

Sprint-I (USN-1)

Date	27 October 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

Uploading the dataset

The screenshot shows the IBM Cognos Analytics with Watson interface. At the top, there's a header with the IBM logo and 'IBM Cognos Analytics with Watson'. Below this, a 'Watson.' banner features a progress bar indicating 'Reading 2 files...' with 'Cancel' and 'Details' buttons. Below the banner, there are two buttons: 'Watch video' and 'Take a product tour'. The main area is titled 'Quick launch' and contains four cards: 'Upload data' (with an icon of an arrow pointing up), 'Prepare data' (with an icon of a cube), 'Exploration' (with an icon of a magnifying glass), and 'Present data' (with an icon of a presentation screen). Each card has a brief description of its function.

Select sources

The screenshot shows the 'Select sources' interface. On the left, there's a sidebar with icons for search, folder, user, clock, list, and upload. The main area is titled 'My content' and has a search bar with the placeholder text 'Type any text to filter items in this folder'. Below the search bar, there's a list of files and folders. The files 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). Each file has an icon representing its type (folder or CSV/XLS file) and a small 'upload' icon.

IBM Cognos Analytics with Watson						
New data module						
Search content						
Properties						
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	
abc 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						
New data module						
Search content						
Properties						
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

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

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