qhzivhvu0

February 15, 2025

24bee106, priyansh panchal , experiment-3 , 12th
Feburary 2025

[11]:
```
import pandas as pd
import numpy as np

url='hatsttpesr://t/rraaiwn..cgsitvh'ubusercontent.com/pcsanwald/kaggle-titanic/refs/heads/

f=pd.read_csv(url)
#f=pd.read_csv("train.csv",header='infer') in jupyter notebook
print(f)
df=pd.DataFrame(f)
print(df.head)
print(df.isnull().values.any())
count=df.isnull().sum().sum()
print(count)
print(df.shape)
print(df.describe())
```

```
survived      pclass                                                name\
0 1
0             3                                    Braund,Mr.OwenHarris
2 31          1        Cumings,Mrs.JohnBradley(FlorenceBriggsTh...
4 ..    1     3                                       Heikkinen,Miss.Laina
88      1     1                  Futrelle,Mrs.JacquesHeath(LilyMayPeel)
6       0     23                                     Allen,Mr.WilliamHenry
88      0     1...
7       1  ...  3                                                   ...
88      0     1                                      Montvila,Rev.Juozas
8       0     3                            Graham,Miss.MargaretEdith
88                          Johnston,Miss.CatherineHelen"Carrie"
9                                                       Behr,Mr.KarlHowell
89                                                     Dooley,Mr.Patrick
0

0        sex     agesibsp      parch              ticket        farecabinembarked
1        male22.0        1 0                  A/521171       7.2500 NaN            S
2      female38.0        1 0               PC1759971.2833          C85             C
3      female26.0        0 0          STON/O2.3101282        7.9250 NaN            S
4      female35.0        1 0  0               113803        53.1000C123            S
..       male35.0        0                    3734508.0500NaN                      S
          ...    ...     ...     ...          ...      ...    ...        ...
```

```
886 male       27.0      0       0             211536   13.0000    NaN       S
887 female 19.0          0 1     0             112053 30.0000 B42           S
88    female NaN         0       2         W./C.660723.4500NaN              S
8        mal 26.0        0       0              111369 30.0000 C148         C
889      e      32.0             0            370376   7.7500     NaN       Q
89       mal
0[8   91   rowes   x   11   columns]
<boundmethodNDFrame.heao        survived   pclass
df  nam\ e  0  1  2  3  4  ..  886  887
888 889 8900 1       3                            Braund,Mr.OwenHarris
             1  1    1   Cumings,Mrs.JohnBradley(FlorenceBriggsTh...
             0       3                              Heikkinen,Miss.Laina
                     1        Futrelle,Mrs.JacquesHeath        (LilyMayPeel)
                     3                              Allen,Mr.WilliamHenry
          ...0    ...   2                                        ...
             1       1                                Montvila,Rev.Juozas
             0       3                          Graham,Miss.MargaretEdith
             1       1                 Johnston,Miss.CatherineHelen"Carrie"
             0       3                                 Behr,Mr.KarlHowell
                                                       Dooley,Mr.Patrick

          sex     agesibsp     parch              ticket          farecabinembarked
0 1 2 3female238.0 4      1    0 0             A/521171    7.2500     NaN         S
..female26.0      886     1                    PC17599    71.2833     C85         C
female 35.0               00          STON/O2.3101282       7.9250 NaN          S
                          1 0                   113803 53.1000  C123          S
887     male35.0          0 0                   37345      8.0500    NaN         S
88         ...    ...     ...     ...0           ...      0...      ...        ...     NaN         S
8       male        27.0      0  0  0           211536      13.0000      B42         S
     889 female 19.0      2                     112053 30.0000 NaN           S
89       female NaN       1    0                W./C.660723.4500C148          C
0          mal      26.0      00                      11136930.0000NaN          Q
           e        32.0      0                 3703767.7500
           mal
[891 rowes x 11 columns]>
True
866
(891, 11)
            survived         pclass            age          sibsp          parch            fare
count 891.000000 891.000000 714.000000 891.000000 891.000000 891.000000
mean       0.383838                    2.308642 29.699118 0.523008        0.381594 32.204208
std        0.486592       0.836071 14.526497               1.102743        0.806057 49.693429
min                          0.000000 1.000000 0.420000 0.000000 0.000000 0.000000
25%       0.000000 2.000000 20.125000 0.000000 0.000000                       7.910400
50%          0.000000                   3.000000 28.000000 0.000000 0.000000 14.454200
75%          1.000000    3.000000 38.000000              1.000000 0.000000 31.000000
max        1.000000      3.000000 80.000000         8.000000         6.000000512.329200
```

[12]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 11 columns):
 #   Column Non-NullCount Dtype
---  ------ -------------- -----
 0   survived 891non-null  int64
 1   pclass 891non-null    int64
 2   name 891non-null      object
 3   sex 891non-null       object
 4   age 714non-null       float64
 5   sibsp 891non-null     int64
 6   parch 891non-null     int64
 7   ticket 891non-null    object
 8   fare 891non-null      float64
 9   cabin 204non-null     object
 10 embarked 889non-null   object
dtypes: float64(2), int64(4), object(5)
memory usage: 76.7+ KB
```

[13]:
```
df1=df.copy()
df2=df.copy()
df3=df.copy()
df4=df.copy()
```

[14]:
```
df1=df1.dropna(axis=0)
print(df1.shape)
print(df.shape)
```

```
(183, 11)
(891, 11)
```

[15]:
```
df2=df2.dropna(axis=1)
print(df2.shape)
print(df.shape)
```

```
(891, 8)
(891, 11)
```

[18]:
```
df3.loc[df3.loc[:,'age'].isna(),'age']=22
df3.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 11 columns):
 #   Column Non-NullCount Dtype
---  ------ -------------- -----
```

```
 0   survived    891 non-null    int64
 1   pclass      891 non-null    int64
 2   name        891 non-null    object
 3   sex         891 non-null    object
 4   age         891 non-null    float64
 5   sibsp       891 non-null    int64
 6   parch       891 non-null    int64
 7   ticket      891 non-null    object
 8   fare        891 non-null    float64
 9   cabin       204 non-null    object
 10  embarked    889 non-null    object
dtypes: float64(2), int64(4), object(5)
memory usage: 76.7+ KB
```

[20]:
```
df3.loc[:,"embarked"].unique()
```

[20]:
```
array(['S', 'C', 'Q', nan], dtype=object)
```

[21]:
```
df3.loc[df3.loc[:,"embarked"].isna(),"embarked"]='df'
df3.info()
df3.loc[:,"cabin"].unique()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 11 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   survived  891 non-null    int64
 1   pclass    891 non-null    int64
 2   name      891 non-null    object
 3   sex       891 non-null    object
 4   age       891 non-null    float64
 5   sibsp     891 non-null    int64
 6   parch     891 non-null    int64
 7   ticket    891 non-null    object
 8   fare      891 non-null    float64
 9   cabin     204 non-null    object
 10  embarked  891 non-null    object
dtypes: float64(2), int64(4), object(5)
memory usage: 76.7+ KB
```

[21]:
```
array([nan, 'C85', 'C123', 'E46', 'G6', 'C103', 'D56', 'A6',
       'C23 C25 C27', 'B78', 'D33', 'B30', 'C52', 'B28', 'C83', 'F33',
       'F G73', 'E31', 'A5', 'D10 D12', 'D26', 'C110', 'B58 B60', 'E101',
       'F E69', 'D47', 'B86', 'F2', 'C2', 'E33', 'B19', 'A7', 'C49', 'F4',
       'A32', 'B4', 'B80', 'A31', 'D36', 'D15', 'C93', 'C78', 'D35',
       'C87', 'B77', 'E67', 'B94', 'C125', 'C99', 'C118', 'D7', 'A19',
       'B49', 'D', 'C22 C26', 'C106', 'C65', 'E36', 'C54',
```

```
'B57 B59 B63 B66', 'C7', 'E34', 'C32', 'B18', 'C124', 'C91', 'E40',
'T', 'C128', 'D37', 'B35', 'E50', 'C82', 'B96 B98', 'E10', 'E44',
'A34', 'C104', 'C111', 'C92', 'E38', 'D21', 'E12', 'E63', 'A14',
'B37', 'C30', 'D20', 'B79', 'E25', 'D46', 'B73', 'C95', 'B38',
'B39', 'B22', 'C86', 'C70', 'A16', 'C101', 'C68', 'A10', 'E68',
'B41', 'A20', 'D19', 'D50', 'D9', 'A23', 'B50', 'A26', 'D48',
'E58', 'C126', 'B71', 'B51 B53 B55', 'D49', 'B5', 'B20', 'F G63',
'C62 C64', 'E24', 'C90', 'C45', 'E8', 'B101', 'D45', 'C46', 'D30',
'E121', 'D11', 'E77', 'F38', 'B3', 'D6', 'B82 B84', 'D17', 'A36',
'B102', 'B69', 'E49', 'C47', 'D28', 'E17', 'A24', 'C50', 'B42',
'C148'], dtype=object)
```

[22]:
```
df3.loc[df3.loc[:,"cabin"].isna(),"cabin"]='D28'
df3.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 11 columns):
 #   Column Non-NullCount Dtype
---  ------ -------------- -----
 0   survived 891non-null int64
 1   pclass 891non-null int64
 2   name 891non-null object
 3   sex 891non-null object
 4   age 891non-null float64
 5   sibsp 891non-null int64
 6   parch 891non-null int64
 7   ticket 891non-null object
 8   fare 891non-null float64
 9   cabin 891non-null object
 10 embarked 891non-null object
dtypes: float64(2), int64(4), object(5)
memory usage: 76.7+ KB
```

[27]:
```
m = np.mean(df4.loc[~df4.loc[:,"age"].isna(),"age"].values)
df4.loc[df4.loc[:,"age"].isna(),"age"]=m
print(m)
df4.info()
```

```
29.69911764705882
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 11 columns):
 #   Column Non-NullCount Dtype
---  ------ -------------- -----
 0   survived 891non-null int64
 1   pclass 891non-null int64
 2   name 891non-null object
```

```
 3    sex    age891non-null          object
 4 sibsp              891 non-null   float64
 5       parch        891 non-null   int64
      6 ticket        891 non-null   int64
 7    fare            891 non-null   object
 8 cabin              891 non-null   float64
        9 embarke 891 non-null       object
 10   d              891non-null     object
dtypes: float64(2), int64(4), object(5)
memory usage: 76.7+ KB
```

[26]:
```python
mc =df4.loc[:,"cabin"].mode()[0]
print(mc)
df4.loc[df4.loc[:,"cabin"].isna(),"cabin"]=mc
df4.info()
```

```
B96 B98
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 11 columns):
 #    Column Non-NullCount Dtype
---------------                      -----
0    survived  891non-null           int64
1    pclass    891non-null           int64
2    name      891non-null           object
3    sex       891non-null           object
4    age       891non-null           float64
5    sibsp     891non-null           int64
6    parch     891non-null           int64
7    ticket    891non-null           object
8    fare      891non-null           float64
9    cabin     891non-null           object
10 embarked 891non-null              object
dtypes: float64(2), int64(4), object(5)
memory usage: 76.7+ KB
```

[25]:
```python
mp =df4.loc[:,"embarked"].mode()[0]
print(mp)
df4.loc[df4.loc[:,"embarked"].isna(),"embarked"]=mp
df4.info()
```

```
S
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 11 columns):
 #    Column Non-NullCount Dtype
---   ------     --------------      -----
0     survived   891non-null         int64
```

```
 1 2   pclass      891non-null      int64
 3 name            891 non-null     object
 4     sex   age891non-null         object
 5     sibsp       891    non-null  float64
 6 parch           891    non-null  int64
 7     ticket      891    non-null  int64
 8 fare            891    non-null  object
 9 cabin           891 non-null     float64
 10 embarke891non-null              object
       d           891non-null      object
dtypes: float64(2), int64(4), object(5)
memory usage: 76.7+ KB
```