ANALYSIS OF HOSPITAL HEALTHCARE DATA

NALAIYA THIRAN PROJECT BASED LEARNING ON

ON

PROFESSIONAL READINESS FOR INNOVATION,

EMPLOYABILITY AND ENTREPRENEURSHIP

A PROJECT REPORT

ARAVINDHAN	312319205016
JAGADISH	312319205049
JAICHARAN	312319205050
JALAGANDESWARAN	312319205051

ST.JOSEPH'S COLLEGE OF ENGINEERING
(An Autonomous Institution)
ST.JOSEPH'S GROUP OF INSTITUTIONS
JEPPIAR INTITUTUIONAL TRUST
OMR CHENNAI - 119

INTERNAL MENTOR

Ms.R.Deepa

INDUSTRIAL MENTOR

Shivam shivare , Hari Prabhu

TABLE OF CONTENTS

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.

By the end of this Project, you will:

- Know fundamental concepts and can work on IBM Cognos Analytics
- Gain a broad understanding of plotting different visualizations to provide the suitable solution.
- Able to create meaningful Visualizations and the Dashboard(s).

2.LITERATURE SURVEY

1)Angela Richard, MS, RN, Meg Kaehny, MSPH, Karis May and Andrew Kramer, MD(2009).Literature Review and Synthesis: Existing Surveys on Health Information Technology, Including Surveys on Health Information Technology in Nursing Homes and Home Health.https://aspe.hhs.gov/reports/literature-review-synthesis-existing-surveys-health-information-technology-including-surveys-health-1

In an effort to better understand the current use and adoption rates of electronic health records and other health information technology (HIT) applications within nursing homes, the Division of Health Care Policy and Research at the University of Colorado Denver has been contracted by the Office of the Assistant Secretary for Planning and Evaluation in the U.S. Department of Health and Human Services (HHS) to develop survey instruments for use in long-term care provider settings. Although numerous survey instruments have been fielded to assess HIT use in nursing homes and long-term care settings, the lack of consistent definitions, terminology, item construction, sampling frames, and measurement criteria render it difficult to accurately gauge current HIT adoption. In this report, we review existing surveys for long-term care and other provider settings pertaining to current HIT use and adoption, barriers to adoption, and recommend issues to consider when developing survey questions to ascertain HIT adoption, use, and barriers to adoption and use in nursing homes.

2) Rojas, E., Munoz-Gama, J., Sepúlveda, M. and Capurro, D., 2016. Process mining in healthcare: A literature review. Journal of biomedical Informatics, 61, pp.224-236.

The provision of quality hospital services depends on the suitable and efficient execution of processes. Healthcare processes are a series of activities aimed to diagnose, treat and prevent any diseases in order to improve a patient's health. These processes are supported by clinical and non-clinical activities, executed by different types of resources (physicians, nurses, technical specialists, dentists, clerks) and can vary from one organization to another. It is known that healthcare processes are highly dynames, complex, ad-hoc, and are increasingly multidisciplinary, making them interesting to analyze and improve. Healthcare processes improvement might have a high impact on the quality of life of patients. However, improving them is not an easy task and several challenges are always present.

3) Liberatore, M.J. and Nydick, R.L., 2008. The analytic hierarchy process in medical and health care decision making: A literature review. European Journal of Operational Research, 189(1), pp.194-207.

The United States continues to devote ever-increasing amounts of its resources to health care. The most recent statistics published by the US government indicate that health care spending was projected to reach $1.7 trillion or 15.3% of its gross domestic product (GDP) in 2003. In addition, this percentage is projected to increase to 18.7% in 10 years (Centers for Medicare and Medicaid Services and US Bureau of the Census, 2004). Total national health expenditures increased by 7.7% in 2003, four times the rate of inflation (Smith et al., 2005). Given the magnitude of these numbes and expenditures, improvement in health care and medical decision making can reap substantial benefits for both patients and health care providers alike. A variety of decision making methods and tools are available to support health care and medical decision making. The purpose of this paper is to review and assess the application of a well-known and widely used decision making methodology, called the analytic hierarchy process (AHP), to important problems in medical and health care decision making.

4) E. Ammenwerth, F. Ehlers, R. Eichstädter, R. Haux, U. Pohl, F. Resch(2002). Systems Analysis in Health Care: Framework and Example.https://www.researchgate.net/publication/11313504_Systems_Analysis_in_Health_Care_Framework_and_Example

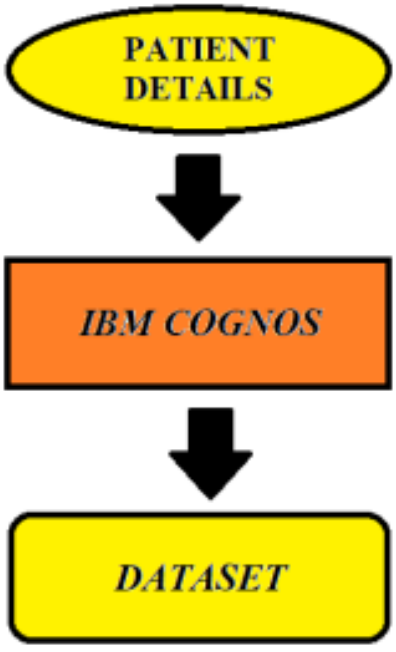
Due to the high complexity of structuresand processes in health care, thorough systems analy-ses in health care run the risk of becoming very com-plex and difficult to handle. Therefore, we aimed tosupport systematic systems analysis in health care bydeveloping a comprehensive framework that presentsand describes potential areas of analysis

2.1.EXISTING PROBLEM

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.

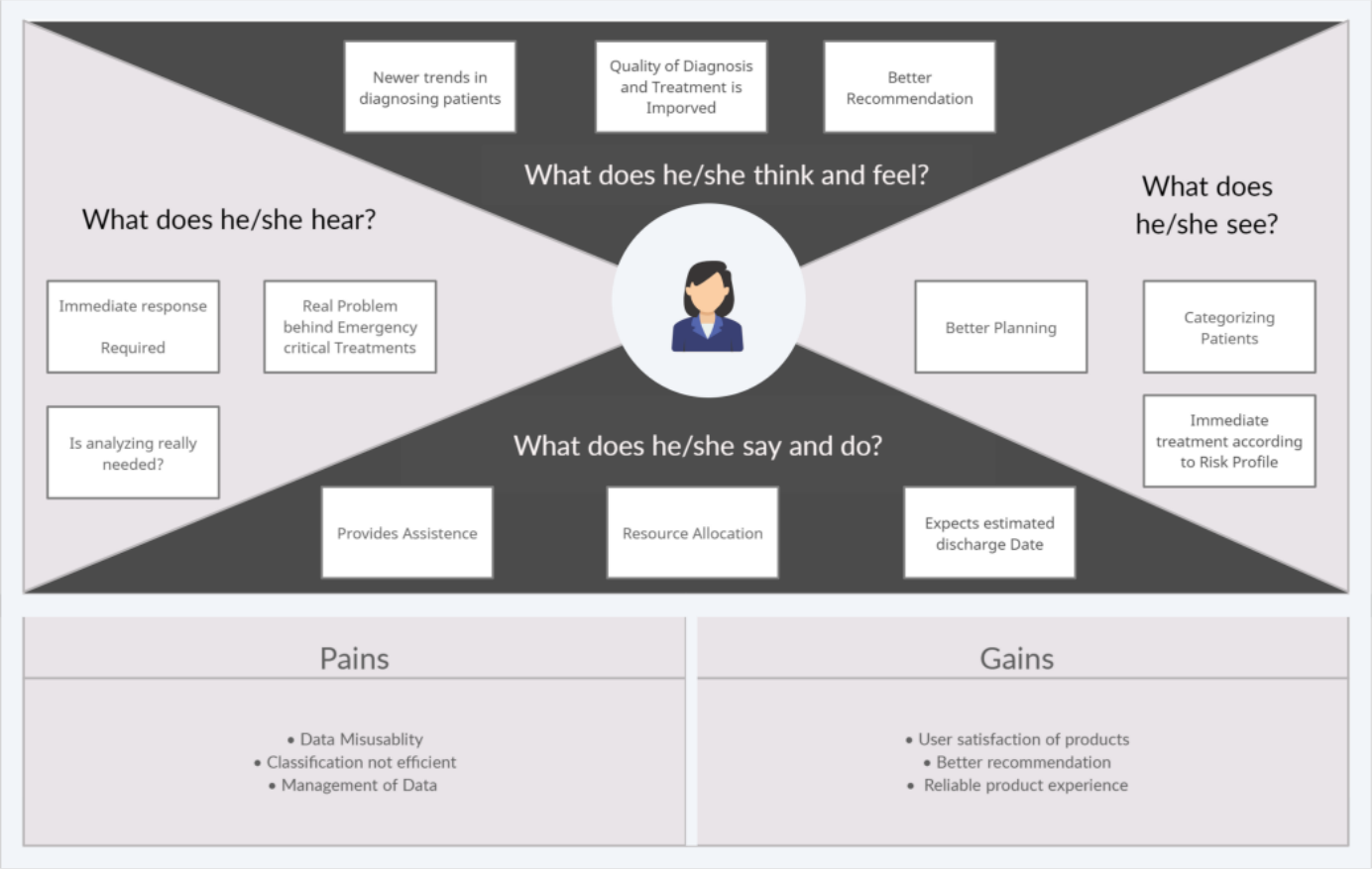
2.3.PROBLEM STATEMENT DEFINITION

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. With the help of the data set we can come to a conclusion that whether which patient is gonna stay for how many days based on their medical report. This is the most important thing in a hospital since based on the stay the resources are utilized accordindly. The length of stay is divided into 11 different classes ranging from 0-10 days or more than 100 days. Here we use multiple data either in the form of charts, tables or graphs and by using analytical visualization required dashboard is filled and this is then finalized in the IBM cognos analytics



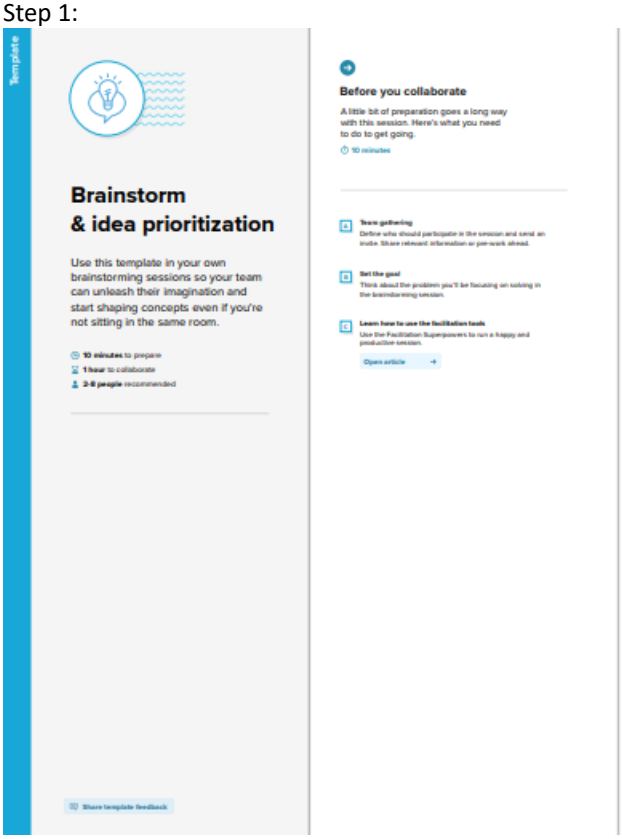
3.IDEATION & PROPOSED SOLUTION

3.1.EMPATHY MAP CANVAS



3.2.IDEATION & BRAINSTORMING

Step 1:



Step 2:

Jagadesh

Healthcare data analytics refers to the collection and analysis of patient data to improve medical care and patient experience.

Patients go through a continuum of caregiving from diagnosis to recovery.

This medical journey is called patient experience (PX).

Artificial intelligence can be applied to this type of analytics to make patient experience data reviews faster, more accurate, and multilingual.

The quality-of-care interactions that patients receive at different points in their journey culminate as their holistic experience.

Uncovering PX using healthcare analytics can help medical providers make required improvements across their enterprise.

Jaicharan

Hospitals can collect patient feedback data using various means

such as customer service calls, online forms, star ratings, interactive voice responses (IVR), and other digital methods.

This data can then be processed to extract the most high-value insights

such as patient sentiments surrounding themes, topics, and entities such as medical staff, wait times, communication, prescription dispensing, etc.

Healthcare data analysis can study past and present patient experiences.

These experiences are used to give medical organizations insights that can be used to improve the quality of care.

Jalagandeswaran

There can be an analytical study of industry-wide data where analysis can unearth patterns.

These patterns are key to understanding and predicting trends.

These trends can be used to control diseases, enhance clinical research, and build better patient management systems

AI-powered health analytics for patient experience can help healthcare providers fully leverage their data to improve the quality of care, patient satisfaction, and safety.

Data mining enables healthcare providers to have a better doctor-patient relationship, and in turn, offer improved patient care

On a global scale, data mining can help in advanced clinical trials and in the prevention of epidemic outbreaks.

Aravindhan

Patient feedback is a vital part of the healthcare data.

When healthcare organizations allow patients to express their opinions and feelings, they feel more engaged.

People express their feedback through various means like surveys, reviews, and electronic medical records (EMRs.).

The first step after distributing surveys and collecting all that patient voice information is to mine the data.

This is the process of extracting the most relevant information and then organizing it in a way that makes it easy to scrutinize for patterns and recurring themes.

These results can also be scored for positive, neutral, or negative patient sentiments toward a variety of healthcare-specific themes such as medical staff, facility hygiene, patient service, waiting times, etc.

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 and break it up into smaller sub-groups.

⏱ 20 minutes

High Value Insights

Medical staff

Wait times

Communication, prescription dispensing

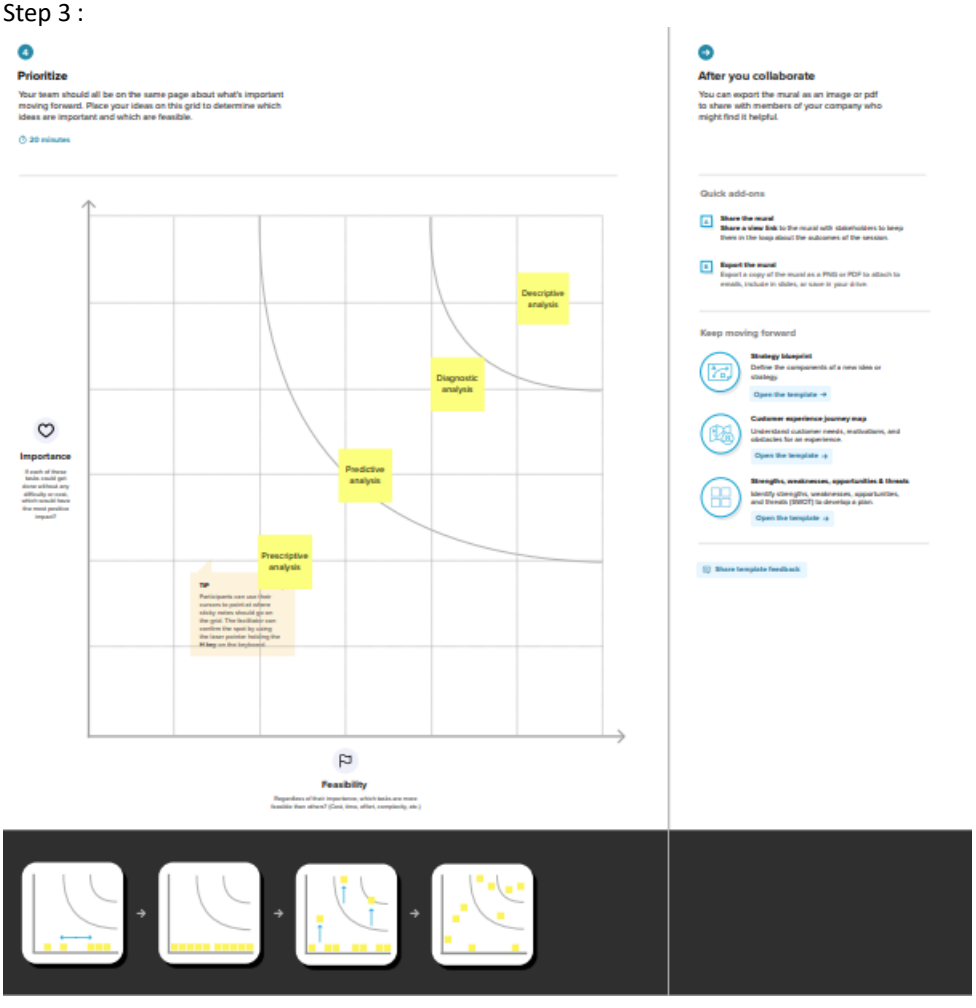
Methods of Patient Feedback

surveys

Reviews

Electronic medical records (EMRs.).

Step 3 :

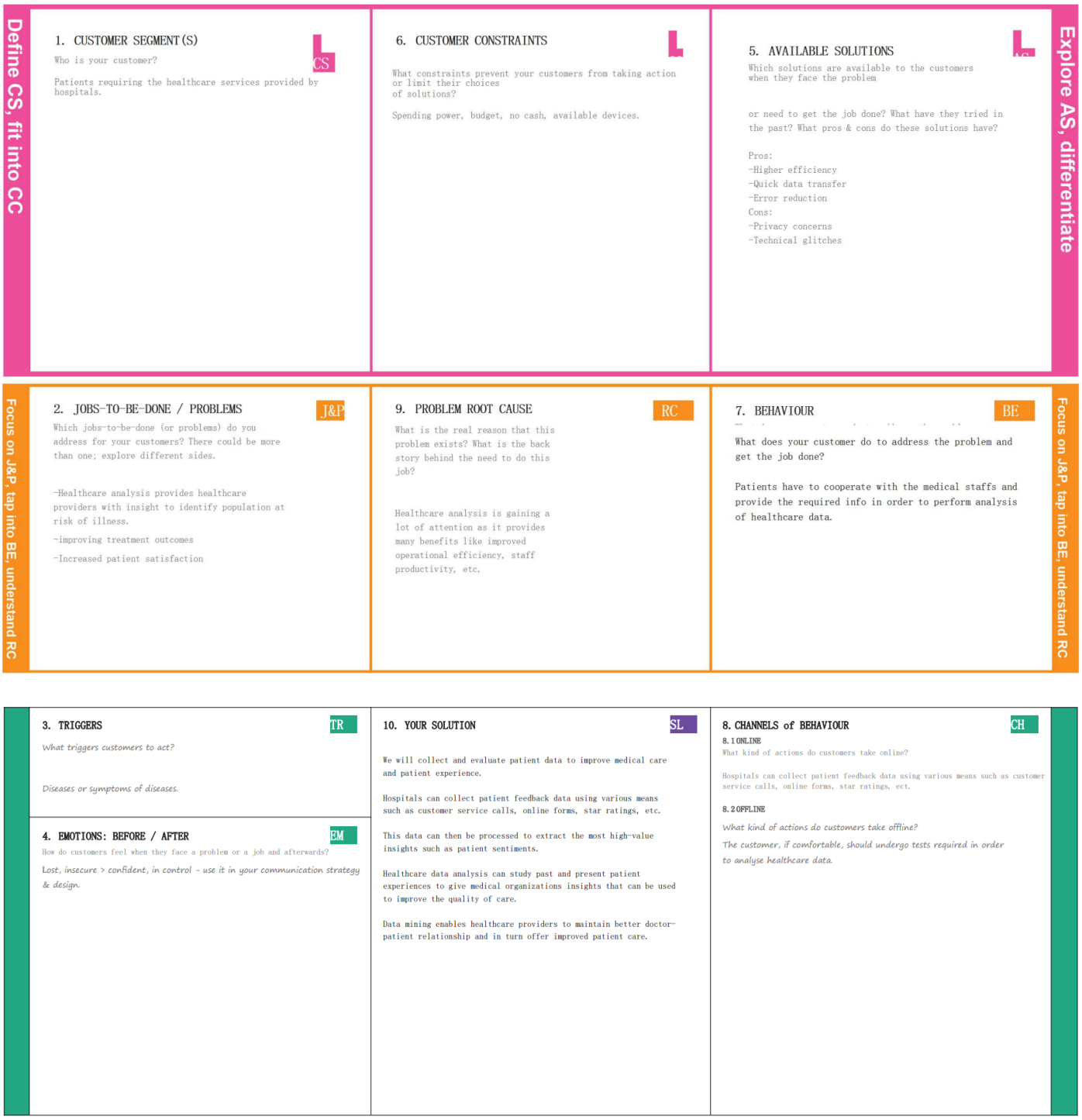


3.3.PROPOSED SOLUTION

Proposed Solution :

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	To perform analysis of hospital health care data
2.	Idea / Solution description	Primary activities that must be undertaken during a data analysis project include identifying the right hospital KPIs to measure, cleaning, scrubbing and meta-tagging healthcare data to line up with the KPIs to measure, and then displaying it visually.
3.	Novelty / Uniqueness	It is the first, to our knowledge, to take a comprehensive review approach and offer a holistic picture of health care analytics and data mining
4.	Social Impact / Customer Satisfaction	Impact of service attributes on customer satisfaction and loyalty in a healthcare context. Customer perceptions at the service attribute level can often be the key to the generation and management of customer satisfaction and loyalty.
5.	Business Model (Revenue Model)	From an enterprise perspective, a healthcare organization's revenue model should have strategic parity with the organization's operating and clinical care models, combining to represent the organization's enterprise business model.
6.	Scalability of the Solution	Prioritising scalability is an essential aspect of digital health solutions, as it facilitates lower maintenance costs, better user experience and can contribute to new models of care and improved health outcomes.

3.4.PROBLEM SOLUTION FIT



4.REQUIREMENT ANALYSIS

4.1.FUNCTIONAL REQUIREMENT

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	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	Check out	Analyzing and monitoring of beds which arerequired are the most important task. Using flawless systems for accurately trackingthe availability of beds.
FR-5	Providing insights of dataset	Raw data collection and sharing of data and systems are essential factors in hospital management. According to these data in appropriate measures can be taken. Providing data set without human error.

4.2.NON-FUNCTIONAL REQUIREMENTS

Following are the non-functional requirements of the proposed solution

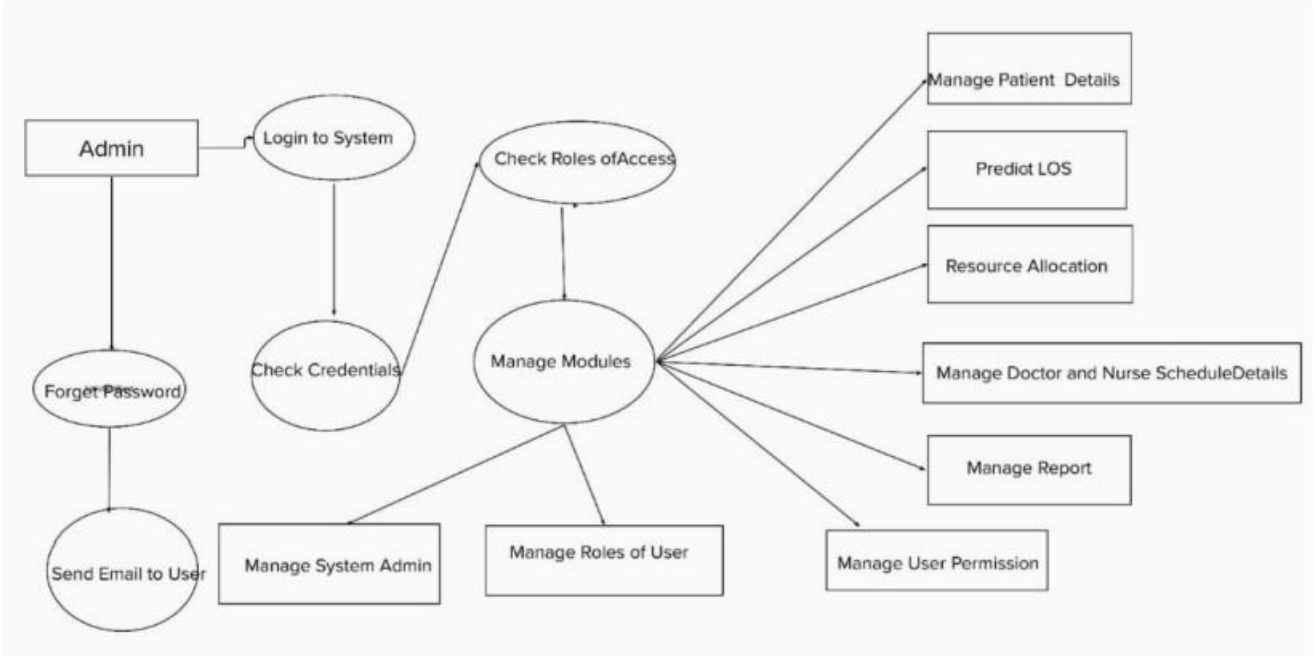
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The designed dashboard provides the Patient's LOS and Severity of illness through graphs, piechart, cross tab and other tools
NFR-2	**Security**	This process of protecting data from unauthorized access and data corruptionthroughout its lifecycle
NFR-3	**Reliability**	The project must have minimal degree of failure under normal usage and how often does the userget access to this work
NFR-4	**Performance**	This dashboard operates quickly, offers highlevels of interactive data, and has a large capacity for data. And data are frequently updated.
NFR-5	**Availability**	Better coordination with the hospital management to provide all its resources accessible when needed. Accessibility of allmedical facilities.
NFR-6	**Scalability**	The project allows multiple users to handle the data at the same time. It is highly scalable since adding features and making advancements in thewebsite is uncomplicated.

5.PROJECT DESIGN

5.1.DATA FLOW DIAGRAM

Data Flow Diagram :

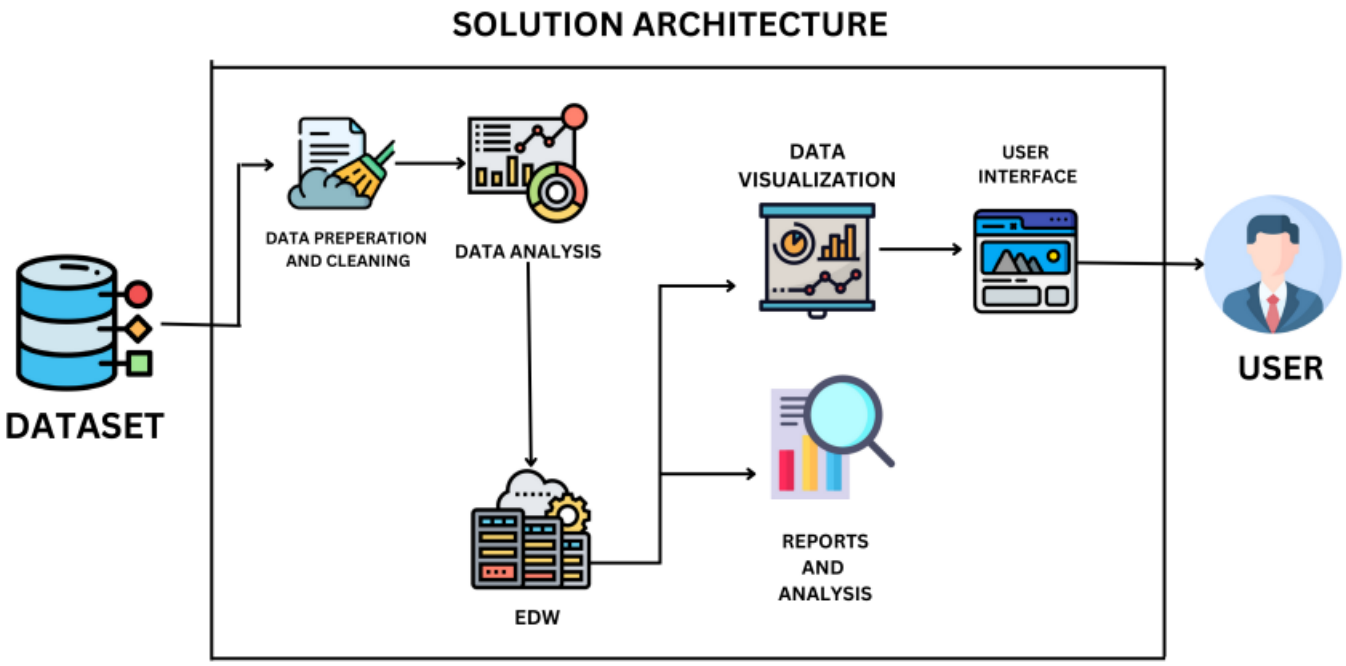
A Data Flow Diagram(DFD) is a traditional visual representation of the information that flows within a system. A neat and clear DFD can depict the right amount of the system requirements graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



5.2.SOLUTION AND TECHNICAL ARCHITECTURE

Solution architecture is a complex process - with many sub-processes - that bridges the gap between business problems and technology solutions. It's goals are to :

- Find the best tech solution to solve existing business problems.
- Describe the structure, characteristics, behaviour, and other aspects of the software to project stakeholders.
- Define features, development phases and solution requirements.
- Provide specifications according to which the solution is defined, managed and delivered.



TECHNOLOGY ARCHITECTURE

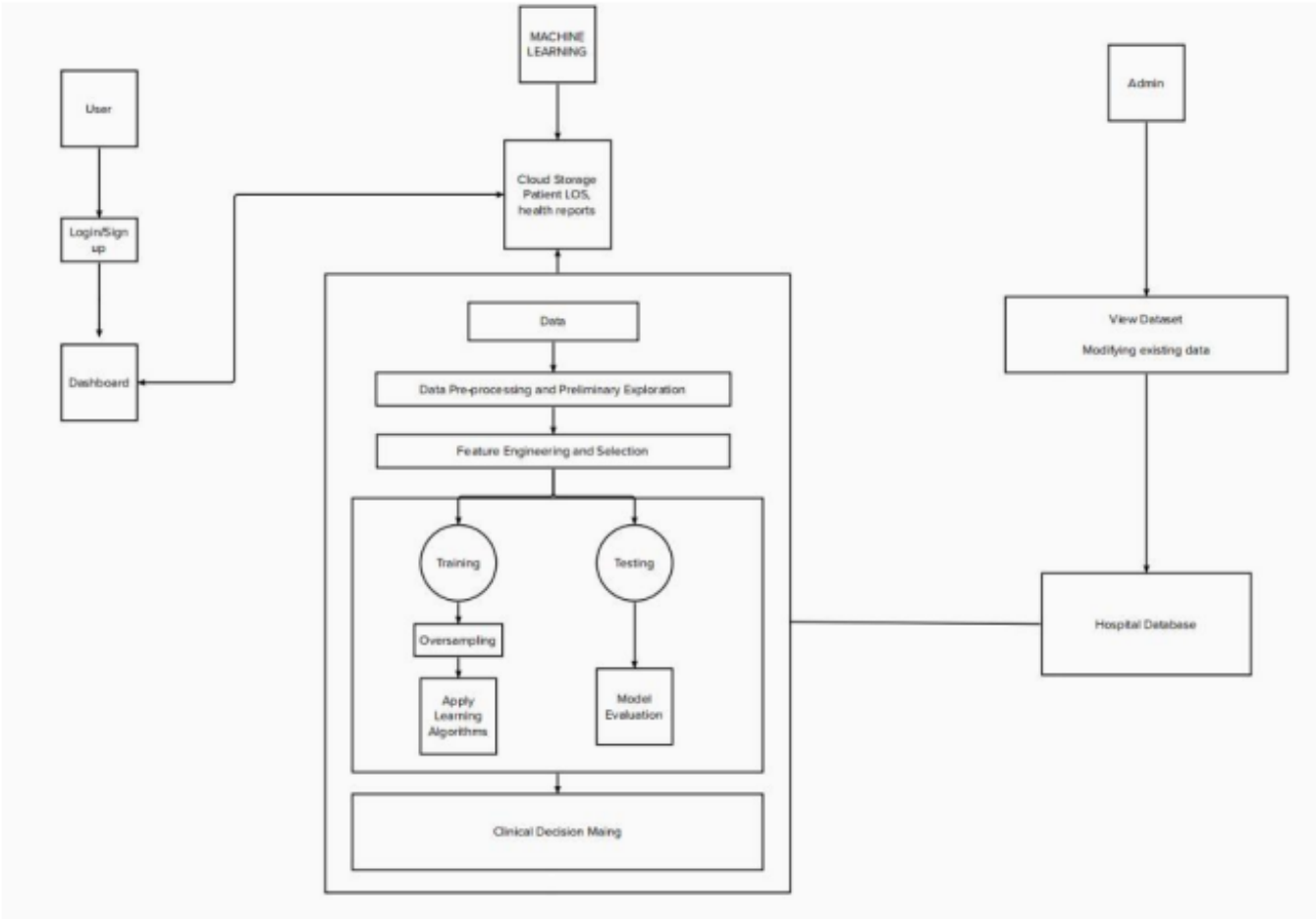


Table-1 : Components & Technologies:

S.No	Component	Description	Technology
1.	User Interface	How user interacts with application	HTML, CSS, JavaScript
2.	Database	Data Type, Configurations	MySQL
3.	Cloud Database	Database Service on Cloud	IBM Cloudant
4.	File Storage	File storage requirements	IBM Block Storage or Other Storage Service
5.	Prediction	ML algorithms are used for predicting the length of stay	ML algorithms - Random forests, Decision trees

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	List the open-source frameworks used	Python
2.	Security Implementations	List all the security / access controls implemented, use of firewalls etc.	Firewall, Antivirus
3.	Scalable Architecture	Justify the scalability of architecture (3 – tier, Micro-services)	Supports higher workloads
4.	Availability	Justify the availability of application	AWS used
5.	Performance	Design consideration for the performance of the application	Dashboard, report, stories

5.3.USER STORIES

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the application through Facebook	I can register & access the dashboard with Facebook Login	Low	Sprint-2
		USN-4	As a user, I can register for the application through Gmail	I can register for the application through Gmail	Medium	Sprint-2
	Login	USN-5	As a user, I can log into the application by entering email & password	I can log into the application by entering email & password	High	Sprint-2
	Dashboard	USN-6	As a user, I can use my account in my dashboard for uploading dataset	I can use my account in my dashboard for uploading dataset	Medium	Sprint-3
Customer (Web user)	Website	USN-7	As a user, I can use my dashboard in website	I can use my dashboard in website	Medium	Sprint-3
Customer Care Executive		USN-8	As a user, I can contact Customer care executive for my login	I can contact Customer care executive for my login	High	Sprint-4
Administrator		USN-9	As a user, I can contact administrator for my queries	I can contact administrator for my queries	High	Sprint-4
Exploration	Dashboard	USN-10	As a user , I can prepare Data by using Exploration techniques	I can prepare Data by using Exploration techniques	High	Sprint-3
Presentation	Dashboard	USN-11	As a user , I can Present Data in my Dashboard	I can Present Data in my Dashboard	High	Sprint-4
Visualization	Dashboard	USN-12	As a user, I can prepare data by using visualization techniques	I can prepare data by using visualization techniques	High	Sprint-3

6.PROJECT PLANNING & SCHEDULING

6.1.MILESTONE AND ACTIVITY LIST

MILESTONES	ACTIVITY LIST
MILESTONE-1	Collecting the data based on the application
MILESTONE-2	Uploading the collected data on the IBM COGNOS platform
MILESTONE-3	Data exploration in the IBM COGNOS platform
MILESTONE-4	Data visualization in the IBM COGNOS platform
MILESTONE-5	Creating an interactive dashboard
MILESTONE-6	Displaying the prepared dashboard
MILESTONE-7	Preparing a standard dataset and removing the unwanted data using the python programming
MILESTONE-8	By using the various algorithm and exploring the result and getting the accurate result with the help of an algorithm which give more accuracy
MILESTONE-9	Displaying the result according to the required format for example displaying the Length Of Stay of a patient
MILESTONE-10	Deployed in the GitHub

6.2.SPRINT DELIVERY PLAN

Product Backlog, Sprint Schedule, and 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 in IBM cloud and the data arecollected.	20	High	2 Members
Sprint-2	Analyze	USN-2	As a health care provider all the data thatare collected is cleaned and uploaded in the database or IBM cloud.	20	Medium	2 Members
Sprint-3	Dashboard	USN-3	As a health care provider I can use my account in my dashboard for uploading dataset.	10	Medium	2 Members
Sprint-3	Visualization	USN-4	As a health care provider I can prepare data for Visualization.	10	High	2 Members
Sprint-4	Visualization	USN-5	As a health care provider I canpresent data in my dashboard.	10	High	2 Members
Sprint-4	Prediction	USN-6	As a health care provider I can predict the length of stay	10	High	2 Members

Project Tracker, Velocity & Burndown Chart:

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

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

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)

7.PROJECT DEVELOPMENT PHASE

Project Development Phase

Sprint-1:
➢ Data Collection ➢ Data Preparation

Sprint-2:
➢ Data Exploration

Sprint-3:
➢ Dashboard creation

Sprint-4:
➢Report Creation ➢ Story Creation

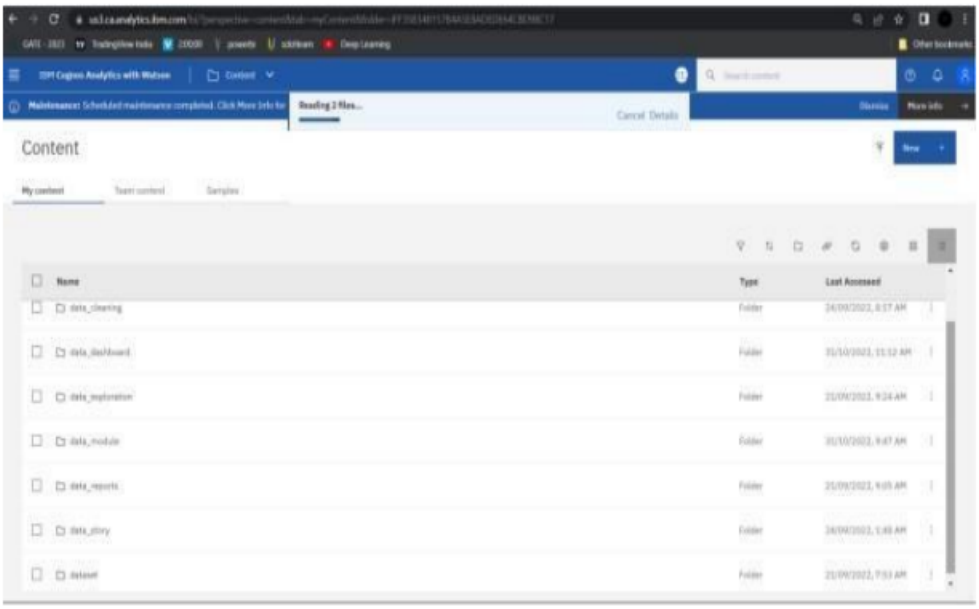
Sprint-1:

Data Collection: Download the Dataset Link:

https://drive.google.com/file/d/1slC0MhsJHeuODVkhIXrdNOX_aBDySSBh/view

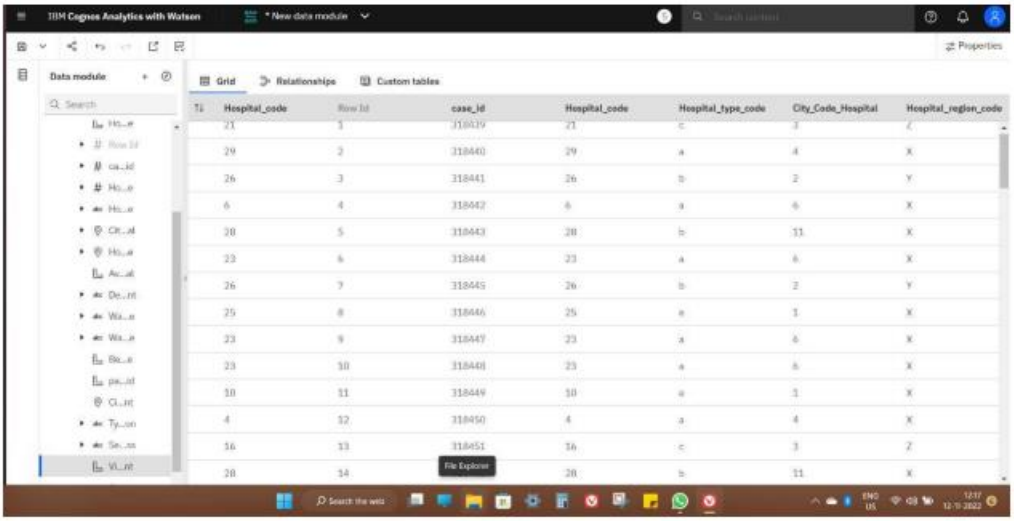
Load the Dataset:

Tool used: IBM Cognos Analytic

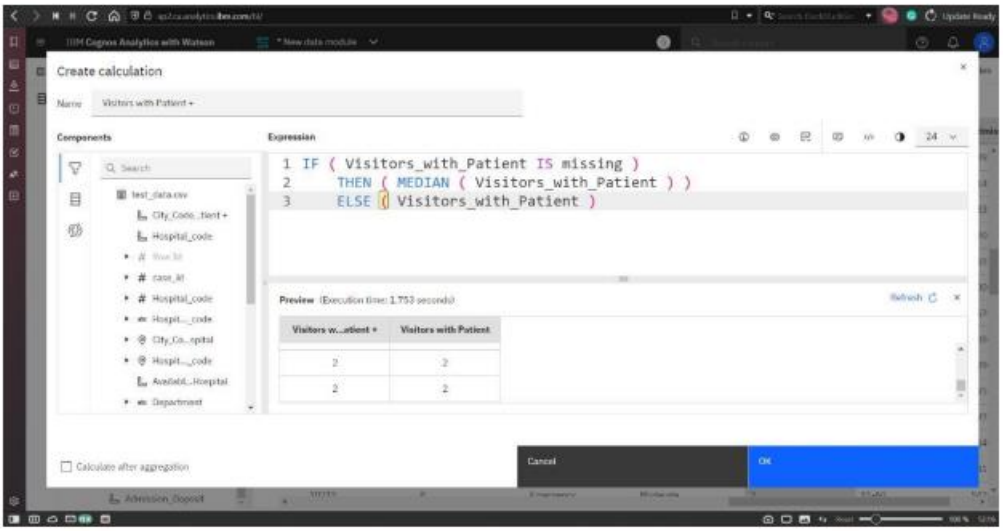


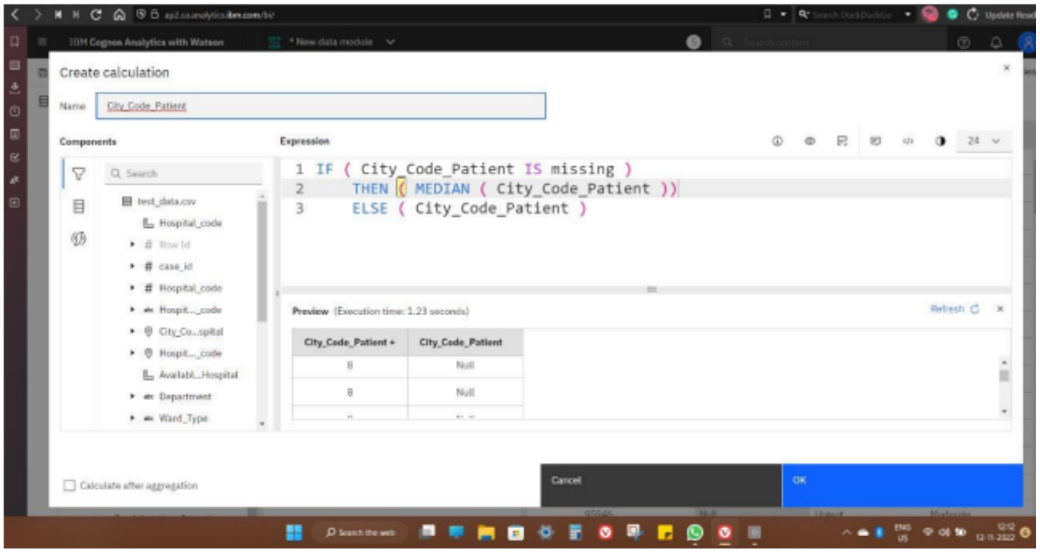
Data Preparation:

Prepare the Dataset

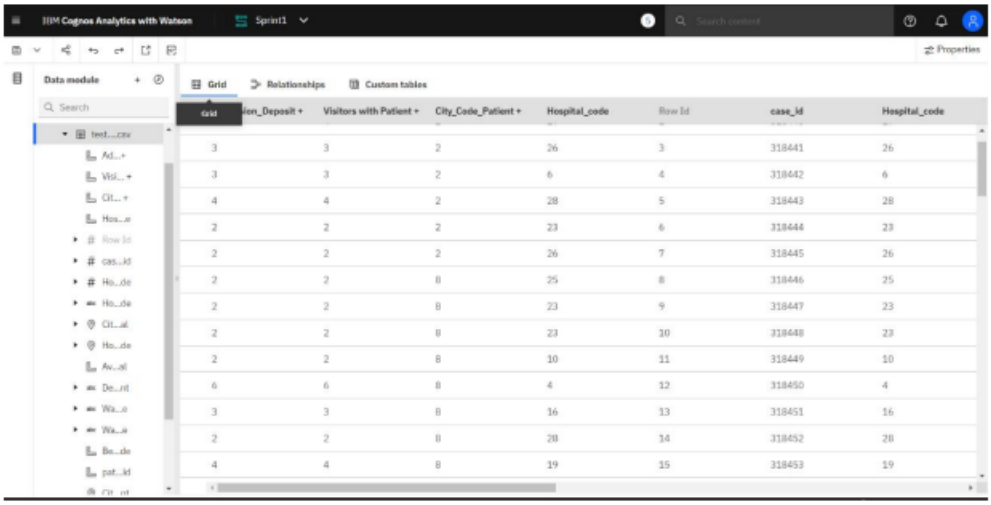


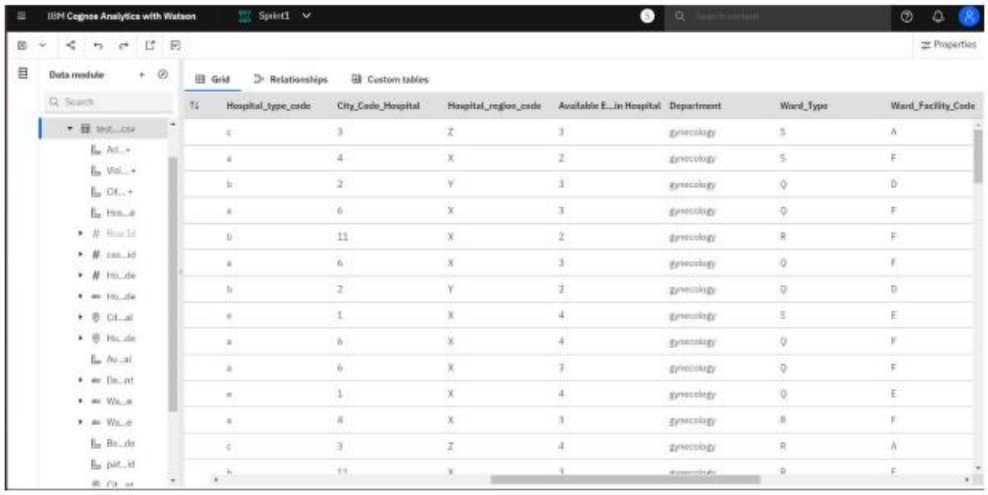
Null Values Cleaning Process:





Hospital Data:





Prepared data link :

https://ap2.ca.analytics.ibm.com/bi/?perspective=ca-modeller&id=iCB6DC2DA80B94C6695EABB38C4383360&objRef=iCB6DC2DA80B94C6695EABB38C4383360&tid=1724333042_4b0456d15d4f4de1ab03ac8afaaea773_sessionTemp

8. CONCLUSION

Thus we have developed a system 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.

With the help of the data set we can come to a conclusion that whether which patient is gonna stay for how many days based on their medical report. This is the most important thing in a hospital since based on the stay the resources are utilized accordingly. The length of stay is divided into 11 different classes ranging from 0-10 days or more than 100 days. Here we use multiple data either in the form of charts, tables or graphs and by using analytical visualization required dashboard is filled and this is then finalized in the IBM cognos analytics.

9.APPENDIX

9.1.REFERENCES

[1] Liberatore, M.J. and Nydick, R.L., 2008. The analytic hierarchy process in medical and health care decision making: A literature review. European Journal of Operational Research, 189(1), pp.194-207.

[2] Rojas, E., Munoz-Gama, J., Sepúlveda, M. and Capurro, D., 2016. Process mining in healthcare: A literature review. Journal of biomedical informatics, 61, pp.224-236.

[3] Schluter, J., Winch, S., Holzhauser, K. and Henderson, A., 2008. Nurses' moral sensitivity and hospital ethical climate: A literature review. Nursing ethics, 15(3), pp.304-321.

[4] Devaraj, S. and Kohli, R., 2000. Information technology payoff in the health-care industry: a longitudinal study. Journal of management information systems, 16(4), pp.41-67.

[5] Wang, Y., Kung, L. and Byrd, T.A., 2018. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. Technological forecasting and social change, 126, pp.3.

[6] Nanibaa'A, G., Sathe, N.A., Antommaria, A.H.M., Holm, I.A., Sanderson, S.C., Smith, M.E., McPheeters, M.L. and Clayton, E.W., 2016. A systematic literature review of individuals' perspectives on broad consent and data sharing in the United States. Genetics in Medicine, 18(7), pp.663-671.

[7] Guerriero, F. and Guido, R., 2011. Operational research in the management of the operating theatre: a survey. Health care management science, 14(1), pp.89-114.

[8] Aboelela, S.W., Larson, E., Bakken, S., Carrasquillo, O., Formicola, A., Glied, S.A., Haas, J. and Gebbie, K.M., 2007. Defining interdisciplinary research: Conclusions from a critical review of the literature. Health services research, 42(1p1), pp.329-346.

[9] Günal, M.M. and Pidd, M., 2010. Discrete event simulation for performance modelling in health care: a review of the literature. Journal of Simulation, 4(1), pp.42-51.

[10] Giacomini, M.K., Cook, D.J., Evidence-Based Medicine Working Group and Evidence-Based Medicine Working Group, 2000. Users' guides to the medical literature: XXIII. Qualitative research in health care A. Are the results of the study valid?. Jama, 284(3), pp.357-362.

GITHUB & PROJECT GITHUB LINK

GITHUB LINK:

GitHub - IBM-EPBL/IBM-Project-8740-1658928427: Analytics for Hospitals' Health-Care Data

PROJECT VIDEO LINK:

https://youtu.be/M4iajjZcjy8