titanic

April 9, 2025

[1]:
```
import numpy as np # biblioteca numérica em python
import pandas as pd # biblioteca de dataframe
import seaborn as sns # biblioteca de visualização de dados
import matplotlib.pyplot as plt #biblioteca gráfica
```

[2]:
```
titanic = pd.read_csv('titanic_train.csv')
```

[3]:
```
titanic.head()
```

[3]:

	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…	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

[4]:
```
titanic.tail()
```

[4]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.45	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75	NaN	Q

```
        Sex   Age  SibSp  Parch      Ticket   Fare Cabin Embarked
886    male  27.0      0      0      211536  13.00   NaN        S
887  female  19.0      0      0      112053  30.00   B42        S
888  female   NaN      1      2  W./C. 6607  23.45   NaN        S
889    male  26.0      0      0      111369  30.00  C148        C
890    male  32.0      0      0      370376   7.75   NaN        Q
```

[5]:
```
titanic.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
```

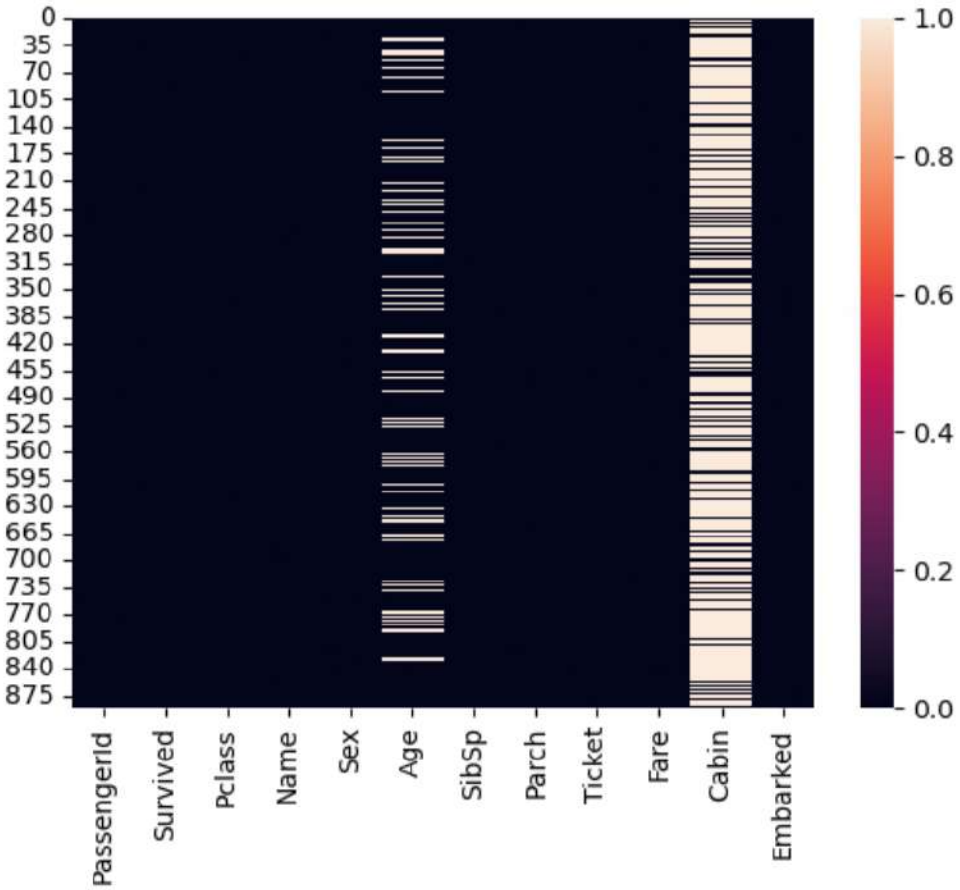
[6]:
```
titanic.isnull()
```

[6]:
```
     PassengerId  Survived  Pclass   Name    Sex    Age  SibSp  Parch  Ticket  \
0          False     False   False  False  False  False  False  False   False
1          False     False   False  False  False  False  False  False   False
2          False     False   False  False  False  False  False  False   False
3          False     False   False  False  False  False  False  False   False
4          False     False   False  False  False  False  False  False   False
..           ...       ...     ...    ...    ...    ...    ...    ...
886        False     False   False  False  False  False  False  False   False
887        False     False   False  False  False  False  False  False   False
888        False     False   False  False  False   True  False  False   False
889        False     False   False  False  False  False  False  False   False
890        False     False   False  False  False  False  False  False   False

      Fare  Cabin  Embarked
0    False   True     False
1    False  False     False
```

```
2    False   True     False
3    False  False     False
4    False   True     False
..     ...    ...       ...
886  False   True     False
887  False  False     False
888  False   True     False
889  False  False     False
890  False   True     False

[891 rows x 12 columns]
```

[7]:
```
sns.heatmap(titanic.isnull())
```

[7]:
```
<AxesSubplot: >
```



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

[9]:
```
176/891
```

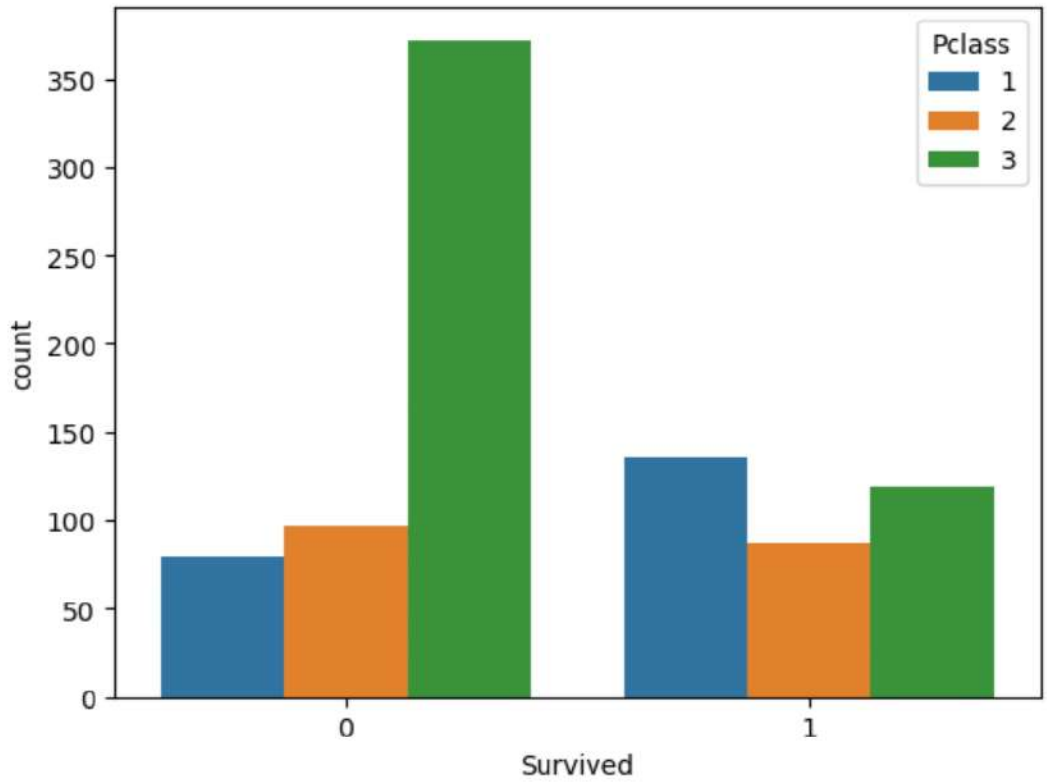
[9]:
```
0.19753086419753085
```

[15]:
```
sns.pairplot(titanic, diag_kind='kde', hue='Survived')
```

[15]:
```
<seaborn.axisgrid.PairGrid at 0x7fcbb7227110>
```

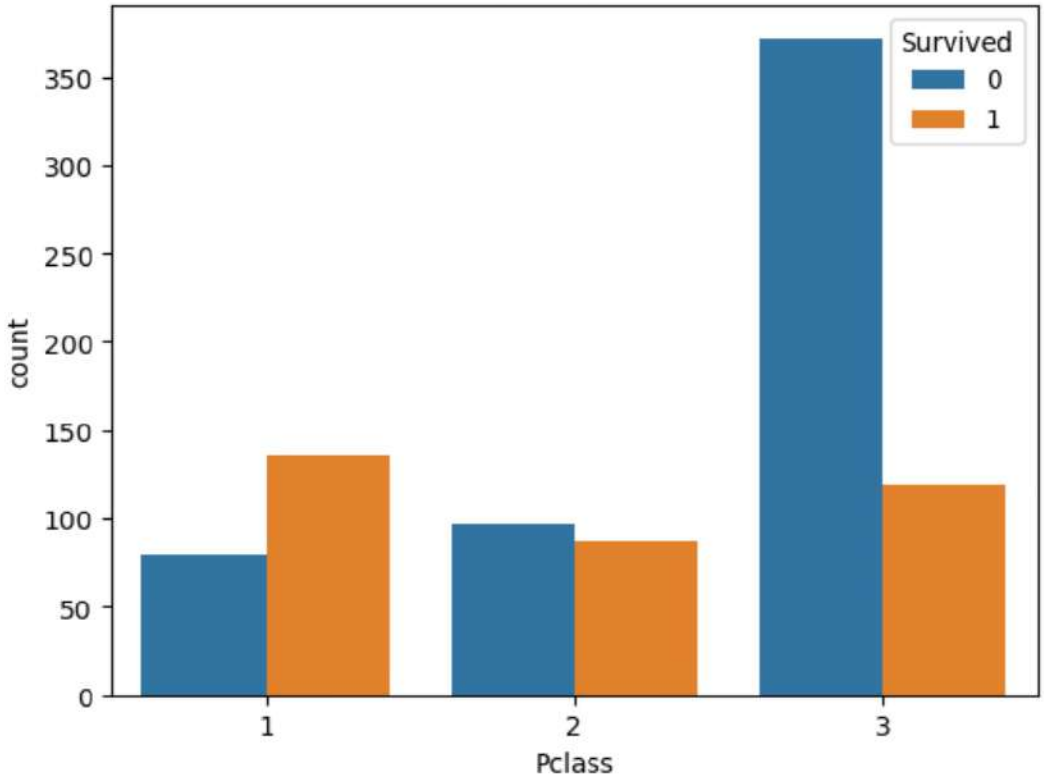
[14]:
```
sns.countplot(x='Survived', hue='Pclass',data=titanic)
```

[14]: <AxesSubplot: xlabel='Survived', ylabel='count'>



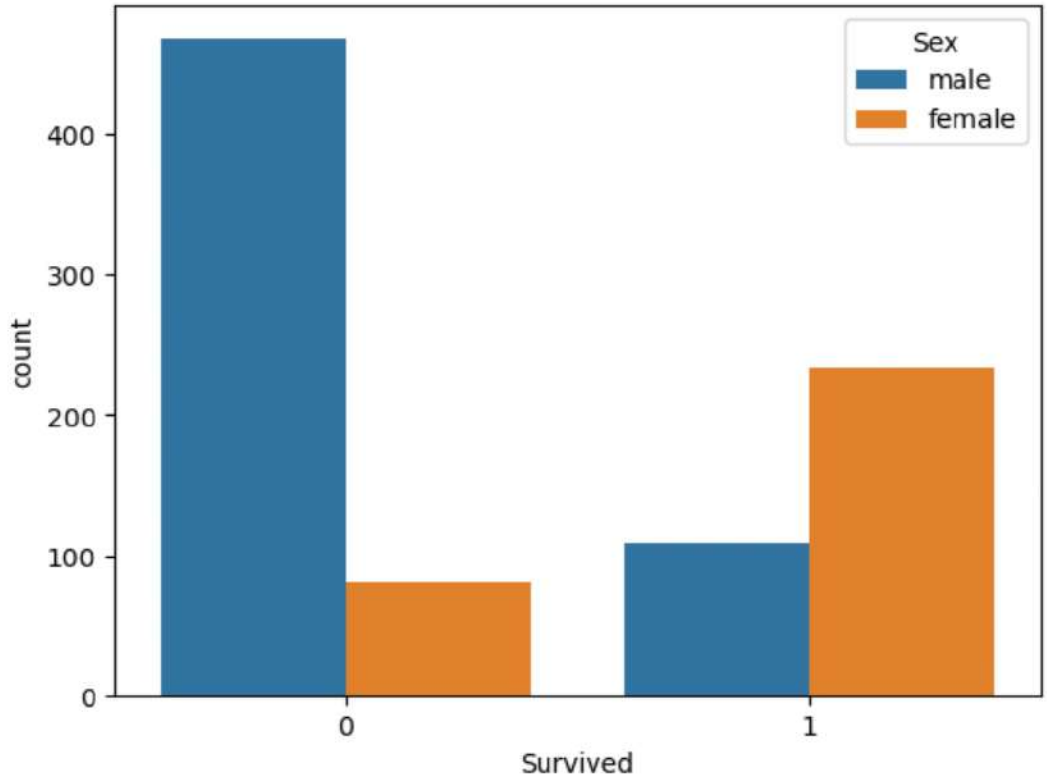
[17]:
```
sns.countplot(x='Pclass',hue='Survived',data=titanic)
```

[17]: <AxesSubplot: xlabel='Pclass', ylabel='count'>



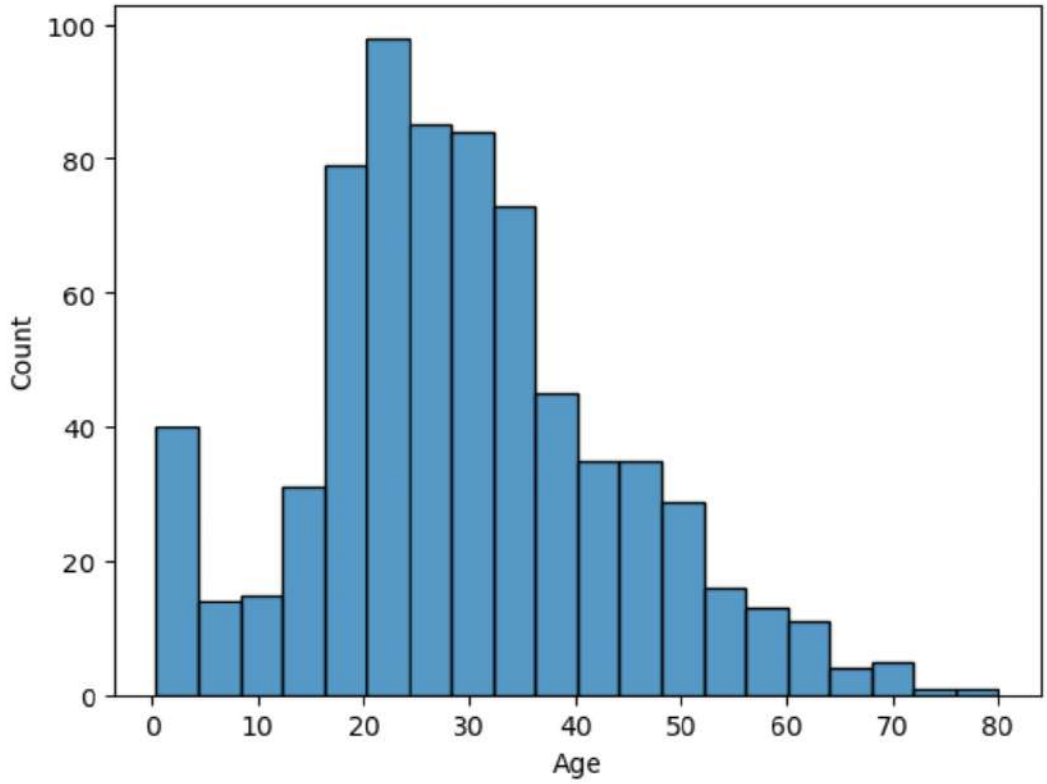
[18]:
```
sns.countplot(x='Survived',hue='Sex',data=titanic)
```

[18]:
```
<AxesSubplot: xlabel='Survived', ylabel='count'>
```



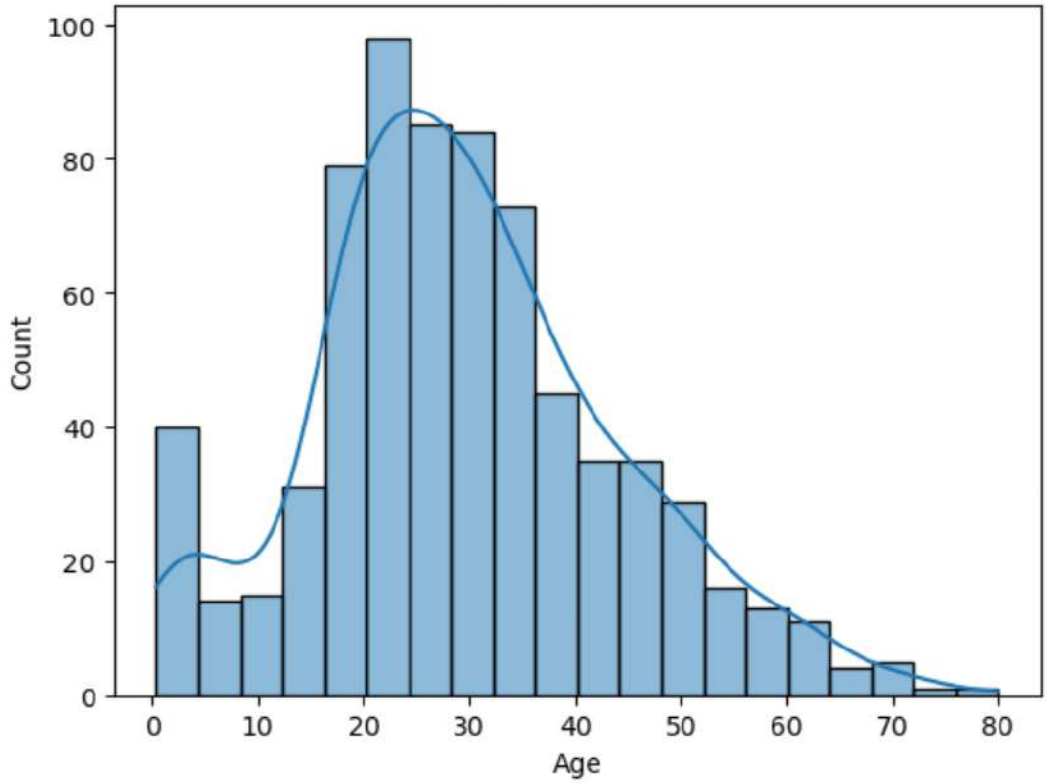
[19]:
```
sns.histplot(titanic['Age'].dropna())
```

[19]: <AxesSubplot: xlabel='Age', ylabel='Count'>



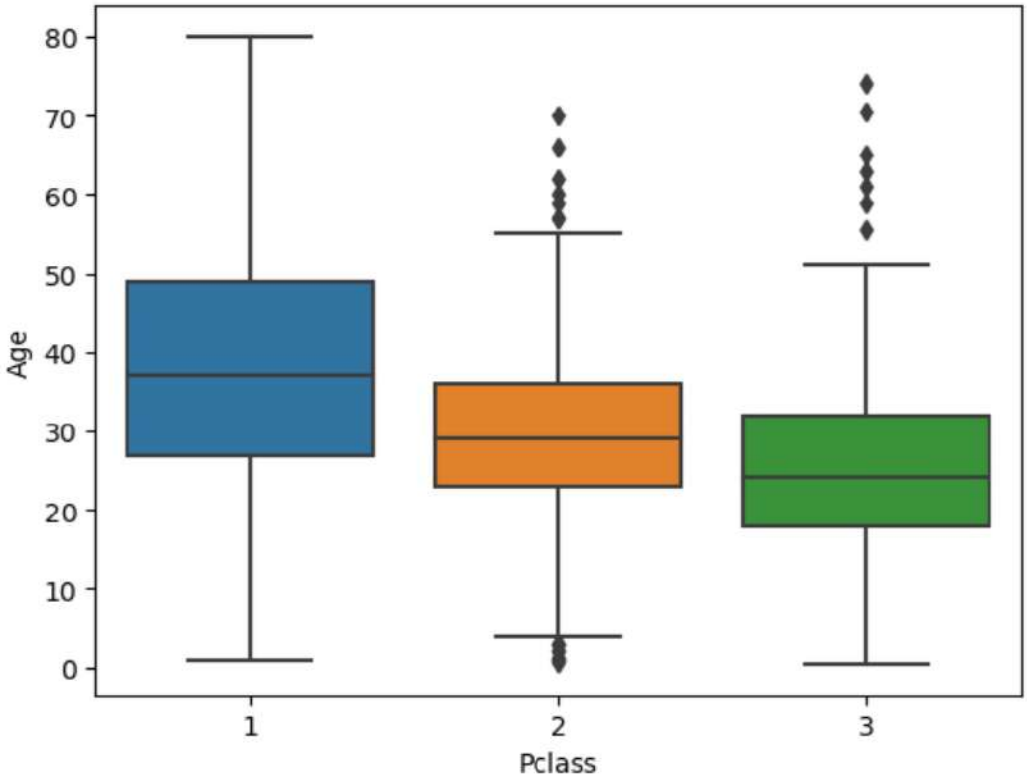
[21]:
```
sns.histplot(titanic['Age'], kde=True)
```

[21]:
```
<AxesSubplot: xlabel='Age', ylabel='Count'>
```



[23]:
```
sns.boxplot(x='Pclass', y='Age', data=titanic)
```

[23]:
```
<AxesSubplot: xlabel='Pclass', ylabel='Age'>
```

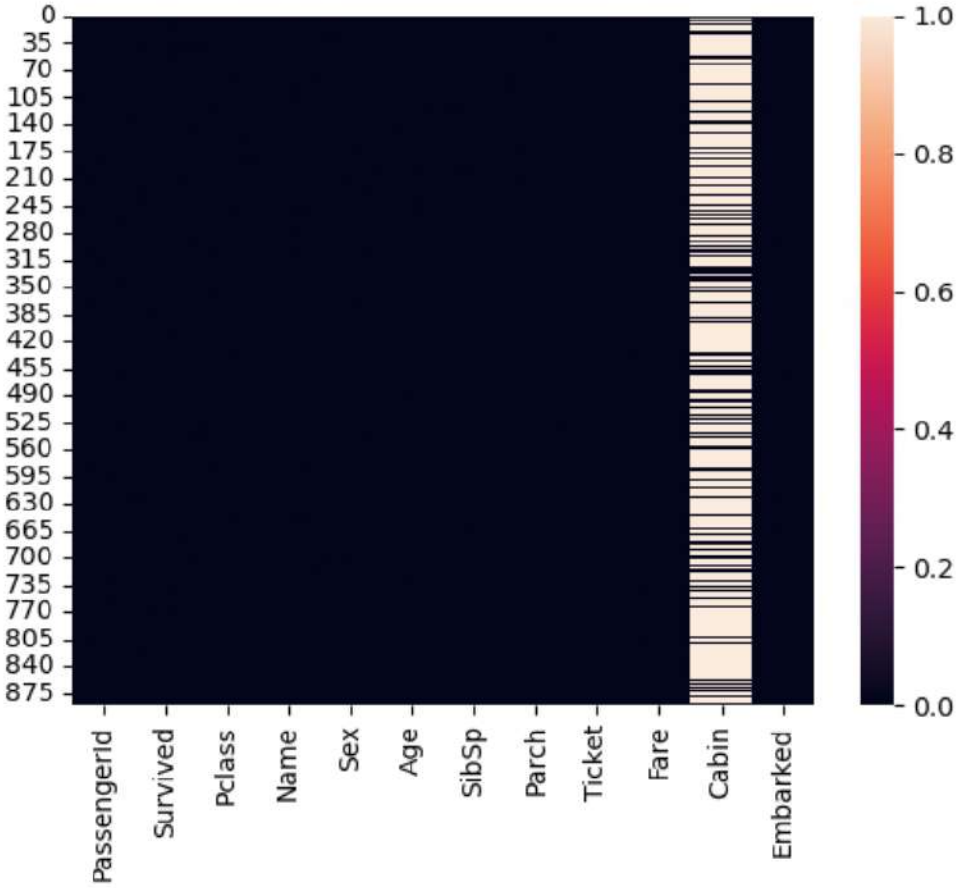


[27]:
```
def input_idade(cols):
    Age=cols[0]
    Pclass=cols[1]

    if pd.isnull(Age):
        if Pclass==1:
            return 39
        elif Pclass==2:
            return 30
        else:
            return 25
    else:
        return Age
```

[30]:
```
titanic['Age']=titanic[['Age','Pclass']].apply(input_idade,axis=1)
```

[31]:
```
sns.heatmap(titanic.isnull())
```

[31]: <AxesSubplot: >



[33]:
```
titanic.drop('Cabin', axis=1, inplace=True)#remover Cabin
```

[35]:
```
sexo = pd.get_dummies(titanic['Sex'], drop_first=True) #booleano sexo
```

[37]:
```
embarque=pd.get_dummies(titanic['Embarked'], drop_first=True)
```

[38]:
```
titanic.drop(['Sex', 'Name', 'Ticket', 'Embarked'], axis=1, inplace=True)
```

[40]:
```
titanic=pd.concat([titanic, sexo, embarque], axis=1)
```

[41]:
```
titanic.head()
```

[41]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	male	Q	S
0	1	0	3	22.0	1	0	7.2500	1	0	1
1	2	1	1	38.0	1	0	71.2833	0	0	0
2	3	1	3	26.0	0	0	7.9250	0	0	1
3	4	1	1	35.0	1	0	53.1000	0	0	1
4	5	0	3	35.0	0	0	8.0500	1	0	1

[42]:
```
casas = pd.read_csv('USA_Housing.csv')
```

[43]:
```
casas.head()
```

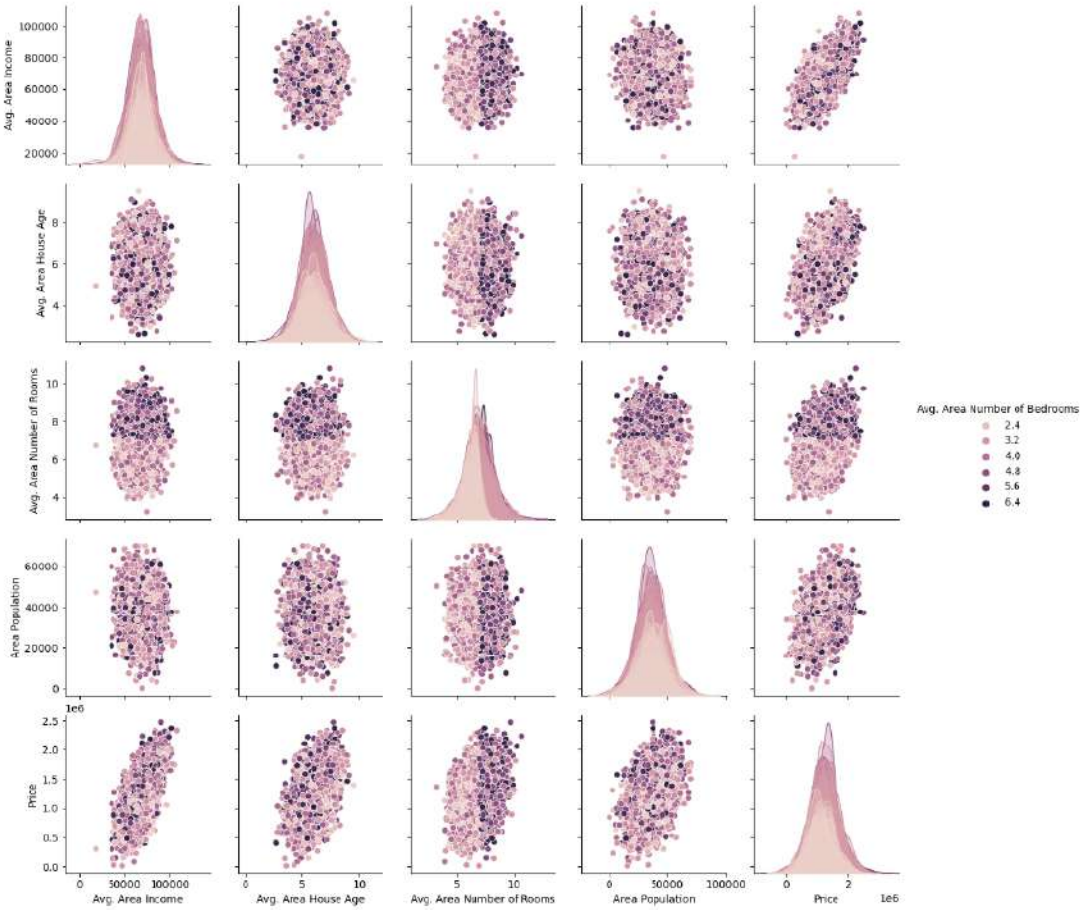
[43]:
```
   Avg. Area Income  Avg. Area House Age  Avg. Area Number of Rooms  \
0      79545.458574             5.682861                   7.009188
1      79248.642455             6.002900                   6.730821
2      61287.067179             5.865890                   8.512727
3      63345.240046             7.188236                   5.586729
4      59982.197226             5.040555                   7.839388

   Avg. Area Number of Bedrooms  Area Population         Price  \
0                          4.09     23086.800503  1.059034e+06
1                          3.09     40173.072174  1.505891e+06
2                          5.13     36882.159400  1.058988e+06
3                          3.26     34310.242831  1.260617e+06
4                          4.23     26354.109472  6.309435e+05

                                             Address
0  208 Michael Ferry Apt. 674\nLaurabury, NE 3701…
1  188 Johnson Views Suite 079\nLake Kathleen, CA…
2  9127 Elizabeth Stravenue\nDanieltown, WI 06482…
3                    USS Barnett\nFPO AP 44820
4                   USNS Raymond\nFPO AE 09386
```

[44]:
```
casas.tail()
```

[44]:
```
      Avg. Area Income  Avg. Area House Age  Avg. Area Number of Rooms  \
4995      60567.944140             7.830362                   6.137356
4996      78491.275435             6.999135                   6.576763
4997      63390.686886             7.250591                   4.805081
4998      68001.331235             5.534388                   7.130144
4999      65510.581804             5.992305                   6.792336

      Avg. Area Number of Bedrooms  Area Population         Price  \
4995                          3.46     22837.361035  1.060194e+06
4996                          4.02     25616.115489  1.482618e+06
4997                          2.13     33266.145490  1.030730e+06
4998                          5.44     42625.620156  1.198657e+06
4999                          4.07     46501.283803  1.298950e+06

                                          Address
4995                 USNS Williams\nFPO AP 30153-7653
4996            PSC 9258, Box 8489\nAPO AA 42991-3352
4997  4215 Tracy Garden Suite 076\nJoshualand, VA 01…
4998                       USS Wallace\nFPO AE 73316
4999  37778 George Ridges Apt. 509\nEast Holly, NV 2…
```

[47]:
```
sns.pairplot(casas, diag_kind='kde', hue='Avg. Area Number of Bedrooms')
```

[47]:
```
<seaborn.axisgrid.PairGrid at 0x7fcbb14ebd90>
```



[50]:
```
casas.corr(numeric_only=True)
```

[50]:

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms
Avg. Area Income	1.000000	-0.002007	-0.011032
Avg. Area House Age	-0.002007	1.000000	-0.009428
Avg. Area Number of Rooms	-0.011032	-0.009428	1.000000
Avg. Area Number of Bedrooms	0.019788	0.006149	
Area Population	-0.016234	-0.018743	
Price	0.639734	0.452543	

```
Avg. Area Number of Bedrooms                  0.462695
Area Population                               0.002040
Price                                         0.335664

                              Avg. Area Number of Bedrooms  Area Population  \
Avg. Area Income                                  0.019788        -0.016234
Avg. Area House Age                               0.006149        -0.018743
Avg. Area Number of Rooms                         0.462695         0.002040
Avg. Area Number of Bedrooms                      1.000000        -0.022168
Area Population                                  -0.022168         1.000000
Price                                             0.171071         0.408556

                                 Price
Avg. Area Income              0.639734
Avg. Area House Age           0.452543
Avg. Area Number of Rooms     0.335664
Avg. Area Number of Bedrooms  0.171071
Area Population               0.408556
Price                         1.000000
```

[53]:
```
sns.heatmap(casas.corr(numeric_only=True), annot=True)
```

[53]:
```
<AxesSubplot: >
```


Avg. Area Income
Avg. Area House Age
Avg. Area Number of Rooms
Avg. Area Number of Bedrooms
Area Population
Price
1
-0.002
-0.011
0.02
-0.016
0.64
-0.002
1
-0.0094
0.0061
-0.019
0.45
-0.011
-0.0094
1
0.46
0.002
0.34
0.02
0.0061
0.46
1
-0.022
0.17
-0.016
-0.019
0.002
-0.022
1
0.41
0.64
0.45
0.34
0.17
0.41
1
1.0
0.8
0.6
0.4
0.2
0.0

[ ]: