



Actividad 4

EXTRACCIÓN *de* CARACTERÍSTICAS

Renata Pilar Gómez Castillo

EXTRACCIÓN DE CARACTERÍSTICAS

Para esta actividad se trabajó con el archivo de Excel de *microretailer_mit_lift_lab.xlsx*. Los datos tiene valores nulos y atípicos por los que primero voy a tratarlos.

Primero comencé separando los datos en dos dataframes diferentes, uno para los datos cualitativos y otro para los datos numéricos.

VALORES NULOS

Para los valores nulos de los datos cualitativos utilice la función `.fillna()` y los rellene con un string '--'. De esta manera no alteró los datos que se tienen.

Mientras que para los valores nulos de los datos cualitativos, con la función de `.fillna()` los rellene con un 0. De igual manera para no afectar los datos que tiene el dataframe original.

OUTLIERS

Para los valores atípicos decidí utilizar el método de cuartiles. De esta manera pude borrar los outliers que se encontraban en los datos numéricos.

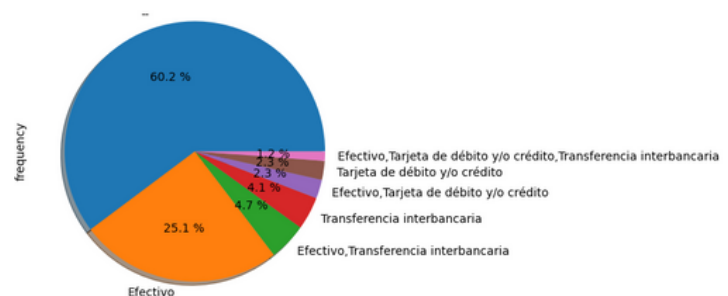
Una vez que los valores nulos y atípicos fueron tratados hice un merge con ambos dataframes (cualitativos y numérico) para tener mi dataframe original.

EXTRACCIÓN DE CARACTERÍSTICAS

Después extraje características a partir de un análisis univariado a 20 variables categóricas.

• 277_PAYMENT_METHOD_SUPPLIERS

	frequency	percentage	cumulative_perc
277_payment_method_suppliers			
--	103	0.602339	0.602339
Efectivo	43	0.251462	0.853801
Efectivo, Transferencia interbancaria	8	0.046784	0.900585
Transferencia interbancaria	7	0.040936	0.941520
Efectivo, Tarjeta de débito y/o crédito	4	0.023392	0.964912
Tarjeta de débito y/o crédito	4	0.023392	0.988304
Efectivo, Tarjeta de débito y/o crédito, Transferencia interbancaria	2	0.011696	1.000000

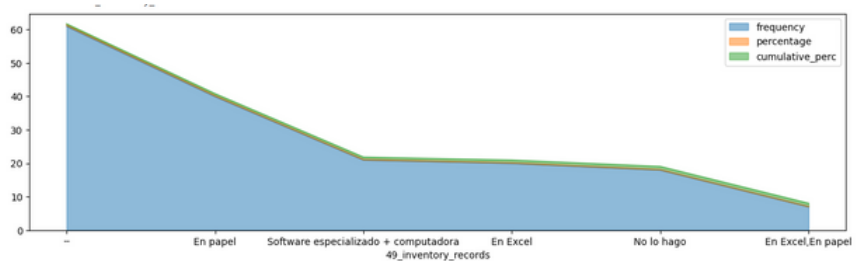


Podemos observar que el proveedor de métodos de pago con mayor frecuencia es efectivos, seguido por la combinación de efectivo con transferencia interbancaria.

El proveedor de métodos de pago menos utilizado es la combinación de efectivo, tarjeta de crédito/débito y transferencia interbancaria.

• 49_INVENTORY_RECORDS

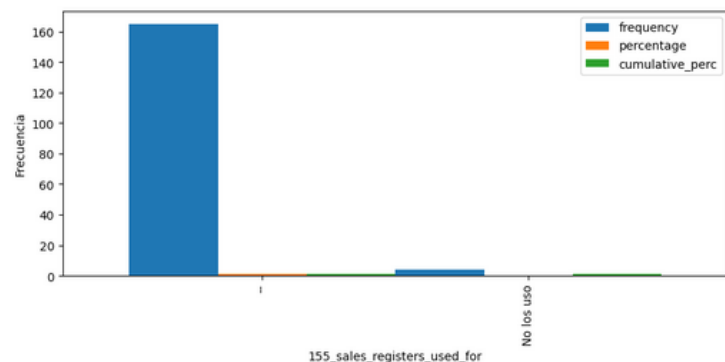
	frequency	percentage	cumulative_perc
49_inventory_records			
--	61	0.356725	0.356725
En papel	40	0.233918	0.590643
Software especializado + computadora	21	0.122807	0.713450
En Excel	20	0.116959	0.830409
No lo hago	18	0.105263	0.935673
En Excel,En papel	7	0.040936	0.976608



Sobre los registros de inventario se observa en la gráfica de área que la mayoría de los registros se encuentran en papel con una frecuencia de 40. Mientras que la segunda forma de registro más usada es software especializado + computadora

• 155_SALES_REGISTERS_USED_FOR

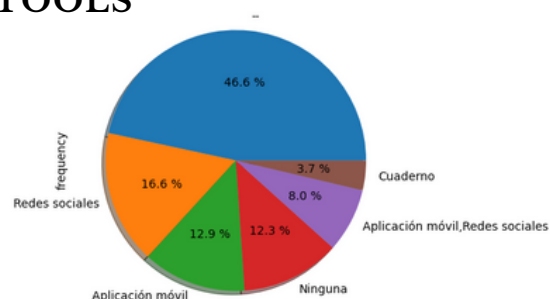
	frequency	percentage	cumulative_perc
155_sales_registers_used_for			
--	165	0.964912	0.964912
No los uso	4	0.023392	0.988304



Para la variable 155 podemos observar que la mayoría de los datos fueron '--' al haber sido nulos y rellenados con un string. Las respuestas que se obtuvieron fue 'no los uso' con una frecuencia de 4.

• 214_CUSTOMER_RELATIONSHIP_TOOLS

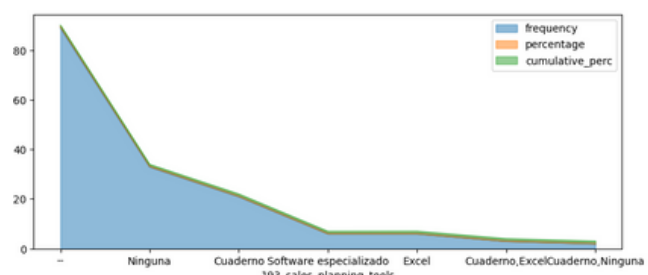
	frequency	percentage	cumulative_perc
214_customer_relationship_tools			
--	76	0.444444	0.444444
Redes sociales	27	0.157895	0.602339
Aplicación móvil	21	0.122807	0.725146
Ninguna	20	0.116959	0.842105
Aplicación móvil,Redes sociales	13	0.076023	0.918129
Cuaderno	6	0.035088	0.953216



En el gráfico de pastel puedo notar que la herramienta de relación que más utilizan los usuarios son las redes sociales, seguido por las aplicaciones móviles. Aunque es importante mencionar que el 12,3% mencionan que no usan ninguna de las herramientas. Y por último la menos utilizada es el cuaderno.

• 193_SALES_PLANNING_TOOLS

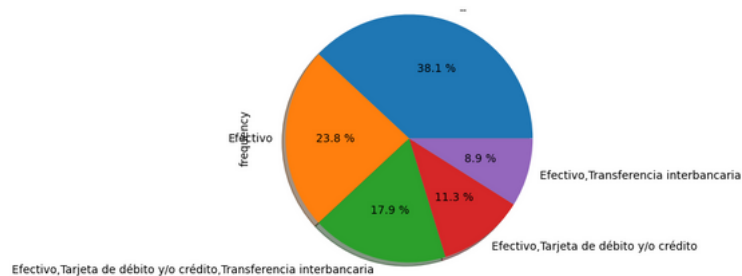
	frequency	percentage	cumulative_perc
193_sales_planning_tools			
--	89	0.520468	0.520468
Ninguna	33	0.192982	0.713450
Cuaderno	21	0.122807	0.836257
Software especializado	6	0.035088	0.871345
Excel	6	0.035088	0.906433
Cuaderno,Excel	3	0.017544	0.923977
Cuaderno,Ninguna	2	0.011696	0.935673



Con esta variable observó que la mayoría de las personas no utilizan ninguna de las herramientas de planeación de ventas, teniendo una frecuencia de 33 registros. Mientras que la herramienta de cuaderno tiene una frecuencia de 21 registros.

• 189 PAYMENT METHODS

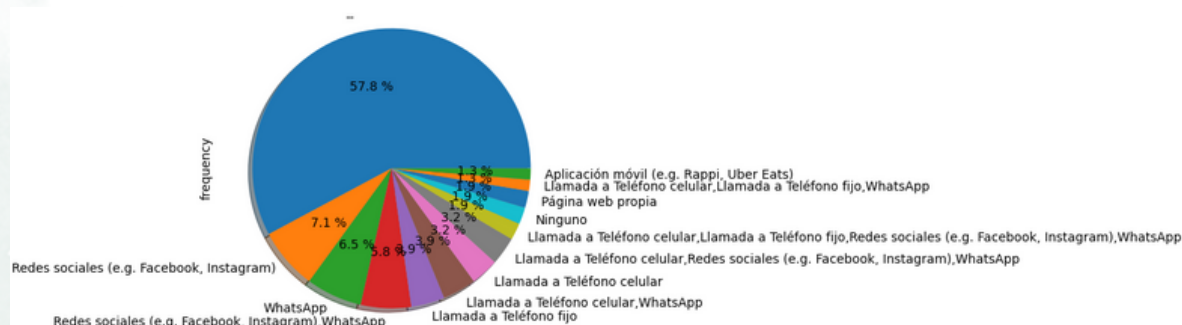
	frequency	percentage	cumulative_perc
189_payment_methods			
--	64	0.374269	0.374269
Efectivo	40	0.233918	0.608187
Efectivo,Tarjeta de débito y/o crédito,Transferencia interbancaria	30	0.175439	0.783626
Efectivo,Tarjeta de débito y/o crédito	19	0.111111	0.894737
Efectivo,Transferencia interbancaria	15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



Ahora veremos sobre los métodos de pago, donde el efectivo es el principal, representando el 23.8%, seguido por la combinación de efectivo, tarjeta de débito / crédito y transferencia interbancaria representando el 17.9%

- 210 SALES CHANNELS

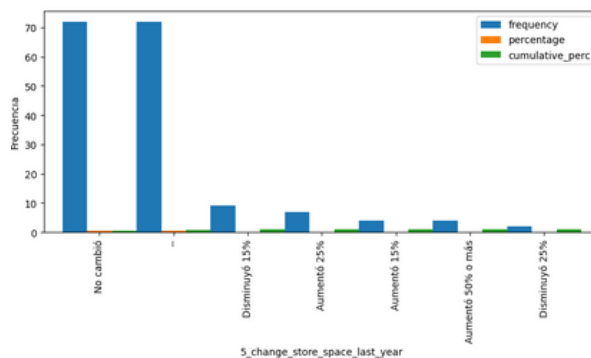
	frequency	percentage	cumulative_perc
210_sales_channels			
--	89	0.520468	0.520468
Redes sociales (e.g. Facebook, Instagram)	11	0.064327	0.584795
WhatsApp	10	0.058480	0.643275
Redes sociales (e.g. Facebook, Instagram),WhatsApp	9	0.052632	0.695906
Llamada a Teléfono fijo	6	0.035088	0.730994
Llamada a Teléfono celular,WhatsApp	6	0.035088	0.766082
Llamada a Teléfono celular	5	0.029240	0.795322
Llamada a Teléfono celular,Redes sociales (e.g. Facebook, Instagram),WhatsApp	5	0.029240	0.824561
Llamada a Teléfono celular,Llamada a Teléfono fijo,Redes sociales (e.g. Facebook, Instagram),WhatsApp	3	0.017544	0.842105
Ninguno	3	0.017544	0.859649
Página web propia	3	0.017544	0.877193
Llamada a Teléfono celular,Llamada a Teléfono fijo,WhatsApp	2	0.011696	0.888889
Aplicación móvil (e.g. Rappi, Uber Eats)	2	0.011696	0.900585



Se puede notar que el canal de ventas más utilizado es redes sociales, seguido por la combinación de redes sociales y whatsapp. Donde la menos utilizada son las aplicaciones móviles, representando únicamente el 1.3% de los datos.

• 5 CHANGE STORE SPACE LAST YEAR

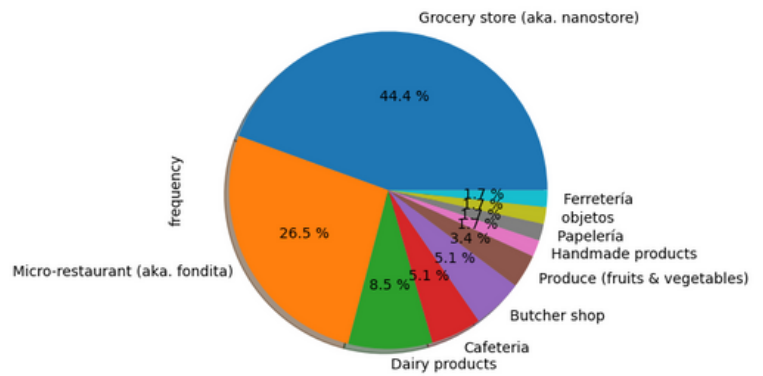
	frequency	percentage	cumulative_perc
5_change_store_space_last_year			
No cambió	72	0.421053	0.421053
--	72	0.421053	0.842105
Disminuyó 15%	9	0.052632	0.894737
Aumentó 25%	7	0.040936	0.935673
Aumentó 15%	4	0.023392	0.959064
Aumentó 50% o más	4	0.023392	0.982456
Disminuyó 25%	2	0.011696	0.994152



En esta variable podemos observar en la gráfica de barras y tabla de frecuencia a las tiendas que cambiaron de espacio/tamaño en el último año. Podemos notar que la mayoría de las tiendas (72 tiendas) no cambiaron, mientras que 9 de ellas disminuyeron un 15%, 7 aumentaron espacio un 25%, 4 tiendas aumentaron 15% y otras 4 un 50% o más de espacio.

• 232_TYPE_OF_STORE

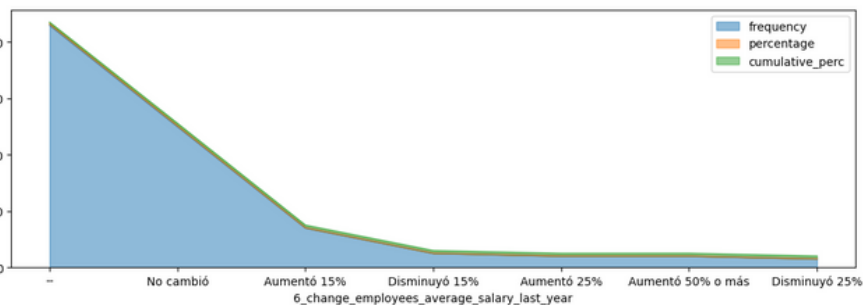
232_type_of_store	frequency	percentage	cumulative_perc
Grocery store (aka. nanostore)	52	0.304094	0.304094
Micro-restaurant (aka. fondita)	31	0.181287	0.485380
Dairy products	10	0.058480	0.543860
Cafeteria	6	0.035088	0.578947
Butcher shop	6	0.035088	0.614035
Produce (fruits & vegetables)	4	0.023392	0.637427
Handmade products	2	0.011696	0.649123
Papeleria	2	0.011696	0.660819
objetos	2	0.011696	0.672515
Ferreteria	2	0.011696	0.684211



Podemos observar que la mayoría de las tiendas 44.4% (52 tiendas) son tiendas de conveniencia, seguido de micro-restaurantes (26.5). El tipo de tiendas que tiene menos frecuencia son las ferreterías, objetos, y papelerías, representando el 1.7%

• 6_CHANGE_EMPLOYEES_AVERAGE_SALARY_LAST_YEAR

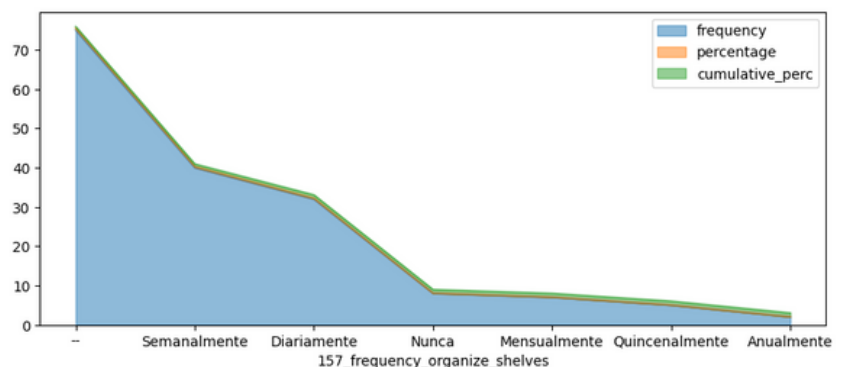
6_change_employees_average_salary_last_year	frequency	percentage	cumulative_perc
--	86	0.502924	0.502924
No cambió	50	0.292398	0.795322
Aumentó 15%	14	0.081871	0.877193
Disminuyó 15%	5	0.029240	0.906433
Aumentó 25%	4	0.023392	0.929825
Aumentó 50% o más	4	0.023392	0.953216
Disminuyó 25%	3	0.017544	0.970760



En cuanto al cambio de salario promedio de los empleados en el último año, 50 de las tiendas no tuvieron un cambio, 14 de las tiendas aumentaron un 15% y 5 de ellas disminuyeron un 15%. De todas las tiendas solo 4 de ellas aumentaron un 50% o más y solo 3 de ellas lo disminuyeron un 25%

• 157_FREQUENCY_ORGANIZE_SHELVES

157_frequency_organize_shelves	frequency	percentage	cumulative_perc
--	75	0.438596	0.438596
Semanalmente	40	0.233918	0.672515
Diariamente	32	0.187135	0.859649
Nunca	8	0.046784	0.906433
Mensualmente	7	0.040936	0.947368
Quincenalmente	5	0.029240	0.976608
Anualmente	2	0.011696	0.988304

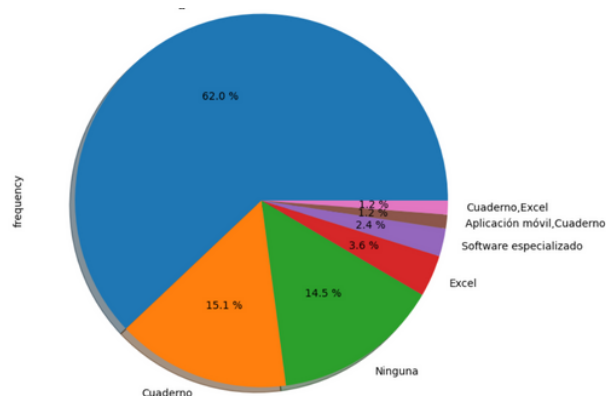


En cuanto a la frecuencia en el que se ordena los estantes la mayoría de las tiendas (40 tiendas) las ordenan semanalmente, seguido de una frecuencia diaria con 32 tiendas.

Puedo observar que 8 de las tiendas mencionan que nunca ordenar los estantes y 2 de ellas lo hacen con una frecuencia anual. Este sería un dato interesante a analizar en comparación con las ventas que tienen las tiendas, para así observar si la frecuencia con el que organizan los estantes tiene algún impacto en sus ventas.

• 192_PROCURUREMENT_PLANNING_TOOLS

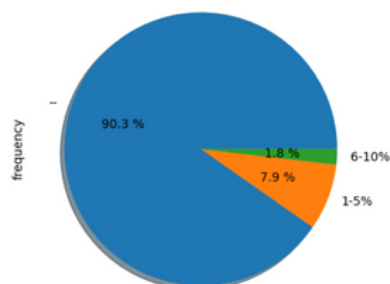
	frequency	percentage	cumulative_perc
192_procurement_planning_tools			
--	103	0.602339	0.602339
Cuaderno	25	0.146199	0.748538
Ninguna	24	0.140351	0.888889
Excel	6	0.035088	0.923977
Software especializado	4	0.023392	0.947368
Aplicación móvil,Cuaderno	2	0.011696	0.959064
Cuaderno,Excel	2	0.011696	0.970760



En cuanto las herramientas de planificación de adquisiciones la mayoría utiliza cuaderno, seguida de ninguna, después de excel (6) y la menos usada una combinación de cuaderno y excel.

• 66_PERCENTAJE_SERVICES_INCOME

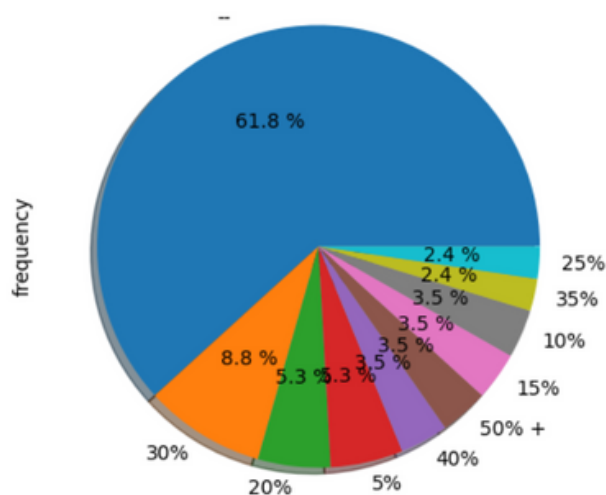
	frequency	percentage	cumulative_perc
66_porcentaje_services_income			
--	149	0.871345	0.871345
1-5%	13	0.076023	0.947368
6-10%	3	0.017544	0.964912



Podemos observar que el porcentaje de ingresos de servicios fue una frecuencia de 13 de 1-5% y una frecuencia de 3 de 6-10%.

• 38_AVERAGE_MARGIN_PROFITS

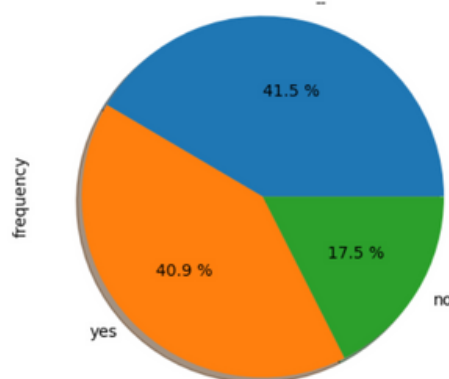
	frequency	percentage	cumulative_perc
38_average_margin_profits			
--	105	0.614035	0.614035
30%	15	0.087719	0.701754
20%	9	0.052632	0.754386
5%	9	0.052632	0.807018
40%	6	0.035088	0.842105
50% +	6	0.035088	0.877193
15%	6	0.035088	0.912281
10%	6	0.035088	0.947368
35%	4	0.023392	0.970760
25%	4	0.023392	0.994152



Ahora tenemos la variable categórica del promedio de margen de ganancia, el que tiene una mayor frecuencia (15) es el 30%, seguido por el 20% y 5% con una frecuencia de 9. El que tiene una menor frecuencia es el de 35% y 25% con una frecuencia de 4.

• 19_TAX_ID

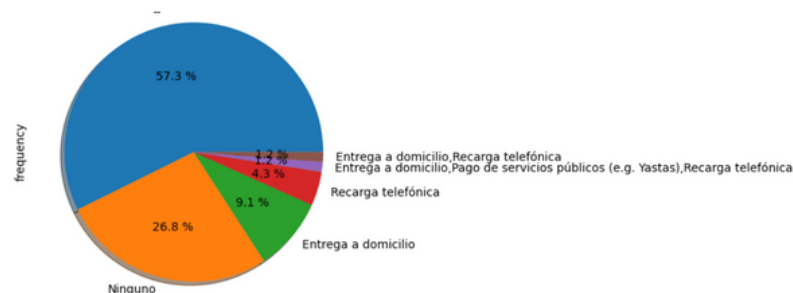
	frequency	percentage	cumulative_perc
19_tax_id			
--	71	0.415205	0.415205
yes	70	0.409357	0.824561
no	30	0.175439	1.000000



Ahora vemos el TAX Id, donde el 40.9% es una respuesta de sí y el 17.5% una respuesta de no. Es importante mencionar que la mayoría de los datos de esta variable son el string '--' al haber sido antes datos nulos.

• 60_SERVICES

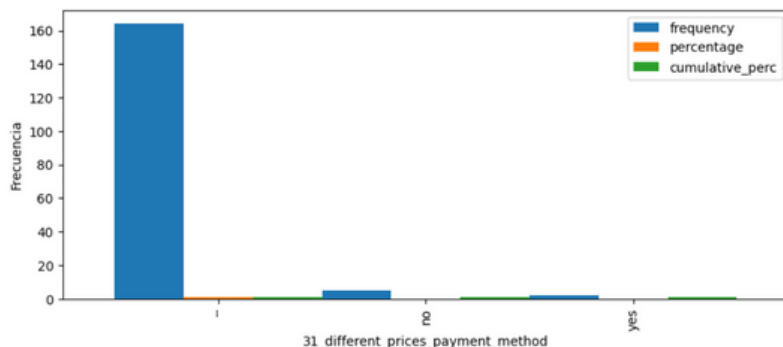
	frequency	percentage	cumulative_perc
60_services			
--	94	0.549708	0.549708
Ninguno	44	0.257310	0.807018
Entrega a domicilio	15	0.087719	0.894737
Recarga telefónica	7	0.040936	0.935673
Entrega a domicilio,Pago de servicios públicos (e.g. Yastas),Recarga telefónica	2	0.011696	0.947368
Entrega a domicilio,Recarga telefónica	2	0.011696	0.959064



Ahora vemos los servicios, donde el 26.8% no utilizan ninguno, seguido de la entrega a domicilio y recarga telefónica. El menos utilizado es la combinación de la entrega a domicilio con la recarga telefónica.

• 31_DIFFERENT_PRICES_PAYMENT_METHOD

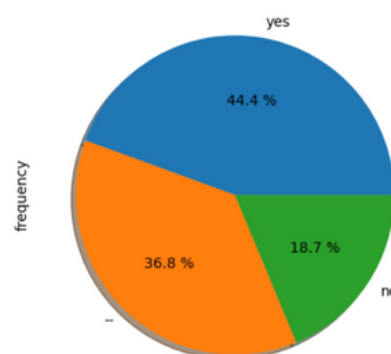
	frequency	percentage	cumulative_perc
31_different_prices_payment_method			
--	164	0.959064	0.959064
no	5	0.029240	0.988304
yes	2	0.011696	1.000000



Después se tiene el de diferentes precios en métodos de pago. Donde la respuesta de no tiene una frecuencia de 5 y la respuesta de sí una frecuencia de 2.

• 186_INTERNET_CONNECTION

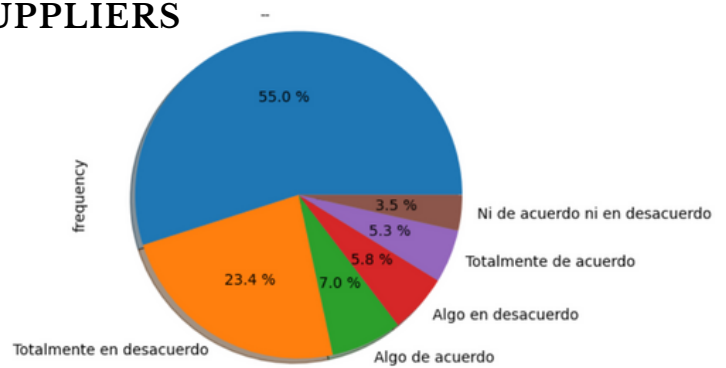
	frequency	percentage	cumulative_perc
186_internet_connection			
yes	76	0.444444	0.444444
--	63	0.368421	0.812865
no	32	0.187135	1.000000



En cuanto a la conexión a internet la mayoría, el 44.4% dicen que sí mientras que el 18.7% dicen que no.

• 45_WILLINGNESS_REORDER_SUPPLIERS

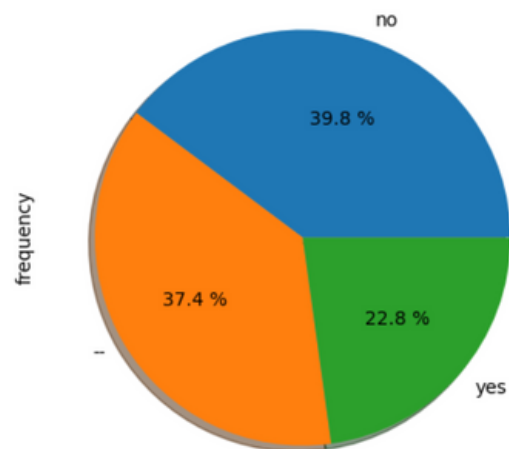
	frequency	percentage	cumulative_perc
45_willingness_reorder_suppliers			
--	94	0.549708	0.549708
Totalmente en desacuerdo	40	0.233918	0.783626
Algo de acuerdo	12	0.070175	0.853801
Algo en desacuerdo	10	0.058480	0.912281
Totalmente de acuerdo	9	0.052632	0.964912
Ni de acuerdo ni en desacuerdo	6	0.035088	1.000000



En cuanto a la disposición de reordenar proveedores el 40 dicen que están totalmente en desacuerdo, representado el 23.4%. Seguido de algo de acuerdo con una frecuencia de 12 que representa el 7%. Por último el que tiene una menor frecuencia es el ni de acuerdo ni en desacuerdo que representa el 3.5%

• 310_BURNOUT

	frequency	percentage	cumulative_perc
310_burnout			
no	68	0.397661	0.397661
--	64	0.374269	0.771930
yes	39	0.228070	1.000000



En cuanto al burnout/agotamiento la mayoría mencionan que no, representando el 39.8%. Mientras que el 22.8% representa que si, siendo una frecuencia de 39.