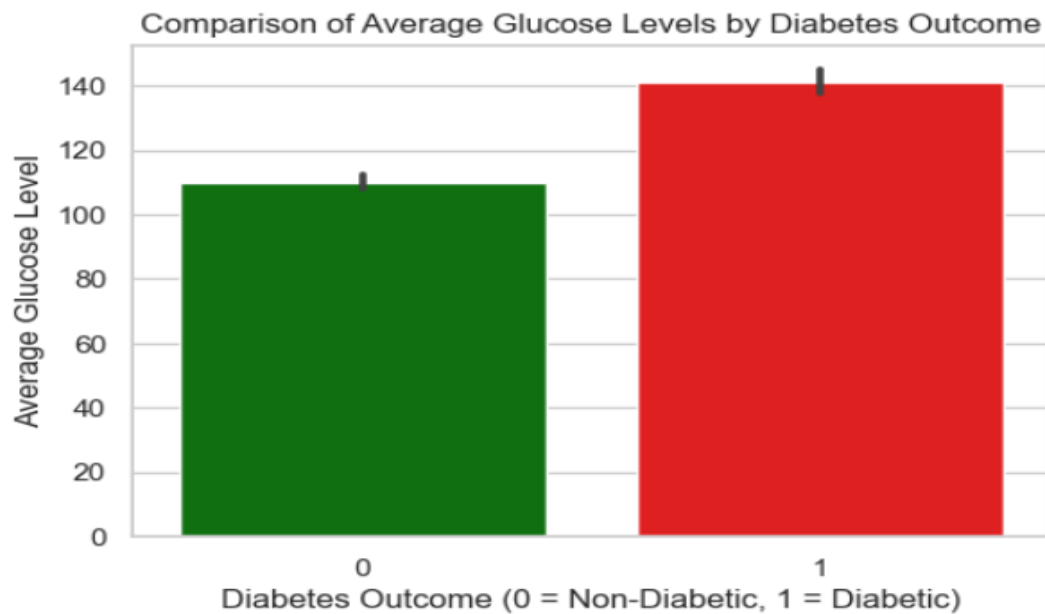


# Interactive Visualization of Diabetes Prevalence in the United States

By [Jirui Zhang]

## 1. Bar Chart: "Comparison of Average Glucose Levels by Diabetes Outcome"



Bars:

- The average glucose level data is presented through vertical bars which differentiate diabetic and non-diabetic patient measurements.
- The data shows diabetic patients in the red bars and non-diabetic patients in green bars to distinguish their results.

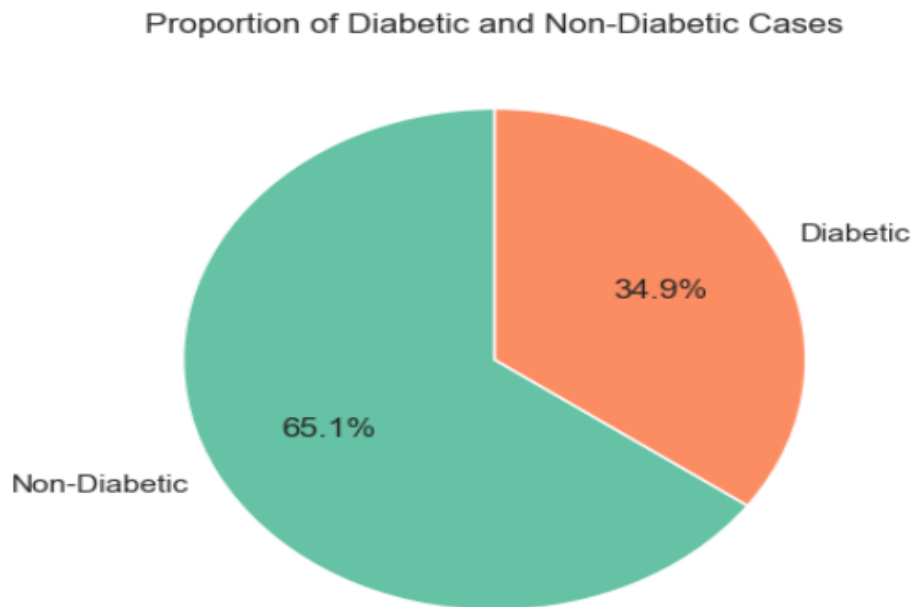
X-Axis:

- Displays the diabetes outcome (0 = Non-Diabetic, 1 = Diabetic).

Y-Axis:

- Indicates the average glucose level.

## 2. Pie Chart: "Proportion of Diabetic and Non-Diabetic Cases"



Colored Slices (Segments):

- Each segment demonstrate the percentages between diabetic and non-diabetic cases.

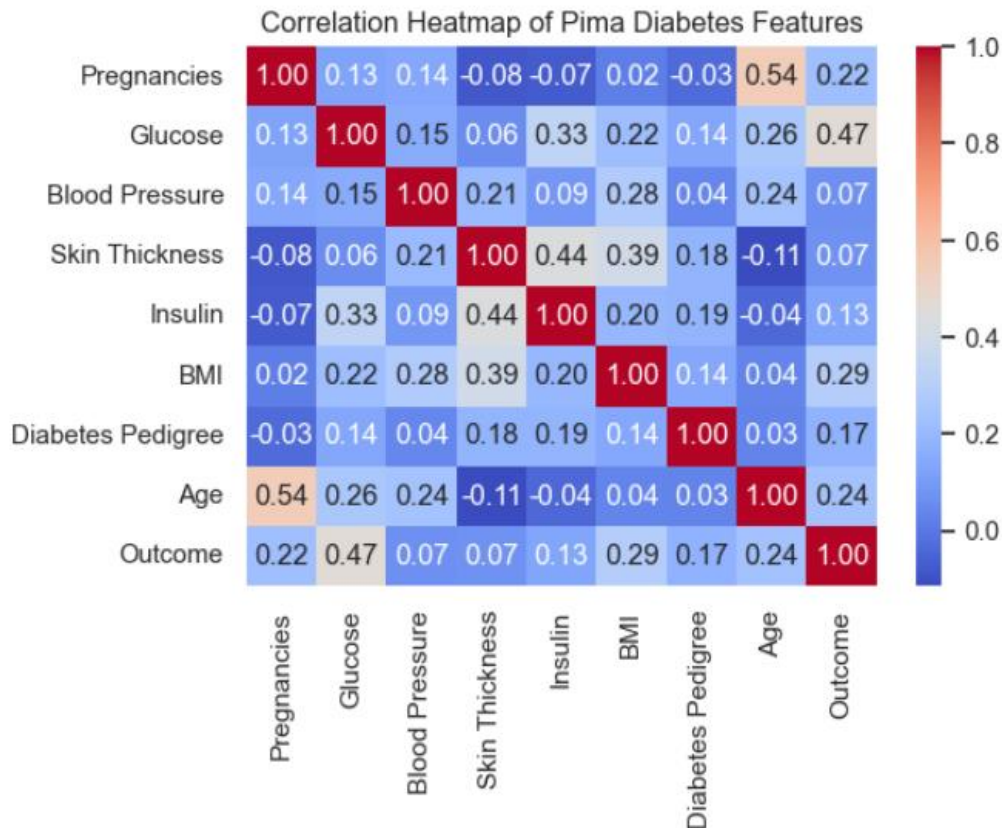
Size of Segments:

- Shows the proportional to the total number of cases in each category.

Labels:

- Display the outcome (Non-Diabetic or Diabetic) while indicating the percentage distribution of cases.

## 3. Correlation Heatmap: "Relationship Between Features"



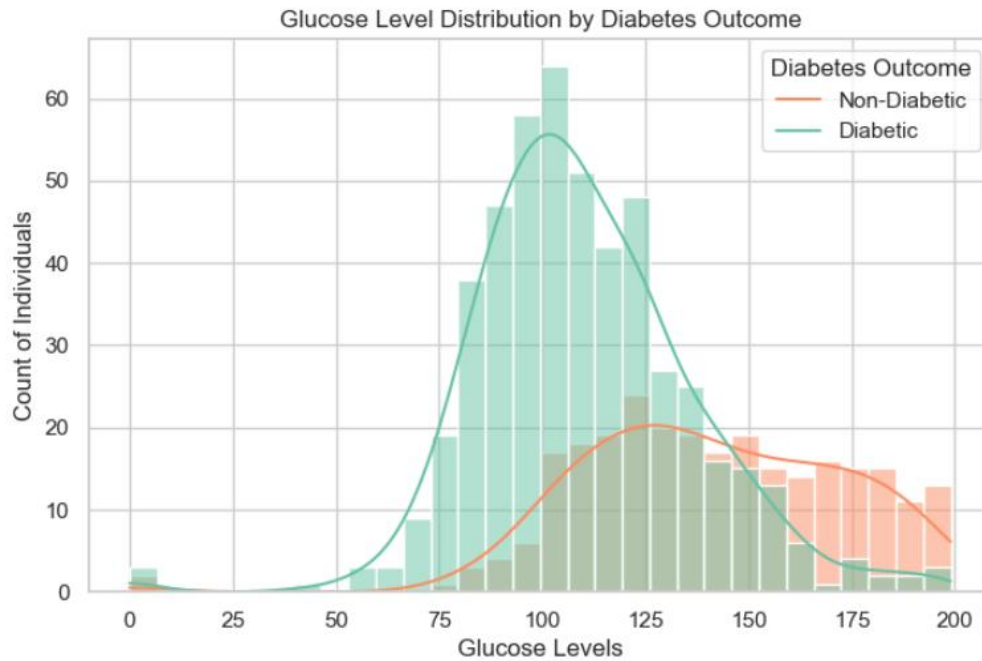
Heatmap:

- Shows correlations between variables in the dataset.
- Strong positive correlations exist between Glucose levels and the measured Outcome values.

Color Intensity:

- Intensities of color shades show correlation strength levels with light tones representing weak connections or no correlations between variables.
- The intensity of color in the visualization serves as an indicator of correlation magnitude.

#### 4. Histogram: "Glucose Level Distribution by Diabetes Outcome"



Bars:

- Each bar represents the count of individuals within a specific glucose range.
- Separate colors represent diabetic and non-diabetic groups.

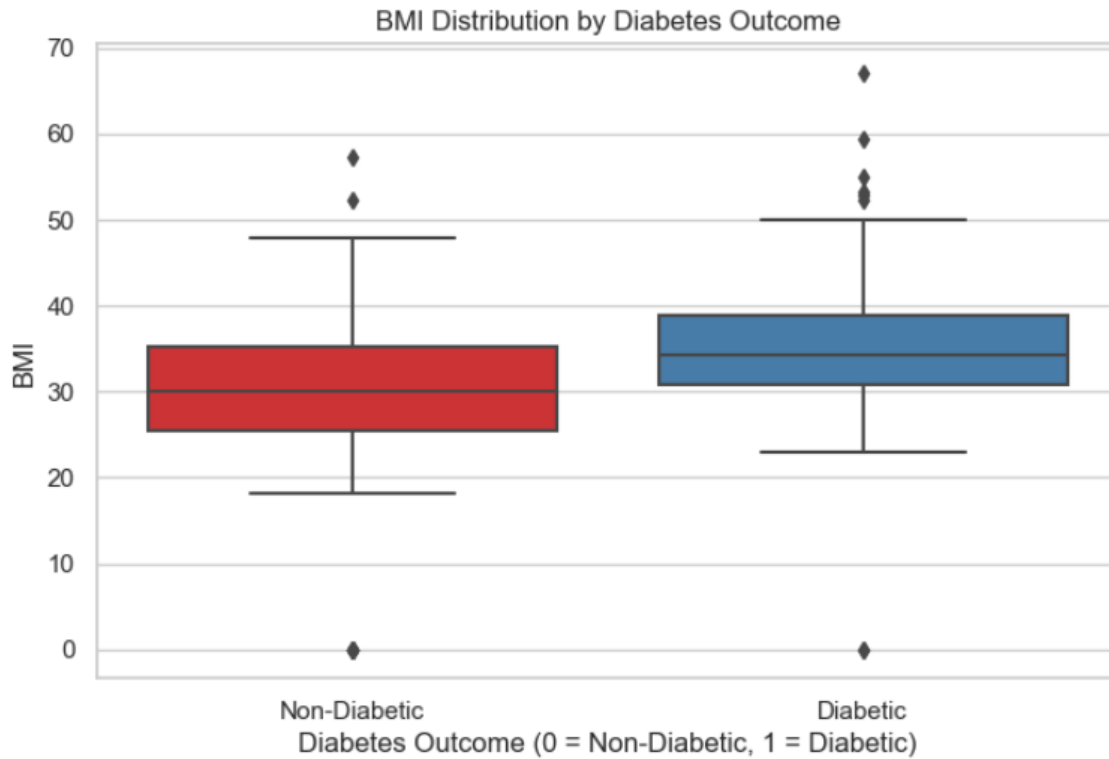
X-Axis:

- Indicates glucose levels.

Y-Axis:

- Indicates count of individuals.

## 5. Box Plot: "BMI Distribution by Diabetes Outcome"



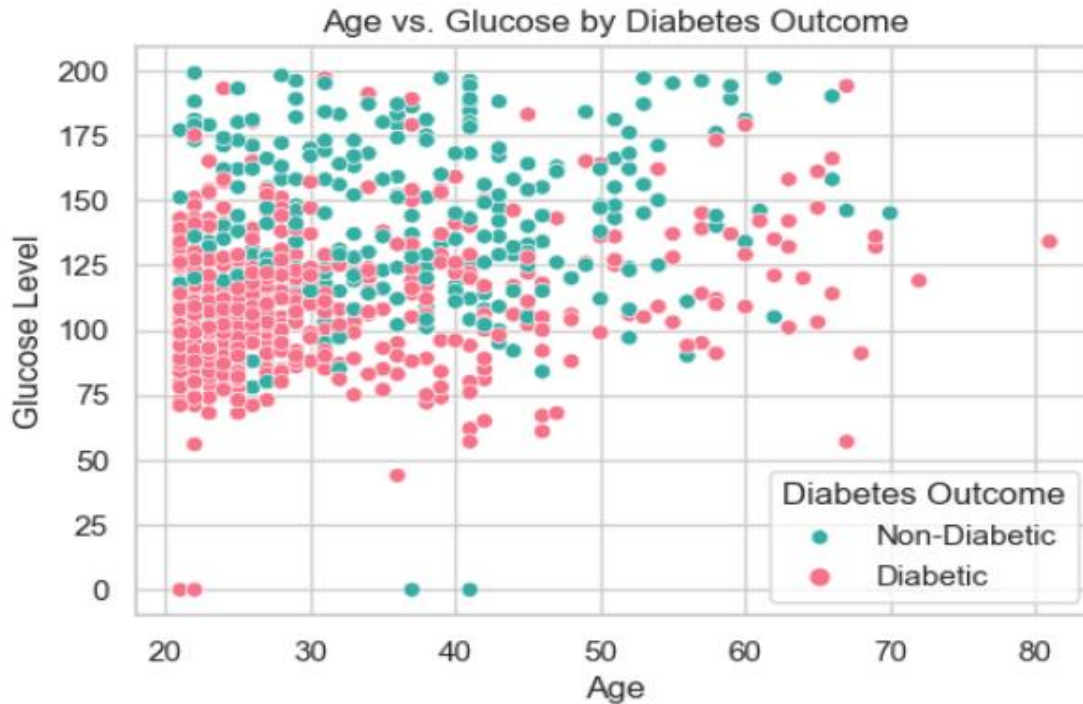
Boxes:

- Display BMI values including ranges, quartiles and median points separately for diabetic and non-diabetic subject groups.

Outliers:

- Points outside the whiskers represent data that significantly differ from the main distribution.

## 6. Scatter Plot: "Age vs. Glucose by Diabetes Outcome"



Points:

- The scatterplot indicates individual cases with age on the x-axis and glucose level on the y-axis.
- Color indicates diabetes outcome (diabetic or non-diabetic).
- Green color shows non-diabetic and red color indicates diabetic.

## 7. Findings Highlights

Summary from histogram, heatmap, and scatterplot Charts:

- Individuals with diabetes usually exhibit elevated glucose measurements and body mass index metrics.
- A distinct shift toward elevated glucose levels appears as the main feature within the histogram among diabetic patients.
- The correlation heatmap reveals that diabetes outcome shows a strong and positive relationship with glucose levels.

- The scatter plot indicates older patients with elevated glucose levels demonstrate higher susceptibility to developing diabetes.

### **Key Insights:**

- The box plot reveals diabetes prevalence increases with BMI value distribution.
- The scatter plot reveals that age by itself does not predict diabetes diagnosis effectively but high glucose levels in older patients attract attention.

Bar chart and pie chart summary:

- The blood sugar measurement of diabetic individuals proves consistently higher than the blood sugar levels of non-diabetic individuals.
- Among all detected cases approximately 34.9% prove to have diabetes.
- People with diabetes generally show elevated BMI readings which suggests that body mass index influences the chance of developing diabetes.

Key Insights from bar and pie Charts:

- The data in the bar chart shows that glucose levels differ significantly between patients who have diabetes and those without.
- The pie chart shows an overview of diabetic cases while illustrating their percentage throughout the database.
- The combination of static with interactive visualizations helps stakeholders develop better insight into diabetes patterns which enables them to outline efficient preventive strategies.

## **8. Data and Methods**

1) Data:

- The data about diabetes prevalence originated in a CSV database that provided information about patient glucose readings, BMI measurements, age, insulin, pregnancies, skin thickness, outcome, blood pressure, and diabetes diagnosis results.

## Methods:

- The researchers determined the average glucose levels between diabetic and non-diabetic patients to show differences.
- Bar chart, pie charts, histograms, and scatter plots were created through python.
- A correlation heatmap showed the relationships which exist between essential health indicators.
- The box plots and scatter plots received visualization.
- The system implements color-coded featured elements which display diabetic population cases as red and non-diabetic cases marked green.

## 9. Significance Statement

- The United States faces an increasing diabetes epidemic which causes extensive disease burden across many millions of people and creates major healthcare expenses.
- The application provides a visualization interface which converts basic statistical information into significant analysis results for tracking diabetes pattern occurrences.
- Public health stakeholders gain better knowledge about diabetes associations between glucose quantities, BMI measurements, age, insulin, pregnancies, skin thickness, outcome, and blood pressure through these visual assessments.
- Healthcare experts and policy-makers use this information for creating precise prevention and intervention plans.
- The platform provides easy-to-use interactive tools which enable users to understand connections between diabetes and essential health parameters.
- Learning about these patterns leads people toward making life changes that promote proactive health care.



- Multiple visual representation methods including bar charts, pie charts, histograms, scatter plots and heatmaps unite in this tool to deliver valuable information to research teams, healthcare experts and the general population hence facilitating data-centric choices in diabetes management.

Github Link: <https://github.com/jirui-zhang/pima-diabetes-analysis>