

Business Problem Summary

I am exploring the factors contributing to elevated diabetes risk among adults in Texas. My business problem is:

“How can Texas public health agencies use self-reported health behavior and demographic data to predict which adults are at 30% or higher risk of developing diabetes within five years, using clinical markers like HbA1c and BMI?”

This topic aligns with my career goal of becoming a data analyst in the healthcare sector, where the early identification of chronic disease risk and targeted prevention strategies are crucial. I want to create value for public health organizations like the Texas Department of State Health Services by delivering predictive insights that support early interventions, reduce healthcare costs, and promote health equity in vulnerable communities.

Primary Persona

Name: Dr. Angela Torres

Role: Director of Chronic Disease Prevention, Texas Department of State Health Services

Responsibilities:

Dr. Torres designs and manages statewide programs aimed at reducing the burden of chronic conditions like diabetes. She allocates resources for screenings, outreach, and education, relying on data to guide decision-making.

Why This Project Matters to Her:

As a public health leader, Dr. Torres must identify at-risk populations and allocate limited resources effectively. With over 2.8 million diagnosed diabetics in Texas and \$25 billion in annual costs, she needs tools to:

- Prioritize high-risk populations for screenings
- Set evidence-based thresholds for intervention
- Coordinate targeted prevention campaigns across rural and underserved areas

This project offers a predictive model she can use to focus efforts where they're most needed maximizing the impact of public health initiatives while addressing disparities in care access.

Potential Datasets

To address this problem, I searched for datasets with:

- Demographics: age, sex, income, education, race
- Health behaviors: smoking, alcohol use, exercise, diet
- Clinical data: BMI, blood pressure, cholesterol, general health status
- Outcome: diabetes classification

Sources reviewed:

- Kaggle: Behavioral Risk Factor Surveillance System (BRFSS)
- UCI Machine Learning Repository: CDC Diabetes Health Indicators
- Kaggle: Diabetes Health Indicators Dataset

Although most datasets lack longitudinal depth, they include rich behavioral and demographic variables that can inform early-stage risk prediction models.

Use of AI Tools

I used You.com to refine my business problem statement. I asked for help making the goal more measurable and actionable. The tool recommended:

- Adding a specific timeline (by 2026)
- Defining “high risk” using thresholds like BMI >30 or HbA1c >5.7%
- Including potential outcomes such as improved health equity and cost reduction

These suggestions helped sharpen my objective, align it with SMART criteria, and frame it in language relevant to public health leaders.

What I’d Like Feedback On

I’d appreciate feedback on whether my target persona (a public health director) is specific and high-level enough, or whether I should focus more narrowly—perhaps on a community clinic manager or epidemiologist. I’m also wondering if I should limit the project scope to specific counties or demographics (e.g., rural Hispanic adults) to make the model more actionable. Any suggestions on accessing datasets with time-series or longitudinal behavior data would also be helpful for refining the predictive model.