

## Sprint-II (USN-3)

Date	4 November 2022
Team ID	PNT2022TMID27483
Project Name	Analytics for Hospitals' Healthcare Data

### DATA PREPARATION AND CLEANING:

**Original Dataset:** [https://drive.google.com/file/d/1sIC0MhsJHeuODVkhIXrdNOX\\_aBDySSBh/view](https://drive.google.com/file/d/1sIC0MhsJHeuODVkhIXrdNOX_aBDySSBh/view)

#### Uploading the dataset

The screenshot shows the IBM Cognos Analytics with Watson interface. At the top, a blue header bar contains the logo, a search bar, and user icons. Below the header, a dark blue banner features the 'Watson.' logo and a progress bar indicating 'Reading 2 files...'. A 'Cancel Details' link is visible next to the progress bar. Below the banner, there are two buttons: 'Watch video' and 'Take a product tour'. Underneath, a 'Quick launch' section displays four cards: 'Upload data' (with an upload icon), 'Prepare data' (with a data module icon), 'Exploration' (with a magnifying glass icon), and 'Present data' (with a presentation icon). Each card includes a brief description of its function.

#### Select sources

The screenshot shows the 'My content' view in IBM Cognos Analytics. A search bar at the top allows filtering items. Below the search bar, a list of data sources is displayed, each with an icon, a name, and a timestamp. The items are: 'Assignment-1' (9/12/2022, 1:58 PM), '50\_Startups.csv' (9/12/2022, 7:56 AM), 'Pharma Sales Dashboard' (10/2/2022, 3:50 PM), 'Pharma\_Monthly\_Sales.csv' (10/2/2022, 2:07 PM), 'test\_data.csv' (11/10/2022, 6:02 AM), 'train\_data.csv' (11/10/2022, 6:03 AM), and 'US Superstore data.xls' (9/12/2022, 8:09 AM). A sidebar on the left contains navigation icons for folders, content, and other features.

IBM Cognos Analytics with Watson							307	Search content	?	🔔	👤
Data module											
Grid Relationships Custom tables											
Row Id	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hosp						
1	318439	21	c	3	Z						
2	318440	29	a	4	X						
3	318441	26	b	2	Y						
4	318442	6	a	6	X						
5	318443	28	b	11	X						
6	318444	23	a	6	X						
7	318445	26	b	2	Y						
8	318446	25	e	1	X						
9	318447	23	a	6	X						

## Cleaning the empty cells

Create calculation

Name City\_Code\_Patient

Components

Expression

Search
train_data.csv
BedGrade
City_Cod...Patient
# Row Id
# case_id
# Hospi...code
# Hospi...code
City_C...spital
1
2
3

```

1 If(City_Code_Patient is missing)
2 THEN (median (City_Code_Patient ))
3 ELSE(City_Code_Patient)

```

Validation Results

✓ The expression is valid.

☐ Calculate after aggregation

Cancel

OK

IBM Cognos Analytics with Watson							307	Search content	?	🔔	👤
Data module											
Grid Relationships Custom tables											
City_Code_Patient	Row Id	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region_code					
7	5	5	26	b	2	Y					
7	6	6	23	a	6	X					
7	7	7	32	f	9	Y					
7	8	8	23	a	6	X					
7	9	9	1	d	10	Y					
7	10	10	10	e	1	X					
7	11	11	22	g	9	Y					
7	12	12	26	b	2	Y					
7	13	13	16	c	3	Z					
7	14	14	9	d	5	Z					
8	15	15	6	a	6	X					
8	16	16	6	a	6	X					
8	17	17	23	a	6	X					

**Create calculation**

Name:

Components

- train\_data.csv
  - BedGrade
  - City\_Cod...Patient
    - # Row Id
    - # case\_id
    - # Hospital\_code
    - Hospital...code
    - City\_C...spital
    - Hospital...code
    - Availabl...ospital
    - Department

Expression

```

1 IF(BedGrade is missing)
2 THEN (median (BedGrade ))
3 ELSE (BedGrade)

```

Validation Results

✓ The expression is valid.

☐ Calculate after aggregation

Cancel OK

IBM Cognos Analytics with Watson

\* New data module

Search content

Properties

Data module

Search

- New data module
- Navigation paths
- train\_data.csv
  - BedGrade
  - City\_Code\_Patient
    - # Row Id
    - # case\_id
    - # Hospital\_code
    - Hospital...pe\_code
    - City\_Cod...Hospital
    - Hospital...on\_code
    - Availabl...Hospital
    - Department
    - Ward\_Type
    - Ward\_Fa...ty\_Code
    - Bed Grade
    - patientid

Grid

	BedGrade	City_Code_Patient	Row Id	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital
2		7	1	1	8	c	3
2		7	2	2	2	c	5
2		7	3	3	10	e	1
2		7	4	4	26	b	2
2		7	5	5	26	b	2
2		7	6	6	23	a	6
3		7	7	7	32	f	9
3		7	8	8	23	a	6
4		7	9	9	1	d	10
3		7	10	10	10	e	1
2		7	11	11	22	g	9
1		7	12	12	26	b	2
3		7	13	13	16	c	3

IBM Cognos Analytics with Watson

\* New data module

Search content

Properties

Data module

Search

- New data module
- Navigation paths
- train\_data.csv
  - BedGrade
  - City\_Code\_Patient
    - # Row Id
    - # case\_id
    - # Hospital\_code
    - Hospital...pe\_code
    - City\_Cod...Hospital
    - Hospital...on\_code
    - Availabl...Hospital
    - Department
    - Ward\_Type
    - Ward\_Fa...ty\_Code
    - Bed Grade
    - patientid

Grid

	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region_code	Available E...in Hospital	Department
1		8	c	3	Z	3	radiotherapy
2		2	c	5	Z	2	radiotherapy
3		10	e	1	X	2	anesthesia
4		26	b	2	Y	2	radiotherapy
5		26	b	2	Y	2	radiotherapy
6		23	a	6	X	2	anesthesia
7		32	f	9	Y	1	radiotherapy
8		23	a	6	X	4	radiotherapy
9		1	d	10	Y	2	gynecology
10		10	e	1	X	2	gynecology
11		22	g	9	Y	2	radiotherapy
12		26	b	2	Y	4	radiotherapy
13		16	c	3	Z	2	radiotherapy

**Prepared Dataset:** [https://us3.ca.analytics.ibm.com/bi/?perspective=ca-modeller&pathRef=.my\\_folders%2FData%2BPreparation](https://us3.ca.analytics.ibm.com/bi/?perspective=ca-modeller&pathRef=.my_folders%2FData%2BPreparation)