

Jupyter Prodigy\_Infotech\_task2\_DS Last Checkpoint: Last Monday at 5:12 PM (autosaved)

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

In [2]:

```
data = pd.read_csv('full.csv')
data.head()
```

Out[2]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Wikid	Name_wiki	Age_wiki	Hometown	B
0	1	0.0	3	Branson, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500		S	691.0	Branson, Mr. Owen Harris	22.0	Bridgnorth, Shropshire, England	South
1	2	1.0	1	Gaines, Mrs. John Bradley (Briscoe) Briggs Th.	female	38.0	1	0	PC 17599	71.2833		C	90.0	Gaines, Mrs. Florence Briggs (née Thayer)	35.0	New York, New York, US	Ch
2	3	1.0	3	Heldren, Mr. & Mrs. Lena Fatelle	female	26.0	0	0	STON/O2 3101282	7.9250		S	885.0	Heldren, Miss Lena Fatelle	26.0	Jyväskylä, Finland	South
3	4	1.0	1	Jacques Heim	female	35.0	1	0	113863	53.1000		S	127.0	Funstie, Miss Ida (née Heim)	35.0	Scituate, Massachusetts, US	South

Jupyter Prodigy\_Infotech\_task2\_DS Last Checkpoint: Last Monday at 5:12 PM (autosaved)

```
In [1]: import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns  
import warnings  
warnings.filterwarnings('ignore')  
  
In [2]: data= pd.read_csv('full.csv')  
data.head()  
Out[2]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked	Wikitid	Name_wiki	Age_wiki	Hometown	B	
0	1	0.0	3	Brundt, Mr. Charles Harris	male	22.0	1	0	A/5 21171	7.2500	...	S	691.0	Brundt, Mr. Owen Harris	22.0	Bridgwater, Devon, England	South
1	2	1.0	1	Cummings, Mrs. John Bradley (Florence Brigitte Th...)	female	38.0	1	0	PC 17599	71.2833	...	C	90.0	Cummings, Mrs. Florence Brigitte (née Thayer)	35.0	New York, New York, US	Ch
2	3	1.0	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2 3101282	7.9250	...	S	865.0	Heikkinen, Miss Laina	26.0	Jyväskylä, Finland	South
3	4	1.0	1	Futrelle, Mrs. Jacques Heikkinen (Lily May Peel)	female	35.0	1	0	113803	53.1000	...	S	127.0	Futrelle, Mrs. Lily May (née Peel)	35.0	Soutou, Massachusetts, US	South
4	5	0.0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	...	S	627.0	Allen, Mr. William Henry	35.0	Birmingham, West Midlands, England	South

5 rows × 21 columns

Jupyter Prodigy\_Infotech\_task2\_DS Last Checkpoint: Last Monday at 5:12 PM (autosaved)

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel) Logout

0 1 0.0 3 Mr. Owen Harris male 22.0 1 0 A/5 21171 7.2500 ... S 691.0 Braund, Mr. Owen Harris 22.0 Bridgewater, Devon, England South

1 2 1.0 1 Cummings, Mrs. John Briggs (Florence Briggs Thayer) female 38.0 1 0 PC 17599 71.2833 ... C 90.0 Cummings, Mrs. Florence Briggs Thayer (Florence Thayer) 36.0 New York, New York, US Ch

2 3 1.0 3 Heikkinen, Miss. Laina female 26.0 0 0 STON/O2 3101282 7.9250 ... S 865.0 Heikkinen, Miss. Laina 26.0 Jyväskylä, Finland South

3 4 1.0 1 Futrelle, Mrs. Jacques Heath (Mrs. Frank Peirce) female 35.0 1 0 113803 53.1000 ... S 127.0 Futrelle, Mrs. Jacques Heath (Mrs. Frank Peirce) 35.0 Scituate, Massachusetts, US South

4 5 0.0 3 Allen, Mr. William Henry male 35.0 0 0 373450 8.0500 ... S 627.0 Allen, Mr. William Henry 35.0 Birmingham, West Midlands, England South

5 rows × 21 columns

In [3]: data.tail()

Out[3]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked	Wikid	Name_wiki	Age_wiki	Hometown	
1304	1305	NaN	3	Spector, Mr. Woolf	male	NaN	0	0	A/5 3236	8.0500	...	S	1227.0	Spector, Mr. Woolf	23.0	London, England So
1305	1306	NaN	1	Oliva y Ocaña, Dona Fernanda	female	30.0	0	0	PC 17758	108.9000	...	C	229.0	am maid, Dofia Fernanda Oliva y Ocaña	39.0	Madrid, Spain Skagen

See other:

Jupyter Prodigy\_Infotech\_task2\_DS Last Checkpoint: Last Monday at 5:12 PM (autosaved)

1306	1306	NaN	1	Ocaya y Dona Fermina	female	30.0	0	0	PC 17758	108.9000	...	C	229.0	and maid. Dofla Fernanda Ocaya y Ocana	39.0	Madrid Spain					
1306	1307	NaN	3	Sæther, Mr Simon Sverresen	male	38.5	0	0	SOTON/CO 3101292	7.2500	...	S	1169.0	Sæther, Mr Simon Sverresen	43.0	Skau, Sel- Tromsø Norway					
1307	1308	NaN	3	Ware, Mr Frederick	male	NaN	0	0	3560309	8.0500	...	S	1289.0	Ware, Mr. Frederick William	34.0	Greenwich, London, England					
1308	1309	NaN	3	Peter, Master Michael J	male	NaN	1	1	2668	22.3583	...	C	702.0	Batas Butrus- Battus Master Michael	4.0	Sarafiki Syria					

5 rows × 21 columns

```
In [4]: data.index
Out[4]: RangeIndex(start=0, stop=1309, step=1)

In [5]: data.shape
Out[5]: (1309, 21)

In [6]: data.size
Out[6]: 27489

In [7]: data.columns
Out[7]: Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked', 'SibSp1', 'NameWith1'])
```

Jupyter Prodigy\_Infotech\_task2\_DS Last Checkpoint: Last Monday at 5:12 PM (autosaved)

In [4]: data.index  
Out[4]: RangeIndex(start=0, stop=1309, step=1)

In [5]: data.shape  
Out[5]: (1309, 21)

In [6]: data.size  
Out[6]: 27489

In [7]: data.columns  
Out[7]: Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp', 'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked', 'WikiId', 'Name\_wiki', 'Age\_wiki', 'Hometown', 'Boarded', 'Destination', 'Lifeboat', 'Body', 'Class'], dtype='object')

In [8]: data.describe()  
Out[8]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	WikiId	Age_wiki	Class
count	1309.000000	891.000000	1309.000000	1046.000000	1309.000000	1309.000000	1308.000000	1304.000000	1302.000000	1304.000000

Jupyter Prodigy\_Infotech\_task2\_DS Last Checkpoint: Last Monday at 5:12 PM (autosaved)

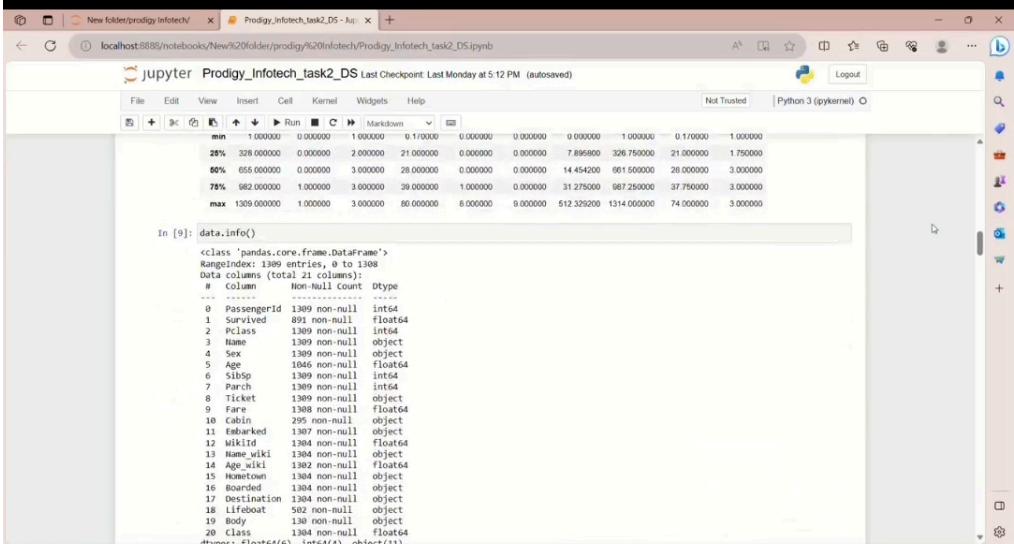
```
In [5]: data.shape
Out[5]: (1309, 21)

In [6]: data.size
Out[6]: 27489

In [7]: data.columns
Out[7]: Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked', 'Wikid', 'Name_wiki',
       'Age_wiki', 'Hometown', 'Boarded', 'Destination', 'Lifeboat', 'Body',
       'Class'],
       dtype='object')

In [8]: data.describe()
Out[8]:
   PassengerId  Survived  Pclass      Age     SibSp    Parch      Fare      Wikid  Age_wiki  Class
count    1309.000000  891.000000  1309.000000  1046.000000  1309.000000  1308.000000  1304.000000  1302.000000  1304.000000
mean     655.000000  0.383834  2.294862  29.881138  0.498854  0.385627  33.295479  658.534509  29.415829  2.291411
std      378.620061  0.486952  0.837836  14.413483  0.416956  0.865680  51.759668  380.377373  13.758954  0.840852
min      1.000000  0.000000  1.000000  0.170000  0.000000  0.000000  0.000000  0.000000  0.170000  1.000000
25%    328.000000  0.000000  2.000000  21.000000  0.000000  0.000000  7.986000  326.760000  21.000000  1.750000
50%    655.000000  0.000000  3.000000  28.000000  0.000000  0.000000  14.454200  661.500000  28.000000  3.000000
75%    982.000000  1.000000  3.000000  39.000000  1.000000  0.000000  31.275000  981.250000  37.750000  3.000000
max    1309.000000  1.000000  3.000000  80.000000  8.000000  9.000000  512.329200  1314.000000  74.000000  3.000000
```

```
In [9]: data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1309 entries, 0 to 1309
```



The screenshot shows a Jupyter Notebook interface running on a Windows operating system. The title bar indicates the notebook is titled "Prodigy\_Infotech\_task2\_DS.ipynb". The main window displays the output of a cell, which is a pandas DataFrame named "data". The code executed was "data.info()", which provides detailed information about the DataFrame's structure.

The output shows the following details:

- Class:** pandas.core.frame.DataFrame
- RangeIndex:** 1308 entries, 0 to 1308
- Data columns (total 21 columns):**

#	Column	Non-Null Count	Dtype
0	PassengerId	1308	int64
1	Survived	891	non-null float64
2	Pclass	1308	non-null int64
3	Name	1308	non-null object
4	Sex	1308	non-null object
5	Age	1046	non-null float64
6	SibSp	1308	non-null int64
7	Parch	1308	non-null int64
8	Ticket	1308	non-null object
9	Fare	1308	non-null float64
10	Cabin	209	non-null object
11	Embarked	1307	non-null object
12	WikitId	1304	non-null float64
13	Name_wiki	1304	non-null object
14	Age_wiki	1304	non-null float64
15	NaN_wiki	1304	non-null object
16	Boarded	1304	non-null object
17	Destination	1304	non-null object
18	Lifeboat	502	non-null object
19	Body	130	non-null object
20	Class	130	non-null float64

- dtypes:** float64(6), int64(11), object(4)

The screenshot shows a Jupyter Notebook interface running on a Windows operating system. The title bar indicates the notebook is titled "Prodigy\_Infotech\_task2\_DS.ipynb". The main window displays a code cell with the following content:

```
# Data Cleaning:  
In [10]: missing_data = data.isnull().sum()  
missing_data
```

The output of the code cell is:

```
Out[10]: PassengerId      0  
Survived        418  
Pclass          0  
Name            0  
Sex            0  
Age         263  
SibSp          0  
Parch          0  
Ticket         0
```

The screenshot shows a Jupyter Notebook interface running on a Windows operating system. The title bar indicates the notebook is titled "Prodigy\_Infotech\_task2\_DS.ipynb". The menu bar includes File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. The toolbar has icons for Run, Cell, Kernel, Widgets, and Help. The status bar shows "Not Trusted" and "Python 3 (ipykernel)".

The notebook content starts with a cell output showing the structure of a DataFrame:

```
In [10]: 15 Hometown 1384 non-null object
          16 Boarded 1384 non-null object
          17 Destination 1384 non-null object
          18 Embarked 1384 non-null object
          19 Body 138 non-null object
          20 Class 1384 non-null float64
          dtypes: float64(6), int64(4), object(11)
          memory usage: 214.9+ KB
```

Below this, a section titled "Data Cleaning:" is shown:

```
In [10]: missing_data = data.isnull().sum()
```

```
Out[10]: PassengerId      0
          Survived       418
          Pclass         0
          Name          0
          Sex           0
          Age          263
          SibSp         0
          Parch         0
          Ticket        0
          Fare          1
          Cabin        1014
          Embarked      2
          Wki          5
          Name_wiki     5
          Age_wiki      7
          Hometown      5
          Boarded       5
          Surname       5
          Destination   887
          Body          1179
          Class          5
```

In [8]: `data.describe()`

Out[8]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	Wiki	Age_wiki	Class
count	1309.000000	891.000000	1309.000000	1046.000000	1309.000000	1308.000000	1304.000000	1322.000000	1304.000000	
mean	655.000000	0.385383	2.294882	29.881138	0.498954	0.395027	33.295479	658.534508	29.415829	2.291411
std	378.620061	0.496582	0.837659	14.413493	1.041658	0.855660	51.759688	360.377373	13.758954	0.848862
min	1.000000	0.000000	1.000000	0.179000	0.000000	0.000000	0.000000	0.170000	1.000000	
25%	328.000000	0.000000	2.000000	21.050000	0.000000	0.000000	7.895000	326.760000	21.000000	1.750000
50%	655.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200	661.500000	28.000000	3.000000
75%	982.000000	1.000000	3.000000	38.000000	1.000000	0.000000	51.750000	981.250000	37.750000	3.000000
max	1309.000000	1.000000	3.000000	60.000000	8.000000	9.000000	512.329200	1314.000000	74.000000	3.000000

In [9]: `data.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1309 entries, 0 to 1308
Data columns (total 11 columns):
 #   Column          Non-Null Count  Dtype    
--- 
 0   PassengerId    1309 non-null   int64    
 1   Survived       891 non-null   float64  
 2   Pclass          1309 non-null   int64    
 3   Name           1309 non-null   object   
 4   Sex            1309 non-null   object   
 5   Age            1046 non-null   float64  
 6   SibSp          1309 non-null   int64    
 7   Parch          1309 non-null   int64    
 8   Ticket         1309 non-null   object   
 9   Fare           1309 non-null   float64  
 10  Cabin          295 non-null   object   
 11  Embarked       1307 non-null   object
```

The screenshot shows a Jupyter Notebook interface running on a Windows operating system. The title bar indicates the notebook is titled "Prodigy\_Infotech\_task2\_DS.ipynb". The notebook window has two code cells visible:

```
In [10]: missing_data = data.isnull().sum()
missing_data
```

```
Out[10]: PassengerId      0
Survived       418
Pclass         0
Name          0
Sex           0
Age        263
SibSp         0
Parch         0
Ticket        0
Fare          1
Cabin       1014
Embarked      2
Wikilid       5
Name_wiki     5
Age_wiki      7
Akerman       5
Boarded       5
Destination   5
Lifeboat      887
Body        1179
Class         5
dtype: int64
```

```
In [11]: data = data.drop_duplicates()
data.head(8)
```

Jupyter Prodigy\_Infotech\_task2\_DS Last Checkpoint: Last Monday at 5:12 PM (autosaved)

In [11]:

```
data = data.drop_duplicates()
data.head(5)
```

Out[11]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Wikit	Name_wiki	Age_wiki	Hometown	E
0	1	0.0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	72.500		S	691.0	Braund, Mr. Owen Harris	22.0	Brigendine, Devon, England	Sout
1	2	1.0	1	Cumings, Mrs. John Bradley (Florence Briggs Th.	female	38.0	1	0	PC 17599	71.2833		C	90.0	Cumings, Mrs. Florence Briggs (new Thayer)	35.0	New York, New York, US	Ch
2	3	1.0	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2 3101282	7.9250		S	865.0	Heikkinen, Miss. Laina	26.0	Jyväskylä, Finland	Sout
3	4	1.0	1	Futrelle, Mrs. Jacques Heyes (Lily May Peel)	female	35.0	1	0	113863	53.1000		S	127.0	Futrelle, Mrs. Lily May Heyes (Lily May Peel)	35.0	Schubate, Massachusetts, US	Sout
4	5	0.0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500		S	627.0	Allen, Mr. William Henry	35.0	Birmingham, West Midlands, England	Sout

Jupyter Prodigy\_Infotech\_task2\_DS Last Checkpoint: Last Monday at 5:12 PM (autosaved)

```
In [11]: data = data.drop_duplicates()
data.head(8)
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked	Wkld	Name_wiki	Age_wiki	Hometown	E
0	1	0.0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	S	691.0	Braund, Mr. Owen Harris	22.0	Bridgeville, Devon, England	South
1	2	1.0	1	Cummings, Mrs. John Bradley (Florence Briggs Th... Th...	female	38.0	1	0	PC 17599	71.2833	C	90.0	Cummings, Mrs. Florence Briggs (née Thayer) Th...	35.0	New York, New York, US	Ch
2	3	1.0	3	Heikkinen, Miss. Laina Fuerte, Mrs. Jacques Heikkinen (Lily May Peel) Fuerte, Mrs. Jacques Heikkinen (Lily May Peel)	female	26.0	0	0	STON/O2 3101262	7.9250	S	865.0	Heikkinen, Miss Laina Fuerte, Mrs. Lily May (née Heikkinen) Fuerte, Mrs. Jacques Heikkinen (Lily May Peel)	26.0	Jyväskylä, Finland	South
3	4	1.0	1	Allen, Mr. William Henry McCarthy, Mr. Timothy J. Allen, Mr. William Henry McCarthy, Mr. Timothy J.	male	35.0	1	0	113803	53.1000	S	127.0	Allen, Mr. William Henry Doherty, Mr. William John (aka "William Moran") McCarthy, Mr. Timothy J.	35.0	Scituate, Massachusetts, US	South
4	5	0.0	3	Allen, Mr. William Henry McCarthy, Mr. Timothy J. Allen, Mr. William Henry McCarthy, Mr. Timothy J.	male	35.0	0	0	373450	8.0500	S	627.0	Allen, Mr. William Henry Doherty, Mr. William John (aka "William Moran") McCarthy, Mr. Timothy J.	35.0	Birmingham, West Midlands, England	South
5	6	0.0	3	Moran, Mr. James McCarthy, Mr. Timothy J. Pahaska, Master. Gosta Moran, Mr. James McCarthy, Mr. Timothy J. Pahaska, Master. Gosta	male	NaN	0	0	330677	8.4583	Q	785.0	Doherty, Mr. William John (aka "William Moran") McCarthy, Mr. Timothy J.	22.0	Cork, Ireland	Quer
6	7	0.0	1	McCarthy, Mr. Timothy J. Pahaska, Master. Gosta McCarthy, Mr. Timothy J. Pahaska, Master. Gosta	male	54.0	0	0	17463	51.9825	S	200.0	McCarthy, Mr. Timothy J.	54.0	Dorchester, Massachusetts, US	South
7	8	0.0	3	Pahaska, Master. Gosta Pahaska, Master. Gosta	male	2.0	3	1	349609	21.0750	S	1108.0	Pahaska, Master. Gosta	2.0	Bjuv, Skåne, Sweden	South

Jupyter Prodigy\_Infotech\_task2\_DS Last Checkpoint: Last Monday at 5:12 PM (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Not Trusted Python 3 (ipykernel)

Futrelle, Mrs. Jacques Heath (Ivy May Peel)

3	4	1.0	1	Jacques	female	35.0	1	0	113803	53	1000	...	S	127.0	Futrelle, Mrs. Jacques (Ivy May Peel)	35.0	Massachusetts, US	South			
4	5	0.0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8	0500	...	S	627.0	Allen, Mr. William Henry	35.0	Birmingham, West Midlands, England	South			
5	6	0.0	3	Mr. Moran	male	NaN	0	0	330877	8	4583	...	Q	785.0	Doherty, Mr. William James (and "James Moran")	22.0	Cork, Ireland	Quarantine			
6	7	0.0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51	.8625	...	S	200.0	McCarthy, Mr. Timothy J	54.0	Dorchester, Massachusetts, US	South			
7	8	0.0	3	Páansson, Master. Gosta Leonard	male	2.0	3	1	349699	21	0750	...	S	1108.0	Páansson, Master. Gosta Leonard	2.0	Bjørvika, Skåne, Sweden	South			

8 rows × 21 columns

In [12]:

```
data['Age'] = pd.to_numeric(data['Age'], errors='coerce')
data['Fare'] = pd.to_numeric(data['Fare'], errors='coerce')
```

Exploratory Data Analysis (EDA):

In [13]:

```
summary_stats = data.describe()
print(summary_stats)
```

PassengerId	Survived	Pclass	Age	sibsp
1	0	3	23.0	1
2	1	1	2	1
3	1	3	35.5	1
4	0	1	30.0	1
5	0	3	33.5	1
6	1	3	35.0	0
7	1	1	36.0	0
8	0	3	46.0	0
9	1	3	47.0	0
10	0	3	54.0	0
11	1	1	54.0	0
12	1	3	55.5	0
13	0	3	56.0	0
14	1	1	56.0	0
15	0	3	56.0	0
16	1	1	56.0	0
17	0	3	56.0	0
18	1	1	56.0	0
19	0	3	56.0	0
20	1	1	56.0	0
21	0	3	56.0	0
22	1	1	56.0	0
23	0	3	56.0	0
24	1	1	56.0	0
25	0	3	56.0	0
26	1	1	56.0	0
27	0	3	56.0	0
28	1	1	56.0	0
29	0	3	56.0	0
30	1	1	56.0	0
31	0	3	56.0	0
32	1	1	56.0	0
33	0	3	56.0	0
34	1	1	56.0	0
35	0	3	56.0	0
36	1	1	56.0	0
37	0	3	56.0	0
38	1	1	56.0	0
39	0	3	56.0	0
40	1	1	56.0	0
41	0	3	56.0	0
42	1	1	56.0	0
43	0	3	56.0	0
44	1	1	56.0	0
45	0	3	56.0	0
46	1	1	56.0	0
47	0	3	56.0	0
48	1	1	56.0	0
49	0	3	56.0	0
50	1	1	56.0	0
51	0	3	56.0	0
52	1	1	56.0	0
53	0	3	56.0	0
54	1	1	56.0	0
55	0	3	56.0	0
56	1	1	56.0	0
57	0	3	56.0	0
58	1	1	56.0	0
59	0	3	56.0	0
60	1	1	56.0	0
61	0	3	56.0	0
62	1	1	56.0	0
63	0	3	56.0	0
64	1	1	56.0	0
65	0	3	56.0	0
66	1	1	56.0	0
67	0	3	56.0	0
68	1	1	56.0	0
69	0	3	56.0	0
70	1	1	56.0	0
71	0	3	56.0	0
72	1	1	56.0	0
73	0	3	56.0	0
74	1	1	56.0	0
75	0	3	56.0	0
76	1	1	56.0	0
77	0	3	56.0	0
78	1	1	56.0	0
79	0	3	56.0	0
80	1	1	56.0	0
81	0	3	56.0	0
82	1	1	56.0	0
83	0	3	56.0	0
84	1	1	56.0	0
85	0	3	56.0	0
86	1	1	56.0	0
87	0	3	56.0	0
88	1	1	56.0	0
89	0	3	56.0	0
90	1	1	56.0	0
91	0	3	56.0	0
92	1	1	56.0	0
93	0	3	56.0	0
94	1	1	56.0	0
95	0	3	56.0	0
96	1	1	56.0	0
97	0	3	56.0	0
98	1	1	56.0	0
99	0	3	56.0	0
100	1	1	56.0	0
101	0	3	56.0	0
102	1	1	56.0	0
103	0	3	56.0	0
104	1	1	56.0	0
105	0	3	56.0	0
106	1	1	56.0	0
107	0	3	56.0	0
108	1	1	56.0	0
109	0	3	56.0	0
110	1	1	56.0	0
111	0	3	56.0	0
112	1	1	56.0	0
113	0	3	56.0	0
114	1	1	56.0	0
115	0	3	56.0	0
116	1	1	56.0	0
117	0	3	56.0	0
118	1	1	56.0	0
119	0	3	56.0	0
120	1	1	56.0	0
121	0	3	56.0	0
122	1	1	56.0	0
123	0	3	56.0	0
124	1	1	56.0	0
125	0	3	56.0	0
126	1	1	56.0	0
127	0	3	56.0	0
128	1	1	56.0	0
129	0	3	56.0	0
130	1	1	56.0	0
131	0	3	56.0	0
132	1	1	56.0	0
133	0	3	56.0	0
134	1	1	56.0	0
135	0	3	56.0	0
136	1	1	56.0	0
137	0	3	56.0	0
138	1	1	56.0	0
139	0	3	56.0	0
140	1	1	56.0	0
141	0	3	56.0	0
142	1	1	56.0	0
143	0	3	56.0	0
144	1	1	56.0	0
145	0	3	56.0	0
146	1	1	56.0	0
147	0	3	56.0	0
148	1	1	56.0	0
149	0	3	56.0	0
150	1	1	56.0	0
151	0	3	56.0	0
152	1	1	56.0	0
153	0	3	56.0	0
154	1	1	56.0	0
155	0	3	56.0	0
156	1	1	56.0	0
157	0	3	56.0	0
158	1	1	56.0	0
159	0	3	56.0	0
160	1	1	56.0	0
161	0	3	56.0	0
162	1	1	56.0	0
163	0	3	56.0	0
164	1	1	56.0	0
165	0	3	56.0	0
166	1	1	56.0	0
167	0	3	56.0	0
168	1	1	56.0	0
169	0	3	56.0	0
170	1	1	56.0	0
171	0	3	56.0	0
172	1	1	56.0	0
173	0	3	56.0	0
174	1	1	56.0	0
175	0	3	56.0	0
176	1	1	56.0	0
177	0	3	56.0	0
178	1	1	56.0	0
179	0	3	56.0	0
180	1	1	56.0	0
181	0	3	56.0	0
182	1	1	56.0	0
183	0	3	56.0	0
184	1	1	56.0	0
185	0	3	56.0	0
186	1	1	56.0	0
187	0	3	56.0	0
188	1	1	56.0	0
189	0	3	56.0	0
190	1	1	56.0	0
191	0	3	56.0	0
192	1	1	56.0	0
193	0	3	56.0	0
194	1	1	56.0	0
195	0	3	56.0	0
196	1	1	56.0	0
197	0	3	56.0	0
198	1	1	56.0	0
199	0	3	56.0	0
200	1	1	56.0	0
201	0	3	56.0	0
202	1	1	56.0	0
203	0	3	56.0	0
204	1	1	56.0	0
205	0	3	56.0	0
206	1	1	56.0	0
207	0	3	56.0	0
208	1	1	56.0	0
209	0	3	56.0	0
210	1	1	56.0	0
211	0	3	56.0	0
212	1	1	56.0	0
213	0	3	56.0	0
214	1	1	56.0	0
215	0	3	56.0	0
216	1	1	56.0	0
217	0	3	56.0	0
218	1	1	56.0	0
219	0	3	56.0	0
220	1	1	56.0	0
221	0	3	56.0	0
222	1	1	56.0	0
223	0	3	56.0	0
224	1	1	56.0	0
225	0	3	56.0	0
226	1	1	56.0	0
227	0	3	56.0	0
228	1	1	56.0	0
229	0	3	56.0	0
230	1	1	56.0	0
231	0	3	56.0	0
232	1	1	56.0	0
233	0	3	56.0	0
234	1	1	56.0	0
235	0	3	56.0	0
236	1	1	56.0	0
237	0	3	56.0	0
238	1	1	56.0	0
239	0	3	56.0	0
240	1	1	56.0	0
241	0	3	56.0	0
242	1	1	56.0	0
243	0	3	56.0	0
244	1	1	56.0	0
245	0	3	56.0	0
246	1	1	56.0	0
247	0	3	56.0	0
248	1	1	56.0	0
249	0	3	56.0	0
250	1	1	56.0	0
251	0	3	56.0	0
252	1	1	56.0	0
253	0	3	56.0	0
254	1	1	56.0	0
255	0	3	56.0	0
256	1	1	56.0	0
257	0	3	56.0	0
258	1	1	56.0	0
259	0	3	56.0	0
260	1	1	56.0</	

Jupyter Prodigy\_Infotech\_task2\_DS Last Checkpoint: Last Monday at 5:12 PM (autosaved)

In [11]:

```
data['Age'] = pd.to_numeric(data['Age'], errors='coerce')
data['Fare'] = pd.to_numeric(data['Fare'], errors='coerce')
```

Exploratory Data Analysis (EDA):

In [13]:

```
summary_stats = data.describe()
print(summary_stats)
```

	PassengerId	Survived	Pclass	Age	SibSp	Class
count	1399.000000	891.000000	1399.000000	1046.000000	1309.000000	
mean	655.000000	0.381838	2.294882	29.881138	0.498854	
std	378.020601	0.486592	0.837836	14.411493	1.041585	
min	1.000000	1.000000	1.000000	0.000000	0.000000	
25%	228.000000	0.000000	2.000000	21.000000	0.000000	
50%	655.000000	0.000000	3.000000	28.000000	0.000000	
75%	982.000000	1.000000	3.000000	39.000000	1.000000	
max	1399.000000	1.000000	3.000000	80.000000	8.000000	

	Parch	Fare	wikid	Age.wiki	Class
count	1399.000000	1308.000000	1304.000000	1302.000000	1304.000000
mean	0.385027	33.295479	658.534509	29.415829	2.291411
std	0.865560	51.750660	380.377373	13.758954	0.840852
min	0.000000	0.000000	1.000000	0.170000	1.000000
25%	0.000000	7.000000	332.700000	17.000000	1.000000
50%	0.000000	14.000000	403.000000	24.000000	2.000000
75%	0.000000	31.000000	571.000000	31.000000	3.000000
max	1.000000	152.000000	1000.000000	512.000000	3.000000

Jupyter Prodigy\_Infotech\_task2\_DS Last Checkpoint: Last Monday at 5:12 PM (autosaved)

In [12]: `data['Age'] = pd.to_numeric(data['Age'], errors='coerce')`  
`data['Fare'] = pd.to_numeric(data['Fare'], errors='coerce')`

**Exploratory Data Analysis (EDA):**

In [13]: `summary_stats = data.describe()`  
`print(summary_stats)`

	PassengerId	Survived	Pclass	Age	SibSp	\
count	1399.000000	891.000000	1399.000000	1846.000000	1309.000000	
mean	659.000000	0.383839	2.254082	29.881138	0.488854	
std	378.020061	0.486592	0.837836	14.411493	1.041658	
min	1.000000	0.000000	1.000000	0.170000	0.000000	
25%	328.000000	0.000000	2.000000	21.000000	0.000000	
50%	655.000000	0.000000	3.000000	28.000000	0.000000	
75%	982.000000	1.000000	3.000000	39.000000	1.000000	
max	1399.000000	1.000000	3.000000	80.000000	8.000000	

	parch	Fare	wikiId	Age.wiki	Class	\
count	1399.000000	1386.000000	1399.000000	1302.000000	1304.000000	
mean	0.389207	33.294470	658.334009	31.415429	2.291411	
std	0.865560	51.758668	380.377373	13.758954	0.848852	
min	0.000000	0.000000	1.000000	0.170000	1.000000	
25%	0.000000	7.895800	326.750000	21.000000	1.750000	
50%	0.000000	14.454200	661.500000	28.000000	3.000000	
75%	0.000000	31.274000	987.250000	37.750000	3.000000	
max	9.000000	512.329200	1314.000000	74.000000	3.000000	

New folder/prodigy Infotech/ Prodigy\_Infotech\_task2\_DS - Jupyter Notebook

Jupyter Prodigy\_Infotech\_task2\_DS Last Checkpoint: Last Monday at 5:12 PM (autosaved)

In [13]: `summary_stats = data.describe()  
print(summary_stats)`

	Pclass	Survived	Age	SibSp	Parch	Fare	WikiId	Age_wiki	Class
count	1309.000000	891.000000	1309.000000	1046.000000	1309.000000	1308.000000	1304.000000	1302.000000	1304.000000
mean	655.000000	0.383638	2.294882	29.881138	0.498854	33.295479	658.534509	29.415829	2.291411
std	378.020000	0.86592	9.837836	14.414401	1.000000	0.000000	51.758668	38.77373	13.770932
min	0.000000	0.000000	1.000000	0.130000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	328.000000	0.000000	2.000000	21.000000	0.000000	7.895800	326.750000	21.000000	1.750000
50%	655.000000	0.000000	3.000000	28.000000	0.000000	14.45200	661.500000	28.000000	3.000000
75%	982.000000	1.000000	3.000000	39.000000	1.000000	31.275000	987.250000	37.750000	3.000000
max	1309.000000	1.000000	3.000000	80.000000	1.000000	512.529200	1314.000000	74.000000	3.000000

In [14]: `# univariate analysis:  
plt.figure(figsize(12, 5))  
plt.subplot(1, 2, 1)  
sns.histplot(data['Age'].dropna(), kde=True)  
plt.title('Age Distribution')  
plt.subplot(1, 2, 2)  
sns.countplot(data['Sex'])  
plt.title('Gender Distribution')`

