**“Caso de Estudio"**

Se entrega el siguiente set de datos de una empresa del sector de salud, **HealthAnalytics.csv**, el cual es un dataset que cuenta con la información de una cadena de hospitales que tiene como objetivo crear la próxima generación de atención médica para sus pacientes, para esto ha contratado a usted para ayudarlo a lograr su visión. La compañía reúne a los mejores médicos y les permite brindar atención médica proactiva a sus pacientes. En este caso, su cliente quiere estudiar alrededor de una de las enfermedades críticas "**Stroke**". El **accidente cerebrovascular** es una enfermedad que afecta las arterias que conducen hacia y dentro del cerebro. Un accidente cerebrovascular ocurre cuando un vaso sanguíneo que transporta oxígeno y nutrientes al cerebro está bloqueado por un coágulo o estalla (o se rompe). Cuando eso sucede, parte del cerebro no puede obtener la sangre (y el oxígeno) que necesita, por lo que mueren las células cerebrales. En los últimos años, el **Cliente** ha capturado varios detalles de salud, demográficos y de estilo de vida sobre sus pacientes. Esto incluye detalles como la edad y el sexo, junto con varios parámetros de salud (por ejemplo, hipertensión, índice de masa corporal) y variables relacionadas con el estilo de vida (por ejemplo, el tabaquismo, el tipo de ocupación).

Las variables que se disponibilizan son:



id Patient ID

gender Género del paciente

age Edad del paciente hypertension Tenencia de hipertensión ?

heart\_disease Tenencia de enfermedad del corazón ? ever\_married Está casado?

work\_type Tipo de ocupación del paciente Residence\_type Tipo de área de residencia avg\_glucose\_level Nivel promedio de glucosa 

bmi Indice de masa corporal smoking\_status El cliente fuma?

**Consignas:**

1.- Definir el problema de la naturaleza que se tiene a continuación, además de los objetivos de negocio bien definidos.

En primer lugar clasificar en grupos y entender la vulnerabilidad o predisposición a sufrir un evento cardiovascular.

En segundo lugar, tomar medidas preventivas basadas en el análisis de probabilidad/riesgo de una persona a tener un evento.

2.- ¿Qué tipo de variables se utilizan en el problema de negocio?

La mayoría son categóricas, pero contamos también con discretas y continuas.

3.- ¿Cómo podríamos resolver este problema de negocio y cumplir con los objetivos planteados a través de la ciencia o analítica de datos?

Buscaríamos determinar la probabilidad de que otros sufran accidentes cardiovasculares utilizando modelos de clasificación como random forest