

Analytics for Hospital's Health Care Data

**NALAIYA THIRAN PROJECT REPORT
2022**

Submitted by

Mohamed Abdul Fazil	312419205059
Abinesh R	312419205005
Jeno Allwyn B	312419205038
John Nesa Kumar S	312419205041

Team ID: PNT2022TMID28411

Content

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

6. PROJECT PLANNING & SCHEDULING

- 6.1 Sprint Planning & Estimation
- 6.2 Sprint Delivery Schedule

7. CODING & SOLUTIONING (Explain the features added in the project along with code)

- 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

Analytics for Hospital's Health Care Data

1. Introduction

1.1 Project Overview

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.

1.2 Purpose

The goal 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.

2. Literature Survey

2.1 Existing Problem

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

2.2 References

1. Proposed application of big data analytics in healthcare at Maharaja Yeshwantrao Hospital

This paper gives an insight of how we can store healthcare data digitally like patient's records as an Electronic Health Record (EHR) and how we can generate useful information from these records by using analytics techniques and tools which will help in saving time and money of patients as well as the doctors. This paper is fully focused towards the Maharaja Yeshwantrao Hospital (M.Y.) located in Indore, Madhya Pradesh, India. M.Y hospital is the central India's largest government hospital. It generates large amount of heterogeneous data from different sources like patients health records, laboratory test result, electronic medical equipment, health insurance data, social media, drug research, genome research, clinical outcome, transaction and from Mahatma Gandhi Memorial medical college which is under MY hospital. To manage this data, data analytics may be used to make it useful for retrieval. Hence the concept of "big data" can be applied. Big data is characterized as extremely large data sets that can be analysed computationally to find patterns, trends, and associations, visualization, querying, information privacy and predictive analytics on large wide spread collection of data. Big data analytics can be done using Hadoop which plays an effective role in performing meaningful real-time analysis on the large volume of this data to predict the emergency situations before it happens. This paper also discusses about the EHR and the big data usage and its analytics at M.Y. hospital

2. A look at challenges and opportunities of Big Data analytics in healthcare

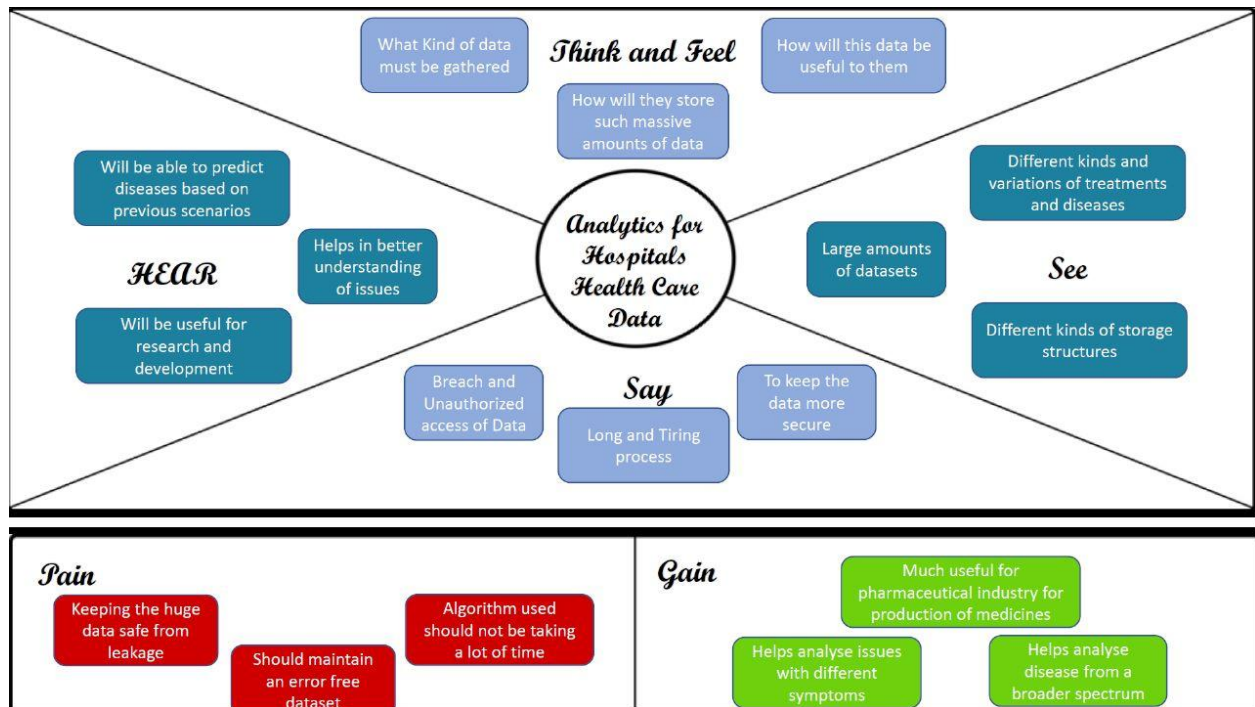
Big Data analytics can revolutionize the healthcare industry. It can improve operational efficiencies, help predict and plan responses to disease epidemics, improve the quality of monitoring of clinical trials, and optimize healthcare spending at all levels from patients to hospital systems to governments. This paper provides an overview of Big Data, applicability of it in healthcare, some of the work in progress and a future outlook on how Big Data analytics can improve overall quality in healthcare systems.

2.3 Problem Statement Definition

The analysis using hospital healthcare data allows us to identify between high risk cases and low risk cases. Prior hospital data provides us with information as to how long the patient was admitted to the hospital and the medication administered to them. Analysing this data can provide us future insight as to how long the patient will stay and helps us in bed allocation and medicine and personnel management.

3. Ideation and Proposed Solution

3.1 Empathy Map



3.2 Ideation and Brainstorming

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

Analytics for Hospitals' Health-Care Data

Brainstorm & idea prioritization

1 Define your problem statement

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 (LOS) is a 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.

Brainstorm

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

10 minutes

	Adithy R	Jeno Alwyn	John Nesa Kumar S	Mohamed Abdul Fazil
Identify the problem statement	Healthcare management	Identify the most common LOS risk factors	Identify the most common LOS risk factors	Identify the most common LOS risk factors
Identify the problem statement	Healthcare management	Identify the most common LOS risk factors	Identify the most common LOS risk factors	Identify the most common LOS risk factors
Identify the problem statement	Healthcare management	Identify the most common LOS risk factors	Identify the most common LOS risk factors	Identify the most common LOS risk factors
Identify the problem statement	Healthcare management	Identify the most common LOS risk factors	Identify the most common LOS risk factors	Identify the most common LOS risk factors
Identify the problem statement	Healthcare management	Identify the most common LOS risk factors	Identify the most common LOS risk factors	Identify the most common LOS risk factors

Step-2: Brainstorm, Idea Listing and Grouping

3

Group Ideas

Take turns sharing your ideas while clustering similar or related notes as you go. In the last 10 minutes, 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



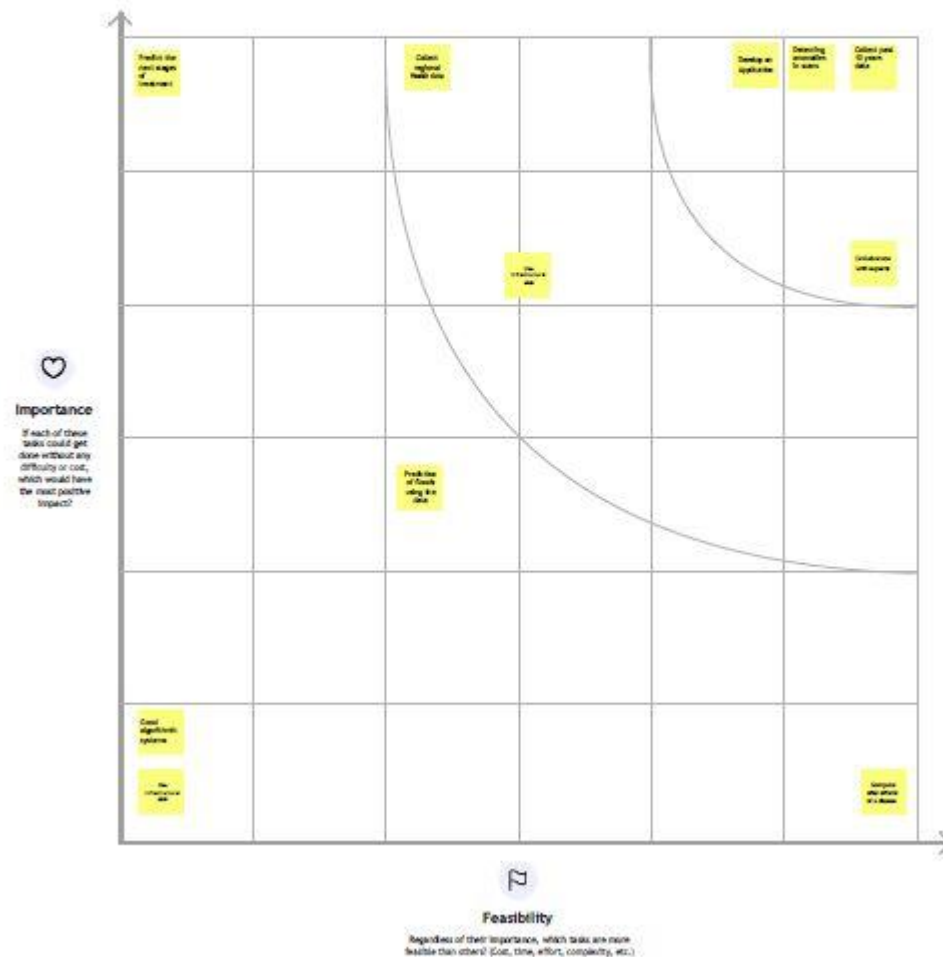
Step-3: Idea Prioritization



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



3.3 Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	The goal 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.
2.	Idea / Solution description	<p>Use various factors which determine the LOS like:</p> <ul style="list-style-type: none"> • Similarity in Patient health conditions • Similarities in previous history of illness • Predictive analysis bases on seasons • Predictive analysis based on age groups • Doctor's Advice regarding LOS • Add bed days for each discharged patient and divide the sum by the number of discharged patients • General predictions based on Oxygen Delivery Index, RBC count, CreatinineLevels etc.
3.	Novelty / Uniqueness	<p>The uniqueness of this proposed system is that data is not gathered by a single source but from multiple departments from the hospitals thereby reducing redundant data</p> <p>The proposed dashboard will make it easy for anyone to analyse and accommodate based on priority and requirements</p>
4.	Social Impact / Customer Satisfaction	<p>The various social impacts / Customer Satisfaction are:</p> <ul style="list-style-type: none"> • Easy to access by the public and prepare accordingly • People need not worry about their LOS and can simply rely on this data

5.	Business Model (Revenue Model)	<p>This system when sold as a business to hospitals as clients will generate huge amounts of revenue as this is a system which will help in resource allocation, budget allocations, pharmaceuticals etc.</p> <p>Previous medical record and records from this dashboard together will make this system reliable at all instances and help in the overall RND of the hospitals and the quality of treatment provided</p>
6.	Scalability of the Solution	<p>Easy to be implemented as this does not involve any special hardware.</p> <p>Data inputs are from the hospitals directly hence no outsourcing is required for datasets</p> <p>In case of sudden surges and sudden mutations of any disease the solution will require a minimum of 14 days to understand the system properly</p>

3.3 Problem Solution Fit

Define CS, fit into CC	1. CUSTOMER SEGMENT(S) CS <ol style="list-style-type: none"> 1. Sign In 2. Upload 3. Modify 4. Review 5. Settings 	6. CUSTOMER CONSTRAINTS CC <ol style="list-style-type: none"> 1. Simple to use and visualize the data. 2. Can work with data in limited Time. 3. It must give real time Overview of Data. 4. Graphically pleasing Display and Very user friendly 	5. AVAILABLE SOLUTIONS AS <ol style="list-style-type: none"> 1. Providing necessary Input to the tool. 2. Avoiding Human Errors. 3. Avoiding Usage in Remote areas. 4. Network Stability. 5. Using Consistent Data. 	Explore AS, differentiate
	2. JOBS-TO-BE-DONE / PROBLEMS J&P <p>Jobs to be done :</p> <ol style="list-style-type: none"> a. Upload the patient dataset b. Prepare Data c. Exploring the data d. Perform the metrics and rules. e. Visualizing the data <p>Problems :</p> <ol style="list-style-type: none"> a. Incorrect input b. Data Latency support c. Poor Network Standard 	9. PROBLEM ROOT CAUSE RC <ol style="list-style-type: none"> 1. The Customer is located far from the City. 2. Misunderstanding of Customer while using the Product tool. 3. Bandwidth of the device does not support the product tool. 4. Lack of Communication Inconsistent Data. 	7. BEHAVIOUR BE <ol style="list-style-type: none"> 1. It can transfer Information Quickly. 2. Visualizes trends and changes in data Over time. 3. Widgets and data Components are Effectively presented. 4. Easily Customizable. Displays Output Clearly. 	

Identify strong TR & EM	3. TRIGGERS TR 1. Redo whole Process. 2. Takes Longer Time than usual.	10. YOUR SOLUTION SL a. Grouping related metrics. b. Using most efficient Visualization. c. Rounding off the numbers in the product. d. Use Size and position to show hierarchy. e. Includes only essential data. f. Short and Precise and is interactive. g. Evolving products from its negatives.	8. CHANNELS OF BEHAVIOUR CH a. Right Visualization to depict the data. b. Choosing Critical data to observe. c. Simple Color Scheme and smart Design elements. d. Incorporating drilldowns to show more Details e. Branding the product	Identify strong TR & EM
	4. EMOTIONS: BEFORE / AFTER EM Before: As expected, to work in time deliverable. After: Delay due to the Problems that were triggered and makes Frustration.			

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 Confirmation via OTP
FR-3	Operability	Share patient data and make it interoperable among the management
FR-4	Accuracy	The dashboard will be able to predict length of stay based on multiple combinations based on input sources with a n accuracy of upto 85%
FR-5	Compliance	The product is to be used within the hospital so any form of data need not be hidden
FR-6	Productivity	The dashboard is believed to improve the predictions of Length of Stay and thereby creating a scenario of providing better solution

4.2 Non-functional Requirements:

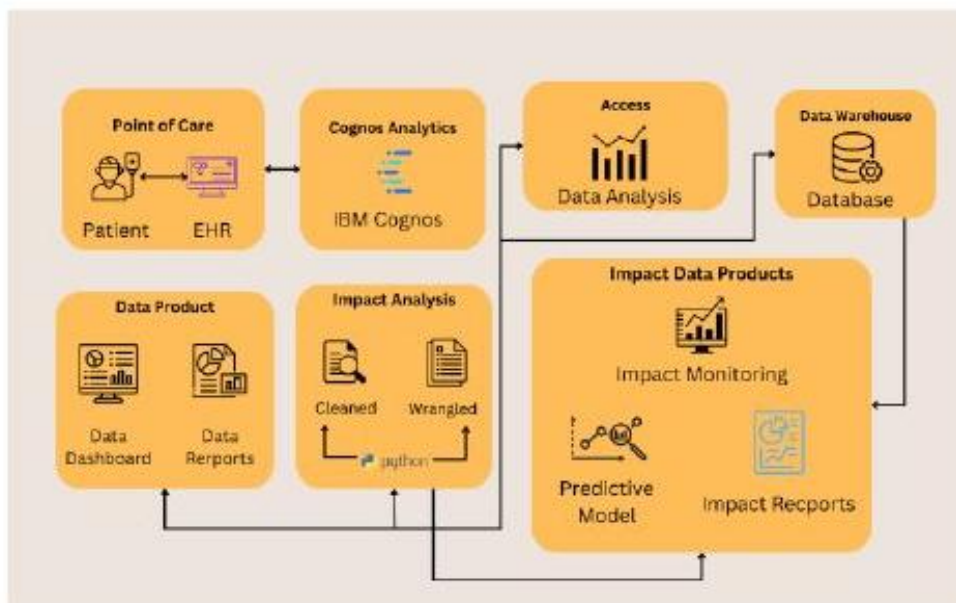
Following are the non-functional requirements of the proposed solution.

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	General industry level security shall be provided

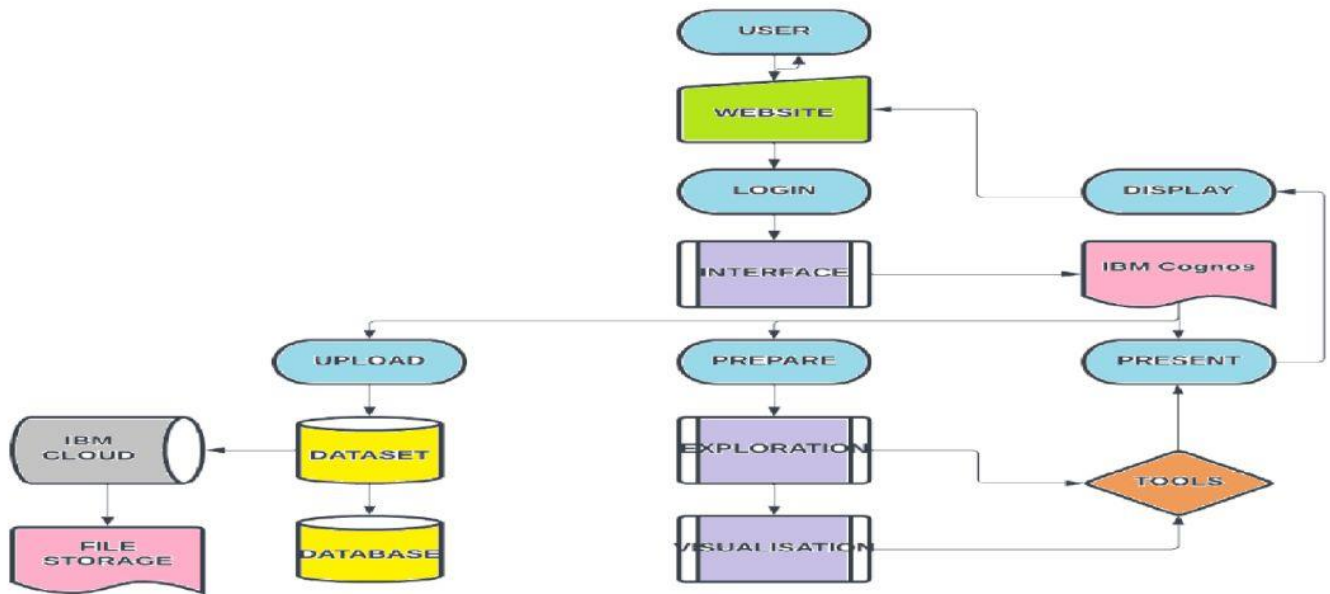
NFR-3	Reliability	This dashboard will be consistent and reliable to the users and helps the user to use in effective ,efficient and reliable manner.
NFR-4	Performance	The dashboard reduces the time needed for analysing data and has an automated system for that which improves the performance
NFR-5	Availability	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	It is a multi tenant system which is capable of rimming on lower level systems as well.

5.Project Design

5.1 Data Flow Diagram



5.2 Solution and Technical Architecture



6. Planning and Execution

Project Planning and Scheduling Script

6.1 Script Planning and Execution

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Retrieve Data	USN-1	As a user, I should get clearer clinical contextfor AIDS patient's unique case	10	High	John Nesa Kumar S
Sprint-1	Visualize the data	USN-2	As a user, I need nicely visualized dashboard ofnumber of beds occupied and number of free beds in hospital.	20	High	Mohamed Abdul Fazil
Sprint-2	Track of patientvisit of Hospital	USN-3	Tracking a patient Health care over years ofvisit and Screening of data they have in hospital.	10	Low	John Nesa Kumar S
Sprint-2	Dashboard	USN-4	As a user, I want the interactive dashboard toanalyse the data. Have the data in terms of Graph.	20	Medium	Mohamed Abdul Fazil
Sprint-3	Detailed EHR's ofpatient	USN-5	Provided greater details in the EHR's of individual patient with clear idea of what to do.	10	High	Abinesh R
Sprint-3	Story Creation	USN-6	As a user, I need the story animation of thedata set with insights	20	High	Jeno Allwyn
Sprint-4	Predict LOS	USN-7	As a user, I want the flawless system to predictthe length of stay of the patients	20	Medium	Abinesh
Sprint-4	Using ML algorithmfor Prediction	USN-8	As a user, I need prior knowledge of LOS can aid in logistics such as room and bedallocation planning.	20	High	Jeno Allwyn
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Retrieve Data	USN-1	As a user, I should get clearer clinical contextfor AIDS patient's unique case	10	High	John Nesa Kumar S
Sprint-1	Visualize the data	USN-2	As a user, I need nicely visualized dashboard ofnumber of beds occupied and number of free beds in hospital.	20	High	Mohamed Abdul Fazil
Sprint-2	Track of patientvisit of Hospital	USN-3	Tracking a patient Health care over years ofvisit and Screening of data they have in hospital.	10	Low	John Nesa Kumar S

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
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022	Sprint-1
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022	Sprint-2
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022	Sprint-3
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022	Sprint-4

7. Coding And Solutioning

```
<!DOCTYPE html>
<html lang="en">

<head>
  <title>Data Analytics</title>
  <meta charset="utf-8">
  <meta name="viewport" content="width=device-width, initial-scale=1">

  <link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/css/bootstrap.min.css">
  <script src="https://ajax.googleapis.com/ajax/libs/jquery/3.6.0/jquery.min.js"></script>
  <script src="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/js/bootstrap.min.js"></script>
</head>

<body>

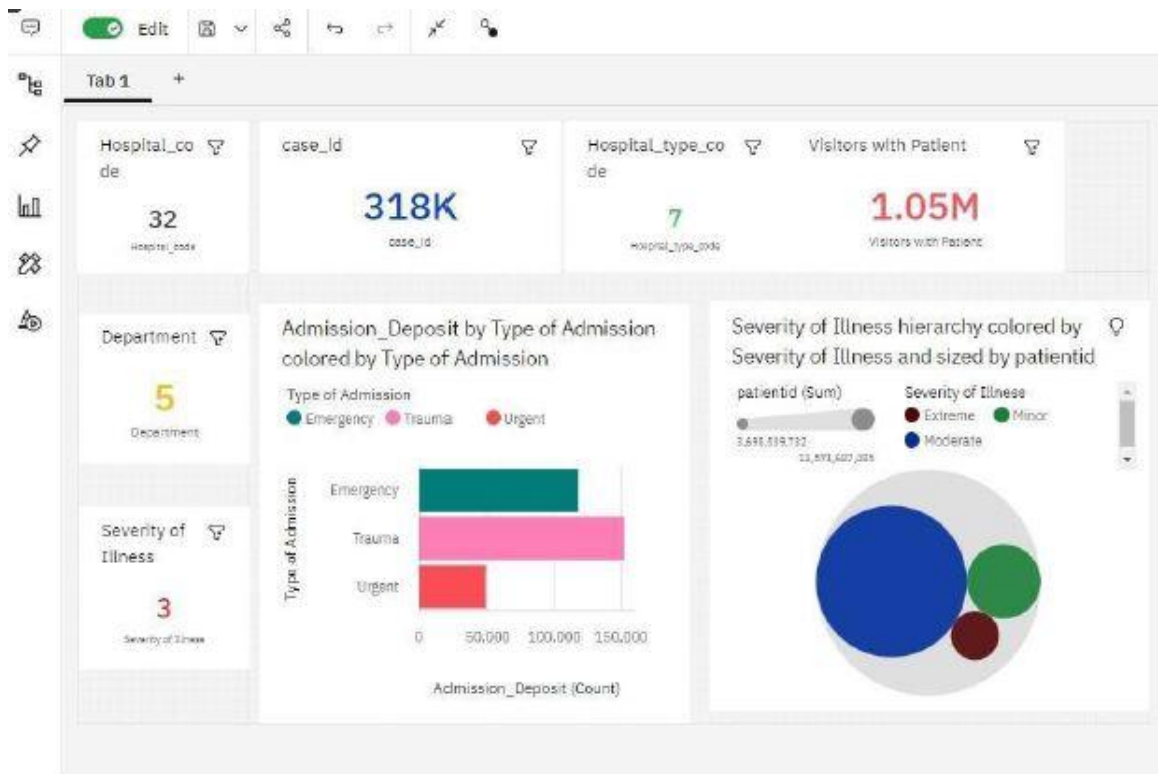
  <nav class="navbar navbar-inverse ">
    <div class="container-fluid">
      <div class="navbar-header">
        <a class="navbar-brand" href="#">Analytics for Hospitals' Health-Care Data</a>
      </div>
      <ul class="nav navbar-nav">
        <li><a href="index.html">Home</a></li>
        <li class="active"><a href="#">Dashboard</a></li>
        <li><a href="report.html">Report</a></li>
        <li><a href="story.html">Story</a></li>
      </ul>
    </div>
  </nav>

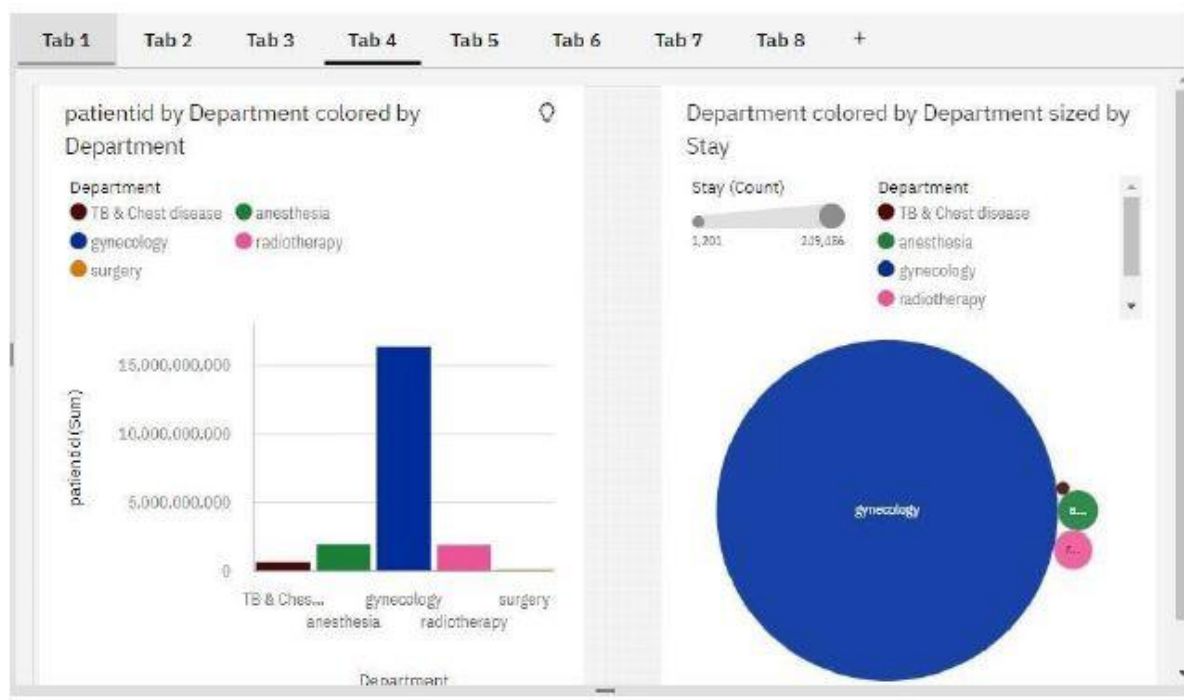
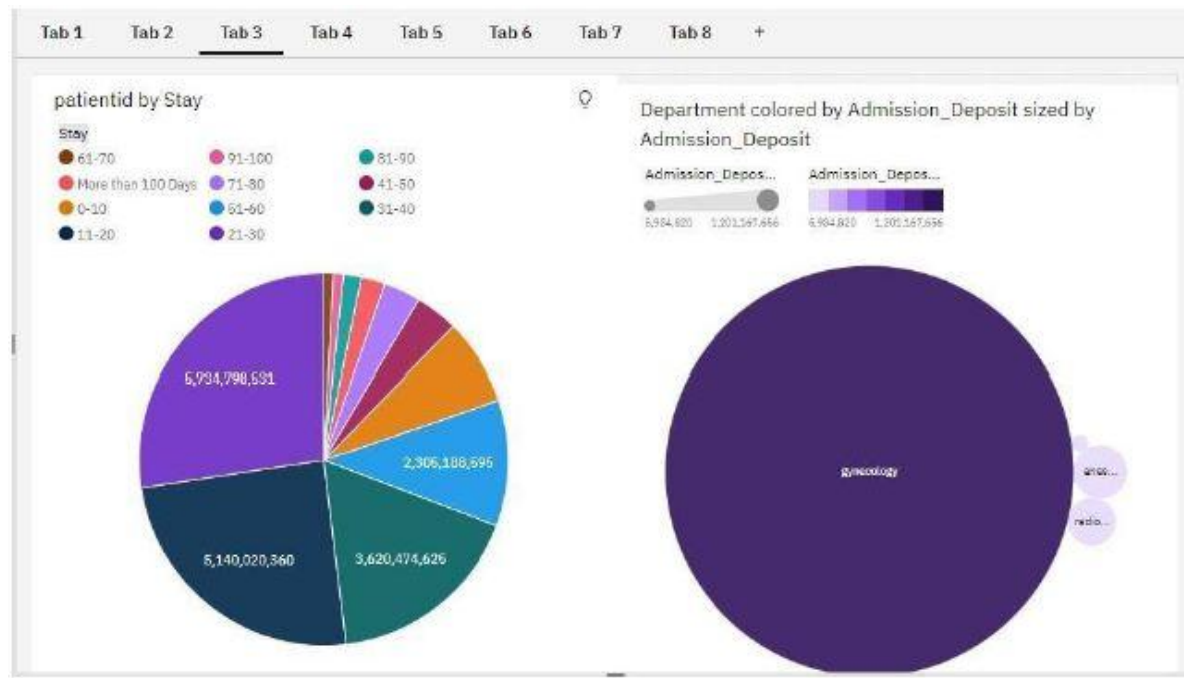
  <div class="container">
    <iframe
src="https://us3.ca.analytics.ibm.com/bi/?perspective=dashboard&pathRef=.my_folders%2FMy%2Bfirst%
2Bdashboard&closeWindowOnLastView=true&ui_appbar=false&ui_navbar=false&shareM
ode=embedded&action=view&mode=dashboard&subView=model000001849013f37c_00000002
"
      width="1020" height="720" frameborder="0" gesture="media" allow="encrypted-media"
allowfullscreen=""></iframe>
    </div>

  </body>

</html>
```

7.1 Dashboard





8. Testing

8.1 User Acceptance Test

Resolution	Severity1	Severity2	Severity3	Severity4	Subtotal
By Design	5	5	3	0	13
Duplicate	0	0	0	0	0
External	7	5	1	0	13
Fixed	11	8	7	5	31
Not Reproduced	1	0	0	0	1
Skipped	0	0	0	0	0
Won't Fix	0	0	0	0	0
Totals	24	18	11	5	58

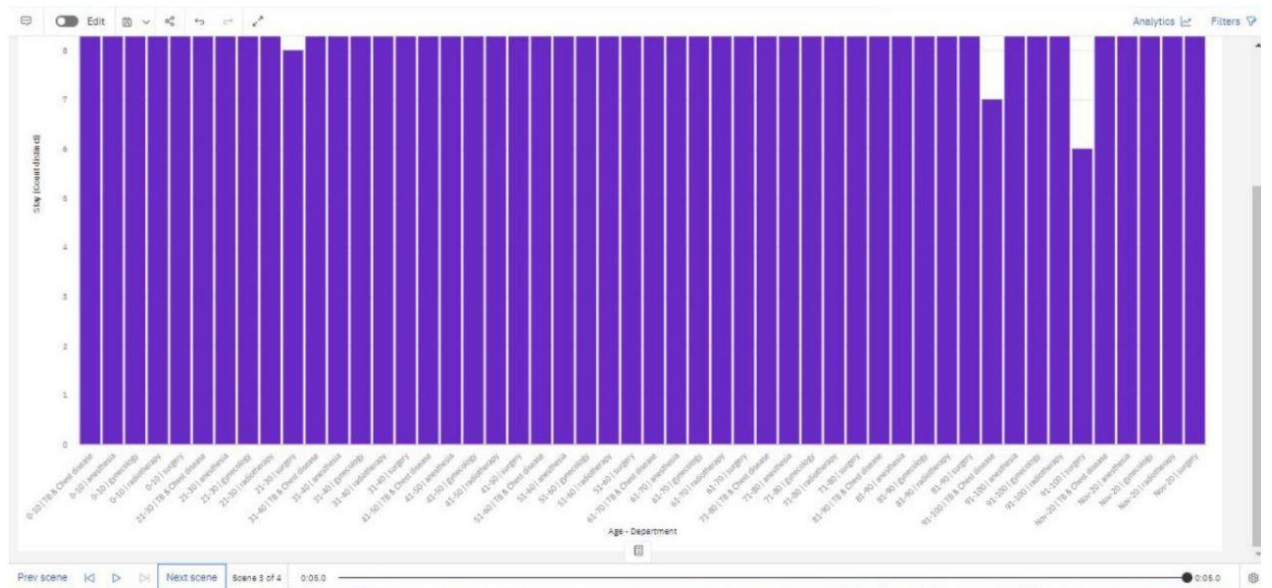
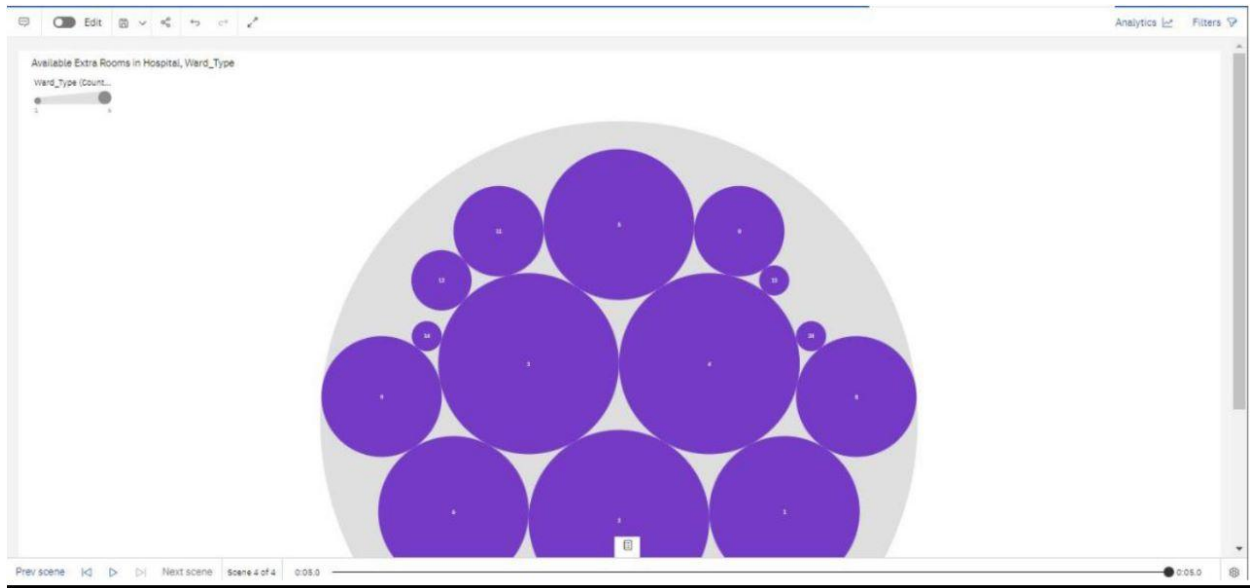
Section	Total Cases	Not Tested	Fail	Pass
Print Engine	0	0	0	0
Client Application	0	0	0	0
Security	0	0	0	0
Outsource Shipping	0	0	0	0
Exception Reporting	0	0	0	0
Final Report Output	0	0	0	0
Version Control	0	0	0	0

8.2 Performance Testing

S.No.	Parameter	Screenshot / Values
1.	Dashboard design	17 / 5
2.	Data Responsiveness	The final output from IBM Cognos With Watson further converted into PDF or Story file, so it can be viewed by all devices.
3.	Amount Data to Rendered (DB2 Metrics)	0 KB.
4.	Utilization of Data Filters	The Utilization of data Filters like Ascending, Descending, Format and so on.
5.	Effective User Story	12
6.	Descriptive Reports	17 / 5

9. Result





10. Advantages & Disadvantages

Advantages:

- Easy to calculate a patient's length of stay.
- It is used to indicate the efficiency.
- Helpful for extension of beds based on patient's length of stay.
- User Friendly
- Easy to understand
- Secure

Disadvantages:

- Need a more dynamic User interface
- Users need to know all the fields.
- Does Not take null value as input
- Does not provide suggestions to the user.

11. Conclusion

There are several complications when it comes to healthcare. We can reduce the risk of complications with better analysis and treatment. The solution that we provide with the analytical data will help the hospitals.

12. Future Scope

There are many possible improvements that could be explored to improve the scalability and accuracy of this prediction system. As we have developed a generalized system, in future we can use this system for the analysis of different data sets. The performance of the health's diagnosis can be improved significantly by handling numerous class labels in the prediction process, and it can be another positive direction of research.

13. Appendix

Source Code:

https://drive.google.com/drive/folders/19vwNM2MeW5Maa2wu8YSis5TOQ5klL3s?usp=share_link

Github Link:

<https://github.com/IBM-EPBL/IBM-Project-21162-1659774449.git>