EXPLORING DIABETES: INSIGHTS INTO LIFESTYLE AND HEALTH RISKS

Swarupa Murala
email: swarupam@buffalo.edu

Tejaswini Chowdam Nallagondappa
email: tchowdam@buffalo.edu

Arunkarthik Periyaswamy
email: Aperiyas@buffalo.edu

I. INTRODUCTION

Diabetes is becoming an increasingly serious global health issue, impacting millions of lives across various regions and socio-economic groups. This chronic condition not only strains individuals but also places immense pressure on healthcare systems worldwide due to its escalating prevalence. Lifestyle factors like poor diet, lack of physical activity, and increasing rates of obesity are major contributors to the rise of type 2 diabetes, particularly as urban living becomes more common. In response, there is a growing need for robust public health initiatives that effectively address these risk factors. This report delves into how lifestyle choices influence diabetes and seeks to provide practical insights that can help combat its spread. By analyzing underlying trends and behaviors, we aim to support the development of policies and personal habits that enhance health outcomes and reduce the burden of diabetes around the world.

II. BACKGROUND AND SIGNIFICANCE

The escalating incidence of diabetes across the globe is a profound public health concern, impacting countless individuals with severe conditions like heart disease, kidney failure, and impaired vision. This challenge is even more pronounced in economically disadvantaged areas, highlighting the pressing need for robust and effective preventative measures.

Diabetes often begins in childhood, fueled by inadequate health education and compounded by the fast-paced nature of contemporary life. Many individuals find themselves in work environments that promote prolonged sitting and urban areas that favor convenience over wellness. These settings typically offer easy access to processed food and lack suitable areas for physical activity, issues that are intensified by our increasingly digital and automated lifestyles.

The economic burden of diabetes is substantial, encompassing continual medical expenses, medications, and possible income loss due to health-related disabilities. Critical health indicators such as BMI, blood pressure, and cholesterol levels are essential for assessing the risk of diabetes, underscoring the importance of early and targeted interventions.

This project transcends mere data analysis; it's about real human lives. By crafting a predictive model centered on lifestyle and health metrics, we aim to pinpoint individuals at risk of diabetes early on. The objective is to facilitate preventative healthcare measures that can stave off or even prevent the onset of diabetes. This initiative is designed to influence not only personal health decisions but also community planning and corporate policies, promoting environments that support active living and healthier eating

habits. In essence, this project represents a proactive step towards a healthier future, aiming to alleviate the global diabetes burden through informed, data-driven health strategies

III. PROJECT POTENTIAL

This project has the potential to contribute to the problem domain of diabetes management and prevention, which is substantial. By developing a predictive model that leverages lifestyle and health metrics to forecast the risk of diabetes, this research aims to intervene at an early stage, potentially preventing the onset of diabetes in at-risk individuals. Such predictive capabilities are crucial for several reasons:

Building Healthier Communities: The insights gained from our model can help shape policies that transform urban areas into spaces that encourage healthy living, making it easier for everyone to access nutritious food and engage in physical activities.

Pushing Frontiers with Technology: By integrating AI and machine learning, this project isn't just about managing diabetes—it's about revolutionizing how we approach healthcare to make it smarter and more effective.

Making Decisions with Confidence: This project empowers healthcare professionals to base their strategies on reliable data, ensuring that every decision made is backed by a solid foundation of evidence.

Impacting Lives Globally: Diabetes knows no borders, and neither do the solutions we're developing. This project has the potential to be adapted and implemented around the world, providing vital health benefits on a global scale.

Customized Care for Everyone: Our project champions personalized healthcare, where treatment and prevention plans are as unique as the individuals they are designed for, enhancing engagement and the success of interventions.

Sharing Knowledge, Shaping Futures: By identifying and analyzing trends in diabetes, this project contributes to the broader understanding of the disease, supporting health initiatives that can change lives worldwide.

Learning for Better Health: The predictive model is more than a tool—it's a source of knowledge that helps patients and doctors alike understand the profound impact of lifestyle on health, inspiring informed choices and better living.

Enhancing Everyday Well-being: By addressing diabetes before it starts, we can help individuals avoid the complications that diminish life quality, helping them lead healthier, fuller lives.

Acting Today, Securing Tomorrow: With early prediction, we can intervene sooner, potentially altering the course of health outcomes and emphasizing the importance of proactive healthcare.

Easing the Economic Burden: Early intervention doesn't just save lives—it saves money too. By catching diabetes early, we can avoid the high costs associated with treating advanced stages of the disease.

Expanding Access to Health Tools: Through digital delivery, our predictive model makes crucial health monitoring accessible to more people, particularly in areas where healthcare resources are limited.

Designing for Health: By understanding how the environment affects health, we can promote changes that make healthy choices the easiest choices, supporting everyone's well-being.

IV. DATASET

The Diabetes prediction dataset comprises medical and demographic information from patients, including their diabetes status (positive or negative). It consists of 1,000,001 rows and 9 columns, featuring variables such as age, gender, BMI, hypertension, heart disease, smoking history, HbA1c level, and blood glucose level. This dataset is useful for developing machine learning models aimed at predicting diabetes based on patient history and demographics. Healthcare professionals can utilize such models to identify at-risk individuals and create personalized treatment strategies. Additionally, researchers can leverage this data to investigate the correlations between medical and demographic factors and the risk of developing diabetes.

URL: **https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset**

V. DATA CLEANING

The data cleaning and preparation phase is foundational for ensuring the accuracy and integrity of our analysis. Here’s how we refined the dataset to lay a robust groundwork for our predictive modeling efforts on diabetes risk:

Removing Duplicates

What We Did: We scoured the dataset to identify and remove any duplicate records.

Why It Matters: This step is crucial for eliminating redundant information that could skew our analysis, ensuring each data point represents a unique individual.

Handling Missing Values

What We Did: We checked for gaps in our data and opted to remove any records with missing information to maintain a high standard of data integrity.

Why It Matters: Clean data is synonymous with reliable results. By ensuring completeness in our dataset, we enhance the accuracy of our predictive outcomes.

Trimming and Standardization

What We Did: We standardized text data within the 'gender' and 'smoking_history' columns by removing excess whitespace and ensuring uniform formatting.

Why It Matters: This meticulous attention to detail prevents minor discrepancies from leading to major misinterpretations in our analysis.

Data Type Conversions

What We Did: We transformed categorical variables like 'gender' and 'smoking_history' into numerical formats using label encoding.

Why It Matters: This conversion enables our statistical models to process and analyze these variables effectively, enriching our data-driven insights.

Feature Engineering

Feature engineering allowed us to enhance our dataset’s utility by creating new features and refining existing ones to reveal patterns and relationships.

1.Age Group Categorization

What We Did: We grouped ages into categories that reflect different life stages, enhancing our analysis across diverse age groups.

Why It Matters: This approach simplifies the analysis, helping us to understand how diabetes risk varies across age demographics.

2.BMI Categorization

What We Did: We categorized BMI values into health-based segments to study their impact on diabetes risk.

Why It Matters: This categorization aids in visualizing how body weight influences diabetes risk, enhancing our understanding of health trends.

3.Interaction Features

What We Did: We created features that capture the combined effects of conditions like hypertension and heart disease on diabetes.

Why It Matters: These combined features offer insights into complex health interactions, potentially unveiling critical risk factors.

4. Health Risk Score

What We Did: We calculated a health risk score to summarize the overall risk based on several health indicators.

Why It Matters: This composite score is a powerful tool for identifying individuals at high risk, enabling targeted preventive measures.

Outlier Management: Ensuring Balanced Analysis

Managing outliers is essential to prevent skewed analyses and ensure the robustness of our predictions.

Identification and Assessment

What We Did: We employed statistical methods and visual tools to pinpoint outliers, evaluating whether they represent errors or rare, significant conditions.

Why It Matters: Accurate identification ensures our dataset truly reflects the general population, avoiding biases introduced by extreme data points.

Handling Techniques

What We Did: Depending on their nature, we removed, transformed, or imputed outliers, ensuring they did not unduly influence our analysis.

Why It Matters: This careful handling maintains the quality of our modeling, ensuring that our predictions are both reliable and applicable.

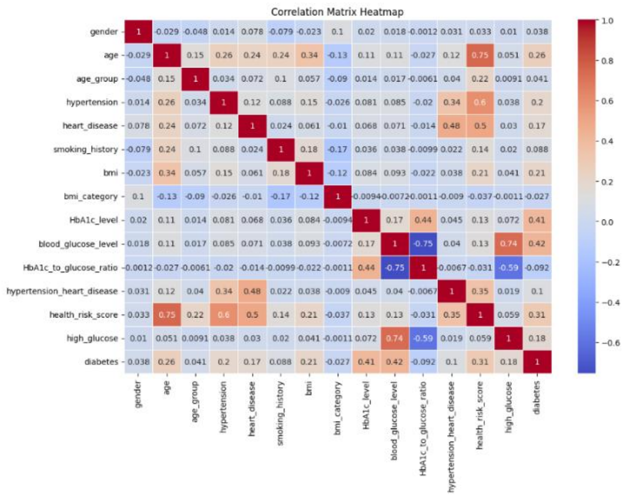
Validation

What We Did: After managing outliers, we validated our methods through cross-validation to ensure they enhanced our model's performance.

Why It Matters: This step confirms that our data preparation efforts have indeed prepared the ground for more accurate and predictive insights.

VI. EXPLORATORY DATA ANALYSIS

1. Correlation between columns



Age and Health Risk Score: The analysis reveals a strong positive correlation between age and the health risk score. This relationship indicates that as individuals age, their cumulative health risks increase, which is an essential consideration for health management and preventative measures in older populations.

Hypertension and Health Risk Score: It was found that hypertension is significantly associated with higher health risk scores. This correlation confirms that hypertension is a critical factor in overall health risks, corroborating its role in exacerbating conditions like diabetes and other severe health issues.

Notable Negative Relationship

Blood Glucose Level and HbA1c to Glucose Ratio: A strong negative correlation was observed, suggesting that higher immediate blood glucose levels often lead to lower HbA1c to glucose ratios. This may indicate episodes of short-term glucose spikes that are not yet reflected in the HbA1c measurements, which are indicative of long-term glucose management.

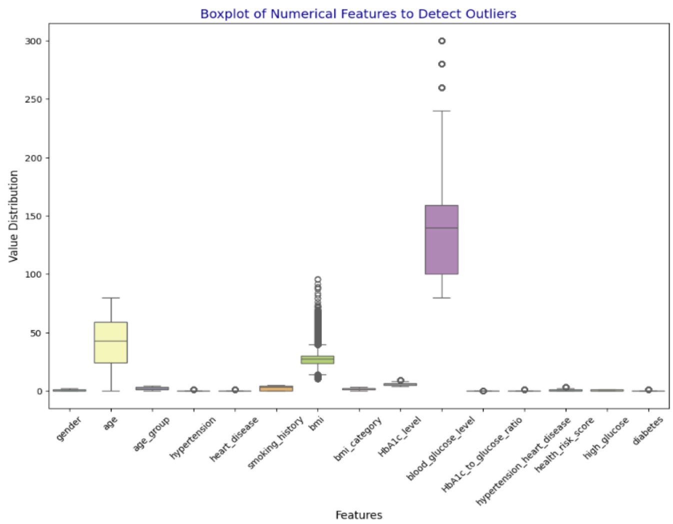
Key Risk Indicators

Health Risk Score Dynamics: While age and hypertension show a strong correlation with the health risk score, the relationship between the health risk score and blood glucose levels is moderate. This suggests that blood glucose levels, while influential, may not be as dominant in the overall health risk score as age and hypertension.

Diabetes and Blood Glucose Level: There is a significant positive correlation indicating that higher blood glucose levels are a robust predictor of diabetes. This finding underscores the importance of regular monitoring of blood glucose levels as a fundamental aspect of diabetes management.

Diabetes and HbA1c Level: The association between elevated HbA1c levels and the presence of diabetes is strongly positive. HbA1c levels are invaluable in the clinical setting, providing a snapshot of an individual's long-term blood sugar levels and playing a crucial role in the diagnosis and management of diabetes

2. Finding the Outliers present in all the columns



Geographic Variation: When analyzing data by geographic region, stark differences emerge, particularly with metrics like BMI. In areas with limited access to recreational facilities or healthier food options, higher instances of extreme BMI values suggest regional disparities that could impact public health strategies.

Temporal Trends in Outliers: Examining how outliers change over time, especially for dynamic metrics like blood glucose levels and HbA1C, reveals potential seasonal variations or shifts in data collection methods. These trends are vital for understanding long-term changes in patient management and the effectiveness of healthcare strategies.

Detailed Feature Analysis:

Diverse Data Range: The analysis shows that numerical features such as age, HbA1C level, and BMI have widely varying distributions. Age, in particular, has a broad spread, indicating a dataset rich in demographic diversity, which is crucial for comprehensive risk assessments.

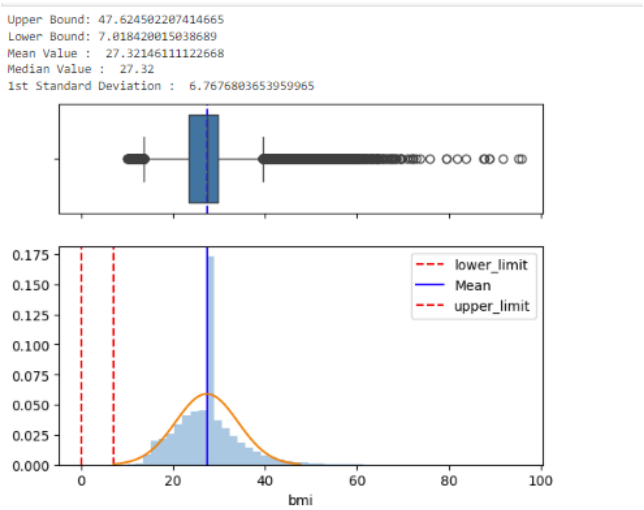
Outlier Detection: Significant outliers in features like BMI and HbA1C level highlight cases of extreme values, which could be due to measurement errors or genuine extremes in patient conditions. Particularly, outliers in HbA1C levels might pinpoint critical scenarios in diabetes management needing closer investigation.

Interactions and Subgroup Dynamics:

Interaction Effects: The interplay between features like hypertension and BMI or age and diabetes unveils complex relationships that could influence diabetes risk. For example, the compounded risks associated with high BMI are notably severe in older adults or those already managing hypertension.

Subgroup Analysis: Analyzing data within specific subgroups—defined by variables such as age, gender, or existing conditions—uncovers nuanced patterns that broad analyses might miss. These findings are especially pronounced in variations in how risk factors affect different genders or age groups, offering tailored insights for targeted interventions.

3. OUTLIERS PRESENT IN BMI FEATURE



The boxplot at the top offers a quick visual summary, showing the distribution's spread and central values like the median, which divides the data into two halves. This gives a clear picture of where most of the data points lie and highlights those that are exceptions or potentially erroneous, standing outside the normal range.

Directly beneath the boxplot, the histogram offers a detailed breakdown of the BMI values, grouping them into bins. This allows us to observe the shape of the distribution—whether it is symmetric, skewed, or perhaps shows multiple peaks, each of which can tell a different story about the population's health.

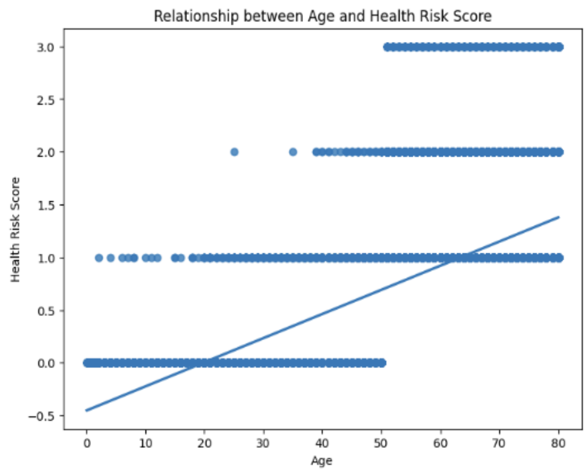
Overlaid on the histogram is a normal distribution curve based on the mean and standard deviation calculated from the data. This curve is useful for comparing the actual distribution of BMI to a theoretical normal distribution, highlighting deviations such as skewness or unusual variability.

The display of key statistics like the mean, median, and standard deviation alongside the visualizations provides a snapshot of the data’s central tendency and variability. This not only helps in understanding what a typical BMI might be but also how varied the BMI values are across the population.

Additionally, the script marks the bounds at three standard deviations from the mean, which helps identify extreme values considered statistical outliers. These are the values that are significantly different from the norm and could indicate special cases or issues with data collection.

By comparing the mean and median, and observing how the data points are distributed around these values, we can assess the skewness of the BMI distribution. A significant difference between these two measures might suggest a skewed distribution, which can have implications for how we interpret the overall health of the population.

4. ANALYSIS OF AGE AND HEALTH RISK SCORE CORRELATION



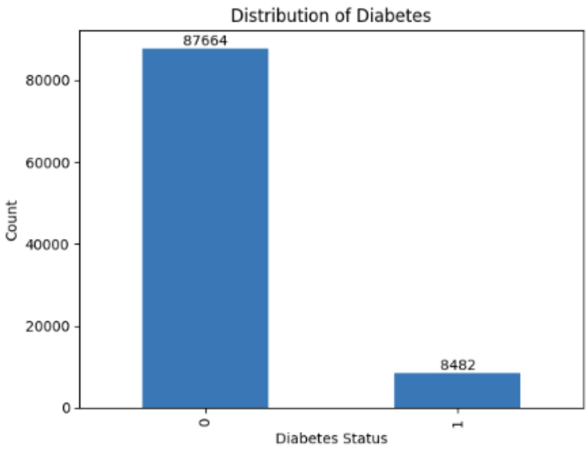
Growing Risks with Age: The graph clearly shows that as people get older, their health risk scores generally tend to rise. This upward trend in the plot underscores a common understanding that health complications increase with age, reflecting the cumulative effect of aging on health.

Discrete Nature of Risk Scores: Interestingly, the health risk scores seem to be grouped into distinct categories like 0, 1, 2, and 3. This pattern suggests that the health risk score is categorized into levels, rather than being a continuous measure. Such a grouping helps in understanding health status in a more segmented way, which can be particularly useful for tailored health interventions.

Higher Risks in the Elderly: People in their sixties and beyond consistently register higher health risk scores, often peaking at the highest category. This observation emphasizes the increased vulnerability among older adults, who are more likely to encounter serious health issues.

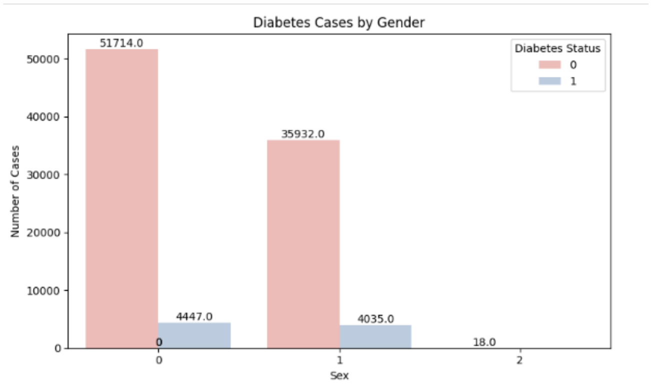
Youthful Variability: There's a noticeable spread in the health risk scores among younger individuals, which points to a diverse array of factors that could influence their health. This variability might be due to different lifestyles, genetic factors, or certain health conditions that emerge early in life, suggesting that young people's health outcomes can be quite unpredictable.

5. DISTRIBUTION OF TARGET VARIABLE



The bar chart shows that the majority of individuals in the dataset have no diabetes (0), with a count of 87,664. A smaller number of individuals have diabetes (1), with a count of 8,482.

6.ANALYSIS OF DIABETES CASES BY GENDER

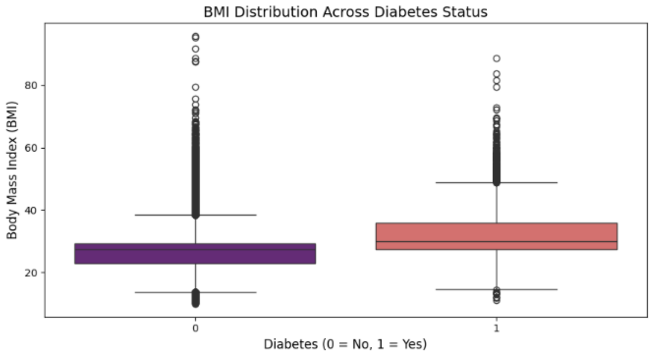


Overall Distribution of Diabetes by Gender: The chart shows that the majority of both male and female participants do not have diabetes, as seen in the significantly taller pink bars for both genders. Specifically, there are 51,714 non-diabetic males and 35,932 non-diabetic females, highlighting a substantial non-diabetic majority within the dataset.

Comparison of Diabetic Cases Between Genders: When it comes to diabetes prevalence, males display a slightly higher number of cases, with 4,447 instances, compared to females, who have 4,035 cases. This suggests a modest but noticeable difference in diabetes occurrence between the two genders.

Identifying Data Anomalies: The data includes an unusual category labeled as "2" under gender, which accounts for only 18 cases. This small number suggests either a potential categorization error or the presence of a non-binary or unspecified gender category that warrants further investigation to ensure data accuracy and inclusivity.

7.BMI DISTRIBUTION ACROSS DIABETES STATUS

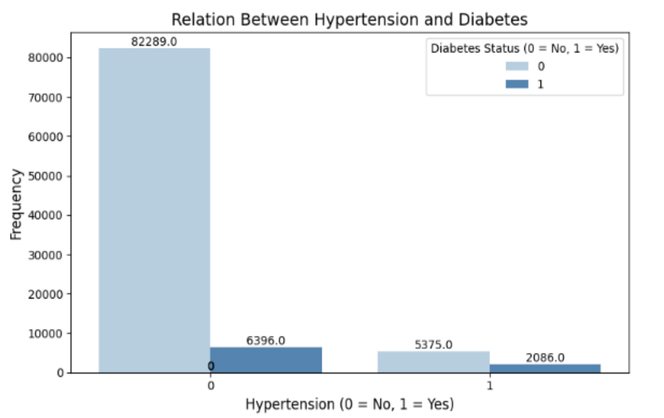


Comparative Analysis of BMI Medians: The boxplot indicates that the median BMI for individuals diagnosed with diabetes (marked as '1') is notably higher than for those without the condition (marked as '0'). This trend reinforces the established link between higher BMI and increased diabetes risk, suggesting that a higher body mass index is a common characteristic among those diagnosed with the condition.

Variability and Presence of Outliers: Both diabetic and non-diabetic groups display a considerable number of outliers, indicating that there are individuals with BMIs significantly higher or lower than the typical range in both categories. Notably, the variability in BMI is more pronounced among individuals with diabetes, as evidenced by a wider spread in their boxplot. This increased variability highlights that while a higher BMI is generally associated with diabetes, individuals with the condition can vary widely in their body mass index.

Overlap in BMI Distributions: Despite notable differences in median BMI values, the boxplots reveal that the interquartile ranges (the middle 50% of values) for both groups overlap considerably. This overlap suggests that while higher BMI can be associated with an increased risk of diabetes, the BMI distributions for individuals with and without diabetes are not entirely distinct. Many individuals without diabetes also present with higher BMI, indicating that BMI alone may not be a definitive predictor of diabetes.

8. RELATION BETWEEN HYPERTENSION AND DIABETES



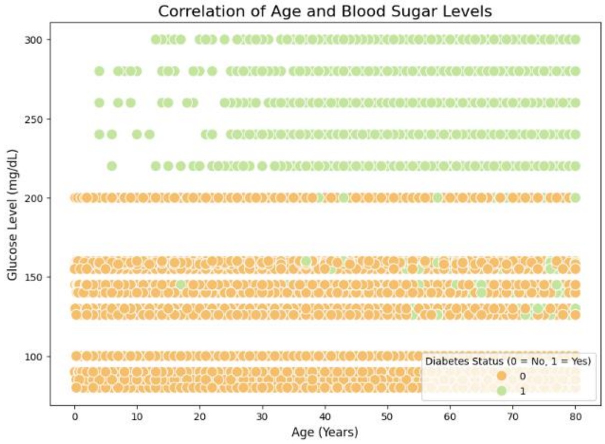
Predominance of Non-Hypertensive and Non-Diabetic Individuals: The most prominent observation from the plot is the large number of individuals who neither have hypertension nor diabetes, with a total count of 82,289. This group represents the majority, highlighting a significant portion of the population that is free from both conditions, which is a positive indicator of health within this demographic.

Intersection of Hypertension and Diabetes: While the occurrence of both hypertension and diabetes within the same individuals is less common, the data still show that 2,086 people are managing both conditions. This co-occurrence, though relatively lower, underscores the importance of monitoring individuals with hypertension for potential diabetic complications, as the intersection of these conditions can complicate treatment and management strategies.

Hypertension without Diabetes: Notably, there are 5,375 individuals who have hypertension but do not have diabetes. This observation suggests that hypertension can occur independently of diabetes and is a prevalent health issue on its own. It highlights the need for targeted hypertension management programs that can operate separately from diabetes interventions.

Diabetes without Hypertension: Among those diagnosed with diabetes, a significant number, 6,396, do not have hypertension. This group illustrates that while diabetes and hypertension often are discussed together due to common risk factors, many individuals with diabetes do not face compounded risks with hypertension. This can influence the approach to diabetes management, emphasizing the need for personalized treatment plans based on individual risk profiles rather than a one-size-fits-all solution.

9.CORRELATION OF AGE AND BLOOD SUGAR LEVELS

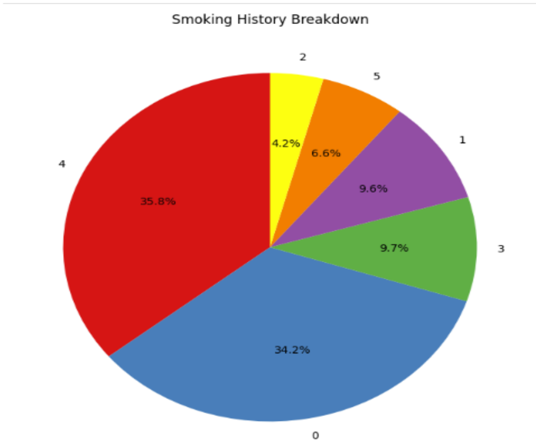


Patterns in Glucose Levels: The visualization highlights noticeable clusters of glucose levels, particularly around the 100 mg/dL, 150 mg/dL, and 200 mg/dL marks. Individuals with diabetes, marked in green, predominantly appear at higher glucose concentrations, typically above 150 mg/dL. In contrast, non-diabetic individuals, shown in orange, are mostly found at lower glucose levels. This pattern suggests that higher glucose levels are a strong marker for diabetes, a crucial insight for medical professionals monitoring for potential diabetic conditions.

Wide Age Range in Glucose Distribution: Interestingly, the scatter plot shows that both diabetic and non-diabetic individuals span a broad age range, from the very young to the elderly, up to 80 years. This widespread distribution underscores that while age contributes to overall health status, it does not singularly predict blood glucose levels. Such a finding emphasizes the complexity of diabetes as a condition influenced by multiple factors beyond age alone.

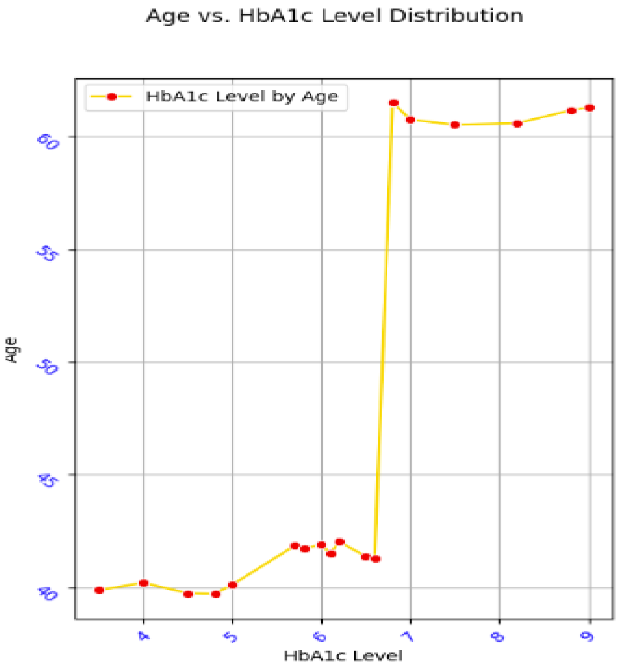
Elevated Glucose Levels Among Diabetics: Consistent with medical understanding, diabetic individuals tend to exhibit higher glucose levels, often clustering between 200–300 mg/dL. Meanwhile, those without diabetes generally maintain levels below 150 mg/dL. This distinction is visually striking in the plot and reinforces the need for vigilant glucose monitoring as part of diabetes management and diagnosis

10.SMOKING HISTORY BREAKDOWN



1. Category 4 (Red):
Represents the largest group, making up 35.8% of the total.
2. Category 0 (Blue):
The second-largest portion, accounting for 34.2%.
3. Category 3 (Green):
Covers 9.7% of the population.
4. Category 1 (Purple):
Comprises 9.6% of the total.
5. Category 5 (Orange):
Makes up 6.6% of the data.
6. Category 2 (Yellow):
The smallest group, contributing 4.2% to the whole.

11.AGE VS HBA1C DISTRIBUTION



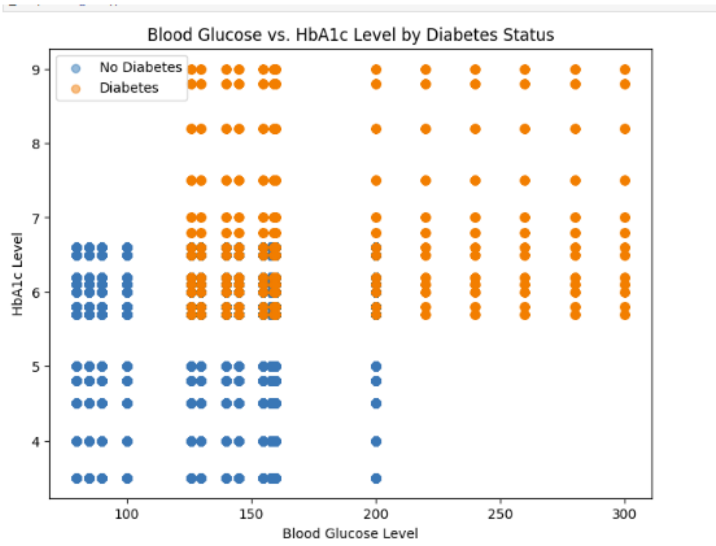
Range of HbA1c Levels: The graph primarily showcases HbA1c levels ranging from 4 to 7, a typical span for this diabetes-related biomarker. Interestingly, there are occasional data points that exceed this usual range, suggesting instances of higher-than-average blood sugar control issues.

Central Age Range: The ages plotted tend to cluster mainly between 40 and 60 years, indicating that the middle-aged population is predominantly represented in this data. This age range often marks a critical period for the onset and management of type 2 diabetes, making these findings particularly relevant for health professionals focusing on this demographic.

Pattern Across Ages: The distribution of HbA1c levels across these ages shows a relatively flat trend within the 40-50 year bracket. This flatness suggests that within this decade, there isn't a significant variation in average HbA1c levels, pointing to a stable phase of glucose management among individuals in this group.

Notable Trend and Variability: A notable spike in HbA1c levels is observed at around the 7 mark. This spike might indicate a threshold beyond which HbA1c levels begin to stabilize or even improve, possibly due to heightened medical intervention or lifestyle adjustments that often follow higher diagnostic readings.

12.BLOOD GLUCOSE VS HBA1C LEVEL BY DIABETES STATUS



Blood Glucose and HbA1c Levels Overview:

Blood Glucose Level: This key metric, displayed along the x-axis, ranges from approximately 50 to 300. It provides a snapshot of an individual's current sugar levels, which can vary significantly throughout the day and with dietary intake.

HbA1c Level: Shown on the y-axis and ranging from 4 to 9, this value reflects the average blood glucose concentration over the past two to three months, offering a longer-term perspective on glucose control.
Visual Color Coding:

Blue Dots: Represent individuals without diabetes, visually coding them in a soothing blue suggests a state of normalcy in terms of blood sugar management.

Orange Dots: Denote individuals diagnosed with diabetes, with the more alerting orange color indicating a need for careful health management.
Key Observations:

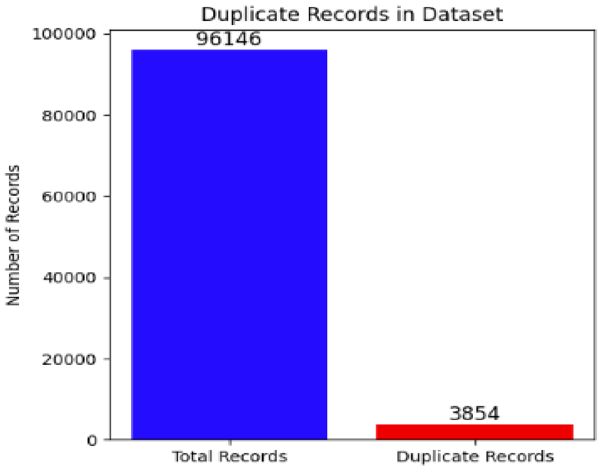
Individuals without Diabetes: These are primarily found in a lower range, with blood glucose levels around 100-150 and HbA1c levels between 4 and 6. This clustering indicates good health control and lower diabetes risk.

Individuals with Diabetes: This group tends to show higher blood glucose levels, typically between 150 and 300, and HbA1c levels from 6 to 9, reflecting ongoing challenges in glucose management.

Overlap Zone: There's a noticeable overlap in the HbA1c range of 6-7, where both groups might meet. However, it's clear that higher blood glucose levels are more commonly associated with diagnosed diabetes, illustrating a gradient of increasing health concern.
Trends and Implications:

The distinct clustering and separation by color in the plot underscore a strong correlation where higher HbA1c and blood glucose levels are indicative of diabetes. Conversely, individuals without diabetes typically feature in the plot's lower regions, suggesting effective glucose control.

13.ANALYSIS OF DUPLICATE RECORDS IN DATASET



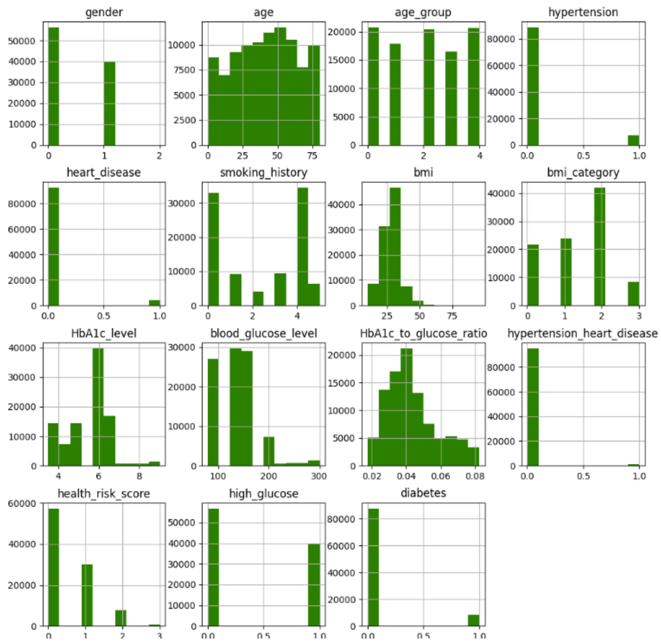
This bar chart explains that the dataset contains a significant number of duplicate records, accounting for approximately 3.85% of the total data points (3,854 out of 100,000). Removing these duplicates can substantially enhance data processing efficiency and optimize model performance.

Dropping duplictes can:
1. Reduce data noise and improve data quality
2. Decrease computational resources required for processing
3. Enhance model accuracy and reliability
4. Improve overall data analysis and insights

14.HISTOGRAM PLOTS

Age Distribution: The age graph illustrates a well-distributed adult population, predominantly spanning from 30 to 70 years old. This shows a mature demographic, with fewer young adults and seniors, suggesting that the data might be skewed towards a working-age population.

Hypertension Insights: The overwhelming majority of the dataset, nearly 90,000 individuals, do not suffer from hypertension. This indicates good cardiovascular health across the majority, with only a small segment of the population affected by this condition, reflecting effective management or lower incidence of this risk factor in this particular group.



Heart Disease Observations: Similarly, heart disease appears to be relatively rare within this group, with most individuals showing no signs of the condition. This is a positive indicator of heart health and may reflect either a healthy lifestyle or successful medical interventions among these individuals.

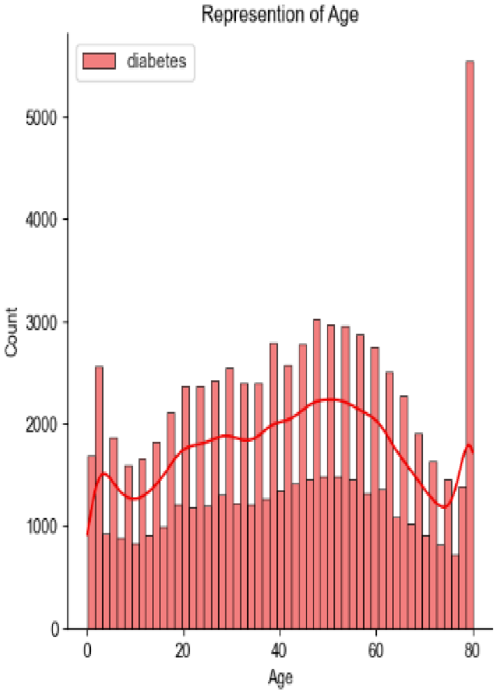
BMI Trends: The BMI data clusters mostly between 20 and 40, suggesting a range from normal to moderately overweight statuses among most individuals. There are very few outliers, indicating that extreme cases of underweight or obesity are not common in this population, pointing to good overall nutritional and health status.

HbA1c Levels: The concentration of HbA1c levels between 5 and 7 suggests generally good glucose control, with notable clusters around the 5 and 6 marks. These levels are typical for people without diabetes or those managing their condition effectively.

Blood Glucose Patterns: The distribution shows that most individuals maintain blood glucose levels within a normal to slightly elevated range (100 to 200 mg/dL), with rare instances of very high levels above 250 mg/dL. This pattern indicates that while there is some variation in glucose control, extreme cases are uncommon.

Diabetes Prevalence: A significant majority of the surveyed population does not have diabetes, with only a minor portion being diabetic. This lower prevalence is encouraging and suggests that diabetes is not a dominant health issue among these individuals, although it remains an important area for ongoing public health focus.

15. REPRESENTATION OF AGE



The bar and line graph vividly illustrate the age distribution within the dataset, showing a prominent concentration of individuals between the ages of 20 and 70, with a peak in the 50 to 60 age range. This suggests a community heavily populated by middle-aged adults, potentially reflective of an active workforce and familial responsibilities.

REFERENCES

[1]https://www.ibm.com/topics/exploratory-data-analysis

[2]https://scikit-learn.org/stable/modules/preprocessing.html

[3]https://www.researchgate.net/publication/308007227_Exploratory_Data_Analysis

[4]https://www.researchgate.net/publication/343678066_Diabetes_mellitus_type_2_Exploratory_data_analysis_based_on_clinical_reading

Peer Evaluation Form for Final Group Work

CSE 487/587B

- Please write the names of your group members.

Group member 1 : **Swarupa Murala**

Group member 2: **Tejaswini Chowdam Nallagondappa**

Group member 3 : **Arunkarthik Periyaswamy**

- Rate each groupmate on a scale of 5 on the following points, with 5 being HIGHEST and 1 being LOWEST.

Evaluation Criteria	Group member 1	Group member 2	Group member 3
How effectively did your group mate work with you?	5	5	5
Contribution in writing the report	5	5	5
Demonstrates a cooperative and supportive attitude.	5	5	5
Contributes significantly to the success of the project.	5	5	5
TOTAL	20	20	20

Also please state the overall contribution of your teammate in percentage below, with a total of all three members accounting for 100% (33.33+33.33+33.33 ~ 100%) :

Group member 1: 33.33

group member 2: 33.33

group member 3: 33.33