

Informe sobre el dataset

- **Práctica 1 - Parte I: Selección del Conjunto de Datos**
 - Introducción: "justificación del conjunto de datos elegido"
 - 1. Justificación de la Selección
 - 2. Relevancia del conjunto de datos
 - 3. Complejidad del conjunto de datos
 - 4. Originalidad
 - 5. Cuestiones a responder con la visualización
 - BIBLIOGRAFÍA
 - ANEXO: **Diabetes Mellitus: Visión General**
 - Contextualization

Práctica 1 - Parte I: Selección del Conjunto de Datos

Autor Anton Barrera Mora (abarreramora@uoc.edu)

Fecha: 20 de abril de 2024

Introducción: Justificación del conjunto de datos seleccionado

En este trabajo presentamos la selección y justificación sobre conjunto de datos elegido para la realización de la práctica 1 de la asignatura de visualización de datos, de acuerdo con los criterios establecidos en el enunciado de la asignatura. El dataset seleccionado es el "[CDC Diabetes Health Indicators](#)" alojado en el repositorio [UCI Machine Learning](#).

1. Justificación de la Selección

La elección de este conjunto de datos pivota en torno a tres ejes:

1. Casos diagnosticados:

Responde a un interés personal, ya contamos diferentes casos de diagnostico en el entorno.

2. Ámbito profesional relevante:

Responde a interés profesional. Como psicólogo clínico especializado en psicología geriátrica. No albergamos dudas sobre el papel que la diabetes y las enfermedades crónicas asociadas representan uno de los mayores desafíos sanitarios a nivel global y nacional, económico y personal, especialmente en personas de edad avanzada. Esto cobra mayor relevancia si tenemos en cuenta el evidente envejecimiento poblacional en los países desarrollados. Es frecuente que el diagnostico de esta dolencia curse comórbido con otros problemas de salud mental en una relación bidireccional especialmente en la depresión y la ansiedad, plagas de salud mental en la actualidad. Por todo lo anterior, creemos que cualquier trabajo en este ámbito puede derivar en mejoras en el tratamiento de la sintomatología y los aspectos psicosociales derivados.

3. Ámbito académico consistente y relevante:

Responde a un necesidad académica de conocer y contar con herramientas para afrontar los retos a diferentes niveles. No cabe duda de que la ciencia de datos es la herramienta perfecta para el desarrollo de modelos predictivos y de clasificación que nos ayuden a conocer mejor los aspectos sobre este problema de salud.

La investigación en diabetes se ha beneficiado enormemente del avance de las herramientas computacionales y la disponibilidad de grandes volúmenes de datos. La minería de datos y las técnicas de aprendizaje automático se emplean cada vez más para desarrollar modelos predictivos y de agrupación, capaces de identificar patrones ocultos y relaciones complejas en conjuntos de datos extensos. Asimismo, el número de publicaciones científicas en esta área ha experimentado un crecimiento constante, reflejo del potencial de estas metodologías para transformar la comprensión y el manejo de la diabetes. Pero cabe oponer no obstante en pocas ocasiones se ha tenido en cuenta la salud mental, en un sentido multicausal o en un sentido correlacional.

4. Ámbito técnico relevante:

Reflejo del potencial de estas metodologías para transformar la comprensión y el manejo de la diabetes, aunque a menudo centrándose en la predicción más que en la exploración visual detallada de las interrelaciones, **existen gran cantidad de abordajes desde la minería de datos** que pueden ayudar a enriquecer en cualquier momento este trabajo, convirtiendo todo en escalable a demanda.

Este trabajo por tanto, podría ser una primera aproximación que habilite futuros abordajes, teniendo en cuenta que ha sido muy tratado en minería de datos como hemos puesto de relieve anteriormente.

Por otro lado, cabe abundar en que el *dataset* seleccionado nace derivado de una encuesta a gran escala como el **BRFSS (Behavioral Risk Factor Surveillance System)** del CDC.

La BRFSS es una encuesta telefónica sobre salud que los CDC recopilan anualmente. Recopila las respuestas de más de 400,000 estadounidenses sobre comportamientos de riesgo para la salud, enfermedades crónicas y el uso de servicios prevención desde 1984.

Contiene preguntas formuladas directamente a los participantes o variables calculadas basadas en las respuestas individuales de cada participante, que abarcarían los factores de riesgos reconocidos como más importantes (hábitos de consumo, indicadores fisiológicos, etc.), incluyendo algunas cuestiones (aunque algo genéricas) sobre salud mental.

Creemos que este *dataset* ofrece una oportunidad única para:

- Aplicar las técnicas de visualización de datos trabajadas en el máster para explorar patrones y relaciones complejas en datos de salud poblacionales.
- Generar *insights* sobre la interrelación entre indicadores de salud, comportamientos, factores socioeconómicos y la prevalencia de diabetes.
- Contribuir en un principio, aunque sea a nivel académico y de forma tentativa a una mejor comprensión de factores asociados a esta condición.

El objetivo final es crear visualizaciones que puedan ser informativas y generar conciencia sobre el problema, mejorando las medidas paliativas a todos los niveles.

2. Relevancia del conjunto de datos

El conjunto de datos "*CDC Diabetes Health Indicators*" presenta una alta relevancia en su contexto por las siguientes razones:

- **Actualidad:** Los datos provienen de encuestas realizadas a partir de 2015 y versiones del dataset han sido actualizadas hasta 2021/2022, cumpliendo el requisito de trabajar con **datos actuales**, lo cual es muy importante en el dinámico campo de la salud pública.
- **Importancia temática:** La diabetes es una enfermedad crónica de alta prevalencia e impacto significativo en la calidad de vida y el gasto sanitario. Analizar indicadores relacionados es fundamental para la vigilancia epidemiológica y la planificación de intervenciones. Afecta a un **colectivo muy amplio y diverso**.
- **Fuente Confiable:** Proviene de los *Centers for Disease Control and Prevention (CDC)* de los USA, una fuente de datos de salud pública de alta reputación y con un alto rigor metodológico (BRFSS).
- **Perspectiva de género:** El dataset incluye la variable **Sex**, lo que **permite explícitamente incorporar y analizar la perspectiva de género** en la prevalencia de diabetes y la distribución de los indicadores de salud asociados, cumpliendo con otro aspecto valorado en el enunciado.
- **Licencia:** Los datos, al ser un producto del gobierno de USA, se consideran de dominio público en dicho país, facilitando su uso para fines académicos con la debida **atribución a la fuente (CDC/BRFSS)**. Asimismo el mismo dataset estaría disponible en diferentes repositorios, como es el caso de [Kaggle](#) en dominio público.

3. Complejidad del conjunto de datos

El dataset cumple con los requisitos de complejidad necesarios para desarrollar un proyecto de visualización interesante:

- **Tamaño (Registros):** Contiene **más de 250,000 registros**, superando ampliamente el mínimo requerido de miles de registros, asegurando la base estadística para el análisis.
- **Tamaño (Variables):** Incluye **21 variables iniciales**. Aunque se encuentra en el rango bajo del requisito de "*decenas de variables*", es un número suficiente para explorar múltiples facetas del problema. Además, es factible **la realización de ingeniería de características** al objeto de generar nuevas variables, eventualmente incrementando la dimensionalidad para alimentar diferentes visualizaciones.
- **Tipos de Datos:** Presenta una **combinación de tipos de datos**, predominantemente:
 - **Binarios (Lógicos):** Múltiples indicadores de sí/no (0/1) como **HighBP**, **HighChol**, **Smoker**, **HeartDiseaseorAttack**, etc.
 - **Ordinales/Categóricos:** Escalas como **GenHlth**, **PhysHlth**, **MentHlth**, **Education**, **Income**. La variable objetivo **Diabetes_binary** (o su versión 012) también es categórica/ordinal.
 - **Cuantitativos:** Incluye **BMI** (Índice de Masa Corporal) y **Age** (categorizada en grupos).
 - Pese a carecer de variables de fecha/hora o cartográficas detalladas de forma inherente (se antoja bastante complejo encontrar este tipo de datos dada la naturaleza del problema), la combinación de variables existente habilita la aplicación de diversas técnicas de visualización.
- **Riqueza tipológica:** La presencia de múltiples indicadores binarios y escalas ordinales, junto con variables demográficas clave, ofrece riqueza para explorar patrones y relaciones complejas entre diferentes aspectos de la salud y el estilo de vida.

Visualizamos las variables disponibles y los metadatos:

```
{'uci_id': 891, 'name': 'CDC Diabetes Health Indicators', 'repository_url':
'https://archive.ics.uci.edu/dataset/891/cdc+diabetes+health+indicators',
'data_url': 'https://archive.ics.uci.edu/static/public/891/data.csv',
'abstract': '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 of some
demographics, lab test results, and answers to survey questions for each
patient. The target variable for classification is whether a patient has
diabetes, is pre-diabetic, or healthy. ', 'area': 'Health and Medicine',
'tasks': ['Classification'], 'characteristics': ['Tabular',
'Multivariate'], 'num_instances': 253680, 'num_features': 21,
'feature_types': ['Categorical', 'Integer'], 'demographics': ['Sex', 'Age',
'Education Level', 'Income'], 'target_col': ['Diabetes_binary'],
'index_col': ['ID'], 'has_missing_values': 'no', 'missing_values_symbol':
None, 'year_of_dataset_creation': 2017, 'last_updated': 'Fri Nov 03 2023',
'dataset_doi': '10.24432/C53919', 'creators': [], 'intro_paper': {'ID':
308, 'type': 'NATIVE', 'title': 'Incidence of End-Stage Renal Disease
Attributed to Diabetes Among Persons with Diagnosed Diabetes – United
States and Puerto Rico, 2000–2014', 'authors': 'Nilka Rios Burrows, MPH;
Israel Hora, PhD; Linda S. Geiss, MA; Edward W. Gregg, PhD; Ann Albright,
PhD', 'venue': 'Morbidity and Mortality Weekly Report', 'year': 2017,
'journal': None, 'DOI': None, 'URL':
'https://www.cdc.gov/mmwr/volumes/66/wr/mm6643a2.htm', 'sha': None,
'corpus': None, 'arxiv': None, 'mag': None, 'acl': None, 'pmid': None,
'pmcid': None}, 'additional_info': {'summary': 'Dataset link:
https://www.cdc.gov/brfss/annual_data/annual_2014.html', 'purpose': 'To
better understand the relationship between lifestyle and diabetes in the
US', 'funded_by': 'The CDC', 'instances_represent': 'Each row represents a
person participating in this study.', 'recommended_data_splits': 'Cross
validation or a fixed train-test split could be used.', 'sensitive_data':
'- Gender\n- Income\n- Education level', 'preprocessing_description':
'Bucketing of age', 'variable_info': '- Diabetes diagnosis\n- Demographics
(race, sex)\n- Personal information (income, educations)\n- Health history
(drinking, smoking, mental health, physical health)', 'citation': None},
'external_url': 'https://www.kaggle.com/datasets/alexteboul/diabetes-
health-indicators-dataset'}
```

	name	role	type	demographic \
0	ID	ID	Integer	None
1	Diabetes_binary	Target	Binary	None
2	HighBP	Feature	Binary	None
3	HighChol	Feature	Binary	None
4	CholCheck	Feature	Binary	None
5	BMI	Feature	Integer	None
6	Smoker	Feature	Binary	None
7	Stroke	Feature	Binary	None
8	HeartDiseaseorAttack	Feature	Binary	None
9	PhysActivity	Feature	Binary	None
10	Fruits	Feature	Binary	None
11	Veggies	Feature	Binary	None
12	HvyAlcoholConsump	Feature	Binary	None
13	AnyHealthcare	Feature	Binary	None
14	NoDocbcCost	Feature	Binary	None
15	GenHlth	Feature	Integer	None

	description	units	missing_values
0	Patient ID	None	no
1	0 = no diabetes 1 = prediabetes or diabetes	None	no
2	0 = no high BP 1 = high BP	None	no
3	0 = no high cholesterol 1 = high cholesterol	None	no
4	0 = no cholesterol check in 5 years 1 = yes ch...	None	no
5	Body Mass Index	None	no
6	Have you smoked at least 100 cigarettes in you...	None	no
7	(Ever told) you had a stroke. 0 = no 1 = yes	None	no
8	coronary heart disease (CHD) or myocardial inf...	None	no
9	physical activity in past 30 days - not includ...	None	no
10	Consume Fruit 1 or more times per day 0 = no 1...	None	no
11	Consume Vegetables 1 or more times per day 0 =...	None	no
12	Heavy drinkers (adult men having more than 14 ...	None	no
13	Have any kind of health care coverage, includi...	None	no
14	Was there a time in the past 12 months when yo...	None	no
15	Would you say that in general your health is: ...	None	no
16	Now thinking about your mental health, which i...	None	no
17	Now thinking about your physical health, which...	None	no
18	Do you have serious difficulty walking or clim...	None	no
19	0 = female 1 = male	None	no
20	13-level age category (_AGEG5YR see codebook) ...	None	no
21	Education level (EDUCA see codebook) scale 1-6...	None	no
22	Income scale (INCOME2 see codebook) scale 1-8 ...	None	no

La exploración inicial refuerza la idea de que, si bien **MentHlth** ofrece un buen punto de partida, para análisis más profundo de la salud mental podría ser conveniente explorar la posibilidad de enriquecer el *dataset* en fases futuras, aunque el foco inicial será extraer el máximo de las variables existentes y las generadas.

5 / 9

Buscamos un enfoque original en el uso de este conjunto de datos:

- **Dataset y enfoque:** Aunque la diabetes es un tema muy estudiado, el uso específico de este dataset reciente de indicadores del CDC para una *visualización de datos exploratoria y relacional* (más allá de dashboards básicos) nos brinda un grado de originalidad. El foco estará en la **interconexión visual de múltiples indicadores de salud y comportamiento teniendo en cuenta la salud mental**, en lugar de los análisis clínicos o predictivos tradicionales.
- **Evitación clásicos y saturados:** El proyecto se aleja necesariamente de los datasets clásicos de aprendizaje automático (Iris, Titanic, etc.) y de temas cuya visualización puede estar más saturada (COVID-19, tráfico, criminalidad general, etc.).
- **Mejora y Enriquecimiento:** Se planea **mejorar el conjunto de datos generando nuevas métricas e indicadores** mediante transformaciones de las variables existentes. Por ejemplo:
 - Crearemos un índice de riesgo cardiovascular combinando **HighBP, HighChol, BMI**.
 - Agruparemos niveles de **Education** o **Income** de formas significativas.
 - Combinaremos indicadores de estilo de vida (dieta, ejercicio, tabaco) en un score de "hábitos saludables".
 - Exploraremos, si es factible y relevante en fases posteriores, **la posibilidad de enriquecer con datos externos agregados a nivel geográfico** si se pudiera identificar dicha variable. Asimismo observamos que existen gran cantidad de conjuntos de datos desde los que **se podría explorar enriquecer este trabajo**.
- **Visualizaciones existentes:** Si bien el CDC y otros organismos publican gráficos básicos con datos del BRFSS, el objetivo de esta práctica es crear una visualización más profunda e interactiva que explore las **relaciones multidimensionales** de una manera que probablemente no ha sido representada visualmente de forma extensa para este dataset derivado específico en el contexto académico de visualización atendiendo a la salud mental.

5. Cuestiones a responder con la visualización

La visualización buscará responder preguntas específicas que aprovechen la estructura y riqueza del dataset, teniendo en cuenta los puntos anteriores y buscando un enfoque novedoso:

- **Preguntas:**
 1. ¿Qué patrones de asociación visual existen entre los auto-reportes de problemas de salud mental (**MentHlth**) y el diagnóstico de diabetes? ¿Se observan perfiles de riesgo distintos cuando se combina **MentHlth** con otros factores (socioeconómicos, de comportamiento, etc.)?
 2. ¿Cómo se interrelacionan visualmente los indicadores de comportamiento (consumo de frutas/vegetales, actividad física, tabaquismo) con la prevalencia de diabetes/prediabetes, y cómo varían estos patrones entre diferentes grupos de edad y género (**Age, Sex**)?
 3. ¿Qué correlaciones visuales existen entre factores socioeconómicos (**Education, Income**) y la acumulación de indicadores de riesgo (**HighBP, HighChol, BMI**), diferenciando entre población con y sin diabetes?
 4. ¿Es posible identificar "clusters" o perfiles visuales de individuos basados en la combinación de múltiples indicadores de salud (e.g. **MentHlth, BMI, PhysActivity, Smoker, GenHlth**), y cómo se asocian estos perfiles con la diabetes?
 5. ¿Cómo se visualiza la relación entre el acceso auto-reportado a servicios de salud (**AnyHealthcare, NoDocbcCost**) y la condición de diabetes, considerando variables demográficas y socioeconómicas? Más allá de predecir quién desarrollará diabetes

6. Aunque carecemos de datos étnicos, se podría abordar la creación de diferentes grupos (e.g. según niveles educativos u otros aspectos socioeconómicos). En caso de poder "perfilar" las instancias, cabría preguntarse ¿Cómo podemos visualizar los diferentes 'caminos' o combinaciones de factores de riesgo (comportamentales, socioeconómicos, salud general y mental) que se asocian con la condición en diferentes grupos poblacionales?

- **Adecuación y relevancia:** Estas preguntas son directamente asumibles con las variables presentes en el dataset o las que se generarían según fuese necesario, sin olvidar que el conjunto de datos se presta a incorporar datos de los múltiples trabajos existentes. Exploran las relaciones entre múltiples factores (complejidad), abordan aspectos relevantes de salud pública y consideran variables demográficas clave (incluyendo género).
- **Originalidad de las preguntas:** Se centran en la exploración visual de la *interconexión* y los *patrones combinados* de indicadores, más allá de análisis univariados o bivariados simples.
- **Diccionario de variables:** Como paso fundamental previo a la visualización, **elaboraremos un diccionario de datos detallado**.

Este diccionario incluiría:

- Nombre de cada variable (original y generada).
- Descripción clara de su significado según la codificación del BRFSS/CDC.
- Tipo de dato (binario, ordinal, cuantitativo, categórico).
- Rango de valores o categorías posibles.
- Identificación de su rol potencial como "hecho" a medir o "dimensión" de análisis.

Este diccionario debería ser clave en la guía tanto del preprocesamiento (ETL) como el diseño de la visualización final, asegurando que las preguntas planteadas puedan ser respondidas con rigor.

BIBLIOGRAFÍA

- Ahmed, A., Khan, J., Arsalan, M., Ahmed, K., Shahat, A. A., Alhalimi, A., & Naaz, S. (2024). Machine Learning Algorithm-Based Prediction of Diabetes Among Female Population Using PIMA Dataset. *Healthcare*, 13(1), 37. <https://doi.org/10.3390/healthcare13010037>
- awesomedata/awesome-public-datasets: A topic-centric list of HQ open datasets. (s. f.). Recuperado 21 de abril de 2025, de <https://github.com/awesomedata/awesome-public-datasets?tab=readme-ov-file#healthcare>
- Behavioral Risk Factor Surveillance System. (2025, abril 17). <https://www.cdc.gov/brfss/>
- Burrows, N. R. (2017). Incidence of End-Stage Renal Disease Attributed to Diabetes Among ... <https://www.cdc.gov/mmwr/volumes/66/wr/mm6643a2.htm>
- Cho, S. B., Kim, S. C., & Chung, M. G. (2019). Identification of novel population clusters with different susceptibilities to type 2 diabetes and their impact on the prediction of diabetes. *Scientific Reports*, 9, 3329. <https://doi.org/10.1038/s41598-019-40058-y>
- Diabetes detailed EDA with conclusion. (s. f.). Recuperado 21 de abril de 2025, de <https://kaggle.com/code/bharat04/diabetes-detailed-eda-with-conclusion>
- Diabetes Prediction. (s. f.). Recuperado 21 de abril de 2025, de <https://kaggle.com/code/zabihullah18/diabetes-prediction>
- Diabetes Diagnosis & Tests | ADA. (s. f.). Recuperado 21 de abril de 2025, de <https://diabetes.org/about-diabetes/diagnosis>

- Diabetes Health Indicators Dataset. (s. f.). Recuperado 21 de abril de 2025, de <https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset>
- Diabetes Health Indicators Dataset Notebook. (s. f.). Recuperado 21 de abril de 2025, de <https://kaggle.com/code/alexteboul/diabetes-health-indicators-dataset-notebook>
- Diabetes Prediction|EDA +Preprocessing +Models. (s. f.). Recuperado 21 de abril de 2025, de <https://kaggle.com/code/abdallahsaadelgendy/diabetes-prediction-eda-preprocessing-models>
- Diabetes Research, Education, Advocacy | ADA. (s. f.). Recuperado 21 de abril de 2025, de <https://diabetes.org/>
- IDF SCHOOL OF DIABETES. (s. f.). Recuperado 21 de abril de 2025, de <https://www.IDFdiabeteschool.org>
- John Clore, K. C. (2014). Diabetes 130-US Hospitals for Years 1999-2008 [Dataset]. UCI Machine Learning Repository. <https://doi.org/10.24432/C5230J>
- Naz, H., & Ahuja, S. (2020). Deep learning approach for diabetes prediction using PIMA Indian dataset. Journal of Diabetes and Metabolic Disorders, 19(1), 391-403. <https://doi.org/10.1007/s40200-020-00520-5>
- CDC. (2017). CDC Diabetes Health Indicators [Dataset]. UCI Machine Learning Repository. <https://doi.org/10.24432/C53919>
- What Is Diabetes? - NIDDK. (s. f.). National Institute of Diabetes and Digestive and Kidney Diseases. Recuperado 21 de abril de 2025, de <https://www.niddk.nih.gov/health-information/diabetes/overview/what-is-diabetes>
- Wykes, T. L., Lee, A. A., Bourassa, K., Kitchen, K. A., & McKibbin, C. L. (2017). Diabetes Knowledge Among Adults with Serious Mental Illness and Comorbid Diabetes Mellitus. Archives of psychiatric nursing, 31(2), 190-196. <https://doi.org/10.1016/j.apnu.2016.09.014>

ANEXO: Diabetes Mellitus: Visión General

Contextualization

La diabetes mellitus (DM) es un grupo heterogéneo de trastornos metabólicos crónicos caracterizados fundamentalmente por la presencia de **hiperglucemia** (niveles elevados de glucosa en sangre). Esta condición resulta de defectos en la secreción de insulina, en la acción de la insulina, o en ambas. La insulina, hormona peptídica producida por las células beta de los islotes pancreáticos es esencial para regular el metabolismo de los carbohidratos, lípidos y proteínas, facilitando la captación de glucosa por parte de las células para su uso como energía. La hiperglucemia crónica asociada a la diabetes no controlada **conduce a daño a largo plazo, disfunción e insuficiencia de diversos órganos**, especialmente los ojos (retinopatía), riñones (nefropatía), nervios (neuropatía), corazón (enfermedad cardiovascular) y vasos sanguíneos.

Clasificación de la Diabetes Mellitus

La clasificación etiológica de la diabetes propuesta por organizaciones como la Asociación Americana de Diabetes (ADA) y la Organización Mundial de la Salud (OMS) distingue varios tipos principales:

1. **Diabetes Mellitus tipo 1 (DM1):** Este tipo se caracteriza por una **destrucción autoinmune** de las células beta pancreáticas, lo que conduce a una deficiencia absoluta de insulina (Diabetes Research, Education, Advocacy | ADA, s. f.). Suele aparecer en la infancia o la adolescencia, aunque puede manifestarse a cualquier edad. Representa aproximadamente entre el 5% y el 10% de todos los casos de diabetes (What Is Diabetes?, s. f.). Los pacientes con DM1 requieren **tratamiento con insulina exógena** de por vida para sobrevivir y mantener un control glucémico adecuado (IDF SCHOOL OF

DIABETES, s. f.). La etiología exacta implica una interacción compleja entre predisposición genética y factores ambientales desencadenantes.

2. **Diabetes Mellitus tipo 2 (DM2):** Es la forma más prevalente de diabetes, constituyendo entre el 90% y el 95% de los casos (What Is Diabetes?, s. f.). Se caracteriza por un **defecto progresivo en la secreción de insulina sobre un fondo de resistencia a la insulina**. La resistencia a la insulina implica que las células del cuerpo (especialmente en músculo, hígado y tejido adiposo) no responden adecuadamente a la acción de la insulina, lo que requiere que el páncreas produzca mayores cantidades de esta hormona para mantener la normoglucemia. Con el tiempo las células beta pueden perder su capacidad para compensar esta resistencia, llevando a una deficiencia relativa de insulina y a la hiperglucemia de facto. Los factores de riesgo incluyen la edad avanzada, la obesidad (particularmente la adiposidad visceral), la inactividad física, antecedentes familiares de diabetes y pertenencia a ciertos grupos étnicos. El abordaje suele incluir modificaciones del estilo de vida (dieta, ejercicio, alimentación, etc.) junto con agentes antidiabéticos orales y en muchos casos eventualmente, insulina.
3. **Diabetes Mellitus gestacional (DMG):** Se define como cualquier grado de intolerancia a la glucosa que se produce por primera vez **durante el embarazo**. La DMG se produce debido a los cambios hormonales propios de la gestación que inducen un estado de mayor resistencia a la insulina. Si el páncreas materno no puede compensar adecuadamente esta resistencia se desarrolla la hiperglucemia. Generalmente se resuelve después del parto pero la DMG confiere un riesgo significativamente mayor tanto para la madre como para el hijo de desarrollar DM2 en el futuro. Su tratamiento es vital para prevenir complicaciones perinatales.
4. **Otros tipos específicos de diabetes:** Incluye formas menos comunes de diabetes debidas a causas específicas, como:
 - Defectos genéticos en la función de las células beta (e.g., MODY (*Maturity-Onset Diabetes of the Young*)).
 - Defectos genéticos en la acción de la insulina.
 - Enfermedades del páncreas exocrino (e.g., pancreatitis, fibrosis quística, etc.).
 - Endocrinopatías (e.g., acromegalia, síndrome de Cushing, etc.).
 - Diabetes inducida por fármacos o productos químicos (e.g., glucocorticoides, algunos tratamientos para el VIH/SIDA, alcoholismo, adicciones).

Prediabetes:

Finalmente, cabe mencionar el estado de **prediabetes**, que es en el cual los niveles de glucosa en sangre son más altos de lo normal, pero no lo suficientemente elevados como para diagnosticar diabetes tipo 2. La prediabetes indica un alto riesgo de progresión a DM2 así como de enfermedad cardiovascular.
