REPORTE TÉCNICO COMPLETO

Análisis de Segmentación de Clientes - MegaMart

---

1. INFORMACIÓN DEL PROYECTO

1.1 Equipo de Trabajo

Nombre	ID Estudiante
César Isao Pastelin Kohagura	A01659947
Luis Emilio Fernández González	A01659517
Eduardo Botello Casey	A01659281

Equipo: Team 3 **Fecha de Entrega:** 07/11/2025 **Institución:** Tecnológico de Monterrey **Curso:** Aplicación de Métodos Multivariados en Ciencia de Datos

1.2 Enlaces y Recursos

- **Video de Presentación:** https://youtu.be/1ee-Rb251Y4
- **Repositorio:** Luis_Emilio-portfolio-ma2003b/case-03-cluster-analysis

---

2. RESUMEN EJECUTIVO

2.1 Contexto del Negocio

MegaMart, empresa retail de gran escala, enfrenta desafíos estratégicos relacionados con campañas de marketing genéricas que no consideran la diversidad del comportamiento de sus clientes. Esto resulta en:

- Baja tasa de engagement en campañas
- Asignación ineficiente de recursos de marketing
- Bajo rendimiento en métricas clave: CLV (Customer Lifetime Value), MROI (Marketing ROI), y Churn
- Incapacidad para detectar clientes de alto riesgo tempranamente

2.2 Solución Propuesta

Implementación de análisis de segmentación basado en datos utilizando técnicas de clustering no supervisado para identificar perfiles de clientes distintivos y traducirlos en estrategias de negocio accionables.

2.3 Impacto Esperado

Cluster	Resultado Esperado
Cluster 0	Mantener baja rotación, alta frecuencia, fuerte tamaño de cesta
Cluster 1	40% reducción en tasa de retorno, 50% aumento en transacciones
Cluster 2	40% aumento en visitas mensuales, 30% crecimiento en compras
Cluster 3	30% aumento en tamaño de cesta, 30% mejora en engagement

---

3. DATASET Y PREPARACIÓN DE DATOS

3.1 Descripción del Dataset

Archivo Principal: `retail_customer_data-1.csv`

Atributo	Valor
Total de Clientes	3,000
Variables Analizadas	9 (+ 1 identificador removido)
Valores Faltantes	0 (dataset completo)
Formato	CSV
Tamaño	~155 KB

3.2 Variables del Dataset

Variable	Descripción	Tipo	Rango
`customer_id`	Identificador único del cliente	string	CUST_0001 - CUST_3000
`monthly_transactions`	Número de transacciones por mes	float64	0.2 - 22.3
`avg_basket_size`	Promedio de ítems por compra	float64	1.0 - 31.6
`total_spend`	Gasto total del cliente	float64	$50 - $8,746
`avg_session_duration`	Duración promedio de sesión de navegación	float64	3.4 - 87.3 min
`email_open_rate`	Tasa de apertura de emails de marketing	float64	0.0 - 0.95
`product_views_per_visit`	Productos vistos por sesión	float64	3.0 - 62.4
`return_rate`	Porcentaje de retorno de productos	float64	0.0 - 0.50
`customer_tenure_months`	Duración como cliente (meses)	int64	1 - 59
`recency_days`	Días desde última compra	int64	1 - 67

3.3 Estadísticas Descriptivas

```
Estadísticas Clave:
- monthly_transactions: μ=6.01, σ=4.78
- avg_basket_size: μ=9.49, σ=7.96
- total_spend: μ=$2,367.38, σ=$2,248.27
- avg_session_duration: μ=38.44, σ=14.33
- email_open_rate: μ=0.44, σ=0.43
- product_views_per_visit: μ=31.36, σ=9.96
- return_rate: μ=0.19, σ=0.16
- customer_tenure_months: μ=17.91, σ=11.24
- recency_days: μ=20.68, σ=12.06
```

3.4 Análisis de Correlaciones

Correlaciones Más Fuertes (Top 5)

Par de Variables	Correlación	Interpretación
`avg_basket_size` ↔ `total_spend`	**0.941**	Altamente positiva - cestas grandes implican mayor gasto
`monthly_transactions` ↔ `total_spend`	**0.764**	Fuerte positiva - más transacciones = más gasto
`monthly_transactions` ↔ `avg_basket_size`	**0.691**	Fuerte positiva - clientes frecuentes compran más por visita

Par de Variables	Correlación	Interpretación
`monthly_transactions` ↔ `recency_days`	-0.632	Moderada negativa - clientes frecuentes tienen baja recencia
`total_spend` ↔ `recency_days`	-0.612	Moderada negativa - gastadores altos compran recientemente

Insight Principal: Las variables de comportamiento de compra (transacciones, tamaño de cesta, gasto) están fuertemente correlacionadas entre sí, mientras que la recencia muestra relación negativa con actividad de compra.

3.5 Preprocesamiento de Datos

3.5.1 Limpieza de Datos

1. **Remoción de `customer_id`:** Variable no informativa para clustering
2. **Validación de valores faltantes:** No se requirió imputación (0 valores faltantes)
3. **Detección de outliers:** Identificados mediante boxplots, mantenidos para análisis

3.5.2 Normalización

Método Utilizado: StandardScaler (sklearn)

```
scaler = StandardScaler()
X_standardized = scaler.fit_transform(df)
```

Justificación:

- Las variables tienen escalas muy diferentes (ej. `return_rate`: 0-0.5 vs `total_spend`: 50-8746)
- Sin normalización, variables de gran escala dominarían el cálculo de distancias
- StandardScaler transforma datos a media=0 y desviación estándar=1

Validación Post-Normalización:

```
Todas las variables normalizadas:
- Media ≈ 0 (±1e-16)
- Desviación estándar ≈ 1.00
- Rango estandarizado: ~[-2.4, +3.4]
```

---

4. METODOLOGÍA DE CLUSTERING

4.1 Técnicas Implementadas

Se utilizaron dos enfoques complementarios:

4.1.1 Clustering Jerárquico (Hierarchical Clustering)

- **Algoritmo:** Aglomerativo (bottom-up)
- **Métodos de Enlace Evaluados:**
  - Single Linkage (vecino más cercano)
  - Complete Linkage (vecino más lejano)
  - Average Linkage (distancia promedio)
  - **Ward's Linkage** ✓ (seleccionado - minimiza varianza intra-cluster)

Selección de Ward:

- Genera clusters compactos y bien definidos
- Minimiza la varianza dentro de cada cluster
- Dendrograma muestra separaciones claras
- Menos susceptible al efecto de encadenamiento

Extracción de Clusters:

```
linkage_matrix = linkage(X_standardized, method='ward')
labels = fcluster(linkage_matrix, n_clusters=4, criterion='maxclust')
```

4.1.2 K-Means Clustering

- **Algoritmo:** Particionamiento iterativo
- **Rango de K evaluado:** 2-10 clusters
- **Parámetros:**
  - `n_init=10` (10 inicializaciones diferentes)
  - `random_state=42` (reproducibilidad)

```
kmeans = KMeans(n_clusters=k, random_state=42, n_init=10)
kmeans.fit(X_standardized)
```

4.2 Determinación del Número Óptimo de Clusters

4.2.1 Método del Codo (Elbow Method)

Análisis de Inercia (WCSS - Within-Cluster Sum of Squares):

K	Inercia	Reducción vs K-1
2	~18,500	-
3	~15,200	17.8%
4	**~13,100**	**13.8%** ← Punto de inflexión
5	~11,500	12.2%
6	~10,200	11.3%

Observación: El "codo" se observa en k=4, donde la reducción marginal de inercia comienza a disminuir.

4.2.2 Análisis de Silueta (Silhouette Analysis)

Scores por Número de Clusters:

K	Silhouette Score	Interpretación	Min Size	Max Size
2	**0.344**	Mejor separación estadística	540	2,460
3	0.295	Buena estructura	540	1,501
4	**0.316**	**Equilibrio óptimo** ✓	420	1,081
5	0.300	Estructura moderada	420	1,081
6	0.248	Separación débil	373	708

Escala de Interpretación:

- 0.71 - 1.0: Estructura fuerte
- 0.51 - 0.70: Estructura razonable
- 0.26 - 0.50: Estructura débil pero presente
- < 0.25: Sin estructura sustancial

4.2.3 Análisis de Dendrograma

- Identificación de gaps verticales grandes en altura de fusión
- Corte óptimo sugerido: ~60 unidades de distancia
- Resultado: 4 ramas principales bien diferenciadas

4.2.4 Decisión Final: K = 4

Justificación Multi-criterio:

1. **Validación Estadística:**
   - Segundo mejor silhouette score (0.316)
   - Balance entre estructura (alta para k=2) y granularidad (baja para k>5)
   - Punto de inflexión claro en método del codo
2. **Interpretabilidad de Negocio:**
   - k=2: Demasiado general, pierde subgrupos importantes
   - k=4: Identifica segmentos accionables con características distintivas
   - k>5: Fragmentación excesiva, dificulta estrategias diferenciadas
3. **Tamaños de Clusters Balanceados:**
   - Cluster más pequeño: 420 clientes (14.0%)
   - Cluster más grande: 1,113 clientes (37.1%)

- Distribución manejable para operaciones de marketing

---

5. RESULTADOS DEL CLUSTERING

5.1 Comparación de Métodos

Métrica	K-Means	Jerárquico (Ward)
Silhouette Score	0.317	0.316
Tiempo de Cómputo	Rápido	Moderado
Estabilidad	Depende de inicialización	Determinístico
Interpretabilidad	Alta (centroides)	Alta (dendrograma)

Conclusión: Ambos métodos convergen a soluciones muy similares (silhouette scores casi idénticos), validando la robustez de la estructura identificada.

5.2 Distribución de Clusters (K-Means - Solución Final)

Cluster ID	N° Clientes	Porcentaje	Nombre Asignado
0	525	17.5%	High-Value Champions
1	929	31.0%	Window Shoppers
2	433	14.4%	Premium Occasional Buyers
3	1,113	37.1%	Low-Engagement Mass Segment

5.3 Perfiles Detallados de Clusters

CLUSTER 0: High-Value Champions

Tamaño: 525 clientes (17.5%)

Métrica	Valor	vs. Promedio General	Ranking
`monthly_transactions`	14.07	**+134.1%**	#1
`avg_basket_size`	22.03	**+132.2%**	#1
`total_spend`	$6,507	**+174.9%**	#1
`recency_days`	8.02	**-61.2%**	#1
`return_rate`	0.10	**-46.7%**	#1
`customer_tenure_months`	26.2	**+46.4%**	#1
`email_open_rate`	0.46	+3.6%	#2
`avg_session_duration`	38.8	+1.0%	#3
`product_views_per_visit`	40.2	+28.2%	#2

Características Distintivas:

- Clientes más valiosos del portfolio (top 17.5%)
- Alta frecuencia de compra (14 transacciones/mes vs 6 promedio)
- Cestas muy grandes (22 ítems vs 9.5 promedio)
- Compran recientemente y de manera consistente
- Baja tasa de retorno (alta satisfacción)
- Mayor antigüedad como clientes (lealtad establecida)

Perfil Psicográfico:

- Compradores leales y comprometidos
- Alta propensión al gasto
- Satisfechos con los productos (bajas devoluciones)
- Responden moderadamente a email marketing

Valor Estratégico: □□□□□ (Máxima prioridad)

CLUSTER 1: Window Shoppers

Tamaño: 929 clientes (31.0%)

Métrica	Valor	vs. Promedio General	Ranking
`monthly_transactions`	1.68	**-72.0%**	#4
`avg_basket_size`	3.05	**-67.8%**	#4
`total_spend`	$423	**-82.1%**	#4
`recency_days`	35.59	**+72.1%**	#4
`return_rate`	0.27	**+47.8%**	#4
`avg_session_duration`	52.31	**+36.1%**	□ #1
`email_open_rate`	0.37	-15.7%	#4
`product_views_per_visit`	33.6	+7.1%	#3
`customer_tenure_months`	15.1	-15.7%	#4

Características Distintivas:

- Navegadores frecuentes pero conversión muy baja
- Sesiones más largas del dataset (52 minutos)
- Alta recencia (no compran desde hace ~36 días)
- Tasa de retorno más alta (insatisfacción/indecisión)
- Bajo engagement con emails de marketing
- Cestas pequeñas cuando compran

Perfil Psicográfico:

- Indecisos, buscan pero no compran
- Posible sensibilidad al precio
- Baja lealtad/compromiso con la marca
- Pueden estar comparando con competencia

Riesgo: Alto (churn potencial)

Valor Estratégico: □□ (Requiere reactivación urgente)

CLUSTER 2: Premium Occasional Buyers

Tamaño: 433 clientes (14.4%)

Métrica	Valor	vs. Promedio General	Ranking
`monthly_transactions`	4.04	-32.8%	#3
`avg_basket_size`	18.17	**+91.6%**	□ #2
`total_spend`	$3,876	+63.7%	□ #2

Métrica	Valor	vs. Promedio General	Ranking
`recency_days`	18.9	-8.6%	#2
`return_rate`	0.24	**+31.7%**	#3
`avg_session_duration`	22.36	**-41.8%**	#4
`email_open_rate`	0.53	+19.5%	#1
`product_views_per_visit`	16.55	**-47.2%**	#4
`customer_tenure_months`	21.6	+20.6%	#2

Características Distintivas:

- Compras poco frecuentes pero de alto valor
- Cestas muy grandes (segundo lugar)
- Sesiones cortas y enfocadas (saben qué quieren)
- Pocos productos vistos (navegación eficiente)
- Alta apertura de emails (receptivos a marketing)
- Antiguos clientes (lealtad de largo plazo)

Perfil Psicográfico:

- Compradores estratégicos/planificados
- Alto poder adquisitivo
- Compran por necesidad específica (no browsing casual)
- Responden bien a comunicación personalizada

Oportunidad: Aumentar frecuencia de compra manteniendo ticket alto

Valor Estratégico: ⭐⭐⭐⭐ (Alto potencial de crecimiento)

CLUSTER 3: Low-Engagement Mass Segment

Tamaño: 1,113 clientes (37.1%)

Métrica	Valor	vs. Promedio General	Ranking
`monthly_transactions`	6.59	+9.6%	#2
`avg_basket_size`	5.56	**-41.4%**	#3
`total_spend`	$1,451	-38.7%	#3
`recency_days`	14.53	**-29.7%**	#3
`return_rate`	0.13	**-30.1%**	#2
`avg_session_duration`	36.7	-4.5%	#2
`email_open_rate`	0.44	-0.8%	#3
`product_views_per_visit`	31.4	+0.1%	#2
`customer_tenure_months`	17.5	-2.3%	#3

Características Distintivas:

- Segmento más grande (37% de la base)
- Frecuencia de transacciones cercana al promedio
- Cestas pequeñas (contribución limitada al revenue)
- Compran con cierta regularidad (recencia baja)
- Baja tasa de retorno (satisfacción aceptable)
- Métricas mayoritariamente "promedio"

Perfil Psicográfico:

- Compradores transaccionales de bajo ticket
- Compran productos básicos/necesarios
- No son heavy users pero tampoco están en riesgo
- Representan la "masa" del mercado

Desafío: Mejorar rentabilidad sin aumentar costos significativamente

Valor Estratégico: ⭐⭐⭐ (Volumen importante, rentabilidad limitada)

5.4 Análisis Comparativo de Clusters

5.4.1 Matriz de Comparación Visual

```
Heatmap de Perfiles de Clusters (valores normalizados):

                              Cluster 0  Cluster 1  Cluster 2  Cluster 3
monthly_transactions             [illegible]      [illegible]        [illegible]        [illegible]
avg_basket_size                  [illegible]      [illegible]        [illegible]        [illegible]
total_spend                      [illegible]      [illegible]        [illegible]        [illegible]
avg_session_duration             [illegible]        [illegible]      [illegible]        [illegible]
email_open_rate                  [illegible]        [illegible]        [illegible]        [illegible]
product_views_per_visit          [illegible]        [illegible]        [illegible]        [illegible]
return_rate                      [illegible]        [illegible]      [illegible]        [illegible]
customer_tenure_months           [illegible]        [illegible]        [illegible]        [illegible]
recency_days                     [illegible]        [illegible]      [illegible]        [illegible]

Leyenda: [illegible] Alto | [illegible] Medio | [illegible] Bajo
```

5.4.2 Ranking de Clusters por Métrica Clave

Métrica	#1	#2	#3	#4
Valor Total	C0	C2	C3	C1
Frecuencia	C0	C3	C2	C1
Tamaño de Cesta	C0	C2	C3	C1
Lealtad (Tenure)	C0	C2	C3	C1
Riesgo de Churn	C1	C2	C3	C0

6. VALIDACIÓN DE CLUSTERING

6.1 Métricas de Calidad

6.1.1 Silhouette Score Global

- **K-Means:** 0.317
- **Jerárquico:** 0.316
- **Interpretación:** Estructura moderada pero clara, típica de datos de comportamiento de clientes

6.1.2 Análisis de Silueta por Cluster

Cluster	Silhouette Promedio	Mín	Máx	Interpretación
0	0.42	-0.08	0.76	Bien definido, pocos mal clasificados
1	0.28	-0.15	0.65	Moderado, algunos casos límite
2	0.35	-0.10	0.70	Bueno, separación clara
3	0.31	-0.12	0.68	Moderado, estructura estable

Observaciones:

- Muy pocos valores negativos (< 2% del dataset) → asignaciones generalmente correctas
- Cluster 0 tiene mejor cohesión interna
- Clusters 1 y 3 tienen más puntos en el límite entre grupos

6.2 Reducción Dimensional (PCA)

6.2.1 PCA 2D

- **Componentes:** 2
- **Varianza Explicada Total:** 62.0%
  - PC1: 38.2%
  - PC2: 23.8%

Interpretación:

- PC1 parece capturar "nivel de engagement/valor del cliente"
- PC2 parece relacionarse con "patrón de navegación"

6.2.2 PCA 3D

- **Componentes:** 3
- **Varianza Explicada Total:** 74.6%
  - PC1: 38.2%
  - PC2: 23.8%
  - PC3: 12.6%

Visualización:

- Los clusters muestran separación razonable en espacio 3D
- Existe overlap en zonas de transición (esperado con silhouette ~0.31)
- Centroides de K-Means claramente diferenciados

Nota Importante: Los datos reales están en espacio 9-dimensional, donde la separación es más marcada que en proyecciones 2D/3D.

6.3 Estabilidad del Clustering

6.3.1 Comparación Jerárquico vs K-Means

Cluster K-Means	Cluster Jerárquico Correspondiente	Overlap
0 (525)	0 (540)	~94%
1 (929)	1 (959)	~91%
2 (433)	2 (420)	~95%
3 (1,113)	3 (1,081)	~96%

Conclusión: Alta concordancia entre métodos (>90% overlap), validando robustez de la segmentación.

---

7. ESTRATEGIAS DE NEGOCIO Y RECOMENDACIONES

7.1 Estrategias por Cluster

CLUSTER 0: High-Value Champions (525 clientes, 17.5%)

Objetivo Estratégico: Retención prioritaria y expansión de valor

Acciones Tácticas:

1. **Programa de Lealtad Premium**
   - Acceso temprano a nuevos productos
   - Descuentos exclusivos (no masivos)
   - Servicios VIP: envío gratis, soporte prioritario
   - Eventos exclusivos/experiencias de marca
2. **Personalización Avanzada con IA**
   - Recomendaciones de productos basadas en ML
   - Emails personalizados con productos complementarios
   - Ofertas dinámicas basadas en historial de compra
3. **Sistema de Alerta de Churn VIP**
   - Monitoreo de cambios en frecuencia de compra
   - Alertas automáticas si recency > 15 días
   - Intervención proactiva del equipo de retención

4. **Soporte Prioritario**

    - Línea directa de atención al cliente
    - Resolución acelerada de problemas
    - Gestor de cuenta dedicado (top spenders)

KPIs a Monitorear:

- Churn rate (objetivo: <5% anual)
- Average Order Value (AOV) - meta: mantener >$6,500
- Purchase frequency (meta: 14+ transacciones/mes)
- NPS (Net Promoter Score)

Presupuesto Sugerido: 40% del budget de marketing (máxima protección)

CLUSTER 1: Window Shoppers (929 clientes, 31.0%)

Objetivo Estratégico: Reactivación y conversión

Acciones Tácticas:

1. **Retargeting Dinámico**

    - Remarketing de productos vistos no comprados
    - Emails de "carrito abandonado" con incentivos
    - Anuncios personalizados en redes sociales
    - Push notifications móviles (si app disponible)

2. **Optimización del Proceso de Checkout**

    - Reducir fricción en el proceso de compra
    - Múltiples métodos de pago
    - Checkout express (1-click)
    - Transparencia en costos de envío desde inicio

3. **Incentivos de Reactivación**

    - Cupones "Bienvenido de vuelta" (descuento 15-20%)
    - Free shipping en primera compra post-inactividad
    - Ofertas limitadas en tiempo (urgencia)
    - Programas "Compra 2, lleva 3"

4. **Email Journeys Basados en Recencia**

    - Día 30: Email de recordatorio
    - Día 45: Email con oferta especial
    - Día 60: Email con descuento significativo
    - Segmentación por categorías de interés

5. **Reducción de Barreras de Retorno**

    - Política de devoluciones clara y generosa
    - Información detallada de productos (reducir incertidumbre)
    - Reviews y ratings prominentes
    - Chat en vivo para resolver dudas

KPIs a Monitorear:

- Conversion rate (objetivo: +50% → 3% conversión)
- Recency days (meta: reducir de 36 a <25 días)
- Return rate (meta: reducir de 27% a 18%)
- Email engagement rate

Presupuesto Sugerido: 25% del budget de marketing (alto potencial)

CLUSTER 2: Premium Occasional Buyers (433 clientes, 14.4%)

Objetivo Estratégico: Aumentar frecuencia de compra

Acciones Tácticas:

1. **Recomendaciones de Productos Complementarios**

    - "Productos que podrían interesarte"
    - Bundles temáticos basados en compras anteriores
    - Cross-selling estratégico post-compra
    - Email con nuevos productos en categorías preferidas

2. **Programas de Suscripción/Recordatorios**

    - Subscripción automática para productos recurrentes
    - Recordatorios de recompra para consumibles
    - "Subscribe & Save" con descuento adicional
    - Calendarios de reposición personalizados

3. **Simplificación del Customer Journey** ⚡

- Listas de favoritos/wishlist
- Reorder con 1-click
- Perfil guardado con preferencias
- Proceso de compra ultra-rápido

4. **Incentivos de Segunda Compra**

   - Cupón en el paquete de primera compra
   - Puntos de lealtad con beneficios tangibles
   - "Compra antes de X días y recibe Y beneficio"
   - Early bird access a ventas especiales

KPIs a Monitorear:

- Purchase frequency (meta: 4 → 6 transacciones/mes)
- Time between purchases (reducir de ~30 a ~20 días)
- Subscription adoption rate
- Repeat purchase rate dentro de 30 días

Presupuesto Sugerido: 20% del budget de marketing (crecimiento estratégico)

CLUSTER 3: Low-Engagement Mass Segment (1,113 clientes, 37.1%)

Objetivo Estratégico: Mejorar rentabilidad con eficiencia de costos

Acciones Tácticas:

1. **Promociones de Volumen y Bundles**

   - "3x2" o "Compra X, lleva Y gratis"
   - Descuentos por cantidad
   - Bundles de productos básicos
   - Ofertas de "kit familiar"

2. **Automatización Masiva de Comunicaciones**

   - Email marketing automatizado (bajo costo)
   - Segmentación básica por categoría
   - Newsletters semanales con ofertas
   - SMS marketing para flash sales

3. **Optimización de Inventario**

   - Focus en productos de alta rotación
   - Stock de items básicos/commodities
   - Reducir SKUs especializados para este segmento
   - Cross-merchandising en tienda física

4. **Detección de Movilidad Ascendente**

   - Identificar clientes con potencial de upgrade
   - Monitorear cambios en avg_basket_size
   - Reasignar a estrategias de clusters superiores
   - Campañas de "graduation" a tier superior

KPIs a Monitorear:

- Avg basket size (meta: 5.56 → 7.5 ítems)
- Campaign cost per conversion (minimizar)
- Segment profitability (margen neto)
- Migration rate a clusters superiores

Presupuesto Sugerido: 15% del budget de marketing (eficiencia operativa)

7.2 Roadmap de Implementación

Q1 - Fundamentos y Activación Inicial (Meses 1-3)

Cluster 0 - Champions:

- □ Diseñar programa de lealtad premium (Mes 1)
- □ Implementar sistema de alertas VIP (Mes 2)
- □ Lanzar beneficios exclusivos (Mes 3)

Cluster 1 - Window Shoppers:

- □ Activar retargeting dinámico (Mes 1)
- □ Optimizar flujo de checkout (Mes 1-2)
- □ Lanzar journeys de reactivación por recencia (Mes 2)
- □ Implementar emails de carrito abandonado (Mes 3)

Cluster 2 - Occasional Buyers:

- □ Configurar recordatorios de recompra (Mes 1)
- □ Crear journeys post-compra (Mes 2)

Cluster 3 - Mass Segment:

- □ Lanzar promociones de volumen (Mes 1)
- □ Automatizar newsletters semanales (Mes 2)

Q2 - Personalización y Retención Inteligente(Meses 4-6)

Cluster 0:

- □ Implementar recomendaciones con ML (Mes 4)
- □ Asignar gestores de cuenta a top 100 clientes (Mes 5)
- □ Lanzar eventos exclusivos (Mes 6)

Cluster 1:

- □ Optimizar incentivos basados en performance Q1 (Mes 4)
- □ A/B testing de ofertas de reactivación (Mes 5)
- □ Implementar chat en vivo (Mes 6)

Cluster 2:

- □ Desarrollar journeys de compra rápida (Mes 4)
- □ Lanzar programa Subscribe & Save (Mes 5)
- □ Campañas de segunda compra (Mes 6)

Cluster 3:

- □ Automatizar comunicaciones masivas (Mes 4)
- □ Implementar sistema de detección de upgrade (Mes 5)

Q3 - Expansión de Valor(Meses 7-9)

Cluster 0:

- □ Lanzar tier superior de lealtad (Mes 7)
- □ Programas de referidos VIP (Mes 8)

Cluster 2:

- □ Campañas agresivas de segunda compra (Mes 7)
- □ Expansión de categorías recomendadas (Mes 8)

Cluster 1:

- □ Campañas avanzadas de reactivación (Mes 7-9)

Cluster 3:

- □ Alinear inventario con patrones de uso (Mes 8-9)

Q4 - Optimización y Escala (Meses 10-12)

Todas las Clusters:

- □ Evaluación completa de impacto (Mes 10)
- □ Re-entrenar modelo de clustering con datos actualizados (Mes 11)
- □ Identificar migraciones entre clusters (Mes 11)
- □ Refinar estrategias basadas en ROI (Mes 12)
- □ Planificación año 2 (Mes 12)

7.3 Métricas de Éxito Proyectadas

7.3.1 Impacto Esperado por Cluster

Cluster	Métrica	Baseline	Objetivo Año 1	Incremento
C0	Churn Rate	10%	<5%	-50%
C0	AOV	$6,507	$7,000+	+7.6%
C1	Conversion Rate	2%	3%	+50%
C1	Return Rate	27%	16%	-40%
C1	Avg Recency	36 días	22 días	-39%

Cluster	Métrica	Baseline	Objetivo Año 1	Incremento
C2	Purchase Frequency	4.04/mes	5.6/mes	+40%
C2	Monthly Visits	~4	~6	+40%
C3	Avg Basket Size	5.56	7.2	+30%
C3	Engagement Rate	15%	20%	+30%

7.3.2 Impacto Global Estimado

Revenue Impact:

- Cluster 0: +$259k (525 clientes × $493 incremento AOV)
- Cluster 1: +$393k (465 conversiones adicionales × $846 AOV promedio)
- Cluster 2: +$265k (433 clientes × 1.56 compras adicionales × $389 AOV)
- Cluster 3: +$200k (1,113 clientes × $180 incremento gasto anual)

Total Revenue Uplift Estimado: ~$1.1M - $1.3M (Año 1)

Marketing Efficiency:

- Reducción de desperdicio en campañas genéricas: 30-40%
- Incremento en ROI de marketing: 50-70%
- Mejora en Customer Lifetime Value (CLV): 25-35%

---

8. IMPLEMENTACIÓN TÉCNICA

8.1 Stack Tecnológico Utilizado

8.1.1 Lenguaje y Entorno

- **Python:** 3.8+
- **Jupyter Notebook:** Análisis interactivo y documentación

8.1.2 Librerías Principales

Manipulación de Datos:

```
import numpy as np           # Operaciones numéricas
import pandas as pd          # Manipulación de dataframes
```

Visualización:

```
import matplotlib.pyplot as plt  # Gráficos base
import seaborn as sns            # Visualizaciones estadísticas
from mpl_toolkits.mplot3d import Axes3D  # Gráficos 3D
```

Clustering:

```
from scipy.cluster.hierarchy import dendrogram, linkage, fcluster
from sklearn.cluster import KMeans
```

Preprocesamiento y Validación:

```
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import silhouette_score, silhouette_samples
from sklearn.decomposition import PCA
```

8.2 Arquitectura del Análisis

```
┌─────────────────────────────────────────────────────────┐
│                   DATA PIPELINE                         │
├─────────────────────────────────────────────────────────┤
│                                                         │
│  1. DATA LOADING                                        │
│     ├─ retail_customer_data-1.csv (3,000 rows × 10 cols)   │
│     └─ pd.read_csv()                                    │
│                                                         │
│  2. DATA CLEANING                                       │
│     ├─ Remove customer_id                               │
│     ├─ Check for nulls (none found)                     │
│     └─ Outlier detection (retained)                     │
│                                                         │
│  3. PREPROCESSING                                       │
│     ├─ StandardScaler.fit_transform()                   │
│     └─ 9 features → standardized (μ=0, σ=1)             │
│                                                         │
│  4. HIERARCHICAL CLUSTERING                             │
│     ├─ linkage(method='ward')                           │
│     ├─ dendrogram visualization                         │
│     ├─ fcluster(n_clusters=4)                           │
│     └─ Silhouette score: 0.316                          │
│                                                         │
│  5. K-MEANS CLUSTERING                                  │
│     ├─ Elbow method (k=2 to 10)                         │
│     ├─ Silhouette analysis                              │
│     ├─ KMeans(n_clusters=4, n_init=10)                  │
│     └─ Silhouette score: 0.317                          │
│                                                         │
│  6. VALIDATION                                          │
│     ├─ Silhouette plots per cluster                     │
│     ├─ PCA projection (2D: 62%, 3D: 74.6% variance)     │
│     └─ Method comparison (Hierarchical vs K-Means)      │
│                                                         │
│  7. PROFILING                                           │
│     ├─ Cluster statistics in original units             │
│     ├─ Heatmap of cluster profiles                      │
│     └─ Comparative analysis vs global means             │
│                                                         │
│  8. VISUALIZATION OUTPUTS                               │
│     ├─ correlation_matrix.png                           │
│     ├─ correlation_scatter_plots.png                    │
│     ├─ dendrograms_comparison.png                       │
│     ├─ ward_dendrogram_detailed.png                     │
│     ├─ cluster_extraction_comparison.png                │
│     ├─ elbow_silhouette_analysis.png                    │
│     ├─ cluster_profiles_heatmap.png                     │
│     ├─ silhouette_plot.png                              │
│     ├─ cluster_visualization_pca.png                    │
│     └─ cluster_visualization_pca_3d.png                 │
│                                                         │
│  9. DATA EXPORT                                         │
│     └─ retail_customer_data_with_labels-1.csv           │
│        (original + Cluster_KMeans + Cluster_Hierarchical)   │
│                                                         │
└─────────────────────────────────────────────────────────┘
```

8.3 Fragmentos de Código Clave

8.3.1 Normalización de Datos

```
# Eliminar identificador no informativo
df = pd.read_csv('retail_customer_data-1.csv')
df.drop('customer_id', axis=1, inplace=True)

# Normalización
scaler = StandardScaler()
X_standardized = scaler.fit_transform(df)

# Validación
print(f"Media post-normalización: {X_standardized.mean():.2e}")
print(f"Std post-normalización: {X_standardized.std():.2f}")
```

8.3.2 Clustering Jerárquico

```
# Generar matriz de enlace con Ward
linkage_matrix = linkage(X_standardized, method='ward')

# Visualizar dendrograma
plt.figure(figsize=(14, 7))
dendrogram(linkage_matrix, no_labels=True, color_threshold=60)
plt.axhline(y=60, color='r', linestyle='--', label='Cut (4 clusters)')

# Extraer clusters
labels = fcluster(linkage_matrix, n_clusters=4, criterion='maxclust')
```

8.3.3 K-Means con Validación

```
# Método del codo
inertias = []
K_range = range(2, 11)

for k in K_range:
    kmeans = KMeans(n_clusters=k, random_state=42, n_init=10)
    kmeans.fit(X_standardized)
    inertias.append(kmeans.inertia_)

# K-Means final
optimal_k = 4
kmeans_final = KMeans(n_clusters=optimal_k, random_state=42, n_init=10)
labels = kmeans_final.fit_predict(X_standardized)

# Validación
silhouette = silhouette_score(X_standardized, labels)
print(f"Silhouette Score: {silhouette:.3f}")
```

8.3.4 PCA para Visualización

```
# Reducción a 2D
pca = PCA(n_components=2)
X_pca = pca.fit_transform(X_standardized)

# Scatter plot
plt.scatter(X_pca[:, 0], X_pca[:, 1], c=labels, cmap='tab10', alpha=0.6)
plt.xlabel(f'PC1 ({pca.explained_variance_ratio_[0]:.1%})')
plt.ylabel(f'PC2 ({pca.explained_variance_ratio_[1]:.1%})')

# Proyectar centroides
centroids_pca = pca.transform(kmeans_final.cluster_centers_)
plt.scatter(centroids_pca[:, 0], centroids_pca[:, 1],
            marker='X', s=200, c='red', edgecolors='black')
```

8.4 Outputs Generados

8.4.1 Visualizaciones

Total de imágenes generadas: **12 archivos PNG**

Archivo	Dimensiones	Propósito
`correlation_matrix.png`	10×8	Mapa de calor de correlaciones
`correlation_scatter_plots.png`	18×5	Scatter plots de top 3 correlaciones
`dendrograms_comparison.png`	16×12	Comparación de 4 métodos de linkage
`ward_dendrogram_detailed.png`	14×7	Dendrograma detallado de Ward
`cluster_extraction_comparison.png`	14×10	Distribución de tamaños para k=2-6
`elbow_silhouette_analysis.png`	14×5	Método del codo + scores de silueta
`cluster_profiles_heatmap.png`	12×6	Perfiles de clusters (heatmap)
`silhouette_plot.png`	10×7	Plot de silueta por cluster
`silhouette_score_comparison.png`	-	Comparación de scores
`cluster_visualization_pca.png`	16×6	PCA 2D (K-Means + Jerárquico)
`cluster_visualization_pca_3d.png`	18×7	PCA 3D (ambos métodos)
`hierarchical_clustering_dendrograms.png`	-	Dendrogramas jerárquicos

Configuración de Calidad:

- DPI: 300 (calidad publicación)
- bbox_inches='tight' (recorte óptimo)

8.4.2 Datos Etiquetados

Archivo: `retail_customer_data_with_labels-1.csv`

Estructura:

```
Columnas originales (9) + 2 nuevas columnas:
- Cluster_KMeans (int: 0-3)
- Cluster_Hierarchical (int: 0-3)

Total: 11 columnas × 3,000 filas
```

Uso: Dataset listo para integración en sistemas CRM/Marketing Automation

---

9. LIMITACIONES Y CONSIDERACIONES

9.1 Limitaciones del Análisis

9.1.1 Limitaciones de Datos

Snapshot Temporal:

- Dataset representa un momento en el tiempo
- No captura tendencias estacionales
- Comportamiento puede cambiar en periodos largos

Variables Ausentes:

```
No incluidas pero potencialmente valiosas:
- Datos demográficos (edad, género, ubicación)
- Canal de adquisición
- Categorías de productos comprados
- Sentimiento en reviews/NPS
- Interacciones con servicio al cliente
- Uso de app móvil vs web
- Respuesta a campañas anteriores
```

Ausencia de Datos Temporales:

- No hay series de tiempo
- Imposible detectar tendencias de crecimiento/decaimiento
- No se puede modelar estacionalidad

9.1.2 Limitaciones Metodológicas

Clustering como Técnica No Supervisada:

- No hay "ground truth" para validar
- Interpretación de clusters requiere juicio experto
- Fronteras entre clusters no siempre nítidas

Limitaciones del Silhouette Score:

- Score de 0.317 indica estructura moderada (no perfecta)
- ~2% de clientes con scores negativos (posiblemente mal clasificados)
- Asume clusters convexos (puede no ser realista)

Reducción Dimensional:

- PCA 2D solo captura 62% de varianza
- Visualizaciones 2D/3D simplifican realidad 9-dimensional
- Overlap visual puede exagerar solapamiento real

Efecto de Encadenamiento (Chaining):

- Single linkage susceptible, por eso se usó Ward
- Incluso Ward puede tener bias en formas de clusters

9.2 Supuestos del Modelo

1. **Estabilidad de Comportamiento:**
    - Asumimos que patrones actuales persisten en corto/mediano plazo
    - Shocks externos (crisis económica, pandemia) pueden invalidar segmentación
2. **Independencia de Observaciones:**
    - Clientes tratados como independientes
    - No se modelan efectos de red (referidos, influencia social)
3. **Linealidad en Correlaciones:**
    - Análisis de correlación de Pearson asume relaciones lineales
    - Relaciones no lineales pueden estar presentes pero no detectadas
4. **Homogeneidad Interna:**
    - Clientes en un cluster comparten características similares
    - Variabilidad intra-cluster aún existe

9.3 Riesgos de Implementación

9.3.1 Riesgos Operacionales

Sobre-Segmentación:

- Riesgo de tratar clusters como grupos rígidos
- Clientes en fronteras pueden recibir estrategias subóptimas
- Solución: Implementar "fuzzy membership" (probabilidades de pertenencia)

Privacy y Ética:

- Uso de datos de comportamiento requiere consentimiento
- Cumplimiento GDPR/CCPA necesario
- Transparencia en uso de datos para segmentación

Resistencia Organizacional:

- Equipos acostumbrados a campañas masivas
- Requiere capacitación en marketing personalizado
- Cambio cultural hacia data-driven decisions

9.3.2 Riesgos Técnicos

Drift de Modelo:

- Comportamiento de clientes evoluciona
- Necesidad de re-entrenamiento periódico (recomendado: trimestral)

Escalabilidad:

- Dataset actual: 3,000 clientes (manejable)
- Si MegaMart crece 10x, clustering jerárquico puede ser lento
- Solución: Mini-batch K-Means o clustering incremental

Integración con Sistemas Existentes:

- CRM puede no soportar multi-cluster strategies
- Marketing automation tools necesitan customización
- ETL pipelines para actualizar etiquetas de clusters

9.4 Trabajo Futuro

9.4.1 Enriquecimiento de Datos

Incorporar Variables Adicionales:

```
# Variables deseables:
- customer_age, customer_gender
- acquisition_channel (SEO, Paid, Social, etc.)
- product_categories_purchased (one-hot encoding)
- customer_satisfaction_score
- response_to_previous_campaigns
- device_type (mobile, desktop, tablet)
```

Análisis Temporal:

- Cohort analysis
- Customer journey mapping
- Time-series clustering (para detectar patrones evolutivos)

9.4.2 Modelos Avanzados

Técnicas Complementarias:

1. **Gaussian Mixture Models (GMM):**
    - Clustering probabilístico (soft assignments)
    - Permite modelar incertidumbre en fronteras
2. **DBSCAN:**
    - Clustering basado en densidad
    - Identifica outliers automáticamente
3. **t-SNE o UMAP:**
    - Visualización no lineal (mejor que PCA para clusters complejos)
4. **Hierarchical K-Means:**
    - Combinar ventajas de ambos enfoques

Machine Learning Supervisado:

- Predecir cluster futuro de nuevos clientes
- Modelo de propensión a migrar entre clusters
- Churn prediction dentro de cada cluster

9.4.3 Personalización Avanzada

Recomendación Híbrida:

- Content-based filtering + Collaborative filtering
- Considera cluster + historial individual

Dynamic Clustering:

- Clustering en tiempo real basado en sesión actual
- Micro-moments targeting

Optimización Multi-objetivo:

- Balancear CLV, churn risk, marketing cost
- Estrategias óptimas por cluster con Reinforcement Learning

10. CONCLUSIONES

10.1 Hallazgos Principales

10.1.1 Estructura de la Base de Clientes

MegaMart presenta una base de clientes **heterogénea** con 4 segmentos claramente diferenciados:

1. **17.5% High-Value Champions:** Núcleo de alto valor que genera la mayor parte del revenue
2. **31.0% Window Shoppers:** Oportunidad significativa de conversión (segmento más grande en riesgo)
3. **14.4% Premium Occasional Buyers:** Potencial de crecimiento en frecuencia
4. **37.1% Low-Engagement Mass:** Mayoría transaccional de bajo margen

Implicación Estratégica: Un enfoque "one-size-fits-all" es **subóptimo**. Se requieren estrategias diferenciadas por segmento.

10.1.2 Validación Metodológica

- **Robustez confirmada:** K-Means y Clustering Jerárquico convergen a soluciones prácticamente idénticas (>90% overlap)
- **Calidad aceptable:** Silhouette score de 0.317 indica estructura moderada pero clara
- **Balance óptimo:** k=4 ofrece mejor trade-off entre granularidad y accionabilidad

10.2 Valor de Negocio Generado

10.2.1 Impacto Financiero Estimado

Revenue Uplift Proyectado (Año 1):

- Cluster 0 (Retención): +$259k
- Cluster 1 (Conversión): +$393k
- Cluster 2 (Frecuencia): +$265k
- Cluster 3 (Tamaño de cesta): +$200k
- **Total:** ~$1.1M - $1.3M (incremento 15-18% en revenue)

Eficiencia de Marketing:

- Reducción de desperdicio: 30-40%
- Incremento en ROI: 50-70%
- Mejora en CLV: 25-35%

10.2.2 Beneficios Cualitativos

1. **Experiencia del Cliente Mejorada:**
    - Comunicaciones relevantes (reduce fatiga de marketing)
    - Ofertas personalizadas (aumenta satisfacción)
2. **Toma de Decisiones Informada:**
    - Asignación de recursos basada en datos
    - Priorización de iniciativas por segmento
3. **Agilidad Estratégica:**
    - Monitoreo de migraciones entre clusters
    - Early warning system para churn (Cluster 0)

10.3 Lecciones Aprendidas

10.3.1 Técnicas

Clustering Jerárquico:

- ✅ Pros: Dendrograma intuitivo, no requiere especificar k a priori
- ❌ Cons: Computacionalmente costoso para datasets grandes, sensible a outliers

K-Means:

- ✅ Pros: Rápido, escalable, centroides interpretables
- ❌ Cons: Requiere especificar k, asume clusters esféricos, sensible a inicialización

Lección: Usar ambos métodos como validación cruzada es altamente recomendable.

10.3.2 Negocio

"El dato por sí solo no basta":

- Clustering identifica patrones, pero interpretación requiere conocimiento del dominio
- Naming de clusters (Champions, Window Shoppers, etc.) crucial para buy-in organizacional

Implementación > Análisis:

- El mejor modelo es inútil sin roadmap de implementación
- Recomendaciones deben ser **específicas, medibles, alcanzables, relevantes y acotadas en tiempo (SMART)**

10.4 Recomendaciones Finales

Para MegaMart:

1. **Priorizar Cluster 0 (Champions):**
    - Implementar sistema de alertas de churn en Q1
    - Este 17.5% es crítico para la sostenibilidad del negocio
2. **Atacar Cluster 1 (Window Shoppers) agresivamente:**
    - Representa 31% de la base con alto potencial de conversión
    - ROI esperado más alto en programas de reactivación
3. **Experimentar con Cluster 2:**
    - Programas de suscripción son la estrategia más prometedora
    - A/B testing para encontrar frecuencia óptima de contacto
4. **Optimizar costos en Cluster 3:**
    - Automatización máxima
    - Focus en detección de "graduandos" a clusters superiores

Para Análisis Futuros:

1. **Incorporar datos temporales:**
    - Clustering de series de tiempo
    - Detectar patrones de evolución
2. **Enriquecer con datos de producto:**
    - Category affinity por cluster
    - Product recommendation engines específicos
3. **Validar con test A/B:**
    - Implementar estrategias en submuestra
    - Medir lift real vs proyecciones

---

11. REFERENCIAS Y RECURSOS

11.1 Documentación Técnica

Scikit-learn Documentation:

- K-Means: https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html
- Silhouette Score: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html
- StandardScaler: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html
- PCA: https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html

SciPy Documentation:

- Hierarchical Clustering: https://docs.scipy.org/doc/scipy/reference/cluster.hierarchy.html
- Linkage Methods: https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.hierarchy.linkage.html

11.2 Librerías y Versiones

```
numpy==1.21.0+
pandas==1.3.0+
matplotlib==3.4.0+
seaborn==0.11.0+
scipy==1.7.0+
scikit-learn==0.24.0+
```

11.3 Estructura de Archivos del Proyecto

```
case-03-cluster-analysis/
│
├── data/
│   ├── retail_customer_data-1.csv              # Dataset original (155 KB)
│   └── retail_customer_data_with_labels-1.csv  # Dataset con labels (205 KB)
│
├── notebooks/
│   └── cluster_analysis.ipynb                   # Notebook principal de análisis
│
├── visualizations/
│   ├── correlation_matrix.png
│   ├── correlation_scatter_plots.png
│   ├── dendrograms_comparison.png
│   ├── ward_dendrogram_detailed.png
│   ├── cluster_extraction_comparison.png
│   ├── elbow_silhouette_analysis.png
│   ├── cluster_profiles_heatmap.png
│   ├── silhouette_plot.png
│   ├── silhouette_score_comparison.png
│   ├── cluster_visualization_pca.png
│   ├── cluster_visualization_pca_3d.png
│   └── hierarchical_clustering_dendrograms.png
│
├── reports/
│   ├── Executive_summary.pdf                    # Resumen ejecutivo (125 KB)
│   ├── Technical_report.pdf                     # Reporte técnico (1.9 MB)
│   └── Team3_Slides (2).pdf                     # Presentación (1.4 MB)
│
└── README.md                                     # Documentación principal (9 KB)
```

11.4 Contacto y Soporte

Equipo de Desarrollo:

- César Isao Pastelin Kohagura - A01659947
- Luis Emilio Fernández González - A01659517
- Eduardo Botello Casey - A01659281

Institución: Tecnológico de Monterrey **Curso:** Aplicación de Métodos Multivariados en Ciencia de Datos **Fecha:** Noviembre 2025

APÉNDICES

Apéndice A: Fórmulas Matemáticas

A.1 Distancia Euclidiana (K-Means)

$$d(p, q) = \sqrt{[\Sigma_i (p_i - q_i)^2]}$$

A.2 Silhouette Score

```
Para cada muestra i:
a(i) = distancia promedio intra-cluster
b(i) = distancia promedio al cluster más cercano

s(i) = [b(i) - a(i)] / max{a(i), b(i)}

Silhouette global = promedio de s(i) para todas las muestras
```

A.3 Inercia (WCSS)

```
WCSS = Σₖ Σᵢ∈Cₖ ||xᵢ - μₖ||²

donde:
- Cₖ = cluster k
- xᵢ = muestra i
- μₖ = centroide del cluster k
```

A.4 Ward's Linkage Distance

```
d(u, v) = √[ (2·|u|·|v|) / (|u| + |v|) ] · ||cᵤ - cᵥ||₂

donde:
- |u|, |v| = tamaños de clusters u y v
- cᵤ, cᵥ = centroides de clusters u y v
```

Apéndice B: Glosario de Términos

Término	Definición
Clustering	Técnica de ML no supervisado que agrupa datos similares
Centroide	Punto medio de un cluster (promedio de todos sus miembros)
Dendrograma	Diagrama de árbol que muestra fusiones jerárquicas
Inercia (Inertia)	Suma de distancias cuadradas a centroides (menor = mejor)
Silhouette Score	Métrica de calidad de clustering (-1 a +1, mayor = mejor)
PCA	Principal Component Analysis, reduce dimensionalidad
Linkage	Método para calcular distancia entre clusters
Churn	Tasa de abandono de clientes
CLV	Customer Lifetime Value, valor total de un cliente
AOV	Average Order Value, valor promedio de pedido
WCSS	Within-Cluster Sum of Squares (ver Inercia)
Recency	Días desde última compra
Tenure	Antigüedad como cliente

Apéndice C: Comandos de Reproducción

C.1 Instalación de Dependencias

```
pip install numpy pandas matplotlib seaborn scipy scikit-learn jupyter
```

C.2 Ejecución del Notebook

```
cd case-03-cluster-analysis/notebooks
jupyter notebook cluster_analysis.ipynb
```

C.3 Generación de Visualizaciones

```
# Todas las visualizaciones se generan automáticamente al ejecutar el notebook
# Los archivos PNG se guardan en la carpeta raíz del proyecto
```

FIN DEL REPORTE TÉCNICO

Generado el: 2025-11-26 Versión: 1.0 Documento: Reporte Técnico Completo - Análisis de Segmentación de Clientes MegaMart