# 📘 README – Health Risk SQL Analysis Project

## 🩺 Project Overview

This SQL project is part of a healthcare analytics initiative to identify patients at risk of chronic illnesses such as diabetes and cardiovascular disease. Using anonymized EHR data, we segment patients by risk, gender, and age, and audit data quality for clinical use and BI dashboarding.

## 📋 Database Table: pateints

The `pateints` table stores the following fields:

|  |  |
| --- | --- |
| Column Name | Description |
| Gender | Patient gender |
| Age | Age in years |
| Hypertension | 1 if present, 0 otherwise |
| Heart\_Disease | 1 if present, 0 otherwise |
| Smoking\_History | Categorical (never, former, current, etc.) |
| BMI | Body Mass Index |
| HbA1c\_level | Glycated hemoglobin level |
| Blood\_Glucose\_level | Blood sugar reading |
| Diabetes | 1 if diabetic, 0 otherwise |
| Age\_Group | Derived age category |
| Lifestyle\_Index | Custom index for habits |
| Is\_At\_Risk | Binary risk flag |
| Normalized\_age | Scaled age for ML |
| Normalized\_BMI | Scaled BMI for ML |

## 🧾 Key SQL Queries & Business Logic

### 🔍 1. High-Risk Patient Detection

Identifies patients with hypertension, elevated HbA1c (≥ 6.5), and blood glucose > 140:

SELECT \* FROM healthcare\_data.pateints  
WHERE hypertension = 1  
 AND HbA1c\_level >= 6.5  
 AND blood\_glucose\_level > 140;

### 📊 2. Gender-Based Diabetes Distribution

Displays diabetes prevalence across gender groups:

SELECT gender, COUNT(\*) AS total\_patients,  
 SUM(diabetes) AS diabetic\_cases,  
 ROUND(SUM(diabetes) \* 100.0 / COUNT(\*), 2) AS diabetes\_rate\_pct  
FROM healthcare\_data.pateints  
GROUP BY gender;

### 👵 3. Elderly High-Risk Patients

Filters patients aged over 60 with high BP and high HbA1c:

SELECT \* FROM healthcare\_data.pateints  
WHERE age > 60  
 AND hypertension = 1  
 AND HbA1c\_level >= 6.5;

### 🧠 4. Risk Level Categorization

Groups patients into "High Risk", "Medium Risk", and "Low Risk":

SELECT age, gender, bmi, HbA1c\_level, blood\_glucose\_level,  
 hypertension, diabetes,  
 CASE  
 WHEN diabetes = 1 OR (hypertension = 1 AND HbA1c\_level >= 6.5) THEN 'High Risk'  
 WHEN HbA1c\_level BETWEEN 5.7 AND 6.4 OR blood\_glucose\_level > 140 THEN 'Medium Risk'  
 ELSE 'Low Risk'  
 END AS risk\_category  
FROM healthcare\_data.pateints;

### ⚠️ 5. Missing Data Detection

Counts NULLs in key EHR fields to audit data quality:

SELECT COUNT(\*) AS total\_records,  
 SUM(CASE WHEN bmi IS NULL THEN 1 ELSE 0 END) AS missing\_bmi,  
 SUM(CASE WHEN HbA1c\_level IS NULL THEN 1 ELSE 0 END) AS missing\_HbA1c,  
 SUM(CASE WHEN blood\_glucose\_level IS NULL THEN 1 ELSE 0 END) AS missing\_glucose,  
 SUM(CASE WHEN smoking\_history IS NULL THEN 1 ELSE 0 END) AS missing\_smoking\_history  
FROM healthcare\_data.pateints;

### 📤 6. Power BI-Ready Summary View

Creates an export-friendly summary view for reporting and dashboarding:

CREATE VIEW patient\_risk\_summary AS  
SELECT age, gender, smoking\_history, bmi, HbA1c\_level,  
 blood\_glucose\_level, hypertension, heart\_disease,  
 diabetes,  
 CASE  
 WHEN diabetes = 1 OR (hypertension = 1 AND HbA1c\_level >= 6.5) THEN 'High Risk'  
 WHEN HbA1c\_level BETWEEN 5.7 AND 6.4 OR blood\_glucose\_level > 140 THEN 'Medium Risk'  
 ELSE 'Low Risk'  
 END AS risk\_category  
FROM healthcare\_data.pateints;

## ✅ Key Outcomes

- Segmented patients by risk using clear clinical thresholds.  
- Identified elderly and hypertensive individuals at higher risk.  
- Quantified diabetes distribution across gender.  
- Audited missing health records to ensure EHR completeness.  
- Built a clean, exportable view for Power BI integration.

## 💡 Recommendations

- Index fields like `diabetes`, `age`, and `HbA1c\_level` for faster query performance.  
- Add a timestamp column for temporal trends in future analyses.  
- Normalize categorical fields (e.g., gender, smoking\_history) using lookup tables.  
- Integrate with dashboards to support hospital decision-makers.