PROJECT REPORT

TEAM ID	**PNT2022TMID23641**
PROJECT NAME	**ANAYTICS FOR HOSPITALS HEALTH -CARE DATA**

TEAM MEMBERS:

Gokula priya S (TL) - 612919104021

Keerthana D (TM1) - 612919104028

Kowsalya K (TM2) - 612919104030

Yuvanthi S (TM3) - 612919104059

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 well-informed 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

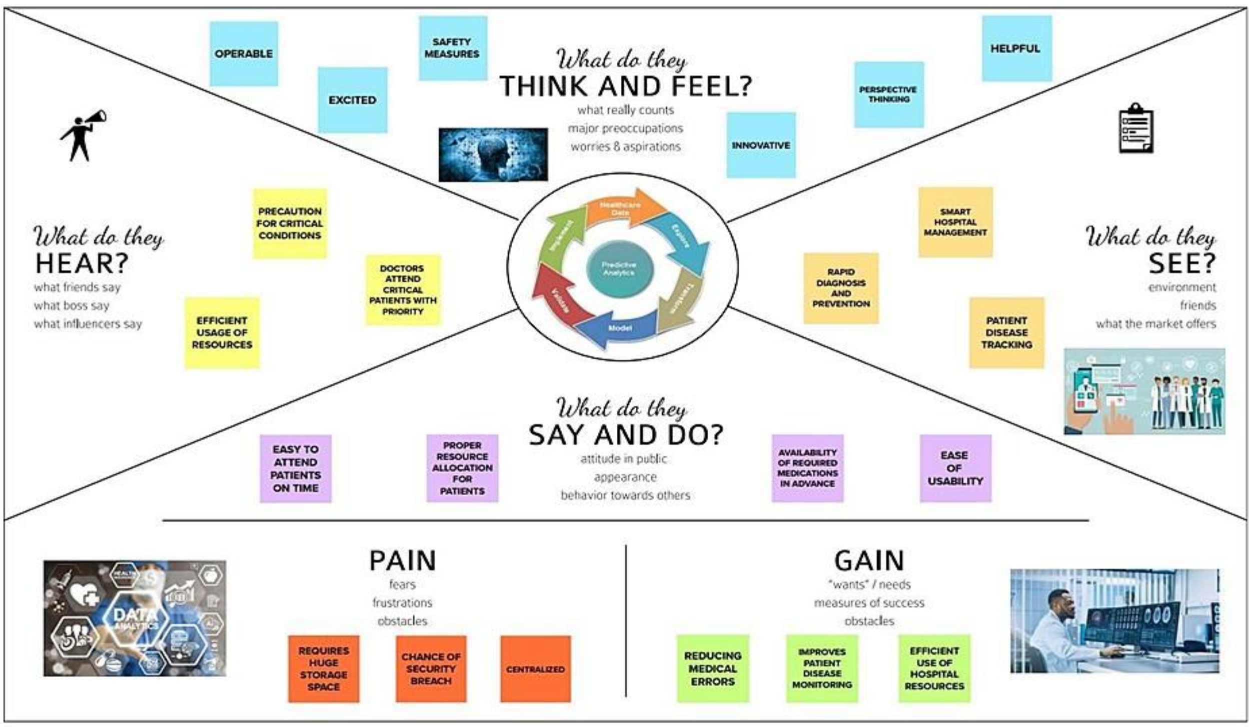
2.3 PROBLEM STATEMENT AND DEFINITION

- 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.

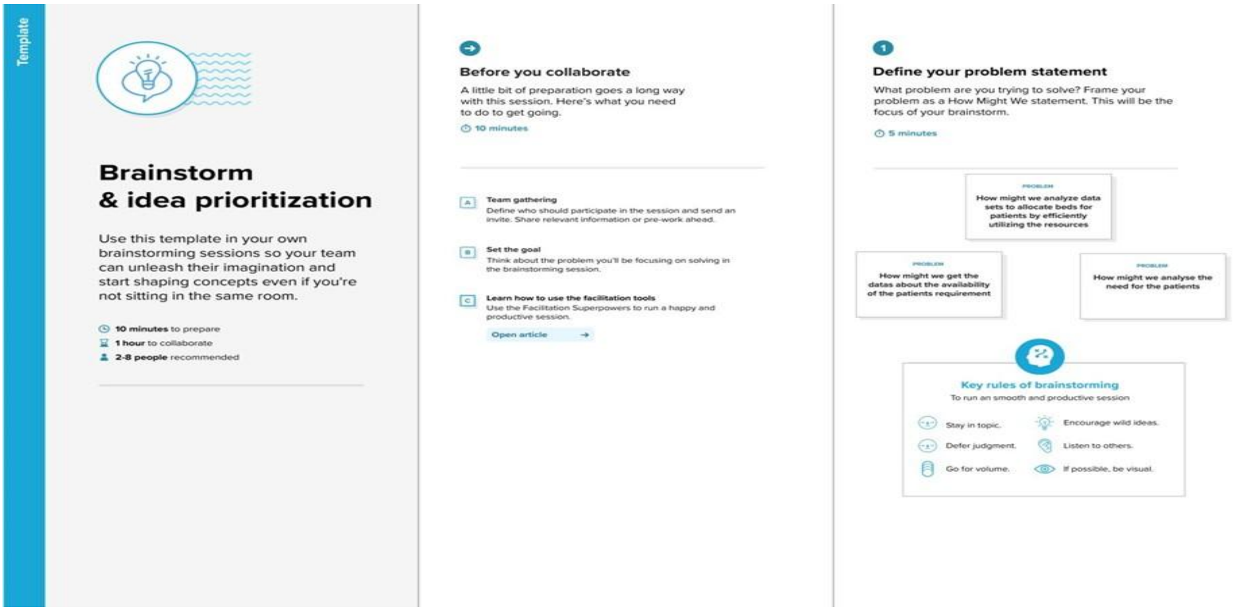
CHAPTER 3

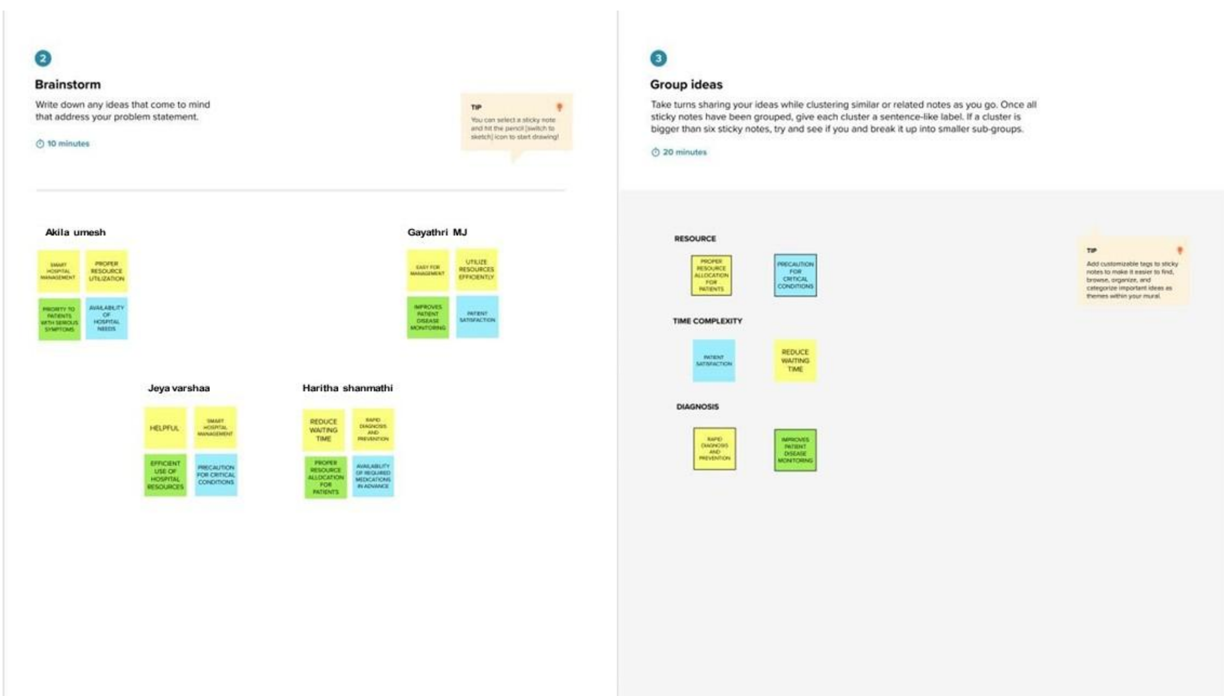
IDEATION & PROPOSED SOLUTION

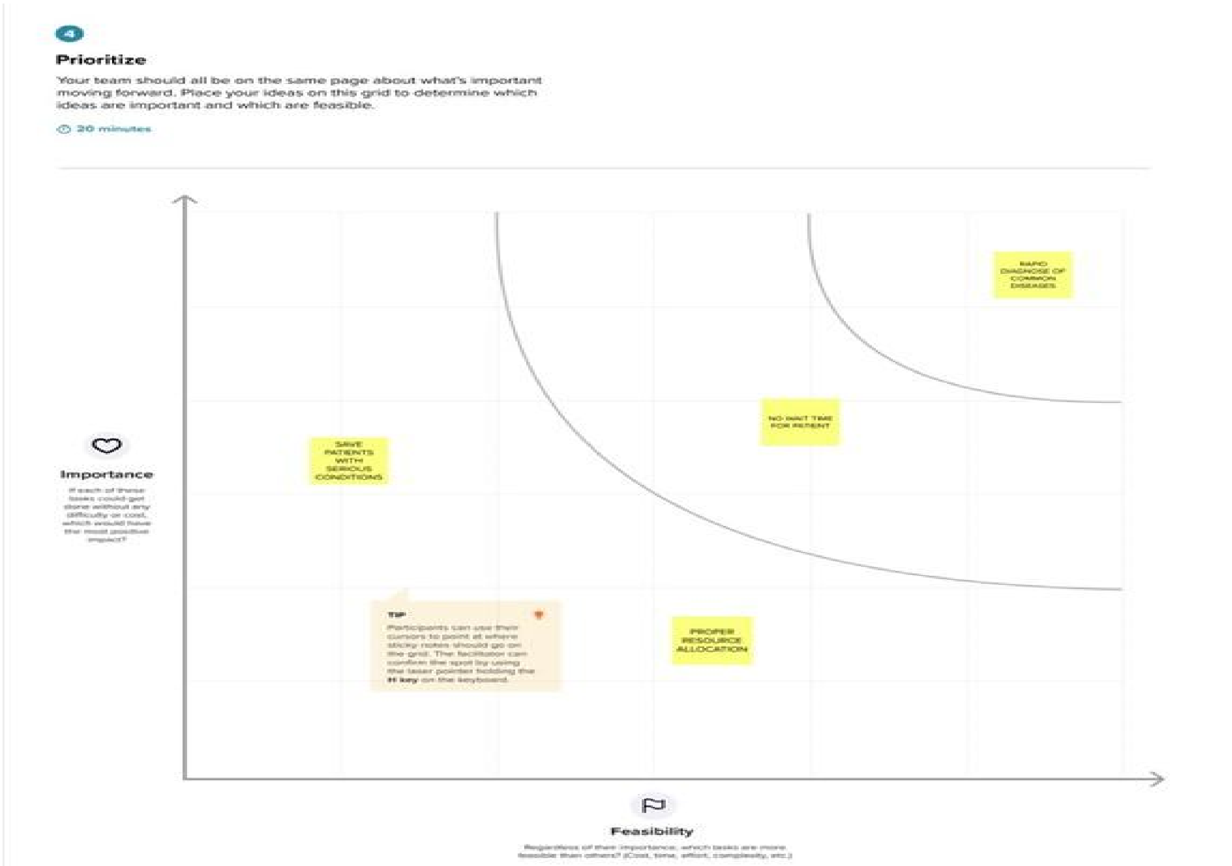
3.1 EMPATHY MAP CAMPUS



3.2 IDEATION & BRAINSTORMING







3.3 PROPOSED 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 patients with more accuracy. As a result proper resources and therapy can be provided.Patients can get proper treatment and better medical care than before which helps them for their faster recovery. So the prediction minimizes the overflow of patients and helps in resource management and optimize their resource utilization. Hence this leads to faster recovery and lower the expenses for treatment. It improves the trust in hospital management.

It avoids the major risk of spreading infection among the hospital staff. This leads to overall safety of hospital staff and patients. Resource consumption is optimized. This model can be used by all government hospitals, private hospitals, and this model is also trained with the real world hospital survey for better prediction small clinics.

Length of the stay will be predicted with more accuracy. This model predicts the length of the stay for all kinds of patients and predicts with more accuracy.

3.4 PROBLEM SOLUTION FIT

1. CUSTOMER SEGMENT(S) ➢ Patients ➢ Hospital Management	**6. CUSTOMER STATE LIMITATIONS** ➢ Inadequate information about availability of required resource	**5. AVAILABLE SOLUTIONS** ➢ Tableau cloud ➢ Text Mining ➢ Information Retrieval
2. PROBLEMS / PAINS ➢ Effective Resource allocation ➢ Reduce Waiting time for patients in Hospitals	**9. ROOT / CAUSE of every problem** ➢ No proper system or less efficient Prediction System	**7. BEHAVIOR** Tracking the information with the available Technologies
3. TRIGGERS TO ACT ➢ Covid-19 Pandemic ➢ Emergency Situations **4. EMOTIONS** ➢ BEFORE: Feeling bad & Frustrated ➢ AFTER: Feeling better &Relaxed	**10. YOUR SOLUTION** Existing: ratio of discharges in given period of time to no. of beds in hospital during the time period Proposing: Using predictive analysis powered by AI	**8. CHANNELS OF BEHAVIOR** ONLINE: Use of data from all region(data Exploration) OFFLINE: Use of data Collect from nearby facilities

CHAPTER 4

REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENT

FR No.	Functional Requirement (Epic)	Sub Requirement (Story/ Sub-Task)
FR-1	**User Registration**	Registration throughForm Registration throughGmail
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.

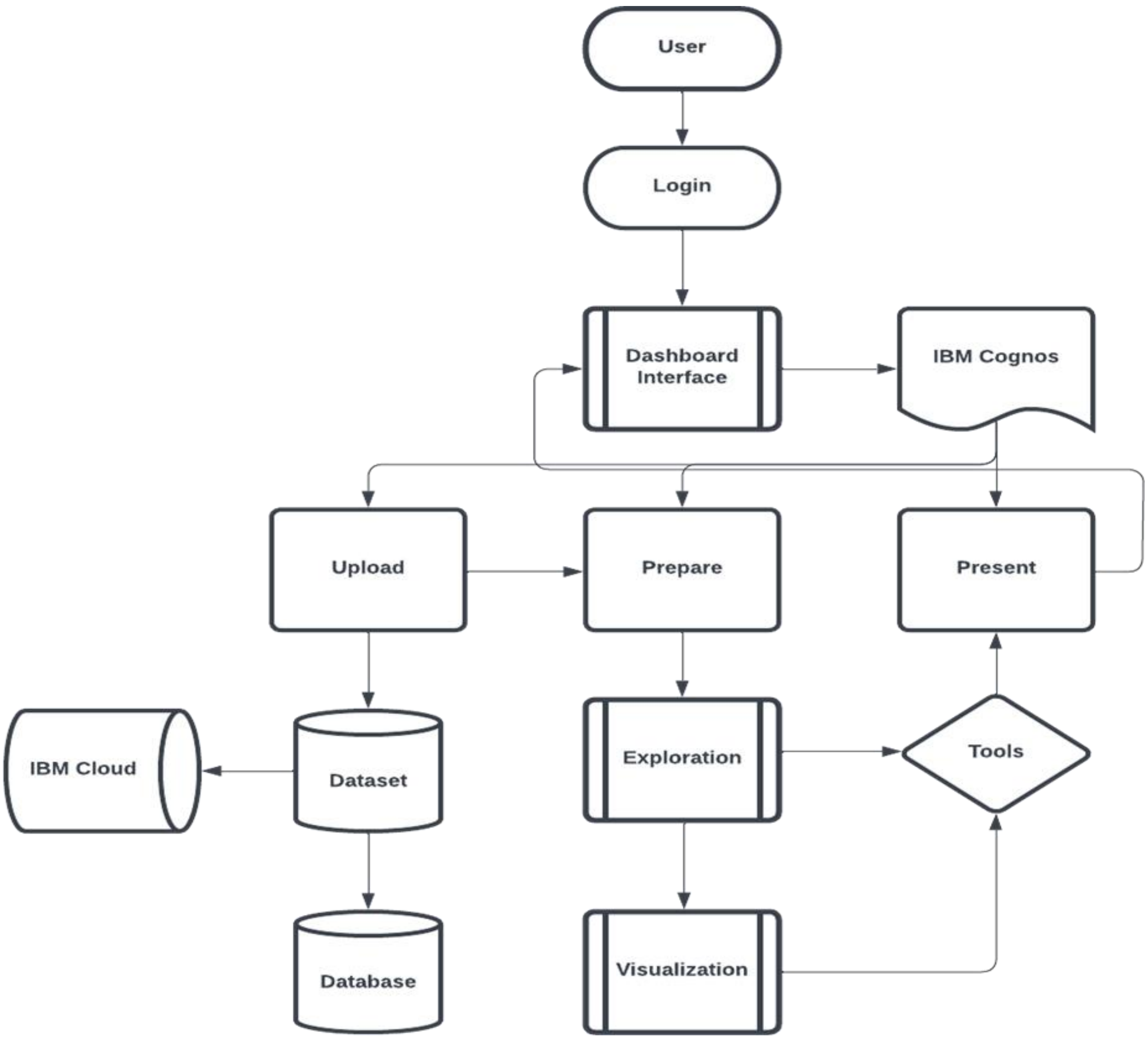
4.2 NON FUNCTIONAL REQUIREMENT

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	This Dashboards are designed to offer a comprehensive overview of patient's LOS, and do sothrough the use of data visualization tools like chartsand 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, efficientand reliable manner.
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.

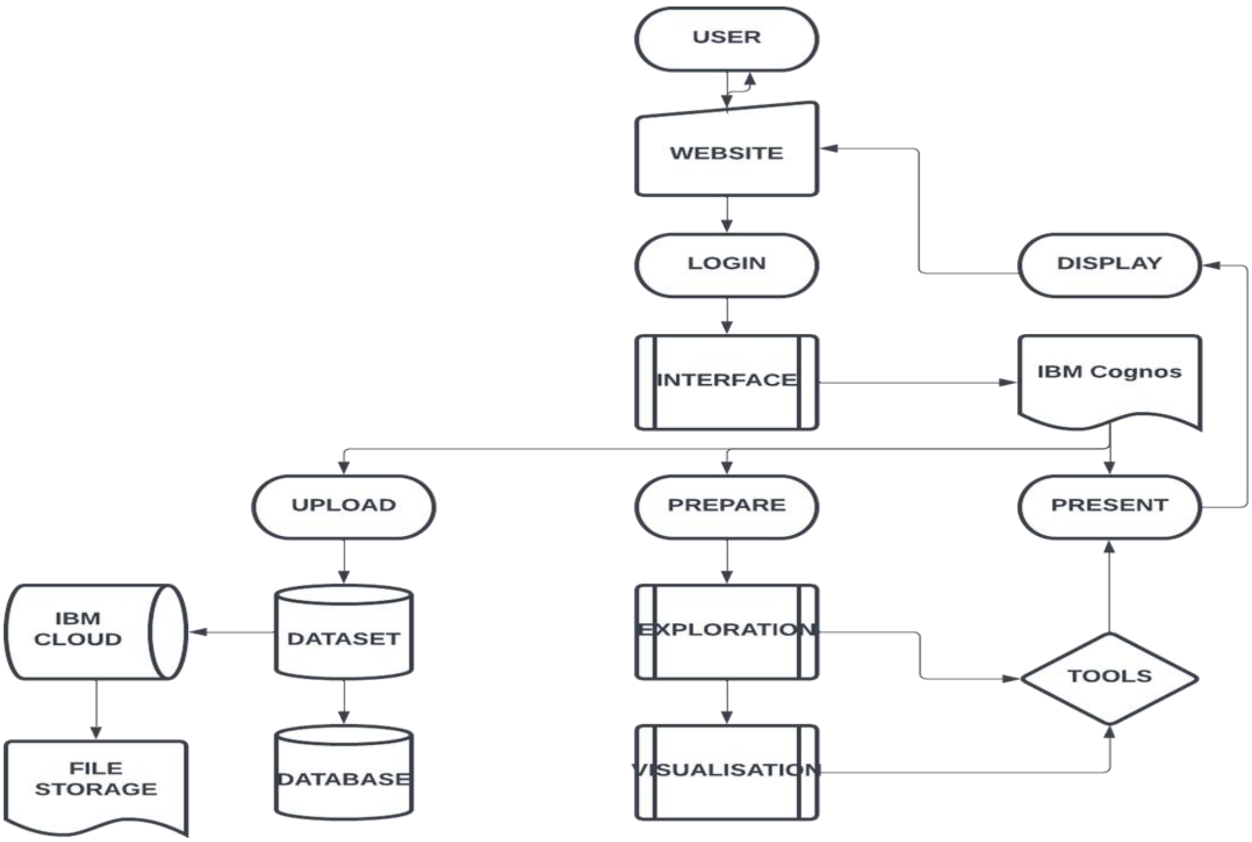
CHAPTER 5

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	Registration	USN-1	As a user i an login to my	I can access	High	Sprint-1

dashboard

dashboard

	Collec t data	USN-1	As a user i can provide my details	I can view my data	Medium	Sprint-1
Admin	Collec t 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
	Uploa d 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	Prepar e 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

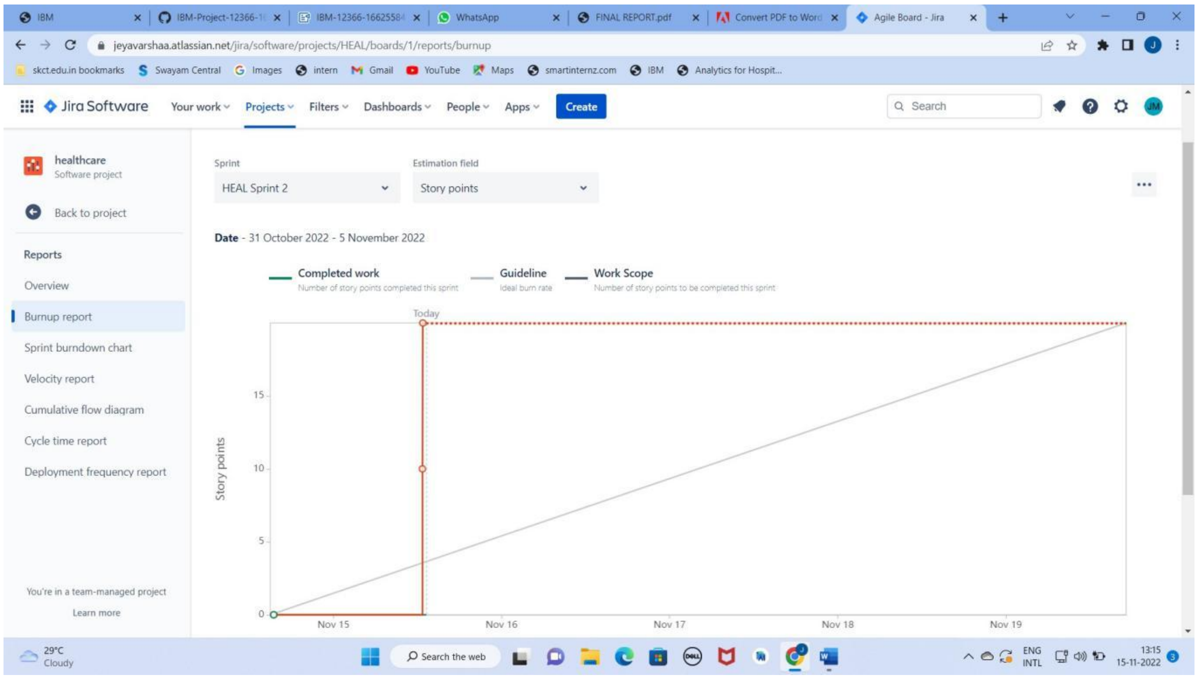
Sprint	Functional Requirement (Epic)	User Story Number	User Story/ Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a health care provider I can create account inIBM cloud and the data are collected.	20	High	Gokula priya S Keerthana D Kowsalya K Yuvanthi S
sprint-2	Analyze	USN-2	As a healthcare provider all data are collected is cleaned and uploaded in the database or IBM cloud	20	medium	Gokula priya S Keerthana D Kowsalya K Yuvanthi S

6.2 SPRINT DELIVERY SCHEDULE

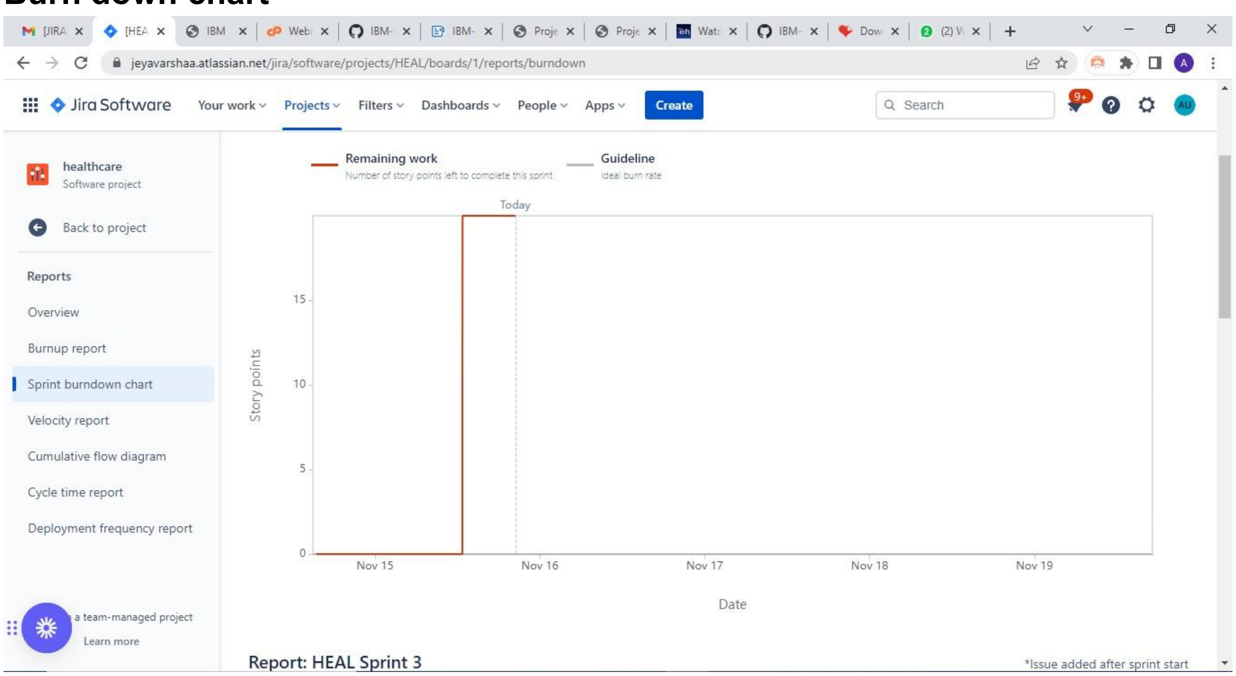
Sprint	Functional Requirement (Epic)	User Story Number	User Story/ Task	Story Points	Priority	Team Members
sprint-3	Dashbord	USN-3	As a healthcare provider I can use my account in my dashboard for uploading dataset	10	medium	Gokula priya S Keerthana D Kowsalya K Yuvanthi S
Sprint-3	Visualization	USN-4	As a health care provider I can prepare data forVisualization.	10	High	Gokula priya S Keerthana D Kowsalya K Yuvanthi S
Sprint-4	Visualization	USN-5	As a health care provider I can present data in my dashboard.	10	High	Gokula priya S Keerthana D Kowsalya K Yuvanthi S
Sprint-4	Prediction	USN-6	As a health care provider I can predict the length of stay	10	High	Gokula priya S Keerthana D Kowsalya K Yuvanthi S

6.3 REPORTS FROM JIRA

Burnup chart

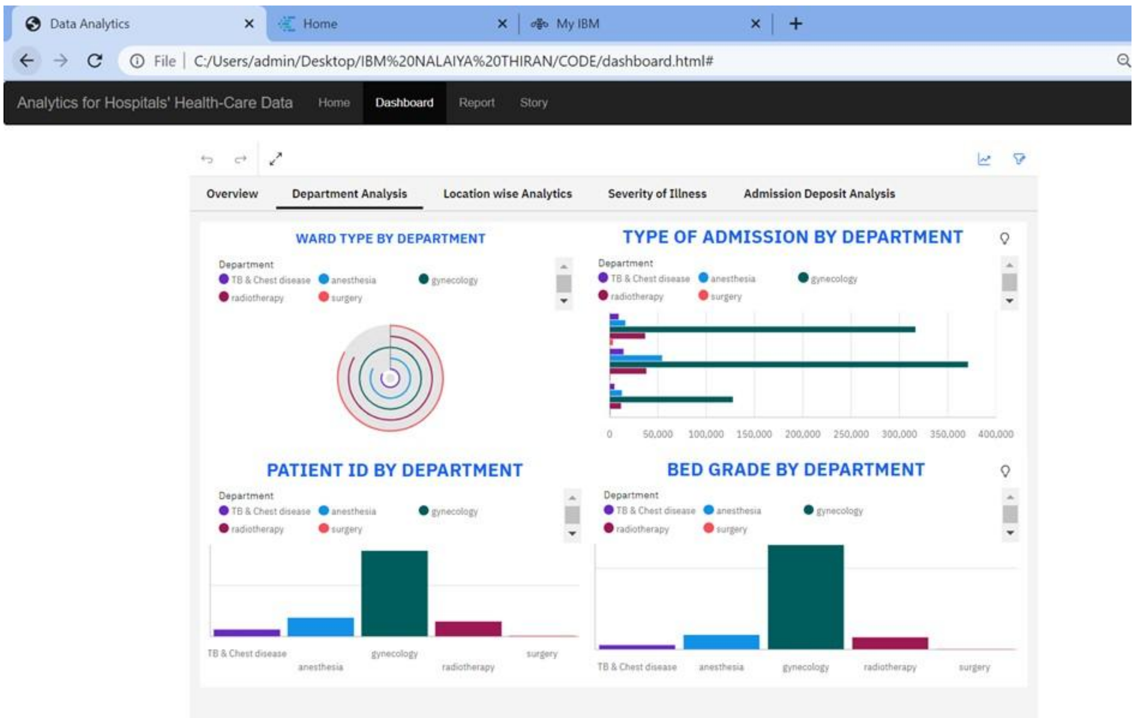


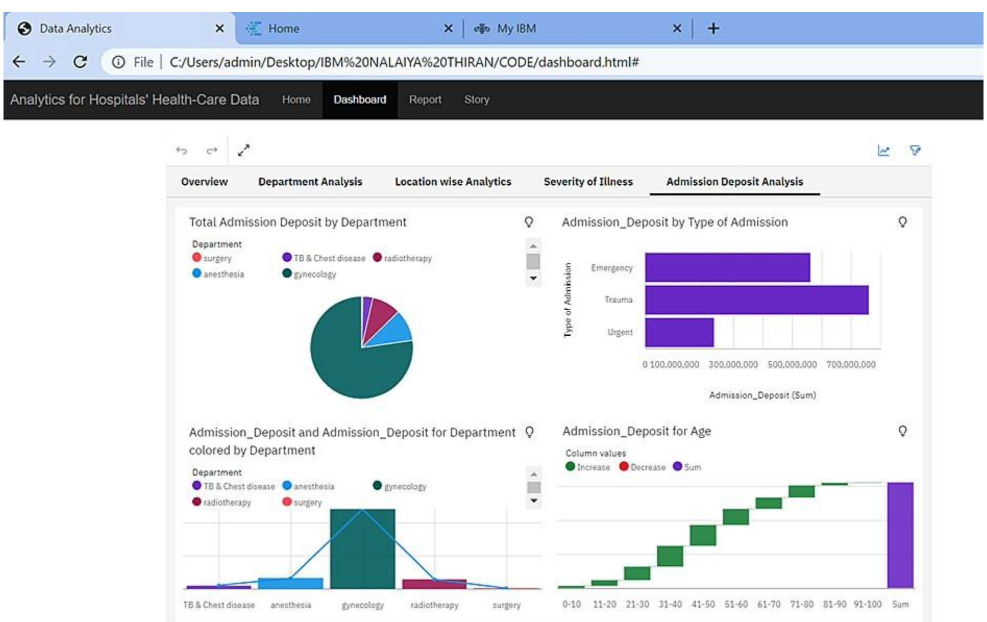
Burn down chart

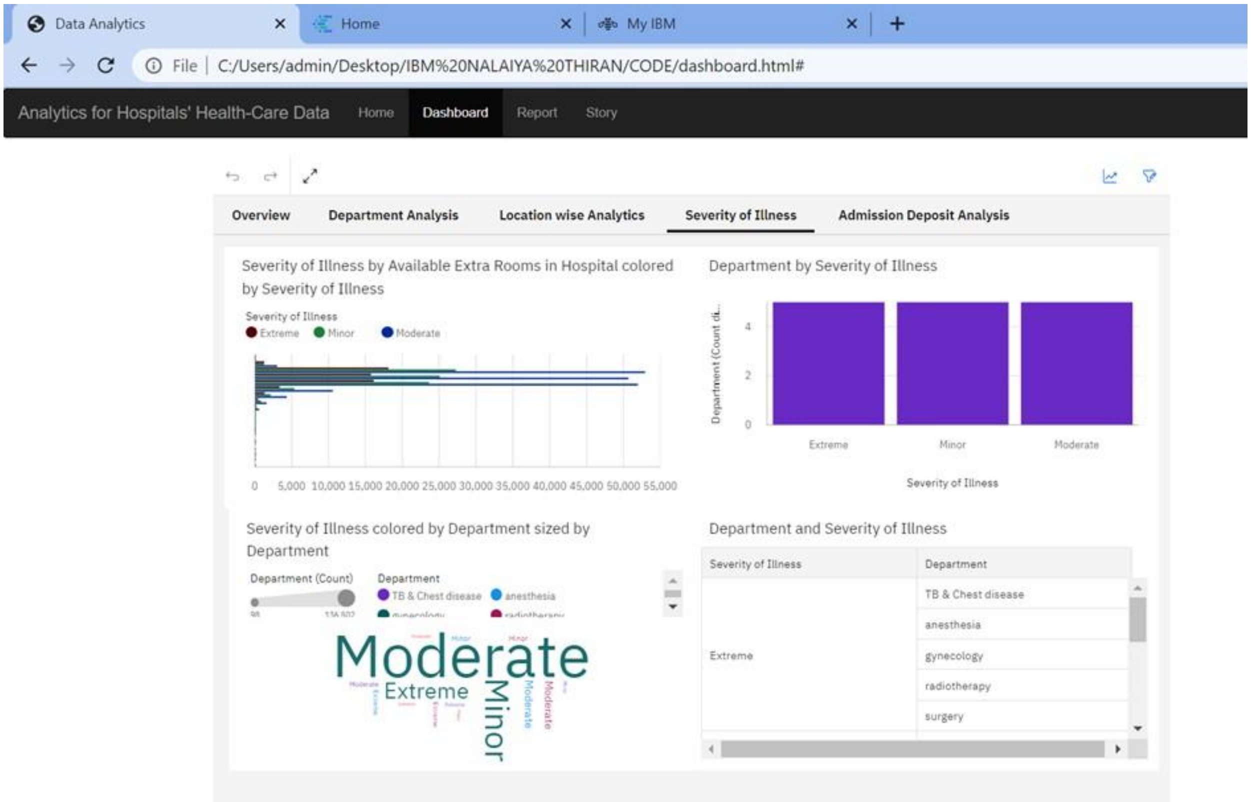


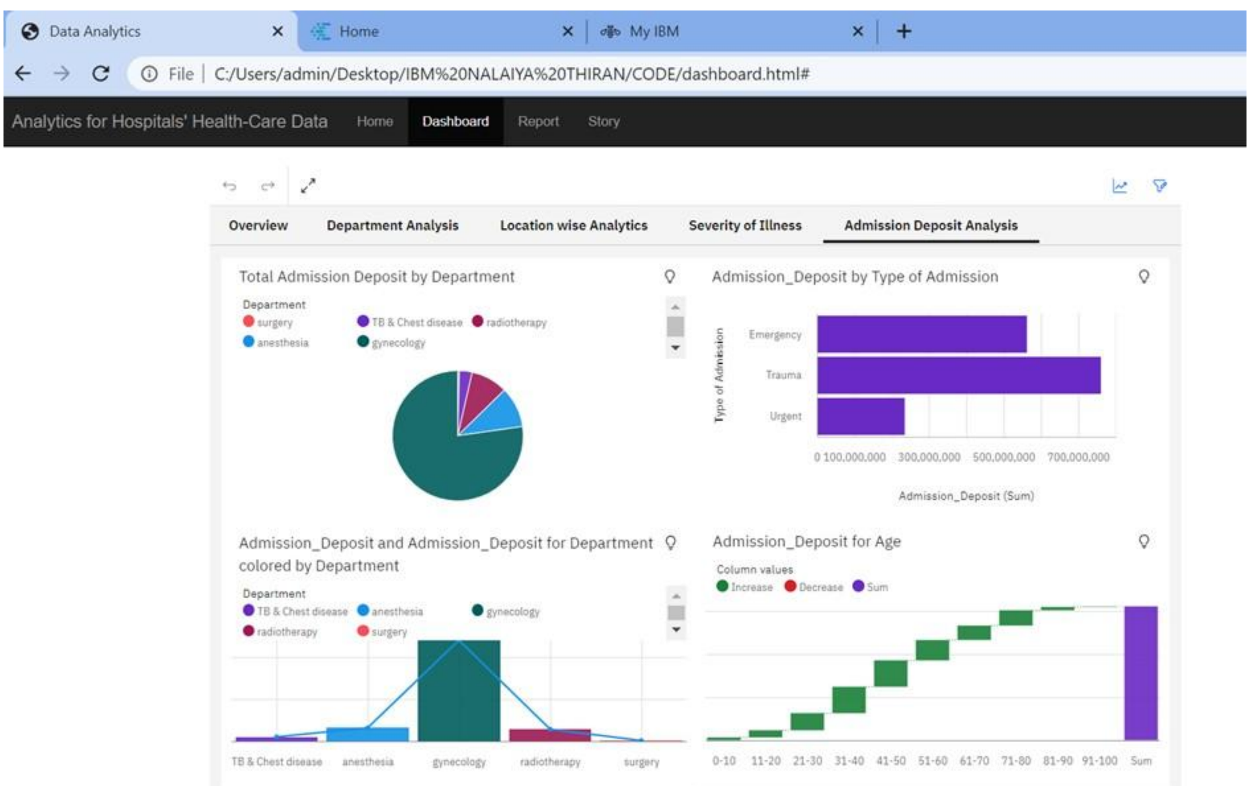
CHAPTER 7
CODING & SOLUTIONING

7.1 FEATURE 1

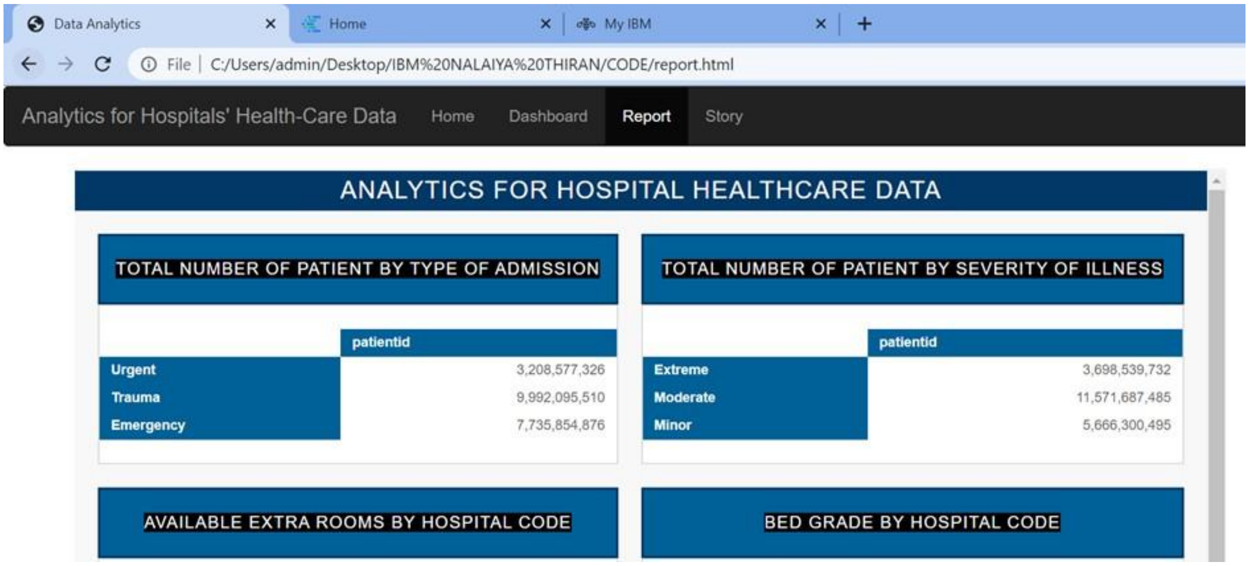


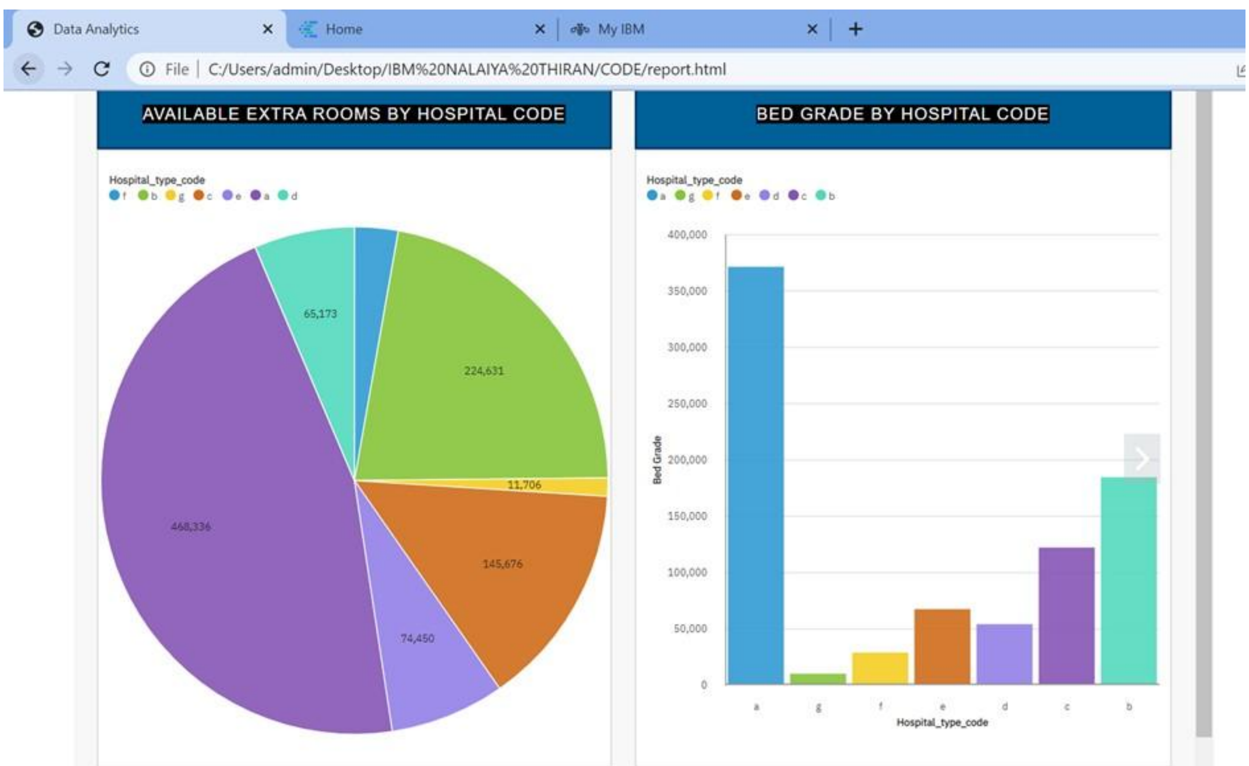


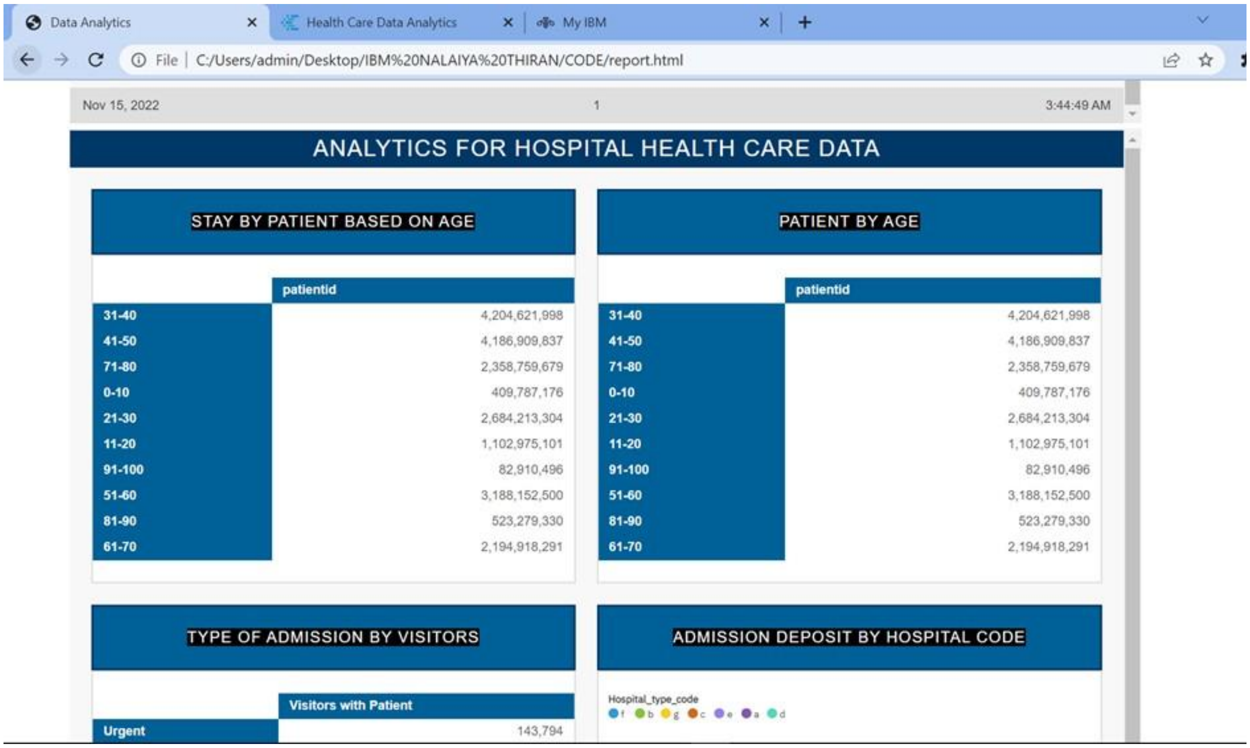

Data Analytics
Home
My IBM
File | C:/Users/admin/Desktop/IBM%20NALAIYA%20THIRAN/CODE/dashboard.html#
Analytics for Hospitals' Health-Care Data
Home
Dashboard
Report
Story
Overview
Department Analysis
Location wise Analytics
Severity of Illness
Admission Deposit Analysis
Severity of Illness by Available Extra Rooms in Hospital colored by Severity of Illness
Severity of Illness
Extreme
Minor
Moderate
0 5,000 10,000 15,000 20,000 25,000 30,000 35,000 40,000 45,000 50,000 55,000
Department by Severity of Illness
Extreme
Minor
Moderate
Severity of Illness
Severity of Illness colored by Department sized by Department
Department (Count)
Department
TB & Chest disease
anesthesia
Moderate
Extreme
Minor
Department and Severity of Illness
Severity of Illness
Department
Extreme
TB & Chest disease
anesthesia
gynecology
radiotherapy
surgery


Data Analytics
Home
My IBM
File | C:/Users/admin/Desktop/IBM%20NALAIYA%20THIRAN/CODE/dashboard.html#
Analytics for Hospitals' Health-Care Data
Home
Dashboard
Report
Story
Overview
Department Analysis
Location wise Analytics
Severity of Illness
Admission Deposit Analysis
Total Admission Deposit by Department
Department
surgery
TB & Chest disease
radiotherapy
anesthesia
gynecology
Admission_Deposit by Type of Admission
Type of Admission
Emergency
Trauma
Urgent
0 100,000,000 300,000,000 500,000,000 700,000,000
Admission_Deposit (Sum)
Admission_Deposit and Admission_Deposit for Department colored by Department
Department
TB & Chest disease
anesthesia
gynecology
radiotherapy
surgery
TB & Chest disease
anesthesia
gynecology
radiotherapy
surgery
Admission_Deposit for Age
Column values
Increase
Decrease
Sum
0-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
91-100
Sum

7.2 FEATURE 2


Data Analytics
Home
My IBM
File | C:/Users/admin/Desktop/IBM%20NALAIYA%20THIRAN/CODE/report.html
Analytics for Hospitals' Health-Care Data
Home
Dashboard
Report
Story
ANALYTICS FOR HOSPITAL HEALTHCARE DATA
TOTAL NUMBER OF PATIENT BY TYPE OF ADMISSION
patientid
Urgent 3,208,577,326
Trauma 9,992,095,510
Emergency 7,735,854,876
TOTAL NUMBER OF PATIENT BY SEVERITY OF ILLNESS
patientid
Extreme 3,698,539,732
Moderate 11,571,687,485
Minor 5,666,300,495
AVAILABLE EXTRA ROOMS BY HOSPITAL CODE
BED GRADE BY HOSPITAL CODE


Data Analytics
Home
My IBM
File | C:/Users/admin/Desktop/IBM%20NALAIYA%20THIRAN/CODE/report.html
AVAILABLE EXTRA ROOMS BY HOSPITAL CODE
Hospital_type_code
f b g c e a d
65,173
224,631
11,706
145,676
74,450
468,336
BED GRADE BY HOSPITAL CODE
Hospital_type_code
a g f e d c b
400,000
350,000
300,000
250,000
200,000
150,000
100,000
50,000
0
Bed Grade
a g f e d c b
Hospital_type_code



ANALYTICS FOR HOSPITAL HEALTH CARE DATA

STAY BY PATIENT BASED ON AGE

	patientid
31-40	4,204,621,998
41-50	4,186,909,837
71-80	2,358,759,679
0-10	409,787,176
21-30	2,684,213,304
11-20	1,102,975,101
91-100	82,910,496
51-60	3,188,152,500
81-90	523,279,330
61-70	2,194,918,291

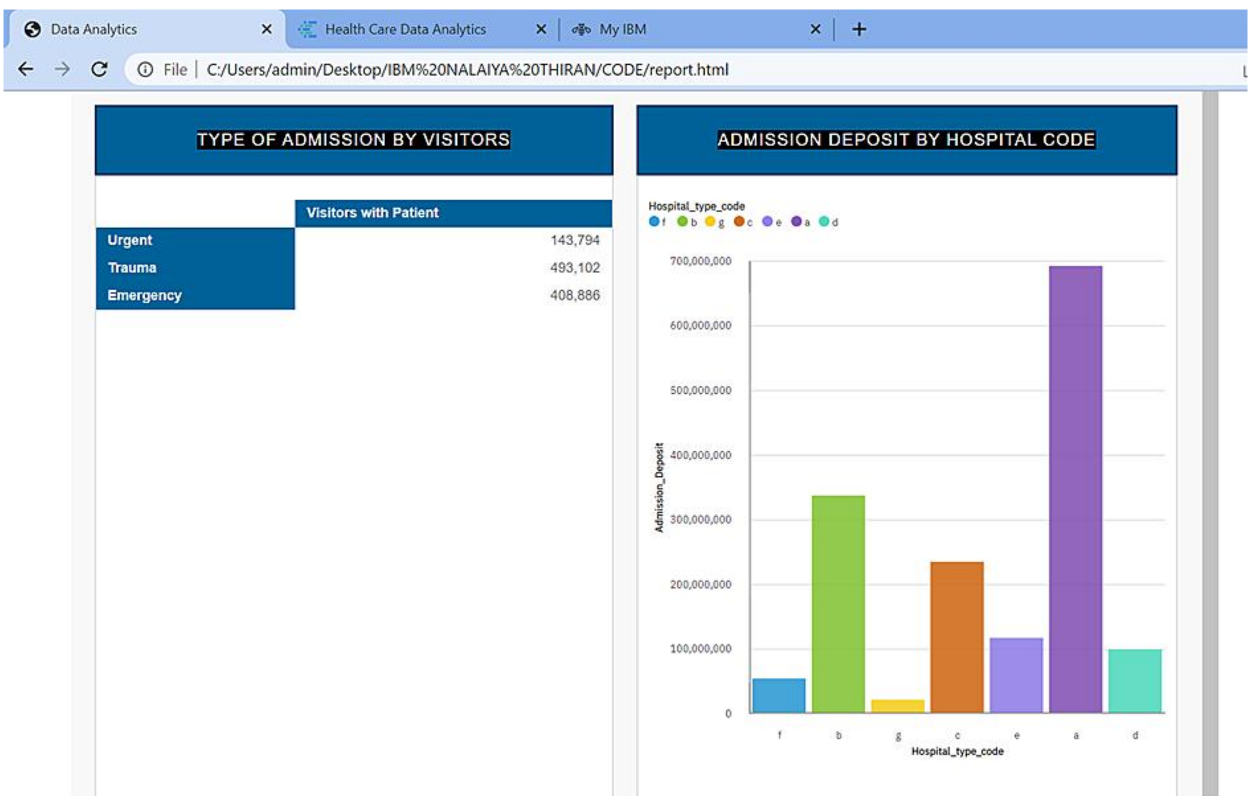
PATIENT BY AGE

	patientid
31-40	4,204,621,998
41-50	4,186,909,837
71-80	2,358,759,679
0-10	409,787,176
21-30	2,684,213,304
11-20	1,102,975,101
91-100	82,910,496
51-60	3,188,152,500
81-90	523,279,330
61-70	2,194,918,291

TYPE OF ADMISSION BY VISITORS

	Visitors with Patient
Urgent	143,794

ADMISSION DEPOSIT BY HOSPITAL CODE



TYPE OF ADMISSION BY VISITORS

	Visitors with Patient
Urgent	143,794
Trauma	493,102
Emergency	408,886

ADMISSION DEPOSIT BY HOSPITAL CODE

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

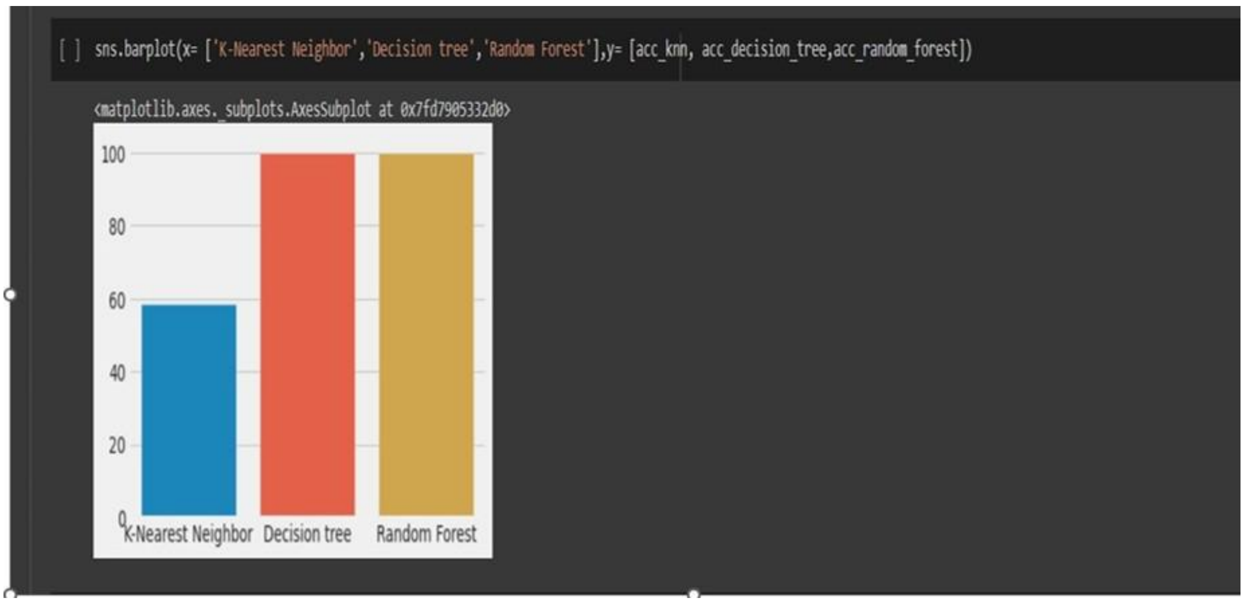
This report shows the number of test cases that have passed, failed, and untested

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

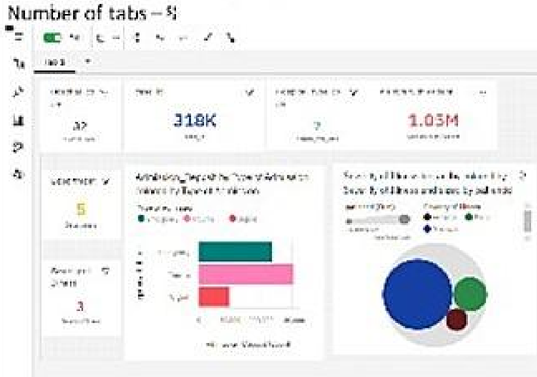
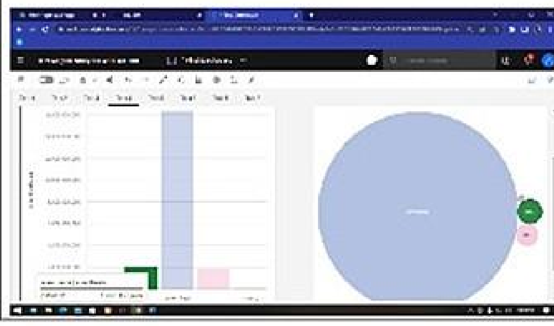
RESULTS

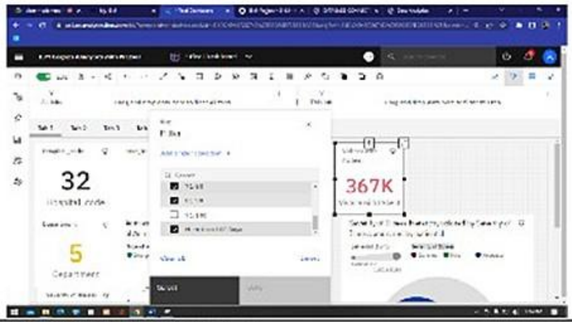
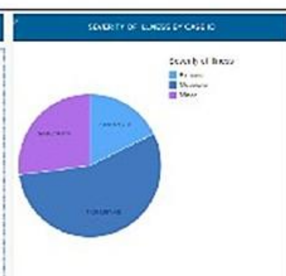
9.1 PERFORMANCE METRICS



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 – 8 
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

CONCLUSION

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

SOURCE CODE

HOME PAGE :

```
<!DOCTYPE html>
<html lang="en">
<head>
  <title>Data Analytics</title>
  <meta charset="utf-8">
  <meta name="viewport" content="width=device-width, initial-scale=1">
  <link rel="stylesheet"
href="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/css/bootstrap.min.css">
  <script src="https://ajax.googleapis.com/ajax/libs/jquery/3.6.0/jquery.min.js"></script>
  <script src="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/js/bootstrap.min.js"></script>
</head>
<body>

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

```
</nav>

<div class="jumbotron">
 <center> <h4><i><b>Team ID : PNT2022TMID20389 </b></i></h4></center>

</div>
<table class="table table-bordered">

    <tbody>
      <tr>
        <td>Team Leader</td>
        <td> Gokula priya S</td>

      </tr>
      <tr>
        <td>Team member</td>
        <td> Keerthana D</td>

      </tr>
      <tr>
        <td>Team member</td>
        <td>Kowsalya K</td>

      </tr>
    <tr>
        <td>Team member</td>
        <td>Yuvanthi S</td>

    </tr>

    </tbody>
  </table>
</body>
</html>
```

About Page :

```
<!DOCTYPEhtml>
<html lang="en">
<head>
  <title>Data Analytics</title>
  <meta charset="utf-8">
  <meta name="viewport"content="width=device-width, initial-scale=1">
  <link rel="stylesheet"href="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/css/bootstrap.min.css">
  <script src="https://ajax.googleapis.com/ajax/libs/jquery/3.6.0/jquery.min.js"></script>
  <script src="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/js/bootstrap.min.js"></script>
</head>
<body>

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


<div class="container">
<b>Analytics For Hospitals' Health-Care Data</b>
<br>

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

<br><br>
<b>Goal:</b>
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.

<br><br>

<b>Technical Architecture:</b>
<br>
<img src="pic.JPG">




</div>
</body>
</html>
```

DASHBOARD PAGE :

```
<!DOCTYPEhtml>
<html lang="en">
<head>
<title>Data Analytics</title>
  <meta charset="utf-8">
  <meta name="viewport"content="width=device-width, initial-scale=1">
  <link rel="stylesheet"href="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/css/bootstrap.min.css">
  <script src="https://ajax.googleapis.com/ajax/libs/jquery/3.6.0/jquery.min.js"></script>
  <script src="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/js/bootstrap.min.js"></script>
</head>
<body>
```

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

<div class="container">
  <iframe
src="https://us1.ca.analytics.ibm.com/bi/?perspective=dashboard&amp;pathRef=.my_folders%2FDashboard%
2FHealth%2BCare%2BData%2BAnalytics&amp;closeWindowOnLastView=true&amp;ui_appbar=false&amp
;ui_navbar=false&amp;shareMode=embedded&amp;action=view&amp;mode=dashboard&amp;subView=mo
del00000184765 84e12_00000000"width="1100" height="600"frameborder="0"gesture="media"
allow="encrypted-media"allowfullscreen=""></iframe>
</div>

</body>
</html>
```

REPORT PAGE :

```
<!DOCTYPEhtml>
<html lang="en">
<head>
 <title>Data Analytics</title>
  <meta charset="utf-8">
  <meta name="viewport"content="width=device-width, initial-scale=1">
  <link rel="stylesheet"href="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/css/bootstrap.min.css">
  <script src="https://ajax.googleapis.com/ajax/libs/jquery/3.6.0/jquery.min.js"></script>
  <script src="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/js/bootstrap.min.js"></script>
```

```
</head>
<body>

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

<div class="container">
  <iframe
src="https://us1.ca.analytics.ibm.com/bi/?pathRef=.my_folders%2FReport%2FHealth%2BCare%2BData%2BAnalytics%2BReport&amp;closeWindowOnLastView=true&amp;ui_appbar=false&amp;ui_navbar=false&amp;shareMode=embedded&amp;action=run&amp;format=HTML&amp;prompt=false"width="1000"
height="900"frameborder="0"gesture="media" allow="encrypted-media"allowfullscreen=""></iframe>
</br>

</div>

</body>
</html>
```

STORY :

```
<!DOCTYPEhtml>
<html lang="en">
<head>
 <title>Data Analytics</title>
  <meta charset="utf-8">
  <meta name="viewport"content="width=device-width, initial-scale=1">
```

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

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

    </ul>
  </div>
</nav>

<div class="container">
  <iframe
src="https://us1.ca.analytics.ibm.com/bi/?perspective=story&amp;pathRef=.my_folders%2FStory%2FHealth
%2Bcare%2Bdata%2Banalytics%2Bstory&amp;closeWindowOnLastView=true&amp;ui_appbar=false&amp;
ui_navbar=false&amp;shareMode=embedded&amp;action=view&amp;sceneId=model000001847a5e7043_00
000001&amp;sceneTime=0" width="1000"  height="900" frameborder="0" gesture="media"  allow="encrypted-
media" allowfullscreen=""></iframe>
</br>

</div>

</body>
</html>
```
-

Importing required Packages

In [72]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
sns.set_style("darkgrid")
plt.style.use("dark_background")
```

Importing the dataset

In [73]:

```
train = pd.read_csv('/content/input/training_data.csv')
test = pd.read_csv('/content/input/testing_data.csv')
Paramters_Description = pd.read_csv('/content/input/parameter_description.csv')
sample = pd.read_csv('/content/input/testing_target.csv')
```

Viewing dataset

In [74]:

```
train.head(5)
```

Out[74]:

	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Hospital_region_code	Available_Extra_Rooms_in_Hospital	Department	Ward_Type	Ward_Facility_Code	Bed_Grad
0	1	8	c	3	Z	3	radiotherapy	R	F	2.(
1	2	2	c	5	Z	2	radiotherapy	S	F	2.(
2	3	10	e	1	X	2	anesthesia	S	E	2.(
3	4	26	b	2	Y	2	radiotherapy	R	D	2.(
4	5	26	b	2	Y	2	radiotherapy	S	D	2.(

Dataset Column Description

```
Paramters_Description
```

	Column	Description
0	case_id	It is identity number given by hospital admini...
1	Hospital_code	It is the code (identity number) given to the ...
2	Hospital_type_code	It is the unique code given to the type of hos...
3	City_Code_Hospital	It is the code given to the city where the hos...
4	Hospital_region_code	It is the code given to the region where the h...
5	Available_Extra_Rooms_in_Hospital	It will display the number of rooms that are s...
6	Department	The department that is overlooking the patient...
7	Ward_Type	The unique code given to the type of ward to w...
8	Ward_Facility_Code	The unique code given to the facility in the w...
9	Bed_Grade	It is the quality or condition of the bed in t...
10	patientid	It is the unique identity value given to the p...
11	City_Code_Patient	It is the unique identity code given to the ci...
12	Type_of_Admission	It is the admission type registered in the hos...
13	Severity_of_Illness	It is the severity level of the patients' illn...
14	Visitors_with_Patient	Number of the visitors with the patients to ta...
15	Age	It is the age of patients. It is given in peri...
16	Admission_Deposit	It is the deposit amount that the patient paid...
17	Stay	It is the Length Of Stay (LOS) of patients. I...

Analysis of dataset

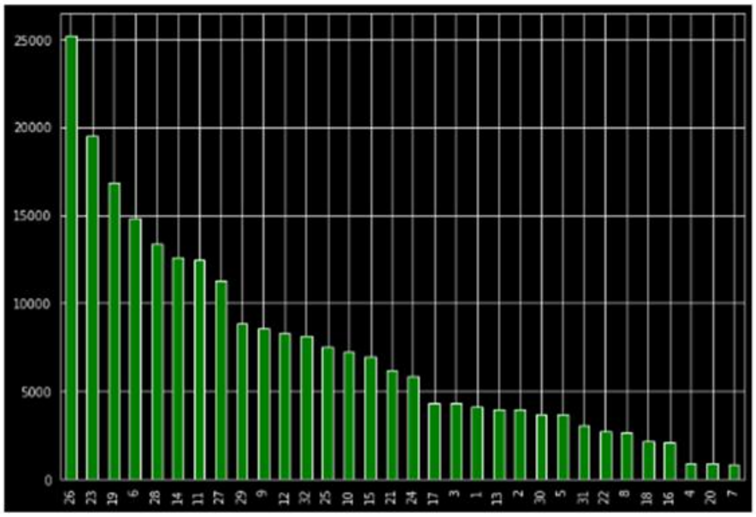
Distribution of values

Hospital_code

```
train.Hospital_code.value_counts()
```

```
26    25225
23    19505
19    16825
6     14847
28    13341
14    12594
11    12454
27    11312
29     8828
9      8558
12     8312
32     8166
25     7529
10     7257
15     6965
21     6226
24     5863
17     4319
3      4308
1      4111
13     3974
2      3940
30     3707
5      3684
31     3051
22     2740
8      2679
18     2164
16     2119
4       937
20      905
7       864
Name: Hospital_code, dtype: int64
```

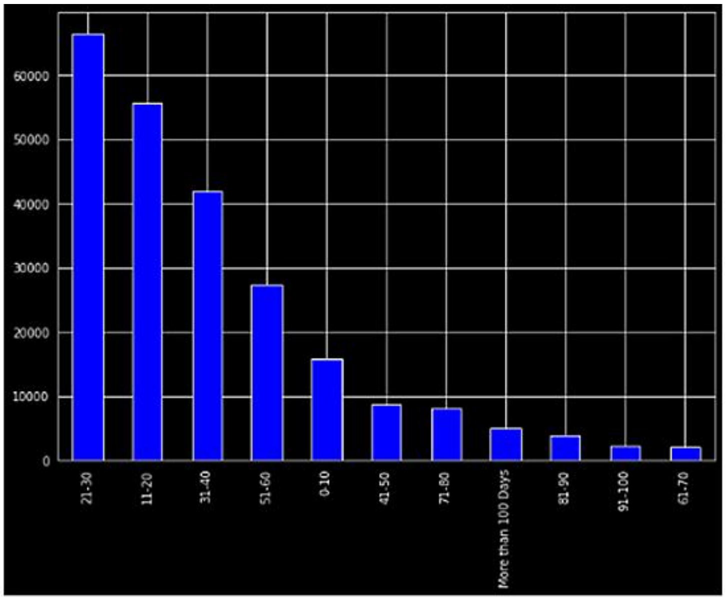
```
plt.figure(figsize=(10,7))
train.Hospital_code.value_counts().plot(kind="bar", color = ['green'])
```



Stay

```
train.Stay.value_counts()
```

```
21-30                 66497
11-20                 55691
31-40                 41951
51-60                 27458
0-10                  15866
41-50                  8665
71-80                  8061
More than 100 Days     5029
81-90                  3821
91-100                 2179
61-70                  2090
Name: Stay, dtype: int64
```



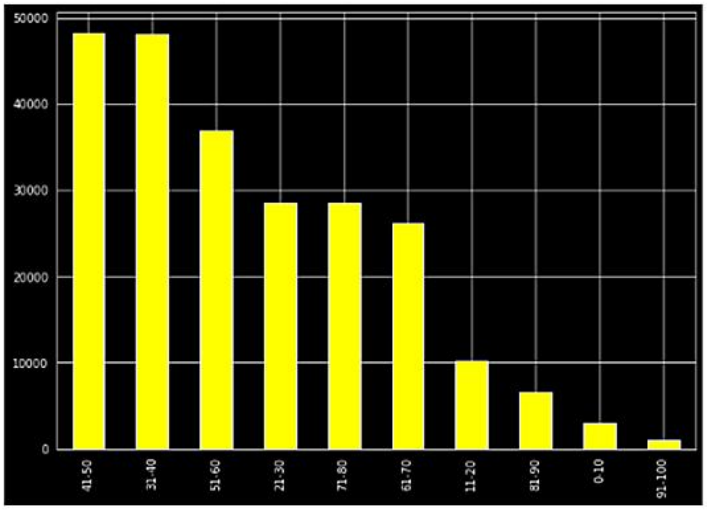
Age

```
train.Age.value_counts()
```

```
41-50     48272
31-40     48106
51-60     36969
21-30     28555
71-80     28552
61-70     26139
11-20     10141
```

```
81-90       6578
0-10        3030
91-100       966
Name: Age, dtype: int64
```

```
#Age distribution
plt.figure(figsize=(10,7))
train.Age.value_counts().plot(kind="bar", color = ['Yellow'])
```



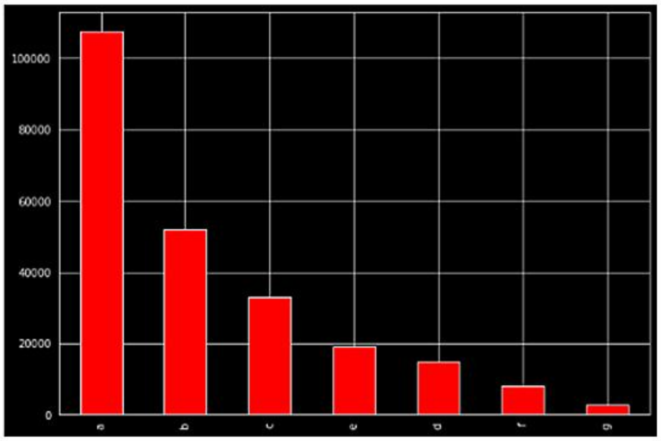
Hospital_type_code

```
train.Hospital_type_code.value_counts()
```

```
a    107545
b     51925
```

```
c      32995
e      19105
d      14833
f       8166
g       2740
Name: Hospital_type_code, dtype: int64
```

```
#Hospital_type_code distribution
plt.figure(figsize=(10,7))
train.Hospital_type_code.value_counts().plot(kind="bar", color = ['Red'])
```

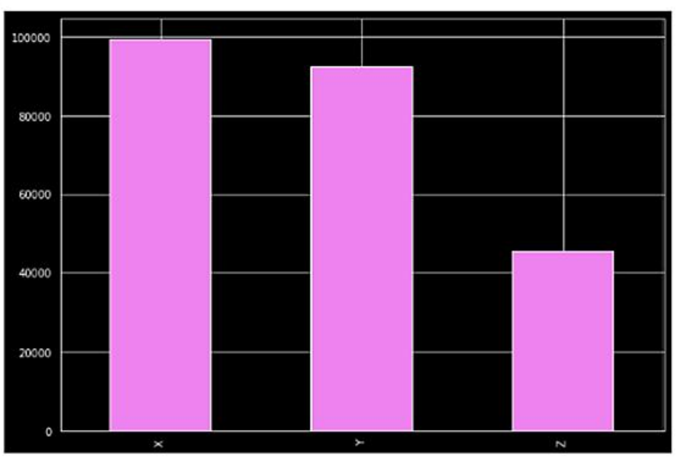


Hospital_region_code

```
train.Hospital_region_code.value_counts()
```

```
X     99568
Y     92214
Z     45527
Name: Hospital_region_code, dtype: int64
```

```
#Hospital_region_code distribution
plt.figure(figsize=(10,7))
train.Hospital_region_code.value_counts().plot(kind="bar", color = ['Violet'])
```



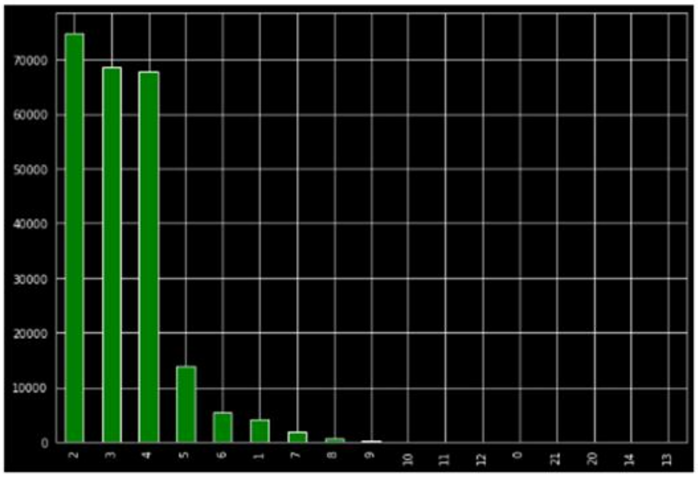
Available_Extra_Rooms_in_Hospital

```
train.Available_Extra_Rooms_in_Hospital.value_counts()
```

```
2      74877
3      68517
4      67756
5      13879
6       5344
1       4208
7       1876
8        622
9        144
10        46
```

```
11     13
12     11
0      11
21      2
20      1
14      1
13      1
Name: Available_Extra_Rooms_in_Hospital, dtype: int64
```

```
#Available_Extra_Rooms_in_Hospital distribution
plt.figure(figsize=(10,7))
train.Available_Extra_Rooms_in_Hospital.value_counts().plot(kind="bar", color = ['green'])
```



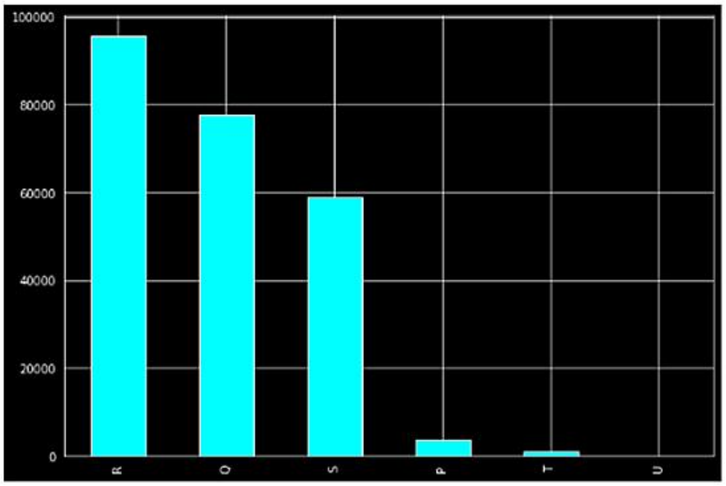
Department

```
train.Department.value_counts()
```

```
gynecology          185062
```

```
R    95788
Q    77707
S    59022
P     3691
T     1092
U        9
Name: Ward_Type, dtype: int64
```

```
#Ward_Type distribution
plt.figure(figsize=(10,7))
train.Ward_Type.value_counts().plot(kind="bar", color = ['cyan'])
```



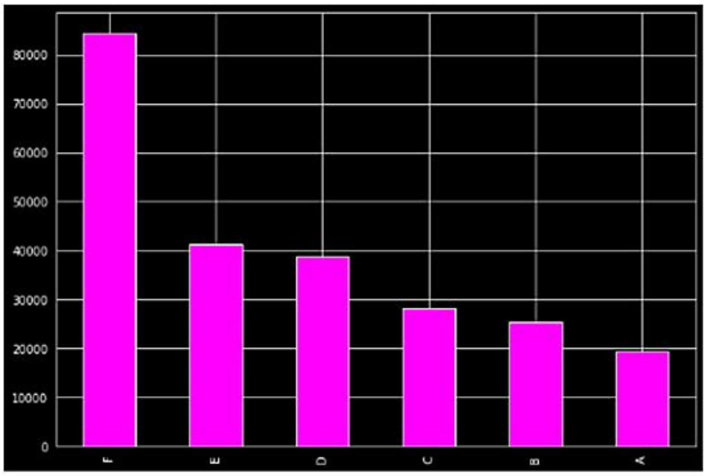
Ward_Facility_Code

```
train.Ward_Facility_Code.value_counts()
```

```
F    84438
E    41246
```

```
D     38584
C     28137
B     25493
A     19411
Name: Ward_Facility_Code, dtype: int64
```

```
#Ward_Facility_Code distribution
plt.figure(figsize=(10,7))
train.Ward_Facility_Code.value_counts().plot(kind="bar", color = ['magenta'])
```



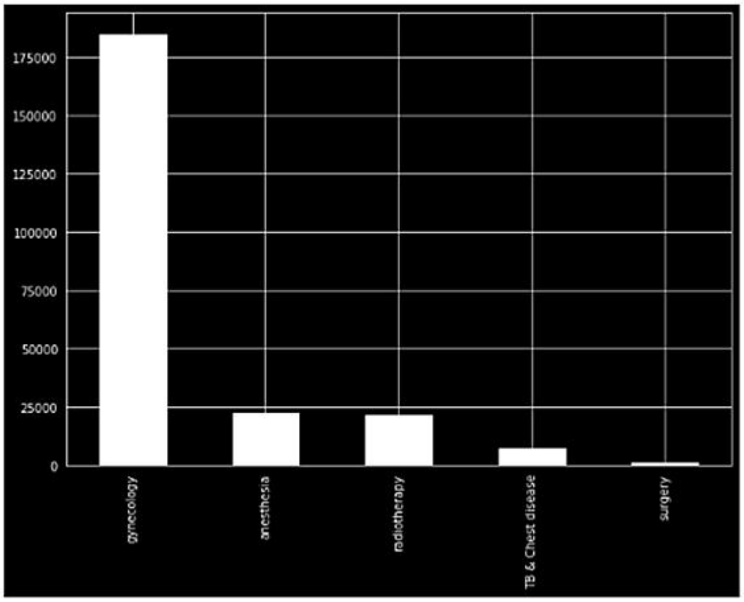
Visitors_with_Patient

```
train.Visitors_with_Patient.value_counts()
```

```
2.0     103037
4.0      59068
3.0      43860
6.0      14211
5.0       6992
```

```
anesthesia              22557
radiotherapy            21725
TB & Chest disease       7017
surgery                   948
Name: Department, dtype: int64
```

```
#Department distribution
plt.figure(figsize=(10,7))
train.Department.value_counts().plot(kind="bar", color = ['white'])
```

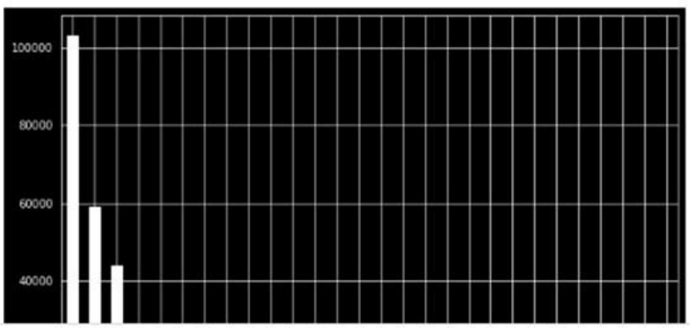


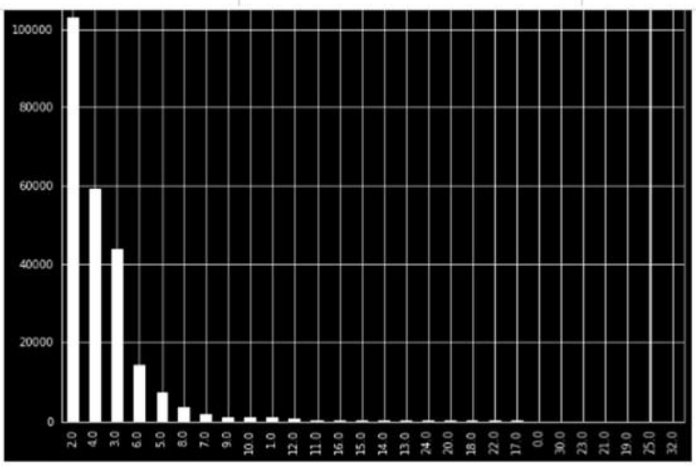
Ward_Type

```
train.Ward_Type.value_counts()
```

```
8.0      3662
7.0      1888
9.0      1024
10.0      882
1.0       871
12.0      757
11.0      242
16.0      220
15.0      146
14.0      138
13.0       84
24.0       63
20.0       46
18.0       35
22.0       16
17.0       15
0.0        13
30.0        9
23.0        8
21.0        8
19.0        6
25.0        6
32.0        1
Name: Visitors_with_Patient, dtype: int64
```

```python
#Visitors_with_Patient distribution
plt.figure(figsize=(10,7))
train.Visitors_with_Patient.value_counts().plot(kind="bar", color = ['white'])
```





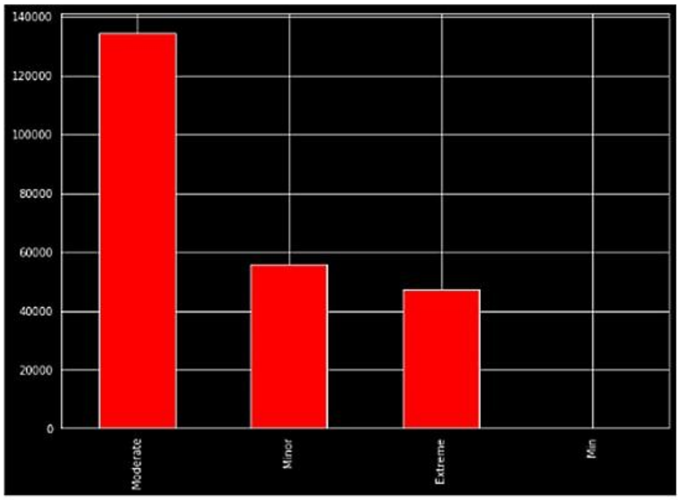
Severity of Illness

]:
```python
train.Severity_of_Illness.value_counts()
```

]:
```
Moderate    134324
Minor        55665
Extreme      47319
Min              1
Name: Severity_of_Illness, dtype: int64
```

]:
```python
#Severity_of_Illness distribution
plt.figure(figsize=(10,7))
train.Severity_of_Illness.value_counts().plot(kind="bar", color = ['red'])
```

]:



Unique values of columns

```
for features in train.columns:
    print('*-----------------------------------------------------------------------------------------------*')
    print(' Unique Values for {}'.format(features))
    print(train[features].unique())
    print('*-----------------------------------------------------------------------------------------------*')
    print()
```

```
*------------------------------------------------------------------------------------*
 Unique Values for case_id
[     1      2      3 ... 237307 237308 237309]
*------------------------------------------------------------------------------------*

*------------------------------------------------------------------------------------*
 Unique Values for Hospital_code
[ 8  2 10 26 23 32  1 22 16  9  6 29 12  3 21 28 27 19  5 14 13 31 24 17
 25 15 11 30 18  4  7 20]
*------------------------------------------------------------------------------------*

*------------------------------------------------------------------------------------*
 Unique Values for Hospital_type_code
['c' 'e' 'b' 'a' 'f' 'd' 'g']
*------------------------------------------------------------------------------------*

*------------------------------------------------------------------------------------*
 Unique Values for City_Code_Hospital
[ 3  5  1  2  6  9 10  4 11  7 13]
*------------------------------------------------------------------------------------*

*------------------------------------------------------------------------------------*
 Unique Values for Hospital_region_code
['Z' 'X' 'Y']
*------------------------------------------------------------------------------------*

*------------------------------------------------------------------------------------*
 Unique Values for Available_Extra_Rooms_in_Hospital
[ 3  2  1  4  6  5  7  8  9 10 12  0 11 20 14 21 13]
*------------------------------------------------------------------------------------*

*------------------------------------------------------------------------------------*
 Unique Values for Department
['radiotherapy' 'anesthesia' 'gynecology' 'TB & Chest disease' 'surgery']
*------------------------------------------------------------------------------------*

*------------------------------------------------------------------------------------*
 Unique Values for Ward_Type
['R' 'S' 'Q' 'P' 'T' 'U']
*------------------------------------------------------------------------------------*

*------------------------------------------------------------------------------------*
 Unique Values for Ward_Facility_Code
['F' 'E' 'D' 'B' 'A' 'C']
*------------------------------------------------------------------------------------*

*------------------------------------------------------------------------------------*
 Unique Values for Bed_Grade
[ 2.  3.  4.  1. nan]
*------------------------------------------------------------------------------------*

*------------------------------------------------------------------------------------*
 Unique Values for patientid
[31397 63418  8088 ... 37502 73756 21763]
*                                                                                    *
```

```
*-------------------------------------------------------------------------------------------------*
 Unique Values for City_Code_Patient
[ 7.  8.  2.  5.  6.  3.  4.  1.  9. 14. nan 25. 15. 12. 10. 28. 24. 23.
 20. 11. 13. 21. 18. 16. 26. 27. 22. 19. 31. 34. 32. 30. 29. 37. 33. 35.
 36.]
*-------------------------------------------------------------------------------------------------*

*-------------------------------------------------------------------------------------------------*
 Unique Values for Type_of_Admission
['Emergency' 'Trauma' 'Urgent']
*-------------------------------------------------------------------------------------------------*

*-------------------------------------------------------------------------------------------------*
 Unique Values for Severity_of_Illness
['Extreme' 'Moderate' 'Minor' 'Min']
*-------------------------------------------------------------------------------------------------*

*-------------------------------------------------------------------------------------------------*
 Unique Values for Visitors_with_Patient
[ 2.  4.  3.  8.  6.  7. 13.  5.  1. 10. 15. 11. 12.  9. 24. 16. 14. 20.
  0. 19. 18. 17. 23. 21. 32. 30. 22. 25. nan]
*-------------------------------------------------------------------------------------------------*

*-------------------------------------------------------------------------------------------------*
 Unique Values for Age
['51-60' '71-80' '31-40' '41-50' '81-90' '61-70' '21-30' '11-20' '0-10'
 '91-100' nan]
*-------------------------------------------------------------------------------------------------*

*-------------------------------------------------------------------------------------------------*
 Unique Values for Admission_Deposit
[4911. 5954. 4745. ... 2710. 2236.   nan]
*-------------------------------------------------------------------------------------------------*

*-------------------------------------------------------------------------------------------------*
 Unique Values for Stay
['0-10' '41-50' '31-40' '11-20' '51-60' '21-30' '71-80'
 'More than 100 Days' '81-90' '61-70' '91-100' nan]
*-------------------------------------------------------------------------------------------------*
```

Data Preprocessing & Feature Engineering

The following features may have relevance with the Length of Stay of a patient

Department: It Relates to the type of disease. Hence it will have impact on the length of stay of the patients

Type of Admission: It Relates to patients' reason of admission to the hospital and definitely it will have impact on length of stay opf the patients

Severity of Illness: It Relates to the curability of disease

Age: Relates to the curability of diseaseThe following features may have relevance with the Length of Stay of a patient

Department: It Relates to the type of disease. Hence it will have impact on the length of stay of the patients

Type of Admission: It Relates to patients' reason of admission to the hospital and definitely it will have impact on length of stay opf the patients

Severity of Illness: It Relates to the curability of disease

Age: Relates to the curability of disease

Ward_Type: Relates to the curability of disease

\

The following features doesn't have relevance with the Length Of Stay(LOS) of Patients

Hospital_region_code: It is code given to the hospital region which is irrelevent to the Length of Stay.

Bed Grade: It is the grade given to the quality of the bed in ward it is also irrelevent to the length of stay.

patientid: It is the identity number or code given for the identification of the patient which is irrelevant to the length of stay.

City_Code_Patient: It is the city code and irrelevant to the length of stay of patients.

```
"""
as 'Hospital_region_code', 'Bed_Grade', 'patientid', 'City_Code_Patient' are irrelevant to the health or
length of stay of patients so lets drop these parameters from training and testing dataset to improve the performace of model (high accurracy)
by reducing the complexity
"""
train = train.drop(['Hospital_region_code', 'Bed_Grade', 'patientid', 'City_Code_Patient'], axis = 1)
test = test.drop(['Hospital_region_code', 'Bed_Grade', 'patientid', 'City_Code_Patient'], axis = 1)
```

```
# Combine test and train dataset for processing
combined = [train, test]
combined
```

```
[        case_id  Hospital_code Hospital_type_code  City_Code_Hospital  \
 0             1              8                  c                   3
 1             2              2                  c                   5
 2             3             10                  e                   1
 3             4             26                  b                   2
 4             5             26                  b                   2
 ...         ...            ...                ...                 ...
 237304   237305             23                  a                   6
 237305   237306             19                  a                   7
 237306   237307              8                  c                   3
 237307   237308             21                  c                   3
 237308   237309              5                  a                   1

         Available_Extra_Rooms_in_Hospital    Department Ward_Type  \
 0                                       3  radiotherapy         R
 1                                       2  radiotherapy         S
 2                                       2    anesthesia         S
 3                                       2  radiotherapy         R
 4                                       2  radiotherapy         S
 ...                                   ...           ...       ...
 237304                                  3    gynecology         R
 237305                                  2    gynecology         R
 237306                                  5    gynecology         Q
 237307                                  4  radiotherapy         S
 237308                                  3    gynecology         Q

        Ward_Facility_Code Type_of_Admission Severity_of_Illness  \
 0                       F         Emergency             Extreme
 1                       F            Trauma             Extreme
 2                       E            Trauma             Extreme
 3                       D            Trauma             Extreme
 4                       D            Trauma             Extreme
 ...                   ...               ...                 ...
 237304                  F            Trauma             Extreme
 237305                  C         Emergency             Extreme
```

```
237306                F       Emergency           Minor
237307                A       Emergency           Minor
237308                E          Trauma             Min

       Visitors_with_Patient   Age  Admission_Deposit   Stay
0                        2.0  51-60             4911.0   0-10
1                        2.0  51-60             5954.0  41-50
2                        2.0  51-60             4745.0  31-40
3                        2.0  51-60             7272.0  41-50
4                        2.0  51-60             5558.0  41-50
...                      ...    ...                ...    ...
237304                   5.0  41-50             4298.0  51-60
237305                   4.0  41-50             4165.0  31-40
237306                   4.0  31-40             5075.0  21-30
237307                   2.0  31-40             5179.0  11-20
237308                   NaN    NaN                NaN    NaN

[237309 rows x 14 columns],
       case_id  Hospital_code Hospital_type_code  City_Code_Hospital  \
0       318439             21                  c                   3
1       318440             29                  a                   4
2       318441             26                  b                   2
3       318442              6                  a                   6
4       318443             28                  b                  11
...        ...            ...                ...                 ...
137052  455491             11                  b                   2
137053  455492             25                  e                   1
137054  455493             30                  c                   3
137055  455494              5                  a                   1
137056  455495              6                  a                   6

       Available_Extra_Rooms_in_Hospital    Department Ward_Type  \
0                                      3    gynecology         S
1                                      2    gynecology         S
2                                      3    gynecology         Q
3                                      3    gynecology         Q
4                                      2    gynecology         R
...                                  ...           ...       ...
137052                                 4    anesthesia         Q
137053                                 2  radiotherapy         R
137054                                 2    anesthesia         R
137055                                 2    anesthesia         R
137056                                 3    gynecology         Q

      Ward_Facility_Code Type_of_Admission Severity_of_Illness  \
0                      A         Emergency            Moderate
1                      F            Trauma            Moderate
2                      D         Emergency            Moderate
3                      F            Trauma            Moderate
4                      F            Trauma            Moderate
...                  ...               ...                 ...
137052                 D         Emergency               Minor
137053                 E         Emergency            Moderate
137054                 A            Urgent               Minor
137055                 E            Trauma               Minor
137056                 F            Trauma             Extreme

       Visitors_with_Patient   Age  Admission_Deposit
0                          2  71-80               3095
1                          4  71-80               4018
2                          3  71-80               4492
3                          3  71-80               4173
4                          4  71-80               4161
...                      ...    ...                ...
137052                     4  41-50               6313
137053                     2   0-10               3510
137054                     2   0-10               7190
137055                     2  41-50               5435
137056                     5  51-60               4702

[137057 rows x 13 columns]]
```

Lets encode the categorical data for traning the model

```python
# Encoding Department
from sklearn.preprocessing import LabelEncoder

for dataset in combined:
    label = LabelEncoder()
    dataset['Department'] = label.fit_transform(dataset['Department'])
combined[1].Department.unique()
```

```
array([2, 1, 0, 3, 4])
```

```python
# Encoding Ward Type, Hospital_type_code, Ward_Facility_Code, Type_of_Admission, Severity_of_Illness
for dataset in combined:
    label = LabelEncoder()
    dataset['Hospital_type_code'] = label.fit_transform(dataset['Hospital_type_code'])
    dataset['Ward_Facility_Code'] = label.fit_transform(dataset['Ward_Facility_Code'])
    dataset['Ward_Type'] = label.fit_transform(dataset['Ward_Type'])
    dataset['Type_of_Admission'] = label.fit_transform(dataset['Type_of_Admission'])
    dataset['Severity_of_Illness'] = label.fit_transform(dataset['Severity_of_Illness'])
```

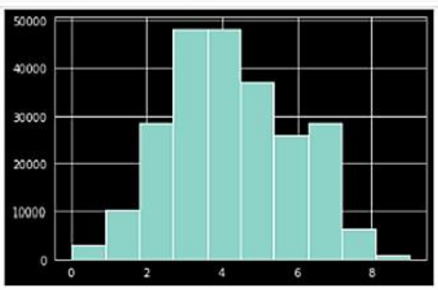
```python
combined[0]
```

	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Available_Extra_Rooms_in_Hospital	Department	Ward_Type	Ward_Facility_Code	Type_of_Admission	Severi
0	1	8	2	3	3	3	2	5	0	
1	2	2	2	5	2	3	3	5	1	
2	3	10	4	1	2	1	3	4	1	
3	4	26	1	2	2	3	2	3	1	
4	5	26	1	2	2	3	3	3	1	
...	...	...	...	...	...	...	...	...	...	
237304	237305	23	0	6	3	2	2	5	1	
237305	237306	19	0	7	2	2	2	2	0	
237306	237307	8	2	3	5	2	1	5	0	
237307	237308	21	2	3	4	3	3	0	0	
237308	237309	5	0	1	3	2	1	4	1	

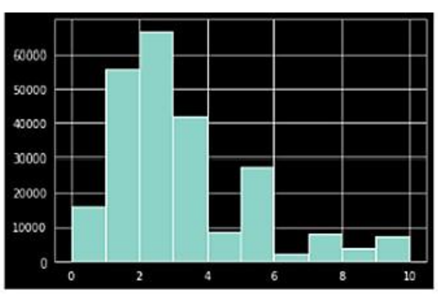
237309 rows × 14 columns

```
combined[1]
```

	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Available_Extra_Rooms_in_Hospital	Department	Ward_Type	Ward_Facility_Code	Type_of_Admission	Severi
0	318439	21	2	3	3	2	3	0	0	
1	318440	29	0	4	2	2	3	5	1	
2	318441	26	1	2	3	2	1	3	0	



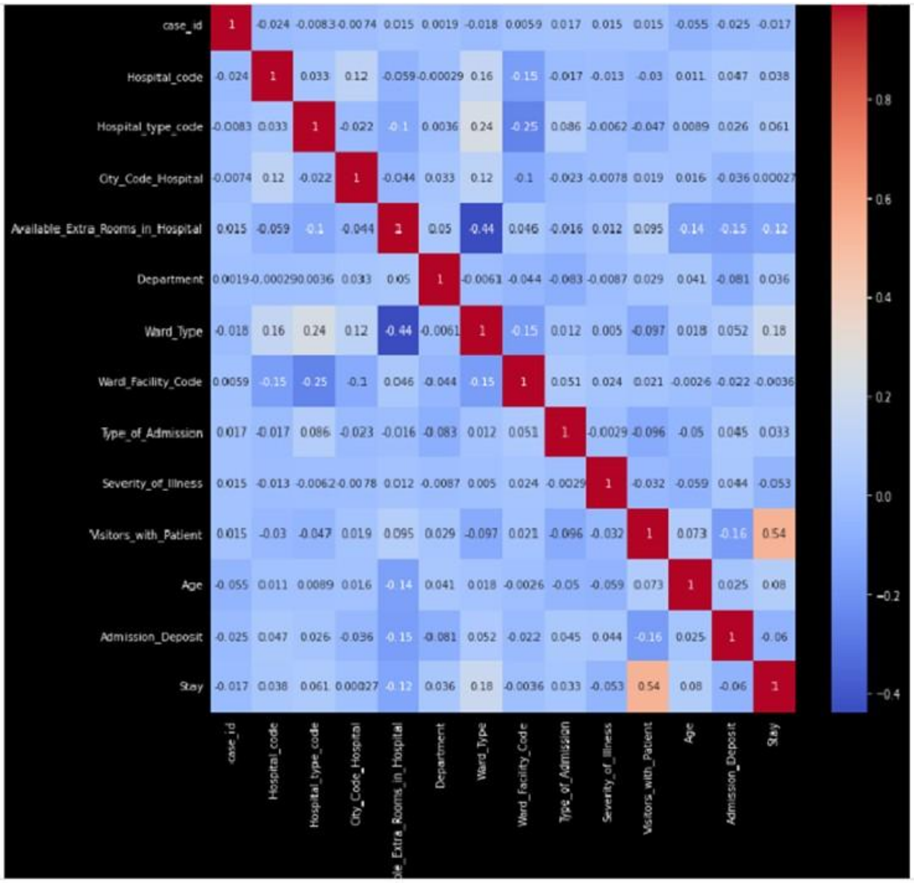

```
combined[0].Stay.hist()
```



shape of combined (train data, test data) dataset

```
for dataset in combined:
    print(dataset.shape)
```

```
(237309, 14)
(137057, 13)
```




```
combined[1]
```

	case_id	Hospital_code	Hospital_type_code	City_Code_Hospital	Available_Extra_Rooms_in_Hospital	Department	Ward_Type	Ward_Facility_Code	Type_of_Admission	Severi
0	318439	21	2	3	3	2	3	0	0	
1	318440	29	0	4	2	2	3	5	1	
2	318441	26	1	2	3	2	1	3	0	
3	318442	6	0	6	3	2	1	5	1	
4	318443	28	1	11	2	2	2	5	1	
...	...	...	...	...	...	...	...	...	...	
137052	455491	11	1	2	4	1	1	3	0	
137053	455492	25	4	1	2	3	2	4	0	
137054	455493	30	2	3	2	1	2	0	2	
137055	455494	5	0	1	2	1	2	4	1	
137056	455495	6	0	6	3	2	1	5	1	

137057 rows × 13 columns

Training the model

```
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC, LinearSVC
from sklearn.ensemble import RandomForestClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.linear_model import Perceptron
from sklearn.linear_model import SGDClassifier
from sklearn.tree import DecisionTreeClassifier
```

```
train = combined[0]
test = combined[1]
```

```
X_train = train.drop(['case_id', 'Stay'], axis=1)
Y_train = train["Stay"]
X_test  = test.drop("case_id", axis=1).copy()
```

```
X_train.shape
```

```
(237309, 12)
```

```
Y_train.shape
```

```
(237309,)
```

```
X_test.shape
```

```
(137057, 12)
```

```
X_test.columns
```

```
Index(['Hospital_code', 'Hospital_type_code', 'City_Code_Hospital',
       'Available_Extra_Rooms_in_Hospital', 'Department', 'Ward_Type',
       'Ward_Facility_Code', 'Type_of_Admission', 'Severity_of_Illness',
       'Visitors_with_Patient', 'Age', 'Admission_Deposit'],
      dtype='object')
```

```
Y_train
```

```
0         0.0
1         4.0
2         3.0
3         4.0
4         4.0
         ...
237304    5.0
237305    3.0
237306    2.0
237307    1.0
237308    NaN
Name: Stay, Length: 237309, dtype: float64
```

```
X_train.fillna(0,inplace=True)
Y_train.fillna(0,inplace=True)
X_test.fillna(0,inplace=True)
```

K-Nearest Neighbor Algorithm

```
knn = KNeighborsClassifier(n_neighbors = 3)
knn.fit(X_train, Y_train)
Y_pred = knn.predict(X_test)
acc_knn = round(knn.score(X_train, Y_train) * 100, 2)
acc_knn
```

```
53.99
```

Descision Tree Algorithm

```
decision_tree = DecisionTreeClassifier()
decision_tree.fit(X_train, Y_train)
Y_pred = decision_tree.predict(X_test)
acc_decision_tree = round(decision_tree.score(X_train, Y_train) * 100, 2)
acc_decision_tree
```

```
99.76
```

Random Forest Algorithm

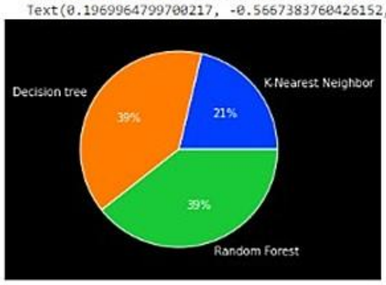
```
random_forest = RandomForestClassifier(n_estimators=100)
random_forest.fit(X_train, Y_train)
Y_pred = random_forest.predict(X_test)
random_forest.score(X_train, Y_train)
acc_random_forest = round(random_forest.score(X_train, Y_train) * 100, 2)
acc_random_forest
```

```
99.76
```

Prediction accuracy comparison

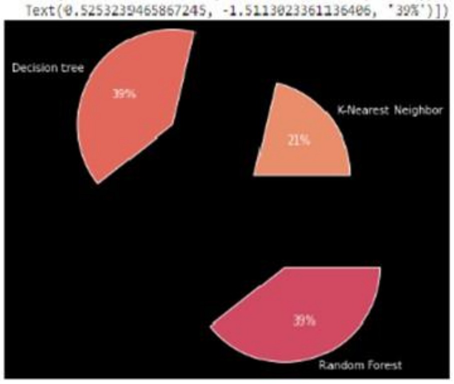
```
palette_color = sns.color_palette('bright')
data=[acc_knn, acc_decision_tree,acc_random_forest]
keys=['K-Nearest Neighbor','Decision tree','Random Forest']

#getting the algorithm with highest accuracy
max_accuracy=max(data)
index=[0,0,0]
j=0;
for i in data:
  if(i==max_accuracy):
    index[j]=1
    j=j+1
  else:
    index[j]=0.01
    j=j+1

plt.pie(data, labels=keys, colors=palette_color, autopct='%.0f%%')
```

```
([,
  ,
  ],
 [Text(0.8628423642631272, 0.682277842548633, 'K-Nearest Neighbor'),
  Text(-0.9277499083745313, 0.590999244932723, 'Decision tree'),
  Text(0.36116021327837317, -1.0390203560781281, 'Random Forest')],
 [Text(0.4706412895980693, 0.3721515504810725, '21%'),
  Text(-0.5060454045679261, 0.322363224508758, '39%'),
  Text(0.1969964799700217, -0.5667383760426152, '39%')])
```




```
palette_color = sns.color_palette('flare')
plt.pie(data, labels=keys, colors=palette_color,explode=index, autopct='%.0f%%')
```

```
 ],
[Text(0.8706863857564283, 0.6884803683899842, 'K-Nearest Neighbor'),
 Text(-1.7711589159877414, 1.1282712857806532, 'Decision tree'),
 Text(0.689487679895076, -1.9835843161491535, 'Random Forest')],
[Text(0.47848531109137044, 0.37835407632242374, '21%'),
 Text(-1.3494544121811365, 0.859635265356688, '39%'),
 Text(0.5253239465867245, -1.5113023361136406, '39%')])
```




```
output = pd.DataFrame({
        "case_id": test["case_id"],
        "Stay": Y_pred
})
```

```
output['Stay'] = output['Stay'].replace(stay_labels.values(), stay_labels.keys())
```

```
output.to_csv('LOS_Prediction.csv', index = False)
```

```
output
```

	case_id	Stay
0	318439	0-10
2	318441	21-30
3	318442	11-20
4	318443	31-40
...	...	...
137052	455491	0-10
137053	455492	0-10
137054	455493	21-30
137055	455494	21-30
137056	455495	51-60

137057 rows × 2 columns

```
data=np.array([[29,0,4,2,2,3,5,1,2,4,7,4018]])
p=random_forest.predict(data)
p
```

```
/usr/local/lib/python3.7/dist-packages/sklearn/base.py:451: UserWarning: X does not have valid feature names, but RandomForestClassifier was fitted wi
th feature names
  "X does not have valid feature names, but"
```

```
array([5.])
```

```
def prediction(p):
  if(p[0]==0):
    print("The predicted LOS of patient is : 0-10")
  elif(p[0]==1):
    print("The predicted LOS of patient is : 11-20")
  elif(p[0]==2):
    print("The predicted LOS of patient is : 21-30")
  elif(p[0]==3):
    print("The predicted LOS of patient is : 31-40")
  elif(p[0]==4):
    print("The predicted LOS of patient is : 41-50")
  elif(p[0]==5):
    print("The predicted LOS of patient is : 51-60")
  elif(p[0]==6):
    print("The predicted LOS of patient is : 61-70")
  elif(p[0]==7):
    print("The predicted LOS of patient is : 71-80")
  elif(p[0]==8):
```

```python
  elif(p[0]==8):
    print("The predicted LOS of patient is : 81-90")
  elif(p[0]==9):
    print("The predicted LOS of patient is : 91-100")
  elif(p[0]==10):
    print("The predicted LOS of patient is : More than 100 Days")
```

```python
data=np.array([[29,0,4,2,2,3,5,1,2,4,7,4018]])
p=random_forest.predict(data)
print(p)
```

```python
prediction(p)
```

```
The predicted LOS of patient is : 51-60
```