```
%matplotlib inline
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import sklearn
import seaborn as sns
```

```
train_data_raw1 = pd.read_csv("train_data.csv")
train_data_raw1
```

	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region_code	Available Extra Rooms in Hospital
0	1	8	c	3	Z	3
1	2	2	c	5	Z	2
2	3	10	e	1	X	2
3	4	26	b	2	Y	2
4	5	26	b	2	Y	2
...	...	...	...	...	...	...
318433	318434	6	a	6	X	3
318434	318435	24	a	1	X	2
318435	318436	7	a	4	X	3
318436	318437	11	b	2	Y	3
318437	318438	19	a	7	Y	5

318438 rows × 18 columns

```
submit = pd.read_csv("lg.csv")
```

```
test_raw = pd.read_csv("test_data.csv")
test_raw
```

	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region_code	Available Extra Rooms in Hospital
0	318439	21	c	3	Z	3
1	318440	29	a	4	X	2
2	318441	26	b	2	Y	3
3	318442	6	a	6	X	3
4	318443	28	b	11	X	[illegible]

```
from sklearn import preprocessing
lb = preprocessing.LabelEncoder()
lb.fit_transform(train_data_raw1.Stay)
submit.Stay = lb.inverse_transform(submit.Stay)
```

```
submit
```

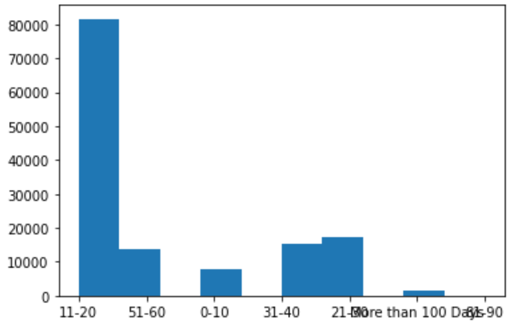
	case_id	Stay
0	318439	11-20
1	318440	51-60
2	318441	0-10
3	318442	11-20
4	318443	31-40
...	...	...
137052	455491	11-20
137053	455492	11-20
137054	455493	11-20
137055	455494	11-20
137056	455495	31-40

137057 rows × 2 columns

```
submit.to_csv("Smote_cat.csv",index=False)
```

```
plt.hist(submit.Stay)
```

```
(array([8.1621e+04, 1.3520e+04, 0.0000e+00, 7.7600e+03, 0.0000e+00,
        1.5456e+04, 1.7309e+04, 0.0000e+00, 1.3490e+03, 4.2000e+01]),
 array([0. , 0.6, 1.2, 1.8, 2.4, 3. , 3.6, 4.2, 4.8, 5.4, 6. ]),
 <BarContainer object of 10 artists>)
```



```
test_raw.isnull().sum()
```

```
case_id                              0
Hospital_code                        0
Hospital_type_code                   0
City_Code_Hospital                   0
Hospital_region_code                 0
```

```
    Available Extra Rooms in Hospital       0
    Department                              0
    Ward_Type                               0
    Ward_Facility_Code                      0
    Bed Grade                              35
    patientid                               0
    City_Code_Patient                    2157
    Type of Admission                       0
    Severity of Illness                     0
    Visitors with Patient                   0
    Age                                     0
    Admission_Deposit                       0
    dtype: int64
```

```
train_data_raw1.fillna(value=np.nan,inplace=True)
```

```
from sklearn.impute import SimpleImputer
si = SimpleImputer(missing_values = np.nan,
                   strategy ='median')
si.fit(train_data_raw1[['Bed Grade','City_Code_Patient']])
bg = si.transform(test_raw[['Bed Grade','City_Code_Patient']])
bg
```

```
    array([[ 2.,  2.],
           [ 2.,  2.],
           [ 4.,  2.],
           ...,
           [ 4., 12.],
           [ 4., 10.],
           [ 4.,  3.]])
```

```
test_raw['Bed Grade'] = bg[:,0]
test_raw['City_Code_Patient'] = bg[:,1]
```

```
test_raw
```

	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region_code	Availabl Extr Rooms i Hospita
0	318439	21	c	3	Z	
1	318440	29	a	4	X	
2	318441	26	b	2	Y	
3	318442	6	a	6	X	
4	318443	28	b	11	X	
...	...	...	...	...	...	..
137052	455491	11	b	2	Y	
137053	455492	25	e	1	X	
137054	455493	30	c	3	Z	
137055	455494	5	a	1	X	
137056	455495	6	a	6	X	

137057 rows × 17 columns

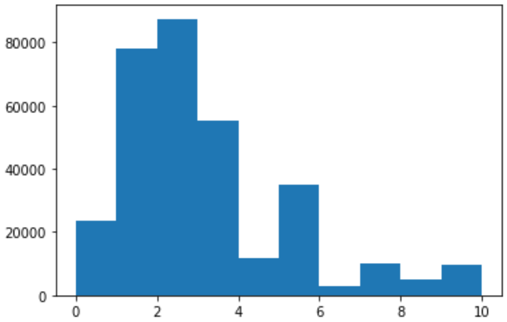
```
from sklearn import preprocessing
lb = preprocessing.LabelEncoder()
```

```
train_data_raw1.Stay = lb.fit_transform(train_data_raw1.Stay)
train_data_raw1.Stay
```

```
0         0
1         4
2         3
3         4
4         4
         ..
318433    1
318434    3
318435    1
318436    1
318437    0
Name: Stay, Length: 318438, dtype: int64
```

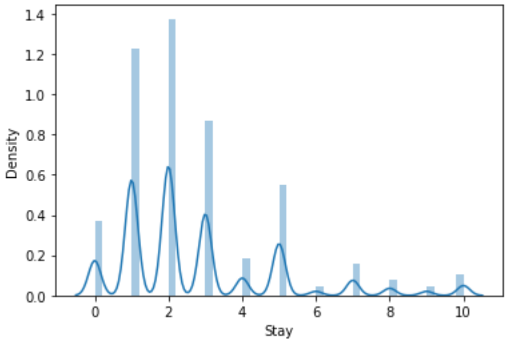
```
plt.hist(train_data_raw1.Stay)
```

```
(array([23604., 78139., 87491., 55159., 11743., 35018.,  2744., 10254.,
         4838.,  9448.]),
 array([ 0.,  1.,  2.,  3.,  4.,  5.,  6.,  7.,  8.,  9., 10.]),
 <BarContainer object of 10 artists>)
```



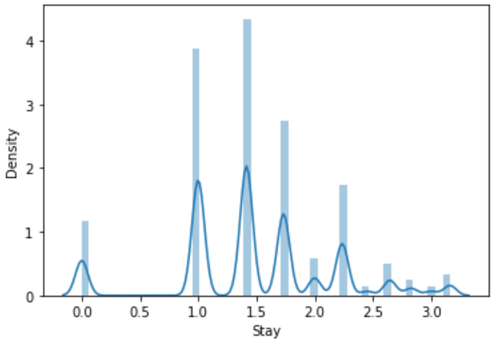
```
sns.distplot(train_data_raw1.Stay)
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a de
  warnings.warn(msg, FutureWarning)
<AxesSubplot:xlabel='Stay', ylabel='Density'>
```



```
sns.distplot(np.sqrt(train_data_raw1.Stay))
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a de
  warnings.warn(msg, FutureWarning)
<AxesSubplot:xlabel='Stay', ylabel='Density'>
```



```
train_data_raw1.columns
```

```
Index(['case_id', 'Hospital_code', 'Hospital_type_code', 'City_Code_Hospital',
       'Hospital_region_code', 'Available Extra Rooms in Hospital',
       'Department', 'Ward_Type', 'Ward_Facility_Code', 'Bed Grade',
       'patientid', 'City_Code_Patient', 'Type of Admission',
       'Severity of Illness', 'Visitors with Patient', 'Age',
       'Admission_Deposit', 'Stay'],
      dtype='object')
```

```
train_data_raw1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 318438 entries, 0 to 318437
Data columns (total 18 columns):
 #   Column                             Non-Null Count   Dtype  
---  ------                             --------------   -----  
 0   case_id                            318438 non-null  int64  
 1   Hospital_code                      318438 non-null  int64  
 2   Hospital_type_code                 318438 non-null  object 
 3   City_Code_Hospital                 318438 non-null  int64  
 4   Hospital_region_code               318438 non-null  object 
 5   Available Extra Rooms in Hospital  318438 non-null  int64  
 6   Department                         318438 non-null  object 
 7   Ward_Type                          318438 non-null  object 
 8   Ward_Facility_Code                 318438 non-null  object 
 9   Bed Grade                          318325 non-null  float64
 10  patientid                          318438 non-null  int64  
 11  City_Code_Patient                  313906 non-null  float64
 12  Type of Admission                  318438 non-null  object 
 13  Severity of Illness                318438 non-null  object 
 14  Visitors with Patient              318438 non-null  int64  
 15  Age                                318438 non-null  object 
 16  Admission_Deposit                  318438 non-null  float64
 17  Stay                               318438 non-null  int64  
dtypes: float64(3), int64(7), object(8)
memory usage: 43.7+ MB
```

```
# type(col)
for i in train_data_raw1.columns:
    print(i, end=' ')
    print(train_data_raw1[i].nunique())
```

```
case_id 318438
Hospital_code 32
Hospital_type_code 7
City_Code_Hospital 11
Hospital_region_code 3
Available Extra Rooms in Hospital 18
Department 5
Ward_Type 6
Ward_Facility_Code 6
Bed Grade 4
patientid 92017
City_Code_Patient 37
Type of Admission 3
Severity of Illness 3
Visitors with Patient 28
Age 10
Admission_Deposit 7300
Stay 11
```

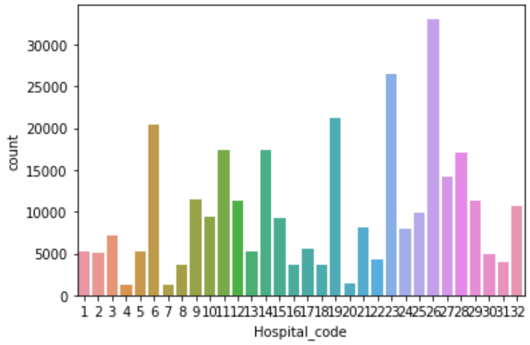
```
cat =  ['Hospital_code', 'Hospital_type_code', 'City_Code_Hospital',
       'Hospital_region_code', 'Available Extra Rooms in Hospital',
       'Department', 'Ward_Type', 'Ward_Facility_Code', 'Bed Grade',
        'City_Code_Patient', 'Type of Admission',
       'Severity of Illness', 'Visitors with Patient', 'Age',
       'Admission_Deposit', 'Stay']
print(len(cat))
train_data_raw1[cat]
```

```
16
```

	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region_code	Available Extra Rooms in Hospital	Departı
0	8	c	3	Z	3	radiothe
1	2	c	5	Z	2	radiothe
2	10	e	1	X	2	anesth
3	26	b	2	Y	2	radiothe
4	26	b	2	Y	2	radiothe
...	...	...	...	...	...	
318433	6	a	6	X	3	radiothe
318434	24	a	1	X	2	anesth
318435	7	a	4	X	3	gynecc
318436	11	b	2	Y	3	anesth

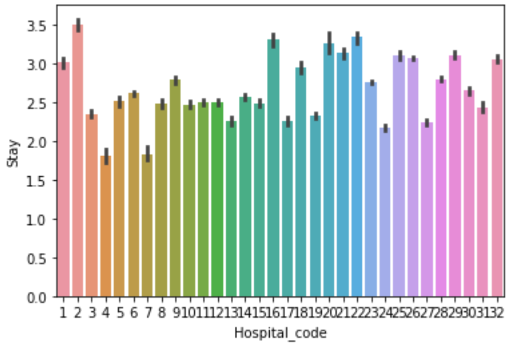
```
sns.countplot(data = train_data_raw1[cat] , x = 'Hospital_code',)
```

```
<AxesSubplot:xlabel='Hospital_code', ylabel='count'>
```



```
sns.barplot(x= "Hospital_code",y= "Stay", data = train_data_raw1)
```

```
<AxesSubplot:xlabel='Hospital_code', ylabel='Stay'>
```



```
gfmlasl = train_data_raw1[train_data_raw1['Hospital_code'] == 26].groupby('Stay')
gfmlasl.get_group(1)
```

	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region_code	Available Extra Rooms in Hospital
11	12	26	b	2	Y	4
26	27	26	b	2	Y	4
107	108	26	b	2	Y	2
156	157	26	b	2	Y	3
163	164	26	b	2	Y	4
...	...	...	...	...	...	..
318079	318080	26	b	2	Y	3
318094	318095	26	b	2	Y	3
318221	318222	26	b	2	Y	5
318262	318263	26	b	2	Y	3
318298	318299	26	b	2	Y	3

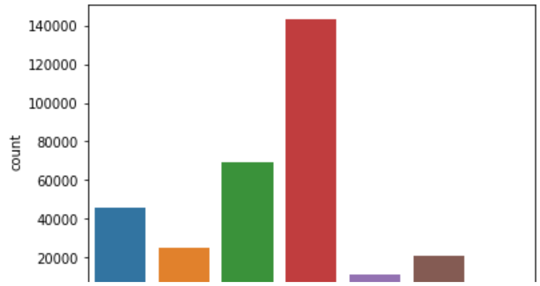
```
train_data_raw1[train_data_raw1['Hospital_code'] == 26] #[train_data_raw1['Hospital_type_code'] == 'b']
```

	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region_code	Available Extra Rooms in Hospital
3	4	26	b	2	Y	2
4	5	26	b	2	Y	2
11	12	26	b	2	Y	4
24	25	26	b	2	Y	4
26	27	26	b	2	Y	4
...	...	...	...	...	...	..
318336	318337	26	b	2	Y	6
318338	318339	26	b	2	Y	4
318340	318341	26	b	2	Y	3
318380	318381	26	b	2	Y	4
318432	318433	26	b	2	Y	2

33076 rows × 18 columns

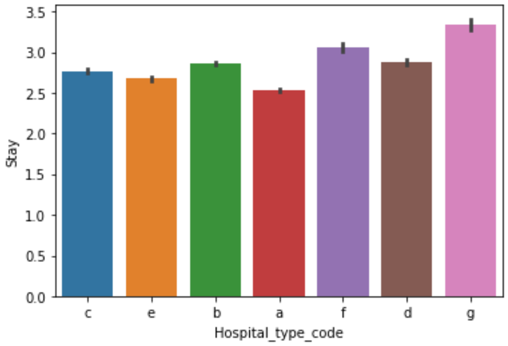
```
sns.countplot(data = train_data_raw1[cat] , x = 'Hospital_type_code')
```

```
<AxesSubplot:xlabel='Hospital_type_code', ylabel='count'>
```



```
sns.barplot(x= "Hospital_type_code",y= "Stay", data = train_data_raw1)
```

```
<AxesSubplot:xlabel='Hospital_type_code', ylabel='Stay'>
```



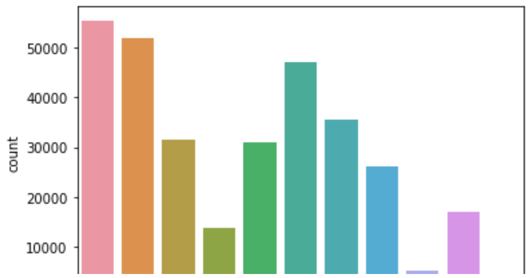
```
train_data_raw1[train_data_raw1['Hospital_type_code'] == 'a']
```

	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region_code	Availabl Extr Rooms i Hospita
5	6	23	a	6	X	2
7	8	23	a	6	X	4
14	15	6	a	6	X	4
15	16	6	a	6	X	3
16	17	23	a	6	X	4
...	...	...	...	...	...	..
318431	318432	23	a	6	X	3
318433	318434	6	a	6	X	3
318434	318435	24	a	1	X	2
318435	318436	7	a	4	X	3
318437	318438	19	a	7	Y	5

143425 rows × 18 columns

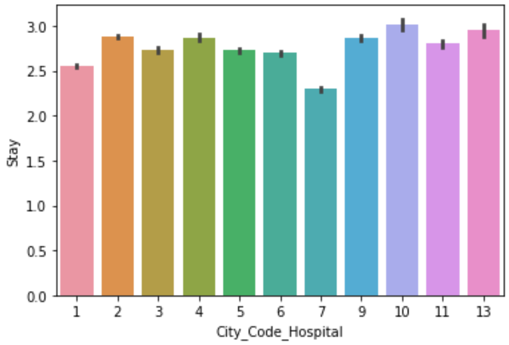
```
sns.countplot(data = train_data_raw1[cat] , x = 'City_Code_Hospital')
```

```
<AxesSubplot:xlabel='City_Code_Hospital', ylabel='count'>
```



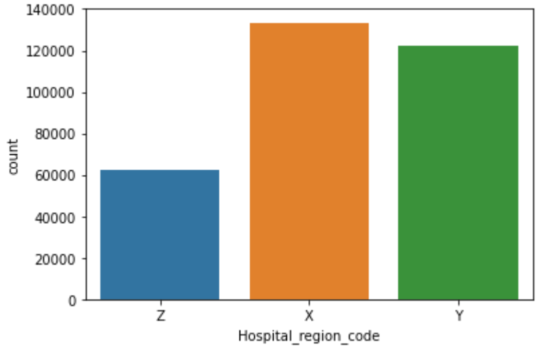
```
sns.barplot(x= "City_Code_Hospital",y= "Stay", data = train_data_raw1)
```

```
<AxesSubplot:xlabel='City_Code_Hospital', ylabel='Stay'>
```



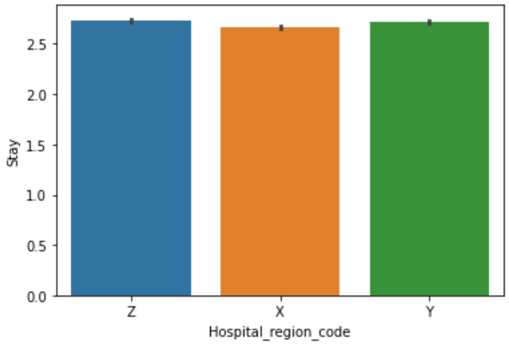
```
sns.countplot(data = train_data_raw1[cat] , x = 'Hospital_region_code')
```

```
<AxesSubplot:xlabel='Hospital_region_code', ylabel='count'>
```



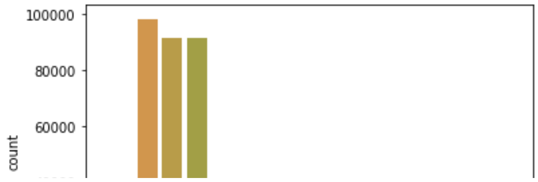
```
sns.barplot(x= "Hospital_region_code",y= "Stay", data = train_data_raw1)
```

```
<AxesSubplot:xlabel='Hospital_region_code', ylabel='Stay'>
```



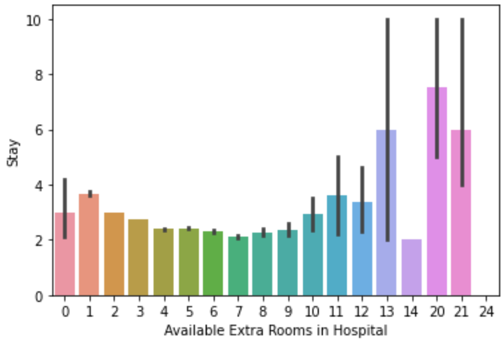
```
sns.countplot(data = train_data_raw1[cat] , x = 'Available Extra Rooms in Hospital')
```

```
<AxesSubplot:xlabel='Available Extra Rooms in Hospital', ylabel='count'>
```



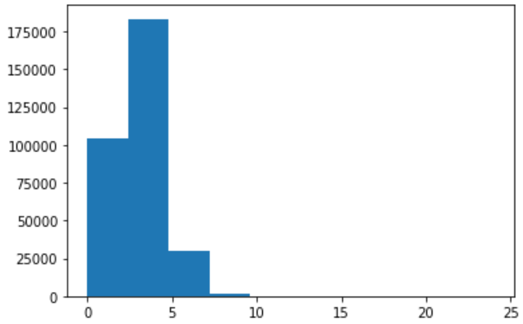
```
sns.barplot(x= "Available Extra Rooms in Hospital",y= "Stay", data = train_data_raw1)
```

```
<AxesSubplot:xlabel='Available Extra Rooms in Hospital', ylabel='Stay'>
```



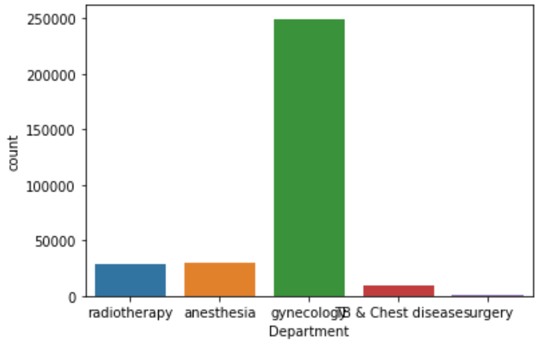
```
plt.hist(train_data_raw1['Available Extra Rooms in Hospital'])
```

```
(array([1.03908e+05, 1.83229e+05, 2.99500e+04, 1.24500e+03, 8.40000e+01,
        1.60000e+01, 0.00000e+00, 0.00000e+00, 5.00000e+00, 1.00000e+00]),
 array([ 0. ,  2.4,  4.8,  7.2,  9.6, 12. , 14.4, 16.8, 19.2, 21.6, 24. ]),
 <BarContainer object of 10 artists>)
```



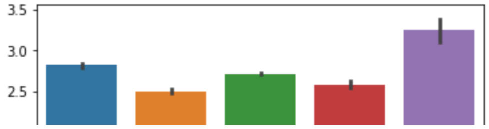
```
sns.countplot(data = train_data_raw1[cat] , x = 'Department')
```

```
<AxesSubplot:xlabel='Department', ylabel='count'>
```



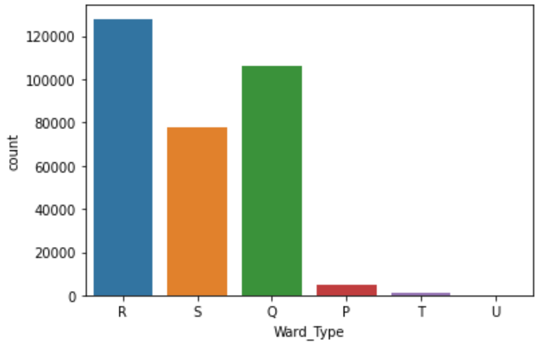
```
sns.barplot(x= "Department",y= "Stay", data = train_data_raw1)
```

```
<AxesSubplot:xlabel='Department', ylabel='Stay'>
```



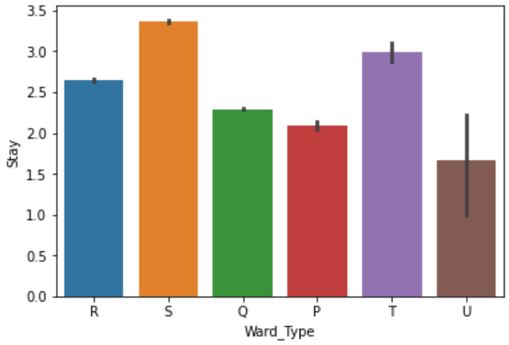
```
sns.countplot(data = train_data_raw1[cat] , x = 'Ward_Type')
```

```
<AxesSubplot:xlabel='Ward_Type', ylabel='count'>
```



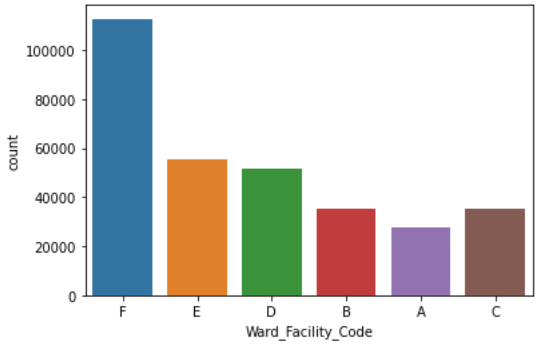
```
sns.barplot(x= "Ward_Type",y= "Stay", data = train_data_raw1)
```

```
<AxesSubplot:xlabel='Ward_Type', ylabel='Stay'>
```



```
sns.countplot(data = train_data_raw1[cat] , x = 'Ward_Facility_Code')
```

```
<AxesSubplot:xlabel='Ward_Facility_Code', ylabel='count'>
```

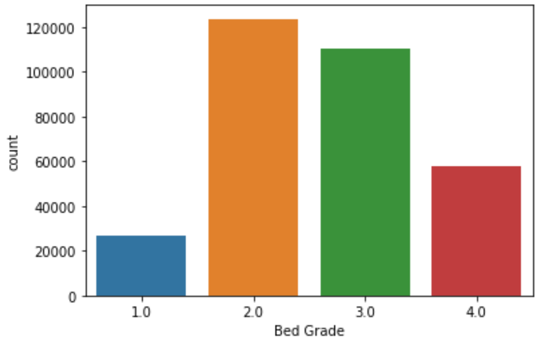


```
sns.barplot(x= "Ward_Facility_Code",y= "Stay", data = train_data_raw1)
```

```
<AxesSubplot:xlabel='Ward_Facility_Code', ylabel='Stay'>
```

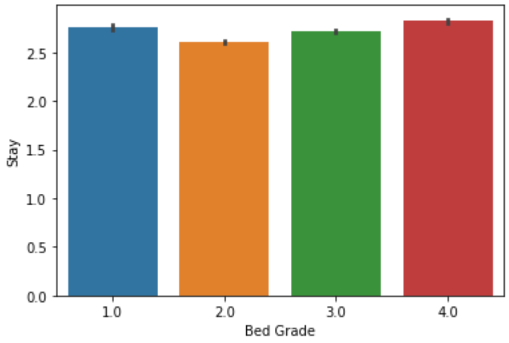
```
sns.countplot(data = train_data_raw1[cat] , x = 'Bed Grade')
```

```
<AxesSubplot:xlabel='Bed Grade', ylabel='count'>
```



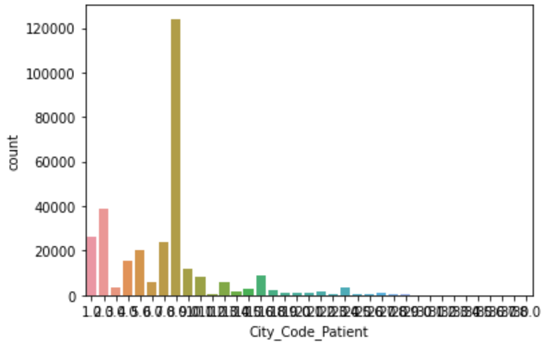
```
sns.barplot(x= "Bed Grade",y= "Stay", data = train_data_raw1)
```

```
<AxesSubplot:xlabel='Bed Grade', ylabel='Stay'>
```



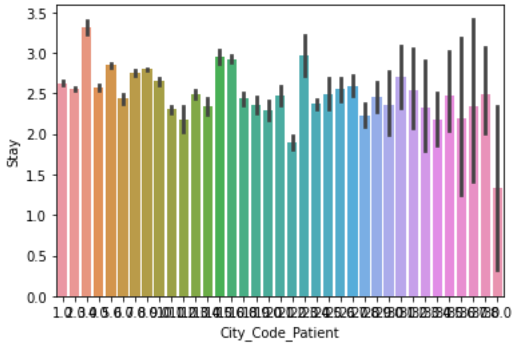
```
sns.countplot(data = train_data_raw1[cat] , x = 'City_Code_Patient')
```

```
<AxesSubplot:xlabel='City_Code_Patient', ylabel='count'>
```



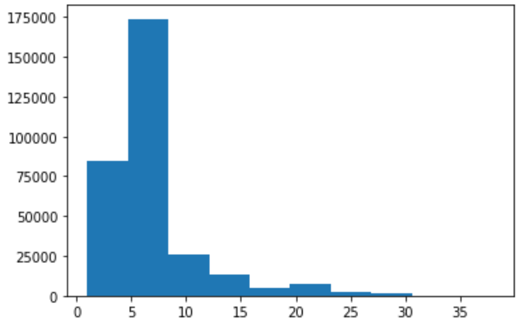
```
sns.barplot(x= "City_Code_Patient",y= "Stay", data = train_data_raw1)
```

```
<AxesSubplot:xlabel='City_Code_Patient', ylabel='Stay'>
```



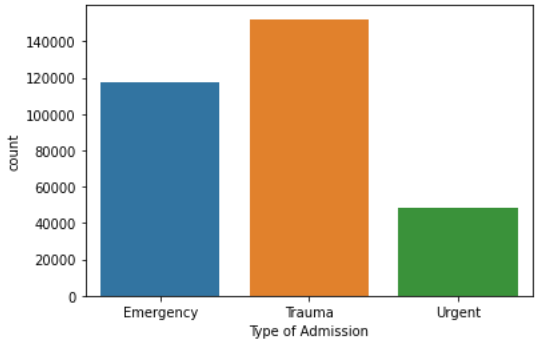
```
plt.hist(train_data_raw1.City_Code_Patient)
```

```
(array([8.43980e+04, 1.73902e+05, 2.62740e+04, 1.35020e+04, 4.68600e+03,
        7.11400e+03, 2.18100e+03, 1.52300e+03, 2.35000e+02, 9.10000e+01]),
 array([ 1. ,  4.7,  8.4, 12.1, 15.8, 19.5, 23.2, 26.9, 30.6, 34.3, 38. ]),
 <BarContainer object of 10 artists>)
```



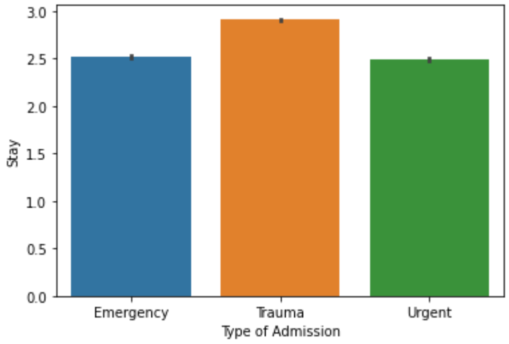
```
sns.countplot(data = train_data_raw1[cat] , x = 'Type of Admission')
```

```
<AxesSubplot:xlabel='Type of Admission', ylabel='count'>
```



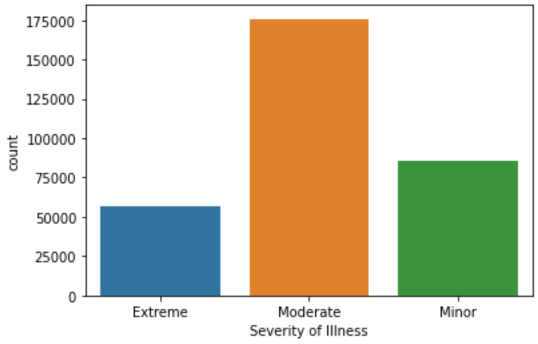
```
sns.barplot(x= "Type of Admission",y= "Stay", data = train_data_raw1)
```

```
<AxesSubplot:xlabel='Type of Admission', ylabel='Stay'>
```



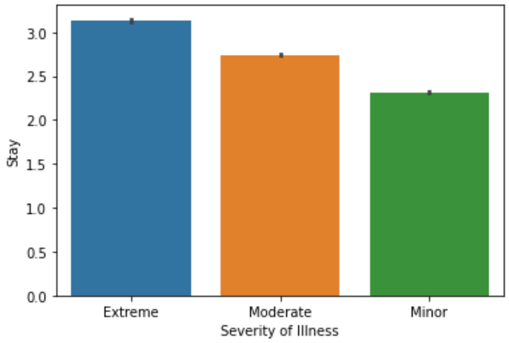
```
sns.countplot(data = train_data_raw1[cat] , x = 'Severity of Illness')
```

```
<AxesSubplot:xlabel='Severity of Illness', ylabel='count'>
```



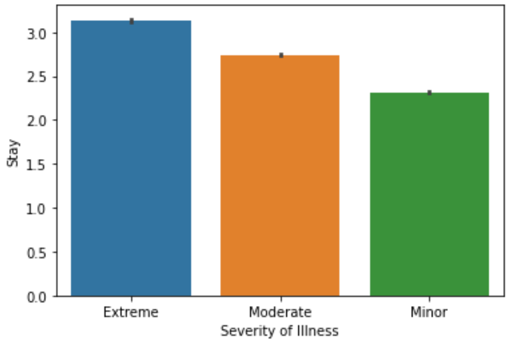
```
sns.barplot(x= "Severity of Illness",y= "Stay", data = train_data_raw1)
```

```
<AxesSubplot:xlabel='Severity of Illness', ylabel='Stay'>
```



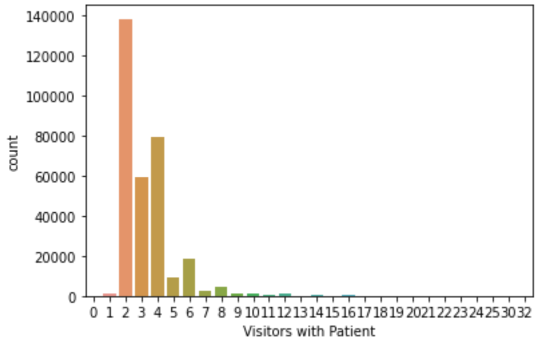
```
sns.barplot(x= "Severity of Illness",y= "Stay", data = train_data_raw1)
```

```
<AxesSubplot:xlabel='Severity of Illness', ylabel='Stay'>
```



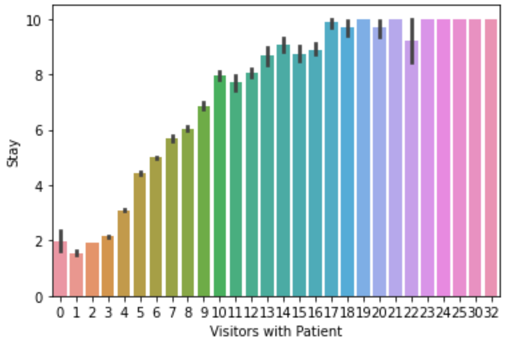
```
sns.countplot(data = train_data_raw1[cat] , x = 'Visitors with Patient')
```

```
<AxesSubplot:xlabel='Visitors with Patient', ylabel='count'>
```



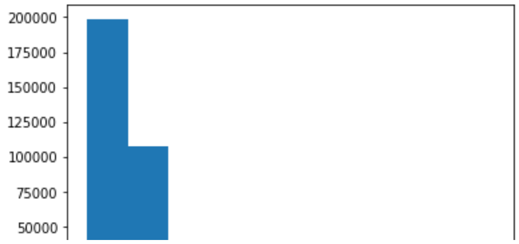
```
sns.barplot(x= "Visitors with Patient",y= "Stay", data = train_data_raw1)
```

```
<AxesSubplot:xlabel='Visitors with Patient', ylabel='Stay'>
```



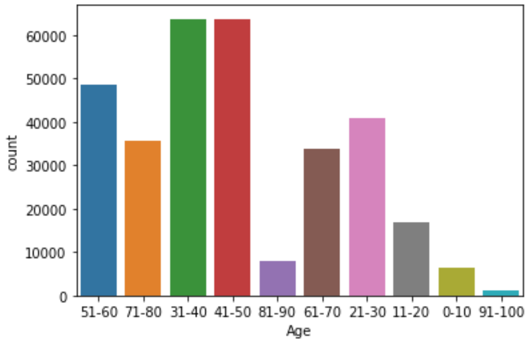
```
plt.hist(train_data_raw1['Visitors with Patient'])
```

```
(array([1.98826e+05, 1.07304e+05, 8.73400e+03, 2.47800e+03, 5.07000e+02,
        3.79000e+02, 9.20000e+01, 1.00000e+02, 0.00000e+00, 1.80000e+01]),
 array([ 0. ,  3.2,  6.4,  9.6, 12.8, 16. , 19.2, 22.4, 25.6, 28.8, 32. ]),
 <BarContainer object of 10 artists>)
```



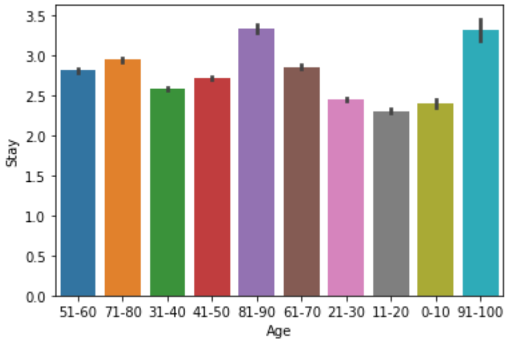
```
sns.countplot(data = train_data_raw1[cat] , x = 'Age')
```

```
<AxesSubplot:xlabel='Age', ylabel='count'>
```



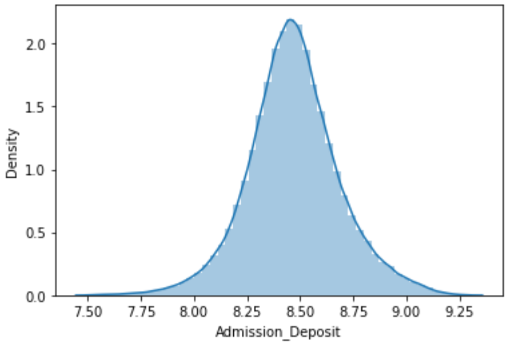
```
sns.barplot(x= "Age",y= "Stay", data = train_data_raw1)
```

```
<AxesSubplot:xlabel='Age', ylabel='Stay'>
```



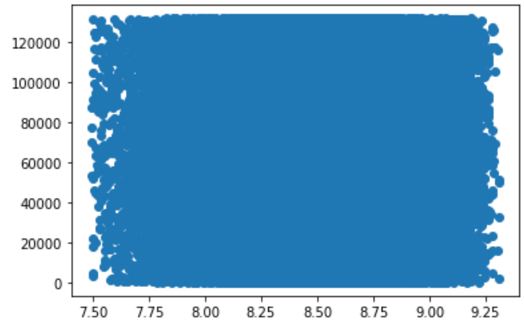
```
#sns.countplot(data = train_data_raw1[cat] , x = 'Admission_Deposit')
sns.distplot(np.log(train_data_raw1.Admission_Deposit))
import scipy
scipy.stats.skew(np.log(train_data_raw1.Admission_Deposit))
scipy.stats.kurtosis(np.log(train_data_raw1.Admission_Deposit))
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a de
  warnings.warn(msg, FutureWarning)
0.8511194351593763
```



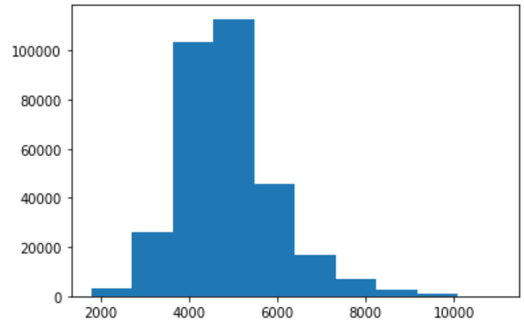
```
plt.scatter(np.log(train_data_raw1.Admission_Deposit),train_data_raw1.patientid)
```

```
<matplotlib.collections.PathCollection at 0x7fcb80166e50>
```



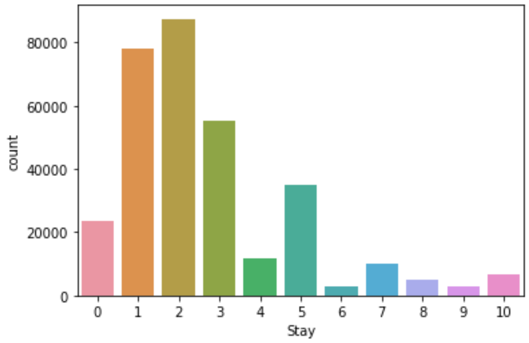
```
plt.hist(train_data_raw1.Admission_Deposit)
```

```
(array([  2975.,  26059., 103190., 112815.,  45913.,  16963.,   6939.,
          2729.,    726.,    129.]),
 array([ 1800. ,  2720.8,  3641.6,  4562.4,  5483.2,  6404. ,  7324.8,
         8245.6,  9166.4, 10087.2, 11008. ]),
 <BarContainer object of 10 artists>)
```



```
sns.countplot(data = train_data_raw1[cat] , x = 'Stay')
```

```
<AxesSubplot:xlabel='Stay', ylabel='count'>
```

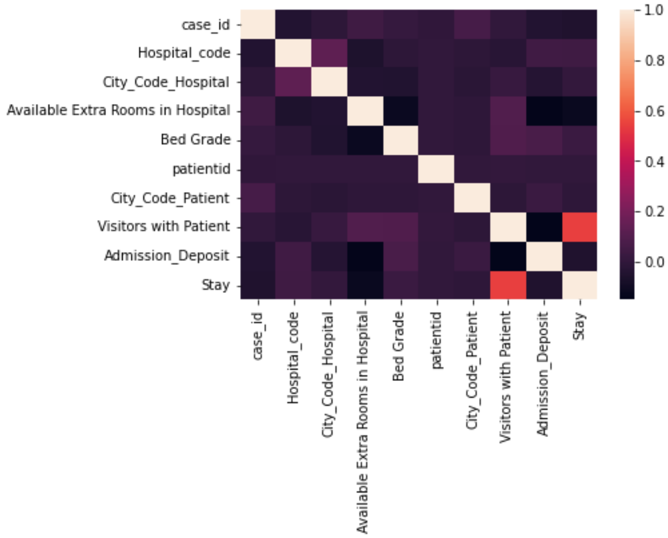


```
train_data_raw1.corr()
```

	case_id	Hospital_code	City_Code_Hospital	Available Extra Rooms in Hospital	Bed Grade	patientid	City_C
case_id	1.000000	-0.043023	-0.011352	0.042580	0.013702	-0.004150	
Hospital_code	-0.043023	1.000000	0.128294	-0.059638	-0.013739	0.002291	
City_Code_Hospital	-0.011352	0.128294	1.000000	-0.045771	-0.049309	0.000750	
Available Extra Rooms in Hospital	0.042580	-0.059638	-0.045771	1.000000	-0.115868	0.000921	

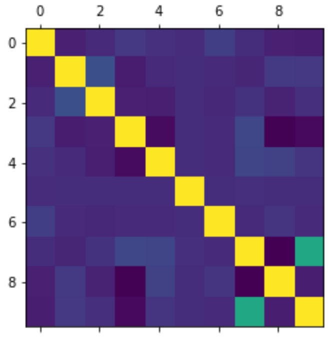
```
sns.heatmap(train_data_raw1.corr())
```

```
<AxesSubplot:>
```



```
plt.matshow(train_data_raw1.corr())
```

```
<matplotlib.image.AxesImage at 0x7fcb7ff28a30>
```



```
from sklearn import preprocessing
col = ['Hospital_code', 'Hospital_type_code', 'City_Code_Hospital',
       'Hospital_region_code', 'Available Extra Rooms in Hospital',
       'Department', 'Ward_Type', 'Ward_Facility_Code', 'Bed Grade',
        'City_Code_Patient', 'Type of Admission',
       'Severity of Illness', 'Visitors with Patient', 'Age',
       ]
lb = preprocessing.LabelEncoder()
for i in col:
    train_data_raw1[i] = lb.fit(pd.concat([train_data_raw1[i],test_raw[i]]))
    test_raw[i] = lb.transform(test_raw[i])
```

```
test_raw.to_csv("Label_Encoded_Test.csv",index=False)
```

```
train_data_raw1
```

	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region_code	Avail Extra R in Hosp
0	1	LabelEncoder()	LabelEncoder()	LabelEncoder()	LabelEncoder()	LabelEncc
1	2	LabelEncoder()	LabelEncoder()	LabelEncoder()	LabelEncoder()	LabelEncc
2	3	LabelEncoder()	LabelEncoder()	LabelEncoder()	LabelEncoder()	LabelEncc
3	4	LabelEncoder()	LabelEncoder()	LabelEncoder()	LabelEncoder()	LabelEncc
4	5	LabelEncoder()	LabelEncoder()	LabelEncoder()	LabelEncoder()	LabelEncc
...	...	...	...	...	...	
318433	318434	LabelEncoder()	LabelEncoder()	LabelEncoder()	LabelEncoder()	LabelEncc
318434	318435	LabelEncoder()	LabelEncoder()	LabelEncoder()	LabelEncoder()	LabelEncc
318435	318436	LabelEncoder()	LabelEncoder()	LabelEncoder()	LabelEncoder()	LabelEncc
318436	318437	LabelEncoder()	LabelEncoder()	LabelEncoder()	LabelEncoder()	LabelEncc
318437	318438	LabelEncoder()	LabelEncoder()	LabelEncoder()	LabelEncoder()	LabelEncc

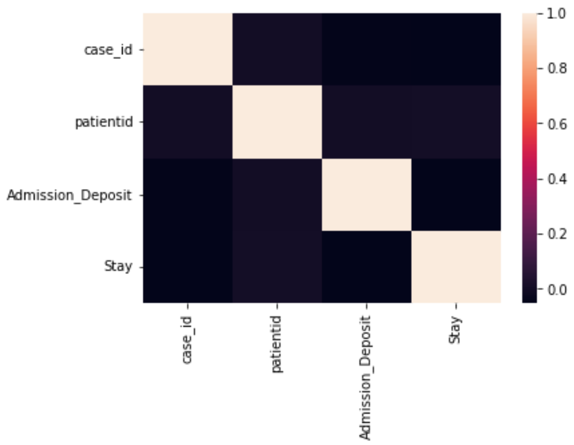
318438 rows × 18 columns

```
#train_data_raw1.to_csv("Label_Encoded_Train.csv")
train_data_raw1.corr()
```

	case_id	patientid	Admission_Deposit	Stay
case_id	1.000000	-0.004150	-0.045972	-0.053882
patientid	-0.004150	1.000000	-0.000877	0.001125
Admission_Deposit	-0.045972	-0.000877	1.000000	-0.052077
Stay	-0.053882	0.001125	-0.052077	1.000000

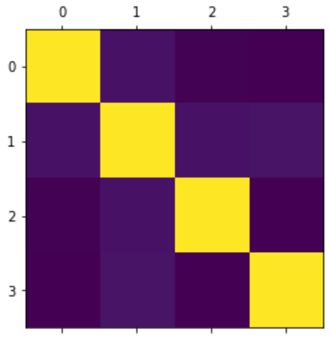
```
sns.heatmap(train_data_raw1.corr())
```

```
<AxesSubplot:>
```



```
plt.matshow(train_data_raw1.corr())
```

```
<matplotlib.image.AxesImage at 0x7fcb7fdc33a0>
```



```
# from sklearn import linear_model
# from xgboost import XGBClassifier
# xg = XGBClassifier()
# lr = linear_model.LinearRegression()
# xg.fit(train_data_raw1[['case_id', 'Hospital_code', 'Hospital_type_code', 'City_Code_Hospital',
#        'Hospital_region_code', 'Available Extra Rooms in Hospital',
#        'Department', 'Ward_Type', 'Ward_Facility_Code', 'Bed Grade',
#        'patientid', 'City_Code_Patient', 'Type of Admission',
#        'Severity of Illness', 'Visitors with Patient', 'Age',
#        'Admission_Deposit']][:200000],train_data_raw1['Stay'][:200000])
# xg.score(train_data_raw1[['case_id', 'Hospital_code', 'Hospital_type_code', 'City_Code_Hospital',
#        'Hospital_region_code', 'Available Extra Rooms in Hospital',
#        'Department', 'Ward_Type', 'Ward_Facility_Code', 'Bed Grade',
#        'patientid', 'City_Code_Patient', 'Type of Admission',
#        'Severity of Illness', 'Visitors with Patient', 'Age',
#        'Admission_Deposit']][:200000],train_data_raw1['Stay'][:200000])
```

```
# xg.score(train_data_raw1[['case_id', 'Hospital_code', 'Hospital_type_code', 'City_Code_Hospital',
#        'Hospital_region_code', 'Available Extra Rooms in Hospital',
#        'Department', 'Ward_Type', 'Ward_Facility_Code', 'Bed Grade',
#        'patientid', 'City_Code_Patient', 'Type of Admission',
#        'Severity of Illness', 'Visitors with Patient', 'Age',
#        'Admission_Deposit']][200000:],train_data_raw1['Stay'][200000:])
```

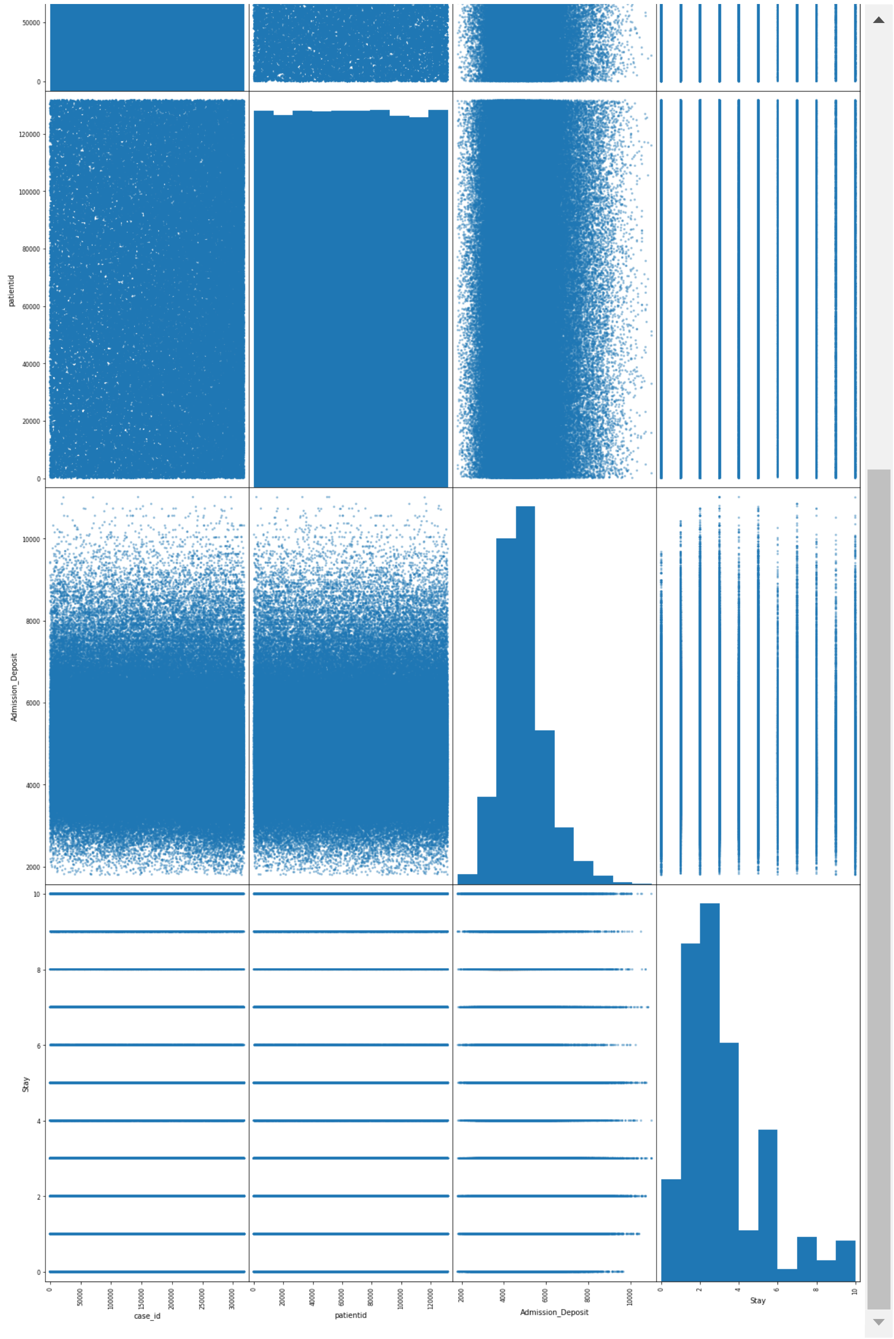
```
train_data_raw1.columns
```

```
    Index(['case_id', 'Hospital_code', 'Hospital_type_code', 'City_Code_Hospital',
           'Hospital_region_code', 'Available Extra Rooms in Hospital',
           'Department', 'Ward_Type', 'Ward_Facility_Code', 'Bed Grade',
           'patientid', 'City_Code_Patient', 'Type of Admission',
           'Severity of Illness', 'Visitors with Patient', 'Age',
           'Admission_Deposit', 'Stay'],
          dtype='object')
```

```
train_data_raw1.City_Code_Patient.unique()
```

```
    array([LabelEncoder()], dtype=object)
```

```
pd.plotting.scatter_matrix(train_data_raw1,figsize=(20,40))
```


50000
0
patientid
120000
100000
80000
60000
40000
20000
0
Admission_Deposit
10000
8000
6000
4000
2000
Stay
10
8
6
4
2
0
case_id
0
50000
100000
150000
200000
250000
300000
patientid
0
20000
40000
60000
80000
100000
120000
Admission_Deposit
2000
4000
6000
8000
10000
Stay
0
2
4
6
8
10