Import Libraries and Load Dataset In [44]: # Import necessary libraries import pandas as pd import numpy as np import matplotlib.pyplot as plt import seaborn as sns In [46]: # Load the Titanic Dataset df = pd.read_csv('Titanic-Dataset.csv') In [48]: df.head() Out[48]: Fare Cabin Embarked PassengerId Survived Pclass Sex Age SibSp Parch Ticket S Braund, Mr. Owen Harris male 22.0 A/5 21171 7.2500 1 Cumings, Mrs. John Bradley (Florence Briggs Th... female 38.0 С PC 17599 71.2833 2 3 3 Heikkinen, Miss. Laina female 26.0 S 0 STON/O2. 3101282 7.9250 Futrelle, Mrs. Jacques Heath (Lily May Peel) female 35.0 113803 53.1000 C123 S 0 S Allen, Mr. William Henry male 35.0 373450 8.0500 Data Cleaning In [51]: # Check for missing values print(df.isnull().sum()) PassengerId Survived 0 Pclass Name 0 Sex 177 Age Parch Ticket Fare 0 Cabin Embarked 2 dtype: int64

In [53]: # Fill missing Age values with mean df['Age']=df['Age'].fillna(df['Age'].mean()) In [55]: print(df.isnull().sum()) PassengerId Survived 0 Pclass 0 Name Age SibSp Parch Ticket Fare Cabin Embarked dtype: int64 In [57]: # Drop irrelevant columns df.drop(columns=['Cabin', 'Ticket', 'Name', 'PassengerId'])

Survived Pclass Sex Age SibSp Parch Fare Embarked 3 male 22.000000 S 0 7.2500 1 female 38.000000 0 71.2833 С 2 3 female 26.000000 0 7.9250 S

1 female 35.000000 0 53.1000 S S 3 male 35.000000 0 8.0500 886 0 13.0000 S 2 male 27.000000 1 female 19.000000 0 30.0000 S 888 3 female 29.699118 2 23.4500 S С 1 male 26.000000 0 30.0000 890 3 male 32.000000 0 7.7500 Q 891 rows × 8 columns In [59]: # Drop rows with missing Embarked df = df.dropna(subset=['Embarked'])

In [61]: print(df.isnull().sum()) PassengerId Survived

687

687

0

Pclass Name Sex Age SibSp Parch Ticket Fare

Cabin Embarked

Survived Pclass Name Sex Age SibSp Parch Ticket Fare Cabin

Embarked Family_Size

dtype: int64 In [63]: # Create New Features/Column # Calculate Family Size df['Family_Size'] = df['SibSp'] + df['Parch'] In [65]: print(df.isnull().sum()) PassengerId

dtype: int64 In [67]: df.head() PassengerId Survived Pclass Sex Age SibSp Parch Fare Cabin Embarked Family_Size **Ticket** Name Braund, Mr. Owen Harris male 22.0 A/5 21171 7.2500 1 Cumings, Mrs. John Bradley (Florence Briggs Th... female 38.0 PC 17599 71.2833 Heikkinen, Miss. Laina female 26.0 S 0 STON/O2. 3101282 7.9250 0 Futrelle, Mrs. Jacques Heath (Lily May Peel) female 35.0 113803 53.1000 C123 S 5 0 S Allen, Mr. William Henry male 35.0 373450 8.0500 NaN

In [70]: # 1.Survival Rate by Age Group # Create Age Group bins = [0, 12, 18, 35, 60, 100]labels = ['Child', 'Teen', 'Adult', 'Middle Age', 'Senior']

Name: Survived, dtype: float64

Embarked C 0.553571 Q 0.389610

Family_Size

0.300935 0.552795

Analysis Questions

df['Age_Group'] = pd.cut(df['Age'], bins=bins, labels=labels) age_group_survival = df.groupby('Age_Group')['Survived'].mean() print("Survival Rate by Age Group:\n", age_group_survival) Survival Rate by Age Group: Age_Group Child 0.579710 Teen 0.428571 0.353271 Adult Middle Age 0.396907 0.190476

In [72]: # 2. Survival Rate by Embarkation Port embark_survival = df.groupby('Embarked')['Survived'].mean() print("\nSurvival Rate by Embarkation Port:\n", embark_survival) Survival Rate by Embarkation Port:

C:\Users\Akshay Besekar\AppData\Local\Temp\ipykernel_1196\1845902966.py:7: FutureWarning: The default of observed=False is deprecated and will be changed to True in a future version of pandas. Pass observed=False to retain

S 0.336957 Name: Survived, dtype: float64 In [74]: # 3. Survival Rate by Family Size family_survival = df.groupby('Family_Size')['Survived'].mean() print("\nSurvival Rate by Family Size:\n", family_survival) Survival Rate by Family Size:

age_group_survival = df.groupby('Age_Group')['Survived'].mean()

current behavior or observed=True to adopt the future default and silence this warning.

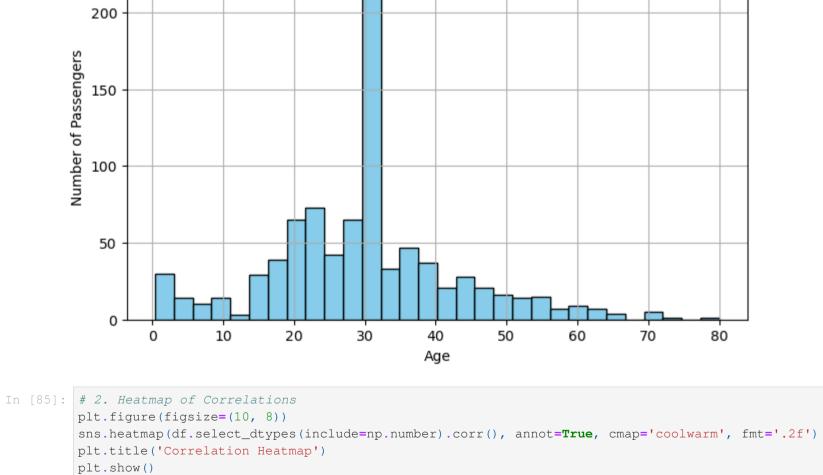
0.578431 0.724138 0.200000 0.136364 0.333333 0.000000 10 0.000000 Name: Survived, dtype: float64

plt.hist(df['Age'], bins=30, color='skyblue', edgecolor='black') plt.title('Age Distribution of Passengers')

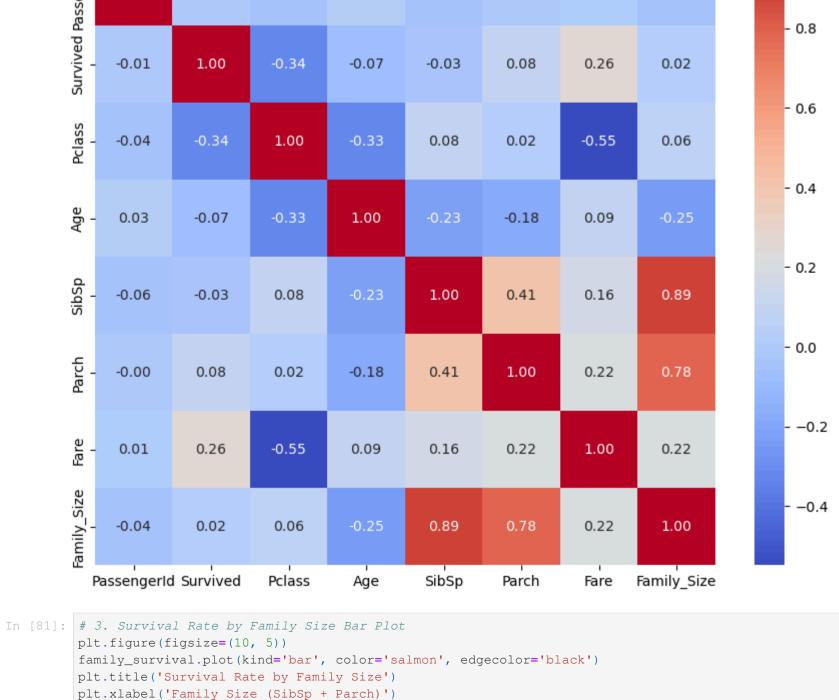
In [77]: # 1. Age Distribution Histogram plt.figure(figsize=(8, 5))

Data Visualizations

plt.xlabel('Age') plt.ylabel('Number of Passengers') plt.grid(True) plt.show() Age Distribution of Passengers 250



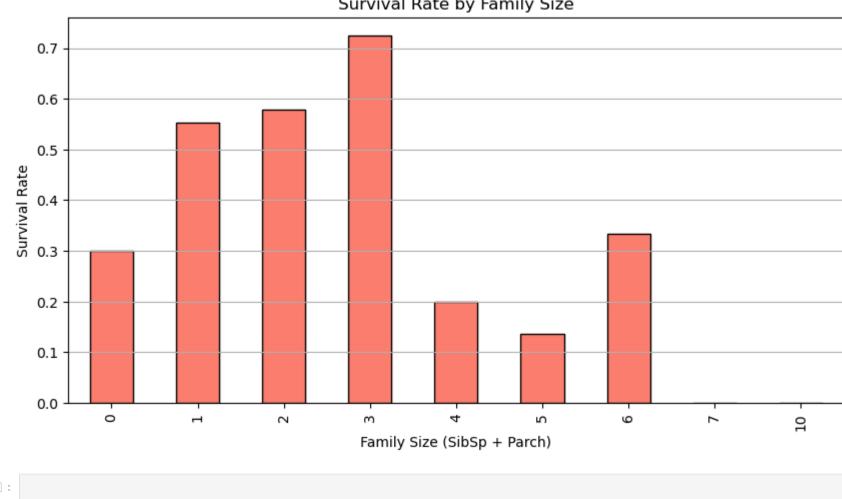
Correlation Heatmap Survived Passengerld -0.01 -0.00 0.01 1.00 -0.04 0.03 -0.06 -0.04



plt.show() Survival Rate by Family Size 0.7

plt.ylabel('Survival Rate')

plt.grid(axis='y')



In []:
In []: