# 

# 

Analítica de datos y herramientas de inteligencia artificial

# 

# **Actividad 4 (Extracción de Características)**

# **Team 5**

José Ignacio hernández Rodríguez A01703130

# Tecnológico de Monterrey, Campus Querétaro

# 28/04/2023

El dataset sobre el cuál se realizó un análisis univariado, es el dataset micro retailer. Mismo que fue proporcionado por el profesor. Dentro de este dataset se encuentra información sobre ciertos procesos de gestión y administración que realizan algunas compañías de diferentes industrias. Dentro del dataset se encuentran en su mayoría variables categóricas, sobre las cuales se realizará el análisis. Es importante mencionar que dentro del dataset se encuentran muchos valores nulos, por lo que es importante trabajar sobre ellos, mediante distintas técnicas de modificación.

El proceso que se realizó para llevar a cabo el análisis univariado consistió en los siguientes pasos:

* Importar el dataset.
* Realizar una exploración del mismo para identificar columnas relevantes y valores nulos.
* Filtrar en el dataset original, las columnas que se van a analizar más adelante.
* Sustituir los valores nulos de las variables seleccionadas. Principalmente se sustituyeron por “Sin registro”.
* Realizar un análisis univariado de cada una de las columnas filtradas.
* Representar mediante un gráfico el análisis univariado de las columnas.

Resultados:

En cuanto a los resultados, se identifica que de las variables seleccionadas, en su gran mayoría la categoría “Sin registro” es la que predomina debido a la falta de información registrada en el dataset. De igual forma se identifica que dentro de las tablas de frecuencia de cada variable, las frecuencias comenzaban muchas veces en 1 y terminaban en valores mayores. Es por esto que de cada tabla de frecuencia se realizaba un filtro para elegir solamente aquellos resultados que fueran superiores a la frecuencia que decidimos utilizar. Esto se hace con la finalidad de que en la visualización, las gráficas no cuenten con tanta información que sea poco relevante y que además impida una buena interpretación de los resultados.

Las gráficas que se muestran en el código, son en su totalidad gráficas de pastel. Ya que en mi opinión, este tipo de gráficas permiten una visualización y entendimiento más claro de los datos.