

Diabetes Health Indicators Dataset Project

Context

- This dataset includes patient profiles with features based on demographics, lifestyle habits, family history, and clinical measurements that are well-established indicators of diabetes risk. All data is generated using statistical distributions inspired by real-world medical research, ensuring privacy preservation while reflecting realistic health patterns.

Attribute Information:

- 1- patient_id: Unique patient identifier
- 2- age: Age of patient in years
- 3- gender : Patient gender
- 4- ethnicity : Ethnic background
- 5- education_level: Highest completed education
- 6- income_level: Income category
- 7- employment_status: Employment type
- 8- smoking_status: Smoking behavior
- 9- alcohol_consumption_per_week: Drinks consumed per week
- 10- physical_activity_minutes_per_week: Physical activity (weekly minutes)
- 11- diet_score: Diet quality (higher = healthier)
- 12- sleep_hours_per_day: Average daily sleep hours
- 13- screen_time_hours_per_day: Average daily screen time hours
- 14- family_history_diabetes: Family history of diabetes
- 15- hypertension_history: Hypertension history
- 16- cardiovascular_history: Cardiovascular history
- 17- bmi: Body Mass Index (kg/m^2)
- 18- waist_to_hip_ratio: Waist-to-hip ratio
- 19- systolic_bp: Systolic blood pressure (mmHg)
- 20- diastolic_bp: Diastolic blood pressure (mmHg)
- 21- heart_rate: Resting heart rate (bpm)
- 22- cholesterol_total: Total cholesterol (mg/dL)
- 23- hdl_cholesterol: HDL cholesterol (mg/dL)
- 24- ldl_cholesterol: LDL cholesterol (mg/dL)
- 25- triglycerides: Triglycerides (mg/dL)

26- glucose_fasting: Fasting glucose (mg/dL)
27- glucose_postprandial: Post-meal glucose (mg/dL)
28- insulin_level: Blood insulin level (μ U/mL)
29- hba1c: HbA1c (%)
30- diabetes_risk_score: Risk score (calculated, 0–100)
31- diabetes_stage: Stage of diabetes
32- diagnosed_diabetes: Target: Diabetes diagnosis

Data Information:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100000 entries, 0 to 99999
Data columns (total 31 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   age              100000 non-null   int64  
 1   gender            100000 non-null   object  
 2   ethnicity         100000 non-null   object  
 3   education_level   100000 non-null   object  
 4   income_level      100000 non-null   object  
 5   employment_status 100000 non-null   object  
 6   smoking_status    100000 non-null   object  
 7   alcohol_consumption_per_week 100000 non-null   int64  
 8   physical_activity_minutes_per_week 100000 non-null   int64  
 9   diet_score         100000 non-null   float64 
 10  sleep_hours_per_day 100000 non-null   float64 
 11  screen_time_hours_per_day 100000 non-null   float64 
 12  family_history_diabetes 100000 non-null   int64  
 13  hypertension_history 100000 non-null   int64  
 14  cardiovascular_history 100000 non-null   int64  
 15  bmi               100000 non-null   float64 
 16  waist_to_hip_ratio 100000 non-null   float64 
 17  systolic_bp        100000 non-null   int64  
 18  diastolic_bp       100000 non-null   int64  
 19  heart_rate          100000 non-null   int64  
 ...
 29  diabetes_stage     100000 non-null   object  
 30  diagnosed_diabetes 100000 non-null   int64  
dtypes: float64(8), int64(16), object(7)
memory usage: 23.7+ MB
```

Data overview:

Index	age	gender	ethnicity	education level	income level	employment status	smoking status	alcohol consumption per week	physical activity minutes per week	dipl score	total cholesterol	ldl cholesterol	hdl cholesterol	triglycerides	glucose fasting	glucose postprandial	insulin level	sbp	diabetes risk score	diabetes stage	diagnosed diabetes
0	58	Male	African	Highschool	Low-Middle	Employed	Never	0	215	67	-	41	90	50	154	206	426	3.10	294	Type 2	1
1	40	Female	White	Highschool	Middle	Employed	Former	1	140	47	-	35	80	35	105	170	300	3.03	233	No Diabetes	0
2	56	Male	Hispanic	Highschool	Middle	Unemployed	Never	1	57	64	-	35	85	35	115	185	300	5.07	231	Type 2	1
3	74	Female	Black	Highschool	Low	Retired	Never	0	49	24	-	35	75	50	105	152	320	3.83	302	Type 2	1
4	46	Male	White	Graduate	Middle	Employed	Never	1	109	32	-	52	125	60	157	184	324	2.26	233	Type 2	1
5	55	Female	White	Graduate	Upper-Middle	Employed	Former	1	155	62	-	40	105	75	115	195	340	5.55	250	No Diabetes	0
6	67	Male	White	Graduate	Middle	Employed	Never	2	165	62	-	55	125	60	165	140	300	5.87	244	No Diabetes	0
7	52	Female	Black	Notional	Upper-Middle	Employed	Former	4	120	55	-	50	110	55	105	130	237	3.21	273	No Diabetes	0
8	49	Female	Black	Highschool	Low-Middle	Retired	Never	3	152	55	-	48	95	75	115	161	285	5.53	264	No Diabetes	0
9	52	Female	White	Postgraduate	Middle	Employed	Never	4	188	37	-	31	95	37	95	142	330	5.73	273	No Diabetes	0
100000 rows x 21 columns																					

Data Information (Cont.):

- data with (100000) record & (32) features (columns).
- Data with (13) float columns & (12) int & (7) strings.

