ANALYTICS FOR HOSPITAL HEALTH CARE DATA

1.INTRODUCTION

1.1 PROJECT OVERVIEW

The introduction to Data Analytics in healthcare will allow to use new technologies both in treatment of patients and health management. It is no secret that data has become important in recent years. As our ability to collect and store information has grown, so has our ability to glean insights from that data. This is particularly true in the healthcare industry. Data analytics can achieve a wide range of objectives, from reducing treatment costs to improving the quality of life.

Healthcare analytics can potentially transform how we manage our health, both on an individual and population level. These insights can help us predict and prevent disease outbreaks and improve the quality of care that patients receive. Healthcare analytics will become increasingly important in a world where the average human lifespan is increasing. We must learn to use data effectively to meet the challenges of an aging population. By harnessing the power of data, we can ensure that everyone has access to the best possible care.

1.2 PURPOSE

We need a centralized, systematic way of collecting, storing and analyzing data so we can use it to our advantage. Healthcare organizations need advanced software as medical tools and methods to transform complex data into insightful information. The data analytics technology thus enables organizations to get maintain a highly-structured data repository that helps them make well-informed decisions for greater efficiency and improved quality of services. Once the healthcare organizations implement data analytics into their systems and get comfortable leveraging the capability of the technology, they start

seeing the bigger picture of the healthcare services by receiving detailed and structured patients' information, which subsequently enables them to offer a completely personalized, holistic and precise treatment for the disease.

2. LITERATURE REVIEW

2.1 EXISTING PROBLEM

1. Capturing Accurate Data

Health care data is assembled from various sources and in different formats, such as structured data, photographs, videos, paper, digital, multimedia, and so on. Capturing data that is clean, accurate, comprehensive, and formatted precisely for use in numerous frameworks is a real challenge for organizations.

2. Fragmented Patient Care

As we mentioned above that most of the data received from various sources is unstructured and undiscovered, making EHR systems more ingenious and interoperable is another challenge. It's pivotal to secure the information of patients, staff, billing, and performance.

Providers must have a clear perspective that which data needs a manual update, and which one needs an automatic update, to avoid downtime of end-users and harming the quality of the dataset.

3. Data Privacy & Security

The HIPAA Security Rule incorporates a list of specialized security for organizations storing protected health information (PHI), including authentication protocols, transmission security, controls over access, auditing, and so on.

Many organizations secure their data with security procedures such as up-to-date antivirus, encrypting sensitive data, multi-factor authentication, but even the most secured data can be taken down due to complicated constraints on their access to data or software.

4. Data Visualization

These days data often need to be visually presented in the form of interactive graphs or charts to be impactful and understood. And, we know it’s quite frustrating and time-consuming to drag information from multiple areas and put it into a reporting tool.

2.2 REFERENCES

[1] Abouelmehdi K, Beni-Hessane A, Khaloufi H. Big healthcare data: preserving security and privacy. J Big Data. 2018. https://doi.org/10.1186/s40537-017-0110-7.

[2] Agrawal A, Choudhary A. Health services data: big data analytics for deriving predictive healthcare insights. Health Serv Eval. 2019. https://doi.org/10.1007/978-1-4899-7673-4_2-1.

[3] Al Mayahi S, Al-Badi A, Tarhini A. Exploring the potential benefits of big data analytics in providing smart healthcare. In: Miraz MH, Excell P, Ware A, Ali M, Soomro S, editors. Emerging technologies in computing—first international conference, iCETiC 2018, proceedings (Lecture

Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST). Cham: Springer; 2018. p. 247–58. https://doi.org/10.1007/978-3-31995450-9_21.

[4] Bainbridge M. Big data challenges for clinical and precision medicine. In: Househ M, Kushniruk A, Borycki E, editors. Big data, big challenges: a healthcare perspective: background, issues, solutions and research directions. Cham: Springer; 2019. p. 17–31.

[5] Bartuś K, Batko K, Lorek P. Business intelligence systems: barriers during implementation. In: Jabłoński M, editor. Strategic performance management new concept and contemporary trends. New York: Nova Science Publishers; 2017. p. 299–327. ISBN: 978-1-53612-681-5.

[6] Bartuś K, Batko K, Lorek P. Diagnoza wykorzystania big data w organizacjach-wybrane wyniki badań. Informatyka Ekonomiczna. 2017;3(45):9–20.

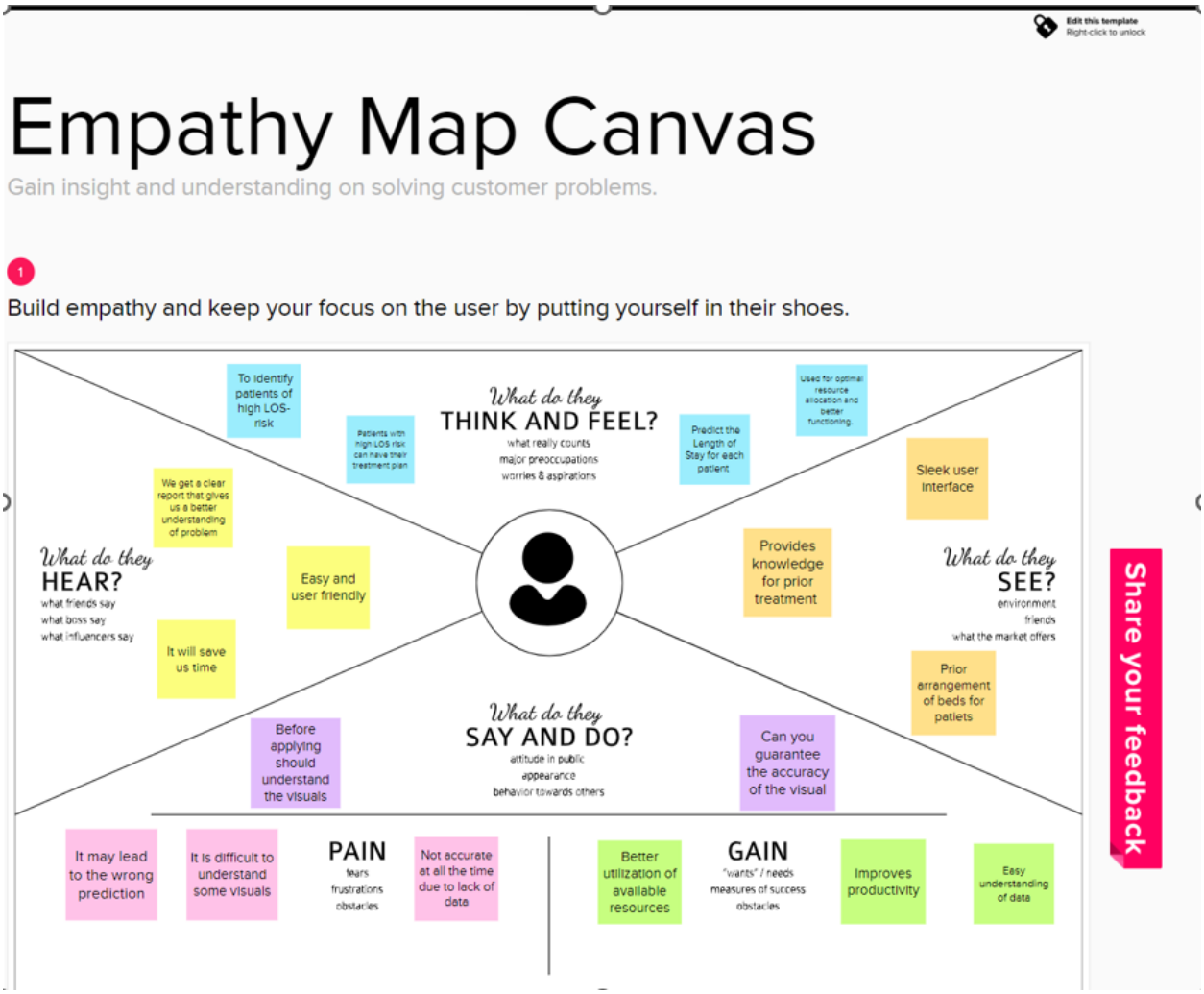
[7] Bartuś K, Batko K, Lorek P. Wykorzystanie rozwiązań business intelligence, competitive intelligence i big data w przedsiębiorstwach województwa śląskiego. Przegląd Organizacji. 2018;2:33–9.

2.3 PROBLEM STATEMENT DEFINITION

What does the problem affect?	It will affect the hospital management
What are the boundaries of the problem?	Inadequate amount of beds, medicines, etc.
What is the issue?	Due to increase in number of patients, the availability of doctors and beds are inadequate
When does the issue occur?	It occurs when the more number of patients are admitted in the hospital
Where does the issue occur?	The issue occurs in the hospital
Why is it important that we fix the problem?	In an emergency situation the patients may not have proper treatment and bed. So, the analytics of providing proper treatment in correct time.
What solution to solve this issue?	Data visualization is used to identify the patients by visualizing the length of stay in the hospital.
What methodology used to solve this issue?	Data analytics techniques are used to identify the patients and suggest the precautions that can be taken for managing the prerequisite of patients.

3. IDEATION & PROPOSED SOLUTION

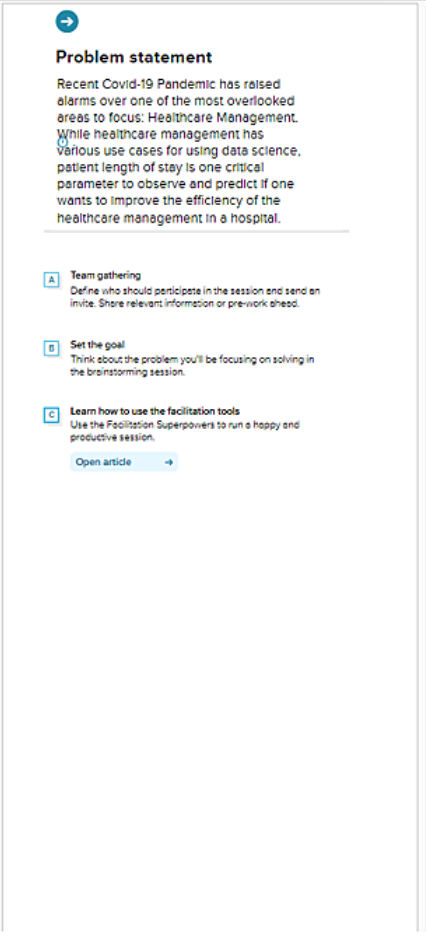
3.1 EMPATHY MAP CANVAS



3.2 IDEATION & BRAINSTORM

Analytics for health care data.

Recent Covid-19 Pandemic has raised alarms over one of the most overlooked areas to focus: Healthcare Management. While healthcare management has various use cases for using data science, patient length of stay is one 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 (patients who will stay longer) 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..



2 Brainstorm

Write down any ideas that come to mind that address your problem statement.

10 minutes

Vinisha S

- Optimal resource allocation
- Able to diagnose disease faster
- Able to treat patients better
- Interaction between doctor and patient is increased
- Better treatment in immediate situation
- Patient health can be monitered wisely

Swarnambigai T

- Help doctor make an effective treatment
- figure out defects in body
- detect the symptoms of a disease
- on time and proper treatment
- handle emergency situation effectively
- doctors get notified about the situation

Dhurga P

- precision medicine
- facilities management
- improved diagnostics
- better health outcomes
- improved patient interactions
- avoid life-long issues

Guna S

- Improve hospital capcity
- New drug discovery
- Maintain health records
- Reduces hospital waiting tme
- Patient satisfaction
- Health monitoring

3 Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

20 minutes

Category 1

- Optimal resource allocation
- facilities management
- Improve hospital capcity

Category 2

- Able to diagnose disease faster
- Able to treat patients better
- detect the symptoms of a disease
- handle emergency situation effectively
- improved diagnostics
- Reduces hospital waiting tme

Category 3

- Interaction between doctor and patient is increased
- doctors get notified about the situation
- improved patient interactions

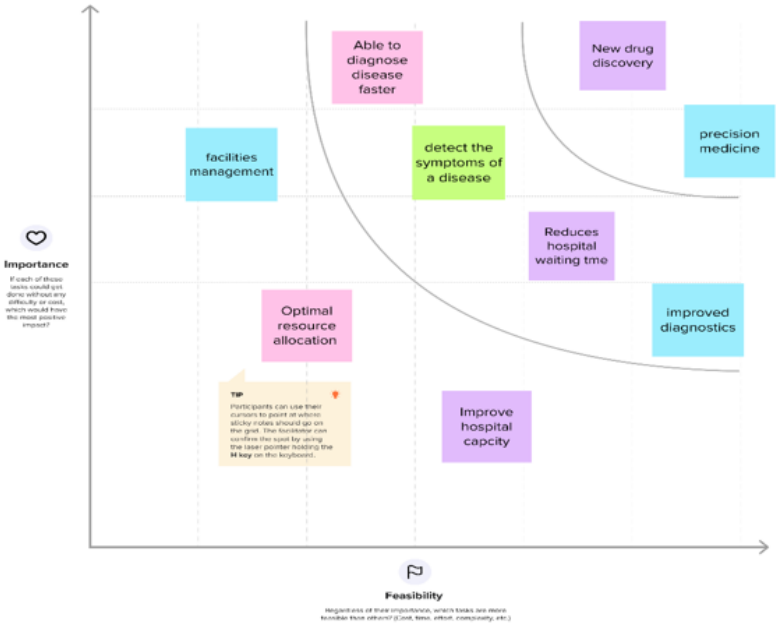
Category 4

- New drug discovery
- precision medicine
- Maintain health records

4 Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

20 minutes



After you collaborate

You can export the mural as an image or pdf to share with members of your company who might find it helpful.

Quick add-ons

- **Share the mural**
  Share a view link to the mural with stakeholders to keep them in the loop about the outcomes of the session.
- **Export the mural**
  Export a copy of the mural as a PNG or PDF to attach to emails, include in slides, or save in your drive.

Keep moving forward

- **Strategy blueprint**
  Define the components of a new idea or strategy.
  Open the template →
- **Customer experience journey map**
  Understand customer needs, motivations, and obstacles for an experience.
  Open the template →
- **Strengths, weaknesses, opportunities & threats**
  Identify strengths, weaknesses, opportunities, and threats (SWOT) to develop a plan.
  Open the template →

Share template feedback

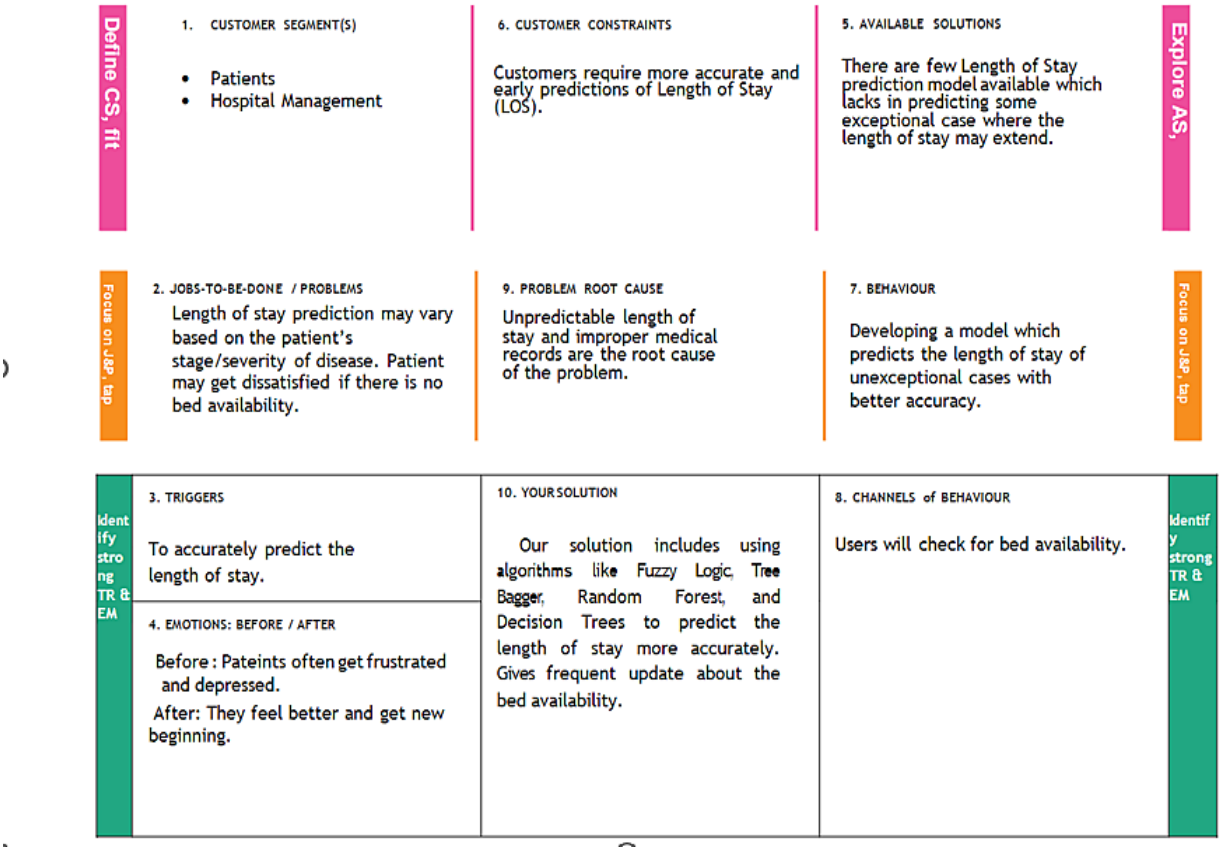
3.3 PROPOSED SOLUTION

S.No	Parameter	Description
1.	Problem Statement (Problem to be solved)	To predict the length of stay of patients.
2.	Idea / Solution description	The length of stay can be predicted using either Fuzzy logic or Tree bagger algorithms. Along with the algorithm certain parameters like age, stage of disease, progression, etc., are used for prediction. IBM Cognos is used for analytics purpose.
3.	Novelty / Uniqueness	It predicts the result with more accuracy using which overstays can be reduced. Proper resources and therapy can be provided.
4.	Social Impact / Customer Satisfaction	Patients can get better treatment and care than before. Length of stay prediction minimize the overflow of patients therefore hospital resource management and utilization will be maximized. Reduces expense for treatment.

5.	Business Model (Revenue Model)	1. This system can be used in all government hospitals, private hospitals and even in small clinics. 2. Activities – Length of stay prediction. 3. Key Resource – Medical records. 4. Bed consumption is low.
6.	Scalability of the Solution	This model will predict the length of stay of all kind of patients.

3.4 PROBLEM SOLUTION FIT

Project Title: Analytics For Hospital's Health-Care Data | Project Design Phase-I - Solution Fit Template | Team ID: PNT2022TMID41255



4. REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENT

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	This Dashboards are designed to offer a comprehensive overview of patient's LOS, and do so through the use of data visualization tools like charts and graphs.
NFR-2	**Security**	The Dashboard helps to indicate the current threat level to the Hospitals; an indication of events and incidents that have occurred; a record of authentication errors; unauthorized access
NFR-3	**Reliability**	This dashboard will be consistent and reliable to the users and helps the user to use in effective, efficient and reliable manner.
NFR-4	**Performance**	This dashboard can scan the backend users and analyzing the frequency in which they visit the dashboard helps understand how useful and helpful the data displayed is for tasks.

Following are the functional requirements of the proposed solution:

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	**User Registration**	Registration through Form Registration through Gmail
FR-2	**User Confirmation**	Confirmation via Email Confirmation via Message
FR-3	**Interoperability**	Dashboard helps to share the patient's information interoperable to the hospitals in timely manner.
FR-4	**Accuracy**	Dashboard helps predict the patient's Health risks accurately based on LOS (Length of Stay).
FR-5	**Compliance**	The compliance of a dashboard is like to use very interactively in real time by the hospitals.
FR-6	**Concise**	These dashboards are clear, intuitive, and customizable and interactive in manner.

4.2 NON-FUNCTIONAL REQUIREMENT

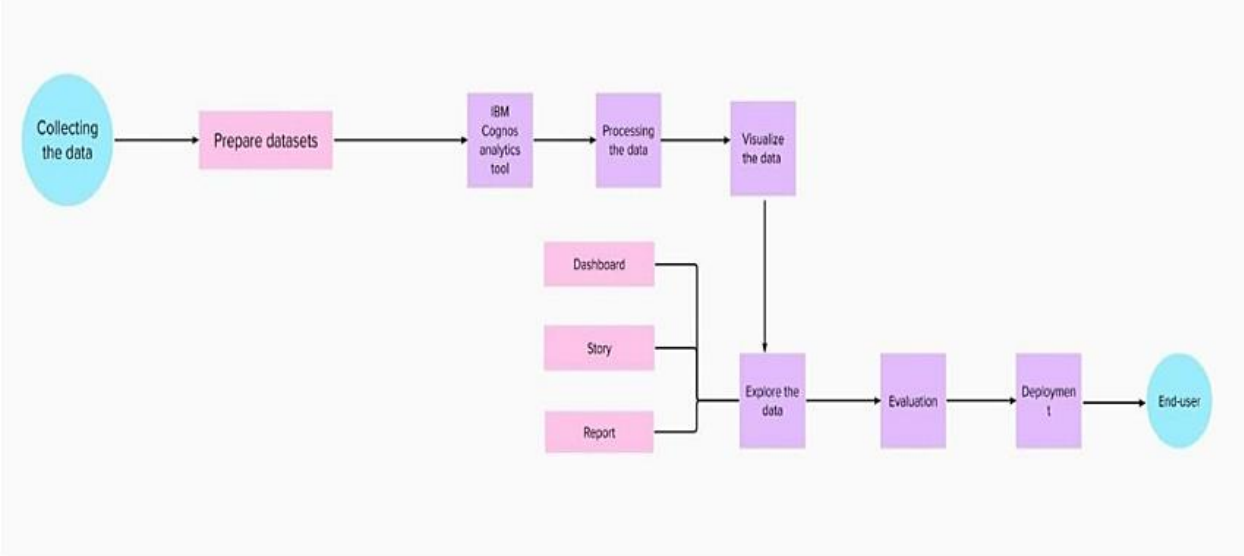
Following are the non-functional requirements of the proposed solution:

NFR-5	**Availability**	The dashboard can available to meet user's demand in timely manner and it is also helps to provide necessary information to the user's dataset
NFR-6	**Scalability**	The layers used in the dashboard are a hosted feature layer, feature layer view, or hosted tile layer.

5.PROJECT DESIGN

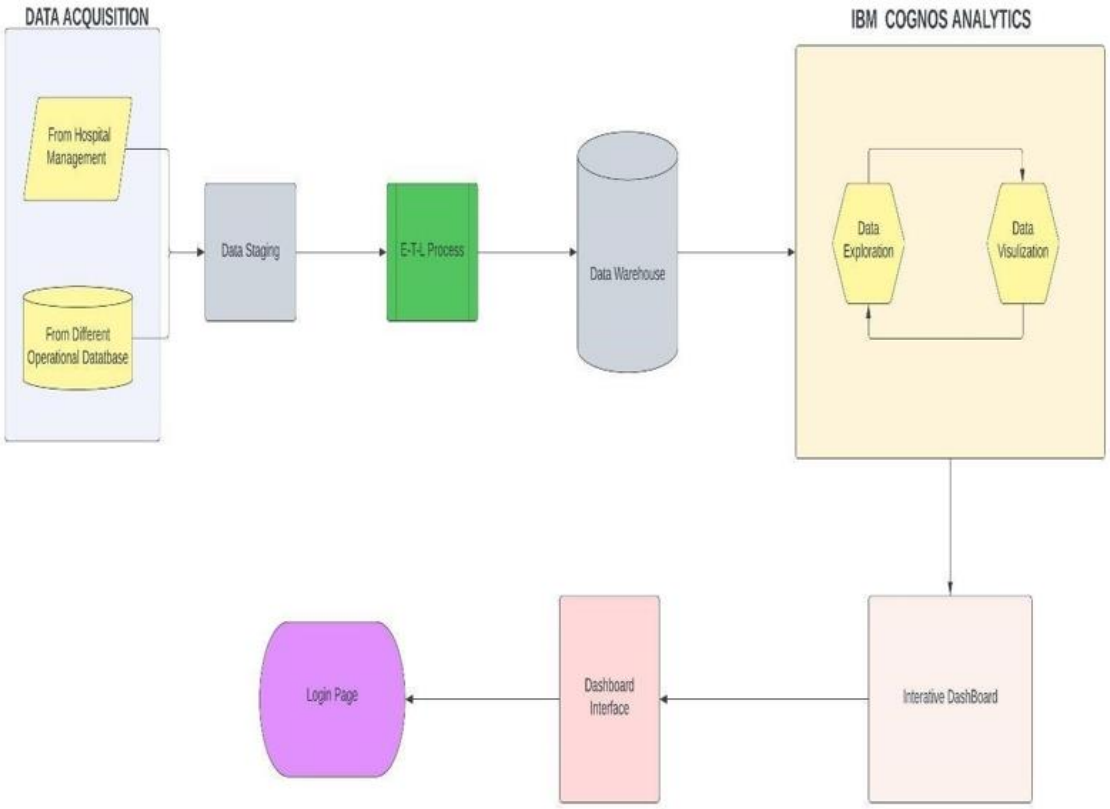
5.1 Data Flow Diagrams

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



5.2 SOLUTION & TECHNICAL ARCHITECTURE

SOLUTION ARCHITECTURE


DATA ACQUISITION
From Hospital Management
From Different Operational Database
Data Staging
E-T-L Process
Data Warehouse
IBM COGNOS ANALYTICS
Data Exploration
Data Visulization
Interative DashBoard
Dashboard Interface
Login Page

TECHNICAL ARCHITECTURE:

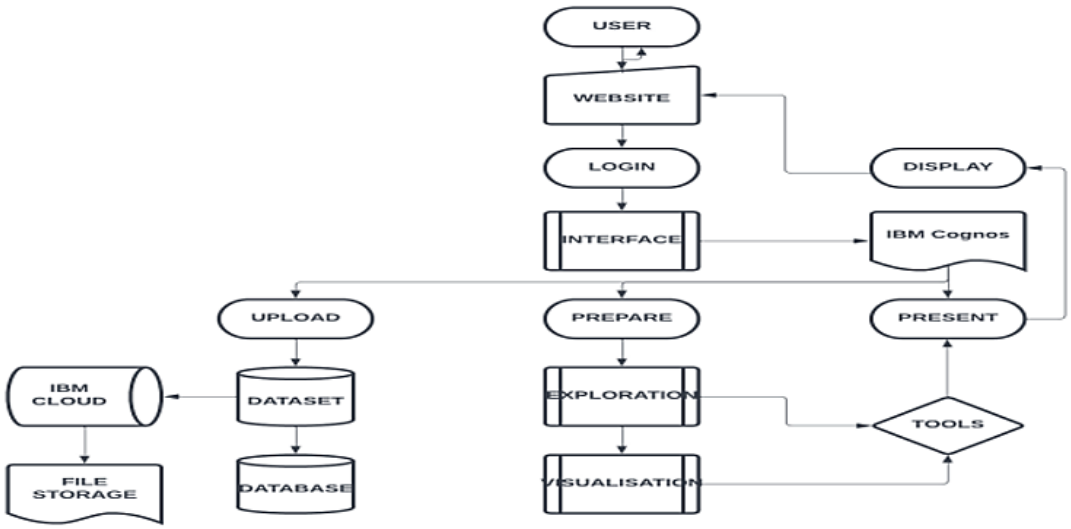


Table-1 : Components & Technologies:

5.3 USER STORIES

Use the below template to list all the user stories for the product:

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 data and the result after the prediction	High	Sprint-2
Admin	Analyse	USN-7	As an admin, I will analyze the given dataset	I can analyze 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

6.PROJECT PLANNING & SCHEDULING

6.1 SPRINT PLANNING & ESTIMATION

MILESTONES	TASKS
MILESTONE - 1	Data Collecting process (Datasets)
MILESTONE - 2	Required Datasets are uploaded on the IBM Cognitive Platform.
MILESTONE - 3	Data Exploration and Data Visualization
MILESTONE - 4	To Create a Interactive Dashboard.
MILESTONE - 5	Display the Insights in the Dashboard
MILESTONE - 6	Construct a Standardized Data Set and use the needed data with the Assistance of a Python Program
MILESTONE - 7	Use of different algorithm with Google Colab to achieve the desired result with more accuracy.
MILESTONE - 8	Making the output simpler and easier to understand and more efficient.
MILESTONE - 9	Deployed in the GitHub and waiting to review it.

6.2 SPRINT DELIVERY SCHEDULE

Project Planning Template (Product Backlog, Sprint Planning, Stories, Story points)

Use the below template to create product backlog and sprint 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 are collected.	20	High	4 Members
Sprint -2	Analyze	USN-2	As a health care provider all the data that are collected is cleaned and uploaded in the database or IBM cloud.	20	Medium	4 Members

Sprint -3	Dashboard	USN-3	As a health care provider, I can use my account in my dashboard for uploading dataset.	10	Medium	4 Members
Sprint -3	Visualization	USN-4	As a health care provider, I can prepare data for Visualization.	10	High	4 Members
Sprint -4	Visualization	USN-5	As a health care provider, I can present data in my dashboard.	10	High	4 Members
Sprint -4	Prediction	USN-6	As a health care provider, I can predict the length of stay	10	High	4 Members

Project Tracker, Velocity & Burndown Chart:

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint -1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022

Sprint -2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint -3	20	6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint -4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

Velocity:

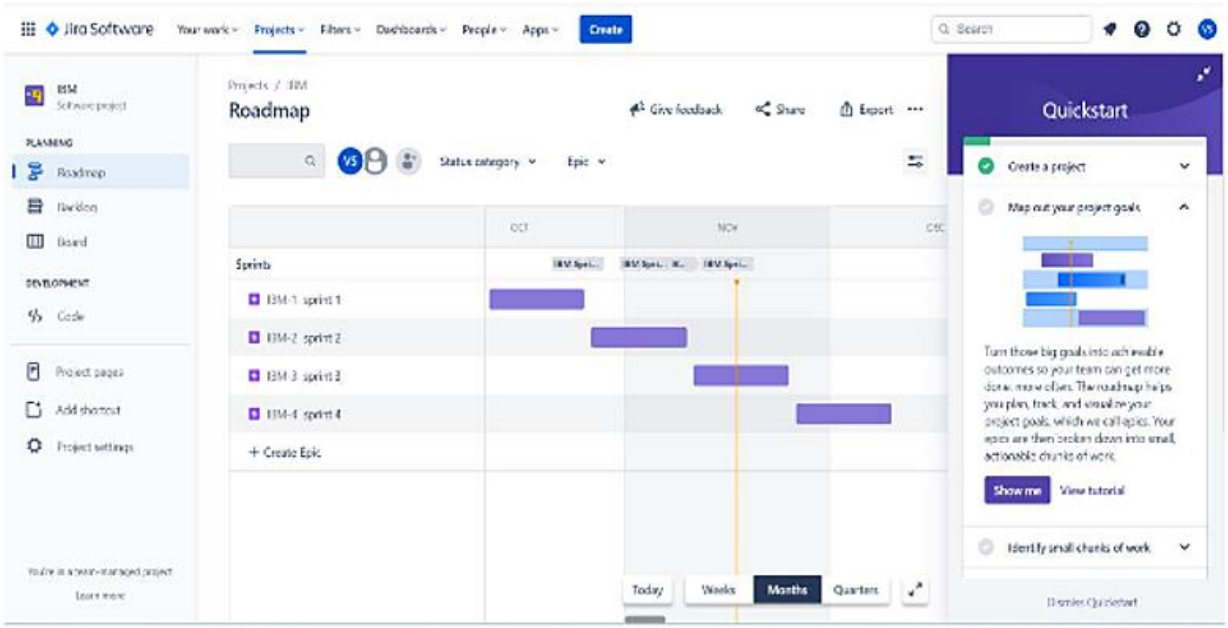
Imagine we have a 10-day sprint duration, and the velocity of the team is 20 (points per sprint). Let's calculate the team's average velocity (AV) per iteration unit (story points per day)

$$AV = \text{sprint duration} / \text{velocity}$$

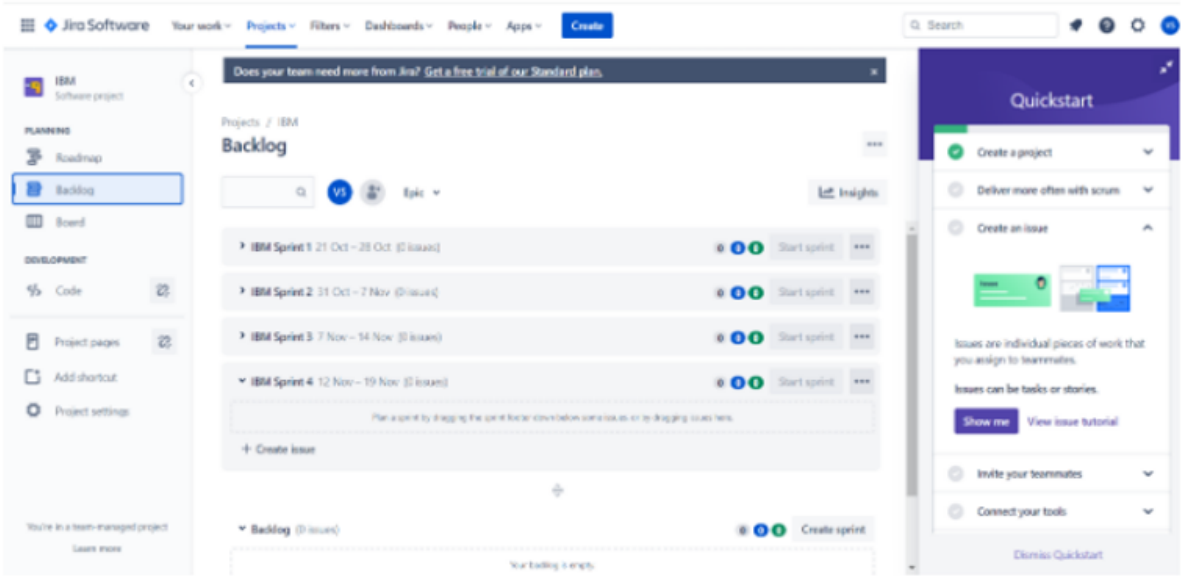
$$= 20 / 10 = 2$$

6.3 REPORTS FROM JIRA

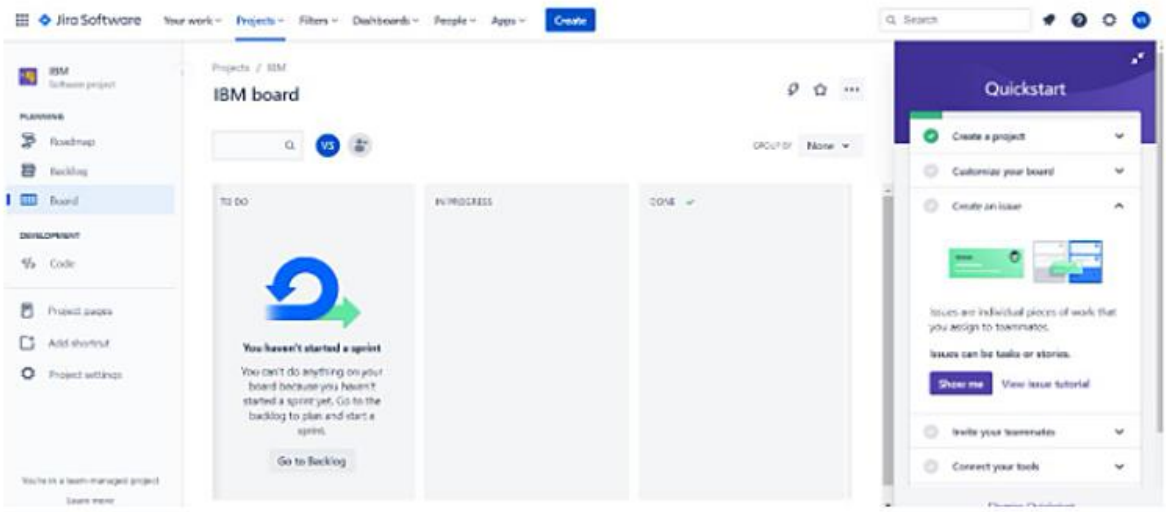
JIRA SOFTWARE



BACKLOG



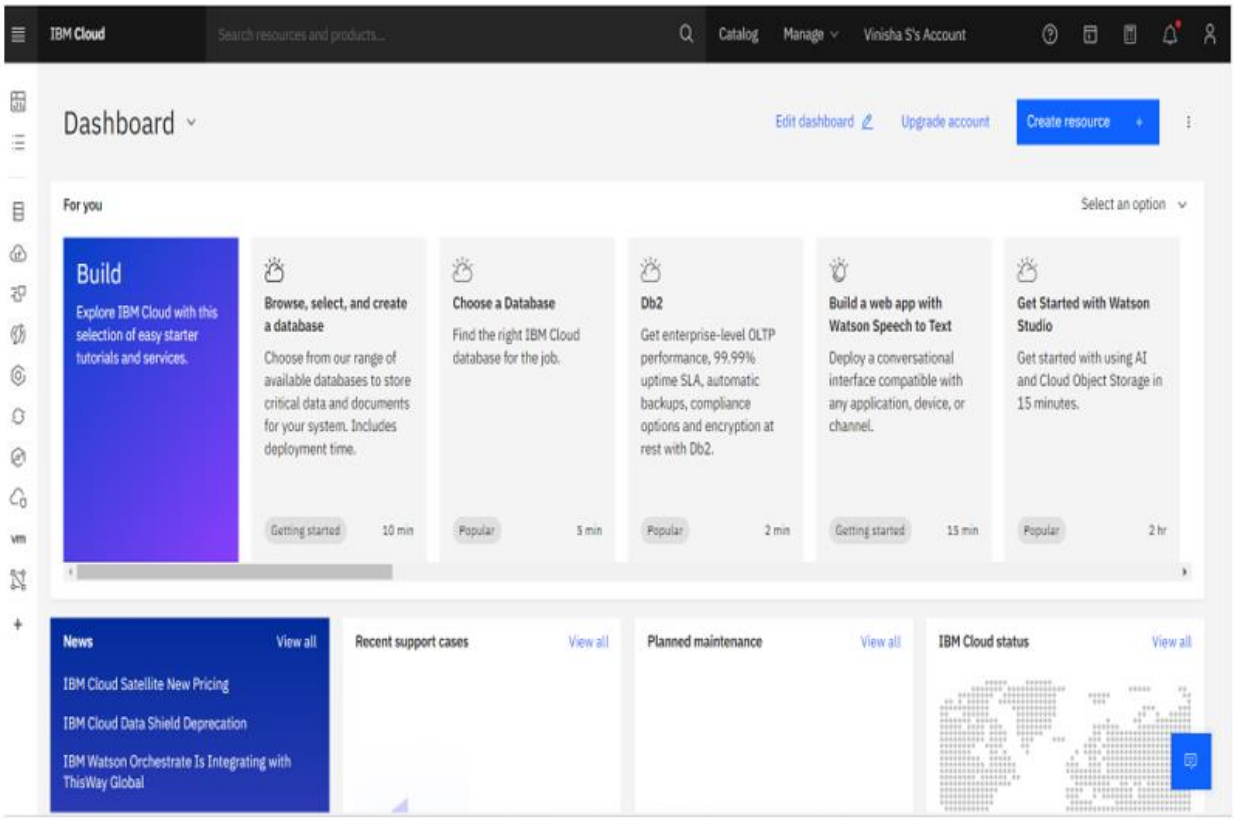
BOARD


Jira Software
Your work
Projects
Filters
Dashboards
People
Apps
Create
Search
IBM
Roadmap
Backlog
Board
Code
Project pages
Add shortcut
Project settings
Projects / IBM
IBM board
GROUP BY
None
TO DO
IN PROGRESS
DONE
You haven't started a sprint
You can't do anything on your board because you haven't started a sprint yet. Go to the backlog to plan and start a sprint.
Go to Backlog
Quickstart
Create a project
Customize your board
Create an issue
Issues are individual pieces of work that you assign to teammates.
Issues can be tasks or stories.
Show me
View issue tutorial
Invite your teammates
Connect your tools

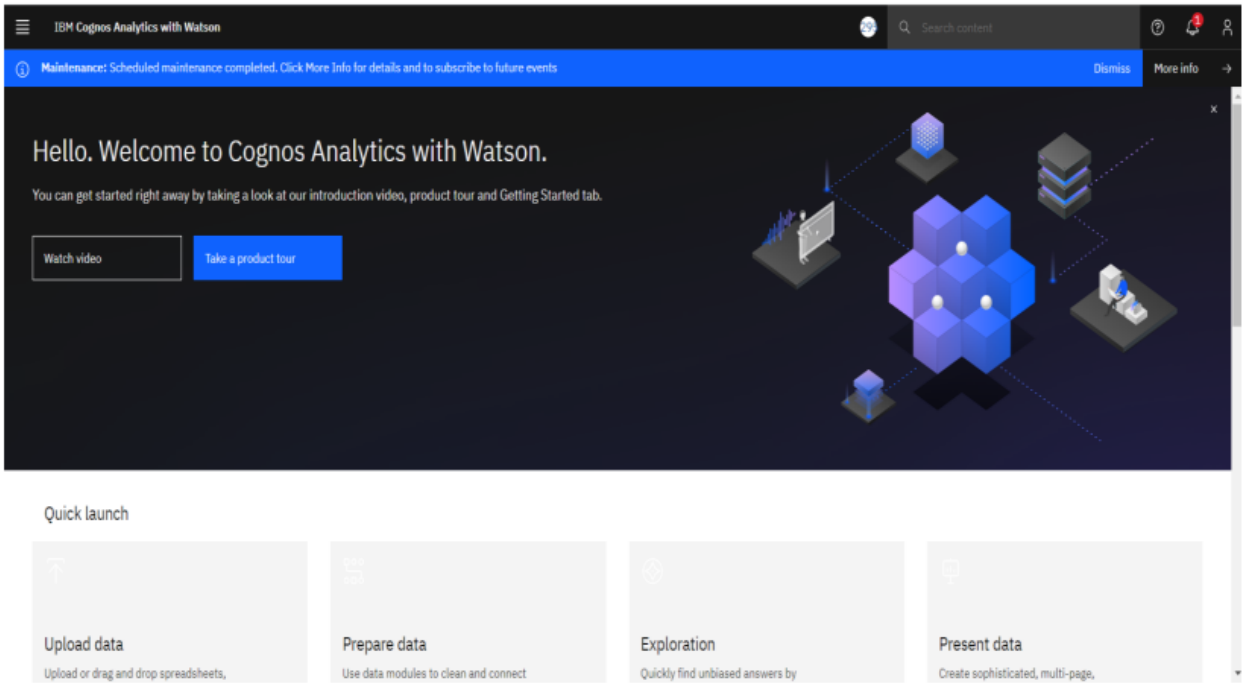
7.CODING & SOLUTIONING

7.1 CREATE AND CONFIGURE IBM CLOUD SERVICES

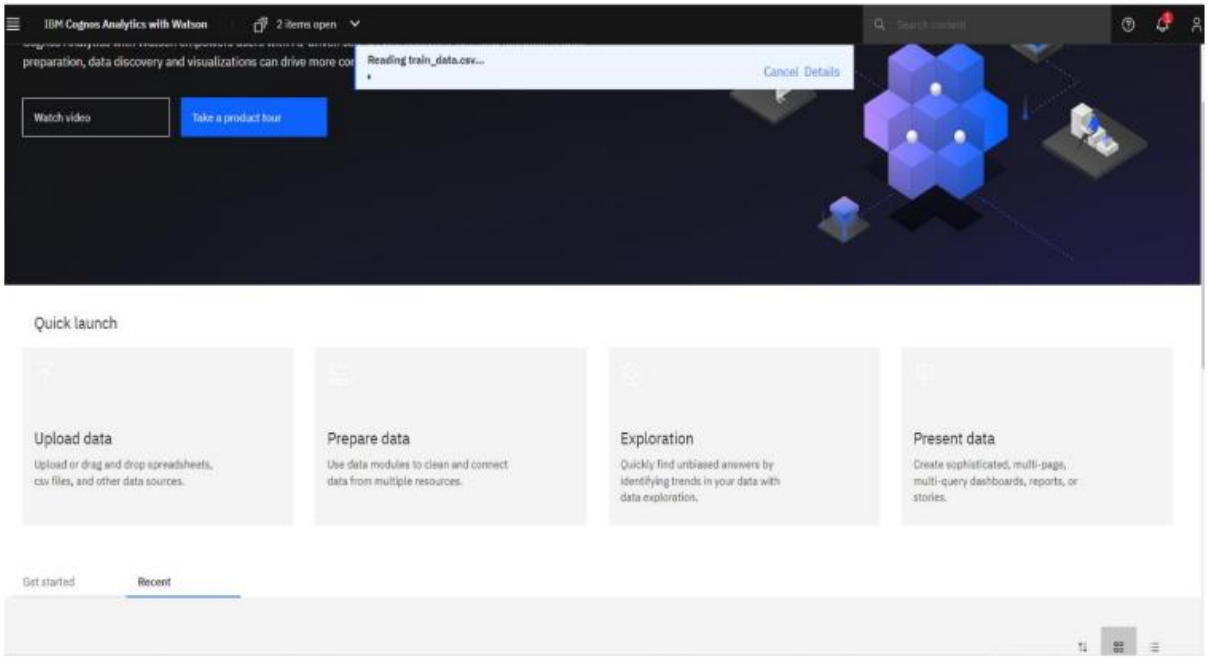
1: As a user, I will create IBM cloud account

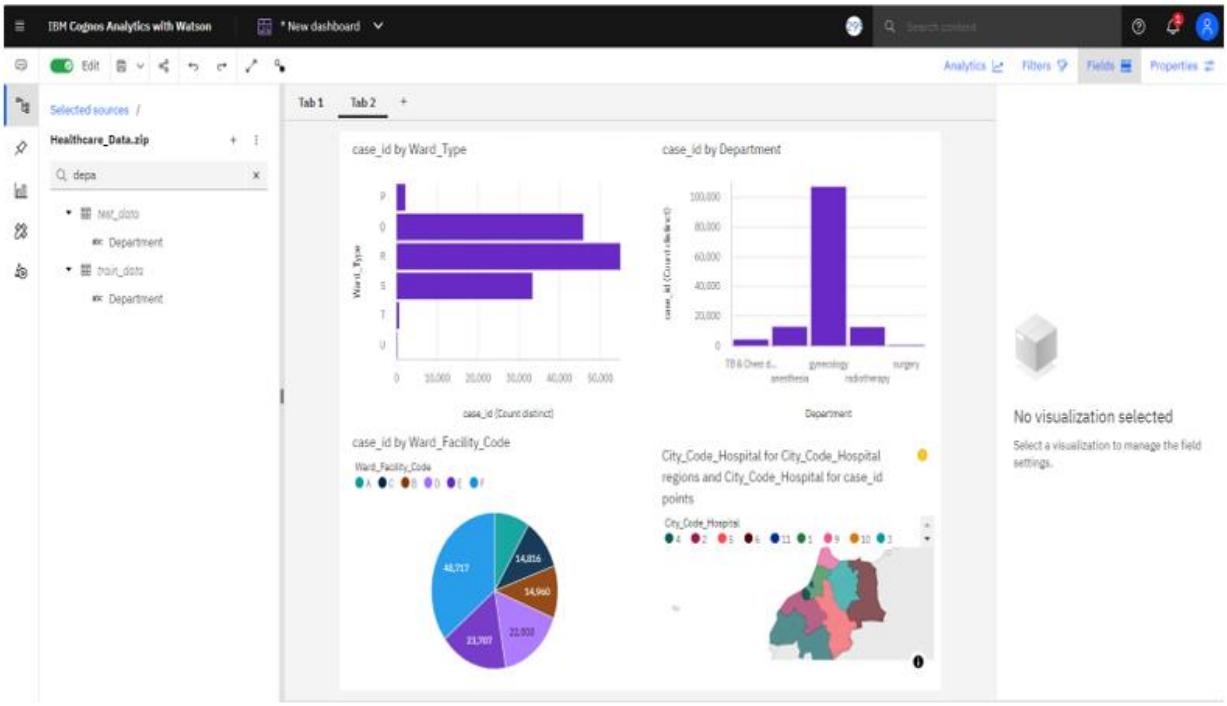


2: As a user, I will create IBM cognos account.

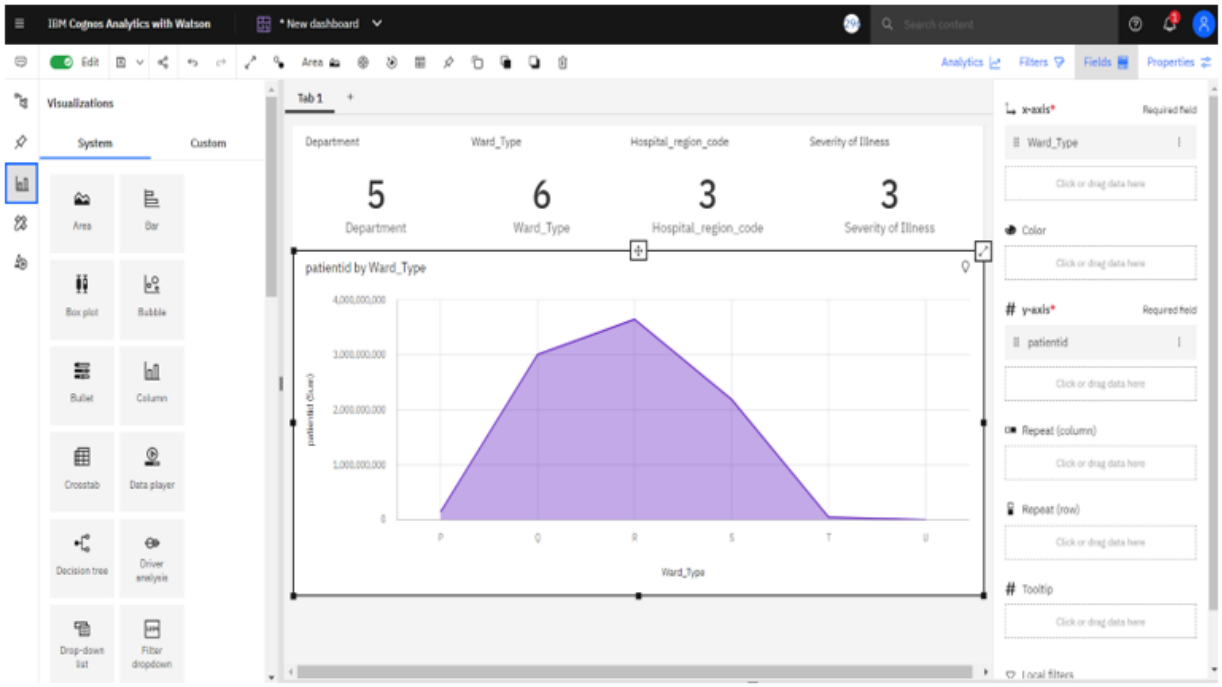


LOAD THE DATASET:

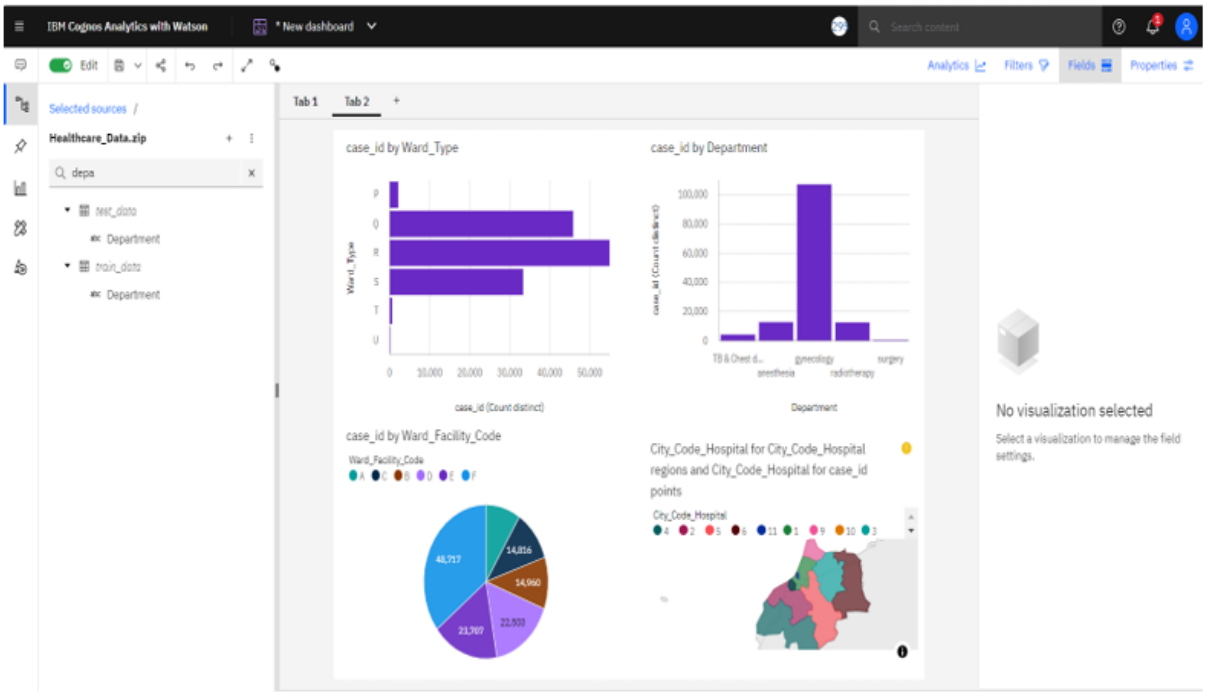




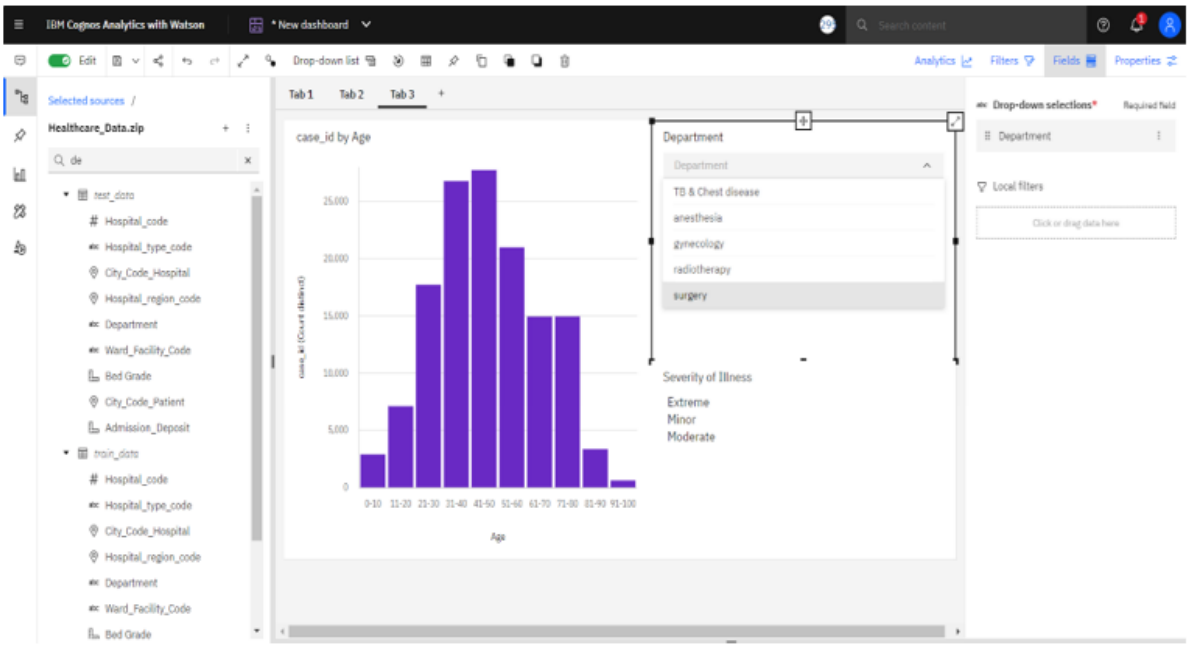
1)Number of Patients By Ward Types



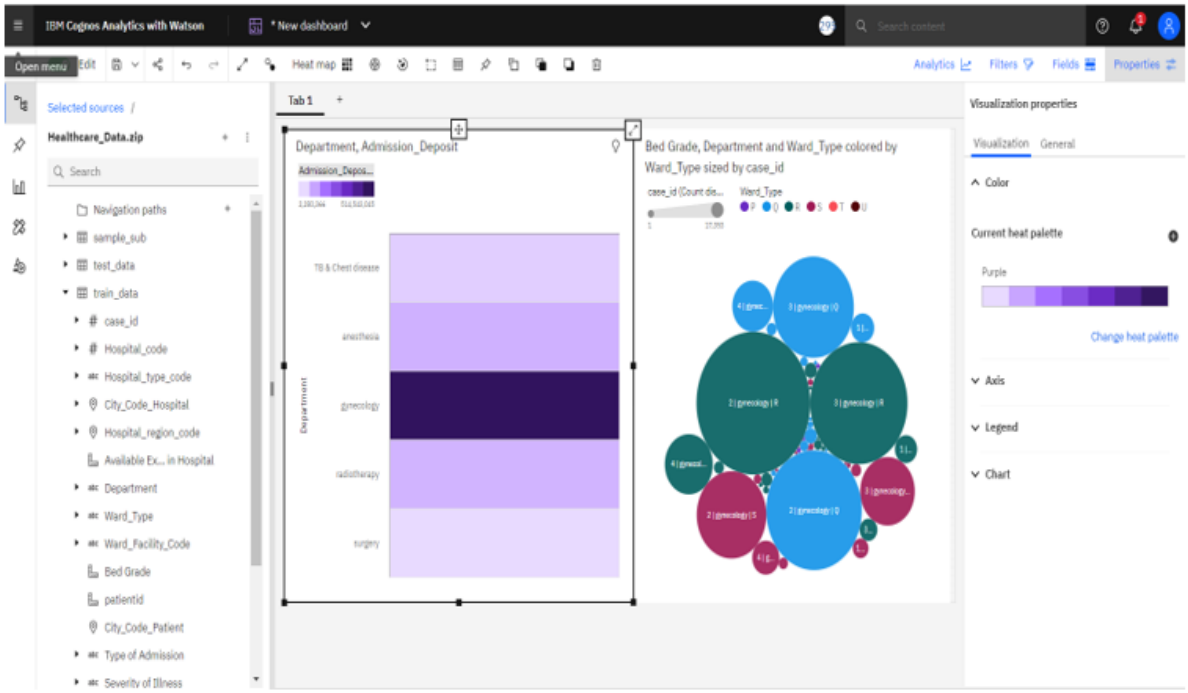
2) Dashboard To Show Number of patients



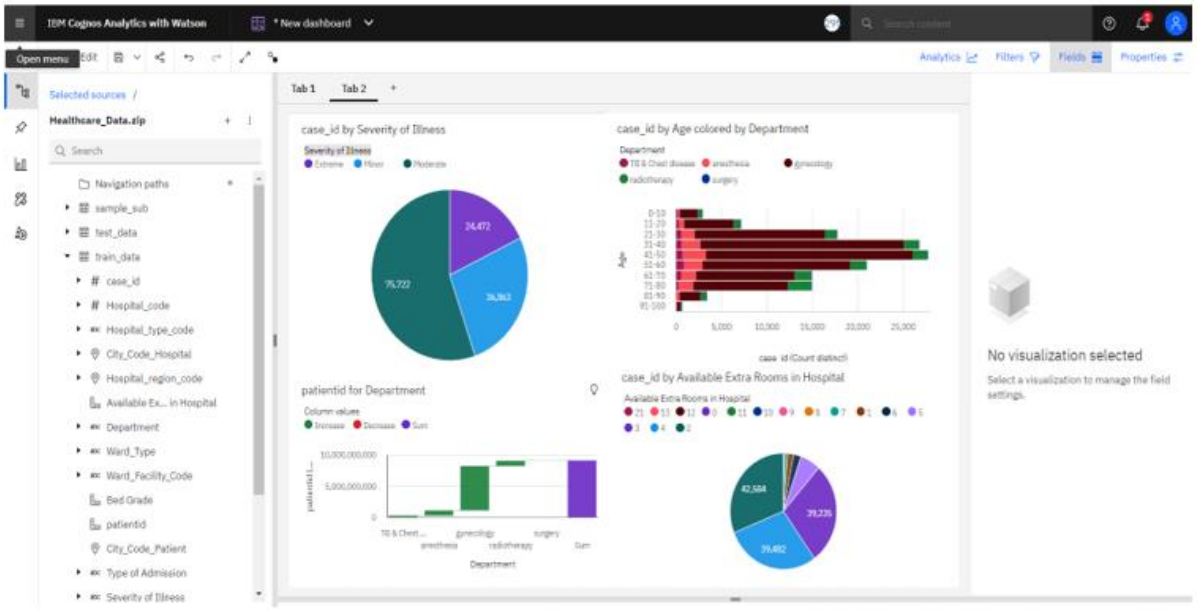
3) Age Wise Patients With Department And Severity Filters



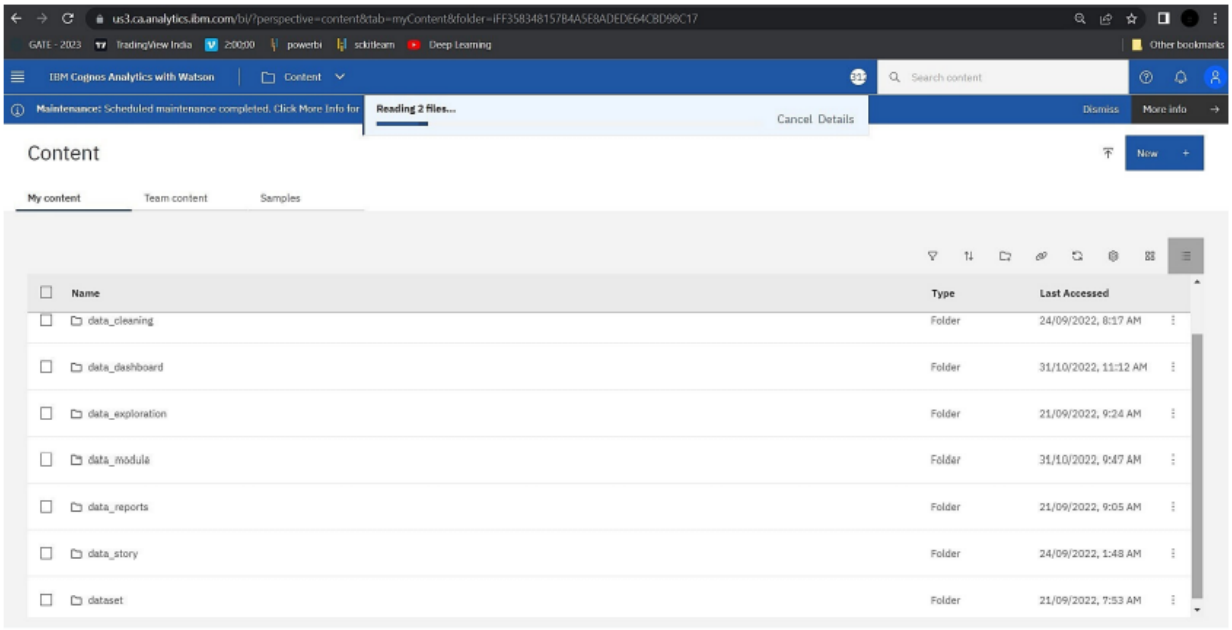
4) Dashboard with Hierarchy Bubble and Heat Map

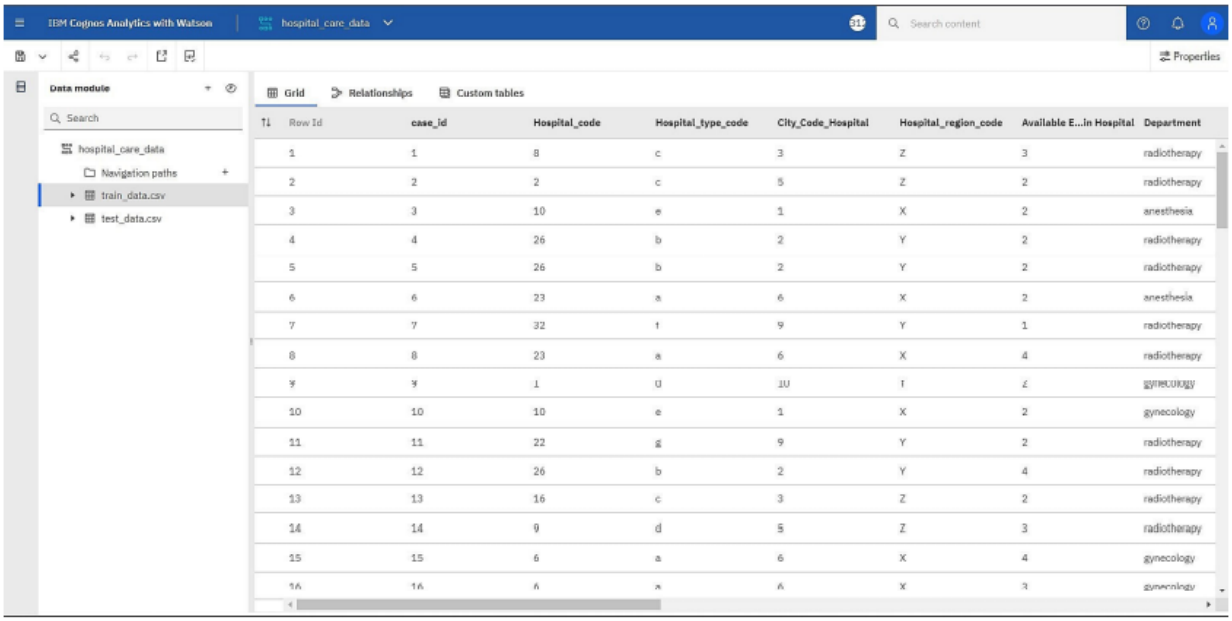


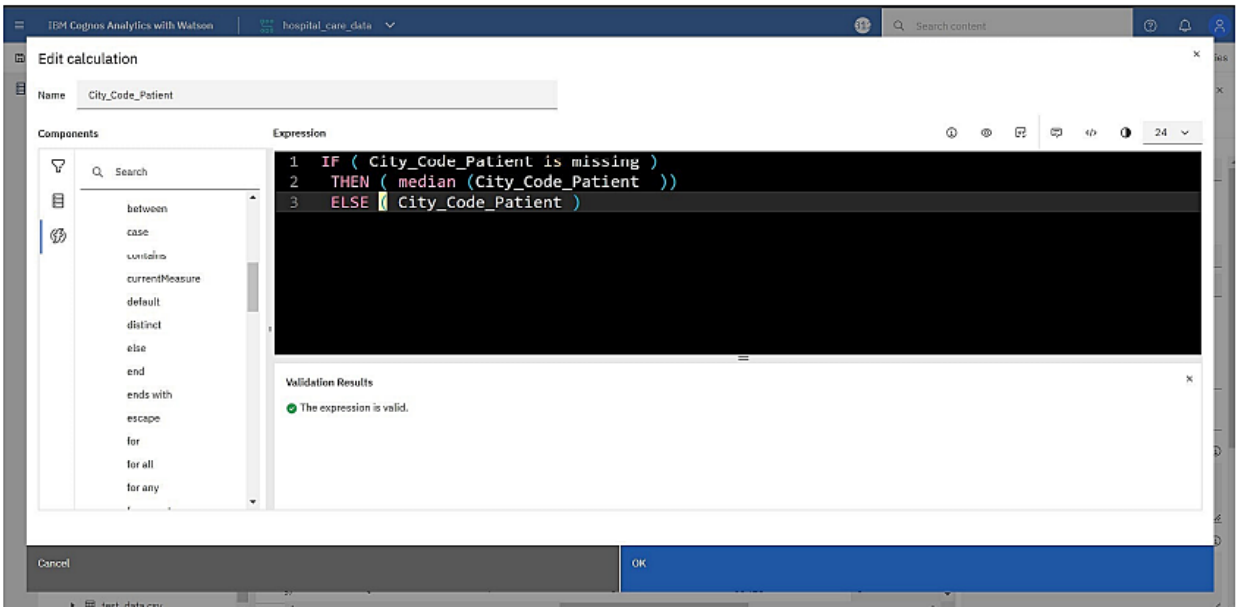
5) Dashboard Showing Pie, Stacked Bar, Waterfall and PieCharts



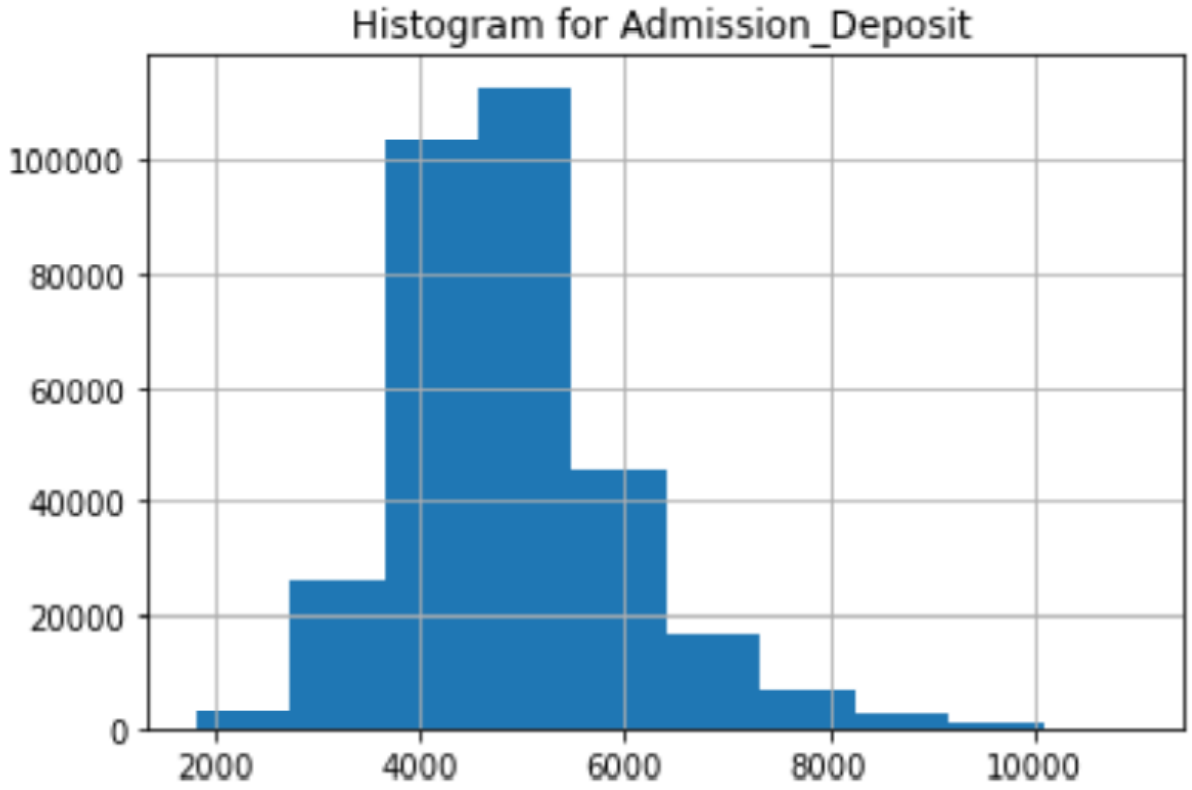
Preparing the dataset:




IBM Cognos Analytics with Watson
hospital_care_data
Search content
Properties
Data module
Search
hospital_care_data
Navigation paths
train_data.csv
test_data.csv
Grid
Relationships
Custom tables
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
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


IBM Cognos Analytics with Watson
hospital_care_data
Search content
Edit calculation
Name
City_Code_Patient
Components
Expression
Search
between
case
currentMeasure
default
distinct
else
end
ends with
escape
for
for all
for any
1 IF (City_Code_Patient is missing)
2 THEN (median (City_Code_Patient))
3 ELSE (City_Code_Patient)
Validation Results
The expression is valid.
Cancel
OK

```
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
```


Histogram for Admission_Deposit
100000
80000
60000
40000
20000
0
2000
4000
6000
8000
10000

8.TESTING

USER ACCEPTANCE TESTING

Purpose of Document :

The purpose of the document is we can analysis the data easily and the method of analysis are detailed in the document we can create an analytics Report, Dashboard and even the slide show story for Data analytics using cognos . We can analysis the data using the Python all the data are elaborated in the document, it is the main Purpose of the document.

Defect Analysis :

This report shows the number of resolved or closed bugs at each severity level,and how they were resolved

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	5	3	2	0	10
Duplicate	0	0	0	1	1
External	2	0	0	1	3
Fixed	6	2	0	0	8
Not Reproduced	0	1	1	0	2
Skipped	0	0	0	0	0
Won't Fix	0	0	0	0	0

Totals	13	6	3	2	24

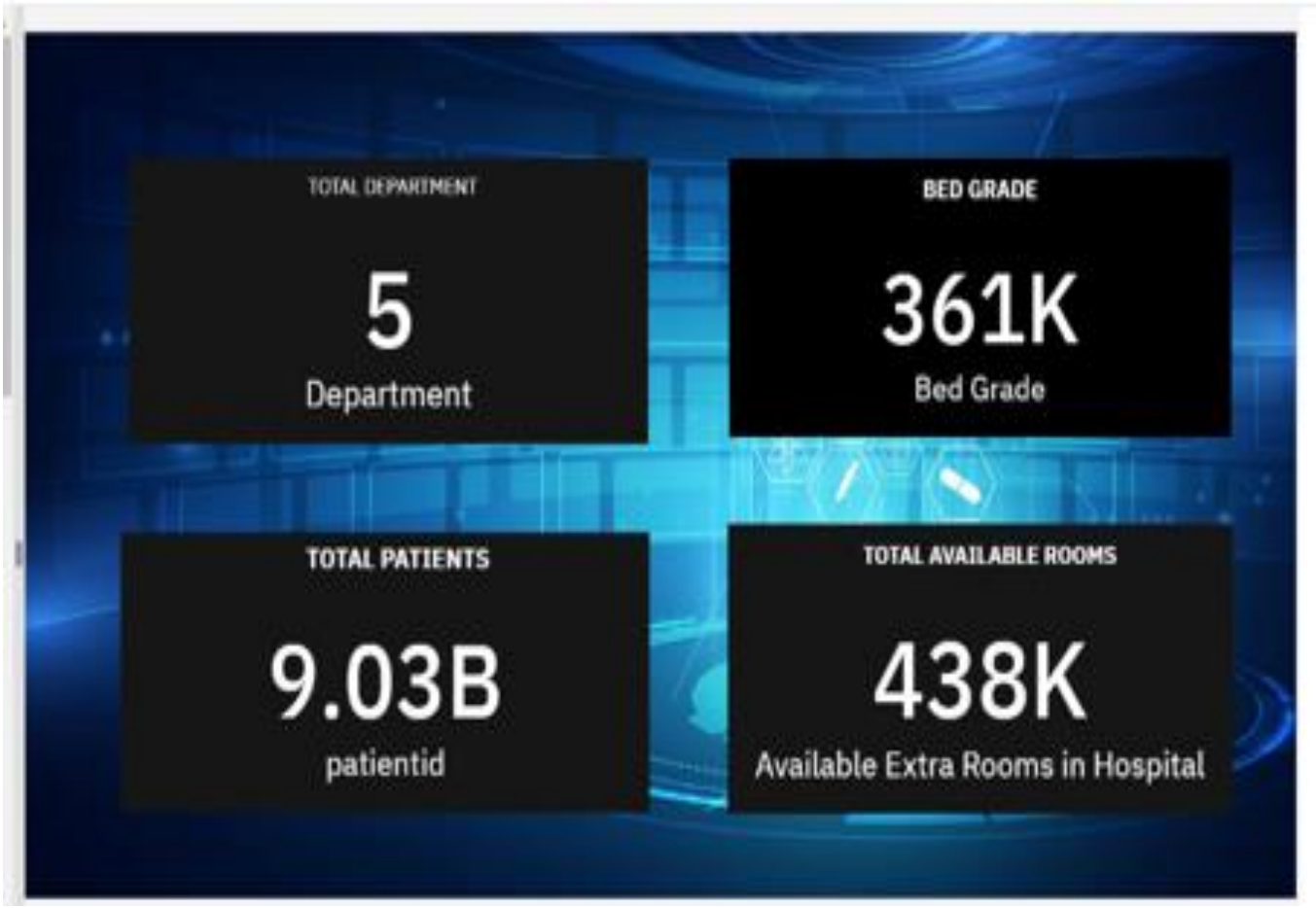
Test Case Analysis :

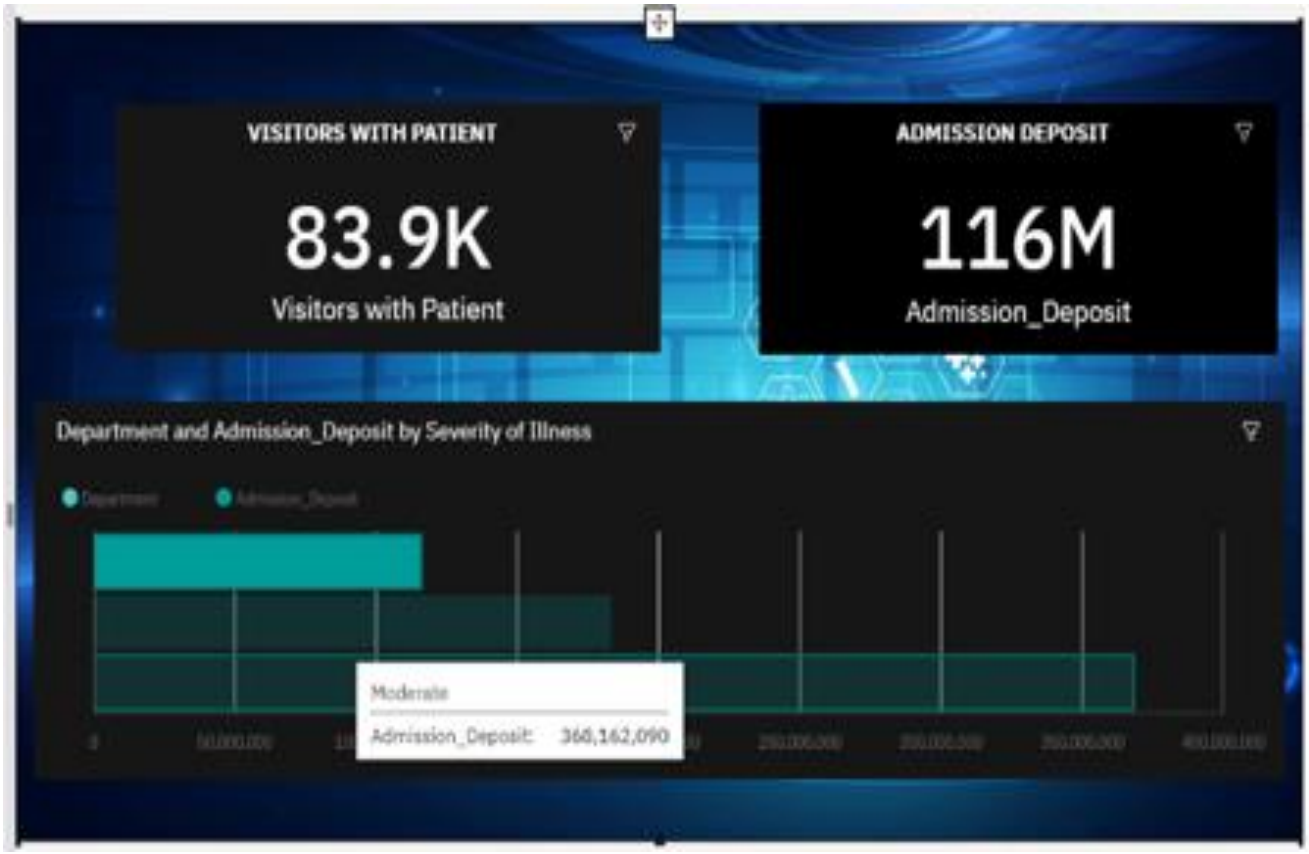
This report shows the number of test cases that have passed, failed, and untested.

Section	Total Cases	Not Tested	Fail	Pass
Print Engine	1	0	0	1
Client Application	2	0	0	2
Security	1	0	0	1
Outsource Shipping	1	0	0	1
Exception Reporting	1	0	0	1
Final Report Output	2	0	0	2
Version Control	2	0	0	2

9. RESULTS

9.1 PERFORMANCE METRICS




VISITORS WITH PATIENT
83.9K
Visitors with Patient
ADMISSION DEPOSIT
116M
Admission_Deposit
Department and Admission_Deposit by Severity of Illness
Moderate
Admission_Deposit: 368,162,090


Department for City_Code_Hospital regions
Department for Admission_Deposit regions
Tangier
Rabat
Casablanca
Fez
Morocco
Marrakesh
Agadir
Tinduf
Bechar

10.ADVANTAGES & DISADVANTAGES

ADVANTAGES

HIGHER-QUALITY CARE

Data draws from a number of sources, including previous doctor, pharmacy visits, social media, and other outside sources, it can create a more complete picture of a patient. A doctor may be able to see underlying causes for a health issue that wouldn't be easily visible with just basic health information.

EARLY INTERVENTION

The overall goal of data in healthcare is to use predictive analysis to find and address medical issues before they turn into larger problems. The doctor can then adjust the treatment to mitigate the risk, thus eliminating the problem before it becomes life threatening.

FRAUD DETECTION

A growing problem in the healthcare and insurance spaces is fraud, or patients submitting false claims in hopes of being paid. Using its advanced algorithms, it can sift through thousands of reports to find mistakes much more quickly than any team of humans could.

DISADVANTAGES

PRIVACY

One of the strongest negatives is the lack of privacy, especially when it comes to confidential medical records. According to reports, the technology takes away individual privacy for the greater good. And the

patient is being monitored continuously, it doesn’t give the patient’s freedom.

REPLACING DOCTORS

While seeing the ability to predict future medical issues as a positive, it also poses the risk of replacing doctors. Data analytics simply isn’t at the point yet where it can be used on its own, and it definitely lacks the personal touch of a human doctor. Some experts fear that the growth of data analytics could potentially undermine doctors and leave patients turning to technology for answers instead of using a licensed doctor.

11. CONCLUSION

Thus, healthcare has experienced much progress in usage and analysis of data. A large-scale digitalization and transparency in this sector is a key statement of almost all countries governments policies. For centuries, the treatment of patients was based on the judgment of doctors who made treatment decisions. In recent years, however, Evidence-Based Medicine has become more and more important as a result of it being related to the systematic analysis of clinical data and decision-making treatment based on the best available information. In the healthcare sector, Data Analytics is expected to improve the quality of life and reduce operational costs. It also helps identify data that provides insightful insights for current as well as future decisions

It is believed that the implementation of data analytics by healthcare organizations could bring many benefits in the upcoming years, including lowering health care costs, better diagnosis and prediction of diseases and their spread, improving patient care and developing protocols to prevent re-hospitalization, optimizing staff, optimizing equipment, forecasting the need for hospital beds, operating rooms, treatments, and improving the drug supply chain.

12. FUTURE SCOPE

The advantages of implementing data analytics have undoubtedly caught the attention of the healthcare industry as they try to find more effective methods for the quality of service.

Data analysts can develop software to automatically inform patients about recommended lifestyle changes to prevent certain conditions. This helps improve performance by delivering data-based quality patient care which, in turn, improves patient satisfaction.

Mobile health applications such as telehealth and wearables like Fit Bit, will also grow and provide consumers with healthcare analytics, services and information instantly.

13. APPENDIX

GITHUB & PROJECT DEMO LINK

GITHUB LINK:

https://github.com/IBM-EPBL/IBM-Project-41157-1660639875.git

PROJECT DEMO LINK:

https://us1.ca.analytics.ibm.com/bi/?perspective=story&pathRef=.my_folders%2FNew%2Bstory&action=view&sceneId=model00000184890eef23_00000000&sceneTime=0