

## **Actividad 4 (Extracción de Características)**

### **HALLAZGOS IMPORTANTES**

Nicole Aryam Rodríguez A00831569

La base de datos contiene 106 columnas siendo estas 81 tipo object, 23 float y 2 tipo int. Los valores nulos de este data frame fueron tratados de la siguiente manera:

1. Se aplica el método fill a toda la database
2. Aquellos datos que se mantuvieron nulos después del primer método, se les relleno con un "-", esto ya que la instancia debe de ser única y no es posible rellenarla con el promedio u otro valor inventado.

Dejando nuestra database sin valores nulos, tratamos los valores atípicos por el procedimiento "cuantiles". Este método lo seleccione ya que el método de los intercuantiles es más robusto y adecuado para detectar valores atípicos extremos. Una vez identificados los valores atípicos en el data frame creado solo con valores numéricos, estos se rellenaron con el promedio de cada columna.

Por consiguiente se concatenaron el data frame de valores cuantitativos con el de cualitativos y es así como tenemos nuestro data frame limpio.

Proseguimos con la etapa de extracción de características donde se extraen características a partir de un análisis univariado a 20 distintas variables categóricas.

Estas fueron las 20 variables categóricas que elegí:

1. 229\_store\_picture
2. 99\_does\_the\_micro\_retailer\_exhibits\_products\_outside\_
3. 108\_does\_the\_micro\_retailer\_has\_a\_barred\_window\_
4. 102\_does\_the\_micro\_retailer\_sells\_fresh\_products\_
5. 2\_current\_permanent\_employees
6. 6\_change\_employees\_average\_salary\_last\_year
7. 49\_inventory\_records
8. 103\_number\_own\_fridges
9. 145\_number\_direct\_competitors
10. 186\_internet\_connection
11. 28\_prefered\_payment\_method
12. 269\_cash\_discount
13. 33\_credit\_to\_customers
14. 227\_agreement\_ecommerce\_platforms
15. 317\_home\_deliveries
16. 312\_payment\_of\_utilities
17. 53\_special\_conditions\_suppliers
18. 277\_payment\_method\_suppliers
19. 279\_changes\_procurement\_pandemic
20. 157\_frequency\_organize\_shelves

## HALLAZGOS DE CADA VARIABLE CATEGÓRICA (con su respectivo filtro y su gráfica)

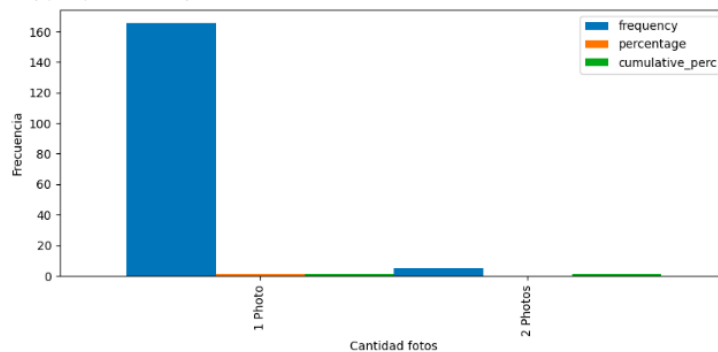
### 1. 229\_store\_picture

```
1 #Ajusto el indice de mi dataframe
2 index1= table1.set_index('229_store_picture')
3 index1
```

	frequency	percentage	cumulative_perc
229_store_picture			
1 Photo	166	0.97076	0.97076
2 Photos	5	0.02924	1.00000

```
[ ] 1 #Realizamos grafico de barras del dataframe filtrado
2 index1.plot(kind = 'bar', width=1, figsize=(10,4))
3 plt.xlabel('Cantidad fotos')
4 plt.ylabel('Frecuencia')
```

Text(0, 0.5, 'Frecuencia')



El 97% de las micro retailers tienen 1 foto de la tienda

### 2. 99\_does\_the\_micro\_retailer\_exhibits\_products\_outside\_

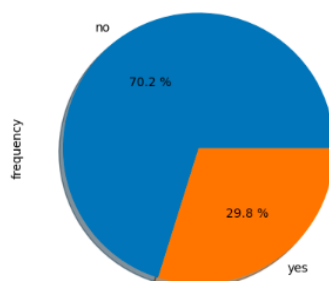
99\_does\_the\_micro\_retailer\_exhibits\_products\_outside\_

```
[ ] 1 table2= freq_tbl(data_clean['99_does_the_micro_retailer_exhibits_products_outside_'])
2 index2= table2.set_index('99_does_the_micro_retailer_exhibits_products_outside_')
3 index2
```

	frequency	percentage	cumulative_perc
99_does_the_micro_retailer_exhibits_products_outside_			
no	120	0.701754	0.701754
yes	51	0.298246	1.000000

```
1 #Realizamos grafico de dispersión del dataframe filtrado
2 index2["frequency"].plot(kind='pie', figsize=(10,5), shadow=True, autopct="%0.1f %%")
```

<Axes: ylabel='frequency'>



El 70% de las micro retailers NO exhiben productos afuera de la tienda, mientras que el 30% SI lo hace.

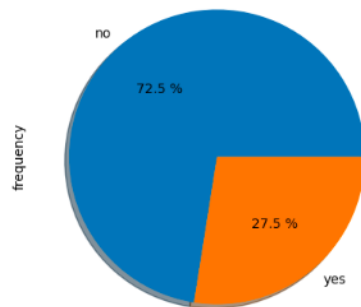
### 3. 108\_does\_the\_micro\_retailer\_has\_a\_barred\_window\_

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

```
[ ] 1 table3= freq_tbl(data_clean['108_does_the_micro_retailer_has_a_barred_window_'])
     2 index3= table3.set_index('108_does_the_micro_retailer_has_a_barred_window_')
     3 index3
```

	frequency	percentage	cumulative_perc
108_does_the_micro_retailer_has_a_barred_window_			
no	124	0.725146	0.725146
yes	47	0.274854	1.000000

```
1 index3["frequency"].plot(kind='pie', figsize=(10,5), shadow=True, autopct="%0.1f %%")
<Axes: ylabel='frequency'>
```



El 72% de las micro retailers NO tienen ventanas con barras para protegerlas mientras que el 28% SI lo hace.

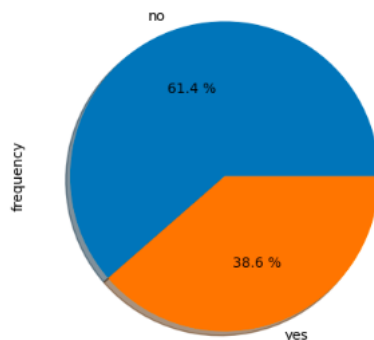
### 4. 102\_does\_the\_micro\_retailer\_sells\_fresh\_products\_

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

```
[ ] 1 table4= freq_tbl(data_clean['102_does_the_micro_retailer_sells_fresh_products_'])
     2 index4= table4.set_index('102_does_the_micro_retailer_sells_fresh_products_')
     3 index4
```

	frequency	percentage	cumulative_perc
102_does_the_micro_retailer_sells_fresh_products_			
no	105	0.614035	0.614035
yes	66	0.385965	1.000000

```
1 index4["frequency"].plot(kind='pie', figsize=(10,5), shadow=True, autopct="%0.1f %%")
<Axes: ylabel='frequency'>
```



El 61% de las micro retailers NO venden productos frescos mientras que el 39% SI lo hace.

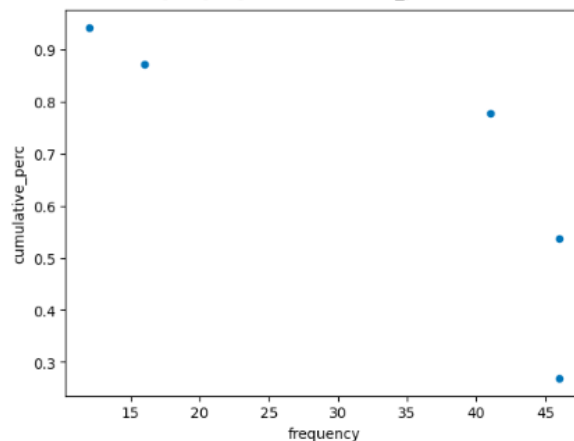
## 5. 2\_current\_permanent\_employees

```
[ ] 1 filtro5= index5[index5['frequency']>11]
    2 filtro5
```

	frequency	percentage	cumulative_perc
2_current_permanent_employees			
1.0	46	0.269006	0.269006
0.0	46	0.269006	0.538012
2.0	41	0.239766	0.777778
4.0	16	0.093567	0.871345
3.0	12	0.070175	0.941520

```
1 filtro5.plot("frequency", "cumulative_perc", kind="scatter")
```

```
<Axes: xlabel='frequency', ylabel='cumulative_perc'>
```



El 94% de los micro retailers tienen menos de 4 empleados permanentes, la gran mayoría tienen entre 0 y 1 empleado permanente.

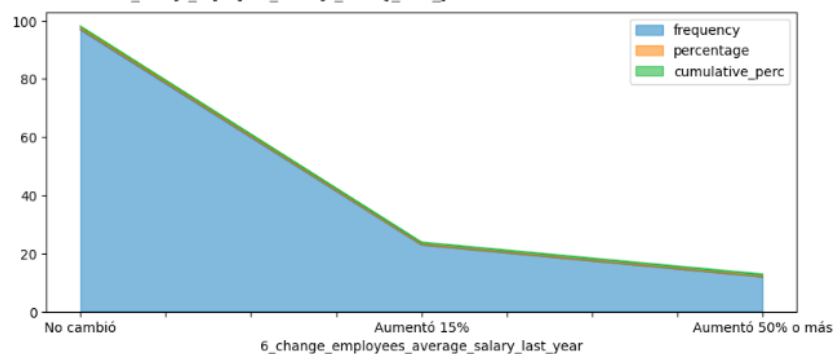
## 6. 6\_change\_employees\_average\_salary\_last\_year

```
] 1 filtro6= index6[index6['frequency']>10]
    2 filtro6
```

	frequency	percentage	cumulative_perc
6_change_employees_average_salary_last_year			
No cambió	97	0.567251	0.567251
Aumentó 15%	23	0.134503	0.701754
Aumentó 50% o más	12	0.070175	0.771930

```
] 1 filtro6.plot(kind='area', figsize=(10,4),alpha = 0.5)
```

```
<Axes: xlabel='6_change_employees_average_salary_last_year'>
```



El 56% de las micro retailers no cambiaron el salario anual promedio de sus empleados, mientras que para un 7% este aumento 50% o más

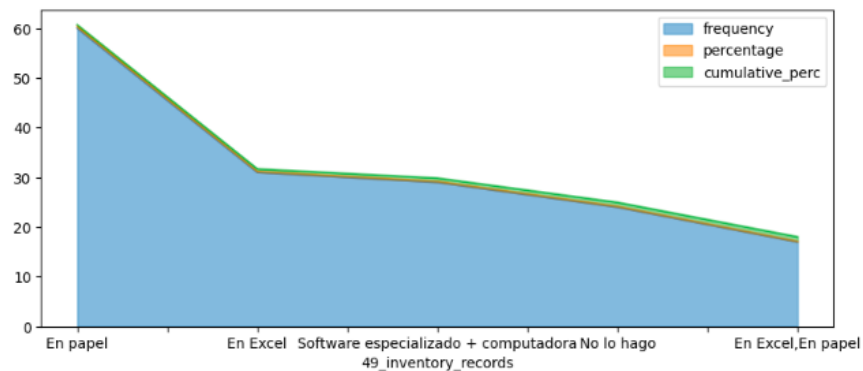
## 7. 49\_inventory\_records

```
[ ] 1 filtro7= index7[index7['frequency']>10]
    2 filtro7
```

	frequency	percentage	cumulative_perc
49_inventory_records			
En papel	60	0.350877	0.350877
En Excel	31	0.181287	0.532164
Software especializado + computadora	29	0.169591	0.701754
No lo hago	24	0.140351	0.842105
En Excel,En papel	17	0.099415	0.941520

```
1 filtro7.plot(kind='area', figsize=(10,4),alpha = 0.5)
```

```
<Axes: xlabel='49_inventory_records'>
```



El 35% de los micro retailers mantiene sus cuentas de inventario en papel , otro 18% en excel, un 16% en un software especializado, y un 14% no mantiene estas cuentas.

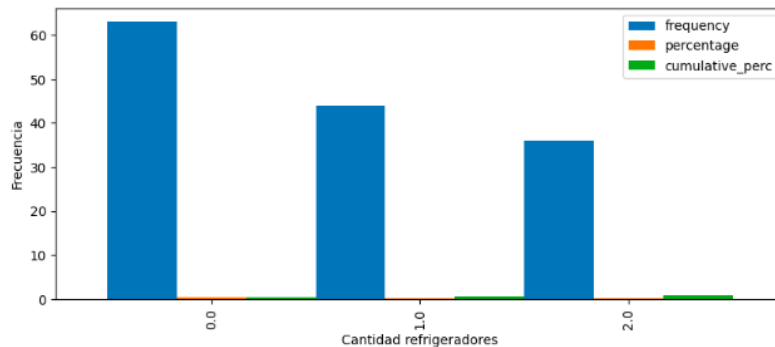
## 8. 103\_number\_own\_fridges

```
[ ] 1 filtro8= index8[index8['frequency']>10]
    2 filtro8
```

	frequency	percentage	cumulative_perc
103_number_own_fridges			
0.0	63	0.368421	0.368421
1.0	44	0.257310	0.625731
2.0	36	0.210526	0.836257

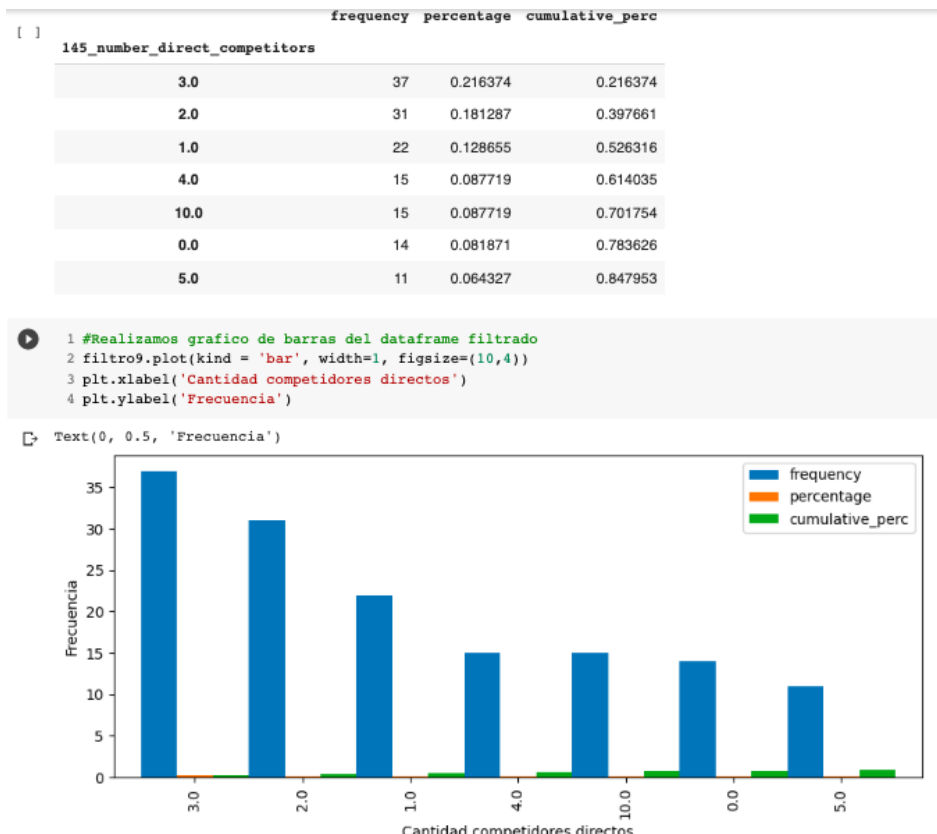
```
1 #Realizamos grafico de barras del dataframe filtrado
2 filtro8.plot(kind = 'bar', width=1, figsize=(10,4))
3 plt.xlabel('Cantidad refrigeradores')
4 plt.ylabel('Frecuencia')
```

```
Text(0, 0.5, 'Frecuencia')
```



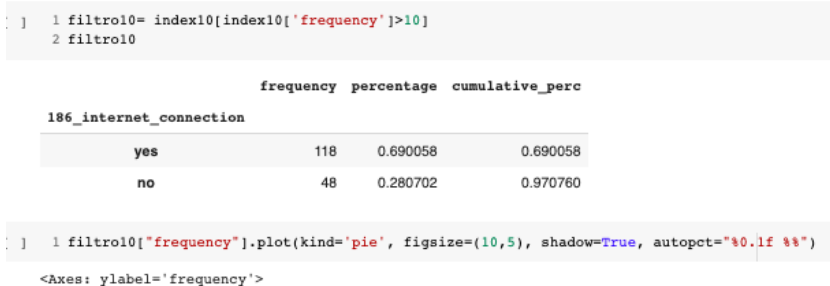
El 36% de los micro retailers tienen 0 refrigeradores, mientras que un 40% tiene ya sea 1 o 2 refrigeradores en tienda.

## 9. 145\_number\_direct\_competitors



El 84% de los micro retailers dicen tener entre 0 y 10 competidores directos. El 21% tiene 3 competidores, el 18% 2 competidores, un 8% dice tener 0 competidores directos y casi el 9% dice tener 10.

## 10. 186\_internet\_connection



Un 69% de los micro retailers tienen conexión a internet, mientras que un 28% no la tienen

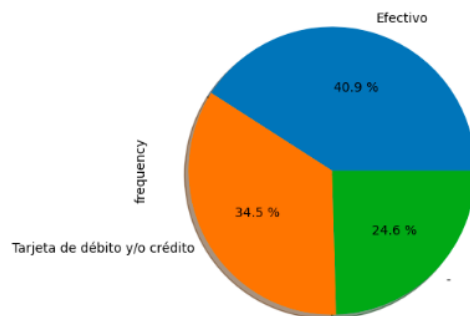
## 11. 28\_prefered\_payment\_method

28\_prefered\_payment\_method

```
[ ] 1 table11= freq_tbl(data_clean['28_prefered_payment_method'])
    2 index11= table11.set_index('28_prefered_payment_method')
    3 index11
```

	frequency	percentage	cumulative_perc
28_prefered_payment_method			
Efectivo	70	0.409357	0.409357
Tarjeta de débito y/o crédito	59	0.345029	0.754386
-	42	0.245614	1.000000

```
1 index11["frequency"].plot(kind='pie', figsize=(10,5), shadow=True, autopct="%0.1f %%")
<Axes: ylabel='frequency'>
```



La mayoría de los micro retailers prefieren el pago en efectivo mientras que un 34.5% prefieren con tarjeta de débito y/o crédito

## 12. 269\_cash\_discount

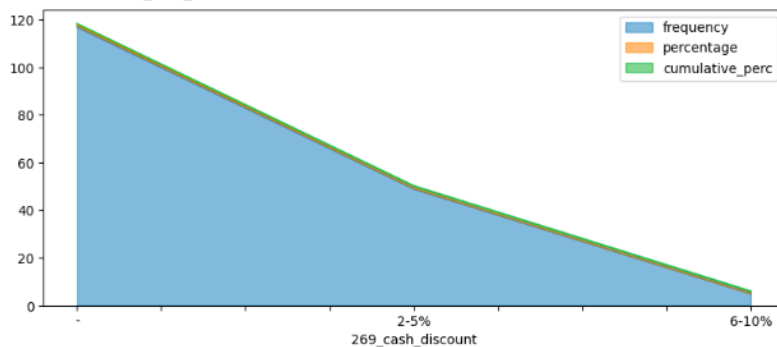
269\_cash\_discount

```
[ ] 1 table12= freq_tbl(data_clean['269_cash_discount'])
    2 index12= table12.set_index('269_cash_discount')
    3 index12
```

	frequency	percentage	cumulative_perc
269_cash_discount			
-	117	0.684211	0.684211
2-5%	49	0.286550	0.970760
6-10%	5	0.029240	1.000000

```
1 index12.plot(kind='area', figsize=(10,4), alpha = 0.5)
```

<Axes: xlabel='269\_cash\_discount'>



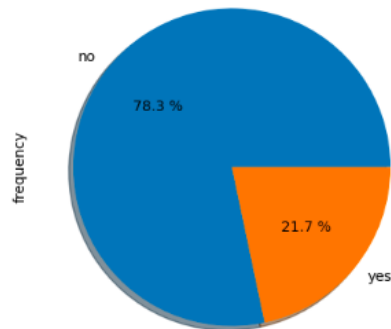
El 28% de los micro retailers dan aproximadamente un 2-5% de cash-discount, mientras que la gran mayoría no tiene registro pero asumiría que no dan cash discount.

### 13. 33\_credit\_to\_customers

```
[ ] 1 filtrol3= index13[index13['frequency']>10]
     2 filtrol3
```

	frequency	percentage	cumulative_perc
33_credit_to_customers			
no	130	0.760234	0.760234
yes	36	0.210526	0.970760

```
[ ] 1 filtrol3["frequency"].plot(kind='pie', figsize=(10,5), shadow=True, autopct="%0.1f %%")
    <Axes: ylabel='frequency'>
```



El 76% de los micro retailers NO dan crédito a los consumidores, mientras que el 21% SI dan crédito.

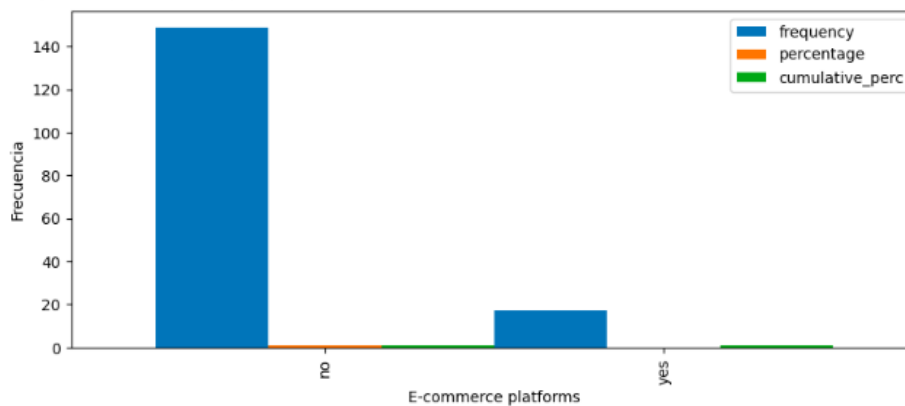
### 14. 227\_agreement\_ecommerce\_platforms

```
[ ] 1 filtrol4= index14[index14['frequency']>10]
     2 filtrol4
```

	frequency	percentage	cumulative_perc
227_agreement_ecommerce_platforms			
no	149	0.871345	0.871345
yes	17	0.099415	0.970760

```
1 #Realizamos grafico de barras del dataframe filtrado
2 filtrol4.plot(kind = 'bar', width=1, figsize=(10,4))
3 plt.xlabel('E-commerce platforms')
4 plt.ylabel('Frecuencia')
```

Text(0, 0.5, 'Frecuencia')



El 87% de los micro retailers NO tienen un acuerdo con plataformas de e-commerce mientras que 9% SI lo tienen.



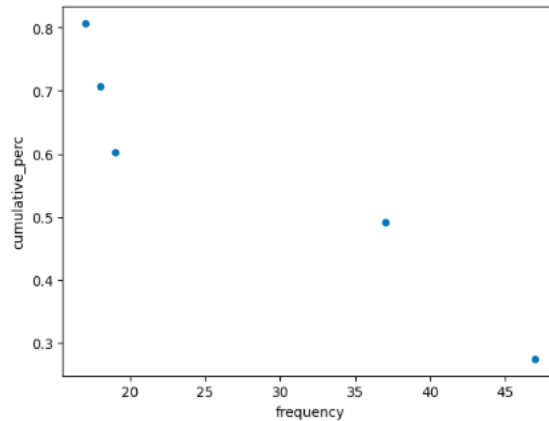
## 15. 317\_home\_deliveries

```
[ ] 1 filtro15= index15[index15['frequency']>10]
    2 filtro15
```

	frequency	percentage	cumulative_perc
317_home_deliveries			
	7.0	47	0.274854
	2.0	37	0.216374
	8.0	19	0.111111
	-	18	0.105263
	40.0	17	0.099415

```
1 filtro15.plot("frequency", "cumulative_perc", kind="scatter")
```

```
<Axes: xlabel='frequency', ylabel='cumulative_perc'>
```



El 27% de los micro retailers hacen 7 home deliveries, mientras que solo 9% hacen 40 home deliveries.

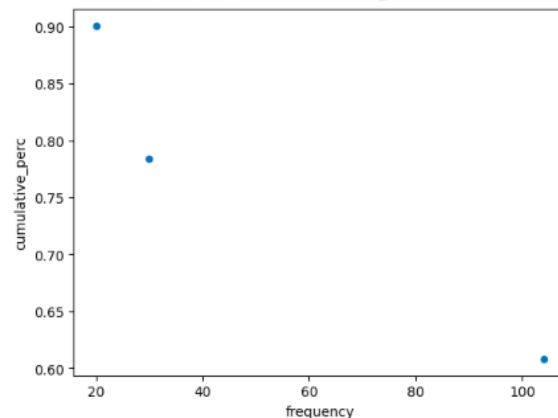
## 16. 312\_payment\_of\_utilities

```
[ ] 1 filtro16= index16[index16['frequency']>10]
    2 filtro16
```

	frequency	percentage	cumulative_perc
312_payment_of_utilities			
Totalmente en desacuerdo	104	0.608187	0.608187
Algo de acuerdo	30	0.175439	0.783626
Totalmente de acuerdo	20	0.116959	0.900585

```
1 filtro16.plot("frequency", "cumulative_perc", kind="scatter")
```

```
<Axes: xlabel='frequency', ylabel='cumulative_perc'>
```



El 60%, la gran mayoría está totalmente de desacuerdo con el pago de utilidades, mientras que un 11% está totalmente de acuerdo con este pago.

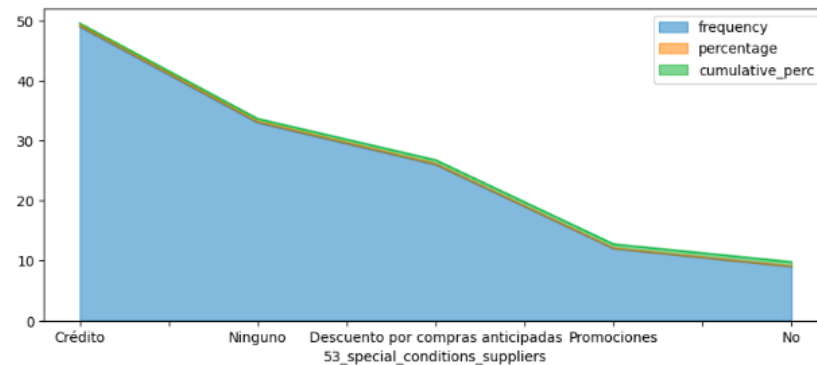
## 17. 53\_special\_conditions\_suppliers

```
1 filtrol7= index17[index17['frequency']>8]
2 filtrol7
```

	frequency	percentage	cumulative_perc
<b>53_special_conditions_suppliers</b>			
Crédito	49	0.286550	0.286550
Ninguno	33	0.192982	0.479532
Descuento por compras anticipadas	26	0.152047	0.631579
Promociones	12	0.070175	0.701754
No	9	0.052632	0.754386

```
1 filtrol7.plot(kind='area', figsize=(10,4),alpha = 0.5)
```

```
<Axes: xlabel='53_special_conditions_suppliers'>
```



El 28% de los micro retailers tienen el crédito como condición especial para sus proveedores, un 19% no tiene ninguna condición especial y un 15% tiene un descuento por compras anticipadas.

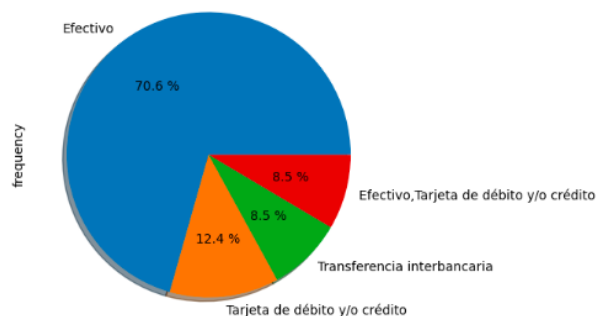
## 18. 277\_payment\_method\_suppliers

```
1 filtrol8= index18[index18['frequency']>10]
2 filtrol8
```

	frequency	percentage	cumulative_perc
<b>277_payment_method_suppliers</b>			
Efectivo	108	0.631579	0.631579
Tarjeta de débito y/o crédito	19	0.111111	0.742690
Transferencia interbancaria	13	0.076023	0.818713
Efectivo,Tarjeta de débito y/o crédito	13	0.076023	0.894737

```
1 filtrol8["frequency"].plot(kind='pie', figsize=(10,5), shadow=True, autopct="%0.1f %%")
```

```
<Axes: ylabel='frequency'>
```



El 63% de los proveedores de los micro retailers tienen como método de pago el efectivo, 11% tarjeta de débito y/o crédito y un 7% transferencia interbancaria.

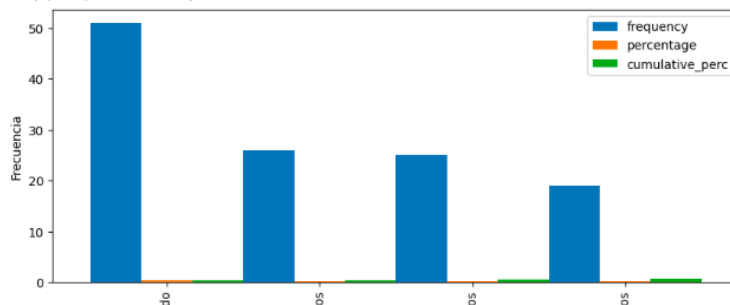
## 19. 279\_changes\_procurement\_pandemic

```
[ ] 1 filtro19= index19[index19['frequency']>18]
    2 filtro19
```

	frequency	percentage	cumulative_perc
279_changes_procurement_pandemic			
Disminuir la cantidad del pedido	51	0.298246	0.298246
Disminución de la variedad de productos,Disminuir la cantidad del pedido,Disminuir la frecuencia de pedidos	26	0.152047	0.450292
Disminuir la frecuencia de pedidos	25	0.146199	0.596491
Disminuir la cantidad del pedido,Disminuir la frecuencia de pedidos	19	0.111111	0.707602

```
1 #Realizamos grafico de barras del dataframe filtrado
2 filtro19.plot(kind = 'bar', width=1, figsize=(10,4))
3 plt.xlabel('Cambios pandemia')
4 plt.ylabel('Frecuencia')
```

Text(0, 0.5, 'Frecuencia')



El 70% de los micro retailers tuvieron cambios de aprovisionamiento debido a la pandemia: entre estos se encuentran; disminuir la cantidad del pedido, la frecuencia del pedido, y la variedad de productos.

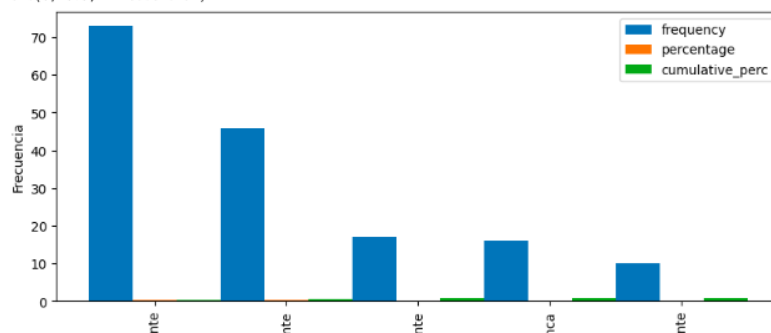
## 20. 157\_frequency\_organize\_shelves

```
[ ] 1 filtro20= index20[index20['frequency']>9]
    2 filtro20
```

	frequency	percentage	cumulative_perc
157_frequency_organize_shelves			
Semanalmente	73	0.426901	0.426901
Diariamente	46	0.269006	0.695906
Mensualmente	17	0.099415	0.795322
Nunca	16	0.093567	0.888889
Quincenalmente	10	0.058480	0.947368

```
1 #Realizamos grafico de barras del dataframe filtrado
2 filtro20.plot(kind = 'bar', width=1, figsize=(10,4))
3 plt.xlabel('Frecuencia de Organizacion Estantes')
4 plt.ylabel('Frecuencia')
```

Text(0, 0.5, 'Frecuencia')



El 42% de los micro retailers organizan sus estantes semanalmente, un 26% diariamente, un 9% mensualmente y otro 9% nunca.