

```
In [4]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [5]: details = pd.read_csv('train.csv')
print(details)

   PassengerId  Survived  Pclass \
0             1         0        3
1             2         1        1
2             3         1        3
3             4         1        1
4             5         0        3
..          ...         ...      ...
886          887         0        2
887          888         1        1
888          889         0        3
889          890         1        1
890          891         0        3

   Name                               Sex  Age  SibSp \
0      Braund, Mr. Owen Harris        male  22.0    1
1  Cumings, Mrs. John Bradley (Florence Briggs Th... female  38.0    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    0
..          ...         ...      ...      ...      ...
886      Montvila, Rev. Juozas         male  27.0    0
887      Graham, Miss. Margaret Edith    female  19.0    0
888  Johnston, Miss. Catherine Helen "Carrie"    female   NaN    1
889      Behr, Mr. Karl Howell          male  26.0    0
890      Dooley, Mr. Patrick            male  32.0    0

   Parch  Ticket   Fare Cabin Embarked
0        0   A/5 21171   7.2500   NaN      S
1        0   PC 17599  71.2833   C85      C
2        0  STON/O2. 3101282   7.9250   NaN      S
3        0  113803  53.1000  C123      S
4        0  373450   8.0500   NaN      S
..      ...         ...      ...      ...      ...
886        0  211536  13.0000   NaN      S
887        0  112053  30.0000  B42      S
888        2   W./C. 6607  23.4500   NaN      S
889        0  111369  30.0000  C148      C
890        0  370376   7.7500   NaN      Q

[891 rows x 12 columns]
```

```
In [6]: details.head()
```

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...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

```
In [7]: details.shape
```

Out[7]: (891, 12)

```
In [8]: details.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  -
0   PassengerId      891 non-null    int64
1   Survived         891 non-null    int64
2   Pclass           891 non-null    int64
3   Name             891 non-null    object
4   Sex              891 non-null    object
5   Age             714 non-null    float64
6   SibSp           891 non-null    int64
7   Parch           891 non-null    int64
8   Ticket           891 non-null    object
9   Fare            891 non-null    float64
10  Cabin           204 non-null    object
11  Embarked         889 non-null    object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

```
In [9]: details.describe()
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

```
In [10]: details.columns.tolist()
```

```
Out[10]: ['PassengerId',
'Survived',
'Pclass',
'Name',
'Sex',
'Age',
'SibSp',
'Parch',
'Ticket',
'Fare',
'Cabin',
'Embarked']
```

```
In [11]: details.isnull().sum()
```

```
Out[11]: 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 [12]: details.dropna(subset=["Embarked"], inplace=True)
```

```
In [13]: details.fillna({'Cabin': 'Unknown'}, inplace = True)
```

```
In [14]: details.fillna({'Age': details['Age'].mean()}, inplace=True)
```

```
In [15]: details.isnull().sum()
```

```
Out[15]: PassengerId    0
Survived              0
Pclass               0
Name                 0
Sex                  0
Age                 0
SibSp                0
Parch               0
Ticket              0
Fare                0
Cabin               0
Embarked            0
dtype: int64
```

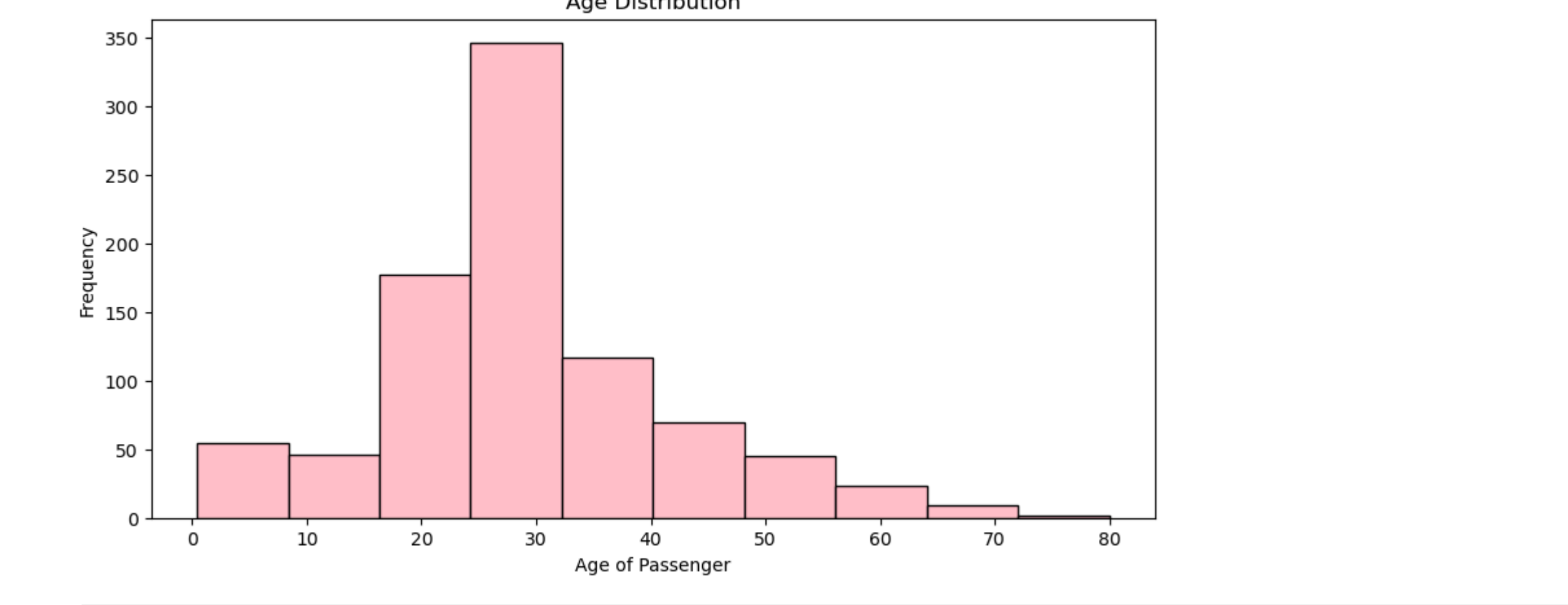
```
In [16]: details.nunique()
```

```
Out[16]: PassengerId    889
Survived              2
Pclass               3
Name                889
Sex                  2
Age                 89
SibSp               7
Parch              7
Ticket             680
Fare              247
Cabin             147
Embarked            3
dtype: int64
```

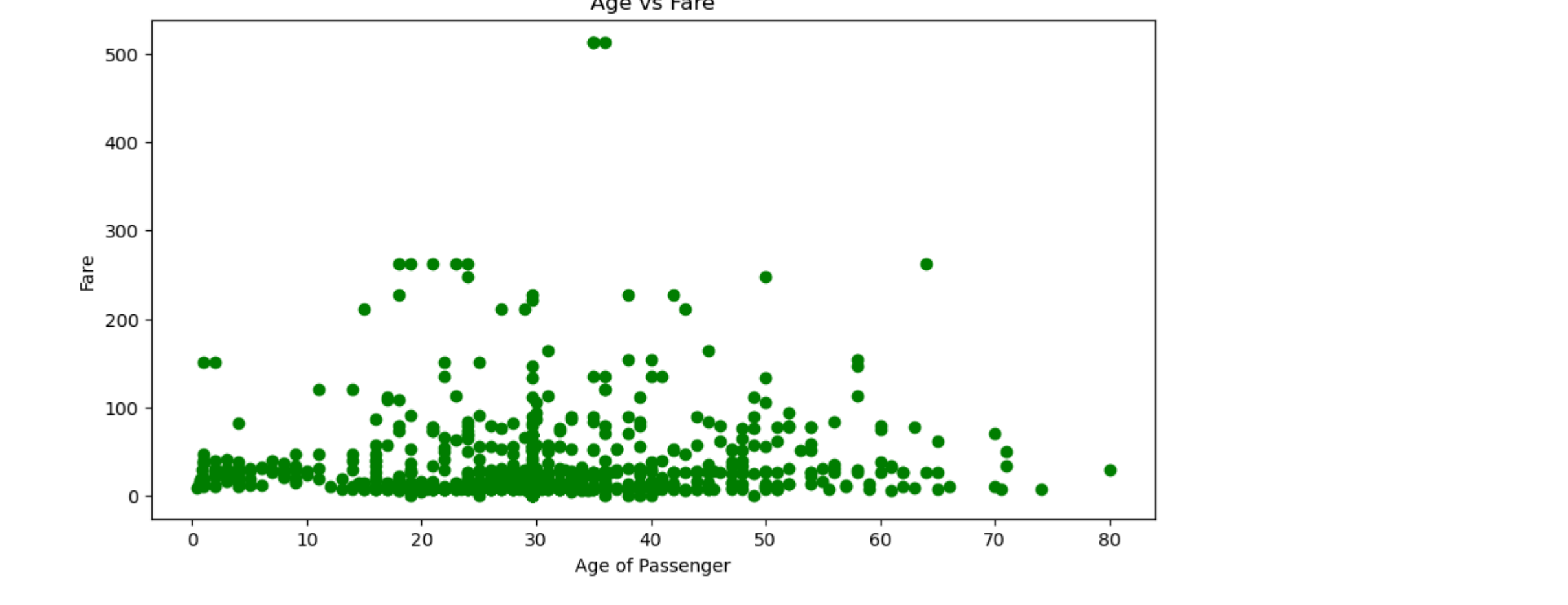
```
In [18]: details.duplicated().sum()
```

Out[18]: 0

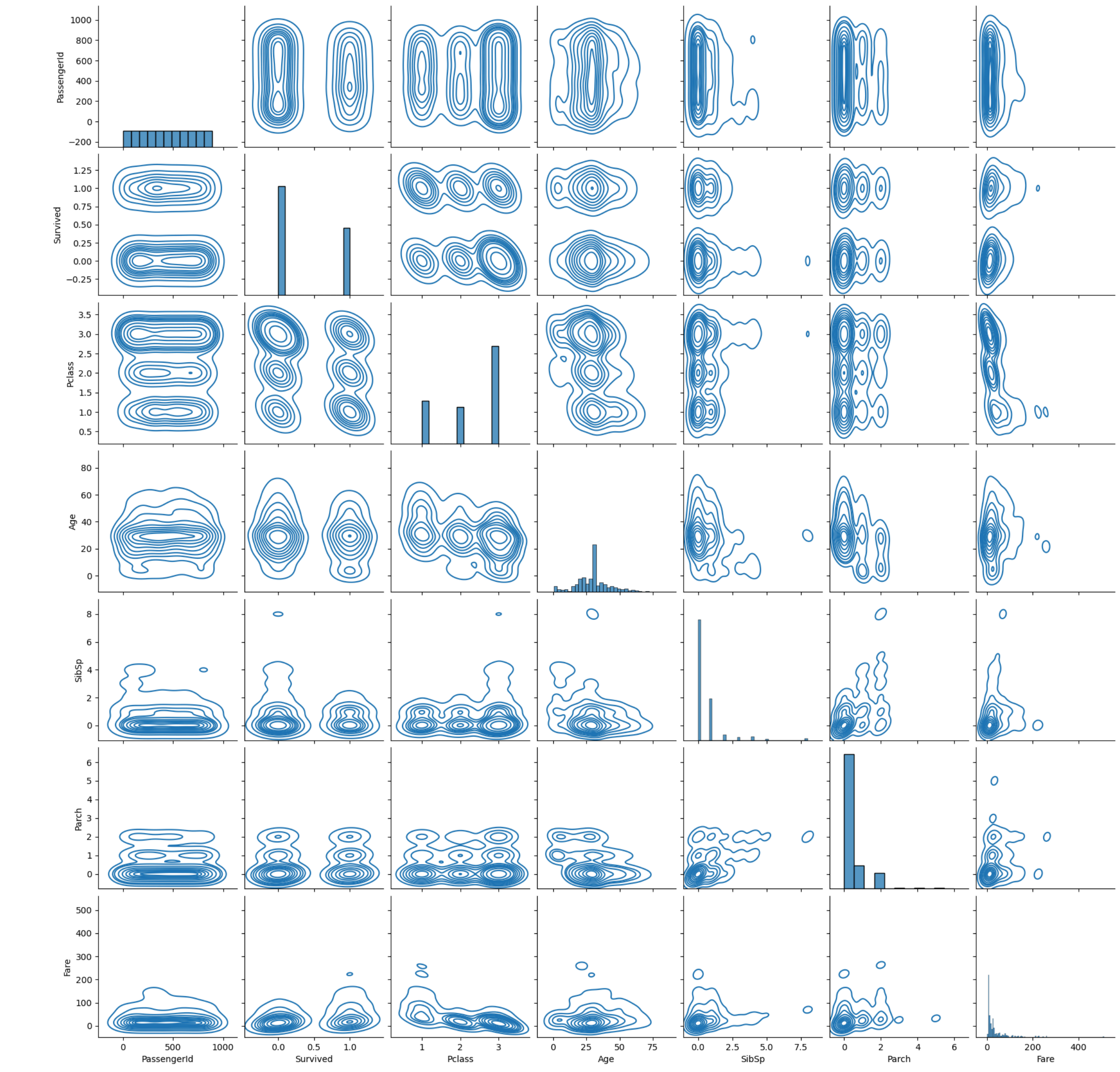
```
In [25]: plt.figure(figsize = (10,5))
plt.hist(details['Age'],color='pink',edgecolor='black',align='mid')
plt.xlabel('Age of Passenger')
plt.ylabel('Frequency')
plt.title('Age Distribution')
plt.show()
```



```
In [34]: plt.figure(figsize=(10,5))
plt.scatter(details['Age'],details['Fare'],color='green')
plt.xlabel('Age of Passenger')
plt.ylabel('Fare')
plt.title('Age vs Fare')
plt.show()
```



```
In [33]: sns.pairplot(details,kind='kde',diag_kind='hist')
```



```
In [ ]:
```