



Tecnológico de Monterrey

Generación de los modelos de análisis de datos (Customers)

Escuela: Instituto Tecnológico de Estudios Superiores de Monterrey

Materia: Desarrollo de proyectos y análisis de datos

Profesor: Alfredo García Suárez

Nivel Académico: Profesional

Ciudad: Puebla

Autores

Omar Eduardo Pelcastre Reyes

Saúl Jesús Cuervo Méndez

Juan José Lara García

Cristian Marino Gutiérrez Jiménez

Kevin Vergara Lara

Marco Ivan Olalde Gonzalez

A01735985@tec.mx

A01735937@tec.mx

A01736667@tec.mx

A01736337@tec.mx

A01735970@tec.mx

A01733378@tec.mx

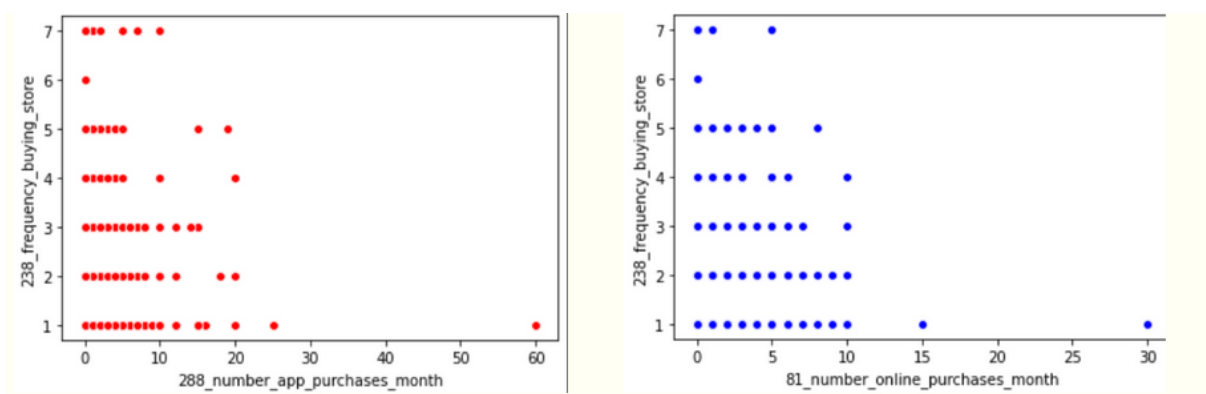
Etapa 3. Generación de modelos de análisis de datos.

Respecto a las regresiones obtenidas con los datos de consumidores obtuvimos resultados lejos de lo ideal, en el caso de las regresiones lineales se utilizaron las únicas variables numéricas que estaban presentes en la base de datos. Estas variables representan cantidades monetarias que los clientes estaban dispuestos a pagar por entregas inmediatas o en el mismo día.

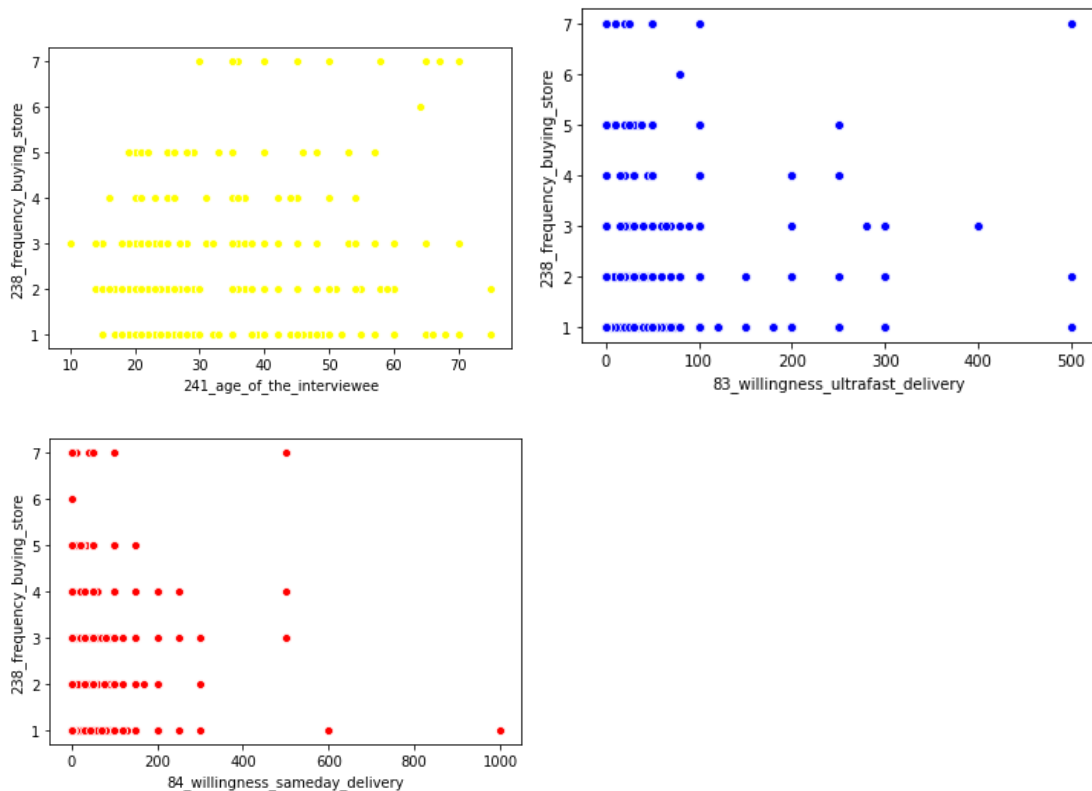
Otra de las variables numéricas utilizadas como variables independientes o predictoras fue la edad de los consumidores. La variable dependiente o de respuesta que utilizamos fue la cantidad de veces que los clientes visitaban la tienda. Consideramos que esta variable es la que nos puede ayudar a analizar qué factores podrían mejorar las microempresas. Desafortunadamente las variables independientes no tienen una relación con la variable dependiente, por lo que desde los gráficos de scatterplot se puede observar que no existe un comportamiento lineal.

Otro de los casos que hicimos de regresión lineal fue utilizando la cantidad de veces que los clientes realizan compras por medio de internet o alguna aplicación de mensajería como whatsapp. En este caso tampoco se encontró una correlación pues los valores de r y r^2 igual son muy pequeños.

En el primer caso de regresión se utilizaron como variables predictoras o independientes la cantidad de veces que los clientes realizan compras por medio de una app, en línea y por medio de whatsapp.



En segundo caso se usaron variables que representan la cantidad de dinero que el cliente está dispuesto a pagar entregas inmediatas o entregas el mismo día. En las gráficas de scatter plot se observa que igual, no existe una relación lineal entre estas variables y la cantidad de veces que el cliente va a la tienda.



Se puede observar que no existe un comportamiento lineal para estas variables por lo tanto los valores de r y r^2 de ambos modelos es muy bajo. Para el primero es de apenas 0.15 y 0.0225 respectivamente. Para el segundo modelo son de 0.12 y 0.0144

Por otro lado en cuanto a las regresión logística se utilizaron las mismas variables que el caso anterior, solo que en esta ocasión la variable dependiente es una variable dicotómica que representa si los clientes encontraron lo que buscaban en la tienda. En estas regresiones es importante recalcar que ambos modelos tuvieron buena exactitud, precisión y sensibilidad. Esto indica que estos modelos son más útiles y correctos que los de regresión lineal para analizar este tipo de variables.