Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
from google.colab import files
import io

# Step 2: Upload CSV File Manually
uploaded = files.upload()
df = pd.read_csv(io.BytesIO(next(iter(uploaded.values()))))

# Step 3: Basic Info
print("\n ♦ Shape of the dataset:", df.shape)
print("\n ♦ Column names:\n", df.columns.tolist())
print("\n ♦ First 5 rows:\n", df.head())
print("\n ♦ Data types:\n", df.dtypes)
print("\n ♦ Missing values:\n", df.isnull().sum())

# Step 4: Summary Statistics
print("\n  Descriptive Statistics:\n", df.describe(include='all'))

# Step 5: Handle missing values (optional cleanup for visuals)
df_cleaned = df.copy()
df_cleaned['Age'] = df_cleaned['Age'].fillna(df_cleaned['Age'].median())
df_cleaned['Fare'] = df_cleaned['Fare'].fillna(df_cleaned['Fare'].median())
df_cleaned['Embarked'] = df_cleaned['Embarked'].fillna(df_cleaned['Embarked'].mode()[0])

# Step 6: Histograms for Numeric Features
df_cleaned.select_dtypes(include=np.number).hist(bins=30, figsize=(15, 10), edgecolor='black')
plt.suptitle("\n  Histograms of Numeric Features", fontsize=16)
plt.tight_layout()
plt.show()

# Step 7: Boxplots for Numeric Features
numeric_cols = df_cleaned.select_dtypes(include=np.number).columns
```

```
plt.figure(figsize=(15, 8))
for i, col in enumerate(numeric_cols, 1):
    plt.subplot(len(numeric_cols)//3 + 1, 3, i)
    sns.boxplot(x=df_cleaned[col], color='skyblue')
    plt.title(f'Boxplot of {col}')
plt.tight_layout()
plt.show()
```

```
# Step 8: Correlation Matrix (Exclude non-numeric columns)
corr = df_cleaned[numeric_cols].corr()
plt.figure(figsize=(12, 8))
sns.heatmap(corr, annot=True, cmap='coolwarm', fmt=".2f", square=True)
plt.title("🔥 Correlation Matrix Heatmap")
plt.show()
```

```
# Step 9: Pairplot (only for first 5 numeric columns to avoid overload)
sns.pairplot(df_cleaned[numeric_cols[:5]])
plt.suptitle("🔗 Pairplot of Numeric Features", y=1.02)
plt.show()
```

```
# Step 10: Plotly Interactive Scatter Matrix (if at least 3 numeric cols)
if len(numeric_cols) >= 3:
    fig = px.scatter_matrix(df_cleaned,
                           dimensions=numeric_cols[:4],
                           color='Survived',
                           title="🎯 Interactive Scatter Matrix (Plotly)")
    fig.show()
```

```
# Step 11: Categorical Insights (Sex vs Survived, Embarked vs Survived)
plt.figure(figsize=(10,4))
sns.countplot(data=df_cleaned, x='Sex', hue='Survived')
plt.title("👤 Sex vs Survival")
plt.show()
```

```
plt.figure(figsize=(10,4))
sns.countplot(data=df_cleaned, x='Embarked', hue='Survived')
plt.title("🚢 Embarkation Port vs Survival")
plt.show()
```

```
# Step 12: Final Message
print("\n✅ EDA Complete! Explore the visual patterns for deeper feature-level insights.")
```

OUTPUT:

Titanic-Dataset.csv(text/csv) - 61194 bytes, last modified: 24/12/2021 - 100% done

Saving Titanic-Dataset.csv to Titanic-Dataset (1).csv

- ◆ Shape of the dataset: (891, 12)

- ◆ Column names:

['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp', 'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked']

- ◆ First 5 rows:

	PassengerId	Survived	Pclass \
0	1	0	3
1	2	1	1
2	3	1	3
3	4	1	1
4	5	0	3

	Name	Sex	Age	SibSp \
0	Braund, Mr. Owen Harris	male	22.0	1
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1
2	Heikkinen, Miss. Laina	female	26.0	0
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1
4	Allen, Mr. William Henry	male	35.0	0

	Parch	Ticket	Fare	Cabin	Embarked
0	0	A/5 21171	7.2500	NaN	S
1	0	PC 17599	71.2833	C85	C
2	0	STON/O2. 3101282	7.9250	NaN	S
3	0	113803	53.1000	C123	S
4	0	373450	8.0500	NaN	S

- ◆ Data types:

PassengerId	int64
Survived	int64
Pclass	int64
Name	object
Sex	object
Age	float64
SibSp	int64
Parch	int64
Ticket	object
Fare	float64
Cabin	object
Embarked	object
dtype:	object

◆ Missing values:

```
PassengerId    0
Survived        0
Pclass          0
Name            0
Sex             0
Age            177
SibSp           0
Parch           0
Ticket          0
Fare            0
Cabin          687
Embarked        2
dtype: int64
```

 Descriptive Statistics:

	PassengerId	Survived	Pclass	Name	Sex \
count	891.000000	891.000000	891.000000		891 891
unique	NaN	NaN	NaN	891	2
top	NaN	NaN	NaN	Dooley, Mr. Patrick	male
freq	NaN	NaN	NaN	1	577
mean	446.000000	0.383838	2.308642		NaN NaN
std	257.353842	0.486592	0.836071		NaN NaN
min	1.000000	0.000000	1.000000		NaN NaN
25%	223.500000	0.000000	2.000000		NaN NaN
50%	446.000000	0.000000	3.000000		NaN NaN
75%	668.500000	1.000000	3.000000		NaN NaN
max	891.000000	1.000000	3.000000		NaN NaN

	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
count	714.000000	891.000000	891.000000	891	891.000000	204	889
unique	NaN	NaN	NaN	681	NaN	147	3
top	NaN	NaN	NaN	347082	NaN	G6	S
freq	NaN	NaN	NaN	7	NaN	4	644
mean	29.699118	0.523008	0.381594	NaN	32.204208	NaN	NaN
std	14.526497	1.102743	0.806057	NaN	49.693429	NaN	NaN
min	0.420000	0.000000	0.000000	NaN	0.000000	NaN	NaN
25%	20.125000	0.000000	0.000000	NaN	7.910400	NaN	NaN
50%	28.000000	0.000000	0.000000	NaN	14.454200	NaN	NaN
75%	38.000000	1.000000	0.000000	NaN	31.000000	NaN	NaN
max	80.000000	8.000000	6.000000	NaN	512.329200	NaN	NaN

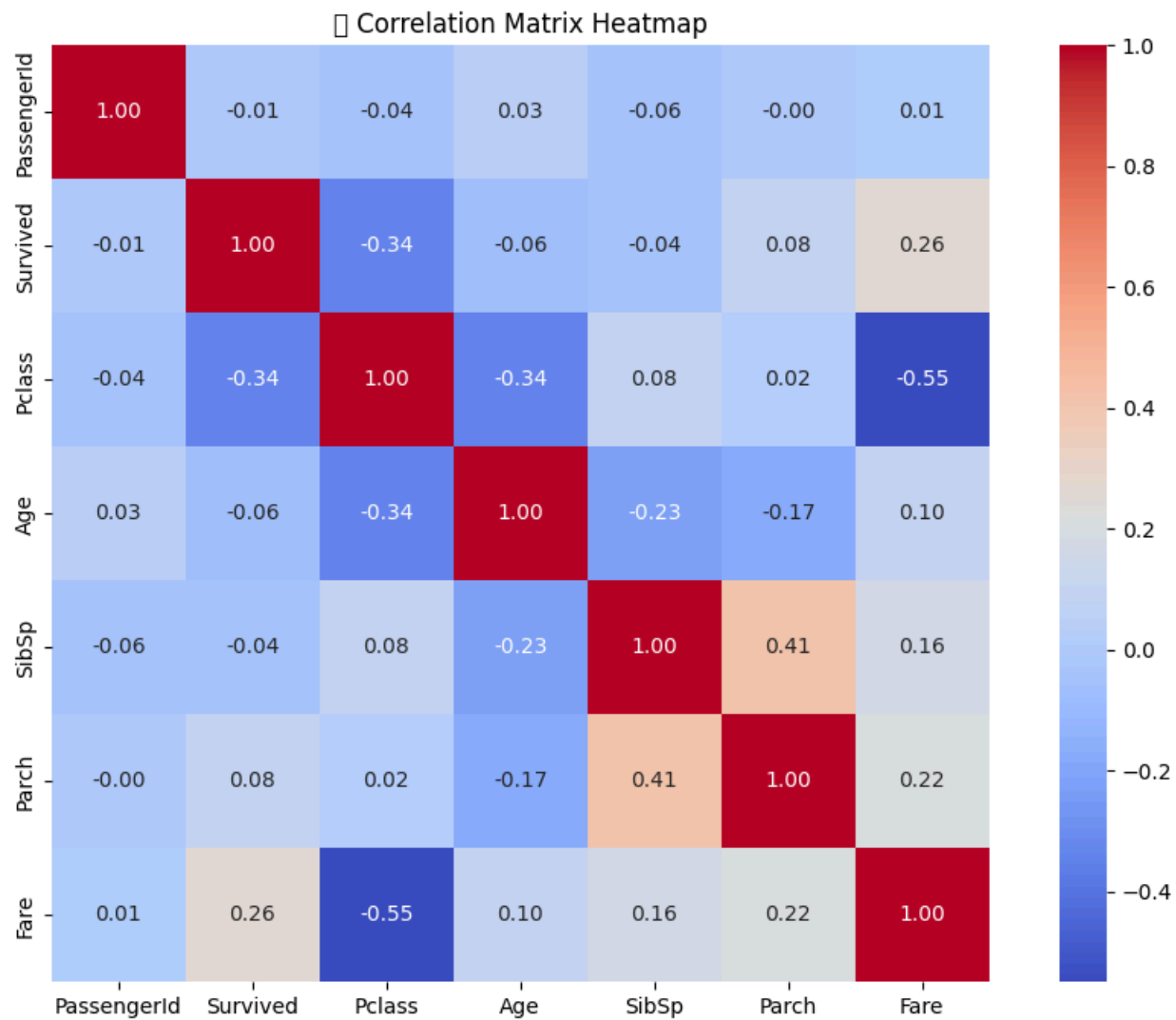
/tmp/ipython-input-2-2987135943.py:33: UserWarning: Glyph 128202 (N{BAR CHART}) missing from font(s) DejaVu Sans.

plt.tight_layout()

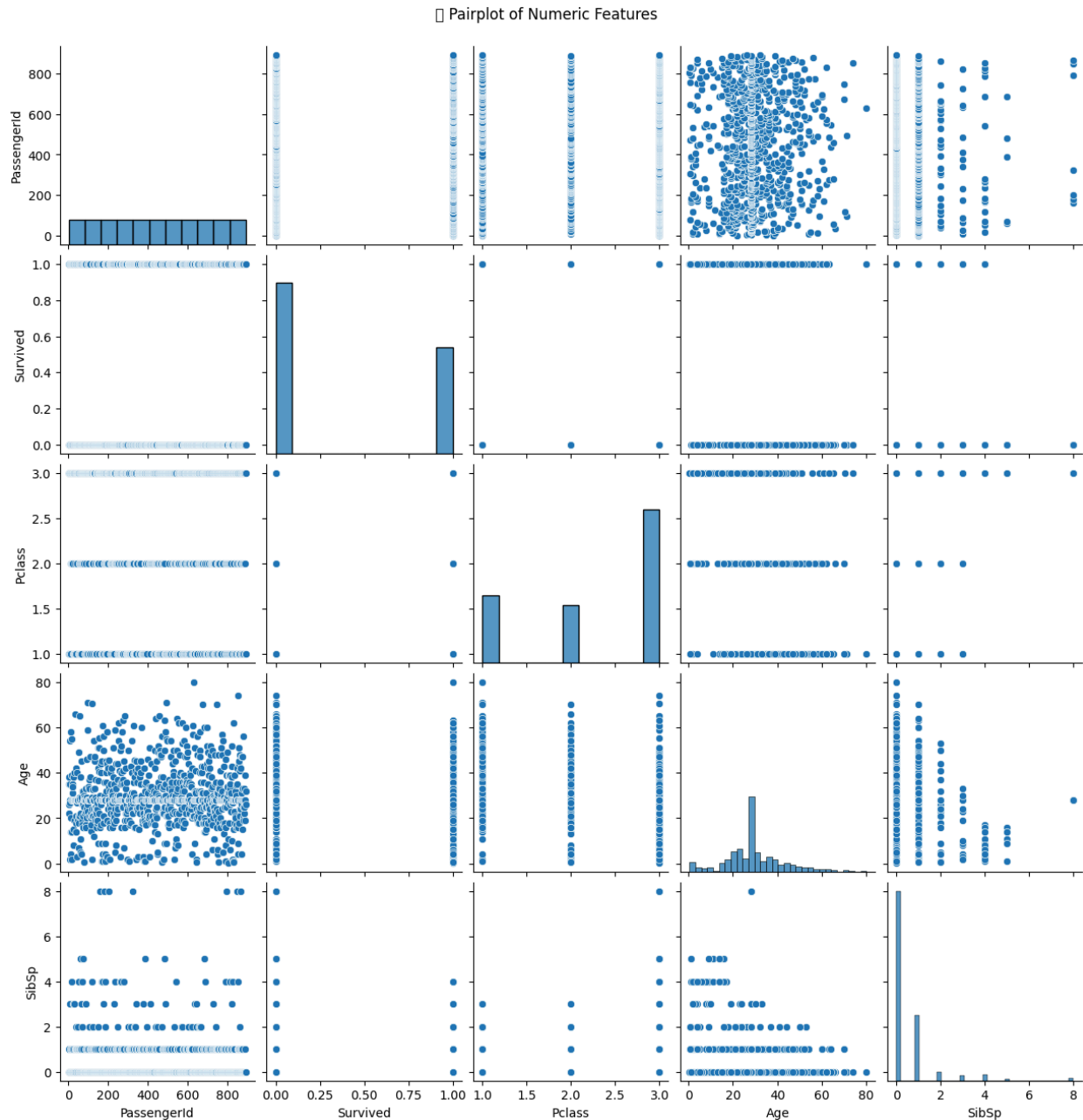
/usr/local/lib/python3.11/dist-packages/IPython/core/pylabtools.py:151: UserWarning: Glyph 128202 ($\bar{\text{BAR CHART}}$) missing from font(s) DejaVu Sans.
fig.canvas.print_figure(bytes_io, **kw)



/usr/local/lib/python3.11/dist-packages/IPython/core/pylabtools.py:151: UserWarning: Glyph 128204 ($\{PUSHPIN\}$) missing from font(s) DejaVu Sans.
fig.canvas.print_figure(bytes_io, **kw)



/usr/local/lib/python3.11/dist-packages/IPython/core/pylabtools.py:151: UserWarning: Glyph 128279 ($\{LINK SYMBOL\}$) missing from font(s) DejaVu Sans.
fig.canvas.print_figure(bytes_io, **kw)



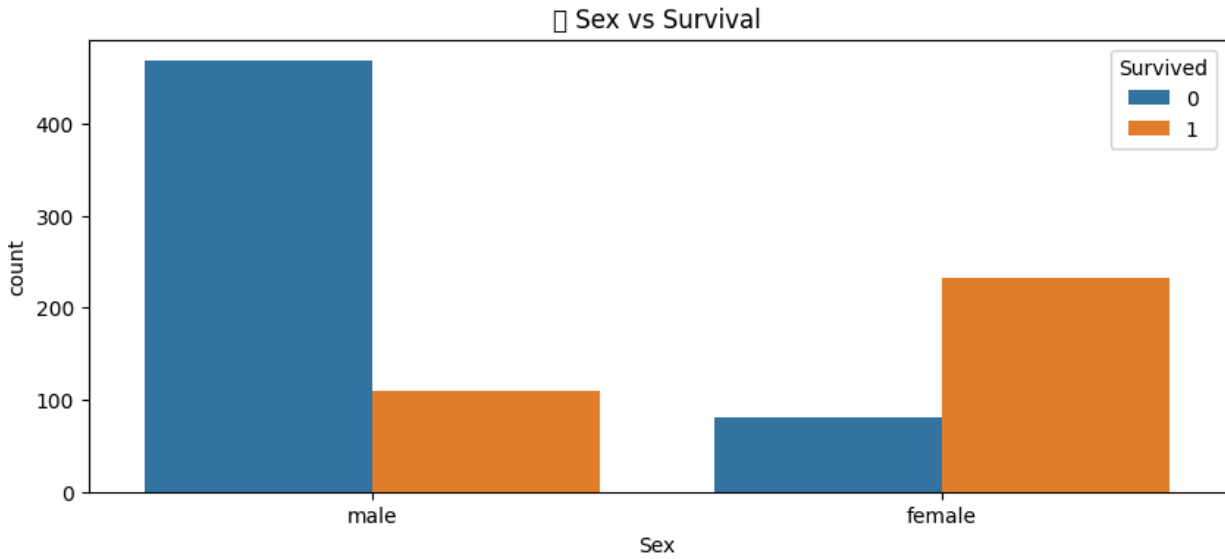
050000.5111.522.53050002040608000.5111.522.53020406080

00.20.40.60.81Survived  Interactive Scatter Matrix

(Plotly)PassengerIdSurvivedPclassAgePassengerIdSurvivedPclassAge

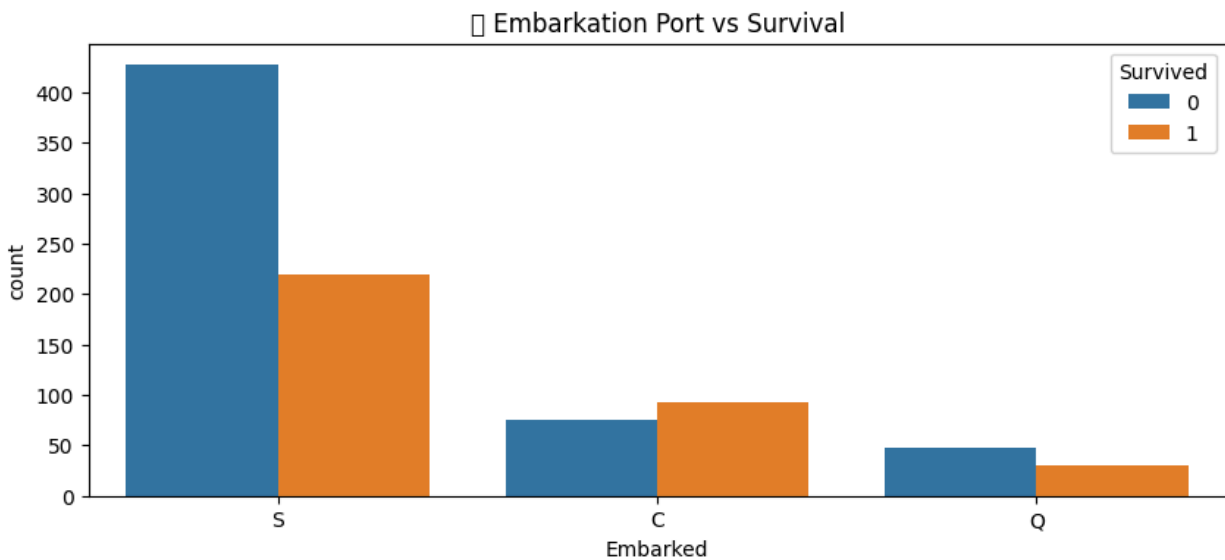
/usr/local/lib/python3.11/dist-packages/IPython/core/pylabtools.py:151: UserWarning:

Glyph 129485 (\N{STANDING PERSON}) missing from font(s) DejaVu Sans.



/usr/local/lib/python3.11/dist-packages/IPython/core/pylabtools.py:151: UserWarning:

Glyph 128755 (\N{PASSENGER SHIP}) missing from font(s) DejaVu Sans.



✓ EDA Complete! Explore the visual patterns for deeper feature-level insights.