



COMILLAS
UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

FACULTAD DE EMPRESARIALES

Psicografía y Big Data en el mundo del marketing

Autor: Marian Domínguez Luqui

5º E-3 Analytics

Director: Prof. Raquel Redondo Palomo

Madrid

Junio 2024

RESUMEN

Este trabajo analiza la interrelación entre la psicografía y el Big Data, analiza su evolución y demuestra su practicidad en el ámbito del marketing.

En primer lugar, se presenta una visión general del marketing como disciplina económica destacando su adaptación a las diferentes tendencias del mercado. Se introduce el concepto de psicografía y se examina su relación con el comportamiento del consumidor.

Igualmente, en un segundo apartado se desarrollan técnicas del Big Data aplicadas al perfilado psicológico para posteriormente ilustrar su estrecha relación mediante un caso práctico.

PALABRAS CLAVE: Marketing, psicografía, OCEAN, análisis de sentimiento, segmentación, cluster.

ABSTRACT

This work analyzes the interrelation between psychographics and Big Data, examines their evolution, and demonstrates their practicality in the field of marketing.

Firstly, it presents an overview of marketing as an economic discipline, highlighting its adaptation to various market trends. The concept of psychographics is introduced, and its relationship with consumer behavior is examined.

Similarly, the second section develops Big Data techniques applied to psychological profiling, subsequently illustrating their close relationship through a practical case study.

KEYWORDS: Marketing, psychographics, OCEAN, sentiment analysis, segmentation, clustering.

ÍNDICE

Listado de abreviaturas -----	6
1.Introducción -----	7
1.1 Objetivos -----	7
1.2 Estructura -----	8
1.3 Metodología -----	9
2. Psicografía: Perfilado Psicológico y Aplicaciones -----	10
2.1 Conceptos fundamentales y claves históricas -----	10
2.2 La psicología detrás del consumidor-----	13
2.3 Las variables psicográficas-----	18
3. Técnicas de Big Data al servicio del perfilado psicológico -----	24
3.1 La obtención de datos-----	24
3.2 Técnicas más relevantes-----	26
3.2.1 Minería de datos-----	26
3.2.2. Análisis de Sentimiento-----	29
3.2.3 Técnica de Clustering-----	31
4. Caso práctico: Always Boutique -----	33
4.1 Análisis de Reseñas y Sentimiento-----	33
4.2 Clustering psicográfico y estrategias de marketing -----	39
5. Conclusiones -----	45
6. Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado -----	47
7. Bibliografía -----	49
8.Anexos -----	52

Listado de abreviaturas

AIO	Actividades, interés y opiniones
DF	Document Frequency
DTM	Document Term Matrix
ID	Identificador
IDF	Inverse Document Frequency
IND	Independent
OCEAN	Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism
PNL	Procesamiento del Lenguaje Natural
RGPD	Reglamento General de Protección de Datos
TF	Term Frequency
TFIDF	Term-Frequency-Inverse Document Frequency
VALS	Values and Lifestyles

1.Introducción

1.1 Objetivos

Este TFG tiene por objeto analizar la relación entre la psicografía y el marketing, para ello se estudiarán técnicas de Big Data como el análisis de sentimientos o la segmentación mediante K-Means, metodologías que permiten interrelacionar ambos conceptos.

Para alcanzar este objetivo, se plantean varias metas específicas; en primer lugar, se busca entender en qué situación se encuentra el sector del marketing a día de hoy, cuales son los retos que el siglo XXI ha generado y qué herramientas permiten su adaptación. La personalización es clave, por ello, otra meta es comprender la mente del consumidor moderno así como las variables más relevantes para su estudio.

Adicionalmente, se busca materializar estos conceptos, darles un enfoque práctico, para ello; otro objetivo a alcanzar es analizar y estudiar las técnicas de Machine Learning que hacen posible un análisis psicográfico.

El objetivo final es integrar todos estos conceptos en una situación real, se desarrollará un caso de estudio que demuestra la aplicabilidad y relevancia de las metodologías en la estrategia de marketing de una empresa.

1.2 Estructura

La finalidad de la estructura seleccionada es guiar al lector por los objetivos propuestos aportando soluciones que van tejiendo una red de conocimiento.

El punto de partida es un análisis histórico, se busca dar respuesta a las siguientes preguntas; ¿cómo hemos llegado a la situación actual? ¿Por qué ha surgido la necesidad de crear métodos de marketing personalizados?

Posteriormente se busca entrar en la mente del consumidor, comprender la psicología detrás de sus decisiones, encontrar patrones y variables determinantes en su proceso de consumo entendiendo como cada variable afecta esta decisión y qué mecanismos existen para su medición de manera que se aporte rigurosidad en un ámbito tan subjetivo como la mente humana.

Una vez estudiada la razón del proyecto y las variables a emplear se llega al análisis técnico. El siguiente paso se centra en recabar diferentes técnicas de Machine Learning que puedan materializar un proyecto de psicografía en el mundo del marketing.

Finalmente, todos los elementos desarrollados quedan unidos en el caso práctico de la tienda de ropa online “Always Boutique”. Empleando amplios datasets sobre las reseñas de productos publicadas en la web así como la personalidad de los clientes se permite dar una respuesta a las siguientes preguntas; ¿Cómo se percibe el negocio de e-commerce? ¿Qué se puede mejorar para aumentar el grado de satisfacción de los clientes? ¿Cómo son los clientes de Always Boutique? Cada parte práctica concluye con una análisis estratégico de marketing sobre las fortalezas, debilidades y oportunidades de Always Boutique.

1.3 Metodología

Para poder responder a las preguntas que constituyen los objetivos de este trabajo se ha realizado un estudio teórico y práctico. En primer lugar, mediante la extracción de información de bibliotecas digitales y buscadores académicos como Dialnet o Google Scholar.

He tenido acceso a las últimas publicaciones en materia de psicografía, marketing personalizado y las técnicas más pioneras de Machine Learning para llevarlo a cabo. Las fuentes consultadas, escritas por eminencias del sector del Marketing y Machine Learning constituyen los pilares sobre los que he cimentado este proyecto.

La internacionalidad de este trabajo queda plasmada en la diversidad y amplitud de sus fuentes. Con el fin de garantizar una investigación relevante y actualizada he llevado a cabo una exhaustiva revisión de publicaciones y estudios de expertos académicos de diversas partes del mundo. Gracias a esta metodología se ha podido aportar una perspectiva global, entendiendo conceptos desde muchos prismas y así fortaleciendo el trabajo.

En lo referente a la obtención de datasets, he empleado la plataforma de análisis de datos Kaggle. Esta fuente proporciona acceso a una amplia colección de datasets en materias diversas, esto aporta un importante valor a mi investigación psicográfica. Gracias a estos datasets he conseguido llevar a cabo un rigurosos análisis de Machine Learning y así llegar al objetivo final, materializar la relación de la psicografía con el Big Data y entender cómo este permite la formación de estrategias de actuación.

2. Psicografía: Perfilado Psicológico y Aplicaciones

2.1 Conceptos fundamentales y claves históricas

El concepto de mercadotecnia proviene del término anglosajón “marketing”, es el proceso de identificación y mantenimiento de consumidores, se busca construir una red de consumidores y fortalecer continuamente dicha red para lograr un beneficio mutuo para ambas partes mediante contactos interactivos, individualizados y de valor añadido durante un periodo de tiempo largo¹.

Schiffman y Kanuk identificaron en la mercadotecnia una serie de etapas claves. Comprender dichas fases permite ofrecer una visión integral de los fundamentos que han moldeado este campo a lo largo de los años.

Durante los años posteriores a la Segunda Guerra Mundial, el foco se ponía en la producción de bienes duraderos. La difícil situación vivida durante los conflictos bélicos y la reconstrucción social forjó a unos consumidores centrados en la calidad del producto y su durabilidad, no se valoraban percepciones estéticas. Un ejemplo claro para entender esta fase es aquella descrita para vender vehículos Ford Modelo T; *“El cliente puede tener cualquier color que quiera, siempre y cuando sea negro”*.²

Con el paso de los años llegó la recuperación económica, las fábricas crecieron y la producción se multiplicó. Este periodo se caracterizó por la importancia de alcanzar los objetivos de venta, las empresas buscaban evitar las pérdidas relacionadas con la sobreproducción como el exceso de inventario, por ello, su estrategia comercial se centró en

¹ Definición de marketing de Gabi y Chalasani (1994)

Carasila, C., Milton, A., “El concepto de Marketing: pasado y presente” en *Revista de Ciencias Sociales*, vol 14, nº2, 2008. (Disponible en https://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1315-95182008000200014; última consulta 1/01/2024).

² Ford, 1909 citado en Kotler, P., Kartajaya, H., y Setiawan, I., *“Marketing 4.0 moving from traditional to digital”* John Wiley & Sons, Inc, New Jersey, 2017, p. 19.

la venta masiva, sin importar las necesidades reales del cliente o su calidad como consumidor fiel. En estos años se emplean prácticas comerciales cuestionables como la obsolescencia programada, la manipulación de precios o el bombardeo constante con mensajes publicitarios. El consumidor queda en una posición vulnerable; la preeminencia de cumplir los objetivos de ventas hizo que la satisfacción del cliente en el largo plazo y su retención quedarán relegadas a un segundo plano.

Con el auge de la economía proliferan multitud de marcas que ofrecían el mismo producto o prestaciones similares, esto lleva a la industria a reenfocar su estrategia de venta, lo importante ya no es lo que ellos ofrecen sino lo que los consumidores quieren comprar, nace así una etapa centrada en escuchar y entender los deseos del consumidor y satisfacerlos con productos y servicios. Esta es una fase más técnica y analítica donde los hábitos de compra y las motivaciones son estudiados, comienza una formación teórica del marketing. Las empresas comienzan a invertir en investigación de mercados, análisis de satisfacción y a seguir las tendencias del momento, este es un punto clave para el mundo del marketing y el Big Data, las empresas comienzan a retroalimentarse directamente con toda aquella información que les proporcionan sus clientes.

En la cuarta etapa de la mercadotecnia, el consumidor juega un papel nuclear, las marcas buscan construir una identidad con la que los individuos puedan sentirse identificados, el branding se constituye como un proceso estratégico para construir una relación entre el productor y el consumidor, la identidad de marca va más allá de los productos en venta y se convierte en un ente con principios. Se busca conectar emocionalmente con el cliente, que cuando se compre un producto o contrate un servicio elementos intangibles como la experiencia de marca tengan más valor que el bien en sí.³

³ Cienfuegos Alvarado, C., Ramos Canales, E., Rábgos Rojas, C., y del Prado Flores, R., “Mercadotecnia integral y branding, comprehensive marketing and branding” en *Revista Anahuac*, 2021, pp 65-82. (Disponible en <https://revistas.anahuac.mx/sintaxis/article/view/1026/929>; última consulta 1/01/2024).

Asimismo, los estudios de mercados son cruciales para entender al consumidor, en este ámbito, la psicografía emerge como un elemento crucial pues aporta una comprensión más profunda del comportamiento del consumidor así como una segmentación de mercado más efectiva.

El marketing moderno se sustenta sobre elementos tradicionales como el marketing mix; los conceptos de precio, producto, distribución y promoción son la base sobre la que se desarrollan ideas más pioneras como las políticas de responsabilidad corporativa o el marketing digital⁴. Actualmente nos encontramos en la fase del marketing 5.0 cuya característica esencial es el entorno digital, la existencia de influencers y el uso de herramientas como la realidad aumentada o la inteligencia artificial mejoran la experiencia de compra del cliente.⁵

Las principales plataformas del marketing del siglo XXI son entre otras; las redes sociales, las páginas web y las aplicaciones móviles, todos ellos son entornos saturados donde el consumidor recibe un impacto publicitario cada 10 segundo, esto equivale a 6.000 impactos diarios, de los cuales, solo es capaz de retener entorno a un 5-10%⁶. Esta estadística hace necesario que las empresas lleven a cabo estrategias con mensajes claros y rápidos, adaptados al cada vez menor tiempo de atención del consumidor. La capacidad de comunicar de manera eficiente y concisa mejora la satisfacción del cliente y optimiza las operaciones de marketing, una buena campaña se caracteriza por su adaptación a las tendencias de un mercado cada vez más dinámico donde los mensajes deben decir mucho en muy poco tiempo⁷.

⁴Grönroos, C., Voima, P., “Making Sense of Value and Value Co-Creation in Service Logic” en *Hanken School of Economics Working Papers*, 2011. (Disponible en https://www.researchgate.net/publication/279495418_Making_Sense_of_Value_and_Value_Co-Creation_in_Service_Logic; última consulta 1/01/2024).

⁵ Kotler, P., Kartajaya, H., Setiawan, I., *Marketing 5.0 Technology for Humanity*, John Wiley & Sons, 2021.

⁶ Barrocal Blanquez, J., “Destacar entre miles de estímulos”, *Wola Publicidad*, 2022. Disponible en <https://wolapublicidad.com/destacar-entre-infinitos-estimulos/>; última consulta 2/05/2024).

⁷ González-Ferriz, F., “El marketing 5.0 y su efecto en la estrategia empresarial del sector industrial en España” en *Revista de Marketing Aplicado (Red Marka)*, vol 25, nº 1, 2021 pp 1-20. Disponible en

La evolución de la mercadotecnia refleja en primer lugar, como las estrategias empresariales han virado de forma radical, del original enfoque puramente transaccional donde vender y alcanzar unos objetivos numéricos era la prioridad al sofisticado proceso de forjar una relación de cercanía y fidelidad con el consumidor, buscando que cada compra sea una experiencia.

Si algo se puede aprender con la evolución del marketing es su constante adaptación al mercado y las necesidades del consumidor, la reinención y creación de nuevas técnicas y estrategias ha sido nuclear para su supervivencia y relevancia. La fase actual, caracterizada por la digitalización, la ciencia de datos, la personalización máxima y el marketing 5.0 invita a reflexionar sobre los nuevos retos e implicaciones éticas que deberán ser tratados para garantizar que el marketing avance con el consumidor y no a su costa.

2.2. La psicología detrás del consumidor

La segmentación de mercado es uno de los pilares de la mercadotecnia moderna, esta estrategia investiga el mercado con el objetivo de encontrar grupos de consumidores similares entre sí y distintos a los demás, se estudian datos como los intereses y aficiones del consumidor o sus valores, buscando individuos que aporten una misma respuesta ante estímulos similares. Una vez se ha obtenido esta información se llevan a cabo planes de marketing personalizados y ajustados a los grupos segmentados⁸.

Debido a la creciente complejidad del consumidor, las segmentaciones clásicas con variables como el nivel socioeconómico o el género son insuficientes para entender los cambios en el mercado. Los criterios tradicionales no recogen el peso de factores como los hábitos diarios,

https://revistas.udc.es/index.php/REDMARKA/article/view/redma.2021.25.1.7848/g7848_pdf; última consulta 1/05/2024).

⁸ Grande, I., “*Dirección de marketing*”, McGraw-Hill, Madrid, 1992.

los intereses o el estilo de vida del comprador, la tecnología y las redes sociales han hecho menos predecibles y más exigentes a los usuarios, por ello, es necesario adoptar un enfoque de segmentación más holístico.

La rama de análisis de comportamiento del consumidor nace con el objetivo de entender su modo de actuar, se fundamenta en diversas teorías, incluyendo la teoría económica, la psicológica, social y la psicoanalítica. Entender la mentalidad de un comprador implica comprender su forma de actuar desde el momento en que surge un problema o nace una necesidad hasta cuando efectúa su compra.

Según la teoría económica, el ser humano tiene deseos ilimitados pero sus recursos; dinero, tiempo y esfuerzo son finitos, de acuerdo a esta teoría, cada vez que un consumidor se encuentra frente a la posibilidad de adquirir un producto o contratar un servicio debe llevar a cabo una evaluación coste-beneficio y determinar qué transacción le proporcione la mayor utilidad posible.

Llevar a cabo este análisis conlleva estudiar costes no monetarios como el tiempo empleado en investigar las opciones y adquirir el producto o el esfuerzo necesario para llevar a cabo una compra, los inconvenientes asociados con la misma o los costes monetarios asociados al precio. En palabras del economista Alfred Marshall, padre del concepto de utilidad marginal, el individuo tratará de adquirir aquel producto que mayor utilidad le proporcione en función del precio que paga por el mismo. Consecuentemente, el consumidor busca maximizar su utilidad mediante una asignación racional de sus limitados recursos entre las opciones disponibles.

A pesar de ello, la realidad muestra cómo esta concepción de individuo plenamente racional no se ajusta al verdadero comportamiento humano. La aversión a las pérdidas, fundamento de la Teoría Prospectiva de Kahneman y Tversky, queda plasmada en el mundo de la

mercadotecnia; los consumidores son mucho más sensibles a las pérdidas que a las ganancias lo que significa que pueden mostrarse reticentes a probar nuevos productos por el riesgo que supone que dichos bienes no cumplan ciertos estándares que sus productos antiguos garantizan, ante la tendencia de tomar decisiones conservadoras, los profesionales del marketing deben diseñar campañas donde la percepción del riesgo quede minimizada y los compradores se sientan seguros con su adquisición.

Asimismo, la toma de decisiones está influenciada por factores ajenos a la razón como son la intuición, la memoria o la percepción, entender la importancia de estos elementos emocionales y cognitivos implica ofrecer servicios que no solo satisfagan las necesidades de los consumidores de manera funcional, sino que también conectan con los usuarios a nivel emocional, de esta forma, es posible generar conexiones profundas con el comprador que se manifiestan mediante la fidelidad a la marca.⁹

La evolución de la mercadotecnia se ha basado en las percepciones cambiantes sobre lo que en cada momento histórico se consideraba esencial; desde una época basada en la producción y venta masiva hasta la concepción actual, centrada en la comprensión de un cada vez menos perceptivo consumidor donde, la anticipación de sus necesidades se posiciona como la clave del éxito.

La teoría psicológica social busca explicar cómo se comporta un individuo en grupo y cómo la presencia de otras personas influye en su forma de pensar y en su toma de decisiones.¹⁰ Esta idea se relaciona con el marketing mediante la Teoría de la Clase Ociosa¹¹ y el concepto de emulación pecuniaria. Según esta teoría, cuando cierta persona consume constantemente y

⁹ Recalde, M.F., Suárez, L., Páez Egúez, J.C., “El comportamiento del consumidor en el ámbito económico y psicológico” en *Revista Conectividad*, vol 2, nº 1, 2020, pp 43-56. (Disponible en <https://revista.ister.edu.ec/ojs/index.php/ISTER/article/view/13/63>; última consulta 2/05/2024).

¹⁰ All port, F. “*Social Psychology*”, Cambridge: Houghton Mifflin, 1968. Revisado el 3 de mayo de 2024.

¹¹ Veblen, T., “*La teoría de la clase ociosa*”, Alianza editorial, 2008. Revisado el 3 de mayo de 2024.

de manera ostentosa, esto se percibe como una muestra de su bienestar económico y consecuentemente, se relaciona el consumismo con riqueza y estatus social.

Bajo esta premisa, surge la emulación pecuniaria; fenómeno por el cual individuos de diversas clases sociales buscan imitar estas prácticas consumistas pues las perciben como selectas y exclusivas. Así nace la creencia de que cuanto mayor sea el precio de un bien o servicio mayor es su calidad y viceversa. Esta concepción puede ser aprovechada por las empresas para resaltar el estatus de sus productos de manera que se perciban como aspiracionales. La técnica del pricing o fijación de precios puede desempeñar un papel nuclear; la adopción de precios elevados puede fortalecer la percepción de calidad y éxito, atrayendo a aquellos consumidores que buscan proyectar su estatus mediante sus elecciones de compra.

Numerosas marcas han forjado su éxito gracias al enfoque aspiracional, para ciertos consumidores la adquisición de un producto ya no se realiza para satisfacer una necesidad funcional sino que buscan adquirir una imagen de prestigio asociada con la marca. Véase como ejemplo de esta dinámica el “fenómeno Birkin”; la prestigiosa firma de lujo Hermès, ha llevado a cabo una estrategia de marketing centrada en la escasez de bolsos Birkin. Para poder adquirir uno de estos bolsos los clientes deben primeramente forjar un vínculo con la marca; esto puede implicar un largo historial de compras como cliente de Hermès o haber efectuado un gasto mínimo, una vez cumplidos estos requisitos, se otorga al cliente la posibilidad de unirse a la lista de espera y eventualmente, adquirir el codiciado bolso.

A primera vista, una estrategia de venta, que dificulta deliberadamente al cliente la adquisición del producto y en muchos casos, limita su capacidad de elección (es el personal de Hermès quien selecciona el bolso que se ofrece al cliente) puede parecer contraintuitiva. Si bien, los datos muestran cómo la variación porcentual acumulada del precio de las acciones

de Hermès ha sido de 34,2% desde febrero de 2023 en comparativa con otras marcas de lujo como LVMH cuya variación porcentual fue de un 6,4% en el mismo periodo.¹²

La estrategia adoptada por la marca francesa no se orienta hacia la masificación de sus productos sino que busca preservar su exclusividad. Hermès es el epítome de lujo, de productos de calidad y estatus hasta el punto que ni siquiera la disponibilidad de recursos financieros garantiza la adquisición de un bolso Birkin. En palabras de Nicole Pollard Bayme: la gente desea aquello que no puede conseguir con facilidad, mediante esta estrategia comercial Hermès ha creado un manual de marketing en el que es la propia marca quien controla la narrativa.

La estrategia de marketing de Hermès se basa en principios fundamentales de la economía, en la teoría de la oferta y la demanda; cuando la oferta es limitada y la demanda es alta la exclusividad y valor percibido de un producto aumenta. Mediante la muy limitada producción de bolsos Birkin, Hermès crea escasez y el atractivo de la marca aumenta.¹³

En definitiva, a la hora de planificar estrategias de marketing se deben entender estos principios, como el ser humano no siempre es racional en sus decisiones, la importancia del factor social y otros elementos subjetivos.

¹² Variación porcentual acumulada del precio de las acciones desde el 23 de febrero de 2023, Datawrapper.

¹³ Berg, M., “ Así ha logrado Hermès convertirse en el símbolo de estatus por excelencia”, Business Insider, 4 marzo 2024. (Disponible en <https://www.businessinsider.es/ha-logrado-hermes-convertirse-simbolo-estatus-excelencia-1368110>; última consulta 3/05/2024).

2.3 Las variables psicográficas

La segmentación psicográfica permite una mayor eficiencia en el marketing mix, hace posible cubrir ciertas lagunas que otros tipos de segmentación no alcanzan a suplir, suponer que consumidores con iguales rasgos demográficos y geográficos comparten los mismos hábitos de consumo implica tener una concepción del marketing incompleta y primitiva.

La segmentación psicográfica permite estudiar al consumidor mediante una visión más subjetiva del mercado, estudiando sus peculiaridades. Si bien no existe una definición cerrada de lo que implica llevar a cabo una segmentación psicográfica, hay ciertos conceptos comunes en las tesis de todo experto en marketing; mediante la psicografía se estudian elementos subjetivos del consumidor como su personalidad, sus intereses, sus actitudes y motivación, sus valores, estilos de vida... en definitiva, aquello que nos permite entender la mente del consumidor¹⁴.

Cuanto más rasgos se estudian sobre el cliente mayor es el grado de precisión de su perfilado, hay que comprender que debido al abultado volumen de datos que se maneja en el mundo del marketing, la optimización de los recursos es esencial para alcanzar objetivos de eficiencia y velocidad. Bajo esta premisa, la psicografía contemporánea ha creado un sistema para agrupar de la manera más certera posible a grandes muestras de consumidores; mediante las variables AIO los expertos en marketing son capaces de llevar a cabo precisos perfiles psicológicos y determinar cómo determinados rasgos psicológicos afectan a los hábitos de consumo¹⁵.

¹⁴ Yanaze, M. H. "Gestión de marketing y comunicación: avances y aplicaciones", ed Sao Paulo: Atlas, 2011, p.196.

¹⁵ Shiffman, L.G., Kanuk, L.L. "*Comportamiento del consumidor*", 9ª edición, Rio de Janeiro, 2012, p. 41. Revisado el 6 de mayo de 2024.

El perfil psicográfico de un consumidor queda determinado por sus respuestas en el cuestionario de Wells y Tigert, el encuestado debe responder a una serie de preguntas indicando su grado de conformidad con ciertas afirmaciones¹⁶, mediante las respuestas aportadas a preguntas sobre las actividades diarias y pasatiempos, información sobre si está inscrito en algún club o a qué evento social acude se busca entender a qué dedica su tiempo el consumidor, cuales son sus intereses como indicador de sus pasiones y qué aspectos de su vida son prioritarios. Finalmente, la percepción de uno mismo así como las opiniones culturales pueden ayudar a determinar cómo aproximarse a un potencial comprador y captar su atención¹⁷. El objetivo de los investigadores es vincular el estilo de vida de los consumidores con sus hábitos de consumo, comprender sus motivaciones y finalmente desarrollar nuevas técnicas de venta más sutiles y efectivas¹⁸.

Los investigadores de la Universidad de Michigan entendieron rápidamente que los compradores no siempre actúan de forma racional en sus hábitos de consumo. Comprendieron que las emociones y valores personales juegan un papel nuclear y por ello desarrollaron el Modelo VALS; Valores Actitudes y Estilo de Vida¹⁹.

Fue la primera teoría psicográfica que incorporaba los valores sociales al parámetro de estudio. El principio detrás de este modelo es que las personas expresan su personalidad por medio de sus comportamientos y esta misma personalidad puede determinar cómo actúan en el mercado. Los valores sociales son la causa que lleva a un individuo a comprar o consumir de cierta manera, siendo este consumo el efecto, la consecuencia.

¹⁶ Wells, W. D., Tigert, D. J., “Activities, interests and opinions” en *Journal of Advertising Research*, 1971. Revisado el 4 de mayo de 2024.

¹⁷ Plummer, J. T., “The Concept and Application of Life Style Segmentation” en *Journal of Marketing*, 1974, pp. 33-37. Revisado el 4 de mayo de 2024.

¹⁸ Davis, J. Schibrowsky, J. Peltier, D., “Interactive Psychographics: Cross-Selling in the Banking Industry” en *Journal of Advertising Research*, vol. 42, nº 2, 2002.

¹⁹ Proviene del inglés “Values and Life Style (VALS) Methodology”.

Aquello que se quiere comprender es la psicología detrás de cada compra, la razón por la que personas con aparentes similitudes socioeconómicas tienen distintas preferencias a la hora de interactuar en el mercado y elegir productos o servicios.

Mediante las respuestas aportadas a un cuestionario, la metodología VALS sitúa a los consumidores en uno de los ocho segmentos. Lo primero que busca el test es entender qué motiva al consumidor, que elementos de sí mismo y del mundo fundamentan su actitud ante la vida, cómo se gobierna y toma decisiones. Los investigadores de SRI International clasificaron en tres categorías las mayores fuentes de motivación; los ideales, los logros y la expresión personal.

Aquellos consumidores cuyos valores impulsan sus vidas tienden a guiarse por elementos como el conocimiento y los principios. De forma paralela, aquellos individuos motivados por los logros y el ascenso social tienden a buscar productos o contratar servicios que muestren a la sociedad su situación de bonanza. Aquellos cuya motivación reside en la expresión personal se encuentran en la constante búsqueda de productos innovadores, servicios que les permitan alejarse de la monotonía y satisfacer sus deseos hedonistas.

Adicionalmente, el modelo VALS cuantifica el nivel de recursos como determinante para la innovación, se diferencian dos tipos de individuos; los que tienen altos y bajos recursos, aquellos con más ingresos tienden a tener una propensión marginal al consumo más alta, es decir, gastan más capital en bienes y servicios. Así mismo, tienen mayor facilidad para acceder al mercado del lujo y son de las primeras personas en adoptar productos o tecnologías punteras e innovadoras que les permita estar a la moda o mantenerse al día con las tendencias del mercado, en definitiva, existe una relación directa entre un poder adquisitivo alto y la innovación. En contraposición, aquellas personas con ingresos medios o bajos tienen una menor capacidad de innovación, su mayor propensión marginal hacia el

ahorro y la menor disponibilidad de fondos los sitúa, por normal general, como consumidores cautelosos y reflexivos con sus decisiones de gasto. Su capacidad de innovación es menor pues tienden a priorizar la satisfacción de necesidades básicas por encima de la innovación, en definitiva, suelen ser más conservadores con sus compras buscando los productos o servicios con mejor relación calidad-precio.

Combinando el factor recursos y motivación Arnold Mitchell consiguió hacer una división de consumidores en cinco perfiles, innovadores, pensadores, creyentes, triunfadores, esforzados, buscadores de experiencias, fabricantes y supervivientes²⁰.

Encontramos un ejemplo práctico del uso de variables psicográficas para motivar decisiones en el escándalo de Cambridge Analytica. Los cinco grandes rasgos de personalidad conocidos como *The Big Five* son patrones de comportamiento que caracterizan a un individuo y están presentes, ya sea en mayor o menor medida, en toda toma de decisiones²¹. Son los siguientes rasgos; apertura entendida como el grado de extroversión de una persona ante nuevas experiencias, conciencia sobre los errores y nivel de perfección que busca una persona en sus acciones, extroversión; mide el grado de sociabilidad, amabilidad entendida como los diferentes niveles de compasión y respeto por los demás y finalmente neuroticismo, mide el grado de facilidad con el que una persona se enfada y cambia su estado de ánimo. Los expertos de Cambridge Analytica emplearon estas variables, conocidas por su acrónimo en inglés como variables OCEAN (openness, conscientiousness, extroversion, agreeableness, neuroticism) para estudiar la personalidad de los usuarios de Facebook. Seis millones de usuarios completaron de manera voluntaria test psicométricos colaborando así con los investigadores. Mediante las respuestas aportadas se llevaron a cabo los perfiles

²⁰ Urbonavičius, S., Kasnauskienė, G., “New Applications of a Traditional Psychographic Segmentation Concept” en *Commerce of Engineering Decisions*, nº 5, 2005, pp 80-86. Recuperado el 10 de mayo de 2024.

²¹ Soto, C. J. “Big Five personality traits” en M. H. Bornstein, M. E. Arterberry, K. L. Fingerman, & J. E. Lansford (Eds.), *The SAGE encyclopedia of lifespan human development*, Thousand Oaks, CA: Sage, 2018, pp 240-241. (Disponible en https://www.researchgate.net/publication/324115204_Big_Five_personality_traits última revisión 12/05/2024)

psicográficos, adicionalmente, existía la posibilidad de compartir el perfil de Facebook, tener acceso a las publicaciones de un usuario, información sobre sus likes y sobre las imágenes que compartía facilitó a los investigadores llevar a cabo una comparativa entre la predicción y la realidad y consecuentemente mejorar de forma sustancial las correlaciones²². Ejemplos de las predicciones que dedujeron fueron; que los usuarios seguidores de Lady Gaga tendían a ser extrovertidos o que aquellas personas con un alto nivel de estabilidad emocional tendían a preocuparse menos por su futura situación económica y consecuentemente eran más propensos a oponerse a políticas de redistribución de riqueza. Este caso ejemplifica cómo gracias a la psicografía es posible comprender motivaciones y comportamientos subyacentes en los consumidores.

El psicometrista y psicólogo computacional Michael Konsinski, profesor de comportamiento Organizacional en la Universidad de Stanford consiguió predecir con una media de 68 “likes” la ideología, inteligencia y afiliación religiosa de un usuario de Facebook ²³. Generar predicciones tan detalladas sobre los individuos con datos, en apariencia, tan nimios como una publicación o un me gusta permitieron a los trabajadores de Cambridge Analytica crear propaganda política que apelara a los miedos o preocupaciones de cada individuo y realmente causar un impacto. Este caso no solo debe estudiarse como un modelo de éxito en el ámbito de la optimización y segmentación publicitaria sino también como una lección ética de la injerencia mediante perfilado psicológico en situaciones como unas elecciones electorales y las implicaciones éticas que puede traer consigo.

En definitiva, la psicografía es una herramienta enormemente útil para comprender las motivaciones del consumidor así como los elementos que guían su comportamiento.

²² Rehman, Ikhlaq ur, "Facebook-Cambridge Analytica data harvesting: What you need to know" en *Library Philosophy and Practice* (e-journal), 2019. (Disponible en <https://digitalcommons.unl.edu/libphilprac/2497> última revisión 12/05/2024).

²³ Kosinski, M., Stillwell, D., Graepel, T., "Private traits and attributes are predictable from digital records of human behavior", *University of California Berkeley*, 11 de marzo de 2013, (Disponible en <https://www.pnas.org/doi/full/10.1073/pnas.1218772110> ; última revisión 12/05/2024).

Mediante el uso de variables psicográficas es posible desentrañar la faceta más irracional de una persona, aquella intrínseca que a menudo es la responsable de patrones de compra y hábitos de consumo implícitos. Entender y saber emplear esta herramienta permite a los especialistas en marketing obtener un visión más completa de su consumidor potencial, hace posible entenderlo más allá de los datos objetivos, revelando sus verdaderos intereses, actitudes, valores y motivaciones.

Una vez perfilado el comportamiento del consumidor, este puede segmentarse en grupos de individuos con tendencias de compra similares. Este targeting permite a los expertos en marketing desarrollar campañas altamente personalizadas, que apelen a los intereses concretos de estos grupos. Al diferenciar a los potenciales consumidores y comprender su comportamiento las empresas tienen más información para acercarse a los intereses y necesidades concretas de sus clientes y alinear sus estrategias de marketing con su audiencia, optimizando de esta forma su presupuesto publicitario.

En definitiva, la psicografía en grandes volúmenes de datos permite a las empresas diseñar y ejecutar campañas de marketing más eficientes. En la sociedad de la inmediatez y la competencia es preferible dirigirse a un público más concreto y adecuadamente segmentado que vaya a responder de manera proactiva, pues es más probable que el mensaje cale y motive la acción de consumo deseada.

Finalmente, el éxito de una campaña publicitaria no se mide simplemente por la cantidad de visualizaciones o por su alcance bruto, el verdadero indicador de efectividad es aquel que prueba que la acción ha conseguido conectar con el público y que este interactúa con la marca, en definitiva, mediante una estrategia de marketing bien segmentada y dirigida no solo maximiza el presupuesto sino que asegura un mayor retorno en la inversión poniendo el foco en aquellos consumidores que van a responder de la forma más positiva.

3. Técnicas de Big Data al servicio del perfilado psicológico

3.1 La obtención de datos

Para llevar a cabo análisis psicográficos y perfilados de clientes es necesario recabar y analizar una gran cantidad de información. La obtención de dichos datos es un asunto enormemente complejo y controversial merecedor de su propio trabajo, si bien, considero esencial hacer una serie de apuntes al respecto, pues todas las técnicas que se mencionan a continuación llevan consigo el uso de dichos datos.

La regulación sobre la protección de datos es una materia cada vez más presente en el marco jurídico europeo, el RGPD entiende la protección de datos como un derecho fundamental y consecuentemente, directamente exigible a nivel europeo²⁴. El artículo 9 del Reglamento regula el tratamiento de categorías especiales de datos personales y prohíbe, con ciertas excepciones, el tratamiento de datos que puedan revelar información sensible de un individuo como es su ideología, convicción religiosa u orientación sexual. Adicionalmente, en el artículo 22 se recoge la elaboración de perfiles, con el consentimiento explícito del interesado es posible llevar a cabo perfilados, si bien, es crucial respetar sus derechos, libertades e intereses legítimos²⁵. Las instituciones europeas son plenamente conscientes de la manipulación existente y por ello, son firmes defensoras del principio de transparencia sobre quién diseña las categorías que segmentan a los consumidores, su significado y el nivel de importancia de dichas categorías²⁶.

²⁴ Pérez Miras, J., “La nueva protección de datos en Europa. El RGPD como nuevo marco jurídico europeo de Privacidad” en *Crónicas Jurídicas*. Disponible en <https://ws168.juntadeandalucia.es/iaap/revista/2018/10/21/la-nueva-proteccion-de-datos-en-europa-el-rgpd-como-nuevo-marco-juridico-europeo-de-privacidad/>; última revisión 20/05/2024).

²⁵ Garriga Domínguez, A., “Reflexiones sobre el marketing político y el fenómeno de la desinformación en el contexto electoral” en *AFD*, vol. XXXVI, 2020, pp 252-287. (Disponible en https://www.boe.es/biblioteca_juridica/anuarios_derecho/abrir_pdf.php?id=ANU-F-2020-10025100287; última revisión 20/05/2024).

²⁶ Lyon, D., “Surveillance Studies. An overview” en *Canadian Journal of Sociology*, 2008, p.186. (Disponible en https://www.researchgate.net/publication/26526923_David_Lyon_Surveillance_Studies_An_Overview; última revisión 20/05/2024).

A pesar de los esfuerzos europeos por proteger a sus consumidores, autoridades extranjeras como la estadounidense carecen de una regulación sólida en materia de protección de datos. Consecuentemente, durante el proceso de transferencia de datos internacionales, la seguridad de dicha información queda gravemente comprometida.

La vulnerabilidad del consumidor se puede apreciar desde múltiples prismas; estudios muestran cómo muchos usuarios no entienden realmente los términos y condiciones a los que consienten cuando ceden sus datos personales. Como consecuencia de ello, aportan información sensible sin comprender realmente las implicaciones que esto conlleva²⁷. Adicionalmente, muchos usuarios no son capaces de apreciar el valor de los datos que ceden ni las implicaciones asociadas a un uso indebido. Aunque los datos cedidos pueden parecer inofensivos de manera aislada, la combinación de todos ellos y un buen análisis psicográfico puede llevar a la creación de perfiles enormemente personalizados capaces de predecir las acciones de un usuario.

Por otra parte, la práctica empresarial de “recoger (datos) primero, preguntar después” exacerba la situación²⁸. Miles de datos personales se recogen de manera indiscriminada sin detallar el propósito de su utilización. Esta técnica expone a los consumidores a situaciones de manipulación, gracias a la información obtenida y las conclusiones recabadas las compañías son capaces de influir en las decisiones de compra de los individuos haciendo uso de sus vulnerabilidades.

En definitiva, es crucial recabar datos personales bajo un enfoque ético y de transparencia donde los consumidores reciban educación sobre la información cedida y sus implicaciones

²⁷ Hoofnagle, C., “Federal Trade Commission Privacy Law and Policy”, Cambridge University Press, 2016, p. 148. Recuperado el 22 de mayo de 2024.

²⁸ Hartzog, W., “Privacy’s Blueprint: The battle to control the design of New Technologies”, Harvard University Press, 2018. Recuperado el 22 de mayo de 2024.

de manera que la decisión de compartir datos se tome de manera consciente y libre²⁹. Adicionalmente, las compañías deben actuar con responsabilidad al tratar información sensible de sus consumidores, mediante la implementación de robustas medidas de seguridad y estrictos protocolos de confidencialidad será posible llevar a cabo un desarrollo ético del Big Data.

3.2 Técnicas más relevantes

Las técnicas de Big Data al Servicio del Perfilado Psicológico combinan la recolección de cantidades masivas de datos, un análisis avanzado y modelos predictivos de Machine Learning.

3.2.1 Minería de datos

El text mining o la minería de datos es un campo que permite a las compañías transformar datos en conocimiento. La información a analizar se encuentra en formato texto, esto es un tipo de dato no estructurado, para poder ser legible por un ordenador es necesario un procesamiento previo, de manera que se obtenga un código que pueda ser utilizado por un algoritmo de Machine Learning.

El tratamiento de textos no está exento de retos; fallos gramaticales, abreviaciones o los sinónimos pueden dificultar la interpretación por parte del algoritmo. Por ejemplo, palabras con faltas de ortografía o abreviadas, elementos típicos de las redes sociales, pueden generar confusiones a la hora de analizar un texto. Otro elemento significativo es la existencia de jerga informal, la lengua y las formas de expresión evolucionan constantemente por lo que es crucial llevar a cabo un entrenamiento constante del algoritmo de forma que pueda adaptarse

²⁹ McSweeney, T. "Psychographics, predictive analytics, artificial intelligence & Bots: Is the FTC Keeping Pace?" en *Georgetown Law Technology Review*, 2018, pp 515-530. (Disponible en <https://georgetownlawtechreview.org/wp-content/uploads/2018/07/2.2-McSweeney-pp-514-30.pdf>; última revisión 22/05/2024).

a los nuevos términos del lenguaje y así mantener un análisis preciso. El contexto también dificulta la minería de datos, una misma palabra puede tener diferentes significados según el contexto en el que se emplea, es necesario que un modelo de Machine Learning pueda evaluar el contexto para predecir de forma correcta. Por todas estas razones, el paso previo de preprocesamiento resulta nuclear, en esta fase se garantiza la estructuración de datos optimizando su utilidad para los futuros análisis.

La representación de documentos se lleva a cabo mediante una matriz llamada Document Term Matrix (DTM), las variables que la conforman son los documentos en las filas y la frecuencia con la que aparece un término en el documento compondrá las columnas³⁰.

Mediante “Bags of Words” se pueden construir las variables de forma sencilla, simplificando la representación de textos y omitiendo factores como el orden de las palabras en un texto o su corrección gramatical. Cada documento constituye una bolsa de palabras, una vez se fija un diccionario de vocabulario se estudia la frecuencia con la que estas palabras aparecen en las diferentes bolsas de palabras, es decir, en los diferentes textos que conforman el corpus. Si la palabra se encuentra en un documento se le asigna el valor “1”, en caso contrario, el valor “0”, esta combinación binaria permite representar de manera simple la estructura de datos. Este enfoque puede resultar incompleto debido a la pérdida de información que genera, para remediar esta situación, suele emplearse conjuntamente con la Representación de la Frecuencia de Términos; con este sistema se asigna un valor numérico a la frecuencia con la que una palabra aparece en un documento, al igual que la Bolsa de Valores, esta medida

³⁰ Qi, Q., Hessen.D.J, Deoskar.T., Van der Heijden.P., “A comparison of latent semantic analysis and correspondence analysis of document-term matrices” en *Cambridge University Press*, 2023. (Disponible en <https://www.cambridge.org/core/journals/natural-language-engineering/article/comparison-of-latent-semantic-analysis-and-correspondence-analysis-of-documentterm-matrices/96316CBB94CD9DD681CB50D9F52E0201>; última revisión 24/05/2024).

busca crear patrones discriminado o infravalorando los términos más inusuales en un corpus de documentos³¹.

Para simplificar al máximo el pre-procesamiento de palabras y el volumen de tokens, entendidos como las palabras que aparecen en la matriz, se llevan a cabo una serie de acciones que reducen el texto; pasar todos los términos a minúsculas, el stemming; eliminación de sufijos de ciertas palabras manteniendo únicamente su raíz común. Eliminar palabras comunes que no aportan valor al documento como son “y, o, si...” simplifica el texto. La lematización, es decir, reducir a palabras comunes verbos conjugados o palabras que surgen de otras, así como la eliminación de espacios o signos de puntuación garantizarán la reducción del texto a sus elementos nucleares, aquellos que le aportan sentido y valor analítico³².

Otra alternativa al método Bag Of Words o al sistema de Term Frequency Representation es el Term-Frequency-Inverse Document Frequency, también conocido como TFIDF. Permite extraer palabras temáticas, vinculadas con los documentos, mediante un diccionario de palabras claves. Su fórmula está compuesta por dos elementos; la Frecuencia del Término o Term Frequency (TF) representa la relevancia de un término dentro de un documento o texto, así mismo, la Frecuencia del Documento, Document Frequency (IDF) aporta una visión más general pues mide en cuantos documentos de un corpus aparece un término, es una medida de dispersión que nos permite conocer cómo de común o particular es un término dentro del corpus, el valor de esta métrica reside en su capacidad para determinar la rareza de una

³¹ Rojas Rubio, L, Meneses Villegas, C, “Una comparación empírica de algoritmos de aprendizaje automático versus aprendizaje profundo para la detección de noticias falsas en redes sociales” en *Ingeniare, Revista Chilena de Ingeniería*, vol 30, n°2, 2022. (Disponible en https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-33052022000200403; última revisión 27/05/2024).

³² Redin, L., “Machine Learning III: Text Mining (parte 1) Nociones básicas”, Universidad Pontificia de Comillas, 2023. Recuperado el 27 de mayo de 2024.

palabra³³. Una DF alta puede deberse al uso de “stopwords” que realmente no aportan significado al texto. Por el contrario, una Frecuencia de Documento baja implica que una palabra no aparece en muchos textos, esto puede deberse a su especificidad o a ser una palabra muy concreta empleada en un ámbito concreto.

Con todo esto, al realizar un análisis TFIDF buscamos una palabra con un alto TF, es decir, que se encuentre muy presente en un texto, pero con inversa DF, es decir, que sea una palabra rara en el corpus solo presente en ciertos documentos. El uso de la inversa se lleva a cabo para descartar palabras comunes (reciben un bajo valor IDF) y resaltar palabras distintivas.

3.2.2. Análisis de Sentimiento

Un usuario de redes sociales es bombardeado diariamente con miles de anuncios, se busca que esta publicidad genere una emoción que lleve a los consumidores a compartir su experiencia. Debido a la viralidad de la publicidad, los expertos en marketing encuentran dificultades a la hora de ejercer un control directo del impacto de la publicidad y la reacción del consumidor, no hay herramientas que permitan medir las emociones que siente un usuario al ver su campaña publicitaria³⁴. Si bien, existe una manera de tasar la respuesta cognitiva; mediante la voluntad del usuario de compartir la publicidad, los comentarios que deja en la publicación o el tiempo dedicado a ver el anuncio es posible analizar la respuesta del consumidor a la campaña. Áreas como la psicolingüística y el procesamiento del lenguaje natural (PNL) han traído consigo herramientas como el Sentiment Analysis que puede medir la fuerza de las opiniones del consumidor expresadas en las redes³⁵.

³³ Kim.S., Gil.J., “Research paper classification systems based on TF-IDF and LDA schemes” en *Human-centric Computing and Information Sciences*, 2019. (Disponible en <https://doi.org/10.1186/s13673-019-0192-7>; última revisión 26/05/2024).

³⁴ Huang, J., Su, S, Zhou, L., Liu,X., “Attitude Toward the Viral Ad: Expanding Traditional Advertising Models to Interactive Advertising” en *Journal of Interactive Marketing*, vol 25, 2013, pp 36-46.(Disponible en <https://www.sciencedirect.com/science/article/abs/pii/S1094996812000333>; última revisión 23/05/2024).

³⁵ Kulkarni, K., Kalro.A, Sharma, D., Sharma,P., “A typology of viral ad sharers using sentiment analysis” en *Journal of Retailing and Consumer Services*, vol 53, 2020.(Disponible en <https://www.sciencedirect.com/science/article/abs/pii/S0969698918307902>; última revisión 23/05/2024).

El primer paso para llevar a cabo un análisis de sentimiento es la recolección de información. En el mundo del marketing, las fuentes principales de información en internet son las redes sociales y las páginas de e-commerce. Plataformas como Instagram o Tiktok proporcionan una gran base de datos sobre cómo los consumidores interactúan con los productos o servicios. Mediante comentarios, likes y compartidos se puede observar el uso real de un producto, las circunstancias en las que se emplea y el nivel de satisfacción de los usuarios. La frecuencia con la que un producto es exhibido o las palabras empleadas para describirlo pueden ser importantes indicadores de la percepción del consumidor.

Un elemento clave de las redes sociales es su espontaneidad, muchos usuarios las consideran un entorno seguro donde expresar opiniones y deseos. Uno de los beneficios de utilizar este tipo de bases de datos es la autenticidad de la información recopilada. En otras plataformas más formales, como las encuestas de satisfacción llevadas a cabo por profesionales, los datos obtenidos pueden ser menos sinceros y la información aportada políticamente correcta debido a la timidez de los encuestados.

Mediante un análisis de sentimiento podemos clasificar de forma masiva las opiniones, emociones o actitudes de los usuarios según su connotación, esta se puede extraer del lenguaje empleado y pueden tener polaridad positiva, negativa o neutra³⁶. Este método puede ser empleado en numerosos sectores y campos de conocimiento, en el ámbito del marketing es especialmente interesante pues puede ayudar a una marca a comprender la percepción que tienen sus consumidores o en el ámbito de las campañas publicitarias optimiza la elección de palabras que suenan más positivas y así generar un mayor impacto.

³⁶ Pang, B., Lee. L. “Minería de opiniones y análisis de sentimientos” en *Fundamentos y tendencias en la recuperación de información* 2008, pp 1-135. Recuperado el 26 de mayo de 2024.

Para llevar a cabo un análisis de sentimiento se pueden emplear varios enfoques; el primero, consiste en emplear un diccionario Lexicón; se trata de un recurso lingüístico construido de manera automática empleando recursos de internet, algunos son de acceso público como SentiWordnet o NTU, los conocimientos del lexicón se usan en el programa de entrenamiento. Un análisis de sentimiento puede centrarse en la polaridad de las palabras o en las emociones. El diccionario AFINN asigna a cada palabra un valor entre el rango -5 a 5; el -5 indica el mayor sentimiento negativo, por el contrario un valor 5 se entiende como un sentimiento enormemente positivo.³⁷

3.2.3 Técnica de Clustering

La técnica de clustering es un método de aprendizaje no supervisado con el cual se puede estudiar la existencia de estructuras en un conjunto de datos. Las observaciones de un dataset con características comunes se engloban en un mismo subgrupo denominado cluster. Los clusters están formados por observaciones similares entre sí, esta similitud se mide mediante la distancia, dos observaciones con una distancia baja se parecen entre sí y por lo tanto pueden estar englobadas en un mismo grupo.

Las técnicas de clustering son múltiples y diversas; si se parte de un estudio individualizado de cada observación entendida como un cluster y se van uniendo distintas observaciones en función de sus similitudes hasta que finalmente se obtiene un único cluster se está llevando a cabo un proceso de clustering aglomerativo. Por el contrario; si se parte de un clúster único compuesto por todo el dataset y progresivamente se divide clusters más pequeños nos encontramos ante el método divisivo. Ambos son sistemas de clustering jerárquico y comparten la característica de no tener un número de clusters preestablecido, es

³⁷ Redin, L., "Machine Learning III: Text Mining (parte 2) Noción", Universidad Pontificia de Comillas, 2023. Recuperado el 27 de mayo de 2024.

el algoritmo quien determina el número óptimo de clusters basados en criterios de similitud o distancia.

El clustering particional se lleva a cabo mediante algoritmos de K-means, K-medoids... A diferencia del clustering jerárquico, las observaciones en K-means se distribuyen en un número de clusters predefinidos. El clustering jerárquico que es más exploratorio, pero, el clustering particional es más adecuado cuando se necesitan grupos concretos desde el inicio, en el mundo del marketing esto es esencial pues permite la segmentación de clientes en grupos concretos optimizando así el uso de los recursos limitados. Según el presupuesto y las posibilidades temporales se podrá hacer una labor de clustering más o menos intensiva, de manera que se creen campañas específicas para cada subgrupo de clientes, por todo ello, el algoritmo que emplearemos es el K-means, el cual, busca minimizar la dispersión dentro de cada cluster formando los grupos más homogéneos posibles.

Una idea clave a la hora de hacer K-means es la dispersión de un cluster, esto es, la distancia de las observaciones que componen un mismo grupo. Lo que se busca medir es como de compactas son las observaciones, cuánto se parecen entre sí, cuanta menor dispersión más homogéneo es el cluster y más parejas sus observaciones, el algoritmo K-means busca que la dispersión total dentro de los clusters sea lo más pequeña posible.³⁸

Responder a la pregunta de en cuántos grupos puedo segmentar a mi clientela de manera que sea lo más representativo posible es clave; encontrar el número óptimo de clusters es un proceso crucial y por ello se recurre a los métodos de la silueta y al método del codo³⁹.

³⁸ Redin, L., "Machine Learning III: Técnicas de Clustering", Universidad Pontificia de Comillas, 2023. Recuperado el 2 de junio de 2024.

³⁹ Provost, F., & Fawcett, T. "Data Science for Business: What you need to know about data mining and data-analytic thinking". O'Reilly Media, Inc, 2013. Recuperado el 2 de junio de 2024.

El sistema de la silueta busca maximizar la media de los índices de silueta de todas las observaciones. Recogiendo valores en un espectro de entre -1 y +1 el índice mide cómo de precisa es la asignación de una observación a un cluster concreto, se compara con el resto de observaciones de ese mismo cluster así como con la similitud de las observaciones en otros grupos. Cuanto más cercano sea el valor a +1, mejor es la asignación, si es negativa la observación está asignada a un cluster totalmente incompatible con ella y si es igual a cero, dicha observación tiene similitudes con observaciones de dos clusters, pudiendo pertenecer a cualquiera de ellos.

De forma paralela, mediante el método del codo se definen los clusters buscando que la variación total intra-cluster sea lo menor posible, dicha variación queda plasmada en una gráfica donde se percibe un cluster que, si se añade uno adicional no mejora sustancialmente la variación intracluster, este cluster previo el óptimo número de subgrupos.

4. Caso práctico: Always Boutique

Always Boutique es una pequeña boutique de ropa online, durante los últimos 5 años ha logrado expandir su presencia en el mercado global ofreciendo una amplia variedad de prendas a su diversa clientela. Durante este periodo la boutique ha recopilado una gran cantidad de datos sobre sus clientes, sus opiniones y hábitos de consumo. Always Boutique ve una oportunidad única para llevar a cabo avanzadas técnicas de análisis de datos y así optimizar sus campañas publicitarias y estrategias de marketing. Todo ello se llevará a cabo empleando técnicas de análisis de sentimiento y clustering.

4.1 Análisis de Reseñas y Sentimientos

En primer lugar, se quiere entender la situación actual, qué percepción tienen los clientes con los servicios prestados y las prendas adquiridas. El dataset aportado consta de las siguientes variables:

- **Id:** Código identificativo de cada reseña
- **Clothing ID:** Código único de cada prenda de ropa, sirve para identificarla

- **Age:** Edad del cliente que escribe la reseña
- **Title:** Resumen del comentario aportado
- **Review Text:** Reseña o valoración sobre el producto
- **Rating:** Valoración del producto, se emplea una escala del 1 al 5, donde 5 indica el mayor nivel de satisfacción.
- **Recommended IND:** Valor binario sobre si el cliente recomendaría el producto o no. 0= o recomendable, 1= recomendado.
- **Positive feedback count:** Número de comentarios positivos que ha recibido una reseña, mide su utilidad para otros clientes.
- **Division Name:** Nombre de la división en la que se encuentra la prenda, puede estar en el área central, general petite o intimate.
- **Department Name:** Indica el departamento en la página web donde se puede encontrar un artículo, las opciones son; dresses, tops, bottoms e intimate.
- **Class Name:** Tipo de prenda adquirida y por la cual se escribe la reseña. Se han evaluado las siguientes prendas;dresses,knits,fine gauge, outerwear, sleep, sweaters, jeans,skirts,lounge,pants, blouses y shorts.

Al importar el dataset observamos que se trata de un archivo con 28.207 observaciones y que en algunos casos no existía contenido para cierta variable, por ello, el primer paso fue limpiar el dataset. Tras realizar este proceso preparatorio; el tamaño del dataset disminuyó a 4.257 observaciones y 11 variables.

Para llevar a cabo el análisis de sentimiento hay que descargar una serie de bibliotecas que permiten el procesamiento de los datos, su interpretación y representación. Se resalta la biblioteca “TidyText” pues contiene el diccionario Afinn, este léxico de sentimientos está compuesto por un lista de términos en inglés clasificados según su polaridad en un espectro de números enteros entre -5 (negativo) y +5 (positivo), a una palabra con connotaciones negativas se le asigna un valor negativo, por el contrario, si una palabra es percibida de manera positiva tendrá un valor positivo en el diccionario Afinn. Véase un ejemplo; el término “happy”, feliz, tiene una puntuación de +3, por el contrario, una palabra como enfermo, “sick” tiene una connotación negativa de -2.

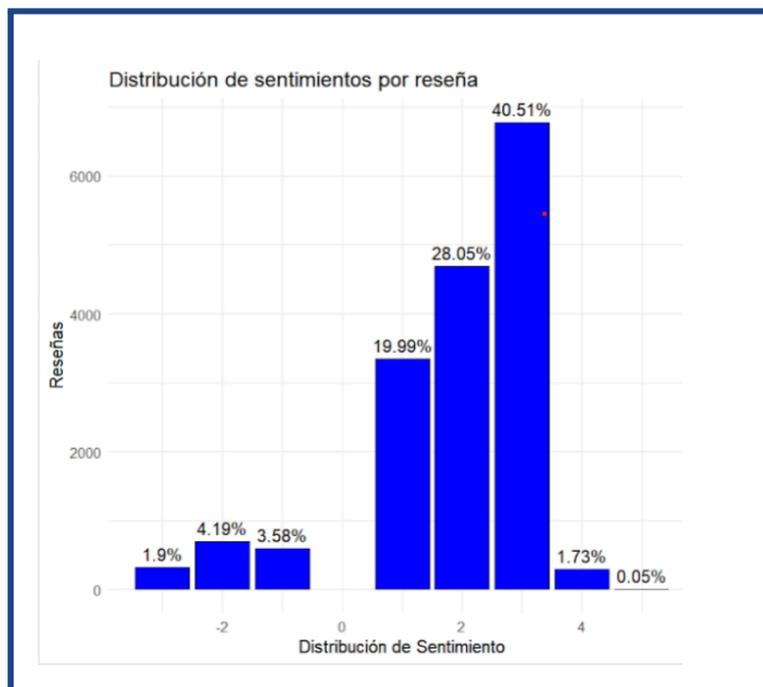
La columna empleada para efectuar el análisis de sentimientos será “Review Text”, en ella, aparece el mensaje completo de la reseña. La razón por la que preferimos emplear esta columna en vez de “Title” es para evitar perder matices. Hay reseñas que no quedan correctamente resumidas en “Title” o pierden valor informativo. Esta situación queda evidenciada con la siguiente reseña; el título es “Not a fan”, en ella, un usuario describe como la prenda adquirida no se ajusta a las características publicitadas ni en la talla ni en composición. El método sentiment analysis no estudia el contexto de una reseña sino cada una de sus palabras, el término “fan” tiene una evaluación de +3, al ser un diccionario basado en unigramas no se capta la negación del sentimiento positivo.

El siguiente paso es el preprocesamiento de la columna “Review Text”, está compuesta por textos completos con palabras redundantes, faltas de ortografía, jerga de internet y abreviaciones. En primer lugar se tokeniza el texto dividiéndolo en unidades más pequeñas, en palabras. La nueva base de datos está compuesta por entorno a 170 mil observaciones, hay múltiples términos que no aportan valor analítico como por ejemplo las palabras “the”, “is” o “i”. Todos estos términos se conocen como stopwords o palabras vacías, se eliminarán para reducir el ruido, tras ello, el listado de observaciones queda reducido a 82 mil, prácticamente a más de la mitad.

Adicionalmente, se llevan a cabo procesos de stemming y lematización de manera que agrupamos las variantes morfológicas de una palabra para tratarla como una única entidad, un ejemplo de stemming es la reducción de la palabra “extremely” a “extrem”. Lematizar las observaciones aporta mayor precisión y legibilidad al análisis, una muestra de lematización es la reducción de la palabra “well” a “good”.

Finalmente, se unen los datos tokenizados, en concreto la columna “lematización” con el diccionario Afinn y se calcula el sentimiento representado en cada reseña.

Para comprender los resultados obtenidos se llevan a cabo una serie de representaciones, en primer lugar el siguiente histograma:

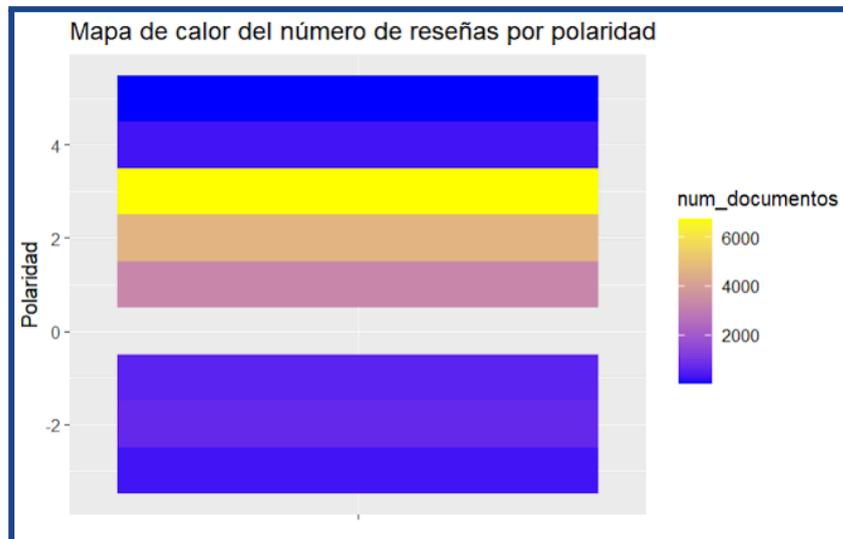


La satisfacción con Always Boutique queda plasmada en sus reseñas, el 90,33% se encuentran en las categorías del 1 al 4. Existe un pico en el nivel de sentimiento 3, se denota que el 40,51% de los clientes incluyen en sus reseñas términos como “love”, “perfect” o “beautiful”. Podríamos entender que el 19,99% de las reseñas en el nivel 1 equivale a valoraciones positivas moderadas, es decir experiencias de compra que cumplen las expectativas básicas de calidad, similitud entre lo publicitado y recibido, puntualidad en la entrega...

El 9,67% de reseñas con sentimientos negativos se encuentran distribuidas en los niveles de -1 a -3. Los niveles más repetidos entre los clientes insatisfechos son los valores -1 y -2, esto sugiere la existencia de problemas menores que, aunque no afectan de forma crítica, si que corrompen la experiencia de compra y alejan a estos consumidores de Always Boutique.

En definitiva, a pesar de la experiencia positiva de la mayoría de los consumidores la boutique no debe infravalorar los comentarios negativos, poner una solución a los problemas en ellos descritos fidelizará a clientes insatisfechos que ven como la Always Boutique se preocupa por sus necesidades y representa una oportunidad de mejora.

La frecuencia con la que se distribuyen las reseñas en el mapa de calor permite identificar una tendencia de consumo generalmente satisfactoria. La mayor concentración de reseñas se encuentran en el nivel 3, es decir, los usuarios tienen una opinión moderadamente positiva. Los extremos de polaridad; personas plétóricas con su compra o enormemente decepcionadas son escasas. Esta distribución es consistente con una distribución normal donde los valores se agrupan con menor frecuencia cuando más se alejan de la media.



Gracias a una nube de palabras se pueden analizar las tendencias de comportamiento de los clientes hora de escribir una reseña. Adjetivos como “comfortable”, “perfect”, “super”, “recommend”, “pretty” o “cute” son muy comunes y justifican ese alto número de reseñas en el nivel 3.

Respecto a los sentimientos negativos, la nube de palabras sugiere que el motivo principal detrás de las reseñas negativas está relacionado con cuestiones de tallaje. Encontramos con frecuencia términos como “boxy”, “wide”, “huge” o “runs” (usado en la frase “runs big/small” para indicar que la prenda es más grande o pequeña de lo anunciado). Con esto, podemos deducir que un punto de mejora en Always Boutique es la precisión con la que se describen las tallas y el patronaje, si los usuarios reciben información certera sobre si una prenda se ajusta a sus medidas o como sienta en el cuerpo es más probable que se encuentran satisfechos con su compra y se reduzcan las críticas negativas.

4.2 Clustering psicográfico y estrategias de marketing

Las variables del apartado anterior nos aportan información objetiva sobre nuestro consumidor típico; individuos con una edad media de 43 años que adquieren diferentes tipos de prendas. A pesar de la objetividad de estas variables, no aportan suficiente información personal para entender la motivación de nuestros clientes. Por ello, se llevará a cabo una segmentación psicográfica utilizando una base de datos que contiene las puntuaciones en el test OCEAN (Openness, Conscientiousness, Extraversion, Agreeableness y Neuroticism) de 1.834 clientes. Este clustering permitirá al equipo de marketing entender y emparejar a clientes con motivaciones similares y con ello personalizar la comunicación de cada grupo.

En primer lugar; las variables que componen el dataset son la siguientes:

Openness; esta variable hace referencia al nivel de apertura que tiene una persona hacia nuevas experiencias, se asocia con el interés por conocer nuevas ideas o puntos de vista totalmente antagonistas. Esta variable suele estar relacionada con la originalidad, la complejidad y el afán de conocimiento. El comportamiento opuesto a la apertura es el convencionalismo.

Conscientiousness; La conciencia hace referencia al comportamiento dirigido a alcanzar metas y conseguir objetivos, un alto nivel de conciencia se relaciona con ideas como la autoexigencia, la disciplina y el orden. La escala diseñada por Judge, Rodell, Klinger, Simon y Crawford posiciona en un polo a las personas desorganizadas, cuyo valor de conciencia será bajo y en el otro extremo a los individuos organizados quienes suelen tener una alta puntuación de conciencia⁴⁰.

Extraversion; Hace referencia a la actitud con la que una persona se asoma al mundo, mide elementos como la sociabilidad o la asertividad. El antónimo conductual es el gregarismo o la mansedumbre.

Agreeableness; La amabilidad se relaciona con los rasgos de franqueza, confianza, sinceridad y modestia. Una persona con alto nivel de amabilidad se caracteriza por tener una

⁴⁰ Judge, Timothy A., Jessica B. Rodell, Ryan L. Klinger, Lauren S. Simon, and Eean R. Crawford, "Hierarchical Representations of the Five-Factor Model of Personality in Predicting Job Performance: Integrating Three Organizing Frameworks with Two Theoretical Perspectives" en *Journal of Applied Psychology* vol 98 n°6, 2013, pp 875–925. (Disponible en <https://doi.org/10.1037/a0033901>, última revisión 15/06/2024).

visión social y comunitaria, por mostrar interés por el resto de personas. Lo contrario a la amabilidad son las tendencias individualistas y egoístas.

Emotional Stability; Equivale a la variable “neuroticism” , mide la capacidad de una persona para manejar situaciones estresantes. Una persona con un nivel de estabilidad emocional baja tiende a ser impulsiva en la toma de decisiones, tiene un comportamiento cambiante y difícil de predecir.

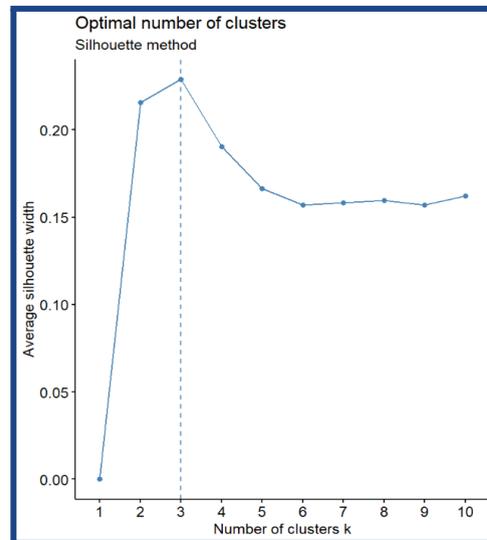
Un total de 1.834 clientes realizaron un test de personalidad basado en las variables OCEAN, respondieron una serie de preguntas que evaluaban elementos de su personalidad. La puntuación de cada variable se encuentra en una escala del 1 al 7, donde una puntuación igual o cercana a 1 denota un nivel bajo en esa característica, por el contrario, valores cercanos a 7 indican que el individuo presenta varias características típicas de una determinada dimensión de la personalidad. Véase un ejemplo; un cliente ha obtenido una puntuación (2) en agreeableness, (3) en emotional stability y (6.5) en extraversion, esta persona es probable que sea extrovertida y sociable, que disfrute de la interacción social y la compañía. Así mismo, su bajo nivel de amabilidad puede revelar una tendencia hacia la competitividad, puede preferir trabajar de manera individual, valorando sus opiniones por encima de las del resto de personas. Finalmente, tiene una estabilidad moderada, esto implica un manejo correcto de situaciones de estrés y situaciones negativas, si bien, dado el valor intermedio de esta variable esta persona puede sufrir fluctuaciones emocionales.

Una vez comprendido el significado detrás de los datos, se llevó a cabo un proceso de clustering. El objetivo del departamento de marketing era mejorar la relación con los clientes mediante la implementación de campañas publicitarias enfocadas a determinados grupos de personas. Se decidió que la división de clientes en tres grupos era lo más adecuado. Se formaron tres clusters con 566, 662 y 606 individuos respectivamente, con ello, se logró una distribución relativamente equilibrada entre los grupos.

Para determinar si se ha escogido el número óptimo de clusters se llevan a cabo varios análisis; en primer lugar el método del codo, cuanto menor sean los valores de la suma de cuadrados dentro de los clusters (Within-Cluster Sum of Squares) menos distancia hay entre cada observación y el centroide y consecuentemente más homogéneo es un cluster. Al formar más y más subgrupos el valor del WSS va disminuyendo, el valor de WSS con un único cluster era de 16197.884 mientras que al formar tres clusters la variación se reduce a

10112.005. La cohesión de las observaciones es mayor, pero, al analizar la diferencia entre formar tres clusters o crear cuatro se ve como la tasa de disminución no es tan significativa, con cuatro clusters la WSS es de 9416.945, por ello, se determina que lo más idóneo es tener tres grupos de clientes.

Este mismo resultado se obtiene por medio del método de la silueta, el Ancho Promedio de la Silueta más cercano a uno, es decir, con mejor estructura de clustering se alcanza al llevar a cabo tres agrupaciones.



Finalmente, empleando otros métodos de evaluación aportados por la librería NbClust se llega al mismo resultado, de los 23 métodos estudiados, 9 determinan que tres es el número óptimo de clusters.

El promedio de todas las observaciones de un cluster se materializa en el centroide, los componentes del centroide deben estudiarse respecto a las dimensiones o variables que configuran nuestro clúster. El centroide tiene un valor para cada una de las variables o dimensiones.

	Openness	Conscientiousness	Extraversion	Agreeableness	Emotional Stability
Cluster 1	5.044.170	5.312.721	2.343.640	4.289.753	5.457.597
Cluster 2	5.229.607	3.823.263	3.100.453	3.885.196	3.320.997
Cluster 3	5.847.360	4.971.947	4.972.772	4.508.251	5.080.033

La interpretación de los centroides de cada cluster nos ayuda a determinar qué rasgo es el dominante en cada grupo. En el cluster nº 1 la variable que más destaca es “Emotional Stability”. En el segundo cluster hay una clara predominancia de la variable “Openness”

frente a la existencia de valores medios en el resto de características. Finalmente, el tercer cluster tiene unos valores elevados en todas las variables, si bien la variable “Openness” también tiene el valor más alto constituyéndose como el rasgo más elevado y significativo de todos los que componen cualquiera de los tres clusters. A continuación, se va a llevar a cabo un análisis más detallado de cada cluster de manera que sea posible entender a los consumidores que los componen.

Si algo destaca del primer cluster es su bajo nivel de extraversión, en este grupo podemos encontrar a los clientes más introvertidos del dataset, sus componentes tienen un perfil social bajo y entre sus intereses no se encuentra la constante interacción social. Son los seres más racionales y emocionalmente estables de la muestra. Esto los configura como personas reflexivas que antes de llevar a cabo una compra evalúan minuciosamente los pros y contras. Adicionalmente, es el cluster con el valor de conciencia (“Conscientiousness”) más elevado, esto los describe como seres ordenados y prudentes, estudios demuestran cómo las personas con un alto grado de conciencia tienden a ser más selectivas con sus decisiones de compra le dan una alta importancia al precio del producto a la hora de valorar si adquirir un bien o servicios, si bien, este colectivo se encuentra entre los más fieles a una marca.⁴¹

La combinación de una elevada estabilidad emocional, un alto nivel de conciencia y una extraversión baja les convierte en el cluster más insensible frente a convenciones sociales, estos clientes son los menos propensos a caer en la emulación pecuniaria. Este cluster se denominará “ **clientes introvertidos y racionales**”.

En el segundo cluster hay un elemento distintivo, el grado de apertura frente a nuevas experiencias (“Openness”) estos clientes tienen una tasa de adaptabilidad superior a la media. Este grupo de consumidores tiene los niveles más bajos de términos de conciencia, amabilidad y estabilidad emocional. Ser el colectivo con la menor conciencia y estabilidad emocional, los identifica como seres con baja autodisciplina y actitud impulsiva, son más susceptibles de sufrir situaciones de ansiedad, debido a su naturaleza nerviosa tienden a sentirse más identificados con emociones negativas. Las personas con un alto nivel de neuroticismo tienen una mayor probabilidad de ver el mundo como un lugar competitivo, las

⁴¹ Ichwan Musa, M., Wardhana Haeruddin, M.I., Maulana Haeruddin, M.I., “Customers’ repurchase decision in the culinary industry: Do the Big-Five personality types matter?” en *Journal of Business and Retail Management Research (JBRMR)*, Vol 13, nº 1, 2018. pp 131-137. (Disponible en <https://eprints.unm.ac.id/28854/1/3.%20%28BEREPUTASI%29%20Customers%E2%80%99%20repurchase%20decision%20in%20the%20culinary%20industry%20Do%20the%20Big-Five%20personality%20types%20matter.pdf>; última revisión 6 de junio de 2024).

persona con una estabilidad emocional baja tiende a prestar más atención a las señales de peligro y competencia, el bajo nivel de amabilidad acentúa esta competitividad y hedonismo, existe una relación indirecta entre estas variables⁴². En definitiva, este cluster está formado por personas que priorizan su bienestar personal, por ello, estos clientes se denominarán los **“clientes competitivos individualistas”**.

Finalmente, nos encontramos con un tercer cluster de clientes compuesto por las personas más amables, extrovertidas y abiertas de todo el grupo de clientes. El alto nivel de apertura implica que son individuos optimistas ante el cambio, curiosos. Este grupo también se caracteriza por ser poco leal, sus ansias por explorar y probar nuevos productos pesan más que cualquier vínculo emocional con una marca. El alto grado de extroversión hace a este grupo especialmente sensible a las reacciones de terceros así como a la influencia de las redes sociales⁴³, son seres que valoran el estatus a la hora de tomar decisiones. La importancia del sentimiento de comunidad también se debe a su alto valor de amabilidad (“Agreeableness”), estos individuos tienen una naturaleza cooperativa, les encanta compartir e interactuar. A este cluster los denominaremos los clientes **“exploradores sociables”**.

Con toda esta información, se procede a idear tres campañas de marketing para Always Boutique.

Para los clientes introvertidos y racionales es esencial diseñar una publicidad que resalte la calidad de las prendas, su buena relación calidad precio u otros elementos objetivos que puedan medir. Son personas que valoran la durabilidad y eficacia, por ello, debemos transmitir la combinación entre funcionalidad y estilo de la marca proporcionada.

Los canales de comunicación más eficaces son aquellos que permiten una interacción individualizada y discreta, los correos electrónicos pueden ser una buena opción. En los correos personalizados que se enviarán, este tipo de consumidores valoran enormemente la información, por ello, deben poder encontrar la guía de los productos, su composición y recomendaciones de cuidado. Así mismo, aportar practicidad sugiriendo cómo combinar las

⁴² Van Hiel, A., Cornelis, I., & Roets, A.,” The intervening role of social worldviews in the relationship between the Five-Factor model of personality and social attitudes” en *European Journal of Personality*, vol 21, 2013, pp 131–148.

⁴³ Gerber, A., Huber, G., Doherty, D., Dowling, C., “The Big Five Personality Traits in the Political Arena” en *Annual Review of Political Science*, vol 14, 2011 pp 265-287. (Disponible en <https://doi.org/10.1146/annurev-polisci-051010-111659> , última revisión 5 de junio de 2024).

diferentes prendas puede empujar a este usuario a incluir ciertos productos en la cesta de compra.

Es especialmente importante captar a este segmento poblacional pues se caracterizan por ser los individuos más fieles a una marca, si quedan satisfechos con el producto es probable que vuelvan a comprar en Always Boutique. Incluir un programa de lealtad que premia las compras repetidas mediante descuentos exclusivos o acceso anticipado a las nuevas prendas puede infundir en estos clientes un sentimiento de pertenencia y aumentar con ello su fidelidad a la marca.

Los clientes competitivos individualistas deben poder sentir como comprar en Always Boutique colma sus deseos personales. La campaña de marketing enfocada a este grupo deben resaltar cómo ser cliente de esta marca es sinónimo de exclusividad y estatus, quien compra en la boutique es diferente y especial. Los mensajes deben convencer de que estas prendas son sinónimo de exclusividad, que el cliente van un paso por delante, no siguen las modas sino que las crean. El lenguaje empleado debe incluir términos como “edición limitada”, “exclusivo” o “en colaboración con”.

La impulsividad de este colectivo hace efectivo el uso de técnicas que animen una compra rápida; señalar el número de unidades que quedan de una prenda, poner una cuenta atrás en un descuento o hacerle saber al potencial cliente que si no adquiere una prenda ahora tal vez luego estará agotada incentiva la compra.

Estos consumidores buscan acercarse a la exclusividad, llegar a ser aspiracionales a ojos de terceros, por ello, las redes sociales son el mejor escaparate. Las campañas publicitarias deben emplear imágenes y videos de influencers o personas exitosas luciendo prendas de la boutique, el pensamiento que queremos sembrar en sus mentes es; “si me compro el mismo jersey que esta persona, a quien admiro, llegaré a ser igual que ella”.

Finalmente, el bajo nivel de amabilidad hace imperativo que la publicidad destaque los beneficios personales que obtendrán al comprar; cómo mejoran su estatus o destacarán como las personas mejores vestidas de una habitación.

Igualmente, la idea de comunidad debe ser nuclear a la hora de dirigirse al tercer cluster, a los exploradores sociales. Son personas que valoran las experiencias casi tanto como el bien en sí, por ello, darles la oportunidad de conocer de primera mano cómo es el proceso de creación de una prenda, asistir a desfiles o eventos de la marca les hará tener una actitud mucho más

positiva y cercana a la boutique. Estos clientes buscan experiencias mucho más holísticas y por ello, comprar una prenda de buena calidad no es suficiente, conocer los valores de Always Boutique o su compromiso con la sociedad y el medio ambiente pueden ser determinantes a la hora de atraer a este grupo social.

La publicidad debe ser dinámica, estos consumidores deben poder verse a sí mismos en multitud de situaciones sociales llevando la marca. El uso de otros clientes como embajadores es una buena manera de acercarse a los exploradores sociales, cuando un usuario comparte en una red social su experiencia de compra los individuos del tercer cluster tienden a prestar atención y empatizar, en definitiva, el testimonio de otros clientes es la mejor forma de convencer a los exploradores sociales. Finalmente, debido a su alto nivel de apertura es el grupo de clientes más difícil de fidelizar, por ello, campañas que fomenten su participación activa como por ejemplo, un concurso de fotos de lo que ser clientes de la Always Boutique pueden ser muy efectivas para mantener la atención de este segmento.

Al comprender la personalidad de sus clientes Always Boutique podrá incorporar en sus campañas publicitarias elementos, que aun pareciendo sencillos atraen a los diferentes compradores y generan una conexión más profunda.

5. Conclusiones

- Es necesario adquirir un enfoque más relacional con el cliente, comprender las necesidades y deseos del consumidor permite a una marca mantener su relevancia en un contexto digital caracterizado por la inmediatez y la disponibilidad. Herramientas como la segmentación del mercado, la investigación e introducción de estrategias psicográficas permiten construir relaciones más duraderas y significativas.
- Comprender la mente del consumidor es fundamental para el éxito de una campaña de marketing. La segmentación debe estudiar variables menos convencionales pero que aportan información valiosa sobre quién es el consumidor. Cualquier estrategia debe diseñarse sobre los pilares de teorías económicas y sociales y entender conceptos como la emulación pecuniaria o la importancia de los aspectos psicológicos, los consumidores son seres complejos y exigentes y así deben ser los métodos para su estudio.

- El uso de herramientas de Big Data es esencial para el éxito en el marketing. Gracias al gran volumen de datos que esta tecnología puede analizar, el Big Data proporciona información valiosa que los departamentos de comunicación y marketing deben emplear para actuar con eficiencia y agilidad en un mercado cambiante y volátil.
- La minería de datos nos permite extraer información personal a partir de las interacciones de un usuario en redes sociales. La recolección de esta información se llevará a cabo de manera ética, protegiendo al consumidor mediante robustos sistemas de seguridad. Integrando técnicas de análisis de sentimiento se pueden obtener visiones detalladas de la reacción de un usuario ante una campaña o un producto, los departamentos de marketing deben hacer uso de estas emociones para crear campañas personalizadas que atraigan a clientes y consigan su fidelización.
- Técnicas de clustering como K-means complementan el análisis de datos agrupando a consumidores e incrementando el grado de conocimiento que se tiene de un colectivo. La segmentación de potenciales clientes es una manera eficiente de optimizar los recursos, obtener una comprensión holística del cliente y fortalecer la relación consumidor-marca.
- El análisis de sentimiento empleado en el caso de Always Boutique se ha constituido como una poderosa herramienta para analizar las debilidades, fortalezas de una empresa. Permite instaurar estrategias de mejora partiendo desde un activo nuclear, el cliente y su percepción de marca.
- La eficiencia de una campaña publicitaria puede incrementarse estudiando a quienes va dirigida, a sus potenciales clientes. La personalización y clustering psicográfico permite comprender elementos subjetivos del consumidor y adaptar el mensaje a cada grupo de consumidores. Este trato personalizado es clave para maximizar el impacto y transformar los estímulos en interacciones duraderas.

6. Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado

ADVERTENCIA: Desde la Universidad consideramos que ChatGPT u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

Por la presente, yo, **María Ángeles Domínguez Luqui**, estudiante de **E3-Analytics** de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "**Psicografía y Big Data en el mundo del marketing**", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. **Brainstorming de ideas de investigación:** Utilizado para idear y esbozar posibles áreas de investigación.
2. **Crítico:** Para encontrar contra-argumentos a una tesis específica que pretendo defender.
4. **Metodólogo:** Para descubrir métodos aplicables a problemas específicos de investigación.
5. **Interpretador de código:** Para realizar análisis de datos preliminares.
7. **Constructor de plantillas:** Para diseñar formatos específicos para secciones del trabajo.
8. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.
9. **Generador previo de diagramas de flujo y contenido:** Para esbozar diagramas iniciales.
10. **Sintetizador y divulgador de libros complicados:** Para resumir y comprender literatura compleja.

13. **Revisor:** Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.

15. **Traductor:** Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: **21/06/2024**

Marian Domínguez Luqui

Firma: _____

7. Bibliografía

Libros de texto

- All port, F. “*Social Psychology*”, Cambridge: Houghton Mifflin, 1968.
- Grande, I., “*Dirección de marketing*”, McGraw-Hill, Madrid, 1992.
- Hartzog, W., “Privacy’s Blueprint: The battle to control the design of New Technologies”, Harvard University Press, 2018.
- Hoofnagle, C., “Federal Trade Commission Privacy Law and Policy”, Cambridge University Press, 2016.
- Kotler, P., Kartajaya, H., y Setiawan, I., “*Marketing 4.0 moving from traditional to digital*” John Wiley & Sons, Inc, New Jersey, 2017.
- Kotler, P., Kartajaya, H., Setiawan, I., “*Marketing 5.0 Technology for Humanity*”, John Wiley & Sons, 2021.
- Mayer-Schoneberger, V., Cukier, K., “*Big Data La revolución de los datos masivos*”, Turner Publicaciones S.L, 2013.
- Shiffman, L.G., Kanuk, L.L. “*Comportamiento del consumidor*”, 9º edición, Rio de Janeiro, 2012.
- Soto, C. J. “Big Five personality traits” en M. H. Bornstein, M. E. Arterberry, K. L. Fingerman, & J. E. Lansford (Eds.), *The SAGE encyclopedia of lifespan human development*, Thousand Oaks, CA: Sage, 2018. (Disponible en https://www.researchgate.net/publication/324115204_Big_Five_personality_traits)
- Veblen, T., “*La teoría de la clase ociosa*”, Alianza editorial, 2008.
- Yanaze, M. H. “Gestión de marketing y comunicación: avances y aplicaciones”, ed Sao Paulo: Atlas, 2011.

Buscadores académicos

- Google Académico: <http://scholar.google.com>
- Dialnet: <https://dialnet.unirioja.es/>

Otras

- Barrocal Blanquez, J., “Destacar entre miles de estímulos”, *Wola Publicidad*, 2022. Disponible en <https://wolapublicidad.com/destacar-entre-infinitos-estimulos>).
- Berg, M., “Así ha logrado Hermès convertirse en el símbolo de estatus por excelencia”, *Business Insider*, 4 marzo 2024. (Disponible en <https://www.businessinsider.es/ha-logrado-hermes-convertirse-simbolo-estatus-excelencia-1368110>).
- Carasila, C., Milton, A., “El concepto de Marketing: pasado y presente” en *Revista de Ciencias Sociales*, vol 14, nº2, 2008. (Disponible en https://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1315-95182008000200014)

- Cienfuegos Alvarado, C., Ramos Canales, E., Rábgos Rojas, C., y del Prado Flores, R., “Mercadotecnia integral y branding, comprehensive marketing and branding” en *Revista Anahuac*, 2021.. (Disponible en <https://revistas.anahuac.mx/sintaxis/article/view/1026/929>).
- Davis, J. Schibrowsky, J. Peltier, D., “Interactive Psychographics: Cross-Selling in the Banking Industry” en *Journal of Advertising Research*, vol. 42, nº 2, 2002.
- Garriga Domínguez, A., “Reflexiones sobre el marketing político y el fenómeno de la desinformación en el contexto electoral” en *AFD*, vol. XXXVI, 2020(Disponible en https://www.boe.es/biblioteca_juridica/anuarios_derecho/abrir_pdf.php?id=ANU-F-2020-1002510028).
- Gerber, A., Huber, G., Doherty, D., Dowling, C., “The Big Five Personality Traits in the Political Arena” en *Annual Review of Political Science*, vol 14, 2011. (Disponible en <https://doi.org/10.1146/annurev-polisci-051010-111659>).
- González-Ferriz, F., “El marketing 5.0 y su efecto en la estrategia empresarial del sector industrial en España” en *Revista de Marketing Aplicado (Red Marka)*, vol 25, nº 1, 2021. Disponible en https://revistas.udc.es/index.php/REDMARKA/article/view/redma.2021.25.1.7848/g7848_pdf).
- Grönroos, C., Voima, P., “Making Sense of Value and Value Co-Creation in Service Logic” en *Hanken School of Economics Working Papers*, 2011. (Disponible en https://www.researchgate.net/publication/279495418_Making_Sense_of_Value_and_Value_Co-Creation_in_Service_Logic).
- Huang, J., Su, S, Zhou, L., Liu, X., “Attitude Toward the Viral Ad: Expanding Traditional Advertising Models to Interactive Advertising” en *Journal of Interactive Marketing*, vol 25, 2013,(Disponible en <https://www.sciencedirect.com/science/article/abs/pii/S1094996812000333>).
- Ichwan Musa, M., Wardhana Haeruddin, M.I., Maulana Haeruddin, M.I., “Customers’ repurchase decision in the culinary industry: Do the Big-Five personality types matter?” en *Journal of Business and Retail Management Research (JBRMR)*, Vol 13, nº 1, 2018. (Disponible en <https://eprints.unm.ac.id/28854/1/3.%20%28BEREPUTASI%29%20Customers%20E2%80%99%20repurchase%20decision%20in%20the%20culinary%20industry%20Do%20the%20Big-Five%20personality%20types%20matter.pdf>).
- Judge, Timothy A., Jessica B. Rodell, Ryan L. Klinger, Lauren S. Simon, and Eean R. Crawford, “Hierarchical Representations of the Five-Factor Model of Personality in Predicting Job Performance: Integrating Three Organizing Frameworks with Two Theoretical Perspectives” en *Journal of Applied Psychology* vol 98 nº6, 2013. (Disponible en <https://doi.org/10.1037/a0033901>).
- Kim, S., Gil, J., “Research paper classification systems based on TF-IDF and LDA schemes” en *Human-centric Computing and Information Sciences*, 2019. (Disponible en <https://doi.org/10.1186/s13673-019-0192-7>).
- Kosinski, M., Stillwell, D., Graepel, T., “Private traits and attributes are predictable from digital records of human behavior”, *University of California Berkeley*, 11 de marzo de 2013, (Disponible en <https://www.pnas.org/doi/full/10.1073/pnas.1218772110>).
- Kulkarni, K., Kalro, A, Sharma, D., Sharma, P., “A typology of viral ad sharers using sentiment analysis” en *Journal of Retailing and Consumer Services*, vol 53, 2020.(Disponible en <https://www.sciencedirect.com/science/article/abs/pii/S0969698918307902>).

- Lyon, D., “Surveillance Studies. An overview” en *Canadian Journal of Sociology*, 2008, (Disponible en https://www.researchgate.net/publication/26526923_David_Lyon_Surveillance_Studies_An_Overview);
- McSweeney, T. “Psychographics, predictive analytics, artificial intelligence & Bots: Is the FTC Keeping Pace?” en *Georgetown Law Technology Review*, 2018. (Disponible en <https://georgetownlawtechreview.org/wp-content/uploads/2018/07/2.2-McSweeney-pp-514-30.pdf>).
- Pang, B., Lee, L. “Minería de opiniones y análisis de sentimientos” en *Fundamentos y tendencias en la recuperación de información* 2008.
- Pérez Miras, J., “La nueva protección de datos en Europa. El RGPD como nuevo marco jurídico europeo de Privacidad” en *Crónicas Jurídicas*. Disponible en <https://ws168.juntadeandalucia.es/iaap/revista/2018/10/21/la-nueva-proteccion-de-datos-en-europa-el-rgpd-como-nuevo-marco-juridico-europeo-de-privacidad/>)
- Plummer, J. T., “The Concept and Application of Life Style Segmentation” en *Journal of Marketing*, 1974.
- Provost, F., & Fawcett, T. “Data Science for Business: What you need to know about data mining and data-analytic thinking”. O'Reilly Media, Inc, 2013.
- Qi, Q., Hessen, D.J, Deoskar, T., Van der Heijden, P., “A comparison of latent semantic analysis and correspondence analysis of document-term matrices” en *Cambridge University Press*, 2023. (Disponible en <https://www.cambridge.org/core/journals/natural-language-engineering/article/comparison-of-latent-semantic-analysis-and-correspondence-analysis-of-document-term-matrices/96316CBB94CD9DD681CB50D9F52E0201>).
- Redin, L., “Machine Learning III: Text Mining (parte 1) Nociones básicas”, Universidad Pontificia de Comillas, 2023.
- Redin, L., “Machine Learning III: Text Mining (parte 2) Noción”, Universidad Pontificia de Comillas, 2023.
- Redin, L., “Machine Learning III: Técnicas de Clustering”, Universidad Pontificia de Comillas, 2023.
- Recalde, M.F., Suárez, L., Páez Egúez, J.C, “El comportamiento del consumidor en el ámbito económico y psicológico” en *Revista Conectividad*, vol 2, nº 1, 2020. (Disponible en <https://revista.ister.edu.ec/ojs/index.php/ISTER/article/view/13/63>).
- Rehman, Ikhlaq ur, "Facebook-Cambridge Analytica data harvesting: What you need to know" en *Library Philosophy and Practice (e-journal)*, 2019. (Disponible en <https://digitalcommons.unl.edu/libphilprac/2497>).
- Rojas Rubio, L, Meneses Villegas, C, “Una comparación empírica de algoritmos de aprendizaje automático versus aprendizaje profundo para la detección de noticias falsas en redes sociales” en *Ingeniare, Revista Chilena de Ingeniería*, vol 30, nº2, 2022. (Disponible en https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-33052022000200403)
- Urbonavičius, S., Kasnauskienė, G., “New Applications of a Traditional Psychographic Segmentation Concept” en *Commerce of Engineering Decisions*, nº 5, 2005.
- Van Hiel, A., Cornelis, I., & Roets, A.,” The intervening role of social worldviews in the relationship between the Five-Factor model of personality and social attitudes” en *European Journal of Personality*, vol 21, 2013.
- Wells, W. D., Tigert, D. J., “Activities, interests and opinions” en *Journal of Advertising Research*, 1971.

Datasets

- Guided Numeric and Text Exploration E-Commerce, Input Data Women's E-Commerce Clothing Reviews, Kraggle. (Disponible en <https://www.kaggle.com/code/nicapotato/guided-numeric-and-text-exploration-e-commerce/input>).
- Top Personality Dataset (MBTI) Myers-Briggs Personality Type Dataset, Kraggle. (Disponible en <https://www.kaggle.com/datasets/arslanali4343/top-personality-dataset>).

8.Anexos

Acceso a la carpeta drive donde se encuentran los datasets y el archivo de R Studio con el análisis de sentimiento y clustering.

<https://drive.google.com/drive/folders/1CgR40VSPWI5PqoQMvg9xqm-EsZ-C-eJY?usp=sharing>

