**Data analysis report: National Health and Nutrition Health Survey csv**

**Prepared by: Warren Shitsama, Data analyst**

**Date: 13th July, 2025**

**Executive summary**

This report presents an in-depth analysis of data revolving around nutrition in relation to diabetes, with attributes such as age, gender, physical activity, body mass index, fasting blood glucose, oral tests, and insulin levels. The analysis tends to uncover if the independent factors such as age, gender, physical activity, body mass index, fasting blood glucose, oral tests, and insulin levels led to diabetes. The key finding here was that no individual factor can lead to diabetes, but a collection.

**Introduction**

Diabetes is a condition where the body fails to convert blood glucose to glycogen, which is kept for later use, thus failing to utilize the glucose. It can either be diabetic type 1 or type 2. And in this case, it is types 2, for most of the metric measured revolve around lifestyle choices. The objective of the analysis was to reveal a connection between the metrics and the probability of causing diabetes.

**Dataset Overview**

The dataset has the following columns, which are 11 in total:

1. Sequence number
2. Age group
3. Age
4. Gender
5. Physical activity
6. Body mass index
7. Fasting blood glucose
8. Diabetic status
9. Diabetic score
10. Oral score
11. Insulin levels

The dataset has been cleaned leading to deletion of incomplete rows or duplicates. The result of the cleaning is 2219 entries.

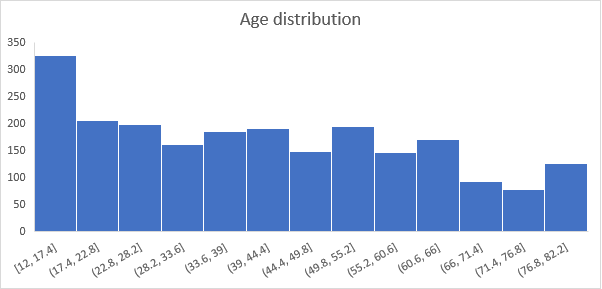
**Methodology**

The dataset was analyzed using techniques such as descriptive analysis, segmentation and benchmarking, and correlational studies. MySQL was used solely in this analysis. The main question while analyzing was, Is there a relationship between the independent factors and the dependent factor (diabetes status)?

**Analysis and Findings**

1. **Descriptive studies**

The study was mostly made up of young people and most of them were female.



Most of them don’t take part in physical activities, this cuts across both the diabetic positive and negative.

1. **Correlation and segmentation analysis**

The findings between the diabetic positive and that negative is the same. There is no correlation between the independent and dependent factors.

However, with a further segmentation of the diabetic positive, it was found that a combination of factors leads to diabetes, not just a single factor.

* Body mass index and diabetes = -0.04 (no correlation)
* Insulin and diabetes = -0.03 (no correlation)
* Fasting blood glucose and diabetes = -0.2 (very weak correlation)

**Recommendation**

1. Given that fact that no single factor leads to diabetes, means strategies covering all the factors is preferrable.
2. Women take up a huge part of those with diabetes, so custom monitoring should be diverted to them.
3. Most of those with diabetes are those between the age of 17 to 50, given that these individuals are the most inclined to adapt to unhealth lifestyles, it is important to focus on them.

**Conclusions**

No single factor leads to diabetes types 2, but a collection.