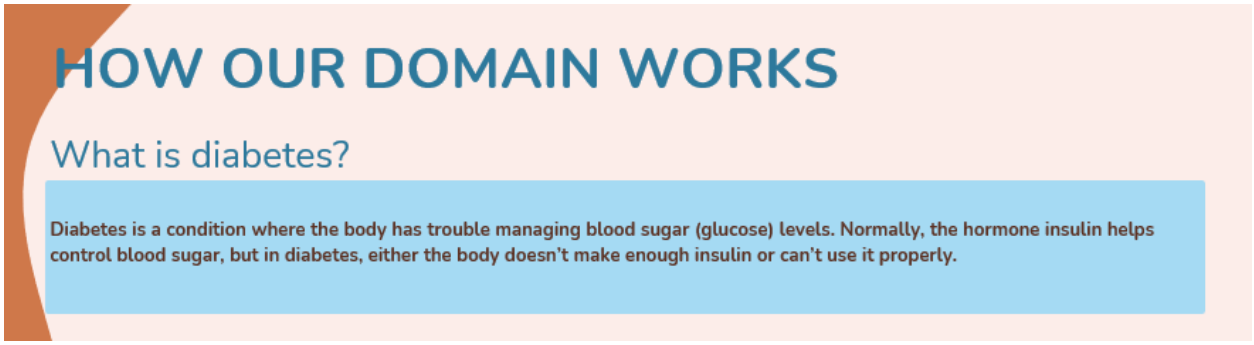
DIABETES PREDICTION for WOMEN ONLY

1. **How does your industry/domain work? (don't make it very generic, try to be specific, quote research/surveys etc., use flow diagrams)**

Diabetes background:



Type 1 Diabetes
Type 1 diabetes is a chronic condition where the pancreas produces little or no insulin due to an autoimmune response that destroys insulin-producing cells. This lack of insulin prevents glucose from entering cells, leading to high blood sugar levels. It typically appears in childhood or adolescence and requires lifelong insulin therapy.

(Insulin Dependent Individuals, pancreas is not excreting insulins anymore)

Type 2 Diabetes

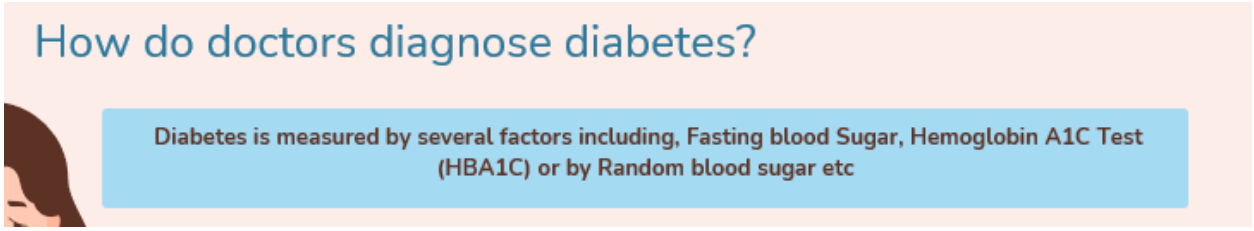
Type 2 diabetes is a chronic condition where the body becomes resistant to insulin, or the pancreas doesn't produce enough insulin to maintain normal blood glucose levels. This leads to high blood sugar levels over time. It is often associated with lifestyle factors such as obesity and inactivity and typically develops in adulthood, though it can also occur in younger individuals. Management includes lifestyle changes, oral medications, and sometimes insulin therapy.

(Individuals who has little insulin being produced by the pancreas, not totally insulin dependent

Gestational Diabetes

Gestational diabetes is a type of diabetes that develops during pregnancy when the body cannot produce enough insulin to meet the increased demands, leading to high blood sugar levels. It typically resolves after childbirth but can increase the risk of developing type 2 diabetes later in life for both the mother and the child. Management includes monitoring blood sugar levels, dietary adjustments, physical activity, and, in some cases, insulin therapy.

(Women who develops diabetes during pregnancy, who may either come out diabetic after pregnancy or not) extra facts: Women who are pregnant and are diabetic have bigger babies than the usual, Pregnant women with diabetes often have higher blood sugar levels, which can pass through the placenta to the baby. This extra glucose causes the baby's pancreas to produce more insulin, leading to increased growth and fat storage, resulting in larger-than-average babies (a condition called macrosomia).



Hospital or clinic setting

As for diagnosis or prediction of diabetes, in a hospital or clinical setting, diabetes is diagnosed through several key tests that assess blood glucose levels and the body's ability to manage sugar over time. When visiting a hospital for a diabetes diagnosis, typically, a doctor will discuss symptoms, conduct a physical examination, and recommend one or more of these tests based on your risk factors and medical history. Here are the main methods:

1. Fasting Blood Sugar Test (FBS) - it should be less than 120 its no eating minimum of 8 hours sometimes up to 12 hours depending on doctors request
2. Hemoglobin A1C Test -its your average sugar for 3 months should be 5.6 and lower for normal, 5.7 upto 6.4 for pre diabetic. And 6.5 higher for diabetic
3. Random Blood Sugar Test - Sugar tesing without fasting. Can be done anytime of the day. 140mg/dl lower for normal. 140-199 prediabetic, diabetic 200 or higher

For self-assessment, at-home tools help people monitor their blood sugar levels and can play a role in diabetes prediction, especially for those at risk. Common self-assessment tools include:

1. Capillary blood glucose (CBG) meters **(this is called a GLUCOMETER) where you will manually prick your finger to get results**

2. **Continuous Glucose Monitors (CGM)** this are real-time continuous glucose monitors (rtCGM). It gets attached to your arms so you could measure anytime

(Only for extra discussion this is the flow chart on how it is done in the hospital) A flowchart that outlines the process to diagnose diabetes is as follows:
Begin Diagnosis:

- If the individual shows symptoms (e.g., increased thirst, frequent urination, unexplained weight loss) or has risk factors (e.g., family history, obesity), proceed to testing.

Select Initial Test:

- Option A: Fasting Blood Sugar Test (FBS)
- Option B: Hemoglobin A1C Test
- Option C: Oral Glucose Tolerance Test (OGTT)
- Option D: Random Blood Sugar Test (for those with symptoms)

Evaluate Test Results:

- FBS:
    - Normal: <100 mg/dL
    - Prediabetes: 100–125 mg/dL
    - Diabetes: ≥126 mg/dL (requires repeat test to confirm)

- A1C:
    - Normal: <5.7%
    - Prediabetes: 5.7%–6.4%
    - Diabetes: ≥6.5% (repeat test needed to confirm)
- OGTT (2 hours after consuming glucose):
    - Normal: <140 mg/dL
    - Prediabetes: 140–199 mg/dL
    - Diabetes: ≥200 mg/dL (confirm with repeat test)
- Random Blood Sugar (for symptomatic individuals):
    - Normal: <200 mg/dL
    - Diabetes: ≥200 mg/dL (if symptoms are present, may not need repeat test)

Confirm Diagnosis:

- If the initial test indicates diabetes (i.e., ≥126 mg/dL for FBS, ≥6.5% for A1C, or ≥200 mg/dL for OGTT), repeat the test on a different day for confirmation.
- For symptomatic individuals with a high random blood sugar test (≥200 mg/dL), additional testing may not be required for diagnosis.

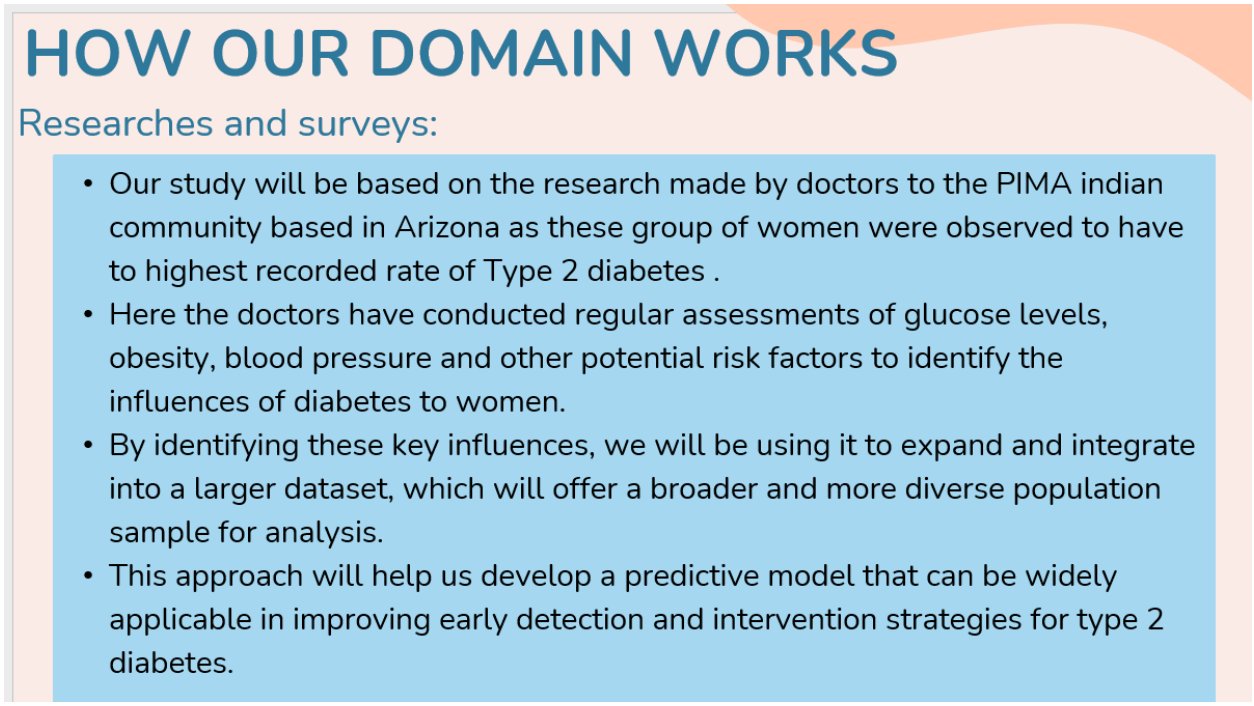
Diagnosis:

- Confirmed Diabetes: Diagnosis confirmed if results on two separate tests are within the diabetic range.
- Prediabetes: If levels indicate prediabetes, suggest lifestyle changes and retest periodically.
- Normal: No diabetes; recommend regular screening if risk factors persist.

Follow-Up:

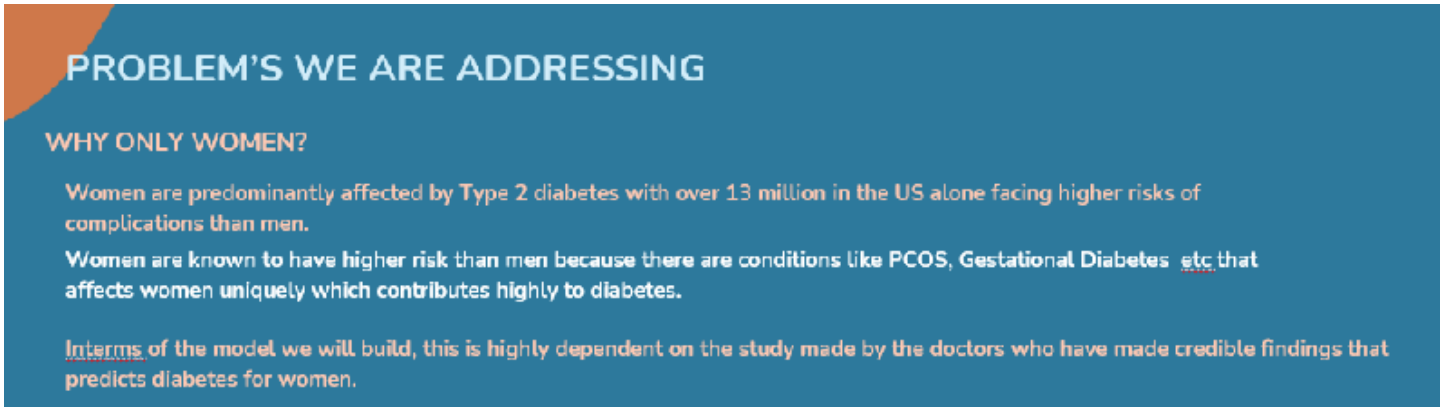
- Provide guidance for diabetes management or preventive measures, including lifestyle adjustments, medication, and monitoring as necessary.

*----------



Problem we are Solving/ Addressing:

High rates of undiagnosed or poorly managed diabetes among women pose a significant public health risk, leading to severe complications, increased healthcare costs, and reduced quality of life**.** Women are uniquely affected by diabetes due to specific risk factors like gestational diabetes and a higher likelihood of developing complications, particularly cardiovascular disease. These risks are exacerbated by delayed diagnosis and inadequate preventive measures, which could otherwise reduce the incidence and severity of diabetes-related complications.



Why are we interested in studying only women for the diabetes prediction?

- Predicting diabetes in women is crucial because they face unique health risks and challenges. ***Women with diabetes are more likely to experience complications*** *like cardiovascular disease, which they are already at a higher risk for than men.*
- **Pregnancy also introduces risks of gestational diabetes, *which can lead to long-term health consequences for both mother and child*,** including an increased chance of developing type 2 diabetes later in life. Early detection allows for interventions that improve health outcomes, reduce complications, and enhance quality of life, making predictive models a valuable tool for preventive healthcare in women.
- Also, if we only study women's diabetes: **Focusing on women in diabetes prediction models can improve both our model accuracy and relevance.** By narrowing to a single demographic, **we reduce data variability tied to gender-specific differences, which allows the model to better capture health patterns unique to women.** Women experience different hormonal cycles, body composition, all of which affect diabetes risk. For instance, factors like pregnancy-related insulin resistance (gestational diabetes), influence diabetes patterns uniquely in women.
- **Studying women exclusively allows the model to learn these specific patterns without interference from unrelated male data, leading to a more refined prediction tailored to women's health needs.** Ultimately, this approach supports targeted healthcare interventions and personalized prevention strategies, which are especially effective in improving outcomes in diabetes care for women.

Our Idea on how we could address the problem:

We will build a model to help predict the chances of a woman being diabetic or not.

Quote research: This research is backed up by doctors who have studied specific group of women who had high diabetes rate in Arizona (PIMA Indians from Arizona, because they were identified to have had the highest type two diabetes for women).

How will it Assist?

- Once the prediction has been done, it will bring awareness if the person is likely diabetic or not
- To assist pregnant women who are at risk of developing Gestational diabetes.
- To assist women who doesn't want to go to hospitals or get theirself checked before having a credible knowledge that they maybe diabetic.
- This will help women create pre emptive measures before available doctor's check up and may ultimately assist in preventing complications
- On the other hand, this will also assist fitness coaches, to tailor fit excercise to women who are at risk developing it
- Insurance makers may also use this to understand if the person they will cover is likely diabetic already or chances of getting one so they could adjust their policies as needed.

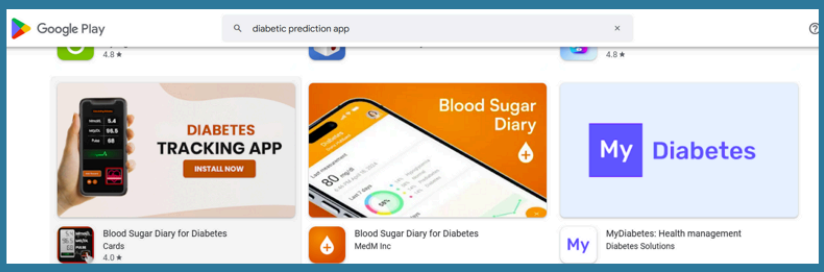
OUR TARGET CONSUMERS

- Women who are wanting to know their risks into becoming diabetic
- Women who are at risk due to a strong family history of diabetes
- Fitness coaches who wanted to offer personalized health plans for women who are at risk of developing diabetes
- Health Insurance companies looking to assess risk factors for diabetes in their policy holders

HOW IS OUR SOLUTION DIFFERENT?

What other solutions are out there?

We have not seen an app that predicts diabetes for women



Our data is diverse Broad and Current

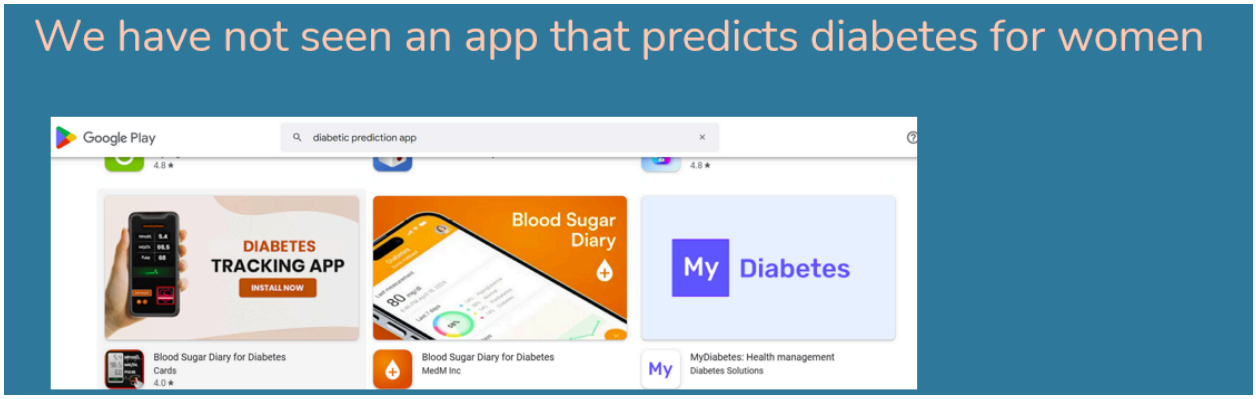
(more information on next slide)

Why do we not have any competitors? Why is this app not available in google play or why hasnt anyone predicted diabetes before:

Building a reliable diabetes prediction app is difficult because of **strict privacy rules**, the **need for high accuracy to avoid risks**, and **complex regulations** that treat such apps as medical tools needing approval. Models also need diverse data to prevent biased results, making development more challenging.

- We will not be impacted by strict privacy rules because we are not going to collect data.
- High accuracy model is what we will aim to have.
- We will not be under any complex regulations because we will only mention highly likely and percentage and unlikely diabetic with percentage.
- Mis diagnosis is highly unlikely with a proven model that actually predicts diabetes because our research is backed up by doctors who have done this study for a long period of time.

Why us and not our competitors: **We dont have a competitor.** Our competitors have focused on health management application. Our focus is wanting to predict if that person is diabetic or not. Or not yet, or people who are worried.



The apps in google play are not predicting if that woman is diabetic or not, **the apps there are just about maintaining diabetes and their food intake with their glucose levels which they could share to the doctors.**

HOW IS OUR OUR PRODUCT DIFFERENT?

PREDICTIVE FOCUS ON KEY RISK VARIABLES VS. STANDARD MONITORING

What is the difference? With the traditional model:

- Current Self-Kit Tools: Traditional self-kits, such as glucose meters, only measure blood glucose levels and offer limited insights into overall diabetes risk.
- Hospital Diagnostics: Hospital tests, like fasting glucose or A1C tests, diagnose diabetes but don't typically factor in other critical variables like BMI, age, or family history until symptoms appear or blood sugar levels are elevated.

With our Model

- Our model incorporates multiple key risk factors specific to diabetes, including pregnancy status, BMI, glucose, blood pressure, age, and family history of diabetes. By analyzing these variables together, our model can predict diabetes risk with more depth and accuracy than self-kits or single-point hospital tests, enabling preventive actions well before full-blown diabetes develops.

WOMEN-CENTRIC APPROACH WITH FOCUS ON PREGNANCY RISK VS. GENERIC PREDICTION MODELS

With the traditional model:

- Most Applications: Many diabetes prediction models are generalized and do not specifically address factors unique to women, such as pregnancy-related diabetes risks (gestational diabetes) and hormonal influences.

With our model:

- Our Model: By incorporating pregnancy status as a variable, our model provides a unique focus on gestational diabetes and the associated risk of developing type 2 diabetes later in life. This makes it especially relevant and accurate for women, who require a more tailored approach for effective prediction and early intervention.

AFFORDABLE PREVENTIVE TOOL VS. EXPENSIVE DIAGNOSTIC OPTIONS

With the traditional model:

Current Self-Kits and Hospital Diagnostics: Many existing self-kits and hospital-based tests for diabetes diagnosis come at a high cost, which can be

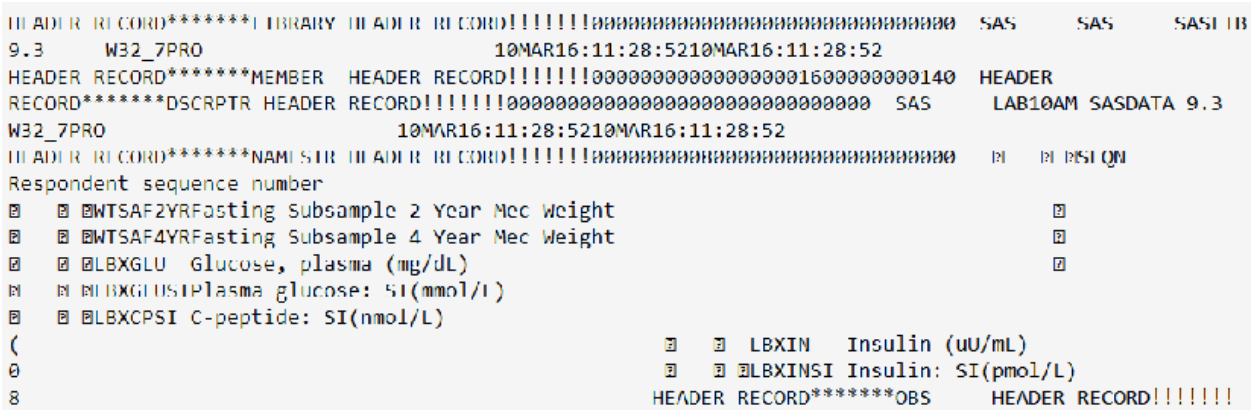
prohibitive for individuals who do not have access to comprehensive health insurance or regular healthcare. Hospital diagnostic tests like fasting glucose, A1C, and glucose tolerance tests may require frequent visits and out-of-pocket expenses, adding a financial burden for those needing routine monitoring.

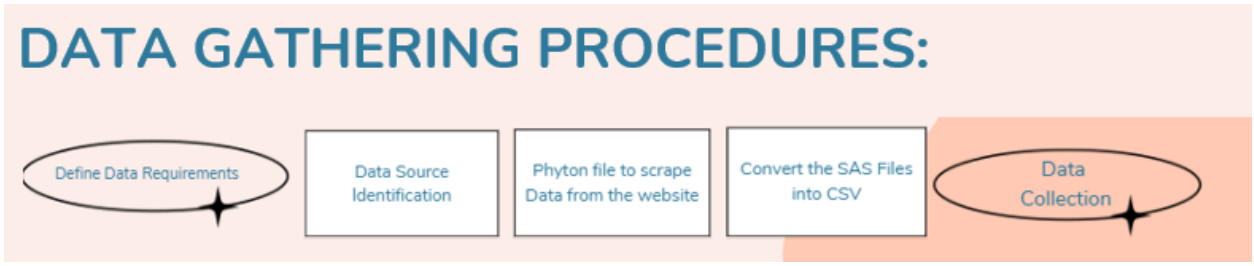
With our model:

Our Model: Design is free but will have ads, this will be an accessible alternative, our model acts as an early-warning system for individuals concerned about their diabetes risk. By identifying high-risk individuals before they require costly clinical testing or ongoing monitoring, it provides an affordable, reliable option for proactive health management. This makes it particularly valuable for individuals in underserved communities or for those who may not have the financial means to access traditional healthcare options.

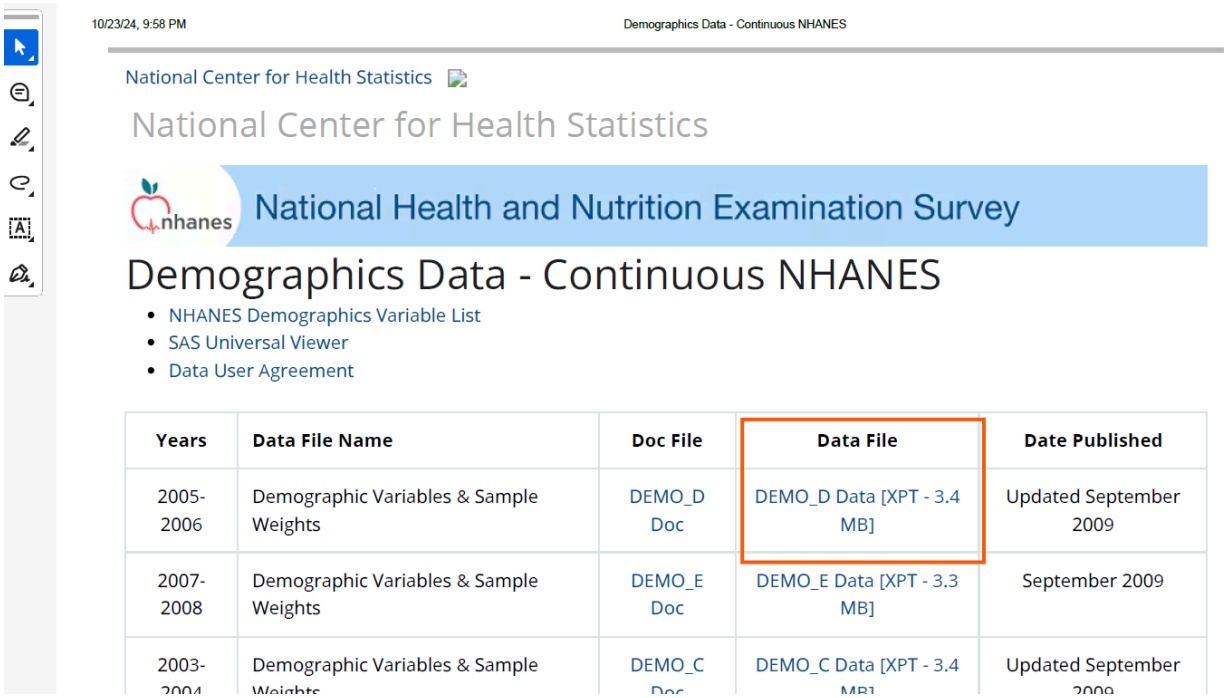
HOW DOES OUR RAW DATA LOOK LIKE?

This is raw data, SAS file





1. For the data requirements, the pages in the website have been converted into pdf, then a python file was created to extract the MB per file, and altogether it has amounted to 38k MB



10/23/24, 9:58 PM Demographics Data - Continuous NHANES

National Center for Health Statistics

National Center for Health Statistics

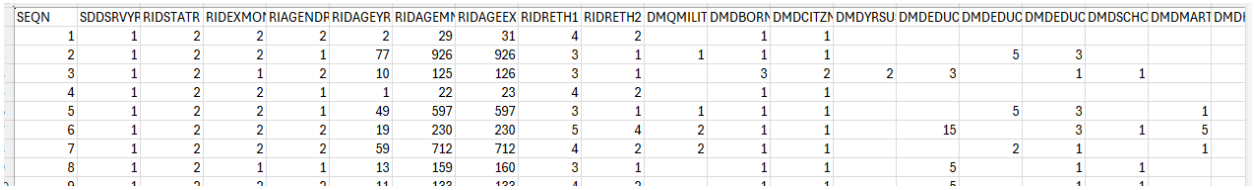
nhanes National Health and Nutrition Examination Survey

Demographics Data - Continuous NHANES

- NHANES Demographics Variable List
- SAS Universal Viewer
- Data User Agreement

Years	Data File Name	Doc File	Data File	Date Published
2005-2006	Demographic Variables & Sample Weights	DEMO_D Doc	DEMO_D Data [XPT - 3.4 MB]	Updated September 2009
2007-2008	Demographic Variables & Sample Weights	DEMO_E Doc	DEMO_E Data [XPT - 3.3 MB]	September 2009
2003-2004	Demographic Variables & Sample Weights	DEMO_C Doc	DEMO_C Data [XPT - 3.4 MB]	Updated September 2009

Then we converted the sas file into CSV, how does the CSV look like? Bunch of numbers where we will do feature engineering



SEQN	SDDSRVYR	RIDSTATR	RIDEXMON	RIAGENDR	RIDAGEYR	RIDAGEMN	RIDAGEEX	RIDRETH1	RIDRETH2	DMQMILIT	DMDBORN	DMDCITZN	DMDYRSU	DMDEDUC	DMDEDUC	DMDEDUC	DMDSCHC	DMDMART	DMDI
1	1	2	2	2	2	29	31	4	2		1	1							
2	1	2	2	1	77	926	926	3	1	1	1	1			5	3			
3	1	2	1	2	10	125	126	3	1		3	2	2	3		1	1		
4	1	2	2	1	1	22	23	4	2		1	1							
5	1	2	2	1	49	597	597	3	1	1	1	1			5	3		1	
6	1	2	2	2	19	230	230	5	4	2	1	1		15		3	1	5	
7	1	2	2	2	59	712	712	4	2	2	1	1			2	1		1	
8	1	2	1	1	13	159	160	3	1		1	1		5		1	1		

POSSIBLE QUESTIONS:

1. **HOW ARE YOU GOING TO PREDICT DIABETES IN WOMEN?**
   We are going to pattern this study to the PIMA Indians Diabetes Study. This study was conducted for a group of women based Arizona as they had one of the highest recorded rates of Type 2 diabetes in the whole world. **On this study the doctors have**

concluded key factors that have affected women in developing diabetes over time. And these key factors are going to be our starting point in the feature engineering part of our data exploration.

These factors are:

- Glucose Levels (mg/dL)
- Insulin Levels (uU/mL)
- BMI
- Blood Pressure and Skin Thickness
- Fasting Status
- Body Weight
- Family history in diabetes

2.How credible is the study?

- The study followed participants over time which allowed researchers to track the development of diabetes in individuals who did not initially have it.
- This longitudinal aspect made it possible to identify patterns in the data and determine which factors were predictive of future diabetes onset.

While it focuses on a specific population (PIMA Indians in Arizona), **the health indicators are relevant across various populations worldwide, making the dataset useful for understanding diabetes risk broadly.**

And that there is Medical research backing:
The features included in the study align with well-established **medical knowledge** about diabetes risk factors. For example:

High **glucose** and **insulin resistance** are known to be central to the development of Type 2 diabetes.
High **BMI** and **blood pressure** are often associated with metabolic syndrome, a precursor to diabetes.
Age and **family history** have been repeatedly identified as significant risk factors in other epidemiological studies on diabetes.

This PIMA Indian data was also the cornerstone in the development of binary classification models

What were developed using the same model set?

Real-life examples of binary classification models, such as those developed using the Pima Indians Diabetes Dataset, have been applied in various healthcare systems and industries to predict whether a person has or is at risk of developing diabetes. Here are some real-life scenarios where binary classification models are being used to predict diabetes or related health conditions:

1. **Electronic Health Records (EHR) in Hospitals**:

Many hospitals and healthcare providers use predictive models to analyze patient data collected in Electronic Health Records (EHR). These models help predict whether a patient is at risk for diabetes based on factors such as age, weight, family history, and lab results (e.g., glucose levels, BMI, blood pressure).

- **Example**: **Mayo Clinic** and **Cleveland Clinic** both use predictive analytics tools integrated into their EHR systems to flag patients who are at high risk of developing diabetes, allowing doctors to intervene earlier with lifestyle changes or medications to prevent the disease.

2. **Wearable Health Technology**:

Devices such as **Fitbit**, **Apple Watch**, and **Garmin** track biometric data, including heart rate, physical activity, and even blood glucose levels (for certain devices). These devices use machine learning models to predict the risk of diabetes or detect irregular patterns in glucose levels.

- **Example**: **Apple Health** integrates various health metrics, and combined with the user's medical history, can help predict conditions such as diabetes or provide early warning signs that the user should consult with a doctor for further testing.

4. **Telemedicine Platforms**:

Telemedicine platforms like **Teladoc** or **Doctor on Demand** often integrate machine learning models to predict the likelihood of diabetes in patients during remote consultations. Based on patient input data such as weight, family history, and recent blood tests, the platform can alert doctors to the possibility of undiagnosed diabetes.

- **Example**: A telemedicine service might use a binary classification model to identify patients who have not yet been diagnosed but are exhibiting high-risk indicators of diabetes. Doctors can then advise these patients to undergo further testing or make lifestyle adjustments.

Why do we not have any competitors? Why is this app not available in google play or why hasnt anyone predicted diabetes before:

Building a reliable diabetes prediction app is difficult because of **strict privacy rules**, the **need for high accuracy to avoid risks**, and **complex regulations** that treat such apps as medical tools needing approval. Models also need diverse data to prevent biased results, making development more challenging.

- We will not be impacted by strict privacy rules because we are not going to collect data.
- High accuracy model is what we will aim to have.
- We will not be under any complex regulations because we will only mention highly likely and percentage and unlikely diabetic with percentage.
- Mis diagnosis is highly unlikely with a proven model that actually predicts diabetes because our research is backed up by doctors who have done this study for a long period of time.

Why us and not our competitors: **We dont have a competitor. Our competitors have focused on health management application**. Our focus is wanting to predict if that person is diabetic or not. Or not yet, or people who are worried.