1.INTRODUCTION

1.1 Project overview

Recent Covid-19 Pandemic has raised alarms over one of the most overlooked areas to focus: Healthcare Management. While healthcare management has various use cases for using data science, patient length of stay is one critical parameter to observe and predict if one wants to improve the efficiency of the healthcare management in a hospital. This parameter helps hospitals to identify patients of high LOS-risk (patients who will stay longer) at the time of admission. Once identified, patients with high LOS risk can have their treatment plan optimized to minimize LOS and lower the chance of staff/visitor infection. Also, prior knowledge of LOS can aid in logistics such as room and bed allocation planning. Suppose you have been hired as Data Scientist of Health Man – a not for profit organization dedicated to manage the functioning of Hospitals in a professional and optimal manner. 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.

2. LITERATURE SURVEY

2.1 Existing Problem

- The already existing model is trained with minimal parameters by leaving the necessary parameter
- Low accuracy in prediction
    - No feature extraction
    - High complexity.

2.2 References

1. Yang J.-J., Li J., Mulder J., Wang Y., Chen S., Wu H., Wang Q., Pan H. Emerging information technologies for enhanced healthcare. *Compute. Ind.* 2015;69:3–11.doi:10.1016/j.compind.2015.01.012. [Cross Ref] [Google Scholar]
2. Cortada J.W., Gordon D., Lenihan B. *The Value of Analytics in Healthcare.* IBM Institute for Business Value; Armonk, NY, USA: 2012. Report No.: GBE03476USEN-00. [Google Scholar]
3. Center for Medicare and Medicaid Services. [(accessed on 1 August 2017)]; Available online:https://www.cms.gov/Research-Statistics-Data-andSystems/Statistics-TrendsandReports /National Health Expend Data/NationalHealthAccountsHistorical.html
4. Berwick D.M., Hackbarth A.D. Eliminating waste in US health care. *J. Am. Med. Assoc.* 2012;307:1513–1516. doi: 10.1001/jama.2012.362. [PubMed] [Cross Ref] [Google Scholar]
5. 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] [Cross Ref] [Google Scholar]
6. Prokosch H.-U., Ganslandt T. Perspectives for medical informatics. *Methods Inf. Med.* 2009;48:38–44. doi: 10.3414/ME9132. [PubMed] [Cross Ref] [Google Scholar]

7. 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] [Cross Ref] [Google Scholar]
8. 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] [Cross Ref] [Google Scholar]

9. 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. [Cross Ref] [Google Scholar]

10. 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

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

12. 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.

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

14. S. Axryd. Why 85% of big data projects fail. [Online]. Available: https://www.digitalnewsasia.com/insights/why-85-big-data-projects-fail. Accessed on: Apr. 16, 2019.

15. W.H. Organization et al., "How can hospital performance be measured and monitored?," in How can hospital performance be measured and monitored? 2003. p. 17–17.

16. Varabyova, J. Schreyögg
International comparisons of the technical efficiency of the hospital sector: panel data analysis of oecd countries using parametric and non-parametric approaches Health Policy, 112 (1) (2013), pp. 70-79
17. D. Heyland, D. Cook, S.M. Bagshaw, A. Garland, H.T. Stelfox, S. Mehta, P. Dodek, J.
Kutsogiannis, K. Burns, J. Muscedere, *et al.*
The very elderly admitted to icu: a quality finish? Crit
Care Med, 43 (7) (2015), pp. 1352-1360

18. Teno JM, Fisher E, Hamel MB, Wu AW, Murphy DJ, Wenger NS, et al. Decisionmaking and outcomes of prolonged icu stays in seriously ill patients. J Am Geriatr Soc 2000; vol. 48, no. S1.

19. A. Hunter, L. Johnson, A. Coustasse

Reduction of intensive care unit length of stay: the case of early mobilization Health
Care Manager, 33 (2) (2014), pp. 128-135

20. ve Maliyetleri S. ”Characteristics, outcomes and costs of prolonged stay icu patients;”2011.

21. A.A. Kramer, J.E. Zimmerman

A predictive model for the early identification of patients at risk for a prolonged intensive care unit length of stay
BMC Med Inform Decis Making, 10 (1) (2010), p. 27

22. A. Pérez, W. Chan, R.J. Dennis

Predicting the length of stay of patients admitted for intensive care using a first step analysis
Health Service Outcomes Res Method, 6 (3–4) (2006), pp. 127-138

23. 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

24. F. Barili, N. Barzaghi, F.H. Cheema, A. Capo, J. Jiang, E. Ardemagni, M. Argenziano, C. Grossi

An original model to predict intensive care unit length-of stay after cardiac surgery in a competing risk framework
Int J Cardiol, 168 (1) (2013), pp. 219-225

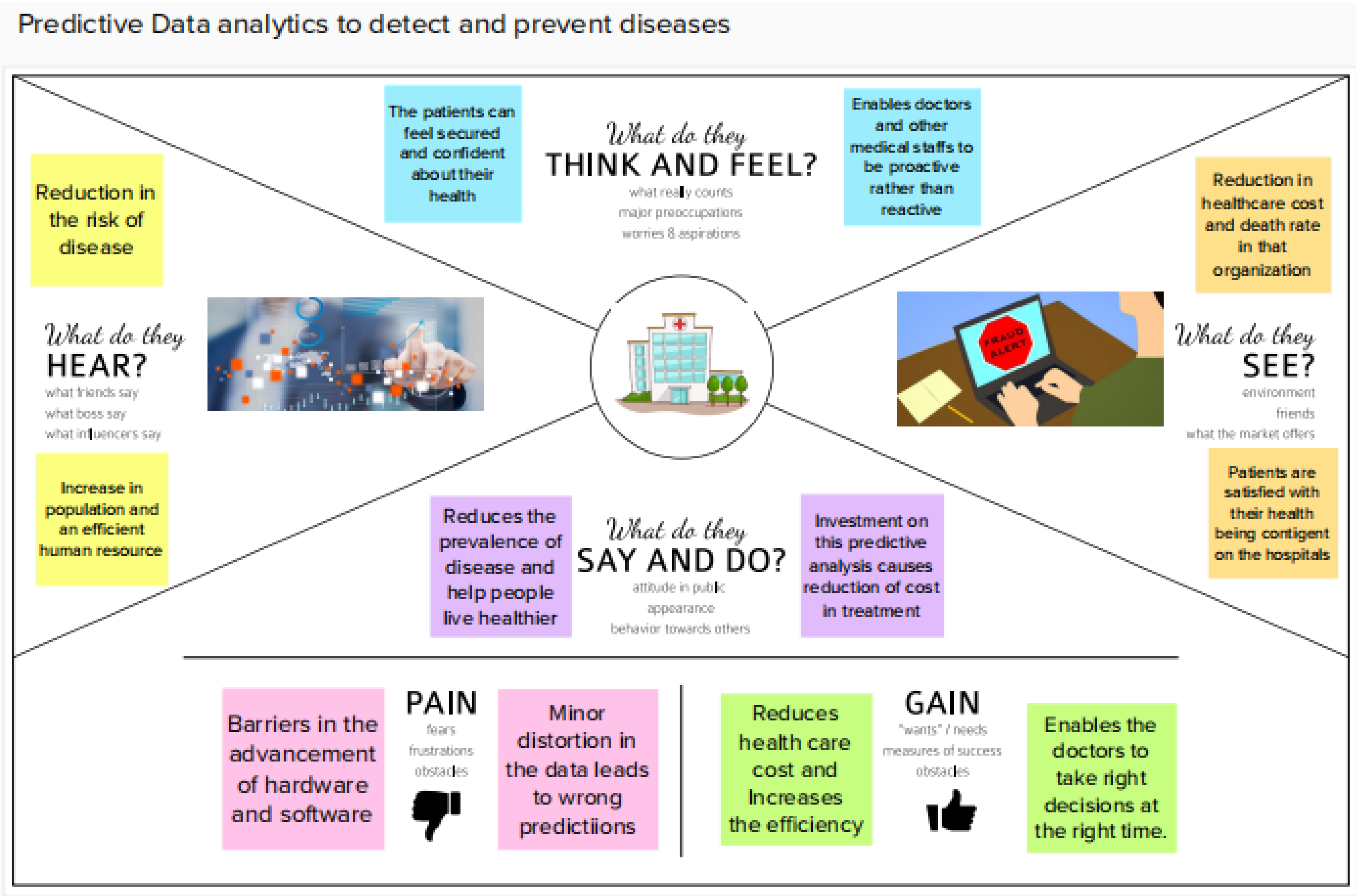
25. M.N. Diringer, N.L. Reaven, S.E. Funk, G.C. Uman
Elevated body temperature independently contributes to increased length of stay in neurologic intensive care unit patients Crit Care Med, 32 (7) (2004), pp.
1489-1495
26. R. Paterson, D. MacLeod, D. Thetford, A. Beattie, C. Graham, S. Lam, D. BellPrediction of in-hospital mortality and length of stay using an early warning scoring system: clinical audit
Clin Med, 6 (3) (2006), pp. 281-284

2.3 Problem Statement Definition

- The goal is to accurately predict the Length of Stay for each patient on case by case basis so that the Hospitals can use this information for optimal resource allocation and better functioning.
- The length of stay is divided into 11 different classes ranging from 010 days to more than 100 days.

3. IDEATION AND PROPOSED SOLUTION

3.1 Empathy Map Canvas



3.2 Ideation & Brainstorming

Brainstorm & Idea Prioritization Template:

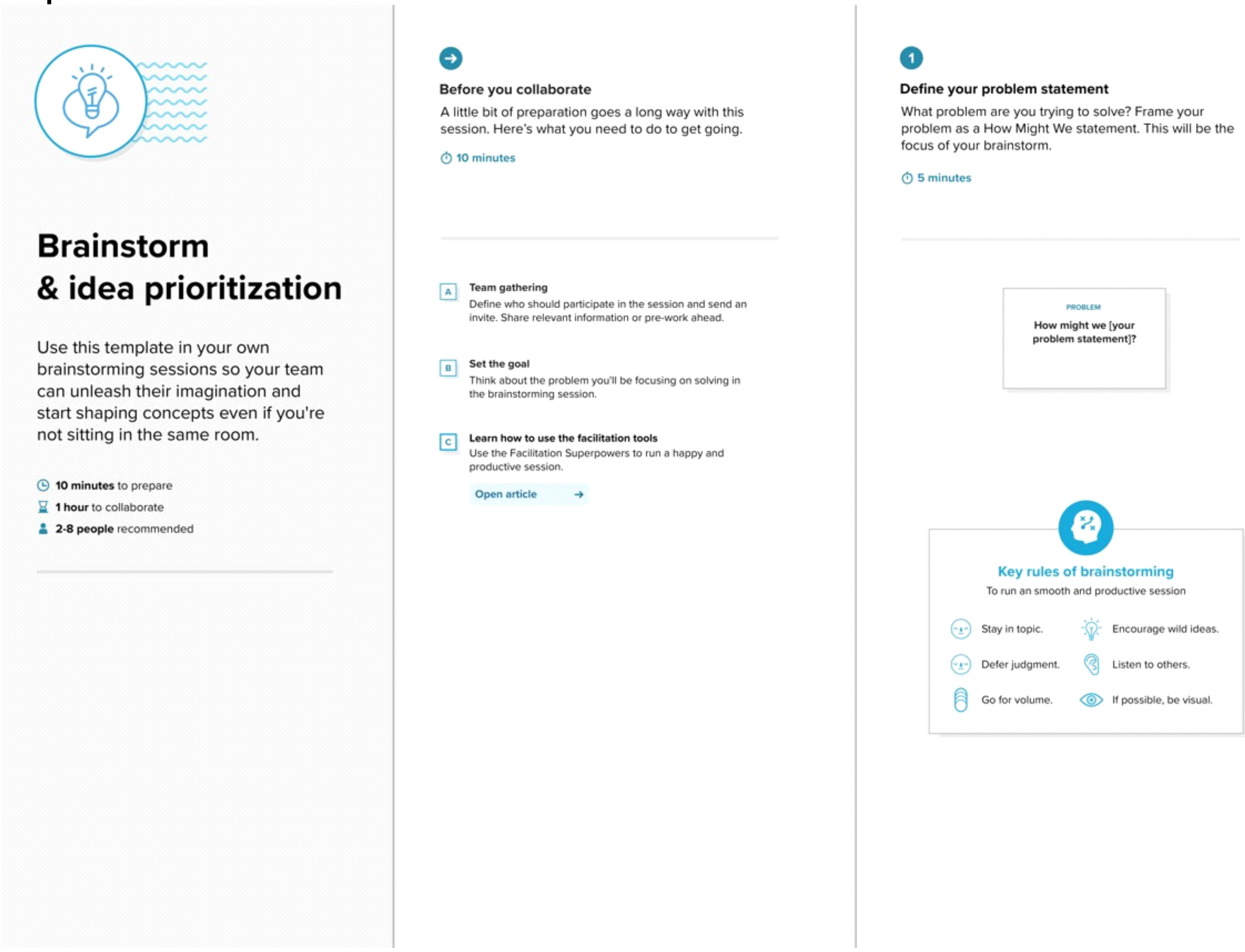
- Brainstorming provides a free and open environment that encourages everyone

Within a team to participate in the creative thinking process that leads to problem solving. Prioritizing volume over value, out-of-the-box ideas are welcome and built upon, and all participants are encouraged to collaborate, helping each other develop a rich amount of creative solutions.

- Use this template in your own brainstorming sessions so your team can unleash

their imagination and start shaping concepts even if you're not sitting in the same room.

Step-1: Team Manesh



ng, Collaboration and Select the Problem Statement

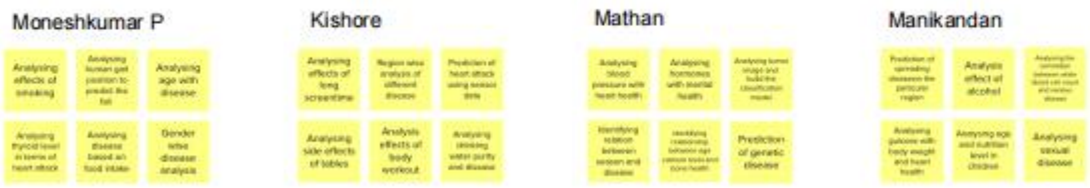
Step-2: Brainstorm, Idea Listing and Grouping

2

Brainstorm

Write down any ideas that come to mind that address your problem statement.

10 minutes



3

Group ideas

Use this space to group similar ideas from the brainstorm. Each group should have a title that describes what the ideas have in common. If a group is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

20 minutes

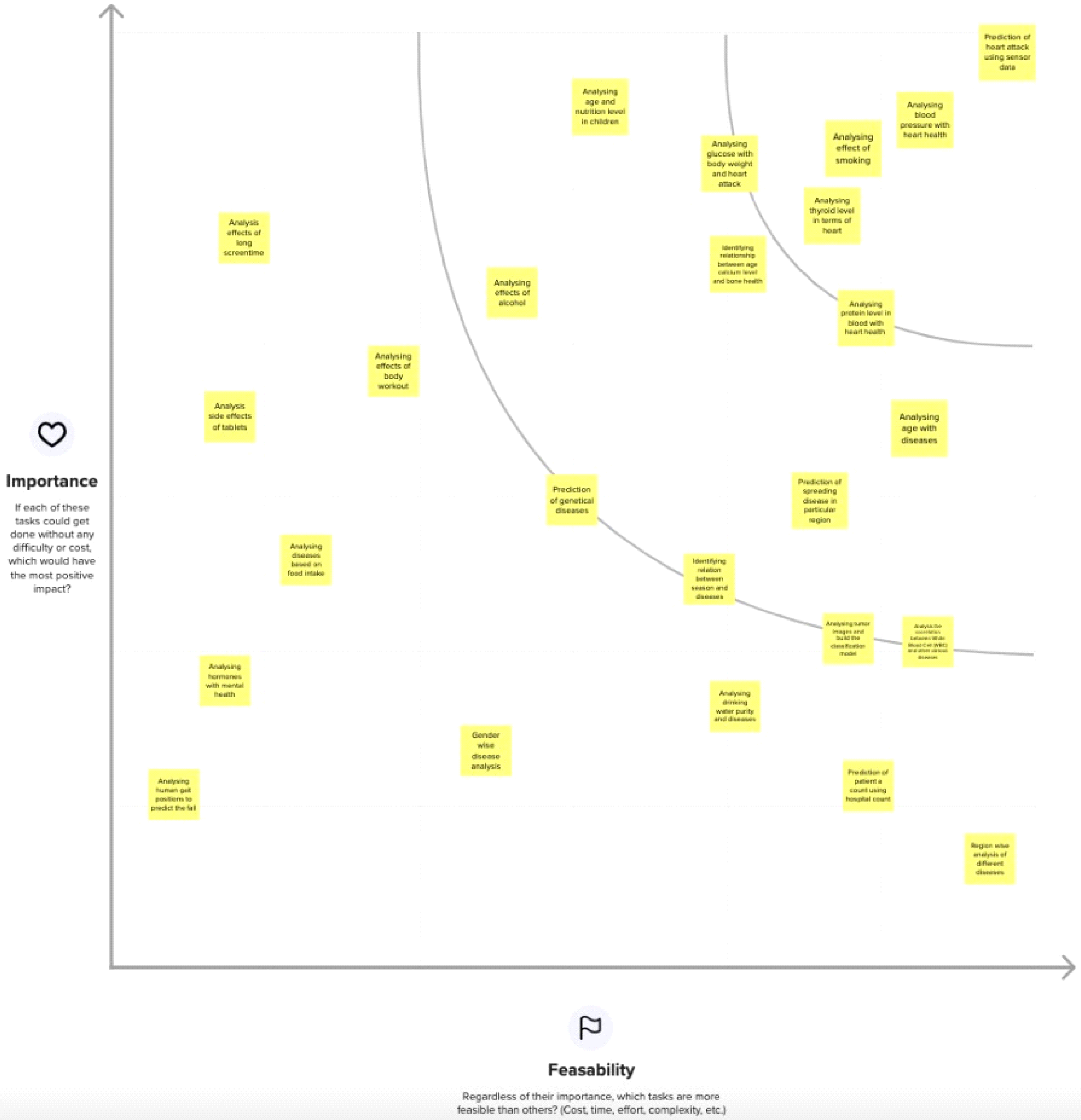
Step-3: Idea Prioritization

4

Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

20 minutes



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 Cog nos will be used for data analytic s. 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 even in The model is 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

Define CS, fit into CC

1. CUSTOMER SEGMENT(S) CS

- Hospital Management
- patient

6. CUSTOMER CONSTRAINTS CC

Customers needs to predict the length of stay of patients with more accuracy during the time of admission.

Maintenance, budget,Human errors in prediction,Unable to predict LOS of patients, No Cost, not sure how to predict.

5. AVAILABLE SOLUTIONS AS

There are few LOS prediction model but with very limited parameters excluding some of the parameters which definitely lead to extension of length of stay of patients

Explore AS, differentiate

Focus on J&P, tap into BE, understand RC

2. JOBS-TO-BE-DONE / PROBLEMS J&P

Job is to predict the length of stay of patients. Unable to predict the LOS of patients leads to improper resource allocation and improper treatment to the patients due to overflow of patients

9. PROBLEM ROOT CAUSE RC

Unable to predict the length of stay of patients with high accuracy. Insufficient medical equipments and bed. Improper maintenance of patients medical history and data

7. BEHAVIOUR BE

Build a model to predict with LOS of patient with higher accuracy. The hospital management should maintain the proper ledger of patients with all the informations about their health, progression and those data can be shared with data analyst to analyse the data

Focus on J&P, tap into BE, understand RC

Identify strong TR & EM

3. TRIGGERS TR

Unable to predict the length of stay of a patient leads to improper allocation of resources.

Hence there is a need to predict the length of stay.

The COVID-19 pandemic proved the impotence of management of hospital resources. So many people struggled due to unavailability of necessary hospital resources for their treatment.

4. EMOTIONS: BEFORE / AFTER EM

Before:

- Improper resource allocation
- Patients unable to get proper treatment and therapy
- Stress and frustration for both patients and hospital management
- unable to promise faster recovery

After:

- Proper resource management and utilization
- Proper treatment and therapy leads to faster recovery
- Proper management and improves trust on the hospital management.

10. YOUR SOLUTION SL

- Collecting data from the trusted source
- Analyze how the length of stay vary with various parameters
- Decide on what are all the parameters impact on the length of stay of patients
- Clean the dataset
- extract the impacting parameters alone to train the model
- train the model to predict the length of stay with various algorithms
- analyze which algorithm is giving better accuracy in predicting the length of stay
- use the algorithm which gives higher accuracy to predict the length of stay

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.

8.CHANNELS of BEHAVIOUR CH

8.1 ONLINE

Handle all the documents and records about the length of stay about the patient and manage them properly. Maintain all the records of medication, treatment, health reports of patients along with the consulting doctors details which can also be used to analyze the length of stay of patients with these details. Properly manage all the patient details.

8.2 OFFLINE

Getting enough medical equipment, checking availability of beds and maintaining in the local electronic ledger or ledger. Checking patients' progress in their health in person and closely monitoring their response to the treatments provided and go for alternative treatments if their body system doesn't respond well to the current treatment.

4. REQUIREMENT ANALYSIS

4.1 Functional requirement

FR No	**Functional Requirement (Epic)**	**Sub Requirement (Story / SubTask)**
FR-1	**Appointments**	**Recurrent appointments and scheduling the available time slots in a regular basis.** **Showing the number of appointments on given day.** **After sign in asking for a ID and phone number to avoid any issues.** **Generating appointment.** **Supporting group appointments and automatically creating a billing charge for completed appointments.** **Appointment Status:** **a. Pending** **b. Confirmed** **c. Cancelled; No Reschedule** **d. Cancelled; Reschedule** **e. No Show** **f. Completed**

FR-2	**Clinical Care**	**The admission of the patient must be examined properly and patients who comes in a critical position should be given immediate treatment.** **Enhanced and improved reliability on reporting the data.** **Access medication history from external sources (ex. Surescripts).** **Predict the length of stay of inpatients**
FR-3	**Patient Records**	**A Proper record or documentations need to be maintained regarding the patients who all consulted and detailed analysis of their health details.** **It should be easily accessible when required.** **Accessible as Standalone function, as well as easily accessible from Progress Note and Evaluation activities.** **Digital records will be more efficient and time saving.**
FR-4	**Bed requirements**	**Analyzing and monitoring of beds which are required are the most important task.** **Using flawless systems for accurately tracking the availability of beds.**

FR-5	Providing insights of dataset	Raw data collection and sharing of data and systems are essential factors in hospital management. According to these data in appropriate measures can be taken. Providing data set without human error.

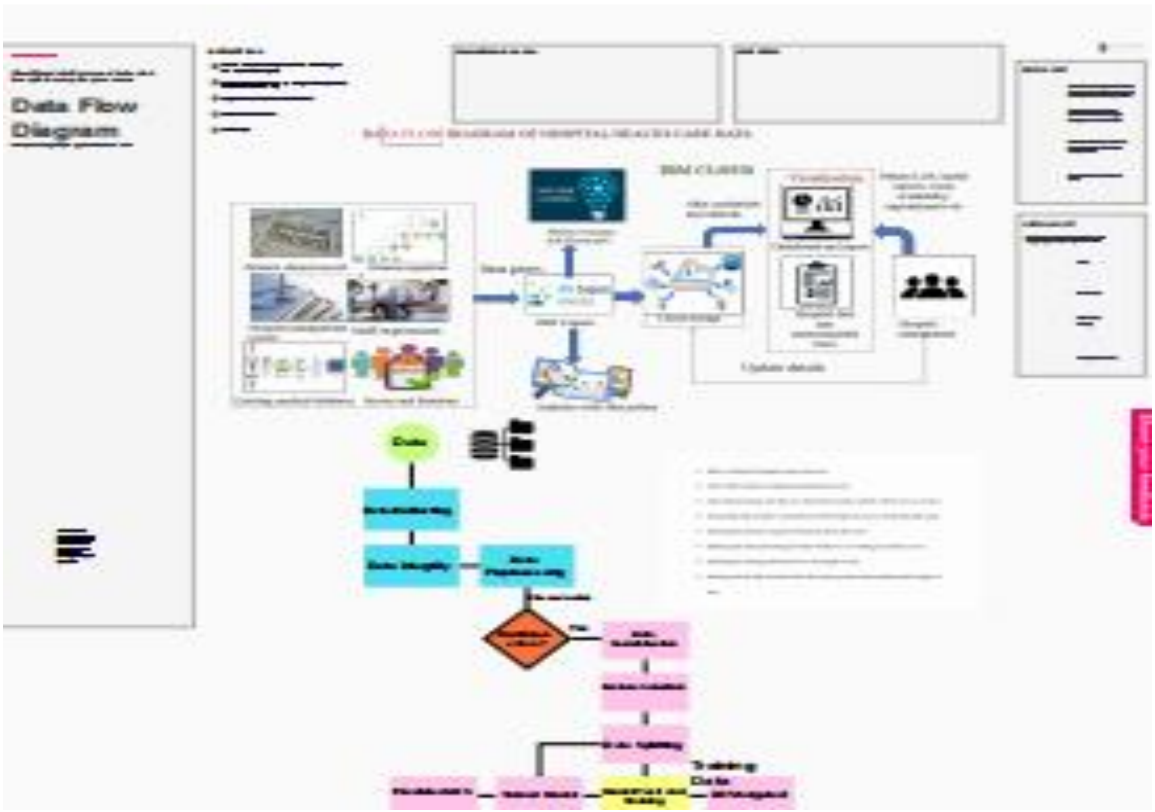
4.3Non-Functional Requirement

Non-Functional Requirement	Description
Usability	Usable systems are straightforward to use by as many people as possible, both in case of either end users or administrators to view the hospital records when needed
Security	Patient identification : To recognize and analyze the patient perfectly
Reliability	Understanding the current trend and working on to it to solve the problem in an efficient manner. Being software as a service, HMS is highly resilient to any technology disruptions, downtime, or crashes experienced by other technology systems.
Performance	Response time: Providing acknowledgment in minimal time about the patient information. Comfortability: To ensure that the guidelines and accessibilities are followed

Availability	**Better coordination with the hospital management to provide all its resources accessible when needed.** **Accessibility of all medical facilities**
Scalability	**Make sure that the work is done in more efficient way with the appropriate resources.** **Make complex decisions understandable with proper data.**

5. PROJECT DESIGN

5.1 Data Flow Diagrams



5.2 Solution & Technical Architecture

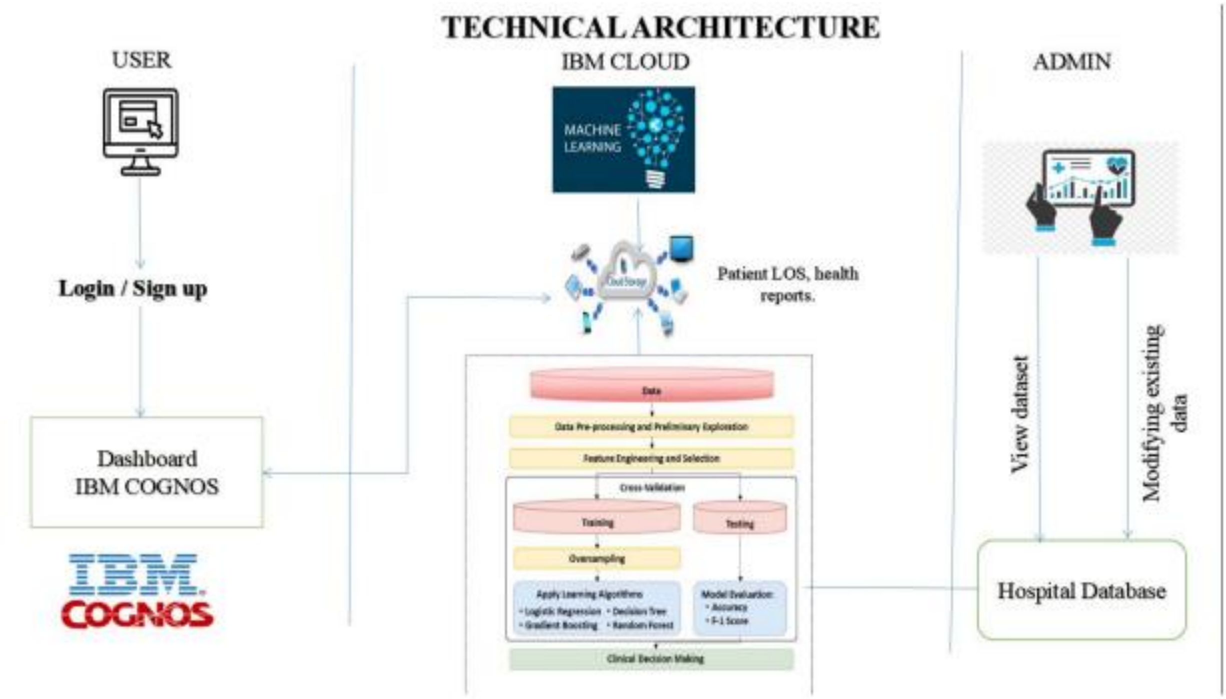


Table-1 : Components & Technologies:

S.No	Component	Description	Technology
1.	User Interface	How user interacts with application e.g. Web UI, Mobile App, Chatbot etc.	HTML, CSS, JavaScript
2.	Application Logic-1	Logic for a process in the application	Python
3.	Application Logic-2	Logic for a process in the application	IBM Watson Assistant
4.	Database	Data Type, Configurations etc.	MySQL
5.	Cloud Database	Database Service on Cloud	IBM Cloud etc.
6.	File Storage	File storage requirements	IBM Block Storage or Other Storage Service or Local Filesystem
7.	External API-1	Purpose of External API used in the application	Aadhar API, etc.
8.	Machine Learning Model	Purpose of Machine Learning Model	Regression Model, etc.
9.	Infrastructure (Server / Cloud)	Application Deployment on Local System / Cloud Local Server Configuration: Cloud Server Configuration :	Local, Cloud Foundry, etc.

5.3 User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer	Dashboard	USN-1	As a user, I can upload the dataset to the dashboard	I can access dashboard	High	Sprint-1
	View	USN-2	As a user, I can view the patient details	I can visualize the data	medium	Sprint-2
Admin	Analysis	USN-3	As a user, I will analysis the given dataset	I can analysis the dataset	High	Sprint-3
	Predict	USN-4	As a user, I will predict the length of stay	I can predict the length of stay	High	Sprint-4
	Collect data	USN-5	As a analyst I need to collect the dataset		High	Sprint-1

	Prepare data	USN-6	As an analyst I need to do feature extraction	I can extract the parameters that have impact the length of stay	High	Sprint-2
Visualization	Dashboard	USN-7	As a user I can prepare data by using visualization technique	I can prepare the data with visualization technique	Medium	sprint -2

6. PROJECT PLANNING

6.1 Sprint Planning & Estimation

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Dataset	USN-1	The user need a complete data about the patient admitted in the hospital and a dataset should be prepared.	2	High	Monesh Kumar p
Sprint-1	Dataset Exploration	USN-2	Data exploration is the first step of data analysis used to explore and visualize data to uncover insight from the start	2	High	Monesh Kumar P Mathan R
Sprint-1	Secondary Exploration	USN-3	The secondary relationship of data is identified here	1	Low	Kishore S Manikandan A

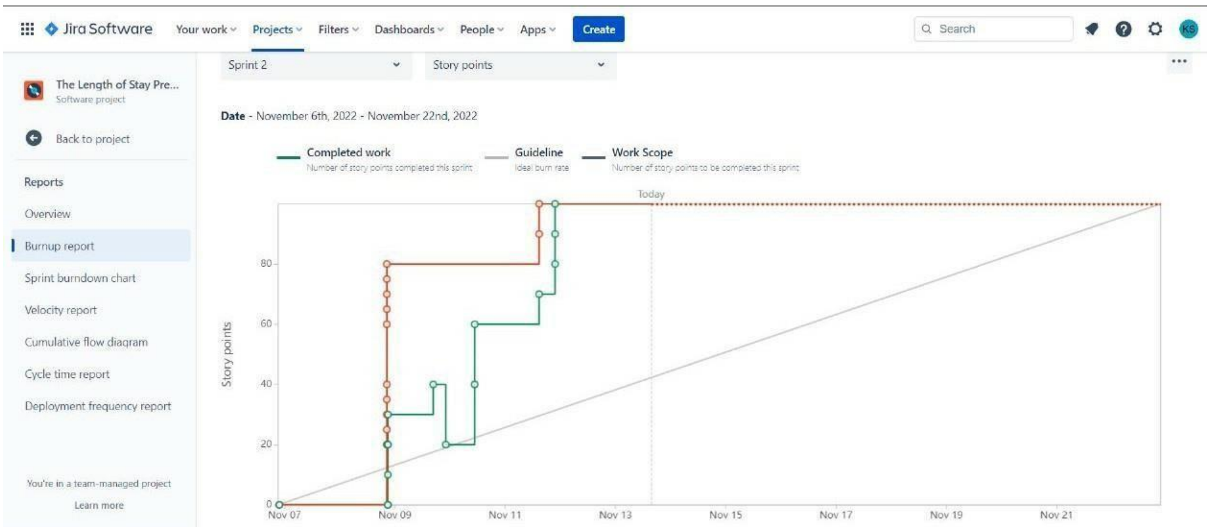
6.2 Sprint Delivery Schedules

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members

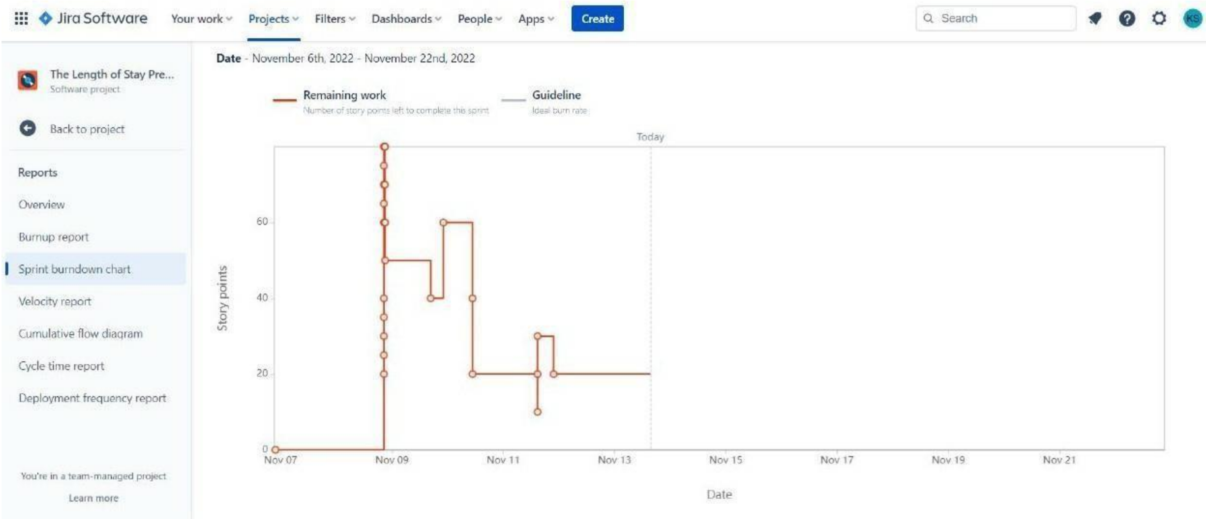
Sprint-2	Data Visualization	USN-4	The patient data are graphically visualized for data verification data to know available resource	2	High	Kishore S Manikandan A
Sprint-3	Dashboard	USN-5	The explore and visualized data are displayed in dashboard	2	High	Mathan R
Sprint-4	Predictive model	USN-6	The predictive analysis on the data performed by modelling the predictive model	2	High	Monesh Kumar P Mathan R

6.3 Reports from JIRA

Burnt Up Chart

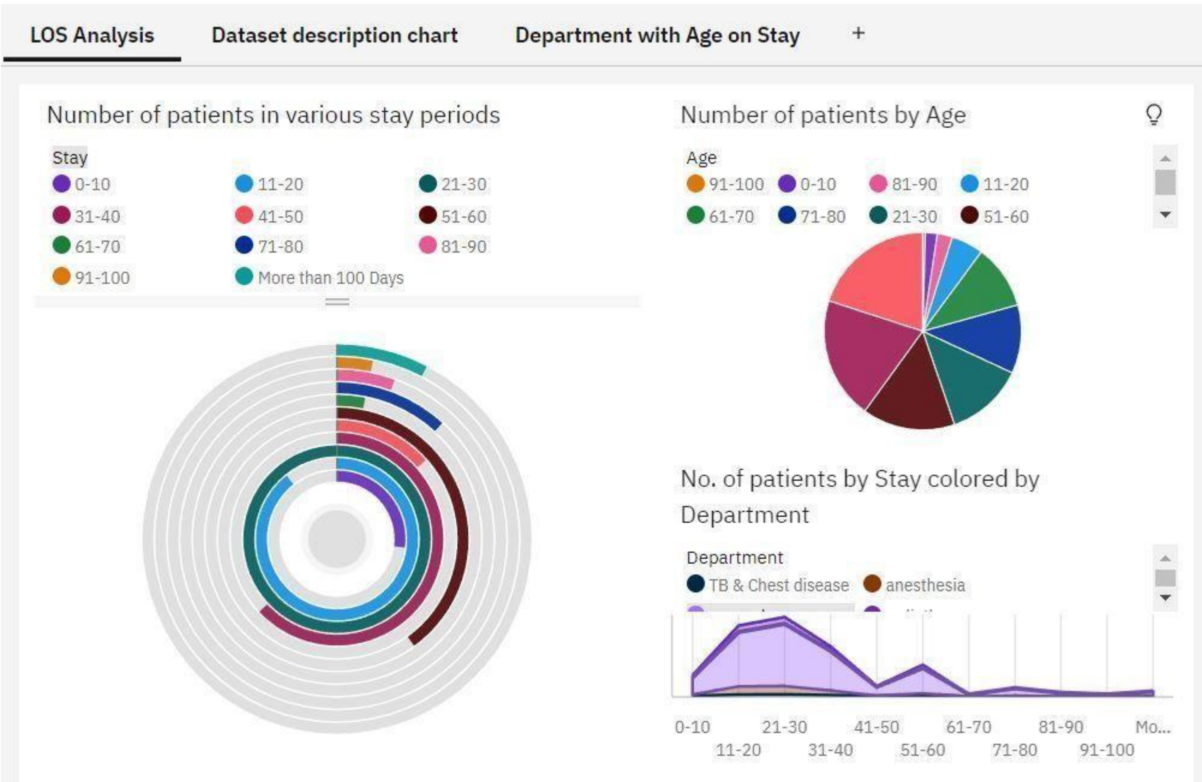


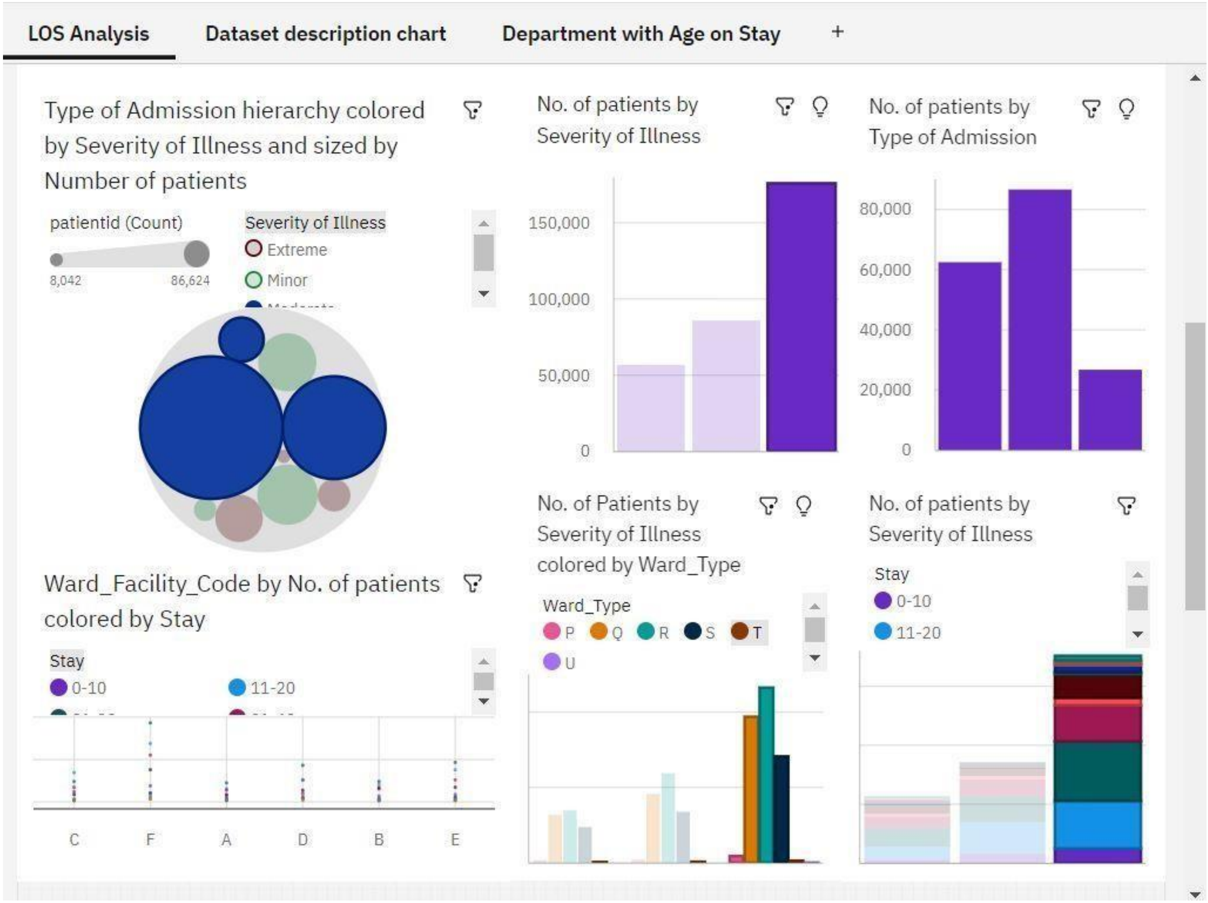
Burnt Down Chart

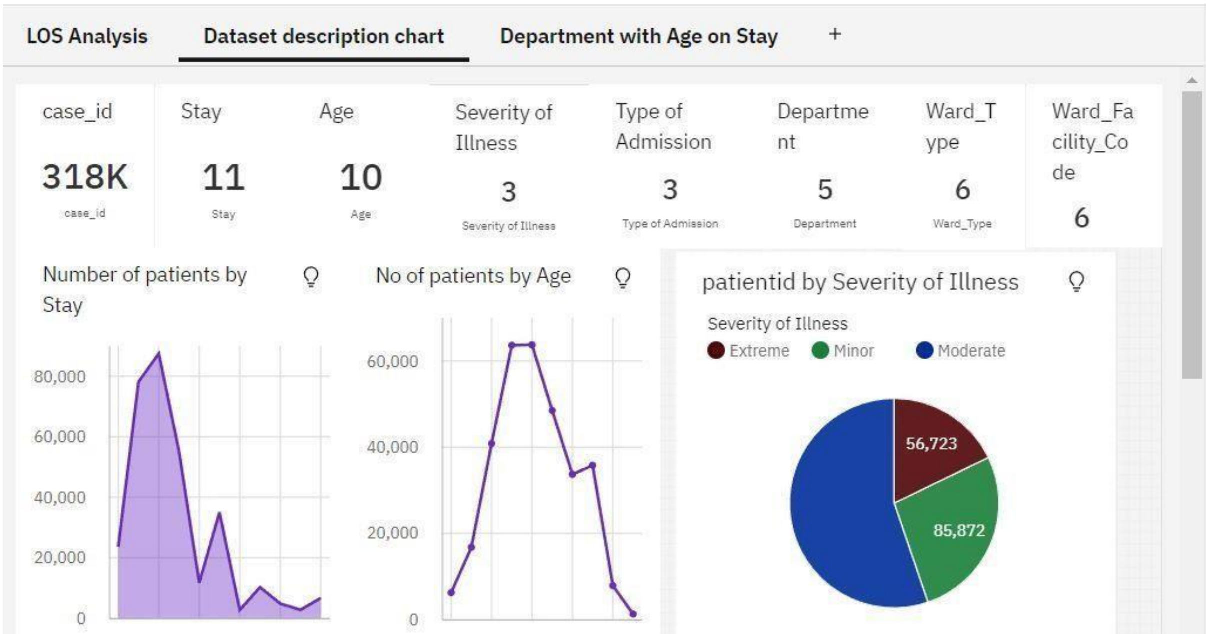


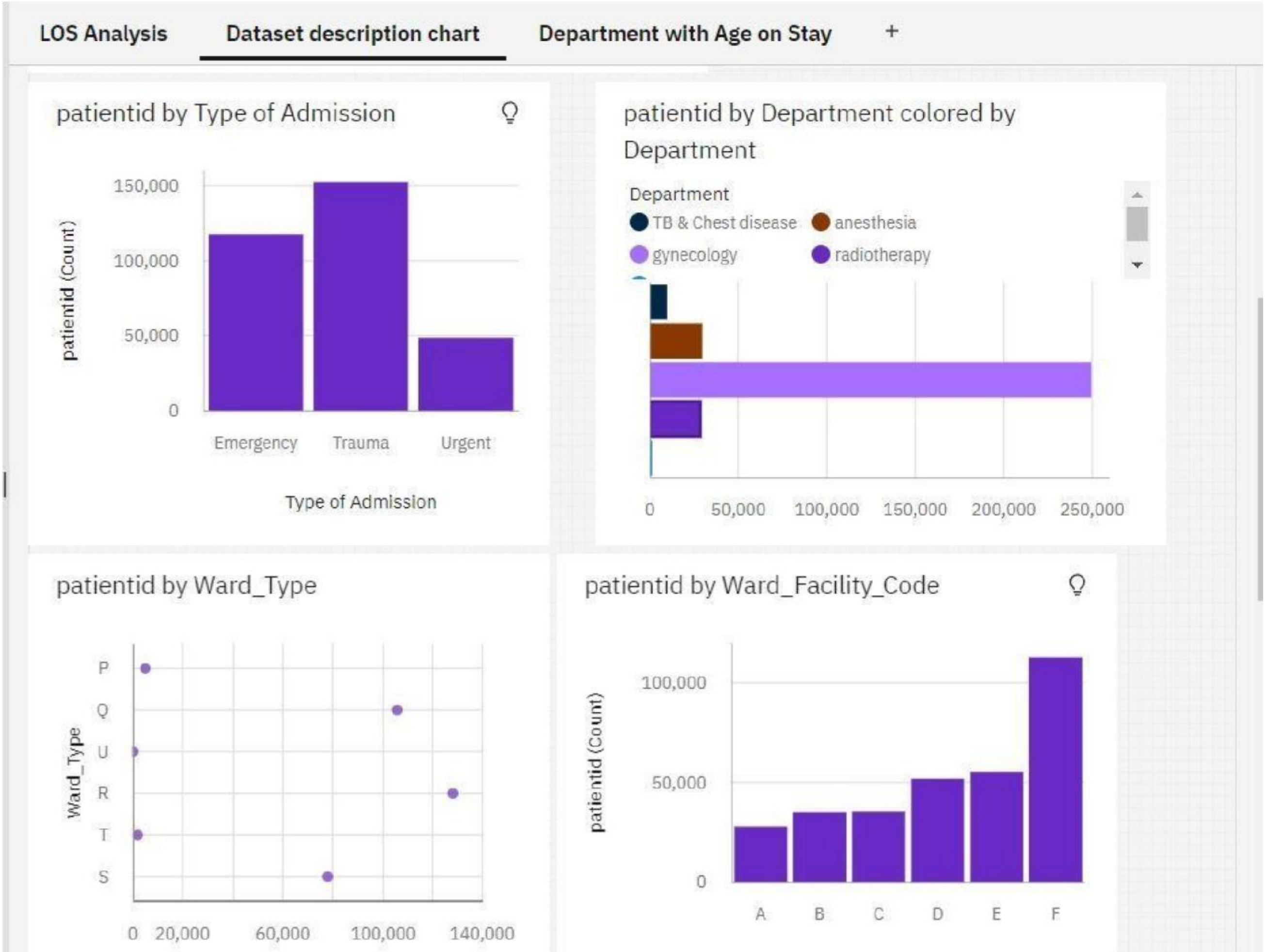
7. CODING & SOLUTIONING (Explain the features added in the project along with code)

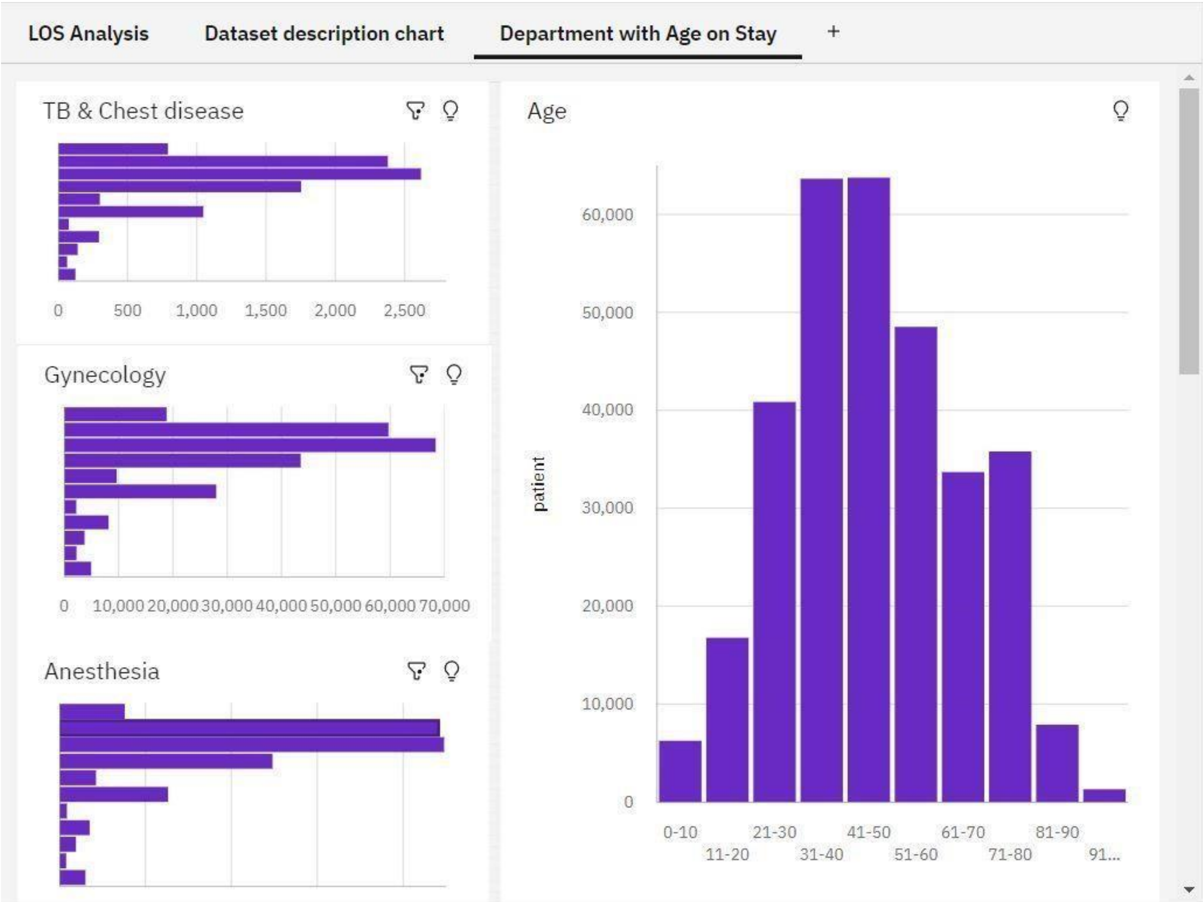
7.1 Feature 1

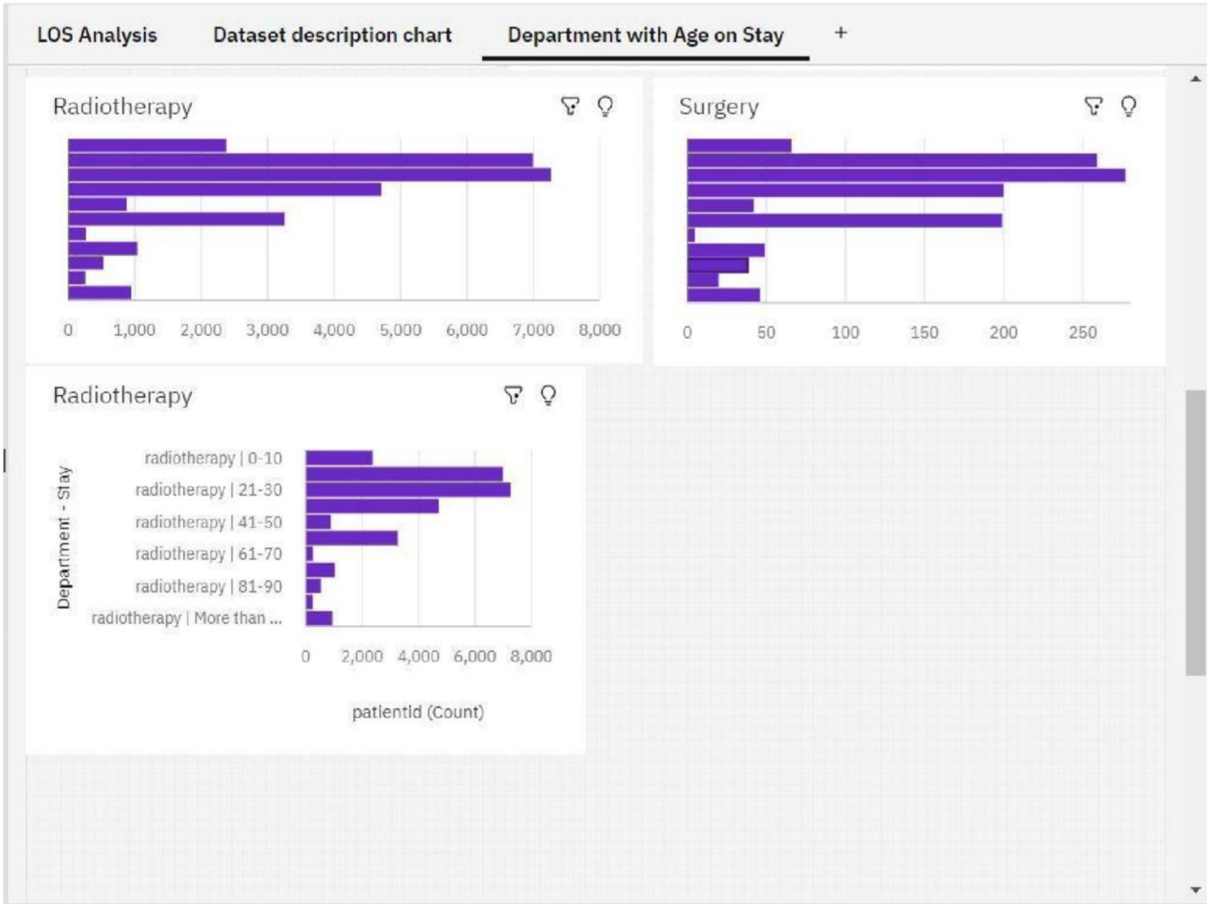



LOS Analysis
Dataset description chart
Department with Age on Stay
+
Type of Admission hierarchy colored by Severity of Illness and sized by Number of patients
patientid (Count)
8,042
86,624
Severity of Illness
Extreme
Minor
No. of patients by Severity of Illness
150,000
100,000
50,000
0
No. of patients by Type of Admission
80,000
60,000
40,000
20,000
0
No. of Patients by Severity of Illness colored by Ward_Type
Ward_Type
P
Q
R
S
T
U
No. of patients by Severity of Illness
Stay
0-10
11-20
Ward_Facility_Code by No. of patients colored by Stay
Stay
0-10
11-20
C
F
A
D
B
E


LOS Analysis
Dataset description chart
Department with Age on Stay
+
case_id
318K
case_id
Stay
11
Stay
Age
10
Age
Severity of Illness
3
Severity of Illness
Type of Admission
3
Type of Admission
Departme nt
5
Department
Ward_T ype
6
Ward_Type
Ward_Fa cility_Co de
6
Number of patients by Stay
80,000
60,000
40,000
20,000
0
No of patients by Age
60,000
40,000
20,000
0
patientid by Severity of Illness
Severity of Illness
Extreme
Minor
Moderate
56,723
85,872


LOS Analysis
Dataset description chart
Department with Age on Stay
patientid by Type of Admission
150,000
100,000
50,000
0
patientid (Count)
Emergency
Trauma
Urgent
Type of Admission
patientid by Department colored by Department
Department
TB & Chest disease
anesthesia
gynecology
radiotherapy
0
50,000
100,000
150,000
200,000
250,000
patientid by Ward_Type
P
Q
U
R
T
S
Ward_Type
0
20,000
60,000
100,000
140,000
patientid by Ward_Facility_Code
100,000
50,000
0
patientid (Count)
A
B
C
D
E
F


LOS Analysis
Dataset description chart
Department with Age on Stay
+
TB & Chest disease
0 500 1,000 1,500 2,000 2,500
Gynecology
0 10,000 20,000 30,000 40,000 50,000 60,000 70,000
Anesthesia
Age
patient
0 10,000 20,000 30,000 40,000 50,000 60,000
0-10 11-20 21-30 31-40 41-50 51-60 61-70 71-80 81-90 91...


LOS Analysis
Dataset description chart
Department with Age on Stay
+
Radiotherapy
0 1,000 2,000 3,000 4,000 5,000 6,000 7,000 8,000
Surgery
0 50 100 150 200 250
Radiotherapy
Department - Stay
radiotherapy | 0-10
radiotherapy | 21-30
radiotherapy | 41-50
radiotherapy | 61-70
radiotherapy | 81-90
radiotherapy | More than ...
0 2,000 4,000 6,000 8,000
patientid (Count)

7.2 Feature 2

```
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

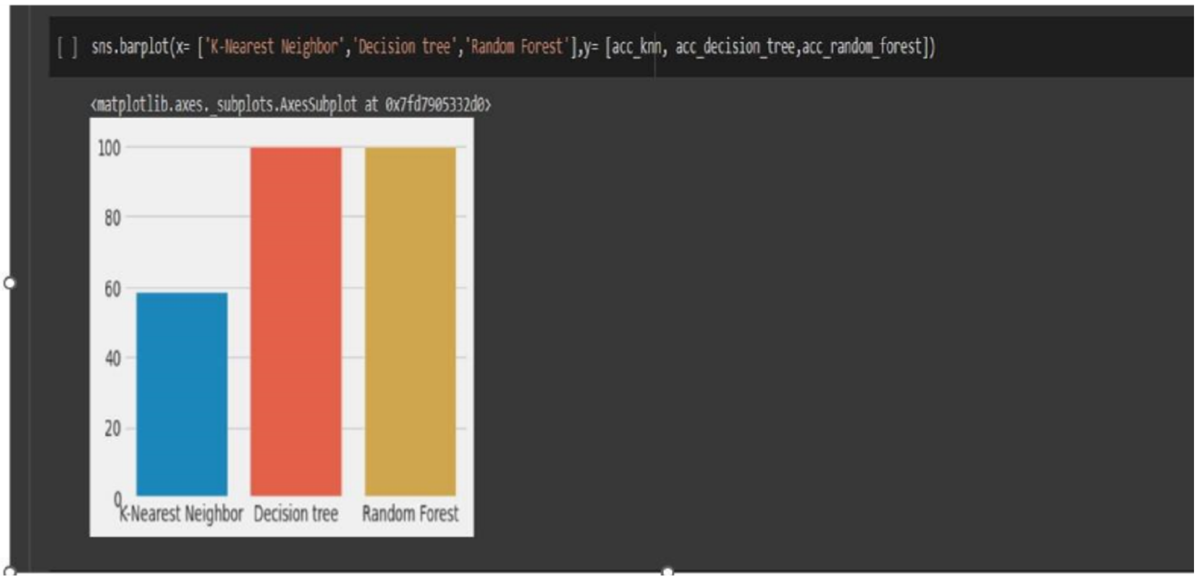
```
random_forest = RandomForestClassifier(n_estimators=100)
random_forest.fit(X_train, Y_train)
Y_pred = random_forest.predict(X_test)
random_forest.score(X_train, Y_train)
acc_random_forest = round(random_forest.score(X_train, Y_train) * 100, 2)
acc_random_forest
```

99.76

Prediction accuracy comparison

8. RESULTS

8.1 Performance Metrics



9. ADVANTAGES & DISADVANTAGES

Advantages

- 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
- Research and prediction of disease.
- Automation of hospital administrative processes.
- 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 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.

10. CONCLUSION

Data analytics is the science of analysing raw datasets in order to derive a conclusion regarding the information they hold. It enables us to discover patterns in the raw data and draw valuable information from them. To some, the domain of healthcare data analytics may look new, but it has a lot of potential, especially if you wish to engage in challenging job roles and build a strong data analytics profile in the upcoming years. In this blog, we have covered some of the major topics such as what is healthcare data analytics, its applications, scope, and benefits, etc. We hope it helps you in your decision-making as a healthcare data analytics professional

11. FUTURE SCOPE

The Future of Healthcare, Intel provides a foundation for big data platforms and AI to advance health analytics. Predictive data analytics is helping health organizations enhance patient care, improve outcomes, and reduce costs by anticipating when, where, and how care should be provided. The future of big data in healthcare will be determined by technological breakthroughs from 2022 to 2030. Complete patient care and cost-effective prescription procedures are required for population health management. To assess clinical and claims data, they must be combined on the same platform.

Countries around the world have started to invest more capital in medical infrastructure, pharmaceuticals, and healthcare smart analytics solutions. The market is growing and will continue to expand, given the benefits of healthcare data analytics. It has also risen as a good career option for fresh data science and data analytics graduates or professionals who wish to build their career in the healthcare sector. Due to the sensitivity of the profession, the salary offers for healthcare data analysts are lucrative around the world. Apart from the remuneration, the opportunities to work with some of the biggest names in the healthcare sector is also worth mentioning. Hence, healthcare data analytics is growing to be one of the most rewarding branches of data analytics in the coming future.

12. APPENDIX

Source Code

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.(

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

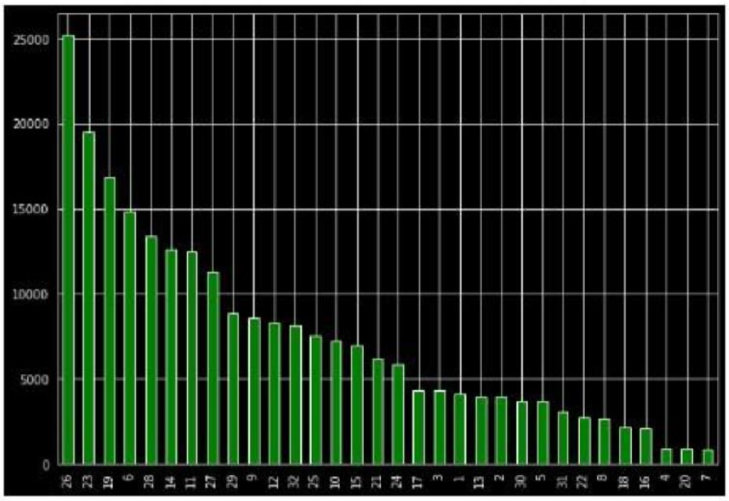
Distribution of values

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
```

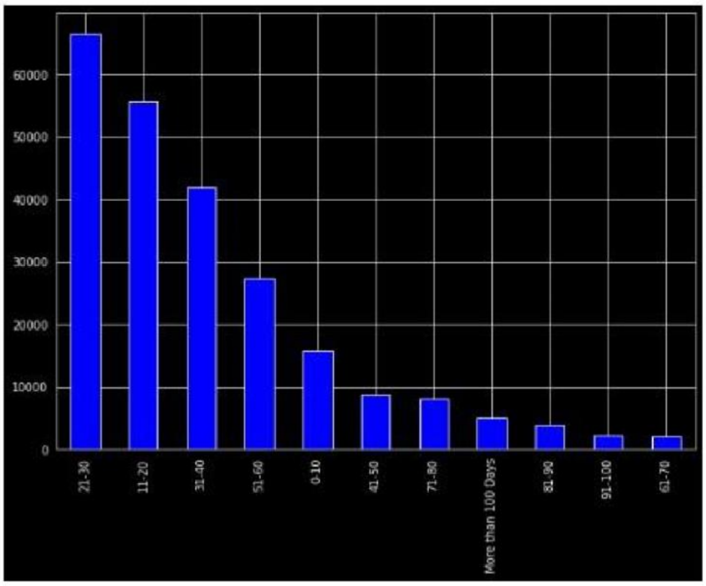
```
plt.figure(figsize=(10,7))
train.Hospital_code.value_counts().plot(kind="bar", color = ['green'])
```



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
```



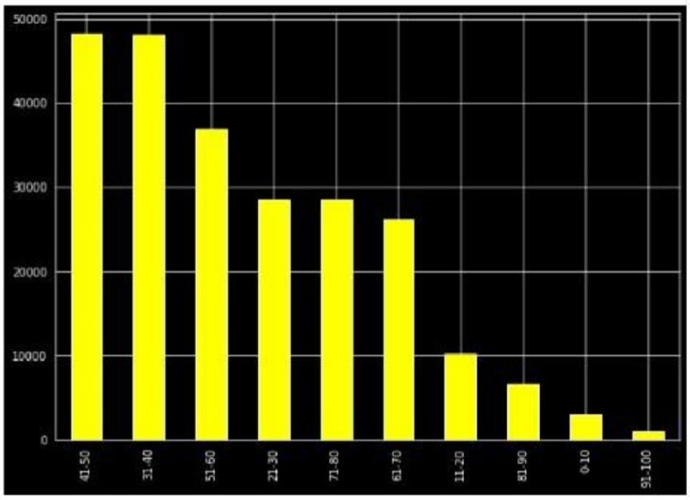
Age

```
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'])
```



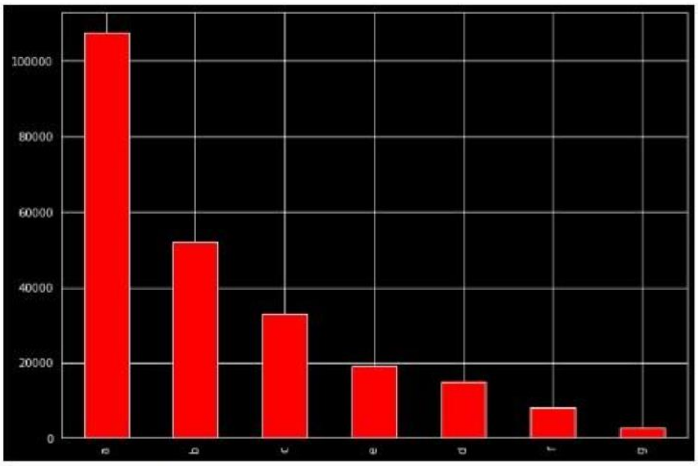
Hospital_type_code

```
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'])
```

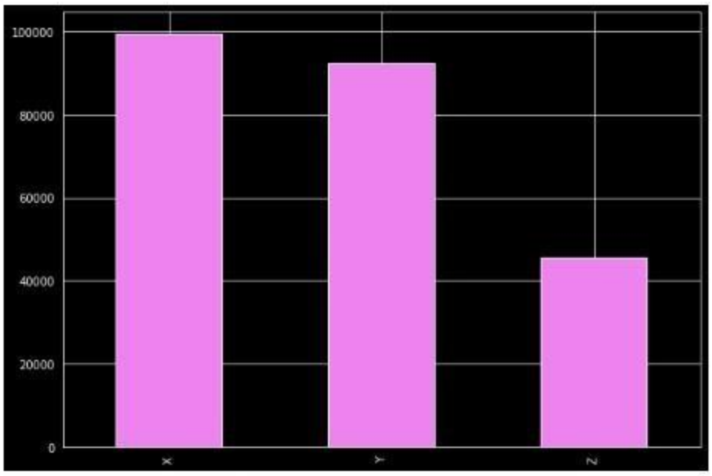


Hospital_region_code

```
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'])
```



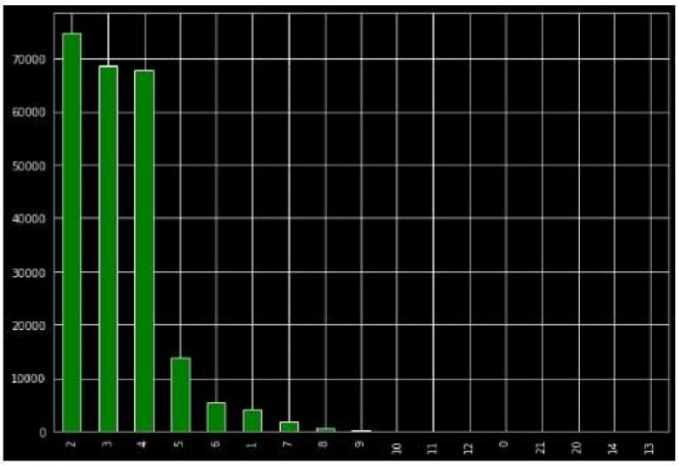
Available_Extra_Rooms_in_Hospital

```
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'])
```



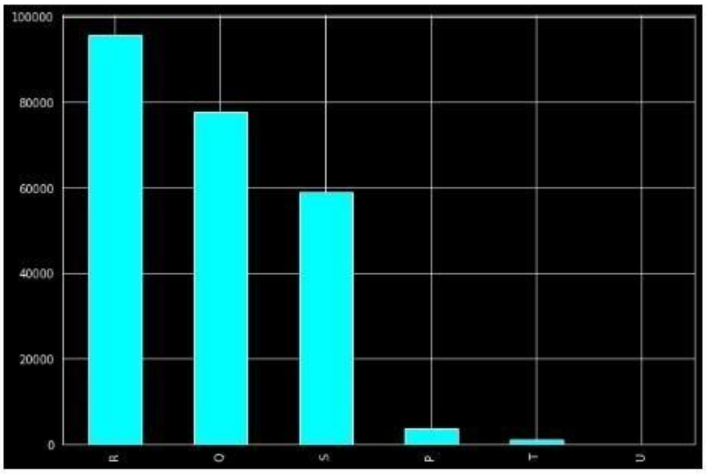
Department

```
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'])
```



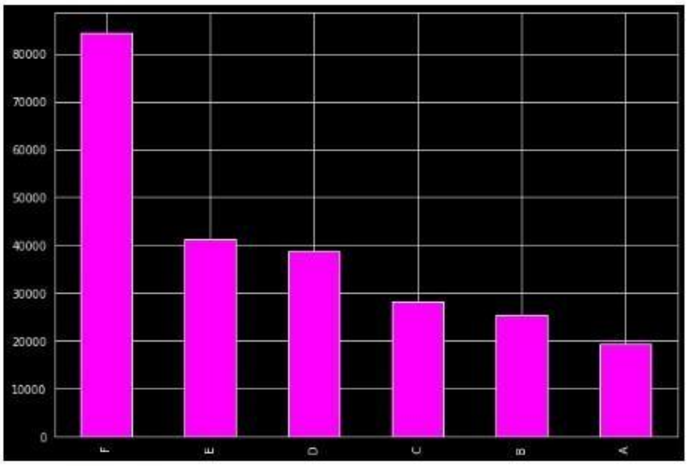
Ward_Facility_Code

```
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'])
```



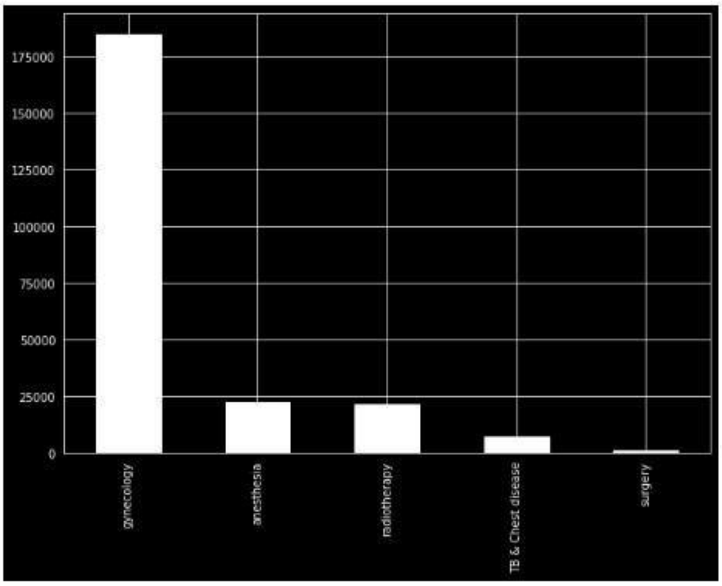
Visitors_with_Patient

```
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'])
```

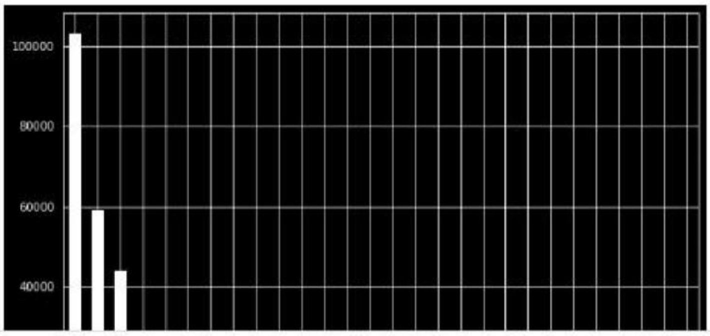


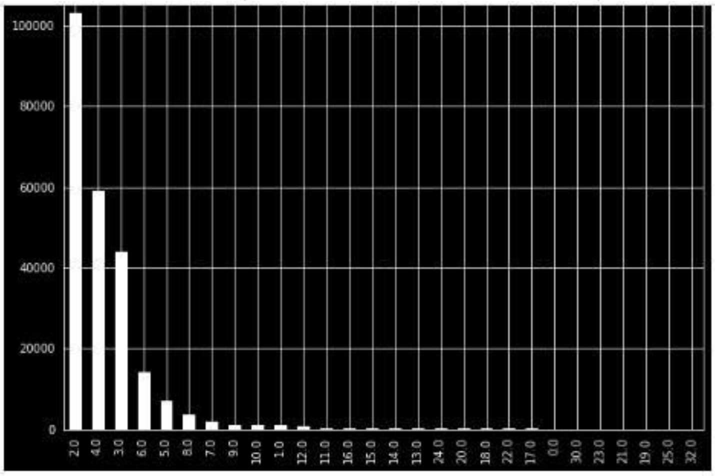
Ward_Type

```
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
```

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



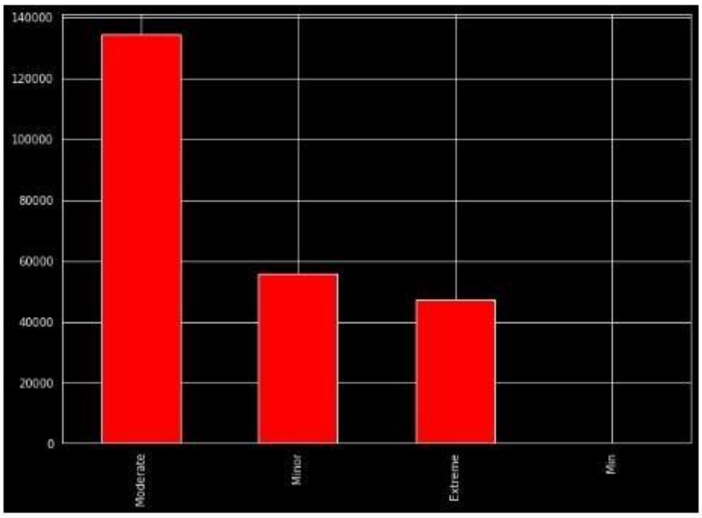


Severity of Illness

```
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'])
```



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'])
```

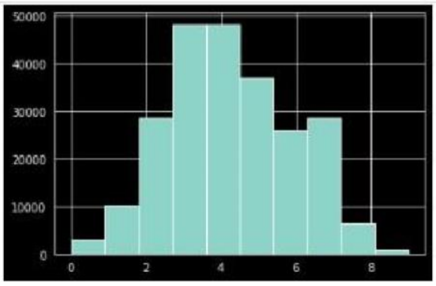
```python
combined[0]
```

	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	

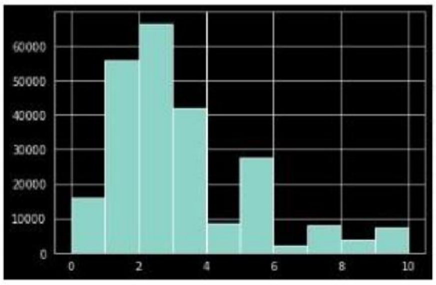
237309 rows × 14 columns

```
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	



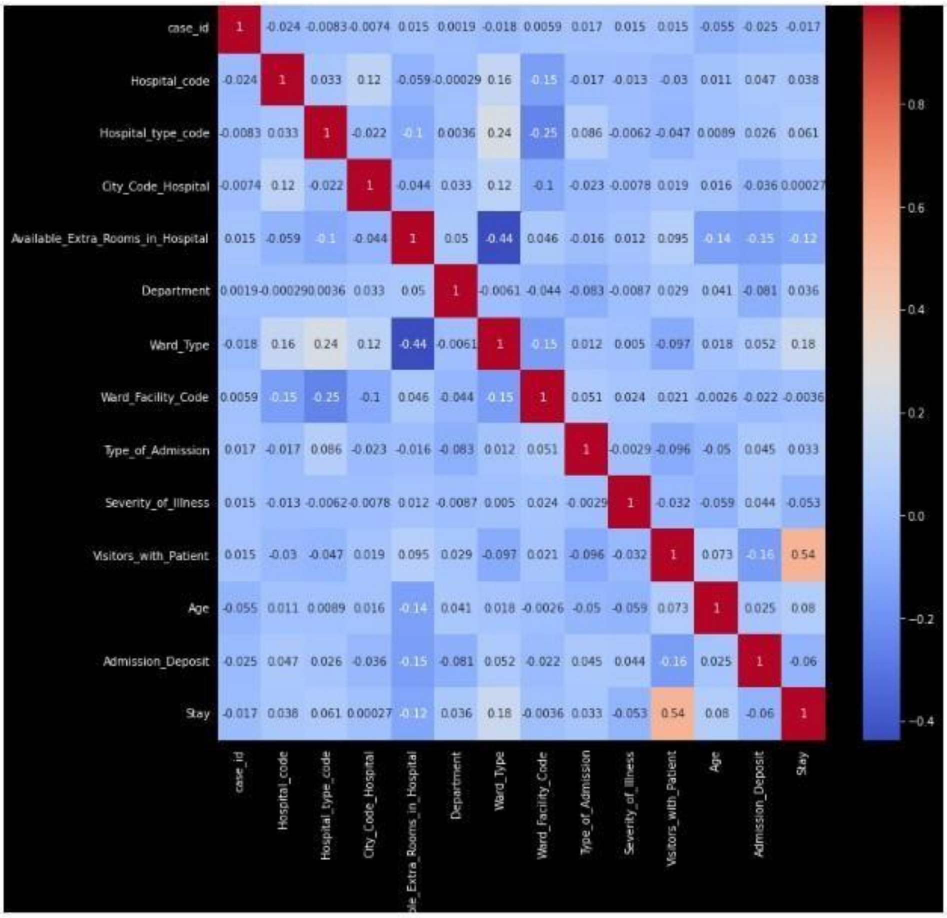

```
combined[0].Stay.hist()
```



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

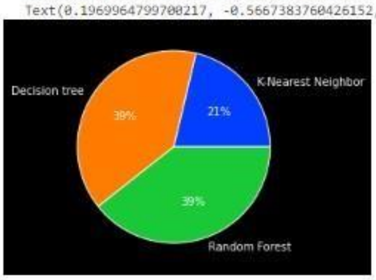
```
random_forest = RandomForestClassifier(n_estimators=100)
random_forest.fit(X_train, Y_train)
Y_pred = random_forest.predict(X_test)
random_forest.score(X_train, Y_train)
acc_random_forest = round(random_forest.score(X_train, Y_train) * 100, 2)
acc_random_forest
```

```
99.76
```

Prediction accuracy comparison

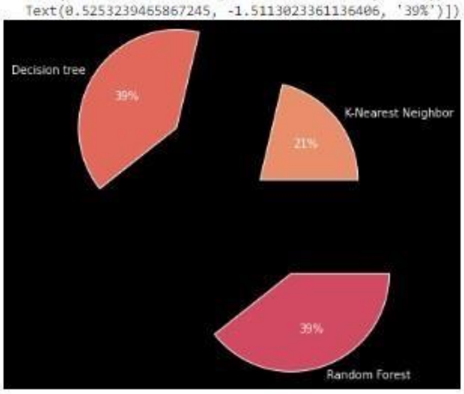
```
palette_color = sns.color_palette('bright')
data=[acc_knn, acc_decision_tree,acc_random_forest]
keys=['K-Nearest Neighbor','Decision tree','Random Forest']

#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):
```

```
  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
```

GitHub & Project Demo Links

GitHub link: https://github.com/IBM-EPBL/IBM-Project-54327-1661851165

Project demo link:

https://colab.research.google.com/drive/1DpGcjD6aJZENhHUiDWnwIjFAbk0I3ux?usp=sharing