

Exploratory Data Analysis

Team ID : PNT2022TMD26934
Date : 05/11/2022

Project Name : Analytics for Hospital's Health Care Data

Required libraries:

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
In [2]: df = pd.read_csv("../Healthcare_data/xxiao_data.csv")
```

```
In [3]: df
```

case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region_code	Extra Rooms in Hospital	Department	Ward_Type	Ward_Facility_Code	Grade	Bed patientid	City_Code_Patient	Admission	Type of Admission	Severity of Illness	Visitors with Illness	Age	Admission
0	1	8	c	3	Z	3	radiotherapy	R	F	2.0	31397	7.0	Emergency	Extreme	2	51-60	
1	2	2	c	5	Z	2	radiotherapy	S	F	2.0	31397	7.0	Trauma	Extreme	2	51-60	
2	3	10	a	1	X	2	anesthesia	S	E	2.0	31397	7.0	Trauma	Extreme	2	51-60	
3	4	26	b	2	Y	2	radiotherapy	R	D	2.0	31397	7.0	Trauma	Extreme	2	51-60	
4	5	26	b	2	Y	2	radiotherapy	S	D	2.0	31397	7.0	Trauma	Extreme	2	51-60	
...
318433	318434	6	a	6	X	3	radiotherapy	Q	F	4.0	86499	23.0	Emergency	Moderate	3	41-50	
318434	318435	24	a	1	X	2	anesthesia	Q	E	4.0	325	8.0	Urgent	Moderate	4	91-99	
318435	318436	7	a	4	X	3	gynecology	R	F	4.0	125235	10.0	Emergency	Minor	3	71-80	
318436	318437	11	b	2	Y	3	anesthesia	Q	D	3.0	91081	8.0	Trauma	Minor	5	11-20	
318437	318438	19	a	7	Y	5	gynecology	Q	C	2.0	21641	8.0	Emergency	Minor	2	11-20	
318438 rows x 18 columns																	

318438 rows x 18 columns

```
In [4]: df.head()
```

case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region_code	Rooms in Hospital	Department	Ward_Type	Ward_Facility_Code	Grade	Bed patientid	City_Code_Patient	Admission	Type of Admission	Severity of Illness	Visitors with Illness	Age	Admission
0	1	8	c	3	Z	3	radiotherapy	R	F	2.0	31397	7.0	Emergency	Extreme	2	51-60	
1	2	2	c	5	Z	2	radiotherapy	S	F	2.0	31397	7.0	Trauma	Extreme	2	51-60	
2	3	10	a	1	X	2	anesthesia	S	E	2.0	31397	7.0	Trauma	Extreme	2	51-60	
3	4	26	b	2	Y	2	radiotherapy	R	D	2.0	31397	7.0	Trauma	Extreme	2	51-60	
4	5	26	b	2	Y	2	radiotherapy	S	D	2.0	31397	7.0	Trauma	Extreme	2	51-60	

df.tail()

Available Extra	Bed	Type of	Severity	Visitors
--------------------	-----	---------	----------	----------

```
In [5]: df.tail()
```

Hospitals																		
318433	318434	6	a	6	X	3	radiotherapy	Q	F	4.0	86499	23.0	Emergency	Moderate	3	41-50		
318434	318435	24	a	1	X	2	anesthesia	Q	E	4.0	325	8.0	Urgent	Moderate	4	81-90		
318435	318436	7	a	4	X	3	gynecology	R	F	4.0	125235	10.0	Emergency	Minor	3	71-80		
318436	318437	11	b	2	Y	3	anesthesia	Q	D	3.0	91081	8.0	Trauma	Minor	5	11-20		
318437	318438	19	a	7	Y	5	gynecology	Q	C	2.0	21641	8.0	Emergency	Minor	2	11-20		

```
In [6]: df.info()
```

```
>>>columns = pandas.core.frame.infoFrom">
NameIndex: 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      318356 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  object
dtypes: float64(2), int64(6), object(9)
memory usage: 43.7+ MB
```

```
In [7]: df.dtypes
```

```
case_id                int64
hospital_code          int64
hospital_type_code     object
city_code_hospital     int64
hospital_region_code   object
available_extra_rooms_in_hospital  int64
department            object
ward_type             object
ward_facility_code     object
bed_grade             float64
patientid             int64
city_code_patient      float64
type_of_admission      object
severity_of_illness     object
visitors_with_patient   int64
age                   object
admission_deposit      float64
stay                  object
dtypes: object          9
memory usage: 43.7+ MB
```

```
In [8]: df.shape
```

```
Out[8]: (318438, 18)
```

Before Null Values checking :

```
In [22]: df.isnull().sum().sum()
```

```
Out[22]: 4445
```

```
In [23]: df.isnull()
```

	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region_code	Available Extra Rooms in Hospital	Department	Ward_Type	Ward_Facility_Code	Grade	Bed patientid	City_Code_Patient	Admission	Type of Admission	Severity of Illness	Visitors with Illness	Age	Admission
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
...
318433	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
318434	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
318435	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
318436	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
318437	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False

318438 rows x 18 columns

```
In [24]: df.describe()
```

	case_id	Hospital_code	City_Code_Hospital	Available Extra Rooms in Hospital	Bed Grade	patientid	City_Code_Patient	Visitors with Patient	Admission_Deposit
count	318438.000000	318438.000000	318438.000000	318438.000000	318325.000000	318438.000000	313606.000000	318438.000000	318438.000000
mean	150219.500000	18.318941	0.717171	3.197027	2.626307	62747.579472	7.221859	3.286099	4880.749392
std	91025.276847	8.633755	3.102235	1.168171	0.873146	37979.326440	4.742266	1.756061	1086.776254
min	1	1	1	1	1	1	1	1	1
25%	79610.250000	11.000000	2.000000	2.000000	2.000000	32847.000000	4.000000	3.000000	4186.000000
50%	150219.500000	19.000000	5.000000	3.000000	3.000000	62724.500000	8.000000	3.000000	4741.000000
75%	236828.750000	26.000000	7.000000	4.000000	3.000000	98470.000000	8.000000	4.000000	5409.000000
max	318438.000000	32.000000	13.000000	24.000000	4.000000	131624.000000	36.000000	32.000000	11008.000000

```
In [27]: df.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             113
patientid             0
city_code_patient      4532
type_of_admission      0
severity_of_illness     0
visitors_with_patient   0
age                   0
admission_deposit      0
stay                  0
dtypes: int64          1
```

```
In [31]: df.corr()
```

	case_id	Hospital_code	City_Code_Hospital	Available Extra Rooms in Hospital	Bed Grade	patientid	City_Code_Patient	Visitors with Patient	Admission_Deposit
case_id	1.000000	-0.043023	-0.011352	0.042580	0.013702	-0.004150	0.065196	0.001309	-0.045972
Hospital_code	-0.043023	1.000000	0.128294	-0.059638	-0.013739	0.002291	-0.015530	-0.028500	0.045446
City_Code_Hospital	-0.011352	0.128294	1.000000	-0.045771	-0.046309	0.000705	-0.023988	0.018184	-0.034455
Available Extra Rooms in Hospital	0.042580	-0.059638	-0.045771	1.000000	-0.110868	0.000621	-0.009681	0.006714	-0.143739
Bed Grade	0.013702	-0.013739	-0.046309	-0.110868	1.000000	0.001645	-0.008105	0.008945	0.073833
patientid	-0.004150	0.002291	0.000705	0.000621	0.001645	1.000000	0.002002	0.006869	-0.000877
City_Code_Patient	0.065196	-0.015530	-0.023988	-0.009681	-0.008105	0.002002	1.000000	-0.012074	0.025837
Visitors with Patient	0.001309	-0.028500	0.018184	0.006714	0.008945	0.000889	-0.012074	1.000000	-0.150358
Admission_Deposit	-0.045972	0.045446	-0.034455	-0.143739	0.073833	-0.000877	0.025837	-0.150358	1.000000

```
In [28]: df.isnull().sum().sum()
```

```
Out[28]: 4445
```

Work With Null Values :

```
In [32]: df["bed_grade"].fillna(df["bed_grade"].mean(), inplace=True)
```

```
In [33]: df["bed_grade"].isnull().sum()
```

```
Out[33]: 0
```

```
In [34]: df.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             0
patientid             0
city_code_patient      4532
type_of_admission      0
severity_of_illness     0
visitors_with_patient   0
age                   0
admission_deposit      0
stay                  0
dtypes: int64          1
```

```
In [35]: df["city_code_patient"].fillna(df["city_code_patient"].mean(), inplace=True)
```

```
In [36]: df["city_code_patient"].isnull().sum()
```

```
Out[36]: 0
```

After Cleaning Process :

Total Null Values Checking :

```
In [37]: df.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             0
patientid             0
city_code_patient      4532
type_of_admission      0
severity_of_illness     0
visitors_with_patient   0
age                   0
admission_deposit      0
stay                  0
dtypes: int64          1
```

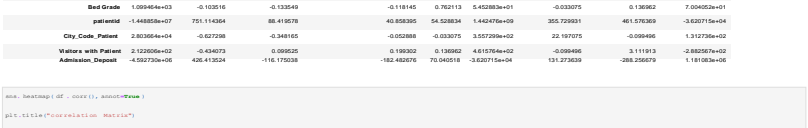
```
In [38]: df.isnull().sum().sum()
```

```
Out[38]: 0
```

```
In [39]: df.corr()
```

	case_id	Hospital_code	City_Code_Hospital	Available Extra Rooms in Hospital	Bed Grade	patientid	City_Code_Patient	Visitors with Patient	Admission_Deposit
case_id	8.450257e+09	-34145.255936	-3237.513037	4572.484177	1089.464209	-1.448858e+07	28036.639476	212.260614	-4.592730e+06
Hospital_code	-3.414526e+04	74.541723	3.436041	-0.056887	-0.103549	7.511146e+02	-0.622708	-0.434073	4.266135e+02
City_Code_Hospital	-3.237513e+03	3.436041	8.625726	-0.160687	-0.133549	8.841059e+01	-0.348165	0.099526	-1.161750e+02
Available Extra Rooms in Hospital	4.572484e+03	-0.001495	-0.160687	1.364624	-0.118145	4.089529e+01	-0.052888	0.193002	-1.824802e+02
Bed Grade	1.089486e+03	-0.102516	-0.133549	-0.118145	1.762113	5.420823e+01	-0.032075	0.136902	7.004520e+01
patientid	-1.448858e+07	751.114364	88.419578	40.858395	54.328834	1.442476e+09	355.729931	461.576369	-3.820715e+04
City_Code_Patient	2.803664e+04	-0.622708	-0.434073	-0.348165	-0.052888	-0.033075	3.557299e+02	22.197075	-0.099496
Visitors with Patient	2.122006e+02	-0.434073	0.099526	0.193002	0.136902	4.615764e+02	-0.099496	3.111913	-2.882576e+02
Admission_Deposit	-4.592730e+06	426.413224	-116.175038	-182.482076	70.040518	-3.820715e+04	131.273639	-285.256679	1.181083e+06

```
In [40]: sns.heatmap(df.corr(), annot=True)
```



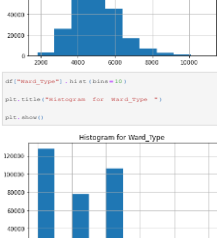
```
In [41]: df["Admission_Deposit"].hist(bins=10)
```

```
plt.title("Histogram for Admission_Deposit ")
plt.show()
```



```
In [42]: df["ward_type"].hist(bins=10)
```

```
plt.title("Histogram for Ward_type ")
plt.show()
```



```
In [43]: df["patientid"].hist(bins=100)
```

```
plt.title("Histogram for patientid ")
plt.show()
```

