



ANÁLISIS DE DATOS EN LIBRERÍA ONLINE DE TUCUMÁN CASO DE APLICACIÓN

LIC. CRISTIAN E. ROCHA
FACULTAD DE CIENCIAS ECONÓMICAS – UNIVERSIDAD NACIONAL DE TUCUMÁN
CROCHA@FACE.UNT.EDU.AR

RESUMEN

La propia actividad de las empresas genera datos: una factura emitida, un pedido online, etc. Estos registros usualmente quedan almacenados en las bases de datos de la organización y no son aprovechadas adecuadamente. Es el caso de un emprendimiento tucumano llamado Infinity – Artículos de librería, cuya nueva plataforma online genera información que no se utiliza para la toma de decisiones.

El presente trabajo de investigación tiene como objetivo, aplicar diferentes métodos de análisis de datos con el fin de proporcionar información oportuna y relevante para la toma de decisiones del emprendimiento.

Para cumplir con el objetivo, se determino que el enfoque que guiara la investigación es el cuantitativo, de diseño no experimental de corte transversal con un alcance correlacional, con el fin de no manipular las variables y generar conocimiento a partir del análisis de los datos recopilados en la empresa.

Palabras Clave: Emprendimiento – Análisis multivariado – Datos

INTRODUCCIÓN

Infinity – Artículos de librería es un emprendimiento tucumano que se dedica a la comercialización de productos como resaltadores, marcadores, lapiceras y elementos que requieren los estudiantes de secundaria y universitarios en su trayecto educativo. Nació en el año 2019 como un emprendimiento de dos universitarios, cuya modalidad de venta se basaba en el contacto a través de redes sociales.

Actualmente cuentan con una tienda online donde se pueden solicitar los productos y un centro de logística que despacha los pedidos ya sea a domicilio o desde el mismo depósito. Sus clientes generalmente son hombre y mujeres, de entre 18 a 35 años de edad que demandan una amplia gama de productos.

Durante su trayecto, invirtieron en distintos rubros relacionados a los insumos requeridos por los estudiantes. Algunos productos fueron mas demandados que otros, lo cual ocasionó una disminución de la rotación en los inventarios en ciertas categorías. Esto implica que esos activos se encuentran estancados y se pierde la oportunidad de generar valor para el emprendimiento.

La nueva plataforma e-commerce permitió recopilar datos de los clientes y llevar un registro de sus movimientos de compras. Cada registro se almacena en una sección que permite la exportación de la información por completo. Por su parte, los dueños reconocen que la información de ventas no se explota adecuadamente.

Por lo que se definen como preguntas de investigación:



- ¿Cuáles son los tipos de productos más demandados?
- ¿Cuáles son los métodos de envío y de pago preferidos?
- ¿Existe relación entre las características de preferencia de los clientes y los productos?
- ¿Qué modelos se podrían aplicar con la finalidad de mejorar la toma de decisiones?

Objetivo

El presente trabajo de investigación tiene como objetivo, aplicar diferentes métodos de análisis de datos con el fin de proporcionar información oportuna y relevante para la toma de decisiones del emprendimiento

Para ello se plantean como objetivos específicos:

- Determinar las preferencias de los clientes de: productos, método de envío y de pago
- Determinar si existe relación entre las características de preferencia de los clientes
- Explotar los datos y proporcionar información valiosa para la toma de decisiones

Metodología

El enfoque del trabajo es de tipo cuantitativo ya que se pretende aplicar métodos estadísticos para analizar la información brindada por los dueños.

El diseño es de tipo no experimental de corte transversal debido que no se realiza manipulación de las variables y se realiza un análisis de la información en un momento determinado. El alcance definido es de tipo correlacional, ya que se evaluará si existe relación entre las variables

Como métodos de recolección de datos primarios, se utilizó la observación directa en la organización y como fuente de datos secundarios, la base de datos brindada por el departamento de comercialización del emprendimiento compuesta por los registros de ventas desde enero 2021 hasta septiembre 2022.

Marco teórico

Big Data

Según García *et al*; en su obra “Ciencia de datos. Técnicas analíticas y aprendizaje estadístico” publicado en 2018, define a Big Data como un conjunto de datos cuyo tamaño supera considerablemente la capacidad de captura, almacenado, gestión y análisis del software convencional de bases de datos. Sin embargo, el concepto no hace referencia simplemente al tamaño de la información, sino también a la variedad del contenido y a la velocidad con la que los datos se generan, almacenan y analizan. Estas dimensiones son las “3V” con las que la empresa Gartner describió big data, es decir volumen, velocidad y variedad de los datos:

- Volumen. Como su propio nombre indica, big data corresponde al gran volumen de datos que se generan diariamente en las empresas y organizaciones de todo el mundo.
- Velocidad. Se trata de los flujos de datos, la creación de registros estructurados y la disponibilidad para el acceso y la entrega.



- Variedad. Big data ha de tener la capacidad de combinar una gran variedad de información digital en los diferentes formatos en los que se puedan presentar. Esta variedad en los datos incluye datos estructurados (bases de datos) y no estructurados, datos jerárquicos, documentos, correo electrónico, datos de medición, vídeo, imágenes fijas, audio, datos de cotizaciones, transacciones financieras, etc., entre otras clases de fuentes generadoras de diferentes tipos de información.

Análisis de correspondencia

El análisis de correspondencias es una técnica descriptiva para representar tablas de contingencia. Los datos de partida para el análisis de correspondencias es una matriz X de dimensiones $n \times k$ que representa las frecuencias absolutas observadas en una tabla de contingencia de dos variables, donde la primera se representa por filas y tiene n categorías y la segunda por columnas y tiene k categorías. Por ejemplo, clasificamos un conjunto de estudiantes en n posibles procedencias geográficas y k posibles opiniones respecto a la docencia. En general, el elemento x_{ij} de la matriz X representa la frecuencia absoluta observada en la casilla (i, j) de la tabla de contingencia.

La metodología la desarrolló Benzecri, a principios de los años 60 del siglo XX en la Universidad de Renner (Francia). En esencia, es un tipo especial de análisis de componentes principales pero realizado sobre una tabla de contingencia y usando una distancia euclídea ponderada llamada chi-cuadrado

Variable estadística

Una variable estadística es una característica de una muestra o población de datos que puede adoptar diferentes valores. Estas se pueden clasificar en:

- Variable cuantitativa: Son variables que se expresan numéricamente.
 - Variable continua: Toman un valor infinito de valores entre un intervalo de datos. El tiempo que tarda un corredor en completar los 100 metros lisos.
 - Variable discreta: Toman un valor finito de valores entre un intervalo de datos. Número de helados vendidos.
- Variable cualitativa: Son variables que se expresan, por norma general, en palabras.
 - Variable ordinal: Expresa diferentes niveles y orden.
 - Variable nominal: Expresa un nombre claramente diferenciado. Por ejemplo el color de ojos puede ser azul, negro, castaño, verde, etc.



Diagnóstico de la empresa

El emprendimiento Infinity, en sus inicios, comercializaba a través de las redes sociales, priorizando el vínculo con el cliente. Poco a poco, fue ganando cuota de mercado lo cual generó la necesidad de sistematizar el proceso de ventas. En el año 2020, invirtieron en la integración de una tienda online que simplificaba el contacto con el cliente.

La tienda online, a medida que se generan ventas, recopila los datos personales de los clientes en una base, los productos solicitados y el método de envío preferido.

El proceso de compra de un cliente sigue los siguientes pasos:

Gráfico 1. Proceso de compra de cliente



Fuente: Elaboración propia

Imagen 1. Captura de pantalla de base de datos

#	Cliente	Pago	Fecha	Método de pago	Método de envío	Envío	Costo del envío	Monto (sin envío)
121	Cristian Emanuel Rocha	Cancelado	25/10/22	Transferencia / Depósito	Punto de Retiro	Pendiente de retiro	-	\$2.782,50
120	Luz Nuñez	Cancelado	22/10/22	Transferencia / Depósito	Punto de Retiro	Pendiente de retiro	-	\$8.147,16
119	Luz Nuñez	Cancelado	20/10/22	Transferencia / Depósito	Punto de Retiro	Pendiente de retiro	-	\$8.809,40
118	Cecilia Barrios	Finalizado	18/10/22	Mercado Pago	Cadetería - Zona Sur	Enviado	\$250,00	\$1.157,52
117	Florencia Ocampos	Finalizado	15/10/22	Mercado Pago	Cadetería - Zona Sur	Enviado	\$250,00	\$1.157,52
116	Jimena Romano	Finalizado	07/10/22	Transferencia / Depósito	Punto de Retiro	Retirado	-	\$1.057,36
115	José Castro	Finalizado	26/09/22	Transferencia / Depósito	Punto de Retiro	Retirado	-	\$1.357,87
114	Celina Castañeda	Finalizado	18/09/22	Mercado Pago	Cadetería - Zona Centro	Enviado	\$200,00	\$2.192,61
113	Celina Castañeda	Cancelado	18/09/22	Mercado Pago	Cadetería - Zona Centro	Pendiente	\$200,00	\$2.125,83
112	Agostina Miragliotta	Finalizado	18/09/22	Transferencia / Depósito	Punto de Retiro	Retirado	-	\$729,02

Fuente: Departamento de Comercialización

Los dueños reconocen que adquieren nuevos productos basados en los comentarios de clientes frecuentes y las nuevas tendencias del mercado. Las inversiones representan el 50% de los beneficios generados anualmente.

Según los indicadores de operaciones, la rotación de algunos productos sobrepasa los 6 meses, es decir, transcurre más de medio año desde que se compró la mercadería hasta que se vendió.



Variables de interés

Debido a lo antes mencionado, es fundamental para los dueños conocer las preferencias de los clientes para disminuir el tiempo de rotación y potenciar el flujo de ingresos.

Dada la base de datos, se pudieron identificar cuatro variables de interés:

- **Sexo:** entendiéndose como el sexo biológico del cliente. Este puede ser femenino o masculino
- **Estado de pago:** Los clientes tienen la opción de cancelar su pedido antes de ser despachado. El estado de pago puede ser finalizado o cancelado.
- **Método de envío:** Algunos clientes prefieren retirar su pedido directamente del depósito (punto de retiro) mientras que otros prefieren recibirlo en su domicilio.
- **Productos:** Al ser tan amplia la oferta, se optó por clasificarlo en rubros:
 - ARTOFIC: Son todos aquellos productos que se utilizan en oficinas como lapiceros, notas adhesivas, abrochadoras y clips.
 - LMR: Son lápices, lapiceras, resaltadores y marcadores. Incluye su repuesto, recarga y relacionados como sacapuntas y correctores.
 - PAPEL: Incluye cuadernos, libretas, agendas y repuestos de hojas.
 - VS: Son todos los productos que quedan fuera de los antes mencionados como cartucheras, mochilas y box de regalos.

Data Cleaning

Un proceso fundamental previo al análisis, es limpiar la base de datos. Esto permite contar con la información de maneca adecuada para procesarla.

Como se observa en la imagen 1, la base de datos cuenta con más de 15 variables.

Imagen 2. Base de datos de ventas original

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	#	Fecha	Cliente	Email	Teléfono	Pago	Método de pago	Envío	Método de envío	Costo de envío	Total (sin envío)	Tipo de envío	Sucursal	Envío - Nombre	Envío - Apellido
2	121	25/10/22	19:42	Cristian Emanuel Rocha	crocha@face.unt.edu.ar	\$43814060615	Cancelado	Transferencia / Depósito	Pendiente de retiro	Punto de Retiro	0	2782,5		Cristian Emanuel	Rocha
3	120	22/10/22	13:25	Luz Nuñez	luz.beauylace@gmail.com	\$43815261665	Cancelado	Transferencia / Depósito	Pendiente de retiro	Punto de Retiro	0	5147,16		Luz	Nuñez
4	119	20/10/22	19:46	Cecilia Barrios	luromanonunez@gmail.com	\$43815261665	Cancelado	Transferencia / Depósito	Pendiente de retiro	Punto de Retiro	0	5809,4		Luz	Nuñez
5	118	18/10/22	19:46	Cecilia Barrios	ceciliarbarrios22@gmail.com	\$43813005008	Finalizado	Mercado Pago	Enviado	Cadetería - Zona Sur	250	1157,52	Domicilio	Cecilia	Barrios
6	117	15/10/22	9:03	Florencia Ocampos	flor.ocampos367@gmail.com	\$43816367935	Finalizado	Mercado Pago	Enviado	Cadetería - Zona Sur	250	1157,52	Domicilio	Florencia	Ocampos
7	116	07/10/22	14:00	Jimena Romano	romanojimena29@gmail.com	\$43813563616	Finalizado	Transferencia / Depósito	Retirado	Punto de Retiro	0	1057,36		Jimena	Romano
8	115	26/09/22	18:02	José Castro	romiliumada12@gmail.com	\$43815528650	Finalizado	Transferencia / Depósito	Retirado	Punto de Retiro	0	1357,87		José	Castro
9	114	18/09/22	22:34	Celina Castañeda	celinadv@gmail.com	\$43815231132	Finalizado	Mercado Pago	Enviado	Cadetería - Zona Centro	200	2192,61	Domicilio	Celina	Castañeda
10	113	18/09/22	22:23	Celina Castañeda	celinadv@gmail.com	\$43845231132	Cancelado	Mercado Pago	Pendiente	Cadetería - Zona Centro	200	2125,83	Domicilio	Celina	Castañeda
11	112	18/09/22	18:27	Agostina Miragliotta	anarivademeira73@gmail.com	\$43813957809	Finalizado	Transferencia / Depósito	Retirado	Punto de Retiro	0	729,02		Agostina	Miragliotta
12	111	17/09/22	21:33	Agostina Miragliotta	anarivademeira73@gmail.com	\$43813957809	Cancelado	Transferencia / Depósito	Pendiente de retiro	Punto de Retiro	0	795,8		Agostina	Miragliotta
13	110	16/09/22	12:55	Florencia Ocampos	flor.ocampos367@gmail.com	\$43816367935	Finalizado	Mercado Pago	Enviado	Cadetería - Zona Sur	250	1535,94	Domicilio	Florencia	Ocampos
14	109	08/09/22	11:26	Noelia Torres	paolatove@gmail.com	\$41133880795	Cancelado	Mercado Pago	Pendiente	Cadetería - Zona Centro	200	601,36	Domicilio	Noelia	Torres
15	108	07/09/22	11:30	Cristian Emanuel Rocha	crocha@face.unt.edu.ar	\$43814060615	Finalizado	Transferencia / Depósito	Retirado	Punto de Retiro	0	511,98		Cristian Emanuel	Rocha
16	107	07/09/22	11:19	Gisela Ruiz	valezpn@gmail.com	\$43814498974	Finalizado	Transferencia / Depósito	Enviado	Cadetería - Zona Centro	200	645,88	Domicilio	Gisela	Ruiz
17	106	06/09/22	17:08	Sofía Agustina Valerio	sofi_agust28@hotmail.com	\$43814900142	Finalizado	Mercado Pago	Retirado	Punto de Retiro	0	5025,67		Sofía Agustina	Valerio
18	105	29/08/22	17:40	Rocio Garcia	mrocogarciasch@gmail.com	\$43815974653	Finalizado	Transferencia / Depósito	Retirado	Punto de Retiro	0	424		Rocio	García

Fuente: Departamento de comercialización

Dado que el propósito del trabajo se relaciona con las variables de interés, se eliminaron las columnas que no son de importancia para este análisis. Solo se trabajó con cuatro variables.



Imagen 3. Base de datos de ventas limpia

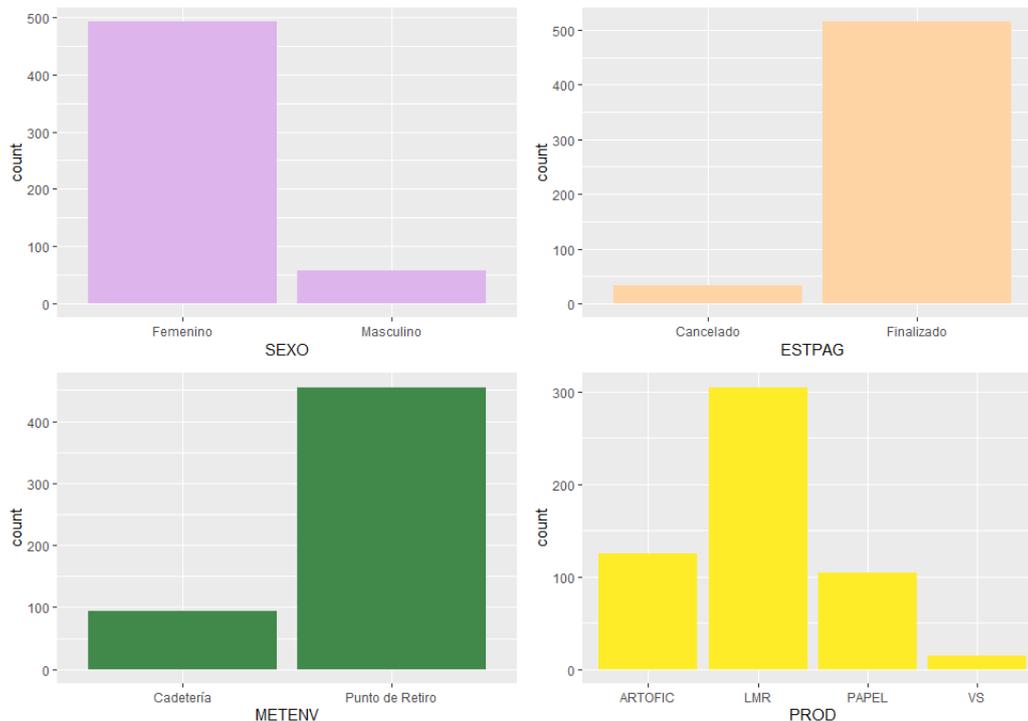
B	C	D	E
SEXO	METPAG	METENV	PROD
Femenino	Mercado Pago	Cadetería - Zona Sur	ARTOFIC
Femenino	Mercado Pago	Punto de Retiro	ARTOFIC
Femenino	Transferencia / Depósito	Punto de Retiro	ARTOFIC
Femenino	Mercado Pago	Cadetería - Zona Centro	ARTOFIC
Femenino	Mercado Pago	Punto de Retiro	ARTOFIC
Femenino	Mercado Pago	Cadetería - Zona Centro	ARTOFIC
Femenino	Mercado Pago	Punto de Retiro	ARTOFIC
Femenino	Transferencia / Depósito	Punto de Retiro	ARTOFIC
Femenino	Mercado Pago	Punto de Retiro	ARTOFIC
Femenino	Transferencia / Depósito	Punto de Retiro	ARTOFIC
Femenino	Mercado Pago	Punto de Retiro	ARTOFIC
Femenino	Transferencia / Depósito	Punto de Retiro	ARTOFIC
Femenino	Mercado Pago	Punto de Retiro	ARTOFIC
Femenino	Transferencia / Depósito	Punto de Retiro	ARTOFIC
Femenino	Mercado Pago	Punto de Retiro	ARTOFIC
Femenino	Transferencia / Depósito	Punto de Retiro	ARTOFIC

Fuente: Elaboración propia

Análisis de datos

Se procedió a analizar la base de datos de la empresa del último año tomando en cuenta las variables de interés mencionadas en el párrafo anterior.

Gráfico 2. Gráfico de barras de variables de interés



Fuente: Elaboración propia

- Es muy frecuente la compra de clientes de sexo femenino



- Del total de pedidos realizados, fueron cancelados un 10%
- Los clientes prefieren retirar desde el deposito mas que recibirlo en sus hogares
- Existe una clara preferencia por el rubro LMR, seguido de ARTOFIC, PAPEL y por último VS

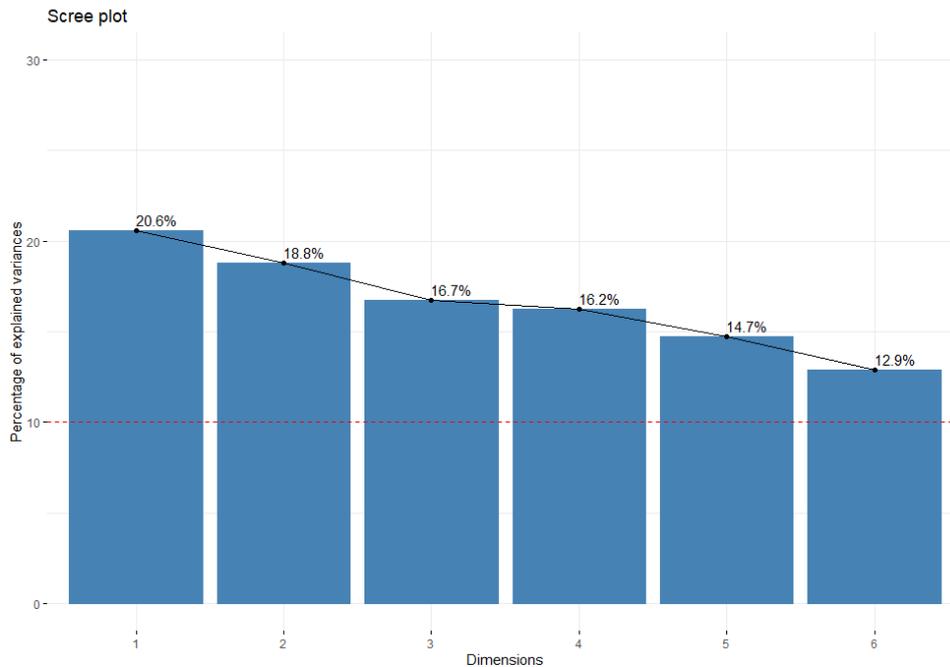
Debido que el objeto del estudio fue analizar si existe una relación entre las distintas variables antes mencionadas, se optó por aplicar un análisis de correspondencias múltiple, ya que las variables son categóricas y lo que se busca es determinar la interdependencia entre las mismas.

Tabla 1. Porcentaje de explicación por dimensión

 	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	0.3085	20.57	20.57
Dim.2	0.2821	18.8	39.37
Dim.3	0.251	16.73	56.11
Dim.4	0.2437	16.25	72.35
Dim.5	0.2211	14.74	87.09
Dim.6	0.1936	12.91	100

Fuente: Elaboración propia

Gráfico 3. Gráfico de barras del porcentaje de explicación por dimensión



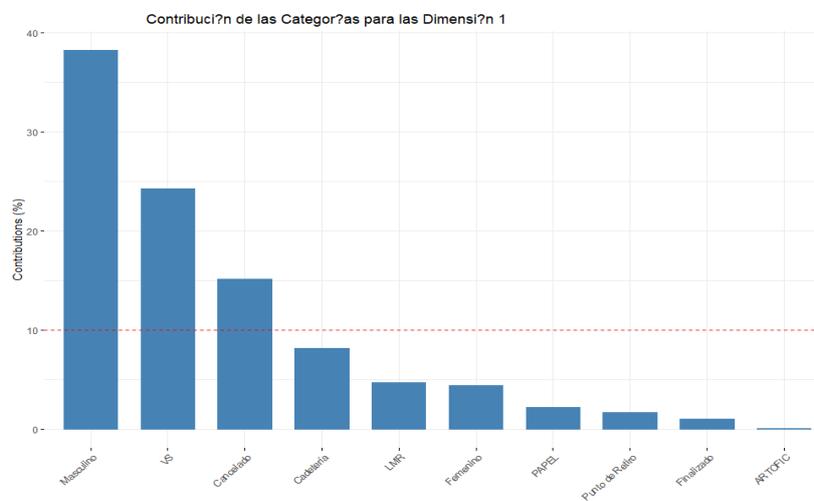
Fuente: Elaboración propia



Al analizar la cantidad de dimensiones que me explicarían el mayor porcentaje de variabilidad en los datos, se puede apreciar que, utilizando dos dimensiones, podría explicar 40% aproximadamente.

Se procedió a determinar el nivel de contribución de las categorías de las variables a la explicación de cada una de las dimensiones definidas.

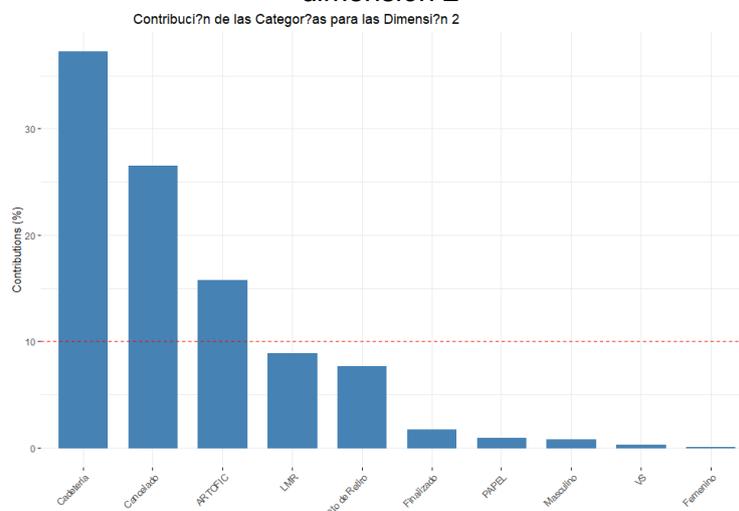
Gráfico 4. Gráfico de barras del porcentaje de contribución de categorías a la dimensión 1



Fuente: Elaboración propia

Se puede observar que el comportamiento de la dimensión 1, se explica por las categorías masculino de sexo, varios de productos y cancelado del estado de pago.

Gráfico 5. Gráfico de barras del porcentaje de contribución de categorías a la dimensión 2

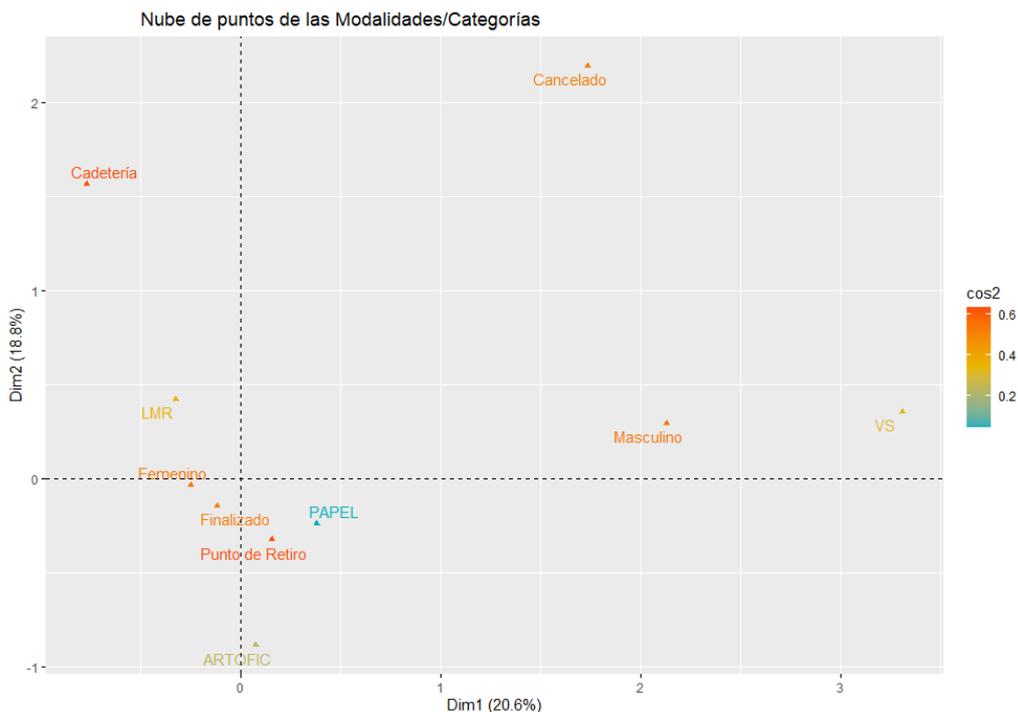


Fuente: Elaboración propia



En cuenta al comportamiento de la dimensión 2, se explica por las categorías Cadetería de Método de envío, Cancelado de estado de pago y ARTOFIC de productos.

Grafico 6. Mapa perceptual de variables



Fuente: Elaboración propia

Respecto al análisis del mapa perceptual, podemos observar lo siguiente:

- Existe una clara relación entre el cliente de sexo femenino con una preferencia por productos como resaltadores, lapiceras, marcadores y papelería.
- Si el cliente es femenino es más probable que finalice su compra y retire el pedido desde el depósito.
- Aquellos clientes que solicitaron productos de papelería como cuadernos o repuestos de hojas no pagan el servicio de Cadetería.
- Es más probable que los clientes que demandan artículos de oficina, retiren sus productos desde el depósito.
- No existe una relación entre el público femenino y el rubro de varios.



Conclusiones y recomendaciones

- Es importante destacar que la información si no es explotada adecuadamente, pierde su valor para la toma de decisiones
- Realizar un análisis descriptivo univariado puede ser apropiado si se busca determinar medidas generales para formar un conocimiento básico, pero como la realidad es compleja, es necesario considerar el análisis multivariado para comprender mejor el fenómeno bajo estudio
- En el caso del emprendimiento, es fundamental que se inviertan en los rubros de lapiceras, lápices, relatadores y marcadores. Son los productos más demandados
- Se recomienda minimizar el ingreso de productos relacionados al rubro varios
- Ya que el cliente femenino representa la mayor proporción de las ventas y son las que tienen mas probabilidad de que no cancele su pedido, se debería implementar un CRM para generar un vínculo más directo y apropiado con este segmento, detectar oportunidades de mejora y generar información valiosa para satisfacer sus necesidades



REFERENCIAS

DE LA GARZA GARCIA, Jorge; MORALES SERRANO, Blanca; GÓNZALEZ CAVAZOS, Beatriz (2013). "Análisis Estadístico Multivariante. Un enfoque teórico y práctico." Mc Graw Hill. México.

GARCÍA, Jesús; MOLINA, José; BERLANG, Antonio; PATRICIA Miguel; BUSTAMANTE, Álvaro; PADILLA, Washington (2018). "Ciencia de datos. Técnicas analíticas y aprendizaje estadístico". Alfaomega. Chile.

HAIR, Joseph; ANDERSON, Rolph; TATHAM, Ronald; BLACK, William (2019). "Análisis Multivariante". Cengage Learning. México.

KELLEHER, Johnr; TIERNEY, Brendan (2021). "Ciencia de Datos". MIT Press. Chile.

URIEL, Ezequiel; ALDÁS MANZANO, Joaquín (2017). Análisis Multivariante Aplicado con R. Alfacentauro. España.