**{Project\_Logo} (**[**Free online logo maker**](https://www.google.com/search?client=firefox-b-1-d&q=free+logo+maker)**)**

### Final\_Project\_Name

### **Group4 Team Members**

|  |  |  |
| --- | --- | --- |
| **Name** | **Email** | **Phone Number** |
| Wei Song | weidongsong1985@gmail.com | 973-818-3138 |
| Jessica Bauman | jbauman815@gmail.com | 732-682-1695 |
| Xiaoyi (Cheryl) Xie | xiaoyixiex@gmail.com | 513-827-8189 |
| Jason Wu |  |  |

### **Background**

Diabetes is a chronic disease that occurs when the pancreas doesn’t produce enough insulin or the body can’t use it as well as it should. Our bodies break down the food we eat into glucose and gets released into the bloodstream. When blood glucose levels go up, insulin should let the glucose into the body’s cells to be used for energy. When insulin is low, too much glucose stays in the bloodstream and can cause life threatening complications. Diabetes can be caused by genetics or lifestyle factors like eating and exercise habits. Over time, diabetes can lead to other serious problems like kidney disease, heart disease or vision loss.

The data file contains 253,680 survey responses to the CDC's BRFSS2015

### **Motivation**

Team members are interested in the healthcare industry or have some medical background. We were inspired to explore diabetes health data using machine learning.

### **Questions to answer**

* After testing the machine learning models, which model has the highest accuracy?
* What is the correlation between the vector and the target variable?

### **Tools/Modules to use**

* Python
* Pandas
* NumPy
* AWS
* Tableau
* pgAdmin 4
* etc.

### **Data sets to use**

[Diabetes Health Indicators Dataset | Kaggle](https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset)

### **Tasks Breakdown**

Wei: create database; clean the data; decision trees model; etc.

Jessica: Dashboard and Visualization (Tableau); etc.

Cheryl: Project Manager; experiment supervised ML; etc.

Jason: Presentation.

### **Tasks and timeline**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Date** | **Task** | **Notes** |
| **Seg-1.1** | **2/28/23** | Topic: Diabetes (Approved),  Dataset: Diabetes Health Indicators Dataset (Approved)  GitHub Repo: Final\_Project\_Group4 (Created)  Added README file.  Created Group chat on Slack. |  |
| **Seg-1.2** | **3/1/23** | Cleaned dataset |  |
| **Seg-1.3** | **3/2/23** | Selected questions.  Added feature branches to main. |  |
| **Seg-1.4** | **3/3/23** | Perform exploratory analysis on the dataset. |  |
| **Seg-1.5** | **3/4/23** |  |  |
| **Seg-1.6** | **3/5/23** | SVM |  |
| **Seg-2.1** | **3/6/23** | Decision Trees Model |  |
| **Seg-2.2** | **3/7/23** | Created a database that stores data with three tables.  Added a project description. |  |
| **Seg-2.3** | **3/8/23** |  |  |
| **Seg-2.4** | **3/9/23** |  |  |
| **Seg-2.5** | **3/10/23** |  |  |
| **Seg-2.6** | **3/11/23** |  |  |
| **Seg-3.1** | **3/12/23** |  |  |
| **Seg-3.2** | **3/13/23** |  |  |
| **Seg-3.3** | **3/14/23** |  |  |
| **Seg-3.4** | **3/15/23** |  |  |
| **Seg-3.5** | **3/16/23** |  |  |
| **Seg-3.6** | **3/17/23** |  |  |
| **Seg-4.1** | **3/18/23** |  |  |
| **Seg-4.2** | **3/19/23** |  |  |
| **Seg-4.3** | **3/20/23** |  |  |
| **Seg-4.4** | **3/21/23** | Mock Presentation |  |
| **Seg-4.5** | **3/22/23** |  |  |
| **Seg-4.6** | **3/23/23** | PROJECT PRESENTATION | 4th Segment and  Self Assessment Due |
| **Seg-5.1** |  |  | All submissions Due |

### **Presentation**

Divide your presentation steps to tasks and assign it to members.

Suggested by:  
Khaled Karman