**MSBA 230 - Final Project**

**Diabetes**

12/01/2023

Akshith Mashetty

Avinash Rajendran

Coby McCaig

Deepika Vemu

Harish Gurrapu

Pragati Khanna

1. **Background**

**Who the company is that you are working for**

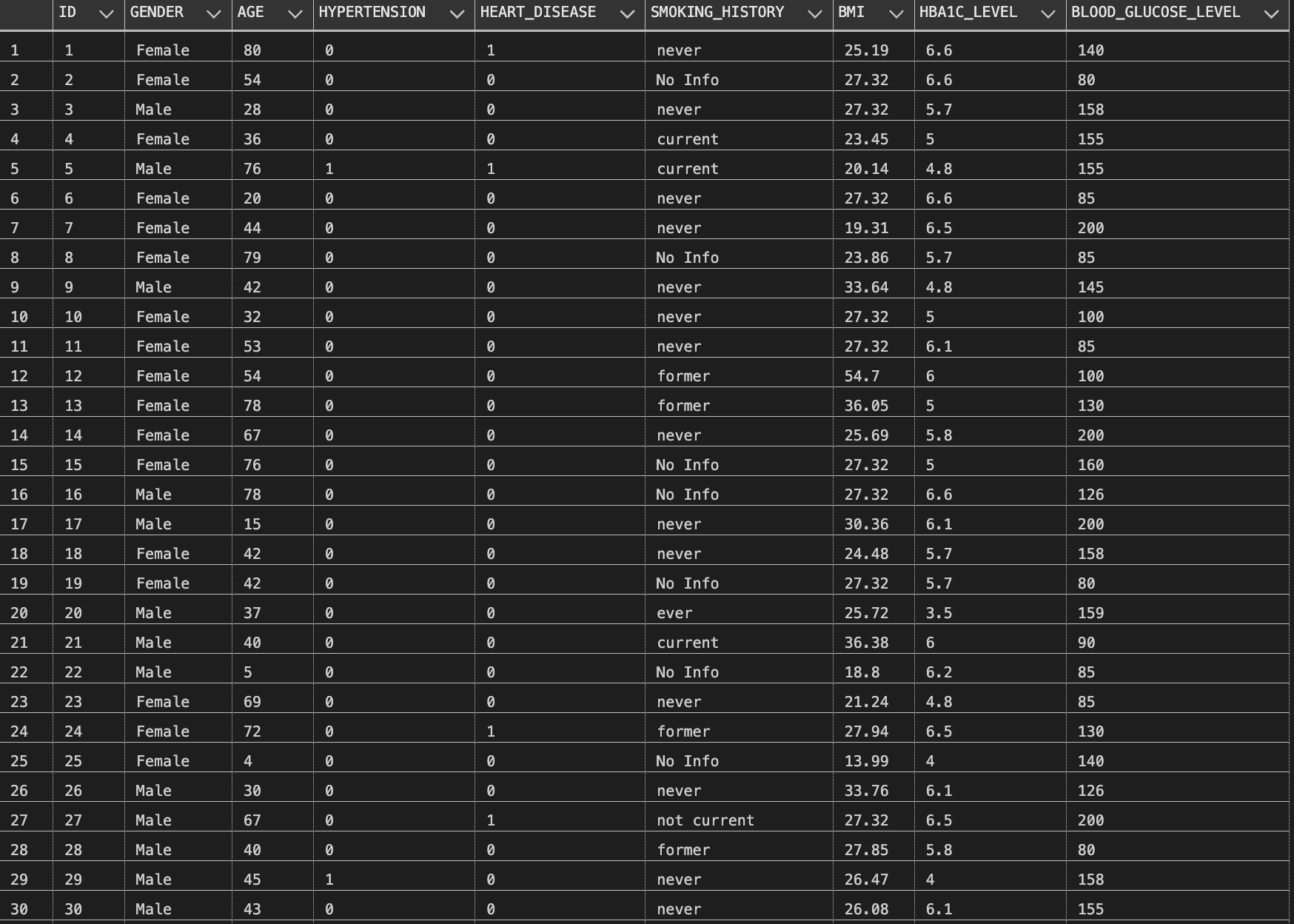
Our team has signed a contract with a diabetes endocrinology center, giving us data to find meaningful insights. The company wants to know how they can put their database to good use and extract any value if any with the help of our database management team. Our database management team can make use of the diabetes database that we are working with and ultimately create business insights that can benefit the company as a whole. The contract is only temporary, so we have used only one dataset.

**The contract is a one-month contract and you must convince the director that they should extend the contract indefinitely.**

* The Diabetes prediction dataset collects medical and demographic data from patients, along with their diabetes status (positive or negative). The data includes age, gender, body mass index (BMI), hypertension, heart disease, smoking history, HbA1c level, and blood glucose level. This dataset can be used to build machine learning models to predict diabetes in patients based on their medical history and demographic information. This can be useful for healthcare professionals in identifying patients who may be at risk of developing diabetes and in developing personalized treatment plans. Additionally, the dataset can be used by researchers to explore the relationships between various medical and demographic factors and the likelihood of developing diabetes.
* The **diabetes\_prediction\_dataset.csv file** contains medical and demographic data of patients along with their diabetes status, whether positive or negative. It consists of various features such as age, gender, body mass index (BMI), hypertension, heart disease, smoking history, HbA1c level, and blood glucose level. The Dataset can be utilized to construct machine learning models that can predict the likelihood of diabetes in patients based on their medical history and demographic details.

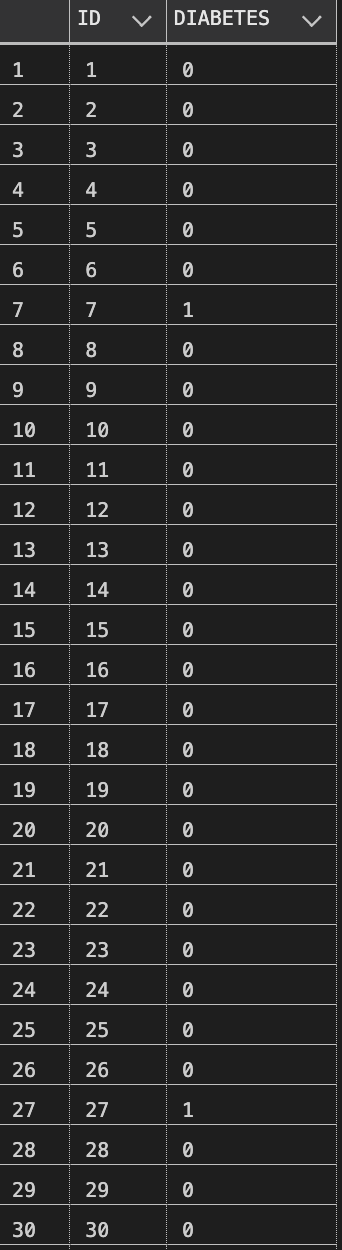
**II. Tables**

**Table 1:** Patient\_Info

****

**Description:**

This table contains essential health-related information about patients enrolled in the study. Each row represents an individual patient, and the columns provide details such as gender, age, hypertension and heart disease status, smoking history, BMI, HbA1c level, and blood glucose level. The diverse data set captures a wide range of patient characteristics, offering a comprehensive view for analysis and insights into potential factors associated with diabetes.

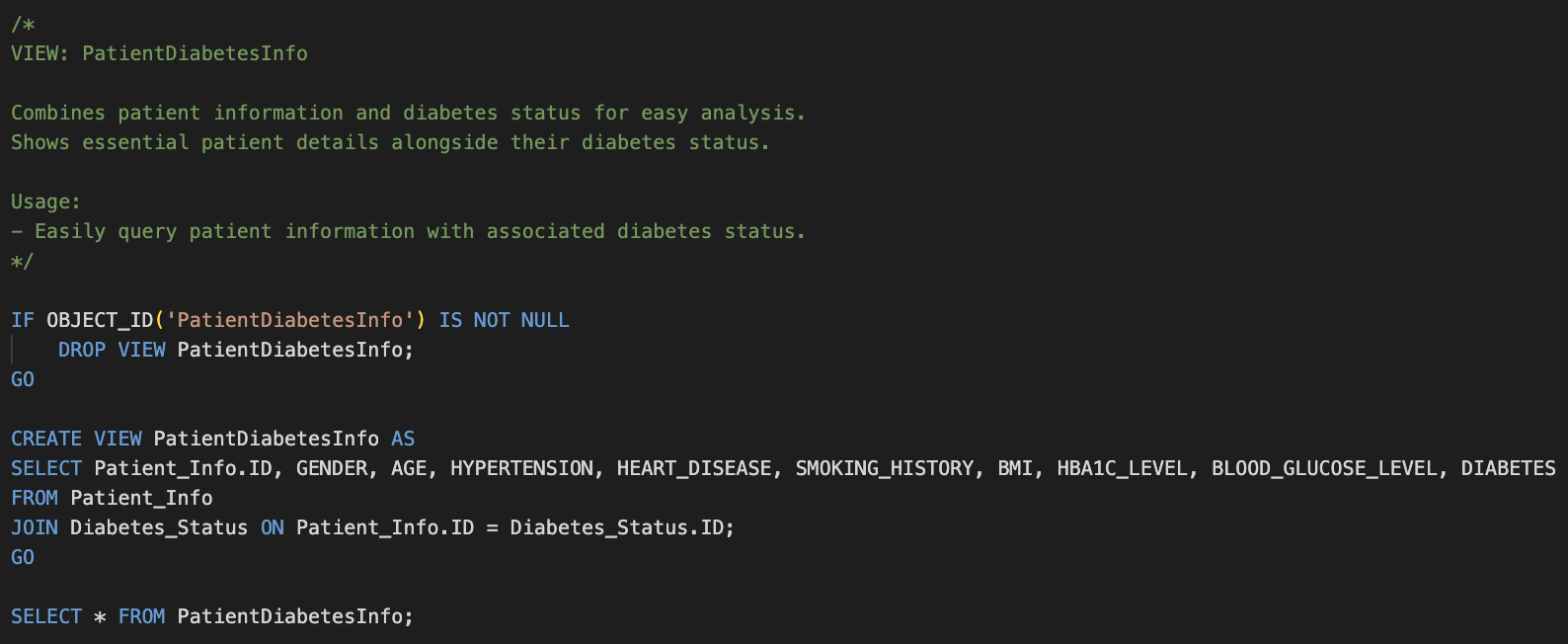


**Table 2:** Diabetes\_Status

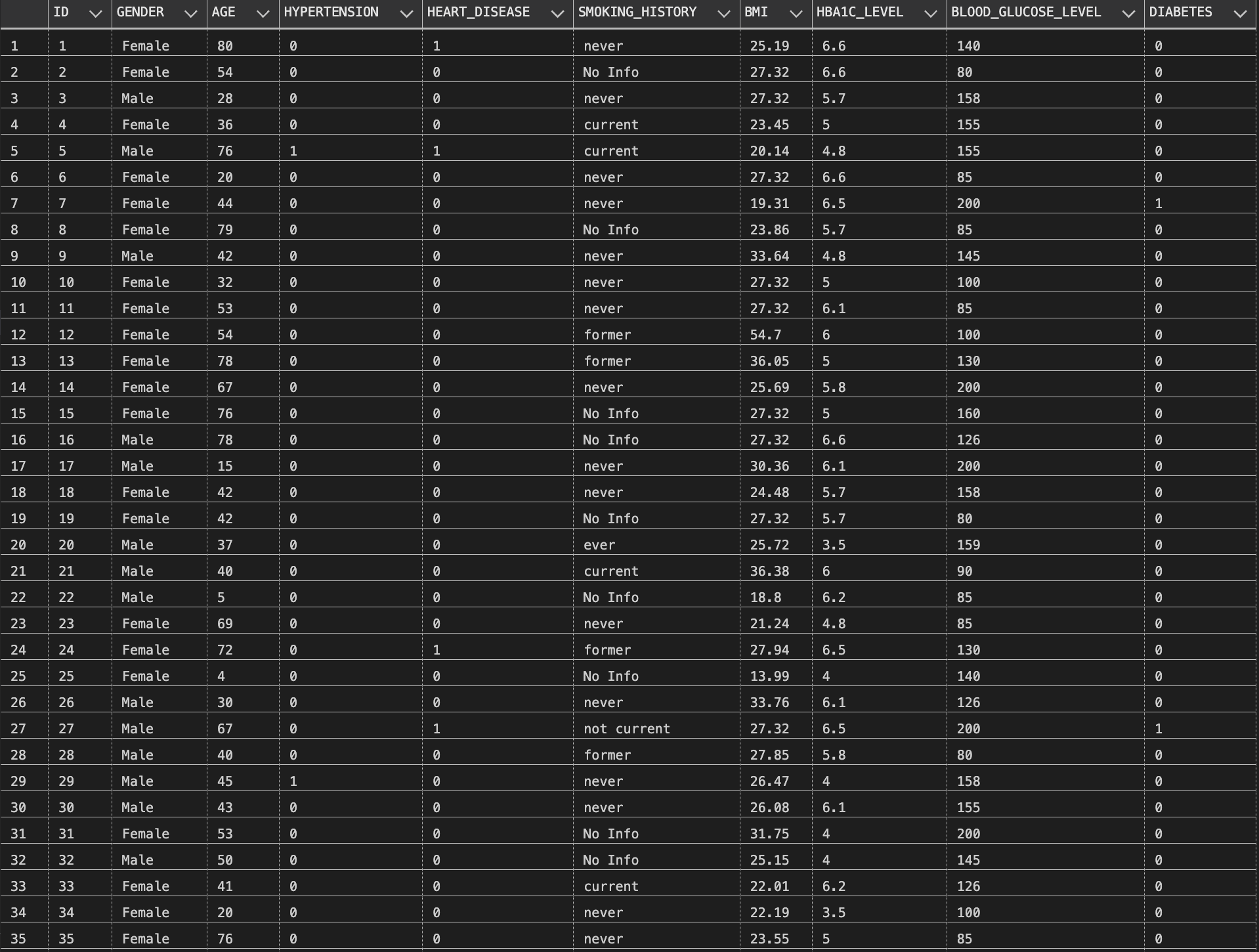
**Description:**

This table provides information on the diabetes status of each patient, indicating whether they have diabetes (1) or not (0). The data in this table is linked to the ID column in the Patient\_Info table, allowing for seamless integration and analysis of diabetes prevalence among the study participants.

**III. Views**

****

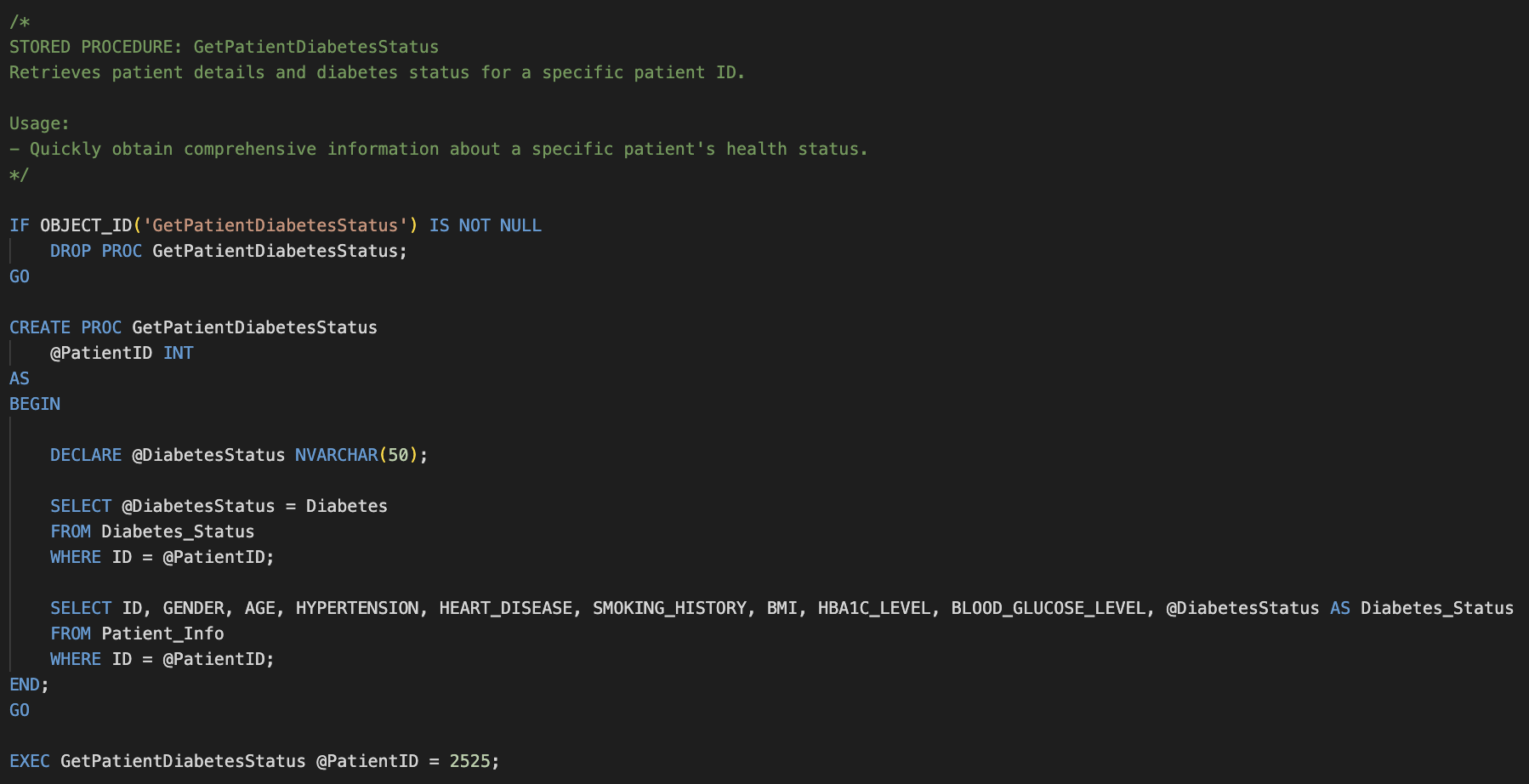
**RESULT:**

****

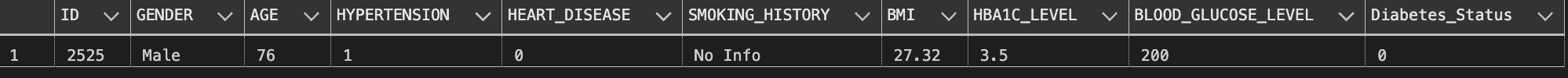
**Description:**

This script creates a view named **PatientDiabetesInfo** that combines patient information and diabetes status for convenient analysis. The view displays essential patient details along with their diabetes status, facilitating easy querying of patient information with associated diabetes status. The usage includes running a SELECT statement to view the contents of the newly created view.

**IV. Stored Procedure**

****

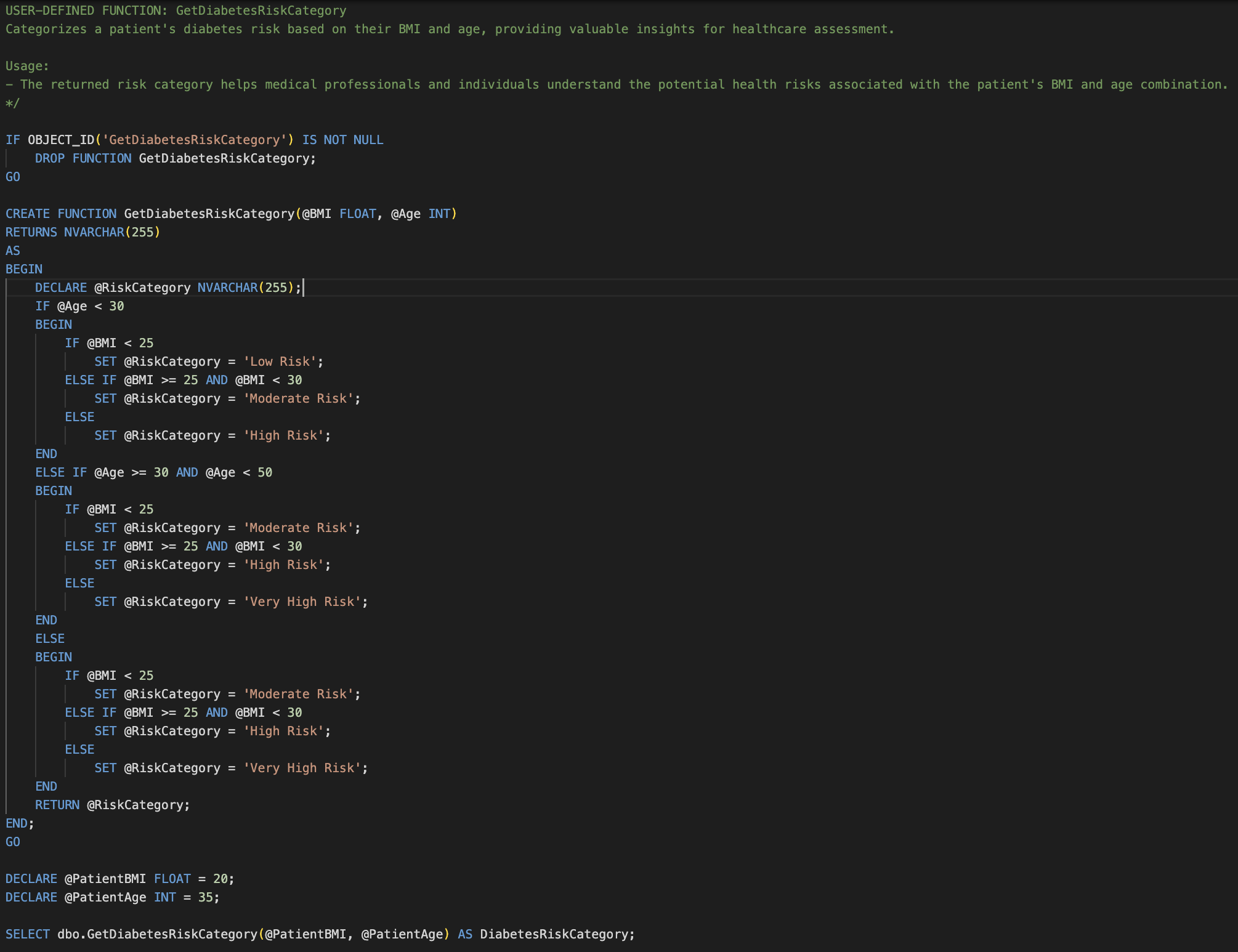
**RESULT:**

****

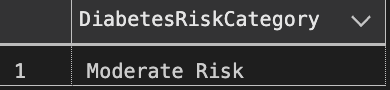
**Description:**

This script creates a stored procedure named GetPatientDiabetesStatus that retrieves patient details and diabetes status for a specific patient ID. The procedure is designed to quickly obtain comprehensive information about a specific patient's health status. The usage includes executing the stored procedure with a specific patient ID (2525 in this case).

**V. User-Defined Function**

****

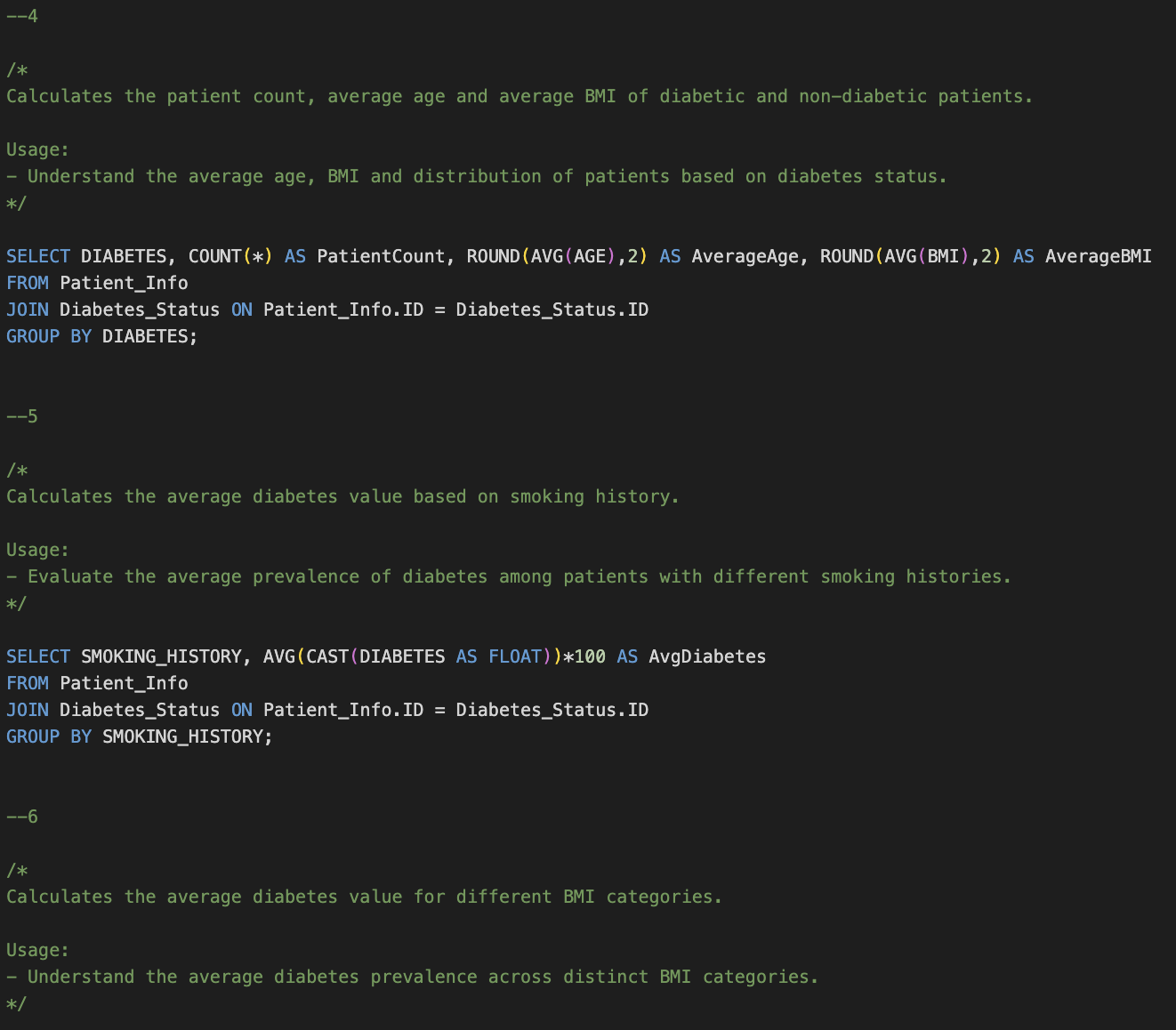
**RESULT:**

****

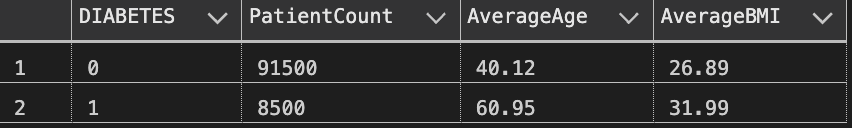
**Description:**

This script creates a user-defined function named GetDiabetesRiskCategory that categorizes a patient's diabetes risk based on their BMI and age. The function returns a risk category, providing valuable insights for healthcare assessment. The usage includes declaring BMI and age variables and then calling the function to obtain the diabetes risk category.

**VI. Query Summaries**

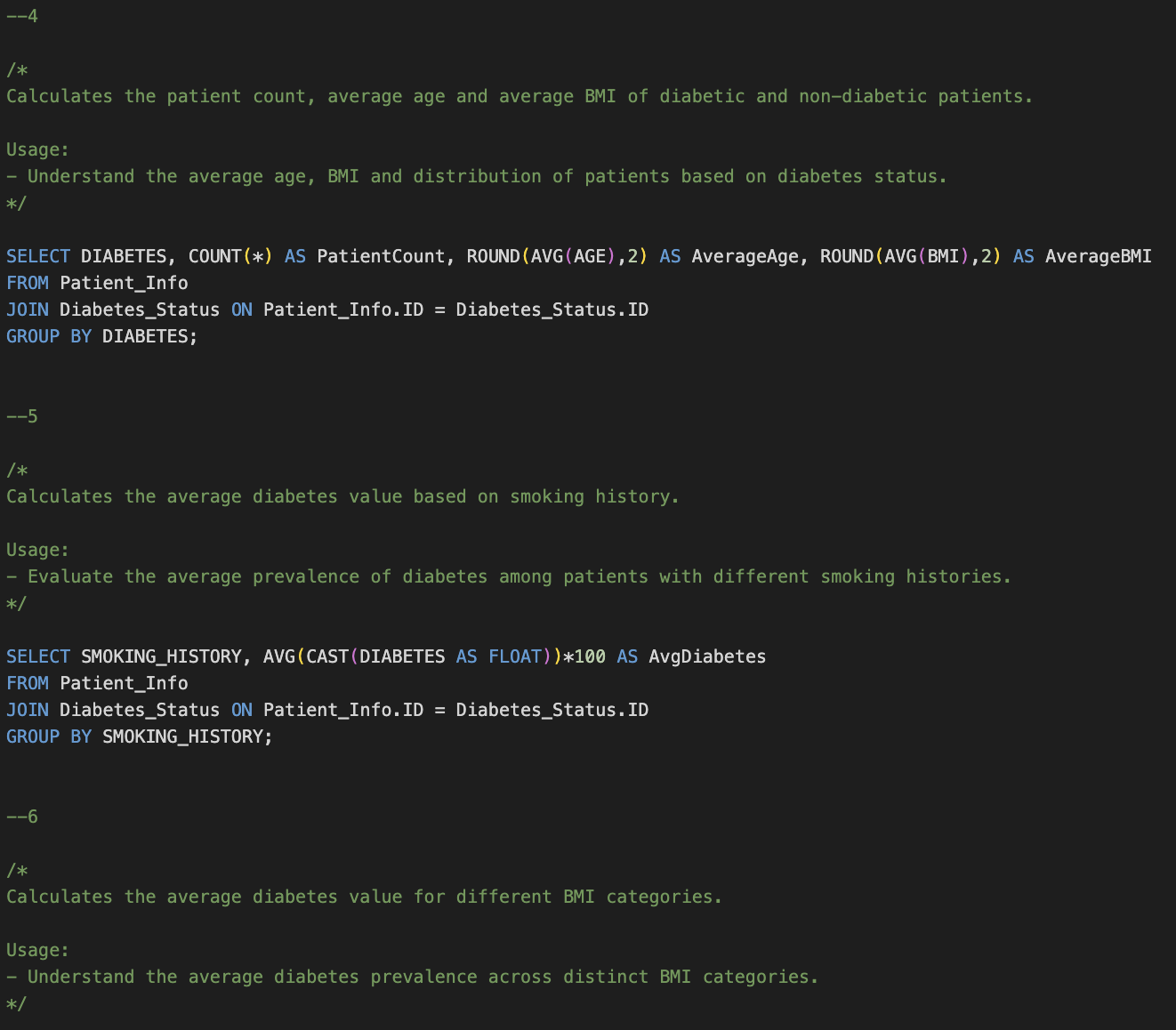
****

**RESULT :**

****

**Description:**

This script calculates the patient count, average age, and average BMI of diabetic and non-diabetic patients. The script provides insights into the distribution of patients based on their diabetes status, allowing users to understand the average age, BMI, and overall distribution of patients with and without diabetes.

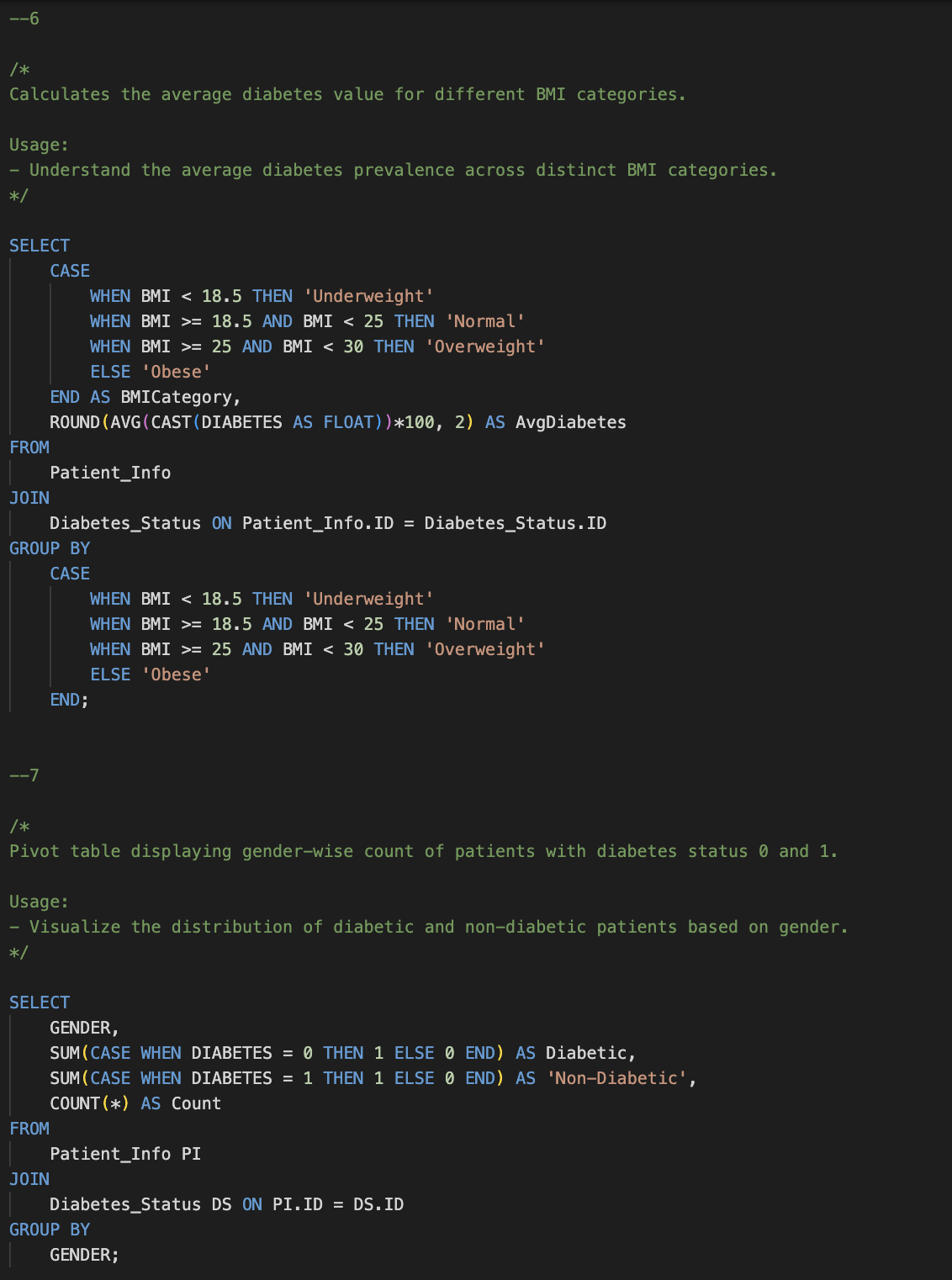
****

**RESULT :**

****

**Description:**

This script calculates the average diabetes value based on smoking history. By evaluating the average prevalence of diabetes among patients with different smoking histories, the script provides insights into the relationship between smoking history and diabetes.

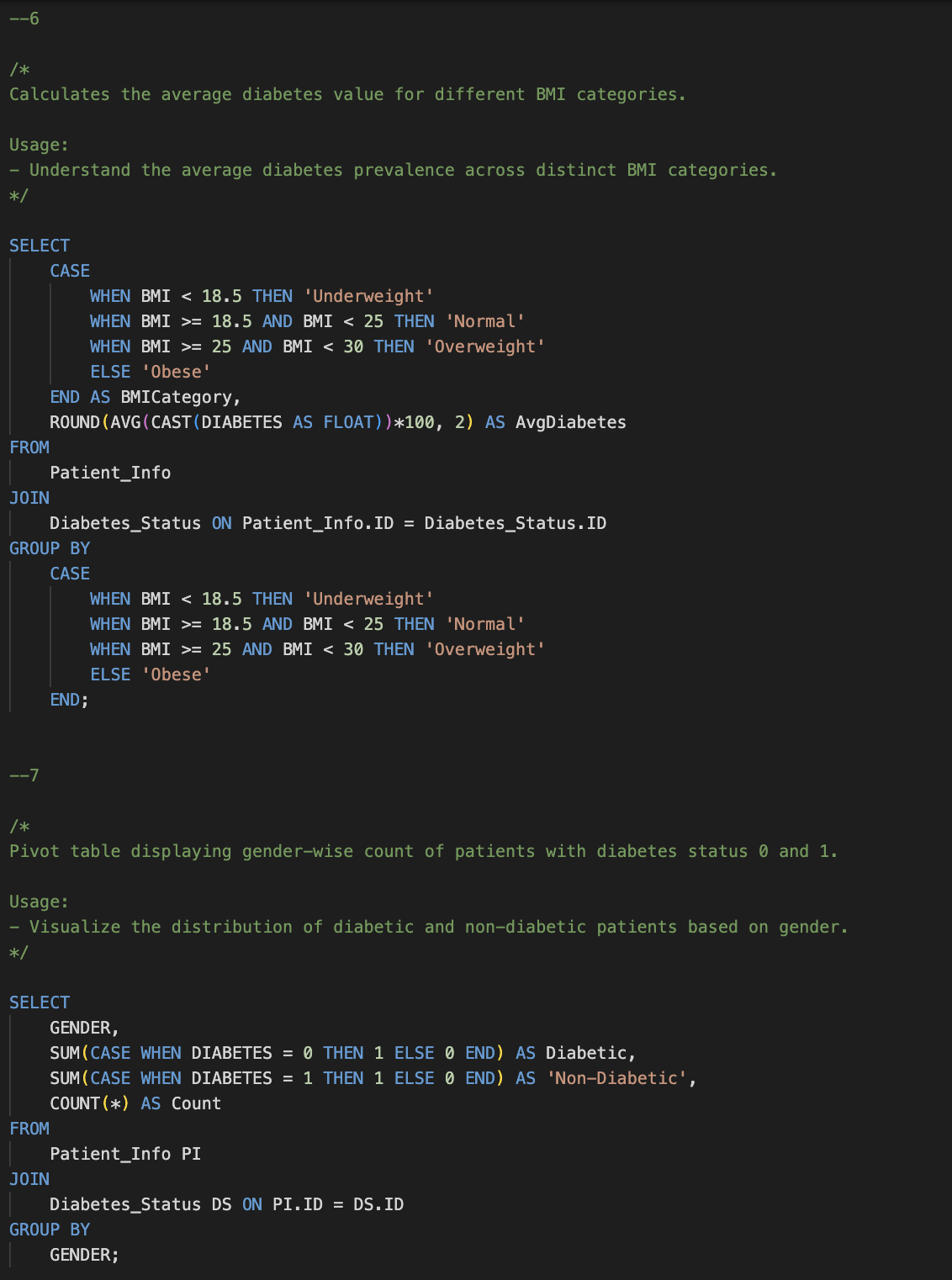
****

**RESULT :**

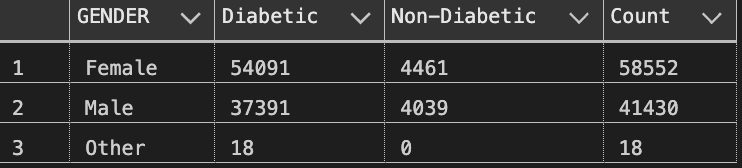
****

**Description:**

This script calculates the average diabetes value for different BMI categories. By categorizing patients into BMI groups and determining the average diabetes prevalence in each category, the script aids in understanding the relationship between BMI and diabetes.

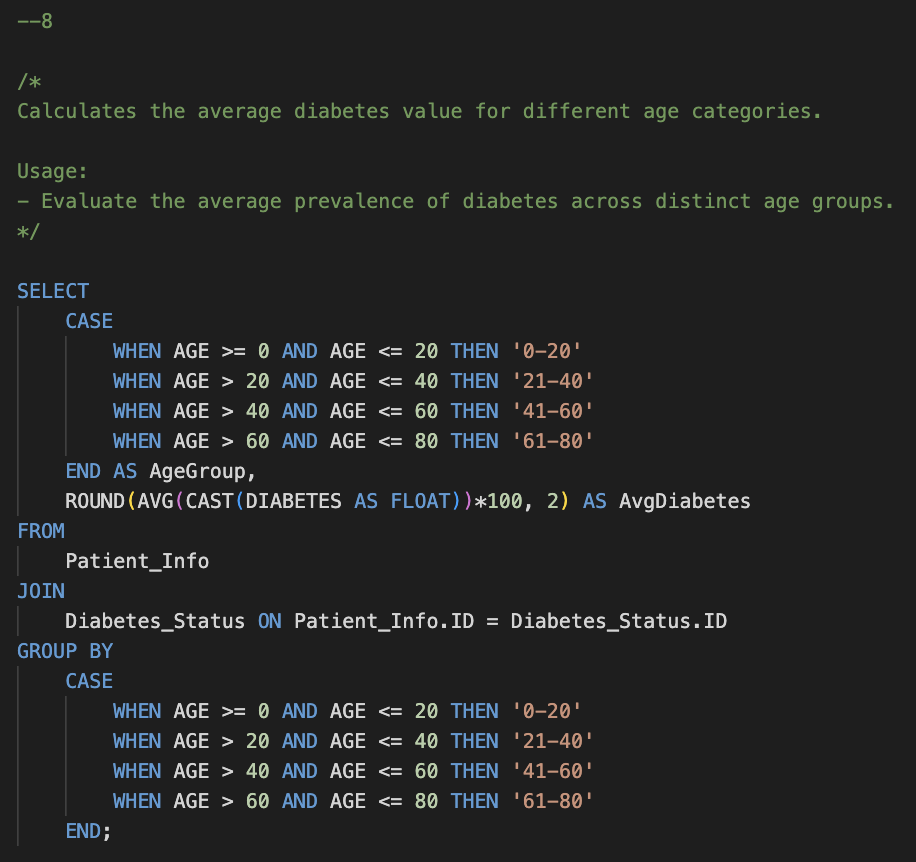
****

**RESULT :**

****

**Description:**

This script creates a pivot table displaying the gender-wise count of patients with diabetes status 0 and 1. It visualizes the distribution of diabetic and non-diabetic patients based on gender, providing a clear representation of how diabetes is distributed among different genders.

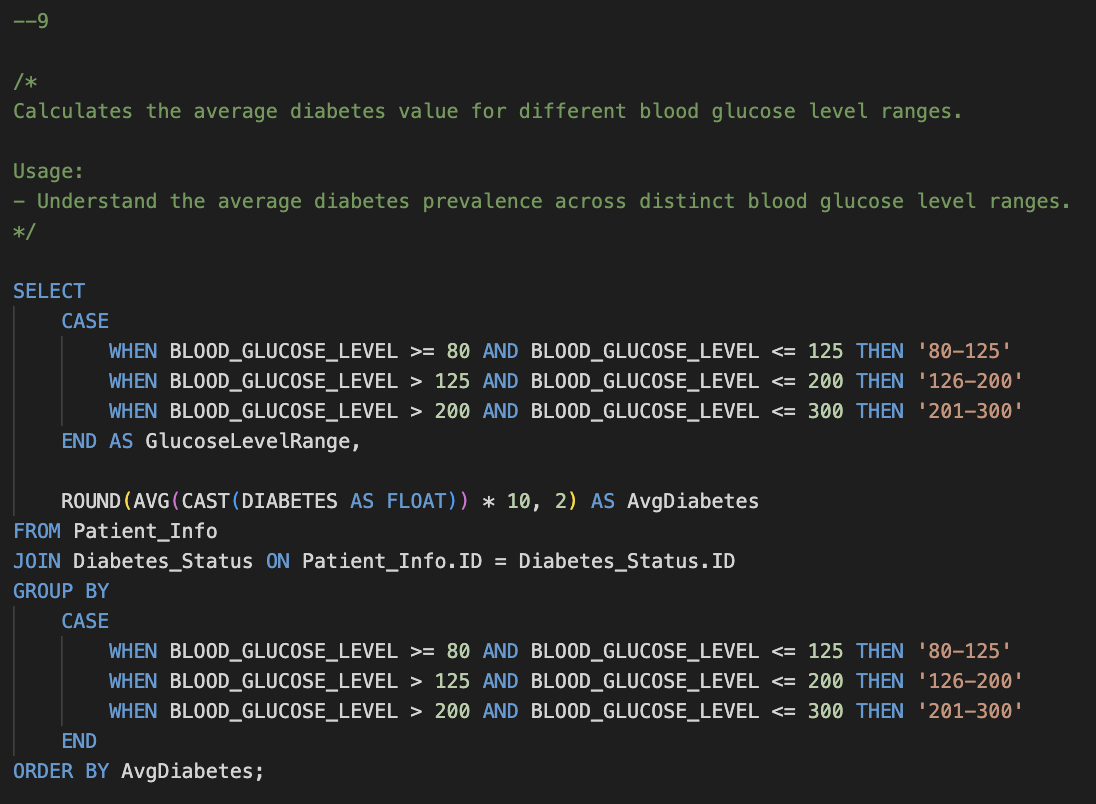
****

**RESULT :**

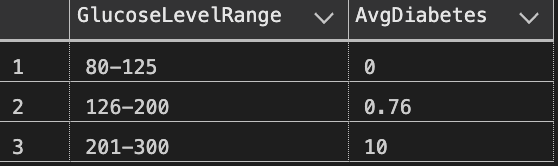
****

**Description:**

This script calculates the average diabetes value for different age categories. By categorizing patients into age groups and determining the average diabetes prevalence in each category, the script allows users to evaluate the distribution of diabetes across distinct age groups.

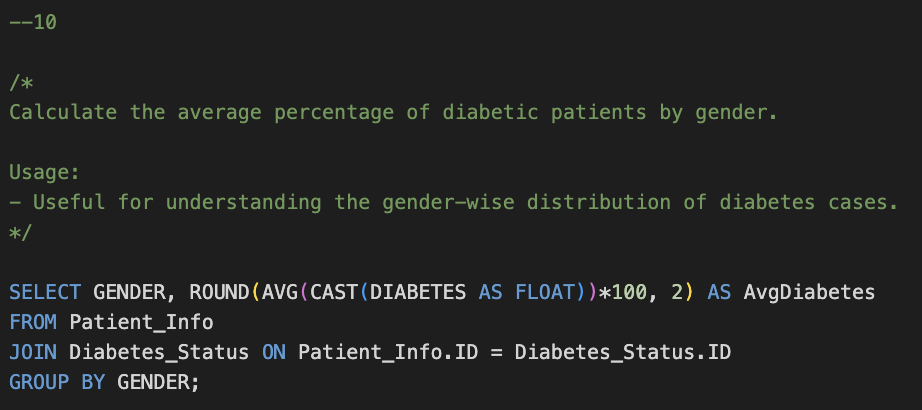
****

**RESULT :**

****

**Description:**

This script calculates the average diabetes value for different blood glucose level ranges. By categorizing patients into blood glucose level ranges and determining the average diabetes prevalence in each range, the script helps users understand the relationship between blood glucose levels and diabetes.

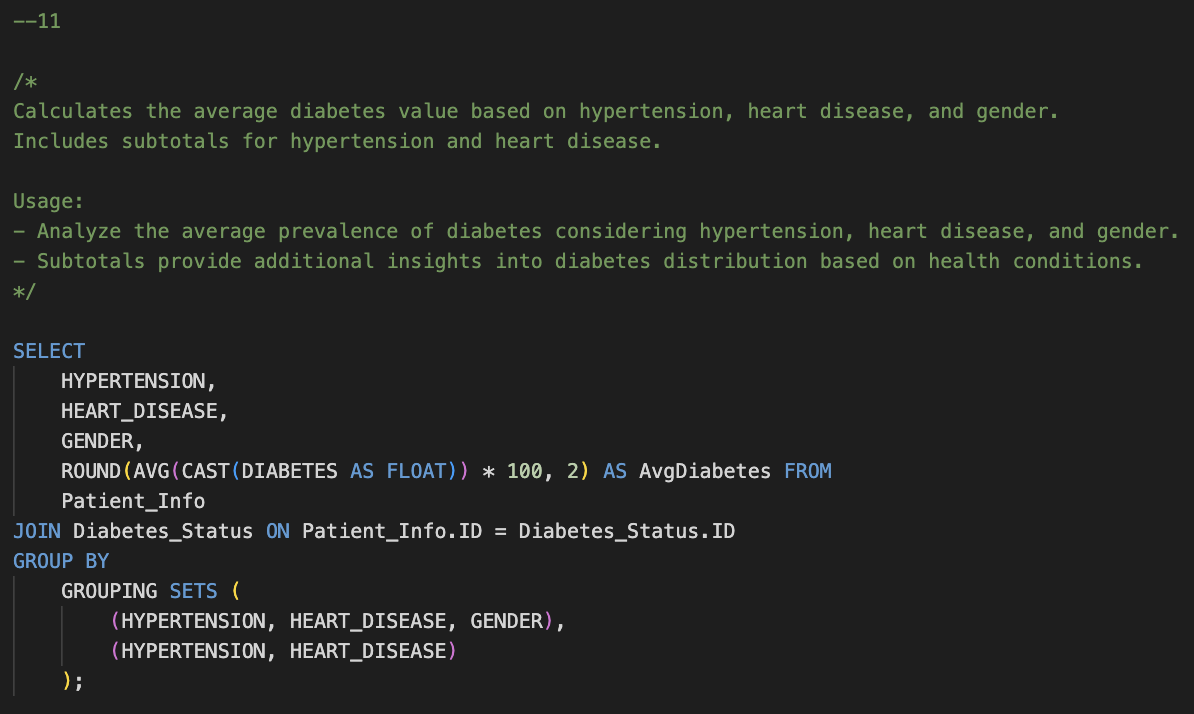
****

**RESULT :**

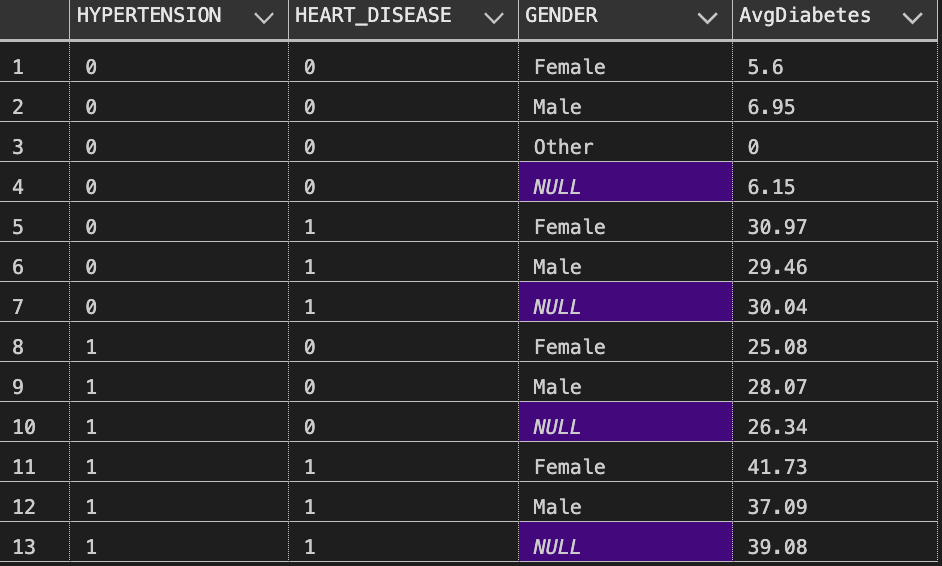
****

**Description:**

This script calculates the average percentage of diabetic patients by gender. Useful for understanding the gender-wise distribution of diabetes cases, the script provides insights into how diabetes prevalence varies among different genders.

****

**RESULT :**

****

**Description:**

This script calculates the average diabetes value based on hypertension, heart disease, and gender. It includes subtotals for hypertension and heart disease, providing additional insights into diabetes distribution based on health conditions.

**VII. Conclusion**

**VIII. Data Source**

“Diabetes Prediction Dataset.” *Kaggle*. N.p., 8 Apr. 2023. Web. <https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset>.

**IX. All Code**

[**https://drive.google.com/file/d/1lL8\_hRtmvwvXLCEvtRCHqYsYVK9yQA5h/view?usp=drive\_link**](https://drive.google.com/file/d/1lL8_hRtmvwvXLCEvtRCHqYsYVK9yQA5h/view?usp=drive_link)