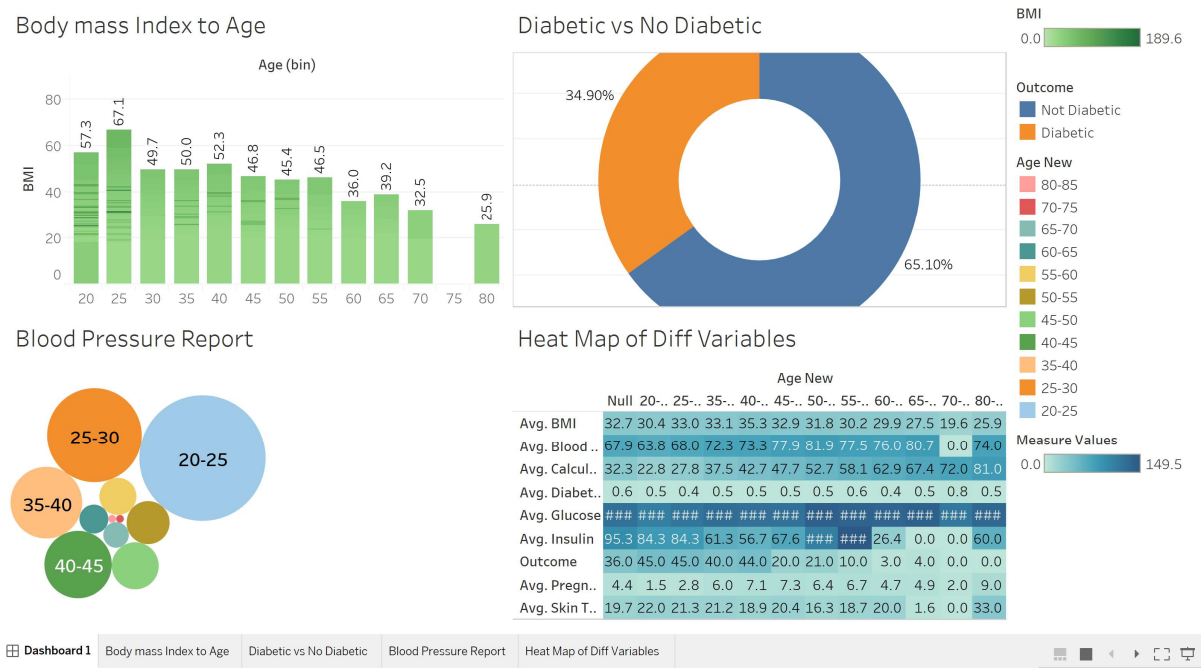
Healthcare Project –

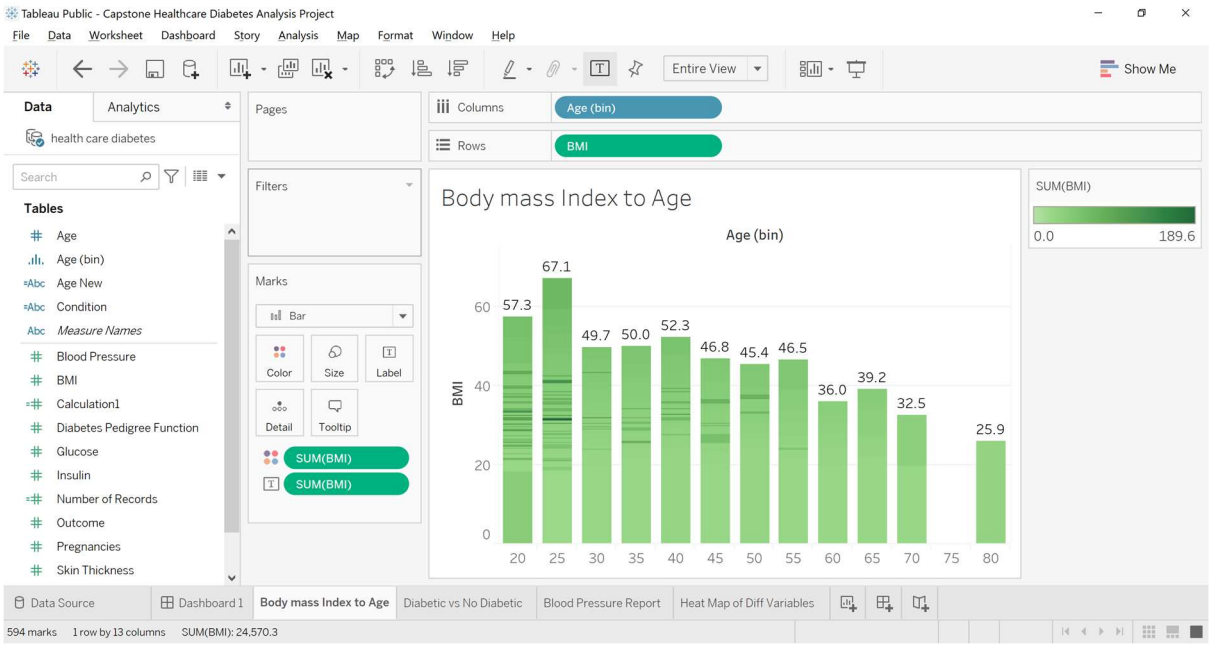
CountMeIn (github.com)

All activities have been completed and submitted as part of python code, results and code PDF and the Tableau outputs. I have attached all the outputs at the end of the file too.

All tasks that have been completed are marked as **Done**

1. **Tableau Outputs**
   Rajesh Nagarjunan - Profile | Tableau Public





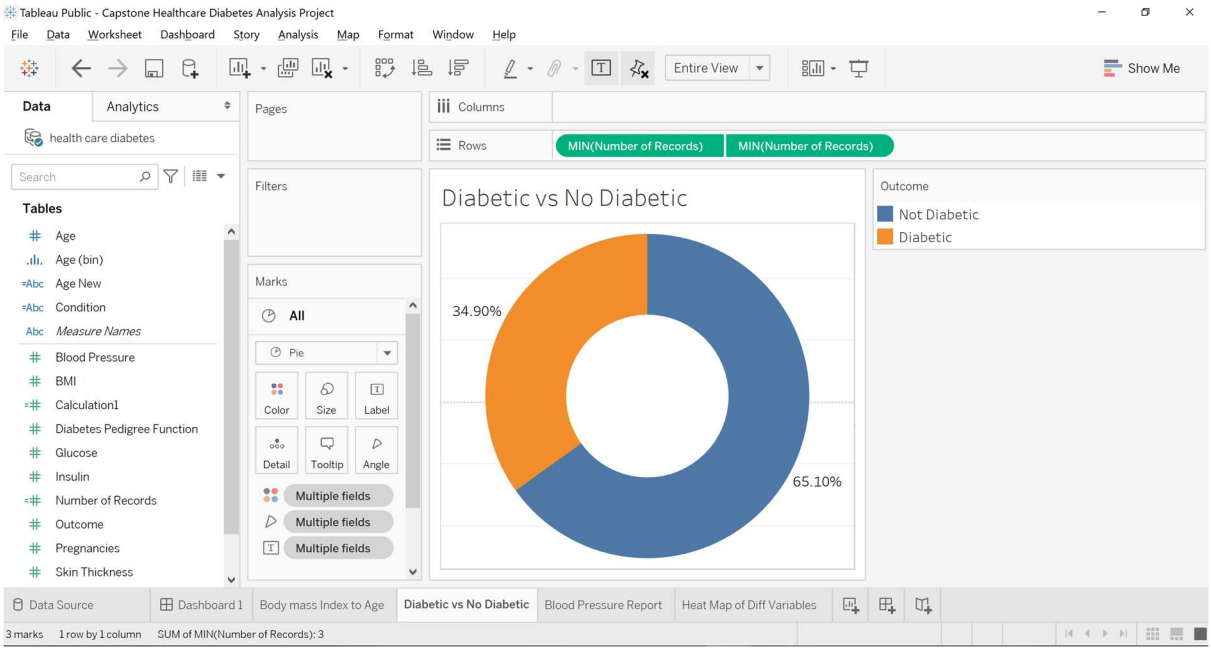


Tableau Public - Capstone Healthcare Diabetes Analysis Project
File Data Worksheet Dashboard Story Analysis Map Format Window Help
Entire View
Show Me
Data
Analytics
health care diabetes
Search
Tables
Age
Age (bin)
Age New
Condition
Measure Names
Blood Pressure
BMI
Calculation1
Diabetes Pedigree Function
Glucose
Insulin
Number of Records
Outcome
Pregnancies
Skin Thickness
Pages
Filters
Marks
All
Pie
Color
Size
Label
Detail
Tooltip
Angle
Multiple fields
Multiple fields
Multiple fields
Columns
Rows
MIN(Number of Records)
MIN(Number of Records)
Diabetic vs No Diabetic
34.90%
65.10%
Outcome
Not Diabetic
Diabetic
Data Source
Dashboard 1
Body mass Index to Age
Diabetic vs No Diabetic
Blood Pressure Report
Heat Map of Diff Variables
3 marks
1 row by 1 column
SUM of MIN(Number of Records): 3

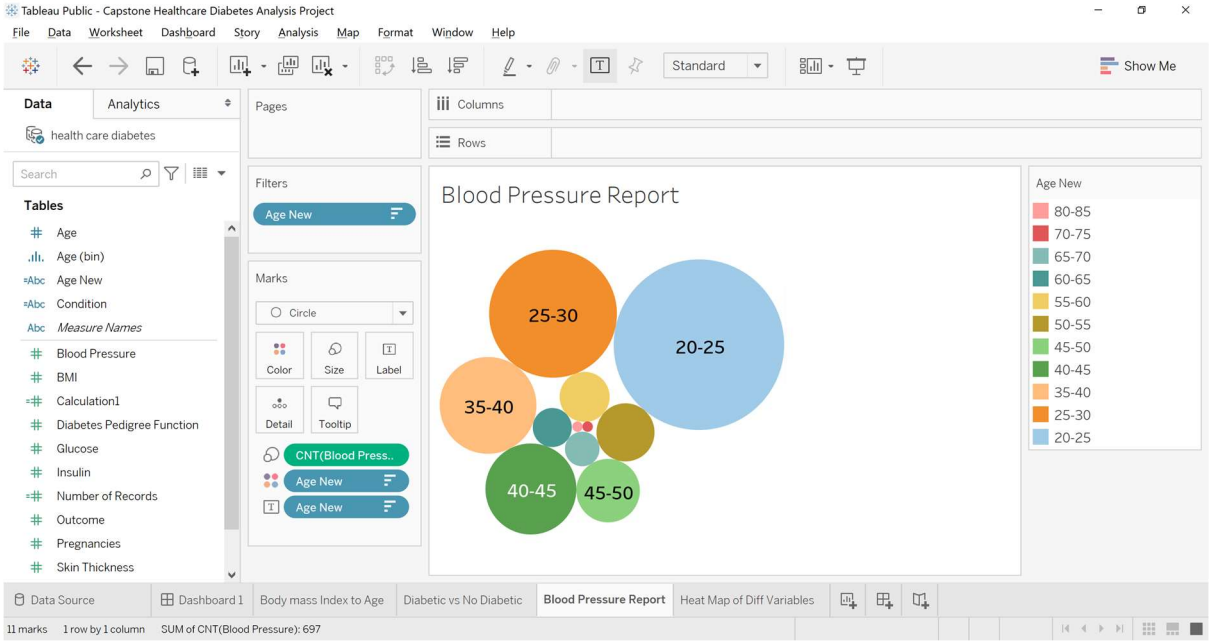
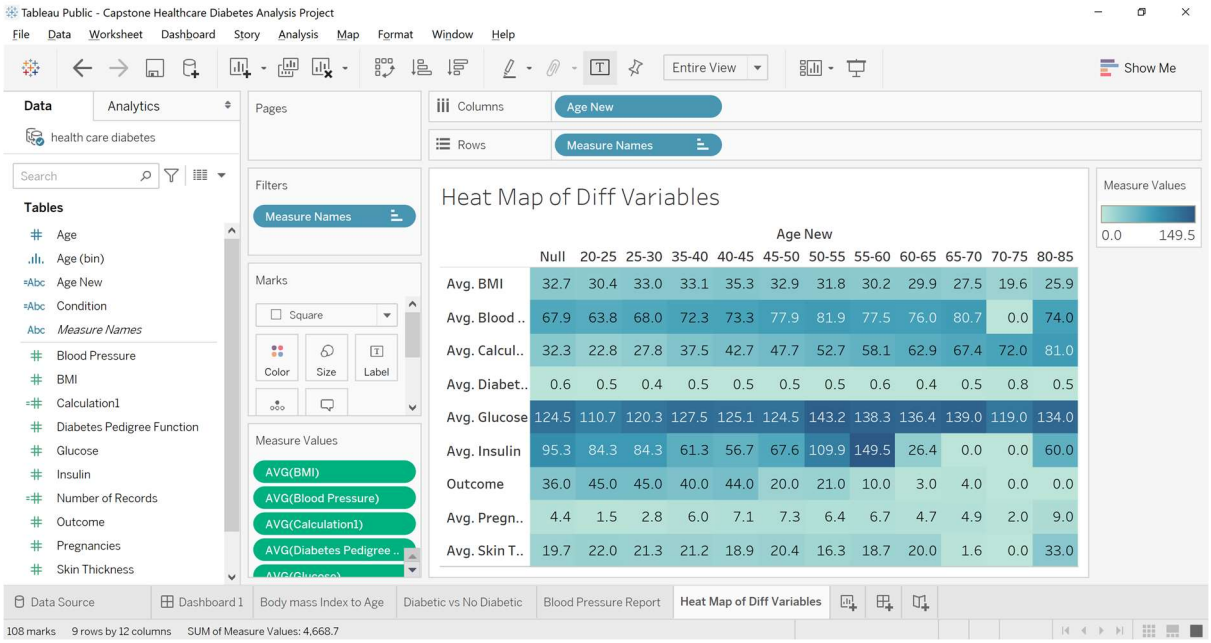


Tableau Public - Capstone Healthcare Diabetes Analysis Project
File Data Worksheet Dashboard Story Analysis Map Format Window Help
Standard
Show Me
Data
Analytics
health care diabetes
Search
Tables
Age
Age (bin)
Age New
Condition
Measure Names
Blood Pressure
BMI
Calculation1
Diabetes Pedigree Function
Glucose
Insulin
Number of Records
Outcome
Pregnancies
Skin Thickness
Pages
Filters
Age New
Marks
Circle
Color
Size
Label
Detail
Tooltip
CNT(Blood Press..
Age New
Age New
Columns
Rows
Blood Pressure Report
25-30
20-25
35-40
40-45
45-50
Age New
80-85
70-75
65-70
60-65
55-60
50-55
45-50
40-45
35-40
25-30
20-25
Data Source
Dashboard 1
Body mass Index to Age
Diabetic vs No Diabetic
Blood Pressure Report
Heat Map of Diff Variables
11 marks
1 row by 1 column
SUM of CNT(Blood Pressure): 697



2. Project Task and Completion Status

Project Task: Week 1

Data Exploration:

1. Perform descriptive analysis. Understand the variables and their corresponding values. On the columns below, a value of zero does not make sense and thus indicates missing value: - **Done**

- Glucose
- BloodPressure
- SkinThickness
- Insulin
- BMI

2. Visually explore these variables using histograms. Treat the missing values accordingly. - **Done**

3. There are integer and float data type variables in this dataset. Create a count (frequency) plot describing the data types and the count of variables. - **Done**

Project Task: Week 2

Data Exploration:

1. Check the balance of the data by plotting the count of outcomes by their value. Describe your findings and plan future course of action. - **Done**

2. Create scatter charts between the pair of variables to understand the relationships. Describe your findings. - **Done**

3. Perform correlation analysis. Visually explore it using a heat map. - **Done**

Project Task: Week 3

Data Modeling:

1. Devise strategies for model building. It is important to decide the right validation framework. Express your thought process. - **Done**

2. Apply an appropriate classification algorithm to build a model. Compare various models with the results from KNN algorithm. - **Done**

Project Task: Week 4

Data Modeling:

1. Create a classification report by analyzing sensitivity, specificity, AUC (ROC curve), etc. Please be descriptive to explain what values of these parameter you have used. - **Done**

Data Reporting:

2. Create a dashboard in tableau by choosing appropriate chart types and metrics useful for the business. The dashboard must entail the following:

a. Pie chart to describe the diabetic or non-diabetic population- **Done**

b. Scatter charts between relevant variables to analyze the relationships- **Done**

c. Histogram or frequency charts to analyze the distribution of the data- **Done**

d. Heatmap of correlation analysis among the relevant variables- **Done**

e. Create bins of these age values: 20-25, 25-30, 30-35, etc. Analyze different variables for these age brackets using a bubble chart. - **Done**