ANALYTICS FOR HOSPITALS' HEALTH-CARE DATA

PROJECT DOCUMENTATION

TEAM MEMBERS:

P. DIVYA (TEAM LEAD)

B. DEEKSHITHA

B. DEEPTHA

PREETHI BASKARAN

CHAPTER 1 INTRODUCTION

This project deals with the analytics for hospital's health care data using data analytics. Data analytics (DA) is the process of examining data sets in order to find trends and draw conclusions about the information they contain. Increasingly, data analytics is done with the aid of specialized systems and software. Data analytics technologies and techniques are widely used in commercial industries to enable organizations to make more-informed business decisions.

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:

Data analytics in health care is vital. It helps health care organizations to evaluate and develop practitioners, detect anomalies in scans and predict

outbreaks in illness, per the Harvard Business School. Data analytics can also lower costs for health care organizations and boost business intelligence. Hospital data analytics can look over patient data and any prescribed medication to alert doctors and patients of incorrect dosages or wrong prescriptions, which lessens human error and the cost to your hospital. This in turn helps in gaining better insights and also enables healthcare practitioners to make wellinformed decisions.

CHAPTER 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 utilized 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 techno-world, the term data is unavoidable and certainly, nothing is possible without its usage. The trends about how to analyze 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

> The already existing model is trained with minimal parameters

> Low accuracy in prediction

> No feature extraction done

> High complexity

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

[6]. Yang J.-J., Li J., Mulder J., Wang Y., Chen S., Wu H., Wang Q., Pan H. Emerginginformation technologies for enhanced healthcare.Comput.ind.2015;69:3-
11.doi:10.1016/j.compimd.2015.01.012. [CrossRef] [Google Scholar]

[7]. Cortada J.W., Gordon D., Lenihan B. The Value of Analytics in Healthcare. IBM Institute for Business Value; Armonk, NY, USA: 2012. Report No.: GBE03476- USEN-
00. [Google Scholar].

[8]. Makary M.A., Daniel M. Medical error-the third leading cause of death in the US. Br. Med. J. 2016;353:i2139. doi: 10.1136/bmj.i2139. [PubMed] [CrossRef] [Google Scholar].

[9]. Prokosch H.-U., Ganslandt T. Perspectives for medical informatics. Methods Inf. Med. 2009;48:38–44. doi: 10.3414/ME9132. [PubMed] [CrossRef] [Google Scholar].

[10]. Simpao A.F., Ahumada L.M., Gálvez J.A., Rehman M.A. A review of analytics andclinical informatics in health care. J. Med. Syst. 2014;38:45. doi: 10.1007/s10916- 014-0045-x. [PubMed] [CrossRef] [Google Scholar].

[11]. Ghassemi M., Celi L.A., Stone D.J. State of the art review: The data revolution incritical care. Crit. Care. 2015;19:118. doi: 10.1186/s13054-015-0801-4. [PMC free article] [PubMed] [CrossRef] [Google Scholar].

[12]. Tomar D., Agarwal S. A survey on Data Mining approaches for Healthcare. Int. J. Bio-Sci. Bio-Technol. 2013;5:241–266. doi: 10.14257/ijbsbt.2013.5.5.25. [CrossRef] [Google Scholar]

[13]. K. Jee and G. H. Kim, "Potentiality of big data in the medical sector: Focus on how to reshape the healthcare system," Healthc. Inform. Res., vol. 19, no. 2, pp. 79–85, Jun. 2013. doi: 10.4258/hir.2013.19.2.79.

[14]. J. King, V. Patel, and M. F. Furukawa, "Physician adoption of electronic health record technology to meet meaningful use objectives: 2009–2012," The Office of the National Coordinator for Health Information Technology, Tech. Rep., Dec. 2012.

[15]. V. Mayer-Schönberger and K. Cukier, Big Data: A Revolution That Will Transform How We Live, Work, and Think. Eamon Dolan, 2014.

[16]. J. Rapoport, D. Teres, Y. Zhao, S. Lemeshow Length of stay data as a guide to hospital economic performance for icu patients Med Care, 41 (3) (2003), pp. 386-397

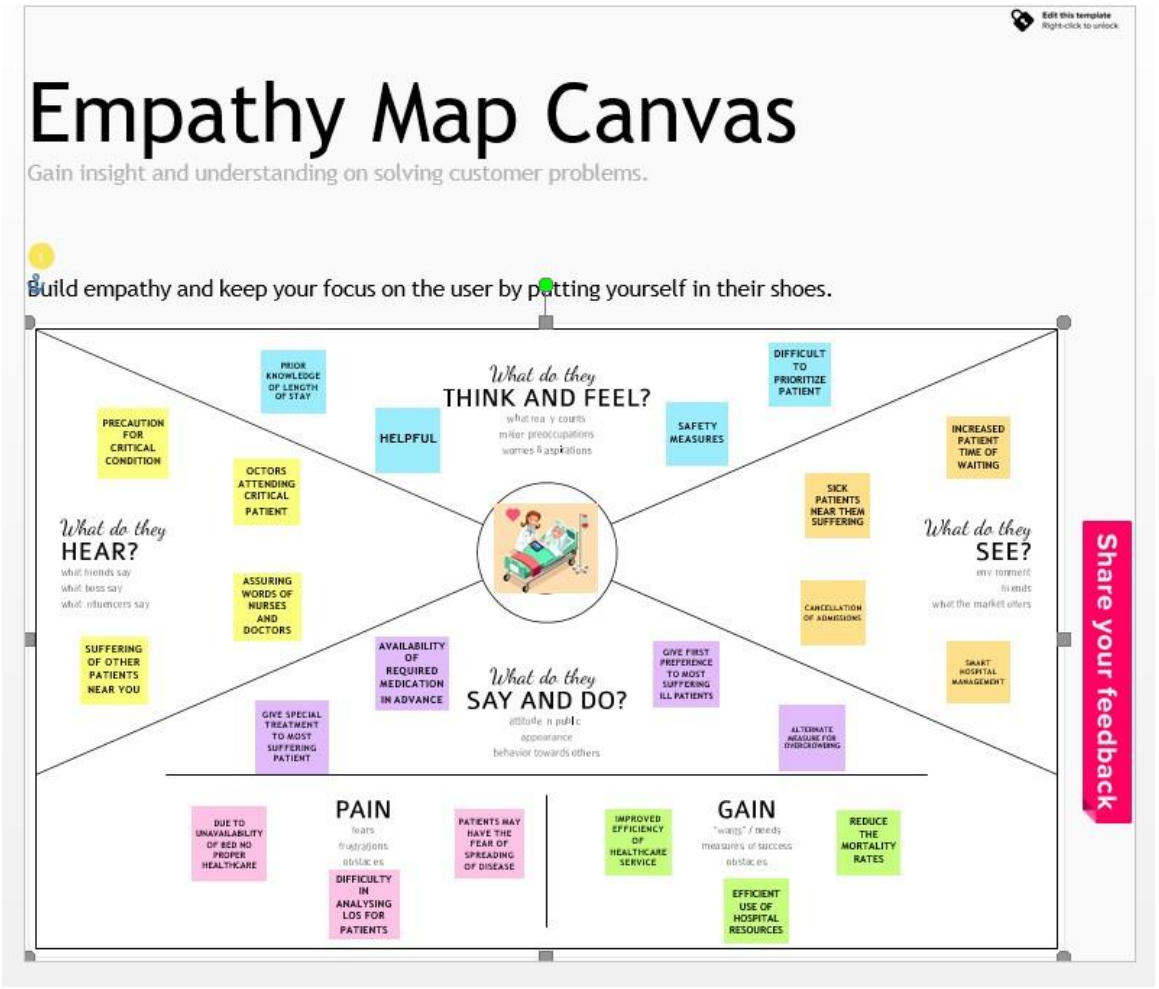
2.3 PROBLEM STATEMENT AND DEFINITION

- The aim 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.

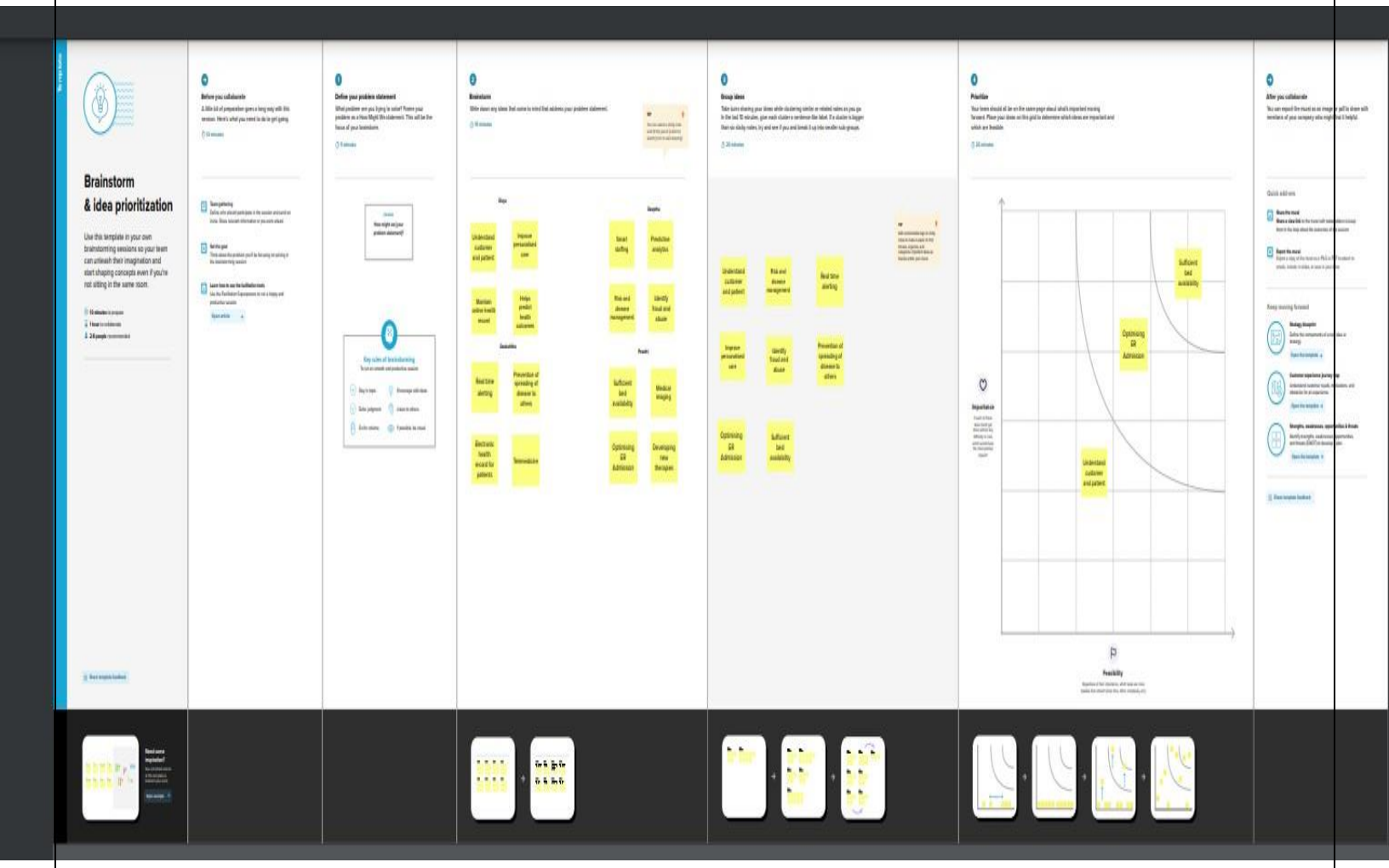
CHAPTER 3

IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAP CAMPUS



3.2 IDEATION & BRAINSTORMING



3.3PROPOSED SOLUTION

Predict the length of stay of patients.

The length of the stay can be predicted using either Random forest or Decision

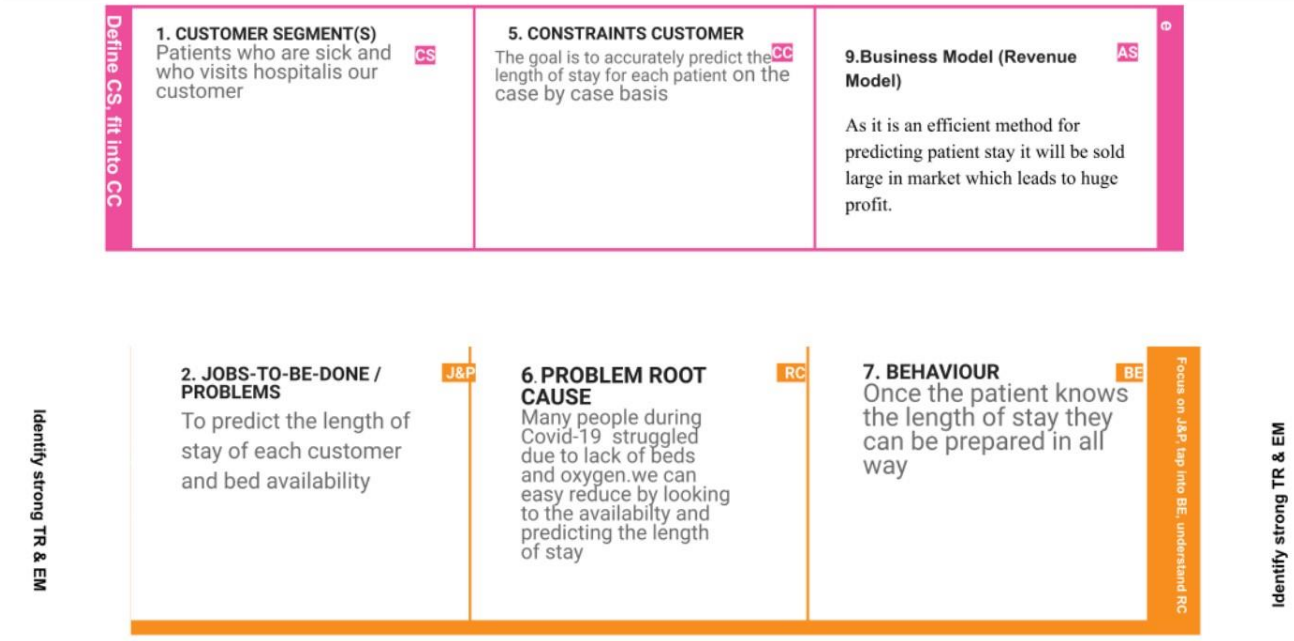
Tree for more accuracy.

Certain parameters like age, stage of the
diseases, disease diagnosis, severity of illness, type of admission, facilities allocated, etc., are used for prediction.

IBM Cognos will be used for data analytics.

The model will be trained using colab.It predicts the length of stay (LOS) of the

3.3 PROBLEM SOLUTION FIT



CHAPTER 4

Solution Requirements (Functional & Non-functional)

Date	16th October 2022
Team ID	PNT2022TMID44120
Project Name	Analytics for Hospital health data
Maximum Marks	4 Marks

Functional Requirements:

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	This Dashboards are designed to offer a Comprehensive overview of patient is 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 effective, efficient and reliable manner.

4.2 NON FUNCTIONAL REQUIREMENT

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.

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
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.

Non-functional Requirements:

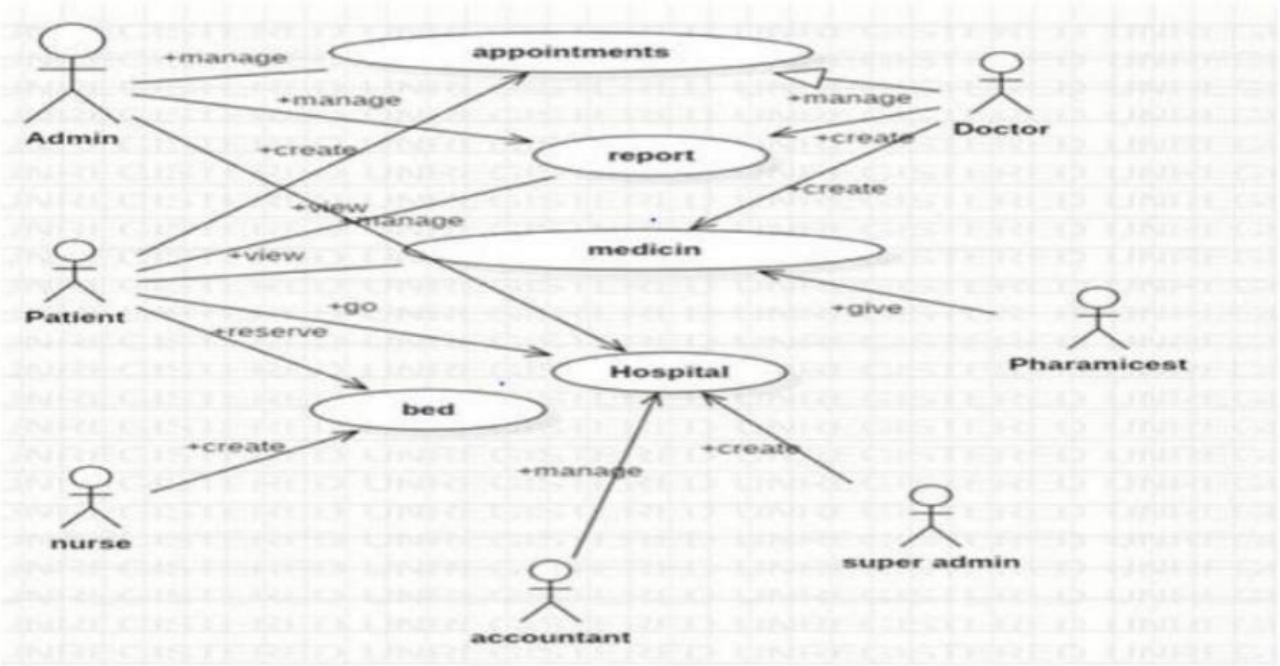
Following are the non-functional requirements of the proposed solution.

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.

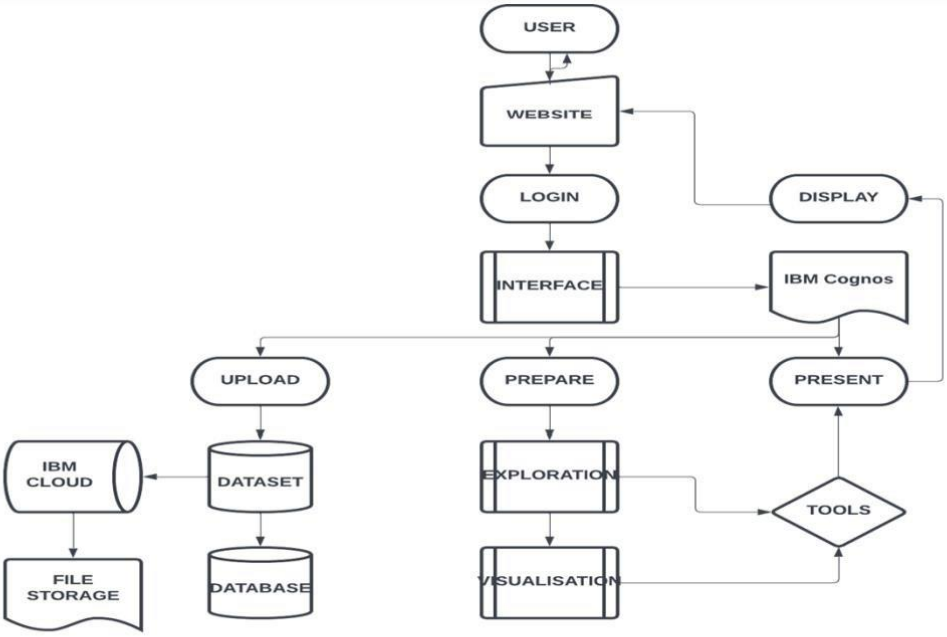
CHAPTER 5

PROJECT DESIGN

5.1 DATA FLOW DIAGRAMS



5.2 SOLUTION & TECHNICAL ARCHITECTURE



5.3 USER STORIES :

User type	Functional requirement (epic)	User story number	User story/task	Acceptance criteria	Priority	Release
Customer	Registrati on	USN-1	As a user i an login to my dashboard	I can access dashboard	High	Sprint-1

	Collect data	USN-1	As a user i can provide my details	I can view my data	Medium	Sprint-1
Admin	Collect data	USN-2	As an analyst i collect the data		High	Sprint-2
	Analyze	USN-2	As an analyst i analyze the given dataset	I can analyze the dataset	High	Sprint-2
	Upload data	USN-3	As an analyst i can upload datasets	I can upload the dataset	Medium	Sprint-3
	Prediction	USN-6	As an analyst i will predict the length of stay of patient	I can predict the length of stay	High	Sprint-4
Visualization	Prepare data	USN-4	As an admin i prepare the data for visualization	I can prepare the data with visualization techniques.	High	Spint-3
	Dashboard	USN-5	As an admin i present the data that is visualized	I can present the result	High	Sprint-4

CHAPTER 6
PROJECT PLANNING & SCHEDULING

6.1 SPRINT PLANNING & ESTIMATION

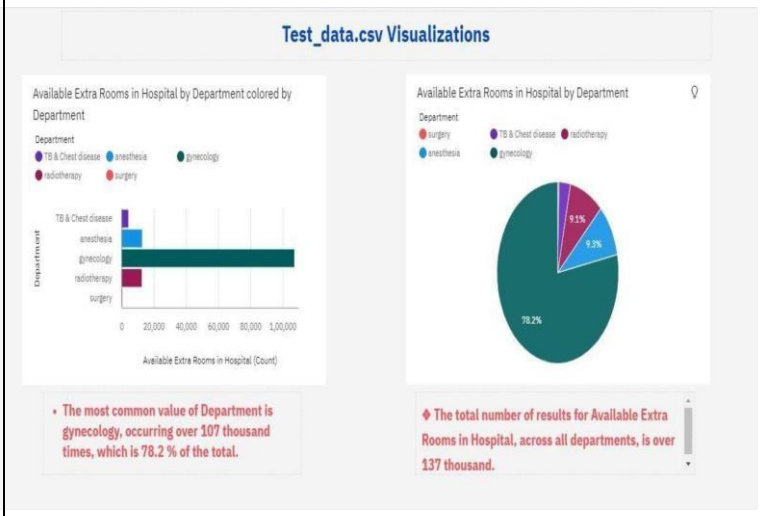
Sprint	Functional Requirement (Epic)	User Story Number	User Story/ Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a health care provider I can create account inIBM cloud and the data are collected.	20	High	Divya.P Deekshitha.B
sprint-2	Analyze	USN-2	As a healthcare provider all data are collected is cleaned and uploaded in the database or IBM cloud	20	medium	Deeptha.B Preethi Baskaran

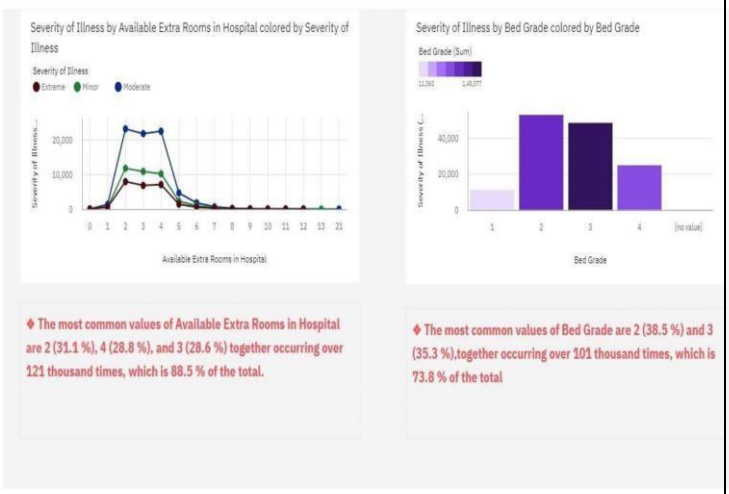
6.2 SPRINT DELIVERY SCHEDULE

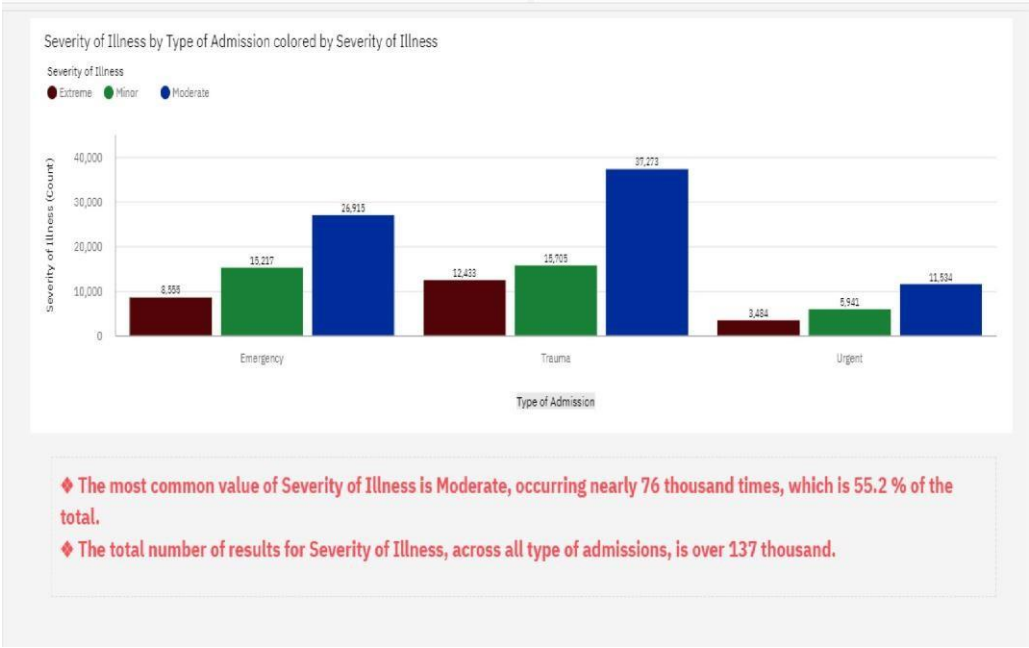
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-2	Data Analysis	USN-3	As a user, I will be performing analysis on the data for making predictions		High	B.Deekshitha P.Divya
Sprint-2	Dashboards	USN-4	As a user, I will be making visualizations and interactive dashboards from the data	10	High	B.Deeptha, Preethi Baskaran
Sprint-3	Story	USN-5	As a user, I will be making stories from the data and the dashboards	20	High	P.Divya Preethi Baskaran
Sprint-4	Report	USN-6	As a user, I will be making a report from the analysis and dashboards	20	High	B.Deeptha B.Deekshitha

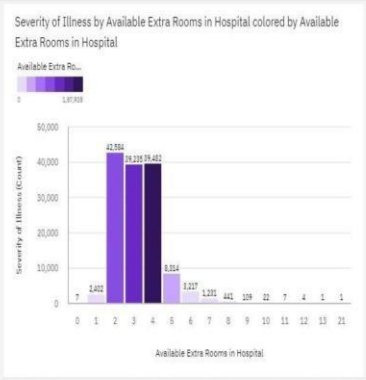
CHAPTER 7
CODING & SOLUTIONING

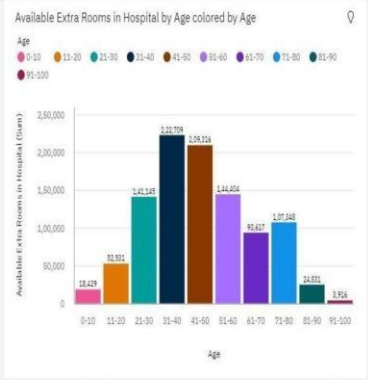
7.1 FEATURE 1


Test_data.csv Visualizations
Available Extra Rooms in Hospital by Department colored by Department
Department
TB & Chest disease
anesthesia
gynecology
radiotherapy
surgery
Available Extra Rooms in Hospital (Count)
Available Extra Rooms in Hospital by Department
9.1%
9.3%
78.2%
• The most common value of Department is gynecology, occurring over 107 thousand times, which is 78.2 % of the total.
♦ The total number of results for Available Extra Rooms in Hospital, across all departments, is over 137 thousand.


Severity of Illness by Available Extra Rooms in Hospital colored by Severity of Illness
Severity of Illness
Extreme
Minor
Moderate
Available Extra Rooms in Hospital
♦ The most common values of Available Extra Rooms in Hospital are 2 (31.1 %), 4 (28.8 %), and 3 (28.6 %) together occurring over 121 thousand times, which is 88.5 % of the total.
Severity of Illness by Bed Grade colored by Bed Grade
Bed Grade (Sum)
Bed Grade
♦ The most common values of Bed Grade are 2 (38.5 %) and 3 (35.3 %),together occurring over 101 thousand times, which is 73.8 % of the total

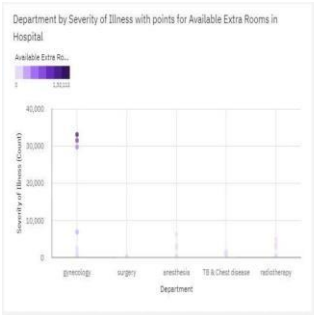

Severity of Illness by Type of Admission colored by Severity of Illness
Severity of Illness
Extreme
Minor
Moderate
Severity of Illness (Count)
8,555
15,217
26,915
12,433
16,705
37,273
3,484
5,941
11,534
Emergency
Trauma
Urgent
Type of Admission
♦ The most common value of Severity of Illness is Moderate, occurring nearly 76 thousand times, which is 55.2 % of the total.
♦ The total number of results for Severity of Illness, across all type of admissions, is over 137 thousand.

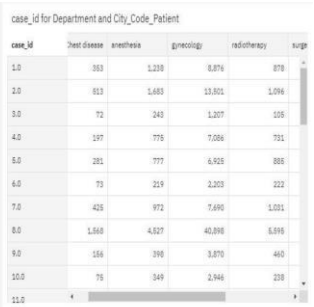




♦ The most common values of Available Extra Rooms in Hospital are 2 (31.1 %), 4 (28.8 %), and 3 (28.6 %), together occurring over 121 thousand times, which is 88.5 % of the total.

♦ Available Extra Rooms in Hospital is most unusual when the combinations of Age and Age are 31-40 and 31-40, 41-50 and 41-50 and 91-100 and 91-100.

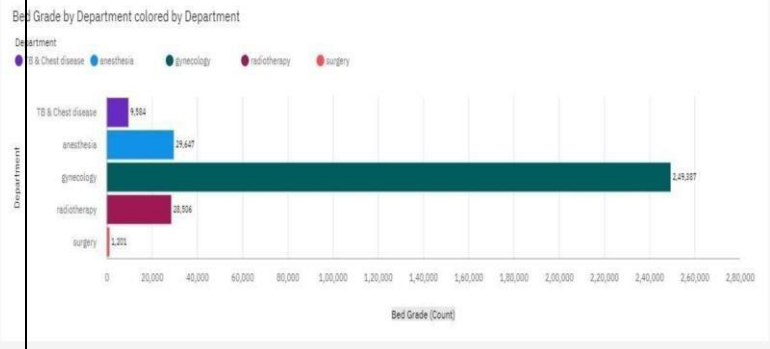




♦ The most common value of Department is gynecology, occurring over 107 thousand times, which is 78.2 % of the total.

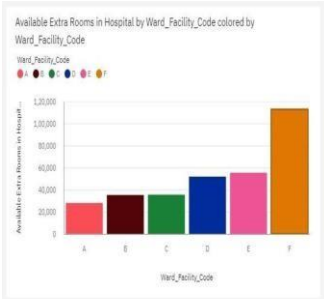
♦ The most common values of City_Code_Patient are 23.0 (3.2 %), 20.0 (3.2 %), 16.0 (3.2 %), 15.0 (3.2 %), and 14.0 (3.2 %), together occurring 25 times, which is 16.2 % of the total.

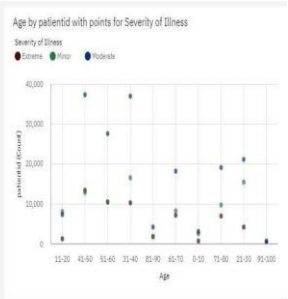
Train_data.csv Visualizations



♦ The total number of results for Bed Grade, across all departments, is over 318 thousand.

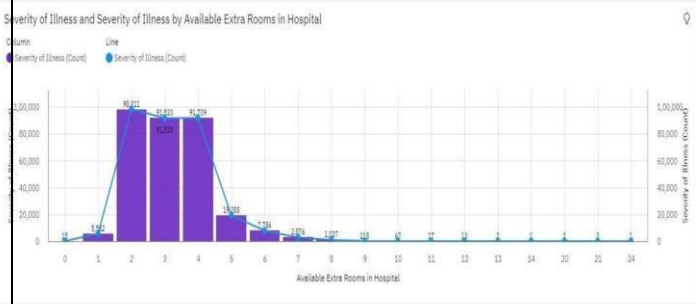
♦ The most common value of Department is gynecology, occurring over 249 thousand times, which is 78.3 % of the total





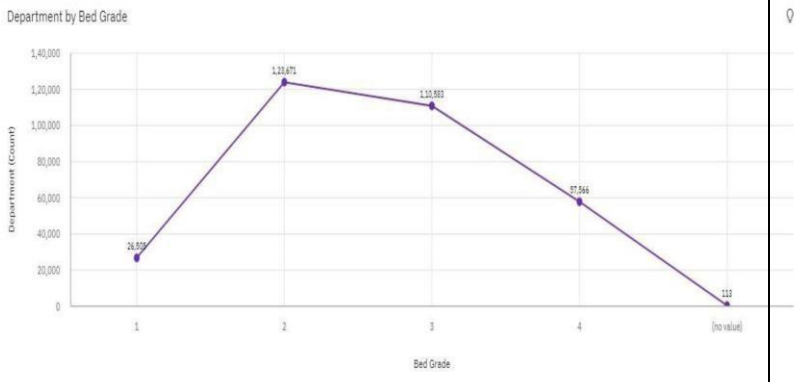
♦ The most common value of Ward_Facility_Code is F, occurring nearly 113 thousand times, which is 35.4 % of the total.

♦ The most common values of Age are 41-50 (20 %) and 31-40 (20 %), together occurring over 127 thousand times, which is 40 % of the total.



♦ The total number of results for Severity of Illness, across all available extra rooms in hospitals, is over 318 thousand.

♦ The most common values of Available Extra Rooms in Hospital are 2 (30.9 %), 4 (28.8 %), and 3 (28.7 %), together occurring nearly 282 thousand times, which is 88.4 % of the total.

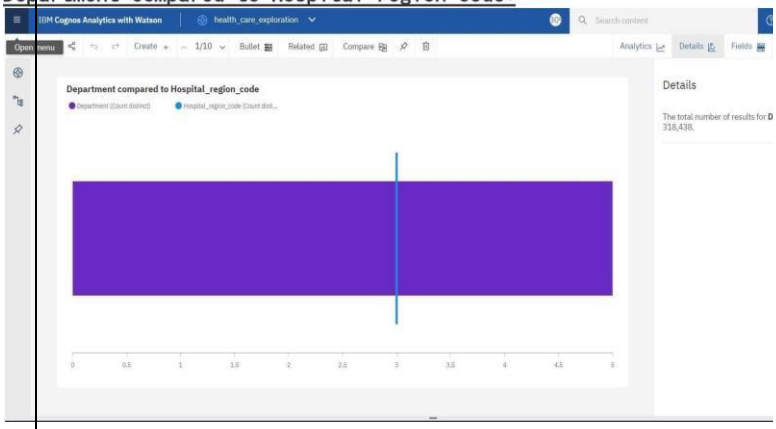


♦ The most common values of Bed Grade are 2 (38.8 %) and 3 (34.7 %), together occurring over 234 thousand times, which is 73.6 % of the total.

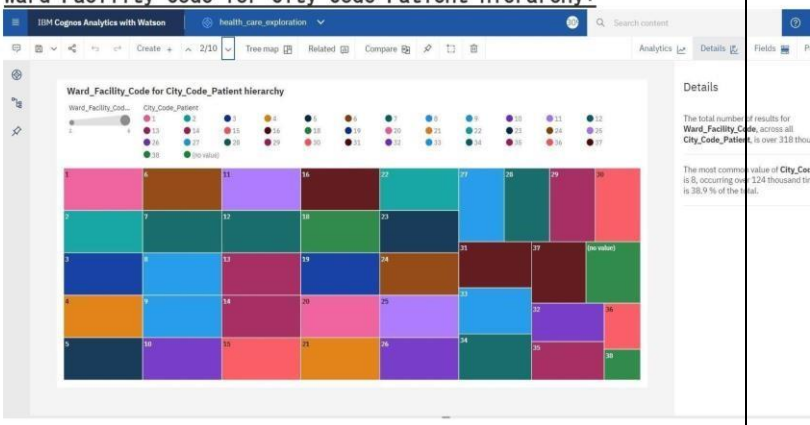
♦ The total number of results for Department, across all bed grades, is over 318 thousand.

7.2 FEATURE 2

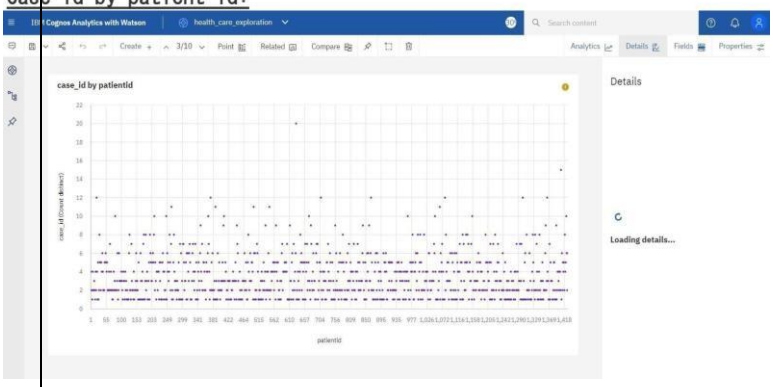
Department compared to Hospital region code:



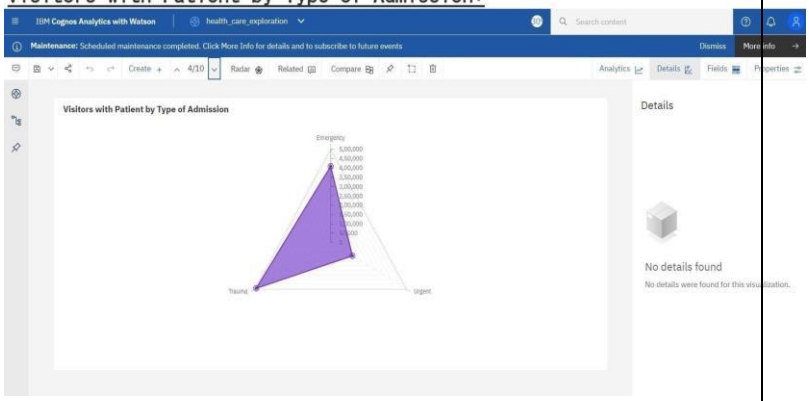
Ward Facility Code for City Code Patient hierarchy:



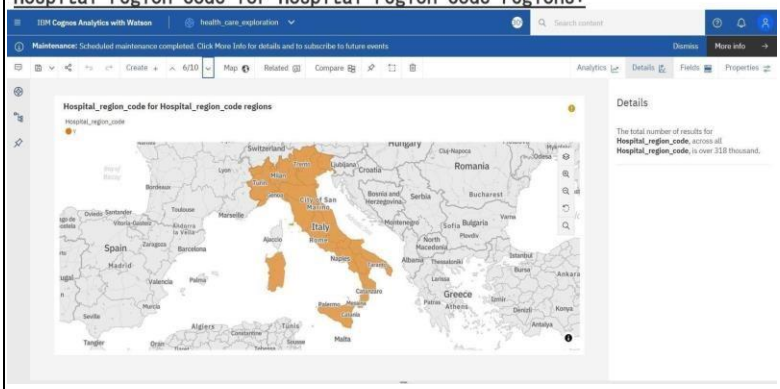
Case id by patient id:



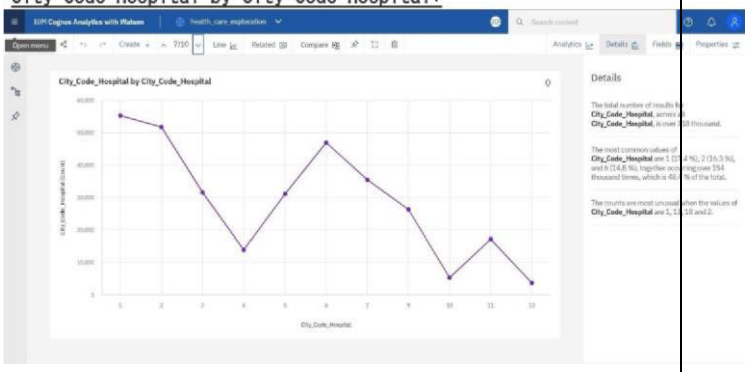
Visitors with Patient by Type of Admission:



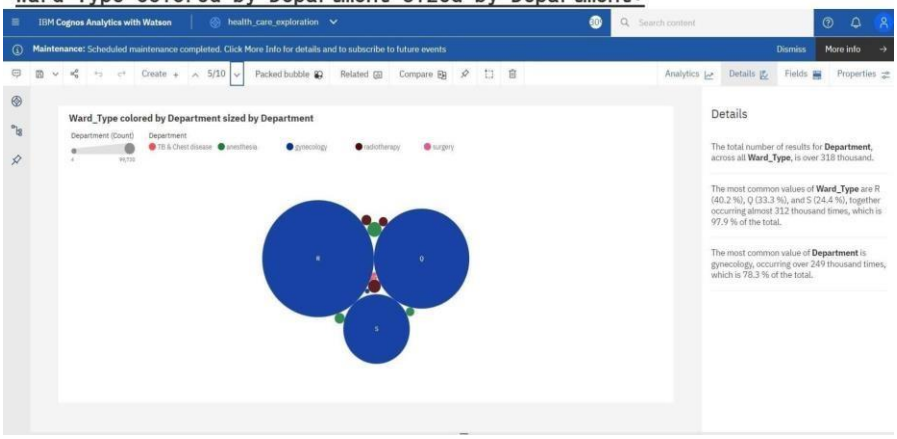
Hospital region code for Hospital region code regions:



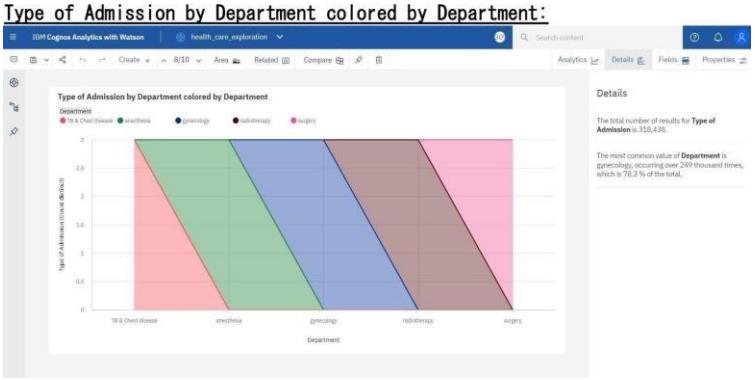
City Code Hospital by City Code Hospital:



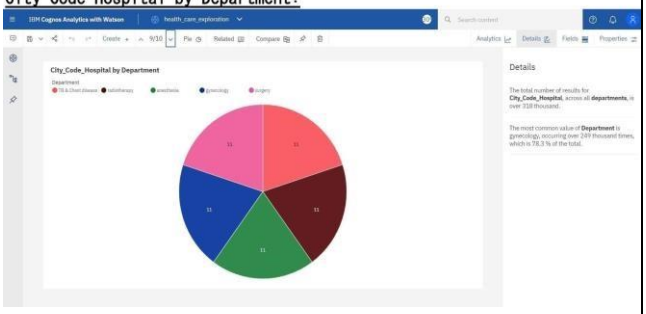
Ward Type colored by Department sized by Department:

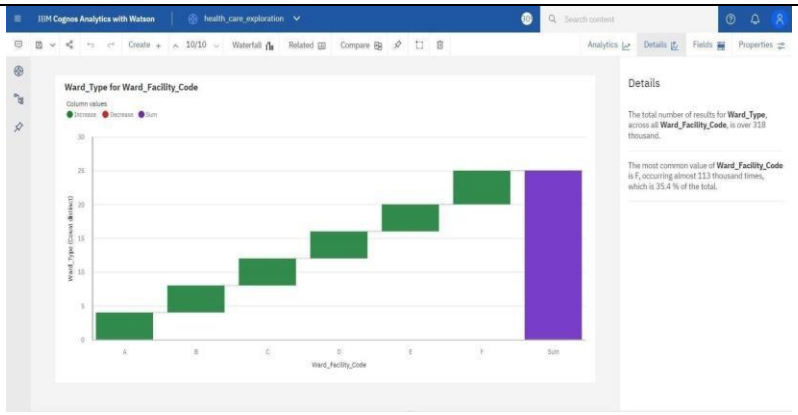


Type of Admission by Department colored by Department:



City Code Hospital by Department:





CHAPTER 8

TESTING

8.1 TEST CASES

- verify user is able to see home page
- verify user is able to see dashboard page
- verify user is able to naivigate to story page
- verify filters are working

8.1 USER 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

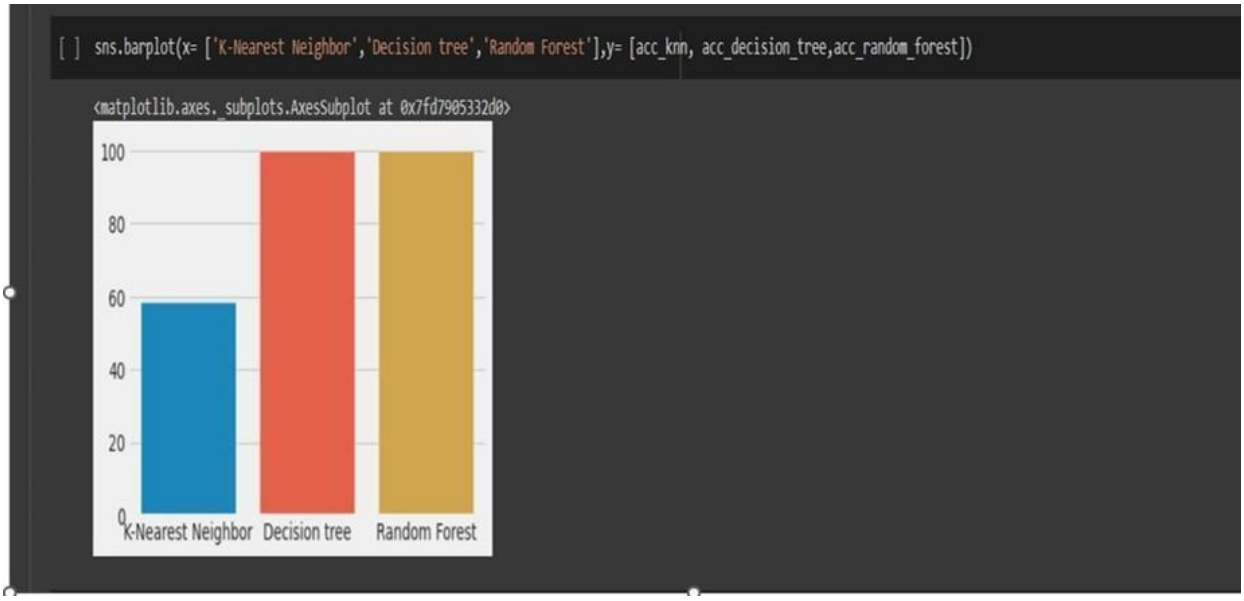
This report shows the number of test cases that have passed, failed, and untested

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

CHAPTER 9

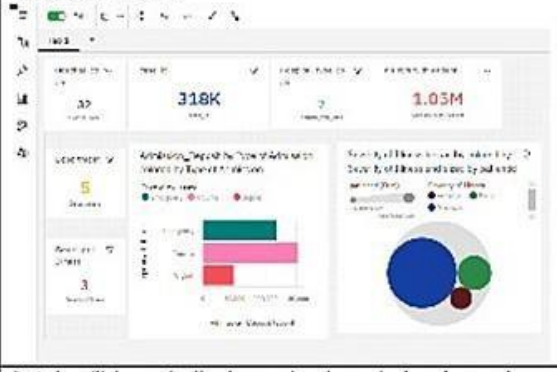
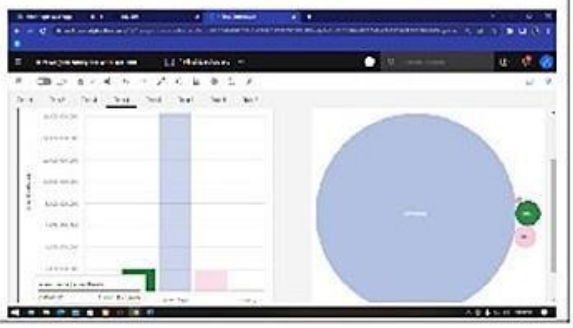
RESULTS

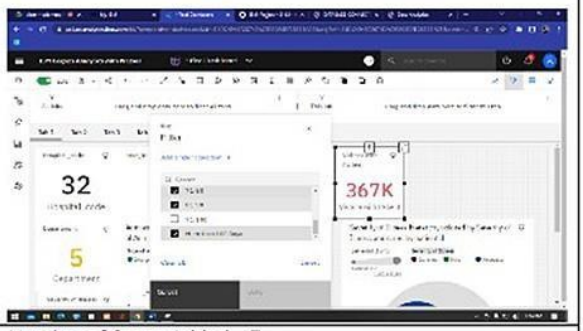
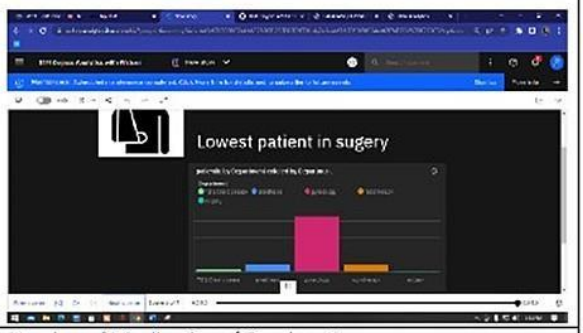
9.1 PERFORMANCE METRICS

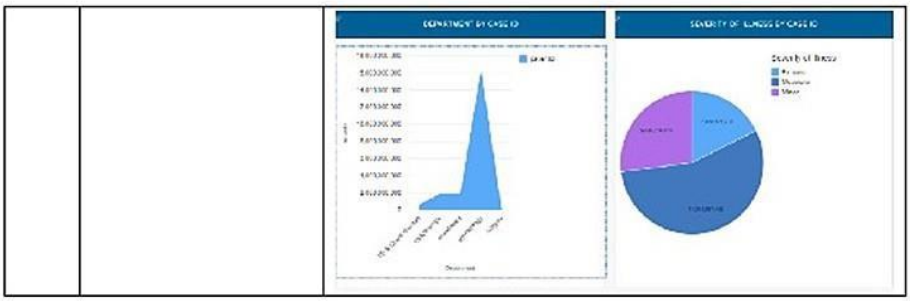


Model Performance Testing:

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 Number of tabs – 5 
2.	Data Responsiveness	Data's will dynamically changed and graph also changed. 
3.	Amount Data to Rendered (DB2 Metrics)	Number of rows read – 318438 Number of rows loaded – 318438 Number of rows rejected – 0 

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



CHAPTER 10

ADVANTAGES AND DISADVANTAGES

ADVANTAGES:

> Cost-effective use of technology

> Improved project management

> Sustaining the improvements in the result

> Boosting hospital capacity

> Enhance the quality and efficiency of healthcare

> benefit areas like emergency preparation, charting, administration, compliance, and financial management.

> Analysing clinical data to improve medical research

> Using patient data to improve health outcomes

> Gaining operational insights from healthcare provider data > Improved staffing through health business management analytics

> Early detection of disease.

> Prevention of unnecessary doctor's visits.

> Discovery of new drugs.

> More accurate calculation of health insurance rates.

> More effective sharing of patient data

DISADVANTAGES:

REPLACING MEDICAL PERSONNEL:

Application of technology in every sphere of human life is improving the way things are done. These technologies are are also posing some threat to world of works. Robotics are replacing human labour.

DATA SAFETY:

Data security is another challenge in applying big data in healthcare. Big data storage is usually targets of hackers. This endangers the safety of medical data. Healthcare organisations are very much concerned about the safety of patients' sensitive personal data. For this, all healthcare applications must meet the requirement for data security and be HIPAA compliant before they can be deployed for healthcare services.

PRIVACY:

One of the major drawbacks in the application of big data in healthcare industry is the issue of lack of privacy. Application of big data technologies involves monitoring of patient's data, tracking of medical inventory and assets, organizing collected data, and visualization of data on the dashboard and the reports. So visualization of sensitive medical data especially that of the patients creates negative impression of big data as it violets privacy

MAN POWER:

Applying big data solutions in healthcare requires special skills, and such kills are scarce. Handling of big data requires the combination of medical, technological and statistical knowledge.

CHAPTER 11

CONCLUSIO

N

The impact of data analytics in healthcare has already made a substantial difference in the ability of healthcare providers to offer patients high-quality care in an efficient, cost-effective manner. However, the role of data analytics in improving patient outcomes and healthcare processes continues to grow and expand as more types of data become available and new tools are developed that make the results of the analytics clear and easy for healthcare professionals to access.

Realizing the potential of data analytics to transform the healthcare industry begins by understanding how the technology can be applied to address healthcare providers' challenges, including staff recruitment and utilization, operational efficiencies, and enhanced patient experiences. Patient-centered healthcare depends on knowing what patients want and need. Data analytics holds the key to unlocking this vital information.

CHAPTER 12

FUTURE SCOPE

Artificial Intelligence (AI) will play a significant role in data analytics in healthcare for the next decade. For example, the field of AI-enabled clinical decision support is just emerging. This type of support can compare patients who fit similar profiles within a system, then it can alert doctors to trends in data that may have been overlooked. The use of big data in healthcare will include testing for drug interactions that small studies are unlikely to catch and prevent patients from taking harmful drug combinations.

Decisions made by physicians, like what test or treatments to give a particular patient, makeup 80-90% of all healthcare spending, so using artificial intelligence to make more educated

decisions will bring down healthcare costs. It's crucial to have informed leaders at the vanguard of these innovations in healthcare.

CHAPTER 13

APPENDIX

GitHub & Project Demo Links:

https://github.com/IBM-EPBL/IBM-Project-30297-1660143653