# Project Development Phase
# Delivery of Sprint-1

| | |
|---|---|
| **Date** | 02 NOV 2022 |
| **Team ID** | PNT2022TMID01342 |
| **Project Name** | Analytics for Hospitals' Health-Care Data |

## Project Development Phase:

**Sprint-1:**

- ➢ Data Collection
- ➢ Data Preparation

**Sprint-2:**

- ➢ Data Exploration

**Sprint-3:**

- ➢ Dashboard Creation

**Sprint-4:**

- ➢ Report Creation
- ➢ Story Creation

**Contributors:**

R.Adithya

K.Charan

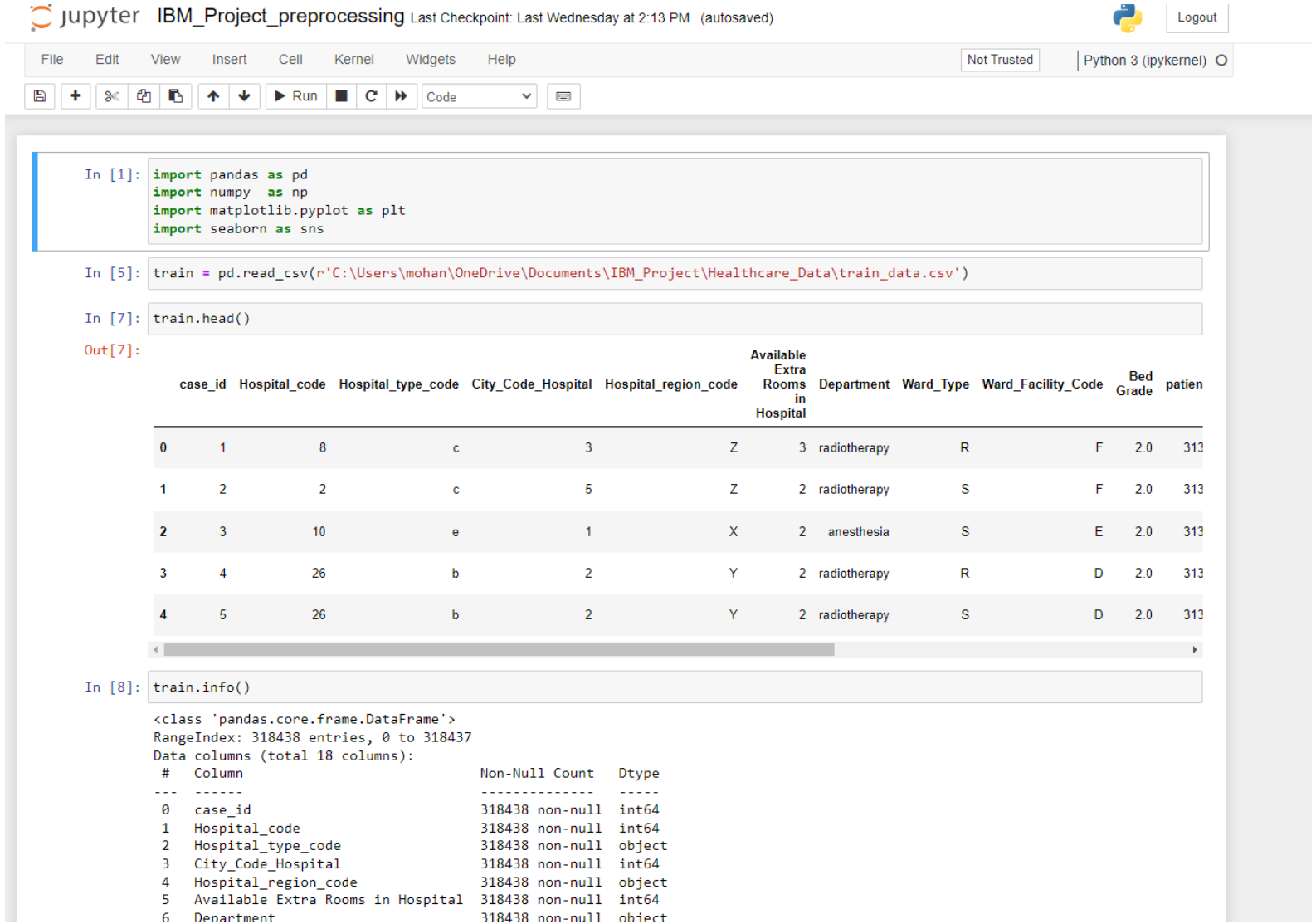
C.Lingeshwaran

V.Mohanraj

# SPRINT-1

## Data Preprocessing:

## Using Jupyter notebook to remove the null values:




| | 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 | patien |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 8 | c | 3 | Z | 3 | radiotherapy | R | F | 2.0 | 313 |
| 1 | 2 | 2 | c | 5 | Z | 2 | radiotherapy | S | F | 2.0 | 313 |
| 2 | 3 | 10 | e | 1 | X | 2 | anesthesia | S | E | 2.0 | 313 |
| 3 | 4 | 26 | b | 2 | Y | 2 | radiotherapy | R | D | 2.0 | 313 |
| 4 | 5 | 26 | b | 2 | Y | 2 | radiotherapy | S | D | 2.0 | 313 |

```
4   5   26   b   2   1   2   Radiotherapy   S   D   2.0   31
```

In [17]:
```
train.isnull().sum()
```

Out[17]:
```
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
```

In [18]:
```
train = train.to_csv('final_train_data.csv', index=False)
```

In [19]:
```
test = pd.read_csv(r'C:\Users\mohan\OneDrive\Documents\IBM_Project\Healthcare_Data\test_data.csv')
```

In [20]:
```
test.head()
```

Out[20]:

| | 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 | patien |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 318439 | 21 | c | 3 | Z | 3 | gynecology | S | A | 2.0 | 170 |
| 1 | 318440 | 29 | a | 4 | X | 2 | gynecology | S | F | 2.0 | 170 |

```
Stay                                  0
dtype: int64
```

In [11]:
```
train['Bed Grade'].value_counts()
```

Out[11]:
```
2.0    123671
3.0    110583
4.0     57566
1.0     26505
Name: Bed Grade, dtype: int64
```

In [12]:
```
train['Bed Grade'].unique()
```

Out[12]:
```
array([ 2.,  3.,  4.,  1., nan])
```

In [13]:
```
train.shape
```

Out[13]:
```
(318438, 18)
```

In [14]:
```
train.dropna(inplace=True)
```

In [15]:
```
train.shape
```

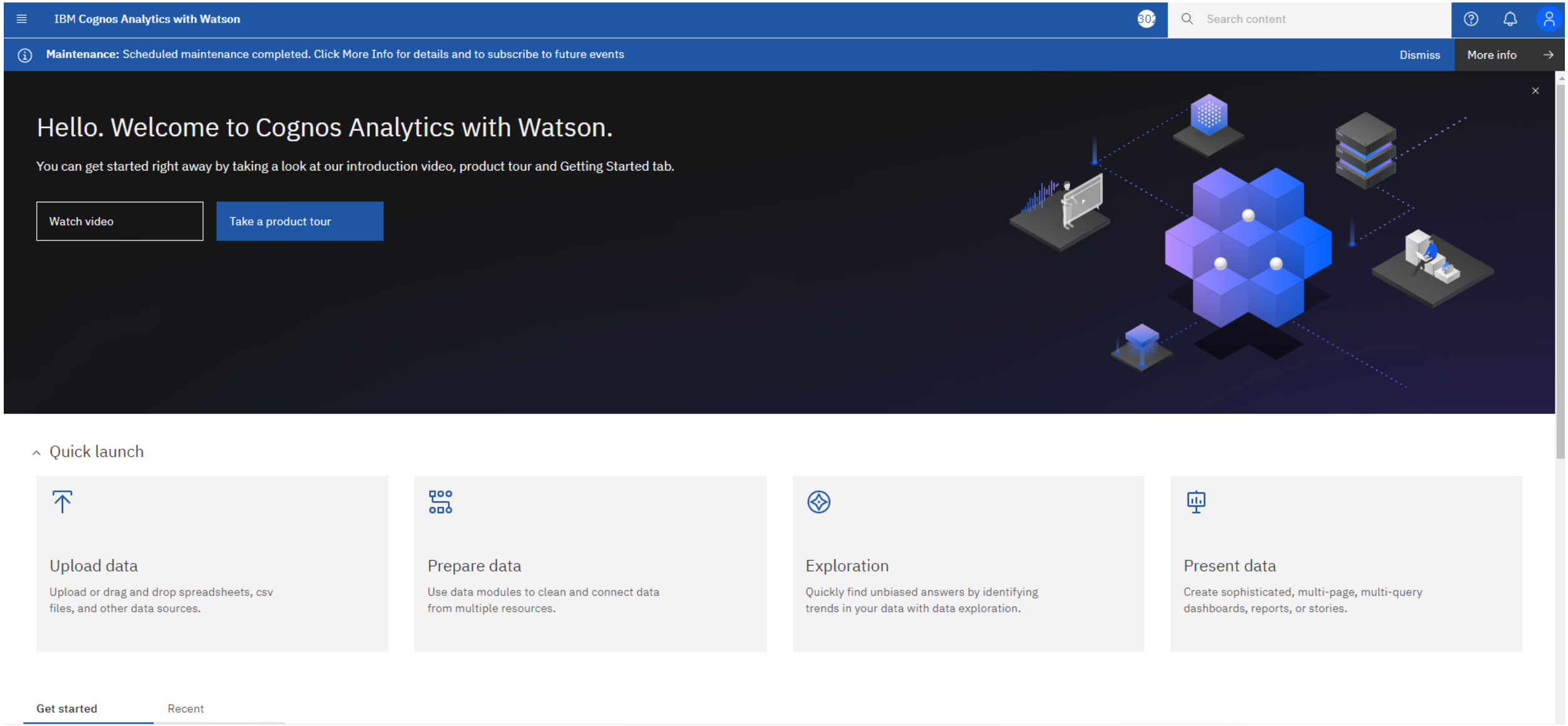
Out[15]:
```
(313793, 18)
```

In [16]:
```
train.head()
```

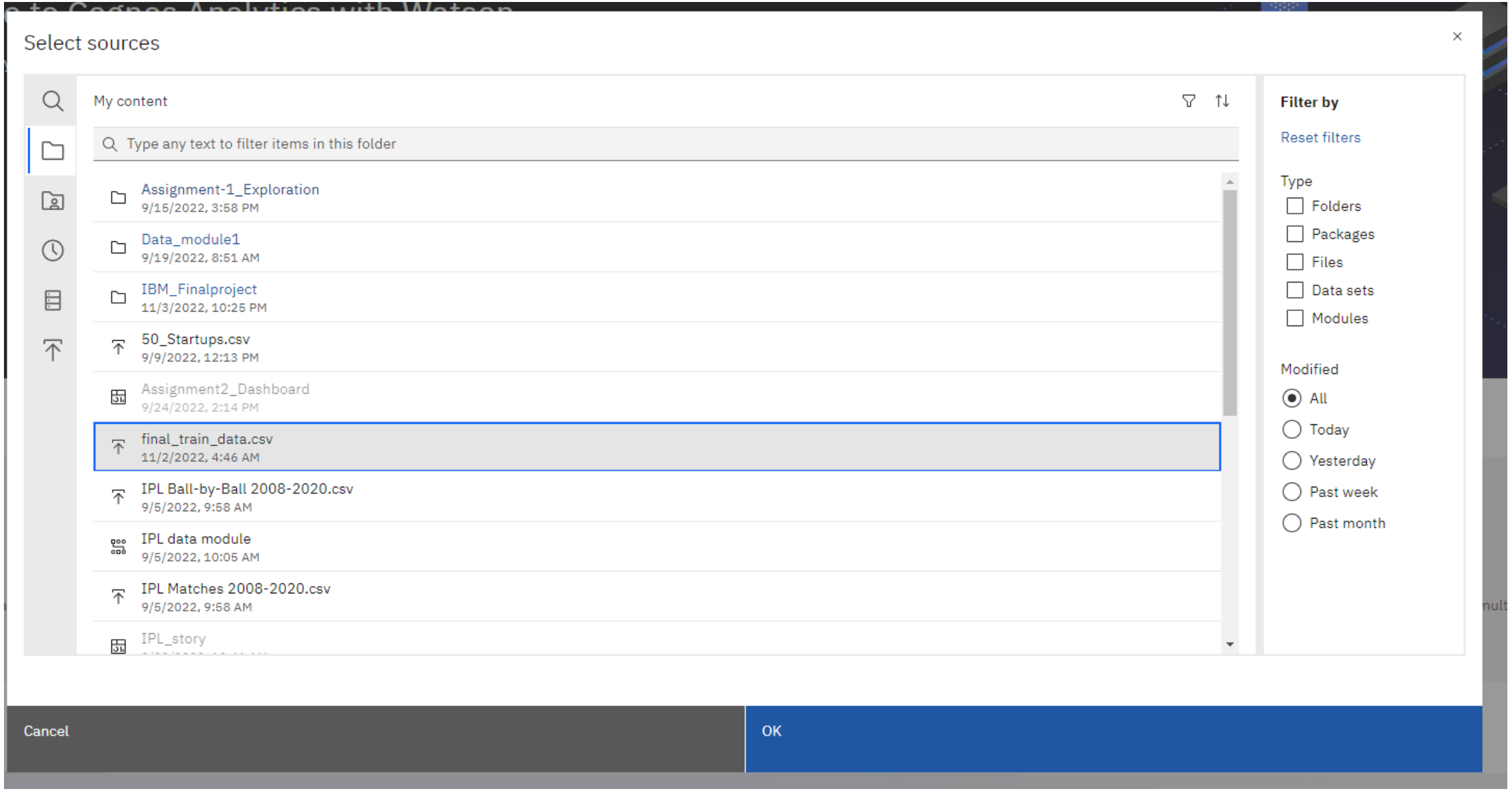
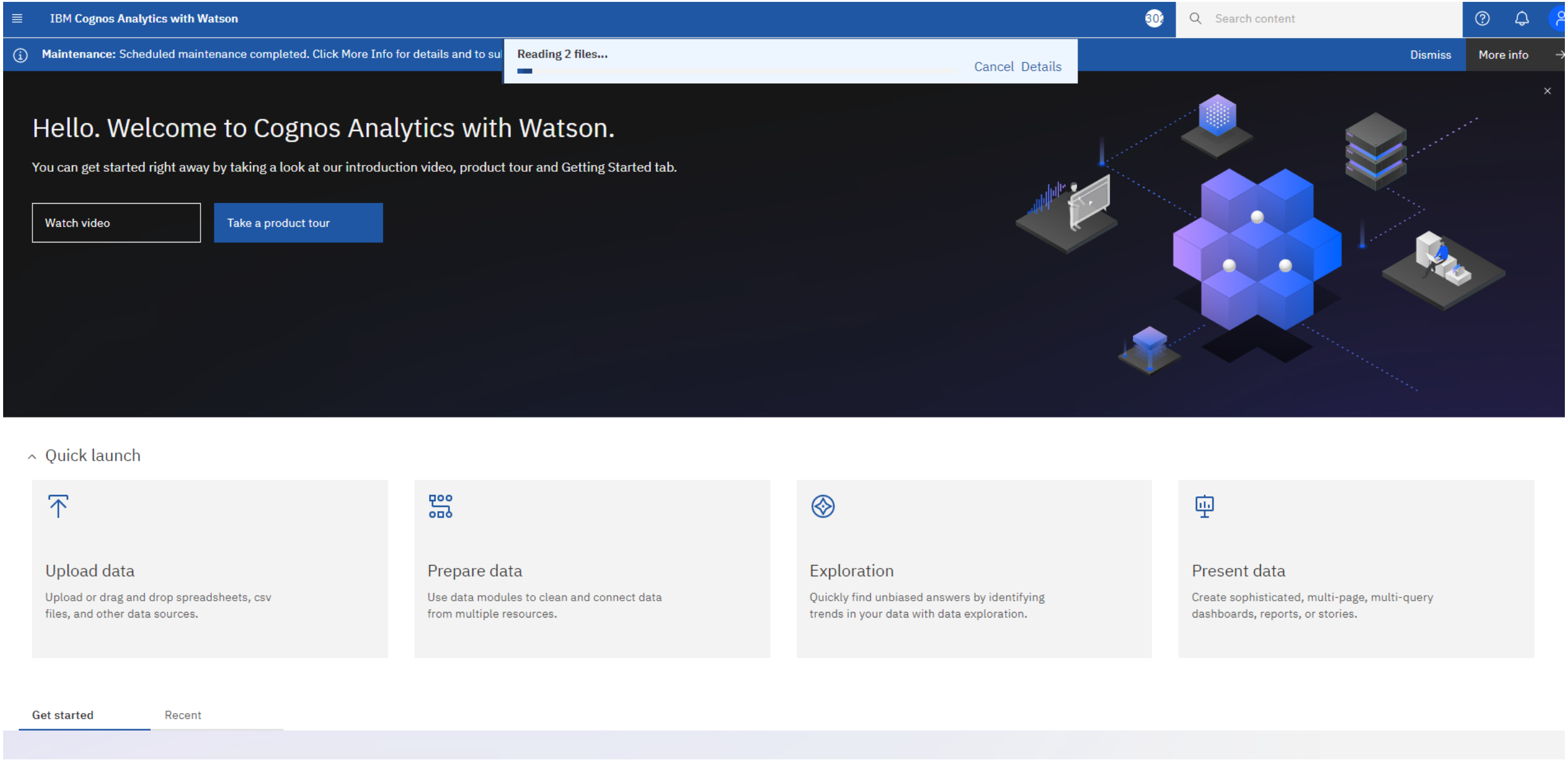
Out[16]:

| | 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 | patien |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 8 | c | 3 | Z | 3 | radiotherapy | R | F | 2.0 | 313 |
| **1** | 2 | 2 | c | 5 | Z | 2 | radiotherapy | S | F | 2.0 | 313 |
| **2** | 3 | 10 | e | 1 | X | 2 | anesthesia | S | E | 2.0 | 313 |
| 3 | 4 | 26 | b | 2 | Y | 2 | radiotherapy | R | D | 2.0 | 313 |

# Signing in to the IBM Cognos tool:



# Selecting the dataset to upload into the Cognos:



# Uploading the dataset to perform the preparation and visualization: