**Outcome of your EDA**

I initiated this analysis to address two key questions:

1. **What role does BMI and Waist Circumference play in relation to Type 2 Diabetes?**
2. **Does having a family history increase the risk of developing diabetes?**

After thorough exploration, the answers to both questions turned out to be **YES**. The analysis revealed that most participants with high BMI also exhibited higher Waist Circumference, suggesting a strong correlation between these factors and the risk of developing Type 2 Diabetes. However, when comparing the impact of family history, its role was not as significant as I initially expected. While it does contribute to the risk, BMI and Waist Circumference appear to have a more direct and pronounced relationship with Type 2 Diabetes.

**What do you feel was missed during the analysis?**

The dataset contains numerous medical and test-related terms, many of which I struggled to fully understand. This limited my ability to interpret the data comprehensively and may have caused me to overlook key insights. I feel that with a clearer understanding of these terms, I could have identified and included additional columns for deeper research. A better grasp of the medical context would have enriched the analysis and improved the overall findings.

**Were there any variables you felt could have helped in the analysis?**

The **Dietary Habits** column currently categorizes individuals as "Healthy" or "Unhealthy." If it were to include actual data, this column could instead display the average daily calorie consumption in numerical values, such as "2000 kcal" or "3000 kcal." Similarly, the **Physical Activity** column presently classifies activity levels into "High," "Moderate," and "Low." If actual metrics were used, it could record the number of active minutes per day, such as "120 minutes" or "30 minutes." This change would provide a more precise and quantitative understanding of dietary patterns and physical activity levels, allowing for detailed data analysis, trends, and correlations. It would also enhance personalization for health recommendations and studies.

**Were there any assumptions made you felt were incorrect?**

1. When I began the analysis, I assumed the dataset represented all types of people. However, as I delved deeper, I discovered a column titled **"Weight Gain During Pregnancy,"** which indicated that the data specifically pertains to women. This realization shifted my perspective on the scope of the analysis. It became clear that the dataset focuses on a particular demographic, likely pregnant or postpartum women.
2. At the start of the analysis, I assumed that family history would have the most significant impact compared to other variables. However, as the analysis progressed, I found that while it does play a role, its influence was not as substantial as I had initially anticipated.

**What challenges did you face, what did you not fully understand?**

1. I noticed that the values in the **“Birth Weight”** column are four digits, which makes them confuse to interpret. They don’t seem to be in standard units like pounds or kilograms, leaving their actual meaning unclear.
2. The dataset lacks critical information, such as the date when the data was collected and the ages of the participants. Having this information would have been invaluable for analyzing trends over time and understanding age-related factors. With these details, I could have better predicted the relative contributions of family history versus BMI to the risk of developing diabetes. This would have allowed for more precise insights into how these variables interact across different age groups.

**Submit a link to your repository to the assignment link during the final week of class.**

<https://github.com/vimal37/DSC530>