

Diabetes Insights

A Comprehensive insight on Diabetes prediction.

Submitted by:

Krishit Panchal

Guided by:

Prof. Chintan Patel



FLY THE NEST

Fly The Nest – Deets Digital

Acknowledgment

I want to sincerely thank my mentor, Prof. Chintan, for his steadfast assistance and knowledgeable direction during this endeavour. His insightful comments and helpful criticism were quite helpful in getting this analysis finished. I also want to express my gratitude to Fly The Nest for giving me the tools and space I needed to complete this project. Their dedication to encouraging creativity and education has greatly influenced this work.

I want to sincerely thank my family and friends for their unwavering support and inspiration throughout this effort. Their confidence in my skills helped me stay motivated and focused. Furthermore, I'd like to thank the data sources and SQL tools that made this study possible, allowing for a thorough examination of diabetes prediction and insights. Lastly, I would like to express my gratitude to everyone who helped make this project a success, whether directly or indirectly.

Many thanks to all of you.

Abstract

Diabetes is a common chronic illness that is impacted by a number of lifestyle choices and health indicators. "Diabetes Insights," a project, examines the likelihood of a diabetes diagnosis based on a number of variables, such as blood pressure, cholesterol, blood glucose levels, BMI, and lifestyle choices. Patterns were found using SQL queries to comprehend the connection between behaviours including smoking, exercise, food consumption, and the incidence of diabetes. The data offers important insights into preventative strategies by highlighting the influence of heart disease, stress levels, and family history on diabetes risk.

The study, which used structured SQL queries for data extraction and pattern identification, was carried out with assistance from Fly The Nest and under the direction of Prof. Chintan Patel. The results aid in early detection and prevention methods by advancing our understanding of diabetes risk factors.

Content

Acknowledgment	2
Abstract.....	3
1.Introduction.....	5
1.1 Purpose of the Project	5
1.2 Background Information	5
1.3 Objectives.....	5
2. Dataset Description.....	6
2.1 Diabetes Dataset	6
2.2 Habits Dataset.....	6
3. Interpretation and Recommendations	7
4. Limitations of the Study	13
5. Conclusion	14
6. References.....	15
6.1 Dataset Sources:	15
6.2 Tools Used:.....	15
6.3 Documentation:	15
6.4 Fly the Nest Materials:.....	15
6.5 Additional Resources:	15

1.Introduction

1.1 Purpose of the Project

The purpose of this project is to analyze the probability of diabetes diagnosis based on various health indicators and lifestyle factors using SQL queries. By examining parameters such as **BMI, blood pressure, cholesterol, glucose levels, and habits like smoking, exercise, and diet**, this study identifies key patterns influencing diabetes risk. The insights derived from the analysis aim to enhance awareness of preventive measures and contribute to better decision-making for early detection and management of diabetes.

1.2 Background Information

Diabetes is a chronic metabolic disorder characterized by elevated blood glucose levels, resulting from either insufficient insulin production or the body's inability to use insulin effectively. It is a major global health concern, affecting millions and leading to severe complications such as heart disease, kidney failure, and nerve damage. The increasing prevalence of diabetes is linked to factors such as **genetics, lifestyle choices, dietary habits, and stress levels**.

With the rise in sedentary lifestyles and processed food consumption, understanding the contributing factors to diabetes has become crucial. Data analysis techniques, particularly **SQL-based insights**, help uncover significant patterns in diabetes risk by evaluating medical history, lifestyle habits, and key health indicators. This project aims to leverage structured queries to identify trends in diabetes occurrence, aiding in preventive strategies and early diagnosis.

1.3 Objectives

- **Analyze Health Indicators:** Identify the impact of BMI, blood pressure, cholesterol, and glucose levels on diabetes diagnosis.
- **Examine Lifestyle Factors:** Evaluate the role of exercise, smoking, diet, and stress in diabetes occurrence.
- **Identify Risk Patterns:** Use SQL queries to detect correlations between habits, medical history, and diabetes probability.
- **Support Preventive Measures:** Provide data-driven insights to aid in early detection and diabetes prevention strategies.

2. Dataset Description

This project utilizes two structured datasets: **Diabetes Dataset** and **Habits Dataset**, containing crucial health and lifestyle factors influencing diabetes risk.

2.1 Diabetes Dataset

- Includes **demographic details** (Age, Gender, Ethnicity, Income).
- Covers **medical attributes** like **BMI, Blood Pressure, Cholesterol, Glucose Levels, HbA1c, and Insulin Resistance**.
- Records **family history, heart disease history, stress levels**, and final **Diabetes Diagnosis (1 = Diagnosed, 0 = Not Diagnosed)**.

2.2 Habits Dataset

- Captures **lifestyle choices**, including **exercise hours, alcohol consumption, smoking status, and physical activity levels**.
- Includes **dietary habits** like **fast food intake, processed food intake, and daily caloric intake**.
- Considers **medication use and sleep patterns**, helping analyze their impact on diabetes risk.

These datasets are analyzed using **SQL queries** to identify trends and patterns in diabetes diagnosis based on various health and lifestyle factors. The findings help in understanding diabetes risk and supporting preventive measures.

3. Interpretation and Recommendations

3.1 Total Number of People Diagnosed with Diabetes

Query Result: 25,104 people are diagnosed with diabetes out of 50,000 total records.

Interpretation: Nearly 50% of the dataset population is diagnosed with diabetes, highlighting a serious health concern.

Recommendations:

- a. Increase public awareness on diabetes prevention and early detection.
- b. Implement lifestyle modification programs to reduce diabetes risk.
- c. Conduct further analysis on factors contributing to high diagnosis rates.

3.2 Total Number of Males Diagnosed with Diabetes

Query Result: 12,434 males are diagnosed with diabetes.

Interpretation: A significant portion of diagnosed individuals are male, indicating potential gender-based health differences.

Recommendations:

- a. Conduct targeted health campaigns for men focusing on lifestyle changes.
- b. Identify key risk factors specific to males contributing to diabetes.
- c. Encourage routine check-ups for early detection.

3.3 Total Number of Females Diagnosed with Diabetes

Query Result: 12,670 females are diagnosed with diabetes.

Interpretation: The number of females diagnosed with diabetes is slightly higher than males. Hormonal factors, pregnancy-related diabetes risks (gestational diabetes), and stress levels may contribute to this trend.

Recommendations:

- a. Implement women-centered diabetes awareness programs.
- b. Encourage early screening during pregnancy to detect gestational diabetes.
- c. Offer stress management programs, as stress may play a role in increased diabetes risk.

3.4 Diabetes Diagnosis Based on Stress Levels

Interpretation: A higher stress level correlates with a greater likelihood of diabetes diagnosis. Chronic stress can lead to hormonal imbalances, increased glucose levels, and unhealthy lifestyle choices, all of which contribute to diabetes risk.

Recommendations:

- a. Introduce stress management programs such as yoga, meditation, and counseling.
- b. Conduct awareness campaigns about the impact of stress on diabetes.
- c. Encourage companies to implement work-life balance policies to reduce stress.

3.5 Diabetes Diagnosis Based on Physical Activity Levels

Interpretation: Individuals with low physical activity have higher diabetes prevalence, highlighting the role of exercise in diabetes prevention.

Recommendations:

- a. Encourage daily physical activity through community fitness programs.
- b. Provide fitness incentives to promote healthier lifestyles.
- c. Educate people on the importance of an active lifestyle in preventing diabetes.

3.6 Impact of Medication Use on Diabetes Diagnosis

Interpretation: The results indicate that a significant number of diabetic individuals rely on medications. This suggests that either diabetes management requires pharmaceutical intervention or that certain medications may contribute to the risk of diabetes. Some medications, such as steroids and certain antipsychotics, have been linked to increased blood sugar levels. Conversely, individuals already diagnosed with diabetes may be using prescribed medication for blood sugar control.

Recommendations:

- a. Conduct further analysis to differentiate between diabetes-causing and diabetes-controlling medications.
- b. Encourage regular monitoring of blood sugar levels for individuals on long-term medication.
- c. Promote lifestyle changes alongside medication for better diabetes management.
- d. Educate patients about potential side effects of medications affecting glucose metabolism.

3.7 Ethnicity-Based Distribution of Diabetes Cases

Interpretation: The results show the distribution of diabetes cases across different ethnic groups. Certain ethnic groups may have a higher prevalence of diabetes due to genetic predisposition, lifestyle factors, dietary habits, or healthcare accessibility. Studies suggest that some populations, such as Asian, Black, and Hispanic individuals, may have a higher risk of developing diabetes due to genetic susceptibility and metabolic differences.

Recommendations:

- a. Conduct targeted awareness campaigns for ethnic groups with higher diabetes prevalence.
- b. Implement personalized prevention strategies considering dietary and genetic risk factors.
- c. Improve access to healthcare and early screening programs in high-risk communities.
- d. Encourage culturally relevant lifestyle modifications to prevent diabetes progression.

3.8 Impact of Family History on Diabetes Diagnosis

Interpretation: The results indicate that individuals with a family history of diabetes are more likely to be diagnosed, suggesting a strong genetic influence on diabetes risk. However, lifestyle choices remain a crucial factor in determining disease onset.

Recommendations:

- a. Encourage early screening and regular check-ups for high-risk individuals.
- b. Promote healthy lifestyle habits to reduce the risk of diabetes development.

3.9 Average Health Metrics of Diabetic Patients

Interpretation: The average values for HbA1c, glucose, cholesterol, blood pressure, and BMI indicate the typical health profile of a diabetic patient, with elevated levels in most categories.

Recommendations:

- a. Encourage lifestyle modifications to improve these health markers.
- b. Regular monitoring can help in early intervention and better disease management.

3.10 Impact of Heart Disease on Diabetes Diagnosis

Interpretation: The results show that individuals with a history of heart disease have a higher likelihood of being diagnosed with diabetes, suggesting a strong correlation between cardiovascular and metabolic health.

Recommendations:

- a. Encourage heart-healthy habits to reduce diabetes risk.
- b. Patients with heart disease should undergo regular diabetes screenings.

3.11 Insulin Resistance in Diabetic vs. Non-Diabetic Individuals

Interpretation: People diagnosed with diabetes have higher insulin resistance compared to non-diabetics, confirming insulin dysfunction as a key factor in diabetes.

Recommendations:

- a. Promote physical activity and a balanced diet to improve insulin sensitivity.
- b. Regular insulin resistance testing for early detection of prediabetes.

3.12 Average Age of Diabetes Diagnosis by Gender

Interpretation: The average age for diabetes diagnosis is around 53 years for both males and females, indicating middle age as a critical period for diabetes onset.

Recommendations:

- a. Implement preventive health checkups for individuals over 40.
- b. Encourage lifestyle changes to delay or prevent diabetes.

3.13 Smoking and Diabetes Risk

Interpretation: Current smokers have the highest risk of diabetes, followed by former smokers, while non-smokers have the lowest risk. This indicates that smoking negatively impacts metabolic health.

Recommendations:

- a. Encourage smoking cessation programs to lower diabetes risk.
- b. Educate individuals about the long-term metabolic effects of smoking.

3.14 Alcohol Consumption and Diabetes Risk

Interpretation: The results suggest that higher alcohol consumption is associated with an increased risk of diabetes, possibly due to its impact on liver function and insulin resistance.

Recommendations:

- a. Advise moderation in alcohol intake to prevent diabetes.
- b. Conduct awareness programs on alcohol's impact on metabolic health.

3.15 Average Caloric Intake of Diabetic vs. Non-Diabetic Individuals

Interpretation: The average daily caloric intake is similar for both diabetic and non-diabetic individuals, indicating that total calorie consumption alone may not be the primary risk factor.

Recommendation:

- a. Focus on the quality of calories consumed rather than just the quantity, emphasizing balanced nutrition.

3.16 Average Sleep Duration in Diabetic vs. Non-Diabetic Individuals

Interpretation: Diabetic individuals tend to sleep more than non-diabetic individuals, possibly due to fatigue and metabolic imbalances.

Recommendation:

- a. Encourage good sleep hygiene and monitoring of sleep patterns for better diabetes management.

3.17 Fast Food & Processed Food Intake and Diabetes Risk

Interpretation: Individuals who consume higher amounts of fast food and processed food are at a greater risk of being diagnosed with diabetes due to unhealthy dietary habits.

Recommendations:

- a. Promote nutrient-dense, whole-food diets to reduce diabetes risk.
- b. Educate on the negative impact of processed food on metabolic health.

3.18 Exercise Levels and Diabetes Risk

Interpretation: Individuals who exercise more than average have a lower chance of being diagnosed with diabetes, emphasizing the role of physical activity in prevention.

Recommendations:

- a. Encourage at least 150 minutes of moderate exercise per week.
- b. Promote strength training as it helps improve insulin sensitivity.

3.19 Top 5 Ethnic Groups with the Highest Diabetes Diagnosis Rates

Interpretation: The Black, White, and Hispanic ethnic groups show the highest diabetes rates, highlighting potential genetic and lifestyle influences.

Recommendations:

- a. Conduct targeted awareness and prevention programs for high-risk ethnic groups.
- b. Promote culturally relevant dietary and lifestyle modifications.

3.20 Identifying High-Risk Patients Based on Multiple Health Conditions

Interpretation: Patients with high BMI, blood pressure, cholesterol, and glucose levels are classified as high-risk individuals for diabetes.

Recommendations:

- a. Implement early intervention programs for high-risk individuals.
- b. Encourage regular monitoring and lifestyle modifications.

3.21 Relationship Between Lifestyle Habits and Diabetes Diagnosis

Interpretation: Physical activity significantly impacts BMI, blood pressure, and glucose levels, influencing diabetes prevalence.

Recommendations:

- a. Encourage regular exercise to maintain healthy metabolic markers.
- b. Implement personalized fitness programs for diabetes prevention.

3.22 Effect of Sleep Duration on Diabetes

Interpretation: Individuals who sleep less than 6 hours or more than 8 hours have a higher diabetes rate, indicating a link between sleep and metabolic health.

Recommendations:

- a. Promote consistent sleep patterns (6-8 hours per night).
- b. Educate on the impact of sleep deprivation on insulin resistance.

3.23 Correlation Between Alcohol, Smoking, and Diabetes Risk

Interpretation: Smokers and individuals with high alcohol consumption tend to have a higher BMI and greater diabetes prevalence.

Recommendations:

- a. Promote smoking cessation programs to lower risk.
- b. Raise awareness of the negative impact of alcohol on metabolic health.

3.24 Identifying Sedentary Patients with High BMI

Interpretation: Individuals with low physical activity and high BMI are at an increased risk of developing diabetes.

Recommendations:

- a. Encourage regular exercise and active lifestyle habits.
- b. Implement community-based fitness programs for at-risk groups.

3.25 Effect of Fast Food Intake on Blood Sugar

Interpretation: Frequent fast food consumption increases glucose levels, placing individuals in high or moderate risk categories for diabetes.

Recommendations:

- a. Promote healthy eating habits and home-cooked meals.
- b. Educate on the impact of fast food on blood sugar regulation.

4. Limitations of the Study

1. Data Accuracy & Completeness:

- The analysis is based on the available dataset, which may have missing or inaccurate data affecting the reliability of insights.

2. Correlation vs. Causation:

- The project identifies associations between habits and diabetes but does not establish causality (e.g., smoking is linked to diabetes but may not directly cause it).

3. Sample Representation Bias:

- If the dataset is not diverse enough, findings may not be generalizable to all populations, limiting the study's applicability.

4. Unaccounted External Factors:

- Certain genetic, environmental, and socio-economic factors influencing diabetes risk are not fully considered in the dataset.

5. Static Data Analysis:

- The dataset provides a snapshot in time rather than a longitudinal study that tracks how lifestyle changes affect diabetes risk over time.

6. SQL Query Limitations:

- SQL helps in structured data analysis but does not provide advanced predictive modeling like machine learning, which could improve insights.

5. Conclusion

This SQL-based Diabetes Insights Project analyzes key factors influencing diabetes diagnosis, focusing on lifestyle habits, medical history, and demographic attributes. By executing various SQL queries, the study uncovers significant correlations between diabetes and factors such as family history, medication use, physical activity, diet, sleep, and smoking or alcohol consumption. The findings reveal that unhealthy habits and pre-existing conditions significantly increase diabetes risk, while active lifestyles and balanced nutrition help mitigate it. The analysis also highlights how different ethnic groups, income levels, and exercise patterns influence diabetes prevalence. Despite limitations like data completeness and the inability to establish direct causation, the project provides valuable data-driven insights that can guide preventive healthcare strategies and policymaking. It also emphasizes the importance of early intervention and lifestyle modifications in reducing the risk of diabetes. Conducted under the guidance of Prof. Chintan Patel and supported by Fly The Nest, this study highlights the role of structured data analysis in understanding and managing diabetes risks effectively.

6. References

6.1 Dataset Sources:

- Source: Diabetes Dataset and Lifestyle Habits Dataset
- Description: The dataset contains patient health records, lifestyle habits, and medical history, providing insights into diabetes diagnosis and its influencing factors.

6.2 Tools Used:

- MySQL: For executing SQL queries to analyze diabetes-related patterns and trends.
- MS Excel: For organizing, cleaning, and performing preliminary data exploration.
- MS Word: For documenting and structuring the final project report.

6.3 Documentation:

- MySQL Documentation: <https://dev.mysql.com/doc/>
- Microsoft Excel Help: <https://support.microsoft.com/excel>

6.4 Fly the Nest Materials:

- Source: "Fly the Nest" Data Science Course – SQL Learning Resources

6.5 Additional Resources:

- SQL Tutorials: [W3Schools SQL Tutorial](#)
- Online Learning Platforms: SQL and Data Science Courses on YouTube and Coursera