

Sprint-I

Date	4 November 2022
Team ID	PNT2022TMID27559
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 large banner area features the 'Watson.' logo and a message: 'You can get started right away by taking a look at our introduction video, product tour and Getting Started tab.' There are two buttons: 'Watch video' and 'Take a product tour'. A progress bar indicates 'Reading 2 files...' with a 'Cancel' and 'Details' link. Below the banner, there's a 'Quick launch' section with four tiles: 'Upload data' (with an upload icon), 'Prepare data' (with a data icon), 'Exploration' (with a magnifying glass icon), and 'Present data' (with a presentation icon). Each tile has a brief description of its function.

Select sources

The screenshot shows the 'Select sources' dialog box. It has a search bar at the top with the text 'My content'. Below the search bar, there's a list of items. The first item is '50_Startups.csv' with a timestamp of '9/17/2022, 11:38 AM'. The second item is 'Age wise patient visualisation' with a timestamp of '11/17/2022, 6:10 AM'. The third item is 'bubbles and radial dashboard' with a timestamp of '11/17/2022, 6:19 AM'. The fourth item is 'dashboard list of patients' with a timestamp of '11/17/2022, 6:14 AM'. The fifth item is 'dashboard showing waterfall, pie, bar' with a timestamp of '11/17/2022, 6:51 AM'. The sixth item is 'Data Preparation' with a timestamp of '11/17/2022, 6:51 AM'. The seventh item is 'Number of patients by ward type' with a timestamp of '11/17/2022, 6:03 AM'. The eighth item is 'test_data.csv' with a timestamp of '11/17/2022, 5:48 AM'. The ninth item is 'train_data.csv' with a timestamp of '11/17/2022, 5:52 AM'. On the right side, there's a 'Filter by' section with a 'Reset filters' link. Under 'Type', there are checkboxes for 'Folders', 'Packages', 'Files', 'Data sets', and 'Modules'. Under 'Modified', there are radio buttons for 'All', 'Today', 'Yesterday', 'Past week', and 'Past month'. The 'All' radio button is selected.

Cancel

OK

IBM Cognos Analytics with Watson						
New data module						
Search content						
Properties						
Data module						
Search						
New data module						
Navigation paths						
test_data.csv						
# Row Id						
# case_id						
# Hospital_code						
Hospital_type_code						
City_Code_Hospital						
Hospital_region_code						
Department						
Ward_Type						

Cleaning the empty cells

Create calculation

Name City_Code_Patient

Components

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

Expression

```
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						
Search						
New data module						
Navigation paths						
train_data.csv						
test_data.csv						

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: BedGrade

Components

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://us1.ca.analytics.ibm.com/bi/?perspective=ca-modeller&pathRef=.my_folders%2FData%2BPreparation