

Sprint 1

Date	29 October 2022
Team ID	PNT2022TMID07721
Project Name	Analytics for Hospital's Health-Care Data

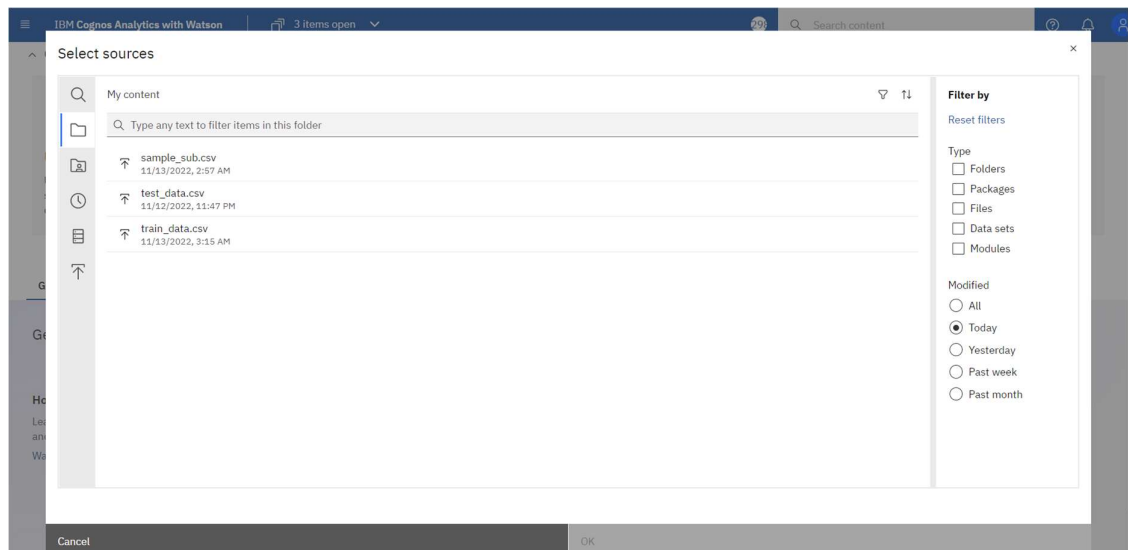
Data collection and Data preparation –

- Hospital data is collected from various resources and hospital database. These data are made into datasets with csv extension
- Refer the dataset in this link:
<https://www.kaggle.com/datasets/dharshanprasathh/hospital-healthcare-data-analytics>

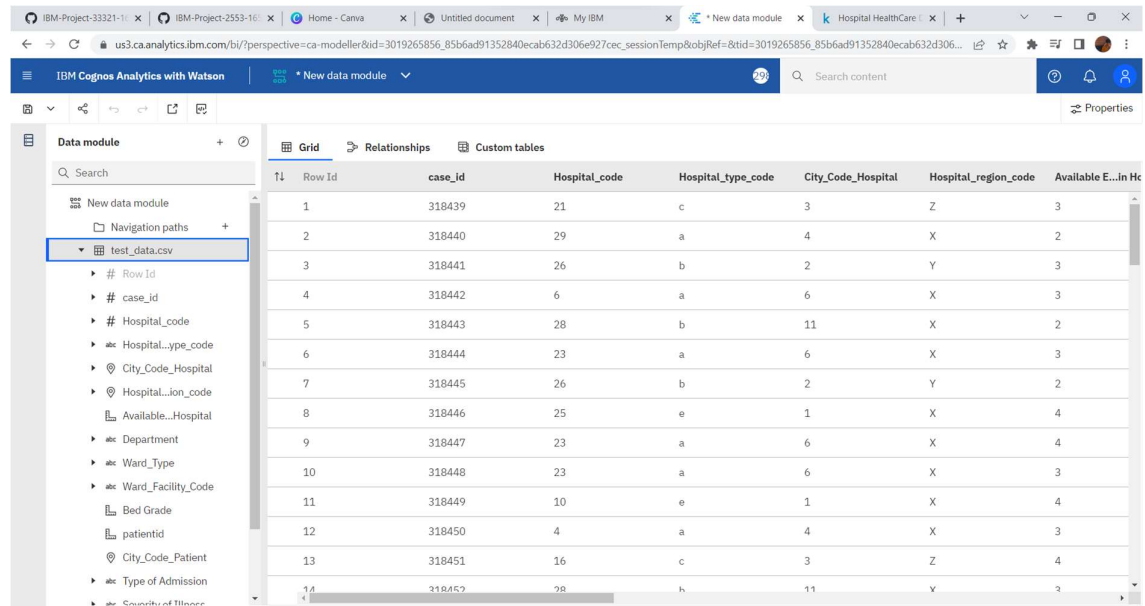
Dataset definition:

- The dataset consists of four csv files:
 1. test_data.csv
 2. sample_sub.csv
 3. test_data.csv
 4. train_data.csv

Uploading dataset in IBM Cognos



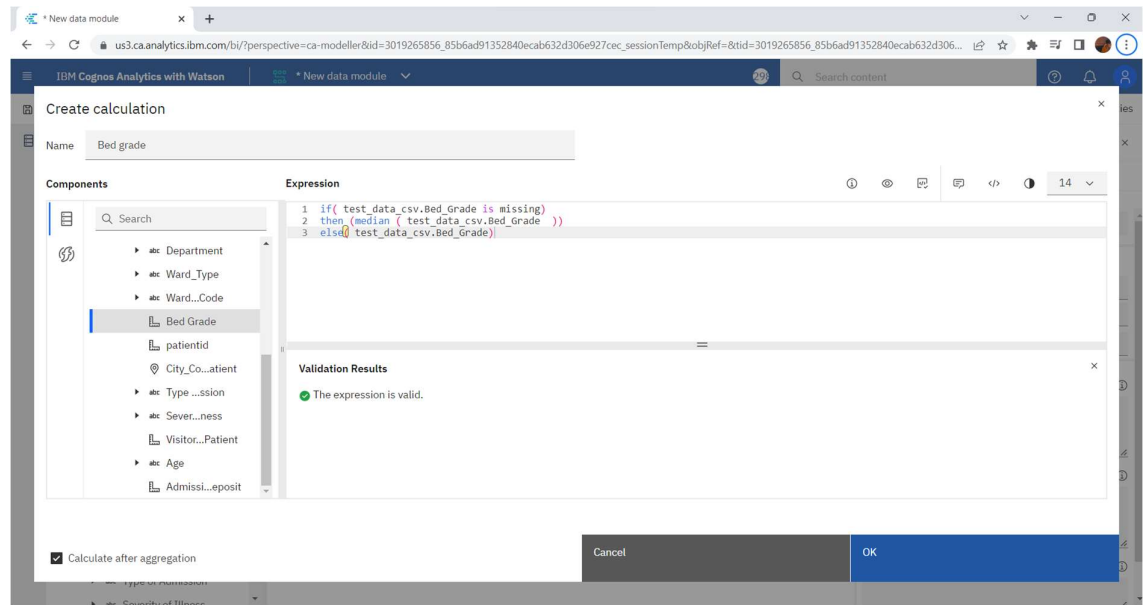
Data Prep using IBM Cognos



The screenshot shows the IBM Cognos Analytics interface. On the left, the 'Data module' pane displays a search bar and a list of fields from the 'test_data.csv' dataset. The main area shows a 'Grid' view of the data. The table has 14 rows and 8 columns: Row Id, case_id, Hospital_code, Hospital_type_code, City_Code_Hospital, Hospital_region_code, and Available E...In Hc.

Row Id	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region_code	Available E...In Hc
1	318439	21	c	3	Z	3
2	318440	29	a	4	X	2
3	318441	26	b	2	Y	3
4	318442	6	a	6	X	3
5	318443	28	b	11	X	2
6	318444	23	a	6	X	3
7	318445	26	b	2	Y	2
8	318446	25	e	1	X	4
9	318447	23	a	6	X	4
10	318448	23	a	6	X	3
11	318449	10	e	1	X	4
12	318450	4	a	4	X	3
13	318451	16	c	3	Z	4
14	318452	28	b	11	Y	3

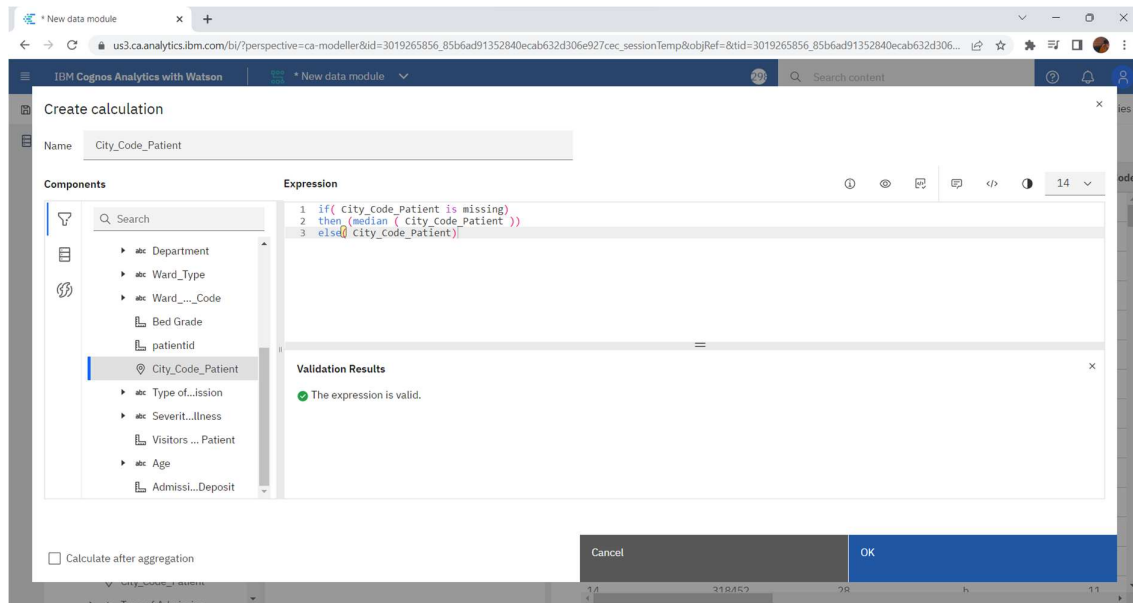
Adding data filters in IBM Cognos



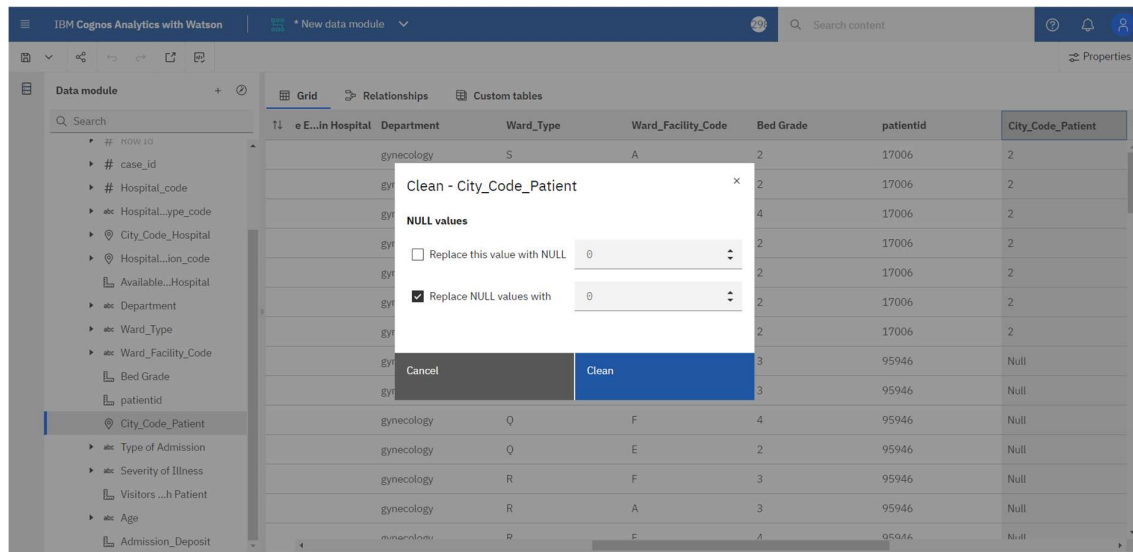
The screenshot shows the 'Create calculation' dialog in IBM Cognos Analytics. The 'Name' field is set to 'Bed grade'. The 'Components' pane on the left shows a list of fields, with 'Bed Grade' selected. The 'Expression' field contains the following code:

```
1 if( test_data_csv.Bed_Grade is missing)
2 then (median ( test_data_csv.Bed_Grade ))
3 else( test_data_csv.Bed_Grade)
```

The 'Validation Results' section shows a green checkmark and the message: 'The expression is valid.' At the bottom, there are 'Cancel' and 'OK' buttons. The 'Calculate after aggregation' checkbox is checked.



Removing Null values in the dataset (Cleaning)



After Data cleaning

IBM Cognos Analytics with Watson

* New data module

Search content

Properties

Data module

Grid Relationships Custom tables

Search

ROW ID

case_id

Hospital_code

Hospital...ype_code

City_Code_Hospital

Hospital...ion_code

Available...Hospital

Department

Ward_Type

Ward_Facility_Code

Bed Grade

patientid

City_Code_Patient

Type of Admission

Severity of Illness

Visitors ...h Patient

Age

Admission_Deposit

ROW ID	Ward_Type	Ward_Facility_Code	Bed Grade	patientid	City_Code_Patient	Type of Admission	Severity of Illness
1	S	A	2	17006	2	Emergency	Mod...
2	S	F	2	17006	2	Trauma	Mod...
3	Q	D	4	17006	2	Emergency	Mod...
4	Q	F	2	17006	2	Trauma	Mod...
5	R	F	2	17006	2	Trauma	Mod...
6	Q	F	2	17006	2	Trauma	Mod...
7	Q	D	2	17006	2	Trauma	Mod...
8	S	E	3	95946	0	Emergency	Mod...
9	Q	F	3	95946	0	Trauma	Mod...
10	Q	F	4	95946	0	Urgent	Mod...
11	Q	E	2	95946	0	Trauma	Mod...
12	R	F	3	95946	0	Emergency	Mod...
13	R	A	3	95946	0	Trauma	Mod...
14	R	F	3	95946	0	Urgent	Mod...

IBM Cognos Analytics with Watson

* New data module

Search content

Properties

Data module

Grid Relationships Custom tables

Search

ROW ID

case_id

Hospital_code

Hospital...ype_code

City_Code_Hospital

Hospital...ion_code

Available...Hospital

Department

Ward_Type

Ward_Facility_Code

Bed Grade

patientid

City_Code_Patient

Type of Admission

Severity of Illness

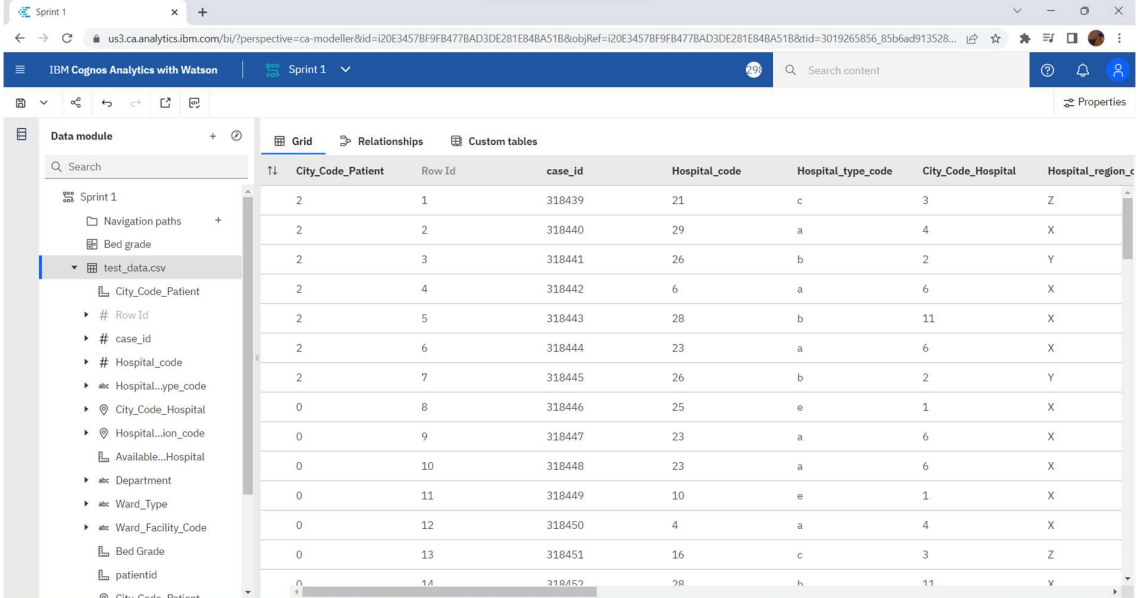
Visitors ...h Patient

Age

Admission_Deposit

ROW ID	Available E...in Hospital	Department	Ward_Type	Ward_Facility_Code	Bed Grade	patientid	City_Code_Pati
1	3	gynecology	S	A	2	17006	2
2	2	gynecology	S	F	2	17006	2
3	3	gynecology	Q	D	4	17006	2
4	3	gynecology	Q	F	2	17006	2
5	2	gynecology	R	F	2	17006	2
6	3	gynecology	Q	F	2	17006	2
7	2	gynecology	Q	D	2	17006	2
8	4	gynecology	S	E	3	95946	0
9	4	gynecology	Q	F	3	95946	0
10	3	gynecology	Q	F	4	95946	0
11	4	gynecology	Q	E	2	95946	0
12	3	gynecology	R	F	3	95946	0
13	4	gynecology	R	A	3	95946	0
14	3	gynecology	R	F	3	95946	0

Prepared Data set after cleaning and removing NULL:



City_Code_Patient	RowId	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region_c
2	1	318439	21	c	3	Z
2	2	318440	29	a	4	X
2	3	318441	26	b	2	Y
2	4	318442	6	a	6	X
2	5	318443	28	b	11	X
2	6	318444	23	a	6	X
2	7	318445	26	b	2	Y
0	8	318446	25	e	1	X
0	9	318447	23	a	6	X
0	10	318448	23	a	6	X
0	11	318449	10	e	1	X
0	12	318450	4	a	4	X
0	13	318451	16	c	3	Z
0	14	318452	28	b	11	Y

Prepared Dataset Link:

https://us3.ca.analytics.ibm.com/bi/?perspective=ca-modeller&id=i20E3457BF9FB477BAD3DE281E84BA51B&objRef=i20E3457BF9FB477BAD3DE281E84BA51B&tid=3019265856_85b6ad91352840ecab632d306e927cec_sessionTemp