VISUALIZATION AND PREDICTING HEART DISEASE WITH AN INTERACTIVE DASHBOARD

TEAM ID:PNT2022TMID53509

GIT REPO ID:

TEAM MEMBERS:

S SAILESH BAABU (TEAM LEADER)
A PRADEEP RAHUL
S RAVICHANDRAN
J SUJEITH KUMAR

Git Hub Repo Link: https://github.com/IBM-EPBL/IBM-Project-20282-1659716490

Demo Video Link: https://drive.google.com/file/d/1P_ZvXFw3pzMbHpyGkMq7Z7W2qPhOZKMS/view?usp=share_link

Repo ID: IBM-EPBL/IBM-Project-20282-1659716490

Project Report Format

1. **INTRODUCTION**
   - 1.1 Project Overview
   - 1.2 Purpose
2. **LITERATURE SURVEY**
   - 2.1 Existing problem
   - 2.2 References
   - 2.3 Problem Statement Definition
3. **IDEATION & PROPOSED SOLUTION**
   - 3.1 Empathy Map Canvas
   - 3.2 Ideation & Brainstorming
   - 3.3 Proposed Solution
   - 3.4 Problem Solution fit
4. **REQUIREMENT ANALYSIS**
   - 4.1 Functional requirement
   - 4.2 Non-Functional requirements
5. **PROJECT DESIGN**
   - 5.1 Data Flow Diagrams
   - 5.2 Solution & Technical Architecture
   - 5.3 User Stories
6. **PROJECT PLANNING & SCHEDULING**
   - 6.1 Sprint Planning & Estimation

INTRODUCTION:

1.1 Project Overview

Cardiovascular diseases (CVDs) are the leading cause of death globally, taking an estimated 17.9 million lives each year The effects of behavioral risk factors may show up in individuals. Identifying those at highest risk of CVDs and ensuring they receive appropriate treatment can prevent premature deaths.

This project is used to predict and visualize the heart diseases of a patient using an interactive dashboard. The data collected from the patients as different attributes such as age , sex , heart rate ... that are used to analyze and predict whether the person it prone to heart disease or not ,Based on the data collected, we will analyze the patients health conditions and come out with an interactive dashboard that will show the patient's condition ,which can be used to take preventive treatment or proactive actions

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 cardiovascular disease is ‘plaque’ which is a hard substance formed in the coronary arteries which is made up of cholesterol (fat), causing 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 attacks 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 amounts of data, so called big data that accommodates hidden knowledge or patterns for decision making. The huge volume of data is used to make decisions which are 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 forecasts 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 2020:** 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 metnes. In the result section, the visualized data shows that the prediction is accurate

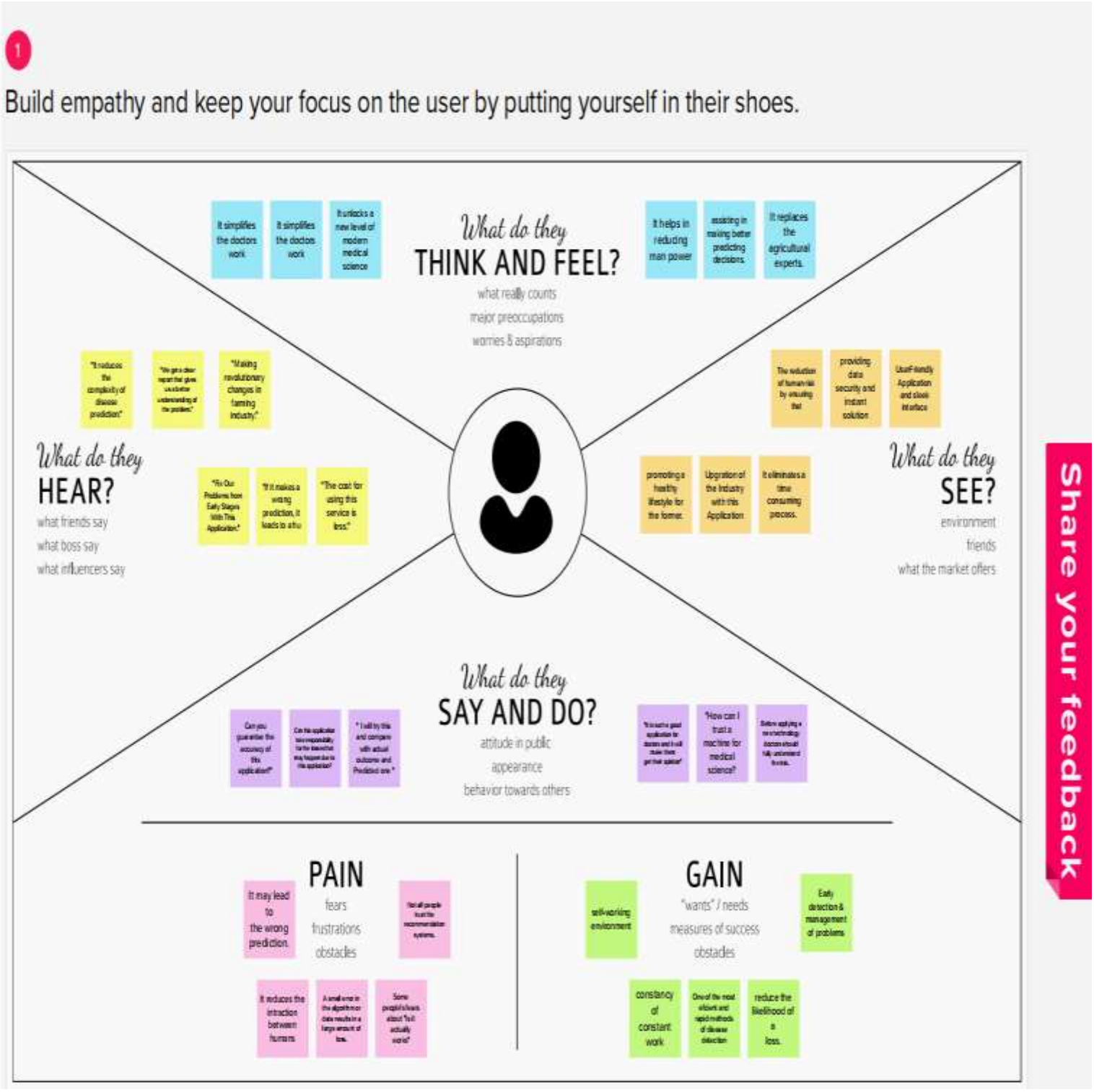
2. **Prediction of heart disease at an early stage using data mining and big data analytics: A survey by N. K. Salma Banu, Suma Swamy:** Several studies have been carried out for developing prediction models using individual techniques and also by combining two or more techniques This paper provides 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 models 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

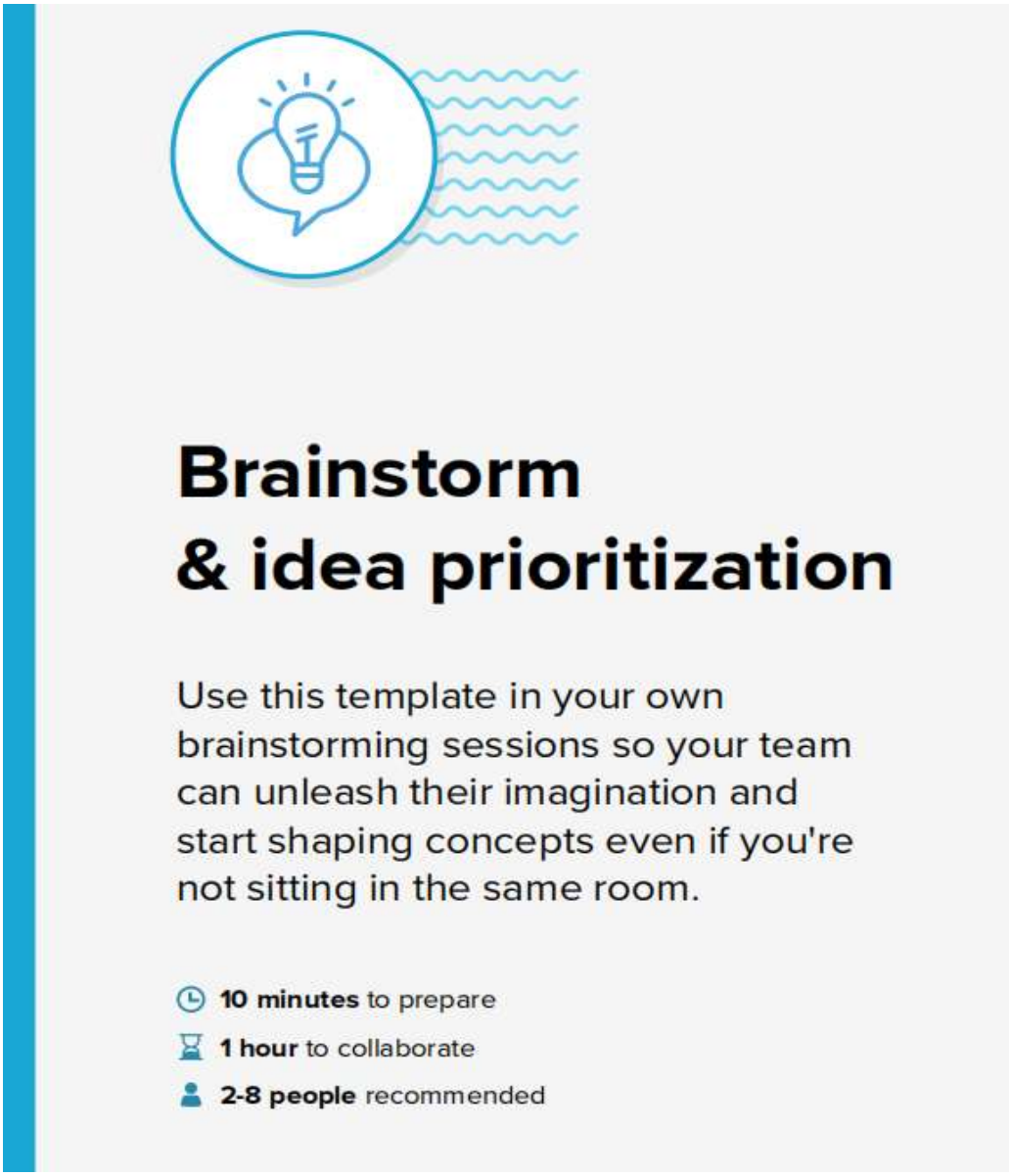
Heart disease can be managed effectively with a combination of lifestyle changes, medicine and,in some cases, surgery. With the right treatment, the symptoms of heart disease can be reduced and the functioning of the heart can be improved. The predicted results can be used to prevent and thus reduce 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 Ideation & Brainstorming



Before you collaborate

A little bit of preparation goes a long way with this session. Here's what you need to do to get going.

10 minutes

A **Team gathering**

Define who should participate in the session and send an invite. Share relevant information or pre-work ahead.

B **Set the goal**

Think about the problem you'll be focusing on solving in the brainstorming session.

C **Learn how to use the facilitation tools**

Use the Facilitation Superpowers to run a happy and productive session.

Open article →

Key rules of brainstorming

To run an smooth and productive session

- Stay in topic.
- Encourage wild ideas.
- Defer judgment.
- Listen to others.
- Go for volume.
- If possible, be visual.

1

Define your problem statement

What problem are you trying to solve? Frame your problem as a How Might We statement. This will be the focus of your brainstorm.

5 minutes

PROBLEM

VISUALIZING AND PREDICTING HEART DISEASE WITH A INTERACTIVE DASHBOARD

3

Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. In the last 10 minutes, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

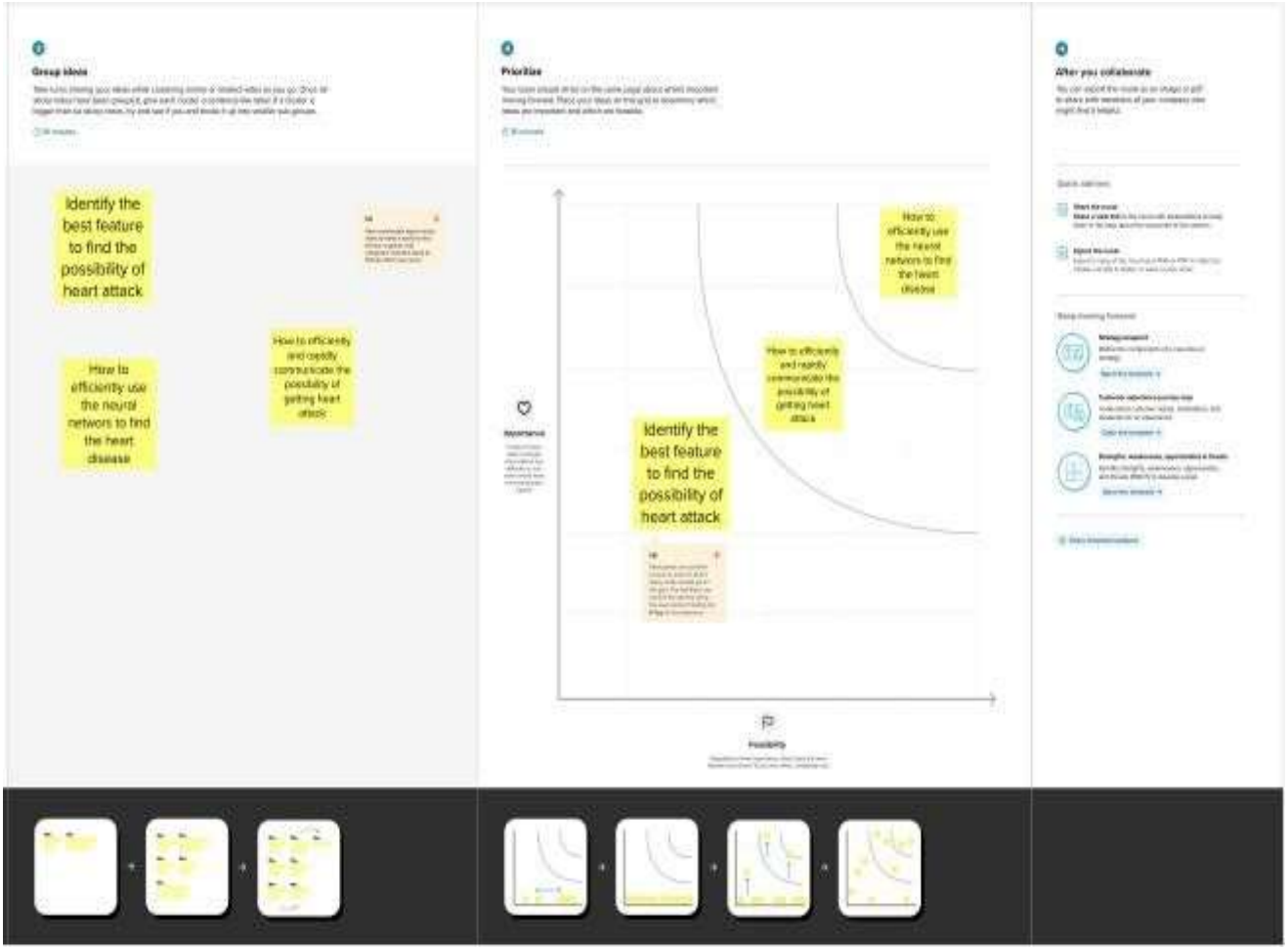
20 minutes

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



After you collaborate

You can export the mural as an image or pdf to share with members of your company who might find it helpful.

Quick add-ons

A **Share the mural**

Share a view link to the mural with stakeholders to keep them in the loop about the outcomes of the session.

B **Export the mural**

Export a copy of the mural as a PNG or PDF to attach to emails, include in slides, or save in your drive.

Keep moving forward

Strategy blueprint

Define the components of a new idea or strategy.

Open the template →

Customer experience journey map

Understand customer needs, motivations, and obstacles for an experience.

Open the template →

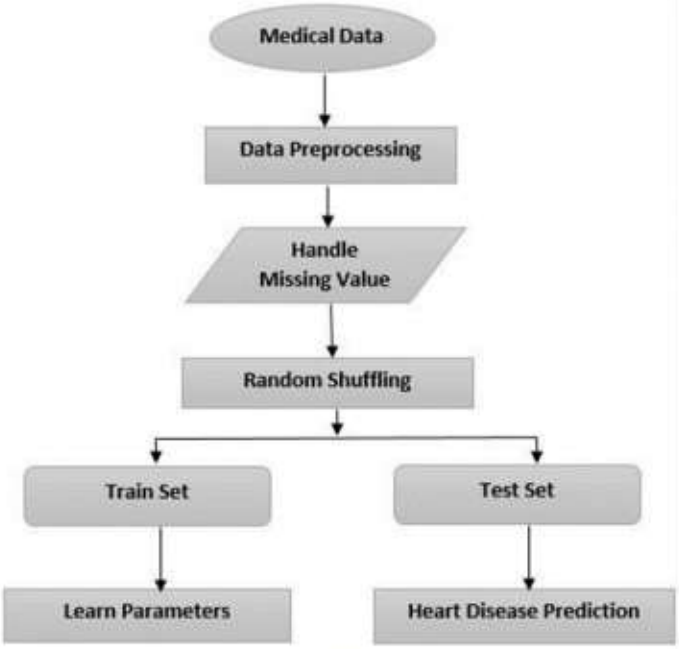
Strengths, weaknesses, opportunities & threats

Identify strengths, weaknesses, opportunities, and threats (SWOT) to develop a plan.

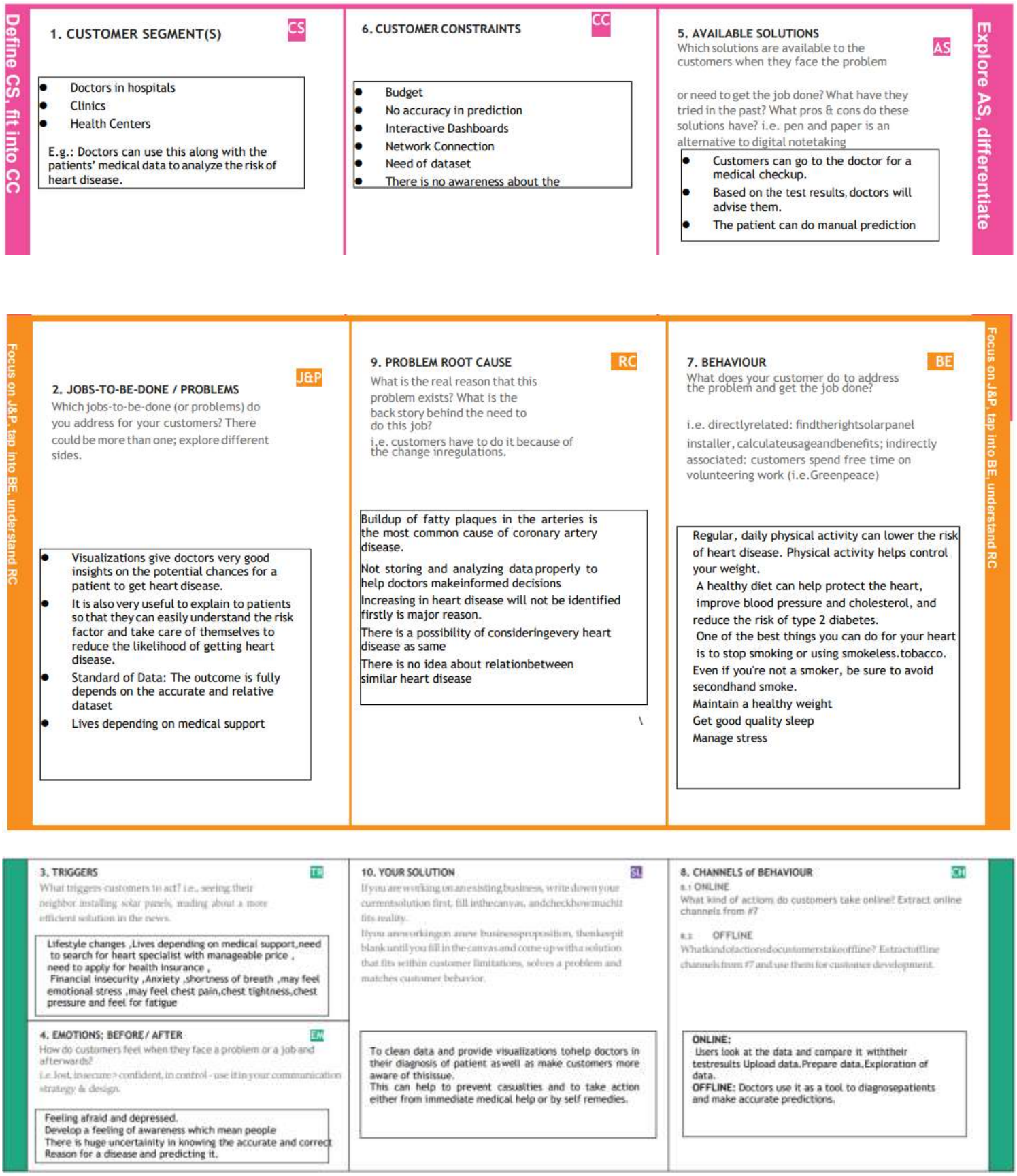
Open the template →

3.3 Proposed Solution

S.No	Parameter	Description
1.	Problem Statement (Problem to be solved)	**Visualizing and Predicting Heart Diseases with an Interactive Dashboard** Heart disease is the leading cause of death for men, women, and people of most racial and ethnic groups in the World. Lack of proper diagnostic tools and accurate results affect the treatment of cardiac patients, the diagnosis and treatment processes are currently quite difficult. Based on a patient's medical history, an expert's symptom analysis report, and physical laboratory results, invasive procedures are used to identify cardiac problems. The goal is to accurately create a data set about Heart patients so that the hospital can use this information to easily visualize and predict the patient details. Furthermore, because of human intervention, delays result in inaccurate diagnoses. Based on a variety of symptoms, including age, gender, pulse rate, physical examination, symptoms, signs of the patient, etc., heart disease can be anticipated.
2.	Idea / Solution Description	The main idea of our project is to use classification and regression techniques in IBM Cognos Analytics Application. The parameters in the data set help hospitals identify the patient heart condition and health condition. An informative and creative dashboard can be created to present the data and utilize it for future use.
3.	Novelty / Uniqueness	The proposed system gets inputs directly from the user for parameters such as age, BP level, cholesterol level, smoker history, heart rate, etc. IBM Cognos Analytics is used for learning relationships among input

		parameters, answering complex queries, with better accuracy, and providing the optimal solution.
4.	Social Impact / Customer Satisfaction	The provision of high-quality services at reasonable prices is a significant problem for healthcare institutions, including hospitals and medical facilities. The provision of high-quality care necessitates accurate patient diagnosis and efficient treatment delivery. Both numerical and categorical data are present in the heart disease database that is accessible. These entries are cleaned and filtered to remove any extraneous data from the database before being subjected to further processing. Complex questions for heart disease diagnosis can be answered by extracting hidden knowledge, i.e., patterns and relationships related to heart illness, from a historical heart disease database. As a result, it may aid medical professionals in making wise clinical judgments.
5.	Business Model (Revenue Model)	
6.	Scalability of the Solution	Easy prediction of the patient details with heart disease. Maintains the best user experience. Reduced Cost. Reduce time management complexity of doctors Faster and more accurate prediction virtually Decrease Mortality rate Reduce the risk of critical factors Analyze in-depth focus on anticipated risks

3.4 Problem Solution fit



4.REQUIREMENT ANALYSIS

4.1 Functional requirement

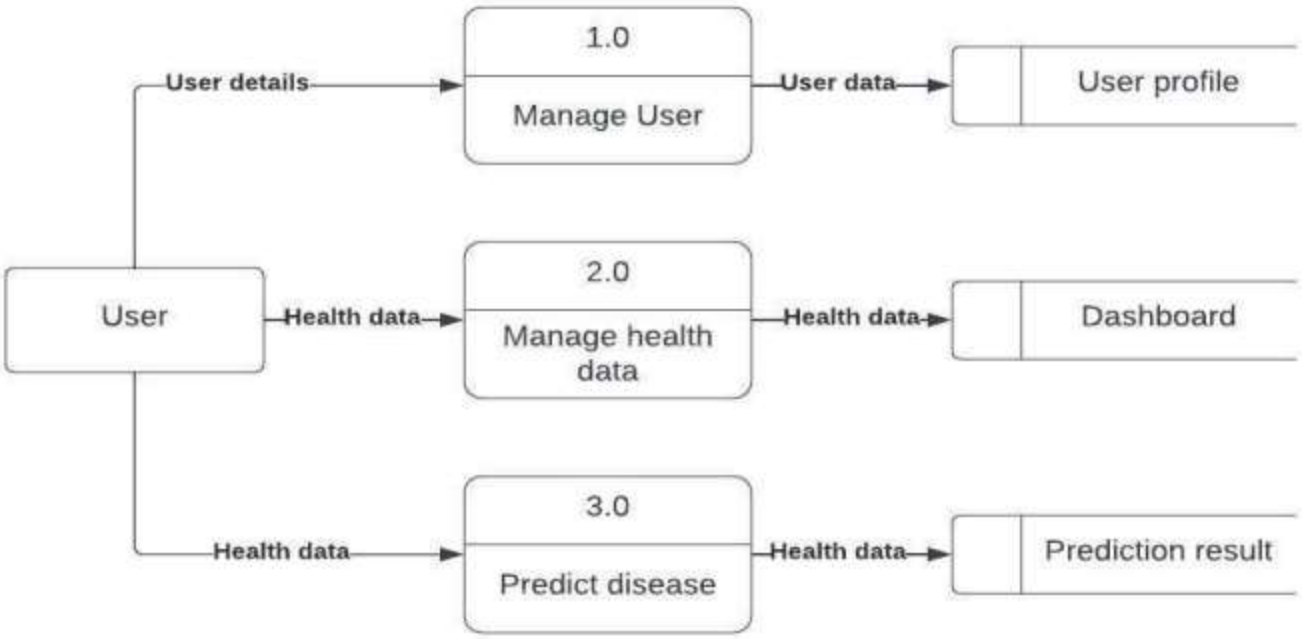
FR No.	Requirement	Description
1.	User Registration	User register into the login portal using the username and password
2.	User Input	User selects the category that they want to visualize and give the inputs
3.	Visualization of data	The data provided will be analyzed and creative dashboards can be visualized by the user
4.	Report generation	Report will be generated based on the analysis and if high risk the user data will be provided to the nearby hospitals

4.2 Non-Functional requirements

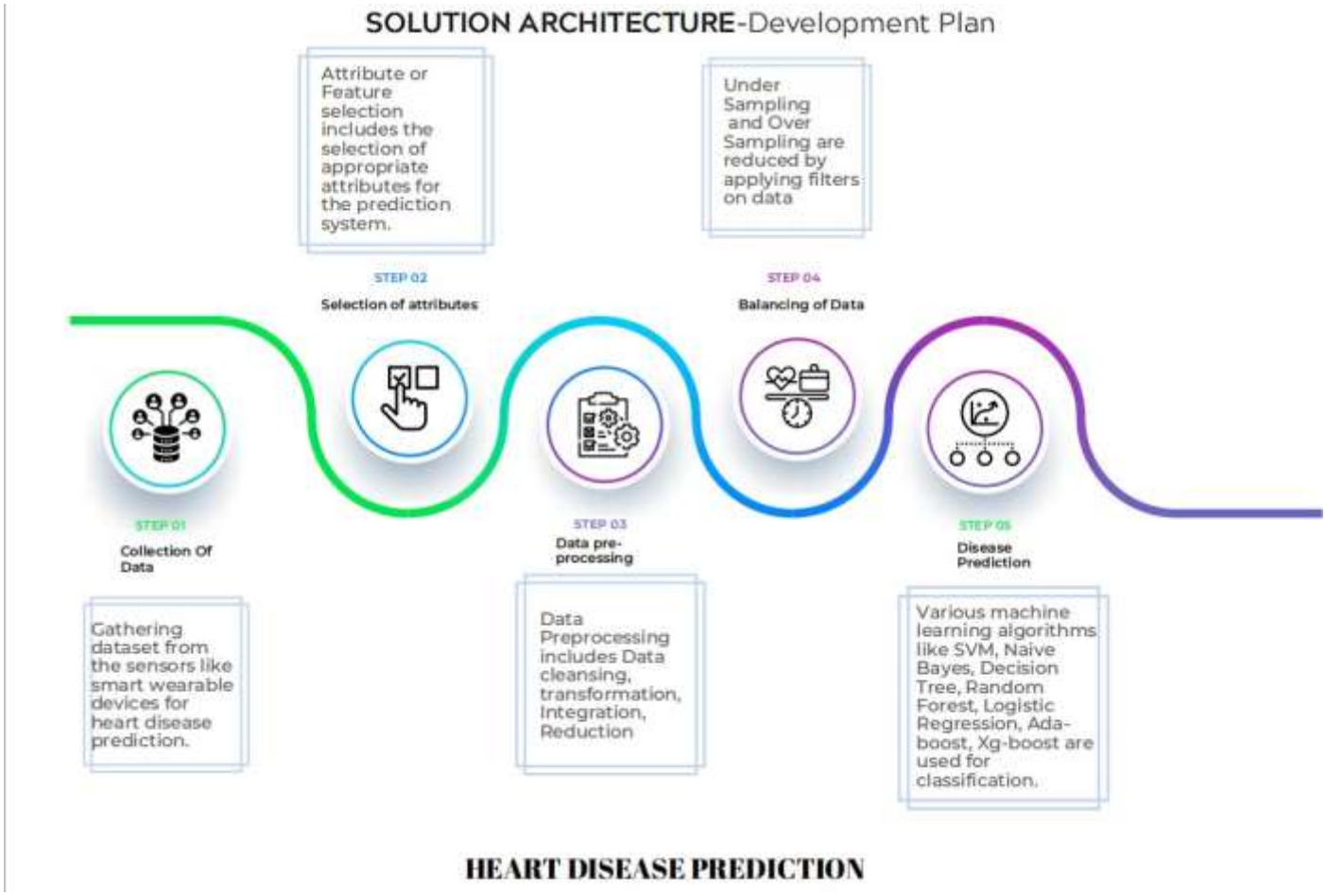
NFR No.	Non-Functional Requirement	Description
1.	Usability	The application will have a simple and user friendly graphical interface. Users will be able to understand and use all the features of the application easily. Any action has to be performed with just a few clicks and the portal will be active
		24hrs
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 safe. Incase of crash, the system should be able to backup and recover the data
3.	Reliability	The application has to be consistent at every scenario and has to work without failure in any environment
4.	Performance	Performance of the application depends on the response time and the speed of the data submission. The response time of the application is direct and faster which depends on the efficiency of implemented algorithm
5.	Availability	The application has to be available 24 x 7 for users without any interruption
6.	Scalability	The application can withstand the increase in the no. of users and has to be able to develop Higher versions

5. PROJECT DESIGN

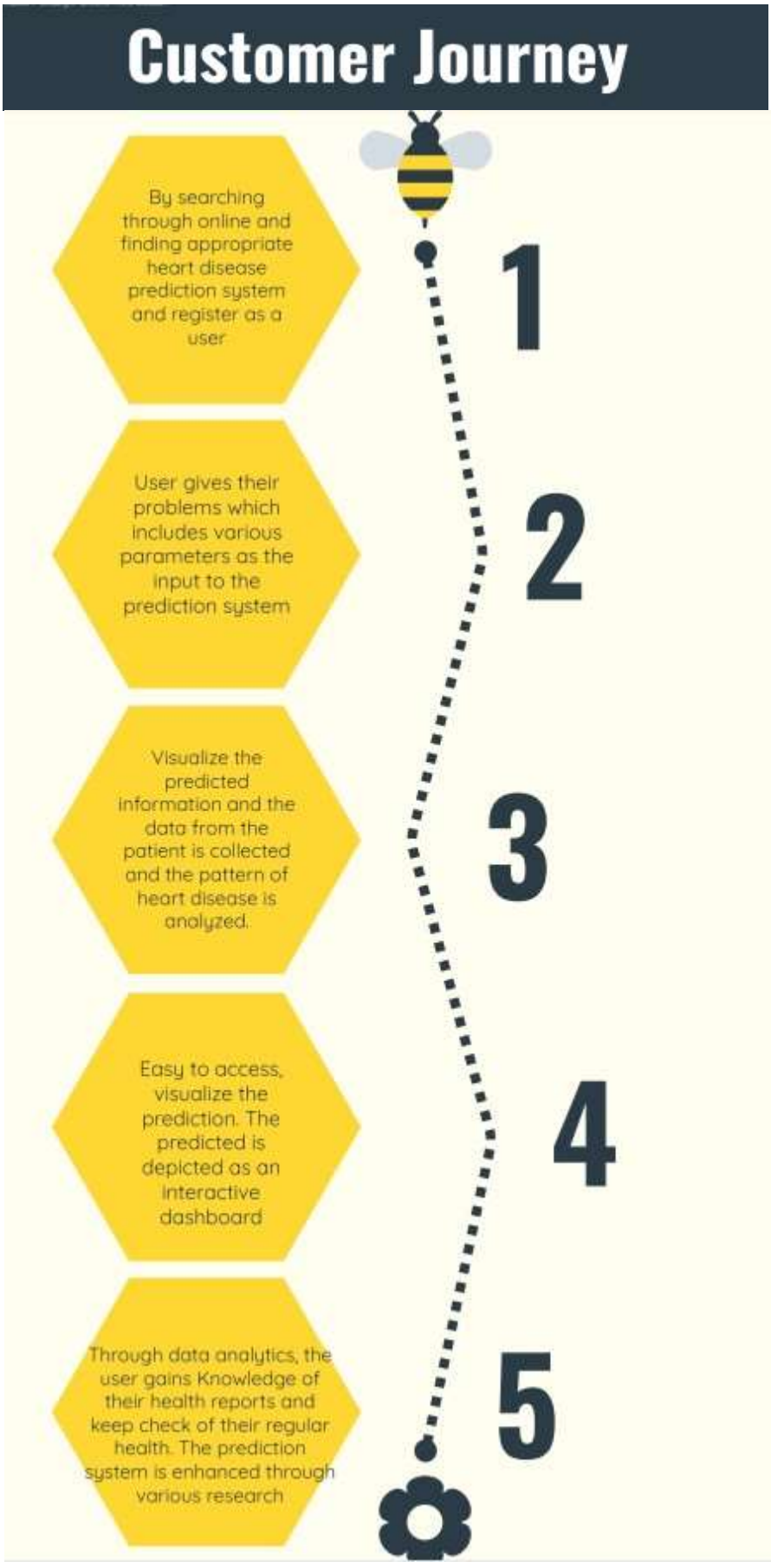
5.1 Data Flow Diagrams



5.2 Solution and Technical Architecture



5.3 User Stories



6. PROJECT PLANNING & SCHEDULING

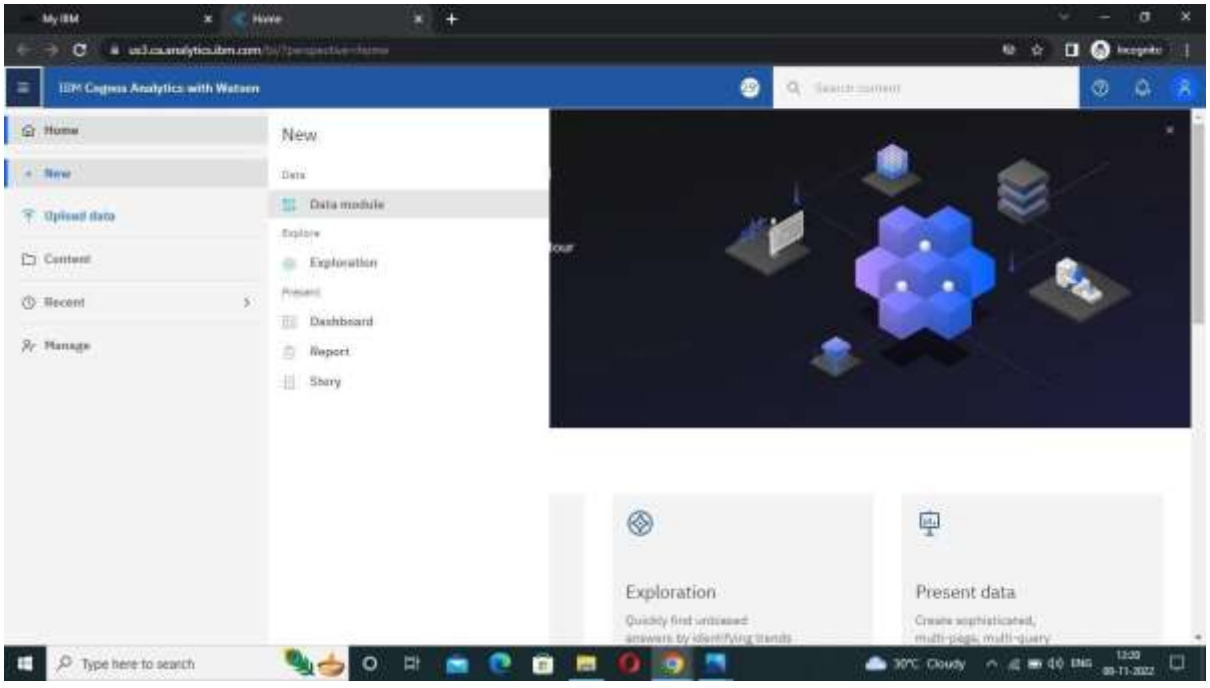
6.1 Sprint Planning & Estimation

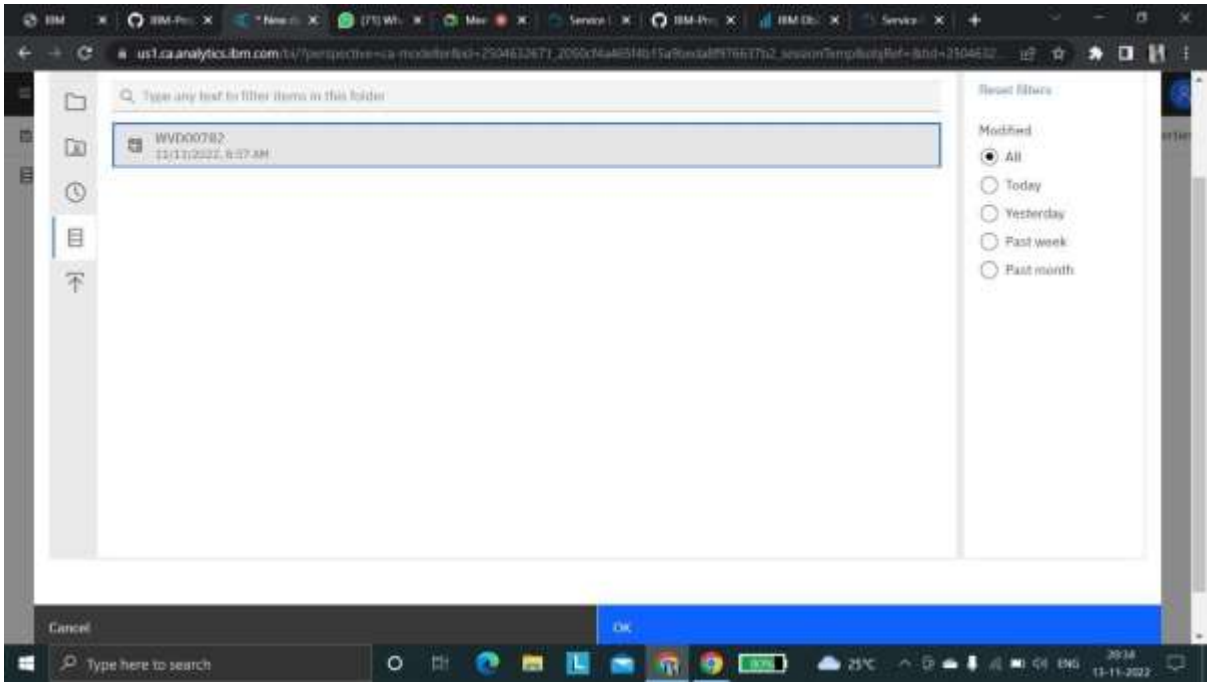
SPRINT 1:

CREATION OF DATA MODULE

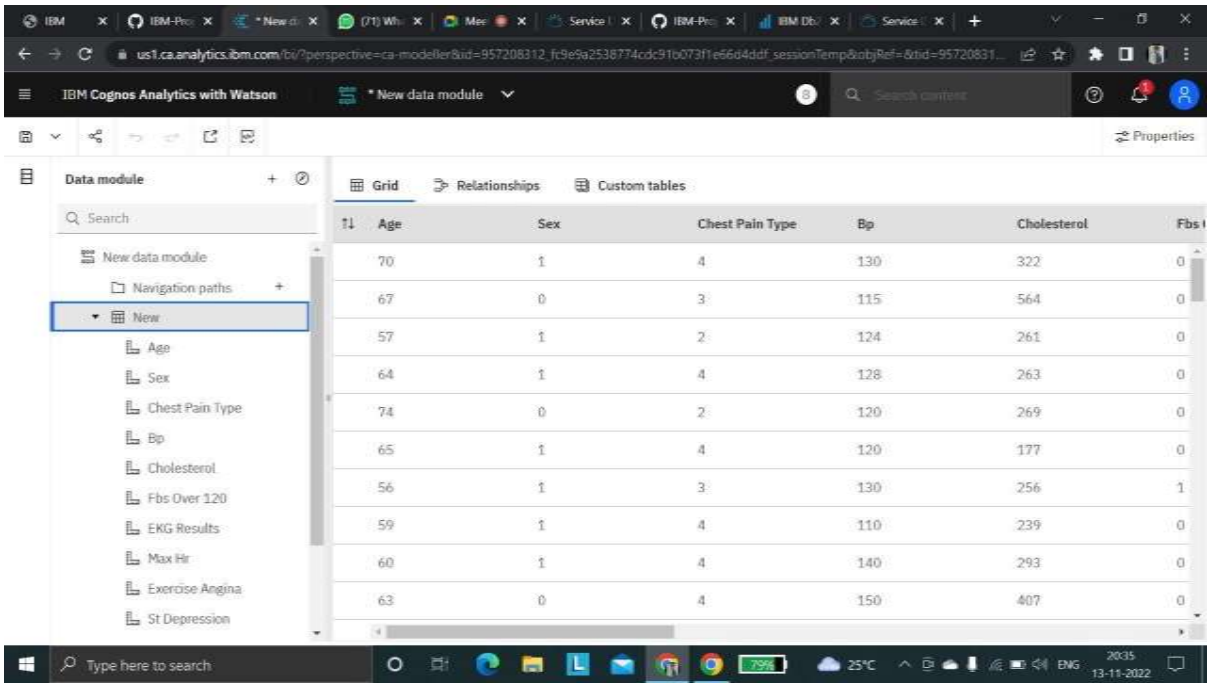
STEP 1: creation of Data Module.

STEP 2: CHOOSING OF Dataset in data server connection in cognos

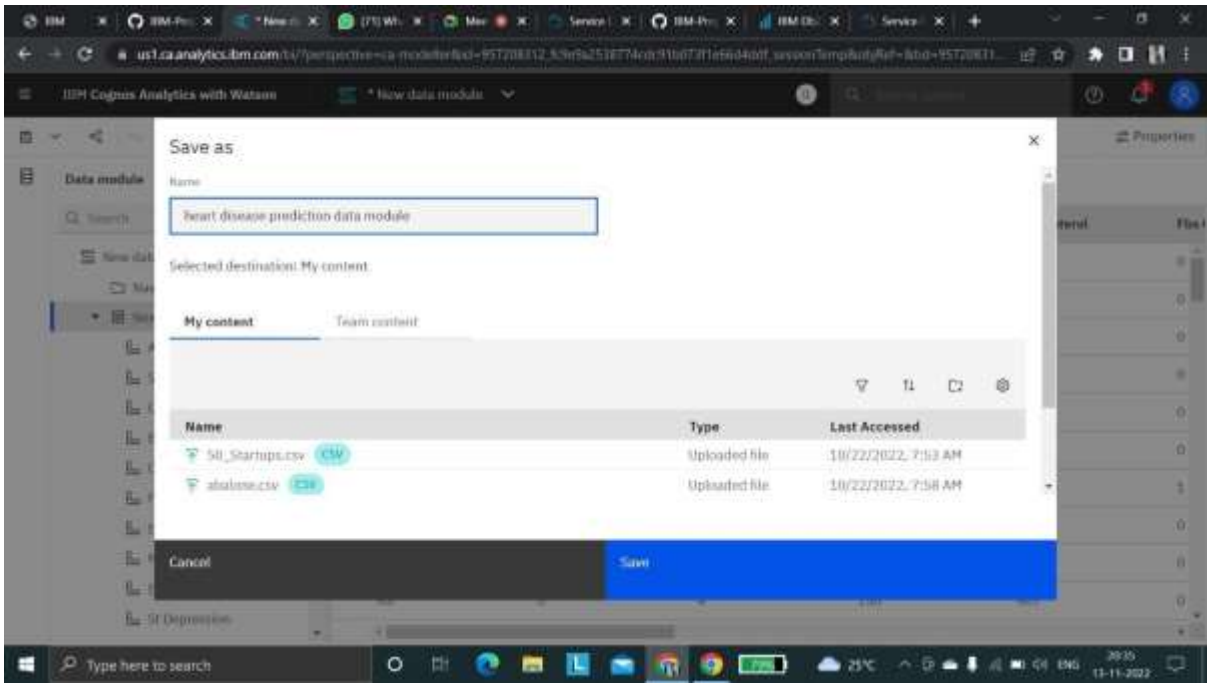




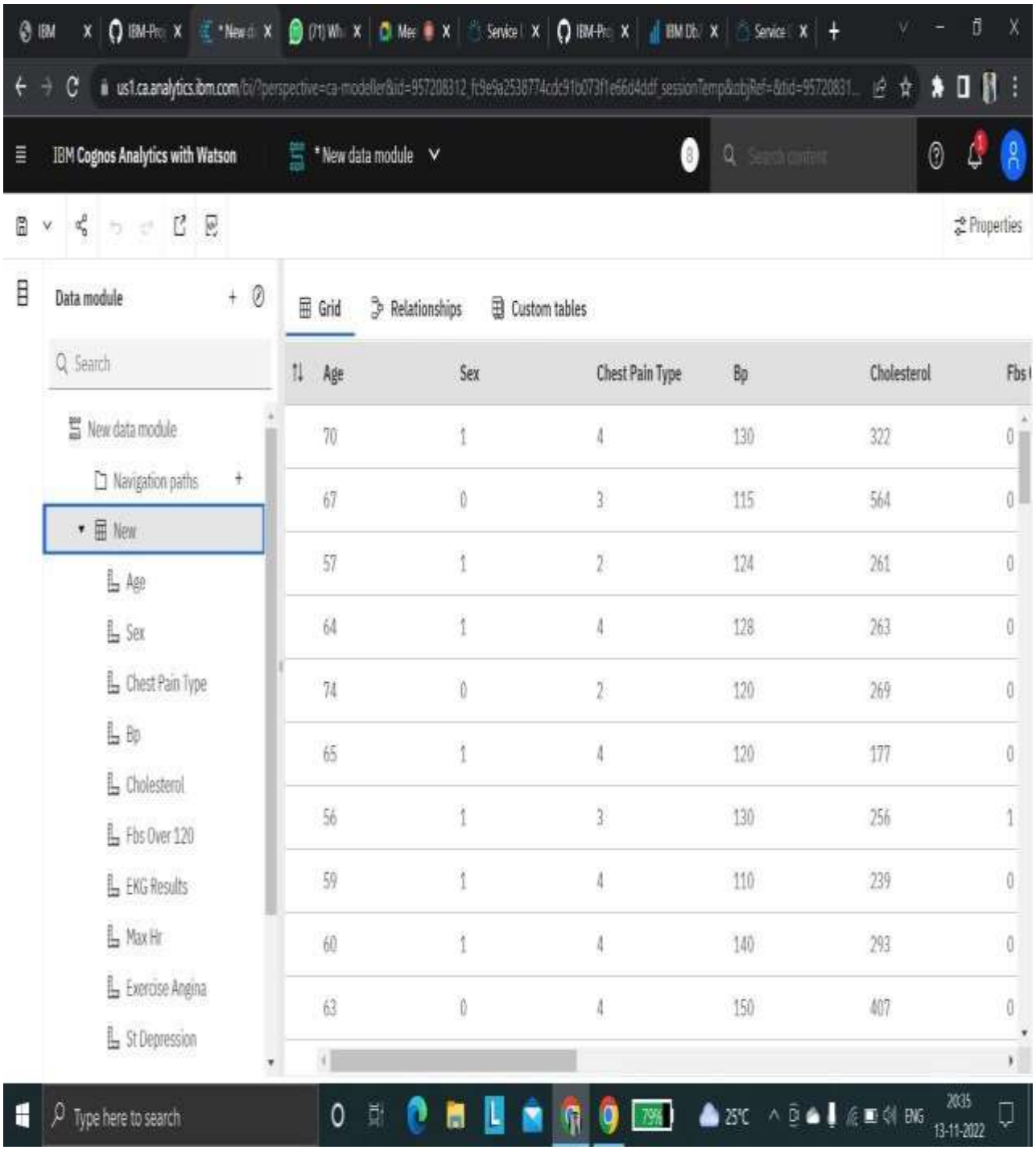
STEP 3: Data Module is created successfully



STEP 4:



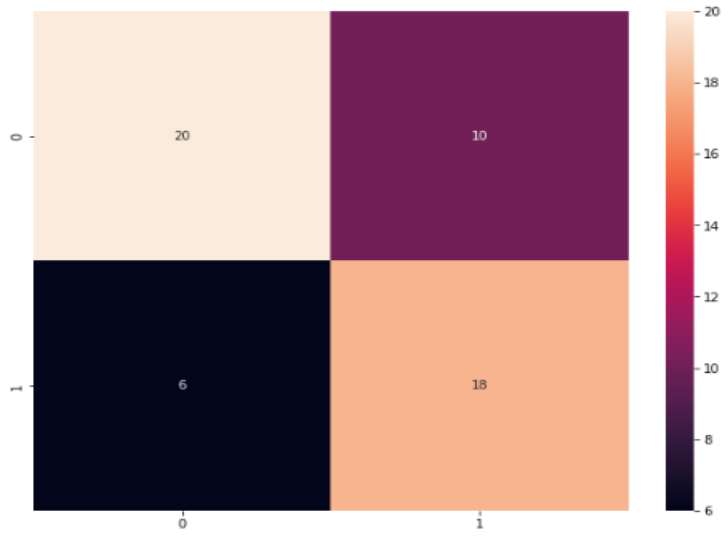
STEP 5: Representation of data module with data for heart disease prediction.



PREDICTION OF HEART DISEASE:

```
y_pred_dte = dt_classifier.predict(X_test)

plt.figure(figsize=(10, 8))
CM=confusion_matrix(Y_test,y_pred_dte)
sns.heatmap(CM, annot=True)

TN = CM[0][0]
FN = CM[1][0]
TP = CM[1][1]
FP = CM[0][1]
specificity = TN/(TN+FP)
loss_log = log_loss(Y_test, y_pred_dte)
acc= accuracy_score(Y_test, y_pred_dte)
roc=roc_auc_score(Y_test, y_pred_dte)
prec = precision_score(Y_test, y_pred_dte)
rec = recall_score(Y_test, y_pred_dte)
f1 = f1_score(Y_test, y_pred_dte)

mathew = matthews_corrcoef(Y_test, y_pred_dte)
model_results =pd.DataFrame([['Decision Tree',acc, prec,rec,specificity, f1,roc, loss_log,mathew]],
               columns = ['Model', 'Accuracy','Precision', 'Sensitivity','Specificity', 'F1 Score','ROC','Log_Loss','mathew_corr

model_results
```

	Model	Accuracy	Precision	Sensitivity	Specificity	F1 Score	ROC	Log_Loss	mathew_corrcoef
0	Decision Tree	0.703704	0.769231	0.666667	0.666667	0.714286	0.708333	19.188653	0.414371



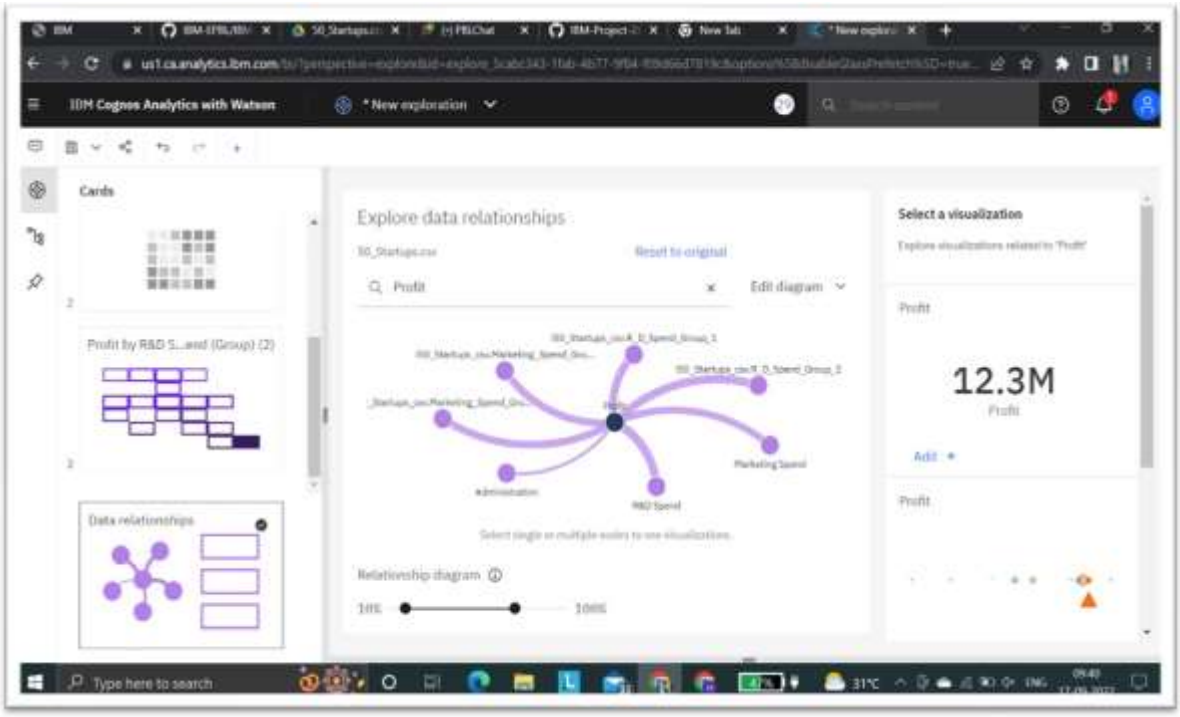

```
Y_pred_dt = np.around(Y_pred_dt)
print(metrics.classification_report(Y_test,Y_pred_dt))
```

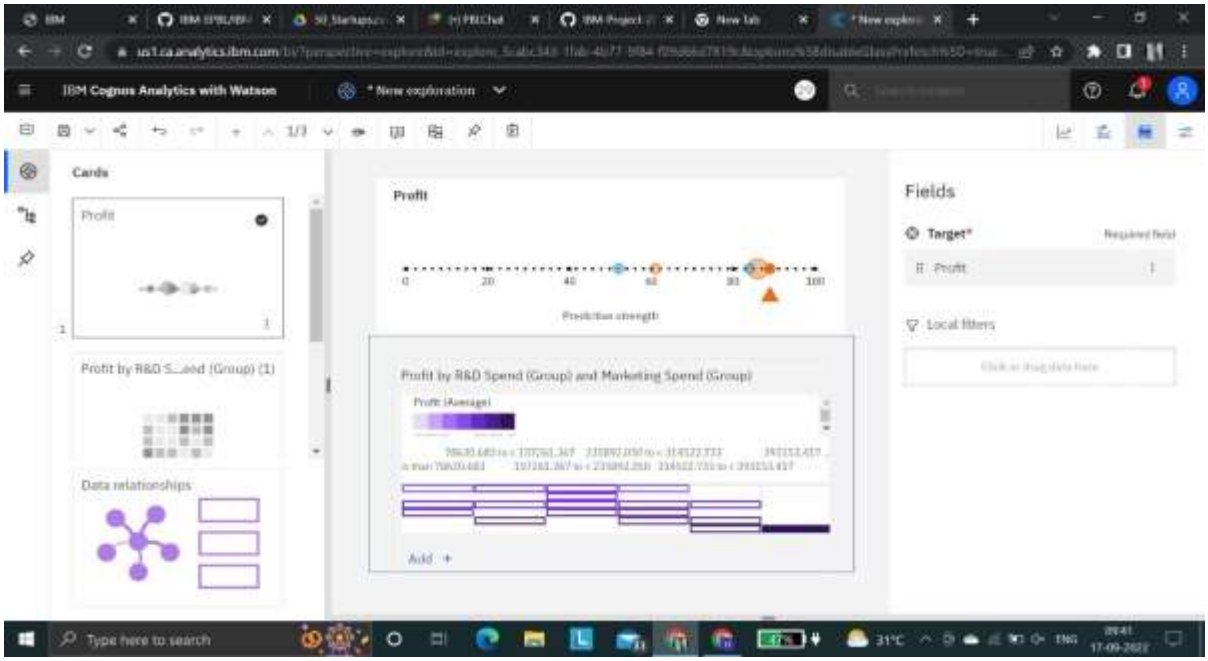
```
              precision    recall  f1-score   support

           1       0.77      0.67      0.71        30
           2       0.64      0.75      0.69        24

    accuracy                           0.70        54
   macro avg       0.71      0.71      0.70        54
weighted avg       0.71      0.70      0.70        54
```

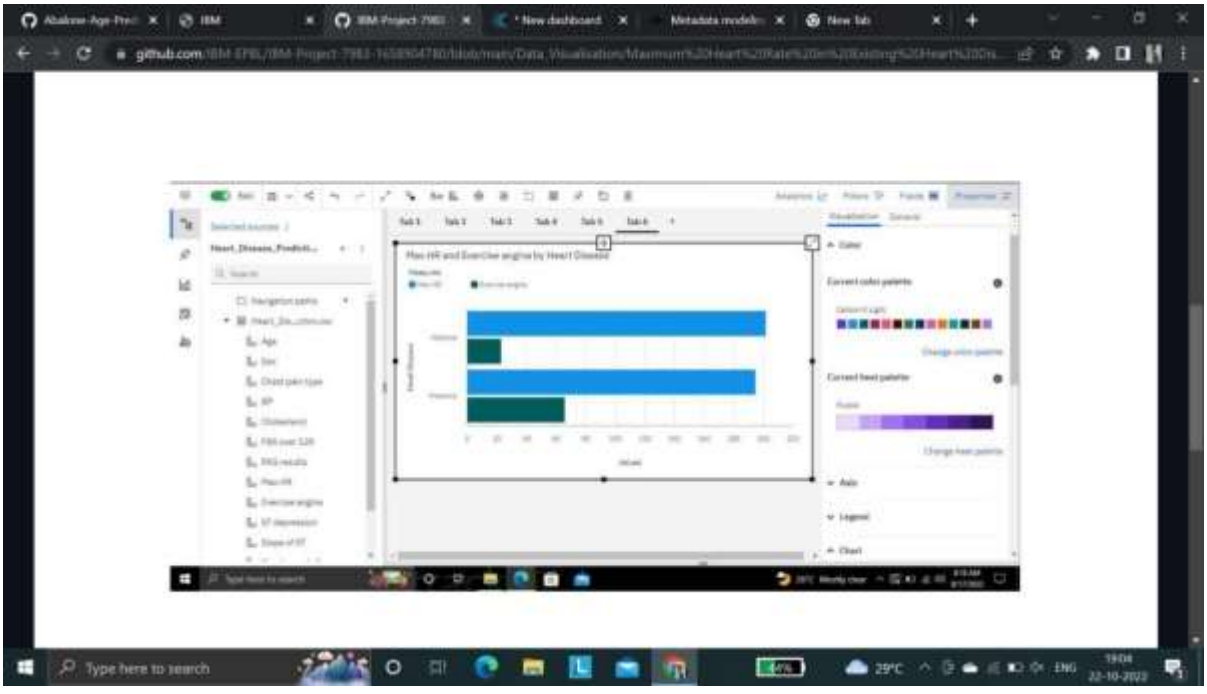
SPRINT 2:

WORKING WITH DATASET

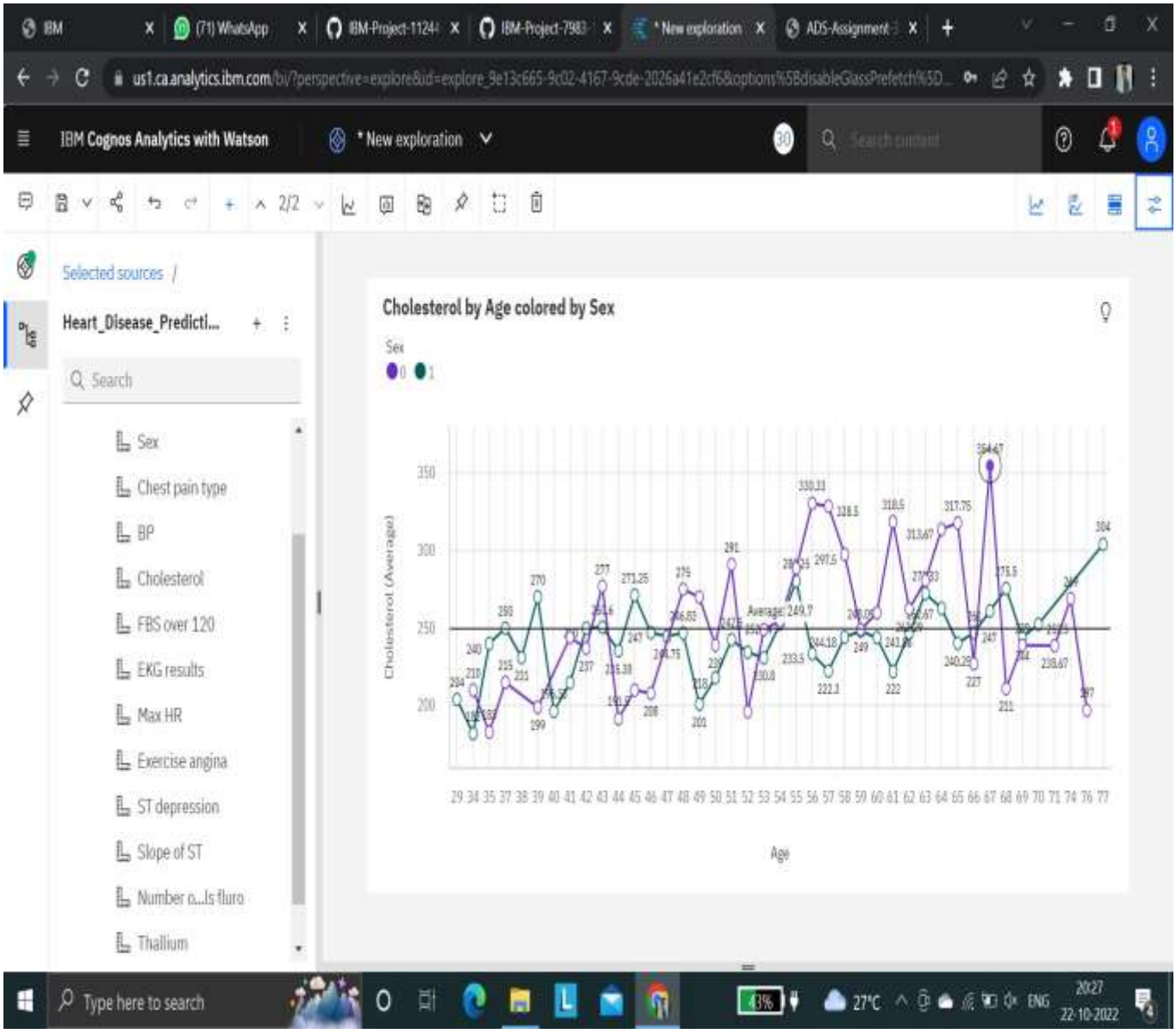
EXPLORING THE DATASET




IBM Cognos Analytics with Watson
* New exploration
Cards
Profit
Profit by R&D S...end (Group) (1)
Data relationships
Profit
Prediction strength
Profit by R&D Spend (Group) and Marketing Spend (Group)
Profit (Average)
Add
Fields
Target*
Profit
Local filters
Type here to search


Type here to search

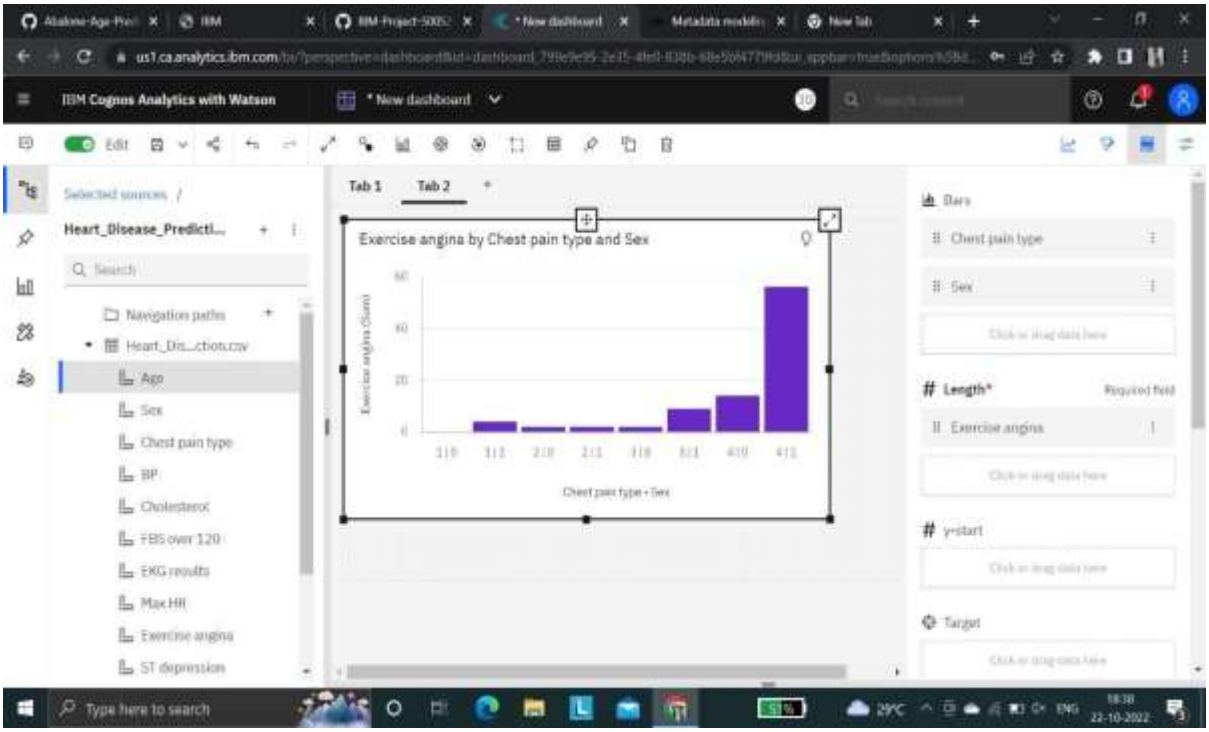
CHOLESTROL BY AFE COLORED BY SEX



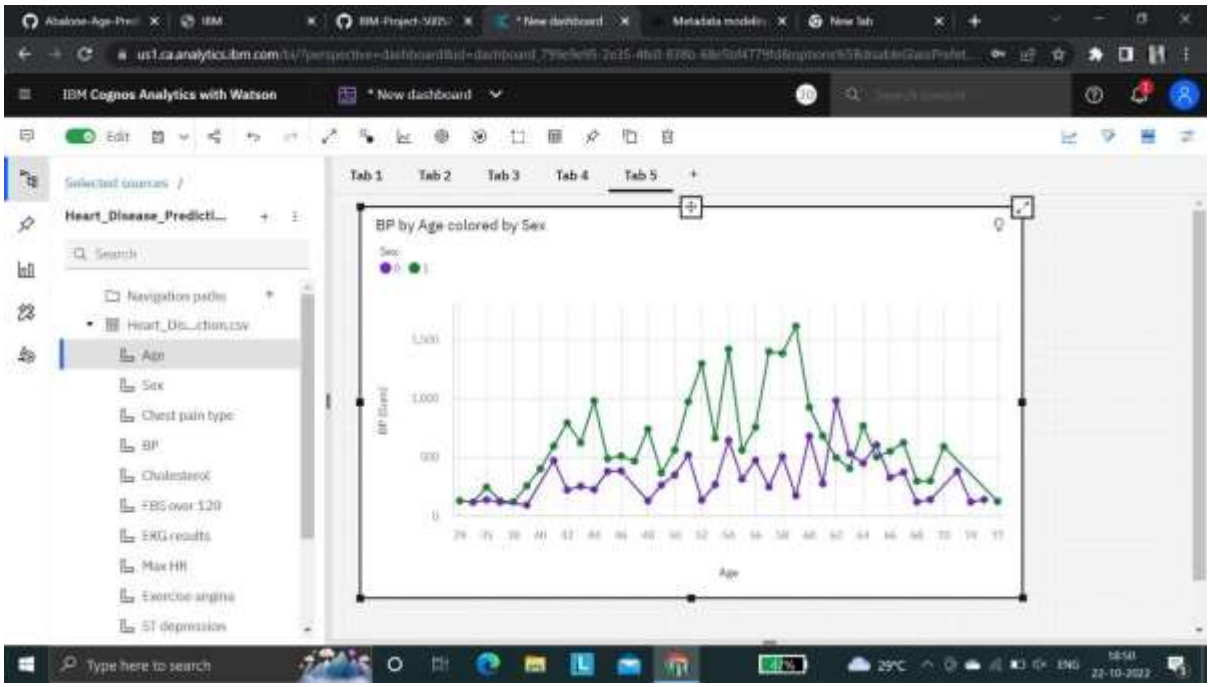
SPRINT 3

DASHBOARD VISUALIZATION

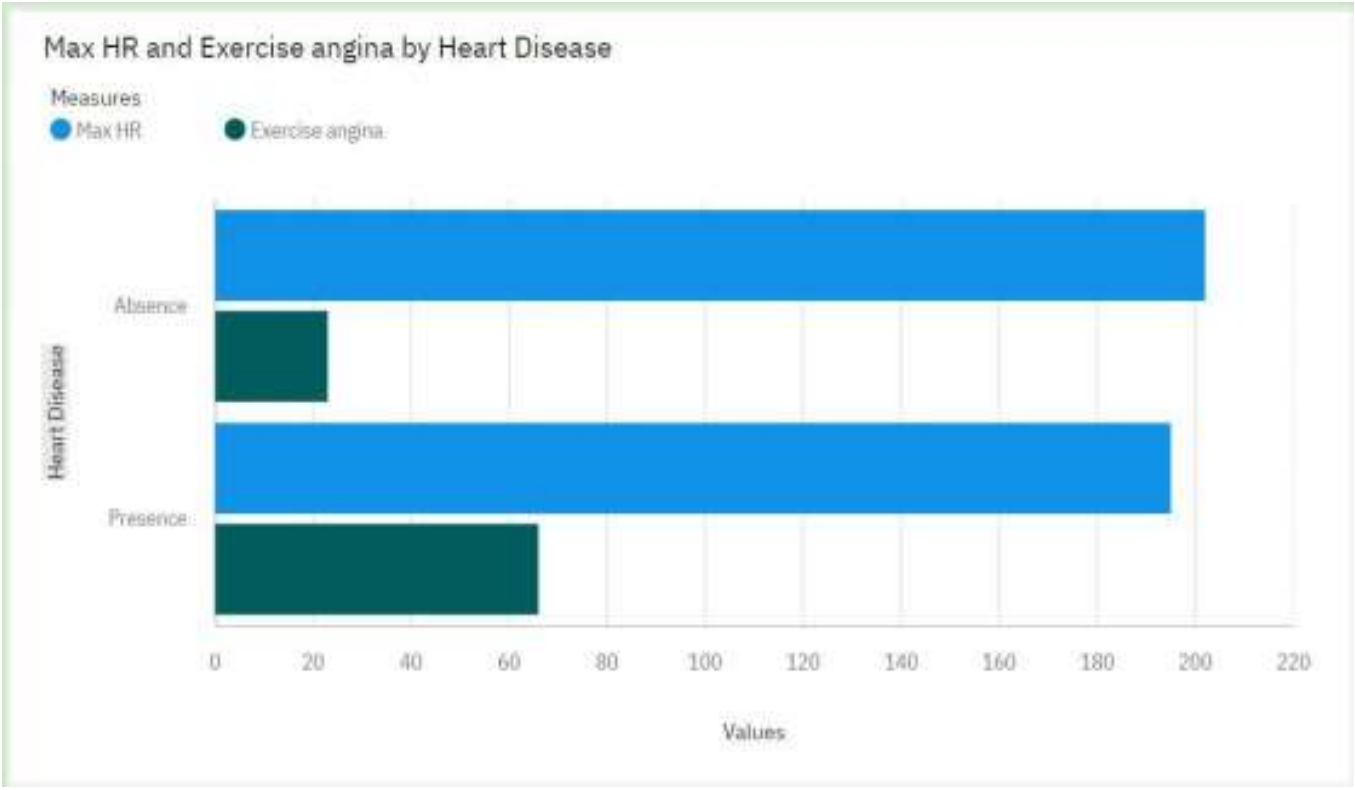
EXERCISE ANGINA BY CHEST PAIN TYPE AND SEX



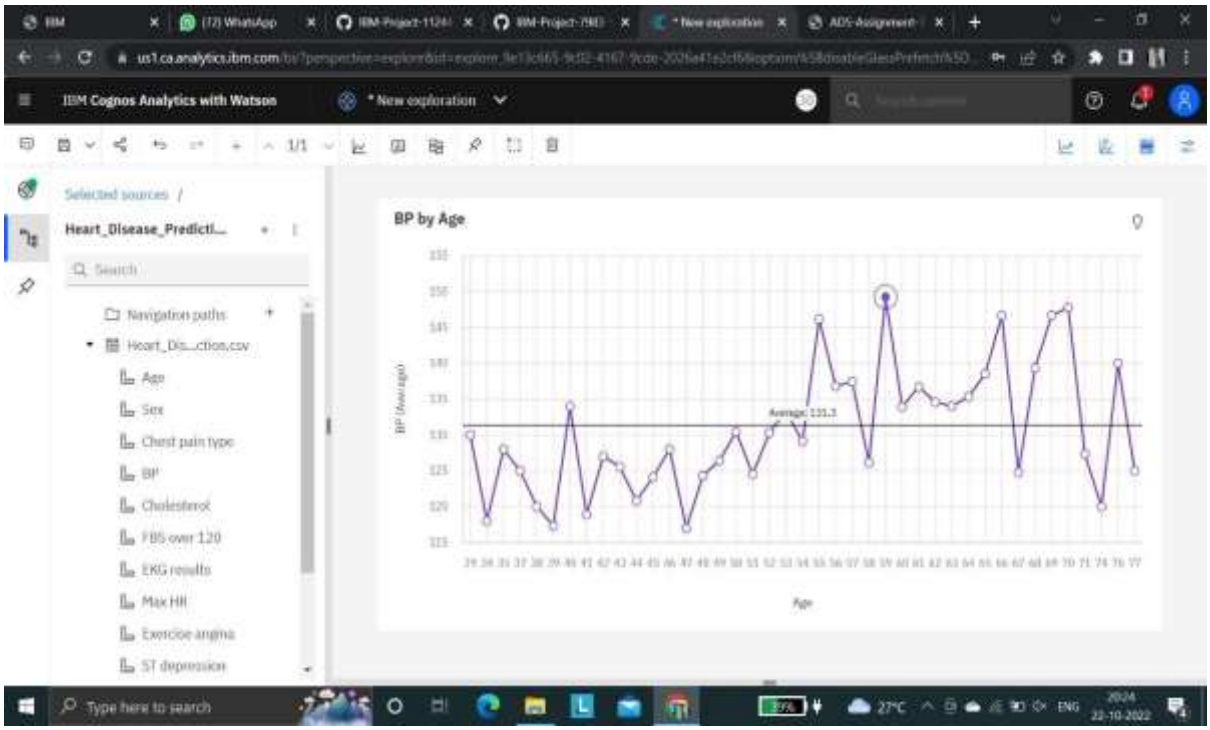
BP BY AGE COLORED BY SEX



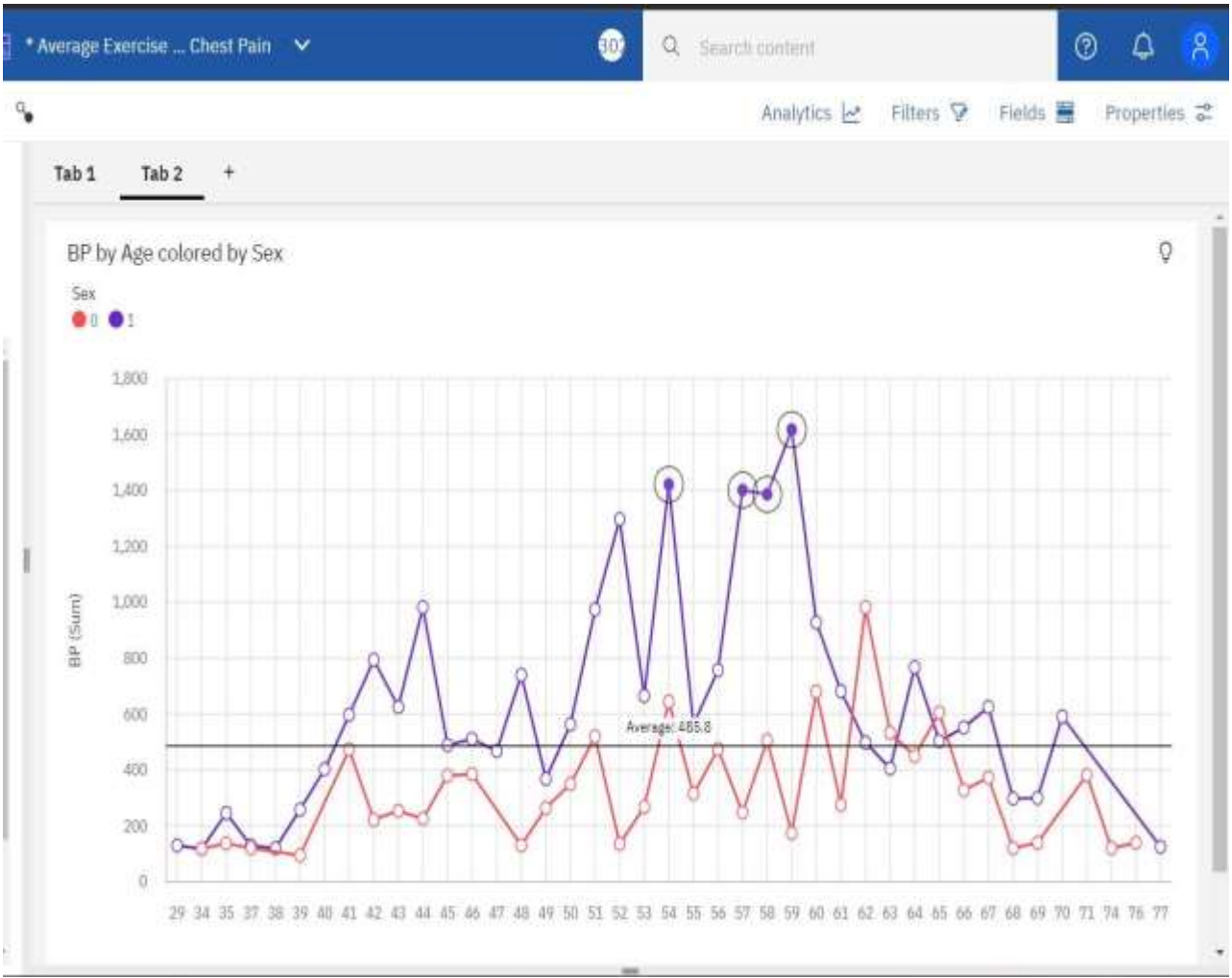
MAX HR AND EXERCISE ANGINA BY HEART DISEASE



BP BY AGE



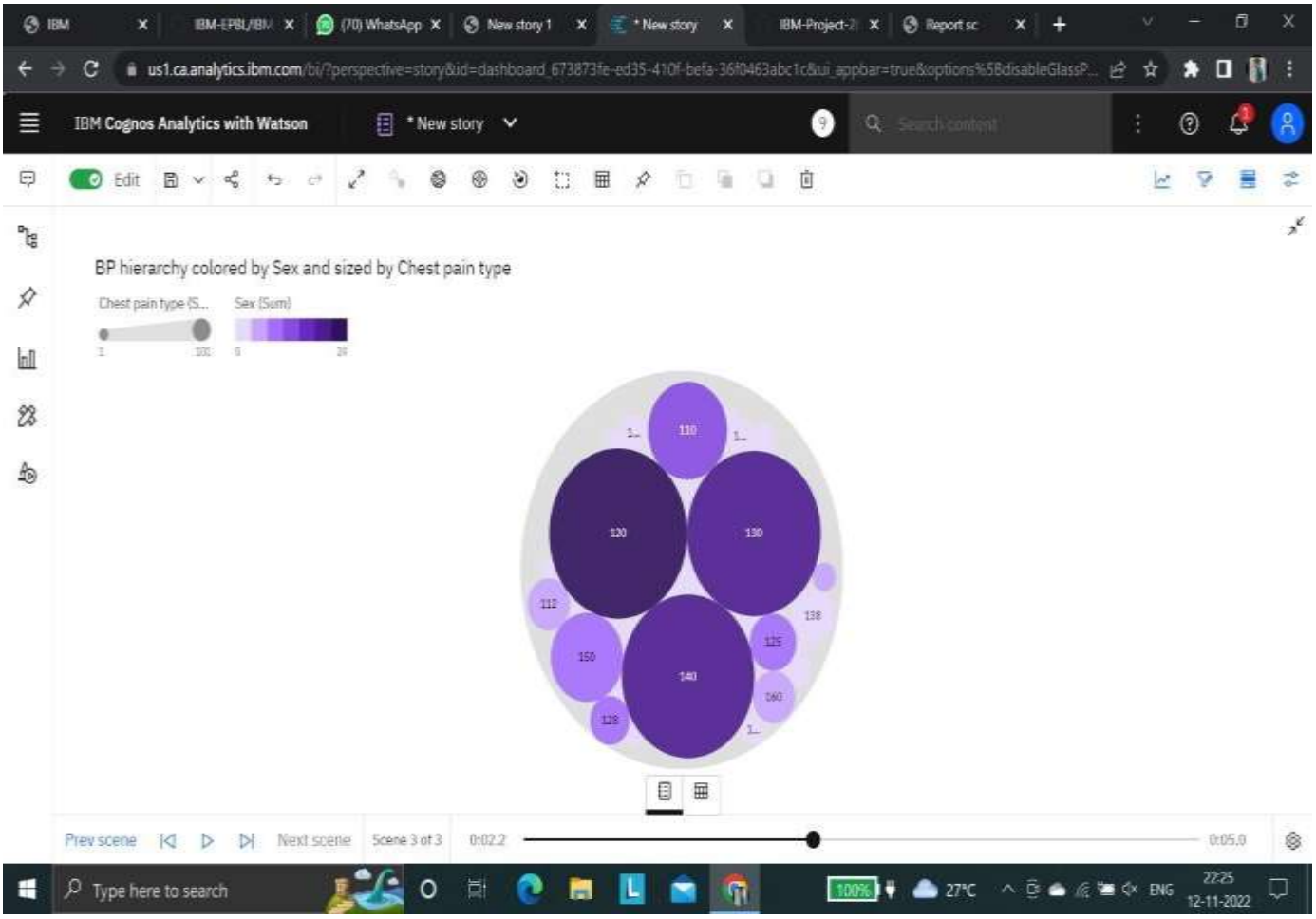
BP BY AGE COLORED BY SEX



STORY VISUALISATION

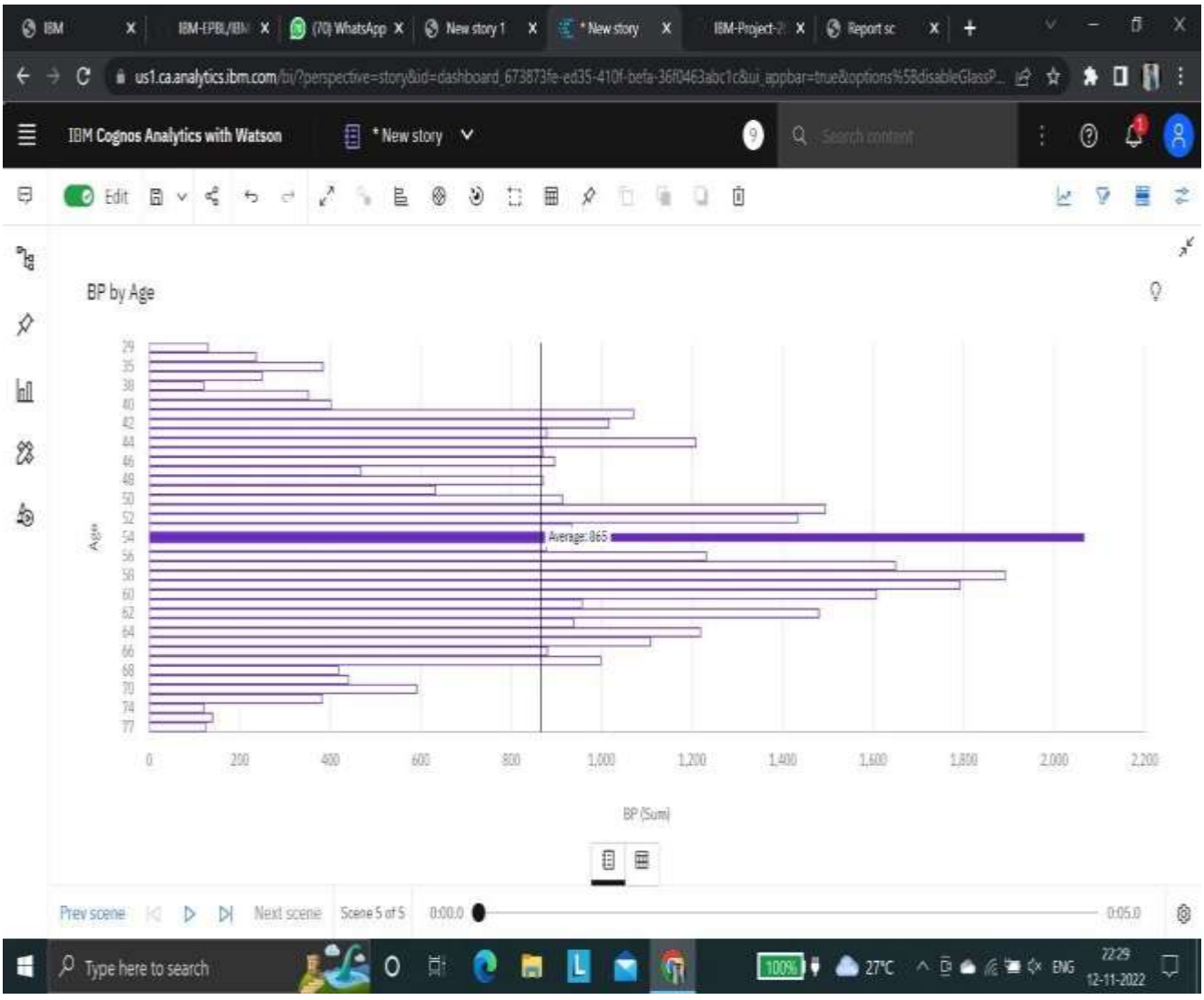
- Over all values of BP, the sum of Chest pain type is 857.
- For Chest pain type, the most significant values of BP are 120, 130, and 140, whose respective Chest pain type values add up to 296, or 34.5 % of the total.
- Chest pain type ranges from 1, when BP is 148, to 101, when BP is 120.

BP VS CHEST PAIN GENDER



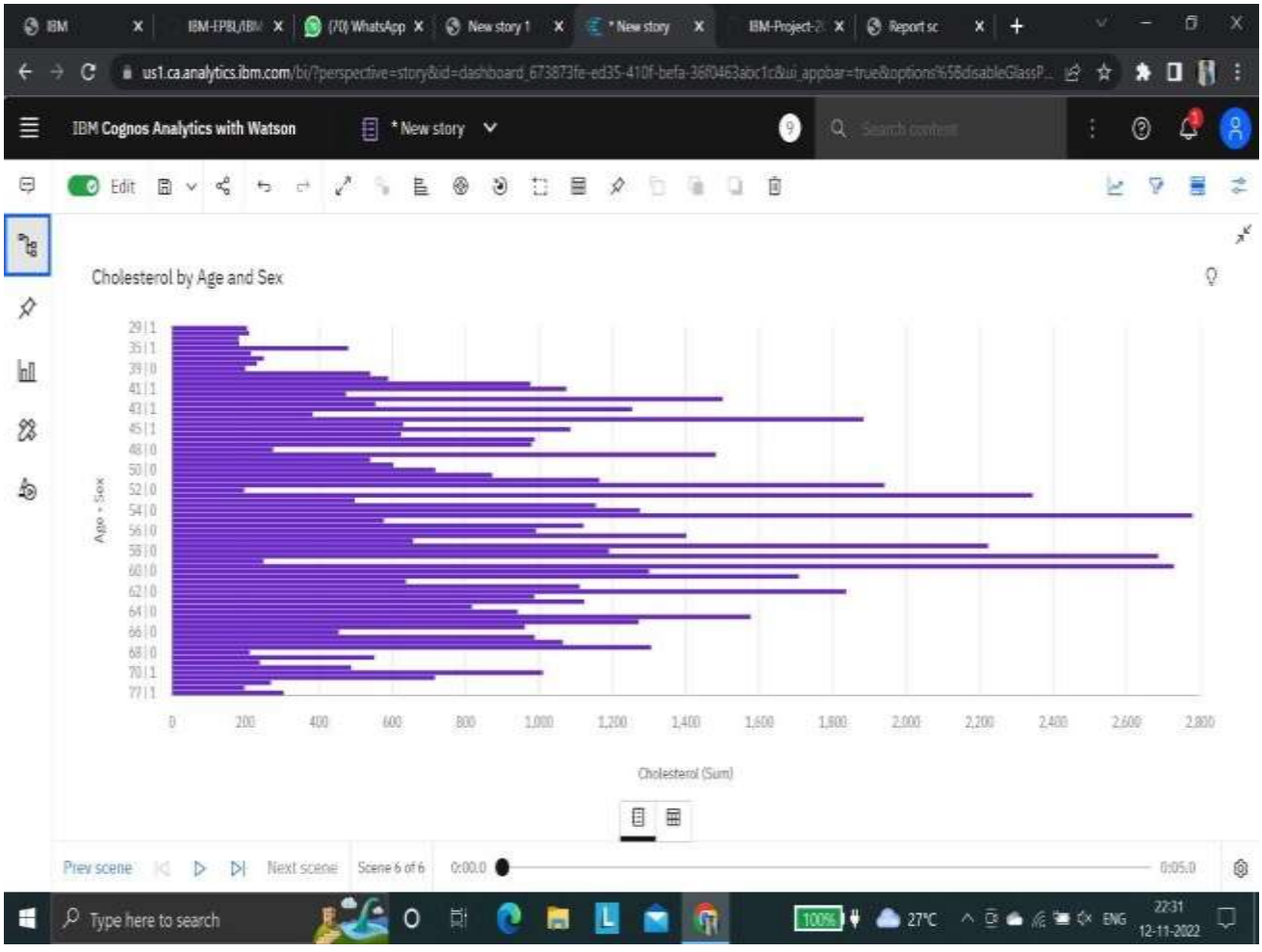
BP BY AGE

- The total number of results for Age is 270.



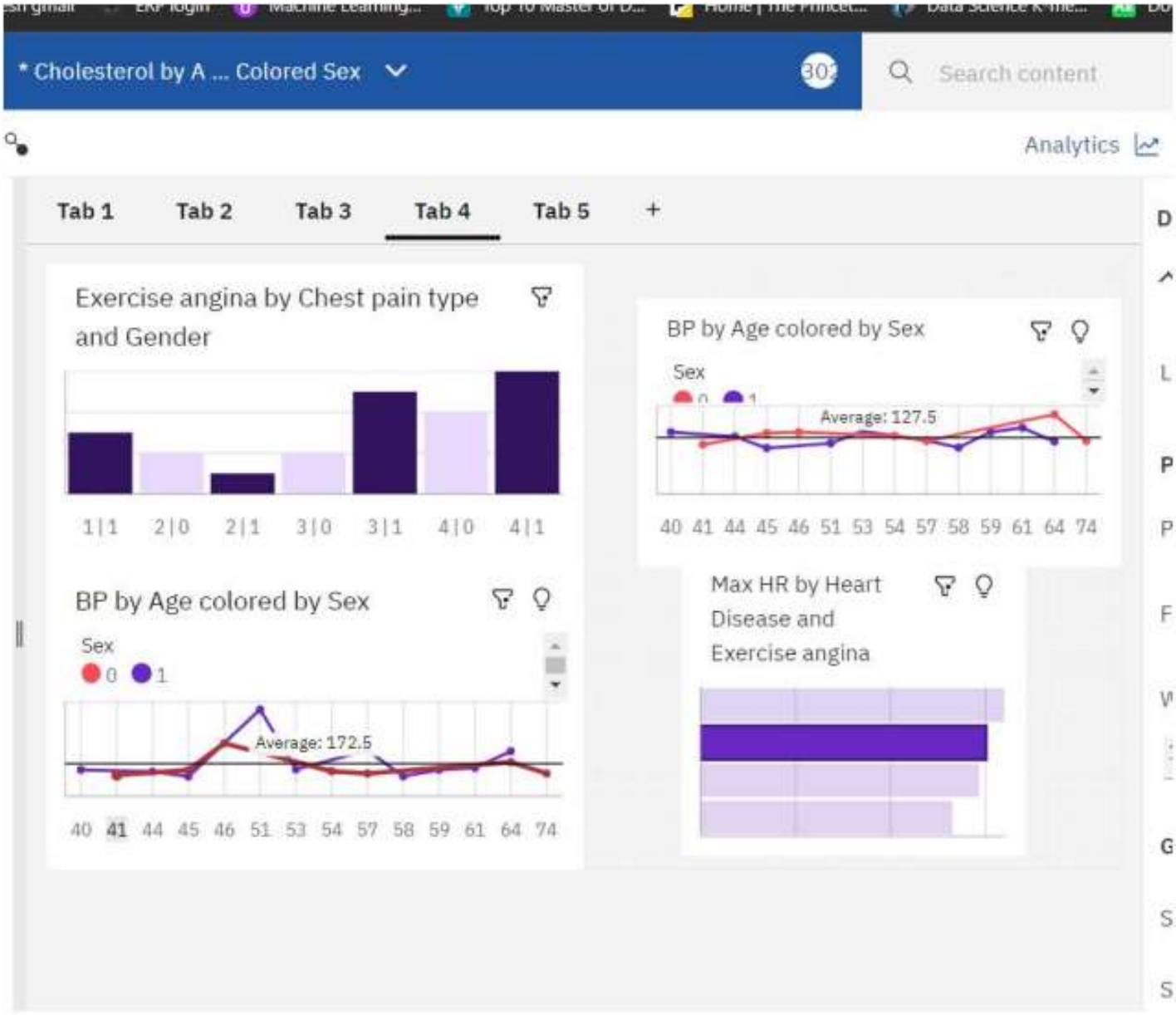
CHOLESTEROL BY AGE AND GENDER

- The Average value of cholesterol is 923.4.
- The value of cholesterol is unusually high when the combinations of Age and Sex are 54+1, 59+1 and 58+1.
- The summed values of Cholesterol range from 182 to nearly 3 thousand.
- Cholesterol is unusually high when Age is 54 and 58.

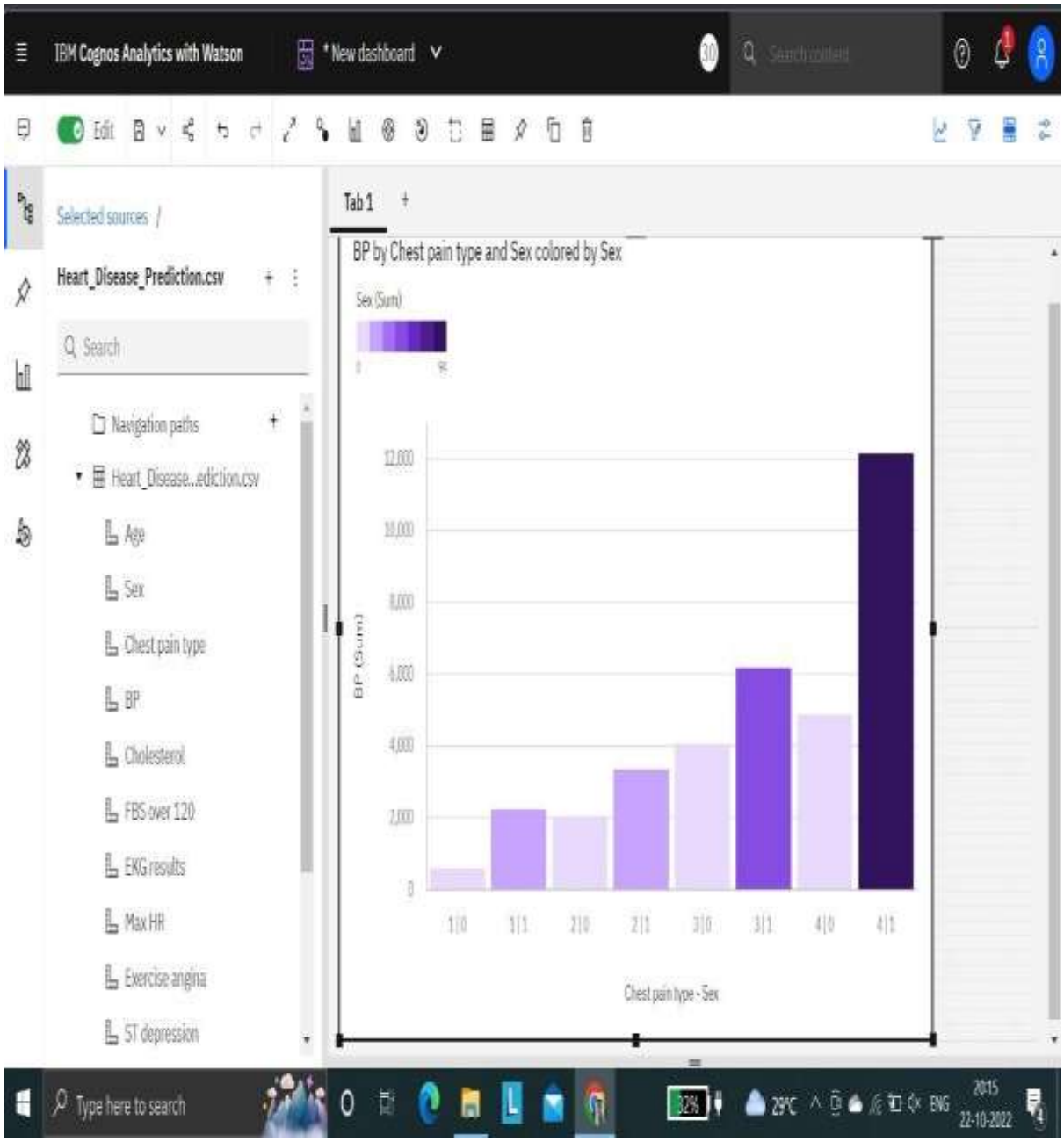


SPRINT 4:

SCREENSHOTS OF DASHBOARD



DASHBOARD VISUALIZATION



REPORT ON DASHBOARD

6.2 SPRINT DELIVERY SCHEDULE:

Sprint	Functional Requirement(Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Preprocessing andPrediction	USN-1	Upload the dataset and understand thedataset.	1	High	1
Sprint-1		USN-2	Perform data preprocessing and datacleaning	2	High	2
Sprint-1		USN-3	Visualize the dataset in various graphs foreasy analysis	2	Low	3
Sprint-1		USN-4	Train the model and predict the chance ofgetting heart disease	2	Medium	3
Sprint-2	Exploration	USN-5	Open a account in IBM Cognos tool	1	High	1
Sprint-2		USN-6	Explore the dataset	1	High	1
Sprint-2		USN-7	Explore BP vs chest pain and gender	2	High	1
Sprint-2		USN-8	Explore heart rate during chest pain	2	High	1
Sprint-2		USN-9	Explore BP by age	2	High	1
Sprint-2		USN-10	Explore Cholesterol by age and gender	1	High	1
Sprint-3	Visualization	USN-11	Visualize Average age for differenttypes of chest pain	1	High	1

Sprint-3		USN-12	Visualize Average exercise Angina DuringChest Pain	2	H
Sprint-3		USN - 13	Visualize BP variation with respect toage	2	
Sprint-3		USN-14	Average age for different types of chest pain in existing heart disease.	1	M
Sprint-3		USN-15	Visualize Serum Cholesterol level vs age	1	H
Sprint-3		USN-16	Maximum heart rate in existing heart disease by exercise angina	2	H
Sprint-4	Dashboard	USN-17	Dashboard showing different types of visuals	2	M

7.CODING AND SOLUTIONS:

1.DECISION TREE CLASSIFIER

TRAIN THE MODEL

```
dt_classifier = DecisionTreeClassifier(
max_depth=20,
min_samples_split=2,
min_samples_leaf=1,
min_weight_fraction_leaf=0.00001,
max_features='auto',
random_state=46)
dt_classifier.fit(X_train, Y_train)
Y_pred_dt=dt_classifier.predict(X_test)
score_dt = round(accuracy_score(Y_pred_dt,Y_test)*100,2)
score_dt
```

MODEL EVALUATION

```
y_pred_dte = dt_classifier.predict(X_test)

plt.figure(figsize=(10, 8))
CM=confusion_matrix(Y_test,y_pred_dte)
sns.heatmap(CM, annot=True)

TN = CM[0][0]
FN = CM[1][0]
TP = CM[1][1]
FP = CM[0][1]
specificity = TN/(TN+FP)
loss_log = log_loss(Y_test, y_pred_dte)
acc= accuracy_score(Y_test, y_pred_dte)
roc=roc_auc_score(Y_test, y_pred_dte)
prec = precision_score(Y_test, y_pred_dte)
rec = recall_score(Y_test, y_pred_dte)
f1 = f1_score(Y_test, y_pred_dte)
mathew = matthews_corrcoef(Y_test, y_pred_dte)
model_results =pd.DataFrame([['Decision Tree',acc, prec,rec,specificity, f1,roc, l
oss_log,mathew]],
   columns = ['Model', 'Accuracy','Precision', 'Sensitivity','Specificity', 'F1
Score','ROC','Log_Loss','mathew_corrcoef'])
```

model_results

2.NAIVE BAYES CLASSIFIER

TRAINING THE MODEL

```
nb_classifier = GaussianNB( var_smoothing=1e-50)
nb_classifier.fit(X_train,Y_train)
nb_classifier.predict(X_test)
Y_pred_nb = nb_classifier.predict(X_test)
score_nb = round(accuracy_score(Y_pred_nb,Y_test)*100,2)
score_nb
```

MODEL EVALUATION

```
_pred_dte = dt_classifier.predict(X_test)

plt.figure(figsize=(10, 8))
CM=confusion_matrix(Y_test,y_pred_dte)
sns.heatmap(CM, annot=True)

TN = CM[0][0]
FN = CM[1][0]
TP = CM[1][1]
FP = CM[0][1]
specificity = TN/(TN+FP)
loss_log = log_loss(Y_test, y_pred_dte)
acc= accuracy_score(Y_test, y_pred_dte)
roc=roc_auc_score(Y_test, y_pred_dte)
prec = precision_score(Y_test, y_pred_dte)
rec = recall_score(Y_test, y_pred_dte)
f1 = f1_score(Y_test, y_pred_dte)

mathew = matthews_corrcoef(Y_test, y_pred_dte)
model_results =pd.DataFrame([['Decision Tree',acc, prec,rec,specificity, f1,roc,
loss_log,mathew]],
          columns = ['Model', 'Accuracy','Precision', 'Sensitivity','Specificity', 'F1
Score','ROC','Log_Loss','mathew_corrcoef'])

model_results
```

FEATURE 2:

3. RANDOM FOREST CLASSIFIER

TRAIN THE MODEL

```
max_accuracy = 0

for x in range(500):
    rf_classifier = RandomForestClassifier(random_state=x)
    rf_classifier.fit(X_train,Y_train)
    Y_pred_rf = rf_classifier.predict(X_test)
    current_accuracy = round(accuracy_score(Y_pred_rf,Y_test)*100,2)
    if(current_accuracy>max_accuracy):
        max_accuracy = current_accuracy
        best_x = x

print(max_accuracy)
print(best_x)
```

PREDICTION:

```
rf_classifier = RandomForestClassifier(random_state=best_x)
rf_classifier.fit(X_train,Y_train)
Y_pred_rf = rf_classifier.predict(X_test)
Y_pred_rf.shape
```

4. K-NEAREST NEIGHBOR CLASSIFIER

TRAIN THE MODEL

```
knn_classifier= KNeighborsClassifier(n_neighbors=31,leaf_size=30)
knn_classifier.fit(X_train,Y_train)
Y_pred_knn = knn_classifier.predict(X_test)
score_knn = round(accuracy_score(Y_pred_knn,Y_test)*100,2)
score_knn
```

MODEL EVALUATION:

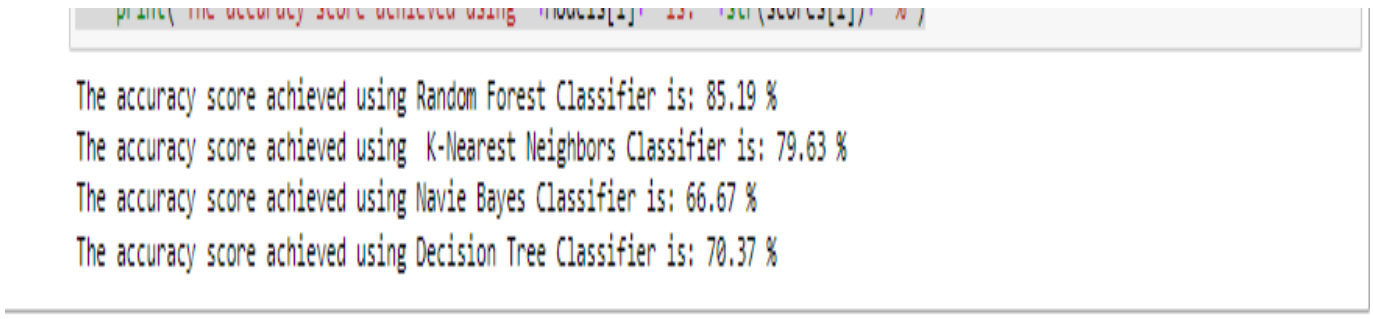
```
y_pred_knne = knn_classifier.predict(X_test)

plt.figure(figsize=(10, 8))
CM=confusion_matrix(Y_test,y_pred_knne)
sns.heatmap(CM, annot=True)
```

```
TN = CM[0][0]
FN = CM[1][0]
TP = CM[1][1]
FP = CM[0][1]
specificity = TN/(TN+FP)
loss_log = log_loss(Y_test, y_pred_knne)
acc= accuracy_score(Y_test, y_pred_knne)
roc=roc_auc_score(Y_test, y_pred_knne)
prec = precision_score(Y_test, y_pred_knne)
rec = recall_score(Y_test, y_pred_knne)
f1 = f1_score(Y_test, y_pred_knne)

mathew = matthews_corrcoef(Y_test, y_pred_knne)
model_results =pd.DataFrame([['K-Nearest Neighbors ',acc, prec,rec,specificity, f1,roc,
loss_log,mathew]],
            columns = ['Model', 'Accuracy','Precision', 'Sensitivity','Specificity', 'F1
Score','ROC','Log_Loss','mathew_corrcoef'])

model_results
```

FINAL:

```
scores = [score_rf,score_knn,score_nb,score_dt]
Models = ["Random Forest Classifier"," K-Nearest Neighbors Classifier","Navie Bayes
Classifier","Decision Tree Classifier"]
for i in range(len(Models)):
   print("The accuracy score achieved using "+Models[i]+" is: "+str(scores[i])+" %")
```

OUTPUT:



8. PERFORMANCE METRICS:

Project Development Phase

Model Performance Test

Date	10 November 2022
Team ID	PNT2022TMID53509
Project Name	Project – Visualizing and Predicting Heart Disease with an Interactive Dashboard
Maximum Marks	10 Marks

Model Performance Testing:

Project team shall fill the following information in model performance testing template.

S.No.	Parameter	Screenshot / Values
1	Dashboard design	No of Visualizations / Graphs - 16
2	Data Responsiveness	The Data is responsiveness since it can be visualizedin different devices and on any screen resolution.

3	Amount Data to Rendered	Data to be loaded consisted of 270 rows and 14 columns, a total of 3780 values(considering each to be integer 1.89kB)
4	Utilization of Data Filters	Data Filtering was used to find out the total number of records in the dataset with different type of fields and filter out one of the types of many.
5	Effective User Story	No of Scene Added – 3No of Report Added - 4
6	Descriptive Reports	No of Visualizations / Graphs – 16

9. ADVANTAGES AND DISADVANTAGES:

ADVANTAGES:

- This is one of the fastest ways to determine if a person is likely to suffer from a heartdisease 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

10. CONCLUSION:

Complications of heart disease include heart attack and stroke. You can reduce the risk of complications with early diagnosis and treatment. So the suggestion that we get from the website might help save patients and take proactive measures for treatment or converting to a healthy lifestyle . It is always important to get treated in the early stages of heart disease. Regularly checking the risk of heart disease by providing the inputs can help people to cure at early stages.

11. FUTURE SCOPE:

We will look into methods to prevent heart diseases altogether other than just predicting it in early stages. To use this website we need to take a lot of tests beforehand. So it would be better if we require less attributes and still give an effective result. Integrating it with Smart watches can help people get results instantaneously , we can link cardiologist network to our website and people can get consultation and immediate medical help incase of high risk alert .

13.APPENDIX:

GITHUB LINK:
https://github.com/IBM-EPBL/IBM-Project-20282-1659716490

PROJECT DEMO LINK:
https://drive.google.com/file/d/1P_ZvXFw3pzMbHpyGkMq7Z7W2qPhOZKMS/view?usp=share_link