

VISUALIZATION AND PREDICTING HEART DISEASE WITH AN INTERACTIVE DASHBOARD

TEAM ID:PNT2022TMID53507

GIT REPO ID:IBM-Project-25207-1659954943

TEAM MEMBERS:

Roshan Vijaragavan B (TEAM LEADER)

Pranavraj P (Team Member 1)

Ugendran L(Team Member 2)

Thanga Tamilselvan V(Team Member 3)

INTRODUCTION:

1.1 Project Overview

In recent times, Heart Disease prediction is one of the most complicated tasks in medical field. In the modern era, approximately one person dies per minute due to heart disease. Data science plays a crucial role in processing huge amount of data in the field of healthcare. As heart disease prediction is a complex task, there is a need to automate the prediction process to avoid risks associated with it and alert the patient well in advance. The main challenge in today's healthcare is provision of best quality services and effective accurate diagnosis . Even if heart diseases are found as the prime source of death in the world in recent years, they are also the ones that can be controlled and managed effectively. The whole accuracy in management of a disease lies on the proper time of detection. The proposed work predicts the chances of Heart Disease and classifies patient's risk level by implementing different data mining techniques such as GaussianNB, Linear REgression, Logistic Regression, KNN and Random Forest Classifier.

1.2 Purpose

Heart disease is perceived as the deadliest disease in human life across the world. In particular, in this type of disease the heart is not capable of pushing the required quantity of blood to the remaining organs of the human body in order to accomplish the regular functionalities . Some of the symptoms of heart disease include physical body weakness, improper breathing, swollen feet, etc. The techniques are essential to identify the complicated heart diseases which result in high risk in turn affecting human life.

Presently, diagnosis and treatment processes are highly challenging due to inadequacy of physicians and diagnostic apparatus that affect

the treatment of heart patients. Early diagnosis of heart disease is significant to minimize the heart related issues and to protect it from serious risks. The invasive techniques are implemented to diagnose heart diseases based on medical history, symptom analysis reports by experts, and physical laboratory reports. Moreover, it causes delay and imprecise diagnosis due to human intervention. It is time consuming, computationally intensive and expensive at the time of assessment. Heart disease can be predicted based on various symptoms such as age, gender, pulse rate etc. Data analysis in healthcare assists in predicting diseases, improving diagnosis, analyzing symptoms, providing appropriate medicines, improving the quality of care, minimizing cost, extending the life span and reduces the death rate of heart patients. ECG (Electrocardiogram) helps in screening irregular heartbeat and stroke with the embedded sensors by resting it on a chest in order to track the patient's heart beat. Heart disease prediction is being done with the detailed clinical data that could assist experts to make decisions. Human life is highly dependent on proper functioning of blood vessels in the heart. The improper blood circulation causes heart inactiveness, kidney failure, imbalanced condition of brain, and even immediate death also. Some of the risk factors that can cause heart diseases are obesity, smoking, diabetes, blood pressure, cholesterol, lack of physical activities and unhealthy diet. Acute Myocardial Infarction (AMI) is the cardiovascular disease that happens due to interruption in the blood flow or circulation in the heart muscle, causing heart muscle to become necrotic (damage or die). The primary reason for this disease is the blockage means that the blood flow to the heart muscle becomes obstructed or reduced. If the blood flow is reduced or

obstructed, the functioning of red blood cells that carry enough oxygen helps in sustaining consciousness and human life have a severe impact. Without oxygen supply for 6 to 8 minutes, heart muscle may get arrested, which in turn results in the patient's death. The significant cause of the cardiovascular disease is 'plaque' means a hard substance formed in the coronary arteries which is made up of cholesterol (fat), causes the blood flow to be reduced or obstructed. Sometimes, it can be formed in the arteries known as atherosclerosis and investigating the cause of it are determined as a chronic inflammation. The increase in the amount of white blood cells causes inflammation and other subsequent disorders such as stroke or reinfarction. Generally, there are two stages of wound healing in terms of monocytes and macrophages, namely, inflammatory and reparative stages.

However, the two stages are compulsory for proper wound healing and if the inflammation is continued too long, then it leads to heart failure. An unusual type of heart disease is the acute spasm or contraction in the coronary arteries.

The spasms become visible in arteries suddenly with no symptom of atherosclerosis. It blocks the blood flow that causes oxygen deprivation in the heart. Male genders are more likely to experience heart attack than females. Moreover, women can experience pain more than an hour and the duration to experience the pain of men is normally less than an hour. The cardiovascular disease has an impact in the complete physiological system, not only in the heart; changes occur everywhere that too in the remote organs such as bone marrow and spleen.

LITERATURE SURVEY

2.1 Existing Problem

Healthcare industries generate enormous amount of data, so called big data that accommodates hidden knowledge or pattern for decision making. The huge volume of data is used to make decision which is more accurate than intuition. Exploratory Data Analysis (EDA) detects mistakes, finds appropriate data, checks assumptions and determines the correlation among the explanatory variables. In the context, EDA is considered as analyzing data that excludes inferences and statistical modeling. Analytics is an essential technique for any profession as it forecast the future and hidden pattern. Data analytics is considered as a cost effective technology in the recent past and it plays an essential role in healthcare which includes new research findings, emergency situations and outbreaks of disease. The use of analytics in healthcare improves care by facilitating preventive care and EDA is a vital step while analyzing data.

2.2 References

1. Heart Disease Prediction using Exploratory Data Analysis R.

Indrakumari, T.Poongodi, Soumya Ranjan Jena: In this paper, the risk factors that causes heart disease is considered and predicted using K-means algorithm and the analysis is carried out using a publicly available data for heart disease. The dataset holds 209 records with 8 attributes such as age, chest pain type, blood pressure, blood glucose level, ECG in rest, heart rate and four types of chest pain. To predict the heart disease K-means clustering algorithm is used along with data analytics and visualization tool. The paper discusses the pre-processing methods, classifier performances and evaluation metrics. In the result section, the visualized data shows that the prediction is accurate.

2. Prediction of heart disease at early stage using data mining and big data analytics: A survey by N. K. Salma Banu, Suma Swamy: Several studies have been carried out for developing prediction model using individual technique and also by combining two or more techniques This paper provides a quick and easy review and understanding of available prediction models using data mining from 2004 to 2016. The comparison shows the accuracy level of each model given by different researchers. A few investigations have been completed for creating expectation model utilizing individual procedure and furthermore by joining at least two strategies This paper gives a speedy and simple survey and comprehension of accessible forecast models utilizing information mining from 2004 to 2016. The correlation shows the precision level of each model given by various analysts

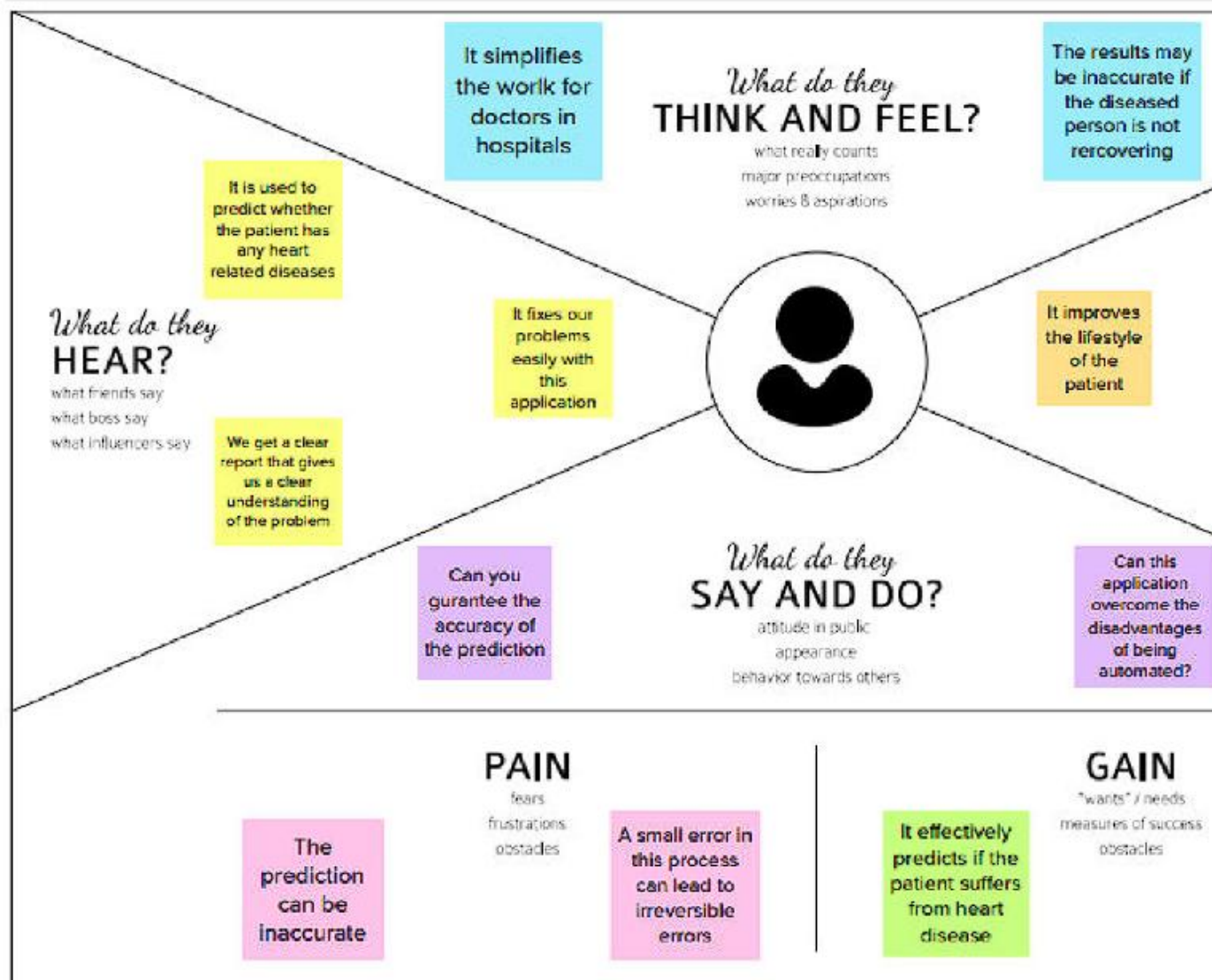
2.3 Problem Statement Definition

Heart disease can be managed effectively with a combination of lifestyle changes, medicine and, in some cases, surgery. With the right treatment, the symptoms of heart disease can be reduced and the functioning of the heart can be improved. The predicted results can be used to prevent and thus reduce cost for surgical treatment and other expenses. The overall objective of our project is to predict accurately with few tests and attribute the presence of heart disease. Attributes considered form the primary basis for tests and give accurate results more or less to predict with faster efficiency the risk of having heart disease. Decisions are often made based on

doctors' intuition and experience rather than on the knowledge rich data hidden in the data set and databases. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients

3.IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas



3.2 Proposed Solution

Proposed Solution Template:

Project team shall fill the following information in proposed solution template.

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Heart Disease prediction is one of the most complicated tasks in machine learning in the modern era, approximately 10 minutes due to heart disease prediction is a complex task. The goal is to automate the prediction of heart disease associated with it and to advance. By using the machine learning which patients are the most likely to suffer from a heart disease in the future.
2.	Idea / Solution description	To accurately predict whether a patient is likely to be suffered from heart disease based on their medical history, gender, age, chest pain, etc.
3.	Novelty / Uniqueness	Using IBM cognos we can analyze the patient data and identify the difficulties, Which process of datasets provided by the hospital.
4.	Social Impact / Customer Satisfaction	The social impact of the project is that Hospitals can use this information to improve patient allocation and better for the patient.
5.	Business Model (Revenue Model)	Used to predict whether a patient is likely to suffer from a heart disease and can start the necessary treatment.

6.	Scalability of the Solution	Can create a health record for each patient. Patients can update their details, their treatments and their Doctor in real time. Doctors can add remarks regarding symptoms, medication, so that the system can monitor medication level and
----	-----------------------------	---

3.3 Problem Solution fit

<p>1. CUSTOMER SEGMENT(S) CS</p> <p>Who is your customer?</p> <ul style="list-style-type: none"> • smokers • people who have high blood pressure • people who have high cholesterol • people who have high lipoprotein • Diabetes patients • people who have lack of regular exercise • Thrombosis patients • people who experience shortness of breath • people who have Chest pain, chest tightness, chest pressure and chest discomfort (angina) • people who have Pain in the neck, jaw, throat, upper belly area or back • people who have pain, numbness, weakness or coldness in the legs or arms if the blood vessels in those body areas are narrowed • people who have overweight 	<p>6. CUSTOMER CONSTRAINTS SL</p> <p>What constraints prevent your customers from taking action or limit their choices of solutions?</p> <ul style="list-style-type: none"> • Lack of knowledge about heart diseases • Negative thoughts of the customer • Personal characteristics and physical stability of the customer • Complex symptoms of heart failure • Psychological problems • Lack of support • Lack of focus in treatment • economical background is major constraints that prevent the customer from taking action • Medical and disease-related limitations 	<p>5. AVAILABLE SOLUTIONS AS</p> <p>Which solutions are available to the customers when they face problem or need to get the job done? What have they tried in the past? What pros & cons do these solutions have?</p> <p>There are various solutions available for the people who are affected with heart diseases. They are,</p> <ul style="list-style-type: none"> • Quit smoking • get cholesterol test periodically • eat plenty of fruits, vegetables and healthy foods with grains, sprouts, nuts etc. • Exercise regularly • Maintain a good physique. <p>If these solutions are properly followed then the people affected with disease can be cured naturally.</p> <ul style="list-style-type: none"> • But, along with these they have to go for regular medical checkup and test for any heart disease. • If disease is found in heart they need to make arrangements under proper medications.
<p>2. JOBS-TO-BE-DONE/PROBLEMS J&P</p> <p>Which jobs-to-be-done (or problems) do you address for your customer? There could be more than one, explore different ideas.</p> <ul style="list-style-type: none"> • I have depending on medical support • Financial insecurity • shortness of breath • may feel chest pain, chest tightness, chest pressure 	<p>9. PROBLEM ROOT CAUSE RC</p> <p>What is the real reason that this problem exists? What is the backstory behind the need to do this job?</p> <ul style="list-style-type: none"> • Buildup of fatty plaques in the arteries is the most common cause of coronary artery disease. • lack of exercise, obesity and smoking • Acute aortic insufficiency (AI). • To cure the diseased patients especially to visualize the heart problems and give relief to them. • One backstory is that many children are now affected with hole in the heart and suffer alot than elders, so this method is initiated • I feel like the final formal paper when human behaviour in this world is a problem in this affects the whole body • Thus, this visualization boards and any such heart diseases is predicted with an interactive dashboard. 	<p>7. BEHAVIOUR</p> <p>What does your customer do to address the problem and get the job done?</p> <ul style="list-style-type: none"> • Regular, daily physical activity can lower the risk of heart disease. Physical activity helps control your weight. • A healthy diet can help protect the heart, improve blood pressure, cholesterol, and reduce the risk of type 2 diabetes. • One of the best things you can do for your heart is to stop smoking using smokeless tobacco. Even if you're not a smoker, be sure to avoid secondhand smoke. • Maintain a healthy weight • Get good quality sleep • Manage stress • High blood pressure and high cholesterol can damage the heart and blood vessels. But without testing for them, you probably won't know whether you have these conditions. Regular screening can tell you what your numbers are and whether you need to take action.

<p>3. TRIGGERS TR</p> <p>What triggers customers to act? I.e. seeing their neighbours installing solar panels, reading about a more efficient solution in the news.</p> <ul style="list-style-type: none"> • Lifestyle changes • I have depending on medical support • need to spend for heart operated with more possible price • need to apply for health insurance • financial insecurity • Anxiety • shortness of breath • may feel emotional stress • may feel chest pain, chest tightness, chest pressure • feel for fatigue <p>4. EMOTIONS: BEFORE / AFTER EM</p> <p>How do customers feel when they face a problem or a job and afterwards?</p> <p>I.e. free, insecure, worried, worried – used in your communication strategy & design</p> <ul style="list-style-type: none"> • before a person knows that he/she is affected with any kind of disease, they are happy and do their work normally • They don't need to worry about their own body for any problems and do their work normally and comfortably • But, after a person comes to know about any kind of problems especially a heart disease, he/she becomes • Eless • unhappy • stressed/depressed • uncomfortable with their daily routines • Lifestyle becomes upside down 	<p>10. YOUR SOLUTION SL</p> <p>If you're working on an existing business, write down your current solution (not, fill in the canvas, and check how much it fits reality.</p> <p>If you're working on a new business proposition, then keep it blank until you fill in the canvas and come up with a solution that fits within customer limitations, solves a problem and matches customer behaviour.</p> <ul style="list-style-type: none"> • Heart disease treatment depends on the cause and type of heart damage. Healthy lifestyle habits – such as eating a low-fat, low-salt diet, getting regular exercise and good sleep, and not smoking – are an important part of treatment. • If lifestyle changes alone don't work, medications may be needed to control heart disease symptoms and to prevent complications. The type of medication used depends on the type of heart disease. • Some people with heart disease may need a procedure or surgery. The type of procedure or surgery will depend on the type of heart disease and the amount of damage to the heart. 	<p>8. CHANNELS of BEHAVIOUR CH</p> <p>8.1 ONLINE</p> <p>What kind of actions do customers take online? Extract online channels from #7</p> <ul style="list-style-type: none"> • Online appointments with doctors • Research about the heart diseases they are diagnosed with • Finding a good natural cause <p>8.2 OFFLINE</p> <p>What kind of actions do customers take offline? Extract offline channels from #7 and use them for customer development</p> <ul style="list-style-type: none"> • Maintaining proper diet and having healthy food • Having adequate amount of sleep • Maintaining a calm and relaxed mind state • Following the suggestions made by the doctors • Doing exercise and maintaining fitness • Taking the right doses of pills at the right time mentioned by doctors
---	--	--

3.4 Problem Solution fit

Proposed Solution Template:

Project team shall fill the following information in proposed solution template.

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	The goal is to accurately create a data set about the Heart patients so that the hospital can use this information to easily visualize and predict the patient details.
2.	Idea / Solution description	The parameters in the data set helps hospitals to identify the patient heart condition and their health condition. A informative and creative dashboard can be created to present the data and utilize it for future use.
3.	Novelty / Uniqueness	Treatment can be easy for the doctors on the basis of the patient heart condition. Time can be saved.
4.	Social Impact / Customer Satisfaction	It help the hospitals to know the health records of the heart patient. It will make the hospital to work efficiently.
5.	Business Model (Revenue Model)	Ad based revenue model - Awareness can be created among the patient through ads.
6.	Scalability of the Solution	Easy prediction of the patient details with heart disease. Maintains best user experience.

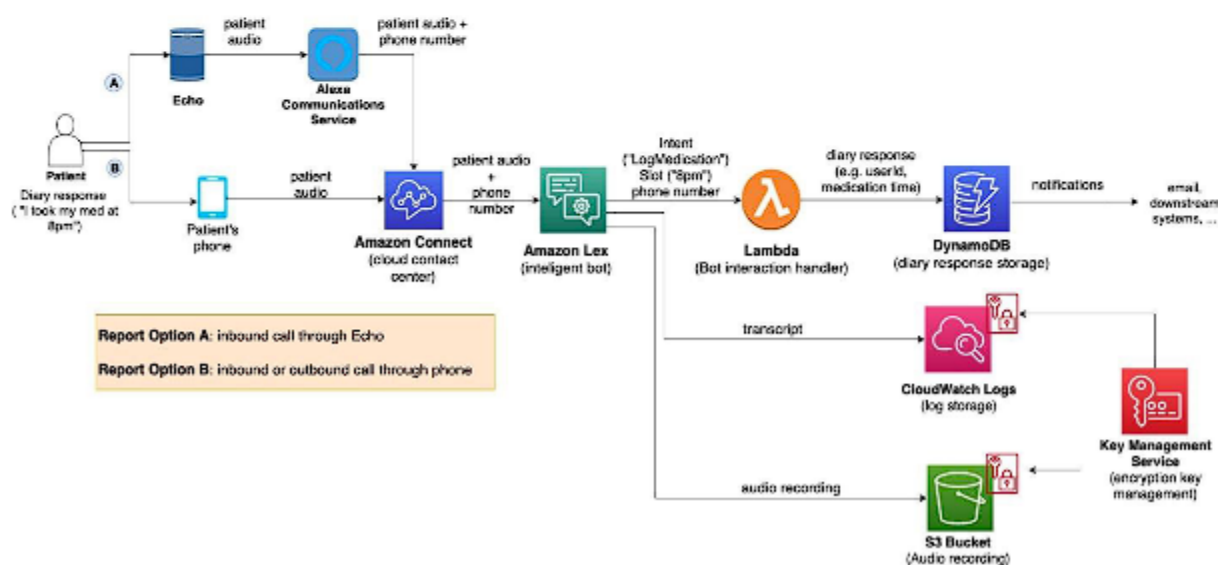
3.5 Solution Architecture

Solution Architecture:

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

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

Example - Solution Architecture Diagram:



4.REQUIREMENT ANALYSIS

4.1 Functional requirement

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Enables user to make registration for the application through Gmail
FR-2	User Confirmation	Once after registration, the user will get confirmation via Email
FR-3	Visualizing Data	User can visualize the trends on the heart disease through Dashboard created using IBM Cognos Analytics
FR-4	Generating Report	User can view his/her health report and can make decisions accordingly

4.2 Non-Functional requirements

NFR No.	Non-Functional Requirement	Description
NFR-1	Usability	The application will have a simple and userfriendly graphical interface. Users will be able to understand and use all the features of the application easily. Any action has to be performed with just a few clicks
NFR-2	Security	For security of the application the technique known as database replication should be used so that all the important data should be kept in multiple locations. In case of crash, the system should be able to take backup and recover the data
NFR-3	Reliability	The application has to be consistent at every scenario and has to work without failure in any environment

5.Customer Journey Map

Customer Journey Map

Use this framework to better understand customer needs, motivations, and obstacles by illustrating a key scenario or process from start to finish. When possible, use this map to document and summarize interviews and observations with real people rather than relying on your hunches or assumptions.

Created in partnership with

Document an Customer Journey Map

Narrow your focus to a specific scenario or process within an product or service. In the **Steps** row, document the step-by-step process someone typically experiences, then add detail to each of the other rows.

<p>Scenario</p> <p>Browsing, booking, attending, and rating a local city tour</p>	<p>Entice</p> <p>How does someone ideally become aware of this process?</p>	<p>Enter</p> <p>What do people experience as they begin the process?</p>
<p>Steps</p> <p>What does the person (or group) typically experience?</p>	<p>Research</p> <p>Research the process, the steps, the obstacles and the opportunities</p>	<p>Define the process</p> <p>Define the process, the steps, the obstacles and the opportunities</p>
<p>Interactions</p> <p>What interactions do they have at each step along the way?</p> <ul style="list-style-type: none"> People: Who do they see or talk to? Places: Where are they? Things: What digital touchpoints or physical objects would they use? 	<p>Identify the touchpoints</p> <p>Identify the touchpoints, the steps, the obstacles and the opportunities</p>	<p>Define the process</p> <p>Define the process, the steps, the obstacles and the opportunities</p>
<p>Goals & motivations</p> <p>On each step, what is a person's primary goal or motivation? (What do they want? or What do they avoid?)</p>	<p>Identify the touchpoints</p> <p>Identify the touchpoints, the steps, the obstacles and the opportunities</p>	<p>Define the process</p> <p>Define the process, the steps, the obstacles and the opportunities</p>
<p>Positive moments</p> <p>What steps does a typical person find enjoyable, productive, fun, motivating, delightful, or exciting?</p>	<p>Identify the touchpoints</p> <p>Identify the touchpoints, the steps, the obstacles and the opportunities</p>	<p>Define the process</p> <p>Define the process, the steps, the obstacles and the opportunities</p>
<p>Negative moments</p> <p>What steps does a typical person find frustrating, confusing, angering, costly, or time consuming?</p>	<p>Identify the touchpoints</p> <p>Identify the touchpoints, the steps, the obstacles and the opportunities</p>	<p>Define the process</p> <p>Define the process, the steps, the obstacles and the opportunities</p>
<p>Areas of opportunity</p> <p>How might we make each step better? What ideas do we have? What have others suggested?</p>	<p>Identify the touchpoints</p> <p>Identify the touchpoints, the steps, the obstacles and the opportunities</p>	<p>Define the process</p> <p>Define the process, the steps, the obstacles and the opportunities</p>

Need some inspiration?

See a collection of 40+ examples to inspire your work.

[View examples](#)

→

→

→

→

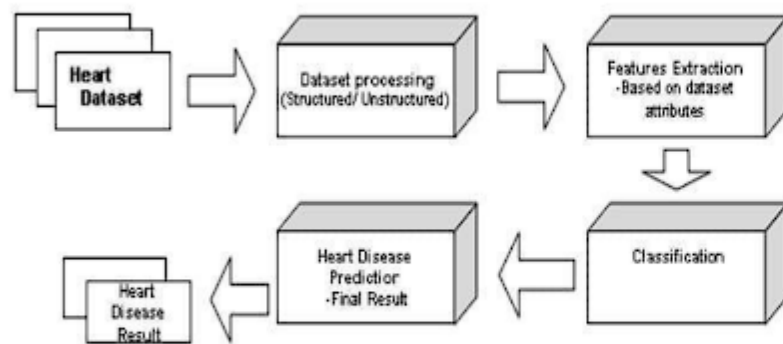
→

5.1 Technology Architecture

Technical Architecture:

The Deliverable shall include the architectural diagram as below and the information as per the table1 and 2

Example: Order processing during pandemics for offline mode



Guidelines:

1. Include all the process (Block)
2. Provide infrastructure
3. Indicate external interface
4. Indicate Data Storage
5. Indicate interface to

Table-1 : Components & Technologies:

S. No	Component	Description
1.	User Interface	How user interacts with application e.g. Web UI, Mobile App, Chatbot etc.
2.	Application Logic-1	Logic for a process in the application
3.	Application Logic-2	Logic for a process in the application
4.	Application Logic-3	Logic for a process in the application
5.	Database	Data Type, Configurations etc.
6.	Cloud Database	Database Service on Cloud

7.	File Storage	File storage requirements
8.	External API-1	Purpose of External API used in the application
9.	External API-2	Purpose of External API used in the application
10.	Machine Learning Model	Purpose of Machine Learning Model
11.	Infrastructure (Server / Cloud)	Application Deployment on Local System / Cloud Local Server Configuration: Cloud Server Configuration :

Table-2: Application Characteristics:

S.No	Characteristics	Description
1.	Open-Source Frameworks	List the open-source frameworks used
2.	Security Implementations	List all the security / access controls implemented, use of firewalls etc.
3.	Scalable Architecture	Justify the scalability of architecture (3 – tier, Micro-services)

4.	Availability	Justify the availability of application (e.g. use of load balancers, distributed servers etc.)
5.	Performance	Design consideration for the performance of the application (number of requests per sec, use of Cache, use of CDN's) etc.

6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

SPRINT 1:

Hello. Welcome to Cognos Analytics with Watson.

You can get started right away by taking a look at our introduction video, product tour and Getting Started tab.

[Watch video](#)[Take a product tour](#)

Quick launch



Upload data

Upload or drag and drop spreadsheets, csv files, and other data sources.



Prepare data

Use data modules to clean and connect data from multiple resources.



Exploration

Quickly find unbiased answers by identifying trends in your data with data exploration.

Search

New data module

New data module

New data module

New data module

New data module

New data module

New data module

New data module

New data module

New data module

New data module

New data module

New data module

New data module

New data module

New data module

New data module

New data module

New data module

New data module

New data module

New data module

New data module

New data module

New data module

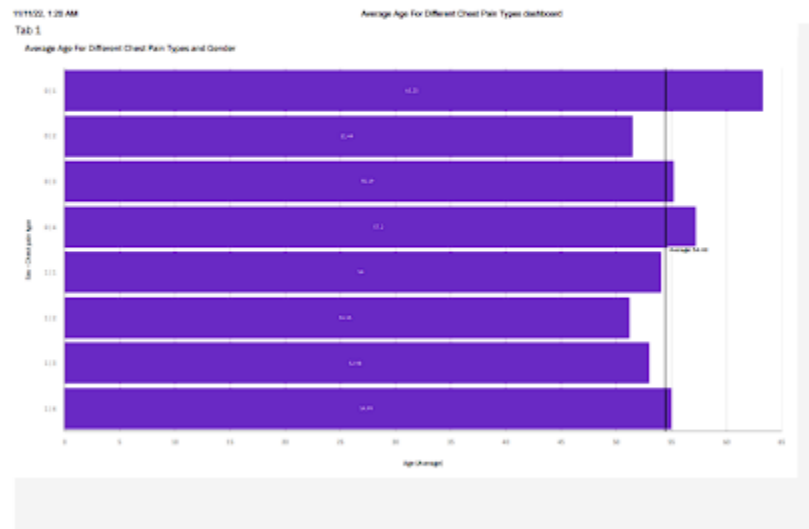
New data module

New data module

Grid Relationships Custom tables

Row Id	Age	Sex	Chest pain type	BP	Cholesterol
1	70	1	4	130	322
2	67	0	3	115	564
3	57	1	2	124	263
4	64	1	4	128	263
5	74	0	2	120	269
6	65	1	4	120	177
7	56	1	3	130	256
8	59	1	4	110	239
9	60	1	4	140	293
10	63	0	4	150	407
11	59	1	4	135	234
12	53	1	4	142	226
13	44	1	3	140	235
14	61	1	1	134	234
15	57	0	4	128	303
16	71	0	4	112	149
17	46	1	4	140	311
18	53	1	4	140	203

DATA EXPLORATION



<https://r1-ca.analytics.ibm.com/hr/perspective-dashboard?d=SCD47K0095044C2B021FABBE034M476&opt=SCD47K0095044C2B021FABBE034M476&opt=N%3D&viewClass=Profile3%3D=true&opt=...> 1/1

Example 1

Heart Disease for Chest pain type and Sex

Heart Disease	1	2	3	4	Summary
0		4	36	32	72
1	16	26	87	90	219
Summary	20	42	123	122	207

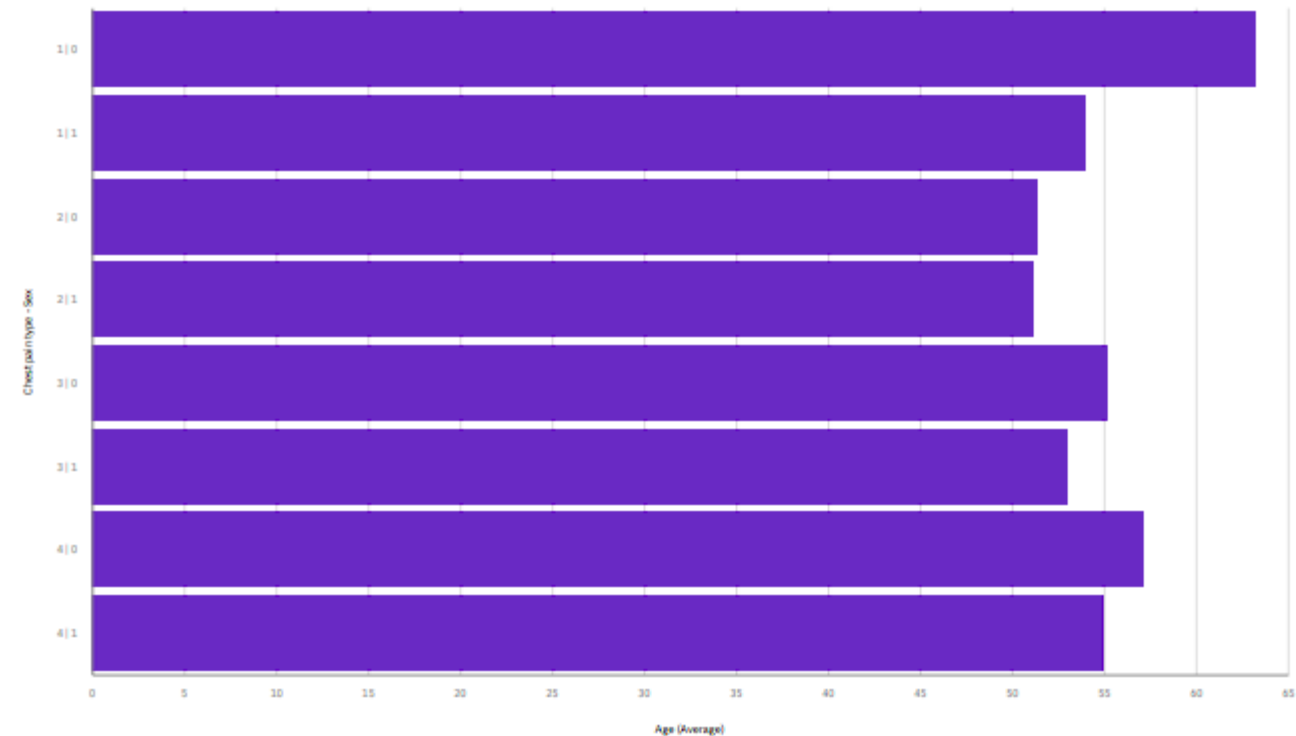
https://a1.analytics.ibm.com/?perspective=full&id=2020F815362457647721448677166&_app=trials&from=VS&activeCaseProject=MS&trials=actions/VScollections/MS/MSnew&date... 11



Sprint 3:

Tab 1

Age by Chest pain type and Sex



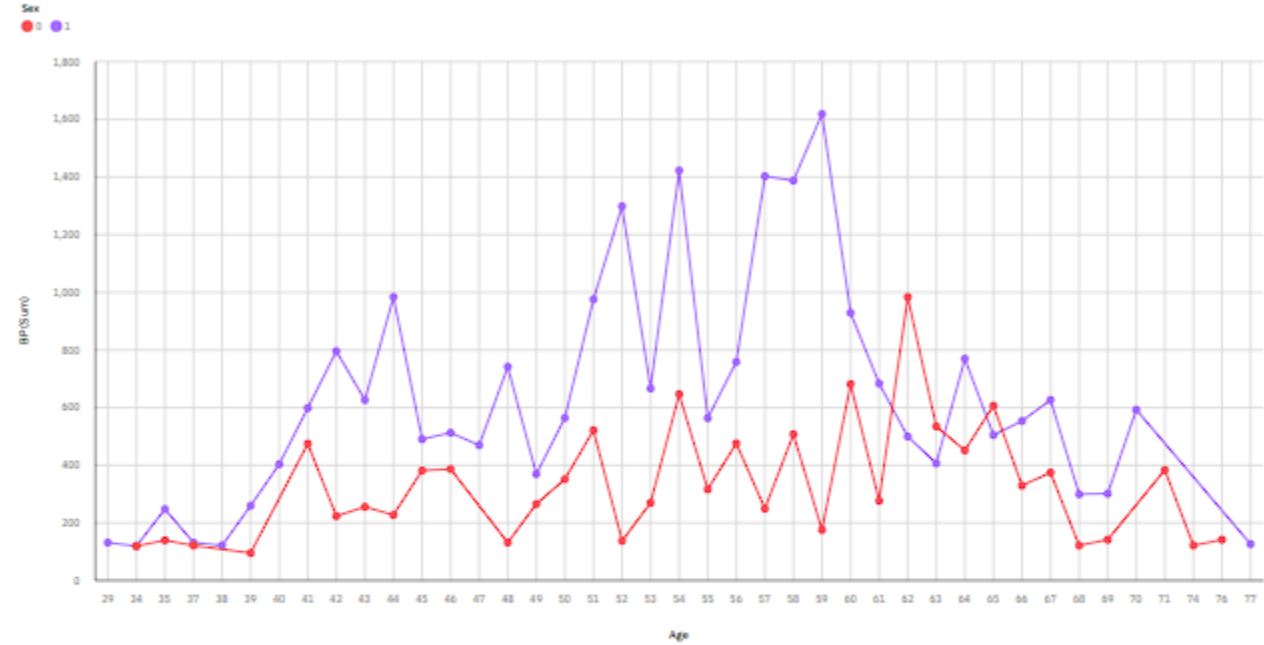
Tab 3

Age for Chest pain type and Sex

Age	1	2	3	4	Summary
0	63.25	51.64	55.19	57.2	55.68
1	54	51.15	52.98	54.99	53.84
Summary	55.85	51.26	53.87	55.59	54.43

Tab 2

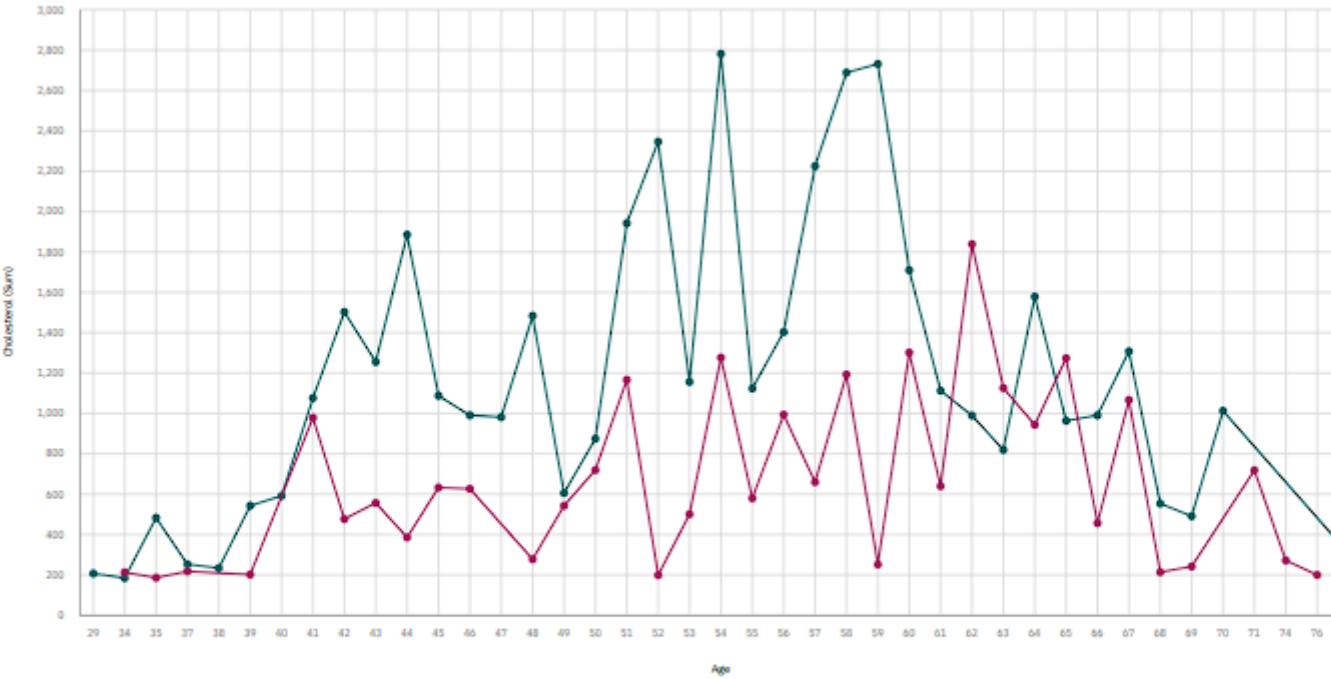
BP by Age colored by Sex



Tab 4

Cholesterol by Age colored by Sex

Sex
0 1



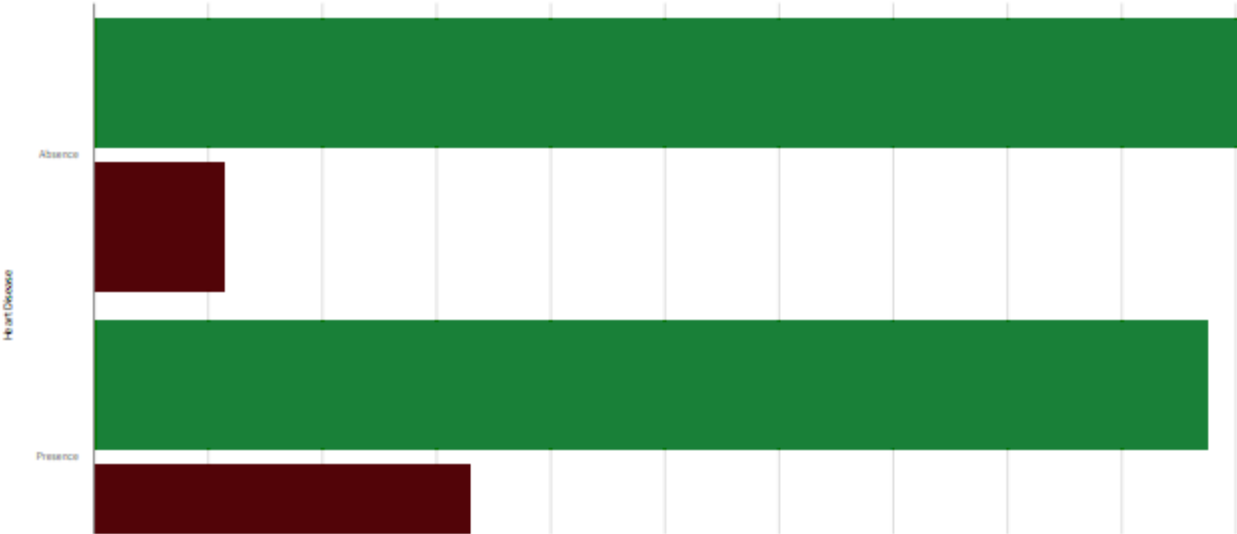
Tab 5

Max HR and Exercise angina by Heart Disease

Measures

● Max HR

● Exercise angina



SPRINT 4:

Visualizing And Predicting Heart Diseases With An Interactive Dash Board

1.1 ABSTRACT

The leading cause of death in the developed world is heart disease. Therefore, there needs to be work done to help prevent the risks of having a heart attack or stroke. Healthcare industries generate enormous amount of data, so called big data that accommodates hidden knowledge or pattern for decision making. The huge volume of data is used to make decision which is more accurate than intuition. Exploratory Data Analysis (EDA) detects mistakes, finds appropriate data, checks assumptions and determines the correlation among the explanatory variables. In the context, EDA is considered as analysing data that excludes inferences and statistical modelling. Analytics is an essential technique for any profession as it forecast the future and hidden pattern. Data analytics is considered as a cost effective technology in the recent past and it plays an essential role in healthcare which includes new research findings, emergency situations and outbreaks of disease. The use of analytics in healthcare improves care by facilitating preventive care and EDA is a vital step while analysing data.

1.2 MOTIVATION & GOAL

The goal is to accurately predict which patients are mostly likely to suffer from a heart disease in the near future using the features given.

1.3 TECHNICAL ARCHITECTURE



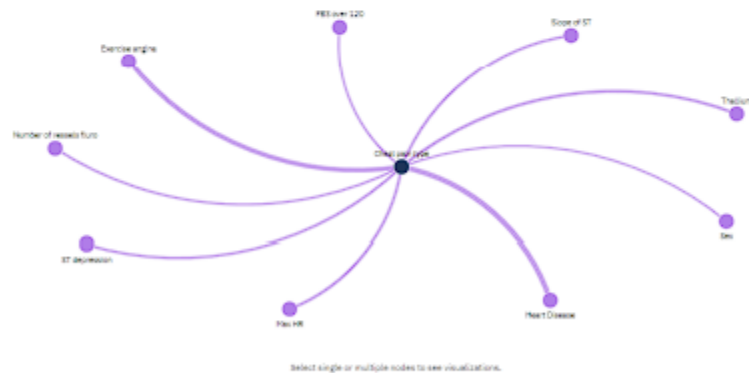
1.4 FINAL DELIVERABLES

1.4.1 SPRINT 1

In First sprint, We took Heart Disease prediction dataset from kaggle and preprocessing both in Jupyter notebook and IBM cognos.To find out null values present in the dataset and change it into a certain value.

Row ID	Age	Sex	Chest pain type	BP	Cholesterol	FBS over 120	CK-MB	Max HR	Exercise eng
1	39	1	4	130	202	0	2	159	0
2	47	0	3	115	164	0	2	140	0
3	57	1	2	134	263	0	6	145	0
4	46	1	6	138	263	0	6	139	1
5	76	0	2	130	269	0	2	125	1
6	45	1	4	130	137	0	6	140	0
7	56	1	3	130	256	1	3	147	1
8	59	1	4	130	229	0	2	147	1
9	40	1	4	140	280	0	2	179	0
10	40	0	4	150	487	0	2	154	0
11	59	1	4	115	234	0	6	146	0
12	51	1	6	142	206	0	2	111	1
13	46	1	3	140	235	0	2	189	0
14	41	1	1	134	234	0	6	140	0
15	57	0	4	135	383	0	2	159	0
16	71	0	4	112	149	0	6	126	0
17	46	1	6	140	103	0	6	129	1
18	50	1	6	140	203	1	2	159	1

After preprocessing the data the dataset is explored in IBM cognos through various graphs and tabular columns by comparing it with different sets of attributes.

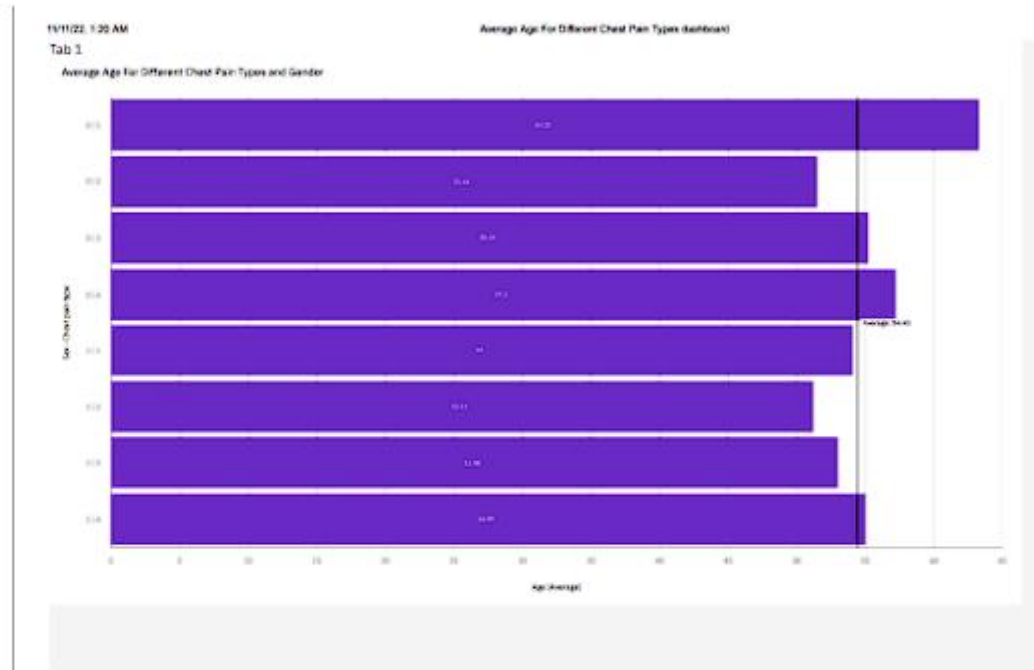


1.4.2 SPRINT 2

After preprocessing and exploring the heart disease prediction data.

Now upload the preprocessed

data to dashboard. In here you can able to create dashboard's for each and single graph representation. This will help to make sure to find the Average availability of beds in the hospitals.



Tab 1

Average Age For Different Types Of Chest Pain In Existing Heart Diseases dashboard

Heart Disease for Chest pain type and Sex

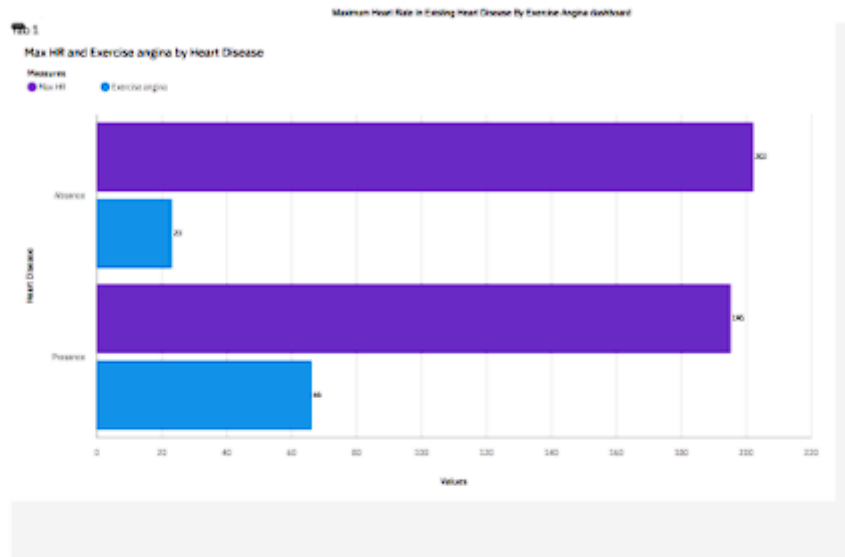
Heart Disease	1	2	3	4	Summary
0	6	16	32	35	87
1	16	26	47	60	149
Summary	20	42	79	129	270

DP Variation With Respect To Age dashboard

59

Serum Cholesterol Levels By Age - Infoboard

2



1.4.3 SPRINT 3

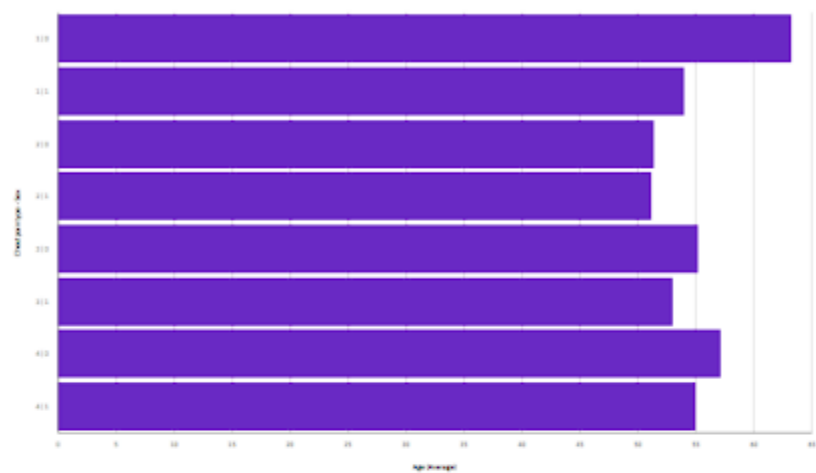
Finally after creating a dashboard for the health-care dataset. Now build a report and story using same method. Report can be build by importing the data exploration templates or templates which was available in the dashboard. This will make sure to create a organized way of report and story of the given dataset even layman can understand.

11/16/22, 9:36 PM

SpineD

Tab 1

Age by Chest pain type and Sex

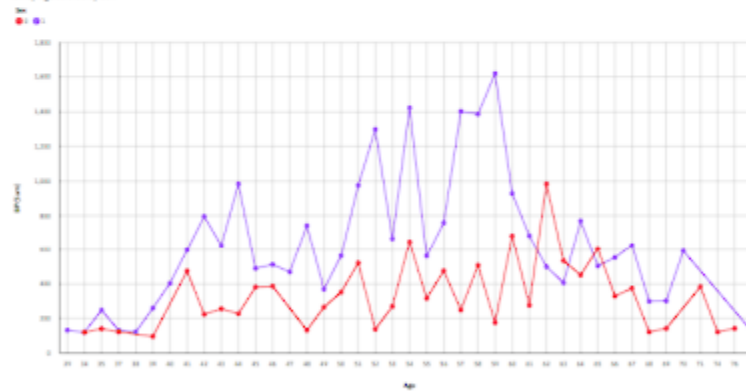


11/16/22, 9:36 PM

SpineD

Tab 2

BP by Age colored by Sex



11/16/22, 9:35 PM

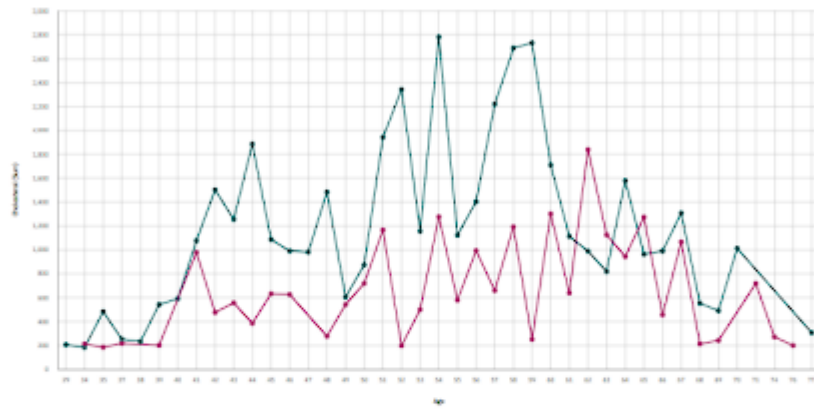
Spine0

Tab 4

Cholesterol by Age colored by Sex

Sex

● M ● F



11/16/22, 9:35 PM

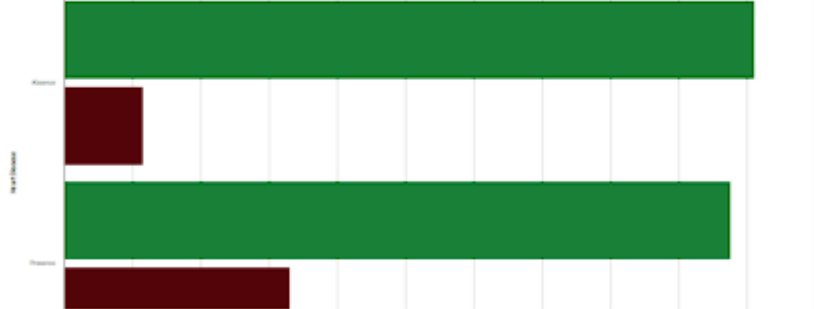
Spine0

Tab 5

Max HR and Exercise angle by Heart Disease

Measure

● Max HR ● Exercise angle



To create a story, which is quite similar to slideshare. We need to build the templates and place it in form of powerpoint and allot them with given time for each slide. So that we can able to create a video representation of the received graph representation.

1.4.4 Sprint 4

In final sprint we will document the process we followed throughout the entire project to give desired results and submit it. Finally, we came to :

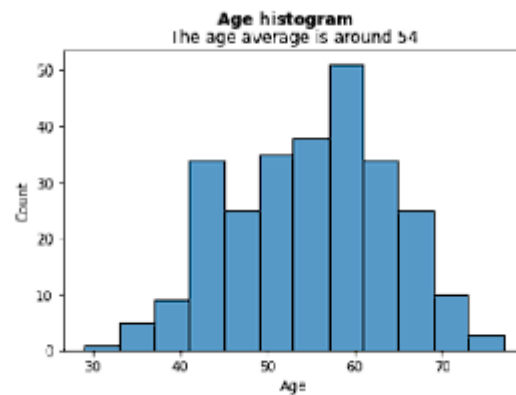
- > Know fundamental concepts and can work on IBM Cognos Analytics
- > Gain a broader understanding of plotting different visualizations to provide the suitable solution
- > Able to create meaningful visualizations and the Dashboard(s).

7. Output Screenshots:

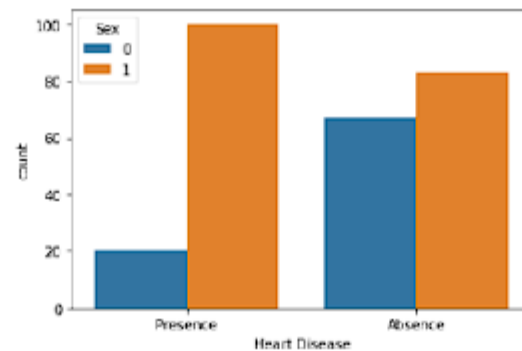
8.

Out[4]:

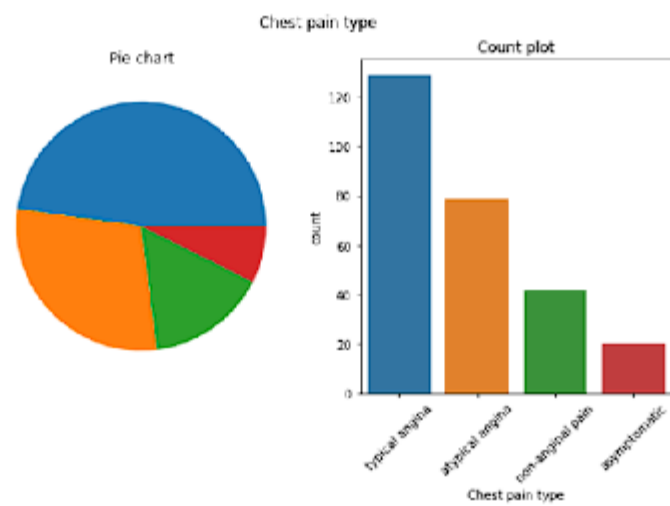
	Age	Sex	Chest pain type	BP	Cholesterol	FBS over 120	EKG results	Max HR	Exercise angina	ST depression	Slope
0	70	1	4	130	322	0	2	109	0	2.4	
1	67	0	3	115	564	0	2	160	0	1.6	
2	57	1	2	124	261	0	0	141	0	0.3	
3	64	1	4	120	253	0	0	105	1	0.2	
4	74	0	2	120	269	0	2	121	1	0.2	



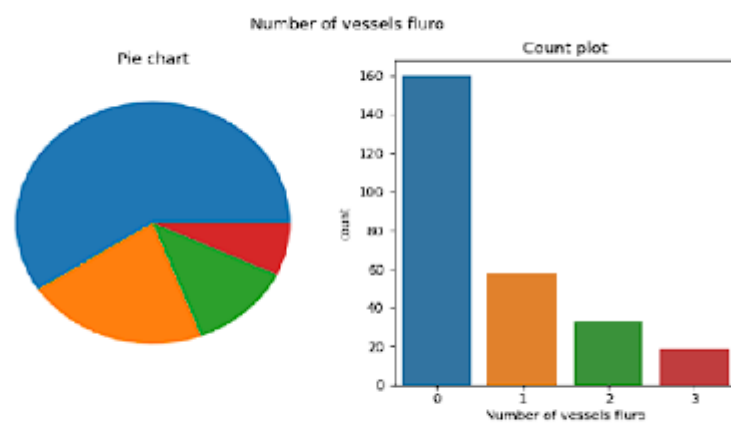
Out[11]: (AxesSubplot:xlabel='Heart Disease', ylabel='count')



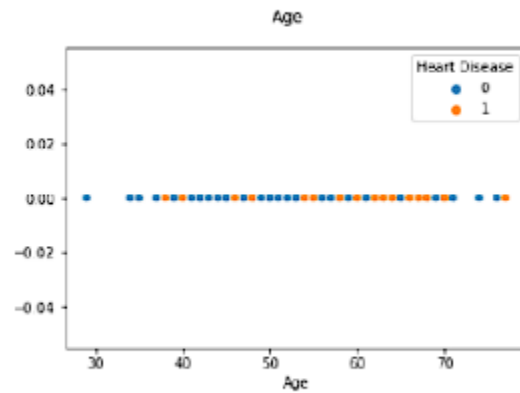
It is observed that Males have more chances to have heart disease than Female.



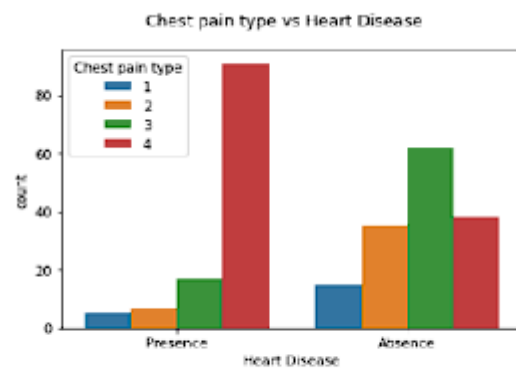
```
Out[12]: 4    129
          3     79
          2     42
          1     20
          Name: Chest pain type, dtype: int64
```



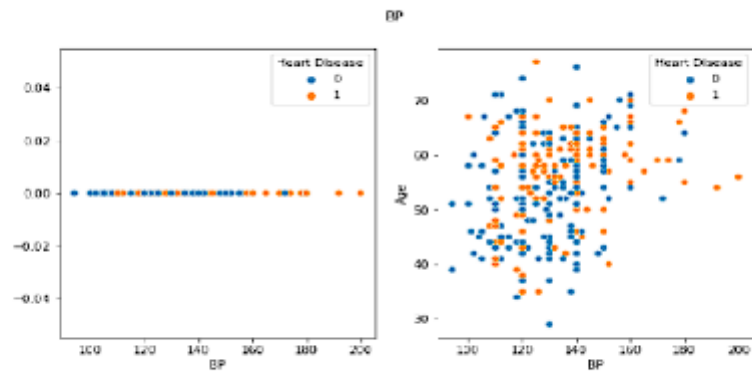
```
Out[15]: 0    160
          1     58
          2     33
          3     19
          Name: Number of vessels fluro, dtype: int64
```



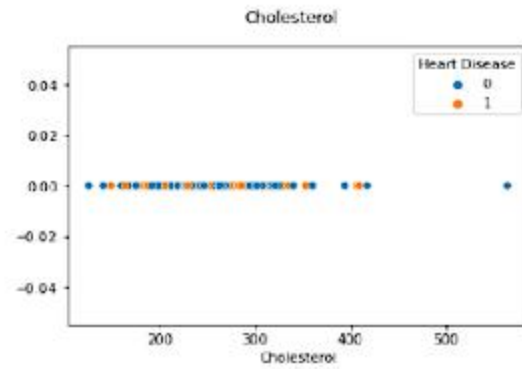
Heart disease based on age – As You Can See That Older people have more chance to have heart disease.



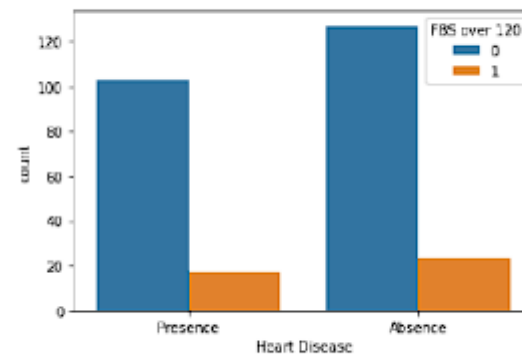
Heart disease based on Chest pain type - 4th type of chest pain dominate in heart disease.



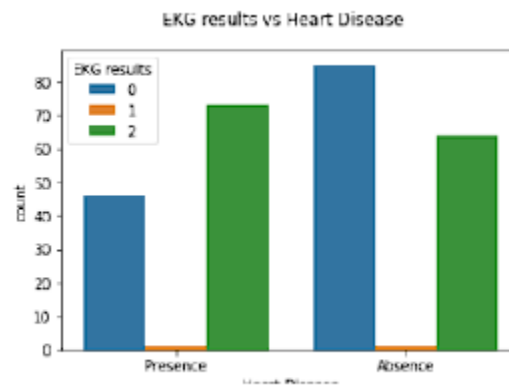
Heart Disease based on BP - Persons with high BP have more chance to get heart disease.



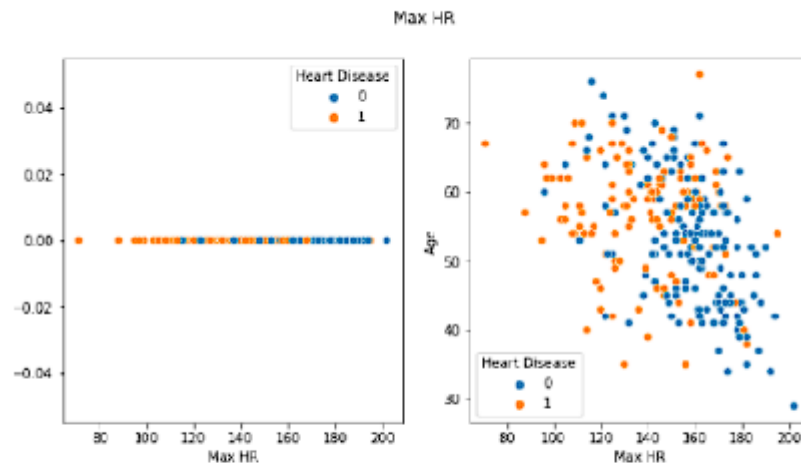
Cholesterol - Higer Cholesterol does not influence on heart disease.



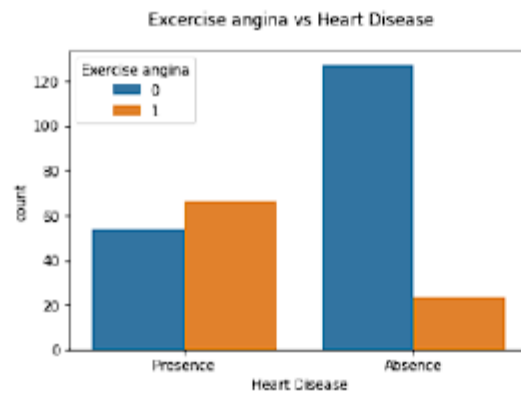
FBS over 120 - Also Increased FBS over 120 does not imply on heart disease prediction.



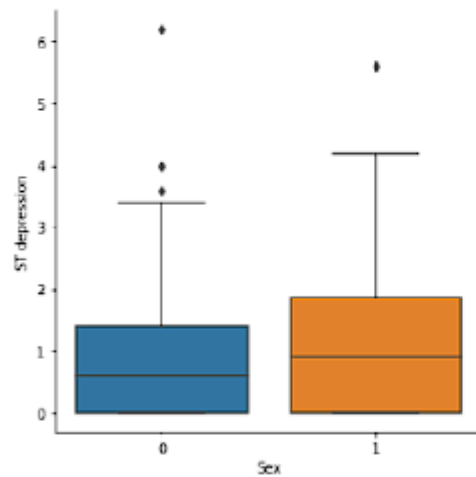
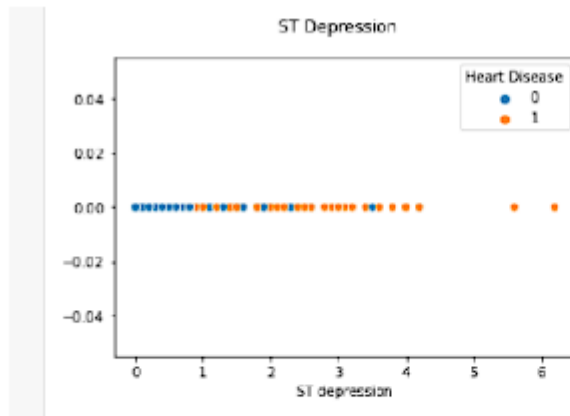
EKG results - The 2nd value of EKG could influence on heart disease prediction.



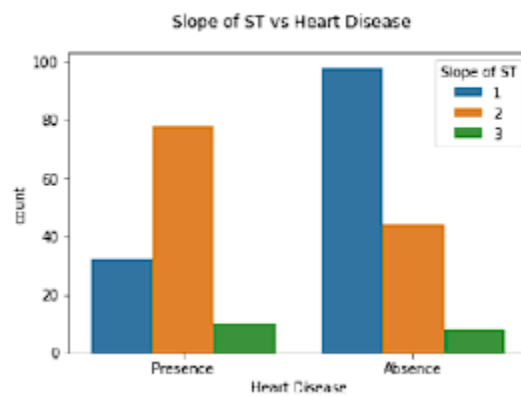
Max HR - From the first graph we can observe that people with lower HR max have a higher likelihood of heart disease than those with higher HR max. Furthermore, we can observe explicit cut off/threshold where below 120 HR max objects have a higher probability to have problem with heart.



Exercise angina - Chest pain after a high exercise intensity or stress can casues a presence of heart failure.



ST depression - Increased ST depression increase heart disease. Nevertheless it can be observe on the bottom figure that males have higher probability of having depression.



8.Models used and Accuracy Scores

Linear Regression:

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

```
Root mean squared error: 0.36416508651208407  
r2: 0.4611150822223089
```

Logistic Regression:

This type of statistical model is often used for classification and predictive analytics. Logistic regression estimates the probability of an event occurring, such as voted or didn't vote, based on a given dataset of independent variables. Since the outcome is a probability, the dependent variable is bounded between 0 and 1. In logistic regression, a logit transformation is applied on the odds—that is, the probability of success divided by the probability of failure.

```
Logistic Regression score: 0.875
```

Gaussian NB:

Gaussian Naive Bayes supports continuous valued features and models each as conforming to a Gaussian (normal) distribution. An approach to create a simple model is to assume that the data is described by a Gaussian distribution with no co-variance (independent dimensions) between dimensions.

```
Gaussian Naive Bayes score: 0.7708333333333334
```

KNeighborsClassifier:

The k-nearest neighbors algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. KNeighborsClassifier looks for the 5 nearest neighbors. We must explicitly tell the classifier to use Euclidean distance for determining the proximity between neighboring points.

```
KNeighbourClassifier score: 0.75
```

Random Forest Classifier:

A random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.

One of the most important features of the Random Forest Algorithm is that it can handle the data set containing continuous variables as in the case of regression and categorical variables as in the case of classification. It performs better results for classification problems.

```
RandomForest score: 0.8333333333333334
```

10.ADVANTAGES AND DISADVANTAGES:

ADVANTAGES:

- This is one of the fastest ways to determine if a person is likely to suffer from a heart disease or not.
- Useful for medical practitioners to easily classify their patients.
- User Friendly
- Easy to understand
- Secure
- Dashboard provides insightful informations

DISADVANTAGES:

- Users need to know all the fields
- Does Not take null value as input
- Does not provide suggestions to user

11.CONCLUSION:

The project involved analysis of the heart disease patient dataset with proper data processing. Then, 5 models were trained and tested with maximum scores as follows:

1. K Neighbors Classifier: 75%
2. Linear Regression: RMSE(0.36), R2-0.46

3. Logistic Regression:87%

4. Random Forest Classifier:83%

5. GaussianNB:77%

13.APPENDIX:

SOURCECODE:<https://www.kaggle.com/datasets/rishidamarla/heart-disease-prediction>

GITHUB LINK: <https://github.com/IBM-EPBL/IBM-Project-25207-1659954943>

PROJECTDEMOLINK:

<https://drive.google.com/file/d/1qOtxWWDV8KVHH5pev7pJJE7dXDYfIHAY/view?usp=drivesdk>