# From Data to Diagnosis: Predicting Diabetes

SDG 3 : Good health and well being.
Member :
- Cheryl Aurellia Valencia || 2702252872
- Cynthia Anne Ramali || 2702254032
- Eirene Michella Tjhan || 2702256630
- Tiffany Felita || 2702265641

# Introduction

Diabetes is a disease caused by high blood sugar levels in the body, where the body does not utilize or process this blood sugar efficiently, resulting in an excess accumulation of blood sugar in the body. This condition eventually leads to diabetes. Diabetes consists of 2 types, there are type I and type II. The difference between them is that type I involves damage to the beta cells in the pancreas, which leads to the inability to produce insulin. This type of diabetes usually affects children. Meanwhile, type II involves the body's insensitivity to insulin processes that can cause obesity, typically affecting adults.

This dataset is originally taken from the National Institute of Diabetes and Digestive and Kidney Diseases with the aim of analyzing and predicting whether someone has diabetes based on the measurements that also available in the dataset. The dataset includes samples from 768 women of Pima Indian from Arizona, United States, who are above 21 years old. Pima Indians themselves are one of the largest populations affected by Type II diabetes in the world.

The dataset uses the Oral Glucose Tolerance Test (OGTT) to observe the body's reaction in processing glucose, which can help to identify insulin resistance and diabetes. Insulin resistance is a condition where the body's cells become less responsive to insulin, leading to an increase in blood sugar levels in the body. Insulin plays a crucial role in regulating blood sugar levels, making it easier for cells to absorb glucose and convert it into energy that the body can use.

Prediction based on the analysis we conducted is crucial because diabetes can lead to many damages in the body, such as hypertension, which has the potential to damage blood circulation throughout the body. When we predict and identify people at risk of diabetes, healthcare professionals can step in with early preventive actions to reduce the chances of issues like high blood pressure and heart problems that come with diabetes.This analysis can significantly improve the overall health outcomes and quality of life for individuals affected by diabetes.

## The Data Says

According to the dataset, we found about 34.9% of women over the age of 21, who took the OGTT test in Arizona, USA are indicated to have a diabetes. This is based on our calculations using the 'Outcome' variable in our dataset.

# Data Description

| Variabel | Explanation |
|---|---|
| Pregnanices | To express the Number of pregnancies. |
| Glucose | Plasma glucose concentration a 2 hours in an oral glucose tolerance test (OGTT) (mg/dL). Distributions of 2-hours blood glucose level in OGTT:<br>- Normal: <140 mg/dL (7.8 mmol/L)<br>- Prediabetes: 140 - 199 mg/dL (7.8 to 11.0 mmol/L)<br>- Diabetes: >200 mg/dL (11.1 mmol/L) |
| BloodPressure | To express the Blood pressure measurement, Diastolic blood pressure (mm Hg). Distributions of diastolic blood pressure:<br>- Normal: Less than 80 mm Hg<br>- Elevated: Less than 80 mm Hg<br>- Hypertension Stage 1: 80-89 mm Hg<br>- Hypertension Stage 2: 90 mm Hg or higher<br>- Hypertensive Crisis: Higher than 120 mm Hg (requires immediate medical attention) |
| SkinThickness | To express the thickness of the skin, Triceps skin fold thickness (mm).<br>- Low: < 10mm<br>- Average: 10-25mm<br>- High: > 25mm |
| Insulin | To express the Insulin level in blood, 2-Hour serum insulin (mu U/ml). |
| BMI | Body Mass Index (weight in kg/(height in m)^2). Adapun distribusi BMI:<br>- Underweight: <18.5 kg/m²<br>- Normal: between 18.5 and 24.9 kg/m²<br>- Overweight: between 25.0 and 29.9 kg/m²<br>- Obesity: >30.0 kg/m² |
| DiabetesPedigreeFunction | To express percentage of risk of developing diabetes based on family history. |
| Age | To express the age. |
| Outcome | To express the final result 1 (diabetical) and 0 (non-diabetical). |

# Result of OGTT
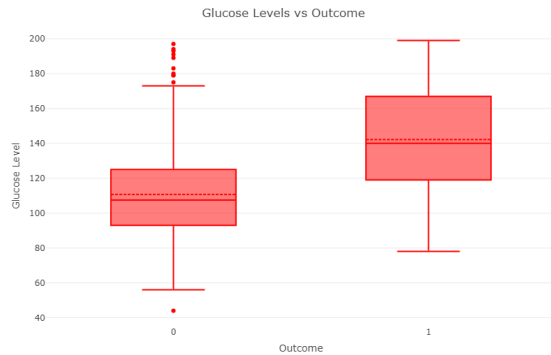


The OGTT results are shown in the graphs. In the first graph, people with a diabetes history keep their glucose levels high even after 2 hours of the oral glucose test, showing there are insulin incapability of regulate glucose concentration in blood, this condition is also known as Insulin Resistance.
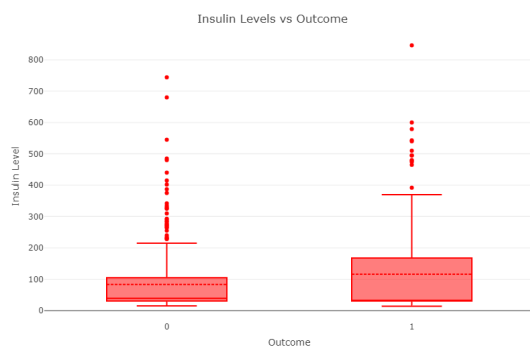


For people with diabetes, the insulin levels may remain elevated or show a delayed response, indicating the failed body's attempt to manage the high blood glucose levels. That makes insulin can't really get the job done breaking down glucose.

# Why DIABETES is a Problem
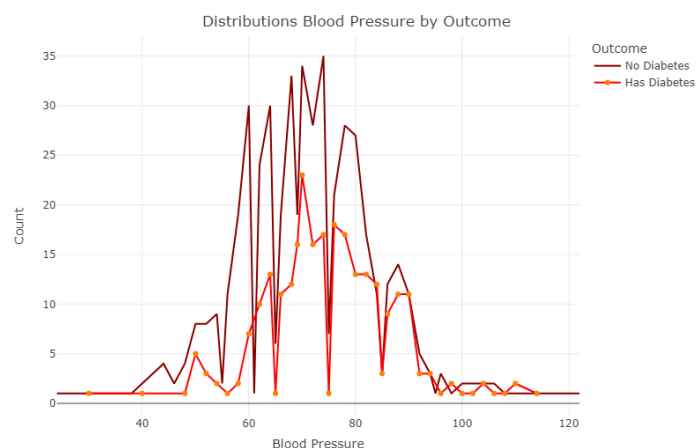


The graph illustrates how the distribution of individuals with and without diabetes correlates with their blood pressure. The brown line represents non-diabetic individuals, while the red line represents individuals with diabetes. The axes on the graph indicate the variables used, where

- The x-axis represents blood pressure values
- The y-axis represents the number of individuals.

From the graph, we can observe that both groups have a similar range of blood pressure values, that's from 40 to 120. Therefore, it shows that the number of non-diabetic individuals with blood pressure between 90 and 100 is higher compared to those with diabetes. This indicates that blood pressure in individuals with diabetes tends to be lower than those without diabetes.

The variation in data and the possible increase of blood pressure in individuals with diabetes indicate that it is crucial for people with diabetes to continue monitoring their blood pressure to prevent further complications.

# Factor That Influence

There are several key factors that can make someone more vulnerable to diabetes.

## Lifestyle
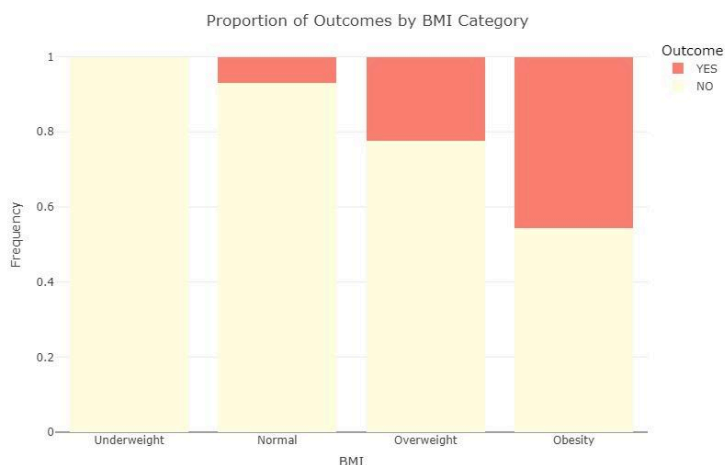


The influential factor is how a person manages their daily lifestyle. Lifestyle, in this aspect, has an impact on the body's limited use of blood sugar due to lack of physical activity. This leads to the body's inability to process glucose and allows stored energy reserves to accumulate. Therefore, unhealthy eating habits such as frequent consumption of high-sugar foods and beverages, as well as less intake of fiber and vitamins that are needed for body production.

## Heredity

```
        Pearson's product-moment correlation

data:  df$DiabetesPedigreeFunction and df$Outcome
t = 4.8858, df = 766, p-value = 1.255e-06
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1043836 0.2416168
sample estimates:
      cor
0.1738441
```

The last factor that can cause diabetes is heredity or genetics.

- **Correlation Coefficient (cor):** The test indicates a statistically significant positive correlation between DiabetesPedigreeFunction and Outcome, though the strength of the correlation is relatively weak (0.174).
- **Significance (p-value):** The p-value of 1.255e-06 is very small (less than 0.05), suggesting strong evidence against the null hypothesis. Therefore, we reject the null hypothesis that there is no correlation between DiabetesPedigreeFunction and Outcome.
- **Confidence Interval:** The 95% confidence interval for the correlation coefficient is [0.1043836, 0.2416168]. This interval indicates that we are 95% confident that the true population correlation coefficient falls between 0.1043836 and 0.2416168.

**Conclusion:**
- There is a statistically significant, albeit weak, positive correlation (r = 0.1738441) between DiabetesPedigreeFunction and Outcome.
- This suggests that as DiabetesPedigreeFunction increases, there tends to be a slight increase in the likelihood of Outcome (which could be diabetes in this context).
- However, the correlation is not strong, as the correlation coefficient is relatively close to zero.

In summary, while there is a statistically significant positive correlation between DiabetesPedigreeFunction and Outcome, DiabetesPedigreeFunction alone may not be a strong predictor of Outcome (diabetes). Other factors likely contribute more significantly to determining the outcome.

# How to Prevent

**Maintain healthy diet**
Maintaining a healthy diet and staying in a healthy weight range are good. Eating a balanced diet means including a variety of foods, like fruits, vegetables, whole grains, lean proteins, and healthy fats, while limiting sugary snacks, processed foods, and high-fat items. This helps body get the nutrients it needs without too many extra calories. Keeping your weight in a healthy range is also important. This can be checked using Body Mass Index (BMI), which

is a measure that compares weight to height. A healthy BMI typically ranges from 18.5 to 24.9. Staying within this range lowers the risk of health problems like diabetes. By eating well and maintaining a healthy weight, you give your body the best chance to stay strong and energetic.

**Excercise regularly**

Exercising regularly and getting some moderate physical activity is key to keeping body fat in check. Whether it's a walk, a bike ride, or a fun workout class, moving your body helps burn calories and keeps you fit. Plus, it makes you feel great and boosts your energy. So, make it a habit to stay active and keep that body fat under control!

**Eating less sugar**

Eating less sugar helps keep your blood sugar from suddenly or sharply increasing too much. Skip the sugary stuff, especially all those processed foods. When you cut back on sugar, your body stays more even-stable, which keeps you feeling full of energy throughout the day. So, say no for those sugary snacks, and you'll feel the difference!

# Conclusion

We conclude that someone can get diabetes because of their daily lifestyle habits. That's why it's really important to maintain a healthy lifestyle every day. However, there's also a chance that we can get diabetes due to family history. So, it's very important to check if we carry diabetes or not and let's also always maintain your glucose blood in normal level for avoiding any further complication