



## Trabajo Fin de Grado

# El análisis de datos: Herramienta clave en el éxito del marketing actual.

Autor

**Luis Morales Blanco**

Directora

**Pilar Olave Rubio**

Facultad de Economía y Empresa

2014

**RESUMEN:**

---

Este trabajo pretende poner *en valor* el gran alcance que tiene la minería de datos y más concretamente, cuales son las técnicas estadísticas que tenemos que utilizar en función de los datos obtenidos, para usar el conocimiento que se deriva de ellos en la toma de decisiones empresariales y, en particular, en las variables estratégicas del marketing. El objetivo del proyecto es demostrar el gran interés que tiene para las empresas el tratamiento de la información para generar conocimiento y valor, en particular en el campo del Marketing y la Investigación de Mercados. Este objetivo lo llevaremos a cabo mediante el tratamiento de una encuesta realizada a empresas de 8 países (Alemania, España, Estados Unidos, Francia, Italia, Polonia, Reino Unido y Suecia) de los sectores químico, acero y equipamiento para el hogar en el año 2.007. A través de la implementación de las técnicas más usuales en análisis de datos multivariantes, constataremos su importancia para la toma de decisiones en el mundo empresarial y justificaremos el peso de la figura del analista de investigación de mercados en la empresa con conocimientos en minería de datos. Así pues, este proyecto consta de 4 partes: marco teórico, principales métodos multivariantes, análisis empírico y conclusiones y la correspondiente bibliografía.

**ABSTRACT:**

---

This essay tries to stand out the great scope the data mining has, and in a deeper way, what statistics techniques should be used on the obtained information, in order to use the knowledge coming from them in corporate decisions, particularly, in the variable marketing strategies. The aim of this project is to show the large interest that companies have in the treatment of information in order to generate knowledge and value, in the marketing field and particularly, in the market research. We will develop this aim through the information from a survey conducted in 8 different countries (Germany, Spain, United States, France, Italy, Poland, United Kingdom and Sweden) of the chemical, steel and Furniture sector in 2007. Thanks to the introduction of usual techniques in multivariate data analysis, we will be able to confirm the importance for taking decisions in the corporate world and we will prove the big significance a data mining analyst in the market research of a company. Therefore, this essay consists of four different parts: theoretical framework, main kinds of multivariate methods, empirical data treatment and conclusions, and the bibliography.

## **INDICE**

PARTE 1: INTRODUCCIÓN .....	5
PARTE 2: HERRAMIENTAS DE ANÁLISIS .....	15
2.1.TABLAS DE CONTINGENCIA Y CORRESPONDENCIAS .....	15
2.2.ANÁLISIS FACTORIAL .....	17
2.3.ANÁLISIS DE CONGLOMERADOS O CLUSTER.....	19
2.4.ANÁLISIS DE REGRESIÓN CON VARIABLE DEPENDIENTE CATEGÓRICA .	21
PARTE 3: DATOS Y METODOLOGIA.....	23
3.1.DATOS .....	23
3.2.SELECCIÓN DE VARIABLES .....	24
3.3.EXPLORANDO LOS DATOS .....	29
3.4.ANÁLISIS DE CORRESPONDENCIAS.....	31
3.5.REDUCCIÓN DE LA DIMENSIÓN. ESTUDIO DE OUTLIERS .....	33
3.6.ANÁLISIS DE CONGLOMERADOS .....	37
3.7.UN MODELO LOGIT SEGÚN INVERSIÓN EN TICS. .....	41
3.8.RESULTADOS DE LAS ESTIMACIONES Y CONCLUSIONES. .....	45
PARTE 4: CONCLUSIONES GENERALES.....	47
BIBLIOGRAFIA .....	48
ANEXOS	

## **INDICE DE ILUSTRACIONES**

Ilustración 1. Datos y fuentes de información .....	8
Ilustración 2. Ejemplo de análisis de correspondencias. ....	16
Ilustración 3. Ejemplo de gráfico en informe de clasificación. ....	20
Ilustración 4. Gráfico sectorial según tipo de mercado, cuota y sector de actividad.....	29
Ilustración 5. Gráfico sectorial según tipo de software. ....	30
Ilustración 6. Gráficos sectoriales según software ERP y tipo de mercado. ....	30
Ilustración 7. Perfil columna: intercambio de datos según tipo de sector. ....	31
Ilustración 8. Perfil fila: intercambio de datos según variación del presupuesto. ....	31
Ilustración 9. Diagrama de asociación entre el tipo de mercado y de intercambio de datos.....	32
Ilustración 10. Gráfico tridimensional del Análisis Factorial.....	36
Ilustración 11. Informe resumen de utilización de software de los distintos cluster.....	40

## **INDICE DE TABLAS**

Tabla 1. Ejemplo de matriz de componentes rotados en análisis factorial.....	18
Tabla 2. Tabla de frecuencias de los cluster.....	37
Tabla 3. Pruebas ómnibus sobre los coeficientes del modelo logit .....	42
Tabla 4. Prueba de Hosmer y Lemeshow .....	42
Tabla 5. Resumen del modelo .....	43
Tabla 6. Variables en la ecuación del modelo .....	43
Tabla 7. Tabla de probabilidades de Análisis Logit, según grupo. ....	44

## **PARTE 1: INTRODUCCIÓN**

Los estudios empíricos sobre creación de valor en la empresa y su relación con las variables específicas de marketing son escasos en muchos sectores, y aunque no es objetivo prioritario de este trabajo estudiar dicha relación, lo que planteamos es analizar la posible creación de valor a través de estrategias (de producción, de tamaño, de líneas de negocio, de cambios de productos, ...) que conforman la organización empresarial, y en este trabajo nos centraremos en aquellas que están más directamente relacionadas con el marketing (aumento de ventas, nuevos productos, satisfacción de clientes, estudio de fugas, ...).

Toda empresa debería tener una planificación a corto y una orientación del marketing a largo plazo, pero el futuro no es predecible. No podemos saber de antemano lo que puede hacer la competencia ante determinadas estrategias de marketing que realiza nuestra empresa, y lo que es más importante, no podemos conocer lo que hará el consumidor o hacia donde se dirigirá en su decisión de compra. Pero, sí podemos analizar modelos en los que aún sin dar resultados cuantitativos, estimemos la probabilidad de aumento en facturación cuando una nueva inversión en TICS se ha realizado en la empresa.

Si con el análisis de datos: ¿hubiera una forma de predecir el comportamiento del consumidor? ¿De predecir las necesidades que va a tener? ¿De conocer los gustos que tiene? Esto podría ahorrar millones y millones de euros a grandes empresas. Desde luego, nunca será esto posible (afortunadamente), pero sí aproximaciones fiables.

Un ejemplo de esto lo encontramos en la industria cinematográfica. Si una productora supiera a través solo del tráiler de la película si esa película va a tener o no éxito, se dirigiría solo a las películas que le aseguran un éxito en las carteleras, y no habría pérdidas millonarias de las películas que no tienen éxito y por tanto no se llevaría a cabo su producción. Esto se convertiría en una gran ventaja competitiva para cualquier empresa.

Algunas empresas ya están llevando a cabo esta “manera de actuar”. Por ejemplo, dentro del sector financiero se pueden analizar transacciones en tiempo real y estimar posibles fraudes para poder tomar acciones rápidas. Las grandes corporaciones ya disponen de departamentos de análisis de datos, para saber, por ejemplo, el análisis de

fugas de clientes, en particular, de compañías de seguros, ya que otro principio del marketing es el mantenimiento y cuidado de un cliente como valor de la empresa, más que la búsqueda de nuevos. Previamente, para determinar la probabilidad de fuga del cliente, realizan una segmentación de la cartera de clientes mediante un análisis cluster, así pueden crear “perfiles” de cada segmento según tipología de hogar, edad de sus componentes, ingresos, etc., para posteriormente determinar sus probabilidades de fuga. Los modelos utilizan variables de todo tipo: sociodemográficas, tipo de producto, de uso del seguro, indicadores de satisfacción de la empresa...

En Reino Unido, los supermercados Tesco consiguieron superar en 2006 la competencia del gigante Wal-Mart gracias al análisis de los hábitos de consumo de sus clientes. Algo que empezó entonces a ser posible gracias a la tarjeta de fidelización Clubcard, en la que a cambio de un pequeño descuento los clientes que hicieron esta tarjeta, quedaron registrados en las bases de datos de la empresa. La empresa llevó a cabo un análisis estadístico para conocer hábitos de compra de sus clientes y vieron una correlación significativa entre la compra de pañales y cerveza. Tesco aprovechó este hecho y reubicó en sus establecimientos ambos productos (situándolos muy próximos), para atraer de esta forma la compra conjunta a través de compras compulsivas. El resultado de esto, fue un gran aumento de las ventas de cerveza y pañales.

Con este ejemplo vemos como las grandes empresas se han convertido en laboratorios de datos para estudiar sus clientes y tomar decisiones. Todo esto se puede hacer mediante el análisis correcto de los datos. Pero nos encontramos en una situación en la que hay grandes masas de datos. Los datos ya están ahí, no hay que generarlos. Todo esto hace que actualmente se diga que el saber analizar y tratar los datos es el “petróleo” del siglo XXI y que las empresas más importantes a nivel internacional inviertan grandes sumas de dinero en poder predecir el comportamiento del cliente a través de los datos.

Este tipo de acciones las llevan a cabo también empresas públicas, algo habitual ya en EEUU. Por ejemplo, en la cárcel de Oregón se creó un modelo de predicción de crímenes. Este modelo, indicaba la probabilidad de que un malhechor sea condenado tras los tres años siguientes a su liberación por un delito. Esto puede ayudar a jueces a tomar decisiones sobre la libertad condicional. Además esta herramienta es de acceso libre. Si se conoce el ID del preso y el delito por el que cumple sentencia, al introducir

esos datos en la web pública de la Comisión de Justicia Criminal de Oregón puedes ver la probabilidad dada del modelo para ese preso, para más detalles [Siegel, E. (2013) en la página 91].

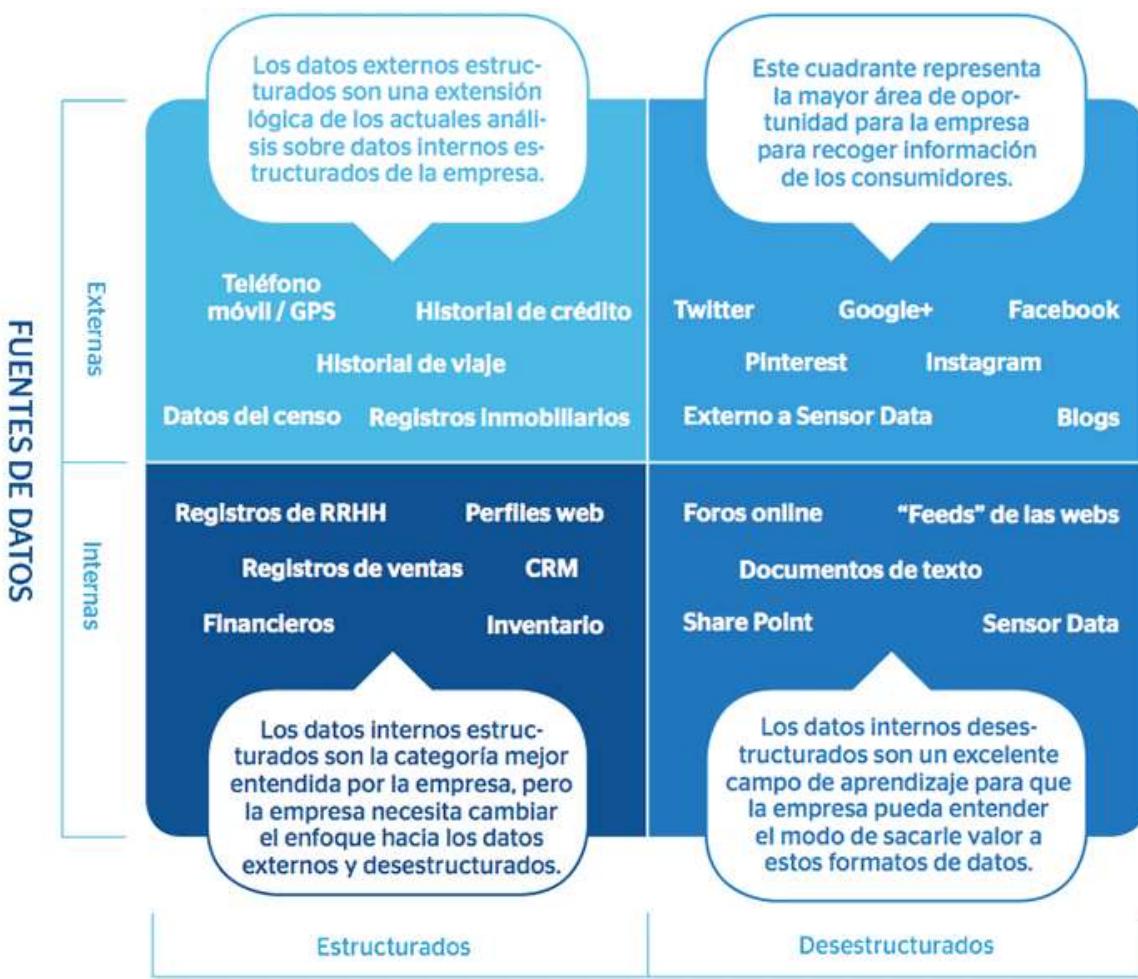
Igual que en marketing encontramos las 4Ps (Product, Price, Placement and Promotion) en los DATOS encontramos las 4Vs (Volumen, Variedad, Velocidad y Veracidad) aunque algunos autores están ampliando a 5Vs esta definición, añadiendo una V más (Valor de Uso). A continuación entraremos en una breve explicación de cada una de ellas.

- Volumen: el volumen del almacenamiento de estos datos, ha pasado de megabytes y gigabytes a “petabytes”. Día a día se generan una gran cantidad de datos. Un ejemplo de esto: Twitter genera 8 terabytes al día, Google procesa más de 20 petabytes al día. Todo esto visto como conjunto, es una grandísima fuente de información de datos a analizar. Nos encontramos en el principio de esta grandísima expansión de datos, ya que crece de forma exponencial. En 2006 la producción digital fue de aproximadamente de 180 exabytes<sup>1</sup>, en 2011 esta cifra se incrementó en 10 veces hasta alcanzar 1.800 exabytes y se prevé que para el año 2020 se alcancen 35.000 exabytes.
- Velocidad: como hemos comentado anteriormente, estos datos proceden del tráfico web y este se actualiza continuamente minuto a minuto. Por lo que el tratamiento y análisis de estas grandes bases de datos tiene que hacerse en tiempo real para que esta información generada de estas bases de datos ayude en la toma de decisiones.
- Veracidad: hace referencia a la calidad de los datos. Un requisito fundamental dentro del Big data es la alta calidad de los mismos. Pero esto no es nada fácil, ya que aunque se dispongan de métodos que puedan eliminar la imprevisibilidad de algunos datos, siempre quedarán otros que no se puede eliminar, como la economía, las decisiones de compra de un cliente, el tiempo...
- Variedad: hace referencia a los distintos formatos que encontramos en red que se pueden almacenar. Al principio eran solamente datos almacenados y estructurados que se guardaban en bases de datos, ahora hay que sumar los datos de cotizaciones bursátiles, los de las redes sociales, los datos en streaming, de la geolocalización, etc. En el gráfico siguiente se aprecia muy bien los distintos

<sup>1</sup> un exabyte= mil millones de gigabytes

datos que encontramos en la red, según si estos están o no estructurados y si son o no pertenecientes de la empresa.

**Ilustración 1. Datos y fuentes de información**



Fuente: Booz & Company | Benefitting from Big Data, 2012

- **Valor de uso:** el Big data, no solo son grandes bases de datos (volumen) con diferentes tipos de información (variedad) en las que los datos aparecen segundo a segundo y hay que analizarlos en tiempo real (velocidad) teniendo en cuenta la calidad de los mismos (veracidad), es algo que da valor a la empresa. Esta información que la empresa puede obtener añade una gran ventaja competitiva a la hora de toma de decisiones. No sólo aporta interés para el área del marketing (optimizando los resultados o procesos hacia el cliente), sino que también aporta

utilidad para la gestión financiera, la colaboración con los empleados, la optimización de operaciones y con nuevos modelos empresariales.

En la situación actual encontramos nuevos datos que son analizados por la empresa. Hace veinte años los negocios podían recolectar datos referentes a: transacciones en el punto de venta, respuesta a promociones con cupones, respuestas a campañas de correo, etc. Pero con las autovías de internet se han generado nuevos datos que pueden ser recolectados por las empresas: de interacciones en redes sociales (likes, comentarios, fotos...), de compras en línea, páginas webs vistas por los usuarios, porcentaje de visitas a la web, tiempo de estancia en una página y el site de la empresa, de uso de dispositivos móviles, de localización geográfica, datos meteorológicos y astronómicos, económicos y bursátiles, archivos fotográficos y de video, etc....

Tenemos que hacer referencia también, al concepto de Web 3.0, que está a punto de llegar y nos va a cambiar el día a día de nuestras vidas. La web 3.0 o internet de las cosas, hace referencia a un futuro próximo en el que estará conectado cualquier aparato a internet (coches, electrodomésticos, mascotas...). Es decir, con la web 3.0. nos inundará una gran oleada de nuevos datos que necesitan ser analizados para poder de esta forma conocer cualquier aspecto relacionado con el cliente. Ya que todo a nuestro alrededor generará datos.

Con todos estos datos disponibles de los usuarios, han surgido temores y opiniones en contra de esta práctica de investigación por poder utilizar los datos traspasando los límites legales y éticos. Y al usarlos de forma abusiva, por ser invasivos e ir en contra de la privacidad de las personas. Es por esto, que la seguridad en estas bases de datos es un fundamento clave.

A continuación, vamos a hacer una pequeña diferenciación entre tres términos muy importantes:

⇒ Open Data: hace referencia a las bases de datos de cualquier tipo (estructurados o no) que circulan libremente por internet y que minuto a minuto siguen incrementándose de forma exponencial debido al tráfico en la web, imágenes y videos, datos de encuestas, formularios, redes sociales, etc. Estos, están disponibles de forma libre para cualquier usuario que sepa descargarlos y analizarlos. Son libres, gratuitos y pueden añadir información a la empresa si se

saben utilizar. Un ejemplo de estos tipos de datos son los disponibles Redes Sociales. Algunas de estas bases de datos, que se consideran libres para cualquier usuario, requieren de cierto conocimiento para su uso.

- ⇒ Big Data: Estas son grandes bases de datos y no solo se entiende lo que son los datos sino también las tecnologías para poder tratarlos, ya que son bases de datos que no pueden ser tratadas de forma convencional porque su volumen supera la capacidad de las herramientas que habitualmente se utilizan para el tratamiento de los mismos. Un ejemplo de estas bases de datos, son las generadas por el INE, el IAest, Banco de España... Por ejemplo en el INE, uno de los organismos mencionados anteriormente, para utilizar y analizar los datos que tiene disponibles, tienes que saber el lenguaje y el formato del software que el investigador utiliza en su trabajo, ya que solo están disponibles los microdatos en determinadas formas. Estas bases de datos, en su mayoría procedentes de ENCUESTAS internacionales y/o nacionales son las que utilizaremos en el apartado 3 de esta memoria de investigación, y en estos casos es fundamental que el analista conozca cómo se han generado los datos y el tipo de muestreo que se ha realizado.
- ⇒ Minería de datos o datamining: es el conjunto de técnicas y procesos utilizados para extraer resultados de grandes bases de datos y de los cuales posteriormente sacar conclusiones útiles y relevantes (tendencias, patrones repetidos, explicaciones de comportamiento de un grupo de consumidores...) para que las empresas puedan tomar medidas. Lo importante no son los datos en sí, sino el conocimiento.

La minería de datos aporta información muy útil para los grandes directivos de empresas, los cuales de esta forma están mejor informados para una toma de decisiones más eficaces. La utilización de estos mecanismos dentro de la empresa, supone una ventaja competitiva frente a los consumidores, ya que supone ir un paso por delante de ellos y aporta una mejora en la competitividad y la productividad de las empresas si se hace correctamente.

Los datos que podemos obtener son muy distintos según su procedencia, según los hayamos generado nosotros o provengan de bases de datos ya generadas dentro o fuera de la empresa:

- ⇒ Open Data/Big Data: Grandes volúmenes de datos que el analista no ha generado, por lo tanto, los objetivos que podemos extraer de estas bases de datos, no están determinados a priori, ya que bases de datos que no generamos nosotros.
- ⇒ Encuesta de investigación: En este caso el analista plantea el modelo, lo que conlleva una selección de variables y de objetivos a priori.

Algunos de los beneficios que podemos desarrollar a partir de conclusiones derivadas de la minería de datos, tratamiento de grandes bases, hacen referencia a las siguientes áreas:

- Personalización de estrategias:

Podemos conocer quiénes son nuestros clientes, donde están, que quieren, como contactar con ellos, conocer sus preferencias, sus hábitos de compra y de esta forma segmentar las diferentes estrategias de marketing en función del perfil de clientes y de los gustos, preferencias y comportamientos de compra aportando así una experiencia personalizada al cliente, fomentando la fidelización del mismo, ya que se les puede enviarles anuncios personalizados que realmente les interesen.

Un ejemplo lo tenemos en la cadena de supermercados Tesco que a través de la información que recogen de las tarjetas de fidelización de sus clientes programan cupones promocionales de descuento adecuándose a cada cliente.

Otro ejemplo lo encontramos en el metro de Londres, que a través de la minería de datos tratan de practicar la personalización del cliente ofreciéndoles información relevante de sus rutas de viaje de horarios, retrasos o información sobre actualización de nuevas rutas. A través del correo electrónico (el usuario previamente se registra) se pueden poner en contacto con el cliente y ofrecerle información específica que afecte a sus rutas de viaje.

Al conocer también que tipo de cliente es y el tipo de productos que busca, podemos personalizarle ofertas o descuentos de otros productos que se adecuen a su perfil de cliente para la empresa y así conseguir un incremento de las ventas cruzadas.

- Optimización de recursos en el marketing y otras áreas de la empresa:

Podemos determinar el gasto óptimo y determinar una mejora continua a través de diferentes análisis, pruebas y mediciones para conseguir la optimización del coste en marketing para la empresa.

También podemos mejorar en otras partes de la empresa para optimizar la cadena de suministro como por ejemplo ha hecho la cadena de supermercados Wal-Mart para mejorar los sistemas de gestión de inventarios por la demanda actual de sus consumidores. En el caso de Wal-Mart, los datos que recogen se encuentran principalmente en los tickets de compra y lo que ha hecho es conectar a través de estos datos sus almacenes con sus proveedores a través de un software en el que a tiempo real los proveedores ven cuando escasea su producto en el inventario y tienen que reponer existencias.

A continuación con breve y sencillo ejemplo práctico [para más detalles, Siegel, E. (2013) p.p. 44-45] vamos a demostrar lo beneficioso que puede ser para una empresa aplicar este tipo de herramientas en el sistema de negocio.

Supongamos que tenemos una base de datos de clientes con un millón de clientes, enviar cada correo electrónico nos cuesta un euro y medio, y sabemos que de cada cien uno nos comprara el producto ( $1.000.000/100=10.000$ ). Por tanto calculemos el beneficio que supondría dicha operación sabiendo que el precio de venta del producto es de 170 euros.

Beneficio total =  $(170 \times 10.000) - (1,50 \times 1.000.000) = 200.000$  euros de beneficio.

Aplicando herramientas de análisis multivariante en la base de datos averiguamos que una cuarta parte de esos clientes tiene el triple de probabilidades de comprar el producto (tres de cada cien) sabiendo que por el tipo de perfil de cada cliente que cuarta parte de la base es. ¿Qué pasaría entonces? ¿Cómo modificaría en este caso a nuestro resultado el aplicar la minería de datos en la estrategia de la empresa?

Beneficio total =  $(170 \times 7.500) - (1,50 \times 250.000) = 900.000$  euros.

Aunque hayamos enviado correos a menos personas, hemos multiplicado nuestro beneficio por 4,5. ¿Cómo? Haciendo uso de la minería de datos. Esto mismo es lo que

están haciendo infinidad de empresas en la actualidad para aumentar sus cifras de negocio.

Y no solo empresas privadas, sino públicas. Con este tipo de análisis, el cuerpo de policía de Santa Cruz (California) puede saber en qué barrios hay más probabilidad de que se cometan delitos y enviar de esta forma a más policías a patrullar estas zonas a través de los datos históricos de los dos años anteriores. Iniciativa que fue galardonada como uno de los 50 mejores inventos por la revista Time [Siegel, E. (2013), página 82].

Este tipo de análisis estadístico nos sirve para dirigir el presupuesto de marketing de la empresa de forma óptima para el segmento de clientes más correcto, aquel segmento que más rentabiliza la inversión en el marketing.

Esto también aporta una serie de problemas o retos que tenemos que hacer frente para la utilización de la información obtenida de las grandes bases de datos en las estrategias de marketing de la empresa:

- En un campo en el que podemos recolectar infinidad de datos, tenemos que seleccionar aquellas variables y datos que son relevantes, útiles e importantes para la consecución de los objetivos de la empresa. Nos encontramos en una situación con infinidad de datos, y tenemos que recolectar y seleccionar aquellos que nos puedan dar información útil y valiosa.
- Hay que saber que técnica de análisis tenemos que utilizar para obtener el tipo de información que estamos buscando. En la solución de este problema es en lo que voy a hacer referencia más adelante.
- El coste de la infraestructura física necesaria para el mantenimiento y la interpretación de grandes volúmenes de datos (cada vez más desestructurados) supone un gran problema para empresas pequeñas que no pueden costearselo. La solución a este problema son las nuevas empresas proveedoras de almacenamiento de datos o de venta de los mismos que actualmente tienen mucho potencial de mercado. Algunas ofrecen a sus clientes espacios en la nube, desde donde estos pueden conseguir resultados rápidos y de forma accesible. Un ejemplo de este tipo de empresas es SAP.
- La información que se obtiene de la minería de datos no solo tiene que ser utilizada por el departamento de marketing. El conocimiento derivado del data

mining tiene que ser conocido por todos campos de la empresa a los que pueda serle útil para mejorar sus acciones.

En el apartado siguiente vamos a desarrollar y explicar cinco tipos de análisis estadísticos diferentes para saber qué tipo de análisis tenemos que utilizar según los objetivos esperados para que nos den información útil y provechosa de la que podamos obtener conocimiento para una toma de decisiones optima dentro de la empresa.

Así pues, como hemos comentado en la introducción el objetivo de este trabajo es desarrollar el gran interés que para las empresas tiene el tratamiento de la información para generar conocimiento y valor, en particular en el campo del marketing y la investigación de mercados.

La memoria de este trabajo la hemos estructurado en cuatro partes, después de la presentación, que nos ha servido como marco teórico para situar el problema, estudiaremos los principales tipos de métodos multivariantes más adecuados en cada problema, para consecuentemente entrar en profundidad en el análisis de un panel real llevado a cabo a nivel mundial en empresas de distintos sectores, y acabaremos con las conclusiones finales del trabajo.

## **PARTE 2: HERRAMIENTAS DE ANÁLISIS**

En una situación en la que tenemos gran cantidad de datos, tenemos que saber cuáles son los métodos más efectivos para analizarlos, y por consiguiente el tipo de resultado que esperamos obtener de ellos para poder llegar a conclusiones y decisiones óptimas para la empresa.

A continuación vamos a explicar cinco métodos de análisis multivalente, y el tipo de resultados que podemos obtener de dicho análisis mediante sencillos ejemplos. Hemos introducido, muy brevemente, estas técnicas para que la memoria fuese auto-contenida.

### **2.1. TABLAS DE CONTINGENCIA Y CORRESPONDENCIAS**

Con este análisis podemos determinar perfiles y/o relaciones de consumidores de un producto o servicio a través de tablas de contingencia. Así pues, esta técnica estudia la relación entre dos características observadas en una muestra, usualmente características demográficas (sexo, edad, salario, localización, estudios, situación laboral,...) y otras de compra del producto de interés. Teniendo en cuenta que el desarrollo del método consiste en realizar un Análisis de Componentes Principales sobre la nube de puntos – fila y puntos-columna, se utiliza la métrica Chi- cuadrado, para recalcular distancias entre filas y columnas. Dicha métrica permitirá medir la inercia de la nube de puntos respecto de sus centros de gravedad. Así, lograremos resumir e interpretar las variables interrelacionadas, ver cómo se sitúan a través de sus distancias y, en definitiva, definir el factor o factores más influyentes y los más relacionados (en el caso de que la variable seleccionada sea de escala habrá que codificarla según el interés del investigador).

Los principales pasos de este análisis son los siguientes:

→ Tablas de contingencia: con este tipo de tablas a través del recuento asociado entre las distintas variables obtenemos de la frecuencia esperada calculada.

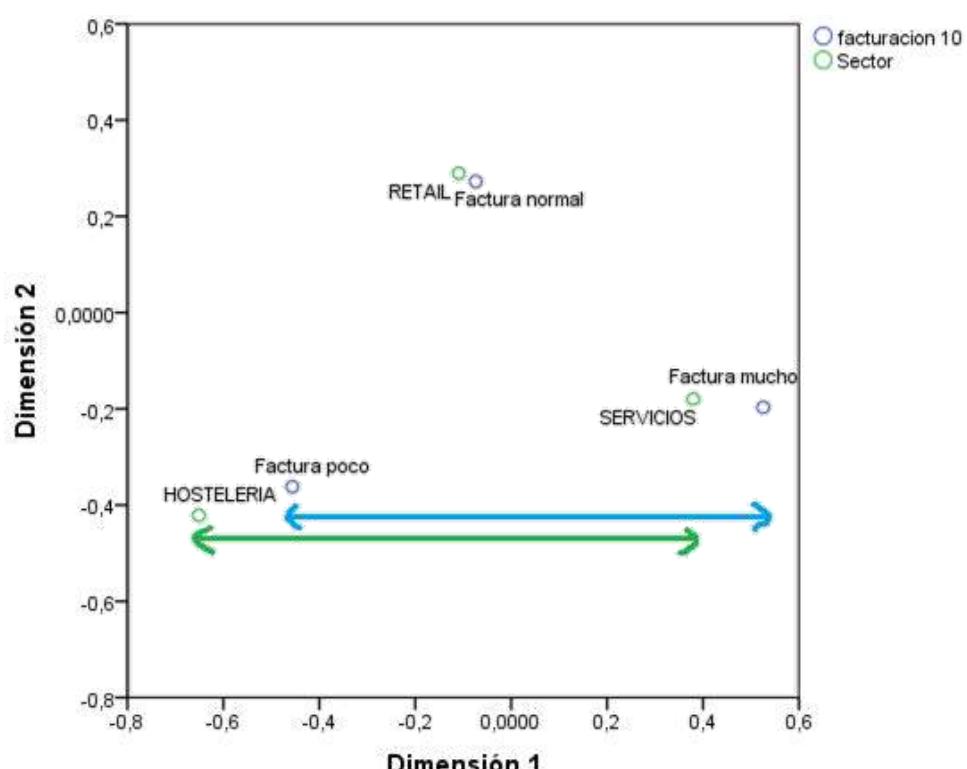
Así obtendremos los residuos estandarizados de cada casilla, si los residuos son superiores a +1,96 hay más recuentos de los que se esperaban, por lo que nos indica por ejemplo la existencia de un grupo de empresas (sector acero que trabajan en mercados internacionales) que compra más software informáticos de lo que se espera por la frecuencia esperada. Esto es lo que denominamos un perfil, esto nos daría un primer indicio para un posterior análisis y por consiguiente enfocar nuestro marketing de la empresa a este tipo de cliente .En caso contrario, si los residuos son inferiores a -1.96, nos indica la existencia de

pocas empresas realizando esa compra, podríamos también afirmar que la compra no se da en empresas que trabajan en mercados regionales. Las conclusiones de marketing que destacamos de este análisis, irán dirigidas a los distintos perfiles que identifiquemos.

Así pues, conoceríamos un primer paso para posteriores análisis de segmentos de consumidores a los que habría que enfocar nuestras acciones de marketing.

- Perfil fila/perfil columna: es un tipo de análisis que en el caso de que nos hayan salido diferencias significativas dentro de la tabla de contingencias nos será de mucha utilidad ya que las podremos ver claramente tanto en tabla como en gráficos. Diferenciando en fila o columna según queramos analizar el perfil del individuo en función de las variables de clasificación que estén colocadas por filas (perfil fila) o las variables colocadas por columnas (perfil columna).
- Diagrama de asociación: Con este tipo de análisis se puede crear un diagrama basado en la asociación de las diferentes variables, en el que se representan las distintas variables analizadas y la proximidad entre estas variables indica el grado de asociación que hay entre ellas .Un ejemplo de este tipo de gráfico que acabo de enunciar es el siguiente:

**Ilustración 2. Ejemplo de análisis de correspondencias.**



Fuente: Elaboración propia

En este ejemplo podemos apreciar la relación de las distintas variables, siendo que en cuanto a la relación entre el tipo de sector y la cantidad de facturación es la siguiente: el sector servicios es el que más factura, el sector retail factura una cantidad intermedia y el sector de la hostelería es el que menos factura. En el apartado tres de este proyecto, veremos su aplicación concreta en la ENCUESTA EUROPEA ANALIZADA.

## **2.2. ANÁLISIS FACTORIAL**

Es una técnica de reducción de la dimensión de variables, y se centra en encontrar unos factores no observables, que el analista definirá teniendo en cuenta las correlaciones entre variables observadas. Busca el número mínimo de dimensiones (factores) para explicar la mayor parte de la información de un modelo. Este análisis estadístico analiza las interrelaciones entre las variables y crea grupos homogéneos de variables que se correlacionan mucho entre sí. El propósito fundamental del Análisis Factorial es simplificar la información. Lo que nos es muy útil con bases de datos con gran cantidad de variables. Dentro de un análisis factorial, hay que diferenciar entre dos fases: un análisis exploratorio para determinar el número de factores y posteriormente un análisis confirmatorio que se centra en implementar y contrastar el número de factores que el análisis previo el analista decide en la exploración de la gran base de datos [véase Hair, J. et al., 2014, nueva edición corregida].

Las fases del análisis exploratorio son,

- Test de Kaiser-Meyer-Olkin y test de esfericidad de Bartlett: son dos indicadores de la idoneidad del modelo. El KMO es un indicador que compara las magnitudes de las correlaciones observadas con las parciales. Los valores de este indicador oscilan entre 0 y 1, siendo valores próximos a 0 un indicador de no adecuación del modelo (KMO indica no correlación entre variables y rechazamos el modelo factorial) y valores próximos a 1 (KMO indica correlación de variables, y a mayor mejor aceptación del análisis) indicador de adecuación del modelo.

El test de esfericidad de Bartlett contrasta como hipótesis cero que la matriz de correlaciones es una matriz identidad, lo que se traduce en la ausencia de correlación entre las distintas variables. Para plantear un modelo de correlaciones, y en definitiva un modelo factorial, el estadístico del test, tiene que dar un p-valor menor que 0.05 para valores peores que ese umbral.

- Comunalidades: este indicador nos indica la proporción de la varianza explicada de cada variable a través del modelo factorial. Este indicador se encuentra entre 0 y 1. Valores de comunalidades próximos a uno indican que la varianza de esa variable está prácticamente explicada en los factores comunes determinados por el modelo de análisis factorial mientras que si se aproxima a cero, la varianza explicada por el modelo implica un mal ajuste para esa variable dentro del modelo y habrá que suprimirla.
- Matriz de componentes rotados: indica el peso que tiene cada variable dentro de cada factor. Los valores están entre cero y uno. Las variables que tengan en un factor un valor próximo a uno, o alto son indicadoras de que esa variable tiene una estrecha relación con el factor. De esta forma indicamos la relación de los factores con las distintas variables. Esta matriz, es prácticamente una matriz de correlaciones factor/variable. Podemos entender esto con el ejemplo siguiente:

**Tabla 1. Ejemplo de matriz de componentes rotados en análisis factorial.**

<b>Matriz de componentes rotados</b>		<b>Componente</b>		
		<b>F1</b>	<b>F2</b>	<b>F3</b>
Satisfacción con el precio		,945	,035	,236
Satisfacción con el producto		,970	,083	,114
Volumen del gasto		,411	-0,11	,794
Nivel de ingresos de cliente		,054	,094	,930
Satisfacción con la entrega		,851	,287	,162
Antigüedad en la empresa		,049	,906	-0,055
Tasa de recompra del cliente		,203	,875	,160

F1= Satisfacción del cliente

F2= Antigüedad del cliente y tasa de recompra

F3= Gasto del cliente e ingresos

Fuente: Elaboración propia.

- Varianza total explicada del modelo: indica la suma total de varianzas que explica el modelo según cuantos factores seleccione el investigador. De esta forma podemos elegir el número de factores óptimo para el análisis exploratorio.
- Gráfico: utilizamos el gráfico de puntos de dispersión utilizando como variables los factores que nos han resultado, para así analizar la muestra y encontrar atípicos dentro de ella. Los atípicos son casos en los que por sus características destacan en el modelo frente al resto de casos.

En la base de datos que presentamos en el apartado 3 de esta memoria de grado detallaremos el procedimiento de forma más práctica.

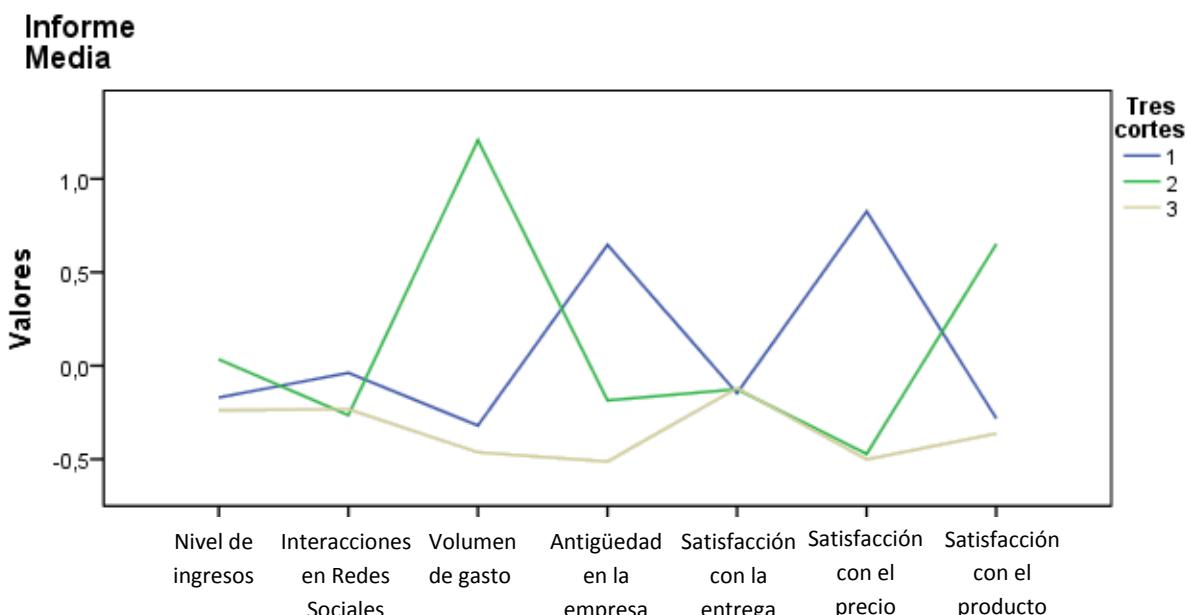
### **2.3. ANÁLISIS DE CONGLOMERADOS O CLUSTER**

Es una técnica diseñada para clasificar diferentes observaciones, tratando de lograr la mayor heterogeneidad entre los distintos grupos y la mayor homogeneidad entre los miembros del mismo grupo. Es de destacar que la composición de los grupos es desconocida a priori, y se definen a partir de las observaciones. En esto último radica la gran diferencia con el análisis discriminante, ya que en éste los grupos son previamente conocidos por el analista.

Este tipo de análisis lo utilizamos para crear grupos dentro de una infinidad de elementos y poder así comparar entre si las características y diferencias de los mismos. Con este análisis podemos clasificar nuestra base de clientes en diferentes tipos, uniéndose en el mismo grupo aquellos con características de consumo más similares:

- Tipificar las variables de diferente escala: En primer lugar, a la hora de realizar este análisis estadístico, tenemos que fijarnos que todas las variables que introduzcamos tengan la misma escala de medida, ya que si no es así, cualquier distancia calculada a partir de esas variables nos llevaran a resultados poco comprensibles. Por lo que si no están en la misma escala, tendremos que estandarizar previamente las variables para poder considerar bueno el análisis.
- Dendograma: es un tipo de gráfico que obtenemos de este análisis y es un diagrama de tipo árbol en el que se pueden apreciar la unión de los diferentes grupos. La posición de la línea vertical, indica a qué distancia se han unido los diferentes grupos. Grupos unidos a distancias lejanas indican grupos creados de forma ficticia ya que estos no tienen apenas nada en común. Y es el propio analista quien debe determinar donde realiza el corte, a la vista del gráfico, del número de grupos (o conglomerados) óptimo. En el caso de encontrar algún atípico, lo recomendable es suprimirlo del análisis para que de esta forma concluyamos con un análisis más eficaz. A veces resulta recomendable un estudio de las frecuencias de cada grupo para que resulten de tamaños similares.
- Informe: una vez tenemos los grupos claramente definidos, podemos mediante un gráfico de líneas, realizado a partir de las medias de los diferentes grupos tipificados de cada una de las variables que hemos definido en el análisis, ver estas diferencias para así poder establecer mejoras o atención especial a aquellos grupos de clientes por ejemplo que estén en peor satisfechos y ver así que podemos mejorar con cada grupo de clientes.

**Ilustración 3. Ejemplo de gráfico en informe de clasificación.**



Fuente: Elaboración propia.

En el ejemplo anterior, podemos apreciar tres grupos de clientes:

- Grupo uno (Azul): es un conjunto de clientes con un nivel medio de ingresos, que son los que más interaccionan con la empresa en redes sociales aunque no son los que más gastan ni tampoco los que menos en productos o servicios de la empresa. También podemos ver que son los clientes más antiguos de la empresa, y en cuanto a niveles de satisfacción, destacar que están muy satisfechos con el nivel de precios.
- Grupo dos (verde): son los clientes con renta más elevada, y los que menos interaccionan con la empresa en las redes sociales aunque tienen un gasto muy elevado en los productos o servicios (estudiaríamos a posteriori si esto se debe a una correlación con su elevado nivel de renta). En cuanto a la antigüedad en la empresa, es media, y en satisfacción es destacable la satisfacción con el producto.
- Grupo tres (amarillo): es el grupo de clientes con menos recursos, que interaccionan muy poco (igual que el grupo dos) con la empresa en redes sociales y son los que menor gasto tienen dentro de la empresa. También cabe destacar que son los clientes más recientes, y en satisfacción son los que menos satisfechos están con las variables estudiadas.

Esta herramienta multivariante nos es muy útil cuando queremos separar una muestra en diferentes grupos, en la cual buscamos conocer que características (demográficas,

económicas, descriptivas...) tienen los integrantes de los grupos en común y las diferencias que se pueden observar entre ellos [para información más detallada de este análisis Hair, J. et al (2014) p.p. 415-472].

## **2.4. ANÁLISIS DE REGRESIÓN CON VARIABLE DEPENDIENTE CATEGÓRICA**

Muchas de las relaciones entre variables de interés se centran en que la variable dependiente es categórica. En este contexto uno de los modelos más utilizados es el modelo de regresión logística (logit).

Este modelo, no es lineal, puesto que la variable dependiente es categórica y por esto, las variables explicativas determinan la acción o no de la variable categórica.

Con este tipo de análisis podemos: utilizar probabilísticamente el efecto que tienen en la variable dependiente diferentes variables explicativas, podemos seleccionar dentro de todas estas variables las más significativas para el análisis y estimar el “riesgo relativo” de las mismas.

Según indica Guisande González, C. y otros (2011) (página 537), este es el tipo de análisis más adecuado para variables cualitativas. La variable dependiente siempre tiene que ser cualitativa, pero las explicativas o independientes pueden ser categóricas (dicotómicas o múltiples), continuas, o mezcla de todas ellas. Por lo que nos permite una gran comparación con cualquier tipo de la que dispongamos en nuestra base de datos, para ver si está o no relacionada con la variable dependiente. En nuestro estudio, vamos a hacer referencia exclusivamente al análisis binomial, para más información sobre el análisis de regresión multinomial consultar el texto antes citado.

En primer lugar hay que explicar que es el riesgo relativo. Es una medida que indica la probabilidad de actuación de un grupo de individuos/datos/observaciones que está expuesto a un factor de riesgo frente al grupo de individuos/datos/observaciones que no lo está.

Con este dato, por tanto, podemos obtener la probabilidad de que ocurra un suceso (acción de compra) de aquel usuario con determinadas características frente al usuario que no las cumple. En el contexto de la base de datos a utilizar es de interés investigar si la utilización de determinados elementos y/o software/hardware de TICs tienen una

influencia en el crecimiento de la empresa (facturación o cuota). Eso es lo que trataremos de ver en la parte empírica de este proyecto.

El riesgo relativo, si diera como resultado 1, indicaría que la presencia del factor de riesgo no modifica la probabilidad de acción, es decir, que el ser adquirir un software o no, no indica una probabilidad diferente según la acción de adquisición. Sin embargo, si el resultado es mayor que 1, indica que hay una asociación positiva, ya que la presencia del factor de exposición indica una mayor probabilidad de que suceda la acción (incremento de facturación), mientras que si es menor que 1 indica lo contrario (asociación negativa), es decir, que el adquirir influya negativamente la probabilidad de crecimiento de la empresa [para más detalles, Hair, J. et al, (2014) p.p. 321-322, donde vemos la relación entre el riesgo relativo (odds ratio) y logodds (logit)].

Para este tipo de análisis, es necesario tener muchas observaciones, y la interpretación de los resultados depende de la codificación que el investigador haya considerado en las variables a priori.

A la hora de plantear el modelo, lo utilizamos simultáneamente con distintas variables para seleccionar las más significativas, por lo que según el número y el tipo de variables que se seleccionen la probabilidad de acción del suceso será diferente.

## **PARTE 3: DATOS Y METODOLOGIA**

### **3.1. DATOS**

La encuesta se encuentra disponible en la página web: [www.ebusiness-watch.org](http://www.ebusiness-watch.org)

La base de datos que vamos a utilizar para la parte empírica del proyecto es un Open Data, aunque no de fácil tratamiento.

La base utilizada corresponde a la ola del año 2.007 y tiene una muestra de 2.121 casos (muchas de las variables tienen un porcentaje de datos faltantes no despreciable, pero no llega a ser imprescindible la imputación). En la que utilizamos, solo tenemos los datos de empresas, de varios sectores (Químico, Equipamiento y Acero) de 8 países (Alemania, España, Francia, Italia, Suecia, Reino Unido, Polonia y Estados Unidos) de los 34 países que aparecen en la encuesta global.

Se estructura de la siguiente forma:

- Pre-encuesta: Información de la base de datos del panel anterior.
- A: Infraestructuras TICs y sistemas de software e-Business
- B: Intercambio de datos automatizado
- C: Normas electrónicas y cuestiones de inter-operatividad
- D: Actividades de innovación de la compañía
- E: Necesidades de competencias TICs y costes.
- F: Impactos, conductores e inhibidores de las TICs
- G: Información básica sobre la competencia.

Las variables seleccionadas (teniendo presente el objetivo último del trabajo en el ámbito del marketing) para este estudio las podríamos dividir en tres grupos distintos:

- Las que hacen referencia al tipo de empresa (cualitativas ,en su mayoría)
- Las que analizan las TICs.
- Las referentes al apartado ventas.

### 3.2. SELECCIÓN DE VARIABLES

#### Variables que hacen referencia al tipo de empresa:

*G1: How many employees does your company have in total in [country], including yourself? (Cuantitativa)*

*G1a: Would you be able to tell me to which of the following size groups your company belongs? (Categórica)*

- |               |                                     |
|---------------|-------------------------------------|
| 1. = “1-9”    | 4. = “250-999”                      |
| 2. = “10-49”  | 5. = “1000+”                        |
| 3. = “50-249” | 6. = “DK (Doesn’t know)/ no answer” |

Para la **variable tamaño (G1 y G1a) de la empresa en función del número de empleados** he creado una nueva variable que voy a definir con el termino tamaño proximado (de aproximación), ya que en la base de datos G1 tiene un porcentaje muy elevado de datos missing, en ese caso la empresa al no saber la respuesta exacta, responde G1a (respuesta categorizada). Nos ha parecido más conveniente al no ser un número excesivo de faltantes utilizar la respuesta categorizada para hacer una Proxi de la G1, utilizando para la estimación de ese dato el último año dentro de cada categoría (G1a).

Esta variable viene dada en la base de datos en dos distintas, por una parte, tenemos el dato tamaño de la encuesta anterior (**z2a y z2b**) y por otra parte, la respuesta del periodo actual (**G1 y G1a**). Esta última es la observación que vamos a utilizar como estimación para la nueva variable ya que tenemos menos datos perdidos.

*G2: When was your company founded, I mean in which year was it formally registered for the first time? (cuantitativa)*

*G3: Would you be able to tell me in which of the following periods your company was founded? Was it ...? (categórica)*

- |                          |                          |
|--------------------------|--------------------------|
| 1. before 1982           | 4. between 2004 and 2007 |
| 2. between 1982 and 1997 | 5. DK / no answer        |
| 3. between 1998 and 2003 |                          |

Respecto a la **variable antigüedad (G2 y G3)** de las diferentes empresas, también utilizaremos el concepto proximado para su creación, ya que como en el caso de la variable

tamaño, nos encontramos empresas con el dato cuantitativo y otras con el dato categórico. Igual que en el caso anterior, utilizaremos para la estimación de faltantes dentro de la variable cuantitativa el menor número de empleados dentro de cada categoría.

*G4: What is your company's most significant market? Is it MAINLY the regional market, the ... market, or international markets which you consider your main sales area? (categórica)*

*G11: Please estimate the percentage share of employees with a college or university degree in your company. (Cuantitativa)*

Conocer el tipo de cualificación que tienen los empleados nos da información del tipo de empresa que es (instituto de investigación, cadenas de montaje...)

*Z1b: Survey Sector Number / Name (Categórica)*

- |                                      |                  |
|--------------------------------------|------------------|
| 1. = “Chemical, rubber and plastics” | 2. = “Steel”     |
|                                      | 3. = “Furniture” |

Hace referencia al sector que trabaja la empresa, nos es muy útil para clasificar las empresas y obtener conclusiones para algún sector destacable.

### **Variables que analizan las TICs:**

*A2: What percentage of your employees have access to the internet at their workplace?*

(cuantitativa)

*A3: What is the maximum bandwidth of your internet connection available for download? Is it less than 144 kilobit per second, between 144 kilobit and 2 megabit per second, or more than 2 megabit per second? (Categórica)*

- |  |  |
|--|--|
| 1. = less than 144 kilobit per second (kb/s)               | 3. = more than 2 megabit per second (Mbit/s) |
| 2. = between 144 kilobit and 2 megabit per second (Mbit/s) | 4. = DK                                      |

*A7: Do you use an ERP system, that is Enterprise Resource Planning (Dicotómica)*

Los sistemas de Planificación de Recursos de la Empresa integran la gestión de algunos sistemas de gestión de la empresa como: ventas/compras, contabilidad, logística, producción, inventarios, control de almacenes, nominas...

Son software generales que optimizan las partes de funcionamiento básico de la empresa.

*A7: Do you use a Supply Chain Management system*

Sistemas SCM son aquellos software que sirven para optimizar la cadena de suministros, desde el proveedor hasta el consumidor final, para coordinar la oferta y la demanda de productos de manera eficiente.

*A7: Do you use a CRM software system, that is a customer relationship management system*

El CRM es un modelo de gestión enfocado a la relación con los clientes (venta relacional). Engloba al conjunto de herramientas que gestionan esa relación, así como aquellos que persiguen la maximización de la satisfacción del consumidor, para establecer relaciones duraderas en el largo plazo. Es muy importante en marketing (marketing relacional).

*A7: Do you use a DMS software system, that is a document management system for storing and archiving electronic documents*

El DMS es un sistema de gestión documental en el que se optimizan grandes cantidades de documentos informáticos.

Un ejemplo de este tipo de base de datos es el utilizado por la Universidad de Zaragoza para la gestión del fondo de biblioteca.

*A7: Do you use a PLM software system, that is software for product lifecycle management?*

El software PLM, permite la gestión eficiente del ciclo completo de vida del producto, desde la concepción del prototipo hasta la puesta en marcha y retirada.

*B1: Does your company use the internet or other computer-mediated networks to order goods or services from suppliers online, not counting manually typed e-mails?*  
(Dicotómica)

*B2: Please estimate the percentage of orders your company places online. Would you say... (Categórica)*

- |                   |                    |
|-------------------|--------------------|
| 1. = less than 5% | 4. = 26 up to 50%  |
| 2. = 5 up to 10%  | 5. = more than 50% |
| 3. = 11 up to 25% | 6. = DK            |

*B3: Can customers order goods or services from your company online on the internet or through other computer-mediated networks? (Dicotómica)*

*B4: Please estimate the percentage of orders you receive online. Would you say... (categorical)*

- |                   |                    |
|-------------------|--------------------|
| 1. = less than 5% | 4. = 26 up to 50%  |
| 2. = 5 up to 10%  | 5. = more than 50% |
| 3. = 11 up to 25% | 6. = DK            |

*B9: Does your company share information on inventory levels or production plans electronically with business partners? (Dicotómica)*

*B11: All in all, which of the following statements best describes the way your company exchanges data with business partners: (Categórica)*

- |                                   |  |
|-----------------------------------|--|
| 1. = mostly electronically        | 3. = electronically processed, but exchanged in paper based format |
| 2. = mostly in paper based format | 4. = DK  |

*E2: How many ICT practitioners does your company employ in ...? (Cuantitativa)*

*E6: In the forthcoming financial year, do you expect your company's ICT budget to increase, decrease or will it roughly stay the same? (Categórica)*

- |             |                          |
|-------------|--------------------------|
| 1. Increase | 3. roughly stay the same |
| 2. Decrease | 4. DK                    |

*E9: Did your company make investments in ICT during the past 12 months, for example for new hardware, software or networks? (Dicotómica)*

*F6: Do you expect that ICT will have a high impact, medium impact, low impact or no impact on the following business functions in your company in the future? What about*

*management and controlling? (Categóricas pero las emplearemos como cuantitativas en los análisis factorial y cluster)*

*What about administration and accounting?*

*What about research and development?*

*What about production processes?*

*What about marketing?*

*What about logistics?*

*What about customer service?*

- |                      |   |
|----------------------|---|
| 1. = “High impact”   | 4. = “no impact”  |
| 2. = “Medium impact” | 5. = “Not relevant - Don't have this business function” |
| 3. = “Low impact”    | 6. = “DK”   |

**Variables referentes al apartado ventas:**

*G6: How large is the market share of your company in the ... market?.* (cuantitativa)

*G7: Has the share of your company in this market increased, decreased, or stayed roughly the same over the past 12 months?* (categórica)

*G9: Has the turnover of your company increased, decreased or stayed roughly the same when comparing the last financial year with the year before?*(categórica)

### 3.3. EXPLORANDO LOS DATOS

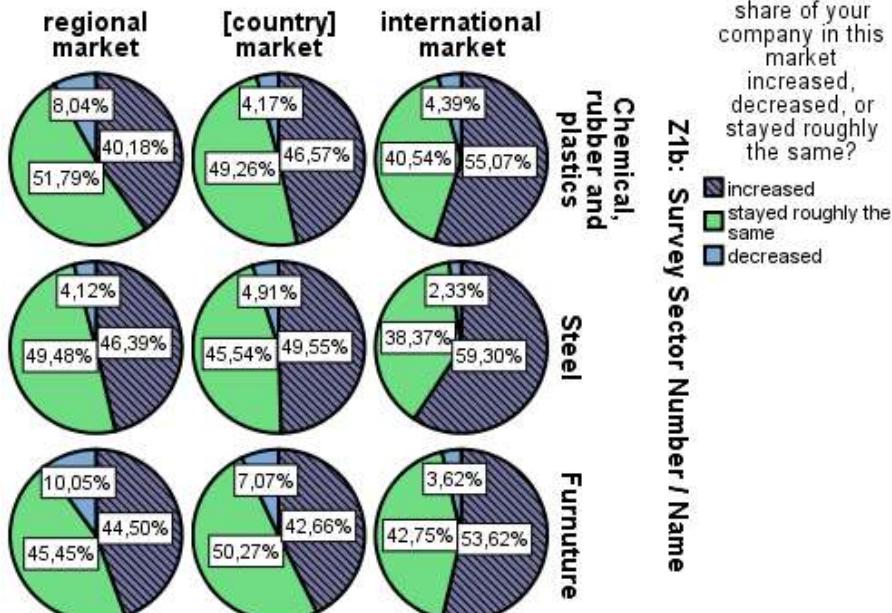
Según el tipo de sector y tipo de mercado al que dirigen la mayor parte de sus ventas (regional, nacional e internacional) no hay aparentes diferencias significativas en el porcentaje de empresas que han aumentado/mantenido su cuota de mercado en todos los sectores. Lo que hace pensar en la introducción de nuevas tecnologías de la comunicación para aquellas empresas cuyos mercados son locales o nacionales para que hubiera una proporción mucho mayor de esas empresas en el aumento de cuota de mercado. Puesto que son mercados en los que observamos menor porcentaje de empresas que han aumentado su facturación en comparación con las internacionales y la introducción de TICs podría ser un factor relevante a la hora de aumentar la facturación optimizando procesos de la empresa.

Hemos comprobado que en el sector “químico, caucho y plástico” el porcentaje de envíos de pedidos online es muy superior comparado con los otros sectores. Sospechamos que en ese sector están incluidas muchas empresas de I+D+i, ya que en el encontramos las empresas con mayor cualificación de empleados del mercado. Consecuentemente, estas empresas suelen trabajar con el máximo ancho de banda.

Una investigación interesante en el contexto del marketing es conocer la utilización que hacen las empresas de TICs (por sectores y/o mercado), así como los tipos de banda en que trabajan y su impacto en la cuota de mercado y facturación, además de analizar como realizan el intercambio de datos (papel, electrónica...).

Ilustración 4. Gráfico sectorial según tipo de mercado, cuota y sector de actividad.

G4: What is your company's most significant market?



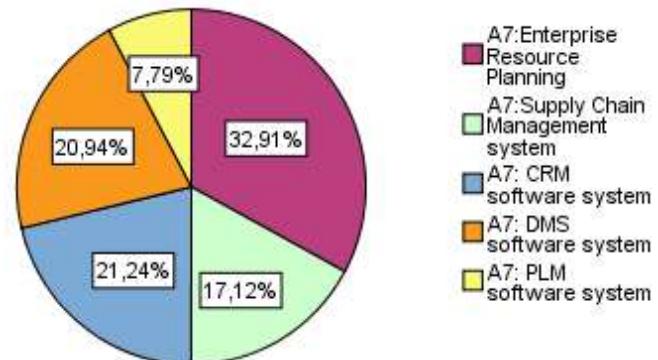
Fuente: Elaboración propia.

Para ello, realizaremos los gráficos sectoriales por fila y columna así como perfiles para ahondar más en el tipo de estrategia a realizar.

Como podemos apreciar en el gráfico de la derecha, el tipo de software más utilizado por las empresas es el EPR, esto podemos suponer que se debe a la generalidad de uso de la herramienta dentro de la empresa. Y por el contrario, encontramos en último lugar al PLM, lo cual puede deberse a la especialización del programa.

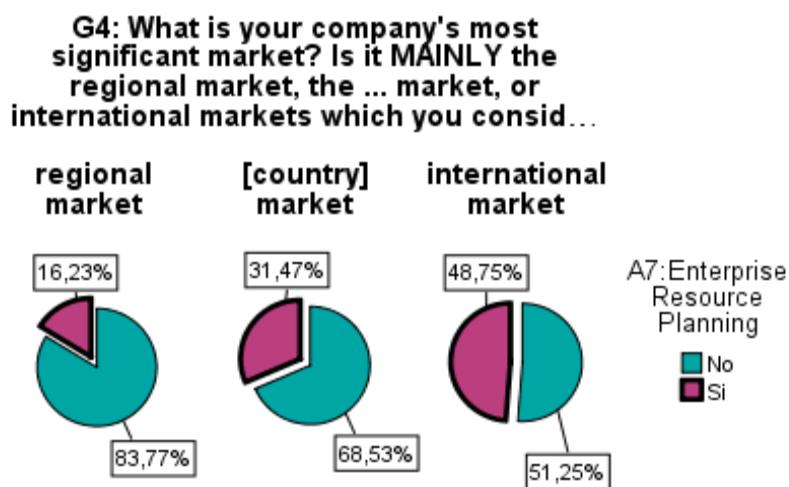
Además existen diferencias significativas aparentes entre el uso de estas herramientas según el tipo de empresa. Podemos afirmar, que para los cinco software distintos que analizan la encuesta, el sector “químico, plástico y caucho” es el que más los utiliza. Sería interesante indagar un poco más en este sector en los análisis que veremos a continuación, y localizar un nicho de mercado potencial en él.

Esto mismo sucede según el tipo de mercado al que dirige la empresa la mayor parte de sus ventas. Conforme aumenta la distancia geográfica entre la empresa y cliente o proveedor (regional→nacional→internacional), aumenta el uso de estos productos.



Fuente: Elaboración propia.

**Ilustración 6. Gráficos sectoriales según software ERP y tipo de mercado.**



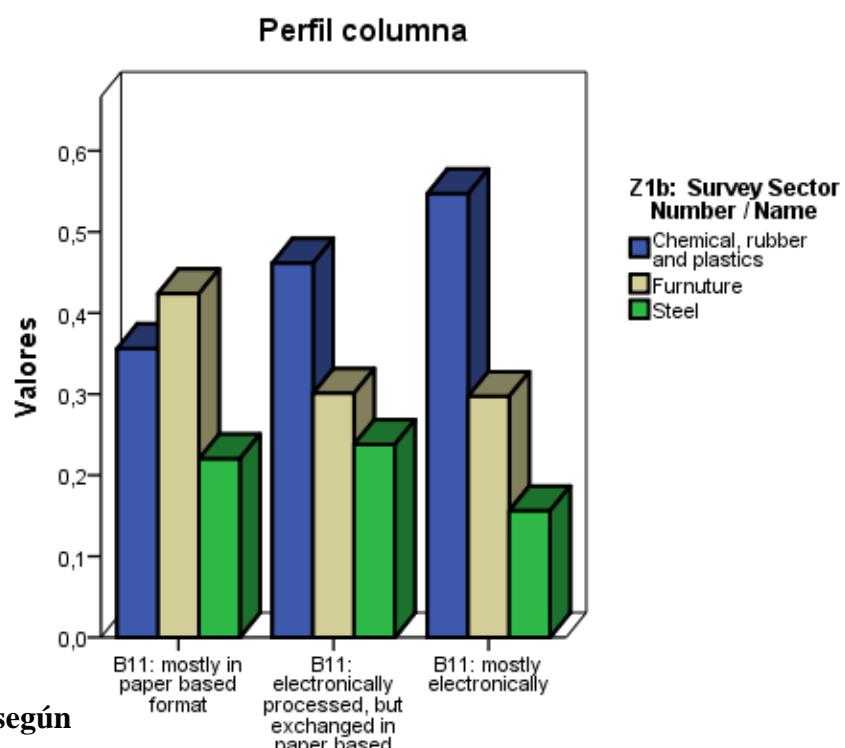
Fuente: Elaboración propia.

### 3.4. ANÁLISIS DE CORRESPONDENCIAS

Como hemos sugerido antes para detectar perfiles en los tipos de empresa vamos a realizar un análisis de correspondencias. A la hora de llevar a cabo el análisis de correspondencias para encontrar diferentes perfiles de empresas, nos vamos a centrar en la variable que indica la forma de transferencia de datos, que de alguna manera tiene bastante relación con la utilización de las TICs dentro de la empresa, que es lo que se pretende con este trabajo (en todos los gráficos de correspondencias a los que se hace referencia en este trabajo el p-valor del test de independencia es, obviamente, menor del 5%).

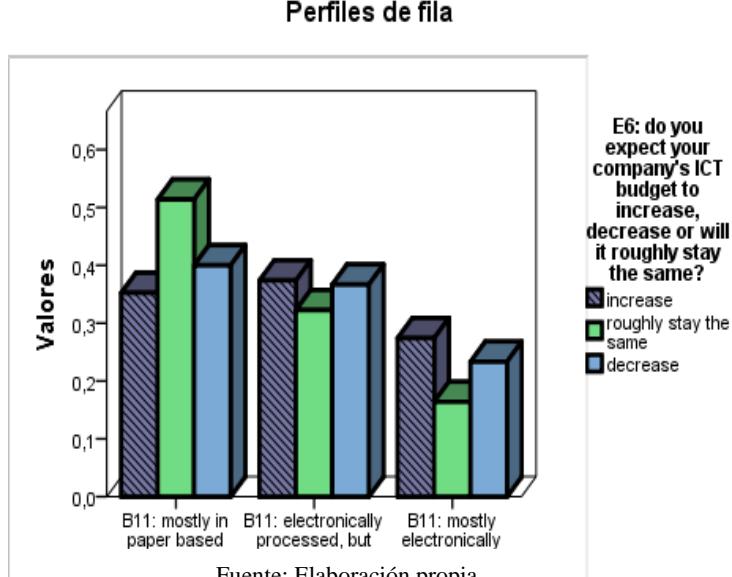
**Ilustración 7. Perfil columna: intercambio de datos según tipo de sector.**

Podemos apreciar en los gráficos siguientes, que las empresas del sector “químico, caucho y plástico” son las que el intercambio de datos lo hacen mayoritariamente de forma electrónica, lo cual viene a corroborar nuestra sospecha razonable de que es un sector en el que se desarrolla I+D+i y las que menos las que trabajan con el acero. (Ilustración 7)



Fuente: Elaboración propia.

**Ilustración 8. Perfil fila: intercambio de datos según variación del presupuesto.**



Fuente: Elaboración propia.

**E6: do you expect your company's ICT budget to increase, decrease or will it roughly stay the same?**

- increase
- roughly stay the same
- decrease

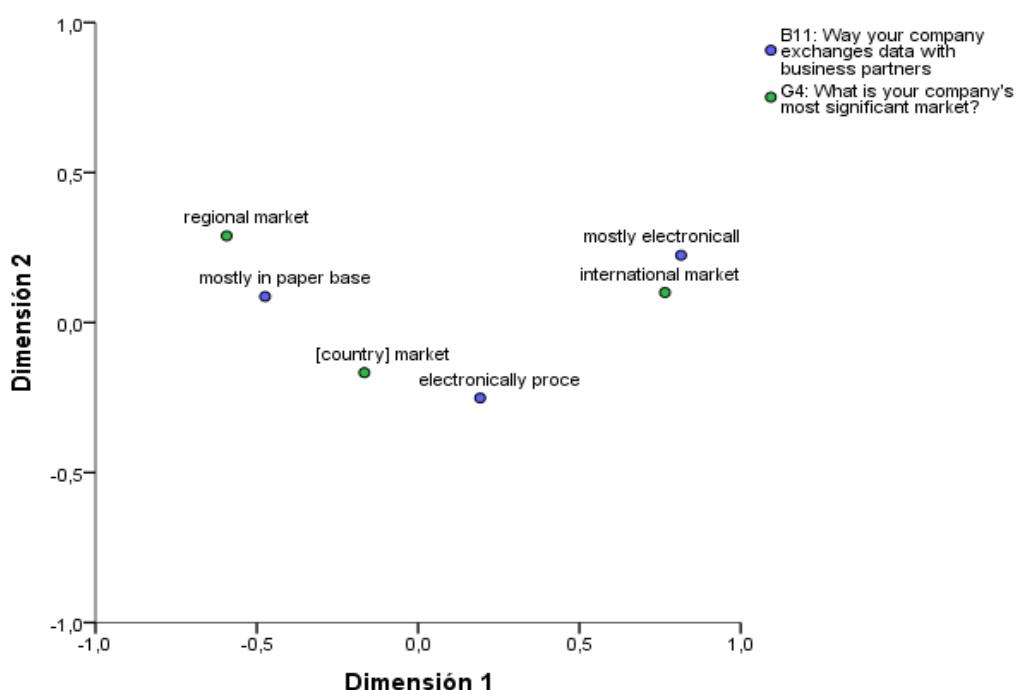
De la misma forma hemos podido constatar que las empresas que reciben más de un 25% de pedidos online son las que tienen implementadas TIC en su empresa, de la misma forma, las que incrementan su presupuesto en TIC son aquellas que el intercambio de datos no lo hacen en papel (Ilustración 8).

En otro orden de ideas, respecto a la importancia del Marketing en los diferentes departamentos de una empresa y la importancia de la utilización de TICs como desarrollo de nuevos productos he considerado de interés la pregunta sobre el tipo de impacto que tienen las TICs en los distintos departamentos. Cuando el intercambio de datos se hace de forma electrónica las empresas opinan que la influencia es muy alta en todos los departamentos, sin embargo, en las que aún realizan intercambio de datos en formato papel, esta influencia es media o baja. Concretamente, analizados los siete departamentos a los que hace referencia la pregunta F6 sobre el impacto de las TICs en ellos, vemos una alta correlación entre la influencia de las TICs en los departamentos y el tipo de transmisión de documentos (electrónico, papel o intermedio).

Desde mi punto de vista, se explica según la experiencia en TICs que tienen las empresas, aquellas que trabajan con las TICs y saben de su potencial, también ven ese potencial en los departamentos existentes en las empresas. Mientras que aquellas que encuentran en el caso contrario, no ven en las TICs una oportunidad potencial para la empresa.

Otro tipo de perfil que se ve muy claro es según el tipo de mercado al que dirigen la mayor parte de las ventas. Aquellas empresas enfocadas a mercados internacionales utilizan en mayor medida las TICs, y conforme se va reduciendo el tipo de mercado (internacional→nacional→regional) también disminuye el uso de las TICs (Ilustración 9).

**Ilustración 9. Diagrama de asociación entre el tipo de mercado y de intercambio de datos.**



Fuente: Elaboración propia.

Si comparamos por la variación de la cuota de mercado, vemos que aquellas empresas que han incrementado la cuota interiorizan más las TICs que las que han permanecido igual que el año anterior. Análogamente, con la variación en la facturación del último año. En el análisis logit final, sería interesante averiguar si el nivel de consolidación que tienen las TICs en la empresa, es un factor de riesgo (pro/contra) para la variación de la facturación o de la cuota de mercado.

### **3.5. REDUCCIÓN DE LA DIMENSIÓN. ESTUDIO DE OUTLIERS**

Una vez realizado el análisis exploratorio y de correspondencias, voy a realizar un análisis factorial exploratorio y confirmatorio. Pretendemos encontrar factores relevantes que puedan describir una base con tantas variables correlacionadas. Finalmente veremos que empresas tienen un comportamiento muy diferenciado de las demás.

Las variables preseleccionadas una vez que hemos visto la comunalidad para cada una de ellas, el estadístico Kaiser-Meyer-Olkin y el test de esfericidad de Bartlett rechazado (véase anexo 3) son:

- Número de empleados con acceso a internet en el puesto de trabajo.
- Número de profesionales en TICs que han sido contratados por la empresa.
- Número de empleados con carrera universitaria.
- Tamaño de la compañía por número de empleados (variable proxy)
- Impacto de las TICs en dirección y control.
- Impacto de las TICs en administración y contabilidad.
- Impacto de las TICs en procesos de producción.
- Impacto de las TICs en marketing.
- Impacto de las TICs en atención al cliente.
- Impacto de las TICs en logística.

Los tres factores son:

- Impacto de TICs (Factor uno): Explica las variables que hacen referencia al impacto de las TICs en los diferentes departamentos de la empresa.
- Tamaño (Factor dos): Viene dado por el tamaño de la empresa (en número de empleados) y el número de empleados que son profesionales en TICs.

- Acceso y cualificación (Factor tres): Hace referencia al número de empleados con acceso a internet y la cantidad de empleados con carrera universitaria en la compañía.

He seleccionado para describir los tres atípicos que más destacan en los 3 factores:

- Empresa nº 11511: Es una empresa polaca, que trabaja el sector químico. Destaca por encima del resto en tamaño de la empresa en número de empleados (3.000 empleados contratados, y el 10% de ellos son profesionales en las TICs). Es una empresa joven de reciente creación (1.996), aspecto que indica una rápida expansión en el negocio. Así mismo, el impacto de las tics, el acceso de los empleados a internet (50%, además utiliza de máximo de banda ancha 2 megabits) y su cualificación (60% con carrera universitaria) no son destacables. Es una empresa que estima un alto impacto en TICs en todos sus departamentos salvo en el de procesos de producción. Invierte en TICs, y que utiliza cuatro de los cinco software de los que se indican en la encuesta. Además, más del 25% de sus pedidos y envíos los realiza de forma online. Trabaja en el mercado nacional, y en el último año ha incrementado la facturación y la cuota de mercado.

Podemos concluir que se trata de una empresa grande y en crecimiento que estima las TICs con un alto potencial. Además, teniendo el 60% de los empleados con estudios superiores (siendo el total de empleados 3.000) podemos suponer que dispone de un amplio departamento de I+D+i (muy necesarios en el tipo de sector trabajado por la empresa). Desde la perspectiva del marketing como empresa especializada en productos TICs, este tipo de empresa sería un cliente de cuenta grande y por consiguiente importante para tenerlo incorporado en nuestra base de clientes. Usaríamos la fuerza de ventas para empezar a contactar y negociar con esta empresa para a través de la venta relacional conseguir una relación sostenible en el tiempo.

- Empresa nº 10.256: Empresa polaca que compite en el mercado químico. Destaca sobre las demás en tamaño (cuenta con 40 empleados, 2 de los cuales son profesionales en TICs), en acceso (95% de los empleados) y cualificación (100% con estudios superiores). La empresa trabaja con el máximo ancho de banda, con software como DMS y PLM y realiza más del 25% de los pedidos de

forma online. En el impacto que tienen las TICs sobre los distintos departamentos de la empresa estima un alto impacto en todos, por lo que son conocedores del gran potencial de las TICs dentro de la empresa. Es una empresa en la que no se comparte información entre los distintos niveles que tiene y con los socios comerciales comparten datos mayoritariamente de forma online. Hizo en el periodo anterior inversiones en TICs y en el periodo siguiente espera mantener ese presupuesto. Es una empresa de 1984 que trabaja en el mercado internacional y tiene un 10% de cuota de mercado (la cual ha incrementado en el último año). El volumen de facturación de esta empresa también ha crecido en el último periodo.

Esta empresa pequeña que al igual que en el caso anterior se trata de una empresa en crecimiento y con conocedora del potencial de las TICs dentro de la empresa. Es una empresa que puede trabajar íntegramente en I+D+i al tener la totalidad de sus miembros con estudios superiores. Desde la perspectiva del marketing, seleccionaríamos productos que dispone la empresa relacionados con I+D (variantes del PLM por ejemplo) para tratar de captar a esta empresa como nuevo cliente.

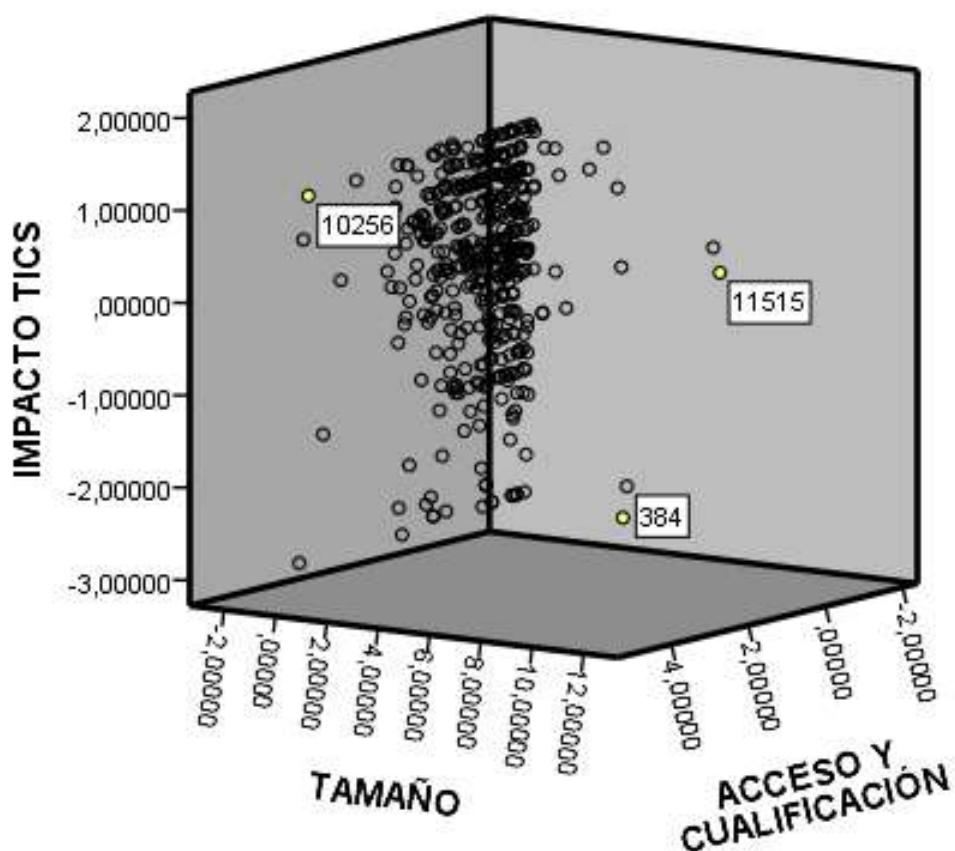
- Empresa nº 384: En este caso nos encontramos con una multinacional del sector químico, que trabaja en Reino Unido, España, Francia e Italia. La empresa atípica se corresponde a la perteneciente a Reino Unido. Destaca ante las demás, por el tamaño de la empresa (igual que en el caso de a 11.515 cuenta con 3.000 empleados, pero solamente un 6,66% son profesionales de las TICs) y el impacto de las TICs en los diferentes departamentos (el cual es bajo o inexistente). Es una empresa en la que todos los empleados tienen acceso a internet pero la mitad de estos tienen estudios superiores. Utiliza como máximo de banda 2 Megabits. No recibe ni envía pedidos de forma online. El intercambio de datos lo realiza en formato papel, por lo que no tiene muy desarrolladas las TICs dentro de su organización. Fue fundada en el 1.994. Es una empresa que trabaja el mercado internacional y tiene un 30% de cuota de mercado, la cual ha permanecido igual en el último año mientras que la facturación ha aumentado.

Este tipo de empresa desconoce el gran potencial que tienen las TICs para su empresa. Para tratar de establecer relaciones comerciales con este cliente, desde

la perspectiva del marketing, tendríamos que enfocar la estrategia en los beneficios que tienen nuestros productos (TICs) para su empresa, en cuanto al incremento de beneficios, utilidad y ahorro de costes para así hacer que sea consciente de la necesidad que tiene de adquirir los software de nuestra empresa.

Como los atípicos seleccionados y más destacables han sido empresas que trabajan el sector químico, el análisis cluster que presentaremos a continuación será solamente sobre las empresas que trabajan en ese sector.

**Ilustración 10. Gráfico tridimensional del Análisis Factorial.**



Fuente: Elaboración propia.

### 3.6. ANÁLISIS DE CONGLOMERADOS

En un trabajo de estas características, el análisis de conglomerados solo tiene interés para destacar similitudes y diferencias entre las empresas a las que se les ha realizado la encuesta, partimos además de una dificultad añadida y es el hecho de que la mayoría de las variables no son de carácter cuantitativo. Aun así, me gustaría destacar que si realizamos el análisis de conglomerados a partir de los resultados que hemos destacado en el punto anterior tiene un valor añadido en este proyecto si encontramos un grupo de empresas cuyo comportamiento es muy diferenciado, y en las que la utilización o no de TICs pueda ser destacable.

Vamos a seleccionar solamente las empresas del sector del químico, para de esta forma ver el impacto que tiene la utilización de los distintos software dentro del sector y poder buscar alguna conclusión interesante en el contexto del marketing para el resto de empresas del sector que no son destacables (atípicos del análisis factorial).

Al realizar el análisis, el resultado óptimo (dado por el dendograma) nos lo proporcionan 5 conglomerados aunque en uno de ellos haya un porcentaje muy elevado de casos (como podemos apreciar en la tabla de abajo).

**Tabla 2. Tabla de frecuencias de los cluster**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1	332	37,3	46,3	46,3
	2	66	7,4	9,2	55,5
	3	165	18,5	23,0	78,5
	4	131	14,7	18,3	96,8
	5	23	2,6	3,2	100,0
Total		717	80,6	100,0	
Perdidos	Sistema	173	19,4		
Total		890	100,0		

Fuente: Elaboración propia.

Vemos que en el conglomerado 1 se encuentran la mayor parte de los casos, por lo que será interesante indagar sobre las diferencias entre los grupos.

- Conglomerado 1: El perfil de empresas que forman este grupo, destacan por ser empresas que valoran la utilización de TICs, ya que estiman una influencia alta en los distintos departamentos de la empresa. El mercado predominante al que dirigen el mayor porcentaje de sus ventas es el nacional. Son empresas en las que predomina un bajo acceso a internet por parte de los empleados y menos del 33% de los empleados tienen estudios superiores (esto puede deberse a empresas del tipo “cadena de montaje”). Es el segundo que más usa los software de la encuesta, por debajo del cluster 3.
- Conglomerado 2: Son las empresas que menos implementadas las TICs, ya que son las que más destacan en la utilización de papel para compartir datos y por no hacer pedidos a través de internet. Además en este grupo predominan aquellas empresas que no hicieron inversiones en TICs el año pasado, y esperan mantener ese mismo presupuesto para el año próximo, es decir, que van a seguir sin invertir. Igual que en el grupo anterior, la cualificación de estudios superiores y el acceso a internet de los empleados son inferiores al 33%. Predominan las empresas pequeñas de 10 a 49 empleados. En este grupo encontramos las empresas que más han disminuido la facturación en el último. Es el grupo que menos utiliza los software que detalla la encuesta.
- Conglomerado 3: En este conglomerado encontramos aquellas empresas con más desarrollo de las TICs de forma interna: destacan por el intercambio de datos de forma electrónica, por el acceso de sus empleados a internet, por la utilización de internet para el pedido y envío de pedidos, y por la influencia futura que estiman que van a tener las TICs dentro de la empresa. Destacan por la utilización de software. Vemos también que son las empresas más fuertes y consolidadas, ya que en este grupo predominan empresas antiguas (anteriores a 1.982) y las más grandes en comparación con el resto de clusters. Añadir también que son el segundo grupo de empresas en cualificación de sus empleados (tienen importantes departamentos de I+D+i), son empresas que hicieron inversiones en TICs en el año anterior y son las que más van a incrementar el presupuesto en esta partida para el posterior.
- Conglomerado 4: Se asemeja en gran parte al conglomerado dos, las diferencias que podemos apreciar entre esos conglomerados, son que en este grupo encontramos empresas que estiman una baja influencia de las TICs en algunos

departamentos de la empresa, utilizan más banda ancha, está formado por empresas más grandes, con mayor incremento de cuota de mercado, sin embargo, menor cualificación. Y en cuanto a la utilización de los diferentes software, destaca por encima de todos al grupo 2 a excepción del CRM.

- Conglomerado 5: Es el grupo formado por menos empresas (tan solo 23). Esta formado en mi opinión por empresas dedicadas a I+D+i, ya que son aquellas que destacan por la cualificación de sus empleados ya que el 92% de estas empresas tienen más del 34% de sus empleados con estudios superiores. Además también destaca por el acceso a internet de sus empleados, y por los software que utilizan (comparten primera posición con el grupo 3 en la utilización de sistemas ERP y SCM, y encabezan en sistemas DMS)

Desde la perspectiva del marketing, podríamos encontrar un nicho de mercado en las empresas que forman los grupos 2 y 4. Son las que menos utilizan Tics, y menos relacionados están con los distintos software. Recomendaría a la empresa dedicada a la comercialización de productos TICs, establecer reuniones con estas empresas, sobre todo con las del grupo 4, que más invierten en TICs. En estas reuniones podríamos presentarles su situación frente a la de la competencia del sector, y como gracias a las TICs, puede mejorar la situación financiera de la empresa. Además, para ayudar a la toma de decisiones, podríamos utilizar estrategias de precios, como precios promocionales, o diseñar un Pack en el que se incluya la formación de los distintos software a los responsables de la utilización de las herramientas. De estos dos grupos, en mi opinión, yo me centraría en el grupo 4 dedicando a esas empresas un mayor interés y atención frente a las del grupo 2, porque son empresas que invierten en TICs, y empiezan a ser conocedoras de su importancia, además de que en el grupo 4 encontramos el doble de empresas.

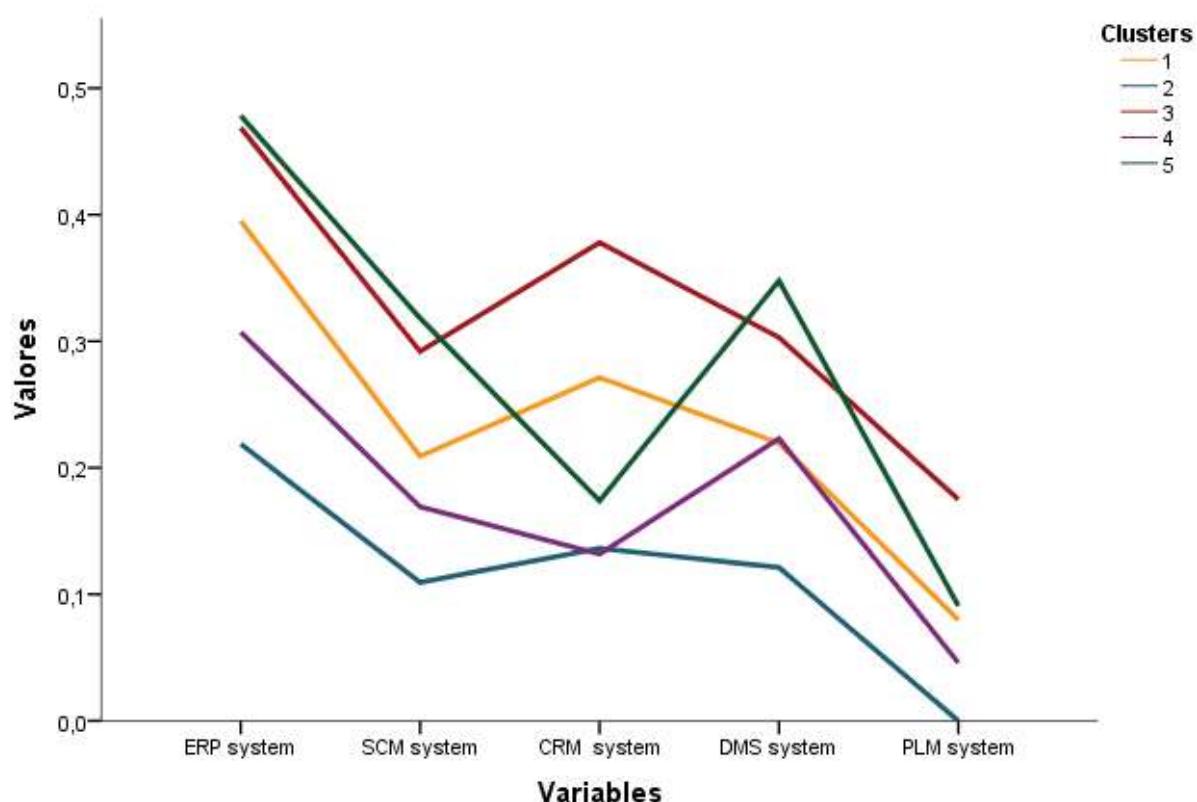
Para las empresas que forman los grupos 3 y 5, que son aquellas que podemos decir que destacan por su implicación con el I+D+i, y podríamos empezar a enfocar la estrategia de marketing en el tipo de productos relacionados con ese campo de la empresa. Este tipo de empresas son clientes importantes con los que convendría conseguir una relación comercial a largo plazo, ya que son los que más van a incrementar su presupuesto en TICs y los que más introducidos están con las nuevas tecnologías. Es importante tener en cuenta además, que las empresas del grupo 3 son conocedoras del valor añadido que proporcionan las TICs a los distintos departamentos de la empresa, mientras que las del grupo 5 no. Por lo que como segundo paso en nuestra estrategia de

marketing, si queremos comercializar en estas empresas no solo software o herramientas que se utilicen en departamentos de I+D+i, en el caso del grupo 3 lo tendríamos más sencillo. A las del tipo 5 habría que formarles sobre los beneficios de las TICs en las distintas partes de la empresa.

El grupo 1, son otro nicho de mercado a explotar, ya que aunque utilicen los software de forma más intensa que los grupos 2 y 4, está formado por empresas condecoradas del potencial de las TICs. Por lo que a este tipo de empresas no hay que presentarles un informe con el beneficio que aportaría un uso más intensivo de TICs, sino que habría que hacer uso de la fuerza de ventas de forma intensiva ya que es un grupo que ha incrementado su facturación, además, casi el 80% de las empresas que lo forman, invirtieron en TICs el año anterior y de estas, el 37,60% van a aumentar el presupuesto destinado a esta partida.

El informe final resumen puede verse en el gráfico adjunto (Ilustración 11).

**Ilustración 11. Informe resumen de utilización de software de los distintos cluster.**



Fuente: Elaboración propia.

Todo lo anterior podría ser de interés para una empresa ficticia que se dedicase a suministrar y comercializar software/hardware informáticos en el mundo B2B<sup>2</sup> al conocer diferentes tipos de nichos de mercado a los que acceder, y poder priorizar entre ellos según los recursos de la propia empresa. Todo ello genera el valor añadido de las 4 Vs (Volumen, Variedad, Velocidad y Veracidad) a las que he hecho referencia en la presentación de este proyecto.

### **3.7. UN MODELO LOGIT SEGÚN INVERSIÓN EN TICS.**

Para finalizar con la parte práctica de este proyecto vamos a llevar a cabo un análisis de regresión con variable dependiente categórica. El modelo que buscamos a priori con este tipo de análisis, es un modelo que explique la importancia de las TICs para la empresa. Por lo tanto la variable dependiente que vamos a estudiar va a ser la variación en facturación (G9), y las variables explicativas son aquellas que están relacionadas con la utilización de las TICs dentro de la empresa y la adquisición de software:

- Las que indican la adquisición de los distintos software: ERP (A7.1), SCM (A7.2), CRM (A7.3), DMS (A7.4) y PLM (A7.5). De esta forma podemos conocer si el disponer de un software en particular, (o en su conjunto) aumenta la probabilidad de la empresa de aumentar la facturación.
- Compartir los datos de forma electrónica o no (B11).
- Si aumentar o no el presupuesto en TICs (E6).
- Hacer inversiones en TICs o no en el último año (E9).

Como es un estudio en el que la variable dependiente es categórica, hemos categorizado la variable G9 en si crece la facturación o no.

El método utilizado es el habitual en estos modelos como en los de regresión con variable dependiente cuantitativa, es decir, regresión por pasos hacia delante (o hacia atrás). Esto es en cada paso va introduciendo aquellas que en ese paso son significativas. En el paso uno, incluye la variable más importante del modelo (en este ejemplo esta variable es el invertir o no en TICs en el periodo anterior). Y en los pasos sucesivos continua incluyendo aquellas variables significativas, mientras que las que no lo son las deja fuera del modelo.

---

<sup>2</sup> Business to Business

Para determinar la calidad del modelo, tenemos que centrar nuestra atención en los siguientes estadísticos:

- Prueba ómnibus: indica la significatividad del modelo según los diferentes pasos en los que se incluyen nuevas variables al modelo. Tenemos que fijarnos en concreto en la significatividad de los distintos pasos para aceptar el modelo.

**Tabla 3. Pruebas ómnibus sobre los coeficientes del modelo logit**

		Chi cuadrado	gl	Sig.
<b>Paso 1</b>	Paso	70,346	1	,000
	Bloque	70,346	1	,000
	Modelo	70,346	1	,000
<b>Paso 2</b>	Paso	28,672	1	,000
	Bloque	99,018	2	,000
	Modelo	99,018	2	,000

Fuente: Elaboración propia.

- Estadístico de Hosmer-Lemeshow: en el caso de que el p-valor para este estadístico fuera menor de 0,05 nos encontraríamos con un mal ajuste del modelo y no nos serviría. En este ejemplo, el estadístico tiene un valor de 0,713, es decir, un buen ajuste.

**Tabla 4. Prueba de Hosmer y Lemeshow**

Paso	Chi cuadrado	gl	Sig.
1	,000	0	.
2	,677	2	,713

Fuente: Elaboración propia.

- -2 log de verosimilitud
- R cuadrado de Cox y Snell: estadístico utilizado para funciones logarítmicas categóricas, no se limita entre 0 y 1.
- R cuadrado de Nagelkerke: es una modificación del estadístico anterior en el que los valores de la R cuadrado si varían entre 0 y 1.

**Tabla 5. Resumen del modelo**

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	2325,773 <sup>a</sup>	,039	,053
2	2297,101 <sup>a</sup>	,055	,074

a. La estimación ha finalizado en el número de iteración 3 porque las estimaciones de los parámetros han cambiado en menos de ,001.

Fuente: Elaboración propia.

Estos tres últimos estadísticos los podemos utilizar para comparar en el modelo las diferentes versiones que podemos obtener de él (ya que en cada paso se introduce una nueva variable). Modelos con la R cuadrado más alta o menor estimación en la función de verosimilitud será el que mejor modelo. En nuestro caso, nos encontramos con un R cuadrado un poco bajo (inferior a 0,1), pero como el estadístico de Homer-Lemeshow nos da un valor bueno, no rechazamos el modelo. No es un modelo perfecto, pero es un modelo realista [para más detalles, consultar Siegel E. (2014)].

El modelo que hemos obtenido es el siguiente:

**Tabla 6. Variables en la ecuación del modelo**

		B	E.T.	Wald	gl	Sig.	Exp(B)
<b>Paso 1<sup>a</sup></b>	E9	,888	,107	68,883	1	,000	2,430
	Constante	-,362	,089	16,484	1	,000	,696
<b>Paso 2<sup>b</sup></b>	E6	,601	,114	27,918	1	,000	<b>1,824</b>
	E9	,800	,109	54,198	1	,000	<b>2,225</b>
	Constante	-,467	,092	25,767	1	,000	,627

a. Variable(s) introducida(s) en el paso 1: E9.

b. Variable(s) introducida(s) en el paso 2: E6.

Fuente: Elaboración propia.

Un dato de especial importancia que encontramos en la tabla superior es el ODDS ratio (columna “Exp (B)”). Este ratio indica el coeficiente entre la probabilidad de que ocurra el suceso que indica la variable dependiente (aumentar la facturación) frente a la probabilidad de que no ocurra en presencia o ausencia del factor (aumentar o no aumentar el presupuesto en TICS).

El valor 1,824 que encontramos en la variable “E6: Aumentar o no el presupuesto en TICS”, nos indica que es 1,824 veces más grande el cociente entre la probabilidad de

aumentar la facturación de la empresa que de no aumentarla, en aquellas empresas que han decidido aumentar el presupuesto en TICs para el año próximo.

De esta misma forma, el valor 2,225 correspondiente a la variable “E9: Invertir o no en TICs en el periodo anterior”, indica que el cociente entre la probabilidad de aumentar la facturación de la empresa que de no aumentarla es 2,225 veces más grande en aquellas empresas que invirtieron en TICs en el año anterior.

**Tabla 7. Tabla de probabilidades de Análisis Logit, según grupo.**

	E6	E9	Probabilidad
<b>Caso 1</b>	1	1	<b>80,23%</b>
<b>Caso 2</b>	0	1	<b>69,00%</b>
<b>Caso 3</b>	1	0	<b>64,59%</b>

Fuente: Elaboración propia.

Donde: E6=1 (aumenta el presupuesto en el año próximo); E6 =0 (No aumenta)

E9=1 (Realizo inversiones en TICs en el último año) E9=0 (No hizo)

De la tabla anterior se deduce:

Caso 1: La empresa que invirtió en TICs el año anterior y este año aumenta el presupuesto en esa partida tiene un 80,23% más de probabilidad de aumentar la facturación que una empresa que ni invirtió en TICs ni aumentará el presupuesto en el próximo año.

Caso 2: La empresa que invirtió en TICs tiene un 69% más de probabilidad de crecer en facturación frente a la que no invirtió.

Caso 3: La empresa que aumenta el presupuesto en TICs para el año próximo, tiene un 64,59% más de probabilidad de incrementar la facturación que aquella que no lo aumenta.

Los resultados obtenidos de este análisis los podríamos utilizar para explotar el nicho de mercado de empresas que aun no están desarrolladas en el campo de las TICs e incrementar la venta de software/hardware a esas empresas. Así pues el analista experto en investigación de mercados presentaría un INFORME a esas empresas para que

cambiaran el tipo de utilización de TICs. Lo cual les hace incrementar su VALOR en el mercado.

**Por tanto este modelo logit justifica de alguna manera la importancia que para las empresas tiene la inversión en TICs si quieren aumentar su facturación.**

### **3.8. RESULTADOS DE LAS ESTIMACIONES Y CONLUSIONES.**

Este pequeño modelo empírico en su conjunto, es una muestra del valor que tiene para cualquier empresa un analista de datos y de los beneficios que supone esta práctica para la empresa. Para concluir con el análisis empírico vamos a recapitular las recomendaciones que desde el punto de vista de una consultora de marketing que cuenta como cliente con una empresa de software/hardware harían que aumentase en gran medida su cartera de clientes y por consiguiente su cuota de mercado y facturación.

Lo más importante es la personalización o segmentación de estrategias. La segmentación que se recomienda llevar a cabo dentro de las empresas que constituyen el sector químico (sector en el que más hemos profundizado en el análisis) es la siguiente:

1. Empresa 11.515: cliente de cuenta grande y por consiguiente importante para tenerlo incorporado en nuestra base de clientes. Utilizar fuerza de ventas y marketing relacional para crear una relación comercial a largo plazo.
2. Empresa 10.256: Esta empresa es probable que sea empresa de investigación, por lo que distribuirle productos que dispone la empresa relacionados con I+D+i (variantes del PLM por ejemplo) a través de la fuerza de ventas, sería la forma más sencilla de intentar captar a este cliente.
3. Empresa 384: Empresa desconocedora del potencial que tienen las TICs para su empresa. Presentar informe con los resultados del análisis logit.
4. Resto: (en orden de importancia)
  - a. Grupo 3 y grupo 5: formados por clientes importantes, por lo que utilizar fuerza de ventas y marketing relacional para establecer relación comercial a largo plazo. Son empresas de investigación, por lo que sería conveniente utilizar promociones con productos relacionados con I+D+i para empezar a captarlos y posteriormente el resto de software/hardware que comercialice la empresa. Grupo 5 es desconocedor del valor de las TICs, por lo que presentar informe logit a las empresas que lo forman.
  - b. Grupo 1 de empresas: Fuerza de ventas intensiva.

- c. Grupo 2 y grupo 4: Presentar informe logit, utilizar fuerza de ventas intensiva y establecer estrategias de precios promocionales que incluyan la formación de los empleados en los software/hardware. Grupo 4 más importante que el grupo 2.

Toda esta información tendría gran interés para aquellas empresas del sector de comercialización de productos informáticos que busquen la expansión de negocio o la búsqueda de nuevos clientes (objetivo cumplido).

## **PARTE 4: CONCLUSIONES GENERALES**

Como hemos podido comprender a lo largo de la parte 3 de este proyecto, el disponer de un analista de investigación de mercados en la empresa y de los software/hardware adecuados de tratamiento y almacenamiento de datos, puede suponer una gran ventaja competitiva para la empresa ya que nos permite conocer mejor el entorno competitivo y a nuestros clientes. Es algo imprescindible para cualquier empresa que este tipo de herramientas se implementen en su sistema de negocio, para explotar y sacar el máximo rendimiento de las bases de datos internas o externas de la empresa y de esta forma optimizar los recursos de la empresa.

La utilización de esta herramienta, no tiene porque ser de un elevado coste para la empresa como hemos podido comprobar. El analista puede recurrir a las bases de datos libres (open data), no siempre hay que generar las propias de la empresa, aunque si no es costoso, sería una herramienta de gran interés.

Viendo los resultados y los beneficios que estas técnicas ejercen en la empresa, es comprensible la gran expansión e introducción que está teniendo dentro de las empresas más importantes. Este tipo de herramientas, no solo afectan a las grandes corporaciones. La pequeña empresa, también puede beneficiarse de la minería de datos. La diferencia radica en la priorización en la segmentación de estrategias, que irán en función de los recursos que tenga la pequeña empresa.

Concretamente, es también importante destacar aquí, el aumento que la implementación de un sencillo modelo logit nos ha permitido cuantificar el aumento en facturación en el corto plazo si la empresa invierte en TICs y/o aumenta su presupuesto en el año próximo. Con este análisis hemos podido comprobar el VALOR que puede tener el uso de este tipo de herramienta de análisis para cualquier empresa. Como se señalaba en la introducción de este proyecto al mencionar las cuatro V (Volumen, Variedad, Velocidad y Veracidad), y la quinta que están incorporando a día de hoy algunos autores (VALOR).

Como señala PORRES, E. en el último informe de Forbes magazine la extracción de información de las grandes bases de datos se está convirtiendo en el nuevo petróleo del siglo XXI. Las empresas saben que ya no es importante analizar el cómo ni el porqué del consumo de este determinado artículo sino sencillamente tener la foto de los datos y de lo que ocurre en el mercado y tomar decisiones.

## BIBLIOGRAFIA

GUINALIU, M. ¿acabará el big data con la investigación tradicional en marketing? [Página web], 14/05/2014 [fecha de consulta: 15/06/2014]

GUISANDE GONZÁLEZ, C. y otros; (2011) *Tratamiento de datos con R, STATISTICA y SPSS*. Díaz de Santos. Madrid.

HAIR, J. y otros, (2008) *Análisis multivariante*. Prentice Hall. Madrid.

HAIR, J. et al, (2014) *Multivariate Data Analysis*. Pearson New International Edition. Edinburgh

PORRES, E.L. The big potential of big data [página web]. Forbes magazine. [www.forbes.com/forbesinsight](http://www.forbes.com/forbesinsight) , 2013 [fecha de consulta: 07/07/2014]

RATNER, B. (2012) *Statistical and machine-learning data mining: techniques for better predictive modeling and analysis of big data*. CRC, Taylor & Francis Group. Boca Raton, FL.

SCHROECK, M. et al, Analytics: el uso de Big Data en el mundo real [página web] <http://www-05.ibm.com>, 2012 [fecha de consulta: 09/05/2014]

SIEGEL, E. (2013) *Predictive analytics: the power to predict who will click, buy, lie, or die*. Wiley. Hoboken.

SIEGEL, E. (2013) *Analítica predictiva: Predecir el futuro utilizando Big Data*. Anaya Multimedia. Madrid.