

## Actividad 4 (Extracción de Características)

Análisis univariado

28 de Septiembre de 2023

ITESM Puebla

Ángel Rubén Vazquez Rivera

(A01735407)



## Índice

<b>Introducción</b>	<b>2</b>
<b>Desarrollo</b>	<b>2</b>
_title	2
_updated_by	3
232_type_of_store	3
108_does_the_micro_retailer_has_a_barred_window_	4
99_does_the_micro_retailer_exhibits_products_outside_	4
102_does_the_micro_retailer_sells_fresh_products_	5
97_number_of_customers_in_store	5
268_number_fridges	6
104_how_many_shelves_does_the_micro_retailer_have	6
2_current_permanent_employees	7
4_number_permanent_employees_last_year	7
20_reviews_finances_monthly	8
49_inventory_records	8
145_number_direct_competitors	9
186_internet_connection	9
214_customer_relationship_tools	10
33_credit_to_customers	10
227_agreement_ecommerce_platforms	11
60_services	11
277_payment_method_suppliers	12
<b>Conclusión</b>	<b>12</b>

## Introducción

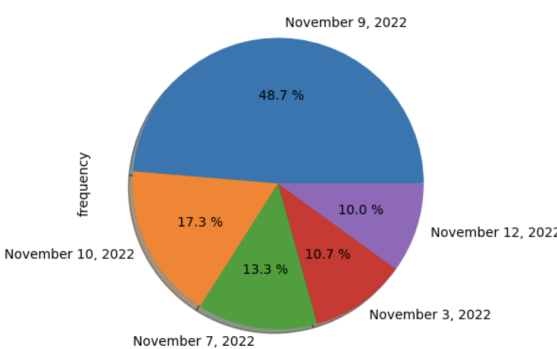
El manejo de datos e información es un área con un potencial enorme en la actualidad, pues llevando a cabo un buen análisis de la información que ya se encuentra al alcance de casi cualquier persona y a través de múltiples medios es posible obtener ventajas y mejoras en los modelos de negocio y predicción de básicamente cualquier situación de la vida real.

En el presente documento contamos con una explicación detallada sobre todos los resultados obtenidos tras aplicar un análisis univariado hacia 20 variables o columnas específicas de una base de datos que fueron seleccionadas en base a criterio propio con visión en la productividad de tratar este tipo de información.

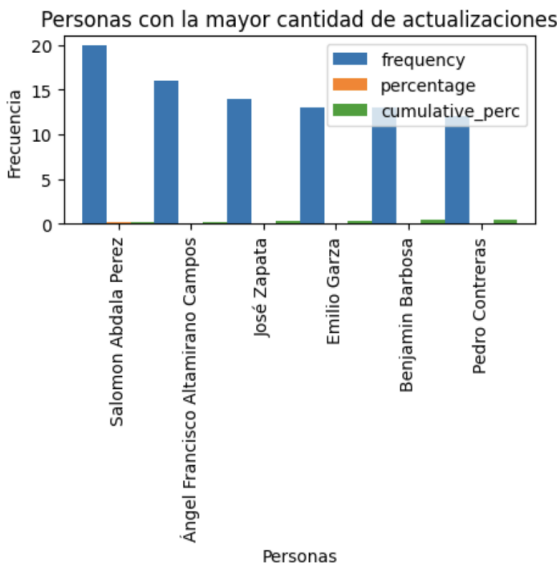
## Desarrollo

A continuación se describen los detalles observados en las gráficas obtenidas tras aplicar una extracción de características y un análisis univariado a cada una de las 20 variables seleccionadas:

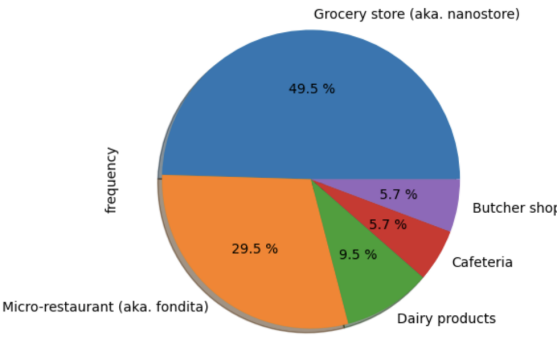
\_title

Gráfica obtenida	Observaciones												
 <table border="1"><thead><tr><th>Fecha</th><th>Porcentaje</th></tr></thead><tbody><tr><td>November 9, 2022</td><td>48.7 %</td></tr><tr><td>November 10, 2022</td><td>17.3 %</td></tr><tr><td>November 7, 2022</td><td>13.3 %</td></tr><tr><td>November 3, 2022</td><td>10.7 %</td></tr><tr><td>November 12, 2022</td><td>10.0 %</td></tr></tbody></table>	Fecha	Porcentaje	November 9, 2022	48.7 %	November 10, 2022	17.3 %	November 7, 2022	13.3 %	November 3, 2022	10.7 %	November 12, 2022	10.0 %	<p>_title se trata del registro de fechas en las que se llevó a cabo la recopilación de información del establecimiento en cuestión, en este caso podemos observar que un 48.7% de los registros fue capturado el 9 de Noviembre de 2020, siendo este un día muy productivo para los encargados de recopilar la información.</p> <p>El resto de registros se divide prácticamente en porcentajes muy similares entre 4 días más.</p>
Fecha	Porcentaje												
November 9, 2022	48.7 %												
November 10, 2022	17.3 %												
November 7, 2022	13.3 %												
November 3, 2022	10.7 %												
November 12, 2022	10.0 %												

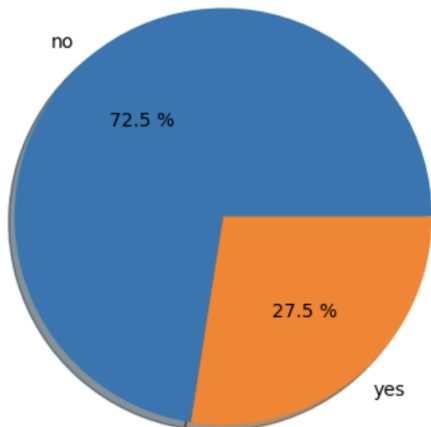
\_updated\_by

Gráfica obtenida	Observaciones
 <p>Personas con la mayor cantidad de actualizaciones</p> <p>Frequency</p> <p>Salomon Abdala Perez</p> <p>Ángel Francisco Altamirano Campos</p> <p>José Zapata</p> <p>Emilio Garza</p> <p>Benjamin Barbosa</p> <p>Pedro Contreras</p> <p>Personas</p> <p>frequency</p> <p>percentage</p> <p>cumulative_perc</p>	<p>_updated_by se trata de las personas que registraron la información y cuenta con una mayor o menor cantidad de actualizaciones a la información, en este caso se puede determinar la actividad de cada uno de los capturadores, llegando a la conclusión de que Salomon es el usuario con mayor actividad y trabajo registrado.</p>

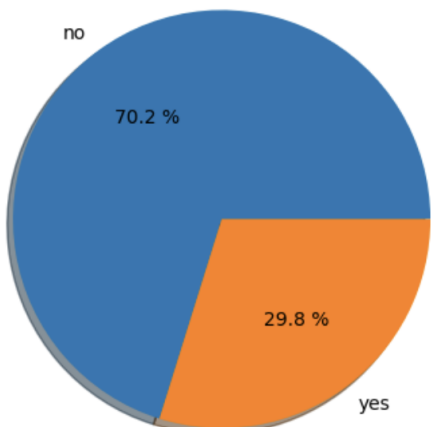
232\_type\_of\_store

Gráfica obtenida	Observaciones
 <p>frequency</p> <p>Grocery store (aka. nanostore)</p> <p>49.5 %</p> <p>Micro-restaurant (aka. fondita)</p> <p>29.5 %</p> <p>Dairy products</p> <p>9.5 %</p> <p>Cafeteria</p> <p>5.7 %</p> <p>Butcher shop</p> <p>5.7 %</p>	<p>232_type_of_store se trata de la clasificación de las tiendas o establecimientos en categorías específicas, llegando a la observación de que un 49.5% de los establecimientos se encuentran dentro del tipo de Grocery store, mientras que el segundo tipo de establecimientos más presente es el Micro-Restaurant, comprendiendo así que existe una mayor cantidad de tiendas pequeñas, pues representa básicamente un 50% de todos los registros.</p>

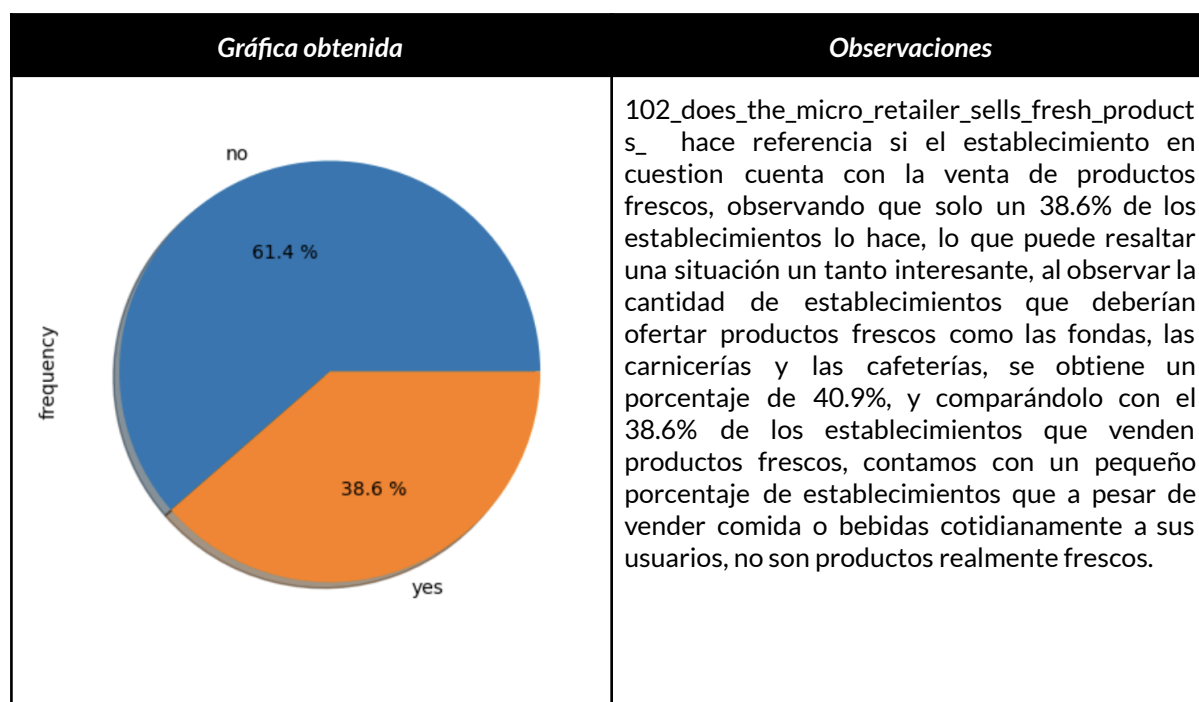
### 108\_does\_the\_micro\_retailer\_has\_a\_barred\_window\_

Gráfica obtenida	Observaciones						
<p>frequency</p>  <table border="1"> <thead> <tr> <th>Response</th> <th>Frequency (%)</th> </tr> </thead> <tbody> <tr> <td>no</td> <td>72.5 %</td> </tr> <tr> <td>yes</td> <td>27.5 %</td> </tr> </tbody> </table>	Response	Frequency (%)	no	72.5 %	yes	27.5 %	<p>108_does_the_micro_retailer_has_a_barred_window_ refiere a si el establecimiento en cuestión cuenta con rejas o protecciones en sus ventanas, y se puede observar en gran medida que solo un 27.5% de los establecimientos cuenta con este tipo de medidas de seguridad, mientras que un 72.5% no está protegido de esta manera, siendo así que es posible detectar una gran vulnerabilidad ante actos delictivos como robos o atracos en una gran mayoría de los establecimientos.</p>
Response	Frequency (%)						
no	72.5 %						
yes	27.5 %						

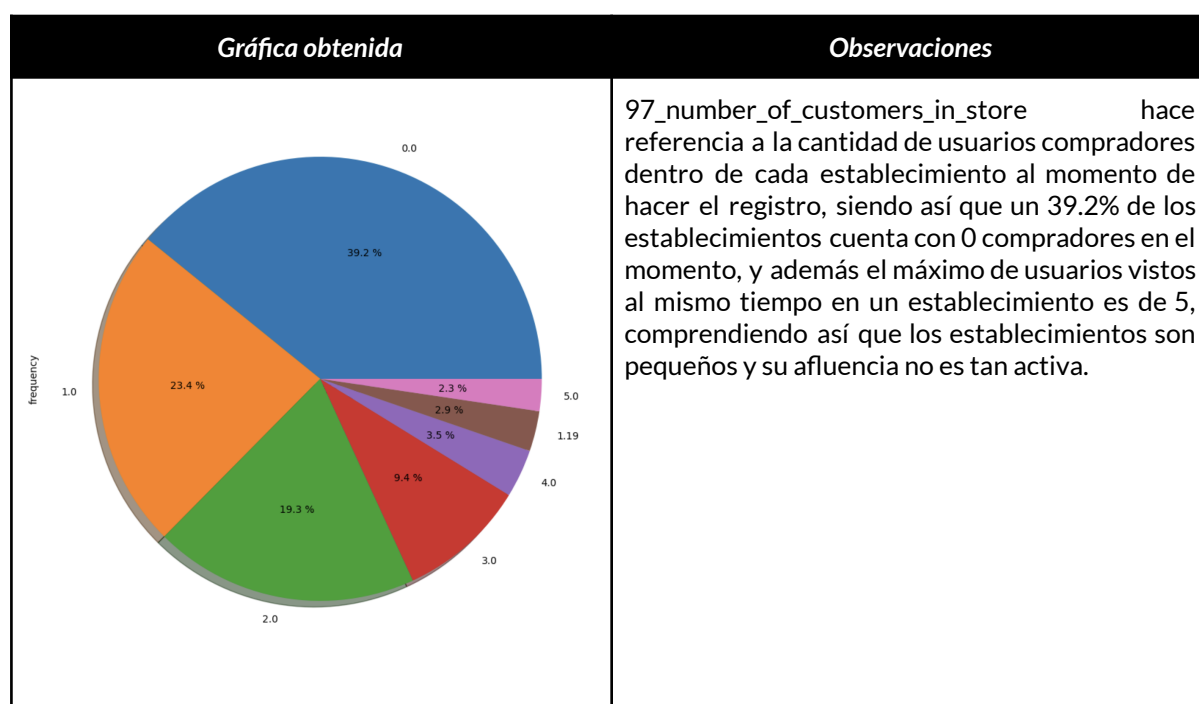
### 99\_does\_the\_micro\_retailer\_exhibits\_products\_outside\_

Gráfica obtenida	Observaciones						
<p>frequency</p>  <table border="1"> <thead> <tr> <th>Response</th> <th>Frequency (%)</th> </tr> </thead> <tbody> <tr> <td>no</td> <td>70.2 %</td> </tr> <tr> <td>yes</td> <td>29.8 %</td> </tr> </tbody> </table>	Response	Frequency (%)	no	70.2 %	yes	29.8 %	<p>99_does_the_micro_retailer_exhibits_products_outside_ se refiere al hecho de que los establecimientos exhiban algunos de sus productos fuera de su local, es decir, en la banqueta, lo que nos muestra que un 70.2% de los establecimientos no lo hace, y solo un 29.8%.</p> <p>De entrada como posible relación podría ser posible que los locales con mejor seguridad se sientan más cómodos al exponer productos fuera, además de que al considerar que la mayoría de los establecimientos son tiendas pequeñas con productos bastante vulnerables ante robos, el 29.8% de esta gráfica es muy probable que se relacione más con la cantidad de restaurantes o fondas que puedan contar con mesas o algo por el estilo en las inmediaciones de su sitio.</p>
Response	Frequency (%)						
no	70.2 %						
yes	29.8 %						

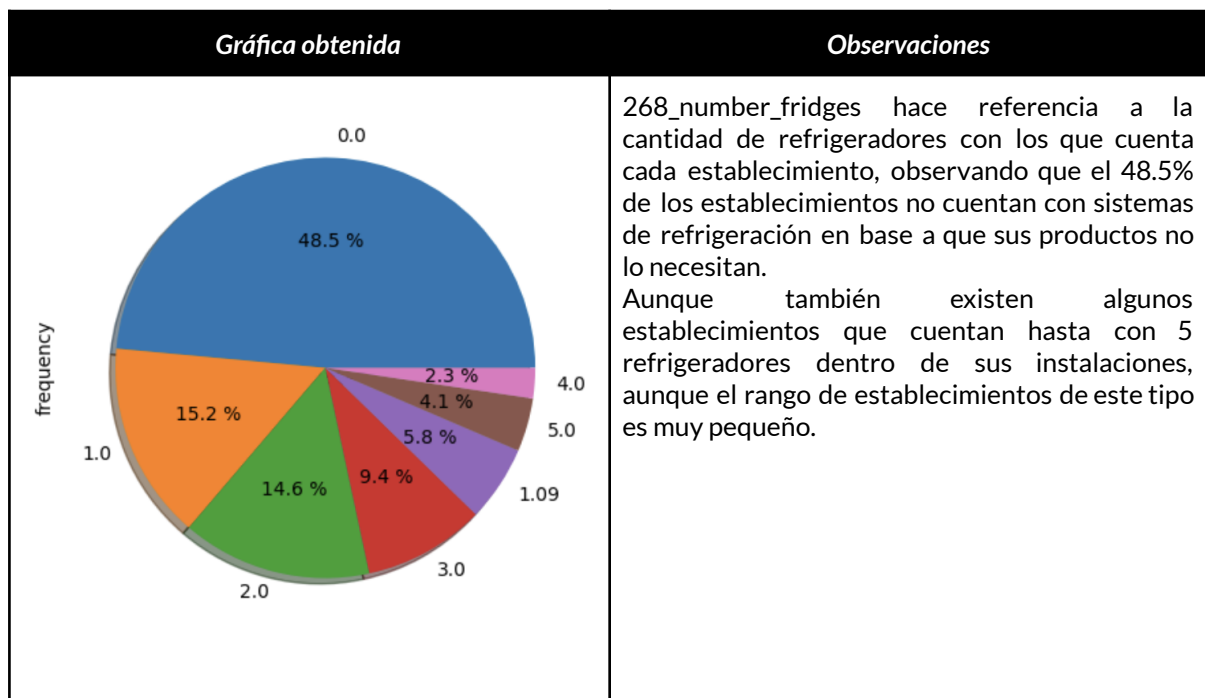
## 102\_does\_the\_micro\_retailer\_sells\_fresh\_products\_



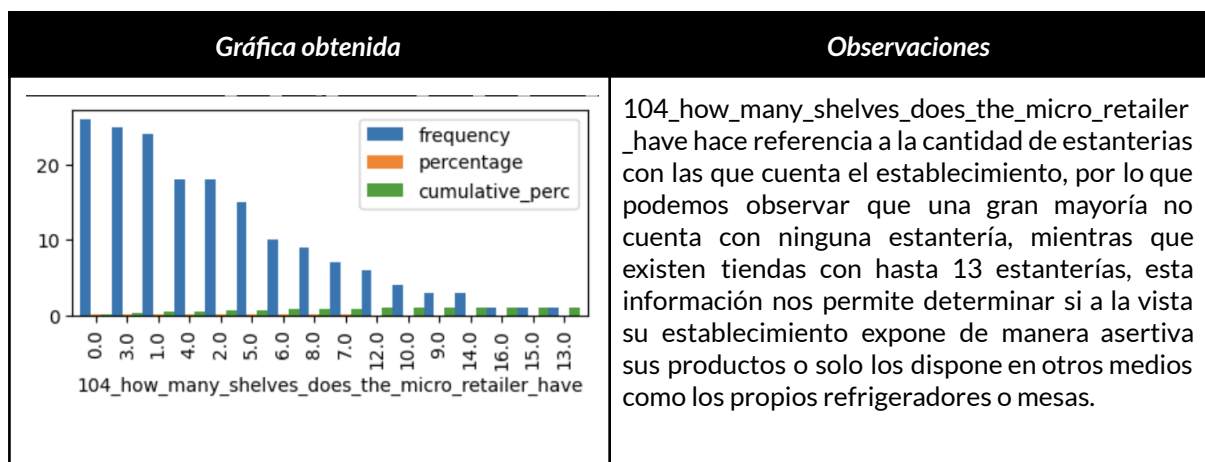
## 97\_number\_of\_customers\_in\_store



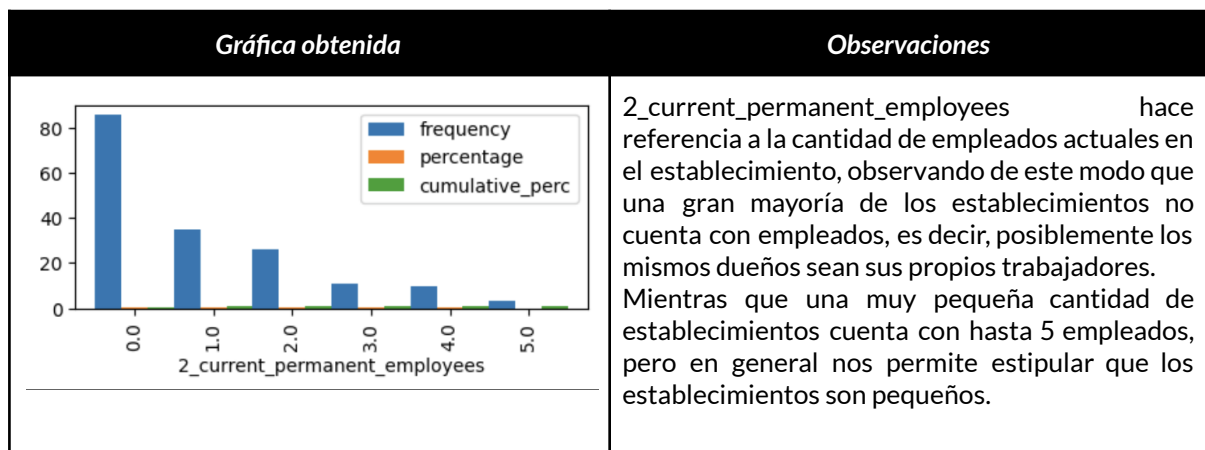
## 268\_number\_fridges



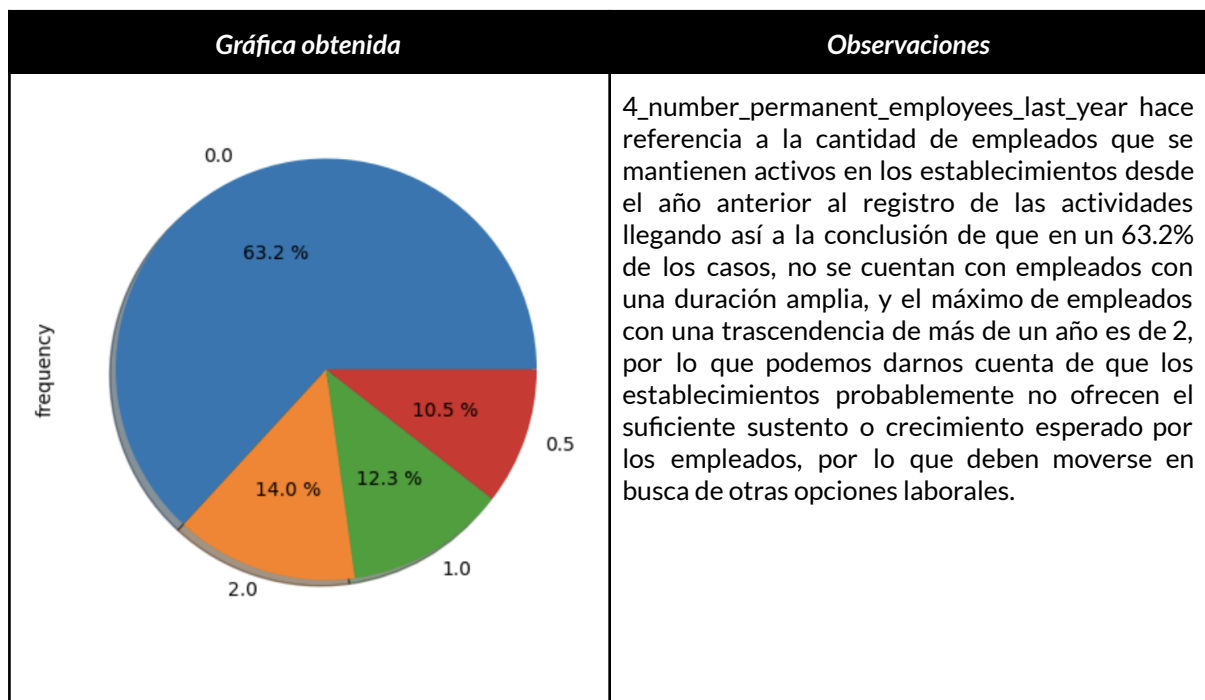
## 104\_how\_many\_shelves\_does\_the\_micro-retailer\_have



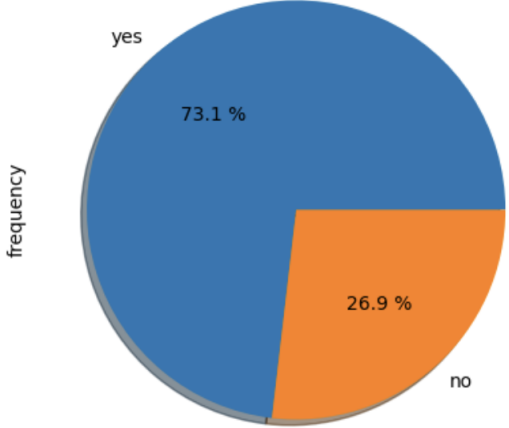
## 2\_current\_permanent\_employees



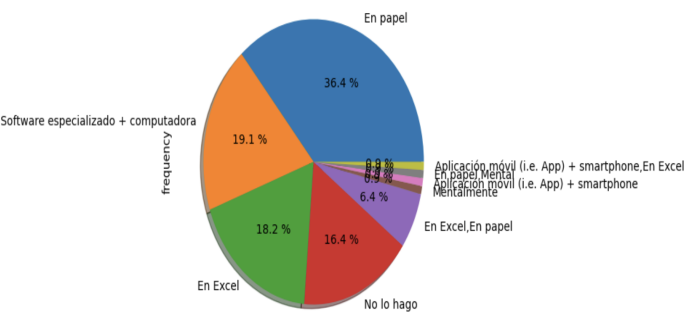
## 4\_number\_permanent\_employees\_last\_year



## 20\_reviews\_finances\_monthly

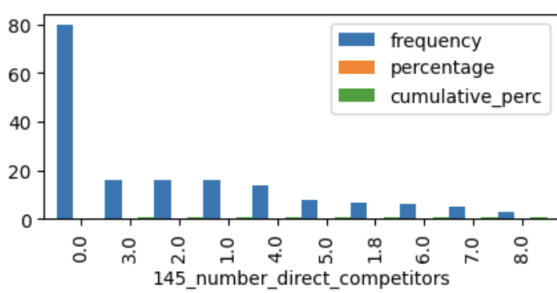
Gráfica obtenida	Observaciones						
 <p>A pie chart titled 'frequency' showing the distribution of responses for '20_reviews_finances_monthly'. The chart is divided into two segments: a large blue segment representing 'yes' at 73.1%, and a smaller orange segment representing 'no' at 26.9%.</p> <table border="1"> <thead> <tr> <th>Response</th> <th>Frequency (%)</th> </tr> </thead> <tbody> <tr> <td>yes</td> <td>73.1 %</td> </tr> <tr> <td>no</td> <td>26.9 %</td> </tr> </tbody> </table>	Response	Frequency (%)	yes	73.1 %	no	26.9 %	<p>20_reviews_finances_monthly hace referencia a la cantidad de establecimientos que realizan de manera periódica una revisión a sus finanzas.</p> <p>Es posible observar que en un 73.1% de los casos, estos si llevan a cabo una revisión, lo cual es de esperarse, pues un establecimiento sin este tipo de chequeos estaría en riesgo.</p> <p>Un 26.9% de los establecimientos no llevan a cabo un chequeo de este tipo, por lo que podemos encontrar una vulnerabilidad en el modelo de negocios de estos, pues no podrían darse cuenta a tiempo de posibles altas o bajas en su rendimiento, o inclusive determinar si su negocio será viable en un futuro a mediano o largo plazo.</p>
Response	Frequency (%)						
yes	73.1 %						
no	26.9 %						

## 49\_inventory\_records

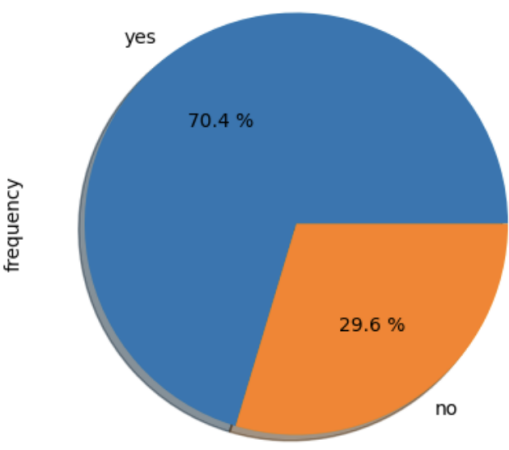
Gráfica obtenida	Observaciones																		
 <p>A pie chart titled 'frequency' showing the distribution of responses for '49_inventory_records'. The chart is divided into nine segments representing different methods of recording inventory. The largest segment is 'En papel' at 36.4%, followed by 'Software especializado + computadora' at 19.1%, 'En Excel' at 18.2%, and 'No lo hago' at 16.4%. Smaller segments include 'En Excel, En papel' (6.4%), and three segments each at 0.9%: 'Aplicación móvil (i.e. App) + smartphone, En Excel', 'En papel, Mentalmente', and 'Aplicación móvil (i.e. App) + smartphone'.</p> <table border="1"> <thead> <tr> <th>Method</th> <th>Frequency (%)</th> </tr> </thead> <tbody> <tr> <td>En papel</td> <td>36.4 %</td> </tr> <tr> <td>Software especializado + computadora</td> <td>19.1 %</td> </tr> <tr> <td>En Excel</td> <td>18.2 %</td> </tr> <tr> <td>No lo hago</td> <td>16.4 %</td> </tr> <tr> <td>En Excel, En papel</td> <td>6.4 %</td> </tr> <tr> <td>Aplicación móvil (i.e. App) + smartphone, En Excel</td> <td>0.9 %</td> </tr> <tr> <td>En papel, Mentalmente</td> <td>0.9 %</td> </tr> <tr> <td>Aplicación móvil (i.e. App) + smartphone</td> <td>0.9 %</td> </tr> </tbody> </table>	Method	Frequency (%)	En papel	36.4 %	Software especializado + computadora	19.1 %	En Excel	18.2 %	No lo hago	16.4 %	En Excel, En papel	6.4 %	Aplicación móvil (i.e. App) + smartphone, En Excel	0.9 %	En papel, Mentalmente	0.9 %	Aplicación móvil (i.e. App) + smartphone	0.9 %	<p>49_inventory_records hace referencia a los tipos de medios que utilizan los establecimientos para llevar a cabo el registro de sus inventarios o existencias.</p> <p>Es importante observar que un 36.4% de los establecimientos hace este recuento de manera manual en papel, dejando en segunda posición el uso de algún software especializado y excel, lo que es bastante común.</p> <p>Según mi interpretación el caso de llevar el registro en papel es posible debido a las dimensiones del establecimiento, pues sus movimientos no deben ser tan grandes, sin embargo, sí debe ser complicado calcular toda la información necesaria en poco tiempo, alentando sus procesos.</p>
Method	Frequency (%)																		
En papel	36.4 %																		
Software especializado + computadora	19.1 %																		
En Excel	18.2 %																		
No lo hago	16.4 %																		
En Excel, En papel	6.4 %																		
Aplicación móvil (i.e. App) + smartphone, En Excel	0.9 %																		
En papel, Mentalmente	0.9 %																		
Aplicación móvil (i.e. App) + smartphone	0.9 %																		



## 145\_number\_direct\_competitors

Gráfica obtenida	Observaciones																				
 <p>Bar chart showing the frequency of direct competitors. The x-axis is labeled '145_number_direct_competitors' and ranges from 0.0 to 8.0. The y-axis ranges from 0 to 80. The chart shows a high frequency for 0 competitors (approx. 80), with frequencies decreasing for 1 through 8 competitors.</p> <table border="1"> <thead> <tr> <th>145_number_direct_competitors</th> <th>frequency</th> </tr> </thead> <tbody> <tr><td>0.0</td><td>80</td></tr> <tr><td>1.0</td><td>15</td></tr> <tr><td>2.0</td><td>15</td></tr> <tr><td>3.0</td><td>15</td></tr> <tr><td>4.0</td><td>10</td></tr> <tr><td>5.0</td><td>5</td></tr> <tr><td>6.0</td><td>5</td></tr> <tr><td>7.0</td><td>5</td></tr> <tr><td>8.0</td><td>5</td></tr> </tbody> </table>	145_number_direct_competitors	frequency	0.0	80	1.0	15	2.0	15	3.0	15	4.0	10	5.0	5	6.0	5	7.0	5	8.0	5	<p>145_number_direct_competitors hace referencia a la cantidad de competidores directos en sus ventas.</p> <p>La mayoría de los establecimientos no tienen identificado a un competidor directo contra el cual vender sus productos, sin embargo en pequeñas proporciones algunos de estos establecimientos pueden llegar a competir hasta con 8 establecimientos o comerciantes diferentes, lo que les complica bastante mantenerse a flote en sus movimientos comerciales.</p> <p>Algo que puedo notar es que al ser establecimientos pequeños, es muy probable que no se encuentren en zonas muy demandantes o llenas de competencia, además de que su mercado posiblemente se limita a zonas pequeñas o moderadas.</p>
145_number_direct_competitors	frequency																				
0.0	80																				
1.0	15																				
2.0	15																				
3.0	15																				
4.0	10																				
5.0	5																				
6.0	5																				
7.0	5																				
8.0	5																				

## 186\_internet\_connection

Gráfica obtenida	Observaciones						
 <p>Pie chart showing the frequency of internet connection. The chart is divided into two segments: 'yes' (70.4%) and 'no' (29.6%). The y-axis is labeled 'frequency'.</p> <table border="1"> <thead> <tr> <th>Internet Connection</th> <th>frequency</th> </tr> </thead> <tbody> <tr><td>yes</td><td>70.4 %</td></tr> <tr><td>no</td><td>29.6 %</td></tr> </tbody> </table>	Internet Connection	frequency	yes	70.4 %	no	29.6 %	<p>186_internet_connection hace referencia a la cantidad de establecimientos que cuentan con conexión a internet.</p> <p>Un 70.4% de los establecimientos cuenta con conexión a internet, esto puede deberse a que es una herramienta útil para ellos o posiblemente el establecimiento se encuentre inclusive en la propiedad personal de los dueños del mismo.</p> <p>Solo un 29.6% de los establecimientos no cuenta con este servicio, por lo que tal vez su tipo o categoría de ventas no lo requiere o se trata de un establecimiento que solo comercia productos pequeños o rápidos con sus usuarios.</p>
Internet Connection	frequency						
yes	70.4 %						
no	29.6 %						

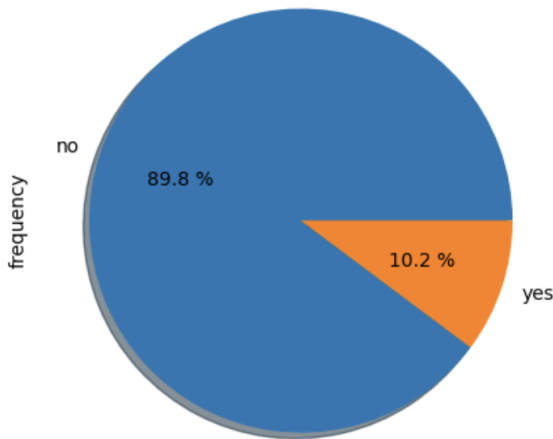
## 214\_customer\_relationship\_tools

Gráfica obtenida	Observaciones
<p>A pie chart titled 'Gráfica obtenida' showing the frequency of tools used for customer relationship management. The chart is divided into five segments: 'Redes sociales' (blue, 31.0%), 'Aplicación móvil' (orange, 24.1%), 'Ninguna' (green, 23.0%), 'Aplicación móvil, Redes sociales' (red, 14.9%), and 'Cuaderno' (purple, 6.9%). The word 'frequency' is written vertically on the left side of the chart.</p>	<p>214_customer_relationship_tools hace referencia a los tipos de herramientas o medios a través de los cuales se entablan relaciones comerciales entre los establecimientos y los compradores potenciales.</p> <p>Las redes sociales con un 31% es una de las herramientas más empleadas para esto, seguidas de aplicaciones móviles, como podría ser el caso de agendas o software de ese estilo, también existe un 23% de establecimientos que decide no usar ninguna herramienta, llegando a la conclusión de que en general solo una cuarta parte de los establecimientos no mantiene un registro o conexión con sus clientes de manera tan asertiva.</p>

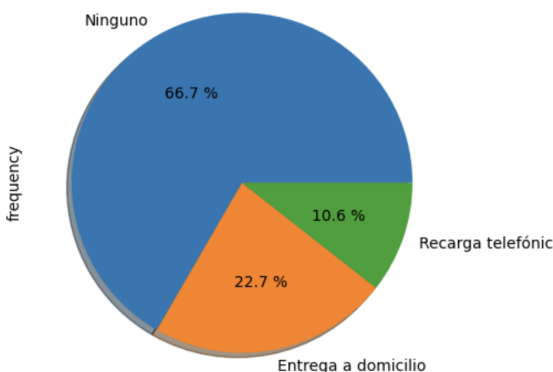
## 33\_credit\_to\_customers

Gráfica obtenida	Observaciones
<p>A pie chart titled 'Gráfica obtenida' showing the frequency of credit offered to customers. The chart is divided into two segments: 'no' (blue, 74.8%) and 'yes' (orange, 25.2%). The word 'frequency' is written vertically on the left side of the chart.</p>	<p>33_credit_to_customers hace referencia al hecho de si un establecimiento le ofrece o no la posibilidad de contar con crédito a sus clientes.</p> <p>En este caso un 74.8% no cuenta con crédito para sus clientes, lo cual es entendible pues al ser establecimientos pequeños no existe la necesidad de costear productos realmente costosos, y además no sería adecuado que cuenten con un sistema de este tipo al necesitar y depender de una inyección de capital constante y dosificada según sus ventas.</p> <p>Un 25.2% si lo hace y esto puede derivarse a que cuentan con una mayor cantidad de clientes, uso de terminales, y posiblemente productos más costosos.</p>

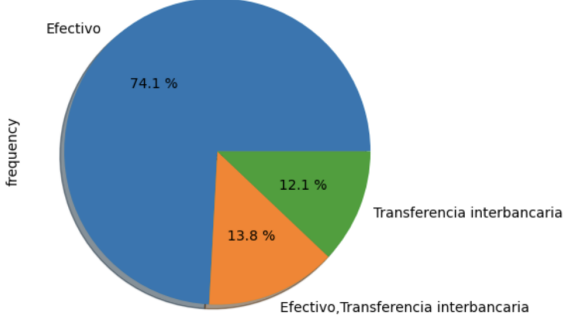
## 227\_agreement\_ecommerce\_platforms

Gráfica obtenida	Observaciones						
 <p>A pie chart titled 'frequency' showing the distribution of establishments using an e-commerce platform. The chart is divided into two segments: a large blue segment representing 'no' at 89.8%, and a smaller orange segment representing 'yes' at 10.2%.</p> <table border="1"> <thead> <tr> <th>Respuesta</th> <th>Frecuencia (%)</th> </tr> </thead> <tbody> <tr> <td>no</td> <td>89.8 %</td> </tr> <tr> <td>yes</td> <td>10.2 %</td> </tr> </tbody> </table>	Respuesta	Frecuencia (%)	no	89.8 %	yes	10.2 %	<p>227_agreement_ecommerce_platforms hace referencia a la cantidad de establecimientos que cuentan con una plataforma de comercio en línea.</p> <p>Un 89.8% de los establecimientos no cuenta con este servicio, mientras que solo un 10.2% si emplea este tipo de recursos.</p> <p>Esto es comprensible, pues aplicar soluciones o estrategias de venta para comercios que no lo requieren sólo sería una inversión sin sentido o mucha productividad.</p>
Respuesta	Frecuencia (%)						
no	89.8 %						
yes	10.2 %						

## 60\_services

Gráfica obtenida	Observaciones								
 <p>A pie chart titled 'frequency' showing the distribution of establishments offering extra services. The chart is divided into three segments: a large blue segment representing 'Ninguno' (None) at 66.7%, an orange segment representing 'Entrega a domicilio' (Home delivery) at 22.7%, and a green segment representing 'Recarga telefónica' (Phone recharge) at 10.6%.</p> <table border="1"> <thead> <tr> <th>Servicio</th> <th>Frecuencia (%)</th> </tr> </thead> <tbody> <tr> <td>Ninguno</td> <td>66.7 %</td> </tr> <tr> <td>Entrega a domicilio</td> <td>22.7 %</td> </tr> <tr> <td>Recarga telefónica</td> <td>10.6 %</td> </tr> </tbody> </table>	Servicio	Frecuencia (%)	Ninguno	66.7 %	Entrega a domicilio	22.7 %	Recarga telefónica	10.6 %	<p>60_services hace referencia a los tipos de servicios extras que puede ofrecer el establecimiento a sus usuarios, servicios que van más allá de su propio producto o productos de venta.</p> <p>un 66.7% de los establecimientos no cuenta con estos servicios, lo cual tiene sentido, pues su mercado, demanda y rendimiento no les permite alcanzar este tipo de factores.</p> <p>solo un 10.6% cuenta con recargas telefónicas, y 22.7% cuenta con entrega a domicilio, siendo así que solo una pequeña parte de los establecimientos es lo suficientemente rentable y grande como para ofertar esos servicios.</p>
Servicio	Frecuencia (%)								
Ninguno	66.7 %								
Entrega a domicilio	22.7 %								
Recarga telefónica	10.6 %								

## 277\_payment\_method\_suppliers

Gráfica obtenida	Observaciones								
 <p>A pie chart illustrating the frequency of payment methods used by suppliers. The chart is divided into three segments: a large blue segment representing 'Efectivo' at 74.1%, an orange segment representing 'Efectivo, Transferencia interbancaria' at 13.8%, and a green segment representing 'Transferencia interbancaria' at 12.1%. The word 'frequency' is written vertically on the left side of the chart.</p> <table border="1"> <thead> <tr> <th>Método de Pago</th> <th>Frecuencia (%)</th> </tr> </thead> <tbody> <tr> <td>Efectivo</td> <td>74.1 %</td> </tr> <tr> <td>Efectivo, Transferencia interbancaria</td> <td>13.8 %</td> </tr> <tr> <td>Transferencia interbancaria</td> <td>12.1 %</td> </tr> </tbody> </table>	Método de Pago	Frecuencia (%)	Efectivo	74.1 %	Efectivo, Transferencia interbancaria	13.8 %	Transferencia interbancaria	12.1 %	<p>277_payment_method_suppliers hace referencia al tipo de servicio de pago que emplea el establecimiento.</p> <p>Un 74.1% de los establecimientos solo acepta efectivo, mientras que un 13.8% recibe efectivo y transferencias bancarias, mientras que un 12.1% se limita solo a transferencias bancarias.</p> <p>En general me interesan mucho los establecimientos que solo usan transferencias bancarias, pues supongo que se tratan de los comercios en línea o con plataforma de venta en línea.</p> <p>Es lógico que la mayoría solo emplee efectivo, pues se tratan de establecimientos pequeños.</p>
Método de Pago	Frecuencia (%)								
Efectivo	74.1 %								
Efectivo, Transferencia interbancaria	13.8 %								
Transferencia interbancaria	12.1 %								

## Conclusión

Al analizar con detenimiento los resultados obtenidos es posible conocer más a fondo el comportamiento del escenario que se está analizando, estipular los factores más importantes y finalmente llegar a conclusiones que nos permitan comprender cómo generar estrategias y plantear posibles comportamientos entre tus unidades observadas.