

PROJECT DEVELOPMENT PHASE

SPRINT DELIVERY -1

Team ID	PNT2022TMID21554
Project Name	Analytics for Hospital's Health-Care Data

Sprint 1 <ul style="list-style-type: none">➤ Collection of Data➤ Loading the data➤ Data Preparation and Visualization	Sprint2 <ul style="list-style-type: none">➤ Exploration of Dataset➤ Visualization of Dataset
Sprint3 <ul style="list-style-type: none">➤ Dashboard Creation➤ Prediction of Length of Stay	Sprint4 <ul style="list-style-type: none">➤ Monitoring the Model/System Accuracy➤ Report Creation

Goal of this project:

The goal is to accurately predict the Length of Stay for each patient on case by case basis so that the Hospitals can use this information for optimal resource allocation and better functioning. The length of stay is divided into 11 different classes ranging from 0-10 days to more than 100 days.

Sprint Delivery1:

Collection of Data:

Hospital Dataset Link:

https://drive.google.com/drive/folders/1qozDgfVw04hGS8Di_SFhUNdUKUd_jkhrT?usp=sharing

UNDERSTANDING THE DATASET:

This project is based on understanding the Hospital's Health Analytics dataset .

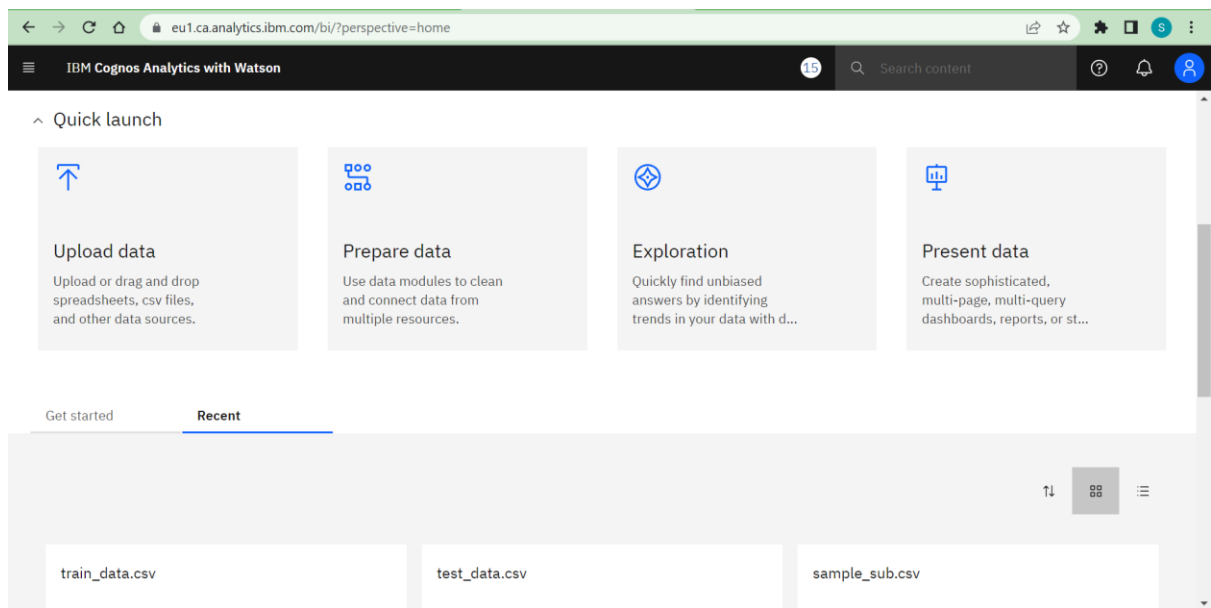
The data is spread across 3 data files (csv) and one data dictionary enclosed.

- 1.Test_data.csv
- 2.Sample_sub.csv
- 3.Train_data_dictionary.csv
- 4.Train_data.csv









The primary data file we use is test_data.csv consist of 17 Columns with 137057 Rows.

Loading the Data to IBM COGNOS ANALYTICS:

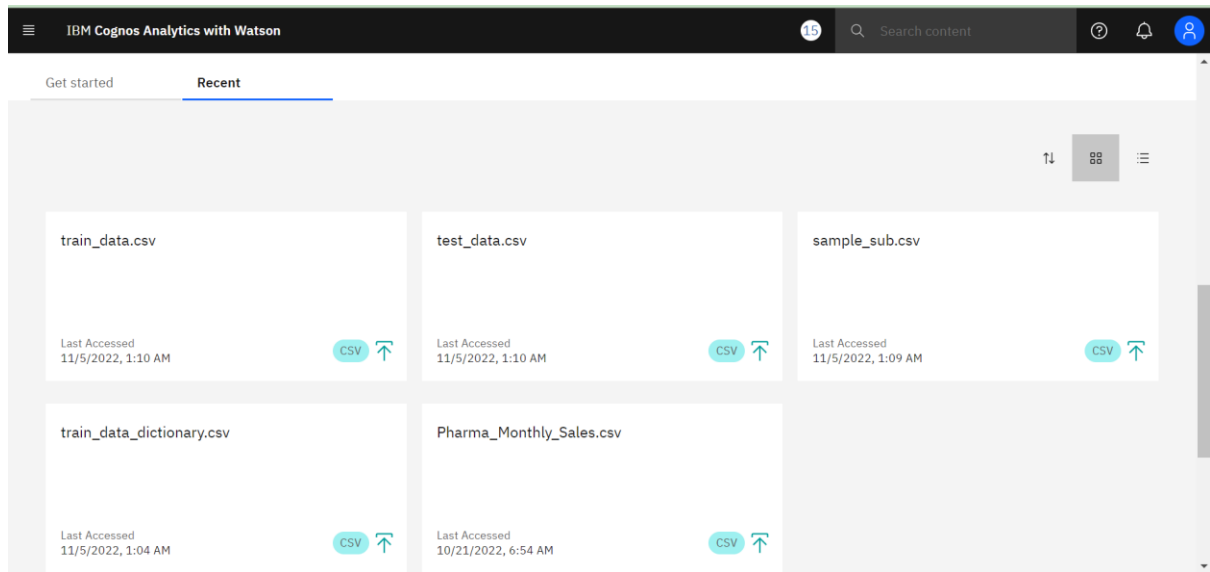
Log in to your IBM Account and click on Upload Data



Select the Data to be uploaded from the Local System

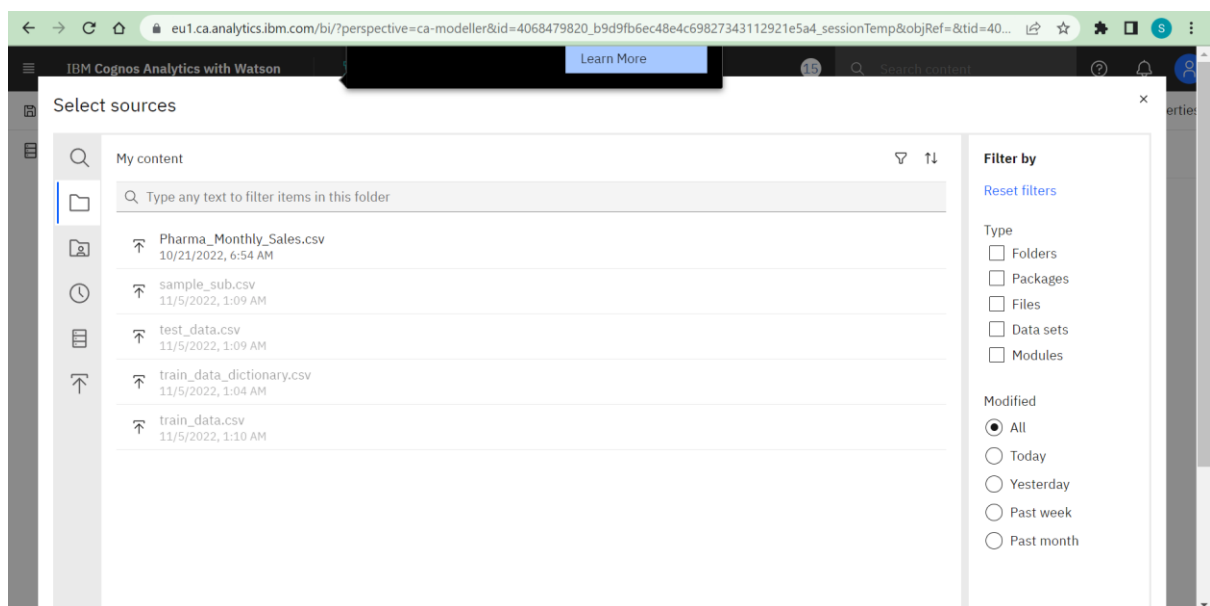
Name	Status	Date modified	Type	Size
 sample_sub		03-11-2022 22:25	Microsoft Excel Com...	1,607 KB
 test_data		03-11-2022 22:25	Microsoft Excel Com...	10,531 KB
 train_data		03-11-2022 22:25	Microsoft Excel Com...	26,285 KB
 train_data_dictionary		03-11-2022 22:25	Microsoft Excel Com...	1 KB

Then under Recent Tabs, The data will get uploaded



Data Preparation:

Select the sources/Files to be Prepared:



IBM Cognos Analytics with Watson

Learn More

15 Search content

Properties

Data module

Search

New data module

Navigation paths

train_data...ionary.csv

Row Id

abc Column

abc Description

train_data.csv

test_data.csv

sample_sub.csv

Row Id

1

2

3

Grid

Relationships

Custom tables

Row Id	Column	Description
1	case_id	Case_ID registered in Hospital
2	Hospital_code	Unique code for the Hospital
3	Hospital_type_code	Unique code for the type of Hospital
4	City_Code_Hospital	City Code of the Hospital
5	Hospital_region_code	Region Code of the Hospital
6	Available Extra Rooms in Hospital	Number of Extra rooms available in the Hospital
7	Department	Department overlooking the case
8	Ward_Type	Code for the Ward type
9	Ward_Facility_Code	Code for the Ward Facility
10	Bed Grade	Condition of Bed in the Ward

Data Cleaning:

IBM Cognos Analytics with Watson

Learn More

15 Search content

Properties

Data module

Search

patientid

City_Code_Patient

Type of Admission

Severity of Illness

Visitors ...h Patient

Age

Admission_Deposit

Stay

test_data.csv

sample_sub.csv

Row Id

1

2

Grid

Relationships

Custom tables

patientid	City_Code_Patient	Type of Admission	Severity of Illness	Visitors with Patient	Age
115513		Emergency	Minor	8	71
115513		Emergency	Minor	4	71
115513		Emergency	Minor	2	71
115513		Emergency	Minor	4	71
115513		Emergency	Minor	2	71
84932		Emergency	Moderate	2	31
84932		Emergency	Moderate	3	31
84932		Emergency	Moderate	3	31
84932	8	Emergency	Moderate	3	31
84932	8	Emergency	Moderate	6	31

Filter...

Create calculation...

Create data group...

Hide from users

Remove

Format data...

Clean...

Sort descending

Sort ascending

Properties

Cleaning bigger values in patient ID column:

The screenshot shows the IBM Cognos Analytics interface. On the left, the 'Data module' pane lists various data sources, including 'patientid', 'City_Code_Patient', and several CSV files. The main area displays a data grid with columns: 'patient', 'Ward_Type', 'Ward_Facility_Code', 'Bed Grade', 'patientid', and 'City_Code_Pa'. The 'patientid' column contains values like 31397. A modal dialog titled 'Clean - patientid' is open, showing options to replace values with NULL or 0. The 'Replace this value with NULL' option is selected, and the value '31397' is entered in the input field.

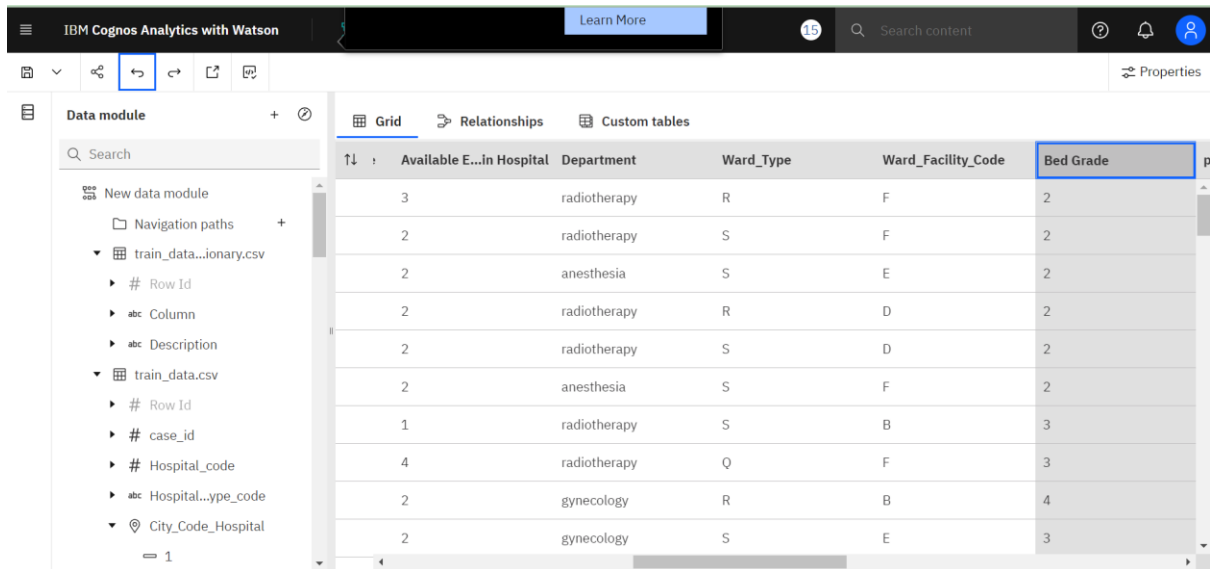
patient	Ward_Type	Ward_Facility_Code	Bed Grade	patientid	City_Code_Pa
rapy	R	F	2	31397	7
rapy	S	F	2	31397	7
sia	S	E	2	31397	7
rapy	R	D	2	31397	7
rapy	S	D	2	31397	7
sia	S	F	2	31397	7
rapy	S	B	3	31397	7
rapy	Q	F	3	31397	7
ogy	R	B	4	31397	7
ogy	S	E	3	31397	7

Data after cleaning:

The screenshot shows the IBM Cognos Analytics interface after cleaning the 'patientid' column. The 'patientid' column now contains 'Null' values for all rows. The data grid columns are: 'patient', 'Department', 'Ward_Type', 'Ward_Facility_Code', 'Bed Grade', 'patientid', and 'City_Code_Pa'. The 'patientid' column is highlighted in blue.

patient	Department	Ward_Type	Ward_Facility_Code	Bed Grade	patientid	City_Code_Pa
radiotherapy	radiotherapy	R	F	2	Null	7
radiotherapy	radiotherapy	S	F	2	Null	7
anesthesia	anesthesia	S	E	2	Null	7
radiotherapy	radiotherapy	R	D	2	Null	7
radiotherapy	radiotherapy	S	D	2	Null	7
anesthesia	anesthesia	S	F	2	Null	7
radiotherapy	radiotherapy	S	B	3	Null	7
radiotherapy	radiotherapy	Q	F	3	Null	7
gynecology	gynecology	R	B	4	Null	7
gynecology	gynecology	S	E	3	Null	7

Sort the Bed grade Column:



IBM Cognos Analytics with Watson

Learn More

15

Search content

Properties

Data module

Search

New data module

Navigation paths

train_data...ionary.csv

Row Id

abc Column

abc Description

train_data.csv

Row Id

case_id

Hospital_code

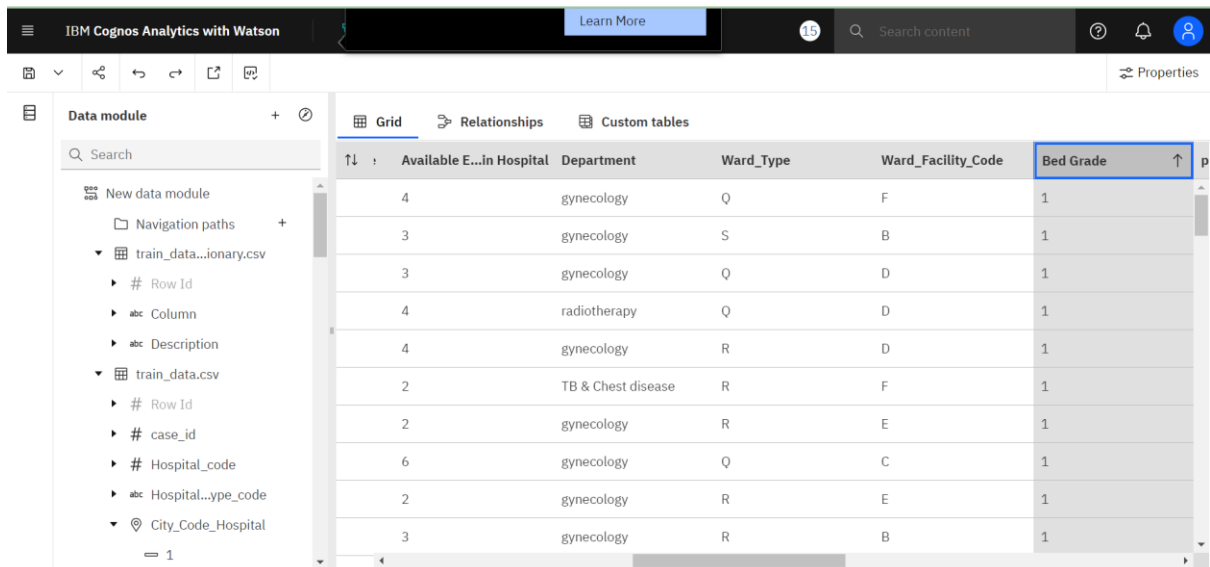
abc Hospital...ype_code

City_Code_Hospital

1

Available E...in Hospital	Department	Ward_Type	Ward_Facility_Code	Bed Grade
3	radiotherapy	R	F	2
2	radiotherapy	S	F	2
2	anesthesia	S	E	2
2	radiotherapy	R	D	2
2	radiotherapy	S	D	2
2	anesthesia	S	F	2
1	radiotherapy	S	B	3
4	radiotherapy	Q	F	3
2	gynecology	R	B	4
2	gynecology	S	E	3

After Sorting in Ascending Order:



IBM Cognos Analytics with Watson

Learn More

15

Search content

Properties

Data module

Search

New data module

Navigation paths

train_data...ionary.csv

Row Id

abc Column

abc Description

train_data.csv

Row Id

case_id

Hospital_code

abc Hospital...ype_code

City_Code_Hospital

1

Available E...in Hospital	Department	Ward_Type	Ward_Facility_Code	Bed Grade
4	gynecology	Q	F	1
3	gynecology	S	B	1
3	gynecology	Q	D	1
4	radiotherapy	Q	D	1
4	gynecology	R	D	1
2	TB & Chest disease	R	F	1
2	gynecology	R	E	1
6	gynecology	Q	C	1
2	gynecology	R	E	1
3	gynecology	R	B	1

IBM Cognos Analytics with Watson [Learn More](#)

Create a data group (numeric style)

Name:

Group names: Groups: Range border values:

Group names	Groups	Range border values
3 to < 4.0		4
2.8 to < 3		3
2.2 to < 2.8		2.3000000000000003
1.6 to < 2.2		2.2
less than 1.6		1.6
		Lower

Removing City_Code_Patient column as it contains some duplicate values

IBM Cognos Analytics with Watson

Learn More

15

Search content

Properties

Data module

Search

train_data.csv

abc Bed Grade (Group)

Row Id

case_id

Hospital_code

abc Hospital...ype_code

City_Code_Hospital

1

2

3

4

5

6

Grid

Relationships

Custom tables

ade

patientid

City_Code_Patient

Type of Admission

Severity of Illness

Visitors with

31397

7

Emergency

Extreme

2

31397

7

Trauma

Extreme

2

31397

7

Trauma

Extreme

2

31397

7

Trauma

Extreme

2

31397

7

Trauma

Extreme

2

31397

7

Trauma

Extreme

2

31397

7

Emergency

Extreme

2

31397

7

Trauma

Extreme

2

31397

7

Trauma

Extreme

2

31397

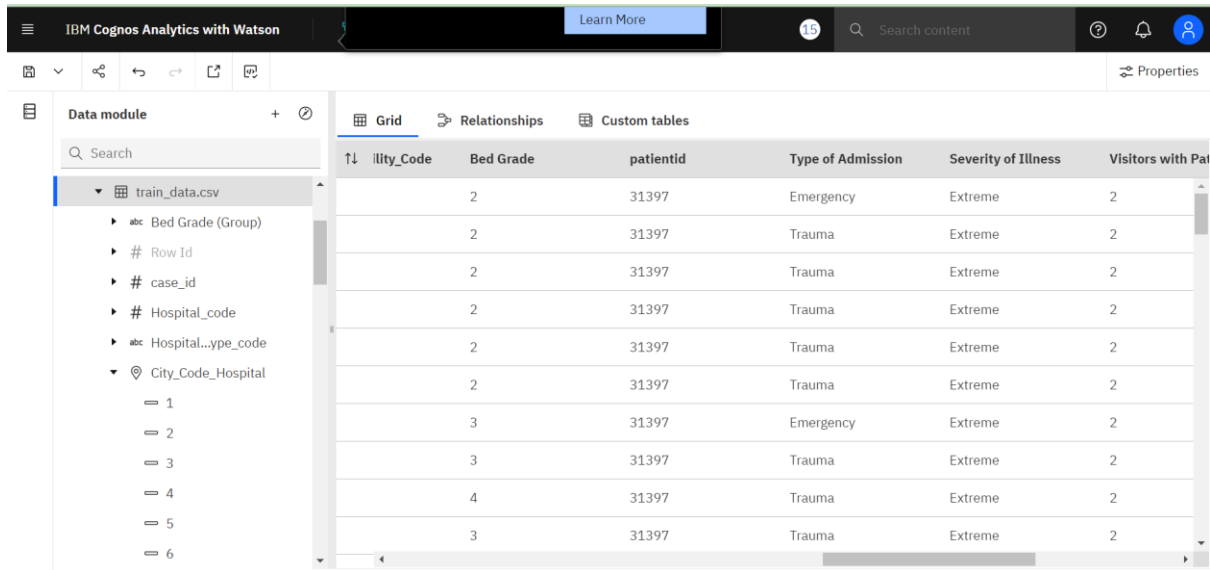
7

Trauma

Extreme

2

After Removing



IBM Cognos Analytics with Watson

Learn More

15 Search content

Properties

Data module

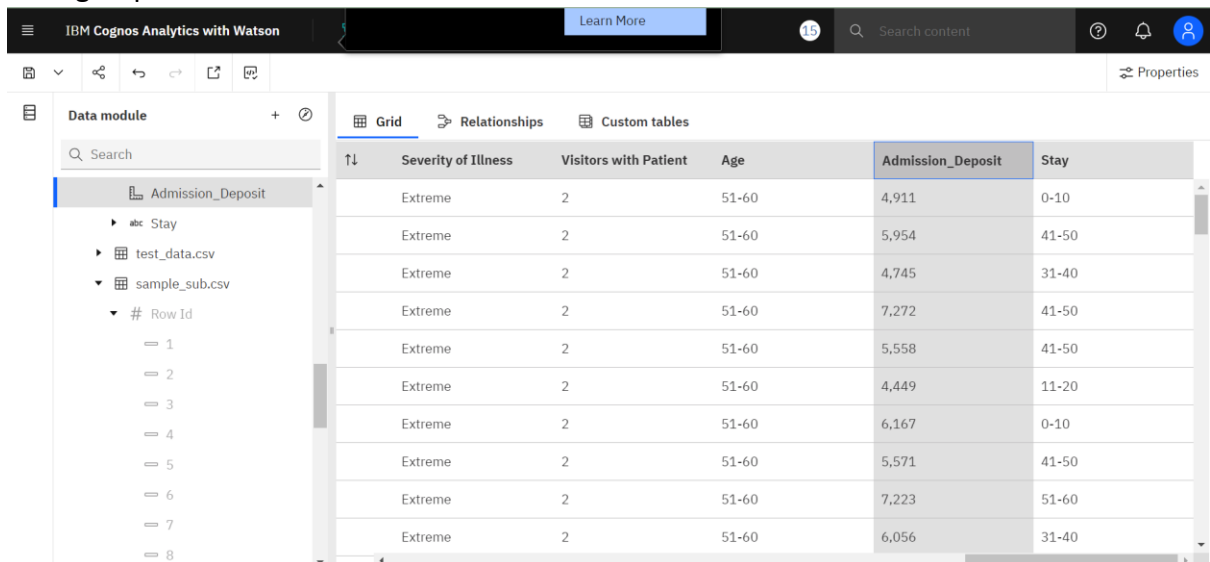
Search

train_data.csv

- Bed Grade (Group)
- Row Id
- case_id
- Hospital_code
- Hospital...ype_code
- City_Code_Hospital
 - 1
 - 2
 - 3
 - 4
 - 5
 - 6

ility_Code	Bed Grade	patientid	Type of Admission	Severity of Illness	Visitors with Pat
	2	31397	Emergency	Extreme	2
	2	31397	Trauma	Extreme	2
	2	31397	Trauma	Extreme	2
	2	31397	Trauma	Extreme	2
	2	31397	Trauma	Extreme	2
	2	31397	Trauma	Extreme	2
	3	31397	Emergency	Extreme	2
	3	31397	Trauma	Extreme	2
	4	31397	Trauma	Extreme	2
	3	31397	Trauma	Extreme	2

Hiding Deposit amount from users



IBM Cognos Analytics with Watson

Learn More

15 Search content

Properties

Data module

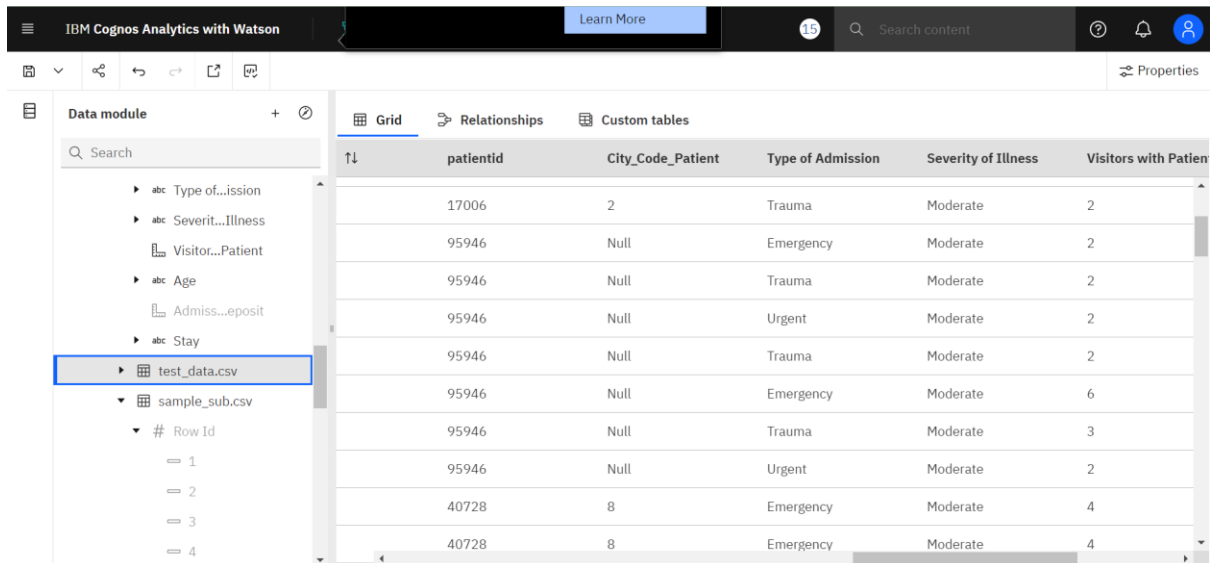
Search

Admission_Deposit

- Stay
- test_data.csv
- sample_sub.csv
- Row Id
 - 1
 - 2
 - 3
 - 4
 - 5
 - 6
 - 7
 - 8

Severity of Illness	Visitors with Patient	Age	Admission_Deposit	Stay
Extreme	2	51-60	4,911	0-10
Extreme	2	51-60	5,954	41-50
Extreme	2	51-60	4,745	31-40
Extreme	2	51-60	7,272	41-50
Extreme	2	51-60	5,558	41-50
Extreme	2	51-60	4,449	11-20
Extreme	2	51-60	6,167	0-10
Extreme	2	51-60	5,571	41-50
Extreme	2	51-60	7,223	51-60
Extreme	2	51-60	6,056	31-40

Remiving Null values in City_Code_Patient



IBM Cognos Analytics with Watson

Learn More

15

Search content

Properties

Data module

Grid Relationships Custom tables

Search

abc Type of...ission

abc Severit...Illness

Visitor...Patient

abc Age

Admiss...eposit

abc Stay

test_data.csv

sample_sub.csv

Row Id

1

2

3

4

patientid	City_Code_Patient	Type of Admission	Severity of Illness	Visitors with Patien
17006	2	Trauma	Moderate	2
95946	Null	Emergency	Moderate	2
95946	Null	Trauma	Moderate	2
95946	Null	Urgent	Moderate	2
95946	Null	Trauma	Moderate	2
95946	Null	Emergency	Moderate	6
95946	Null	Trauma	Moderate	3
95946	Null	Urgent	Moderate	2
40728	8	Emergency	Moderate	4
40728	8	Emergency	Moderate	4

Replacing null values with constant 1

Clean - City_Code_Patient

NULL values

- ☐ Replace this value with NULL 0
- ☒ Replace NULL values with 1

Cancel

Clean

IBM Cognos Analytics with Watson

Learn More

15

Search content

Grid

Relationships

Custom tables

City_Co...atient

abc Type of...ission

abc Severit...Illness

Visitor...Patient

abc Age

Admiss...eposit

sample_sub.csv

Row Id

1

2

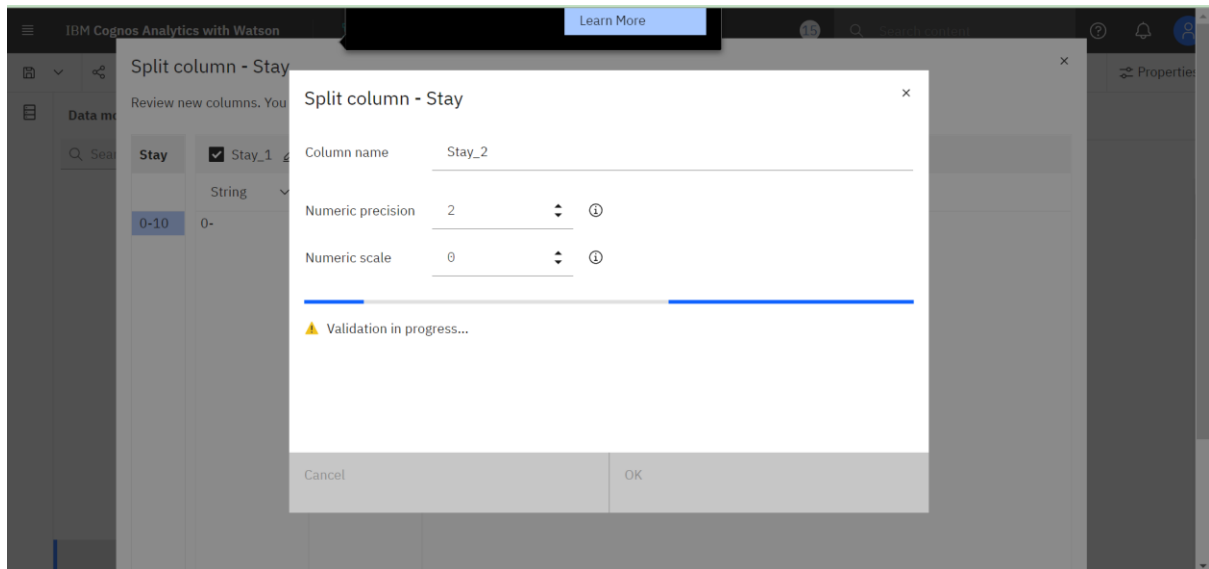
3

4

5

City_Code	Bed Grade	patientid	City_Code_Patient	Type of Admission	Severity of Illr
2	17000	2	Trauma	Moderate	
2	17006	2	Trauma	Moderate	
3	95946	1	Emergency	Moderate	
3	95946	1	Trauma	Moderate	
4	95946	1	Urgent	Moderate	
2	95946	1	Trauma	Moderate	
3	95946	1	Emergency	Moderate	
3	95946	1	Trauma	Moderate	
4	95946	1	Urgent	Moderate	
2	40728	8	Emergency	Moderate	

Splitting Column Stay



IBM Cognos Analytics with Watson

Learn More

15

Search content

Properties

Data module

Search

318443

318444

318445

318446

318447

318448

318449

318450

318451

318452

318453

Load more ...

abc Stay

Row Id	case_id	Stay	Stay_1	Stay_2
1	318439	0-10	0-	0
2	318440	0-10	0-	0
3	318441	0-10	0-	0
4	318442	0-10	0-	0
5	318443	0-10	0-	0
6	318444	0-10	0-	0
7	318445	0-10	0-	0
8	318446	0-10	0-	0
9	318447	0-10	0-	0
10	318448	0-10	0-	0

IBM Cognos Analytics with Watson

Learn More

15

Search content

Properties

Data module

Search

318443

318444

318445

318446

318447

318448

318449

318450

318451

318452

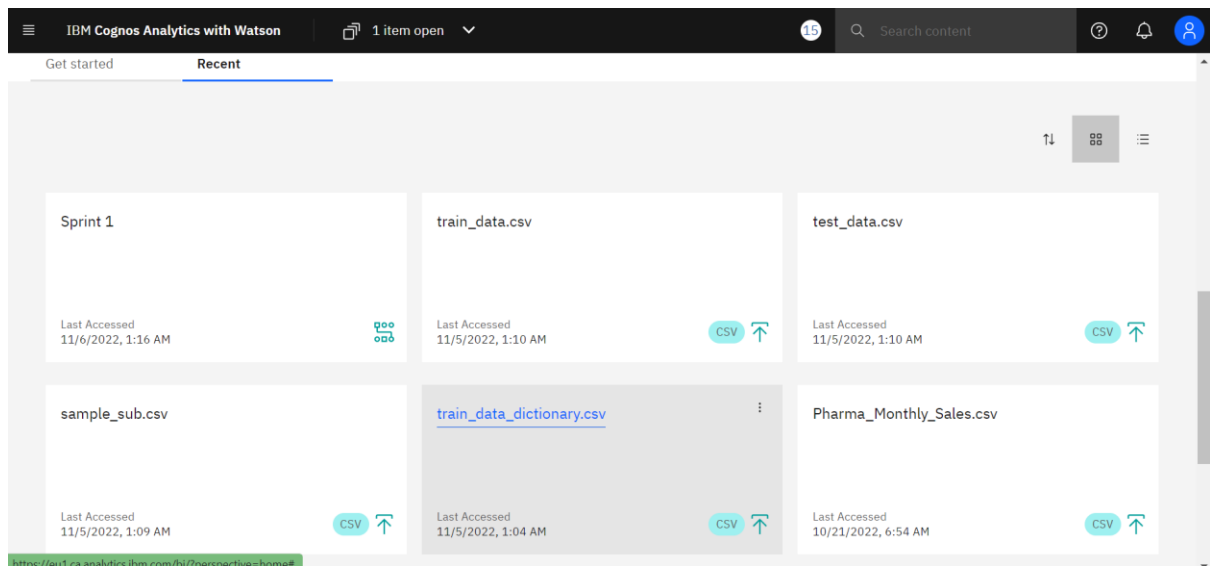
318453

Load more ...

abc Stay

Row Id	case_id	Stay
1	318439	0-10
2	318440	0-10
3	318441	0-10
4	318442	0-10
5	318443	0-10
6	318444	0-10
7	318445	0-10
8	318446	0-10
9	318447	0-10
10	318448	0-10

Prepared Data(SPRINT 1):



Prepared Dataset Link:

https://eu1.ca.analytics.ibm.com/bi/?perspective=ca-modeller&pathRef=.my_folders%2FSprint%2B1