

# Analytics for Hospitals Health-Care Data

TEAM ID: PNT2022TMID16292

## IMPORT LIBRARIES

```
Home Page - Select or create a notebook | final code ibm - Jupyter Notebook | +
localhost:8888/notebooks/final%20code%20ibm.ipynb#
File Edit View Insert Cell Kernel Widgets Help Notebook saved Not Trusted Python 3 (ipykernel)

In [ ]: 1 import os
        2 for dirname, _, filenames in os.walk('input'):
        3     for filename in filenames:
        4         print(os.path.join(dirname, filename))

In [2]: 1 import pandas as pd
        2 import numpy as np
        3 import matplotlib.pyplot as plt
        4 from matplotlib.colors import ListedColormap
        5 import seaborn as sns
        6 from warnings import filterwarnings
        7 filterwarnings('ignore')
        8 pd.options.display.max_columns = None
        9 pd.options.display.max_rows = None
        10 pd.options.display.float_format = '{:.6f}'.format
        11 from sklearn.model_selection import train_test_split
        12 import statsmodels
        13 import statsmodels.api as sm
        14 from sklearn.preprocessing import StandardScaler
        15 from sklearn import metrics
        16 from sklearn.linear_model import LogisticRegression
        17 from sklearn.metrics import classification_report
        18 from sklearn.metrics import cohen_kappa_score
        19 from sklearn.metrics import confusion_matrix
        20 from sklearn.metrics import roc_curve
        21 from sklearn.metrics import accuracy_score
        22 from sklearn.tree import DecisionTreeClassifier
        23 from sklearn.ensemble import RandomForestClassifier
        24 from sklearn import tree
        25 from sklearn.model_selection import GridSearchCV
        26 from sklearn.ensemble import AdaBoostClassifier, GradientBoostingClassifier
        27 from catboost import CatBoostClassifier
        28 from sklearn.feature_selection import RFE
        29 plt.rcParams['figure.figsize'] = [15, 8]
```

```
Home Page - Select or create a notebook | final code ibm - Jupyter Notebook | +
localhost:8888/notebooks/final%20code%20ibm.ipynb#
File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

In [3]: 1 train = pd.read_csv("D:/HealthCare/train_data.csv")
        2 test = pd.read_csv("D:/HealthCare/test_data.csv")

In [4]: 1 train.head()

Out[4]:
   case_id  Hospital_code  Hospital_type_code  City_Code_Hospital  Hospital_region_code  Available
   Extra Rooms  Department  Ward_Type  Ward_Facility_Code  Bed Grade  pati
   Hospital
0      1         8         c         3         Z         3  radiotherapy  R         F  2.000000  S
1      2         2         c         5         Z         2  radiotherapy  S         F  2.000000  S
2      3        10         e         1         X         2  anesthesia    S         E  2.000000  S
3      4        26         b         2         Y         2  radiotherapy  R         D  2.000000  S
4      5        26         b         2         Y         2  radiotherapy  S         D  2.000000  S

In [5]: 1 print(train.shape)
        2 print(test.shape)

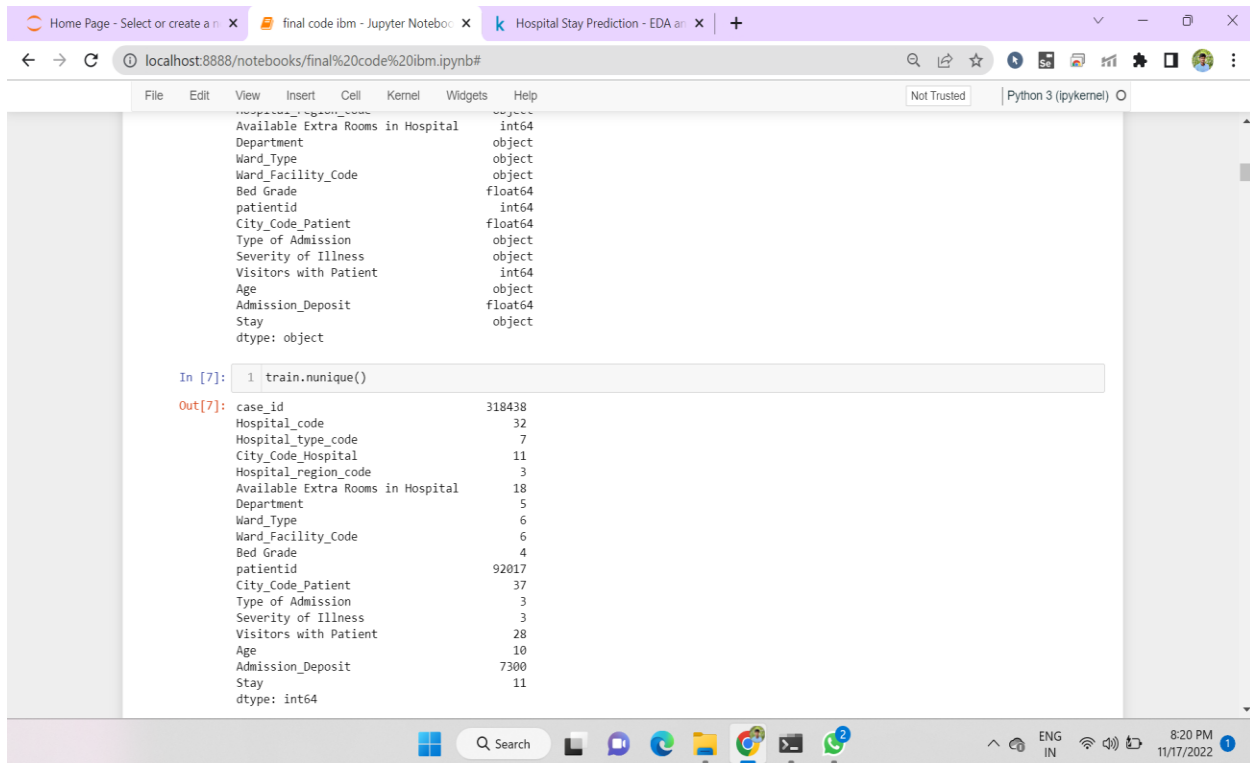
(318438, 18)
(137057, 17)

In [6]: 1 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



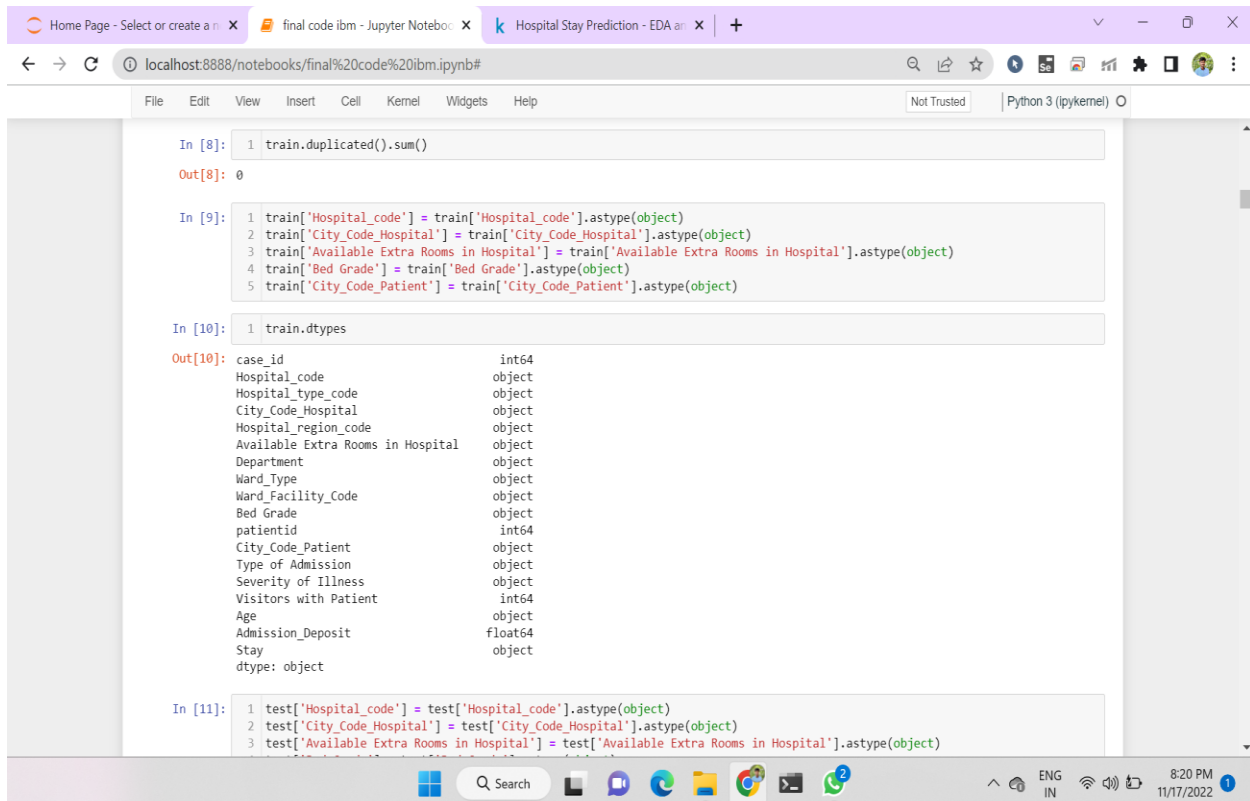
The screenshot shows a Jupyter Notebook interface with a browser window at the top. The notebook has a menu bar with File, Edit, View, Insert, Cell, Kernel, Widgets, and Help. The toolbar shows 'Not Trusted' and 'Python 3 (ipykernel)'. The code cell contains the following:

```
In [7]: 1 train.nunique()
```

The output shows the number of unique values for each feature:

```
Out[7]: case_id          318438
Hospital_code          32
Hospital_type_code      7
City_Code_Hospital     11
Hospital_region_code    3
Available Extra Rooms in Hospital  18
Department             5
Ward_Type              6
Ward_Facility_Code      6
Bed Grade              4
patientid             92017
City_Code_Patient       37
Type of Admission       3
Severity of Illness     3
Visitors with Patient   28
Age                   10
Admission_Deposit      7300
Stay                  11
dtype: int64
```

## CHANGE DATA TYPES



The screenshot shows the same Jupyter Notebook interface with the following code cells:

```
In [8]: 1 train.duplicated().sum()
Out[8]: 0
```

```
In [9]: 1 train['Hospital_code'] = train['Hospital_code'].astype(object)
2 train['City_Code_Hospital'] = train['City_Code_Hospital'].astype(object)
3 train['Available Extra Rooms in Hospital'] = train['Available Extra Rooms in Hospital'].astype(object)
4 train['Bed Grade'] = train['Bed Grade'].astype(object)
5 train['City_Code_Patient'] = train['City_Code_Patient'].astype(object)
```

```
In [10]: 1 train.dtypes
Out[10]: 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
Stay                  object
dtype: object
```

```
In [11]: 1 test['Hospital_code'] = test['Hospital_code'].astype(object)
2 test['City_Code_Hospital'] = test['City_Code_Hospital'].astype(object)
3 test['Available Extra Rooms in Hospital'] = test['Available Extra Rooms in Hospital'].astype(object)
```

# Analytics for Hospitals Health-Care Data

TEAM ID: PNT2022TMID16292

## REMOVE INSIGNIFICANT VARIABLES

```
Home Page - Select or create a n x final code ibm - Jupyter Notebo... k Hospital Stay Prediction - EDA an x +
localhost:8888/notebooks/final%20code%20ibm.ipynb#
File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel) O

In [11]: 1 test['Hospital_code'] = test['Hospital_code'].astype(object)
2 test['City_Code_Hospital'] = test['City_Code_Hospital'].astype(object)
3 test['Available Extra Rooms in Hospital'] = test['Available Extra Rooms in Hospital'].astype(object)
4 test['Bed Grade'] = test['Bed Grade'].astype(object)
5 test['City_Code_Patient'] = test['City_Code_Patient'].astype(object)

In [12]: 1 test.dtypes

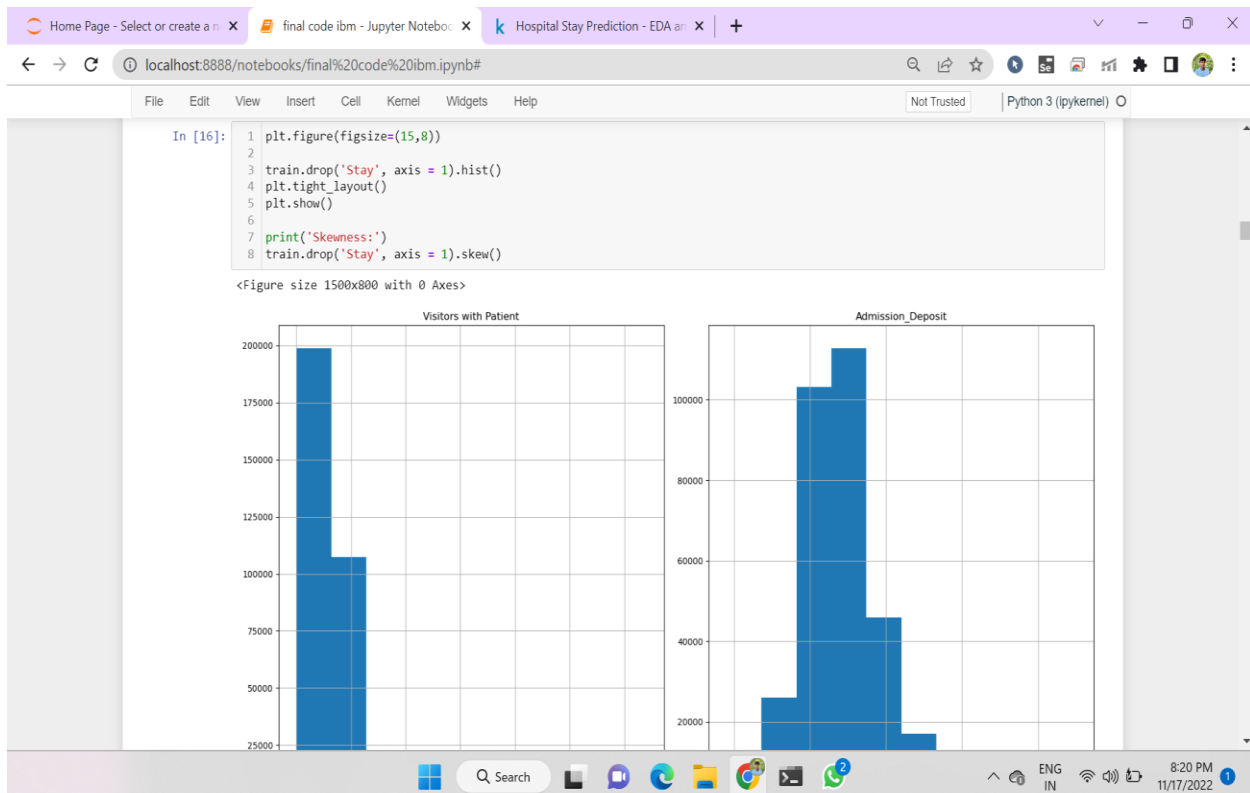
Out[12]: case_id          int64
Hospital_code      object
Hospital_type_code  object
City_Code_Hospital  object
Hospital_region_code object
Available Extra Rooms in Hospital object
Department         object
Ward_Type          object
Ward_Facility_Code  object
Bed Grade          object
patientid          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]: 1 train.drop(['case_id', 'patientid'], axis=1, inplace=True)

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

In [15]: 1 train['Stay'] .replace ('More than 100 Days', '100+', inplace=True)

In [16]: 1 plt.figure(figsize=(15,8))
```



# Analytics for Hospitals Health-Care Data

TEAM ID: PNT2022TMID16292

