Texto

Descripción generada automáticamente

**Posgrado**

Analítica de Datos

Análisis predictivo del comportamiento del Consumidor

**Asignatura**

Métodos Estadísticos

**Presenta**

Nicolas Alejandro Fernandez Espinosa

John Alejandro Méndez Carrillo

|  |  |
| --- | --- |
| Bogotá D. C., Colombia | Marzo 2023 |

# **Introducción**

El mundo avanza a pasos agigantados y para las empresas comercializadoras de productos es preciso seguir el paso, conocer sus fortalezas y debilidades, ventajas y desventajas en un mercado altamente competitivo. Esto se logra conociendo de primera mano a sus propios consumidores, sus necesidades y características. Esto les permite diseñar productos y servicios que satisfagan sus demandas y generar mayores ventas y fidelidad de los clientes.

Sin embargo, conocer las necesidades de los consumidores y predecir su comportamiento puede ser un desafío. Los clientes son seres complejos y cambiantes, con necesidades y deseos que pueden variar con el tiempo, la situación y el contexto. Por lo tanto, las empresas necesitan recopilar y analizar grandes cantidades de datos para comprender los patrones de comportamiento de ellos.

En este sentido, los métodos predictivos y el uso de técnicas que permitan un acercamiento rápido y certero a los clientes se han convertido en una herramienta valiosa para las empresas. Estos métodos utilizan datos históricos para predecir el comportamiento futuro de los clientes, lo que permite a las empresas tomar decisiones más informadas y acertadas sobre el diseño de productos y servicios, el marketing y la estrategia empresarial en general.

# 

# **Marco Teórico**

La segmentación de clientes nace como una necesidad del marketing en la década de 1950, en la industria de los automóviles, donde las empresas comenzaron a diseñar modelos de automóviles para satisfacer las necesidades específicas de diferentes grupos de clientes. Fue tan exitosa que se extendió a sectores como alimentos y bebidas, detectando diferentes grupos de consumidores y adaptando su oferta de productos y servicios para sus clientes. En 1970, las técnicas se sofisticaron, la estadística comenzó a mejorar el proceso y la investigación de mercados tomó relevancia. Con ayuda de la tecnología y la utilización de bases de datos más sectores se unieron, como servicios. No solo se hablaba de compras, ahora se contaba con información de niveles de gasto, ingresos, frecuencia y otros factores. En el 2000, el uso de internet y redes sociales catapultó la segmentación de clientes a otro nivel, las empresas comenzaron a recopilar datos en línea y utilizar técnicas de segmentación para personalizar la oferta de productos y servicios.

La segmentación de los clientes debe permitirles a los usuarios:

* Diferencias en las necesidades de los clientes: Los clientes tienen diferentes necesidades y expectativas con respecto a los productos y servicios que ofrecen las empresas. La segmentación de los clientes permite a las empresas comprender mejor estas diferencias y diseñar productos y servicios que satisfagan las necesidades específicas de cada grupo.
* Comportamiento del consumidor: Los clientes tienen diferentes comportamientos de compra y patrones de consumo. Al segmentar los clientes, las empresas pueden identificar estos comportamientos y adaptar su oferta de productos y servicios para satisfacer las necesidades específicas de cada grupo.
* Competencia: Es un factor importante a considerar en cualquier estrategia de marketing. La segmentación de clientes permite a las empresas mantener una ventaja competitiva al diseñar productos y servicios que satisfagan mejor las necesidades de los clientes.
* Personalización: Es una tendencia creciente en la industria del marketing. La segmentación de clientes permite a las empresas personalizar su oferta de productos y servicios para satisfacer las necesidades específicas de cada grupo. Esto puede mejorar la experiencia del cliente y fomentar la lealtad del cliente.
* Tecnología: Ha cambiado la forma en que las empresas interactúan con sus clientes. La segmentación de clientes se ha vuelto más fácil y precisa con el uso de la tecnología. Las empresas pueden recopilar datos sobre el comportamiento de los clientes y utilizarlos para segmentarlos con mayor precisión.

# **Objetivo general**

Comprender mejor a los clientes y sus necesidades para desarrollar una estrategia de marketing más efectiva. Identificar grupos de clientes con comportamientos similares para ofrecer productos y servicios específicos que satisfagan sus necesidades. Mejorar la rentabilidad de la empresa al enfocarse en segmentos de clientes más rentables y reducir el costo de adquisición de clientes.

## **Objetivos específicos**

* Identificar patrones de comportamiento y preferencias de los clientes para personalizar los mensajes y ofertas de marketing.
* Identificar segmentos de clientes con mayor potencial de crecimiento y oportunidades de expansión.
* Evaluar la eficacia de las campañas de marketing en diferentes segmentos de clientes.
* Identificar las necesidades insatisfechas de los clientes y desarrollar productos y servicios para satisfacerlas.
* Identificar las tendencias del mercado y la competencia para adaptar la estrategia de marketing de la empresa.

# **Análisis de descriptivo**

Se cuenta con una base de datos de los consumidores de un supermercado. Tienen 2.240 registros y 29 columnas. Cuenta con información personal como:

* ID: identificador único del cliente
* Year Birth: año de nacimiento del cliente
* Educación: nivel de educación del cliente
* Marital Status: estado civil del cliente
* Income: ingresos familiares anuales del cliente
* Kidshome: Número de niños en el hogar del cliente
* Teen Home: Número de adolescentes en el hogar del cliente
* Dt Cliente: Fecha de alta del cliente en la empresa
* Recency: Número de días desde la última compra del cliente
* Complain: 1 si el cliente se quejó en los últimos 2 años, 0 en caso contrario

Información de consumo:

* MntWines: Cantidad gastada en vino en los últimos 2 años
* MntFruits: Cantidad gastada en frutas en los últimos 2 años
* MntMeatProducts: cantidad gastada en carne en los últimos 2 años
* MntFishProducts: cantidad gastada en pescado en los últimos 2 años
* MntSweetProducts: cantidad gastada en dulces en los últimos 2 años
* MntGoldProds: cantidad gastada en oro en los últimos 2 años

Información de promociones:

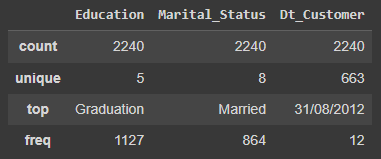
* NumDealsPurchases: Número de compras realizadas con descuento
* AcceptedCmp1: 1 si el cliente aceptó la oferta en la 1.ª campaña, 0 en caso contrario
* AcceptedCmp2: 1 si el cliente aceptó la oferta en la 2ª campaña, 0 en caso contrario
* AcceptedCmp3: 1 si el cliente aceptó la oferta en la 3ra campaña, 0 en caso contrario
* AcceptedCmp4: 1 si el cliente aceptó la oferta en la 4ª campaña, 0 en caso contrario
* AcceptedCmp5: 1 si el cliente aceptó la oferta en la 5ª campaña, 0 en caso contrario
* Respuesta: 1 si el cliente aceptó la oferta en la última campaña, 0 en caso contrario

Información de la forma de compra:

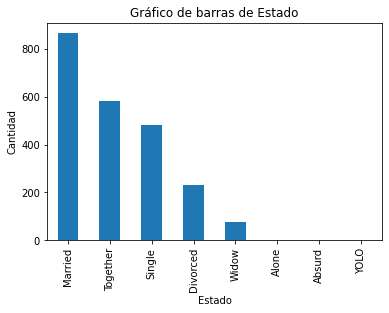
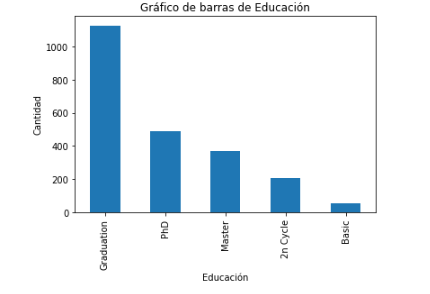
* NumWebPurchases: Número de compras realizadas a través de la web de la empresa
* NumCatalogPurchases: Número de compras realizadas utilizando un catálogo
* NumStorePurchases: Número de compras realizadas directamente en las tiendas
* NumWebVisitsMonth: Número de visitas al sitio web de la empresa en el último mes

**Análisis de las variables**

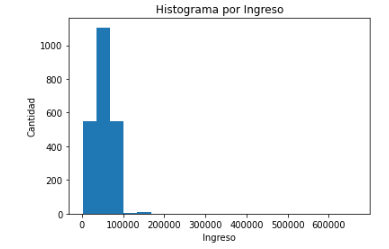
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Tipo** | **Únicos** | **Nulos** | **Cantidad** |
| ID | Integer | 2240 | 0 | 2240 |
| Year\_Birth | Integer | 59 | 0 | 2240 |
| Education | String | 5 | 0 | 2240 |
| Marital\_Status | String | 8 | 0 | 2240 |
| Income | Float | 1974 | 24 | 2240 |
| Kidhome | Integer | 3 | 0 | 2240 |
| Teenhome | Integer | 3 | 0 | 2240 |
| Dt\_Customer | String | 663 | 0 | 2240 |
| Recency | Integer | 100 | 0 | 2240 |
| MntWines | Integer | 776 | 0 | 2240 |
| MntFruits | Integer | 158 | 0 | 2240 |
| MntMeatProducts | Integer | 558 | 0 | 2240 |
| MntFishProducts | Integer | 182 | 0 | 2240 |
| MntSweetProducts | Integer | 177 | 0 | 2240 |
| MntGoldProds | Integer | 213 | 0 | 2240 |
| NumDealsPurchases | Integer | 15 | 0 | 2240 |
| NumWebPurchases | Integer | 15 | 0 | 2240 |
| NumCatalogPurchases | Integer | 14 | 0 | 2240 |
| NumStorePurchases | Integer | 14 | 0 | 2240 |
| NumWebVisitsMonth | Integer | 16 | 0 | 2240 |
| AcceptedCmp3 | Integer | 2 | 0 | 2240 |
| AcceptedCmp4 | Integer | 2 | 0 | 2240 |
| AcceptedCmp5 | Integer | 2 | 0 | 2240 |
| AcceptedCmp1 | Integer | 2 | 0 | 2240 |
| AcceptedCmp2 | Integer | 2 | 0 | 2240 |
| Complain | Integer | 2 | 0 | 2240 |
| Z\_CostContact | Integer | 1 | 0 | 2240 |
| Z\_Revenue | Integer | 1 | 0 | 2240 |
| Response | Integer | 2 | 0 | 2240 |

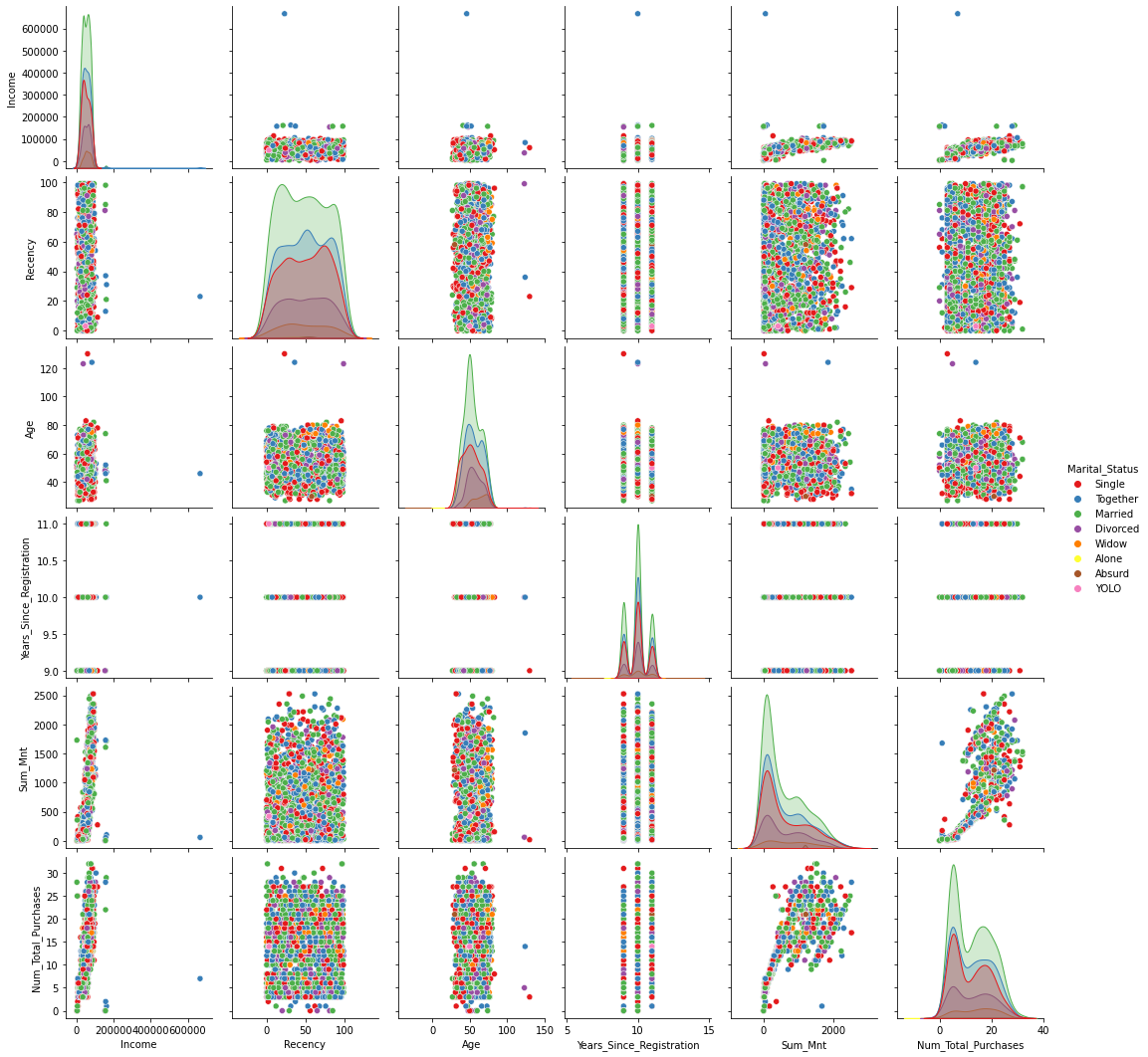


Revisando las variables tenemos:

* El 50% de los consumidores es profesional.
* El 38.5% de los consumidores está casado.

No hay un comportamiento particular del año de nacimiento de los clientes.

* Ingresos:  
  El ingreso medio de los clientes es de 52.000 dólares. 
* Relación de variables con el estado civil



# **Metodologías a Implementar**

Se usarán métodos de análisis de datos de Machine Learning como aprendizaje supervisado y no supervisado.

En un inicio se realizará un análisis exploratorio de datos que permita identificar las características de las variables, posibles relaciones entre estas y detectar valores atípicos. Se puede realizar mediante el uso de gráficos, tablas de frecuencia y estadísticas descriptivas. En las descriptivas anteriormente mostradas, se obtuvieron estadísticas descriptivas y visualizaciones de las variables a estudiar.

Luego se hará uso del Análisis de Componentes Principales (PCA) para reducir la dimensionalidad del conjunto de datos, transformando las variables originales en un conjunto más pequeño y de componentes principales.

Se realizará clasificación para entrenar el algoritmo y detectar una compra futura por los clientes, así si un nuevo cliente se presenta en el algoritmo, este clasificará al cliente en el grupo de "comprará" o "no comprará".

Por último, se realizará una comparación con un modelo de regresión lineal para determinar la relación entre el ingreso (variable dependiente) y otra u otras variables que afectan el comportamiento de compra.

**Bibliografía**

GONZALO, A. (2019). *Segmentación utilizando K-means en Python*. Obtenido de <https://machinelearningparatodos.com/segmentacion-utilizando-k-means-en-python/>

FABIEN, D. (2019). *Kaggle*. Obtenido de <https://www.kaggle.com/code/fabiendaniel/customer-segmentation>

VANDERPLAS, J. (2016). *Python Data Science Handbook*.

RASCHKA, S. & MIRJALILI, V (2019). *Python Machine Learning*.

WAYNE, L. (2018). *Marketing Analytics: Essential Tools for Data-Driven Decisions*.