

REPORTE DE HALLAZGOS

ACTIVIDAD 4. Extracción de características

María Fernanda Martínez Ríos
A01067198

La base de datos que analizamos se trata de datos de tiendas de distintos tipos que recaudaron ciertos alumnos anteriormente, es importante mencionar que la base de datos cuenta con muchos datos inexistentes o nulos. Además, tiene 106 columnas y 107 filas, siendo 81 de ellas de tipo objeto, 23 de tipo ‘float’ y 2 de entero.

Para este análisis y práctica de la extracción de características, elegí 20 columnas de las 81 que corresponden a variables de tipo cualitativa o categórico. Siendo los siguientes los nombres de dichas columnas:

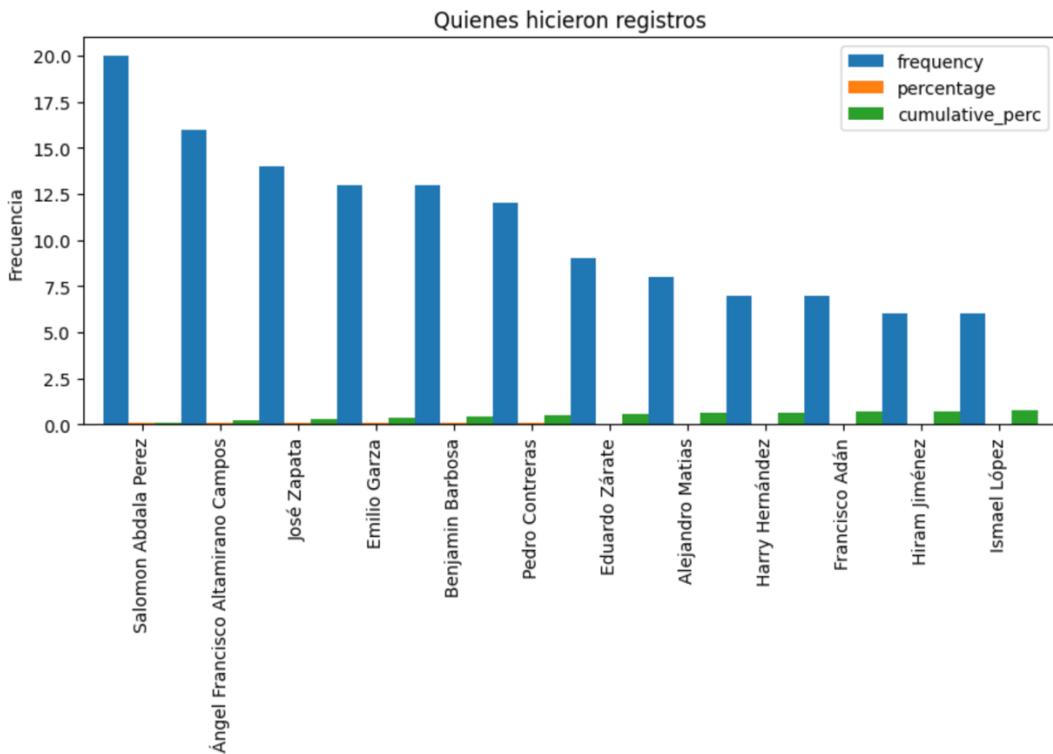
- _updated_by
- 232_type_of_store
- 184_store_devices
- 6_change_employees_average_salary_last_year
- 49_inventory_records
- 18_sales_records
- 186_internet_connection
- 210_sales_channels
- 189_payment_methods
- 214_customer_relationship_tools
- 33_credit_to_customers
- 35_interest_rate
- 312_payment_of_utilities
- 69_why_not_services
- 313_home_deliveries
- 44_willingness_inventory_management_suppliers
- 46_willingness_reorder_suppliers_variety
- 157_frequency.organize_shelves
- 250_technology_investment_operations
- 261_technology_organization

Aun y con la falta de información, realice el análisis de cada una de las columnas anteriores, llenando los valores nulos con el dato ‘Sin registro’.

La primer columna, ‘_updated_by’ nos da información sobre la persona que hizo el registro de la información, en total solo 31 personas distintas agregaron información, sin embargo de estas 31, hubo quienes hicieron mas de 9 registros y otras que únicamente uno como podemos observar en la siguiente tabla:

<u>_updated_by</u>	frequency	percentage	cumulative_perc
Salomon Abdala Perez	20	0.116959	0.116959
Ángel Francisco Altamirano Campos	16	0.093567	0.210526
José Zapata	14	0.081871	0.292398
Emilio Garza	13	0.076023	0.368421
Benjamín Barbosa	13	0.076023	0.444444
Pedro Contreras	12	0.070175	0.514620
Eduardo Zárate	9	0.052632	0.567251
Alejandro Matías	8	0.046784	0.614035
Harry Hernández	7	0.040936	0.654971
Francisco Adán	7	0.040936	0.695906
Hiram Jiménez	6	0.035088	0.730994
Ismael López	6	0.035088	0.766082

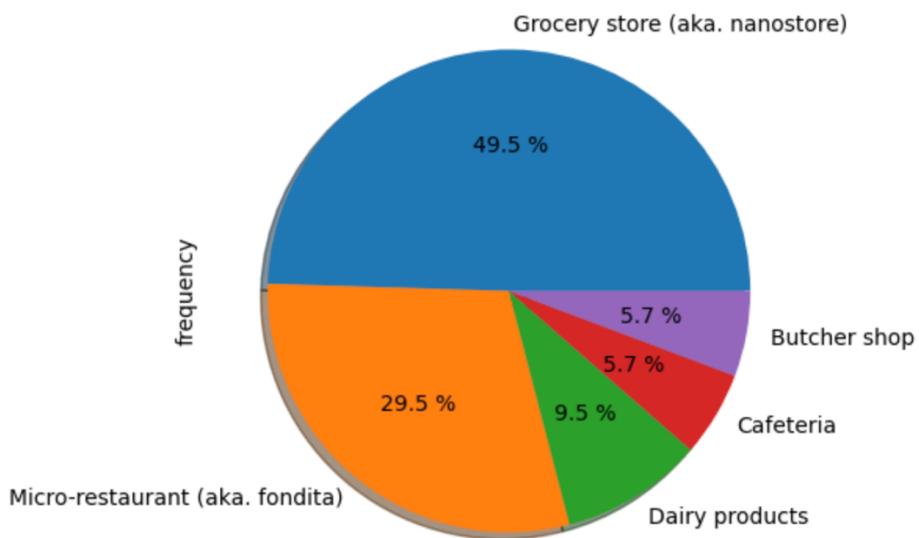
Ahora con la gráfica podemos ver mas visualmente la participacion de las personas que tenemos en la tabla, los registros de estas personas representan el 76% del total de la columna y las de Salomon Abdala el 11% del total.



La segunda columna ‘232_type_of_store’ trata acerca de el tipo de tienda que eran las tiendas registradas a partir del producto o servicio que ofrecen. Como mencione, hay varios tipos, la frecuencia de algunas categorías podría ser mayor, pero al registrarla pusieron nombres que no coinciden pero se trata de la misma categoría.

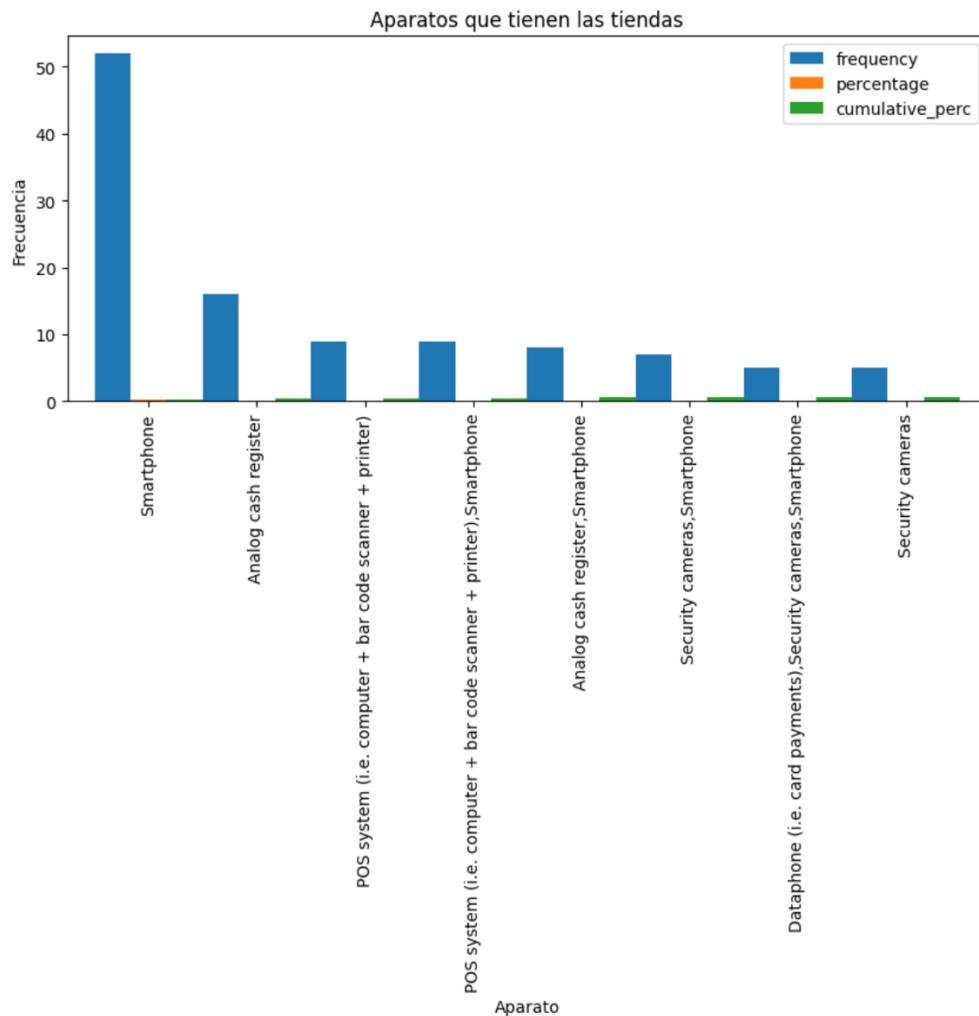
	232_type_of_store	frequency	percentage	cumulative_perc
0	Grocery store (aka. nanostore)	52	0.304094	0.304094
1	Micro-restaurant (aka. fondita)	31	0.181287	0.485380
2	Dairy products	10	0.058480	0.543860
3	Cafeteria	6	0.035088	0.578947
4	Butcher shop	6	0.035088	0.614035

Las tiendas de alimentos son las más comunes, seguido de las de pequeños restaurantes. La primer categoría abarca el 30% del total de los datos y la segunda, el 18%. Pero si consideramos que estas fueron el 100% de los datos, su porcentaje de participación sería el siguiente:



La tercer columna que escogí y analicé fue la de 184_store_devices que nos da información sobre la tecnología o aparatos que tienen estas tiendas para realizar sus operaciones. Los de la tabla fueron los más comunes, pero el teléfono inteligente fue el de mayor frecuencia, 30% tienen ‘smartphone’. De igual forma, otros tienen teléfono mas otros aparatos como cámaras de seguridad o cajas registradoras.

	frequency	percentage	cumulative_perc
184_store_devices			
Smartphone	52	0.304094	0.304094
Analog cash register	16	0.093567	0.397661
POS system (i.e. computer + bar code scanner + printer)	9	0.052632	0.450292
POS system (i.e. computer + bar code scanner + printer),Smartphone	9	0.052632	0.502924
Analog cash register,Smartphone	8	0.046784	0.549708
Security cameras,Smartphone	7	0.040936	0.590643
Dataphone (i.e. card payments),Security cameras,Smartphone	5	0.029240	0.619883
Security cameras	5	0.029240	0.649123



La cuarta columna nos habla de ‘6_change_employees_average_salary_last_year’, es decir, sobre el cambio de salario de los empleados de estas tiendas. Como punto negativo para nusro

análisis, el 50% de los datos fueron nulos (sin registro), así que la información puede no ser tan representativa.

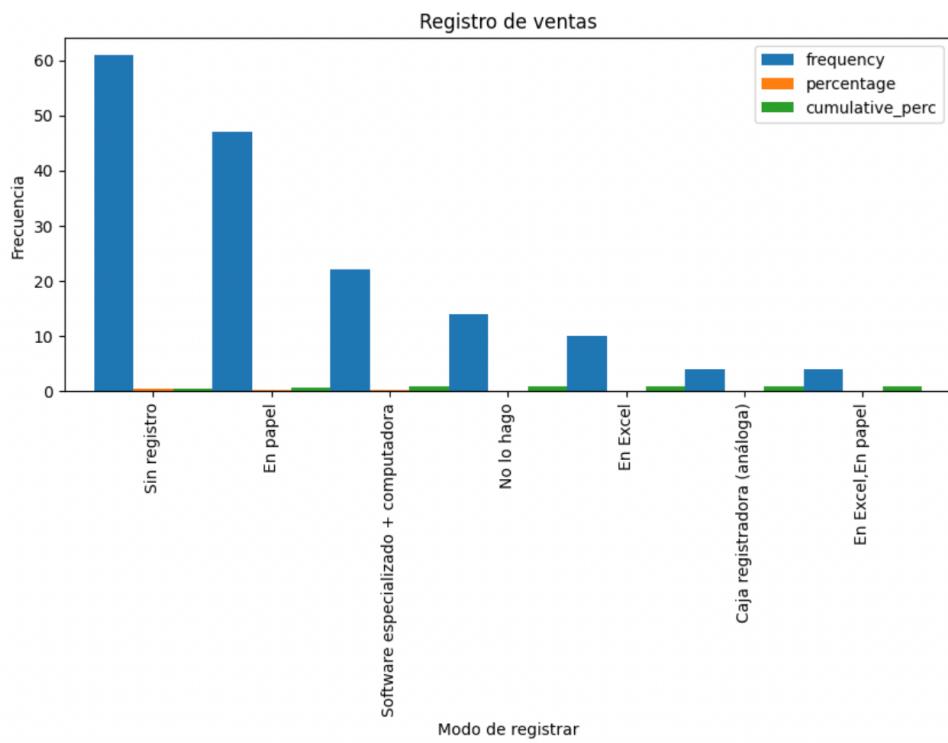
		frequency	percentage	cumulative_perc
6_change_employees_average_salary_last_year				
Sin registro	86	0.502924	0.502924	
No cambió	50	0.292398	0.795322	
Aumentó 15%	14	0.081871	0.877193	
Disminuyó 15%	5	0.029240	0.906433	
Aumentó 25%	4	0.023392	0.929825	
Aumentó 50% o más	4	0.023392	0.953216	

Mientras otro 29% no tuvo ningun cambio en su salario, 8% de los datos registraron que tuvo un aumento del 15% y menor porcentaje, alredor del 2% aumentaron el salario a; hasta 50% o más.

La siguiente columna, nos habla de la manera en que las tiendas tienen sus registros, donde los anotan o guardan. Esta columna de igual forma resulto tener muchos datos sin registro, pero se sabe que el otro 23% lo hace en papel, 12% a traves de una computadora y 11% en excel.

		frequency	percentage	cumulative_perc
49_inventory_records				
Sin registro	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	

La sexta columna es la de registros, pero esta vez de los registros que se hacen sobre las ventas de las tiendas. Sigue siendo más alta la barra de los datos sin registro y como segunda más frecuente los registros en papel.



La columna 7 es una columna categórica con información de si y no de acuerdo a las tiendas que tienen internet y las que no tienen.



La siguiente columna, ‘210_sales_channels’, los canales de venta de las tiendas incluyen redes sociales y páginas web. También predominaron los datos sin registro (52%), pero las redes sociales fueron el segundo con mayor frecuencia (6.4%), seguido de mensajes instantáneos de WhatsApp (5.8%).

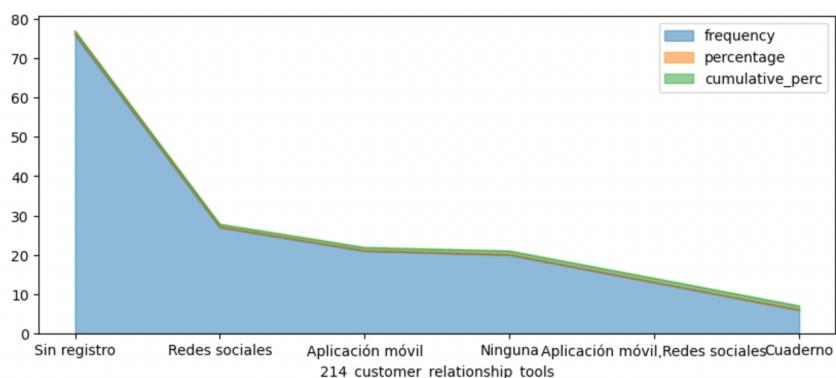
		210_sales_channels	frequency	percentage	cumulative_perc
0		Sin registro	89	0.520468	0.520468
1	Redes sociales (e.g. Facebook, Instagram)		11	0.064327	0.584795
2		WhatsApp	10	0.058480	0.643275
3	Redes sociales (e.g. Facebook, Instagram), WhatsApp		9	0.052632	0.695906
4		Llamada a Teléfono fijo	6	0.035088	0.730994
5		Llamada a Teléfono celular, WhatsApp	6	0.035088	0.766082
6		Llamada a Teléfono celular	5	0.029240	0.795322
7	Llamada a Teléfono celular, Redes sociales (e.g. Facebook, Instagram)		5	0.029240	0.824561

La columna de métodos de pago nos da la siguiente información:

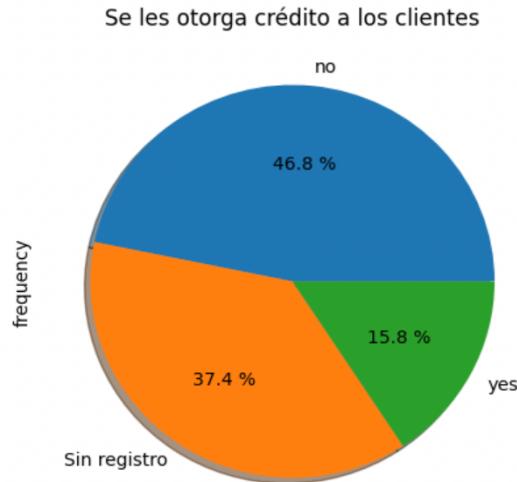
		189_payment_methods	frequency	percentage	cumulative_perc
0		Sin registro	64	0.374269	0.374269
1		Efectivo	40	0.233918	0.608187
2	Efectivo, Tarjeta de débito y/o crédito, Transferencia interbancaria		30	0.175439	0.783626
3		Efectivo, Tarjeta de débito y/o crédito	19	0.111111	0.894737
4		Efectivo, Transferencia interbancaria	15	0.087719	0.982456
5	Aplicación móvil (e.g. CODI), Efectivo, Tarjeta de débito y/o crédito		1	0.005848	0.988304
6	Aplicación móvil (e.g. CODI), Efectivo, Tarjeta de débito y/o crédito		1	0.005848	0.994152
7		Tarjeta de débito y/o crédito	1	0.005848	1.000000

El 23% paga en efectivo y 39% en otros medios electrónicos o tarjetas. Pero si no consideramos los 64 sin registro que hay, se puede decir que queda la participación quedando como el 37% de los pagos son en efectivo y el 62.61% restante con otros medios.

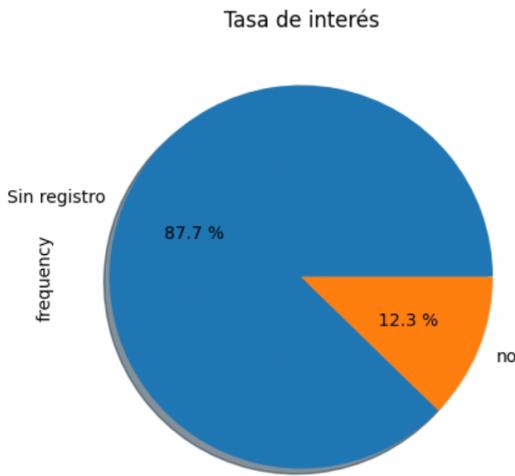
La columna 10 es acerca de las herramientas que se utilizan para hacer ‘engagement’ con los clientes y las más populares fueron las redes sociales y aplicaciones móviles.



Sabemos gracias a la columna de ‘33_credit_to_customers’ el porcentaje de tiendas que otorgan créditos a sus clientes y las que no:



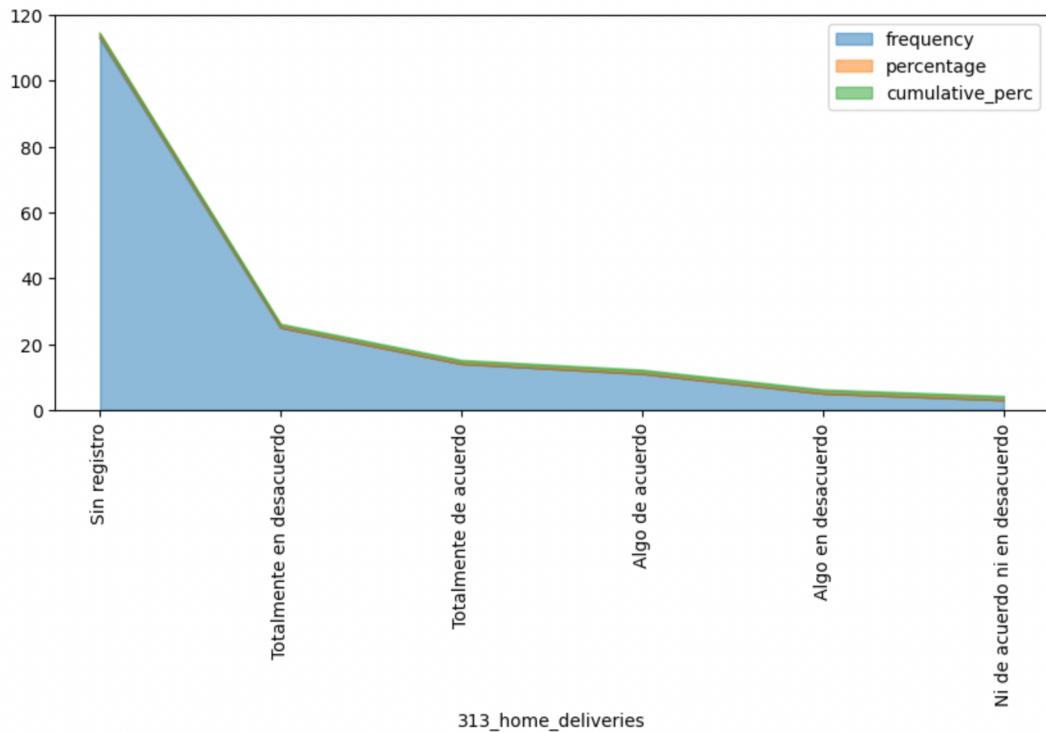
La siguiente gráfica nos muestra las tiendas que cobran intereses y las que no de las que si otorgan créditos a sus clientes. De las respuestas que no fueron nulas, ninguna tienda menciono que si cobrará intereses.



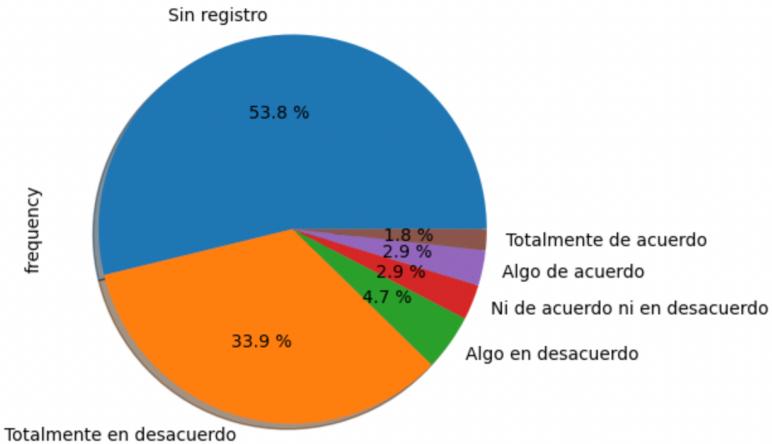
Otra columna analizado fue el nivel de satisfacción con el pago de utilidades, pero mas de la mitad de los datos se encontraron sin registro, sin embargo, de los que si hubo respuesta, sabemos que la mayoría esta totalmente en desacuerdo, esta categoría es la de mayor frecuencia con 46 respuestas así y las respuestas restantes también mostraron desacuerdo.

	312_payment_of_utilities	frequency	percentage	cumulative_perc
0	Sin registro	105	0.614035	0.614035
1	Totalmente en desacuerdo	46	0.269006	0.883041
2	Totalmente de acuerdo	9	0.052632	0.935673
3	Algo de acuerdo	5	0.029240	0.964912
4	Ni de acuerdo ni en desacuerdo	4	0.023392	0.988304
5	Algo en desacuerdo	2	0.011696	1.000000

La siguiente columna es igual acerca del nivel de satisfacción pero acerca de las entregas a casas, nuevamente había muchos datos nulos y así como en la columna anterior, la mayoría estuvo en desacuerdo. El 66% fueron sin registro y 14% totalmente en desacuerdo.



Después en la siguiente gráfica de barras igual podemos ver la la satisfacción en cuanto a la disposición a la gestión de inventario y proveedores. De las respuestas que hay en dicha columna, 33.9% están en total desacuerdo.

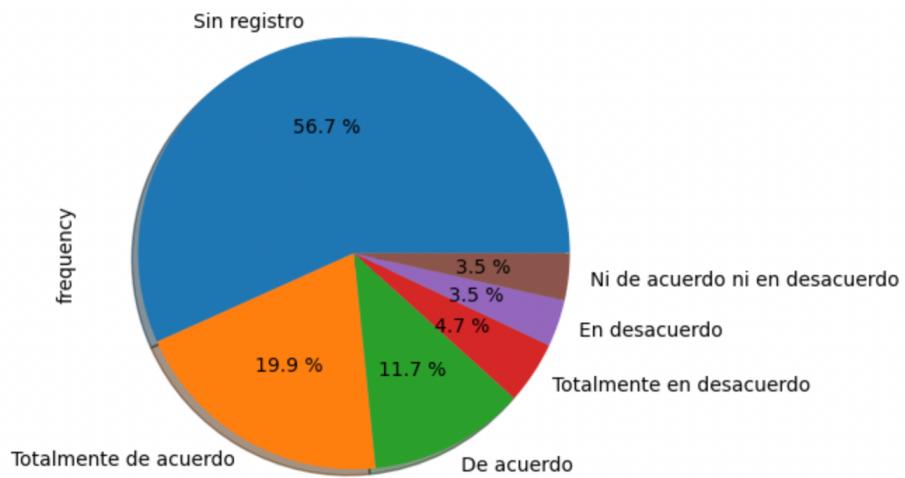


Esta tabla muestra la frecuencia con la que las tiendas organizan sus estantes y la mayoría lo hace cada semana, el 32% todos los días, el 4% nunca lo hace, 2.9% quincenalmente y los restantes lo hacen en períodos mensuales o mayores.

	frequency_organize_shelves	frequency	percentage	cumulative_perc
0	Sin registro	75	0.438596	0.438596
1	Semanalmente	40	0.233918	0.672515
2	Diariamente	32	0.187135	0.859649
3	Nunca	8	0.046784	0.906433
4	Mensualmente	7	0.040936	0.947368
5	Quincenalmente	5	0.029240	0.976608
6	Anualmente	2	0.011696	0.988304
7	Semestralmente	1	0.005848	0.994152
8	No se tiene	1	0.005848	1.000000

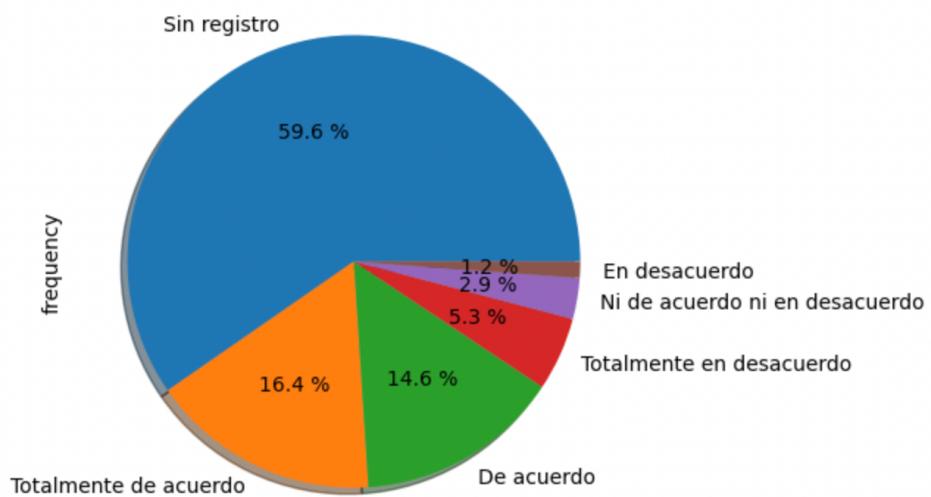
La penúltima columna es acerca del nivel de satisfacción con las operaciones de inversión en tecnología y los resultados se distribuyeron de la siguiente forma:

Operaciones de inversión en tecnología



Por ultimo, esta variable es sobre la satisfacción u opinión acerca de la organización de la tecnología de las tiendas, mas de la mitad de las respuestas son datos sin registro, pero si predominó una satisfacción positiva a comparación de otras columnas.

Organización de la tecnología



Para finalizar, puedo decir que fue un análisis sencillo, mas no obstante algo ambiguo, pues en todas las columnas faltó información por registrar y por lo tanto, se analizo una muestra mucho más pequeña de tiendas, lo que podría indicar que la información resultante no fuera representativa, sin embargo, el análisis de columnas categóricas también representa una parte fundamental en el entendimiento de una base de datos, aun y cuando en la mayoría de los casos no se puedan obtener medidas de tendencia central, de dispersión o de posición, nos proporcionan información relevante acerca de opciones y frecuencias.