**Proyecto diabetes**

**Introducción**

La diabetes es una enfermedad crónica que afecta a millones de personas en todo el mundo. Se caracteriza por niveles elevados de glucosa en sangre debido a una deficiencia en la producción de insulina o una resistencia a esta.

La base de datos proporciona información sobre los distintos tipos de diabetes, incluyendo: la diabetes inducida por esteroides, la diabetes mellitus neonatal (NDM), la prediabetes, la diabetes tipo 1 y 2, así como lo del síndrome de Wolfram. La información recopilada abarca una amplia gama de atributos médicos, genéticos y de estilo de vida.

**Objetivo general**

Procesar un conjunto de datos de los tipos de diabetes y sus atributos para analizar la relación entre los tipos de diabetes, la edad, los factores sociodemográficos y los factores biológicos.

**Planteamiento del problema**

La diabetes es una de las principales causas de muerte, con un número creciente de personas que padecen la enfermedad. Las complicaciones de la diabetes afectan la calidad de vida de los pacientes. El objetivo de este proyecto es procesar un conjunto de datos de los tipos de diabetes y sus atributos.

**Posibles preguntas de investigación**

1. ¿Qué tipo de diabetes es el más común en la base de datos?
2. ¿Cuáles son los marcadores genéticos más frecuentes asociados con la diabetes tipo 1?
3. ¿Existe una correlación entre los antecedentes familiares de diabetes y el desarrollo de la diabetes tipo 2?
4. ¿Hay diferencias significativas en el IMC entre pacientes con diabetes tipo 1 y diabetes tipo 2?
5. ¿Cómo es la distribución de los tipos de diabetes?
6. ¿Cómo es la distribución de la diabetes en cuanto a la edad?
7. ¿Cómo es la distribución de la diabetes en cuanto a los hábitos dietéticos, se relaciona con el IMC?
8. ¿Es más probable (qué tan probable es) que padezcan diabetes los fumadores y los consumidores de alcohol?
9. ¿Cuál es la edad más común?
10. ¿Cuál de todas las diabetes demuestra el nivel más alto de colesterol?
11. ¿Cuántos pacientes con diabetes tipo 2 son fumadores?
12. ¿Cuál de todas las diabetes demuestra el nivel más alto de colesterol?

**Posibles soluciones**

El procesamiento de estos datos permitirá que más adelante podamos realizar un análisis más detallado de éste.

**Consideraciones futuras**

Una vez que la base de datos esté limpia y procesada, permitirá realizar un análisis más profundo, como la visualización de los tipos de diabetes 1 y 2, la edad, . En futuras etapas, se podrían utilizar estos datos para implementar modelos predictivos de la diabetes ya que podrían agregar otras condiciones médicas que nos permite identificar causas de la diabetes.

**Conclusión**

Una base de datos limpia y ordenada, nos permite tener una mejor visualización y claridad de lo que se está o se va a trabajar, y esto a su vez nos garantiza la claridad y utilidad de la información que vamos a comunicar. En este proyecto, nos enfocamos primero en encontrar una base de datos con el tema que mayor atención nos llamó, una vez que encontramos la base de datos, comenzamos con la limpieza para posteriormente dar respuesta a una serie de preguntas que se plantearon en base al objetivo general y a la información de la base de datos.