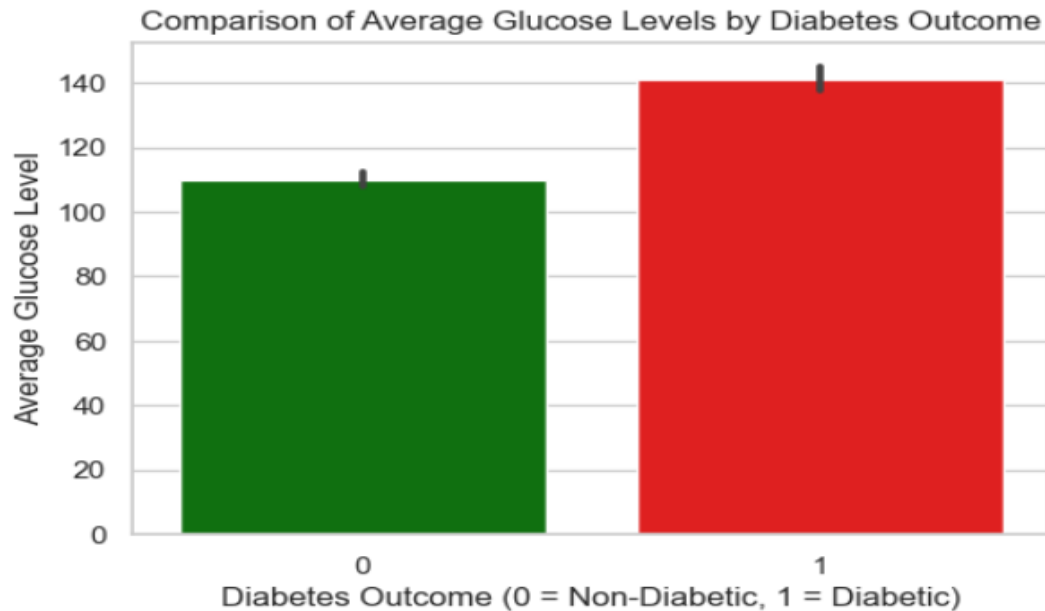


Interactive Visualization of Diabetes Prevalence in the United States

By [Jirui Zhang]

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:

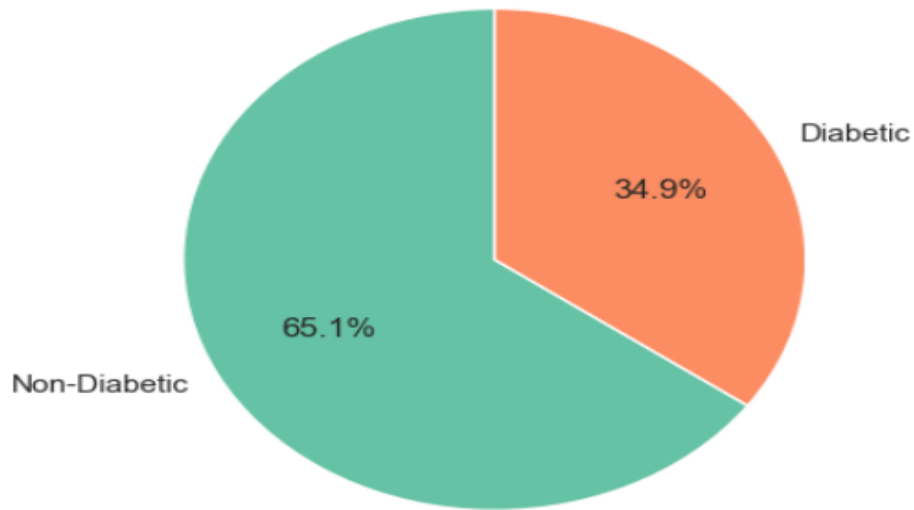
- Displays the diabetes outcome (0 = Non-Diabetic, 1 = Diabetic).

Y-Axis:

- Indicates the average glucose level.

Pie Chart: "Proportion of Diabetic and Non-Diabetic Cases"

Proportion of Diabetic and Non-Diabetic Cases



Colored Slices (Segments):

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

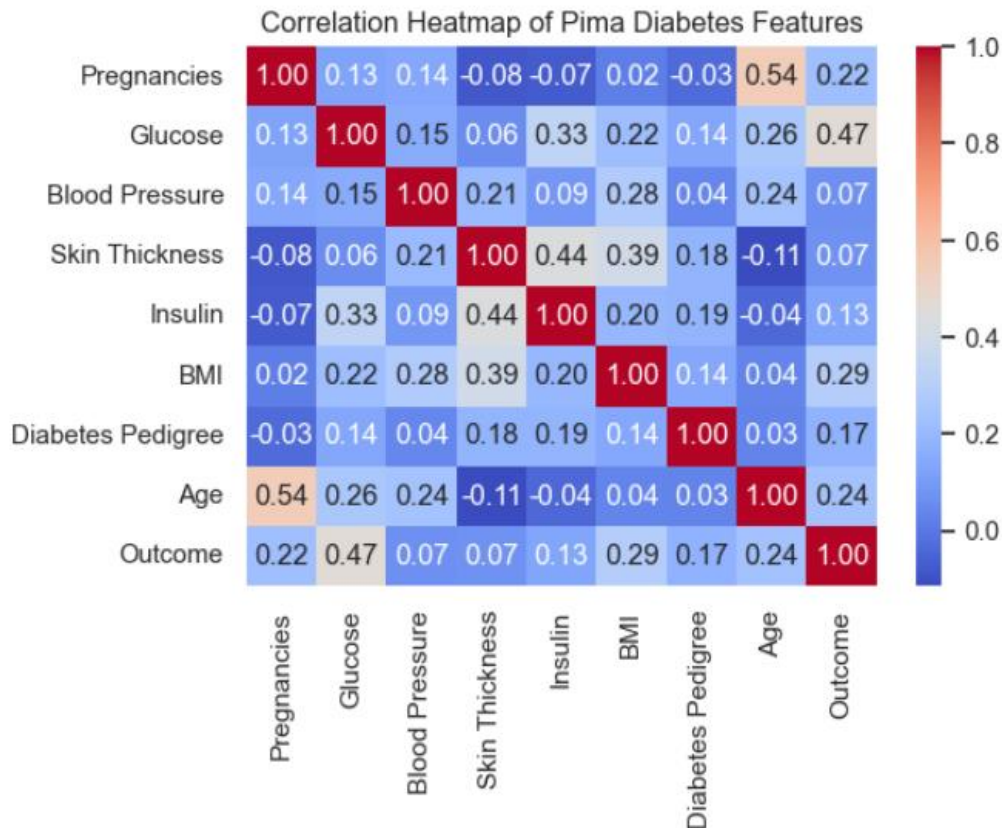
Size of Segments:

- 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.

Correlation Heatmap: "Relationship Between Features"



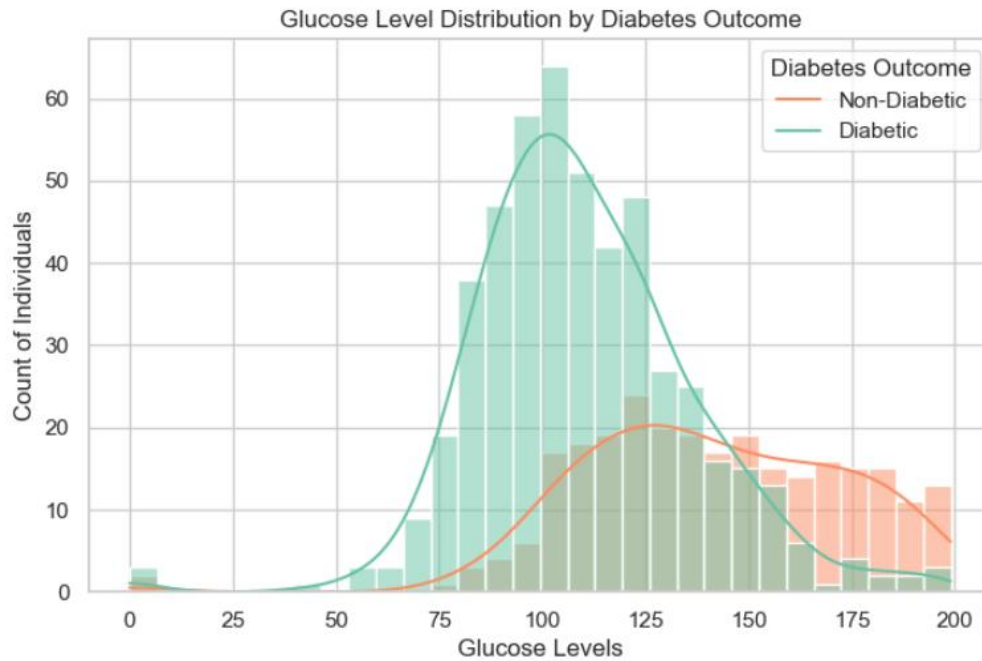
Heatmap:

- 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.

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.

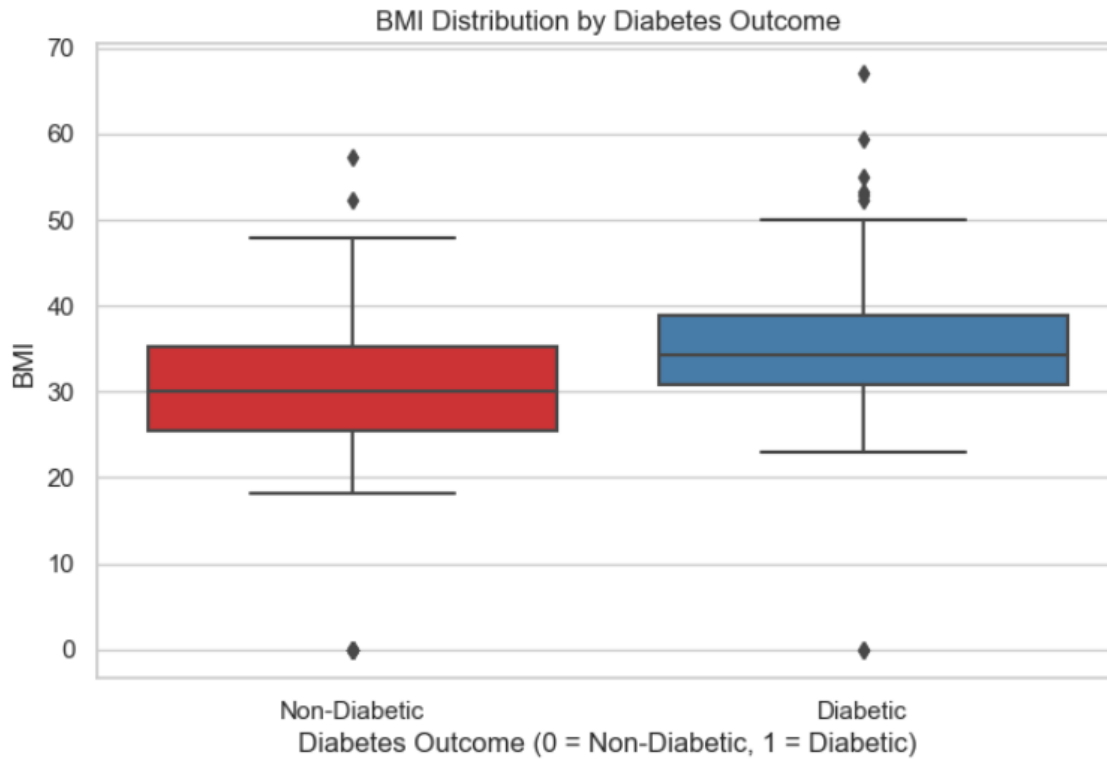
X-Axis:

- Indicates glucose levels.

Y-Axis:

- Indicates count of individuals.

Box Plot: "BMI Distribution by Diabetes Outcome"



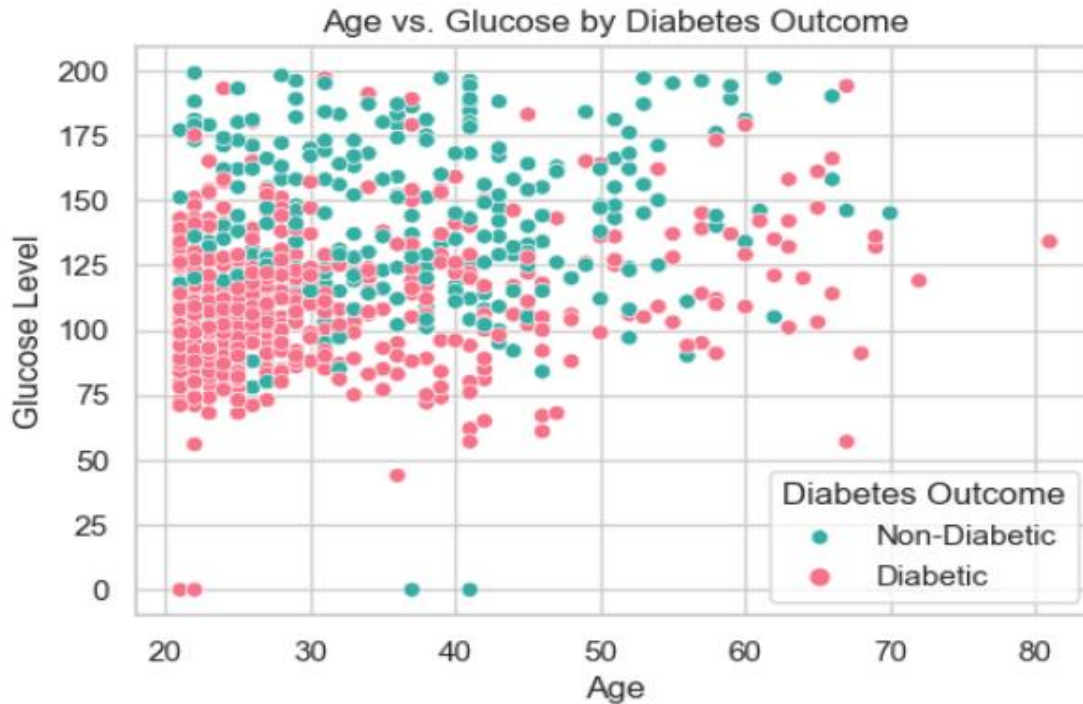
Boxes:

- 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.

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.

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.

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.

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