

Tipo de documento: Tesis de Maestría



Departamento de Economía. Maestría en Economía Aplicada
**Métodos para mejorar la asertividad de
pronósticos de ventas en empresas de
consumo masivo: Caso Ecuatoriano**

Autoría: Jurado Santillán, Carlos Alberto

Año: 2024

¿Cómo citar este trabajo?

Jurado Santillá, C. (2024) "*Métodos para mejorar la asertividad de pronósticos de ventas en empresas de consumo masivo: Caso Ecuatoriano*". [Tesis de Maestría. Universidad Torcuato Di Tella]. Repositorio Digital Universidad Torcuato Di Tella
<https://repositorio.utdt.edu/handle/20.500.13098/13235>

El presente documento se encuentra alojado en el Repositorio Digital de la Universidad Torcuato Di Tella bajo una licencia Creative Commons Atribución-No Comercial-Compartir Igual 4.0 Argentina (CC BY-NC-SA 4.0 AR)
Dirección: <https://repositorio.utdt.edu>



**UNIVERSIDAD
TORCUATO DI TELLA**

**UNIVERSIDAD TORCUATO DI TELLA
MAESTRÍA EN ECONOMÍA APLICADA**

Métodos para mejorar la asertividad de pronósticos de ventas en empresas de
consumo masivo: Caso Ecuatoriano

PROYECTO INTEGRADOR

Previo la obtención del Título de Posgrado

Presentado por:

➤ CARLOS ALBERTO JURADO SANTILLAN

GUAYAQUIL - ECUADOR

Año: 2024

DEDICATORIA

El presente trabajo de análisis exploratorio lo dedico principalmente a Dios, por ser el inspirador y darme fuerza para continuar en este proceso de obtener uno de los anhelos más deseados; y a mis padres Carlos y Sandra, por su amor, trabajo y sacrificio en todos estos años. Es un orgullo y privilegio ser su hijo, son los mejores padres.

*Carlos Alberto Jurado
Santillán.*

AGRADECIMIENTOS

Me gustaría agradecer principalmente al Mgtr. Alejandro Aguilar Oreo (Gerente General en Universal Sweet Industries), por cada día mostrarme que los retos se pueden superar, solo es cuestión de creernos capaces.

Agradezco al Phd. Cesar Ciappa. Por su tiempo en todo el proceso de realización de esta tesis y a la UTDT.

DECLARACIÓN EXPRESA

“Los derechos de titularidad y explotación, nos corresponde conforme al reglamento de propiedad intelectual de la institución; *Carlos Alberto Jurado Santillán* y doy mi consentimiento para que la UTDT realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual”

Carlos Jurado S.

Resumen

En Ecuador la falta de análisis de los perfiles de cliente ha causado que las empresas sigan operando sus negocios basados en modelos no predictivos lo cual incurre en tomar decisiones erradas e influyen directamente en el estado de resultados como potenciales perdidos. Considerando que, la falta de información es una limitación para el desarrollo una operación.

Este estudio nace de la necesidad de establecer las características de mayor relevancia de los clientes de la empresa “Universal Sweet Industries”, para así establecer reportes e información relevante que permitan identificar tendencias del comportamiento de dichos clientes y tomar que maximicen ventas y rentabilidad (RGM)¹

Para lograr esto se necesitó determinar una base histórica de ventas por cliente con la mayor información posible para eliminar el sesgo por variables no observables, la metodología que implementamos es un modelo Arima en el que en función a comportamientos históricos del cliente se pronostica el movimiento futuro de los mismos.

Los resultados demuestran que aplicar análisis de forecasting para el comportamiento de cliente y traducirlo a una estimación de demanda permite reducir el impacto de perdidas por una estimación no sincerada, adicional a esto nos permite identificar oportunidades de crecimiento en una coyuntura post pandemia donde ante cualquier shock el mercado ecuatoriano se contrae, las firmas necesitan anticiparse a estos comportamientos para saber como reaccionar.

Palabras Clave: Estimación de demanda, Post Pandemia, Perfil de clientes, Ecuador, Clúster de clientes, Probabilidad de Compra.

¹ Revenue Growth Management

ÍNDICE GENERAL

Resumen	5
ÍNDICE GENERAL	6
CAPÍTULO 1	12
1.1. Introducción	12
1.2. Descripción del Problema	13
1.3. Justificación del problema	14
1.4. Objetivos	14
1.4.1. Objetivo General	14
1.4.2. Objetivos Específicos.....	14
CAPÍTULO 2	16
2.1. Marco Teórico	16
2.1.1. La Empresa.....	16
2.1.2. Canales de Distribución	16
Figura 1. Product Flow de La Universal.....	16
2.2. Minería de Datos.....	16
2.2.1. Análisis de Clusters	17
2.2.2. Agrupamiento de particiones	17
Agrupación K-means.....	17
Ventajas	18
Desventajas	18
Método clásico de partición: Algoritmo K-means	18
Agrupación de K-medoids o PAM	19
2.2.3. Método para Medir Distancias	19
2.2.4. Determinación del número óptimo de clusters	20
Método del Codo	20

2.2.5.	Medidas Internas para Validación de los clusters	20
	Criterio de la Silueta	20
	Índice de Dunn	21
2.2.6.	Arboles de decisión.....	22
	Algoritmo CHAID	22
2.3.	Revisión Bibliográfica	22
2.4.	Modelos de Aprendizaje Estadístico	25
	Predicción	25
2.5.	Métodos Paramétricos: Modelo de regresión lineal.	26
	Ajuste del Modelo de Regresión Lineal.....	27
2.6.	Métodos No Paramétricos: K-Vecinos más cercanos (KNN) y Árboles de Regresión y Clasificación (CART).	28
	Algoritmo K-Vecinos más Cercanos (KNN).....	28
	Arboles de Clasificación y Regresión (CART).....	28
	Elección del Mejor Árbol.....	29
2.7.	Análisis de Correspondencia	29
2.8.	Métricas de evaluación	30
	Error cuadrático Medio (RSME)	30
	Coefficiente de Determinación (R-Square).....	30
	Error Porcentual Absoluto Medio (MAPE).....	31
CAPÍTULO 3		32
3.1.	Metodología	32
3.1.1.	Enfoque de la Investigación	32
3.1.2.	Datos	32
3.2.	Diseño de la Investigación.....	33
3.3.	Modelo logístico binario	33
3.3.1.	Análisis para Cocoa	34

Matriz de confusión del modelo.....	35
Interpretación de los coeficientes y la razón de probabilidades (OR)	36
3.3.2. Análisis para Manicho	40
Interpretación de los coeficientes y la razón de probabilidades (OR)	41
3.3.3. Análisis para Manicho con Galleta.....	44
Interpretación de los coeficientes y la razón de probabilidades (OR)	46
3.3.4. Análisis para Manicho con Galleta.....	49
Interpretación de los coeficientes y la razón de probabilidades (OR)	51
3.4. Análisis por análisis de correspondencia.....	55
Análisis de correspondencia entre diferentes categorías	57
CAPÍTULO 4	60
4.1. Conclusiones	60
4.2. Recomendaciones	61
Bibliografía	63
ANEXOS	65

Tabla 1. Frecuencia de Cocoa	34
Tabla 2. Matriz de confusión de Cocoa	36
Tabla 3. Métricas del Modelo para Cocoa	36
Tabla 4. Coeficientes y OR para preferir Cocoa.....	37
Tabla 5. Coeficientes y OR Para No Preferir Cocoa	38
Tabla 6. Frecuencia de Manicho	40
Tabla 7. Matriz de Confusión de Manicho	41
Tabla 8. Métricas del Modelo de Manicho	41
Tabla 9. Coeficientes y OR para preferir Manicho.....	41
Tabla 10. Coeficientes y OR para NO preferir Manicho	43
Tabla 11. Frecuencia de Manicho con Galleta	44
Tabla 12. Matriz de Confusión de Manicho con Galleta	45
Tabla 13. Métricas del Modelo	46
Tabla 14. Coeficientes y OR para preferir Manicho con Galleta.....	46
Tabla 15. Coeficientes y OR para NO preferir Manicho con Galleta	48
Tabla 16. Frecuencia de Menta Glacial.....	50
Tabla 17. Matriz de Confusión de Menta.....	51
Tabla 18. Métricas del Modelo	51
Tabla 19. Coeficientes y OR para preferir Menta Glacial	51
Tabla 20. Coeficientes y OR para No Preferir Menta Glacial	53
Tabla 21. Relación entro los cluster y la marca.....	59

Gráfico 1. Frecuencia de Cocoa.....	35
Gráfico 2. Razón de Probabilidad para Preferir Cocoa	37
Gráfico 3. Razón de Probabilidad para No Preferir Cocoa	39
Gráfico 4. Frecuencia de Manicho.....	40
Gráfico 5. Razón de Probabilidad para Preferir Manicho	42
Gráfico 6. Razón de Probabilidad para No Preferir Manicho.....	43
Gráfico 7. Frecuencia de Manicho con Galleta.....	45
Gráfico 8. Razón de probabilidad para preferir Manicho con Galleta	47
Gráfico 9. Razón de Probabilidad para No Preferir Manicho con Galleta.....	48
Gráfico 10. Frecuencia de Menta Glacial	50
Gráfico 11. Razón de Probabilidad para preferir Menta Glacial	52
Gráfico 12. Razón de Probabilidad para No Preferir Menta Glacial	54
Gráfico 13. Asociación entre marca y agencia	55
Gráfico 14. Análisis de Correspondencia Múltiple: Clustering de Variables	57
Gráfico 15. Número óptimo de Clusters - Método del Codo.....	58
Gráfico 16. Clusters de Clientes en el Espacio MCA	59

Anexo 1. Coeficientes del modelo de Cocoa.....	65
Anexo 2. Coeficientes del modelo de Manicho.....	66
Anexo 3. Coeficientes del modelo de Manicho con Galleta	68
Anexo 4. Coeficientes del modelo de Menta Glacial	69

CAPÍTULO 1

1.1. Introducción

En la última década la precisión de demanda se ha convertido en algo fundamental para el comercio. Los datos generados por los clientes durante la última década se han convertido en un recurso valioso para las firmas a la hora de tomar decisiones, los mismos se usan para identificar patrones de consumo o preferencias de clientes (Dahlén C & Siwerz R, 2017). Latinoamérica particularmente posee características diferenciadas de las sociedades de países desarrollados. Ecuador es un país donde el comercio basado en análisis de datos aún no se encuentra desarrollado, con lo cual se hace presente la necesidad de innovar en el mercado. La empresa ecuatoriana que tomaremos como referencia para realizar esta investigación es “Universal Sweet Industries”, una empresa con más de 135 años en el mercado ecuatoriano, reconocida por poseer marcas representativas como Manicho, Cocoa, Menta glacial, Bandido, entre otras. Cada una de estas es líder de mercado en sus categorías, por lo que su presencia en cada uno de los canales de distribución no era algo desconocido. A principios del 2020 el mercado ecuatoriano tuvo una contracción en ventas producto de las restricciones sanitarias para minorizar el crecimiento de casos positivos de Covid-19, esto obligó a las empresas ecuatorianas a buscar fuentes de información que ayuden a predecir el comportamiento de sus clientes puesto que necesitaban minimizar los riesgos, normalmente estos shocks se presentan como una contracción, sin embargo en el caso de la firma ecuatoriana que estamos analizando las estrategias que se plantearon permitieron desarrollar el modelo de negocio obteniendo un crecimiento frente al 2019.

El periodo de este estudio serán los meses comprendidos en el 2021, 2022 y 2023. La evidencia empírica muestra que producto de la imposición de restricciones de movilidad humana en el Ecuador, las personas se vieron incentivadas a realizar compras de productos de primera necesidad y dejaron de realizar compras de productos por impulso, normalmente en el mercado de productos de baja necesidad se presentan fluctuaciones estacionales posterior a eso vuelven a su comportamiento habitual como por ejemplo compra de juguetes en navidad, en ese sentido es interesante analizar como una categoría

por impulso tuvo un crecimiento en ventas cuando el mercado ecuatoriano sufría una contracción por restricciones de presupuesto de los consumidores que priorizaban sus gastos en productos de mayor necesidad.

Universal Sweet Industries logró instaurar el análisis de mercado como una rutina para la toma de decisiones (RGM) en el portafolio, de forma que estas palancas² tengan un impacto positivo en el P&L, estas sesiones se volvieron rutinas hasta instaurar un área específica que se dedicara a levantar información relevante que determine las decisiones de consumo de los clientes atendidos por la empresa, minimizando pérdidas por producto caducado dado que su giro de negocio se basa en la venta de productos de la categoría Chocolates y Candy mismos que tienen una durabilidad promedio de 8 meses.

1.2. Descripción del Problema

La industria de alimentos y bebidas de consumo masivo representa una de las actividades más representativas del mercado ecuatoriano y a nivel mundial, esto se da debido a la tasa de crecimiento poblacional. Hoy en día la demanda de productos listos para consumir “**ready prepared foods**”, está creciendo de debido a un aumento de ingresos per cápita, una mayor urbanización y un número creciente de mujeres que ingresan al mundo laboral (C. A. da Silva, 2009).

Las empresas de consumo masivo requieren anticipar el comportamiento del mercado para realizar su planificación de producción, así como también ser más precisos en sus proyecciones de ventas. Perfeccionar las técnicas de estimación de demanda y consolidar este proceso como una rutina de seguimiento continuo con periodicidad mensual permite anticipar eventualidades y sincerar expectativas basado en datos históricos. Por esta razón es prioridad identificar las variables o aquellas sesiones que permitan explicar aquellos one

² Ideas basadas en un análisis en el se busque crecimiento de margen bruto o ventas de algún producto o canal de distribución

timme³ que afectan la venta dado que estas dos ayudan a robustecer el modelo y tomar mejores decisiones evitando así pérdidas por altos días de inventario de productos con fecha de caducidad determinada (Thron, 2007).

1.3. Justificación del problema

Determinar un modelo de precisión de demanda basado en una proyección sensibilizada por el área comercial (proyección de ventas) recopilando información que ayude a prevenir shocks positivos como negativos en la venta de las empresas de consumo masivo en el segmento de alimentos, para este caso particular utilizaremos la información de la empresa “Universal Sweet Industries”, una empresa que basa su operación en la comercialización de productos de chocolates, polvo de cacao, confites y galletas. Esta información abrirá una pauta para que los tomadores de decisiones relacionados a la planificación y análisis estratégico del giro de negocio de esta empresa para maximizar su rentabilidad y minimizar las pérdidas por producto en inventario con vida útil no apta para su comercialización (Eguiguren Calisto, 2019).

1.4. Objetivos

1.4.1. Objetivo General

Identificar las variables cualitativas y cuantitativas con mayor poder explicativo para mejorar la precisión de demanda de la empresa “Universal Sweet Industries”

1.4.2. Objetivos Específicos

1. Establecer variables relacionadas al consumo de las categorías que comercializa la empresa basada en lectura previa.

³ Eventualidades que se dan por variables externas al modelo, gestión o rutina normal del giro de negocio, la premisa de uso de este termino es que la eventualidad no se presente de manera recurrente y sea identificada oportunamente para ser prevista en futuras ocasiones evitando así su recurrencia.

2. Determinar un modelo de precisión de demanda, con la finalidad de establecer un parámetro como meta de mejora continua basado en técnicas de Clusterización de clientes para el canal de distribución Indirecto CBC⁴.
3. Evaluación mensual de las acciones promocionales en las marcas core de la empresa Universal Sweet Industries.

⁴ CBC: The Central America Bottling Corporation, Empresa dedicada a la distribución de productos de consumo masivo, principalmente bebidas gaseosas y carbonatadas, esta empresa es parte del mismo holding de empresas que Universal Sweet Industries.

CAPÍTULO 2

2.1. Marco Teórico

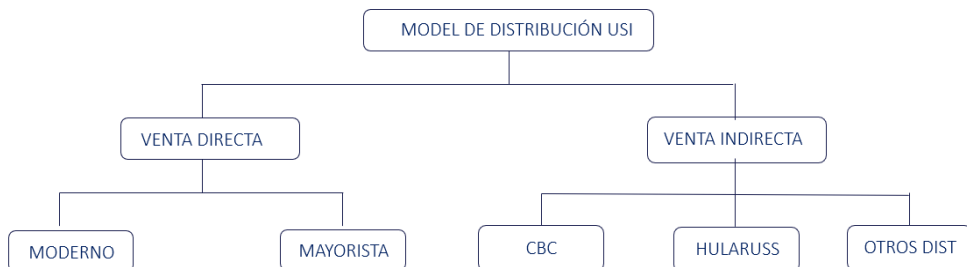
2.1.1. La Empresa

La Empresa Universal Sweet Industries es una empresa ecuatoriana cuya actividad económica consiste en la comercialización al por mayor de productos de la categoría snacks and candy y polvo de Cacao, posee Cobertura a nivel nacional y tiene mas de 135 años en el mercado ecuatoriano.

2.1.2. Canales de Distribución

Posee una distribución directa e indirecta, los canales Mayorista y Moderno son atendidos de manera directa, mientras que el canal tradicional posee una venta indirecta dado que es atendido con distribuidores, entre ellos están dos cuentan con cobertura nacional (CBC y Hularuss⁵)

Figura 1. Product Flow de La Universal.



2.2. Minería de Datos

La Minería de datos se puede definir como un conjunto de estrategias y componentes que permiten interpretar los datos obtenidos de Sistemas de Procesamiento Transaccionales (SPT) u otras fuentes de diferentes formatos

⁵ Hularuss: Empresa dedicada a la distribución de productos de consumo masivo en el mercado ecuatoriano, esta es una empresa extrema del Holding de empresas del grupo Mariposa (holding al que pertenece La Universal).

(estructurados, semi-estructurados y no estructurados) los cuales tienen un potencial para contribuir en los aspectos principales del negocio (Coelho, 2016).

Un ejemplo de los beneficios que aporta la minería de datos para el giro de un negocio es la propuesta de (Cálad, 2015) donde aplicó técnicas: K-means y árboles de decisión para determinar un criterio de segmentación de clientes de Tiendacol S.A una empresa del sector de modas ubicada en la ciudad de Medellín, como resultado logró identificar tres segmentos de clientes diferentes: los mejores, los intermedios y los peores. Los mejores que representan a los clientes consecutivos con alta predisposición de gasto, los intermediarios son clientes con una recompra espaciada, pero con predisposición de gasto y los peores que son clientes poco fidelizados y de baja predisposición de gasto.

2.2.1. Análisis de Clusters

“El propósito del análisis de clusters es agrupar elementos en grupos homogéneos en función de las similitudes entre ellos”. (Peña, 2002), para el presente estudio esta técnica se utilizará para determinar segmentos de clientes atendidos por el canal de distribución de Tesalia.

2.2.2. Agrupamiento de particiones

De acuerdo con lo planteado por (Kassambara, 2015) en su recomendación de agrupamiento de particiones especifica el uso de dos métodos relevantes, los cuales son: Agrupación K-means y Agrupación de K-medoids o PAM.

Agrupación K-means

La agrupación K-means sirve para hacer agrupaciones de objetos o individuos en base a la similitud de sus características (se busca formar clusters compactos y bien definidos).

Ventajas

Es útil cuando se busca segmentar grandes volúmenes de datos.
Computacionalmente es menos costoso que el clustering Jerárquico.

Desventajas

Se necesita conocer la cantidad de cluster a formar
Debido a que crea los grupos en base a las distancias entre individuos, solo se puede aplicar con variables cuantitativas.

La cantidad de los clusters depende en gran medida de los valores con los que se inicia el algoritmo.

El resultado puede ser un óptimo localmente pero no necesariamente es el óptimo global.

El resultado se encuentra afectado por outliers.

Método clásico de partición: Algoritmo K-means

El objetivo del algoritmo K-means se basa en determinar una muestra de n elementos con p variables para dividir dicha muestra en un número de grupos predeterminados, G .

Este este algoritmo requiere de cuatro etapas:

1. Seleccionar G puntos como centros de los grupos iniciales. Esto puede hacerse:
 - a) asignando aleatoriamente los objetos a los grupos y tomando los centros de los grupos así formados;
 - b) tomando como centros los G puntos más alejados entre sí;
 - c) construyendo los grupos con información a priori, o bien seleccionando los centros a priori.
3. Calcular las distancias euclídeas de cada elemento al centro de los G grupos, y asignar cada elemento al grupo más próximo. La asignación se realiza

secuencialmente y al introducir un nuevo elemento en un grupo se recalculan las coordenadas de la nueva media de grupo.

4. Definir un criterio de optimalidad y comprobar si reasignando uno a uno cada elemento de un grupo a otro mejora el criterio.
5. Si no es posible mejorar el criterio de optimalidad, terminar el proceso.

Según (Peña, 2002) para definir el criterio G se debe aplicar ley de Parsimonia.

Agrupación de K-medoids o PAM

Según (Hung, Lien, & Ngoc, 2019) es el algoritmo de aprendizaje en el cada grupo está representado por uno de los objetos de los grupos (mediana). PAM es un método que aísla en cierto grado los outliers en comparación con k-medias.

2.2.3. Método para Medir Distancias

Este paso determina la similitud de los elementos (x, y) y de igual forma influye en la creación de los grupos.

Los métodos clásicos que existen para medir las distancias son: Euclidiana y Manhattan, los mismos están representados en las siguientes ecuaciones:

1. Distancia Euclidiana

$$d_{euc}(x, y) = \sqrt{\sum_{i=1}^n (x_i - x_j)^2}$$

(1)

2. Distancia Manhattan

$$d_{man}(x, y) = \sum_{i=1}^n (x_i - x_j)$$

(2)

2.2.4. Determinación del número óptimo de clusters

Existen tres métodos: Método del Codo, Método de silueta promedio y el estadístico de brecha, pero como se mencionó anteriormente en el presente estudio por parsimonia aplicaremos el método del Codo, sin embargo, en el programa R studio se validará este método todos los índices que se encuentran en el paquete de la función ipak (en la misma se encuentran más de 30 índices) con lo cual por selección de la mayoría podremos determinar con mayor confiabilidad el número óptimo de clústeres.

Método del Codo

Es un método intuitivo donde se identifica el punto en el que la varianza intra-clúster disminuye, este efecto se interpreta como cambios marginales, es decir el aumento del número de clústeres no mejorará significativamente el poder explicativo de los mismos.

2.2.5. Medidas Internas para Validación de los clusters

Criterio de la Silueta

El coeficiente de la silueta se calcula utilizando la distancia media dentro de un grupo (a) y la distancia media del grupo más cercano (b) para cada muestra x_i del conjunto X .

$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}}$$

- Si x_i es un punto aislado entonces $s_i = 0$.

- Si s_i es cercano a 1 significa que los clusters son compactos y están bien separados.
- Si s_i es cercano a 1 significa que el individuo se encuentra asignado correctamente en el cluster.
- Si s_i es cercano a 0 significa que el individuo se encuentra en el límite entre dos clusters.
- Si s_i es cercano a -1 significa que el individuo no se encuentra en el cluster correcto.

Índice de Dunn

Se calcula de la siguiente forma:

- Para cada grupo se calcula la distancia entre cada uno de los objetos que lo conforman y los objetos de los grupos restantes.
- Use la separación mínima, es decir la menor distancia por pares como la separación entre grupos.
- En cada grupo se debe calcular la distancia entre objetos del mismo grupo.
- Se Utiliza la distancia máxima como la compacidad dentro-clusters.

El Índice de Dunn está representado en la siguiente ecuación:

$$D = \frac{\min Separación}{\max Diámetro}$$

Esta formula se puede interpretar de la siguiente manera:

Si el conjunto de datos contiene grupos compactados y bien clasificados, se espera que el diámetro de los grupos sea corto y que la distancia entre los grupos sea amplia. Por lo tanto, este índice debe maximizarse.

2.2.6. Árboles de decisión

Un Árbol de decisión es una representación de un conjunto de condiciones organizadas en una estructura priorizada, de forma que el resultado final se determina siguiendo las condiciones que dan desde el inicio hasta el final del árbol.

Algoritmo CHAID

Su nombre proviene del estadístico Chi-Cuadrado y de una detección de interacción automática (Chi-Square Automatic Interaction Detection). Se asemeja a la regresión múltiple con la diferencia que este procedimiento *x es una variable Cuantitativa, Discontinua o Discreta; y es la variable dependiente.* En cada paso CHAID elige la variable independiente (Predictora) que presenta la interacción más fuerte con la variable dependiente. El propósito de este método es obtener tipologías y perfiles de los individuos pertenecientes al conjunto de datos.

Luego de dividir la Población en dos o mas grupos distintos basados en categorías del mejor predictor de una variable dependiente. Se divide estos grupos en subgrupos pequeños basados en variables de otros predictores, este proceso de división continua termina hasta que no se encuentren más predictores estadísticamente significativos, el algoritmo presenta los resultados en un diagrama de árbol de decisión.

2.3. Revisión Bibliográfica

La industria de alimentos en el mercado ecuatoriano generalmente se ha basado en análisis temporales, los mismos no toman en consideración patrones de consumo por alguna variable exógena, en la actualidad el mercado ecuatoriano se ha enfrentado ha varios de estos como: Paros Nacionales, Pandemia, Crisis energética, variaciones en el impuesto al valor agregado, entre

otros. Dichos cambios afectan la decisión de consumo de los ecuatorianos priorizando sus preferencias de acuerdo con su restricción presupuestaria. En estos escenarios los métodos de aprendizaje estadístico han logrado mostrar otro enfoque a las firmas teniendo excelentes resultados con este tipo de pronósticos.

Las empresas de la industria de alimentos que adaptan una cultura y áreas de trabajo basados en datos logran mejorar en 6% su rentabilidad de aquellas que no lo hacen. Los pronósticos de ventas permiten tener un mejor control de inventario y pedidos considerando que estos activos en piso son de gran relevancia puesto que su correcta administración evitaría pérdidas por caducidad de producto (Siwerz & Dahlén, 2017)

Los Autores (Siwerz & Dahlén, 2017) en su estudio de predicción de ventas de una tienda de alimentos comparan métodos de aprendizaje estadístico: Perceptrón Multicapa (MLP), SVM y Red de función de base radial (RBFN), donde obtuvieron como resultado que el SVM posee mayor poder explicativo gracias a que posee menor error que los otros dos. Ellos compararon los resultados en función del Error Porcentual Absoluto Medio (MAPE) y el Error Cuadrático Medio (RMSE).

Es importante considerar que estas metodologías requieren variables con alto poder explicativo, en el mercado ecuatoriano las variables macroeconómicas están correlacionadas con el desempeño de todas las actividades comerciales del país. Un estudio enfocado al comercio minorista (Wang, Liu, & Liu, 2019) logró obtener que la metodología SVM tiene una alta precisión para pronosticar comportamientos de productos perecederos, sin embargo, esta precisión mejoraba si se añadían variables como: PIB, tasa de inflación, Desempleo y nivel socio económico de clientes enrolados.

El autor (Palacios Utreras, 2020) utilizó el método de árboles de clasificación y regresión (CART) donde obtuvo del 76% al 99% de precisión en su pronóstico de precios de proyectos inmobiliarios, estos resultados se asemejan a la metodología MLP que es un modelo Robusto.

Por otro lado (Díaz Sepúlveda & Correa, 2015) en su investigación comparan los modelos de regresión lineal sea cuadrático o trigonométrico con CART con una simulación, obteniendo que el error de predicción de la regresión lineal es menor que el de CART cuando se conoce la forma funcional de los datos sin embargo, también concluyen que CART es una metodología que se puede usar cuando no se conoce la forma funcional de los datos.

2.4. Modelos de Aprendizaje Estadístico

Predicción

Un modelo de predicción es una representación estadística que utiliza datos históricos de variables o individuos para predecir resultados futuros. Estos Modelos tienen como base los patrones, tendencias y correlaciones entre variables para determinar un comportamiento futuro.

Los Autores (James, Witten, Hastie, & Tibshirani, 2013) determinan que, si el error promedio es cero, se puede predecir Y con la siguiente ecuación:

$$\hat{Y} = \hat{f}(x)$$

\hat{f} : Estimación para f

\hat{Y} : Resultado de estimación de Y

La precisión de \hat{Y} como predicción de Y depende de dos cantidades, el error deducible y error irreducible representado por la siguiente ecuación:

$$\begin{aligned} E(Y - \hat{Y})^2 &= E[f(x) + \epsilon - \hat{f}(x)]^2 \\ &= E[f(x) - \hat{f}(x)]^2 + Var(\epsilon) \end{aligned}$$

$E(Y - \hat{Y})^2$: Es la diferencia del valor esperado al cuadrado entre el valor predicho y el real de Y .

$E[f(x) - \hat{f}(x)]^2$: Es el error Reducible

$Var(\epsilon)$: es el error Irreducible o Varianza asociada al término del error.

Para el error reducible existen métodos de aprendizaje estadístico caracterizados como paramétricos y no paramétricos según (James, Witten, Hastie, & Tibshirani, 2013).

2.5. Métodos Paramétricos: Modelo de regresión lineal.

La regresión lineal encuentra la relación entre dos o más variables independientes y su correspondiente variable dependiente.

La ecuación de regresión lineal tiene la siguiente forma:

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + x_{im}\beta_m + \dots + \epsilon_i, i = 1, \dots, n$$

Esta ecuación se representa de la siguiente forma matricialmente:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

El vector de observaciones de Y esta compuesto por:

$$y = (y_1, y_2, \dots, y_n)'$$

El vector de parámetros β esta compuesto por:

$$\beta = (\beta_1, \beta_2, \dots, \beta_n)'$$

La matriz X se denomina matriz de diseño

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix}$$

El vector de desviaciones está representado por:

$$\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)'$$

El modelo está representado por la siguiente ecuación en su notación matricial compacta:

$$y = \mathbf{X}\beta + \epsilon$$

Supuestos del Modelo de Regresión Lineal

Las desviaciones estándar o errores ϵ_i del modelo lineal se aproximan a n variables aleatorias con media 0, no Correlacionadas (multicolinealidad) e igualdad de varianzas de los residuos.

Se debe cumplir que:

- $E(\epsilon_i) = 0, i = 1, \dots, n$
- $E(\epsilon_i \epsilon_j) = 0, i \neq 1, \dots, n$
- $var(\epsilon_i) = \sigma^2, i = 1, \dots, n$

Si estas condiciones se cumplen se puede concluir que:

$$E(\epsilon) = 0, \Sigma_{\epsilon} = \sigma^2 I_p$$

Luego de seleccionar el modelo, se requiere aplicar el método de Mínimos Cuadrados Ordinarios (MCO), el cual consiste en estimar el vector de parámetros β , estos coeficientes minimizan la suma cuadrática del error entre el valor observado y el valor estimado.

Ajuste del Modelo de Regresión Lineal

Las medidas de ajuste recurrentes son el Error Estándar Residual (RSE) y el Coeficiente de determinación R^2 . Si $R^2 \approx 1$ significa que el modelo explica con gran proporción la varianza en la variable de respuesta, el efecto contrario se da cuando $R^2 \approx 0$. Por otro lado RSE esta definido por:

$$RSE = \sqrt{\frac{1}{n - p - 1} RSS}$$

Donde RSS es la suma de residuos cuadrados, esto se interpreta como una relación inversa con p , si $RSS \downarrow \rightarrow p \uparrow$, si esta relación se da $\rightarrow RSE \downarrow$ el modelo se ajusta mejor a los datos.

2.6. Métodos No Paramétricos: K-Vecinos más cercanos (KNN) y Árboles de Regresión y Clasificación (CART).

De acuerdo con (Quispe, Calla, Yangali, Rodríguez, & Pucamayo, 2019) en varios casos no se puede determinar la forma funcional f por lo que se utiliza un enfoque con mayor flexibilidad, el cual permite realizar la regresión. El problema con los métodos no paramétricos es que no reduce la estimación de f a un pequeño número de parámetros, por ende, es necesario una gran cantidad de observaciones para tener una estimación significativa de f .

Algoritmo K-Vecinos más Cercanos (KNN)

Este algoritmo realiza clasificaciones o predicciones de la agrupación basado en un punto. Generalmente se utiliza este algoritmo para como una metodología de clasificación, basado en la premisa de que se pueden encontrar puntos similares si los mismos están cercanos.

Si se tiene un conjunto de entrenamiento (x_i, c_j) donde $i = 1, 2, \dots, n$ y las etiquetas $j = 1, 2, \dots, g$. A una observación x se le asigna una clasificación c_j si esta es la clase con mayor frecuencia entre las k observaciones de entrenamiento mas cercanos. El parámetro k se ajusta hasta encontrar la mejor clasificación con el conjunto de entrenamiento.

Arboles de Clasificación y Regresión (CART)

Esta Metodología del Machine Learning utiliza datos históricos para crear arboles de clasificación o regresión, su función principal es clasificar o predecir nuevos datos.

La ventaja de los arboles CART se encuentra en la robustez a outliers, la invarianza en la estructura de sus arboles de clasificación o de regresión a transformaciones monótonas de las variables independientes y su interpretabilidad.

Un Modelo de árbol de Regresión es una descripción Condicional de Y dado X . Los dos Componentes fundamentales del modelo son: un árbol binario b , nodos Terminales y el vector de parámetros $\theta = (\theta_1, \theta_2, \dots, \theta_b)$ donde el parámetro θ_b está asociado al nodo terminal N_i , Un árbol de regresión crea un modelo explicativo y predictivo para una variable cuantitativa dependiente basada en variables explicativas cuantitativas y cualitativas. (Díaz & Correa, 2013).

Elección del Mejor Árbol

Para determinar el mejor árbol se debe estimar la tasa de error $R(T)$. La misma cuenta con la siguiente forma funcional:

$$R^{cv}(T) = \frac{R^1(T), \dots, R^k(T)}{k}$$

Donde $R^1(T), \dots, R^k(T)$ representan las estimaciones de k .

2.7. Análisis de Correspondencia

El Objetivo de este método también conocido como análisis de Homogeneidad, es determinar cuantificaciones que sean óptimas en el sentido de que las categorías estén separadas entre sí tanto como sea posible. Esto implica que los objetos de la misma categoría se trazan uno cerca del otro y que los objetos de diferentes categorías se trazan lo más separados posible. El término homogeneidad también hace referencia al hecho de que el análisis será más satisfactorio cuando las variables son homogéneas; es decir, cuando dividen los objetos en clústeres con las mismas categorías o categorías similares. Esta técnica, enmarcada dentro de los métodos estadísticos factoriales, está diseñada para estudiar las relaciones entre cualquier número de modalidades de distintas variables categóricas. En esto presenta dos claras diferencias con respecto a otro método factorial, el Análisis de Componentes Principales, que estudia las relaciones entre variables y no entre valores de éstas y que se nutre de variables de tipo cuantitativo. (Parra, 1996)

2.8. Métricas de evaluación

Usualmente las tres métricas que se utilizan para medir un modelo de regresión son: RMSE, Coeficiente de determinación R-Square y MAPE. La forma de interpretabilidad esta dada por: A mayor R^2 y menor valor de RSME y MAPE mejor será la predictibilidad del modelo.

Error cuadrático Medio (RSME)

Su forma funcional es la siguiente:

$$RSME = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_i - \hat{y}_i)^2}$$

Donde \hat{y}_i representa la predicción i-ésima de Y . La forma de interpretar los resultados es la siguiente:

A menor valor de RSME; Las respuestas predichas son muy cercanas a las repuestas verdaderas. RSME tendrá un valor alto si algunas observaciones difieren drásticamente entre las respuestas verdaderas y las predichas.

Coeficiente de Determinación (R-Square)

El coeficiente R^2 mide el nivel de precisión de los datos estimados de la regresión sobre los datos reales. Si el valor es cercano a 1 indica que el modelo se asemeja a los valores reales, un valor cercano a 0 indica que el modelo no posee ningún valor predictivo. R^2 se define de la siguiente forma:

$$R^2 = \frac{TSS - RSS}{RSS} = 1 - \frac{RSS}{TSS}$$

$RSS = \sum_{t=1}^n (y_i - \hat{y}_i)^2$ es la suma de residuos cuadrados.

$TSS = \sum (y_i - \widehat{y}_i)^2$ es la suma total de los cuadrados.

Error Porcentual Absoluto Medio (MAPE)

Es el porcentaje de errores promedio entre el valor real y el estimado, se define de la siguiente forma:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_i - \widehat{y}_i}{y_i} \right|$$

Donde y_i, \widehat{y}_i es el valor real y predicho respectivamente y n es el número de ejemplos de entrenamiento.

CAPÍTULO 3

3.1. Metodología

3.1.1. Enfoque de la Investigación

El estudio busca entender las preferencias de los clientes en las marcas de "Cocoa", "Manicho", "Manicho Con Galleta" o "Menta Glacial", en relación con otras variables de interés. Donde propondremos modelos de clasificación, series temporales, análisis de cohortes y análisis de correspondencia con clustering. Estos métodos permitirán obtener una visión integral de los patrones de preferencia y las dinámicas del mercado, proporcionando información valiosa para optimizar estrategias de marketing y segmentación de clientes.

3.1.2. Datos

Esta base de datos de la empresa Universal presenta un tamaño de muestra de $n = 1404749$, abarcando un período de 2021 hasta 2024 (actual).

Variables

Variable	Descripción
Marca	Indica la preferencia del cliente: Cocoa, Manicho, Manicho con Galleta y Menta Gracial
Etiqueta Cliente	Tipo de cliente basado en el sector o tipo de establecimiento: Educativo Privado, Mini Super, Tienda Autoservicio, Tienda Barrotes Externo, Tienda Barrotes Interno, Tienda De Mostrador, Tienda Frutas Y Verdura, Educativo Público, Independiente Tradic, y Panadería.
Estrella Validada	Calificación de calidad del producto en una escala de 2 a 5 estrellas, y A+ para excelente.
Categoría de producto	Categoría del producto, dividida en 3 niveles: "Chocolates", "Confitería", "Modificadores".

Agencia ALI	Agencia o ubicación específica de distribución del producto: Portoviejo, Sto Domingo, Manta, Chone, Quevedo, Gye_Sur, Playas, Gye_Norte, Milagro, Salinas, Quito Norte, Quito Sur, Ibarra, Tulcan, Ambato, Cuenca, Loja, Machala, Latacunga, y Riobamba.
Código de cliente	Codificación de cada cliente que compra producto de la marca cliente
Atributo	Atributo temporal relacionado con el periodo en el que se realizó la transacción, desde enero-21 hasta mayo-24.
Valor	Valor numérico asociado a la transacción del monto de compra.

3.2. Diseño de la Investigación

El objetivo de este estudio es desarrollar cuatro modelos independientes de regresión logística binaria para clasificar las preferencias de los clientes hacia diferentes marcas de productos. Específicamente, buscamos determinar si un cliente prefiere "Cocoa", "Manicho", "Manicho Con Galleta" o "Menta Glacial". Cada modelo está diseñado para predecir la probabilidad de preferencia hacia una marca específica frente a las demás combinadas. Para lograr esto, hemos creado cuatro conjuntos de datos binarios, uno para cada marca de interés, en los cuales se comparan la preferencia por la marca seleccionada contra todas las demás categorías combinadas. Por ejemplo, el primer modelo clasificará a los clientes que prefieren "Cocoa" frente a aquellos que prefieren cualquier otra marca "No Cocoa".

3.3. Modelo logístico binario

$$\log\left(\frac{p}{1-p}\right) = b_0 + B_{1x9} * \text{Etiqu.Clie} + B_{1x4} * \text{Estr.Vali} + B_{1x3} * \text{cat.prod} + B_{1x19} * \text{Agen.ALI} + b_{36} * \text{cod.clie} + b_{37} * \text{Valor}$$

Coefficientes:

$B_{1 \times 9} = [b_1, b_2, \dots, b_9]$: Vector de coeficientes para cada categoría de etiqueta de clientes.

$B_{1 \times 4} = [b_{10}, b_{11}, b_{12}, b_{13}]$: Vector de coeficientes para cada categoría de estrella validada.

$B_{1 \times 19} = [b_{14}, b_{15}, \dots, b_{32}]$: Vector de coeficientes para cada categoría de agencia ALI.

$B_{1 \times 3} = [b_{33}, b_{34}, b_{35}]$: Vector de coeficientes para cada categoría de producto.

b_{36} : coeficiente asignado para cada cliente que ha realizado una transacción.

b_{37} : coeficiente asignado al valor de la transacción asociado a un monto de compra.

No se consideró las variables "Región" por su alta correlación con la variable "agencia ALI" provocando que el modelo no consiga encontrar una clasificación estable en la clasificación de las "Marcas". Tampoco se consideró la variable "Atributo temporal" dado que la naturaleza de variables temporales es reservada para análisis de serie temporales.

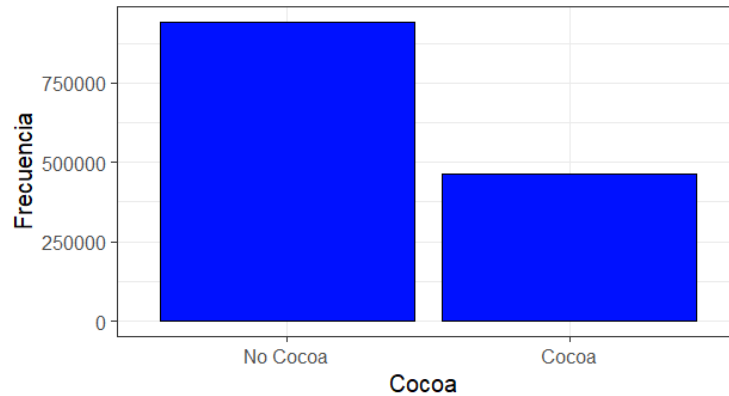
3.3.1. Análisis para Cocoa

Desarrollamos un modelo de regresión logística binaria para clasificar si un cliente prefiere la marca "Cocoa" en comparación con otras marcas, incluyendo "Manicho", "Manicho Con Galleta" y "Menta Glacial". Transformamos la variable "Marca" en una variable binaria donde "Cocoa" es la categoría de interés y todas las demás marcas se agrupan en una sola categoría denominada "No Cocoa".

Tabla 1. Frecuencia de Cocoa

	Frecuencia	Proporción
No Cocoa	941648	0.670
Cocoa	463101	0.329

Gráfico 1. Frecuencia de Cocoa



Para la marca de “cocoa”, la variable “Producto” se decidió no incluirlas en el modelo de regresión logística porque su inclusión causaba problemas que impedían que el modelo encontrara una clasificación estable, especialmente la categoría de “Producto modificadores” que presenta una alta correlación con la variable “Marca cocoa”. Al excluir estas variables, el modelo consiguió encontrar una clasificación estable.

El modelo propuesto es:

$$\log \left(\frac{P(Y = cacao)}{P(Y = No\ marca)} \right) = b_0 + B_{1x9} * Etiq.Clie + B_{1x4} * Estr.Vali + B_{1x19} * Agen.ALI + b_{33} * cod.clie + b_{34} * Valor$$

Matriz de confusión del modelo

La matriz de confusión se utiliza para evaluar el rendimiento del modelo de clasificación mostrando la comparación entre las predicciones del modelo y los valores reales observados en el conjunto de datos de prueba. La matriz de confusión para nuestro modelo es la siguiente:

Tabla 2. Matriz de confusión de Cocoa

Predicción	Real	
	Cocoa	No Cocoa
Cocoa	3090	1545
No Cocoa	135812	280978

Tabla 3. Métricas del Modelo para Cocoa

Acuracidad (aciertos)	0.674
Sensibilidad	0.022
Especificidad	0.995

El modelo muestra una acuracidad del 67.41% para clasificar correctamente la preferencia de la marca "Cocoa" y "No Cocoa". Sin embargo, la baja sensibilidad 2.22% indica que el modelo no está identificando correctamente la mayoría de los verdaderos positivos (clientes correctamente clasificados como "Cocoa"). Por otro lado, la alta especificidad 99.45%, muestra que el modelo es muy efectivo para identificar los verdaderos negativos (clientes correctamente clasificados como "No Cocoa"). Indicando que el modelo tiene dificultades para detectar correctamente los casos de "Cocoa".

Interpretación de los coeficientes y la razón de probabilidades (OR)

En la regresión logística, los coeficientes estimados representan el cambio en el logaritmo de las probabilidades $\log\left(\frac{p}{1-p}\right)$, de la variable dependiente por una unidad de cambio en la variable independiente. Los OR (razón de probabilidad) transforman estos coeficientes en una métrica más intuitiva que refleja el cambio multiplicativo en las probabilidades.

$$OR = \frac{P(Y = cacao)}{P(Y = No marca)} = e^{-b}$$

Tabla 4. Coeficientes y OR para preferir Cocoa

Categoría	Coeficiente (<i>b</i>)	OR (<i>e^b</i>)
Etiqueta Cliente: Panadería	2.648	14.1
Etiqueta Cliente: Tienda Frutas Y Verdura	1.489	4.4
Etiqueta Cliente: Tienda Barrotes Exte	1.310	3.7
Etiqueta Cliente: Tienda De Mostrador	1.302	3.7
Etiqueta Cliente: Tienda Barrotes Inte	1.289	3.6
Etiqueta Cliente: Tienda Autoservicio	1.236	3.4
Etiqueta Cliente: Independiente Tradic	1.209	3.4
Etiqueta Cliente: Mini Super	1.102	3.0
Agencia ALI: Chone	1.022	2.8
Agencia ALI: Portoviejo	0.894	2.4

Nota: Tabla completa con todos los OR en anexo 1

Gráfico 2. Razón de Probabilidad para Preferir Cocoa



Los coeficientes positivos y sus correspondientes OR, indican que los tipos de clientes listados tienen una probabilidad significativamente mayor de preferir la marca "Cocoa" en comparación con "No Cocoa". Por ejemplo:

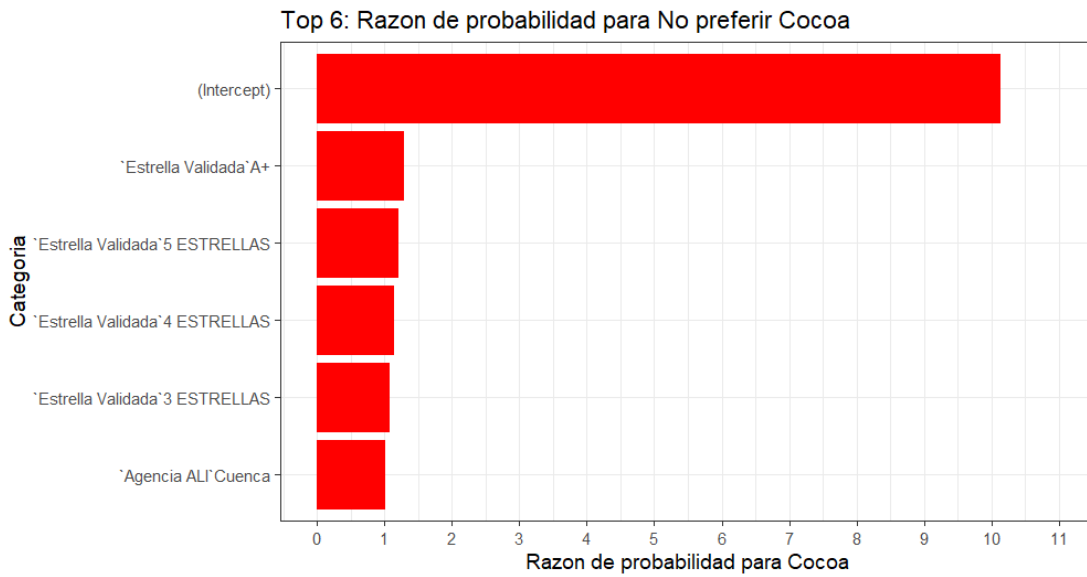
- Para Los clientes identificados como "Panadería" tienen aproximadamente 14.12 veces más probabilidades de preferir "Cocoa" en comparación con "No preferir Cocoa", manteniendo constantes todas las demás variables.
- Los clientes de "Tienda Frutas Y Verdura" tienen aproximadamente 4.43 veces más probabilidades de preferir "Cocoa" en comparación con "No Cocoa".
- Los clientes de "Tienda Barrotes Exte" tienen aproximadamente 3.70 veces más probabilidades de preferir "Cocoa" en comparación con "No Cocoa".
- Los clientes de "Tienda De Mostrador" tienen aproximadamente 3.68 veces más probabilidades de preferir "Cocoa" en comparación con "No Cocoa".
- Los clientes de "Tienda Barrotes Inte" tienen aproximadamente 3.63 veces más probabilidades de preferir "Cocoa" en comparación con "No Cocoa".

Así sucesivamente la interpretación para las demás categorías.

Tabla 5. Coeficientes y OR Para No Preferir Cocoa

Categoría	Coeficiente (<i>b</i>)	1/OR (e^{-b})
Constante	-2.316	10.1
Estrella Validada: A+	-0.254	1.3
Estrella Validada: 5 ESTRELLAS	-0.192	1.2
Estrella Validada: 4 ESTRELLAS	-0.128	1.1
Estrella Validada: 3 ESTRELLAS	-0.067	1.1
Agencia ALI: Cuenca	-0.013	1.0

Gráfico 3. Razón de Probabilidad para No Preferir Cocoa



Los coeficientes negativos y sus correspondientes $(OR)^{-1}$, indican que las categorías listadas disminuyen la probabilidad de que un cliente prefiera "Cocoa" en comparación con "No Cocoa". Por ejemplo:

- La constante del modelo indica que preferir "No Cocoa" cuando todas las variables son cero. Es decir, en ausencia de otras influencias, las probabilidades de preferir "No Cocoa" son aproximadamente 10.14 veces más altas que las de preferir "Cocoa".
- Para productos con una calificación de A+, las probabilidades de preferir "No Cocoa" son aproximadamente 1.29 veces más altas que las de preferir "Cocoa". Esto indica una ligera preferencia por "No Cocoa" en productos con esta calificación.
- Para productos con una calificación de 5 estrellas, las probabilidades de preferir "No Cocoa" son aproximadamente 1.21 veces más altas que las de preferir "Cocoa".

Así sucesivamente la interpretación para las demás categorías.

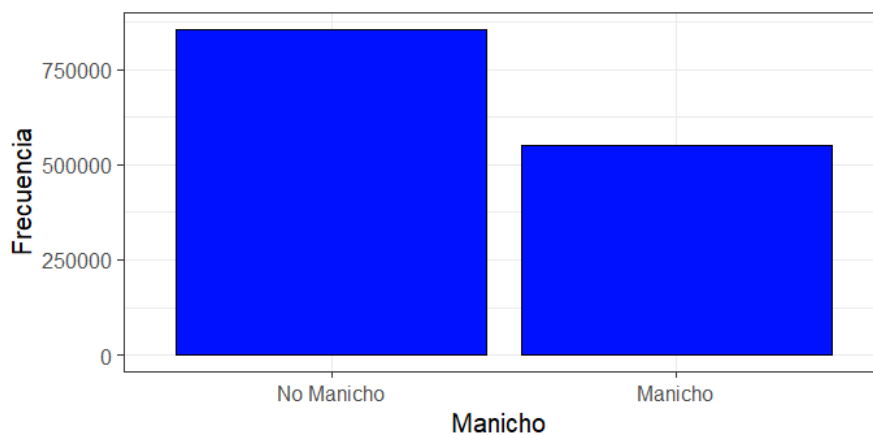
3.3.2. Análisis para Manicho

Transformamos a una variable binaria "Manicho", siendo la categoría de interés y todas las demás marcas se agrupan en una sola categoría denominada "No Manicho".

Tabla 6. Frecuencia de Manicho

	Frecuencia	Proporción
No Manicho	855516	0.609
Manicho	549233	0.391

Gráfico 4. Frecuencia de Manicho



A diferencia de "Cocoa", para la marca "Manicho" la variable "Producto" se decidió incluirlas en el modelo de regresión logística porque no se evidencio que presente problema en la clasificación del modelo, incluso se evidencio que mejoro las métricas en las clasificaciones de la "marca".

El modelo propuesto es:

$$\log \left(\frac{P(Y = Manicho)}{P(Y = No Manicho)} \right) = b_0 + B_{1x9} * Etiq.Clie + B_{1x4} * Estr.Vali + B_{1x3} * cat.prod + B_{1x19} * Agen.ALI + b_{36} * cod.clie + b_{37} * Valor$$

Tabla 7. Matriz de Confusión de Manicho

Predicción	Real	
	Manicho	No Manicho
Manicho	150450	9600
No Manicho	14579	246796

Tabla 8. Métricas del Modelo de Manicho

Acuracidad (aciertos)	0.943
Sensibilidad	0.912
Especificidad	0.963

El modelo para clasificar la preferencia por "Manicho" muestra una alta acuracidad de (94.26%), sensibilidad (91.17%) y especificidad (96.26%). Esto indica que el modelo es efectivo para identificar tanto los clientes que prefieren "Manicho" como aquellos que no lo prefieren.

Interpretación de los coeficientes y la razón de probabilidades (OR)

$$OR = \frac{P(Y = Manicho)}{P(Y = No Manicho)} = e^{-b}$$

Tabla 9. Coeficientes y OR para preferir Manicho

Categoría	Coeficiente (<i>b</i>)	OR (<i>e^b</i>)
Etiqueta Cliente: Panadería	2.419	11.24
Etiqueta Cliente: Independiente Tradic	1.959	7.09
Etiqueta Cliente: Tienda Frutas Y Verd	1.953	7.05
Etiqueta Cliente: Tienda Barrotes Exte	1.826	6.21
Etiqueta Cliente: Tienda Barrotes Inte	1.758	5.80
Etiqueta Cliente: Mini Super	1.719	5.58
Etiqueta Cliente: Tienda De Mostrador	1.668	5.30

Etiqueta Cliente: Tienda Autoservicio	1.424	4.15
Etiqueta Cliente: Educativo Público	0.768	2.16
Agencia ALI: Tulcan	0.645	1.91

Nota: Tabla completa con todos los OR en anexo 2

Gráfico 5. Razón de Probabilidad para Preferir Manicho



Los coeficientes positivos y sus correspondientes OR, indican que los tipos de clientes listados tienen una probabilidad significativamente mayor de preferir la marca "Manicho" en comparación con "No Manicho". Por ejemplo:

- Los clientes identificados como "Panadería" tienen aproximadamente 11.24 veces más probabilidades de preferir "Manicho" en comparación con "No Manicho", manteniendo constantes todas las demás variables.
- Los clientes de "Independiente Tradic" tienen aproximadamente 7.09 veces más probabilidades de preferir "Manicho" en comparación con "No Manicho".
- Los clientes de "Tienda Frutas Y Verd" tienen aproximadamente 7.05 veces más probabilidades de preferir "Manicho" en comparación con "No Manicho".
- Los clientes de "Tienda Barrotes Exte" tienen aproximadamente 6.21 veces más probabilidades de preferir "Manicho" en comparación con "No Manicho".

- Los clientes de "Tienda Barrotes Inte" tienen aproximadamente 5.80 veces más probabilidades de preferir "Manicho" en comparación con "No Manicho".

Así sucesivamente la interpretación para las demás categorías.

Tabla 10. Coeficientes y OR para NO preferir Manicho

Categoría	Coeficiente (<i>b</i>)	1/OR (e^{-b})
Categoría: Modificadores	-3.88E+10	∞
Categoría: Confitería	-3.93E+06	∞
Constante	-3.108	22.38
Agencia ALI: Manta	-0.631	1.88
Estrella Validada:A+	-0.598	1.82
Agencia ALI: Portoviejo	-0.467	1.60
Agencia ALI: Loja	-0.433	1.54
Agencia AL: Quito Norte	-0.425	1.53
Estrella Validada: 5 ESTRELLAS	-0.374	1.45

Gráfico 6. Razón de Probabilidad para No Preferir Manicho



Los coeficientes negativos y sus correspondientes $(OR)^{-1}$, indican que las categorías listadas disminuyen la probabilidad de que un cliente prefiera "Manicho" en comparación con "No Manicho". Por ejemplo:

- Un coeficiente extremadamente grande y negativo con un $1/OR$ infinito indica que la categoría "Modificadores" prácticamente garantiza que un cliente no prefiera "Manicho". Es decir, es casi imposible que un cliente prefiera "Manicho" si el producto es un modificador.
- Similar a la categoría anterior, un coeficiente extremadamente grande y negativo con un $1/OR$ infinito indica que la categoría "Confitería" también garantiza que un cliente no prefiera "Manicho". Es decir, es casi imposible que un cliente prefiera "Manicho" si el producto es una confitería.
- La constante del modelo indica que preferir "No Manicho" cuando todas las variables independientes son cero. Es decir, la ausencia de otras influencias, las probabilidades de preferir "No Manicho" son aproximadamente 22.38 veces más altas que las de preferir "Manicho".
- Los productos distribuidos por la agencia en Manta tienen aproximadamente 1.88 veces más probabilidades de ser preferidos como "No Manicho" que como "Manicho". Esto indica una ligera preferencia por "No Manicho" en productos distribuidos por esta agencia.

Así sucesivamente la interpretación para las demás categorías.

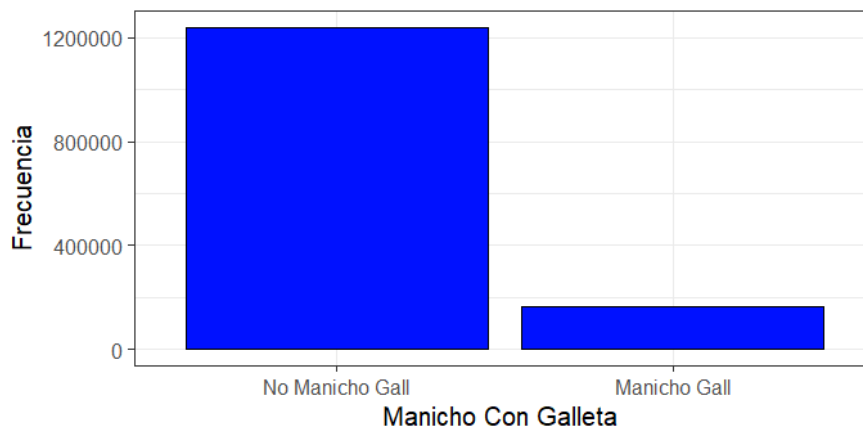
3.3.3. Análisis para Manicho con Galleta

Transformamos a una variable binaria "Manicho con Galleta", siendo la categoría de interés y todas las demás marcas se agrupan en una sola categoría denominada "No Manicho con Galleta".

Tabla 11. Frecuencia de Manicho con Galleta

	Frecuencia	Proporción
No Manicho Galleta	1238705	0.882
Manicho Galleta	16604	0.118

Gráfico 7. Frecuencia de Manicho con Galleta



A diferencia de “Cocoa”, para la marca “Manicho con Galleta” la variable “Producto” se decidió incluirlas en el modelo de regresión logística porque no se evidencio que presente problema en la clasificación del modelo, incluso se evidencio que mejoro las métricas en la clasificación de la “marca”

El modelo propuesto es:

$$\log \left(\frac{P(Y = \text{Manicho Gall})}{P(Y = \text{No Manicho Gall})} \right) = b_0 + B_{1x9} * \text{Etiqu. Clie} + B_{1x4} * \text{Estr. Vali} + B_{1x3} * \text{cat. prod} + B_{1x19} * \text{Agen. ALI} + b_{36} * \text{cod. clie} + b_{37} * \text{Valor}$$

Tabla 12. Matriz de Confusión de Manicho con Galleta

Predicción	Real	
	Manicho con Galleta	No Manicho con Galleta
Manicho con Galleta	30752	5824
No Manicho con Galleta	18907	365942

Tabla 13. Métricas del Modelo

Acuracidad (aciertos)	0.941
Sensibilidad	0.619
Especificidad	0.841

El modelo para clasificar la preferencia por "Manicho Con Galleta" muestra una alta acuracidad (94.13%) y especificidad (98.43%). Esto indica que el modelo es muy efectivo para identificar los clientes que no prefieren "Manicho Con Galleta". Sin embargo, la sensibilidad (61.93%) es menor, lo que sugiere que el modelo tiene dificultades para identificar correctamente a todos los clientes que prefieren "Manicho Con Galleta".

Interpretación de los coeficientes y la razón de probabilidades (OR)

$$OR = \frac{P(Y = \text{Manicho Gall})}{P(Y = \text{No Manicho Gall})} = e^{-b}$$

Tabla 14. Coeficientes y OR para preferir Manicho con Galleta

Categoría	Coefficiente (<i>b</i>)	OR (e^b)
Constante	3.108	22.38
Agencia ALI: Manta	0.631	1.88
Estrella Validada: A+	0.598	1.82
Agencia ALI: Portoviejo	0.467	1.60
Agencia ALI: Loja	0.433	1.54
Agencia ALI: Quito Norte	0.425	1.53
Estrella Validada: 5 ESTRELLAS	0.374	1.45
Agencia ALI: Quito Sur	0.364	1.44
Agencia ALI: Sto Domingo	0.349	1.42
Agencia ALI: Cuenca	0.254	1.29

Nota: Tabla completa con todos los OR en anexo 3.

Gráfico 8. Razón de probabilidad para preferir Manicho con Galleta



Los coeficientes positivos y sus correspondientes OR, indican que los tipos de clientes listados tienen una probabilidad significativamente mayor de preferir la marca "Manicho con Galleta" en comparación con "No Manicho con Galleta". Por ejemplo:

- La constante del modelo indica de preferir "Manicho Con Galleta" cuando todas las variables independientes son cero. Es decir, una ausencia de otras influencias, las probabilidades de preferir "Manicho Con Galleta" son aproximadamente 22.38 veces más altas que las de preferir "No Manicho Con Galleta".
- Los productos distribuidos por la agencia en Manta tienen aproximadamente 1.88 veces más probabilidades de ser preferidos como "Manicho Con Galleta" que como "No Manicho Con Galleta". Esto indica una ligera preferencia por "Manicho Con Galleta" en productos distribuidos por esta agencia
- Para productos con una calificación de A+, las probabilidades de preferir "Manicho Con Galleta" son aproximadamente 1.82 veces más altas que las de preferir "No Manicho Con Galleta". Esto indica una ligera preferencia por "Manicho Con Galleta" en productos con esta calificación.

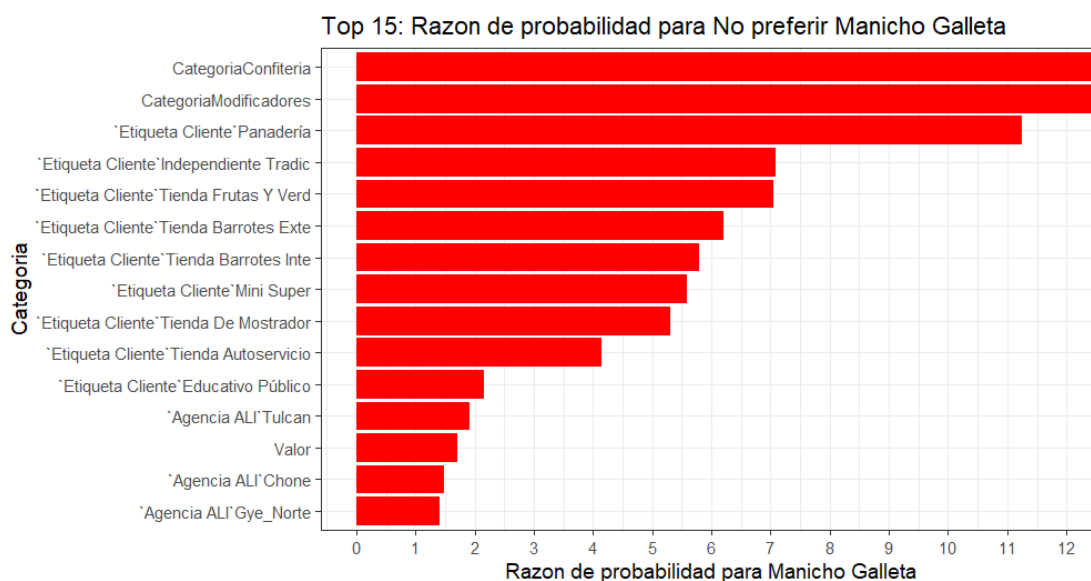
- Los productos distribuidos por la agencia en Portoviejo tienen aproximadamente 1.60 veces más probabilidades de ser preferidos como "Manicho Con Galleta" que como "No Manicho Con Galleta".
- Los productos distribuidos por la agencia en Loja tienen aproximadamente 1.54 veces más probabilidades de ser preferidos como "Manicho Con Galleta" que como "No Manicho Con Galleta".

Así sucesivamente la interpretación para las demás categorías.

Tabla 15. Coeficientes y OR para NO preferir Manicho con Galleta

Categoría	Coeficiente (<i>b</i>)	1/OR (e^{-b})
CategoriaConfiteria	-3.52E+10	∞
CategoriaModificadores	-9.10E+09	∞
Etiqueta Cliente: Panadería	-2.419	11.24
Etiqueta Cliente: Independiente Tradic	-1.959	7.09
Etiqueta Cliente: Tienda Frutas Y Verd	-1.953	7.05
Etiqueta Cliente: Tienda Barrotes Exte	-1.826	6.21
Etiqueta Cliente: Tienda Barrotes Inte	-1.758	5.80

Gráfico 9. Razón de Probabilidad para No Preferir Manicho con Galleta



Los coeficientes negativos y sus correspondientes $(OR)^{-1}$, indican que las categorías listadas disminuyen la probabilidad de que un cliente prefiera "Manicho con Galleta" en comparación con "No Manicho con Galleta". Por ejemplo:

- Un coeficiente extremadamente grande y negativo con un $1/OR$ infinito indica que la categoría "Confitería" prácticamente garantiza que un cliente no prefiera "Manicho Con Galleta". Es decir, es casi imposible que un cliente prefiera "Manicho Con Galleta" si el producto es una confitería.
- Similar a la categoría anterior, un coeficiente extremadamente grande y negativo con un $1/OR$ infinito indica que la categoría "Modificadores" también garantiza que un cliente no prefiera "Manicho Con Galleta". Es decir, es casi imposible que un cliente prefiera "Manicho Con Galleta" si el producto es un modificador.
- Los clientes identificados como "Panadería" tienen aproximadamente 11.24 veces más probabilidades de preferir "No Manicho Con Galleta" en comparación con "Manicho Con Galleta". Esto indica una fuerte preferencia por "No Manicho Con Galleta" entre estos clientes.
- Los clientes de "Independiente Tradic" tienen aproximadamente 7.09 veces más probabilidades de preferir "No Manicho Con Galleta" en comparación con "Manicho Con Galleta".

Así sucesivamente la interpretación para las demás categorías.

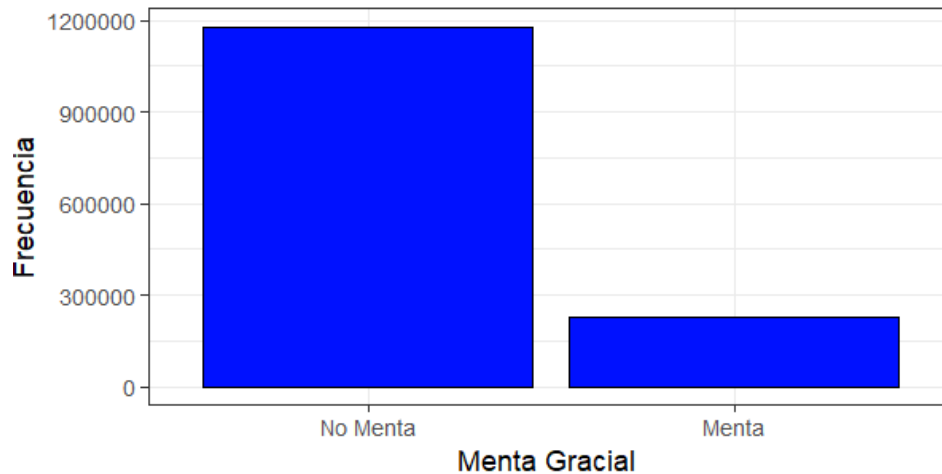
3.3.4. Análisis para Manicho con Galleta

Transformamos la variable "Marca" en una variable binaria donde "Menta Gracial" es la categoría de interés y todas las demás marcas se agrupan en una sola categoría denominada "No Menta Gracial".

Tabla 16. Frecuencia de Menta Glacial

	Frecuencia	Proporción
No Menta Glacial	1178378	0.839
Menta Glacial	226371	0.161

Gráfico 10. Frecuencia de Menta Glacial



Para la marca de “Menta Glacial”, la variable “Producto” se decidió no incluirlas en el modelo de regresión logística porque su inclusión causaba problemas que impedían que el modelo encontrara una clasificación estable, especialmente la categoría de “Producto modificadores” que presenta una alta correlación con la variable “Marca Menta Glacial”. Al excluir estas variables, el modelo consiguió encontrar una clasificación estable.

El modelo propuesto es:

$$\log \left(\frac{P(Y = menta)}{P(Y = No menta)} \right) = b_0 + B_{1x9} * Etiq.Clie + B_{1x4} * Estr.Vali + B_{1x19} * Agen.ALI + b_{33} * cod.clie + b_{34} * Valor$$

Tabla 17. Matriz de Confusión de Menta

Predicción	Real	
	Menta Gracial	No Menta Gracial
Menta Gracial	0	4
No Menta Gracial	67835	353586

Tabla 18. Métricas del Modelo

Acuracidad (aciertos)	0.839
Sensibilidad	0.000
Especificidad	1.000

El modelo para clasificar la preferencia por "Menta Glacial" muestra una precisión general del 83.9%. Sin embargo, esta alta acuracidad es engañosa debido a la desbalanceada distribución de las clases en los datos. La sensibilidad del 0% indica que el modelo no es capaz de identificar ningún cliente que prefiera "Menta Glacial", mientras que la especificidad del 100% muestra que el modelo identifica correctamente todos los clientes que no prefieren "Menta Glacial".

Interpretación de los coeficientes y la razón de probabilidades (OR)

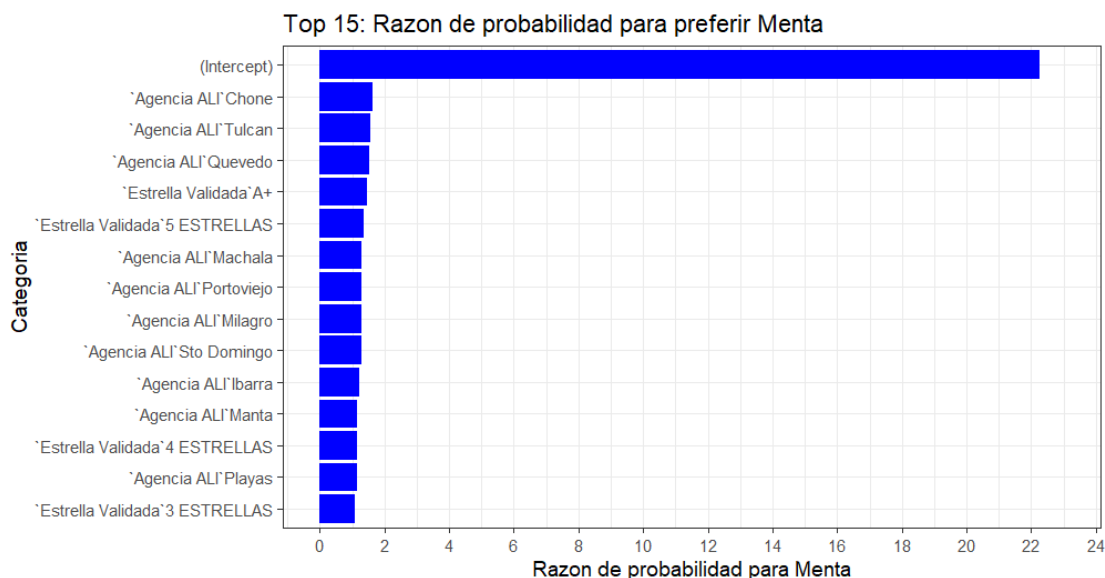
$$OR = \frac{P(Y = menta)}{P(Y = No menta)} = e^{-b}$$

Tabla 19. Coeficientes y OR para preferir Menta Glacial

Categoría	Coeficiente (<i>b</i>)	OR (<i>e^b</i>)
Constante	3.103	22.26
Agencia ALL: Chone	0.491	1.63
Agencia ALL: Tulcan	0.436	1.55
Agencia ALL: Quevedo	0.415	1.51
Estrella Validada: A+	0.377	1.46
Estrella Validada5: ESTRELLAS	0.306	1.36

Nota: Tabla completa con todos los OR en anexo 4.

Gráfico 11. Razón de Probabilidad para preferir Menta Glacial



Los coeficientes positivos y sus correspondientes OR, indican que los tipos de clientes listados tienen una probabilidad significativamente mayor de preferir la marca "Menta Gracial" en comparación con "No Menta Gracial". Por ejemplo:

- La constante del modelo indica que preferir "Menta Glacial" cuando todas las variables independientes son cero. Es decir, que la ausencia de otras influencias, las probabilidades de preferir "Menta Glacial" son aproximadamente 22.26 veces más altas que las de preferir "No Menta Glacial".
- Los productos distribuidos por la agencia en Chone tienen aproximadamente 1.63 veces más probabilidades de ser preferidos como "Menta Glacial" que como "No Menta Glacial". Esto indica una ligera preferencia por "Menta Glacial" en productos distribuidos por esta agencia.
- Los productos distribuidos por la agencia en Tulcán tienen aproximadamente 1.55 veces más probabilidades de ser preferidos como "Menta Glacial" que como "No Menta Glacial".

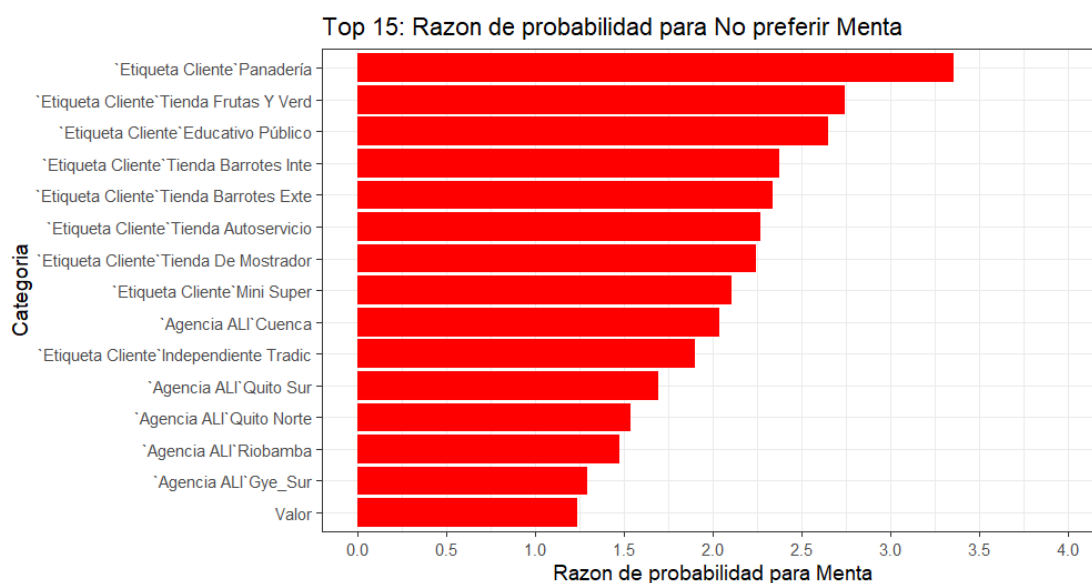
- Los productos distribuidos por la agencia en Quevedo tienen aproximadamente 1.51 veces más probabilidades de ser preferidos como "Menta Glacial" que como "No Menta Glacial".
- Para productos con una calificación de A+, las probabilidades de preferir "Menta Glacial" son aproximadamente 1.46 veces más altas que las de preferir "No Menta Glacial". Esto indica una ligera preferencia por "Menta Glacial" en productos con esta calificación.

Así sucesivamente la interpretación para las demás categorías.

Tabla 20. Coeficientes y OR para No Preferir Menta Glacial

Categoría	Coeficiente (<i>b</i>)	1/OR (e^{-b})
Etiqueta Cliente: Panadería	-1.211	3.36
Cliente: Tienda Frutas Y Verd	-1.010	2.74
Cliente: Educativo Público	-0.974	2.65
Cliente: Tienda Barrotes Inte	-0.865	2.38
Cliente: Tienda Barrotes Exte	-0.850	2.34
Cliente:Tienda Autoservicio	-0.818	2.27

Gráfico 12. Razón de Probabilidad para No Preferir Menta Glacial



Los coeficientes negativos y sus correspondientes $(OR)^{-1}$, indican que las categorías listadas disminuyen la probabilidad de que un cliente prefiera "Menta Gracial" en comparación con "No Menta Gracial". Por ejemplo:

- Los clientes identificados como "Panadería" tienen aproximadamente 3.36 veces más probabilidades de preferir "No Menta Glacial" en comparación con "Menta Glacial". Esto indica una fuerte preferencia por "No Menta Glacial" entre estos clientes.
- Los clientes de "Tienda Frutas Y Verd" tienen aproximadamente 2.74 veces más probabilidades de preferir "No Menta Glacial" en comparación con "Menta Glacial".
- Los clientes de "Educativo Público" tienen aproximadamente 2.65 veces más probabilidades de preferir "No Menta Glacial" en comparación con "Menta Glacial".

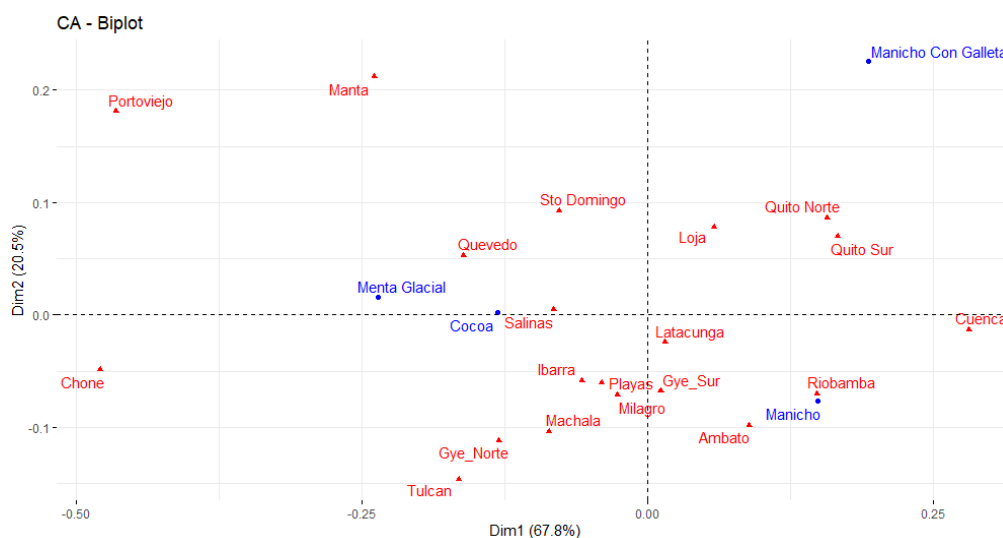
Así sucesivamente la interpretación para las demás categorías.

3.4. Análisis por análisis de correspondencia

En este estudio, buscamos entender las preferencias de los clientes en las marcas "Cocoa", "Manicho", "Manicho con Galleta" y "Menta Glacial". El análisis de correspondencia nos permite explorar cómo estas preferencias se relacionan con otras variables categóricas clave, como la etiqueta del cliente, la región, la etiqueta del cliente y entre otras características. Al representar gráficamente las asociaciones entre categorías permite realizar un clustering de las variables, agrupando categorías similares y facilitando la interpretación de los resultados.

Comenzamos haciendo una asociación entre la marca del producto y las agencias.

Gráfico 13. Asociación entre marca y agencia



El gráfico muestra la relación entre dos variables categóricas: Marca y Agencia ALI.

- Los puntos azules representan diferentes Marcas.
- Los puntos rojos representan diferentes Agencias ALI.

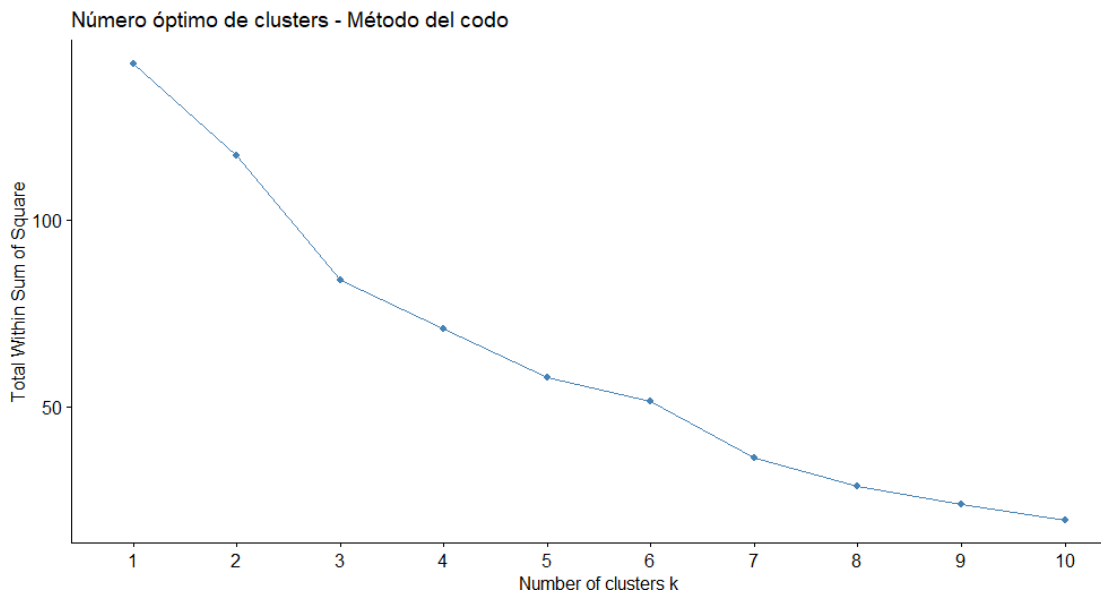
El eje Dim1 (Dimensión 1) y el eje Dim2 (Dimensión 2) son las principales dimensiones de variabilidad en los datos. En este caso, Dim1 explica el 67.8% de la variabilidad y Dim2 explica el 20.5%.

En conjunto, Dim1 y Dim2 explican el 88.3% (67.8% + 20.5%) de la variabilidad en los datos, lo cual es una proporción bastante alta. Esto indica que el gráfico captura la mayor parte de la información importante sobre las relaciones entre las marcas y las agencias.

Donde la proximidad entre puntos de diferentes categorías indica una asociación más fuerte. Entre los resultados de la gráfica son los siguientes:

- Manicho Con Galleta está cerca de la esquina superior derecha, lo que sugiere que está fuertemente asociado con algunas agencias ubicadas en esa dirección.
- Menta Glacial y Cocoa están relativamente cerca una de la otra en la parte izquierda del gráfico, lo que sugiere que estas marcas tienen una distribución similar entre ciertas agencias.
- Las agencias como Portoviejo, Manta, Quevedo, Sto Domingo, y Chone están más dispersas, lo que indica una distribución variada de las marcas en estas agencias.
- Agencias como Gye_Norte, Tulcan, y Ibarra están más agrupadas cerca de ciertas marcas, lo que sugiere una relación específica con estas marcas.
- Manicho Con Galleta y Manicho parecen tener diferentes asociaciones con las agencias, lo que puede indicar diferentes estrategias de distribución o preferencias de los consumidores en distintas regiones.
- Manicho Con Galleta está más cerca de Quito Norte y Quito Sur, lo que podría indicar una mayor preferencia o presencia en estas agencias.
- Menta Glacial y Cocoa están asociados con agencias como Salinas y Quevedo.

Gráfico 15. Número óptimo de Clusters - Método del Codo



Del gráfico arriba se observa que los números de clúster más adecuados usando el método del codo, serían 3, 5 o 7 clúster. Escoger 5 o más clúster no tendría sentido dado que solo estamos evaluando 4 productos, lo escoger más clúster podría dificultar la interpretación de los asociados a las variables y la marca.

Además, la interpretación de los clústeres debe considerar la baja variabilidad explicada. Aunque los clústeres pueden ayudar a identificar grupos similares, la baja variabilidad implica que estos grupos pueden no ser completamente representativos de todas las relaciones en los datos.

Del clúster observamos que la categoría del clúster “educativo privado” y “educativo público” están alejados del clúster de manicho, Manicho con galleta, cocoa y menta. De la misma manera, el clúster con la categoría de Tulcán también está alejado de los otros dos clústeres.

Gráfico 16. Clusters de Clientes en el Espacio MCA

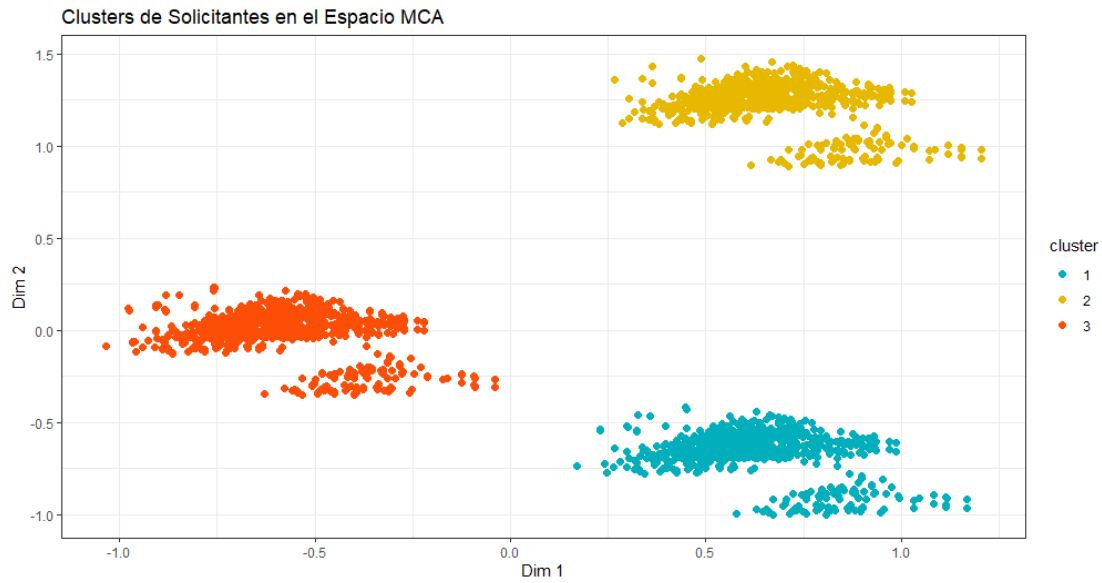


Tabla 21. Relación entre los cluster y la marca

Cluster	Cocoa	Manicho	Manicho Con Galleta	Menta Glacial
1	463101	0	0	0
2	0	0	0	226371
3	0	549233	166044	0

Finalmente, cuando agrupamos en cluster a los clientes, se evidencia que los clientes están distribuidos exactamente las marcas de “Cocoa”, “Manicho y Manicho con chocolate” y “Menta Gracial”, tal como se había esperado, dado que las marcas están dentro de la categoría del producto.

CAPÍTULO 4

4.1. Conclusiones

La importancia de este trabajo se basa en mostrar como la evidencia práctica y las herramientas estadísticas pueden otorgar una mejor asignación de recursos llegando a una granularidad por cliente, ya que el enfoque de toda empresa es maximizar los ingresos con menor inversión de recursos. Esta demostrado que en la industria de dulces y snacks existe una estacionalidad dentro de los meses de septiembre y noviembre, el complejo mercado ecuatoriano obliga a las firmas a buscar generar crecimientos frente a la venta de periodos pasados en los meses comprendidos entre enero y junio, dado que esto significa una mayor utilidad al cierre del periodo. Para lograr esto las firmas deben focalizar sus recursos en sus mayores áreas de oportunidad, en este caso puntual nos focalizamos en captar mas clientes, desarrollar clusters de clientes con características similares e identificar clientes con mayor probabilidad de compra y no compra. Puesto con estas directrices claras se desarrollan estrategias comerciales para grupos específicos, se optimizan los recursos de inversión y se pronostica con mayor asertividad la demanda de los productos que comercializa la empresa Universal Sweet Industries.

Por este motivo, tener evidencia estadística que permita a los tomadores de decisiones saber en que clientes invertir una acción promocional que permita captar nuevos clientes basado en características semejantes y probabilidades de compra, ayuda a mejorar el performance del negocio.

Los resultados mostrados en el capítulo anterior evidencian una relación con las variables identificadas en literatura previa estas reflejan las similitudes entre los clientes que compran Manicho, Cocoa, Manicho con Galleta y Menta Glacial. Estas Cualidades son provenientes de un comportamiento histórico comprendido desde enero del 2021 hasta abril del 2024

En conjunto estas características pueden brindar una propensión de comprar un producto de la empresa o no, ya que, si bien no haber comprado estos productos en el periodo de análisis de este estudio no significa que no adquiera el producto en un mes futuro, si se la puede interpretar como una deficiencia del modelo de distribución actual de la empresa.

Cabe recalcar que el presente estudio posee ciertas debilidades. Si bien no es un estudio causal en estricto rigor la implementación de varios controles en el uso de las herramientas estadísticas reduce los sesgos por causas observacionales, sin embargo, el enfoque de este estudio es brindar metodologías que permitan a la firma identificar variables que ayuden a mejorar la asertividad de la demanda en el mercado ecuatoriano.

4.2. Recomendaciones

Con las técnicas aplicadas es recomendable utilizar nuevos modelos econométricos para reducir sesgos no observables y obtener estimaciones causales respecto a las propensiones de comprar o no un producto, se sugiere levantar más información de los clientes como Nivel de compra de productos competidores, días de crédito del cliente, días de recompra del cliente y tamaño de la tienda, puesto que hoy las variables con las que se realizó el presente estudio en su mayoría son Cualitativas. Incorporar variables Cuantitativas permitirá realizar otras metodologías de clusterización como K-means y de esta forma tener un comparativo de los modelos para seleccionar el de mayor poder explicativo.

De acuerdo con los resultados obtenidos por las categorías antes mencionadas, sería factible que la empresa implemente una herramienta analítica interna automatizada para la generación de pronósticos de los diferentes grupos de clientes, esto permitiría tomar decisiones para cada clúster de clientes logrando así ahorros significativos en el trade spend.

Bibliografía

- C. A. da Silva, D. B.-d.-C. (2009). *Agro-industries for Development*. Cambridge University Press.
- Cálad, F. (2015). Segmentación de clientes automatizada a partir de técnicas de minería de datos (K-Means Clustering). *Universidad EIA*.
- Coelho, D. M. (2016). Towards of a Business Intelligence Platform.
- Dahlén C, & Siwerz R. (2017). Predicting sales in a food store department using machine learning.
- Díaz Sepúlveda, J. F., & Correa, J. C. (2015). Comparación entre árboles de regresión CART y regresión lineal. *Comunicaciones En Estadística*.
- Díaz, J., & Correa, J. (2013). Comparación entre árboles de regresión CART y regresión lineal. *Universidad Santo Tomas*.
- Eguiguren Calisto, .. (2019). Modelos de proyección de demanda para productos de alta volatilidad y bajo volumen en ventas dentro de una empresa de alimentos.
- Hung, P. D., Lien, N. T., & Ngoc, N. D. (2019). Customer Segmentation Using Hierarchical Agglomerative Clustering. *International Journal of Computer Applications*.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction. *Springer New York Heidelberg Dordrecht London*.
- Kassambara, A. (2015). Alboukadel Kassambara Practical Guide To Cluster Analysis in R.
- Palacios Utreras, C. (2020). Análisis y predicción de las tendencias de venta en el mercado usando. *Universidad San Francisco De Quito. Bachelor's thesis, Quito*.
- Peña, D. (2002). Analisis de Datos Multivariantes. *McGraw-Hill*. Obtenido de https://www.researchgate.net/profile/Daniel_Pena4/publication/40944325_Analisis_de_Datos_Multivariantes/links/549154880cf214269f27ffae/Analisis-de-Datos-Multivariantes.pdf?origin=publication_detail
- Quispe, A., Calla, K., Yangali, J., Rodríguez, J., & Pucamayo, I. (2019). Estadística no paramétrica aplicada a la Investigación Científica. *Editorial EIDEC*.

- Siwerz, R., & Dahlén, C. (2017). Predicting sales in a food store. *KTH ROYAL INSTITUTE OF TECHNOLOGY*.
- Thron, T. N. (2007). Evaluating alternative supply chain structures for perishable products. *The International Journal of Logistics*.
- Wang, J., Liu, G. Q., & Liu, L. (2019). A Selection of Advanced Technologies for Demand Forecasting in the Retail Industry. *IEEE International Conference on Big Data Analytics*.

ANEXOS

Anexo 1. Coeficientes del modelo de Cocoa

Categoría	Coeficiente (<i>b</i>)	OR (<i>e^b</i>)
Etiqueta Cliente: Panadería	2.648	14.12
Etiqueta Cliente: Tienda Frutas Y Verd	1.489	4.43
Etiqueta Cliente: Tienda Barrotes Exte	1.310	3.70
Etiqueta Cliente: Tienda De Mostrador	1.302	3.68
Etiqueta Cliente: Tienda Barrotes Inte	1.289	3.63
Etiqueta Cliente: Tienda Autoservicio	1.236	3.44
Etiqueta Cliente: Independiente Tradic	1.209	3.35
Etiqueta Cliente: Mini Super	1.102	3.01
Agencia ALI: Chone	1.022	2.78
Agencia ALI: Portoviejo	0.894	2.44
Etiqueta Cliente: Educativo Público	0.696	2.01
Agencia ALI: Gye_Norte	0.619	1.86
Agencia ALI: Manta	0.575	1.78
Agencia ALI: Salinas	0.494	1.64
Agencia ALI: Gye_Sur	0.372	1.45
Agencia ALI: Latacunga	0.334	1.40
Agencia ALI: Quevedo	0.282	1.33
Agencia ALI: Playas	0.250	1.28
Agencia ALI: Machala	0.244	1.28
Agencia ALI: Sto Domingo	0.206	1.23
Agencia ALI: Milagro	0.199	1.22
Agencia ALI: Loja	0.143	1.15
Agencia ALI: Tulcan	0.136	1.15
Agencia ALI: Riobamba	0.126	1.13
Agencia ALI: Ibarra	0.116	1.12

Agencia ALI: Quito Sur	0.106	1.11
Agencia ALI: Quito Norte	0.076	1.08
Valor	0.000	1.00
Cod_Solicitante	0.000	1.00
Agencia ALI: Cuenca	-0.013	0.987
Estrella Validada: 3 ESTRELLAS	-0.067	0.936
Estrella Validada: 4 ESTRELLAS	-0.128	0.880
Estrella Validada: 5 ESTRELLAS	-0.192	0.825
Estrella Validada: A+	-0.254	0.776
(Intercept)	-2.316	0.099

Anexo 2. Coeficientes del modelo de Manicho

Categoría	coeficiente	OR
` Etiqueta Cliente ` Panadería	2.419	11.24
` Etiqueta Cliente ` Independiente Tradic	1.959	7.09
` Etiqueta Cliente ` Tienda Frutas Y Verd	1.953	7.05
` Etiqueta Cliente ` Tienda Barrotes Exte	1.826	6.21
` Etiqueta Cliente ` Tienda Barrotes Inte	1.758	5.80
` Etiqueta Cliente ` Mini Super	1.719	5.58
` Etiqueta Cliente ` Tienda De Mostrador	1.668	5.30
` Etiqueta Cliente ` Tienda Autoservicio	1.424	4.15
` Etiqueta Cliente ` Educativo Público	0.768	2.16
` Agencia ALI ` Tulcan	0.645	1.91
Valor	0.532	1.70
` Agencia ALI ` Chone	0.398	1.49
` Agencia ALI ` Gye_Norte	0.346	1.41
` Agencia ALI ` Ibarra	0.168	1.18

` Agencia ALI ` Machala	0.084	1.09
` Agencia ALI ` Salinas	0.041	1.04
` Agencia ALI ` Riobamba	0.029	1.03
` Agencia ALI ` Playas	0.003	1.00
Cod_Solicitante	0.000	1.00
` Estrella Validada ` 3 ESTRELLAS	-0.054	0.95
` Agencia ALI ` Milagro	-0.077	0.93
` Agencia ALI ` Latacunga	-0.123	0.88
` Estrella Validada ` 4 ESTRELLAS	-0.125	0.88
` Agencia ALI ` Gye_Sur	-0.167	0.85
` Agencia ALI ` Quevedo	-0.216	0.81
` Agencia ALI ` Cuenca	-0.254	0.78
` Agencia ALI ` Sto Domingo	-0.349	0.71
` Agencia ALI ` Quito Sur	-0.364	0.70
` Estrella Validada ` 5 ESTRELLAS	-0.374	0.69
` Agencia ALI ` Quito Norte	-0.425	0.65
` Agencia ALI ` Loja	-0.433	0.65
` Agencia ALI ` Portoviejo	-0.467	0.63
` Estrella Validada ` A+	-0.598	0.55
` Agencia ALI ` Manta	-0.631	0.53
(Intercept)	-3.108	0.04
CategoriaConfiteria	-3.93E+06	Infinito
CategoriaModificadores	-3.88E+10	Infinito

Anexo 3. Coeficientes del modelo de Manicho con Galleta

Categoría	coeficiente	OR
(Intercept)	3.108	22.38
` Agencia ALI` Manta	0.631	1.88
` Estrella Validada` A+	0.598	1.82
` Agencia ALI` Portoviejo	0.467	1.60
` Agencia ALI` Loja	0.433	1.54
` Agencia ALI` Quito Norte	0.425	1.53
` Estrella Validada` 5 ESTRELLAS	0.374	1.45
` Agencia ALI` Quito Sur	0.364	1.44
` Agencia ALI` Sto Domingo	0.349	1.42
` Agencia ALI` Cuenca	0.254	1.29
` Agencia ALI` Quevedo	0.216	1.24
` Agencia ALI` Gye_Sur	0.167	1.18
` Estrella Validada` 4 ESTRELLAS	0.125	1.13
` Agencia ALI` Latacunga	0.123	1.13
` Agencia ALI` Milagro	0.077	1.08
` Estrella Validada` 3 ESTRELLAS	0.054	1.06
Cod_Solicitante	0.000	1.00
` Agencia ALI` Playas	-0.003	1.00
` Agencia ALI` Riobamba	-0.029	0.97
` Agencia ALI` Salinas	-0.041	0.96
` Agencia ALI` Machala	-0.084	0.92
` Agencia ALI` Ibarra	-0.168	0.85
` Agencia ALI` Gye_Norte	-0.346	0.71
` Agencia ALI` Chone	-0.398	0.67
Valor	-0.532	0.59
` Agencia ALI` Tulcan	-0.645	0.52
` Etiqueta Cliente` Educativo Público	-0.768	0.46
` Etiqueta Cliente` Tienda Autoservicio	-1.424	0.24

` Etiqueta Cliente `Tienda De Mostrador	-1.668	0.19
` Etiqueta Cliente ` Mini Super	-1.719	0.18
` Etiqueta Cliente `Tienda Barrotes Inte	-1.758	0.17
` Etiqueta Cliente `Tienda Barrotes Exte	-1.826	0.16
` Etiqueta Cliente `Tienda Frutas Y Verd	-1.953	0.14
` Etiqueta Cliente ` Independiente Tradic	-1.959	0.14
` Etiqueta Cliente ` Panadería	-2.419	0.09
CategoriaModificadores	-9.10E+09	0.00
CategoriaConfiteria	-3.52E+10	0.00

Anexo 4. Coeficientes del modelo de Menta Glacial

Categoría	coeficiente	OR
(Intercept)	3.103	22.26
`Agencia ALI`Chone	0.491	1.63
`Agencia ALI`Tulcan	0.436	1.55
`Agencia ALI`Quevedo	0.415	1.51
`Estrella Validada`A+	0.377	1.46
Estrella Validada`5 ESTRELLAS	0.306	1.36
`Agencia ALI`Machala	0.263	1.30
`Agencia ALI`Portoviejo	0.255	1.29
`Agencia ALI`Milagro	0.248	1.28
`Agencia ALI`Sto Domingo	0.244	1.28
`Agencia ALI`Ibarra	0.208	1.23
`Agencia ALI`Manta	0.140	1.15
Estrella Validada`4 ESTRELLAS	0.137	1.15
`Agencia ALI`Playas	0.136	1.15
Estrella Validada`3 ESTRELLAS	0.078	1.08
Cod_Solicitante	0.000	1.00
`Agencia ALI`Loja	-0.020	0.98

`Agencia ALI`Gye_Norte	-0.123	0.88
`Agencia ALI`Salinas	-0.132	0.88
`Agencia ALI`Latacunga	-0.201	0.82
Valor	-0.211	0.81
`Agencia ALI`Gye_Sur	-0.258	0.77
`Agencia ALI`Riobamba	-0.387	0.68
`Agencia ALI`Quito Norte	-0.430	0.65
`Agencia ALI`Quito Sur	-0.529	0.59
Cliente`Independiente Tradic	-0.641	0.53
`Agencia ALI`Cuenca	-0.711	0.49
`Etiqueta Cliente`Mini Super	-0.745	0.47
a Cliente`Tienda De Mostrador	-0.808	0.45
a Cliente`Tienda Autoservicio	-0.818	0.44
Cliente`Tienda Barrotes Exte	-0.850	0.43
Cliente`Tienda Barrotes Inte	-0.865	0.42
eta Cliente`Educativo Público	-0.974	0.38
Cliente`Tienda Frutas Y Verd	-1.010	0.36
`Etiqueta Cliente`Panadería	-1.211	0.30
