# ANALYTICS FOR HOSPITAL HEALTH DATA
**TEAM ID**: PNT2022TMID38968

## A PROJECT REPORT

## SUBMITTED BY

***SATHISH.S (TEAM LEADER)***
**KARTHIK RAJA.S (TEAM MEMBER)**
**MOHANDA JIKIRIYA.R(TEAM MEMBER)**
**RANJITH KUMAR.R (TEAM MEMBER)**

**DEPARTMENT OF ELECTRONICS & COUMMUNICATION ENGINEERING**

## KRISHNASAMY COLLEGE OF ENGINEERING & TECHNOLOGY

## ANNA UNIVERSITY:2019-2023

## <u>TABLE OF CONTENT</u>

1. Introduction

   The healthcare industry is multidimensional, with multiple data sources involving healthcare systems, health insurers, clinical researchers, social media, and government , generating different types and massive amounts of data. It is impossible to handle this big data with traditional software and hardware and the existing storage methods and tools. Data analytics is the process of the analysis of data to identify trends and patterns to gain valuable insights. The data generated in the health industry are characterized by the four Vs of big data, namely volume, velocity, variety, and veracity, which play crucial roles in health data analytics. Also, evidence-based decision making has gained importance, which involves the sharing of data among various data repositories.This is due to the increased importance of personalized medicine, the use of advanced technologies, the demand for new payment models, improvement and expansion of care delivery sites, and competition. Various research attempts, based on big data, have provided strong evidence that the efficiency of healthcare applications is dependent upon the

basic architecture, techniques, and tools used. Statistical data and reports can be generated with the use of patient records, aiding in knowledge discovery, and thereby influencing value added services to the patients, improving healthcare quality, the making of timely decisions, and minimizing the costs incurred. Hence, there is a need to incorporate and integrate big data analytics into existing healthcare systems. Despite healthcare analytics having massive potential for value-added change, there are many technological, social, organizational, economic, and policy barriers associated with its application.

## 1.1 Project Overview

Recent Covid-19 Pandemic has raised alarms over one of the most overlooked areas to focus on, which is absolutely Healthcare Data 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. The goal is to accurately predict the Length of Stay for each patient on a 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. The tools that we are using for data analytics is Cognos Analytics from IBM.

## 1.2 Purpose

Data analytics in clinical settings attempts to reduce patient wait times via improved scheduling and staffing, give patients more options when scheduling appointments and receiving treatment, and reduce readmission rates by using population health data to predict which patients are at greatest risk.Insurance firms use data analytics to confirm that they comply with ever-changing regulations; analyze claims and prescriptions to target prevalent health maladies; and compare pricing data with quality metrics to identify high-value, low-cost health providers. Insurance firms also use predictive analytics to spot the potential for fraudulent claims and notify

providers of at-risk claims. Public health professionals increasingly emphasize prediction and prevention over response and treatment. Predictive analytics is used to identify patients at highest risk of chronic illness at the early stages of the disease. Analysis of lab testing, claims data, data that patients themselves generate, and various social factors reduces the risk of long-term illness, which lowers overall healthcare costs and improves patient outcomes.

2. **Literature Survey**

The main aim of this paper is to provide a deep analysis on the research field of healthcare data analytics. This paper is analyzing the previous studies and works in this research area, as well as highlighting some of guidelines and gaps. This study has used seven popular databases and selected most relevant papers, in order to conduct this paper. The paper has listed some data analytics tools and techniques that have been used to improve healthcare performance in many areas such as: medical operations, reports, decision making, and prediction and prevention system. Moreover, the systematic review has showed an interesting demographic of fields of publication, research approaches, as well as outlined some of the possible reasons and issues associated with healthcare data analytics, based on geographical distribution theme[1].

This part deals with the advanced analytical methods focused on healthcare. This includes the clinical prediction models, temporal data mining methods, and visual analytics. Integrating heterogeneous data such as clinical and genomic data is essential for improving the predictive power of the data that will also be discussed. Information retrieval techniques that can enhance the quality of biomedical search will be presented. Data privacy is an extremely important concern in healthcare. Privacy-preserving data publishing techniques will therefore be presented.[2]. One of the promises of the growing critical mass of clinical data accumulating in electronic health record (EHR) systems is secondary use (or re-use) of the data for other purposes, such as quality improvement and clinical research.1 The growth of such data has increased dramatically in recent years due to incentives for EHR adoption in the US funded by the Health Information Technology for Economic and Clinical Health (HITECH) Act.2-3 In the meantime, there has also seen substantial growth in other kinds of health-related data, most notably through efforts to sequence genomes and other biological structures and functions.4 The

analysis of this data is usually called analytics (or data analytics). This chapter will define the terminology of this field, provide an overview of its promise, describe what work has been accomplished, and list the challenges and opportunities going forward[3].

Clinicians, healthcare providers-suppliers, policy makers and patients are experiencing exciting opportunities in light of new information deriving from the analysis of big data sets, a capability that has emerged in the last decades. Due to the rapid increase of publications in the healthcare industry, we have conducted a structured review regarding healthcare big data analytics. With reference to the resource-based view theory we focus on how big data resources are utilised to create organization values/capabilities, and through content analysis of the selected publications we discuss: the classification of big data types related to healthcare, the associate analysis techniques, the created value for stakeholders, the platforms and tools for handling big health data and future aspects in the field. We present a number of pragmatic examples to show how the advances in healthcare were made possible. We believe that the findings of this review are stimulating and provide valuable information to practitioners, policy makers and researchers while presenting them with certain paths for future research[4]. In this modern technoworld, the term data is unavoidable and certainly, nothing is possible without its usage. The trends about how to analyse the data are the need of the hour. Data analytics is becoming a future escalating tool of all industries including medicine, robotics, etc. This article briefly explains how data analytics is used in healthcare systems. Health care is the process of maintaining and improving the health of an individual by preventing, diagnosing and treating the diseases, illness and other physical and mental imbalances in people. Data analytics is classified into four types and they are descriptive, diagnostic, predictive and prescriptive analysis. Health care makes use of prescriptive analysis to arrive at the best results and make better decisions. Big data plays a major role in data analytics. It helps the data analysts to collect data from the patients and store them efficiently. After the completion of this whole article, the reader will be able to get the collective idea about health care analytics.[5] **2.1 Existing problem**

Due to lack of effective data governance procedures, capturing data is one of the biggest obstacles for healthcare organizations. To use data more efficiently, it must

be clean, precise, correctly formatted so that it can be used across various healthcare systems.

## 2.2 References

1) Mohammad Alkhatib , Amir Talaei-Khoei (University of Nevada,Reno)Amir Talaei-Khoei University of Nevada, Reno | UNR · Department of Accounting and Information Systems PhD of Information Systems-Amir Ghapanchi

2) From:"Book of Data Analytics" Chandank Reddy(Wayne State University) Charu C.Aggarwal(Watson Research Center)

3) From: Hoyt,RE,Yoshihashi,A,Eds.(2014).Health Informatics:Practical Guide for Healthcare and formation Technology Professionals,Sixth Edition.Pensacola,FL,Lulu.com.

4) Panagiota Galetsia , Korina Katsaliakia , Sameer Kumarb,* a School of Economics, Business Administration & Legal Studies, International Hellenic University, 14th km Thessaloniki-N. Moudania, Thessaloniki, 57001, Greece b Opus College of Business, University of St. Thomas Minneapolis Campus, 1000 LaSalle Avenue, Schulze Hall 435, Minneapolis, MN 55403, USA

5) from"n book: Innovative Data Communication Technologies and Application (pp.83-96)" P. Nagaraj-Professor (Assistant) at Kalasalingam University

## 2.3 Problem Statement Definition

During the covid-19 pandemic, we have faced one of the difficult times of our life. Everyone seeks to survive from the great disaster. At the time of pandemic, noone get to know about which hospital has vacant beds(free beds) to admit themselves or others infected by covid. This situation made the death rate higher.

# 3. IDEATION & PROPOSED SOLUTION

## 3.1 Empathy Map Canvas
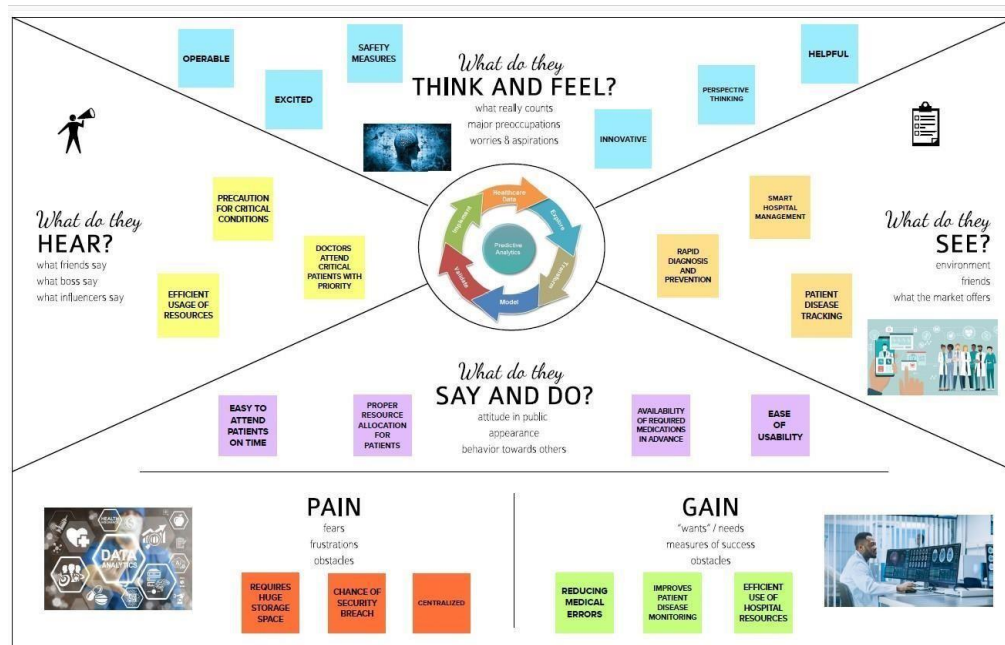
## 3.2 Ideation & Brainstorming

Step-1: Team Gathering, Collaboration and Select the Problem Statement



## 3.3 PROPOSED SOLUTION

| S.No. | Parameter | Description |
|-------|-----------|-------------|
|       |           |             |

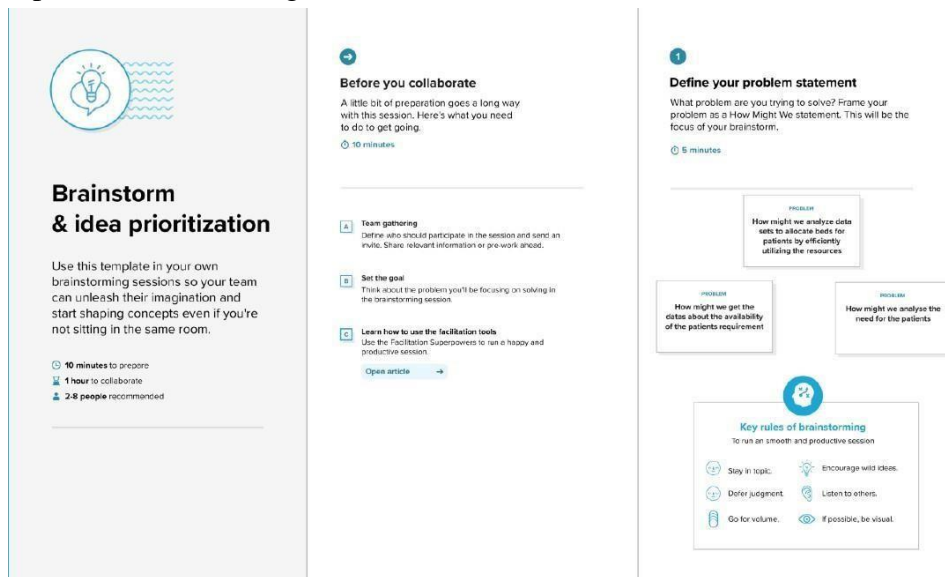| 1. | Problem Statement (Problem to be solved) | During the covid-19 pandemic, we have faced one of the difficult times of our life. Everyone seeks to survive from the great disaster. At the time of pandemic, noone get to know about which hospital has vacant beds(free beds) to admit themselves or others infected by covid. This situation made the death rate higher. |
|---|---|---|

| 2. | Idea / Solution description | Predictive analytics can create patient journey dashboards and disease trajectories that helps us to know about the patient's period of stay. It improves effective allocation of beds and other resources, treatment delivery, improves efficiencies, and so on. |
|---|---|---|
| 3. | Novelty / Uniqueness | Healthcare data frequently resides in several locations. The Collected data should be stored in central system(like centralized storage). This data becomes accessible and usable when it is combined into a single, central system, such as an enterprise data warehouse (EDW). Uniqueness of our project is that we can able to use data for different things such as which medicine is more effective and for understanding behavioural pattern of particular disease. |
| 4. | Social Impact / Customer Satisfaction | effective use of resource<br>Enhanced diagnosis  Improved Treatment<br>enhancing the<br>overall quality of treatment and<br>life of patients |
| 5. | Business Model (Revenue Model) | With the gathered data, redirecting the patients to particular hospital based on the vacancy, leading retailers used methods like market-basket analysis to discover insights about consumer purchase behaviour and used these insights to optimize the physical store experience, target relevant ads and streamline the supply chain, among other strategic initiatives. |

| 6. | Scalability of the Solution | A variety of institutions must store, evaluate, and take action on the massive amounts of data being produced by the health care sector as it expands quickly. India is a vast, culturally varied nation with a sizable population that is increasingly able to access centralised healthcare services. |
|---|---|---|

## **3.4** **Problem Solution fit**

1. **CUSTOMER SEGMENTS**

   ❖ Patients
   ❖ Hospital Management

2. **PROBLEMS**

   ❖ Effective resource allocation
   ❖ Reduce waiting time for patients in hospital

3. **TRIGGERS TO ACT**

   ❖ Covid pandemic
   ❖ Emergency situations

4. **EMOTIONS**

   ❖ BEFORE: feeling bad & frustrated
   ❖ AFTER: feeling better & relaxed

5. **AVAILABLE SOLUTIONS**

   ❖ Tableau cloud
   ❖ Text mining
   ❖ Information retrieval

6. **CUSTOMER STATE LIMITATIONS**

   ❖ Inadequate information about availability of required resources

7. **BEHAVIOR**

   ❖ Tracking the information with available technologies

8. **CHANNELS OF BEHAVIOR**

   ❖ ONLINE: use of data from all region (data exploration)
   ❖ OFFLINE: use of data collected from nearby facilities

9. **CAUSE OF EVERY PROBLEM**

   ❖ No proper system or less effective prediction system

10. **SOLUTION**

- ❖ EXISTING: ratio of discharges in given period of time to no of beds in hospital during the time period
- ❖ PROPOSED: using predictive analysis powered by AI

## 4. REQUIREMENT ANALYSIS

This part of this document contains the functional and the non-functional requirements of this project.

### 4.1 Functional requirement

Following are the functional requirements of the proposed solution.

| FR No. | Functional Requirement | Sub Requirement (Story / SubTask) |
|--------|------------------------|-----------------------------------|
| 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 requirements

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

| NFR No. | Non-Functional Requirement | Description |
|---------|----------------------------|-------------|
| NFR-1 | Usability | These Dashboards are designed to offer a comprehensive overview of |

| | | |
|---------|----------------------------|-------------|
| | | a 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 an 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. |

| 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

| User Type | Functional Requirement (Epic) | User Story Number | User Story / Task | Acceptance criter |
|---|---|---|---|---|
| Customer | Dashboard | USN-1 | As a user, I can upload the datasets to the dashboard | I can access variou operations |
| | View | USN-2 | As a user, I can view the patient details | I can view the visu and the result after prediction |
| Admin | Analyse | USN-3 | As an admin, I will analyse the given dataset | I can analyse the d |
| | Predict | USN-4 | As an admin, I will predict the length of stay | I can predict the le stay |

dataenters and leaves the system, what changes the information, and where data is stored
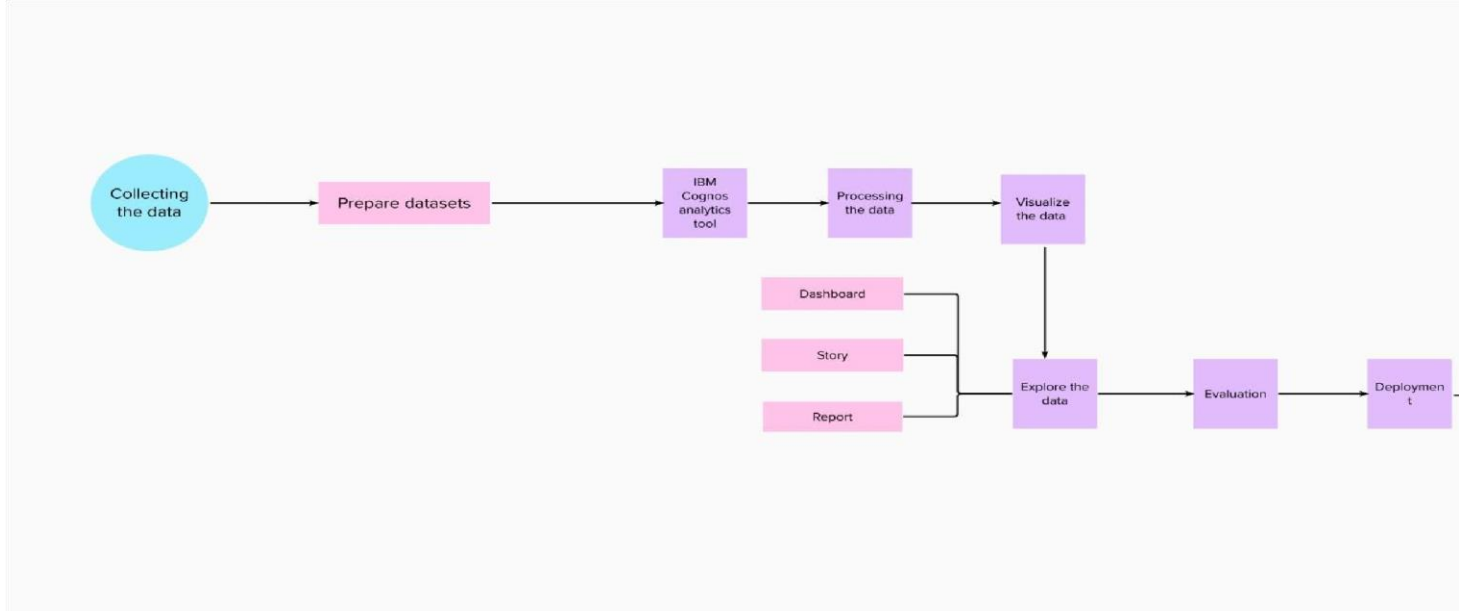


**User Stories**

| S.No | Component | Description | Techno |
|------|-----------|-------------|--------|
| 1. | User Interface | The user interacts with application using Dashboard | IBM Co |
| 2. | Data Processing | The data from the dataset is pre-processed | IBM Co |
| 3. | Cloud Database | The clean dataset is stored on IBM Cloud | IBM Clo |
| 4. | Data visualization | The data is visualized into different forms | IBM Co |
| 5. | Prediction | ML algorithms are used for predicting the length of stay | ML algo Bagger |

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

## 5.3 SOLUTION &TECHNOLOGY ARCHITURE
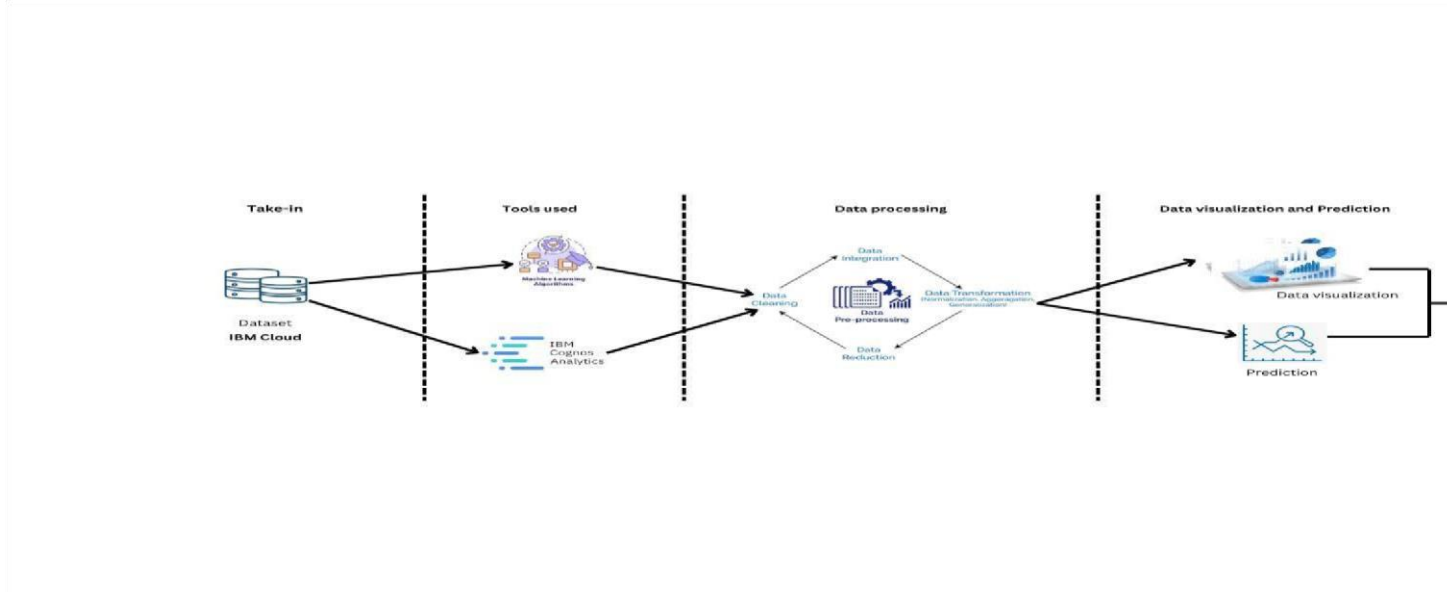
**Table-1 : Components & Technologies:**



**Table-2: Application Characteristics:**

| S.No | Characteristics | Description | Technol |
|------|----------------|-------------|---------|
| 1. | Open-Source Frameworks | Open-source frameworks used | IBM Cog |
| 2. | Security Implementations | Request authentication using Encryptions | Encryptio |
| 3. | Scalable Architecture | Scalability consists of 3-tiers | Dashboa Applicat Databas |
| 4. | Availability | The application is available for cloud users | IBM Clou |
| 5. | Performance | The user can know the prolonged period of stay | ML algor |

**5.3 User Stories**

| User Type | Functional Requirement (Epic) | User Story Number | User Story / Task | Acceptance criteria | Priority | |
|-----------|-------------------------------|-------------------|-------------------|---------------------|----------|--|

| Customer (Mobile user) | Registration | USN-1 | As a user, I can register for the dashboard by entering my email, and password, and confirming my password. | I can access my account in the dashboard | High |
|---|---|---|---|---|---|
| | | USN-2 | As a user, I will receive a confirmation email once I have registered for the dashboard | I can receive a confirmation email & click confirm | High |
| | | USN-3 | As a user, I can register for the dashboard through Social Media | I can register & access the dashboard with Social Media Login | Low |
| | | USN-4 | As a user, I can register for the dashboard through Gmail | I can register and access dashboard with Gmail | Medium |
| | Login | USN-5 | As a user, I can log into the application by entering email & password | I can login to the account in my email login. | High |
| | Dashboard | USN-6 | As a user ,I can use my account in my dashboard for uploading dataset. | I can login to the account for uploading dataset. | Medium |
| Customer (Web user) | Website | USN-7 | As a user ,I can use my dashboard in website | I can login into the dashboard by visiting website. | Medium |

| | | | | | |
|---|---|---|---|---|---|
| Customer Care Executive | | USN-8 | As a user ,I can contact Customer care Executive for my login. | I can contact customer executive for my login. | High |
| Administrator | | USN-9 | As a user ,I can contact administrator for my queries. | I can contact administrator for solving my queries. | High |

| | | | | | |
|---|---|---|---|---|---|
| Exploration | Dashboard | USN-10 | As a user, I can prepare data by using Exploration Techniques. | I can prepare data by using Exploration Techniques. | High |
| Presentation | Dashboard | USN-11 | As a user, I can Present data in my dashboard. | I can present data by using my account in dashboard. | High |
| Visualization | Dashboard | USN-12 | As a user, I can Prepare Data by using Visualization Techniques. | I can prepare data by using Visualization Techniques. | High |

## 6. PROJECT PLANNING & SCHEDULING

| Sprint | Functional Requirement (Epic) | User Story Number | User Story / Task | Story Points | Priority | |
|---|---|---|---|---|---|---|
| Sprint-1 | API and Database connectivity | USN-1 | As an analyser , I use Kaggle  API to gather the dataset | 3 | Medium | |
| | | USN-2 | As an analyzer , I can create database connectivity using IBM Cloud and IBM Cognos Analytics. | 5 | High | |
| | | USN-3 | AS an analyzer , I can preprocess the data and Explore it . | 5 | High | |
| Sprint-2 | Dashboard | USN-4 | As an analyser, I can add the created visualizations to the dashboard (Creating dashboard) | 8 | High | |
| Sprint-3 | Report, Story | USN-5 | As an analyser, I can create reports for the given dataset (Creating report) | 8 | High | |
| | | USN-6 | As an analyser, I can create story for the given dataset | 8 | High | |
| Sprint | Functional Requirement (Epic) | User Story Number | User Story / Task | Story Points | Priority | |
| | | | (Creating story) | | | |
| Sprint-4 | Documentation | USN-7 | We create our entire project documentation. | 10 | High | |

| Sprint | Total Story Points | Duration | Sprint Start Date | Sprint End Date (Planned) | Story Points Completed (as o Planned End Da |
|---|---|---|---|---|---|
| Sprint-1 | 13 | 12 Days | 24 Oct 2022 | 04 Nov 2022 | |
| Sprint-2 | 8 | 2 Days | 05 Nov 2022 | 06 Nov 2022 | |
| Sprint-3 | 16 | 2 Days | 07 Nov 2022 | 08 Nov 2022 | |

| Sprint-4 | 10 | 4 Days | 09 Nov 2022 | 12 Nov 2022 | |

**Project Tracker, Velocity & Burndown Chart: (4 Marks)** **Velocity:**

$$AV = \frac{sprint\ duration}{velocity} = \frac{20}{10} = 2$$

Average

velocity for

sprint-

1:

AV =

13/12 =
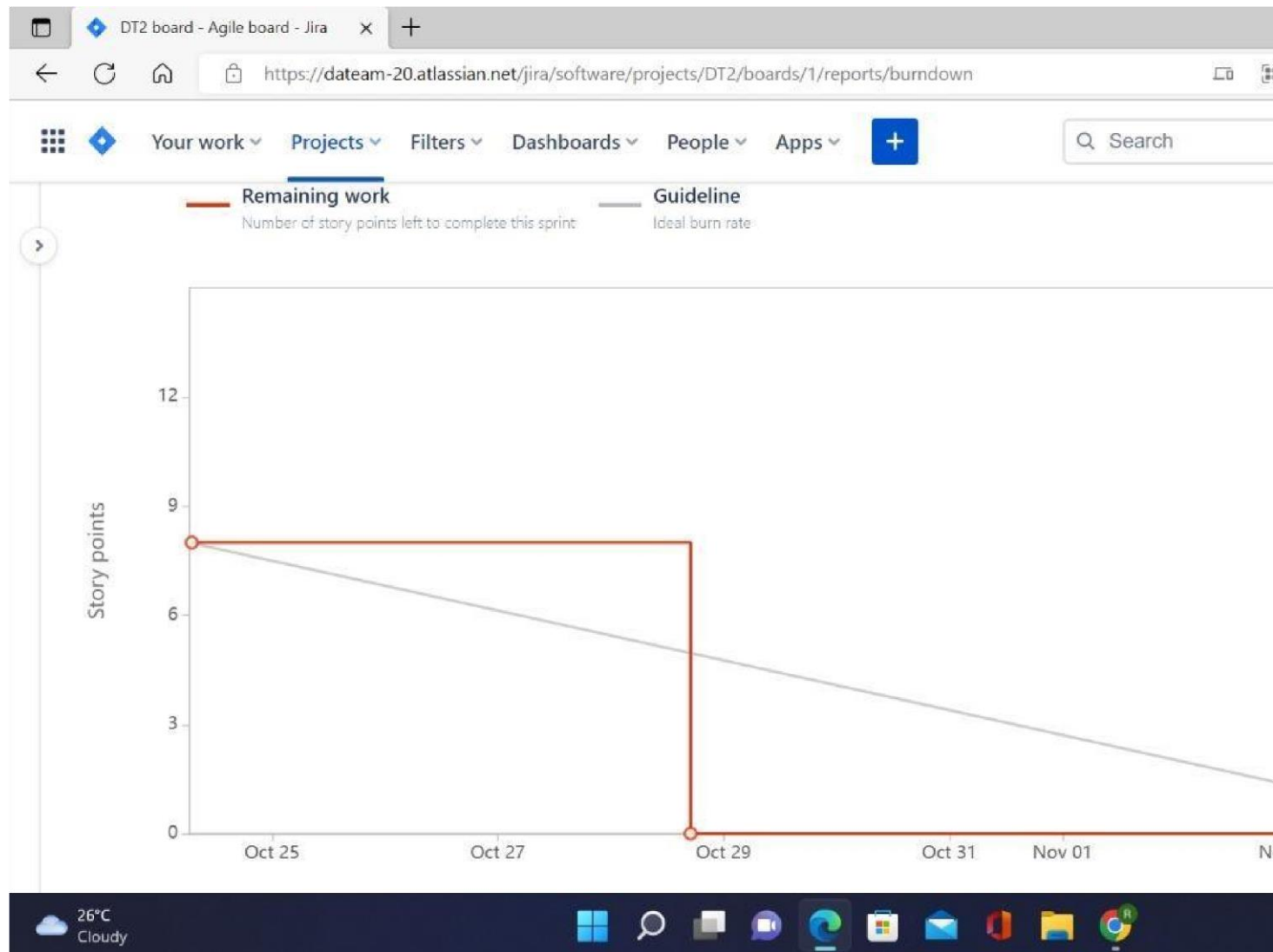
1.08

Average

velocity for

sprint-

2:

AV = 8/2
= 4

Average
velocity for
sprint3:

AV =
16/2 = 8

Average
velocity for
sprint4:

AV =
10/4 =
2.5

**Burndown Chart:**

Your work ∨    Projects ∨    Filters ∨    Dashboards ∨    People ∨    Apps ∨    +    🔍 Search

—— **Remaining work**
Number of story points left to complete this sprint

—— **Guideline**
Ideal burn rate



Story points

12

9

6

3

0

Oct 25          Oct 27          Oct 29          Oct 31          Nov 01          N

26°C
Cloudy

# 7.TESTING

## 7.1Test Cases

- ➢ Verify user is able to see Home page.
- ➢ Verify user is able to see Dashboard page.

➤ Verify user is able to navigate to ➤ Report page. Verify user is able to ➤ navigate to story page. Verify filters are working

## *7.2User Acceptance Testing*

### 1. Purpose of Document

The purpose of this document is to briefly explain the test coverage and open issues of the [ProductName] project at the time of the release to User Acceptance Testing (UAT).

### 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 | Subtotal |
|---|---|---|---|---|---|
| By Design | 8 | 5 | 0 | 3 | 16 |
| Duplicate | 1 | 0 | 5 | 0 | 6 |
| External | 0 | 3 | 2 | 1 | 6 |
| Fixed | 13 | 4 | 3 | 16 | 36 |
| Not Reproduced | 0 | 1 | 0 | 0 | 1 |
| Skipped | 0 | 1 | 0 | 1 | 2 |
| Won't Fix | 1 | 4 | 2 | 1 | 8 |
| Totals | 23 | 18 | 12 | 22 | 75 |

### 3. 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 | 9 | 0 | 0 | 9 |
| Client Application | 43 | 0 | 0 | 43 |
| Security | 1 | 0 | 0 | 1 |
| Outsource Shipping | 1 | 0 | 0 | 1 |
| Exception Reporting | 9 | 0 | 0 | 9 |
| Final Report Output | 10 | 0 | 0 | 10 |
| Version Control | 1 | 0 | 0 | 1 |
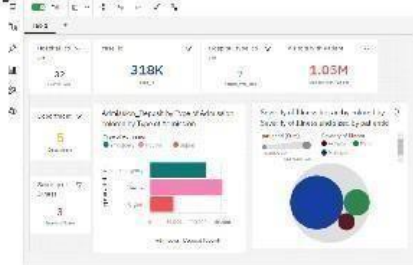
# 8.RESULTS
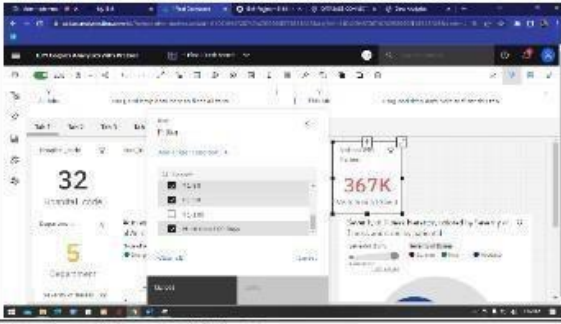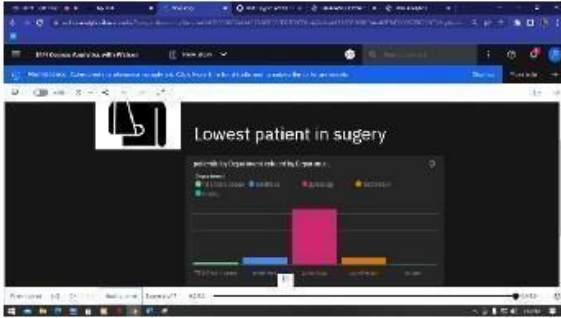
## 8.1 Performance Metrics

**Model Performance Testing:**

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

| S.No. | Parameter | Screenshot/Values |
|-------|-----------|-------------------|
| 1. | Dashboard design | Number of Visulizations / Graphs – 22<br>Number of tabs – 8<br> |
| 2. | Data Responsiveness | Data's will dynamically changed and graph also changed.<br> |
| 3. | Amount Data to Rendered (DB2 Metrics) | Number of rows read – 318438<br>Number of rows loaded – 318438<br>Number of rows rejected – 0<br> |

| 4. | Utilization of Data Filters | We created filters for Dashboards which is perfectly working. |
|---|---|---|
| | |  |
| 5. | Effective User Story | Number of Scene Added – 7 Animations are perfectly displayed. Images are perfectly rendered. |
| | |  |
| 6. | Descriptive Reports | Number of Visulizations / Graphs – 6 |
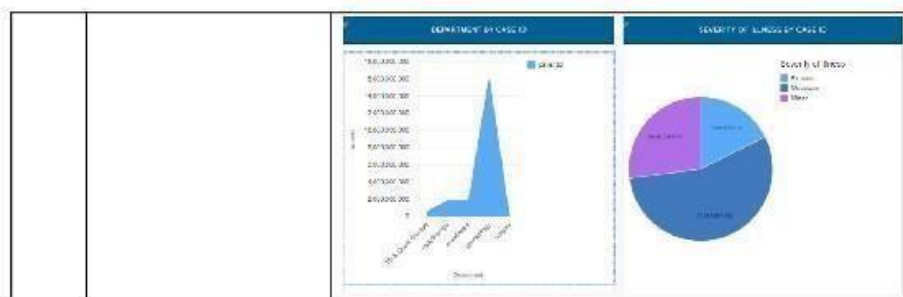


## 9.ADVANTAGES

- ➤ Improved research efforts Improved
- ➤ health outcomesObtain operational
- ➤ insightsImproved staffing ➤
Informed strategic planning

- ➤ *Higher-Quality Care*

## 10.DISADVANTAGES

- *Privacy*
- *Replacing Doctors*
- Frustration with poor implementation.
- Cybersecurity risks
- Healthcare Regulatory Changes.
- Healthcare Staffing Shortages

## 11.CONCLUSION

- It also means describing how health plans, health care organizations, and clinicians should be accountable to patients and society and conversely. How ➤ individuals can take appropriate responsibility for their own health.
- Data analytics is the science of analyzing raw datasets in order to derive a conclusion regarding the information they hold.
- It enables us to discover patterns in the raw data and draw valuable information from them.

## 12.FUTURE SCOPE

- **Improved Decision Making**: Data Analytics eliminates guesswork and manualtasks. 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 satisfactionData analytics . ➤ achieve business goals of pharmaceutical companies, payers,  insurance companies, physicians, hospitals, medical equipment companies, sales reps, and other stakeholders in the healthcare business, need for this haveonly increased after the Affordable Act came into being