TEAM ID : PNT2022TMID07242

Exploratory Data Analysis:

Required libraries:

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

```python
df= pd.read_csv("C:/Users/spara/OneDrive/Desktop/Healthcare_Data/train_data.csv")
```

```python
df
```

	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	A
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	E	2.0	31397	7.0	Trauma	Extreme	2	51-60	
2	3	10	e	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	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	

318438 rows × 18 columns

```python
df.head()
```

	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	Admissi
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	E	2.0	31397	7.0	Trauma	Extreme	2	51-60	
2	3	10	e	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	

```python
df.tail()
```

	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	Admissi
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	

```python
df.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  object
dtypes: float64(3), int64(6), object(9)
memory usage: 43.7+ MB
```

```python
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
dtype: object
```

```python
df.shape
```

```
(318438, 18)
```

Before Null Values checking :

```python
df.isnull().sum().sum()
```

```
4645
```

```python
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	Bed Grade	patientid	City_Code_Patient	Type of Admission	Severity of Illness	Visitors with Patient	Age	A
0	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	
2	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	
4	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	
318434	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	
318436	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	

318438 rows × 18 columns

```python
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	313906.000000	318438.000000	318438.000000
mean	159219.500000	18.318841	4.771717	3.197627	2.625807	65747.379472	7.251869	3.284099	4880.745092
std	91925.276967	8.633756	3.102530	1.168171	0.873146	37915.936441	4.745384	1.764561	1086.776254
min	1.000000	1.000000	1.000000	0.000000	1.000000	1.000000	1.000000	0.000000	1800.000000
25%	79610.250000	11.000000	2.000000	2.000000	2.000000	32847.000000	4.000000	2.000000	4186.000000
50%	159219.500000	19.000000	5.000000	3.000000	3.000000	65724.000000	8.000000	3.000000	4741.000000
75%	238828.750000	26.000000	7.000000	4.000000	3.000000	98470.000000	8.000000	4.000000	5409.000000
max	318438.000000	32.000000	15.000000	24.000000	4.000000	131624.000000	38.000000	32.000000	11008.000000

```python
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
dtype: int64
```

```python
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	-0.011362		0.042580	0.013702	-0.004100	0.092196	0.001309	-0.045972
Hospital_code	-0.043023	1.000000	0.136204		-0.059938	-0.013739	0.002201	-0.015330	0.003655	-0.046446
City_Code_Hospital	-0.011362	0.136204	1.000000		-0.045771	-0.040009	0.000527	-0.023986	0.019194	-0.149759
Available Extra Rooms in Hospital	0.042580	-0.059938	-0.045771		1.000000	-0.116968	0.003521	-0.009881	0.086714	-0.143759
Bed Grade	0.013702	-0.013739	-0.040009		-0.116968	1.000000	0.001645	0.033250	-0.009295	0.075653
patientid	-0.004100	0.002201	0.000527		0.003521	0.001645	1.000000	0.003202	0.000088	-0.000877
City_Code_Patient	0.092196	-0.015330	-0.023986		-0.009881	0.033250	0.003202	1.000000	-0.012574	-0.026937
Visitors with Patient	0.001309	0.003655	0.019194		0.086714	-0.009295	0.000088	-0.012574	1.000000	-0.100368
Admission_Deposit	-0.045972	-0.046446	-0.149759		-0.143759	0.075653	-0.000877	-0.026937	-0.100368	1.000000

```python
df.isnull().sum().sum()
```

```
4645
```

Work With Null Values :

```python
df["Bed Grade"].fillna(df["Bed Grade"].mean(),inplace=True)
```

```python
df["Bed Grade"].isnull().sum()
```

```
0
```

```python
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
dtype: int64
```

```python
df["City_Code_Patient"].fillna(df["City_Code_Patient"].mean(),inplace=True)
```

```python
df["City_Code_Patient"].isnull().sum()
```

```
0
```

After Cleaning Process :

Total Null Values Checking :

```python
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                       0
Type of Admission                       0
Severity of Illness                     0
Visitors with Patient                   0
Age                                     0
Admission_Deposit                       0
Stay                                    0
dtype: int64
```

Total Null Values :

```python
df.isnull().sum().sum()
```

```
0
```

```python
df.cov()
```

	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	-34145.255539	-3237.613037	4572.484177	1095.464209	-1.446869e+07	26036.626476	212.260614	-4.592739e+09
Hospital_code	-3.414526e+04	74.541723	3.438341	-0.599241	-0.601455	-0.103316	7.011144e+02	-0.627268	-4.294125e+02
City_Code_Hospital	-3.237613e+03	3.438341	9.625726	-0.165887	-0.110464	6.188106e+01	-0.350857	0.106154	-1.161756e+03
Available Extra Rooms in Hospital	4.572484e+03	-0.599241	-0.165887	1.364624	-0.119146	1.560500e+02	-0.082689	0.180302	-1.824674e+02
Bed Grade	1.095464e+03	-0.601455	-0.110464	-0.119146	0.762113	5.425899e+01	0.112951	-0.014314	7.004453e+01
patientid	-1.446869e+07	751.114364	86.419578	40.690085	54.528634	1.442676e+09	550.723091	461.007639	-3.620771e+04
City_Code_Patient	2.603664e+04	-0.627268	-0.346160	-0.082689	0.112951	5.507209e+02	22.107075	-0.099496	-1.312739e+02
Visitors with Patient	2.123606e+02	-0.458273	0.439427	0.089852	-0.014314	4.610764e+02	-0.099496	3.114097	-2.882067e+02
Admission_Deposit	-4.592739e+09	-429.412524	-119.175638	-182.467476	70.044532	-3.620771e+04	-131.273938	-288.206678	1.181363e+06

```python
sns.heatmap(df.corr(),annot=True)
plt.title("correlation Matrix")
plt.show()
```


correlation Matrix

```python
df["Admission_Deposit"].hist(bins=5)
plt.title("Histogram for Admission_Deposit ")
plt.show()
```


Histogram for Admission_Deposit

```python
df["Ward_Type"].hist(bins=5)
plt.title("Histogram for Ward_Type ")
plt.show()
```


Histogram for Ward_Type

```python
df["patientid"].hist(bins=5)
plt.title("Histogram for patientid ")
plt.show()
```


Histogram for patientid