

TEAM ID	PNT2022TMID19527
PROJECT TITLE	Analytics for Hospital Health-Care Data

Analytics for Hospital Health-Care Data Report

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

7. SOLUTIONING

- 7.1 Feature 1
- 7.2 Feature 2

8. TESTING

- 8.1 Test Cases

9. RESULTS

- 9.1 Performance Metrics

10. ADVANTAGES & DISADVANTAGES

11. CONCLUSION

12. FUTURE SCOPE

13. APPENDIX

GitHub & Project Demo Link

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. In addition, prior knowledge of LOS can aid in logistics such as room and bed allocation planning. What if I 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. 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.

1.2 PURPOSE

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.

2.LITERATURE SURVEY

2.1 EXISTING PROBLEM

S.NO	TITLE	AUTHOR	DESCRIPTION
1	Data-driven Methods for Typical Treatment Pattern Mining	Chonghui Guo and jingfeng Chen	A huge volume of digitized clinical data is generated and accumulated rapidly since the widespread adoption of Electronic Medical Records [EMRs]

2	Big Data in Supply Chain Management	Aniket Nargundkar and J.Kulkarni	A robust model is designed to help hospital administration to predict patient's Length Of Stay [LOS] to resolve the issues faced by hospital
3	A Systematic Mapping study	Anil Pandit and Sharvari Shukla	The current study performs a Systematic Literature Review (SLR) to synthesise prior research on the applicability of Big Data Analytics (BDA) in healthcare

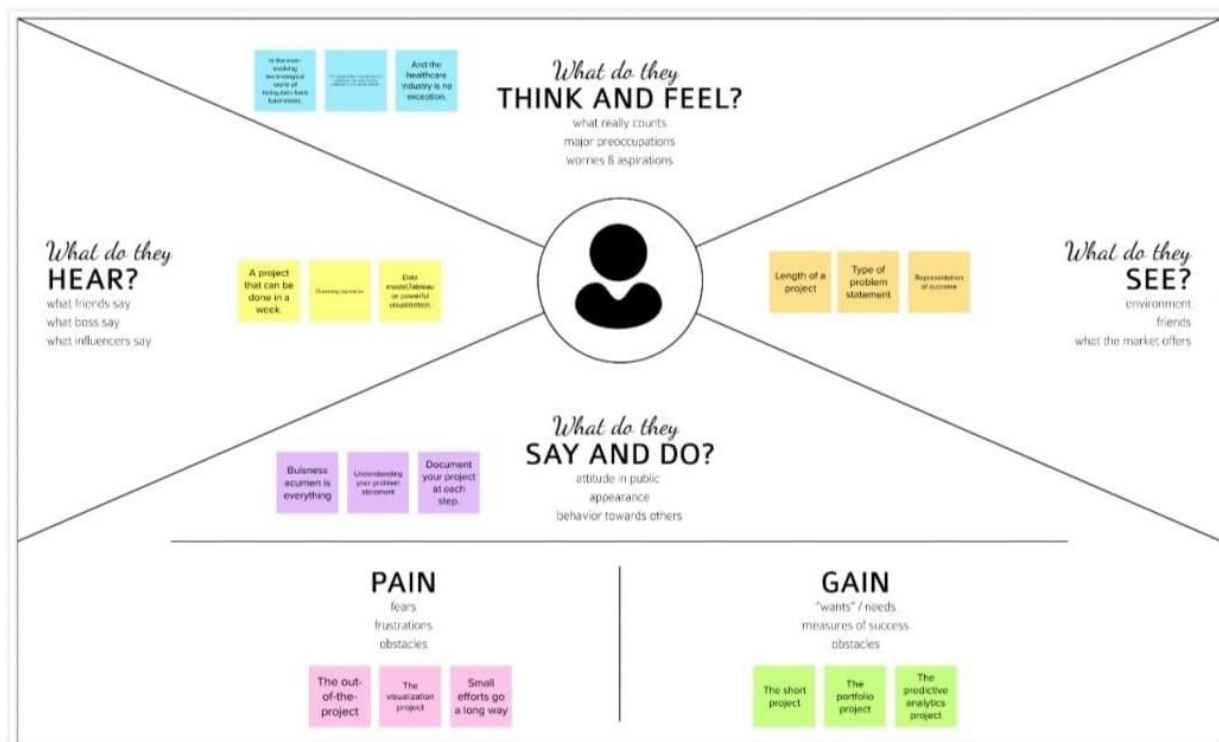
2.2 REFERENCE

2.3 PROBLEM STATEMENT

Create a problem statement to understand your customer's point of view. The Customer Problem Statement template helps you focus on what matters to create experiences people will love. A well-articulated customer problem statement allows you and your team to find the ideal solution for the challenges your customers face. Throughout the process, you'll also be able to empathize with your customers, which helps you better understand how they perceive your product or service.

3. IDEATION AND PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS



3.2 IDEATION AND BRAINSTORMING

Brainstorm & idea prioritization

Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room.

- ⌚ 40 minutes to prepare
- 🕒 1 hour to collaborate
- 👥 2-8 people recommended

 [Share template feedback](#)



Before you collaborate

A little bit of preparation goes a long way with this session. Here's what you need to do to get going.

① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩ ⑪ ⑫ ⑬ ⑭ ⑮ ⑯ ⑰ ⑱ ⑲ ⑳ ㉑ ㉒ ㉓ ㉔ ㉕ ㉖ ㉗ ㉘ ㉙ ㉚ ㉛ ㉜ ㉝ ㉞ ㉟ ㊱ ㊲ ㊳ ㊴ ㊵ ㊶ ㊷ ㊸ ㊹ ㊺ ㊻ ㊼ ㊽ ㊾ ㊿ ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩ ⑪ ⑫ ⑬ ⑭ ⑮ ⑯ ⑰ ⑱ ⑲ ⑳ ㉑ ㉒ ㉓ ㉔ ㉕ ㉖ ㉗ ㉘ ㉙ ㉚ ㉛ ㉜ ㉝ ㉞ ㉟ ㊱ ㊲ ㊳ ㊴ ㊵ ㊶ ㊷ ㊸ ㊹ ㊺ ㊻ ㊼ ㊽ ㊾ ㊿

- A Team gathering**
Briefing, where shared participants to the problem, each about a problem. Share relevant information or pre-work shared.
 - B Set the goal**
Think about the problem you'll be focusing on solving in the brainstorming session.
 - C Learn how to use the facilitation tools**
Use the Facilitation Superpowers to run a happy and productive session.

[Open article](#) →

[Open a article](#) →

1 Define your problem statement

Recent Covid-19 Pandemic has raised alarms over one of the most overlooked areas to focus: Healthcare. The goal is to accurately predict the Length of Stay for each patient on case by case basis so that the Hospitals Run 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 problem is to accurately predict the Length of Stay for each patient.

Key rules of brainstorming

To run an efficient and productive session

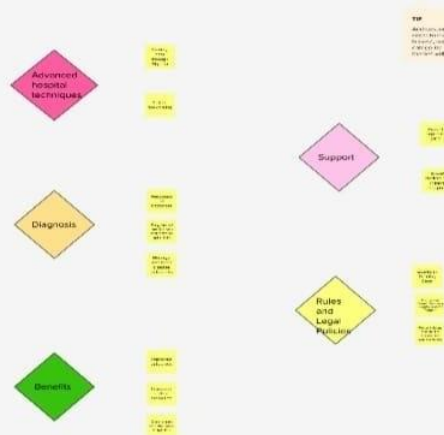
- | | | | |
|---|------------------|---|-------------------------|
| | Stay in facts. | | Encourage wild ideas. |
|  | Defer judgment. |  | Listen to others. |
|  | Go for quantity. |  | If possible, be visual. |

[illegible][illegible]

3 Group ideas

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

© 20 minutes



TIP
 Ankle sprains, which are the most common type of injury, can be prevented by wearing proper footwear and using proper technique when jumping or landing.

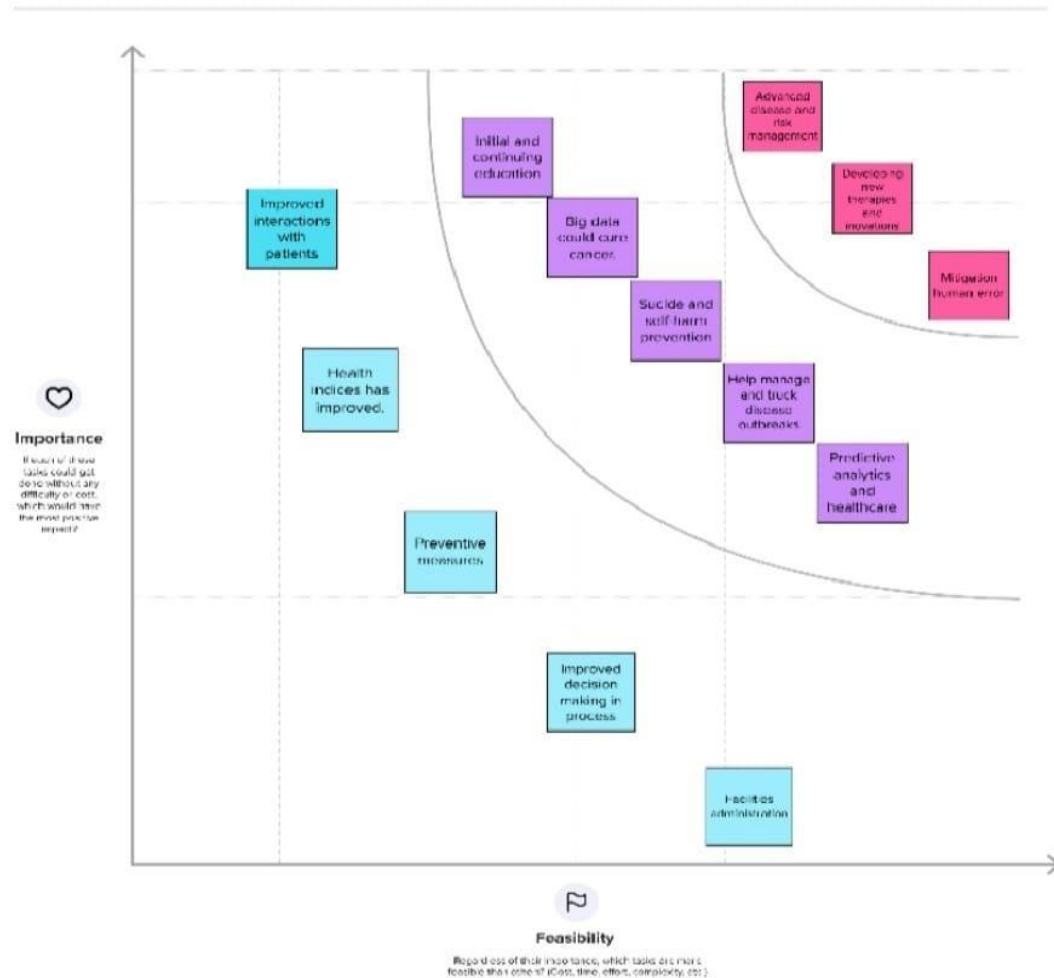


4

Prioritize

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

⌚ 20 minutes



3.3 PROPOSED SOLUTION

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	EHR data matched patient-reported data in 23.5 percent of records in a study at an ophthalmology practise. Patients' EHR data did not agree in any way when they reported having three or more eye health complaints.
2.	Idea / Solution description	Predictive analytics can create patient journey dashboards and disease trajectories that can lead to effective, and result-driven healthcare. It improves treatment delivery, cuts costs, improves efficiencies, and so on.
3.	Novelty / Uniqueness	Healthcare data frequently resides in several locations. from various departments, such as radiology or pharmacy, to various source systems, such as EMRs or HR software. The organisation as a whole contributes to the data. This data becomes accessible and usable when it is combined into a single, central system, such as an enterprise data warehouse (EDW).
4.	Social Impact / Customer Satisfaction	Enhanced diagnosis Improved medical treatment Improved health results Improved relationships with patients More positive health indicators
5.	Business Model (Revenue Model)	The two factors that have the biggest negative effects on hospital income are claim denials and patient incapacity to pay their part. 90% more uncollectible claim denials were written off by hospitals and healthcare systems in 2017 compared to the preceding six years.
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

Define CS, fit into CC	1. CUSTOMER SEGMENT(S) CS Who is your customer? i.e. working parents of 0-5 y.o. kids	6. CUSTOMER CONSTRAINTS CC What constraints prevent your customers from taking action or limit their choices of solutions? i.e. spending power, budget, no cash, network connection, available devices.	5. AVAILABLE SOLUTIONS AS Which solutions are available to the customers when they face the problem or need to get the job done? What have they tried in the past? What price & come do these solutions bring? i.e. pen and paper is an alternative to digital notetaking	Explore AS, differentiate
	Various patient demographics, including risk level and insurance status, can be used to segment the patients. It is the method of classifying patients usually by age, gender, illness, belief, lifestyle	Avoidable medical errors. Low treatable mortality rates. Lack of transparency. Difficulty finding a good doctor. High maintenance costs. The lack of insurance coverage. The shortage of nurses and doctors. A different perspective on solving the shortage crisis.	Higher taxes on alcohol and tobacco. Improve fitness standards. Improve research. Transnational support. Reduction in consumption. Recycle and reuse. Reduce corruptive actions. Promote vaccinations.	
Focus on J&P, tap into BE, understand RC	2. JOBS-TO-BE-DONE / PROBLEMS J&P Which jobs-to-be-done (or problems) do you address for your customers? There could be more than one, explore different sides.	9. PROBLEM ROOT CAUSE RC What is the real reason that this problem exists? What is the back story behind the need to do this job? i.e. customers have to do it because of the change in regulations	7. BEHAVIOUR BE What does your customer do to address the problem and get the job done? i.e. directly related: find the right solar panel installer, calculate usage and benefits; indirectly associated: customer's spend free time on volunteering work (i.e. Greenpeace)	Focus on J&P, tap into BE, understand RC
	The fact that the responsibility for managing patients is split between their insurer and numerous healthcare providers presents one of the largest hurdles in the deployment of healthcare data analytics. Problems: 1. poor infrastructure 2. inadequate workforce 3. unmanageable patient burden 4. Ambiguous quality of service 5. high expense	Disease caused by Viruses, Bacteria, Fungi and Parasites How they cause damage: They invade living, normal cells and use those cells to multiply and produce other viruses like themselves Solutions: Handle & Prepare Food Safely Wash Hands Often Clean & Disinfect Commonly Used Surfaces Cough & Sneeze Into Your Sleeve Don't Share Personal Items Get Vaccinated	Disruptive conduct as they've an altered intellectual degree of worry of being sick, stressful approximately out of the pocket cost, alteration of way of life if suffered from a continual illness	
Define CS, fit into CL	3. TRIGGERS TR What triggers customers to act? i.e. seeing their neighbour installing solar panels, reading about a more efficient solution in the news.	10. YOUR SOLUTION SL What kind of solution suits Customer scenarios the best? Adjust your solution to fit Customer behaviour, use Triggers, Channels & Emotions for marketing and communication.	8.1 ONLINE CHANNELS CH What kind of actions do customers take online? Extract online channels from box #7 Behaviour	Explore AS, differentiate
	The most common triggers were unscheduled contact with physician or nurse moderate/severe pain, moderate/severe worry, anxiety, suffering, existential pain and /or psychological pain 4. EMOTIONS: BEFORE / AFTER EM How do customers feel when they face a problem or a job and afterwards? i.e. first, ensure a confident, in-control - use it in your communication strategy & design. Before: Worrying approximately your health, feeling irritating or overwhelmed, depression, worry and sadness After: Fear that hassle will come back Memory and concentration Improving reminiscence and concentration Feeling alone	Hand Hygiene Checklist Avoid abbreviations. Rapid Response System. Promote reporting. Enforce strict disinfection protocols. Use superior tracking equipment. Verify all scientific procedures. Observe care in dealing with medicines. Review staffing policies. Work with depended on providers. If you are working on an existing business, write down your current solution first, fill in the canvas, and check how much fits reality. If you are working on a new business proposition, then keep it blank until you fill in the canvas and come up with a solution that fits within customer limitations, solves a problem and matches customer behaviour.	8.2 OFFLINE CHANNELS CH What kind of actions do customers take offline? Extract offline channels from box #7 Behaviour and use them for customer development. Re-engineer health center discharges Prevent significant line-related blood movement infections Prevent venous thromboembolism	

4. REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENT

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form
FR-2	User Confirmation	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 ID number, etc
FR-4	Report Generation	The Hospital Management System generates a report on every patient regarding various information like patients name, Phone number, bed number, the doctor's name whom its assigns, ward name, and more. The Hospital Management system also helps in generating reports on the availability of the bed regarding information like bed numbers 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 patient checkout 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.

4.2 NON-FUNCTIONAL REQUIREMENTS

FR No.	Non-Functional Requirement	Description
NFR-1	Usability	The effectiveness, efficiency and satisfaction with which specific users can achieve a specific set of tasks in a particular environment.
NFR-2	Security	This process of protecting data from unauthorized access and data corruption throughout its lifecycle

NFR-3	Reliability	A highly reliable system has a lower risk of errors and process failures that can cause patients harm
NFR-4	Performance	performance measurements include: <ul style="list-style-type: none">• Quality and efficiency of patient care• Cost of healthcare services• 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 controlled conditions to be expanded under real world conditions to reach a greater proportion of the eligible population, while retaining effectiveness.

5. PROJECT DESIGN

5.1 DATA FLOW DIAGRAMS

5.2 SOLUTION AND TECHNICAL ARCHITECTURE

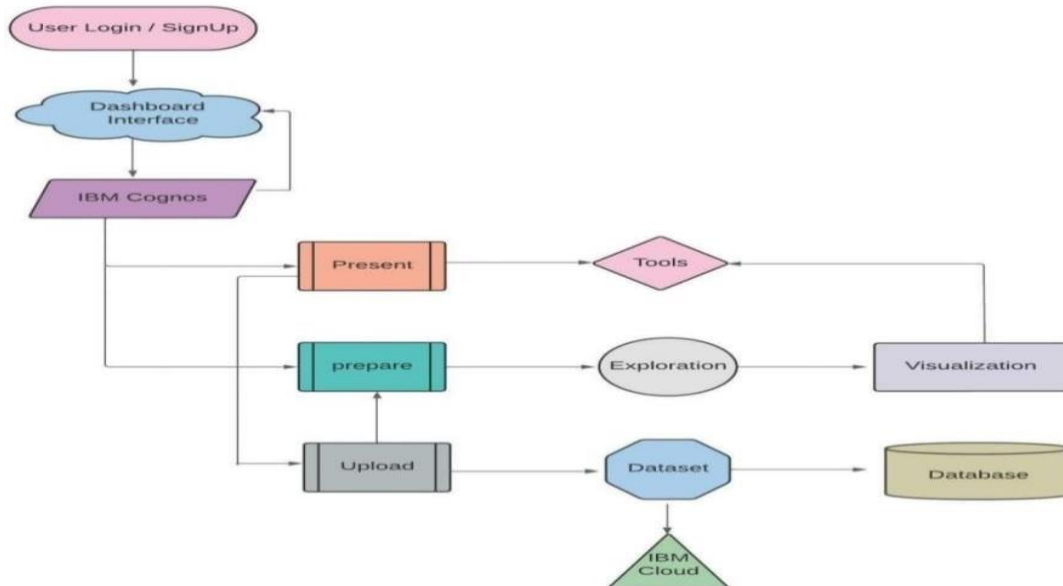


TABLE 1:COMPONENTS & TECHNOLOGIES

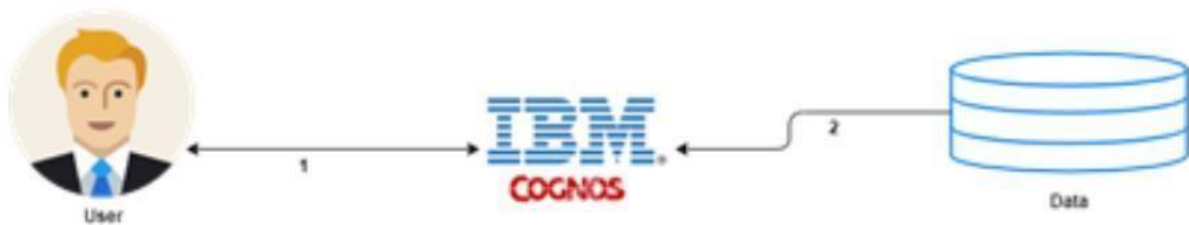
S.No	Component	Description	Technology
1.	User Interface	How user interacts with application e.g. Web UI, Mobile App, Chatbot etc.	HTML, CSS, JavaScript
2.	Application Logic-1	Logic for a process in the application	Python
3.	Application Logic-2	Logic for a process in the application	IBM Watson STT service
4.	Application Logic-3	Logic for a process in the application	IBM Watson Assistant

5.	Database	Data Type, Configurations etc.	MySQL, NoSQL, etc.
6.	Cloud Database	Database Service on Cloud	IBM DB2, IBM Cloudant etc.
7.	File Storage	File storage requirements	IBM Block Storage or Other Storage Service or Local Filesystem
8.	External API-1	Purpose of External API used in the application	IBM Weather API, etc.
9.	External API-2	Purpose of External API used in the application	Aadhar API, etc.
10.	Infrastructure (Server / Cloud)	Application Deployment on Local System / Cloud Local Server Configuration: Cloud Server Configuration :	Local, Cloud Foundry

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	List the open-source frameworks used	Technology of Opensource framework
2.	Security Implementations	List all the security / access controls implemented, use of firewalls etc.	e.g. SHA-256, Encryptions, IAM Controls, OWASP etc.
3.	Scalable Architecture	Justify the scalability of architecture (3 – tier, Micro-services)	Technology used
4.	Availability	Justify the availability of application (e.g. use of load balancers, distributed servers etc.)	Technology used
5.	Performance	Design consideration for the performance of the application (number of requests per sec, use of Cache, use of CDN's) etc.	Technology used

SOLUTION ARCHITECTURE:



5.3 USER STORIES

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Data Collection	USN-1	The User needs a complete data about the patients admitted in the hospital and a dataset should be prepared.	20	Medium	RANJANI M, VASUKI M
Sprint-1	Data Exploration	USN- 2	As a user, I need nicely visualized dashboard of number of beds occupied and number of free beds in hospital.	10	High	Sandhana k, sandhiya G
Sprint-2	Track of patient visit of Hospital	USN-3	Tracking a patient Health care over years of visit and Screening of data they have in hospital.	10	Medium	Sandhiya G, Ranjani M

Sprint -2	Dashboard	USN - 4	As a user, I want the interactive dashboard to analyse the data. Have the data in terms of Graph.	20	High	Ranjani M, vasuki M, Sandhiya G
Sprint-3	Detailed EHR's of patient	USN-5	Provided greater details in the EHR's of individual patient with clear idea of what to do.	10	Medium	vasuki M, Sandhiya G
Sprint- 3	Story Creation	USN-6	As a user, I need the story animation of the data set with insights	20	High	Sandhana k, sandhiya G
Sprint-4	Predict LOS	USN-7	As a user, I want the flawless system to predict the length of stay of the patients	10	High	Ranjani M, Vasuki M
Sprint-4	Using ML algorithm for Prediction	USN-8	As a user, I need prior knowledge of LOS can aid in logistics such as room and bed allocation planning.	20	High	Ranjani M, vasuki M, Sandhiya G, Sandhana K

6. PROJECT PLANNING AND SCHEDULING

6.1 SPRINT PLANNING AND ESTIMATION

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

Velocity:

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

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

6.2 SPRINT DELIVERY SCHEDULE

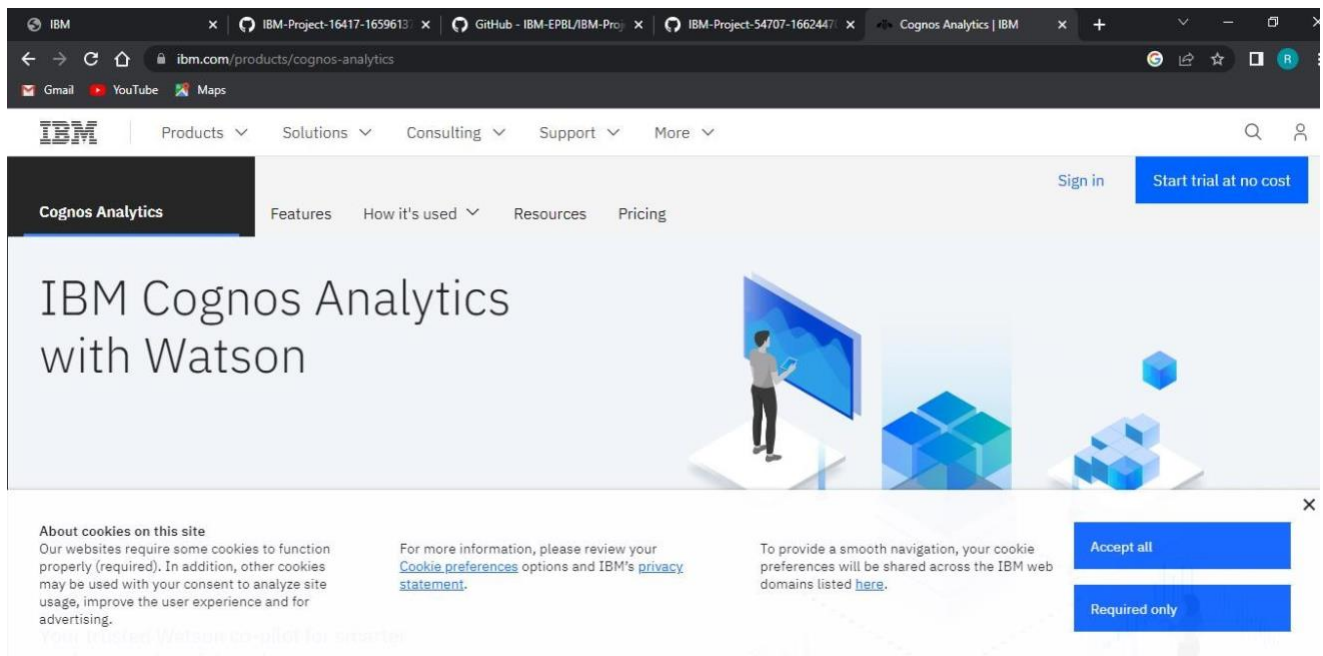
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Data Collection	USN-1	The User needs a complete data about the patients admitted in the hospital and a dataset should be prepared.	20	Medium	RANJANI M, VASUKI M
Sprint-1	Data Exploration	USN- 2	As a user, I need nicely visualized dashboard of number of beds occupied and number of free beds in hospital.	10	High	Sandhana k, sandhiya G
Sprint-2	Track of patient visit of Hospital	USN-3	Tracking a patient Health care over years of visit and Screening of data they have in hospital.	10	Medium	Sandhiya G, Ranjani M

Sprint -2	Dashboard	USN - 4	As a user, I want the interactive dashboard to analyse the data. Have the data in terms of Graph.	20	High	Ranjani M, vasuki M, Sandhiya G
Sprint-3	Detailed EHR's of patient	USN-5	Provided greater details in the EHR's of individual patient with clear idea of what to do.	10	Medium	vasuki M, Sandhiya G
Sprint- 3	Story Creation	USN-6	As a user, I need the story animation of the data set with insights	20	High	Sandhana k, sandhiya G
Sprint-4	Predict LOS	USN-7	As a user, I want the flawless system to predict the length of stay of the patients	10	High	Ranjani M, Vasuki M
Sprint-4	Using ML algorithm for Prediction	USN-8	As a user, I need prior knowledge of LOS can aid in logistics such as room and bed allocation planning.	20	High	Ranjani M, vasuki M, Sandhiya G, Sandhana K

7. SOLUTIONING

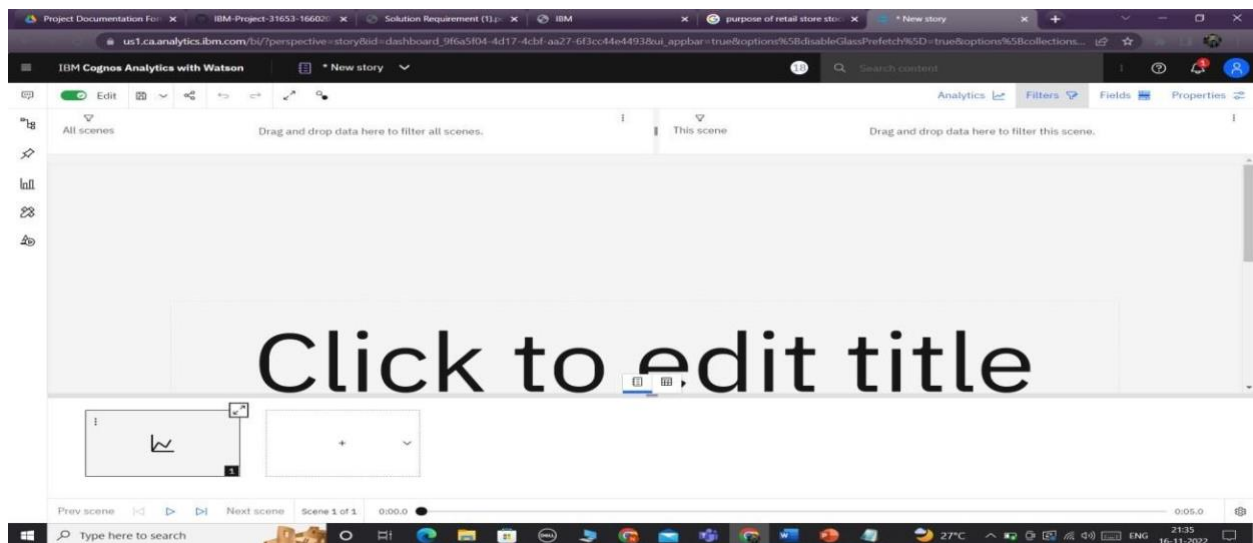
7.1 FEATURE 1

DASHBOARD AND REPORT:



7.2 FEATURE 2

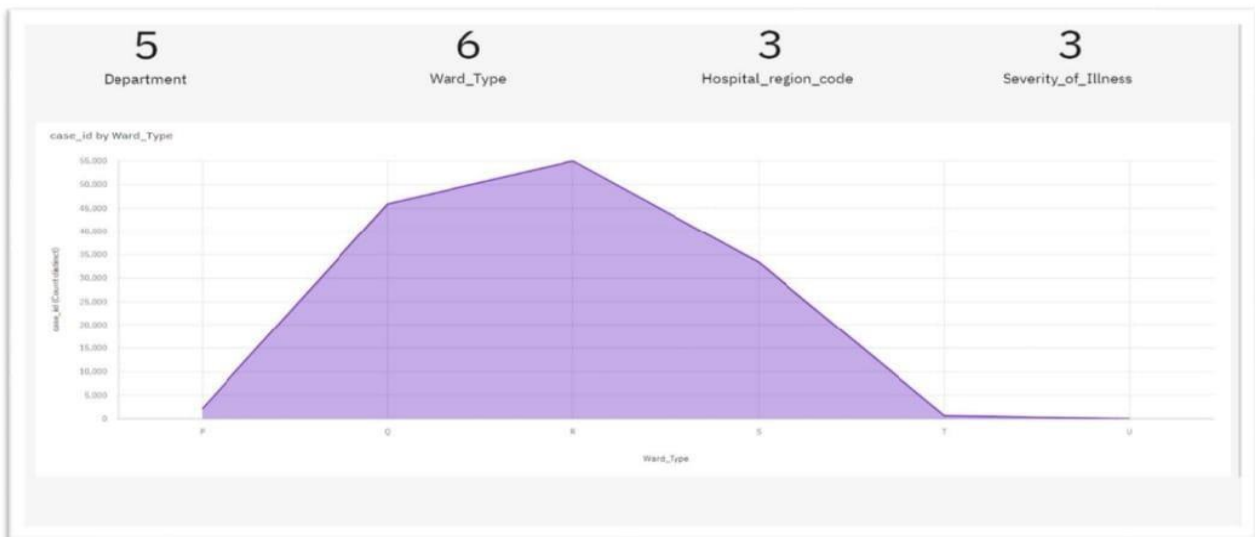
STORY OF THE DATASET:



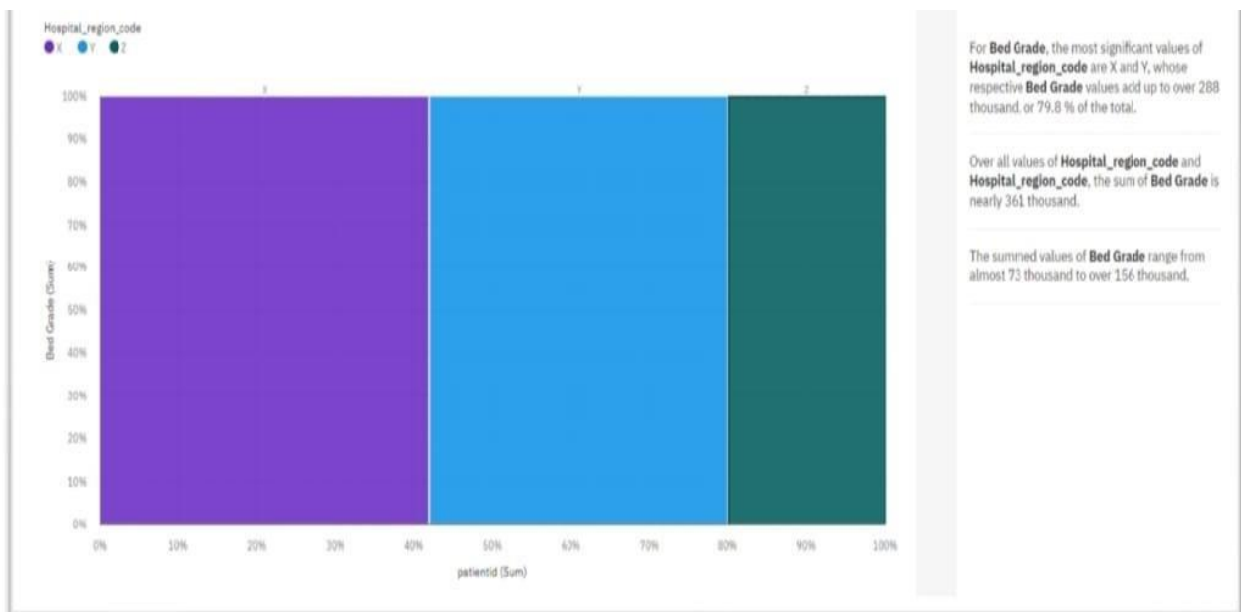
8. TESTING

8.1 TEST CASES

1.EXPLORATION OF DATA



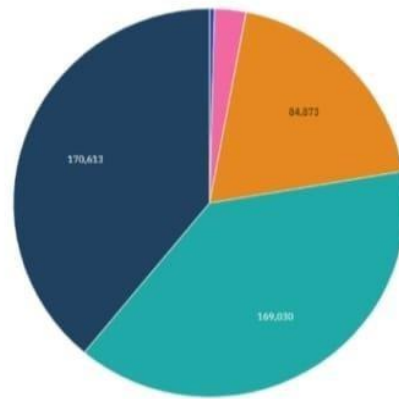
2. Bed Grade and Patient for Hospital_region_code coloured by Hospital_region_code



3. Available Extra Rooms in Hospital by Ward_Type

Ward_Type

U T P S R Q



Over all values of **Ward_Type**, the sum of **Available Extra Rooms in Hospital** is nearly 438 thousand.

For **Available Extra Rooms in Hospital**, the most significant values of **Ward_Type** are Q, R, and S, whose respective **Available Extra Rooms in Hospital** values add up to nearly 425 thousand, or 97 % of the total.

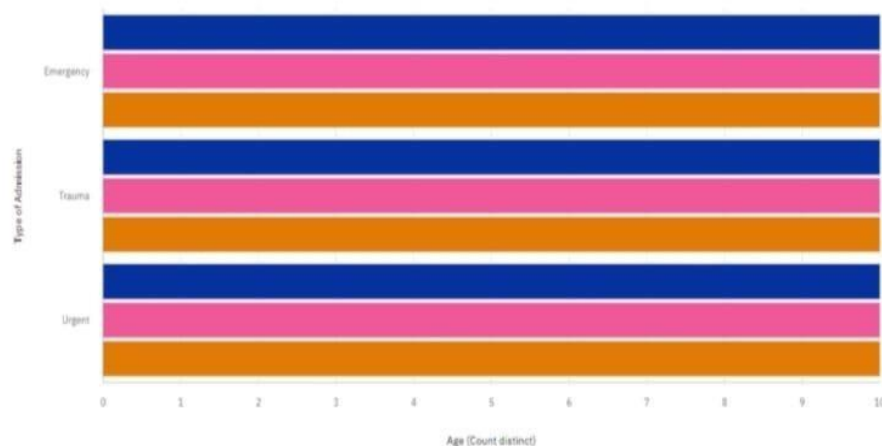
Available Extra Rooms in Hospital ranges from 7, when **Ward_Type** is U, to almost 171 thousand, when **Ward_Type** is Q.

Available Extra Rooms in Hospital is unusually high when **Ward_Type** is Q and R.

4. Department coloured by Age sized by Bed Grade

Severity of Illness

Extreme Minor Moderate

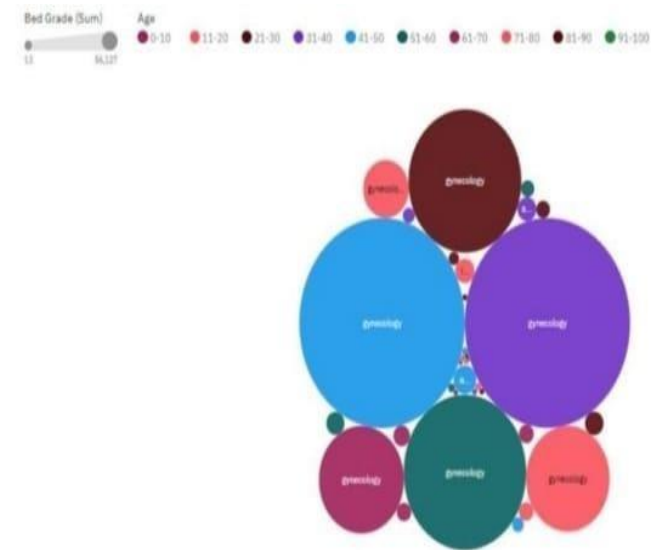


The total number of results for **Age** is 137,057.

The most common values of **Type of Admission** are Trauma (47.7 %) and Emergency (37 %), together occurring over 116 thousand times, which is 84.7 % of the total.

The most common value of **Severity of Illness** is Moderate, occurring nearly 76 thousand times, which is 55.2 % of the total.

5. Age by Type Admission coloured by Severity of Illness



Bed Grade is almost 361 thousand.

For **Bed Grade**, the most significant value of **Department** is gynecology, whose respective **Bed Grade** values add up to over 277 thousand, or 76.8 % of the total.

For **Bed Grade**, the most significant values of **Age** are 41-50, 31-40, 51-60, 21-30, and 71-80, whose respective **Bed Grade** values add up to almost 282 thousand, or 78 % of the total.

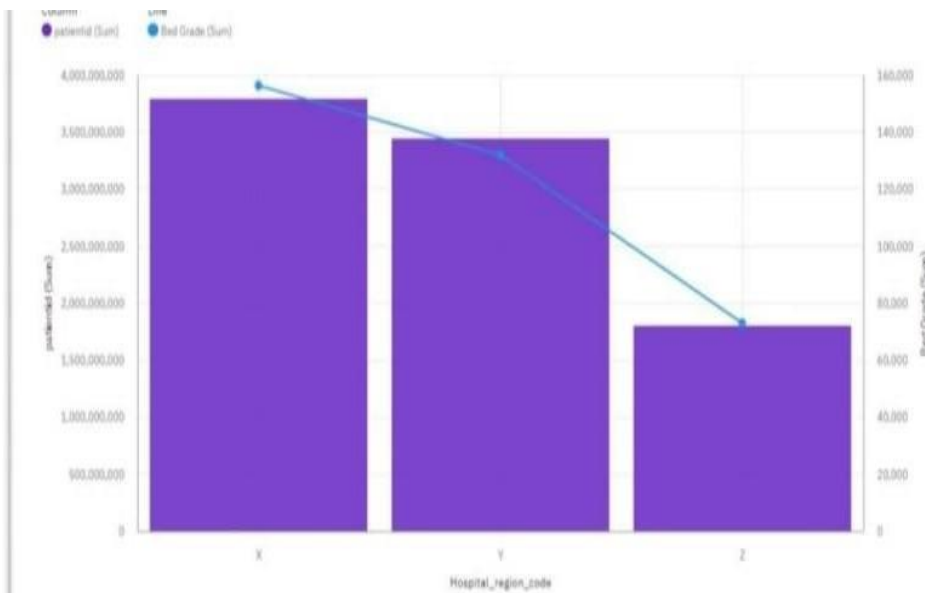
The summed values of **Bed Grade** range from 13 to over 56 thousand.

Bed Grade is unusually high when the combinations of **Department** and **Age** are gynecology and 41-50 and gynecology and 31-40.

Bed Grade is unusually high when **Department** is gynecology.

Bed Grade is most unusual when **Age** is 91-100, 41-50 and 31-40.

6. Bed Grade and Patient by Hospital_region_code



the sum of **patientid** is over 9.0 billion.

For **patientid**, the most significant values of **Hospital_region_code** are X and Y, whose respective **patientid** values add up to over 7.2 billion, or 80.1 % of the total.

patientid ranges from nearly 1.8 billion, when **Hospital_region_code** is Z, to nearly 3.8 billion, when **Hospital_region_code** is X.

patientid is unusually low when **Hospital_region_code** is Z.

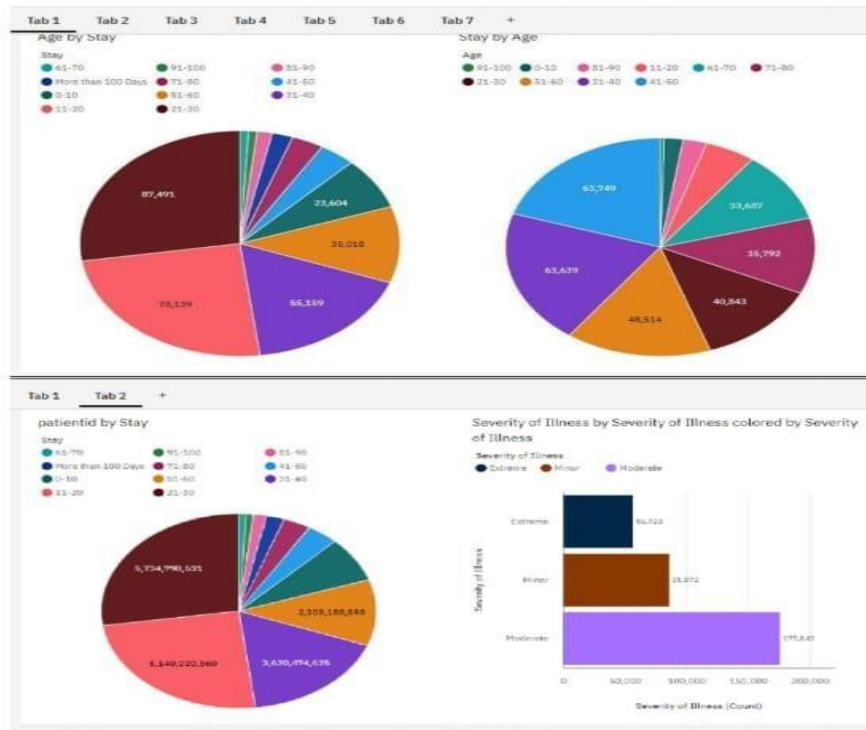
Over all values of **Hospital_region_code**, the sum of **Bed Grade** is nearly 361 thousand.

For **Bed Grade**, the most significant values of **Hospital_region_code** are X and Y, whose respective **Bed Grade** values add up to over 288 thousand, or 79.8 % of the total.

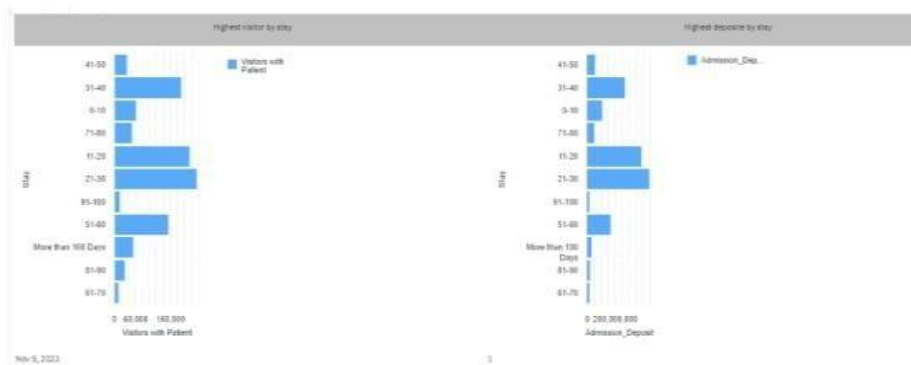
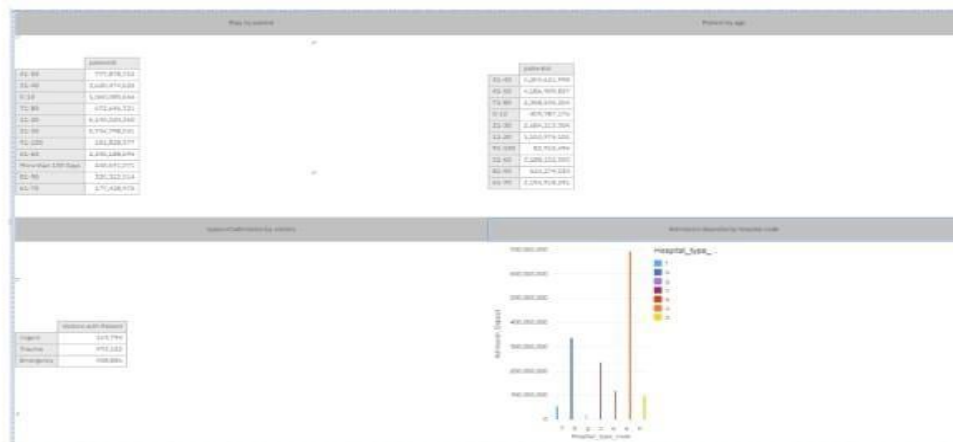
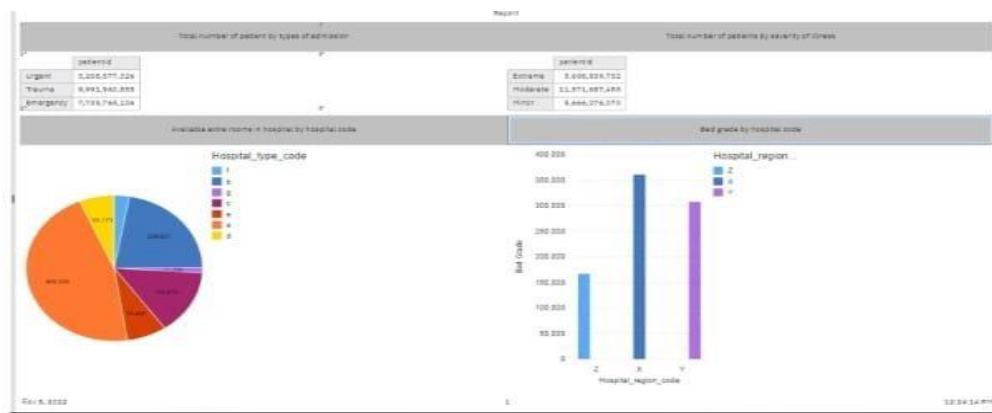
Bed Grade ranges from almost 73 thousand, when **Hospital_region_code** is Z, to over 156 thousand, when **Hospital_region_code** is X.

9. RESULTS

9.1 PERFORMANCE METRICS DASHBOARD:



REPORT:



STORY:



10. ADVANTAGES:

Cost Reduction ☐ Improves Data Security ☐ Less paperwork ☐ Better Collaboration and Communication ☐
Error-free administration ☐ Improved Clinical decision -making

DISADVANTAGES:

- Cybersecurity Risks.
- Patient Isolation .
- Frustration with poor Implementation.

11. CONCLUSION

In hospital health care management , we need a healthy lifestyle to build up a healthy immune system and to avoid diseases.

12. FUTURE SCOPE

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.

13. APPENDIX

Github project link - <https://github.com/IBM-EPBL/IBM-Project-31813-1660205338>