



Desarrollo de un Modelo de Clusterización para Segmentar Consumidores de Aguacate

Autor (es)
Laura Restrepo Morales
Estefania Valencia Cuartas

Especialización en Big Data e Inteligencia de Negocios

Facultad de Ingenierías y Arquitectura

Universidad Católica Luis Amigó

Medellín, Colombia

2024

Dedicatoria

Este trabajo de grado es dedicado a mi Familia quienes son el pilar fundamental de mi vida y mi crecimiento personal acompañándome en cada uno de mis retos, Gracias por estar siempre a mi lado, por ser mi fuente de inspiración y por darme las alas para seguir soñando.

Laura Restrepo Morales

Este trabajo de grado es dedicado a mi mamá, su apoyo constante y su capacidad para superar retos, han sido una inspiración inigualable. Gracias por estar siempre a mi lado, celebrando cada uno de mis logros y siendo mi mayor motivación.

A mi abuela, que ha sido mi pilar en cada etapa de mi vida. Tu presencia inquebrantable y tu sabiduría han sido una guía en los momentos de dificultad y una alegría en cada logro alcanzado.

A mi alma gemela, que ha sido mi fuente de inspiración y mi apoyo incondicional. Tu presencia ha sido esencial en cada paso de este camino.

Y a mi madrina, cuya influencia y apoyo han contribuido significativamente a mi proyecto de vida. Tu orientación y amor han sido fundamentales para alcanzar mis metas.

Estefanía Valencia Cuarta

Agradecimientos

Agradezco profundamente a mi familia, cuyo amor y apoyo incondicional han sido fundamentales para mi crecimiento y éxito, También quiero expresar mi agradecimiento a mis maestros, por su dedicación, enseñanza. Sus lecciones han sido esenciales en este proceso.

Laura Restrepo Morales

Agradezco a todas las personas que han sido fundamentales en la culminación de este trabajo de grado. Primero, a mi familia y madrina, cuyo amor y apoyo han sido el pilar sobre el cual he construido cada uno de mis logros. Y a mis docentes, cuyo conocimiento, dedicación y orientación han sido esenciales en mi formación. Este trabajo es el reflejo del esfuerzo, la dedicación y el cariño que ustedes me han brindado. A cada una de ustedes, mi más sincero agradecimiento.

Estefanía Valencia Cuartas

Contenido

Introducción.....	5
Motivación.....	7
Planteamiento del problema.....	8
Justificación.....	9
Objetivos.....	10
Marco metodológico.....	11
Marco teórico.....	16
Marco conceptual.....	19
Marco normativo.....	21
Desarrollo del Proyecto.....	22
Discusión.....	55
Conclusión.....	75
Referencias.....	76

Introducción

El aguacate es un producto que, gracias a sus propiedades nutricionales, su sabor y su versatilidad ha tenido un aumento significativo en el consumo en el mundo. En Colombia, este aumento se traduce en una gran oportunidad para las empresas exportadoras a este sector, como es el caso de Cartama, empresa exportadora número uno de aguacate Hass en el país. En este sentido, se hace necesario conocer las características y los hábitos de consumo de los clientes alrededor del aguacate, con el fin de diseñar estrategias de marketing que permitan el incremento de las ventas. Por lo anterior, el presente trabajo de grado tuvo como objetivo general construir un modelo de clusterización para identificar comportamientos de compra, preferencias y demografía de los consumidores de aguacate en el mercado colombiano. En este sentido, se plantean objetivos específicos que incluyen: i) la recolección de información demográfica y psicográfica, ii) la transformación de la información, iii) desarrollar un modelo de clusterización con técnicas de aprendizaje no supervisada y iv) validación del modelo propuesto con información real de la empresa.

La clusterización, justifica el presente proyecto ya que va a permitir segmentar a los consumidores en grupos con características similares estas agrupaciones le permitirán a Cartama distinguir a sus clientes potenciales y los determinantes en su proceso de toma de decisiones a la hora de realizar una compra. Esto permitirá relacionar la demanda con la oferta de una manera adecuada enfocada en el buyer persona correcto. De la misma manera, las variedades de aguacate, formas y momentos de consumo son claves para determinar la forma correcta para acercarnos al cliente y lograr entender si Cartama necesita lanzar productos nuevos, mejorar sus campañas de marketing o cualquier otra estrategia pertinente para el negocio. Como se

mencionó con anterioridad el consumo de aguacate ha aumentado con los años y se ha convertido en tendencia no solo el consumo de fruta sino también de sus derivados o productos fabricados a base de este con productos tales como Aceites, postres, productos cosméticos entre otros.

A Pesar de que el consumo de aguacate es de carácter global , poco se ha desarrollado a la hora de conocer el perfil del consumidor y en el caso de Colombia existe muy poca información que permita construir un buyer persona por esta Razón Cartama ha realizado diversas encuestas en sus diferentes puntos de venta con el fin de obtener información real de las personas que actualmente ya son clientes de la marca y entender que buscan estas personas y que las motiva a consumir aguacate Por lo tanto realizar este proceso de clusterización es vital para el desarrollo comercial de la compañía.

Motivación

El interés por la perspectiva que encierra el aguacate, un alimento sano que ha comenzado a cobrar fuerza en el mercado internacional y con la creciente popularidad que la fruta ha experimentado en los últimos años ha surgido la necesidad de comprender cuáles han sido las dinámicas del mercado y conocer cuáles son los consumidores por lo tanto el desarrollo de este modelo no solo constituye un reto académico sino también una oportunidad para lograr un impacto real sobre una empresa del sector agro en este caso Cartama qué es la principal empresa de exportación de aguacate Hass en Colombia que en busca de lograr impactar el mercado colombiano necesita estrategias de venta que le generen un valor agregado a la fruta para que esta sea aprovechada en su totalidad y que los colombianos seamos parte de este crecimiento y consumo pues la mayoría de la fruta que se produce en Colombia termina en mercados extranjeros. Con la segmentación de los clientes esperamos que Cartama pueda mejorar sus estrategias de marketing y promover un enfoque en el consumo del producto en el mercado nacional.

Planteamiento del problema

Para este trabajo de grado se buscó desarrollar un modelo de clusterización con el fin de identificar cuáles son los patrones de compra, preferencias y características demográficas de los consumidores de aguacate. Esto es crucial para empresas como Cartama, líder en la exportación de aguacate Hass en Colombia, que busca expandir su mercado interno. El proceso clave es la construcción del modelo de clusterización que permita segmentar a los consumidores de aguacate y así optimizar las estrategias de marketing. La necesidad de comprender quienes son los consumidores de esta fruta es debido al creciente interés mundial por el aguacate, tanto por sus beneficios nutricionales como por su versatilidad culinaria y cosmética.

Para este ejercicio es clave analizar los datos demográficos y psicográficos recolectados, para identificar factores que influyen en la compra del aguacate, tales como la variedad, la forma y los momentos de consumo. Estos conocimientos permitirán a Cartama adaptar su oferta de manera más efectiva, utilizando canales de distribución adecuados y campañas de marketing dirigidas.

Estudios previos, como el de Wu et al. (2023), han demostrado que "los ingresos, la edad, el origen étnico y la educación se correlaciona significativamente con las ventas de aguacate", lo que subraya la importancia de la segmentación del mercado para llegar de manera efectiva al consumidor final. Además, el modelo podría ser un instrumento clave para incrementar la competitividad de Cartama y mejorar la rentabilidad (Wu et al., 2023).

Justificación

Con el constante crecimiento del consumo de aguacate a nivel mundial es necesario entender cuál es el perfil del consumidor de esta fruta y realizar la clusterización de este para entender las preferencias y necesidades de los clientes para optimizar las estrategias de marketing que permita mejorar las ventas de las empresas dedicadas a la comercialización de este fruto, en este caso Cartama compañía exportadora Número uno de aguacate Hass de Colombia que busca desarrollar la venta interna de la fruta impulsando el consumo dentro del país para esto es fundamental entender el mercado nacional y las decisiones de compra de los consumidores para atender de forma correcta la demanda de este producto

La construcción de este modelo permitirá segmentar los consumidores en grupos con características y comportamientos similares permitiéndole identificar a Cartama quienes son sus verdaderos clientes potenciales y qué factores son determinantes para la compra del aguacate incluyendo atributos como Variedades , formas de consumo , disponibilidad , momentos de consumo y demás atributos a considerar , estos insights permitirán que Cartama adapte su oferta al mercado, seleccionando los canales de distribución adecuados , campañas de marketing efectivas que logren un impacto directo sobre las ventas y la rentabilidad de la empresa.

Objetivos

Objetivo general

Construir un modelo de clusterización que facilite la identificación de los comportamientos de compra, preferencias y demografía de los consumidores de aguacate

Objetivos específicos

- Extraer datos demográficos y psicográficos sobre el consumo de aguacate
- Transformar los datos recopilados para la construcción del modelo
- Desarrollar un modelo de clusterización con técnicas de aprendizaje no supervisado
- Evaluar el modelo construido con datos reales de la compañía

Marco metodológico

En este capítulo se describe de forma clara el paso a paso para dar cumplimiento a cada uno de los objetivos específicos, así como la descripción de las diferentes actividades, métodos y técnicas que permitan desarrollar los pasos de la metodología CRISP-DM y los objetivos propuestos para dar solución a la problemática previamente descrita.

Objetivo específico	Actividad	Entregable	Fase CRISP-DM
Extraer datos demográficos y psicográficos sobre el consumo de aguacate	Diseñar y aplicar encuestas realizadas a clientes Cartama en los puntos de venta de las ciudades de Medellín y Bogotá y redes sociales	Base de datos, la cual es una encuesta diseñada con preguntas cerradas con el fin de tener la información lo más precisa posible donde encontramos las siguientes variables demográficas, psicográficas y de tipo comportamental.	Comprensión del Negocio / Comprensión de los Datos

<p>Transformar los datos recopilados para la construcción del modelo</p>	<p>Revisión de los datos recopilados a través de encuestas en puntos de venta y redes sociales.</p> <p>Análisis exploratorio para identificar patrones, faltantes y posibles errores en los datos.</p> <p>Limpieza y preprocesamiento de los datos, incluyendo normalización y codificación de variables categóricas.</p>	<p>Documento con la caracterización de la BD.</p> <p>Documento donde se evidencie el proceso de limpieza</p>	<p>Preparación de los Datos</p>
<p>Desarrollar un modelo de clusterización con técnicas de aprendizaje no supervisado</p>	<p>Realización de un modelo no supervisado con la base de datos "Encuestas punto de venta", mediante el lenguaje de programación Python que ya está depurada en el paso anterior.</p> <p>Descripción de los cluster creados, para identificar patrones dentro de cada segmento.</p> <p>Construir el ciclo de vida de cada clúster de acuerdo con las características anteriormente encontradas.</p>	<p>Modelo de clustering en la herramienta Google Colab.</p> <p>descripción de los segmentos arrojados por el modelo de clusterización.</p>	<p>Modelado</p>

<p>Evaluar el modelo construido con datos reales de la compañía</p>	<p>Diseñar estrategias de retención por medio de los resultados obtenidos en el modelo.</p> <p>Realizar la evaluación técnica bajo conceptos propios de los resultados obtenidos del modelo.</p>	<p>Un análisis detallado de cada uno de los segmentos de clientes que fueron creados.</p> <p>Propuestas de estrategias para que los clientes permanezcan por más tiempo en la marca.</p>	<p>Evaluación</p>
---	--	--	-------------------

Los objetivos previamente descritos con sus actividades y entregables se desarrollarán a partir de la metodología Crisp-Dm (Cross-Industry Standard Process for Data Mining). Es un método validado, de los más utilizados; incluye un modelo y guía. Para la aplicación de esta metodología, se tienen en cuenta seis fases que dependen entre sí y en cada fase se encuentran los siguientes elementos

1. Comprensión del Negocio.

Esta fase se centra en definir los objetivos del proyecto y entender el contexto comercial. Se identifican las partes interesadas y sus expectativas, se recopila información relevante sobre el mercado y el consumidor, y se evalúan los recursos disponibles y las restricciones del proyecto.

2. Comprensión de los datos.

En esta fase, se recopilan datos de fuentes internas y externas. Se realiza una descripción de los datos disponibles, se lleva a cabo un análisis exploratorio para identificar patrones y anomalías, y se evalúa la calidad de los datos, preparando un plan para su limpieza.

3. Preparación de los Datos.

Aquí se seleccionan las variables relevantes para el análisis y se limpia la información, eliminando o corrigiendo datos inconsistentes o faltantes. También se transforma la información a un formato adecuado para el modelado y se pueden construir nuevas variables o estructuras que mejoren el análisis.

4. Modelado.

Esta fase implica la selección de técnicas de modelado adecuadas, como K-means o DBSCAN. Se construye el modelo utilizando herramientas de análisis de datos y se evalúan sus resultados, ajustando los parámetros según sea necesario. También se valida que el modelo cumpla con los objetivos establecidos.

5. Evaluación.

En esta etapa, se revisa el proceso y los resultados obtenidos. Se comparan los hallazgos con los objetivos del negocio y se proporcionan recomendaciones basadas en los resultados. Además, se toma una decisión sobre si proceder con la implementación del modelo o realizar ajustes.

6 implementación.

Finalmente, se desarrolla un plan para aplicar las soluciones obtenidas. Se establecen métricas para medir el éxito del modelo en la práctica y se proporciona soporte para realizar ajustes según sea necesario. Además, se documenta todo el proceso y los resultados para futuras referencias.

Estas fases se interrelacionan y pueden requerir iteraciones para afinar el enfoque y los resultados, lo que hace que la metodología CRISP-DM sea especialmente útil para proyectos de minería de datos y análisis, como el desarrollo del modelo de clusterización en esta investigación

Marco teórico

La aceptación mundial del aguacate como un alimento nutritivo y versátil se ha visto impulsada por sus numerosos beneficios nutricionales, incluidos ácidos grasos monoinsaturados, vitaminas y antioxidantes, como lo señalan Wang et al., 2020. La popularidad de esta fruta en Colombia ha ido creciendo, en gran parte debido a su exposición en las redes sociales, lo que ha resultado en su promoción en diversos platos como guacamole, ensaladas y sándwiches. El aguacate se ha convertido en un símbolo de modernidad y salud, atrayendo a consumidores que buscan opciones de alimentos saludables chen et al., 2023.

El comportamiento del consumidor es el estudio de cómo y por qué las personas realizan compras. Los factores demográficos y psicográficos, así como la edad, el género y los factores socioeconómicos, son esenciales para determinar las preferencias de compra en este contexto. La demanda de productos como el aguacate está fuertemente influenciada por estas variables, por lo que es importante segmentar el mercado para entender las necesidades de los consumidores, como lo demuestran Wu et al., 2023.

Se pueden configurar estrategias de marketing eficaces mediante el uso de clustering, un método de aprendizaje no supervisado que agrupa a los consumidores en segmentos con características similares. El uso de un modelo de clustering permitirá a la empresa Cartama identificar los factores clave que influyen en las decisiones de compra de aguacate, permitiendo una oferta que satisfaga las necesidades del mercado de manera más efectiva, como lo señalan Kogan et al. en 2018.

El consumo de aguacates ha experimentado un aumento significativo a nivel mundial. La tendencia se ha visto influenciada en gran medida por el deseo de encontrar alimentos nutritivos

y la disponibilidad generalizada de aguacate en las redes sociales. El guacamole es un aderezo favorito para los aguacates en los EE. UU., pero también existe la opción de comerlos en diferentes momentos del día, como el desayuno Cavaletto en 2015. El mercado colombiano puede aprender de estas tendencias e implementar estrategias de marketing más efectivas.

A pesar del aumento del consumo, los pequeños productores colombianos aún enfrentan importantes desafíos para acceder a los mercados internacionales. Su rentabilidad y sostenibilidad se ven obstaculizadas por la falta de recursos y la competencia desleal. La investigación e implementación de estrategias que mejoren la cadena de valor del aguacate es fundamental para mejorar la calidad de vida en las zonas rurales donde son parte importante del proceso de producción, como afirma Anesbury et al., 2020.

Además, la creciente demanda de aguacate no se debe sólo a su valor nutricional, sino también al impacto medioambiental que conlleva su producción. La importancia de la agricultura sostenible y el impacto del comportamiento del consumidor en el medio ambiente va en aumento. Para abordar este cambio de perspectiva, es imperativo examinar cómo la producción y comercialización del aguacate pueden alinearse mejor con los valores establecidos en Cavaletto sourcing (2015) y Feller et al. et al. (2016).

El impacto de las redes sociales en el comportamiento del consumidor ha sido transformador y tiene un impacto significativo en sus decisiones de compra. Estas plataformas pueden servir como un medio para conectar emocionalmente al consumidor con la marca, lo cual debe tenerse en cuenta a la hora de determinar estrategias de marketing para el aguacate en Colombia anesbury et al., 2020.

En definitiva, el marco teórico anterior proporciona una base sólida para el estudio de los patrones de compra y venta de los consumidores de aguacate en Colombia. Se pueden formular estrategias efectivas que no solo aumenten las ventas de Cartama sino que también promuevan el desarrollo sostenible en la producción y distribución de aguacate si se comprenden las tendencias del mercado, el perfil del consumidor y los desafíos que enfrentan los productores. El objetivo de este estudio es llenar el conocimiento que actualmente falta del consumidor colombiano y mejorar las condiciones de vida de las comunidades involucradas en la cadena de valor del aguacate.

Marco conceptual

- **Aguacate:** El aguacate es una fruta de alto valor nutricional, rica en grasas saludables, vitaminas (como la E, K y B), minerales y antioxidantes. Su popularidad ha crecido debido a la percepción de ser un alimento saludable y su versatilidad en la cocina (Wang et al., 2020).
- **Comportamiento del Consumidor:** El comportamiento del consumidor se refiere al proceso de toma de decisiones que involucra la búsqueda, evaluación y selección de productos o servicios. Este comportamiento está influenciado por factores demográficos, psicográficos y culturales, y puede variar según el contexto y las tendencias del mercado (Kotler & Keller, 2016).
- **Segmentación de Mercado:** La segmentación de mercado es el proceso de dividir un mercado en grupos de consumidores con características o comportamientos similares. Esta práctica permite a las empresas adaptar sus estrategias de marketing para satisfacer mejor las necesidades de cada segmento (Smith, 1956). En el caso del aguacate, la segmentación puede basarse en variables demográficas (edad, ingreso, educación) y psicográficas (valores, estilo de vida).
- **Clusterización:** La clusterización es una técnica de análisis de datos que agrupa objetos o individuos en conjuntos (clúster) basados en similitudes. En el contexto de esta investigación, se utilizará para identificar grupos de consumidores de aguacate con preferencias y comportamientos de compra similares. Este enfoque permite a las empresas como Cartama entender mejor a su audiencia y optimizar sus estrategias de marketing (Jain & Dubes, 1988).

- **Modelos No Supervisados:** Los modelos no supervisados son técnicas de aprendizaje automático que permiten descubrir patrones en datos sin etiquetas predefinidas. A diferencia del aprendizaje supervisado, donde se utilizan datos etiquetados para entrenar modelos, los modelos no supervisados identifican estructuras ocultas en los datos. En esta investigación, se emplearán para desarrollar un modelo de clusterización que facilite la segmentación del mercado de aguacate en Colombia (Hastie et al., 2009).
- **Marketing Relacional:** El marketing relacional se centra en construir relaciones duraderas con los clientes, más allá de una simple transacción. Este enfoque es fundamental en la industria alimentaria, donde las emociones y las experiencias de consumo juegan un papel crucial en la lealtad del cliente (Berry, 1983). Las campañas de marketing dirigidas a los "amantes" del aguacate pueden fortalecer esta relación.
- **Tendencias del Mercado:** Las tendencias del mercado son patrones observables en el comportamiento de los consumidores y en las dinámicas de la industria. En el caso del aguacate, el auge de las redes sociales ha promovido su imagen como un producto gourmet, lo que ha influido en las decisiones de compra y ha abierto nuevas oportunidades en subproductos (Chen et al., 2023).

Marco normativo

La Ley 9 de 1979 establece normas de sanidad agropecuaria, asegurando la calidad y seguridad del aguacate. La Resolución 590 de 2016 regula las condiciones fitosanitarias, definiendo requisitos para el manejo de plagas.

En el ámbito comercial, el Código de Comercio (Decreto 410 de 1971) establece las reglas para la comercialización, mientras que la Ley 1480 de 2011 (Estatuto del Consumidor) protege los derechos de los consumidores, garantizando información clara sobre los productos.

La Ley 1581 de 2012 regula la protección de datos personales, asegurando la confidencialidad de la información recopilada, complementada por el Decreto 1377 de 2013, que establece directrices para su tratamiento.

Finalmente, los acuerdos de libre comercio facilitan la exportación de aguacate, permitiendo a empresas como Cartama acceder a nuevos mercados internacionales.

Desarrollo del proyecto

El presente proyecto tiene como objetivo analizar el consumo de aguacate a través de datos demográficos y psicográficos, utilizando la metodología CRISP-DM como marco de trabajo para estructurar el análisis.

objetivo específico I: Extraer datos demográficos y psicográficos sobre el consumo de aguacate. (Comprensión del Negocio)

Comprensión de negocio

Como lo mencionamos anteriormente Cartama siendo líder en la exportación de aguacate Hass en Colombia, busca expandirse en el mercado local buscando ofrecer su fruta con un valor agregado para ello es necesario comprender que consumen los colombianos y cuáles son sus patrones de compra con el fin de diseñar estrategias comerciales por lo que realizaremos un modelo de clusterización para segmentar a sus consumidores en grupos con características similares y tratar de identificar si existe un buyer person en particular basándonos en basado en datos demográficos, psicográficos y de consumo.

La información obtenida le permitirá a Cartama mejorar su eficiencia en marketing, distribución de producto, desarrollo de nuevas propuestas alimenticias explorando oportunidades en productos derivados del aguacate, aumentando así su presencia en el mercado colombiano.

Diccionario

Se realizan encuestas a clientes Cartama en los puntos de venta de las ciudades de Medellín y Bogotá, es una encuesta diseñada con preguntas cerradas con el fin de tener la información lo más precisa posible.

Donde encontramos las siguientes variables demográficas:

- ¿A qué rango de edad perteneces?:

Rango de edad de las personas encuestadas: 18 a 24, 25 a 34, 35 a 44, 45 a 54, 55 o más Variable tipo Object, que nos permite conocer la edad de los clientes actuales.

- Género: Femenino, Masculino Variable tipo Object nos permite conocer el género de los clientes actuales de aguacate.
- Número de integrantes en el Hogar: 1, 2, 3, 4 o 5 personas Variable tipo float 64, permite la caracterización de los hogares según su número de integrantes
- Grado de escolaridad: Bachiller, Técnico Tecnólogo, Profesional, Posgrado, Maestría. Variable tipo Object que permite caracterizar a los clientes según su grado de escolaridad
- Tiene Empleo: Si o No Variable de tipo Object, que permite la caracterización del cliente conociendo si este se encuentra empleado o no
- Ciudad: Medellín o Bogotá Variable tipo Object que permite la caracterización del cliente según la zona en la que habita

También hallamos las siguientes Variables psicográficas:

- ¿Qué variedad de aguacate consumes?:

Antillano, Choquete, Fuerte, Hass, lorena, Papelillo, Santana

Variable tipo Object que permite segmentar el mercado y conocer la variedad de aguacate más consumida.

- ¿Con qué frecuencia consumes Aguacate?:

Cada semana, Cada 15 días, cada mes, todos los días, ocasionalmente

Variable de tipo Object que permite identificar la frecuencia de consumo de la fruta en este caso Aguacate

- ¿Qué te motiva a consumir aguacate?:

Su sabor, su versatilidad en la cocina, sus beneficios para la salud, tendencias alimenticias. Variable de tipo Object que permite la caracterización del cliente según su motivación para el consumo del aguacate

- ¿Qué factores influyen más en tu decisión de compra de aguacates?: Calidad,

Precio Variable tipo object que permite la caracterización del cliente según sus preferencias al seleccionar y comprar aguacate

- ¿Te gustaría ver más variedades de productos de aguacate en el mercado? (como productos de cuidado personal productos horneados etc.): Si, No

Variable de tipo Object que permite la caracterización de los subproductos derivados del aguacate y conocer la aceptación de éstos por parte de los clientes.

también se encontraron variables de tipo comportamental:

- Cada vez que consumes aguacate ¿Cuántos kilos compras?: 1,2,3,4,5,6,7,8,9,10
Variable tipo float64 Permite la caracterización de clientes y de hogares según la cantidad de kilos de aguacate que compra
- Donde compras el aguacate: Almacenes de Cadena, Carretillas, Legumbrieras, Supermercados, Tienda de Barrio. Variable tipo Object, que permite la caracterización de los lugares donde los clientes compran aguacate.
- ¿Consumes guacamole?: Si, No, A veces Variable tipo Object que permite identificar el consumo del guacamole como producto derivado del aguacate
- ¿Con qué frecuencia consumes guacamole? Cada 15 días, cada mes, cada semana, ocasionalmente, no aplica. Variable tipo object que permite la caracterización de la frecuencia con la que el cliente consume guacamole
- ¿Cuáles son los momentos en que consumes Aguacate: ¿Almuerzo, Cena, Desayuno? Variable tipo Object que permite la caracterización del cliente según los momentos en los que consume aguacate
- ¿Qué otros productos de aguacate consumes?: Aceite de Aguacate, Guacamole, Salsa de aguacate, Ninguno. Variable tipo object que permite conocer si existen otros productos derivados del aguacate que al cliente le gusten consumir
- ¿Cómo sueles consumir aguacate principalmente?: Como ingrediente en platos principales (sopas sándwiches etc.), En recetas dulces (batidos postres etc.), En recetas saladas (guacamole, ensaladas, tacos etc.) variable de tipo Object que permite conocer las preferencias del cliente al consumir el aguacate

- ¿Has probado alguna vez productos de aguacate no convencionales como helado de aguacate o brownies de aguacate?: Si, No Variable de tipo Object que nos permite conocer si el cliente ha probado productos derivados del aguacate

El resultado de esta actividad fue una base de datos que proporciona una visión inicial del perfil del consumidor.

Al recibir la base de datos, se realizó un análisis preliminar para comprender mejor el contexto y las características de los consumidores. Se identificaron las principales variables que afectan el consumo de aguacate y se establecieron hipótesis sobre los patrones de comportamiento de los consumidores.

Para dar inicio al desarrollo de nuestro modelo comenzamos con la carga de la base de datos a Google Colab, que será la herramienta utilizada para la ejecución del código.

Realizamos la visualización de las 5 filas superiores del Data set para tener una vista preliminar

```
d = pd.read_csv("/content/Encuestas punto de venta.csv")
# visualizamos 5 filas superiores
d.head(5)
```



	¿A que rango de edad perteneces?	Género	Número de integrantes en su hogar	Grado de escolaridad	tiene empleo	ciudad	¿Que variedad de aguacate consumes?	¿Con qué frecuencia consumes Aguacate?	Cada que consumes aguacate ¿Cuántos kilos compras ?	Donde compras el aguacate
0	25 a 34	Femenino	2.0	Profesional	Si	Medellin	Hass	Todos los días	5.0	Legumbreteria
1	18 a 24	Femenino	1.0	Profesional	Si	Medellin	Hass	Todos los días	5.0	Tienda de Barrio
2	25 a 34	Femenino	2.0	Profesional	Si	Medellin	Hass	Todos los días	5.0	Tienda de Barrio
3	25 a 34	Masculino	2.0	Tecnologo	Si	Medellin	Hass	Cada semana	6.0	Supermercados
4	25 a 34	Femenino	2.0	Posgrado	Si	Medellin	Hass	Cada mes	2.0	Carretillas

Objetivo específico II: Transformar los datos recopilados para la construcción del modelo (Comprensión de los Datos).

Una vez que se obtuvo la base de datos, se procedió a la fase de procesamiento y limpieza de datos con el fin de realizar un Modelo tipo Clúster para determinar cuál es el perfil del consumidor de aguacate, comenzamos con el análisis exploratorio de la base de datos "Encuestas punto de venta". Se revisaron los datos obtenidos de las encuestas, realizando un análisis exploratorio para identificar patrones, datos faltantes y posibles errores. En este caso contamos

con la ventaja de que la base de datos original se recopiló con un formulario de Google con preguntas cerradas lo que influyó en la limpieza de datos logrando que esta fuera mucho más sencilla.

Para dar inicio procedemos a la revisión de los datos recopilados a través de encuestas en puntos de venta y redes sociales. Visualizamos las últimas 5 filas

```
[ ] #visualizamos las 5 filas inferiores
d.tail(5)
```



	¿A que rango de edad perteneces?	Género	Número de integrantes en su hogar	Grado de escolaridad	tiene empleo	ciudad	¿Que variedad de aguacate consumes?	¿Con qué frecuencia consumes Aguacate?	Cada que consumes aguacate ¿Cuántos kilos compras ?	Donde compras el aguacate
11192	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
11193	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
11194	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
11195	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
11196	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Se puede evidenciar que hay filas vacías las cuales se eliminarán más adelante, comprobamos los tipos de datos de nuestra DB y podemos visualizar que la mayoría de variables son categorías de tipo objeto.

```

#comprobamos los tipos de datos
d.dtypes

```

		0
¿A que rango de edad perteneces?	object	
Género	object	
Número de integrantes en su hogar	float64	
Grado de escolaridad	object	
tiene empleo	object	
ciudad	object	
¿Que variedad de aguacate consumes?	object	
¿Con qué frecuencia consumes Aguacate?	object	
Cada que consumes aguacate ¿Cuántos kilos compras ?	float64	
Donde compras el aguacate	object	
¿Consumes guacamole?	object	
¿Con qué frecuencia consumes guacamole?	object	
¿Cuáles son los momentos en que consumes Aguacate	object	
¿Qué otros productos de aguacate consumes?	object	
¿Qué te motiva a consumir aguacate?	object	
¿Cómo sueles consumir aguacate principalmente?	object	
¿Has probado alguna vez productos de aguacate no convencionales como helado de aguacate o brownies de aguacate?	object	
¿Qué factores influyen más en tu decisión de compra de aguacates?	object	
¿Te gustaría ver más variedades de productos de aguacate en el mercado? (como productos de cuidado personal productos horneados etc)	object	

```

dtype: object

```

Continuamos con la exploración del data set, analizando la cantidad de valores nulos en cada variable

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11197 entries, 0 to 11196
Data columns (total 19 columns):
# Column
---
0 ¿A que rango de edad perteneces?
1 Género
2 Número de integrantes en su hogar
3 Grado de escolaridad
4 tiene empleo
5 ciudad
6 ¿Que variedad de aguacate consumes?
7 ¿Con qué frecuencia consumes Aguacate?
8 Cada que consumes aguacate ¿Cuántos kilos compras ?

```

	Non-Null Count	Dtype
0 ¿A que rango de edad perteneces?	11192 non-null	object
1 Género	11192 non-null	object
2 Número de integrantes en su hogar	11192 non-null	float64
3 Grado de escolaridad	11192 non-null	object
4 tiene empleo	11192 non-null	object
5 ciudad	11192 non-null	object
6 ¿Que variedad de aguacate consumes?	11192 non-null	object
7 ¿Con qué frecuencia consumes Aguacate?	11192 non-null	object
8 Cada que consumes aguacate ¿Cuántos kilos compras ?	11192 non-null	float64

Después de analizar la información anterior, encontramos que todas las columnas son relevantes para el modelo y se procede con la eliminación de las filas nulas y visualizamos las últimas 5 filas para confirmar la eliminación de los nulos

d.tail(5)

	¿A que rango de edad perteneces?	Género	Número de integrantes en su hogar	Grado de escolaridad	tiene empleo	ciudad	¿Que variedad de aguacate consumes?	¿Con qué frecuencia consumes Aguacate?	Cada que consumes aguacate ¿Cuántos kilos compras ?	Donde compras el aguacate	¿Consumes guacamole?	¿Con qué frecuencia consumes guacamole?
11187	25 a 34	Masculino	4.0	Tecnico	Si	Medellin	Lorena	Cada mes	1.0	Almacenes de Cadena	A veces	Ocasionalmente
11188	25 a 34	Masculino	4.0	Tecnico	Si	Bogota	Hass	Todos los dias	10.0	Tienda de Barrio	No	No aplica
11189	35 a 44	Femenino	5.0	Profesional	Si	Bogota	Choquete	Cada semana	5.0	Almacenes de Cadena	A veces	Ocasionalmente

Procedemos con la transformación de nuestros datos en el caso de los datos numéricos: procedemos a transformar los datos de la variable Número de integrantes en su hogar y variable Cada que consumes aguacate ¿Cuántos kilos compras? por número entero, ya que se encuentran en decimal, dándole un formato incensario al Data Set

```
d['Número de integrantes en su hogar'] = d['Número de integrantes en su hogar'].astype(int)
d.head(5)
```

	¿A que rango de edad perteneces?	Género	Número de integrantes en su hogar	Grado de escolaridad	tiene empleo	ciudad	¿Que variedad de aguacate consumes?	¿Con qué frecuencia consumes Aguacate?	Cada que consumes aguacate ¿Cuántos kilos compras ?	Donde compras el aguacate	¿Consumes guacamole?	¿Con qué frecuencia consumes guacamole?
0	25 a 34	Femenino	2	Profesional	Si	Medellin	Hass	Todos los dias	5.0	Legumbreteria	Si	Ocasionalmente
1	18 a 24	Femenino	1	Profesional	Si	Medellin	Hass	Todos los dias	5.0	Tienda de Barrio	Si	Cada semana
2	25 a 34	Femenino	2	Profesional	Si	Medellin	Hass	Todos los dias	5.0	Tienda de Barrio	Si	Cada semana
3	25 a 34	Masculino	2	Tecnologo	Si	Medellin	Hass	Cada semana	6.0	Supermercados	Si	Cada semana

Una vez eliminadas las filas nulas, se procede con el cambio de nombre de algunas columnas para que sea más sencilla su visualización

```
d = d.rename(columns={"¿A que rango de edad perteneces?": "Edad", "Número de integrantes en su hogar":
"# Int en el Hogar", "Grado de escolaridad": "Escolaridad", "tiene empleo": "Empleo",
"¿Que variedad de aguacate consumes?": "Variedad_aguacate",
"¿Con qué frecuencia consumes Aguacate?":
"Frecuencia consumo aguacate", "Cada que consumes aguacate ¿Cuántos kilos compras ?":
"¿Cuántos kilos compras ?", "¿Con qué frecuencia consumes guacamole?":
"Frecuencia consumo guacamole", "¿Cuáles son los momentos en que consumes Aguacate":
"Momentos consumo del aguacate",
"¿Has probado alguna vez productos de aguacate no convencionales como helado de aguacate
"Has probado alguna vez productos de aguacate no convencionales",
"¿Qué factores influyen más en tu decisión de compra de aguacates?":
"Factores de compra",
"¿Te gustaría ver más variedades de productos de aguacate en el mercado? (como productos
d.head()
```

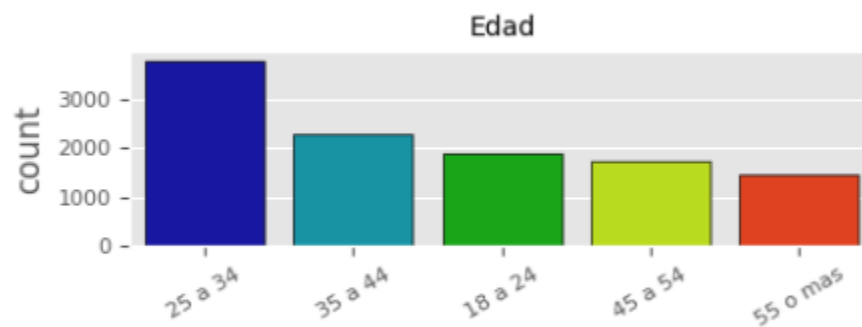
Procedemos analizar las variables numéricas del Data set en este caso solo tenemos dos, # int en el hogar y ¿cuántos kilos compras?

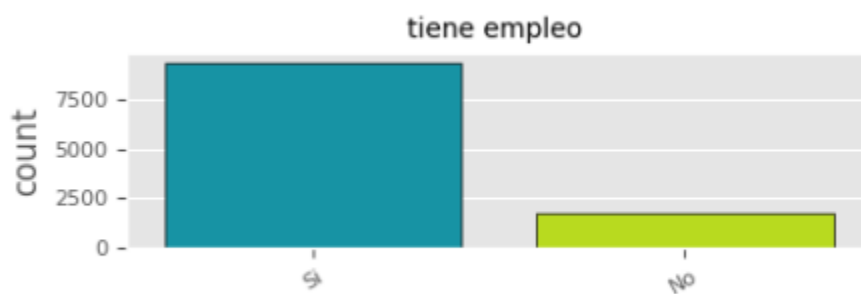
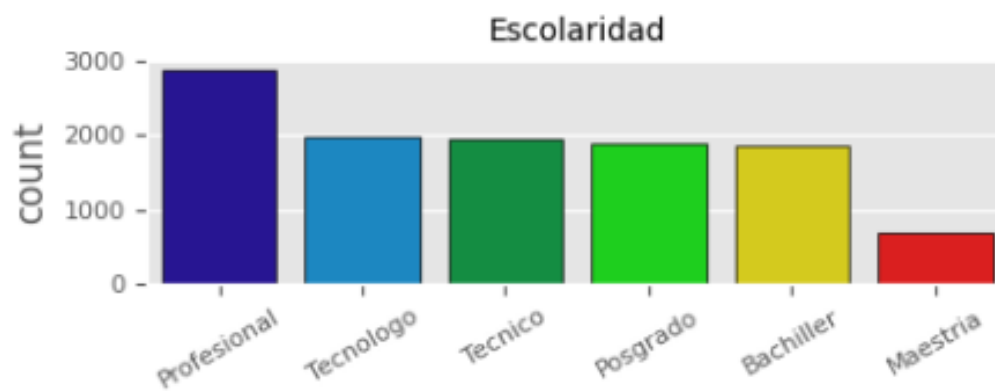
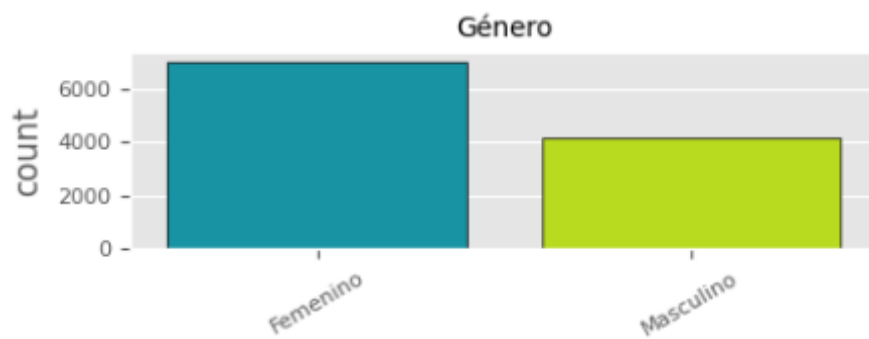
```
numCols=d.select_dtypes(include = ['float64','int32','int64']).columns.tolist()
d[numCols].head(2)
```

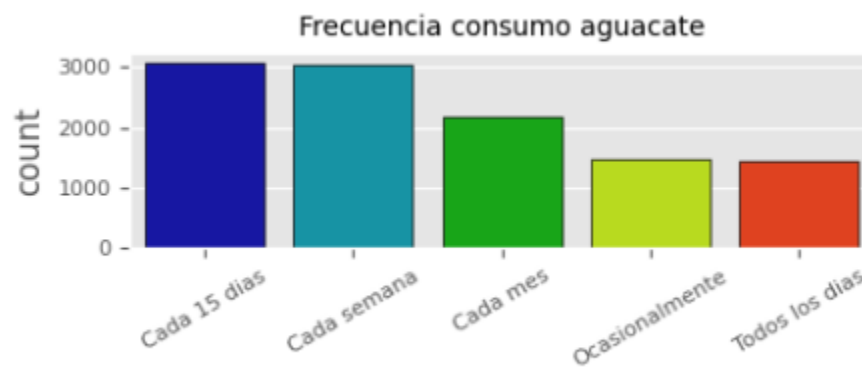
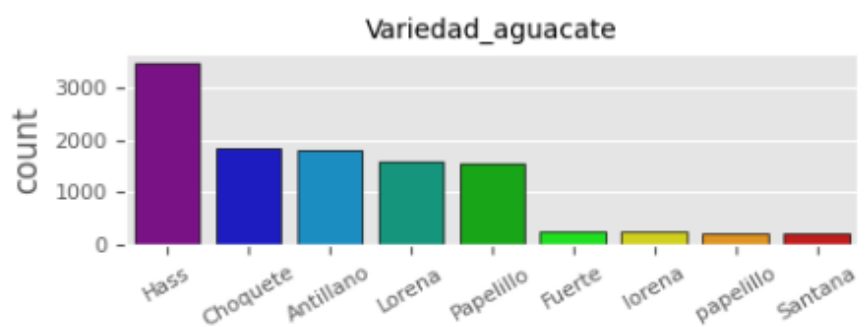
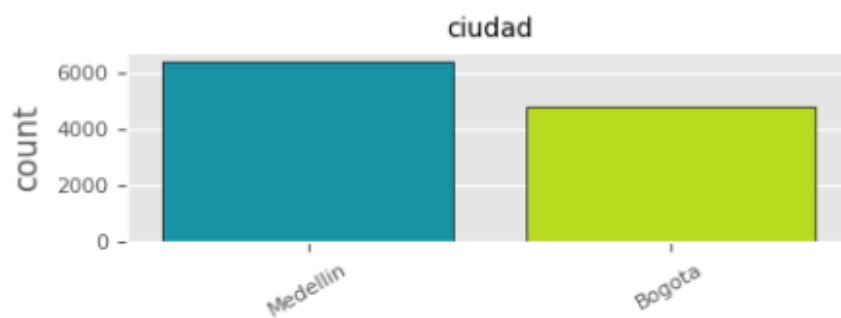
	# Int en el Hogar	¿Cuántos kilos compras ?
0	2	5
1	1	5

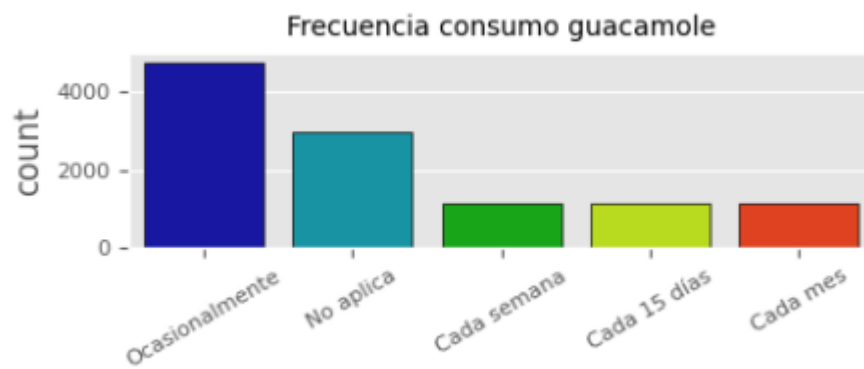
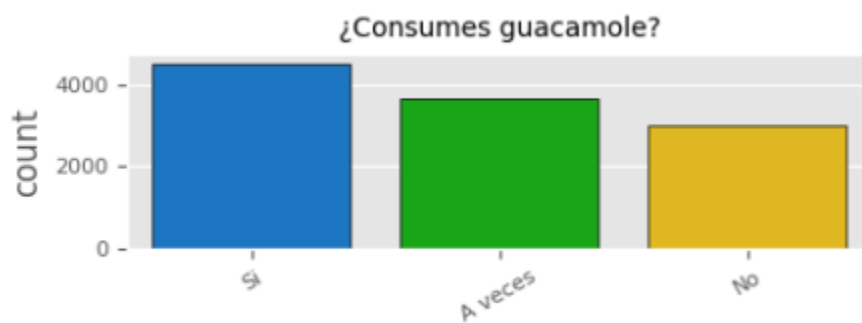
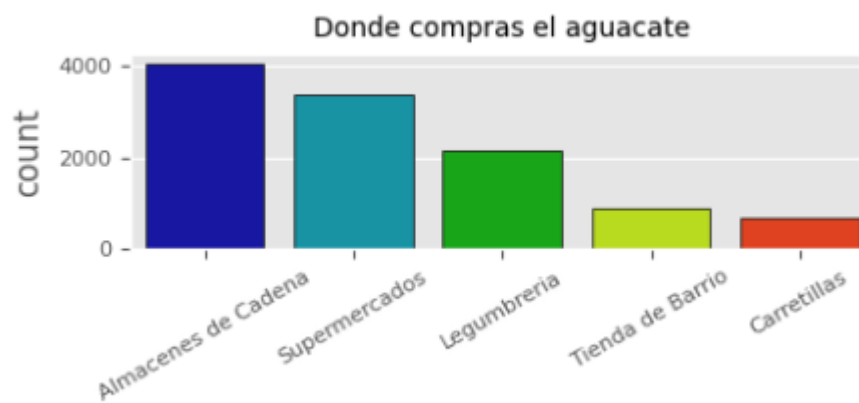
También analizamos las variables categorías y las exploramos por medio de gráficos, que nos permiten una mejor visualización de los datos y ver el comportamiento de estos

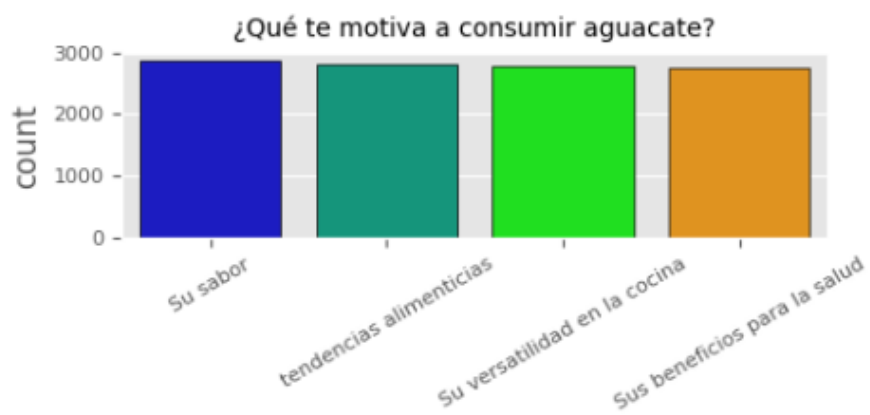
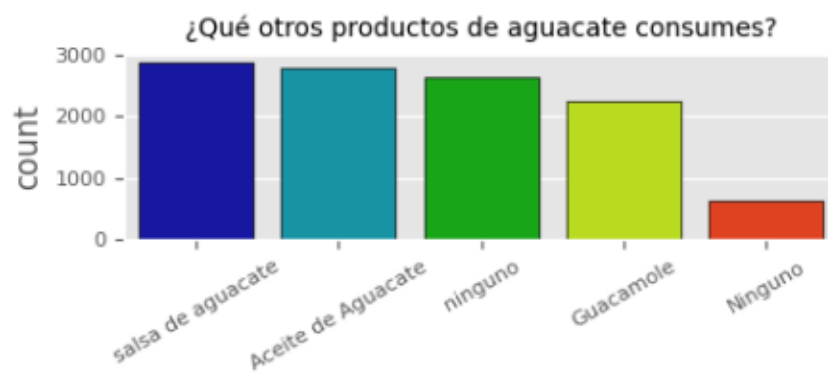
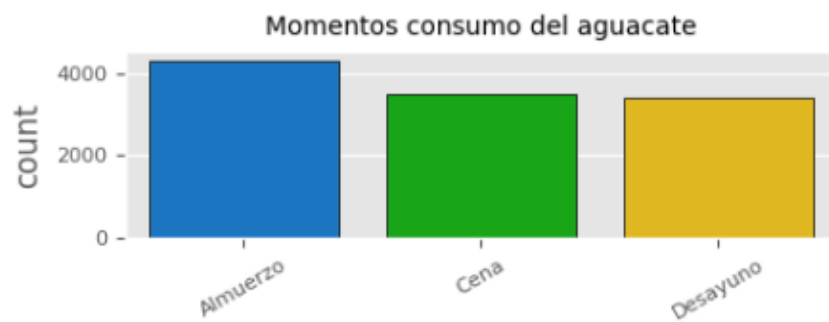
```
#Visualización de frecuencia de instancias para variables categóricas
multiple_plot(3, d , catCols, None, 'countplot', 'Frecuencia de instancias para variables categóricas',30)
```

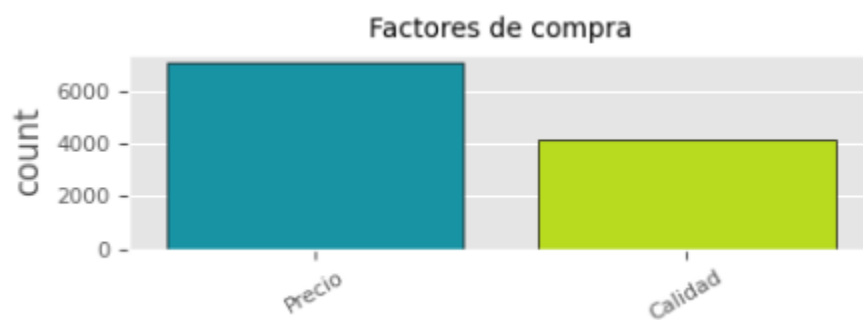
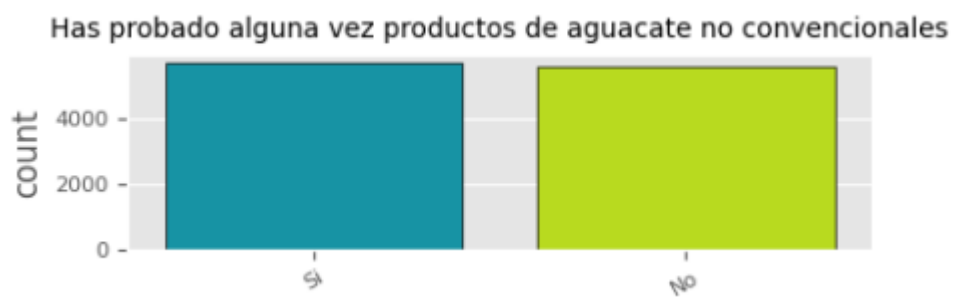
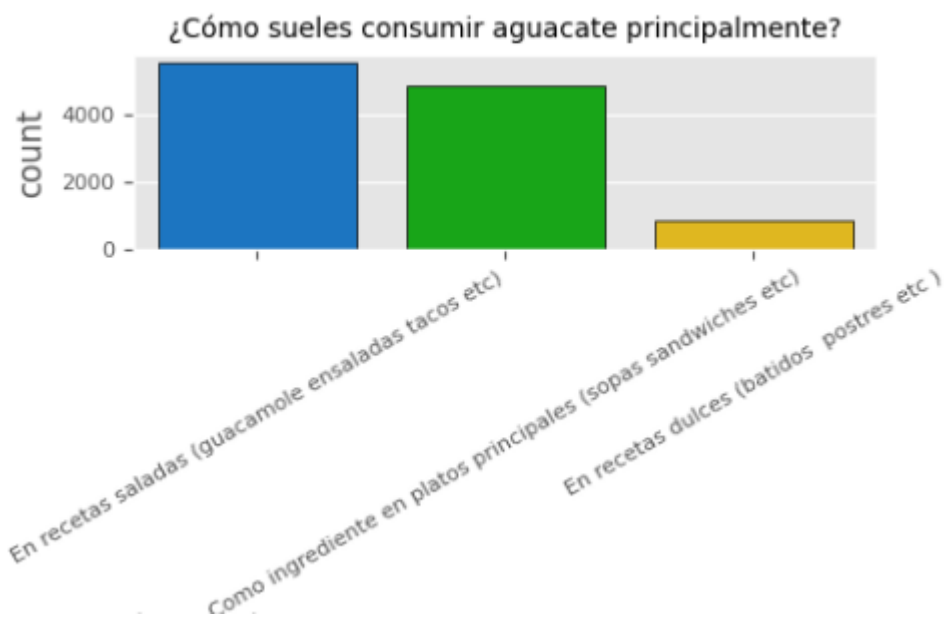


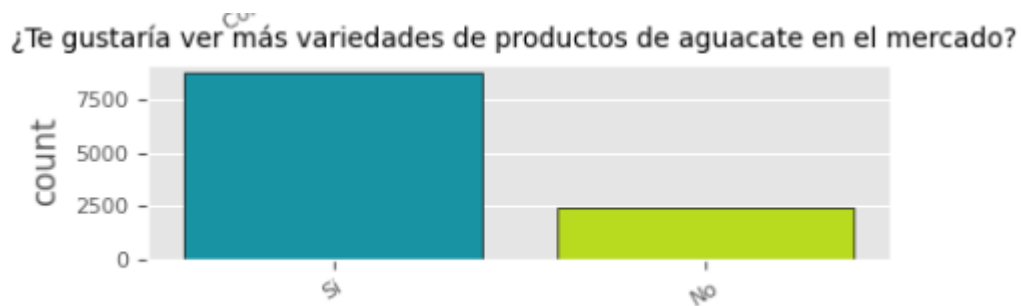












Aunque las gráficas son una buena forma de visualizar la información, es bueno identificar por cada variable el número de registros que tiene cada una, para tener una información más detallada y clara

```
for col in catCols:
    print("="*20 + f" {col} " + "="*20)
    print(d[col].value_counts())
    print()
```

```
===== Edad =====
```

```
Edad
25 a 34    3804
35 a 44    2286
18 a 24    1884
45 a 54    1748
55 o mas   1470
Name: count, dtype: int64
```

```
===== Género =====
```

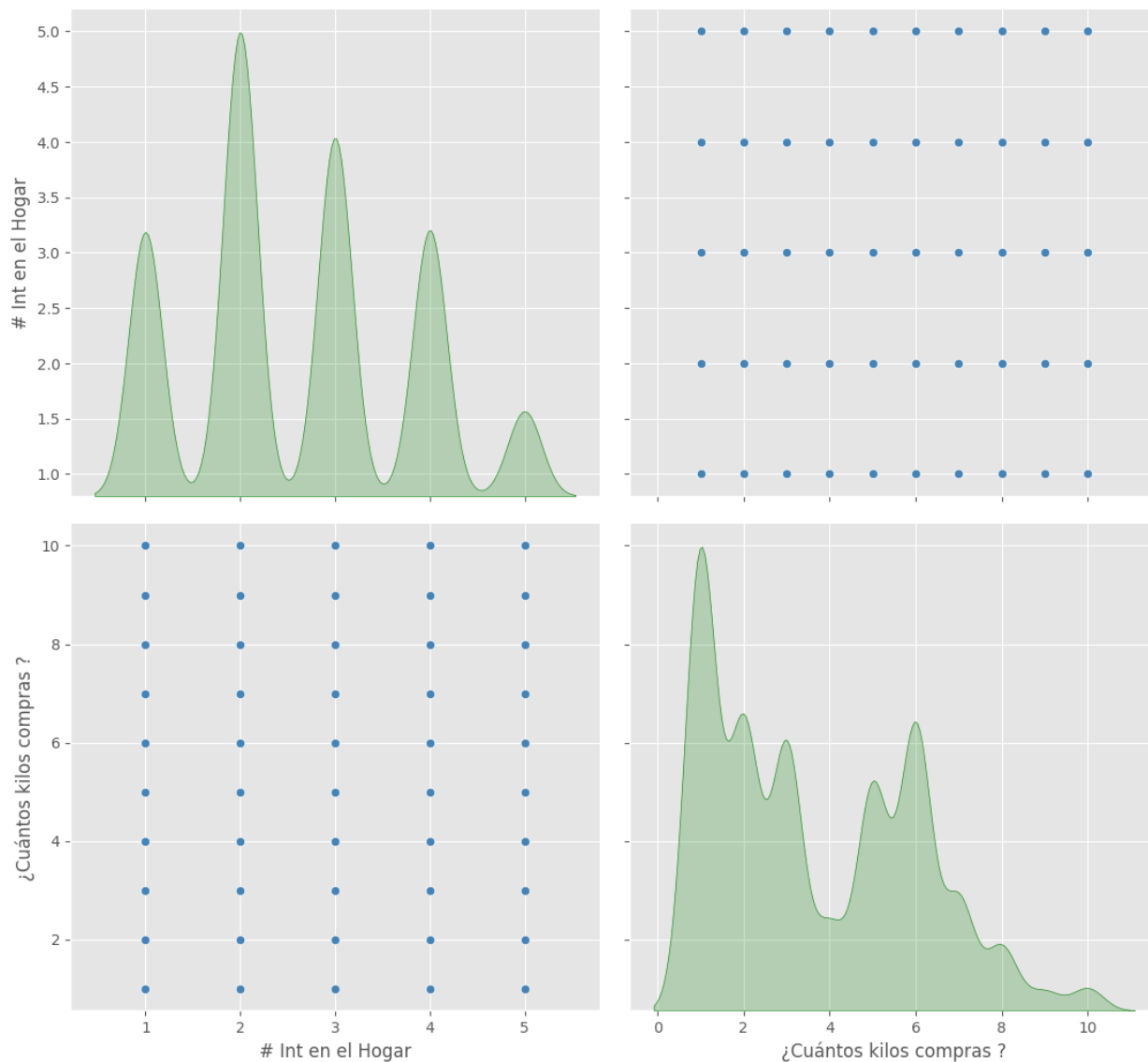
```
Género
Femenino    7048
Masculino   4144
Name: count, dtype: int64
```

```
===== Escolaridad =====
```

```
Escolaridad
Profesional    2873
Tecnologo     1960
Tecnico        1930
Posgrado       1896
Bachiller      1857
Maestria        676
Name: count, dtype: int64
```

Procedemos con la visualización de las variables numéricas y la relación entre estas dos

Relación entre las variables numéricas



Por medio de este gráfico que nos permite comparar las dos variables numéricas: el número de integrantes del hogar ("# Int en el Hogar") y los kilos de aguacate comprados ("¿Cuántos kilos compras?"). observamos que existen picos que sugieren que los hogares están

conformados por un número de personas probablemente entre 1,2 y 5 y el promedio de compra de kilos está entre los 0 y los 6 kilos, pero parece no haber mucha relación entre estas dos variables ambos factores tienen distribuciones particulares, pero no fuertemente relacionadas por lo que pensamos que otros factores son los que influyen en la compra del aguacate

Una vez analizadas las variables y visualizadas cada una de ellas, procedemos a pasar todas nuestras variables categóricas a numéricas de forma manual con el fin de evitar que se creen nuevas variables que puedan afectar el modelo por lo que las variables se transforman de la siguiente manera:

Edad:

de 18 a 24 años: 0

25 a 34 años: 1

de 35 a 44 años: 2

45 a 54 años: 3

55 o más años: 4

Género:

Femenino: 0

Masculino: 1

Escolaridad

Bachiller: 0,

Maestría: 1,

Postgrado: 2,

Profesional: 3,

Técnico: 4,

Tecnólogo: 5

Frecuencia consumo aguacate

Cada 15 días: 0,

Cada mes: 1,

Cada semana: 2,

Ocasionalmente: 3,

Todos los días: 4

Donde compras el aguacate

Almacenes de Cadena: 1,

Carretillas: 2,

Legumbres: 3,

Supermercados: 4,

Tienda de Barrio: 5

Ciudad

Bogotá: 0,

Medellín: 1

Variedad de aguacate

antillano: 0,

Choquete: 1,

fuerte: 2,

Hass: 3,

lorena: 4,

papelillo: 5,

santana: 6

Consumes Guacamole

A veces: 0,

No: 1,

Si: 2

Frecuencia consumo guacamole

Cada 15 días: 0,

Cada mes: 1,

Cada semana: 2,

No aplica: 3,

Ocasionalmente: 4

Momentos de consumo del aguacate

Almuerzo: 0,

Cena: 1,

Desayuno: 2

¿Qué otros productos de aguacate consumes?

aceite de aguacate: 0,

guacamole: 1,

ninguno: 2,

salsa de aguacate: 3

¿Qué te motiva a consumir aguacate?'

Su sabor: 0,

Su versatilidad en la cocina: 1,

Sus beneficios para la salud: 2,

tendencias alimenticias: 3

¿Cómo sueles consumir aguacate principalmente?'

Como ingrediente en platos principales (sopas sándwiches etc.): 0,

En recetas dulces (batidos postres etc.): 1,

En recetas saladas (guacamole ensaladas tacos etc.): 2

¿Has probado alguna vez productos de aguacate no convencionales?'

Si: 1,

No: 0

Factores de compra

Calidad: 0,

Precio: 1

¿Te gustaría ver más variedades de productos de aguacate en el mercado?'

Si: 1,

No: 0

Tiene empleo

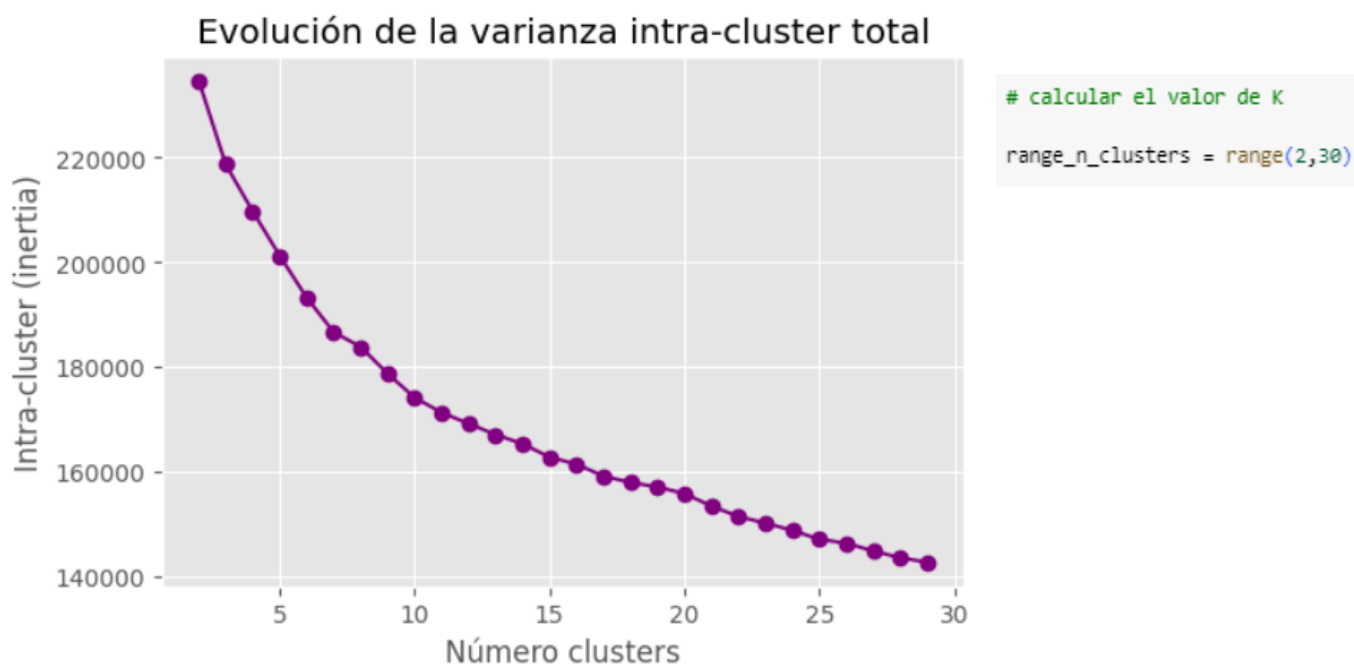
No:0

Si:1

Objetivo específico III: Desarrollar un modelo de clusterización con técnicas de aprendizaje no supervisado (Modelado).

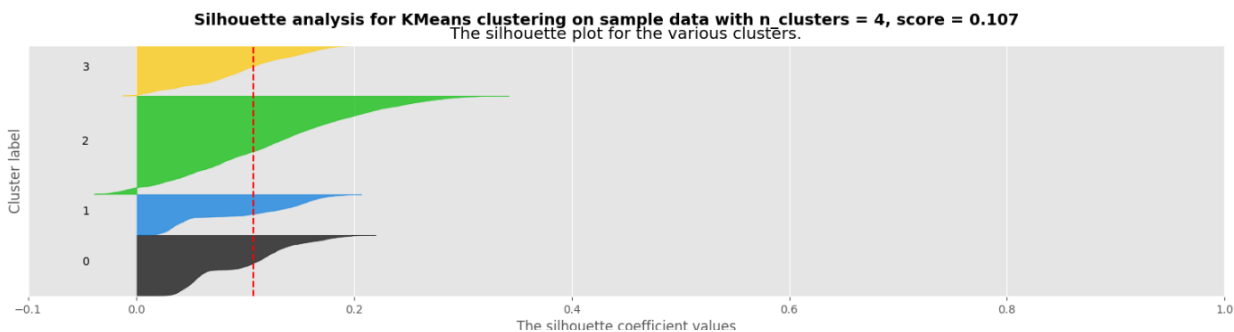
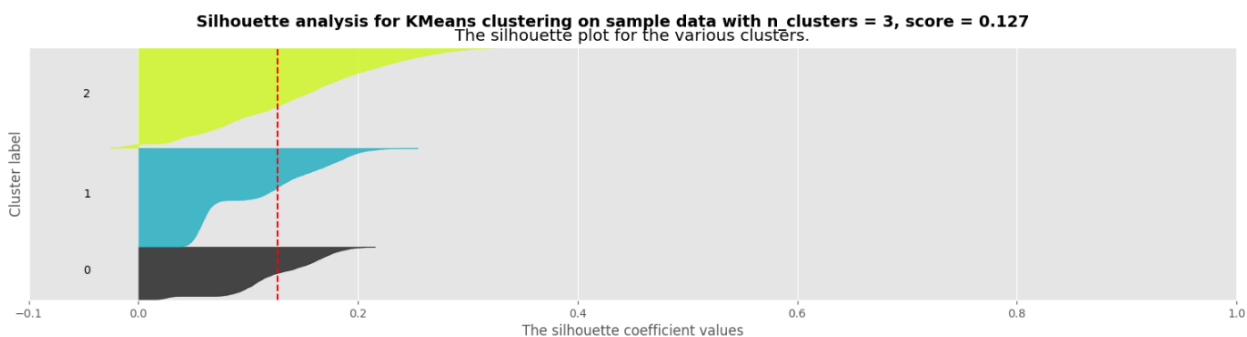
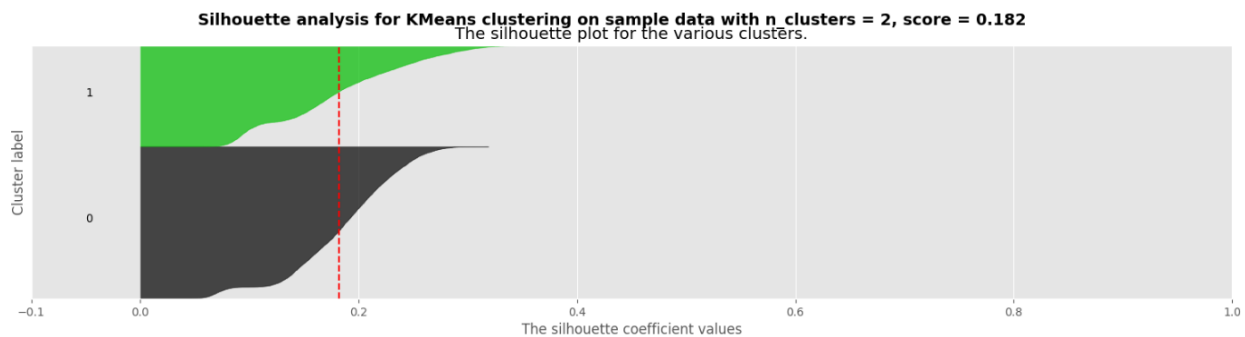
En esta fase, se desarrollará un modelo de clusterización utilizando técnicas de aprendizaje no supervisado. Se seleccionarán técnicas adecuadas, como K-means, y se implementará el modelo utilizando Google Colab con la base de datos procesada se realiza la creación del modelo y se ajustará el número de clústeres mediante técnicas como el método del codo y también el de la silueta analizando las características de cada clúster para comprender mejor los comportamientos de compra.

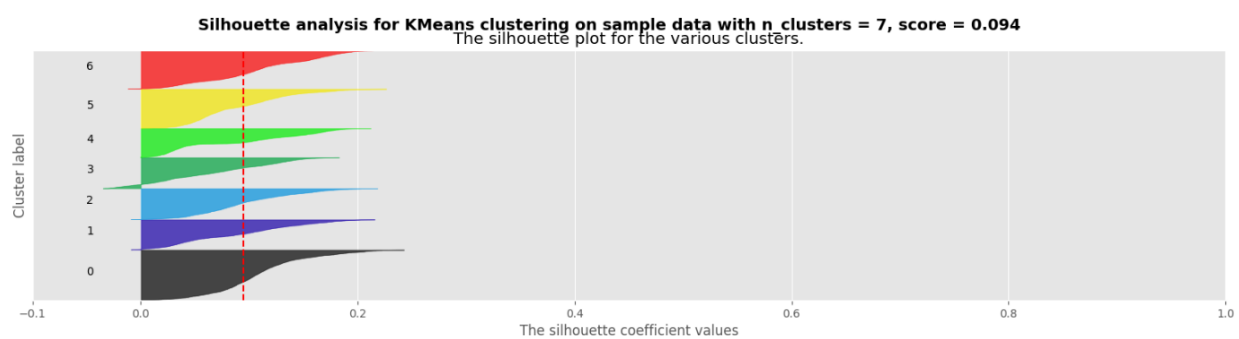
Iniciamos con la creación de nuestro modelo , como primera instancia calcularemos el valor de k Utilizando el método del codo y el método de la silueta , e identificar cual sería el número ideal de clusters.

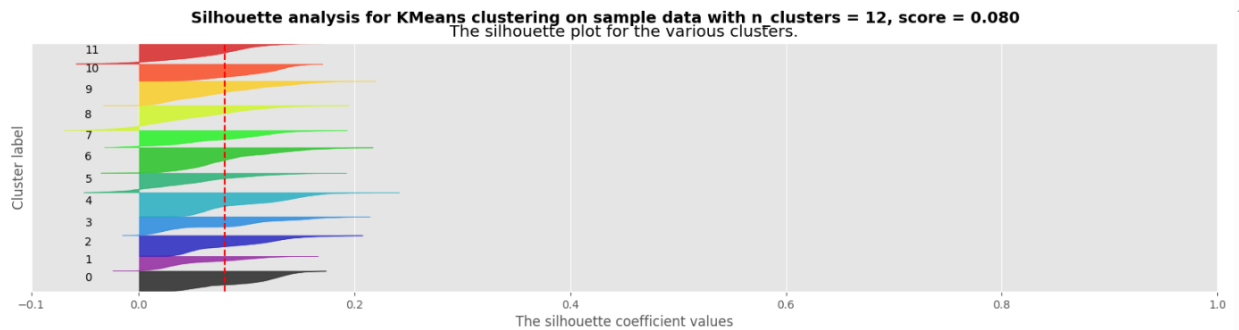
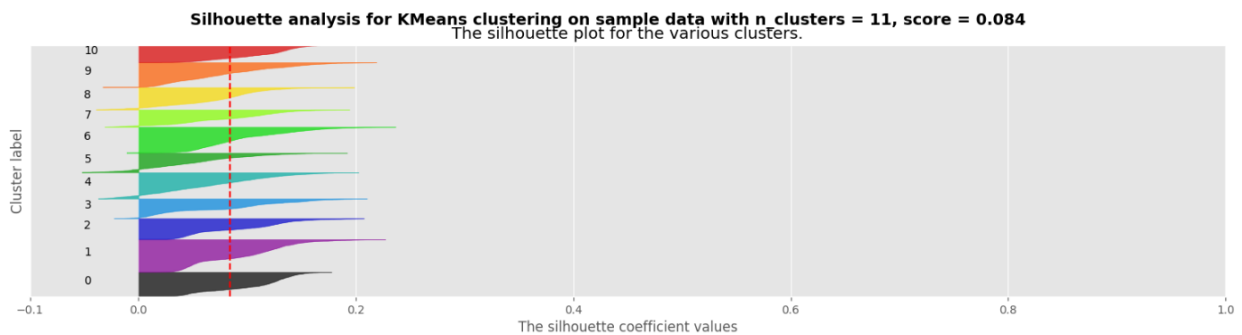
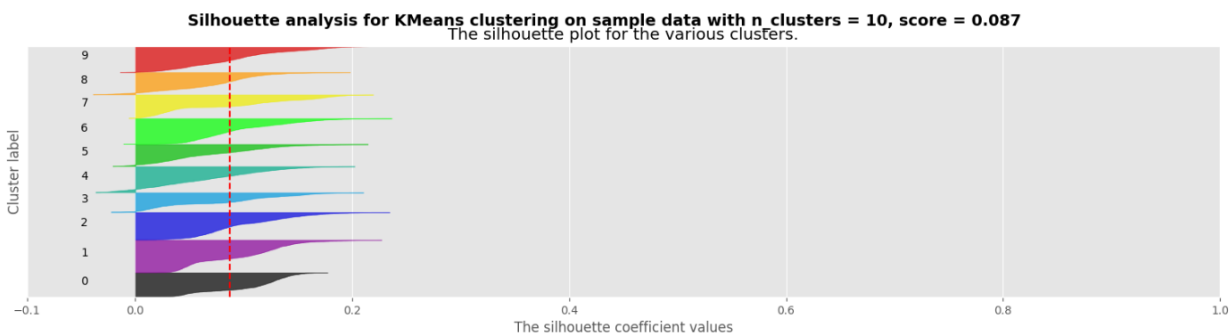


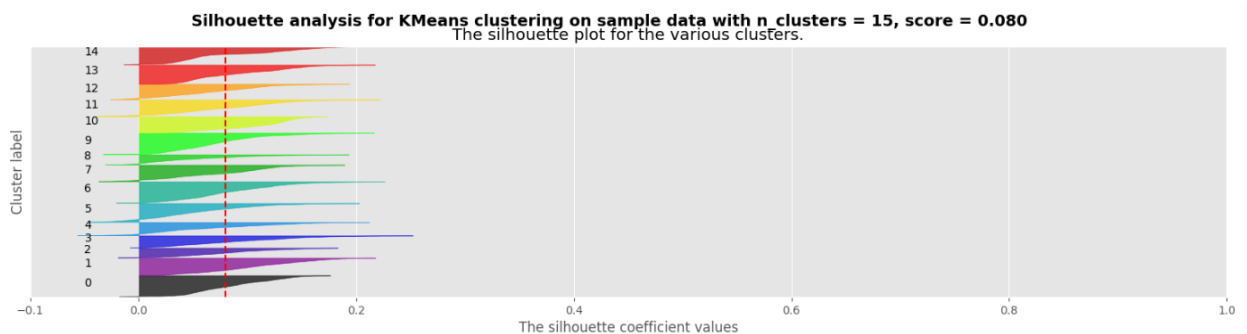
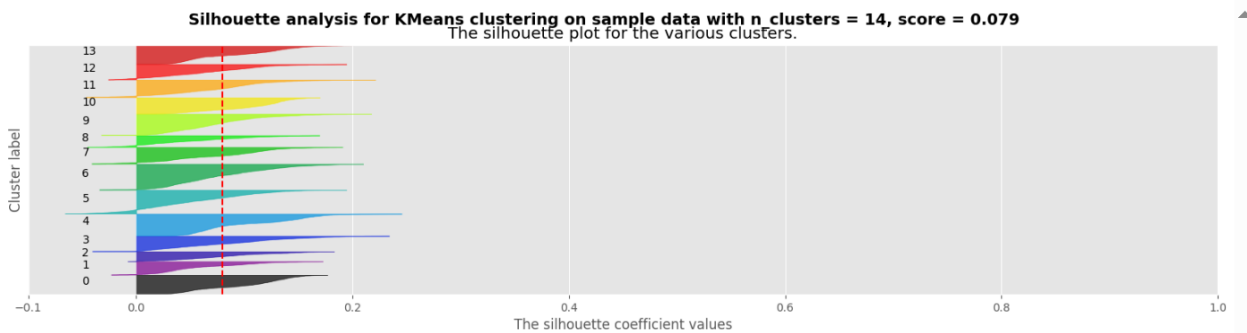
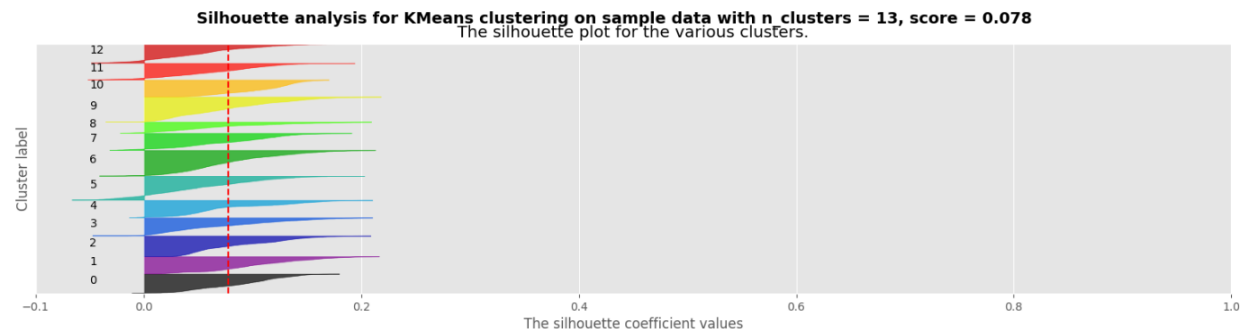
Al realizar el método del codo podemos evidenciar que posee un ligero corte en el número 7 pero con el fin de tener otra perspectiva acerca de la cantidad de clusters a utilizar en el método aplicamos el método de la silueta

```
silhouette_analysis(d, range_n_clusters)
```









Para el desarrollo de este ejercicio al evaluar los resultados de la silueta este nos arroja que el número óptimo de cluster a utilizar debería ser 2, pero consideramos que para la base de datos que tenemos y para lo que queremos realizar en este caso una segmentación de clientes para conocer el cliente del aguacate este número no es significativo, porque regresamos al resultado anterior con el método del codo y desarrollamos el modelo con 7 cluster

Decidimos trabajar con un porcentaje de datos del 80% para entrenamiento - 20% para la evaluación del modelo.

```
[48] # Dividimos los datos en 80% para entrenamiento y 20% para evaluación  
d_train, d_test = train_test_split(d, test_size=0.20, random_state=42)
```

```
▶ # Aplicar KMeans sobre los datos de entrenamiento  
n_clusters = 7  
kmeans = KMeans(n_clusters=n_clusters, random_state=42)  
kmeans.fit(d_train)
```



```
KMeans  
KMeans(n_clusters=7, random_state=42)
```

Como primera visualización a los cluster tenemos los grupos segmentados, con su respectivo centroide en este caso aplicamos T-SNE con el fin de reducir la dimensionalidad ya que al tener tantas variables se hace difícil la visualización de estos y buscamos la detección de los patrones y agrupación de los datos.

```

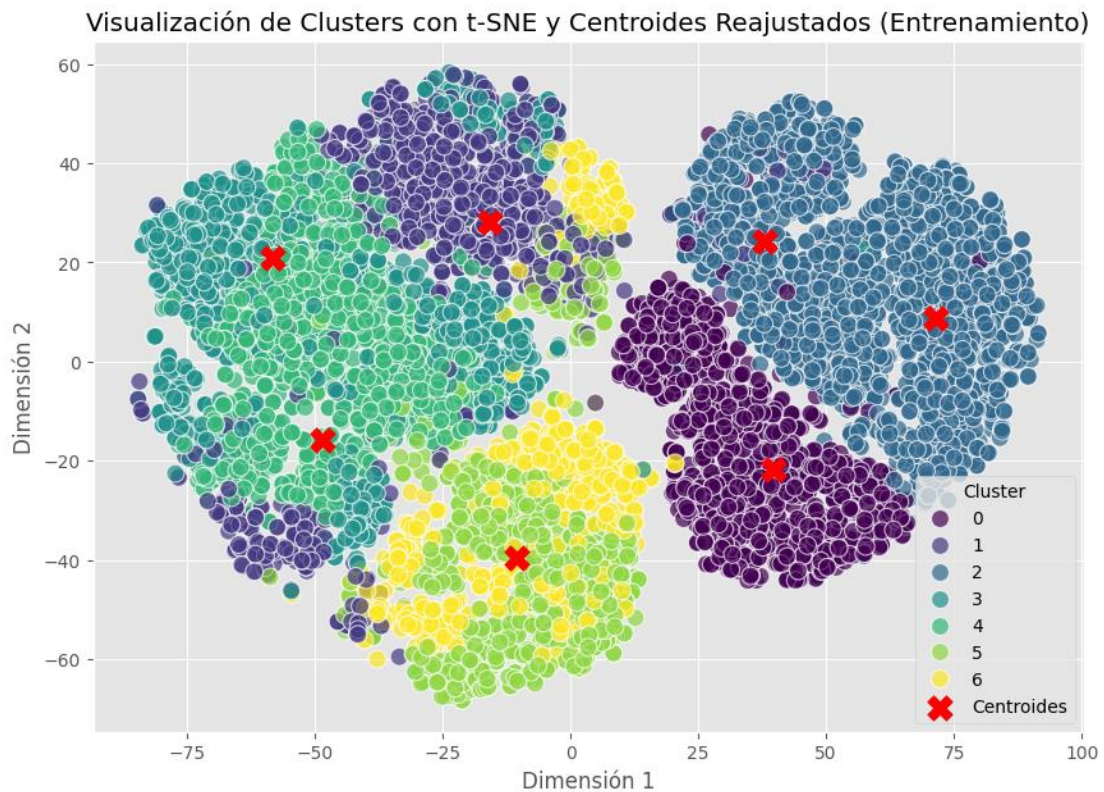
# Aplicar t-SNE sobre los datos de entrenamiento
tsne = TSNE(n_components=2, random_state=10)
d_tsne_train = tsne.fit_transform(d_train)

# Reajustar los centroides con KMeans usando los datos transformados con t-SNE
kmeans_tsne = KMeans(n_clusters=n_clusters, random_state=42)
kmeans_tsne.fit(d_tsne_train)
centroids_tsne = kmeans_tsne.cluster_centers_

# Graficar los clusters y los centroides recalculados en el espacio t-SNE
plt.figure(figsize=(10, 7))
sns.scatterplot(x=d_tsne_train[:, 0], y=d_tsne_train[:, 1], hue=dClus_train['cluster'], palette='viridis', s=100, alpha=0.7)

# Graficamos los centroides recalculados
plt.scatter(centroids_tsne[:, 0], centroids_tsne[:, 1], c='red', s=200, marker='X', label='Centroides')
# Detalles de la gráfica
plt.title('Visualización de Clusters con t-SNE y Centroides Reajustados (Entrenamiento)')
plt.xlabel('Dimensión 1')
plt.ylabel('Dimensión 2')
plt.legend(title='Cluster')
plt.show()

```

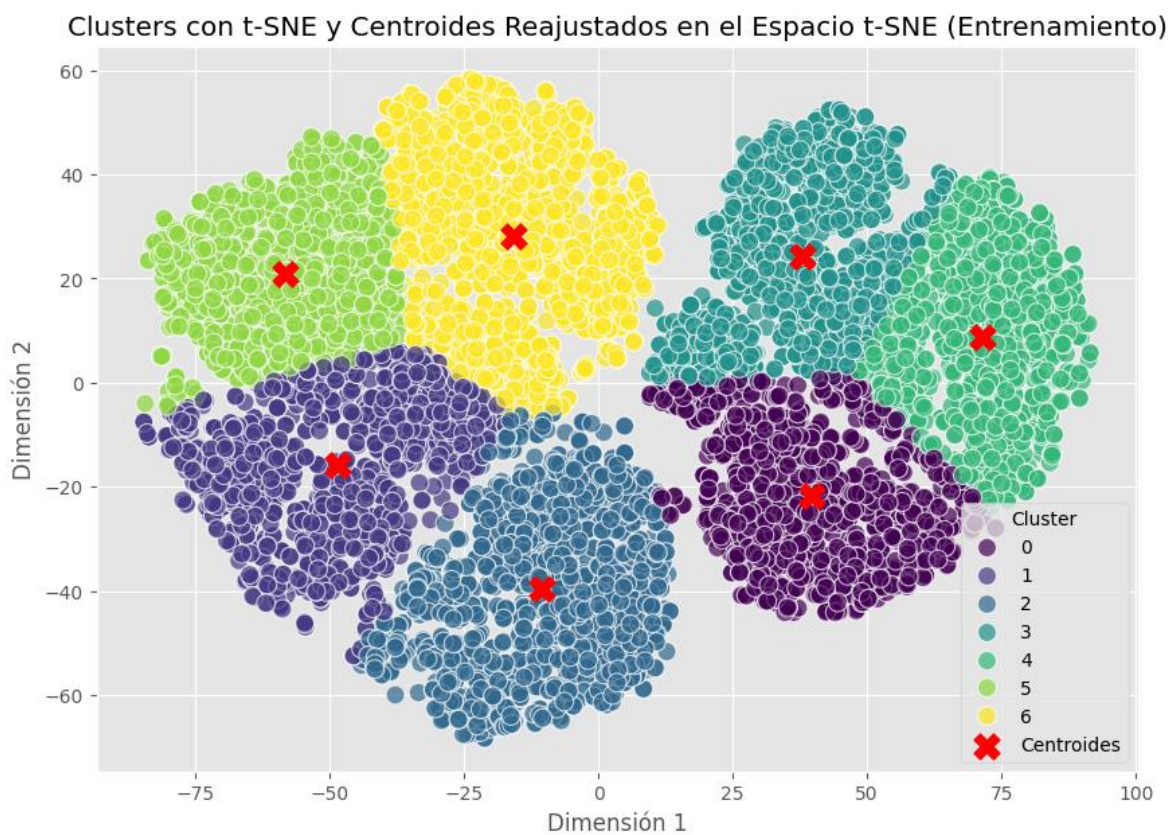


Se puede evidenciar que el modelo no se encuentra bien ajustado, y los grupos no son muy claros por lo que es necesario realizar un ajuste, realizando una reducción de datos y recalculando los centroides, obteniendo el siguiente resultado.

```
# Volver a aplicar KMeans sobre los datos reducidos con t-SNE (entrenamiento)
kmeans_tsne = KMeans(n_clusters=n_clusters, random_state=42)
dClus_train['cluster_tsne'] = kmeans_tsne.fit_predict(d_tsne_train)
# Visualizar los clusters en el espacio t-SNE
plt.figure(figsize=(10, 7))
sns.scatterplot(x=d_tsne_train[:, 0], y=d_tsne_train[:, 1], hue=dClus_train['cluster_tsne'], palette='viridis', s=100, alpha=0.7)

# Graficar los centroides recalculados en el espacio t-SNE
centroids_tsne = kmeans_tsne.cluster_centers_
plt.scatter(centroids_tsne[:, 0], centroids_tsne[:, 1], c='red', s=200, marker='X', label='Centroides')

# Añadir detalles de la gráfica
plt.title('Clusters con t-SNE y Centroides Reajustados en el Espacio t-SNE (Entrenamiento)')
plt.xlabel('Dimensión 1')
plt.ylabel('Dimensión 2')
plt.legend(title='Cluster')
plt.show()
```



Es claro que los puntos pudieron acomodarse de forma correcta a su grupo y centroide más cercano por lo que consideramos que este es el modelo correcto para nuestra segmentación.

Una vez terminado el modelo procedemos a evaluar nuestros segmentos, para determinar nuestro cliente ideal basados en las variables de nuestro data set, la relación entre estas y la agrupación realizada con el modelo.

Como lo mencionamos anteriormente las variables de integrantes en el hogar y la cantidad de kilos que compran no presentan una alta relación entre sí por lo que es necesario evaluar otras variables e identificar cuáles tienen mayor peso para nuestro buyer persona

Objetivo específico IV: Evaluar el modelo construido con datos reales de la compañía

Procedemos con la evaluación del modelo con el 20% de los datos de que dejamos para esto aplicamos nuevamente el código con los datos de prueba

Obteniendo los siguientes resultados:

```

# Aplicar t-SNE sobre los datos de prueba
d_tsne_test = tsne.fit_transform(d_test)

# Predecir clusters con el modelo de KMeans previamente entrenado
dClus_test = d_test.copy()
dClus_test['cluster'] = kmeans.predict(d_test)

# Visualizar los clusters en el espacio t-SNE para los datos de prueba
plt.figure(figsize=(10, 7))
sns.scatterplot(x=d_tsne_test[:, 0], y=d_tsne_test[:, 1], hue=dClus_test['cluster'], palette='viridis', s=100, alpha=0.7)

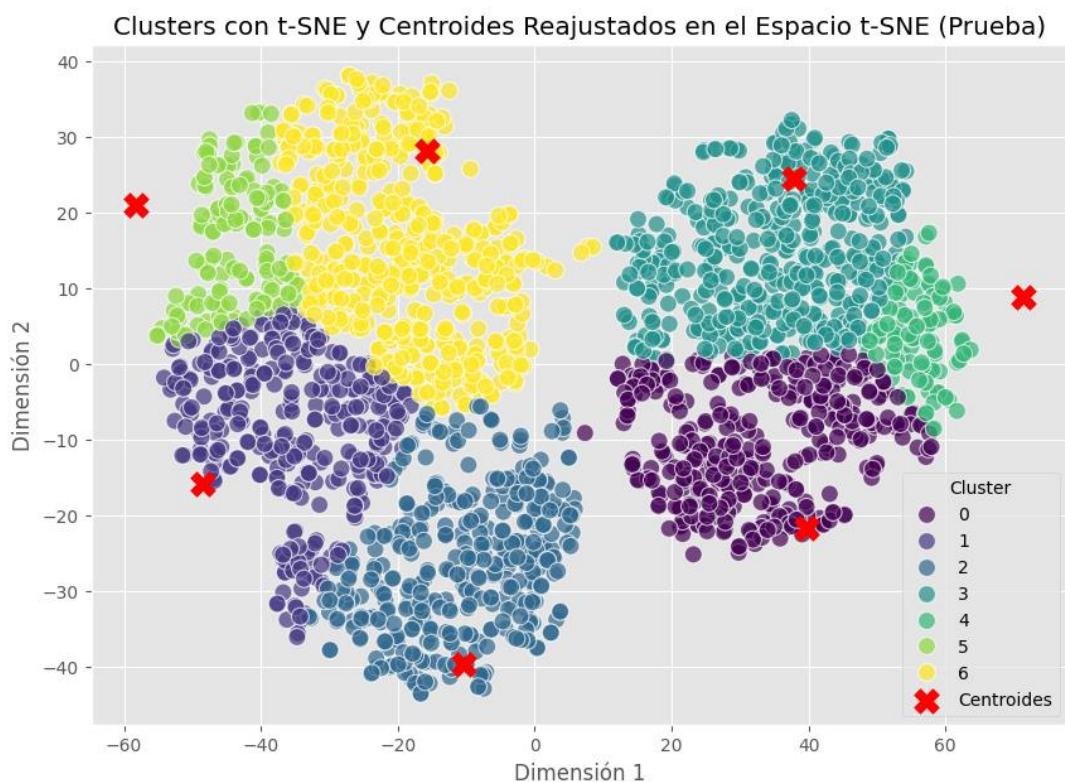
# Predecir los clusters en el espacio t-SNE usando el modelo de KMeans entrenado
dClus_test['cluster_tsne'] = kmeans_tsne.predict(d_tsne_test)

# Visualizamos los clusters en el espacio t-SNE para los datos de prueba
plt.figure(figsize=(10, 7))
sns.scatterplot(x=d_tsne_test[:, 0], y=d_tsne_test[:, 1], hue=dClus_test['cluster_tsne'], palette='viridis', s=100, alpha=0.7)

# Graficar los centroides recalculados en el espacio t-SNE (prueba)
plt.scatter(centroids_tsne[:, 0], centroids_tsne[:, 1], c='red', s=200, marker='X', label='Centroides')

# Añadir detalles de la gráfica
plt.title('Clusters con t-SNE y Centroides Reajustados en el Espacio t-SNE (Prueba)')
plt.xlabel('Dimensión 1')
plt.ylabel('Dimensión 2')
plt.legend(title='Cluster')
plt.show()

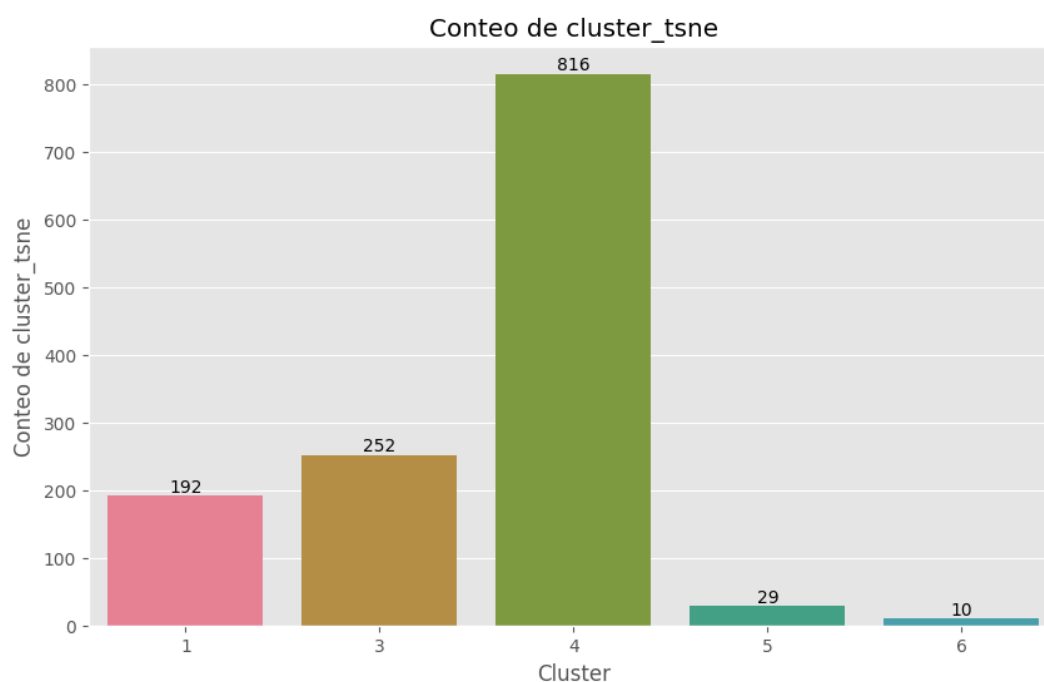
```



La porción de datos que dejamos para realizar la prueba pudo acomodarse de forma sencilla a los centroides o grupos ya establecidos en el entrenamiento del modelo por lo que al agregar nueva información al modelo este podrá actualizarse sin problema y realizar la correcta segmentación del nuevo cliente.

Discusión

Una vez realizado nuestro modelo obtenemos los siguientes resultados



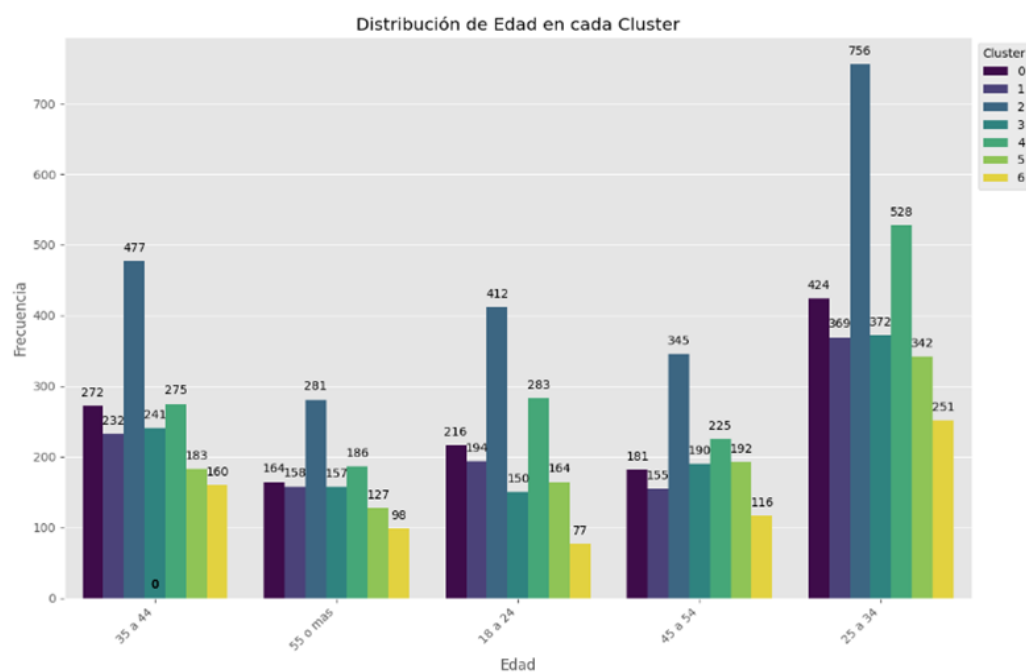
Según el T-SNE el clúster con una mayor participación de los encuestados es el cluster 4 porque usaremos este cluster para la construcción del Buyer Person, el cual nombraremos como

Clúster 4 - "Consumidores jóvenes ocasionales"

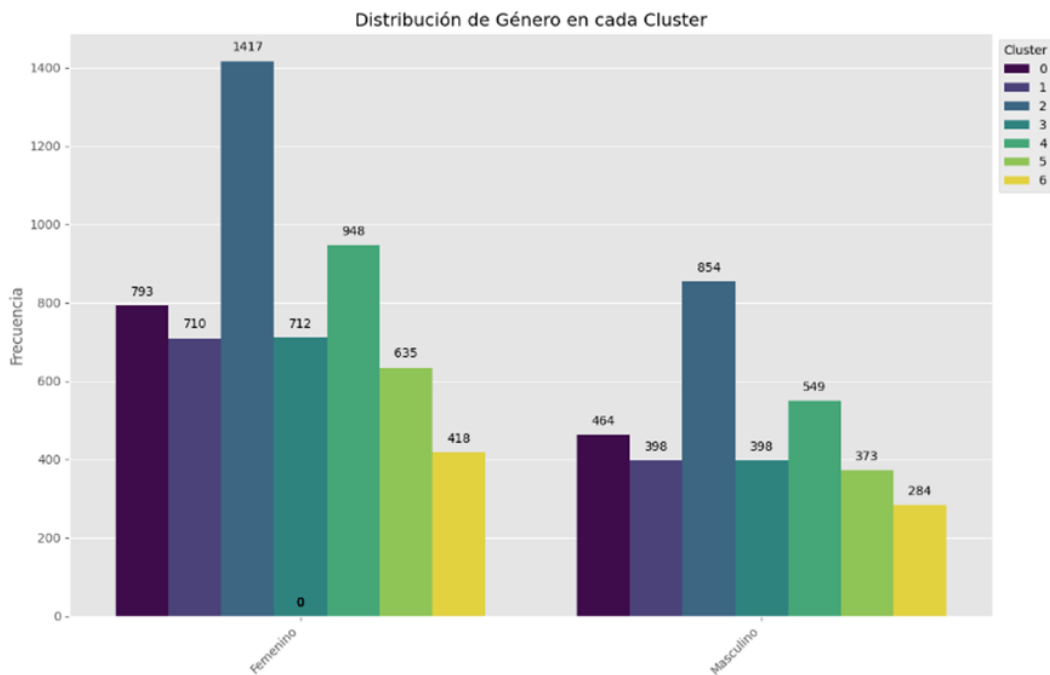
Edad: Predominantemente jóvenes entre 25 y 34 años. De género femenino con nivel de escolaridad tecnológico y profesional, son personas empleadas, prefieren el aguacate Hass con un Consumo de cada 15 días, suele comprar el producto en supermercados y almacenes de cadena al momento de la compra están fuertemente motivados por el precio de la fruta, consumen productos derivados como el guacamole de forma ocasional, Prefieren consumir el aguacate en

horas del almuerzo les gusta consumirlo en recetas saladas, pero no suele consumir otros productos derivados además del guacamole pero a pesar de esto están abiertos a probar y conocer nuevos productos Se encuentran motivados por los beneficios para la salud que tiene el aguacate Hass.

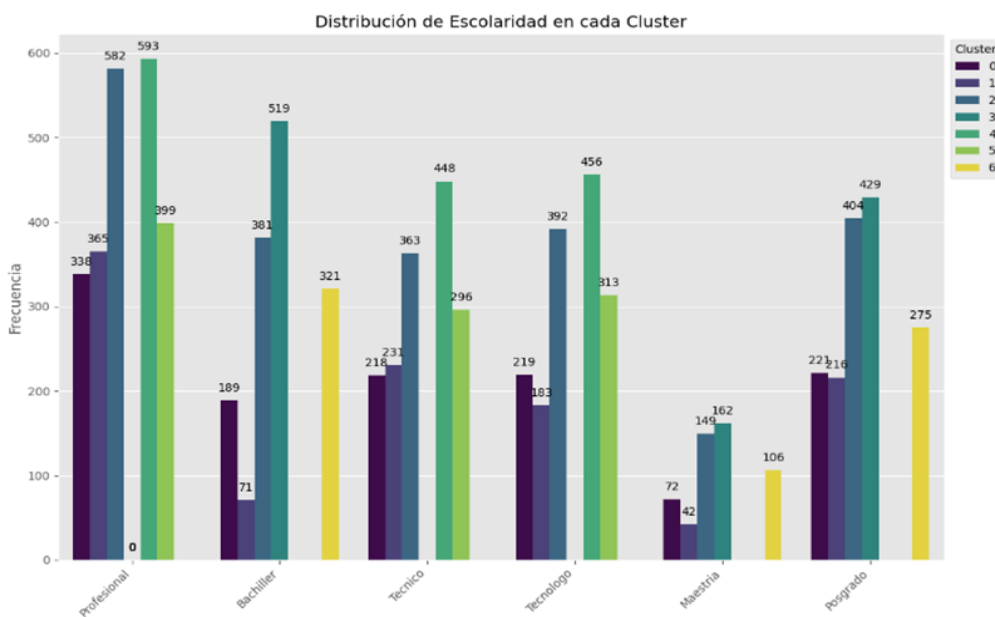
Además de encontrar el Buyer Person se realiza la segmentación de los demás grupos con el fin de realizar las respectivas recomendaciones para la empresa, se presentan los resultados generales de cada variable para proceder con la interpretación de los clusters.



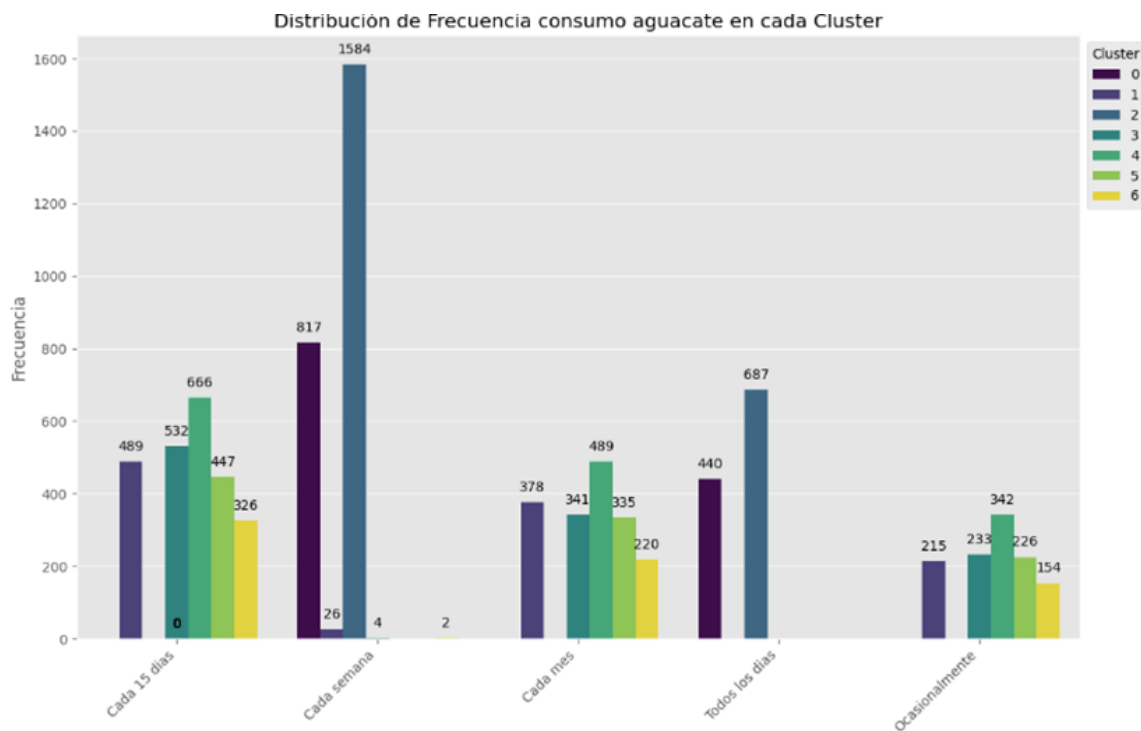
La edad presenta variación entre los diferentes clusters pero la que tiene mayor representación está entre los 25 y los 34 años perteneciendo a la generación Millennial.



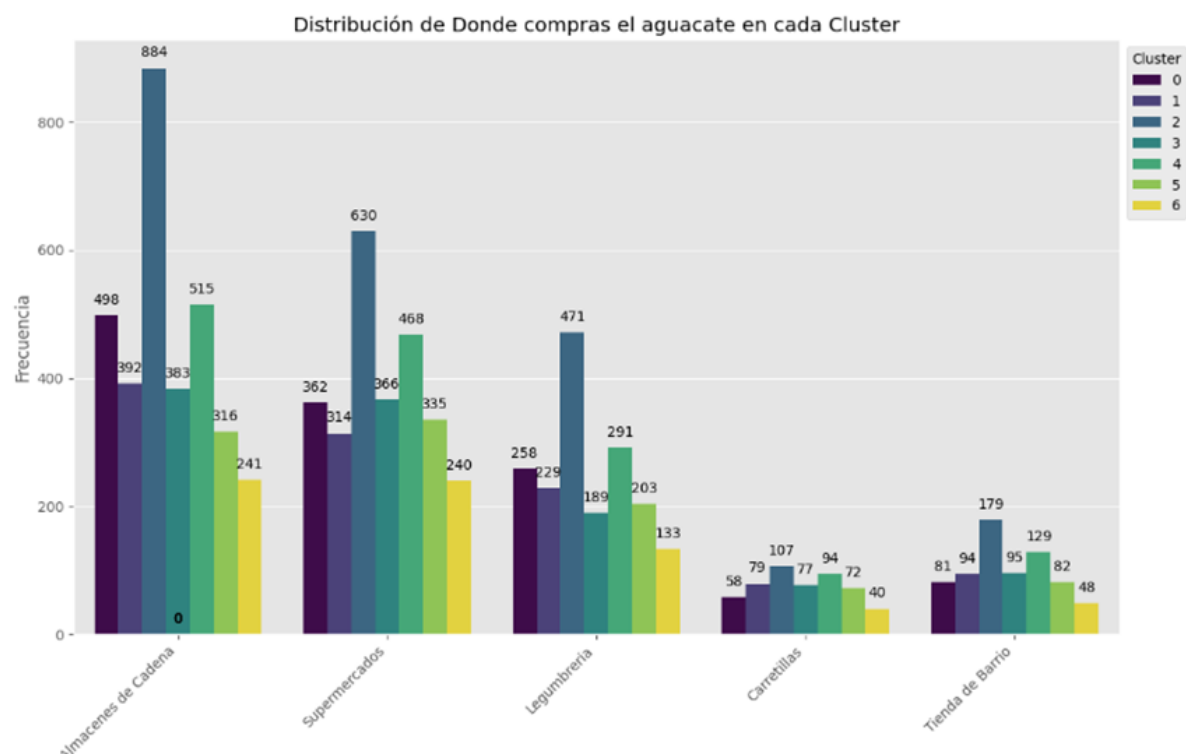
Es clave resaltar que el género femenino tiene un alto impacto en las encuestas es importante mencionar que las mujeres tienden a ser más participativas en concurso de las redes sociales y realizar las compras del hogar por lo que estas razones pudieron influir en los resultados.



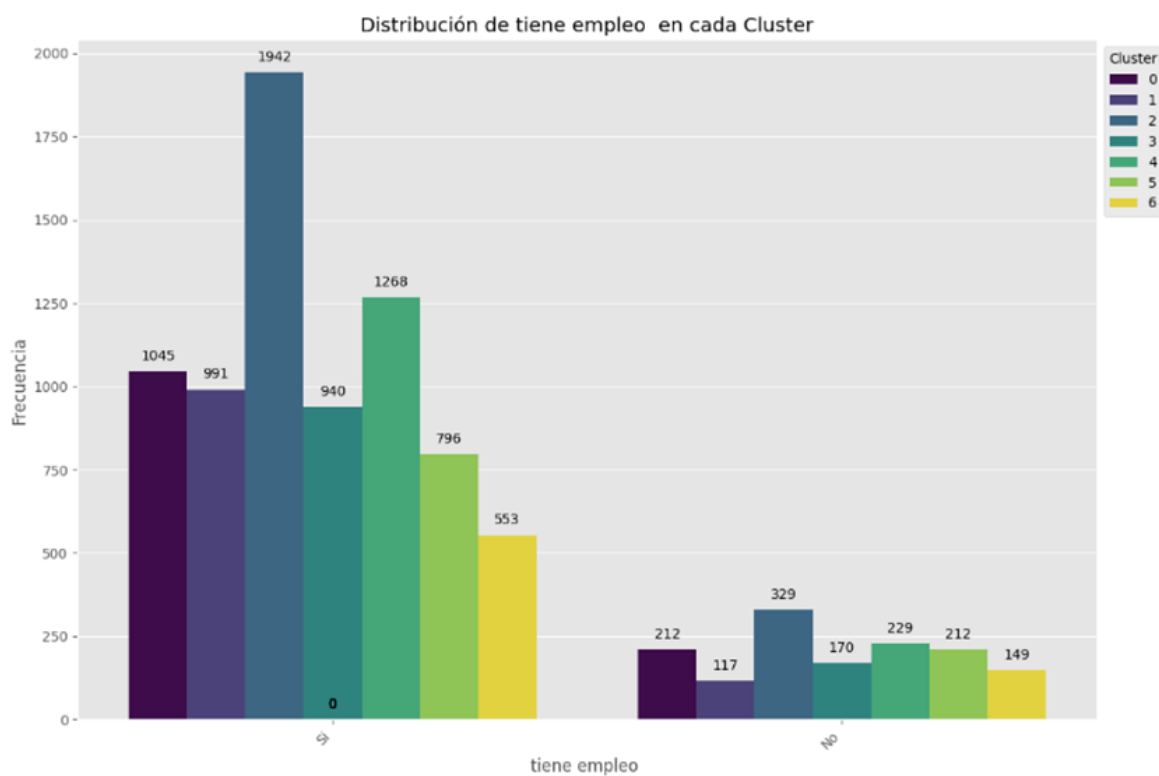
La distribución de la escolaridad es muy variable pero la mayoría de las personas encuestadas han realizado técnicas, tecnologías o son Profesionales mientras que las personas con maestría no son muy influyentes en los resultados.



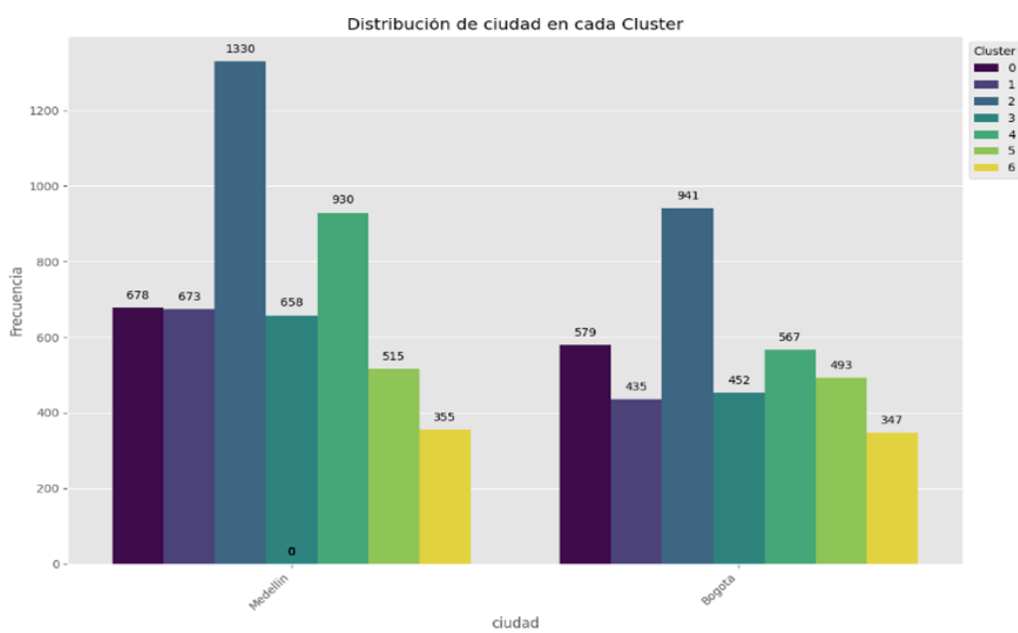
Se logra evidenciar que solo el cluster 2 suele consumir aguacate de forma semanal y muy pocas personas consumen esta fruta de forma diaria, en todos los resultados encontramos que los valores más representativos se encuentran en frecuencias de cada 15 días o de forma mensual.



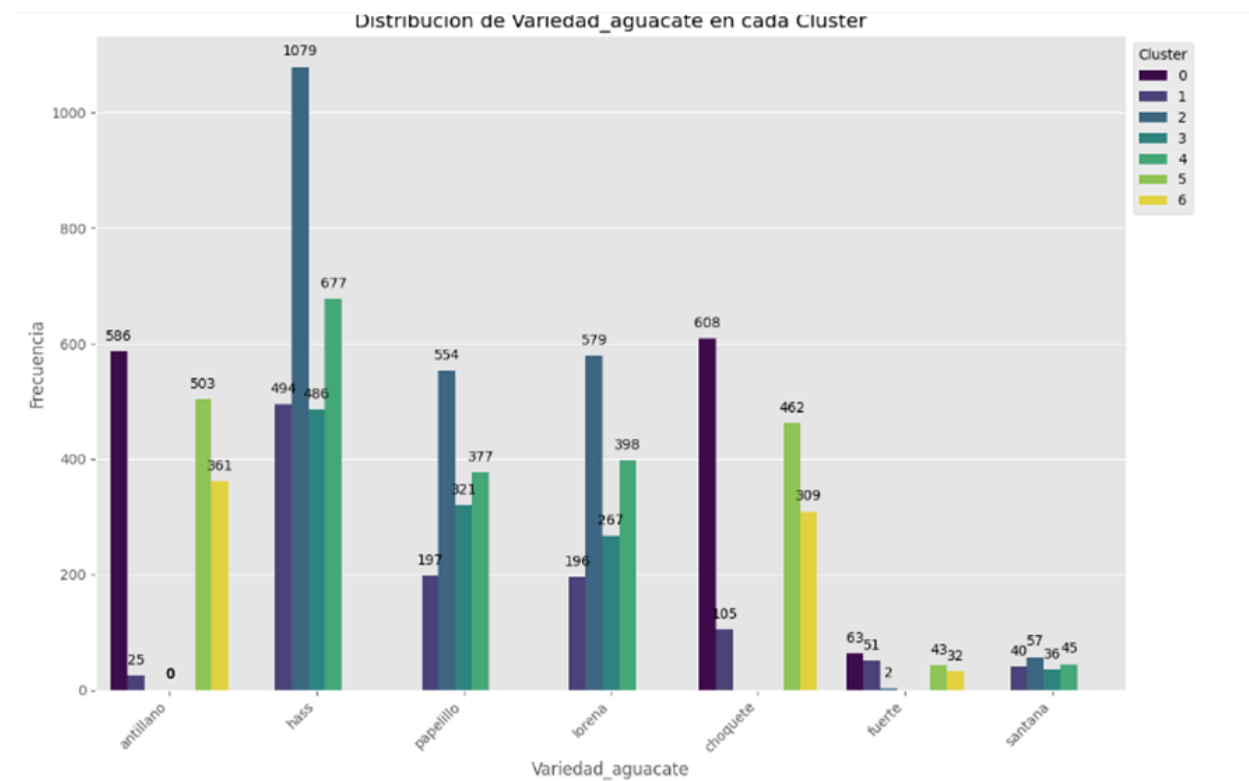
En la gráfica anterior evidenciamos que las personas suelen comprar su aguacate en almacenes de cadena y supermercados actualmente Cartama solo logra impactar en almacenes de cadena pues no ha logrado entrar con precios competitivos a los supermercados estos serían un buen nicho para la venta.



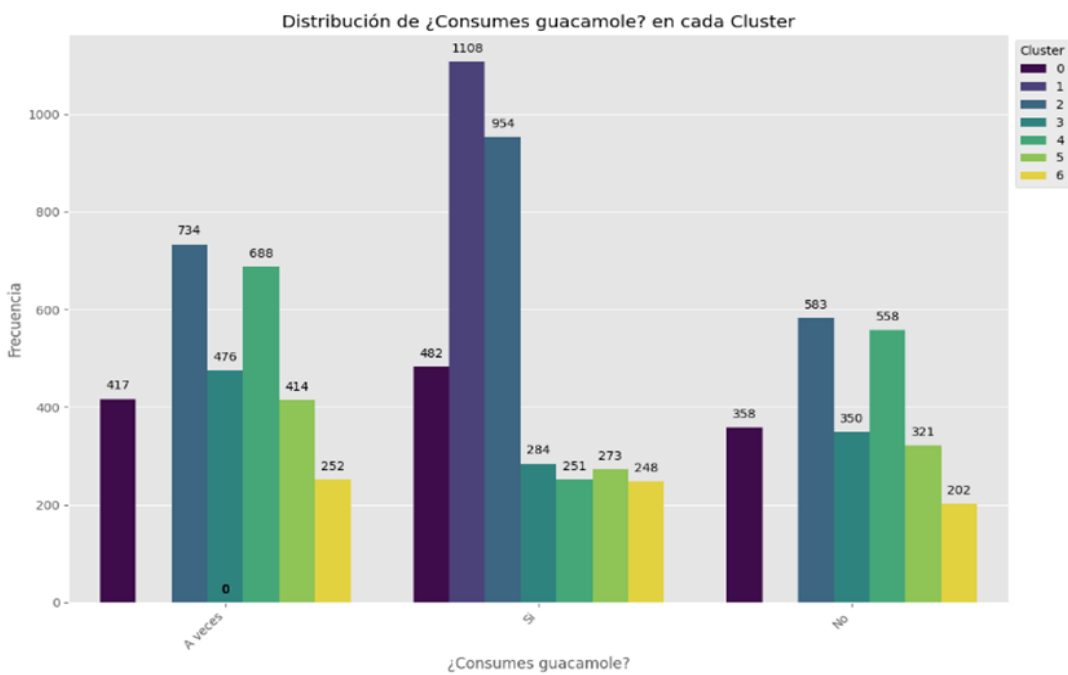
Evidentemente la porción más grande se la llevan las personas empleadas con un total de 7.535 personas.



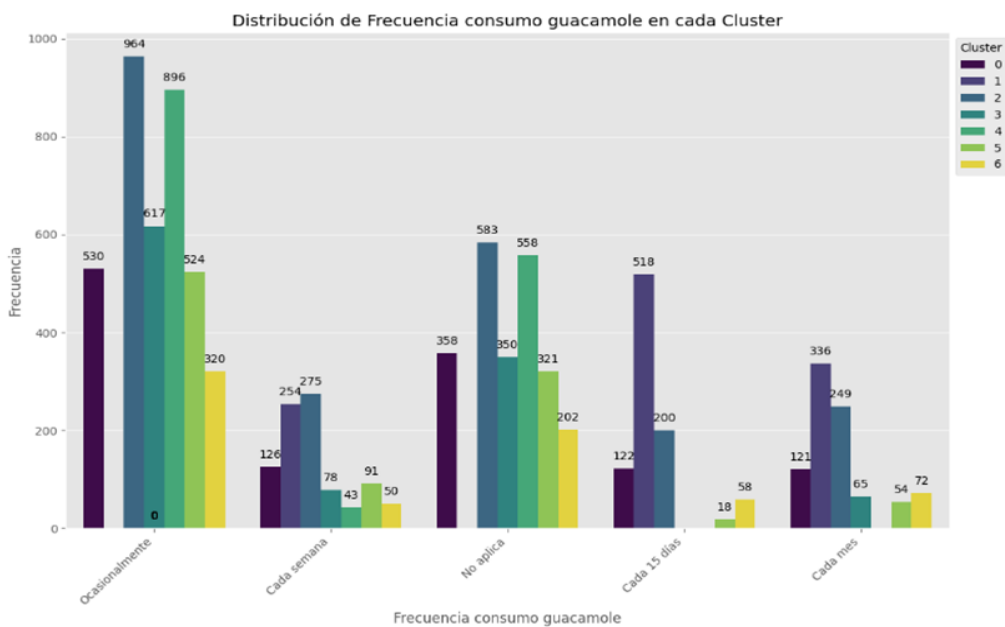
Cartama actualmente cuenta con distribución en dos ciudades del país Medellín y Bogotá los resultados en realidad muestran mucha homogeneidad por lo que no son tan representativos los resultados, se podría comenzar a explorar nuevas ciudades



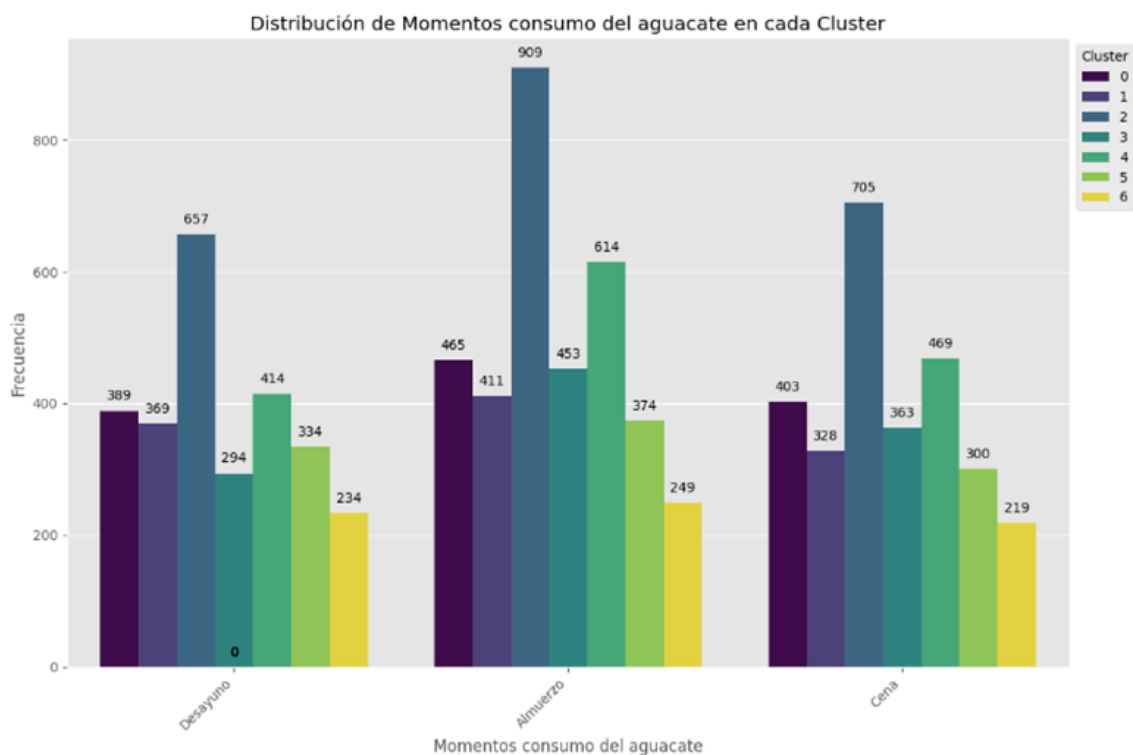
El aguacate Hass es una de los preferidos por los encuestados siendo un buen resultado para Cartama con 2.736 respuestas a favor ya que solo produce esta variedad pero el aguacate papellillo y lorena son su más fuerte competidor si sumamos ambos resultados es un total de 2.889 resultados estas dos variedades son muy parecidas en cuanto a sus características físicas por lo que suelen confundirse, estas variedades son competencia directa para la empresa



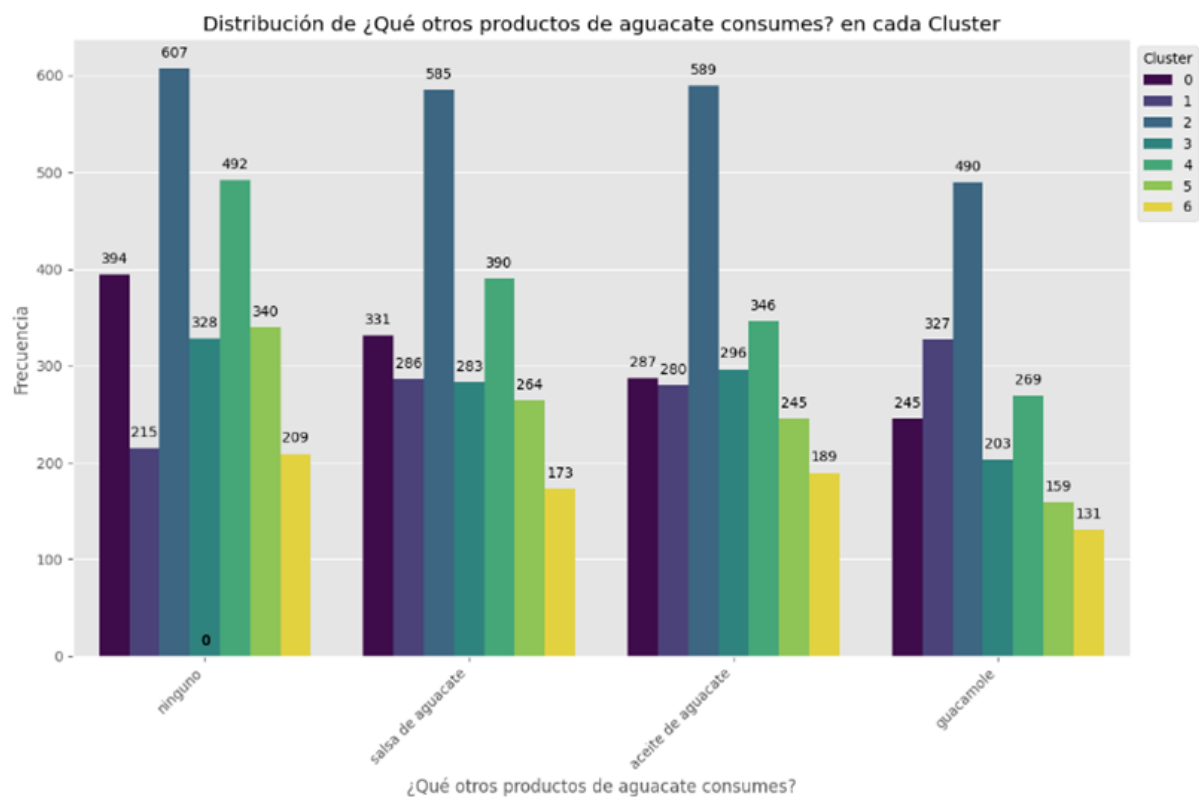
Las personas suelen consumir guacamole pues podemos tomar la respuesta a veces como un consumo afirmativo del guacamole, por lo que este derivado podría representar una buena participación en ventas.



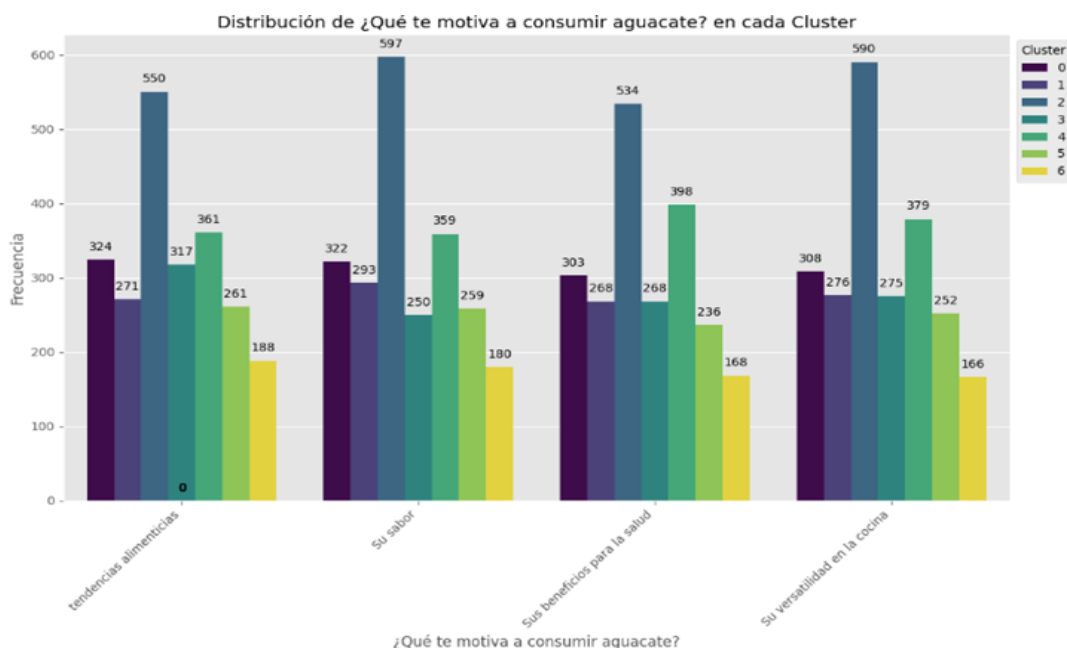
La frecuencia de consumo del guacamole como derivado de la fruta se encuentra en un promedio de forma ocasional, por lo que se debe incentivar el consumo de este mostrando sus diferentes bondades.



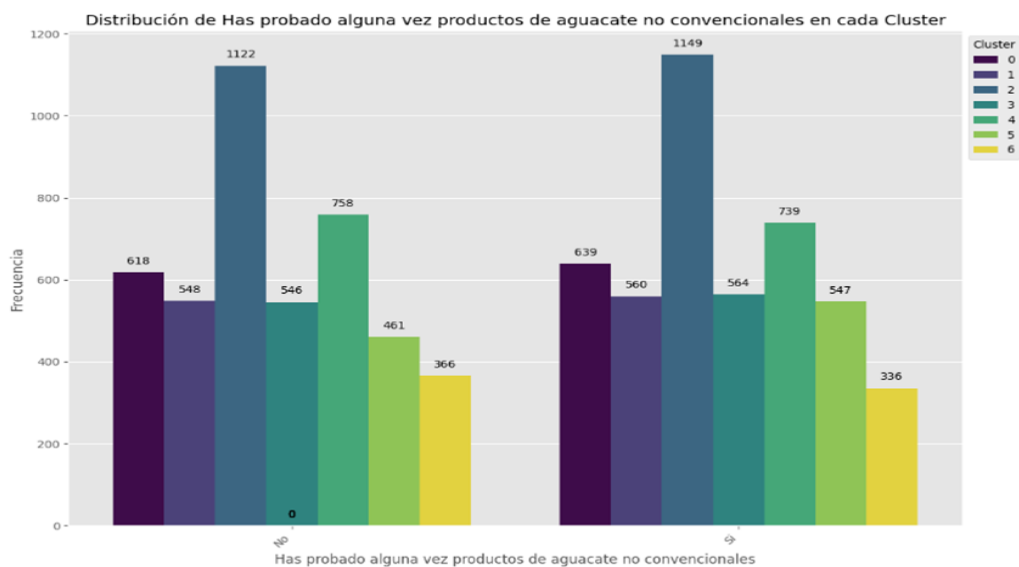
Los momentos de consumo son claves para el desarrollo de nuevas campañas de marketing, aunque claramente existe poca variedad en los resultados el almuerzo sigue siendo el momento de consumo con mayor participación, se pueden explorar los demás momentos impactando a la audiencia con recetas o campañas publicitarias enfocadas en esos dos momentos del día.



Para esta encuesta se establecen posibles productos que puede están consumiendo las personas, se obtuvieron buenos resultados muy homogéneos para la salsa, aceite y guacamole, pero hay un número de 2.585 personas que no consume ninguno de los productos es un número muy amplio de personas que se pueden impactar con las campañas adecuadas.

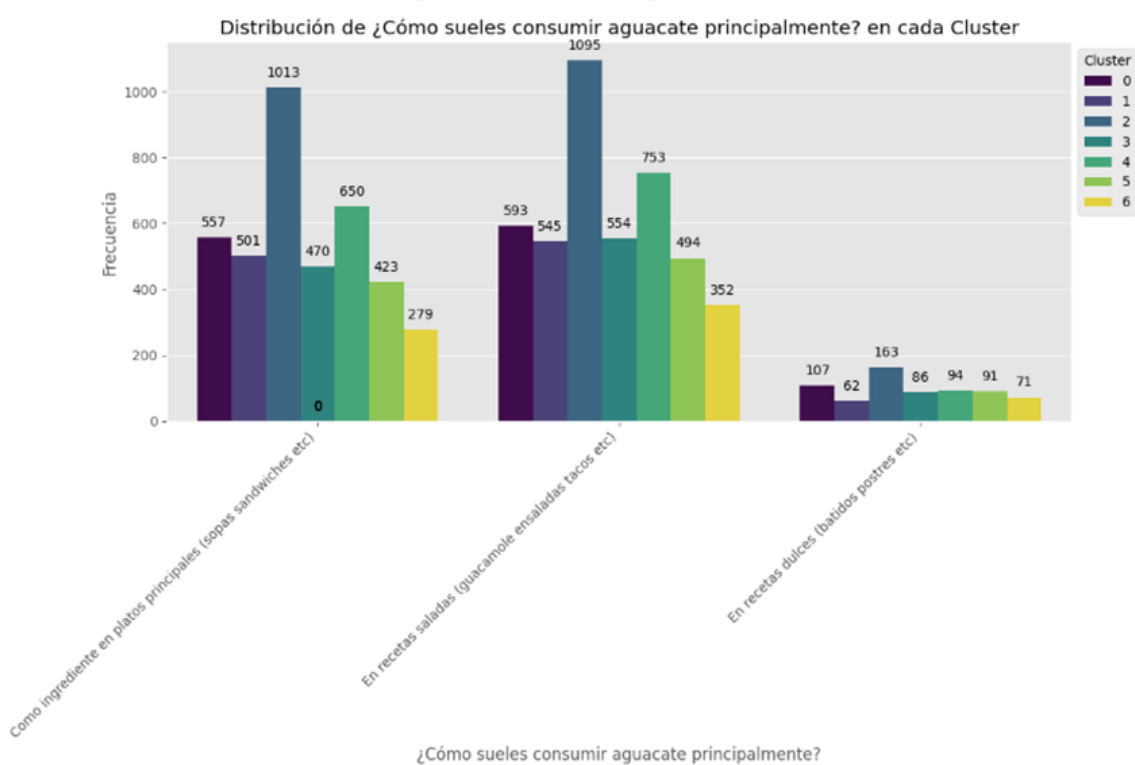


Al igual que la gráfica anterior las respuestas no presentan mucha variabilidad, pero es clave que lo que más motiva a los consumidores son las tendencias alimenticias con 2.273 respuestas y el sabor de la fruta con 2.010, resaltando que el aguacate Hass el mejor en sabor respecto a las demás variedades.

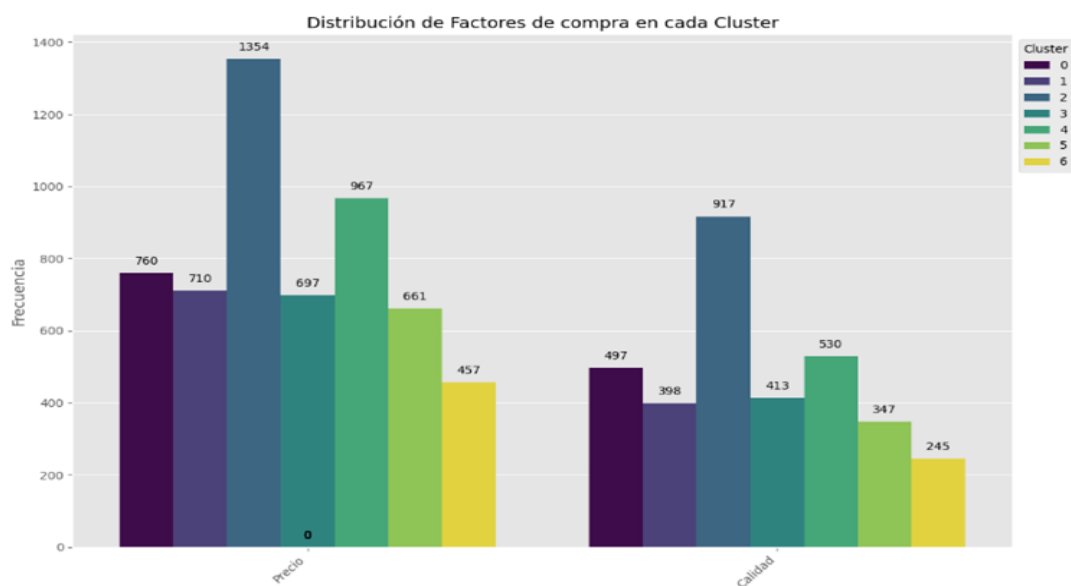


para este caso los resultados fueron demasiado cercanos, los encuestados reportan que

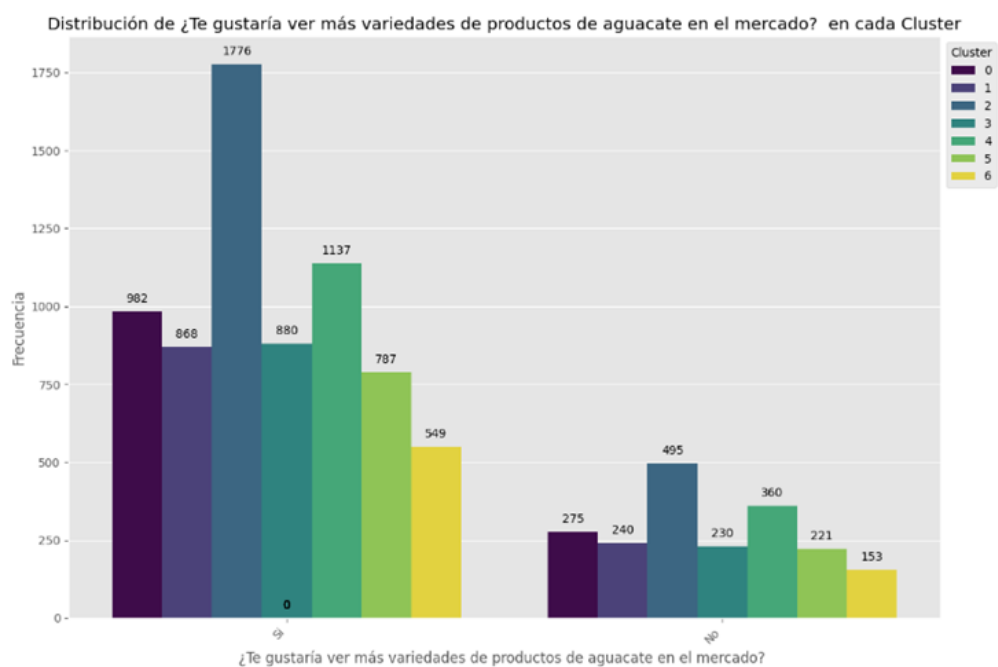
4419 personas no han probado productos no convencionales mientras que 4534 personas reportan que sí, es claro que la porción de personas que no la han probado es muy alta es una oportunidad para expandir el mercado para Cartama.



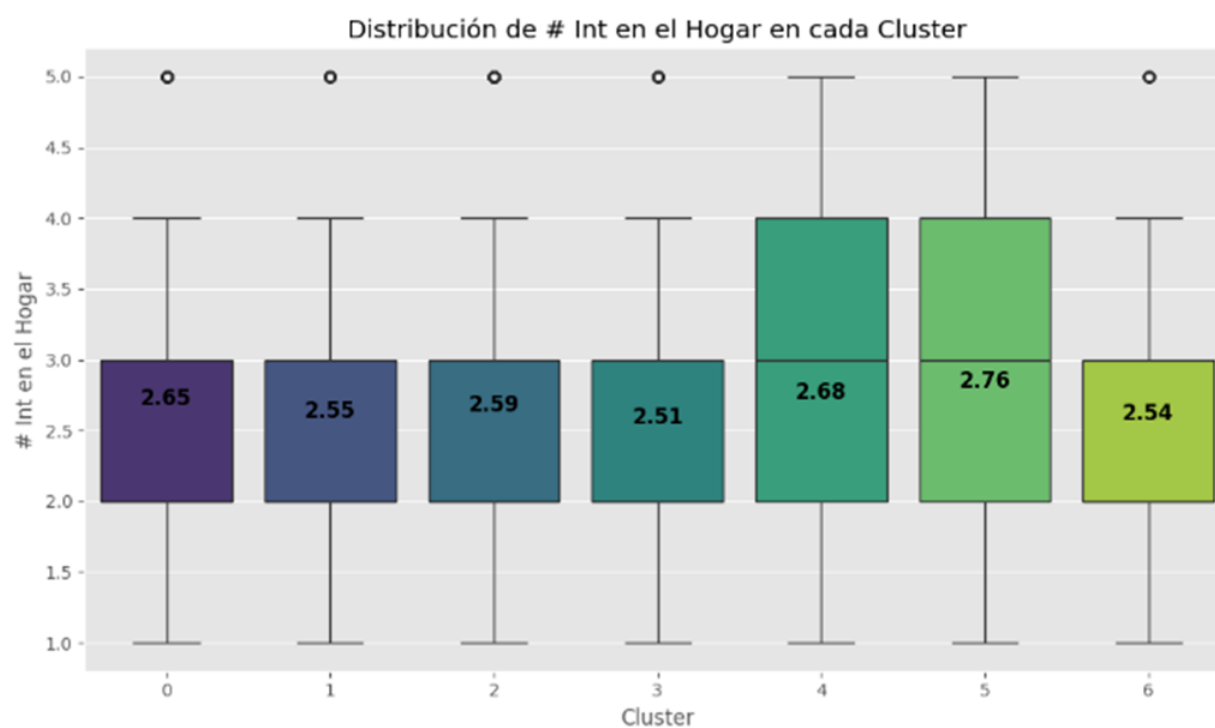
Las recetas dulces no son un fuerte para el aguacate como los batidos y postres, las personas prefieren consumir esta fruta en platos principales y en recetas saladas.



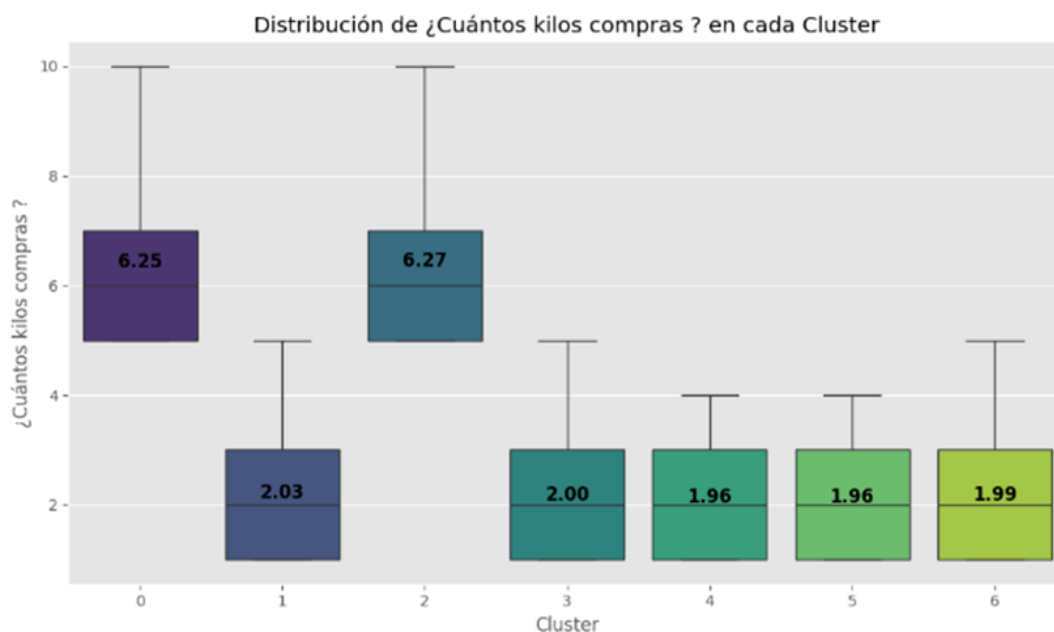
Uno de los factores con mayor participación para los encuestados es el precio, por lo que este es un factor clave para la economía de los encuestados sin dejar a un lado la calidad que también fue un factor de gran peso.



A pesar de que muchos encuestados no han probado productos no convencionales del aguacate es claro que están dispuestos a probarlos con un total de 6.919 personas dispuestas a consumirlos.



La distribución de las personas que conforman los hogares se encuentra en un promedio de 2 a tres personas.



Por último la cantidad de kilos que consumen cada uno de esos hogares se encuentra en un promedio de dos kilos, teniendo en cuenta que un aguacate Hass de calidad se encuentra entre los 165 gramos es un promedio de compra de 10 a 12 aguacates por hogar .

Una vez analizamos todos los resultados por variable procedemos a la descripción de los demás grupos, exceptuando el clúster 4 el cual ya fue descrito con anterioridad como nuestro Buyer Person.

Resaltamos que todas nuestras recomendaciones se encuentran acompañadas de estrategias de Marketing que es lo que se busca realizar con el modelo

Clúster 0 - "Consumidores Prácticos"

Variedad de edades, pero con una mayoría en rangos de 35 a 54 años, tanto hombres como mujeres, profesionales y empleados con frecuencia de consumo semanal, suele realizar sus

compras en almacenes de cadena y supermercados. Suelen consumir variedad: Antillano y Choquete en almuerzos y cenas

El 50% de las personas les gusta consumir guacamole y lo hacen de forma ocasional
Motivaciones de consumo: Preferencia sabor y versatilidad prefieren las recetas saladas además del guacamole no suelen consumir más productos derivados de esta fruta, le motiva consumir aguacate por su sabor y las tendencias alimenticias y por el precio del aguacate lo cual se evidencia en la variedad que consumen ya que el antillano y Choquete son mucho más económicos que el aguacate Hass.

Recomendaciones:

Realizar campañas de contenido que destaque los beneficios nutricionales del aguacate Has y realizar comparativas con las otras variedades para enfatizar su calidad y sabor superior y a que se debe su calidad como las buenas prácticas agrícolas y sostenibles que actualmente maneja Cartama, desarrollar campañas de recetas con colaboraciones con influencers o chefs locales que muestran la facilidad de consumir aguacate en los 3 momentos del día con recetas innovadoras

Clúster 1 “Consumidores Tradicionales”

Edad entre los 25 a los 34 años, Profesionales, empleados, consumen aguacate de forma quincenal, suelen comparar en almacenes de cadena y supermercados, Prefieren el aguacate Hass, suelen consumir guacamole con una frecuencia de cada 15 días , prefieren el consumo al

Almuerzo , además del guacamole suelen consumir salsas de aguacate , los motiva la versatilidad en la cocina y solo un 50 por ciento ha consumido productos no convencionales del aguacate , suelen consumirlo en recetas saladas , motivados por el precio de la fruta a la hora de la compra, les gustaría ver productos innovadores de aguacate en el mercado

Recomendaciones:

Crear estrategias diseñadas para abordar las preocupaciones de precio y preferencia destacando la calidad y el valor superior del aguacate Hass frente a otras variedades con el fin de justificar su precio para ello es necesario educar al consumidor sobre ello , ofrecer campañas con recetas que sean económicas con el fin de que comiencen a cambiar sus hábitos y estilo de vida e incluyen el aguacate como parte de su dieta semanal.

Clúster 2: "Exploradores Jóvenes del Aguacate Hass"

En cuanto a l edad se encuentran concentrados en el rango de 25-44 años, Técnicos y profesionales empleados consumen aguacate forma quincenal , suelen consumir en almacenes de cadena , supermercados y Legumbrierias , suelen consumir aguacate Hass en recetas Saladas , consumen guacamole con frecuencia mensual o quincenal , suelen hacer en desayunos y almuerzos, además del guacamole consumen salsa de aguacate, dentro de las motivaciones para consumir aguacate se encuentran su sabor y la versatilidad en la cocina, solo el 50% ha probado productos no convencionales pero se encuentran dispuestos a probarlos.

Recomendaciones:

Se podría introducir productos nuevos derivados del aguacate pues están dispuestos a explorar y son receptivos a los productos no convencionales, esto ofrece una gran oportunidad de venta, se pueden dirigir campañas innovadoras en redes sociales y experiencias en punto de venta, para incentivar el consumo diario.

Clúster 3 “Jóvenes Saludables”

Edad entre los 18 y los 34 años , profesionales y tecnólogos, empleados consumen aguacate cada mes o cada 15 días, suelen comprar en almacenes de cadena, consumen aguacate Hass y Lorena, suelen consumir guacamole de forma poco frecuente, prefieren consumir aguacate al almuerzo y las cenas, suelen consumir aceite de aguacate, motivados por los beneficios para la salud y la versatilidad en la cocina, consumen el aguacate en recetas saladas y como plato principal , motivados por el precio de compra del aguacate , les gustaría ver más productos derivados en el mercado

Recomendaciones:

Realizar campañas enfocadas en la salud y los beneficios del aguacate representa una gran oportunidad para impulsar el consumo frecuente de aguacate Hass, adicional se pueden introducir productos innovadores derivados de la fruta, que puedan ser consumidos en almuerzos y cenas, para posicionar la fruta y recargar su versatilidad en la vida cotidiana.

Clúster 5: “Consumidores Clásicos”

Edad entre los 25 y los 44 años Profesionales y tecnólogos, empleados con consumo de aguacate cada 15 días suele realizar su compra en supermercados, consumen variedad antillano y

Choquete, suelen consumir guacamole de forma ocasional , consumen la fruta en desayunos y almuerzos, además del guacamole no consumen más productos derivados del aguacate , se encuentran motivados por las tendencias alimenticias, consumen el aguacate en recetas saladas y uno de los factores de compra es el precio , les gustaría ver más productos de aguacate en el mercado

Recomendaciones:

Para este grupo en particular que busca opciones económicas se pueden desarrollar campañas de promociones y descuentos especiales en puntos de venta resaltando la calidad de la fruta y sus beneficios para que los clientes se motiven por la compra de esta variedad , también se puede ofrecer productos derivados de excelente calidad y premiar la lealtad de compra de los consumidores como parte de la oferta económica.

Clúster 6 “Buscadores de opciones económicas”

Edad entre los 25 y 34 años, de escolaridad Bachiller empleadas con frecuencia de consumo del aguacate de cada 15 días o un mes , consumo en supermercados y Legumbrieras, consumen variedad antillano y Choquete , suelen consumir guacamole de forma ocasional , consumen aguacate el almuerzo les gusta consumirlo en recetas saladas , además del guacamole no suelen consumir más derivados de la fruta, motivados por las tendencias alimenticias e impulsados más por el precio que por la calidad, a pesar de que no consumen más derivados del aguacate si les gustaría conocer y probar nuevos productos.

Recomendaciones:

Para este grupo que prioriza la economía y no consume aguacate Hass pero que está dispuesto a probar productos no convencionales, recomendamos el lanzamiento de productos derivados de precios accesibles para el público, ofreciendo promociones, experiencias y realizando campañas que se enfoquen en la calidad de aguacate Hass enfatizando la relación Calidad Precio con el fin de aumentar el consumo de esta fruta.

Conclusión

La realización de este modelo tipo cluster para Cartama nos permite evidenciar el aguacate Hass y sus derivados si tiene un perfil del consumidor definido y que esta fruta suele ser consumida en grandes cantidades , que los consumidores prefieren comprar en supermercados y almacenes de cadena y están abiertos al consumo de sus derivados no solo en fruta si no en guacamole y otros productos nuevos que no son muy comunes en el mercado , estos resultados podría ayudar a enfocar mejor las estrategias de marketing y desarrollo de productos con el fin de darle valor a la fruta nacional que dispone la empresa en este momento pues los consumidores buscan calidad y buenos precios en sus compras y quieren seguir un estilo de vida saludable el cual pueden lograr gracias al consumo de aguacate por lo que recomendamos a la empresa que siga su proceso de comercialización de fruta en los puntos de venta actuales como lo es el canal Retail y expanda sus productos a grandes supermercados donde se encuentra su público objetivo, adicional Cartama ya cuenta con un portafolio de productos derivados del aguacate como lo es la pulpa y el guacamole, es importante que construya sus campañas de marketing en este portafolio pues es claro que es un producto con gran potencial según el perfil del consumidor y adicional para el crecimiento de la marca puede explorar mercados de productos alimenticios basados en este fruto puesto que los clientes tienen toda su atención en probar nuevos productos a base de aguacate, para las personas que aún no consumen el aguacate Hass es clave el desarrollo de campañas de Marketing que resalten la calidad de esta Variedad frente a las demás con el fin de educar al consumidor e incentivarlo a que cambie sus preferencias e impulsar la venta del Hass aumentando de forma significativa las ventas para la empresa pues es claro que es un público grande que necesita ser captado

Referencias

Wu, D., Xu, Z., & Li, J. (2023). Search Data and Geodemographics Determinants of the Avocado Sales in the US Markets. *Journal of Business and Management*, 29(1), 23–56.

[https://doi.org/10.6347/JBM.202307_29\(1\).0002](https://doi.org/10.6347/JBM.202307_29(1).0002)

Chen, Z., Zeng, G., Zhong, S., & Wang, L. (2023). From the exotic to the everyday: The Avocado crossing borders via cyberspace. *Appetite*, 180.

<https://doi.org/10.1016/j.appet.2022.106362>

Cavaletto, G., & Sourcing, V. P. (2015). *The avocado market in the United States*.
<http://www.californiaavocado.com/>

Anesbury, Z. W., Talbot, D., Day, C. A., Bogomolov, T., & Bogomolova, S. (2020). The fallacy of the heavy buyer: Exploring purchasing frequencies of fresh fruit and vegetable categories. *Journal of Retailing and Consumer Services*, 53.

Anesbury, Z., Sweeney, S., & Zuo, M. (2020). *Consumer Insights on Avocado Consumption and Market Trends*. *Journal of Consumer Studies*.

Cavaletto, J., & Sourcing, M. (2015). *Trends in Avocado Consumption: Insights from the U.S. Market*. *International Journal of Food Sciences*.

Chen, Y., Huang, Y., & Wu, J. (2023). *The Role of Social Media in Promoting Healthy Eating: The Case of Avocado*. *Nutrition Reviews*.

Feller, I., Wilkins, L., & Aris, J. (2022). *Sustainability in Avocado Production: Challenges and Opportunities*. Journal of Agricultural Research.

Kogan, A., Heller, A., & Fuchs, H. (2018). *Data Mining Techniques for Market Segmentation: A Case Study of Avocado Consumers*. Journal of Marketing Research.

Wang, J., Zhang, Z., & Li, T. (2020). *Nutritional Benefits of Avocado Consumption: A Review*. Advances in Nutrition.

Wu, Q., Zhao, W., & Li, R. (2023). *Demographic Factors Influencing Avocado Consumption: Evidence from Global Studies*. Journal of Consumer Behavior.