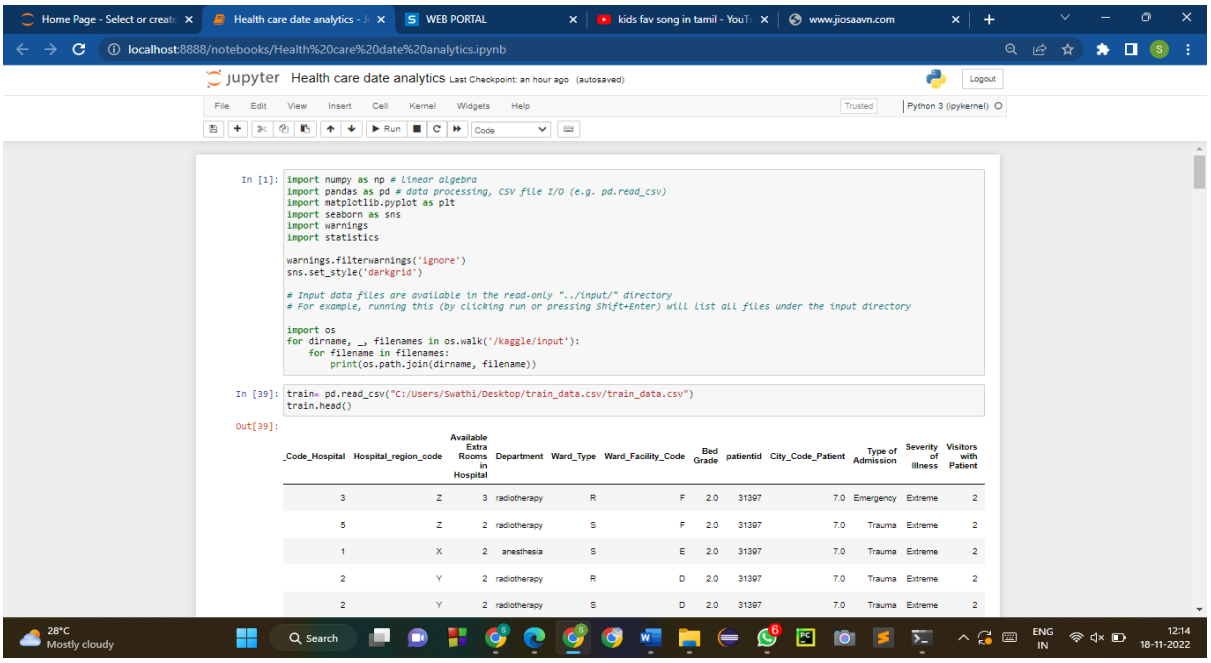
ANALYTICS FOR HOSPITALAND HEALTHCARE DATA

TEAM ID : **PNT2022TMID16345**

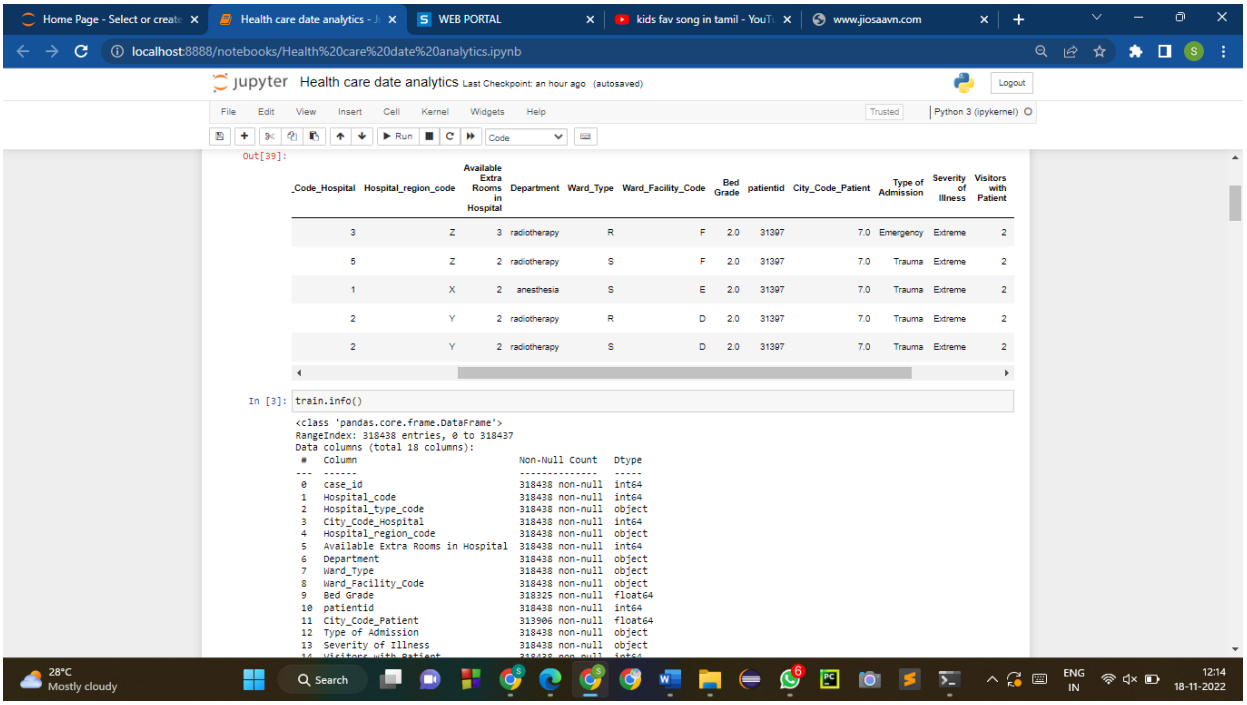
IMPORT LIBRARIES :



```
In [1]: import numpy as np # linear algebra
        import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
        import matplotlib.pyplot as plt
        import seaborn as sns
        import warnings
        import statistics

        warnings.filterwarnings('ignore')
        sns.set_style('darkgrid')

        # Input data files are available in the read-only "../input/" directory
        # For example, running this (by clicking run or pressing Shift+Enter) will list all files under the input directory

        import os
        for dirname, _, filenames in os.walk('/kaggle/input'):
            for filename in filenames:
                print(os.path.join(dirname, filename))

In [39]: train= pd.read_csv("C:/Users/Swathi/Desktop/train_data.csv/train_data.csv")
         train.head()
```

Out[39]:

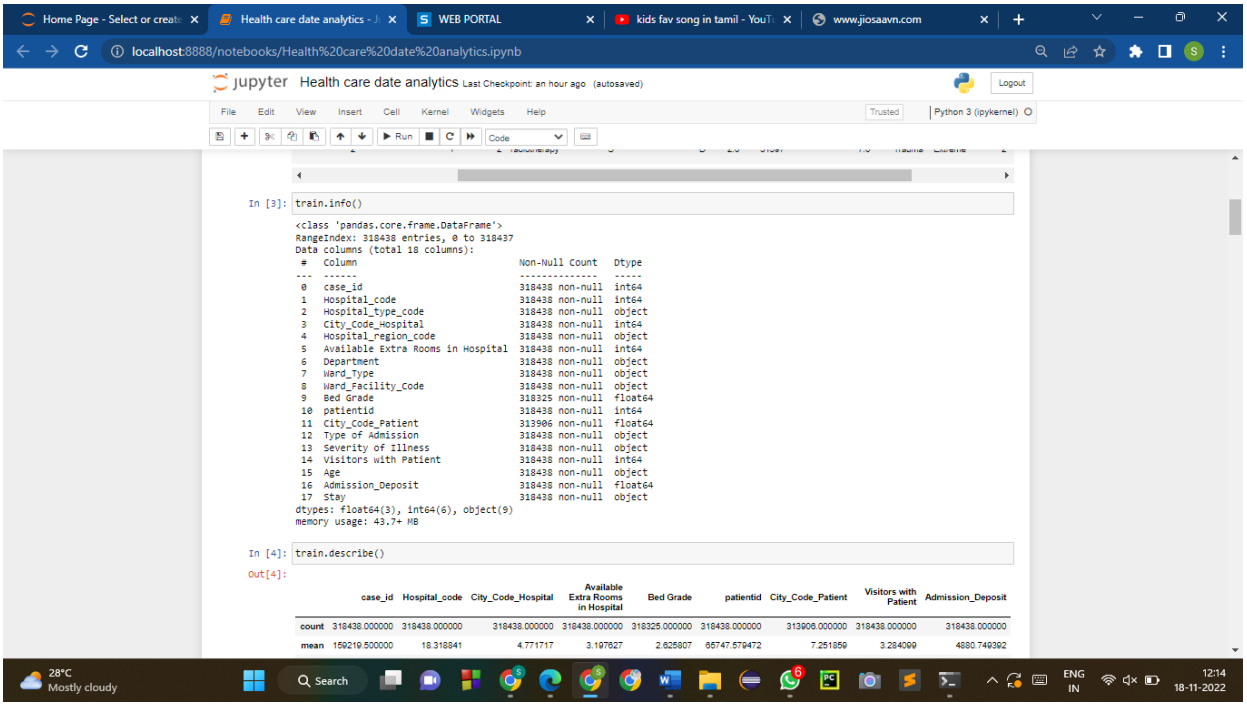
_Code_Hospital	Hospital_region_code	Available Extra Rooms in Hospital	Department	Ward_Type	Ward_Facility_Code	Bed Grade	patientid	City_Code_Patient	Type of Admission	Severity of Illness	Visitors with Patient
3	Z	3	radiotherapy	R	F	2.0	31397	7.0	Emergency	Extreme	2
5	Z	2	radiotherapy	S	F	2.0	31397	7.0	Trauma	Extreme	2
1	X	2	anesthesia	S	E	2.0	31397	7.0	Trauma	Extreme	2
2	Y	2	radiotherapy	R	D	2.0	31397	7.0	Trauma	Extreme	2
2	Y	2	radiotherapy	S	D	2.0	31397	7.0	Trauma	Extreme	2

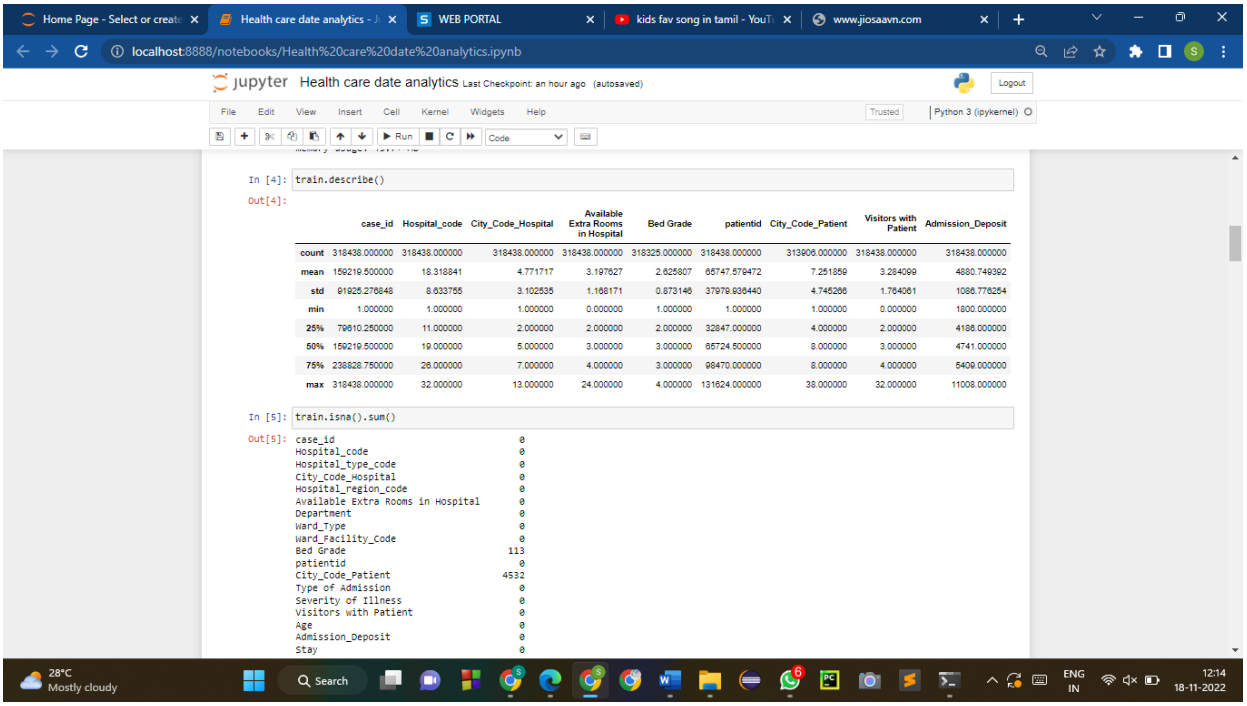


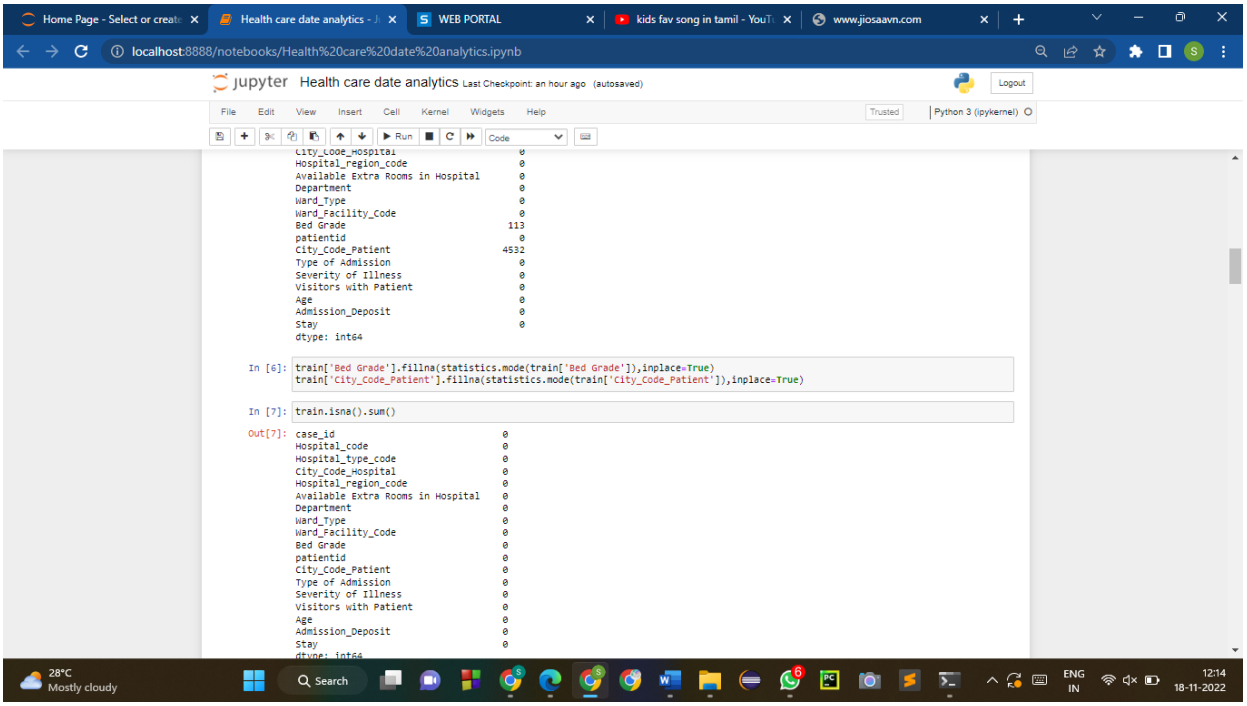
Out[39]:

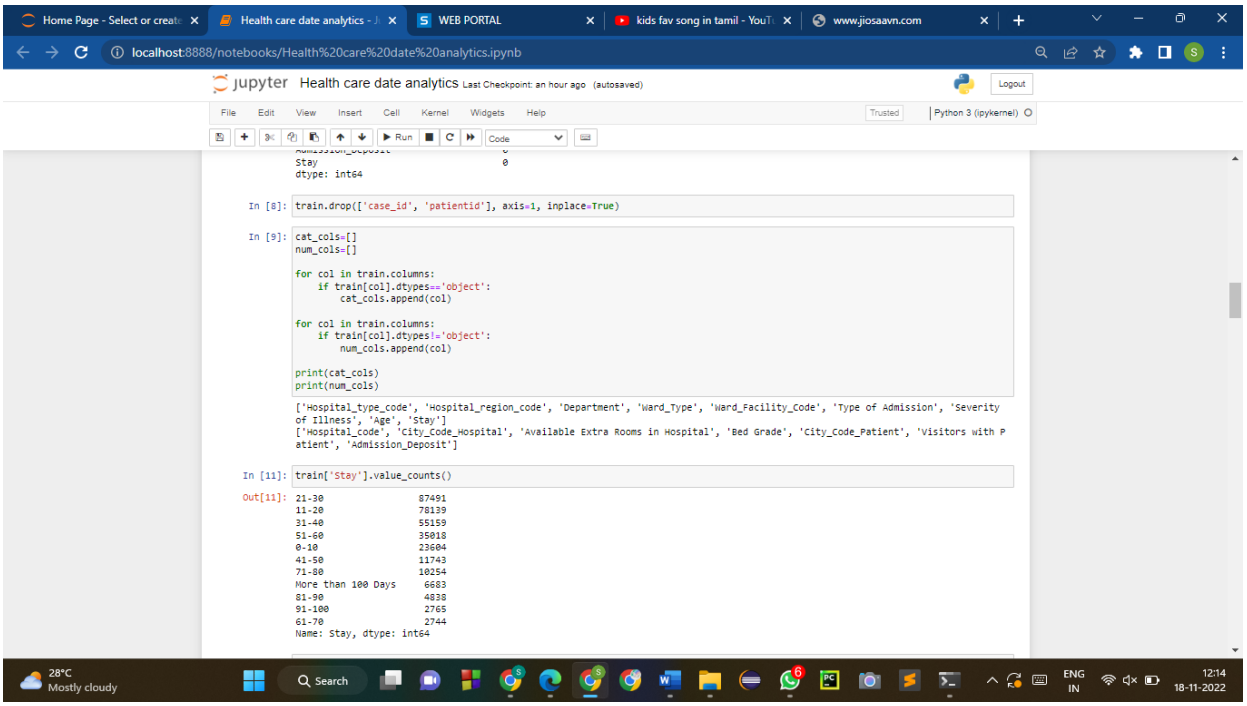
_Code_Hospital	Hospital_region_code	Available Extra Rooms in Hospital	Department	Ward_Type	Ward_Facility_Code	Bed Grade	patientid	City_Code_Patient	Type of Admission	Severity of Illness	Visitors with Patient
3	Z	3	radiotherapy	R	F	2.0	31397	7.0	Emergency	Extreme	2
5	Z	2	radiotherapy	S	F	2.0	31397	7.0	Trauma	Extreme	2
1	X	2	anesthesia	S	E	2.0	31397	7.0	Trauma	Extreme	2
2	Y	2	radiotherapy	R	D	2.0	31397	7.0	Trauma	Extreme	2
2	Y	2	radiotherapy	S	D	2.0	31397	7.0	Trauma	Extreme	2

```
In [3]: train.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
```


Home Page - Select or create
Health care date analytics - J
WEB PORTAL
kids fav song in tamil - YouT
www.jiosaavn.com
localhost:8888/notebooks/Health%20care%20date%20analytics.ipynb
jupyter Health care date analytics Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [3]: train.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
In [4]: train.describe()
Out[4]:
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 2.625907 65747.579472 7.251859 3.284099 4880.749392
28°C Mostly cloudy
Search
ENG IN
12:14 18-11-2022


Home Page - Select or create
Health care date analytics - J
WEB PORTAL
kids fav song in tamil - YouT
www.jiosaavn.com
localhost:8888/notebooks/Health%20care%20date%20analytics.ipynb
jupyter Health care date analytics Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
In [4]: train.describe()
Out[4]:
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 2.625907 65747.579472 7.251859 3.284099 4880.749392
std 91925.276848 8.633755 3.102535 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 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 65724.500000 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 [5]: train.isna().sum()
Out[5]:
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
28°C Mostly cloudy
Search
ENG IN
12:14 18-11-2022


Home Page - Select or create
Health care date analytics - J
WEB PORTAL
kids fav song in tamil - YouT
www.jiosaavn.com
localhost:8888/notebooks/Health%20care%20date%20analytics.ipynb
jupyter Health care date analytics Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
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 [6]:
train['Bed Grade'].fillna(statistics.mode(train['Bed Grade']),inplace=True)
train['City_Code_Patient'].fillna(statistics.mode(train['City_Code_Patient']),inplace=True)
In [7]:
train.isna().sum()
Out[7]:
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
28°C
Mostly cloudy
Search
ENG
IN
12:14
18-11-2022


Home Page - Select or create
Health care date analytics - J
WEB PORTAL
kids fav song in tamil - YouT
www.jiosaavn.com
localhost:8888/notebooks/Health%20care%20date%20analytics.ipynb
jupyter Health care date analytics Last Checkpoint: an hour ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
Stay 0
dtype: int64
In [8]:
train.drop(['case_id', 'patientid'], axis=1, inplace=True)
In [9]:
cat_cols=[]
num_cols=[]
for col in train.columns:
if train[col].dtypes=='object':
cat_cols.append(col)
for col in train.columns:
if train[col].dtypes!='object':
num_cols.append(col)
print(cat_cols)
print(num_cols)
['Hospital_type_code', 'Hospital_region_code', 'Department', 'Ward_Type', 'Ward_Facility_Code', 'Type of Admission', 'Severity of Illness', 'Age', 'Stay']
['Hospital_code', 'City_Code_Hospital', 'Available Extra Rooms in Hospital', 'Bed Grade', 'City_Code_Patient', 'Visitors with Patient', 'Admission_Deposit']
In [11]:
train['Stay'].value_counts()
Out[11]:
21-30 87491
11-20 78139
31-40 55159
51-60 35018
0-10 23604
41-50 11743
71-80 10254
More than 100 Days 6683
81-90 4838
91-100 2765
61-70 2744
Name: Stay, dtype: int64
28°C
Mostly cloudy
Search
ENG
IN
12:14
18-11-2022