```
In [ ]: # Titanic EDA Notebook
        # Task 5 - Data Analyst Internship
        # --- 1. Import Libraries ---
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns
        import os
        # Set plot style
        sns.set(style="whitegrid")
        # --- 2. Load Dataset ---
        print("Current Working Directory:", os.getcwd())
        df = pd.read csv("train.csv")
        # --- 3. Basic Info ---
        print("\n--- Dataset Info ---")
        df.info()
        print("\n--- First 5 Rows ---")
        print(df.head())
        print("\n--- Missing Values ---")
        print(df.isnull().sum())
        print("\n--- Summary Statistics ---")
        print(df.describe())
        # --- 4. Handle Missing Values (simple fix for demo) ---
        # Fill Age with median, Embarked with mode
        df['Aqe'].fillna(df['Age'].median(), inplace=True)
        df['Embarked'].fillna(df['Embarked'].mode()[0], inplace=True)
        # Drop Cabin (too many missing values)
        df.drop(columns=['Cabin'], inplace=True)
        # --- 5. Univariate Analysis ---
        # Categorical: Gender distribution
        sns.countplot(x='Sex', data=df)
        plt.title("Gender Distribution")
        plt.show()
        # Numerical: Age distribution
        df['Age'].hist(bins=30, edgecolor='black')
        plt.title("Age Distribution")
        plt.xlabel("Age")
        plt.ylabel("Count")
        plt.show()
        # Boxplot: Fare
        sns.boxplot(x='Fare', data=df)
        plt.title("Fare Boxplot")
        plt.show()
        # --- 6. Bivariate Analysis ---
        # Survival rate by gender
        sns.countplot(x='Survived', hue='Sex', data=df)
        plt.title("Survival by Gender")
        plt.show()
        # Survival rate by class
        sns.countplot(x='Survived', hue='Pclass', data=df)
        plt.title("Survival by Passenger Class")
        plt.show()
        # Scatter plot: Age vs Fare (colored by survival)
        sns.scatterplot(x='Age', y='Fare', hue='Survived', data=df)
        plt.title("Age vs Fare by Survival")
        plt.show()
        # --- 7. Multivariate Analysis ---
        # Pairplot
        sns.pairplot(df[['Survived', 'Pclass', 'Age', 'Fare']], hue='Survived')
        plt.suptitle("Pairplot of Numeric Features", y=1.02)
        plt.show()
        # Heatmap (correlation)
        numeric_df = df.select_dtypes(include=['int64', 'float64'])
```

```
plt.figure(figsize=(10,6))
         sns.heatmap(numeric_df.corr(), annot=True, cmap='coolwarm')
         plt.title("Correlation Heatmap")
         plt.show()
In [10]: import pandas as pd
         import matplotlib.pyplot as plt
         import seaborn as sns
         import os
         # Set plot style
         sns.set(style="whitegrid")
         # --- 2. Load Dataset ---
         print("Current Working Directory:", os.getcwd())
         df = pd.read_csv("train.csv")
         # --- 3. Basic Info ---
         print("\n--- Dataset Info ---")
         df.info()
         print("\n--- First 5 Rows ---")
         print(df.head())
         print("\n--- Missing Values ---")
         print(df.isnull().sum())
         print("\n--- Summary Statistics ---")
         print(df.describe())
         # --- 4. Handle Missing Values (simple fix for demo) ---
         # Fill Age with median, Embarked with mode
         df['Age'].fillna(df['Age'].median(), inplace=True)
         df['Embarked'].fillna(df['Embarked'].mode()[0], inplace=True)
         # Drop Cabin (too many missing values)
         df.drop(columns=['Cabin'], inplace=True)
         # --- 5. Univariate Analysis ---
         # Categorical: Gender distribution
         sns.countplot(x='Sex', data=df)
         plt.title("Gender Distribution")
         plt.show()
         # Numerical: Age distribution
         df['Age'].hist(bins=30, edgecolor='black')
         plt.title("Age Distribution")
         plt.xlabel("Age")
         plt.ylabel("Count")
         plt.show()
         # Boxplot: Fare
         sns.boxplot(x='Fare', data=df)
         plt.title("Fare Boxplot")
         plt.show()
         # --- 6. Bivariate Analysis ---
         # Survival rate by gender
         sns.countplot(x='Survived', hue='Sex', data=df)
         plt.title("Survival by Gender")
         plt.show()
         # Survival rate by class
         sns.countplot(x='Survived', hue='Pclass', data=df)
         plt.title("Survival by Passenger Class")
         plt.show()
         # Scatter plot: Age vs Fare (colored by survival)
         sns.scatterplot(x='Age', y='Fare', hue='Survived', data=df)
         plt.title("Age vs Fare by Survival")
         plt.show()
         # --- 7. Multivariate Analysis ---
         sns.pairplot(df[['Survived', 'Pclass', 'Age', 'Fare']], hue='Survived')
         plt.suptitle("Pairplot of Numeric Features", y=1.02)
         plt.show()
         # Heatmap (correlation)
         numeric_df = df.select_dtypes(include=['int64', 'float64'])
```

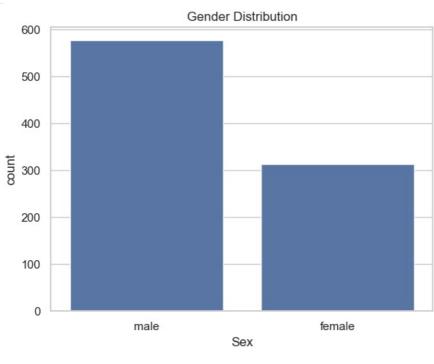
```
plt.figure(figsize=(10,6))
sns.heatmap(numeric_df.corr(), annot=True, cmap='coolwarm')
plt.title("Correlation Heatmap")
plt.show()
```

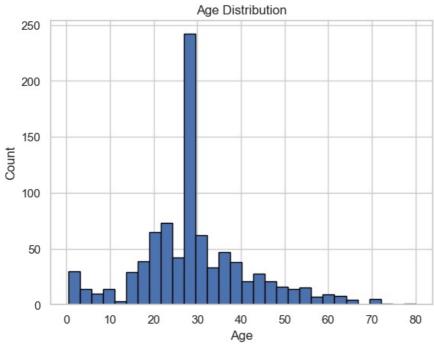
```
--- Dataset Info ---
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
#
     Column
                  Non-Null Count
                                   Dtype
- - -
     -----
                  -----
                  891 non-null
                                   int64
0
     PassengerId
1
     Survived
                  891 non-null
                                   int64
                  891 non-null
2
     Pclass
                                   int64
3
     Name
                  891 non-null
                                   object
     Sex
                  891 non-null
                                   object
5
     Age
                  714 non-null
                                   float64
 6
                  891 non-null
     SibSp
                                   int64
                  891 non-null
     Parch
7
                                   int64
 8
     Ticket
                  891 non-null
                                   object
9
                  891 non-null
                                   float64
     Fare
 10
    Cabin
                  204 non-null
                                   object
                  889 non-null
11 Embarked
                                   object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
--- First 5 Rows ---
   PassengerId Survived
                          Pclass
0
                       0
             1
                                3
1
             2
                       1
                                1
2
             3
                                3
                       1
3
             4
                                1
                       1
             5
4
                       0
                                3
                                                           Sex
                                                                      SibSp
                                                  Name
                                                                 Aae
0
                              Braund, Mr. Owen Harris
                                                          male
                                                                22.0
                                                                           1
1
  Cumings, Mrs. John Bradley (Florence Briggs Th...
                                                                38.0
                                                        female
                                                                           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
   Parch
                    Ticket
                                Fare Cabin Embarked
0
                 A/5 21171
                              7.2500
                                       NaN
                                                  S
       0
1
                  PC 17599
                             71.2833
                                       C85
                                                   C
2
       0
          STON/02. 3101282
                             7.9250
                                       NaN
                                                   S
3
       0
                     113803
                             53.1000
                                      C123
                                                   S
                    373450
                                                   S
4
       0
                             8.0500
                                       NaN
--- Missing Values ---
PassengerId
                 0
Survived
                 0
Pclass
                 0
                 0
Name
Sex
                 0
               177
Age
SibSp
                 0
Parch
                 0
Ticket
                 0
Fare
                 0
Cabin
               687
Embarked
                 2
dtype: int64
--- Summary Statistics ---
       PassengerId
                      Survived
                                     Pclass
                                                               SibSp \
                                                     Age
        891.000000 891.000000
                                 891.000000
                                             714.000000
                                                          891.000000
count
        446.000000
                      0.383838
                                                            0.523008
mean
                                   2.308642
                                              29.699118
                                   0.836071
std
        257.353842
                      0.486592
                                              14.526497
                                                            1.102743
                      0.000000
                                   1.000000
min
          1.000000
                                               0.420000
                                                            0.000000
                                                            0.000000
        223.500000
                      0.000000
                                   2.000000
                                               20.125000
25%
50%
        446.000000
                      0.000000
                                   3.000000
                                              28.000000
                                                            0.000000
                      1.000000
                                   3.000000
                                              38.000000
                                                            1.000000
75%
        668.500000
        891.000000
                       1.000000
                                   3.000000
                                              80.000000
                                                            8.000000
max
            Parch
                          Fare
count 891.000000
                   891.000000
         0.381594
                    32.204208
mean
std
         0.806057
                    49.693429
min
         0.000000
                     0.000000
         0.000000
                     7.910400
25%
                    14.454200
50%
         0.000000
                    31.000000
75%
         0.000000
```

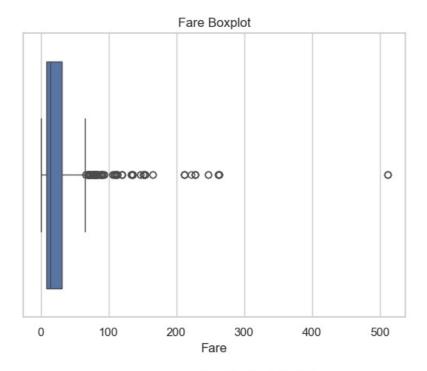
6.000000

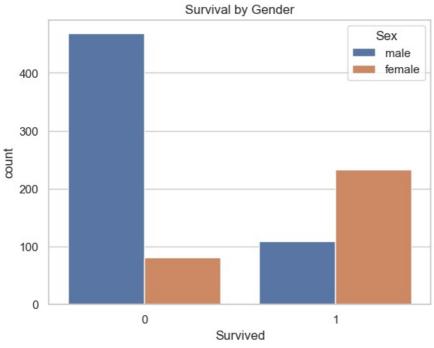
max

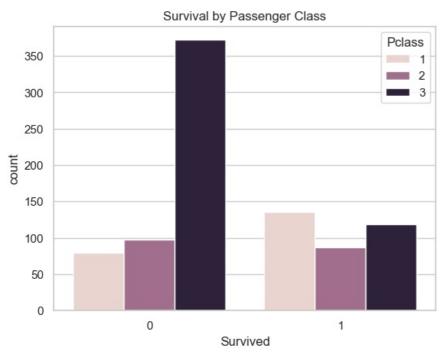
512.329200

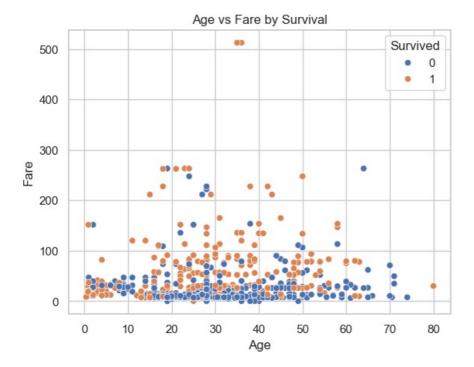


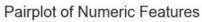


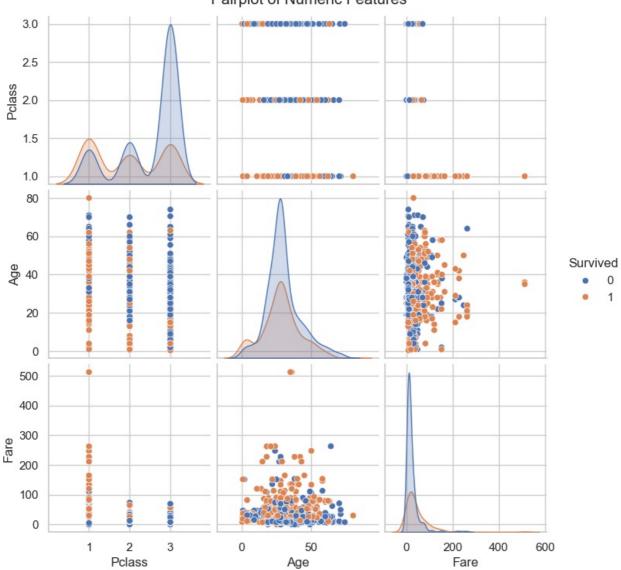












Correlation Heatmap									1.0
Passengerld	1	-0.005	-0.035	0.034	-0.058	-0.0017	0.013		- 0.8
Survived	-0.005	1	-0.34	-0.065	-0.035	0.082	0.26		- 0.6
Pclass	-0.035	-0.34	1	-0.34	0.083	0.018	-0.55		- 0.4
Age	0.034	-0.065	-0.34	1	-0.23	-0.17	0.097		- 0.2
SibSp	-0.058	-0.035	0.083	-0.23	1	0.41	0.16		- 0.0
Parch	-0.0017	0.082	0.018	-0.17	0.41	1	0.22		0.2
Fare	0.013	0.26	-0.55	0.097	0.16	0.22	1		0.4
	Passengerld	Survived	Pclass	Age	SibSp	Parch	Fare		_