

#### Actividad 4: Extracción de Características

Se trabajó con la base de datos “microretailer\_mit\_lift\_lab.xlsx”, la cual consiste en un dataframe de 106 columnas con 171 entradas o filas con variables de tipo flotante, entero y objeto. Para el análisis de esta actividad se seleccionaron 20 columnas categóricas y realizarles análisis univariado a cada una de ellas.

Para ello, primero se limpiaron los datos sustituyendo primero los valores nulos por la mediana, en el caso de las variables numéricas y por “desconocido” en el caso de variables cualitativas; para posteriormente tratar los valores atípicos de las columnas cuantitativas, que en esta ocasión fueron tres variables de las veinte que se seleccionaron.

##### Análisis univariado

###### 1. 186\_internet\_connection

Esta variable presentaba una gran cantidad de nulos, por ello es por lo que un 36.8% de los datos son “desconocido”; sin embargo, se dice que es una variable tipo booleana ya que solo tiene dos respuestas, si o no. Aun así, se observa que la mayoría de los negocios si cuentan con internet dentro de su establecimiento, al menos la mitad de la población.

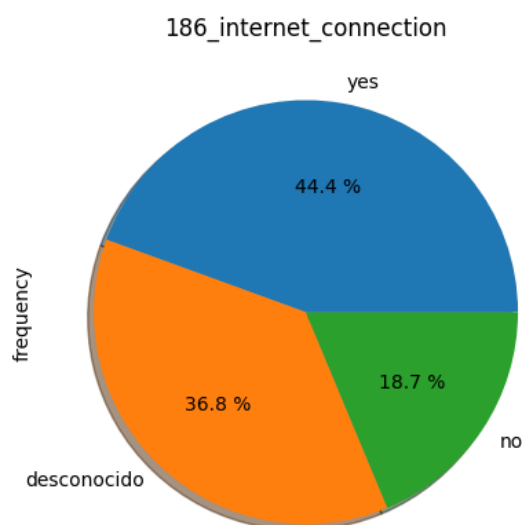
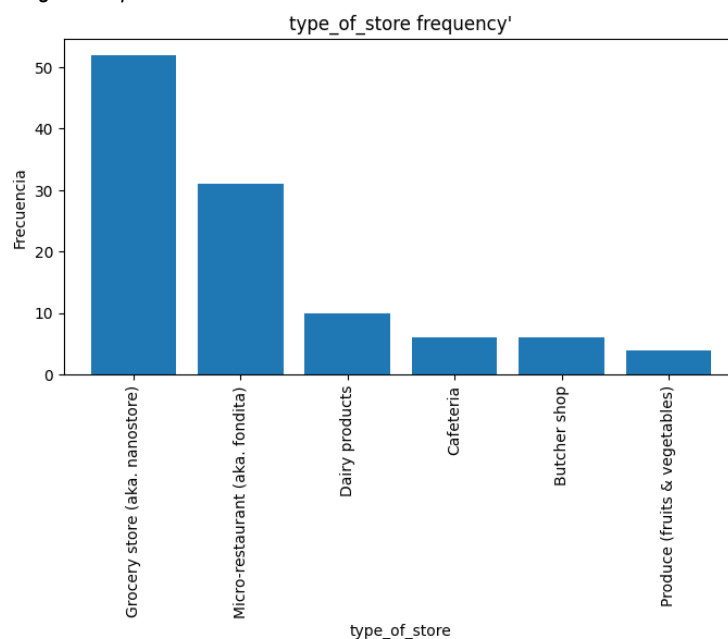


Imagen 1.

## 2. 232\_type\_of\_store

Esta columna describe los tipos de tiendas que tienen los micro\_retailers, de un total de 64 categorías, se disminuyó a solo 6 categorías con base a las que tuvieran más de 3 registros. Como se observa en la imagen 2, las tiendas que más frecuentes son: Grocery store (nanostore, micro-restaurant (fondita), dairy products, cafetería, butcher shop y Produce. Por lo que se puede decir que las tiendas más recurrentes son las relacionadas a comida, además, de ser las tienditas más pequeñas las que más presencia tienen.

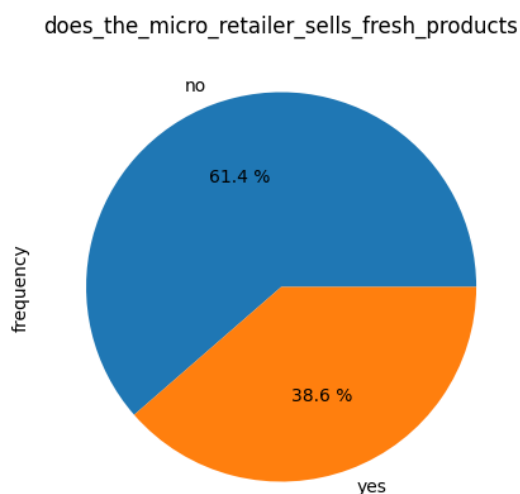
Imagen 2. Tipos de tiendas



## 3. 102\_does\_the\_micro\_retailer\_sells\_fresh\_products\_

Esta columna a diferencia de la primera no contaba con valores nulos, por lo que su análisis es mucho más sencillo. Describe si los micro\_retailers venden o no productos frescos y se observa que la gran mayoría con un 61.4% no vende; a pesar de que en la variable anterior se observó que las tiendas más recurrentes son las que tienen que ver con comida.

Imagen 3. Vende productos frescos



#### 4. 184\_store\_devices

Esta columna hace referencia a los dispositivos tecnológicos con los que cuenta el negocio. Originalmente la variable contaba con 39 categorías; sin embargo, dentro de todas esas categorías, se repetían varias si es que el negocio contaba con más de un dispositivo. Por ello, se decidió mostrar en primer lugar aquellas categorías que tuvieran más de 5 registros y posteriormente agruparlas por sus repeticiones. Obteniendo así, que una gran parte de los negocios cuentan con al menos un teléfono celular, ya que juntando todas las repeticiones arroja que al menos es un 40% de la población; asimismo, en una menor cantidad, pero existen retailers que también cuentan con caja registradora, computadoras o scanner, y cámaras de seguridad.

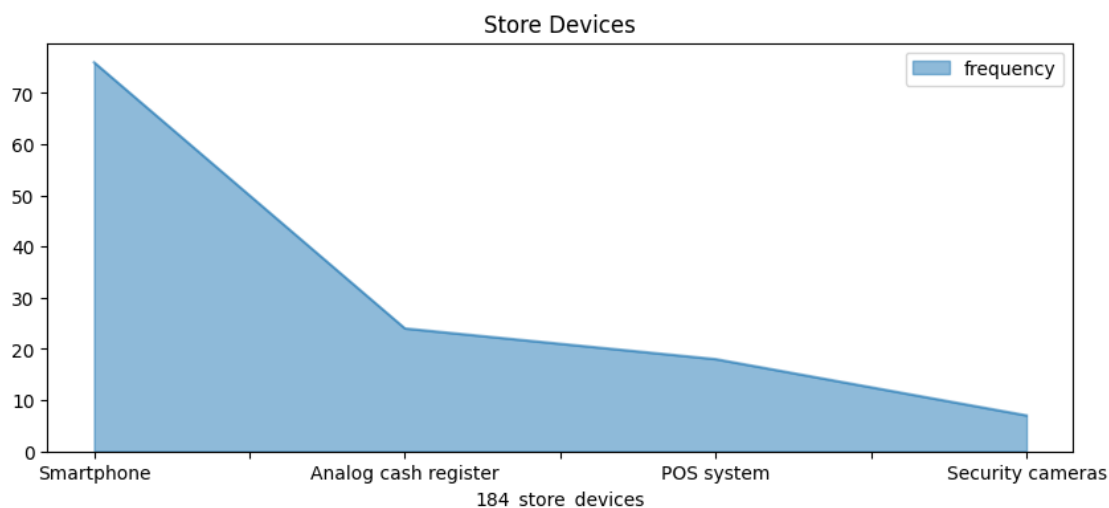


Imagen 4

5. 2\_current\_permanent\_employees

La siguiente tabla habla de los empleados permanentes actuales en los negocios, al ser una variable tipo flotante se le trataron los valores nulos, así como los datos atípicos. Al solo tener 4 categorías, se dejaron todas.

Se observa que un 63% de los negocios solo cuentan con un empleado permanente, el máximo de empleados es de 3 y en otros casos no hay empleados. Lo cual llega a tener sentido al saber que trata sobre microempresas.

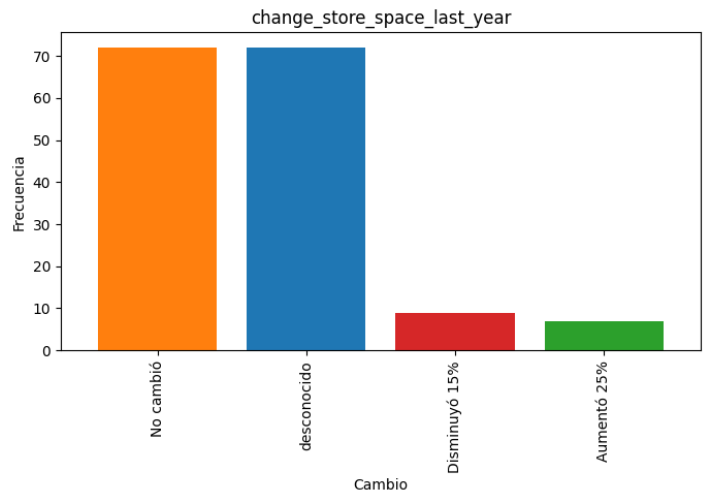
	frequency	percentage	cumulative_perc
2_current_permanent_employees			
1.0	108	0.631579	0.631579
2.0	26	0.152047	0.783626
0.0	26	0.152047	0.935673
3.0	11	0.064327	1.000000

Tabla 1.

6. 5\_change\_store\_space\_last\_year

Esta variable describe si el espacio de la tienda cambió en el ultimo año, haciendo referencia a las categorías de: no cambió, aumentó o disminuyó en diferentes porcentajes. Para el análisis se seleccionó aquellas que tuvieran más de 5 repeticiones, obteniendo que en la mayoría no hubo ningún cambio y que en otra gran cantidad existen datos desconocidos.

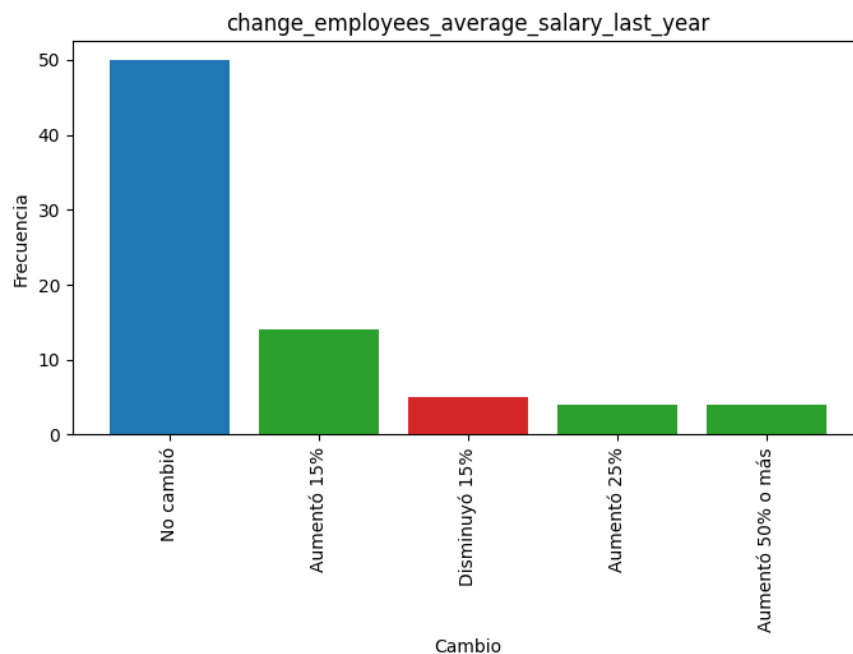
Imagen 5. Cambio en el espacio



#### 7. 6\_change\_employees\_average\_salary\_last\_year

Esta variable parecida a la anterior tiene categorías que refieren a un porcentaje de cambio, en este caso si el promedio del salario de los empleados cambió, con la diferencia que originalmente contaba con 12 categorías. Para este análisis de filtro por aquellas observaciones con más de 3 apariciones. Aunque el 50% de los datos son desconocidos, para fines visuales se quitó esa categoría. Teniendo como resultado que un 30% de los salarios no cambió, pero mínimamente en otros negocios aumentó, como se observa en la imagen 6.

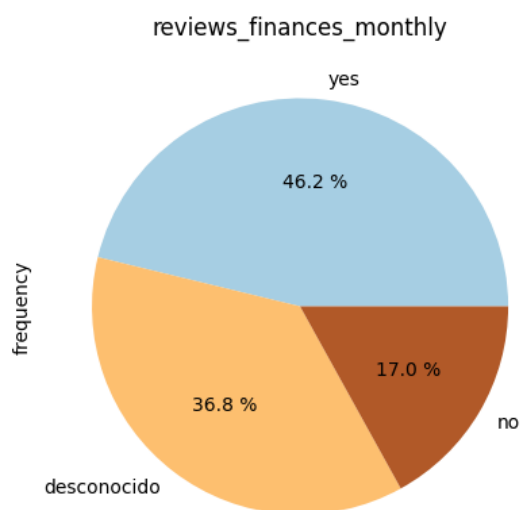
Imagen 6. Cambio en el salario en el último año



#### 8. 20\_reviews\_finances\_monthly

La imagen 7 representa la proporción de los negocios que cada mes revisan sus finanzas, como se observa es una variable que arrojó una gran cantidad de datos nulos; sin embargo, también se observa que la mayoría de los empresarios si revisan mes con mes sus finanzas, y solo un 17% no lo hace.

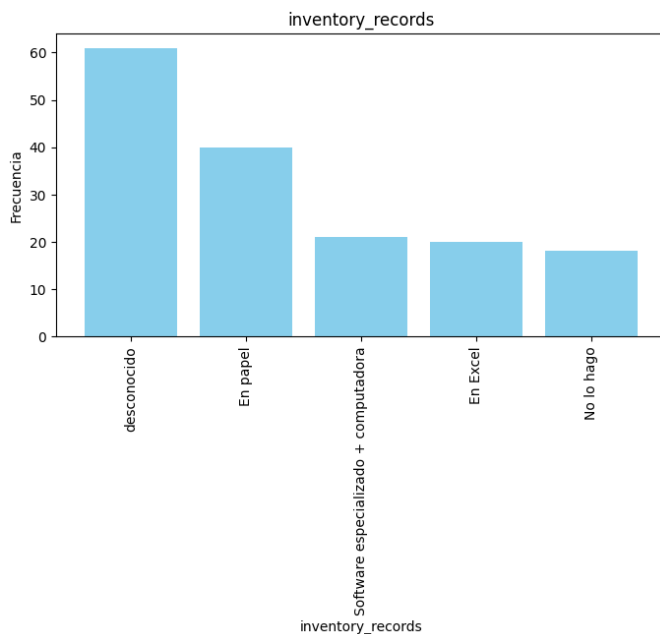
Imagen 7. Chequeo de finanzas mensual



#### 9. 49\_inventory\_records

Esta variable representa cómo registran el inventario en cada uno de los negocios, originalmente se contaba con 10 categorías entre ellas algunas repetidas si es que el retailer cuenta con más de una manera de registro. Se filtró el dataframe por aquellas observaciones que su frecuencia fuera igual o mayor a 10. Como se observa en la imagen 8, una gran parte de la población registra su inventario en papel o incluso no lo hace, y en casi la misma medida se ayudan de algo más tecnológico como Excel o computadora.

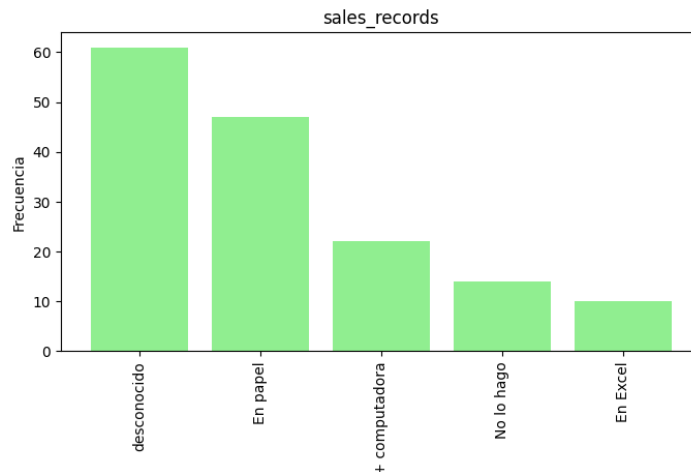
Imagen 8. Registro de inventario



#### 10. 18\_sales\_records

Igual que la variable anterior, esta mide cómo las microempresas registran sus ventas y también se filtró por aquellas frecuencias mayores o iguales a 10. Obteniendo que todavía una mayor parte lo hace en papel o no lo hace, pero una mayor parte las registra ya sea en computadora o con un software especializado.

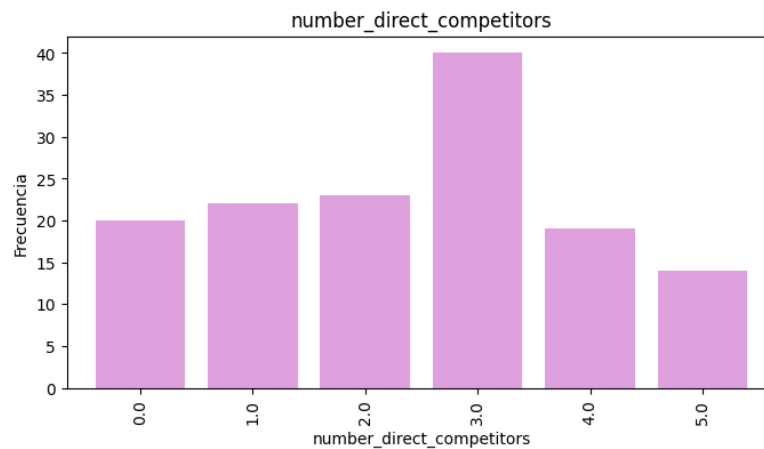
*Imagen 9. Sales records*



#### 11. 145\_number\_direct\_competitors

Esta variable al ser tipo flotante se cambió a tipo objeto para poder hacer el análisis y poder hacerla categórica. Mide el número de competidores directos y solo se muestra aquellas categorías que superan las 10 repeticiones. La imagen 10 representa aquellas que más se repitieron y se puede observar que una gran parte de los negocios tienen al menos 3 competidores directos, pero también existen algunos que tampoco cuentan con algún competidor.

*Imagen 10. Competidores directos*



## 12. 210\_sales\_channels

Esta variable representa los canales de venta, para el análisis se filtraron aquellas observaciones que aparecen más de 5 veces en el dataframe; aunque originalmente se contaba con 30 categorías, muchas de ellas se repiten como se observa en la tabla 2. Aun así, se muestra aquellos canales más populares sin contar que más del 50% de la información es nula; se observa que las microempresas prefieren llegar a sus clientes a través de redes sociales, incluyendo WhatsApp.

Tabla 2. Canales de venta

	frequency	percentage	cumulative_perc
210_sales_channels			
desconocido	89	0.520468	0.520468
Redes sociales (e.g. Facebook, Instagram)	11	0.064327	0.584795
WhatsApp	10	0.058480	0.643275
Redes sociales (e.g. Facebook, Instagram),WhatsApp	9	0.052632	0.695906
Llamada a Teléfono fijo	6	0.035088	0.730994
Llamada a Teléfono celular,WhatsApp	6	0.035088	0.766082

## 13. 189\_payment\_methods

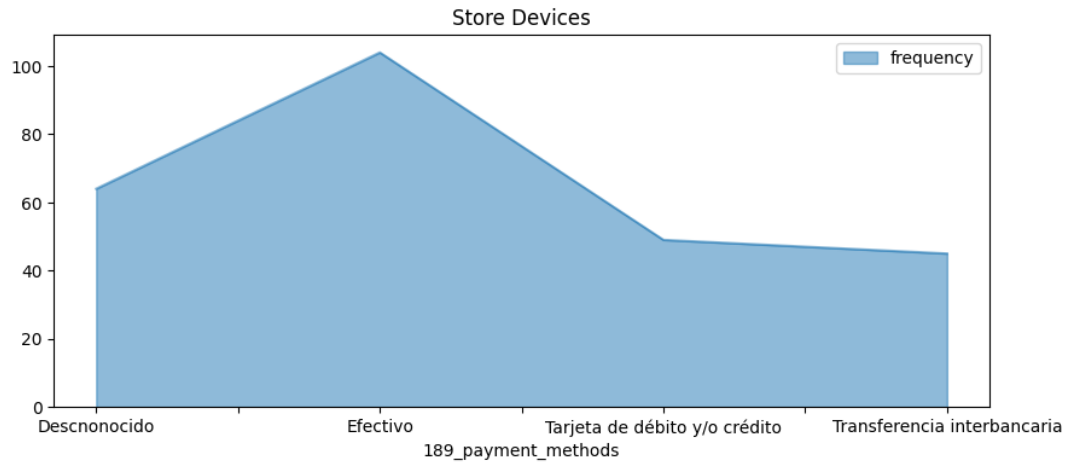
Así como en columnas anteriores, esta variable cuenta con diferentes categorías, pero algunas de ellas se repiten dentro de las mismas. Se seleccionaron aquellas categorías que cuentan con una frecuencia mayor a 5 para conocer aquellas más populares. Como se observa en la tabla 2 y en la imagen 11 el método de pago preferido es el efectivo por mucho y casi en la misma proporción las tarjetas o transferencias.

Tabla 3. Métodos de pago

	frequency	percentage	cumulative_perc
189_payment_methods			
desconocido	64	0.374269	0.374269
Efectivo	40	0.233918	0.608187
Efectivo,Tarjeta de débito y/o crédito,Transferencia interbancaria	30	0.175439	0.783626
Efectivo,Tarjeta de débito y/o crédito	19	0.111111	0.894737
Efectivo,Transferencia interbancaria	15	0.087719	0.982456



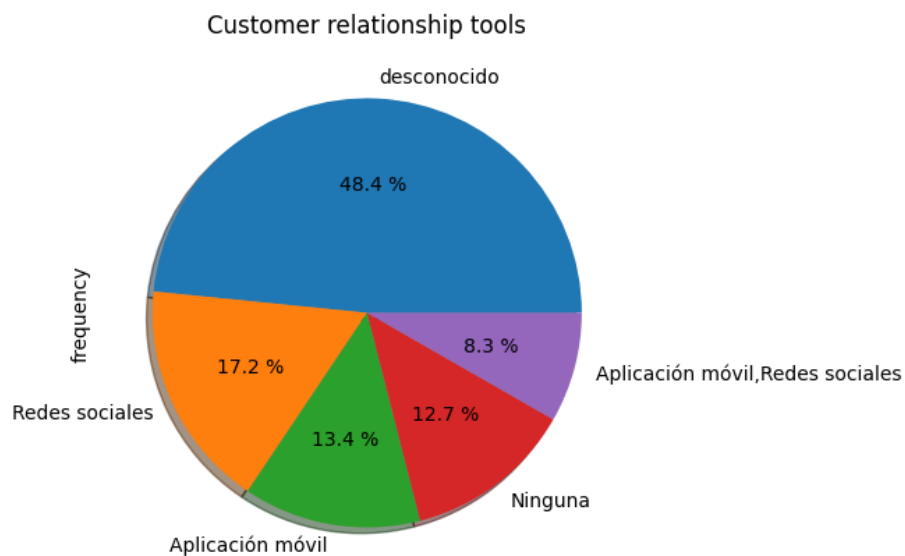
Imagen 11. Métodos de pago



#### 14. 214\_customer\_relationship\_tools

Esta variable describe aquellas herramientas con las que las empresas se relacionan con sus clientes. Originalmente esta columna contaba con 14 categorías, pero solo se seleccionaron aquellas que contaban con una repetición mayor a 10 para solo analizar las más populares. Aunque casi la mitad de la información se desconoce, las herramientas preferidas son las redes sociales, y aplicaciones móviles.

Imagen 12. Herramientas para relacionarse con el cliente.



15. 38\_average\_margin\_profits

Esta variable contaba originalmente con 11 columnas, pero se seleccionaron solo aquellas que se repitieran más de 10 veces. Aunque el 61% de la información se desconoce, en la tabla 4, se observa que solo el 8% de los microempresarios ganan 30% de utilidad.

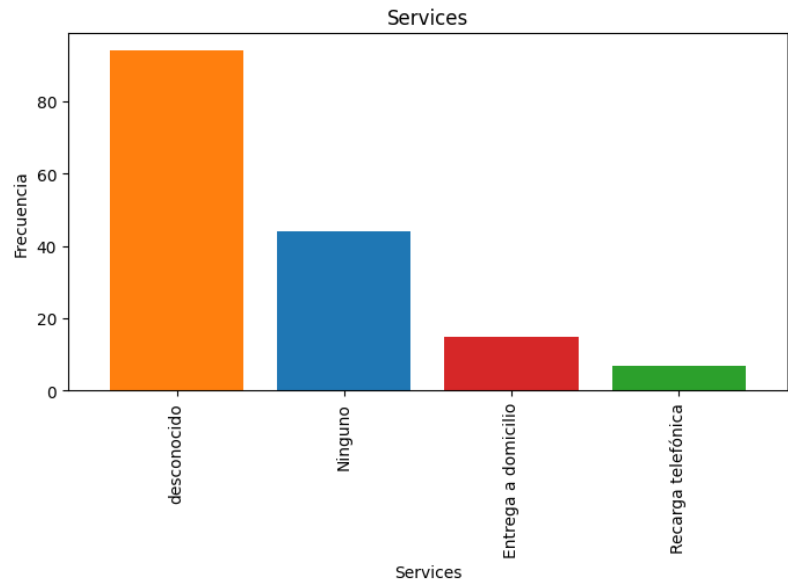
Tabla 4. Margen de utilidad

	frequency	percentage	cumulative_perc
38_average_margin_profits			
desconocido	105	0.614035	0.614035
30%	15	0.087719	0.701754

16. 60\_services

Esta variable contaba originalmente con 13 categorías, sin embargo, solo se seleccionaron aquellas que se repiten más de 5 veces, para así obtener aquellos servicios más populares entre las microempresas de retail. Aunque el 54% de la información es nula, en la imagen 13 se observa que gran parte de las empresas no ofrece ningún servicio adicional; sin embargo, los servicios más recurrentes son entrega a domicilio y recargas telefónicas.

Imagen 13. Servicios

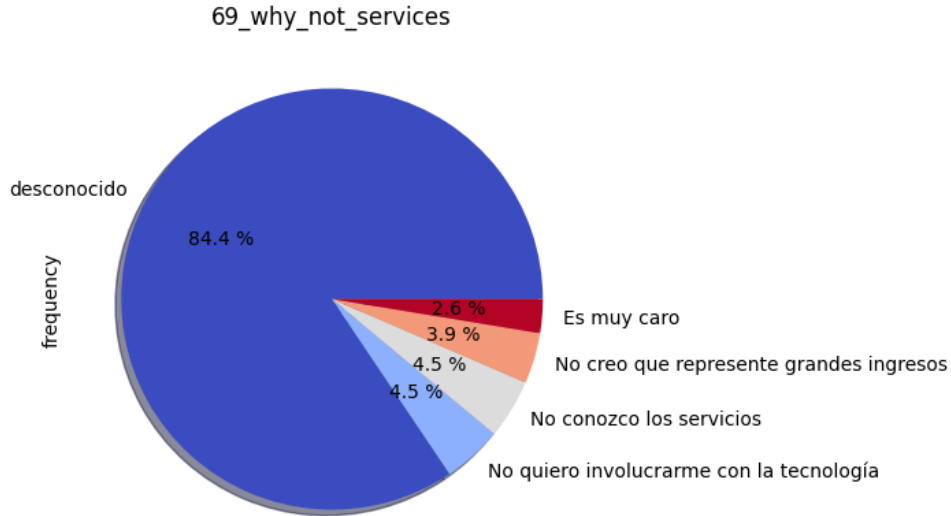


17. 69\_why\_not\_services

Esta variable describe la respuesta a la anterior, ya que describe el porqué las empresas no ofrecen u ofrecerían ningún servicio. Al solo contar con 5 categorías, se utilizaron todas. Como se observa en la imagen 14, a pesar de que se desconoce

gran información, se sabe que no escogen ofrecer servicios por ser caro, no representa grandes ingresos, no conocen de servicios o no se quiere ser parte de las tecnologías (esta como la más importante).

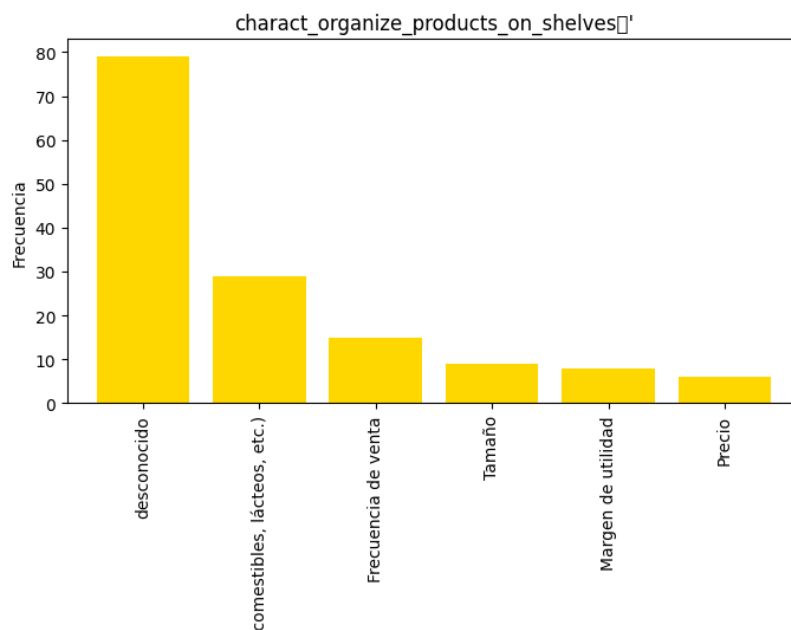
Imagen 14. No servicios



#### 18. 158\_charact\_organize\_products\_on\_shelves

Esta variable describe el cómo los micro retailers escogen organizar sus estantes. Aunque el 46% de la información se desconoce, tal como se observa en la imagen 15 los métodos favoritos son: por categoría, frecuencia de venta, tamaño, utilidad y precio.

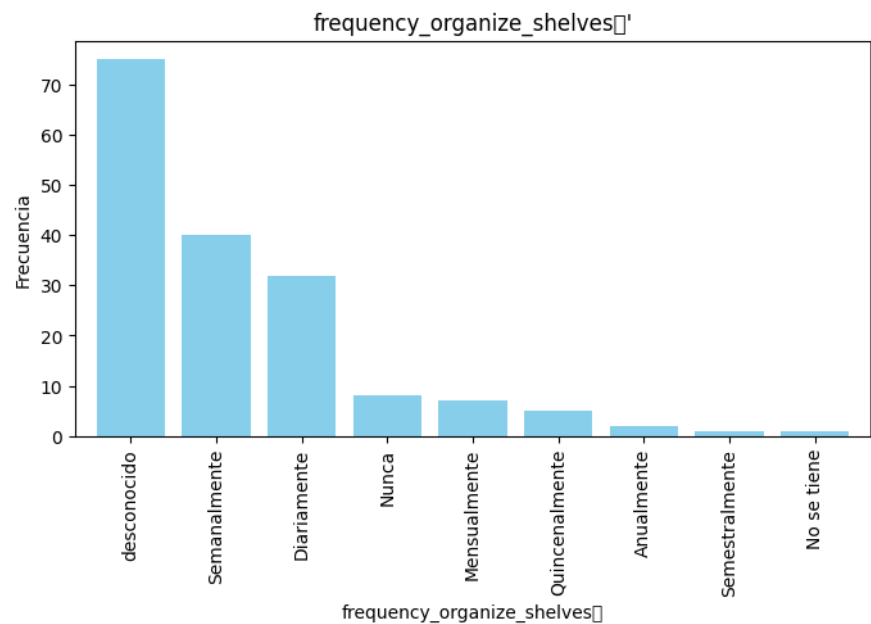
Imagen 15. Criterio para organizar estantes



19. 157\_frequency\_organize\_shelves

Esta variable describe la frecuencia con la que lo empresarios organizan sus repisas. Como en la figura 16 se muestra, la mayoría de las empresas prefieren hacerlo seguido, ya sea semanal o diariamente.

Imagen 16. Frecuencia en organización



20. 161\_actions\_stockouts

Esta variable describe las acciones que los micro\_retailers hacen cuando tienen mercancía agotada, aunque originalmente se contaba con 14 categorías, solamente se dejaron aquellas que se repetían más de 5 veces. Como se observa en la tabla 5, el 46% de la información se desconoce, pero prefieren solicitar al cliente el producto agotado para poder conseguirlo, así como ofrecer un producto sustituto; algunas otras respuestas entraron en que prefieren perder la venta.

Tabla 5. Acciones por stockout

	frequency	percentage	cumulative_perc
161_actions_stockouts			
desconocido	79	0.461988	0.461988
Solicitar tiempo al cliente para conseguir el producto	31	0.181287	0.643275
Ofrecer un producto sustituto	20	0.116959	0.760234
Ofrecer un producto sustituto,Solicitar tiempo al cliente para conseguir el producto	12	0.070175	0.830409
Perder la venta	8	0.046784	0.877193