IDENTIFICATION OF TRENDS IN RACIAL/ETHNIC DISPARITIES OF PATIENTS WITH DIABTES MELLITUS USING WEB-BASED TABLEAU DASHBOARD- A CROSS SECTIONAL STUDY

A Thesis presented to

the Faculty of the Graduate School

at the University of Missouri

In Partial Fulfilment

Of the Requirements for the Degree

Master of Science

By

VISHWA BHAYANI

Dr. IRIS ZACHARY, Thesis Supervisor

DECEMBER 2021

The undersigned appointed by the Dean of the Graduate School, have examined the dissertation entitled:

IDENTIFICATION OF TRENDS IN RACIAL/ETHNIC DISPARITIES OF PATIENTS WITH DIABETES MELLITUS USING WEB-BASED TABLEAU DASHBOARD- A CROSS SECTIONAL STUDY

presented by Vishwa Bhayani,

a candidate for the degree of Master of Science and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Iris Zachary

Dr. Sue Boren

Dr. Uzma Khan

DEDICATION

I dedicate this thesis to my family, my fiancée, and my professor Dr. Iris Zachary for supporting me and believing in me.

ACKNOWLEDGEMENTS

This study would not have been possible with the constant support of my advisor Dr. Iris Zachry for standing with me at every step throughout my academic program. Special thanks to Dr. Uzma Khan for taking immense interest in my thesis and Dr. Sue Boren for making sure this study was going on a right track.

Contents

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Many ways of creating data visualizations using the Business Intelligence tools have been implemented and new ways are continuously evolving. In this project, I have used Tableau Public software to analyze the trends of Race and Ethnic disparities in Diabetes Mellites patients by creating individual dashboards. A desktop application was created using the web programming languages like HTML and CSS to create more interactive and application embedded dashboard. The dataset selected was Health Facts® database, which is contributed of day-to-day patient care encounters. Data preparation and data pre-processing and data visualization were the two major steps to get the required results. In Data pre-processing, data mining tools such as SQL and Python were used to create the final dataset with 7 variables. An infographic was created with a dashboard with three views embedded into web pages. These three dashboard views were created based their specific information like one dashboard view showing the insights about the distribution of Hispanic with Diabetes Mellitus, another dashboard view showing the trends of African American and Caucasian races and the third view showing the distribution of patients with Diabetes Mellitus and specific cancers based on race and ethnicity. All the data obtained were analyzed and compared with the CDC data and the results confirmed. Moreover, to measure the usability of the desktop application, SUS (System Usability Score) was used. The result of the application clearly indicated the way such application can be used strategize the programs and interventions to reduce the disparities for Diabetes Mellitus amongst the African American population.

CHAPTER 1: INTRODUCTION

1.1. DIABETES MELLITUS INTRODUCTION

The medical definition of diabetes is "a chronic disease associated with abnormally high levels of sugar or glucose in the blood"[1]. There are three types of diabetes based on their cause: Type 1 Diabetes Mellitus, Type 2 Diabetes Mellitus and Gestational diabetes. Type 1 Diabetes Mellitus is caused due to the inadequate production of the insulin. Type 2 Diabetes Mellitus is caused due to inadequate sensitivity of cells to action of insulin. Gestational diabetes is the result of an increased in the blood glucose level for the first time during the pregnancy.

Insulin is a peptide hormone secreted by the β cells of the pancreatic islets of Langerhans and maintains normal blood glucose levels by facilitating cellular glucose uptake, regulating carbohydrate, lipid and protein metabolism and promoting cell division and growth through its mitogenic effects. [19] Glucose acts as an energy to the cells of the body. Insulin facilitates the absorption of the glucose in the cells of muscles, fat and liver. Diabetes (increase in the blood sugar) occurs when the body either does not secrete enough insulin or when the body cannot use the insulin that has secreted effectively.

Type 1 Diabetes Mellitus is called early-onset diabetes as it is caused at a very young age and Type 2 Diabetes Mellitus is called late-onset diabetes as it is caused at a latter age of life. The common symptoms seen in the patients of diabetes, despite of its type are excess thirst, excess eating, and frequent urination. Diabetes mellitus is also associated with various complications such as heart disease, stroke, amputations, blindness, and kidney diseases.

1.2. SIGNIFICANCE OF DIABETES

According to the National Conference of state legislatures, Diabetes Mellitus is the seventh leading cause of death in U.S[18]. Thecost to treat an individual with diabetes is more than 200 percent higher than the cost to treat a patient without diabetes. According to the American Diabetes Association, people with diagnosed diabetes incur an average medical expenditure of $16,752 per year, of which about $9,601 is attributed to diabetes. On average, people with diagnosed diabetes have medical expenditures approximately 2.3 time higher than what expenditure would be in the absence of diabetes[21]. According to the National Diabetes Statistics Report 2020[22], the crude estimates for population with diabetes mellitus in 2018 were:

• 26.9 million people of all ages—or 8.2% of the US population—had diagnosed diabetes.

• 210,000 children and adolescents younger than age 20 years—or 25 per 10,000 US youths—had diagnosed diabetes. This includes 187,000 with type 1 diabetes.

• 1.4 million adults aged 20 years or older—or 5.2% of all US adults with diagnosed diabetes—reported both having type 1 diabetes and using insulin.

• 2.9 million adults aged 20 years or older—or 10.9

1.3. RACIAL/ETHNIC DISPARITIES IN DIABETES MELLITUS

Diabetes Mellitus occurs due to the demographic inheritance and environmental factors like age, stress, obesity, diet, sleep, and lack of physical activity. People of certain race and ethnicity predominated the genes for diabetes and the predisposition carries forward to the next generation. The environmental facts like unhealthy diet, stress and age will trigger it to cause diabetes. Racial and ethnic minorities have significantly higher rates of diabetes-related complications. For example, African American are 2-4 times more prone to renal disease, blindness, amputations, and amputation-related mortality than non-Hispanic whites with diabetes.

According to the CDC National Diabetes Statistics Report 2020 the key findings among the US adults aged 18 years or older, age-adjusted data for 2017-2018 indicate the following[22]:

- Prevalence of diagnosed diabetes was highest among American Indians/Alaska Natives (14.7%), people of Hispanic origin (12.5%), and non-Hispanic blacks (11.7%), followed by non-Hispanic Asians (9.2%) and non-Hispanic whites (7.5%)
- American Indians/Alaska Natives had the highest prevalence of diagnosed diabetes for women (14.8%)
- American Indian/Alaska Native men had a significantly higher prevalence of diagnosed diabetes (14.5%) than non-Hispanic black (11.4%), non-Hispanic Asian (10.0%), and non-Hispanic white (8.6%) men
- Among adults of Hispanic origin, Mexicans (14.4%) and Puerto Ricans (12.4%) had the highest prevalence, followed by Central/South Americans (8.3%) and Cubans (6.5%)
- Among non-Hispanic Asians, Asian Indians (12.6%) and Filipinos (10.4%) had the highest prevalence's, followed by Chinese (5.6%). Other Asian groups had a prevalence of 9.9%
- Among adults, prevalence varied significantly by education level, which is an indicator of socioeconomic status. Specifically, 13.3% of adults with less than a high school education had diagnosed diabetes versus 9.7% of those with a high school education and 7.5% of those with more than a high school education.

Racial and ethnicity also play an important role in the seeking of diabetes care. Difference in the health status or access to healthcare among the racial, ethnic, geographic, and socioeconomic groups are referred to as health disparities. Racial and ethnic minorities bear a disproportionate burden of the diabetes epidemic. Minorities have higher rates of serious health conditions that affect communities of color [synonym of disproportionate]. Certain race/ethnicity minorities are also not privileged to have a health insurance coverage to seek care and treatment for the diabetes and its related complications. Despite of high prevalence of the condition, minorities experience lower quality of care, and great barriers to self-management compared to Caucasian patients. These disparities can result in the shorter lifespans and lower quality of life, with lack of proper resources such as diet, education, employment, safe and healthy neighborhood.

1.4. PROBLEM STATEMENT

Racial and ethnic disparities for diabetes mellitus is commonly seen, but the ways to represent them in the form of user-friendly infographics and visualization is very limited. Some are due to

the missing and disoriented data, and some are due to the limited availability of visualization of health disparities.

1.5. AIM

The aim of this project is to analyze the demographics of the diabetes patient's data from a large database using data mining tools and create and a web-based visualization focusing on clarity and usability features.

CHAPTER 2: USEFULNESS OF THE TABLEAU DASHBOARD AS A VISUALIZATION TOOL IN HEALTHCARE

2.1. INTRODUCTION

Data visualization is booming in the health sector in today's era where the industry uses dashboards and a variety of visualization tools to facilitate the understanding of the complex and dynamic problems. It supports health informatics by addressing specific needs. Visualization is offered in many ways such as single graph, charts, animations, dashboards, maps and so on. One such tool that performs the visual analytics and gives the interactive results is Tableau software. Tableau is a software that can help users explore and understand their data by creating interactive visualizations. The advantages are that it can be used on any database, and it is as easy as dragging and dropping the needed rows and columns for creating an interactive visualization.

Although Tableau is getting its name in health care, it is still not widely used in the healthcare sector. This review is created to show the use of tableau dashboard in the healthcare sector. Previous systematic reviews were carried out by Auliya, R. S., 2018[17] where they aimed to identify trends, issues, methods, and frameworks in the development of healthcare dashboards but, this review is different in terms that qualitative analysis of the studies that are done based on the dataset, purpose and their used of real time data and web-based Tableau dashboard. Previously no specific study was done with focus only on Tableau dashboard. This study is focused on the studies that only used Tableau dashboard for data visualization.

2.2. METHODOLOGY

Data source and method

Two databases were systematically searched to identify relevant publications: MEDLINE and PubMed. Included were studies that evaluate and review applications that were published between 2017 to 2021. The search terms consisted of four constructs. The first construct was related to the Tableau, the second construct was related to the Dashboard. The third construct was based public health. The final construct was on the search limit of the above three constructs. The search strategy was limited to English research that were on health-related data and visualization with tableau tool, published between 2017 till 2021.

Search Terms used were Data visualization, Tableau, Epidemiology, Public health, Demographic, Dashboard.

Inclusion criteria

Included studies (N=11) have used the Tableau tool for visualization purpose for any healthcare data.

Exclusion criteria

Excluded studies included those studies who have just mentioned the importance of visualization with Tableau tool without using any health-related data or have mentioned about the explanation of the visual basic through Tableau.

Quality Assessment

To assess quality and bias of the included studies, the purpose of the use of the Tableau, category of its use and its use as Web-based tool along with type of data used were analyzed.

The following information was extracted from the included articles:

Category of Tableau: Studies were classified based on the type of category Tableau was used in their study: monitoring, evaluation, analysis.

Purpose of the study: Describes various purposes related to visualizing the data to explore, evaluated, compare, and predict the health-related data.

Data: Includes the specific disease targeted or the dataset used to visualize

Web-Based category: Studies are categorized into using the visualization as a Web-based portal.

Real-time visualization category: Studies are categorized into taking the real-time data while visualizing the Tableau dashboard.

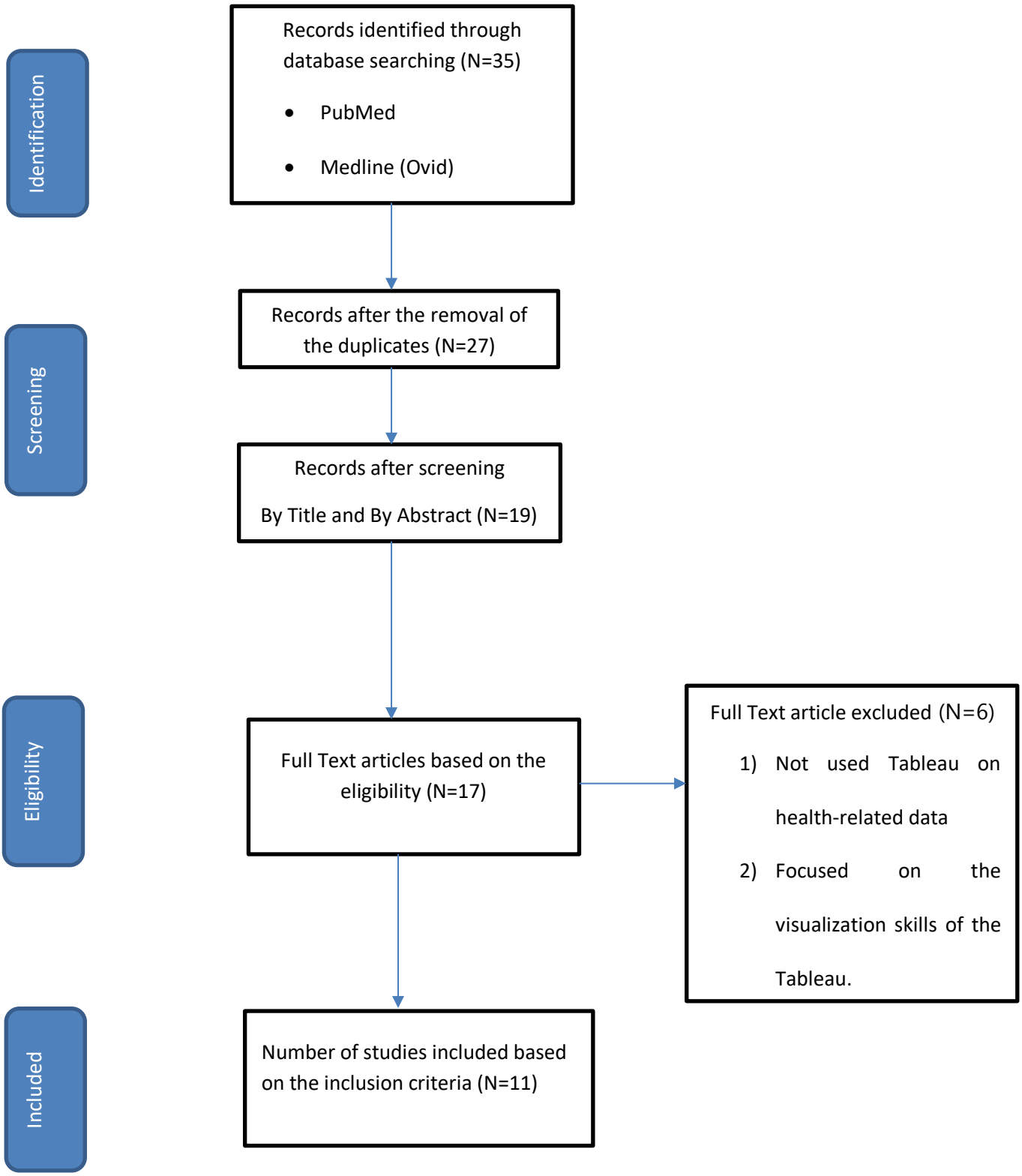


Figure 1 Flow chart showing selection of articles

2.3. RESULTS

General Characteristics

The database search using the key search terms resulted in total 35 articles. After reviewing the titles and abstracts, 16 studies were excluded. The titles of remaining 19 studies were reviewed, and 2 studies were excluded based on the inclusion criteria and redundancy. After examination, 11 studies (see Table 1) fully met the inclusion criteria and were included in the final review. The 11 included studies were published between 2017 and 2021. All these studies are in English language.

The category of Tableau was divided into the three: evaluation, analysis and monitoring. An "analysis" is an examination of the elements or structure of something, as a basis for discussion or interpretation. An "evaluation" is your conclusion about a source, based on evidence as to what you hold to be most important or effective, and "monitoring" is the systematic process of collecting, analyzing, and using information to track a program's progress (as shown in Figure 1) towards reaching its objectives and to guide management. Out of 11 studies, N=2 have used Tableau dashboard for the evaluation purpose. N=7 have used Tableau for analysis purpose and N=1 have used for monitoring purpose. One study has use Tableau for both, monitoring and evaluation purposes.

Table 1 Based on inclusion criteria and category of visualization

Author, year	**Category**
William Martinez et al, 2018[11]	Evaluation
Hamish ROBERTSON, 2017[15]	Analysis
Olga Strachna MS, 2021[16]	Monitoring
Luna Khirfan, 2020[9]	Analysis
Timothy C. Huber, Arun Krishnaraj, 2018	Analysis
Benjamin A. Goldstein, PhD, MPH; 2020[7]	Evaluation
Sundas Khan, 2019[8]	Monitoring and Evaluation
Inseak Ko, 2018[10]	Analysis
Ibrahim Dadari, 2021[4]	Analysis
Dancy Scott[5]	Analysis
Yeonkyeong Park, 2020[13]	Analysis

Evaluation Methods:

One of the key objectives of this review was to find out the purpose of use of Tableau dashboard. Tableau connects different data sources and helps the users to create a variety of data visualization. The purpose of the visualization varies from study to study and also depends on the data set (Table 2). Yeonkyeionj [(13)] and Robertson [(15)] have used Tableau visualization for creating the spatial pattern for an air pollution and Alzheimer's disease, respectively. Martinez[11] has used to implement the

sprint methodology in the diabetes dashboard that is already embedded in the patient portal. Luna[9], Dadari[4] have use Tableau to show the mappings of the geographical locations. Analyzing the statistics and machine learning through Tableau is also an advantage. Benjamin has used tableau to show the real-time data of the predicted values obtained from the random forest (machine learning algorithm). Visualizing the frequency can be made easy with tableau by creating the line graphs and word clouds. Dancy-Scott[5] did a similar study to understand the frequency of the HIV related terminologies by visualizing its frequency and word clouds on Tableau. Moreover, Tableau data can be a good visualization tool to monitor a large volume of patient data. In the study by Sundas[8] and Ola[16], monitoring the real time data was carried out using Tableau visualization.

In addition to that, predicting the trends based on the insights created in the Tableau is useful in strategic planning and implementation of health policies. The study by Inseak[8] have created dashboard to visualize trends of the analyzed prescription data from National claims dataset.

Table 2. Based on the purpose of the study and data used

Author, year	**Purpose of the Study**	**Data Used**
William Martinez et al, 2018[11]	To apply design sprint methodology paired with mixed-methods, task-based usability. testing to design and evaluate an innovative, patient-facing diabetes dashboard embedded in an existing patient portal and integrated. into an electronic health record.	Diabetes
Hamish ROBERTSON, 2017[15]	To explore how the issues associated with the dementias can be better understood and responded to using visual data environments.	Alzheimer's disease
Olga Strachna MS, 2021[16]	To develop an electronic PROs (ePROs) program that meets a range of clinical needs across a head and neck multidisciplinary disease management team.	Head and Neck Cancer
Luna Khirfan, 2020[9]	To identify different tools for data analysis	DayLightening case
Timothy C. Huber, Arun Krishnaraj, 2018	To better visualize and analyze trends in this data, an interactive data visualization dashboard was created using a commercially available data visualization platform.	Radiology data (CDS)
Benjamin A. Goldstein, PhD, MPH; 2020[7]	To evaluate the development and performance of a clinical decision support tool to inform resource utilization for elective procedures.	Coronavirus data
Sundas Khan, 2019[8]	We aimed to monitor and evaluate the adoption and trigger rates of the tool and assess whether ongoing tool modifications would improve adoption rates.	CDS
Inseak Ko, 2018[10]	This study suggests the procedures of effective and efficient visualization of big data for general healthcare researchers	Korean National Insurance Claims data

Ibrahim Dadari, 2021[4]	To summarize the first mapping of the implementation and report strategies from Gavi- supported counties country reports.	Immunization records (Gavi Joint Appraisals (JAs))
Dancy Scott[5]	To visualize the frequency metrics associated with the terms as line graphs and word clouds	HIV termiology
Yeonkyeong Park, 2020[13]	To introduce an example of easy and e_ective web-based visualization of research findings, relying on predicted concentrations of particulate matter _ 10 _g/m3. (PM10) and nitrogen dioxide (NO2) obtained from our previous study in South Korea and Tableau software.	Air Pollution data

As mentioned earlier, Tableau can be implemented on a variety of health care data (Table 2). N=6 [(11,15,16,7,4)] studies have used data related to disease, N=2[(13,9)] has used the health data related to Environmental hazards, N=1[(5)] has used the data with health-related terminology, N=1[(10)] study has used the Insurance claims data and N=1[(8)] has used the electronic health record data (CDS tool).

Table 3. Based on the implementation as Web-based and using real-time data in dashboard

Author, year	**Web-Based dashboard**	**Use Real-time data**
William Martinez et al, 2018[11]	Yes	No
Hamish ROBERTSON, 2017[15]	No	No
Olga Strachna MS, 2021[16]	No	No
Luna Khirfan, 2020[9]	No	No
Timothy C. Huber, Arun Krishnaraj, 2018	No	Yes
Benjamin A. Goldstein, PhD, MPH; 2020[7]	No	Yes
Sundas Khan, 2019[8]	No	Yes
Inseak Ko, 2018[10]	No	No
Ibrahim Dadari, 2021[4]	Yes	Yes
Dancy Scott[5]	No	No
Yeonkyeong Park, 2020[13]	Yes	No

Apart from the type of dataset, how frequently the data is updated is also important in healthcare because it helps in creating an awareness and the data are updated regularly. Real-time data is immediately delivered after collection. They are useful in making the strategic business decision. Table 3 shows the studies that have used real-time data in their visualization. Few (N=4) of the studies have implemented the real-time data in their Tableau dashboard. Out of which N=2 have used the real-time data to evaluate the performance of the tool, whereas N=1 have used for the strategical purposes. N=1 have used to access the impact of the decision support tool.

Another striking feature of using Tableau visualization is using the same dashboard created on Web and creating as web-based dashboard (Table 3). So, the dashboard will work dynamically where the

users can drill out more information about a particular piece of data. Out of 11 studies, N=3 have used web-based Tableau dashboard. N=2 have used the mapping strategies showing a particular geographical regain using the web-based dashboard and one study has used the diabetes patient information from the EHR and created a dashboard embedded in a web portal.

2.4. DISCUSSION

Health care data is very complex. Especially when it comes to making important decisions about patient health and resources focusing on population health by creating a data-driven decision approach is important. The best way to communicate with the general population and the healthcare experts is through visualization of the data. Visualizing the health-related data will allow the experts to present the key trends via groups, charts, animation and many such visualization. A healthcare dashboard is a modern analytic tool that is used to analyze the trends, evaluate the strategies, or/and monitor the program in a dynamic and interactive way. A single dashboard contains the many visuals that are arranged in such an away that the whole dashboard will provide with different insights about the project/program and necessary arguments, or decisions are made. There are several data visualization tools in the market. Tableau is one such interactive data visualization tool that is fastest growing and powerful in the Business Intelligence industry. Tableau can be linked with a variety of data sources and can create beautiful insights to visualize the trends. The literature review used 11 studies that have used Tableau as a visualization tool to display the problems related to healthcare. All the studies have used Tableau for different purposes like analyzing the trends, visualizing the geographic location, predicting the results, counting the metrics, and evaluating the project. Many of these studies have implemented the web-based dashboard to help the user get the specific data in detail. This aids the non-technical users to gather the information about a particular source and create the timeline of the collected data. Many studies have tried to incorporate the real-time data in the dashboard that will change the pattern when the data is updated. This will help in identifying the pin-point issues that will arise now and can give the insights soon and help in creating the up-to-data data sights.

Limitation

There are few limitations to this study:

1) Type of the Tableau account: To create a dashboard in Tableau, a Tableau account is required. These studies did not mention about the type of the Tableau used. As there are many versions of Tableau present like Tableau premium(paid) and Tableau Public (free). Paid version is quick costly and free version do not give all the facilities that are available in the free version.
2) No mention regarding the security of the data. When the data is linked with the Tableau, it will be projected on the browser. There is no mention about how the data is secured on the web platform.

2.5. CONCLUSION

Visualization is a useful way to understand the impact of various diseases on racial groups and help reach more nuanced understanding of those diseases amongst various communities. Visualization helps in solving health disparities by allowing users to absorb data and visualizing new ways to solve

the problem. Thus, it aids in positive decision making. Tableau is a one of the best tools for data visualization in the health care sector which can be used for a variety of purposes and can be linked to any data related to healthcare. Its GIU (Graphical user Interface) can create a creative interactive dashboard that is compatible even for the mobile interface and quite easy to implement. It has a capacity to create a large unorganized data into real insights which drives decision making.

CHAPTER 3: METHOD

3.1 DATA SOURCE

The data used was CERNER Health Facts®data. CERNER Health Facts® is a Longituninal electronic health record (EHR) patient data which captures clinical records with time-stamped and ordered information on pharmacy, laboratory, admission and billion data from all designated locations (including Cerner and non-Cerner participating facilities). Cerner Health Facts is a HIPAA- compliant database collected from participating clinical facilities. It is collected from 90 health systems and over 600 health facilities. Cerner Health Facts database captures and stores de-identified, longitudinal electronic health records (EHR) patient data, and aggregates and organizes these data into consumable datasets to facilitate analysis and reporting. The data are generated from Cerner and non-Cerner participating contributing facilities and go back as far as 2000. Cerner Health Facts focuses on providing the information regarding the five health outcomes: clinical, economic, process, functional and satisfaction. Specifically, the database includes the data on demographics, encounters, diagnosis, prescriptions, procedures, laboratory tests, locations of services/patients (e.g., clinic, ED, ICU), hospital information and billing. It is designed to track drugs or device's usage across diagnosis and major procedures, as well as by geographic region and hospital type.

Currently, Cerner Health Facts®[23] contains data from:

- Over 158,300,000 patients
- Over 1.3 billion laboratory results
- Over 84 million acute admissions, emergency and ambulatory visits
- Over nine years of detailed pharmacy, laboratory, billing and registration data
- More than 151 million orders for nearly 4,500 drugs by name and brand, and
- 100% of Patients in Orchid, Keck Care and KIDS

The data set was obtained by requesting from the Office of Research at University of Missouri-Columbia. Health Facts data is available free of cost for the researchers and the students at the University of Missouri- Columbia. November 2018 folder with more than 10 million patients was selected. Diabetes was the disease of choice for this study. Database ER diagram was used to find out to fetch out the data based on the ICD-9 diagnosis codes of the diabetes disease. Tables like Encounter, Patients and Diagnosis were used to fetch the data. SAS software was used to collect the data from the folder.

3.2 DATA PREPROCESSIONG

Data preprocessing is one of the most important steps in any analysis. This is because from the dataset with there are unwanted variables and information that are unnecessary for the project.

Data preprocessing is defined as the process of converting the raw useful data into the useful and meaningful data. This project has used SQL and Python programming for the data preprocessing.

SQL (Structured Query Language) is a programming language designed for managing data in a relational database. It has been around since the 1970s and is the most common method of accessing data in databases today. SQL has a variety of functions that allow its users to read, manipulate, and change data. It is easy to use on the larger database.

Python is a versatile programming language for the datamining. Its producers define the Python language as “an interpreted, an object-oriented, high-level programming language with dynamic semantics. It's high-level built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for rapid application development as well as for use as a scripting or glue language to connect existing components.” [6]

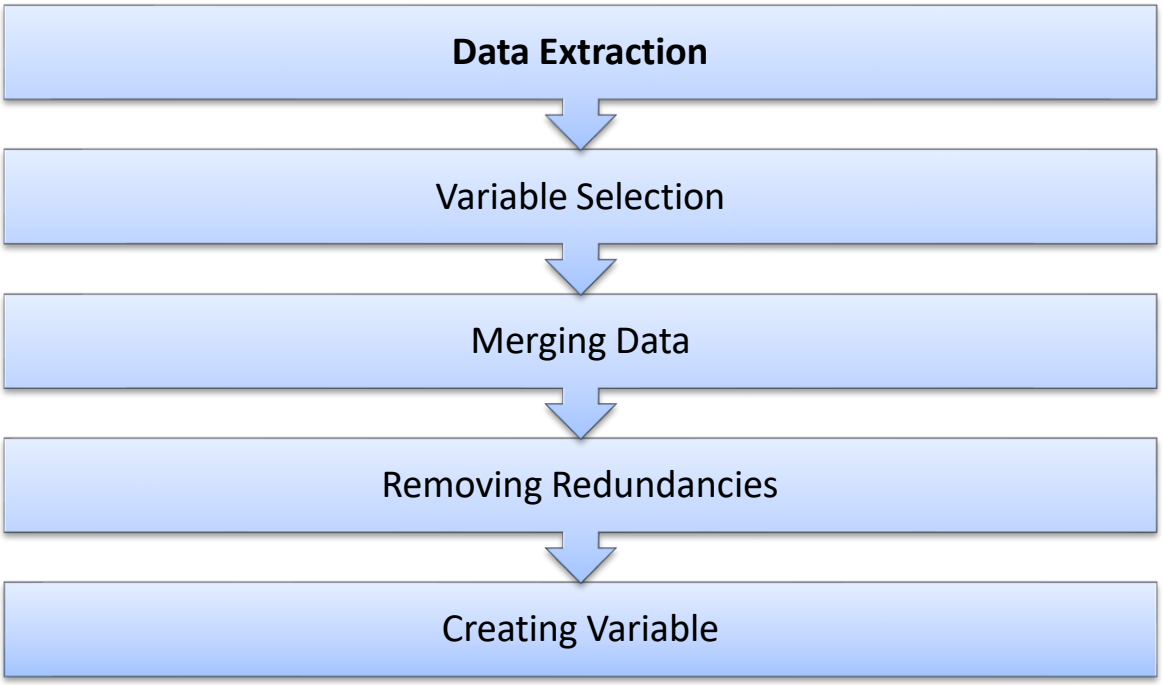


Figure 2 Data Pre-processing steps

DATA EXTRACTION

SAS software was used to collect the data from the folder.

SAS stands for Statistical Analysis System. SAS software was used on the health facts data to fetch the patient demographic information along with their diabetes information. Queries was used to extract the ENCOUNTER_ID between 2500000 till 100000000 for extracting the details of the patient demographics. In the similar way, patients with the diagnosis code were 250. 250.01,

250.02, 250.03 were queried with the ENCOUNTER_ID between 2500000 till 100000000. SO, table 1 contained column names: ENCOUNTER_ID, PATIENT_ID, AGE_IN_YEARS, RACE, MARITAL_STATUS and GENDER, ADMISSION_RATE whereas, Table 2 contains following columns: ECOUNTER_ID, DIAGOSIS_CODE, DISGNOSIS_DESCRIPTION. Both the tables were converted into the csv file.

3.3VARIABLE SELECTION

All the variables present in the dataset created using the SAS are not important for analysis. In these, there are duplicate variables, deprecated variables and variables which are irrelevant for project or which cannot help in identifying the trends. Out of 10 variables 4 variables were selected for the identification of the trends in the diabetes mellitus: RACE, AGE_IN_YEARS, GENDER, and DIAGNOSIS_CODE.

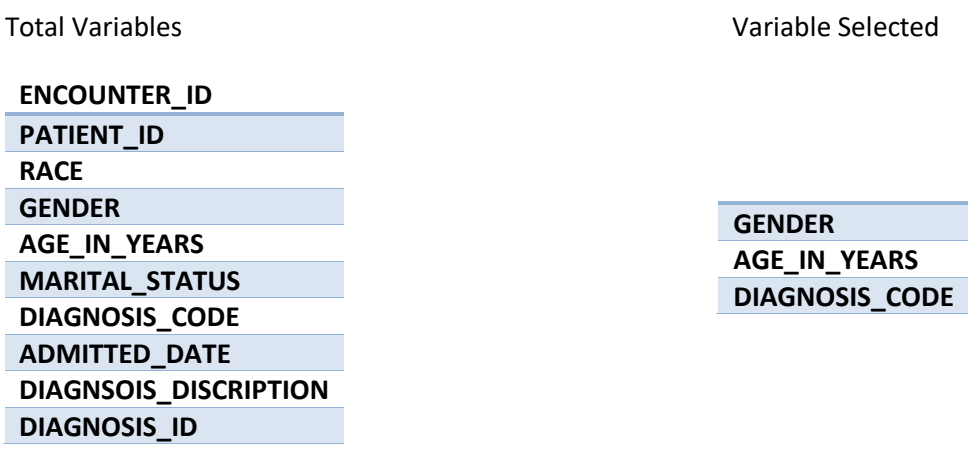


Figure 3 Diagram showing the variables extracted from Table 1 and Table 2 using the SAS software

3.4 MERGING THE DATA

After selecting the data variables for the project, next comes the process of merging the data based on the common column names and removal of the duplicate values. As the patients have the unique PATIENT_ID, SQL was used to merge the data and create a dataset with unique PATIENT_ID.

SQL stands for Structured Query Language. It is used to convert the database and help in inserting, filtering, manipulating data in different ways. The csv files created were loaded individually on the SQL database. Various queries were performed on it to merge the tables into one using the **"INNER JOIN**" function. Inner join function was used on the ENCOUNTER_ID column which is present in both the tables and the tables were merged into the single table. The columns in that table were

following: ENCOUNTER_ID, PATIENT_ID, AGE_IN_YEARS, RACE, MARITAL_STATUS, GENDER, DIAGNOSIS_CODE and DIAGNOSIS_DESCRIPTION. Unique PATIENT_IDs were generated using the "DISTINCT ()" function and a final table with 86,767 patients were generated. But not all the patients had unique PATIENT_IDs. There were some redundancies in the Patient_ID.

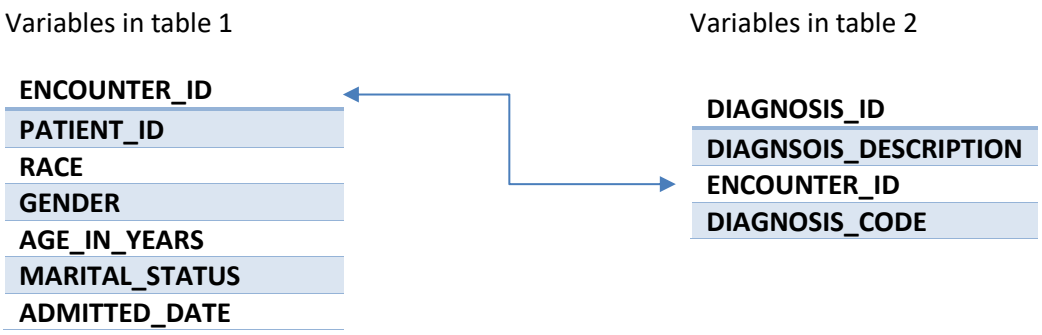


Figure 4 Showing the relation between two tables

3.5 REMOVING THE REDUNDANCIES

After merging and filtering the dataset using SQL, there were still some redundancies that need to be removed. These redundances were based on the patients with same Patient ID but, different age based on the different encounters to the healthcare facility. These redundancies in the PATIENT_ID were removed using the Python.

Python is one of the fastest growing programming languages. Data frames were created and redundant PATIENT_IDs were identified. Total of 58,541 patients were collected through this process. Later, the data frame was converted into the excel file for easy analysis.

3.6 CREATING THE VARIABLE

Microsoft excel was used in creating the variable from the variables selected for easy analysis of the data. The excel file created were loaded on the MS Excel. Various modifications I the data were created like:

1) AGE_IN_YEARS were converted based on the categories like <18 years, 18-44 years, 45-64 years, and >=65 years.
2) Diagnosis_code was grouped into TYPE1 (250.01, 250.03) and Type 2(250, 250.02)
3) Race/Ethnicity only included the patients how were either Caucasians, African Americans or Hispanics using filters.

With this the total number of patients were narrowed down to 56,845.

Pivot tables were used to analyze the data.

Age range	Variable
Less than 18 years	**<18**
Between 18 to 44 years	**18-44**
Between 45 to 64 years	**45-64**
Greater than or equal to 65 years	**>=65**

Figure 5 Variables created based on age groups.

Diagnosis_code	Variable
250	**Type_2**
250.02	
250.02-1	**Type_1**
250.03	

Figure 6 Variables created based on Diagnosis_code.

3.7 CREATING DASHBOARD

Data visualization was created using the Tableau Public dashboard. Compared to most of the BI tool options, Tableau public gives the advantage of superior visualization. The tool is designed in an intuitive manner that promotes a user-friendly interface enabling users to utilize the app's functionality. Using a drag-and-drop feature, users can easily perform complex data visualization without in-depth coding knowledge. It gives the facility to create the interactive visual analytics in the form of dashboards. A dashboard is a type of graphical user interface which often provides at-a-glance views of key performance indicators relevant to a particular objective or business process. Such dashboards created is used for both technical as well as non-technical users.

The data was loaded in the tableau dashboard in the form of excel sheets. Based on the types of data: 4 dashboards were created. One dashboard is focusing on the patients with only DM (including Type 1 and Type 2) based on the race, second dashboard included the data of patients with DM based on ethnicity and the third dashboard is based on the data of patients with diabetes mellitus and common cancers associated with diabetes. The last dashboard included the data information of the Health facts data and compare it with that of CDC data. Various bar graphs, pie charts, cluster graphs were used to create the dashboard.

3.8 CREATING INDIVIDUAL WEB PAGES

A web page is a hypertext document provided by a website and displayed to a user in a web browser. A website typically consists of many web pages linked together in a coherent fashion. HTML (Hypertext Markup Language) and CSS (Cascading Style Sheets) were used to create the individual web page, which in turn lead to creating the website.

The plan of the project was to create the web-based dashboard which means that the interactive dashboard embedded into the web pages. To create the web-based dashboard, 3 web pages were created using HTML and CSS: Home, about data (subdivided into: Health Facts and CDC) , 3D Portal report (later subdivided into: diabetes and cancer reports). Home web page gives the idea about the number of diabetes cases all over United states with the information about the Health Facts data compared to CDC data. About the data web page gives the information about the type of the data used to create the analysis (Health Facts and CDC). The health facts report webpage will give the option for three web pages, “Based on race”, “Based on ethnicity”, and “Based on cancer” reports.

All the reports have a key-messages that gives the idea about the important findings from the dashboards on the top of the web page and necessary references to dig in deep into the data at the footer part of the webpage.

4. RESULTS

4.1. ANALYSIS OF THE DATASET

The data set with 55,081 patient data were analyzed using the pivot table function of the Microsoft Excel tool. There were 7 columns in the final dataset. RACE, GENDER, <18, 18-44, 45-64,>=65, Type 1 and Type 2.

Below is the analysis based on individual variables.

4.1.1. BASED ON RACE

The dataset was filtered to non-Hispanic African-Americana and non-Hispanic Caucasian. After filtering the dataset, there were 11,373 African Americans and 43,711 Caucasians in the dataset.

Race	Total (N)
African American	11,373
Caucasian	43,711

Table 4. Based on the Race (N) in the data

4.1.2. BASED ON GENDER

The data filtered based on the gender of the patients and were also categorized based on the race. The proportion of female were 56.2% (N=31,300) whereas, 43.7% (N=24,078) were male. The below table shows the number of the African American and Caucasians based on gender.

Gender	Total (N)	Race	Total based on Race (N)
Female	31,003	African American	7,397
		Caucasian	23,606
Male	24,078	African American	3,976
		Caucasian	20,102

Table 5. Based on the Gender (N) and Race (N)

4.1.3. BASED ON AGE

The dataset was analyzed based on the categories of age for both Male and Female patients, individually. The age category followed the age-category mentioned by CDC as: <18 years, 18-44 years, 45-64 years and >=65 years.

Analysis of the Female patients based on the age-category:

Race	Type 1	Type 2
African American	39	34
Caucasian	84	66

Table 6. Number of Female below 18 years

Table 8:

Race	Type 1	Type 2
African American	219	1287
Caucasian	594	2627

Table 7. Number of Female between 18-44 age-group

Race	Type 1	Type 2
African American	155	3073
Caucasian	573	8023

Table 8. Number of Female between 45-64 age-group

Race	Type 1	Type 2
African American	101	2489
Caucasian	633	11006

Table 9. Number of Female greater than or equal to 65 years

Analysis of the Male patients based on the age-category:

Race	Type 1	Type 2
African American	20	14
Caucasian	74	57

Table 10. Number of Male below 18 years

Race	Type 1	Type 2
African American	90	635
Caucasian	261	1649

Table 11. Number of Male between 18-44 age-group

Race	Type 1	Type 2
African American	97	1965
Caucasian	517	7759

Table 12. Number of Male between 45-64 age-group

Race	Type 1	Type 2
African American	35	1120
Caucasian	392	9396

Table 13. Number of Male greater than or equal to 65 years

4.1.4. BASED ON HISPANICS

Age range	Female	Male
<18	8	2
18-44	231	180
45-64	356	363
>=65	326	295

Table 14. Number of Hispanic Male and Female based on age-groups

<18	1
18-44	19
45-64	8
>=65	5

Table 15. Based on number of Females with Type 1 DM

<18	7
18-44	212
45-64	348
>=65	321

Table 16. Based on number of Females with type 2 DM

<18	1
18-44	17
45-64	18
>=65	1

Table 17. Based on number of Males with Type 1 DM

<18	1
18-44	163
45-64	345
>=65	294

Table 18. Based on number of Males with Type 2 DM

4.1.5. BASED ON TYPE OF DIABETES

As mentioned, there are two type of Diabetes: Type 1 Diabetes and Type 2 Diabetes. Majority of the population suffered from the type 2 DM (96%).

Type of Diabetes Mellitus	Total (N)
Type 1 DM	3,954
Type 2 DM	52,891

Table 19. Total number based on type of Diabetes Mellitus

4.2 TABLEAU DASHBOARD

Tableau public was used for creating the dashboards. The data findings from the MS excel pivot table were loaded and connected to each other to create the dashboards. Total 4 tableau dashboards were created using the Tableau Public tool.

Dashboard 1: Showing the comparison of the CDC data with the Health Facts data.

This dashboard gives the brief introduction to the type of the data used. The data for the total number of diabetes mellitus cases, number of cases of Type 2 Diabetes mellitus, Type 1 Diabetes mellitus cases were added. As the project is based on the race and ethnic disparities in Diabetes Mellitus patients, Number of populations with diabetes mellites based on race (African American and Caucasian) and the number of Hispanic populations with Diabetes Mellitus were added. To compare the Health Facts data with the overall population with Diabetes Mellites across the United States, CDC data were used. In addition to those data, the data regarding the patients below 20 years with Type 1 and Type 2 Diabetes Mellites were also added to give the idea regarding the Juvenile Diabetes Mellitus (Type 1).

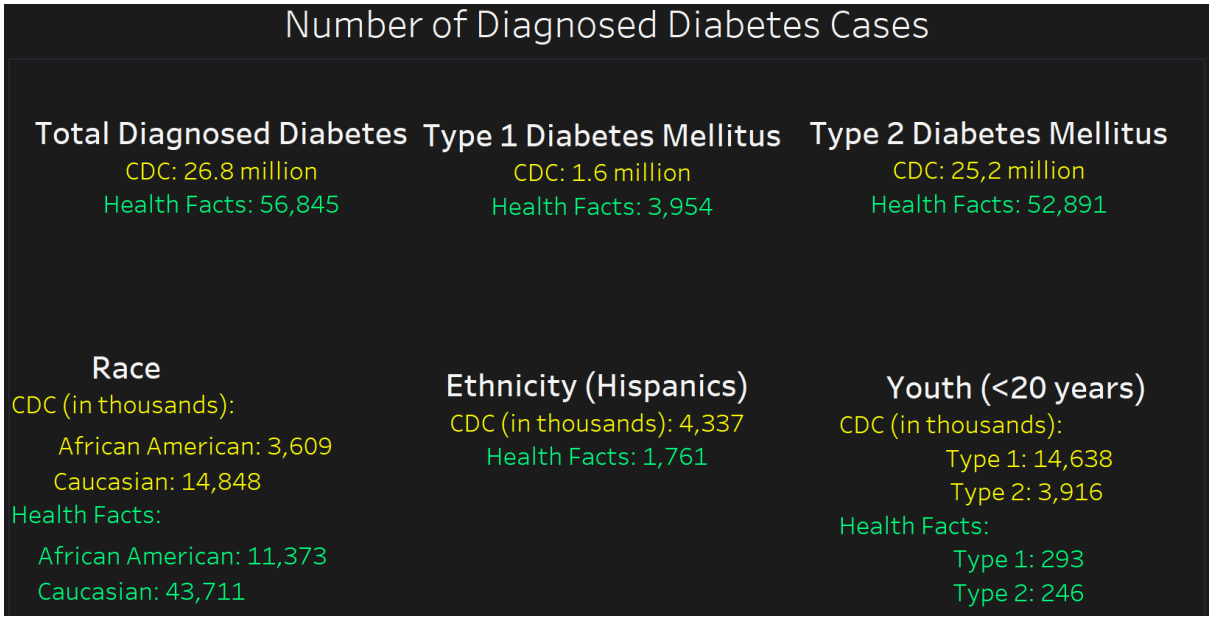


Figure 7. Screenshot of the home dashboard showing total cases

Usability feature: Different font colors were used to distinguish to compare different data sources (for example, yellow color for CDC data and green color for Health Facts data).

Findings: As the data from the Healthcare facilities (collected in Health Facts) is compared with the CDC data, there are more proportion of the patients with clinically diagnosed diabetes mellites.

Dashboard 2: Based on the Race

This dashboard will show the tends of the patients with type 1 and type 2 Diabetes Mellitus based on Race. Horizontal bar graphs have been used throughout the dashboard to maintain its consistency. The following graphs have been used : Number of Females based on Race, Number of Males based on Race, Number of American-American and Caucasian below 18 years for both Male and Female based on Type of Diabetes Mellitus, Number of American-American and Caucasian

between 18-44 years for both Male and Female based on Type of Diabetes Mellitus, Number of American-American and Caucasian between 45-64 years for both Male and Female based on Type of Diabetes Mellitus, Number of American-American and Caucasian greater than or equal to 65 years for both Male and Female based on Type of Diabetes Mellitus.

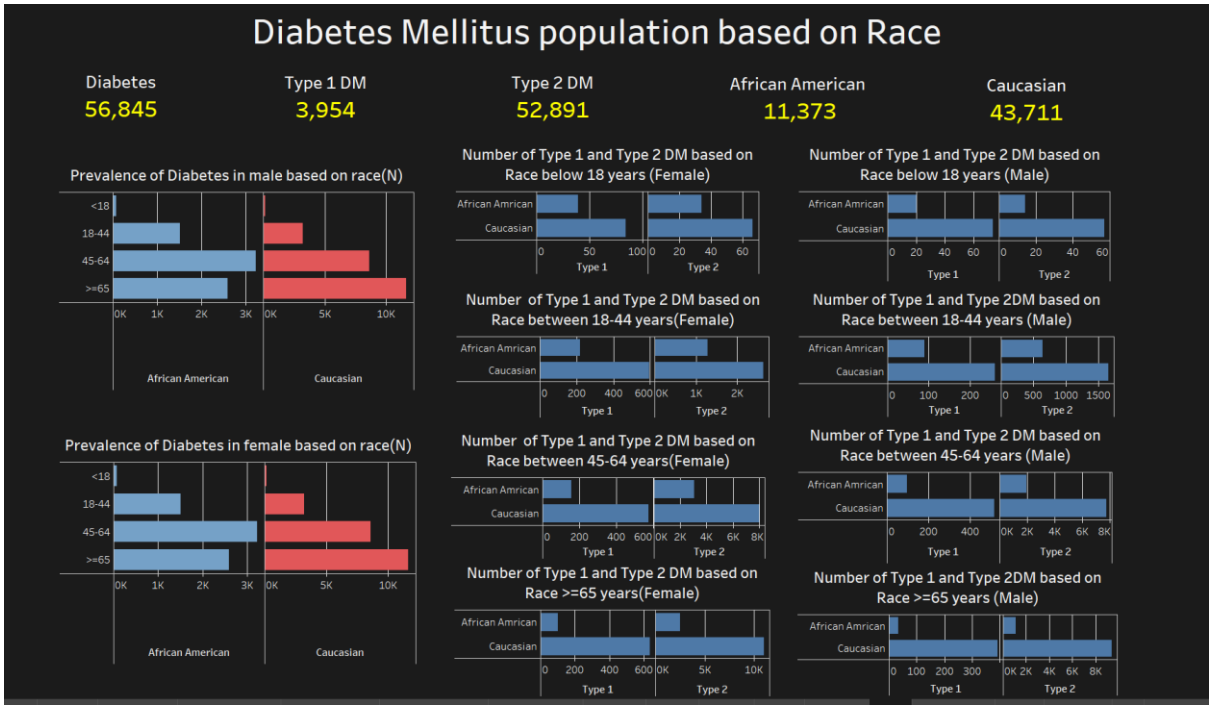


Figure 8. Screenshot of the DM population based on Race

Findings:

- The proportion of African American between age group 18-44 years with critically diagnosed DM is 4.05% as compared to Caucasian with proportion 9.31%.
- The proportion of African American between age group 45-64 years with critically diagnosed DM is 9.6% as compared to Caucasian with proportion 30.62%.
- The proportion of African American between age greater than 64 years with clinically diagnosed DM is 6.72% as compared to Caucasian with proportion 38.89%.
- The proportion of African American between age group below 18 years with clinically diagnosed DM is 0.2% as compared to Caucasian with proportion 0.5%.

Usability feature: The dashboard has all the graphs as horizontal bar graphs to avoid confusion to the eyes. The top-most row of the dashboard has the total number of the patients with total number of African American and Caucasian race to give a brief idea regarding the racial disparities from the graphs.

Dashboard 3: Based on the Ethnicity

This dashboard shows the trends of the patients with Type 1 and Type 2 Diabetes Mellitus based on the Ethnicity. Hispanic population were shown in the graphs. Horizontal bar graphs were used to show the population with type 1 and type 2 Diabetes Mellitus for Male and Female base on the age-groups. Vertical stacked graphs were shows to categorize individually the Hispanic population for

Type 1 and Type 2 Diabetes Mellitus. The tables with its values can be found on the bottom-left side of the dashboard.

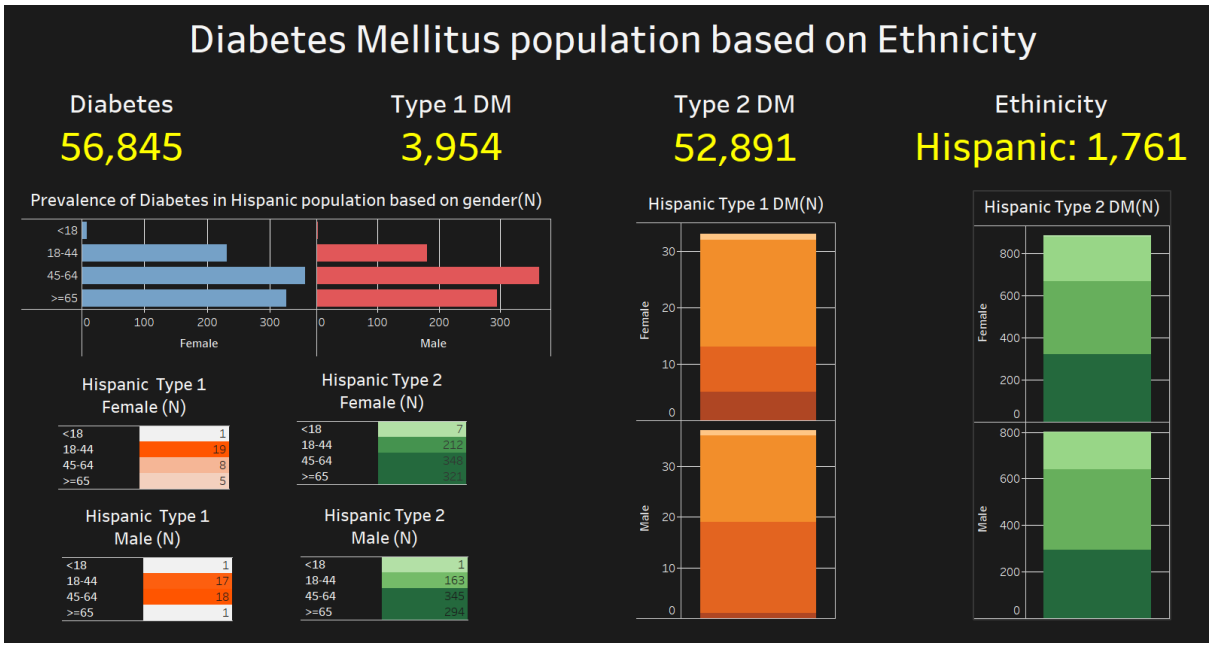


Figure 9. Screenshot of the DM population based on Ethnicity

Findings:

- According to CDC, a Hispanic/Latino American adult has more than 50% chance to develop DM at a younger age.
- About 40.8% of the Hispanic population is between age 45-64 years.
- The proportion of Hispanic female is 52.2%, whereas Hispanic male is 47.4%.
- The population of Hispanics between age group 18-44 years id 51.4% (highest in Type 1 DM) and between 45-64 years is 40.98% (highest in Type 2 DM).

Usability feature: The color of the vertical stacked graph with orange is the same to the tables with the values showing the same components. The similar goes to the vertical stacked graph with green.

Dashboard 4: Based on the trends for the patients with Diabetes Mellitus and the Cancers

As there are various studies showing that there is a link of diabetes mellitus and cancers due to the hyperinsulinemia d hyperglycemia. There are few cancers that are common for the diabetes patients: Breast, Bladder, Colon, Pancreas, Liver and Uterus. Most of these cancers are associated with Type 2 Diabetes Mellitus.

Horizontal bar graphs were used to show the patients with specific cancers based on the Race (African American and Caucasian), Pie chart was used to show the population of Hispanic with cancers caused in them. The vertical bar graph was used to show the patients with specific cancers based on the type of diabetes mellitus.

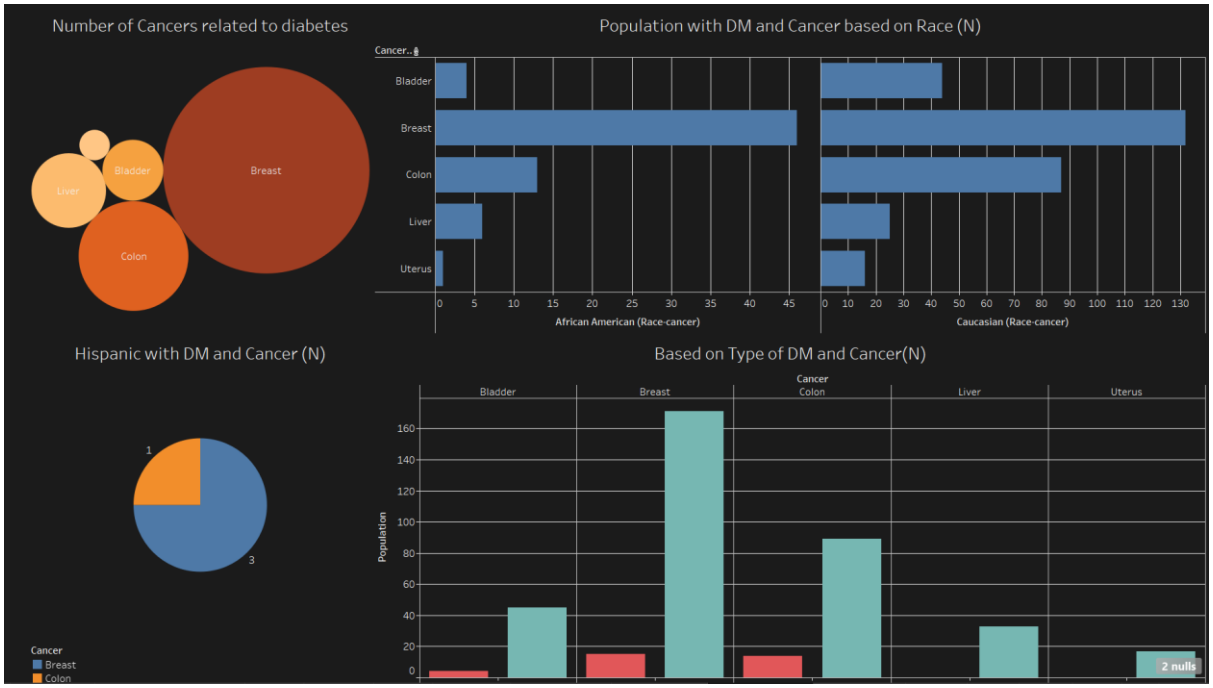


Figure 10. Screenshot of the DM population with Cancer

Findings:

- About 18.7% of the total population are African American and 81.2% are Caucasian.
- Breast cancers constitute about 47.9% of the total cancers.
- Hispanic population are associated with Breast and Colon cancers.
- Uterus cancer and Liver cancer are only associated with Type 2 DM.
- The proportion of Type 2 DM patients with Liver cancer is 9.2% and Uterus cancer is 4.7%.

Usability feature: The four graphs used were placed in each corners of the dashboard to have a good vision and scope of understanding the data better.

Another dashboard was created as a sub-category for the Cancer-Diabetes mellitus dashboard. The dashboard shows the comparison of African American race based on the types of cancer and diabetes with the Caucasian race.

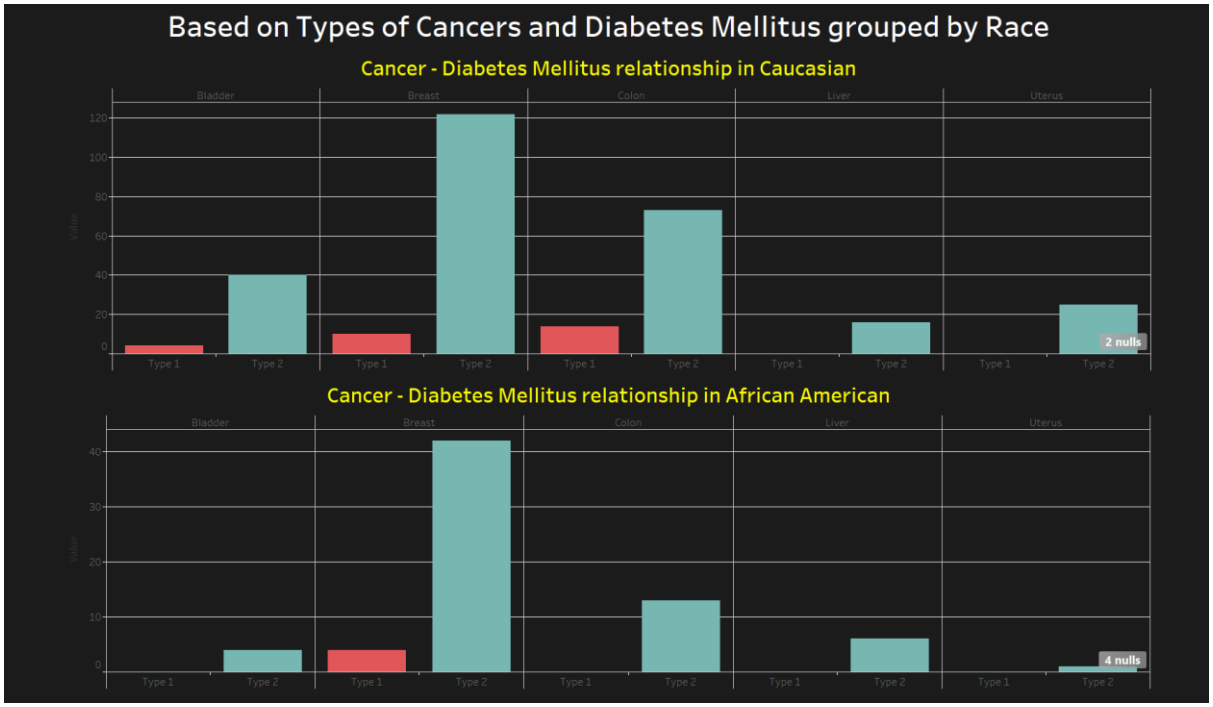


Figure 11. Screenshot of the population with DM and Cancer based on Race

4.3. WEBSITE (INSERT THE DIAGRAMS OF EACH WEB PAGE)

The Tableau dashboard embedded website is given the name of 3D Portal. 3D stands for Diabetes Demographics Dashboard. It has the navigation bar and the webpages attached to it. Here is the individual webpage described in detail.

4.3.1. INTRODUCTION PAGE

An introduction page was created to introduce the purpose of the website and the explain its contents. It gives a brief introduction regarding the data source used on the website, the number of web pages and its contents.

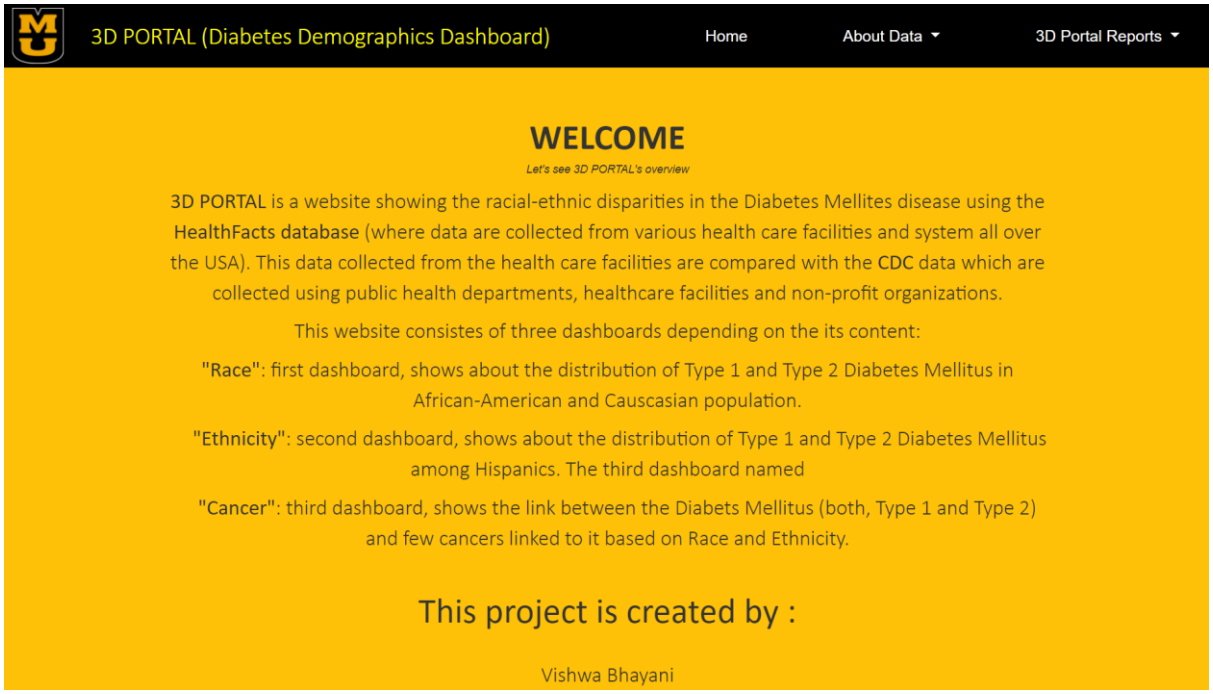


Figure 12. Screenshot of the Introduction page

4.3.2. HOME PAGE

The home page section gives the introduction about the 3D portal with an image of the US Diabetes Map which shows the percentage of adults who reported being told by a health professional that they have diabetes. The source of this map is CDC, Behavioral Risk Factor Surveillance System.

Below that there is the Dashboard with the information about the number of diabetes mellitus patients in CDC and Health Facts data.

The footer part of the index webpage contains the reference links such as CDC-Diabetes data & Statistics, National Diabetes Statistic Report, 2020 and 2019 Diabetes Report Card.

Usability feature: The color for the CDC data is coaded in Yellow font whereas, the color for the Health Facts data is located in the Green fonts.

4.3.3. ABOUT DATA

About the data web page has a drop down with two web pages: Health Facts data and CDC data.

The Health Facts data web page will have all the information about the data like what is that data, what are the benefits of using the data for researchers and how to get access to the data.

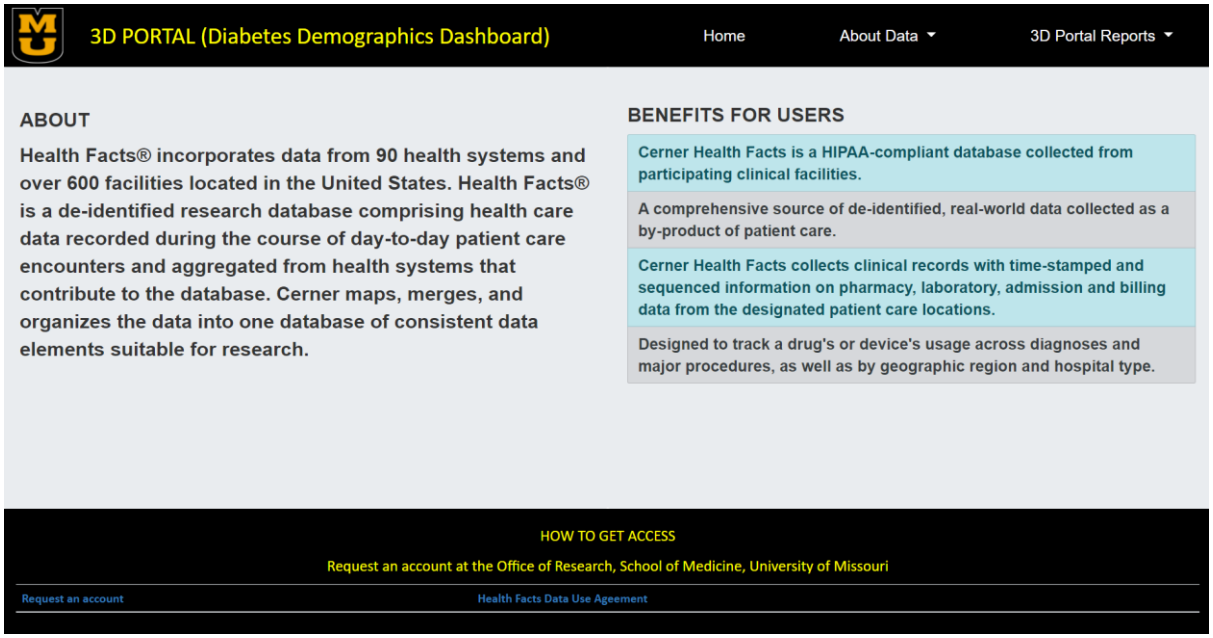


Figure 13. Screenshot of the Health Facts web page

CDC data web page will also give the similar information regarding what is that data, what are the advantages of using the data and how to get access to the data.

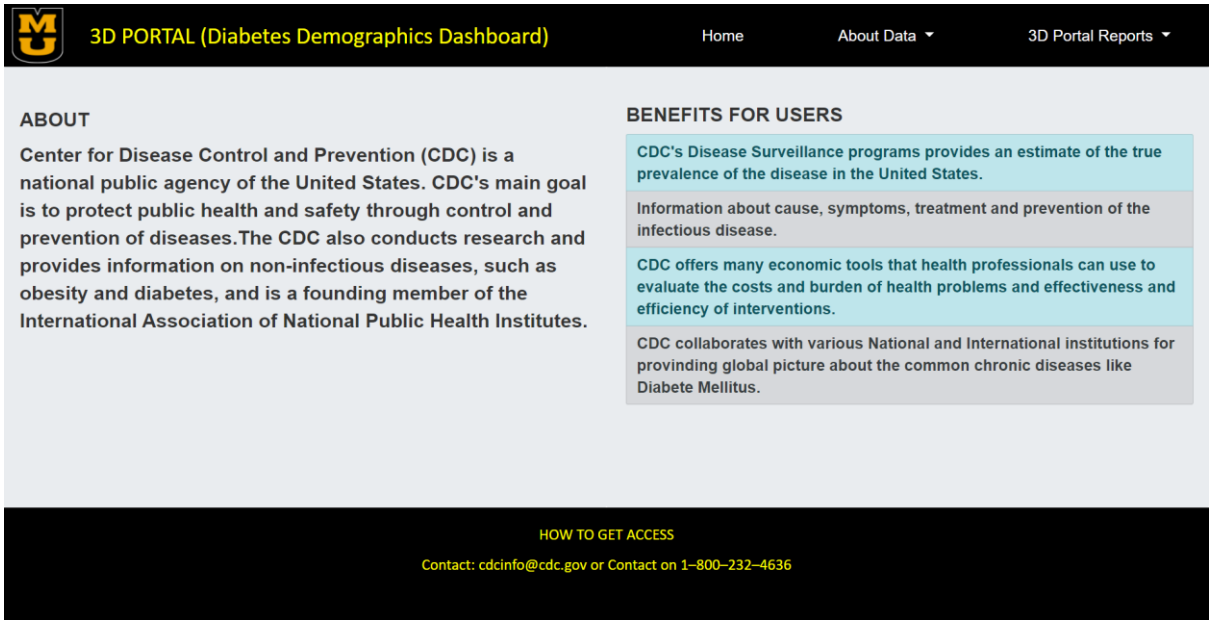


Figure 14. Screenshot of the CDC web page

4.3.4. 3D PORTAL REPORTS

The health Facts reports constitutes of three webpages: Race, Ethnicity and Cancer.

All these three web pages will have the similar template to that of the home page of the about the data pages. The construction is shown un the below diagram.

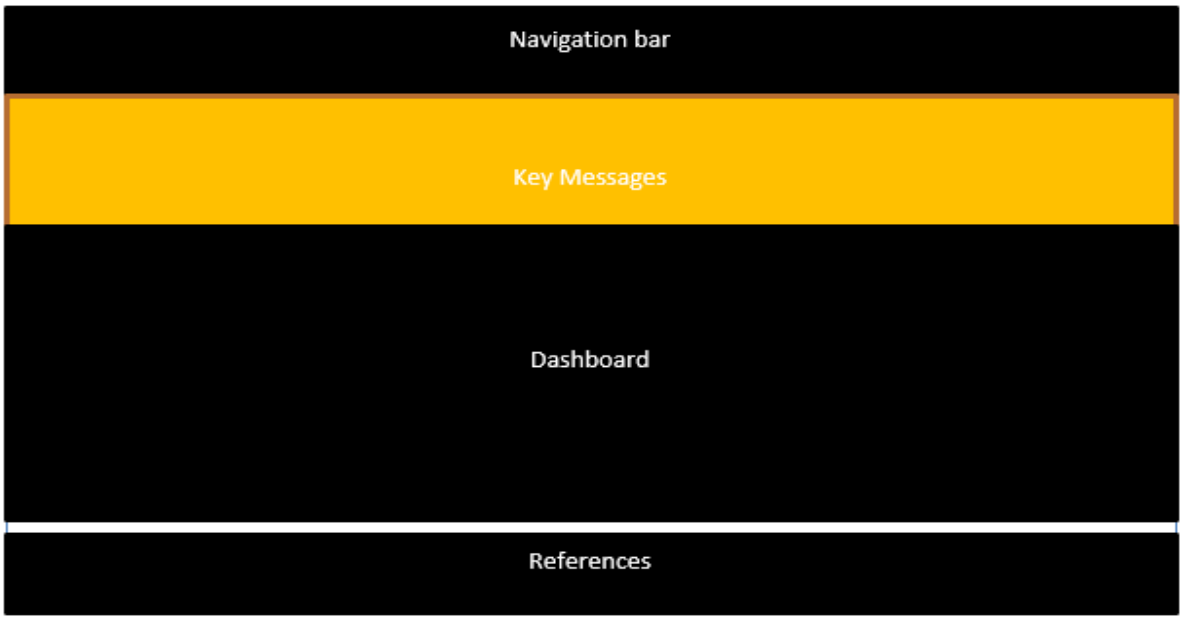


Figure 15. Image showing the template of the report web pages

There will be a navigation bar at the top, followed by that there will be an area for the key findings and the individual dashboard attached to that web page. Below that there will be an area for the references used.

Sample of race web page is shown below:

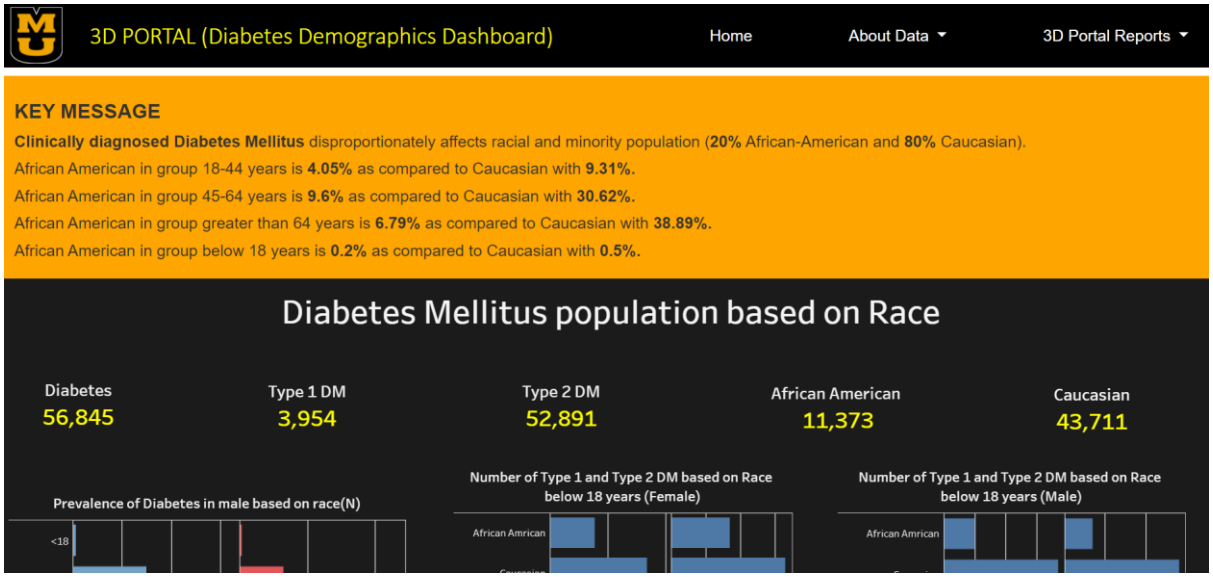


Figure 16. Top sceenshot of the Race web page

4.4. EVAULATION OF WEBSITE BASED ON USABILITY

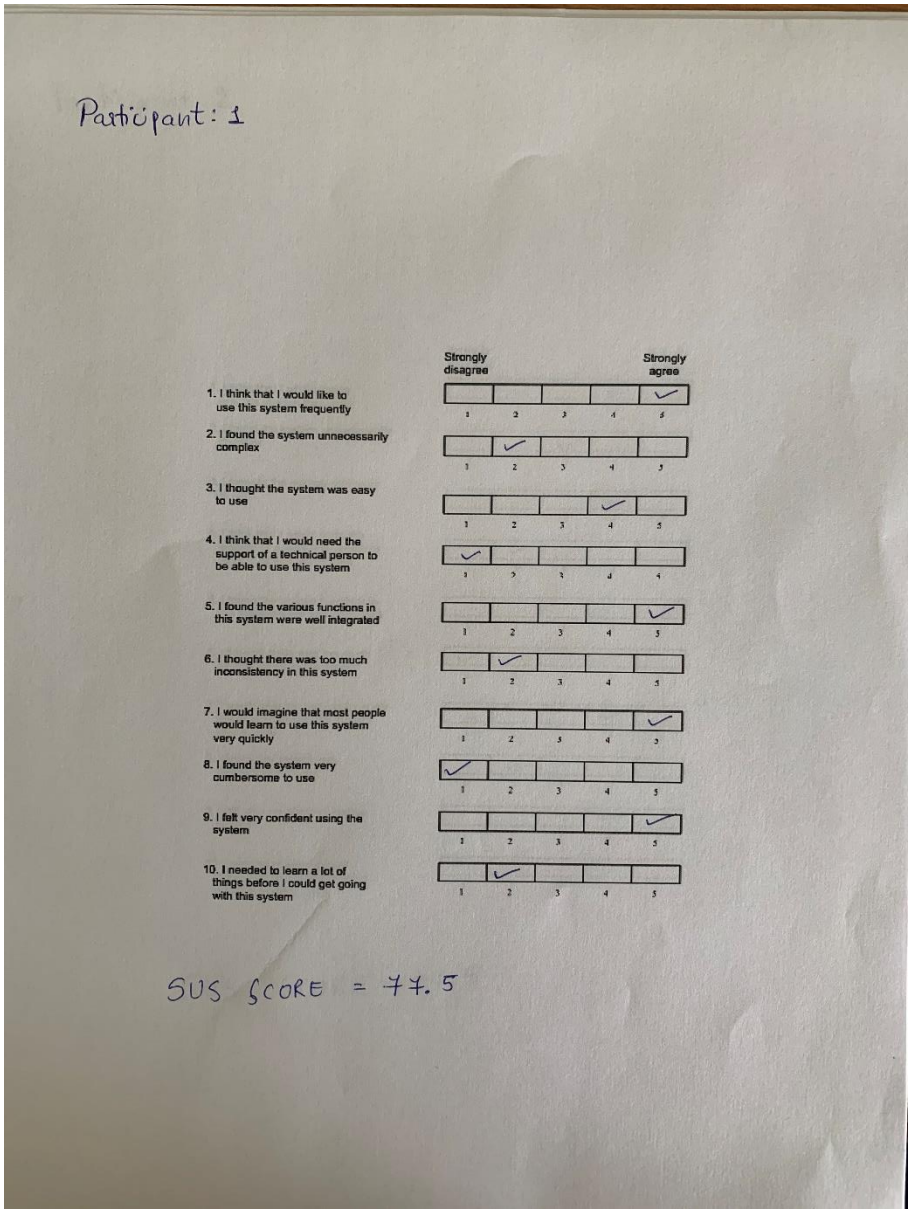
The usability of the website was done using the System usability Scale (SUS). The System Usability Scale (SUS) provides a “quick and dirty”, reliable tool for measuring the usability. It consists of a 10-item questionnaire with five response options for respondents; from Strongly agree to Strongly disagree. Originally created by John Brooke in 1986, it allows you to evaluate a

wide variety of products and services, including hardware, software, mobile devices, websites and applications. Based on research, a SUS score above a 68 would be considered above average and anything below 68 is below average,

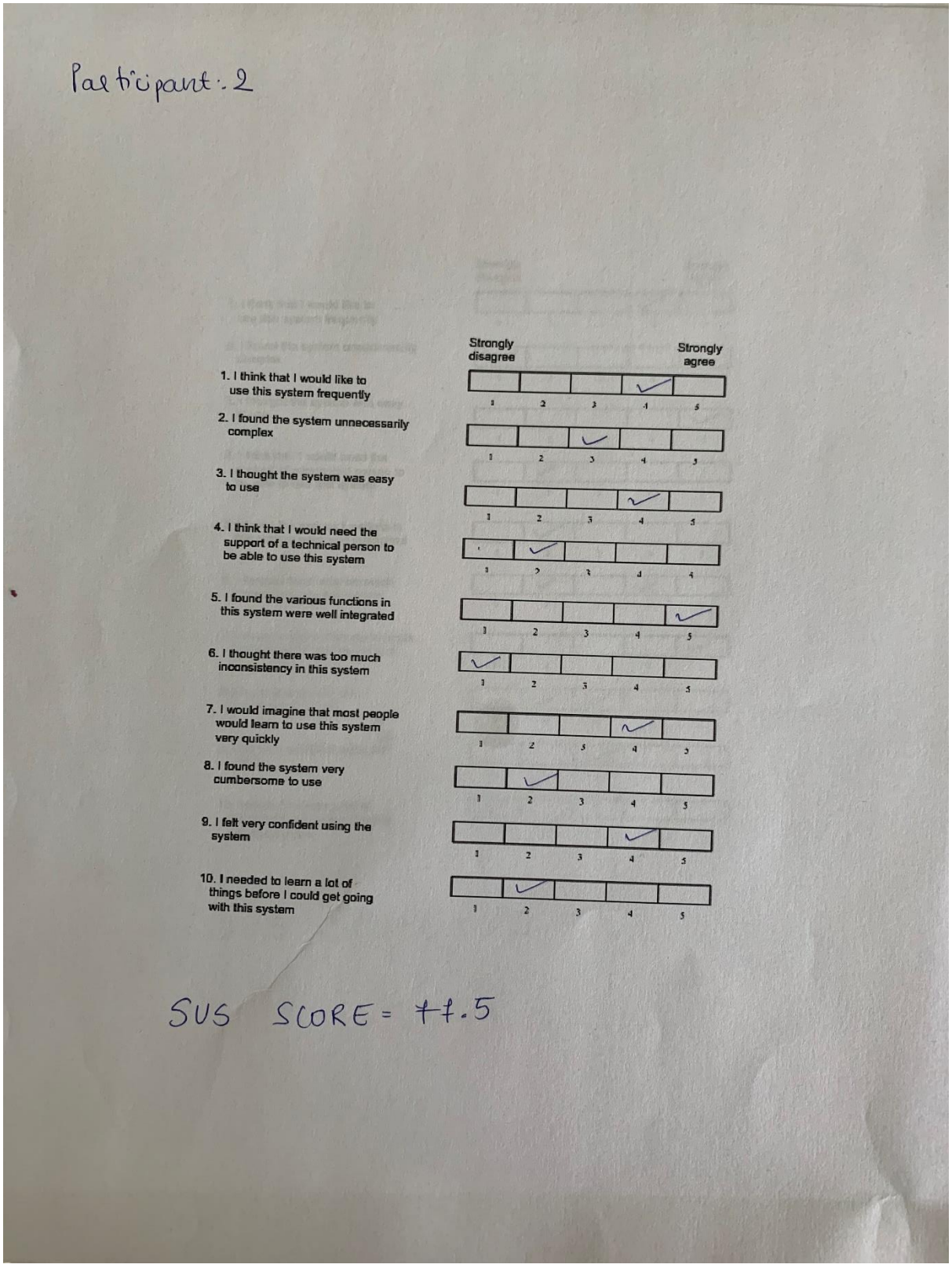
Method: 3 participants were asked to perform the task of checking out the number of patients with Type 2 Diabetes Mellites based on the race who are females and are between 45-64 years of age and the number of people with Type 2 Diabetes Mellites and Breast cancer.

Evaluation forms by participant: As this website is targeted to the population health researchers or epidemiologists, well-educated participants with the education or skill in any scientific field were selected.

Participant 1:



Participant 2:

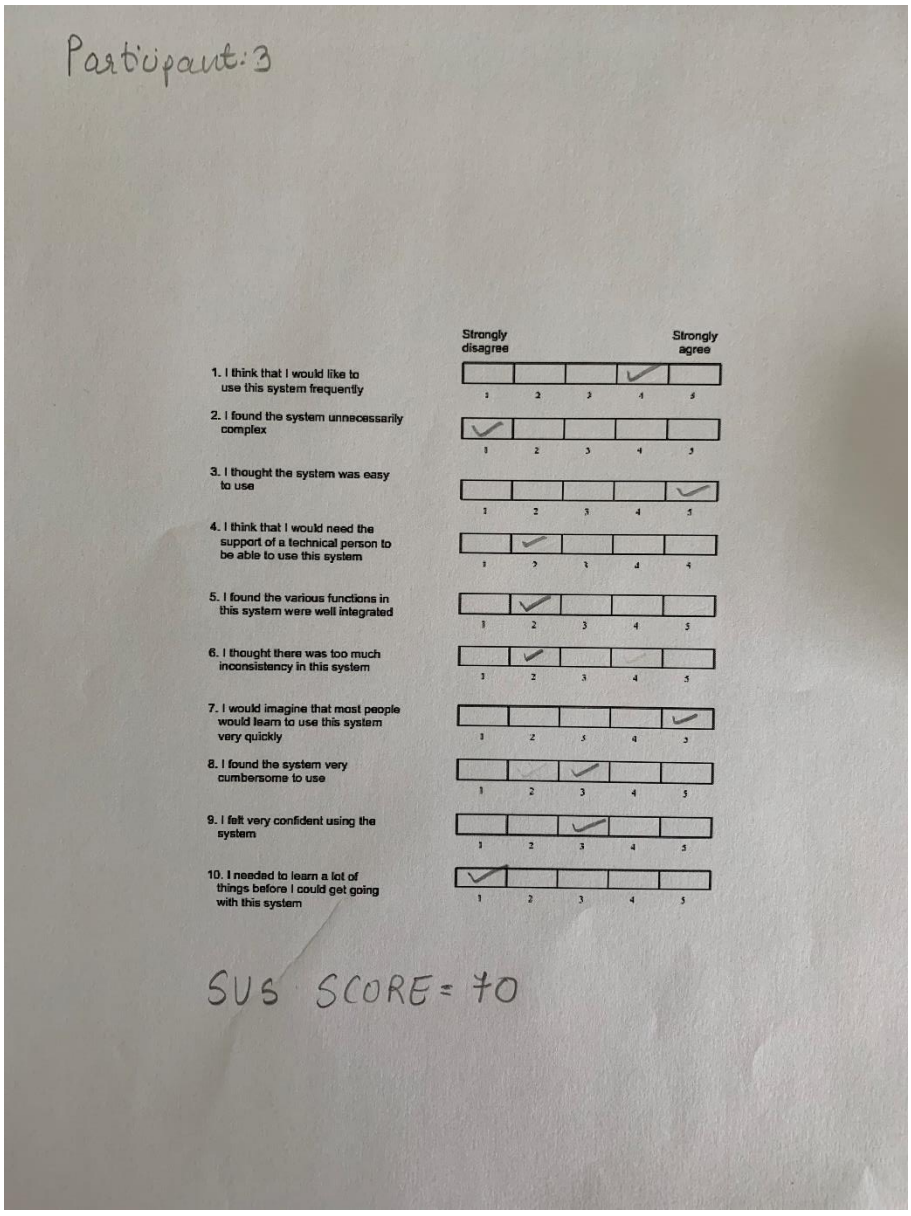


Participant. 2

	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently	1	2	3	✓ 4	5
2. I found the system unnecessarily complex	1	2	✓ 3	4	5
3. I thought the system was easy to use	1	2	3	✓ 4	5
4. I think that I would need the support of a technical person to be able to use this system	✓ 1	2	3	4	5
5. I found the various functions in this system were well integrated	1	2	3	4	✓ 5
6. I thought there was too much inconsistency in this system	✓ 1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	1	2	3	✓ 4	5
8. I found the system very cumbersome to use	1	✓ 2	3	4	5
9. I felt very confident using the system	1	2	3	✓ 4	5
10. I needed to learn a lot of things before I could get going with this system	1	✓ 2	3	4	5

SUS SCORE = 77.5

Participant 3:



Average SUS Score of the three participants: 75

From the usability point of view, we can say that this website can be comfortably used by the people with some scientific knowledge.

5. DISCUSSION

BASED ON DATA PREPROCESSING

Complex relational databases make data mining a difficult job. These databases are usually for the operational transactions but not for analytics. To mine the data and create the dataset with unique values, data mining tools like SQL is helpful. SQL will help in filtering the data by creating the tables with the limited values and variables just with few quick queries. It is very helpful in cleaning up the data by removing the redundant values and the missing values from the dataset. The dataset used in this project is from a complex relational database with a large schema. SQL is a very decent tool in generating the tables and merging them. Data wrangling tasks like removing the missing values, eliminating unwanted columns, merging the tables on a common variable, getting unique IDs were extracted from the raw data collected.

Sometimes, filtering using just one tool is not enough. Many of the redundances cannot be removed if the values for the other variables are different. If these redundancies stay for the final evaluation or analysis, false analysis report can be generated and can misguide the findings. Python programming language was used for further filtering out and manipulate the data. Python and SQL go in hand in hand in data pre-processing. Python has a variety of in-built libraries like NumPy and Pandas, that is specially designed for data analysis and manipulation. Creating the data frames from the data set will help in analysing the data based on a particular variable or condition. So, to get the final dataset without any duplicate data, Pandas' library of python programming was used. Jupiter Notebook was used for the same. Data frames were created, and codes were implemented to identify the total number of the duplicates and loop function was applied to remove and get the final dataset.

Once the dataset was created, it was ready for the final analysis. To analyze the data, identifying the important variables and selecting the dependent variables are important tasks. Sometimes, the interpretation is difficult or is not possible with the existing variables. Creating variables by categorizing a particular variable will give more insight to that dataset. To analyze the data properly, the dataset created with python was loaded to the MS Excel tool. Various variables were created like base done the category of the age, race and diabetes mellite types. Filtering feature of the MS Excel tool helps inn focusing on the specific information and height the important part. Pivot tables helps to summarize, sort, group and in turn analyze the specified data and their relationship with other variables in the dataset. Different arithmetic functions like mean, count, sum, multiply, etc functions can be performed using the pivot tables. The data were analyzed using the pivot tables and each analysis was stored in individual sheets of the MS excel.

The data is ready for its visualization.

BASED ON DATA VISUALIZATION

Tableau is a great tool for data visualization. Once you have designed in Tableau in the form of a dashboard, you can easily publish that to the web browser. Creating the individual dashboards and merging them with the individual web pages is an innovative way to use the visualization at a new level. Here, it not only focuses on the visualizations, but helps in evaluation, usability and making a

more interactive website, which can give the information about the various information as well. This is possible by adding various links into the website.

The dashboard contains the details of the patients with Diabetes Mellitus based on the Race and Ethnicity. Racial and ethnicity also play an important role in the seeking the diabetes care. Racial and ethnic minorities bear a disproportionate burden of the diabetes epidemic. They are the higher rate of serious health conditions that affect communities of color. They have been adding to the worse diabetes control and higher complications. Despite of high prevalence of the condition, minorities experience lower quality of care, and great barriers to self-management compared to Caucasian patients. These disparities can result in the shorter lifespans and lower quality of life, with lack of proper resources like diet, education, employment, safe and healthy neighborhood. Compared with the white adults, risk of diabetes in African American is 77% higher and in Hispanic it is 66% higher. Each dashboard gave the unique trends about the occurrences of only Type 1 and Type 2 Diabetes Mellites or with specific cancers different race and ethnic group.

The trends have indicated that African Americans are less in the dataset, this is because of the following reasons:

1) The proportion of the African American with health insurance is very less. This data is collected in the healthcare facility so, there are races like Caucasian who have easy access to the health insurance.
2) The obesity is directly linked to the diabetes mellites, but still while extracting the information, just few of the Caucasian and African-American were mentioned as obese. So, to determine the cause of the diabetes mellites due to obesity, clinicians should enter that as diagnosis code to track the progress of their weight and BMI.
3) The Hispanic population are very less and not privileged like the Caucasian with health Insurance.

Moreover, as an extension of this project, the racial and ethnic disparities were also identified in the diabetes with Diabetes Mellite and specific cancers. The report of the cancers associated with the diabetes will signify, which cancers are common with the person with diabetes mellitus. There have been the studies saying that type 2 DM is mostly associate with the cancer related to the colon, breast, uterus, liver and pancreas. There is a theory of how the person uncontrolled hyperinsulinemia will affect the cellular proliferation and reduce the apoptosis of the cell, triggering the cancer growth. As you can see the breast cancer rate is more as compared to the other cancers. Also, the liver and uterus cancers are associated only with the Type 2 DM because of the medication and the insulin interaction with the estrogen levels cause it to trigger the cancer cell proliferation.

Scope of the dashboard: This type of dashboard can be used by the agencies implementing the programs like National diabetes prevention program and Native Diabetes Wellness programs which is the partnership of the public and private organization working together to deliver affordable, evidence-based lifestyle change program to help people with diabetes. To identify the particular group of the racial and ethnic minority to get them enroll in such programs at a lower rate.

All these three dashboards were embedded in a website named 3D portal. The main purpose of this website is to make it intuitive, easy to navigate and help users to return to use it frequently. The

dashboards created using the Tableau were embed on the website as individual web pages, along with other information about the data used and the purpose of the website to help the users easy to navigate the required information. As the data was complex, small data visualizations in the form of tableau dashboards for specific purposes were created to make it simple into manageable chunk. Moreover, not too much information or less information were provided on the website. The information that are required by the us showing the demographics of the patients with diabetes mellites were included.

In addition to that, specific links were mentioned into the website that helped the used to dig deeper into the data and get clarification. They are in the form of hyperlinks to an external site.

Another feature of improving the navigation was adding the key findings for each dashboard at the top before the dashboard to give the idea about the findings to the users so that users are not confused while looking at the dashboard, which is likely to be the case if the position the key finding, and dashboard were vice versa.

While designing the website, the scientific audience, who do not have much computer skills were kept in mind. The design of the dashboard was done while keeping the usability features into consideration like the following:

1) The font size was appropriate to look from a distance of minimum 3 feet.
2) The graphs used in all the three dashboards were similar to avoid confusing.
3) The template deceased for all three dashboards were similar to maintain uniformity.
4) The dashboard has maintained enough white spaces between the graphs.
5) The information regarding the data is mentioned at the top to give the user idea about what is there in the data and what to expect.

The SUS score is a measure of reliability for measuring the usability for a tool of software. Here the tool we used is a website. The results of the SUS scores proved that this website is usable to someone with a scientific background (should not be from the healthcare specifically).

Usefulness of this website:

1) Writing grants: Grant writing is an important task and the first step to get enough funding to achieve specific goals. This website can give the information about the racial disparities in the patients with diabetes mellitus and the need for certain intervention programs or healthcare strategies in specific areas or around the US county.
2) Identify the racial-ethnic disparities: The data and graphs give the idea about the trends of the health disparities in the people with color. It provokes the need to change the strategies to implement the cost-effective solutions for the minority population with diabetes mellitus.
3) Collaborating with Non -profit organizations: Once the need for the interventions identified, by collaborating with the non-profit organizations who work in providing the care to the people I the terms of nutrition food pantry services, healthy shelter homes, educating regarding the diseases for the betterment of the community.

LIMITATIONS:

1) The number of variables were less in creating the dashboard. If there are availability of more variables like their socio-economic status, education zip codes, county data, plotting on the geo maps would have been easier. If the variable related to the smoking or alcohol drinking would have been present, the reason for their disease would have been known from the demographics point of view.
2) In evaluation of the usability, there were only 3 participants. To make the SUS interpretation reliable, there should be at least 15 participants.
3) Due to the lack of the obesity data, the dashboard were created without incorporating the direct link between the diabetes mellitus and obesity.

FUTURE WORK:

1) Performing the geo mapping using the zip codes to find the geographical area with the prevailing racial and ethnic disparities in Diabetes Mellitus.
2) Focusing more on the trends of involving the population with breast cancer and diabetes mellitus in African American and Caucasian.

6. BIBLIOGRAPHY

1. Definition of Diabetes mellitus. (n.d.). Retrieved May 25, 2021, from https://www.medicinenet.com/diabetes_mellitus/definition.htm
2. *Advanced Science Letters*, *24*(11), 8632–8639. https://doi.org/10.1166/asl.2018.12315
3. Catalano, M. M., Vaughn, P., & Been, J. (2017). Using Maps to Promote Data-Driven Decision-Making: One Library's Experience in Data Visualization Instruction. *Medical Reference Services Quarterly*, *36*(4), 415–422. https://doi.org/10.1080/02763869.2017.1369292
4. Chishtie, J. A., Marchand, J.-S., Turcotte, L. A., Bielska, I. A., Babineau, J., Cepoiu-Martin, M., Irvine, M., Munce, S., Abudiab, S., Bjelica, M., Hossain, S., Imran, M., Jeji, T., & Jaglal, S. (2020). Visual Analytic Tools and Techniques in Population Health and Health Services Research: Scoping Review. *Journal of Medical Internet Research*, *22*(12). https://doi.org/10.2196/17892
5. Dadari, I., Higgins-Steele, A., Sharkey, A., Charlet, D., Shahabuddin, A., Nandy, R., & Jackson, D. (2021). Pro-equity immunization and health systems strengthening strategies in select Gavi-supported countries. *Vaccine*. https://doi.org/10.1016/j.vaccine.2021.03.044
6. Dancy-Scott, N., Dutcher, G. A., Keselman, A., Hochstein, C., Copty, C., Ben-Senia, D., Rajan, S., Asencio, M. G., & Choi, J. J. (2018). Trends in HIV Terminology: Text Mining and Data Visualization Assessment of International AIDS Conference Abstracts Over 25 Years. *JMIR Public Health and Surveillance*, *4*(2), e8552. https://doi.org/10.2196/publichealth.8552
7. *Developing an Interactive Data Visualization Tool to Assess the Impact of Decision Support on Clinical Operations*. (n.d.). Retrieved April 9, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6148824/.
8. Goldstein, B. A., Cerullo, M., Krishnamoorthy, V., Blitz, J., Mureebe, L., Webster, W., Dunston, F., Stirling, A., Gagnon, J., & Scales, C. D. (2020). Development and Performance of a Clinical Decision Support Tool to Inform Resource Utilization for Elective Operations. *JAMA Network Open*, *3*(11), e2023547. https://doi.org/10.1001/jamanetworkopen.2020.23547
9. Khan, S., Richardson, S., Liu, A., Mechery, V., McCullagh, L., Schachter, A., Pardo, S., & McGinn, T. (2019). Improving Provider Adoption with Adaptive Clinical Decision Support Surveillance: An Observational Study. *JMIR Human Factors*, *6*(1), e10245. https://doi.org/10.2196/10245
10. Khirfan, L., Mohtat, N., Peck, M., Chan, A., & Ma, L. (2020). Dataset for assessing the scope and nature of global stream daylighting practices. *Data in Brief*, *33*. https://doi.org/10.1016/j.dib.2020.106366
11. Ko, I., & Chang, H. (2018a). Interactive data visualization based on conventional statistical findings for antihypertensive prescriptions using National Health Insurance claims data. *International Journal of Medical Informatics*, *116*, 1–8. https://doi.org/10.1016/j.ijmedinf.2018.05.003
12. Martinez, W., Threatt, A. L., Rosenbloom, S. T., Wallston, K. A., Hickson, G. B., & Elasy, T. A. (2018). A Patient-Facing Diabetes Dashboard Embedded in a Patient Web Portal: Design Sprint and Usability Testing. *JMIR Human Factors*, *5*(3), e26. https://doi.org/10.2196/humanfactors.9569
13. Moll, M. C., Decavel, F., & Merlet, C. (2009). Tableau de bord d'évaluation du système qualité des pôles en établissement de santé: Un outil pédagogique. *Recherche en soins*

infirmiers, *N° 98*(3), 19–27.
14. Park, Y., Song, I., Yi, J., Yi, S.-J., & Kim, S.-Y. (2020). Web-Based Visualization of Scientific Research Findings: National-Scale Distribution of Air Pollution in South Korea. *International Journal of Environmental Research and Public Health*, *17*(7). https://doi.org/10.3390/ijerph17072230
15. Qu, Z., & Hullman, J. (2018). Keeping Multiple Views Consistent: Constraints, Validations, and Exceptions in Visualization Authoring. *IEEE Transactions on Visualization and Computer Graphics*, *24*(1), 468–477. https://doi.org/10.1109/TVCG.2017.2744198
16. Robertson, H., Nicholas, N., Dhagat, A., & Travaglia, J. (2017). A Spatial Dashboard for Alzheimer's Disease in New South Wales. *Studies in Health Technology and Informatics*, *239*, 126–132.
17. Strachna, O., Cohen, M. A., Allison, M. M., Pfister, D. G., Lee, N. Y., Wong, R. J., McBride, S. M., Mohammed, R. R., Kemeny, E., Polubriaginof, F. C. G., Kassa, A., Hannon, M., & Cracchiolo, J. R. (2021a). Case study of the integration of electronic patient-reported outcomes as standard of care in a head and neck oncology practice: Obstacles and opportunities. *Cancer*, *127*(3), 359–371. https://doi.org/10.1002/cncr.33272
18. Auliya, R. S., Aknuranda, I., & Tolle, H. (2018). A Systematic Literature Review on Healthcare Dashboards Development: Trends, Issues, Methods, and Frameworks.
19. Insulin and Insulin Resistance. (n.d.). Retrieved May 25, 2021, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1204764/
20. Python for Data Science and Data Analysis. (2019, March 5). Simplilearn.Com. https://www.simplilearn.com/why-python-is-essential-for-data-analysis-article
21. The Cost of Diabetes | ADA. (n.d.). Retrieved May 25, 2021, from https://www.diabetes.org/resources/statistics/cost-diabetes
22. National Diabetes Statistical report 2020
23. https://sc-ctsi.org/