

IBM ANALYTICS FOR HOSPITAL HEALTH CARE DATA

TEAM ID: PNT2022TMID37244

GROUP MEMBERS:

ABDULLAH AMBERKHANI

MURALIDHARAN

MUHAMMED ABDUL BASIDH

GAURAV RANKA

Project 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. **RESULTS**
 - 7.1 Visualizations
8. **ADVANTAGES & DISADVANTAGES**
9. **CONCLUSION**
10. **FUTURE SCOPE**
11. **APPENDIX**
 - 12.1 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. 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.

1.1 Purpose

The purpose 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

Recent Covid-19 Pandemic has raised alarms over one of the most overlooked areas to focus: Healthcare

Management. While healthcare management has various use cases for using data science, patient length of stay is one critical parameter to observe and predict if one wants to improve the efficiency of the healthcare management in a hospital.

This parameter helps hospitals to identify patients of high LOS-risk (patients who will stay longer) at the time of admission. Once identified, patients with high LOS risk can have their treatment plan optimized to minimize LOS and lower the chance of staff/visitor infection. Also, prior knowledge of LOS can aid in logistics such as room and bed allocation planning.

Suppose you have been hired as Data Scientist of Health Man – a not for profit organization dedicated to manage the functioning of Hospitals in a professional and optimal manner.

2.2 References

[IBM-Project-11497-1659331419/literature survey.pdf at main · IBM-EPBL/IBM-Project-11497-1659331419 \(github.com\)](#)

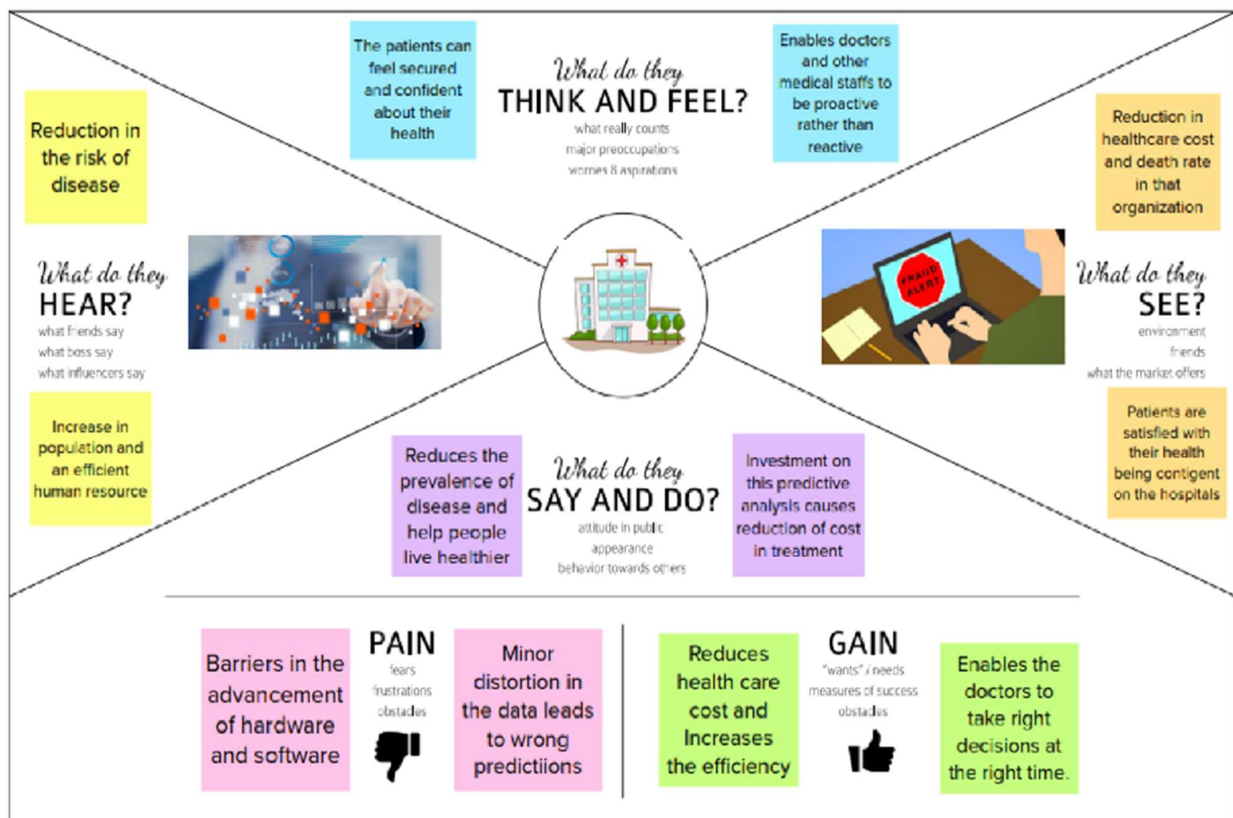
[IBM-Project-11497-1659331419/literature survey.pdf at main · IBM-EPBL/IBM-Project-11497-1659331419 \(github.com\)](#)

2.2 problem statement definition

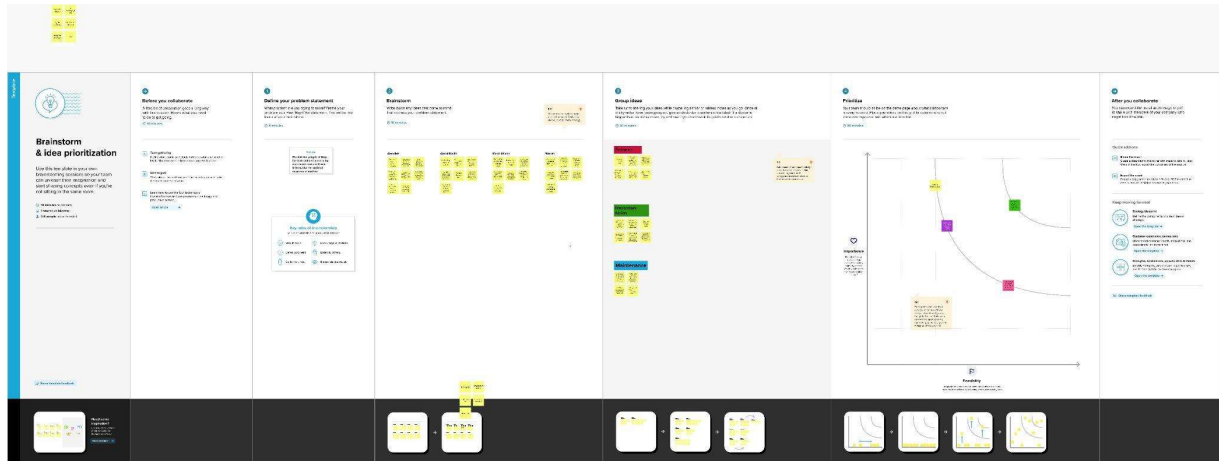
The goal is to accurately predict the Length of Stay for each patient on case-by-case basis so that the Hospitals can use this information for optimal resource allocation and better functioning. The length of stay is divided into 11 different classes ranging from 0-10 days to more than 100 days.

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas



3.2 Ideation and Brain Storming



3.3 Proposed Solution

S. No	Parameter	Description
1	Problem Statement (Problem to be solved)	<p>Analytics for Hospitals Health-Care Data:</p> <p>Hospitals have some main challenges such as deficient infrastructure, deficient manpower, unmanageable patient load, etc. so people can benefit if these problems are solved by adhering to certain software or some notes to maintain them all.</p> <p>The goal is to accurately predict the Length of Stay for each patient on a 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.</p>

2	Idea/Solution description	We are able to predict the length of stay of patients with data from the movement they entered the hospital and are diagnosed with an accuracy of ~70%. Such a model has the ability to profoundly improve hospital management and patient well-being.
3	Novelty/Uniqueness	Length of stay in the hospital differs based upon the critical in their health situation it can range between 2 to 3 days or even up-to 10- 20 days so based on the exploratory analysis of various patients we can accurately predict the length of stay of patients and can allocate optimum resource allocation.
4	Social Impact/Customer satisfaction	With Exploratory analysis using different methods to predict the length of stay creates a way the patients to know the vacancy of beds in the hospitals and also paved a way in their critical times to secure their better life.
5	Business Model (Revenue Model)	Using this model The usage of length of stay of patients in the hospitals has increased among the people and it is free of cost to get the details about the vacancy. It doesn't affect the revenue model.
6	Scalability of the Solution	<p>It is a easily scalable method using dataset of previous patients we can able to predict the LOS</p> <ul style="list-style-type: none"> · Increased productivity among the users · Decreased stress level · Possibility of getting the detailed list of vacancy

3.4 Problem Solution Fit

1. CUSTOMER SEGMENT(S)

This project is mainly for patients and hospital who wants to know the length of staying of existing patients so that they can get admitted into that hospital.

2. JOBS-TO-BE-DONE / PROBLEMS

The main goal is to accurately predict the length of stay of the patients in the hospital so that the out patients can know whether they can admitted into the hospital otherwise they can switch over to other hospital.

3. TRIGGERS

For estimating better prediction of length of stay of patients accurate estimation is needed

4. EMOTIONS: BEFORE / AFTER

Patients feels restless and they struggled to know where they can get admitted with a bed

5. AVAILABLE SOLUTIONS

When patients are facing the problem about the vacancy, using some existing data to accurately predict the availability, with exploratory analysis, etc.

6. CUSTOMER CONSTRAINTS

Network connection is a major issue while searching for availability of hospitals Also Budget is also a main constraints for majority of the peoples

7. BEHAVIOR

Use of some exploratory analysis to accurately predict the availability of vacancy can really helpful to the patients.

8. CHANNELS of BEHAVIOR

Patients and hospitals are the Customers. In online patients c–an able to check the availability with some data models. If they are nearby to the hospital they can directly come offline to the hospital.

9. PROBLEM ROOT CAUSE

Due to the lack of staffs to take care of the patients, accurate prediction is needed to predict accurately the length of stay of existing patients.

10. SOLUTION

To accurately predict the length of stay of patients in the hospital we can use the previous datasets of the patients based on that datasets we can able to predict the availability.

4. Requirement Analysis

4.1 Functional Requirements

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	The User has his/her own ID to get registered in the portal.
FR-2	User Confirmation	Confirmation via email and confirmation via OTP
FR-3	Dashboard	The collected data are found in visualized format and The prior data are analysed.
FR-4	Dataset	The patients record and staffs record are collected and consolidated as dataset.
FR-5	Report Generator	The periodic reports of patients and the LoS are reported.
FR-6	Exploration	The data exploration on available dataset

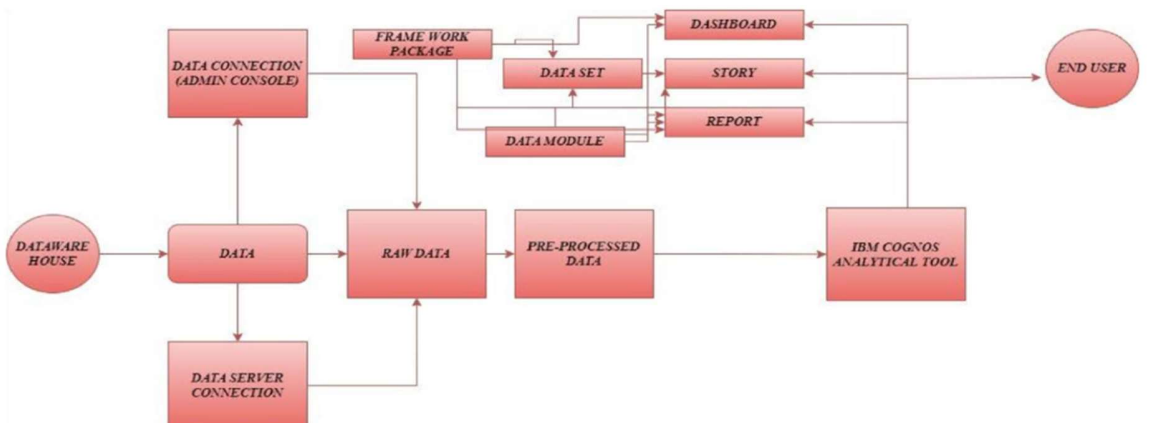
4.2 Non - Functional Requirements

FR No.	Non-Functional Requirement	Description
--------	----------------------------	-------------

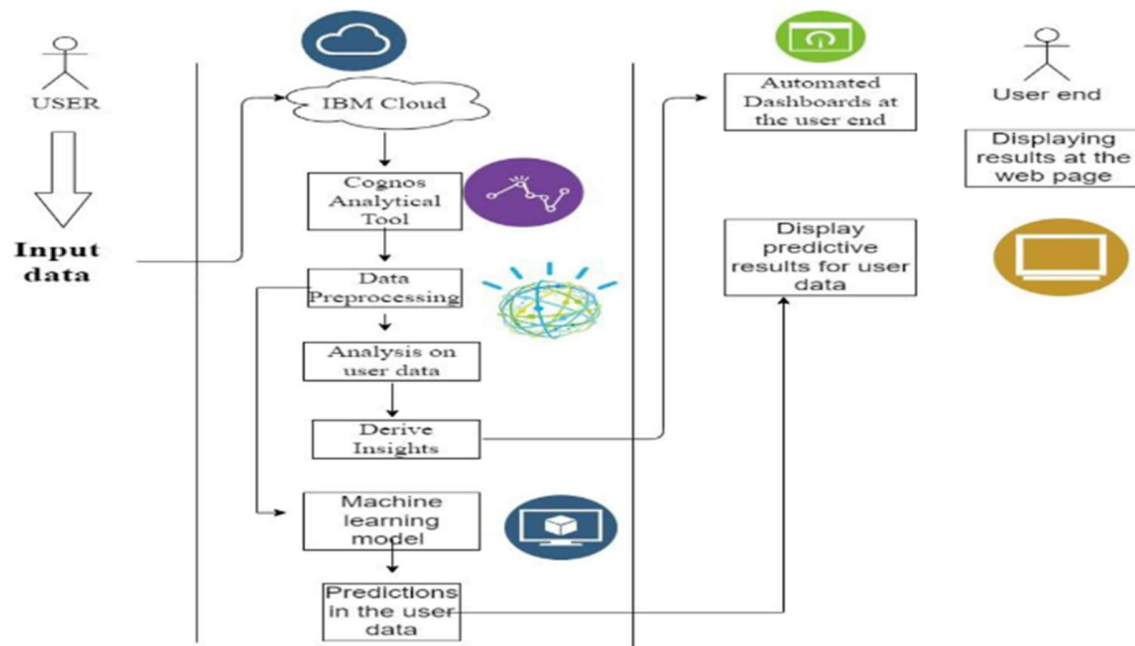
NFR-1	Usability	No prior experience required to use the dashboard. People with basic understanding can use the system.
NFR-2	Security	Only registered user can use this application.
NFR-3	Reliability	The Analytics system ensures the reliability
NFR-4	Performance	Gets updated regularly to improve the performance of the application.

5. Project Design

5.1 Data Flow Diagram



5.2 Solution & Technical Architecture



5.3 User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Patient	Hospitalization	USN-1	Patients are required to hospitalize if they have any medical issues.	Direct Hospitalization	High	Sprint-1
	Treatment Report	USN-2	Patients should collect them treatment report and get further doctor consult.	They can receive the report from hospital.	High	Sprint-1
Hospital Management	Resource Allocation	USN-3	Hospital Management should allocate the Necessary resource for treating the Patients.	Should be ready for any Circumstance.	High	Sprint-2

	Predicting Length of Stay	USN-4	The Doctors should be aware of condition of Patients to predict the LoS.	Exploring the data about the patient health condition and predicting LoS.	High	Sprint-1
	Resource Availability	USN-5	The Hospital Staff should be aware of available resources in hospital.	Visualizing the about the resource availability	High	Sprint-1

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 user, I can register for the IBM Cognos application by entering my email, and password, and confirming my password.	1	High	<ul style="list-style-type: none"> Abdullah Amberkhani Gaurav Ranka Mohammed Abdul basidh Muralidharan
Sprint-1	Confirmation Mail	USN-2	As a user, I will receive a confirmation email once I have registered for the application	1	High	<ul style="list-style-type: none"> Abdullah Amberkhani Gaurav Ranka Mohammed Abdul basidh Muralidharan
Sprint-1	Login	USN-3	As a user, I can log into the application by entering my email & password	1	High	<ul style="list-style-type: none"> Abdullah Amberkhani Gaurav Ranka Mohammed Abdul basidh Muralidharan
Sprint-1	Dashboard	USN-4	Upload dataset to IBM Cognos	1	High	<ul style="list-style-type: none"> Abdullah Amberkhani Gaurav Ranka Mohammed Abdul basidh Muralidharan
Sprint-2	Exploration	USN-5	Explore the dataset and find attributes and inconsistencies	1	High	<ul style="list-style-type: none"> Abdullah Amberkhani Gaurav Ranka Mohammed Abdul basidh Muralidharan
Sprint-2	Cleaning	USN-6	Clean the dataset for inconsistencies and errors	1	High	<ul style="list-style-type: none"> Abdullah Amberkhani Gaurav Ranka Mohammed Abdul basidh Muralidharan

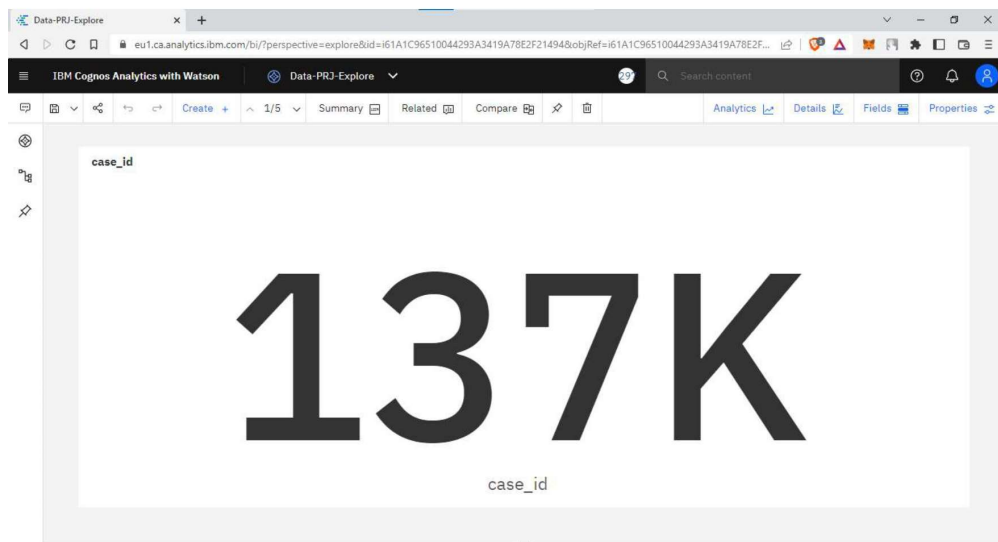
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-3	Data visualization	USN-7	Create visualization for the following: <ul style="list-style-type: none"> Length of Stay for each case of a patient. Stay by Patient ID using Column Chart Severity of illness by Patient-Id using Tree Map 	1	High	<ul style="list-style-type: none"> Abdullah Amberkhani Gaurav Ranka Mohammed Abdul basidh Muralidharan
Sprint-4	Data Visualization	USN-8	Create visualization for the following: <ul style="list-style-type: none"> Age, Department Wise Patient using Table Room Availability by Pie Chart Dashboard Creation Department-wise no. of admissions by Waterfall Chart 	1	High	<ul style="list-style-type: none"> Abdullah Amberkhani Gaurav Ranka Mohammed Abdul basidh Muralidharan
Sprint-4	Save the visualization to the Dashboard	USN-9	The Visualization created is saved to the Dashboard		High	<ul style="list-style-type: none"> Abdullah Amberkhani Gaurav Ranka Mohammed Abdul basidh Muralidharan

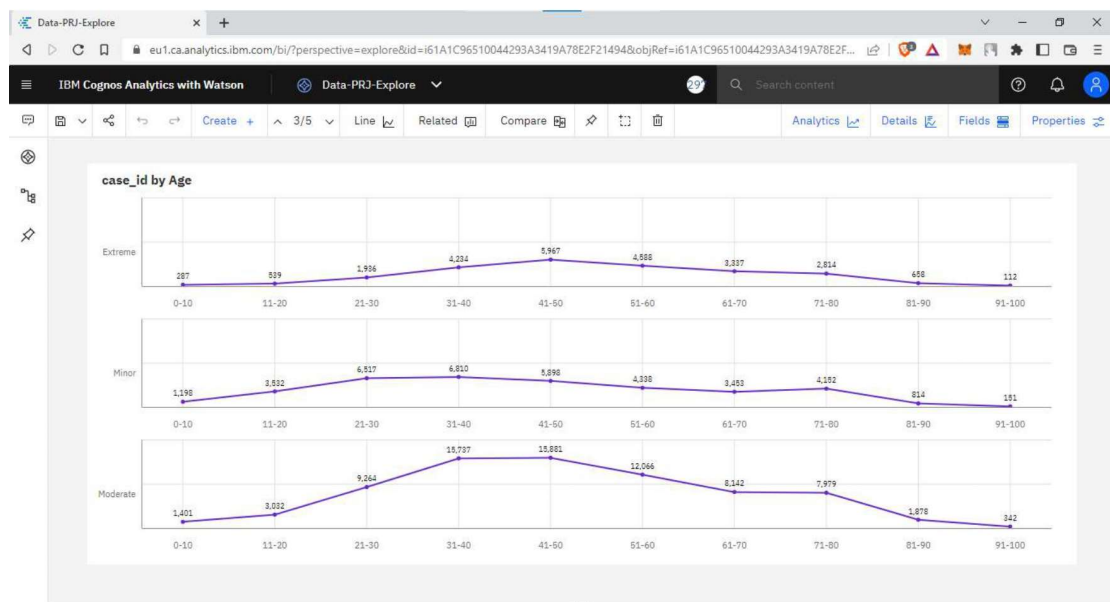
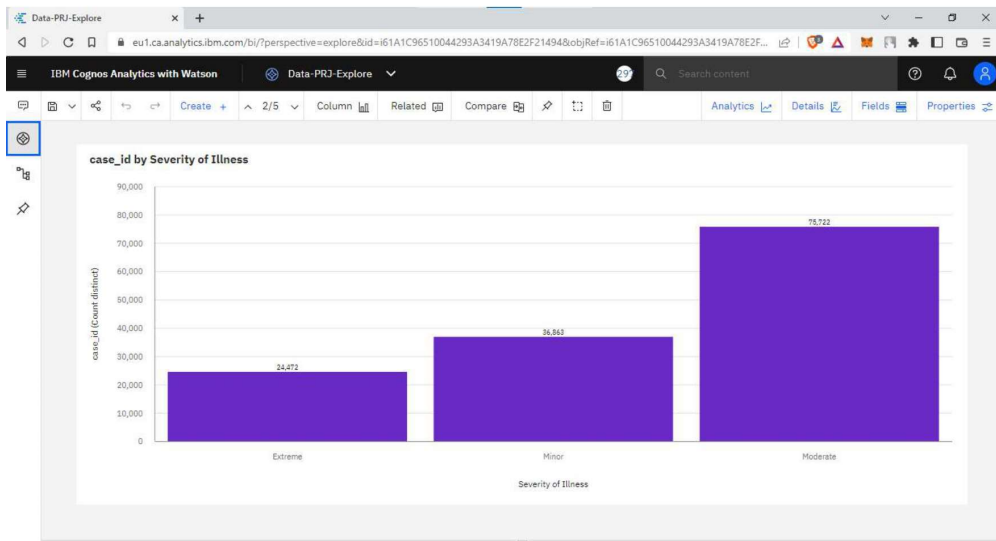
6.2 Sprint Delivery Schedule

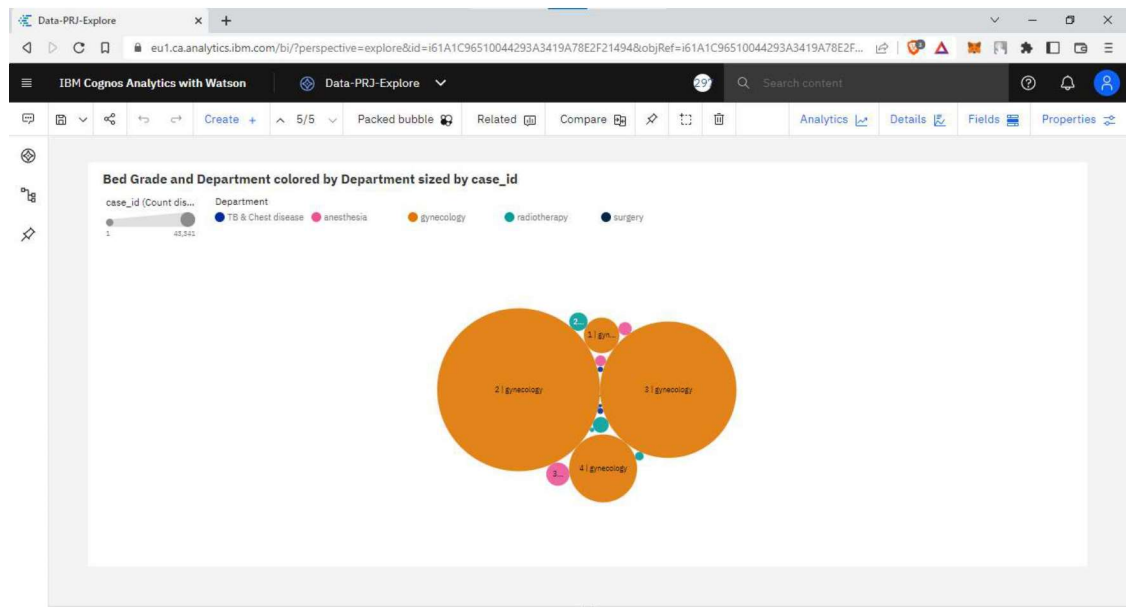
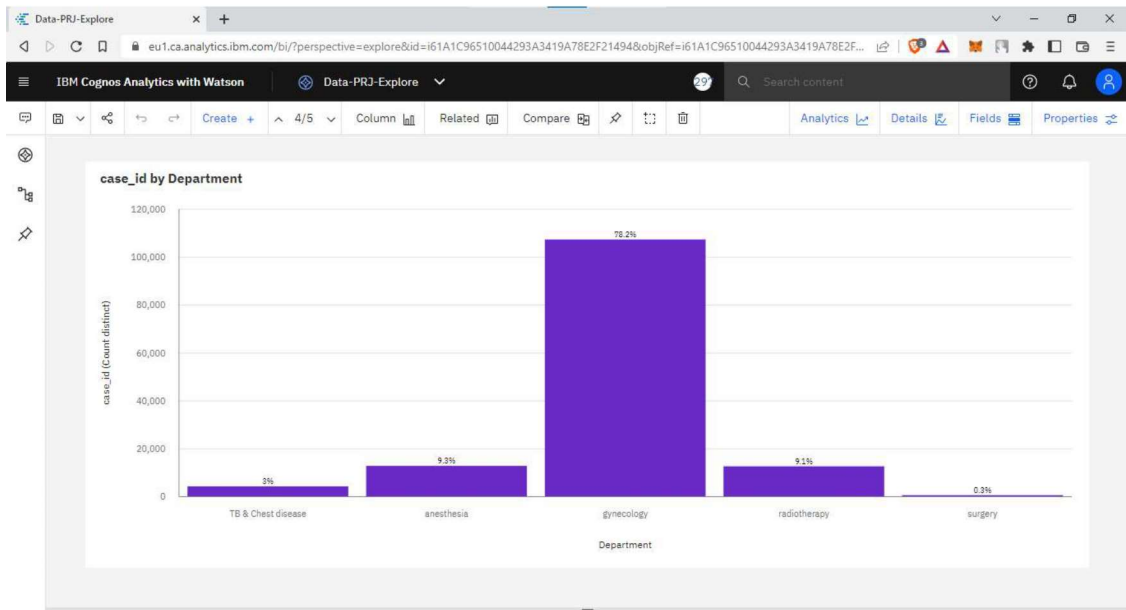
Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022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

7. Results

7.1 Virtualizations







8. Advantages and Disadvantages

Advantages

- **Data analytics helps an organization make better decisions**

Lot of times decisions within organizations are made more on gut feel rather than facts and data. One of the reasons for this could be lack of access to quality data that can help with better decision making. Analytics can help with transforming the data that is available into valuable information for executives so that better decisions can be made. This can be a source of competitive advantage if fewer poor decisions are made since poor decisions can have a negative impact on a number of areas including company growth and profitability.

- **Increase the efficiency of the work**

Analytics can help analyse large amounts of data quickly and display it in a formulated manner to help achieve specific organizational goals. It encourages a culture of efficiency and teamwork by allowing the managers to share the insights from the analytics results to the employees. The gaps and improvement areas within a company become evident and actions can be taken to increase the overall efficiency of the workplace thereby increasing productivity.

- **The analytics keeps you updated of your customer behavioural changes**

In today's world, customers have a lot of choices. If organizations are not tuned to customer desires and expectations, they can soon find themselves in a downward spiral. Customers tend to change their minds as they are continuously exposed to new information in this era of digitization. With vast amount of customer data, it is practically impossible for organizations to make senses of all the changes in customer perception data without using the power of analytics. Analytics gives you insights into how your target market thinks and if there is any change. Hence, being aware of shift in customer behaviour can provide a decisive advantage to companies so that they can react faster to the market changes.

- **Personalization of products and services**

Gone are the days where a company could sell a standard set of products and services to customers. Customers crave products and services that can meet their individual needs. Analytics can help companies keep track of what kind of service, product, or content is preferred by the customer and then show the recommendations based on their preferences. For example, in social media, we usually see what we like to see, all of this is made possible due to the data collection and analytics that companies do. Data analytics can help provide targeted services to customers based on their individual requirements.

- **Improving quality of products and services**

Data analytics can help with enhancing the user experience by detecting and correcting errors or avoiding non-value-added tasks. For example, self-learning systems can use data to understand the way customers are interacting with the tools and make appropriate changes to improve user experience. In addition, data analytics can help with automated data cleansing and improving the quality of data and consecutively benefiting both customers and organizations.

Disadvantages

- **Lack of alignment within teams**

There is a lack of alignment between different teams or departments within an organization. Data analytics may be done by a select set of team members and the analysis done may be

shared with a limited set of executives. However, the insights generated by these teams are either of not much value or are having limited impact on organizational metrics. This could be due to a “silos” way of working with each team only using their existing processes disconnected from other departments. The analytics team should be focussed on answering the right questions for the business and the results generated by data analytics teams needs to be properly communicated to the right employees to drive the right set of actions and behaviours so that it can have an positive impact on the organization.

- **Lack of commitment and patience**

Analytics solutions are not difficult to implement, however, they are costly, and the ROI is not immediate. Especially, if existing data is not available, it may take time to put processes and procedures in place to start collecting the data. By nature, the analytics models improve accuracy over time and require dedication to implement the solution. Since the business users do not see results immediately, they sometimes lose interest which results in loss of trust and the models fail. When an organization decides to implement data analytics methods, there needs to be a feedback loop and mechanism in place to understand what is working and what is not, and corrective actions are required to fix things that are broken. Without this closed loop system, senior management may decide that analytics is not working or much valuable and may abandon the entire exercise.

- **Low quality of data**

One of the biggest limitations of data analytics is lack of access to quality data. It is possible that companies already have access to a lot of data, but the question is do they have the right data that they need? A top down approach is required where the business questions that need to be answered need to be known first and what data is required to answer these questions can then be determined. In some cases, data may have been collected for historical reasons may not be suitable to answer the questions that we ask today. At other times, even though we have the right metrics that we are collecting data on, the quality of the data collection may be poor. There can be instances where adequate data is not available or is missing for proper analytics to be done. As they say, garbage-in garbage-out. If the data quality is poor, the decision made by using this data is also going to be poor. Hence, actions must be taken to fix the quality of the data before it can be effectively used within organizations.

- **Privacy concerns**

Sometimes, data collection might breach the privacy of the customers as their information such as purchases, online transactions, and subscriptions are available to companies whose services they are using. Some companies might exchange those datasets with other companies for mutual benefit. Certain data collected can also be used against a person, country, or community. Organizations need to be cautious of what sort of data they are collecting from customers and ensure the security and confidentiality of the data. Only the data required for the analysis needs to be captured and if there is sensitive data, it needs to be anonymized so that sensitive data is protected. Data breaches can cause customers to lose trust in the organizations which may result in a negative impact on the organization.

- **Complexity & Bias**

Some of the analytics tools developed by companies are more like a black box model. What

is inside the black box is not clear or the logic the system uses to learn from data and create a model is not readily evident. For example, a neural network model that learns from various scenarios to decide who should be given a loan and who should be rejected. The usage of these tools may be easy but the logic of how decisions are made is not clear to anyone within the company. If companies are not careful and a poor quality data set is used to train the model, there may be hidden biases in the decisions made by these systems which may not be readily evident and organizations may be breaking the law by discriminating against race, gender, sex, age etc.

9. Conclusion

Data analysis is an important process of research or simply discovering information related to any work. Data derived from the observation, experiment, and other primary and secondary data collection methods is large and cannot be taken as it is. Not all data is relevant, neither can it directly signify any trends, relations, facts, and associations within the data. To find out those required trends and relations, the data needs to be reconstructed in the relevant form and modified. This process is called data analysis. Data analysis and conclusion take forward the research.

10. Future Scope

The future scope of data analytics in India is bright due to several reasons. A career in business analytics is rewarding and offers a wide range of personal and professional development opportunities. Understanding statistical approaches, mathematical capability, business learning, logical thinking and Big Data are just a few of the natural talents needed to be a Data Analytics professional. The ability to analyze business circumstances and develop creative solutions is also necessary.

In India, the widespread usage of Big Data assures a high level of employment, better wages, and rising demand in future. Analytics has the potential to transform the current business environment by capturing a significant amount of data, expanding business models, igniting creative processes, and promoting the overall growth and development of a company confirming why the future scope of data analytics in India presents brilliant career options.

11. Appendix

11.1 GitHub Project and Demo Link

- GitHub : <https://github.com/IBM-EPBL/IBM-Project-11497-1659331419>
- Demo : [VID-20221118-WA0002.mp4 - Google Drive](#)

