

f



Tecnológico de Monterrey

Instituto tecnológico de estudios superiores de Monterrey

Actividad 4 Extracción de Características

Unidad de formación

Desarrollo de Proyectos de Análisis de Datos (Gpo 302)

Docentes

Gabriela Guadalupe Reyes Zarate
Alfredo García Suarez

Estudiante

Santiago Olivera Vaquero A01736797

Fecha

19 nov 2023

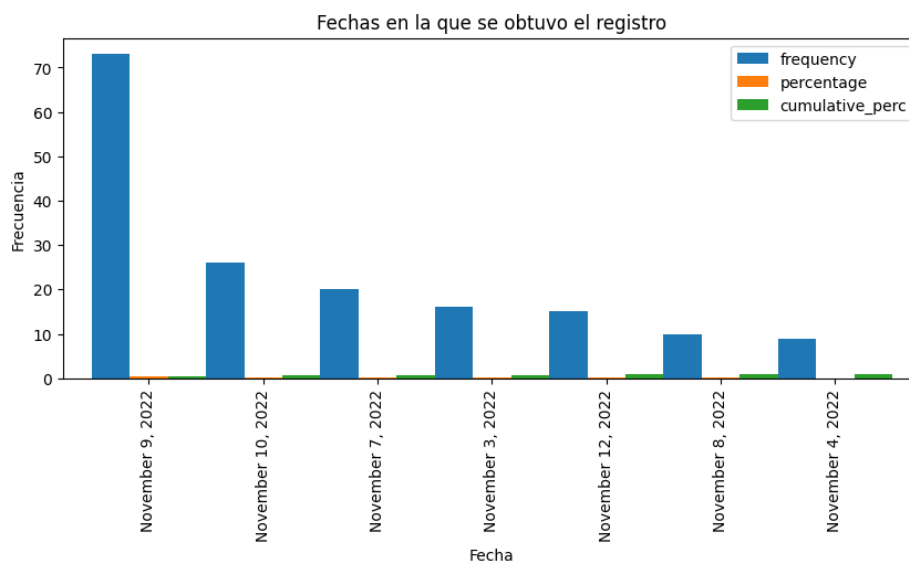
Resumen

Este trabajo consiste de un ejercicio con una base de datos de micro negocios retail con 171 entradas armada por estudiantes del Tecnológico de Monterrey Campus Puebla de semestres pasados. El ejercicio consistió en preparar los datos y hacer un análisis de frecuencias univariado con 20 variables categóricas a través de histogramas, diagramas de pastel y tablas de frecuencia. Las variables categóricas son aquellas que se pueden tener datos repetidos en su columna.

Análisis

Fecha en la que se obtuvo el registro:

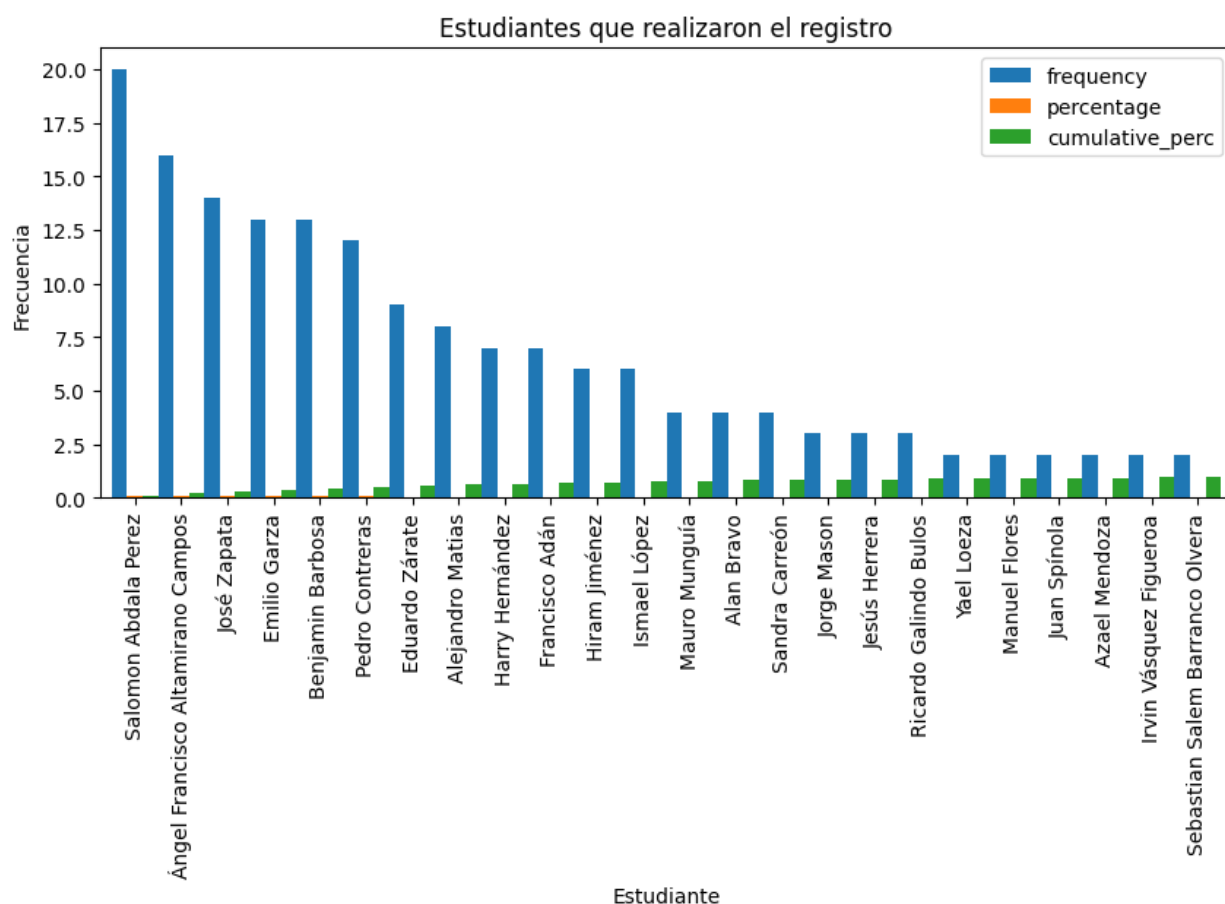
	frequency	percentage	cumulative_perc
_title			
November 9, 2022	73	0.426901	0.426901
November 10, 2022	26	0.152047	0.578947
November 7, 2022	20	0.116959	0.695906
November 3, 2022	16	0.093567	0.789474
November 12, 2022	15	0.087719	0.877193
November 8, 2022	10	0.058480	0.935673
November 4, 2022	9	0.052632	0.988304



De las 171 entrevistas totales que se realizaron, alrededor del 99% se realizaron en las fechas de noviembre que se muestran en la tabla y la gráfica. De estas la mayoría se realizaron el 9 de noviembre con un 43% de las entrevistas seguido por un 15% el 10 de noviembre. Es decir, en el día en que se realizaron la mayor cantidad de entrevistas se realizaron casi el triple de entrevistas que en el día en segundo lugar. El 9 de noviembre cae en miércoles, lo cual puede explicar porque una gran parte de las entrevistas se realizaron ese día. Esto es debido a que los miércoles son los días en los que los estudiantes de Innovación y Transformación tienen más tiempo libre dado que la única clase en ese día es un bloque.

Estudiante que realizó el registro:

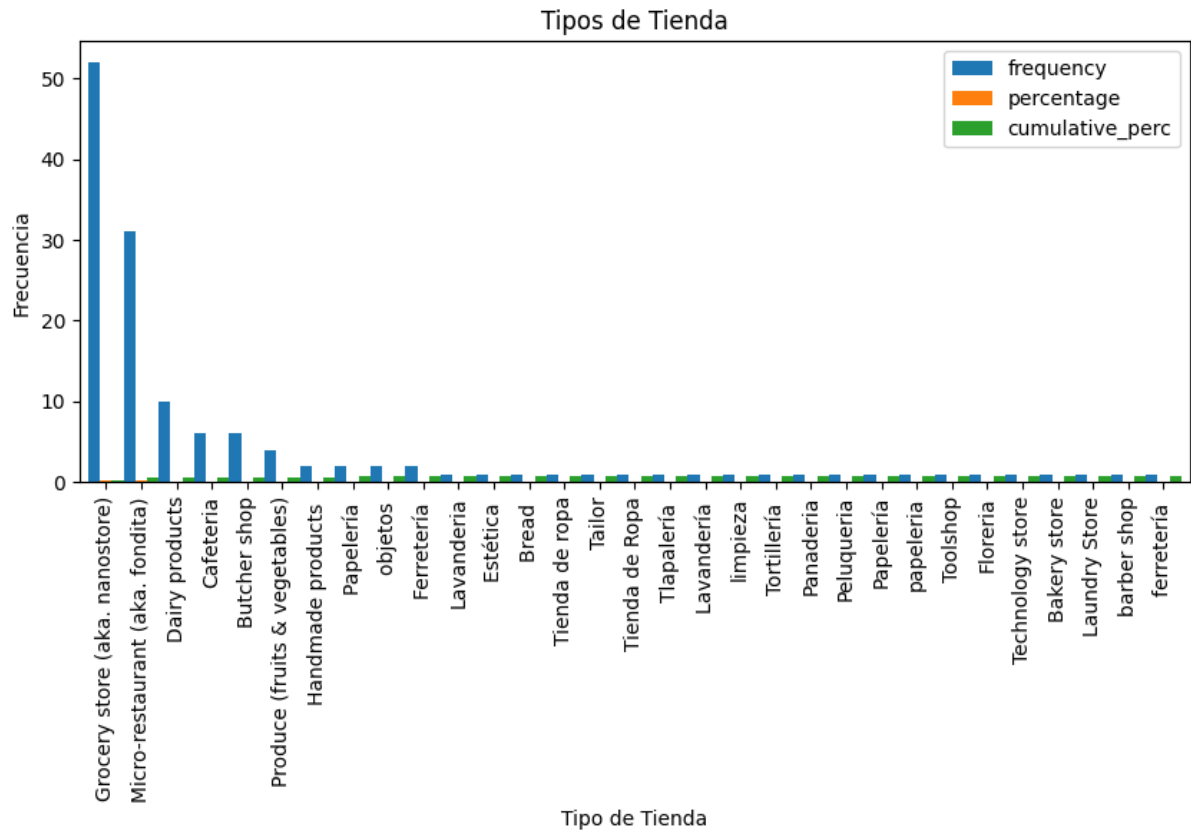
	frequency	percentage	cumulative_perc
_updated_by			
Salomon Abdala Perez	20	0.116959	0.116959
Ángel Francisco Altamirano Campos	16	0.093567	0.210526
José Zapata	14	0.081871	0.292398
Emilio Garza	13	0.076023	0.368421
Benjamin Barbosa	13	0.076023	0.444444
Pedro Contreras	12	0.070175	0.514620
Eduardo Zárate	9	0.052632	0.567251
Alejandro Matias	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
Mauro Munguía	4	0.023392	0.789474
Alan Bravo	4	0.023392	0.812865
Sandra Carreón	4	0.023392	0.836257
Jorge Mason	3	0.017544	0.853801
Jesús Herrera	3	0.017544	0.871345
Ricardo Galindo Bulos	3	0.017544	0.888889
Yael Loeza	2	0.011696	0.900585
Manuel Flores	2	0.011696	0.912281
Juan Spinola	2	0.011696	0.923977
Azael Mendoza	2	0.011696	0.935673
Irvin Vásquez Figueroa	2	0.011696	0.947368
Sebastian Salem Barranco Olvera	2	0.011696	0.959064



Alrededor del 96% de las entrevistas fueron realizadas por los 24 estudiantes que se muestran en la tabla y en la gráfica. Al observar la gráfica se observa una relación escalonada descendente desde los estudiantes que realizaron más entrevistas a los que realizaron menos. A partir del estudiante en la novena posición en número de entrevistas realizadas, es decir “Harry Hernández”, empiezan a haber pares y después tercias de estudiantes que realizaron el mismo número de entrevistas. Esto puede ser indicativo de una forma de organización en los equipos pero se debería investigar más a detalle para concluir al respecto.

Tipos de Tienda:

	frequency	percentage	cumulative_perc
232_type_of_store			
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
Papelería	2	0.011696	0.660819
objetos	2	0.011696	0.672515
Ferretería	2	0.011696	0.684211
Lavandería	1	0.005848	0.690058
Estética	1	0.005848	0.695906
Bread	1	0.005848	0.701754
Tienda de ropa	1	0.005848	0.707602
Tailor	1	0.005848	0.713450
Tienda de Ropa	1	0.005848	0.719298
Tlapalería	1	0.005848	0.725146
Lavandería	1	0.005848	0.730994
limpieza	1	0.005848	0.736842
Tortillería	1	0.005848	0.742690
Panadería	1	0.005848	0.748538
Peluquería	1	0.005848	0.754386
Papelería	1	0.005848	0.760234
papelería	1	0.005848	0.766082
Toolshop	1	0.005848	0.771930
Florería	1	0.005848	0.777778
Technology store	1	0.005848	0.783626
Bakery store	1	0.005848	0.789474
Laundry Store	1	0.005848	0.795322
barber shop	1	0.005848	0.801170

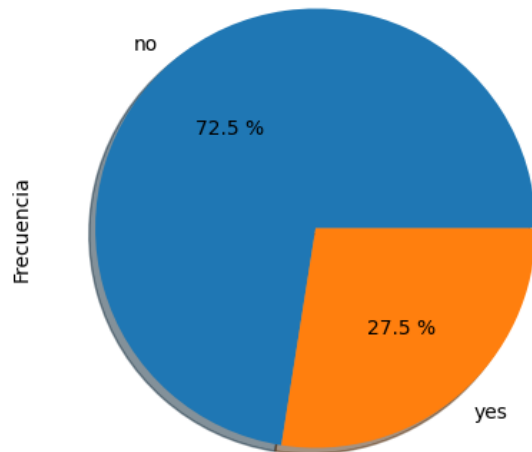


El 68% de las tiendas entrevistadas se encuentran dentro de 10 tipos, con el resto teniendo una frecuencia idéntica del 0.58%. Esto habla de como varias tiendas se concentran en el mismo tipo con el tipo de tienda más frecuente representando el 30% de la muestra mientras que con las tiendas menos frecuentes habla de una distribución equitativa en la que si se combinan serían igual de frecuentes que la tienda más frecuente.

Tiendas con ventanas abarrotadas:

	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

Frecuencia de Tiendas con ventanas con barrotes

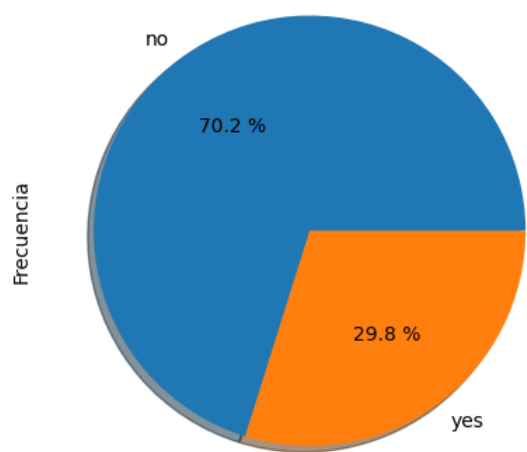


De la muestra de tiendas entrevistadas el 72.5% no tienen ventana abarrotadas mientras que el resto si. Esto puede deberse a varias causas como la falta de necesidad desde la perspectiva de los dueños o la falta de capital suficiente para invertir en su instalación. Habría que analizar más al respecto para concluir acerca de las causas de esta distribución.

Tiendas que exhiben sus productos afuera

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

Frecuencia de Tiendas que Exhiben Productos Afuera

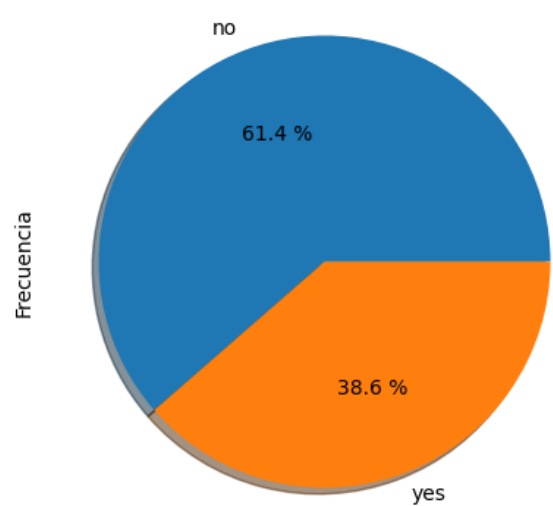


El 70.2% de las tiendas no exhiben productos afuera de sus casas. Esto representa una mayoría relativamente significativa y podría deberse a una causa de falta de confianza en la seguridad de sus productos pero habría que investigar más para concluir al respecto.

Tiendas que venden productos frescos:

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

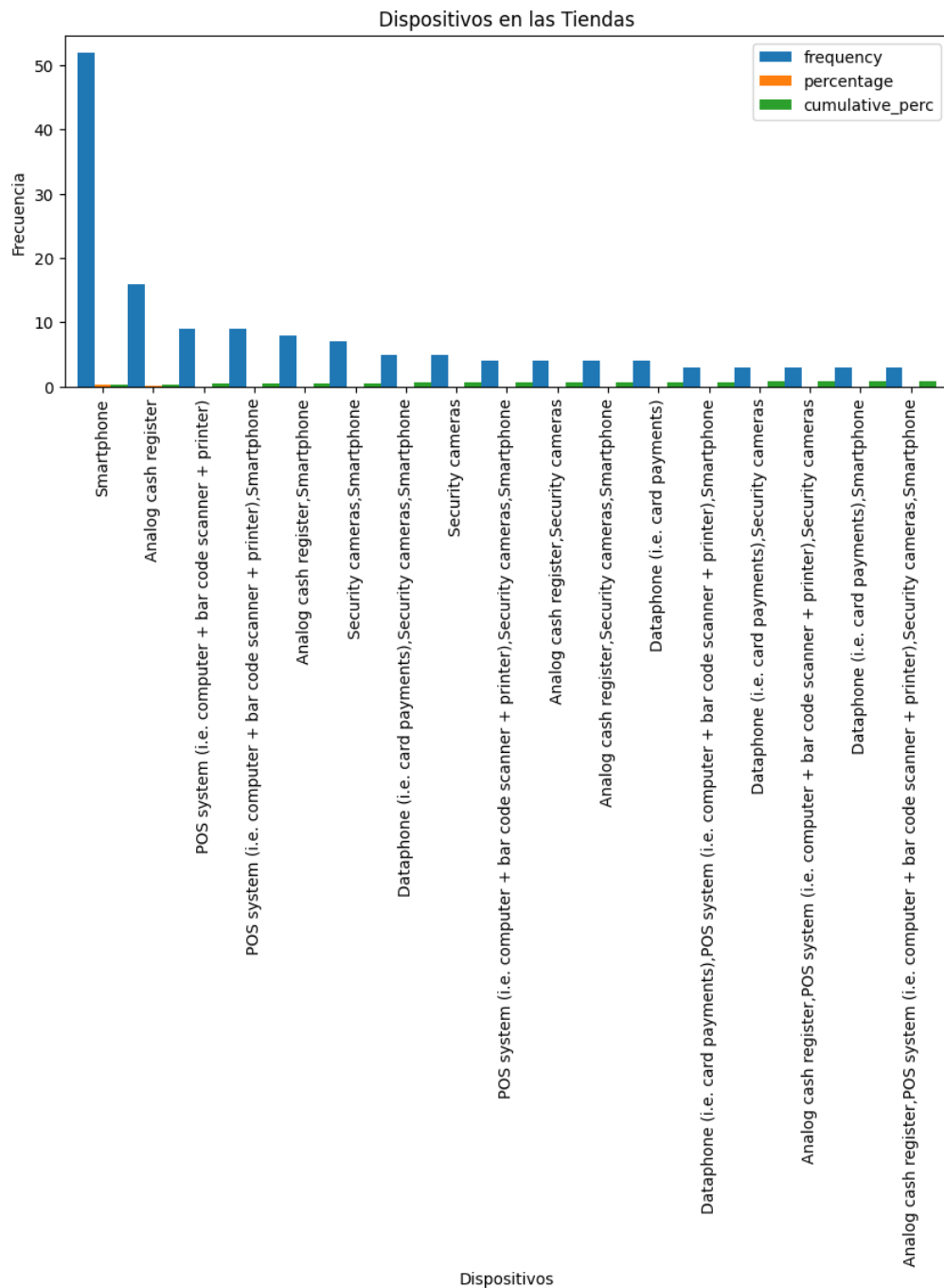
Frecuencia de Tiendas que Venden Productos Frescos



La mayoría de las tiendas no venden productos frescos con una presencia en la muestra del 61.4%.

Dispositivos electrónicos en las tiendas:

	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
POS system (i.e. computer + bar code scanner + printer),Security cameras,Smartphone	4	0.023392	0.672515
Analog cash register,Security cameras	4	0.023392	0.695906
Analog cash register,Security cameras,Smartphone	4	0.023392	0.719298
Dataphone (i.e. card payments)	4	0.023392	0.742690
Dataphone (i.e. card payments),POS system (i.e. computer + bar code scanner + printer),Smartphone	3	0.017544	0.760234
Dataphone (i.e. card payments),Security cameras	3	0.017544	0.777778
Analog cash register,POS system (i.e. computer + bar code scanner + printer),Security cameras	3	0.017544	0.795322
Dataphone (i.e. card payments),Smartphone	3	0.017544	0.812865
Analog cash register,POS system (i.e. computer + bar code scanner + printer),Security cameras,Smartphone	3	0.017544	0.830409

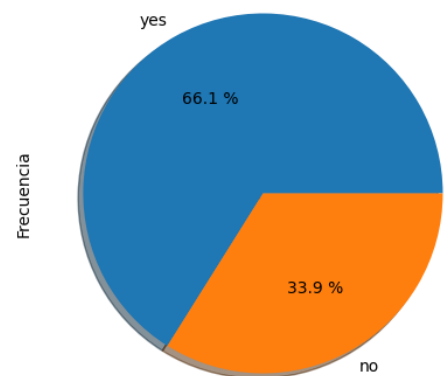


De todas las tiendas entrevistadas hay una alta concentración que usan Smartphone con un 30% de la muestra seguida por una caja registradora con un 9%. A partir de allí la distribución empieza a igualarse cada vez más lo cual se puede observar con él como las barras empiezan a tener tamaños más similares mientras más pequeñas son.

Estudiantes que dijeron “Hola, somos estudiantes del Tec de Monterrey estamos realizando una encuesta:

	frequency	percentage	cumulative_perc
hola_somos_estudiantes_del_tec_de_monterrey_estamos_realiz			
yes	113	0.660819	0.660819
no	58	0.339181	1.000000

Frecuencia de Estudiantes que Dijeron "Hola, somos Estudiantes del Tec de Monterrey estamos Realizando una Encuesta"

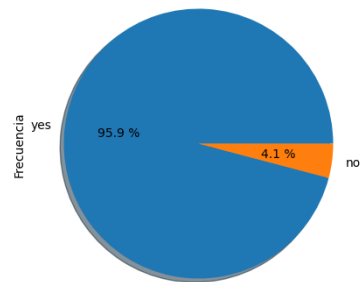


En la mayoría de las entrevistas realizadas, los estudiantes si mencionaron que eran estudiantes del Tecnológico de Monterrey con un 66.1% de la muestra.

Negocios que comentaron que la información es confidencial:

	frequency	percentage	cumulative_perc
le_comento_que_toda_esta_informacin_es_confidencial			
yes	164	0.959064	0.959064
no	7	0.040936	1.000000

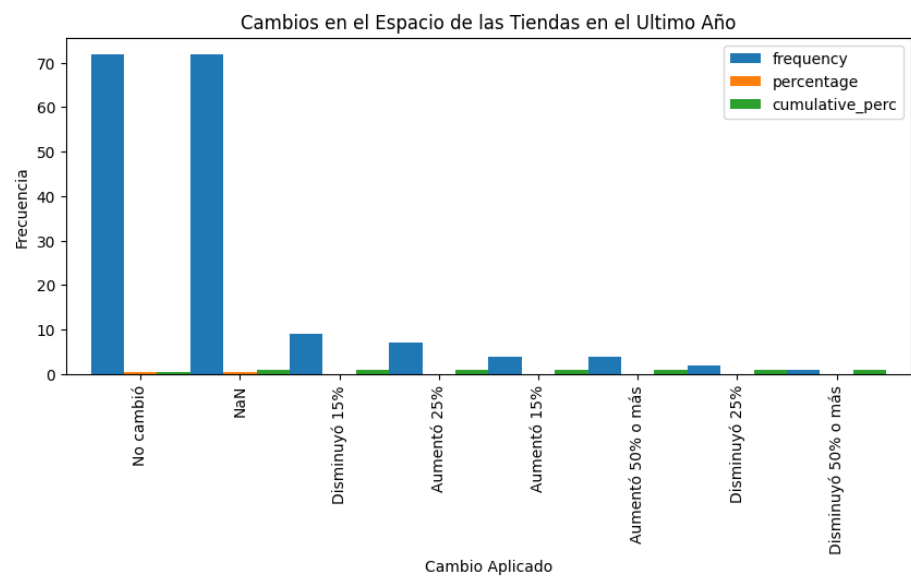
Frecuencia de Negocios que Comentaron que la Información es Confidencial



De los negocios entrevistados la gran mayoría dijeron que la información es confidencial con un 95.9%. Esto pudo haber representado una dificultad para obtener información precisa acerca de los negocios.

Cambios en el espacio en el último año:

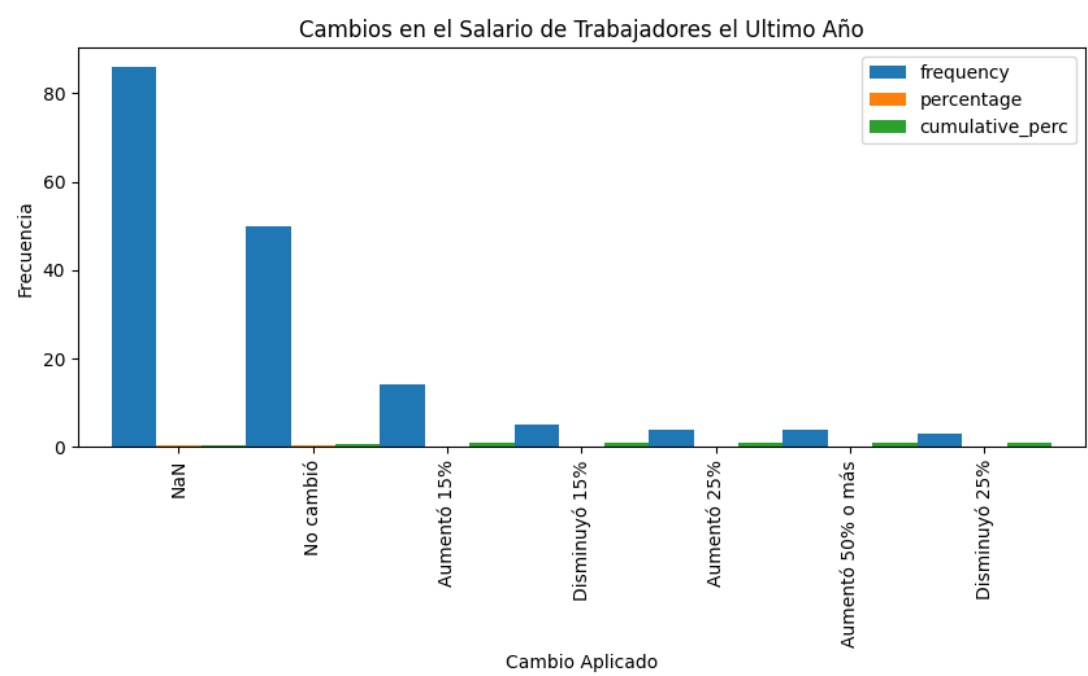
	frequency	percentage	cumulative_perc
5_change_store_space_last_year			
No cambió	72	0.421053	0.421053
NaN	72	0.421053	0.842105
Disminuyó 15%	9	0.052632	0.894737
Aumentó 25%	7	0.040936	0.935673
Aumentó 15%	4	0.023392	0.959064
Aumentó 50% o más	4	0.023392	0.982456
Disminuyó 25%	2	0.011696	0.994152
Disminuyó 50% o más	1	0.005848	1.000000



De los negocios entrevistados sólo alrededor del 16% aplicaron un cambio en el espacio de su tienda durante el último año. El 42% no ha aplicado ningún cambio y del del otro 42% no se tiene información al respecto. Esto puede significar que no hay mucho interés en cambiar el espacio de las tiendas por parte de los negocios pero dado que gran parte de la información se desconoce no se puede concluir al respecto. Si se desprecian los valores nulos, las tiendas que han aplicado algún cambio en el espacio de sus tiendas en el último año representarían poco menos de un tercio de un tercio de la nueva muestra.

Cambios en el salario de empleados el último año:

	frequency	percentage	cumulative_perc
6_change_employees_average_salary_last_year			
NaN	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

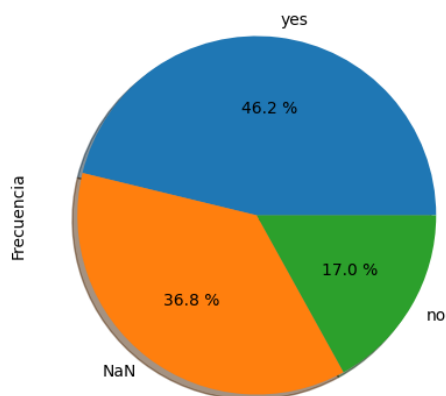


De todos los negocios entrevistados solo se pudo obtener información de la mitad con relación a cambios que se han aplicado en los ingresos de los empleados. De esa mitad poco más de la mitad no aplicaron un cambio representando un 29% de la muestra total y de los que sí aplicaron un cambio poco menos de la mitad aumentaron los ingresos, siendo estos un 15% de la muestra total mientras que el resto de los datos se encuentran dentro del 13% de la muestra total.

Negocios que revisan sus finanzas cada mes:

	frequency	percentage	cumulative_perc
20_reviews_finances_monthly			
yes	79	0.461988	0.461988
NaN	63	0.368421	0.830409
no	29	0.169591	1.000000

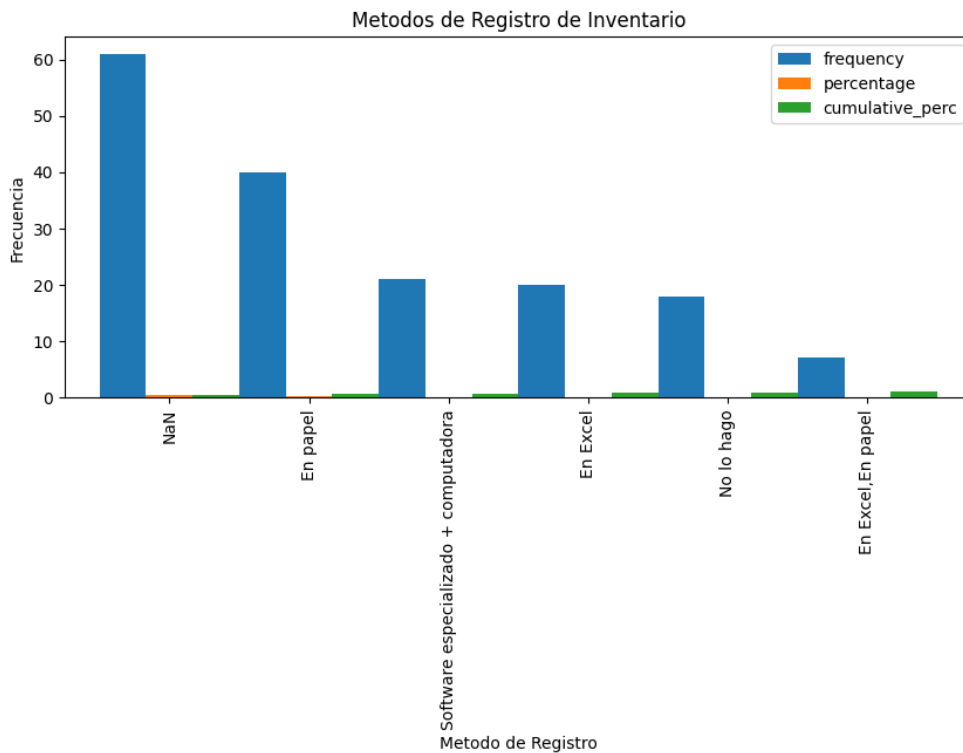
Frecuencia de Tiendas que Revisan sus Finanzas al Mes



De todos los negocios entrevistados, el 46.2% si revisan sus finanzas al mes, el 17% no lo hace y del 36.8% restante no se tiene información. Esto puede significar que la mayoría de los negocios si revisan sus finanzas al mes, siendo estos más del doble de los que se obtuvo información al respecto, pero dado la alta presencia de valores nulos del 36.8% estos datos no son suficientes para hacer conclusiones al respecto sin averiguar porque esos valores son nulos.

Métodos en los que los negocios registran su inventario:

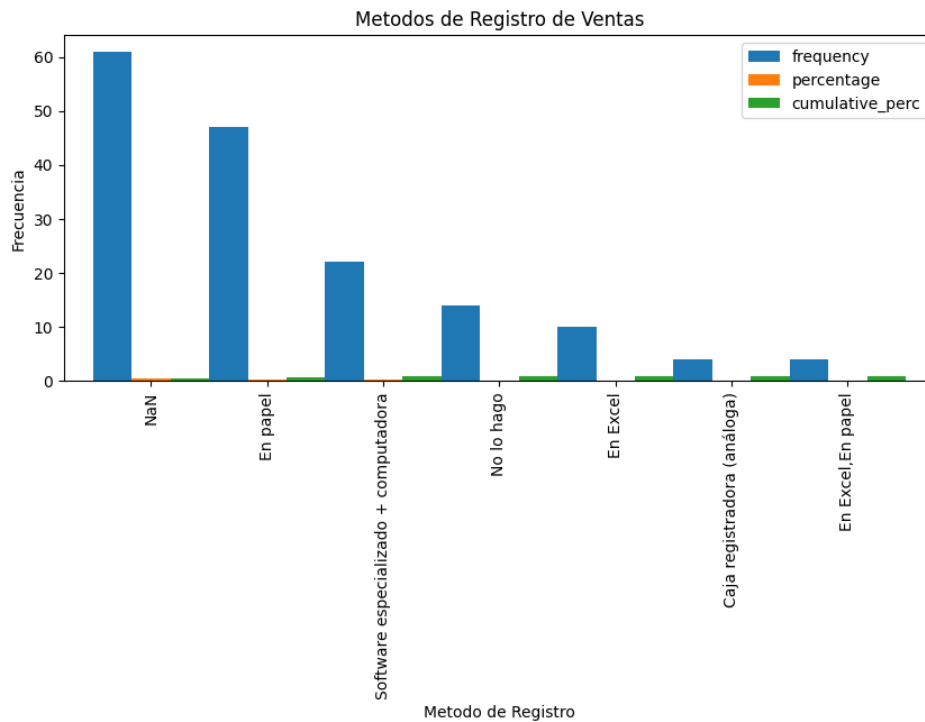
	frequency	percentage	cumulative_perc
49_inventory_records			
NaN	61	0.356725	0.356725
En papel	40	0.233918	0.590643
Software especializado + computadora	21	0.122807	0.713450
En Excel	20	0.116959	0.830409
No lo hago	18	0.105263	0.935673
En Excel,En papel	7	0.040936	0.976608



De todos los negocios entrevistados solo se pudo obtener información del 65%. con relación a los métodos de registro de inventario. De los que sí se obtuvo información, la mayoría lo hacen en papel siendo estos un 23% de la muestra. A partir de allí los otros métodos se encuentran distribuidos de forma similar, incluyendo a los que no lo hacen, con la frecuencia de cada uno siendo de alrededor de la mitad del método más aplicado siendo esta de alrededor del 11%. Esto exceptúa al método menos frecuente que se observa en la muestra, siendo este la combinación de excel y papel, que representa un 4% y los demás métodos restantes que fueron despreciados que representan un 3% en conjunto.

Métodos en los que los negocios registran sus ventas:

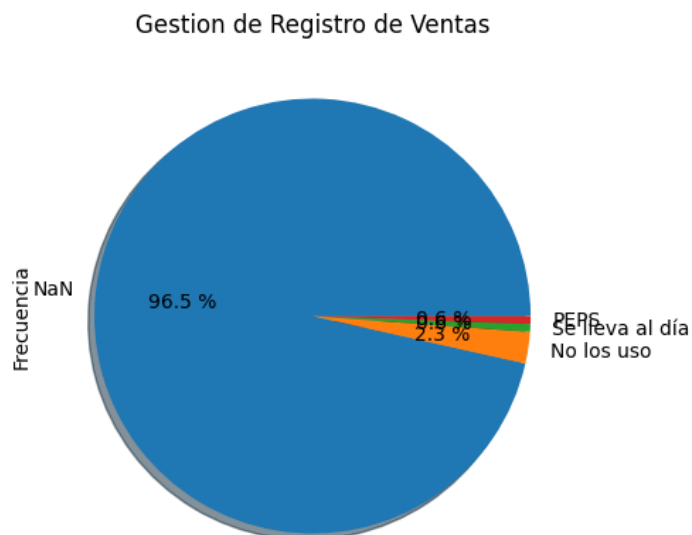
	frequency	percentage	cumulative_perc
18_sales_records			
NaN	61	0.356725	0.356725
En papel	47	0.274854	0.631579
Software especializado + computadora	22	0.128655	0.760234
No lo hago	14	0.081871	0.842105
En Excel	10	0.058480	0.900585
Caja registradora (análoga)	4	0.023392	0.923977
En Excel,En papel	4	0.023392	0.947368



De todos los negocios entrevistados solo se pudo obtener información del 65% con relación a los métodos de registro de ventas. De los que sí se obtuvo información la mayoría lo hacen en papel representando un 27% de la muestra total. Este diagrama es muy similar al diagrama anterior de métodos de registro de inventario, siendo que el porcentaje de información que se logró obtener de la muestra para ambas variables es el mismo y los métodos utilizados son los mismos en el mismo orden de mayor a menor frecuencia exceptuando la caja registradora que solo se encuentra en la variable de registro de ventas y los métodos con excel y no lo hacen que cambian de orden de la variable de registro de inventario a registro de ventas.

Gestión de registro de ventas:

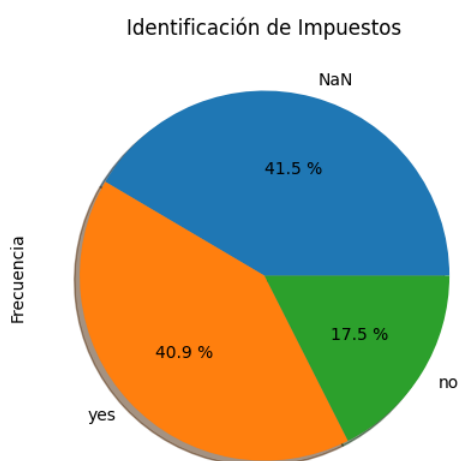
	frequency	percentage	cumulative_perc
155_sales_registers_used_for			
NaN	165	0.964912	0.964912
No los uso	4	0.023392	0.988304
Se lleva al día	1	0.005848	0.994152
PEPS	1	0.005848	1.000000



De todos los negocios entrevistados solo se pudo obtener información acerca de su metodo de registro de ventas del 4%. De los que sí se obtuvo información, la mitad no lo hacen y del resto de la muestra la mitad llevan el registro al día y la otra mitad usan el método PEPS. Esto puede implicar que hay poco interés en la gestión de registro de ventas de parte de los negocios pero dada la alta presencia de valores nulos no se puede concluir al respecto.

Negocios con identificación de impuestos:

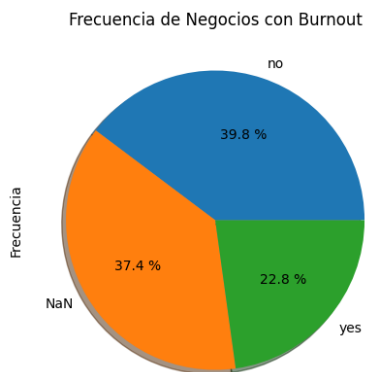
	frequency	percentage	cumulative_perc
19_tax_id			
NaN	71	0.415205	0.415205
yes	70	0.409357	0.824561
no	30	0.175439	1.000000



De todos los negocios entrevistados solo se pudo obtener información con relación a si tienen identificación de impuestos del 60%. De esos, más de la mitad si tienen identificación de impuestos siendo estos un 40% de la muestra total mientras que el 17% de la muestra total confirmaron que no lo tienen. Aún así la presencia de valores nulos es demasiado alta para hacer conclusiones al respecto.

Negocios con burnout:

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

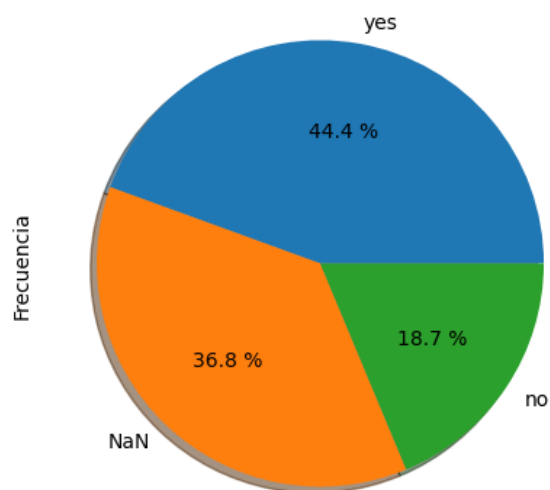


De los negocios entrevistados solo se pudo obtener información del 63% con relación a si sufren de burnout. De estos, casi 2 tercios confirmaron que no lo experimentan, siendo estos del 39.8% de la muestra total y el 22.8% restante de la muestra confirmaron que sí lo experimentan. En la gráfica se observa que la frecuencia de negocios que sí experimentan burnout, los que no y de los que no se obtuvo información están relativamente distribuidos.

Negocios con conexión a internet:

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

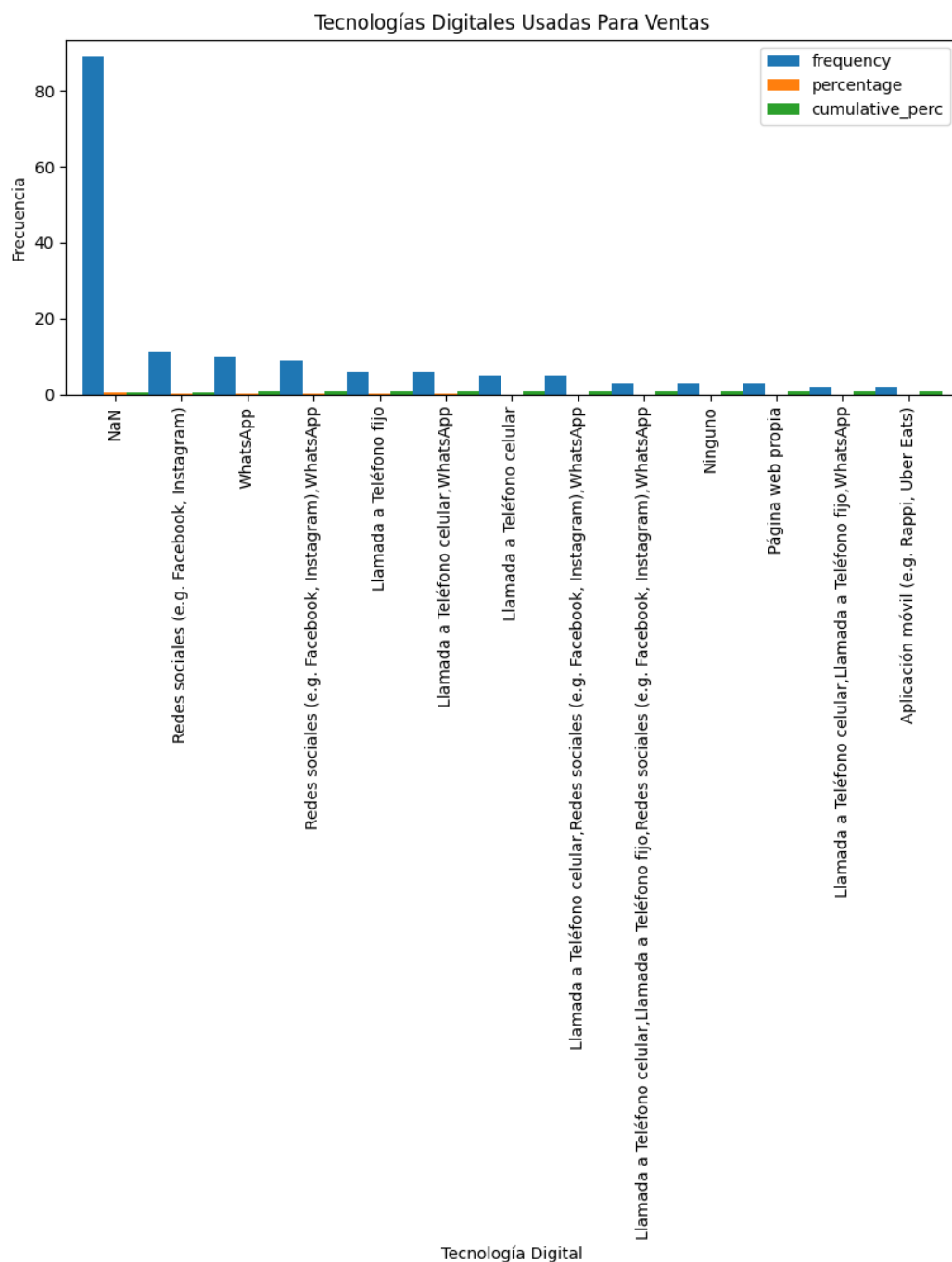
Frecuencia de Negocios con Conexión a Internet



De los negocios entrevistados solo se pudo obtener información del 64% con relación a si tienen conexión de internet. De estos más de la mitad si la tienen con un 44.4% de la muestra total contra un 18.7% de la muestra que no la tiene. Esto puede significar que los negocios en Puebla si tienen en su mayoría acceso a internet pero hay demasiados valores nulos para hacer conclusiones al respecto.

Tecnologías digitales para ventas:

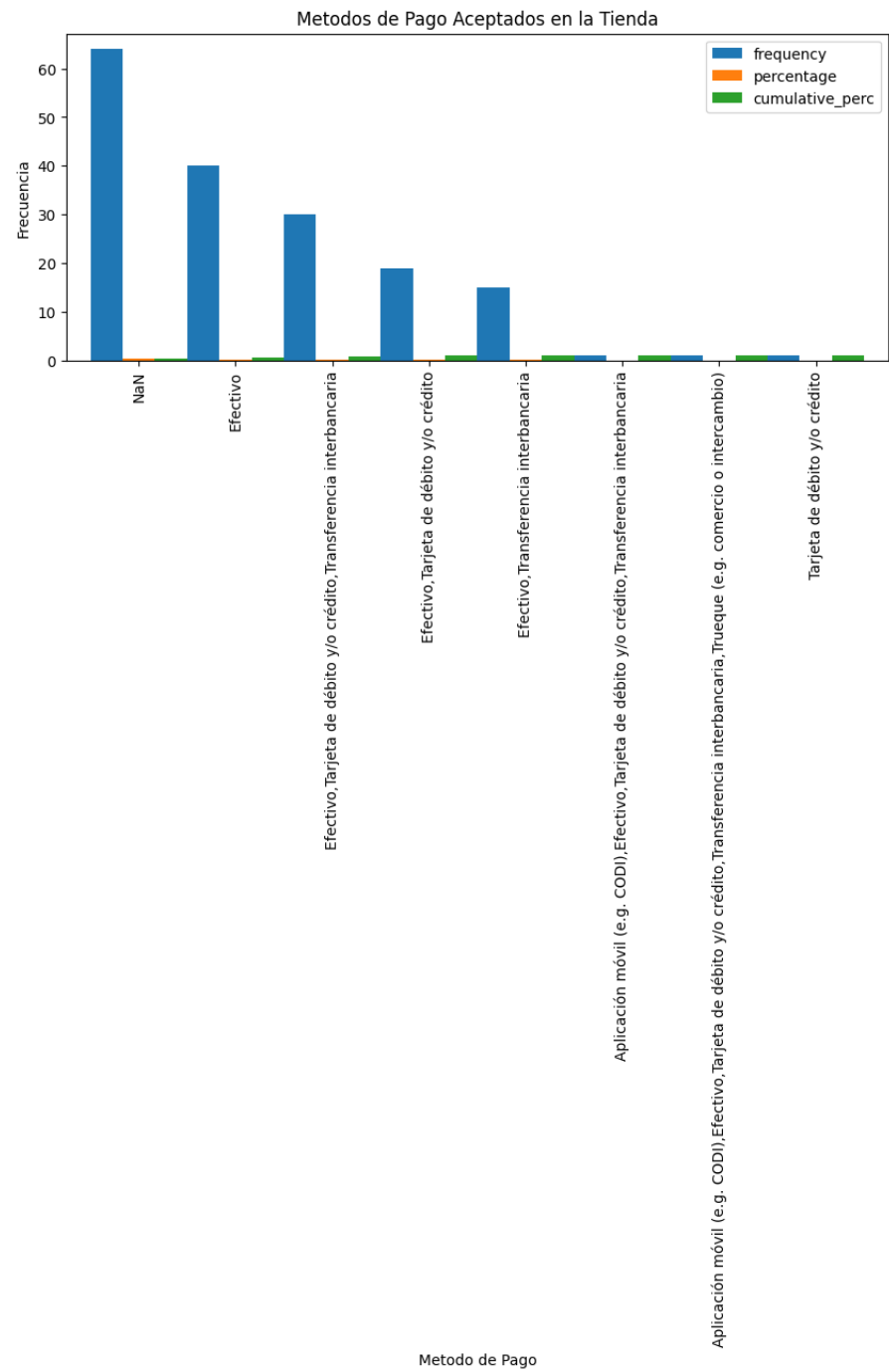
	frequency	percentage	cumulative_perc
210_sales_channels			
NaN	89	0.520468	0.520468
Redes sociales (e.g. Facebook, Instagram)	11	0.064327	0.584795
WhatsApp	10	0.058480	0.643275
Redes sociales (e.g. Facebook, Instagram),WhatsApp	9	0.052632	0.695906
Llamada a Teléfono fijo	6	0.035088	0.730994
Llamada a 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



De todos los negocios entrevistados solo se pudo obtener información de la mitad con relación al tipo de tecnologías digitales que utilizan en ventas. De estos la mayoría usan whatsapp, redes sociales, llamada a teléfono fijo o llamada a teléfono celular, o algunas de estas en combinación con otras. El uso de una página propia o aplicación móvil es poco frecuente siendo esta frecuencia de a lo mucho del 1.7%.

Métodos de pago:

	frequency	percentage	cumulative_perc
189_payment_methods			
NaN	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
Aplicación móvil (e.g. CODI), Efectivo, Tarjeta de débito y/o crédito, Transferencia interbancaria	1	0.005848	0.988304
Aplicación móvil (e.g. CODI), Efectivo, Tarjeta de débito y/o crédito, Transferencia interbancaria, Trueque (e.g. comercio o intercambio)	1	0.005848	0.994152
Tarjeta de débito y/o crédito	1	0.005848	1.000000



De los negocios entrevistados solo se pudo obtener información con relación al tipo de pago que aceptan del 63%. De los que sí se obtuvo información la mayoría solo aceptan efectivo con un 23% de la muestra. A partir de allí la diferencia entre las frecuencias de los diferentes métodos de pago tiende a ser similar hasta llegar al quinto método de pago más aceptado representando casi un 9% de la muestra cuya diferencia con el cuarto método más aceptado es menor que la diferencia del cuarto con el tercero. El resto de los pagos tienen a lo mucho una frecuencia de a lo mucho el 0.5% cada uno lo cual representa una caída drástica del 9% del quinto método más frecuente la cual se puede observar en el histograma.

Conclusión y Comentarios

Para muchas variables de la base de datos con la que se trabajó, se presentaron varias variables nulas. Esto puede significar que hubo una dificultad al realizar las entrevistas por parte de los estudiantes en semestres pasados lo cual pudo limitar la información que se logró obtener para su análisis. Para poder obtener información más a detalle acerca de los datos que se obtuvieron valdría la pena analizar porque varias de estas variables son nulas.