

Análisis de Ventas en una Empresa Agroindustrial por medio del Reconocimiento de Patrones en los Datos.

Daniel Arturo Olivares-Vera ^{a*}, Ma. de los Angeles Gómez-Castro ^a, José Israel Hernández-Vázquez ^b,
José Omar Hernández-Vázquez ^b, Angela del Pilar Olivares-Vera ^c

^a Tecnológico Nacional de México/Instituto Tecnológico de León. Departamento de Ciencias Económica Administrativa. Av. Tecnológico S/N, Fracc. Julián de Obregón, León, Gto., México.

^b Tecnológico Nacional de México/Instituto Tecnológico de León. Departamento de Ingeniería Industrial. Av. Tecnológico S/N, Fracc. Julián de Obregón, León, Gto., México.

^c Escuela Nacional de Estudios Superiores Unidad León (UNAM-ENES León). Licenciatura en Economía Industrial. Blvd. UNAM #2011, Predio El Saucillo y, Comunidad de los Tepetates, El Potrero, 37684 León, Gto., México.

Resumen

La presente investigación muestra el análisis del comportamiento de las ventas de una empresa de la agroindustria, donde se busca encontrar patrones por medio del análisis de datos aplicados a las ventas de siete años (2016-2022), con la finalidad de realizar estrategias dirigidas y cluster de productos en base a los resultados encontrados por medio de modelos de regresión lineal, análisis de varianza (ANOVA), mapas de calor, clusterización por medio de k-means, dendogramas.

Abstract

The present investigation shows the analysis of the behavior of the sales of an agribusiness company, where it is sought to find patterns through the analysis of data applied to the sales of seven years (2016-2022), with the purpose of carrying out strategies directed and cluster of products based on the results found through linear regression models, analysis of variance (ANOVA), heat maps, clustering through k-means, dendrograms.

Palabras claves: Análisis de Datos, Tendencia de Ventas, ANOVA, Clusterización y k-means.

Keywords: Data Analysis, Sales Trend, ANOVA, Clustering and k-means.

1. INTRODUCCIÓN

La investigación se realizó tipo caso en una empresa del sector agroindustrial, donde se analizaron siete años de ventas de diversos productos.

A. Estrategias de mercados

Las empresas han tenido ventajas competitivas importantes, tales como las que demuestran [1], de las cuales se podrían mencionar: “mejora de la gestión empresarial, velocidad de la toma de decisiones”, las estrategias mercadológicas compiten para que sus productos sean competitivos y accesibles al mercado en donde está su enfoque. También destaca [1] que: “existen múltiples beneficios y áreas de oportunidad que las pequeñas empresas y nuevos negocios pueden atacar en busca de la permanencia en el mercado y del camino a la cima” las estrategias de mercado dan un realce para dar apoyo en la actividad a ser estudiada y llevar un análisis para la estructura de su FODA, los pronósticos, como herramienta de fases metodológicas, recolectando información del entorno actual, ya que en un sistema globalizado es también importante entender la demanda del producto, empleando estrategias de mercado. Se atienden los esquemas competitivos del país, ya que conllevan importantes actualizaciones de diversas posiciones en el ambiente comercial, así lo han demostrado. [2] Incluye un comparativo de las unidades producidas en diferentes estados del país. [3] Conceptualiza “la capacidad de la empresa para producir y mercadear productos en mejores condiciones de precio, calidad y oportunidad” Ya que este considera la importancia de que el país siendo productivo y teniendo un desarrollo

gradual también el progreso se obtendrá en lo empresarial. Los esquemas aplicados a los diversos cambios en las estructuras de la comercialización afectan y/o benefician la adquisición de los consumidores, dependiendo el estatus que represente este sector y los beneficios y oportunidades de cada condición económica. El apoyo del gobierno para comercializar en la actualidad es sumamente importante para este ramo de la industria, la empresa se encargará de crear y actualizar el diseño de sus estrategias. En la actualidad la creación y crecimiento de las empresas de calzado en nuestro país representa una aportación de fuentes de trabajo, así como la oportunidad de desarrollo económico de las regiones en donde se maquila el calzado. Una propuesta interesante del economista Adam Smith” para el correcto funcionamiento de una economía de mercado, es que el Estado debe regular e intervenir para garantizar el acceso a todos los agentes económicos a los mecanismos de mercado, posibilitando una economía de mercado abierta a todos los hombres” (pp. 203-207).

B. La Ciencia de Datos

La competencia entre organizaciones en la búsqueda de las preferencias de sus clientes es cada vez más demandada. Las necesidades cambian, como ejemplo esta el tiempo de vida de la moda, la cual ha demostrado ser corta.

Según [4] comenta que la clasificación de los datos se divide en:

- Estructurados (Datos tradicionales, Datos con campos fijos como hojas de cálculo, archivos, etc.).
- Semiestructurados (No tienen formatos fijos, pero tienen etiquetas o marcadores (XML y HTML).
- No Estructurados (Datos Big Data se almacena como documentos u objetos no tienen campos fijos: audio, video, fotografías, artículos, libros, SMS, WhatsApp)

Así mismo es importante conocer la clasificación de los datos en función a su volumen, la ciencia de los datos es el área responsable de la interpretación y manipulación de los datos; la evolución de esta ha sido desde la Estadística pasando por la Minería de Datos siguiendo al Análisis Predictivo a la Ciencia de Datos. La Minería de datos aplicada en los pequeños negocios, es posible, pero se debe tomar varios aspectos, tal como comenta [5,6,7,8,9,10,11]. Este creciente interés por la adopción de minería de datos en las pequeñas y medianas empresas tiene una tendencia al alza, tal como lo comenta [12]. La ejecución de la ciencia de datos implica, tener un amplio conocimiento en ciencias de la computación, conocimientos en matemáticas, estadísticas y gestión de negocios. En ese mismo sentido el proceso a seguir implica una transición que de manera general van en la búsqueda de información, preparación de datos, uso de modelos matemáticos.

C. La analítica de datos como ventaja competitiva en las organizaciones

Gracias al mundo cambiante y a las múltiples exigencias del mercado empresarial actual, las multinacionales, pequeñas y medianas empresas y nuevos negocios han comenzado a cambiar su modelo de negocio y adaptarlo a las nuevas prácticas tecnológicas de hoy en día. El Análisis de datos se ha convertido en una de las llaves principales en la permanencia en el mercado empresarial. En el presente trabajo teórico se estudian los múltiples beneficios, principales características y las decisiones más importantes que las empresas han tomado y deben tomar en su camino a la actualización empresarial. El objetivo es presentar y demostrar, con base en la literatura revisada, la importancia del Análisis de datos y su impacto real en la creación de ventajas competitivas para las empresas.

D. Proyección de Ventas

[13] Habla sobre la importancia de realizar Análisis entre factores que afecten el desempeño de las organizaciones bajo la ciencia de datos para identificar los conglomerados de clientes en función a su comportamiento incluyendo herramientas de visualización para claridad de la interpretación de los datos y las relaciones entre variables las variables con la finalidad de realizar propuestas eficaces.

E. Comportamiento de las Ventas

Existen distintos estudios para determinar el tamaño del mercado y el arquetipo de compradores en función al análisis de datos, como el realizado por [14], para conocer el mercado por medio de herramientas Estadísticas y proyecciones incluyendo herramientas de visualización y algoritmos de modelación, que pueden apoyar a encontrar aspectos como mejores compradores, segmento de clientes, tipos de productos en función a las ventas para definir tamaño del mercado latente a atender.

F. Pronósticos

[15] Explica como la ciencia de datos por medio del manejo de grandes cantidades de datos ha apoyado a anticiparse por medio de una correcta toma de decisiones, usada también en el sector agrícola, de la misma manera [16], muestra como la utilización de datos en el sector agroindustrial, puede apoyar a la realización de proyecciones y pronósticos, por medio de construir modelos explicativos con métodos univariados.

G. Análisis de Mercado

Los Análisis de mercado son importantes en función de conocer cómo se comportan y de esta manera poder ofrecer productos específicos, dentro de ellos destaca la competitividad sistémica y factores económicos, socio-culturales, ecológico-ambientales, tecnológico-infraestructura, y políticos-institucionales como lo comenta [17], donde realiza un Análisis de Componentes Principales sobre los factores, con el fin de caracterizar y jerarquizar los factores que mayor a la competitividad, donde el conocimiento de mercado por medio de la innovación brinda una ventaja competitiva, como lo comenta explica [18].

2. METODOLOGÍA

El desarrollo de la investigación se llevará a cabo conforme a la Figura 1. Metodología. El estudio usa una base de 453,240 datos, compuestos en 9 variables de las cuales están compuestos 8 son variables independientes (Año, Mes, Zona ventas, Estado, Municipio, Producto, Vendedor, Cliente) y una variable dependiente que son las Ventas mensuales por producto, la cual es denominada como “Ventas”.



Figura 1. Metodología.

Para los Análisis se utilizaron diversas herramientas de ciencias de datos, tales como modelos de regresión lineal, Análisis de varianza (ANOVA), mapas de calor, clusterización por medio de k-means, dendogramas, con la finalidad de realizar una representación clara del comportamiento de los datos en sus diferentes dimensiones.

3. DESARROLLO

A. Obtener la Información

La Información se genera del ERP (Enterprise Resource Planning) de la empresa, en la cual se obtiene una base de datos de la empresa del caso, en ella se obtienen las 8 variables independientes y una variable dependiente mencionadas anteriormente como la resultante de las ganancias de las ventas.

B. Tratamiento de Datos

Los datos fueron tratados para poder ser analizados, dentro de los ajustes de estos, están los realizados para convertir en factor a aquellos que estaban catalogados como carácter, con la finalidad de poder hacer agrupaciones, de distinta índole, como lo es por estado, producto, fecha, etc., tal como se muestra en el Tabla I. Base de Datos de Ventas.

Tabla I. Base Datos de Ventas

A	Mes	Z.	Est.	Mun.	Prod.	Vend.	Cli.	Total
2016	ene	Bj.	BC.	Mxli	P.17	V.10	C. 01	320796
2016	ene	Bj.	BC.	Mxli	P.25	V.10	C. 01	568548.94
2016	ene	Bj.	BC.	Mxli	P.32	V.10	C. 01	223488
2016	ene	Bj.	BC.	Mxli	P.09	V.10	C. 01	29952
...
2022	jul	S.E.	Ver.	J.R.	P.30	V.02	C.322	19609.98

C. Análisis Estadístico

Para poder definir los factores que afectan a las ventas, se procedió a realizar un análisis de varianza (ANOVA) para conocer la significancia de cada uno de los factores a analizar, los resultados se muestran en el Tabla II. ANOVA de Datos de Ventas, se observan los resultados, en función a su significancia: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1, donde se puede observar que solo la variable vendedor no es significativa con un 95% de nivel de confianza.

Tabla II. ANOVA de Datos de Ventas

Cli.	Df	Sum Sq	Mean Sq	Pr(>F)	
A	6	6.24E+12	1.04E+12	<2e-16	***
Mes	11	3.42E+11	3.11E+10	0.00509	**
Zona	16	1.19E+14	7.43E+12	<2e-16	***
Est.	29	1.42E+13	4.91E+11	<2e-16	***
Mun.	142	2.88E+13	2.03E+11	<2e-16	***
Prod.	90	1.77E+14	1.96E+12	<2e-16	***
Vend.	5	2.09E+10	4.18E+09	0.89772	
Cli.	133	3.18E+13	2.39E+11	<2e-16	***
Fecha	61	1.24E+12	2.04E+10	0.00234	**
Residuals	50130	6.43E+14	1.28E+10		

D. Agrupar y Clusterizar

- 1) *Análisis de Ventas por Estado:* Los análisis realizados de las ventas por estado, demuestran una similitud en su comportamiento, los cuales se agrupan como se muestran en el Tabla III. Análisis de Ventas por Estado, muestra la clasificación resultante del listado de estados.

Tabla III. Análisis de Ventas por Estado

Estado	Grupo
NORTE	a
BAJA CALIFORNIA	b
SONORA	b
BAJA CALIFORNIA SUR	bc
CHIHUAHUA	c
NUEVO LEON	cd
VERACRUZ	cd
COLIMA	cd
HIDALGO	cd
QUERETARO	cd
SINALOA	cd
JALISCO	cd
PUEBLA	cd
HERMOSILLO	cd
AGUASCALIENTES	d
MORELOS	d
DURANGO	d
CHIAPAS	d
TABASCO	d
MICHOACAN	d
ZACATECAS	d
CDMX	d
CAMPECHE	d
TAMAULIPAS	d
COAHUILA	d
GUANAJUATO	d
SAN LUIS POTOSI	d
EDO. DE MEXICO	d
OAXACA	d
YUCATAN	d
GUERRERO	d
QUINTANA ROO	d

- 2) *Análisis de Ventas por Año:* Posteriormente se continuará con realizar un análisis por medio de técnicas de clusterización, como es el caso del k-means, donde se obtiene que es adecuado considerar cuatro clústeres, para encontrar algún tipo de patrón para poder agrupar el comportamiento de ventas por estado, tal como se muestra en la Figura 2. Numero de k por Estado.

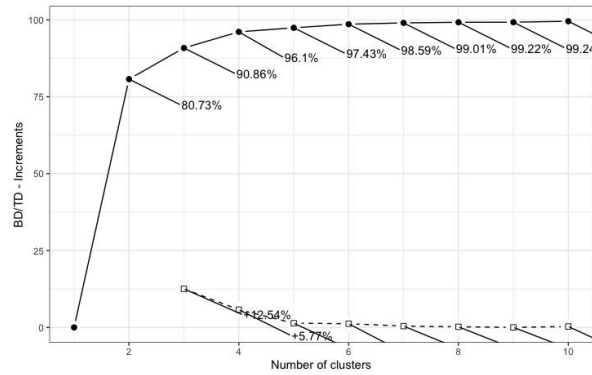


Figura 2. Numero de k por Estado.

Profundizando en los Análisis del comportamiento de las ventas, se procedió a revisar el número de cluster para Municipio, Producto, Cliente y Mes como se muestra en la Figura 3. a) detalla en el concepto municipio y el comportamiento de las ventas, en donde se encontro que ocho cluster son los adecuados considerar, en el caso de los productos se definio la cantidad de grupos o cluster que mejor explicaban el comportamiento de las ventas, en donde se encontro que cinco cluster son los adecuados a considerar para los productos, en el caso de los clientes se encontró que cinco cluster son los adecuados considerar, finalmente para el caso de los meses, se encontró que cuatro cluster son los adecuados considerar tal como se muestra respectivamente en la Figura 3. Numero de k cluster por: a) Municipio, b) Producto, c) Cliente y d) Mes.

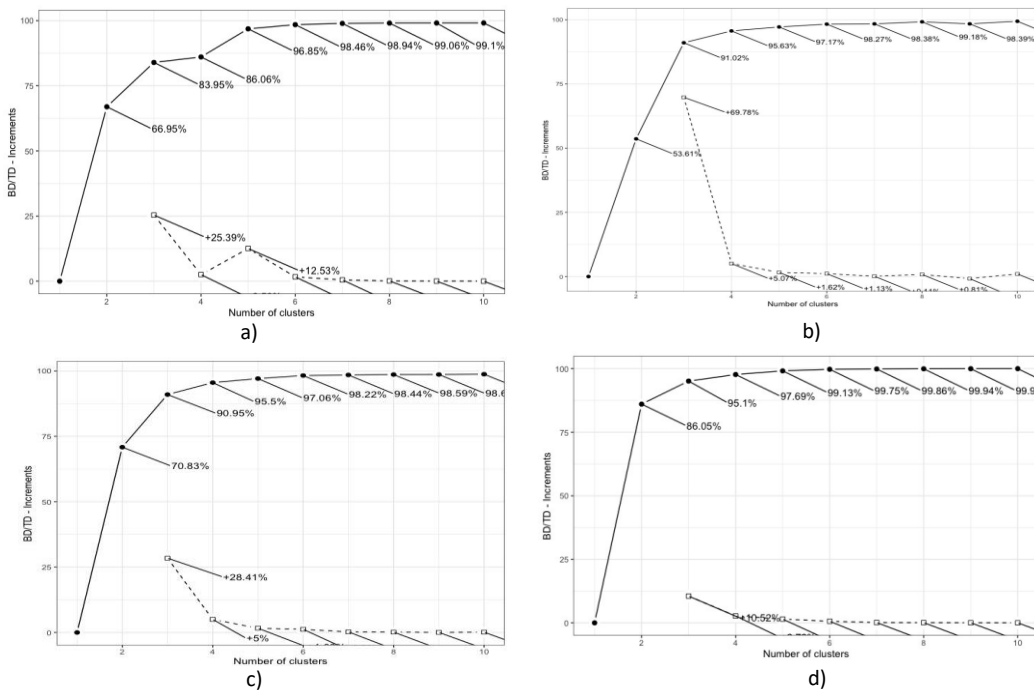


Figura 3. Numero de k cluster por: a) Municipio, b) Producto, c) Cliente y d) Mes.

4. RESULTADOS

Para revisar el crecimiento de las ventas, se procedió a revisar la tendencia de los ultimos siete años, siendo el 2022 solo una parcialidad dada la fecha en que se toman los datos los cuales son hasta el mes de julio 2022, tal como se muestra en la Figura 4. a) Tendencia de Ventas por Año donde se obtuvo el modelo de regresión lineal, como se muestra en la ecuación 1 el modelo resultante es muy preciso, ya que obtenemos un Coeficiente de

correlación, $r = 0.989149203$ y un Coeficiente de determinación, $R^2 = 0.978416146$, de la misma manera se procedió a realizar un dendograma para agrupar las ventas por estado tal como se muestra en la Figura 4. b) Dendograma por Estado, donde resultan cuatro grupos para el comportamiento de las ventas de esta manera se pueden realizar campañas específicas y de esta manera orientar de mejor manera los recursos de la empresa, por otro lado se procedió a representar el comportamiento de las ventas en el país según su densidad, tal como se muestra en la Figura 4. c) Mapa de Calor del Comportamiento de las Ventas en el País.

$$Ventas = -9.94e + 10 + (4.940e - 07 * Año) \tag{1}$$

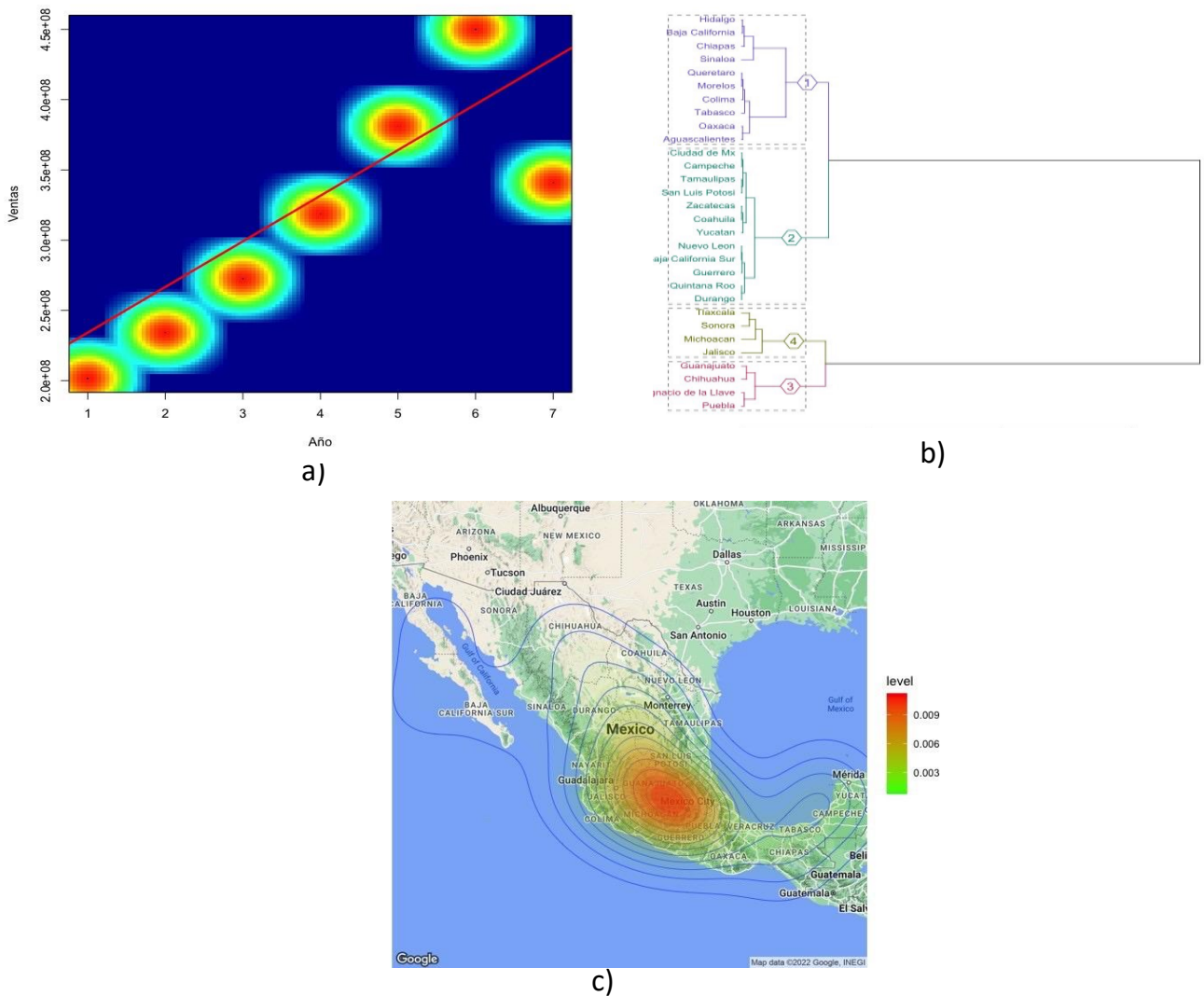


Figura 4. Numero de k cluster por: a) Tendencia de Ventas por Año, b) Dendograma por Estado y c) Mapa de Calor del Comportamiento de las Ventas en el País.

Finalmente se buscaron agrupar por medio del uso de dendogramas para cada uno de los conceptos, en el se agrupa el comportamiento de las ventas con respecto a su localización en un análisis más detallado ahora por municipio, dando como resultado el dendograma de ocho cluster, se muestra en la Figura 5. Dendograma de ventas: a) Municipio. En el caso de los productos, los datos obtenidos muestran por medio de la tecnica de k-means, que cinco clusters son los adecuados para agrupar el comportamiento de las ventas, tal como se muestra en la Figura 5. Dendograma de ventas: b) Producto. Así mismo el comportamiento de las ventas por

clientes, también fue agrupada en función a su comportamiento, este se puede observar en la Figura 5. Dendograma de ventas: c) Clientes, esta representa ocho clusters a considerar, finalmente se representan los resultados obtenidos del agrupamiento por mes, tal como se expresa en el dendograma de la clasificación donde se observa un comportamiento no esperado en el cluster 3, donde agrupa a los meses de Junio, Enero y Febrero los cuales sabemos que no están cercanos pero que esta información sirvió para generar estrategias en dichos meses en particular, tal como se muestra Figura 5. Dendograma de ventas: d) Mes.

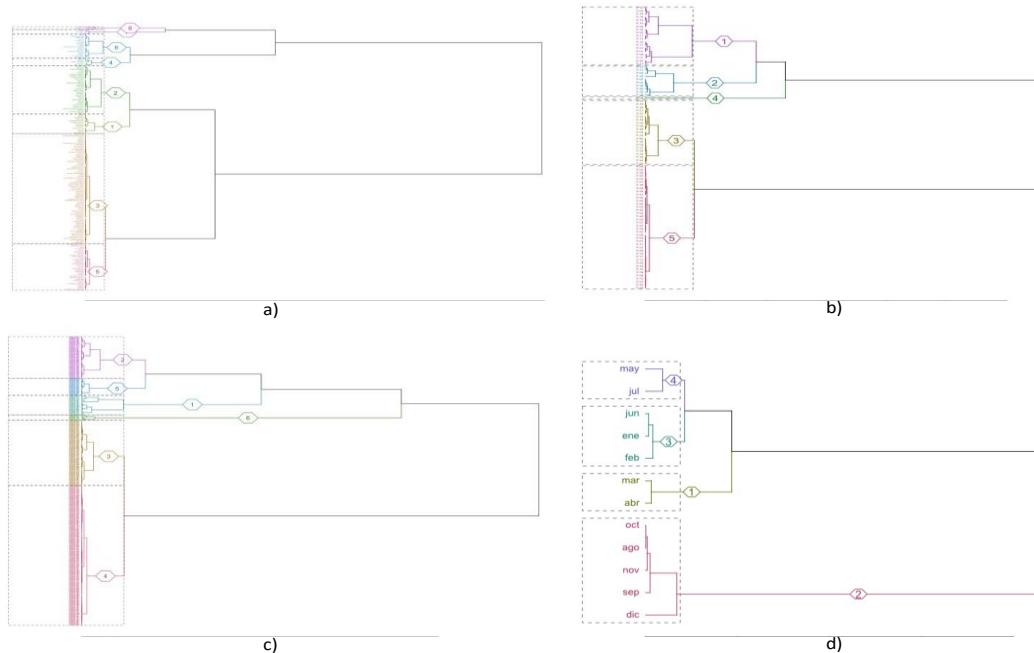


Figura 5. Dendograma de ventas: a) Municipio, b) Producto, c) Cliente y d) Mes.

5. CONCLUSIONES

El Análisis por medio de cluster en especial con herramientas como el k-means, permite generar cluster en base a criterios específicos de clasificación, los cuales ayudan a realizar mejores estrategias de ventas, como se mostro en la figura del volumen de ventas por Estado que mostro el dendograma, donde lo más conveniente resultado el hacer cuatro grupos segun el comportamiento de ventas anuales.

El Análisis arrojó que la estrategia de la empresa sera en enfocarse en 4 Estados (Sonora, Tlaxcala, Michoacan y Jalisco), que es donde la tendencia de las ventas anuales han tenido crecimiento y en determinados productos, dentro del mismo se encuentra el grupo de municipios que tienen mayores ventas, como es el caso de Hermosillo conformado por el cluster número 8 y donde menos ventas existen esta compuesta por 73 municipios de los 175 considerados en el estudio. Las ventas se concentran de manera muy importante en el cluster 6, los cuales están compuestos por los clientes: Cliente 073, Cliente 074, Cliente 303, Cliente 215, Cliente 180 y Cliente 016. Con los datos anteriores, la empresa generara estrategias claramente dirigidas en función de los tipos de productos que son aceptados, por determinados clientes, en determinados, Estados y municipios y considerando la temporalidad en términos de meses que más se demandan, con la finalidad de incrementar las ventas y utilizar de mejor.

REFERENCIAS

- [1] Laynez, M., Estudio del Big Data. Tesis de Licenciatura. Universidad de Sevilla, 2017.
- [2] Chan Canto, J. F., Morales González, M. A., Canto Esquivel, A. M. Panorama de la industria del calzado en el municipio de Ticul, Yucatán y su análisis desde la perspectiva de la competitividad sistémica. Repositorio De La Red Internacional De Investigadores En Competitividad, 13, 188–206, 2020.
- [3] Navarro Silva, O., Gracián Moran, G., Ferrer Reyes, W., La calidad del capital humano una vertiente positiva en la competitividad de las micros, pequeñas y medianas empresas enfocada en la empresa Cristy viajes. Universidad y Sociedad , 10 (2), 26 – 33, 2018.
- [4] Huyen. Phan. An Exploration of Big Data and Analytics Software. PhD thesis, University (PhD), 2020.
- [5] T. C. Loh and S. C.L. Koh. Critical elements for a successful enterprise resource planning implementation in small- And medium-sized enterprises. In International Journal of Production Research, volume 42, pages 3433–3455. Taylor Francis Group, sep 2004.
- [6] Hai Ying Xie. The research on knowledge management of small and medium-sized IT enterprises based on data mining. In 2009 International Conference on Information Management, Innovation Management and Industrial Engineering, ICIII 2009, volume 4, pages 127–130, 2009.
- [7] Ju Yeon Lee, Joo Seong Yoon, and Bo Hyun Kim. A big data analytics platform for smart factories in small and medium-sized manufacturing enterprises: An empirical case study of a die casting factory. International Journal of Precision Engineering and Manufacturing, 18(10):1353–1361, oct 2017.
- [8] Nihan Gulsoy and Sinem Kulluk. A data mining application in credit scoring processes of small and medium enterprises commercial corporate customers. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(3):e1299, may 2019.
- [9] Marina Nevskaya, Sergey Seleznev, Vladimir Másloboev, Elena Klyuchnikova, and Diminriy Makarov. Involving small and medium-sized mining industry businesses in mining waste processing in the Russian Federation. Mineral Economics, pages 1–6, apr 2020.
- [10] Siti Aishah Mohd Selamat, Simant Prakoonwit, and Wajid Khan. A review of data mining in knowledge management applications - findings for transportation of small and medium enterprises. SN Applied Sciences, 2(5):1–15, may 2020.
- [11] Olivares Vera, D. A., & Hernández Palafox, F., Ciencia de datos en empresas pequeñas y medianas (PYMES) en la zona Bajío. Tecnotrend, 4(13), 1–27. <http://tecnotrend.delasalle.edu.mx/uploads/ao6n11/daniel.pdf>, 2021.
- [12] Shirley Y. Coleman. Data-mining opportunities for small and medium enterprises with official statistics in the UK. Journal of Official Statistics, 32(4):849–865, dec 2016.
- [13] Arias Aybar, I. D., Cueva Angulo, R. B., Llanos Donayre, G. M., Pipa Ayala, E., Valdez Brizuela, U. A. Aplicación de ciencia de datos para identificar los segmentos de clientes de Grupo Deltron, jul 2021.
- [14] Pérez Paspuel, B. G., Análisis de datos aplicando las técnicas de Data Mining (Reglas de Asociación y Clustering) para fortalecer el comercio electrónico descubriendo hábitos de compra de productos y accesorios de bicicletas en la ciudad de Tulcán [Tesis de pregrado, Universidad Técnica del Norte], 2022.
- [15] Mompo Serrano, A., Usos de la Ciencia de Da- tos aplicados al sector agrícola (Doctoral dissertation, Universitat Politècnica de Valencia), 2022.
- [16] Carrera, R. A., Pronósticos del volumen de venta de mango mediante métodos univariados en series temporales / Mango sales volume forecast by using univariate methods in time series. Universidad ciencia, 9(3), 214-229, 2020.
- [17] León Naranjo, J. R., Análisis del mercado de productos de la cadena de agroindustrial de la miel de abejas en la provincia del Guayas (Bachelor's thesis, Facultad de Ciencias Económicas. Universidad de Guayaquil), 2021.
- [18] Herrera Rodríguez, J. M., Ojalora Gómez, L. M., López, Año Rueda, O. M., La innovación en marketing de las empresas agroindustriales de la Provincia de Sugamuxi. Tendencias, 21(2), 45-62., 2020.

Correo electrónico autor: danielarturo.olivares@leon.tecnm.mx