Analytics for Hospitals' Health-Care Data

Team ID : PNT2022TMID33091

Project Name : Analytics for Hospitals' Health-Care Data

Team Size : 4

Team Members:

Team Leader : SANJAY.V
Team Member 1 : SIVASUBRAMANIYAN.A
Team Member 2 : NAVEEN.A
Team Member 3 : MOHAN SALMAN MYDEEN.A

Project Report

FUTURE SCOPE
APPENDIX
Source Code
GitHub & Project Demo Link

INTRODUCTION

Project Overview:

Healthcare analytics is the process of analyzing current and historical industry data to predict trends, improve outreach, and even better manage the spread of diseases. The field covers a broad range of businesses and offers insights on both the macro and micro level.Data analytics in clinical settings attempts to reduce patient wait times via improved scheduling and staffing, give patients more options when scheduling appointments and receiving treatment, and reduce readmission rates by using population health data to predict which patients are at greatest risk.Healthcare industry is no exception. Working in the healthcare industry for almost a year now, I've realized that the healthcare industry is notorious for being conservative when it comes to technology adoption.

PURPOSE:

The rise of data analytics technologies in the healthcare industry including machine learning, deep learning, artificial intelligence, and others have derived the development of Population Health Management (PHM) service in the industry. The rise of PHM has re-shifted the focus of healthcare services from treatment and response to prediction and prevention. The powerful predictive analytics technology is able to identify risk patterns for chronic diseases in patients at early stages, thereby, giving

healthcare practitioners a chance to act in time and avoid complexities of the later stages. The predictive analysis isn't just a life-saving technology, but it also helps in lowering the cost of the treatment and cure for patients.

Literature survey

Introduction

Poverty is the real context of India. Three fourths of the population live below or at subsistence levels. This means 70-90 per cent of their incomes goes towards food and related consumption. In such a context social security support for health, education, housing etc. becomes critical. Ironically, India has one of the largest private health sectors in the world with over 80 per cent of ambulatory care being supported through out-of-pocket expenses. The public health services are very inadequate. The public curative and hospital services are mostly in the cities where only25 percent of the one billion population reside. Rural areas have mostly preventive and promotive services like family planning and immunisation. The private sector has a virtual monopoly over ambulatory curative services in both rural and urban areas and over half of hospital care. Further, a very large proportion of private providers are not qualified to provide modern health care because they are either trained in other systems of medicine (traditional Indian systems like ayurveda, unani, siddha, and homoeopathy) or worse, do not have any training. These, however, are the providers from whom the poor are most likely to seek health care. This adds to the risk faced by the already impoverished population. The healthcare sector is widely considered as one of the most important industries in information technology (Wager 2005). More and more, information technology has been considered as a practice that facilitates healthcare performance through using data and information efficiently within the healthcare sectors. Therefore, Wager et al (2005) said that in order to understand the relation between information technologies and healthcare, we first need to understand what are the technologies used in healthcare.The healthcare sector is widely considered as one of the most important industries in information technology (Wager 2005). More and more,

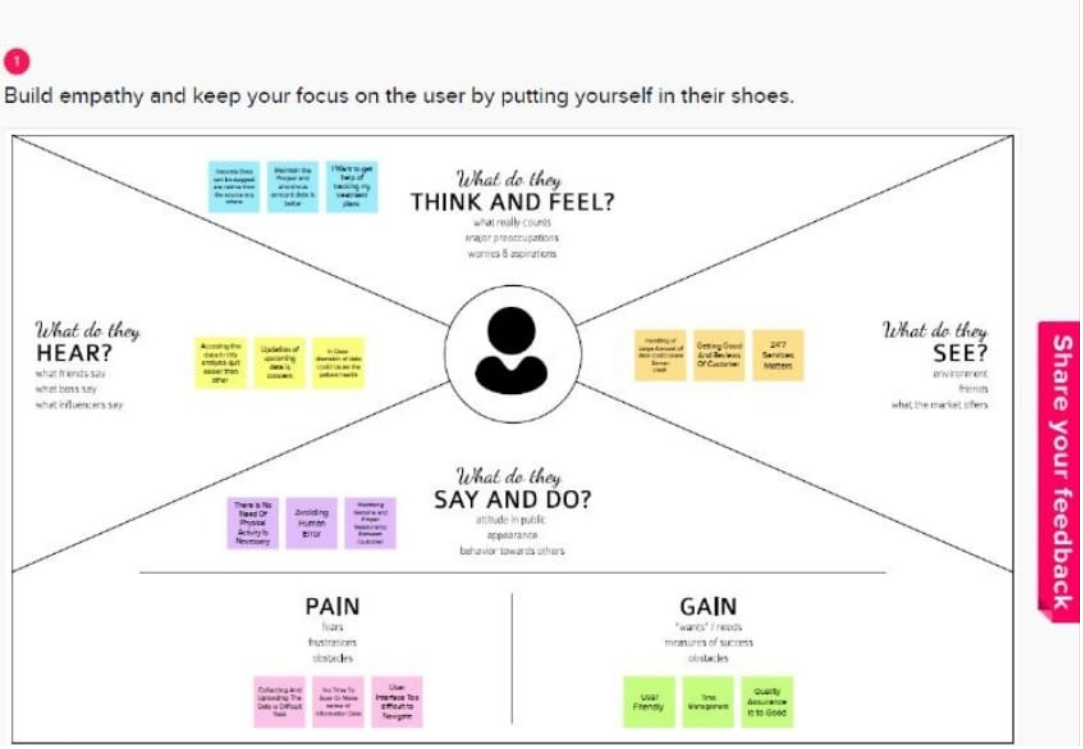
information technology has been considered as a practice that facilitates healthcare performance through using data and information efficiently within the healthcare sectors. Therefore, Wager et al (2005) said that in order to understand the relation between information technologies and healthcare, we first need to understand what are the technologies used in healthcare

The Proposed Frame

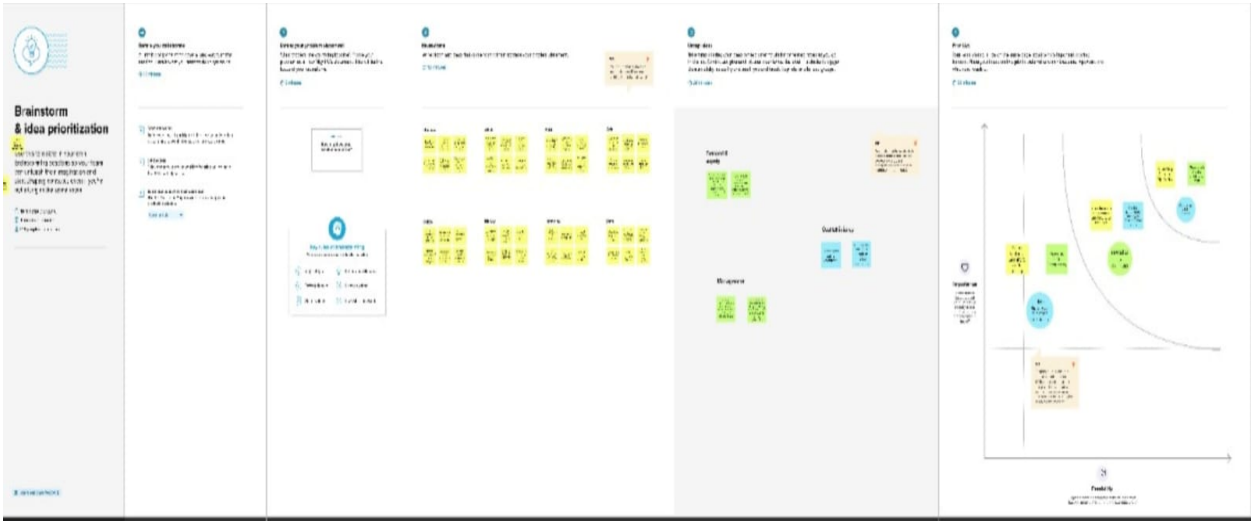
The healthcare industry has generated large amount of data generated from record keeping, compliance and patient related data. In today's digital world, it is mandatory that these data should be digitized. To improve the quality of healthcare by minimizing the costs, it's necessary that large volume of data generated should be analysed effectively to answer new challenges. Similarly government also generates petabytes of data every day. It requires a technology that helps to perform a real time analysis on the enormous data set. This will help the government to provide value added services to the citizens. Big data analytics helps in discovering valuable decisions by understanding the data patterns and the relationship between them with the help of machine learning Algorithms (1).This paper provides an overview of big data analytics in healthcare and government systems. It describes about big data generated by these systems, data characteristics, security issues in handling big data and how big data analytics helps to gain a meaningful insight on these data set. The healthcare industry has generated large amount of data generated from record keeping, compliance and patient related data. In today's digital world, it is mandatory that these data should be digitized. To improve the quality of healthcare by minimizing the costs, it's necessary that large volume of data generated should be analysed effectively to answer new challenges. Similarly government also generates petabytes of data every day. It requires a technology that helps to perform a real time analysis on the enormous data set.

IDEATION & PROPOSED SOLUTION

EMPATHY MAP CANVAS



IDEATION & BRAINSTORMING



PROPOSED SOLUTION

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	To predict the length of stay of patients .
2.	Idea / Solution description	● The length of stay can be predicted using either Fuzzy logic or Tree bagger algorithms. ● Along with the algorithm certain parameters like age, stage of disease, progression, etc., are used for prediction. ● IBM Cognos is used for analytics purpose.
3.	Novelty / Uniqueness	● It predicts the result with moreaccuracyusing whichoverstays can be reduced. ● Proper resources and therapy can beprovided.
4.	Social Impact/ Customer Satisfaction	● Patients can get better treatment and care than before. ● Length of stay prediction minimize the overflow of patients there fore hospital resource utilization will be maximized. ● Reduces expense for treatment.
5.	Business Model(Revenue Model)	● This system can be used in all government hospitals, private hospitals and even in small clinics. ● Activities – Length of stay prediction. ● Key Resource – Medical records . ● Bed consumption is low.
6.	Scalability of the Solution	This model will predict the length of stay of all kindof patients.

PROBLEM SOLUTION FIT

Project Title: Analytics For Hospital's Health-Care Data | Project Design Phase-I - Solution Fit Template | Team ID: PNT2022TMID33091

Define CS, fit

1. CUSTOMER SEGMENT(S)

- Patients
- Hospital Management

6. CUSTOMER CONSTRAINTS

Customers require more accurate andearly predictions of Length of Stay (LOS).

5. AVAILABLE SOLUTIONS

There are few Length of Stay prediction model available whichlacks in predicting some exceptional case where the length of stay may extend.

Explore AS,

Focus on J&P, tap

2. JOBS-TO-BE-DONE / PROBLEMS

Length of stay prediction may varybased on the patient's stage/severity of disease. Patientmay get dissatisfied if there is no bed availability.

9. PROBLEM ROOT CAUSE

Unpredictable length of stay and improper medicalrecords are the root causeof the problem.

7. BEHAVIOUR

Developing a model which predicts the length of stay of unexceptional cases with better accuracy.

Focus on J&P, tap

Identify strong TR & EM

3. TRIGGERS

To accurately predict thelength of stay.

4. EMOTIONS: BEFORE / AFTER

Before : Pateints often get frustratedand depressed.

After: They feel better and get new beginning.

10. YOUR SOLUTION

Our solution includes using algorithms like Fuzzy Logic, Tree Bagger, Random Forest, and Decision Trees to predict the length of stay more accurately. Gives frequent update about the bed availability.

8. CHANNELS of BEHAVIOUR

Users will check for bed availability.

Identify strong TR & EM

REQUIREMENT ANALYSIS

FUNCTIONAL REQUIREMENT

FR No.	Functional Requirement(Epic)	Sub Requirement (Story/ Sub-Task)
FR-1	**User Registration**	Utilizing a Form for Registration, Signingup with Gmail
FR-2	**UserConfirmation**	Email confirmation required
FR-3	**Interoperability**	A dashboard makes it possible to quicklyand inter-operably transmitpatient information with hospitals.
FR-4	**Accuracy**	Based on LOS (LenRth of Stay) the dashboard accurate y predicts thepatient'shealth risks.
FR-5	**Compliance**	The useof a dashboard for compliance byhospitals is quitedynamic and takesplacein real time.
FR-6	**Concise**	These dashboards are easy to understand,simpleto customize, and interactive.

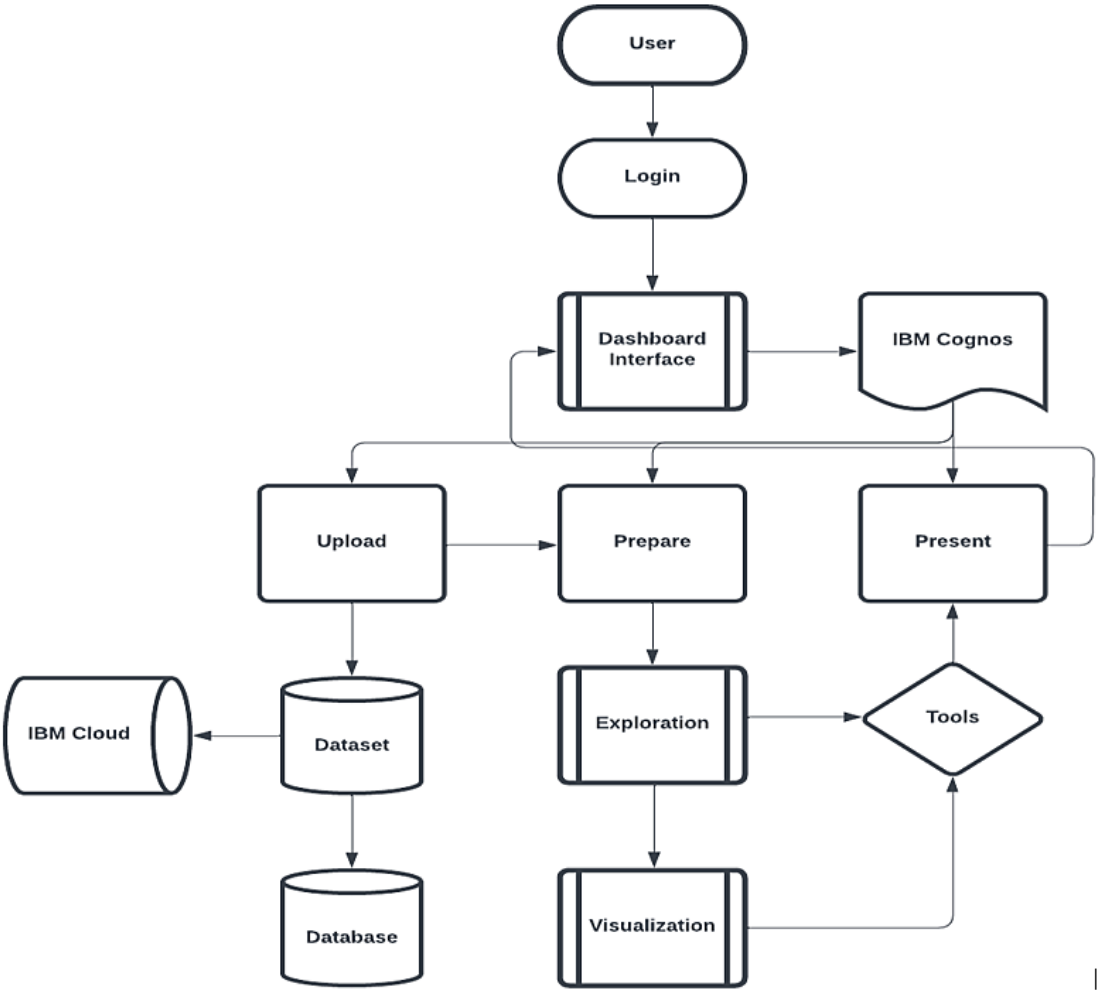
NON-FUNCTIONAL REQUIREMENT

FR No.	Non functional Requirement	Description
NFR-1	**Usability**	In order to provide a through visualrepresentation of the patient's LOS, this dashboard makesuse of datavisualization techniques including charts andgraphs.
NFR-2	**Security**	The Dashboard aids in indicating the level of threat that currently exists forthe hospitals, aswell as past occurrences and incidents, authentication mistakes, scans, probes,and unwanted access.
NFR-3	**Reliability**	Users willfind this dashboard to be consistent, dependabl t and helpful in using in an effective,erricient, and dependable manner.
NFR-4	**Performance**	This dashboard may scan backend users, andexamining how frequentlythey visitthe dashboard might revealrelevant informationabout thejobs the data is beneficial for.
NFR-5	**Availability**	The dashboard is able to promptly satisfy userneeds and aids in givingthe user's datasetthe relevant information.
NFR-6	**Scalability**	A hostedfeature layer, featurelayer view, or hostedtile layer are the layersthat areused inthedashboard.

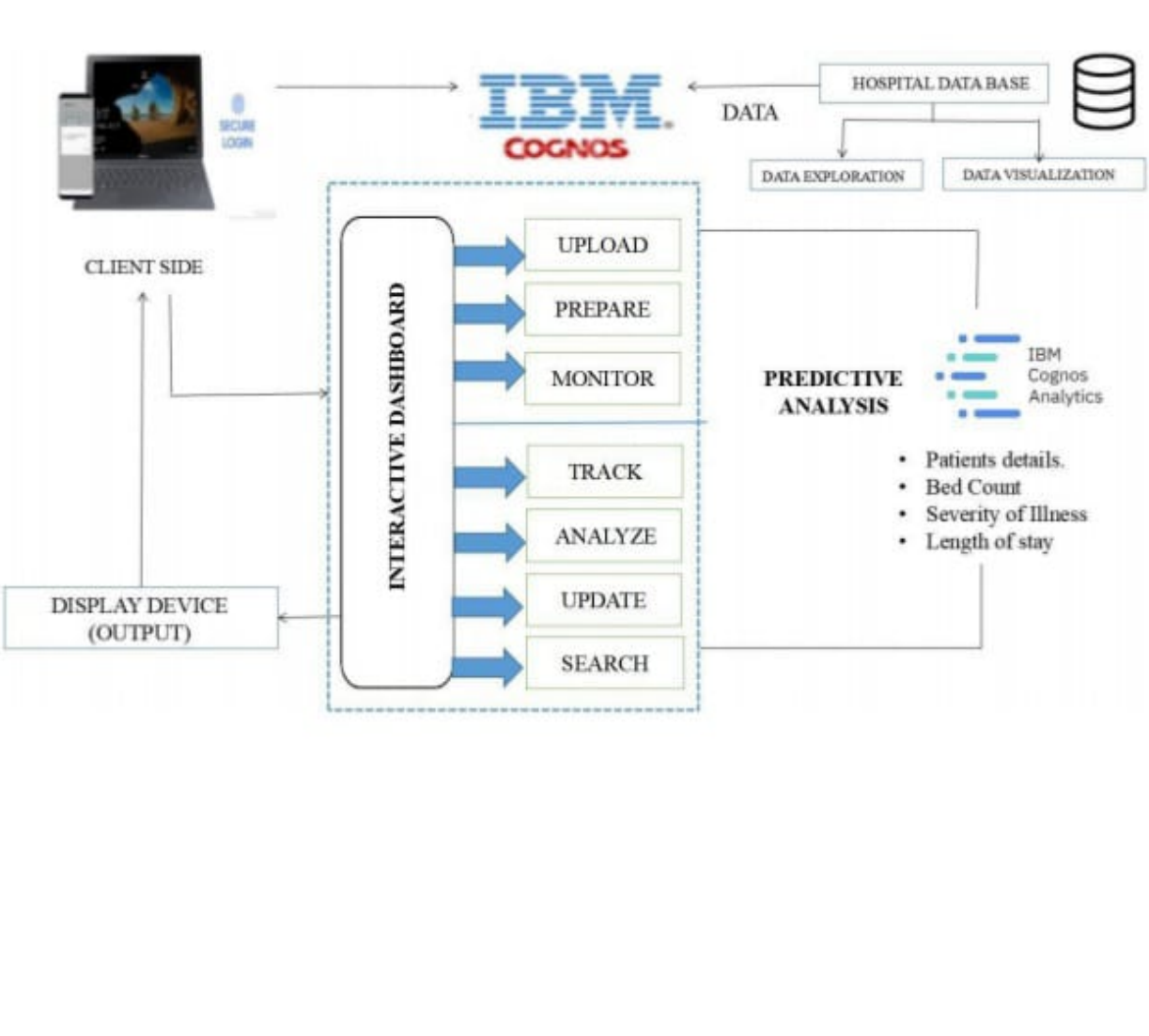
PROJECT DESIGN

DATA FLOW DIAGRAM

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



SOLUTION ARCHITECTURE



USER STORIES

User Type	Functional Requirement(Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the dashboard byentering my email, and password, and confirming my password.	I can access my account inthe dashboard	High	Sprint-1
		USN-2	As a user, I will receivea confirmation emailonce I haveregistered for the dashboard	I can receive a confirmation email & clickconfirm	High	Sprint-1
		USN-3	As a user, I can register for the dashboardthrough Social Media	I can register & access thedashboard with Social Media Login	Low	Sprint-2
		USN-4	As a user, I can register for the dashboardthrough Gmail	I can register and accessdashboard with Gmail	Medium	Sprint-2
	Login	USN-5	As a user, I can log into the application by entering email& password	I can login to the account inmy emaillogin.	High	Sprint-2
	Dashboard	USN-6	As a user ,I can use my accountin mydashboard for uploading dataset.	I can login to the accountfor uploading dataset.	Medium	Sprint-3
Customer (Webuser)	Website	USN-7	As a user ,I can use my dashboard in website	I can logininto the dashboard by visitingwebsite.	Medium	Sprint-3
Customer Care Executive		USN-8	As a user ,I can contactCustomer care Executive for my login.	I can contact customerexecutive for my login.	High	Sprint-4
Administrator		USN-9	As a user ,I can contact administrator for my queries.	I cancontact administratorforsolving my queries.	High	Sprint-4
Exploration	Dashboard	USN-10	As a user, I can preparedata by usingExploration Techniques.	I can prepare databy using Exploration Techniques.	High	Sprint-3

Presentation	Dashboard	USN-11	As a user, I can Present data in my dashboard.	I can present data by using my account in dashboard.	High	Sprint-4
Visualization	Dashboard	USN-12	As a user, I can Prepare Data by usingVisualization Techniques.	I can preparedata by usingVisualizationTechniques.	High	Sprint-3

PROJECT PLANNING & SCHEDULING

SPRINT PLANNING & SCHEDULING

Sprint	Functional Requirement(Epic)	User StoryNumber	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Retrieve Data	USN-1	As a user, I should get clearer clinical context for AIDS patient's uniquecase	10	Medium	Sanjay.V Naveen.A
Sprint-1	Visualize the data	USN- 2	As a user,I neednicely visualized dashboard of number of beds occupied and number of free beds in hospital.	20	High	Sanjay.V Naveen.A
Sprint-2	Track of patientvisit of Hospital	USN-3	Tracking a patient Health care overyears ofvisitand Screening	10	Medium	Sanjay.V Nave

			of data theyhave in hospital.			en.A Sivasubramaniyan.A
Sprint -2	Dashboard	USN - 4	As a user , I want the interactive dashboard to analyze the data. Have the data in termsof Graph.	20	High	Sanja y.V Nave en.A Sivasubramaniyan.A
Sprint-3	Detailed EHR's ofpatient	USN-5	Provided greater details in the EHR's of individual patient with clear idea of what to do.	10	Medi um	Sanja y.V Nave en.A Mohamed SalmanMydeen.A
Sprint- 3	Story Creation	USN-6	As a user , I need the story animation of the dataset with insights	20	High	Sanja y.V Nave en.A Mohamed SalmanMydeen.A
Sprint-4	Predict LOS	USN-7	As a user, I want theflawless system to predictthelength of stay of the patients	20	High	Sanja y.V Nave en.A

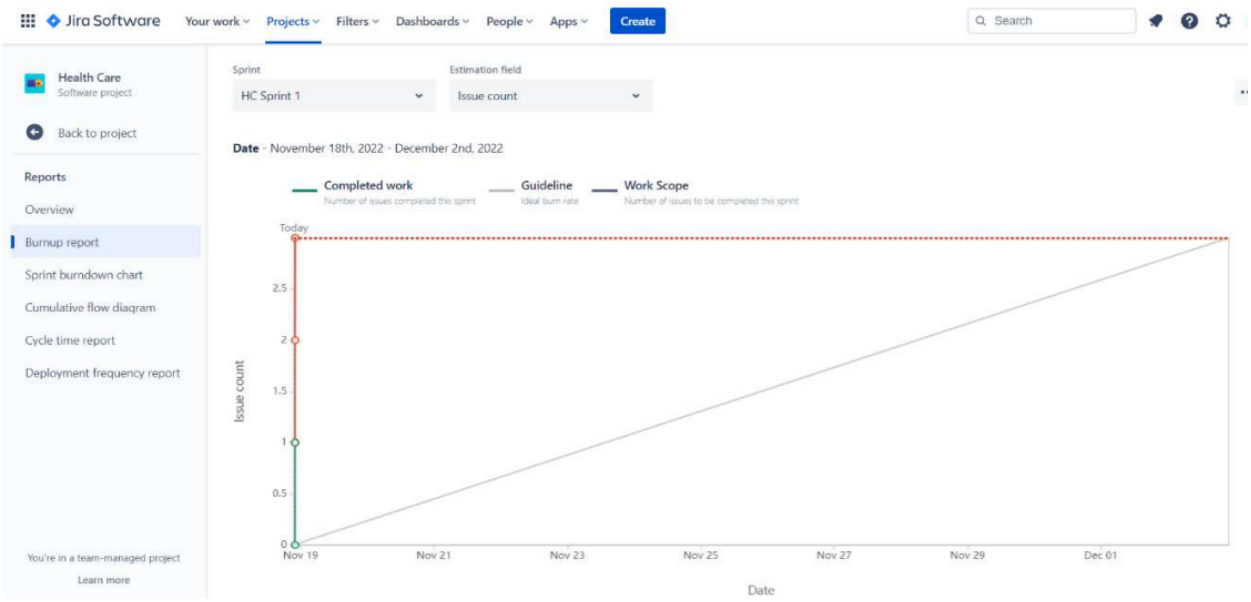
Sprint-4	Using ML algorithm for Prediction	USN-8	As a user,I need prior knowledge of LOS can aid in logistics such as room and bedallocation planning.	20	High	Sanjay.V Naveen.A

SPRINT DELIVERYSCHDULE

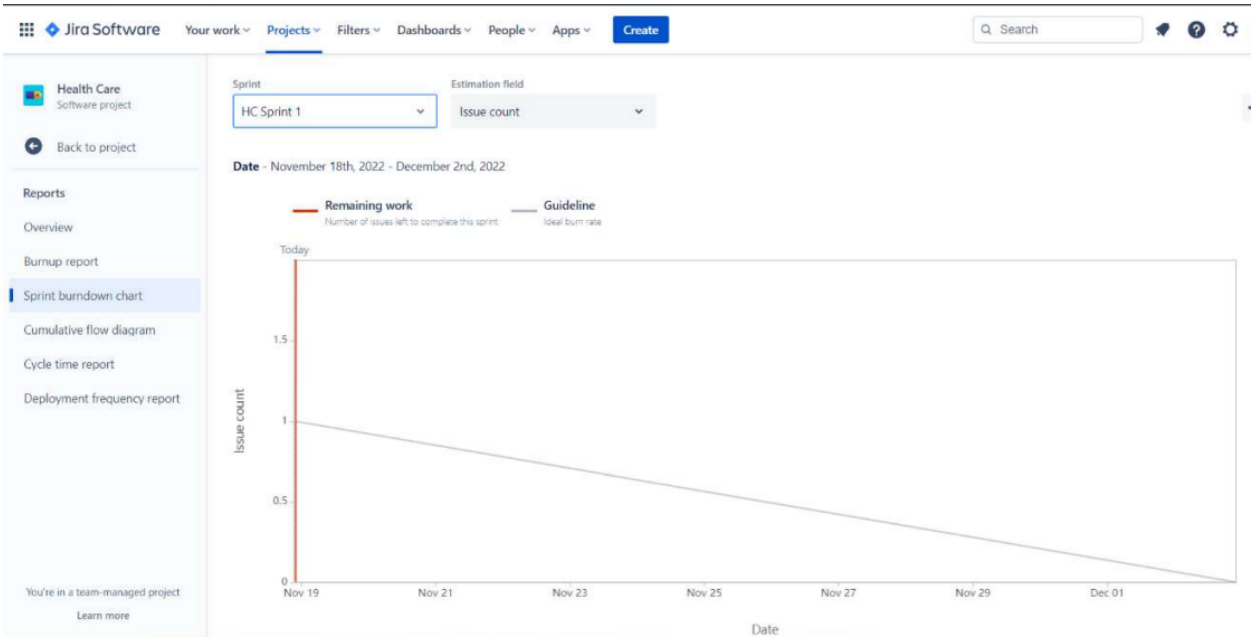
Sprint	Total Story Points	Duration	Sprint StartDate	Sprint End Date(Planned)	Story Points Completed (as on Planned End Date)	Sprint ReleaseDate (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12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

REPORTS FROM JIRA

BURNUP CHART

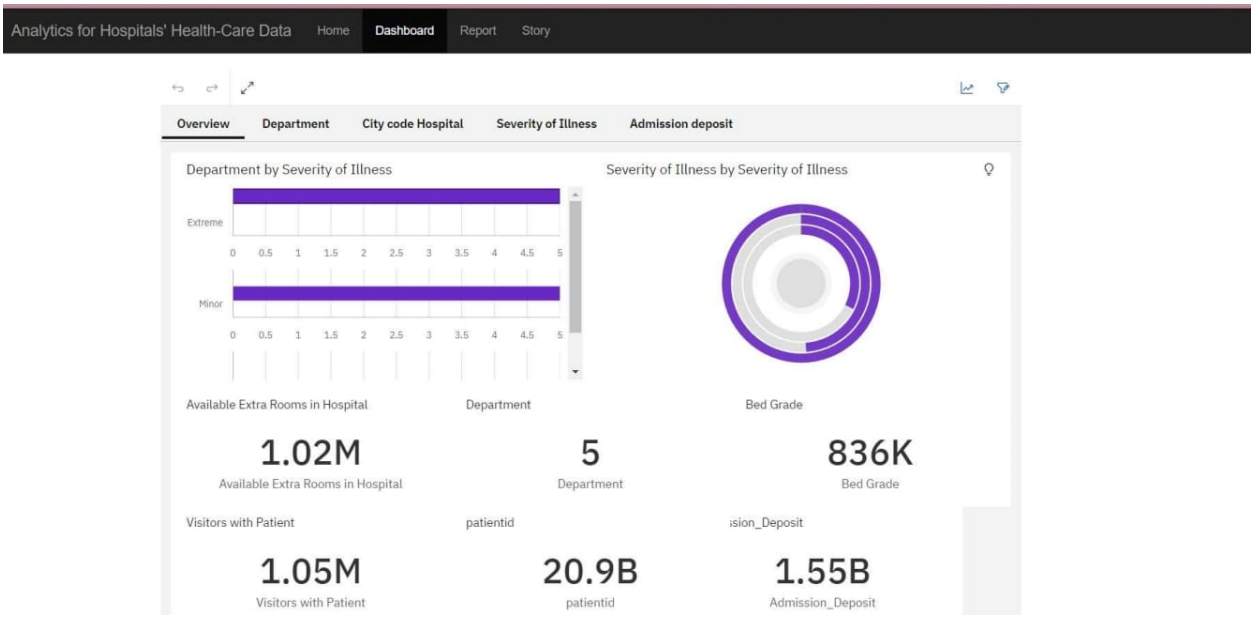


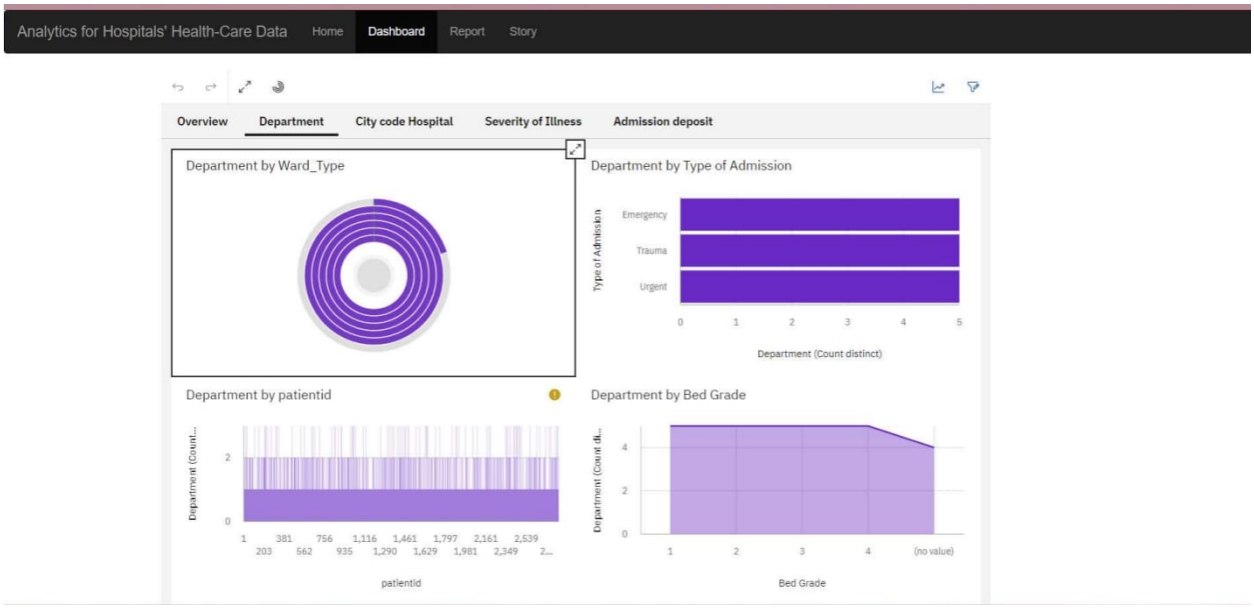
BURNDOWN CHART

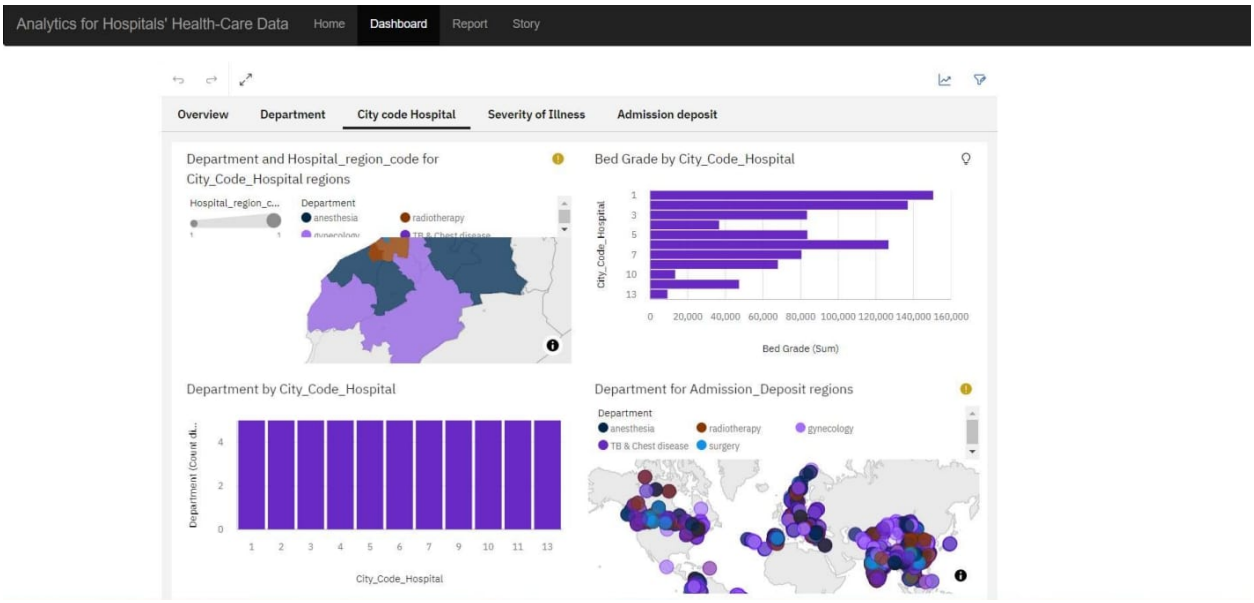


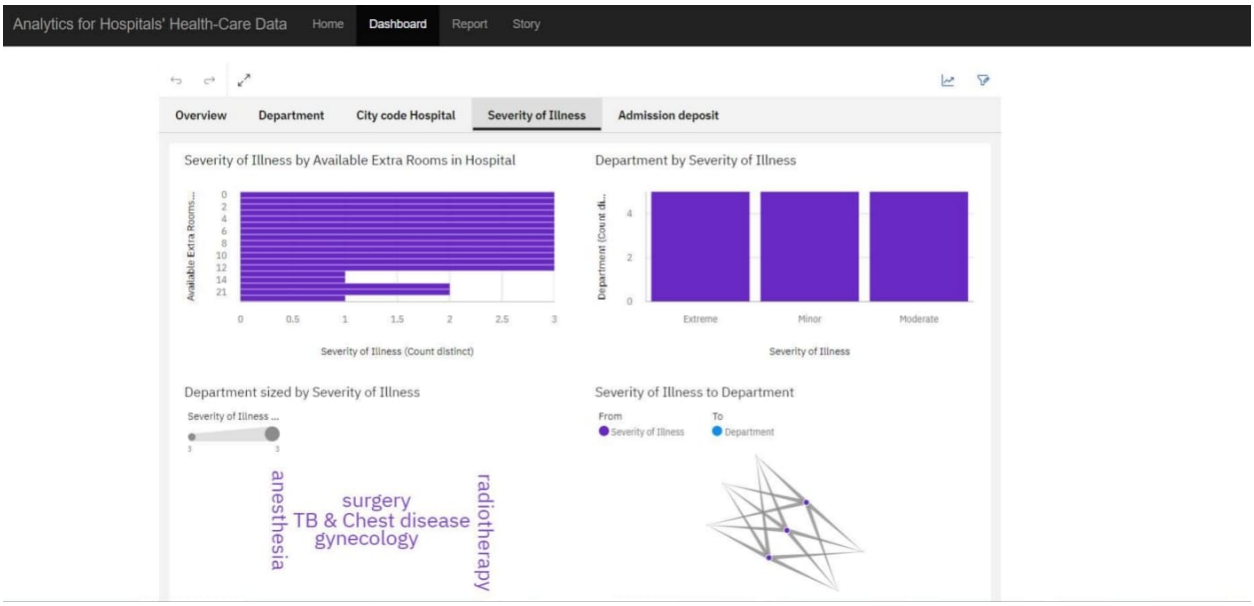
CODING AND SOLUTIONING

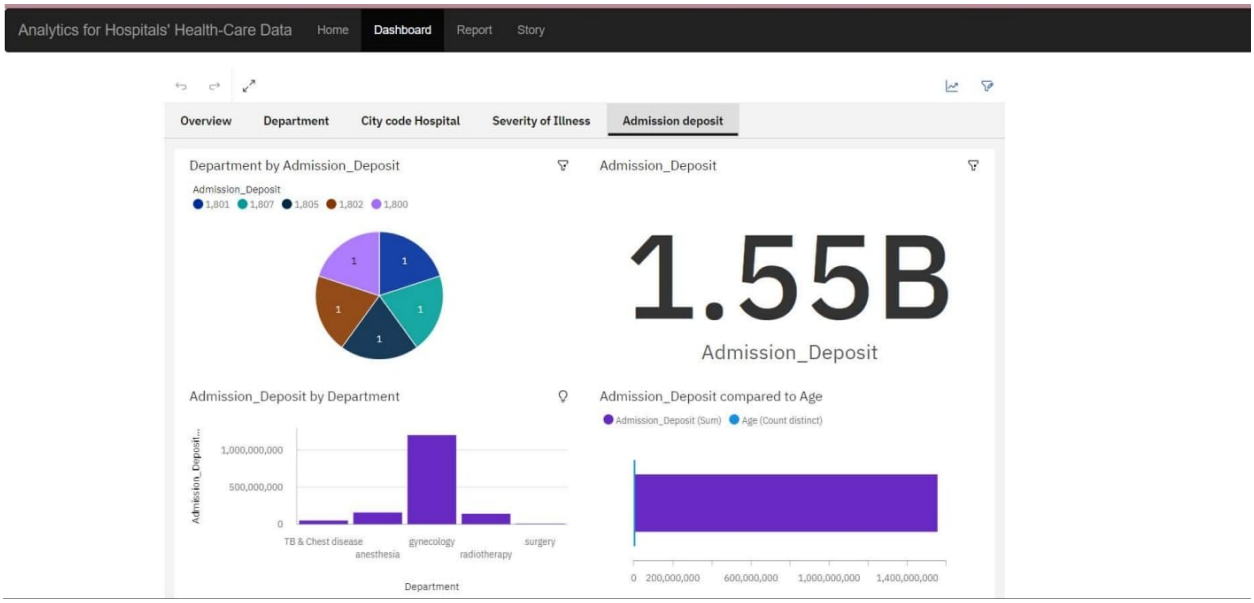
FEATURE 1
STORYBOARD CREATION



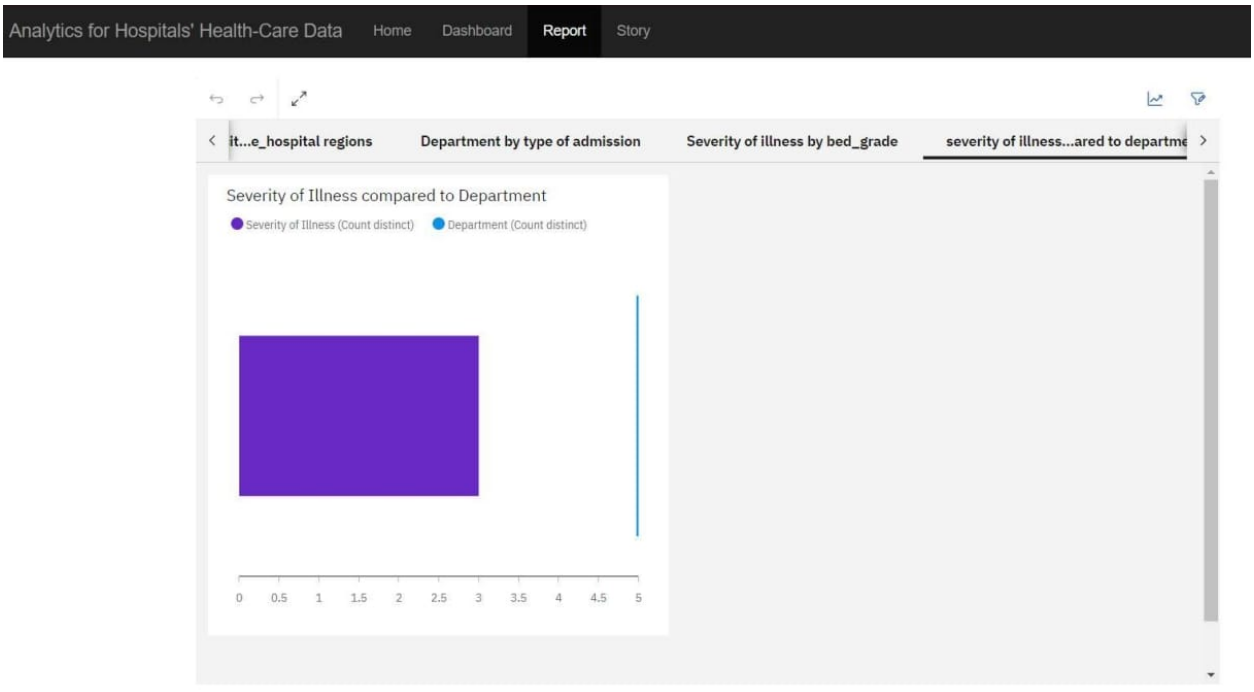


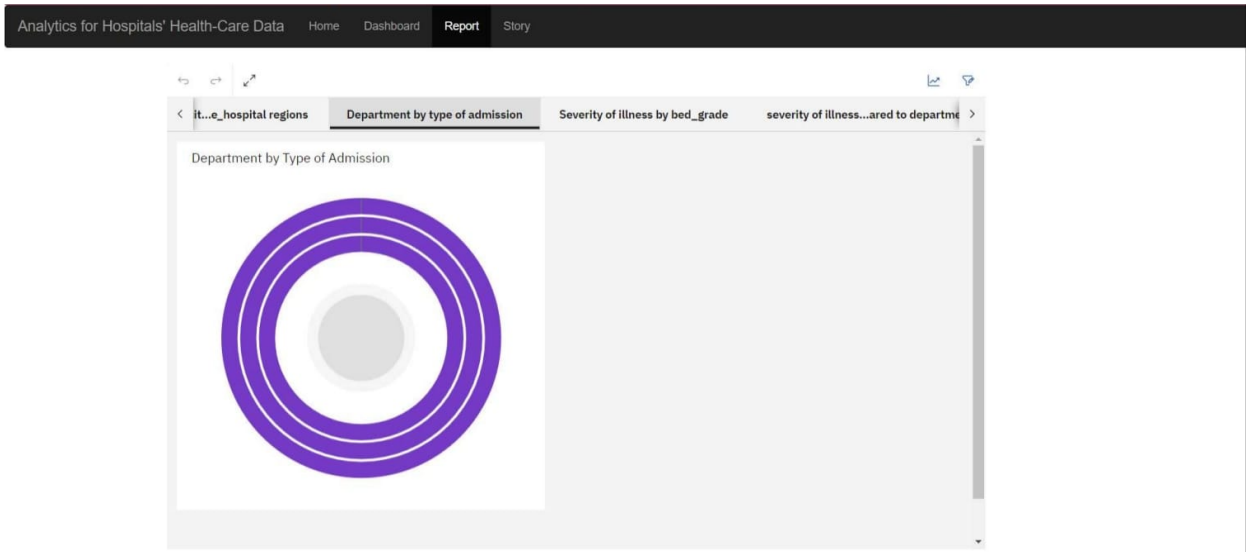

Analytics for Hospitals' Health-Care Data
Home
Dashboard
Report
Story
Overview
Department
City code Hospital
Severity of Illness
Admission deposit
Department and Hospital_region_code for City_Code_Hospital regions
Hospital_region_c...
Department
anesthesia
radiotherapy
Bed Grade by City_Code_Hospital
City_Code_Hospital
Bed Grade (Sum)
Department by City_Code_Hospital
Department (Count di...
City_Code_Hospital
Department for Admission_Deposit regions
Department
anesthesia
radiotherapy
gynecology
TB & Chest disease
surgery

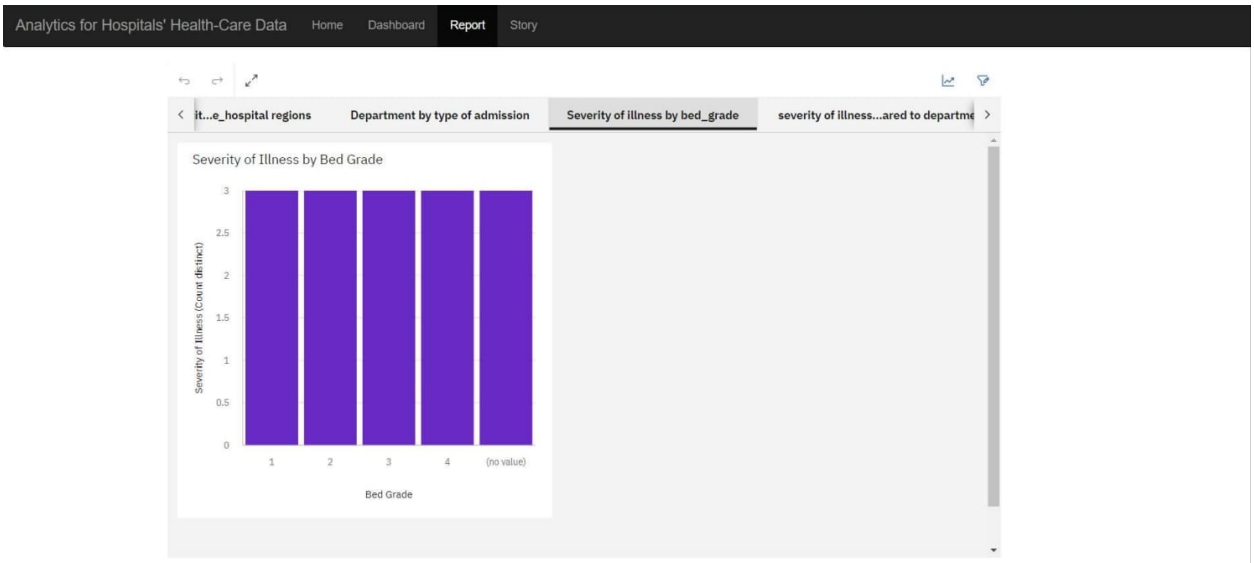

Analytics for Hospitals' Health-Care Data
Home
Dashboard
Report
Story
Overview
Department
City code Hospital
Severity of Illness
Admission deposit
Severity of Illness by Available Extra Rooms in Hospital
Available Extra Rooms...
Severity of Illness (Count distinct)
Department by Severity of Illness
Department (Count di...
Extreme
Minor
Moderate
Severity of Illness
Department sized by Severity of Illness
Severity of Illness ...
anesthesia
surgery
TB & Chest disease
gynecology
radiotherapy
Severity of Illness to Department
From
Severity of Illness
To
Department

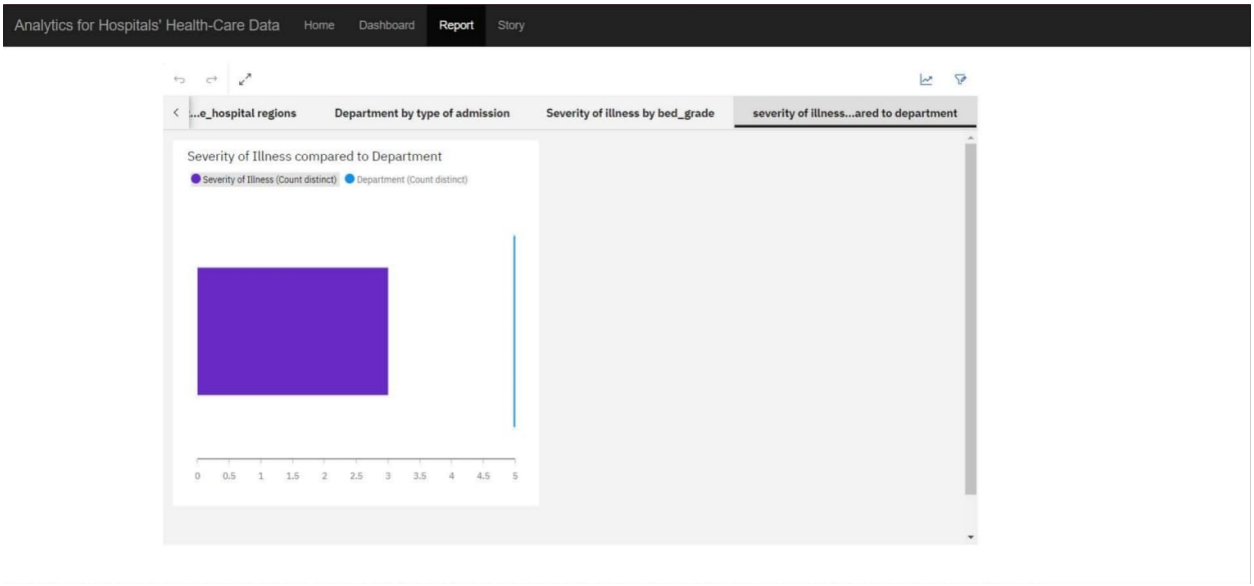


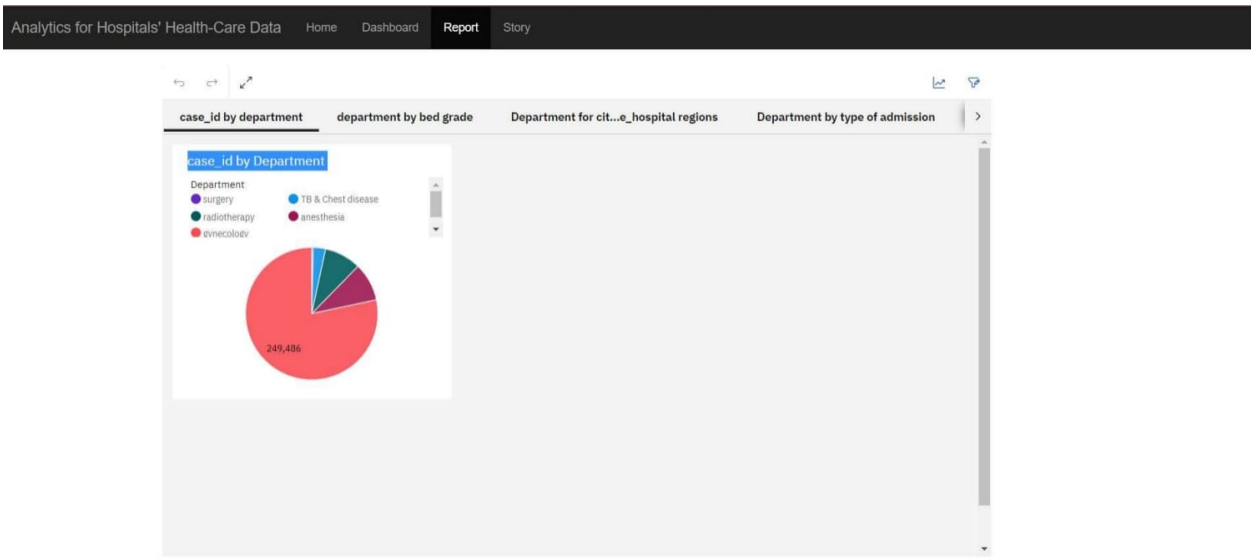
FEATURE 2- REPORT



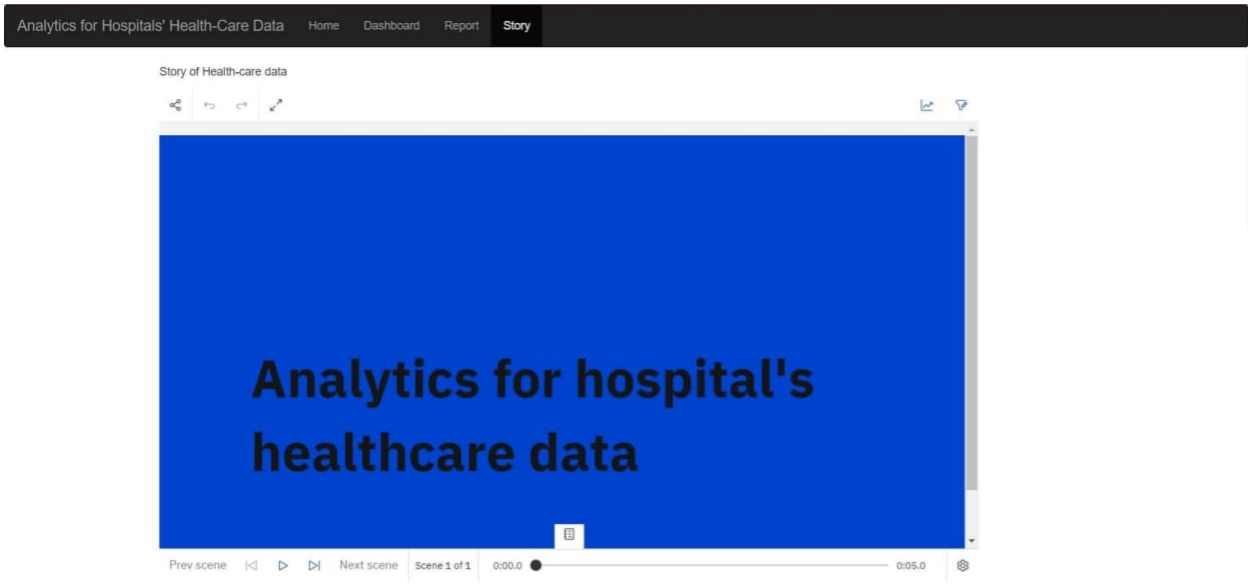

Analytics for Hospitals' Health-Care Data
Home
Dashboard
Report
Story
it...e_hospital regions
Department by type of admission
Severity of illness by bed_grade
severity of illness...ared to departme
Department by Type of Admission

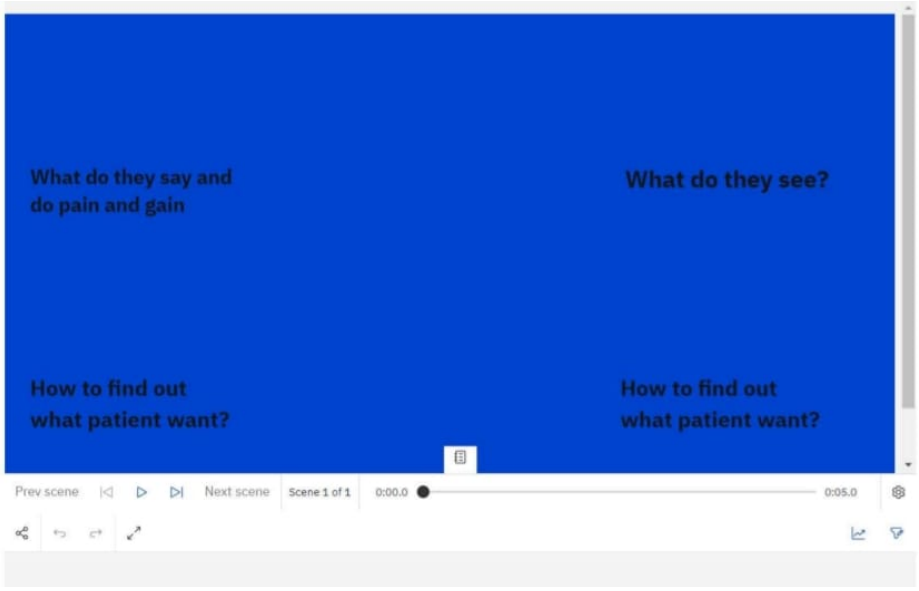

Analytics for Hospitals' Health-Care Data
Home
Dashboard
Report
Story
it...e_hospital regions
Department by type of admission
Severity of illness by bed_grade
severity of illness...ared to departme
Severity of Illness by Bed Grade
Severity of Illness (Count distinct)
3
2.5
2
1.5
1
0.5
0
1
2
3
4
(no value)
Bed Grade

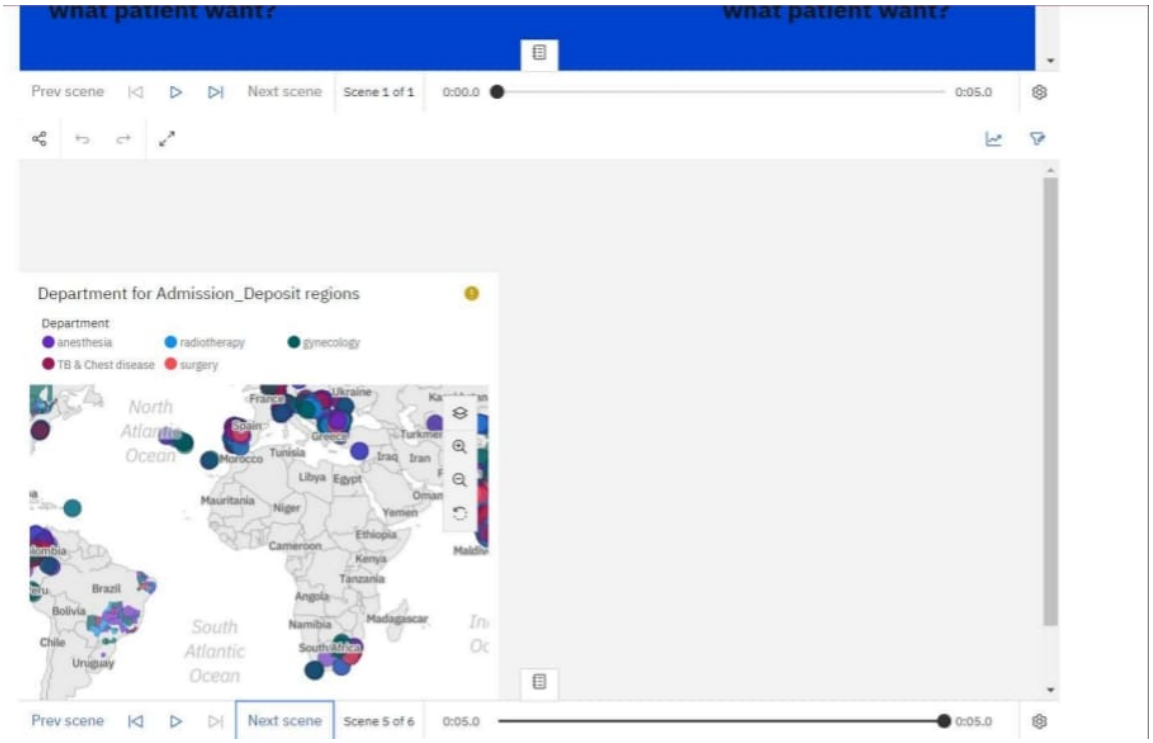

Analytics for Hospitals' Health-Care Data
Home
Dashboard
Report
Story
...e_hospital regions
Department by type of admission
Severity of illness by bed_grade
severity of illness...ared to department
Severity of Illness compared to Department
Severity of Illness (Count distinct)
Department (Count distinct)
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5

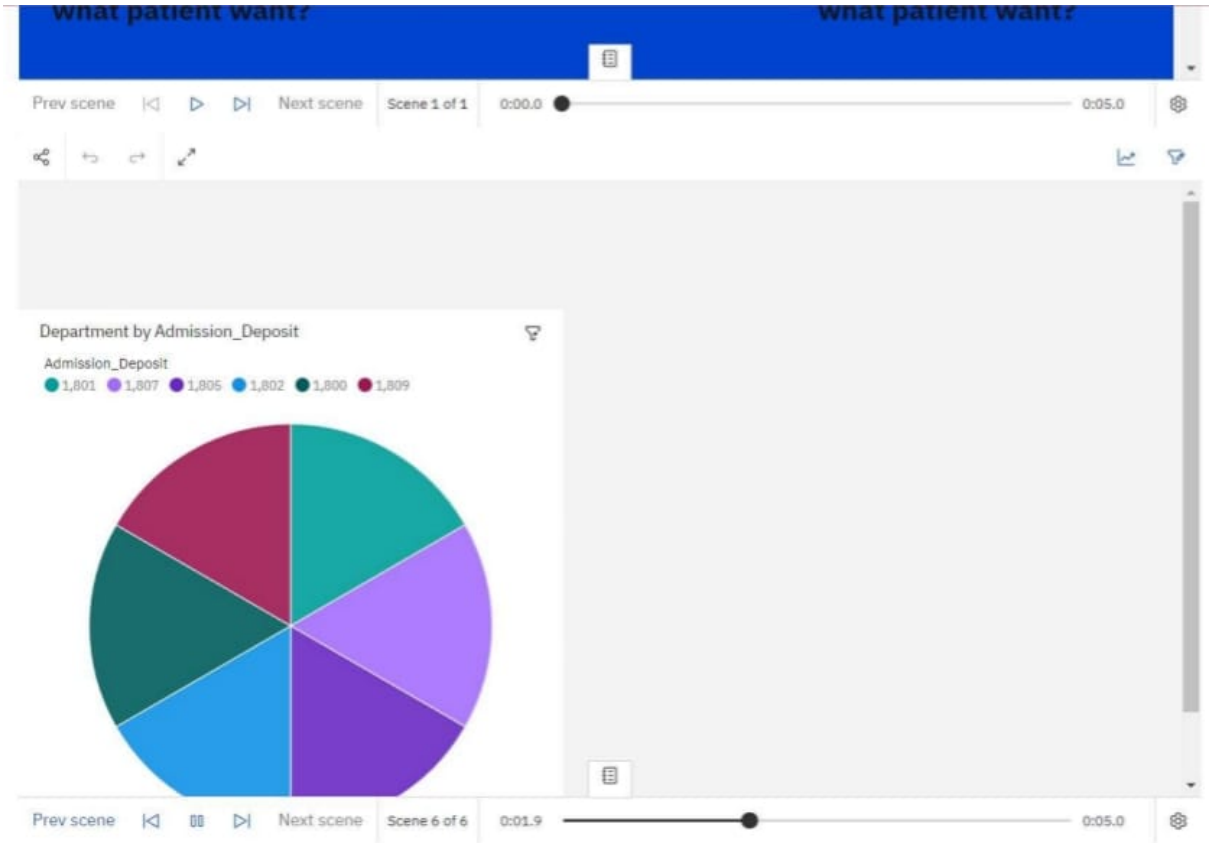

Analytics for Hospitals' Health-Care Data
Home
Dashboard
Report
Story
case_id by department
department by bed grade
Department for cit...e_hospital regions
Department by type of admission
case_id by Department
Department
surgery
TB & Chest disease
radiotherapy
anesthesia
249,486

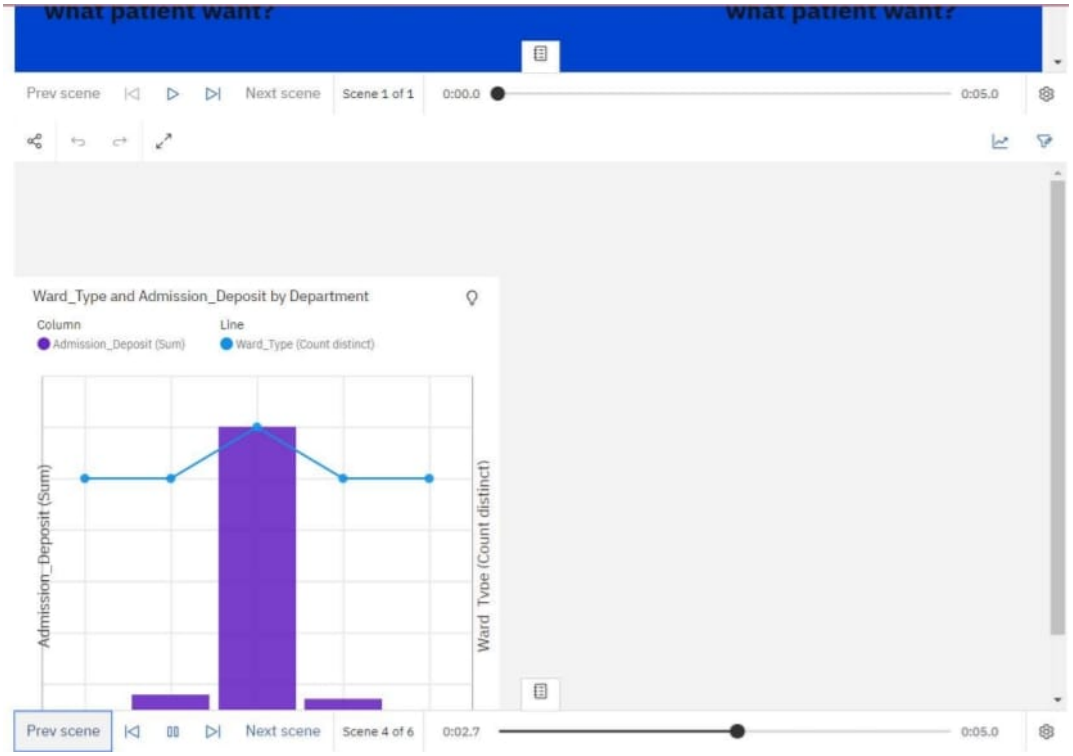
FEATURE-STORY

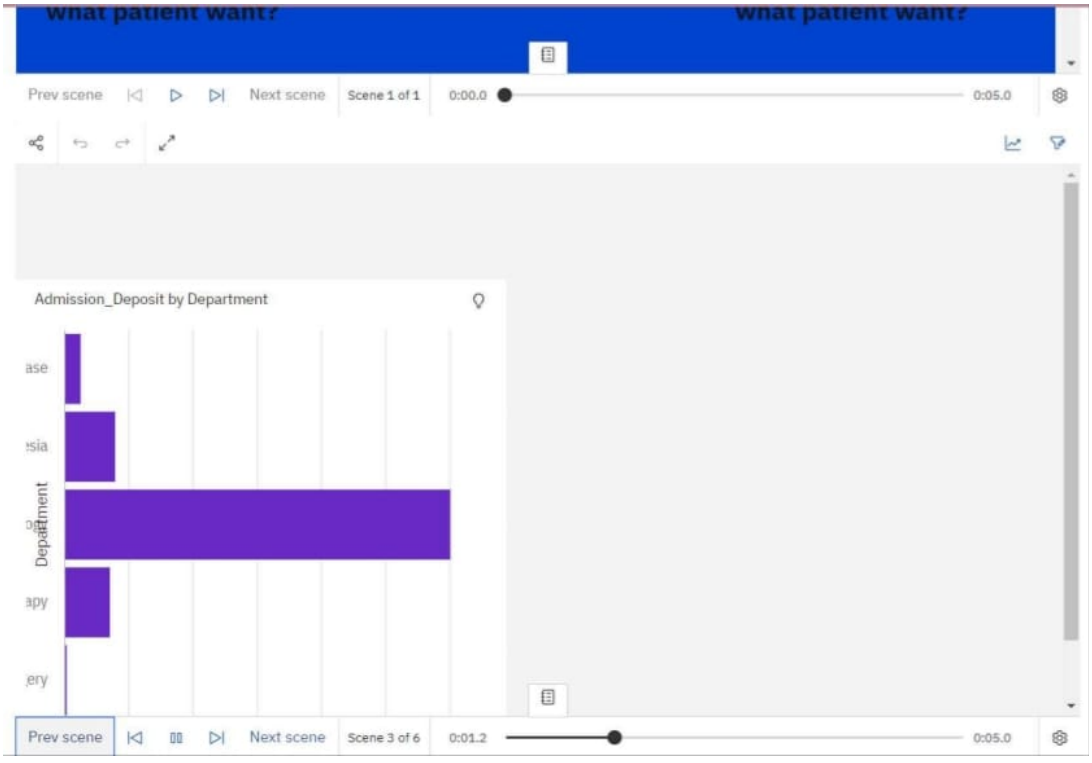





what patient want?
what patient want?
Prev scene
Next scene
Scene 1 of 1
0:00.0
0:05.0
Department for Admission_Deposit regions
Department
anesthesia
radiotherapy
gynecology
TB & Chest disease
surgery
North Atlantic Ocean
South Atlantic Ocean
France
Ukraine
Spain
Greece
Turkmen
Tunisia
Iraq
Iran
Morocco
Libya
Egypt
Oman
Mauritania
Niger
Yemen
Ethiopia
Cameroon
Kenya
Tanzania
Angola
Namibia
Madagascar
South Africa
Brazil
Bolivia
Chile
Uruguay
Prev scene
Next scene
Scene 5 of 6
0:05.0
0:05.0


what patient want?
what patient want?
Prev scene
Next scene
Scene 1 of 1
0:00.0
0:05.0
Department by Admission_Deposit
Admission_Deposit
1,801
1,807
1,805
1,802
1,800
1,809
Prev scene
Next scene
Scene 6 of 6
0:01.9
0:05.0


what patient want?
what patient want?
Prev scene
Next scene
Scene 1 of 1
0:00.0
0:05.0
Ward_Type and Admission_Deposit by Department
Column
Admission_Deposit (Sum)
Line
Ward_Type (Count distinct)
Admission_Deposit (Sum)
Ward_Type (Count distinct)
Prev scene
Next scene
Scene 4 of 6
0:02.7
0:05.0


what patient want?
what patient want?
Prev scene
Next scene
Scene 1 of 1
0:00.0
0:05.0
Admission_Deposit by Department
ase
sia
apy
ery
Prev scene
Next scene
Scene 3 of 6
0:01.2
0:05.0

TESTING

BI.TESTING

- verify user is able tosee home page
- verify user is able to see dashboardpage
- verify user is able to naivigateto story page
- verify filters are working

USER ACCEPTANCE TESTING

1.PURPOSE OF DOCUMENT

The purpose of this document is to briefly explain the test coverage and open issuesof the [ProductName] project at the time of the release to User Acceptance Testing(UAT).

2.DEFECT ANALYSIS

This report showsthe number of resolved or closed bugs at each severity level, and how they were resolved

Resolution	Severity 1	Severity2	Severity3	Severity4	Subtotal
By Design	8	5	0	3	16
Duplicate	1	0	4	0	7
External	0	3	5	1	5
Fixed	13	4	3	18	32
NotReproduced	0	1	0	1	2
Skipped	1	2	0	0	1
Won't Fix	0	5	2	1	8
Totals	23	14	13	26	75

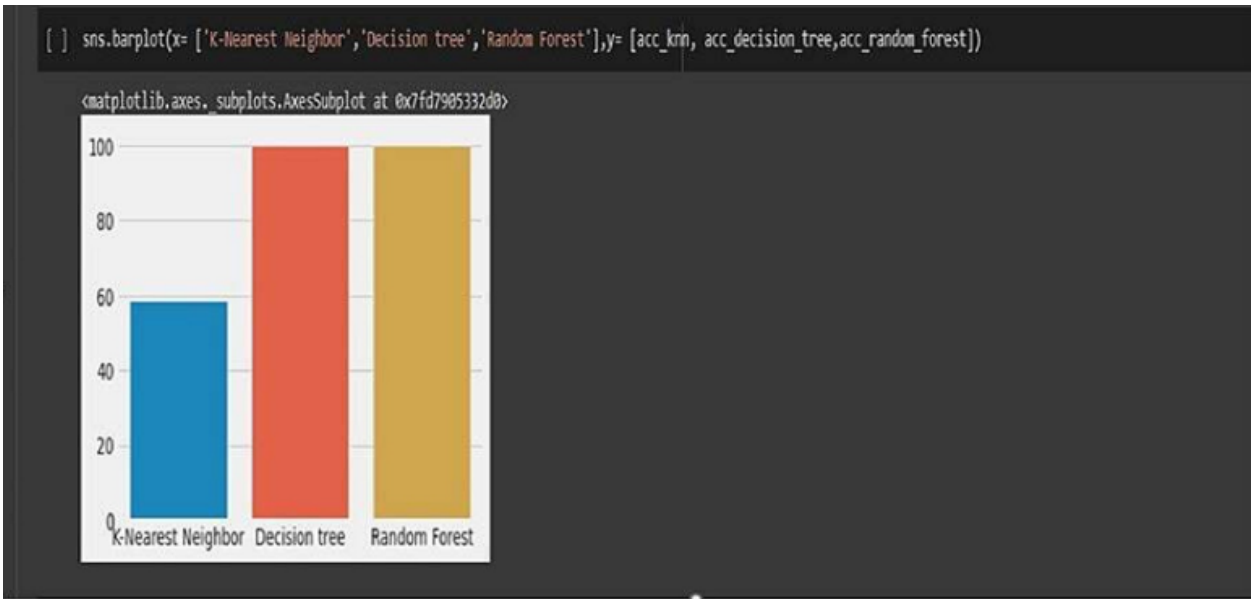
3.TEST CASE ANALYSIS

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

Section	Total Cases	Not Tested	Fail	Pass
Print Engine	6	0	0	6
Client Application	51	0	0	51
Security	1	0	0	1
Outsource Shipping	3	0	0	3
Exception Reporting	6	0	0	6
Final Report Output	2	0	0	2
Version Control	1	0	0	1

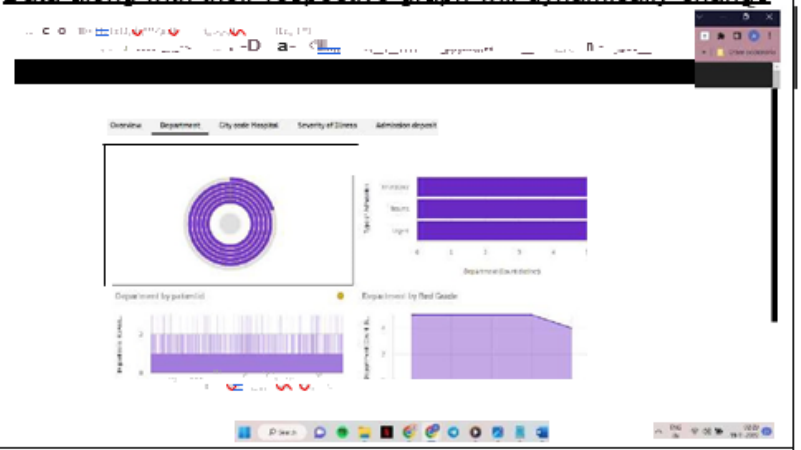
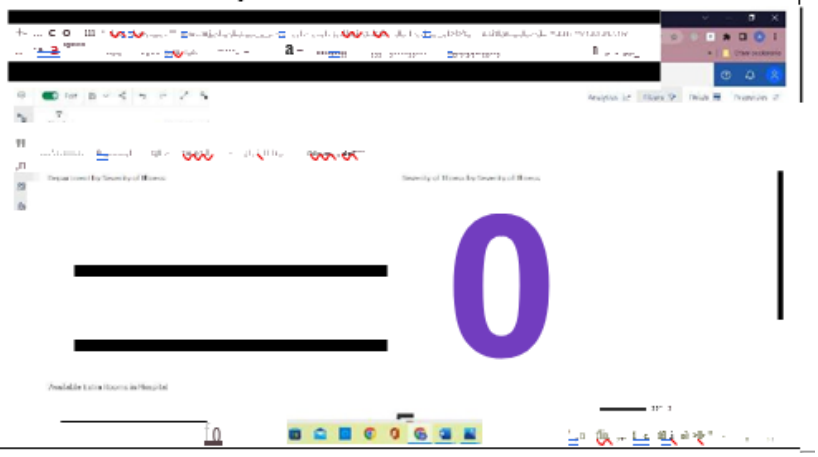
RESULT

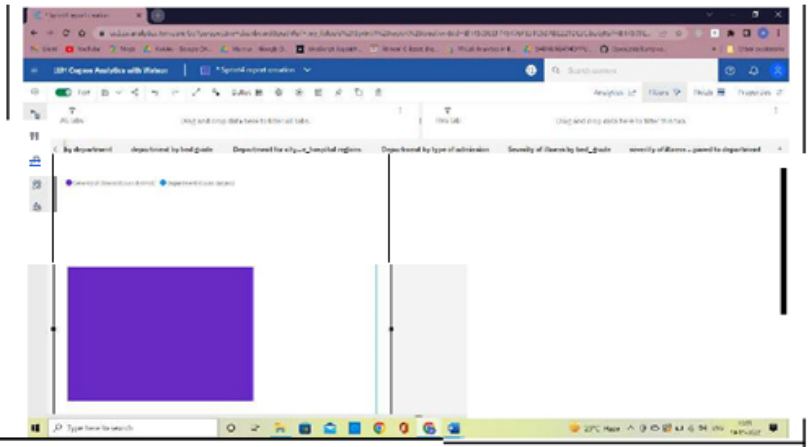
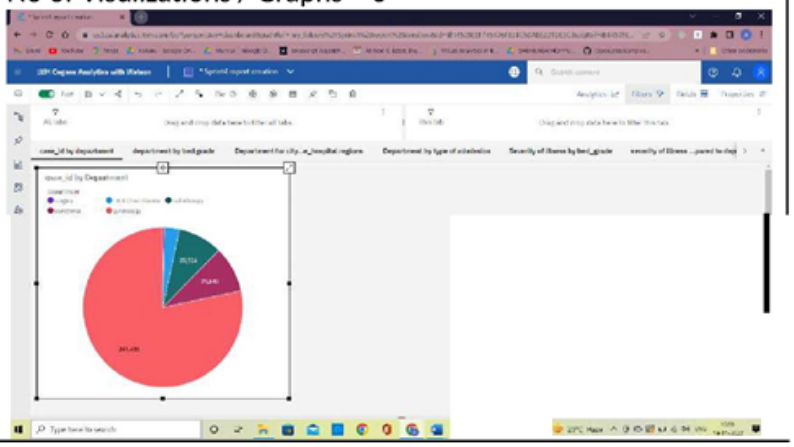
PERFORMANCE METRICS



Model Performance Testing:

Project team shall till the following information in model performance testing template

S.No.	Parameter	Screenshot / Values
1.	Dashboard design	Number of Visulizations / Graphs - 24 No of tabs-  ö
2.	Data Responsiveness	Data along with their respective graph will dynamically change 
3.	Amount Data to Rendered (DB2 Metrics)	Number of rows read-318438 Number of rows loaded-318438 Number of rows rejected-a 

4.	Utilization of Data Filters	We created filters for Dashboards 
5.	Effective User Story	No of Scene Added -6 Animations are perfectly displayed Images are perfectly rendered. 
6.	Descriptive Reports	No of Visulizations / Graphs – 6 

ADVANTAGE AND DISADVANTAGES

ADVANTAGES:

- Cost-effective use of technology
- Improved project management
- Boosting hospital capacity
- Enhance the qualityand efficiency of healthcare
- benefit areas like emergency preparation, charting, administration,compliance, and financial management.
- Analysing clinicaldata to improvemedical research
- Using patient data to improve health outcomes
- Gaining operational insights from healthcareprovider data
- Improved staffing throughhealth business management analytics
- Early detection of disease.

- Prevention of unnecessary doctor's visits.
- Discovery of new drugs.
- More accurate calculation of health insurancerates.
- 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 challengein applying big data in healthcare. Big datastorage is usually targets of hackers. This endangers the safety of medical data.Healthcare organisations are very much concerned about the safety of patients' sensitivepersonal data. For this, all healthcare applications must meet therequirement for data security and be HIPAA compliant before they can be deployedfor healthcare services.

PRIVACY:

One of the major drawbacks in the application of big data in healthcare industry is the issueof lack of privacy. Application of big data technologies involves monitoring of patient's data, tracking of medical inventoryand 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 itviolets 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.

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 effective manner . However, the role of data analytics in improving patientoutcomes 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 patientexperiences. Patient-centered healthcare depends on knowing what patients want and need. Data analytics holds the key to unlockingthis vital Information.

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 profileswithin a system, then it can alert doctorsto trends in data that may havebeen 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 takingharmful 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.

APPENDIX

SOURCE CODE

HOME PAGE:

Index.html

```
<!DOCTYPE html>
<html lang="en">
<head>
  <title>Data Analytics</title>
  <meta charset="utf-8">
  <meta name="viewport"
content="width=device-width,
initialscale=l">
  <link rel="stylesheet"
href="https:/ / maxcdn.bootstra pcdn.com/
bootstrap/3.4.1/css/bootstrap.min.css">
  <script
src= "https:// ajax .googleapis.coma/jax/ libs/ jque ry/ 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="about.html">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: PNT2022TMID33091
</b></i></h4></center>

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

```
                    <tbody>
                      <tr>
                        <td>Team Leader</td>
                        <td>SANJAY.V</td>
</tr>
<tr>
                        <td>Team member</td>
                        <td>SIVASUBRAMANIYAN.A</td>
</tr>
<tr>
                        <td>Team member</td>
                        <td>NAVEEN.A</td>
</tr>
                          <tr>
                        <td>Team member</td>
                        <td>MOHAMED SALMAN MYDEEN.A</td>
                    </tr>
                    </tbody>
                    </table>
                    </body>
                    </html>
```

About page;

about.html

```
<!DOCTY
PE html>
<html lang="en">
<head>
  <title>Data Analytics</title>

  <meta charset="utf-8">
  <meta name="viewport" content=
"width=device-width, initialscale=1">
  <link rel="stylesheet"
href="https://maxcdn.bootstrapcdn.com/bootstra
p/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/bo
otstrap.mi n.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' HealthCare
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><br>

Recent Covid-19Pandemic has raised alarms over
one of the mostoverlooked areas to focus:
HealthcareManagement.
While healthcare management has various use
cases for using data science,patient length of
stay is one criticalparameter to observe and
```

predictif one wants to improve the efficiency of the healthcare management in a hospital. This parameter helps hospitals to identify patientsof high LOS-risk (patients who will stay longer) at the time of admission. Once identified, patients with high LOSrisk can have their treatmentplan optimized to minimize LOS and lower the chance of staff/visitor infection. Also, priorknowledge of LOS can aid in logisticssuch as roomand 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.

Goal:

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 resourceallocation and better functioning. The length of stay isdivided into 11 different classes rangingfrom 0-10 days to more than

100 days.

```
<br>

<br><br>

<b>Technical Architecture: </b>
<br>
<br><br>
<img src= "Technical_architecture.png">
</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=l">
  <link   rel="stylesheet"
```

```
  href="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/css/b
  oot
  <script
  src="https://ajax.googleapis.com/ajax/libs/jquery/3.6.0/jquery.
  min.js"></script>
  <script
  src="https://maxcdn.bootstrapcdncom/bootstrap/3.4.1/js/boots
  trap.min.js"><f s


</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://us3.ca.analytics.ibm.com/bi/?perspective=dashboa
rd&amp;pathRef=.my_fol
dded&amp;action=view&amp;mode=dashboard&amp;subView
=model000001848bea4a5eallowfullscreen=""></iframe>
</div>


</body>
</html>
```

REPORT PAGE:

```
  <!DOCTYPE html>
<html lang="en">
<head>
 <title>Data Analytics</title>
  <meta charset="utf-8">
  <meta name="viewport" content="width=device-width, initial-
  scale=l">
  <link rel="stylesheet" href="htt ps://
  maxcdn.bootstrapcdn.com/ bootstrap/ 3.4.1/css/boot
  <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"><f s
</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://us3.ca.analytics.ibm.com/
bi/?perspective=dashboard&amp;pathRef=.my_fold
embedded&amp;action=view&amp;mode=dashboard&amp;sub
View=model000001848e4
```

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

STORY:

```
<!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/boot
  strc
  <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/bootstr
  ap.min.js"></scriJ
</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">
<p>Story of Health-care data</p>
<iframe
src="https://us3.ca.analytics.ibm.com/bi/?perspective=story&amp;pat
hRef=.my_folders%2F
mbedded&amp;action=view&amp;sceneId=model00000l848e2967c2
_00000002&amp;sceneallowfullscreen=""></iframe>
</br>
</div>
<div class="container">
<iframe
src="https://us3.ca.analytics.ibm.com/bi/?perspective=story&amp;pa
thRef=.my_folders%2F
embedded&amp;action=view&amp;sceneId=model00000l848e2f216
f_00000000&amp;scem allowfullscreen=""></iframe>
</br>
```

<iframe src="https://us3.ca.analytics.ibm.com/bi/?perspective=story&pathRef=.my_folders%2Fembedded&action=view&sceneId=model00000l848e33666e_00000002&scen allowfullscreen=""></iframe>

</br>

</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_Grade
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.(

Analysis of dataset

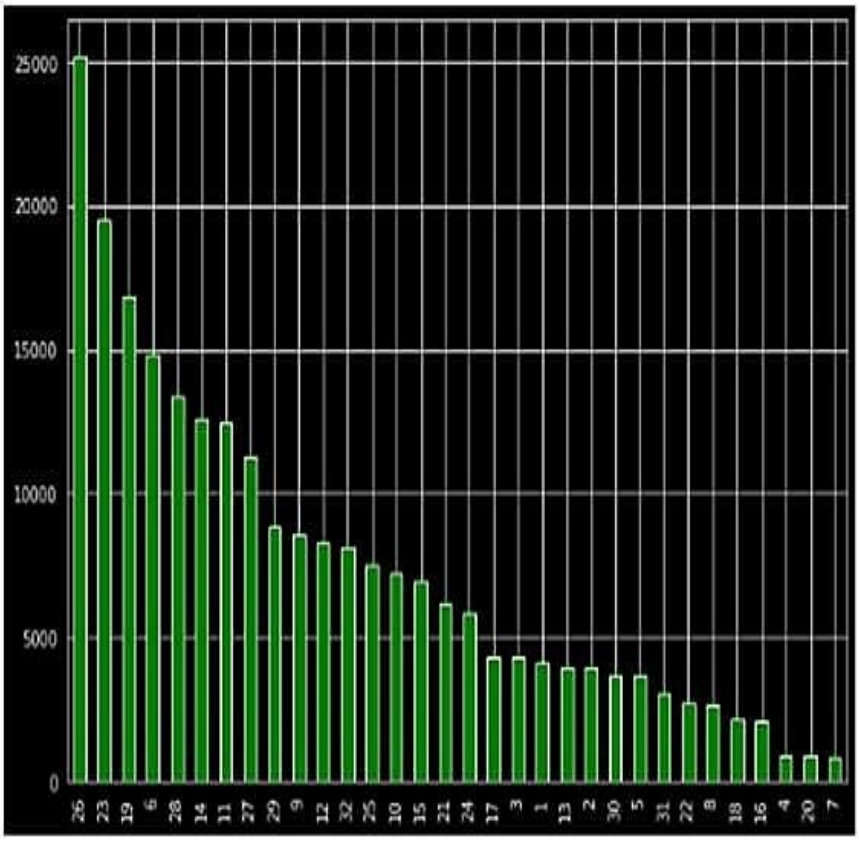
Distribution of values

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
```

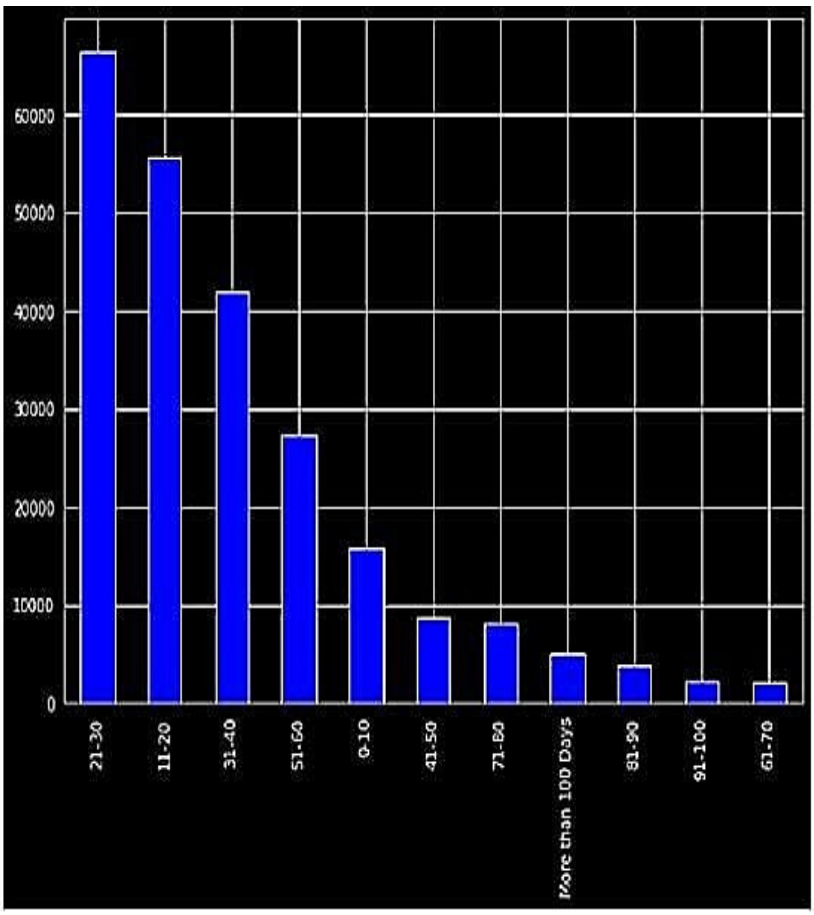
```
plt.figure(figsize=(10,7))
train.Hospital_code.value_counts().plot(kind="bar", color = ['green'])
```



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
```



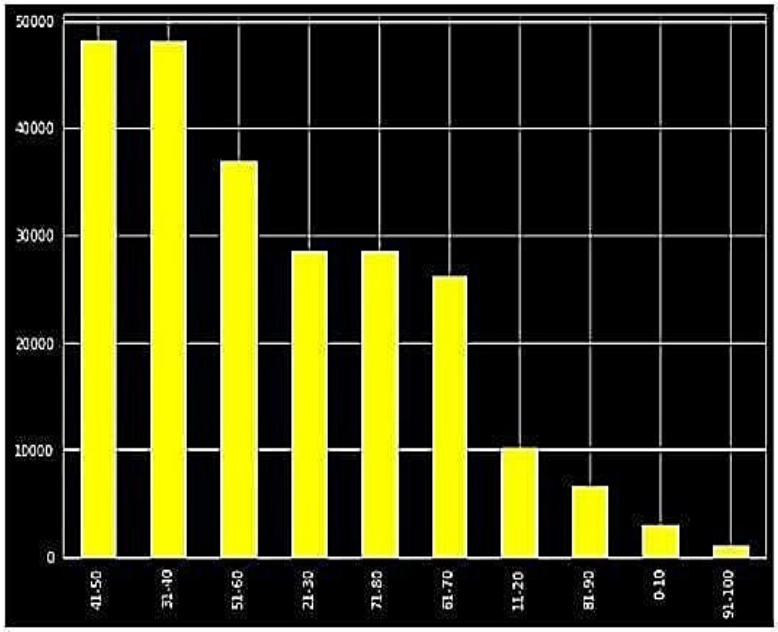
Age

```
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'])
```



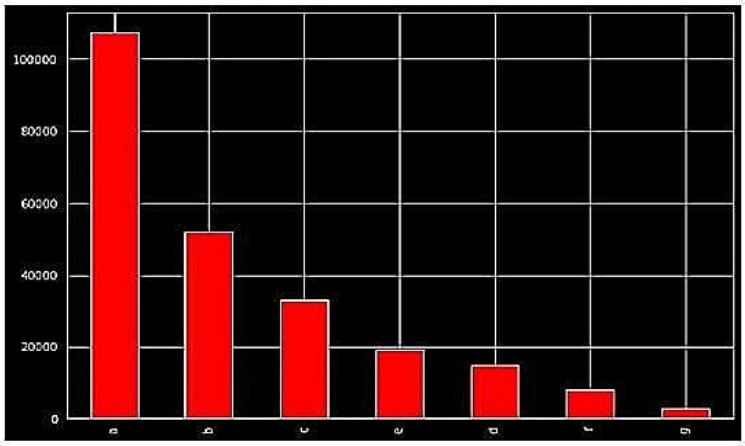
Hospital_type_code

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

```
a    107545
b     51925
```

```
c     32995
e     19105
d     14833
f      8166
g      2748
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'])
```

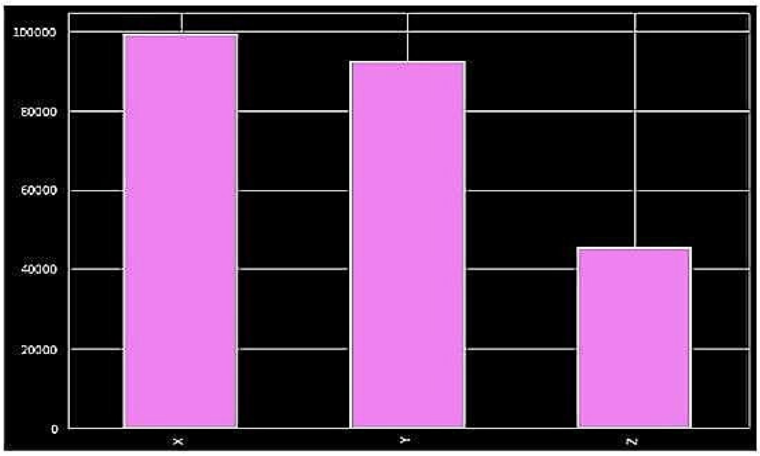


Hospital_region_code

```
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'])
```



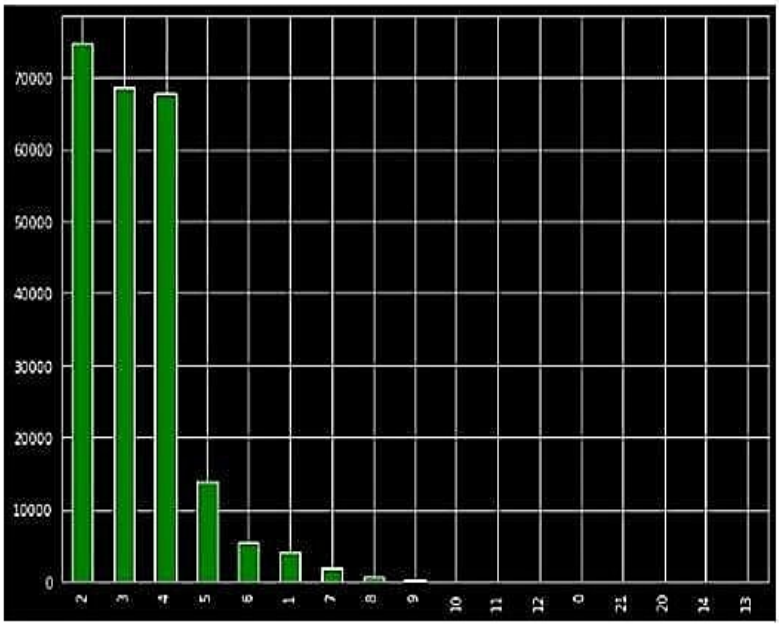
Available_Extra_Rooms_in_Hospital

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

```
2     74877
3     68517
4     57755
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'])
```



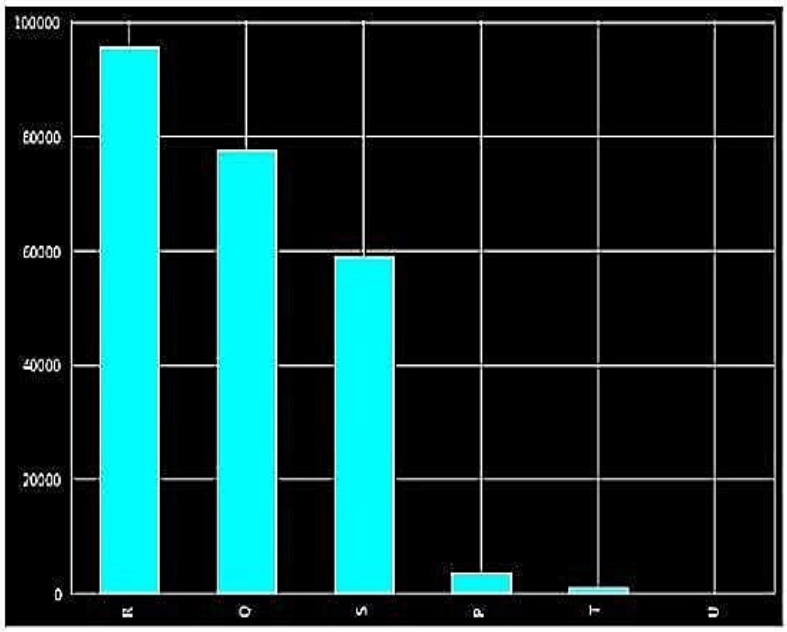
Department

```
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'])
```



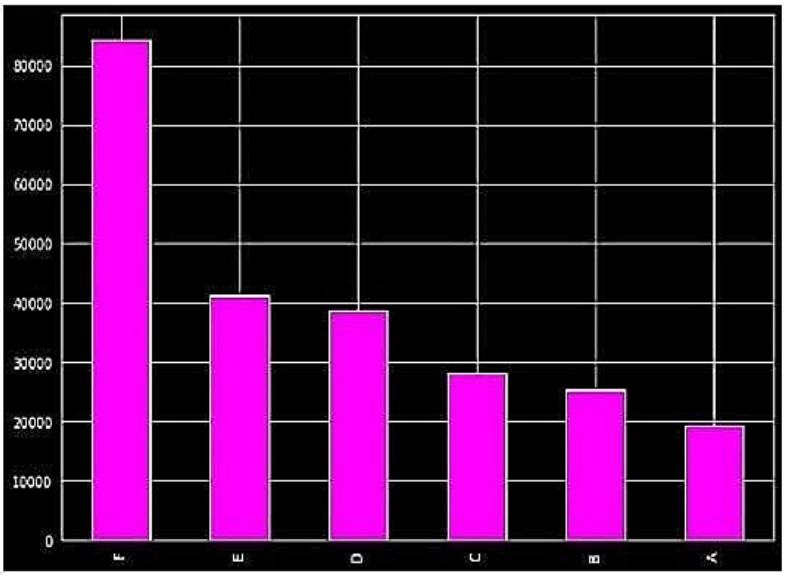
Ward_Facility_Code

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

```
F    84438
E    41245
```

```
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'])
```



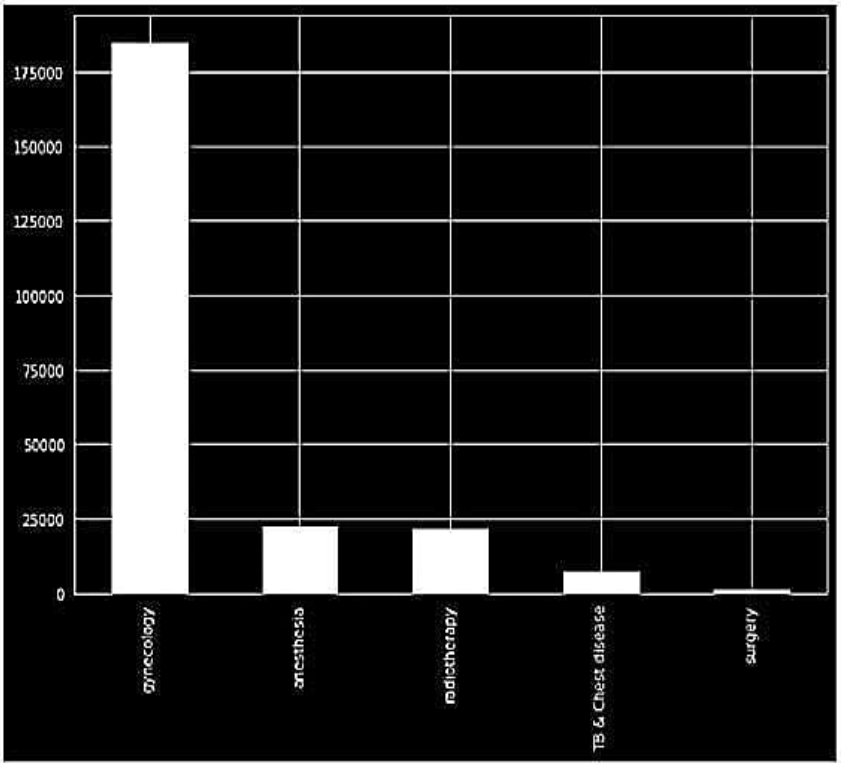
Visitors_with_Patient

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

```
2.0    103037
4.0     59058
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'])
```

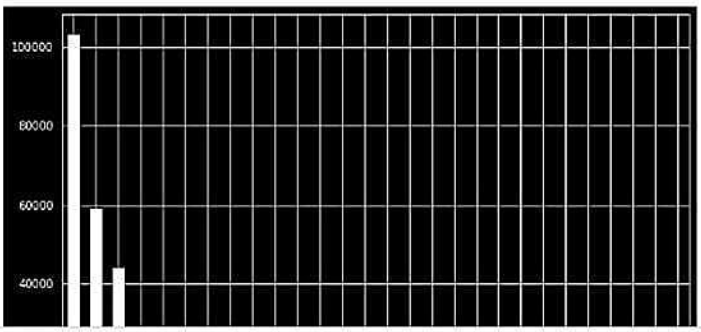


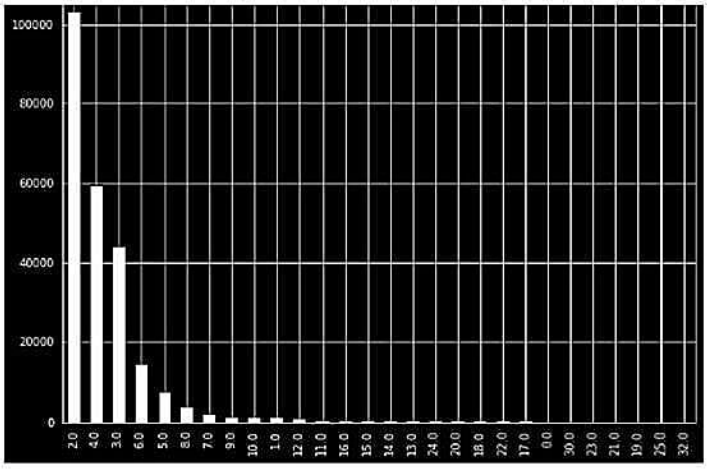
Ward_Type

```
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
```

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





Severity of Illness

```
train.Severity_of_Illness.value_counts()
```

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

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

```
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
```

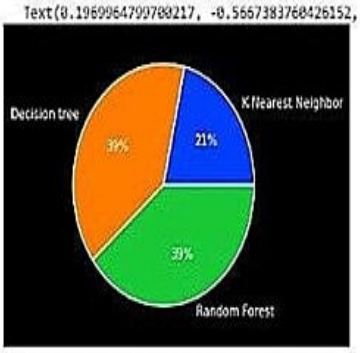
Prediction accuracy comparison

```
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(-0.9277499883745313, 0.590999244932723, 'Decision tree'),
  Text(0.361160213278373l7, -1.0390203560781281, 'Random Forest')],
 [Text(0.4706412895080693, 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%%')
```

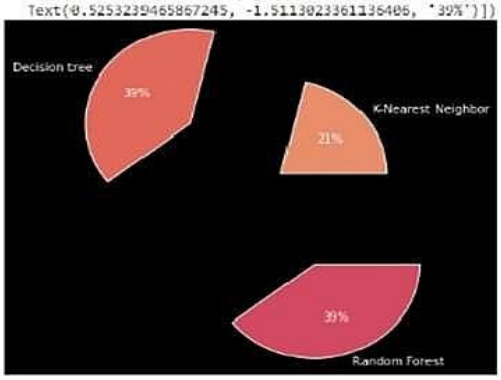
```
..
  elif(p[0]==8):
    print("The predicted LOS of patient is : 81-90")
  elif(p[0]==9):
    print("The predicted LOS of patient is : 91-100")
  elif(p[0]==10):
    print("The predicted LOS of patient is : More than 100 Days")
```

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

```
prediction(p)
```

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

```
 ],
[Text(0.8706863857564283, 0.6884803683859842, 'K-Nearest Neighbor'),
 Text(-1.7711589159877414, 1.1282712857806532, 'Decision tree'),
 Text(0.689487679895076, -1.0835843161491535, 'Random Forest')],
[Text(0.4784853110913704, 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):
```

GITHUB LINKS:

GitHub link: https://github.com/IBM-EPBL/IBM-Project-38824-1660385799.git