WIH3001
DATA SCIENCE
PROJECT

EARLY-STAGE DIABETES RISK PREDICTION

Nurul Asyiqin binti Mohd Harmizi
17065343/2

INTRODUCTION

DISCUSSION POINTS

- Diabetes Definition
- Types of Diabetes
- Symptoms
- Risk factors
- Complications of diabetes

DIABETES DEFINITION

1. According to Centres for Disease Control and Prevention (CDC), diabetes is a chronic health condition that affects how the human body turns food into energy.
2. It occurs either when the pancreas does not produce enough insulin OR when the body cannot effectively use the insulin it produces

Reference: https://www.who.int/news-room/fact-sheets/detail/diabetes

Types of diabetes

Type 1 Diabetes

An illness characterized by the body's inability to produce insulin due to the autoimmune destruction of the beta cells in the pancreas.

Type 2 Diabetes

A well-known diabetes disease where it occurs when the human body becomes resistant to insulin, and the sugar builds up in your blood.

Prediabetes

A serious health condition where blood sugar levels are higher than normal, but not high enough yet to be diagnosed as Type 2 diabetes

Gestational Diabetes

A condition where the blood glucose values above normal but below those diagnostics of diabetes, which occurs during pregnancy.

SYMPTOMS

- Frequent urination
- Fatigue
- Unexplained weight loss
- Slow healing cuts and wounds
- Increased hunger and thirst
- Tingling or numbness in the hands or feet

RISK FACTORS

- Weight
- Inactivity
- Family history
- Age
- High Blood Pressure
- Abnormal cholesterol levels

COMPLICATIONS

- Cardiovascular disease
- Nerve damage (neuropathy)
- Eye damage (retinopathy)
- Depression

PROBLEM UNDERSTANDING

DISCUSSION POINTS

- Problem Statement
- Questions being Addressed
- Objectives

PROBLEM STATEMENT

1. According to World Health Organization (WHO), it is reported that between year 2000 and 2016, there was a 5% increase in premature mortality from diabetes globally (Diabetes, 2021).

2. There is an increased number of patients whom suffered Type 2 diabetes among children, adolescents and young adults (Nadeau et al., 2016).

3. In year 2020, the emergence of COVID-19 coronavirus disease has affected more people with pre-existing medical conditions, including diabetes (COVID-19 and Diabetes, 2021).

QUESTIONS BEING ADDRESSED

1. What **risk factors** might occur to diabetic patients that can be predicted the most?
2. What are people's **lifestyle habits** that the user mostly likely for them to get diabetes?
3. How to **differentiate between prediabetes with Type 2 diabetes** in terms of symptoms or any risk factor?
4. Does having **health complications might increase the chance of getting diabetes,** or vice versa?
5. What can we develop from our findings and analyses in this project to raise awareness about diabetes prevention to the user?

Objectives

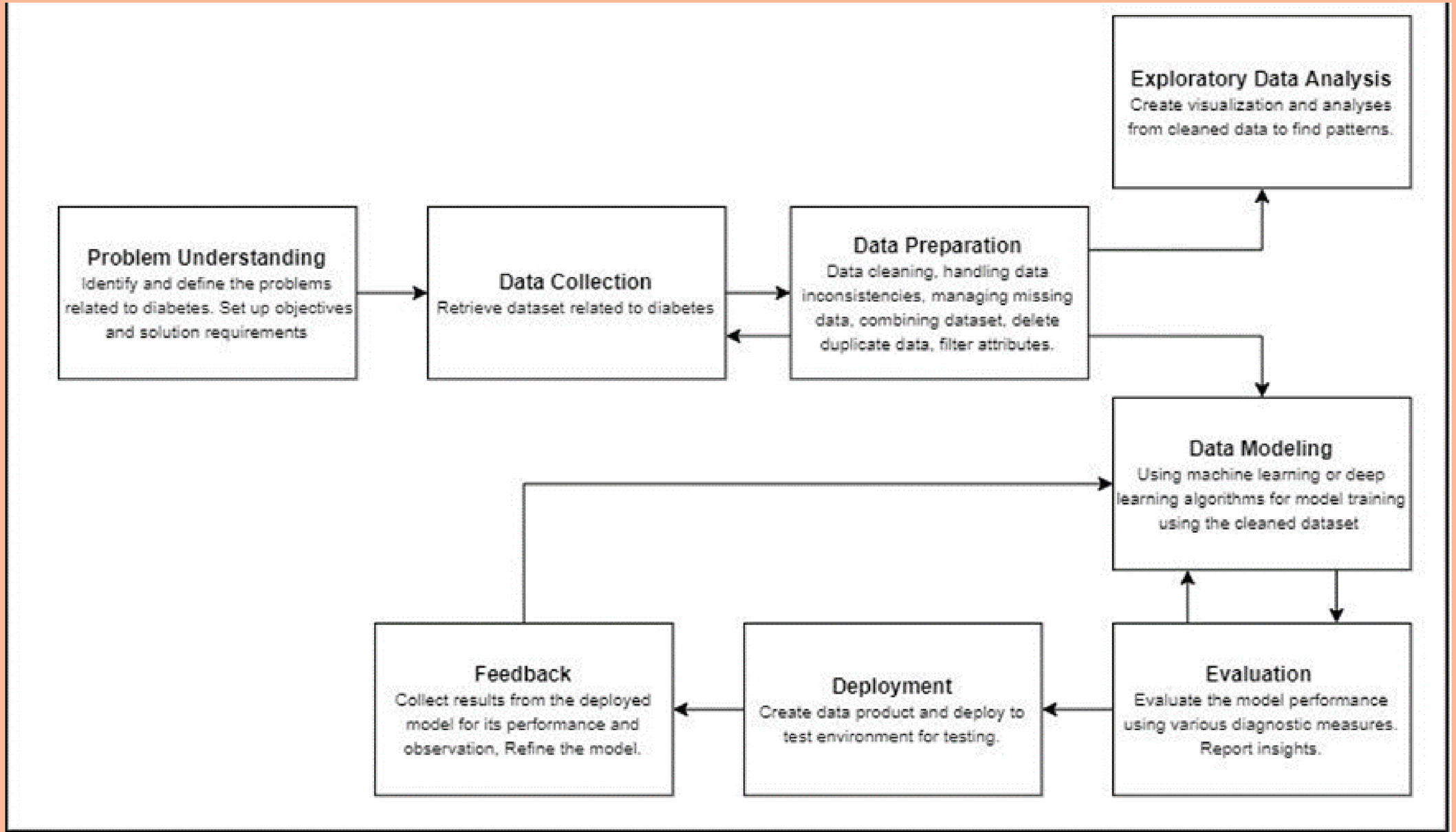
1. To indicate which **symptoms or risk factors** are high to occur among diabetic patients.
2. To **differentiate the healthy people and diabetic patients** based on the symptoms of the people who have diabetes or any health conditions that can be linked to it.
3. To analyze **which health complications are highly associated with diabetes**.
4. To **raise awareness** by creating a suitable application for the user to check their health status.

DATA SCIENCE METHODOLOGY

DISCUSSION POINTS

- Data Collection
- Data Preparation
- Exploratory Data Analysis
- Predictive Modeling
- Model Evaluation
- Testing and Deployment

Data Science Methodology Pipeline



DATA COLLECTION

- Collected from Behavioral Risk Factor Surveillance System (BRFSS) Centers for Disease Control and Prevention (CDC).
    - **2019 BRFSS Survey Data and Documentation**.
- Has 418,268 entries (rows) and 343 total columns
- We need to convert the data file from SAS Transport format to CSV file format.

Centers for Disease Control and Prevention

Behavioral Risk Factor Surveillance System

2019 Data

The 2019 BRFSS data continue to reflect the changes initially made in 2011 for weighting methodology (raking) and adding cell-phone-only respondents. The aggregate BRFSS combined landline and cell phone data set is built from the landline and cell phone data submitted for 2019 and includes data for 49 states, the District of Columbia, Guam, and Puerto Rico.

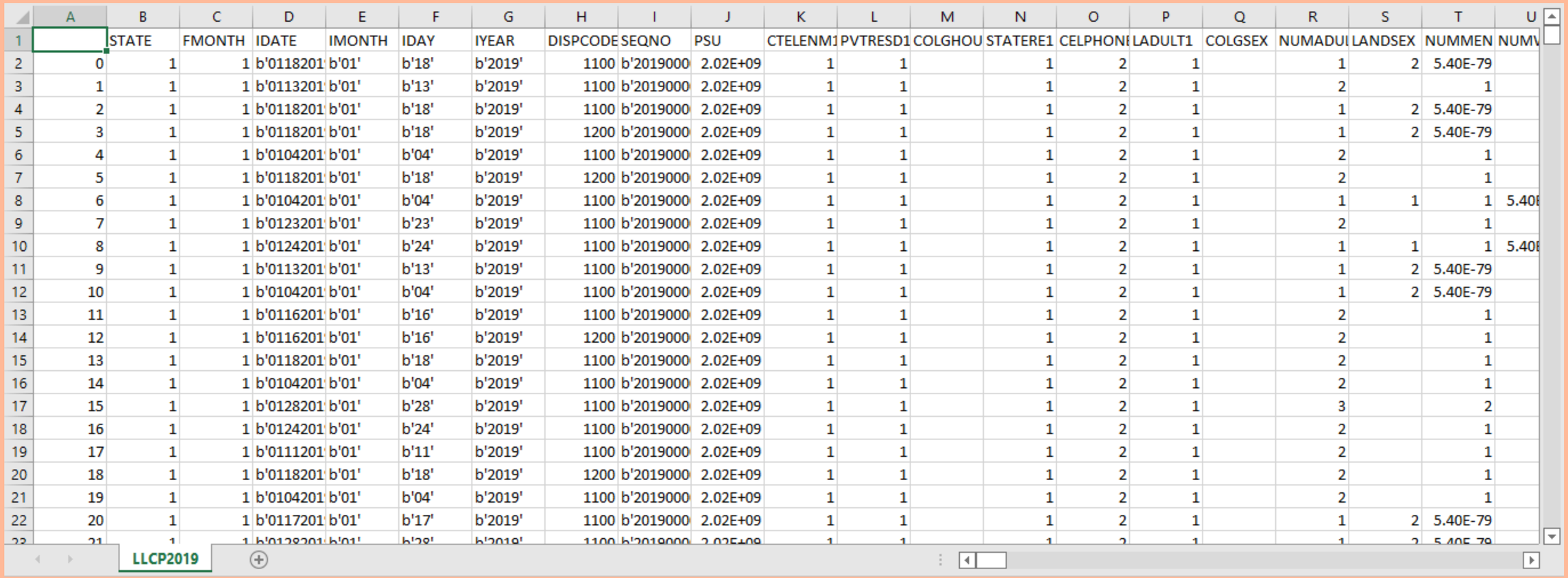
2019 Survey Data Information

2019 BRFSS Overview CDC [PDF – 298 KB]
Provides information on the background, design, data collection and

2019 BRFSS Data (ASCII) [ZIP – 66.2 MB]
August, 2020
This file for the combined landline and cell phone data set is in ASCII format. It has a fixed record length of 2158 positions.

2019 BRFSS Data (SAS Transport Format) [ZIP – 101 MB]
August, 2020
This file for the combined landline and cell phone data set was exported from SAS V9.3 in the XPT transport format. This file

After file conversion:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1		STATE	FMONTH	IDATE	IMONTH	IDAY	IYEAR	DISPCODE	SEQNO	PSU	CTELENM1	PVTRESD1	COLGHOU	STATERE1	CELPHON	LADULT1	COLGSEX	NUMADUI	LANDSEX	NUMMEN	NUMV
2	0	1	1	b'0118201	b'01'	b'18'	b'2019'	1100	b'2019000	2.02E+09	1	1		1	2	1		1	2	5.40E-79	
3	1	1	1	b'0113201	b'01'	b'13'	b'2019'	1100	b'2019000	2.02E+09	1	1		1	2	1		2		1	
4	2	1	1	b'0118201	b'01'	b'18'	b'2019'	1100	b'2019000	2.02E+09	1	1		1	2	1		1	2	5.40E-79	
5	3	1	1	b'0118201	b'01'	b'18'	b'2019'	1200	b'2019000	2.02E+09	1	1		1	2	1		1	2	5.40E-79	
6	4	1	1	b'0104201	b'01'	b'04'	b'2019'	1100	b'2019000	2.02E+09	1	1		1	2	1		2		1	
7	5	1	1	b'0118201	b'01'	b'18'	b'2019'	1200	b'2019000	2.02E+09	1	1		1	2	1		2		1	
8	6	1	1	b'0104201	b'01'	b'04'	b'2019'	1100	b'2019000	2.02E+09	1	1		1	2	1		1	1	1	5.40
9	7	1	1	b'0123201	b'01'	b'23'	b'2019'	1100	b'2019000	2.02E+09	1	1		1	2	1		2		1	
10	8	1	1	b'0124201	b'01'	b'24'	b'2019'	1100	b'2019000	2.02E+09	1	1		1	2	1		1	1	1	5.40
11	9	1	1	b'0113201	b'01'	b'13'	b'2019'	1100	b'2019000	2.02E+09	1	1		1	2	1		1	2	5.40E-79	
12	10	1	1	b'0104201	b'01'	b'04'	b'2019'	1100	b'2019000	2.02E+09	1	1		1	2	1		1	2	5.40E-79	
13	11	1	1	b'0116201	b'01'	b'16'	b'2019'	1100	b'2019000	2.02E+09	1	1		1	2	1		2		1	
14	12	1	1	b'0116201	b'01'	b'16'	b'2019'	1200	b'2019000	2.02E+09	1	1		1	2	1		2		1	
15	13	1	1	b'0118201	b'01'	b'18'	b'2019'	1100	b'2019000	2.02E+09	1	1		1	2	1		2		1	
16	14	1	1	b'0104201	b'01'	b'04'	b'2019'	1100	b'2019000	2.02E+09	1	1		1	2	1		2		1	
17	15	1	1	b'0128201	b'01'	b'28'	b'2019'	1100	b'2019000	2.02E+09	1	1		1	2	1		3		2	
18	16	1	1	b'0124201	b'01'	b'24'	b'2019'	1100	b'2019000	2.02E+09	1	1		1	2	1		2		1	
19	17	1	1	b'0111201	b'01'	b'11'	b'2019'	1100	b'2019000	2.02E+09	1	1		1	2	1		2		1	
20	18	1	1	b'0118201	b'01'	b'18'	b'2019'	1200	b'2019000	2.02E+09	1	1		1	2	1		2		1	
21	19	1	1	b'0104201	b'01'	b'04'	b'2019'	1100	b'2019000	2.02E+09	1	1		1	2	1		2		1	
22	20	1	1	b'0117201	b'01'	b'17'	b'2019'	1100	b'2019000	2.02E+09	1	1		1	2	1		1	2	5.40E-79	
23	21	1	1	b'0128201	b'01'	b'28'	b'2019'	1100	b'2019000	2.02E+09	1	1		1	2	1		1	2	5.40E-79	

LLCP2019

DATA PREPARATION

1. Filter the required columns from 2019 BRFSS Dataset.
2. Remove any rows that contains null values (in categorical columns)
3. Fill in missing rows that with column mean values (in numeric columns)
4. Rename all column headers
5. Replace necessary values for each categorical column.*
6. Filter the values that contains "Don't Know" and "Refused".*
7. Format the numerical values in column "BMI", "Weight" and "Height"

(*) - need to refer to 2019 BRFSS Codebook CDC

Filtered columns

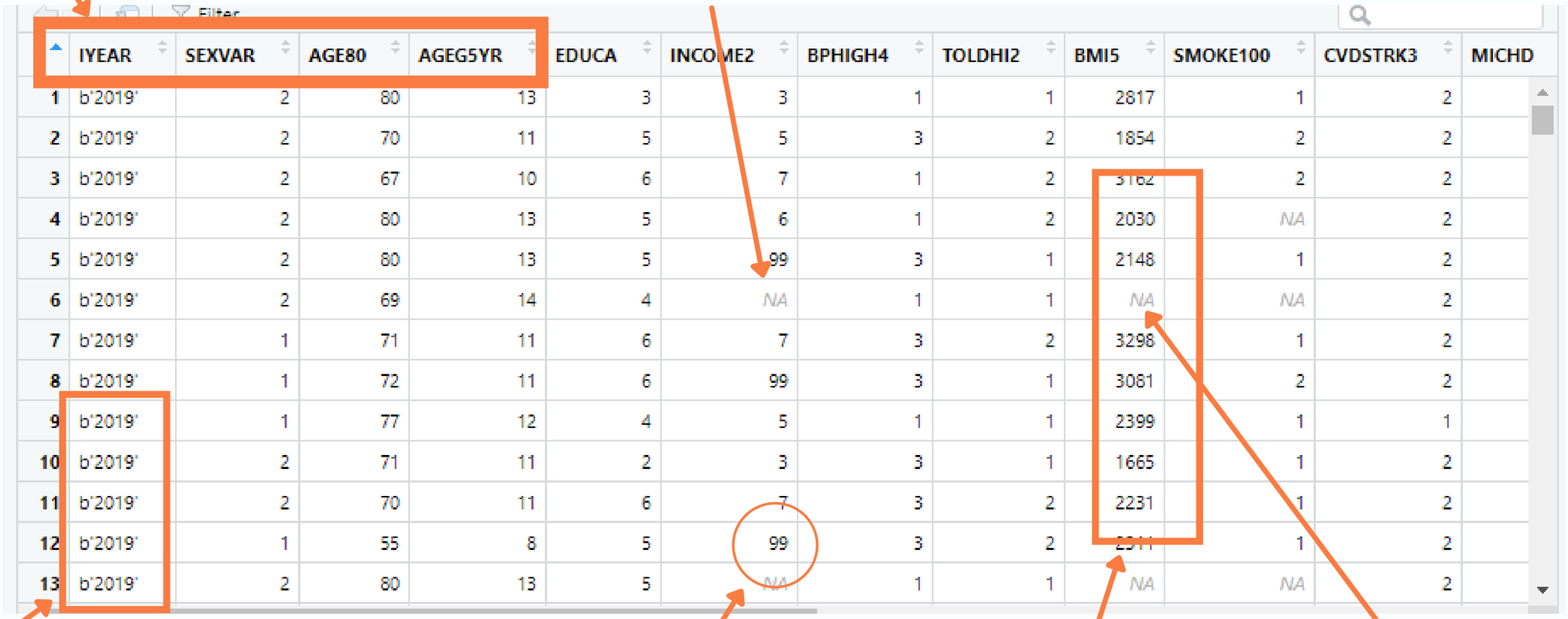
- **SEXVAR** (Gender)
- **_AGEG5YR** (Age Category)
- **EDUCA** (Education Level)
- **INCOME2** (Monthly Income)
- **DIABETE4** (Diabetes confirmation)
- **BPHIGH4** (High Blood Pressure confirmation)
- **TOLDHI2** (High cholesterol confirmation)
- **_BMI5** (User's Body Mass Index (BMI))
- **WTKG3** (User's weight in kilogram)
- **SMOKE100** (Smoking)
- **CVDSTRK3** (Cardiovascular Disease or Stroke)

- **_MICHD** (Heart disease)
- **_TOTINDA** (Physical Activity)
- **_RFDRHV6** (Alcohol Consumption)
- **HLTHPLN1** (Any Healthcare Nearby)
- **MEDCOST** (Accessibility to healthcare due to cost)
- **GENHTLH** (User's general health rating)
- **MENTHLTH** (User's mental health for past 30 days)
- **PHYSHLTH** (User's physical Health for past 30 days)
- **DIFFWALK** (Difficulty on Walking)

Before data cleaning

Rename all the column header

Remove rows that contains null values in categorical column



	IYEAR	SEXVAR	AGE80	AGEG5YR	EDUCA	INCOME2	BPHIGH4	TOLDHI2	BMI5	SMOKE100	CVDSTRK3	MICHD
1	b'2019'	2	80	13	3	3	1	1	2817	1	2	
2	b'2019'	2	70	11	5	5	3	2	1854	2	2	
3	b'2019'	2	67	10	6	7	1	2	3162	2	2	
4	b'2019'	2	80	13	5	6	1	2	2030	NA	2	
5	b'2019'	2	80	13	5	99	3	1	2148	1	2	
6	b'2019'	2	69	14	4	NA	1	1	NA	NA	2	
7	b'2019'	1	71	11	6	7	3	2	3298	1	2	
8	b'2019'	1	72	11	6	99	3	1	3081	2	2	
9	b'2019'	1	77	12	4	5	1	1	2399	1	1	
10	b'2019'	2	71	11	2	3	3	1	1665	1	2	
11	b'2019'	2	70	11	6	7	3	2	2231	1	2	
12	b'2019'	1	55	8	5	99	3	2	2511	1	2	
13	b'2019'	2	80	13	5	NA	1	1	NA	NA	2	

Replace values

Filter values that contains code for "Don't Know" or "Refused"

Formatting the numerical column in 2 decimal places

Fill in null values for numeric column

- Sample codes for replacing values in multiple column:

Source on Save

```
# Remove all 7 (don't knows)
# Remove all 9 (refused)
db["diabetes"][db["diabetes"] == 3] <- "Healthy"
db["diabetes"][db["diabetes"] == 2] <- "Gestational Diabetes"
db["diabetes"][db["diabetes"] == 1] <- "Diabetes"
db["diabetes"][db["diabetes"] == 4] <- "Prediabetes"
db <- filter(db, (diabetes!=7) & (diabetes!=9))
unique(db$diabetes)

#High Blood Pressure - Replace values to only 2 options:
#     0 for no blood pressure
#     1 for yes high blood pressure
#     2 for during pregnancy only
#     Remove all 7 (don't knows)
#     Remove all 9 (refused)
db$blood_pressure[db["blood_pressure"] == 1] <- "Yes"
db$blood_pressure[db["blood_pressure"] == 2] <- "Yes, During Pregnancy"
db$blood_pressure[db["blood_pressure"] == 3] <- "No"
db$blood_pressure[db["blood_pressure"] == 4] <- "Prehypertensive"
db <- filter(db, (blood_pressure!=7) & (blood_pressure!=9))

#High Cholesterol - replace value 2 to 0, since 2 represents No High Choleste
# Remove all 9 (refused)
db$high_chol[db["high_chol"] == 1] <- "Yes"
db$high_chol[db["high_chol"] == 2] <- "No"
db <- filter(db, (high_chol!=7) & (high_chol!=9))

#Stroke - Replace value 2 to 0 because it is a No
# Remove all 7 (don't knows) and 9 (refused)
db$stroke[db["stroke"] == 1] <- "Yes"
db$stroke[db["stroke"] == 2] <- "No"
```

```
#remove rows that has null values in categorical columns
library(tidyr)
db_2019$bmi[is.na(db_2019$bmi)] <- mean(db_2019$bmi, na.rm = TRUE)
db_2019$weight[is.na(db_2019$weight)] <- mean(db_2019$weight, na.rm = TRUE)
db_2019$height[is.na(db_2019$height)] <- mean(db_2019$height, na.rm = TRUE)
db_2019 <- db_2019%>% drop_na()

db <-db_2019

####### === Diabetes and Other Related Disease Section  === #######
#Diabetes - Replace values to be more suitable for ML algorithms
```

- Fill in the missing rows that contains in numerical columns with column mean values.
- Remove rows that contains null values in categorical data

After data cleaning

Filter

	year	gender	age	weight	height	age_category	education	income	blood_pressure	high_chol	bmi	smoking
1	2019	Female	80	69.8500	1.57	>80	Some High School	15,000 - 20,000	Yes	Yes	28.17000	Yes
2	2019	Female	70	48.9900	1.63	70-74	College student	25,000 - 35,000	No	No	18.54000	No
3	2019	Female	67	86.1800	1.65	65-69	College Graduate	50,000 - 75,000	Yes	No	31.62000	No
4	2019	Male	71	113.4000	1.85	70-74	College Graduate	50,000 - 75,000	No	No	32.98000	Yes
5	2019	Female	71	42.6400	1.60	70-74	Elementary Level	15,000 - 20,000	No	Yes	16.65000	Yes
6	2019	Female	70	58.9700	1.63	70-74	College Graduate	50,000 - 75,000	No	No	22.31000	Yes
7	2019	Male	76	70.3100	1.65	75-79	College Graduate	> 75,000	No	Yes	25.79000	Yes
8	2019	Female	51	77.5600	1.73	50-54	High School Graduate	35,000 - 50,000	No	No	26.00000	No
9	2019	Female	73	61.2300	1.63	70-74	High School Graduate	35,000 - 50,000	No	No	23.17000	Yes
10	2019	Female	80	61.2300	1.65	>80	College student	35,000 - 50,000	No	No	22.46000	No
11	2019	Female	68	81.6500	1.63	65-69	College student	50,000 - 75,000	Yes	No	30.90000	No
12	2019	Male	68	81.6500	1.75	65-69	College student	> 75,000	Yes	Yes	26.58000	No
13	2019	Female	77	90.7200	1.60	75-79	High School Graduate	35,000 - 50,000	Yes	Yes	35.43000	No
14	2019	Female	48	81.6500	1.70	45-49	College Graduate	> 75,000	Yes	Yes	28.19000	No
15	2019	Female	44	65.7700	1.70	40-44	College Graduate	> 75,000	No	No	22.71000	No
16	2019	Female	66	97.0700	1.65	65-69	College Graduate	20,000 - 25,000	Yes	Yes	35.61000	No
17	2019	Male	80	99.7900	1.75	>80	College Graduate	25,000 - 35,000	Yes	No	32.49000	No

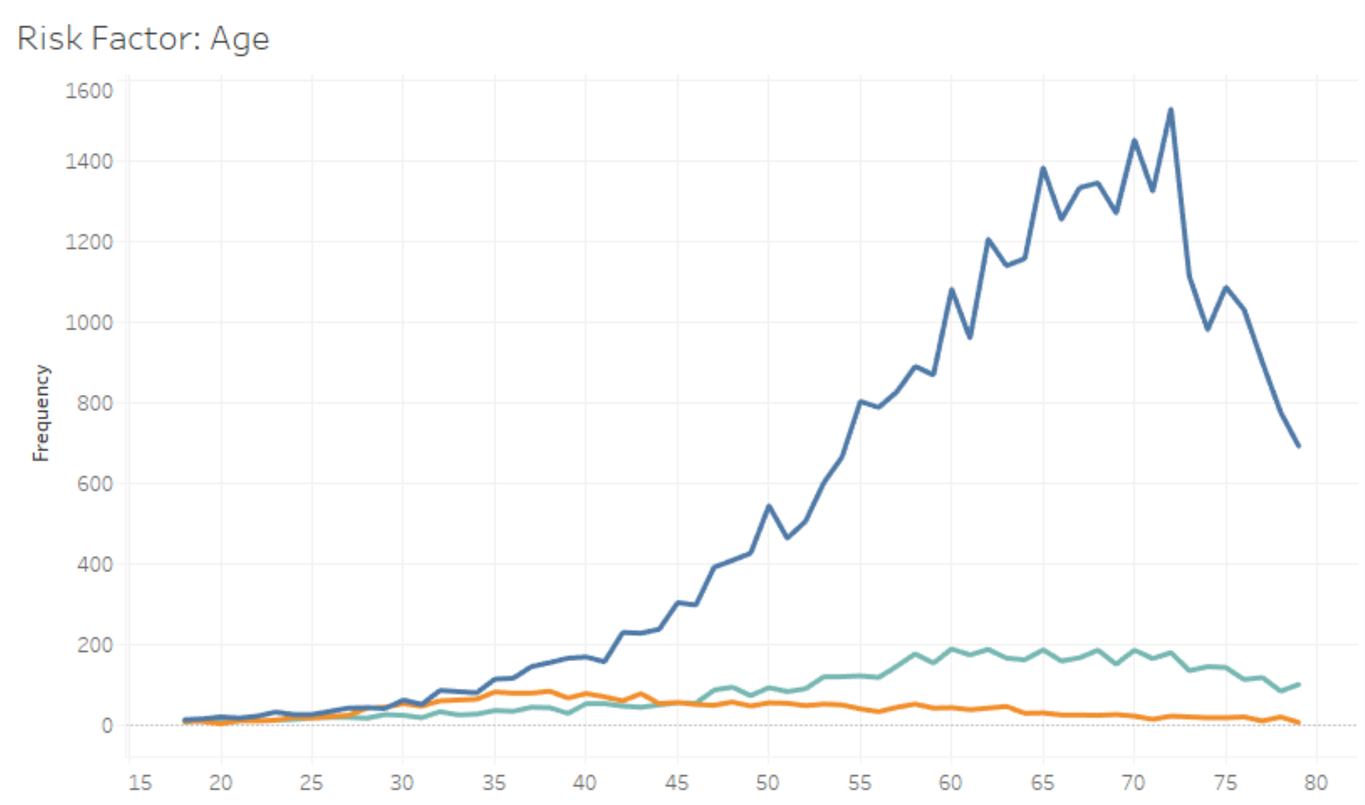
DATA TRANSFORMATION

Few steps before we train the model:

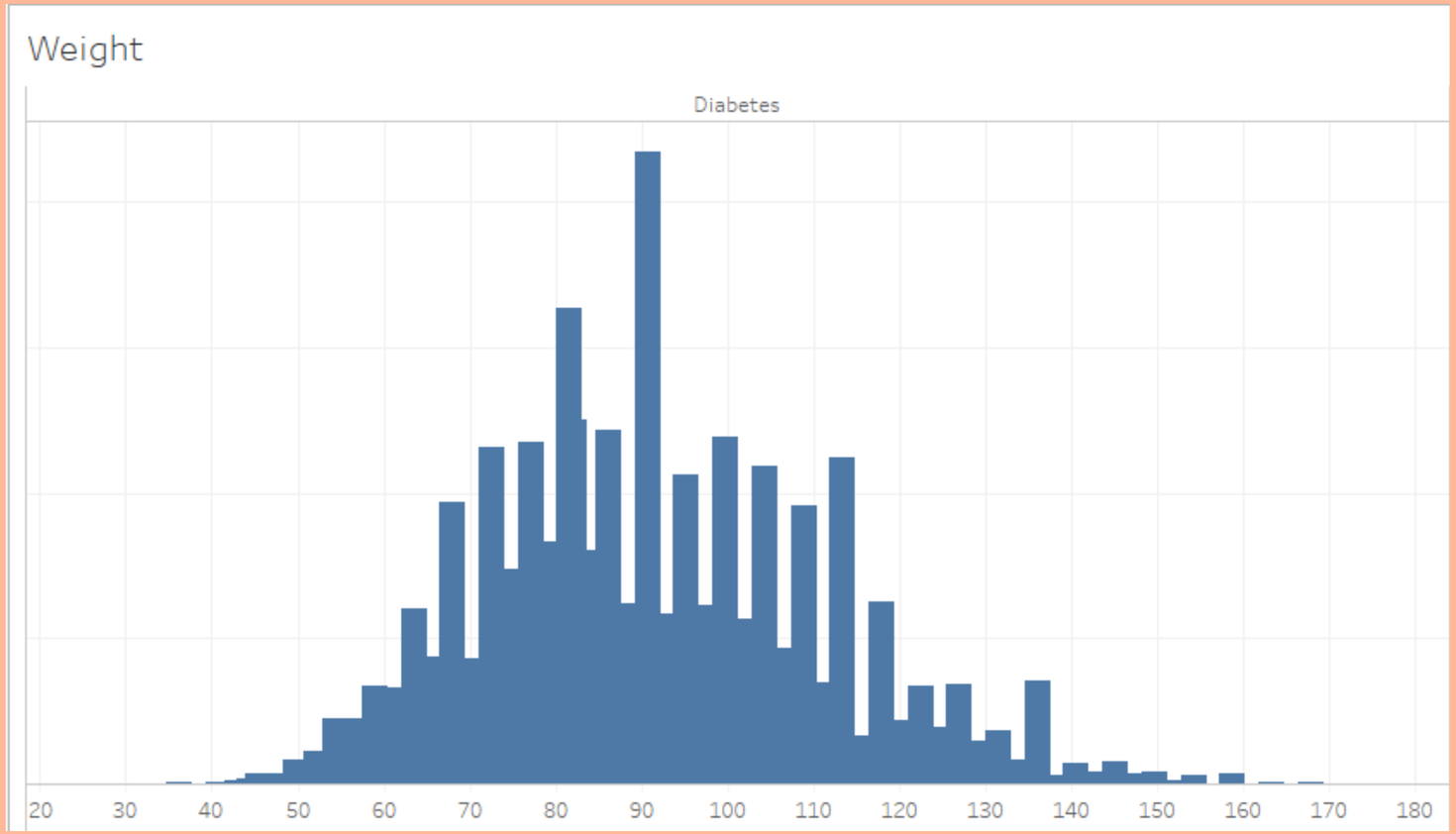
1. Create new data frame for predictive modeling
2. Filter out columns that have less importance.
3. Standardize all numerical inputs
4. Label all categorical inputs using LabelEncoder()
5. Split the dataset into training set and testing set (80:20)
6. Random sampling the training data

EXPLORATORY DATA ANALYSIS

Question 1: What are the **risk factors** might occur to diabetic patients that can be predicted the most?



From the graph trend, as the age increases, the number of people suffered diabetes increased.



The number of people getting diabetes increases as the weight value increases. It is proven that being overweight raises your risk for type 2 diabetes

Blood Pressure

Diabetes	High	Normal	Prehypertensive
Diabetes	27,581	9,320	259
Gestational Diab..	673	1,485	19
Prediabetes	3,484	2,174	207

- People who have high blood pressure would likely to get the risk of getting diabetes.
- This has proven that diabetes are related with high blood pressure.

Cholesterol

Diabetes	High Chol	Low Chol
Diabetes	23,500	13,781
Gestational Diabetes	716	1,619
Prediabetes	3,308	2,596

- Based on our findings, people with high cholesterol are also will having diabetes risk.
- In our analysis, people with high cholesterol have a higher number of people who also getting diabetes compared to people with low cholesterols

Physical Activity

Diabetes	Have Physical Activity	No Physical Activity
Diabetes	23,376	13,905
Gestational Diabetes	1,786	549
Prediabetes	4,107	1,797

- However, physical activity are can be quite hard to predict if the person have the risk of diabetes, due to different metabolisms on each person might have.

Metabolism - Better Health Channel. (2020, April 30). Better Health Channel. https://www.betterhealth.vic.gov.au/health/conditionsandtreatments/metabolism

Question 2: What are people's **lifestyle habits** that the user mostly likely for them to get diabetes?

It is found that lifestyle habits such as **No Physical Activity** is the most likely for the people to get diabetes.

LifeStyle Values

Diabetes	Have Physical Activity	No Physical Activity
Diabetes	23,935	14,381
Gestational Diabetes	1,942	602
Prediabetes	4,192	1,858

LifeStyle Values

Diabetes	Heavy Drinker	Not Alcohol Drinker
Diabetes	1,019	37,297
Gestational Diabetes	161	2,383
Healthy	16,028	216,053
Prediabetes	283	5,767

LifeStyle Values

Diabetes	Do smoking	Not Smoker
Diabetes	19,053	19,263
Gestational Diabetes	927	1,617
Healthy	95,385	136,696
Prediabetes	2,781	3,269

Question 3: How to differentiate between prediabetes with Type 2 diabetes in terms of symptoms or any risk factor?

- Based on our findings, prediabetes and diabetes may have many similarities in terms of symptoms and any risk factors that may contribute to it.
- It is reported from CDC that people with prediabetes has high blood sugar levels compared to normal person, but it is not high enough to be diagnosed as diabetes.

Risk Factor: Cholesterol

Diabetes	High Cholesterol	Low Cholesterol
Diabetes	23,984	14,332
Gestational Diabetes	760	1,784
Prediabetes	3,369	2,681

Risk Factor: Physical Activity

Diabetes	Have Physical Activity	No Physical Activity
Diabetes	23,935	14,381
Gestational Diabetes	1,942	602
Prediabetes	4,192	1,858

About Prediabetes and Type 2 Diabetes | National Diabetes Prevention Program | Diabetes | CDC. (2021, December 21). Centers for Disease Control and Prevention. https://www.cdc.gov/diabetes/prevention/about-prediabetes.html#:%7E:text=What%20are%20Prediabetes%20and%20Diabetes,Prediabetes%20can%20often%20be%20reversed.

Question 4: Does having health complications might increase the chance of getting diabetes, or vice versa?

Illness: Stroke

Illness	Diabetes	Diabetes: Gestational Diabetes	Diabetes: Prediabetes
No	34,427	2,489	5,654
Yes	3,889	55	396

Illness: Heart Disease or Attack

Illness	Diabetes	Diabetes: Gestational Diabetes	Diabetes: Prediabetes
No	29,976	2,440	5,285
Yes	8,340	104	765

Illness: High Blood Pressure

Illness	Diabetes	Diabetes: Gestational Diabetes	Diabetes: Prediabetes
No	9,633	1,625	2,237
Prehypertensive	263	20	206
Yes	28,293	719	3,566
Yes, During Pregnancy	127	180	41

For health complications such as **stroke** and **heart disease**, having such health complications may not increase the chance of getting diabetes.

For **high blood pressure (hypertension)**, the people who have this complication might increase their chance of getting diabetes.

In this case, we can see that having health complication may not increase your chance of getting diabetes, but having diabetes may increase your chance of having health complications

PREDICTIVE MODELING

1. Generalized Logistic Regression

- Using the R package named **caret** and **glm**
- The glm() function returns a model object, therefore we may apply extractor functions
- Predicting only 2 class **(Healthy, Diabetes)**

```
#For Cross validation to 10 fold
tr_control <- trainControl(method = "cv", number = 10, savePredictions = TRUE)

#Model training
fit_under <- train(form = diabetes ~ .,
                   data = tr_under,
                   trControl = tr_control,
                   method = "glm",
                   family = "binomial" )
fit_under
confusionMatrix(fit_under)
#Accuracy = 0.7079008, Kappa = 0.4158016
```

PREDICTIVE MODELING (CONT')

3. Decision Tree Classifier

- Using the R package named **rpart** for model training and **rpart.plot** for plotting the decision trees for binary classification

```
# Create the tree (for undersampled data)
output.tree <- rpart(diabetes ~ ., data = tr_under, method = 'class')
# Plot the tree
rpart.plot(output.tree, extra = 106)
accuracy_tune(output.tree)

# Create the tree (for oversampled data)
output.tree2 <- rpart(diabetes ~ ., data = tr_over, method = 'class')
# Plot the tree
rpart.plot(output.tree2, extra = 106)
accuracy_tune(output.tree2)
```

MODEL EVALUATION

Generalized Logistic Regression (using **caret** and **glm**)

```
Generalized Linear Model

71888 samples
   12 predictor
    2 classes: '0', '1'

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 64699, 64700, 64700, 64699, 64700, 64699,
Resampling results:

  Accuracy   Kappa
  0.7101881  0.4203762

> confusionMatrix(fit_under)
Cross-Validated (10 fold) Confusion Matrix

(entries are percentual average cell counts across resamples)

          Reference
Prediction    0    1
         0 35.1 14.1
         1 14.9 35.9

 Accuracy (average) : 0.7102
```

Decision Tree Classifier (using **rpart**)

```
Confusion Matrix and Statistics

          Reference
Prediction Diabetes Healthy
  Diabetes     6380   16591
  Healthy      2590   28020

               Accuracy : 0.642
                 95% CI : (0.6379, 0.6461)
    No Information Rate : 0.8326
    P-Value [Acc > NIR] : 1

                  Kappa : 0.209

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.7113
            Specificity : 0.6281
         Pos Pred Value : 0.2777
         Neg Pred Value : 0.9154
             Prevalence : 0.1674
         Detection Rate : 0.1191
   Detection Prevalence : 0.4287
      Balanced Accuracy : 0.6697

       'Positive' Class : Diabetes
```

MODEL EVALUATION (CONT')

ACCURACY

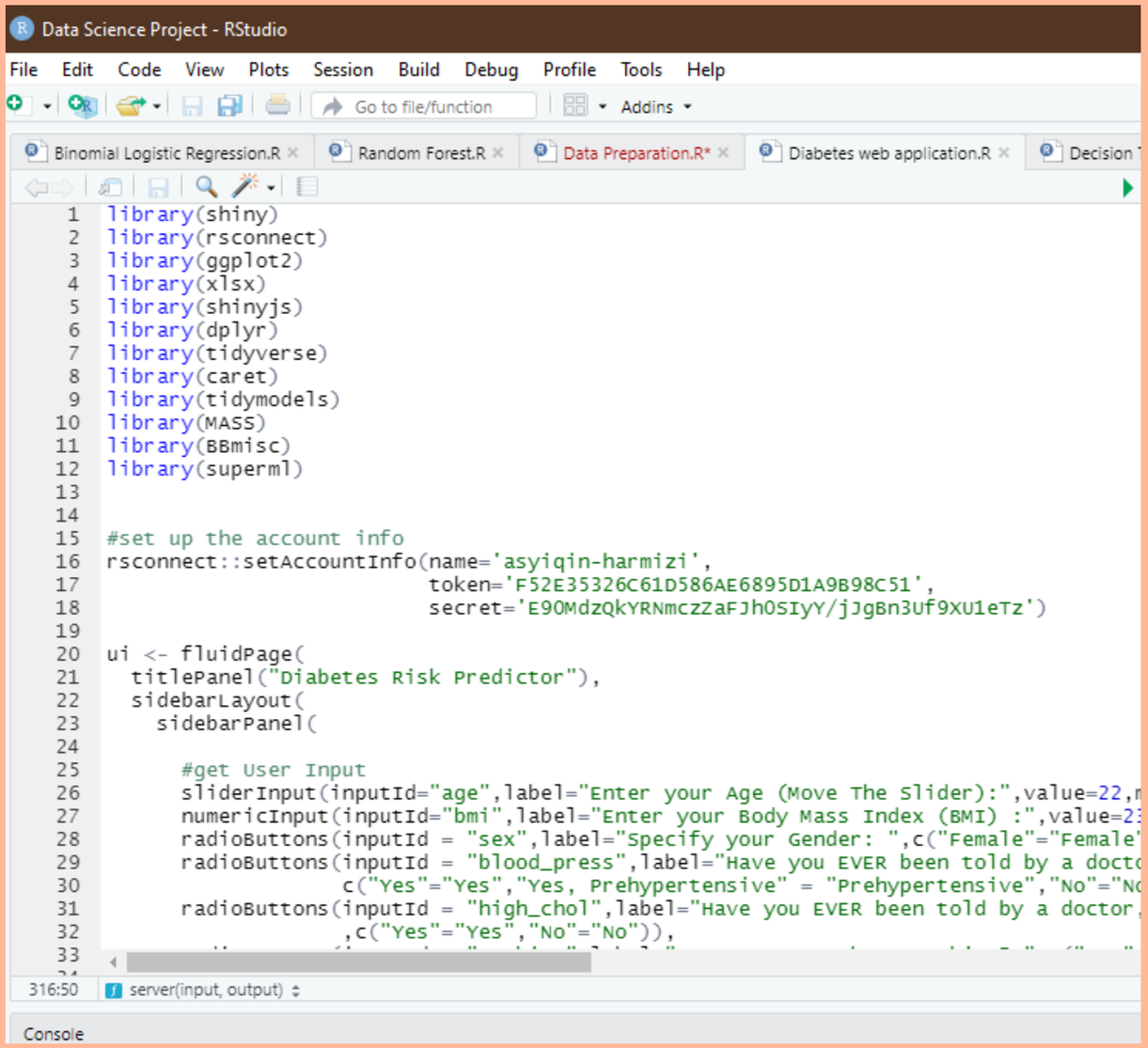
- Generalized Logistic Regression
  -> **0.7102**
- Decision Trees Classifier
  -> **0.642**

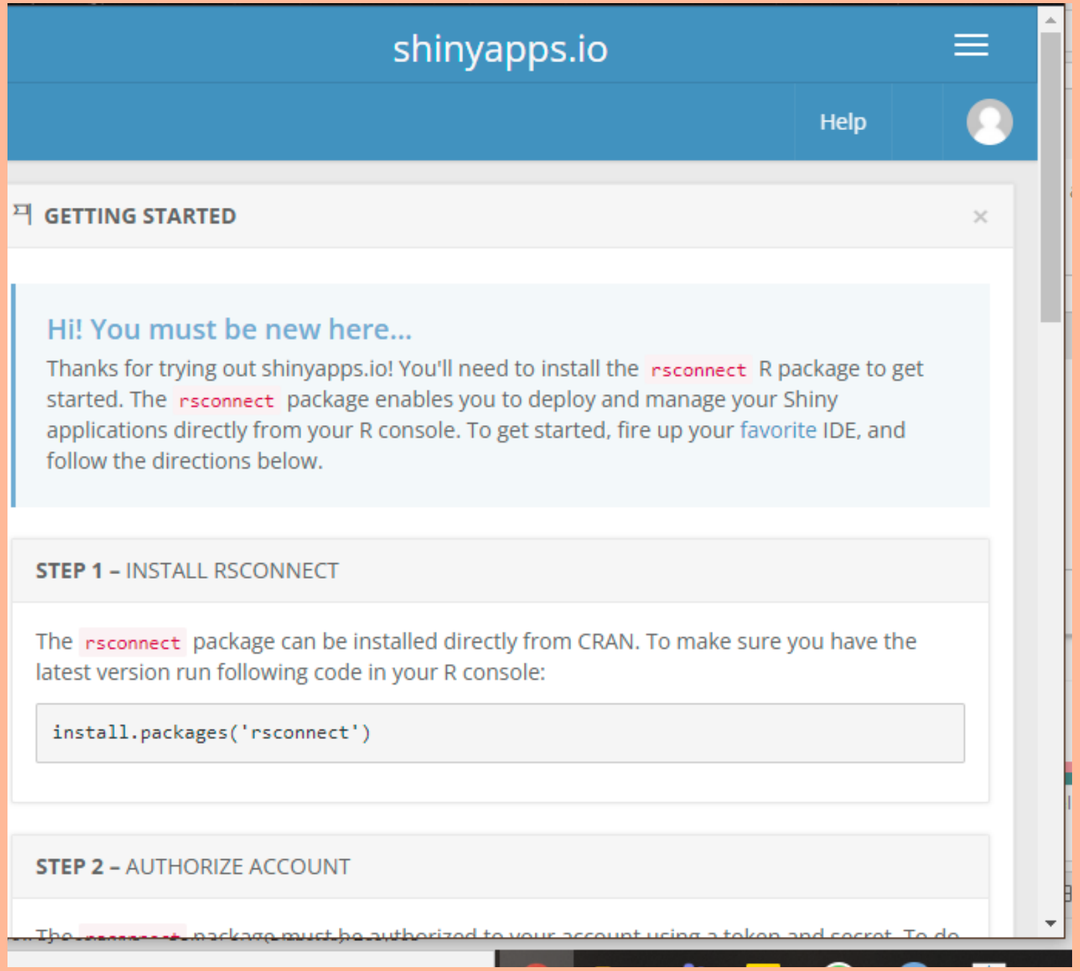
KAPPA

- Generalized Logistic Regression
  -> **0.42037**
- Decision Trees Classifier
  -> **0.209**

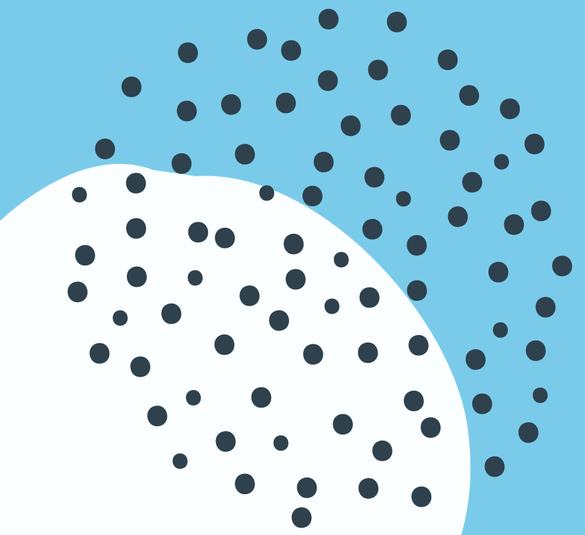
DEPLOYMENT

- Create a web application using R Shiny Apps
- Apply the model with better performance into Diabetes Risk Predictor App


Data Science Project - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Binomial Logistic Regression.R
Random Forest.R
Data Preparation.R*
Diabetes web application.R
Decision
1 library(shiny)
2 library(rsconnect)
3 library(ggplot2)
4 library(xlsx)
5 library(shinyjs)
6 library(dplyr)
7 library(tidyverse)
8 library(caret)
9 library(tidymodels)
10 library(MASS)
11 library(BBmisc)
12 library(superml)
15 #set up the account info
16 rsconnect::setAccountInfo(name='asyiqin-harmizi',
17 token='F52E35326C61D586AE6895D1A9B98C51',
18 secret='E9OMdzQkYRNmczZaFJhOSIyY/jJgBn3Uf9XU1eTz')
20 ui <- fluidPage(
21 titlePanel("Diabetes Risk Predictor"),
22 sidebarLayout(
23 sidebarPanel(
25 #get User Input
26 sliderInput(inputId="age",label="Enter your Age (Move The Slider):",value=22,
27 numericInput(inputId="bmi",label="Enter your Body Mass Index (BMI) :",value=2
28 radioButtons(inputId = "sex",label="Specify your Gender: ",c("Female"="Female"
29 radioButtons(inputId = "blood_press",label="Have you EVER been told by a docto
30 c("Yes"="Yes","Yes, Prehypertensive" = "Prehypertensive","No"="No
31 radioButtons(inputId = "high_chol",label="Have you EVER been told by a doctor,
32 ,c("Yes"="Yes","No"="No")),
316:50
server(input, output)
Console


shinyapps.io
Help
GETTING STARTED
Hi! You must be new here...
Thanks for trying out shinyapps.io! You'll need to install the rsconnect R package to get started. The rsconnect package enables you to deploy and manage your Shiny applications directly from your R console. To get started, fire up your favorite IDE, and follow the directions below.
STEP 1 – INSTALL RSCONNECT
The rsconnect package can be installed directly from CRAN. To make sure you have the latest version run following code in your R console:
install.packages('rsconnect')
STEP 2 – AUTHORIZE ACCOUNT

TOOLS USED & CODES



RSTUDIO

- Data Preparation
- Predictive modeling and analysis

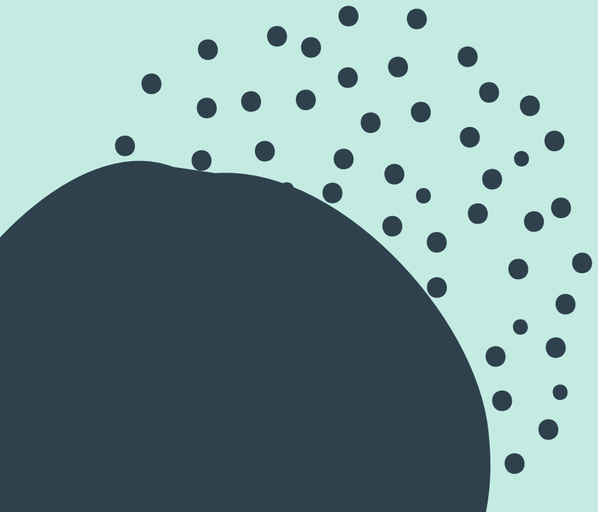
TABLEAU

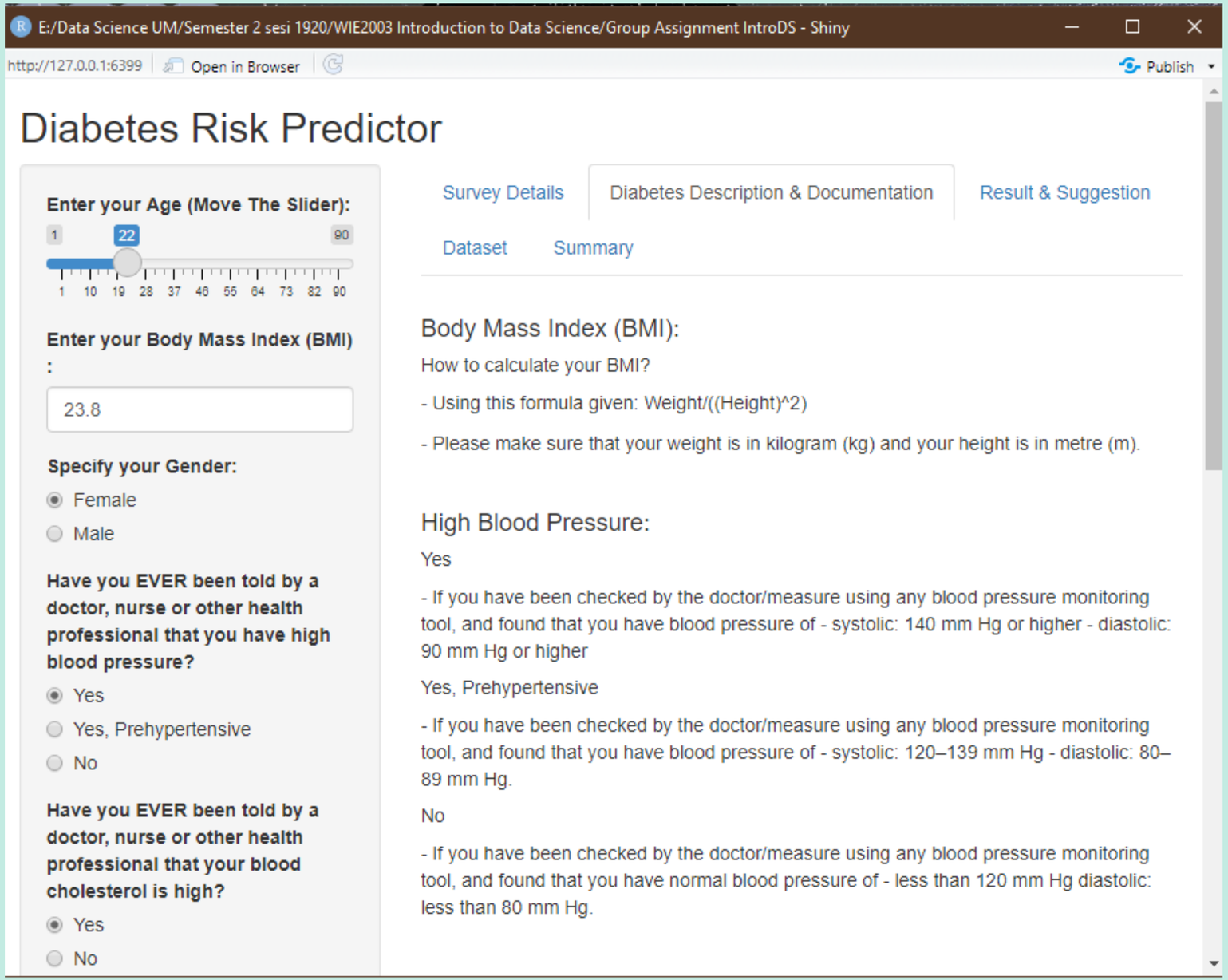
- Exploratory Data Analysis

RSHINY APPS

- Creating web application

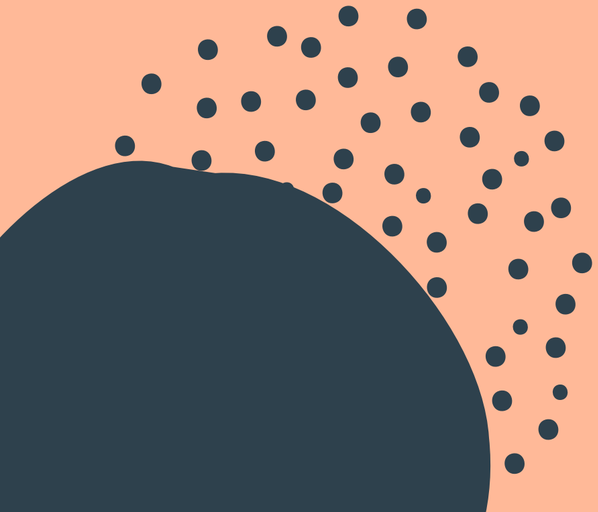
APPLICATION DEMONSTRATION

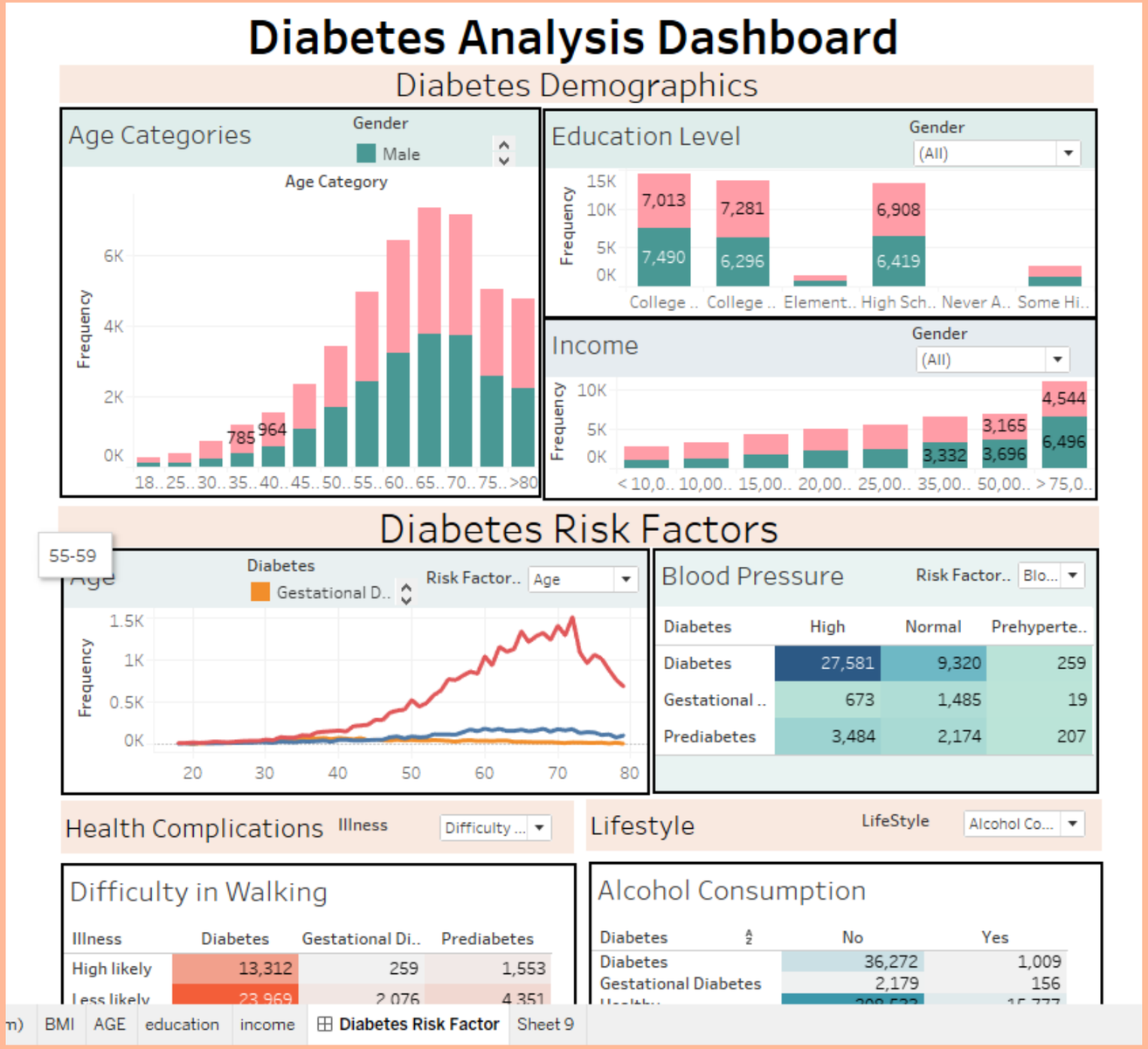




Diabetes Risk Predictor App using RShiny Apps

DATA STORY AND VISUALIZATION





Diabetes Analysis Dashboard

- Consist of analyses based on people who have different diabetes types (Diabetes, Prediabetes and Gestational Diabetes)
- Contains:
  - Demographic of people who have different diabetes types
  - Diabetes risk factors
  - Health complication
  - Lifestyle that contributed to diabetes.

CONCLUSION

- Lifestyle habits such as less physical activities can cause the person to have risk of getting diabetes.
- Age, weight, blood pressure and cholesterol are the risk factors that can be used to predict either a person having the risk of getting diabetes.
- The person that has health complications may not getting diabetes, but getting diabetes may increased the risk of having health complications.
- It is very important for us to get the confirmation from health professionals either we having the diabetes risk and get treated.

THANK YOU

References :-

1. *About Prediabetes and Type 2 Diabetes | National Diabetes Prevention Program |Diabetes | CDC*. (2021, December 21). Centers for Disease Control and Prevention. https://www.cdc.gov/diabetes/prevention/about-prediabetes.html#:%7E:text=What%20are%20Prediabetes%20and%20Diabetes,Prediabetes%20can%20often%20be%20reversed.
2. *COVID-19 and diabetes*. (2021, May 20). International Diabetes Federation. https://www.idf.org/aboutdiabetes/what-is-diabetes/covid-19-and-diabetes/1-covid-19-and-diabetes.html
3. *Diabetes*. (2021, November 10). World Health Organization. https://www.who.int/news-room/fact-sheets/detail/diabetes
4. *Diabetes Symptoms*. (2021, April 27). Centers for Disease Control and Prevention. https://www.cdc.gov/diabetes/basics/symptoms.html

6. *Diabetes - Symptoms and causes*. (2020, October 30). Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/diabetes/symptoms-causes/syc-20371444
7. *Metabolism - Better Health Channel*. (2020, April 30). Better Health Channel. https://www.betterhealth.vic.gov.au/health/conditionsandtreatments/metabolism
8. Nadeau, K. J., Anderson, B. J., Berg, E. G., Chiang, J. L., Chou, H., Copeland, K. C., Hannon, T. S., Huang, T. T. K., Lynch, J. L., Powell, J., Sellers, E., Tamborlane, W. V., & Zeitler, P. (2016). *Youth-Onset Type 2 Diabetes Consensus Report: Current Status, Challenges, and Priorities*. Diabetes Care, 39(9), 1635-1642. https://doi.org/10.2337/dc16-1066
9. Watson, S. (2020, February 27). *Everything You Need to Know About Diabetes*. Healthline. https://www.healthline.com/health/diabetes