

# Business Intelligence Data Visualization for Diabetes Health Prediction

## Data Analytics and Insights for Diabetes Prediction

Samantha Siow Jia Qi, Sarasvathi Nagalingham

Faculty of Data Science and Information Technology  
INTI International University, Nilai, Negeri Sembilan, Malaysia

**Abstract**—In today's environment, Business Intelligence (BI) is transforming the world at a rapid pace across domains. Business intelligence has been around for a long time, but when combined with technology, the results are astounding. BI is also playing an important role in the healthcare domain. Centers for Disease Control and Prevention (CDC) is the largest science-based, data-driven service provider in the country for public health protection. For over 70 years, has been using science to fight disease and keep families, businesses, and communities healthy. However, research indicates that the prevalence of diabetes in the US is rising alarmingly. As a result, if diabetes is not treated, it can lead to life-threatening complications such as heart disease, loss of feeling, blindness, kidney failure, and amputations. As a result, this study was conducted to analyze people's health conditions and daily lifestyles in order to predict which type of diabetes they would most likely diagnose with the implementation of business intelligence using Tableau dashboard. Furthermore, background research is conducted on CDC to understand their work, challenges, and opportunities. By the end of the project, the information obtained and visualized should be able to enhance business choices and make better decisions on controlling diabetes in the future.

**Keywords**—Diabetes; business intelligence; prediction; dashboard visualization; data analysis; centers for disease control and prevention

### I. INTRODUCTION

#### A. Business Intelligence Methodology

Healthcare analytics and business intelligence (BI) are two emerging technologies that offer analytical capability to aid the healthcare sector in enhancing service quality, lowering expenses, and managing risks. With the increasing volume of data and the desire to learn from it, demand for BI applications for healthcare continues to rise. The need for data management and analysis expertise in healthcare is rapidly increasing. With the help of Business Intelligence, companies can make better decisions by displaying current and historical data within the context of their business. Analysts may use BI to generate the company's performance and competitive standards, which will help the companies operate more smoothly and effectively as well as identify the market trends for enhancement. In any BI project, a consistent methodology and approach must be defined as they can let decision-makers develop, support, and integrate best management practices within the organization. Moreover, it can achieve a higher chance of success, save time and effort, eliminate unnecessary

operations, and ensure accurate reporting and analysis. Therefore, the Agile BI methodology will be adopted in the context of our domain, Centers for Disease Control and Prevention (CDC).

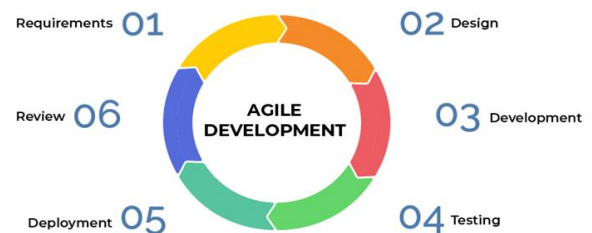


Fig. 1. Phases of agile methodology.

The Agile BI methodology is a project management strategy that focuses on continuous improvement and is suitable for those that demand speed and flexibility to satisfy customer requirements. It is considerably more accessible than it looks to apply Agile BI methodology to an organization's internal business processes [5]-[6]. The following are the phases of the developing an Agile BI methodology (Fig. 1):

- **Requirement:** In this phase, the project team specifies the requirements. They should outline the business opportunities of the project that can result in profit and business growth, along with estimating the time and effort required to complete it. Technical and economic feasibility can be determined as well as whether the project is worth pursuing based on this information.
- **Design:** This phase required working with the company's stakeholders to determine the requirements once the project had been identified. Diagrams such as flow diagrams and high-level UML diagrams are used to demonstrate how the new features perform and how they integrate into the existing system.
- **Development:** The works begin once the project team has identified the requirements based on stakeholder feedback. UX designers and developers start working on the project's initial iteration to deliver a viable product. The product will go through several phases of improvement to meet the stakeholder's requirements.

- **Testing:** The Quality Assurance team will check the product's performance and address any defects and bugs that produce unexpected results. The software engineer team will finalize the system and user documentation once the product is satisfied.
- **Deployment:** This phase includes ongoing software release support, especially when the project team must keep the system working well and demonstrate users on using the system. The production phase comes to an end when the support has ended.
- **Review:** The final phase is to get feedback from the stakeholder once the product has been released. The project team will receive feedback on the development and work through it to improve their work for the next project.

### B. Literature Review

The Centers for Disease Control and Prevention (CDC) is the United States' main public health agency. It is a government agency of the United States that belongs to the Department of Health and Human Services and is headquartered in Atlanta, Georgia. The agency's primary aim is to protect public health and safety by controlling and preventing disease, injury, and disability in the United States and around the world. The CDC concentrates national attention on developing and implementing disease control and prevention strategies [7].

The Behavioral Risk Factor Surveillance System (BRFSS) is a health-related telephone survey conducted annually by the Centers for Disease Control and Prevention (CDC). The survey collects data from over 400,000 Americans each year on health-related risk behaviors, chronic health issues, and use of preventative treatments. It has been conducted every year since 1984. According to the CDC, 34.2 million Americans have diabetes and 88 million have prediabetes as of 2018. Furthermore, the CDC indicates that one in every five diabetics and about eight out of ten pre diabetics are unaware of their risk. To improve preventive care, EHR is necessary, which can monitor the patient's state of health and track treatment progress to make preventative care easier.

The main purpose of this business is to implement electronic health records (EHR) to give better care to patients. EHR and the ability to exchange health information electronically can help patients receive better quality and safer care while also providing substantial profit to the business. It automates a multitude of operations for the practice and allows for instant access to patient records for better coordinated and efficient care when compared to manually inserting, updating, or deleting data using a csv file. To improve productivity and work-life balance, physicians are allowed to share electronic information with one another remotely and in real time [1].

Furthermore, early diagnosis is essential since it can lead to lifestyle changes and more effective treatment, making diabetes risk prediction models important tools for the general population and public health officials. Health analytics are given by applying EHR, which can help to recognize patterns, predict diagnoses, and recommend potential treatment

alternatives. Rather of relying on trial-and-error methods, these analytics lead to more successful overall patient results the first time.

### C. Challenges faced by CDC

Various changes are occurring in the Centers for Disease Control and Prevention (CDC), creating new challenges to both large and small medical businesses. To name a few, the integration of services, technological developments, and patient preferences have created a new environment in which performing a medical treatment is no longer solely about treating patients.

First, big data development is well received in the medical world, but implementation may not be as easy. Non-relational databases combine patient data from several sources to provide meaningful metrics. The technology appears at the ideal time to meet the demand for newly available patient information warehouses. Relational databases have traditionally been used by healthcare providers to manage and store patient records. Nevertheless, relational databases are incapable of managing unstructured data, such as medical documentation and transcripts. Only a small fraction of healthcare providers has successfully transitioned from relational to non-relational databases using standard electronic health records (EHRs). Most firms that effectively employ non-relational information systems are big and financially secure.

Besides that protecting the devices that protect public health is important as many developments in healthcare technology use Internet connectivity. However, this convenience extends to cyber attackers as well. Malicious cyber-attacks will become more common as the Internet-connected medical device industry evolves. The US authorities warned people in 2015 that hackers may command infusion pumps to deliver harmful drug doses. The leak raised the prospect of malevolent programmers infiltrating medical devices and causing harm to people. Moreover, hackers utilize medical equipment to invade care provider networks, stealing research and clinical trial data as well [13]. Presently, no medical equipment has been infiltrated by hackers, resulting in a fatal event. Yet, cyber security experts warn that an assault on an untrained care provider can harm an organization in a variety of ways.

Furthermore, as patients are becoming more liable for a growing amount of their medical expenses, health providers are listing patient collections as their top revenue cycle management challenge. Providers must comply with patient payment preferences to urge patients to make payments on time. Invoice statements should be patient-friendly to meet the expectations of patients and enhance their user experience. For example, e-Statements and a range of payment alternatives, such as credit cards, etc. through an online patient portal can be used. However, setting up such billing and payment processing systems in-house can be difficult and expensive for medical offices. They must not only discuss agreements with each payment processor and develop the infrastructure, but they must also bear the continuous administrative costs of such technology. Besides, healthcare professionals must adhere to stringent criteria to preserve patient information.

You must ensure that your payment interface and processing system are entirely compliant, or you may face a severe penalty.

Lastly, in recent years, there have been substantial changes in the health insurance sector of the CDC. As more patients bear a greater share of their medical costs, they naturally expect greater services from their providers. The CDC will confront increased competition in gaining and maintaining patients who seek a level of customer service comparable to that of other retail chains. For example, they expect a streamlined patient experience in which they can "self-service" to settle most doubts, problems, or concerns whenever, wherever, and however they see fit [2], [14]-[15].

#### D. Opportunities of CDC

After looking into the difficulties or problems that arise in the disease control and prevention centers, we have identified some alternative solutions as well as opportunities to ensure the organization remains sustainable. The opportunities of disease control and prevention centers are listed as follows.

Due to the Centers for Disease Control and Prevention (CDC) currently using the traditional spreadsheet approach to store and manage patient records, it slows down the workflow of the organization. Therefore, CDC needs to transition from using a spreadsheet like Microsoft Excel to an electronic health records (EHR) system. The EHR system helps their CDC staff to capture and access information gathered during patient appointments with just a click. It is more user-friendly than a spreadsheet since the process of adding, updating, viewing, and deleting the patient's data can be done systematically and more simply. Moreover, the EHR system allows multiple staff to have simultaneous access to health records anytime and anywhere. And most significantly, it also enables the staff to share information with other practices suchlike emergency facilities, specialists, and laboratories, to contain all the information clinicians provide for patient care [9]-[10].

Besides that CDC will be able to expand the view of patient care and ensure every patient has gotten the medical services they need. With the use of the EHR system, it has provided a variety of useful functions like the capability to generate the analytical report based on clinical data collected from ongoing patient care. In healthcare, the analytical report is used to detect or predict early signs of patient deterioration and aids in the more definitive patient diagnosis, followed by the appropriate treatment of the identified indication. Accurate preventive measures have been shown to reduce mortality and morbidity rates in diagnosed patients.

Lastly, a high level of security management retains patients as EHR systems provide better security of confidential records. Certain users might be granted varying levels of access to patients' data to ensure that the sensitive files are protected and kept safe. The unauthorized individuals who aim to access the system will be restricted. The probability of valuable records being stolen by theft or hacker is lower. As a result, CDC has gained trust and built loyalty among patients.

## II. PROPOSING BI SOLUTIONS

Predictive analytics is one of the top business intelligence trends for the past two years in a row, but the potential applications extend well beyond business and far into the future. CDC can build a database for predictive analytics tools that would enhance care delivery. This is especially important in the case of individuals who have a complicated medical history and are suffering from various illnesses. The purpose of healthcare internet business intelligence is to assist doctors in making data-driven choices in seconds and improving patient care. The purpose of healthcare internet business intelligence is to assist health care agencies in making data-driven choices in seconds and improving the health of the United States citizens. CDC can also use BI tools to make a predictive diabetes analysis in detail. With these, we can clearly know the health of the United States citizens and identify the key factors about who are getting diseases. Then, CDC can find the best solution to reduce the Americans' diseases. However, new BI solutions and tools would also be able to anticipate who is at risk of diabetes, so it is advised on extra testing or weight control [8].

In addition to predictive analysis, BI tools may assist firms in analysing clinical data such as assessing lab test results, the incidence of unfilled prescriptions, and so on. This allows CDC to track disease entails determining what causes individuals to get ill as well as the most efficient measures to avoid illness. It can also assist local committees in determining which regions require further funding.

Moreover, business intelligence is a relatively recent concept that refers to the gathering and analysis of data to better business operations and strategic planning. This similar paradigm underpins healthcare business intelligence, but the data in issue is patient data obtained through several sources. Using the BI tools can improve the decision making and ensure the data quality [12]. Using BI tools can also quickly generate accurate reports. BI tools make it simpler and easier for businesses to develop and share dashboards and gather information [11]. For example, Tableau includes several ready-made templates for users in the healthcare agency, which aids in installation by allowing firms to easily drill down into their data. Healthcare BI software is a subset of BI software that is aimed specifically at the healthcare sector. These technologies increase the ability of medical experts to examine data obtained from various sources. These sources might include patient files and medical data, but they can also include extra information.

## III. DATASET ANALYSIS

The dataset used in this project is retrieved from National Health and Nutrition Examine Survey (NHANES) by CDC [3]. The author did some analysis on the dataset and found out that the dataset is a health record of diabetes chronic disease representing the United States population of all ages. The dataset consists of attributes describing various blood testing, body mass index (BMI) and other symptoms that could cause diabetes. Moreover, it is also found that the dataset does not contain any individual's personal information as health records are confidential information. Due to the NHANES Data Release and Access Policy, these records are designated

as classified official documents because they include private and personal information. Therefore, it must be given a high level of data protection.

### A. Description of Dataset

This dataset contains around 200,000 records with 22 attributes. The attributes and its data type as well as description are shown below (Table I):

TABLE I. DESCRIPTION OF DATASET

No	Attribute Name	Data Type	Description
1	Diabetes_012	Numerical (Interval Scale)	Disease status 0 = no diabetes 1 = prediabetes 2 = diabetes
2	HighBP	Boolean	Blood pressure 0 = no high BP 1 = high BP
3	HighChol	Boolean	Cholesterol 0 = no high cholesterol 1 = high cholesterol
4	CholCheck	Boolean	Cholesterol check 0 = no cholesterol check in 5 years 1 = yes, having cholesterol check in 5 years
5	BMI	Numeric (Continuous)	Value of body Mass Index based on patient weight and height
6	Smoker	Boolean	Have ever smoked more than 100 cigarettes (5 packs) in a lifetime 0 = no 1 = yes
7	Stroke	Boolean	Have ever experienced a stroke 0 = no 1 = yes
8	HeartDiseaseorAttack	Boolean	Have ever experiencing coronary heart disease (CHD) or myocardial infarction (MI) 0 = no 1 = yes
9	PhysActivity	Boolean	Physical activity within the past 30 days 0 = no 1 = yes
10	Fruits	Boolean	Consume fruit at least one or more times per day 0 = no 1 = yes
11	Veggies	Boolean	Consume vegetables at least one or more times per day 0 = no 1 = yes
12	HvyAlcoholConsump	Boolean	Heavy drinkers • Adult men – having more than 14 drinks per week • Adult women – having more than 7 drinks per week 0 = no 1 = yes
13	AnyHealthcare	Boolean	Have any type of health care coverage (health insurance, prepaid plans like HMO)

			0 = no 1 = yes
14	NoDocbcCost	Boolean	In the past 12 months, have you had a need to see a doctor but couldn't because of the cost? 0 = no 1 = yes
15	GenHlth	Numerical (Interval Scale)	Would you say that your overall health is: Scale 1-5 1 = excellent 2 = very good 3 = good 4 = fair 5 = poor
16	MentHlth	Numerical (Interval Scale)	In the last 30 days, how many days did you have poor mental health, including stress, depression, emotional issues? Scale 1-30 days
17	PhysHlth	Numerical (Interval Scale)	In the last 30 days, how many days did you have poor physical health, including physical illness and injury? Scale 1-30 days
18	DiffWalk	Boolean	Have trouble walking or climbing stairwells? 0 = no 1 = yes
19	Sex	Nominal Scale	Gender 0 = female 1 = male
20	Age	Numerical (Ratio Scale)	Level age category 1 = Age 18 to 24 2 = Age 25 to 29 3 = Age 30 to 34 4 = Age 35 to 39 5 = Age 40 to 44 6 = Age 45 to 49 7 = Age 50 to 54 8 = Age 55 to 59 9 = Age 60 to 64 10 = Age 65 to 69 11 = Age 70 to 74 12 = Age 75 to 79 13 = Age 80 or older
21	Education	Numerical (Ratio Scale)	Education level Scale 1-6 1 = Never attended school or only kindergarten 2 = Grades 1 through 8 (Elementary) 3 = Grades 9 through 11 (Some high school) 4 = Grade 12 or GED (High school graduate) 5 = College 1 year to 3 years (Some college or technical school) 6 = College 4 years or more (College graduate)
22	Income	Numerical (Ratio Scale)	Income Scale 1-8 1 = less than \$10,000 5 = less than \$35,000 8 = \$75,000 or more

### B. Data Cleaning

The practice of correcting or deleting inaccurate, damaged, improperly formatted, duplicate, or incomplete data from a dataset is known as data cleaning. There are numerous ways for data to be duplicated or incorrectly categorized when merging multiple data sources. Even if results and algorithms appear to be correct, they are unreliable if the data is inaccurate. Because the procedures will differ from dataset to dataset, there is no one definitive way to specify the precise phases in the data cleaning process. But it is essential to create a template for your data cleaning procedure so you can be sure you are carrying it out correctly each time. The following techniques demonstrate how the dataset is cleaned by using Microsoft Excel and RStudio.

Microsoft Excel provides a lot of functions to the user and allows the user to show the data analysis results in a variety of ways. The data analysis results can be displayed as charts that emphasize the significant points in the data, and the audience will immediately comprehend what you want to project in the data. Excel allows us to visualize the data to reveal hidden data and identify data patterns. Several features have been picked from the vast array to be used in this project.

- Checking Missing Values

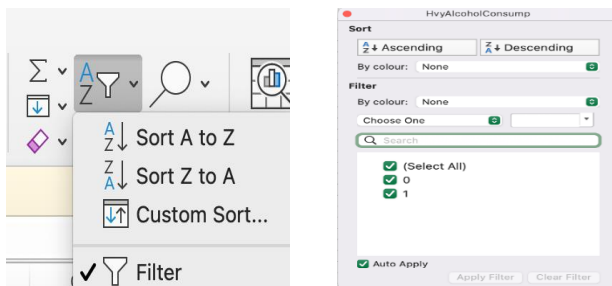


Fig. 2. Filter function in excel.

Filter function is used to filter a range of data depending on the criteria you define (Fig. 2). The filter function allows users to filter out all the data in a column, allowing them to see an overview of the data values. Furthermore, use this method to determine whether there is a null value in the column.

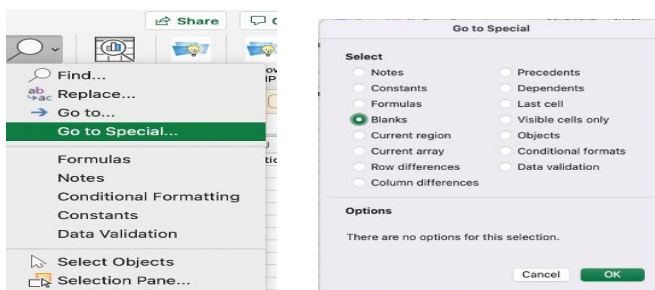


Fig. 3. Find function in excel.

The find function allows the user to check the missing value (Fig. 3). After that put all the missing values as blank. It means the dataset occurs missing value if it manages to be replaced as a blank field. So, it is required to handle it when

there is a null value to ensure data accuracy and improve data quality.

- Remove Duplicate Data

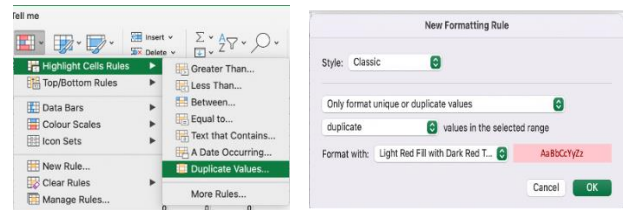


Fig. 4. Remove duplicate function.

Duplicate data might be valuable in some cases, but it can also make it difficult to interpret your data (Fig. 4). To discover and highlight duplicate data, conditional formatting can be used. That way, you may go through the duplicates and determine whether to eliminate them. It allows you to select the formatting you wish to apply to the duplicate values in the box next to values with.

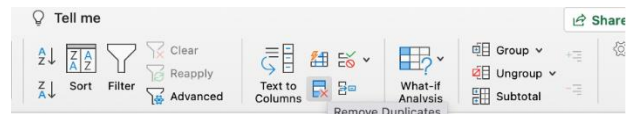


Fig. 5. Remove duplicates feature.

The duplicate data will be permanently removed if you utilize the Remove Duplicates feature (Fig. 5). Before deleting the duplicates, it's a good practice to transfer the original data to another worksheet so you don't lose any information inadvertently. It is also more convenient for users to remove all the duplicate data. It will remove all the duplicate values and remain unique.

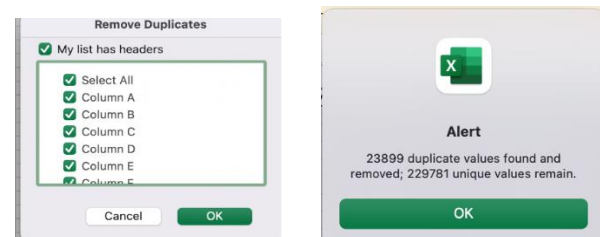


Fig. 6. Remove all duplicates.

As a result, used this method to retrieve the unique value efficiently and quickly from the specified column (Fig. 6). Besides, this function can ensure all the data quality to generate a better virtualization. Duplicating tough checks by humans is inefficient and time consuming. In the worst-case scenario, it is the result of human error.

Furthermore, RStudio is used as well for data cleaning as is an Integrated Development Environment (IDE) that provides a one-stop solution for all statistical computation and graphics. It is a powerful and simple way to engage with R programming. The RStudio is a more advanced version of R that has a multi-pane window setup that allows users to access all the important info on a single screen (such as source, console, environment & history, files, photos, graphs). RStudio was used in this project because various activities

required the use of the R language to help in the visualization and pre-processing of data. The data may be noisy, contain outliers, or simply contain inaccuracies that must be dealt with to increase completeness and compellability. R has excellent data wrangling support. Packages such as dplyr and readr can convert unstructured data into structured data. Therefore, it can improve the efficiency of data pre-processing.

- Handle missing value

```
#####Identify missing value#####
#Calculate total missing values
sum(is.na(diabetes))
mean(is.na(diabetes))#Percentage of missing values
colSums(is.na(diabetes))
```

Fig. 7. Handle missing values.

This function identifies the missing value in our dataset (Fig. 7). The sum(is.na()) function shows us how many total missing values there are in this data frame. The mean(is.na()) function displays the overall proportion of missing values in this data collection. Before deal with missing values, must first figure out which columns have missing values. Therefore, may use the colSums(is.na()) function to check.

- Replace missing value with mean

```
#Replacing the missing value with mean
meanAge <- mean(diabetes$Age, na.rm=TRUE)
meanAge
format(round(meanAge, 0), nsma11 = 0)
diabetes$Age[is.na(diabetes$Age)] <-format(round(meanAge, 1), nsma11 = 1)
```

Fig. 8. Replace missing value.

This function will be used when wants to replace the missing value with mean (Fig. 8). It'll start by calculating the mean of the value, then use the format() function to round it to the nearest integer. The is.na() function checks each value to see whether it is NA, and if it is, it replaces the NA values with mean.

- Remove missing value

```
#Remove the missing value
diabetes <- diabetes %>% drop_na(Diabetes_binary)
```

Fig. 9. Remove missing value.

This function will remove any missing values in a specific column (Fig. 9). Pipe operator (%>%) is used to connect the datasets and the select function. The drop\_na() function drops rows that have a missing value in the Diabetes binary columns.

- Remove duplicated data

```
#####Check duplicate data#####
duplicated(diabetes)
sum(duplicated(diabetes))
#delete duplicated rows
diabetes <- diabetes %>% distinct()
```

Fig. 10. Remove duplicated data function.

This function removes any duplicated rows from the data frame (Fig. 10). The sum(duplicated ()) function displays the total number of duplicated data in this data frame. The distinct() function can be used to keep just unique or distinct

rows from a data frame. Only the first row is kept if there are duplicate rows.

- Remove outliers of data

```
#####Identify the OUTLIERS#####
outvals = boxplot(diabetes$BMI, main = "Boxplot of BMI")$out
#Remove outliers
q1 <- quantile(diabetes$BMI, 0.25)
q3 <- quantile(diabetes$BMI, 0.75)
iqr <- IQR(diabetes$BMI)
diabetes <- subset(diabetes, diabetes$BMI > (q1 - 1.5 * iqr) & diabetes$BMI < (q3 + 1.5 * iqr))
```

Fig. 11. Remove outlier data.

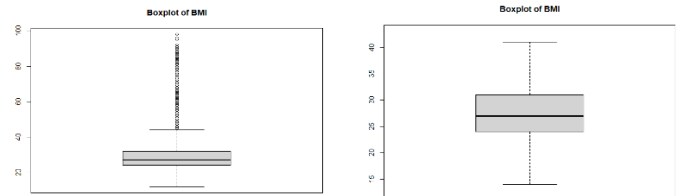


Fig. 12. Boxplot of BMI before and after removing outliers.

Box plots graphically represent the distribution of numerical data and skewness by displaying data quartiles and averages, which can be used to identify outliers within a data collection (Fig. 11 and 12). When analyzing a box plot, an outlier is defined as a data point that lies outside the whiskers of the box plot. The subset() function is then used to remove the BMI values that are more than 1.5 times the interquartile range above the upper quartile and less than 1.5 times the lower quartile (Q1 - 1.5 \* IQR) & (Q3 + 1.5 \* IQR).

- Analysis of data frame's value

```
summary(diabetes)
table(diabetes$Diabetes_012)
```

Fig. 13. Summary function.

```
> table(diabetes$Diabetes_012)
 0      1      2
190055 4629 35097
```

Fig. 14. Table function.

The summary() and table() functions are used to construct and provide the summarized result to the user (Fig. 13 and 14). It is especially handy when the user wants to know the frequency of occurrence of each value in a specific column or data frame. According to the figure above, the table() function is used to analyze the Diabetes 012 column by generating a result in terms of how frequently each value occurs.

```
> summary(diabetes)
Diabetes_012      HighBP      Highchol      cholCheck      BMI      Smoker      Stroke
Min. :0.00000 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :12.00 Min. :0.0000 Min. :0.0000
1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:1.0000 1st Qu.:24.00 1st Qu.:0.0000 1st Qu.:0.0000
Median :0.0000 Median :0.0000 Median :0.0000 Median :1.0000 Median :27.00 Median :0.0000 Median :0.0000
Mean :0.2969 Mean :0.429 Mean :0.4241 Mean :0.9027 Mean :28.38 Mean :0.4432 Mean :0.04057
3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:31.00 3rd Qu.:1.0000 3rd Qu.:0.00000
Max. :2.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :98.00 Max. :1.0000 Max. :1.00000
HEARTDISEASERATE: PhysActivity: Fruits Veggies mya1CcholConsum Anyhealthcare NoDocccost
Min. :0.00000 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000
1st Qu.:0.00000 1st Qu.:1.0000 1st Qu.:1.0000 1st Qu.:1.0000 1st Qu.:0.0000 1st Qu.:1.0000 1st Qu.:0.00000
Median :0.00000 Median :1.0000 Median :1.0000 Median :1.0000 Median :0.0000 Median :1.0000 Median :0.00000
Mean :0.09419 Mean :0.7585 Mean :0.8343 Mean :0.8114 Mean :0.0562 Mean :0.9511 Mean :0.08418
3rd Qu.:0.00000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:0.00000
Max. :1.00000 Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.00000
```

Fig. 15. Summary function result of diabetes dataset.

The summary() function can be used to generate a summary of the entire data frame by listing out its columns

and displaying the data quartiles, min, max, and average for each variable (Fig. 15).

- Read and Write function

```
#####Read Data#####  
diabetes <- read.csv("C:\\Users\\User\\Downloads\\diabetes_012_health_indicators_BRFSS2015.csv")  
#####Write Data#####  
write.csv(diabetes,"Diabetes.csv",row.names = FALSE)
```

Fig. 16. Read and write function.

These functions are used to read a csv file from the folder that was imported into RStudio and save it to a variable, as well as to export an analyzed data frame from RStudio to a csv file and save it to our computer (Fig. 16). The read.csv function in the code above attempts to read the csv file from the folder and load it into the diabetes variable as a data frame. After completing the data pre-processing, the write.csv function may be used to write the clean diabetes data frame into a csv file called "Diabetes.csv" and exported to a folder.

### C. Data Modeling

Data modelling is the act of developing a visual representation of an entire information system or certain components of it to convey relationships between various data points and organizational structures. The objective is to provide examples of the different types of data that are used and stored inside the system, their relationships, possible groupings and organizational structures, formats, and attributes.

An algorithm is a group of calculations and heuristics used in data mining (also known as machine learning) to build a model from data. The algorithm initially examines the data you submit, searching for kinds of patterns or trends before building a model. In numerous iterations, the algorithm uses the findings of this research to choose the best parameters for building the mining model. The full data collection is then subjected to these criteria to extract useful patterns and thorough statistics. In this section, the algorithm used include decision tree, support vector machine (SVM) and Naïve Bayes algorithm.

1) *Decision tree*: A supervised learning approach that works with both discrete and continuous variables is a decision tree. The dataset is divided into subgroups based on the dataset's most important attribute. The algorithms determine how this attribute is identified by the decision tree and how this splitting is carried out. The root node, which represents the most important predictor, splits into decision nodes, which are sub-nodes, and terminal or leaf nodes, which do not further split.

2) *Support Vector Machine (SVM)*: Finding a hyperplane in an N-dimensional space (N is the number of features) that categorizes the data points clearly is the goal of the support vector machine algorithm (Fig. 17). There are a variety of different hyperplanes that might be used to split the two classes of data points. Finding a plane with the greatest margin, that is, the greatest separation between data points from both classes. Maximizing the margin distance adds some support,

increasing the confidence with which future data points can be categorized.

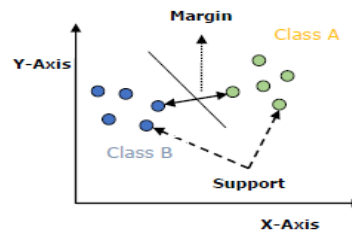


Fig. 17. SVM algorithm.

3) *Naïve Bayes algorithm*: Naïve Bayes algorithm is a classification method built on the Bayes Theorem and predicated on the idea of predictor independence. A Naive Bayes classifier, to put it simply, believes that the presence of one feature in a class has nothing to do with the presence of any other feature. The formula for Bayes' theorem is as follows (Fig. 18):

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Fig. 18. Calculation for Bayes' theorem.

- The posterior probability, or P(A|B), measures the likelihood that a given hypothesis (A) will really occur.
- P(B|A) stands for Likelihood Probability, which measures how likely it is based on the evidence at hand that a given hypothesis is correct.
- Priority probability, or P(A), is the likelihood of a theory before seeing the evidence.
- The probability of evidence is marginal probability, or P(B).

## IV. BUSINESS INTELLIGENCE ARCHITECTURE

A business intelligence architecture is the structure that an organization uses to operate business intelligence and analytics applications. It covers the information technology systems and software tools used to gather, combine, store, and analyze BI data before presenting it to high-level executives and other business users as information on daily operations and statistics. The underlying BI architecture is a critical component in implementing an effective business intelligence program that makes data analysis and reporting to assist an organization in tracking business performance, optimizing business processes, identifying new revenue opportunities, improving strategic planning, and providing better decisions overall [4].

Putting such a framework in place helps the healthcare BI team to operate in a coordinated and organized manner to construct an organizational BI solution that fulfils the data analytics requirements of its company. The BI architecture also assists BI and data managers in developing an effective method for processing and handling data that is delivered into the environment (Fig. 19).

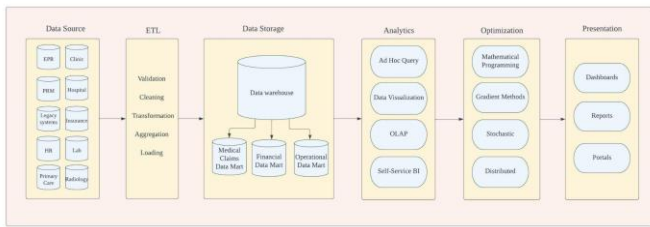


Fig. 19. Proposed BI architecture.

- Data source

These are all the sources that gather and store the data specified as important for the enterprise BI program such as EPR, PRM, radiology, insurance and so on. Secondary sources, such as patient databases from third-party information providers, might also be included. As a result, both internal and external data sources are frequently included in BI architectures. Data relevance, data validity, data quality, and the amount of information in the accessible data sets are all important factors in the data source selection process. Furthermore, to fulfil the data analysis and decision-making requirements of executives and other business users, a combination of structured, semi - structured and unstructured data types may be necessary.

- ETL (Extract, Transform and Load)

ETL is a data integration process that integrates data from numerous data sources into a single, consistent data repository that is then put into a data warehouse. ETL cleanses and organizes data using a set of business rules to fulfil business intelligence objectives, such as monthly reporting, but it may also handle more complex analytics to enhance back-end operations or end user experiences.

- Data Storage

This contains all the repositories where BI data is stored and handled. The most common is a data warehouse that holds structured data in a relational or multidimensional database and allows easy access for querying and analysis. Data warehouse can be linked to smaller data marts which are created for departments data customized to their BI requirements.

- Analytics

In the step, the focus will be on data analysis after handling, processing, and cleaning the data in the previous steps with the aid of a data warehouse. The analytics layer is a series of steps that comprise a toolbox that can be used for any form of analytics. BI application tools are used to meet the pervasive demand for successful analysis to enable organizations of all sizes to develop and earn profit. The four big data analytics techniques include classification, prediction, clustering, and association rules. Besides, a set of technologies may be implemented into a BI architecture to evaluate data and deliver information to business users such as ad hoc query, OLAP and data mining, BI. Specifically in the case of ad hoc query analysis, which allows for higher freedom, flexibility as well as usability in conducting analysis and assisting in the rapid and correct response to crucial business

problems. Moreover, the increased use of self-service BI tools allows managers and business analysts to execute queries on their own rather than depending on members of the BI team to do so. In addition, data visualization tools are also included in BI software and can be used to produce graphical representations of data such like graphs, charts, diagrams and so on to show patterns, outlier elements, and trends in datasets.

- Optimization

In general, an optimization process refers to any process that systematically proposes a better solution and results than previously used solutions. It is the practice of fine-tuning a process to optimize a collection of parameters while keeping within a series of constraints. The main purpose of process optimization is to provide more options for modifying the analytics layer's findings. The optimization block comprises a variety of approaches ranging from mathematical programming to gradient methods and stochastic to distributed.

- Presentation

This layer contains tools that present information to different users in a variety of forms. In the presentation layer, the type of technology includes dashboards, reports, and portals. All these information delivery tools allow business users to see the results and insights of BI and analytics applications for additional data analysis through built-in data visualization and the usual self-service capabilities. Executives and managers who want a broad picture of their organization's performance might use data visualization tools such like dashboards. A dashboard is a handy tool that lets users view data using graphs or charts, colored metrics as well as tables. Furthermore, users may also enable users to visualize more specific information regarding key performance indicators (KPIs) in their organizations. With the help of dashboards, they can more effectively track their progress towards setting goals. Besides, web portals refer to software that makes surfing the internet easier. By using the proposed BI architecture, they can retrieve files in file systems belonging to the Healthcare company or information given by web servers in private networks. Dashboards and web portals may both be configured to enable real-time data access with flexible views and drill-down capabilities. Reports have a more static framework for presenting data.

## V. DATAWAREHOUSE AND OLAP MODEL

The dimensional model above determined the simplest type of data warehouse schema, which is the star schema. This schema is commonly used to design or construct a data warehouse and dimensional data marts. In a data warehouse, a star schema can include one fact table and a series of interconnected dimension tables in the center. In the star schema above, the fact table includes keys to the five-dimension table such as Patient\_ID, Living\_ID, Disease\_ID, HealthCheck\_ID, Service\_ID and other attributes like BMI (Body Mass Index) and diagnostic status. The following are the list of attributes in each of the dimension table:



- Patient dimension table: Sex, Age, Education and Income
- Health service dimension table: AnyHealthcare, NoDocbcCost, CholCheck
- Health check dimension table: High BP, HighChol, GenHlth, MentalHlth, PhysHlth, DiffWalk
- Disease dimension table: Stroke, HeartDiseaseOrAttack
- Living habit dimension table: Smoking, PhysActivity, Fruits, Veggies, HeavyAlcoholConsump.

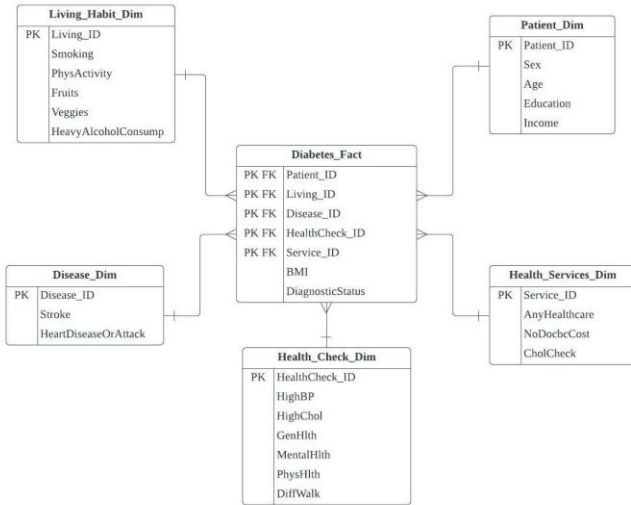


Fig. 20. Star schema.

The advantage of star schema is that all OLAP systems implement the star schema to efficiently create OLAP cubes (Fig. 20). In contrast, many OLAP systems have a ROLAP mode of operation that allows users to use a star schema as a source without having to create a cube structure. Moreover, star schema gains in query performance when compared to fully normalized schemas as it can give improvements in performance for read-only reporting applications.

### VI. DASHBOARD VISUALIZATION

There are two dashboards created for this project which are the diabetes analytics dashboard and patient analysis dashboard. By providing it with a visual context via maps or graphs, data visualization helps us understand what the information means. As a result, it is simpler to spot trends, patterns, and outliers in enormous data sets since the data is easier for the human mind to understand.

#### A. Dashboard for Diabetes Analysis

A dashboard refers to an electronic tracking tool that organizations use to present and summarize their data collection. The dashboard in Fig. 21 shows the information about the overall diabetes analytics according to the Centers for Disease Control and Prevention (CDC). CDC may use a dashboard to determine the rates of diagnosed diabetes, prediabetes, and no diabetes in the general population as well as the leading causes of diabetes. For example, it displays the

number of patients’ diagnostic statuses based on their status of smoking, stroke, heart disease or attack, blood pressure and cholesterol as well as physical activity. Moreover, the dashboard enables CDC to keep track of their patient’s health conditions, so that they can predict the risk and early signs of diabetes and provide the appropriate treatment.

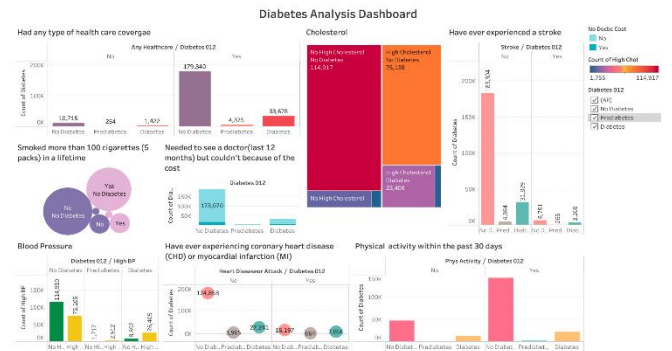


Fig. 21. Diabetes analysis dashboard.

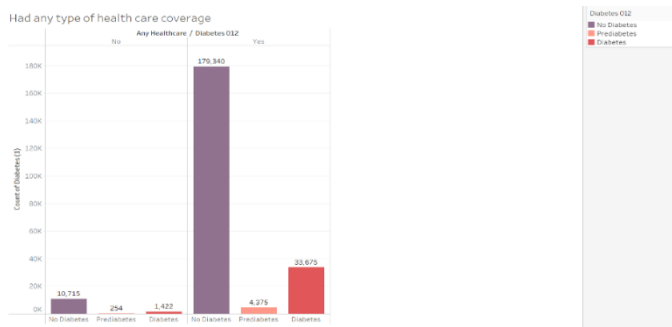


Fig. 22. Healthcare coverage on patient’s diagnostic status.

According to the Fig. 22, a side-by-side bar graph has been used to categorize and present data that results from classifying a group of things based on two or more factors. It shows the number of patients who have or have not had health care coverage based on the patient's diagnostic status (no diabetes, prediabetes, diabetes). The bar graph depicts the patient's diagnostic status using three different colors. Patients without diabetes are represented by purple, those with prediabetes by light pink, and those with diabetes by red. Based on the graph, the number of patients without diabetes with healthcare coverage is 179,340, which is dramatically more than the number of patients without diabetes with healthcare coverage, which is 10,715. As a result, can conclude that patients who have healthcare coverage, such as health insurance or prepaid plans like HMO, can help with diabetes prevention.

The stacked bar graph in Fig. 23 is used to show the count of the patients’ diagnostic status (no diabetes, prediabetes, diabetes) based on the NoDocbcCost in the last 12 months. A stacked bar graph is a type of graph which is utilized to divide and compare parts of a whole. Each bar in the graph represents a whole, and each section represents a different part or category of that whole. The various categories in the bar are represented by different colors. Light blue for no (the patient could see the doctor without worrying about the cost) and dark blue for yes (the patient could not see a doctor because of the

cost). Based on the graph, the number of patients without diabetes who could see a doctor without being concerned about the cost is 173,070, which is noticeably higher than the number of patients without diabetes who could not see a doctor due to the cost, which is 16,985. In addition to that the number of diabetic patients who could see a doctor without worrying about the cost is 31,355, which is higher than the number of diabetic patients who could not see a doctor because of the cost, which is 3,742. As a result, can conclude that most patients do not have to be worried about the cost of seeing a doctor, and that it has little impact on a patient's risk of developing diabetes.

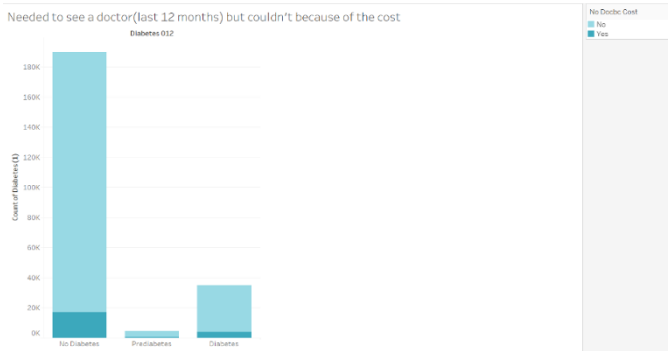


Fig. 23. Count of patients' diagnostic status based on medical fees.

183304 out of 219497, and no diabetes has the highest count within stroke which is 6751 out of 10284. Besides that, can see that the number of patients with diabetes and prediabetes who have no stroke is higher than the number of patients who have stroke, reaching 36193 respectively, compared to the number of patients who have heart disease and attack, which is 3798. To conclude, stroke is not the main reason to get diabetes because the number of patients with non-stroke is higher than the patients with stroke.

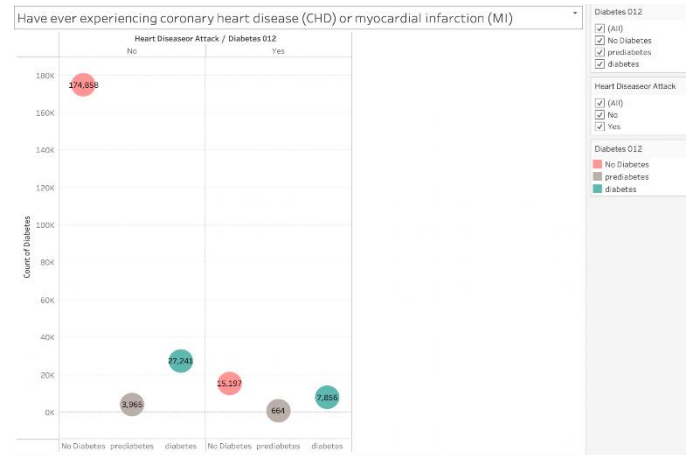


Fig. 25. Patients' diagnostic status based on heart disease.

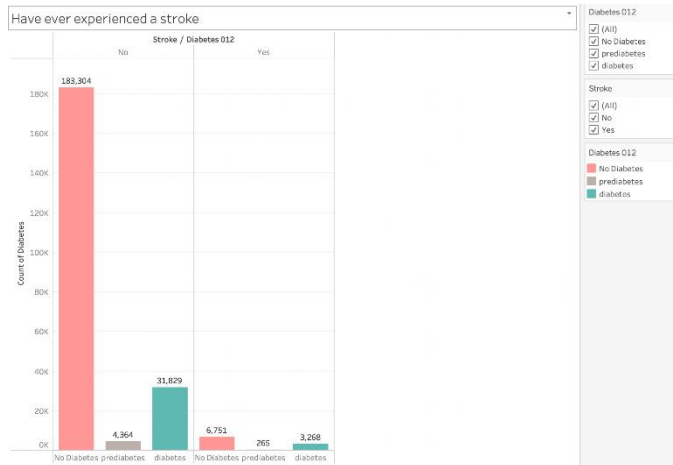


Fig. 24. Patients' diagnostic status based on stroke.

Fig. 24 represents the patients' diagnostic status which includes no diabetes, prediabetes, diabetes based on stroke and non-stroke. There are two categorical variables which are displayed using the side-by-side column graph. A side-by-side column graph is used to illustrate two category variables. The results collected when the patients are classified according to diagnostic status (no diabetes, prediabetes, diabetes) and the state of stroke (has stroke, no stroke) are displayed in the side-by-side column graph above. The bar graph uses three unique colors to indicate the patient's diagnostic status which is categorized by the diabetes status. Pink color for no diabetes, grey color for prediabetes and green color for diabetes. The filter option in the upper right corner can help you select and display only the data you want to view, making comparison and analysis easier as well. Furthermore, also observe that no diabetes has the highest count within no diabetes which is

Fig. 25 represents the patients' diagnostic status which includes no diabetes, prediabetes, and diabetes, based on whether the patient has heart disease and attack, and non-heart disease and attack. A Multi-Category Chart is useful when you have data for components that fall into multiple categories. With this, can show comparative data in more than one category. You can distinguish various states based on their varied colors using the card in the upper right corner. Based on the graph above, pink represents no diabetes, grey represents prediabetes, and green represents diabetes. It shows that the highest count inside no diabetes is 174 858 out of 206064, while the greatest count within stroke is 15197 out of 23717. Patients who get heart disease and attack with no diabetes accounted half of the total. However, can observe that the count of diabetes and prediabetes in no heart disease or attack is higher than the patients who had heart disease or attack, reaching 31377 respectively, compared to the count of those who have heart disease and attack, which is 8520. From this, it can be stated that heart disease rarely affects diabetes.



Fig. 26. Patients' diagnostic status based on physical activity.

Fig. 26 depicted the count of patients' diagnostic status (no diabetes, prediabetes, diabetes) based on physical activity and no physical activity. The two categorical variables are displayed using side-by-side column graph. A side-by-side column graph can be used to organize and present data that results from categorizing a group of people or things using two or more criteria. For example, the side-by-side column graph above displays the data obtained when patients are categorized according to diagnostic status (no diabetes, prediabetes, diabetes) and state of exercise (has physical activity, no physical activity). The bar graph uses three unique colors to indicate the patient's diagnostic status. According to the legend, the pink column represents patients who do not have diabetes, the green represents patient who have prediabetes, and yellow represent patients who have diabetes. The filter option in the right corner can assist in filtering and displaying only the data that you want to see, making comparison and analysis easier. Based on the graph, it demonstrates that the number of patients without diabetes who engage in physical activity is 143312, which is significantly higher than the number of patients without diabetes who do not engage in physical activity, which is 46743. From this, it can be assumed that physical activity can aid in the development of resistance and the prevention of diabetes.

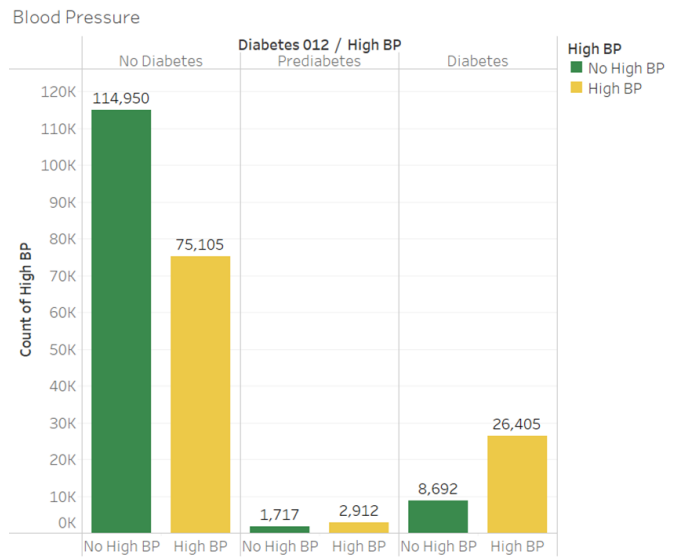


Fig. 28. High blood pressure status based on patient's diagnostic status.

Fig. 28 shows the count of high blood pressure and no high blood pressure based on the patient's diagnostic status (no diabetes, prediabetes, diabetes). A side-by-side bar graph is used to organize and present data that results from categorizing a group of things according to two or more criteria. The bar graph displays two distinct colors to depict whether the patient has blood pressure. Patients with no high blood pressure are represented by green, while those with high blood pressure are represented by yellow. From this graph, the number of patients without diabetes and having high blood pressure is 75105 which is much lower than the number of patients without diabetes and no high blood pressure at 114950. However, it depicts that for prediabetic and diabetic patients, the count of high blood pressure is relatively high, reaching 2912 and 26405 respectively, compared to the count of no high blood pressure, which is 1717 and 8692.

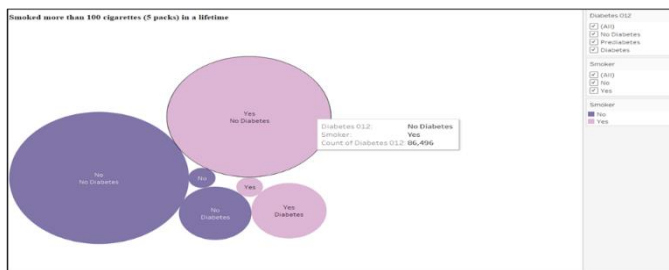


Fig. 27. Association between the diagnostic status and smoking.

In Fig. 27 above, a cluster of circles represents the association between the diagnostic status (no diabetes, prediabetes, diabetes) and smoking. To make the best use of space, the bubbles are packed in as tightly as possible. The individual bubbles are defined by the category field, and the bubble size is represented by the value field. Above, it shows at diagnostic status and smoking. While, can't control how the bubbles are arranged, but can control how big they are by putting a measure on size, in this case, and use the count of diagnostic status. The size of the bubbles represents the count of diagnostic status for various combinations of diagnostic and smoking status. This function is especially beneficial since the size of the bubble clearly shows the difference between them. The larger the circle, the larger is the proportion of the total. When a legend field is assigned to the Packed Bubble chart, the grouping mode is selected by default. It colorizes and divides the bubbles further based on the legend field. The color scheme is classified into two categories: smokers and non-smokers. According to the graph, most diabetic patients are heavy smokers (18223), while the majority of non-diabetic patients are non-smokers. Therefore, may conclude that smoking has a significant impact on causing diabetes.

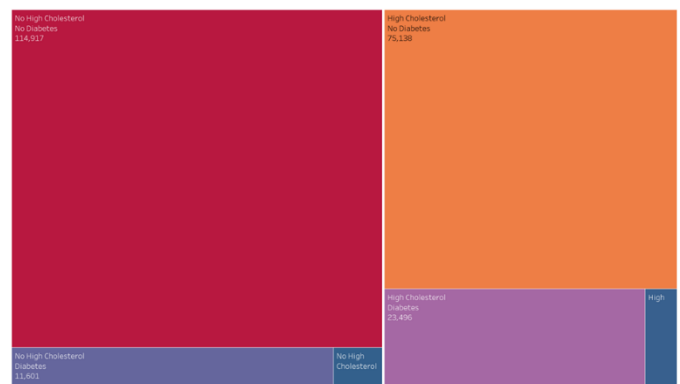


Fig. 29. High cholesterol status based on patient's diagnostic status.

As depicted in Fig. 29 above, it shows high cholesterol and non-cholesterol counts according to the patient's diagnostic status. The tree map is a visual representation consisting of nested rectangles. These rectangles indicate different categories within a given dimension and are arranged in a tree-like hierarchy. The tree map chart displays various distinct colors to distinguish between high cholesterol and non-

cholesterol in different diagnostic statuses. When looking for insights in a tree map, the largest box represents the largest portion of the entire, while the smallest box represents the smallest portion.

Fig. 30 below depicts the total number of no high cholesterol and high cholesterol for each diagnostic status such as no diabetes, prediabetes, and diabetes. From this table, we can analyze whether high cholesterol has a strong effect on causing diabetes.

High Chol	No Diab..	Prediab..	Diabetes
No High Cholesterol	114,917	1,755	11,601
High Cholesterol	75,138	2,874	23,496

Fig. 30. Total number of high cholesterol status for each diagnostic status.

### B. Dashboard for Patient Analysis

The spreadsheets and dashboards shown below are mainly focused on patient-related data, starting from the patient's gender, age group, education level and their income. The purpose is to obtain a broad perspective of patient's personal information and acquire relevant knowledge for Centers for Disease Control and Prevention (CDC) management to analyze whether these personal factors will cause diabetes and may use that as a reference to make a better decision for prevention.

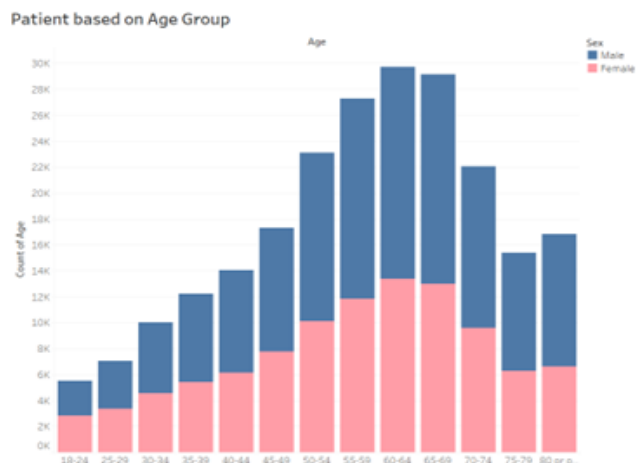


Fig. 31. Patient based on age group.

In Fig. 31 above represented the horizontal bar of patient based on their age group separated by gender. The stacked bar graph can differentiate the gender based on their distinct colors, which is described at the card on the top right corner. If click one of the genders in the horizontal bar chart, it will direct us to another worksheet which is specifically filter out based on the gender have chosen by using filter action.

The figures are the age group filtered out by gender (Fig. 32). The left figure represents the age group by male while the right figure represents the age group by female. Tableau action filters enable us to transfer information between worksheets (Fig. 33). Generally, when you choose marks from one worksheet, that information is sent to other worksheets, resulting in the display of the relevant information. Behind the scenes, action filters deliver data

values from appropriate source fields to the destination worksheet as filters. By implementing this function, the CDC management will be able to identify and focus the age groups based on gender.

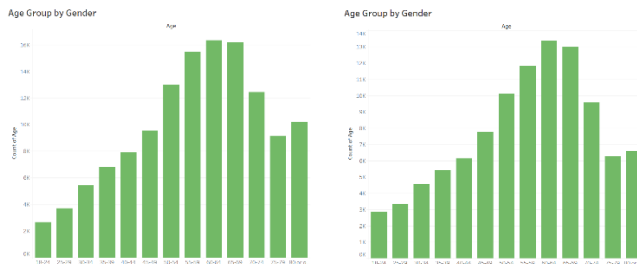


Fig. 32. Total number of high cholesterol status for each diagnostic status.

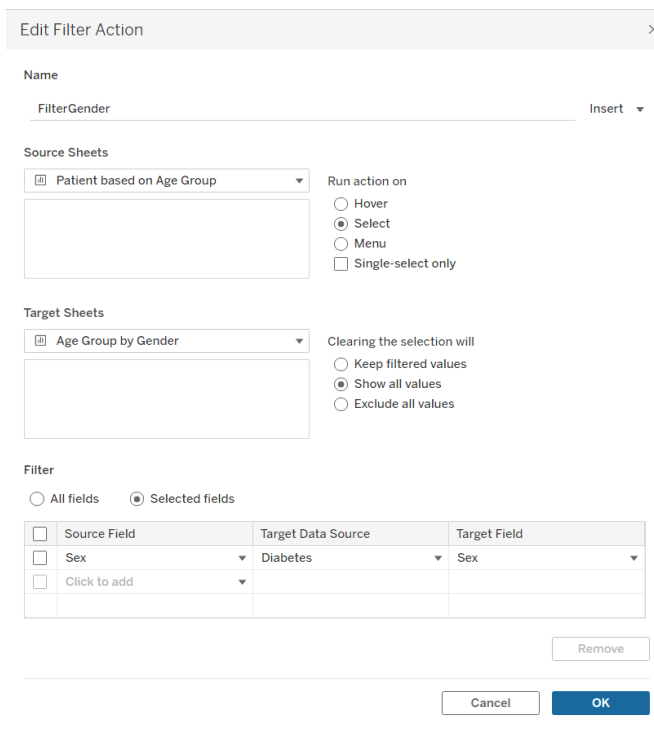


Fig. 33. Filter action in tableau.

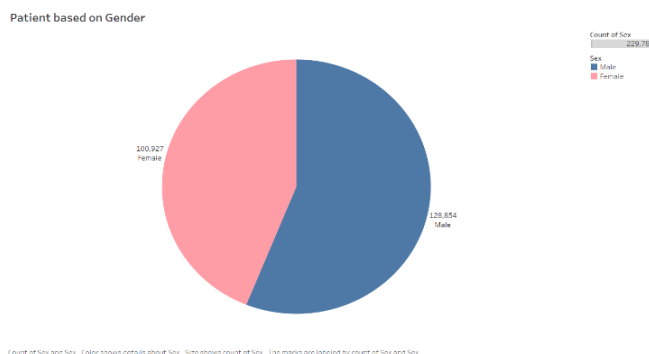


Fig. 34. Patient based on gender.

The pie chart above (Fig. 34) shows the number of patients based on gender. The pie chart displays two different colors, each representing a gender. Male is represented by blue, while female is represented by pink. This provides a general

estimate of the proportion of male and female customer patient collected from the database. The figure can clearly find that male has a greater number than female, which male has several 128,854 and 100,927 for females.

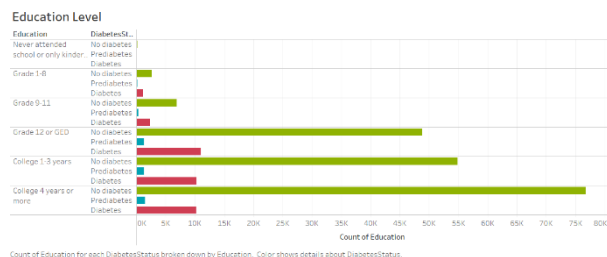


Fig. 35. Education level.

Fig. 35 above determines the highest education level patient had taken. There are six level of education which are never attended school or only kindergarten, Grade 1 through 8, Grade 9 through 11, Grade 12 or General Education Diploma (GED), College 1 year to 3 years, College 4 years or more. It basically shows whether the level of education play a role in diabetes. For instance, the number of patients whose highest education level is 4 years college or more and does not diagnose with diabetes are 76,746 people. Studies shows that education levels may increase the adoption of health behaviors such as sufficient eating and medication adherence. As a result, it is likely that education levels work as a fundamental cause of disease by using resources such as knowledge, which have a substantial effect on people's ability to decrease risks that may prevent or delay diabetes or better treat the disease once it comes.

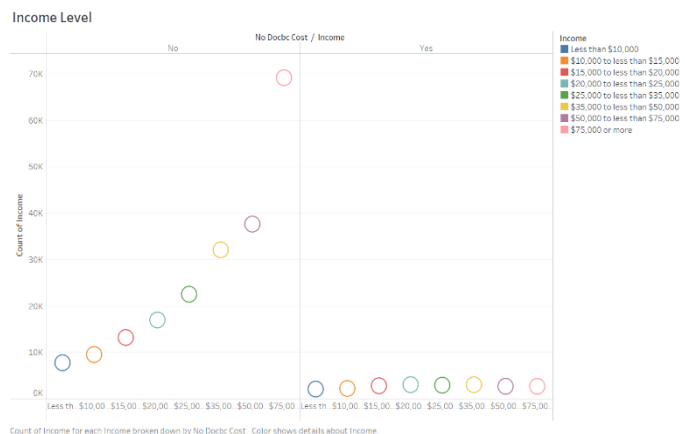


Fig. 36. Income level.

Fig. 36 above determines the whether the patient has a time in the past 12 months when needed to see a doctor but could not be due to the cost of treatment based on their income level. Studies shows that people who have low income likely to have a higher risk of diagnose with diabetes. It can find there is a big different between the side-by-side circle visualization. As a result, this can be assumed that most patients who have an income of \$75,000 or more do not have to worry about the treatment cost, that is, 69,117 people (see Fig. 37).

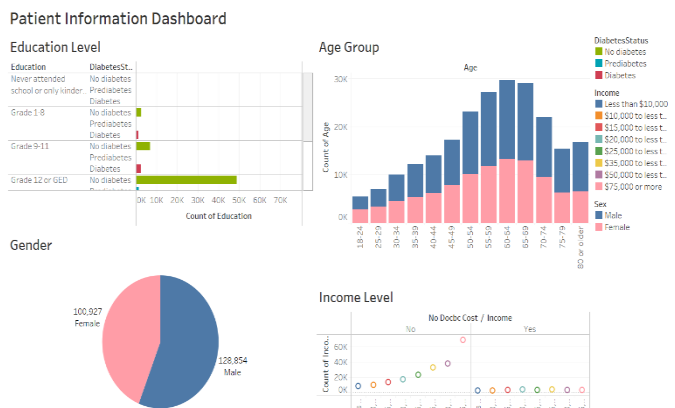


Fig. 37. Patient analysis dashboard.

### VII. CONCLUSION

The goal of Business Intelligence is to assist and enable better business decisions. BI provides organizations with access to information that is crucial to the success of a variety of areas and departments. Effectively integrating BI will give the organization with more actionable data, valuable insights into market trends, and a more strategically orientated decision-making approach. In this project, the author has investigated Centers for Disease Control (CDC) to create a BI solution for them to make better decision after visualization the insight from the dataset.

Before visualizing the data, author must do some in-depth research on the domain background, their problem, and opportunities, create a project charter, data identification, designing BI logical architecture as well as data warehouse model. Once all of this has been done, the author will start building dashboards that can develop insights to the executive and assist them in making a better decision by using Tableau. Tableau is a powerful tool for quickly creating interactive data visualizations and user-friendly. The author has created 2 dashboards which contain patient information analysis and health information analysis.

Finally, BI solution is important as it aids in the production of reliable reports by collecting data directly from the data source. Today's BI solutions minimize the time-consuming effort of manually aggregating data since BI technologies provide up-to-date data, executives can monitor firms in real time. As a result, the author thinks that the project's results might be improved further for developing a best-fit BI solution and data visualization for CDC. With the enhancement and improvement, it will undoubtedly assist CDC in striving for greater opportunities in terms of patients, health and so on.

### REFERENCES

- [1] "Benefits of Switching to an Electronic Health Record (EHR)," *Practice Fusion*, Nov. 22, 2016. <https://www.practicefusion.com/health-informatics-practical-guide-page-1> (accessed Sep. 21, 2022).
- [2] "4 Health Care Data Challenges and How to Overcome Them," *Corporate Compliance Insights*, Jun. 28, 2018. <https://www.corporatecomplianceinsights.com/4-health-care-data-challenges-overcome> (accessed Sep. 21, 2022).

- [3] “NHANES - About the National Health and Nutrition Examination Survey,” *Centers for Disease Control and Prevention*, 2019. [https://www.cdc.gov/nchs/nhanes/about\\_nhanes.htm](https://www.cdc.gov/nchs/nhanes/about_nhanes.htm)
- [4] “What is Business Intelligence Architecture (BI Architecture)?,” *SearchBusinessAnalytics*. <https://www.techtarget.com/searchbusinessanalytics/definition/business-intelligence-architecture>
- [5] A. Menon, “What is Agile Business Intelligence? - A Guide,” *Techfunnel*, Aug. 08, 2021. <https://www.techfunnel.com/information-technology/agile-business-intelligence> (accessed Sep. 21, 2022).
- [6] JavaTpoint, “Agile Model - javatpoint,” *www.javatpoint.com*, 2011. <https://www.javatpoint.com/software-engineering-agile-model>
- [7] Wikipedia Contributors, “Centers for Disease Control and Prevention,” *Wikipedia*, Dec. 09, 2018. [https://en.wikipedia.org/wiki/Centers\\_for\\_Disease\\_Control\\_and\\_Prevention](https://en.wikipedia.org/wiki/Centers_for_Disease_Control_and_Prevention)
- [8] “Business Intelligence in Healthcare,” *Villanovau.com*, 2022. <https://www.villanovau.com/resources/bi/business-intelligence-in-healthcare/#:~:text=By%20providing%20a%20foundation%20for> (accessed Sep. 21, 2022).
- [9] HealthIT.gov, “What are the advantages of electronic health records?,” *HealthIT.gov*, Mar. 08, 2022. <https://www.healthit.gov/faq/what-are-advantages-electronic-health-records>
- [10] D. Dugar, “Benefits of Electronic Health Records | EHR Advantages & Disadvantages,” Sep. 19, 2021. <https://www.selecthub.com/medical-software/benefits-of-ehr-systems> (accessed Sep. 21, 2022).
- [11] “What you need to know about BI dashboards,” *Tableau Software*. <https://www.tableau.com/learn/articles/business-intelligence/bi-dashboards>
- [12] University of Massachusetts Dartmouth, “Decision-making process - UMass Dartmouth,” *Umassd.edu*, 2021. <https://www.umassd.edu/fycm/decision-making/process/>
- [13] A. B. I. BV, “Where do Clinical Evaluation and Clinical Investigation meet?,” *www.qservegroup.com*. <https://www.qservegroup.com/eu/de/i319/where-do-clinical-evaluation-and-clinical-investigation-meet>
- [14] Heather, “7 Major Challenges Facing the Healthcare Industry in 2021,” *MailMyStatements*, Oct. 27, 2020. <https://mailmystatements.com/2020/10/27/2019challenges/>
- [15] R. College, “4 Major Challenges Facing Today’s Health Care Industry,” *Regis College Online*, Mar. 10, 2018. <https://online.regiscollege.edu/blog/4-challenges-facing-the-health-care-industry/>