

# Healthcare Data Analysis Report

February 02, 2025

# Table of Contents

Cover.....2

Executive Summary.....3

    Overview.....3

    Key Findings.....3

    Key Conclusions.....3

1. Correlation Of Blood Glucose, Bmi, And Age With Diabetes Outcome...4

    Analysis Overview.....4

    Statistical Evidence.....4

    Conclusions and Recommendations.....4

2. Diabetes Pedigree Function Distribution And Relationship With Pregnancies Across Diabetes Outcome Groups...5

    Analysis Overview.....5

    Statistical Evidence.....5

    Conclusions and Recommendations.....5

Limitations & Next Steps.....6

    Limitations.....6

    Next Steps.....6

# Executive Summary

## Overview

This report presents a comprehensive analysis of the provided data, highlighting key patterns, trends, and actionable insights derived from the analysis.

## Key Findings

- This analysis examines the correlation between blood glucose levels, BMI, and age with diabetes outcome using descriptive statistics. The data reveals a complex relationship, with higher average glucose and BMI associated with a higher probability of a positive diabetes outcome. However, the analysis is limited by the lack of inferential statistics and visualization, preventing a definitive conclusion on the strength and significance of these correlations, and the interaction with age.
- This analysis examines the distribution of Diabetes Pedigree Function (DPF) and its relationship with the number of pregnancies in two outcome groups: individuals without diabetes (outcome\_0) and individuals with diabetes (outcome\_1). The analysis utilizes descriptive statistics (mean, median, mode) provided for both DPF and Pregnancies in each outcome group. The findings reveal significant differences in both DPF and pregnancy counts between the two groups, suggesting a potential association between these factors and the development of diabetes.

## Key Conclusions

- **Finding: High average blood glucose and BMI.**

Impact: Indicates a high prevalence of hyperglycemia and obesity, increasing the risk of diabetes.

Recommendation: Implement public health initiatives to address obesity and improve blood glucose control.

- **Finding: Significant proportion of positive diabetes outcomes.**

Impact: Highlights the need for effective diabetes prevention and management strategies.

Recommendation: Invest in early detection programs and improve access to diabetes care.

- **Finding: Higher mean DPF in the diabetic group.**

Impact: Suggests DPF could be a risk factor for diabetes.

Recommendation: Further investigate the relationship between DPF and diabetes risk using statistical tests (e.g., t-test, ANOVA) and controlling for other confounding variables.

- **Finding: Higher mean number of pregnancies in the diabetic group.**

Impact: Suggests a potential correlation between pregnancies and diabetes risk.

Recommendation: Investigate this correlation further using regression analysis to determine the strength and significance of the relationship, while accounting for other factors like age, BMI, and family history.

# 1. Correlation Of Blood Glucose, Bmi, And Age With Diabetes Outcome

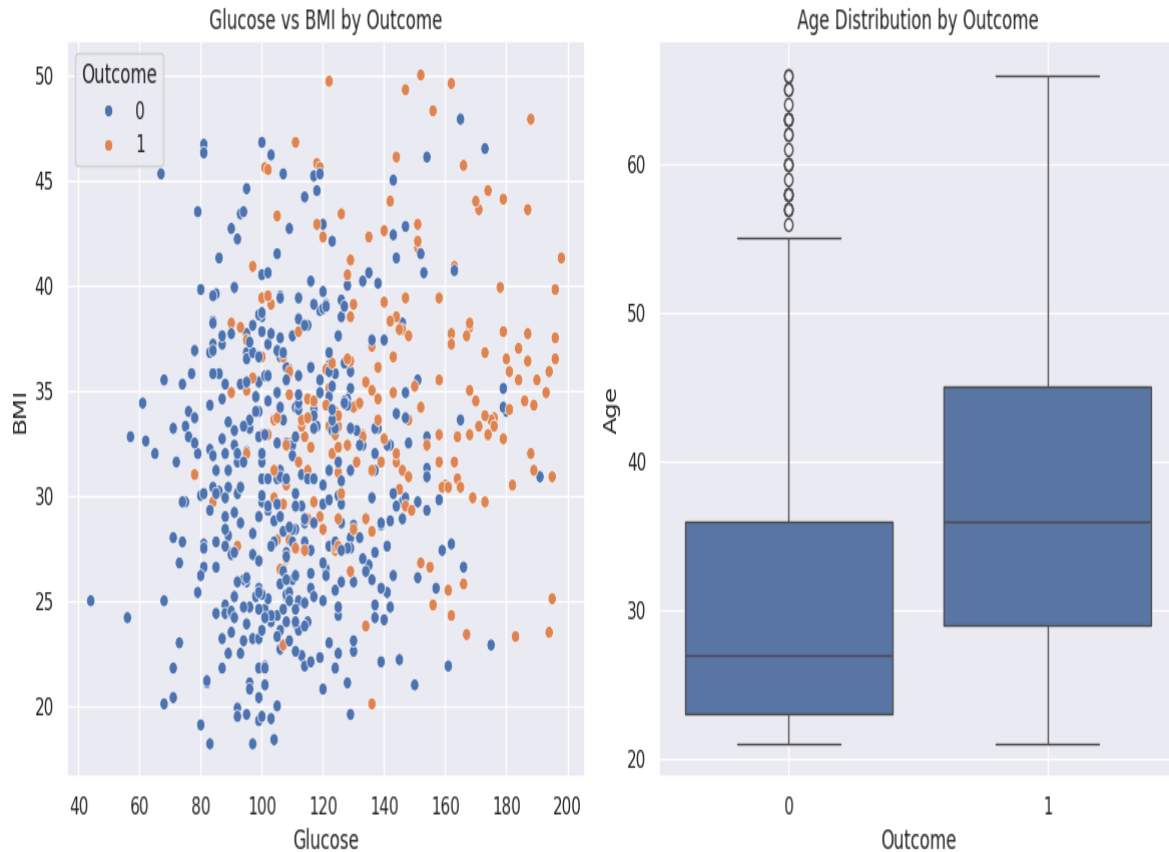


Figure 1: Correlation Of Blood Glucose, Bmi, And Age With Diabetes Outcome

## Analysis Overview

This analysis examines the correlation between blood glucose levels, BMI, and age with diabetes outcome using descriptive statistics. The data reveals a complex relationship, with higher average glucose and BMI associated with a higher probability of a positive diabetes outcome. However, the analysis is limited by the lack of inferential statistics and visualization, preventing a definitive conclusion on the strength and significance of these correlations, and the interaction with age.

- **Average Blood Glucose Level: 119.55 mg/dL (Elevated average glucose suggests a potential risk factor for diabetes.)**
- **Average BMI: 32.076 (High average BMI indicates a significant portion of the population is obese, increasing diabetes risk.)**
- **Average Age: 32.92 years (The average age provides context but doesn't reveal age-specific patterns without further analysis.)**

- **Diabetes Outcome (Proportion): 0.3196 (32%) (A substantial proportion of the population exhibits a positive diabetes outcome.)**

## Statistical Evidence

The provided data only includes descriptive statistics (mean, median, mode) for blood glucose, BMI, age, and diabetes outcome. The mean blood glucose (119.55) is above the generally accepted normal range, suggesting a high prevalence of hyperglycemia. Similarly, the mean BMI (32.076) falls within the obese range, indicating a high prevalence of obesity. The mode for diabetes outcome is 0, indicating that more individuals in the dataset do not have diabetes, but the mean of 0.3196 shows a significant proportion with a positive outcome. The median values provide additional insights into the central tendency, showing that half the population has a BMI of 32 or less and a blood glucose level of 114.5 or less. The modes suggest that 100 mg/dL is the most frequent blood glucose level and 31.6 is the most frequent BMI. However, without standard deviations or other measures of dispersion, it's impossible to assess the variability within each feature and the strength of any correlations.

## Conclusions and Recommendations

Based on the limited descriptive statistics, there's a suggestion of a positive correlation between higher blood glucose levels, higher BMI, and a positive diabetes outcome. The high prevalence of hyperglycemia and obesity in the dataset supports this observation. However, this is purely observational; further analysis is needed to confirm the strength and statistical significance of these correlations and to explore potential interactions with age. The lack of a visual representation of the data also hinders a comprehensive analysis.

- **Finding:** High average blood glucose and BMI.

**Impact:** Indicates a high prevalence of hyperglycemia and obesity, increasing the risk of diabetes.

**Recommendation:** Implement public health initiatives to address obesity and improve blood glucose control.

- **Finding:** Significant proportion of positive diabetes outcomes.

**Impact:** Highlights the need for effective diabetes prevention and management strategies.

**Recommendation:** Invest in early detection programs and improve access to diabetes care.

## 2. Diabetes Pedigree Function Distribution And Relationship With Pregnancies Across Diabetes Outcome Groups

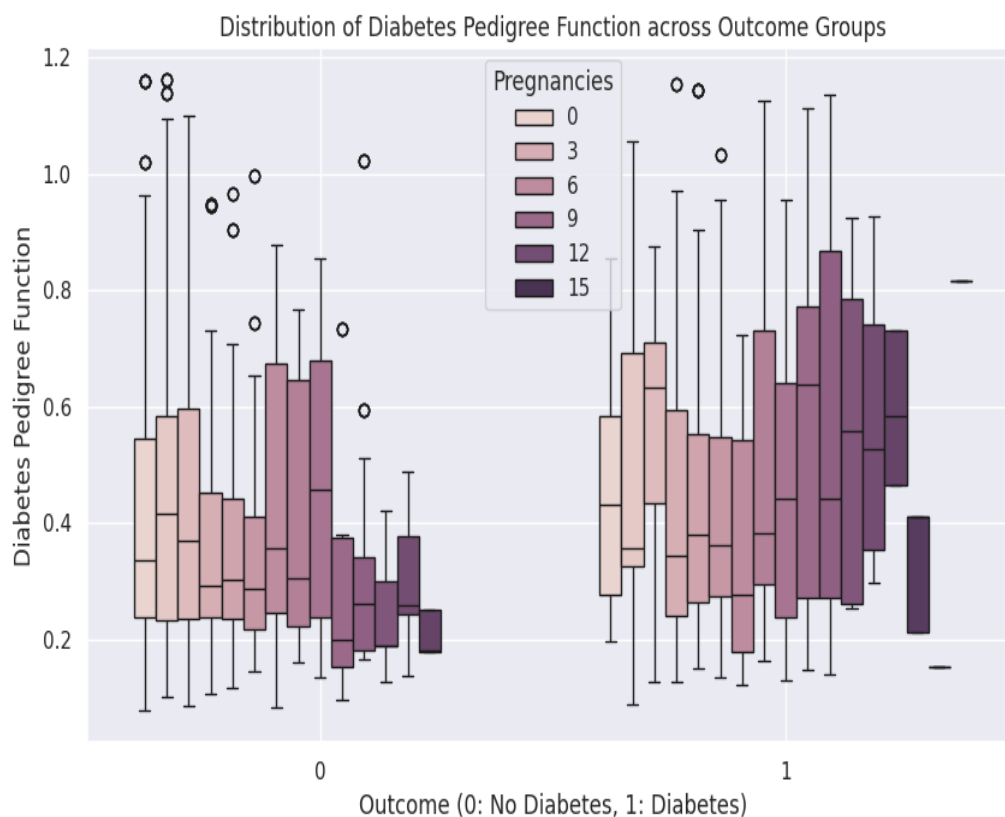


Figure 2: Diabetes Pedigree Function Distribution And Relationship With Pregnancies Across Diabetes Outcome Groups

### Analysis Overview

This analysis examines the distribution of Diabetes Pedigree Function (DPF) and its relationship with the number of pregnancies in two outcome groups: individuals without diabetes (outcome\_0) and individuals with diabetes (outcome\_1). The analysis utilizes descriptive statistics (mean, median, mode) provided for both DPF and Pregnancies in each outcome group. The findings reveal significant differences in both DPF and pregnancy counts between the two groups, suggesting a potential association between these factors and the development of diabetes.

- **Mean DPF (outcome\_0): 0.4035 (Average DPF for non-diabetic individuals.)**
- **Mean DPF (outcome\_1): 0.4794 (Average DPF is higher for diabetic individuals, suggesting a potential risk factor.)**

- **Mean Pregnancies (outcome\_0): 3.1991 (Average number of pregnancies for non-diabetic individuals.)**
- **Mean Pregnancies (outcome\_1): 4.9754 (Average number of pregnancies is significantly higher for diabetic individuals, indicating a potential correlation with diabetes.)**
- **Mode DPF (both groups): 0.0 (The most frequent DPF value is 0.0 in both groups, indicating a potential data skew or a significant number of individuals with a low DPF.)**

## Statistical Evidence

The descriptive statistics highlight a notable difference in the mean DPF between the two outcome groups. The mean DPF for individuals with diabetes (0.4794) is higher than the mean DPF for those without diabetes (0.4035). This suggests that a higher DPF might be associated with an increased risk of developing diabetes. Similarly, the mean number of pregnancies is substantially higher in the diabetic group (4.9754) compared to the non-diabetic group (3.1991). This indicates a potential correlation between a higher number of pregnancies and a greater likelihood of developing diabetes. The mode of 0.0 for DPF in both groups suggests a right-skewed distribution, meaning that a large portion of the data is concentrated at lower values of DPF. Further investigation into the distribution using histograms or other visualization techniques would be beneficial to confirm this and understand the underlying patterns better.

## Conclusions and Recommendations

Based on the provided data, there is evidence suggesting a potential association between higher DPF values, a greater number of pregnancies, and the development of diabetes. However, correlation does not equal causation. Further analysis is needed to establish a causal relationship and to understand the underlying mechanisms.

- **Finding:** Higher mean DPF in the diabetic group.

**Impact:** Suggests DPF could be a risk factor for diabetes.

**Recommendation:** Further investigate the relationship between DPF and diabetes risk using statistical tests (e.g., t-test, ANOVA) and controlling for other confounding variables.

- **Finding:** Higher mean number of pregnancies in the diabetic group.

**Impact:** Suggests a potential correlation between pregnancies and diabetes risk.

**Recommendation:** Investigate this correlation further using regression analysis to determine the strength and significance of the relationship, while accounting for other factors like age, BMI, and family history.





# Limitations & Next Steps

## Limitations

- Only descriptive statistics are provided; inferential statistics are needed to confirm significance.
- The sample size and its representativeness are unknown.
- Other potential confounding variables are not considered in this analysis (e.g., age, BMI, physical activity, diet).
- The mode of 0.0 for DPF in both groups suggests potential data skew or outliers that need further investigation.
- Age-specific patterns cannot be determined without further analysis.
- Only descriptive statistics are provided; inferential statistics (e.g., correlation coefficients, regression analysis) are needed to determine the strength and significance of relationships.
- The dataset's size and sampling method are unknown, potentially affecting the generalizability of the findings.
- No visualization is provided, limiting the ability to identify patterns and potential outliers.

## Next Steps

- Perform regression analysis to model the impact of glucose, BMI, and age on diabetes outcome.
- Assess the dataset's size and sampling method to evaluate the generalizability of the results.
- Assess the sample size and its representativeness to ensure the generalizability of the findings.
- Visualize the data using histograms and box plots to understand the distribution of DPF and Pregnancies in each group.
- Conduct inferential statistical tests (t-tests, ANOVA, regression analysis) to determine the statistical significance of the observed differences and correlations.
- Stratify the analysis by age group to investigate age-specific effects.
- Conduct correlation analysis to quantify the relationships between variables.
- Include other relevant variables in the analysis to control for confounding factors and build a more comprehensive model.

- Create visualizations (scatter plots, box plots) to explore the data and identify patterns.