

Exploratory Data Analysis:

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("C:/Users/sgraw/Desktop/Healthcare_Data/train_data.csv")

In [3]: df
```

Out[3]:

| | 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 | F | 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 | 125335 | 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 | 20 | |

318438 rows x 18 columns

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

Out[4]:
```

| | 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 | Admisal |
|--|---------|---------------|--------------------|--------------------|----------------------|---|----------------|-----------|--------------------|--------------|-----------|-------------------|----------------------|---------------------------|-----------------------------|-------|---------|
| | 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 | e | 1 | X | 2 anesthesia | S | E | 2.0 | 31397 | 7.0 | Trauma | Extreme | 2 | 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 | |

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

Out[5]:
```

| | 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 |
|--|---------|---------------|--------------------|--------------------|----------------------|---|----------------|-----------|--------------------|--------------|-----------|-------------------|----------------------|---------------------------|-----------------------------|-------|---|
| | 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 | 125335 | 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()

<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              318438 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
```

```
In [7]: df.dtypes

Out[7]:
```

| | |
|-----------------------------------|---------|
| 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 |

```
In [8]: df.shape

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

Before Null Values checking :

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

Out[22]: 4645

In [23]: df.isnull()

Out[23]:
```

| | 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 x 18 columns

```
In [26]: df.describe()

Out[26]:
```

| | 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 | 156219.500000 | 18.318481 | 4.777177 | 3.197627 | 2.825807 | 65747.579472 | 7.251859 | 3.284099 | 4880.748392 |
| std | 91925.276847 | 8.633755 | 3.102536 | 1.168171 | 0.873146 | 37979.936440 | 4.745266 | 1.764061 | 1086.776254 |
| min | 1.000000 | 1.000000 | 1.000000 | 0.000000 | 1.000000 | 1.000000 | 1.000000 | 0.000000 | 1086.000000 |
| 25% | 79910.250000 | 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 | 65724.500000 | 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 | 13.000000 | 24.000000 | 4.000000 | 131624.000000 | 38.000000 | 32.000000 | 11008.000000 |

```
In [27]: df.isnull().sum()

Out[27]:
```

| | |
|-----------------------------------|-------|
| 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 |

```
In [32]: df.corr()

Out[32]:
```

| | 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 | 0.065198 | 0.001309 |
| | Hospital_code | -0.043023 | 1.000000 | 0.128294 | -0.059638 | -0.013739 | 0.002291 | -0.015530 | -0.028500 |
| | City_Code_Hospital | -0.011352 | 0.128294 | 1.000000 | -0.045771 | -0.040309 | 0.000750 | -0.023988 | 0.018184 |
| | Available Extra Rooms in Hospital | 0.042580 | -0.059638 | -0.045771 | 1.000000 | -0.115868 | 0.000921 | -0.009681 | 0.096714 |
| | Bed Grade | 0.013702 | -0.013739 | -0.043039 | -0.115868 | 1.000000 | 0.001645 | -0.008105 | 0.088945 |
| | patientid | -0.004150 | 0.002291 | 0.000750 | 0.000921 | 0.001645 | 1.000000 | 0.002002 | 0.006889 |
| | City_Code_Patient | 0.065198 | -0.015530 | -0.023988 | -0.009681 | -0.008105 | 0.002002 | 1.000000 | -0.012074 |
| | Visitors with Patient | 0.001309 | -0.028500 | 0.018184 | 0.096714 | 0.088945 | 0.006889 | -0.012074 | 1.000000 |
| | Admission_Deposit | -0.045972 | 0.045446 | -0.034455 | -0.143739 | 0.075833 | -0.000877 | 0.025837 | -0.150358 |

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

Out[28]: 4645
```

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()

Out[34]:
```

| | |
|-----------------------------------|-------|
| 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 |

```
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()

Out[37]:
```

| | |
|-----------------------------------|-------|
| 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 :

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

Out[38]: 0

In [39]: df.cov()

Out[39]:
```

| | 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 | 8.460252e+00 | -3.614525e+00 | -3.237513e+01 | 4.572484e+07 | 1099.466209 | -1.448859e+07 | 28036.639476 | -4.592730e+08 |
| | Hospital_code | -3.614526e+04 | 74.541723 | 3.436541 | -0.601495 | -0.103516 | 7.511446e+02 | -0.627298 | -0.434073 |
| | City_Code_Hospital | -3.237513e+03 | 3.436541 | 9.625726 | -0.165887 | -0.133549 | 8.841959e+01 | -0.348165 | 0.099525 |
| | Available Extra Rooms in Hospital | 4.572484e+03 | -0.601495 | -0.165887 | 1.364624 | -0.118145 | 4.065839e+01 | -0.052888 | 0.199302 |
| | Bed Grade | 1.099464e+03 | -0.103516 | -0.133549 | -0.118145 | 0.762113 | 5.462883e+01 | -0.033075 | 0.136962 |
| | patientid | -1.448858e+07 | 7.5114364 | 88.419578 | 40.858395 | 54.528834 | 1.462476e+09 | 355.72931 | 461.576389 |
| | City_Code_Patient | 2.803684e+04 | -0.627298 | -0.348165 | -0.052888 | -0.033075 | 3.557299e+02 | 22.197075 | -0.099496 |
| | Visitors with Patient | 2.122606e+02 | -0.434073 | 0.099525 | 0.199302 | 0.136962 | 4.815764e+02 | -0.099496 | 3.111913 |
| | Admission_Deposit | -4.592730e+08 | 4.26.413524 | -116.175038 | -182.482676 | 70.040518 | -3.620715e+04 | 131.273639 | -288.256879 |

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

plt.title("Correlation Matrix")

plt.show()
```



```
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()
```

