

STATISTICAL ESSENTIALS

Laboratorio Practico N°02:

- 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). El Cliente quiere que usted prediga la probabilidad de que ocurra un accidente cerebrovascular a sus pacientes. Esto ayudará a los médicos a tomar medidas de salud proactivas para estos pacientes. Las variables que se disponibilizan son:

Variable	Definition
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?
stroke	0 - No stroke, 1 - Sufrió Stroke

Entregables del Laboratorio:

- 1.- Definir el problema de la naturaleza que se tiene a continuación, además de los objetivos de negocio (Objetivos de ciencias de la salud) bien definidos.
- 2.- ¿Qué tipo de variables se utilizan en el problema? De un ejemplo de cada uno de los tipos de variables estudiados en clase.
- 3.- De lo antes explorado, cuáles los principales factores que influyen en una enfermedad cardiovascular. Aplique una prueba de medias paramétrico y no paramétrico y comentes sus principales resultados.

