20250412_01

April 12, 2025

[2]:
```
from my_utils import clean_titanic_data, plot_count, age_category

df = clean_titanic_data('original_titanic.csv')
df.head(10)
```

[2]:

	ID	Survived	Class	Name	Gender	Age	Fare
0	1	0	3	Braund, Mr. Owen	male	22.000000	7.2500
1	2	1	1	Cumings, Mrs. John	female	38.000000	71.2833
2	3	1	3	Heikkinen, Miss. Laina	female	26.000000	7.9250
3	4	1	1	Futrelle, Mrs. Jacques	female	35.000000	53.1000
4	5	0	3	Allen, Mr. William	male	35.000000	8.0500
5	6	0	3	Moran, Mr. James	male	28.111111	8.4583
6	7	0	1	McCarthy, Mr. Timothy	male	54.000000	51.8625
7	8	0	3	Palsson, Master. Gosta	male	2.000000	21.0750
8	9	1	3	Johnson, Mrs. Oscar	female	27.000000	11.1333
9	10	1	2	Nasser, Mrs. Nicholas	female	14.000000	30.0708

[4]:
```
subset = df[(df['Class'] == 3) & (df['Fare'] > 30) & (df['Survived'] == 1)]
print(subset)
```

```
Empty DataFrame
Columns: [ID, Survived, Class, Name, Gender, Age, Fare]
Index: []
```

[6]:
```
subset = df[(df['Age'] < 18) & (df['Fare'] > 20) & (df['Survived'] ==0)]
print(subset)
```

	ID	Survived	Class	Name	Gender	Age	Fare
7	8	0	3	Palsson, Master. Gosta	male	2.0	21.075

[8]:
```
Fare_by_Class = df.groupby('Class')['Fare'].mean()
print(Fare_by_Class)
Age_by_Gender = df.groupby('Gender')['Age'].mean()
print(Age_by_Gender)
Fare_by_Gender = df.groupby('Gender')['Fare'].mean()
print(Fare_by_Gender)
```

```
Class
1    58.7486
```

```
2    30.0708
3    10.6486
Name: Fare, dtype: float64
Gender
female    28.000000
male      28.222222
Name: Age, dtype: float64
Gender
female    34.70248
male      19.33916
Name: Fare, dtype: float64
```

[10]:
```
Age_and_Fare_by_Gender = df.groupby('Gender').agg({'Age': 'mean','Fare':␣
↪'mean'})
print(Age_and_Fare_by_Gender)
```

```
              Age      Fare
Gender
female  28.000000  34.70248
male    28.222222  19.33916
```

[12]:
```
Fare_by_Class_and_Gender = df.groupby(['Class','Gender'])['Fare'].mean()
print(Fare_by_Class_and_Gender)
```

```
Class  Gender
1      female    62.191650
       male      51.862500
2      female    30.070800
3      female     9.529150
       male      11.208325
Name: Fare, dtype: float64
```

[14]:
```
Fare_by_Class_and_Gender = df.groupby(['Class','Gender'])['Fare'].mean().
↪reset_index()
print(Fare_by_Class_and_Gender)
```

```
   Class  Gender       Fare
0      1  female  62.191650
1      1    male  51.862500
2      2  female  30.070800
3      3  female   9.529150
4      3    male  11.208325
```

[16]:
```
df['AgeGroup'] = df['Age'].apply(age_category)
df.head(10)
```

[16]:
```
   ID  Survived  Class                     Name  Gender        Age     Fare  \
0   1         0      3        Braund, Mr. Owen    male  22.000000   7.2500
1   2         1      1      Cumings, Mrs. John  female  38.000000  71.2833
```

```
2   3         1      3  Heikkinen, Miss. Laina  female  26.000000   7.9250
3   4         1      1  Futrelle, Mrs. Jacques  female  35.000000  53.1000
4   5         0      3      Allen, Mr. William    male  35.000000   8.0500
5   6         0      3        Moran, Mr. James    male  28.111111   8.4583
6   7         0      1   McCarthy, Mr. Timothy    male  54.000000  51.8625
7   8         0      3  Palsson, Master. Gosta    male   2.000000  21.0750
8   9         1      3     Johnson, Mrs. Oscar  female  27.000000  11.1333
9  10         1      2   Nasser, Mrs. Nicholas  female  14.000000  30.0708

  AgeGroup
0    Adult
1    Adult
2    Adult
3    Adult
4    Adult
5    Adult
6   Senior
7    Child
8    Adult
9    Child
```

[20]:
```
import numpy as np

conditions = [df['Fare'] < 10, (10 <= df['Fare']) & (df['Fare'] < 30), 30 <=␣
 ↪df['Fare']]
choices = ['Low', 'Medium', 'High']

df['FareTier'] = np.select(conditions, choices)
df.head(10)
```

[20]:
```
   ID  Survived  Class                     Name  Gender        Age     Fare  \
0   1         0      3        Braund, Mr. Owen    male  22.000000   7.2500
1   2         1      1     Cumings, Mrs. John  female  38.000000  71.2833
2   3         1      3  Heikkinen, Miss. Laina  female  26.000000   7.9250
3   4         1      1  Futrelle, Mrs. Jacques  female  35.000000  53.1000
4   5         0      3      Allen, Mr. William    male  35.000000   8.0500
5   6         0      3        Moran, Mr. James    male  28.111111   8.4583
6   7         0      1   McCarthy, Mr. Timothy    male  54.000000  51.8625
7   8         0      3  Palsson, Master. Gosta    male   2.000000  21.0750
8   9         1      3     Johnson, Mrs. Oscar  female  27.000000  11.1333
9  10         1      2   Nasser, Mrs. Nicholas  female  14.000000  30.0708

  AgeGroup FareTier
0    Adult      Low
1    Adult     High
2    Adult      Low
3    Adult     High
```

```
4    Adult      Low
5    Adult      Low
6   Senior     High
7    Child   Medium
8    Adult   Medium
9    Child     High
```