



**TESIS DE GRADO
EN INGENIERIA INDUSTRIAL**

**REDES NEURONALES APLICADAS A
LOS NEGOCIOS: ANALISIS DE UN CASO**

Autor: Manuel Sebastián Gómez de Lima

**Directores de Tesis:
M. Ing. Paola Britos
M. Ing. Enrique Fernández
Dr. Ramón García Martínez**

2006

Resumen Ejecutivo

Actualmente las empresas generan una gran cantidad de información que, por desconocimiento del valor de la misma o de las técnicas para aprovecharla, no utilizan a pleno potencial. Sin embargo, dentro de esta información subyacen patrones y tendencias que pueden indicar las necesidades implícitas de los consumidores y representar oportunidades de mejora del negocio.

El propósito de este trabajo es desarrollar un modelo que permita el reconocimiento de patrones de negocio. Para resolver este problema de una manera simple y eficiente se utilizarán redes neuronales artificiales, que mediante el aprendizaje supervisado simplifican y aceleran el proceso de procesamiento de la información, con el propósito de encontrar patrones implícitos dentro de la masa de la información.

El objetivo planteado es el de aplicar las redes neuronales artificiales a un problema real de negocio, con el fin de encontrar la solución al principal problema de los negocios, es decir, la satisfacción de las necesidades del cliente. Específicamente, se trabajará sobre una base de datos de ventas de una confitería, analizando los patrones y comportamientos de ventas con el fin de optimizar la oferta en función de los clientes.

Utilizando herramientas de redes neuronales artificiales se estudiarán los patrones de consumo de los clientes de la confitería (estudio de clasificación) y la composición de dicha clientela (estudio de segmentación o “clustering”). En base a los resultados de la misma, se analizarán posibles mejoras a la oferta del local. Como conclusión, se intentará demostrar la viabilidad de la aplicación de las redes neuronales al mundo de los negocios.

Palabras Clave: redes neuronales artificiales, aplicación a negocios, patrones de ventas, clasificación, segmentación, clustering.

Executive Brief

Nowadays, companies generate a great mass of information which is not fully taken advantage of, either by ignorance of its real value or the lack of techniques to process it. However, therein lie patterns and trends that may reveal the implicit needs of customers and represent business improvement opportunities.

The purpose of this paper is to develop a pattern recognition model for businesses. To solve this problem in a simple and efficient manner, artificial neural networks will be used with the intention of finding implicit patterns within the mass of information. Using supervised learning, artificial neural networks simplify and speed up information processing.

The objective is the application of artificial neural networks to a real business problem, with the aim of finding solutions to the main problem of the business world, that is, the satisfaction of customer needs. More specifically, a data base containing sales information of a coffee store will be used, analyzing patterns and sales trends in order to optimize supply based on demand (clients' needs).

Using artificial neural networks tools, the consumption patterns of the coffee store clients (classification study) and the composition of the group of clients (segmentation or clustering study) will be studied. Based on the results of these studies, improvements to the product offer of the coffee store will be analyzed. Finally, as a conclusion, this paper will try to demonstrate the viability of the application of artificial neural networks to the business world.

Keywords: artificial neural networks, business application, sales patterns, classification, segmentation, clustering.

Agradecimientos

En primer lugar, a mis padres, por hacer todo lo posible, y más, por darme la mejor educación, y a Julia, por aguantarme, alentarme y apoyarme siempre. Segundo, a los socios y empleados del Coffee Store (local Paseo del Ángel) por facilitar los datos y por estar siempre dispuestos a ayudar. Y finalmente, al Dr. Ramón García Martínez, M. Ing. Paola Britos e Ing. Enrique Fernández por ayudarme a definir el tema del trabajo y por sus valiosos comentarios y correcciones.

INDICE

1. INTRODUCCION.....	1
1.1. Breve descripción del problema	1
1.2. Motivación y objetivos	1
1.3. Pasos a seguir.....	2
1.4. Criterios de éxito	3
2. ESTADO DE LA CUESTION.....	5
2.1. Redes Neuronales Artificiales	5
2.2. Aplicaciones de las redes neuronales	6
2.3. Breve historia de las redes neuronales artificiales	7
2.4. La inspiración biológica	8
2.5. El modelo artificial básico	9
2.6. Recursos	12
2.7. Publicaciones	13
3. DEFINICION DEL PROBLEMA.....	15
3.1. Descripción del problema.....	15
3.2. Caracterización del negocio	16
3.2.1. The Coffee Store.....	16
3.2.2. Barrio de Belgrano	17
3.2.3. Ubicación del local	20
4. SOLUCION PROPUESTA.....	23
4.1. Introducción.....	23
4.2. Herramientas a utilizar	23
4.2.1. NNClust.....	24
4.2.2. NNClass	25
4.2.3. Resumen.....	27
4.3. Base de datos	28
4.3.1. Depuración de los datos.....	28
4.4. Estudio de segmentación	30
4.4.1. Uso de la herramienta y armado de la base de datos	30
4.4.2. Aplicación de la herramienta	31
4.4.3. Análisis de los resultados.....	36
4.4.4. Conclusiones del uso de la herramienta	41
4.5. Estudio de Clasificación	42
4.5.1. Uso de la herramienta y preparación de la base de datos	42
4.5.2. Aplicación de la herramienta	43
4.5.3. Análisis del Código de Tramo	43
4.5.4. Análisis de la Franja Horaria	50
4.5.5. Conclusiones del uso de la herramienta	58
4.6. Propuestas de mejora generales.....	59
5. CONCLUSIONES.....	63
5.1. Futuras líneas de investigación	64
6. REFERENCIAS	65

7. ANEXOS.....	69
Anexo 1 – Datos del Coffee Store	69
Anexo 2 – Estadísticas Ciudad de Buenos Aires	71
Anexo 3 – Macros de Visual Basic y fórmulas Excel.....	79
Anexo 4 – Gráficos adicionales del estudio de segmentación (herramienta NNClust).....	82
Anexo 5 – Gráficos adicionales del estudio de clasificación (herramienta NNClass).....	83



1. INTRODUCCION

1.1. Breve descripción del problema

El problema que se intenta solucionar en el presente trabajo es el del aprovechamiento eficiente de la información generada por las empresas. El creciente uso de la informática hace que las empresas almacenen grandes cantidades de información que en general no son aprovechadas, sea por desconocimiento del valor de la misma o de las técnicas para poder sacarle provecho.

Son pocas las empresas que analizan dicha información en búsqueda de patrones o conclusiones útiles para el negocio. En caso de hacerlo, en general se utilizan algoritmos lineales sin realimentación ni “aprendizaje”, los cuales insumen más tiempo y requieren mayor preparación previa y análisis posterior, arrojando resultados menos eficientes.

El propósito de este trabajo es desarrollar un modelo que permita el reconocimiento de patrones de negocio. Cabe destacar que al ser el problema relativamente nuevo, no existe un gran número de soluciones para el mismo. Se estudiará puntualmente el caso de una confitería, en la cual se analizarán los patrones de consumo de los clientes. De esta manera se intentará optimizar la oferta de forma tal de adaptarla a la demanda, lo cual constituye una de las claves en el mundo actual de los negocios.

Para resolver este problema de una manera simple y eficiente se utilizarán herramientas de redes neuronales, las que mediante el aprendizaje supervisado simplifican y aceleran el proceso de procesamiento de la información. Dentro de esa gran cantidad de información se buscarán patrones y tendencias que puedan revelar las necesidades implícitas de los consumidores y representar oportunidades de mejora.

1.2. Motivación y objetivos

Más allá de la solución del problema particular, es decir el análisis de los patrones de venta de una confitería, la verdadera motivación de este trabajo radica en demostrar la utilidad y la facilidad de la aplicación de las redes neuronales a los negocios. El objetivo a largo plazo es el de colaborar para constituir un nuevo paradigma en el análisis de datos en dicho contexto, tanto para grandes empresas como para PYMEs.



Si bien no es nueva la aplicación de las redes neuronales en el mundo, su uso en el ámbito empresarial se encuentra menos extendido. En general sus aplicaciones son más técnicas, desde usos militares hasta reconocimiento de voz y escritura, pasando por diagnósticos médicos y manejo de maquinaria (ver capítulo 2). Esto puede deberse a que los usuarios de las redes neuronales poseen en general una formación más científica o técnica, un tanto alejada de los negocios. El objetivo general es demostrar la utilidad de las redes neuronales aplicadas a los negocios, a través del enfoque de la ingeniería industrial que abarca tanto campos técnicos como empresariales.

Por otro lado, existe poco conocimiento entre el público de las redes neuronales. Fuera del ámbito técnico, la gente que conoce las redes neuronales tiene en general la creencia que las mismas son difíciles de usar, que requieren conocimientos avanzados o que son lentas e ineficientes. Probablemente esto se deba a los fracasos en dicho campo en el pasado, que postergaron su desarrollo. Actualmente, con el avance de la informática, el uso de las redes neuronales se encuentra al alcance de cualquier persona que posea una computadora de escritorio. Por lo tanto, un segundo objetivo de este trabajo es el de desmitificar los conceptos arraigados acerca de las redes neuronales, demostrando su velocidad y facilidad de uso.

En Argentina, en tanto, el desconocimiento de las redes neuronales por parte de la gente es aún mayor que en el mundo. Su conocimiento se limita al ámbito académico y la cantidad de publicaciones existentes al respecto son limitadas, generalmente enfocado a aplicaciones técnicas. El tercer objetivo de este trabajo es el de contribuir al conocimiento de las redes neuronales en el país, demostrando que su aplicación es posible en ámbitos empresariales.

Por último, este trabajo intentará demostrar que la aplicación de las redes neuronales es viable en PYMEs, microemprendimientos o en cualquier ámbito comercial o personal donde sea necesario el análisis de datos. Utilizando la base de datos de una confitería se demostrará que dicha tecnología no es exclusiva de grandes empresas.

1.3. Pasos a seguir

En primer lugar, será necesario una interiorización con el tema mediante la búsqueda de recursos y la lectura de bibliografía. Segundo, se buscarán herramientas de redes neuronales artificiales aptas para el uso requerido. Luego se aprenderá el manejo de las herramientas, y en función de las mismas se adecuará y depurará la base de datos existente. A continuación se aplicarán las herramientas a los datos y se analizarán los resultados obtenidos. Por último, se



extraerán conclusiones, evaluando el nivel de éxito del trabajo, enumerando las limitaciones y proponiendo futuras líneas de investigación.

1.4. Criterios de éxito

El éxito del presente trabajo de investigación dependerá en primer lugar del grado de facilidad con que se operen las herramientas y tiempo necesario para realizar el análisis. La obtención de conclusiones válidas y de fácil comprensión sin la necesidad de acudir a conocimientos avanzados de estadística representará un importante criterio de éxito.

Otra medida es el grado de cumplimiento de los objetivos previamente planteados: generar conciencia de la utilidad de las redes neuronales a los negocios y verificar su aplicación en distintos ámbitos empresariales, desmitificando las creencias arraigadas respecto a esta tecnología.





2. ESTADO DE LA CUESTION

2.1. Redes Neuronales Artificiales

Las redes neuronales artificiales ofrecen un paradigma atractivo para el diseño y el análisis de sistemas adaptativos inteligentes para un amplio rango de aplicaciones en inteligencia artificial [Fiszelew y García-Martínez, 2001].

Una Red Neuronal Artificial (RNA) es un paradigma de procesamiento de información que está inspirado en la forma en que los tejidos nerviosos biológicos, tales como el cerebro, procesan la información. El elemento clave de este paradigma es la estructura novedosa del sistema de procesamiento de la información. Se compone de un gran número de elementos de procesamiento altamente interconectados (neuronas) funcionando al unísono para resolver problemas específicos

Las RNA, al igual que las personas, aprenden mediante el ejemplo. Una RNA está configurada para una aplicación específica, tal como el reconocimiento de patrones y clasificación de datos, a través de un proceso de aprendizaje. El aprendizaje en sistemas biológicos involucra ajustes en las conexiones sinápticas que existen entre las neuronas. Esto es cierto también para las RNA [Stergiou y Sigamos, 1996].

Las RNA, con su habilidad para extraer significado de datos complicados o imprecisos, pueden ser usadas para extraer patrones y determinar tendencias que son muy complejas como para poder ser detectadas tanto por humanos como por otras técnicas informáticas. Puede considerarse a una red neuronal entrenada como un “experto” en la categoría de información a la que ha sido asignada para analizar. Este experto puede ser utilizado entonces para proveer proyecciones dadas nuevas situaciones de interés y contestar preguntas del tipo “qué pasaría si”.

Otras ventajas incluyen:

1. Aprendizaje adaptativo: La habilidad de aprender cómo hacer tareas basado en los datos dados para entrenamiento o como experiencia inicial.
2. Auto-Organización: Una RNA puede crear su propia organización o representación de la información que recibe durante el tiempo de aprendizaje



3. Operación en Tiempo Real: Las computaciones de una RNA pueden ser realizadas en paralelo. Actualmente, se está diseñando y fabricando hardware especial para aprovechar esta capacidad.
4. Tolerancia a Fallas vía Codificación de la Información Redundante: La destrucción parcial de una red lleva a la correspondiente degradación de la performance. Sin embargo, algunas capacidades de la red pueden ser retenidas a pesar de daños importantes a la red. [Stergiou y Siganos]

El éxito de las redes neuronales puede ser atribuido a ciertos factores clave [Statsoft]:

- Potencia: Las redes neuronales son técnicas capaces de modelizar funciones extremadamente complejas. En particular, las redes neuronales son no lineales. Durante muchos años, la modelización lineal fue la técnica más comúnmente usada. Cuando la aproximación lineal no era válida, lo cual frecuentemente era el caso, el modelo tenía serios inconvenientes. Las redes neuronales también controlan el problema de dimensionalidad que frustra los intentos de modelizar funciones no lineales con gran número de variables.
- Facilidad de uso: Las redes neuronales aprenden con el ejemplo. Las mismas recogen información representativa y luego utilizan algoritmos de entrenamiento para “aprender” la estructura de los datos. Si bien el usuario necesita tener conocimiento heurístico de cómo seleccionar y preparar los datos, cómo seleccionar una red neuronal apropiada o de cómo interpretar los resultados, el nivel de conocimiento requerido para aplicar exitosamente las redes neuronales es mucho menor del necesario para aplicar otros métodos estadísticos no lineales tradicionales.

2.2. Aplicaciones de las redes neuronales

Las redes neuronales son aplicables a prácticamente cualquier situación en la cual existe una relación entre variables independientes y dependientes. Dado que el fuerte de las redes neuronales es el de identificar patrones o tendencias en los datos, su uso está extendido en casos de predicción o pronóstico. Proyecciones de ventas, control de procesos industriales, estudios de mercado, validación de datos, evaluación de riesgo y marketing son algunos de los ejemplos de aplicaciones.

Sin embargo, su uso va más allá de simples proyecciones de datos. Las redes neuronales han revolucionado el reconocimiento de patrones aplicado a diversas



áreas. Algunos ejemplos incluyen: reconocimiento de voz, diagnósticos médicos, identificación de minas submarinas, análisis de texturas, reconocimiento tridimensional de objetos, manejo de maquinaria, reconocimiento de escritura a mano, asignación de crédito y reconocimiento facial [Stergiou y Siganos, 1996]. Desde usos militares en misilística hasta controles de calidad, desciframiento del código genético e incluso la prevención de ataques terroristas [Davis, 2005], el campo de acción es amplio.

2.3. Breve historia de las redes neuronales artificiales

Las redes neuronales nacieron gracias a la investigación en Inteligencia Artificial. Más específicamente, surgieron de los intentos de imitar la tolerancia a fallas y capacidad de aprendizaje de los sistemas neurales biológicos a través de modelar la estructura de bajo nivel del cerebro [Statsoft, 2003]. Si bien esta tecnología parece moderna, sus orígenes se remontan a la primera mitad del siglo XX. La siguiente tabla (tabla 2.3-1) [Bullinaria, 2004] muestra la evolución de las redes neuronales:

1943	McCulloch y Pitts proponen el modelo neuronal “McCulloch-Pitts”.
1949	Hebb publica su libro “La organización del comportamiento”, en el cual se propone la regla Hebbiana de aprendizaje.
1958	Roseblatt introduce las redes de una sola capa hoy llamadas “Perceptrones”.
1969	El libro “Perceptrones” de Minsky y Papera demuestra las limitaciones de los perceptrones de una sola capa, y prácticamente la disciplina entera queda postergada.
1982	Hopfield publica una serie de trabajos sobre las redes de Hopfield.
1982	Kohonen desarrolla los mapas auto-organizativos que hoy llevan su nombre.
1986	El algoritmo de aprendizaje para perceptrones de múltiples capas “Back Propagation” es redescubierto y la disciplina vuelve a surgir.
1990s	El subcampo de “Redes de funciones de base radial” es desarrollado
2000s	La potencia de los conjuntos de redes neuronales y los “Support Vector Machines” se vuelve notable.

Tabla 2.3-1. Hechos significativos de la historia de las redes neuronales



2.4. La inspiración biológica

Si bien la informática está aún lejos de lograr un sistema comparable al cerebro humano, contrastarlos sirve para mensurar las similitudes y diferencias entre ambos.

El cerebro humano posee unas diez mil millones de células nerviosas, o neuronas. En promedio, cada neurona se conecta a otras a través de unas diez mil conexiones, o “sinapsis”. La red neuronal del cerebro forma un sistema masivo de procesamiento paralelo de información, en contraste con las computadoras en las cuales un único procesador ejecuta linealmente una serie de instrucciones [Schraudolph y Cummins, 1999].

Por otra parte, la velocidad de procesamiento representa una diferencia importante. Mientras que el cerebro opera a unos 100 Hz (100 operaciones por segundo), una computadora convencional procesa varias centenas de millones de operaciones por segundo. A pesar del “hardware” lento con el que está construido, el cerebro posee una serie de capacidades notables:

- Su performance tiende a degradarse lentamente bajo daño parcial. En contraste, la mayoría de los programas y sistemas son frágiles: si se remueve alguna parte arbitrariamente, es muy probable que el sistema entero deje de funcionar.
- Puede aprender (reorganizarse) a partir de la experiencia.
- La recuperación de los daños es posible si unidades sanas logran reemplazar a las unidades dañadas, aprendiendo y asumiendo funciones de éstas.
- Ejecuta masivas computaciones paralelas de forma extremadamente eficiente. Por ejemplo, la percepción visual compleja, que equivale a 10 etapas de procesamiento, ocurre en menos de 100 milisegundos.
- Sustenta a la inteligencia y la conciencia. La ciencia todavía no ha logrado determinar cómo ocurre esto exactamente.

A modo de poder comparar mejor, la tabla siguiente [Schraudolph y Cummins, 1999] muestra un resumen de lo explicado previamente y agrega otros datos comparativos (figura 2.4-1).





		
Elementos de procesamiento	10^{14} sinapsis	10^8 transistores
Tamaño	10^{-6} m ³	10^{-6} m ³
Uso de energía	30 W	30 W (CPU)
Velocidad de procesamiento	100 Hz	10^9 Hz
Tipo de procesamiento	Paralelo, distribuido	En serie, centralizado
Tolerancia a fallas	Sí	No
Posib. de aprendizaje	Sí	Muy elemental
Inteligente / Conciencia	Sí	No

Tabla 2.4-1. Comparación del cerebro y los sistemas informáticos

2.5. El modelo artificial básico

De forma tal de capturar la esencia de los sistemas neuronales biológicos, una neurona artificial se puede describir de la siguiente manera [Statsoft, 2003]:

- Recibe una serie de “entradas” (sean datos originales o “salidas” de otras neuronas). Cada entrada llega a través de una conexión que tiene una cierta “fuerza”, o “peso”, que equivale a la eficiencia sináptica de una neurona biológica. Cada neurona tiene también un determinado valor de umbral. La suma ponderada de las entradas menos el valor de umbral componen la “activación” de la neurona (también conocida como Potencial Post-Sináptico –PPS- de la neurona).
- La señal de activación pasa a través de una función de activación, o función de transferencia, para producir la “salida” de la neurona.

Esto describe a una neurona individual. La pregunta siguiente es: ¿Cómo se deben conectar las neuronas para formar una red? Para que la red sea útil, debe haber como mínimo entradas (que toman los valores de las variables de interés externas) y salidas (que generan predicciones, o señales de control). Sin embargo, raramente las neuronas de entrada y salida están directamente conectadas. En el hombre, las neuronas que corresponden a los ojos (entrada) no están directamente conectadas a las manos (salida), por citar un ejemplo. Se requiere de un procesamiento interno para generar una salida. Por esto existen neuronas “ocultas”, que actúan internamente en la red [Statsoft, 2003]. Cabe aclarar que existen redes artificiales útiles de una sola capa, e incluso de un solo elemento, pero sus aplicaciones son más bien básicas [Anderson y McNeil, 1992].



La topología, o estructura, de la red es solo una de las maneras de clasificar a las redes neuronales artificiales. Existen otras clasificaciones importantes tales como la aplicación de las mismas, el tipo de conexión y el método de aprendizaje [Makhfi, 2006]:

- Topología
 - Capa simple (single layer)
 - Capas múltiples (multilayer)
 - Recurrente
 - Autoorganizada
 - etc.
- Aplicación
 - Clasificación
 - Clustering
 - Aproximación de funciones
 - Predicción
 - etc.
- Tipo de conexión
 - Estática (feedforward)
 - Dinámica (feedback)
- Método de aprendizaje
 - Supervisado
 - No supervisado

Las conexiones entre neuronas son una característica fundamental de las redes neuronales artificiales. En algunas redes, por ejemplo, es necesario que una neurona inhiba a otras neuronas de la misma capa. Esto se llama “inhibición lateral”, o competencia. En general, la diferencia principal reside en la existencia o no de retroalimentación. Básicamente, existen dos tipos de redes: las estáticas, o “Feedforward”, aquellas en que las conexiones son en un solo sentido desde la entrada hacia la salida, y las dinámicas, o “feedback”, en las cuales existe retroalimentación. Las redes “feedforward” son las que tienen hasta el momento mayores aplicaciones prácticas [Statsoft, 2003].

Luego de preparar los datos de entrada, se debe entrenar a la red. Existen dos enfoques para que la red “aprenda” – en forma supervisada o no supervisada. En pocas palabras, el aprendizaje supervisado significa proveer a la red de la salida deseada, ya sea “calificando” manualmente a la performance de la red o proveyendo a la misma con los resultados esperados. En el aprendizaje no



supervisado, la red es provista de los datos de entrada pero no los de salida, y el sistema debe “decidir” qué características usar para agrupar los datos.

La mayoría de las redes utilizan el aprendizaje supervisado. El aprendizaje no supervisado se utiliza generalmente para realizar algunas caracterizaciones iniciales a los datos de entrada. Un modelo básico de esto utiliza los mapas autoorganizativos de Kohonen. En cuanto a la posibilidad de verdadero aprendizaje por parte de estos sistemas, aún es una promesa lejos de lograrse, e incluso de entenderse. Esta posibilidad todavía está relegada al laboratorio [Anderson y McNeil, 1992].

Por último, las redes neuronales tienen un amplio rango de aplicaciones. Sin embargo, los problemas de clasificación y los de predicción (o regresión) son los más comunes. Otras aplicaciones son la asociación de datos, la conceptualización de datos, el “clustering”, y el filtrado de datos. Existe una importante cantidad de tipos de redes asociadas a estas aplicaciones. La tabla siguiente (tabla 2.5-1) [Anderson y McNeil, 1992] muestra un resumen de las mismas:

Aplicación	Tipo de red (en inglés)	Usos de la red (ejemplos)
Predicción	<ul style="list-style-type: none"><input type="checkbox"/> Back-propagation<input type="checkbox"/> Delta Bar Delta<input type="checkbox"/> Extended Delta Bar Delta<input type="checkbox"/> Directed Random Search<input type="checkbox"/> Higher Order Neural Networks<input type="checkbox"/> Self-organizing map into Back-propagation	Predecir variables de salida en función de valores de entrada de variables relacionadas (ejemplo: selección de acciones en la bolsa, identificación de pacientes con enfermedades, predicción del tiempo).
Clasificación	<ul style="list-style-type: none"><input type="checkbox"/> Learning Vector Quantization<input type="checkbox"/> Counter-propagation<input type="checkbox"/> Probabalistic Neural Networks	Clasificar a los datos de entrada (ejemplo: identificación de escritura, clasificación según características).
Asociación de datos	<ul style="list-style-type: none"><input type="checkbox"/> Hopfield<input type="checkbox"/> Boltzmann Machine<input type="checkbox"/> Hamming Network<input type="checkbox"/> Bidirectional associative Memory<input type="checkbox"/> Spation-temporal Pattern Recognition	Como la clasificación, pero con reconocimiento de errores (ejemplo: no solo reconocer escritura, sino hacerlo también cuando el escáner funciona mal).



Aplicación	Tipo de red (en inglés)	Usos de la red (ejemplos)
Conceptualización de datos	<input type="checkbox"/> Adaptive Resonance Network <input type="checkbox"/> Self Organizing Map	Analizar las entradas para agrupar relaciones que se pueden inferir (ejemplo: extraer de una base de datos los sujetos más proclives a comprar un producto particular).
Filtrado de datos	<input type="checkbox"/> Recirculation	Suavizar una señal de entrada (ejemplo: quitar el ruido a una señal telefónica).

Tabla 2.5-1. Aplicaciones de las redes neuronales

2.6. Recursos

Las aplicaciones de las redes neuronales y sistemas de reconocimiento de patrones tienen un gran potencial para ser aplicados al mundo de los negocios. Estos sistemas tienen especial utilidad en la proyección, clasificación, segmentación y el análisis de datos.

Sin embargo, cuando se realiza una búsqueda en Internet no se encuentra una gran cantidad de aplicaciones específicas de estos temas a los negocios. Primordialmente, se encuentran muchas páginas académicas, de universidades o centros de investigación, que se dedican al estudio de estos temas. Las aplicaciones son más bien teóricas o para usos tecnológicos e informáticos.

Por otro lado, se pueden encontrar programas (software) dedicados a realizar algunas de las tareas específicas antes mencionadas, tales como clasificar, predecir o pronosticar y analizar datos. Estos programas suelen ser específicos para una de las aplicaciones y generalmente no combinan estas actividades.

En cuanto al reconocimiento de patrones, existen una variedad de programas de computación dedicados a esta tarea, pero su número es escaso y pocos están aplicados al mundo de los negocios. La mayoría de los programas aplicados a los negocios son paquetes de software grandes y caros. En general, los programas de reconocimiento de patrones que se pueden encontrar en Internet están más relacionados con la lectura óptica de caracteres (OCR, del inglés "Optical Character Reading") y con el reconocimiento de patrones en imágenes.

La poca difusión del software de redes neuronales para aplicaciones comerciales es llamativa, especialmente considerando el estado actual de la tecnología informática. Hoy en día, una simple computadora de escritorio puede manejar sin



problemas todos los tipos de software relacionados con redes neuronales. Sin embargo, no se encuentra una gran cantidad de software de redes neuronales. Puede atribuirse este desconocimiento a que la tecnología de redes neuronales es relativamente nueva y sus aplicaciones a los negocios son, por el momento, escasas.

Otro problema es que en el pasado las redes neuronales tuvieron algunos inconvenientes limitantes tales como la necesaria preparación previa de los datos. Para que una red arroje resultados adecuados se requiere un buen pre-procesamiento de los datos, la selección correcta de la arquitectura y el adecuado entrenamiento de la red. En el pasado, la mayor parte de las fallas al usar redes neuronales provenían de las inadecuadas habilidades de los usuarios al preparar los datos y el diseño de una red neuronal. Hoy en día, los beneficios de las redes neuronales para el mundo de los negocios siguen siendo subestimados debido a la dificultad para pre-procesar adecuadamente los datos y construir correctamente la red [OzGrid, 2006].

2.7. Publicaciones

En el ámbito académico, por otra parte, la cantidad de recursos disponibles sobre redes neuronales es amplio, aunque las aplicaciones directas a negocios son menores.

Internet es una fuente importante de recursos sobre el tema de redes neuronales. Existe un gran número de páginas que contienen publicaciones sobre redes neuronales, pero éstas hacen foco principalmente en cuestiones técnicas de las mismas. A modo de ejemplo, los enlaces siguientes poseen colecciones de publicaciones referidas a temas generales sobre las redes neuronales:

- <http://www.cs.queensu.ca/home/browse/Publications/NNP1.htm> [Browse, 2006]
- <http://www.cs.man.ac.uk/~jls/Publications/node4.html> [Shapiro, 2004]
- <http://www.calsci.com/Applications.html#Pattern%20Recog%20Applications> [BrainMaker, 2006]
- <http://www.neurocolt.com/abs/2002/2002.html> [Neurocolt, 2002]
- <http://www.cs.york.ac.uk/arch/NeuralNetworks/references.htm> [Austin, 2006]

Referidos específicamente a los negocios, el número de recursos disponibles es considerablemente menor. Una gran cantidad de los trabajos encontrados, tales como [Shachmurove, 2002], [Vellido *et al*, 1999], [Zhang, *et al*, 1998], [Raghupathi y Shekhar, 1996] y [García Vega, 2002], consisten en repasos históricos de publicaciones previas sobre redes neuronales aplicadas a los negocios. Estas



publicaciones constituyen una buena fuente de trabajos relacionados. [Li, 1994], por su parte, escribe sobre la factibilidad de aplicar redes neuronales a los negocios y propone posibles aplicaciones. [Britos, 2005] relaciona la aplicación de sistemas inteligentes, entre los cuales se encuentran las redes neuronales, con objetivos de negocio.

En el caso de las publicaciones más aplicadas, se encuentran [Shachmurove y Witkowska, 2000] que utilizan redes neuronales para predecir la evolución de la bolsa de valores. Probablemente esta sea la aplicación más extendida en el campo de los negocios. [Puppio, 2002], por otro lado, propone la aplicación de las redes neuronales a la tasación inmobiliaria.

Se aprecia que la mayoría de los trabajos encontrados son en inglés. Los que están en español provienen de varios países de Latinoamérica, tales como Colombia, Venezuela, México, Chile y Argentina, y de España. No se aprecia que haya un país que tenga preponderancia en cuanto a la cantidad de trabajos publicados. Esto demuestra que la aplicación de redes neuronales es una disciplina incipiente en nuestro país.



3. DEFINICION DEL PROBLEMA

3.1. Descripción del problema

El problema que se intentará solucionar es el del aprovechamiento eficiente de la información generada por las empresas. Como objetivo principal, se buscará estudiar la demanda de un negocio de forma tal de optimizar la demanda en función de las necesidades del cliente, cuestión principal en el mundo de los negocios.

El auge de la informática de las últimas décadas ha introducido a las computadoras en prácticamente todos los ámbitos del mundo de los negocios. Esto ha simplificado y acelerado muchos procesos internos, generando por otro lado una gran cantidad de información. Sea por desconocimiento o inacción, muchas veces esta masa de información es archivada o eliminada sin ningún tipo de procesamiento. Sin embargo, estos datos contienen valiosa información sobre el negocio, que podría ser utilizada para generar mayor rentabilidad y ganancias, lo cual representa la meta de todo empresario.

En un mundo cada vez más competitivo, los empresarios no pueden descartar esta fuente gratuita de información disponible. El capital intelectual se está convirtiendo en un activo cada vez más importante en el mundo. Las grandes empresas del mundo lo saben, y están comenzando a estudiar, analizar y procesar la gran cantidad de información que generan.

El objetivo final de este trabajo es demostrar que no es necesario poseer equipos sofisticados, contratar expertos en estadística ni ser una gran multinacional para poder aprovechar la información, sino que la tecnología está disponible para todo aquel que posea una computadora de escritorio y mínimos conocimientos de los datos a analizar.

Específicamente, se estudiará una base de datos de consumo obtenida de una confitería, "Coffee Store", la cual fue facilitada por los socios del local. Se aplicarán redes neuronales artificiales, combinando distintas herramientas con el fin de detectar tendencias de consumo y segmentar a la clientela. A través de este análisis se propondrán modificaciones y mejoras a la oferta, para así adecuarla a la demanda y las necesidades implícitas de los clientes.



3.2. Caracterización del negocio

3.2.1. The Coffee Store

The Coffee Store es una cadena de confiterías creada en Argentina en noviembre de 1998. Nace para satisfacer un nicho del mercado que valora al café no solo como una bebida sino como un estilo de vida, lo cual se ve reflejado en su misión: “The Coffee Store nace para rendir homenaje al café en todas sus formas. Reflejando su espíritu con intensidad y pasión. Ofreciendo una amplia variedad de excelentes cafés del mundo tratados con nuestro exclusivo proceso de tostado”. Este concepto de confiterías no era nuevo en el mundo pero sí en el mercado argentino.

En sus locales se venden 16 variedades de café de diversos orígenes, todas de grano arábigo, incluyendo cinco variedades saborizadas, dos “blends” y una descafeinada. Si bien su origen es argentino, The Coffee Store se ha expandido también en Uruguay y en España. En Argentina existen actualmente 25 locales, de los cuales sólo 2 se encuentran fuera de Capital y Gran Buenos Aires (La Plata y Rosario), mientras que hay uno en Uruguay (Punta del Este) y 4 en Madrid, España. Por la ubicación de los locales (ver anexo 1.1), se puede apreciar que las confiterías apuntan a un segmento con un nivel socioeconómico alto. Todos los locales mantienen un ambiente y arquitectura similar.

La empresa propietaria, Kaldi S.A., opera mediante el sistema de franquicias. El anexo 1.2 muestra los valores de las franquicias y los principales indicadores del negocio. La compra de una franquicia incluye los siguientes servicios:

- Derecho al Uso de la Marca
- Apoyo con Selección de la Ubicación del Punto de Venta
- Desarrollo del Proyecto Arquitectónico, Imagen y Diseño
- Manuales Operativos, Técnicos, del Empleado y de Identidad Corporativa.
- Provisión de Productos (sin montos mínimos de compra)
- Preferencia Zonal
- Años de licencia en función del contrato de locación o concesión
- Apoyo Continuo
- Management Integral (optativo)

El sistema de franquicias permitió a la marca una rápida expansión, afianzando su imagen entre el público. Aunque el número de cadenas de confiterías similares se ha incrementado notablemente en los últimos años, The Coffee Store sigue siendo una de las cadenas líderes en el sector.



3.2.2. Barrio de Belgrano

“Belgrano es un barrio residencial y distinguido. Su nombre evoca la figura de Manuel Belgrano, uno de los más importantes protagonistas de la historia argentina. Sus calles son silenciosas y arboladas, conglomeran comercios, restaurantes, reliquias arquitectónicas y grandes espacios verdes. Las barrancas de Belgrano constituyen uno de los lugares mas destacados del barrio. Su diseño, al igual que otros espacios importantes de la ciudad, es obra del famoso paisajista Carlos Thays, creador del Jardín Botánico. Desde su formación, esta zona tuvo un uso residencial e institucional, y los habitantes pertenecían a un grupo social ligado a las clases dirigentes e intelectuales de Buenos Aires, por lo que entre las construcciones, se distinguen las importantes casonas, embajadas y petit-hotels” [Ripio, 2004].

El barrio de Belgrano tiene, según un estudio del año 2001, una población de 138.942 habitantes, siendo el séptimo barrio más habitado sobre los 47 de la ciudad. La tabla 3.2.2-1 muestra el detalle de los datos de población [GCBA, 2005].

Total	Sexo	
	Varones	Mujeres
138.942	62.354	76.588

Tabla 3.2.2-1. Población del barrio de Belgrano

Los anexos 2.1 y 2.2 contienen una serie de figuras y tablas que reproducen información acerca de las distintas comunas en las que se encuentra dividida la ciudad. Las mismas fueron obtenidas de la Encuesta Anual de hogares del año 2004 [Registro Civil, 2004]. Estas tablas contienen datos disgregados según los distintos centros de Gestión y Participación de la ciudad de Buenos Aires. Cabe remarcar que el barrio de Belgrano se encuentra casi en su totalidad dentro del CGP número 13, aunque una parte pequeña del barrio pertenece al CGP 14 Oeste. Por su parte, el local analizado se encuentra ubicado en el CGP 13, aunque a una cuadra del límite con el CGP 14 Oeste.

Se puede apreciar en las tablas del anexo 2.3 los distintos indicadores que posicionan a los CGP cercanos a Belgrano como los que tienen los mejores niveles de calidad de vida en la ciudad de Buenos Aires. A continuación, se muestra un resumen de alguna de estas tablas (tablas 3.2.2-2 a 3.2.2-8) para el CGP 13.



a) Educación

	Primario Incompleto	Primario Completo	Secundario Incompleto	Secundario Completo	Superior / Universitario Incompleto	Superior / Universitario Completo
CGP N° 13	2,4	8,5	8,5	19,3	22,4	38,9
Ciudad de Bs. As.	4,7	16,7	13,1	21,5	16,5	27,6

Tabla 3.2.2-2. Máximo nivel de instrucción alcanzado

		Estatal / Público	Privado	NS / NC
Asiste actualmente	CGP N° 13	47,7	52,3	0
	Ciudad de Bs. As.	59,9	39,4	0,7
No asiste pero asistió	CGP N° 13	72	27,8	0,2
	Ciudad de Bs. As.	80,1	19,5	0,5
Total	CGP N° 13	64,7	35,2	0,1
	Ciudad de Bs. As.	73,9	25,6	0,5

Tabla 3.2.2-3. Sector al que pertenece o perteneció el establecimiento educativo

Se puede observar que el CGP 13 tiene indicadores de educación muy superiores a la media de la ciudad. Más del 80% de la población de dicha zona terminó el secundario y más del 61% asistió a la universidad. Los valores de la ciudad son del 65,6% y 44,1% respectivamente.

En cuanto al tipo de establecimiento educativo al que se concurre también se encuentran diferencias sustanciales. De la población que actualmente asiste, el 52% de la población de Belgrano concurre a establecimientos privados, contra 39,4% en el resto de la población de la ciudad.

**b) Salud y vivienda**

	Sistema Público y Plan Médicos de Cabecera	Obra Social	Medicina Prepaga	Otros
13	9,2	47,6	28,5	14,7
Ciudad de Bs. As.	22	54	14,4	9,5

Tabla 3.2.2-4. Tipo de cobertura médica

	Sin hacinamiento	Con hacinamiento	Con hacinamiento crítico
13	99	0,6	0,4
Ciudad de Bs. As.	94,8	3,4	1,8

Tabla 3.2.2-5. Situación de hacinamiento

Como se aprecia en el cuadro 3.2.2-4, los habitantes de Belgrano poseen algún sistema de medicina prepaga en una porcentaje que prácticamente duplica al promedio de la ciudad. Por otra parte, sólo el 9,2% utiliza el sistema público de salud, en comparación con el 22% del resto de la ciudad.

En cuanto a la situación de hacinamiento en las viviendas, el cuadro muestra que el CGP 13 tiene valores menores que el promedio. Sin embargo, el verdadero valor de esta estadística se aprecia al analizar el cuadro completo (ver anexo 2.3). Allí se puede ver que el CGP 13 tiene los menores valores entre todos los CGP.

c) Empleo y salario

	Ocupados	Desocupados	Inactivos
CGP N° 13	72,7	3,2	24,1
Ciudad de Bs. As.	69,1	3,9	27

Tabla 3.2.2-6. Condición de actividad

	Patrón o Empleador	Trabajador por cuenta propia	Asalariado	Trabajador Familiar
CGP N° 13	8,9	20,4	70,2	0,6
Ciudad de Bs. As.	5,6	19,5	73,7	1,1

Tabla 3.2.2-7. Tipo de actividad



Se puede observar en las tablas de empleo previas que tanto la tasa de desocupación como la población inactiva son menores que el promedio de la ciudad. Por otra parte, el porcentaje de patrones y empleadores del CGF 13 es 58% mayor que el promedio de la ciudad. Este dato es más significativo si se tiene en cuenta que ese porcentaje es el mayor de todos los CGP de la ciudad (ver anexo 2.3).

	Patrón o Empleador	Trabajador por cuenta propia	Asalariado	Total
CGP N° 13	4462	1313,5	1163,6	1501,2
Ciudad de Bs. As.	2401,1	940,8	933,1	1011,3

Tabla 3.2.2-8. Nivel de salarios por tipo de actividad

Sin duda que el cuadro de salarios marca la diferencia más importante con respecto al resto de la ciudad. En promedio, el CGP 13 tiene salarios que son 48% mayores a los del resto de la ciudad. En el caso de los patrones y empleadores la diferencia llega al 85,8%. Tanto esta categoría como el total general tienen los salarios más altos de toda la ciudad. Los otros dos grupos también son ampliamente superiores que el promedio, superando el 24% de diferencia en el peor caso, aunque ambos se colocan en el segundo lugar del ranking de salarios.

Todos estos cuadros, sumados a aquellos del anexo 2.3, muestran que Belgrano es uno de los barrios con mejor calidad de vida, sino acaso el mejor. Sus habitantes pertenecen a una minoría que se ubica, considerando incluso a todo el país, en el tope de la escala social y económica. Cualquier análisis e intento de mejora debe tener en cuenta este hecho.

3.2.3. Ubicación del local

El local analizado se encuentra en el barrio de Belgrano, más específicamente en la calle Mariscal Antonio José de Sucre al 2145. Fue inaugurado en marzo de 2004, en el complejo “El Paseo del Ángel”. Este paseo de compras al aire libre fue inaugurado a fines de los años 90 en el sitio donde se ubicaba la “Casa del Ángel”, una construcción de estilo normando con un mirador en el que se encontraba la figura de un ángel, que fue la casa del residente francés Carlos Delcasse, renombrado esgrimista y promotor de este deporte en Argentina [Folklore, 2005]. Este paseo de compras contiene tiendas de marcas caras, las cuales apuntan al público de la zona. La figura 3.2.3-1 muestra un plano del Barrio de Belgrano y la ubicación del Coffee Store (en azul).



Figura 3.2.3-1. Mapa del barrio de Belgrano y ubicación del local del Coffee Store (en azul).

La ubicación del local tiene las siguientes características:

- Está situado en el Paseo del Ángel, exclusivo paseo de compras de la zona.
- Se encuentra enclavado en una zona residencial de alto valor, cuyos residentes pertenecen a un nivel socioeconómico alto.
- Está muy próximo a la avenida Cabildo. Esta avenida, especialmente entre José Hernández y Congreso tiene una alta concentración de galerías y locales a la calle, constituyendo una de las más importantes áreas comerciales de la ciudad.
- Se encuentra a dos cuadras del centro de Belgrano, la plaza General Manuel Belgrano, donde hay una iglesia, varios museos importantes y una gran concentración de comercios. La esquina de Cabildo y Juramento es una de las más transitadas de la zona. Se estima que los sábados entre las 10 y las 16 circulan por la zona unas 45.000 personas [Mibelgrano, 2006].



- A 3 cuadras se encuentran las Barrancas de Belgrano, un parque de 3,2 hectáreas. Junto con la plaza Manuel Belgrano, la iglesia y los museos conforman una zona turística de importancia.
- A menos de 3 cuadras se encuentra el Complejo Multiplex Belgrano, con 6 salas de cine. Además, en un radio de 10 cuadras hay una gran cantidad de teatros, auditorios y otros cines.
- Se encuentra a 5 cuadras de la estación Belgrano C, que tiene una gran afluencia de pasajeros y una importante zona comercial. A una distancia de 10 cuadras se encuentra la estación Belgrano R y el barrio homónimo, una zona residencial de alto nivel socioeconómico que además posee poca oferta gastronómica.
- Las estaciones de subterráneo José Hernández y Juramento de la línea D se hallan a una distancia de 5 y 4 cuadras respectivamente. Solamente por la estación Juramento circulan diariamente unas 60.000 personas [Clarín, 1999].

Como se puede ver, el local tiene una ubicación privilegiada. Existen en su entorno áreas muy pobladas y calles muy transitadas. El hecho de que hayan cines y teatros en las cercanías ayuda a incrementar el potencial de clientes. Por último, el eje comercial de la avenida Cabildo y las estaciones de tren y subte aseguran un flujo importante de personas por la zona.



4. SOLUCION PROPUESTA

4.1. Introducción

Para poder solucionar el problema de la aplicación de las redes neuronales a las bases de datos del Coffee Store, es necesario definir primero las herramientas a utilizar. En virtud de que el presente trabajo consiste en un desarrollo académico, y no posee un presupuesto para la compra de una herramienta de exploración de información comercial, se deberá realizar una tarea de búsqueda de herramientas gratuitas.

A continuación, se deberá adecuar la base de datos para posibilitar la correcta aplicación de las herramientas. Una vez definidas las herramientas y depurada la base de datos, se aplicarán las herramientas.

Por otra parte, se mantendrán reuniones con dueños y empleados del local para poder conocer mejor el funcionamiento del local. De esta forma se podrá evaluar el desempeño de las herramientas en función de los resultados, y se podrán estudiar propuestas de mejora para la oferta del local.

4.2. Herramientas a utilizar

Las herramientas que se utilizarán para el análisis de los datos serán los programas NNClass y NNClust. Estas dos aplicaciones son en realidad planillas de Excel que ejecutan un conjunto de macros. El hecho de que sean planillas de Excel facilita la compatibilidad con otros programas, lo cual simplifica la implementación y el análisis de la información. La elección de estas dos herramientas reside, en primer lugar, a la experiencia previa con las mismas y, segundo, a que son software de distribución gratuita por internet. Existen otros programas de minería de datos con redes neuronales, pero la gran mayoría son pagos o son muy específicos.

Las dos herramientas mencionadas se pueden descargar de internet a través de la página web de su creador, Angshuman Saha, 2006: <http://www.geocities.com/adotsaha/index.html>. Allí también se encuentran descripciones de las herramientas, así como un listado de propiedades y limitaciones que se transcriben más adelante.



4.2.1. NNClust

Esta herramienta permite el análisis automático de datos mediante el uso de mapas autoorganizados de Kohonen. Los mapas autoorganizados de Kohonen, SOM (Self Organizing Maps) o SOFM (Self Organizing Feature Maps), fueron introducidas por Teuvo Kohonen y son un tipo especial de redes neuronales artificiales de aprendizaje no supervisado que ha sido exitosamente aplicado como una herramienta de Data Mining. Las ventajas de los mapas autoorganizativos radica en que son capaces de preservar la topología del espacio de los datos, proyectan datos altamente dimensionales a un esquema de representación de baja dimensión y tienen la habilidad de encontrar similitudes en los datos [Salas].

Las SOM pueden ser descritas formalmente como un mapeamiento no-lineal, ordenado, suave de los datos de entrada altamente dimensionales hacia los elementos de un arreglo regular de baja dimensión. El algoritmo de las SOM consiste en un procedimiento iterativo capaz de representar la estructura topológica del espacio de entrada (discreto o continuo) por medio de un conjunto discreto de prototipos de vectores de peso las que son asociadas a neuronas de la red. Las SOM mapean los patrones de entradas vecinos a neuronas vecinas [Salas].

En particular, la herramienta NNClust utiliza los mapas de Kohonen para el “Clustering”, o segmentación de los datos. La misma se utilizará para dividir a la clientela en distintos segmentos de acuerdo a su consumo, de forma tal de poder extraer conclusiones y proponer mejoras válidas y efectivas.

Propiedades de la herramienta:

- Simplicidad de uso, compatibilidad con Excel y Windows, fácil instalación y tamaño reducido.
- Se pueden modificar los parámetros de entrenamiento y, durante el mismo, presentar los datos al mapa en forma aleatoria.
- Los pesos de las corridas previas son almacenados y se pueden utilizar para nuevas estudios.
- El programa genera una serie de informes con los parámetros estadísticos de los clusters y dibuja un mapa de los mismos.
- Se pueden guardar los resultados de las corridas en planillas separadas.



Limitaciones de la herramienta:

- Restricción de tamaño: El programa no puede manejar más de 50 variables, 100 neuronas y 5000 datos. Sin embargo, para el experimento a realizar sólo la restricción de datos representa una limitación.
- Restricción de ciclos: El programa no puede superar los 1000 ciclos de entrenamiento. De todas formas, se realizaron corridas sucesivas con 100 ciclos para que el programa “aprenda”.
- Velocidad de procesamiento: Esta herramienta no es particularmente rápida para el análisis con muchos datos. Tiene principalmente dos cuellos de botella:
 - Preprocesamiento de los datos antes del ciclo de entrenamiento.
 - Dibujo del mapa luego de la finalización del ciclo de entrenamiento.
- Gráficos: Esta herramienta no tiene la posibilidad de generar una gran variedad de gráficos que habitualmente se relacionan con el SOM.
- Otras limitaciones técnicas
 - La aplicación no puede manejar grillas de neuronas que no sean cuadradas.
 - El programa sólo puede usar una función de vecindad Gaussiana.
 - Los vectores de peso son iniciados al azar con valores de coordenadas entre -1 y 1.
 - Las variables de entrada están limitadas a valores entre -1 y 1.
 - Las distancias de plano utilizadas para computar similitudes son Euclidianas. La aplicación no puede utilizar otro tipo de distancias.

4.2.2. NNClass

Esta herramienta permite el análisis automático de datos mediante el uso de redes neuronales de retropropagación. La red de retropropagación (Backpropagation), se debe a Rumelhart, Hinton y Williams (1986) y consiste en una red multicapa en la que luego de obtenida una salida en la propagación hacia adelante se comprueba el error cometido, es decir, la diferencia entre la salida real y la deseada. Este error se propaga hacia atrás por toda la red distribuyéndolo entre todos los nodos de cada capa. La distribución consiste en variar los pesos de cada conexión sináptica hasta conseguir después de muchas iteraciones, una cota de error inferior a la estipulada previamente [Nueva Esparta, 2006].



Uno de los grandes avances logrados con la Backpropagation es que esta red aprovecha la naturaleza paralela de las redes neuronales para reducir el tiempo requerido por un procesador secuencial para determinar la correspondencia entre unos patrones dados. Además el tiempo de desarrollo de cualquier sistema que se este tratando de analizar se puede reducir como consecuencia de que la red puede aprender el algoritmo correcto sin que alguien tenga que deducir por anticipado el algoritmo en cuestión [Acosta et al, 2000].

El NNClass permite clasificar los datos en función de una determinada variable. En particular, se estudiarán los consumos de distintos productos en función del día u hora de la semana en que se produjo la compra. Se buscarán patrones que permitan obtener conclusiones que a su vez permitan mejorar la oferta del negocio.

Propiedades de la herramienta:

- Simplicidad de uso y compatibilidad con Excel y Windows, fácil instalación y tamaño reducido.
- Es posible separar entre datos de entrenamiento y datos de validación.
- Se puede entrenar a la red tanto secuencialmente como en lote.
- Los pesos iniciales pueden ser establecidos al azar o especificados por el usuario. Estos pesos son guardados y se pueden utilizar en corridas sucesivas. Además, los mismos se pueden presentar en orden aleatorio.
- La aplicación puede manejar datos categóricos automáticamente, lo cual es necesario para el presente estudio.
- El programa rastrea los errores de clasificación e informa de los resultados.
- Se generan gráficos útiles para el análisis posterior.
- Los resultados pueden ser guardados en planillas de Excel separadas.



Limitaciones de la herramienta:

- Restricción de tamaño: El programa no puede manejar más de 50 variables de entrada, 2 capas ocultas con menos de 20 neuronas y 2000 datos. Sin embargo, la cantidad de datos fue aumentada mediante el editor de macros de Visual Basic.
- Restricción de ciclos: El programa no puede superar los 500 ciclos de entrenamiento. De todas formas, se realizaron corridas sucesivas con 100 ciclos para que el programa “aprenda”.
- Velocidad de procesamiento: Esta herramienta no es particularmente rápida para el análisis con muchos datos.
- Gráficos: Esta herramienta no tiene la posibilidad de generar una gran variedad de gráficos.
- La aplicación no realiza ningún tipo de preprocesamiento. El usuario debe verificar previamente que:
 - Los distintos campos sean linealmente independiente (es decir, que no sean una combinación lineal de otros campos).
 - La presencia de datos “outliers” (muy alejados de la media) pueden afectar los resultados.
 - Una gran cantidad de categorías puede derivar en resultados menos robustos.

4.2.3. Resumen

Estas herramientas tienen una serie de propiedades que justifican su elección como programas de análisis para el presente estudio. La simplicidad de uso de las mismas y su compatibilidad con Excel son esenciales para cumplir con los objetivos propuestos en el trabajo.

Por otra parte, poseen una serie de limitaciones propias de los programas de circulación libre. Sin embargo, ambas herramientas son lo suficientemente potentes para el estudio que se pretende realizar. Parte del atractivo del mismo es demostrar que el análisis mediante redes neuronales no requiere de software avanzados, grandes inversiones o amplios conocimientos, y que las herramientas para realizarlo están al alcance de cualquier persona.



4.3. Base de datos

La base de datos corresponde a las ventas de un local de “The Coffee Store”. Luego de una depuración de los datos (que se comenta en el ítem “Depuración de los datos”), los campos definitivos que se utilizan se encuentran definidos en la tabla 4.3-1. Cada fila en la base de datos corresponde a un ítem dentro de una factura. Por lo tanto, cada factura – es decir, cada consumo de cada cliente – puede comprender varias filas dentro de la base de datos.

Campo	Descripción
Artículo	Contiene un código para cada tipo de artículo vendido en cada ticket
Cantidad	La cantidad comprada del artículo en cuestión en ese ticket
Fecha	Fecha de la venta
Hora	Hora de la venta
Nº de compra	Es el número que identifica a cada ticket
Memo	Es un campo en el que se ingresó información adicional. Normalmente comentarios sobre pedidos no estándar
Nº de renglón	Es un contador de líneas dentro de un ticket. En el caso que se hallan comprado dos artículos diferentes en un solo ticket el primero de ellos tendrá en este campo el valor 1 mientras que el segundo tendrá un 2,
Precio	Precio del artículo
Total de la factura	Contabiliza el total a abonar en cada ticket
Código de rubro	Los artículos se agrupan en rubros diferentes
Descripción del rubro	Cada código de rubro tiene una descripción en palabras. Por ejemplo "Pocillo"
Código de mozo	Se refiere al mozo que atendió la mesa en cuestión
Código de mesa	Es el número de la mesa
Código de tramo	El código de tramo es un número que hace referencia al momento de la semana en que se realizó la compra
Tramo	Descripción del tramo

Tabla 4.3-1. Campos de la base de datos

4.3.1. Depuración de los datos

La base original se encontraba en el formato propietario de los productos TANGO®, por lo cual se decidió su migración a Excel para un manejo estándar del formato de los datos y para que pudiera ser compatible con los programas de análisis disponibles. Esta migración generó una cantidad de errores que debieron ser solucionados para el correcto uso de la base de datos. A continuación se mencionan los problemas encontrados y las soluciones planteadas (tabla 4.3-2):



Problema encontrado	Solución planteada
La base de datos original constaba de más de 100 tablas de información diferentes, de las cuales la mayoría se encontraban en blanco, alguna de ellas poseían sólo 2 o 3 campos, nombraban de distinta manera a un mismo campo	Se debió eliminar algunos campos completamente debido a que no se pudo identificar con certeza el significado del dato. Se unificó todas las tablas en una sola planilla con 15 campos y más de 24.000 datos quedando la misma con los datos descriptos en la tabla 4.3-1.
Muchos datos numéricos se encontraban en formato de texto (Campos artículo, hora, mozo)	Se transformó los datos textos a formato numérico.
Corrupción de ciertas filas de datos (Datos ilegibles o sin sentido)	Las filas debieron ser eliminadas completamente debido a que era imposible determinar sus datos originales
Faltante de datos en los campos "Precio Unitario" y "Total Factura".	<p>Si el error consistía en un precio faltante dentro de una factura con el importe total correcto, éste se podía obtener por diferencia. Si el error consistía en el faltante del importe total, el mismo se podía obtener por sumatoria de los ítems individuales de dicha factura. Si se daban ambos errores a la vez, se analizaba:</p> <p>Si el producto faltante se podía encontrar en otras filas de la base de datos utilizando el código de artículo, y su precio era homogéneo en el tiempo, se podía obtener dicho precio. Si el producto tenía varios precios distintos (debido a los adicionales), entonces no se podía determinar el precio correcto y la factura debía ser eliminada.</p> <p>En otro caso, el artículo se repetía en toda la base de datos sólo una vez, lo cual no era suficiente para determinar el precio, debiéndose eliminar dicho artículo –y la factura correspondiente– completamente. Si el artículo faltante era un "adicional" dentro de una factura de 6 ítems, si se borraba la factura completa significaba perder una cantidad significativa de datos de consumo. Debido a que los adicionales eran un artículo de muy bajo precio (no superaban en promedio el valor de \$1,50), se decidió eliminar el adicional y conservar la factura restante para preservar los datos.</p>
Faltante de los datos de la columna "Rubro", donde los ítems que estaban fuera del menú aparecían con la leyenda de Excel "#N/A".	Se modificó dicha leyenda por un nuevo rubro denominado "FUERA_MENU".

Tabla 4.3-2. Depuración de los datos



Finalizada la depuración de los datos, la base totalizó 23.642 filas, correspondientes a aproximadamente 5 semanas de ventas, desde el 14/6/2005 hasta el 25/7/2005. Esta base de datos tiene la particularidad de que en la fecha 13/7/2005 (fila 16.240 aproximadamente) ocurre un aumento de precios en prácticamente la totalidad de los rubros. Dicho aumento afecta los rubros de confitería (té, tortas, dulces y cuadrados), pero principalmente al café en todas sus variedades (pocillo, jarrito, capuchino, doble y con leche). Cabe mencionar que el 68,7% de los datos corresponden al período de tiempo anterior al aumento de precios.

4.4. Estudio de segmentación

4.4.1. Uso de la herramienta y armado de la base de datos

Debido a que la herramienta soporta solamente datos numéricos, el número de campos utilizables se ve reducido. De todas maneras, para el estudio que se propone realizar, los tres campos requeridos son del tipo numérico. Estos campos son “Cantidad”, “Precio” y “Total Factura”. La cantidad hace referencia al número de unidades compradas, el precio es el unitario del artículo comprado y el total de la factura es precio total que paga el cliente, incluyendo todos los ítems en la factura multiplicados por su respectiva cantidad.

Sin embargo, es evidente que la base de datos en su estado original no es útil para el análisis de segmentación que se desea hacer. La segmentación, o clustering, se debe realizar sobre los distintos clientes, es decir, para cada factura. La base de datos original posee la factura –el consumo de un cliente– disgregada entre los distintos ítems de la misma. De correr el programa en el estado original de la base, se segmentaría en base a los artículos y no la los clientes. Por esta razón es necesario modificar la base de datos para poder realizar el análisis correspondiente.

En primer lugar, se agruparon las distintas facturas en una sola línea para que cada una representara un cliente. Para evitar la tediosa tarea de realizar esta operación a mano, se creó una macro de Excel (programada en Visual Basic) que agrupaba las facturas de acuerdo al número de factura (ver anexo 3.1). Esto generó una nueva base de datos con 11.912 filas. La macro eliminaba las cantidades individuales de cada artículo y generaba un campo que contabilizaba la cantidad total de ítems dentro de la factura, y agregaba un nuevo campo que contabilizaba el promedio de precio por artículo.

Sin embargo, se detectó que en ocasiones una misma factura se encontraba dividida en dos momentos distintos del tiempo, con lo cual la macro generaba dos



o mas facturas con el mismo número pero incompletas. En total, se detectaron 142 facturas duplicadas. Para resolver esta situación, se ordenaron los datos por número de factura y se generó otra macro (ver anexo 3.2) que juntaba las facturas duplicadas en una sola. Finalmente, se obtuvo la base de datos final con 11.770 filas.

4.4.2. Aplicación de la herramienta

Introducción

Como se señaló anteriormente, se utilizaron tres variables numéricas para el estudio. Estas fueron la cantidad de ítems en cada factura, el total de dicha factura y el promedio de precio por artículo. Se utilizaron 5000 datos de la base de datos elegidos al azar mediante una macro de Excel (ver anexo 3.3): Se utilizaron 5 neuronas, lo cual permite una diferenciación máxima de 25 clusters, y 100 ciclos de entrenamiento, lo cual le da un buen parámetro de aprendizaje al programa sin extender exageradamente el tiempo de ejecución.

Las primeras pruebas se realizaron utilizando tanto los valores normalizados en corridas previas como los pesos obtenidos por el programa. Esto derivó en resultados poco satisfactorios ya que los clusters obtenidos eran muy parecidos entre sí, lo cual no permitía realizar una discriminación de los distintos grupos de clientes. El estudio se realizó nuevamente utilizando los pesos de corridas anteriores pero volviendo a normalizar los datos, lo cual entregó resultados coherentes y útiles para el análisis.

Resultados de la primera corrida

La primera corrida del programa dio como resultado tres clusters principales y una serie de “outliers” ubicados en dos sectores secundarios. La figura 4.4.2-1 a continuación muestra los resultados de esta primera corrida.

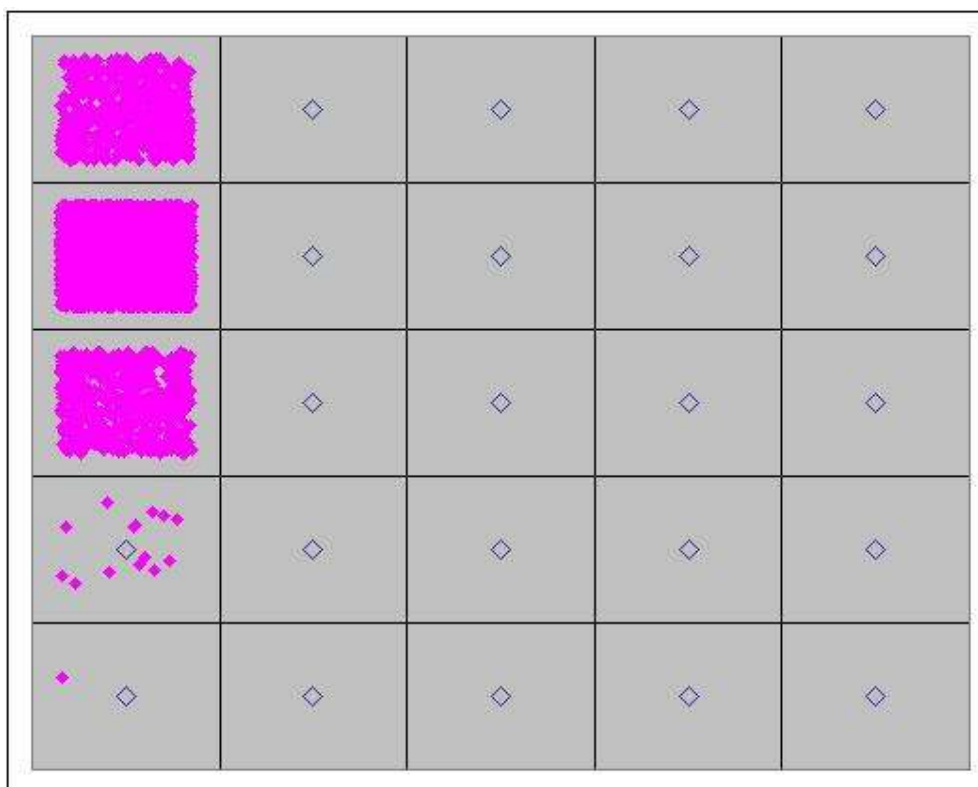


Figura 4.4.2-1. Gráfico de la segunda corrida del NNClust:

Se puede observar que existe una alta concentración de datos en un cluster en particular. Esto es esperable ya que el programa aún no alcanzado un nivel avanzado de aprendizaje. El detalle de los resultados se encuentra en la tabla 4.4.2-1 a continuación.

	General	Cluster				
		1	2	3	4	5
Columna en el mapa		1	1	1	1	1
Fila en el mapa		1	2	3	4	5
Tamaño del Cluster	5000	1	14	515	3919	551
TOTAL de la factura	10,2	132,5	91,9	30,9	6,3	16,0
CANTIDAD de ítems	2,0	37,0	11,6	4,0	1,5	3,3
PRECIO por ítem	4,9	3,6	9,9	8,8	4,3	5,2

Tabla 4.4.2-1. Resultados de la primera corrida del NNClust

Todos los valores de las tablas son promedios de cada cluster. Es decir, el “total de la factura” es el valor promedio de todas las facturas de dicho cluster. Lo mismo ocurre con la cantidad de ítems por factura y el precio promedio de los ítems.



Claramente los clusters 1 y 2 son outliers, y corresponden a ciertas facturas con gran cantidad de ítems y con alto valor total. Probablemente correspondan a grupos numerosos de personas que almuerzan o cenan en el local.

El cluster 4 es de el mayor población, concentrando un 80% de los datos. Tiene un consumo bajo en cantidad y en la suma total. Lo más probable es que sea gente que toma solo un café. Los otros dos clusters tienen una población similar, con un consumo similar de ítems pero con distintos valores de precio.

Como se puede apreciar, esta primera corrida posee una alta concentración de datos, lo cual atenta contra el objetivo de la misma, es decir, la discriminación de los distintos segmentos de la clientela. Por esta razón, es necesario realizar una nueva corrida del programa.

Resultados del la segunda corrida

El gráfico siguiente (figura 4.4.2-2) muestra los resultados de la segunda corrida.

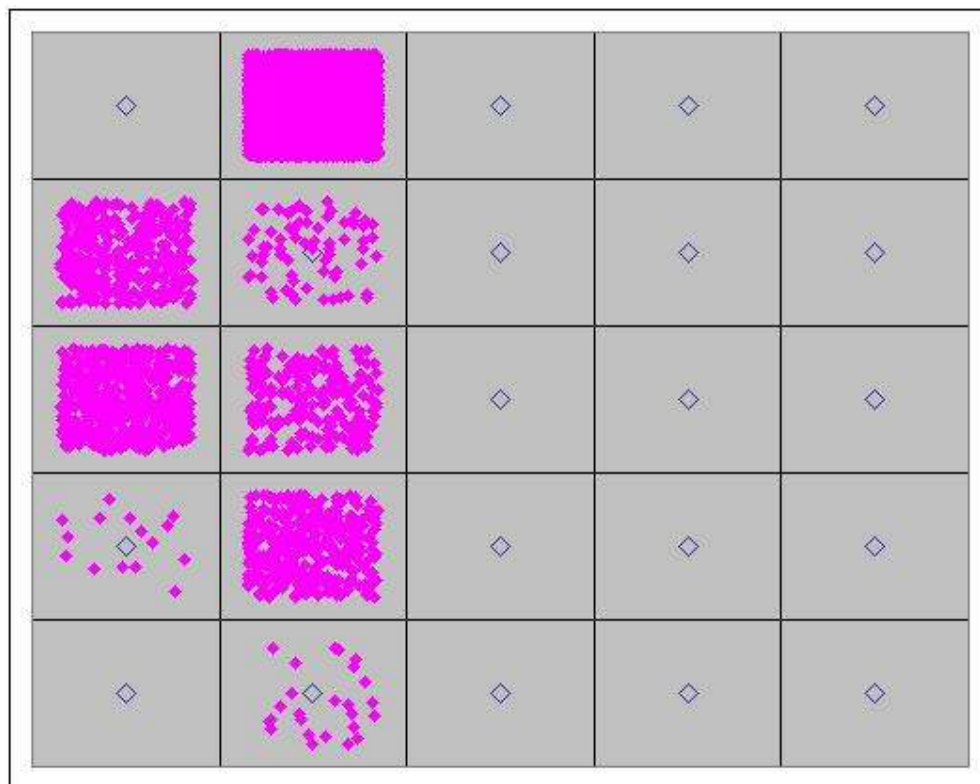


Figura 4.4.2-2. Gráfico de la segunda corrida del NNClust:

Se pueden apreciar en este caso 5 clusters principales y 3 secundarios, dando un total de 8 clusters. Se aprecia nuevamente un cluster muy denso. Los resultados de esta corrida se encuentran en la tabla 4.4.2-2 a continuación.



	General	Cluster							
		1	2	3	4	5	6	7	8
Columna en el mapa		1	1	1	2	2	2	2	2
Fila en el mapa		2	3	4	1	2	3	4	5
Tamaño del Cluster	5000	15	489	309	25	372	195	69	3526
TOTAL de la factura	10,2	94,6	31,5	13,2	16,0	17,9	12,5	13,4	5,5
CANTIDAD de ítems	2,0	13,3	4,0	1,6	2,6	3,4	3,0	2,7	1,5
PRECIO por ítem	4,9	9,4	9,0	8,7	6,6	5,8	4,3	5,0	3,9

Tabla 4.4.2-2. Resultados de la segunda corrida del NNClust

Esta corrida posee un poder de discriminación superior, puesto que el número de clusters permite visualizar algunos grupos representativos. El cluster mayor (cluster 8), que en este caso concentra al 70% de los clientes, sigue representando a los consumidores de baja facturación y cantidad. Por otra parte, el cluster 1 agrupa aquellas facturas con alto valor total y cantidad. En el medio de éstas existe toda una gama de facturas de consumos medios con distintas combinaciones de cantidad de ítems y promedios de precios.

Esta corrida sin duda podría ser útil para analizar a la clientela del local. Es evidente que el nivel de aprendizaje del programa es alto ya que en la segunda corrida se obtienen resultados aceptables. Si bien los resultados de esta corrida son válidos, se decide realizar una nueva corrida para verificar si se logra una mayor segmentación de los clientes.

Resultados de la tercera corrida

En esta corrida se obtuvieron 12 clusters, de los cuales 7 pueden ser considerados principales. El gráfico (figura 4.4.2-3) y la tabla 4.4.2-3 siguiente muestran los resultados.

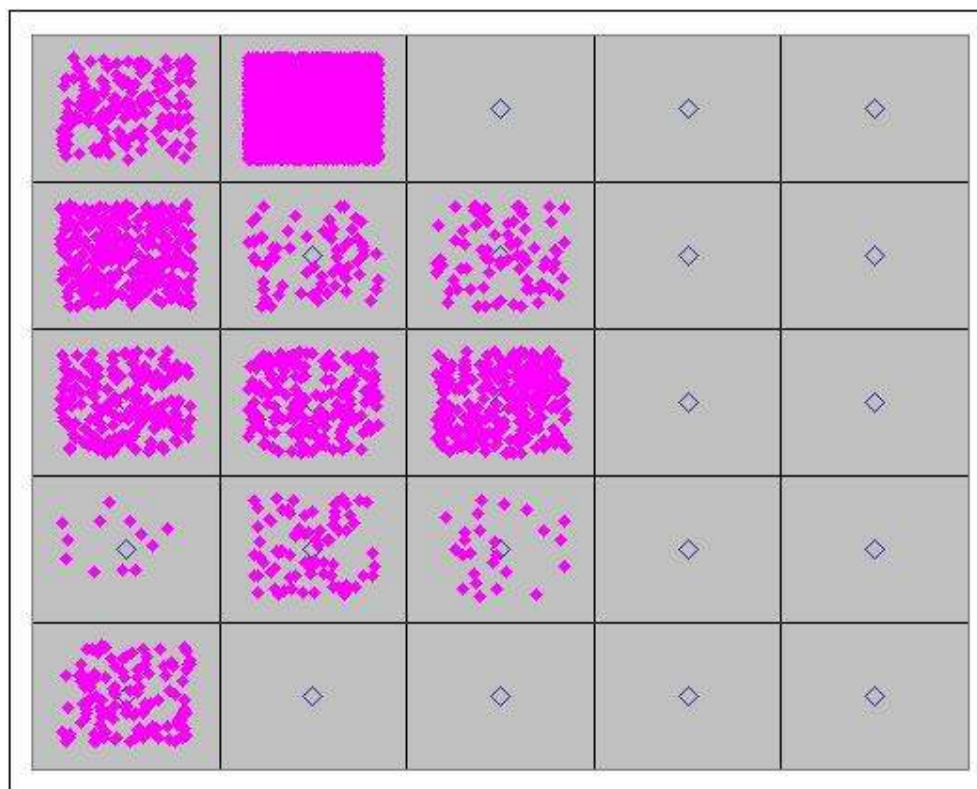


Figura 4.4.2-3. Gráfico de la tercera corrida del NNClust.

	General	Cluster											
		1	2	3	4	5	6	7	8	9	10	11	12
Columna en el mapa		1	1	1	1	1	2	2	2	2	3	3	3
Fila en el mapa		1	2	3	4	5	2	3	4	5	2	3	4
Tamaño del Cluster	5000	117	12	191	309	151	79	194	69	3526	25	252	75
TOTAL de la factura	10,2	45,8	101,9	30,9	13,2	22,0	16,3	12,5	13,4	5,5	16,0	18,0	24,8
CANTIDAD de items	2,0	4,0	13,8	3,8	1,6	4,3	3,0	3,0	2,7	1,5	2,6	3,4	4,1
PRECIO por item	4,9	13,8	10,3	8,7	8,7	5,4	5,7	4,3	5,0	3,9	6,6	5,9	7,0

Tabla 4.4.2-3. Resultados de la tercera corrida del NNClust.

Puede observarse que muchos de los clusters encontrados en esta corrida son prácticamente idénticos a la corrida anterior. Este es el caso de los clusters 2, 4, 7, 8, 9 y 10, que corresponden a los clusters 1, 3, 6, 7, 8 y 4 de la segunda corrida. Esto significa que los restantes 2 clusters de la corrida anterior (2 y 5) se dividieron en 6 nuevos en la tercera corrida (1, 3, 5, 6, 11 y 12).

Ante este panorama se hace difícil dilucidar si la tercera corrida realmente aporta al objetivo del trabajo. Se hace necesario analizar detalladamente los nuevos clusters para arribar a una conclusión. La tabla 4.4.2-4 muestra todos los clusters que varían entre la segunda y la tercera corrida.



	Clusters 2a. corrida		Clusters 3a corrida					
	2	5	1	3	5	6	11	12
Columna en el mapa	1	2	1	1	1	2	3	3
Fila en el mapa	3	2	1	3	5	2	3	4
Tamaño del Cluster	489	372	117	191	151	79	252	75
TOTAL de la factura	31,5	17,9	45,8	30,9	22,0	16,3	18,0	24,8
CANTIDAD de ítems	4,0	3,4	4,0	3,8	4,3	3,0	3,4	4,1
PRECIO por ítem	9,0	5,8	13,8	8,7	5,4	5,7	5,9	7,0

Tabla 4.4.2-4. Comparación de clusters diferentes entre segunda y tercera corrida

Se puede apreciar en la tabla que los nuevos clusters de la tercera corrida son diferentes entre sí, y a su vez difieren de los de la segunda corrida. Estas diferencias se aprecian sobre todo en los campos TOTAL y PROMEDIO. Por otro lado, la mayoría de nuevos clusters supera los 100 datos, lo cual les otorga la representatividad necesaria. De esta manera se puede concluir que la tercera corrida es útil para el análisis ya que aporta nueva información acerca de los clientes de mayor consumo.

Se realizó una cuarta corrida que dio como resultado 16 clusters. En esta ocasión 10 clusters permanecieron constantes y, al igual que en la tercera corrida, los restantes 2 clusters de la corrida anterior se dividieron en 6. Sin embargo, en este caso los nuevos clusters tenían un tamaño muy reducido (la mayoría no superaba los 60 datos), lo cual significaba que la segmentación era tan grande que algunos clusters perdían representatividad. Por esta razón se decidió descartar esta última corrida y utilizar la tercera corrida para realizar el análisis. Los datos de la cuarta corrida se encuentran en el anexo 4.

4.4.3. Análisis de los resultados

Debido al nivel de representatividad de los clusters de la tercera corrida, los mismos fueron seleccionados para realizar el análisis de la clientela del local. La tabla 4.4.3-1 muestra los resultados de dicha corrida, y además agrega tres filas que representan el porcentaje de la población de cada cluster sobre el total de los datos, la facturación total de cada cluster en el período de análisis, expresada como el consumo promedio de cada factura por la población del cluster, y por último el porcentaje de la facturación.



	General	Cluster											
		1	2	3	4	5	6	7	8	9	10	11	12
Tamaño del Cluster	5000	117	12	191	309	151	79	194	69	3526	25	252	75
Porcentaje Clientes		2,3%	0,2%	3,8%	6,2%	3,0%	1,6%	3,9%	1,4%	70,5%	0,5%	5,0%	1,5%
TOTAL de la factura	10,2	45,8	101,9	30,9	13,2	22,0	16,3	12,5	13,4	5,5	16,0	18,0	24,8
CANTIDAD de ítems	2,0	4,0	13,8	3,8	1,6	4,3	3,0	3,0	2,7	1,5	2,6	3,4	4,1
PRECIO por ítem	4,9	13,8	10,3	8,7	8,7	5,4	5,7	4,3	5,0	3,9	6,6	5,9	7,0
Facturación	50766	5364	1223	5901	4079	3327	1287	2420	923	19433	401	4544	1863
Porcentaje Fact.		11%	2%	12%	8%	7%	3%	5%	2%	38%	1%	9%	4%
<i>Relación Fact./Pobl.</i>		4,5	10,0	3,0	1,3	2,2	1,6	1,2	1,3	0,5	1,6	1,8	2,4

Tabla 4.4.3-1. Detalle de los clusters

Por su parte, la última fila es un índice que expresa, en términos relativos, cuán rentable es un cluster en función de su población. Se calcula como el porcentaje de facturación sobre el porcentaje de clientes. A continuación se analizan los clusters detalladamente. Cabe aclarar que ciertas inferencias en cuanto a los consumos se realizan en base a conversaciones con los responsables del local, aunque están respaldadas por los resultados de las corridas.

- Cluster 9: Este cluster concentra el 70,5% de las facturas y casi el 40% de la facturación. Son aquellos consumos que conforman la base de la clientela del local. Corresponden a facturas de bajo valor total y pocos ítems consumidos (entre 1 y 2). Seguramente responde a personas solas que toman un café con algún dulce o parejas que toman solamente un café cada uno.
- Clusters 7 y 8: Estos dos clusters son muy parecidos en cuanto a sus valores característicos. Corresponden a consumos con un promedio de 3 ítems de bajo precio. Probablemente corresponden a una persona sola que desayuna o merienda o a un grupo de personas reducido que toman solamente un café.
- Clusters 6, 10 y 11: Estos tres clusters tienen valores similares del total de la factura y precio promedio de los ítems. Cabe recalcar que entre los tres clusters representan el 13% de la facturación. No es fácil discernir qué productos consumen ya que todos sus valores son medios. Probablemente son grupos de 2 personas que toman el té con algún acompañamiento mediano tal como los tostados o las medialunas.
- Cluster 5: Este cluster posee un valor intermedio en cuanto al total, pero con una cantidad mayor de ítems de bajo precio. No es un cluster despreciable ya que con el 3% de la clientela representa el 7% de la facturación. Seguramente agrupa a grupos de tres o cuatro personas que toman solo café.



- Cluster 12: Parecido al cluster anterior, con la diferencia de que el promedio de precio de los ítems es mayor. Podrían ser dos o tres personas que toman algo más que un café pero menos que un almuerzo, por ejemplo, un sándwich con una bebida. Con solo 1,5% representa el 4% de la facturación.
- Clusters 3 y 4: Dos clusters muy importantes en cuanto a sus valores, ya que juntos representan el 10% de los clientes y el 20% de la facturación. Por los altos valores promedio de los ítems seguramente son personas que almuerzan en el local. La diferencia reside en la cantidad de ítems: el cluster 4 son personas que almuerzan solas mientras que el cluster 3 son grupos de 2 o 3 personas.
- Cluster 1: Este cluster representa por sí mismo el 11% de la facturación, teniendo sólo el 2,3% de los clientes. Tiene un precio promedio por ítem muy alto, con 4 ítems por factura y cuyo total tiene una media de casi \$46. Puede concentrar a distintos clientes: parejas que cenar, almuerzos grandes o grupos de personas que consumen picadas acompañadas por bebidas alcohólicas. Independientemente del tipo de consumo, este cluster es de alta importancia para el local y ofrece gran potencial para su explotación.
- Cluster 2: Este cluster agrupa a una serie de “outliers” que poseen un muy alto valor total en las facturas, que supera los \$100. También es notable la gran cantidad de ítems por factura, que supera los 13 productos, dando como promedio unos \$10 por ítem. Este cluster puede representar a grupos numerosos de personas que almuerzan o cenar en el local. Es interesante notar que si bien este cluster sólo representa el 2% de la facturación, lo hace con sólo un 0,2% de las facturas. Esto significa que cada factura tiene un peso en la facturación de 10 veces su población.

Una observación interesante se basa en el comportamiento de los clientes, que responden a una distribución de Pareto. Esto se puede observar en el hecho de que la mayoría de los clientes generan una parte relativamente baja de la facturación y viceversa. Con solo observar el cluster 9 se puede apreciar que 70,5% de los clientes le representan al local solamente el 38% de las ventas. Los tres clusters siguientes en tamaño (4, 7 y 11), que representan el 15,1% de la población, poseen a su vez el 22% de la facturación. Y por último, los 8 clusters restantes, con solo un 14,1% de la población, generan el 40% de la facturación. La tabla 4.4.3-2 y la figura 4.4.3-1 ilustran esta situación.



	A	B	C
Cant. de Clusters	1	3	8
Población	70,5%	15,1%	14,4%

Tabla 4.4.3-2. Agrupamiento de clusters

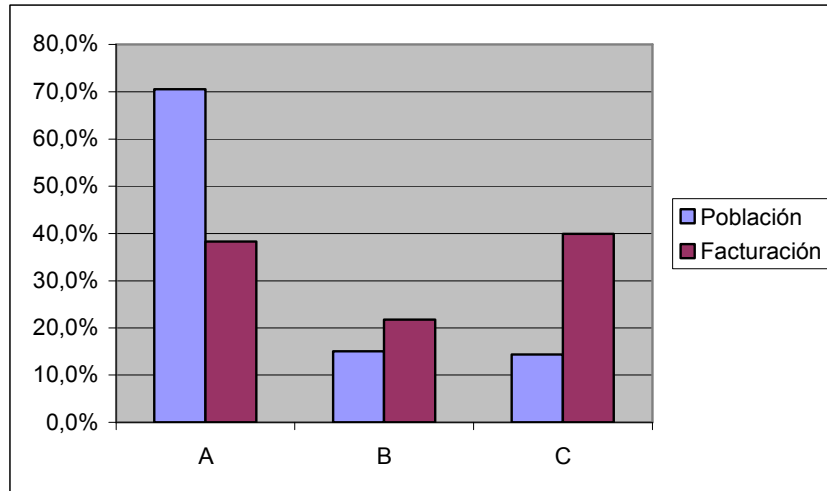


Figura 4.4.3-1. Facturación y Población de los grupos de clusters

Se han denominado a los tres grupos como A, B y C, tal como se realiza en la teoría de Pareto. El principio de Pareto establece que en una población dada existen una serie de grupos que se pueden dividir entre los “pocos vitales y los muchos triviales”. Sin duda que en este caso no se puede tildar al grupo A de “trivial”, ya que concentra al 70% de los clientes. Pero por otro lado no se puede dejar de lado el hecho de que un grupo reducido de clientes como el C, con un 14% de los clientes, genera mayor facturación que el primer grupo. Evidentemente, esto da una pauta de hacia donde deben dirigirse los esfuerzos de promoción.

Sin ir más lejos, el cluster 1 presenta un interesante potencial para ser explotado. Si bien no está perfectamente definido el perfil de los clientes, se trata sin duda de un grupo único. Tanto parejas que cenar como grupos de personas que consumen picadas, tablas y bebidas alcohólicas, todos los clientes de este sector pertenecen probablemente a un estrato social alto. La relación de porcentajes de facturación a población es de 4,5, la segunda de todos los clusters y la primera para los clusters principales (aquellos con más de 100 datos). Algo similar ocurre con el cluster 2, aunque con números mucho menores. Con sólo 0,2% de la población genera un 2% de facturación, lo cual significa un índice de relación de 10, el cual coloca a este cluster, con gran margen, como el más rentable de todos los clusters.



Es muy interesante traducir el potencial de estos dos clusters en números. Juntos facturan \$ 6587, que representan el 13% de la facturación total. Ahora bien, duplicar esta facturación significaría incorporar 129 nuevas facturas (o incrementar el consumo de facturas similares de los clientes actuales). Por otra parte, generar un incremento de \$ 6587 en el cluster 9 significaría incorporar 1195 facturas. Sin duda, influir sobre casi 1200 clientes requerirá un esfuerzo de marketing mucho mayor, tanto mental como monetario, que hacerlo sobre 129 clientes.

Si bien no es el objetivo puntual de este trabajo, se pueden utilizar los resultados para generar algunas propuestas de mejora en base a los resultados. Por ejemplo, en el caso de los clusters anteriores, se podría pensar en posicionar al Coffee Store como una alternativa viable para cenar o tomar un “drink”. Una forma de hacer esto es ampliar el menú incorporando platos de mayor categoría. Incluso podría ambientarse un sector del local como un restaurante, moviendo a la gente a adoptar al Coffee Store como una alternativa para cenar. Por otra parte, no se puede dejar de lado el potencial del Coffee Store como lugar de reunión, tanto de grupos de amigos como grupos de negocios. En este sentido, la opción de las picadas, un producto con alto rendimiento, debe ser tenida en cuenta. Se puede ampliar la oferta de picadas y tablas e incorporar otras alternativas novedosas, tales como las “tapas” españolas.

Por otra parte, no es posible dejar de lado al cluster 9, que contiene a la masa de clientes del local. Además de esto, este cluster nuclea a los clientes que consumen el producto insignia de la franquicia, el café. Un incremento de tan sólo \$ 1 (18% del valor promedio actual) en el consumo de todos los clientes genera \$ 3526 de facturación adicionales, lo cual representa un 7% del total. No es despreciable el potencial para crear nuevos negocios dentro de este grupo.

Podría suceder que los dueños del local no desearan cambiar el posicionamiento del local sino atraer nuevos clientes que pertenecieran a este cluster. En este caso se podría, por ejemplo, crear promociones relacionadas con el café, asociando productos, generando ventas atadas y aumentando la facturación. Un ejemplo de esto podría ser ofrecer una tostada, medialuna, barra de cereal o cualquier otro producto pequeño con el pocillo a un precio menor que los dos productos por separado.

Otro grupo interesante está conformado por los clusters 3 y 4. Estos dos clusters representan juntos el 10% de los clientes y el 20% de la facturación. Tal concentración de valor no puede ser pasada por alto en el análisis. Por lo visto anteriormente, la gente que conforma estos clusters seguramente almuerza en el local. El cluster 4, el más numeroso de los dos, agrupa personas que almuerzan



solas, mientras que en el cluster 3 son parejas o grupos de tres personas. Así como para los clusters 1 y 2, se requiere en este caso generar alternativas para la hora del almuerzo en el local, de tal forma de captar nuevos clientes y de fidelizar a los actuales. Teniendo en cuenta que mucha gente almuerza todos los días en el local, se podría crear un “menú del día” rotativo. Cada día, los clientes dispondrían de una o dos opciones distintas para comer, y no se “aburrirían” de almorzar todos los días en el local.

Todos los clusters restantes también deben ser tenidos en cuenta. No se puede obviar el hecho de que con el 15% de los clientes representan un cuarto de la facturación total del local. Las características de estos grupos son variados, desde personas solas que desayunan hasta grupos de dos o tres que almuerzan ligero. Las acciones a tomar en estos casos serían totalmente distintas, dependiendo del cluster en particular, pero todas deberían tener un objetivo en común: crear nuevos posicionamientos que a su vez generen nuevas necesidades en los clientes.

4.4.4. Conclusiones del uso de la herramienta

El análisis de los clusters tiene como objetivo segmentar a los clientes para poder analizar sus características y actuar en consecuencia. Esto es particularmente útil a la hora de decidir cambios e introducir promociones. Como se ha visto anteriormente, si se hace foco sobre un grupo reducido de clientes se pueden obtener los mismos o mejores resultados que tratando de abarcar a la masa de clientes, con un costo notablemente menor. Esto muestra la importancia de realizar un análisis exhaustivo de los datos existentes de forma tal de no malgastar recursos.

Otro punto a tener en cuenta al analizar los clusters es el factor humano. No se pueden estudiar los segmentos como meros números, sino que se deben tener en cuenta las características e idiosincrasia de los clientes. Como ya se ha explicado anteriormente, este Coffee Store se encuentra ubicado en Belgrano, uno de los barrios con mayor ingreso per cápita de la ciudad. Es evidente que los vecinos del barrio no aceptarán al Coffee Store como una opción de primera para cenar, ya que seguramente lo hacen en restaurantes especializados de alta categoría. El desafío es establecer la idea del local como una segunda opción para cenar: no para el sábado a la noche, si no para esos días en los que no se tienen ganas de cocinar. Existen muchas maneras de lograr este posicionamiento. La clave es entender al cliente, cómo piensa y qué necesita.



4.5. Estudio de Clasificación

4.5.1. Uso de la herramienta y preparación de la base de datos

La herramienta NNClass se utilizó con el fin de encontrar patrones de consumo de distintos productos en función al día y horario en que los clientes concurren al local.

Los productos existentes en la base de datos se encuentran en la tabla 3. En la preparación de la base de datos se detectó la existencia del rubro 53, “Adicionales”, que contaba con unas pocas líneas en toda la base de datos. Debido a que su presencia no permitía correr la herramienta correctamente, se decidió juntar al rubro 53 con el 10, que también era “Adicionales”, pero con una cantidad de datos mucho mayor. Al rubro 99 se lo designó “Fuera Menú”, tal como lo indicaba el número de artículo. Por otra parte, se puede observar que hay dos rubros que hacen referencia a “sándwiches” (rubros 21 y 46). Sin embargo, tal como fue constatado con los contactos del local, estos dos productos son totalmente distintos y no es conveniente juntarlos, como se verá más adelante. Lo mismo ocurre con las ensaladas (rubros 27 y 41). Por lo tanto, en total quedan 39 rubros los mismos se encuentran definidos en la tabla 4.5-1.

Código	Rubro / Producto
2	POCILLO
3	JARRITO
4	DOBLE
5	CAFE CON LECHE
6	TE
8	CAPUCCINO ITALIANO
9	CAPUCCINO
10	ADICIONALES
11	TRAGOS CON CAFÉ
14	BEBIDAS S/ ALCOHOL
16	DULCES
18	TORTAS
21	SANDWICHES
22	PEQUEÑAS DEGUSTACIONES
23	BRUSCHETTAS
24	TABLITAS
25	QUICHES
26	CREPES
27	ENSALADAS
32	MERCHANDISING
33	CAFE AL PESO
36	CUADRADOS



Código	Rubro / Producto
37	BEBIDAS C/ ALCOHOL
38	LACTEOS
40	CARNES
41	ENSALADAS
42	VARIOS
43	PASTAS
44	PESCADOS
45	POLLO
46	SANDWICHES
47	VERDURAS
48	MERIENDA
49	DESAYUNO
50	SOPAS
51	LICUADOS MILKSHAKE
52	HELADOS
54	PROMOCIONES KALDI
99	FUERA_MENU

Tabla 4.5-1. Rubros de productos

4.5.2. Aplicación de la herramienta

En primer lugar se utilizó el campo “código de tramo” como variable categórica. El código de tramo indica el momento de la semana en que ocurre la compra: lunes a jueves, viernes y sábado, y domingo, diferenciados entre AM y PM. Este estudio dio como resultado patrones más bien regulares, que no permitían obtener conclusiones aceptables. A continuación se decidió crear un nuevo campo que indicara la hora del día. Para esto se dividió al día en “franjas” (mañana, media mañana, mediodía, tarde, noche y traspasado). Este estudio dio resultados satisfactorios, los cuales permitieron obtener conclusiones. En las secciones siguientes se detallan los dos estudios realizados y las conclusiones generales derivadas del uso de la herramienta de clasificación NNClass.

4.5.3. Análisis del Código de Tramo

El primer estudio se realizó utilizando el campo “Código de Tramo”. El código de tramo es un campo que ya existía en la base de datos y que divide a la semana en 6 secciones. A continuación, en la tabla 4.5.3-1, se muestran los distintos tramos.



Código de Tramo	Tramo
1	Lun a Jueves AM
2	Lun a Jueves PM
3	Viernes y Sábados AM
4	Viernes y Sábados PM
5	Domingo AM
6	Domingo PM

Tabla 4.5.3-1. Descripción de tramos

El corte entre la mañana (AM) y la tarde (PM) ocurre a las 16 horas de cada día, y los horarios de madrugada se computan como la noche del día anterior. Para ambos estudios se realizaron corridas consecutivas del programa hasta lograr los resultados entre las mismas no sufriesen grandes variaciones. Esto ocurrió en la tercer corrida. Cabe mencionar que en las primeras corridas de ambos estudios el tramo 6 no fue identificado por el programa, aunque su presencia fue correcta en las siguientes corridas. Probablemente esto se deba al funcionamiento del programa, que utiliza ciertos datos para validar los resultados. Tal vez el hecho de que el tramo 6 es el que aparece último en la base de datos haya influido en su ausencia de la primera corrida. Sin embargo, no es clara la razón real de este hecho. De todas maneras, como se verá a continuación, esta anomalía no afecta los resultados.

La primera parte, es decir los datos anteriores al aumento de precios, totalizaba unas 16800 filas, mientras que la segunda parte contenía aproximadamente 6900 filas. Luego de depurar los datos se utilizaron 6812 datos de la segunda parte. De tal forma de que los estudios fueran comparables, se decidió utilizar la misma cantidad de datos para la primera parte. Para esto se implementó una macro de Visual Basic para seleccionar los datos al azar. La misma se encuentra en el anexo 3.4.

Resultados obtenidos del primer análisis

Los resultados para la primera parte (antes del aumento de precios) se encuentran en la figura 4.5.3-1.

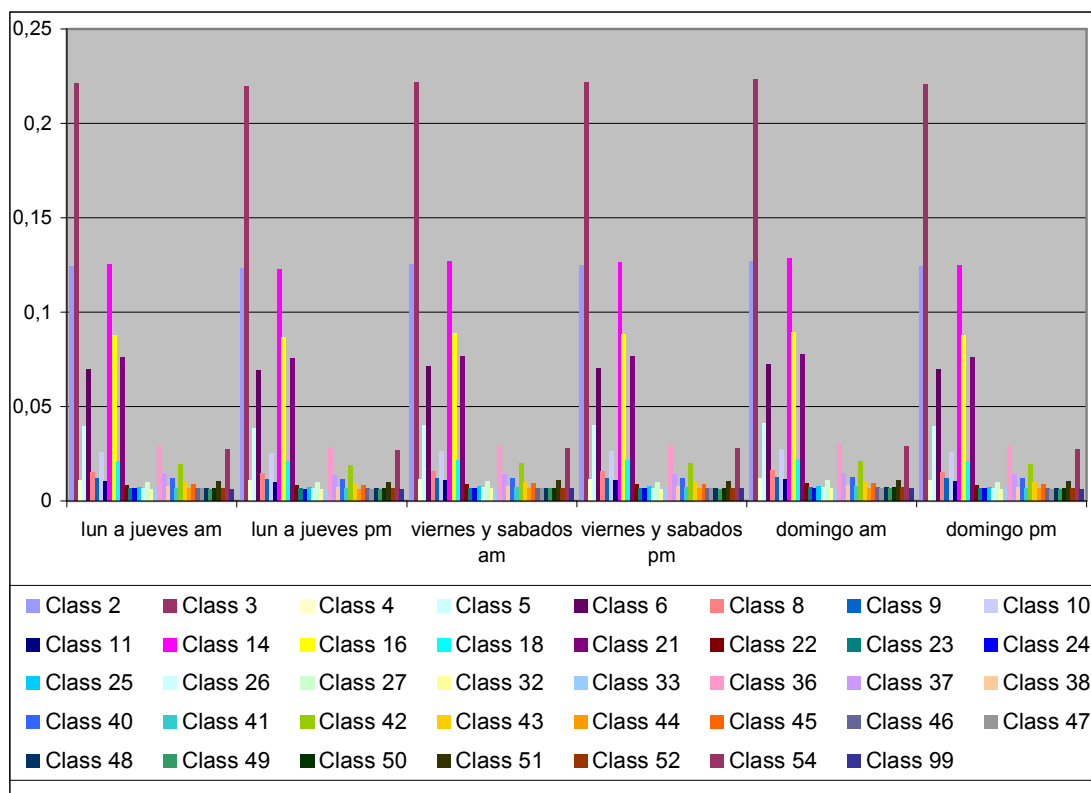


Figura 4.5.3-1. Resultados del primer análisis del NNClass para el código de tramo, con selección de datos aleatoria

Como se puede observar en el gráfico, no es posible definir un patrón demasiado claro en este estudio, ya que prácticamente no hay diferencias entre los distintos tramos. En mayor o menor medida, todos los rubros siguen el mismo patrón regular. Primero, para un mismo día el consumo es siempre mayor en el tramo AM (mañana-mediodía) que en PM (tarde-noche) correspondiente. Segundo, el menor consumo se da en los tramos de lunes a jueves. Tercero, el domingo AM tiene el mayor consumo del fin de semana, mientras que el domingo PM tiene el menor. La única variación puede surgir en si el tramo domingo PM es mayor o menor que los tramos de lunes a jueves. Sin embargo la diferencia es tan pequeña que los resultados de analizar esto no serían significativos. En el Anexo 5.1.1 se pueden encontrar todos los gráficos detallados por grupos de rubros.

En general los resultados obtenidos son acordes a la lógica. Los días de semana hay menor consumo que en el fin de semana debido a que la gente se encuentra trabajando. En general, los habitantes de Belgrano trabajan en el centro de la ciudad. Por otra parte, el hecho de que haya más consumo durante la primera parte del día es atendible, ya que es más común que la gente almuerce en una confitería a que cene.

La razón de la regularidad en los patrones no es clara. Probablemente el hecho de que los datos hayan sido elegidos al azar puede haber influido en este patrón.



También podría ocurrir que los parámetros del programa no estén correctos y deban ser modificados. Sin embargo, se realizaron dos pruebas adicionales con otros parámetros que dieron resultados casi idénticos. En las condiciones actuales, esta parte del estudio no sirve para obtener conclusiones válidas ya que no hay diferencias mayores entre rubros.

Como consecuencia de esto se decidió realizar un segundo estudio. Ante la posibilidad de que la selección aleatoria de datos haya influido negativamente en los resultados, se tomaron los primeros 6812 datos de la base de datos. De esta forma se aseguraba que los datos fueran homogéneos ya que eran datos consecutivos. Sin embargo, los resultados, como lo demuestra la figura 4.5.3-2 más adelante, fueron muy similares al caso anterior, lo cual demuestra que probablemente el patrón regular sea correcto. Resta analizar la segunda parte del estudio, lo cual se encuentra en la siguiente sección.

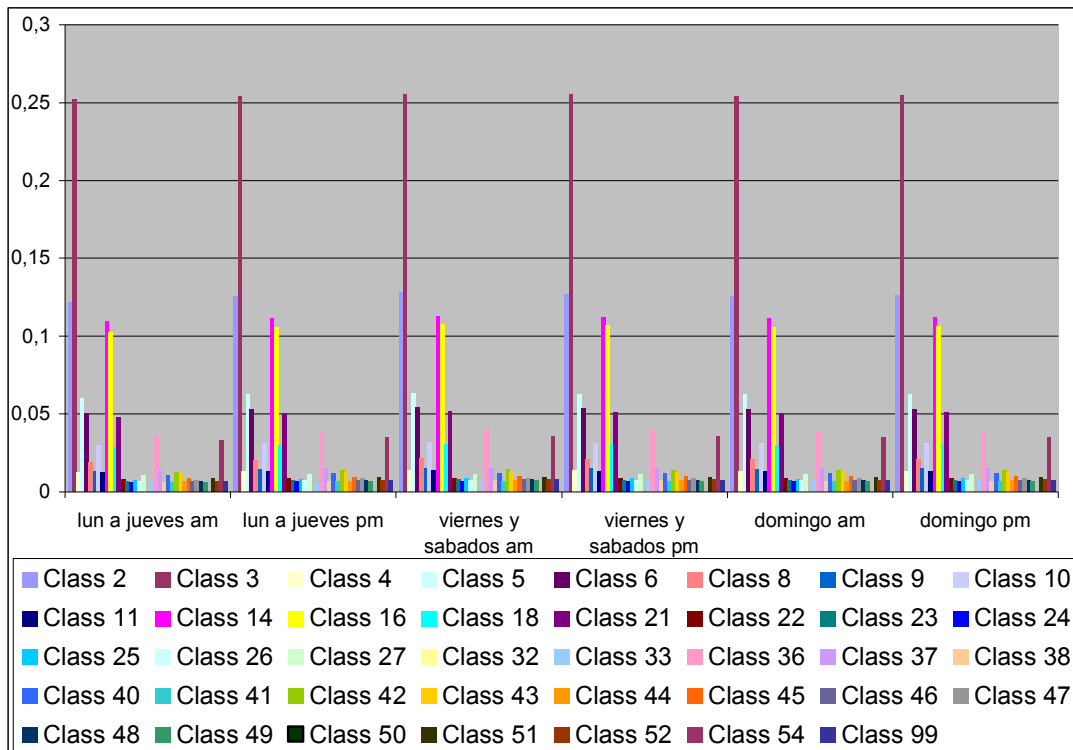


Figura 4.5.3-2. Resultados del primer análisis del NNClass para el código de tramo, con selección de datos continuos.



Resultados obtenidos del segundo análisis

La segunda parte del estudio derivó en mejores resultados que la parte anterior, ya que no sólo el patrón deja de ser regular, sino que además existen diferencias entre los distintos rubros. La figura 4.5.3-3 muestra los resultados obtenidos.

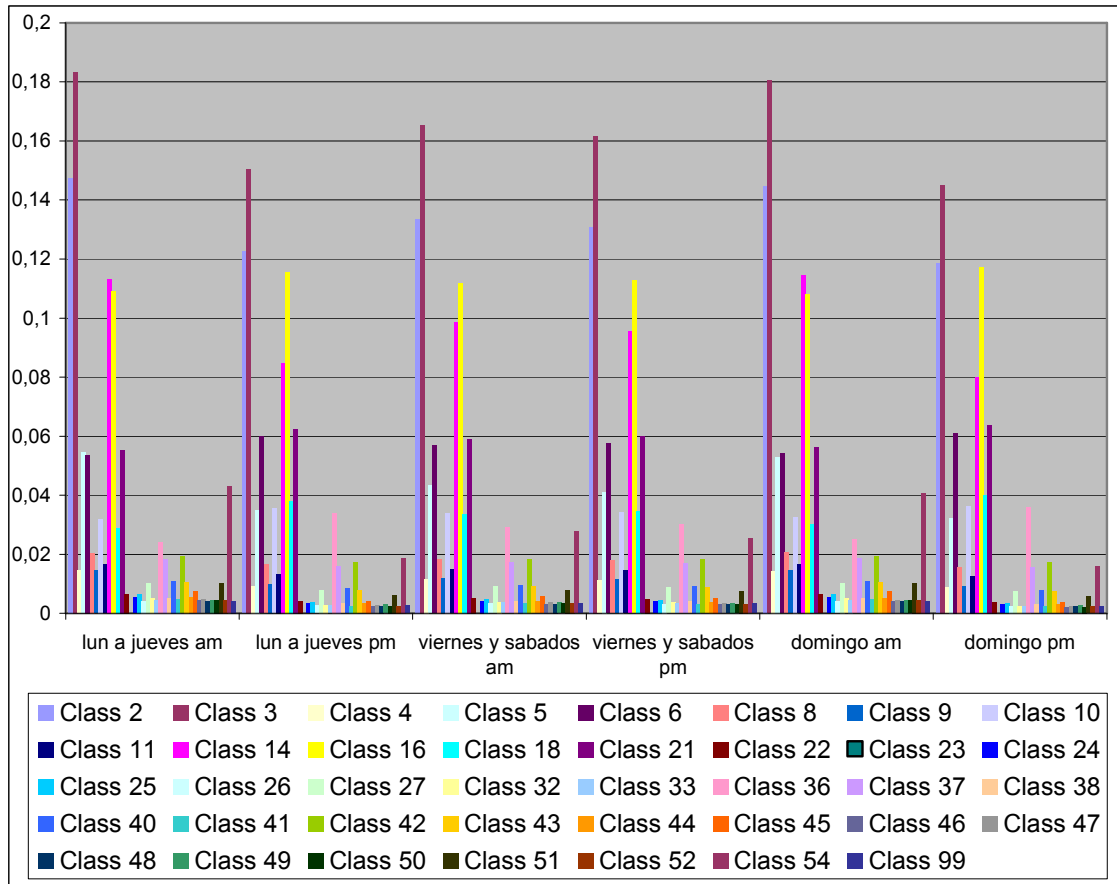


Figura 4.5.3-3. Resultados del segundo análisis del NNClass para el código de tramo

Se pueden apreciar en el gráfico algunos productos destacables. Este es el caso del pocillo (clase 2) y el jarrito (clase 3). Estos dos productos no solo son los primeros dos en relevancia, sino que además presentan el patrón más notable. En todos los tramos el jarrito posee un consumo mayor al del pocillo. Si se observan los dos gráficos de la primera parte (figuras 4.5.3-1 y 4.5.3-2) se puede apreciar que la brecha entre pocillo y jarrito todavía existe, pero se ha achicado notablemente. Se puede explicar este comportamiento teniendo en cuenta el aumento de precios ocurrido entre las dos partes. Luego de dicho suceso, mucha gente probablemente comienza a consumir el pocillo debido a su menor costo. Este análisis da la pauta de que los resultados de la primera parte, si bien no son muy útiles por sí mismos, sirven para sacar conclusiones contrastándolos con la segunda parte del estudio.

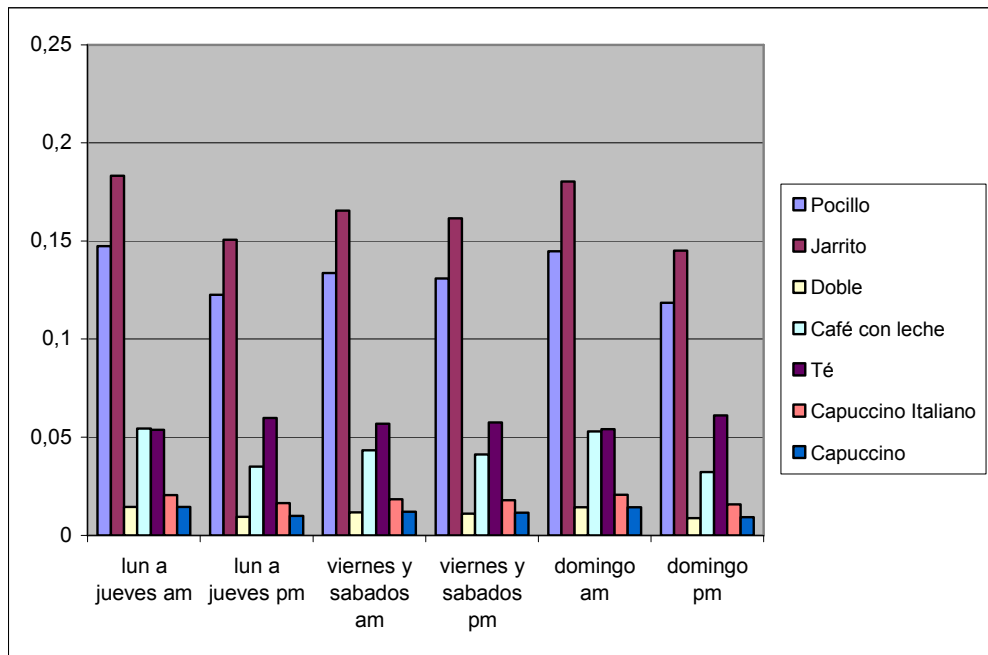


Figura 4.5.3-4. Detalle de las infusiones por código de tramo.

Por otra parte, se puede observar que los tramos AM estos productos siempre poseen un consumo mayor a sus correspondientes tramos PM. Es interesante notar el caso del té (clase 6), que sigue un patrón inverso como se puede observar en el gráfico 4.5.3-4 con mayor detalle. Esto es entendible ya que el té es comúnmente consumido durante la tarde mientras que el café es la bebida preferida durante la mañana. Además, el café se consume habitualmente como sobremesa luego de las comidas. Como se ha dicho anteriormente, el local recibe más clientes durante el almuerzo que en la cena, lo cual refuerza aún más la validez de los resultados.

Otros rubros importantes son las bebidas sin alcohol (clase 14), los dulces (clase 16), los sándwiches (clase 21) y las promociones KALDI (clase 54). Las promociones KALDI consisten en un menú light compuesto por “quiches” (tartas pequeñas), ensalada bebida y café. La figura 4.5.3-5 a continuación muestra a estos rubros en detalle. Se observa que las Promociones KALDI tienen un alto consumo en los tramos AM y bajo en los tramos PM, sobre todo en los días de semana. Esto es lógico ya que dicho menú tiene un tamaño reducido y es más apto para un almuerzo que para una cena. Es común que la gente que trabaja en la zona, sobre todo muchas mujeres que atienden locales en los alrededores, opten por este menú light en los días de semana.

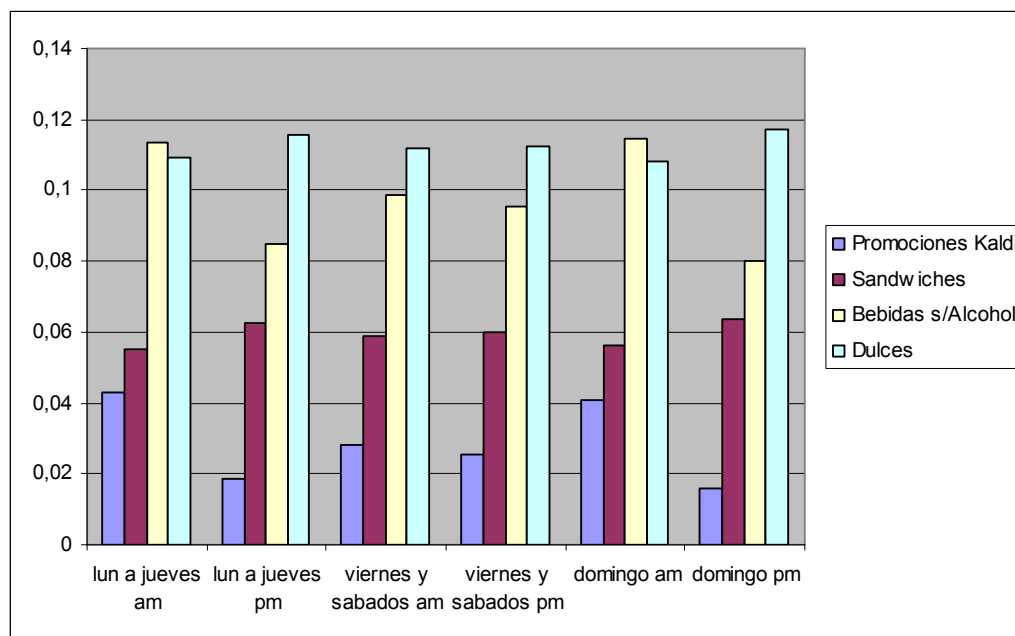


Figura 4.5.3-5. Detalle de los rubros principales por código de tramo.

Se puede observar que tanto los sándwiches como los dulces siguen un patrón diferente a gran parte del resto de los productos (ver gráficos en anexo 5.1.2). A diferencia de la mayoría, estos dos productos tienen su consumo más fuerte durante los tramos PM. En ambos casos, el aumento más elevado ocurre el domingo a la tarde. En el caso de los sándwiches esto podría deberse a que la gente concurre al local a la tarde-noche para comer algo ligero, evitando tener que cocinar en sus casas. Por su parte, los dulces, junto con la torta (clase 18) y los cuadrados (clase 36), presentan un comportamiento similar. Esto puede deberse al hecho que los dulces y tortas se consumen sobre todo con la merienda y no tanto con el desayuno.

Tanto el resto de las comidas como las bebidas con alcohol (clase 37), se puede observar en los gráficos del anexo 5.1.2 que los mismos siguen el patrón general con mayor consumo en los tramos AM, y una fuerte baja el domingo PM. Esto es entendible ya que es más probable que los clientes concurren a la confitería a almorzar que a cenar. Esto se acentúa el domingo al mediodía, en los cuales la gente prefiere salir a comer que cocinar en sus casas.

El estudio de los tramos sirvió para brindar algunas conclusiones interesantes, pero no llega a ser completamente satisfactorio, ya que no revela grandes diferencias entre los distintos productos. Además, el análisis de la primera parte (antes del aumento de precios) no entregó resultados correctos. Por esta razón se decidió crear una nueva clasificación para realizar un nuevo estudio.



4.5.4. Análisis de la Franja Horaria

La categorización “Franja Horaria” no existía en la base de datos original. Fue creada para estudiar los patrones de consumo diarios, en base al horario en que se realizó la compra. Cada día fue dividido en distintas “franjas” de acuerdo a la hora, como muestra la tabla 4.5.4-1, a continuación. En el proceso se detectó que el campo hora no se encontraba en formato numérico, sino en formato de texto. Para ahorrar tiempo se decidió utilizar una macro de Visual Basic para realizar la conversión, la cual se encuentra en el anexo 3.5. En el anexo 3.6 se puede encontrar la macro utilizada para crear el campo “Franja” en la base de datos.

Franja Horaria	Desde	Hasta
Mañana	7:00	10:59
Media Mañana	11:00	12:59
Mediodía	13:00	15:59
Tarde	16:00	18:59
Noche	19:00	22:59
Trasnoche	23:00	02:00

Tabla 4.5.4-1. Descripción de las franjas horarias.

Si bien el horario habitual del local es de 8:00 a 22:00, existían algunos consumos anteriores a la hora de apertura y posteriores a la hora de cierre. Esto se debe a que el horario del local no es exacto, los empleados llegan antes de la hora de apertura y abren el local en caso de haber clientes esperando. Esos consumos anteriores a la fecha oficial de apertura se han incorporado a la franja “mañana”. Lo mismo ocurre con el horario de cierre, ya que no se puede “echar” a los clientes a las 22:00 si estos todavía no han finalizado de consumir. Por esa razón, la franja “noche” se extiende hasta las 23:00. Por otra parte, el local permanece abierto hasta más tarde los días del fin de semana, principalmente viernes y sábados. Además de las obvias razones, el local se encuentra cerca de un complejo de cines, cuyas funciones de “trasnoche” aumentan la cantidad de clientes los fines de semana. Debido a que probablemente el consumo de dicho horario sea diferente del de la noche, se decidió diferenciar a esta franja de horario como “Trasnoche”.

Como en el estudio anterior, el análisis fue realizado en dos partes: antes del aumento de precio y después. Para ambas partes se utilizaron parámetros idénticos. Se usaron dos variables de entrada, o “inputs”. La variable “Franja” es la variable categórica y la cantidad (es decir, cuántos ítems de cada producto son adquiridos) es la variable continua. La única salida, o “output”, es el código de rubro. Se ejecutó el programa repetidamente, hasta el momento en que dos corridas sucesivas no tuvieran mayores diferencias.



En ambos casos, la primera corrida no identificó la franja “trasnoche”, al igual que en el estudio anterior. Probablemente esto es debido a que su extensión es la menor y a que aparece en ultimo lugar. En las segundas corridas sí aparece esta franja y el gráfico tiene una forma más definida. Los terceros gráficos son muy parecidos a los segundos, con lo cual no se realizaron más corridas. A continuación se encuentra el detalle de los resultados y las conclusiones del estudio.

Resultados obtenidos del primer análisis

En la figura 4.5.4-1 se muestran los resultados obtenidos directamente en el programa NNClass. La misma muestra los consumos de los distintos productos por franja horaria.

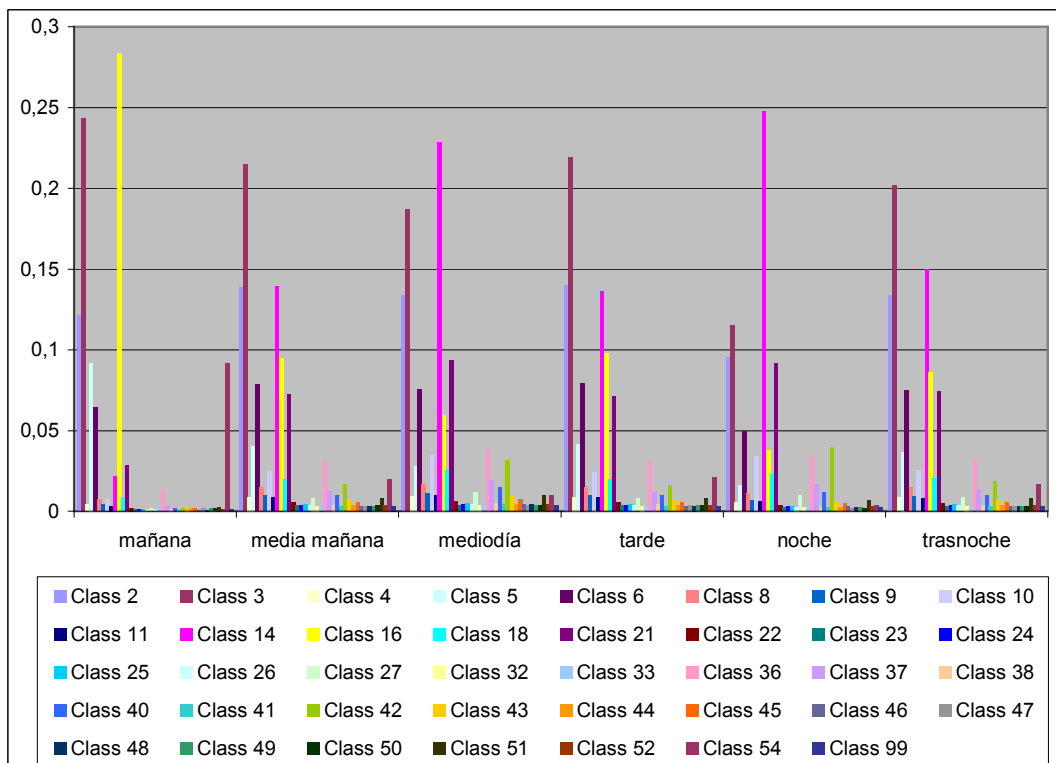


Figura 4.5.4-1. Resultados del primer análisis del NNClass para la franja horaria.

Existe una serie de productos que tienen un alto nivel de representatividad en el gráfico. Los mismos se encuentran detallados en la figura 4.5.4-2. Los casos más notables son los de los dulces (clase 16), que tienen un muy fuerte consumo en la mañana y un marcado descenso el resto del día, con excepción de la tarde y la trasnoche; las bebidas sin alcohol (clase 14), que tienen altos consumos durante el mediodía y la noche, y muy bajos durante la mañana; y las promociones “KALDI” (clase 54), con un pico de consumo durante la mañana. Otro producto de



alto peso en el estudio, los sándwiches (clase 21), tienen un patrón esperable, con consumos altos durante el mediodía y la noche, y bajo durante la mañana.

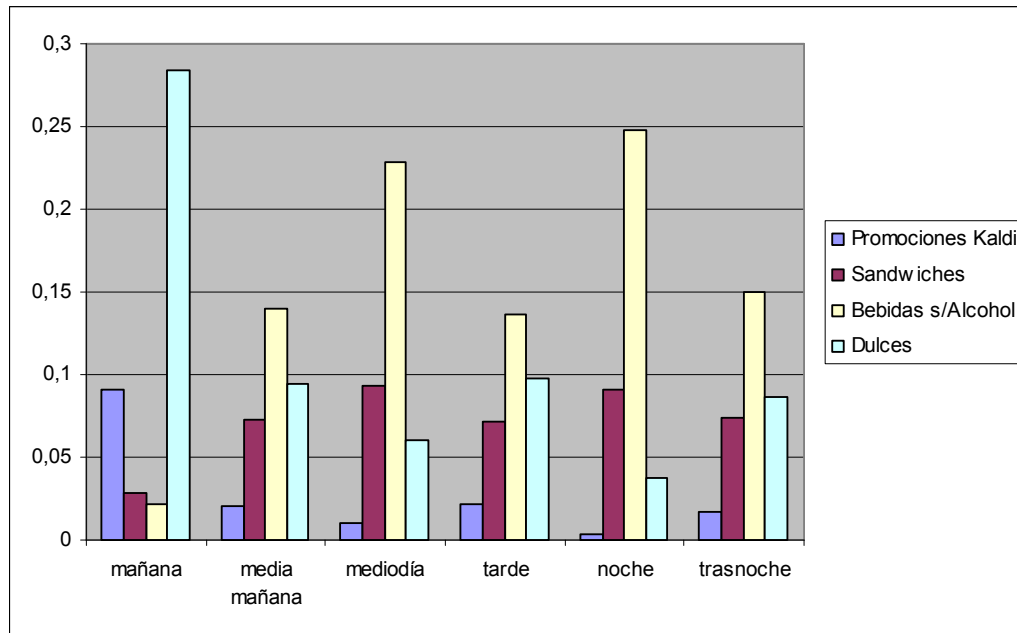


Figura 4.5.4-2. Detalle de los rubros principales por franja horaria.

El caso de los dulces (clase 16) es llamativo, sobre todo si se lo compara con los resultados del estudio de los tramo. En aquél, los dulces registraban un mayor consumo en los tramos de la tarde. Sin embargo, en este estudio existe un pico de consumo durante la mañana, y un descenso marcado en las demás franjas. Lo más curioso es que si se compara a los dulces con los cuadrados (clase 36) y las tortas (clase 18), éstos poseen un patrón inverso. De todas formas, cabe aclarar que no se dispone de información acerca de los distintos tipos de subproductos dentro del rubro “dulces”. Podría ocurrir que el rubro sea muy abarcativo, incluyendo productos que se consuman tanto a la mañana como a la tarde.

Lo mismo ocurre con las bebidas sin alcohol (clase 14). Los altos consumos durante el mediodía y la noche son esperables debido a que estos son los horarios de almuerzo y cena, respectivamente. Durante la media mañana, la tarde y la traspnoche su consumo es intermedio, y a la mañana su consumo es bajo, lo cual es atendible ya que a la mañana la gente generalmente desayuna con café o té. Por otra parte, es llamativo que las bebidas con alcohol (clase 37 - ver Anexo 5.2.1, gráfico D) tengan un patrón similar. Sería esperable que las mismas tuvieran altos consumos durante el mediodía, la noche y la traspnoche, y bajos en el resto de los horarios. Sin embargo, el consumo de bebidas con alcohol durante la media mañana y la tarde es comparable a los horarios antes mencionados. A primera vista, esta anomalía parece ser errónea. No obstante, según fuentes



consultadas en el local, el consumo de bebidas con alcohol de hecho tiene un patrón cuasi-constante (excepto a la mañana). Es interesante constatar que esta peculiaridad es real, debido mayormente a que la clientela principal de las franjas de media mañana y tarde son grupos mujeres de entre 30 y 50 años, de muy alto poder adquisitivo, que consumen regularmente bebidas espirituosas, tales como vino y principalmente champagne. Cabe destacar que los tragos con café siguen un patrón similar al de las bebidas con alcohol, aunque con valores inferiores.

Las promociones KALDI (clase 54), por su parte, aparentemente tendrían también un patrón erróneo, ya que no se explica que una comida “salada” posea un alto consumo durante la mañana. Sin embargo, si se realiza un estudio pormenorizado del consumo, no sólo se puede encontrar que este patrón es correcto sino que además se encuentra que el mayor consumo de esta especie de “almuerzo liviano” ocurre durante el horario de 10 a 11 de la mañana. Una posible explicación podría ser gente que desayuna tarde y desea consumir algo más abundante que un desayuno normal. También podría ser gente que debe almorzar temprano, y que luego consume nuevamente a media tarde. Lo más probable, sin embargo, es que haya gente que combine el desayuno y el almuerzo en una sola comida, lo que en otros países se denomina “brunch”.

Un párrafo aparte merecen las infusiones, que son la base del negocio de la confitería. La figura 4.5.4-3 muestra más detalladamente los resultados de las infusiones. Los dos cafés tradicionales y de mayor consumo, el pocillo (clase 2) y el jarrito (clase 3), tienen un comportamiento similar, con un consumo prácticamente constante durante el día. Ambos poseen un leve descenso durante la noche (el del jarrito más pronunciado) y además tienen una única diferencia a la mañana: mientras que el pocillo tiene un consumo levemente inferior a su promedio durante las horas de la mañana el jarrito tiene un consumo mayor al suyo. Esto se puede explicar teniendo en cuenta que por su pequeño tamaño, el pocillo es un café que se suele beber después de comer, mientras que el jarrito es más grande, y por lo tanto más adecuado para el desayuno. Sin embargo, esto no explica el porqué del descenso del consumo durante la noche, ya que en esta franja podría ocurrir un pico de consumo debido a que la gente concurre al local para cenar. Una razón para esto podría ser que durante la franja “noche” se percibe un aumento en el consumo de tortas (clase 18 – ver Anexo 5.2.1, gráfico C) y cuadrados (clase 36). Estos postres podrían desplazar al café como el último consumo de la noche. Otros tipos de postres, bebidas o productos no están identificados podrían desplazar al café. Dichos productos podrían encontrarse agrupados en el rubro “Varios” (clase 42 – ver anexo 5.2.1, gráfico C), el cual sufre un incremento importante durante la franja noche (y en la franja mediodía también, lo cual refuerza la teoría). Otra razón podría ser el hecho de que mucha



gente prefiere no consumir café durante las horas de la noche para que el efecto estimulante de la cafeína no interfiera con su sueño.

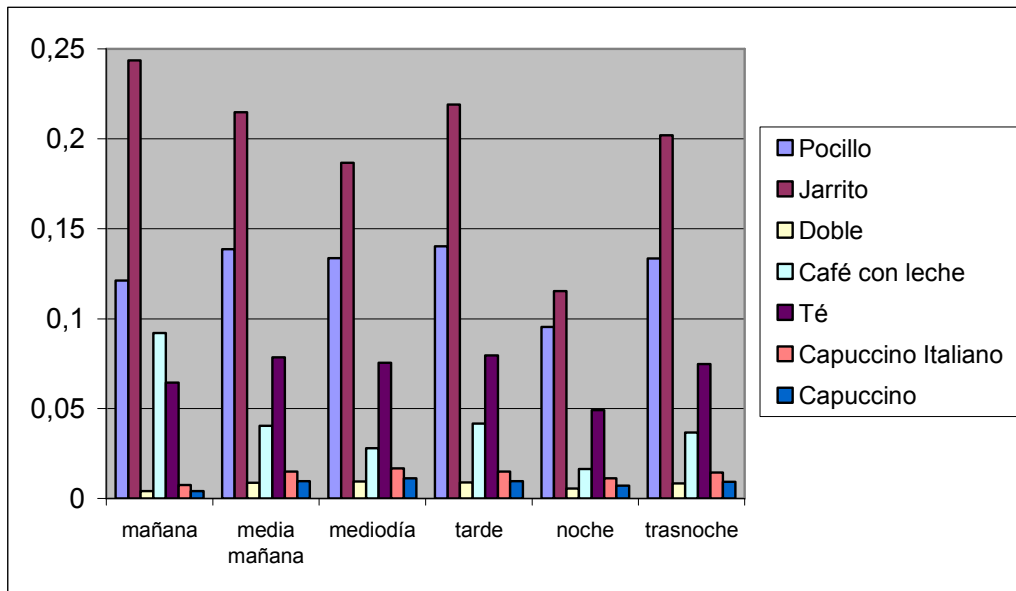


Figura 4.5.4-3. Detalle de las infusiones por franja horaria para el primer análisis.

El té (clase 6), y en menor medida el café doble (clase 4), tienen un comportamiento muy parecido al del pocillo, mientras que el café con leche (clase 5) es similar al jarrito, aunque con un pico mucho más pronunciado a la mañana. Esto se debe a que el café con leche es para el consumidor argentino una infusión típica del desayuno. Los dos capuccinos (clases 8 y 9) tienen un comportamiento parecido al del pocillo, aunque con un descenso menor a la noche, lo cual se puede explicar teniendo en cuenta que, como los postres, su consumo es común durante la sobremesa de la cena.

Los restantes productos tienen poca incidencia en el consumo. Los gráficos de los mismos se encuentran en los gráficos E y F del anexo 5.2.1. En general, la mayoría de los productos, incluso aquellos de distintos rubros, responden al siguiente patrón: bajo consumo a la mañana, muy alto al mediodía, intermedio a media mañana, tarde y traspnoche, y un poco más bajo a la noche. Esto se puede deber a que la gente considera a la confitería un lugar más adecuado para almorzar que para cenar. La gente cena habitualmente en restaurantes más caros, pero suele almorzar en confiterías, más baratas y con una atención más rápida. Esto se acentúa debido a la ubicación de la confitería, ya que los habitantes del barrio de Belgrano (y especialmente en la zona del Paseo del Ángel) pertenecen sobre todo a una clase socioeconómica media-alta a alta.



Dentro de las comidas, todos los productos responden al patrón antes citado, con la excepción de la carne (clase 40) y las ensaladas (clase 41), tal vez las comidas más habituales en la cena, que aumentan durante la noche. Por otra parte las verduras sufren un descenso mayor al promedio, lo cual es difícil de explicar sin mayores datos. En cuanto a los dos tipos de sándwiches (clases 21 y 46), el gráfico G del anexo 5.2.1 muestra que no son el mismo tipo de producto. El sándwich 21 tiene un incremento de consumo en la franja noche, mientras que el otro posee un descenso. Además, este último tiene un menor consumo en la mañana. Según las fuentes consultadas, hay una gran variedad de sándwiches en el local, agrupando la clase 21 a una gran variedad de ellos con un rango de precios medio, mientras que la clase 46 son sándwiches más caros, y más comúnmente consumidos en el mediodía y la tarde. Por estas razones no se agruparon ambos tipos de sándwiches en un mismo rubro. Lo mismo ocurre con los dos rubros de ensaladas (27 y 41), gráfico H del mismo anexo. El primer caso corresponde las variedades más económicas, mientras que el segundo rubro son aquellas más caras y se consumen en menor cantidad. Tanto en el caso de los sándwiches como el de las ensaladas, aquellos con son más caros son de mayor tamaño y constituyen un plato por sí mismos. Esa es la razón por la cual su consumo baja durante la noche, ya que en la cena los consumidores optan por combinaciones de platos más elaborados.

Todos los demás productos, en mayor o menor medida, responden al patrón explicado en párrafos anteriores. Esto no deja de ser llamativo en algunos casos particulares, que se muestran en el gráfico F del anexo 5.2.1. Todos los productos de este gráfico siguen un patrón similar al indicado anteriormente, con poco consumo a la mañana, muy alto al mediodía y bajo a la noche. Resulta llamativo que tanto los desayunos como las meriendas sigan este patrón, ya que sería lógico, por ejemplo, que los desayunos tuvieran un alto consumo a la mañana y las meriendas durante la tarde. Las razones para esto son múltiples. En primer lugar, estos rubros son productos que poseen poca cantidad de datos dentro de la base de datos. Podría ocurrir también que debido a los pocos datos existentes el programa no pueda extraer un patrón representativo de cada producto, y como resultado muestre el patrón promedio general para estos productos. De todas formas, analizando la base de datos se pueden encontrar ciertas anomalías dentro de los datos, como que los desayunos son consumidos durante la noche y que otros productos que aparecen en el gráfico a la mañana no tienen de hecho consumos durante esa franja. Sin duda, los rubros con pocos datos no solo no son representativos, sino que además son más permeables a errores humanos. Más aún, estos rubros no pueden ser analizados correctamente por el programa, por lo cual son separados del análisis del presente estudio.



Resultados obtenidos del segundo análisis

A continuación se encuentra el gráfico con los resultados del programa para la segunda parte de la base de datos, es decir, después del aumento de precios. Los patrones son, en general, parecidos aunque se detectan algunas diferencias en ciertos rubros representativos.

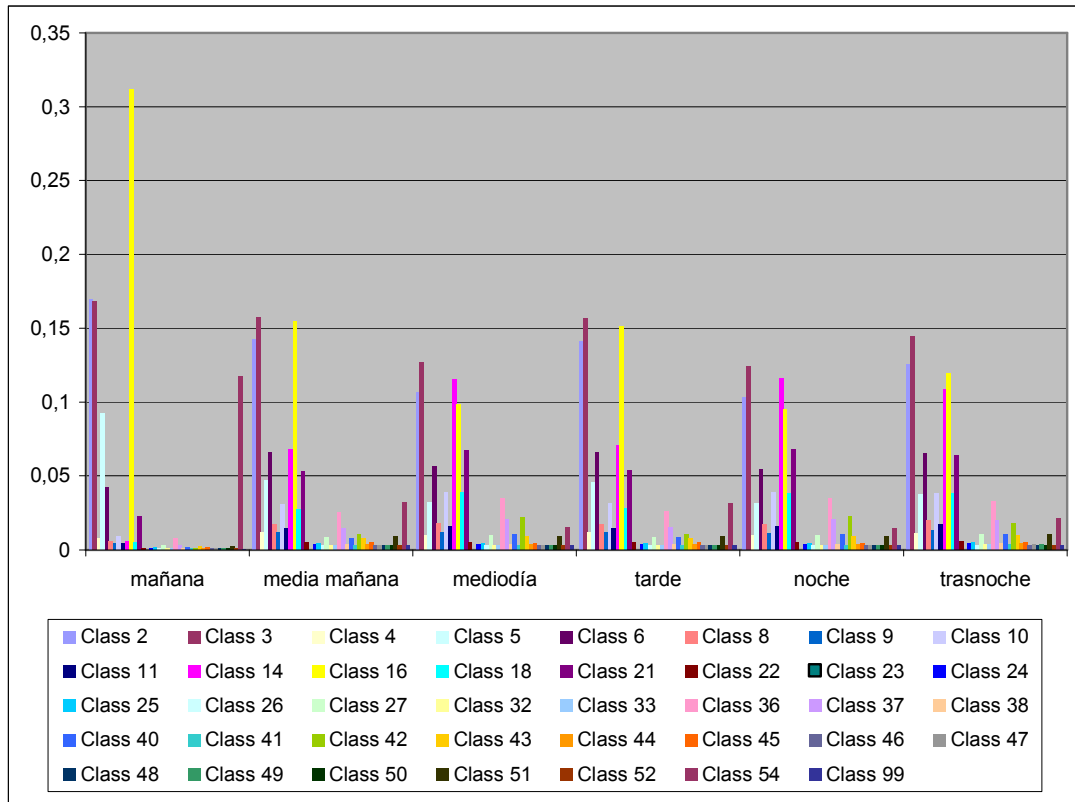


Figura 4.5.4-4. Resultados del segundo análisis del NNClass para la franja horaria.

Los gráficos detallados para los distintos subgrupos se encuentran en el anexo 5.2.2. Los rubros principales, es decir los dulces (clase 16), las bebidas sin alcohol (clase 37), las promociones KALDI (clase 54) y los sándwiches (clase 21), no sufren cambios importantes.

Entre las infusiones, por otra parte, sí existen cambios en las clases más representativas. Como se puede apreciar en la figura 4.5.4-5 a continuación, el pocillo (clase 2) gana terreno frente a otros productos, y su consumo durante la mañana se convierte en el más alto. De hecho, si se compara con los gráficos de la primera parte del estudio, se puede apreciar que este incremento ocurre a costas de una fuerte disminución en el consumo del jarrito (clase 3) y más leve en el caso del té (clase 6). Además, el pocillo aumenta su consumo durante el mediodía y la noche. Esto estaría indicando que a raíz de los aumentos de



precios, los clientes optan por una opción más económica para paliar el aumento de costo.

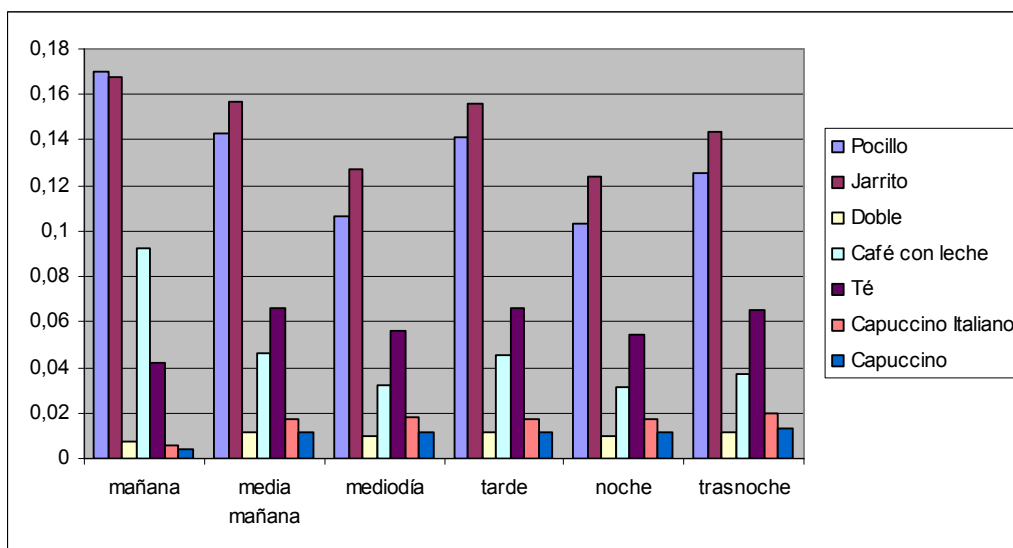


Figura 4.5.4-5. Detalle de las infusiones por franja horaria para el segundo análisis.

Las comidas (Anexo 5.2.2, gráfico E) no muestran grandes variaciones, aunque se percibe un aumento del consumo de las mismas durante los tramos de trasnoche. Es poco probable que este cambio se deba al cambio de precios. En este caso podría influir la época del año, con la llegada de las vacaciones de invierno puede ocurrir que alguna gente cene más tarde. Además, esta período viene acompañado por una gran cantidad de estrenos de películas en el cine que podrían elevar las cenas tardías y, como se ha explicado anteriormente, el consumo del local está relacionado con el complejo de cines cercano.

Las bebidas alcohólicas (Anexo 5.2.2, gráfico D), por su parte, también registran variaciones respecto a la primera parte que merecen ser analizadas. Las bebidas con alcohol en general (clase 37), y sobre todo los tragos con café (clase 11), presentan incrementos importantes en su consumo en prácticamente todas las franjas. El consumo de bebidas con alcohol presenta incrementos en las franjas media mañana, tarde, noche y trasnoche. Los tragos con café tienen también incrementos durante el mediodía. En ambos casos, el crecimiento más importante se da en la franja trasnoche. Resulta difícil relacionar el aumento de precio con el aumento de consumo: el aumento de precio debería disminuir el consumo, y es poco probable que el alza de precios aumente el alcoholismo. Tal vez se pueda explicar esto en relación a otros productos, por la aparición de un efecto de sustitución. Sin embargo es difícil afirmar esto sin mayores datos.



Los demás rubros, que son poco representativos como se ha explicado anteriormente, muestran un cambio general en el patrón de consumo. El mismo se muestra más regular entre las franjas posteriores a la mañana. Las pequeñas degustaciones (clase 22) y las tablitas (clase 24), sobre todo, muestran un cambio mayor en la franja trasnoche (Anexo 5.2.2, gráfico F). Este resultado es más coherente que el de la primera parte. Sin embargo, las anomalías del desayuno y la merienda siguen existiendo, lo cual demuestra una vez más que los productos con datos limitados son poco fiables para el análisis.

4.5.5. Conclusiones del uso de la herramienta

El uso de la herramienta NNClass permite obtener conclusiones útiles acerca de los patrones de consumo de los clientes que concurren a la confitería. Si bien no es el objetivo concreto del presente trabajo, existen muchos casos en que los resultados de estos estudios pueden servir para proponer mejoras.

Un ejemplo de esto es el caso del jarrito (clase 3). El estudio de franjas muestra que el consumo de este producto cae en los horarios de mediodía y noche. Para incentivar el consumo de los jarritos en estos horarios se podría, por ejemplo, asociarlos a los almuerzos y cenas creando un menú que incluyera plato principal y café. Otra opción aún mejor podría asociar al jarrito con los cuadrados y tortas, los cuales tienen altos consumos en dichos horarios, creando una opción alternativa al postre para después de las comidas. Lo mismo podría hacerse con los dulces, que también tienen un consumo bajo al mediodía y a la noche. Podría asociárselos como acompañamiento al jarrito o al pocillo, que es un típico producto de sobremesa, produciendo un efecto de “venta atada”.

Esta asociación de productos puede aplicarse también al caso de los capuccinos (clases 8 y 9), principalmente en el horario de la cena. Cabe señalar que este tipo de café tiene una alta rentabilidad, lo cual motiva aún más aumentar su consumo. Otro producto de alta rentabilidad, el té (clase 6), podría también apalancarse en los horarios del mediodía y la cena, publicitándolo como una opción sana para después de las comidas. Podría incluso introducirse el té digestivo como una alternativa al café. Por supuesto que en este caso debe cuidarse de que no ocurra un efecto de “canibalización” del consumo, es decir, el aumento del consumo del producto a costa de la reducción de otro. Los esfuerzos de publicitar un nuevo té deben estar dirigidos a aquel cliente que habitualmente no consume ni té ni café. De esta manera, el aumento de consumo será realmente beneficioso para el local.

Otro caso concreto es el de las promociones KALDI (clase 54). Dichas promociones poseen un alto consumo en los horarios matutinos (sobre todo a la mañana), independientemente del día, aunque menor en los viernes y sábados.



Como se ha explicado antes, el mayor consumo matutino podría deberse a que esta promoción consiste en una comida ligera, lo cual la hace poco apta para la cena. Una opción para aumentar el consumo de esta promoción es el de agrandarlo para que sea atractivo como un almuerzo completo o incluso como cena. Sin embargo, esto podría afectar el consumo de la promoción original, apreciada por quienes desean un almuerzo ligero. Por lo tanto sería más adecuado crear un menú completamente nuevo. Este nuevo menú “light” podría contener comidas “light” pero con platos más abundantes, como por ejemplo milanesas de soja, y un postre “light”, como podría ser una fruta o gelatina.

4.6. Propuestas de mejora generales

Si bien no es el objetivo de este trabajo, la presente sección intenta demostrar que las redes neuronales pueden ayudar a obtener conclusiones valiosas del consumo del local sin requerir de conocimientos extensos de estadística o marketing. Con los resultados de las herramientas de redes neuronales utilizadas y algunos datos obtenidos de reuniones con dueños y empleados del local, se proponen a continuación algunas mejoras concretas que podrían aumentar la clientela y facturación.

Además de las conclusiones surgidas del análisis individual de cada herramienta, la combinación de ambas puede generar propuestas adicionales. Como se ha expresado anteriormente, es fundamental conocer a la clientela para poder modificar el menú adecuadamente. Algunos casos particulares merecen ser analizados con mayor detalle:

Mujeres de 40-50 años

Un caso particular es el de las mujeres de mediana edad y alto nivel económico que se juntan en el local y consumen picadas y bebidas alcohólicas de alto valor. El estudio de segmentación ubicaría a este grupo en un alguno de los clusters con alto valor total por factura. El estudio de clasificación y las fuentes del local las ubican en los franjas horarias de la media mañana y la tarde mediodía.

Este grupo tiene gran potencial para ser explotado. Se podrían armar distintas tablitas y picadas de las ya existentes, e ir variando el menú para incentivar que concurran al local más seguido. Se podrían introducir las tapas y realizar degustaciones de vinos. Esto no solo sirve para el grupo de mujeres mencionado. Se podría crear un nuevo nicho para la gente que aprecia el vino y las picadas. Todos estos productos son de alta rentabilidad y sin duda crearán alto valor para el local.



Niños

Actualmente el local está dirigido especialmente a un público adulto. No existen productos específicos para los niños más allá de los usuales (tostados, submarinos). Sin dejar de atender al target principal, ya que probablemente mucha gente valore el posicionamiento actual del local, se podrían introducir variantes para satisfacer las necesidades de los jóvenes que acompañan a sus padres.

Una opción viable sería introducir menús especiales para el almuerzo y la cena. Es común que los niños coman poca cantidad, por lo cual estas comidas podrían ser comidas simples de tamaño reducido con el acompañamiento de bebidas gaseosas, las cuales tienen un muy alto rendimiento. Pizzetas, milanesas, rebozados de muzzarella o hamburguesas podrían ser opciones posibles. Otra opción sería crear menús con comidas más sanas, como contraposición a los locales de comida rápida, muy populares entre los niños, que venden comidas con una alta cantidad de calorías y grasas. Como valor agregado, se podría incluir algún tipo de regalo, un dulce o un juguete por ejemplo, en los menús para niños, de tal forma de atraerlos al local.

En definitiva, realizando estas modificaciones se podría posicionar efectivamente al coffee store como una alternativa viable para el almuerzo o cena familiar.

Oficinistas

Se ha visto que en los alrededores existe un gran número de oficinas y locales comerciales. Por otra parte, el menú light que se ofrece es de tamaño reducido y no llega a constituir un almuerzo por sí mismo. El estudio de segmentación ha demostrado que las personas que almuerzan en el local representan entre un 8 y un 20% de la facturación total del local. Por esta razón es importante ampliar la oferta para la hora del almuerzo.

Debido a que el horario de almuerzo no suele ser extenso, estos menús deberían ser rápidos en su entrega y su consumo. Para hacer esto sin reducir la oferta a sándwiches, debería establecerse uno o dos platos específicos para el día, y que los mismos estén pre-elaborados. Este “menú del día” debería variar todas las jornadas para así atraer a la gente diariamente a almorzar.

Por otra parte, podría introducirse un sistema de “abono mensual”. Una persona que almuerza todos los días en un restaurant o bar gasta gran parte de su sueldo en comer. Mucha gente desea controlar el gasto de antemano para no excederse. Este abono mensual tendría ese beneficio para el cliente y además podría



ofrecerse a un precio menor que todas las comidas por separado. De esta manera, si bien el Coffee Store estaría resignando parte de sus ganancias al bajar el precio, estaría asegurando que el cliente asista al local por un mes entero. Y en el largo plazo, estaría fidelizando al cliente al local, lo cual es aún más importante que las ganancias de corto plazo.

Noche del domingo

Además de los citados menús para niños, los almuerzos y cenas ofrecen otras posibilidades de mejora. El estudio de segmentación muestra que las comidas sufren un importante descenso en su consumo durante el tramo de la noche del domingo. Conociendo a la gente de la zona se puede encontrar aquí una interesante oportunidad para explotar.

Una gran cantidad de personas de nivel socioeconómico medio y alto contratan personal doméstico durante la semana, el cual, entre otras cosas, cocina el almuerzo y cena. Durante el fin de semana el personal suele retirarse, con lo cual los residentes deben cocinar ellos mismos, comer las sobras de la semana, pedir comida por delivery o ir a cenar afuera. Aún para la gente que no tiene gente que le cocine, el domingo suele ser un día de descanso en el que no se desea cocinar. Aquí se presenta la oportunidad de ofrecer un producto único y dirigido específicamente a satisfacer una necesidad concreta.

Con un lema publicitario como “No cocine, venga al Coffee” se podría incentivar a la gente a concurrir al local para cenar. La oferta debería incluir comida similar a la de una cena casera, y con precios razonables de tal forma de competir con el delivery de comidas. No sólo podría aplicarse al domingo a la noche, sino también a los días de semana donde existe también un descenso notable en el consumo de comidas. Sin embargo, el tramo del domingo a la noche posee el descenso más marcado de todos los días y además posee las características enumeradas anteriormente.

Cine

El cine atrae un número importante de espectadores sobre todo en los horarios de la tarde y la noche, especialmente los viernes y sábados. Mucha gente egresa del mismo a la hora de la cena o desea completar el programa tomando un café. Se podrían realizar promociones conjuntas con el complejo de cines cercano para aumentar las ventas de ambos locales. Por ejemplo, podría ser que al presentar la entrada de cine de la fecha se obtenga un descuento en el consumo del Coffee Store, o que consumiendo en el Coffee Store se otorgue un voucher para descuento en entradas de cine o en pochoclo.



Se podría incluso diseñar un sistema de recompensas en el cual la gente posea una tarjeta con la cual gane puntos con cada consumo realizado. Si bien estos sistemas ya existen en una gran cantidad de comercios, podría introducirse en combinación con el cine y con otros locales del paseo de compras del Paseo del Ángel o de la avenida Cabildo, creando un sistema exclusivo para la gente del barrio de Belgrano.



5. CONCLUSIONES

Los objetivos puntuales de análisis de la base de datos de la confitería han alcanzado, e incluso superado, las expectativas. El uso de solamente dos herramientas de redes neuronales artificiales permitió extraer un amplio número de conclusiones de los datos. Asimismo, se pudieron realizar numerosas propuestas de mejora de la oferta del local. Más allá del éxito de estas propuestas, lo cual excede el objetivo de la tesis, las mismas estuvieron totalmente fundamentadas por el análisis.

El presente trabajo ha demostrado la simpleza del uso y la utilidad de las herramientas de redes neuronales, así como su disponibilidad para cualquier persona con acceso a Internet. Se han desmitificado muchas de las creencias acerca de las redes neuronales, como por ejemplo que las redes neuronales son lentas. Todas las corridas de ambas herramientas no tomaron más que un par de horas en total. En cuanto al concepto de que las redes neuronales son difíciles de utilizar, se puede apreciar que en el análisis de los resultados no se usaron más que nociones básicas de estadística.

Por otra parte, es cierto que las herramientas utilizadas son bastante básicas y que requirieron de cierto procesamiento previo de los datos. Sin embargo, a los efectos de los objetivos concretos del trabajo, los programas han demostrado ser de gran utilidad. Para análisis más detallados existe una importante variedad de software, tanto pago como gratuito, que pueden ser obtenidos a través de Internet.

Más aún, el trabajo ha demostrado el potencial de las redes neuronales para su aplicación a los negocios, particularmente para negocios pequeños o medianos. Se ha visto que no se requieren equipos poderosos de computación ni grupos numerosos de analistas para utilizar estas herramientas. Cualquier empresa grande, pyme o empresario puede valerse de esta tecnología para optimizar su oferta y cumplir con sus objetivos satisfaciendo las necesidades de los clientes.

La aplicación exitosa de las redes neuronales al ámbito empresarial en Argentina está todavía lejos de ser la ideal. De todas formas, este trabajo cumple con otro de sus objetivos al aportar conocimiento práctico sobre el tema. Es deseable que otros trabajos continúen echando luz sobre este asunto, generando así un nuevo paradigma en los modelos de análisis de datos.



5.1. Futuras líneas de investigación

Este trabajo es una aplicación particular de las redes neuronales a las empresas, centrándose en una base de datos perteneciente a una confitería. Futuros trabajos relacionados con el tema deberían abarcar otros rubros y negocios distintos. Estos pueden incluir comercios no relacionados con la gastronomía, empresas de servicio, industria manufacturera, empresas de energía, compañías financieras, para citar algunas.

Otros trabajos podrían utilizar herramientas más sofisticadas que no posean las limitaciones ya mencionadas. Existe un gran número de herramientas relacionadas con las redes neuronales, no solo de clasificación y clustering, que podrían ser aplicadas a los negocios. Sería interesante también comparar a las redes neuronales con otros métodos de análisis de datos, contrastando la facilidad de uso y la performance de las mismas.



6. REFERENCIAS

[Acosta, M; *et al*; 2000]: Acosta, M.I., Balazar, H. y Zuluaga, C., “Tutorial de Redes Neuronales”, Universidad Tecnológica de Pereira, Colombia, 2000. Disponible en <http://ohm.utp.edu.co/neuronales/Capitulo2/Backpropagation/AntecedentesB.htm>, página web vigente al 31/05/2006.

[Anderson D. y McNeil, G.; 1992]: Anderson, D., y McNeil, G., “Artificial Neural Networks Technology”, Data & Analysis Center for Software, 1992. Disponible en internet en [http://www.dacs.dtic.mil/techs/neural/neural ToC.html](http://www.dacs.dtic.mil/techs/neural/neural%20ToC.html), página web vigente al 02/06/2006

[Austin, J. ; 2006]: Austin, J., “Research Papers”, enlaces a publicaciones del autor sobre Redes Nuronales, 1986-2006. Sitio del departamento de Ciencias de la Computación de la “University of York” del Reino Unido. Página web vigente al 15/06/2006

[Britos, P, 2005] Britos, P., “Objetivos de Negocio y Procesos de Minería de Datos Basados en Sistemas Inteligentes”, “Reportes Técnicos en Ingeniería del Software” núm. 7, págs. 26-29, 2005.

[Bullinaria, J.; 2004]: Bullinaria, J., “Introduction to Neural Networks”, 2nd Year UG, MSc in Computer Science, 2004. Disponible en internet en <http://www.cs.bham.ac.uk/~jxb/inn.html>, página web vigente al 26/05/2006.

[BrainMaker, 2006]: “BrainMaker Neural Network Application Examples”, enlaces a trabajos de varios autores relacionados con Redes Neuronales. Sitio de la empresa Brainmaker, dedicada a aplicaciones de Redes Neuronales. Página web vigente al 15/06/2006

[Browse, R.A., 2006]: Browse, R.A., “Neural Networks Papers”, enlaces a publicaciones del autor sobre Redes Nuronales, 1994-1999. Sitio de la Facultad de Computación de la “Queens University” de Canadá. Página web vigente al 15/06/2006

[Clarín, 1999]: “El subte ya llega al corazón del barrio de Belgrano”, artículo del diario Clarín (online) de la fecha 22/06/1999, disponible en internet en <http://www.clarin.com/diario/1999/06/22/e-03701d.htm>, página web vigente al 04/06/2006.



[Davis, B; 2005]: Davis, B.E., "The Pattern of Knowledge, Pattern Recognition", White Paper, Kaiteur Institute For Knowledge Management, 2005. Disponible en http://www.kikm.org/pattern_recog.htm, página web vigente al 19/11/2005

[Fiszelew, A y García-Martínez, R; 2001]: Fiszelew A., García-Martínez R., "Generación automática de redes neuronales con ajuste de parámetros basado en algoritmos genéticos"., 2001. Revista Instituto Tecnológico Buenos Aires; Número 26. Páginas 76-101. EMEDE S.A.

[Folklore, 2005]: "Belgrano", artículo de "Folklore Tradiciones", disponible en internet en <http://www.folkloretradiciones.com.ar/barrios/belgrano.htm>, página web vigente al 04/06/2006.

[García Vega, A; 2002]: García Vega, A., "La inteligencia artificial en los negocios", Universidad de Veracruz, 2002.

[GCBA, 2005]: Dirección General de Estadística y Censos, Gobierno de la ciudad de Buenos Aires, "Informe de Resultados N°163: Estimación de la población de la Ciudad Autónoma de Buenos Aires al 1 de julio del 2001", 2005. Disponible a través del portal del Gobierno de la Ciudad de Buenos Aires, en http://www.buenosaires.gov.ar/areas/hacienda/sis_estadistico/nuevoinforme/N163.pdf, página web vigente al 04/06/2006.

[Li, E; 1994]: Li, E., "Artificial Neural Networks and Their Business Applications", "Information & Management" núm. 27, págs. 303-313, 1994

[Makhfi, P.; 2006]: Makhfi, P., "Introduction to Artificial Neural Networks", 2001-2006, disponible en internet en <http://www.makhfi.com/tutorial/introduction.htm>, página web vigente al 02/06/2006

[Mibelgrano, 2006]: "Una esquina emblemática", artículo de "Mi Belgrano", portal del barrio de Belgrano, disponible en internet en <http://www.mibelgrano.com.ar/cabildo.htm>, página web vigente al 04/06/2006.

[Neurocolt, 2002]: "NeuroCOLT Technical Reports 2002", enlaces a trabajos de varios autores relacionados con Redes Neuronales. Sitio de NeuroCOLT (Neural Networks and Computational Learning Theory). Página web vigente al 15/06/2006

[Nueva Esparta, 2006]: Universidad Nueva Esparta, "Aprendizaje por descubrimiento en la enseñanza a distancia: Conceptos y un caso de estudio", 2004. Disponible en internet en



<http://www.alfa.une.edu.ve/biblio/BiblioGenreral/E/E/ense%F1anza.asp>, página web vigente al 31/05/2006.

[OzGrid, 2006]: Página de la empresa OzGrid, dedicada a aplicaciones para planillas de Excel: <http://www.ozgrid.com/Services/neural-network-technology.htm>, página web vigente al 26/01/2006.

[Puppio, R; 2002]: Puppio, R.P., “Redes Neurales aplicadas al Avalúo Inmobiliario”, “Sociedad de Ingeniería de Tasación de Venezuela (SOITAVE)”, 2002.

[Raghupathi, W y Shekhar, S; 1996]: Raghupathi, W. y Shekhar, S., « Neural Networks in Business: An Introduction to the Minitrack”, “Proceedings of the 29th Annual Hawaii International Conference on System Sciences – 1996”.

[Registro Civil, 2004]: “Encuesta Nacional de Hogares 2004”, disponible a través del portal del Registro Civil del Gobierno de la Ciudad de Buenos Aires en http://www.registrocivil.gov.ar/areas/hacienda/sis_estadistico/tabulados_basicos_EAH_2004_corregidos.pdf, página web vigente al 04/06/2006.

[Ripio, 2004]: “Barrio de Belgrano”, artículo de “RipioTurismo”, portal de internet de turismo, disponible en <http://www.ripioturismo.com.ar/buebelgrano.htm>, página web vigente al 04/06/2006.

[Salas, 2006]: Salas, M., “Mapas Autoorganizativas de Kohonen (SOM)”, Departamento de Computación, Universidad de Valparaíso. Disponible en http://www.inf.utfsm.cl/~rsalas/Pagina_Investigacion/docs/Apuntes/Redes%20SO M.pdf, página web vigente al 28/5/2006.

[Schraudolph, N; y Cummins; F; 1999]: Schraudolph, N. y Cummins, F., “Introduction to Neural Networks”, Istituto Dalle Molle di Studi sull'Intelligenza Artificiale, Lugano, CH, disponible en <http://www.idsia.ch/NNcourse/brain.html>, página web vigente al 28/05/2006.

[Shachmurove, Y; 2002]: Shachmurove, Y., “Applying Artificial Neural Networks to Business, Economics and Finance”, Departments of Economics, The City College of the City University of New York and The University of Pennsylvania, 2002.

[Shachmurove Y; y Witkowska; D; 2000]: Shachmurove, Y. y Witkowska, D., “Utilizing Artificial Neural Network Model to Predict Stock Markets”, 2000.



[Shapiro, 2004]: Shapiro, J., "Papers on Neural Networks", enlaces a publicaciones del autor sobre Redes Nuronales, 1988-2002. Sitio de la Facultad de Ciencias de la Computación de la "Manchester University", del Reino Unido. Página web vigente al 15/06/2006

[Statsoft, 2003]: Statsoft Electronic Textbook, 2003. Disponible en internet en <http://www.statsoft.com/textbook/stneunet.html>, página web vigente al 26/05/2006.

[Stergiou C. y Siganos, D; 1996]: Stergiou, C. y Siganos, D., "Neural Networks". Department of Computing, London Imperial College. Disponible en internet en http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html, página web vigente al 26/01/2006

[Vellido A; *et al*; 1999]: Vellido, A., Lisboa, P.J.G. y Vaughan, J., "Neural networks in business: a survey of applications (1992-1998)", "Expert Systems with Applications" núm.17, págs. 51–70, 1999.

[Zhang, G.; *et al*; 1998]: Zhang, G., Patuwo, B.E. y Hu, M.Y., "Forecasting with artificial neural networks: The state of the art", "International Journal of Forecasting" núm. 14, págs. 35–62, 1998.



7. ANEXOS

Anexo 1 – Datos del Coffee Store

Anexo 1.1 – Ubicación de los locales

<i>Nombre del local</i>	<i>Zona/Barrio/Ciudad</i>	<i>Dirección</i>
ARGENTINA		
Abasto Shopping	Abasto	Av. Corrientes y Agüero
Acassuso	Acassuso	Av. Del Libertador 14849/55
Alto Palermo Shopping	Palermo	Santa Fe y Coronel Díaz
Belgrano R	Belgrano	Juramento 3000
Buenos Aires Design	Recoleta	Pueyrredón 2501
Casa del Angel	Belgrano	Sucre 2145
Hoyts Cinema (Unicenter)	Martínez	Paraná 3745 Martínez
La Cuadra	Belgrano	Jorge Newbery 1651
La Plata	La Plata	48 y Diag. 80
Microcentro	Microcentro	Florida y Lavalle
Centro Cultural Borges	Microcentro	San Martín esquina Viamonte
Galerías Pacífico	Microcentro	San Martín 768. Local 133
Olivos	Olivos	Av. Maipú 2209
Palermo	Palermo	Salguero 2294
Patio Bullrich	Recoleta	Av. Del Libertador y Callao
Puerto Madero	Puerto Madero	Alicia M. de Justo 292
Puerto Madero Este	Puerto Madero	Alicia M. de Justo 1788
Pilar	Pilar	Shopping Torre del Sol
Recoleta	Recoleta	Las Heras 1822
Recoleta	Recoleta	Pueyrredón 2180
Tren de la Costa	San Isidro	Lasalle 653
Tribunales	Tribunales	Viamonte 1393
Unicenter Shopping	Martínez	Paraná 3745 Martínez
Del Parque Shopping Center	Villa del Parque	Nazarre 3135, Local 103
Rosario	Rosario	Alto Rosario Shopping
URUGUAY		
Punta del Este	Punta del Este	A. Roosevelt y Parada 6
ESPAÑA		
Centro Comercial Xanadú Arroyomolinos	Madrid	
Centro Comercial La Vaguada	Madrid	
Centro Comercial La Moraleja Green Norte	Madrid	
Centro Comercial la Moraleja Green Sur	Madrid	

Fuente: The Coffee Store

<http://www.thecoffeestore.com.ar/sr800/home.asp?secc=locales>



Anexo 1.2 – Sistema de Franquicias

Argentina

Concepto	The Coffee Store	The Coffee Store Express
Derecho por el uso de la Marca	\$ 25.000	\$ 25.000
Fee Mensual	Del 5 % sobre las ventas netas, según escala	
Capital Requerido	\$195.000 / \$225.000	\$160.000 / \$180.000
Recupero de la inversión	24 a 28 meses	
Facturación mensual Bruta (esperada)	\$ 65.000	\$ 39.000
Utilidad Bruta mensual (estimada)	\$ 9.000 / \$ 11.000	\$ 5.500 / \$ 7.000
T.I.R. anual (esperada)	32% a 38%	32% a 38%

Valores estimados + IVA sujetos a variación, en Pesos Argentinos.

España

Concepto	The Coffee Store Ad hoc
Derecho por Entrada	€ 12.000
Canon de Mantenimiento	5% sobre las ventas netas
Capital Requerido	€ 170.000 / € 200.000 (Incluye el Derecho de Entrada, Obra civil, equipamiento y mercancía)
Recupero de la inversión	24 a 30 meses
Facturación mensual Bruta (esperada)	€ 32.000 / € 42.000
Utilidad Bruta mensual (estimada)	€ 7.500 / € 9.500
T.I.R. anual (esperada)	28% a 33%

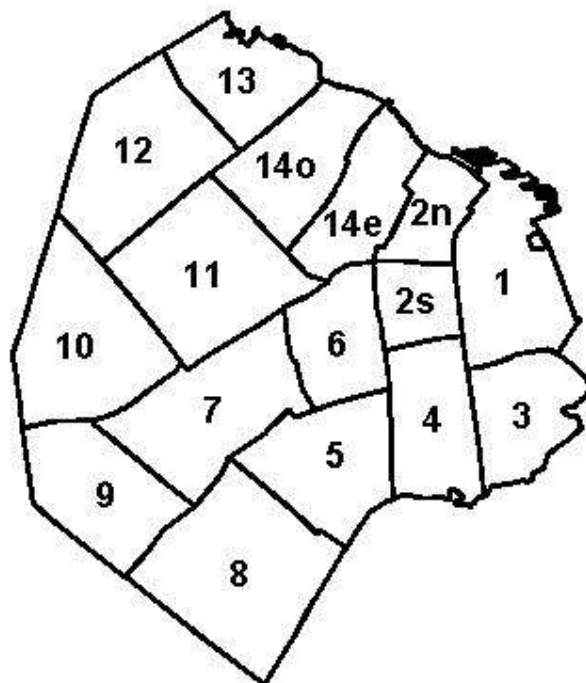
Valores estimados + IVA sujetos a variación, en Euros.

Fuente: The Coffee Store

<http://www.thecoffeestore.com.ar/sr800/home.asp?secc=franquicias&item=indica>

**Anexo 2 – Estadísticas Ciudad de Buenos Aires****Anexo 2.1 – Barrios comprendidos por Centro de Gestión y Participación (CGP)**

CGP	Barrios comprendidos
1	Constitución, Puerto Madero, Recoleta, Retiro, SanTelmo, Montserrat, San Nicolás
2 Norte	Recoleta
2 Sur	Almagro, Balvanera
3	La Boca, Barracas
4	Boedo, Nueva Pompeya, Barracas, San Cristóbal, Parque Patricios
5	Villa Soldati, Nueva Pompeya
6	Caballito, Almagro, Boedo
7	Floresta, Villa Luro, Vélez Sarsfield, Flores, Caballito, Parque Avellaneda, Parque Chacabuco
8	Lugano, Riachuelo, Villa Soldati
9	Liniers, Mataderos
10	Floresta, Monte Castro, Vélez Sarfield, Versalles, Villa Devoto, Villa Luro, Villa del Parque, Villa Pueyrredón, Villa Real, Villa Santa Rita
11	Agronomía, Villa Gral. Mitre, Villa Santa Rita, Villa del Parque, Villa Ortúzar, Paternal, Villa Crespo
12	Coghlan, Saavedra, Belgrano R, Villa Urquiza, Villa Pueyrredón
13	Nuñez, Belgrano
14 Este	Palermo, Villa Crespo
14 Oeste	Palermo, Belgrano, Colegiales, Chacarita, Villa Crespo

Anexo 2.2 – Mapa de los CGP

Fuente: Gobierno de la Ciudad de Buenos Aires (<http://www.buenosaires.gov.ar>)



Anexo 2.3 – Estadísticas obtenidas de la Encuesta Anual de Hogares 2004 (EAH 2004).

CGP	Edad								Total
	0 - 9	10 - 19	20 - 29	30 - 39	40 - 49	50 - 59	60 - 69	70 y más	
1	8,2	12,1	19,6	15,6	10,5	12,0	10,3	11,7	100,0
2 norte	8,3	9,0	26,5	13,1	9,7	10,8	9,8	12,7	100,0
2 sur	10,4	13,6	19,5	13,9	10,4	13,1	7,0	12,2	100,0
3	15,8	16,0	15,8	13,0	13,8	9,6	6,9	9,1	100,0
4	12,7	16,0	15,5	14,4	10,1	12,4	9,0	9,9	100,0
5	15,8	16,8	15,5	13,8	12,2	9,8	6,3	9,8	100,0
6	10,0	10,2	14,3	15,0	12,8	12,1	11,4	14,3	100,0
7	12,2	10,4	15,8	12,7	11,2	12,3	10,5	15,0	100,0
8	17,3	18,2	15,4	13,3	11,7	8,8	7,4	8,0	100,0
9	12,8	12,5	14,1	12,5	10,6	10,9	12,4	14,3	100,0
10	9,8	12,7	15,9	12,0	12,8	12,3	8,9	15,7	100,0
11	10,5	13,5	15,9	13,8	11,7	12,8	9,8	11,9	100,0
12	10,1	12,8	14,5	13,9	13,4	12,4	11,6	11,2	100,0
13	10,3	9,1	17,7	14,8	14,3	11,1	11,5	11,3	100,0
14 oeste	10,4	10,4	16,8	12,7	14,5	14,7	9,0	11,6	100,0
14 este	10,1	11,2	18,1	16,6	14,0	11,0	9,5	9,6	100,0
Ciudad de Bs. As.	11,3	12,5	17,0	13,8	12,1	11,8	9,5	12,0	100,0

Fuente: Encuesta Anual de Hogares. Año 2004. Dirección General de Estadística y Censos (GCBA).

CGP	Condición de asistencia escolar			Total
	Asiste	No asiste pero asistió	Nunca asistió	
1	28,4	70,9	0,7	100,0
2 norte	34,9	64,9	0,2	100,0
2 sur	31,3	67,5	1,1	100,0
3	31,7	66,8	1,6	100,0
4	32,0	66,5	1,5	100,0
5	32,2	65,8	2,0	100,0
6	25,3	74,3	0,4	100,0
7	29,1	70,3	0,6	100,0
8	33,5	63,8	2,6	100,0
9	27,8	70,8	1,3	100,0
10	28,4	70,9	0,7	100,0
11	30,1	69,0	0,9	100,0
12	29,8	69,5	0,7	100,0
13	30,2	69,5	0,3	100,0
14 oeste	30,9	68,8	0,3	100,0
14 este	33,0	66,9	0,1	100,0
Ciudad de Bs. As.	30,4	68,8	0,9	100,0

Fuente: Encuesta Anual de Hogares. Año 2004. Dirección General de Estadística y Censos (GCBA).



Condición de asistencia	CGP	Sector			
		Estatal / Público	Privado	Ns/Nc	Total
Asiste actualmente	1	60,8	38,1	1,1	100,0
	2 norte	53,9	45,5	0,6	100,0
	2 sur	69,4	30,1	0,6	100,0
	3	70,2	28,0	1,8	100,0
	4	71,9	27,9	0,2	100,0
	5	74,0	25,5	0,5	100,0
	6	56,7	41,7	1,6	100,0
	7	54,8	44,6	0,6	100,0
	8	74,5	25,0	0,5	100,0
	9	60,8	36,5	2,8	100,0
	10	57,3	42,0	0,7	100,0
	11	59,8	39,4	0,8	100,0
	12	49,7	50,0	0,4	100,0
	13	47,7	52,3	0,0	100,0
	14 oeste	50,6	49,4	0,0	100,0
	14 este	55,1	44,7	0,3	100,0
	Ciudad de Bs. As.	59,9	39,4	0,7	100,0
No asiste pero asistió	1	76,6	23,3	0,1	100,0
	2 norte	70,8	28,9	0,3	100,0
	2 sur	84,3	15,6	0,1	100,0
	3	87,1	12,8	0,1	100,0
	4	88,8	11,2	0,0	100,0
	5	89,0	10,7	0,3	100,0
	6	80,3	18,3	1,4	100,0
	7	81,9	17,9	0,3	100,0
	8	88,4	11,5	0,1	100,0
	9	80,5	18,2	1,4	100,0
	10	76,8	21,4	1,8	100,0
	11	82,7	17,2	0,1	100,0
	12	78,9	20,7	0,4	100,0
	13	72,0	27,8	0,2	100,0
	14 oeste	75,8	24,2	0,0	100,0
	14 este	72,3	27,3	0,3	100,0
	Ciudad de Bs. As.	80,1	19,5	0,5	100,0
Total	1	72,0	27,5	0,4	100,0
	2 norte	64,9	34,7	0,4	100,0
	2 sur	79,6	20,2	0,2	100,0
	3	81,7	17,7	0,7	100,0
	4	83,3	16,6	0,1	100,0
	5	84,1	15,6	0,3	100,0
	6	74,3	24,2	1,5	100,0
	7	73,9	25,7	0,3	100,0
	8	83,6	16,2	0,2	100,0
	9	74,9	23,3	1,8	100,0
	10	71,2	27,3	1,5	100,0
	11	75,7	23,9	0,3	100,0
	12	70,1	29,5	0,4	100,0
	13	64,7	35,2	0,1	100,0
	14 oeste	68,0	32,0	0,0	100,0
	14 este	66,6	33,0	0,3	100,0
	Ciudad de Bs. As.	73,9	25,6	0,5	100,0

Fuente: Encuesta Anual de Hogares. Año 2004. Dirección General de Estadística y Censos (GCBA).



C G P	Nivel Máximo Alcanzado							Total
	Hasta Primario Incompleto	Primario Completo	Secundario Incompleto	Secundario Completo	Superior / Universitario Incompleto	Superior / Universitario Completo	Sin dato	
1	3,6	15,7	12,8	20,0	17,6	30,3	0,0	100,0
2 norte	2,3	9,6	5,6	20,2	19,4	42,9	0,0	100,0
2 sur	5,5	14,5	16,3	20,9	17,3	25,4	0,0	100,0
3	9,7	21,5	19,4	21,6	12,7	14,9	0,1	100,0
4	7,5	23,7	16,0	20,0	14,9	17,8	0,0	100,0
5	10,3	26,0	21,0	22,1	10,8	9,8	0,0	100,0
6	4,2	13,6	10,1	24,7	15,3	31,9	0,1	100,0
7	4,2	17,0	13,1	24,9	16,4	24,4	0,1	100,0
8	10,7	30,5	20,2	21,2	9,3	8,2	0,0	100,0
9	5,2	25,5	16,6	24,4	15,2	12,9	0,1	100,0
10	3,2	19,7	13,1	23,7	16,2	24,1	0,0	100,0
11	4,7	19,1	16,3	18,5	17,0	24,3	0,0	100,0
12	3,6	15,1	12,8	22,4	16,9	29,1	0,0	100,0
13	2,4	8,5	8,5	19,3	22,4	38,9	0,0	100,0
14 oeste	2,2	9,9	10,2	20,0	18,4	39,3	0,0	100,0
14 este	2,3	10,1	7,2	18,0	18,6	43,9	0,0	100,0
Ciudad de Bs. As.	4,7	16,7	13,1	21,5	16,5	27,6	0,0	100,0

Fuente: Encuesta Anual de Hogares. Año 2004. Dirección General de Estadística y Censos (GCBA).

CGP	Cantidad de personas por ambiente				
	Hasta 0.99	De 1 a 1.99	De 2 a 3	Más de 3	Total
1	41,8	48,6	7,0	2,6	100,0
2 norte	57,1	39,2	2,3	1,4	100,0
2 sur	41,1	51,4	4,6	2,9	100,0
3	34,0	51,4	8,5	6,2	100,0
4	37,8	54,3	5,4	2,5	100,0
5	33,0	52,3	9,0	5,7	100,0
6	52,7	45,1	1,2	1,0	100,0
7	46,0	50,9	2,6	0,6	100,0
8	31,6	54,5	9,4	4,5	100,0
9	46,0	49,2	4,0	0,8	100,0
10	49,7	48,9	1,0	0,3	100,0
11	44,2	52,9	2,2	0,7	100,0
12	49,8	48,2	1,9	0,1	100,0
13	55,1	43,9	0,6	0,4	100,0
14 oeste	56,4	40,3	2,1	1,3	100,0
14 este	55,7	41,6	1,2	1,5	100,0
Ciudad de Bs. As.	47,1	47,8	3,4	1,7	100,0

Fuente: Encuesta Anual de Hogares. Año 2004. Dirección General de Estadística y Censos (GCBA).



C G P	Años de escolarización								Mediana	Media
	Hasta 3	de 4 a 6	7	de 8 a 10	de 11 a 12	13	14 y más	Total		
1	1,8	2,1	15,9	10,3	23,7	3,7	42,5	100,0	12,0	12,4
2 norte	2,1	0,5	9,9	4,2	22,8	2,1	58,5	100,0	15,0	13,7
2 sur	3,3	2,8	15,9	12,6	25,1	4,1	36,1	100,0	12,0	11,8
3	5,6	4,5	22,2	15,4	26,8	2,5	22,9	100,0	11,0	10,4
4	4,7	3,7	23,7	13,7	23,0	4,3	26,9	100,0	12,0	10,8
5	5,6	5,2	27,0	16,6	26,2	2,4	17,0	100,0	10,0	9,9
6	1,2	3,1	14,1	7,2	28,7	2,9	42,7	100,0	12,0	12,5
7	2,4	2,0	17,1	10,2	28,6	3,4	36,4	100,0	12,0	11,9
8	6,0	5,7	31,3	16,1	25,3	2,3	13,2	100,0	9,0	9,4
9	2,4	3,2	25,7	13,7	28,8	2,9	23,4	100,0	12,0	10,7
10	2,3	1,5	20,2	10,7	26,3	3,4	35,6	100,0	12,0	11,7
11	2,8	2,3	19,1	11,9	24,8	2,3	36,7	100,0	12,0	11,7
12	2,3	1,8	15,1	10,3	27,1	2,5	41,0	100,0	12,0	12,3
13	1,0	1,5	8,8	6,7	24,9	3,3	53,7	100,0	14,0	13,5
14 oeste	0,7	1,7	9,9	8,2	23,1	2,6	53,9	100,0	14,0	13,4
14 este	1,3	1,0	10,2	5,8	21,0	2,2	58,5	100,0	15,0	13,8
Ciudad de Bs. As.	2,6	2,5	17,0	10,3	25,5	2,9	39,2	100,0	12,0	12,0

Nota: se excluyen aquellos casos en los cuales no se dispone de suficiente información para determinar la cantidad de años de escolarización.

No se contabilizan años por las escuelas especiales no primarias

A los primarios especiales completos se les asigna 7 años. A los secundarios acelerados completos se les asigna 5 años.

Fuente: Encuesta Anual de Hogares. Año 2004. Dirección General de Estadística y Censos (GCBA).

CGP	Tipo de cobertura					Total
	Sistema Público y Plan Médicos de Cabecera	Obra Social	Medicina pre paga	Otros	Ns/Nc	
1	23,8	55,1	14,1	6,9	0,0	100,0
2 norte	13,5	45,7	23,7	17,1	0,0	100,0
2 sur	24,2	58,6	8,7	8,5	0,0	100,0
3	37,7	47,4	7,4	7,6	0,0	100,0
4	34,8	53,6	7,2	4,4	0,0	100,0
5	45,0	45,2	5,4	4,3	0,1	100,0
6	18,0	55,4	16,8	9,8	0,0	100,0
7	16,8	64,1	12,5	6,4	0,1	100,0
8	47,2	45,4	3,2	4,2	0,1	100,0
9	26,2	57,1	9,0	7,6	0,0	100,0
10	15,5	61,1	14,2	9,2	0,0	100,0
11	17,3	58,5	15,3	8,9	0,0	100,0
12	19,2	55,5	16,0	9,3	0,0	100,0
13	9,2	47,6	28,5	14,7	0,0	100,0
14 oeste	12,6	47,9	21,8	17,7	0,1	100,0
14 este	10,7	53,6	22,0	13,5	0,1	100,0
Ciudad de Bs. As.	22,0	54,0	14,4	9,5	0,0	100,0

Fuente: Encuesta Anual de Hogares. Año 2004. Dirección General de Estadística y Censos (GCBA).



CGP	Hacinamiento			Total
	Sin hacinamiento	Con hacinamiento	Con hacinamiento crítico	
1	90,3	7,0	2,6	100,0
2 Norte	96,4	2,3	1,4	100,0
2 Sur	92,5	4,6	2,9	100,0
3	84,9	8,4	6,7	100,0
4	92,0	5,4	2,6	100,0
5	85,1	9,0	5,9	100,0
6	97,8	1,2	1,0	100,0
7	96,9	2,6	0,6	100,0
8	85,8	9,4	4,8	100,0
9	95,3	4,0	0,8	100,0
10	98,7	1,0	0,3	100,0
11	97,1	2,2	0,7	100,0
12	98,0	1,9	0,1	100,0
13	99,0	0,6	0,4	100,0
14 Oeste	96,7	2,1	1,3	100,0
14 Este	97,3	1,2	1,5	100,0
Ciudad de Bs.As.	94,8	3,4	1,8	100,0

Fuente: Encuesta Anual de Hogares. Año 2004. Dirección General de Estadística y Censos (GCBA).

C G P	Condición de Actividad			Total
	Ocupados	Desocupados	Inactivos	
1	68,1	5,3	26,6	100,0
2 Norte	65,2	2,6	32,3	100,0
2 Sur	69,6	5,9	24,5	100,0
3	69,6	5,9	24,5	100,0
4	68,9	3,9	27,2	100,0
5	64,9	5,2	29,9	100,0
6	71,3	3,6	25,1	100,0
7	68,1	2,0	29,9	100,0
8	63,6	5,9	30,5	100,0
9	63,3	5,0	31,8	100,0
10	65,5	2,3	32,3	100,0
11	72,6	3,4	24,0	100,0
12	71,0	3,3	25,6	100,0
13	72,7	3,2	24,1	100,0
14 Oeste	70,6	4,3	25,0	100,0
14 Este	74,7	2,9	22,4	100,0
Ciudad de Bs. As.	69,1	3,9	27,0	100,0

Fuente: Encuesta Anual de Hogares. Año 2004. Dirección General de Estadística y Censos (GCBA).



Sexo	CGP	Tasa			
		Actividad	Empleo	Desocupación	Subocupación
Varones	1	70,8	63,8	9,9	8,1
	2 norte	71,0	67,7	4,7	3,3
	2 sur	72,7	66,7	8,3	9,8
	3	72,0	64,0	11,0	13,9
	4	70,1	63,1	10,0	11,1
	5	66,6	60,5	9,2	10,0
	6	73,2	68,5	6,4	4,7
	7	71,0	68,0	4,2	5,1
	8	67,9	60,8	10,4	12,3
	9	67,1	63,2	5,8	8,1
	10	71,0	68,3	3,9	11,0
	11	70,4	66,3	5,8	9,2
	12	72,3	67,7	6,3	7,3
	13	73,9	70,5	4,6	7,2
	14 oeste	71,4	67,4	5,6	5,7
	14 este	76,2	74,1	2,8	10,0
	Ciudad de Bs.As.	71,3	66,6	6,5	8,2
Mujeres	1	56,6	52,6	7,0	14,5
	2 norte	56,9	53,7	5,6	13,5
	2 sur	56,1	48,7	13,2	13,3
	3	55,5	49,9	10,1	24,3
	4	49,8	43,9	11,8	17,9
	5	46,9	40,1	14,4	17,3
	6	55,4	51,9	6,4	11,2
	7	50,7	45,0	11,3	8,5
	8	47,6	39,3	17,3	21,9
	9	45,9	39,7	13,4	16,8
	10	48,9	43,0	12,1	18,3
	11	54,1	49,9	7,8	15,8
	12	53,6	49,1	8,3	15,2
	13	59,0	54,1	8,3	8,2
	14 oeste	57,6	52,9	8,1	12,2
	14 este	62,6	58,1	7,2	15,3
	Ciudad de Bs.As.	53,9	48,7	9,6	14,6
Total	1	63,1	57,7	8,5	11,2
	2 norte	62,8	59,6	5,1	8,7
	2 sur	63,4	56,7	10,7	11,5
	3	62,8	56,1	10,6	19,0
	4	58,9	52,5	10,8	14,3
	5	56,2	49,7	11,5	13,2
	6	63,3	59,3	6,4	7,9
	7	59,8	55,3	7,5	6,7
	8	57,2	49,5	13,4	16,5
	9	55,3	50,1	9,3	12,1
	10	58,8	54,3	7,7	14,3
	11	61,4	57,2	6,8	12,4
	12	62,0	57,5	7,3	11,0
	13	65,5	61,3	6,5	7,7
	14 oeste	64,1	59,7	6,8	8,8
	14 este	68,6	65,2	5,0	12,7
	Ciudad de Bs.As.	61,7	56,7	8,0	11,3

Fuente: Encuesta Anual de Hogares. Año 2004. Dirección General de Estadística y Censos (GCBA).



CGP	Categoría Ocupacional					Total
	Patrón o empleador	Trabajador por cuenta propia	Asalariado	Trabajador familiar	Ns/Nc	
1	5,9	18,4	75,3	0,5	0,0	100,0
2 norte	7,0	18,9	73,0	1,1	0,0	100,0
2 sur	3,8	20,0	75,1	1,1	0,0	100,0
3	2,8	19,3	77,3	0,5	0,0	100,0
4	3,0	18,4	77,1	1,4	0,0	100,0
5	2,4	21,6	74,3	1,6	0,1	100,0
6	7,1	18,1	73,6	1,1	0,0	100,0
7	6,1	16,1	76,0	1,8	0,0	100,0
8	2,8	18,9	77,3	0,9	0,1	100,0
9	2,3	23,5	73,0	1,2	0,0	100,0
10	5,2	22,7	71,3	0,6	0,2	100,0
11	5,8	20,4	72,7	1,1	0,0	100,0
12	6,0	23,6	68,9	1,5	0,0	100,0
13	8,9	20,4	70,2	0,6	0,0	100,0
14 oeste	7,6	15,3	75,4	1,7	0,0	100,0
14 este	6,8	20,4	72,1	0,7	0,0	100,0
Ciudad de Bs. As.	5,6	19,5	73,7	1,1	0,0	100,0

Fuente: Encuesta Anual de Hogares. Año 2004. Dirección General de Estadística y Censos (GCBA).

CGP	Categoría ocupacional					Total
	Patrón o empleador	Trabajador por cuenta propia	Asalariado	Trabajador familiar		
1	2932,7	846,6	912,3	.		1014,0
2 Norte	2803,0	1434,8	1175,8	.		1326,2
2 Sur	2034,3	792,2	918,8	.		938,6
3	1441,6	600,8	707,5	.		705,6
4	1308,5	632,9	737,2	.		736,9
5	1409,0	583,5	660,5	.		661,8
6	1794,2	821,8	1029,0	.		1045,7
7	1630,9	807,9	907,6	.		932,8
8	1517,0	538,8	585,6	.		603,2
9	1521,7	730,7	770,9	100,0		778,0
10	2220,9	841,1	831,7	.		906,0
11	1639,7	912,3	937,4	.		974,9
12	2016,9	883,1	910,3	.		955,1
13	4462,0	1313,5	1163,6	.		1501,2
14 Oeste	2966,1	1475,9	1162,3	.		1321,0
14 Este	3404,2	1532,9	1127,9	.		1336,8
Ciudad de Bs.As.	2401,1	940,8	933,1	100,0		1011,3

Fuente: Encuesta Anual de Hogares. Año 2004. Dirección General de Estadística y Censos (GCBA).



Anexo 3 – Macros de Visual Basic y fórmulas Excel

Anexo 3.1 – Macro que agrupa las facturas

```
Sub Facturas()  
Dim i As Integer, j As Integer, pp As Integer  
  
j = 1  
pp = 1  
  
For i = 2 To 23642  
If Sheets("Tabla").Range("E" & i + 1).Value <>  
Sheets("Tabla").Range("E" & i).Value Then  
fecha = Sheets("Tabla").Range("C" & i).Value  
nro = Sheets("Tabla").Range("E" & i).Value  
cant = pp  
tot = Sheets("Tabla").Range("I" & i).Value  
prom = tot / cant  
Sheets("Facturas").Select  
j = j + 1  
Range("A" & j).Value = fecha  
Range("B" & j).Value = nro  
Range("C" & j).Value = tot  
Range("D" & j).Value = cant  
Range("E" & j).Value = prom  
pp = 1  
Sheets("Tabla").Select  
Else  
pp = pp + 1  
End If  
Next i  
  
End Sub
```

Anexo 3.2 – Macro que agrupa las facturas duplicadas

```
Sub agrupadas()  
Dim x As Integer, i As Integer  
  
For i = 2 To 11912  
If Range("B" & i + 1).Value = "" Then  
Exit For  
End If  
If Range("B" & i + 1).Value = Range("B" & i).Value Then  
w = Range("D" & i + 1).Value  
Range("D" & i + 1).Value = Range("D" & i).Value + w  
Range("E" & i + 1).Value = Range("C" & i + 1).Value  
/ Range("D" & i + 1).Value  
End If  
Next i  
  
End Sub
```



```
Range("E" & i + 1).Select  
Rows(i & ":" & i).Select  
Application.CutCopyMode = False  
Selection.Delete Shift:=xlUp  
i = i - 1  
End If  
Next i  
  
End Sub
```

Anexo 3.3 – Macro que selecciona aproximadamente 5000 facturas al azar

```
Sub random()  
  
Randomize Timer  
  
For i = 2 To 8101  
    x = Rnd  
    If x < 0.44 Then  
        Range("f" & i).Value = "X"  
    End If  
Next i  
  
End Sub
```

Anexo 3.4 – Macro que selecciona aproximadamente 6812 datos al azar

```
Sub random2()  
  
Randomize Timer  
  
For i = 2 To 16211  
    x = Rnd  
    If x < 0.4202 Then  
        Range("Q" & i).Value = "X"  
    End If  
Next  
  
End Sub
```



Anexo 3.5 – Macro que convierte datos de texto a valores numéricos

```
Sub hora()  
  
For i = 2 To 23643  
    x = Range("D" & i).Value  
    Range("D" & i).Select  
    ActiveCell.FormulaR1C1 = CInt(x)  
Next i  
  
End Sub
```

Anexo 3.6 – Fórmula de Excel que genera el campo “Franja”

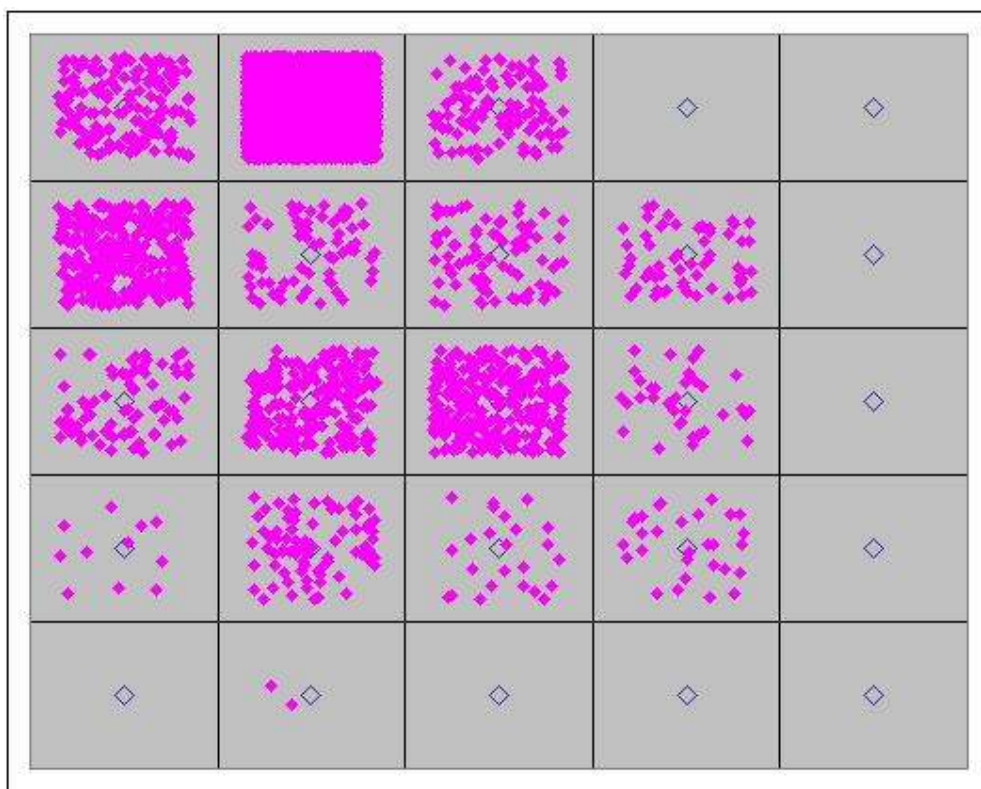
Siendo D2 el campo numérico que representa la hora:

```
=SI(Y(D2>200;D2<1059),"Mañana";SI(Y(D2>1100;D2<1259),"Media  
Mañana";SI(Y(D2>1300;D2<1559),"Mediodía";SI(Y(D2>1600;D2<1859),"Tarde";SI(Y(D2>1900;D2<2259  
),"Noche";"Trasnoche")))))
```



Anexo 4 – Gráficos adicionales del estudio de segmentación (herramienta NNClust)

Anexo 4.1 – Gráfico de la cuarta corrida



Anexo 4.2 – Datos de la cuarta corrida

	General	Cluster							
		1	2	3	4	5	6	7	8
Columna en el mapa		1	1	1	1	2	2	2	2
Fila en el mapa		2	3	4	5	1	2	3	4
Tamaño del Cluster	5000	11	81	309	151	2	79	194	69
TOTAL de la factura	10,2	106,7	28,1	13,2	22,0	49,0	16,3	12,5	13,4
CANTIDAD de ítems	2,0	13,5	3,9	1,6	4,3	3,5	3,0	3,0	2,7
PROMEDIO por ítem	4,9	10,9	7,5	8,7	5,4	14,3	5,7	4,3	5,0

	General	Cluster							
		9	10	11	12	13	14	15	16
Columna en el mapa		2	3	3	3	3	4	4	4
Fila en el mapa		5	2	3	4	5	2	3	4
Tamaño del Cluster	5000	3526	25	252	75	105	29	34	58
TOTAL de la factura	10,2	5,5	16,0	18,0	24,8	32,3	60,5	38,0	43,3
CANTIDAD de ítems	2,0	1,5	2,6	3,4	4,1	3,5	4,7	2,2	5,6
PROMEDIO por ítem	4,9	3,9	6,6	5,9	7,0	10,1	15,8	18,2	8,8



Anexo 5 – Gráficos adicionales del estudio de clasificación (herramienta NNClass)

Anexo 5.1 – Clasificación según Código de Tramo

Anexo 5.1.1 – Resultados del primer análisis

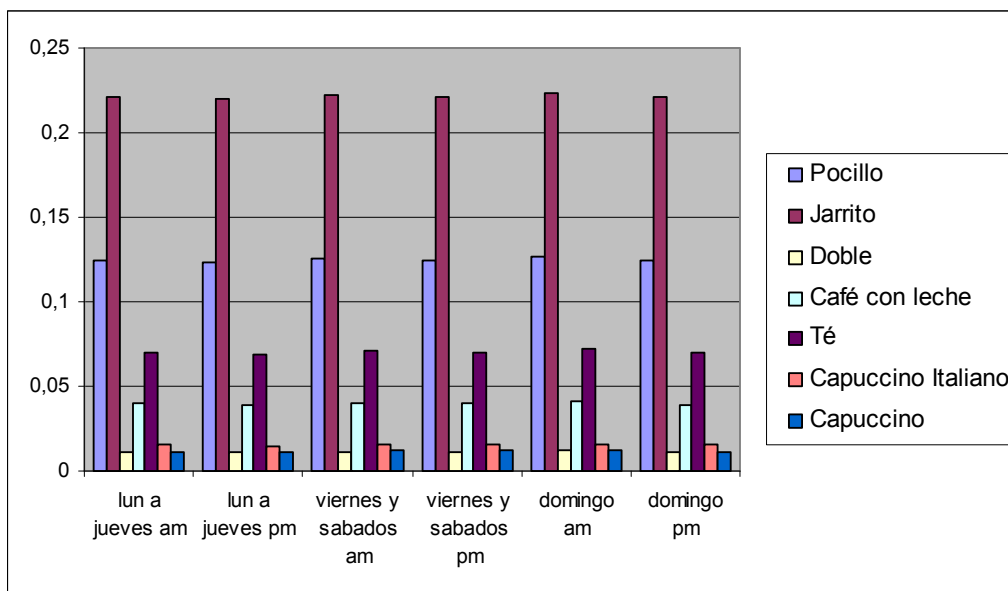


Gráfico A

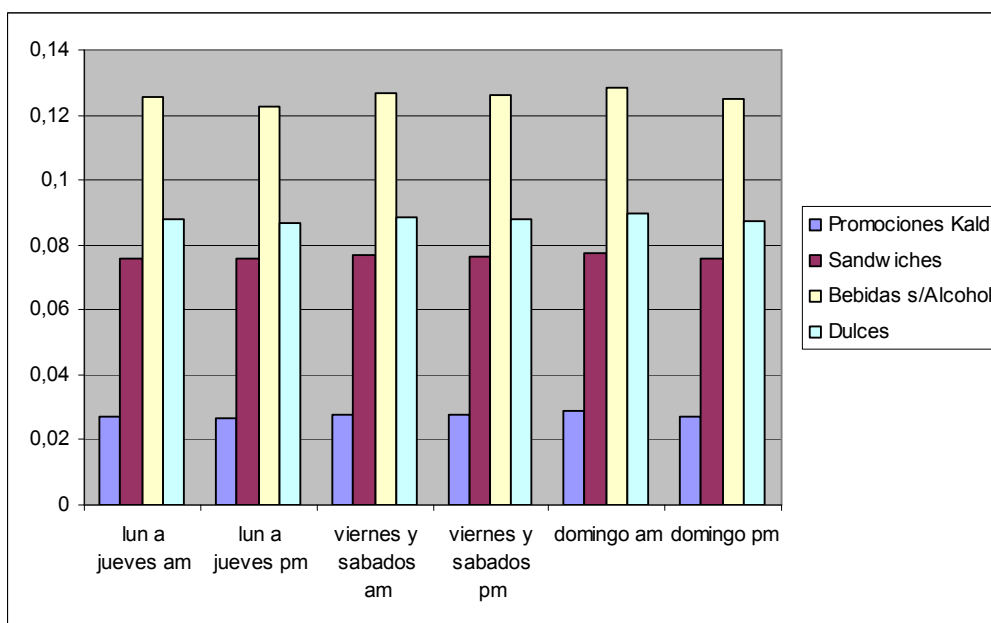


Gráfico B

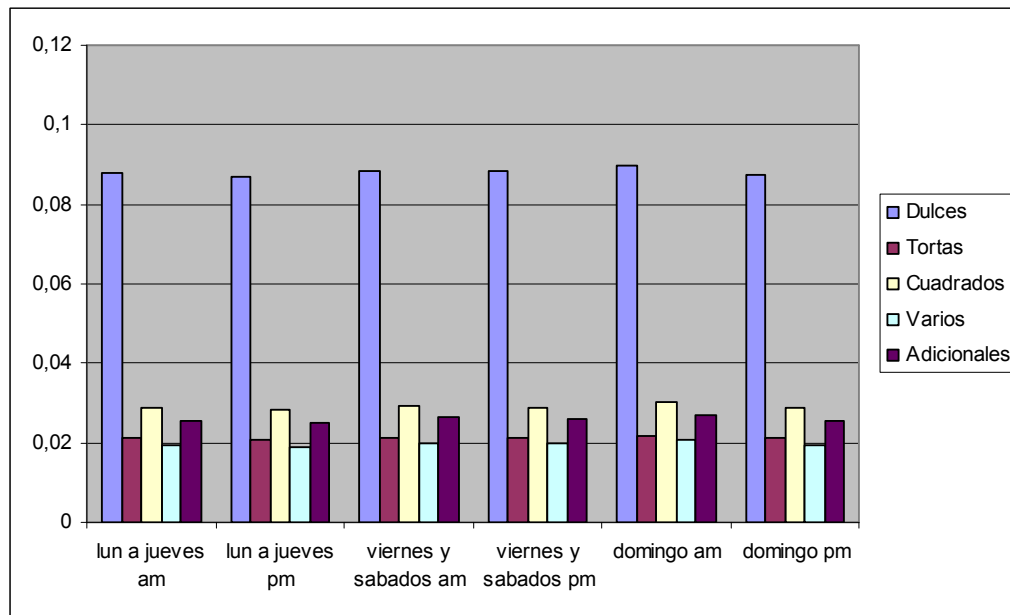


Gráfico C

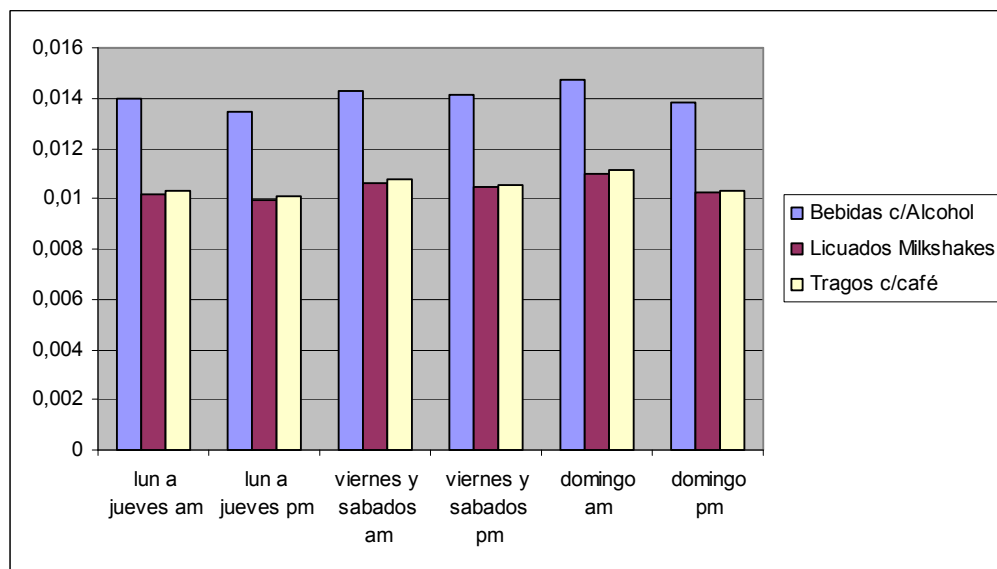


Gráfico D

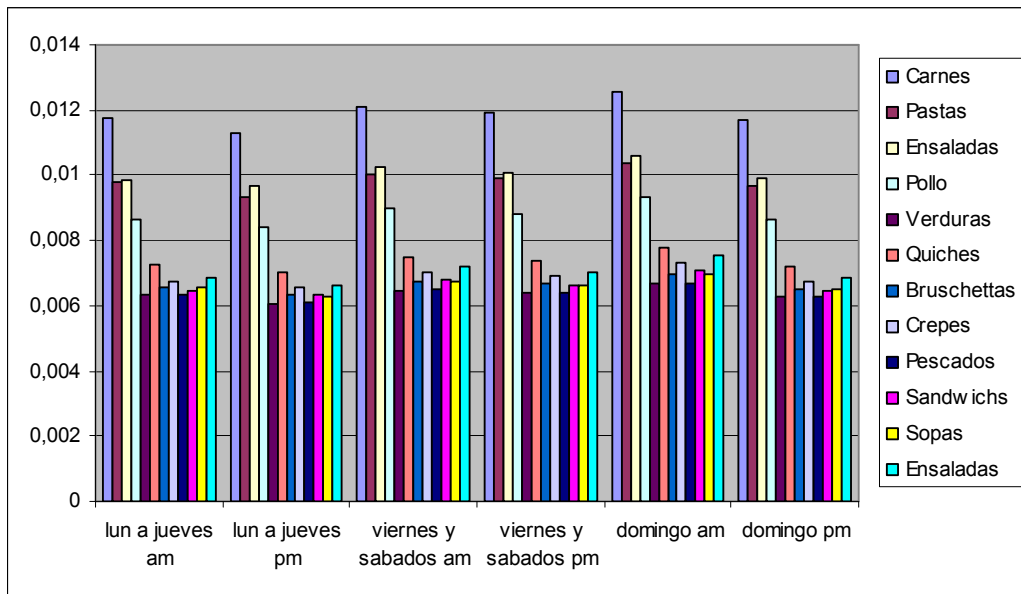


Gráfico E

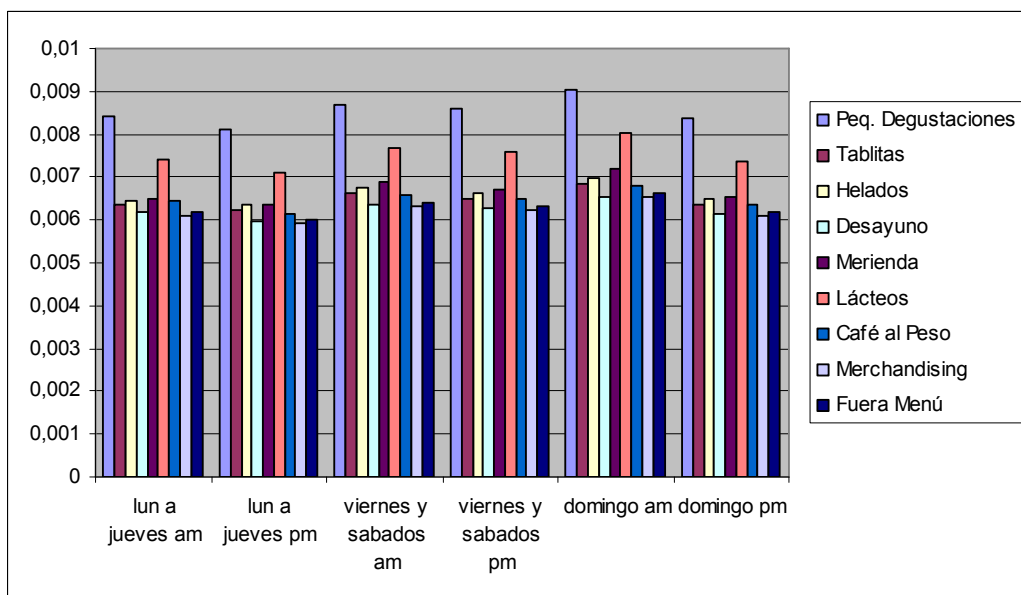


Gráfico F



Anexo 5.1.2 – Resultados del segundo análisis

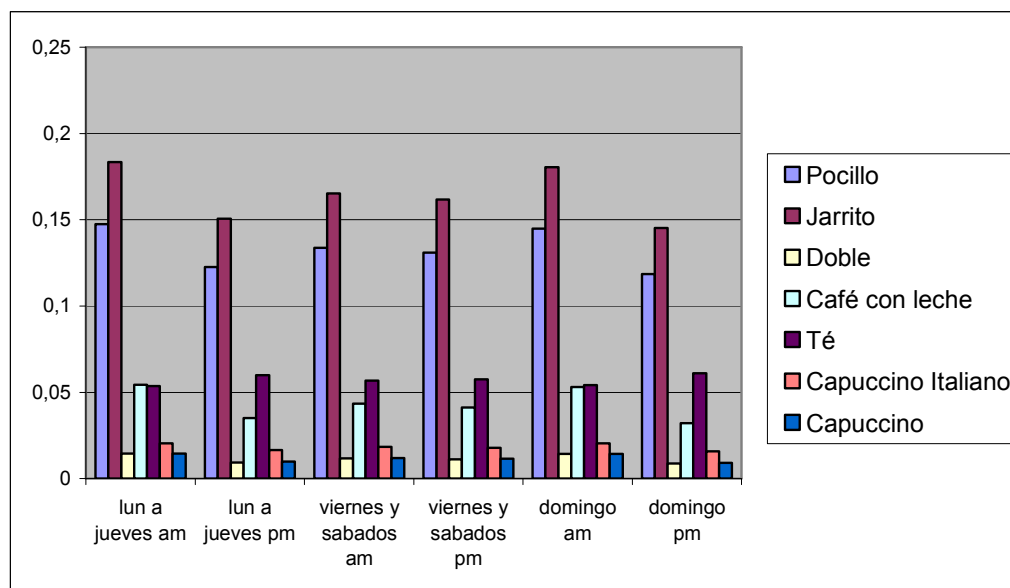


Gráfico A

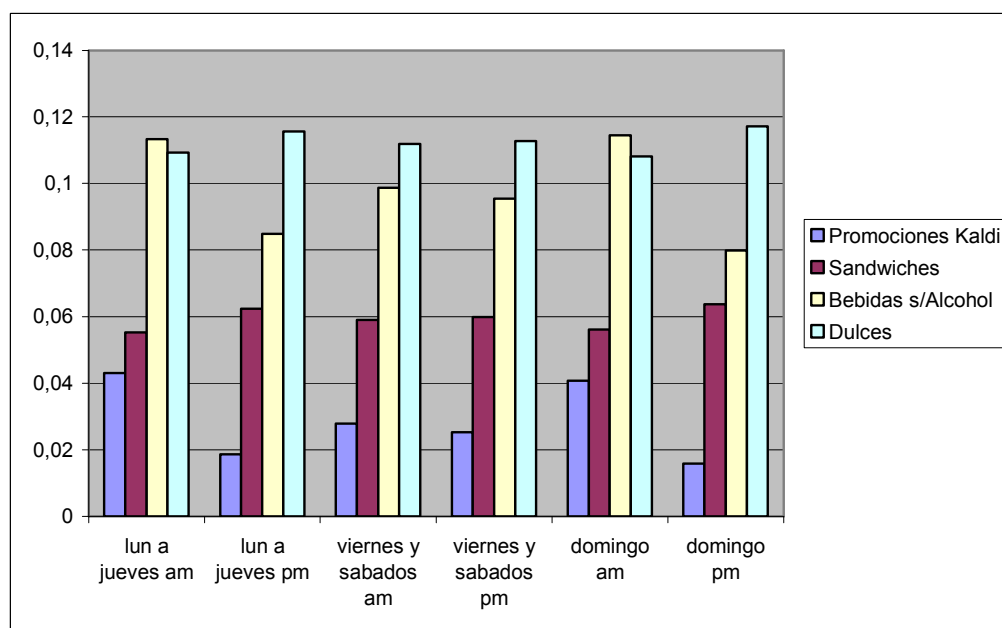


Gráfico B

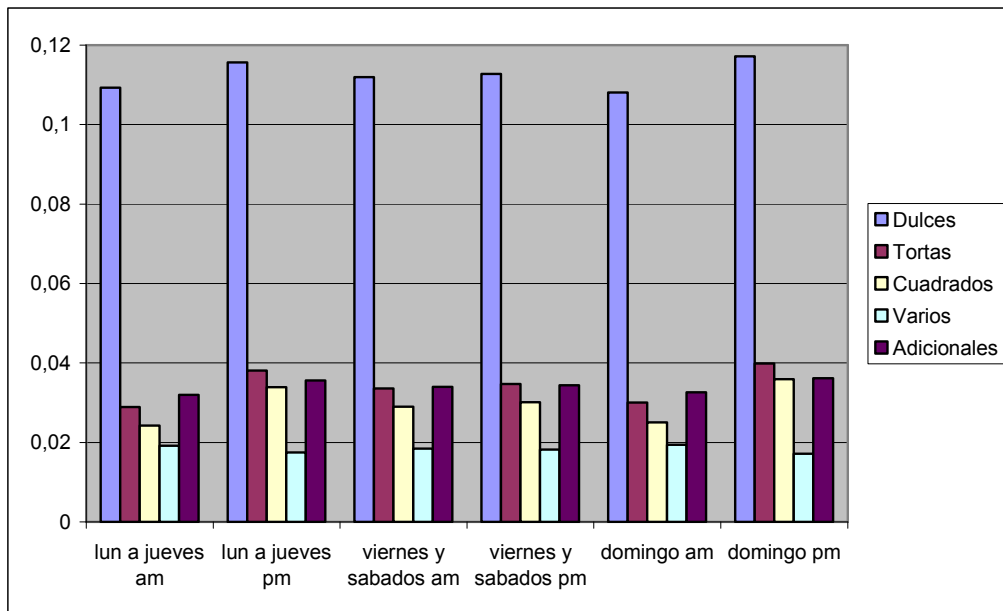


Gráfico C

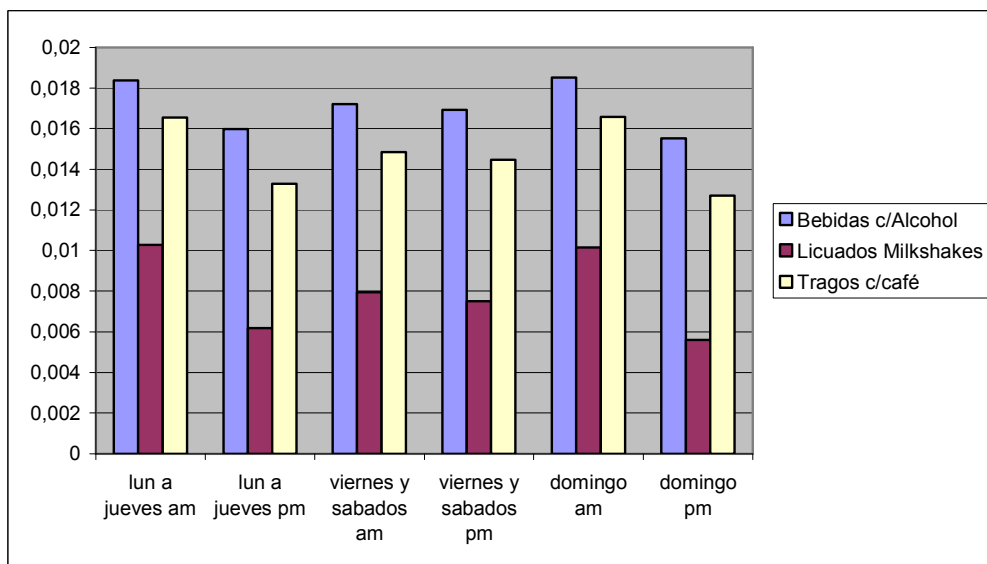


Gráfico D

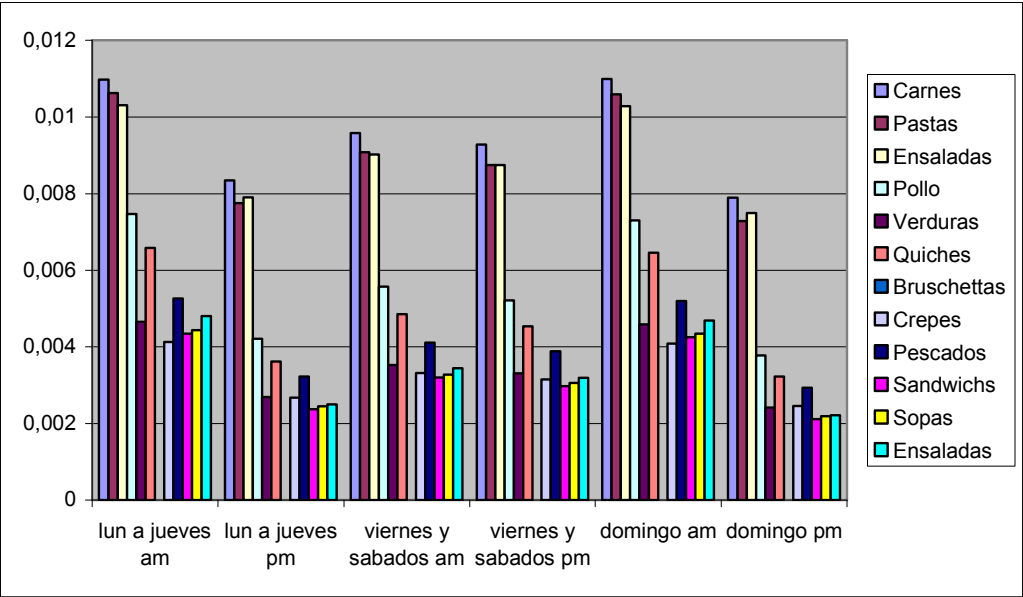


Gráfico E

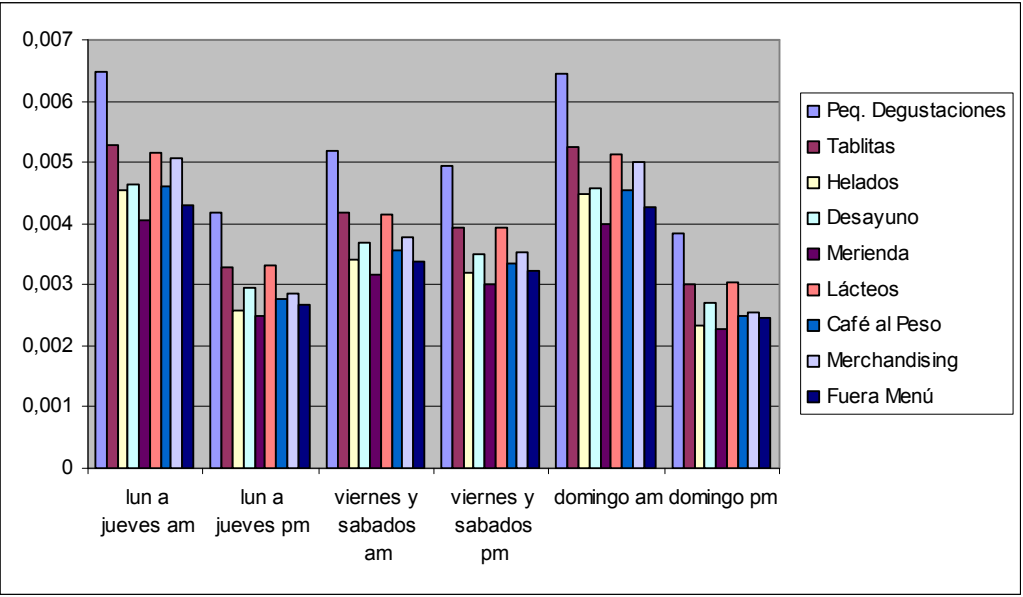


Gráfico F



Anexo 5.2 – Clasificación según Franja Horaria

Anexo 5.2.1 – Resultados del primer análisis

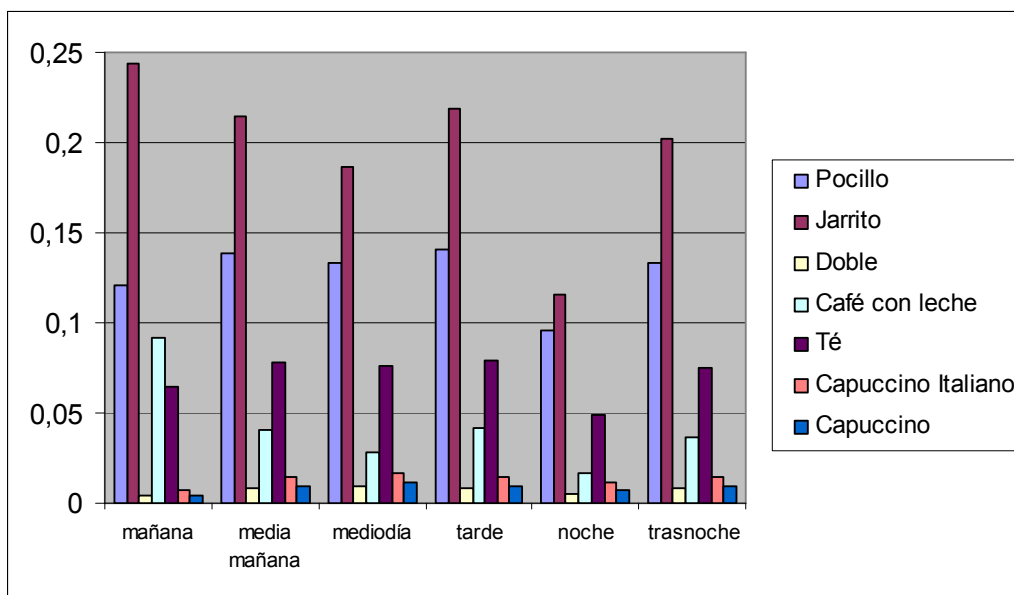


Gráfico A

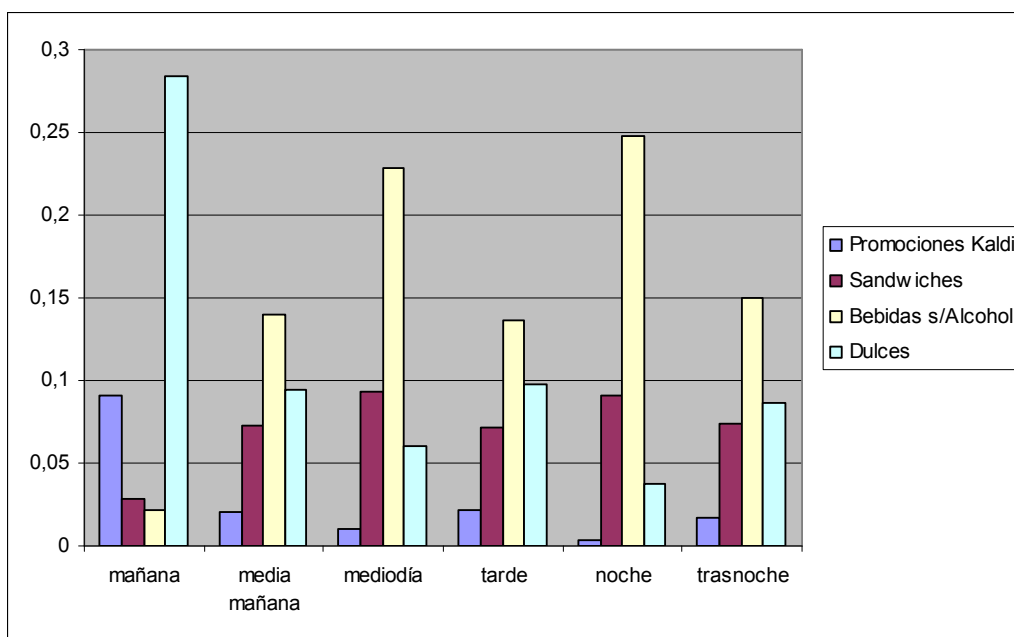


Gráfico B

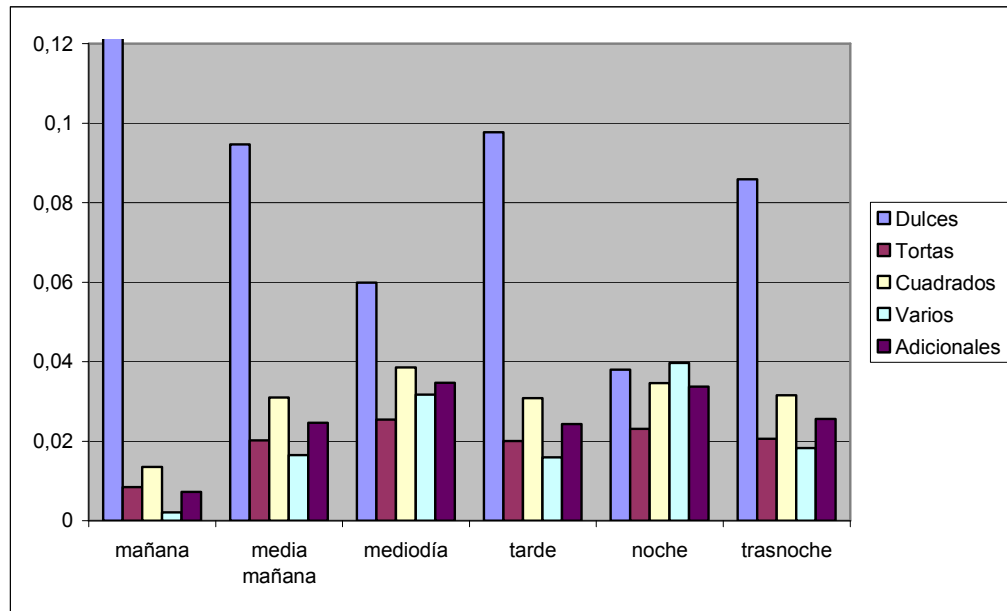


Gráfico C

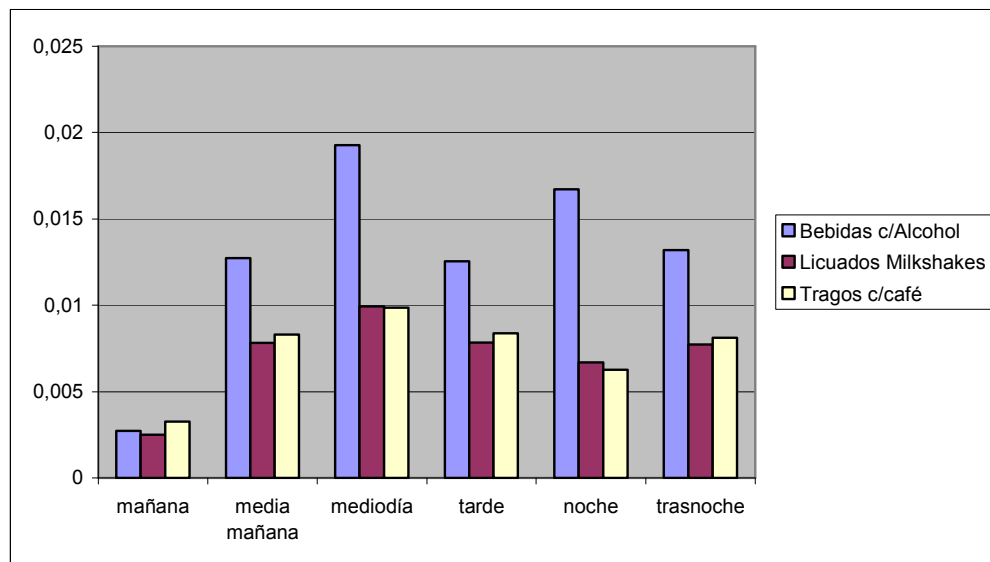


Gráfico D

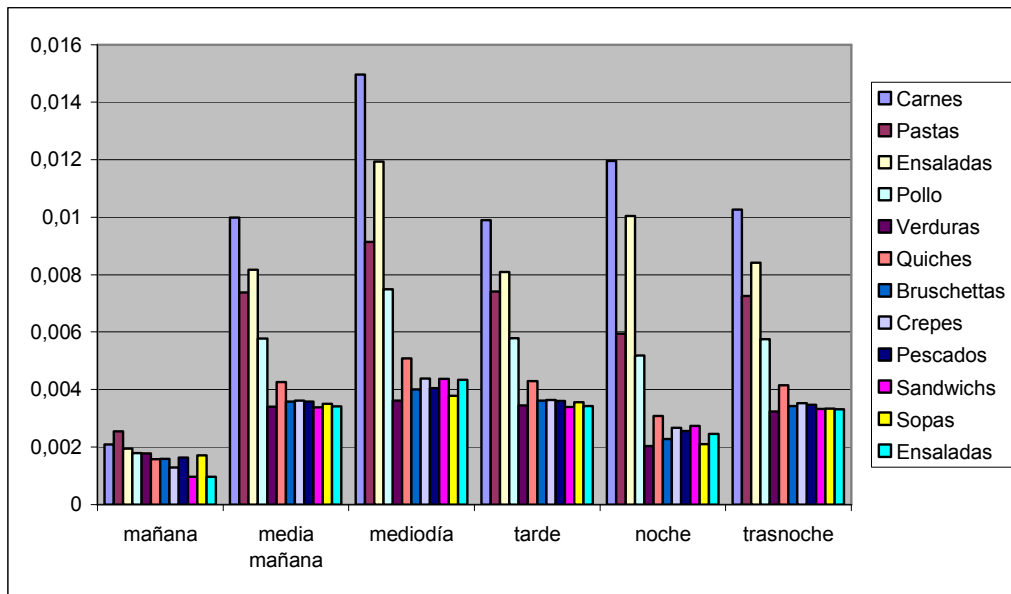


Gráfico E

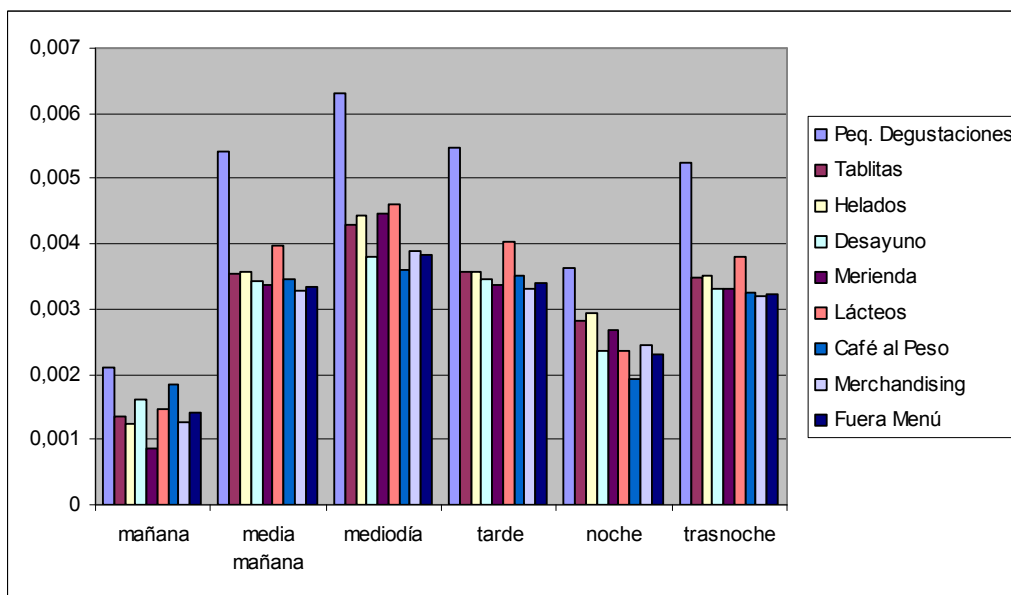


Gráfico F

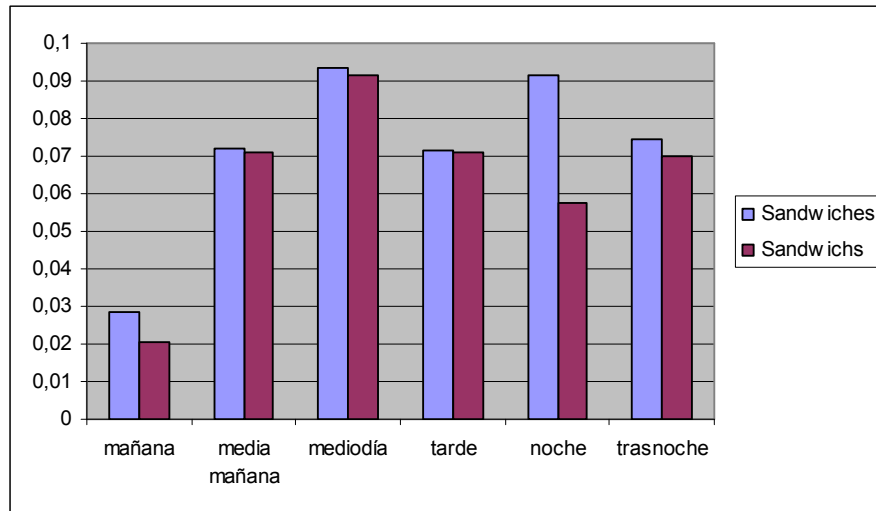


Gráfico G

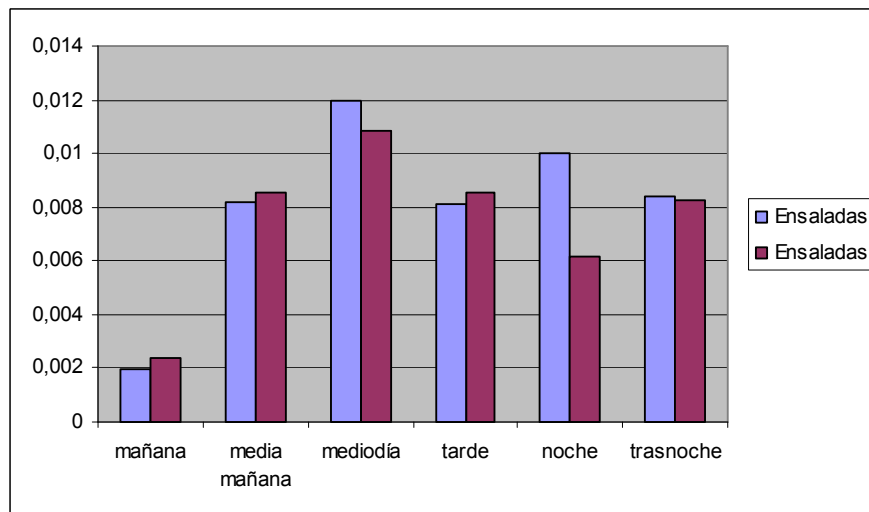


Gráfico H



Anexo 5.2.2 – Resultados del segundo análisis

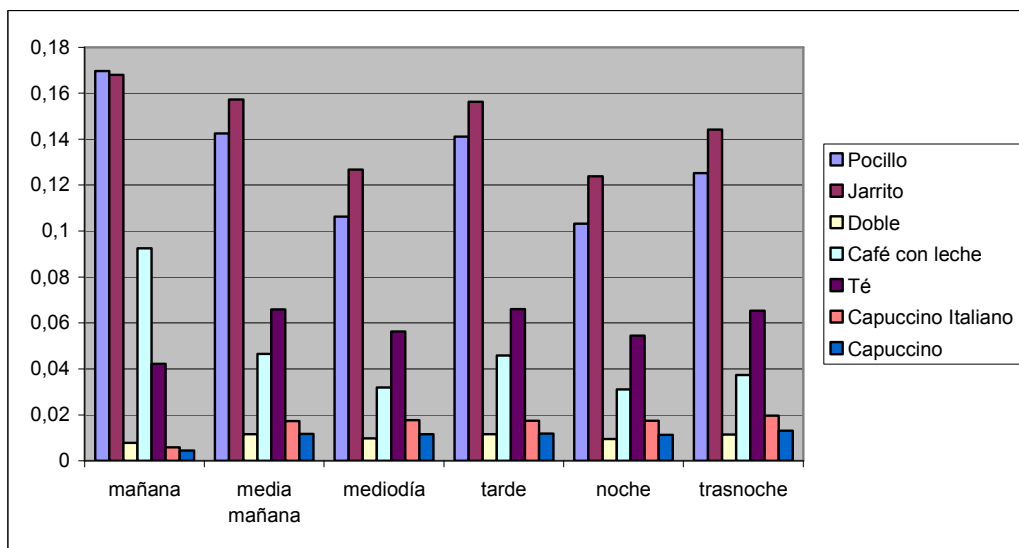


Gráfico A

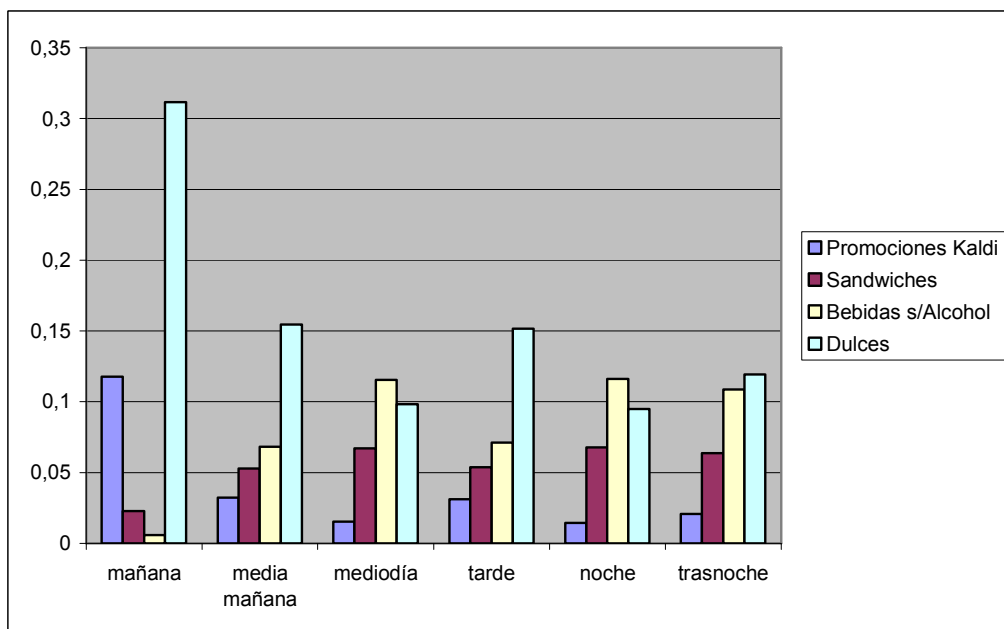


Gráfico B

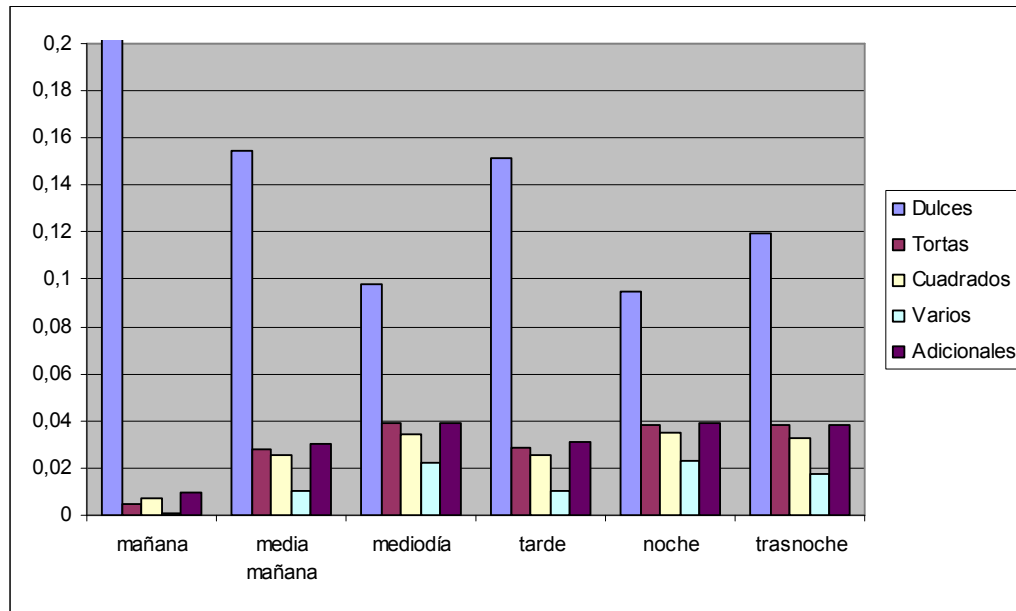


Gráfico C

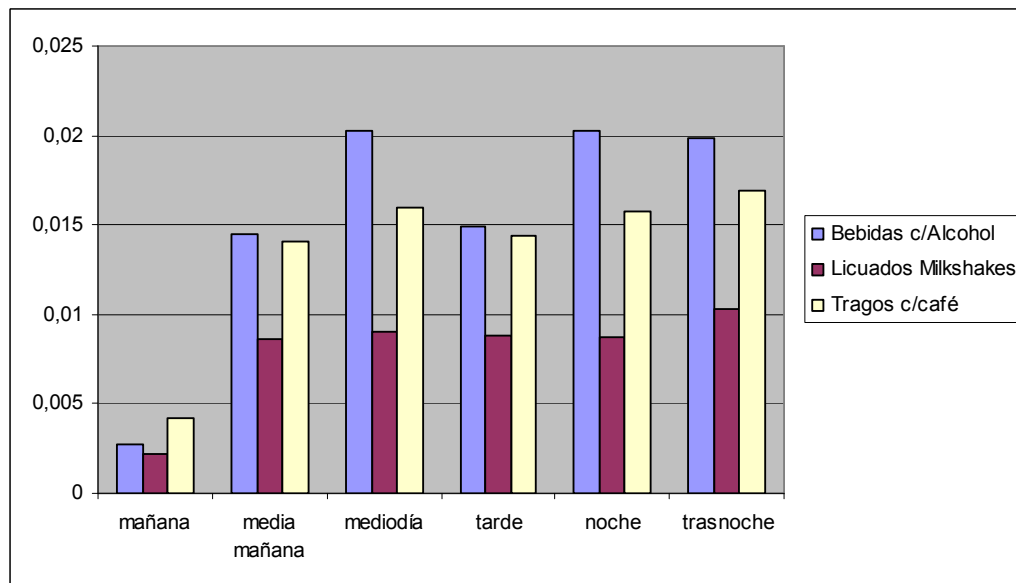


Gráfico D

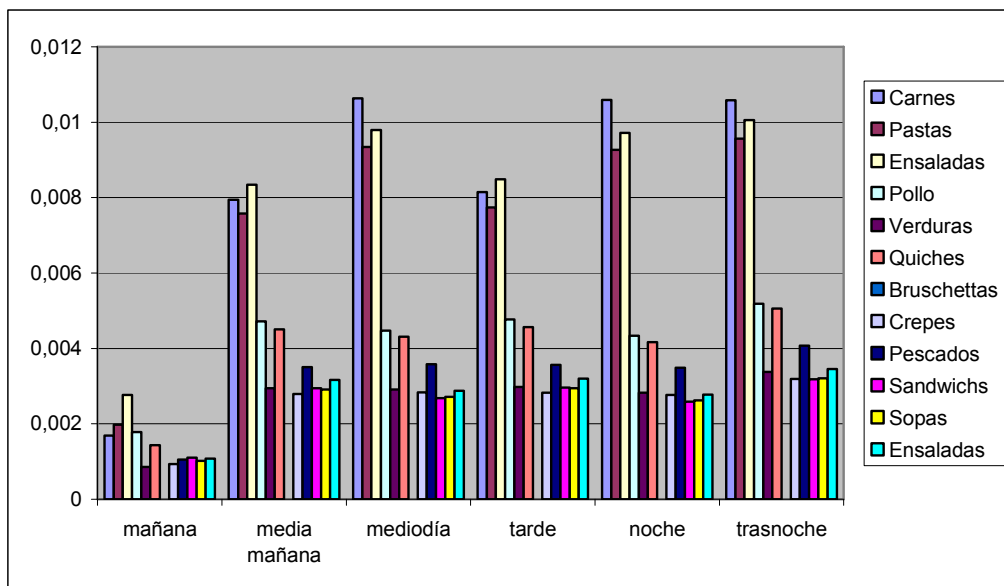


Gráfico E

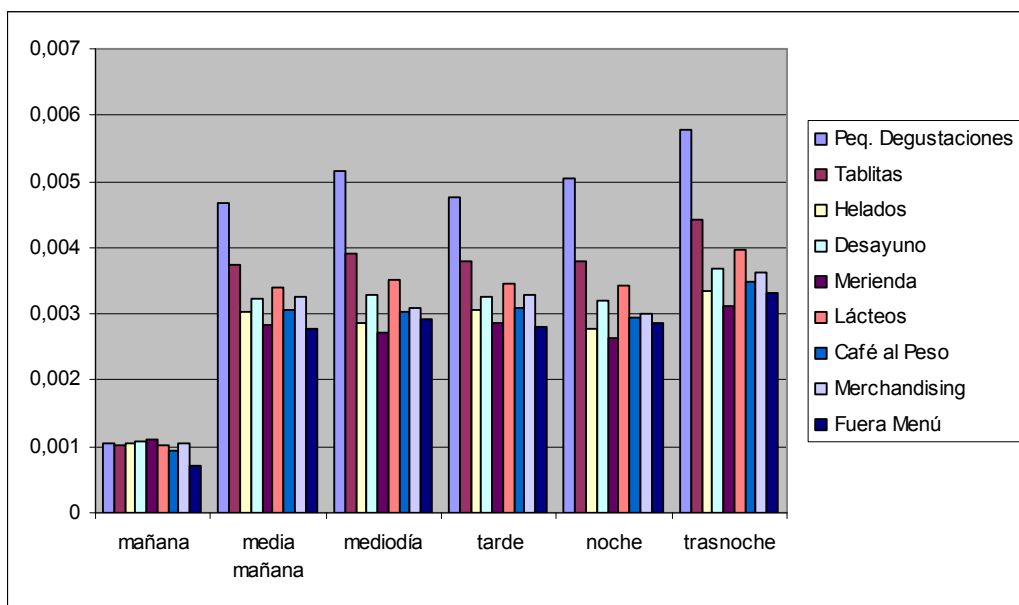


Gráfico F

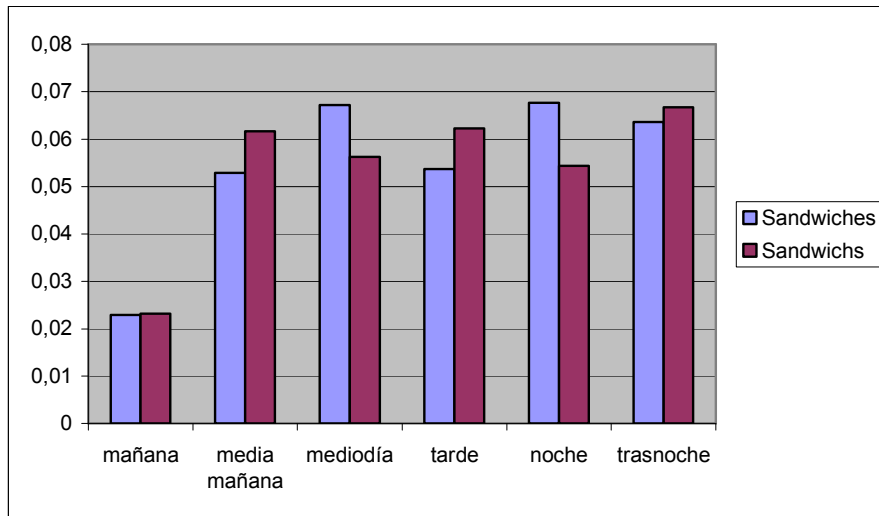


Gráfico G

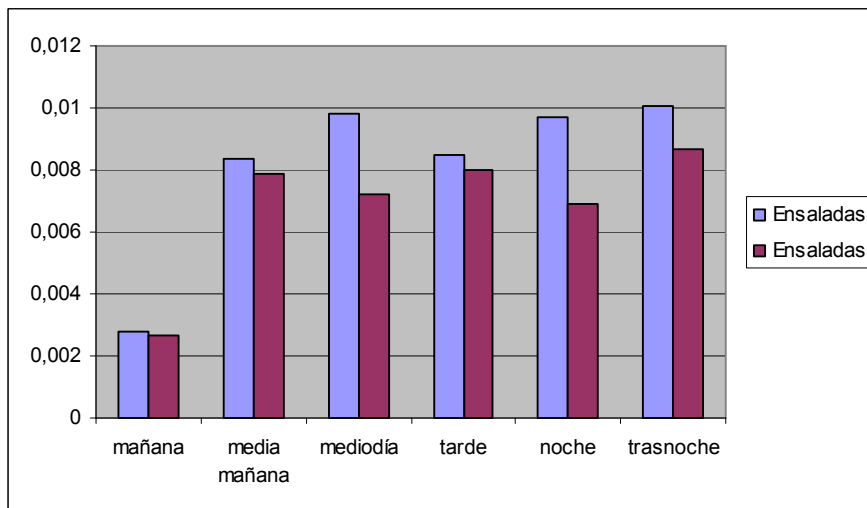


Gráfico H