Diabetes Risk Analysis and Visualization

Ayush Patel – 202318036

Mitul Dudhat – 202318024

Siddharth Kadam – 202318015

Sreejesh S. Nair – 202318001

Project Goal

- **Goal:**
- To analyze health and demographic data to identify key risk factors associated with diabetes.
- Use data visualization to highlight high-risk groups, supporting targeted diabetes prevention strategies.
- **Objectives:**
- Identify critical factors (age, BMI, gender, chronic conditions) contributing to diabetes risk.
- Examine combined effects of multiple risk factors on diabetes likelihood.
- Illustrate how different risk factors affect diabetes risk across demographics.

Dataset Description

- Tabular format with 100,000 records.
- Each record includes health and demographic information impacting diabetes risk.

Sr. No.	Attribute Name	Description	Data Type
1	Age	Age of the individual, ranging from 0 to 80 years	Integer
2	Gender	Gender of the individual (Male/Female)	Categorical
3	BMI	Body Mass Index of the individual, ranging from 10 to 96	Float
4	Smoking History	History of smoking with 6 distinct levels	Categorical
5	Hypertension	Indicates presence of hypertension (Yes/No)	Categorical
6	Heart Disease	Indicates presence of heart disease (Yes/No)	Categorical
7	Blood Glucose	Blood glucose level, ranging from 80 to 300	Integer
8	diabetes	Indicates presence of Diabetes (Yes/No)	Categorical
9	HbA1c	Hemoglobin A1c level, ranging from 3.0 to 10.0	Float
		Derived Variables	
10	Chronic Risk Score	Derived cumulative score combining smoking, hypertension, and heart disease risk factors (0 to 1)	Float
11	Smoking Category	Derived binary category indicating smoking history presence (Yes/No)	Categorical
12	Age Group	Derived age categorization (Infant, Children, Adolescents, Adult, Old)	Categorical
13	BMI Categories	Derived BMI categorization (Under 25, 25-30, Over 30)	Categorical
14	Weight Category	Derived weight classification (Underweight, Normal, Overweight, Obese, Severely Obese)	Categorical

Mustafa, M. (n.d.). *Diabetes Health Indicators Dataset* [Data set]. Kaggle. https://www.kaggle.com/datasets/iammustafa

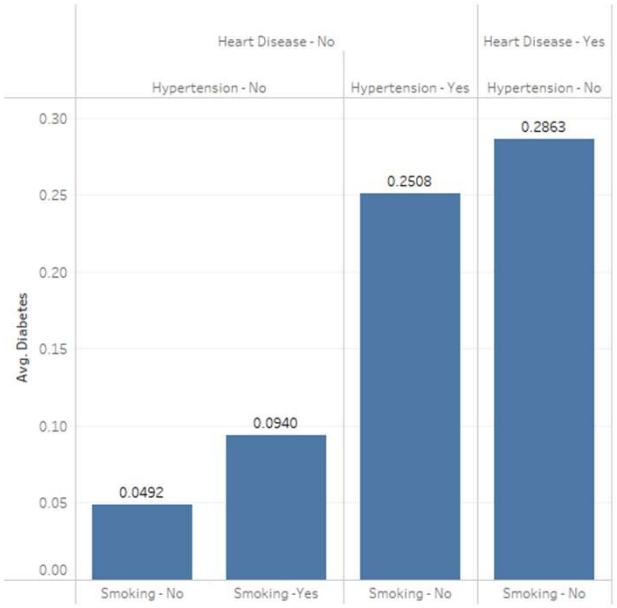
Problems

- **Key Questions to Explore:**

1. **How do cardiovascular health and lifestyle factors affect diabetes risk?**
   - *Hypothesis (H1):* Individuals with a higher "Chronic Risk Score"—a combination of hypertension, heart disease, and smoking—are more likely to develop diabetes, with heart disease having the strongest individual impact.
2. **Does BMI significantly impact diabetes risk across different demographics?**
   - *Hypothesis (H2):* Individuals with a BMI over 30 are at a higher risk of diabetes compared to those with a BMI under 25, regardless of age and gender.
3. **Does gender play a role in influencing diabetes-related biomarkers?**
   - *Hypothesis (H3):* Gender and obesity levels significantly affect blood glucose and HbA1c levels, with males and those classified as severely obese showing higher levels. Furthermore, HbA1c may serve as a more reliable predictor of diabetes risk compared to blood glucose alone.

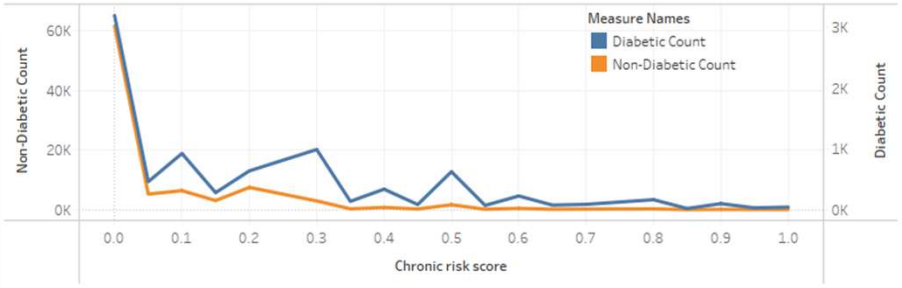
Visualizing the Relationship Between Risk Factors and Diabetes Incidence

Impact of Smoking and Hypertension on Diabetes Prevalence



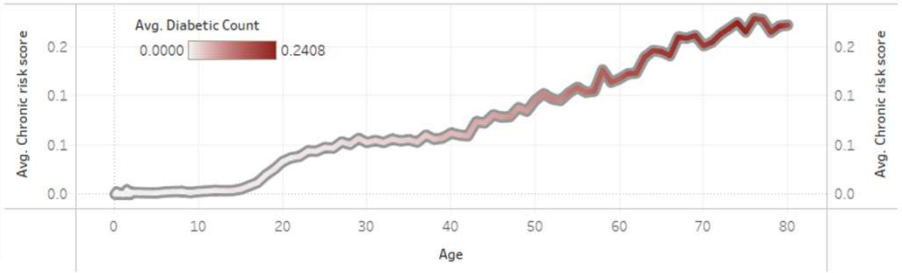
Compares the average diabetes rates across groups with no risk factors, smoking only, hypertension only, and heart disease only, to show which factor alone has the most impact on diabetes.

Relationship Between Chronic Risk Score and Diabetes



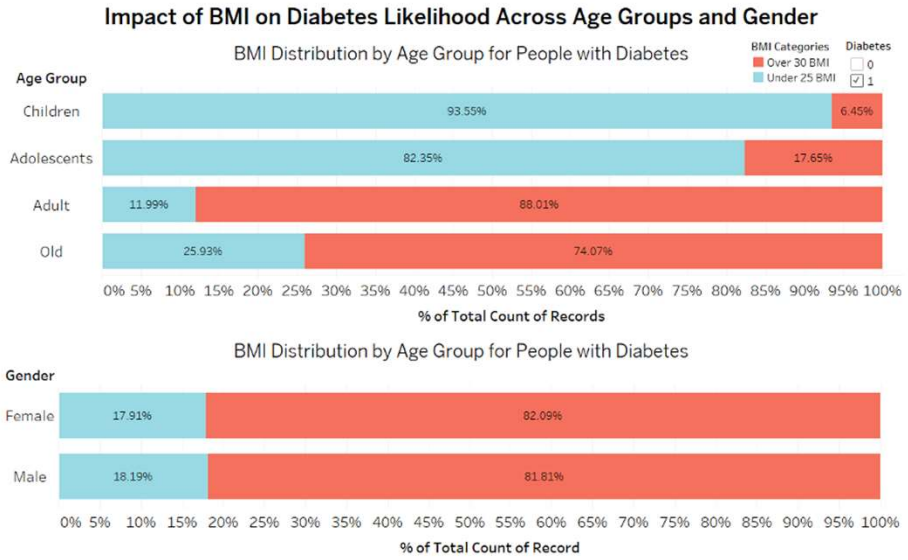
Compares diabetic and non-diabetic counts across different Chronic Risk Scores.

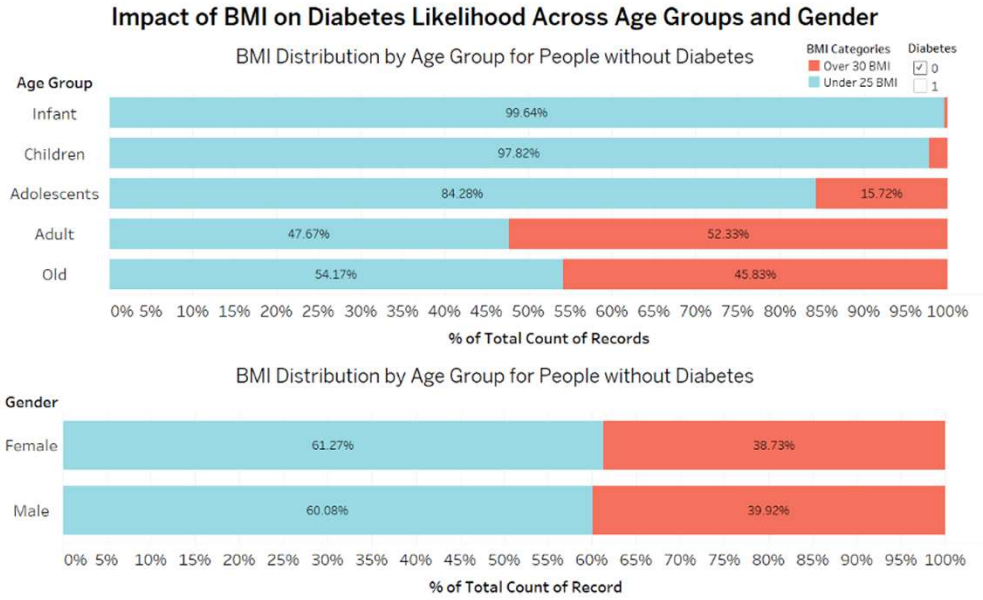
Average Chronic Risk Score Across Age Groups



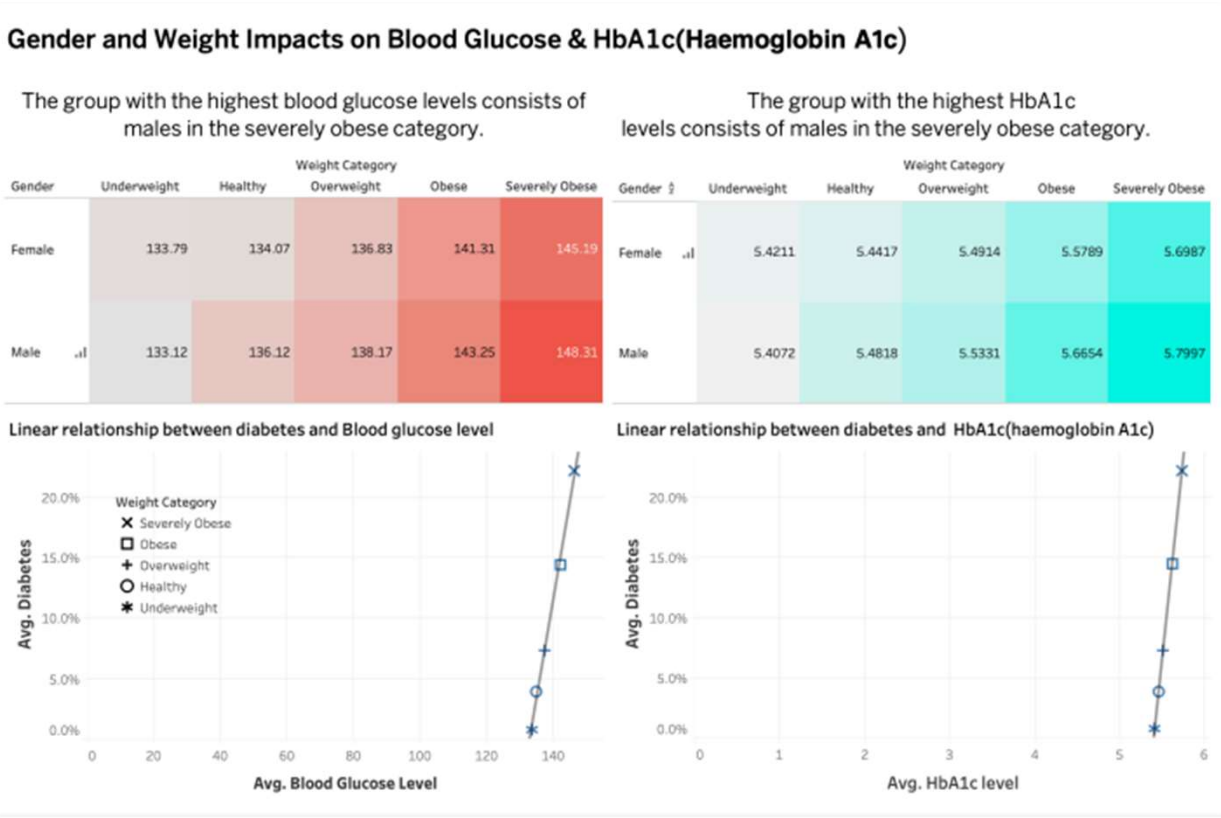
Shows how the Chronic Risk Score varies by age, with a color gradient indicating diabetes prevalence.

- **Task** : Analyse how individual, combined, and age-related risk factors impact diabetes prevalence.
- **Findings:**
  1. Heart disease and hypertension are the strongest predictors of diabetes, with prevalence highest among those with heart disease.
  2. Higher chronic risk scores correlate with greater diabetes incidence, confirming cumulative risk factor impact.
  3. Chronic risk scores and diabetes prevalence rise with age, particularly after 60, indicating age-related risk accumulation.





- **Task:** Analyze the impact of BMI on diabetes likelihood across age groups and gender.
- **Findings:**
  1. Adults and older individuals with diabetes predominantly have a BMI over 30, regardless of gender.
  2. Non-diabetic individuals, especially in younger age groups, are more likely to have a BMI under 25.



- **Task:** Analyze the impact of gender and weight on blood glucose and HbA1c levels.
- **Findings:**
  1. Severely obese males have the highest blood glucose levels, with males generally showing higher levels across all weight categories.
  2. Severely obese males have the highest HbA1c levels, and males consistently show higher levels than females in all weight categories.
  3. Higher blood glucose levels correlate with increased diabetes prevalence, particularly in severely obese individuals.
  4. HbA1c has a stronger correlation with diabetes prevalence than blood glucose, with a significantly lower p-value.

Results and Conclusion

- **Results:**
  - Higher chronic risk scores, particularly due to heart disease, are positively correlated with increased diabetes incidence. Heart disease has the strongest impact on diabetes risk.
  - Individuals with a BMI over 30, especially adults and older individuals, are more likely to have diabetes, with this trend consistent across genders.
  - Males and severely obese individuals show higher blood glucose and HbA1c levels. HbA1c is a slightly stronger predictor of diabetes risk than blood glucose.
- **Conclusion:**

  The Diabetes Risk Analysis Dashboard project provides critical insights into diabetes risk factors, supporting healthcare professionals in risk assessment. This report outlines the approach taken, highlights findings from the data visualizations, and offers actionable recommendations based on the analysis. By examining the influence of individual and combined health conditions—such as heart disease, hypertension, smoking, and BMI—the dashboard underscores significant predictors of diabetes, with HbA1c emerging as a particularly strong indicator.

Thank You