Understanding Diabetes Risk Through Data Analytics

An Interactive Dashboard Report for Strategic Health Insights

Prepared by: Calista Dikeh

April 2025

Introduction

Diabetes is a chronic, lifelong condition that affects how the body regulates blood sugar (glucose). It has become one of the most pressing public health challenges of the 21st century. With its complications ranging from cardiovascular disease to kidney failure, blindness, and amputations, diabetes places a significant burden on healthcare systems and economies worldwide. As global prevalence continues to rise, driven by sedentary lifestyles, high-calorie diets, increased stress levels, and ageing populations, the need for early diagnosis and prevention strategies has never been more critical.

This report provides a data-driven analysis of patient health records using an interactive Power BI dashboard. The dataset includes multiple biometric and demographic attributes, including age, body mass index (BMI), glucose concentration, insulin levels, and pregnancy history. By transforming raw clinical data into meaningful visual insights, the dashboard enables healthcare professionals, public health analysts, and policymakers to understand the underlying patterns associated with diabetes risk.

Through dynamic visualisations and segmentation of patient groups, we aim to uncover correlations between patient characteristics and diabetes outcomes. These insights can be used to develop targeted intervention strategies, optimise resource allocation in screening programs, and enhance patient education efforts. Ultimately, the goal is to support data-informed decision-making that helps reduce the incidence and impact of diabetes through early identification, lifestyle adjustments, and proactive clinical engagement.

Dataset Overview

Dataset Summary

- **Source**: Diabetes Dataset from Kaggle
- **Total Records**: 768 patients
- **Variables**: 9 health-related features
- **Data Type**: Tabular
- **Patient Demographics**: Female patients aged 21–81

Key Columns Analysed

Column	Description	Type
Pregnancies	Number of pregnancies	Numeric
Glucose	Plasma glucose concentration	Numeric
BloodPressure	Diastolic blood pressure	Numeric
SkinThickness	Triceps skin fold thickness	Numeric
Insulin	2-Hour serum insulin	Numeric
BMI	Body mass index	Numeric
DiabetesPedigree	Genetic predisposition	Numeric
Age	Patient age	Numeric
Outcome	1 = Diabetic, 0 = Non-Diabetic	Binary Categorical

Data Quality Assessment

To ensure the integrity of the dashboard analysis, data preprocessing was performed to clean and validate records.

Data Transformation Highlights

1. **Missing Values**:

- No nulls were present in the dataset, but zero entries for all columns except pregnancies were replaced with column medians to avoid distortion of health metrics

2. **Normalization**:

- All numeric features were scaled where appropriate for visual clarity.

3. **Categorical Splits**:

- Patients were grouped by outcome (Diabetic vs Non-Diabetic) and segmented further by age, pregnancy history, and glucose range.

4. **New Derived Columns:**

- Age brackets: 20–30, 31–40, ..., 71–81
- Pregnancy Categories: 0–2, 3–5, 6+
- Glucose and insulin groupings were created for clearer pattern detection in scatter and box plots.

Key Performance Indicators (KPIS)

KPI	Value
Total Patients	768
Percentage Diabetic	34.9%
Average Glucose Level	121.68 mg/dL
Average BMI	32.45
Minimum Patient Age	21 years
Maximum Patient Age	81 years

Dashboard Objectives

- To assess the prevalence of diabetes within the population.
- To explore age distribution and its correlation with diabetes.
- To examine the influence of pregnancy history on diabetes risk.
- To analyse glucose levels as a predictive factor for diabetes.
- To study the relationship between BMI and age in the context of diabetes.

- To understand insulin value patterns across diagnosis groups.
- To investigate the relationship between glucose and insulin levels in patients.

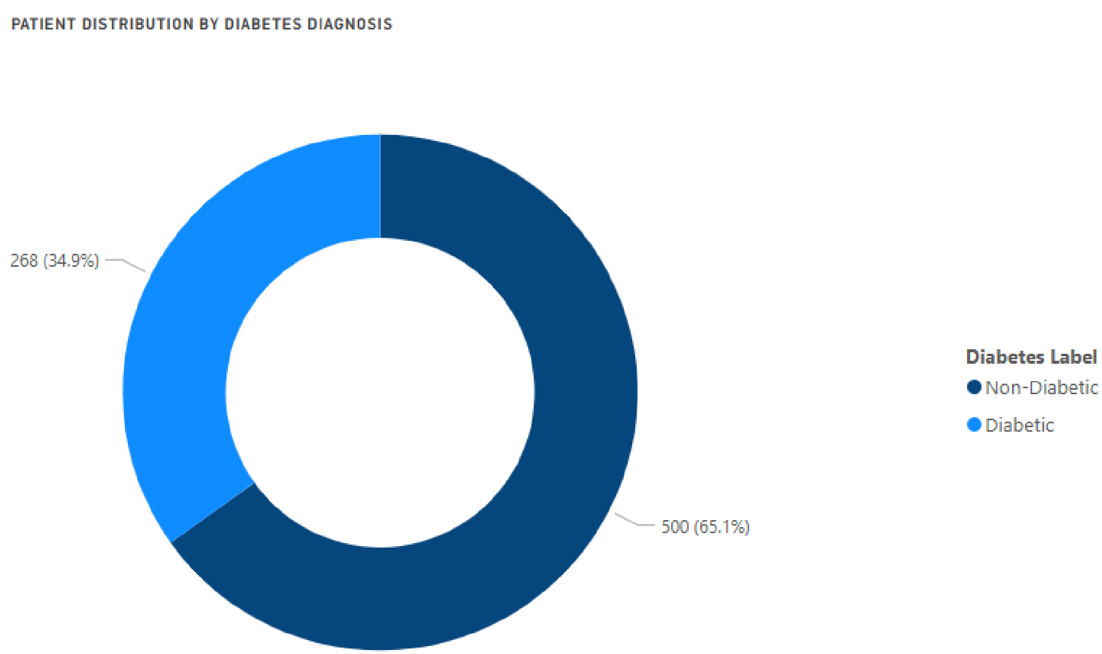
Analysis and Visual Insights

Objective 1: Diabetes Prevalence in the Population

- 65.1% of the patients are Non-Diabetic(Dark blue)
- 34.9% are Diabetic(Sky blue)

Visual:

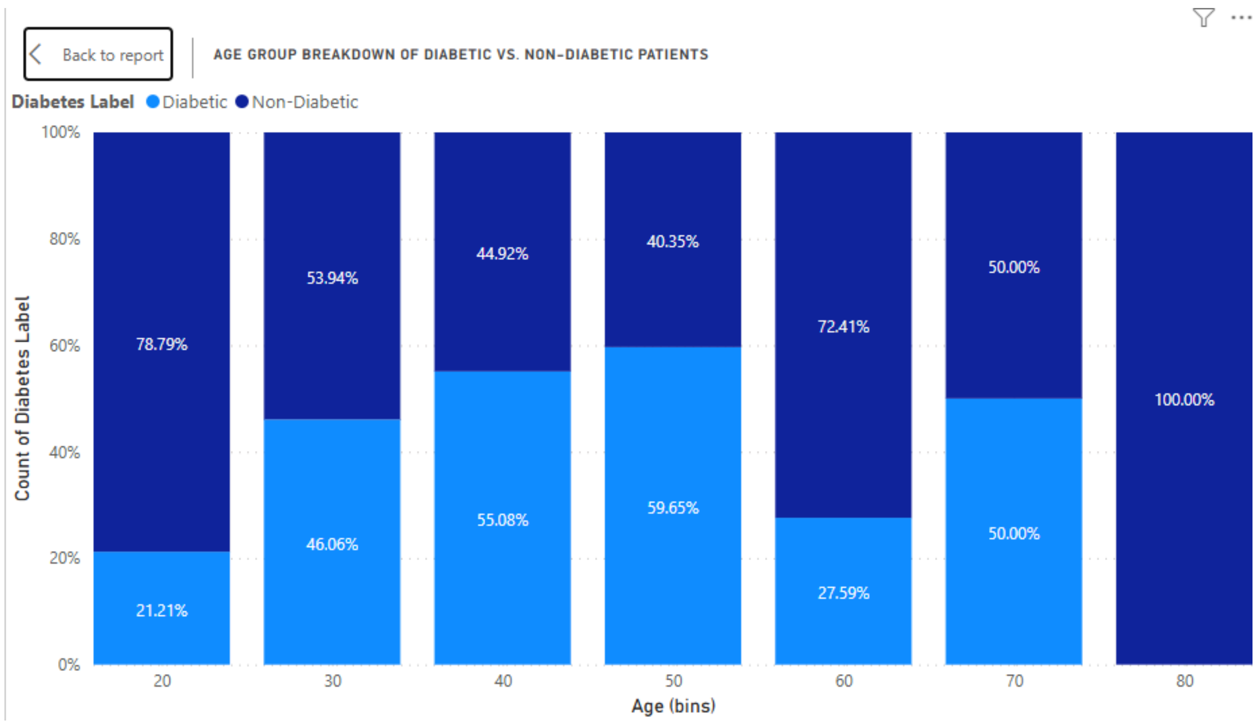
Pie chart showing overall diabetes distribution.



Objective 2: Age Distribution and Diabetes Risk

- Younger individuals (20–40) are predominantly non-diabetic.
- As patients age, especially between 41 – 60, the percentage of diabetics increases notably.
- By age 80, the diabetic vs. non-diabetic split becomes equal.

Visuals: Stacked bar chart: Diabetes by age group

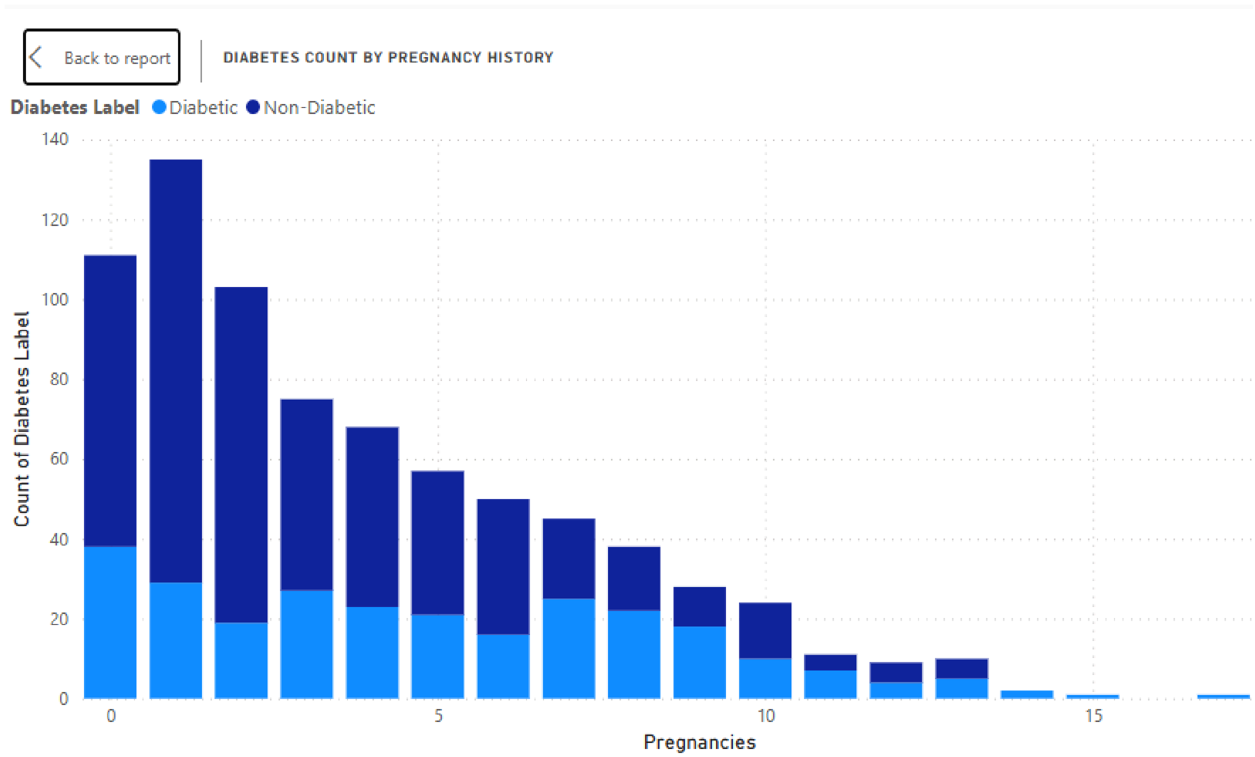


Objective 3: Pregnancy and Diabetes Correlation

- Higher pregnancy counts correlate with increased diabetes risk.
- Diabetics are more frequent in groups with 3+ pregnancies.
- Non-diabetics are distributed more evenly across all pregnancy categories.

Visuals:

Column chart: Diabetes vs. pregnancy count

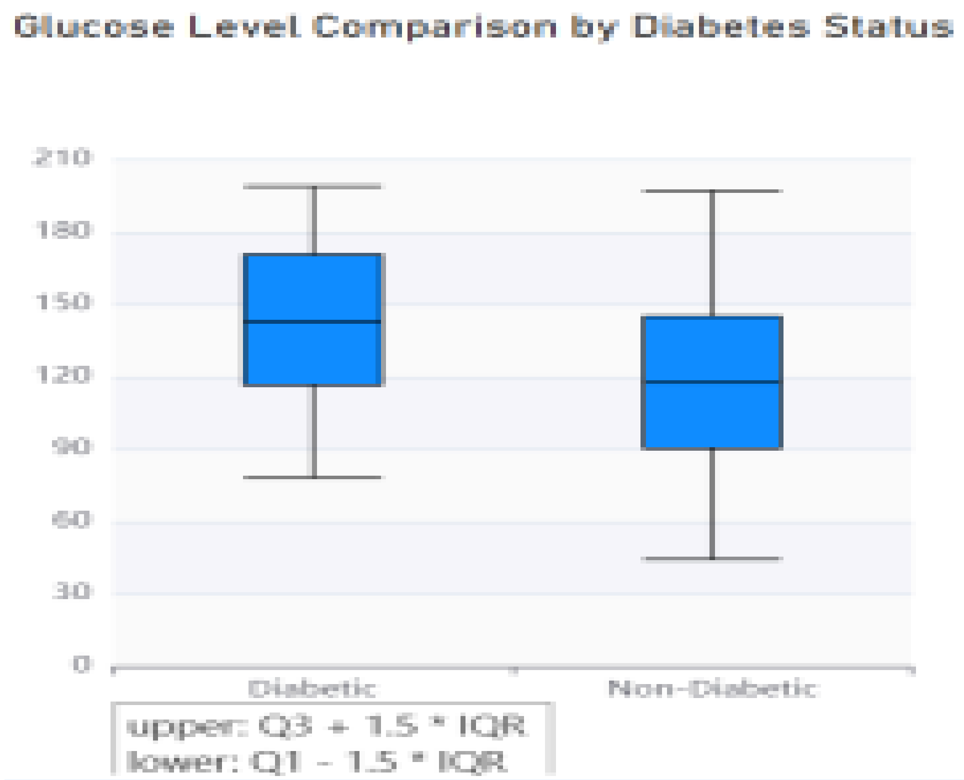


Objective 4: Glucose Levels as an Indicator

● Diabetic patients have higher median and wider spread of glucose values.

● Non-diabetics have fewer extreme glucose readings, reinforcing glucose as a key indicator.

Visual: Box plot comparing glucose levels by diabetes status



Objective 5: Body Mass Index (BMI) and Age Risk

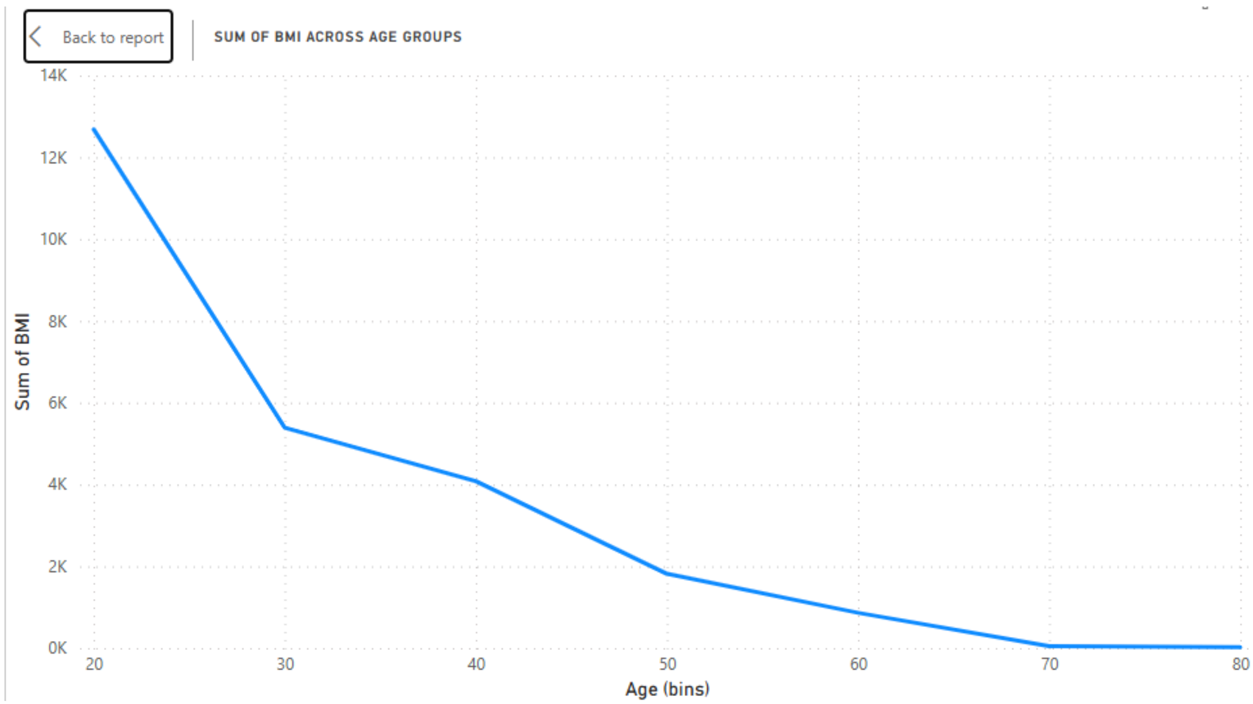
- Highest total BMI values are observed in the 20–30 age group.

- BMI contribution declines as age increases.

- Obesity in younger populations may signal early diabetes risk.

Visual:

Line chart: Sum of BMI by age group



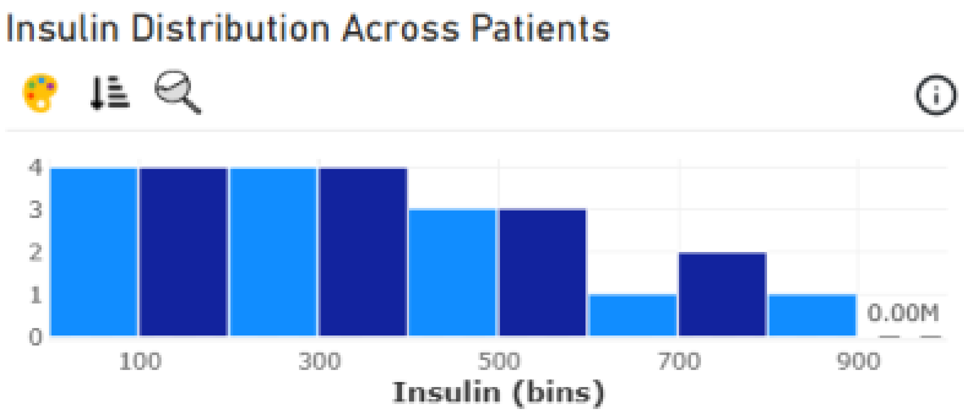
Objective 6: Insulin Patterns Across Diagnosis Groups

- Diabetic patients exhibit a broader and higher insulin range.
- Some diabetic patients exceed 800+ insulin units — flagged as potential clinical outliers.

- Non-diabetics typically stay below 300 insulin units.

Visuals:

Histogram: Insulin values by diabetes outcome

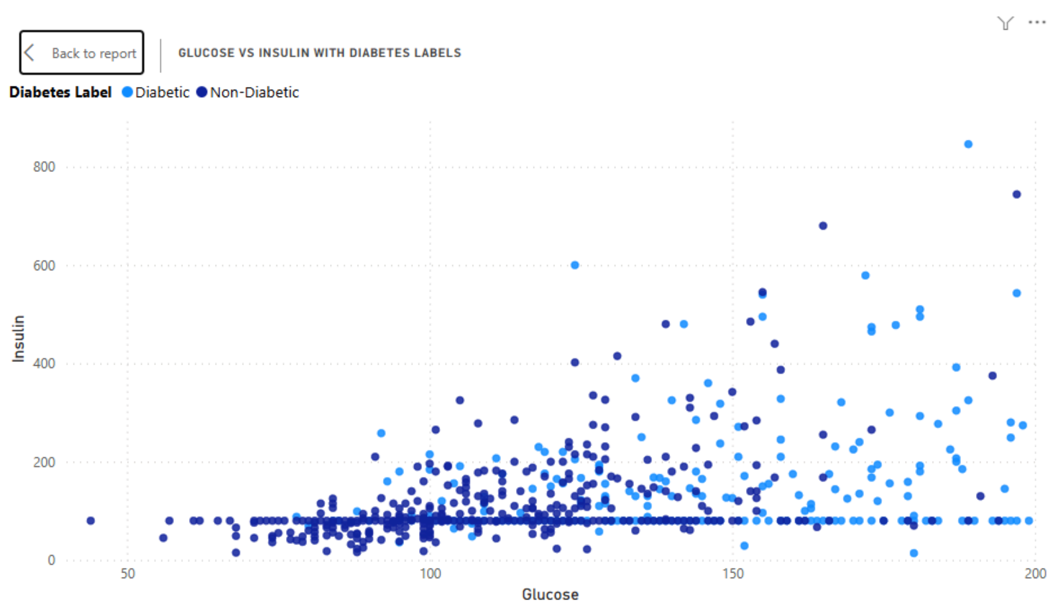


Objective 7: Relationship Between Glucose and Insulin

- Diabetics cluster in the upper-right quadrant of the glucose-insulin scatter plot.
- Non-diabetics remain in the lower-left region.
- This confirms a strong interplay between insulin and glucose levels in identifying risk.

Visual:

Scatter plot: Glucose vs. Insulin by Diabetes Label



Strategic Recommendations

1. **Proactive Monitoring**
    - ○ Target early screenings in patients aged 30–50 with high BMI or glucose levels.
2. **Pregnancy-Focused Screening**
    - ○ Monitor women with 3+ pregnancies for early signs of diabetes.
3. **Clinical Threshold Alerts**
    - ○ Flag insulin readings above 300 for deeper clinical review and intervention.
4. **Enhanced Risk Models**
    - ○ Use machine learning models based on glucose, insulin, BMI, and pregnancy history to predict future diabetes cases.
5. **Expand Variables**
    - ○ Integrate additional features like blood pressure, genetic predisposition, and lifestyle factors for better accuracy

Conclusion

This dashboard presents a powerful tool for understanding and predicting diabetes risks based on patient health profiles. It confirms well-known medical patterns, such as the role of glucose and BMI, while also highlighting the significance of pregnancy history and insulin metrics.

With proper integration into health systems, such dashboards can drive proactive care, reduce late diagnoses, and ultimately help lower the public health burden of diabetes.