**Diabetes Prediction Analysis**

**Introduction**

The **Diabetes prediction dataset** comprises medical and demographic information of patients along with their diabetes status (positive or negative). It includes variables such as; age, gender, body mass index (BMI), hypertension, heart disease, smoking history, HbA1c level, and blood glucose level. The dataset can be beneficial for healthcare providers to identify individuals who might be susceptible to diabetes and devise customized treatment approaches.

My analysis of the dataset was centered around the impact of age groups, gender, BMI, smoking history, HbA1c level, and heart disease on the probability of developing diabetes.

**Data**

The complete dataset can be accessed on the following webpage: <https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset>.

**Data Processing**

The analyzed data were cleaned using the following procedures:

* The original CSV data was converted to Excel format.
* Duplicate data was removed.
* Age, BMI, and HbA1c level were categorized into groups to facilitate analysis and improve visualization.
* Pivot tables and charts were utilized to analyze and visualize the data.

In my data analysis, I adopted the framework of "Question-Visualization-Observations" to examine specific attributes of the data.

1. What is the distribution of diabetes across different age groups and genders?

**Observations:** The likelihood of developing diabetes generally increases as age increases. The infants’ age group has the lowest likelihood of developing diabetes. Diabetes is more prevalent in older age groups for both males and females. Females have a higher prevalence of diabetes compared to males.

1. How does the BMI of an individual correlate with their likelihood of developing diabetes?

**Observations:** As BMI increases, there is a corresponding increase in the likelihood of developing diabetes. There is a noticeable difference in the likelihood of developing diabetes between individuals that are Obese and those that are overweight. Women generally have a higher risk of developing diabetes than men at any given BMI level.

1. What is the association between smoking history and the development of diabetes?

**Observations:** Non-smokers have a higher likelihood of developing diabetes compared to smokers. The association between smoking history and diabetes is more pronounced in females compared to males.

1. How does the HbA1c level impact the likelihood of developing diabetes?

**Observations:** The prevalence of HbA1c level is higher in the adults, aged, and middle-aged groups, with the adult group having the highest proportion of individuals with prediabetes, followed by the aged group. It's interesting to note that adolescents and the pediatrics group had normal levels of HbA1c, but the prevalence was very minimal.

1. What is the Impact of Age and BMI on Heart Disease Prevalence?

**Observations:** Older age groups tend to have higher rates of heart disease, regardless of BMI.

Among individuals with overweight or obese BMI levels, the prevalence of heart disease is highest in the adult group, followed by the aged and middle-aged groups. The prevalence of heart disease is highest among individuals with overweight BMI levels. The prevalence of heart disease tends to be lower among individuals with underweight BMI levels.

**Recommendations**

* Encourage individuals to maintain a healthy weight through a balanced diet and regular exercise as it is an important factor in reducing the likelihood of developing diabetes.
* Increase diabetes screening and prevention efforts among older individuals, especially women, who have a higher risk of developing diabetes.
* Develop tailored smoking cessation programs for individuals at risk of developing diabetes, with a particular focus on females.
* Encourage routine monitoring of HbA1c levels among adults and the aged group, as they have a higher prevalence of prediabetes.
* Implement regular heart disease screening programs among older individuals, regardless of their BMI levels.
* Increase awareness about the higher prevalence of heart disease among individuals with overweight BMI levels, and encourage weight management strategies to reduce the risk of heart disease.
* Promote healthy lifestyle choices and regular health check-ups, especially for older individuals, to reduce the risk of developing diabetes and heart disease.

**Conclusion**

The analysis of the **Diabetes prediction dataset** revealed that age, gender, BMI, smoking history, HbA1c level, and heart disease are all significant factors that influence the likelihood of developing diabetes. Age and BMI were found to have a strong association with diabetes, while gender and smoking history also played a role. The prevalence of heart disease was also found to be higher in older age groups, particularly among individuals with overweight or obese BMI levels. The findings of this analysis can be used to identify patients who may be at risk of developing diabetes and to develop personalized treatment plans**.**