



# CDC Diabetes Health Indicators

External

Linked on 9/25/2023

The Diabetes Health Indicators Dataset contains healthcare statistics and lifestyle survey information about people in general along with their diagnosis of diabetes. The 35 features consist...



## Dataset Characteristics

Tabular, Multivariate

## Associated Tasks

Classification

## # Instances

253680

## Subject Area

Health and Medicine

## Feature Type

Categorical, Integer

## # Features

21

## Dataset Information



### For what purpose was the dataset created?

To better understand the relationship between lifestyle and diabetes in the US

### Who funded the creation of the dataset?

The CDC

### What do the instances in this dataset represent?

Each row represents a person participating in this study.

### Are there recommended data splits?

Cross validation or a fixed train-test split could be used.

### Does the dataset contain data that might be considered sensitive in any way?

- Gender
- Income
- Education level

### Was there any data preprocessing performed?

Bucketing of age

## Additional Information

Dataset link: [https://www.cdc.gov/brfss/annual\\_data/annual\\_2014.html](https://www.cdc.gov/brfss/annual_data/annual_2014.html)

## Has Missing Values?

No

## Introductory Paper

### [Incidence of End-Stage Renal Disease Attributed to Diabetes Among Persons with Diagnosed Diabetes — United States and Puerto Rico, 2000–2014](#)

By Nilka Rios Burrows, MPH; Israel Hora, PhD; Linda S. Geiss, MA; Edward W. Gregg, PhD; Ann Albright, PhD. 2017

Published in Morbidity and Mortality Weekly Report

## Variables Table

| Variable Name   | Role    | Type    | Demographic | Description  | Units | Missing Values |
|-----------------|---------|---------|-------------|--|-------|----------------|
| ID              | ID      | Integer |             | Patient ID   |       | no             |
| Diabetes_binary | Target  | Binary  |             | 0 = no diabetes 1 = prediabetes or diabetes                              |       | no             |
| HighBP          | Feature | Binary  |             | 0 = no high BP 1 = high BP   |       | no             |
| HighChol        | Feature | Binary  |             | 0 = no high cholesterol 1 = high cholesterol                             |       | no             |
| CholCheck       | Feature | Binary  |             | 0 = no cholesterol check in 5 years 1 = yes cholesterol check in 5 years |       | no             |

| Variable Name        | Role    | Type    | Demographic | Description  | Units | Missing Values |
|----------------------|---------|---------|-------------|--|-------|----------------|
| BMI                  | Feature | Integer |             | Body Mass Index  |       | no             |
| Smoker               | Feature | Binary  |             | Have you smoked at least 100 cigarettes in your entire life? [Note: 5 packs = 100 cigarettes] 0 = no 1 = yes |       | no             |
| Stroke               | Feature | Binary  |             | (Ever told) you had a stroke. 0 = no 1 = yes   |       | no             |
| HeartDiseaseorAttack | Feature | Binary  |             | coronary heart disease (CHD) or myocardial infarction (MI) 0 = no 1 = yes                                    |       | no             |
| PhysActivity         | Feature | Binary  |             | physical activity in past 30 days - not including job 0 = no 1 = yes   |       | no             |
| Fruits               | Feature | Binary  |             | Consume Fruit 1 or more times per day 0 = no 1 = yes   |       | no             |
| Veggies              | Feature | Binary  |             | Consume Vegetables 1 or more times per day 0 = no 1 = yes  |       | no             |

| Variable Name     | Role    | Type    | Demographic | Description  | Units | Missing Values |
|-------------------|---------|---------|-------------|--|-------|----------------|
| HvyAlcoholConsump | Feature | Binary  |             | Heavy drinkers (adult men having more than 14 drinks per week and adult women having more than 7 drinks per week) 0 = no 1 = yes |       | no             |
| AnyHealthcare     | Feature | Binary  |             | Have any kind of health care coverage, including health insurance, prepaid plans such as HMO, etc. 0 = no 1 = yes                |       | no             |
| NoDocbcCost       | Feature | Binary  |             | Was there a time in the past 12 months when you needed to see a doctor but could not because of cost? 0 = no 1 = yes             |       | no             |
| GenHlth           | Feature | Integer |             | Would you say that in general your health is: scale 1-5 1 = excellent 2 = very good 3 = good 4 = fair 5 = poor                   |       | no             |

| Variable Name | Role    | Type    | Demographic | Description  | Units | Missing Values |
|---------------|---------|---------|-------------|--|-------|----------------|
| MentHlth      | Feature | Integer |             | Now thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good? scale 1-30 days |       | no             |
| PhysHlth      | Feature | Integer |             | Now thinking about your physical health, which includes physical illness and injury, for how many days during the past 30 days was your physical health not good? scale 1-30 days                |       | no             |
| DiffWalk      | Feature | Binary  |             | Do you have serious difficulty walking or climbing   |       | no             |

| Variable Name | Role    | Type    | Demographic     | Description   | Units | Missing Values |
|---------------|---------|---------|-----------------|---|-------|----------------|
|               |         |         |                 | stairs? 0 = no 1 = yes  |       |                |
| Sex           | Feature | Binary  | Sex             | 0 = female 1 = male   |       | no             |
| Age           | Feature | Integer | Age             | 13-level age category (_AGEG5YR see codebook) 1 = 18-24 9 = 60-64 13 = 80 or older  |       | no             |
| Education     | Feature | Integer | Education Level | Education level (EDUCA see codebook) scale 1-6 1 = Never attended school or only kindergarten 2 = Grades 1 through 8 (Elementary) 3 = Grades 9 through 11 (Some high school) 4 = Grade 12 or GED (High school graduate) 5 = College 1 year to 3 years (Some college or technical school) 6 = College 4 years or more (College graduate) |       | no             |

| Variable Name | Role    | Type    | Demographic | Description  | Units | Missing Values |
|---------------|---------|---------|-------------|--|-------|----------------|
| Income        | Feature | Integer | Income      | Income scale (INCOME2 see codebook) scale 1-8 1 = less than \$10,000 5 = less than \$35,000 8 = \$75,000 or more |       | no             |

Rows per page 25 0 to 23 of 23

Additional Variable Information

- Diabetes diagnosis

- Demographics (race, sex)

- Personal information (income, educations)...

SHOW MORE

Class Labels

- Diabetes

- Pre-diabetes

- Healthy

Reviews

4.0

★

★

★

★

★

☆

(1 rating)

LOGIN TO WRITE A REVIEW

Alpha Guya

★

★

★

★

☆

-

1/4/2024



IMPORT IN PYTHON

” 1 citations

👁 109479 views

## Citations/Acknowledgements

If you use this dataset, please follow the acknowledgment policy on the [original dataset website](#).

## Keywords

Survey

## DOI

10.24432/C53919

## License

See linked dataset for licensing information.



**UC Irvine**  
Machine Learning  
Repository

### THE PROJECT

About Us

CML

National Science Foundation

### NAVIGATION

Home

View Datasets

Donate a Dataset

### LOGISTICS

Contact



[Privacy Notice](#)

[Feature Request or Bug Report](#)