# PROJECT REPORT

TEAM ID: PNT2022TMID23028

PROJECT NAME: ANALYTICS FOR HOSPITAL'S HEALTH CARE DATA

TEAM LEADER: NISHANTHI.V (913119104061)

**TEAM MEMBERS:** 

MATHUMITHA M (913119104053)

KEERTHIKA R (913119104044)

ISHA PARASU B (913119104033)

# **INDEX**

### 1. INTRODUCTION

- 1.1 Project Overview
- 1.2 Purpose

### 2. LITERATURE SURVEY

- 2.1 Existing problem
- 2.2 References
- 2.3 Problem Statement Definition

# 3. IDEATION & PROPOSED SOLUTION

- 3.1 Empathy Map Canvas
- 3.2 Ideation & Brainstorming
- 3.3 Proposed Solution
- 3.4 Problem Solution fit

### 4. REQUIREMENT ANALYSIS

- 4.1 Functional requirement
- 4.2 Non-Functional requirements

### 5. PROJECT DESIGN

- 5.1 Data Flow Diagrams
- 5.2 Solution & Technical Architecture
- 5.3 User Stories

### 6. PROJECT PLANNING & SCHEDULING

- 6.1 Sprint Planning & Estimation
- 6.2 Sprint Delivery Schedule
- 6.3 Reports from JIRA

### 7. CODING & SOLUTIONING

- 7.1 Feature 1
- 7.2 Feature 2

### 8. TESTING

- 8.1 Test Cases
- 8.2 User Acceptance Testing

# 9. RESULTS

9.1 Performance Metrics

### 10. ADVANTAGES & DISADVANTAGES

- 11. CONCLUSION
- 12. FUTURE SCOPE
- 13. APPENDIX

Source Code

GitHub & Project Demo Link

# **Project Report**

### 1. INTRODUCTION

### 1.1 Project Overview

The healthcare industry generates a tremendous amount of data but struggles to convert that data into insights that improve patient outcomes and operational efficiencies. Data analytics in healthcare is intended to help providers overcome obstacles to the widespread application of data-derived intelligence:

- -Making healthcare data easier to share among colleagues and external partners, and easier to visualize for public consumption
- -Providing accurate data-driven forecasts in real time to allow healthcare providers to respond more quickly to changing healthcare markets and environments
- -Enhancing data collaboration and innovation among healthcare organizations to convert analytics-ready data into business-ready information by automating low-impact data management tasks

The tools used in analytics fall into three general categories:

- -Software that acquires the data from sources that include patient surveys, case files, and machine-to-machine data transfers
- -Programs that clean, validate, and analyze the data in response to a specific research question
- -Software that builds on the results of the analysis to suggest various actions to achieve specific healthcare goals

Along with collecting, analyzing, and interpreting data, analytics software must secure the data and the analysis results while ensuring that the healthcare professionals who'll benefit from the insights have ready access to the information in a form that they can easily use in their work.

### 1.2 Purpose

Data analytics in healthcare can be applied to every aspect of patient care and operations management. The analyses investigate methods of improving the provision of clinical care, enhancing disease prevention, and measuring the effectiveness of various treatment options.

The ability of data analytics to convert raw healthcare data into actionable intelligence is expected to have the greatest impact on these areas of healthcare:

- -Research and prediction of disease
- -Automation of hospital administrative processes
- -Early detection of disease
- -Prevention of unnecessary doctor's visits
- -Discovery of new drugs
- -More accurate calculation of health insurance rates
- -More effective sharing of patient data
- -Personalization of patient care

An example of the use of data analytics in a clinical setting is the library of molecular and clinical data that Tempus is creating to give healthcare professionals a clearer clinical context for a cancer patient's unique case. The data is collected from lab reports, clinical notes, radiology and pathology images, and

accelerated cancer research. The project's goal is to provide oncologists with tools to help them personalize their patients' treatments.

Public health officials are using data analytics to improve the prevention of disease and identify members of the population who are most at risk. For example, Linguamatics has created an analytics engine that uses natural language processing to analyze the unstructured patient data in health records to identify lifestyle factors that its predictive models associate with high-risk patients.

### 2. LITERATURE SURVEY

### 2.1 Existing problem

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.

#### 2.2 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.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.
- 4.Castro EM, Van Regenmortel T, Vanhaecht K, Sermeus W, Van Hecke A. Patient empowerment, patient participation and patient-centeredness in hospital care: a concept analysis based on a literature review. Patient Educ Couns. 2016;99(12):1923–39.
- 5.Corsi A, de Souza FF, Pagani RN, et al. Big data analytics as a tool for fighting pandemics: a systematic review of literature. J Ambient Intell Hum Comput. 2021;12:9163–80. https://doi.org/10.1007/s12652-020-02617-4.
- 6.Fredriksson C. Organizational knowledge creation with big data. A case study of the concept and practical use of big data in a local government context. 2016.

https://www.abo.fi/fakultet/media/22103/fredriksson.pdf.

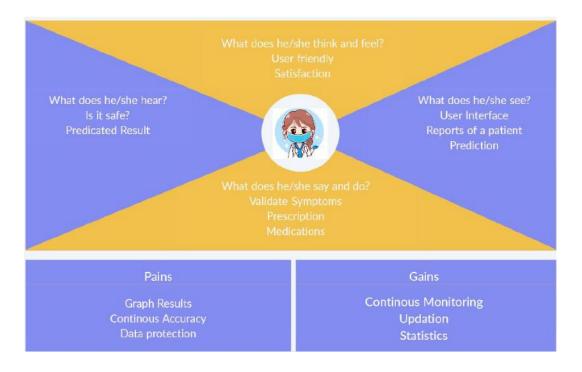
### 2.3 Problem Statement Definition

The main purpose 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.

#### 3. IDEATION & PROPOSED SOLUTION

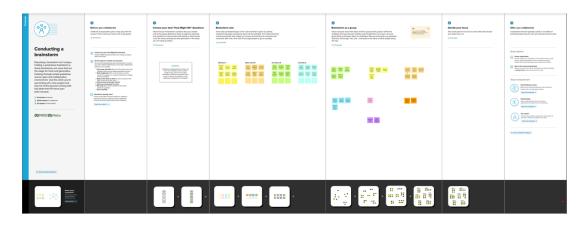
### 3.1 Empathy Map Canvas

An empathy map is a collaborative tool teams can use to gain a deeper insight into their customers. Much like a user persona, an empathy map can represent a group of users, such as a customer segment. The empathy map was originally created by Dave Gray and has gained much popularity within the agile community.



### 3.2 Ideation & Brainstorming

Ideation is the process where you generate ideas and solutions through sessions such as Sketching, Prototyping, Brainstorming, Brainwriting, Worst Possible Idea, and a wealth of other ideation techniques.

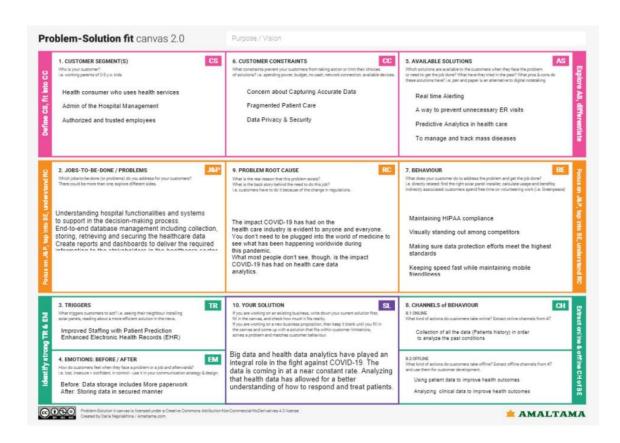


# **3.3 Proposed Solution**

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	The impact COVID-19 has had on the health care industry is evident to anyone and everyone. You don't need to be plugged into the world of medicine to see what has been happening worldwide during this pandemic. What most people don't see, though, is the impact COVID-19 has had on health care data analytics. "Big data tools have played an increasingly significant role in health care decision-making" says health care analytics. It is not just providers, but lawmakers and researchers who are turning to big data analytics and predictive models to help allocate resources, predict surges, improve patient care and outcomes and employ preventive measures.
2.	Idea / Solution description	Big data and health data analytics have played an integral role in the fight against COVID-19. The data is coming in at a near constant rate. Analyzing that health data has allowed for a better understanding of how to respond and treat patients.
3.	Novelty / Uniqueness	Deploying a healthcare analytics suite can help healthcare providers leverage data for insights in several areas of operations. One major area where using analytics can optimize efforts is the management of hospital and foundation

		donations and grants.  For many healthcare providers, donations are the basis of their yearly budgets, so organizing and tracking expenses and activity is vital for setting appropriate goals.  Moreover, it can help track donor engagement, retention, and previous contributions.
4.	Social Impact / Customer Satisfaction	Health care is expensive and those costs only continue to increase across the board. We are, however, seeing a shift from fee-for-service payment models to value-based care.  Through the use of predictive and prescriptive analytics, health care organizations and practitioners can get detailed models for lowering costs and patient risk. In addition to the patient-centric benefits mentioned above, health data analytics can reduce appointment no-shows, manage supply chain costs, prevent equipment breakdowns and decrease fraud.
5.	Business Model (Revenue Model)	Your salary, much like your role responsibilities, will vary based on the industry or organization for which you choose to work. Payscale lists the median salary for a health data analyst as \$63,000 per year. The U.S. Bureau of Labour Statistics doesn't have an entry specific to Health Data Analysts, however, the Medical and Health Services Manager page does list "health information manager." The median pay listed here is closer to \$100,000.
6.	Scalability of the Solution	Big Data Analytics can provide insight into clinical data and thus facilitate informed decision-making about the diagnosis and treatment of patients, prevention of diseases or others. Big Data Analytics can also improve the efficiency of healthcare organizations by realizing the data potential

### 3.4 Problem Solution fit



### 4. REQUIREMENT ANALYSIS

# **4.1 Functional requirement**

A function of software system is defined in functional requirement and the behavior of the system is evaluated when presented with specific inputs or conditions which may include calculations, data manipulation and processing and other specific functionality.

- Our system should be able to load air quality data and preprocess data.
- It should be able to analyze the air quality data.
- It should be able to group data based on hidden patterns.
- It should be able to assign a label based on its data groups.
- It should be able to split data into trainset and testset.
- It should be able to train model using trainset.
- It must validate trained model using testset.
- It should be able to display the trained model accuracy.
- It should be able to accurately predict the air quality on unseen data.

FUNCTI ON NUMBER	FUNCTIONAL REQUIREMENT	SUB REQUIREMENT
FN.NO-1	USER REGISTRATION/LOGIN	REGISTER AFTER OPENING THE APPLICATION
FN.NO-2	USER DETAIL	PROVIDE DETAIL ABOUT HEALTH STATUS AFTER REGISTRATION BY SELECTING AND SPECIFY ABOUT IT
FN.NO-3	CAPTURING IMAGE	CAPTURE THE IMAGE AND CHECK THE PARAMETERS OF THE CAPTURED IMAGE.
FN.NO-4	IMAGE PROCESSING	UPLOAD THE IMAGE FOR PROCESSING
FN.NO-5	IMAGE IDENTIFICATION	IDENTIFY THE DISEASE FROM THE IMAGE WHICH HAS BEEN CAPTURED
FN.NO-6	IMAGE DESCIPTION	PROVIDE THE PRESCRIPTION FOR THE PATIENT

# **4.2 Non-Functional requirements**

Nonfunctional requirements describe how a system must behave and establish constraints of its functionality. This type of requirements is also known as the system's quality attributes. Attributes such as performance, security, usability, compatibility are not the feature of the system, they are a required characteristic. They are "developing" properties that emerge from the whole

arrangement and hence we can't compose a particular line of code to execute them. Any attributes required by the customer are described by the specification. We must include only those requirements that are appropriate for our project.. Some Non-Functional Requirements are as follows:

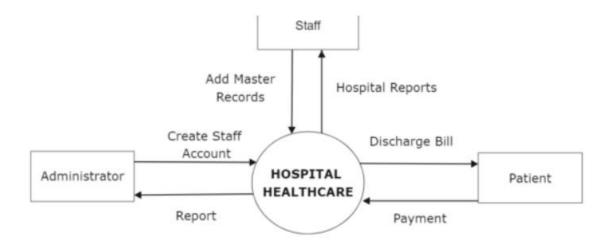
- Reliability
- Maintainability
- Performance
- Portability
- Scalability
- Flexibility

NON - FUNCTI ON NUMBER	NON- FUNCTIONAL REQUIREMENT	SUB REQUIREMENT
NFN.NO-1	USABILITY	DATASET OF ALL USERS TO IDENTIFY THE STATUS AND DETAILS OF THE PATIENT.
NFN.NO-2	SECURITY	THE INFORMATION AND THE HEALTH STATUS OF USER AND DETAILS ARE SECURED HIGHLY.
NFN.NO-3	RELIABILITY	THE IMAGE QUALITY IS IMPORTANT TO PROVIDE THE HEALTH DETAILS ABOUT THE PATIENT.
NFN.NO-4	PERFORMANCE	PERFORMANCE IS BASED ON THE PATIENT STATUS AND THE DISEASE THAT IS SCANED.
NFN.NO-5	AVAILABILITY	IT IS AVAILABLE FOR ALL THE USER TO CHECK THE STATUS ABOUT THEIR HEALTH AND GET THE PRESCRIPTION.
NFN.NO-6	SCALABILITY	INCREASE THE PREDICTION OF DETAILS OF THE PATIENTS.

### 5. PROJECT DESIGN

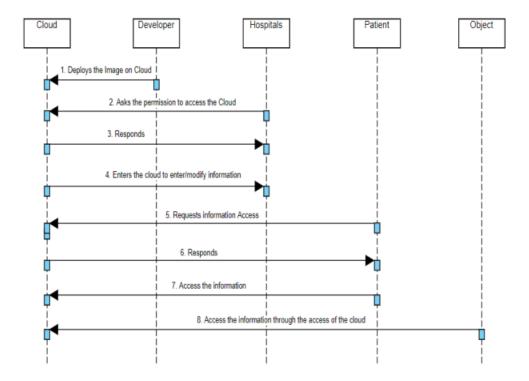
### **5.1 Data Flow Diagrams**

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:



# Technical Architecture:

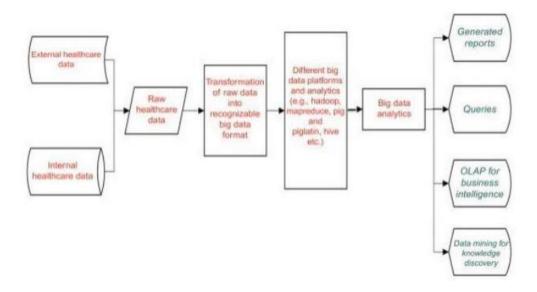


Table-1 : Components & Technologies:

S.No			
	Component	Description	Technology
1.	User Interface	The user interacts with application using Web UI	HTML, CSS, JavaScript
2.	Data Processing	The data from the dataset is pre-processed	IBM Cognos Analytics
3.	Cloud Database	The clean dataset is stored on IBM Cloud	IBM Cloud
4.	Data visualization	The data is visualized into different forms	The data is visualized into different forms
5.	Prediction	These Algorithm techniques are used to predict the proper way to make the stock in store.	ML algorithms –Logistic Regression, Linear Regression, Random Forest,ABC Techniques.

Table-2: Application Characteristics:

S.No			
	Characteristics	Description	Technology
1.	Open-Source Frameworks	Open-source frameworks used	IBM Cognos Analytics, Python
2.	Security Implementations	Request authentication using Encryptions	Encryptions
3.	Scalable Architecture	Scalability consists of 3-tiers	Web Server – HTML, CSS, Javascript Application Server – Python Database Server – IBM Cloud
4.	Availability	The application is available for cloud users	IBM Cloud Hosting
5.	Performance	The user can know how to maintain the inventory to increase profits.	ML algorithms.

# **5.3** User Stories

.

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Administrator (Admin)	Login		As a admin, I can add users details.	I can provide username and password for the users.	High	Sprint-1
User	Login	USN-1	As a user, I can login into the website/ application using username and password	I can access Dashboard and user account.	High	Sprint-1
	Dashboard	USN-2	As a user, I can add Patient Details like Patient name, contact number, age etc.	I can view and edit patient details.	Medium	Sprint-1
	Dashboard	USN-3	As a user, I can add bed details, Doctor details and other hospital detail.	I can monitor Staff and hospital performance.	Medium	Sprint-1
	Dashboard	USN-4	As a user, I can upload patient medical reports.	I can upload and Modify patient reports	High	Sprint-1

# 6. PROJECT PLANNING & SCHEDULING

# **6.1 Sprint Planning & Estimation**

Sprint	Functional Requirem ent (Epi c)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	2	High	
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	1	High	
Sprint-1		USN-3	As a user, I can register for the application through Gmail	2	Medium	
Sprint-1	Login	USN-4	As a user, I can log into the application by entering email	1	High	
	Dashboard	USN-5	As a user/patient I can request for Organ/Blood Donation .	2	Medium	
		USN-6	As a user/donor I can view my appointments	1	High	

# **6.2 Sprint Delivery Schedule**

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-	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov2022
Sprint-	20	6 Days	07 Nov 2022	12 Nov 2022	20	07 Nov 2022
Sprint-	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

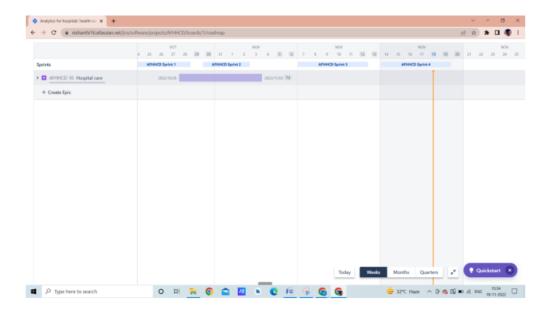
### **Burndown Chart:**

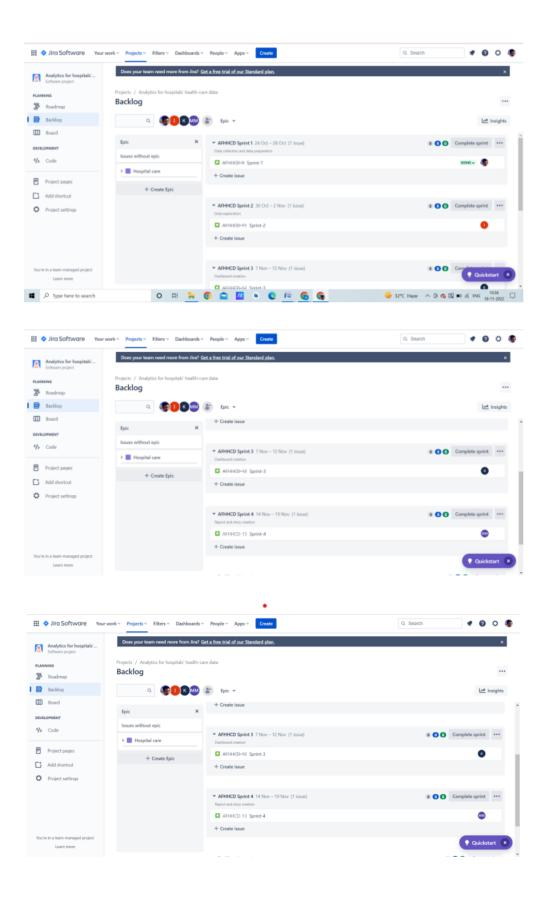
A burn down chart is a graphical representation of work left to do versus time. It is often used in agile software development methodologies such as Scrum. However, burn down charts can be applied to any project containing measurable progress over time



# **6.3 Reports from JIRA**

# Road Maps:



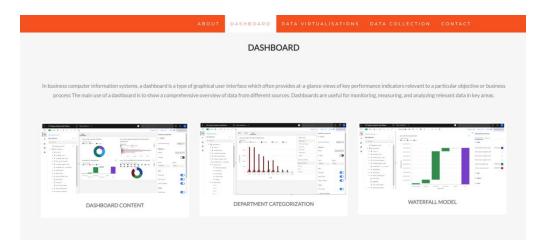


### 7. CODING & SOLUTIONING

### 7.1 Feature 1 – Dashboard

In business computer information systems, a dashboard is a type of graphical user interface which often provides at-a-glance views of key performance indicators relevant to a particular objective or business process

The main use of a dashboard is to show a comprehensive overview of data from different sources. Dashboards are useful for monitoring, measuring, and analyzing relevant data in key areas.



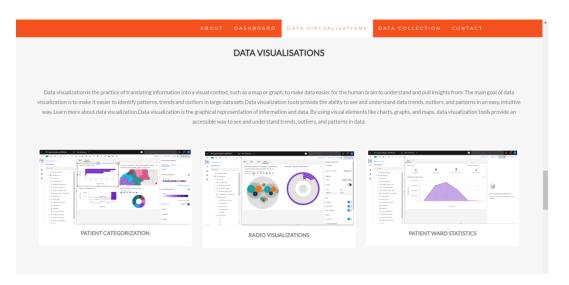
### Code:

```
<div id="services" class="container-fluid text-center bg-grey">
  In business computer information systems, a dashboard is a type of graphical user interface
which often provides at-a-glance views of key performance indicators relevant to a particular
objective or business process
    The main use of a dashboard is to show a comprehensive overview of data from different sources.
Dashboards are useful for monitoring, measuring, and analyzing relevant data in key areas.</p>
  <div class="col-sm-4">
     <div class="thumbnail">
       <img src="C:\Users\gowtham s\OneDrive\Desktop\code\dashboard.png" alt="db" width="500"</pre>
height="400">
        <strong>DASHBOARD CONTENT</strong>
    <div class="col-sm-4">
       <div class="thumbnail">
         <img src="C:\Users\gowtham s\OneDrive\Desktop\code\dept.png" alt="dept" width="500"</pre>
          <strong>DEPARTMENT CATEGORIZATION</strong>
     <div class="col-sm-4">
       <div class="thumbnail">
```

#### **7.2** Feature 2 – Data Visualizations

Data visualization is the practice of translating information into a visual context, such as a map or graph, to make data easier for the human brain to understand and pull insights from. The main goal of data visualization is to make it easier to identify patterns, trends and outliers in large data sets

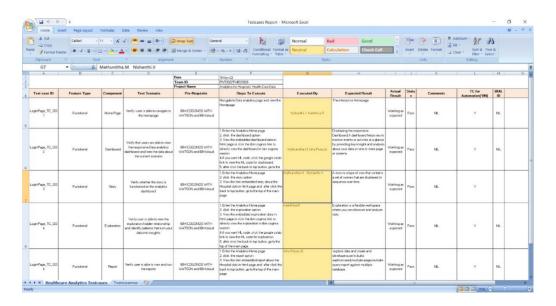
Data visualization tools provide the ability to see and understand data trends, outliers, and patterns in an easy, intuitive way. Learn more about data visualization. Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.



### Code:

### 8. TESTING

### **8.1** Test Cases



# **8.2** User Acceptance Testing

# 1. 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.

### 2.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	Subtot al
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

# 3. Test Case Analysis:

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

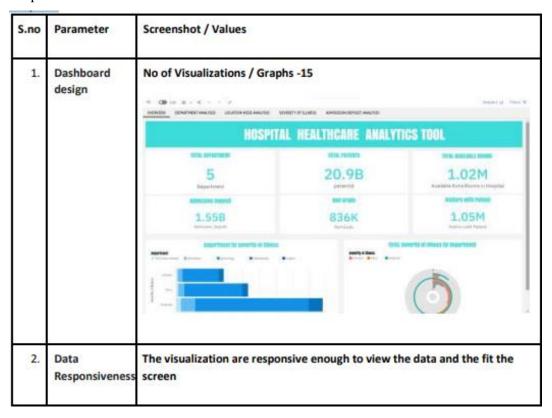
Section	Total Cases	Not Tested	Fai 1	Pass
Print Engine	1	0	0	1
Client Application	2	0	0	2
Security	1	0	0	1
Outsource Shipping	1	0	0	1
Exception Reporting	1	0	0	1
Final Report Output	2	0	0	2
Version Control	2	0	0	2

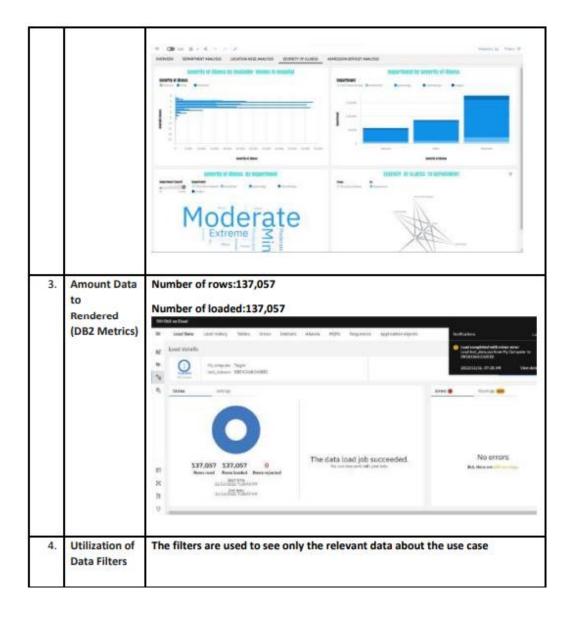
# 9. RESULTS

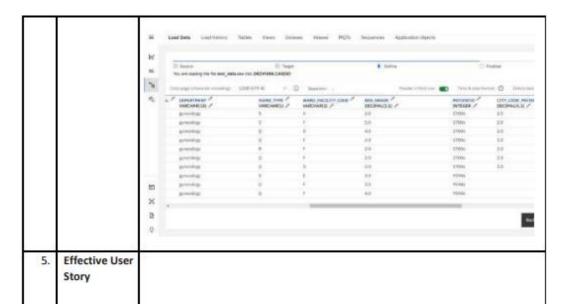
# **9.1 Performance Metrics**

Model Performance Testing:

Project team shall fill the following information in model performance testing template.









### 10. ADVANTAGES & DISADVANTAGES

### Advantages:

- Patients' records can be accessed at a click.
- Doctors can be able to monitor and consult patients remotely giving doctors less physical examinations and desk work.
- Reduces waiting time for patients to consult with doctors (ER Visits)
- Healthcare personnel can receive alert on emergencies and acute medical cases and respond immediately
- Hospitals can cut costs
- Efficient healthcare delivery
- Healthcare delivery becomes available to the general population

### Disadvantages:

- Man power
- Privacy
- Quality of Data Input which says about the accuracy and reliability of information.
- Data Safety where big storage is the target of hackers.
- Replacing medicine personal such as replacing with robots.

### 11. CONCLUSION

Big data analytics has the potential to transform the wayhealthcare providers use sophisticated technologies togain insight from their clinical and other data repositories and make informed decisions. In the future we'll see the rapid, widespread implementation and use of bigdata analytics across the healthcare organization and thehealthcare industry. To that end, the several challengeshighlighted above, must be addressed. As big data analytics becomes more mainstream, issues such as guaranteeing privacy, safeguarding security, establishing standardsand governance, and continually improving the tools andtechnologies will garner attention. Big data analytics andapplications in healthcare are at a nascent stage of development, but rapid advances in platforms and tools can accelerate their maturing process.

# 12. FUTURE SCOPE

Data Analytics eliminates guesswork and manual tasks. Be it choosing the right content, planning marketing campaigns, or developing products. Organizations can use the insights they gain from data analytics to make informed decisions. Thus, leading to better outcomes and customer

satisfaction. 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. The enormous potential of predictive analytics includes helping identify patients at risk for chronic conditions, developing evidence-based best practices, and proactively spotting potential obstacles to care plan adherence.

### 13. APPENDIX

# Source Code:-

https://github.com/IBM-EPBL/IBM-Project-23033-1659864641/blob/main/Final%20deliverables/Testing/Source%20Code/code.pdf

GitHub & Project Demo Link:-

https://github.com/IBM-EPBL/IBM-Project-23033-1659864641