Interactive Data Dashboard for Diabetes Risk Profiling

Author: Pradnesh kapte, Prathmesh Yadav
Tools Used: Power BI, DAX, Power Query, Data Modeling
Dataset Size: 10,000 rows, 21 columns
Data Type: Semi-synthetic, pre-engineered and manually refined for class balance and interpretability

Objective

This project aimed to build a multi-layered, interactive Power BI dashboard that goes beyond basic visualization to uncover hidden risk factors driving the prevalence of diabetes. By integrating demographic, clinical, and lifestyle data, the goal was to:

- Identify critical intersections of risk based on age, BMI, lifestyle habits, and ethnicity
- Create a balanced, ML-ready dataset for unbiased predictive modeling
- Provide actionable insights for healthcare professionals, public health policymakers, and healthtech developers

The final output is a **visually powerful**, **data-scientifically robust**, and **domain-relevant dashboard** that translates raw data into clear, compelling insights.

Dataset Overview

The dataset comprises 10,000 rows and 21 columns, covering a wide range of diabetes-related variables:

Demographic Data:

- Age (numeric, binned into AgeBands)
- Sex (Male, Female)
- Ethnicity (Asian, Black, Hispanic, White)

Clinical Indicators:

- BMI (Body Mass Index)
- Fasting Glucose (mg/dL)
- HbA1c (%) – marker of chronic blood sugar exposure
- Blood Pressure (Systolic & Diastolic)
- GGT, Serum Urate – metabolic enzymes

Lifestyle Features:

- Physical Activity Level (Low, Moderate, High)
- Alcohol Consumption (None, Moderate, Heavy)
- Smoking Status (Never, Former, Current)

Medical History:

- Family History of Diabetes
- Previous Gestational Diabetes

Engineered Fields:

- Glucose_Category (Normal, Prediabetes, Diabetes)
- Is_Diabetic (binary logic based on Fasting Glucose ≥ 126 or HbA1c ≥ 6.5%)

Note: The dataset was **intentionally balanced** across the three diagnosis classes to ensure fairness, eliminate prediction bias, and enable accurate supervised learning.

Dashboard Design & Navigation

The report is structured into 4 functional pages, each offering unique insights and user-driven exploration capabilities.

Page 1: Population Overview

Purpose: Offer a demographic and diagnostic-level snapshot.

Visuals & Metrics:

- KPI cards: Avg BMI (29.18), Avg Glucose (132.48 mg/dL), Diabetic Prevalence (~85%)
- Donut Chart: Sex distribution
- Bar Chart: Diabetic cases by Age Group (18–29, 30–44, 45–59, 60+)
- Clustered Bar: Diabetic counts by Ethnicity

Insights:

- Diabetic burden is highest in 30–44 age group, followed closely by 45–59.
- Gender-based distribution is nearly even, suggesting universal risk.
- Hispanic and Black populations show slightly elevated diabetic counts

Page 2: Risk Factor Explorer

Purpose: Uncover deeper medical and clinical drivers behind diabetes.

Key Visuals:

- Bubble Chart: BMI vs Glucose (bubble size = HbA1c)
- Clustered Column Charts: Avg HbA1c by AgeBand, BMI Category, and Ethnicity
- Matrix Heatmap: Diabetic % across AgeBand × BMI Category
- Pie Chart: Glucose_Category distribution

Insights:

- HbA1c reveals high chronic risk even in Normal/Overweight BMI individuals.
- Hispanic and Black groups show consistently higher HbA1c regardless of BMI.
- Bubble chart proves glucose increases with BMI, but HbA1c differentiates hidden risk.
- Glucose_Category pie chart validates dataset is balanced across diagnostic stages.

Page 3: Behavioral & Demographic Impact

Purpose: Examine how lifestyle and age interact with ethnicity to influence risk.

Visuals:

- Matrix: Diabetic % by Physical Activity × Alcohol Consumption
- Line Chart: Avg Glucose by AgeBand & Ethnicity
- Bar Chart: Max Avg of BMI, Glucose, HbA1c by Ethnicity
- Bar Chart: Count of Diabetics by AgeBand

Insights:

- Low activity & heavy alcohol = 91% diabetic rate – the riskiest combination.
- AgeBand 45–59 shows glucose peaks across most ethnicities.
- HbA1c is elevated across all ethnicities – chronic sugar exposure is widespread.
- Diabetic volume is densest between 30–59 – prime zone for preventive care.

Page 4: Dataset Fairness & Model Integrity

Purpose: Justify model readiness and fairness through balanced data design.

Visuals:

- Pie chart: Equal distribution of diagnostic classes
- Matrix: Diabetic % by AgeBand × BMI Category

Insights:

- Obese 30–44 = 89% diabetic rate – highest risk zone
- Underweight seniors (60+) show 33% prevalence – busting the weight-only myth
- Even Normal BMI individuals show ~85% prevalence in some age bands

Why It Matters: Ensures unbiased predictive modeling, ideal for AI/ML healthcare solutions.

Key Takeaways

- **Diabetes isn't always visible**: Normal BMI can still mean high HbA1c.
- **Behavioral risk is critical**: Activity + Alcohol intake amplify metabolic risk.
- **Ethnic patterns matter**: Hispanic and Black groups require customized outreach.
- **Middle age is the tipping point**: Ages 30–59 carry most volume & risk.

Preventive Recommendations

- **Universal screening from age 30+,** regardless of BMI
- **Behavior-focused intervention campaigns**, especially targeting inactive drinkers
- **Culturally sensitive health communication** for ethnic subgroups
- **Fairness-first model development** using balanced datasets for training

Why This Dashboard Matters

- **Insightful**: Merges behavioral, biological, and demographic data into one view
- **Interactive**: Enables multidimensional filtering and real-time drilldown
- **Balanced**: Engineered for fair modeling, ideal for AI and EHR integration
- **Scalable**: Can be expanded with real-time data feeds or mobile health inputs
- **Industry-Relevant**: Aligns with WHO, CDC, and digital health analytics frameworks

Conclusion

This Power BI dashboard isn't just a visualization project — it's a full-fledged diagnostic lens. It captures how diabetes hides in behavior, appears differently across ethnic lines, and spikes at life stages when people often feel the healthiest. It's predictive, fair, and ready to be plugged into real-world decision-making.