

# SPRINT-1

TEAM ID	PNT2022TMID35169
PROJECT NAME	Analytics for Hospitals Health-Care Data

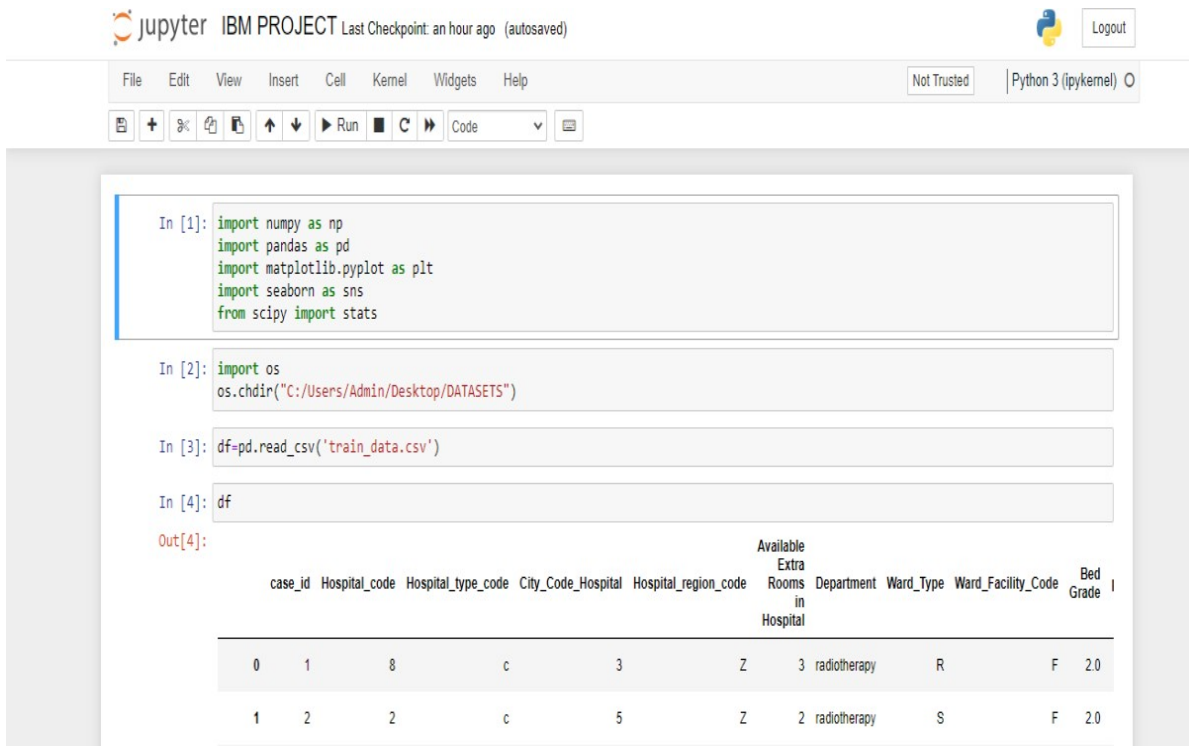
## Steps to be done

- Collection of data
- Data preprocessing
- Upload the dataset

## COLLECTION OF DATA:

[https://www.dropbox.com/s/eqe4g6nvg6qcsmu/Healthcare\\_Data.zip?dl=0](https://www.dropbox.com/s/eqe4g6nvg6qcsmu/Healthcare_Data.zip?dl=0)

## DATA PREPROCESSING:



The image shows a Jupyter Notebook interface with the following components:

- Header:** jupyter IBM PROJECT Last Checkpoint: an hour ago (autosaved) Python 3 (ipykernel) Logout
- Menu Bar:** File Edit View Insert Cell Kernel Widgets Help
- Toolbar:** Includes icons for file operations, running code, and a dropdown menu set to 'Code'.
- Code Cells:**
  - In [1]:** Imports libraries: `import numpy as np`, `import pandas as pd`, `import matplotlib.pyplot as plt`, `import seaborn as sns`, and `from scipy import stats`.
  - In [2]:** Imports `os` and sets the directory: `os.chdir("C:/Users/Admin/Desktop/DATASETS")`.
  - In [3]:** Reads the CSV file: `df=pd.read_csv('train_data.csv')`.
  - In [4]:** Displays the dataframe: `df`.
- Output [4]:** A preview of the dataframe with the following columns: `case_id`, `Hospital_code`, `Hospital_type_code`, `City_Code_Hospital`, `Hospital_region_code`, `Available Extra Rooms in Hospital`, `Department`, `Ward_Type`, `Ward_Facility_Code`, and `Bed Grade`. The first two rows of data are shown:

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
0	1	8	c	3	Z	3 radiotherapy	R	F	2.0
1	2	2	c	5	Z	2 radiotherapy	S	F	2.0

jupyter IBM PROJECT Last Checkpoint: an hour ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

318438 rows x 18 columns

```
In [5]: #Summary of the dataframe
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  ...
```

jupyter IBM PROJECT Last Checkpoint: an hour ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

```
In [14]: df.case_id

Out[14]: 0      1
         1      2
         2      3
         3      4
         4      5
         ...
318433  318434
318434  318435
318435  318436
318436  318437
318437  318438
Name: case_id, Length: 318438, dtype: int64

In [15]: df.Hospital_code

Out[15]: 0      8
         1      2
         2     10
         3     26
         4     26
         ..
318433    6
```

**Use Jupyter notebook to remove the null values**



In [22]: df.Ward\_Type

```
Out[22]: 0      R
         1      S
         2      S
         3      R
         4      S
         ..
        318433  Q
        318434  Q
        318435  R
        318436  Q
        318437  Q
        Name: Ward_Type, Length: 318438, dtype: object
```

In [23]: df.Ward\_Facility\_Code

```
Out[23]: 0      F
         1      F
         2      E
         3      D
         4      D
         ..
        318433  F
        318434  E
```



In [42]: df['Bed Grade'].unique()

```
Out[42]: array([ 2.,  3.,  4.,  1., nan])
```

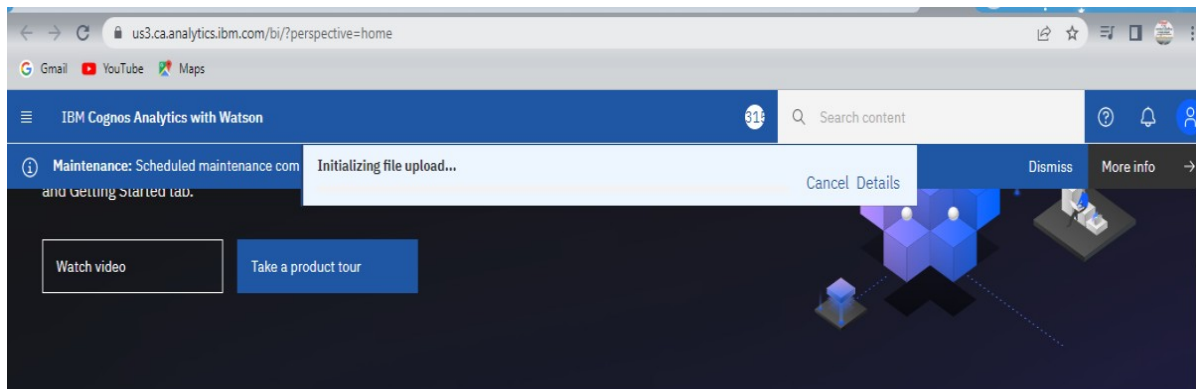
In [43]: df['Visitors with Patient'].unique()

```
Out[43]: array([ 2,  4,  3,  8,  6,  7, 13,  5,  1, 10, 15, 11, 12,  9, 24, 16, 14,
                20,  0, 19, 18, 17, 23, 21, 32, 30, 22, 25], dtype=int64)
```


In [44]: df['Severity of Illness']

```
Out[44]: 0      Extreme
         1      Extreme
         2      Extreme
         3      Extreme
         4      Extreme
         ...
        318433  Moderate
        318434  Moderate
        318435  Minor
        318436  Minor
        318437  Minor
        Name: Severity of Illness, Length: 318438, dtype: object
```

# Upload the dataset:




## Quick launch




**Upload data**

Upload or drag and drop spreadsheets, csv files, and other data sources.




**Prepare data**

Use data modules to clean and connect data from multiple resources.



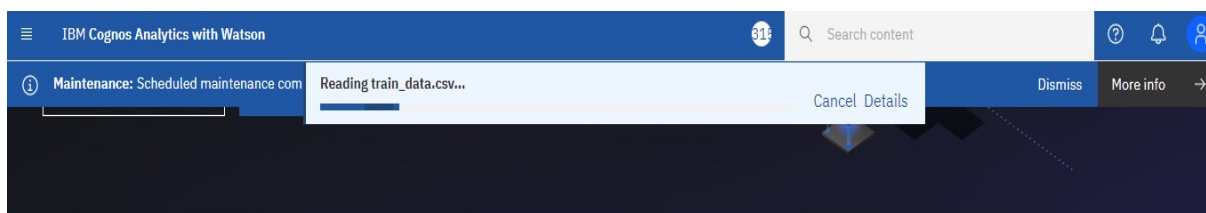
**Exploration**

Quickly find unbiased answers by identifying trends in your data with data explor...




**Present data**

Create sophisticated, multi-page, multi-query dashboards, reports, or stori...




## Quick launch




**Upload data**

Upload or drag and drop spreadsheets, csv files, and other data sources.




**Prepare data**

Use data modules to clean and connect data from multiple resources.



**Exploration**

Quickly find unbiased answers by identifying trends in your data with data explor...



**Present data**

Create sophisticated, multi-page, multi-query dashboards, reports, or stori...

Get started

Recent