Exploratory Analysis of Diabetes Health Indicators

October 2, 2025

1 Diabetes Health Indicators Analysis

1.1 Background

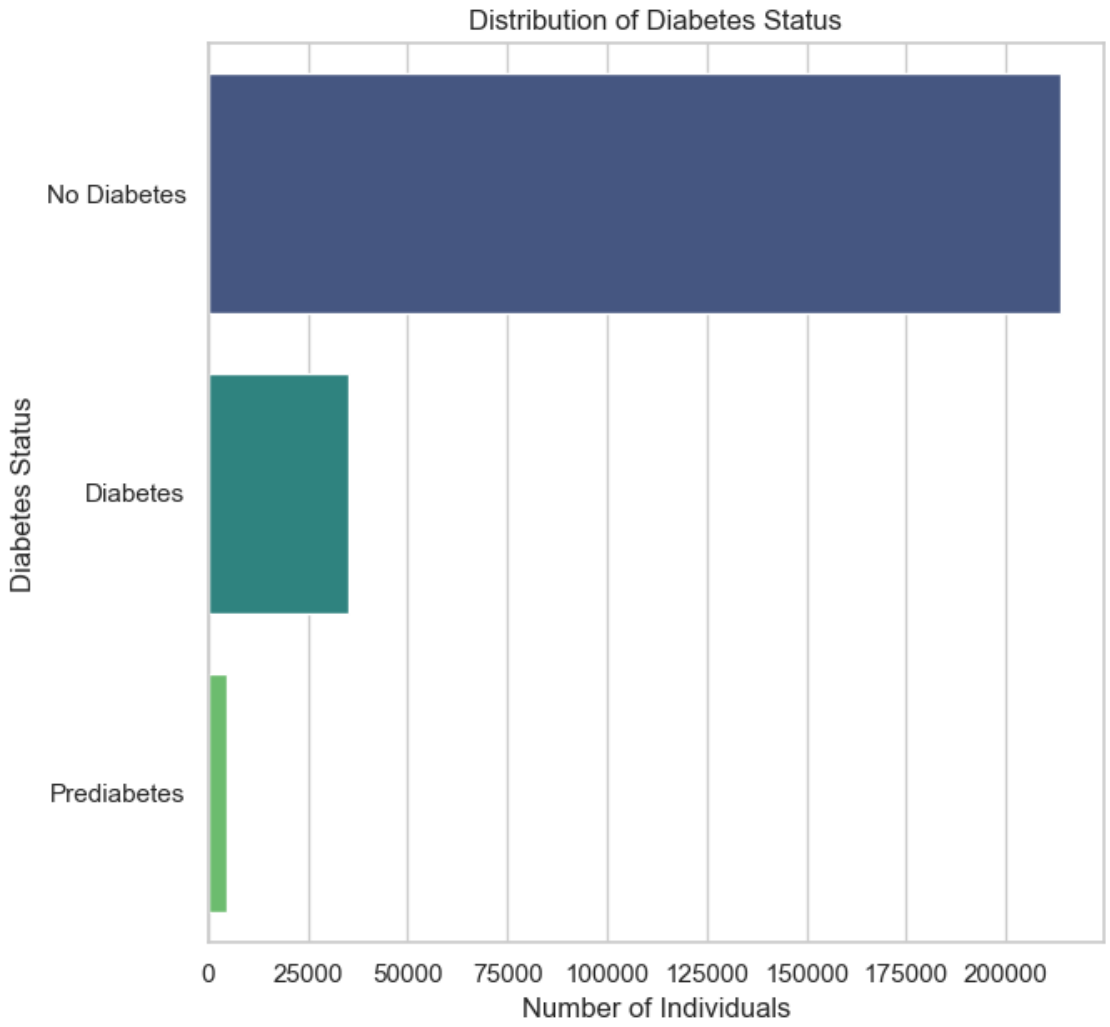
Diabetes is a chronic condition that affects millions worldwide. This project examines a set of health indicators to understand factors linked with diabetes presence and risk.

1.2 Objectives

This analysis seeks to:

1. Identify the health indicators most strongly associated with diabetes status
2. Examine how demographic factors relate to diabetes risk
3. Explore links between modifiable behaviors and diabetes
4. Highlight potential intervention points for prevention and early management

The dataset shows a clear imbalance in diabetes status. About 85% of individuals do not have diabetes, 13% have diabetes, and 2% have prediabetes. Although diabetes affects a smaller share of the population, the large sample size makes it a meaningful health burden. The contrast across groups provides a solid basis for comparing risk factors and demographic patterns linked with the condition.



BMI distributions differ clearly by diabetes status. Individuals with diabetes and prediabetes have higher median BMI values (about 30 and 29) than those without diabetes (about 25). The density curve for people without diabetes peaks near a BMI of 25, while those with diabetes show a broader spread with more values in the overweight (25–29.9) and obese (30+) ranges. This supports BMI as an important risk factor, with higher values strongly linked to diabetes.

Age is a major factor in diabetes prevalence. Rates begin to rise after about age 45 and increase further in the 65–69, 70–74, 75–79, and 80+ groups. While fewer than 10% of adults under 45 have diabetes, the rate more than doubles in older groups. This pattern supports age-based screening and early risk counseling starting in midlife.

People with diabetes have higher rates of other health conditions. High blood pressure (73%), high cholesterol (67%), and heart disease (22%) are more common among those with diabetes compared with those without (38%, 35%, and 7%). People with prediabetes also show higher rates than those without diabetes, suggesting a progression of related conditions. These findings show how metabolic and cardiovascular risks often cluster, pointing to the need for coordinated management.

Self-reported general health differs across diabetes groups. Individuals without diabetes most often report very good health (38%). Those with diabetes more often report good (27%) or fair (28%) health, with only a small share reporting excellent health (3%). Prediabetes falls between the two, consistent with a transitional risk state. These patterns suggest diabetes affects daily well-being and perceived quality of health.

Lifestyle behaviors vary meaningfully by diabetes status. People without diabetes report higher rates of physical activity (78%), fruit intake (63%), and vegetable intake (81%) compared with those with diabetes (63%, 60%, and 75%). Heavy alcohol use is low overall but slightly higher among those without diabetes. These findings suggest lifestyle changes may help both prevention and management, with physical activity as a key target.

The correlation matrix summarizes links across variables. Diabetes status shows its strongest positive correlations with general health rating (0.30), high blood pressure (0.27), BMI (0.22), and difficulty walking (0.22). Physical activity is negatively correlated (-0.12), suggesting a protective role. These results reinforce earlier findings and show that diabetes risk involves multiple interacting factors.

Diabetes prevalence is similar between females and males. Both groups show comparable rates (roughly 13–15% diabetes and 1–2% prediabetes). This suggests sex alone is not a major independent driver of risk, although it may still interact with other factors. Prevention and management strategies should address both groups equally while focusing on shared modifiable risks.

Education level shows an inverse relationship with diabetes. People with more years of schooling, especially college graduates, have lower diabetes prevalence (about 10% among those with 4+ college years) compared with those who never attended school (about 27%). Income shows a similar pattern: rates fall from 25% in the lowest bracket to about 8% in the highest. These trends highlight the influence of social and economic conditions and suggest that broader policies in education and economic stability may support diabetes prevention.

The combined risk score illustrates how multiple factors accumulate differently across diabetes groups. Individuals with diabetes have higher median scores and wider variation, indicating that several risks often occur together. The BMI and physical activity comparison further shows that regular activity is linked with lower BMI distributions regardless of diabetes status. This supports interventions that address clusters of risk rather than single factors in isolation.

The scatter plot of age, BMI, and diabetes status shows that higher BMI values are more often linked with diabetes across age groups. The spread suggests that risk thresholds may shift with age. Individuals who are both older and have elevated BMI appear especially vulnerable, indicating a priority group for screening and targeted counseling.

2 Conclusion

This exploratory analysis shows that diabetes relates to a set of interlinked factors spanning age, body weight, comorbid conditions, lifestyle behaviors, and social context. Clear differences in BMI patterns, age distribution, associated conditions, and socioeconomic gradients point to multiple opportunities for prevention and early intervention. The findings emphasize the value of promoting physical activity, healthy weight, and access to education and resources. These results can inform clinical risk assessment and public health strategies aimed at reducing the overall burden of diabetes.