ANALYTICS FOR HOSPITAL AND HEALTH-CARE DATA

A PROJECT REPORT

Submitted by

AARTHICKRAJA A.P (RollNo: 922119104001)

MENAKA.C (RollNo: 922119104023)

PRADEESHYUVAN .P (Roll No: 922119104032)

VINO JOEL.R (Roll No: 922119104049)

TEAM ID : PNT2022TMID33111

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

SSM INSTITUTE OF ENGINEERING AND TECHNOLOGY

DINDIGUL

ABSTRACT

ABSTRACT

This paper is mainly streamed towards hospitals and their health-care data. Due to recent covid-19 pandemic has raised alarms over one of the most overlooked areas to focus. Healthcare management has various use cases for data science, patients health details and therir past history with data records. Analyizing the data's with a module and exploring the visualization can improve the dataset. In order to discuss health data analytics and the role it plays in the health care sector, we must first understand the data that is being collected and analyzed. There is data being collected on the processes and procedures of the business side of health care, but there is also an enormous amount of health data being gathered, stored and analyzed. Health data is any data relating to the health of an individual patient or collective population. This information is gathered from a series of health information

systems (HIS) and other technological tools utilized by health care professionals, insurance companies and government organizations. Consider the impact this has had on the COVID-19 pandemic. The data being collected is analyzed in real time to understand the effects of the virus better and predict future trends so we may slow the spread and prevent future outbreaks. Health care data management has the potential to lead to better care if used properly. With centralized datasets, there is immediate access to necessary information whenever and wherever it is needed. The addition of big data analytics improves efficiency on all fronts. Better data leads to better care.

TABLE OF CONTENTS

TABLE OF CONTENTS

INTRODUCTION

CHAPTER 1

INTRODUCTION

The introduction about the analytics for hospital and health-care data with IBM-Cognos and analytics.

1.1 PROJECT OVERVIEW

Recent Covid-19 Pandemic has raised alarms over one of the most overlooked areas to focus on Healthcare. While Healthcare management has various use cases for using data science, patient length of stay is on critical parameter to observe and predict if one wants to improve the efficiency of the healthcare management in a hospital. This parameter helps

hospitals to identify patients of high LOS-risk at the time of admission. Once identified, patients with high LOS risk can have their treatment plan optimized to minimize LOS and lower the chance of staff/visitor infection. Also, prior knowledge of LOS can aid in logistics such as room and bed allocation planning. Suppose you have been hired as Data Scientist of Health Man a not for profit organization dedicated to manage the functioning of Hospitals in a professional and optimal manner. The goal is to accurately predict the length of stay of 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.

1.2 PURPOSE

a. This type of analysis is used to investigate why an event happened.

b. This form of analysis is used to forecast something that will happen in the future. For example, a hospital might predict, based on trends observed over the past decade, that incoming cardiac patients will most likely increase by 20% this year.

c. This is possibly the most important form of analysis in healthcare and the trend that is growing quickest. This form of analysis takes pre-existing data and implements treatment plans. For example, a healthcare provider might use a smart device to automatically analyze a patient's vital signs, preemptively alert them that they're at risk for developing a medical condition, and instruct them to visit their healthcare provider.

d. While healthcare data analytics is highly advantageous, it can get pretty complicated, too. Whether the data was collected by assessing important real-time signs or through electronic health records (EHR), it needs to be derived from various sources by following proper government regulations, thus making the process precarious and complex.

e. Anything from clinical data to patient behaviour, medical expenses, healthcare, or pharmaceuticals data analytics can be employed at the micro and macro level to evidently enhance operations, boost patient care, and even tackle the overall expenses.

Fig 1.1 Analytics for healthcare

LITERATURE SURVEY

CHAPTER 2

LITERATURE SURVEY

The introduction about the literature survey gone throughfor the projectare briefly discussedin this chapter.

2.1 EXISTING PROBLEM

As we all know health care organisation will understand of big data analytics. Data mining offers novel information regarding health care helpful for making administrative as well as prediction disease, selection of treatment, health isurance policy. The novel corona virus pandemic outbreak is seriously threatening human health. Security optimization implementaion and testing on real world patients Hospitalization cost and the insured population all show a trend of increasing year by year. The users to help to see understand the valuable information provided by data care visual analytics huge amount of structured and unstructured and semi structured data have been generated by various institutions around the world.

This research demonstrates to address lack, this study examines the historical development, architectural design and component functionalitites of big data analytics. Data sets can gain unwanted attendion from hackers and important information can leaked to competitors. As each and every patient records are important to the hospital organisatoin and the data should be protected with secutiry measures. The health industry sector had been confronted bu the need to manage the big data being produced by various sources, which are well known for producing high volumes of heterogeneous data lack of standardisation methods and electronic tools. In recent years, there has been much research in medical big data, mainly targeting data collecting data, data analysis and visualisation.

2.2 REFERENCES

[1] Big data analytics for healthcare industry.

Authors: R Sunil Kumar, A Daniel
Published in: 2015 IEEE.

The user to help able to information provided by the healthcare data in any need with the help of system the is developed using the AI with chats analysis and visualization. Analysing the dats with the better understand and current practices, capabilities and challenges related to clinical data analytics. Multiview data analytics requires advanced machine learning techniques.

Merits: The data analysis will help to the hospital organisation with the structured data access at anytime through systems.

Demerits: Data should be maintained and protected as leakage of data could lead to unwanted issues to the organisation in privacy concerns.

[2] Intelligent and health care management.

Authors: Yinchuan Wang, .et.al, Zhihan Lv, .et.al,

Published in: 2017 IEEE.

Hospitalization cost and the insured and polulation all show a trend of increasing year by year. Data analytics become a future escalting tool of all industries including medicine, robotics, etc.,

Merits: As the population increases the productivity also increases so a system can manage this work loads.
Demerits: There is not mandatory to set a employee for data handling as system does it with AI build intelligence.

[3] IoT ENABLED SMART HEALTHCARE SYSTEM.

Authors: Syed Rooh Ullah, Divya Tomer, Imran Ahmed.

Published in: 2021 IEEE.

IoT enabled devices realtime, interoperability with prediction of disease real data set focusing on different pandemic symptoms. Unable to predict the data with the right chart and visualaization.

Merits: Data prediction will be viewed in the right visualization and chart with different kind of analysis.

Demerits: Accurate data is needed and analysis the pattern.

[4] A FRAMEWORK FOR DATA ANALYTICS BASED ON SYSTEMS.

Author: Alejandro Bal dominos .et.al.,

Published in: 2014 IEEE.

The visual analytics pattern is undertaken and the framework is designed based on the system with the data care information. The user to help the provided by healthcare and information is collected from the hospital organisation. Lack of oraganisational alignment and strategy for data analyticsad strandardized methods and report formation.

Merits: Data accessing is flexible with the designed framework.

Demerits: Maintainance of data fomat should be structured and grouped.

2.3 PROBLEM STATEMENT DEFINITION

There are multiple problems in the hospital health care organisation as some of the them are mentioned and solved though analysis of the various problem statements.

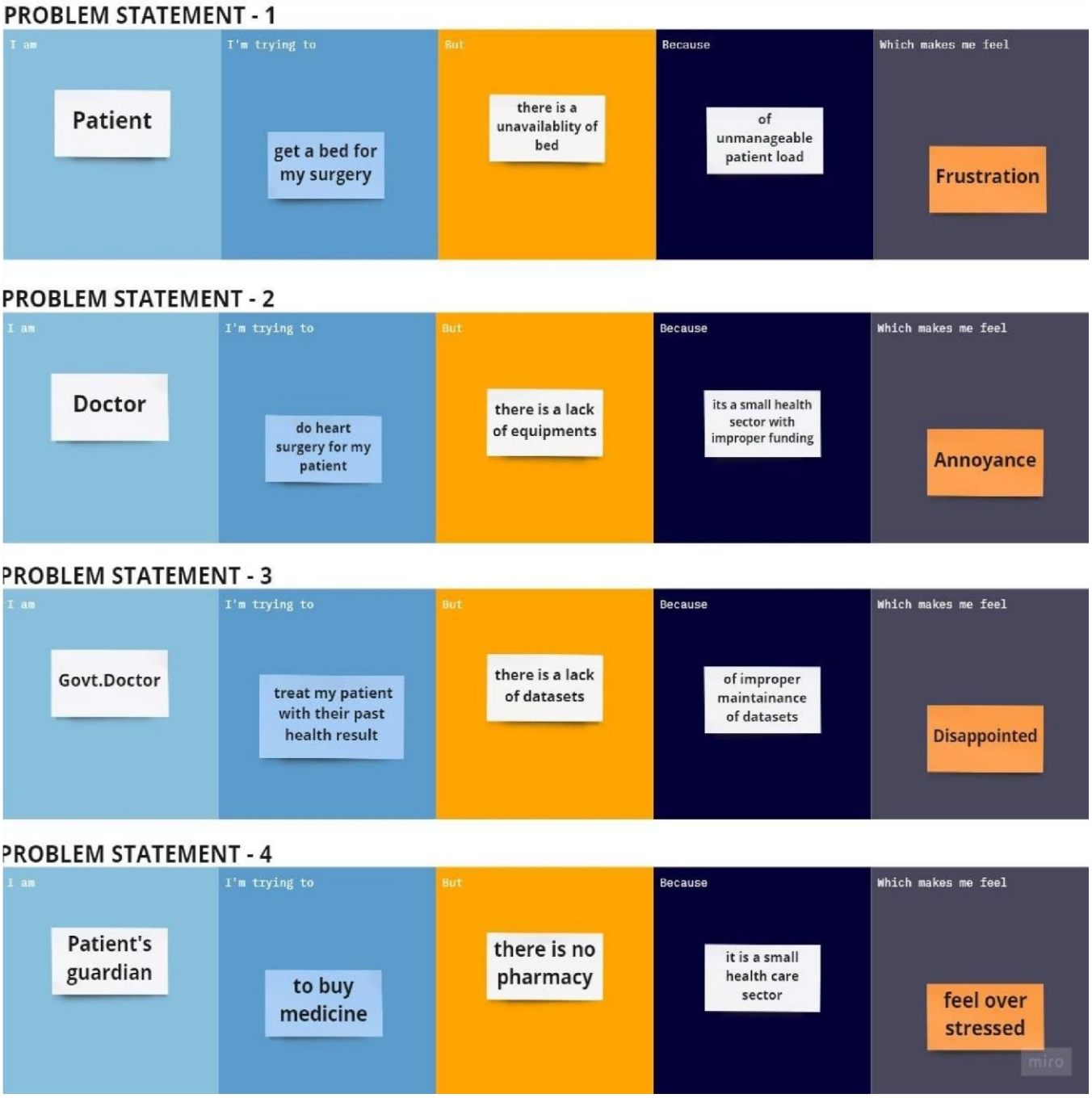


Fig 2.1 ProblemStatement Definition

IDEATION & PROPOSED SOLUTION

CHAPTER 3

IDEATION & PROPOSEDSOLUTION

3.1 EMPATHY MAP CANVAS

An empathy map is a simple,easy –to-digest visual that captures knowledge about a user's behaviors and attitudes. It is a useful tool to helps teams better understand their users..

Creating an effective solution requires understanding the true problem and the person who is experiencing it.The exercise of creating the map helps participants consider thingsfrom the user'sperspective along with his or her goals and challenge.

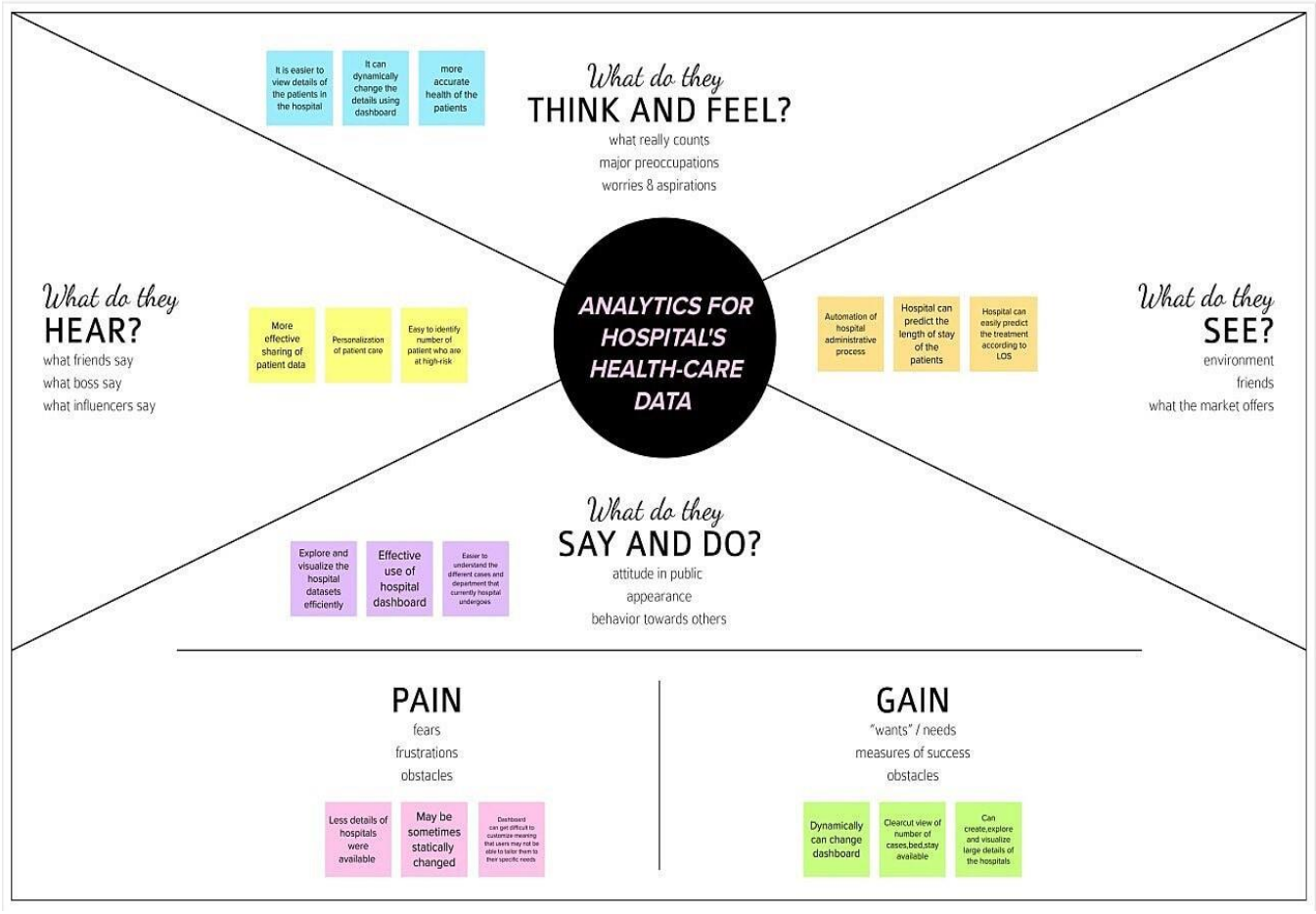


Fig 3.1 Empathy Map Canvas

3.2 IDEATION & BRAINSTORMING

Brainstorming provides a free and open environment that encourages everyonewithin a team to participate in the creativethinking process that leads to problem solving. Prioritizing volume over value, out-of-the-box ideas are welcome and built upon, and all participants are encouraged to

collaborate,helping each other develop a rich amount of creative solutions.

Template

Brainstorm & idea prioritization

Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room.

- **10 minutes** to prepare
- **1 hour** to collaborate
- **2-8 people** recommended

Share template feedback

Need some inspiration?

See a finished version of this template to kickstart your work.

Open example →

Before you collaborate

A little bit of preparation goes a long way with this session. Here's what you need to do to get going.

10 minutes

A Team gathering
Define who should participate in the session and send an invite. Share relevant information or pre-work ahead.

B Set the goal
Think about the problem you'll be focusing on solving in the brainstorming session.

C Learn how to use the facilitation tools
Use the Facilitation Superpowers to run a happy and productive session.

Open article →

1 Define your problem statement

What problem are you trying to solve? Frame your problem as a How Might We statement. This will be the focus of your brainstorm.

5 minutes

PROBLEM

How might we [your problem statement]?

Key rules of brainstorming

To run an smooth and productive session

- Stay in topic.
- Defer judgment.
- Go for volume.
- Encourage wild ideas.
- Listen to others.
- If possible, be visual.

Fig 3.2 Brainstorming 1

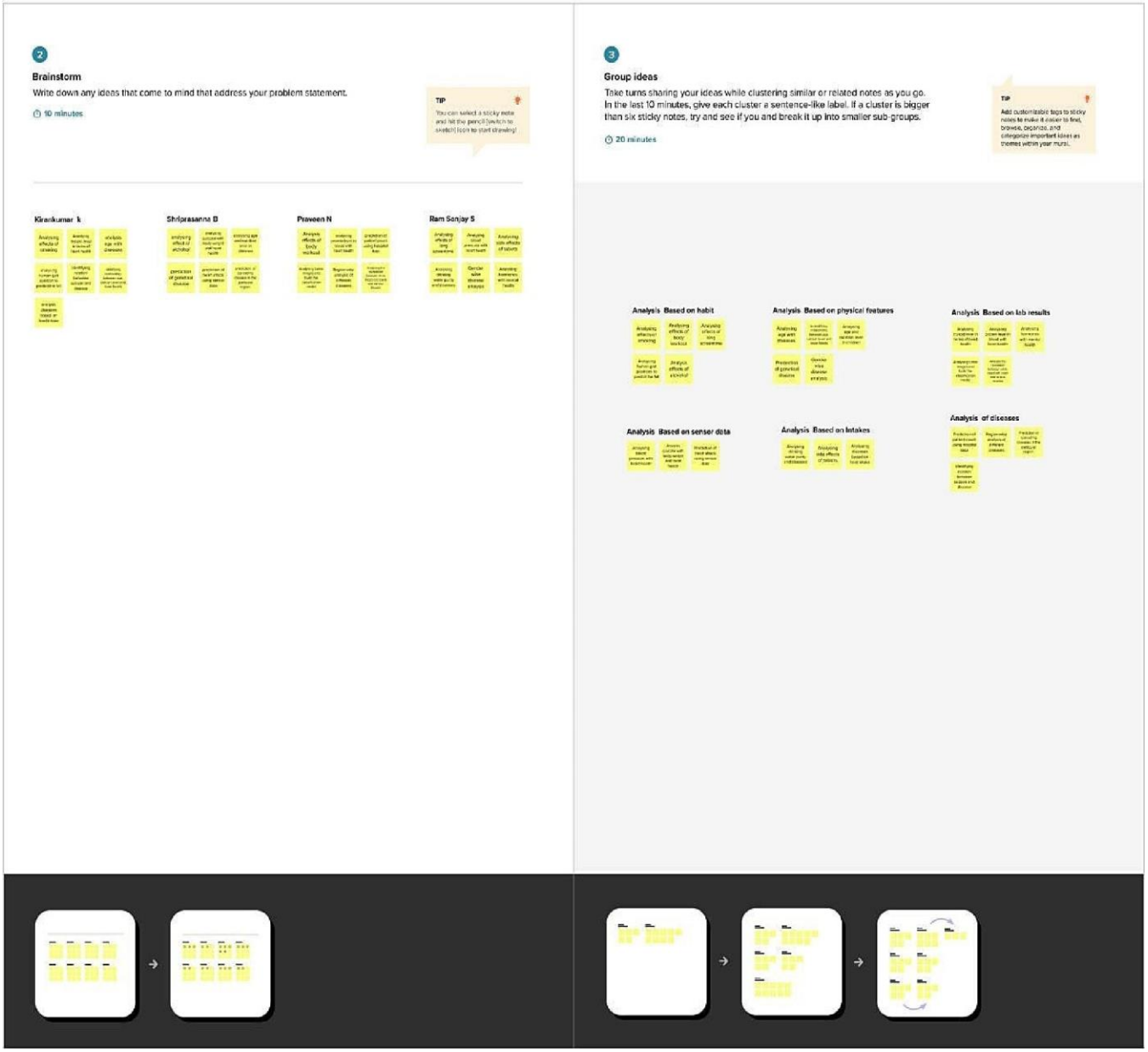


Fig 3.3 Brainstorming 2

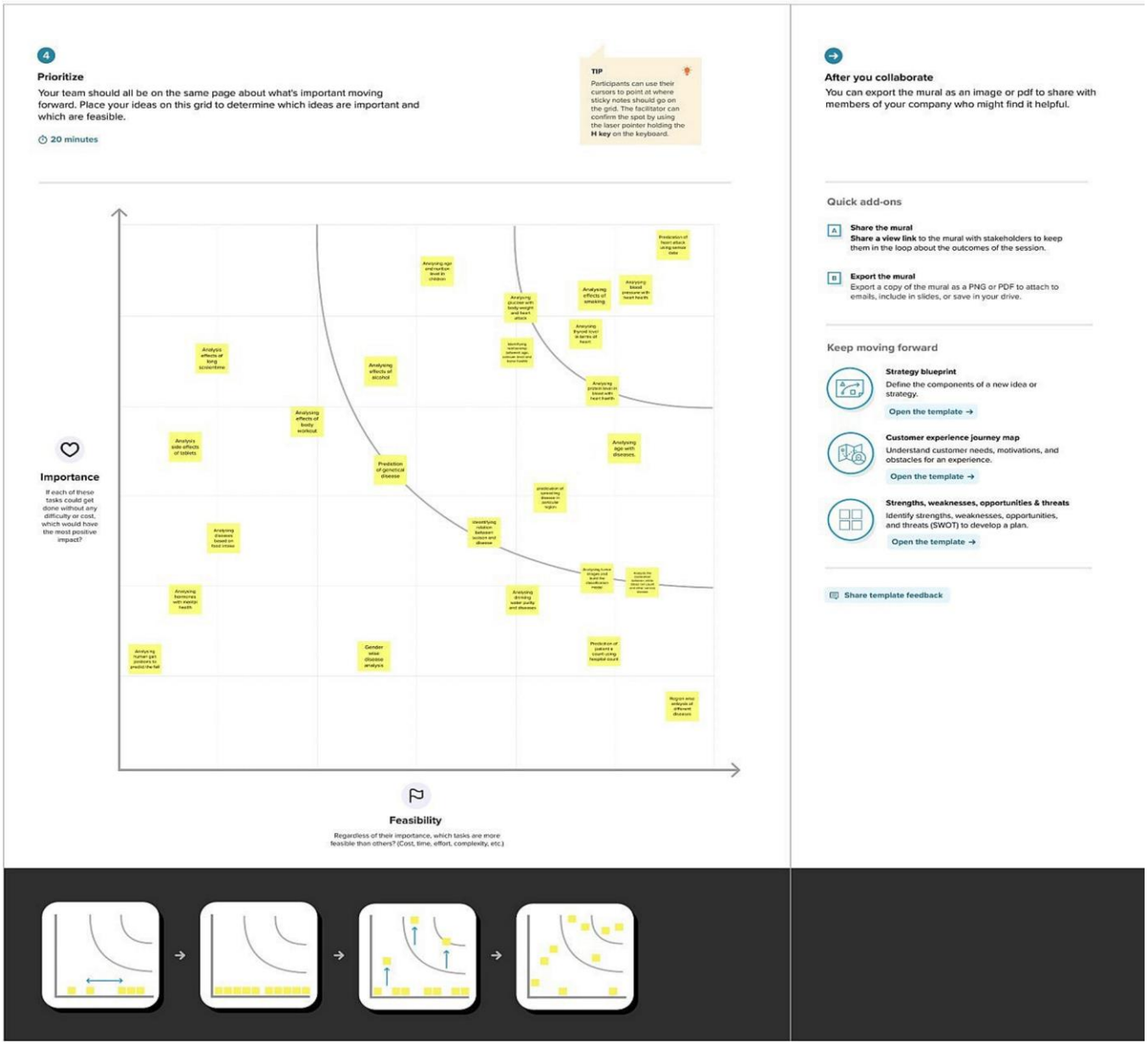


Fig 3.4 Brainstorming 3

3.3 PROPOSED SOLUTION

S.No.	Parameter	Descrip on
1.	Problem Statement (Problem to be solved)	EHR data matched pa ent-reported data in 23.5 percent of records in a study at an ophthalmology prac se.Pa ents' EHR data did not agree in any way when they reported having three or more eye health complaints.
2.	Idea / Solu on descrip on	Predic ve analy cs can create pa ent journey dashboards and disease trajectories that can lead to effec ve, and resultdriven healthcare. It improves treatment delivery, cuts costs, improves efficiencies, and so on.
3.	Novelty / Uniqueness	Healthcare data frequently resides in several loca ons. from various departments, such as radiology or pharmacy, to various source systems, such as EMRs or HR so ware. The organisa on as a whole contributes to the data. This data becomes accessible and usable when it is combined into a single, central system, such as an enterprise data warehouse (EDW).
4.	Social Impact / Customer Sa sfac on	Enhanced Improved diagnosis medical treatment Improved health results Improved rela onships with pa ents More posi ve health indicators
5.	Business Model (Revenue Model)	The two factors that have the biggest nega ve effects on hospital income are claim denials and pa ent incapacity to pay their part. 90% more uncollec ble claim denials were wri en off by hospitals and healthcare systems in 2017 compared to the preceding six years.
6.	Scalability of the Solu on	A variety of ins tu ons must store, evaluate, and take ac on on the massive amounts of data being produced by the health care sector as it expands quickly. India is a vast, culturally varied na on with a sizable popula onthat is increasingly able to access centralised healthcare services.

PROBLEM SOLUTION FIT

1. Customer Segments

+ Hospital Management

+ Patients

6. C ustomer L imita on

Can't assure the effective utilization and allocation of resources

5. Available Solu on Text mining

Information retrieval

2. Problems/Pains

1. Proper allocation of resources
2. Predicting the length of stay of COVID patients
3. Proper utilization and treatment to patients

9. Problem

Efficient less calculation and prediction of occurring situations

7. Behaviour

Data tracking with available methodologies such as text mining and information retrieval

3. Triggers to Act Prevaling emergency situations and Pandemic period

10. Your Solu on

Using predictive analysis powered by the Artificial intelligence which is used in analytics technique

8. Channels of Behaviour

1. Online: Usage of data exploration

4. Emo ons

Tensed and perplexed mind set to get rectified from the pandemic period		**2. Offline:** Preparing the dataset on the COVID patients.

Fig 3.5 Problem Solution Fit

REQUIREMENT ANALYSIS

CHAPTER 4

REQUIREMENT ANALYSIS

In this chapter,the requirement analysisof the proposed system has been discussed along with the brief explanation about its advantages.

4.1 FUNTIONAL REQUIREMENT

Following are the functional requirements of the proposed solution.

FR No.	Func onal Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	Collect data	Data from various sources are collected using different methods in order to provide op mized results.
FR-2	Data Cleaning and Wrangling	When combining mul ple data sources, there are many opportuni es for data to be duplicated or mislabeled hence we cleanse the data
FR-3	Crea ng data model	The process of analyzing and defining all the data, as well as the rela onships between those bits of data comes under this.
FR-4	Predic on and Analysis	The hidden trends are analyzed and the final results are predicted using machine learning and AI algorithms.

4.2 NON-FUNCTIONAL REQUIREMENT

Following are the non-functional requirements of the proposedsolution.

FR No.	Non-Func onal Requirement	Descrip on

NFR-1	**Usability**	The project must be easy to use. The user needs to have a good experience while working with the interface.
NFR-2	**Security**	Every user can access the website only if they posses the password. The database is secured with encryp on techniques which provides high levels of security
NFR-3	**Reliability**	The project must have minimal degree of failure under normal usage and how o en does the user get access to this work
NFR-4	**Performance**	The project must respond quickly to the user's ac ons or even if the user has to wait the wai ng period must be short.
NFR-5	**Availability**	The project is pla orm independent. It runs perfectly on almost every pla orm.
NFR-6	**Scalability**	The project allows mul ple users to handle the data at the same me. It is highly scalable since adding features and making advancements in the website is uncomplicated.

This chapter dealt with the funtional and non-functional requirement analysis of proposed system.

PROJECT DESIGN

CHAPTER 5

PROJECT DESIGN

5.1 DATA FLOW DIAGRAMS

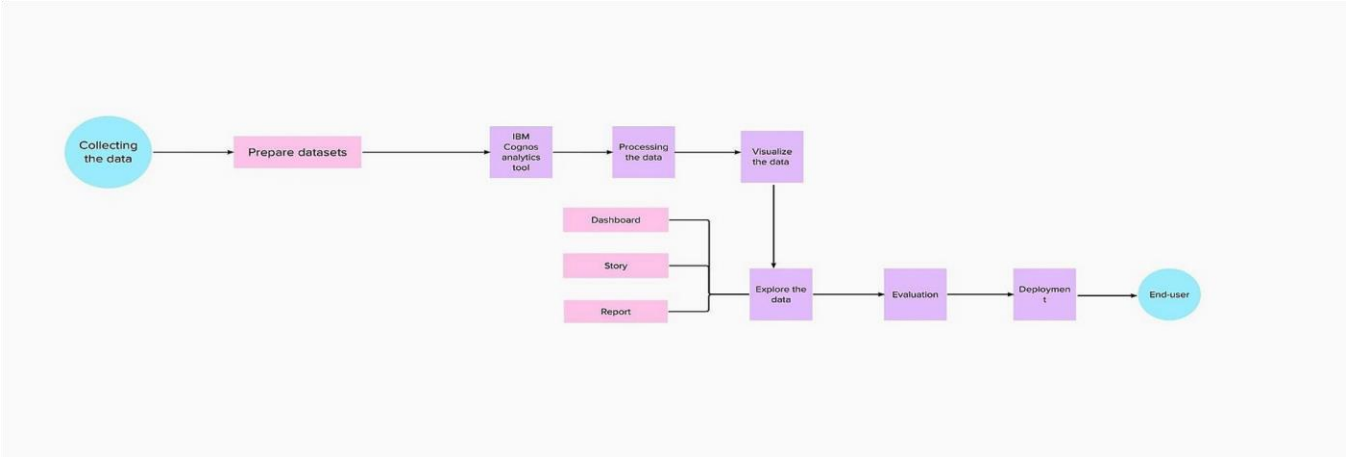


Fig 5.1 Dataflow Diagram

5.2 SOLUTION & TECHNICAL ARCHITECTURE

5.2.1 SOLUTION ARCHITECTURE

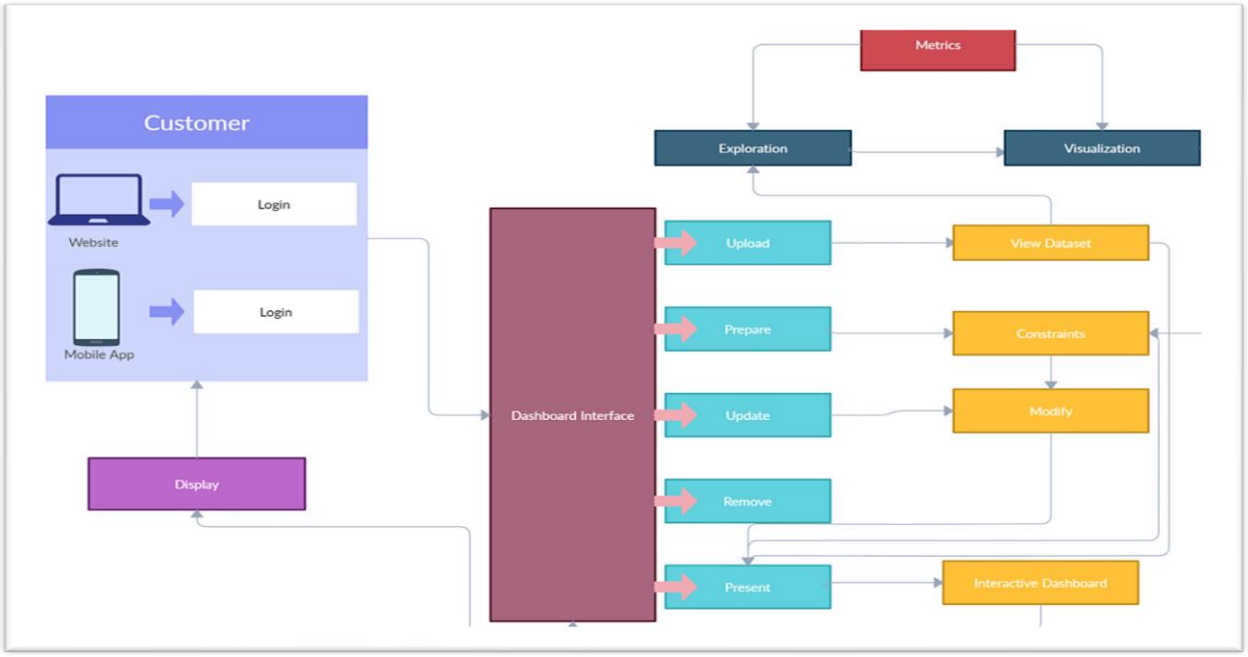


Fig 5.2 Solution Architecture Diagram

5.2.2 TECHNICAL ARCHITECTURE

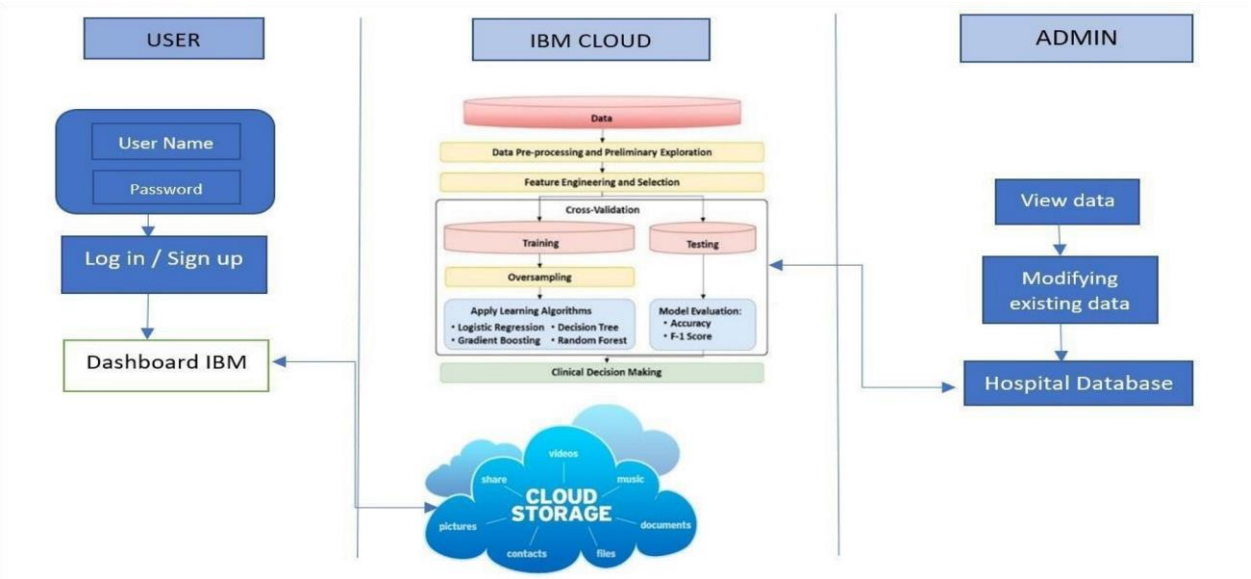


Fig 5.3 Technical Architecture Diagram

5.3 USER STORIES

User Type	Functional Requirement (Epic)	User Story Number		User Story / Task	Acceptance criteria	Priority	Release
Customer (Web user)		USN-1		As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2		As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3		As a user, I can register for the application through Gmail	I can register & access the dashboard	Medium	Sprint-1

	Login	USN-4		As a user, I can log into the application by entering email & password	I can access the dashboard	High	Sprint-1
	Dashboard	USN-5		As a user, I can upload the datasets to the dashboard	I can access various operations	High	Sprint-1
	View	USN-6		As a user, I can view the patient details	I can view the visual dataand the result after the prediction	High	Sprint-2
Admin	Analyse	USN-7		As an admin, I will analyse the given dataset	I can analyse the dataset	High	Sprint-2
	Predict	USN-8		As an admin, I will predict the length of stay	I can predict the length of stay	High	Sprint-2

Table 5.1 User Stories

PROJECT PLANNING & SCHEDULING

CHAPTER 6

PROJECT PLANNING & SCHEDULING

6.1 SPRINT PLANNING & ESTIMATION

MILESTONES	ACTIVITY LIST
MILESTONE-1	Collecting the data based on the application
MILESTONE-2	Uploading the collected data on the IBM COGNOS platform
MILESTONE-3	Data exploration in the IBM COGNOS platform
MILESTONE-4	Data visualization in the IBM COGNOS platform
MILESTONE-5	Creating an interactive dashboard
MILESTONE-6	Displaying the prepared dashboard
MILESTONE-7	Preparing a standard dataset and removing the unwanted data using the python programming
MILESTONE-8	By using the various algorithm and exploring the result and getting the accurate result with the help of an algorithm which give more accuracy
MILESTONE-9	Displaying the result according to the required format for example displaying the Length Of Stay of a patient

MILESTONE-10	Deployed in the GitHub

Table 6.1 Sprint Planning and Estimation

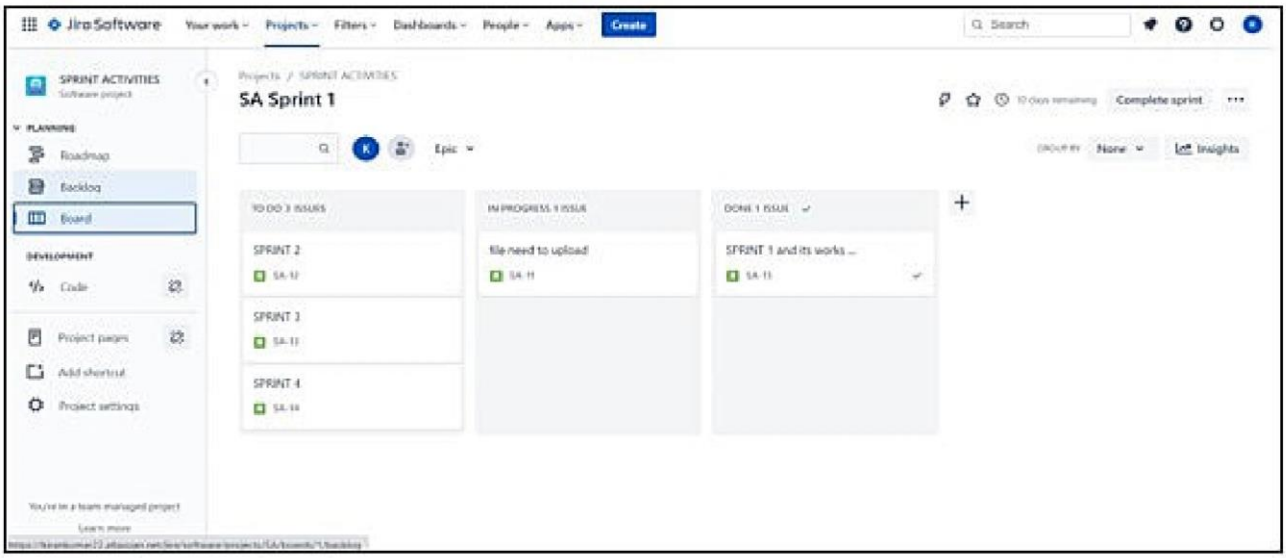
6.2 SPRINT DELIVERY SCHEDULE

Sprint	**Functional Requirement (Epic)**	**User Story Number**	**User Story / Task**	**Story Points**	**Priority**	**Team Members**
Sprint-1	Registration	USN-1	As a health care provider I can create account in IBM cloud and the data arecollected.	20	High	2 Members
Sprint-2	Analyze	USN-2	As a health care provider all the data thatare collected is cleaned and uploaded in the database or IBM cloud.	20	Medium	2 Members
Sprint-3	Dashboard	USN-3	As a health care provider I can use my account in my dashboard for uploading dataset.	10	Medium	2 Members
Sprint-3	Visualization	USN-4	As a health care provider I can prepare data for Visualization.	10	High	2 Members
Sprint-4	Visualization	USN-5	As a health care provider I canpresent data in my dashboard.	10	High	2 Members

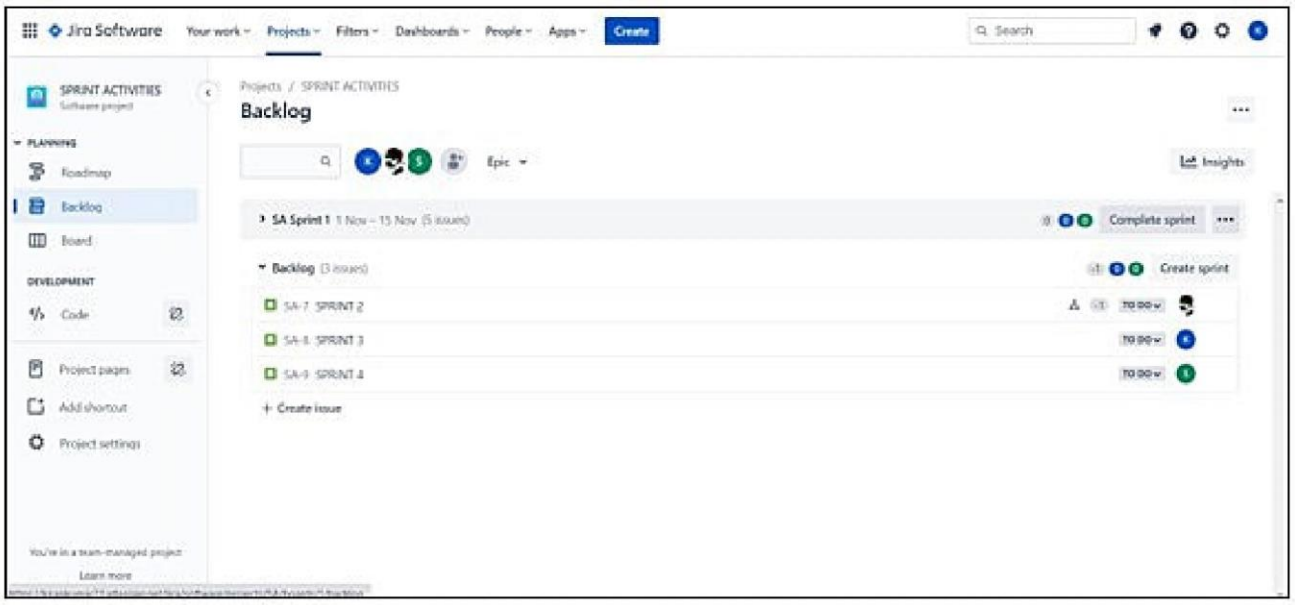
Table 6.2 Sprint Delivery Schedule

6.3 REPORTS FROM JIRA

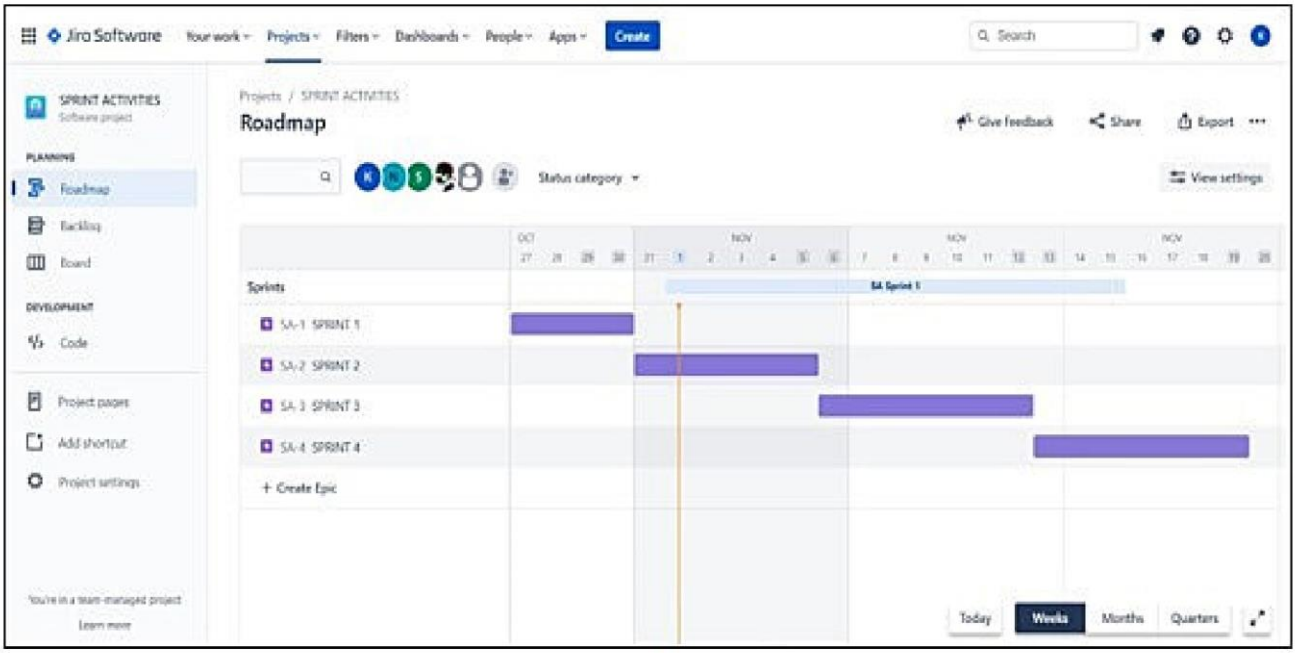
JIRA BOARD



BACKLOGS



ROADMAP


Jira Software
Your work
Projects
Filters
Dashboards
People
Apps
Create
Search
SPRINT ACTIVITIES
Software project
PLANNING
Roadmap
Backlog
Board
DEVELOPMENT
Code
Project pages
Add shortcut
Project settings
You're in a team-managed project
Learn more
Projects / SPRINT ACTIVITIES
Roadmap
Give feedback
Share
Export
Status category
View settings
OCT
NOV
NOV
NOV
Sprints
SA Sprint 1
SA-1 SPRINT 1
SA-2 SPRINT 2
SA-3 SPRINT 3
SA-4 SPRINT 4
+ Create Epic
Today
Weeks
Months
Quarters

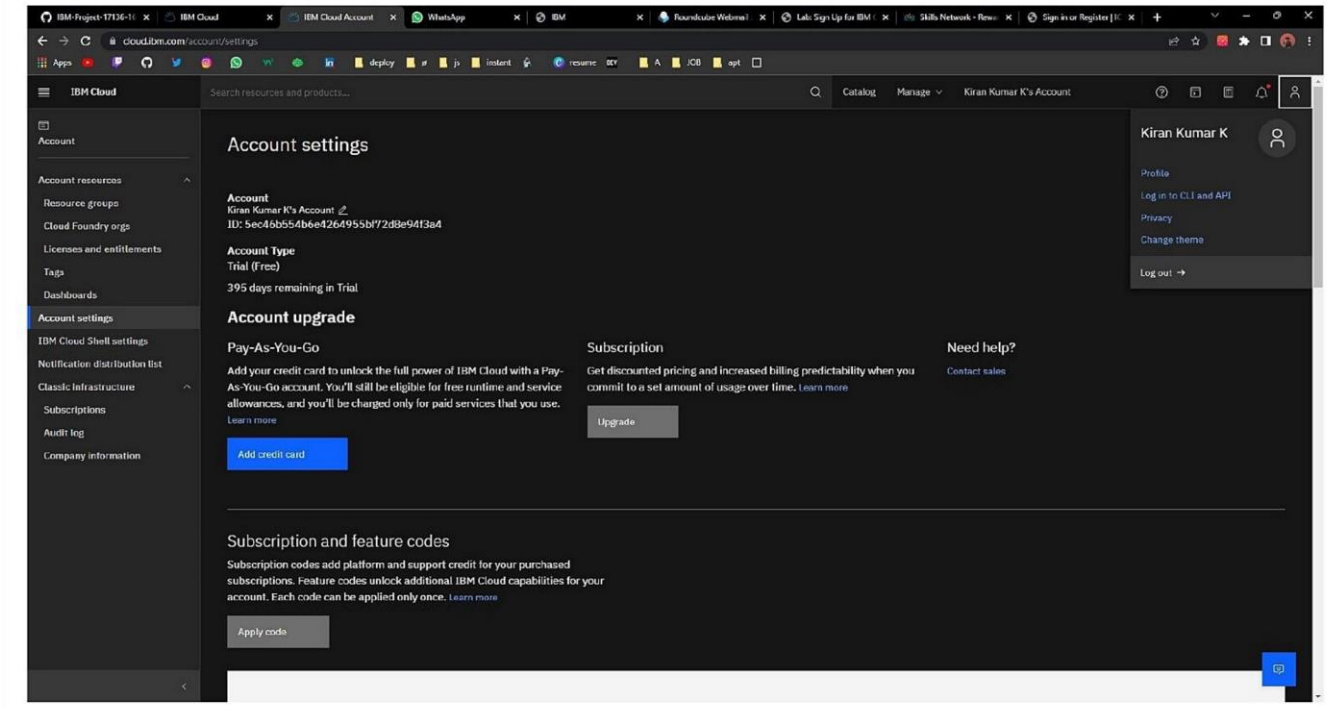
CODING AND SOLUTIONING

CHAPTER 7

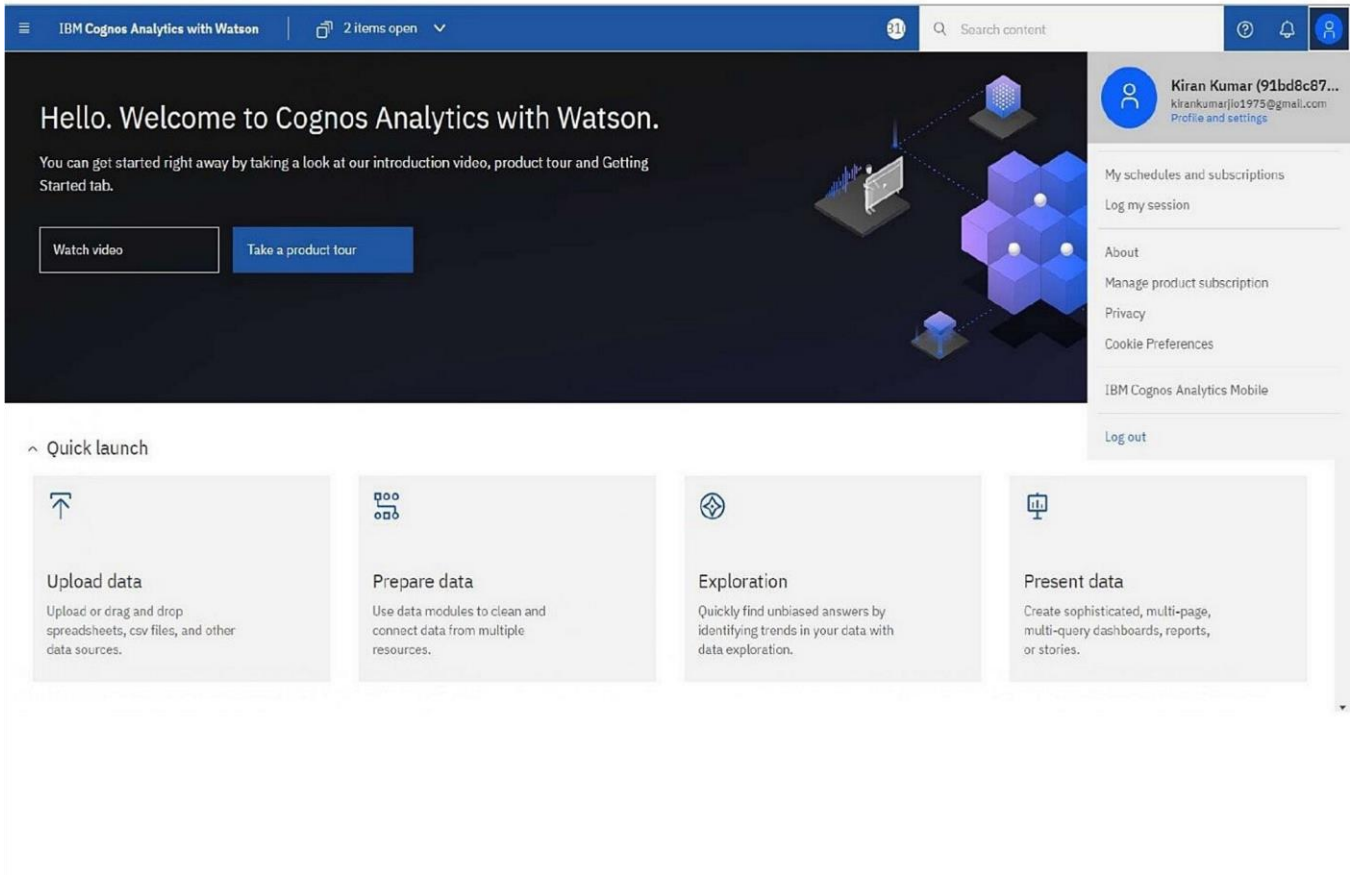
CODING AND SOLUTIONING

7.1 CREATE AND CONFIGURE IBM CLOUD SERVICES

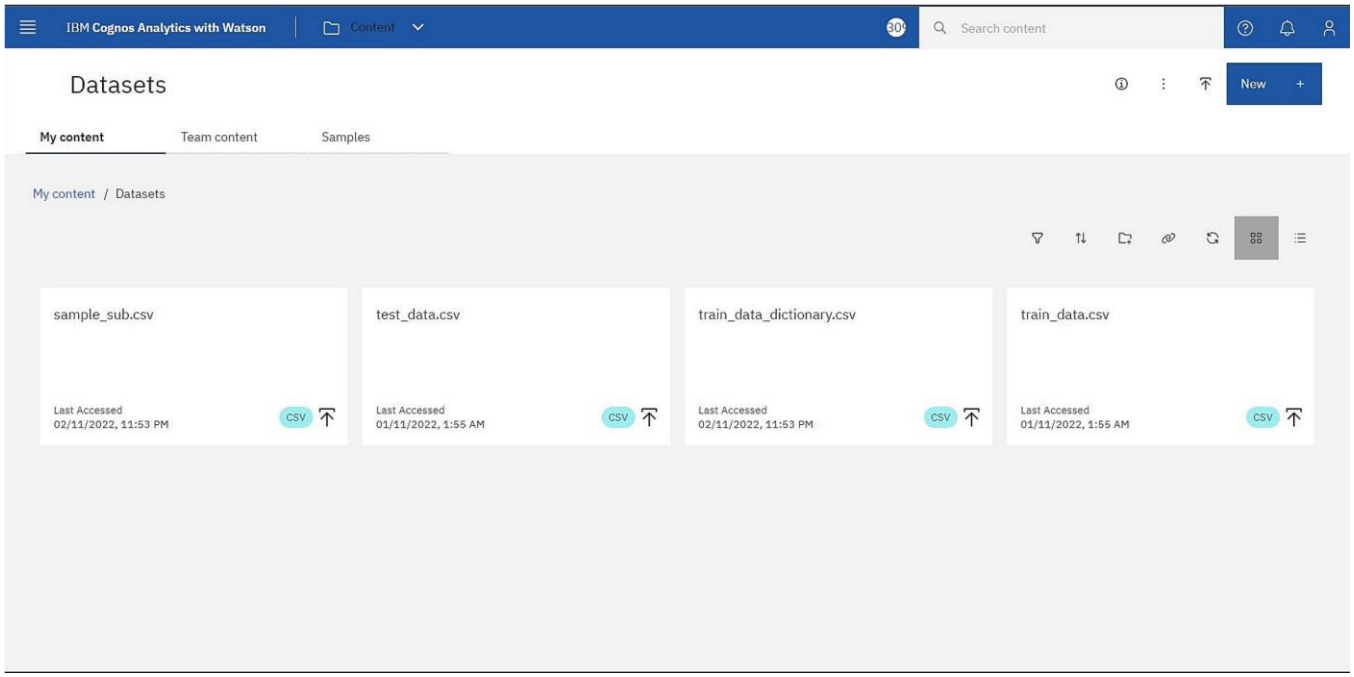
USN 1: As a user, I will create IBM cloud account.



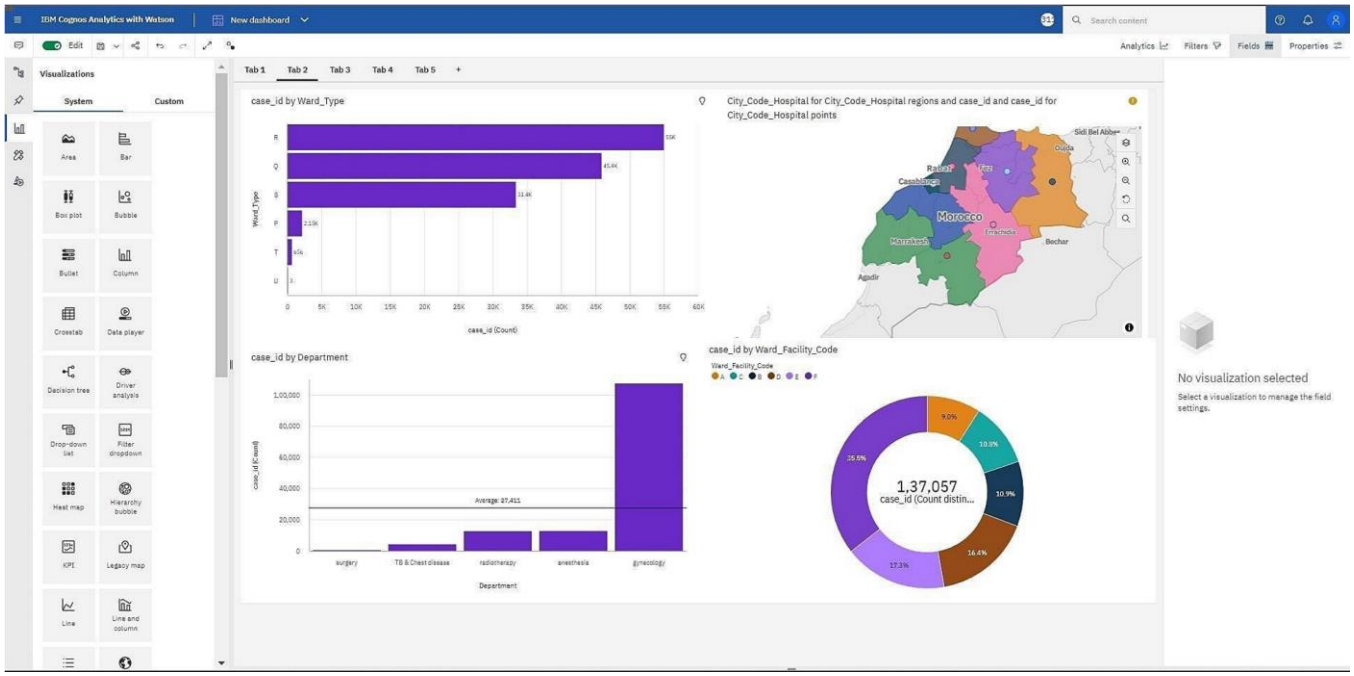
USN 2: As a user, I will create IBM cognos account.



USN 3: Loading the datasets

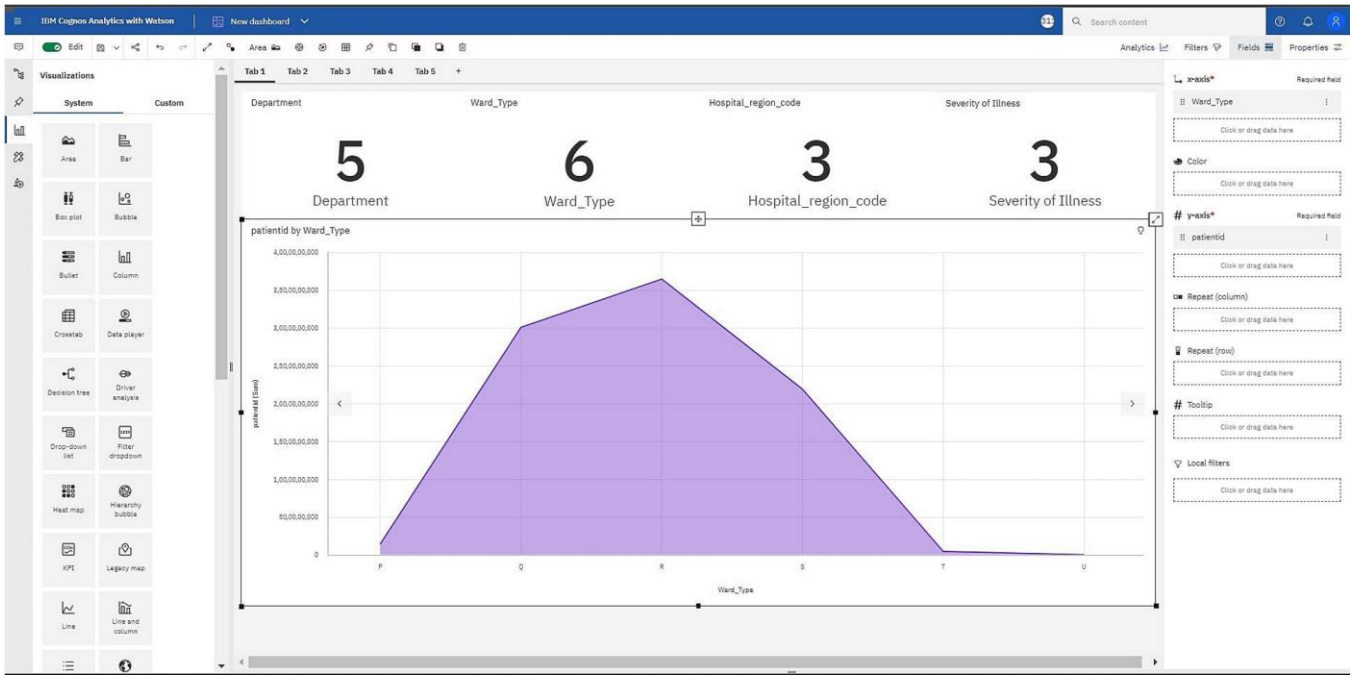


USN 4: Data analysing with different visualisatioin chart

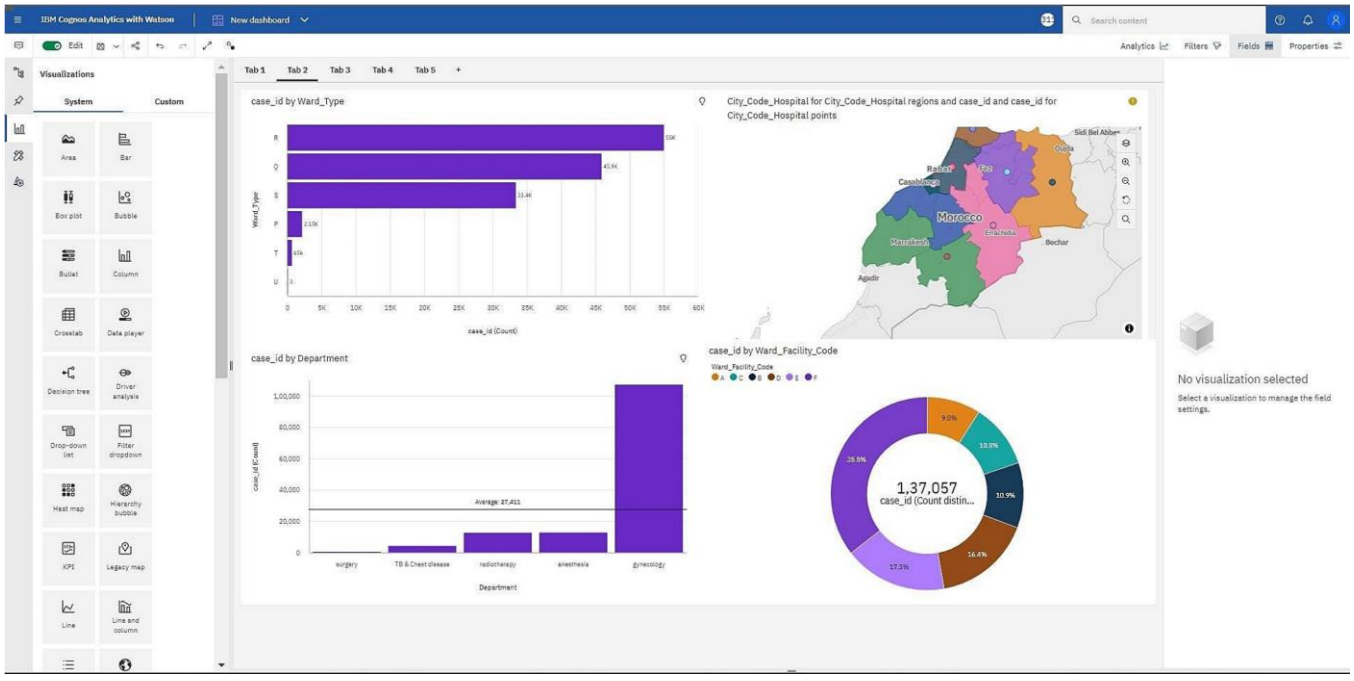


7.2 Data Visualizations

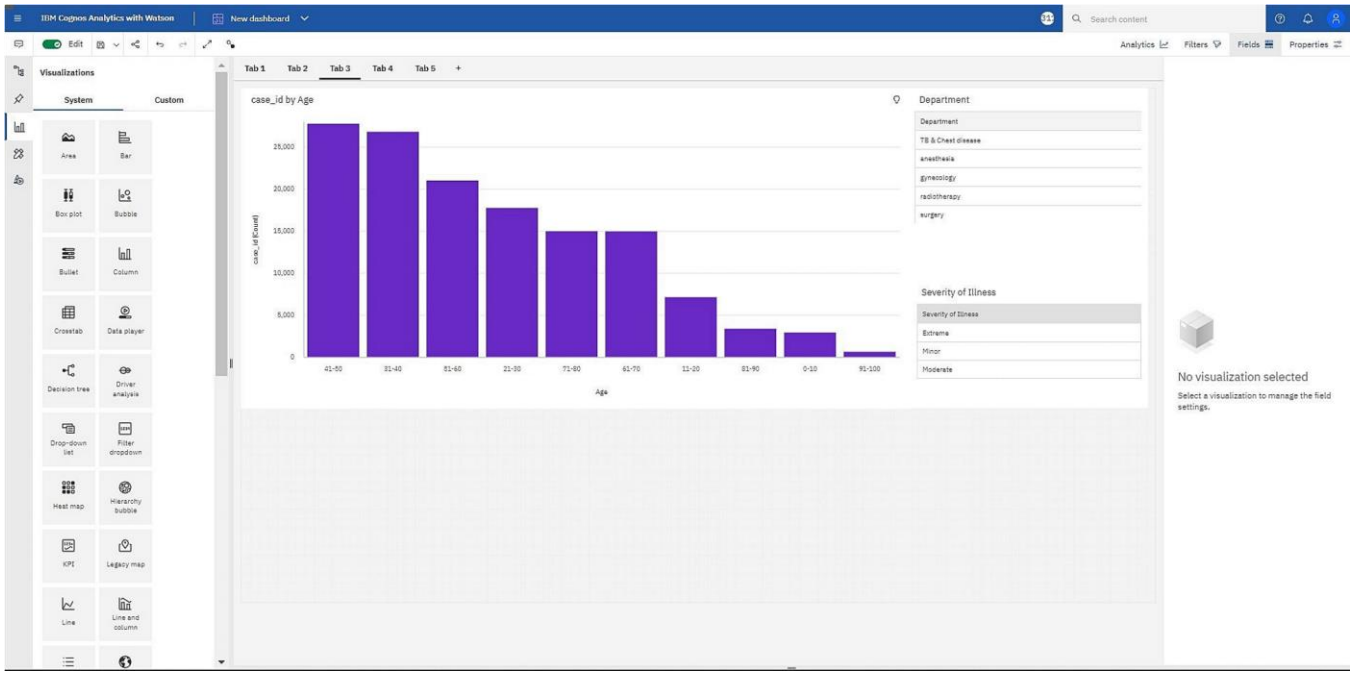
USN 5: Number of Patients By ward types.



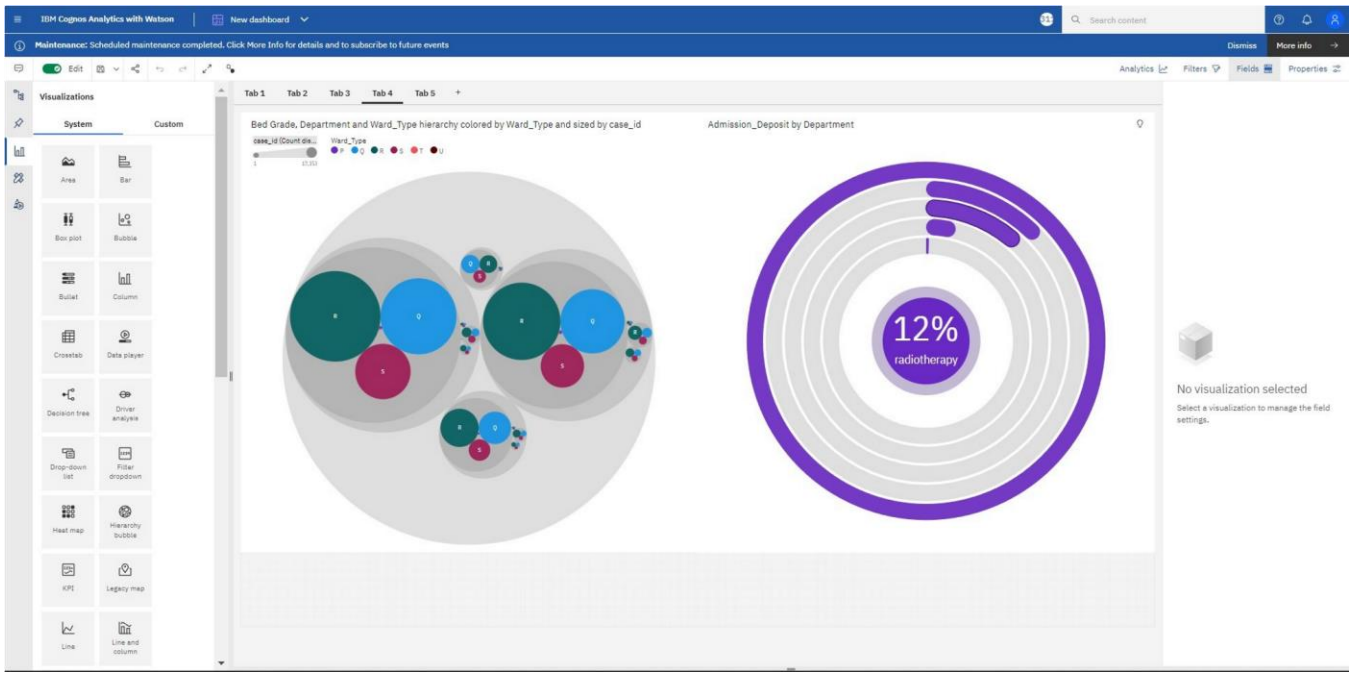
USN 6: Dashboard to show number of patients.



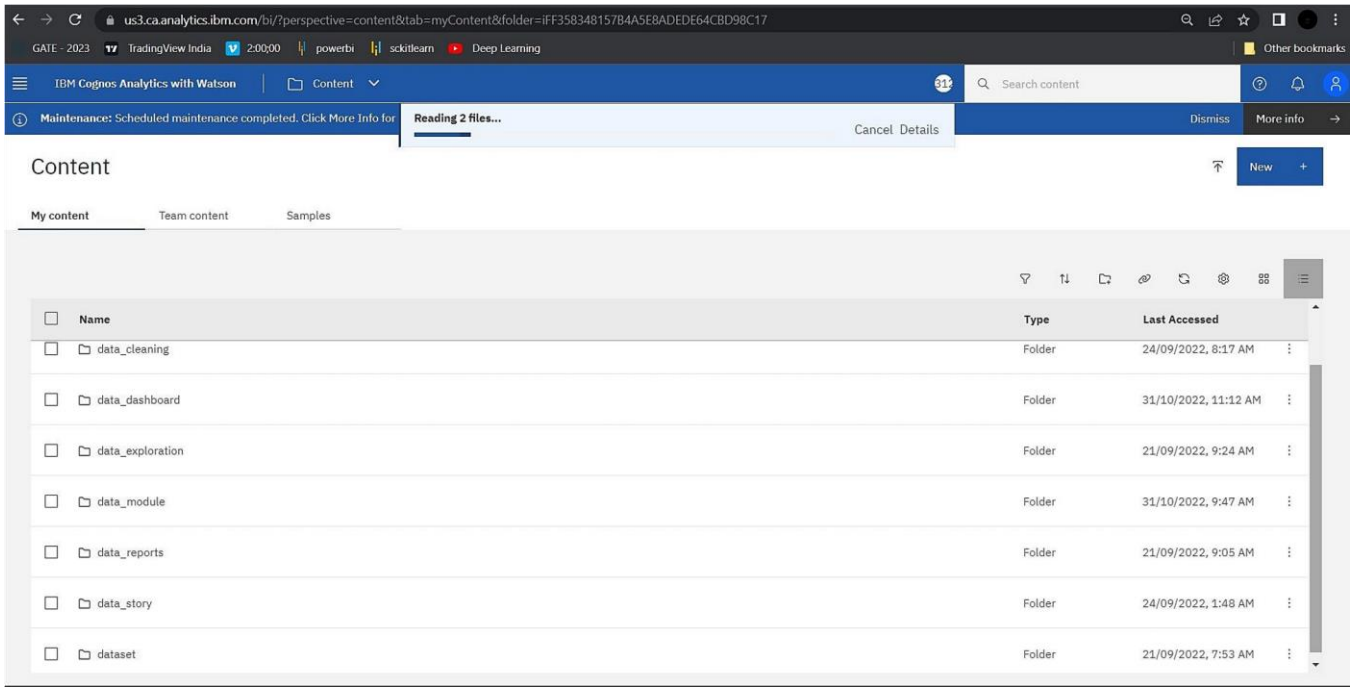
USN 7 : Age wise patients with department and severity filters.



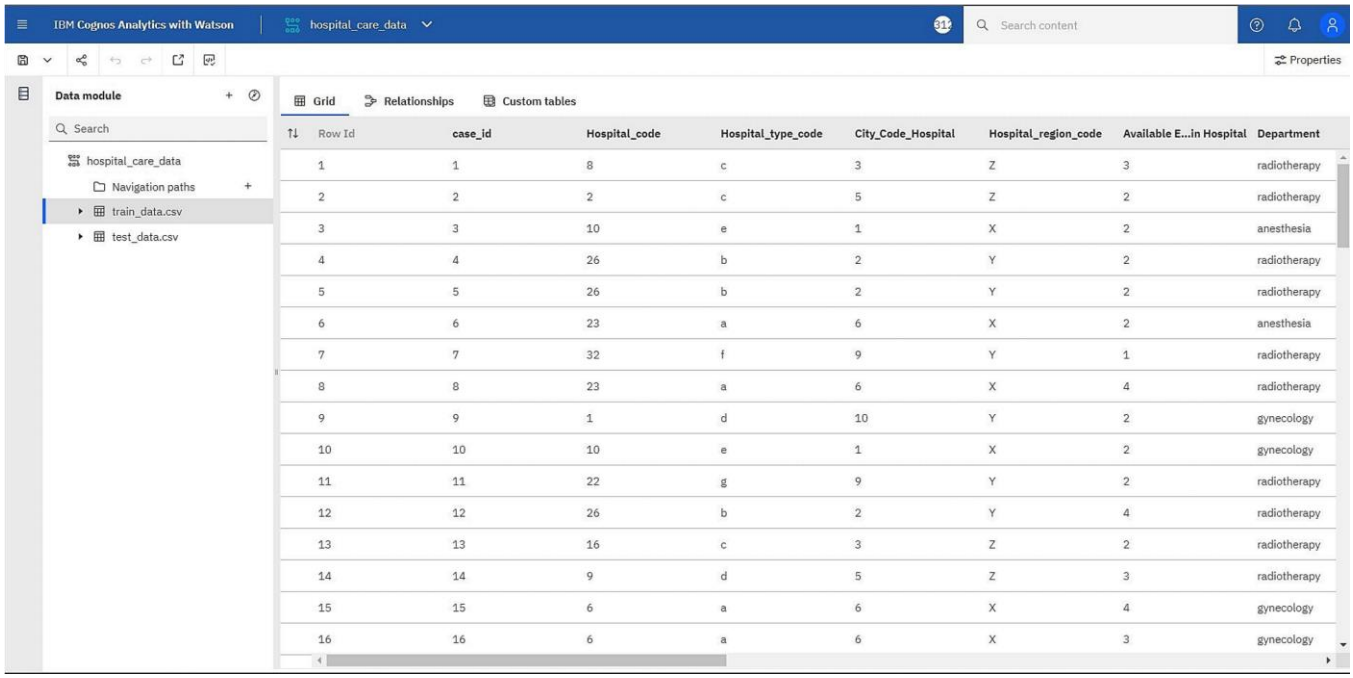
USN 8: Dashboard with Hierarchy bubble and radial visuals.



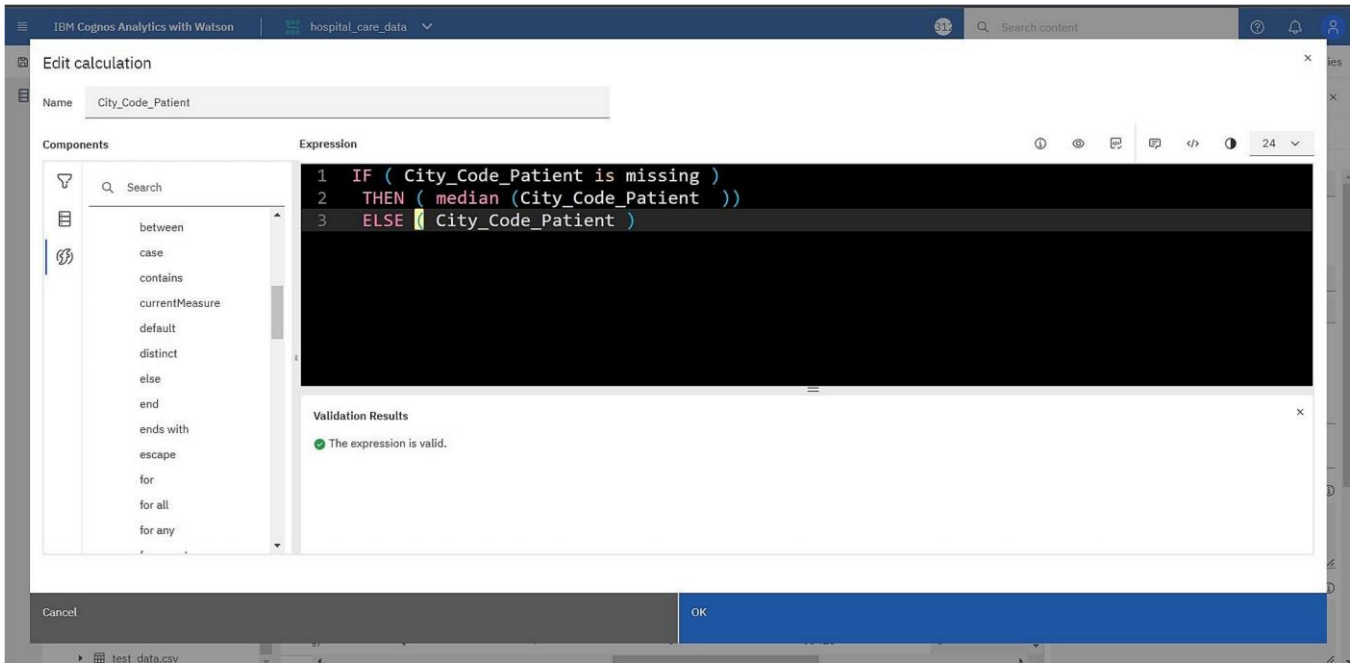
USN 9: Select the dataset.

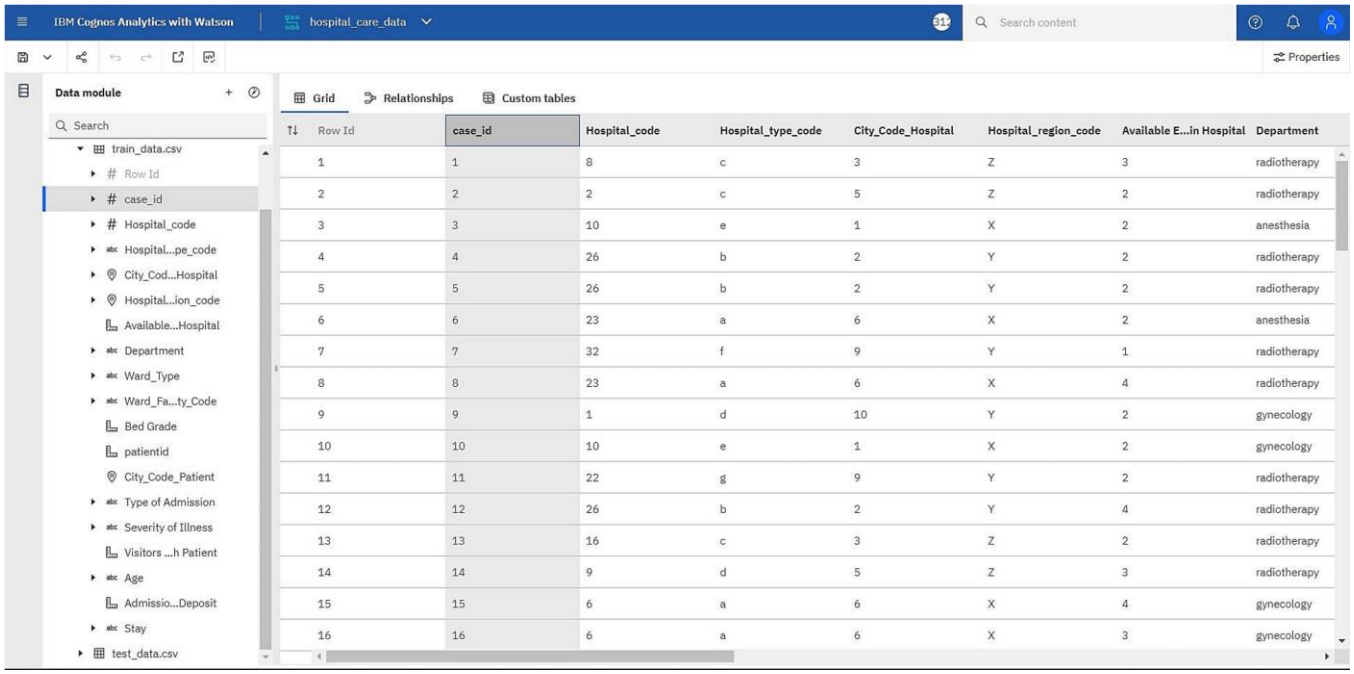


USN 10: Prepare the dataset.



USN 11: Null values cleaning process.



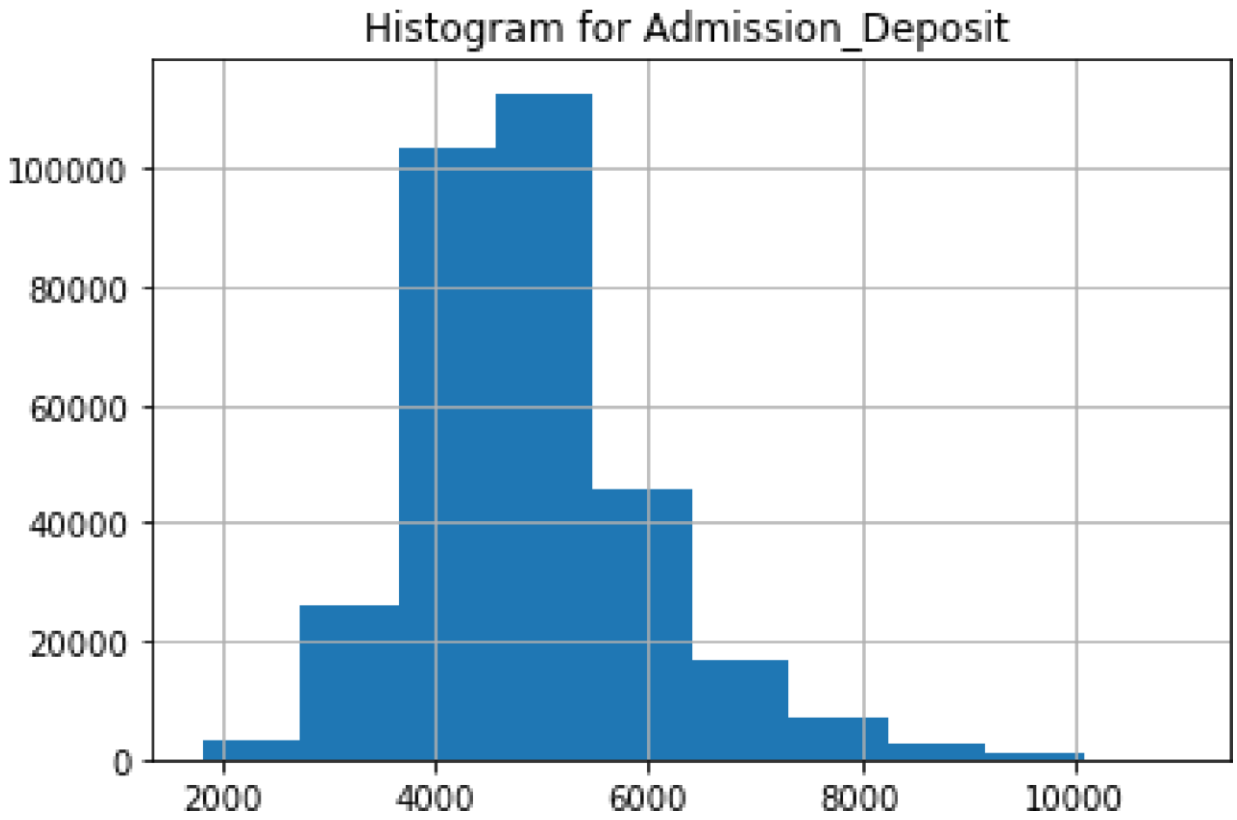


Row Id	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region_code	Available E...in Hospital	Department
1	1	8	c	3	Z	3	radiotherapy
2	2	2	c	5	Z	2	radiotherapy
3	3	10	e	1	X	2	anesthesia
4	4	26	b	2	Y	2	radiotherapy
5	5	26	b	2	Y	2	radiotherapy
6	6	23	a	6	X	2	anesthesia
7	7	32	f	9	Y	1	radiotherapy
8	8	23	a	6	X	4	radiotherapy
9	9	1	d	10	Y	2	gynecology
10	10	10	e	1	X	2	gynecology
11	11	22	g	9	Y	2	radiotherapy
12	12	26	b	2	Y	4	radiotherapy
13	13	16	c	3	Z	2	radiotherapy
14	14	9	d	5	Z	3	radiotherapy
15	15	6	a	6	X	4	gynecology
16	16	6	a	6	X	3	gynecology

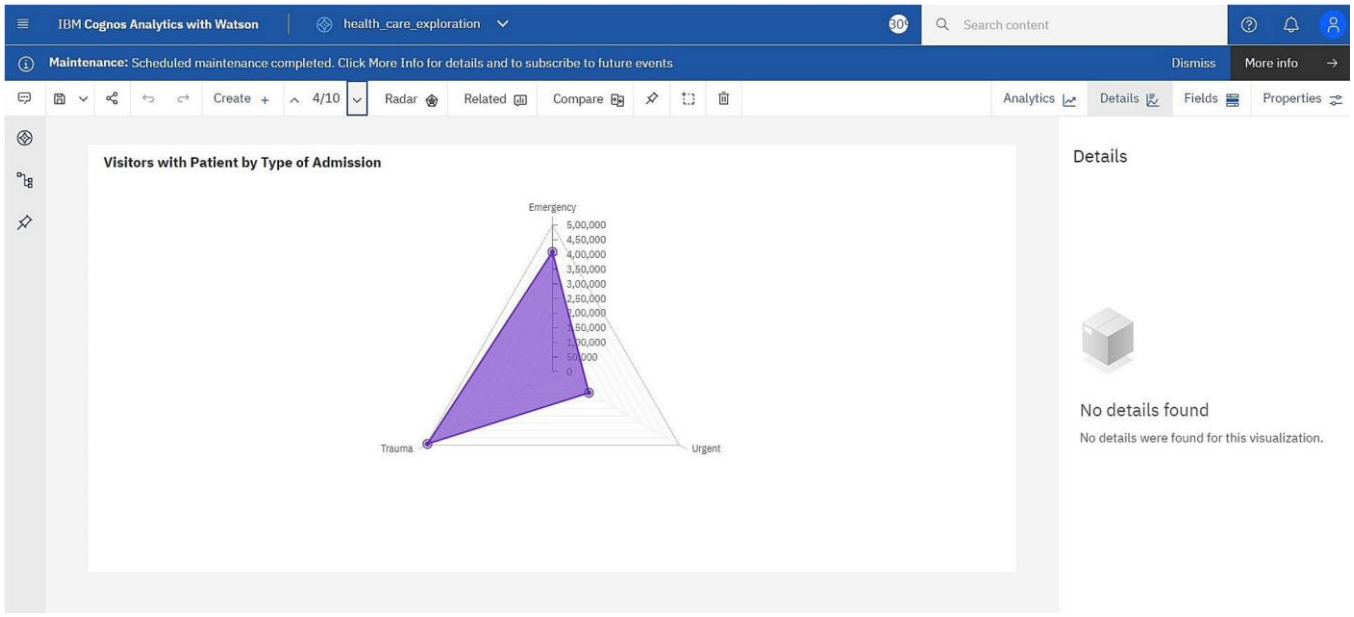
USN 12: Python pandas with numpy.

```
import pandas as pd

import numpy as np
```

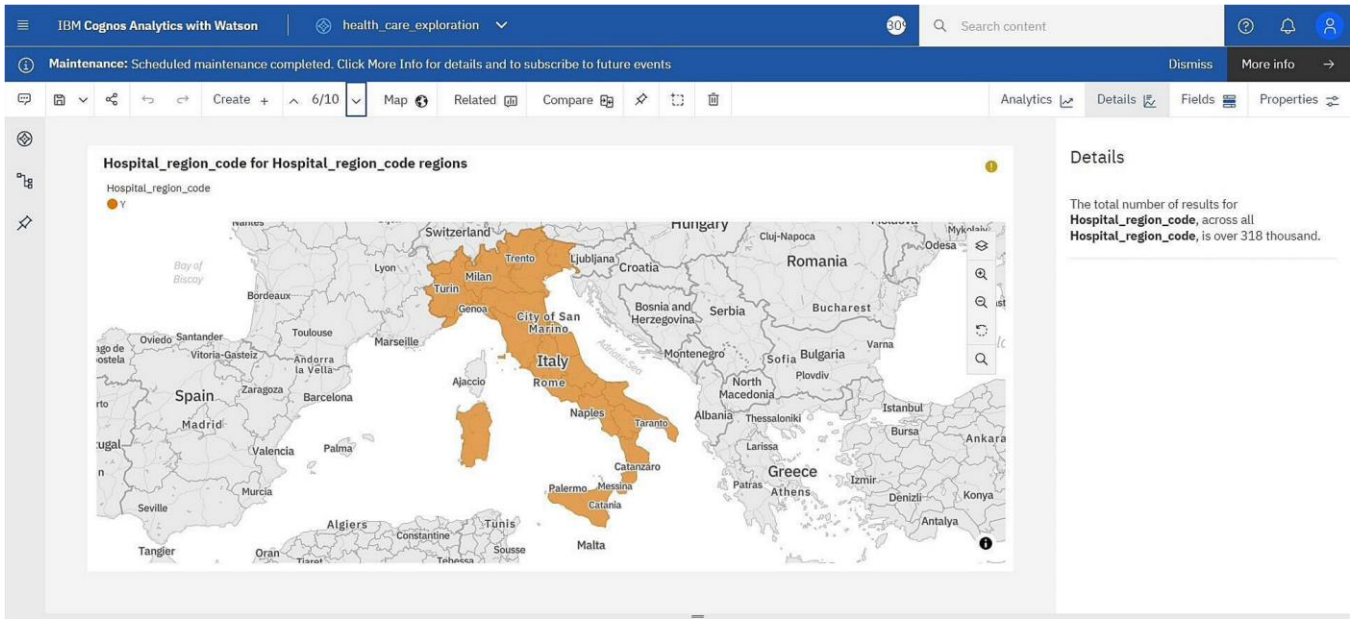
```
import matplotlib.pyplot as plt

import seaborn as sns

%matplotlib inline

df= pd.read_csv("C:/Users/nprav/OneDrive/Desktop/
Healthcare_Data/train_data.csv")

df
```



USN 13: Visitors with patient by type of admission.

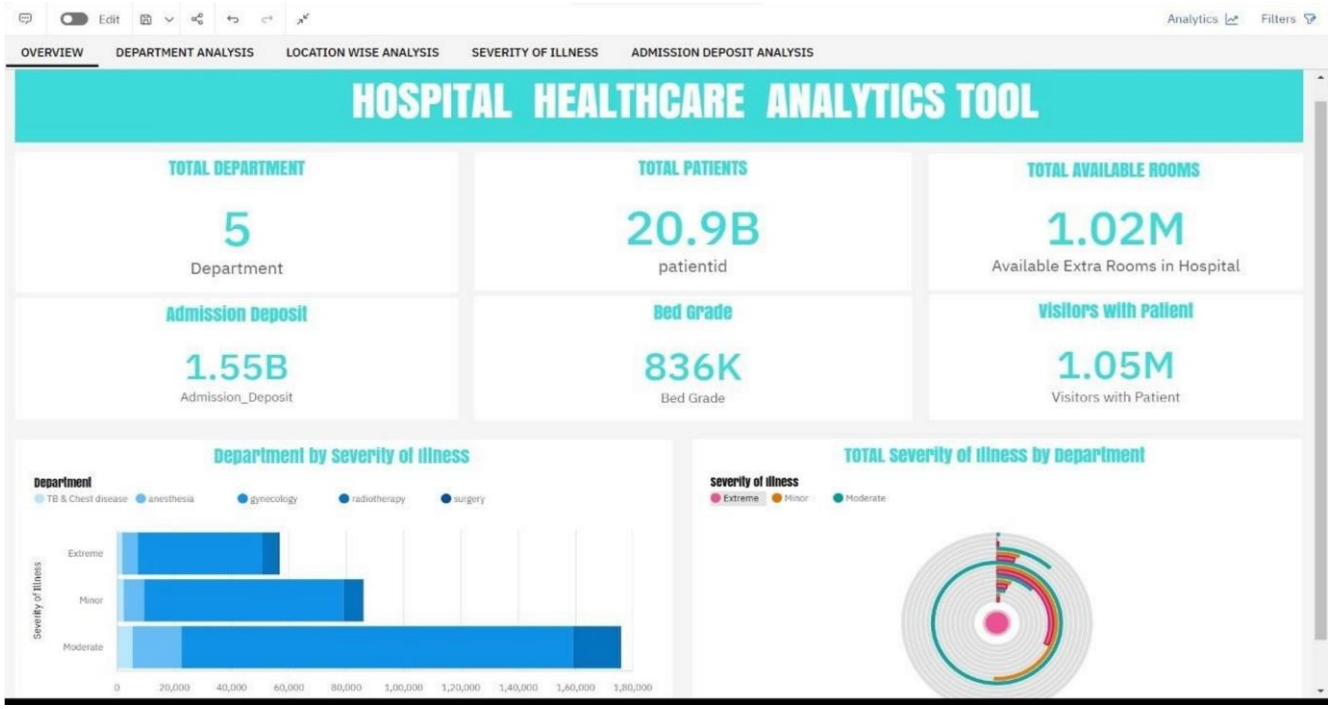


USN 14: Hospital region code for hospital region code regions.

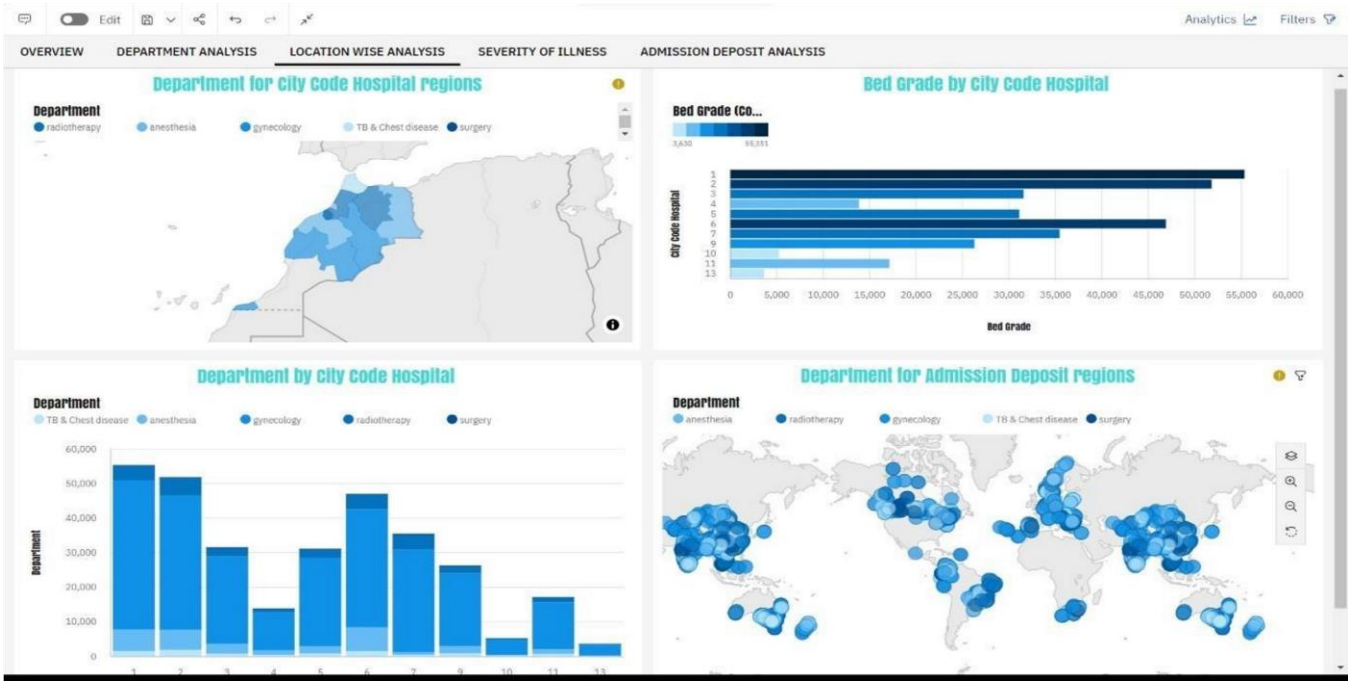


7.3 DASHBOARD CREATION WITH DATASETS.

OVERVIEW



LOCATION WISE ANALYSIS



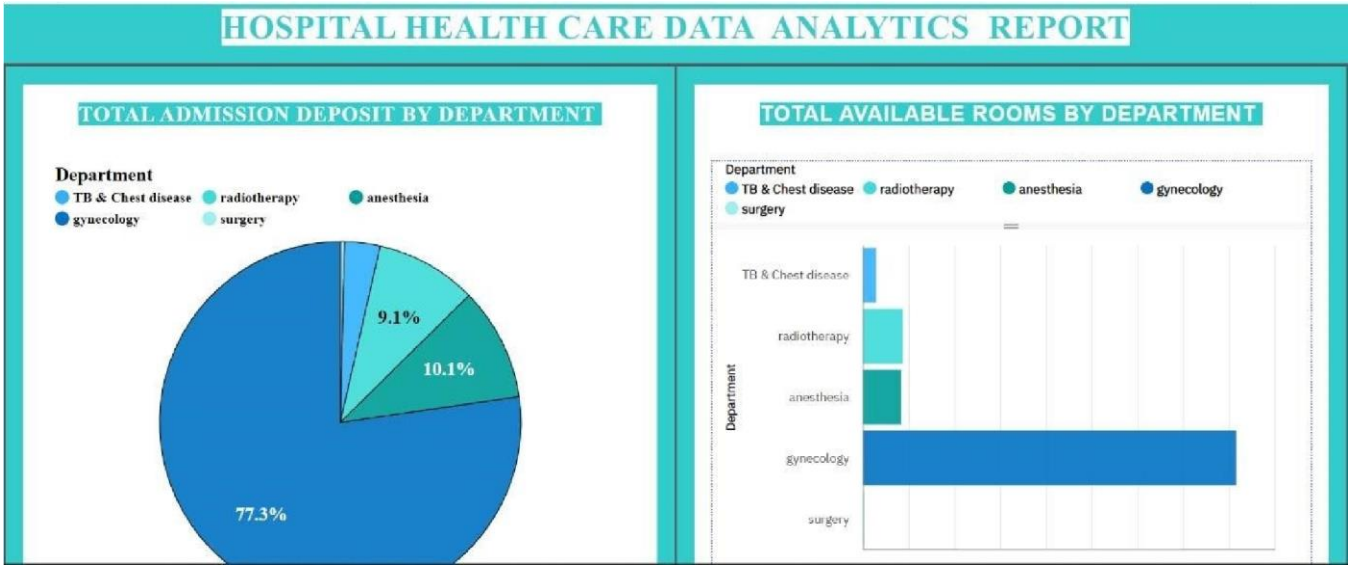
SEVERITY OF ILLNESS

ADMISSION DEPOSIT ANALYSIS

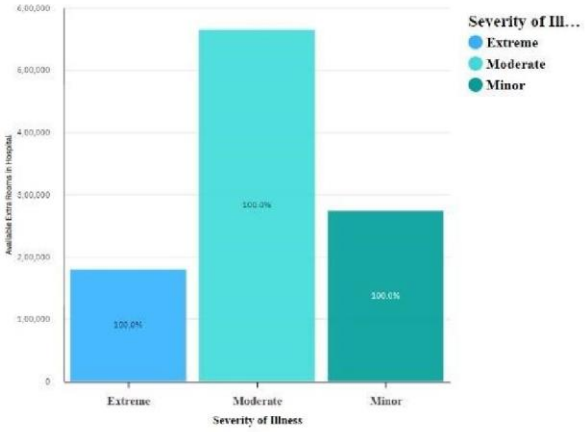
RESULTS

CHAPTER 8

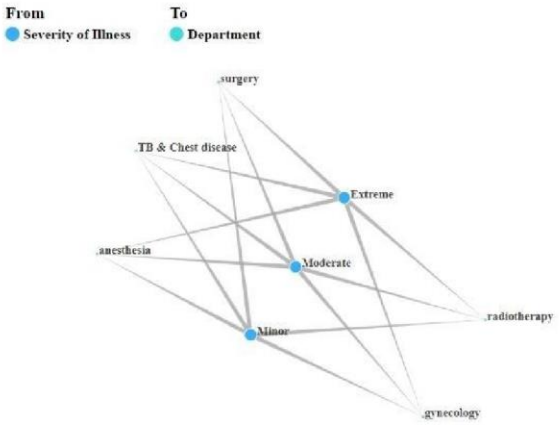
RESULT



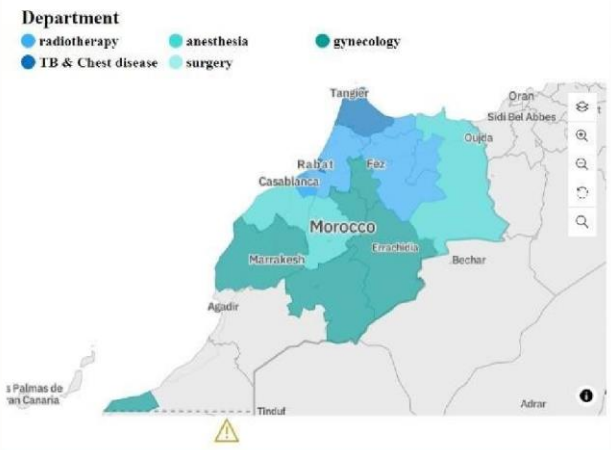
AVAILABLE EXTRA ROOMS FOR ILLNESS



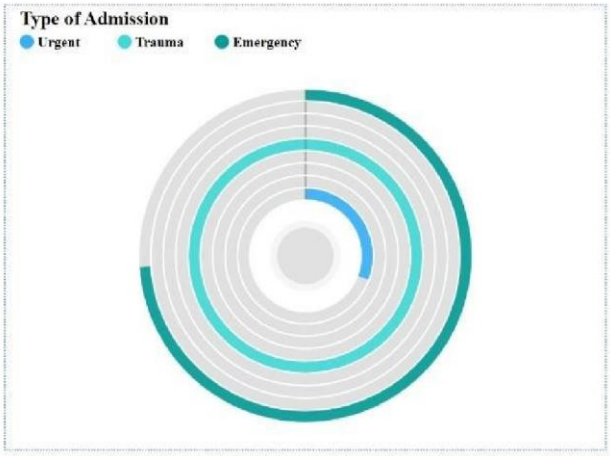
SEVERITY OF ILLNESS TO DEPARTMENT



CITY CODE HOSPITAL BY DEPARTMENT



TYPE OF ADMISSION BY ADMISSION DEPOSIT



ADVANTAGES AND DISADVANTAGES

CHAPTER

9

ADVANTAGES AND DISADVANTAGES

9.1 ADVANTAGES

- As the internet reaches the far ends of our world, so does digital health. With a simple internet connection, anyone can access **patient health records** online without visiting the medical center.
- Gone are those days of securely storing all the handwritten prescriptions and test reports. With **electronic health record** apps, you can store all the relevant health data in one place without worrying about losing one.
- Personal digital health tools like fitness bands let you know your health-related data on a real-time basis.
- Those technological marvels constantly track your vitals and auto-dial emergency numbers in case anything wrong happens to you.

9.2 DISADVANTAGES

- Adapting to new technologies has always been a challenge for senior citizens.
- Adapting to new technologies has always been a challenge for senior citizens. Some of them prefer the old-school treatment methods instead of getting used to digital health facilitators.
- If not done properly, these apps can often crash, resulting in an inconvenience.

CONCLUSION

10

CONCLUSION

This research demonstrates Analytics for hospital and health care data with data visualization and anlaytics.This information is gathered from a series of health information systems (HIS) and other technological tools utilized by health care

professionals, insurance companies and government organizations. Consider the impact this has had on the COVID-19 pandemic. The data being collected is analyzed in real time to understand the effects of the virus better and predict future trends so we may slow the spread and prevent future outbreaks. Health care data management has the potential to lead to better care if used properly.

Healthcare analytics can be understood as the gathering and analysis of healthcare sector data with the purpose of deriving insights and prompting decision-making. Ranging from main areas like medical expenses, clinical data, patient behavior, or pharmaceuticals, healthcare data analytics can be employed at both the macro and micro level to sufficiently boost operations, enhance patient care, and curtail overall expenses.

Nevertheless this data, while being highly advantageous is also pretty complicated. Be it the data from electronic health records (EHR) or the data gained by assessing real-time vital signs, the data is not only derived from a number of varying sources, but it is also required to follow government regulations, making it a complex and precarious process.

FUTURE SCOPE

11

FUTURE SCOPE

The data analytics market in the healthcare space has only increased over the last few years. Considering the rising costs of medical treatments globally, a proper body of knowledge was needed to reduce the costs at the business-level as well as the professional-level. McKinsey, in one of its reports, states that healthcare expenses constitute 17.6 percent of the GDP in the USA, which is approximately US$600 billion, more than what is the set benchmark for the ideal size of population in the country. This is a serious indicator of bigger trouble. Hence, the usage of healthcare data analytics is being promoted these days.

To some, the domain of healthcare data analytics may look new, but it has a lot of potential, especially if you wish to engage in challenging job roles and build a strong data analytics profile in the upcoming years. In this blog, we have covered some of the major topics such as what is healthcare data analytics, its applications,

scope, and benefits, etc. We hope it helps you in your decision-making as a healthcare data analytics professional.