



**Tecnológico
de Monterrey**

Abner palma Garcia

Actividad 4 :Extracción de Características

Análisis general de la base de datos

micro retailer mit lift lab

▸ cargar el archivo

4,7 celdas ocultas

▼ filtro

```
[ ] 1 categoricas= Micro_Retailer[['232_type_of_store','108_does_the_micro_retailer_has_a_barred_window_',
2                                'pictures_of_shelves_if_possible','5_change_store_space_last_year',
3                                '6_change_employees_average_salary_last_year','155_sales_registers_used_for',
4                                '49_inventory_records','18_sales_records',
5                                '210_sales_channels','189_payment_methods',
6                                '28_prefered_payment_method','270_card_extra_charge',
7                                '214_customer_relationship_tools','38_average_margin_profits'
8                                , '193_sales_planning_tools','60_services',
9                                '79_delivery_vehicle','312_payment_of_utilities',
10                               '185_place_orders_suppliers','157_frequency_organize_shelves']]
11 categoricas.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 171 entries, 0 to 170
Data columns (total 20 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   232_type_of_store                    171 non-null    object
1   108_does_the_micro_retailer_has_a_barred_window_  171 non-null    object
2   pictures_of_shelves_if_possible      19 non-null     object
3   5_change_store_space_last_year       99 non-null     object
4   6_change_employees_average_salary_last_year  85 non-null     object
```

Filtro

El primer paso después de cargar el archivo para leerlo es filtrar las columnas que utilizaremos en este caso se escogieron 20 columnas las cuales consideramos más relevantes para su posterior análisis.

Sustitución de nulos

Posteriormente se utiliza la función `isnull` para visualizar los valores nulos de cada columna y posteriormente se hace la sustitución de los valores nulos por el string sin registro ya que de esta manera podemos visualizar de manera exacta las carencias de valores dentro del data frame a diferencia que si realizamos un promedio podremos estar manipulando la información y al momento de realizar la interpretación esta no se apagará

▼ nulos

```
1 #Checar si hay valores nulos por columna
2 valores_nulos=categoricas.isnull().sum()
3 valores_nulos
```

```
232_type_of_store                0
108_does_the_micro_retailer_has_a_barred_window_  0
pictures_of_shelves_if_possible  152
5_change_store_space_last_year   72
6_change_employees_average_salary_last_year  86
155_sales_registers_used_for     165
49_inventory_records             61
18_sales_records                 61
210_sales_channels               89
189_payment_methods              64
28_prefered_payment_method       169
270_card_extra_charge            169
214_customer_relationship_tools   76
38_average_margin_profits        105
193_sales_planning_tools         89
60_services                      94
79_delivery_vehicle              169
312_payment_of_utilities         105
185_place_orders_suppliers       100
157_frequency_organize_shelves   75
dtype: int64
```

lo más posible a la realidad por lo que se decidió conservar estos registros sin valor para contabilizarlos en una nueva categoría

232_type_of_store

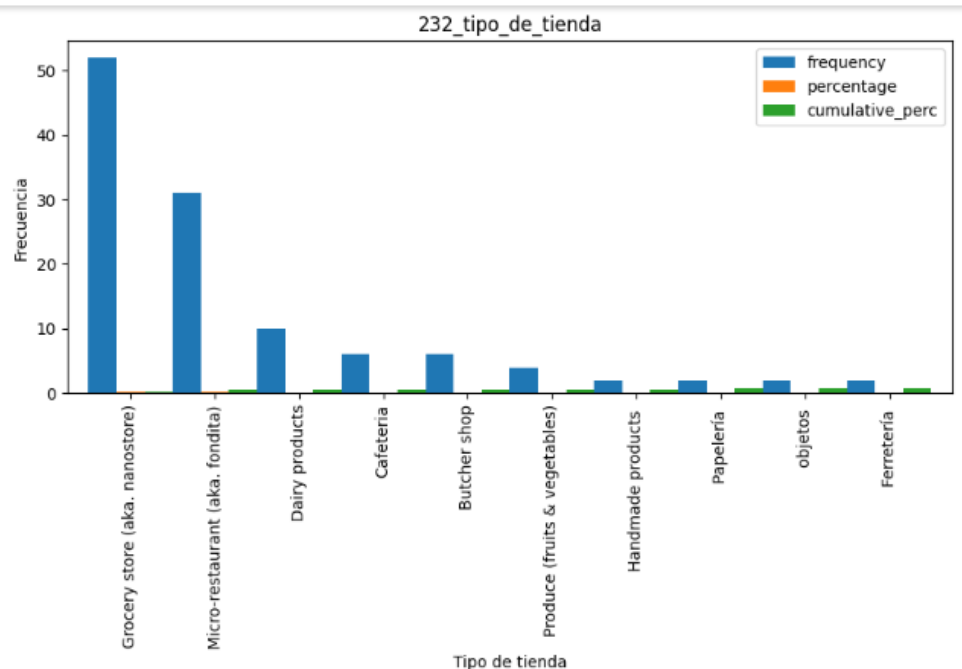
en esta tabla podemos observar la frecuencia del tipo de tiendas que contestaron las encuestas en donde en su mayoría son supermercados con una frecuencia de 52 puntos seguida de los micro restaurantes o las conocidas fondas con 31 puntos sello de cafeterías y carnicerías, por lo que podemos observar que en su mayoría las tiendas que respondieron esta encuesta son del sector primaria de alimentos.

▼ sustitucion de valores nulos

```
[ ] 1 #sustituir valores nulos por un string en concreto
    2 categoricas=categoricas.fillna('sin registro')
```

```
[ ] 1 #Checar si hay valores nulos por columna
    2 valores_nulos=categoricas.isnull().sum()
    3 valores_nulos
```

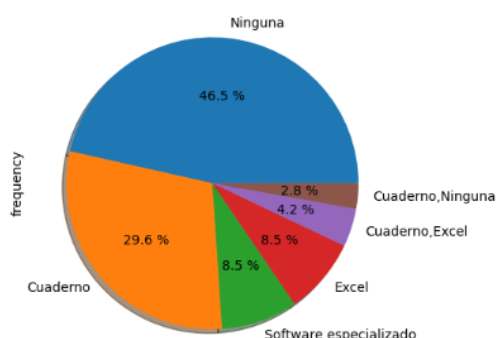
```
232_type_of_store      0
108_does_the_micro_retailer_has_a_barred_window_      0
pictures_of_shelves_if_possible      0
5_change_store_space_last_year      0
6_change_employees_average_salary_last_year      0
155_sales_registers_used_for      0
10_furniture_items      0
```



193_sales_planning_tools

```
1 #Realizamos grafico de pastel del dataframe filtrado
2 Filtro_index["frequency"].plot(kind='pie', figsize=(10,5), s
```

<Axes: ylabel='frequency'>



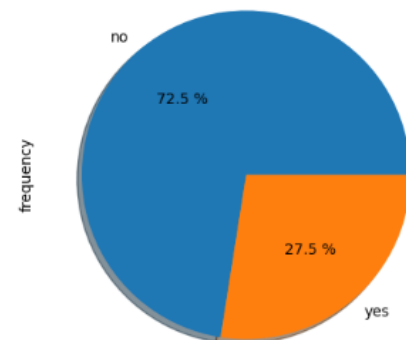
En la gráfica anterior podemos observar que en su mayoría no hay ningún registro una herramienta para la planificación de ventas ya que el 46% de las tiendas encuestadas no utiliza ninguna herramienta mientras que el 29% utiliza un cuaderno

mientras que solamente el 8.5% utiliza o software especializado o Excel por lo que vemos que las tiendas que en su mayoría contestaron la encuesta son carecen de estas herramientas y funciona de manera tradicional

108_does_the_micro_retailer_has_a_barred_window_

en este gráfico podemos observar que el 72% de las tiendas que contestaron la encuesta no cuentan con una ventana restringida por lo que en cierto caso podría representar un problema de seguridad , sin embargo esta información podemos relacionarla con la temporalidad de venta que tiene cada establecimiento ya que al no necesitar una ventana restringida podemos decir que no vende en horarios posteriores en los que sí se ocuparía por lo que podemos decir que el 27% de las tiendas que contestaron esta encuesta venden en horas extra de lo permitido puede ser 24/7.

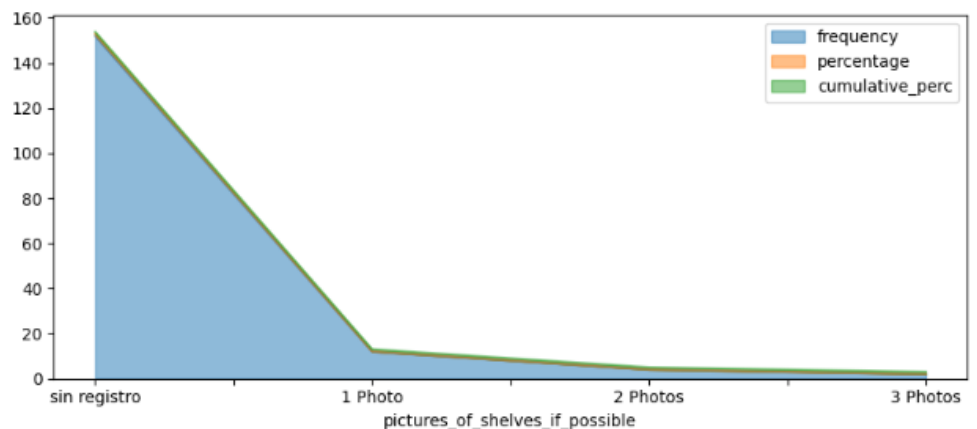
```
1 #Realizamos grafico de pastel del dataframe filt
2 Filtro_index["frequency"].plot(kind='pie', figsi
<Axes: ylabel='frequency'>
```



pictures_of_shelves_if_possible

En esta gráfica podemos observar que gran parte o casi en su totalidad de las tiendas que contestaron la encuesta no cuentan con fotografías de los estantes de sus productos y menos de 10 tiendas cuentan con por lo menos una a tres fotos.

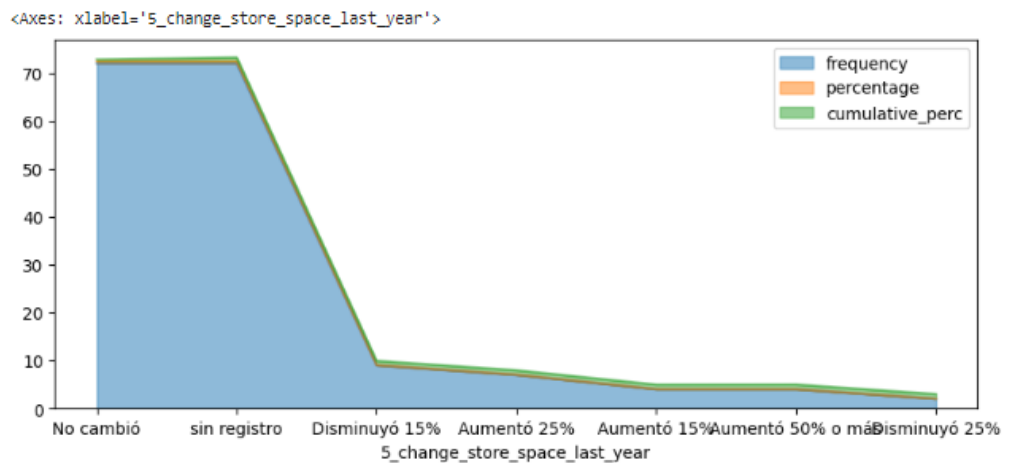
```
1 #Realizamos grafico de área del dataframe filtrado
2 Filtro_index.plot(kind='area', figsize=(10,4),alpha = 0.5)
<Axes: xlabel='pictures_of_shelves_if_possible'>
```



5_change_store_space_last_year

```
1 #Realizamos grafico de área del dataframe filtrado
2 Filtro_index.plot(kind='area', figsize=(10,4),alpha = 0.5)
```

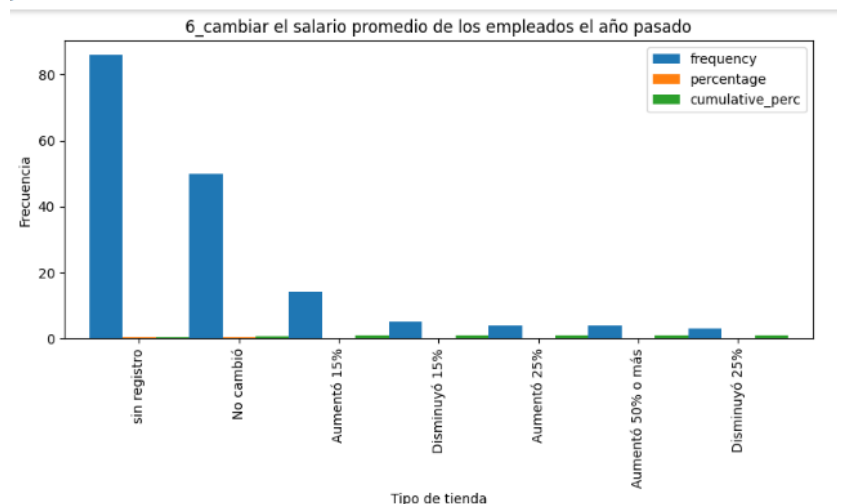
en esta gráfica podemos observar que el cambio de espacio disponible en la tienda respecto al del año pasado no ha cambiado en su gran mayoría ya que 70 de las tiendas que contestaron el



registro no cambió o no tiene ningún ninguna información respecto a esta categoría, 10 de las tiendas que contestaron el registro disminuyó en un 15% y menos de la de 10 tiendas que contestaron el registro aumentaron por lo menos 15, 25 y 50% más por lo que podemos observar que en su mayoría no hubo ningún tipo de cambio en el espacio disponible las tiendas sin embargo hay algunas que aumentaron sus espacios más de las que disminuyeron los suyos.

6_change_employees_average_salary_last_year

Esa gráfica nos dice el cambio del salario promedio de los empleados respecto al del año pasado al igual que la gráfica anterior en su mayoría no se nota ningún cambio y considerando que el 80% de las tiendas que contestaron la encuesta no pusieron ninguna información respecto a esta teoría

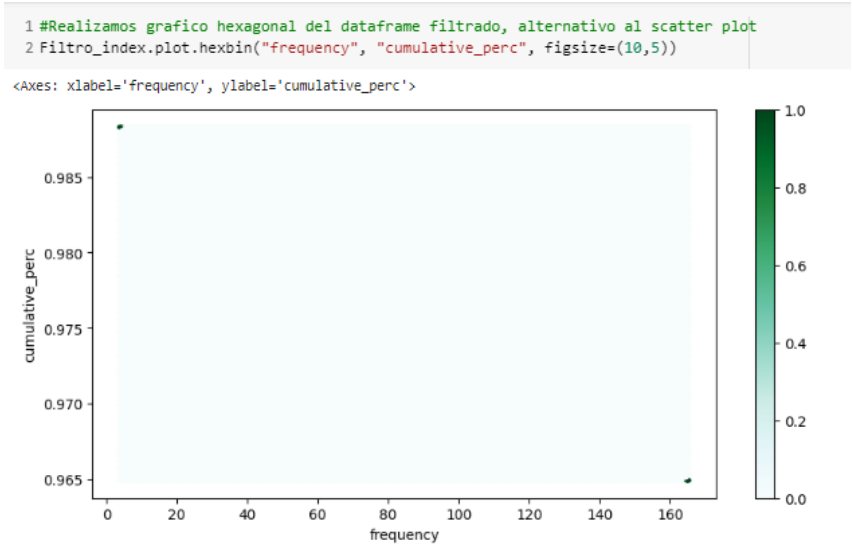


podemos decir que con información restante en su mayoría hubo un aumento por lo menos del 15% en el salario en una frecuencia mayor que los descuentos o la disminución de salarios a partir del 15% es decir

que se aumentó el salario más en más tiendas que en las tiendas que se disminuyó.

155_sales_registers_used_for

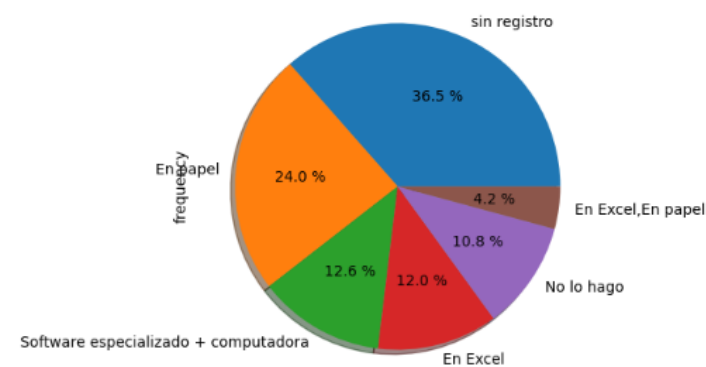
En esta gráfica observamos que el registro usado por las ventas es nulo ya que los valores que se presentan en la gráfica son muy extremos debido a que en su mayoría más del 90% de los encuestados no puso ninguna información respecto a esta categoría y menos del 3% puso que no registra ningún uso.



49_inventory_records

En esta gráfica podemos observar los inventarios de los registros que se realizan en las tiendas en su mayoría no tiene ninguna información respecto a esta categoría con un 36%,posteriormente la categoría que predomina es la de registros en papel

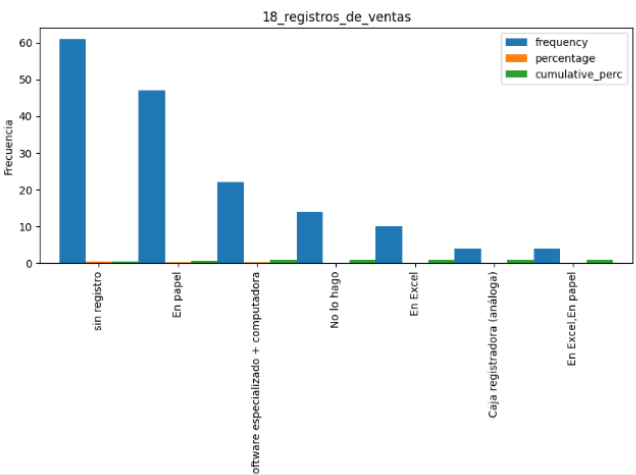
<Axes: ylabel='frequency'>



con un 24% un 12% se realizan en computadora o un software especializado y otro 12% en Excel, y solo un 10% no realiza ningún tipo de registro por lo que podemos observar que en su mayoría cerca del 75% de las tiendas no cuenta con un sistema de registros de inventarios por lo que presenta una necesidad y un problema por esas tiendas.

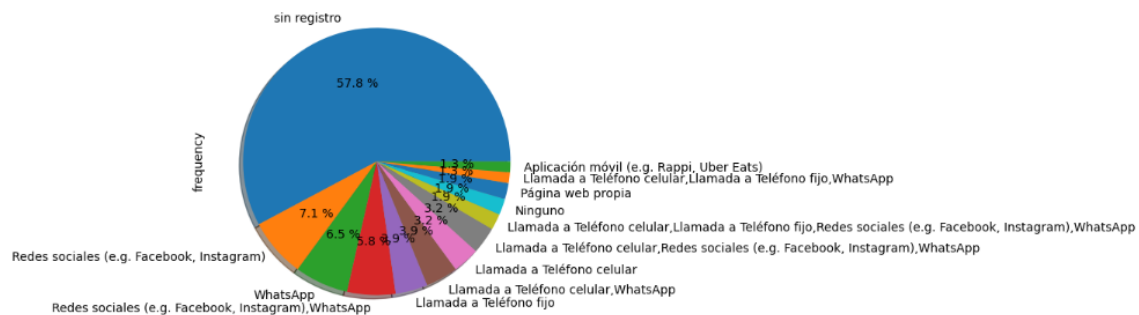
18_sales_records

al igual que la gráfica anterior los registros de las ventas en su



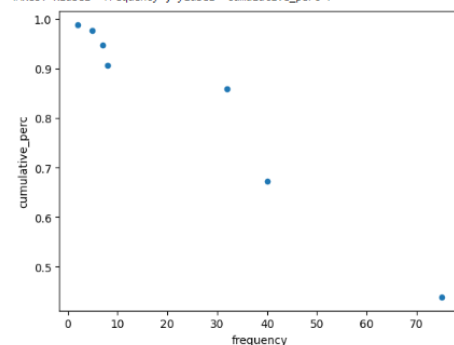
mayoría se hacen utilizando papel y menos del 25% se realizan a través de un software especializado computadora un programa de Excel y el 10% no realiza ningún tipo de registro de ventas

210_sales_channels

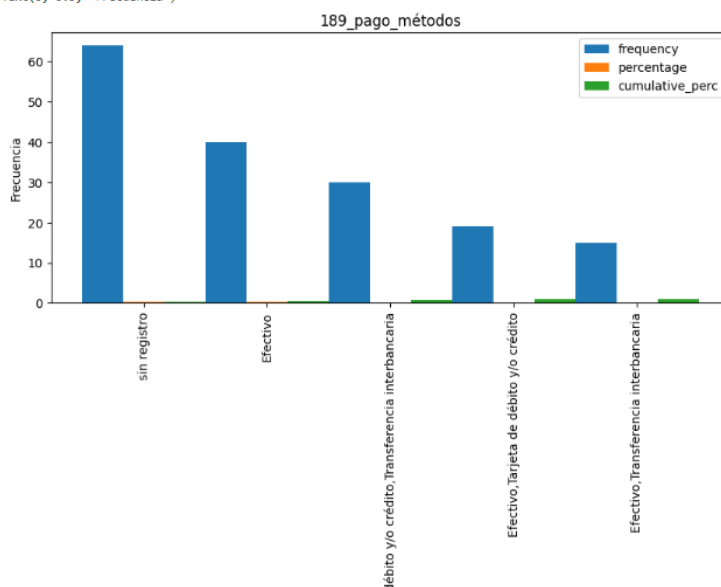


Dentro de la información registrada en el canal de venta podemos observar que en su mayoría el 7% utiliza redes sociales como instagram Facebook y un 6% adicional utiliza WhatsApp el restante utiliza métodos llamada telefónica y solamente el 1.3% utiliza aplicaciones móviles como Uber eats .

```
obj_1 = #Realizamos grafico de dispersion del dataframe filtrado
2 Filtro_index.plot("frequency", "cumulative_perc", kind="scatter")
<Axes: xlabel='frequency', ylabel='cumulative_perc'>
```



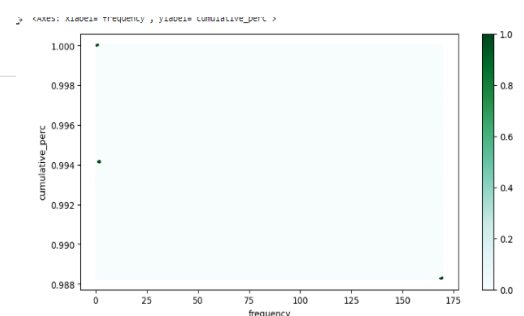
Text(0, 0.5, 'Frecuencia')



crédito o débito.

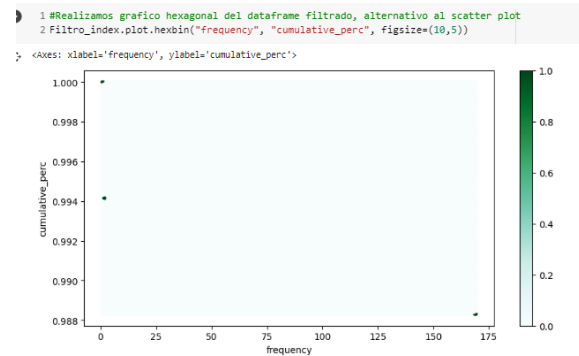
189_payment_methods

Los métodos de pago en su mayoría son en efectivo sin embargo las tarjetas de crédito y débito no tiene una amplia diferencia ante el efectivo. Por lo que un poquito menos de la mitad tienen la capacidad para recibir ventas con tarjetas de



28_prefered_payment_method

Significa podemos observar que los valores son nulos ya que cuentan sin registro o los pocos registros que hay son demasiado poco

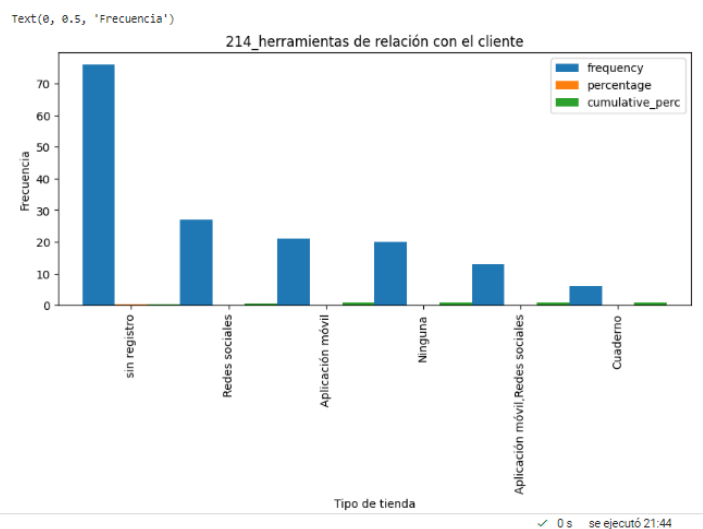


En esta tabla podemos observar que en su mayoría las tiendas ya utilizan herramientas digitales para relacionarse con el cliente sin embargo 20 tiendas no utilizan ningún tipo de relación con el cliente o herramienta y 10 tiendas solo utiliza un cuaderno como relación con el cliente

270_card_extra_charge

Al igual que la gráfica anterior en su mayoría son registros sin información los pocos registros que hay son muy pocos para tomarlos en cuenta

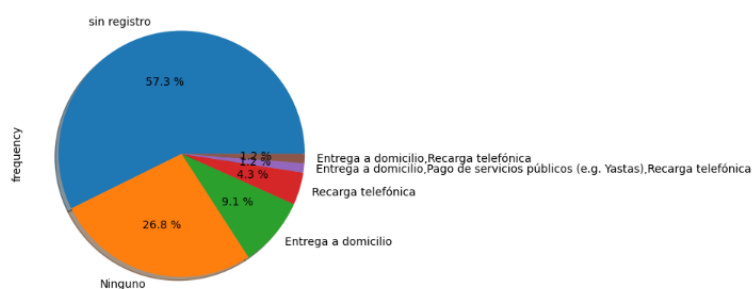
214_customer_relationship_tools



38_average_margin_profits

0

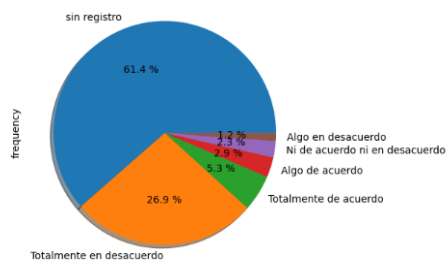
60_services



cuenta con recarga telefónica.

En esta gráfica podemos observar que el 10% de las tiendas registradas cuenta con entrega a domicilio mientras que solo el 5%

<Axes: ylabel='frequency'>



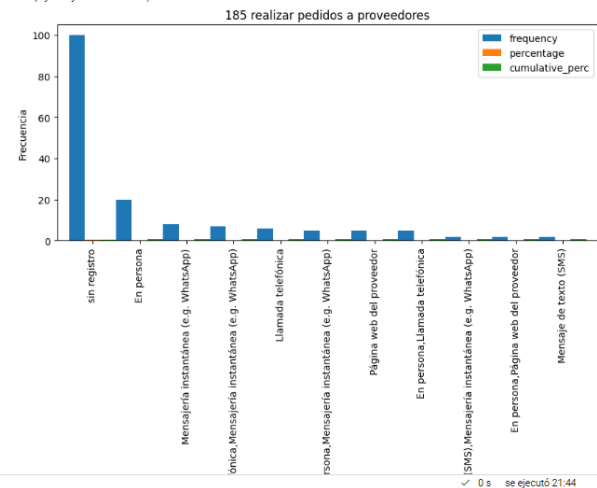
En serio que podemos observar las herramientas que utilizan las tiendas para realizar sus pedidos observando que en su gran mayoría las tienes ya utilizan dispositivos digitales para realizar sus pedidos y solamente 20 de las tiendas los pedidos se realizan en persona únicamente

312_payment_of_utilities

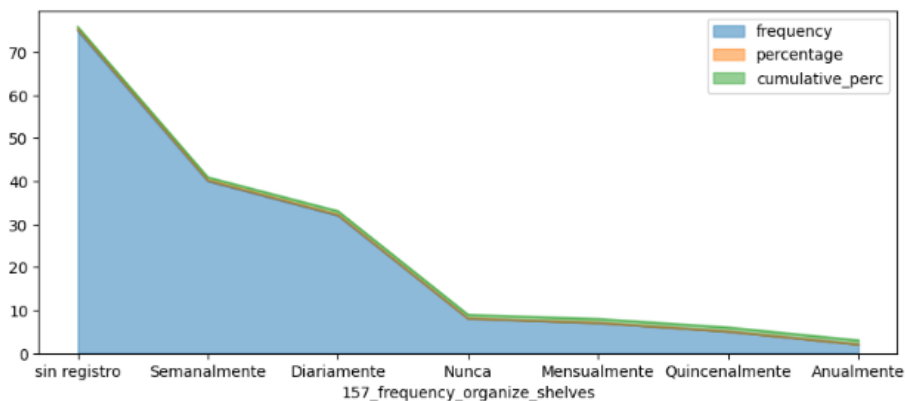
En ese grupo podemos observar que en su mayoría recién están en desacuerdo con el poca utilidades y solamente el 5% está totalmente de acuerdo

185_place_orders_suppliers

Text(0, 0.5, 'Frecuencia')



<Axes: xlabel='157_frequency_organize_shelves'>



157_frequency_organize_shelves Podemos observar la frecuencia en la que se organizan los estantes en su mayoría se realizan semanalmente ya que 40 de los registros de las tiendas lo realizan 35 lo realiza diariamente , 10 tiene lo realiza mensualmente, otras Hilda realizan quincenalmente y otras 10 solo lo realizan anualmente mientras que los que nunca realizan ningún organización son 10 de esta hacienda registradas.

