**AI Phase 3 Report**

**Step 1: Importing Libraries**

At the beginning of the analysis, we import the necessary Python libraries to work with data, perform visualizations, and apply data preprocessing techniques.

**Step 2: Loading the Dataset**

We load the dataset from a CSV file into a Pandas DataFrame. This DataFrame, named df, will be our primary data source for analysis and preprocessing.

**Step 3: Understanding the Dataset**

* Summary Statistics

We calculate summary statistics for the dataset. This includes measures like the mean, standard deviation, minimum, and maximum values for each numerical feature. It provides an initial understanding of the data distribution.

* Data Information

We examine the data types and check for any missing values within the dataset. This step helps us understand the structure of the data and identify if any data cleaning is needed.

* Missing Value Analysis

We explicitly check for the presence of missing values in the dataset. Fortunately, it appears that there are no missing values, which simplifies our preprocessing.

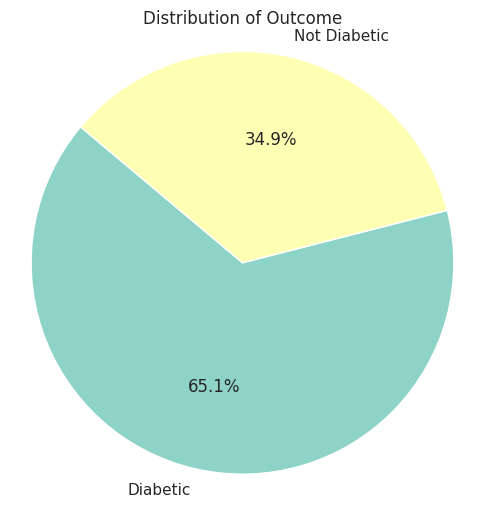
**Step 4: Visualizing the Dataset**

* Histograms of Numeric Columns

We create histograms for the numeric columns to visualize the distribution of each feature. This helps us understand the shape of the data and identify potential patterns or outliers.

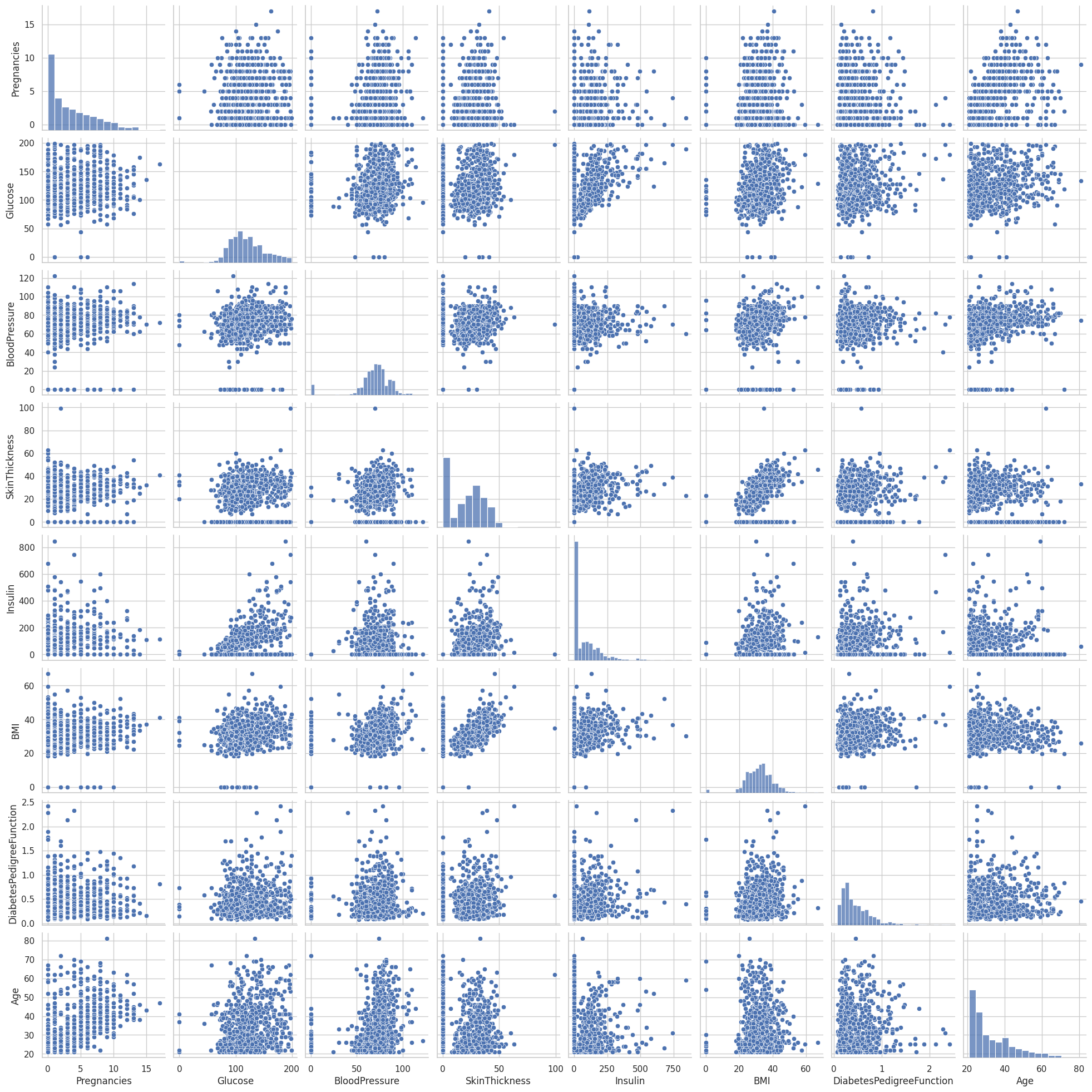
* Distribution of the Outcome

To understand the distribution of the target variable, 'Outcome', we create a pie chart. This chart visually represents the proportion of individuals with and without diabetes in the dataset.



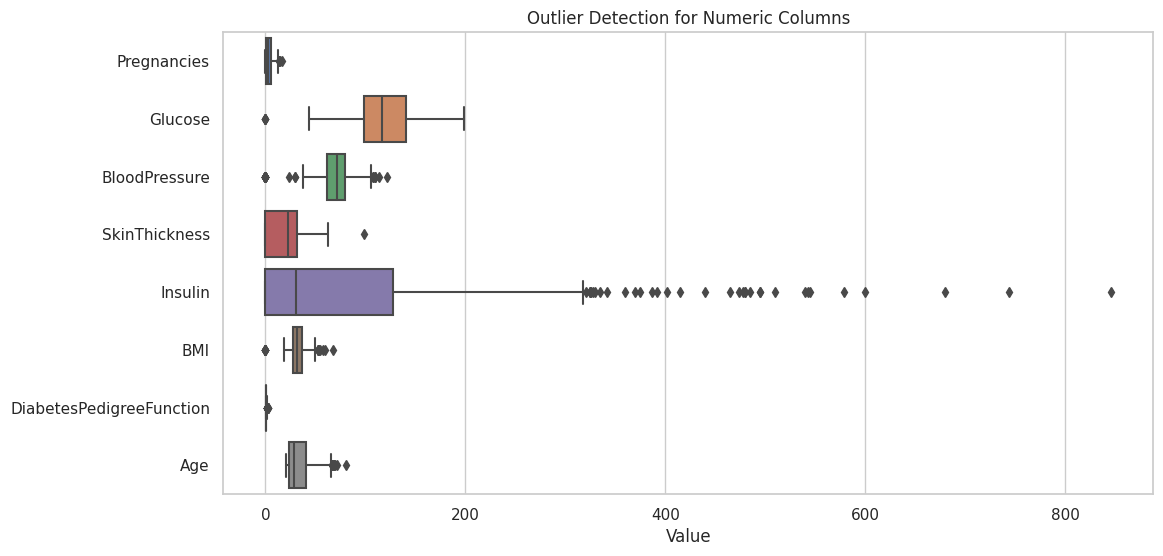
* Pairwise Relationships

A pairplot is generated to show scatterplots and histograms of numeric columns against each other. This aids in recognizing potential correlations or associations between features.



* Outlier Detection

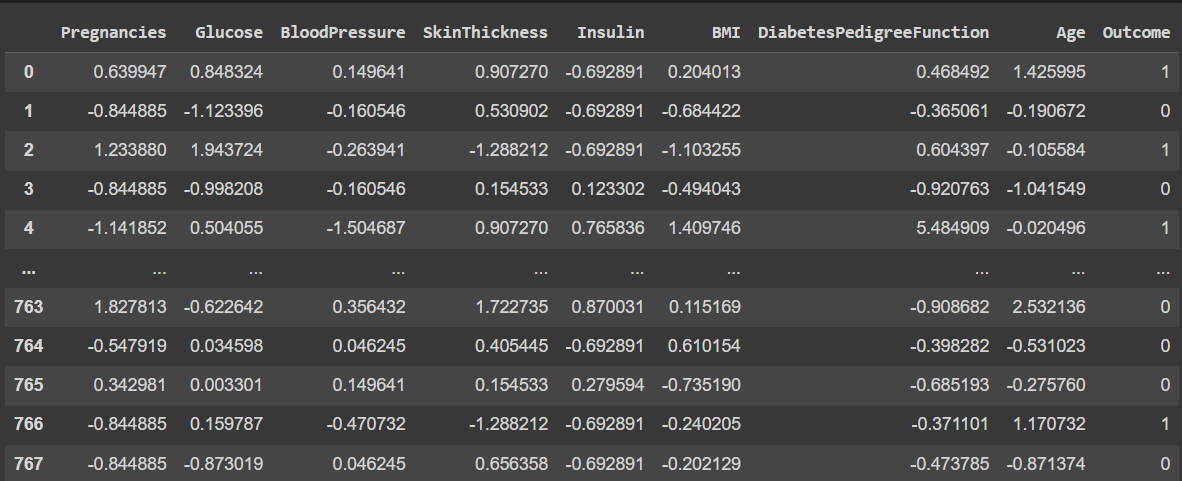
We create boxplots for the numeric columns to detect any potential outliers. Outliers are data points that significantly differ from the majority of the data and can impact the analysis.



**Step 5: Pre-Processing the Dataset**

* Scaling Numeric Features

In preparation for machine learning or further analysis, we apply two types of scaling to the numeric columns: Min-Max scaling and standard scaling. These techniques ensure that all features have similar scales, making them more suitable for many machine learning algorithms.

****