



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

ANÁLISIS DE DATOS DE MOVIMIENTO DE CLIENTES EN UN SUPERMERCADO

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL

CÉSAR MIGUEL FERREIRO PÉREZ

PROFESOR GUÍA:
MARCEL GOIC FIGUEROA

MIEMBROS DE LA COMISIÓN:
MARCELO OLIVARES ACUÑA
ANDRES MUSALEM SAID

SANTIAGO DE CHILE
2017

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL
POR: CÉSAR MIGUEL FERREIRO PÉREZ
FECHA: 2017
PROF. GUÍA: MARCEL GOIC FIGUEROA

ANÁLISIS DE DATOS DE MOVIMIENTO DE CLIENTES EN UN SUPERMERCADO

La creciente incidencia de la tecnología en los procesos de compra ha facilitado poder hacer seguimiento de los clientes a través de las distintas etapas que componen una compra, tanto en formatos físicos como digitales. En particular en una tienda física es posible conocer la ubicación del cliente a largo de su visita a la sala de ventas.

En el presente trabajo se cuenta con la secuencia de ubicaciones seguidas por los clientes al interior de un supermercado en la ciudad de Santiago, estos fueron capturados a través de sensores dispuestos en la sala de ventas que detectan aquellos dispositivos móviles con la opción de wifi encendida. El principal objetivo del presente trabajo es entender como esta nueva fuente de información puede ayudar a enriquecer los análisis que se hacen actualmente sobre comportamiento de compra, donde generalmente solo se usan datos transaccionales. Para esto el análisis se divide en 3 partes: temporal, espacial y relacional.

Primero los datos son pre procesados aplicando filtros sobre estos para restringir el análisis sobre secuencias de ubicaciones que correspondan a viajes de compra. En particular se definen criterios sobre las zonas en que son detectadas las observaciones y la duración de las visitas a la sala de ventas. Una vez hecho esto se analiza como varía la frecuencia de observación de acuerdo con las distintas trayectorias observadas. De esta parte se concluye que una gran fracción de las observaciones no puede ser asignada a un viaje de compra.

Posteriormente se analizan los datos en su dimensión temporal a través de las variables hora de llegada y visitas por día. Estos datos se contrastan con variables análogas obtenidas a través de datos transaccionales. El principal resultado que se obtiene de esta comparación es que los patrones observados en datos transaccionales son consistentes con los obtenidos a través de datos móviles. Adicionalmente se logra estimar el tiempo que están los clientes al interior de la sala de ventas lo que no es posible con los datos transaccionales.

Luego se analizan los datos en su dimensión espacial, esto es como se mueven los clientes al interior de la sala de ventas. En particular se construye el indicador de velocidad y se estudia como este varía de acuerdo con distintas variables. Una vez hecho esto se analiza cómo cambian los resultados al introducir correcciones a las trayectorias capturadas.

Finalmente se comparan las visitas a ciertas zonas de la sala de ventas con las compras registradas en dichas zonas. El principal resultado de esta parte es la detección de brechas entre las compras y las visitas en distintas zonas de la sala de ventas. Esta información tiene el potencial de ser usada para mejorar la oferta de la sala de ventas en layout, ubicación de producto y ofertas entre otros.

We don't really want what we think we desire.

(Slavoj Žižek)

Agradecimientos

Con este trabajo se da por finalizada una de las etapas más intensas que me ha tocado vivir, la que no habría podido concluir sin la compañía de muchas personas.

Quiero agradecer a mi familia por siempre haber sido un espacio acogedor y de apoyo.

También a mis amigos industriales Pancho, Fran, Gurú, Asian, Veloso, Peraca, Juano, Chino y tantos más, con los que compartí cursos, trabajos, tareas, carretes y sin duda muy buenos momentos.

A los amigos que me dio la vida que siempre fueron un espacio de distención, buenos momentos y apoyo. Simón, Misterio y mis queridos Tigres.

Finalmente agradecer a los miembros de mi comisión en especial a mi profesor guía Marcel Goic por la buena disposición a discutir avances y todo el apoyo en el desarrollo de este trabajo.

Tabla de Contenido

Introducción	1
1. Metodología	5
2. Procesamiento de datos	9
2.1. Filtro lógicos	9
2.2. Frecuencia de observaciones	12
2.3. Procesamiento de datos de POS	14
3. Análisis temporal	15
4. Análisis espacial	19
4.1. Descripción variables	19
4.2. Análisis de la muestra	21
4.3. Análisis de velocidad	24
4.4. Testeo de hipótesis de comportamiento	26
4.5. Corrección de trayectorias	31
5. Brechas entre visitas y compras	35
5.1. Comparación datos transaccionales y de posición	35
5.2. Reglas de asociación de visitas	36
5.3. Reglas de asociación de transaccionales	37
5.4. Comparación reglas	39
6. Conclusiones y trabajos futuros	42
A. Mapa de la sala de ventas	44
B. Productos por zonas definidas 1	45
C. Productos por zonas definidas 2	46
Bibliografía	47

Índice de Tablas

1.1. Ejemplo regla	7
2.1. Ejemplo datos	9
2.2. Resumen de filtros aplicados	11
2.3. Descripción ob y frec	12
3.1. Comparación hora de llega y hora emisión boletas	16
3.2. Tiempo total con y sin filtro	18
4.1. Descripción dtotal y vel	20
4.2. Resultados test de medias.	23
4.3. Descripción velocidades por tipo de transición	24
4.4. Estimación modelo 4.1.	30
4.5. Estimación modelo 4.2	31
4.6. Estimación modelo 4.1 con corrección de velocidad.	33
4.7. Estimación modelo 4.2 con corrección de velocidad.	34
5.1. Descripción ob y frec	35
5.2. Comparación compras y visitas	36
5.3. Reglas con mayor soporte	37
5.4. Reglas con mayor Lift	37
5.5. Reglas con Lift<1	37
5.6. Reglas transaccionales con mayor Soporte	38
5.7. Reglas transaccionales con mayor Lift	38
5.8. Reglas transaccionales con Lift<1	38
5.9. Mayores brechas de soporte	39
5.10. Mayores brechas de confianza	40

Índice de Ilustraciones

1.	Ejemplo de camino seguido por un cliente.	3
2.1.	Zona de entrada y salida	10
2.2.	Degradación datos al aplicar filtros.	11
2.3.	Relación observaciones y frecuencia	12
2.4.	Concentración observaciones para distintas frecuencias	13
2.5.	División de la sala de ventas en zonas.	14
3.1.	Distribución hora de llega y hora emisión boletas	15
3.2.	Desfase hora de llega y hora emisión boletas	16
3.3.	Visitas por día	17
3.4.	Tiempo total de observación	17
3.5.	Tiempo total de observación con filtro	18
4.1.	Observaciones registradas	19
4.2.	Concentración de observaciones	20
4.3.	Histogramas dtotal y vel	20
4.4.	Comparación submuestras	22
4.5.	Ejemplo división sala de ventas	24
4.6.	Velocidades según tipo de transición	25
4.7.	Relación velocidad frecuencia	25
4.8.	Efecto de la frecuencia en la velocidad	26
4.9.	Tipos de transiciones corregidas	32
5.1.	Comparación soporte reglas transaccionales y de visitas	39
5.2.	Comparación confianza reglas transaccionales y de visitas	40

Introducción

Actualmente existe una tendencia global al alza de los canales de venta online como complemento de la tienda física, esto debido principalmente al aumento de la incidencia de la tecnología en los procesos de compra [1].

Este contexto presenta una gran oportunidad para los formatos físicos en el diseño de experiencias atractivas de consumo. Ya que si bien la principal variable al explicar la elección de una tienda física es la ubicación de esta, al momento de explicar fidelidad a una determinada marca, la experiencia de compra es lo más importante. [2].

En los últimos años investigadores han planteado que el ambiente en que se desarrolla el consumo tiene un efecto tanto o más importante en el comportamiento de los clientes que los elementos clásicos de análisis tales como: precio, surtido de productos y promociones [12]. La principal explicación a este fenómeno es que el ambiente constituye el medio a través del cual el cliente interactúa con los productos. Esto afecta variables tan relevantes como el tiempo que se gasta en la tienda, la forma en que esta es visitada y la atención que se presta a los diversos productos y categorías.

Más aun, existe evidencia que variables tales como layout y localización de productos tienen un efecto significativo en el proceso de decisión de compra de los clientes [7]. Por ejemplo, se ha demostrado que el orden en que se construye la canasta, el que depende en parte del layout, produce un efecto conocido como *licensing*, que en el contexto de compras en la tienda física se refiere a la idea de que comprar categorías “sanas” (frutas, verduras, etc.) , compensa el efecto negativo asociado a categorías “no sanas” (cerveza, dulces, etc.) y viceversa [11]. De este modo si es que un cliente agrega a la canasta un categoría sana aumenta su probabilidad de agregar una que es no sana durante su visita a la tienda. [4].

También se han estudiado las trayectorias que siguen los clientes al interior de la sala de ventas y como estas varían con respecto a un camino mínimo. En esta línea se ha encontrado que caminos con mayores desvios en comparación a un camino mínimo están asociados a canastas de productos mas grandes mientras que clientes con trayectorias más parecidas a un camino óptimo tienden a comprar más en las categorías que tienen mayor frecuencia de venta [9].

Estos patrones de comportamiento han sido verificados a través de experimentos de campo donde a los clientes se les entregan cupones de descuento de distintos productos ubicados al interior de la sala de ventas. Se ha encontrado como resultado que mientras más hace el

cupón que el cliente se desvíe de su trayectoria planeada, mayor es el monto de compras no planeadas [5].

Otro fenómeno que se ha evidenciado es que la concentración de clientes en la sala afecta tanto la probabilidad de visita como de compra en la zona congestionada. Es así, que cuando un cliente detecta un punto de alta concentración de clientes dentro de la sala, aumenta su probabilidad de visitar esta zona. Pero disminuye su probabilidad de comprar productos presentes en esta zona [4].

Debido a esta importancia del ambiente en el proceso de compra y al surgimiento de nuevas tecnologías, es que actualmente se habla de la tercera ola en la inteligencia de mercados: la gestión de experiencia de clientes (customer experience management) [12], cuya principal novedad es la posibilidad de seguir en tiempo real a los clientes en la tienda, ya sea física o virtual.

Esto genera la oportunidad de enriquecer los análisis que actualmente se hacen sobre comportamiento del consumidor. Ya que al considerar las variables en que se desarrolla el proceso de compra, es posible medir el impacto que tiene el ambiente en los patrones de comportamiento del cliente.

Una de las formas en que es posible seguir a los clientes dentro de la tienda física, es a través de sensores que detectan los dispositivos móviles que tienen la opción wifi encendida, y entregan la posición del dispositivo dentro de la sala a través del tiempo.

Esta nueva fuente de información permite el desarrollo de nuevos análisis sobre como los consumidores se desenvuelven dentro de la tienda física integrando nuevas variables que sin esta tecnología no son posibles de obtener.

Por ejemplo permite identificar cuando cierta zona es visitada sin que necesariamente se haya comprado algún producto de esta. Una posible aplicación podría ser descomponer la eficacia de puntos promocionales en visitas como en compras. Y también estudiar como la promoción afecta las visitas y compras a otros productos.

También permite obtener la secuencia de zonas visitadas por el cliente, y de este modo entender como va construyendo la canasta y que variables influyen en esto. Potencialmente se puede diferenciar entre zonas en que los clientes pasan más rápido o más lento y esto relacionarlo con el nivel de atención que prestan a los productos.

En el presente trabajo se cuenta con datos sobre la posición de los clientes al interior de la sala de ventas de un supermercado de la ciudad de Santiago los datos fueron capturados con la tecnología Cisco CMX. A través de estos datos se construyen los caminos seguidos por los clientes al interior del supermercado.

Formalmente un camino se define como el movimiento consistente de un agente en un medio físico o simulado que es observable. Para caracterizarlo se requieren 3 elementos [8]:

- Ambiente: es el medio a través del cual se mueve el agente. En este caso corresponde a

la sala de ventas en estudio.

- Agente: es quien realiza el movimiento. En este caso corresponde a los clientes cuyos dispositivos móviles tienen la opción wifi encendida y son detectados por los sensores dispuestos en la sala de ventas.
- Posición: corresponde a las coordenadas del agente en el ambiente para distintos instantes de tiempo.

En la Figura 1 se muestra el camino seguido por un cliente, donde las líneas representan la secuencia seguida dentro de su visita y las intersecciones de estas son los puntos capturados por los sensores.

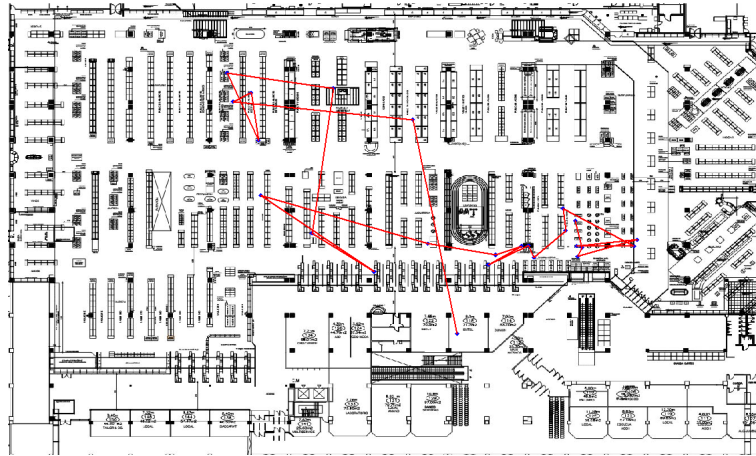


Figura 1: Ejemplo de camino seguido por un cliente.

Al observar la imagen se vislumbran algunos de los desafíos de trabajar con esta nueva fuente de información. Por ejemplo, las mediciones no son equiespaciadas, por lo que dependiendo de la zona en que se mida varía la frecuencia de las observaciones. Esto además tiene como consecuencia que se produzcan “saltos” entre observaciones consecutivas, lo que plantea como desafío poder inferir los caminos efectivos a partir de la información disponible.

El principal objetivo de este trabajo es analizar esta nueva fuente de información, para así poder dimensionar el alcance que tiene esta tecnología. La forma en que se abordará el estudio será a través de 3 aristas: temporal, espacial y relacional.

Primero se pre procesan los datos aplicandoles filtros lógicos que permitan recuperar los caminos seguidos por los clientes y eliminar aquellas observaciones que no sean atribuibles a un viaje de compra. Luego se analizan variables temporales tales como la hora de llegada a la sala de ventas, la duración de las visitas y el número de visitas por día. Estos datos son comparados con datos de venta para ser validados.

Posteriormente se analizan la forma en que los clientes se mueven al interior de la sala de ventas. Para esto se usan las variables espaciales obtenidas por los sensores y se construyen variables de desplazamiento como velocidad y distancia recorrida, con las que se testean hipótesis de comportamiento sobre la velocidad con que los clientes se mueven al interior de

la sala de ventas.

Finalmente se estudia la relación entre las visitas a las distintas zonas de la sala de ventas. Para esto se levantan reglas de asociación las que son comparadas con reglas de asociación obtenidas a través de datos transaccionales, para identificar brechas entre ventas y visitas.

Objetivos

Objetivo general

Analizar patrones de movimiento de clientes en un supermercado usando información de los caminos seguidos por estos en la sala de ventas.

Objetivos específicos

- Definir filtros sobre los datos para recuperar caminos seguidos por clientes.
- Comparar variables temporales obtenidas con los sensores de movimiento con variables análogas obtenidas de datos transaccionales.
- Analizar espacialmente los datos y testear hipótesis de comportamiento.
- Levantar reglas de asociación de visitas a las distintas zonas de la sala de ventas y compararlas con reglas de asociación obtenidas a través de datos transaccionales.

Capítulo 1

Metodología

Etapas del estudio

Las principales etapas del estudio son cuatro: pre procesamiento de datos, análisis temporal, análisis espacial y análisis relacional de los datos.

Pre procesamiento datos

En esta etapa se pretende depurar la base de datos disponible para eliminar todas aquellas observaciones que no sean útiles para los desarrollos posteriores.

Primero se aplicarán filtros lógicos para poder diferenciar entre las observaciones que correspondan a clientes y las que no. En particular se eliminan las observaciones que no constituyen caminos y además se definen zonas de entrada y salida de la sala de ventas para identificar a los clientes.

Una vez aplicados los filtros se construirán variables temporales, las que tienen relación con el momento del día en que realiza la visita al supermercado y la duración que tiene esta, y variables de desplazamiento, que caracterizan el movimiento al interior de la sala de ventas.

Luego se analiza la calidad de los datos en cuanto la frecuencia de medición. En particular se estudia como decae la frecuencia promedio de medición a medida que se tienen más observaciones por camino. Además se analiza como varía la concentración de observaciones al interior de la sala de ventas para distintos rangos de valor de la frecuencia.

Finalmente se explica el procesamiento de datos transaccionales los que son usados posteriormente para compararlos con las variables temporales y en la evaluación de reglas de asociación. Para este último punto se define un nivel de agregación de los pasillos y se asignan los productos a estos.

Análisis Temporal

En esta etapa se analizarán las variables temporales: hora de llegada a la sala de ventas, número de visitas por día y tiempo total de observación. Las primeras dos variables serán comparadas con variables análogas obtenidas a través de datos transaccionales. Mientras que para la variable duración se analizará como varía esta al aplicarle filtros sobre la duración de observación mínima exigida. El principal objetivo de esta etapa es concluir si es que los datos móviles son consistentes con los datos transaccionales.

Análisis Espacial

En esta etapa se analizan las variables espaciales que corresponden a la posición de los dispositivos móviles al interior de la sala de ventas. Con estas se construyen las métricas de desplazamiento : distancia total recorrida y velocidad.

Primero se analiza la distribución de las observaciones al interior de la sala de ventas, esto con dos objetivos. El primero es identificar si es que existen irregularidades en las mediciones al interior de la sala de ventas. Y el segundo detectar puntos donde exista mayor concentración de observaciones, y si es que esta concentración de puntos es consistente en el contexto de un supermercado.

Luego se analiza la muestra con la que se está trabajando para concluir si es que presenta algún tipo de sesgo. Para esto se construye la variable dt que representa el tiempo entre observaciones consecutivas (transición) y se realiza sobre esta distintos test de medias para identificar si que varía en relación a otras variables.

Posteriormente se modela la sala de ventas en pasillos y corredores, esto para poder caracterizar distintos tipos de transiciones, y poder caracterizar diferencias entre estas. Una vez hecho esto se plantean modelos lineales para testear hipótesis de comportamiento de clientes al interior de la sala de ventas.

Finalmente se vuelven a estimar los mismos modelos anteriores, pero esta vez haciendo correcciones a las trayectorias presentes en la muestra sobre la que se trabaja. En particular se busca corregir las trayectorias que presentan saltos discontinuos entre zonas no adyacentes de la sala de ventas.

Análisis Relacional: Reglas de asociación

Las reglas de asociación estudian la relación entre eventos y como se correlaciona la ocurrencia de estos. Usualmente son usadas en el análisis de canasta realizado por los supermercados con el fin de detectar que productos tienden a ser comprados juntos y cuales no. De este modo se pueden tomar decisiones de promociones y distribución de productos dentro de la sala de ventas.

Se tiene $I = I_1, I_2, \dots, I_m$ un conjunto de atributos binarios denominados *items* y T una base de datos de transacciones. Con esto se define una regla de asociación como una implicancia de la forma $X \rightarrow I_j$, donde X es un conjunto de *items* contenidos en I denominado antecedente y I_j es un *item* de I que no está contenido en X llamado consecuente [10].

Para este caso los atributos son las visitas a las distintas categorías registradas por los sensores dispuestos en la sala de ventas y la taxonomía es la siguiente:

$$\{\text{Visitazona } i\} \rightarrow \{\text{Visitazona } j\}$$

Cabe destacar que este enfoque permite construir reglas de mayor complejidad que incluyan más elementos en el antecedente, por ejemplo $\{\text{Visitazona } i, \text{Visitazona } h\} \rightarrow \{\text{Visitazona } j\}$.

Para evaluar las reglas de asociación existen 3 indicadores básicos: soporte, confianza y lift [3]. Los que se explicarán mediante un ejemplo.

Se tiene la regla mostrada en la Tabla 1.1. A continuación se detalla el valor de los indicadores y su respectiva interpretación.

Antecedente	Consecuente	Soporte	Confianza	Lift
Visita Zona de Dulces	Visita Zona de Fruta y Verdura	0.44	0.73	1.17

Tabla 1.1: Ejemplo regla

- **Soporte:** Proporción de visitas que contienen al antecedente y consecuente.

$$\text{Soporte} = \frac{\text{Visitas que contienen}(\text{Dulces} \cup \text{Fruta y Verdura})}{\text{Total Visitas}} = 44\%$$

Esto quiere decir que el 44% de las visitas registradas visitan la zona de Dulces y la de Fruta y Verdura.

- **Confianza:** Probabilidad de que aparezca el consecuente y el antecedente en una transacción, dado que aparece el antecedente.

$$\text{Confianza} = \frac{\text{Visitas que contienen}(\text{Dulces} \cup \text{Fruta y Verdura})}{\text{Visitas que contienen}(\text{Dulces})} = 73\%$$

Significa que la probabilidad de visitar la zona de Fruta y Verdura dado que se visitó Dulces es de 73%.

- **Lift:** Puntaje de aumento en la probabilidad de que ocurra el consecuente dado que ocurrió el antecedente.

$$\text{Lift} = \frac{\text{Confianza}(\text{Dulces} \cup \text{Fruta y Verdura})}{\text{Soporte}(\text{Fruta y Verdura})} = 1,17$$

Esto significa que si se visitó Dulces, la probabilidad de visitar Fruta y Verduras aumenta en un 17%.

La razón de definir el Lift es que cuando un *item* consecuente tiene un soporte muy alto (está presente en muchas visitas) , el resultado del indicador de confianza puede verse afectado debido a que la cantidad de transacciones en que aparezca el antecedente y consecuente será similar a las que aparezca solo el consecuente. Esto tiene como resultado una confianza alta no debido a una relación entre *items*, sino a que el soporte del antecedente se vuelve relativamente insignificante frente al del consecuente.

Finalmente para evaluar las reglas de asociación estas se compararán con reglas generadas a través de datos transaccionales. La forma en que se compararán será a través de la detección de brechas en los indicadores presentados anteriormente para las distintas reglas levantadas. El principal objetivo de esta parte es identificar diferencias entre las reglas generadas que puedan servir para mejorar la oferta del supermercado.

Capítulo 2

Procesamiento de datos

2.1. Filtro lógicos

Los datos usados para el desarrollo del presente estudio provienen de mediciones tomadas en un supermercado de Santiago entre el 2 y 15 de Noviembre de 2016. Previo a la aplicación de filtros, se cuenta con 2.275.843 registros correspondientes 183.883 dispositivos móviles distintos.

Cada registro indica la posición de un dispositivo móvil dentro de la sala de ventas para un instante dado de tiempo, tal como se muestra en la Tabla 2.1. Estos datos fueron tomados usando la tecnología Cisco CMX, la que en teoría entrega una frecuencia de 1 punto cada 8 segundos con una precisión de 1-3 metros.

MAC	Fecha	Tiempo	X	Y
94:b1:0a:14:f6:5b	9	18.12	225.13	232.64
94:b1:0a:14:f6:5b	9	18.14	211.81	227.86

Tabla 2.1: Ejemplo datos

Con el fin de aplicar filtros a los datos se construye una base de datos auxilia que contiene las siguientes variables:

- **MAC:** Dirección MAC del dispositivo móvil.
- **Fecha:** Fecha en que fue capturada la observación.
- **ob:** Número de observaciones asociadas a la dirección MAC en la fecha dada.
- **tmin:** Tiempo en que se registra la primera observación de una dirección MAC dada en una fecha dada.
- **(xmin , ymin):** Coordenadas asociadas al tiempo tmin.

- **tmax:** Tiempo en que se registra la última observación de una dirección MAC dada en una fecha dada.
- **(xmax , ymax):** Coordenadas asociadas al tiempo tmax.
- **ttotal:** Tiempo total de registro calculado como la diferencia entre tmax y tmin.
- **dtotal:** Distancia recorrida por el cliente durante su tiempo de registro.
- **vel:** Velocidad promedio del cliente durante su visita. Se calcula como $vel = \frac{dtotal}{ttotal}$.
- **frecuencia:** Promedio de observaciones por minuto, se calcula como $frecuencia = \frac{ob}{ttotal}$.
- **distanciaprom:** Promedio de distancia entre observaciones consecutivas, se calcula como $distanciaprom = \frac{dtotal}{ob}$.
- **dtprom:** Tiempo promedio entre observaciones.
- **ddistanciaprom:** Distancia promedio entre observaciones.

Sobre la base de datos auxiliar se aplican filtros lógicos con el objetivo de capturar aquellas observaciones que corresponden a clientes y descartar las que no lo son. El primer filtro aplicado consiste en eliminar todos los caminos que tengan solamente una observación. Con esto se pasa de 183.339 dispositivos móviles en 2.275.843 registros a 103.288 dispositivos móviles en 2.195.792 registros.

Con el objetivo de capturar las observaciones que visitan la sala de ventas y no solamente las tiendas que se encuentran afuera de esta. Se aplica un filtro sobre la zona donde es detectada la primera y la última observación de cada dispositivo móvil. Específicamente se conservan solo los caminos cuya primera observación se encuentra dentro de la zona verde o naranja y la última observación es registrada bajo la línea naranja, de acuerdo como se muestran en la Figura 2.1. Luego de aplicar este filtro se conservan 29.341 dispositivos móviles en 654.035 registros.

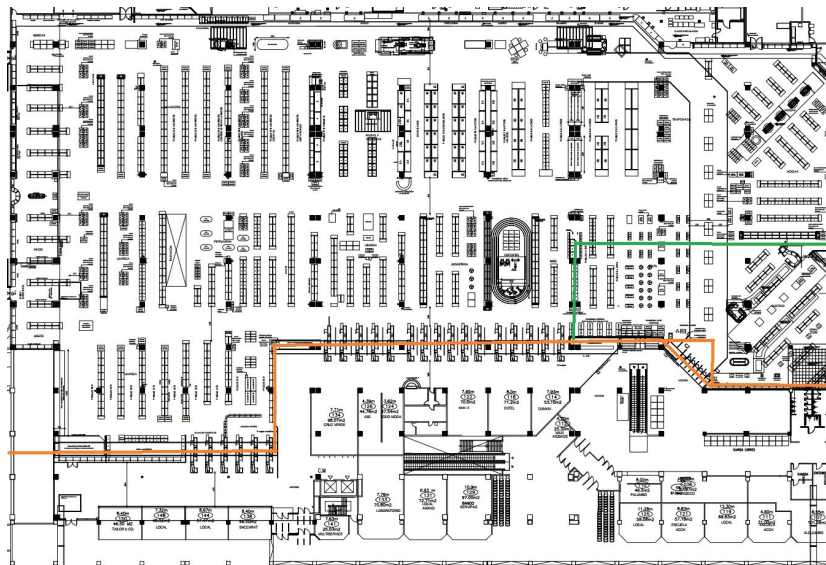


Figura 2.1: Zona de entrada y salida

Luego para diferenciar entre clientes y personal del supermercado tales como reponedores,

promotores, vendedores, etc. Se filtra por el tiempo total de observación, conservando solo las que tengan menos de 3 horas de registro, quedando 27.009 dispositivos móviles en 518.410 registros.

Finalmente se eliminan todos los caminos que estén compuestos solamente de dos observaciones. Esto debido a que, al haber aplicado el filtro de zona de entrada y salida, si es que un camino registra solamente dos observaciones, se encontrarán en estas zonas respectivamente y no entregarán información sobre el recorrido del cliente. Con esto quedan 23.481 dispositivos móviles en 511.354 registros.

En la Figura 2.2 se muestra como se degradan los datos al aplicarle filtros. Se observa que el filtro de zonas de entrada y salida es el que hace perder mas observaciones, tanto en dispositivos móviles como en registros. Esto es posiblemente a que fuera de la sala de ventas hay distintas tiendas cuyos clientes también están siendo capturados por los sensores.

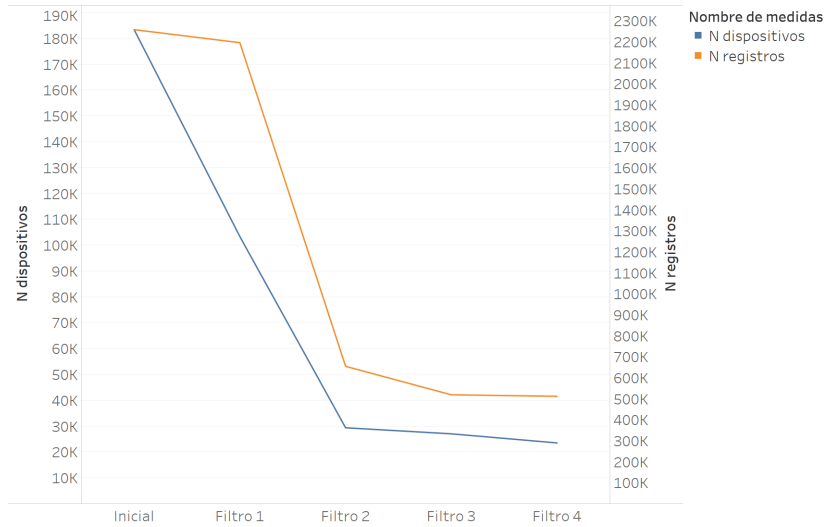


Figura 2.2: Degradación de datos al aplicar filtros.

Filtro	N Dispositivos móviles	N Registros
Sin Filtro	183.339	2.275.843
Filtro 1	103.288	2.195.792
Filtro 2	29.341	654.035
Filtro 3	27.009	518.410
Filtro 4	23.481	511.354

Tabla 2.2: Resumen de filtros aplicados

2.2. Frecuencia de observaciones

Al analizar la frecuencia y el número de puntos por camino de los datos obtenidos luego de aplicar los filtros lógicos se tienen los datos mostrados en la Tabla 2.3. En la Figura 2.3 se muestra la cantidad de puntos por camino con la respectiva frecuencia promedio del camino, donde puede apreciar que los caminos con mayor frecuencia promedio tienden a estar compuestos en su mayoría por menos observaciones y que a medida que aumenta la cantidad de observaciones por camino disminuye la dispersión en la frecuencia promedio.

Métrica	ob	frec
max	656	44.3
min	2	0.02
promedio	21.8	0.02
mediana	13	0.64
desviación	27.5	3.07

Tabla 2.3: Descripción ob y frec

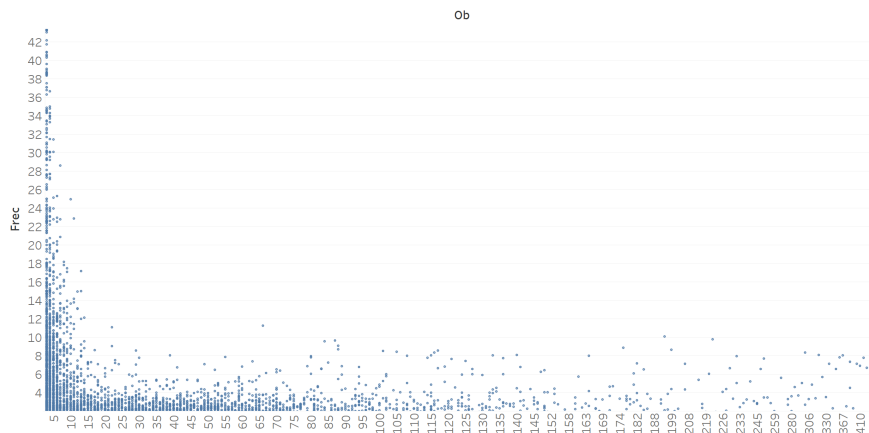
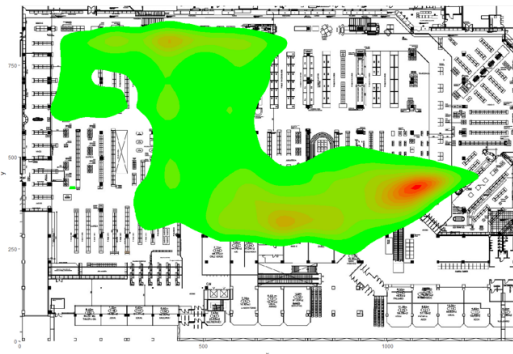


Figura 2.3: Relación observaciones y frecuencia

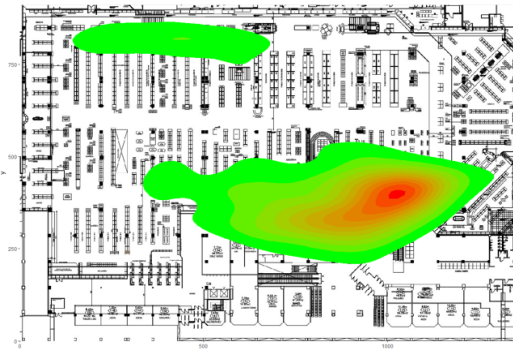
En la Figura 2.4 se muestra la concentración de observaciones para distintos rangos de valores de la frecuencia. Es posible observar que a las observaciones con una frecuencia entre 0 y 10 tienden a distribuirse más dentro de la sala, mientras que las mayores a este rango están más concentradas en la zona de entrada y cajas.



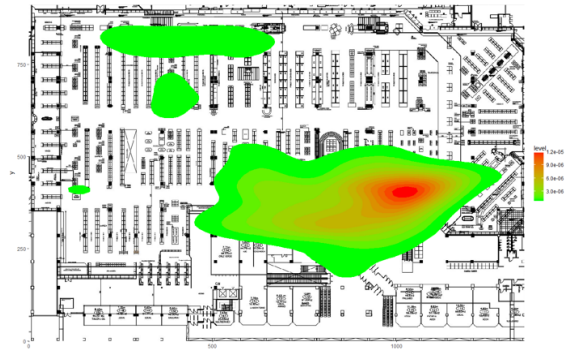
(a) frec 0-10



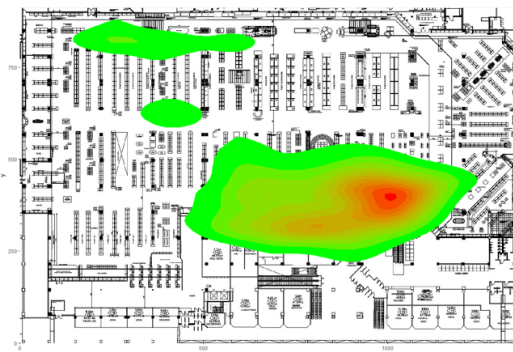
(b) frec 10-20



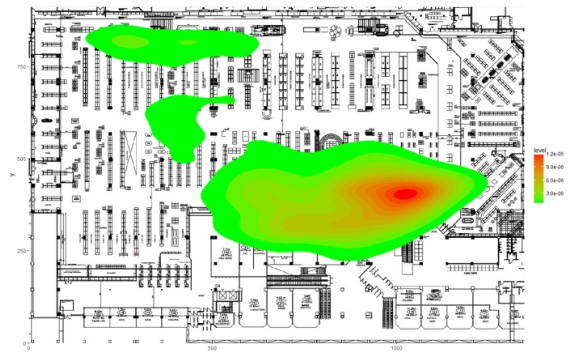
(c) frec 20-30



(d) frec 30-40



(e) frec 40-50



(f) frec 50-64

Figura 2.4: Concentración observaciones para distintas frecuencias

Lo mostrado en las Figuras 2.3 y 2.4 sugiere que la frecuencia de medición no es igual en todas las zonas de la sala de ventas. Esto debido a que clientes que recorren más la sala de ventas tienen mayor probabilidad a pasar por zonas donde la frecuencia de medición es menor, y de este modo disminuir la frecuencia promedio del camino registrado. Se concluye que existe un *trade off* entre cantidad y calidad de información disponible.

2.3. Procesamiento de datos de POS

Se cuenta con los datos transaccionales del supermercado para el mismo periodo en el que se tienen los datos de los caminos seguidos por los clientes (2-15 de Noviembre). En total se registran 29.015 productos presentes en 115.456 boletas.

Estos productos se asignan a distintas zonas de la sala de ventas, de acuerdo al nivel de agregación definido en la Figura 2.5. De este modo se logran clasificar 25.592 productos quedando 102.059 canastas. En los anexos B y C se muestra el detalle de los productos pertenecientes a cada zona definida. La principal dificultad en la clasificación de productos es que no se cuenta con información detallada de los planogramas, por lo que esta debe ser hecha inspeccionando la sala de ventas lo que hace perder precisión.

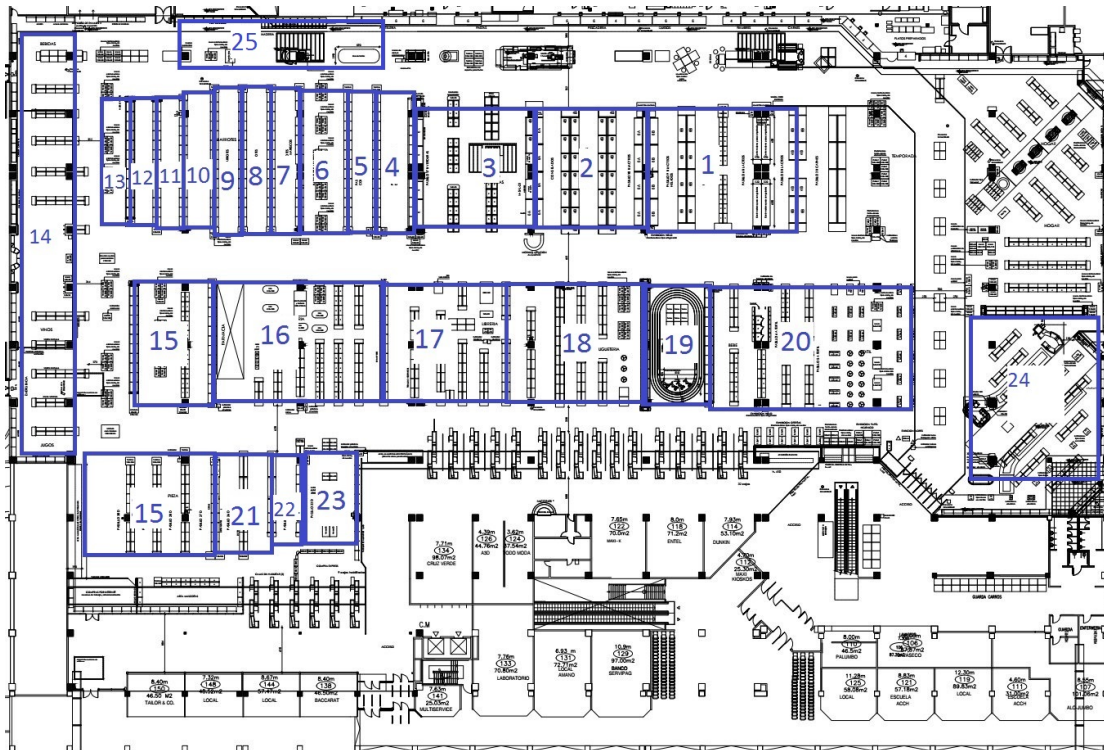


Figura 2.5: División de la sala de ventas en zonas.

Capítulo 3

Análisis temporal

En este capítulo se analizan 3 variables temporales: tiempo de llegada a la sala de ventas, visitas registradas por día y tiempo total de observación del cliente. Las primeras dos se comparan con variables análogas obtenidas a partir de datos transaccionales para así verificar si es que los datos móviles son consistentes con los transaccionales.

En la Figura 3.1 se compara la distribución de la hora de llegada de los clientes a la sala de ventas, (hora en que se obtiene la primera observación del dispositivo móvil correspondiente) y la hora de emisión de las boletas. Es posible observar que ambas tienen una distribución similar, presentando dos puntos de alta congestión: uno en torno a las 13:00 y otro entre 19:00 y 21:00 aproximadamente. Estos rangos de alta congestión corresponden a el horario de almuerzo y el horario de salida de oficinas respectivamente, por lo que es esperable que aumente el tráfico de clientes en estos puntos.

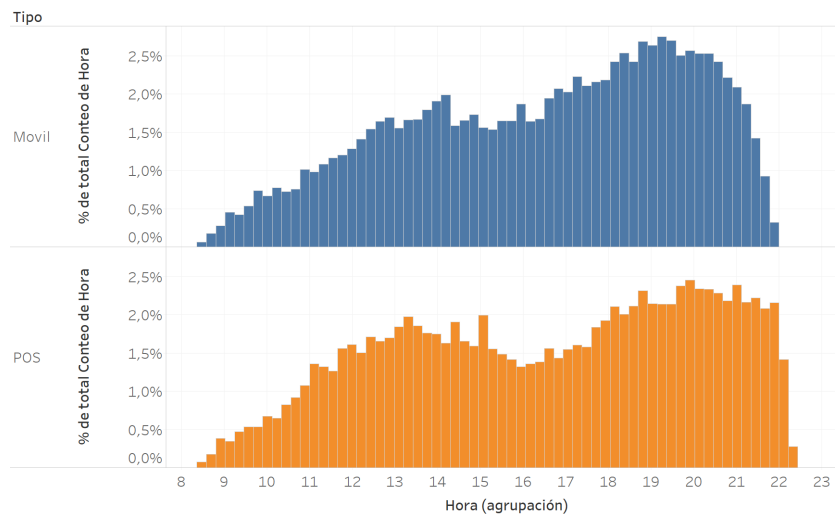


Figura 3.1: Distribución hora de llega y hora emisión boletas

	Movil	POS
max	22:00	22:45
min	08:30	08:30
promedio	16:28	16:34
mediana	16:58	16:55
desviación	3.3	3.6

Tabla 3.1: Comparación hora de llega y hora emisión boletas

Además, es esperable que exista un desfase entre la hora de ingreso a la sala y la hora de emisión de las boletas explicado por el tiempo que demoran los clientes al interior de la sala de ventas en la construcción de sus canastas. Este desfase se muestra en la Figura 3.2. De esta figura tambien es posible observar la magnitud del procesos de sampleo, donde aproximadamente se obtiene un camino cada 100 boletas registradas.

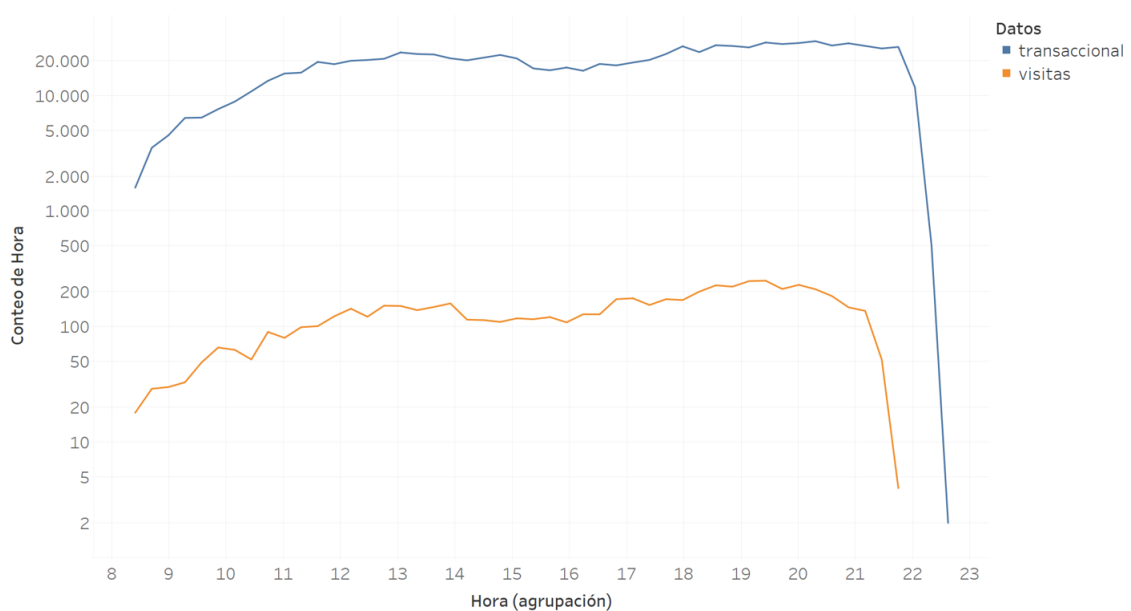


Figura 3.2: Desfase hora de llega y hora emisión boletas

En la Figura 3.3 se compara el porcentaje de visitas registradas con el número de boletas emitidas para cada día de la semana. Tal como se puede ver los datos moviles en general siguen las misma tendencia que los datos transaccionales.

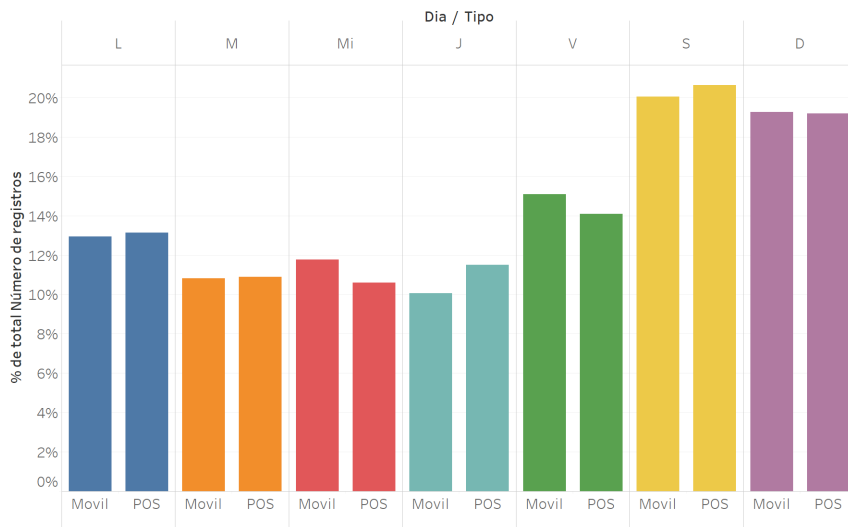


Figura 3.3: Visitas por día

En base a lo mostrado en las Figuras 3.1 y 3.3 se puede concluir que si bien los datos moviles son menores en volumen que los transaccionales en general son consistentes en la dimensión temporal (hora de llegada y visitas por día).

Finalmente se calcula el tiempo total de visita en la sala de ventas como la diferencia de hora entre la primera y la última observación. En la Figura 3.4 se puede ver que gran parte de las observaciones está en un rango de duración de entre 0 y 3 minutos de medición. Esto posiblemente se debe a dispositivos moviles que no constituyen visitas a la sala de ventas.

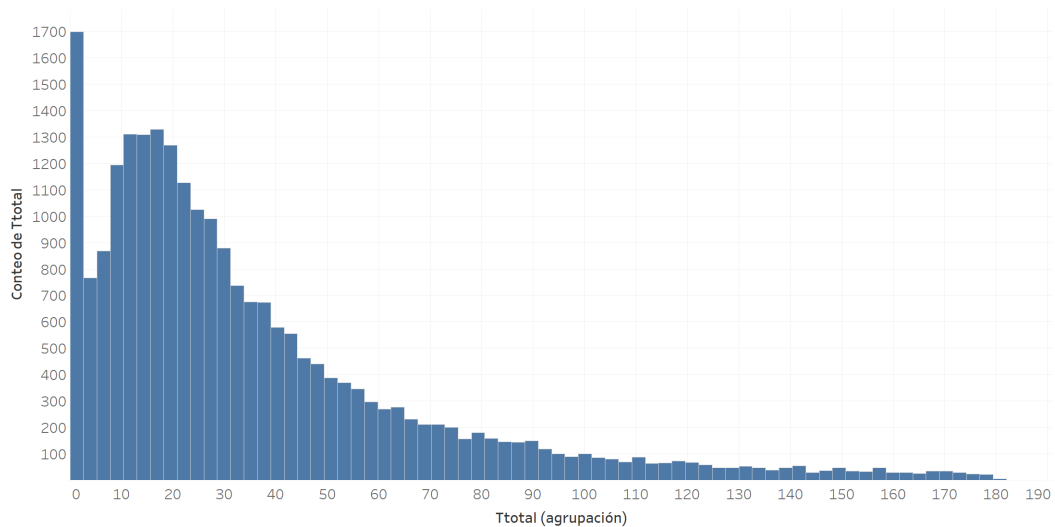


Figura 3.4: Tiempo total de observación

En la tabla 3.2 se describe el tiempo total de observación al aplicarle un filtro que considere solo las mediciones con un tiempo de observación mayor a 3 minutos. Al aplicar este filtro se eliminan el 16 % de los dispositivos moviles registrados. Esto provoca un aumento en la mediana y el promedio del tiempo total de observación.

	Sin Filtro	Con Filtro
max	180	180
min	0.1	3
promedio	36.06	38.98
mediana	25.54	27.98
desviación	33.89	33.67
N	27009	22842

Tabla 3.2: Tiempo total con y sin filtro

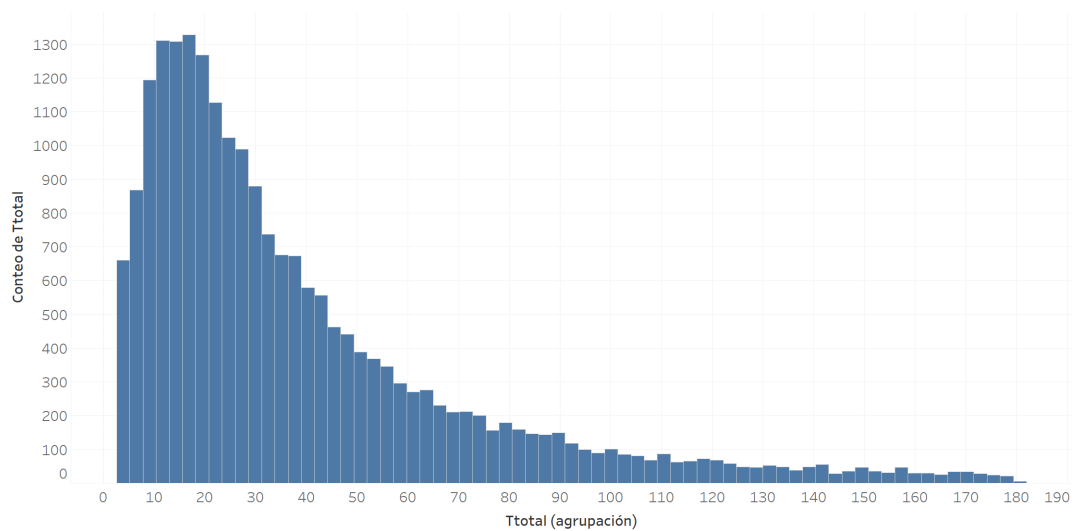


Figura 3.5: Tiempo total de observación con filtro

A partir de este resultado se vislumbra una posible forma de mejorar los filtros usados en la sección anterior. La idea sería seleccionar aquellos caminos que sean consistentes con distribuciones empíricas de variables conocidas, tales como las mostradas en esta sección.

Capítulo 4

Análisis espacial

4.1. Descripción variables

Variables espaciales

Corresponden a las coordenadas x e y asociadas a cada observación registrada. En la Figura 4.1 se muestran todas las observaciones conservadas luego de aplicar los filtros señalados en el Capítulo 2. Es posible observar que en las zonas marcadas con azul la precisión de las observaciones disminuye notablemente con respecto al resto de la sala de ventas.

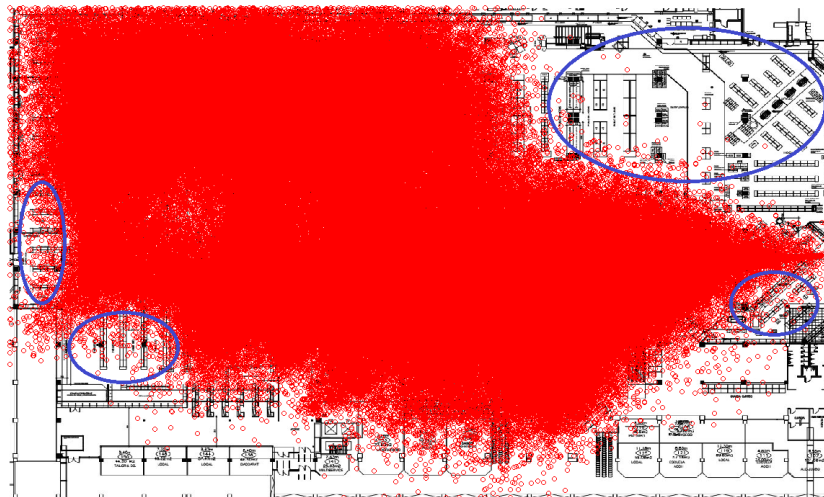


Figura 4.1: Observaciones registradas

En la Figura 4.2 se muestra la concentración de las observaciones, donde destacan 2 zonas de alta concentración. La primera es en la parte superior izquierda, que se explica debido a que es ahí donde se encuentra la panadería. Y la segunda en la parte inferior del mapa, se divide en dos, por un lado están las cajas y por el otro la entrada a la sala de ventas, zonas que deben ser visitadas por todos los clientes.

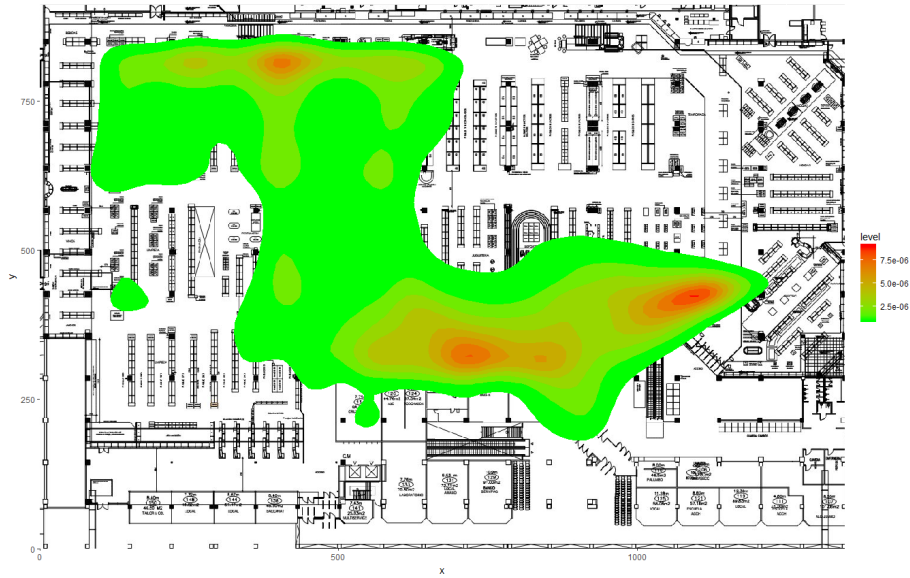


Figura 4.2: Concentración de observaciones

Variables de desplazamiento

A continuación se detallan las variables de desplazamiento distancia total y velocidad promedio.

Métrica	dtotal [m]	vel [m/min]
max	5062.5	1170.3
min	0.0	0.0
promedio	281.52	23.3
mediana	217.36	11.4
desviación	305.1	51.2

Tabla 4.1: Descripción dtotal y vel

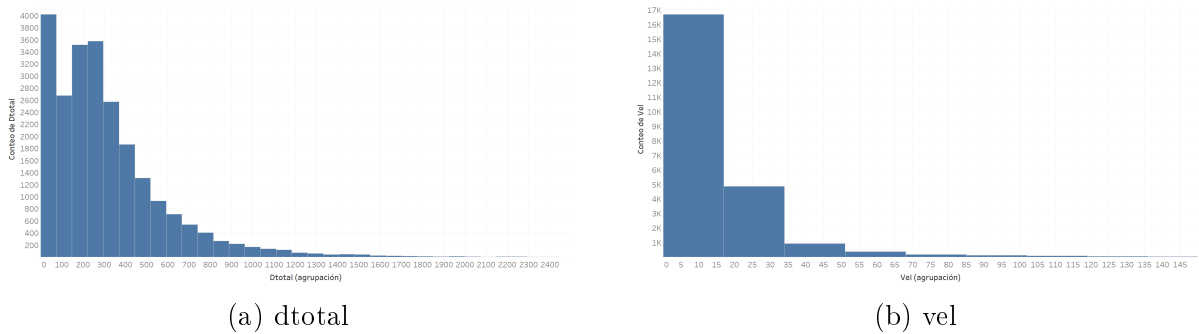


Figura 4.3: Histogramas dtotal y vel

La distancia total recorrida fue calculada como la suma de las distancias entre observacio-

nes consecutivas para cada camino registrado. Para la velocidad se dividió la distancia total por el tiempo total de registro. Debido a los puntos ciegos y a la irregularidad de las mediciones en las distintas zonas de la sala de ventas, es que la distancia total está posiblemente subestimada y por lo tanto la velocidad promedio también.

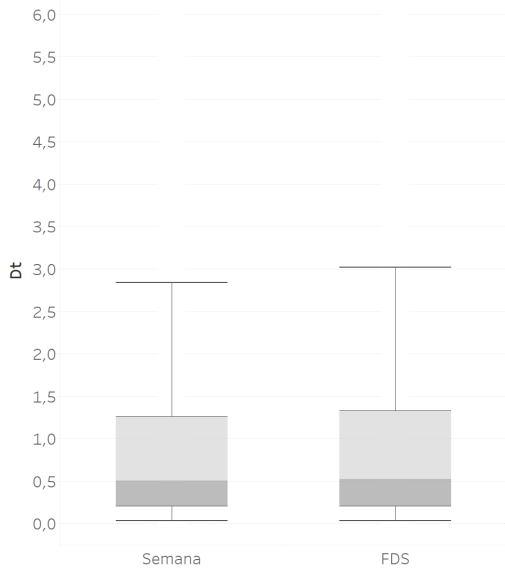
Una potencial solución a este sesgo es la corrección de trayectorias a través de algoritmos de *Map-Matching*. Para el uso de estos algoritmos se modela el espacio de medición (sala de ventas en este caso) como un grafo. Una vez hecho esto su función es dada una trayectoria medida por los sensores encontrar el camino en el grafo que más se asemeje al camino real seguido por el dispositivo medido [13].

4.2. Análisis de la muestra

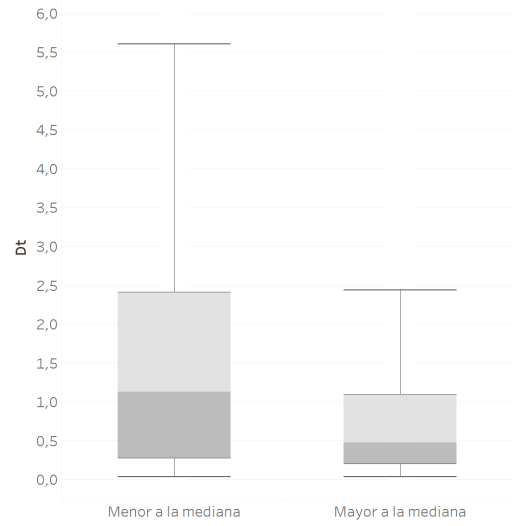
En esta sección se busca analizar si la muestra que se tiene es independiente e idénticamente distribuida (iid) para entender si es que existe algún tipo de sesgo en las mediciones. Para esto se construye la variable dt que representa el tiempo de una transición (dos observaciones consecutivas). Con esta variable se realizan test de medias dividiendo la muestra de acuerdo a distintos criterios:

- **Día:** se divide la muestra en las observaciones que ocurren en día de semana y aquellas que ocurren en fin de semana.
- **Distancia total:** se divide la muestra en dos, de acuerdo a si la distancia total del camino al que pertenece la transición es mayor o menor a la mediana.
- **Tiempo total:** se divide la muestra en dos, de acuerdo a si el tiempo total del camino al que pertenece la transición es mayor o menor a la mediana.
- **Observaciones:** se divide la muestra en dos, de acuerdo a si el número de observaciones del camino al que pertenece la transición es mayor o menor a la mediana.
- **Hora peak:** Se divide la muestra en dos, de acuerdo a si la transición ocurrió en el rango 12:00-14:00 o 18:00-21:00.

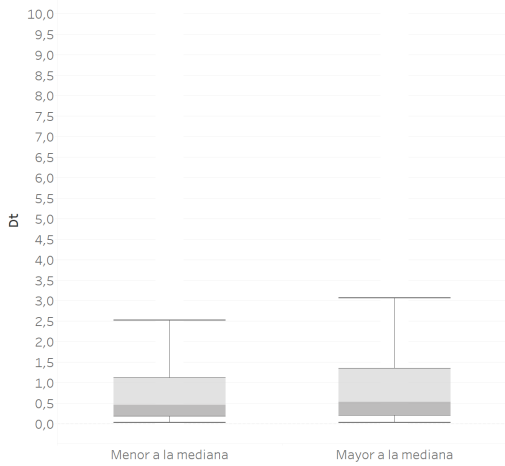
En la Figura 4.4 se compara la distribución de las distintas particiones hechas a la muestra. Se puede concluir que las particiones que presentan mayores diferencias son las de distancia total y número de observaciones.



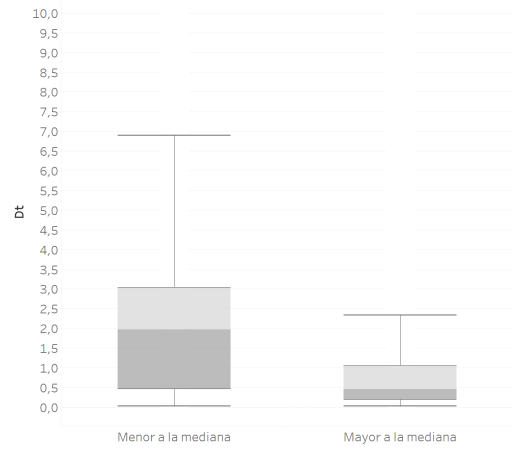
(a) División por fin de semana



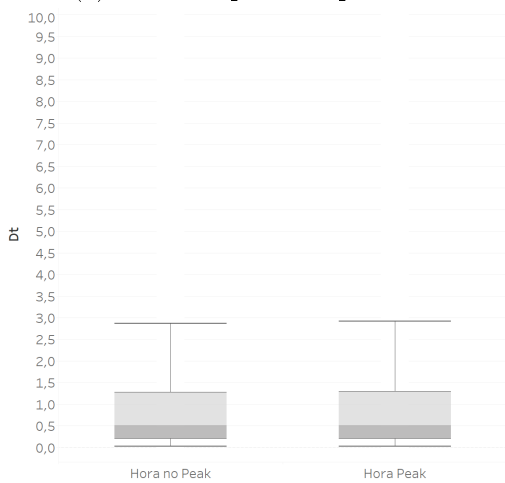
(b) División por distancias total



(c) División por tiempo total



(d) División por observaciones



(e) División por horario

Figura 4.4: Comparación submuestras

En la Tabla 4.2 se muestran los resultados obtenidos al realizar test de medias a las distintas particiones. Para el criterio de día el grupo 1 contiene las transiciones que son en día de semana mientras que el grupo 2 aquellas que ocurren en fines de semana. Para los criterios de distancia total, tiempo total y número de observaciones el grupo 1 contiene aquellas transiciones que se encuentran bajo la mediana y el grupo 2 las que se encuentran por sobre la mediana. Finalmente para el criterio de hora peak el grupo 2 contiene a las transiciones que ocurren entre 12:00-14:00 o 18:00-21:00 y el grupo 1 las que están fuera de estos rangos.

Criterio	Promedio grupo 1	Promedio grupo 2	p-valor
Semana vs Fin de Semana	1.24	1.30	0.01
Distancia total	2.82	1.05	0.00
Tiempo total	0.98	1.37	0.00
Número observaciones	3.20	1.01	0.00
Hora peak vs Hora valle	1.28	1.25	0.17

Tabla 4.2: Resultados test de medias.

De los resultados anteriores se concluye que existen diferencias significativas en todos los criterios excepto en el de hora peak. Los que presentan las mayores diferencias son por número de observaciones y distancia total respectivamente.

Para el criterio de número de observaciones, una posible explicación es que la frecuencia de medición en estos casos es menor y por lo tanto las transiciones tienden a estar más espaciadas. En el caso del criterio de distancia total es similar, ya que caminos con menor frecuencia va a tender a sub estimar la distancia total recorrida por el dispositivo y con esto el tiempo entre observaciones será mayor.

Finalmente para el criterio de tiempo total, donde se tiene que los caminos con mayor tiempo total asociado tienen mayor tiempo entre observaciones, probablemente se deba a que si un cliente gasta más tiempo en la sala de ventas es más probable que transite por zonas donde los sensores no capturan su dispositivo móvil y con esto se generan "saltos" los que aumentan el tiempo entre observaciones consecutivas.

Las mayores diferencias presentadas son bajo los criterios de números de observaciones y distancia total recorrida, ambas posiblemente afectadas por una baja frecuencia de medición. Una posible forma de solucionar este problema sería a través de la imputación de puntos en los caminos que presenten saltos i.e baja frecuencia promedio de medición, de este modo la distancia no quedaría subestimada y se aumentaría el número de puntos por camino.

4.3. Análisis de velocidad

Para estudiar el impacto que tiene la disposición física de la tienda analizará cómo varía la velocidad con que se mueven los clientes en distintas zonas de la sala de ventas, esta se separa en corredores y pasillos, tal como se ejemplifica en la Figura 4.5. Una vez hecho esto se definen 5 tipos de transiciones de acuerdo a la ubicación de puntos consecutivos en un mismo camino:

- **Pasillo-Pasillo (PP):** Transiciones que ocurren en un mismo pasillo.
- **Corredor-Corredor (CC):** Transiciones que ocurren en un mismo corredor.
- **Pasillo-Corredor (PC):** Transiciones donde una observación está en un pasillo (corredor) y la siguiente en un corredor (pasillo).
- **Pasillo distinto (P2):** Transiciones donde observaciones consecutivas están en distintos pasillos.
- **Corredor distinto (C2):** Transiciones donde observaciones consecutivas están en corredores distintos.

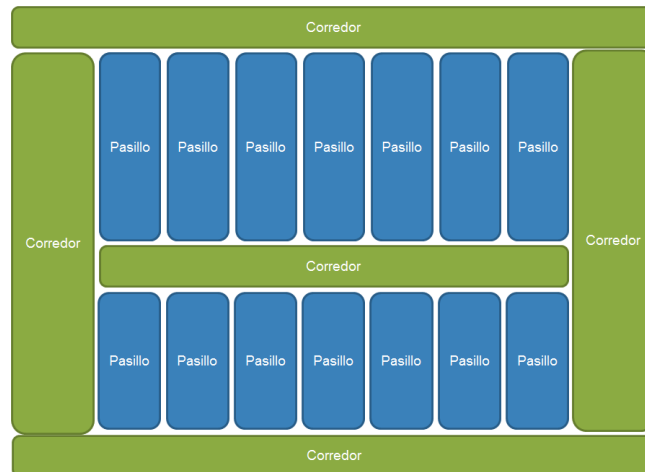


Figura 4.5: Ejemplo división sala de ventas

De la Tabla 4.3 se puede observar que en promedio la velocidad en pasillos es menor que en corredores. Esto tiene sentido ya que es en los pasillos donde se ubican la mayor cantidad de productos, mientras que los corredores son usados para traslados entre pasillos y es menor la exposición a productos.

	PP	CC	PC	P2	C2
max	773.2	1161.1	1361.4	2216.3	969.1
min	0.0	0.0	0.1	0.03	0.2
promedio	29.2	33.71	44.2	50.4	46.08
mediana	13.7	15.61	21.5	23.9	23.3
desviación	43.6	53.45	69.6	80.7	81.9

Tabla 4.3: Descripción velocidades por tipo de transición

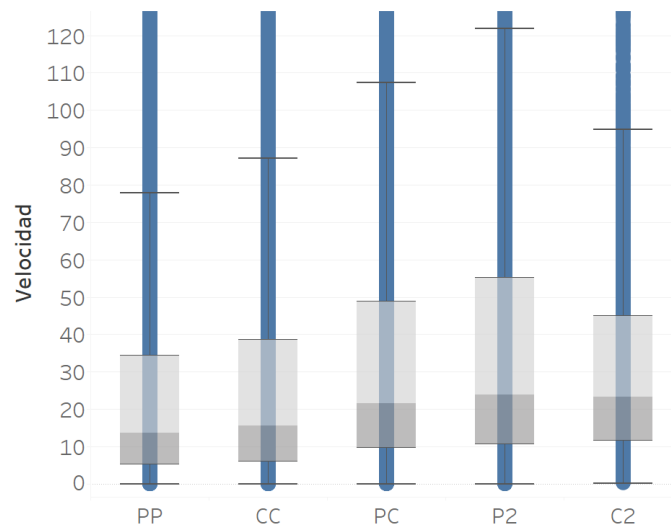


Figura 4.6: Velocidades según tipo de transición

Una variable relevante en la estimación de la velocidad es la frecuencia de medición. Tal como se muestra en la Figura 4.7 existe una correlación positiva entre estas dos variables. Esto se debe principalmente a que si por ejemplo se tienen dos caminos con mismo inicio y final en un mismo rango de tiempo como en la Figura 4.8. El segundo camino que tiene mayor frecuencia y por lo tanto mas observaciones en el mismo rango de tiempo se le atribuirá una mayor distancia recorrida en el mismo rango de tiempo y por la tanto una mayor velocidad.

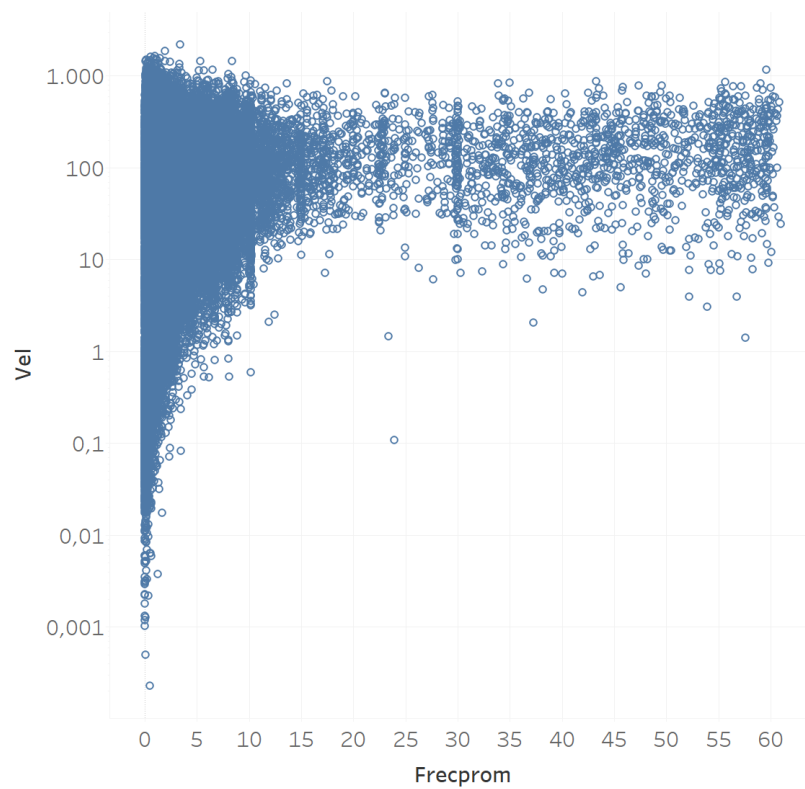


Figura 4.7: Relación velocidad frecuencia



Figura 4.8: Efecto de la frecuencia en la velocidad

Nuevamente una forma en que se podría corregir este sesgo en que los caminos con mayor frecuencia presentan mayor velocidad es imputando observaciones a los caminos de modo que estos queden con una frecuencia promedio regular. Para esto es importante que las observaciones imputadas respeten las restricciones espaciales de la sala de ventas y sean coherentes con las trayectorias observadas.

4.4. Testeo de hipótesis de comportamiento

Una vez definidas los tipos de transición se busca testear distintas hipótesis de comportamiento sobre la velocidad en que los clientes se mueven al interior de la sala de ventas. Se definen las siguientes hipótesis.

H1: Transiciones de mayor distancia son realizadas a mayor velocidad.

Esta hipótesis busca capturar el hecho de que si un cliente sabe que tiene que recorrer una distancia mayor para encontrar un determinado producto lo hará más rápido que si dicho producto se encuentra más cerca.

H2: Clientes que recorren más distancia lo hacen a una mayor velocidad.

Con esta hipótesis se busca validar que los clientes con mayor recorrido de la sala de ventas tienen una velocidad mayor. Esto debido a que un mayor recorrido está asociado a una canasta mas grande, por lo tanto es esperable que clientes con mayores compras pierdan menos tiempo comparando productos.

H3: Clientes que visitan el supermercado el fin de semana son más lentos que los que visitan en día de semana. En particular se espera que clientes que visiten la sala de ventas los fines de semana tengan una menor velocidad debido a que cuentan con mayor tiempo disponible para realizar sus compras.

Se busca verificar si es que existe una diferencia en el comportamiento de los clientes de

los fines de semana con los de semana en cuanto a la velocidad que recorren la sala de ventas.

H4: En horarios de mayor congestión los clientes se mueven mas lento.

Se busca verificar si es que la congestión tiene un efecto en la velocidad en que se mueven los clientes. Para poder cuantificar la congestión se construye una variable que para cada camino registrado entrega la cantidad de clientes que han entrado a la sala en un rango de una hora desde que se registra la primera observación del camino.

H5: Clientes se mueven más lento en pasillos que en corredores.

Esta hipótesis busca verificar que los clientes en los pasillos se encuentran evaluando los distintos productos, con lo que su velocidad disminuye, mientras que los corredores son usados como conexión entre pasillos y por lo tanto su velocidad aumenta en estos.

H6: Clientes que están más tiempo en la sala de ventas son más lentos.

Se busca verificar si es que la velocidad de los clientes esta correlacionada con el tiempo que estos permanecen dentro de la sala de ventas. En este caso existen dos posibilidades, la primera es que los clientes que permanecen más tiempo es debido a mayores recorridos (posiblemente canastas con más productos) y por lo tanto no necesariamente presentan menor velocidad. La segunda es que los clientes que permanecen más tiempo no tienen necesariamente recorridos más largos asociados y por lo tanto su tiempo de permanencia se debe a una menor velocidad.

H7: Mientras mayor es el tiempo restante para abandonar la sala de ventas menor es la velocidad.

Se busca capturar la idea de que los clientes cuentan con un tiempo estimado a gastar dentro de la sala de ventas, y a medida que este tiempo se va agotando comienzan a aumentar su velocidad para cumplir con el tiempo presupuestado. Esto se puede ver como una versión de la teoría de *goal-gradient* que postula que a medida que los clientes se acercan a una determinada meta, estos tienden a aumentar su esfuerzo en llegar a esta [6], lo que en este contexto se traduce en un aumento de la velocidad. Además esta hipótesis se separa en fin de semana y día de semana, para constatar si este efecto difiere entre estos tipos de días.

Para testear estas hipótesis se definen dos modelos lineales cuya única diferencia es que el primero no considera efectos fijos a nivel individual y el segundo si.

$$v_{ic} = \beta_0 + \beta_1 \text{frec}_c + \beta_2 \text{dtotal}_c + \beta_3 \text{tttotal}_c + \beta_4 \text{FDS}_i + \beta_5 d_i + \beta_6 \text{tipo}_i + \beta_7 \text{TimetoleaveFDS}_i + \beta_8 \text{TimetoleaveSEMANA}_i + \beta_9 \text{congestion}_c \quad (4.1)$$

$$v_{ic} = \beta_c + \beta_1 \text{frec}_c + \beta_2 \text{dtotal}_c + \beta_3 \text{tttotal}_c + \beta_4 \text{FDS}_i + \beta_5 d_i + \beta_6 \text{tipo}_i + \beta_7 \text{TimetoleaveFDS}_i + \beta_8 \text{TimetoleaveSEMANA}_i + \beta_9 \text{congestion}_c \quad (4.2)$$

Donde:

- v_{ic} : velocidad de la transición i perteneciente al camino c .
- $frec_c$: frecuencia promedio del camino c .
- $dtotal_c$: distancia total registrada por el camino c .
- $ttotal_c$: tiempo total registrado por el camino c .
- FDS_i : indica si es que la transición fue realizada en fin de semana.
- d_i : distancia de la transición i .
- $tipo_i$: tipo de transición (PP, CC, PC, P2 o C2). En los modelos de regresión se omiten las que son del tipo CC, para no tener problemas de multicolinealidad.
- $Timetoleave_i$: diferencia entre la hora de la última observación del camino al que pertenece la transición i y la hora a la que ocurre i . En los modelos se utiliza $TimetoleaveFDS_i$ y $TimetoleaveSEMANA_i$ cuando la transición ocurre en fin de semana o fin de semana respectivamente. El objetivo de esto es poder diferenciar este patron para distintos tipos de días.
- $congestion_c$: número de clientes registrados en un intervalo de una hora desde que ocurre la primera observación de c .

Debido a los potenciales sesgos detectados anteriormente, los modelos se estimarán sobre 4 conjuntos de datos para de este modo poder analizar la consistencia de los resultados.

- **Conjunto 1:** Toda la muestra.
- **Conjunto 2:** Se consideran solo las transiciones pertenecientes a caminos que están en la mitad superior de la variable distancia total.
- **Conjunto 3:** Se consideran solo las transiciones pertenecientes a caminos que están en la mitad superior de la variable número de observaciones.
- **Conjunto 4:** Se consideran solo las transiciones pertenecientes a caminos cuyo mayor tiempo entre observaciones es menor $\bar{dt} + 2\sigma_{dt}$.

En las Tablas 4.4 y 4.5 se resumen los resultados obtenidos al estimar los modelos 4.1 y 4.2 respectivamente. A continuación se analizan los resultados obtenidos para las distintas hipótesis de comportamiento planteadas.

H1: Transiciones de mayor distancia son realizadas a mayor velocidad.

Esta hipótesis se válida para todas las regresiones realizadas excepto para la que usa el conjunto 4 de datos en el modelo de efecto fijo a nivel individual. En este caso si bien el coeficiente tiene el signo esperado, no logra ser estadísticamente significativo.

H2: Clientes que recorren más distancia lo hacen a una mayor velocidad.

En este caso la hipótesis se válida para ambos modelos en todos los conjuntos de datos usados para la estimación.

H3: Clientes que visitan el supermercado el fin de semana son más lentos que los que lo visitan en día de semana .

Para esta hipótesis los resultados difieren considerablemente entre el modelo con y sin efecto fijo. En el modelo sin efecto fijo solo resulta ser significativa la variable para el conjunto 4 de datos, aunque en todos los otros conjuntos se obtiene el signo deseado. Mientras que al incluir efectos fijos por individuo la variable no resulta ser significativa en ningún caso y su signo varía según el conjunto de datos usados en la estimación. Es posible que esto se deba a que la decisión de que día visitar el supermercado está siendo capturada por el efecto fijo con lo que la variable resulta no significativa.

H4: En horarios de mayor congestión los clientes se mueven más lento.

No se logra verificar esta hipótesis para ninguno de los modelos estimados, siendo no significativa en todas las estimaciones realizadas.

H5: Clientes se mueven más lento en pasillos que en corredores.

Se logra verificar la hipótesis para todos los modelos planteados, siendo esta significativa al 0.1 % en todos los modelos.

H6: Clientes que están más tiempo en la sala de ventas se mueven más lento.

En esta hipótesis se tienen resultados contradictorios. En ambos modelos para los tres primeros conjuntos de datos se verifica lo contrario a lo planteado en la hipótesis y en el conjunto 4 de datos se verifica la hipótesis. Esto pueda deberse a existencia de heterogeneidad entre los clientes con respecto a esta variable, donde algunos que gastan más tiempo es debido a una mayor canasta y otros a menor velocidad.

H7: Mientras mayor el tiempo restante para abandonar la sala de ventas menor es la velocidad.

En el caso del modelo sin efecto fijo la hipótesis se válida para los primeros tres conjuntos de datos, sin mostrar diferencias significativas en el caso de fin de semana y día de semana. Para el modelo de efecto fijo se tiene que la hipótesis se valida en el caso de fin de semana para los tres primeros conjuntos de datos mientras que en el caso de semana solo para el primero.

Variable	Conjunto 1	Conjunto 2	Conjunto 3	Conjunto 4
<i>Intercepto</i>	-0.11 (***)	-0.09 (***)	-0.08 (***)	-0.18 (***)
<i>frec</i>	0.33 (***)	0.41 (***)	0.4 (***)	0.31 (***)
<i>dtotal</i>	0.06 (***)	0.03 (***)	0.033 (***)	0.09 (***)
<i>FDS</i>	-0.002	-0.001	-0.004	-0.01 (.)
<i>d</i>	0.09 (***)	0.13 (***)	0.15 (***)	0.12 (***)
<i>Timetoleave_FDS</i>	-0.02 (***)	-0.019 (***)	-0.019 (***)	0.001
<i>Timetoleave_SEMANA</i>	-0.02 (***)	-0.02 (***)	-0.019 (***)	0.008
<i>congestion</i>	0.002	0.002	0.003	-0.002
<i>ttotal</i>	0.01 (.)	0.016 (***)	0.015 (***)	-0.11 (***)
<i>PC</i>	0.11 (***)	0.11 (***)	0.1 (***)	0.11 (***)
<i>C2</i>	0.11 (***)	0.1 (***)	0.1 (***)	0.11 (***)
<i>P2</i>	0.18 (***)	0.18 (***)	0.17 (***)	0.18 (***)
<i>PP</i>	-0.05 (***)	-0.05 (***)	-0.04 (***)	-0.05 (***)
R^2	0.068	0.077	0.081	0.095
<i>N</i>	173.821	153.159	154.713	115.844

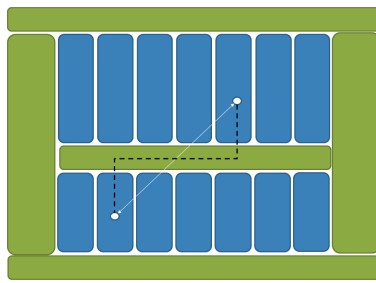
Tabla 4.4: Estimación modelo 4.1.

Variable	Conjunto 1	Conjunto 2	Conjunto 3	Conjunto 4
<i>frec</i>	0.32 (***)	0.41 (***)	0.4 (***)	0.34 (***)
<i>dtotal</i>	0.06 (***)	0.04 (***)	0.04 (***)	0.08 (***)
<i>FDS</i>	-0.005	0.008	0.012	-0.028
<i>d</i>	0.01 (***)	0.13 (***)	0.15 (***)	0.13
<i>Timetoleave_FDS</i>	-0.02 (**)	-0.019 (*)	-0.019 (*)	-0.002
<i>Timetoleave_SEMANA</i>	-0.02 (*)	-0.011	-0.01	0.03
<i>congestion</i>	-0.003	-0.005	-0.004	0.009
<i>ttotal</i>	0.003	0.015 (*)	0.013 (.)	-0.13 (***)
<i>PC</i>	0.1 (***)	0.09 (***)	0.09 (***)	0.08 (***)
<i>C2</i>	0.1 (**)	0.1 (***)	0.08 (***)	0.065
<i>P2</i>	0.17 (***)	0.16 (***)	0.15 (***)	0.16 (***)
<i>PP</i>	-0.06 (***)	-0.06 (***)	-0.06 (***)	-0.09 (***)
<i>R²</i>	0.068	0.077	0.08	0.094
<i>N</i>	173.821	153.159	154.713	115.844

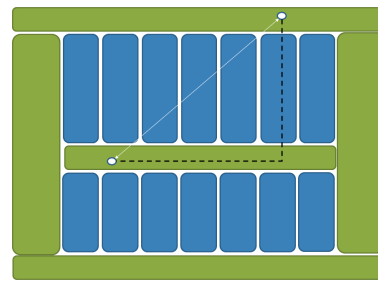
Tabla 4.5: Estimación modelo 4.2

4.5. Corrección de trayectorias

Los caminos capturados en general presentan transiciones que no son consistentes con patrones de movimiento de clientes, por ejemplo, existen “saltos” entre pasillos y corredores. Para tener una aproximación más realista del movimiento de los clientes es que se corrigen algunos tipos de transiciones bajo un criterio de camino mínimo factible, en donde a las transiciones que son entre pasillos y corredores no adyacentes se les atribuye la distancia del camino mínimo que une las observaciones pero respetando las restricciones espaciales de la sala de ventas tal como es muestra en la Figura 4.9.



(a) Pasillos no adyacentes.



(b) Corredores no adyacentes.

Figura 4.9: Tipos de transiciones corregidas

En total se corrigen 12.143 correcciones correspondientes al 7.03 % de la base de datos usada en las estimaciones anteriores. En promedio las velocidades de las transiciones corregidas aumentan en un 12.4 % reportando como máximo 41.4 % de aumento. Sobre estos datos se estiman los mismos modelos planteados en la sección anterior. En las Tablas 4.6 y 4.7 se muestra el detalle de los resultados obtenidos.

En general no se detectan grandes diferencias en los resultados salvo en dos casos. El primero es en el modelo con efecto fijo a nivel individual donde la variable *Timetoleave_SEMANA* pasa de ser significativa solamente bajo el conjunto 1 a serlo en los cuatro conjuntos de datos. Obteniendo el signo de esperado en todas las estimaciones, excepto en la que se realiza sobre el conjunto 4.

El segundo cambio detectado es también en el modelo de efecto fijo a nivel individual en la variable *ttotal* donde al agregar la corrección de velocidad esta variable pasa a ser significativa bajo los 4 conjuntos de datos. El segundo caso en que varían

Variable	Todos	Conjunto 1	Conjunto 2	Conjunto 3
<i>Intercepto</i>	-0.11 (***)	-0.1 (***)	-0.09 (***)	-0.17 (***)
<i>frec</i>	0.34 (***)	0.38 (***)	0.38 (***)	0.33 (***)
<i>dtotal</i>	0.047 (***)	0.033 (***)	0.032 (***)	0.07 (***)
<i>FDS</i>	-0.001	-0.001	-0.003	-0.008
<i>d</i>	0.09 (***)	0.13 (***)	0.14 (***)	0.12 (***)
<i>Timetoleave_FDS</i>	-0.018 (***)	-0.018 (***)	-0.018 (***)	0.0024
<i>Timetoleave_SEMANA</i>	-0.02 (***)	-0.019 (***)	-0.018 (***)	0.009
<i>congestion</i>	0.001	0.002	0.002	-0.024
<i>ttotal</i>	0.01 (***)	0.017 (***)	0.016 (***)	-0.1 (***)
<i>PC</i>	0.11 (***)	0.1 (***)	0.1 (***)	0.11 (***)
<i>C2</i>	0.11 (***)	0.1 (***)	0.1 (***)	0.11 (***)
<i>P2</i>	0.18 (***)	0.17 (***)	0.17 (***)	0.18 (***)
<i>PP</i>	-0.044 (***)	-0.044 (***)	-0.041 (***)	-0.05 (***)
R^2	0.068	0.076	0.078	0.096
N	173.821	153.159	154.713	115.844

Tabla 4.6: Estimación modelo 4.1 con corrección de velocidad.

Variable	Conjunto 1	Conjunto 2	Conjunto 3	Conjunto 4
<i>frec</i>	0.33 (***)	0.39 (***)	0.38 (***)	0.33 (***)
<i>dtotal</i>	0.06 (***)	0.036 (***)	0.038 (***)	0.07 (***)
<i>FDS</i>	-0.001	0.012	0.01	-0.03
<i>d</i>	0.09 (***)	0.13 (***)	0.15 (***)	0.13 (***)
<i>Timetoleave_FDS</i>	-0.02 (*)	-0.016 (***)	-0.017 (***)	0.0019
<i>Timetoleave_SEMANA</i>	-0.016 (*)	-0.01 (***)	-0.009 (***)	0.035 (*)
<i>congestion</i>	-0.006	-0.01	-0.01	0.1
<i>ttotal</i>	0.008	0.015	0.01	-0.13 (***)
<i>PC</i>	0.09 (***)	0.09 (***)	0.09 (***)	0.08 (***)
<i>C2</i>	0.86 (*)	0.09 (*)	0.07	0.06
<i>P2</i>	0.16 (***)	0.16 (***)	0.15 (***)	0.16 (***)
<i>PP</i>	-0.06 (***)	-0.05 (***)	-0.05 (**)	-0.08 (***)
R^2	0.068	0.076	0.078	0.095
N	173.821	153.159	154.713	115.844

Tabla 4.7: Estimación modelo 4.2 con corrección de velocidad.

Capítulo 5

Brechas entre visitas y compras

En este capítulo se muestra un ejemplo de una posible aplicación de esta nueva fuente de información. El principal objetivo es identificar brechas entre ventas y visitas en distintas zonas de la sala de ventas. Esto es potencialmente útil para tomar decisiones de localización de productos.

Para seleccionar los caminos que se utilizarán en el análisis se observan la Figura 2.4 buscando obtener los caminos que tengan mayor recorrido por la sala. Es por esto que se descartan todo los caminos que tengan menos de 25 puntos, ya que de este modo se consigue una frecuencia promedio con un rango que va desde 0.14 a 11.22 (Tabla 5.1) ,obteniendo así clientes que presentan un mayor recorrido a través de la sala de ventas. Al aplicar este filtro se conservan 6.079 dispositivos móviles en 334.205 registros.

Métrica	ob	frec
max	656	11.22
min	25	0.14
promedio	51.64	1.35
mediana	40	1.05
desviación	37.65	1.11

Tabla 5.1: Descripción ob y frec

5.1. Comparación datos transaccionales y de posición

Se cuenta con la información transaccional del supermercado para las mismas fechas en que se han tomado las mediciones de posición de clientes (2 - 15 de Noviembre 2016). En este periodo se registran 115.456 boletas, las que contienen 29.015 productos distintos. A estos productos se les asignan zonas de acuerdo al nivel de agregación de la Figura 2.5. En total se lograron asignar 25.592 productos con lo que se conservan 102.059 canastas.

En la Tabla 5.2 se compara el porcentaje de visitas y compras para cada zona. Se considera como visita a una zona cuando una observación se encuentra dentro de los límites de la zona definida.

Los números que se encuentran en rojo son las zonas que presentan mayor porcentaje de compras que visitas, esto probablemente se debe a errores en la asignación de los productos las zonas, a problemas de los sensores en la detección de dispositivos móviles en ciertas zonas o a exclusión de trayectorias relevantes en los filtros usados. Estas diferencias entre visitas y compras entregan una primera aproximación de la conversión que tienen las zonas definidas.

Zona	Transaccional	Visitas	Diferencia
1	29.45 %	7.57 %	-21.88 %
2	14.86 %	36.23 %	21.36 %
3	25.22 %	62.11 %	36.89 %
4	0.08 %	32.10 %	32.02 %
5	10.62 %	39.83 %	29.22 %
6	25.41 %	60.28 %	34.88 %
7	3.07 %	31.10 %	28.03 %
8	11.11 %	28.46 %	17.35 %
9	13.49 %	31.79 %	18.30 %
10	7.30 %	31.75 %	24.46 %
11	9.88 %	26.62 %	16.74 %
12	10.61 %	23.81 %	13.20 %
13	9.99 %	26.24 %	16.25 %
14	39.78 %	14.58 %	-25.21 %
15	14.86 %	22.77 %	7.91 %
16	23.28 %	48.24 %	24.96 %
17	7.64 %	40.05 %	32.41 %
18	5.36 %	61.06 %	55.70 %
19	1.13 %	32.72 %	31.60 %
20	8.88 %	73.56 %	64.68 %
21	5.89 %	6.91 %	1.02 %
22	1.57 %	9.51 %	7.94 %
23	1.22 %	18.28 %	17.06 %
24	2.38 %	16.55 %	14.17 %
25	45.16 %	26.59 %	-18.57 %

Tabla 5.2: Comparación compras y visitas

5.2. Reglas de asociación de visitas

Se levantaron reglas de asociación de visitas a zonas de acuerdo a la taxonomía:

$$\{Visitazona i\} \rightarrow \{Visitazona j\}$$

En total se generaron 452 reglas. En la tabla 5.3 se muestran las reglas con mayor soporte y confianza. En general son reglas entre las zonas más visitadas de la sala de ventas.

Regla	Soporte	Confianza	Lift
jyc→vestuario	0.47	0.77	1.05
dulces→fyv	0.44	0.73	1.17
jyc→fyv	0.42	0.69	1.11
dulces→jyc	0.40	0.66	1.08
farmacia→dulces	0.34	0.71	1.18

Tabla 5.3: Reglas con mayor soporte

En la tabla 5.4 se muestran las reglas con mayor Lift. Estas son entre zonas adyacentes que en general tienen bajo porcentaje de visitas. La interpretación de este resultado, es que las zonas presentes en estas reglas tienen una muy alta correlación en cuanto a visitas conjuntas.

Regla	Soporte	Confianza	Lift
mascotas→ferretería	0.03	0.48	5.06
ferretería→mascotas	0.03	0.35	5.06
mascotas→automovil	0.04	0.61	3.33
ferretería→automóvil	0.05	0.57	3.13
automóvil→ferretería	0.05	0.30	3.13

Tabla 5.4: Reglas con mayor Lift

En la Tabla 5.5 se muestran las reglas que tienen Lift menor que 1. Esto en el contexto de análisis de canasta se interpreta como productos sustitutos o que no tienden a comprarse juntos (ejm : mantequilla y margarina). Para este caso se puede ver que estas reglas corresponden a zonas que son no adyacentes y en general mezclan tipos de productos que no son complementarios.

Regla	Soporte	Confianza	Lift
ferretería→deportes	0.03	0.29	0.88
ferretería→ongelados	0.03	0.34	0.93
líquidos→vestuario	0.10	0.69	0.93
electro→cereales	0.04	0.27	0.94
mascotas→vestuario	0.05	0.70	0.95

Tabla 5.5: Reglas con Lift<1

5.3. Reglas de asociación de transaccionales

En la tabla 5.6 se muestran las reglas transaccionales con mayor soporte. Análogamente al caso anterior, donde las reglas con mayor soporte correspondían a las zonas de mas visitas,

en este caso son reglas entre las zonas que mas trasacciones registran.

Regla	Soporte	Confianza	Lift
panaderia→liquidos	0.20	0.45	1.14
lacteos→panaderia	0.18	0.60	1.32
lacteos→liquidos	0.16	0.53	1.33
panaderia→fyv	0.15	0.34	1.34
dulces→panadería	0.14	0.54	1.19

Tabla 5.6: Reglas transaccionales con mayor Soporte

En la tabla 5.7 se muestran las reglas con mayor Lift. En este caso son todas entre zonas que se encuentran relativamente cerca y contienen productos similares, todos son del tipo abarrotes. Este resultado tiene sentido debido a la complementariedad de los productos presentes en estas zonas.

Regla	Soporte	Confianza	Lift
aliños→tallarines	0.05	0.49	5.01
tallarines→aliños	0.05	0.53	5.01
tallarines→abarrotes	0.03	0.35	4.85
abarrotes→aliños	0.04	0.50	4.74
conservas→tallarines	0.01	0.42	4.26

Tabla 5.7: Reglas transaccionales con mayor Lift

Finalmente en la tabla 5.8 se muestran las reglas con $Lift < 1$. En este caso corresponde a tipos de productos no complementarios.

Regla	Soporte	Confianza	Lift
vestuario→panaderia	0.03	0.32	0.72
vestuario→liquidos	0.03	0.30	0.77
jyc→lacteos	0.01	0.26	0.87
vestuario→lacteos	0.02	0.26	0.89
libreria→panaderia	0.03	0.41	0.90

Tabla 5.8: Reglas transaccionales con $Lift < 1$

Al comparar ambos tipos de reglas de acuerdo con los criterios anteriores. Se puede observar que para el caso de las reglas con mayor soporte se tienen como elemento común que ambos casos son reglas que corresponden a las zonas más visitadas y más compradas de acuerdo con el tipo de datos usados (datos móviles o transaccionales). Al comparar las reglas con mayor Lift se tiene que en ambos casos zonas adyacentes y de productos complementarios. Finalmente, en el caso de las reglas con menor Lift, se tiene lo contrario a lo anterior, productos no complementarios y de zonas no adyacentes. Cabe destacar que si bien es posible hacer este paralelo entre ambos tipos de reglas, en general no coinciden el antecedente y consecuente resultante entre ambos tipos de datos.

5.4. Comparación reglas

Se escogen aquellas reglas que presenten igual antecedente y consecuente tanto en las transaccionales como de visitas, con esto quedan 139 reglas seleccionadas, las que se analizan a continuación de acuerdo a las brechas generadas en los indicadores de soporte y confianza.

En la Figura 5.1 se compara el soporte de las reglas transaccionales con las reglas de visitas. Los puntos que están sobre la diagonal corresponden a pares de zonas entre las que ocurren mas transacciones conjuntas que visitas conjuntas. Esta inconsistencia puede deberse a lo mismo que las detectadas en las brechas de las zonas (Tabla 5.2). Puede ser por errores en la asignación de productos a zonas y/o problemas de medición de visitas en ciertas zonas.

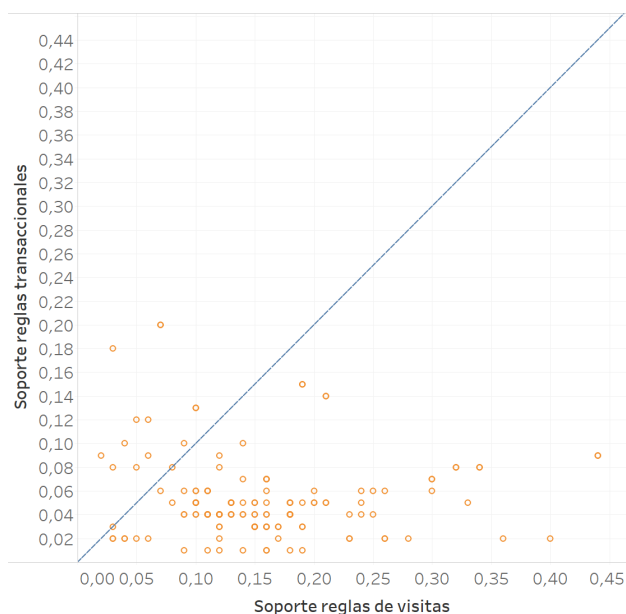


Figura 5.1: Comparación soporte reglas transaccionales y de visitas

Regla	Soporte visitas	Soporte transacciones	Brecha
jyc→dulces	0.4	0.02	0.38
fyv→dulces	0.44	0.09	0.35
vestuario→farmacia	0.36	0.02	0.34
farmacia→dulces	0.34	0.08	0.26

Tabla 5.9: Mayores brechas de soporte

El identificar estas brechas puede ayudar a mejorar la oferta del supermercado y a potenciar oportunidades de venta cruzada dependiendo del tipo de productos de la zona. Por ejemplo, la regla con la mayor brecha de las presentadas en la Tabla 5.9 es entre “juguetes y cumpleaños” y “dulces”. Según esta regla ambas zonas se visitan mucho en conjunto, pero no se compra mucho en conjunto, siendo que tienen algún tipo de relación en los productos que venden, relacionados con el mundo infantil. Por otro lado, hay reglas que no resultan tan

útiles debido a la diferencia del tipo de productos, por ejemplo, la que relaciona “vestuario” y “farmacia”. Otro punto interesante en este cruce de reglas es que se aprecia que las zonas que más aparecen son “farmacia” y “dulces”, esto puede deberse a que estas zonas están siendo usadas como pasillo, sin que los clientes tengan intención de compra al interior de estas. Nuevamente esto presenta una oportunidad para localizar productos atractivos en estas zonas de modo de aprovechar el tránsito de clientes.

Analogamente a lo anterior en la Figura 5.2 se compara la confianza de las reglas. Los puntos que se encuentran bajo la diagonal indican que la probabilidad de visitar la zona consecuente dado que se visitó la antecedente es mayor que comprar en la consecuente dado que se compró en la antecedente. Este punto presenta una potencial oportunidad para localizar productos complementarios al interior de la sala de ventas. En la tabla 5.10 se muestran las zonas con mayores brechas en Confianza.

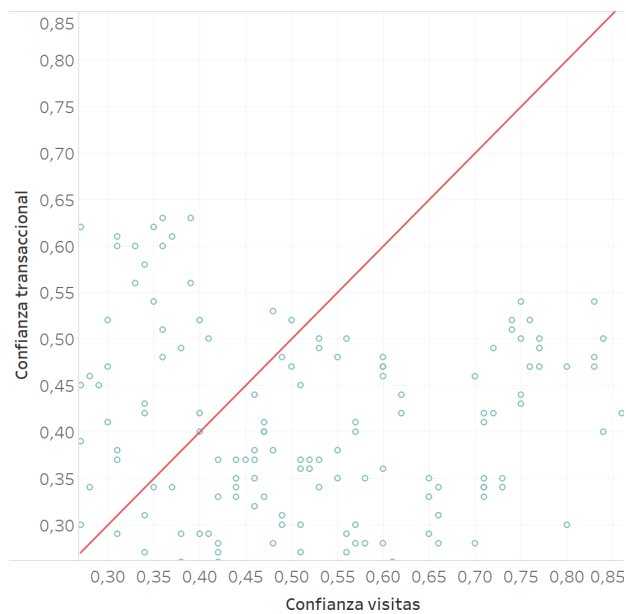


Figura 5.2: Comparación confianza reglas transaccionales y de visitas

Regla	Confianza visitas	Confianza transacciones	Brecha
panadería→dulces	0.8	0.30	0.50
mascotas→farmacia	0.84	0.40	0.44
librería→fyv	0.7	0.28	0.42
dulces-→fyv	0.72	0.34	0.39

Tabla 5.10: Mayores brechas de confianza

El identificar estas brechas tiene el potencial beneficio de mejorar la localización de productos al interior de la sala de ventas, ya que al conocer la probabilidad de visitar la zona consecuente se pueden localizar productos o promociones en esta que sean más atractivos dado la zona antecedente que se visitó.

En la presente sección se ha mostrado que existen brechas entre las visitas a las categorías y las compras que se realizan en estas. Si bien puede que estas estén sobre estimadas debido a la

muestra de datos usadas, entrega una posible aplicación de esta nueva fuente de información, donde se permite generar indicadores de conversión a nivel de categorías que relacionen visitas con compras.

Capítulo 6

Conclusiones y trabajos futuros

En el presente trabajo se han estudiado los datos de las trayectorias seguidas por los clientes al interior de un supermercado a través de tres aristas: temporal, espacial y relacional. Además de mostrar posibles filtros para capturar las observaciones que corresponden a clientes del supermercado.

Del pre-procesamiento de datos es posible concluir que es necesario diseñar filtros que permitan separar las observaciones de clientes de los que no lo son, debido a que los sensores no son capaces de realizar esa diferenciación. La principal dificultad detectada en los filtros es que debe haber un balance entre la exigencia y el volumen de datos esperado, ya que se constató que la degradación de los datos es muy fuerte a medida que se les exige mayor calidad a los datos. En particular se tiene que a medida que se exigen mayor cantidad de observaciones por caminos la frecuencia de medición cae fuertemente.

En el análisis temporal se logra constatar que los patrones detectados por los sensores móviles son consistentes con los datos transaccionales. La única diferencia que presentan es el volumen de datos, siendo los datos móviles de mucho menor volumen. Nuevamente esto se debe a los filtros aplicados, ya que estos hacen que se pierda gran cantidad de datos. De esta sección surge la posibilidad de un nuevo criterio para aplicar filtros, donde se seleccione caminos de modo que la distribución de sus variables temporales tiempo de llegada y visitas por día coincida con la distribución de las variables análogas obtenidas a partir de datos transaccionales. Esto podría en combinación con los filtros anteriores podría ayudar a obtener mejores caminos, pero no hay claridad el efecto que tendría en el volumen de datos resultante.

Luego en el análisis espacial se constata que la muestra presenta diferencias significativas en el tiempo entre observaciones al dividir la muestra de acuerdo con distintas variables tales como: tiempo total registrado, número de observaciones por camino, distancia total recorrida y día de la semana. Si bien las diferencias son significativas, en algunos casos como cuando se compara las mediciones de fin de semana con las de día de semana, estas son pequeñas y no representan mayores problemas. Pero en el caso de las otras variables se detectan diferencias más grandes que pueden traer problemas al momento de hacer inferencia estadística a partir de los datos. Cabe destacar que la única partición que no presenta diferencias significativas

es la que divide las observaciones que fueron realizadas en horario peak y las que no. Con esto se muestra que no existe interferencia entre dispositivos al momento de las mediciones.

La detección de problemas en la captura de datos permite plantear distintas soluciones. Una es aplicar más filtros a los datos de modo que la muestra sea uniforme. El principal problema de esto es que al ser más exigente con los filtros el volumen de datos disminuirá aún más. Por esto en la medida que no se cuente con un mayor volumen de datos inicial, esta opción no parece ser la más viable. La segunda opción es agregar un paso más en el pre procesamiento de datos, donde a través de algoritmos de map matching se puedan regularizar las observaciones imputando puntos a las distintas trayectorias observadas. De este modo se lograría una frecuencia de medición uniforme.

Al testear hipótesis de comportamiento se pudo constatar que hay diferencias significativas en la velocidad que los clientes recorren la sala de ventas de acuerdo a distintas variables, tales como tiempo total de observación, distancia total, tiempo restante para salir, etc. Potencialmente se pueden extender los análisis anteriores diferenciado no solo entre pasillos y corredores, sino que clasificando los pasillos de acuerdo con la velocidad que son transitados como medida de la atención que prestan los clientes en estos y el nivel de involucramiento que tienen al comprar en las categorías presentes en estos.

Finalmente, al analizar cómo se relacionan las compras con las visitas a las categorías se constata que si existen brechas en estas. Si bien las brechas pueden estar siendo sobre estimadas debido a la exigencia de los filtros usados, plantean una posible línea de trabajo que genere indicadores de conversión que relacionen visitas y compras. Esto actualmente no es posible realizarlo, ya que los supermercados solo cuentan con información de los productos vendidos mas no de las zonas visitadas que no derivaron en compras.

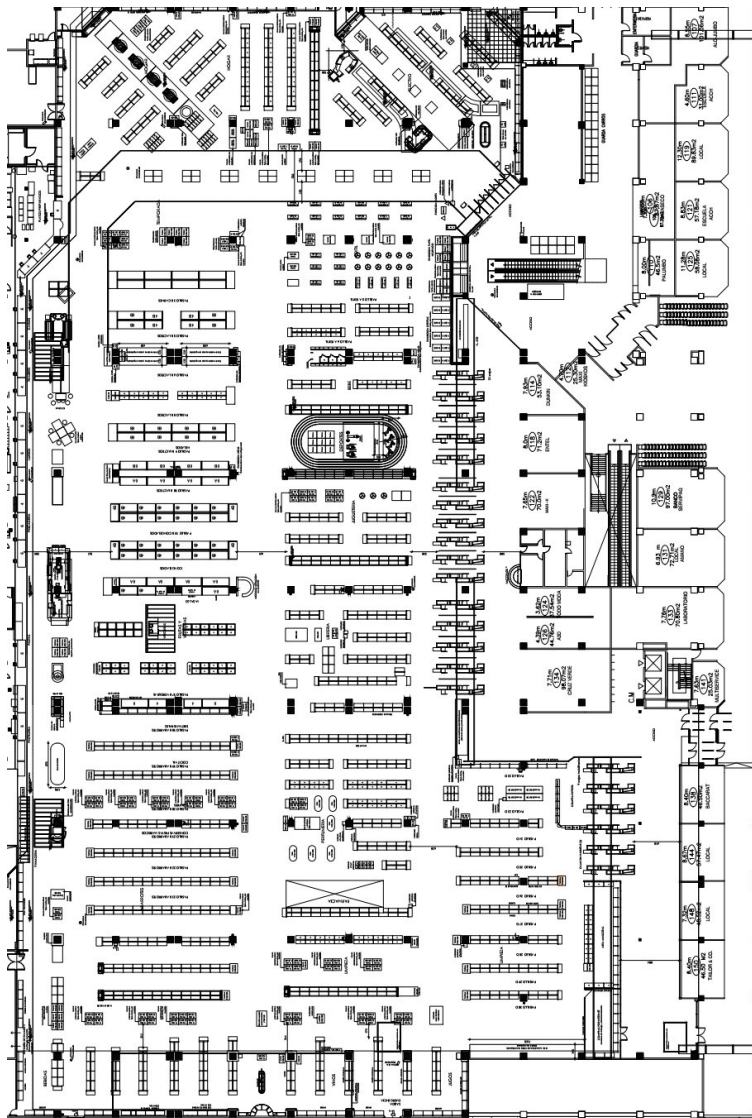
A modo de conclusión general se constata que los datos usados son potencialmente útiles para enriquecer los análisis que actualmente se hacen sobre el comportamiento de los clientes al interior de la sala de venta, permitiendo entender como el ambiente en que se desarrolla el proceso de compra influye en las decisiones del consumidor. Con esto es posible generar nuevos indicadores de conversión y de comportamiento al interior de la sala de ventas, como es la velocidad de desplazamiento.

Es así como dentro de los posibles trabajos futuros que se pueden generar a partir de este surgen dos temáticas que están ligadas entre si. Por un lado, la mejora de las trayectorias capturadas por los sensores y por el otro el establecimiento de relaciones entre trayectorias y canastas compradas por los clientes.

En la medida que se obtengan trayectorias más regulares es posible mejorar los análisis hechos anteriormente, ya que los caminos presentaran menos saltos y será posible saber la posición del cliente en más partes de su visita a la tienda. Una vez que se cuente con trayectorias más regulares una línea de trabajo interesante sería poder predecir canastas a partir de las trayectorias. Esto debido a que si se establece una relación entre trayectorias y canastas es posible explorar el desarrollo de promociones personalizadas, mejorando así la experiencia de compra ofrecida por el supermercado a los clientes.

Apéndice A

Mapa de la sala de ventas



Apéndice B

Productos por zonas definidas 1

Zona	Productos
1	Lácteos
2	Congelados
3	Fruta y Verdura (fyv)
4	Light: Galletas integrales y productos dietéticos.
5	Condimentos y encurtidos.
6	Dulces: Galletas, chocolates, etc.
7	Conservas: Conversas y leche condensada.
8	Cereales: Cereales, leche en polvo y comida bebe.
9	Azucar, café, infusiones y mermeladas.
10	Abarrotes: Harina, legumbres, postres para preparar, puré y sopas.
11	Tallarines: Tallarines, queso rallado y salsa de tomate.
12	Aliños: Aceite, arroz, sal y vinagre.
13	Snacks

Apéndice C

Productos por zonas definidas 2

Zona	Productos
14	Líquidos
	Limpieza hogar:
15	Aromatizadores, detergentes, insecticidas, lavalozas, limpiadores líquidos, cloro, confort, toallas de papel, servilletas, escobas, esponjas, fosforos, limpia pisos, mopas y plumeros.
16	Farmacia
17	Librería:
	Carpetas, cuadernos, artículos oficina, papelería, lápices, tempera, libros y mochilas.
	Juguetes y cumpleaños (jyc):
18	Juegos de mesa, muñecas, juguetes bebe, autos, juguetes de acción, cumpleaños y disfraces.
19	Deportes y Camping
20	Vestuario
21	Mascotas
22	Ferretería
23	Automoviles
24	Electro
25	Panadería

Bibliografía

- [1] *Informacion de tendencias del retail en Chile*. Departamento de estudios de la Camara de comercio de Santiago, 2016.
- [2] J. Josko Brakus. Brand experience: What is it? how is it measured? does it affect loyalty? *Journal of Marketing*, 73(3):52–68, 2009.
- [3] Vanja Demetrio. *Levantamiento de patrones de comportamiento de precios en una cadena de supermercado*. Universidad de Chile, 2012.
- [4] Sam K Hui. Testing behavioral hypotheses using an integrated model of grocery store shopping path and purchase behaviour. *Journal of Consumer Research*, 36(1):478–493, 2009.
- [5] Jacob Suher Sam K Hui J. Jeffrey Inman, Yanliu Huang. The effect of in-store travel distance on unplanned spending: Applications to mobile promotion strategies. *Jorunar of Marketing Research*, 77(2):1–16, 2013.
- [6] Oleg Urminsky Kivetz, Ran and Yuhuang Zheng. The goal-gradient hypothesis resurrec-
ted: Purchase acceleration, illusionary goal progress, and customer retention. *Journal of Marketing Research*, 43(1):39–58, 2006.
- [7] D.K.C. Oulton D.P Newman A.J., Yu. New insights into retail space and format planning from customer tracking data. *Journal of Retailing and Consumer Services*, 9(1):252–258, 2002.
- [8] Sam K Hui Peter S Fader, Eric T Bradlow. Path data in marketing: An integrative framework and prospectus for model building. *Marketing Science*, 28(1):320–335, 2009.
- [9] Sam K Hui Peter S Fader, Eric T Bradlow. The traveling salesman goes shopping: The systematic deviations of grocery paths from tsp-optimality. *Marketing Science*, 28(3): 566–572, 2009.
- [10] A. Swami R. Agrawal, T. Imielinski. Mining association rules between sets of items in large databases. *ACM SIGMOD*, 73(3):207–216, 1993.
- [11] Khan Uzma Ravi Dhar. Licensing effect in consumer choice. *Marketing Research*, 43(1): 259–266, 2006.

- [12] Burke Raymond. The third wave of marketing intelligence. *Retailing in the 21st Century Current and Future Trends*, 1(1):113–125, 2005.
- [13] Y. Zheng X. Xie W. Wang Y. Lou, C. Zhang and Y. Huang. Map-matching for low-sampling-rate gps trajectories. *ACM SIGSPATIAL GIS*, 1(1):352–361, 2009.