**Actividad 5 (Extracción de Características)**

**Analítica de datos y herramientas de inteligencia artificial II (Gpo 501)**

IIS



**PROFESOR:**

Alfredo García Suárez

**PRESENTA:**

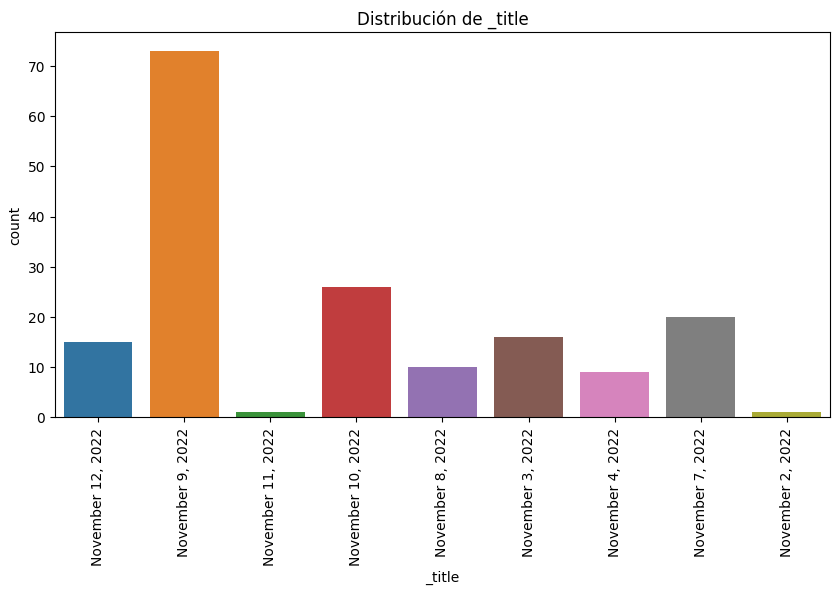
Ricardo Saldaña Muñoz | A01707504

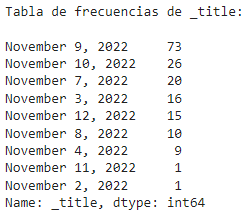
14 de octubre de 2023

**Campus Querétaro**

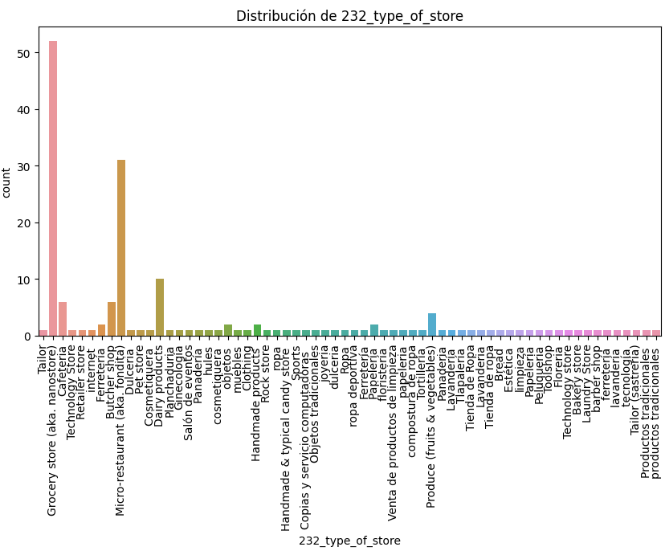
**Introducción**

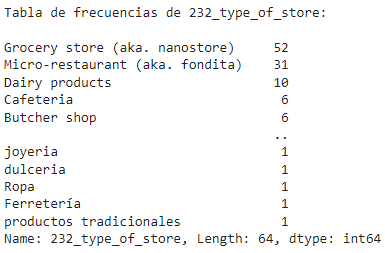
En el presente documento se documentarán los hallazgos obtenidos de cada una de las variables categóricas analizadas del dataset microretailer\_mit\_lift\_lab.xlsx.

***Variable \_title***

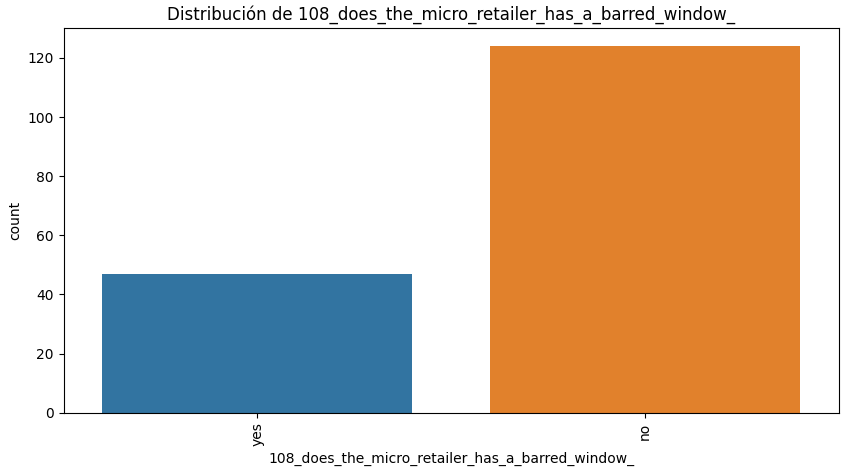
******

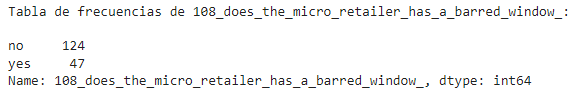
De esta variable no se contaban con ningún valor nulo, y nos muestra la fecha. Podemos observar que más del 50% pertenecen a las fechas del 9 y 10 de noviembre del 2022; teniendo 99 datos de 171 en total, mientras que las fechas de noviembre 2 y 11 de 2022 solamente tienen un registro respectivamente.

***Variable 232\_type\_of\_store***

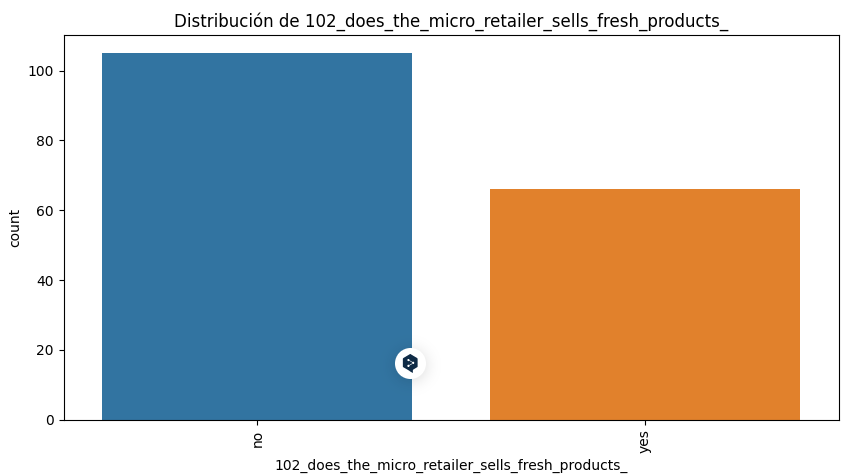


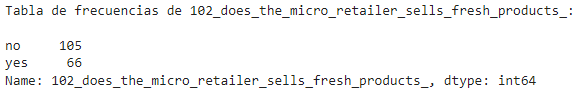
De esta variable no se contaban con ningún valor nulo, y nos muestra el tipo de tienda. Podemos observar que los primeros 3 registros los cuales son Grocery store, Micro-restaurant y Dairy products tienen una gran cantidad de frecuencia en el dataset ya que después la mayoría de los registros se mantiene por debajo de 6 registros.

 ***Variable 108\_does\_the\_micro\_retailer\_has\_a\_barred\_window\_***

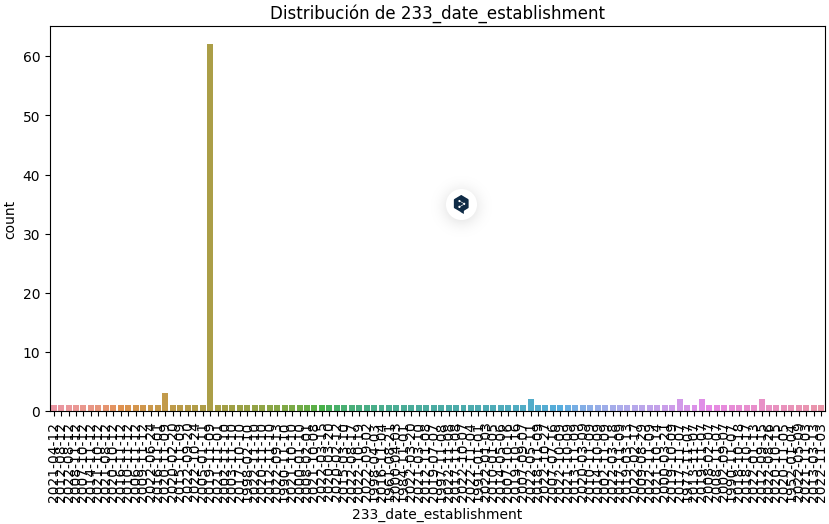
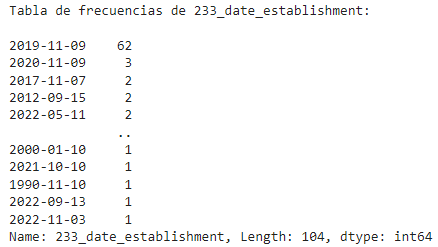


De esta variable no se contaban con ningún valor nulo, y nos indica si el minorista tiene una ventana con rejas. Podemos observar que el 72% de los minoristas no cuentan con una ventana con rejas, mientras que el 28% sí.

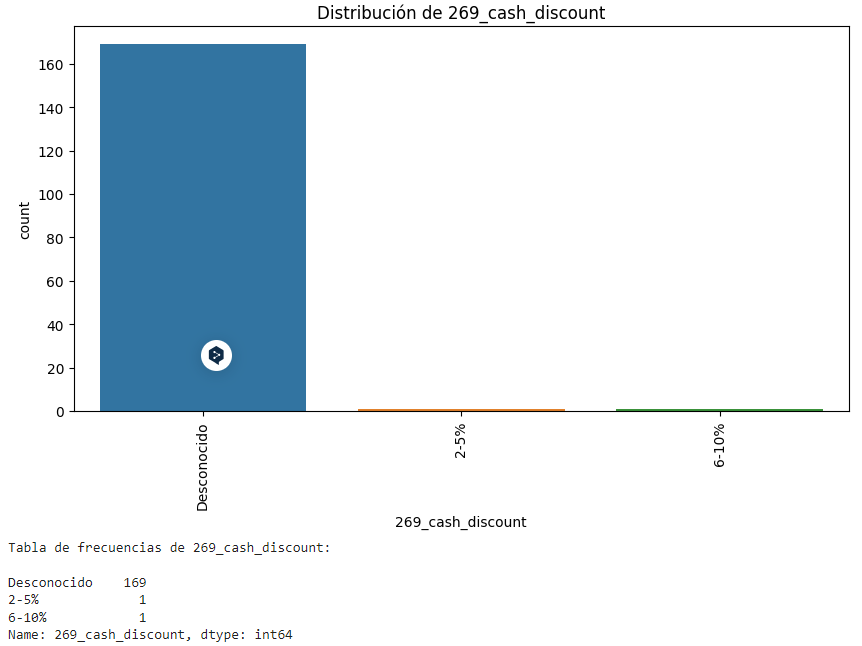
***Variable 102\_does\_the\_micro\_retailer\_sells\_fresh\_products\_***

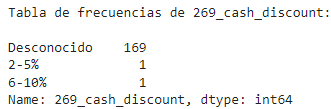


De esta variable no se contaban con ningún valor nulo, y nos indica si el minorista vende productos frescos. Podemos observar que 105 de 171 minoristas no venden sus productos frescos.

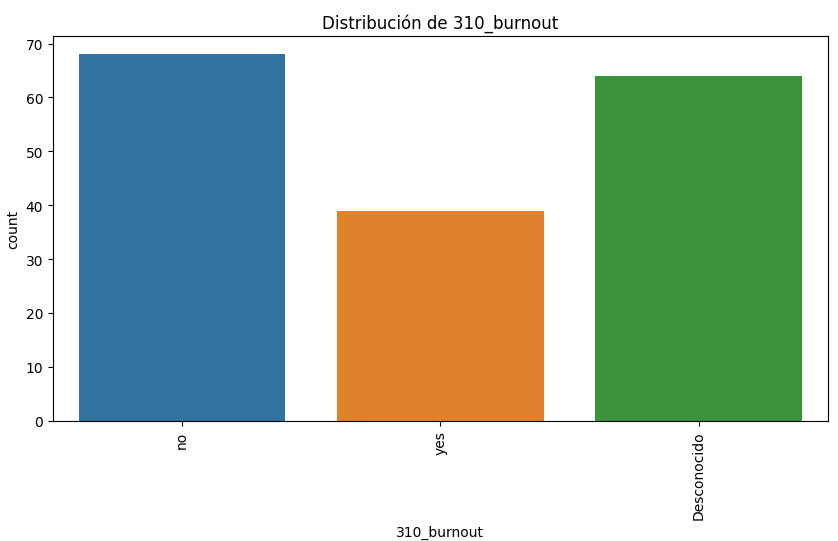
***Variable 233\_date\_establishment***

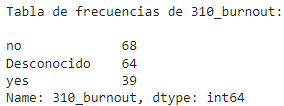
De esta variable se contaban con 59 valores nulos y se trataron con la mediana, y nos indica la fecha de establecimiento de la tienda. Podemos observar que 62 tiendas ocupan el 36.2% de los datos siendo en el día del 2019-11-09, y todos los demás registros contienen de 2 registros para abajo, a excepción del 2020-11-09 que contiene 3 registros.

***Variable 269\_cash\_discount***

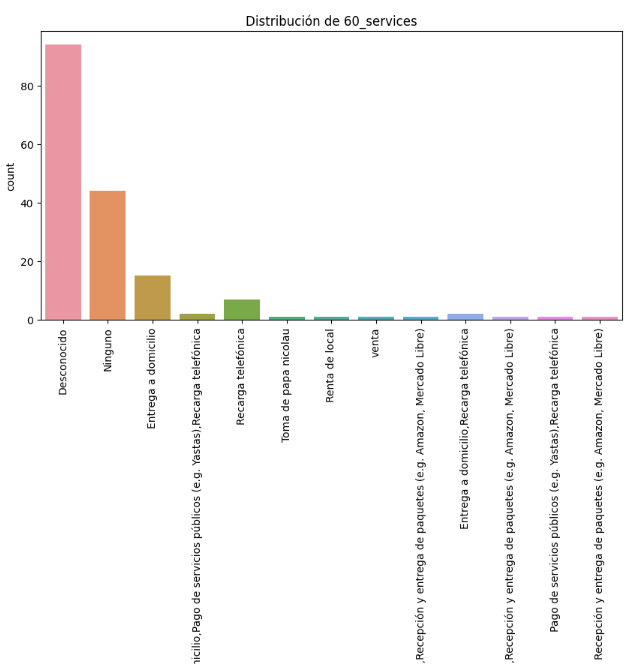


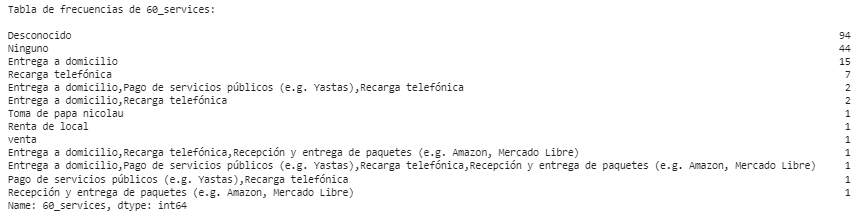
De esta variable se contaban con 169 de 171 valores nulos y se trataron, definiendo los valores con el texto “Desconocido” ya que no sería adecuado hacerlo con algún método ya que solo se contaban con 2 registros. Esta variable nos indica el descuento por pago en efectivo y además es posible que todos aquellos que no contaban con registro fuera de 0, pero al tenerlo como object se puso con Desconocido.

***Variable 310\_burnout***

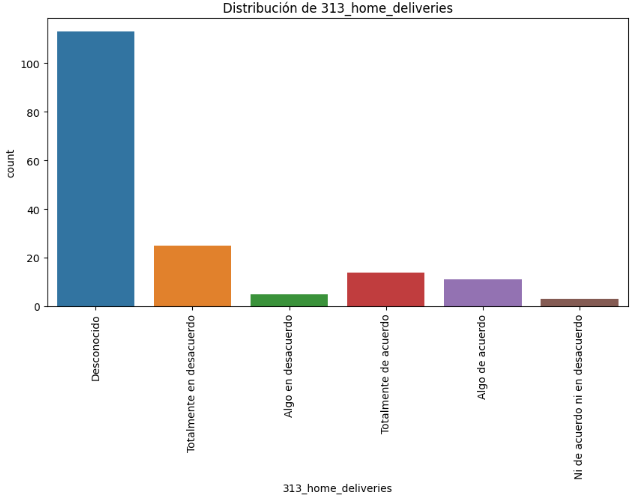
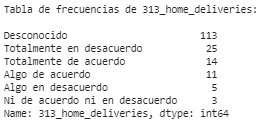


De esta variable se contaban con 64 de 171 valores nulos y se trataron, definiendo los valores con el texto “Desconocido” ya que no sería adecuado hacerlo con algún método ya que se contaban con menos de la mitad de los datos. Esta variable indica información sobre el agotamiento o burnout. Y vemos que 68 registros muestran que no, 39 que sí, y 64 registros son desconocido.

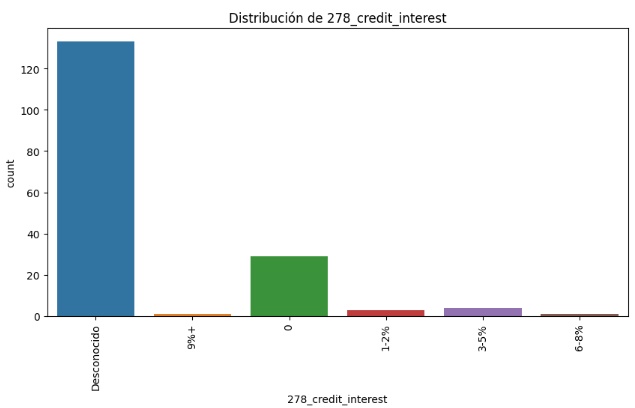
***Variable 60\_services***

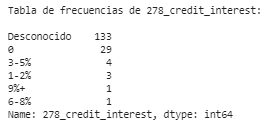


De esta variable se contaban con 94 de 171 valores nulos y se trataron, definiendo los valores con el texto “Desconocido” ya que no sería adecuado hacerlo con algún método ya que se contaban con menos de la mitad de los datos. Esta variable indica servicios ofrecidos por la tienda. Y vemos que 44 tiendas no cuentan con ningún servicio, 15 servicio a domicilio, 7 recarga telefónica y los demás registros contaban con menos de 2 registros por servicio a excepción de los que fueron desconocidos con un total de 94 registros.

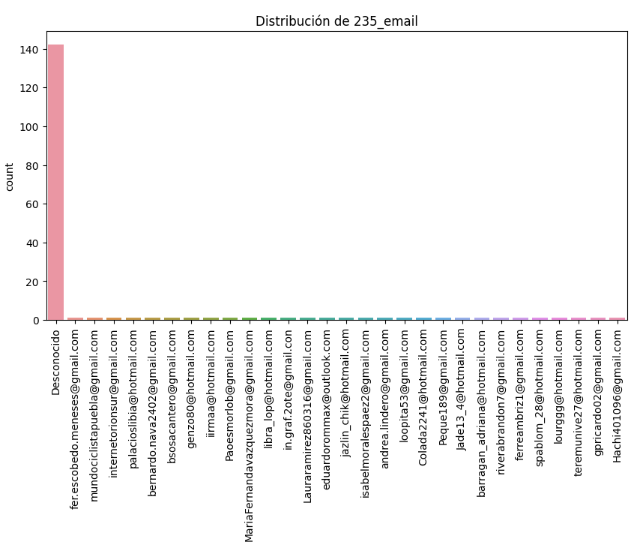
*********Variable 313\_home\_deliveries***

De esta variable se contaban con 113 de 171 valores nulos y se trataron, definiendo los valores con el texto “Desconocido” ya que no sería adecuado hacerlo con algún método ya que se contaban con menos de la mitad de los datos. Esta variable indica si están de acuerdo con entregas a domicilio. Y vemos 113 tiendas no se conoce su respuesta, 25 están en desacuerdo y 14 en acuerdo, mientras que los demás se podría decir que se mantienen un poco neutros.

***Variable 278\_credit\_interest***

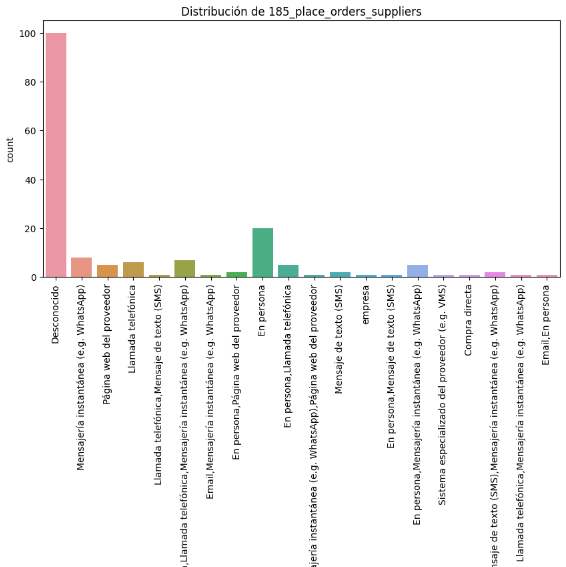


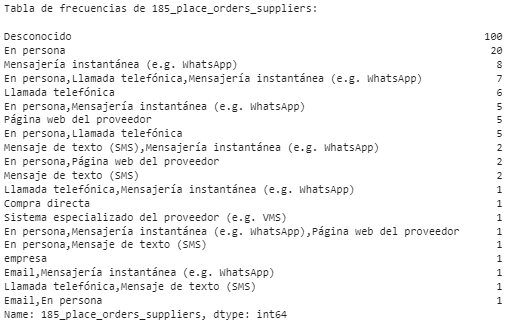
De esta variable se contaban con 133 de 171 valores nulos y se trataron, definiendo los valores con el texto “Desconocido” ya que no sería adecuado hacerlo con algún método ya que se contaban con menos de la mitad de los datos. Esta variable indica los intereses de crédito. Y vemos 133 tiendas no se conoce su respuesta, 29 no hay interés de crédito, 8 registros se encuentran entre el 3% y 9% y solo 1 arriba del 9%.

***Variable 235\_email***

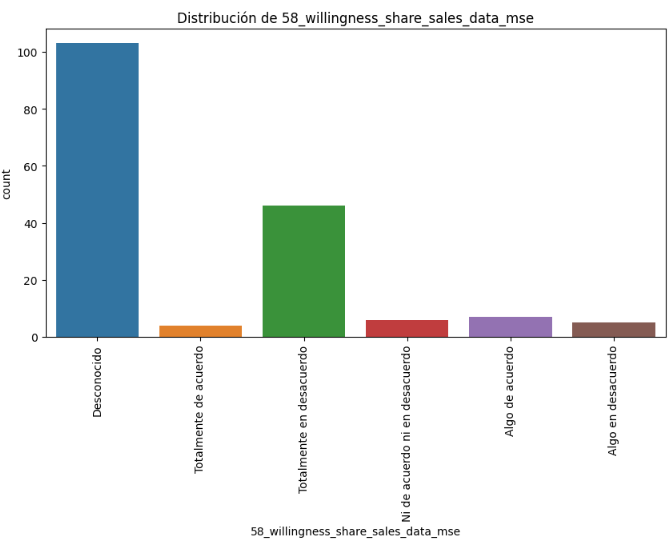


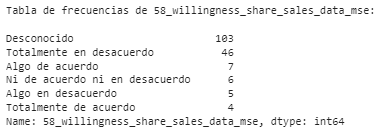
De esta variable se contaban con 142 de 171 valores nulos y se trataron, definiendo los valores con el texto “Desconocido” ya que no sería adecuado hacerlo con algún método ya que se contaban con menos de la mitad de los datos. Esta variable indica la dirección de correo electrónico. Y vemos que 142 son desconocidos mientras que los demás son valores únicos.

***Variable 185\_place\_orders\_suppliers***

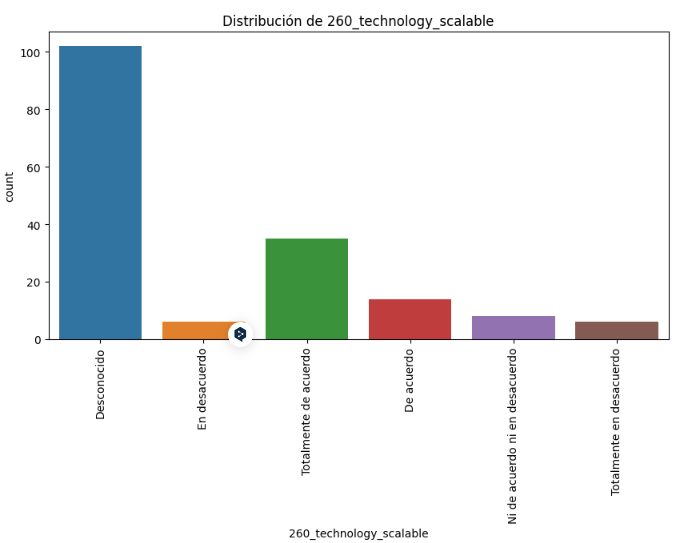


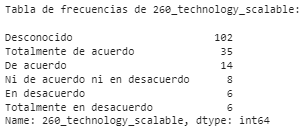
De esta variable se contaban con 100 de 171 valores nulos y se trataron, definiendo los valores con el texto “Desconocido” ya que no sería adecuado hacerlo con algún método ya que se contaban con menos de la mitad de los datos. Esta variable indica la colocación de pedidos con proveedores. Y vemos que 100 son desconocidos, 20 en persona, 8 en mensajería instantánea, y los registros faltantes tienen menos de 7 registros.

***Variable 185\_place\_orders\_suppliers***

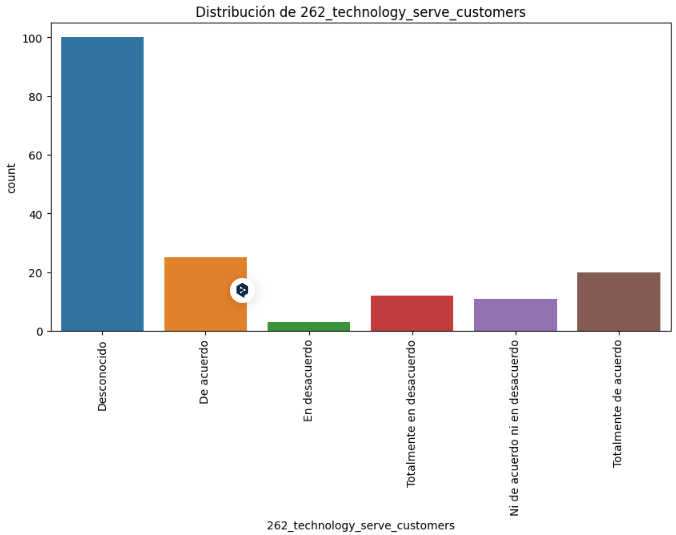


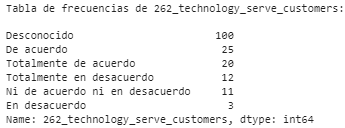
De esta variable se contaban con 103 de 171 valores nulos y se trataron, definiendo los valores con el texto “Desconocido” ya que no sería adecuado hacerlo con algún método ya que se contaban con menos de la mitad de los datos. Esta variable indica la disposición a compartir datos de ventas con la MSE (Micro and Small Enterprises). Y es posible observar que de 103 registros se desconocía la respuesta, 46 están en desacuerdo y solo 4 están en total acuerdo, mientras que el resto se podría decir que están en un acuerdo neutro.

***Variable 260\_technology\_scalable***

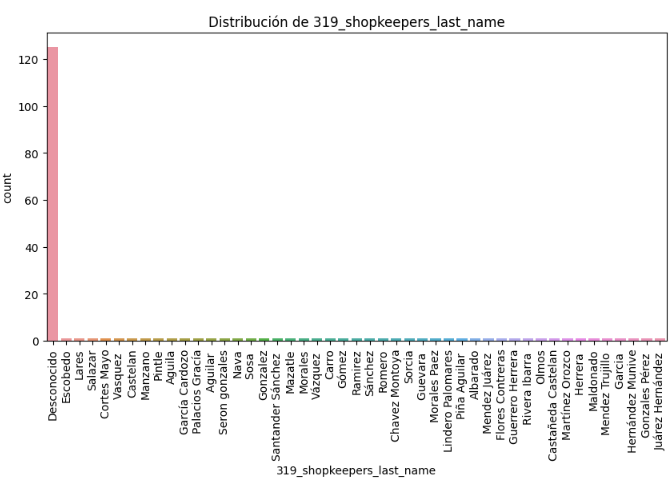


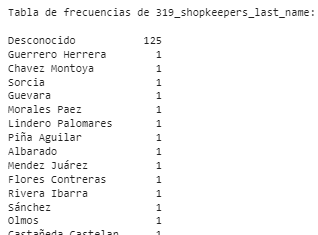
De esta variable se contaban con 102 de 171 valores nulos y se trataron, definiendo los valores con el texto “Desconocido” ya que no sería adecuado hacerlo con algún método ya que se contaban con menos de la mitad de los datos. Esta variable indica si están de acuerdo con la tecnología escalable. Y es posible observar que de 102 registros se desconocía la respuesta, 35 están de acuerdo y solo 6 en desacuerdo, mientras que el resto se podría decir que están en un acuerdo neutro.

***Variable 260\_technology\_scalable***

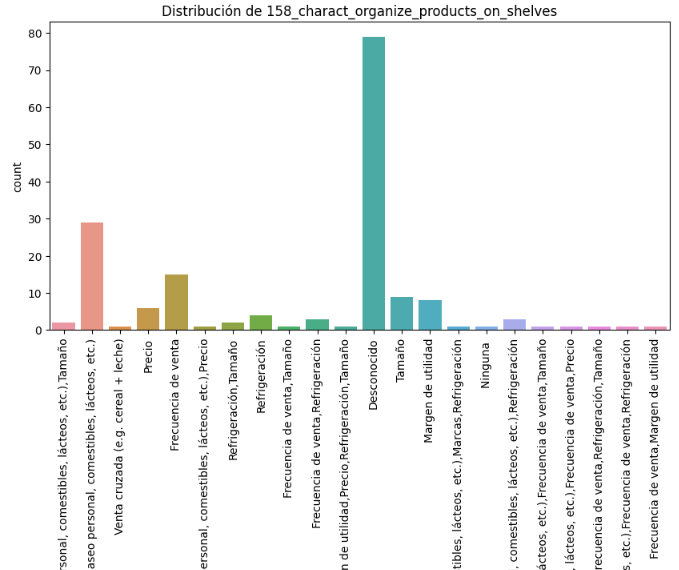


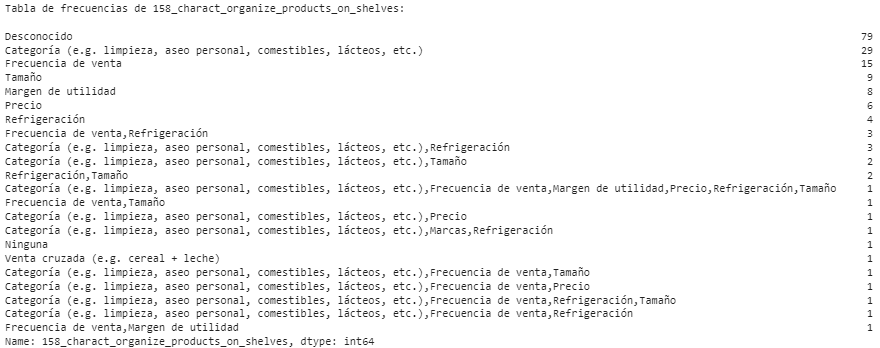
De esta variable se contaban con 100 de 171 valores nulos y se trataron, definiendo los valores con el texto “Desconocido” ya que no sería adecuado hacerlo con algún método ya que se contaban con menos de la mitad de los datos. Esta variable indica si están de acuerdo con que se utilice la tecnología para servir a los clientes. Y es posible observar que de 100 registros se desconocía la respuesta, 45 están de acuerdo y 15 más apegados a que se encuentran en desacuerdo, mientras que el resto se podría decir que están en un acuerdo neutro.

***Variable 260\_technology\_scalable***

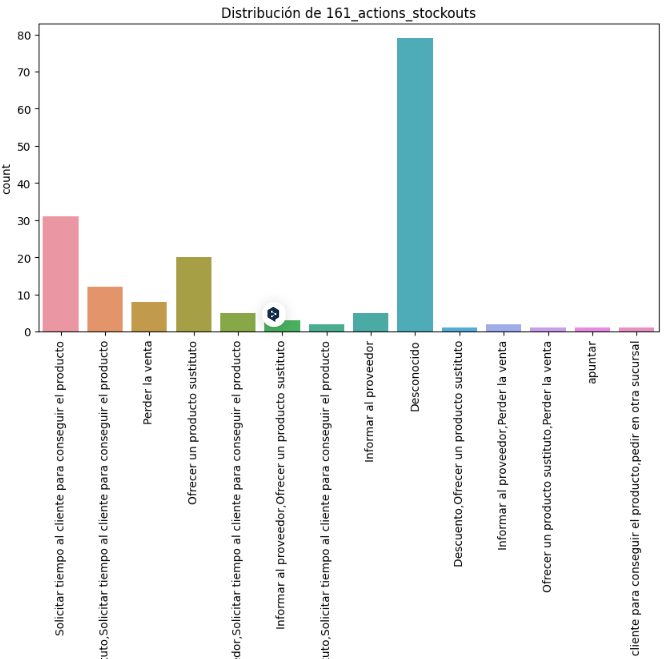
******

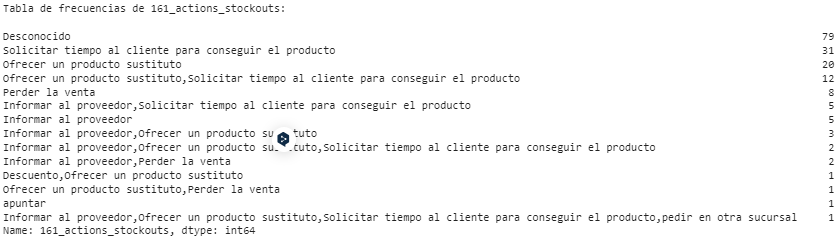
De esta variable se contaban con 125 de 171 valores nulos y se trataron, definiendo los valores con el texto “Desconocido” ya que no sería adecuado hacerlo con algún método ya que se contaban con menos de la mitad de los datos. Esta variable indica el apellido de tendero. Y vemos que 125 son desconocidos mientras que los demás son valores únicos.

***Variable 158\_charact\_organize\_products\_on\_shelves***

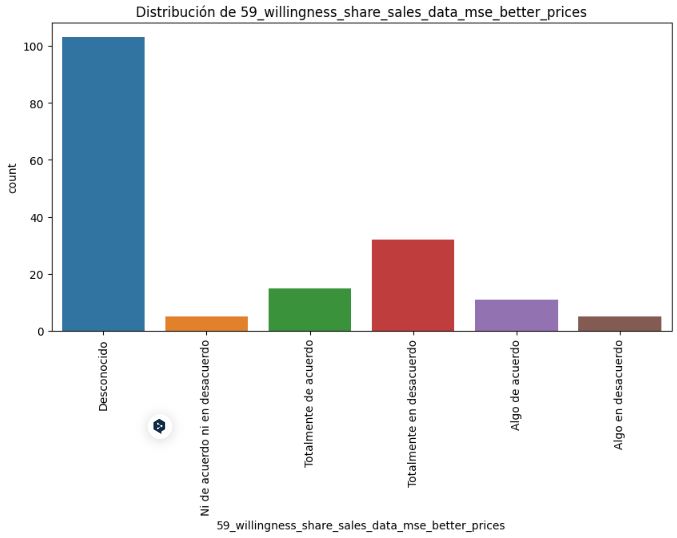


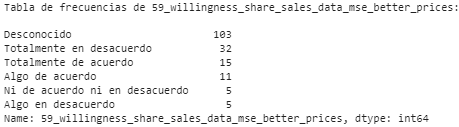
De esta variable se contaban con 79 de 171 valores nulos y se trataron, definiendo los valores con el texto “Desconocido” ya que no sería adecuado hacerlo con algún método ya que se contaban con menos de la mitad de los datos. Esta variable indica las características de la organización de productos en estantes. En esto podemos observar que 79 datos son desconocidos, 29 son de categoría, 15 frecuencia de venta, tamaño 15 y los demás bajan de 9 registros.

***Variable 161\_actioons\_stockouts***

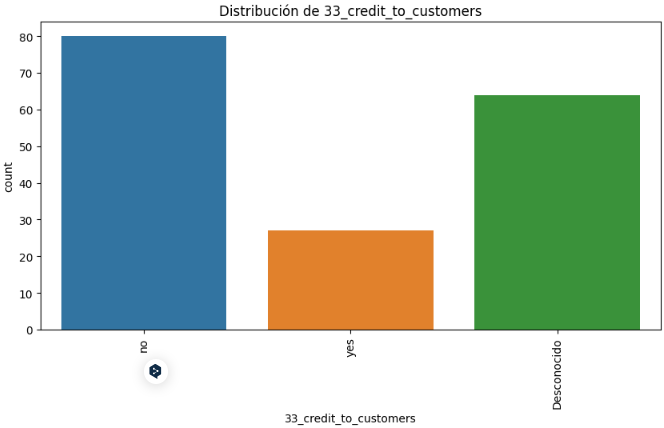


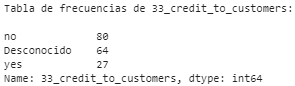
De esta variable se contaban con 79 de 171 valores nulos y se trataron, definiendo los valores con el texto “Desconocido” ya que no sería adecuado hacerlo con algún método ya que se contaban con menos de la mitad de los datos. Esta variable indica las acciones para abordar faltantes de stock. En esto podemos observar que 79 datos son desconocidos, 31 solicitan tiempo al cliente para registrar el producto, 51 se basan principalmente en ofrecer un producto sustituto, 12 pierden la venta y las demás características bajan de 8 registros.

***Variable 59\_willingness\_share\_sales\_data\_mse\_better\_prices***



De esta variable se contaban con 103 de 171 valores nulos y se trataron, definiendo los valores con el texto “Desconocido” ya que no sería adecuado hacerlo con algún método ya que se contaban con menos de la mitad de los datos. Esta variable indica si están de acuerdo con que se utilice la tecnología para servir a los clientes. Y es posible observar que de 103 registros se desconocía la respuesta, 32 están en desacuerdo, 15 en acuerdo y los demás se podría decir que se encuentran en un punto neutro.

***Variable 33\_credit\_to\_customers***



De esta variable se contaban con 64 de 171 valores nulos y se trataron, definiendo los valores con el texto “Desconocido” ya que no sería adecuado hacerlo con algún método ya que se contaban con menos de la mitad de los datos. Esta variable indica si le dan crédito a los clientes. Y vemos que 80 registros muestran que no, 27 que sí, y 64 registros son desconocido.