



**UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL**

**DEFINICIÓN DE POLÍTICAS DE MARKETING GEOREFERENCIADO
PARA UN SUPERMERCADO MAYORISTA**

**MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL
INDUSTRIAL**

GABRIELA CATALINA LAZO MÉNDEZ

PROFESOR GUÍA:
LUIS ABURTO LAFOURCADE

MIEMBROS DE LA COMISIÓN:
ALEJANDRA PUENTE CHANDÍA
RICARDO MONTOYA MOREIRA

**SANTIAGO DE CHILE
OCTUBRE 2011**

DEFINICIÓN DE POLÍTICAS DE MARKETING GEOREFERENCIADO PARA UN SUPERMERCADO MAYORISTA

El *Customer Relationship Management* permite caracterizar a los distintos consumidores de un negocio para así aplicar medidas de marketing focalizado. El marketing georeferenciado enriquece la información disponible considerando la ubicación geográfica de este consumidor para identificar comportamientos de compra asociados a esta dimensión.

El objetivo general de este trabajo, se enmarca en ambos temas. Busca modelar la relación entre la ubicación geográfica y la demanda de un cliente en un supermercado mayorista, estimando su monto y frecuencia en la sala permitiendo identificar dónde viven los mejores clientes, dónde están quienes pueden aumentar su valor y la ubicación de las áreas donde la competencia está afectando más las ventas.

El análisis se hace en 2 locales, estudiando a 3.000 clientes en la sala de Maipú y 2.000 en Temuco aproximadamente. Se trabaja con las compras realizadas entre el 2008 y 2010, para graficar en mapas la intensidad de las variables: densidad de clientes, monto y frecuencia, identificándose barrios donde se dan comportamientos de compra particulares. Posteriormente se comparan estimaciones de demanda hechas con regresiones geográficas ponderadas (GWR) y se comparan los resultados con una regresión lineal (LR).

Los resultados indican que pese a que LR y la GWR tuvieron bondades de ajuste y MAPEs muy parecidos (R^2 sobre 0.5 y MAPE bordeando incluso el 50%), la GWR identifica mejor los locales de la competencia que reducen las ventas del mayorista, identificando 8 salas contra sólo 3 que arrojó el método lineal para un local. La GWR además sugiere que la competencia tiene mayor efecto en disminuir las visitas de los clientes en otra sala, más que quitar monto de ventas, dado que la persona ya asistió.

Respecto al efecto de la ubicación en la frecuencia, se encontró que no existe correlación, habiendo entre 2 y 4 visitas mensuales en la mayoría de los casos sin importar distancia ni ubicación. El monto sí resultó estar relacionado con los barrios de residencia de los clientes, donde en vecindarios de GSE de menores ingresos se compraba hasta un 50% más que en otros.

Se propone para trabajos futuros incorporar la variable GSE y comprobar si es que ésta juega un rol fundamental en el monto comprado, así como también ampliar el estudio georeferenciado a otras dimensiones, como análisis de canastas por barrio, de modo de complementar mejor promociones personalizadas para cada cliente según el lugar donde viva.

Agradecimientos

A todos los que hicieron posible que este trabajo se realizara.

A mis papás que no entendían cómo el computador podía pasar un día entero prendido corriendo modelos.

A mis amigos con quienes pude compartir, discutir y comparar resultados, quienes además me acompañaron emocionalmente en el largo camino de la titulación. Le agradezco especialmente a todos los que estuvieron en el curso del F y se fueron de viaje conmigo a San Pedro en medio del desarrollo: Felipe, Paty y Marce. Especialmente para mi amigo Mauri que además fue mi compañero de trabajo por medio año y en quien deposité mi absoluta confianza.

A mi pololo Felipe alias furbina que me prestó su Mac para correr más rápido los modelos que se quedaban pegados en mi pc común.

A Yunny por su paciencia y siempre buena disposición para contestar mis infinitas dudas acerca del proceso de titulación y todos sus relacionados.

A mis profes del curso que me ayudaron a fortalecer mi tema, en especial a “Manreyes” que me dio la idea de ocupar modelos espaciales.

A Luis Aburto por orientarme las veces en que sentía que perdía el rumbo y por su inmensa creatividad y entusiasmo que me proyectó para continuar con mi memoria.

A todo el equipo de *data mining* de Penta Analytics, especialmente a Pedro por confiar en mi y darme la posibilidad de desarrollar el tema de memoria que yo quise.

Finally thanks a lot to myheatmaps¹ team, the software that I used to understand and visualize customer’s behavior. Especially to Jay McAliley for explain me in detail everything about the heat maps creation, besides developing for me some ideas about how to visualize the data in other creative ways.

¹ <http://www.myheatmap.com/>

Índice

1.	Introducción	1
2.	Justificación del proyecto	4
3.	Descripción del proyecto	6
4.	Objetivos	6
4.1	Objetivo general	6
4.2	Objetivos específicos.....	6
5.	Resultados esperados	7
6.	Alcances	7
7.	Marco Conceptual	8
7.1	Geomarketing.....	8
7.1.1	<i>Heat maps</i>	9
7.2	<i>Spatial analysis</i>	10
7.2.1	Modelos de predicción	11
8.	Metodología	16
9.	Desarrollo metodológico	18
9.1	Maipú.....	18
9.1.1	Datos	18
9.1.2	Análisis de la situación actual	21
9.1.3	Modelos de estimación para el monto y la frecuencia	30
9.1.4	Barrios de interés y ubicación de la publicidad	48
9.2	Temuco	51
9.2.1	Datos	51
9.2.2	Análisis de la situación actual	53
9.2.3	Modelos de estimación para el monto y la frecuencia	60
9.2.4	Barrios de interés y ubicación de la publicidad	67
9.3	Discusión y comparación resultados local Maipú y Temuco	70
10	Conclusiones	72
11	Bibliografía	74
12	Anexo.....	76
Anexo A:	Comunas de Santiago	76
Anexo B:	Frecuencia y ticket promedio	77
Anexo C:	Scatter plots de las compras.....	79

Anexo D: Errores.....	81
Anexo E: Gasto por barrio	84

Índice de gráficos

Gráfico 1: Variación de las ventas acumuladas de los supermercados	1
Gráfico 2: Facturación anual de supermercados	1
Gráfico 3: Participación de mercado de supermercados mayoristas.....	2
Gráfico 4: Medios que utilizan los clientes para informarse de las promociones.....	5
Gráfico 5: Histograma de distancias de los clientes en local Maipú	20
Gráfico 6: <i>Box plot</i> de distancias de los clientes del local Maipú.....	20
Gráfico 7: Origen de los clientes del local Maipú.....	22
Gráfico 8: Días entre compras de los clientes del local Maipú	23
Gráfico 9: Ticket promedio de los clientes de Maipú, acercamiento.....	25
Gráfico 10: Monto y distancia en <i>scatter plot</i> , local Maipú.....	32
Gráfico 11: Monto y distancia en <i>box plot</i> , local Maipú	33
Gráfico 12: Monto y días entre compras, local Maipú	34
Gráfico 13: Monto y ticket promedio, local Maipú.....	34
Gráfico 14: Días entre compras y distancia, local Maipú.....	35
Gráfico 15: Ajuste de la regresión lineal para el monto, local Maipú	38
Gráfico 16: Ajuste de la regresión lineal para los días entre compras, local Maipú.....	40
Gráfico 17: Ajuste de la GWR para el monto, local Maipú.....	43
Gráfico 18: Ajuste de la GWR para los días entre compras, local Maipú	45
Gráfico 19: Zoom del ajuste de la GWR para los días entre compras, local Maipú.....	46
Gráfico 20: Ranking de competencia significativa para el monto, local Maipú	47
Gráfico 21: Ranking de competencia significativa para los días entre compras, local Maipú.....	47
Gráfico 22: Histograma de las distancias de los clientes, local Temuco	52
Gráfico 23: <i>Box plot</i> de las distancias de los clientes, local Temuco.....	52
Gráfico 24: Origen de los clientes del local Temuco	54
Gráfico 25: Días entre compras de los clientes del local Temuco	55
Gráfico 26: Ticket promedio de los clientes del local Temuco	56
Gráfico 27: Monto y distancia en <i>scatter plot</i> , local Temuco	61
Gráfico 28: Monto y distancia en <i>box plot</i> , local Temuco	62
Gráfico 29: Monto y días entre compras en <i>scatter plot</i> , local Temuco	62
Gráfico 30: Monto y ticket promedio en <i>scatter plot</i> , local Temuco	63
Gráfico 31: Ajuste regresión lineal para el monto, local Temuco.....	64
Gráfico 32: Ajuste de la GWR para el monto, local Temuco	66
Gráfico 33: Ranking de competencia significativa, local Temuco	66
Gráfico 34: Días entre compras de clientes local Maipú.....	77
Gráfico 35: Ticket promedio de clientes local Maipú	77
Gráfico 36: Días entre compras de los clientes, local Temuco.....	78
Gráfico 37: Ticket promedio de los clientes, local Temuco.....	78
Gráfico 38: <i>Scatter plot</i> de montos de clientes, local Maipú (en MM\$).....	79
Gráfico 39: <i>Scatter plot</i> de montos de clientes, local Temuco (en MM\$).....	79
Gráfico 40: <i>Scatter plot</i> para días entre compras, local Temuco.....	80
Gráfico 41: Histograma del MAPE en la estimación del monto con LR, local Maipú	81
Gráfico 42: Histograma del MAPE en la estimación de los días entre compras con LR, local Maipú	81
Gráfico 43: Histograma del MAPE en la estimación del monto con GWR, local Maipú..	82

Gráfico 44: Histograma del MAPE en la estimación de los días entre compras con GWR, local Maipú	82
Gráfico 45: Histograma del MAPE para la estimación del monto con GWR, local Temuco	83
Gráfico 46: Distribución del gasto según <i>cluster</i> de barrio, local Maipú	84
Gráfico 47: Distribución del gasto según <i>cluster</i> de barrio, local Temuco	84

Índice de tablas

Tabla 1: Comparación entre análisis espacial global y local	11
Tabla 2: Datos clientes de Maipú	19
Tabla 3: Regresión lineal para el monto, local Maipú	37
Tabla 4: Regresión lineal con <i>recency</i> para días entres compras, local Maipú	39
Tabla 5: Regresión lineal sin <i>recency</i> para los días entre compras, local Maipú	39
Tabla 6: Regresión geográfica ponderada para el monto, local Maipú	42
Tabla 7: Regresión geográfica ponderada para los días entre compras, local Maipú	44
Tabla 8: Regresión geográfica ponderada para los días entre compras, sin <i>Recency</i> , local Maipú	44
Tabla 9: Datos clientes de Temuco	51
Tabla 10: Regresión lineal para el monto, local Temuco	63
Tabla 11: Regresión geográfica ponderada para el monto, local Temuco	65

Índice de figuras

Figura 1: Características de una manzana censal en Santiago	3
Figura 2: Cobertura de restaurantes McDonald's en Estados Unidos	3
Figura 3: Elementos constituyentes de un sistema de Geomarketing	8
Figura 4: <i>Heat map</i> de la altura de la ola del tsunami en Japón.....	10
Figura 5: Análisis espacial estacionario y no estacionario.....	11
Figura 6: Malla de regresión de ventanas móviles	12
Figura 7: Función w de ponderación	13
Figura 8: <i>Fixed Spatial Kernel</i>	14
Figura 9: <i>Adaptive Spatial Kernel</i>	14
Figura 10: Porcentaje del monto de las compras explicado por la distancia, local Maipú	21
Figura 11: <i>Heat map</i> del valor en local Maipú, comunas Maipú y Cerrillos	26
Figura 12: <i>Heat map</i> del valor en local Maipú, comunas Padre Hurtado y Peñaflores	26
Figura 13: <i>Heat map</i> de los días entre compras en local Maipú.....	28
Figura 14: <i>Heat map</i> de la densidad de clientes en local Maipú, comunas Maipú y Cerrillos	28
Figura 15: <i>Heat map</i> de la densidad de clientes en local Maipú, comunas Padre Hurtado y Peñaflores	29
Figura 16: Competencia mayorista de local Maipú.....	30
Figura 17: Competencia significativa para el local Maipú.....	48
Figura 18: Barrios de interés para el local Maipú	49
Figura 19: Porcentaje del monto de las compras explicado por la distancia, local Temuco	53
Figura 20: <i>Heat map</i> del valor, local Temuco	57
Figura 21: <i>Heat map</i> de los días entre compras, local Temuco	57
Figura 22: <i>Heat map</i> de la densidad de clientes, local Temuco	58
Figura 23: Competencia mayorista de local Temuco	59
Figura 24: Acercamiento competencia para el local Temuco	59
Figura 25: Competencia significativa para el local Temuco.....	67
Figura 26: Áreas de competencia fuerte para el local Temuco	68
Figura 27: Áreas con clientes que compran más en local Temuco	69
Figura 28: Áreas con clientes que compran menos en local Temuco	70
Figura 29: Mapas con las comunas de Santiago.....	76

Índice de ecuaciones

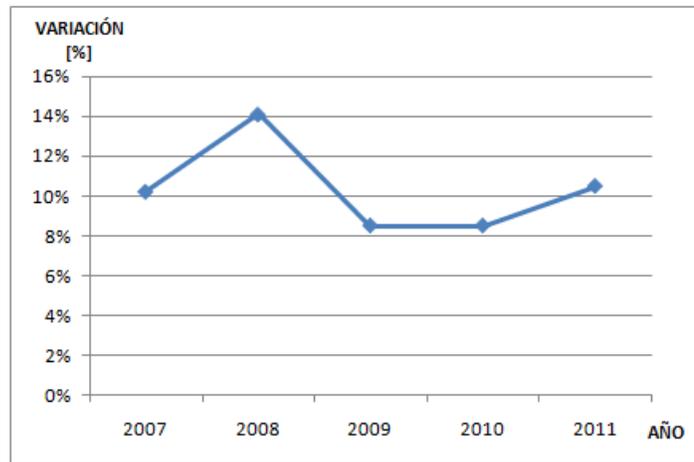
Ecuación 1: Función Gaussiana.....	13
Ecuación 2: Residuos de una regresión.....	15
Ecuación 3: Residuos de una regresión, usando el promedio como estimador	15
Ecuación 4: Coeficiente de determinación	15
Ecuación 5: Coeficiente de determinación ajustado.....	15
Ecuación 6: Error porcentual absoluto promedio (MAPE)	16
Ecuación 7: Modelo de estimación del monto para un cliente i, local Maipú.....	36
Ecuación 8: Modelo de estimación para los días entre compras para un cliente i, local Maipú.....	38
Ecuación 9: Modelo de estimación para el monto para un cliente i, local Temuco.....	60

1. Introducción

El retail, es una de las actividades más importantes dentro de la economía chilena, representa un 20% del PIB del país [2], lo que equivale a ventas por más de US\$48.000 millones.

Una participación importante dentro de este grupo, la tienen los supermercados, quienes perciben ingresos por más de US\$11.000 al año y que han tenido un aumento sostenido de sus ventas durante los últimos años, acumulando un 10.5% de crecimiento a finales de julio de 2011.

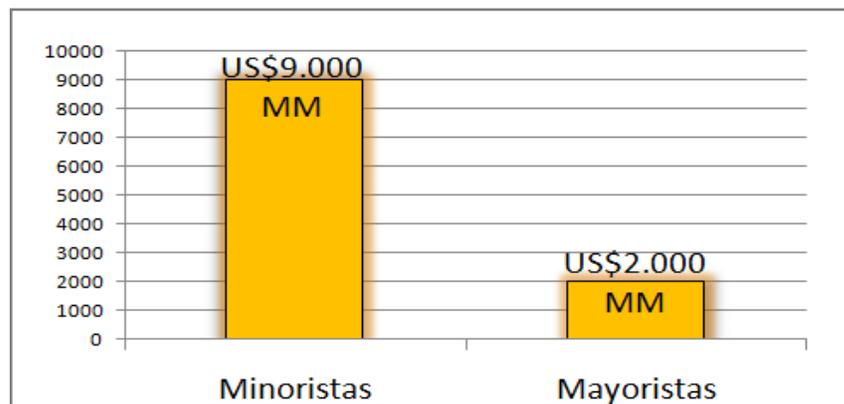
Gráfico 1: Variación de las ventas acumuladas de los supermercados



Fuente: Elaboración propia. Datos: INE [14]

Un tipo de formato dentro de los supermercados lo constituyen los mayoristas, que fundamentalmente van dirigidos a los almaceneros. Este sector vende más de US\$2.000 millones, equivalente a más de 1/5 de lo que venden los supermercados minoristas.

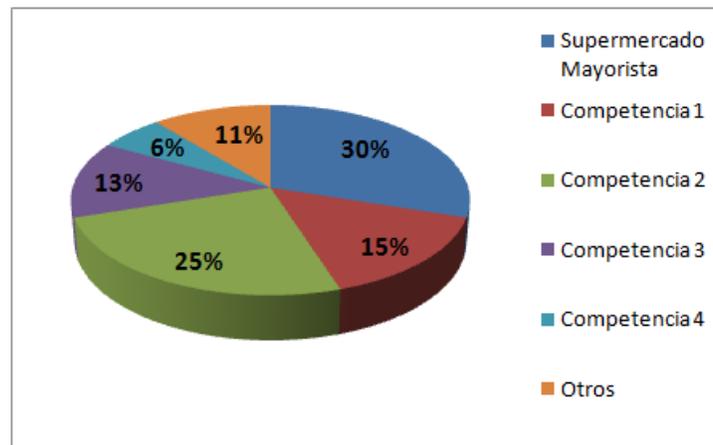
Gráfico 2: Facturación anual de supermercados



Fuente: Elaboración propia. Datos Diario La Estrategia, 2009 [17]

Los supermercados mayoristas pertenecen a una industria de alta competencia, donde cada cadena pone todo su esfuerzo para concentrar la mayor cantidad de ingresos. Pese a que el mayorista de este proyecto concentra la preferencia de los clientes, existen otros 3 actores que tienen más de un 10% de la participación de mercado.

Gráfico 3: Participación de mercado de supermercados mayoristas



Fuente: Elaboración propia. Datos Diario La Estrategia, 2009 [17]

Al ser un sector de gran valor y fuerte competencia, se hace indispensable entender el comportamiento de los clientes, con el fin de asegurar las ventas y fidelizar las compras a través del tiempo.

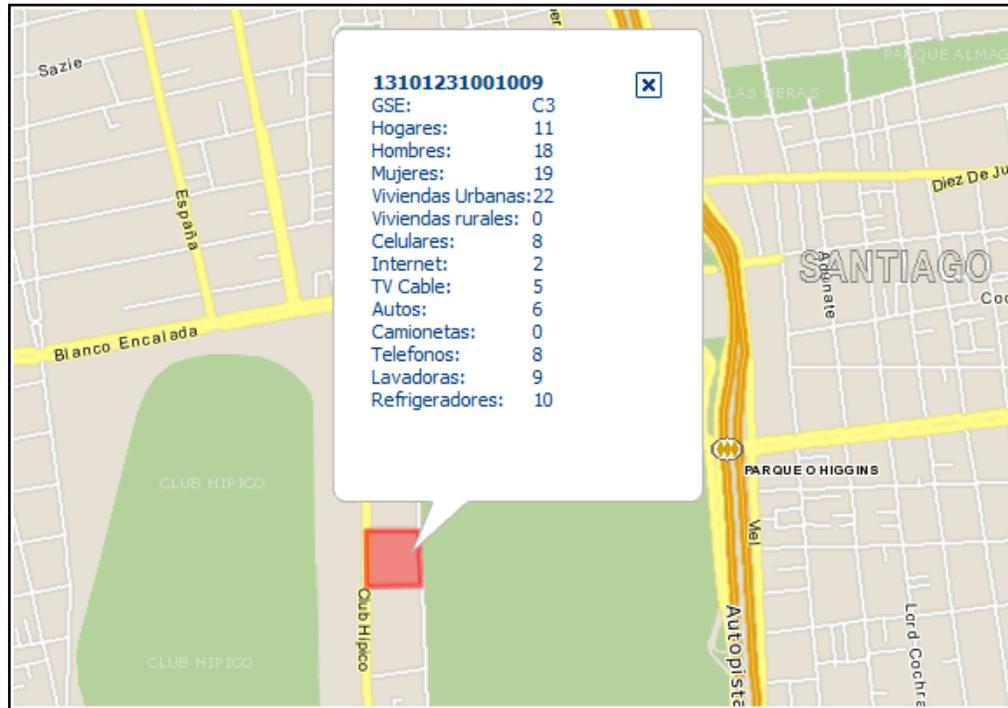
Una de las variables que explican el comportamiento de una persona es la geográfica, a partir de la ubicación, se tiene acceso a una serie de singularidades que afectan la conducta, tales como el clima o la cultura, o simplemente su barrio de residencia, permitiéndonos conocer qué idioma habla una persona, cuál es su consumo energético o cuál es su poder adquisitivo, entre otras.

El geomarketing, es una rama del marketing que utiliza esta variable espacial para la toma de decisiones de negocios. A través de mapas, se puede representar el panorama general en cuanto a clientes, competencia, puntos de distribución, tiendas propias, etc. permitiendo:

- Estudiar localizaciones óptimas para nuevas sucursales.
- Obtener conocimiento de los clientes, a partir de las características de su lugar de residencia.
- Identificar dónde se encuentran ubicados los clientes más y menos valiosos.
- Reconocer áreas de mayor competencia.
- Encontrar espacios óptimos para emplazar la publicidad.
- Optimizar la ubicación de las partes de la cadena de suministro.

Por ejemplo, a nivel de manzana censal se puede llegar a conocer desde el grupo socioeconómico de los habitantes de esas viviendas, hasta cuántos celulares se usan o cuántos autos se tienen.

Figura 1: Características de una manzana censal en Santiago



Fuente: MapCity, 2011 [15]

O también se podrían percibir las áreas de mayor influencia de las diferentes tiendas de una cadena de restaurantes en cada zona geográfica. Tal como aparece en la figura 2, las zonas más iluminadas representan áreas donde las personas tienen mejor alcance a los locales de McDonald's, mientras que los terrenos más tenues simbolizan sectores donde la franquicia tiene menos dominio.

Figura 2: Cobertura de restaurantes McDonald's en Estados Unidos



Fuente: Data Pointed, 2009 [8]

Este trabajo de titulación se enmarca en el área del marketing georeferenciado, el objetivo es modelar la relación que existiría entre la demanda de un cliente de un supermercado mayorista y su ubicación geográfica, esperando encontrar zonas donde los vecinos tengan comportamientos de compra similares.

Además se espera conocer cuáles y dónde están los locales de la competencia que son más importantes, como también descubrir oportunidades de crecimiento en ventas en sectores geográficos específicos.

2. Justificación del proyecto

El CRM o *Customer Relationship Management* se ha hecho cargo de estudiar a las personas, de modo de identificar cuáles son sus necesidades e incluso sus anhelos y aspiraciones, para así poder ofrecerle un producto que responda y se adapte 100% a sus deseos. Este conocimiento y gestión de clientes incluye variables socio-demográficas como el sexo, la edad o el ingreso de la persona, transaccionales como cuánto gasta en cada visita a una tienda, u otras más específicas relacionadas con estilos de vida, como saber si esa persona es amante de los deportes, o preocupada de comer sano o de cuidar a su familia. Todas estas variables enriquecen el conocimiento del cliente y permiten averiguar qué es lo que necesitan comprar, cuánto y cómo.

Sin embargo, ¿qué entendimiento se tiene acerca de cómo es la demanda de una persona dado sus variables geográficas?, ¿será que las personas que están más cerca de una tienda compran más?, ¿o las que están más lejos disminuyen su frecuencia?, o incluso, ¿existirán ciertas ubicaciones geográficas que puedan favorecer ciertos comportamientos de compra?

Este proyecto incorpora la variable espacial al conocimiento de los clientes. Esto permitirá enriquecer la información disponible, permitiendo identificar claramente cómo se comportan las personas de cada barrio y como van cambiando sus variables de compra como la frecuencia o el monto adquirido dependiendo si se ubican al norte o sur del supermercado, si están a más de 1 km. de distancia de la tienda o incluso como respondería un cliente si está ubicado a menos de una cuadra de un local de la competencia.

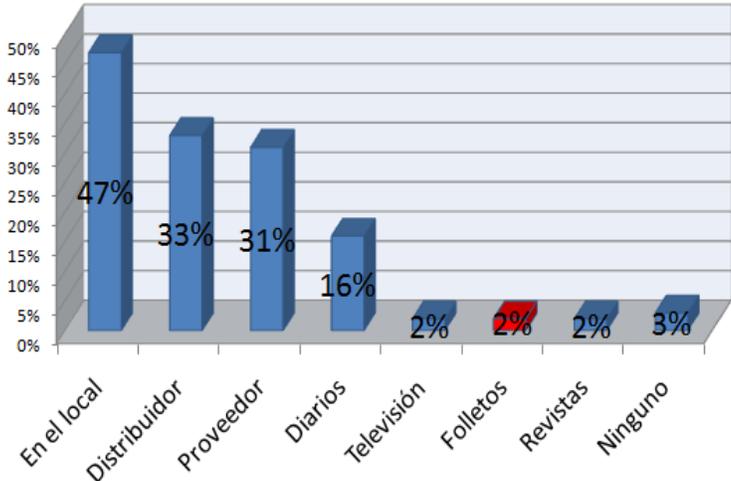
Con el complemento de la geografía al conocimiento de clientes, se podría conocer cómo se distribuyen geográficamente los buenos y los malos compradores, dónde hay barrios con concentración de clientes de iguales características y más importante aun dónde viven las personas susceptibles a aumentar sus ventas dado su lugar de residencia.

También se podrían reconocer las áreas de influencia de cada uno de los locales y responder: ¿hasta qué distancia máxima se hacen efectivas las visitas a un supermercado? Incluso se podría conocer con exactitud qué efecto causa cada uno de los locales de la competencia sobre las ventas de un supermercado, permitiendo fichar a las marcas rivales más fuertes en quitar ingresos a cierta sala.

Todo este conocimiento adicional ayuda a delimitar pautas para apoyar en la toma de decisiones donde esta variable geográfica es su eje fundamental. Ejemplo de esto es la publicidad, a través del conocimiento de las áreas con clientes susceptibles a aumentar sus compras, identificación de sectores donde la competencia es más agresiva o simplemente conocer dónde está la mayoría de las personas que asisten a cierto local, se puede reconocer dónde convendría publicitar o dónde se podrían captar con promociones a clientes que en un futuro podrían aumentar su valor.

A modo de representación del valor de esta información, en el caso de la publicidad, suponemos que si el supermercado mayorista gastara sólo un 1% de sus ventas en hacer difusión e invirtiera un 2% en hacer revistas o volantes informativos², dado que hoy sólo un 2% de las personas compraron dado que recibieron la publicidad, se está perdiendo un 98% de revistas que es el equivalente a \$40MM aproximadamente.

Gráfico 4: Medios que utilizan los clientes para informarse de las promociones



Fuente: Elaboración propia. Datos Penta Analytics, 2009.

Esto ocurre porque el supermercado no sabe dónde están sus clientes, ni menos aun conoce dónde están los buenos, los malos y los influenciables. Por lo tanto, este estudio entregará esta información de modo de poder distribuir mejor los esfuerzos de publicidad, así como también aporta al conocimiento de clientes, permitiendo conocer donde existen personas con potencial de comprar más, dado su ubicación geográfica o distancia a la tienda, y donde están los locales de la competencia que más le quitan clientes al mayorista, para así poder aumentar las ventas del supermercado.

²Distribución según datos ACHAP 2009.

3. Descripción del proyecto

Se desea conocer el comportamiento de compra de los clientes, a través de variables espaciales como la ubicación y la distancia a un local, además de otras variables de tipo transaccional, con el fin de percibir si la ubicación geográfica explica el monto y la frecuencia de un cliente en un supermercado.

Se busca entender la situación actual y estimar el monto y frecuencia de compra a través de modelos de regresión. Se utilizarán de por medio herramientas del marketing georeferenciado, como regresiones geográficas ponderadas para las estimaciones, además de mapas especializados para la visualización de la situación actual.

Con la información entregada por el análisis de la situación actual, en cuanto a los clientes y su geografía, además de la localización de la competencia, junto con los resultados arrojados por los modelos de regresión, se identificarán cuáles son los barrios con buenos y malos clientes, dónde hay villas con oportunidad de crecimiento de ventas, cuáles son los locales de la competencia que más afectan al mayorista y basado en esto, donde se podrían enfocar los esfuerzos de publicidad estática.

4. Objetivos

4.1 Objetivo general

Modelar la relación que existe entre la ubicación y la demanda de los clientes de un supermercado mayorista estimando frecuencia y monto de la compra para apoyar decisiones de marketing georeferenciado.

4.2 Objetivos específicos

- Visualizar geográficamente la situación actual de los locales del supermercado, graficando mediante mapas, a los clientes y su demanda, además de los supermercados de la competencia.
- Reconocer las variables que puedan explicar la demanda de los clientes, además de la ubicación y los datos contenidos en los tickets de compra.
- Realizar modelos que expliquen la demanda en función de la ubicación y las demás variables. Comparar los resultados de un modelo de regresión geográfica ponderada con uno de regresión lineal.
- Determinar cuáles son los locales de la competencia que afectan las ventas del mayorista.
- Identificar zonas geográficas donde residan clientes con oportunidad de crecimiento en ventas.

- Distinguir puntos favorables donde emplazar la publicidad estática del supermercado.

5. Resultados esperados

- *Heat maps* para el monto, la frecuencia y la densidad de clientes, para visualizar los diferentes comportamientos de compra entre barrios. Además de mapas que muestren donde estén los competidores del supermercado.
- Listado de variables que afecten la demanda en un supermercado, conociendo la importancia y efecto en la compra de cada una de ellas.
- Valor estimado de la compra para cada cliente, tanto del monto como la frecuencia, a través de modelos de regresión.
- Nombres de los locales de la competencia significativa para el mayorista.
- Barrios con concentración de buenos y malos clientes, definido según su monto y/o frecuencia de compra, donde existan clientes que puedan aumentar su valor.
- Distinción de áreas donde ubicar la publicidad, dado el conocimiento de los clientes y su variable geográfica.

6. Alcances

El estudio se acotará a la sucursal ubicada en Maipú. Además se aplicará la metodología a otra sede en Temuco, con el fin de validar las posibles tendencias de compra basadas en la georeferenciación, sin embargo no se harán propuestas de publicidad para esta última sede.

Se definirán los sectores geográficos más importantes, tanto porque tienen clientes que compran mucho o muy poco, proponiendo estas mismas localizaciones como ubicaciones propicias donde realizar la publicidad. No se medirá el efecto que tendría ésta en las futuras compras.

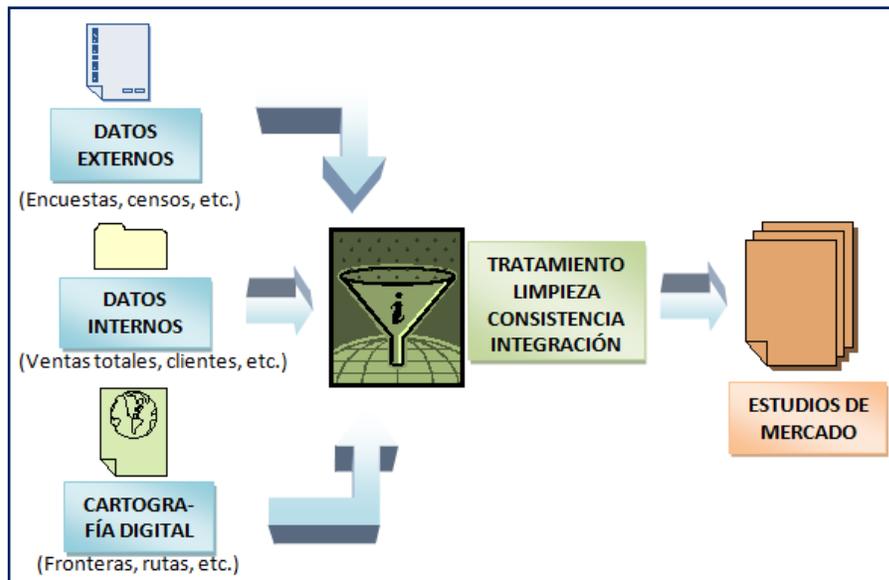
7. Marco Conceptual

7.1 Geomarketing

El geo-marketing o marketing georeferenciado es una ciencia relativamente nueva, que nace de la confluencia de dos ramas: la geografía y el marketing. Es un sistema integrado por datos, programas informáticos, estadística y representaciones gráficas, que permiten analizar la realidad económica-social desde un punto de vista geográfico, ayudando en la toma de decisiones de negocios y poniendo al descubierto cuáles son las localizaciones clave de las partes de un comercio, además de contestar la pregunta: “quién compra donde” [5].

Pese a la dependencia que existe entre esta variable geográfica con el marketing, no es hasta los años 90 que esta rama empieza a ser más desarrollada, debido al mejor acceso que se tiene a tecnologías como los GIS³ (Sistemas de Información Geográfica), GPS (Sistema de Posicionamiento Global) o imágenes satelitales, además de la información geo-referenciada que se podría extraer de los censos o encuestas. No obstante, estos no son los únicos elementos constitutivos de un sistema de geomarketing. A continuación se enuncian en la figura 3 y en seguida se explican los datos que lo forman.

Figura 3: Elementos constituyentes de un sistema de Geomarketing



Fuente: Elaboración propia

³ Los GSI o Sistemas de Información Geográfica, son sistemas de software que se usan para almacenar, recuperar, plotear y analizar la información geográfica. Los lugares son almacenados en forma de coordenadas como latitud/longitud u otra, y a través de etiquetas o colores se representan en un mapa las características que se quieran visualizar, ya sea ventas de las sucursales de una tienda, densidad o ingreso de una población, profundidad de un lago, etc.[10].

Datos externos: vienen de otras empresas y difunden bases de datos con contenido de tipo socio-económico. Un ejemplo de este tipo de institución sería el INE: Instituto Nacional de Estadísticas, de Chile.

Datos internos: son con los que cuenta la propia empresa, en las bases de datos internas se encuentra toda la información del cliente, las compras, los locales, etc.

Cartografía digital: mapas que permiten visualizar y tratar estadísticamente los datos. Necesita de por medio un software eficaz de GSI que trabaje los datos recopilados de fuentes internas y externas, para mostrarlos de una manera eficiente y explicativa.

Con los 3 elementos anteriores se procede al tratamiento de toda la información. Las relaciones que existirían entre los datos y la geografía, son detectadas a través de un AEDE (Análisis Exploratorio de Datos Espaciales), que cuando es aplicado a grandes volúmenes de información es conocido como *spatial data mining*.

Finalmente, con todo el análisis realizado, se puede llegar a un mapa que contenga información relevante para el negocio, convirtiéndose éste en una poderosa herramienta de estudio de mercado, que permitirá determinar donde conviene ubicar una tienda nueva, visualizar donde están los clientes más valiosos, entender por qué un local está bajando sus ventas, etc.

7.1.1 Heat maps

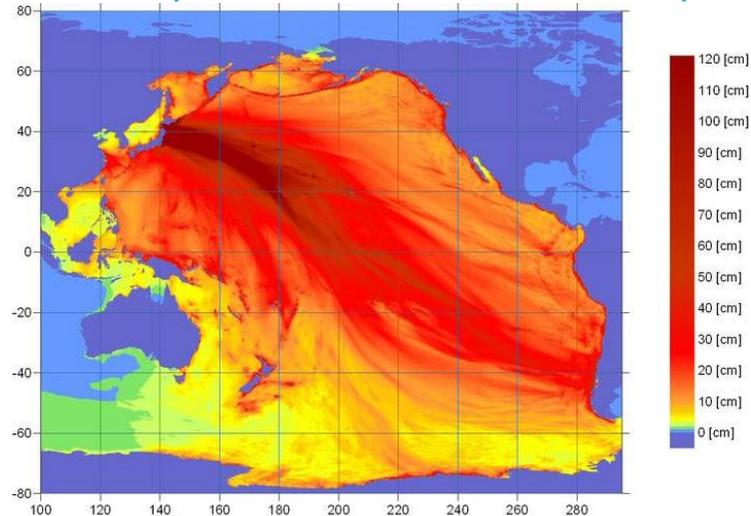
Dentro de los planos posibles de obtener para el análisis de la información, se encuentran los *heat maps* o mapas de calor. Estos son mapas donde la información es mostrada mediante una escala de colores. Ejemplo de esto, son los mapas que muestran las profundidades de un océano, donde los colores azules más oscuros representan un nivel más sumergido y los colores celestes los niveles más superficiales de las aguas.

Además de sus usos en geografía, es posible aplicar esta herramienta en materias de marketing. Por ejemplo, se podrían graficar las compras que hacen los clientes en una tienda, la competencia que presenta un negocio, las áreas con mayor cantidad de consumidores o el nivel de visitas en las sucursales. Todo esto sería apreciado a través de las áreas de calor del mapa: los colores más “calientes” representan zonas de mayor flujo, ya sea mayor concentración de clientes, más locales de la competencia u otro, mientras que las zonas más “frías” demuestran una ausencia o un valor muy bajo de la variable que se quiere graficar, por ejemplo lugares donde los clientes van poco a un supermercado, o áreas donde la competencia sea escasa.

Muchos software GIS poseen herramientas para poder generar este tipo de mapas. La entrada de datos requerida son las coordenadas geográficas de los puntos a graficar (latitud y longitud por ejemplo) y una tercera variable que se quiera medir, por ejemplo GSE de una persona.

A modo de ejemplo, se observa a continuación un *heat map* que midió la altura de la ola provocada por el tsunami en Japón. Las zonas de color rojo más oscuro son donde la ola alcanzó la mayor altura, mientras que las zonas más frías de color celeste representan el nivel más bajo del mar.

Figura 4: *Heat map* de la altura de la ola del tsunami en Japón



Fuente: Documenting reality, 2011 [9]

7.2 Spatial analysis

El marketing georeferenciado se encuentra inserto en lo que se conoce como *spatial analysis*. El análisis espacial, consiste en usar las características geográficas, topográficas o geométricas de los objetos para realizar cualquier tipo de investigación. Estos estudios se desarrollan en distintas áreas de la ciencia, como la botánica, la demografía, la epidemiología, la climatología, etc.

Un caso muy ejemplar de los diferentes análisis espaciales existentes, lo cita Fotheringham [11]. Si se quisiera analizar el clima de Estados Unidos, por un lado se pueden consultar los valores promedio de las lluvias, las temperaturas, la humedad, etc. Sin embargo, esta mirada no nos permite conocer acerca de los climas locales de cada estado, no se puede saber la diferencia del clima que existe entre el estado de Florida y Washington, por ejemplo. En cambio, se puede hacer un análisis estado por estado, que nos daría acceso a información mucho más valiosa y exacta que la primera.

La diferencia entre estas dos miradas es lo que se conoce como análisis global y análisis espacial local, mientras que el primero establece medidas homogéneas para todas las partes geográficas, el segundo es más apegado a la realidad local de cada topografía. Esto también está muy relacionado con el entendimiento de los procesos estacionarios y no estacionarios. El análisis global podría ser un tipo de análisis espacial estacionario, es decir la distribución de probabilidad conjunta es constante en todo el espacio. En cuanto al análisis espacial no estacionario, esta probabilidad va variando de acuerdo a las características geográficas locales. Mirar detalle de la tabla 1

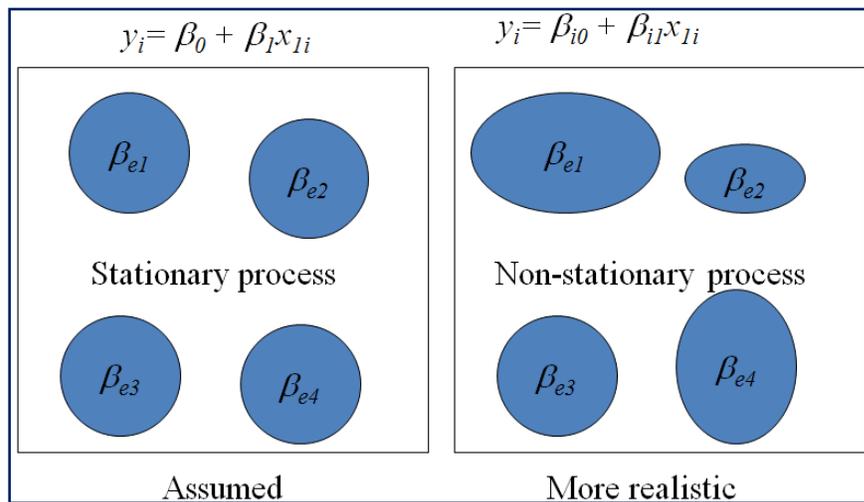
con las diferencias entre estos análisis y la figura 5 con una idea gráfica acerca de la distinción entre análisis espacial estacionario v/s no estacionario.

Tabla 1: Comparación entre análisis espacial global y local

<i>Global</i>	<i>Local</i>
Resume la data para una región entera	Desagrega las estadísticas globales en locales
Valores estadísticos únicos	Múltiples valores estadísticos
Limitado espacialmente	Espacial
Enfatiza las similitudes a través del espacio	Enfatiza las diferencias a través del espacio
Busca patrones regulares, reglas comunes	Busca las excepciones locales
Ejemplo: Regresiones	Ejemplo: Regresión Geográfica Ponderada

Fuente: Elaboración propia

Figura 5: Análisis espacial estacionario y no estacionario



Fuente: Dept. of Geog., UWM [24]

7.2.1 Modelos de predicción

Dentro del análisis espacial existen modelos que permiten hacer estimaciones de cualquier variable que pueda ser modelada a través de la geografía, ejemplo de esto son las regresiones geográficas ponderadas.

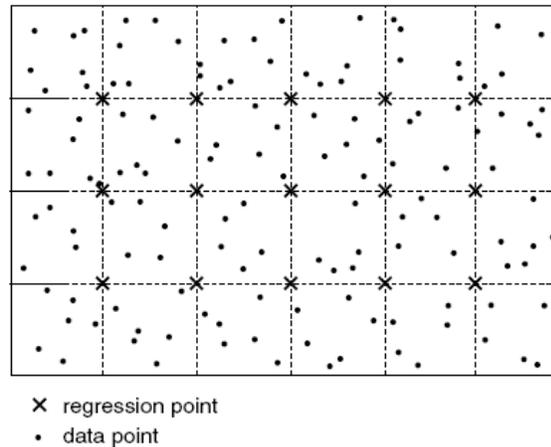
7.2.1.1 Regresión geográfica ponderada (GWR)

Dentro del análisis espacial local, se encuentra el modelo de regresión geográfica ponderada, éste trabaja sobre la base de una regresión normal, pero incorporando el carácter local de la data, considerando que existen correlaciones entre los datos de ciertas localidades (haciendo que los errores de esta regresión

también estén correlacionados). Este modelo es una adaptación de la regresión de ventanas móviles, pero la GWR incluye el concepto de vecino cercano o lejano desde el punto de regresión.

En la Regresión de Ventanas Móviles o *Moving Window Regression* (MWR), se trabaja con una malla con datos, donde cada espacio es un área de regresión, es decir es un conjunto de regresiones locales, donde cada una de ellas es calibrada con los datos que están dentro de su zona.

Figura 6: Malla de regresión de ventanas móviles



Fuente: Fotheringham "et al", "Geographically Weighted Regression" [11]

Cada una de estas zonas puede ser ajustada a cualquier forma: cuadrada, circular, hexagonal, etc., siendo la más común la cuadrada, como la que se muestra en la figura. El tamaño de cada ventana puede ser libremente ajustado, haciendo que matrices con ventanas más anchas den forma a superficies mucho más suaves en los *heat maps* y las más pequeñas cambios en la superficie mucho más bruscos. Este ancho de ventana es conocido como *bandwidth*⁴ y es uno de los parámetros más importantes dentro de la regresión.

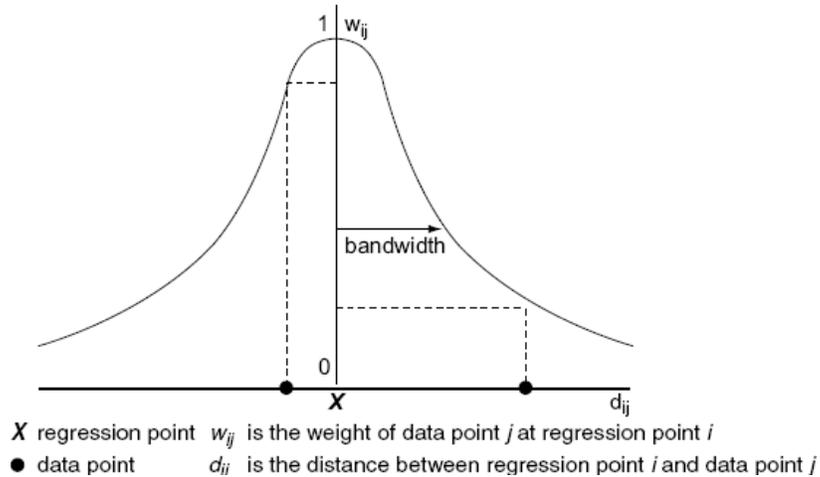
La Regresión Geográfica Ponderada funciona como una *Moving Window Regression*⁵, pero esta vez, los puntos que quedan dentro de cada área de regresión son ponderados por su distancia al punto de regresión, siendo máximo el peso cuando el dato coincide con la localización del punto de regresión, igual a cero si queda fuera de la ventana y entre 1 y 0 dentro de la ventana.

La función de ponderación más común es la Gaussiana. A continuación se muestra como disminuye la ponderación, conforme va aumentando la distancia del dato con el punto de regresión:

⁴ Ancho de banda.

⁵ Regresión de ventanas móviles.

Figura 7: Función w de ponderación



Fuente: Fotheringham "et al", "Geographically Weighted Regression" [11]

En la campana de Gauss esta función de ponderación w queda definida como:

Ecuación 1: Función Gaussiana

$$w_{ij} = \exp \left(-\frac{1}{2} \left(\frac{d_{ij}}{b} \right)^2 \right)$$

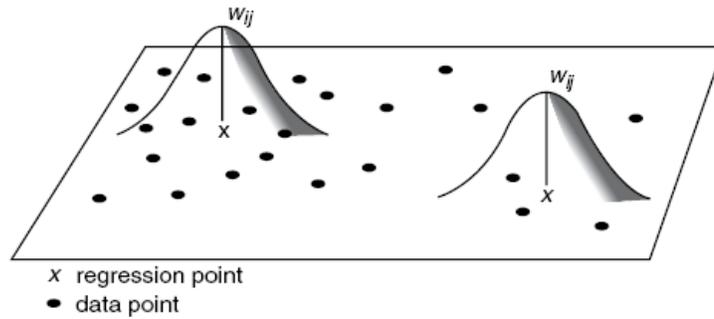
Donde:

d_{ij} : distancia entre el punto de regresión i con el dato j

b : ancho de banda

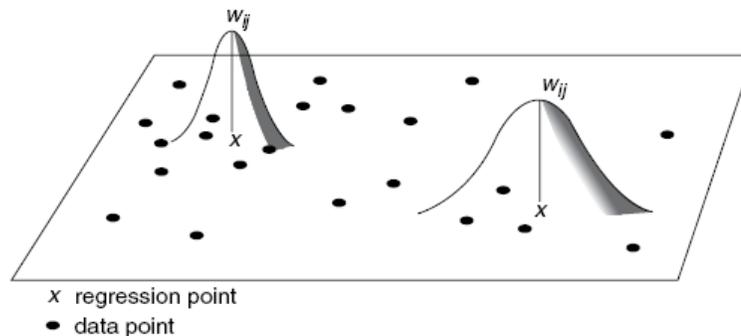
Cabe destacar que esta función de ponderación, junto con su *bandwidth*, puede ser igual para todas las ventanas, lo que se conoce como *Fixed Spatial Kernel* o diferente, "*Adaptive Spatial Kernel*". El segundo método tiene por objetivo evitar que para zonas poco densas de clientes, se hagan calibraciones con muy poco datos, por lo que en vez de establecer barrios del mismo tamaño (iguales *bandwidth*) se establezcan porcentajes de vecinos más cercanos con quienes calibrar. Por ejemplo, un *Adaptive Spatial Kernel* de 0.4 significa que cada ventana trabajará con el 40% de los clientes que le queden más cercanos.

Figura 8: Fixed Spatial Kernel



Fuente: Fotheringham "et al", "Geographically Weighted Regression" [11]

Figura 9: Adaptive Spatial Kernel



Fuente: Fotheringham "et al", "Geographically Weighted Regression" [11]

7.2.1.2 Regresión lineal (LR)

Es un método matemático que explica la relación que existe entre una variable dependiente y con una variable independiente x , o en el caso de una regresión lineal múltiple, dos o más variables independientes x_i . Tiene diferentes aplicaciones. Se ocupa tanto en materias de negocios como en medicina. Por ejemplo el modelo de CAPM (*Capital Asset Pricing Model*) utiliza de la regresión lineal para predecir el retorno de una acción en un portafolio diversificado, a partir del conocimiento de su riesgo. En medicina existen diferentes modelos lineales que explicarían la relación del tabaquismo con el cáncer de pulmón.

Una de las maneras más comunes de ajustar una regresión lineal, es a través de mínimos cuadrados ordinarios (OLS), ésta consiste en minimizar la suma de los residuos al cuadrado, entendiéndose como residuo a la diferencia del valor real de y con su estimación.

Esto queda definido como:

Ecuación 2: Residuos de una regresión

$$SS_{ERROR} = \sum_i (y_i - \hat{y})$$

Por otro lado, si también se considera que el promedio de los datos pueden ser una buena estimación para una variable y , puede definirse el error de ésta como:

Ecuación 3: Residuos de una regresión, usando el promedio como estimador

$$SS_{TOT} = \sum_i (y_i - \bar{y})$$

Evaluación

Con las dos sumatorias anteriores se construye una medida de bondad para la regresión. Este es conocido como R^2 o coeficiente de determinación, que mide el porcentaje de la varianza en y que es explicada por el modelo regresivo. Este coeficiente toma valores entre 0 y 1, donde 0 representa el peor ajuste del modelo a los datos y 1 el mejor.

Ecuación 4: Coeficiente de determinación

$$R^2 = 1 - \frac{SS_{ERR}}{SS_{TOT}}$$

Es importante tener en cuenta que cuando se ocupa el método de mínimos cuadrados ordinarios para ajustar la regresión, el valor de R^2 aumenta conforme lo hace el número de variables independientes en el modelo. En este caso, si se quiere medir la bondad del modelo al agregar más variables, lo correcto es observar el coeficiente de determinación ajustado, ya que este \bar{R}^2 sólo crece si la variable agregada aumenta el valor explicativo del modelo.

El coeficiente ajustado se calcula como:

Ecuación 5: Coeficiente de determinación ajustado

$$\bar{R}^2 = \frac{R^2 - (k - 1)}{(n - k)(1 - R^2)}$$

Donde:

n : n° de observaciones

k : n° de variables independientes

Este coeficiente R^2 no es la única herramienta que existe para saber que tan bueno o malo es el modelo. También es posible medir el error porcentual absoluto medio o MAPE, este indicaría cuánto se equivoca en promedio la estimación en cada valor. Se calcula de la siguiente manera:

Ecuación 6: Error porcentual absoluto promedio (MAPE)

$$MAPE = \sum_{i=1}^n \frac{|\hat{y}_i - y_i|/y_i}{n}$$

Para poder aplicar una regresión lineal a los datos que se quieren predecir, es necesario que se cumplan los siguientes supuestos:

1. Existe una relación lineal entre la variable independiente y con la variable dependiente x.
2. Independencia de los errores. Si existiera alguna correlación entre los errores, en teoría se podría predecir un error en base a otro, lo que deja un espacio para mejorar la predicción de y.
3. Homocedasticidad de los errores. Es decir la varianza de ellos es constante.
4. Normalidad de los errores.

8. Metodología

Recopilación de los datos

Se comenzará con la recopilación de toda la información necesaria para el proyecto. Los detalles transaccionales serán sacados de las bases de datos MySQL y la información de la competencia, como sucursales y localización, de sus propias páginas webs.

La información geográfica de los clientes como su ubicación en coordenadas de latitud y longitud, serán obtenidas desde Google Maps, con un programa que automatiza la extracción de éstas.

Los datos del supermercado serán limpiados previamente antes de comenzar a utilizarse, borrando las filas que posean valores NULL, además de eliminar los *outliers*, los que serán identificados con gráficos de histogramas y *box plots* con el *software* estadístico SPSS 17.0.

Análisis previo de la situación

Se hará un estudio previo de la situación, identificando en un mapa la sucursal del supermercado mayorista y su competencia mayorista. Se ocupará para este fin los mapas de Google Maps. Además se hará otro mapa con los clientes, de modo de identificar a cuántos km. a la redonda se encuentra qué porcentaje de las ventas (ejemplo: “a 2 km de a la redonda del local, se encuentran los clientes que representan el 50% de las ventas del supermercado”).

Para un análisis local de los clientes se usarán los *heat maps* del software on-line MyHeatMaps y se graficará el monto que gastan, su frecuencia y la densidad de ellos alrededor de la sala.

Para complementar el estudio preliminar, se hará un análisis descriptivo de todas las variables que describan a los clientes y que serían posibles de incorporar en los modelos de estimación, tales como monto comprado, ticket promedio, frecuencia, comuna de residencia, distancia al supermercado, etc.

Consolidación de los datos

Se consolidarán los datos con todas las variables necesarias en una planilla de Microsoft Excel, aprovechando que la cantidad de clientes no es tan grande (menos de 10.000 para cada sucursal del mayorista) y se guardará en formato .txt con tabulaciones.

Ejecución de los modelos de estimación

Se correrá una regresión lineal para pronosticar el monto comprado y la frecuencia esperada en la sala. Se utilizará el *software* para *data mining*, Clementine, el que recibirá de input el archivo .txt. Luego se comparará la bondad de este modelo con el de regresión geográfica ponderada, modelo que será corrido en el *software* R, que también requiere de entrada una archivo .txt. Cabe destacar que en ambos modelos se desea analizar el efecto de la distancia, trabajándose con la distancia euclidiana para el modelo de la regresión lineal y agregándose las coordenadas de latitud-longitud para la regresión geográfica.

Evaluación de zonas geográficas y recomendaciones

Basado en los resultados obtenidos en el análisis previo de la situación y en los modelos, se identificarán los barrios de interés para el mayorista, tanto porque concentran buenos o malos clientes, como también porque existen personas que son potenciales de aumentar su valor. Se identificará la ubicación de los competidores que más le quitan ventas a la sala y en base a esta información se entregarán recomendaciones de puntos donde ubicar la publicidad. Esto se realizará de forma cualitativa, entregándose los resultados en mapas de Google Maps.

9. Desarrollo metodológico

Se comienza desarrollando la metodología para el local de Maipú, para posteriormente hacer el análisis para Temuco.

9.1 Maipú

9.1.1 Datos

9.1.1.1 Datos disponibles

Se poseen los datos de todos los clientes del supermercado mayorista que hayan entregado su rut, ya sean socios o no. Se trabajará con los datos de los clientes asociados al local de Maipú, del cual se tienen las transacciones hechas desde enero 2008 hasta el presente. Para este trabajo se usarán los datos desde el 2008 hasta el 31 de diciembre de 2010, lo que incluye los datos de 4.102 clientes.

Por otro lado, para no sesgar el modelo a sólo clientes del supermercado, y a modo de incorporar a los no-clientes, se trabajó con los clientes de 3 sucursales más del mayorista: Las Condes, Puente Alto y Quilicura, personas que según el registro no han comprado jamás en Maipú y que por lo tanto serían lo equivalente a clientes de la “competencia”. Pese a que estas personas no están dentro de las áreas de influencia estudiadas, se encuentran a menos de 20 km. del supermercado, que fue el límite impuesto para estudiar a los clientes de la sala.

A todos estos clientes y no-clientes se les extrajo sus coordenadas de latitud y longitud (se usó un programa que automatizara la extracción de estos datos desde Google Maps), para posteriormente calcular las distancias a los locales.

9.1.1.2 Limpieza previa y datos finales

Del total de clientes existentes se trabajó con 3.146 correspondiente a un 72% del total. Los filtros se aplicaron según valores de distancia de la persona a la tienda y frecuencia de compra, además de eliminar clientes con valores blancos o no válidos (sin dirección completa).

Tabla 2: Datos clientes de Maipú

TOTAL CLIENTES: 4.102	CLIENTES ANALIZADOS: 2.946 (72%)	Datos correctos: 2823 (96%) Datos corregidos: 123 (4%)
	CLIENTES FILTRADOS: 1.156 (23%)	Clientes con campos en blanco o Direcciones ilegibles: 738 (77%) <i>Outliers</i> (distancias, frecuencias y montos fuera del rango establecido): 418 (23%)

Fuente: Elaboración Propia

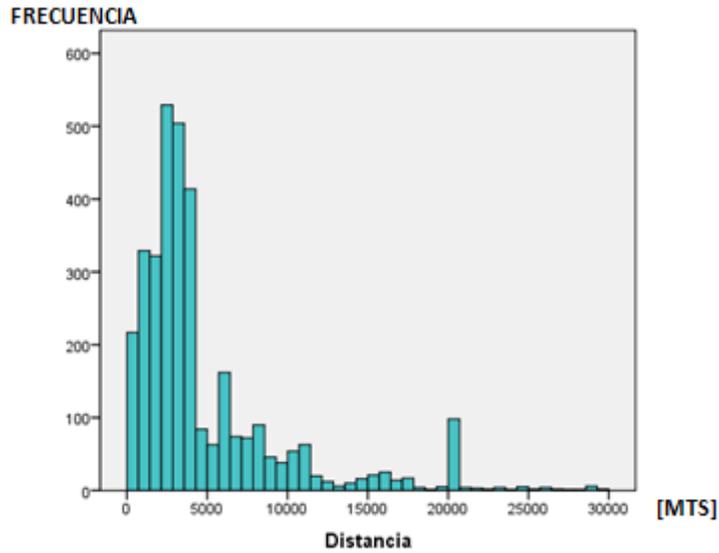
Para limpiar la base de datos se establecieron filtros previos para la distancia y la frecuencia. Si la distancia era superior a 30 km., los datos fueron excluidos del modelo, ya que no se considerará en el estudio hacer publicidad del supermercado para localidades en extremos apartadas del establecimiento.

Con respecto a la frecuencia de visitas al establecimiento, se eliminó del estudio a los clientes con frecuencias mensuales mayores a 30, suponiendo que como máximo alguien podría ir todos los días al supermercado y que una frecuencia mayor a esto, podría deberse a suplantaciones de identidad, tal como darle el rut a un conocido para que realice sus compras.

Para detectar los *outliers*, se hicieron histogramas para estudiar dentro de qué rango se encuentra la mayoría de los datos y además se hicieron *box plots* donde se observaron claramente los valores *outliers*. Se analizó la variable distancia, para decidir hasta qué radio se estudiarían los clientes.

De acuerdo al histograma, hay una alta concentración de clientes hasta los 5 km. y cerca de los 20 km, también hay otra agrupación interesante. Los clientes ubicados a más de 20 km. no se estudiaron ya que corresponderían a una minoría inferior al 5%.

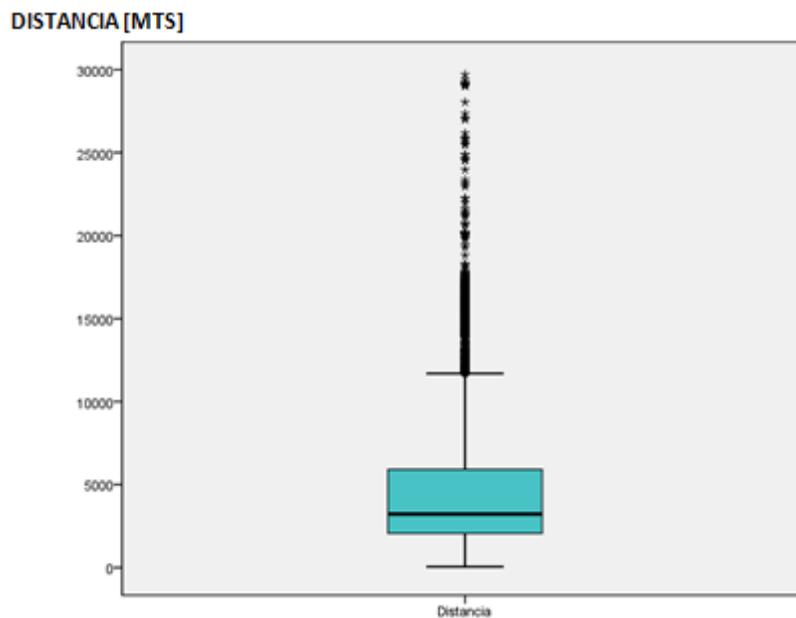
Gráfico 5: Histograma de distancias de los clientes, local Maipú



Fuente: Elaboración propia

Complementando esta información con el *box plot*, más del 90% de los clientes de Maipú se emplaza a menos de 12 km. del supermercado. Pero se observa nuevamente un núcleo a 20 km. del supermercado. Por lo tanto, se estudiará a todos los clientes que vivan a menos de 20 km. del supermercado.

Gráfico 6: *Box plot* de distancias de los clientes, local Maipú



Fuente: Elaboración propia

Como último criterio de eliminación de *outliers*, y debido a que en un periodo de 3 años existen muchos clientes fugados, se exigió que las personas debían tener un monto superior a \$360.000 durante el periodo analizado, lo que equivale a compras de \$10.000 mensuales por todos los meses, lo que descartó a 200 clientes del estudio.

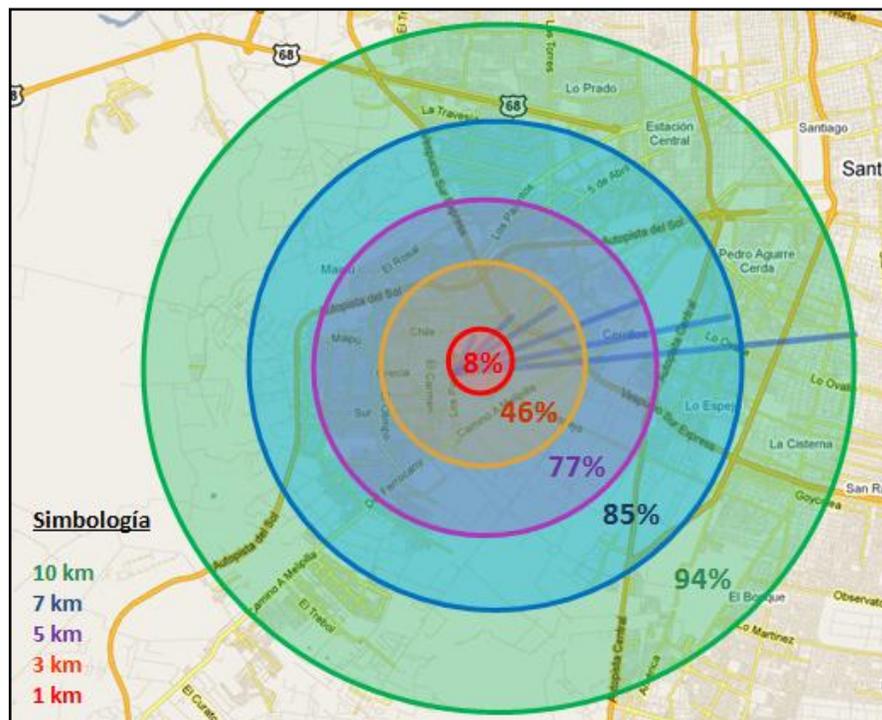
9.1.2 Análisis de la situación actual

Antes de comenzar con la confección de los modelos de estimación para los montos y la frecuencia de compra, se hará un análisis de la situación actual de los clientes, en cuanto a su distancia euclidiana al supermercado, comunas de origen, valor de sus compras, frecuencia y ticket promedio. Además se harán *heat maps* con la georeferenciación de los montos, la frecuencia y la cantidad de los clientes. Este estudio dará las primeras luces acerca de cuáles son las tendencias de compra por sector geográfico.

9.1.2.1 Distancia de los clientes al local Maipú

Se analizó qué porcentaje del monto total de las compras está explicado por las distintas distancias euclidianas al supermercado. Se analizaron los límites de 1km., 3km., 5 km., 7 km. y 10 km. a la redonda. Encontrándose que prácticamente un 50% del monto total está concentrado en los clientes que viven a menos de 3 km. del supermercado. Mientras que si aumentamos esta distancia a 7 km., ya se explica el 77% de las compras totales.

Figura 10: Porcentaje del monto de las compras explicado por la distancia, local Maipú



Fuente: Elaboración propia

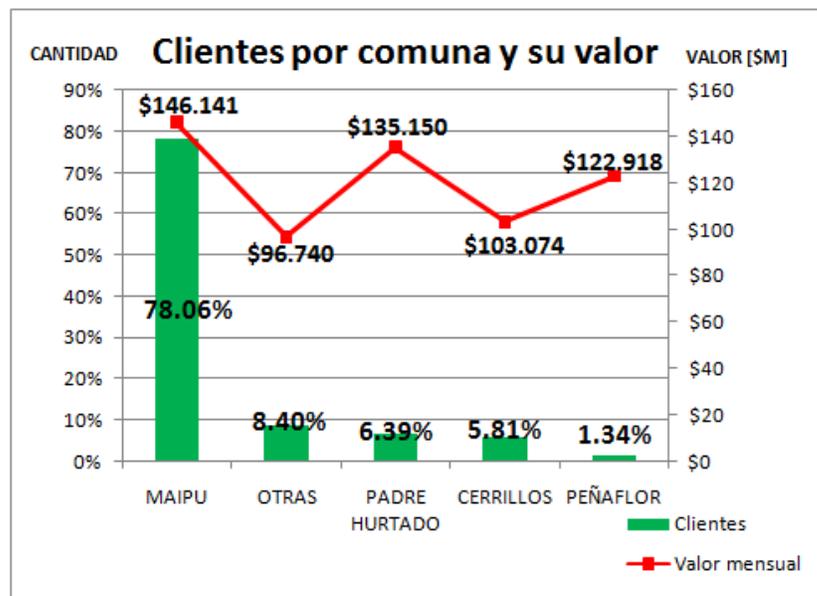
Se observa un comportamiento especial en la zona más cercana al supermercado, a menos de 1 km. de la sala sólo se explica un 8% de las compras. Sin embargo, a 3 km. de éste el monto explicado aumenta a casi la mitad y de ahí hacia la

periferia el aumento de las compras cada vez es menor, verificándose un crecimiento menguado luego de los 5 km. de lejanía.

9.1.2.2 Comunas de origen de los clientes y su valor

El local analizado se encuentra situado en la comuna de Maipú, en Av. Pajaritos 1637. Pese a que seguramente la mayoría de los clientes pertenezcan a esta comuna, de todos modos es interesante analizar otros orígenes de los consumidores. Es decir, más allá del análisis global recién hecho con los anillos en el mapa, es importante ir acercándose a un análisis a nivel más local.

Gráfico 7: Origen de los clientes del local Maipú



Fuente: Elaboración propia

Observando el gráfico 7, se aprecia que en efecto casi un 80% de los clientes reside en la misma comuna del supermercado. Sin embargo, también se observan otras comunas con una participación no despreciable. En segundo lugar se encuentra Padre Hurtado, localidad más rural que Maipú y que está ubicada al sur de ésta. Peñaflores, comuna contigua a Padre Hurtado por el sur, también aparece en el ranking pero con un menor porcentaje, lo que podría ser explicado por la mayor lejanía a la sala y también porque fue detectada más competencia en esta comuna.

La masa de clientes aparece ligada al valor de la compra mensual de ellos, exceptuando el grupo "otras" (comunidades), a medida en que hay más clientes en una localidad, las compras también son mayores, siendo Maipú el sector más poblado de clientes y el que presenta las compras más considerables, seguido de Padre Hurtado y Peñaflores.

Cerrillos, que limita por el noreste con Maipú, se lleva otro porcentaje relevante en cuanto a cantidad de clientes y valor, superando al del grupo "otras". Por lo tanto Padre

Hurtado, Peñaflores y Cerrillos, además de Maipú, serán analizadas como una localidad aparte en los modelos de estimación.

9.1.2.3 Frecuencia y ticket promedio de los clientes

Se analizó la frecuencia y el ticket promedio para las 4 comunas con más clientes, mencionadas anteriormente.

Para medir la frecuencia, se utilizó el indicador de días entre compras⁶, que se calcula de la siguiente manera:

$$\text{Días entre compras} = \frac{c_f - c_i}{(n - 1)}$$

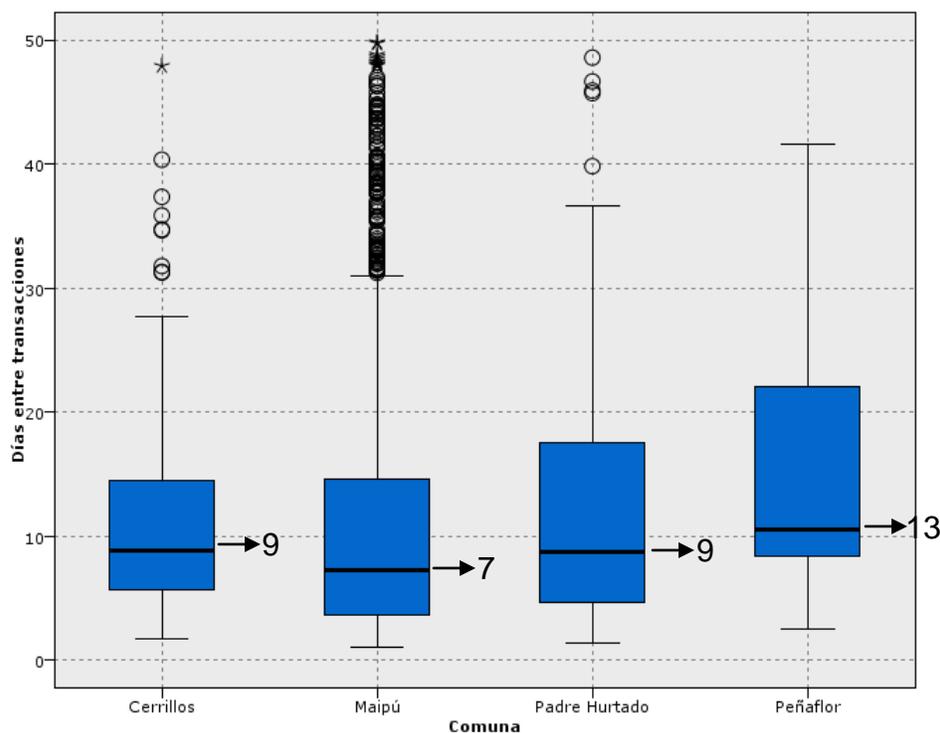
Donde:

c_f : día en que se efectuó la última compra

c_i : día en que se efectuó la primera compra

n : cantidad de compras realizadas entre c_f y c_i

Gráfico 8: Días entre compras de los clientes, local Maipú



Fuente: Elaboración propia

⁶ Entendiendo una compra como una visita.

De acuerdo al gráfico recién presentado, no existen diferencias significativas en la frecuencia. Sin embargo, en la comuna de Maipú es donde la gente va más seguido al supermercado, es decir posee un indicador “días entre compras” bajo. En tanto, en la comuna de Peñaflores, las visitas se hacen de manera más espaciada. Esto cobra sentido si pensamos que los habitantes de Maipú tienen el supermercado a una distancia mucho menor que los peñaflorinos.

Este fenómeno se hace presente para las 4 comunas: a mayor distancia, mayor indicador de días entre compras. Si se ordena a las comunas de la más cercana al supermercado a la más lejana, se obtiene el mismo orden que para las visitas al supermercado más y menos seguidas. Este es: Maipú-Cerrillos-Padre Hurtado-Peñaflores.

Por otro lado, llama la atención, que los clientes de la comuna de Maipú, son los menos fieles al supermercado, o los con más tendencia a fugarse, lo que se refleja en la gran cantidad de *outliers* para los días entre compras sobre 30. Es decir hay muchas personas de Maipú, que casi ya no están yendo al supermercado, o lo están haciendo cada 50 o 100 días, lo que hace pensar que han abandonado definitivamente al mayorista, o realizan sólo parte de sus compras en éste y el resto de la mercancía la compran en la competencia⁷.

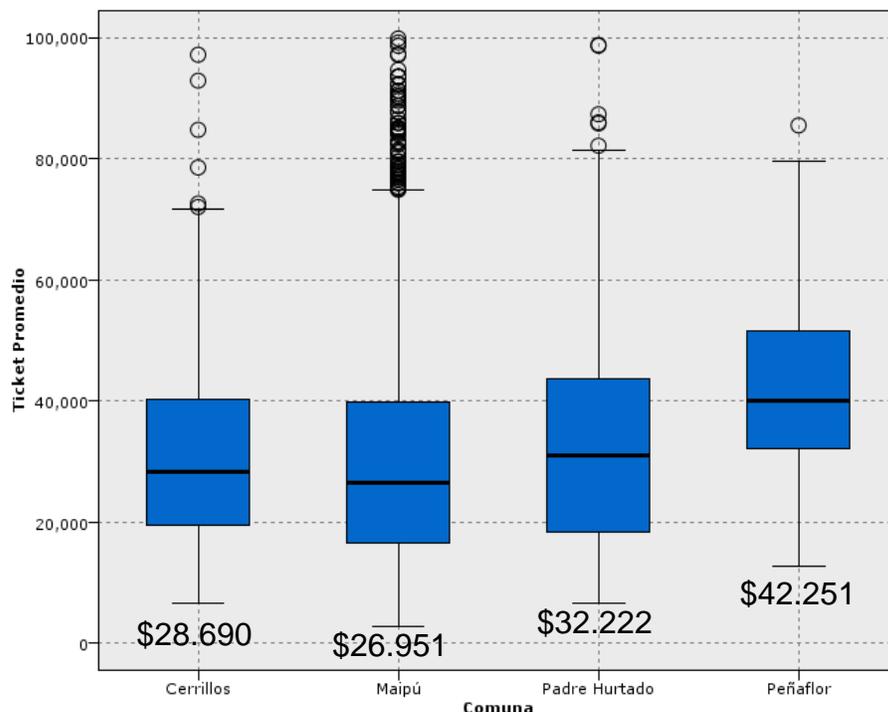
Pese a que los habitantes de Maipú, son los que van más seguido al supermercado, los de Peñaflores, que tienen menos visitas al supermercado, son finalmente los que poseen mayor ticket promedio que cualquier otro cliente. En promedio en cada visita gasta más de \$42.000, mientras que los maipucinos sólo \$27.000 aproximadamente.

Pese a esto, nuevamente se observa mucha heterogeneidad en los comportamientos del cliente de Maipú, presentándose muchos *outliers*, es decir aunque el ticket promedio sea bajo, existen algunos clientes en Maipú, que gastan grandes sumas en cada visita.⁸

⁷Para mirar el *box plot* con todos los *outliers* dirigirse al Anexo B.

⁸Detalle de *outliers* en el Anexo B.

Gráfico 9: Ticket promedio de los clientes de Maipú, acercamiento



Fuente: Elaboración propia

Debido a que el ticket promedio y los días entre compras, son variables diferenciadoras entre un cliente y otro, serán incluidas en los modelos de predicción, además de la comuna de residencia del consumidor.

9.1.2.4 Heat maps del valor, los días entre compras y la densidad de clientes⁹

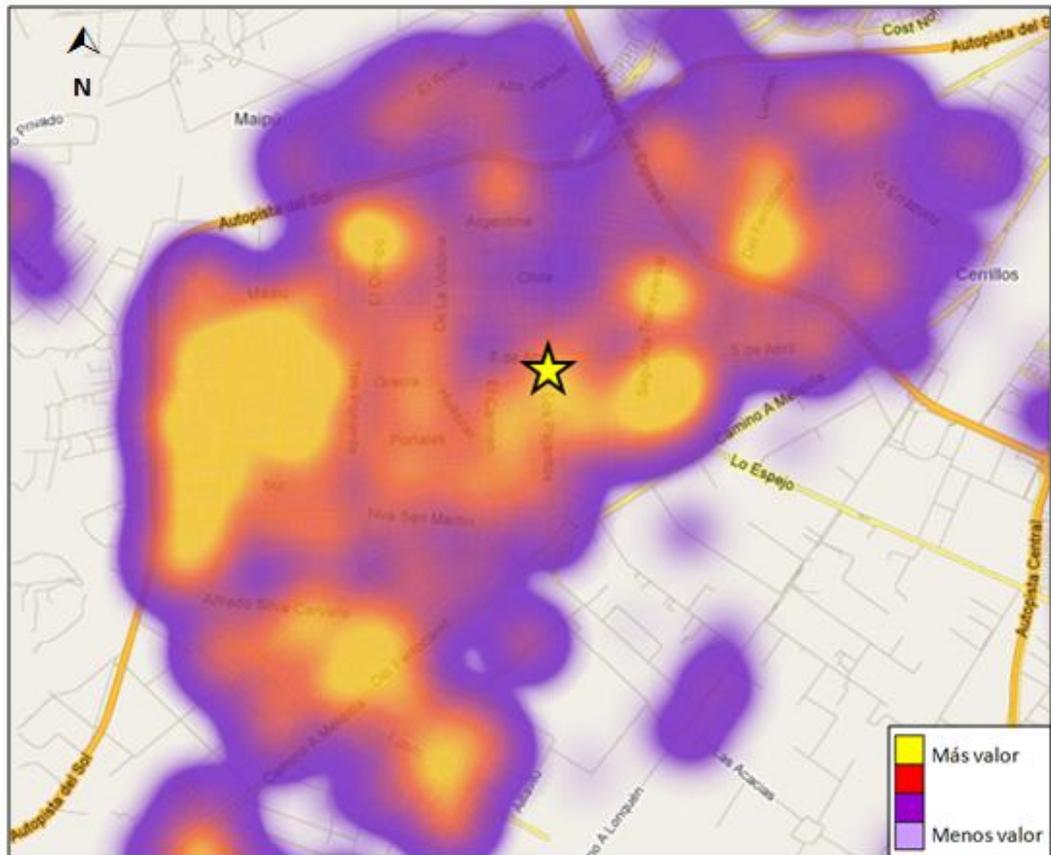
Para hacer un análisis de la situación actual desde un enfoque local, es decir dejar de lado el análisis por anillos a la redonda y ver el comportamiento de los clientes por barrios específicos, se graficó el monto, la frecuencia de las compras y la densidad de los clientes a través de *heat maps*¹⁰.

⁹ Los colores de los siguientes *heat maps* representan intensidad y no valores, siendo que los colores más cálidos, como el amarillo, reflejan mayor intensidad para esa variable y los más fríos, como el morado, menor intensidad.

¹⁰ Se graficaron los datos de los clientes del local Maipú con las compras hechas desde el año 2008 al 2010.

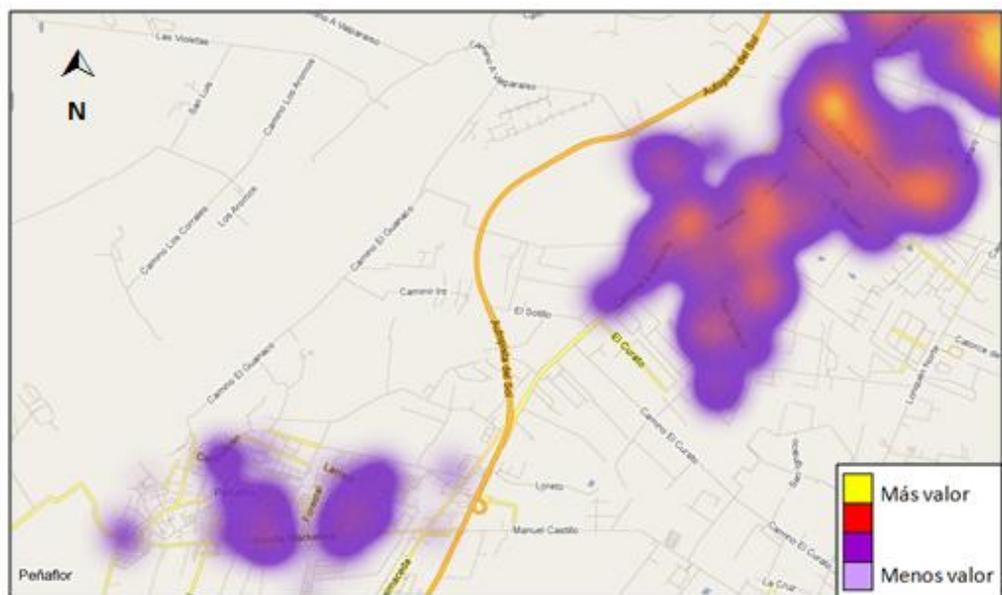
Valor

Figura 11: Heat map del valor en local Maipú, comunas Maipú y Cerrillos



Fuente: Elaboración propia

Figura 12: Heat map del valor en local Maipú, comunas Padre Hurtado y Peñaflo



Fuente: Elaboración propia

En el mapa se aprecia claramente que la mayor cantidad de ventas se produce en Maipú. En esta comuna, las ventas no se distribuyen uniformemente alrededor del supermercado, sino que se ven sujetas al plano de la ciudad y a los barrios dentro de la comuna. Por ejemplo, la zona amarilla de las ventas se ve cortada abruptamente por Camino a Melipilla por el sur-orientado y por la Autopista del Sol en el poniente y en el norte. Además cabe destacar que justo al norte del supermercado hay un hueco en las ventas, éste corresponde al sector de Alto Jahuel, que es una zona residencial de mayores ingresos con respecto al resto de la comuna, lo que se escapa un poco del *target* del negocio mayorista que va más enfocado a la clase media-baja.

En cambio, al poniente del supermercado, se ve la mayor intensidad en compras. Este sector que colinda con la Autopista del Sol al poniente, es una zona con gran cantidad de villas y poblaciones. En estos lugares no existe ningún supermercado y los habitantes en su mayoría son personas del GSE D, son precisamente ellos los que adquieren un mayor monto en el supermercado¹¹.

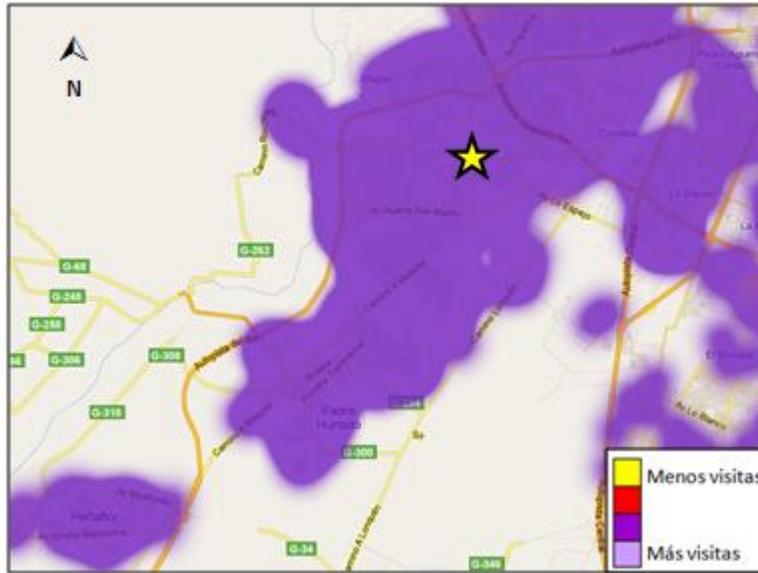
Otros sectores importantes de ventas, son las comunas de Padre Hurtado y Peñaflo, más la primera que la segunda, que no sólo posee un radio de acción ancho, sino que también tiene zonas “calientes”, es decir sectores donde el monto de la compra es alto, especialmente el lado norte de esta comuna que es el más cercano a Maipú.

Días entre compras

Debido a que las diferencias entre los días entre compras no eran muy grandes entre comunas, variando desde 7 a 13 entre clientes del local Maipú, al llevar esto a *heat maps*, las diferencias son muy insignificantes como para poder apreciarse diferencias de color entre barrios. Tal como se aprecia a continuación, el mapa muestra una única masa de color morado, lo que indicaría que todos asisten más o menos con la misma regularidad a comprar. Siendo los maipucinos los que van más seguidos, teniendo una frecuencia semanal, mientras que los de Peñaflo son los que asisten menos veces, esto es entre 2 y 3 veces al mes aproximadamente.

¹¹ Para ver detalles acerca de cuánto gasta cada cliente según su posición geográfica ir a *scatter plot* del monto en el Anexo C.

Figura 13: *Heat map* de los días entre compras en local Maipú



Fuente: Elaboración propia

Esto hace pensar que es más interesante mirar la diferencia en la concentración de clientes por barrios, más que apreciar un *heat map* para la frecuencia de la compra.

Densidad de clientes

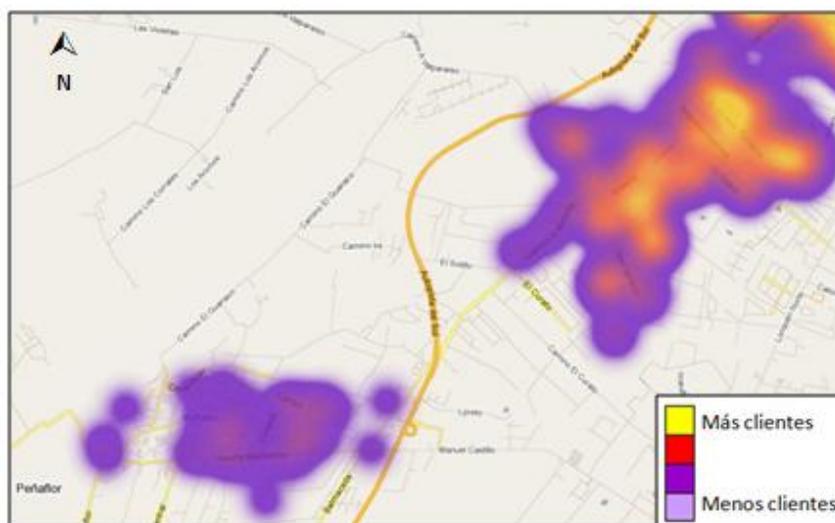
A continuación un *heat map* para la densidad de clientes del local Maipú.

Figura 14: *Heat map* de la densidad de clientes en local Maipú, comunas Maipú y Cerrillos



Fuente: Elaboración propia

Figura 15: Heat map de la densidad de clientes en local Maipú, comunas Padre Hurtado y Peñaflo



Fuente: Elaboración propia

Comparando este mapa con el del valor, hay ciertas semejanzas como que la mayor concentración, en este caso de clientes, está en la comuna de Maipú, seguida de Padre Hurtado en segundo lugar.

Sin embargo, en esta oportunidad hay una distribución de clientes mucho más uniforme alrededor del supermercado, en comparación al monto. La mayor discontinuidad de clientes se da sólo al norte del supermercado, comprobándose que cerca del mayorista sí hay hartos clientes, pero que no compran tanto como se demostró anteriormente. Por lo tanto estos barrios cercanos al supermercado, con gran concentración de clientes pero bajo monto, serían sectores interesantes de analizar, contenido que será abordado en la sección 9.1.4.1 Barrios de interés.

9.1.2.5 Competencia

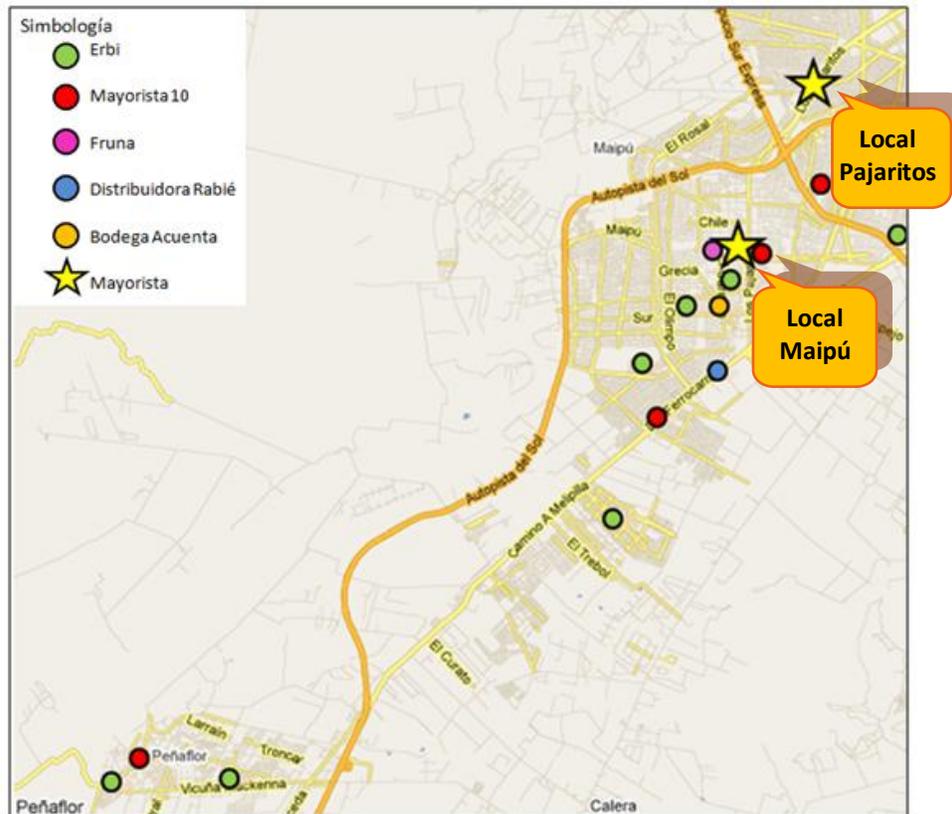
Se trabajó con los locales de la competencia que ya manejaba el supermercado y además se actualizó el listado existente con nuevas salas que se han ido instalando.

Se consideró como competencia de la sala Maipú, a todo supermercado mayorista, ubicado en la comuna de Maipú o en las comunas colindantes con ésta¹² (Padre Hurtado, Cerrillos, Pudahuel y Estación Central) y localizados a menos de 10 km del mayorista. Notar que si existía un supermercado mayorista a menos de 10 km, pero ubicado en una comuna diferente a las mencionadas, se dejó fuera del análisis. Por otro lado, también se consideró como competencia a los mayoristas ubicados en la comuna de Padre Hurtado y Peñaflo, debido a clientes de la sala Maipú que viven en estas zonas.

¹² Para más detalles, mirar en Anexo A un mapa con las comunas de Santiago.

En total se encontraron 14 locales mayoristas, los que se pueden observar a continuación.¹³

Figura 16: Competencia mayorista de local Maipú



Fuente: Elaboración propia

9.1.3 Modelos de estimación para el monto y la frecuencia

Se confeccionarán dos modelos de regresión para predecir el monto de la compra y dos más para los días entre compras. Los modelos a realizar cubrirán el enfoque global y local del análisis espacial. Se hará desde el punto de vista global una regresión lineal, la que necesitará como *input* la distancia euclidiana del cliente al supermercado, despreciando la localización exacta en el espacio. Mientras que para el enfoque geográfico local se realizará un regresión geográfica ponderada, que trabajará con la posición (x,y) del local y de cada cliente. La distancia será incluida para ver si hay efectos de ésta a nivel local.

Se comparará la bondad de ajuste de ambos modelos, el error MAPE, cantidad de variables significativas y se elegirá el mejor según estos criterios.

¹³ Notar que el análisis se hace desde la perspectiva del local de Maipú, por lo que la sala de Pajaritos sería considerado una competencia.

9.1.3.1 Variables a incluir

Basado en los modelos gravitacionales, se incorporarían como base, las variables distancia del cliente al supermercado y tamaño de la sala. Sin embargo, como se trabajará con los datos de compra de sólo los clientes de este supermercado y no su competencia, el tamaño de la sala no será considerado ya que es una constante.

Variables base del modelo:

- Distancia del cliente al supermercado
- Distancia del cliente a los supermercados de la competencia

Además de estas variables, se incluirán otras que permitan caracterizar de manera diferenciadora a un grupo importante de personas.

Variables posibles de incluir en el modelo:

- Ticket promedio
- Frecuencia de compra
- nº de compras
- *Recency*
- Cercanía
- Cantidad de supermercados de la competencia cercanos al cliente
- Comuna

Descripción de las variables

A continuación se hace una breve descripción del listado de las últimas variables que se incluirán en los modelos de regresión:

- Ticket promedio: cantidad promedio gastada en cada visita al supermercado.
- Frecuencia de compra: entendida como días entre compras, se refieren a cuántos días pasan entre visita y visita al mayorista.
- nº de compras: cantidad de visitas hechas al supermercado en el periodo de tiempo total considerado (3 años).
- *Recency*: cantidad de días que han pasado desde la última visita del cliente.
- Cercanía: variable binaria que vale 1 si el cliente vive a menos de 2 km del supermercado mayorista¹⁴.

¹⁴ Más adelante se explicará por qué se consideró una distancia de 2 km. y no más ni menos.

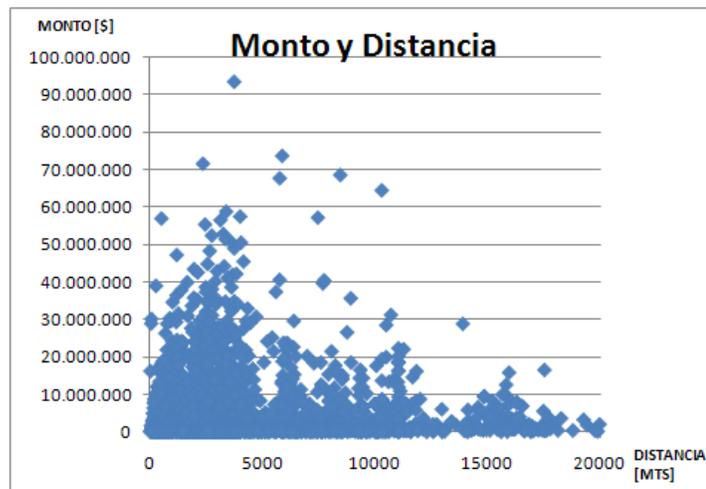
- Cantidad de supermercados de la competencia cercanos al cliente: cantidad de supermercados de la competencia que quedan a menos de 2km. del cliente.
- Comuna: conjunto de variables binarias que indican en qué comuna reside el cliente.

9.1.3.2 Análisis descriptivo de las variables

Se hizo un estudio preliminar de las variables propuestas, para ver si permitían discriminar a las personas según su monto de compra o frecuencia.

Para las variables, frecuencia y distancia se observaron comportamientos diferenciados contra la variable dependiente monto. Cumpléndose que montos mayores de compra ocurren a menor distancia de la sala y mayores frecuencias (menor coeficiente “días entre compras”) también están asociados a una mayor cercanía al supermercado. Esto se comprueba con los gráficos de dispersión y *box plots* que se muestran a continuación:

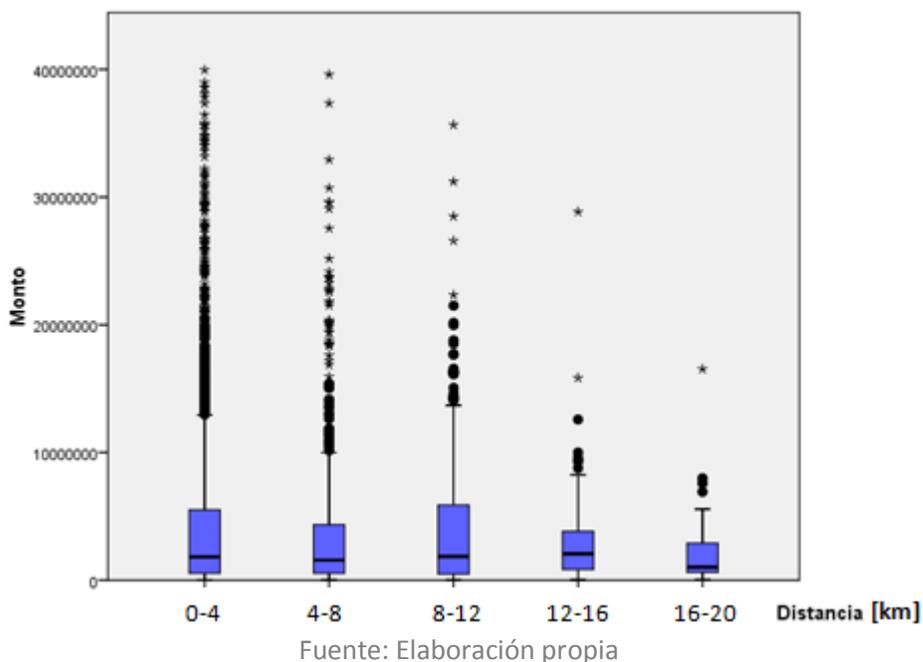
Gráfico 10: Monto y distancia en *scatter plot*, local Maipú



Fuente: Elaboración propia

En el gráfico 10, se aprecia una disminución del monto a medida que va aumentando la distancia, sin embargo se puede notar una gran dispersión para distancias menores a 5km., es decir cerca del supermercado existe gente que pueda comprar poco o mucho, lo que ya fue evidenciado en los *heat maps*, donde en la zona norte y oriente, muy cercanas al supermercado, no se encontraron grandes montos de compra, mientras que hacia el poniente sí.

Gráfico 11: Monto y distancia en *box plot*, local Maipú

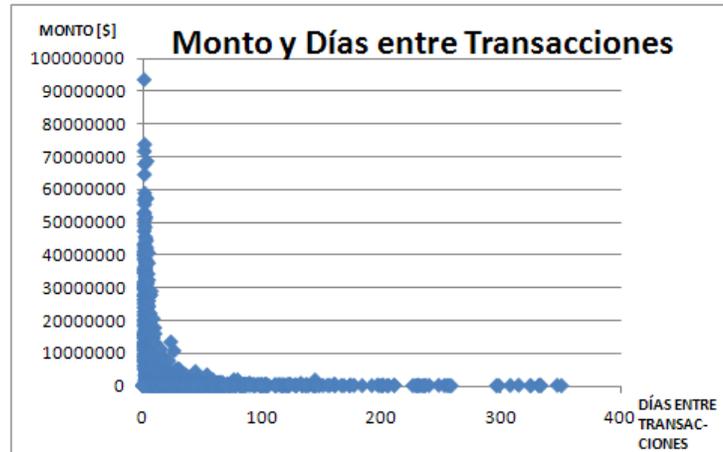


El *box plot* muestra que en promedio todos estarían comprando aproximadamente lo mismo, pero que en las zonas más cercanas al supermercado (*box 1 y 2*), existen personas que compran mucho más que el resto.

Otro aspecto interesante, resulta al notar que entre las *boxes 2 y 3* el monto hace un salto. Este último *cluster*, que cubre entre 8 y 12 km., coincide con la distancia promedio a la que se encuentra la comuna de Padre Hurtado respecto al supermercado, es decir acá se ve representado que el núcleo de esta comuna incluso compra más que otras zonas de Maipú.

Respecto a los días entre compras, tal como fue mencionado anteriormente, las personas que realizan más visitas a las salas, son las que compran más. Por lo que está presente el desafío de hacer asistir a las personas más seguido al supermercado.

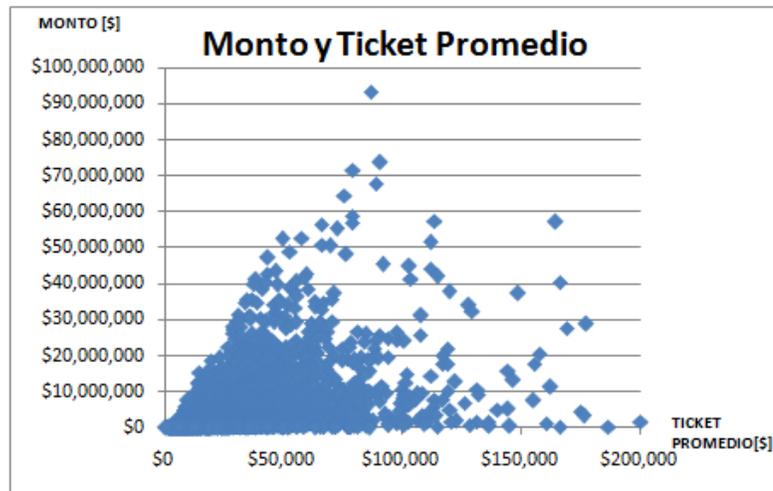
Gráfico 12: Monto y días entre compras, local Maipú



Fuente: Elaboración propia

Para la variable ticket promedio también se realizó un gráfico de dispersión, pero no se observó una tendencia tan pronunciada como en el caso anterior, aunque claramente tickets bajos son incompatibles con montos altos, en especial considerando que más de la mitad de las personas van entre 2 y 4 veces al supermercado.

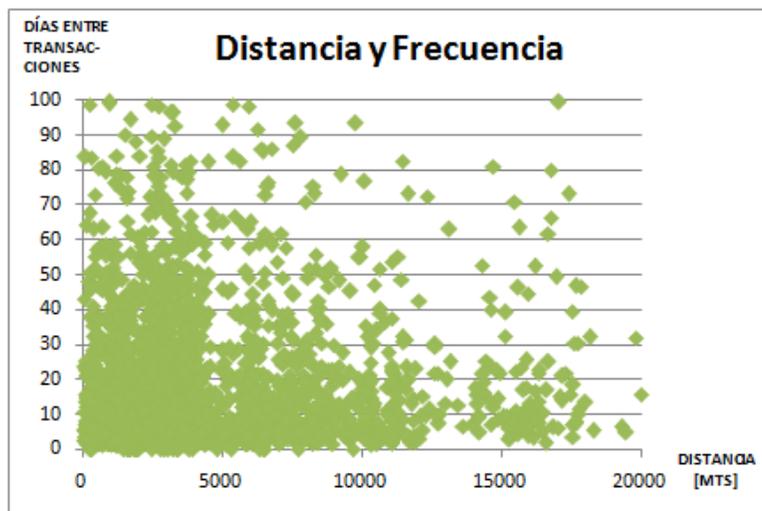
Gráfico 13: Monto y ticket promedio, local Maipú



Fuente: Elaboración propia

Finalmente es interesante apreciar que para la variable frecuencia (días entre compras) en función de la distancia, no hay cambios significativos en el comportamiento de las personas, tendencia que ya fue observada en el *heat map* donde toda la masa presentó un mismo color, haciendo pensar que las personas asisten al supermercado con cierta frecuencia que no está o está poco determinada por la distancia.

Gráfico 14: Días entre compras y distancia, local Maipú



Fuente: Elaboración propia

9.1.3.3 Modelo de regresión lineal

Se realizó un modelo de regresión lineal para estimar el monto y los días entre compras. Se ocuparon variables de tipo transaccional como lo es el ticket promedio o el *recency* y además se incorporaron variables que son características para cada cliente, como la cantidad de supermercados que le quedan cerca o la comuna de residencia.

Para calcular la importancia de cada variable en el modelo, se calcularon los coeficientes betas con una muestra de calibración igual al 80% de los datos, y luego se hizo un testeó de los betas arrojados, para el 20% de los datos restantes. El error MAPE de las regresiones fue calculado sólo sobre la muestra de testeó.

Estimación del monto

Para predecir el monto de la compra de cada cliente, se realizó un modelo con las siguientes variables:

- Distancia al supermercado
- Cercanía al supermercado
- Distancia a los supermercados de la competencia
- Cantidad de competencia cercana
- Comuna de residencia
- Ticket promedio
- Días entre compras
- nº compras
- *Recency*

Ecuación 7: Modelo de estimación del monto para un cliente i , local Maipú

$$\begin{aligned} \text{Monto}_i = & \beta_{\text{maipu } i} \cdot \text{maipu} + \beta_{\text{cerrillos } i} \cdot \text{cerrillos} + \beta_{\text{padre h } i} \cdot \text{padre hurtado} \\ & + \beta_{\text{peñaflor } i} \cdot \text{peñaflor} + \beta_{\text{díasentre } i} \cdot \text{días entre compras} \\ & + \beta_{\text{ncompras } i} \cdot \text{cantidad de compras} + \beta_{\text{recency } i} \cdot \text{recency} \\ & + \beta_{\text{ticket } i} \cdot \text{ticket promedio} + \beta_{\text{cercania } i} \cdot \text{cercania} \\ & + \beta_{\text{distancia } i} \cdot \text{distancia a local maipu} \\ & + \beta_{\text{distanciacompetenciaj } i} \cdot \text{distancia a competencia } j \\ & + \beta_{\text{Qcomp } i} \cdot \text{cantidad de competencias cercanas} \end{aligned}$$

$$\forall i = 1 \dots 2.946$$

La elección acerca de qué distancia se consideraría como cercana, ya sea 1 km., 2 km., u otra, requirió correr el modelo repetidas veces con las diferentes variables confeccionadas, considerando como cercano: 1km, 1.5 km., 2 km y 3 km. Se comparó el coeficiente de R^2 para ver qué distancia ajustaba mejor en el modelo. El resultado arrojado fue que considerando 2 km. como cercano la bondad del ajuste fue mejor, por lo que la variable “cercanía al supermercado” es una binaria, que vale 1 si la persona se ubica a menos de 2 km. de la sala y “cantidad de competencias cercanas” es la cantidad de supermercados mayoristas, que quedan a menos de 2 km. del cliente.

Al igual que “cercanía al supermercado”, la variable “comuna de residencia” también es una variable binaria. Para evitar construir tantas variables binarias como comunas de residencia, se tomaron en cuenta las 4 comunas con más presencia de clientes del supermercado de Maipú (Maipú, Cerrillos, Padre Hurtado y Peñaflor), construyendo 4 variables binarias para las comunas, donde el valor de las 4 *dummies* fijadas en cero, corresponde a un cliente que vive en una comuna diferente a las mencionadas. Las variables de comuna de residencia son las primeras 4 del modelo de la ecuación 8.

Por otro lado, las variables “distancia a competencia j ” que marcan la distancia del cliente a un supermercado j de la competencia, en esta ocasión serán 13, ya que son 13 los supermercados ubicados a menos de 5km. Corriendo otras regresiones con los supermercados a una distancia mayor que ésta, ninguna variable resultó significativa.

Utilizando el método *stepwise* para regresiones lineales en el software Clementine, la bondad de ajuste, los coeficientes obtenidos para la distancia y las variables que resultaron significativas son los siguientes:

Tabla 3: Regresión lineal para el monto, local Maipú

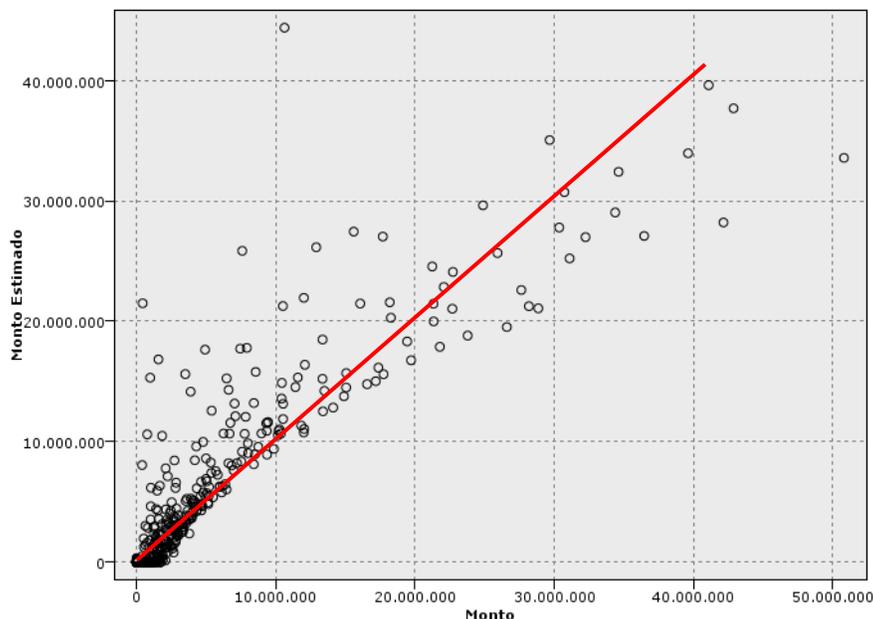
Variable	Coef.Beta	Valor	Sig.
Maipú	β_{maipu}	-0.404	0.000
Cerrillos	$\beta_{cerrillos}$	-0.128	0.000
Padre Hurtado	$\beta_{padre h}$	-0.137	0.000
Peñaflor	$\beta_{peñaflor}$	-0.085	0.000
Ticket Promedio	β_{ticket}	0.669	0.000
nº compras	$\beta_{ncompras}$	0.697	0.000
Cercanía	$\beta_{cercania i}$	-0.034	0.000
Mayorista 10 Peñaflor	$\beta_{distanciaM 10P}$	0.015	0.012
R² ajustado = 0.764 MAPE=50%			

Fuente: Elaboración propia

Estudiando los coeficientes, se ve que las variables de tipo transaccional pesan más en importancia al momento de estimar el monto, que las de información de cliente, como las comunas de residencia. Las variables ticket promedio y nº de compras tienen un coeficiente beta sobre el nivel 0.5, donde un aumento de éstas explica un aumento en el monto final también. Cabe destacar que la distancia no resultó significativa en este modelo, pero sí la cercanía, aunque con un coeficiente estandarizado de un nivel muy bajo, restándole importancia al efecto final en el monto. Lo mismo ocurre con un local de la competencia que resultó significativo, su coeficiente beta no es lo suficientemente alto como para generar un impacto alto en el valor de la compra.

Todas estas variables en conjunto explican casi un 80% de la varianza que se da en el monto de la compra, lo que representa un muy buen ajuste. Sin embargo, también es interesante estudiar el error que se comete en la estimación del monto de cada cliente y si acaso el modelo de regresión lineal, se equivoca más en clientes de ciertas características que otros.

Gráfico 15: Ajuste de la regresión lineal para el monto, local Maipú



Fuente: Elaboración propia

En el gráfico anterior la línea roja representa al modelo de estimación que hubiese sido perfecto en la predicción y los puntos lo que ocurrió en la realidad. Se ve que el modelo subestima los montos de las personas que compran más de \$20.000.000 y en general sobreestima a los que van entre \$0 y \$20.000.000. Además, para las personas que compran menos de \$10.000.000 el modelo erra en menor magnitud que para montos mayores. Se obtiene un MAPE total del 50%.¹⁵

Estimación de los días entre compras

Para predecir los días entre compras se utilizaron las mismas variables que para predecir el monto, salvo que el monto ahora fue tomado como un *input*.

Ecuación 8: Modelo de estimación para los días entre compras para un cliente i , local Maipú

$$\begin{aligned}
 \text{Días entre compras}_i = & \beta_{\text{maipu } i} \cdot \text{maipu} + \beta_{\text{cerrillos } i} \cdot \text{cerrillos} \\
 & + \beta_{\text{padre h } i} \cdot \text{padre hurtado} + \beta_{\text{peñaflor } i} \cdot \text{peñaflor} \\
 & + \beta_{\text{monto } i} \cdot \text{monto} + \beta_{\text{ncompras } i} \cdot \text{cantidad de compras} \\
 & + \beta_{\text{recency } i} \cdot \text{recency} + \beta_{\text{ticket } i} \cdot \text{ticket promedio} \\
 & + \beta_{\text{cercania } i} \cdot \text{cercania} + \beta_{\text{distancia } i} \cdot \text{distancia a local maipu} \\
 & + \beta_{\text{distancia a competencia } j} \cdot \text{distancia a competencia } j \\
 & + \beta_{Q_{\text{comp } i}} \cdot \text{cantidad de competencias cercanas}
 \end{aligned}$$

$$\forall i = 1 \dots 2.946$$

¹⁵ Para detalles acerca de los errores, ir al Anexo D para ver la cantidad y distribución de los errores porcentuales absolutos.

Haciendo una regresión con el método *stepwise* nuevamente, las variables que resultaron ser significativas, sus coeficientes y la bondad de ajuste, son observadas a continuación:

Tabla 4: Regresión lineal con *recency* para días entre compras, local Maipú

Variable	Coef.Beta	Valor	Sig.
Maipú	β_{maipu}	-0.016	0.000
Cerrillos	$\beta_{cerrillos}$	-0.006	0.000
Padre Hurtado	$\beta_{padre h}$	-0.006	0.000
Peñaflor	$\beta_{peñaflor}$	-0.003	0.000
Competencias cercanas	β_{Qcomp}	0.002	0.001
<i>Recency</i>	$\beta_{recency}$	0.986	0.000
Ticket promedio	β_{ticket}	0.003	0.000
R² ajustado = 0.998 MAPE=51%			

Fuente: Elaboración propia

La variable *recency* tiene un coeficiente beta extremadamente alto en relación a las otras variables, haciendo pensar que los días entre compras sólo están explicados por esta dimensión. Por esta razón se corre una segunda regresión sin incorporarla.

Tabla 5: Regresión lineal sin *recency* para los días entre compras, local Maipú

Variable	Coef.Beta	Valor	Sig.
Maipú	β_{maipu}	-0.760	0.000
Cerrillos	$\beta_{cerrillos}$	-0.222	0.000
Padre Hurtado	$\beta_{padre h}$	-0.238	0.000
Peñaflor	$\beta_{peñaflor}$	-0.097	0.000
Ticket promedio	β_{ticket}	-0.137	0.000
Local Pajaritos	$\beta_{distanciaPajaritos}$	-0.152	0.001
Erbi Cerrillos	$\beta_{distanciaErbiC}$	-0.137	0.001
R² ajustado = 0.915 MAPE=301%			

Fuente: Elaboración propia

El R² cambia a 0.915, baja poco significativa, encontrándose otras variables con coeficientes betas relevantes que explicarían la frecuencia de un cliente en el

supermercado. Sin embargo el MAPE aumenta considerablemente hasta un 300%, mostrando que el primer modelo sería más exacto en su predicción que el segundo¹⁶.

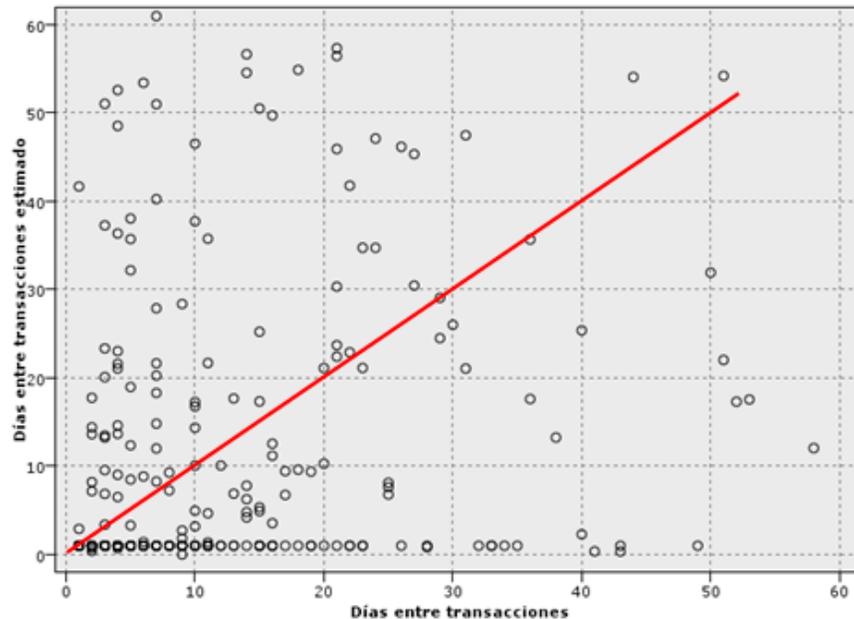
Pese a esto, el hecho de sacar la variable que más explica la frecuencia de compra, permite visualizar mejor cuáles son los otros factores que influyen en la periodicidad de compra de las personas.

Las variables de mayor importancia (mayor coeficiente beta) son las comunas de residencia, donde vivir en Maipú causaría el mayor efecto en la continuidad de las visitas (beta alto y negativo que indica que los días entre compras son menores). En tanto vivir en Peñaflores significaría asistir con menos regularidad al supermercado, tal como fue apreciado anteriormente en los *box plot*.

Otras variables de coeficiente beta negativo son el ticket promedio, donde personas de ticket promedio más bajo asistirían menos al supermercado y dos locales de la competencia, donde uno es de la misma cadena del mayorista en estudio, pudiendo comprobarse que existe canibalización entre las sedes.

A continuación se muestra el ajuste del segundo modelo (sin *recency*), la línea roja representa lo que hubiese ocurrido si la predicción hubiese sido perfecta. Dado que el error fue bastante alto, no es posible predecir con exactitud cada cuanto asisten las personas al supermercado, sin embargo el alto coeficiente R^2 permite entender bastante bien cuáles son las variables que influyen en el comportamiento de la frecuencia.

Gráfico 16: Ajuste de la regresión lineal para los días entre compras, local Maipú



Fuente: Elaboración propia

¹⁶ Para ver la distribución de los errores ir al Anexo D.

9.1.3.4 Modelo de regresión geográfica ponderada

Con el fin de mejorar las predicciones que se hacen a nivel global con las regresiones lineales, se correrán regresiones geográficas ponderadas que sean capaces de estimar valores para el monto y los días entre compras a nivel local. A través de este tipo de regresiones, es posible captar información que una regresión lineal no lo hace. Por ejemplo, en éstas últimas es poco probable que los locales de la competencia sean significativos, ya que éstos son importantes para cierto barrio en particular y no para una ciudad entera, que es el efecto que capta la RL. En cambio la GWR, como trabaja a nivel de barrios, es capaz de obtener qué competencia es importante, ya que calibra sólo a los vecinos cercanos a este local, y no al mapa completo.

La regresión geográfica ponderada, calcula sus estimaciones, a través de esta calibración por ventanas o por barrios y además trabaja con el concepto de cercanía y lejanía, ya que los vecinos más cercanos tienen un peso mayor que los más lejanos para una calibración.

Por esto mismo, es prioritario que el ancho del barrio o *bandwidth* sea ajustado minuciosamente de modo de mejorar la predicción.

Cálculo del *bandwidth*

Mientras que la función de ponderación dentro de cada barrio, afecta poco el resultado final de las estimaciones, el *bandwidth* es el parámetro más importante dentro de la GWR. Un valor incorrecto de éste, produce grandes errores en la predicción, por lo que es prioritario obtener su valor óptimo.

Se usa el software estadístico R y su método “*gwr.sel*”, para calcular el *bandwidth* que minimice la sumatoria de los residuos al cuadrado. Los resultados obtenidos son los siguientes:

Para el monto:

Bandwidth= 3.313804 km.

Para los días entre compras:

Bandwidth = 0.3613906 km.

Cabe mencionar, que para éstos cálculos se ocupa al 100% de los clientes, ya que no es posible particionar la muestra en una de calibración y otra de testeo, ya que el método necesita del 100% de los datos para los cálculos, debido a que los coeficientes betas y el R^2 es calculado para cada cliente y ningún cliente es igual a otro, por lo que los coeficientes de uno no pueden ser aplicados a otro.

Estimación del monto

Debido a que no es posible particionar la muestra en clientes para testeo y clientes para calibración, el cálculo de los coeficientes beta se hace sobre toda la población pero sólo sobre dos periodos de compra (año 2008 y 2009), mientras que el error MAPE se calcula sobre la predicción hecha para el año siguiente.

Tabla 6: Regresión geográfica ponderada para el monto, local Maipú

Variable	Coef.Beta	Valor	Sig.
Maipú	β_{maipu}	-0.058	0.000
Padre Hurtado	$\beta_{padre h}$	-0.020	0.021
Peñaflor	$\beta_{peñaflor}$	-0.018	0.031
Ticket promedio	β_{ticket}	0.707	0.000
Días entre compras	$\beta_{diasentre}$	-0.063	0.000
nº compras	$\beta_{ncompras}$	0.755	0.000
Local Pajaritos	$\beta_{distanciaPajaritos}$	0.255	0.000
M10 Maipú	$\beta_{distanciaM 10M}$	0.422	0.000
M10 Peñaflor	$\beta_{distanciaM 10P}$	0.338	0.000
Fruna	$\beta_{distanciaFruna}$	0.074	0.000
R² ajustado = 0.792 MAPE=62%			

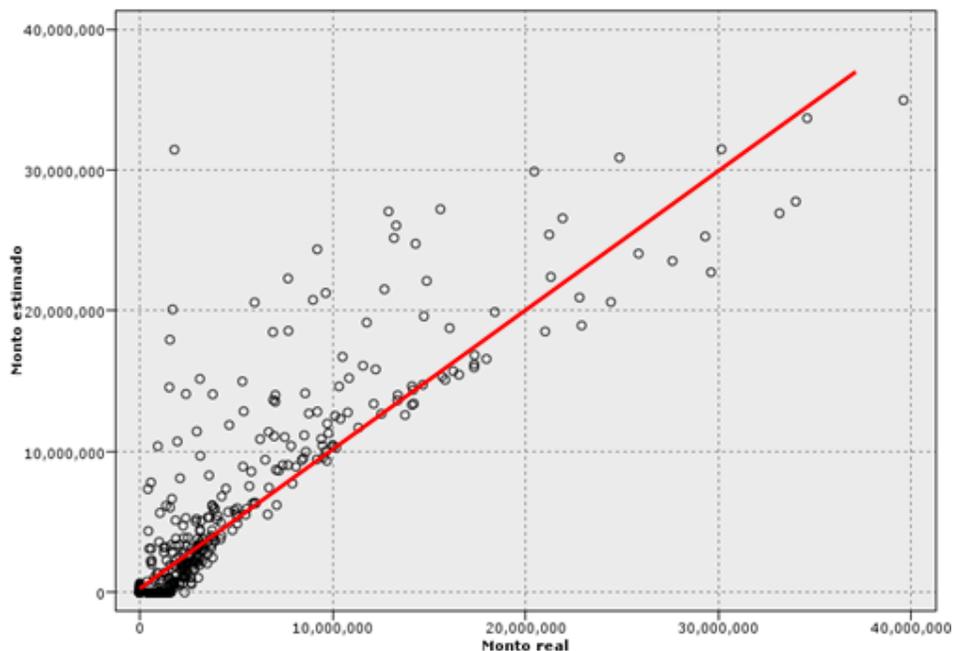
Fuente: Elaboración propia

Comparando los resultados de la GWR con la RL, lo primero que destaca es que en esta ocasión se agregan 4 locales de la competencia que resultan significativos. Las tiendas con coeficiente beta positivo, indican que mientras más alejado esté el cliente de esta competencia es mejor, porque comprarían más en el local del mayorista de Maipú. En este sentido el local que presentó un mayor efecto en disminuir el valor de las compras sería Mayorista 10 Maipú ubicado a 5 cuadras del supermercado en estudio. También destaca la aparición del local Pajaritos de la misma cadena analizada, lo que demuestra que existe canibalización entre los locales.

Además son significativas otras variables como los días entre compras, de signo negativo que indica que la gente que tiene días entre compras más altos, es decir asisten más espaciadamente al supermercado, gasta menos. Las variables comuna de residencia, ticket promedio y nº de compras son igual de significantes como en la regresión lineal y tienen la misma interpretación.

Estudiando el ajuste de la regresión se obtiene lo siguiente:

Gráfico 17: Ajuste de la GWR para el monto, local Maipú



Fuente: Elaboración propia

El ajuste reflejado en el gráfico recién mostrado es prácticamente igual que con la regresión lineal, al igual que el coeficiente de determinación y el MAPE¹⁷. Sin embargo, la información acerca de otras variables significativas, en especial la de los locales de la competencia, le aporta un gran valor a la estimación con las regresiones geográficas ponderadas.

Estimación de los días entre compras

Con el mismo modelo que para la regresión lineal, se predicen los días entre compras con GWR. Los resultados obtenidos son los siguientes:

¹⁷ Para detalles acerca de los errores, ir al Anexo D para ver la cantidad y distribución de los errores porcentuales absolutos.

Tabla 7: Regresión geográfica ponderada para los días entre compras, local Maipú

Variable	Coef.Beta	Valor	Sig.
Maipú	β_{maipu}	-0.019	0.021
Recency	$\beta_{recency}$	0.983	0.000
Acuenta	$\beta_{distanciaAcuenta}$	-0.264	0.000
Rabié	$\beta_{distanciaRabie}$	-0.611	0.000
M10 Cerrillos	$\beta_{distanciaM10C}$	-0.317	0.000
Erbi Maipú 1	$\beta_{distanciaErbiM1}$	-0.749	0.000
M10 Maipú	$\beta_{distanciaM10M}$	-0.038	0.000
Erbi Peñaflo	$\beta_{distanciaErbiP}$	-0.362	0.000
R² ajustado = 0.998			
MAPE = 46%			

Fuente: Elaboración propia

Nuevamente el *recency* es la variable que más pesa al momento de determinar la frecuencia de visitas al supermercado, donde a mayor *recency* menor es la frecuencia. Debido a su alto coeficiente beta y siguiendo el mismo procedimiento que para la regresión lineal, se decide sacar esta variable con el fin de estudiar mejor el efecto de las otras dimensiones.

Tabla 8: Regresión geográfica ponderada para los días entre compras, sin Recency, local Maipú

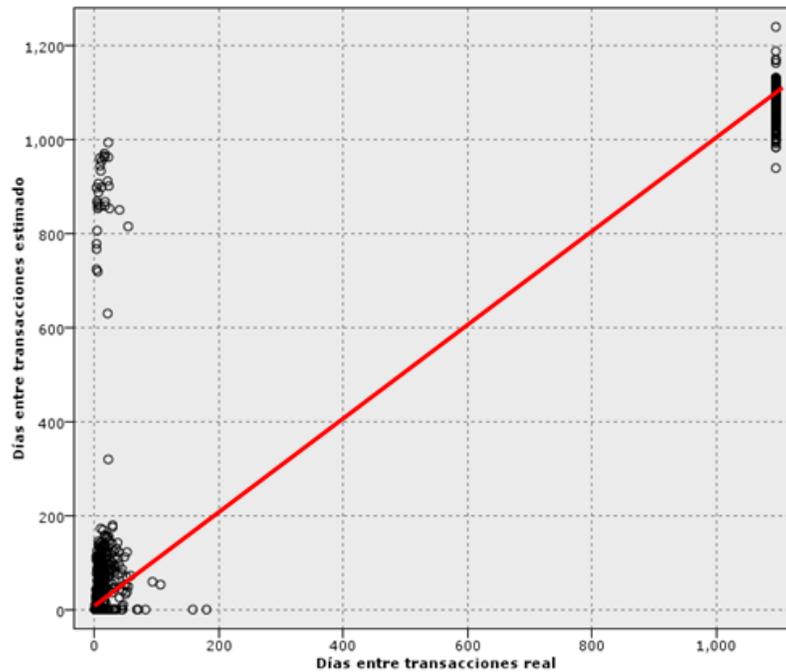
Variable	Coef.Beta	Valor	Sig.
Maipú	β_{maipu}	-0.639	0.000
Cerrillos	$\beta_{cerrillos}$	-0.194	0.000
Peñaflo	$\beta_{peñaflo}$	-0.079	0.000
Padre Hurtado	$\beta_{padre h}$	-0.193	0.000
Ticket promedio	β_{ticket}	-0.232	0.000
nº compras	$\beta_{ncompras}$	-0.191	0.000
Local Pajaritos	$\beta_{distanciaPajaritos}$	-0.061	0.000
Acuenta	$\beta_{distanciaAcuenta}$	-0.461	0.000
M10 Maipú 2	$\beta_{distanciaM10M2}$	-0.233	0.000
M10 CAM	$\beta_{distanciaM10CAM}$	-0.401	0.000
Erbi Cerrillos	$\beta_{distanciaErbiC}$	-0.207	0.000
Fruna	$\beta_{distanciaFruna}$	-0.493	0.000
R² ajustado = 0.927			
MAPE = 265%			

Fuente: Elaboración propia

Acerca de los locales de la competencia, hay 6 locales que producen algún efecto en la frecuencia de compra. Los locales con coeficiente beta negativo, son la competencia fuerte del mayorista, ya que a menor distancia de estos locales, mayor es el valor de “días entre compras” en la sucursal Maipú, es decir la gente va menos seguido.

Estudiando el ajuste, se obtiene lo siguiente:

Gráfico 18: Ajuste de la GWR para los días entre compras, local Maipú

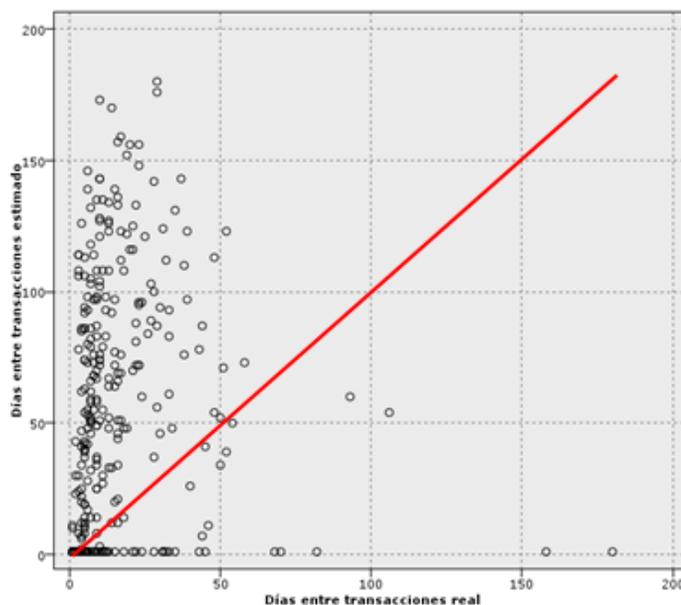


Fuente: Elaboración propia

Se observan principalmente dos grupos de puntos, uno cercano al cero donde la predicción parece subestimar y sobreestimar los valores y otro cercano al mil que es el correspondiente a las personas que no asisten al supermercado y que como puede observarse el modelo fue capaz de captarlos.

En el gráfico 19 al hacer un acercamiento a la predicción de las personas que sí asisten al supermercado se tiene que en la mayoría de los casos el modelo está subestimando la verdadera frecuencia, es decir tiende a predecir días entre compras mayores al real (menor asistencia).

Gráfico 19: Zoom del ajuste de la GWR para los días entre compras, local Maipú



Fuente: Elaboración propia

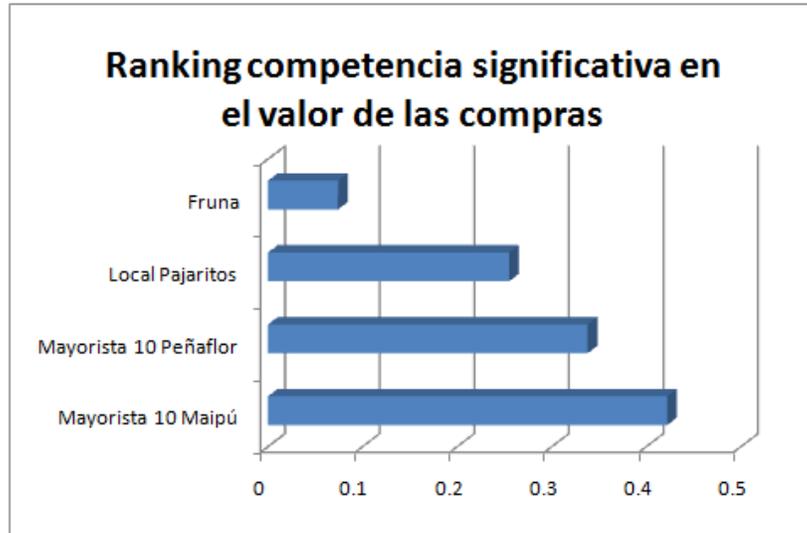
En resumen, y al igual que para la predicción del monto, el ajuste representando en el gráfico reciente, el coeficiente R^2 y el MAPE¹⁸, no muestran mayores diferencias con la regresión lineal, sin embargo, se destaca la información recabada acerca del efecto de 8 competencias respecto de la regresión lineal, que entregó sólo 3 locales significativos.

Competencia significativa para el local Maipú

A continuación se hace un resumen con los locales de la competencia que provocan algún efecto negativo en las ventas del mayorista, tanto en el monto como en la frecuencia de sus clientes.

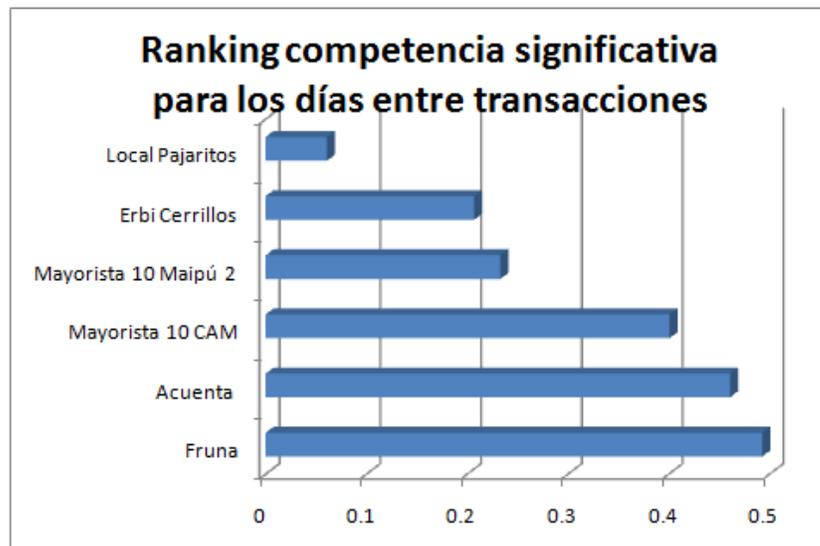
¹⁸Detalles acerca de los errores porcentuales absolutos en el Anexo D.

Gráfico 20: Ranking de competencia significativa para el monto, local Maipú



Fuente: Elaboración propia

Gráfico 21: Ranking de competencia significativa para los días entre compras, local Maipú

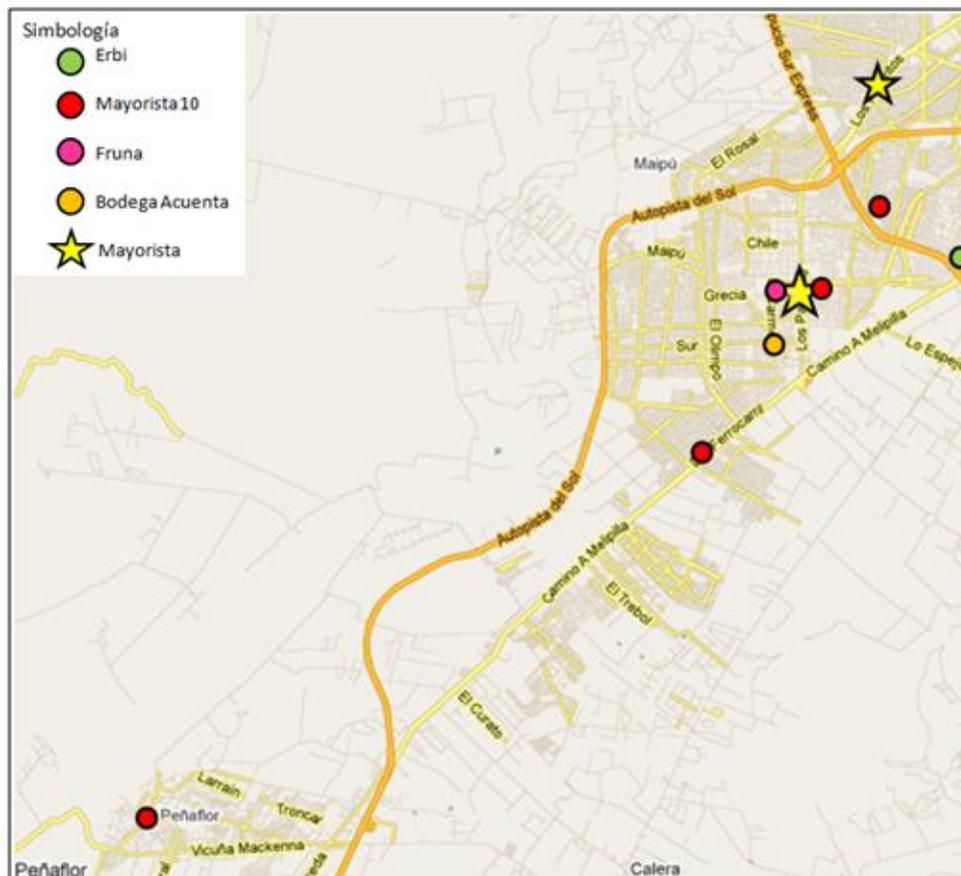


Fuente: Elaboración propia

Basado en esto, se puede obtener información geográfica relevante, acerca de cuáles son las áreas donde los clientes son menos fieles al supermercado en estudio. Con los datos anteriores, se dice que en las zonas de los locales que fueron definidos como competencia fuerte, hay clientes del mayorista que son poco fieles, ya que precisamente acá es dónde las ventas o la frecuencia de los clientes disminuyen debido a la existencia de la competencia.

Se revisa en los mapas que siguen la ubicación geográfica de los locales que resultaron ser significativos ya sea para el monto o la frecuencia. Se puede destacar que la mayor cantidad de locales de competencia relevante están muy cercanos al mayorista, excepto por sólo 1 que se ubica en la comuna de Peñaflores.

Figura 17: Competencia significativa para el local Maipú



Fuente: Elaboración propia

9.1.4 Barrios de interés y ubicación de la publicidad

Con la información obtenida a través de los *heat maps*, se mostrarán cuáles son los barrios que anidan clientes con comportamientos de compra similares, así como con la información entregada por las regresiones geográficas, se analizarán los barrios con mayor cantidad de competencia significativa.

Después de identificar los barrios de interés se propondrán acciones de publicidad para algunos de ellos.

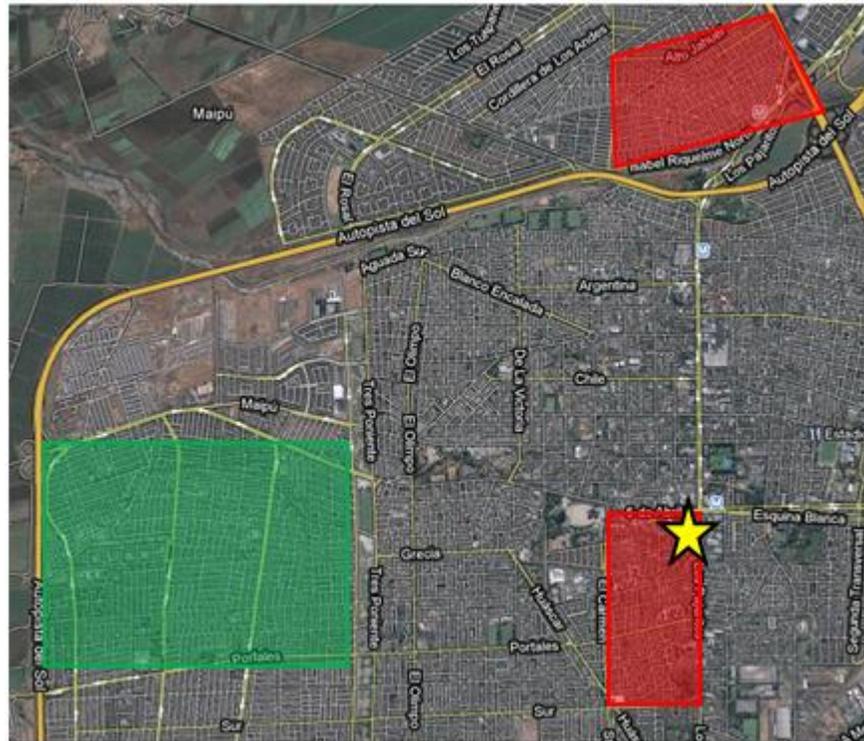
9.1.4.1 Barrios de interés

Se encontraron 3 barrios importantes en la comuna de Maipú, uno de éstos corresponde al cuadrante donde está ubicado el supermercado mayorista y donde además se encuentra la mayor cantidad de competencia significativa: 3 locales.

Por el lado poniente de la comuna y marcado en verde, están los clientes que tienen el mayor monto de compra en el supermercado, comprando cerca de \$2.1 MM anualmente. Estas personas en su mayoría, pertenecen a la Población San Luis y otras villas de similares características en la comuna de Maipú.

Finalmente, al norte del supermercado y marcado en rojo, se encuentran los barrios de Alto Jahuel y La Sinfonía, sus habitantes pertenecen al GSE ABC1 y C2¹⁹ y son los clientes que menos compran en el mayorista, adquiriendo casi la mitad menos que sus vecinos de la Población San Luis, con montos cercanos a los \$1.4 MM al año.

Figura 18: Barrios de interés para el local Maipú



Fuente: Elaboración propia

Cabe destacar que las diferencias entre el barrio de los buenos clientes contra el de los malos, es que además en el área de la Población San Luis existe al menos un 30% más de clientes comparado con un área de igual dimensión en el Alto Jahuel.

¹⁹ La información acerca del grupo socioeconómico por manzanas fue consultado en las estadísticas de Map City.

9.1.4.2 Acciones de publicidad

Como menciona Cliquet [7], las campañas de publicidad gastan enormes sumas de dinero con el fin de crear conciencia de marca, generar preferencias por un producto, invitar a las personas a comprar, etc. Estas campañas involucran el desarrollo de dos elementos: el contenido y la difusión del mensaje. Dentro de la divulgación de este mensaje, se cuenta la elección del medio correcto y llegar a las personas adecuadas. El Geomarketing, ayuda a la elección óptima de estas personas, a través del conocimiento de los clientes por la variable espacial. Específicamente, el Geomarketing de Zonas, permitiría identificar áreas de clientes con características singulares de acuerdo a su ubicación geográfica, tanto socio-demográficas como de comportamiento.

Publicidad con foco en los buenos barrios

Se comprobó que hay ciertas ubicaciones geográficas, específicamente barrios y villas, donde hay una tendencia de sus habitantes por comprar más. Sin embargo, emplazados en estos mismos lugares, existen vecinos que no se están comportando como lo hace el resto. Se estima que si las personas que compran bajo la mediana aumentaran en sólo un 5% su monto de compra, el mayorista vería aumentadas sus ventas en \$10MM al año, por cada barrio bueno. Por esta razón, se propone enfocar acciones de publicidad sobre los buenos barrios donde existirían personas alejadas del comportamiento de compra general y que podrían, dada su variable espacial, aumentar el monto.²⁰

Publicidad con foco en zonas de alta competencia

Con los resultados de las regresiones geográficas ponderadas, fue posible rescatar cuáles son los locales de la competencia que son realmente importantes para el mayorista, ya que le quitan ventas o visitas de sus clientes. Específicamente se encontró una zona, donde está la mayoría de los locales que resultaron significativos. Justamente por las pérdidas que éstos estarían ocasionando para el supermercado, se propone esta área como foco de publicidad estática, la que podría incorporar paneles con promociones en las afueras del mismo supermercado, como también la repartición de volantes o revistas a su alrededor.

²⁰ Detalle de la distribución de la compra entre un barrio bueno y uno malo, se puede ver en el Anexo E. Esta comparación se hizo entre el sector de la Población San Luis y el de Alto Jahuel.

9.2 Temuco

Con el fin de probar si acaso existen comportamientos de compra explicados por la ubicación geográfica para cualquier sucursal, se hace un estudio georeferenciado de los clientes del mayorista ubicado en Temuco. Se desean encontrar, al igual que para Maipú, barrios con buenos y malos clientes, de acuerdo al monto adquirido y/o su frecuencia de compra.

9.2.1 Datos

9.2.1.1 Datos disponibles

Se cuenta con los datos de 5.540 clientes del local Temuco, que han realizado compras entre Enero de 2008 y Diciembre de 2010.

9.2.1.2 Limpieza previa y datos finales

Se trabajará sólo con el 40% de los datos aproximadamente, debido a los siguientes filtros impuestos:

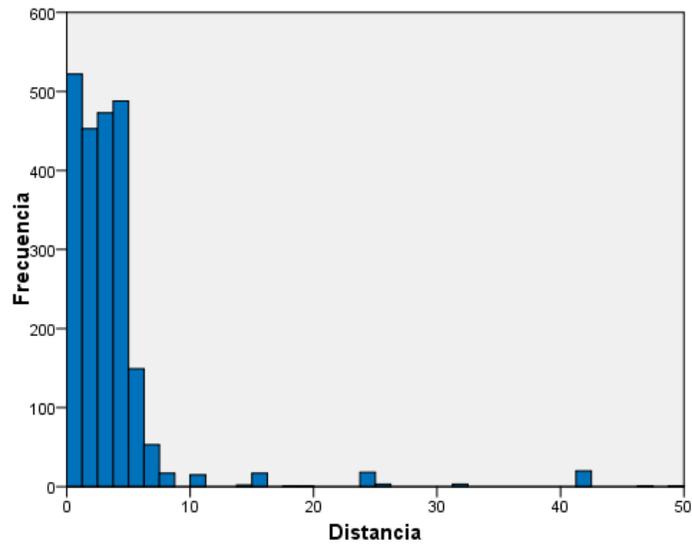
Tabla 9: Datos clientes de Temuco

TOTAL CLIENTES: 5.540	CLIENTES ANALIZADOS: 2.125 (38%)	Datos correctos: 2.125 (100%)
	CLIENTES FILTRADOS: 3.415 (62%)	Cientes sin coordenadas: 1.954 (57%) <i>Outliers</i> (distancias y frecuencias fuera del rango establecido): 1.461(43%)

Fuente: Elaboración propia

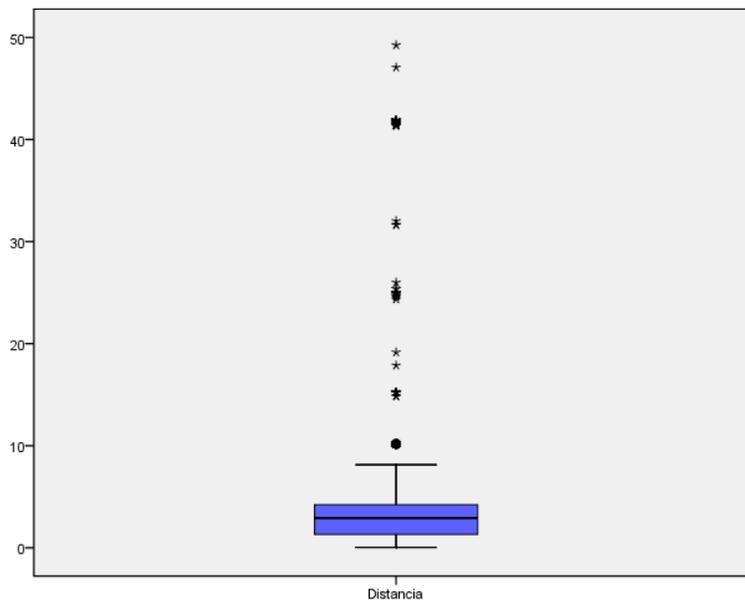
El mayor filtro, fue que para muchos clientes no se encontraron sus coordenadas de latitud y longitud, principalmente porque sus direcciones estaban mal escritas. Además, se trabajó con los clientes ubicados a menos de 10km. del local que resultaron ser una minoría tal como se observa en el histograma y *box plot* siguientes.

Gráfico 22: Histograma de las distancias de los clientes, local Temuco



Fuente: Elaboración propia

Gráfico 23: Box plot de las distancias de los clientes, local Temuco



Fuente: Elaboración propia

También se eliminaron los clientes que tenían sólo 1 visita al supermercado en los 3 años, pero se trabajó con todos los montos mayores a cero, sin eliminación de *outliers* al contrario de como se hizo en Maipú donde se estableció un monto mínimo de compra anual. En este caso se tendrá la ventaja de poder evidenciar mejor geográficamente a los clientes que compran más o menos que el promedio, pero se afectará el valor del MAPE que podría ser mayor debido a los *outliers* por monto presentes.

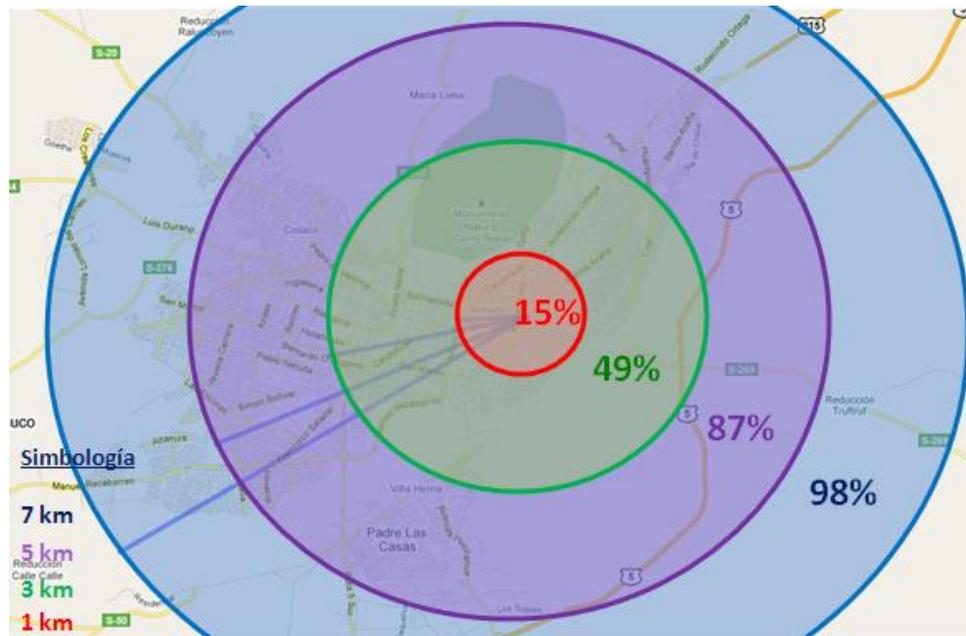
9.2.2 Análisis de la situación actual

Se analizan los clientes del supermercado Temuco en cuanto a su comuna de origen, distancia a la sala, monto y frecuencia de compra y su distribución, a través de *heat maps* y otros mapas que permiten entender la situación actual.

9.2.2.1 Distancia de los clientes al local Temuco

En promedio, los clientes del local Temuco, se encuentran a una distancia de 3 km. aproximadamente. Analizando qué porcentaje de las ventas es explicado por cada distancia, se obtienen resultados muy similares a los de Maipú, esto es que a menos de 3km. se explica casi la mitad de las ventas totales del supermercado y que a 1 km. a la redonda del local, este porcentaje alcanza el orden del 10%.

Figura 19: Porcentaje del monto de las compras explicado por la distancia, local Temuco



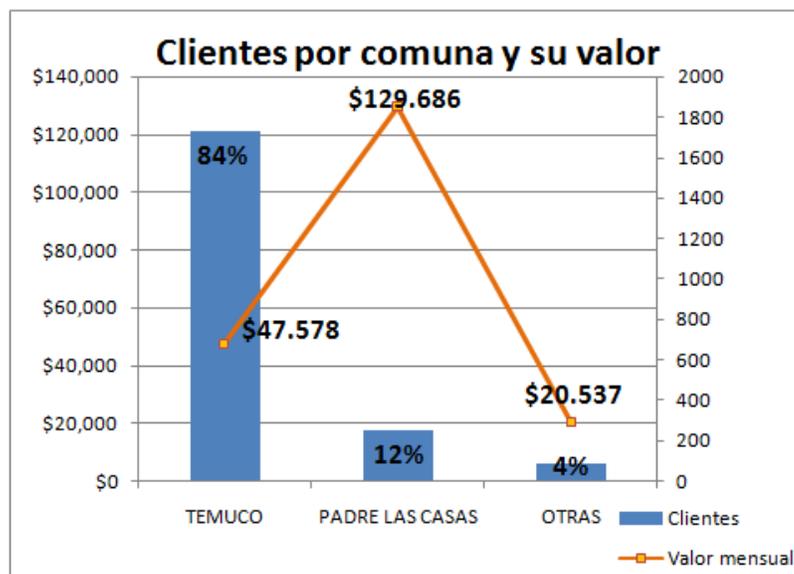
Fuente: Elaboración propia

Debido a que la superficie habitable de Temuco es menor que la de Maipú, casi el 100% de las ventas es explicado a una distancia menor, igual a 7km., contra 10 km. del local anterior.

9.2.2.2 Comunas de origen de los clientes y su valor

El local que se estudia se encuentra emplazado en la calle Balmaceda 1583 en la comuna de Temuco. Por este motivo, más del 80% de los clientes vive en esta localidad. Sin embargo, hay un 12% que proviene de la comuna Padre Las Casas, ubicado a 4km. del supermercado, por lo que este barrio será analizado como una comuna aparte para la estimación.

Gráfico 24: Origen de los clientes del local Temuco



Fuente: Elaboración propia

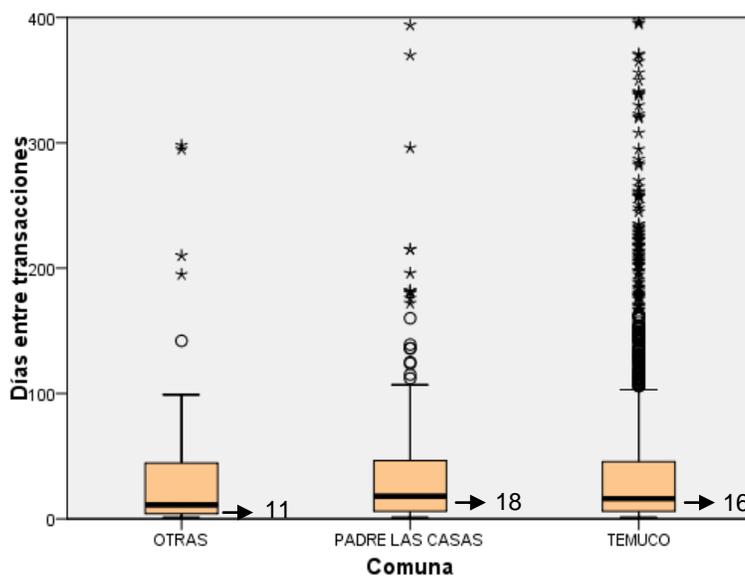
Especialmente la comunidad de Padre Las Casas es la que presenta un valor promedio de compra mucho superior a las otras comunas, inclusive Temuco que es donde está emplazado el supermercado, obteniendo casi el doble de la compra mensual promedio de los temuquenses.

9.2.2.3 Frecuencia y ticket promedio de los clientes

Se estudian ambas variables para las comunas de Temuco, Padre Las Casas y otras.

Acercas de la frecuencia, específicamente el indicador días entre compras, se observa que no hay grandes diferencias entre la asistencia de estas 3 agrupaciones de comunas, ocurriendo que los clientes asisten entre 2 y 3 veces al mes al supermercado aproximadamente. Llama la atención que las personas que asisten menos seguido a la sala son las que más compran en promedio, esto corresponde a los clientes de Padre Las Casas.

Gráfico 25: Días entre compras de los clientes del local Temuco²¹



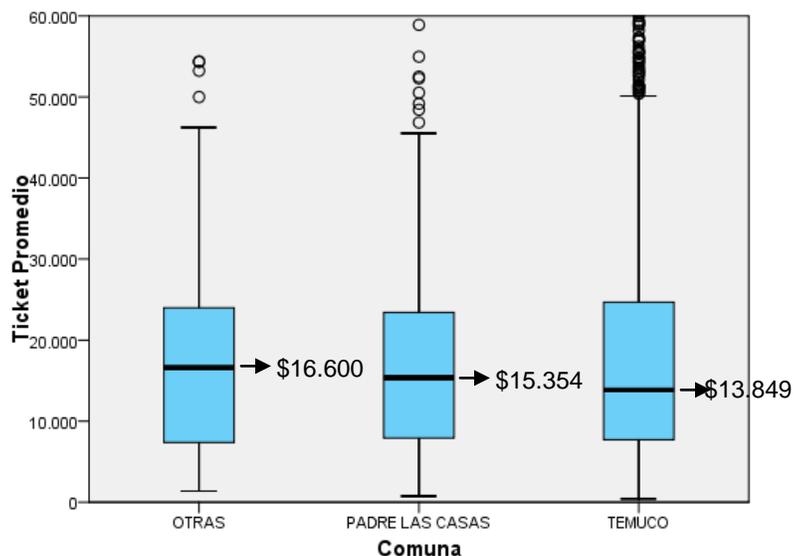
Fuente: Elaboración propia

Los habitantes de Temuco, pese a ser los más cercanos al supermercado, no son los que más asisten, contrario a lo que pasaba en Maipú, que eran precisamente ellos los que iban semanalmente al local. Los temuquenses en cambio, sólo van 2 meses al mes contra 3 visitas que tienen los clientes de otras comunas.

El ticket promedio de estas localidades tampoco difiere mucho y nuevamente se cumple que los habitantes de la comuna donde está el local, no son los mejores clientes. En este caso, tienen el ticket promedio más bajo igual a \$13.849, mientras que los clientes de otras comunas gastan \$16.600 en cada visita al supermercado.

²¹ Acercamiento a *box plot* de días entre compras, para una vista más general ir al Anexo B.

Gráfico 26: Ticket promedio de los clientes del local Temuco²²



Fuente: Elaboración propia

9.2.2.4 Heat maps del valor, los días entre compras y la densidad de clientes

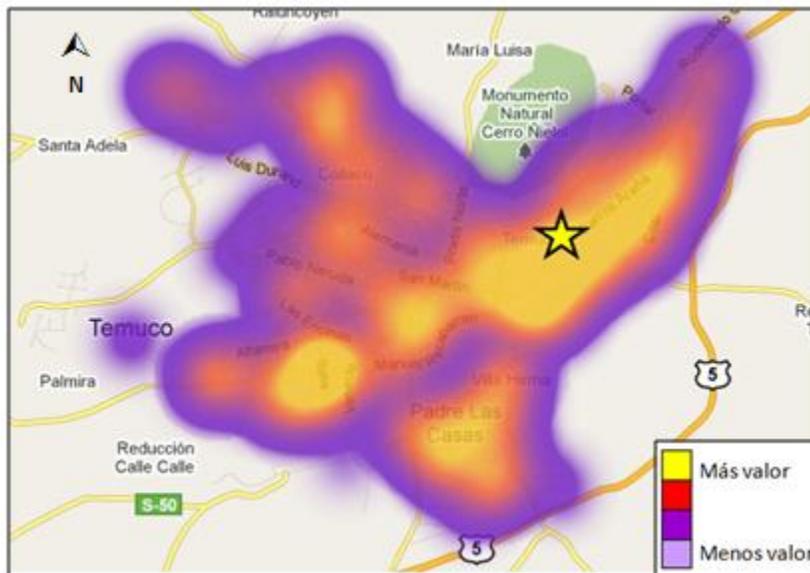
Con los datos transaccionales desde el año 2008 hasta el 2010, se representa mediante *heat maps* la nube de clientes del local Temuco, para su densidad, días entre compras y monto de compra.

Valor

El monto de los clientes del local Temuco presenta distintos matices por barrios. Las zonas con mayor intensidad en ventas son las cercanas al supermercado, por el lado sur, además de otros núcleos al sur-poniente de Temuco y en la comuna de Padre Las Casas. La zona más fría, es decir con menores montos de compras, son las del norponiente de Temuco. Más adelante en la sección 9.2.4.1, se analizará a qué barrios y qué características tienen los habitantes de esta zona.

²² Para ver vista más general ir a Anexo B.

Figura 20: Heat map del valor, local Temuco

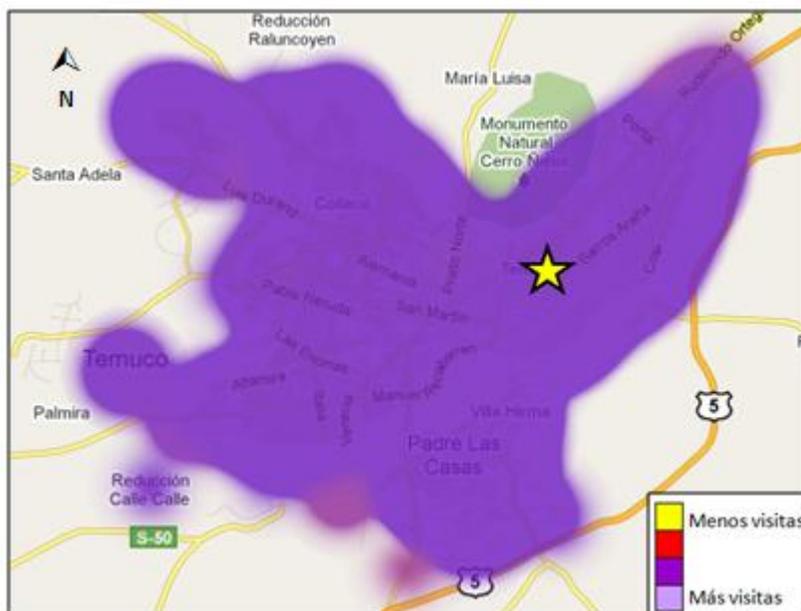


Fuente: Elaboración propia

Días entre compras

Tal como arrojó el *box plot*, no existen mayores diferencias entre los días entre compras de una localidad y otra. El color morado uniforme del siguiente *heat map*, indica que cada persona asiste en promedio la misma cantidad de veces al supermercado, que como se mencionó anteriormente es entre 2 y 3 veces al mes.

Figura 21: Heat map de los días entre compras, local Temuco



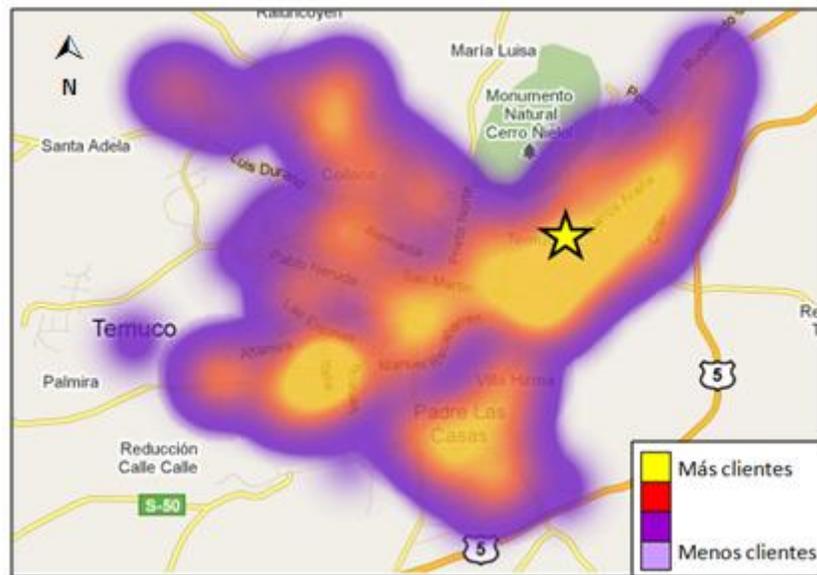
Fuente: Elaboración propia

Densidad de clientes

Para la densidad de clientes se observa que la distribución es más uniforme comparado con el monto de las compras. En general existen pocas discontinuidades, encontrándose que el color predominante es el amarillo, excepto en la zona delimitada por las calles Pablo Neruda y Las Encinas, además del barrio al norte y al poniente de la calle Prieto Norte que posee una tonalidad más fría, es decir de menores compras.

Destaca la zona caliente al sur del mayorista, correspondiente a Padre Las Casas, cuyos pobladores representan más de un 10% de los clientes de todo el local Temuco.

Figura 22: Heat map de la densidad de clientes, local Temuco



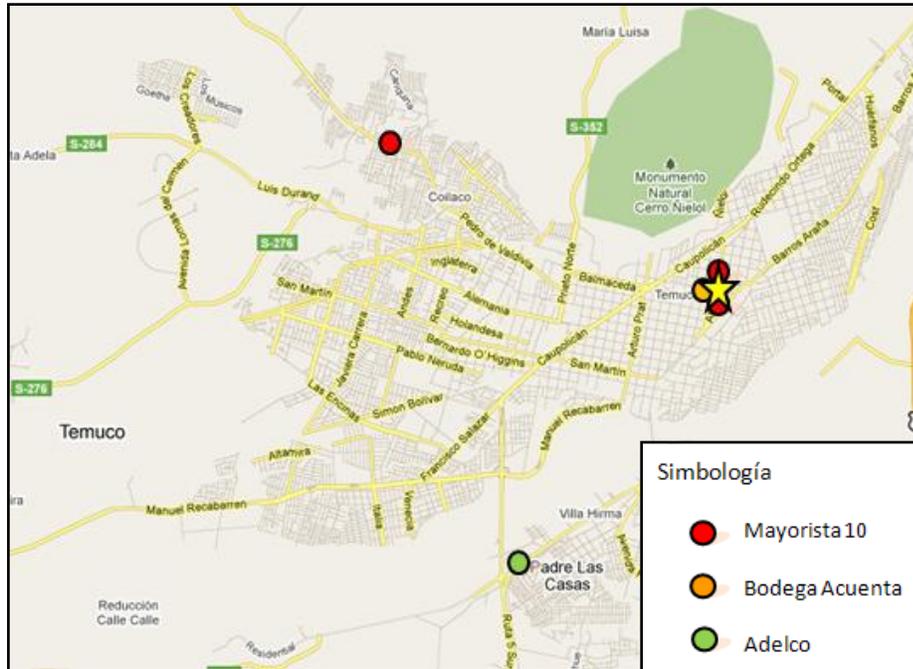
Fuente: Elaboración propia

9.2.2.5 Datos de la competencia

Debido a que Temuco presenta una superficie de terreno pequeña (10 km. en su parte más ancha), se considerará como competencia a todo local ubicado en esta ciudad, sin considerar a la distancia que se encuentre del mayorista.

Se encontraron 5 locales de la competencia, 3 Mayorista 10, un Adelco y una Bodega Acuenta, donde 3 de estos 5 establecimientos están a menos de 2 cuadras del supermercado de interés.

Figura 23: Competencia mayorista de local Temuco



Fuente: Elaboración propia

Figura 24: Acercamiento competencia para el local Temuco



Fuente: Elaboración propia

9.2.3 Modelos de estimación para el monto y la frecuencia

A través de regresiones geográficas ponderadas se estimó el monto para cada cliente del local Temuco. En esta ocasión no se correrán regresiones lineales, ya que se comprobó que el ajuste y el MAPE son prácticamente los mismos y más aun, las regresiones geográficas, permiten identificar más variables significativas que las lineales. Además, no serán estimados los días entre compras ya que no tienen una distribución ordenada a través del espacio. En el Anexo C se observa un *scatter plot* donde se comprueba que la distribución de la frecuencia es absolutamente al azar y no responde a parámetros explicados por la geografía ni a variables transaccionales.²³

9.2.3.1 Variables a incluir

Se incluirán las mismas variables ocupadas para el local Maipú, esto comprende las variables transaccionales, la ubicación geográfica del cliente y la distancia a cada local de la competencia.

El modelo de estimación para el monto de la compra, queda definido como sigue²⁴:

Ecuación 9: Modelo de estimación para el monto para un cliente i , local Temuco

$$\begin{aligned} \text{Monto}_i = & \beta_{\text{temuco } i} \cdot \text{temuco} + \beta_{\text{pcasas } i} \cdot \text{PadreLasCasas} \\ & + \beta_{\text{diasentre } i} \cdot \text{días entre compras} \\ & + \beta_{\text{ncompras } i} \cdot \text{cantidad de compras} + \beta_{\text{recency } i} \cdot \text{recency} \\ & + \beta_{\text{ticket } i} \cdot \text{ticket promedio} + \beta_{\text{cercania } i} \cdot \text{cercania} \\ & + \beta_{\text{distancia } i} \cdot \text{distancia a local temuco} \\ & + \beta_{\text{distanciacompetenciaj } i} \cdot \text{distancia a competencia } j \\ & + \beta_{\text{Qcomp } i} \cdot \text{cantidad de competencias cercanas} \end{aligned}$$

$$\forall i = 1 \dots 2.125$$

Las primeras dos variables corresponden a binarias que indican si acaso el cliente vive en alguna de esas dos comunas o en otra distinta (representado con valor cero para ambas *dummies*). El resto de las variables tiene la misma interpretación hecha con anterioridad en la sección 9.1.3.1.

²³ Fue probada una regresión geográfica para estimar los días entre compras, con variables geográficas y transaccionales, pero el coeficiente R^2 sólo alcanzó un valor del orden del 0.01

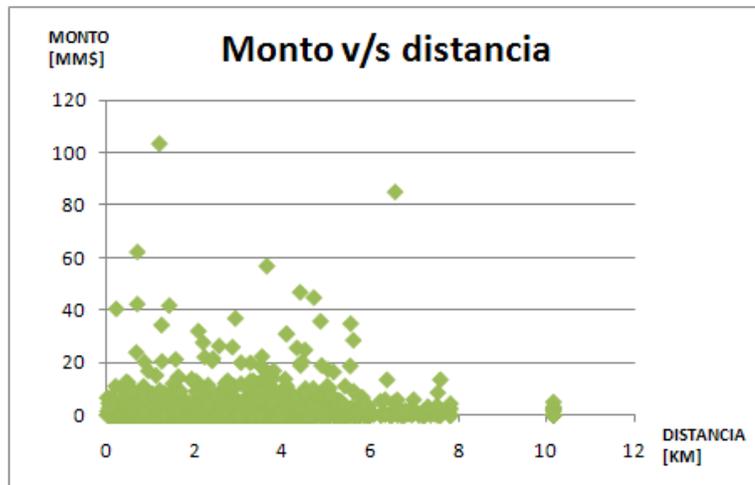
²⁴ La ubicación geográfica va implícita en la fórmula (coordenadas de latitud y longitud).

9.2.3.2 Análisis descriptivo de las variables

Se estudia el monto en función de algunas variables independientes de la regresión, para ver si existe alguna relación entre ellas que explique el comportamiento de compra de los clientes.

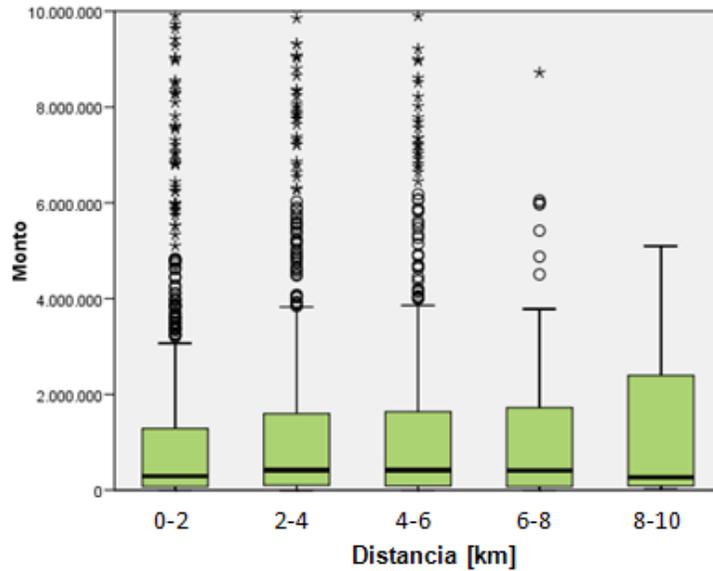
Al analizar el monto en función de la distancia, se encuentra que a mayor cercanía al local las personas no compran un monto mayor, esto viene a contradecir la tendencia de Maipú. Es de esperar por lo tanto, que esta variable no sea significativa en la regresión, lo que no quita que existan comportamientos de compra explicados por la geografía, ya que bien podría pasar que dos barrios distintos, pero ubicados a igual distancia del local, compren montos totalmente distintos.

Gráfico 27: Monto y distancia en *scatter plot*, local Temuco



Fuente: Elaboración propia

Gráfico 28: Monto y distancia en *box plot*, local Temuco

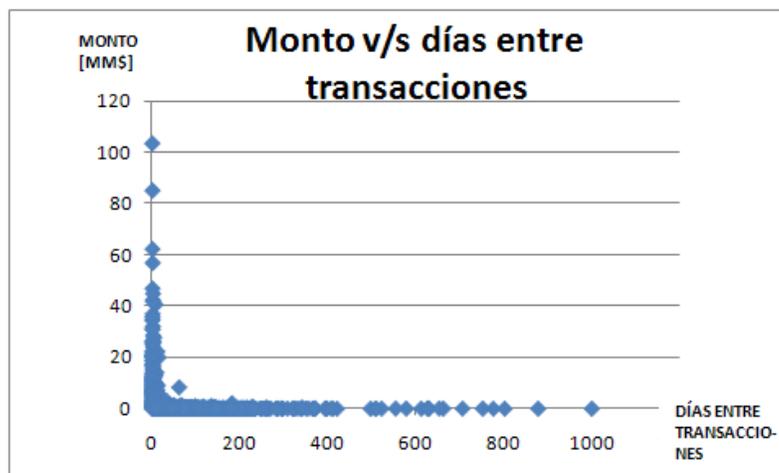


Fuente: Elaboración propia

Tal como se aprecia en el *scatter plot* y *box plot* para el monto en función de la distancia, no existe una tendencia de aumento en la compra a medida que el cliente viva más cerca del supermercado.

En cambio, los días entre compras sí explican la variable monto, cumpliéndose que para las personas que van más seguido a Temuco, es decir tienen un días entre compras menor, los montos de compra son mayores, mientras que si asisten menos gastan menos. La tendencia se observa en el siguiente *scatter plot*.

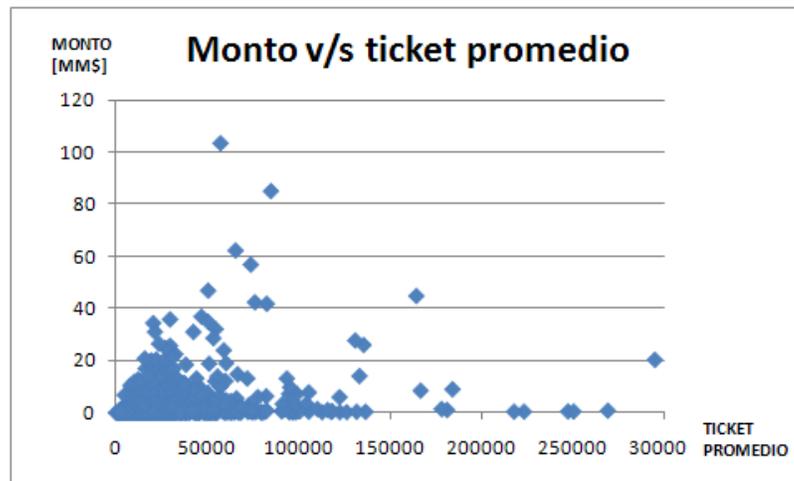
Gráfico 29: Monto y días entre compras en *scatter plot*, local Temuco



Fuente: Elaboración propia

Finalmente para el monto explicado por el ticket promedio, se repite la misma tendencia del local Maipú, esto es que ticket promedios bajos son incompatibles con montos altos, no cumpliéndose lo contrario, es decir hay personas que algunas veces compraron mucho, pero debido a que asistieron poco esto les dio un monto total bajo.

Gráfico 30: Monto y ticket promedio en *scatter plot*, local Temuco



Fuente: Elaboración propia

9.2.3.3 Regresión lineal

Estimación del monto

Al igual que para el local de Maipú se realiza un método de análisis global (regresión lineal) para compararlo con los resultados de la regresión geográfica ponderada.

Se calibran los coeficientes betas con un 80% de los datos, mientras que el MAPE es calculado sobre el 20% de la población restante.

Tabla 10: Regresión lineal para el monto, local Temuco

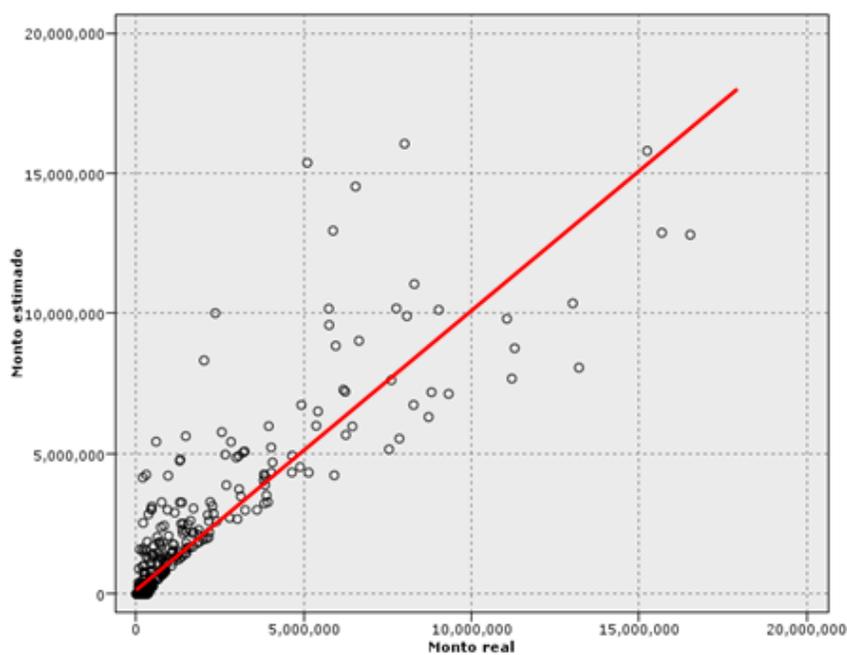
Variable	Coef. Beta	Valor	Sig.
nº compras	$\beta_{ncompras}$	0.732	0.000
Ticket promedio	β_{ticket}	0.297	0.000
R² ajustado = 0.628 MAPE=111%			

Fuente: Elaboración propia

Sólo dos variables son significativas, permitiendo explicar más de un 60% de la varianza. El n° de compras y el ticket promedio tienen una relación dependiente con el monto, donde un mayor n° de estas 2 variables indica que las personas adquieren montos mayores en la sala.

El MAPE resulta del orden del 100%, observándose que el modelo tienda a errar más para montos superiores a \$5MM, donde puede subestimar o sobreestimar indistintamente.

Gráfico 31: Ajuste regresión lineal para el monto, local Temuco



Fuente: Elaboración propia

9.2.3.4 Regresión geográfica ponderada

Se desarrolla el modelo de regresión geográfica ponderada para estimar el monto de compra agregado entre los años 2008 y 2010 para el local Temuco. Antes de calcular los parámetros de las variables dependientes, se destaca que fue necesario calcular el ancho de barrios o *bandwidth* para los datos de Temuco, ya que el parámetro es único para cada conjunto de datos y por lo tanto es incorrecto usar el valor obtenido para Maipú.

Con el método que minimiza la sumatoria de los residuos al cuadrado, el *bandwidth* obtenido para el monto de Temuco es el siguiente:

$$\text{Bandwidth} = 2.978718 \text{ km.}$$

Estimación del monto

Se estimó el monto calculando los coeficientes betas con los datos del año 2008 y 2009 y calculando el MAPE sobre las compras del 2010. Los resultados se muestran en la tabla 11.

Tabla 11: Regresión geográfica ponderada para el monto, local Temuco

Variable	Coef. Beta	Valor	Sig.
Padre Las Casas	β_{PCasas}	0.060	0.003
nº compras	$\beta_{ncompras}$	0.732	0.000
Ticket promedio	β_{ticket}	0.250	0.000
Adelco	β_{Adelco}	0.274	0.000
Acuenta	$\beta_{Acuenta}$	0.941	0.000
R² ajustado = 0.597 MAPE=112%			

Fuente: Elaboración propia

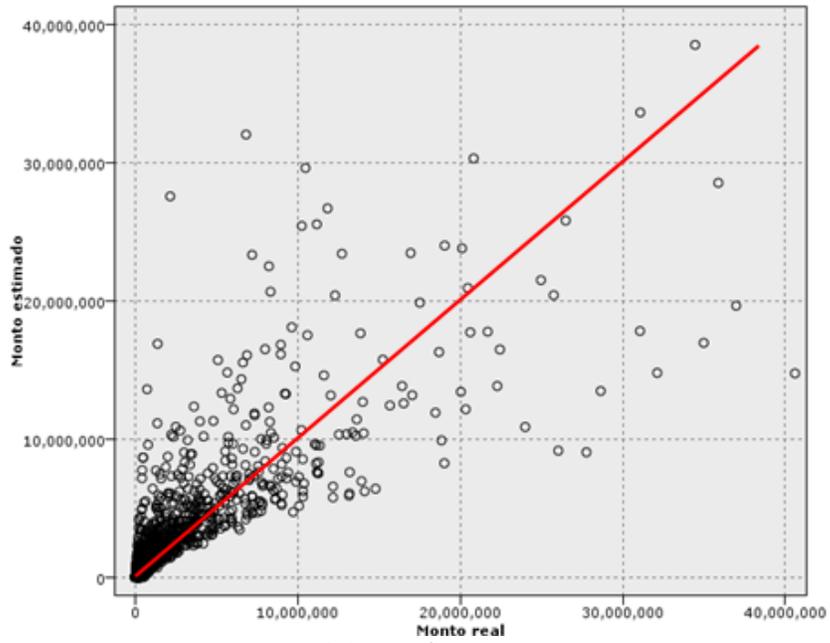
Se obtiene un buen ajuste cercano al 0.6, con un MAPE²⁵ que bordea el 100%, al igual que para la regresión lineal. Sin embargo, la GWR es capaz de captar dos locales de la competencia que son significativos, teniendo uno de ellos un coeficiente beta del nivel del 0.9 que implica una gran explicación del monto adquirido en el mayorista. Este local es una Bodega Acuenta ubicada a sólo unos pasos del Local Temuco. Esta sala, de acuerdo a los resultados recién obtenidos sería la variable que explica mejor el monto de los temuquenses, valor que resultó ser el más bajo de las dos comunas con más clientes.

La otra competencia significativa es Adelco, ubicada en la comuna de Padre Las Casas, localidad donde residen los mejores clientes del mayorista, en cuanto a monto mensual adquirido. Este local sería el segundo actor importante que repercute en las ventas, por lo que estos dos emplazamientos de competencia son sumamente importantes de analizar con el fin de evitar disminuciones de ingresos por compra.

Sobre el ajuste de la GWR, en el gráfico 32 se ve que para montos mayores a \$15MM el modelo subestima las compras, mientras que para la gente que compra menos a \$10MM la predicción señala que esas personas deberían comprar más de lo que lo hacen actualmente.

²⁵ Para ver detalles de la distribución del MAPE, ir a Anexo D.

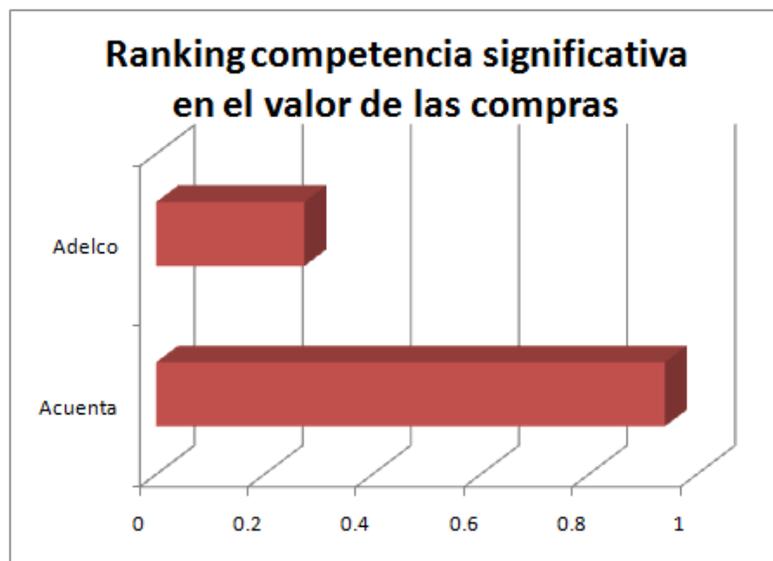
Gráfico 32: Ajuste de la GWR para el monto, local Temuco



Competencia significativa para el local Temuco

Según los resultados obtenidos en la GWR se hace el listado con la competencia fuerte para el mayorista y se representa en un mapa la ubicación de éstas:

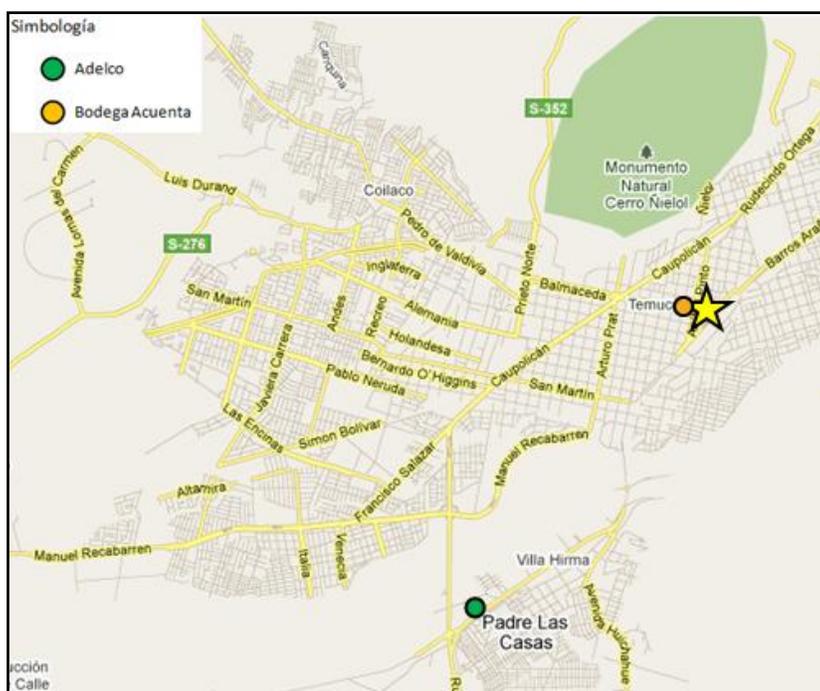
Gráfico 33: Ranking de competencia significativa, local Temuco



Fuente: Elaboración propia

Se observa donde están ubicados los locales que resultaron ser significativos. En esta oportunidad no se presentaron zonas con acumulación de competencias, sino que sólo se presentaron 2 locales aislados. De los 3 locales que estaban a tan sólo unos metros del mayorista, sólo uno trascendió, correspondiente a la cadena Acuenta. El supermercado Adelco en cambio, se ubica a una distancia mucho mayor de la sala, ya que está emplazado en la comuna de Padre Las Casas.

Figura 25: Competencia significativa para el local Temuco



Fuente: Elaboración propia

9.2.4 Barrios de interés y ubicación de la publicidad

Basado en las ubicaciones de la competencia fuerte y los barrios donde están los clientes que compran más y menos que el promedio de los clientes, se identifican las áreas geográficas seguidas a continuación.

9.2.4.1 Barrios de interés

Se encontraron 4 zonas relevantes, dos debido a la presencia de competidores importantes y otras dos porque hay clientes que tienen comportamientos de compra en común respecto al monto que adquieren en el mayorista.

El área contigua al supermercado Temuco, corresponde a la zona donde está ubicada la competencia Acuenta, zona con clientes que comprarían menos debido a que este local les queda cerca.

La zona al sur en Padre Las Casas, es donde se encuentra la competencia Adelco, que tiene el mismo efecto en los clientes que el local Acuenta anterior.

Figura 26: Áreas de competencia fuerte para el local Temuco



Fuente: Elaboración propia

La zona marcada en verde de la figura que sigue, área ubicada al sur del local Temuco, corresponden a los asentamientos con los clientes que compran montos mayores, esto fue apreciado también en el *heat map* del monto donde se presentó una gran zona caliente al sur del local, mientras que hacia el lado poniente aparecieron áreas más frías. Las personas de la zona verde compran casi un 40%²⁶ más que los clientes de la peor zona, visualizada en rojo en la figura 26.

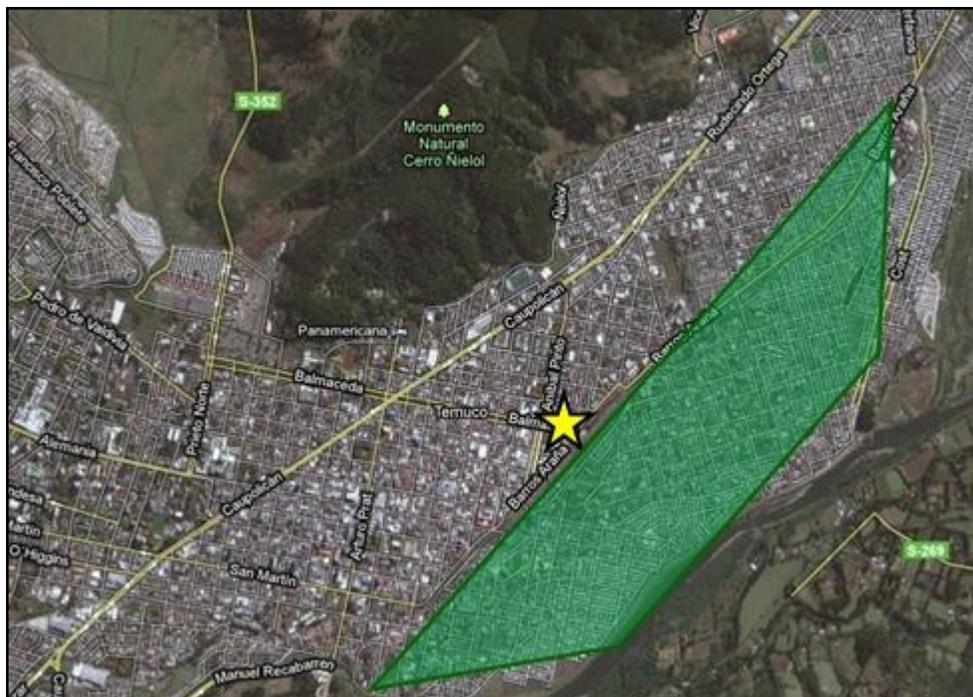
Las personas que viven en el sector en verde, pertenecen a diferentes villas y poblaciones con habitantes que en su mayoría pertenecen al GSE D y E. En cambio, los lugares donde viven los clientes de la zona roja, corresponden a los barrios Inglés, Los Conquistadores y las Lomas de Mirasur, sectores de la clase acomodada de Temuco. En estos últimos barrios, además de ocurrir que las personas compran menos,

²⁶ Ir a Anexo E para más detalles.

también hay menor cantidad de clientes, representando éstos sólo 1/3 de las personas de los sectores con los mejores clientes²⁷.

En el Anexo C se puede ver el *scatter plot* con todos los clientes del mayorista y el monto que han adquirido desde el año 2008 al 2010. Se puede comprobar que en efecto la concentración y monto mayor de los clientes al sur del local es notoria contrastado con el lado poniente de la ciudad.

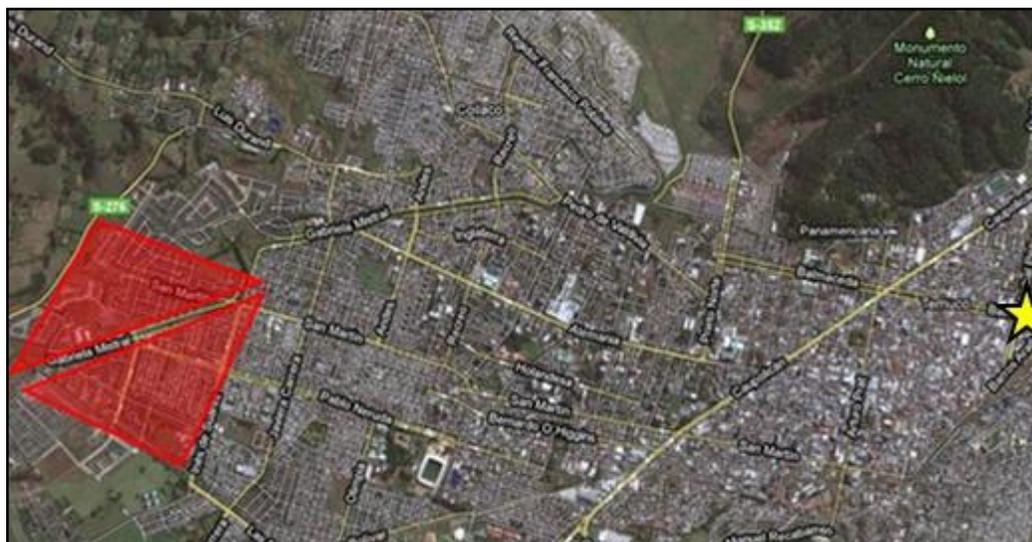
Figura 27: Áreas con clientes que compran más en local Temuco



Fuente: Elaboración propia

²⁷ El cálculo fue hecho en un área de igual tamaño ubicada en la zona “buena” y otra en la zona “mala”. Los resultados fueron que en la zona buena viven 356 clientes y en la mala sólo 112.

Figura 28: Áreas con clientes que compran menos en local Temuco



Fuente: Elaboración propia

9.3 Discusión y comparación resultados local Maipú y Temuco

Comportamientos de compra pudieron ser explicados por la geografía en ambos locales analizados. Se encontró que el monto adquirido va directamente ligado con el barrio de residencia de los clientes, donde a menor GSE mayor era la cantidad comprada. Resultado alineado con la estrategia del supermercado mayorista que va dirigido principalmente a personas del segmento C3 y D, poseedores de pequeños negocios como almacenes, bazares, panaderías, entre otros.

Pese a que el monto sí fue explicado por las variables geográficas en el local de Maipú y Temuco, se evidenció que ni la distancia ni la ubicación geográfica definieron el indicador “días entre compras” en la sala. En el caso de Maipú el modelo de regresión geográfica obtuvo un buen ajuste con un R^2 de 0.9 pero con un MAPE cercano al 300% para la regresión sin *recency*, mientras que para el caso de Temuco la regresión geográfica ponderada no alcanzó a obtener el nivel de 0.1 en su coeficiente R^2 .

Las diferencias en los ajustes obtenidos se deben a la variación en el tratamiento de los datos que se hizo en cada caso. Mientras que para Maipú se incluyó a los no-clientes (clientes que compran en otras sedes pero no en Maipú), en Temuco sólo se trabajó con los clientes efectivos. Esto produjo que para el caso de Maipú el modelo predijera muy bien a quienes no eran clientes (frecuencia nula en el supermercado), pero errando bastante en quienes sí asistían a la sala y por lo tanto obteniendo un error MAPE extremadamente elevado. En cambio en Temuco sólo se trabajó con los clientes reales del supermercado, afirmándose que el modelo se equivoca muchísimo cuando se trata de predecir con cuanta periodicidad asiste una persona a hacer las compras.

Por otro lado, una segunda diferencia importante detectada en los modelos entre ambas sedes, está relacionada con la estimación del monto. Pese a que en ambos casos se obtuvo buenos ajustes (entre el 0.6 y 0.8 de R^2 para Maipú y Temuco respectivamente), en el caso de la sede del sur el MAPE fue mucho mayor, superando el 100%, mientras que en Maipú bordeó el 60%. Esto nuevamente se debe a la calidad de los datos. En el caso de Maipú se hizo una mejor limpieza de los *outliers*, procurando sacar a los clientes con compras muy esporádicas (imponiendo un mínimo de \$10.000 mensual). En cambio en Temuco se dejó a todos los clientes, independiente hubiesen comprado sólo \$500 o \$3MM en un plazo de 3 años.

Esto provocó que se perdiera precisión en la estimación en Temuco, pero se obtuvo como ventaja poder apreciar mejor las diferencias en los *heat maps* para detectar donde están las personas que compran más o menos.

En resumen la limpieza de *outliers* e incorporación de casos positivos y negativos (clientes y no clientes) a la data jugaron un rol fundamental en la obtención de buenos ajustes y bajos errores en el modelo de regresión. Acerca de los resultados para ambos locales se afirma que el monto de la compra aparece sumamente ligado a la geografía o barrio de residencia de una persona, mientras que la frecuencia de las compras es un parámetro mucho más difícil de detectar, donde la mayoría de las personas van entre 2 y 4 veces al mes sin importar que tan cerca o lejos vivan.

10 Conclusiones

El Geomarketing es una rama del marketing que incorpora elementos de la geografía para ayudar en la toma de decisiones de negocios, ya sea desde la elección de dónde colocar una tienda nueva, hasta la optimización de la ubicación de cada una de las partes de la cadena de suministro. En esta oportunidad se utilizó el Geomarketing para complementar el conocimiento de clientes a través de su variable espacial. El objetivo era modelar la relación que existe entre la demanda de un cliente de un supermercado mayorista con su ubicación y distancia a la tienda. Específicamente, se buscó estimar el monto y la frecuencia de los clientes en un supermercado, con el fin de identificar cuáles son las áreas, que dada su variable geográfica, admiten una oportunidad de crecimiento en ventas, así como también identificar cuáles son los locales de la competencia que más afectan el desempeño de las ventas del mayorista.

Para las estimaciones se ocupó un modelo de análisis espacial local llamado regresión geográfica ponderada (GWR) y se compararon los resultados con un modelo de enfoque global como una regresión lineal (LR). Se encontró que en general el ajuste, representado por el coeficiente de determinación y el MAPE, son prácticamente iguales para ambos métodos, con R^2 superiores a 0.6 y MAPEs bordeando el 50%. Sin embargo, la GWR resultó mucho más efectiva encontrando los locales de la competencia que le quitan ventas o visitas al supermercado analizado, identificándose 8 locales significativos contra 3 que encontró el método lineal, para una misma sucursal.

La primera conclusión al respecto indica que la competencia tiende a disminuir más el tiempo entre visitas en otro supermercado, que el monto de la compra. Es decir, el hecho que haya un supermercado de la competencia cerca, provocaría con mayor probabilidad que esa persona aleje más sus visitas y como consecuencia deje de venir a esta sala, que en vez que dado que ya vino, compre menos. Más aun, se puede considerar que la disminución del monto de la compra es consecuencia de que asistió menos al supermercado, ya que se encontró que el nº de visitas al supermercado y su frecuencia tienen una correlación positiva con el monto adquirido.

Acerca de las variaciones en el monto de la compra, se obtuvieron buenos ajustes y bajos errores en modelos de regresión con variables geográficas (R^2 cercano al 0.8 y MAPE del 50% en la mejor de las salas). Permitiendo conocer cuánto compraría una persona en el supermercado mayorista con sólo saber en qué barrio vive y donde están las competencias a su alrededor. Para barrios de GSE más bajos las compras fueron mayores que para barrios ABC1 y C2 donde sus habitantes compraron incluso casi un 50% menos que la cantidad de sus otros vecinos.

Es decir, el monto estaría altamente correlacionado con el par ordenado (latitud, longitud) de cada cliente, pero no así con la distancia. No se cumplió que para hogares más cercanos a la sala las personas compraran más. Por lo tanto, análisis espaciales globales del tipo “todas las personas a 1km. a la redonda comprarán \$x, no se cumplieron“, haciendo que la distancia no tenga repercusión en la demanda, no así la ubicación como ya se mencionó anteriormente.

Respecto a la frecuencia y su relación con la distancia y la ubicación geográfica de cada cliente, no se encontraron dependencias significativas. Es decir las personas asistirían con una regularidad que varía desde la visita semanal hasta la bimensual, sin importar si están cerca o lejos de la sala. Más aun, la frecuencia no respondió a patrones geográficos como sí ocurrió con el monto. Sólo en el local de Maipú se cumplió que los habitantes de esta misma comuna asistían con una regularidad levemente mayor que los clientes de otras comunas: 4 veces al mes versus 3 visitas mensuales de los otros consumidores, sin embargo la diferencia no fue lo suficientemente relevante como para que la variable distancia haya resultado significativa.

Cabe destacar que pese a que se encontraron barrios específicos donde existen concentraciones de personas que compran montos más elevados que sus pares, dentro de estos mismos vecindarios existen clientes que adquieren cantidades considerablemente menores que el resto, es decir barrios que permitirían una oportunidad de crecimiento en ventas dado el comportamiento promedio de sus habitantes, por esta razón se propusieron acciones de publicidad para estas zonas, sugiriéndose como línea de investigación futura, medir el efecto en las ventas que provocaría el uso de publicidad focalizada sobre estas localidades.

Para este mismo fin y como trabajo futuro, se plantea el uso de promociones sobre estos clientes con el fin de estudiar si esta acción acaso podría igualar las compras de ellos con las de sus pares que adquieren una mayor cantidad. Además se sugiere incorporar la variable de GSE a las estimaciones, para comprobar si en efecto la pertenencia a estos segmentos condiciona el gasto de las personas en el supermercado. También se plantea la posibilidad de extender el trabajo de georeferenciación a otras áreas de estudio relacionadas con surtido de tiendas, a través de análisis de canasta por manzana censal o por barrio de interés, se puede identificar cuáles son los productos más consumidos por cierta localidad y por lo tanto de qué productos conviene abastecer más o menos a la tienda y cuáles productos nuevos podrían venderse basado en las compras hechas en tiendas con ubicaciones geográficas de las mismas características que ésta.

Finalmente, aprovechando el amplio desarrollo que han tenido las tecnologías de información para georeferenciación, como por ejemplo los GPS en la telefonía celular, podría utilizarse la información brindada por programas como *Foursquare*²⁸, que registran los lugares que ha visitado una persona, para estudiar el flujo de clientes en cierta tienda y con esto poder establecer temporadas altas, horarios peaks, bloques con menor cantidad de gente, etc. con el fin de regular mejor la dotación y *scheduling* de los trabajadores.

²⁸ Aplicación para celulares que permite a sus usuarios registrarse en los lugares que visitaron, permitiendo entregar recomendaciones y ver quiénes más estuvieron o están ahí y que opiniones han dejado. A través de su página web se puede consultar por lugares específicos y saber quiénes se registraron cuándo y dónde.

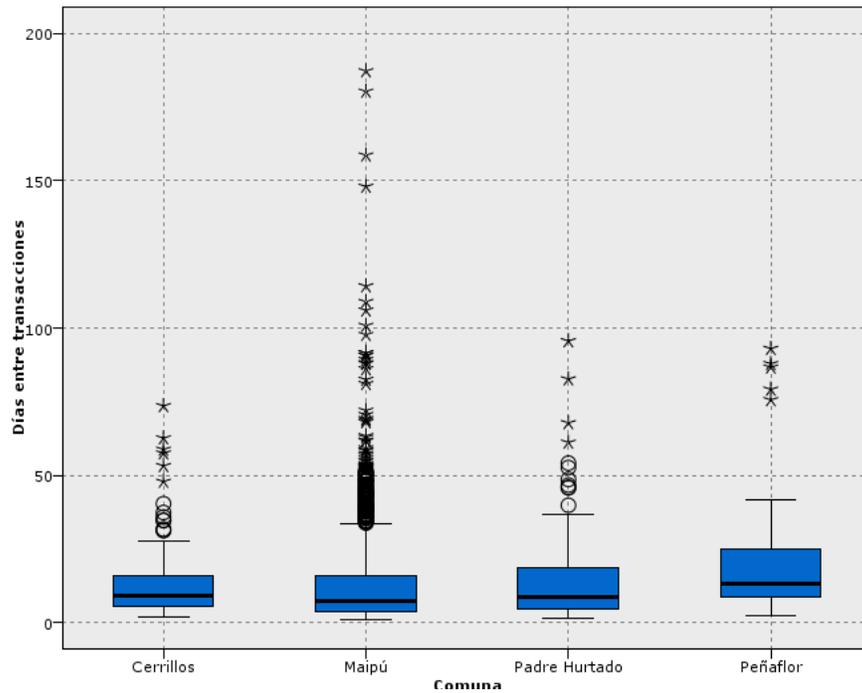
11 Bibliografía

- [1] ALBORNOZ, M. 2007. Modelamiento de la elección de supermercados en el gran Santiago. Memoria para Ingeniero Civil Industrial. Santiago, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.
- [2] CERET. 2009. El boom del retail en Chile. [en línea]. <<http://www.ceret.cl/noticias/el-boom-del-retail-en-chile/>>. [consulta: 3 marzo 2011].
- [3] CHASCO, M. 1997. Modelos de determinación de áreas de mercado del comercio al por menor, Madrid.
- [4] CHASCO, M. [199-]. Estimación de un modelo logit y aplicación a la determinación de áreas de mercado. Comparación con el modelo de Huff, Madrid.
- [5] CHASCO, M. [200-]. El geomarketing y la distribución comercial, Madrid.
- [6] CHASCO, P. 2000. Modelos de gravitación comercial: una aplicación al anuario comercial de España, Madrid.
- [7] CLIQUET, G. 2006. Geomarketing: Methods and strategies in spatial marketing. ISTE LTD, London.
- [8] DATA POINTED. 2009. Distance to McDonald's. [en línea]. <<http://www.datapointed.net/visualizations/maps/distance-to-nearest-mcdonalds/>>. [consulta: 21 septiembre 2011].
- [9] DOCUMENTING REALITY. 2011. [en línea]. <<http://www.documentingreality.com/forum/f181/2011-t-hoku-earthquake-tsunami-japan-11-march-2011-a-77259/>>. [consulta: 15 julio 2011].
- [10] ESRI ESPAÑA. 2010. ¿Qué es un SIG?. [en línea]. <<http://www.esri.es/es/formacion/que-es-un-sig-/>>. [consulta: 15 febrero 2011].
- [11] FOTHERINGHAM, S. BRUNSDON, C. y CHARLTON, M. 2002. Geographically weighted regression: the analysis of spatially varying relationships. John Wiley&Sons, LTD.
- [12] GEOMARKETING ÁREAS de influencia (I) Modelos gravitacionales. 2008. [en línea]. <<http://geomarketingspain.blogspot.com/2008/05/reas-de-influencia-i-modelos.html>>. [consulta: 15 noviembre 2010].
- [13] GOOGLE MAPS. 2011. [en línea]. < <http://maps.google.com/>>. [consulta: 15 enero 2011].

- [14] INE. 2011. Índice de ventas de supermercados. [en línea]. <http://www.ine.cl/canales/chile_estadistico/estadisticas_economicas/supermercados/series_estadisticas/series_estadisticas.php>. [consulta: 20 septiembre 2011].
- [15] MAP CITY. 2011. [en línea]. <www.mapcity.cl>. [consulta: 1 marzo 2011].
- [16] MASHINI, D. 2009. Geomarketing: guía de consumidores urbanos. [en línea]. <<http://www.plataformaurbana.cl/archive/2009/09/04/geomarketing-guia-de-consumidores-urbanos/>>. [consulta: 15 enero 2011].
- [17] MERCADO MAYORISTA en Chile Factura US\$2.000 Millones al Año. 2009. [en línea] La Estrategia on line. 19 de octubre, 2009. <http://www.estrategia.cl/detalle_noticia.php?cod=24085>. [consulta: 10 noviembre 2010].
- [18] PROGRAMA PARA heat maps [en línea]. <<http://www.myheatmap.com/>>. [consulta: 1 marzo 2011].
- [19] TIBCO. 2010. Location Data + Business Intelligence = LocationIntelligence. [en línea]. <<http://spotfireblog.tibco.com/?p=4070>>. [consulta: 20 febrero 2011].
- [20] TORRES, M. 2008. Herramientas estadísticas para la predicción de variables georeferenciadas. Memoria para Estadístico. Valparaíso, Pontificia Universidad Católica de Valparaíso, Facultad de Ciencias.
- [21] VIDAL, J. 2009. Modelo para la localización de supermercados orientada al cliente. Memoria para Ingeniero Civil Industrial. Santiago, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas. 67p.
- [22] WIKIMAPIA. 2011. Let's describe the whole world. [en línea]. <<http://wikimapia.org/#lat=-33.5101981&lon=-70.7877016&z=15&l=3&m=b>>. [consulta: 8 junio 2011].
- [23] WIKIPEDIA. 2011. Gran Temuco. [en línea] <http://es.wikipedia.org/wiki/Gran_Temuco>. [consulta: 11 junio 2011].
- [24] Yu, D. Yehua, D. [200-]. Geographically Weighted Regression. [en línea]. <pages.csam.montclair.edu/~yu/GISDay_GWR.ppt>. [consulta: 5 abril 2011].

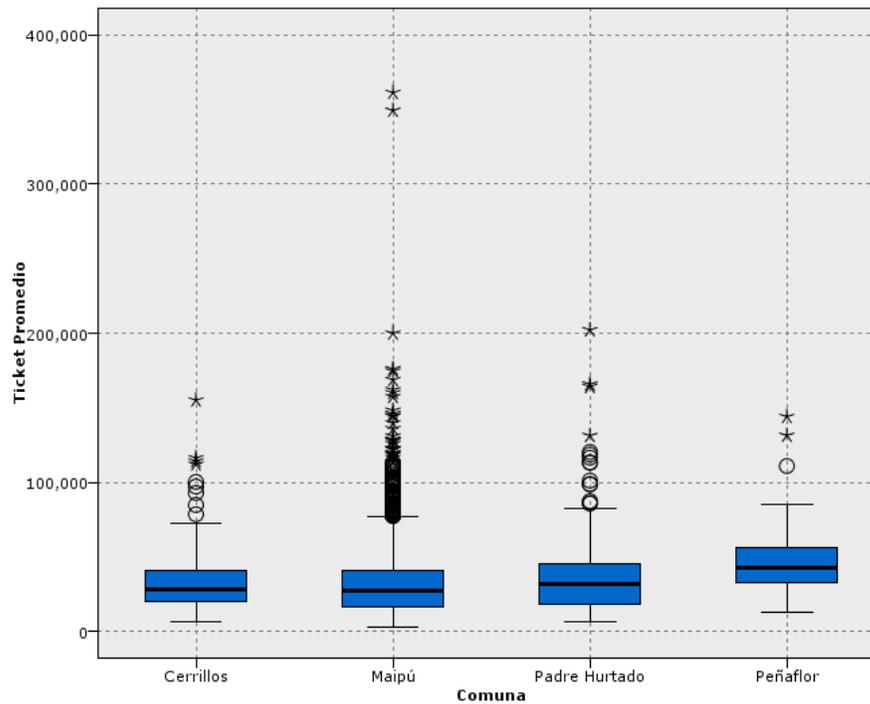
Anexo B: Frecuencia y ticket promedio

Gráfico 34: Días entre compras de clientes local Maipú



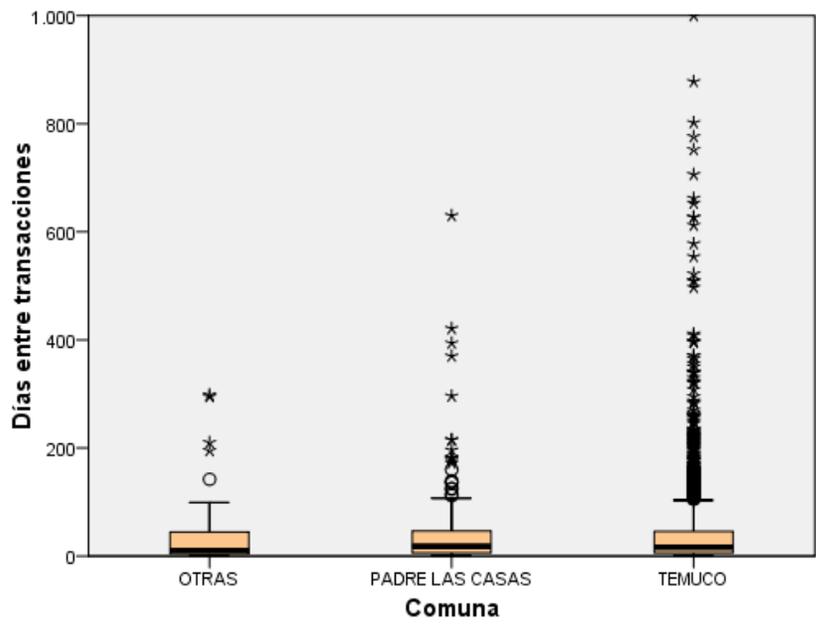
Fuente: Elaboración propia

Gráfico 35: Ticket promedio de clientes local Maipú



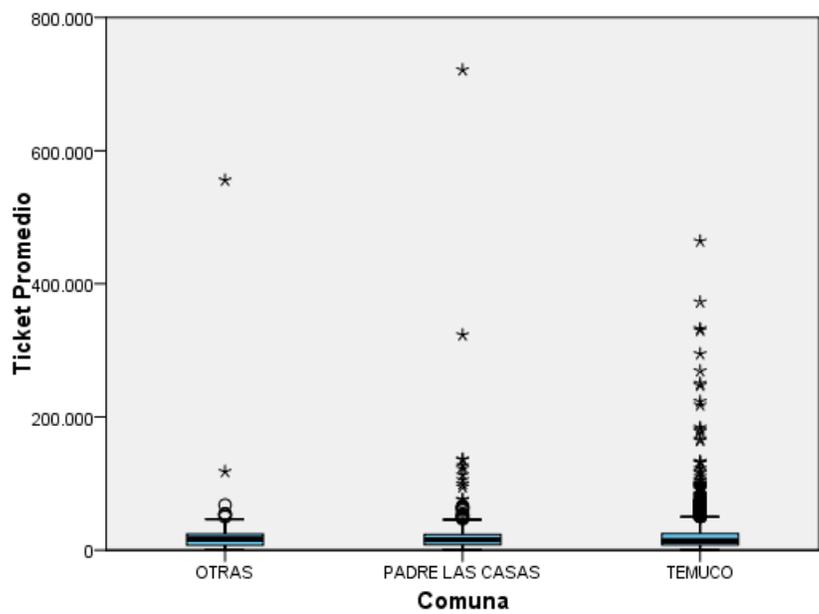
Fuente: Elaboración propia

Gráfico 36: Días entre compras de los clientes, local Temuco



Fuente: Elaboración propia

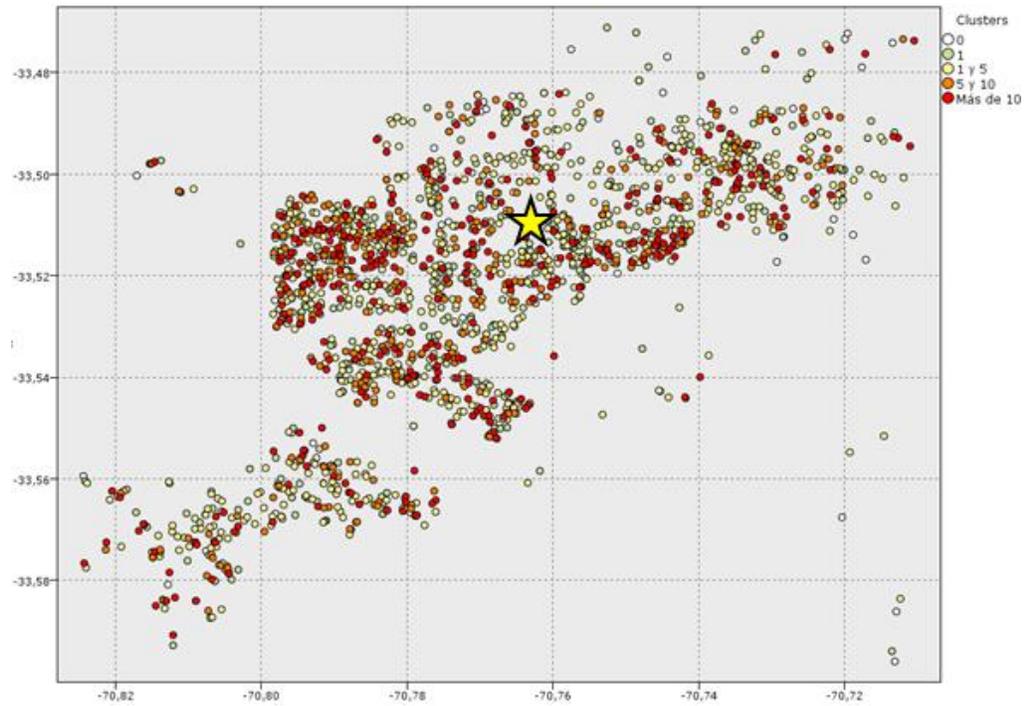
Gráfico 37: Ticket promedio de los clientes, local Temuco



Fuente: Elaboración propia

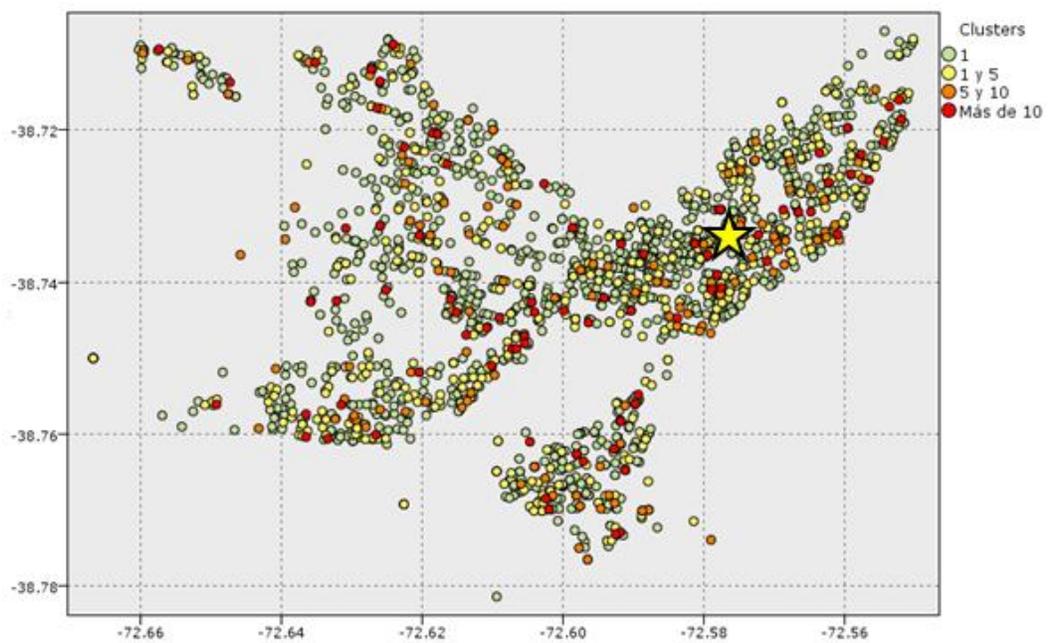
Anexo C: Scatter plots de las compras

Gráfico 38: *Scatter plot* de montos de clientes, local Maipú (en MM\$)



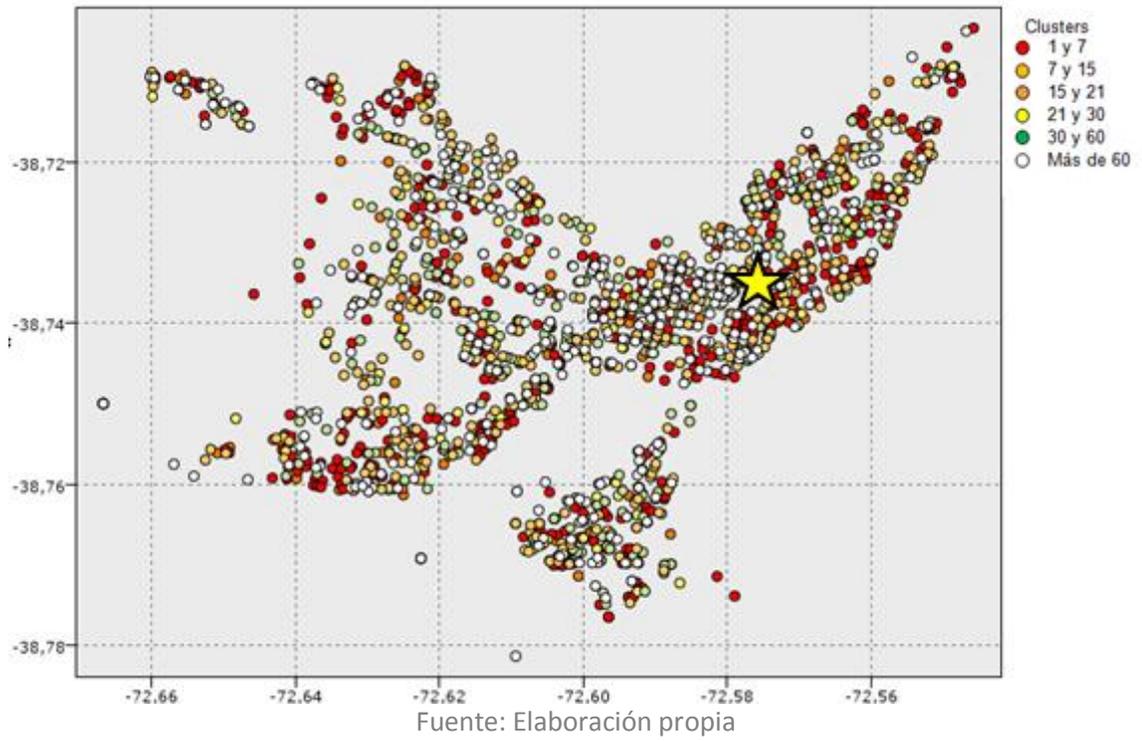
Fuente: Elaboración propia

Gráfico 39: *Scatter plot* de montos de clientes, local Temuco (en MM\$)



Fuente: Elaboración propia

Gráfico 40: Scatter plot para días entre compras, local Temuco



Anexo D: Errores

Gráfico 41: Histograma del MAPE en la estimación del monto con LR, local Maipú

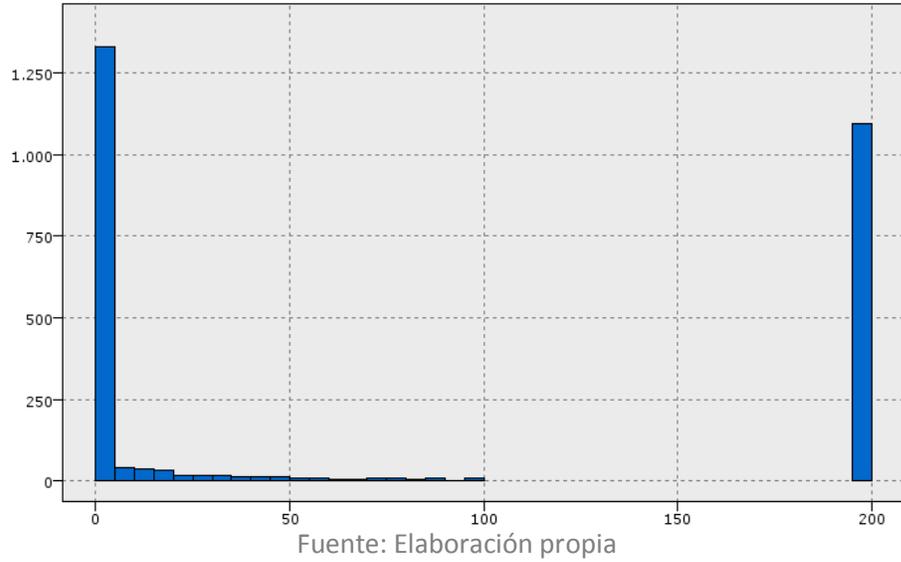


Gráfico 42: Histograma del MAPE en la estimación de los días entre compras con LR, local Maipú

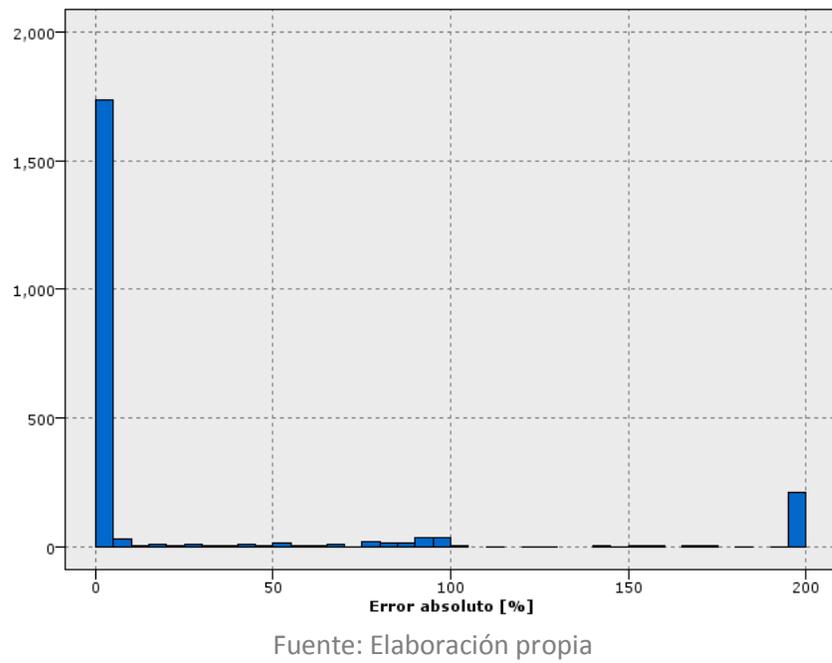


Gráfico 43: Histograma del MAPE en la estimación del monto con GWR, local Maipú

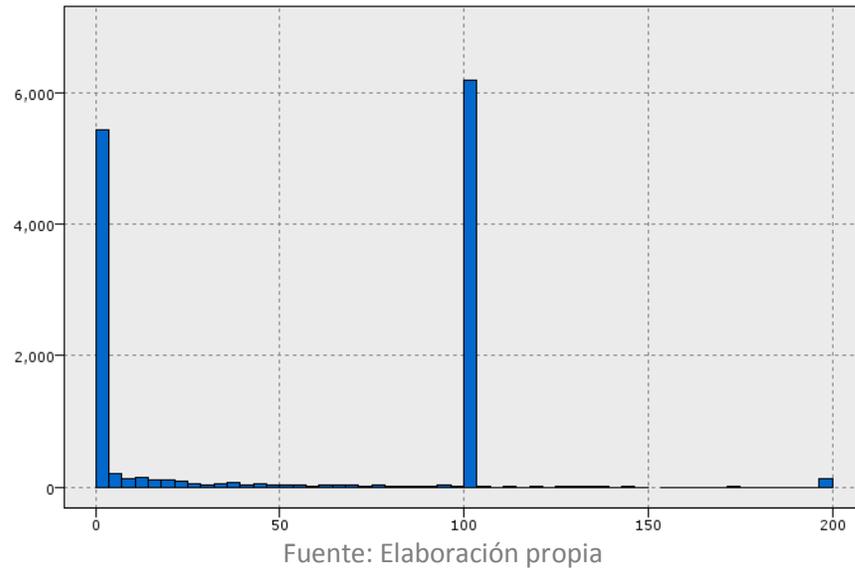


Gráfico 44: Histograma del MAPE en la estimación de los días entre compras con GWR, local Maipú

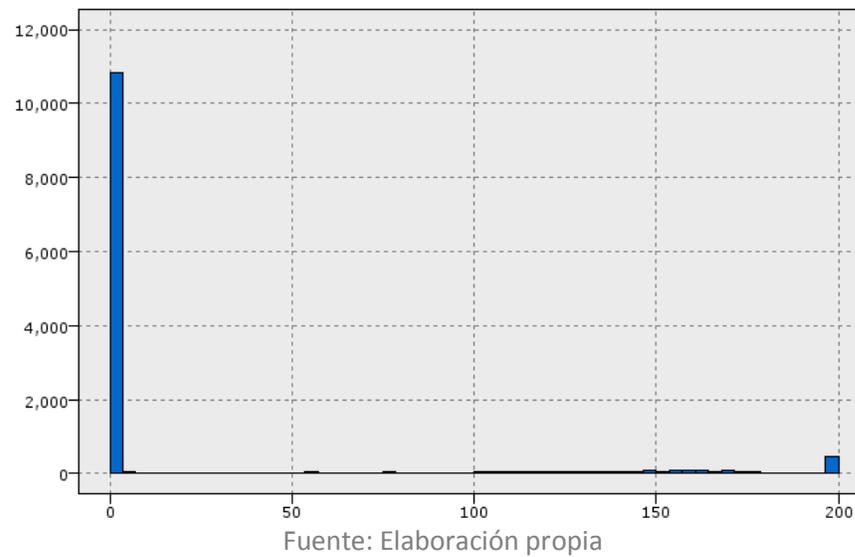
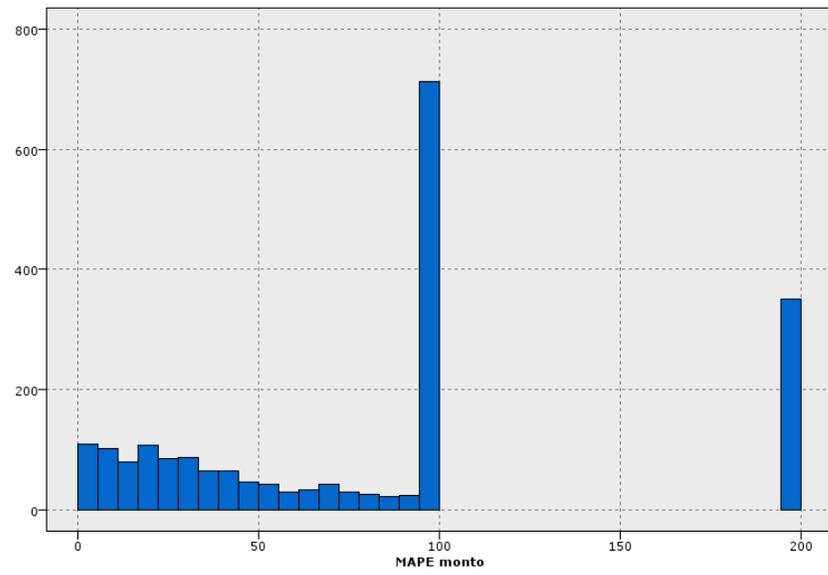


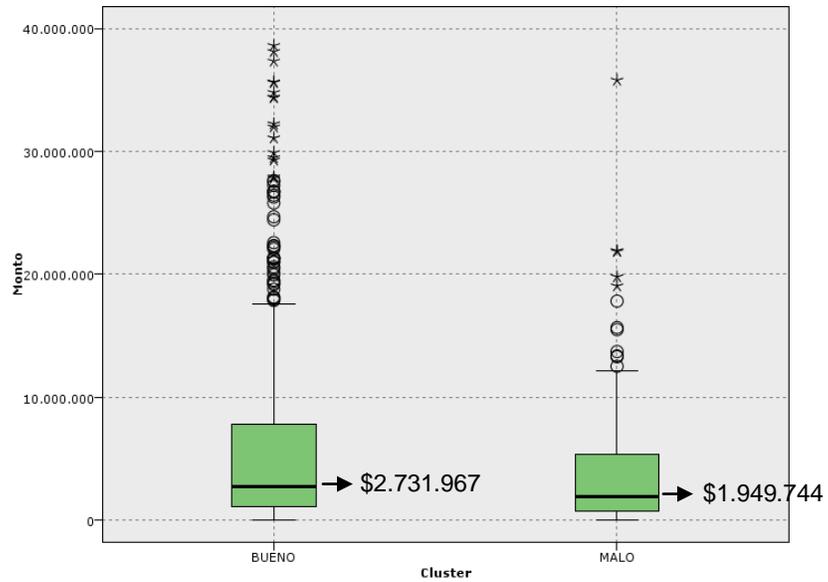
Gráfico 45: Histograma del MAPE para la estimación del monto con GWR, local Temuco



Fuente: Elaboración propia

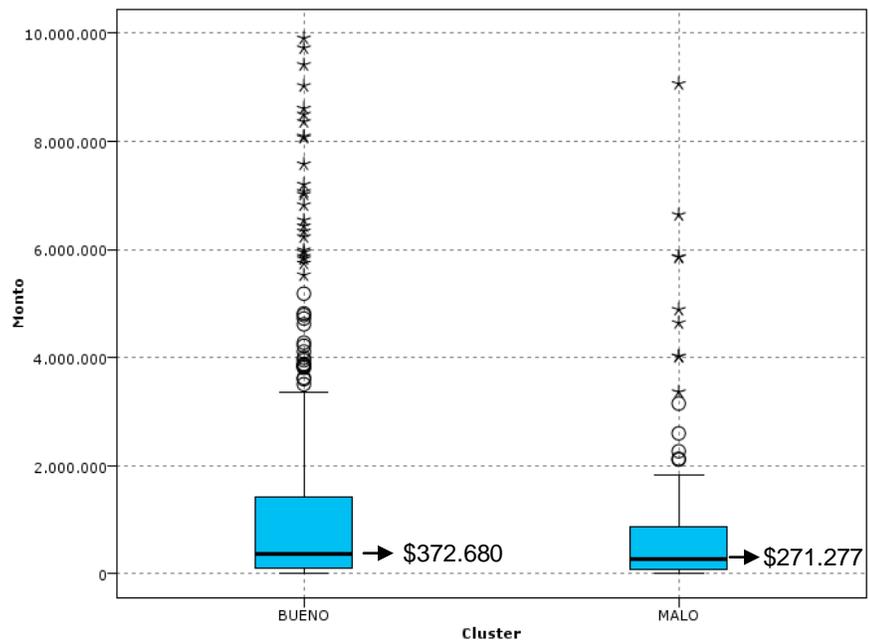
Anexo E: Gasto por barrio

Gráfico 46: Distribución del gasto según *cluster* de barrio, local Maipú



Fuente: Elaboración propia

Gráfico 47: Distribución del gasto según *cluster* de barrio, local Temuco



Fuente: Elaboración propia