

Segmentación de Entrega por Volumen para Optimizar la distribución en una empresa de consumo masivo

Autores:

Carlo Paolo Jordan Cueva

Yannyk Quimis Rivera

ESAI Business School

Maestría en Inteligencia de Negocios y Ciencia de Datos

Junio 14, 2025

Palabras clave: Segmentación de datos, clusters, distribución logística, análisis de datos y consumo masivo.

Abstract

El presente proyecto tuvo como objetivo segmentar las entregas de productos por volumen para optimizar la distribución logística en una empresa de consumo masivo dedicada a la comercialización de bebidas, licores y abarrotes. El principal problema identificado fue la asignación ineficiente de rutas y sectores, lo que ocasiona entregas tardías, aumento de costos y pérdida de satisfacción del cliente.

A partir de una base de datos de clientes, se seleccionaron tres variables cuantitativas clave (volumen actual, volumen promedio de 3 meses y venta del mismo mes del año anterior), las cuales fueron estandarizadas para realizar un análisis estadístico robusto. Se aplicaron técnicas de segmentación como el análisis de clúster jerárquico y el algoritmo de K-medias, permitiendo identificar tres grupos naturales de clientes con comportamientos diferenciados de compra.

Los resultados permitieron clasificar los clientes en segmentos de alto, medio y bajo volumen de compra, lo cual facilitó el rediseño de rutas más eficientes, asignación de recursos logísticos y toma de decisiones estratégicas basadas en datos. La validez de la segmentación fue confirmada mediante análisis de varianza (ANOVA) y tablas cruzadas con variables como sector, categoría de cliente y frecuencia de visita.

Como resultado, se logró mejorar la planificación operativa, reduciendo la dispersión de entregas y justificando rutas de camiones con volúmenes balanceados por sector. Este proyecto demuestra el valor de la analítica avanzada para resolver desafíos logísticos mediante el uso de herramientas como SPSS y Power BI.

A. Resumen Ejecutivo

La empresa ¨Corpsa S.A ¨ es una organización que se dedica a la comercialización en la industria de consumo masivo de bebidas carbonatadas, aguas, licores y abarrotes. Esta es una de las empresas que cuenta con abastecimiento de cobertura de zonas con su fuerza de ventas, sin embargo, la mayor oportunidad de mejora está en alcanzar el 100% en efectividad de entrega, generando la errónea segmentación de sectores y asignación de rutas distribuidas por volumen de ventas.

Actualmente, según el Anexo 1, podemos observar el mapa geográfico de asignación de visitas a clientes de entrega de un solo camión. Teniendo más de 60 puntos por parar en un polígono bastante disperso, donde también hay puntos de entrega en Vía Salitre, Aurora y Mucho Lote 2. Este gráfico sirve para demostrar el problema que genera la insatisfacción del cliente al no recibir sus pedidos a tiempo, la disminución de prestigio de la empresa y el aumento de costo de transporte por la alta dispersión de puntos.

B. Resumen Ejecutivo Gráfico

Tabla 1.

Estadísticos Descriptivos de las variables estandarizadas

Estadísticos descriptivos

	N	Mínimo	Máximo	Media	Desv. estándar
Puntuación Z: VOLUMEN MES ACTUAL CAJAS	1471	-,50491	14,37899	,0000000	1,00000000
Puntuación Z: VENTA AÑO ANTERIOR (MISMO MES) CAJAS	1471	-,40807	20,39707	,0000000	1,00000000
Puntuación Z: VOLUMEN PROMEDIO 3 M CAJAS	1471	-,46959	16,60446	,0000000	1,00000000
N válido (por lista)	1471				

Tabla 2.

Estadísticos de los datos

Estadísticos

		SECTOR	FRECUENCIA DE VISITA	CATEGORIA CLIENTE
N	Válido	1471	1471	1471
	Perdidos	0	0	0

Según la tabla 2, nos indica que no hay ningún valor perdido y que todos los 1471 son válidos para su tratamiento.

C. Detalle Técnico

1. Fuente de Datos

- SECTOR: Representa la zona geográfica o área de cobertura.

  Tipo de variable: Categórica nominal

- SUPERVISOR: Indica el nombre o tipo de supervisor responsable.

  Tipo de variable: Categórica nominal

- VENDEDOR: Número o ID del vendedor asignado a cada cliente.

  Tipo de variable: Numérica discreta

- FRECUENCIA DE VISITA: Código que indica con qué frecuencia se visita al cliente (por ejemplo, *V*, *R*, *J*). Representa días específicos o patrones como "Viernes", "Jueves", etc.

  Tipo de variable: Categórica ordinal

- NOMBRE: Nombre del cliente o razón social.

  Tipo de variable: Categórica nominal

- CATEGORÍA CLIENTE: Clasificación comercial del cliente según su volumen o importancia.

  Tipo de variable: Categórica ordinal

- VOLUMEN MES ACTUAL CAJAS: Número de cajas vendidas al cliente en el mes actual.

  Tipo de variable: Cuantitativa continua

- VENTA AÑO ANTERIOR (MISMO MES) CAJAS: Número de cajas vendidas a ese cliente en el mismo mes del año anterior.

  Tipo de variable: Cuantitativa continua

- VOLUMEN PROMEDIO 3 M CAJAS: Promedio de cajas vendidas al cliente en los últimos tres meses.

  Tipo de variable: Cuantitativa continua

  La tabla busca hacer un análisis comparativo de ventas por cliente, con foco en:

- Evolución del volumen vendido (mes actual vs. año anterior).
- Tendencias de compra por cliente (volumen promedio de 3 meses).
- Gestión comercial (frecuencia de visita, categoría del cliente).
- Apoyo para la toma de decisiones comerciales y de planificación de visitas o promociones.

Por último para todo este desarrollo se seleccionaron las tres variables cuantitativas que son Volumen mes actual en cajas, Volumen promedio 3 M Cajas y Venta año anterior (mismo mes) cajas porque representan de manera directa el comportamiento de compra del cliente tanto en el presente como en el pasado reciente.

2. Técnica

2.1 Análisis Estadístico

El análisis estadístico nos permite recopilar, describir y analizar datos con el fin de obtener conclusiones válidas para la toma de decisiones (Perez Juste, 2012, #). En este caso, nos servirá para observar y resumir los datos históricos de las ventas para poder entender el comportamiento de los clientes en términos de volumen, frecuencia de compra, ubicación, etc; ya que nos ayuda a detectar variables claves y que serán usadas en su posterior segmentación. Esto nos aporta información clara del comportamiento logístico actual de la empresa Comecarba.

2.2 Estandarización de datos

La estandarización de variables es un paso importante y mandatorio en técnicas de análisis de clusters ya que nos ayuda a eliminar el efecto del rango o la escala de medida, lo que nos garantiza que cada variable tenga el mismo peso en el cálculo de distancias. (Pardo & Pardo Merino, 2002, #).

Con lo antes mencionado, para poder segmentar de manera correcta y realizar los clusters jerárquicos y K-medias, es necesario primero estandarizar las tres variables cuantitativas: volumen del mes actual, volumen promedio de 3 meses y ventas del mismo

mes del año anterior. Este proceso se realizó utilizando la transformación Z la cual convierte los valores originales en una escala común con media 0 y desviación estándar 1.

Este paso es importante ya que evitará que una variable dominará el análisis debido a su escala numérica y garantizará que todas las variables tengan el mismo peso que permite una segmentación más justa y relevante del comportamiento de los clientes.

2.3 Cluster jerárquicos

Es un método de agrupamiento que construye una jerarquía de grupos sin que sea necesario especificar previamente cuántos clusters se desean. Además que se puede utilizar el método de Enlace de Ward con Distancia Euclidiana cuadrada que nos ayuda a minimizar la varianza dentro de cada grupo, generando clusters compactos y de tamaño similar. (Universidad de Granada, n.d.)

Además, esta técnica de clusters sirve para identificar y agrupar datos para luego segmentar los puntos de entrega en grupos. Adicionalmente, el dendrograma nos permite ver la cantidad de clusters naturales puede realizarse en nuestro trabajo.

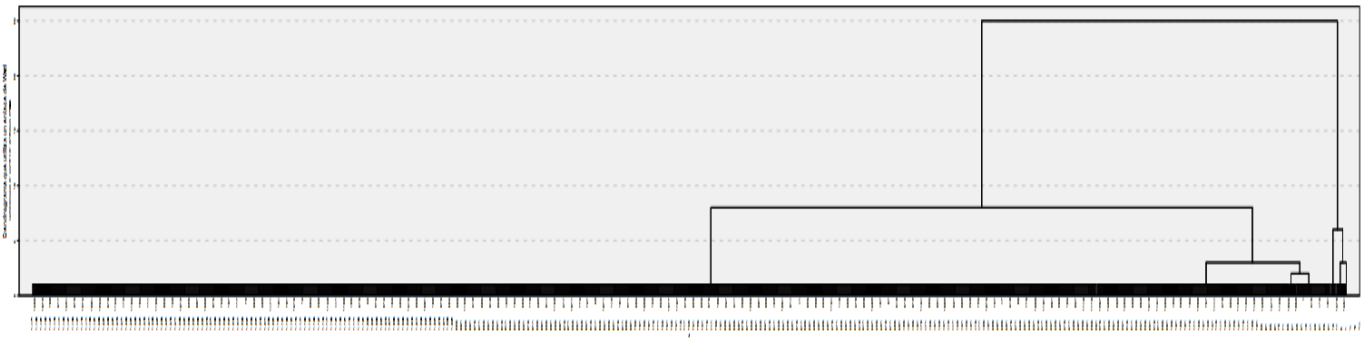
2.4 Cluster K-Medias (K-means)

Por último, la técnica de K-Medias nos permite dividir un conjunto de datos en grupos de tal forma que los elementos de cada grupo sean similares entre sí y diferentes a los otros grupos. Además nos permite a nosotros poder definir el número de grupos (clusters) que queramos realizar. (Peña, 2002, 228)

D. Resultados

Figura 1.

Dendrograma que utiliza el Enlace de Ward



Este gráfico nos muestra 3 cortes principales, es decir, 3 clusters naturales ya que permite visualizar cómo los clientes se agrupan progresivamente y estima el número óptimo de clusters, que a su vez son tres grandes grupos de clientes. Se eligió el método de Ward (revisar Anexo 2) por su capacidad para minimizar la varianza interna, lo cual produce clusters homogéneos y consistentes. Esto favorece decisiones más consistentes en la planificación de entregas del producto.

Adicionalmente, este cluster ayudó a descubrir que los clientes no son uniformes ya que hay grupos bien definidos que comparten niveles similares de compra, lo cual justifica segmentarlos para mejorar la eficiencia de cobertura y entregas de producto.

Tabla 3.

Centros de clusters iniciales de K-Media

Centros de clústeres iniciales

	Clúster		
	1	2	3
Puntuación Z: VOLUMEN MES ACTUAL CAJAS	3,88647	-,50491	11,73337
Puntuación Z: VENTA AÑO ANTERIOR (MISMO MES) CAJAS	20,39707	-,40807	-,40807
Puntuación Z: VOLUMEN PROMEDIO 3 M CAJAS	8,54283	-,46959	16,60446

La tabla 3 nos permitió clasificar a los clientes en tres grupos bien diferenciados: uno con alto volumen mensual y promedio alto, otro con comportamiento intermedio y frecuencia variable y el tercero con bajo volumen y poco historial. Está segmentación servirá para asignar rutas y recursos más eficientes.

Tabla 4.

Números de casos por cluster

Número de casos en cada clúster

Clúster	1	21,000
	2	1444,000
	3	6,000
Válidos		1471,000
Perdidos		,000

La tabla 4 nos enseña las características claras por cada grupo de cluster, donde el cluster 1 posee un volumen bajo de ventas pero con un comportamiento de menor quiere decir que son clientes pasivos o de baja rotación. El cluster 2 posee un volumen alto de ventas con clientes de potencial crecimiento y son estratégicos para la empresa.

Por último, el cluster 3 tiene un volumen intermedio de ventas, posee clientes estables. Puede optimizarse en rutas combinadas con otros grupos.

De acuerdo a la tabla ANOVA del Anexo 3, nos muestra que todas las variables utilizadas tienen diferencias significativas entre los grupos con valores de significancia <0.05, por lo tanto se rechaza la hipótesis nula para cada variable. Esto significa que los promedios de volumen son considerablemente diferentes entre los grupos, por ende, estas variables explican correctamente la segmentación realizada mediante clusters.

Tabla 6.

Tabla Cruzada de Cluster con Categoría Cliente

Tabla cruzada Número de caso de clúster*CATEGORIA CLIENTE

			CATEGORIA CLIENTE						
			0	2 PLUS	3 PLUS	4 PLUS	5 PLUS	MAX	Total
Número de caso de clúster	1	Recuento	0	0	0	0	3	18	21
		% dentro de Número de caso de clúster	0,0%	0,0%	0,0%	0,0%	14,3%	85,7%	100,0%
	2	Recuento	7	140	233	454	296	314	1444
		% dentro de Número de caso de clúster	0,5%	9,7%	16,1%	31,4%	20,5%	21,7%	100,0%
	3	Recuento	0	0	0	0	2	4	6
		% dentro de Número de caso de clúster	0,0%	0,0%	0,0%	0,0%	33,3%	66,7%	100,0%
Total		Recuento	7	140	233	454	301	336	1471
		% dentro de Número de caso de clúster	0,5%	9,5%	15,8%	30,9%	20,5%	22,8%	100,0%

La tabla 6 permite identificar qué categorías de cliente predominan en cada cluster. Se puede observar una asociación entre los clusters y el nivel comercial de los clientes, lo cual valida la segmentación realizada y nos permite planificar rutas diferenciadas según su valor estratégico.

Tabla 7.

Tabla Cruzada de Cluster con Sector

Tabla cruzada Número de caso de clúster*SECTOR

			SECTOR						
			5 DICIEMBRE - PENITENCIARIA	AURORA	BASTION BLOQUE 1	BASTION BLOQUE 10	BASTION BLOQUE 3	BASTION BLOQUE 5	BASTION BLOQUE 7
Número de caso de clúster	1	Recuento	0	2	0	1	0	1	0
		% dentro de Número de caso de clúster	0,0%	9,5%	0,0%	4,8%	0,0%	4,8%	0,0%
	2	Recuento	25	31	78	12	21	103	23
		% dentro de Número de caso de clúster	1,7%	2,1%	5,4%	0,8%	1,5%	7,1%	1,6%
	3	Recuento	1	0	0	0	0	0	1
		% dentro de Número de caso de clúster	16,7%	0,0%	0,0%	0,0%	0,0%	0,0%	16,7%
Total		Recuento	26	33	78	13	21	104	24
		% dentro de Número de caso de clúster	1,8%	2,2%	5,3%	0,9%	1,4%	7,1%	1,6%

BASTION BLOQUE 9	LA JOYA	LAS LOJAS	MUCHO LOTE 1	MUCHO LOTE 2	ORQUIDEAS	PASCUALES	PUENTE LUCIA	SAMBORONDON PUEBLO	SAN FRANCISCO	TARIFA
0	2	0	1	1	2	1	1	4	2	1
0,0%	9,5%	0,0%	4,8%	4,8%	9,5%	4,8%	4,8%	19,0%	9,5%	4,8%
26	83	32	127	72	103	115	82	105	125	139
1,8%	5,7%	2,2%	8,8%	5,0%	7,1%	8,0%	5,7%	7,3%	8,7%	9,6%
0	0	0	0	0	0	0	0	2	0	0
0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	33,3%	0,0%	0,0%
26	85	32	128	73	105	116	83	111	127	140
1,8%	5,8%	2,2%	8,7%	5,0%	7,1%	7,9%	5,6%	7,5%	8,6%	9,5%

VIA SALITRE	VILLA BONITA	VILLA CLUB	VILLAS DEL REY	Total
0	1	0	1	21
0,0%	4,8%	0,0%	4,8%	100,0%
31	35	24	52	1444
2,1%	2,4%	1,7%	3,6%	100,0%
1	0	0	1	6
16,7%	0,0%	0,0%	16,7%	100,0%
32	36	24	54	1471
2,2%	2,4%	1,6%	3,7%	100,0%

En la tabla 7, se puede visualizar cómo se distribuyen los sectores geográficos dentro de cada cluster y esto aporta valor para la planificación logística, ya que nos ayuda a identificar zonas con concentración de clientes similares y así optimizar la asignación de rutas de reparto y cantidad de camiones por ruta. Esta tabla junto con la 6 se puede

confirmar que son las más influyentes ya que se puede visualizar de mejor manera la segmentación de los clusters y el comportamiento de los clientes para mejorar la ruta de entrega del producto.

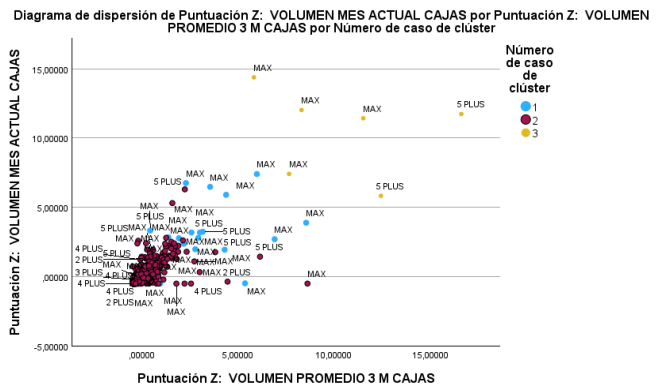
El anexo 4 nos permite analizar la relación entre la frecuencia de visita a los clientes con los clusters donde se evidencia que los clientes con mayor volumen, pertenecen al cluster 2 y nos dice que el modelo de atención actual tiene una respuesta parcialmente al comportamiento de compra.

Además, el anexo 5 nos ayuda a identificar qué supervisores tienen un mayor número de clientes en cada cluster. Esta data nos ayuda a evaluar la carga operativa por supervisor y que puedan tomar decisiones de distribución si es necesario.

Por último, el anexo 6 nos enseña la distribución de clusters entre vendedores, lo que permite conocer que vendedores atienden clientes de alto, medio y bajo volumen, esto nos ayuda para poder definir incentivos adicionales para mejorar las ventas de los volúmenes bajos.

Figura 2.

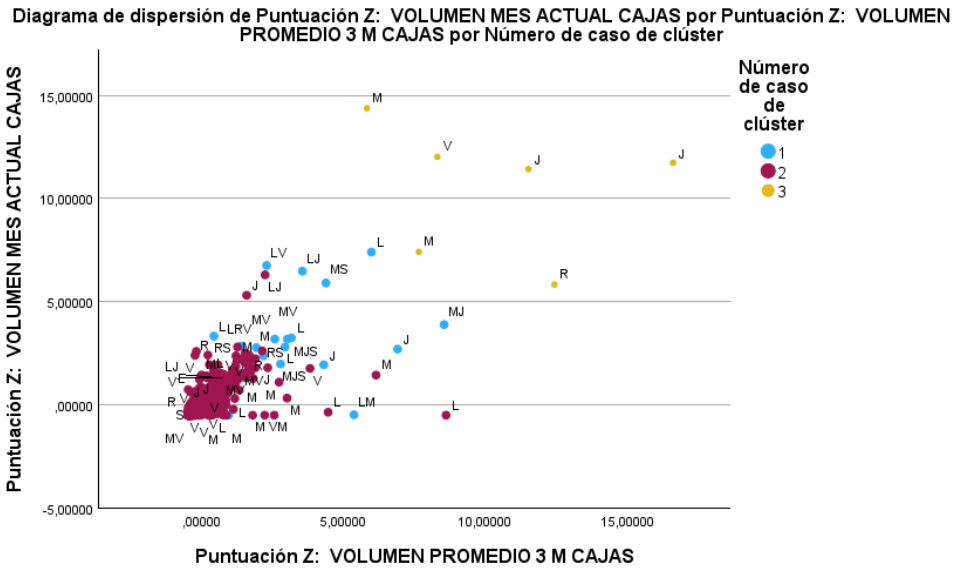
Scatterplot de Puntuación Z Volumen Mes Actual por Puntuación Z Volumen Promedio 3 M Cajas por Número de caso de cluster con etiqueta Categoría Cliente



El scatterplot de la figura 2 muestra cómo se agrupan los clientes según sus valores estandarizados de volumen actual y promedio. Se puede observar que los clientes de categoría alta como MAX y 3 PLUS tienden a concentrarse en los clusters de mayor volumen, validando así la información antes presentada en tablas anteriores.

Figura 3.

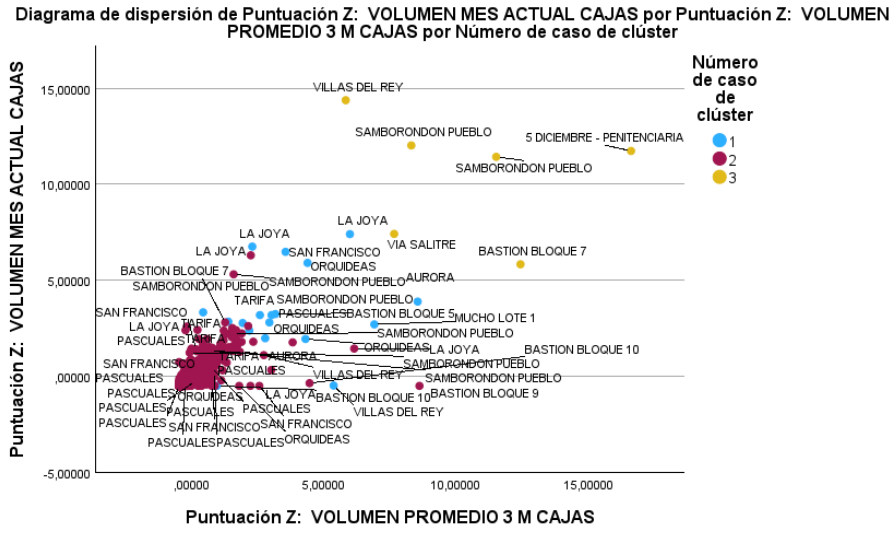
Scatterplot de Puntuación Z Volumen mes actual por Puntuación Z Volumen Promedio 3 M Cajas por Número de caso de cluster con etiqueta de Frecuencia de Visita



Con la figura 3, se observa que las visitas más frecuentes se relacionan con los clientes de clusters de mayor volumen, esto refuerza la idea de que la frecuencia de visita podría usarse como variable adicional para optimizar rutas logísticas.

Figura 4.

Scatterplot de Puntuación Z Volumen mes actual por Puntuación Z Volumen Promedio 3 M Cajas por Número de caso de cluster con etiqueta de Sector



Finalmente, con la figura 4 podemos identificar que si existe concentración geográfica por cluster ya que ciertos sectores aparecen agrupados dentro de un cluster, se puede aprovechar esa distribución para reestructurar las rutas de entrega y volverlas más eficientes por zona, pudiendo reducir las entregas tardías a los clientes.

Ante todos estos resultados, podemos concluir que a lo largo de este proyecto se logró clasificar a los clientes de la empresa mediante un proceso de segmentación basado en análisis de datos con la aplicación de técnicas estadísticas como lo son los clusters jerárquico y K-medias donde se identificó tres grupos diferentes de clientes en función a su historial de compra mensual, promedio histórico y registros del año anterior.

Este tipo de agrupamiento permite entender mejor el perfil de cada segmento y tomar decisiones eficientes donde se identificó patrones consistentes que pueden ser aprovechados

para volver a diseñar las rutas de entrega, distribuir mejor los recursos de ventas y mejorar la eficiencia operativa logística en general.

Teniendo como resultados, asignaciones de camiones sustentado con un volumen de 250 a 300 cajas por cada uno, designado a estar fijo dentro de un solo sector. Tal es el caso mostrado en el anexo 7, que el Coordinador de entrega segmentó el polígono solo con Via Salitre, por lo que él mismo justifica 279 cajas. Siendo el contraste en mejora de segmentación de sectores que se mostró en el anexo 1, el cual un solo camión contaba con tres sectores de entrega.

Todo es logrado en base a los análisis estadísticos realizados por la herramienta SPSS y la data geográfica por sector utilizada para apoyo de segmentación con Power Bi mostrado en el anexo 8.

E. Referencias Bibliográficas

Pardo, A., & Pardo Merino, A. (2002). *Spss 11 Guia Para el Analisis de Datos*. McGraw-Hill/Interamericana de Espana, S.A.

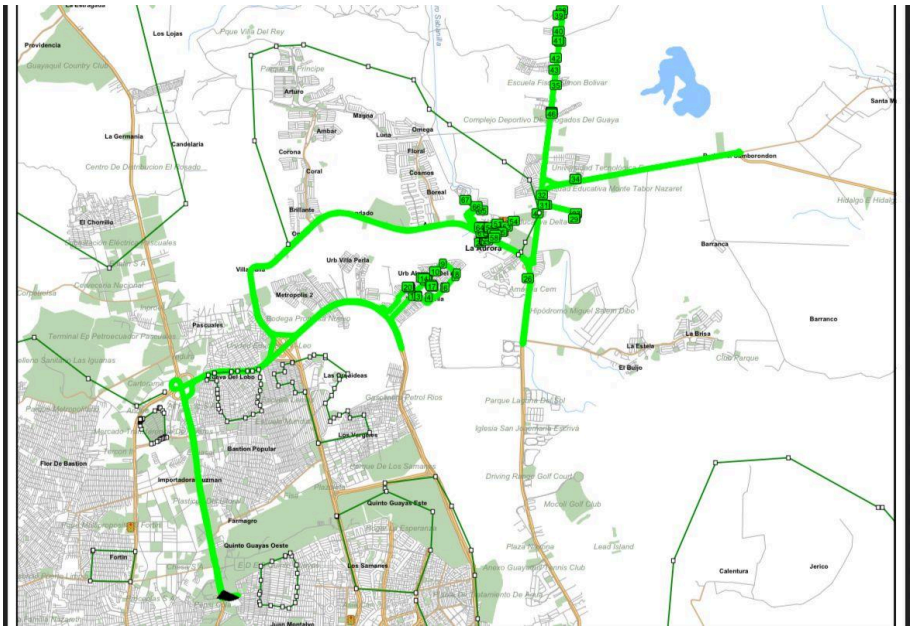
Peña, D. (2002). *Análisis de datos multivariantes*. McGraw-Hill.

Perez Juste, R. (2012). *Curso 0 de Estadistica Aplicada*. Universidad Nacional de Educación a Distancia (UNED). http://creativecommons.org/licenses/by-nc-nd/3.0/es/

Universidad de Granada. (n.d.). *Práctica 8 | Estadística*. Universidad de Granada. Retrieved May 31, 2025, from https://wpd.ugr.es/~bioestad/guia-spss/practica-8/

F. Anexos

Anexo 1. Mapa volumétrico de ruta de entrega de clientes



Anexo 2. Resumen de procesamiento de casos de cluster jerárquico.

Resumen de procesamiento de casos[a,b]

Casos					
Válido		Perdidos		Total	
N	Porcentaje	N	Porcentaje	N	Porcentaje
1471	100,0	0	,0	1471	100,0

a. Distancia euclídea al cuadrado utilizada

b. Enlace de Ward

Anexo 3. *Tabla de ANOVA*

ANOVA

	Clúster		Error			
	Media cuadrática	gl	Media cuadrática	gl	F	Sig.
Puntuación Z: VOLUMEN MES ACTUAL CAJAS	428,931	2	,417	1468	1028,642	<,001
Puntuación Z: VENTA AÑO ANTERIOR (MISMO MES) CAJAS	402,680	2	,453	1468	889,407	<,001
Puntuación Z: VOLUMEN PROMEDIO 3 M CAJAS	431,550	2	,413	1468	1043,855	<,001

Las pruebas F sólo se deben utilizar con fines descriptivos porque los clústeres se han elegido para maximizar las diferencias entre los casos de distintos clústeres. Los niveles de significación observados no están corregidos para esto y, por lo tanto, no se pueden interpretar como pruebas de la hipótesis de que los medias de clúster son iguales.

Anexo 4. Tabla Cruzada de Cluster con Frecuencia de Visita

Tabla cruzada Número de caso de clúster*FRECUENCIA DE VISITA

			FRECUENCIA DE VISITA															
			J	JS	L	LJ	LM	LRV	LV	M	MJ	MJS	MS	MV	R	RS	S	V	Total
Número de caso de clúster	1	Recuento	2	0	4	1	1	1	1	2	1	2	1	2	0	1	0	2	21
		% dentro de Número de caso de clúster	9,5%	0,0%	19,0%	4,8%	4,8%	4,8%	4,8%	9,5%	4,8%	9,5%	4,8%	9,5%	0,0%	4,8%	0,0%	9,5%	100,0%
	2	Recuento	214	3	185	100	0	9	6	240	3	6	6	79	188	72	139	194	1444
		% dentro de Número de caso de clúster	14,8%	0,2%	12,8%	6,9%	0,0%	0,6%	0,4%	16,6%	0,2%	0,4%	0,4%	5,5%	13,0%	5,0%	9,6%	13,4%	100,0%
	3	Recuento	2	0	0	0	0	0	0	2	0	0	0	0	1	0	0	1	6
		% dentro de Número de caso de clúster	33,3%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	33,3%	0,0%	0,0%	0,0%	0,0%	16,7%	0,0%	0,0%	16,7%	100,0%
Total		Recuento	218	3	189	101	1	10	7	244	4	8	7	81	189	73	139	197	1471
		% dentro de Número de caso de clúster	14,8%	0,2%	12,8%	6,9%	0,1%	0,7%	0,5%	16,6%	0,3%	0,5%	0,5%	5,5%	12,8%	5,0%	9,4%	13,4%	100,0%

Anexo 5. *Tabla Cruzada de Cluster con Supervisor*

Tabla cruzada Número de caso de clúster*SUPERVISOR

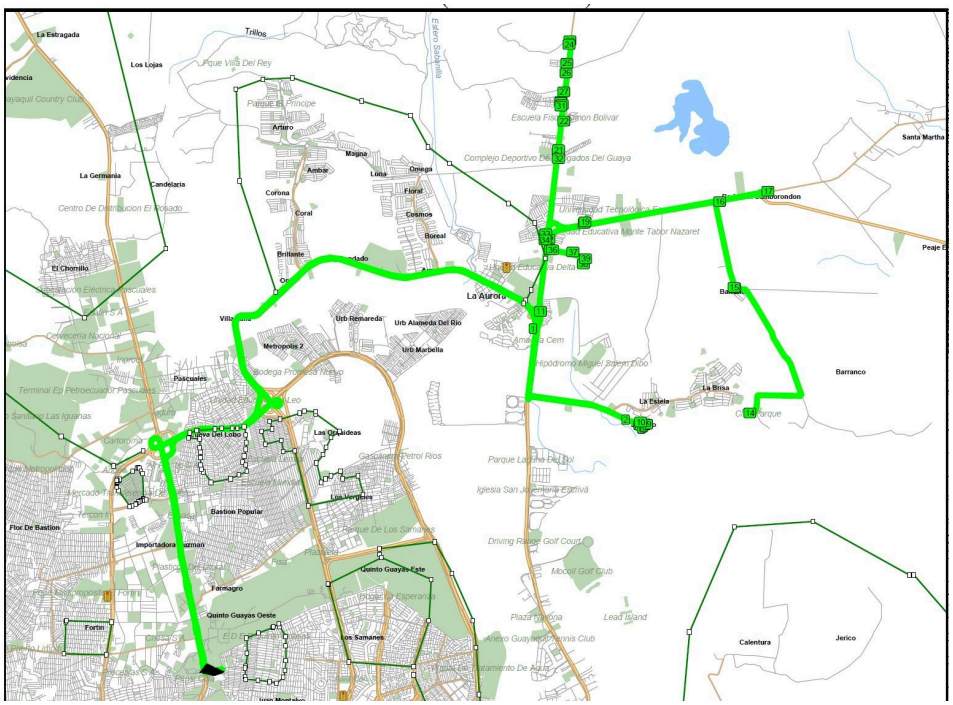
			SUPERVISOR FORANEO-METRO	Total
Número de caso de clúster	1	Recuento	21	21
		% dentro de Número de caso de clúster	100,0%	100,0%
	2	Recuento	1444	1444
		% dentro de Número de caso de clúster	100,0%	100,0%
	3	Recuento	6	6
		% dentro de Número de caso de clúster	100,0%	100,0%
Total		Recuento	1471	1471
		% dentro de Número de caso de clúster	100,0%	100,0%

Anexo 6. *Tabla Cruzada de Cluster con Vendedor*

Tabla cruzada Número de caso de clúster*VENDEDOR

			VENDEDOR								
			1	2	3	4	5	6	7	8	Total
Número de caso de clúster	1	Recuento	0	2	2	1	0	1	7	8	21
		% dentro de Número de caso de clúster	0,0%	9,5%	9,5%	4,8%	0,0%	4,8%	33,3%	38,1%	100,0%
	2	Recuento	264	255	185	243	226	154	57	60	1444
		% dentro de Número de caso de clúster	18,3%	17,7%	12,8%	16,8%	15,7%	10,7%	3,9%	4,2%	100,0%
	3	Recuento	0	0	2	1	2	0	1	0	6
		% dentro de Número de caso de clúster	0,0%	0,0%	33,3%	16,7%	33,3%	0,0%	16,7%	0,0%	100,0%
Total		Recuento	264	257	189	245	228	155	65	68	1471
		% dentro de Número de caso de clúster	17,9%	17,5%	12,8%	16,7%	15,5%	10,5%	4,4%	4,6%	100,0%

Anexo 7. Mapa volumétrico de ruta de entrega clusterizado



Anexo 8. Dashboard Geográfico de Volumen de Cajas

