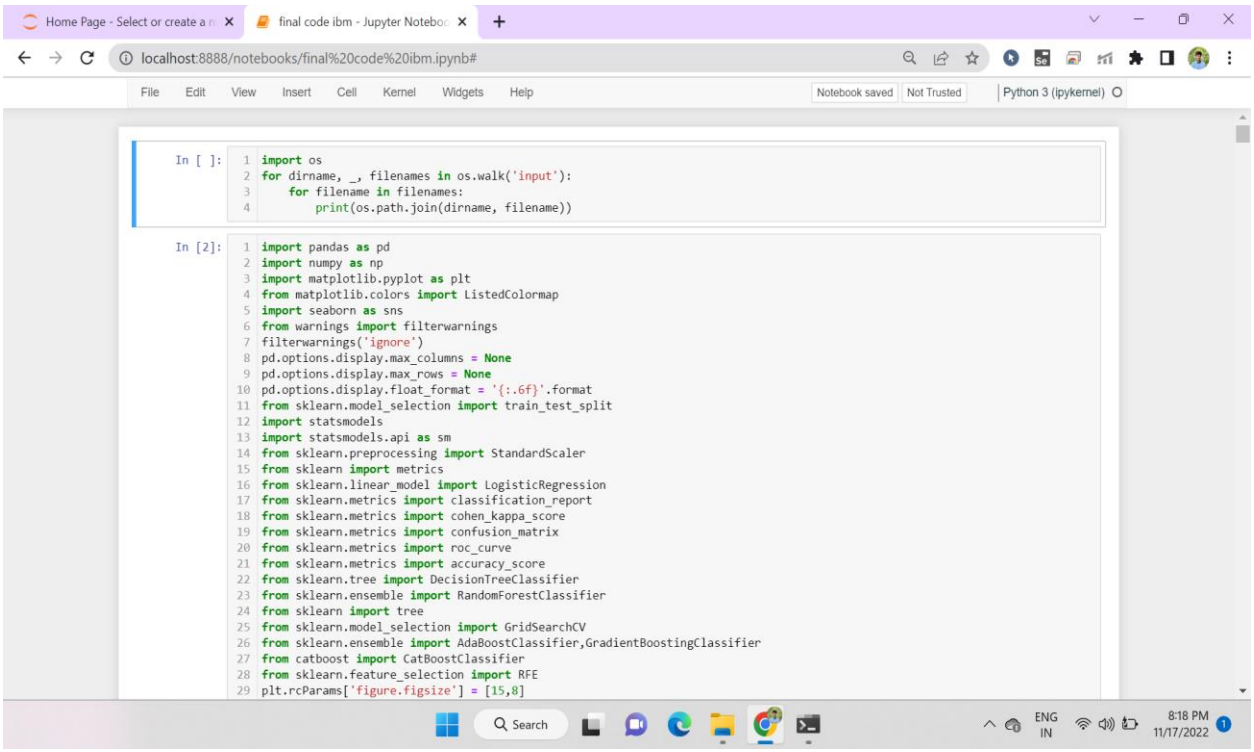
Analytics for Hospitals Health-Care Data

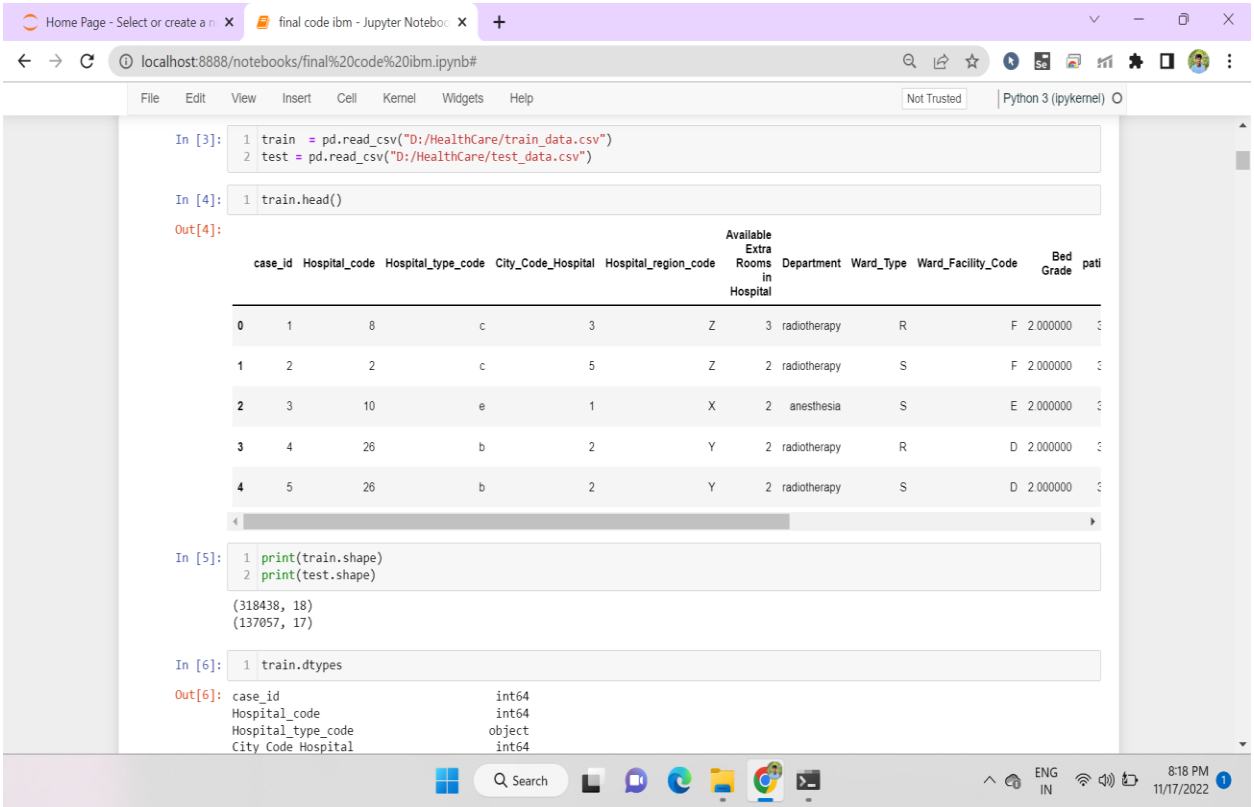
TEAM ID: PNT2022TMID16292

IMPORT LIBRARIES



```
In [ ]:
import os
for dirname, _, filenames in os.walk('input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

```
In [2]:
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.colors import ListedColormap
import seaborn as sns
from warnings import filterwarnings
filterwarnings('ignore')
pd.options.display.max_columns = None
pd.options.display.max_rows = None
pd.options.display.float_format = '{:.6f}'.format
from sklearn.model_selection import train_test_split
import statsmodels
import statsmodels.api as sm
from sklearn.preprocessing import StandardScaler
from sklearn import metrics
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report
from sklearn.metrics import cohen_kappa_score
from sklearn.metrics import confusion_matrix
from sklearn.metrics import roc_curve
from sklearn.metrics import accuracy_score
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn import tree
from sklearn.model_selection import GridSearchCV
from sklearn.ensemble import AdaBoostClassifier,GradientBoostingClassifier
from catboost import CatBoostClassifier
from sklearn.feature_selection import RFE
plt.rcParams['figure.figsize'] = [15,8]
```



```
In [3]:
train  = pd.read_csv("D:/HealthCare/train_data.csv")
test = pd.read_csv("D:/HealthCare/test_data.csv")
```

```
In [4]:
train.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	pati
0	1	8	c	3	Z	3	radiotherapy	R	F	2.000000	
1	2	2	c	5	Z	2	radiotherapy	S	F	2.000000	
2	3	10	e	1	X	2	anesthesia	S	E	2.000000	
3	4	26	b	2	Y	2	radiotherapy	R	D	2.000000	
4	5	26	b	2	Y	2	radiotherapy	S	D	2.000000	

```
In [5]:
print(train.shape)
print(test.shape)
```

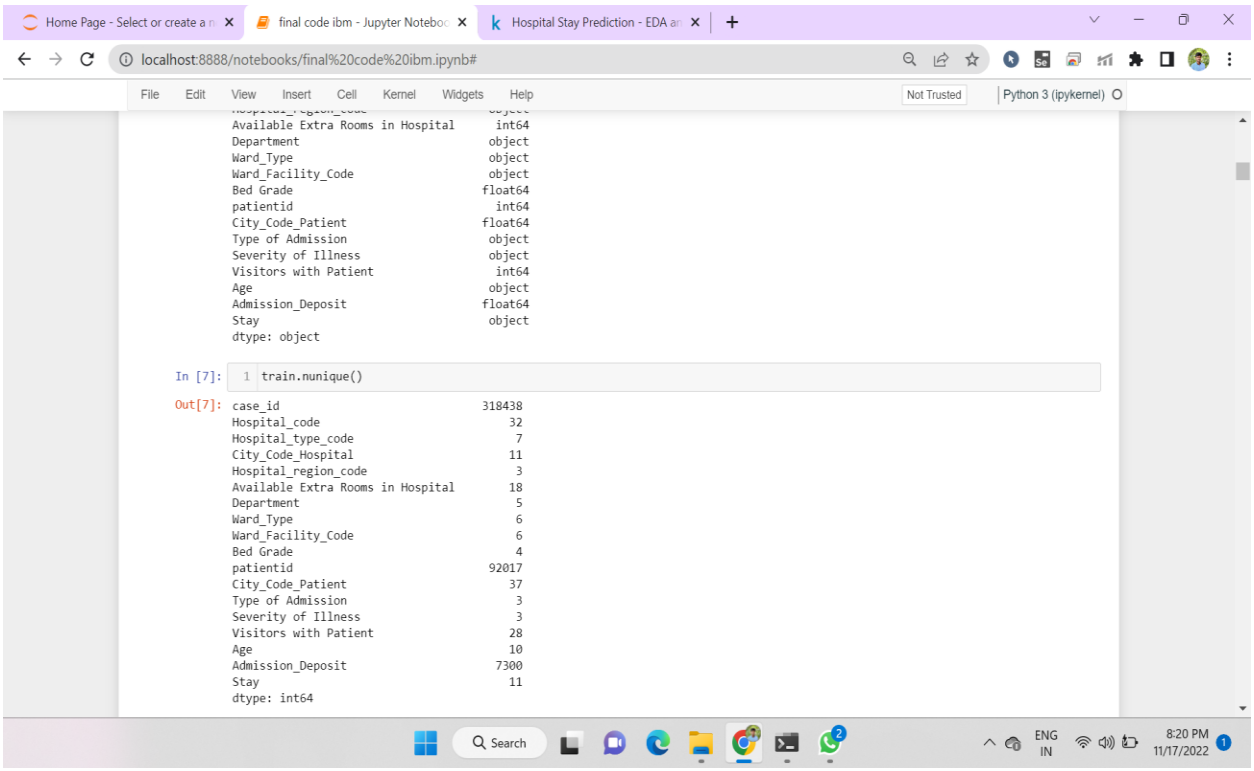
```
(318438, 18)
(137057, 17)
```

```
In [6]:
train.dtypes
```

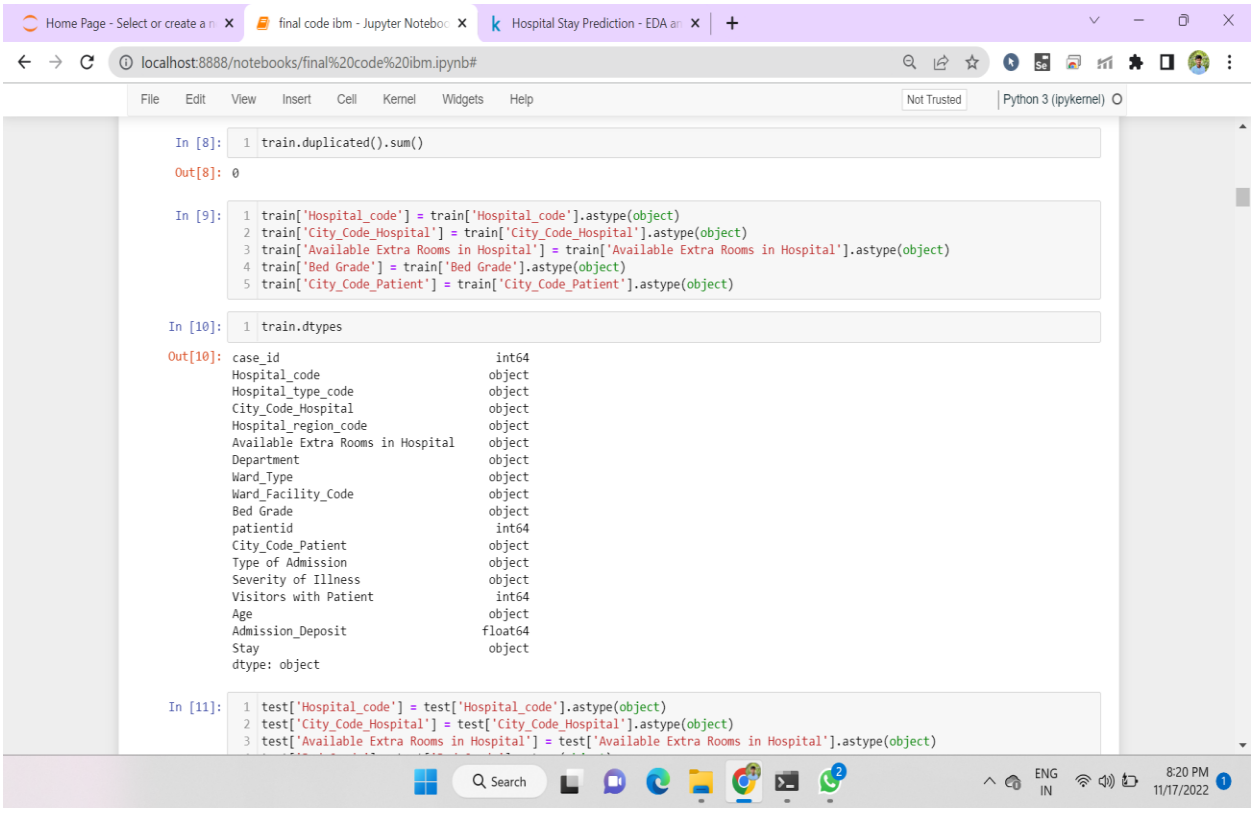
```
Out[6]: case_id                int64
Hospital_code             int64
Hospital_type_code       object
City_Code_Hospital        int64
```

Analytics for Hospitals Health-Care Data

TEAM ID: PNT2022TMID16292



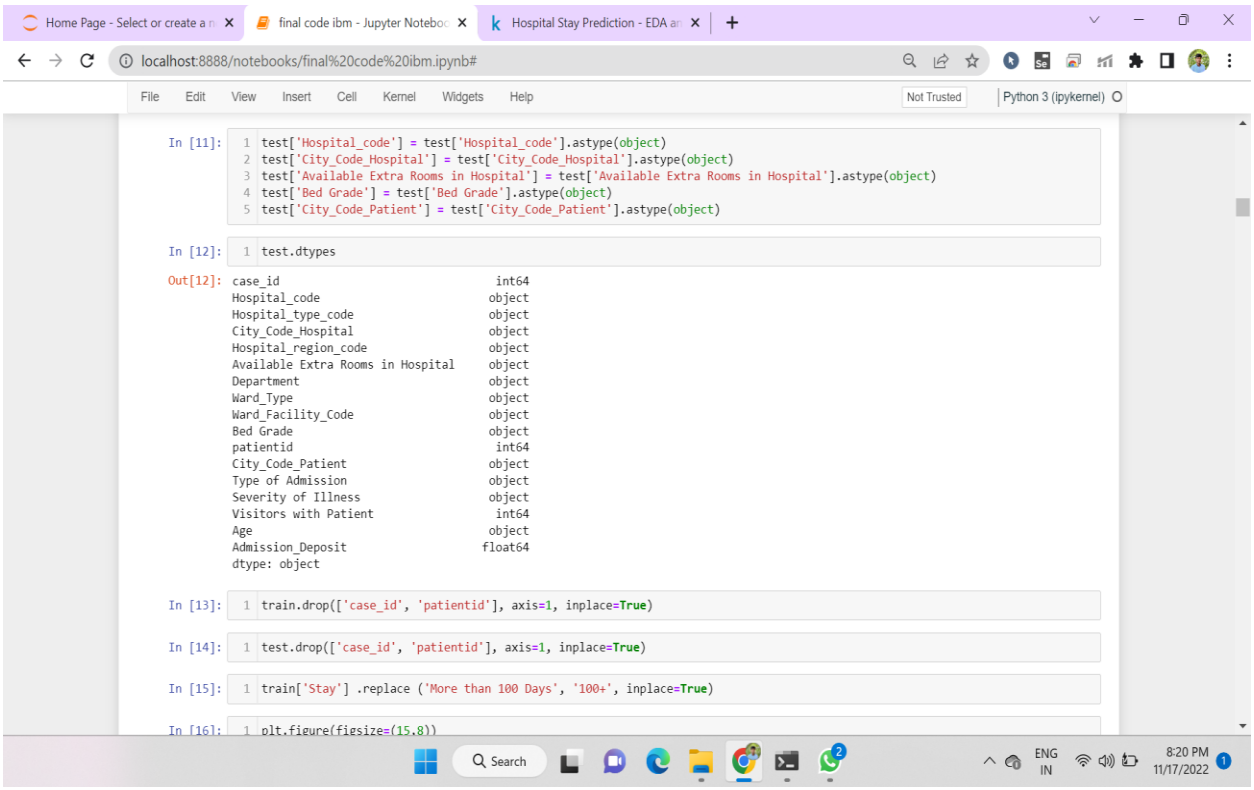
CHANGE DATA TYPES



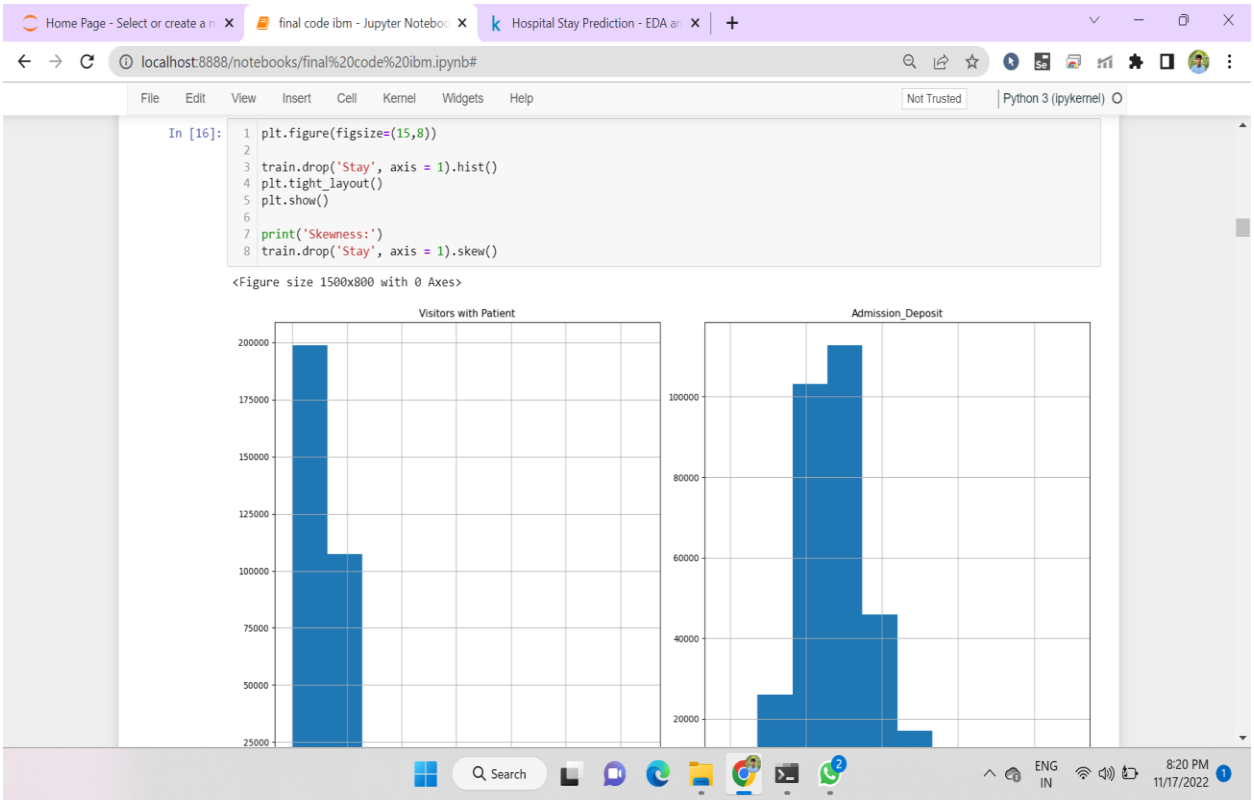
Analytics for Hospitals Health-Care Data

TEAM ID: PNT2022TMID16292

REMOVE INSIGNIFICANT VARIABLES



```
In [11]:
test['Hospital_code'] = test['Hospital_code'].astype(object)
test['City_Code_Hospital'] = test['City_Code_Hospital'].astype(object)
test['Available Extra Rooms in Hospital'] = test['Available Extra Rooms in Hospital'].astype(object)
test['Bed Grade'] = test['Bed Grade'].astype(object)
test['City_Code_Patient'] = test['City_Code_Patient'].astype(object)

In [12]:
test.dtypes

Out[12]:
case_id                             int64
Hospital_code                      object
Hospital_type_code                 object
City_Code_Hospital                 object
Hospital_region_code               object
Available Extra Rooms in Hospital  object
Department                         object
Ward_Type                          object
Ward_Facility_Code                 object
Bed Grade                          object
patientid                           int64
City_Code_Patient                  object
Type of Admission                  object
Severity of Illness                object
Visitors with Patient               int64
Age                                object
Admission_Deposit                 float64
dtype: object

In [13]:
train.drop(['case_id', 'patientid'], axis=1, inplace=True)

In [14]:
test.drop(['case_id', 'patientid'], axis=1, inplace=True)

In [15]:
train['Stay'] .replace ('More than 100 Days', '100+', inplace=True)

In [16]:
plt.figure(figsize=(15,8))
```



```
In [16]:
plt.figure(figsize=(15,8))

train.drop('Stay', axis = 1).hist()
plt.tight_layout()
plt.show()

print('Skewness:')
train.drop('Stay', axis = 1).skew()

<Figure size 1500x800 with 0 Axes>
```

Analytics for Hospitals Health-Care Data

TEAM ID: PNT2022TMID16292