

Name: Shiv kumar paul
Dataset: Titanic Dataset

Tools: Python (Jupyter Notebook)
Email: rubelpaul990@gmail.com

Summary:

The Titanic dataset contains information about **1309 passengers** aboard the RMS Titanic, and is commonly used for classification tasks such as predicting survival. The dataset includes both numerical and categorical features.

Column	Description
PassengerId	Unique ID for each passenger
Survived	Survival (0 = No, 1 = Yes)
Pclass	Ticket class (1 = 1st, 2 = 2nd, 3 = 3rd)
Name	Name of the passenger
Sex	Gender
Age	Age in years
SibSp	siblings/spouses aboard
Parched	parents/children aboard
Ticket	Ticket number
Fare	Ticket fare
Embarked	Port of embarkation (C = Cherbourg, Q = Queenstown, S = Southampton)

Analyzing the data in Jupyter Notebook

1. Importing Libraries

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

2. Reading the files with the alias name 'df1' & and identifying by using df1.head()

```
df1 = pd.read_csv('gender_submission.csv')
```

```
df1.head()
```

	PassengerId	Survived
0	892	0
1	893	1
2	894	0
3	895	0
4	896	1

3. Finding missing value

```
49]: df1.isnull().sum()

49]: PassengerId      0
     Survived       0
     dtype: int64

61]: df1.duplicated()

61]: 0      False
     1      False
     2      False
     3      False
     4      False
     ...
    413     False
    414     False
    415     False
    416     False
    417     False
Length: 418, dtype: bool
```

4. Reading the files with the alias name 'df2' & and identifying by using df2.head()

```
df2 = pd.read_csv('test.csv')

df2.head()

  PassengerId  Pclass          Name   Sex  Age  SibSp  Parch  Ticket  Fare  Embarked
0         892     3  Kelly, Mr. James  male  34.5     0     0  330911  7.8292      Q
1         893     3  Wilkes, Mrs. James (Ellen Needs)  female  47.0     1     0  363272  7.0000      S
2         894     2  Myles, Mr. Thomas Francis  male  62.0     0     0  240276  9.6875      Q
3         895     3        Wirz, Mr. Albert  male  27.0     0     0  315154  8.6625      S
4         896     3  Hirvonen, Mrs. Alexander (Helga E Lindqvist)  female  22.0     1     1  3101298  12.2875      S
```

5. Finding the missing value.

```
[69]: df2.isnull().sum()  
  
[69]: PassengerId      0  
Pclass            0  
Name              0  
Sex              0  
Age             86  
SibSp            0  
Parch            0  
Ticket           0  
Fare             1  
Cabin          327  
Embarked         0  
dtype: int64  
  
[75]: df2.duplicated()  
  
[75]: 0    False  
1    False  
2    False  
3    False  
4    False  
...  
413   False  
414   False  
415   False  
416   False  
417   False  
Length: 418, dtype: bool
```

6. Removing the missing values.

```
[89]: # Removing missing values from df2 column Age and cabin  
  
df2.dropna(axis=1, thresh=0.5*len(df2), inplace=True)  
  
[95]: df2.isnull().sum()  
  
[95]: PassengerId      0  
Pclass            0  
Name              0  
Sex              0  
Age             0  
SibSp            0  
Parch            0  
Ticket           0  
Fare             0  
Embarked         0  
dtype: int64
```

Name: Shiv kumar paul
Dataset: Titanic Dataset

Tools: Python (Jupyter Notebook)
Email: rubelpaul990@gmail.com

7. Reading the files with the alias name 'df3' & and identifying by using df3.head()

[77]:	df3 = pd.read_csv('train.csv')														
[79]:	df3.head()														
[79]:	PassengerId	Survived	Pclass	Name			Sex	Age	SibSp	Parch	Ticket		Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN		S		
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th... Heikkinen, Miss. Laina	female	38.0	1	0	PC 17599	71.2833	C85		C		
2	3	1	3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN		S		
3	4	1	1	Allen, Mr. William Henry	male	35.0	1	0	113803	53.1000	C123		S		
4	5	0	3						373450	8.0500	NaN		S		

8. Identifying null values

df3.isnull().sum()	
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	

In the age column there are around 177 missing values and in Cabin column there are around 687 and in Embarked there are 2 missing values.

9. Removing the missing values.

```
[97]: # Removing missing values from df3 column age and cabin
```

```
df3.dropna(axis = 1, thresh = 0.5*len(df3), inplace=True)
```

```
[180]: df3.dropna(inplace=True)
```

```
[103]: df3.isnull().sum()
```

```
[103]: PassengerId      0
        Survived        0
        Pclass          0
        Name           0
        Sex            0
        Age            0
        SibSp          0
        Parch          0
        Ticket         0
        Fare           0
        Embarked       0
        dtype: int64
```

```
[85]: df3.duplicated().sum()
```

```
[85]: 0
```

10. Merging all the datasets

```
•[105]: # Merging the data df1 & df2
```

```
merge = df1.merge(df2) # default inner join
```

11. Converting the column ‘Sex’ to a binomial distribution, ‘male’ = 1 & ‘female’ = 0

Name: Shiv kumar paul
Dataset: Titanic Dataset

Tools: Python (Jupyter Notebook)
Email: rubelpaul990@gmail.com

```
[170]: # Changing the Gender column into binomial values
merge['Sex'] = merge['Sex'].map({'male':1, 'female':0})

[178]: merge
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	892	0	3	Kelly, Mr. James	1	34.5	0	0	330911	7.8292	Q
1	893	1	3	Wilkes, Mrs. James (Ellen Needs)	0	47.0	1	0	363272	7.0000	S
2	894	0	2	Myles, Mr. Thomas Francis	1	62.0	0	0	240276	9.6875	Q
3	895	0	3	Wirz, Mr. Albert	1	27.0	0	0	315154	8.6625	S
4	896	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	0	22.0	1	1	3101298	12.2875	S
...
326	1301	1	3	Peacock, Miss. Treastell	0	3.0	1	1	SOTON/O.Q. 3101315	13.7750	S
327	1303	1	1	Minahan, Mrs. William Edward (Lillian E Thorpe)	0	37.0	1	0	19928	90.0000	Q
328	1304	1	3	Henriksson, Miss. Jenny Lovisa	0	28.0	0	0	347086	7.7750	S
329	1306	1	1	Oliva y Ocana, Dona. Fermina	0	39.0	0	0	PC 17758	108.9000	C
330	1307	0	3	Saether, Mr. Simon Sivertsen	1	38.5	0	0	SOTON/O.Q. 3101262	7.2500	S

331 rows × 11 columns

12. Merging all 3 datasets

```
# Merging the data df1, df2 & df3

final_merge = merge.merge(df3)

final_merge.head()
```

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
-------------	----------	--------	------	-----	-----	-------	-------	--------	------	----------

It is observed that after merging all the datasets, 'df3' is unable to provide an output because the row count in 'df2' ends at 891, while the row count in 'df3' begins at 892. This discrepancy prevents the output from being produced.

13. Converting the column 'Sex' to a binomial distribution, 'male' = 1 & 'female' = 0

Name: Shiv kumar paul
Dataset: Titanic Dataset

Tools: Python (Jupyter Notebook)
Email: rubelpaul990@gmail.com

[182]:	df3['Sex'] = df3['Sex'].map({'male':1, 'female':0})																																																																																																																																																
[288]:	df3																																																																																																																																																
[288]:	<table><thead><tr><th></th><th>PassengerId</th><th>Survived</th><th>Pclass</th><th>Name</th><th>Sex</th><th>Age</th><th>SibSp</th><th>Parch</th><th>Ticket</th><th>Fare</th><th>Embarked</th></tr></thead><tbody><tr><td>0</td><td>1</td><td>0</td><td>3</td><td>Braund, Mr. Owen Harris</td><td>1</td><td>22.0</td><td>1</td><td>0</td><td>A/5 21171</td><td>7.2500</td><td>S</td></tr><tr><td>1</td><td>2</td><td>1</td><td>1</td><td>Cumings, Mrs. John Bradley (Florence Briggs Th... Heikkinen, Miss. Laina</td><td>0</td><td>38.0</td><td>1</td><td>0</td><td>PC 17599</td><td>71.2833</td><td>C</td></tr><tr><td>2</td><td>3</td><td>1</td><td>3</td><td>Futrelle, Mrs. Jacques Heath (Lily May Peel)</td><td>0</td><td>26.0</td><td>0</td><td>0</td><td>STON/O2. 3101282</td><td>7.9250</td><td>S</td></tr><tr><td>3</td><td>4</td><td>1</td><td>1</td><td>Allen, Mr. William Henry</td><td>1</td><td>35.0</td><td>0</td><td>0</td><td>113803</td><td>53.1000</td><td>S</td></tr><tr><td>4</td><td>5</td><td>0</td><td>3</td><td>Rice, Mrs. William (Margaret Norton)</td><td>1</td><td>35.0</td><td>0</td><td>0</td><td>373450</td><td>8.0500</td><td>S</td></tr><tr><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr><tr><td>885</td><td>886</td><td>0</td><td>3</td><td>Montvila, Rev. Juozas</td><td>0</td><td>39.0</td><td>0</td><td>5</td><td>382652</td><td>29.1250</td><td>Q</td></tr><tr><td>886</td><td>887</td><td>0</td><td>2</td><td>Graham, Miss. Margaret Edith</td><td>1</td><td>27.0</td><td>0</td><td>0</td><td>211536</td><td>13.0000</td><td>S</td></tr><tr><td>887</td><td>888</td><td>1</td><td>1</td><td>Behr, Mr. Karl Howell</td><td>0</td><td>19.0</td><td>0</td><td>0</td><td>112053</td><td>30.0000</td><td>S</td></tr><tr><td>889</td><td>890</td><td>1</td><td>1</td><td>Dooley, Mr. Patrick</td><td>1</td><td>26.0</td><td>0</td><td>0</td><td>111369</td><td>30.0000</td><td>C</td></tr><tr><td>890</td><td>891</td><td>0</td><td>3</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>370376</td><td>7.7500</td><td>Q</td></tr></tbody></table> <p>712 rows × 11 columns</p>		PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked	0	1	0	3	Braund, Mr. Owen Harris	1	22.0	1	0	A/5 21171	7.2500	S	1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th... Heikkinen, Miss. Laina	0	38.0	1	0	PC 17599	71.2833	C	2	3	1	3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	0	26.0	0	0	STON/O2. 3101282	7.9250	S	3	4	1	1	Allen, Mr. William Henry	1	35.0	0	0	113803	53.1000	S	4	5	0	3	Rice, Mrs. William (Margaret Norton)	1	35.0	0	0	373450	8.0500	S	885	886	0	3	Montvila, Rev. Juozas	0	39.0	0	5	382652	29.1250	Q	886	887	0	2	Graham, Miss. Margaret Edith	1	27.0	0	0	211536	13.0000	S	887	888	1	1	Behr, Mr. Karl Howell	0	19.0	0	0	112053	30.0000	S	889	890	1	1	Dooley, Mr. Patrick	1	26.0	0	0	111369	30.0000	C	890	891	0	3	370376	7.7500	Q
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked																																																																																																																																						
0	1	0	3	Braund, Mr. Owen Harris	1	22.0	1	0	A/5 21171	7.2500	S																																																																																																																																						
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th... Heikkinen, Miss. Laina	0	38.0	1	0	PC 17599	71.2833	C																																																																																																																																						
2	3	1	3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	0	26.0	0	0	STON/O2. 3101282	7.9250	S																																																																																																																																						
3	4	1	1	Allen, Mr. William Henry	1	35.0	0	0	113803	53.1000	S																																																																																																																																						
4	5	0	3	Rice, Mrs. William (Margaret Norton)	1	35.0	0	0	373450	8.0500	S																																																																																																																																						
...																																																																																																																																						
885	886	0	3	Montvila, Rev. Juozas	0	39.0	0	5	382652	29.1250	Q																																																																																																																																						
886	887	0	2	Graham, Miss. Margaret Edith	1	27.0	0	0	211536	13.0000	S																																																																																																																																						
887	888	1	1	Behr, Mr. Karl Howell	0	19.0	0	0	112053	30.0000	S																																																																																																																																						
889	890	1	1	Dooley, Mr. Patrick	1	26.0	0	0	111369	30.0000	C																																																																																																																																						
890	891	0	3	370376	7.7500	Q																																																																																																																																						

Statistical Analysis

14. Describing the data using .describe() function for 'df3' dataset

[186]:	df3.describe().round()																																																																																	
[186]:	<table><thead><tr><th></th><th>PassengerId</th><th>Survived</th><th>Pclass</th><th>Sex</th><th>Age</th><th>SibSp</th><th>Parch</th><th>Fare</th></tr></thead><tbody><tr><td>count</td><td>712.0</td><td>712.0</td><td>712.0</td><td>712.0</td><td>712.0</td><td>712.0</td><td>712.0</td><td>712.0</td></tr><tr><td>mean</td><td>449.0</td><td>0.0</td><td>2.0</td><td>1.0</td><td>30.0</td><td>1.0</td><td>0.0</td><td>35.0</td></tr><tr><td>std</td><td>259.0</td><td>0.0</td><td>1.0</td><td>0.0</td><td>14.0</td><td>1.0</td><td>1.0</td><td>53.0</td></tr><tr><td>min</td><td>1.0</td><td>0.0</td><td>1.0</td><td>0.0</td><td>0.0</td><td>0.0</td><td>0.0</td><td>0.0</td></tr><tr><td>25%</td><td>223.0</td><td>0.0</td><td>1.0</td><td>0.0</td><td>20.0</td><td>0.0</td><td>0.0</td><td>8.0</td></tr><tr><td>50%</td><td>445.0</td><td>0.0</td><td>2.0</td><td>1.0</td><td>28.0</td><td>0.0</td><td>0.0</td><td>16.0</td></tr><tr><td>75%</td><td>677.0</td><td>1.0</td><td>3.0</td><td>1.0</td><td>38.0</td><td>1.0</td><td>1.0</td><td>33.0</td></tr><tr><td>max</td><td>891.0</td><td>1.0</td><td>3.0</td><td>1.0</td><td>80.0</td><td>5.0</td><td>6.0</td><td>512.0</td></tr></tbody></table>		PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Fare	count	712.0	712.0	712.0	712.0	712.0	712.0	712.0	712.0	mean	449.0	0.0	2.0	1.0	30.0	1.0	0.0	35.0	std	259.0	0.0	1.0	0.0	14.0	1.0	1.0	53.0	min	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	25%	223.0	0.0	1.0	0.0	20.0	0.0	0.0	8.0	50%	445.0	0.0	2.0	1.0	28.0	0.0	0.0	16.0	75%	677.0	1.0	3.0	1.0	38.0	1.0	1.0	33.0	max	891.0	1.0	3.0	1.0	80.0	5.0	6.0	512.0
	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Fare																																																																										
count	712.0	712.0	712.0	712.0	712.0	712.0	712.0	712.0																																																																										
mean	449.0	0.0	2.0	1.0	30.0	1.0	0.0	35.0																																																																										
std	259.0	0.0	1.0	0.0	14.0	1.0	1.0	53.0																																																																										
min	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0																																																																										
25%	223.0	0.0	1.0	0.0	20.0	0.0	0.0	8.0																																																																										
50%	445.0	0.0	2.0	1.0	28.0	0.0	0.0	16.0																																																																										
75%	677.0	1.0	3.0	1.0	38.0	1.0	1.0	33.0																																																																										
max	891.0	1.0	3.0	1.0	80.0	5.0	6.0	512.0																																																																										

Name: Shiv kumar paul
Dataset: Titanic Dataset

Tools: Python (Jupyter Notebook)
Email: rubelpaul990@gmail.com

15. Describing the data using .describe() function for 'merge' dataset

	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Fare
count	331.0	331.0	331.0	331.0	331.0	331.0	331.0	331.0
mean	1100.0	0.0	2.0	1.0	30.0	0.0	0.0	41.0
std	123.0	0.0	1.0	0.0	14.0	1.0	1.0	61.0
min	892.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
25%	992.0	0.0	1.0	0.0	21.0	0.0	0.0	8.0
50%	1100.0	0.0	2.0	1.0	27.0	0.0	0.0	16.0
75%	1210.0	1.0	3.0	1.0	39.0	1.0	1.0	41.0
max	1307.0	1.0	3.0	1.0	76.0	8.0	6.0	512.0

16. Finding the value count using value_counts() function in 'merge' and 'df3' dataset

```
[190]: merge.value_counts()

[190]:   PassengerId  Survived  Pclass      Name          Sex  Age  SibSp  Parch     Ticket      Fare  Embarked
    892            0        3  Kelly, Mr. James       1  34.5    0    0  330911  7.8292    Q      1
   1186            0        3  Wittevrongel, Mr. Camille  1  36.0    0    0  345771  9.5000    S      1
   1183            1        3  Daly, Miss. Margaret Marcella Maggie  0  30.0    0    0  382650  6.9500    Q      1
   1179            0        1  Snyder, Mr. John Pillsbury  1  24.0    1    0  21228   82.2667   S      1
   1177            0        3  Dennis, Mr. William       1  36.0    0    0  A/5 21175  7.2500    S      1
                                                ..
   1029            0        2  Schmidt, Mr. August       1  26.0    0    0  248659  13.0000   S      1
   1028            0        3  Zakarian, Mr. Mapriededer  1  26.5    0    0  2656   7.2250    C      1
   1027            0        3  Carlsson, Mr. Carl Robert  1  24.0    0    0  350409  7.8542    S      1
   1026            0        3  Dintcheff, Mr. Valtcho      1  43.0    0    0  349226  7.8958    S      1
   1307            0        3  Saether, Mr. Simon Sivertsen  1  38.5    0    0  SOTON/O.Q. 3101262  7.2500    S      1
Name: count, Length: 331, dtype: int64

[192]: df3.value_counts()

[192]:   PassengerId  Survived  Pclass      Name          Sex  Age  SibSp  Parch     Ticket      Fare  Embarked
    1            0        3  Braund, Mr. Owen Harris       1  22.00    1    0  A/5 21171  7.2500    S      1
   622            1        1  Kimball, Mr. Edwin Nelson Jr  1  42.00    1    0  11753  52.5542    S      1
   595            0        2  Chapman, Mr. John Henry      1  37.00    1    0  SC/AH 29037  26.0000   S      1
   596            0        3  Van Impe, Mr. Jean Baptiste  1  36.00    1    1  345773  24.1500   S      1
   598            0        3  Johnson, Mr. Alfred       1  49.00    0    0  LINE   0.0000    S      1
                                                ..
   298            0        1  Allison, Miss. Helen Loraine     0  2.00    1    2  113781  151.5500   S      1
   300            1        1  Baxter, Mrs. James (Helene DeLaudeniere Chaput)  0  50.00    0    1  PC 17558  247.5208   C      1
   303            0        3  Johnson, Mr. William Cahoon Jr      1  19.00    0    0  LINE   0.0000    S      1
   306            1        1  Allison, Master. Hudson Trevor     1  0.92    1    2  113781  151.5500   S      1
   891            0        3  Dooley, Mr. Patrick       1  32.00    0    0  370376  7.7500    Q      1
Name: count, Length: 712, dtype: int64
```

17. Retrieving all the information using the info() function for ‘merge’ and ‘df3’ dataset

```
[194]: merge.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 331 entries, 0 to 330
Data columns (total 11 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   PassengerId 331 non-null    int64  
 1   Survived     331 non-null    int64  
 2   Pclass       331 non-null    int64  
 3   Name         331 non-null    object  
 4   Sex          331 non-null    int64  
 5   Age          331 non-null    float64 
 6   SibSp        331 non-null    int64  
 7   Parch        331 non-null    int64  
 8   Ticket       331 non-null    object  
 9   Fare          331 non-null    float64 
 10  Embarked     331 non-null    object  
dtypes: float64(2), int64(6), object(3)
memory usage: 28.6+ KB
```

```
[196]: df3.info()

<class 'pandas.core.frame.DataFrame'>
Index: 712 entries, 0 to 890
Data columns (total 11 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   PassengerId 712 non-null    int64  
 1   Survived     712 non-null    int64  
 2   Pclass       712 non-null    int64  
 3   Name         712 non-null    object  
 4   Sex          712 non-null    int64  
 5   Age          712 non-null    float64 
 6   SibSp        712 non-null    int64  
 7   Parch        712 non-null    int64  
 8   Ticket       712 non-null    object  
 9   Fare          712 non-null    float64 
 10  Embarked     712 non-null    object  
dtypes: float64(2), int64(6), object(3)
memory usage: 66.8+ KB
```

18. Finding the total rows and columns for both datasets

```
[198]: merge.shape
```

```
[198]: (331, 11)
```

```
[200]: df3.shape
```

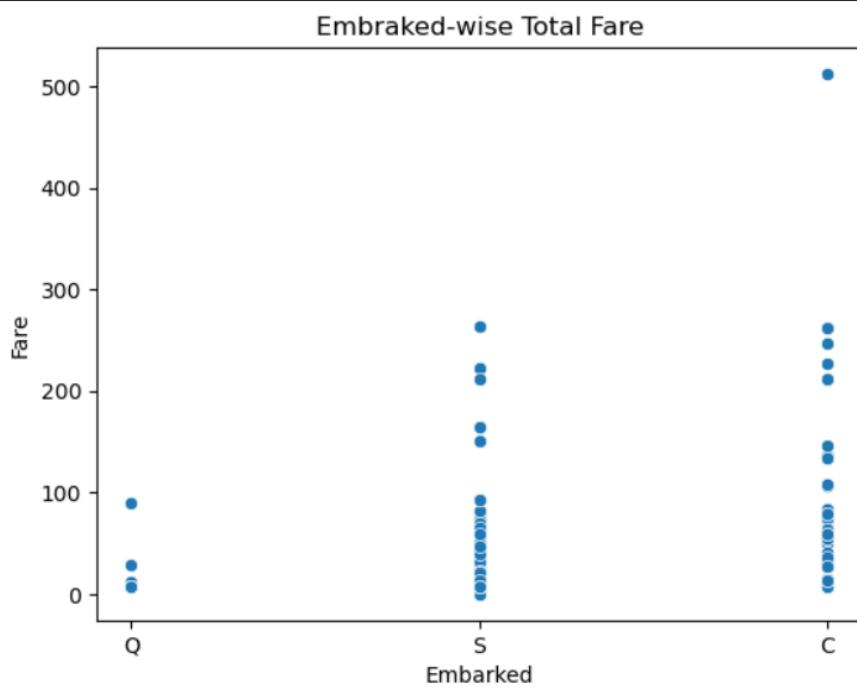
```
[200]: (712, 11)
```

In the merge data 331 is the total rows and 11 is the total columns. Similarly df3 has 712 total rows and 11 columns.

Visualizing the data

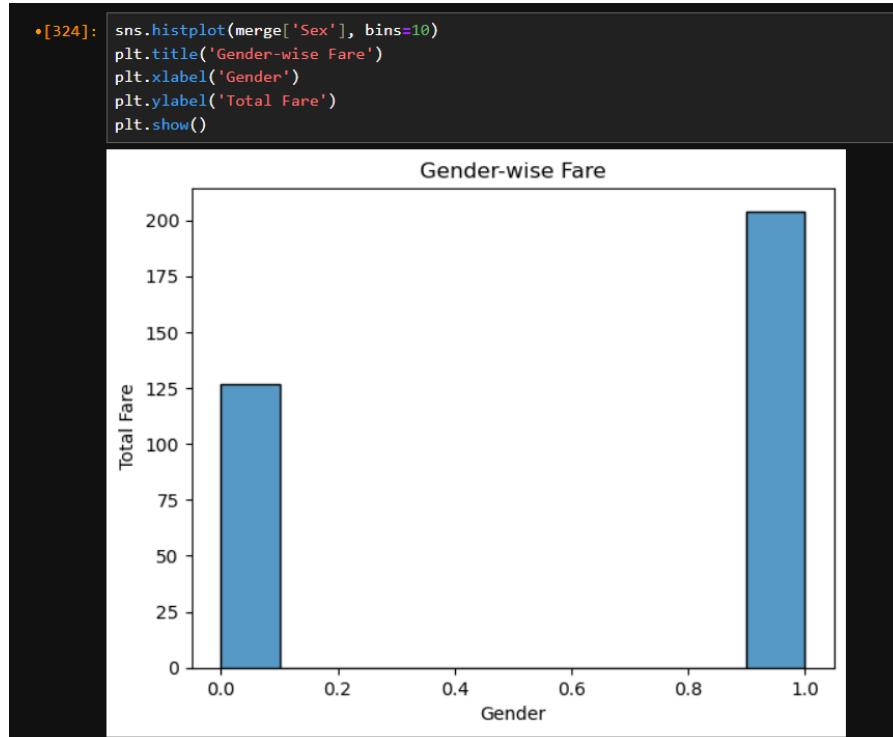
19. Scatterplot

```
: # Scatterplot visualization
sns.scatterplot(x = 'Embarked', y = 'Fare', data=merge)
plt.title('Embraked-wise Total Fare')
plt.show()
```



This graph analyzes the Embarked-wise total fare (Port of embarkation (C = Cherbourg, Q = Queenstown, S = Southampton)), which shows that C =Cherbourg has the highest total fare.

20. Histogram visualization



This graph represents the Gender-wise total fare, on the x-axis, 0 represents ‘Female’ and 1 represents ‘Male’. According to the analysis, we can observe that there are more travellers than females.

21. Heatmap visualization

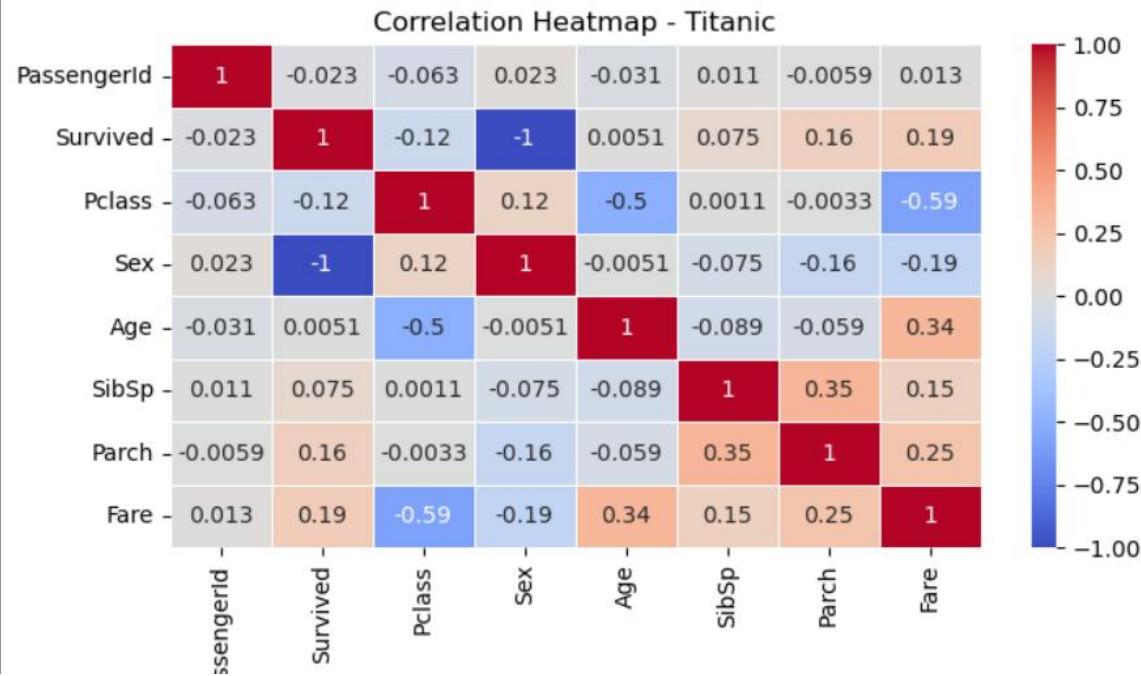
Name: Shiv kumar paul
Dataset: Titanic Dataset

Tools: Python (Jupyter Notebook)
Email: rubelpaul990@gmail.com

```
[360]: # Plot the heatmap for merge dataset
```

```
    numeric_df = merge.select_dtypes(include = 'number')
    corr_matrix = numeric_df

    plt.figure(figsize=(8, 4))
    sns.heatmap(corr_matrix.corr(), annot=True, cmap='coolwarm', linewidths=0.5)
    plt.title('Correlation Heatmap - Titanic')
    plt.show()
```



- It shows how strongly pairs of variables are **linearly related**.
- Values range from:
 - **+1** = Perfect positive correlation
 - **-1** = Perfect negative correlation
 - **0** = No correlation

The color scale (coolwarm) highlights:

- ● Red = Strong positive correlation
- ● Blue = Strong negative correlation
- ○ White/gray = Weak or no correlation

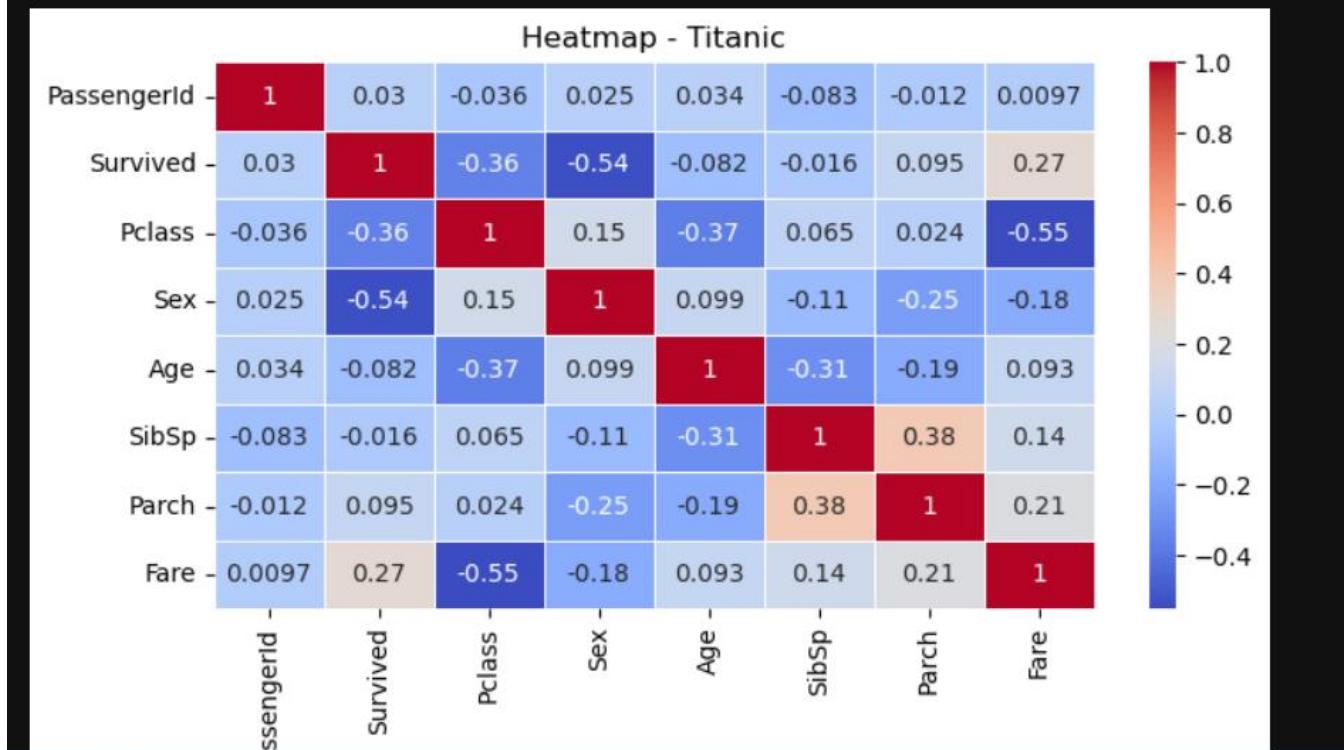
Pair	Correlation	Meaning
Fare & Pclass	-0.59	Strong negative correlation – higher class → higher fare.
Sex & Survived	-1.00	Strong negative correlation – males (0) are less likely to survive, females (1) are more likely
Parch & SibSp	0.35	Moderate positive correlation – families often traveled together.
Fare & Survived	0.19	Slight positive – passengers paying more had slightly higher survival chance.
Age & Fare	0.34	Mild positive – older passengers may have paid more
Pclass & Survived	-0.12	Slight negative – higher class passengers more likely to survive
PassengerId	Not meaningful	Just an identifier, should be excluded from analysis.

22. Heatmap visualization for df3 data

```
# Plot heatmap for df3 dataset

numeric_df = df3.select_dtypes(include='number')
matrix = numeric_df

plt.figure(figsize=(8, 4))
sns.heatmap(matrix.corr(), annot=True, cmap='coolwarm', linewidths=0.5)
plt.title("Heatmap - Titanic")
plt.show()
```



- It shows how strongly pairs of variables are **linearly related**.
- Values range from:
 - **+1** = Perfect positive correlation
 - **-1** = Perfect negative correlation
 - **0** = No correlation

The color scale (coolwarm) highlights:

- ● Red = Strong positive correlation
- ● Blue = Strong negative correlation

- White/gray = Weak or no correlation

Pair	Correlation	Meaning
Fare & Pclass	-0.55	Strong negative correlation – higher class → higher fare.
Sex & Survived	-0.54	Strong negative correlation – males (0) are less likely to survive, females (1) are more likely
Parch & SibSp	0.27	Moderate positive correlation – families often traveled together.
Fare & Survived	0.38	Slight positive – passengers paying more had slightly higher survival chance.
Age & Fare	0.36	Mild positive – older passengers may have paid more
Age & SibSp or Parch	Not meaningful	Just an identifier, should be excluded from analysis.

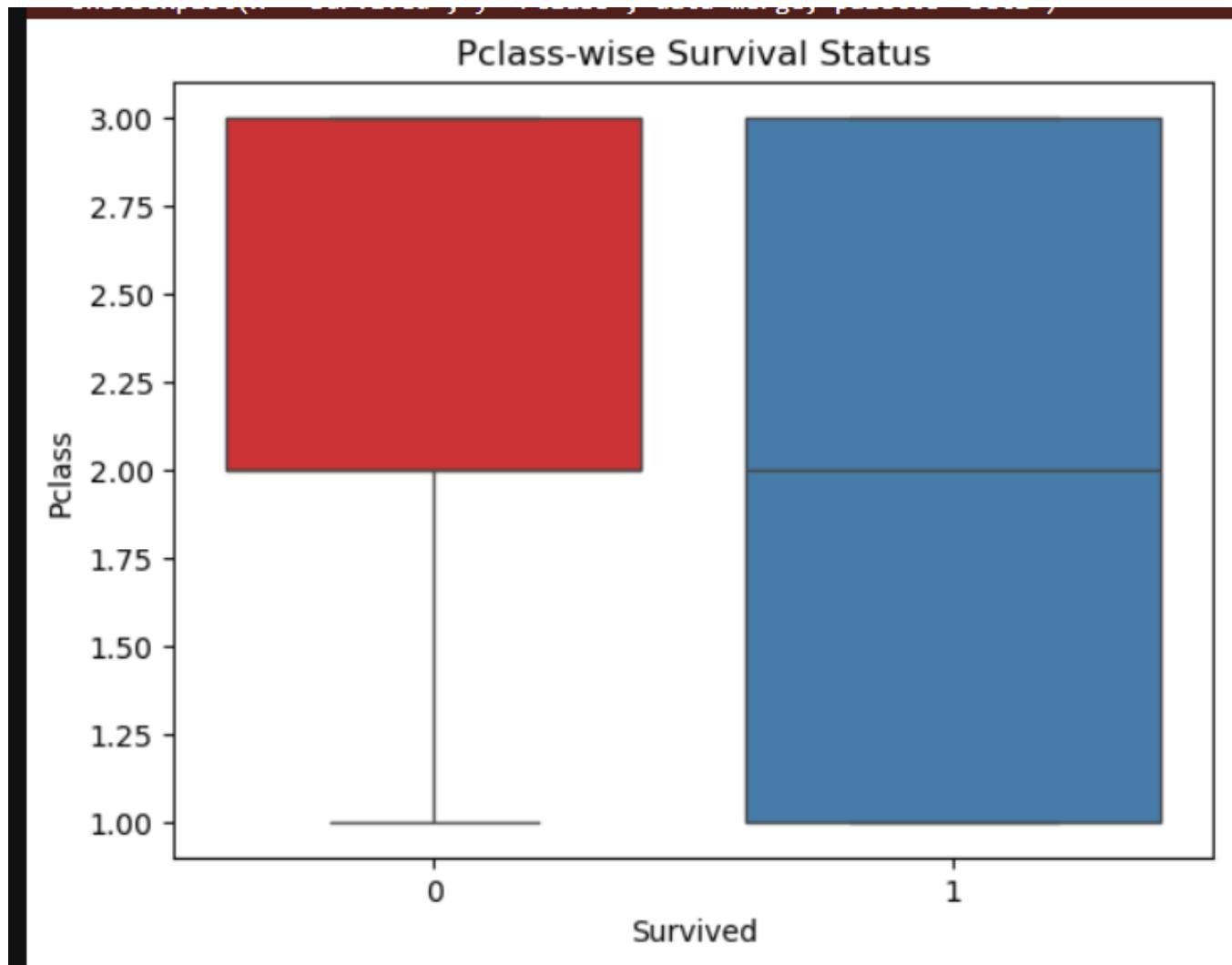
23. Boxplot Visualization

Input:

```
[350]: # survival status pclass

sns.boxplot(x= 'Survived', y='Pclass', data=merge, palette='Set1')
plt.title('Pclass-wise Survival Status')
plt.show()
```

Output:



Survived = 0 (Did not survive):

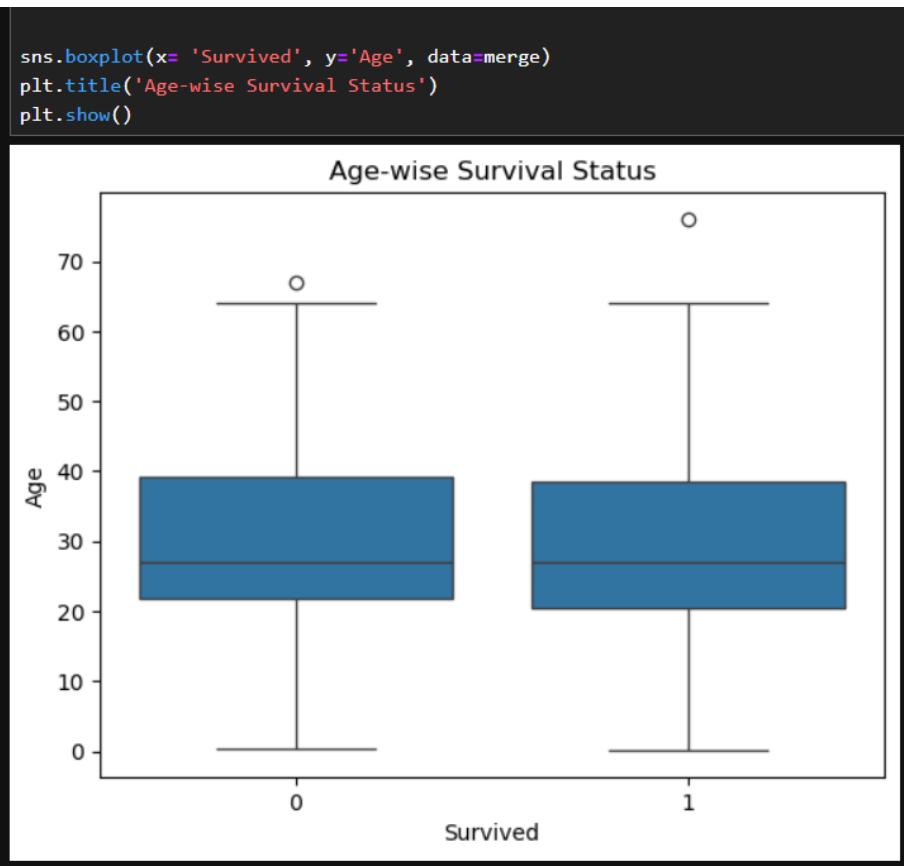
- Most non-survivors belonged to Pclass 3 (lower class).

- There's a wider range from **Pclass 1 to 3**, but the **median is around Pclass 3**.
- Indicates that **lower-class passengers had lower survival rates**.

Survived = 1 (Survived):

- Survivors were more likely from **Pclass 1 and 2**.
- The median Pclass is around **2**, and the lower whisker reaches **1** (upper class).
- This shows that **higher-class passengers had better survival chances**.

24. Age-wise Survival Status Boxplot visualization



X-axis (Survived):

- 0 = Did not survive
- 1 = Survived

Y-axis (Age): Age of passengers

General Observations:

- The **age distributions** for both survivors and non-survivors are **fairly similar**.
- The **median age** is roughly the same for both groups (around **25–30 years**).
- Both groups show a wide range of ages, from infants to around **70 years old**.
- A few **outliers** (dots above the box) are present in both groups, indicating passengers who were significantly older than the rest (above 65–70).

Name: Shiv kumar paul
Dataset: Titanic Dataset

Tools: Python (Jupyter Notebook)
Email: rubelpaul990@gmail.com