
Análisis de Clúster para la generación de campañas de mercadeo en la distribución de la toxina US Botulinum en una compañía cosmética

Cristhian Amaya, Andres Beltrán, Gloria Ramos y Sergio Rojas

Maestría MIAD Uniandes

Resumen

El proyecto está enfocado en desarrollar un análisis de clusters para una compañía que produce y distribuye la toxina U.S. botulinum, un grupo de exotoxinas que interfieren la transmisión neural causando parálisis muscular (Nigam & Nigam, 2010). Este es un producto inyectable que los usuarios solicitan en centros médicos con especialidades en estética. El mercado de Estados Unidos para este producto está valorado en US\$4.6 Billones, y se espera que crezca año a año al 5% (Fortune Business Insights, 2023), lo que crea una oportunidad para esta compañía de capturar mayores porciones de mercado. La compañía ha venido recogiendo datos de los usuarios que visitan su página web por medio de la herramienta Google Analytics. La base de datos contiene información de las sesiones (las visitas de los usuarios a la página web), e información relacionada con los intereses, edad, sexo y demás atributos de los usuarios que visitan la página. Con base en esta base de datos se espera desarrollar la segmentación de audiencias con el fin de maximizar el impacto de las campañas de mercadeo que la compañía lanzará por medio de otras plataformas digitales.

Introducción

Una compañía del sector cosmético cuyo mercado se enfoca a Estados Unidos (USA) produce fármacos inyectables para el rostro, para todos los consumidores que deseen preservar el estado de su piel (prevenir tener más arrugas) o estirar la piel del rostro para mantener un aspecto joven. De acuerdo con (Fortune Business Insights, 2023) en 2022 el mercado de la toxina U.S. botulinum (un componente principal para la fabricación de estos inyectables) estuvo valorado en US\$4.6 billones, y para el 2030 se espera que crezca a US\$6.7 billones, lo que se traduce en oportunidades de mercado para este tipo de productores.

Las compañías de este sector desarrollan distintas estrategias de mercadeo en dos frentes (B2B y B2C) para capturar porciones de este mercado, dado que, estos inyectables son distribuidos y aplicados por medio de compañías del sector salud con capacidades para realizar procedimientos de belleza (e.g. Hospitales, centros dermatológicos, centros de inyectores, entre otros). Es decir, el consumidor final (los usuarios que se inyectan) no pueden adquirir estas toxinas como se adquiere cualquier producto de un distribuidor, sino que se acercan a un centro de salud, donde les inyectan los productos que ellos demandan. En este sentido, las compañías que ofertan dichas toxinas deben desarrollar estrategias de mercadeo para corporaciones (lo que se llama mercadeo

B2B), y para consumidores finales (o más conocido como B2C). El presente trabajo estará enfocado en el segmento B2C (sin dejar de lado que el análisis de clusters también puede desarrollarse para el segmento B2B). Las bases de datos suministradas son capturadas desde sitios web enfocados en el segmento B2C.

La compañía cosmética ha recogido información de los usuarios que visitan su página web mediante Google Analytics, “una plataforma que recoge datos de [los] sitios web ... para crear informes que proporcionan estadísticas para el provecho de [la] empresa” (Google, 2023). Dicha información está compuesta por datos de rango de edad, genero, intereses de los usuarios, ente otros. Es importante resaltar que la plataforma no distribuye información personal identificable (PII por sus siglas en ingles), por lo que no se usa datos de carácter sensible en este análisis. Con los datos proveídos, la compañía espera segmentar a los usuarios que visitan su página web en un número reducido de clusters.

Dentro de las ventajas que se tienen por incluir el análisis de clusters en las estrategias de mercadeo, se encuentran: (1) ayudar a segmentar a los consumidores basados en sus características, preferencias, comportamientos, o necesidades heterogéneas (dado que estas últimas están cambiando constantemente); permitiendo incrementar el entendimiento de la base de consumidores para adaptar los esfuerzos de mercadeo y maximizar el impacto de recursos limitados (LinkedIn, 2023; Dibb, Market segmentation: strategies for success, 1998), (2) ofrecer recomendaciones personalizadas, venta cruzada, cupones, descuentos, o promociones a los usuarios de manera enfocada (LinkedIn, 2023), (3) incrementar la eficiencia las estrategias de focalización (targeting), al añadir claridad al proceso de desarrollo de estrategias de mercadeo (Fonseca, 2011; Dibb, Market segmentation: strategies for success, 1998), (4) identificar los segmentos no explorados para crear una ventaja competitiva, desarrollando productos y servicios llamativos únicos para dichas audiencias (Fonseca, 2011; Dibb, Market segmentation: strategies for success, 1998), (5) soportar los lanzamientos de nuevos productos (Saunders, 1980), (6) mejorar el rendimiento organizacional (Dibb, Stern, & Robin, 2002), entre otras.

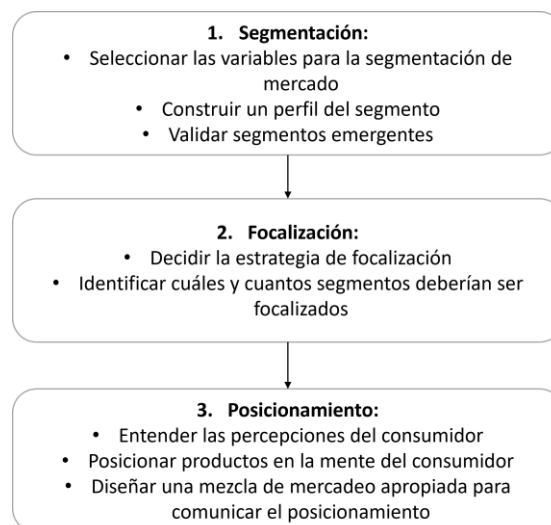


Ilustración 1 Adaptado de (Dibb, Market segmentation: strategies for success, 1998)

El análisis de clusters para mercadeo comprende tres fases: segmentación, focalización y posicionamiento (llamado STP por sus siglas en inglés) (Dibb, 1998).

El presente trabajo estará destinado en cubrir la primera fase de este análisis de clusters, segmentación.

Revisión preliminar de antecedentes en la literatura

En un artículo de 1980 (Saunders, 1980) considera que los métodos de clusters más apropiados para desarrollar segmentos de mercado, experimentación y posicionamiento de producto son Segmentación de beneficios y agrupamiento jerárquico, este último apoyado por un análisis de factor R en caso de que los datos sean apropiados para la estandarización. En caso contrario la distancia euclidiana fue recomendada. Se proponen alternativas al agrupamiento jerárquico como el método de factor de Q, particiones (optimización), búsqueda de densidad, Aglomeración (*clumping* en inglés) y agrupamiento restringido.

De acuerdo con (Dibb, Stern, & Robin, 2002) el análisis de cluster es la herramienta más efectiva para incrementar el entendimiento del consumidor, e impactar en el rendimiento organizacional. Esto comparando distintos métodos como “Las Estadísticas Descriptivas”, “Tabulación Cruzada”, “Correlación”, “Regresión de dos variables”, “Regresión múltiple”, “Análisis de Factor”, “Análisis discriminante”, “Análisis de Conjoint”, entre otros. (Dibb, Stern, & Robin, 2002) concluye esto tras recopilar las percepciones de distintos académicos expertos en esta área.

Por otro lado, cuando se hace un análisis de cluster para segmentar mercados, (Lin, 2002) propone que no se considere únicamente variables demográficas, sino también variables psicográficas, es decir, variables que permitan entender el estilo de vida y la personalidad de los consumidores, explorar modelos de consumo e identificar características relevantes de la marca. Esto es justificado por la idea que, para crear submercados homogéneos, las variables tradicionales demográficas (como edad, sexo y ubicación), no son suficientes para agrupar a los usuarios dentro de segmentos, pues siguen teniendo necesidades heterogéneas. Para demostrar esto, el estudio añadió variables de clasificación tipo VALS2 (Valores y estilos de vida) y LOV (lista de valores), y comparo su rendimiento con un estudio que solo incluye variables demográficas, demostrando así que más información psicográfica debe ser añadida en este tipo de análisis.

Finalmente, (Fonseca, 2011) analiza las ventajas competitivas de construir un análisis de cluster apropiado para segmentar mercados y ofrecer productos diferenciados a estos submercados resultantes, de manera que se pueda llenar de manera más precisa las demandas de los clientes, y por lo tanto extraer mayores porcentajes de utilidad del mercado. En este último, (Fonseca, 2011) desarrolla un ejemplo con una base de datos de modalidad mixta, seleccionando el modelo más apropiado con criterio de AIC. El modelo más apropiado del estudio fue un modelo latente de dos clases.

Descripción detallada de los datos

Información basada en el sitio de Google Analytics Data API Google Developers (Developers, 2023) donde describen la generación de los datos generados desde Google Analytics.

1. **‘brandingInterest’**: Esta variable muestra el tipo de interés que tuvieron los usuarios que se encuentran en etapas avanzadas del proceso de compra. Los usuarios se registran en

diferentes categorías según sus intereses. Por ejemplo, “Estilos de vida y hobbies/Amantes de los animales” o “Viajeros/Fans de los viajes/Viajeros de playa”.

2. **‘userAgeBracket’**: Esta variable se utiliza para agrupar a los usuarios en categorías o rangos de edad. Por ejemplo, los rangos podrían ser "Menores de 18 años", "De 18 a 25 años", "De 26 a 35 años" y así sucesivamente. Esta variable ayuda a comprender la distribución de edades de los usuarios en el conjunto de datos.
3. **‘userGender’**: Esta variable indica el género o sexo de los usuarios. Puede tomar valores como "Masculino", "Femenino", "No binario" u otras categorías que representen la identidad de género de los usuarios. El propósito de esta variable es analizar la composición de género de la audiencia o realizar análisis específicos relacionados con el género.
4. **‘country’**: Esta variable indica el país de origen o ubicación geográfica de los usuarios. Cada valor corresponde a un país específico, como "Estados Unidos", "Canadá", "Reino Unido", etc. El país de origen es útil para realizar análisis geográficos, como la segmentación de usuarios por ubicación geográfica.
5. **‘region’**: Muestra la zona geográfica de donde viene el usuario, basándose en su IP.
6. **‘sessions’**: Muestra el número de sesiones que superaron los 10 segundos, lograron una conversión o tuvieron 2 o más visitas.

En total se cuentan con 14316 registros, realizando un análisis detallado no tiene ningún valor nulo, la variable ‘brandingInterest’ tiene 116 valores únicos, hay 7 grupos etarios comprendidos en la variable ‘userAgeBracket’, 3 géneros diferenciados en ‘userGender’, todos los datos comprenden solo un país y 20 regiones.

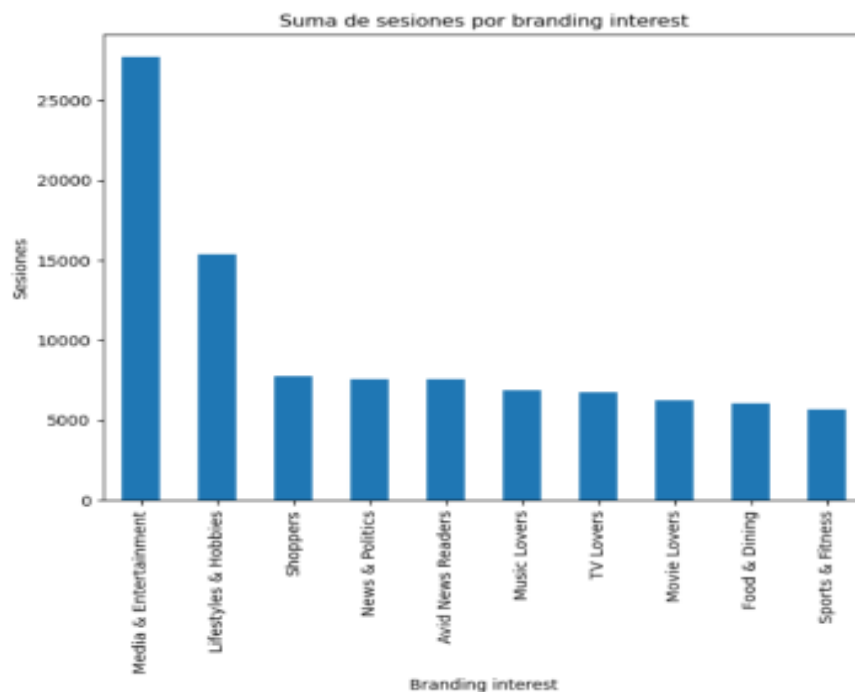


Ilustración 2. Grafico de Barras con suma de sesiones por interes

En la variable 'brandingInterest' se puede observar que "Media & Entertainment" es el interés más alto, seguido por "Lifestyles & Hobbies". El top 10 comprende el 47% del total de las sesiones.

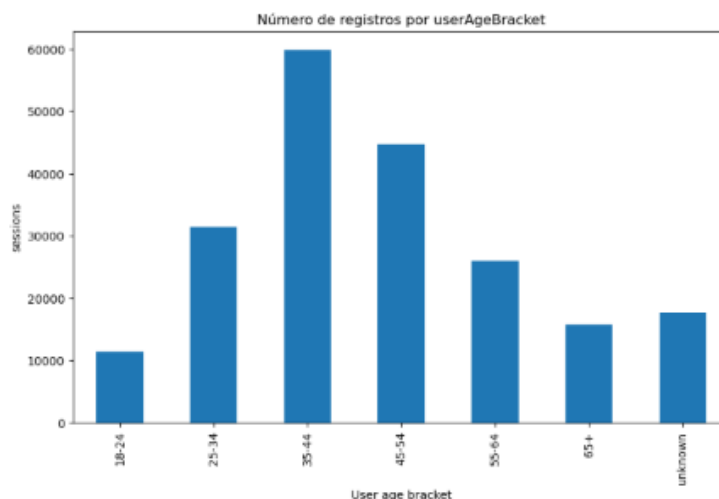


Ilustración 3. Número de sesiones por grupo etario,

El grupo etario que tiene mayor número de sesiones es de 35 a 44 años, que comprende un total 59.887 de sesiones que comprende un 28.98% de los datos

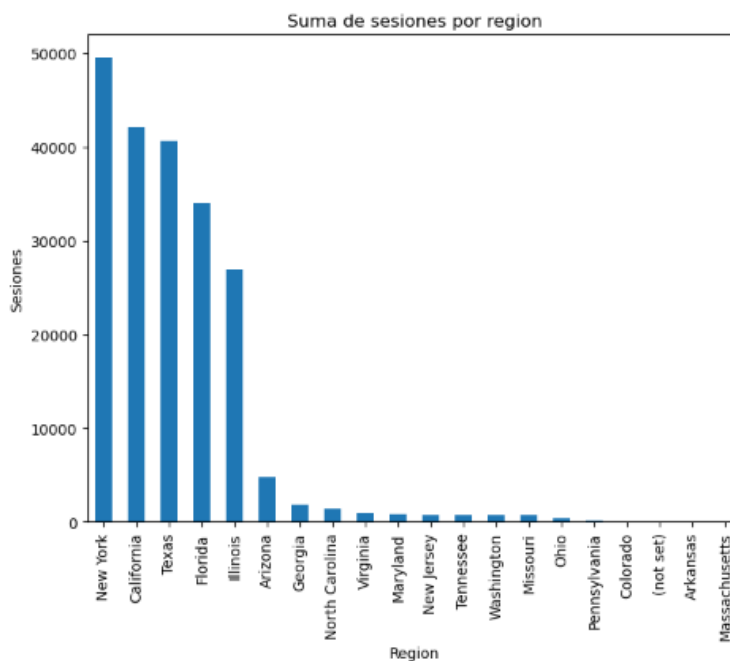


Ilustración 4 Número de rsiones por región

La región con numero de sesiones es de New York con 49,502 sesiones de un total de 206,644 sesiones en el total de registros.

Propuesta metodológica

Objetivos:

- Aplicar la clusterización jerárquica como técnica de aprendizaje no supervisado para agrupar datos en clústeres.
- Identificar patrones y estructuras ocultas en el conjunto de datos sin requerir etiquetas previas.
- Generar conocimientos valiosos a partir de la interpretación de los resultados de la clusterización.

Metodología Propuesta:

1. Recopilación y Preprocesamiento de Datos: Obtener el conjunto de datos de origen y realizar una exploración inicial. Realizar tareas de limpieza de datos, tratamiento de valores faltantes y normalización de ser necesario.
2. Selección de Variables (Opcional): Si el conjunto de datos es grande o contiene muchas características, considerar la selección de variables relevantes para simplificar el análisis y mejorar la eficiencia computacional.
3. Configuración de Parámetros: Determinar el método de enlace (single, complete, average, etc.) y la métrica de distancia (euclidiana, Manhattan, etc.) más apropiados para la clusterización jerárquica.
4. Aplicación de Clusterización Jerárquica: Utilizar el algoritmo de clusterización jerárquica para agrupar las muestras en clústeres de manera jerárquica.

Construir el dendrograma inicial para visualizar la estructura jerárquica de los clústeres.

5. Evaluación del Número Óptimo de Clústeres: Emplear métodos como el codo (elbow method), índice de silueta u otros métodos de selección de clústeres para determinar el número adecuado de grupos.
6. Interpretación de Resultados: Evaluar y describir la coherencia de los clústeres identificados. Explorar y documentar las características y patrones comunes dentro de cada clúster.
7. Validación de Resultados (Opcional): Si se dispone de etiquetas verdaderas, se pueden utilizar medidas de validación como el índice de Rand o la pureza para evaluar la calidad de la clusterización.
8. Documentación y Comunicación de Resultados: Generar un informe detallado que incluya hallazgos, conclusiones y visualizaciones que respalden la interpretación de los resultados.

Conclusiones:

Esta metodología propuesta permite aplicar el aprendizaje no supervisado a través del algoritmo de clusterización jerárquica para explorar y descubrir estructuras ocultas en un conjunto de datos. Los resultados obtenidos proporcionarán una mejor comprensión de la naturaleza de los datos y podrán ser utilizados en diversos campos, desde la segmentación de clientes hasta la clasificación

de elementos en taxonomías. La estructura detallada de la metodología facilita un análisis sistemático y una interpretación sólida de los resultados.

Bibliografía

- Developers, G. (16 de Ago de 2023). *Google Analytics*. Obtenido de Google Analytics:
<https://developers.google.com/analytics/devguides/reporting/data/v1/api-schema?hl=es-419>
- Dibb, S. (1998). Market segmentation: strategies for success. *Marketing intelligence & planning*, 394-406.
- Dibb, S., Stern, P., & Robin, W. (2002). Marketing knowledge and the value of segmentation. *Marketing Intelligence & planning*, 113-119.
- Fonseca, J. R. (2011). Why does segmentation matter? Identifying Market segments through a mixed methodology. *European Retail Research*, 1-26.
- Fortune Business Insights. (2023). *U.S. Botulinum Toxin Market Size, Share & Industry Analysis, By Application (Therapeutics), and Aesthetics, By Type, By End User and Forecast, 2023-2030*. USA: Fortune Business Insights.
- Google. (2023). *How Google Analytics works*. Obtenido de Google Support:
<https://support.google.com/analytics/answer/12159447?hl=en>
- Lin, C.-F. (2002). Segmenting customer brand preference: demographic or psychographic. *Journal of Product and Brand Management*, 249-268.
- LinkedIn. (2023). *How do you use cluster analysis to segment your customers*. Obtenido de LinkedIn advice: <https://www.linkedin.com/advice/3/how-do-you-use-cluster-analysis-segment-your>
- Nigam, P., & Nigam, A. (2010). Botulinum toxin. *National Library of Medicine*, 8-14.
- Saunders, J. (1980). Cluster Analysis for Market segmentation. *European Journal of Marketing*, 422-435.