Date Page

08) Exploratory Data Analysis (EDA): Visualize datasets using Matplotlib & Seaborn, Identify trends, Outliers & Correlations.

```
Import pandas as pd
Import matplotlib.pyplot as plt
Import seaborn as sns.

sns.set_style('whitegrid')
plt.rcParam['figure.figsize'] = (10,6)
df = sns.load_dataset('titanic')
print('Dataset Head:')
print(df.head())
print("\nDataset Info:")
print(df.info())
print("\nSummary Statistics:")
Print(df.describe())
Print("\nMissing Values:")
Print(df.isnull().sum())
Plt.figure()
sns.histplot(df['age'], bins=30, kde=True,
  color='blue')
plt.title('Age distribution')
Plt.xlabel('Age')
Plt.ylabel('Count')
Plt.show()
Plt.figure()
sns.countplot(x='class', data=df, palette=
'Set2')
Plt.title('Passenger class distribution')
Plt.xlabel('Passenger class')
```

```
plt.ylabel('Count')
plt.show
plt.figure
sns.boxplot(x='class', y='age', data=df,
palette='Set3')
plt.title('Age distribution by Passenger
          class')
plt.show()
Plt.figure()
Sns.lineplot(x='age', y='Survived', data=df)
plt.title('Survival Trend by Age')
Plt.xlabel('Age')
Plt.ylabel('Survival Rate')
Plt.show()
numerical_cols=df.select_dtypes(include=
  ['float64', 'int64']).columns
Corr=df[numerical_cols].corr()
plt.figure()
Sns.heatmap(Corr, annot=True, cmap=
'coolwarm', center=0)
plt.title('fare vs Age (colored by Survival)')
plt.show()
```

OUTPUT:-

Dataset Head:

	Survived	Pclass	sex	age	parch	fare	E. C
0	0	3	male	22.0	0	7.25000	S Th
1	1	1	female	28.0	0	71.2833	C fi
2	1	3	female	26.0	0	7.9250	S Th
3	1	1	female	35.0	0	53.1000	S fi
4	0	3	male	35.0	0	8.05000	S Th

Date Page

	who	adult_male	deck	embark_town	alive	alone
0	man	True	Nan	Southampton	no	False
1	woman	False	C	Cherbourg	yes	False
2	woman	False	Nan	Southampton	yes	True
3	woman	False	C	Southampton	yes	False
4	man	True	Nan	Southampton	no	True

Dataset Info:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 891 entries, 0 to 890

Data columns (total 15 columns):

#	Column	Non-Null Count	Dtype
0	Survived	891 non-null	Int64
1	pclass	891 non-null	Int64
2	sex	891 non-null	Object
3	age	714 non-null	float64
4	sibsp	891 non-null	Int64
5	parch	891 non-null	Int64
6	fare	891 non-null	float64
7	Embarked	889 non-null	Object
8	class	891 non-null	Category
9	who	891 non-null	Object
10	adult_male	891 non-null	bool
11	alive	891 non-null	Object
12	alone	891 non-null	bool
13	deck	203 non-null	Category
14	embark_town	889 non-null	Object

dtypes: bool(2), Category(2), float64(2), Int64(4), Object(5).

memory usage : 80.7+ KB

None

Summary Statistics :

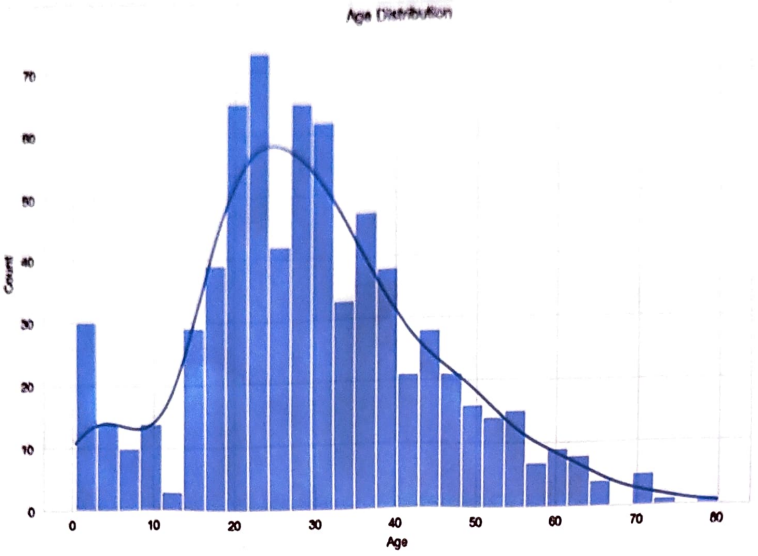
	Survived	pclass	age	sibsp	parch	fare
count	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	1.000000	4.000000	38.000000	1.000000	0.000000	31.000000
max	1.000000	5.000000	80.000000	8.000000	6.000000	512.329200

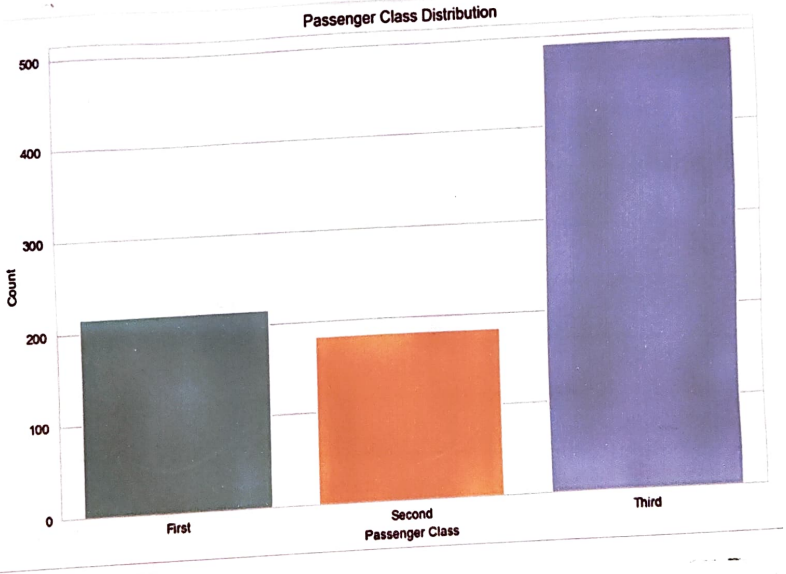
Date Page

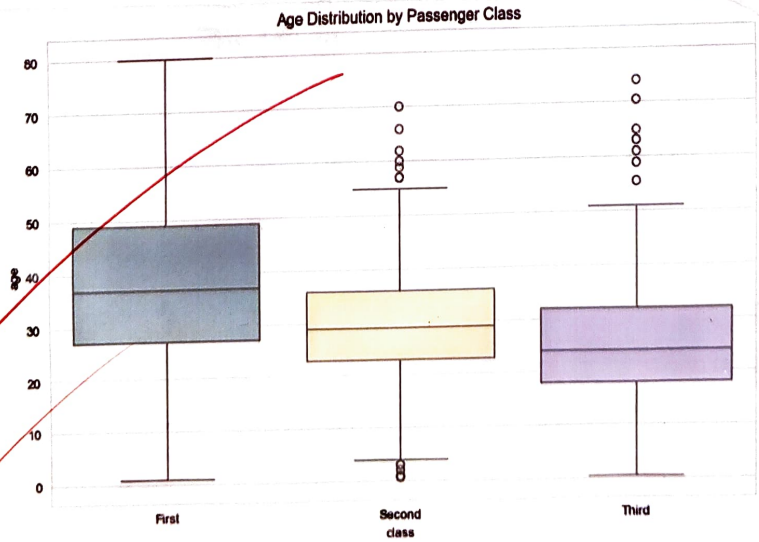
Missing values :

Survived	0
pclass	0
sex	0
age	177
sibsp	0
parch	0
fare	0
Embarked	2
class	0
who	0
adult_male	0
deck	688
Embark_town	2
alive	0
alone	0

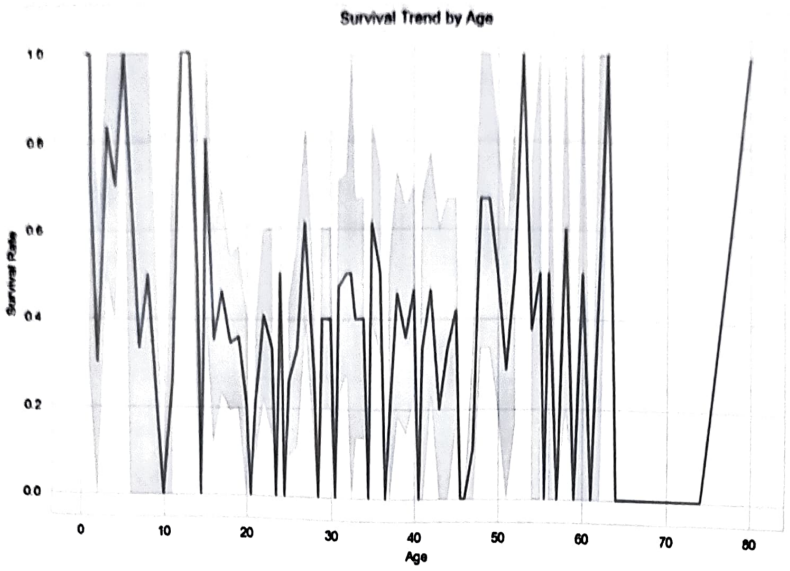
dtype : int64

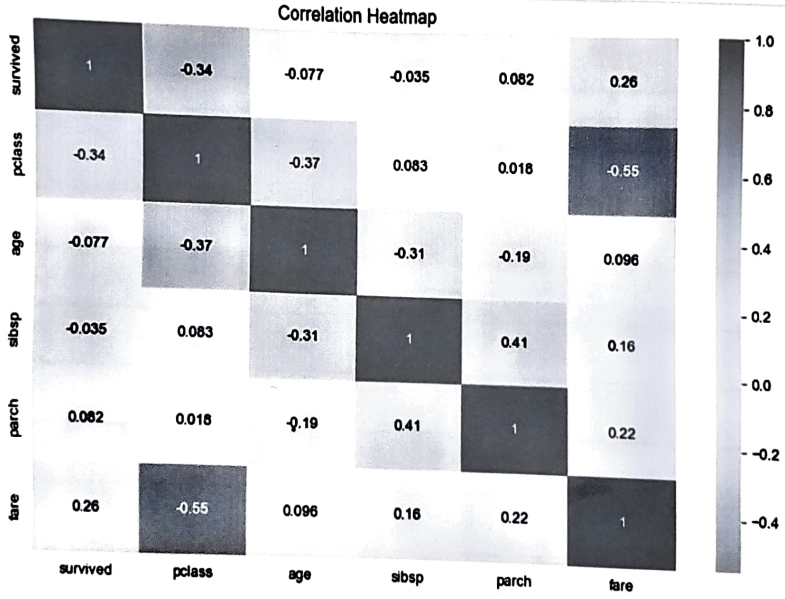

Age Distribution
Count
Age
0
10
20
30
40
50
60
70
80

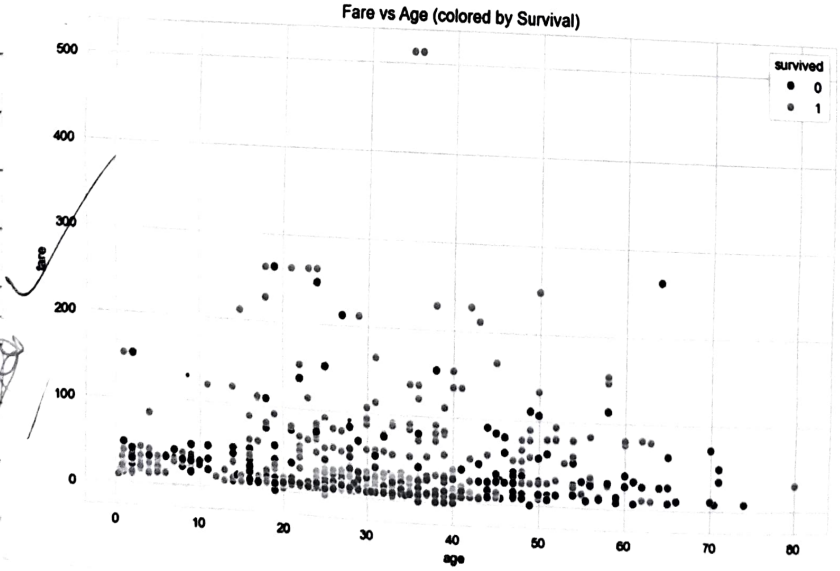

Passenger Class Distribution
Count
Passenger Class
First
Second
Third
0
100
200
300
400
500


Age Distribution by Passenger Class
age
First
Second class
Third
0
10
20
30
40
50
60
70
80

Date Page


Survival Trend by Age
Survival Rate
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
60
70
80
Age


Correlation Heatmap
survived
pclass
age
sibsp
parch
fare
1
-0.34
-0.077
-0.035
0.082
0.26
-0.34
1
-0.37
0.083
0.018
-0.55
-0.077
-0.37
1
-0.31
-0.19
0.096
-0.035
0.083
-0.31
1
0.41
0.16
0.082
0.018
-0.19
0.41
1
0.22
0.26
-0.55
0.096
0.16
0.22
1
survived
pclass
age
sibsp
parch
fare
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4


Fare vs Age (colored by Survival)
survived
0
1
fare
500
400
300
200
100
0
0
10
20
30
40
50
60
70
80
age