# PROJECT REPORT

| | |
|---|---|
| **Team ID** | **PNT2022TMID26228** |
| **Project Name** | **Analytics For Hospital Health–Care Data** |

## 1.INTRODUCTION

### 1.1 PROJECT OVERVIEW

Healthcare data analytics refers to the process of working on raw datasets related to healthcare and analyzing them to find hidden patterns, trends, etc., thus paving a way for further improvements at patient-level as well as business-level.

Since we talked about healthcare-related data in the above definition, let us now understand what healthcare data is. Healthcare data is nothing but any data that is related to the patient and the healthcare facilities such as medical records, scan and test reports, hospital records, etc. Different tools are used to collect this data. Some of the important tools and ways are electronic health records (EHRs), patient portals, master patient indexes (MPIs), online health-related mobile applications, etc. Not only does this help in data-driven informed decision-making, but it also helps in providing a personalized experience and treatment to the patients. The Covid-19 pandemic has created an urgent need for application of data analytics in the healthcare industry. Many hospitals and healthcare organizations have successfully utilized the benefits of Data Analytics to tackle the pandemic.

### 1.2 PURPOSE

Healthcare analytics refers to the utilization of vast amounts of collected data to supply organizations with actionable insights. These insights are developed through analytical disciplines to handle fact-based deciding. In turn, these decisions better planning, management, measurement and learning. While healthcare management has various use cases for creating use of data science, patient length of stay is one critical parameter to note and forecast if one looks to boost the efficiency of the healthcare management in a very hospital. IBM Cognos Analytics is employed to integrate hospitals reporting, modeling, analysis, dashboards, stories, and event management. This parameter helps hospitals to spot 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 reduce LOS and lower the prospect of staff/visitor infection. Also, prior knowledge of LOS can aid in logistics like room and bed allocation planning. The goal is to accurately predict the Length of Stay for every patient on a case by basis so the Hospitals can use this information for optimal resource allocation and better functioning

# 2. LITERATURE SURVEY

## 2.1 EXISTING SYSTEM

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

M.S.SIMI AND K.S.NAYAKI GAVE A REVIEW ON “DATA ANALYTICS IN HEALTH CARE” AT INTERNATIONAL CONFERENCE ON CIRCUIT, POWER AND COMPUTING TECHNOLOGIES(2019). The knowledge discovery in database (KDD) is alarmed by the advancement of strategies and systems for making the utilization of data. A standout amongst the most vital stride of the KDD is the data analytics. Data analytics is one of the extensively researched areas inoperable from the wide impact showed by this computational strategy on differing fields, such as, Artificial Intelligence (AI), databases, statistics, and visualization. It has unlimited applications and ways to deal with analytics the data in suitable ways. Both the data analytics and medicine have raised some of dependable early discovery frameworks and different medical services related frameworks from the medical data. Medical data analytics is a dynamic interdisciplinary area of research that is viewed as the result of applying artificial intelligence and data analytics concepts to the field of clinical and medical services. We have reviewed the different papers intricate in this field in terms of technique, algorithms and results. The aim of this research work is to give a review on the foundation benchmarks in analytics

of infertility, and present the findings and results of past researches on utilizing data analytics procedures to analyze electronic health records.

M.AMBIGAVATHI AND D.SRIDHARAN,"BIG DATA ANALYTICS IN HEALTH CARE,"AT 2018 TENTH INTERNATIONAL CONFERENCE ON (ICOAC). The pace of both digital innovation and technology disruption is refining the healthcare industry at an exponential rate. The large volume of healthcare data continues to mount every second, making it harder and very difficult to find any form of useful information. Recently, big data is shifting the traditional way of data delivery into valuable insights using big data analytics method. Big data analytics provides a lot of benefits in the healthcare sector to detect critical diseases at the initial stage and deliver better healthcare services to the right patient at the right time so that it improves the quality of life care. Big data analytics tools play an essential role to analyze and integrate large volumes of structured, semi-structured and unstructured vital data rapidly produced by the various clinical, hospitals, other social web sources and medical data lakes. However, there are several issues to be addressed in the current health data analytics platforms that offer technical mechanisms for data collection, aggregation, process, analysis, visualization, and interpretation. Due to lack of detailed study in the previous literature, this article inspects the promising field of big data analytics in healthcare. This article examines the unique characteristics of big data, big data analytical tools, different phases followed by the healthcare economy from data collection to the data delivery stage. Further, this article briefly summarizes the open research challenges with feasible findings, and then finally offers the conclusion.

Z.FEI, Y.RYEZNIK, A.SVERDLOW, C.W.TAN AND W.K.WONG GAVE "AN OVERVIEW OF HEALTH CARE DATA ANALYTICS WITH APPLICATION TO THE COVID-19 PANDEMIC",2020. In the era of big data, standard analysis tools may be inadequate for making inference and there is a growing need for more efficient and innovative ways to collect, process, analyze and interpret the massive and complex data. We provide an overview of challenges in big data problems and describe how innovative analytical methods, machine learning tools and meta heuristics can tackle general healthcare problems with a focus on the current pandemic. In particular, we give applications of modern digital technology, statistical methods,data platforms and data integration systems to improve diagnosis and treatment of diseases in clinical research and novel epidemiological tools to tackle infection source problems, such as finding Patient Zero

in the spread of epidemics. We make the case that analyzing and interpreting big data is a very challenging task that requires a multi-disciplinary effort to continuously create more effective methodologies and powerful tools to transfer data information into knowledge that enables informed decision making.

S.KAPOOR, L.KASAR, A.MANDOLE AND J.MAHAJAN GAVE AN APPROACH IN “HEART DISEASE PREDICTION USING MACHINE LEARING AND DATA ANALYTICS APPROACH”(2019). In recent times, Machine Learning has played a significant role in the healthcare industry and amongst all of the major diseases, heart disease is one of the significant and most critical diseases to predict. There is a rapid increase in the number of cases each day. It has been observed that in every minute, 4 people between the age group of 30-50 get a stroke, so we are using machine learning algorithms to mitigate this problem. Kaggle used the heart disease dataset used for this project. This paper demonstrates the prediction of heart disease using multiple machine learning classification algorithms such as Naive Bayes, Random Forest, SVM etc., and compares their accuracy scores. Later on, Stacking Ensemble Learning Technique is used to boost our classification models' performance.

S.NOUREDDINE,Z.BAARIR,A.TOUMI,A.BETKA,N.KAZAR AND A.N.BEHARKAT. “SMART BREAST CANCER PREDICTION USING DATA MINING PROCESS”.(2017) In terms of mortality, breast cancer occupies the 5th place in the world. According to the World Health organization (WHO), breast cancer is the principal cause of cancer death in the last years, especially among women. Following this very alarming finding, research was developed using predictive approaches based on the analysis of a large volume of data. In the literature, data analysis was carried out through data mining using several techniques, especially meta heuristics, to aid in the preventive or predictive decision of breast cancer. In this paper, a recent meta heuristic called Symbiotic Organisms Search (SOS) based on a bio-inspired phenomenon is combined with an Artificial Neural Network (ANN), will be utilized in data mining for a predictive study of breast cancer disease. This paper contains also an intelligent application using the SOS algorithm. The experiments generated very confident results in terms of predictive decision following the classification of data.

A.T.NAGI, M.JAVED AWAN, R.JAVED AND N.AYESHA. “A COMPARISION OF TWO-

STAGE CLASSIFIER ALGORITHM WITH ENSEMBLE TECHNIQUES ON DETECTION OF DIABETIC RETINOPATHY"(2018). The Diabetic retinopathy is disease of the human eye that causes retinal damage in diabetic patients. It further leads to the blindness. The machine learning techniques plays an important rule to predict the early diabetic retinopathy which avoided from the intensive labor. In this paper we used the novel technique, the Two Stage Classifier, an ensemble technique which combines various machine learning algorithms for classification. In the subject paper, the classifier is applied to predict Diabetic retinopathy (DR), a disease of the human eye that causes retinal damage in diabetic patients and ultimately lead to complete blindness. The problem lies in the fact that it is time consuming to detect this disease but an early detection of the disease is essential to avoid complete blindness. We apply machine learning algorithms to determine the existence of DR and compare the accuracies of the applied techniques. The Two Stage Classifier, turns out to be better not only in terms of parallelism but also in terms of accuracy.

A.ARJUN, A.SRINATH AND B.R.CHANDAVARKAR. "PREDICTIVE ANALYTICS AND DATA MINING IN HEALTHCARE".AT 2021 12TH INTERNATIONAL TECHNOLOGIES. Machine Learning and Data Mining for healthcare. There has been an enormous growth in the field of HIT (health information technology) in the recent years. Be it detection of certain diseases, scanning of organs, finding tumors, these machine oriented operations without human intervention, have certainly increased the quality of medical attention one can get, and the technology required has come a long way. Health data tends to be inherently complex with exceptions in almost all cases. Data mining is the technique converting raw data into a meaningful format. Analysis and prediction on such data, although computationally and algorithmic complex, is an emerging technology that is a small step to more proactive and preventive automated treatment options. There are various data mining techniques such as classification, clustering, association, regression, prediction, pattern recognition . Even the efficiency of certain medicines can be found using machine learning techniques, which is a life saving and cost effective method. In this paper, we are going to use machine learning as a tool for predictive analysis to predict chronic kidney diseases based on the Chronic disease dataset taken from UCI ML repository. We will be applying machine learning algorithms,specifically decision trees, to build a classifier to predict if a person has the disease or not. This paper shows the issue that specific machine learning algorithms need to be tailor-made to specific nature of medical data. "ANALYSIS OF RESEARCH IN HEALTHCARE DATA ANALYTICS"(2017): The main aim of this paper is to provide a deep analysis on the research field of healthcare data analytics., as well as highlighting some of guidelines

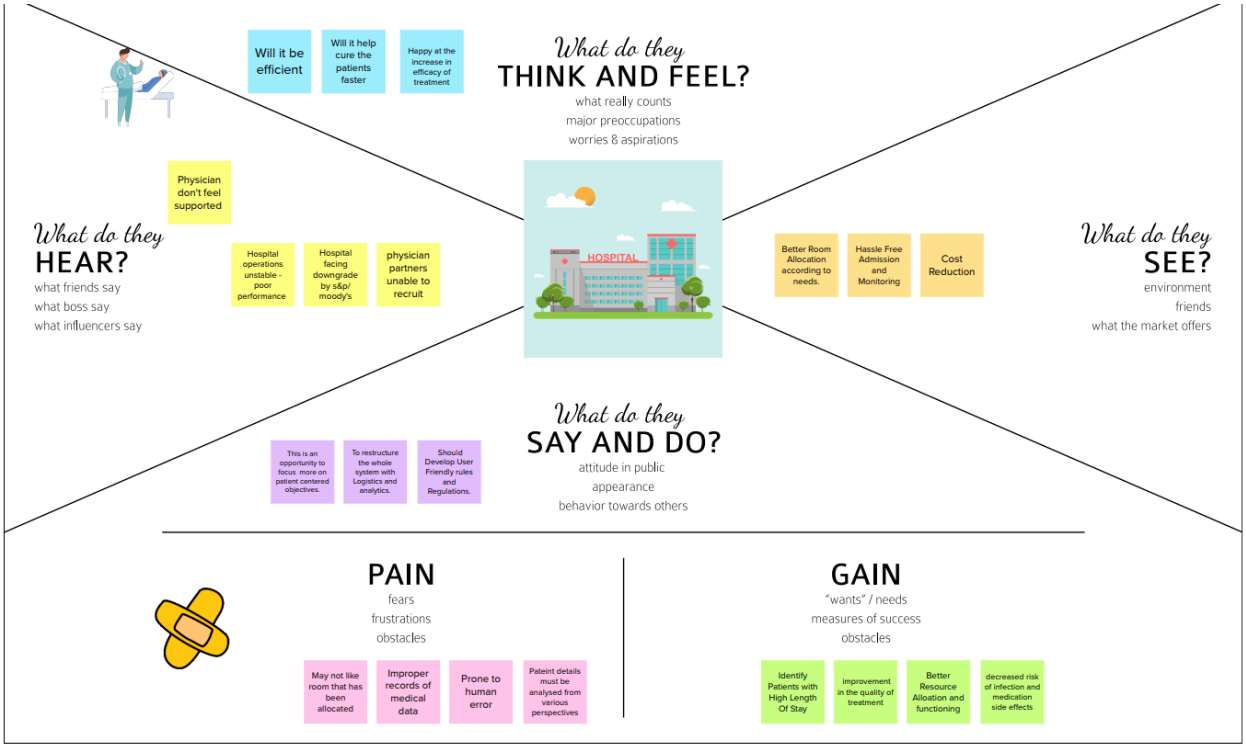
and gaps in previous studies. This study has focused on searching relevant papers about healthcare analytics by searching in seven popular databases such as google scholar and springer using specific keywords, in order to understand the healthcare topic and conduct our literature review. The paper has 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.

## 2.3 PROBLEM STATEMENT DEFINITION

At minimum, a data analytics platform in healthcare must support the key functions necessary for processing the data. The criteria for platform evaluation may include availability, continuity, ease of use, scalability, ability to manipulate at different levels of granularity, privacy and security enablement, and quality assurance [6, 29, 32]. In addition, while most platforms currently available are open source, the typical advantages and limitations of open source platforms apply. To succeed, big data analytics in healthcare needs to be packaged so it is menu-driven, user-friendly and transparent. Real-time big data analytics is a key requirement in healthcare. The lag between data collection and processing has to be addressed. The dynamic availability of numerous analytics algorithms, models and methods in a pull-down type of menu is also necessary for large-scale adoption. The important managerial issues of ownership, governance and standards have to be considered. And woven through these issues are those of continuous data acquisition and data cleansing. Health care data is rarely standardized, often fragmented, or generated in legacy IT systems with incompatible formats [6]. This great challenge needs to be addressed as well.

# 3. IDEATION AND PROPOSED SOLUTION

## 3.1 EMPATHY MAP CANVAS

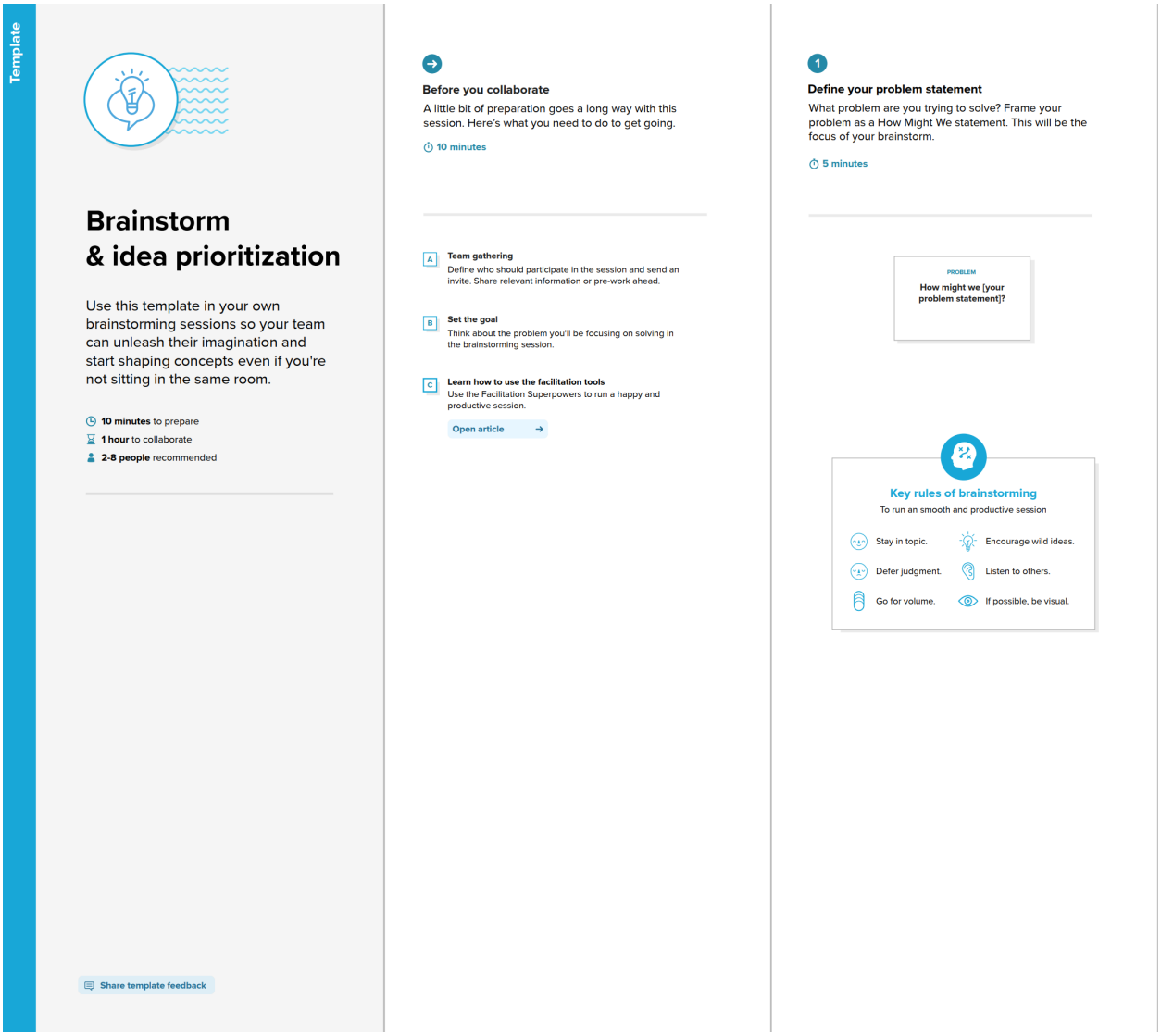



## 3.2 IDEATION AND BRAINSTORMING

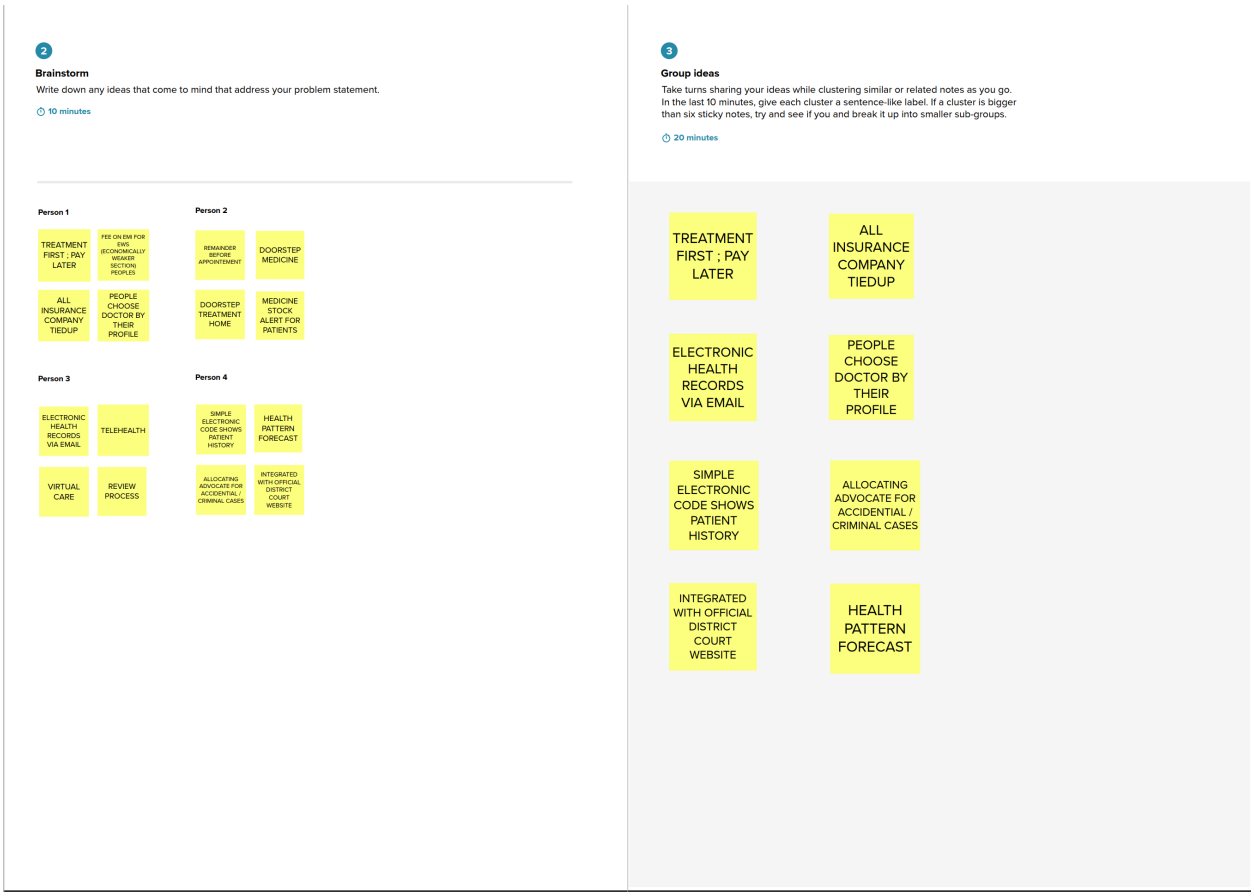
### Brainstorm & Idea Prioritization Template:

Brainstorming provides a free and open environment that encourages everyone within a team to participate in the creative thinking 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.

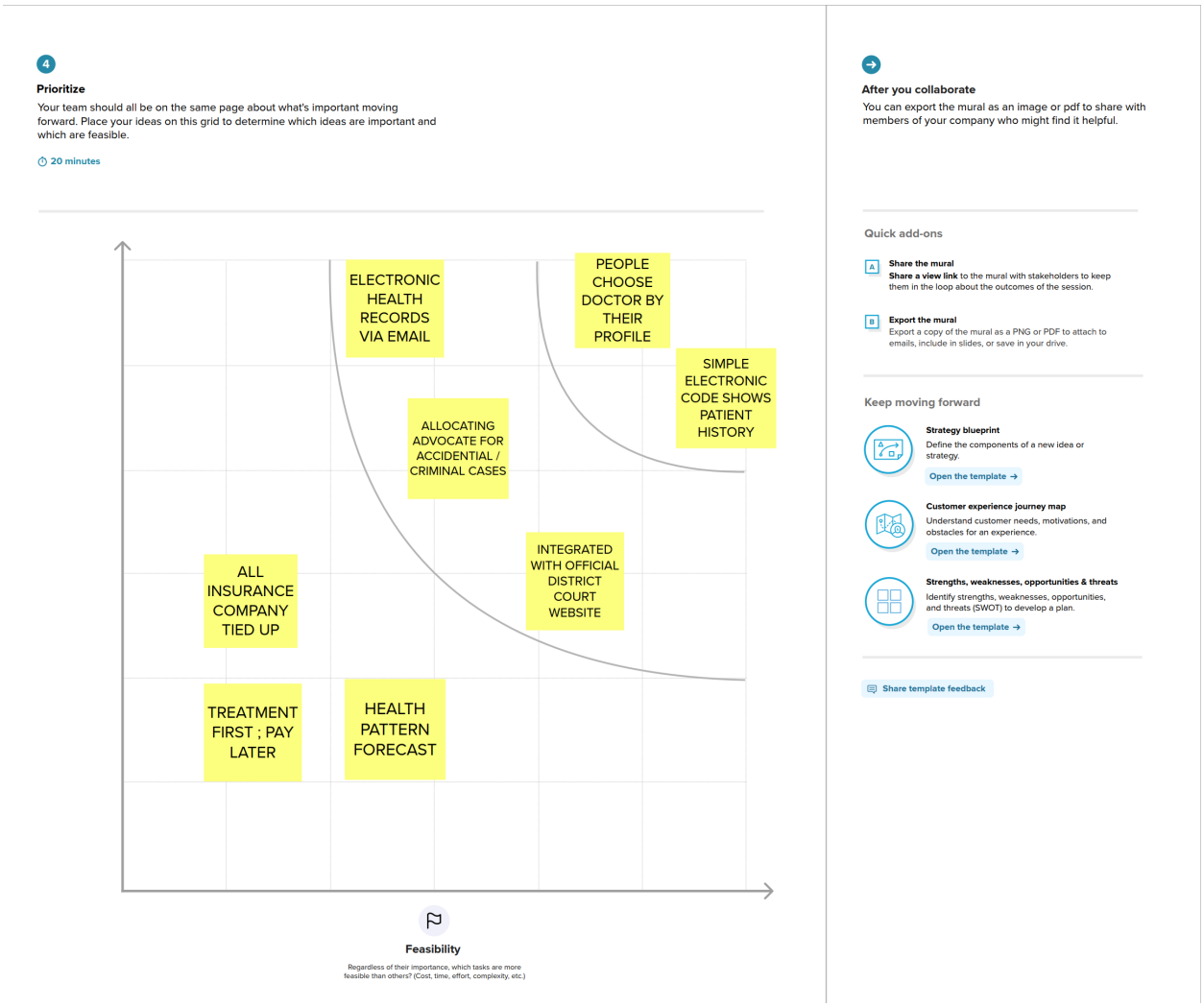
# Step-1: Team Gathering, Collaboration and Select the Problem Statement




# Step-2:Brainstorm, Idea Listing and Grouping




# Step-3 Idea Prioritization



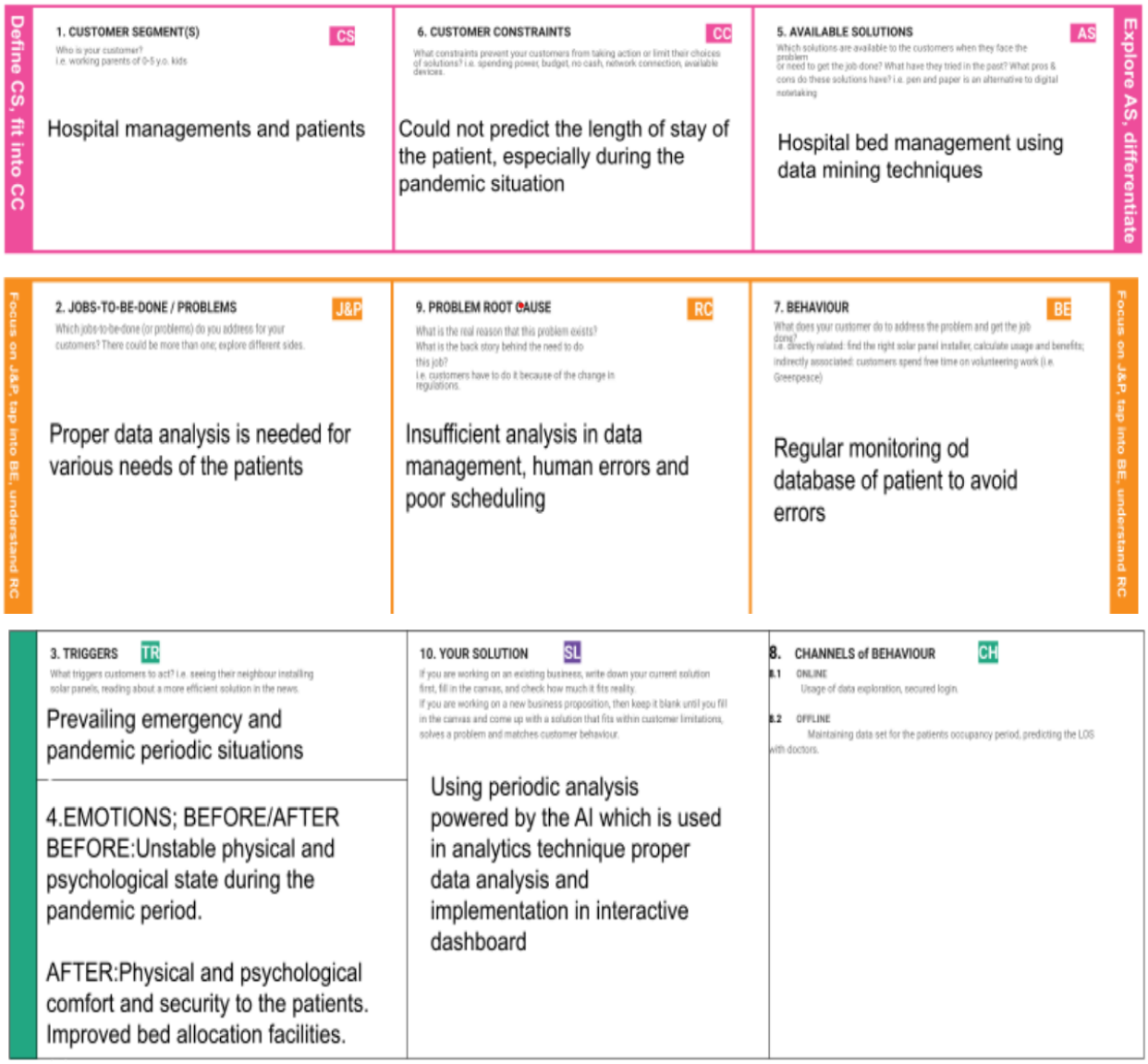

# 3.3 PROPOSED SOLUTION

## Proposed Solution Template:

| S.No. | Parameter | Description |
|---|---|---|
| 1. | Problem Statement (Problem to besolved) | To Predict the Length of the stay for each patient on case by case basis so that the hospital can use the information for optimal<br>resource allocation and better functioning |

| | | |
|---|---|---|
| 2. | Idea / Solution description | To create a dashboard for monitoring the lengthof stay of patients and also include the necessary data which were provided by the hospital to create an effective virtual dashboardusing Psychoanalytical. |
| 3. | Novelty / Uniqueness | By Using Data Visualization techniques, the dashboard supports clinicians and hospitalmanagers in viewing and exploring data onprocesses and outcomes of care in an interactive manner. |
| 4. | Social Impact / Customer Satisfaction | The hospitals can use this dashboard to view their hospitality records and they can update the given inputs and get the expected output effectively. |
| 5. | Business Model (Revenue Model) | While using this dashboard the hospitals can easily get regular updates on the patients and this was widely applicable in all departments ofthe hospitals. The Hospital staff can easily logininto the dashboard and view the risk rate of the patients according to the length of stay in the hospital and can give proper treatment according to that. |
| 6. | Scalability of the Solution | The scalability of this project was a. The hospitals can view the length of stay of thepatient case by case basis b. The patient id,department name, other hospital-related details, etc. c. It also helped to visualize an interactive dashboard efficiently |

## 3.4 PROBLEM SOLUTION




# 4.REQUIREMENT ANALYSIS

### 4.1 Functional Requirements:

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 Registration through LinkedIN |
| FR-2 | User Confirmation | Confirmation via Email<br>Confirmation via OTP |
| FR-3 | Database | Every patient has some necessary data like phone number, their first and last name, personal health number, postal code, country, address, city, 'patient's IDnumber, etc |
| FR-4 | Report Generation | The Hospital Management System generates a report on every patient regarding various information like patientsname, Phone number, bed number, the doctor's name whom its assigns, ward name, and more. The Hospital Management system also helps in generating reports onthe availability of the bed regarding information like bednumbers unoccupied or occupied, ward name, and more. |
| | Check Out | The staff in the administration section of the ward can delete the patient ID from the system when the patientcheckout from the hospital. The Staff in the administration section of the ward can put the bed empty in the list of beds available. |
| | Adding Patients | The Hospital Management enables the staff at the front desk to include new patients in the system. |

## Non-functional Requirements:

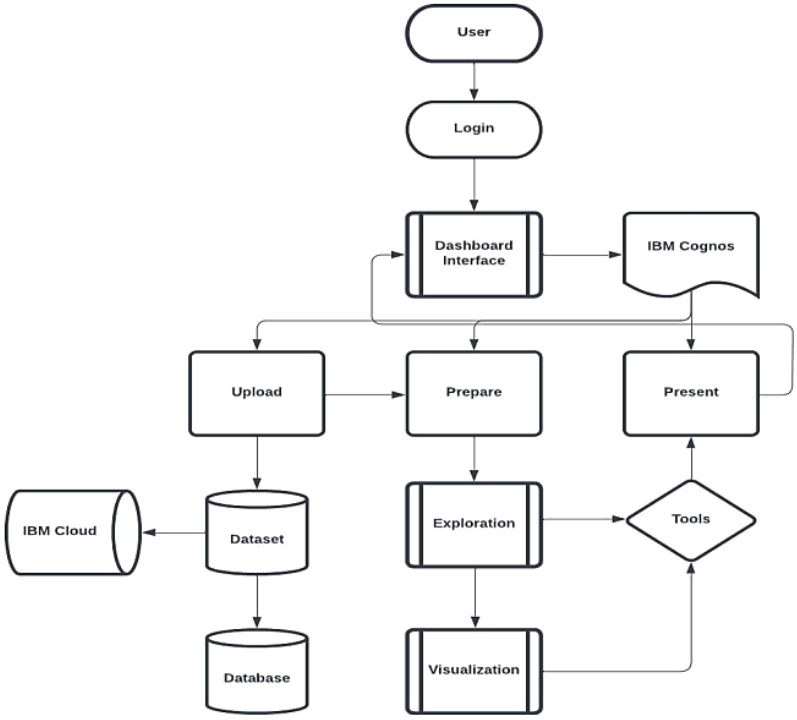
Following are the non-functional requirements of the proposed solution.

| FR No. | Non-Functional Requirement | Description |
|---|---|---|
| NFR-1 | **Usability** | The effectiveness, efficiency and satisfaction with<br>which specific users can achieve a specific set oftasks in a particular environment. |
| NFR-2 | **Security** | This process of protecting data from unauthorized<br>access and data corruption throughout its lifecycle |
| NFR-3 | **Reliability** | A highly reliable system has a lower risk of errorsand process failures that can cause patients harm |
| NFR-4 | **Performance** | performance measurements include:<br>1. Quality and efficiency of patient care<br>2. Cost of healthcare services<br>3. Disparities in performance Care outcomes |
| NFR-5 | **Availability** | inpatient, outpatient, pharmacy, and enrollment |
| NFR-6 | **Scalability** | The ability of a health intervention shown to be efficacious on a small scale and/or under controlledconditions to be expanded under real world conditions to reach a greater proportion of the<br>eligible population, while retaining effectiveness |

# 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.




## 5.3 USER STORIES

| Sprint | Functional Requirement(Epic) | User StoryNumber | User Story / Task | Story Points | Priority | Team Members |
|---|---|---|---|---|---|---|
| Sprint-4 | Using ML algorithm for Prediction | USN-8 | As a user, I must have prior knowledge of LOS can aid in logistics such as room and bed allocation planning. | 20 | High | Shrinivas S S |

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

| Sprint | Total Story Points | Duration | Sprint StartDate | Sprint End Date(Planned) | Story Points Completed (as on Planned End Date) | Sprint ReleaseDate (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 |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |

## 6. PROJECT PLANNING & SCHEDULING

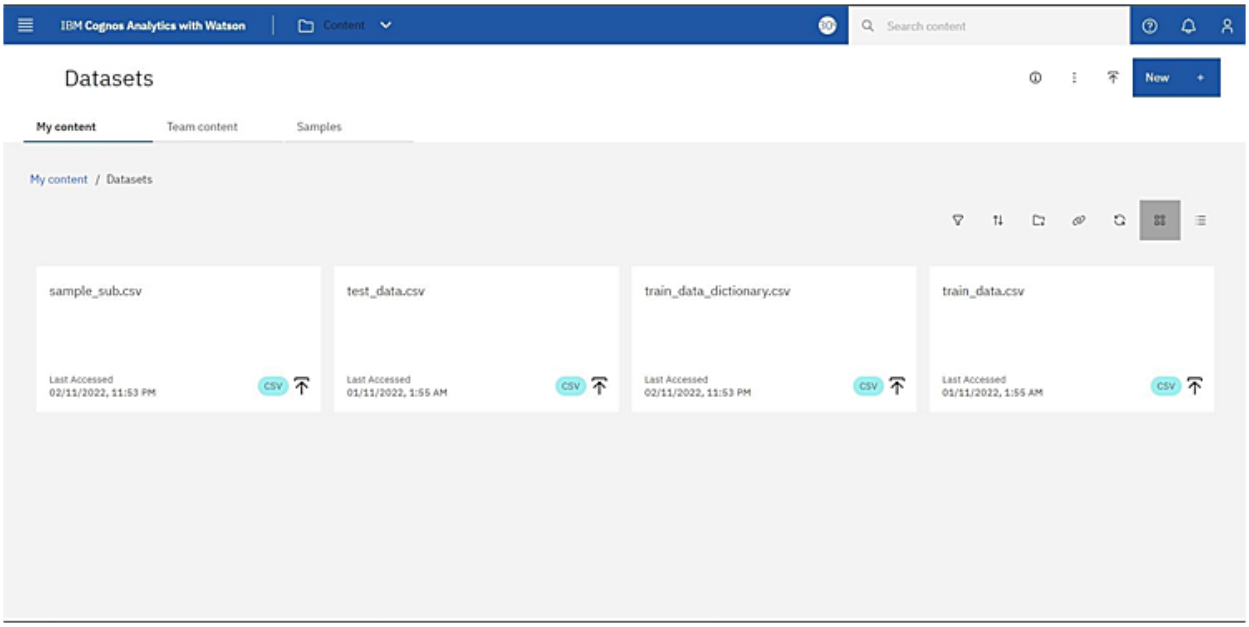
### Sprint Planning & Estimation
### Product backlog and sprint schedule:

| Sprint | Functional Requirement(Epic) | User StoryNumber | User Story / Task | Story Points | Priority | Team Members |
|---|---|---|---|---|---|---|
| Sprint-1 | Retrieve Data | USN-1 | As a user, I should get a clearer clinical contextfor AIDS patient's unique case | 10 | Medium | Vishali, Nanditha |
| Sprint-1 | Visualize the data | USN- 2 | As a user, I need a nicely visualized dashboard of the number of beds occupied and the number of free beds in a hospital. | 20 | High | Vaidegi, Divya darshani, Vishali |
| Sprint-2 | Track of patient visit of the Hospital | USN-3 | Tracking a patient's Health care over years of visits and screening of data they have in hospital. | 10 | Medium | Vaidegi, Nanditha |

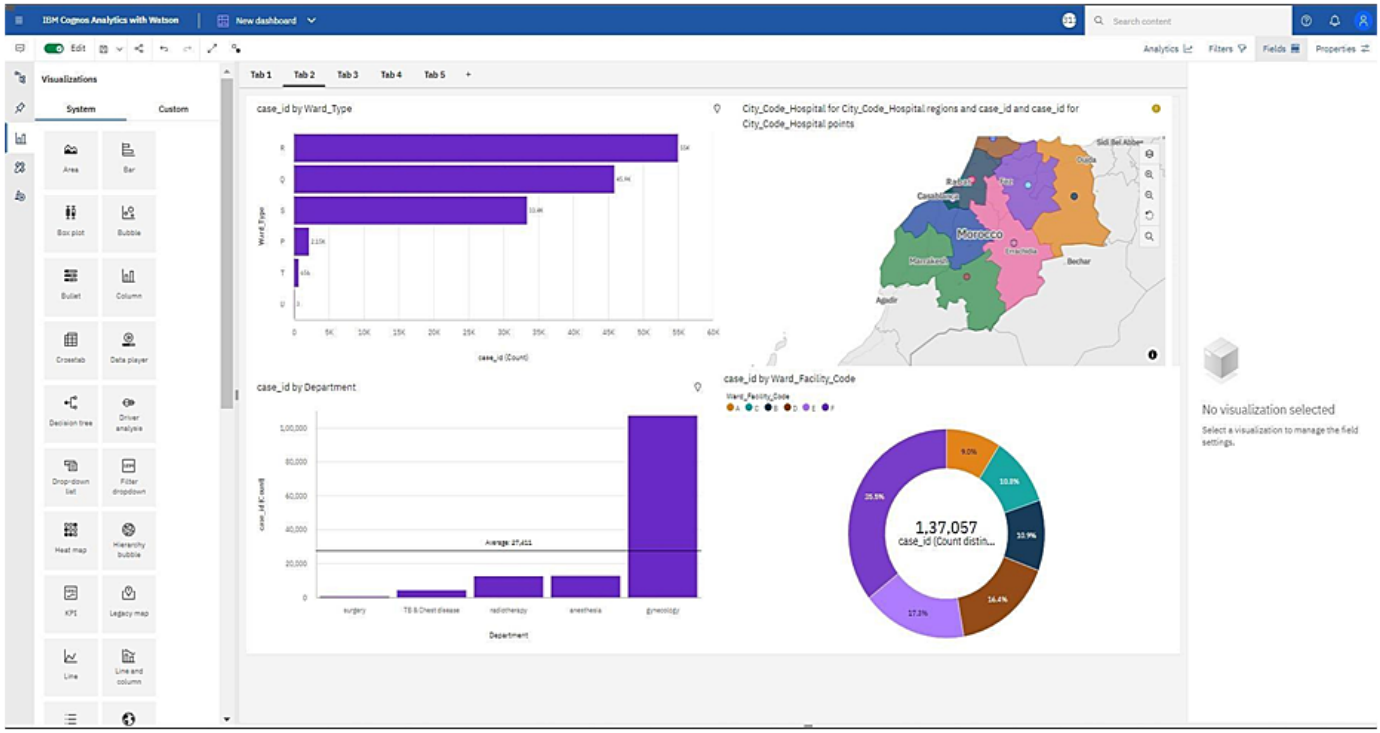
| | | | | | | |
|---|---|---|---|---|---|---|
| Sprint -2 | Dashboard | USN - 4 | As a user , I want the interactive dashboard to analyze the data & Have the data in terms of Graph. | 20 | High | Nanditha, Divya darshani |
| Sprint-3 | Detailed EHR's of patient | USN-5 | Provided greater details in the EHR's of the individual patient with clear idea of what to do. | 10 | Medium | Vishali, Divya darshani |
| Sprint-3 | Story Creation | USN-6 | As a user , I need the story animation of thedataset with insights | 20 | High | Vaidegi |
| Sprint-4 | Predict LOS | USN-7 | As a user, I want a flawless system to predictthe length of stay of the patients | 20 | High | Vishali, Nanditha, Divya darshani |

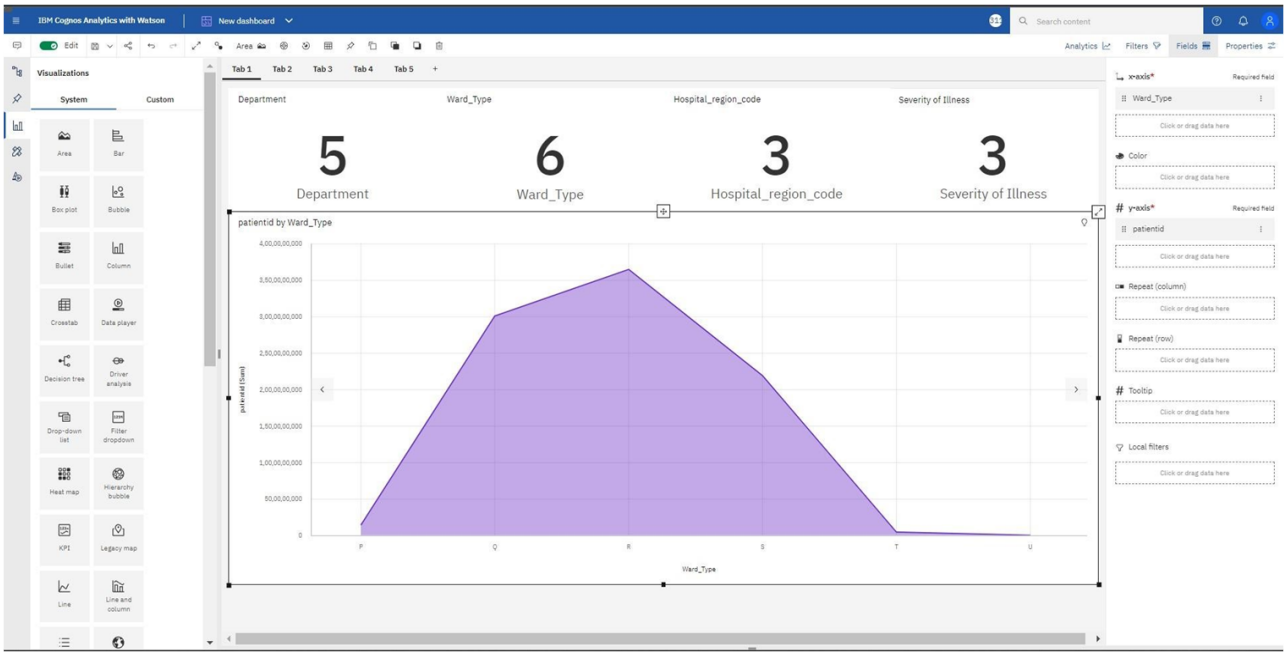
| | | | | | | |
|---|---|---|---|---|---|---|
| Sprint-4 | Using ML algorithm for Prediction | USN-8 | As a user, I must have prior knowledge of LOS can aid in logistics such as room and bed allocation planning. | 20 | High | Nanditha |

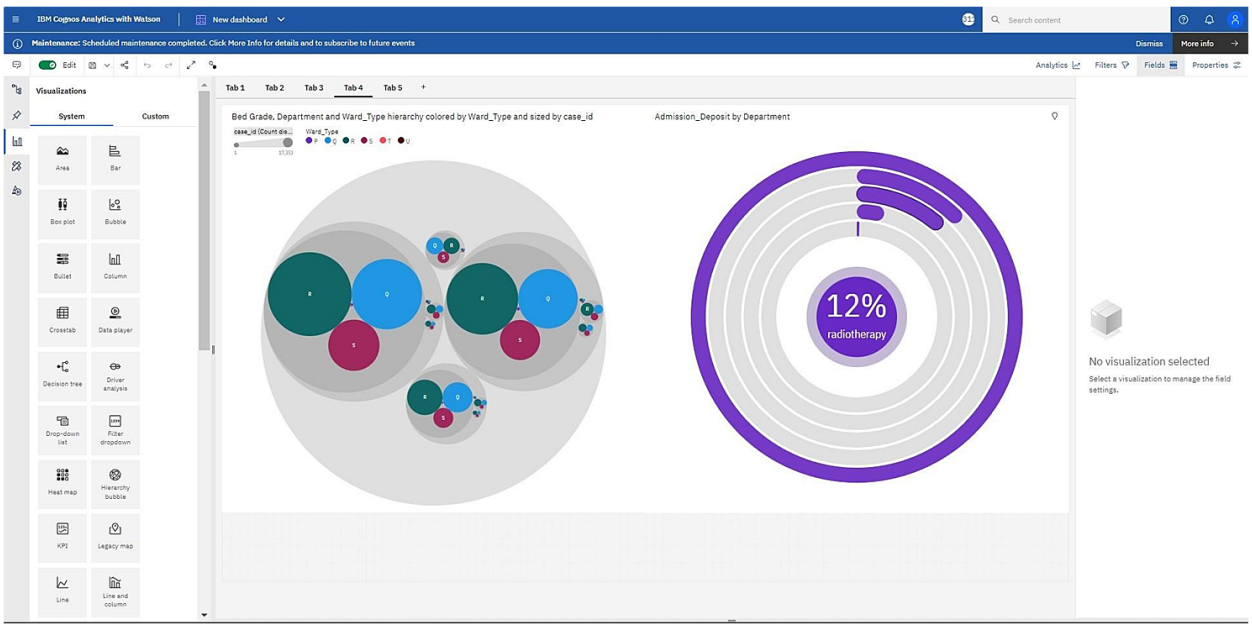
# 7. CODING AND SOLUTIONING




## EXPLORATION OF DATA


IBM Cognos Analytics with Watson
New dashboard
Search content
Edit
Analytics
Filters
Fields
Properties
Visualizations
System
Custom
Tab 1
Tab 2
Tab 3
Tab 4
Tab 5
case_id by Ward_Type
City_Code_Hospital for City_Code_Hospital regions and case_id and case_id for City_Code_Hospital points
Morocco
case_id by Department
surgery
TB & Chest disease
radiotherapy
anesthesia
gynecology
Department
case_id by Ward_Facility_Code
1,37,057
case_id (Count distin...
No visualization selected
Select a visualization to manage the field settings.


IBM Cognos Analytics with Watson
New dashboard
Search content
Edit
Analytics
Filters
Fields
Properties
Visualizations
System
Custom
Tab 1
Tab 2
Tab 3
Tab 4
Tab 5
Department
Ward_Type
Hospital_region_code
Severity of Illness
5
Department
6
Ward_Type
3
Hospital_region_code
3
Severity of Illness
patientid by Ward_Type
Ward_Type
x-axis*
Required field
Color
y-axis*
patientid
Repeat (column)
Repeat (row)
Tooltip
Local filters




# 8. RESULT

## 8.1 PERFORMANCE METRICS

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

.

# 9.ADVANTAGES AND DISADVANTAGES:

## 9.1 ADVANTAGES

- Improved diagnostics
- Better patient care
- Better health outcomes
- Improved patient interactions
- Increased health indicators

## 9.2 DISADVANTAGES

- Man Power
- Privacy
- Quality of Data Input
- Data Safety
- Replacing Medical Personnel

# 10. CONCLUSION

Big data analytics in healthcare is evolving into a promising field for providing insight from very large data sets and improving outcomes while reducing costs. Its potential is great; however there remain challenges to overcome.

# 11.FUTURE SCOPE

Electronic Health Records (EHR's)- This is the most common application of big data analytics in healthcare in the United States, where many hospitals have adopted EHRs. One of the key factors of an EHR is that health information can be created and managed by authorized healthcare providers in a digital format capable of being shared with other providers across more than one healthcare organization.

- Enhancing Patient Engagement- Patients use devices like AppleWatch and FitBit to monitor their health data. The healthcare field is wide open to data leaders who can leverage preventative analytics to help patients change to healthier lifestyles.
- Predictive Analytics- aims to alert healthcare professionals, clinicians, caregivers, etc., of the probability and outcomes before they occur, helping them to prevent or cure health issues. Predictive analytics uses techniques from healthcare data mining, statistics, predictive modeling, machine learning, and artificial intelligence to analyze current data and make predictions about the future.