

PROJECT DEVELOPMENT PHASE

DELIVERY OF SPRINT 1

TEAM ID	PNT2022TMID13523
DATE	29/10/2022
PROJECT NAME	Analytics for Hospital's Health Care Data

Project Development Phase :

Sprint 1 :

- Data Collection
- Data Preparation

Sprint 2 :

- Data Exploration

Sprint 3 :

- DashBoard Creation

Sprint 4:

- Report creation
 - Story creation
-

SPRINT 1 :

Data Collection : The data set used is shared in this drive link

https://drive.google.com/file/d/1sIC0MhsJHeuODVkhIXrdNOX_aBDySSBh/view

1. Load the dataset : Using IBM cognos analytics

The screenshot shows the IBM Cognos Analytics with Watson interface. On the left, there's a sidebar titled "Data module" with a search bar and a list of datasets: "New data module", "Navigation paths", "train_data.csv", and "test_data.csv". The "test_data.csv" item is highlighted with a blue selection box. The main area is a "Grid" view displaying a table with 14 rows of data. The columns are: Row Id, case_id, Hospital_code, Hospital_type_code, City_Code_Hospital, Hospital_region_code, Available_E...in Hospital, and Department. The data shows various hospital codes (a, b, c, d, e) and department names (gynecology). A watermark for "Activate Windows" is visible at the bottom right.

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

1. Save the datasets inside a **Data Module**:

<https://us3.ca.analytics.ibm.com/bi/?perspective=content>

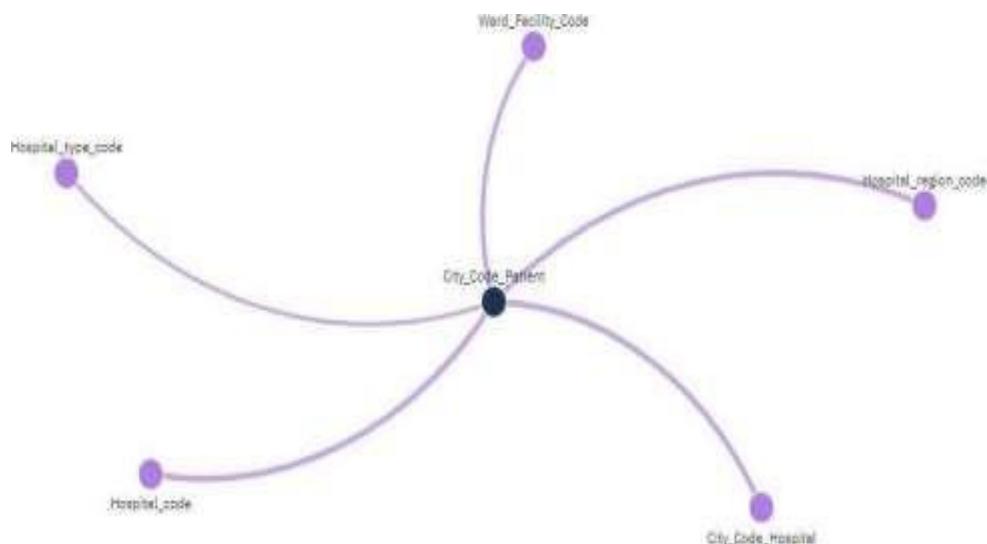
The screenshot shows a web browser window with the URL <https://us3.ca.analytics.ibm.com/bi/?perspective=content> in the address bar. The page title is "IBM Cognos Analytics with Watson". The main content area is titled "Data Module" and contains a single item named "Health care Dataset". Below the item name, it says "Last Accessed 14/11/2022, 11:46". There are several small icons to the right of the dataset name, including a magnifying glass, a gear, and a person icon. At the bottom right of the page, there is a message: "Activate Windows Go to Settings to activate Windows."

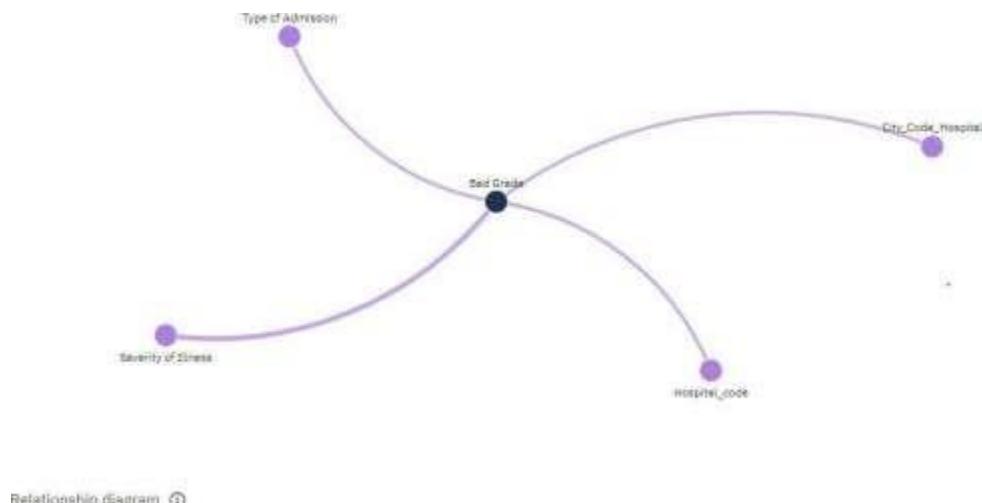
3. Data Preparation (includes data cleaning) :

Prepare the dataset in the aim to find right insights

Screenshot of IBM Cognos Analytics with Watson interface showing a data module for a "Health care Dataset". The data grid displays 18 rows of data with columns: Row Id, case_id, Hospital_code, Hospital_type_code, City_Code_Hospital, Hospital_region_code, Available_E...in Hospital, and Department. The "Department" column consistently shows "gynecology". The "Hospital_type_code" column shows values b, a, b, e, a, a, e, a, c, b, a, 11, 6, 2, 1, 6, 11, 7, 2, 7, 2. The "City_Code_Hospital" column shows values 11, X, Y, 1, X, X, X, Z, X, 1, X, 4, 3, 2, 4, 1, 5, 2, 3. The "Hospital_region_code" column shows values 2, 3, 2, 4, 4, 3, 4, 3, 4, 1, 3, 4, 3, 3, 1, 5, 2, 3. The "Available_E...in Hospital" column shows values 2, 3, 2, 4, 4, 3, 4, 3, 4, 1, 3, 4, 3, 3, 1, 5, 2, 3.

Row Id	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region_code	Available_E...in Hospital	Department
5	318443	28	b	11	X	2	gynecology
6	318444	23	a	6	X	3	gynecology
7	318445	26	b	2	Y	2	gynecology
8	318446	25	e	1	X	4	gynecology
9	318447	23	a	6	X	4	gynecology
10	318448	23	a	6	X	3	gynecology
11	318449	10	e	1	X	4	gynecology
12	318450	4	a	4	X	3	gynecology
13	318451	16	c	3	Z	4	gynecology
14	318452	28	b	11	X	3	gynecology
15	318453	19	a	7	Y	1	gynecology
16	318454	26	b	2	Y	5	gynecology
17	318455	19	a	7	Y	2	gynecology
18	318456	26	b	2	Y	3	gynecology





Relationship diagram ①

1. Null values cleaning process :

Edit calculation

Name: City_Code_Patient

Components	Expression
<ul style="list-style-type: none"> main_data.csv # case_id # Hospital_code ** Hospital_code ?? City_Co...spital ?? Hospital_code Available_Hospital ** Department ** Ward_Type ** Ward_...Code 	<pre> 1. IF(City_Code_Patient is missing) 2. then(median [City_Code_Patient]) 3. else(City_Code_Patient) </pre>
	Validation Results The expression is valid.

Prepared Data Set Link : https://us3.ca.analytics.ibm.com/bi/?perspective=ca-modeller&id=iD4D82396B2CB4F4DA139074AFF61A2D9&objRef=iD4D82396B2CB4F4DA139074AFF61A2D9&tid=3019265856_7786e234de564f078541990a392c4722_sessionTemp