**Módulo: ADM1929 - Business understanding: Pensamiento analítico basado en datos. - (A51)**

**Actividad: Reto de aprendizaje 15. Entendimiento de la data**

**Nombre: Roberto Mora Balderas**

**Asesor: José Carlos Soto Monterrubio**

**Fecha: 24 de julio de 2023**

**Objetivo:**

Comprender las características de la data con la que trabajará el caso de aplicación con la finalidad de definir el programa de operación de minería de datos.

**Instrucciones:**

**1.** Importa la [data proporcionada](https://uvanahuac.instructure.com/courses/5434/files/2408043/download) a Rapid Miner, utiliza la pestaña de estadística para identificar las variables, sus valores estadísticos generales, sus niveles de correlación. En un documento de word, desarrolla los siguientes elementos:

* Número de registros de la base de datos y número de variables.
* Nombre y tipo de variables con explicación del significado de cada variable.
* Valores estadísticos descriptivos para las variables que aplique.
* Matriz de correlación con las variables vigentes.
* Identificación de variables de baja calidad con Auto Model.
* Explicación narrativa en un párrafo de la base de datos (Comenta con tus palabras un resumen de qué aspectos son los más importantes que captas en el análisis de la example set)

**Desarrollo**:

A screenshot of a computer

Description automatically generated

Se puede apreciar, que existen 200 registros y 5 variables, CustomerID, Gender, Age, Annual Income, Spending Score.

Aquí podemos observar algunas estadísticas descriptivas de cada una de las variables.

A screenshot of a computer

Description automatically generated

A continuación una breve descripción del significado de cada variable.

* CustomerID: Identificación única del cliente.
* Gender: Género del cliente (por ejemplo, Masculino o Femenino).
* Age: Edad del cliente.
* Annual Income (k$): Ingresos anuales del cliente en miles de dólares.
* Spending Score (1-100): Puntuación de gasto del cliente en una escala del 1 al 100, que refleja el comportamiento de compra.

A continuación se muestran las variables utilizadas por el predict del automodel.

A screenshot of a computer

Description automatically generated

Se muestran los resultados de la predicción.

A screenshot of a data

Description automatically generated

Podemos observar como el gender es aquella variable que no puede ser significativa, además se muestran los resultados de correlaciones.

A screenshot of a computer

Description automatically generated

Conclusiones

El análisis nos arroja información de la relación que tienen las variables, podemos observar como el genero es la variable que no puede ser considerada para nuestro modelo, podemos observar como la edad es aquella variable que mas se relaciona con el spending score. El data set nos da para trabajar mucho ya sea en métodos predictivos o de agrupamiento, los cuáles pueden ser explotados por Rapid Miner.

**Referencias**

* Anáhuac Online. (2019). *Aplicando el caso y data understanding*[Archivo de video]. [Contenido creado para Anáhuac Online].