

Visualizaciones analíticas para la toma de decisiones en pequeñas y medianas empresas utilizando Data Mining*

Analytical Visualizations for Decision-Making in Small and Medium Sized Enterprises through Data Mining

Paola Andrea Noreña Cardona**
Gustavo Adolfo Quiceno Brand***
Luis Armando Uribe Marín****



Tipo de artículo: resultado de investigación

Recibido: 11 de diciembre de 2015

Aceptado: 10 de marzo de 2016

Resumen

Las visualizaciones analíticas se definen en el contexto de la inteligencia de negocios, como instrumentos que permiten analizar anticipadamente grandes cantidades de información en diferentes bases de datos y ayudar a tomar decisiones más estratégicas. Las pequeñas y medianas empresas que no poseen plataformas de inteligencia de negocios BI, especialmente en explotación de información se ven afectadas en diferentes momentos, debido a la manipulación y a la observación de estos tamaños de información que dificultan sus procesos. Para disminuir esta problemática, se pretenden realizar visualizaciones analíticas, que permitan la toma de decisiones en pequeñas y medianas empresas utilizando la metodología más apropiada para el análisis de bases de datos y explotación de información a partir de Data Mining. Esto facilita el acceso de estas compañías a módulos similares de explotación de información, a los que ofrecen las grandes plataformas de Inteligencia de Negocios. Además se desea validar el servicio del visor analítico en una empresa.

Palabras clave: visualizaciones analíticas; CRISP-DM; inteligencia de negocios; toma de decisiones; Data-Mining

Abstract

Analytical visualizations are defined, within the context of business intelligence, as tools allowing the previous analysis of large amounts of information in different databases, to help in the making of more strategic decisions. Small and medium sized enterprises do not have Business Intelligence (BI) platforms. They are affected at different times, because of the manipulation and observation of such amounts of information which make their processing difficult. Analytical visualizations are intended to reduce such problem, to allow decision-making in small and medium sized enterprises by using the most appropriate methodology for database analysis from Data Mining. This makes the access easier for these companies into Information Exploitation modules, which are databases to those offered by big Business Intelligence platforms. In addition, the validation of the analytical viewer service in a company is also intended.

Keywords: Analytical visualizations; CRISP-DM; Business Intelligence; decision-making; Data-Mining.

* Este artículo se deriva como resultado del proyecto visor analítico para la toma de decisiones gerenciales en pequeñas y medianas empresas utilizando inteligencia de negocios.

** M.Sc en Ingeniería de Sistemas. Docente ocasional. Tecnológico de Antioquia- Institución Universitaria. Correo electrónico: panorena@tdea.edu.co

*** Ingeniero de Software. Tecnológico de Antioquia- Institución Universitaria. Correo electrónico: gquicen2@correo.tdea.edu.co

**** Ingeniero de Software. Tecnológico de Antioquia-Institución Universitaria. Correo electrónico: luribem@correo.tdea.edu.co

Introducción

Las visualizaciones analíticas ofrecen un valor especial de exploración y técnicas de visualización de datos, que permiten tener factores diferenciadores en el comportamiento de los reportes y tableros al transmitir información sobre las futuras posibilidades, lo que se deriva en una toma de decisión más estratégica al generar mayor conocimiento y previsión de la información. Esto genera beneficios a las empresas, ya que puede prever y analizar los movimientos de sus datos. Un ejemplo de visualización analítica, podría ser los ingresos económicos en las ventas de unas botas femeninas en una geografía particular y ayudar en la obtención del valor de su crecimiento en un tiempo determinado (SAS Institute, 2013).

Actualmente las grandes empresas utilizan herramientas de inteligencia de negocios para analizar grandes cantidades de información y tomar adecuadas decisiones en el negocio (Calzada, L & Abreu, J. 2009; Microsoft Corporation, 2010; Sauter, 2014). Pequeñas y medianas empresas importan datos utilizando herramientas que no son especializadas y se pierde información que puede ser valiosa en el negocio (Hernández *et al.*, 2007). Algunas de ellas usan otras básicas, que permiten operaciones en sistemas de información con bases de datos relacionales que ayudan al gerenciamiento o la administración de sus organizaciones y permiten visualizaciones de datos, no permiten explotar la información para hacer acercamientos futuros de indicadores o márgenes de error en la toma de decisiones. Se debe proporcionar herramientas adecuadas para la explotación y análisis de los datos que logren obtener el conocimiento necesario en el proceso de toma de decisiones estratégicas (Soto, 2001).

Para lograr una adecuada toma de decisiones en pequeñas y medianas empresas, se pretende utilizar visualizaciones analíticas utilizando herramientas de inteligencia de negocio como el Data Mining, DM, que permita la explotación

de la información (Ferreira *et al.*, 2005) en las empresas. El DM traducido al español como minería de datos, es un mecanismo de explotación consistente en la búsqueda de información valiosa en grandes volúmenes de datos que proporcionan la información histórica con la cual los algoritmos de minería de datos tienen la información necesaria para la toma de decisiones (Soroush y Bameni, 2009). Además, el DM analiza archivos y bitácoras de transacciones, trabaja a nivel del conocimiento, con el fin de descubrir patrones, relaciones, reglas, asociaciones o incluso excepciones útiles para la toma de decisiones (Vallejo, 2006). De lo anterior se concibe la necesidad de crear visualizaciones analíticas para pequeñas y medianas empresas con ayuda de una metodología. Las metodologías, permiten implementar técnicas de análisis y explotación de información mediante el proceso de minería de datos, en forma sistemática y ayuda a las organizaciones a entender el proceso de descubrimiento de conocimiento con una guía para la planificación y ejecución de los proyectos, de esta manera generar un resultado que beneficie a las pequeñas y medianas empresas. (Moine, 2001a).

Este artículo se presenta de la siguiente manera: en la sección 2 se condensan las metodologías que permiten la explotación de información. En la sección 3 se exhibe la metodología más apropiada, a partir del análisis de varios autores. En la sección 3 se presentan los resultados de las visualizaciones analíticas utilizando la metodología seleccionada. En la sección 4 se realiza una discusión, a partir de los resultados obtenidos. Finalmente, en la sección 5 se presentan las conclusiones.

Antecedentes

Como consecuencia de la experiencia algunos autores han desarrollado metodologías para la explotación de información que permiten la toma de decisiones como: SAS Institute (2002) propone SEMMA, que permite la exploración y el modelado de volúmenes grandes de información para revelar patrones de negocio por medio de cinco fases: muestreo, exploración, modificación,

modelado y verificación. Sin embargo no tiene en cuenta la implementación y comprensión del negocio que son indispensables para las empresas especialmente las pequeñas y medianas. Pyle (2003) *presenta* la metodología Catalyst, conocida como P3TQ la cual tiene en cuenta el precio, tiempo y producto en la explotación de la información, pero no tiene en cuenta análisis del negocio. *Berry y Linoff (2004) desarrolla la metodología Berry y Linoff*, que permite la relación de atributos a medir a la hora de evaluar los algoritmos de minería de datos, pero su nivel de abstracción en el negocio debilita su proceso. Knowledge Discovery in Databases KDD o extracción de conocimiento en bases de datos se define como el proceso de identificación de patrones válidos, útiles y comprensibles a partir de los datos (Casillasy Delgado, 2005) (Giraldo y Builes, 2013). La extracción de patrones no comprensibles dificulta o imposibilita su interpretación, revisión, validación y uso en la toma de decisiones. Además una información incomprensible no proporciona conocimiento. KDD y SEMMA son catalogados como modelos y no llegan a ser una metodología, debido a que dejan a criterio del equipo de trabajo la definición de las actividades a realizar en cada etapa del proyecto. Particularmente SEMMA excluye dos etapas importantes del proceso como son el análisis del negocio y la difusión del nuevo conocimiento, evidenciando que el modelo está orientado especialmente aspectos técnicos. El Modelo de Explotación de Información ,MIII, proporciona una guía de pasos para la construcción y ejecución de modelos de minería de datos, a partir del Modelo de Negocio (MII). (Moine *et al.*, 2011a). Para representar módulos en base a Data Mining, es indispensable tener riguroso

análisis de la información del negocio, es así como SPSS, NCR y Daimler Chrysler presentan la metodología CRISP-DM, la cual constituye la guía de referencia más utilizada en el desarrollo de proyectos de Explotación de Datos (Moine *et al.*, 2011a). El presente trabajo se enmarca en esta metodología para atender a las necesidades de toma de decisiones en pequeñas y medianas empresas.

Metodología

La metodología para esta investigación consistió en realizar la revisión de literatura, seleccionar la metodología de trabajo en Data Mining, adicionalmente se propone mediante cubos OLAP (cubos de procesamiento analítico en línea) hacer visualizaciones analíticas para dar solución al análisis y explotación de información y se evalúa en una empresa. Para el desarrollo de las visualizaciones se utiliza la metodología CRISP-DM (*Cross Industry Standard Process for Data Mining*) la cual consiste en un modelo jerárquico de procesos, constituido por un conjunto de tareas en seisniveles de abstracción organizados de forma jerárquica en tareas, que van desde el nivel más general, hasta los casos más específicos y organiza el desarrollo de un proyecto de Data Mining, en una serie de seis fases. La sucesión de fases no es necesariamente rígida. Cada fase es estructurada en varias tareas generales de segundo nivel. Las tareas generales se proyectan a tareas específicas, donde finalmente se describen las acciones que deben ser desarrolladas para situaciones específicas, pero en ningún momento se propone como realizarlas (Chapman *et al.*, 2000).

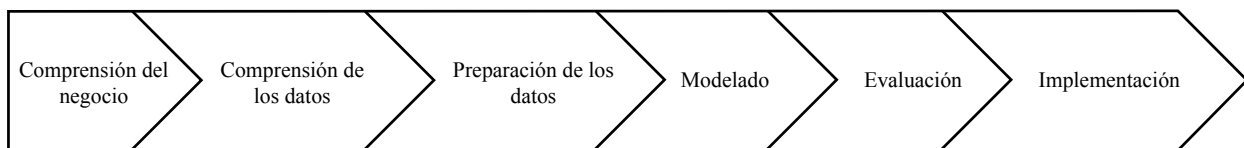


Figura 1. Modelo CRISP-DM.
Fuente: Chapman *et al.* (2000)

Fase 1. Comprensión del negocio o problema. Es probablemente la más importante y aglutina las tareas de comprensión de los objetivos y requisitos del proyecto desde una perspectiva empresarial o institucional, con el fin de convertirlos en objetivos técnicos y en un plan de proyecto. Para obtener el mejor provecho de DM, es necesario entender de la manera más completa el problema que se desea resolver, esto permitirá recolectar los datos correctos e interpretar correctamente los resultados.

Fase 2. Comprensión de los datos. Comprende la recolección inicial de datos, con el objetivo de establecer un primer contacto con el problema, familiarizándose con ellos, identificar su calidad y establecer las relaciones más evidentes mediante recolección de datos iniciales, evaluación de la situación y verificación de la calidad de los datos.

Fase 3. Preparación de los datos. Se procede a su preparación para adaptarlos a las técnicas de Data Mining que se utilicen posteriormente, tales como técnicas de visualización de datos, de búsqueda de relaciones entre variables u otras medidas para exploración de los datos.

Fase 4. Modelado. Se seleccionan las técnicas de modelado más apropiadas en función de los siguientes criterios: ser apropiada al problema, disponer de datos adecuados, cumplir los requisitos del problema, tiempo adecuado para obtener un modelo, conocimiento de la técnica. Chapman *et al.* (2000).

Fase 5. Evaluación. En esta fase se evalúa el modelo, teniendo en cuenta el cumplimiento de los criterios de éxito del problema. Debe considerarse además, que la fiabilidad calculada para el modelo se aplica solamente para los datos sobre los que se realizó el análisis.

Fase 6. Implementación. Para implementar el resultado de DM en la organización, esta tarea toma los resultados de la evaluación y concluye una estrategia para su implementación. Si un procedimiento general se ha identificado para crear el modelo, este procedimiento debe ser documentado para su posterior implementación.

Resultados

Para dar solución al análisis y explotación de información anticipadamente, se propone la existencia de visualizaciones analíticas a partir de los Cubos OLAP, que permiten almacenar los datos multidimensionales en bases de datos siguiendo la metodología CRISP-DM y se prueba en la empresa Francisco Murillo S.A.S de la ciudad de Medellín en Colombia.

En la *comprensión del negocio*, se incluyeron la exploración de los interesados de la empresa y las necesidades de su dominio. Al explorarlo se observan los interesados con sus roles como se determina en la Tabla 1 y los requisitos del dominio en marcados en la compra y venta de artículos electrónicos como se observan en la Tabla 2.

Tabla 1. Interesados

| Nombre | Cargo | Influye Proyecto |
|---------------------------------------|------------------|------------------|
| Álvaro Cuartas Cuartas(Interno) | Subgerente | SI |
| Silvania (Externo) | Proveedor | NO |
| Legrand(Externo) | Proveedor | NO |
| Luis Armando Uribe Marín (Externo) | Ingeniero | SI |
| Gustavo Adolfo Quiceno Brand(Interno) | Ingeniero | SI |
| Programas y Asesorías(Externo) | Casa de software | SI |

Fuente: elaboración propia (2015)

Tabla 2. Requisitos

| Nombre | Descripción |
|---|---|
| 1. Ingreso a las visualizaciones analíticas | Permite el ingreso al sistema Visor Analítico, utilizando un usuario y su contraseña del sistema ERP (sistemas de planificación de recursos empresariales) Internamente se conoce las claves de base de datos y se adicionan al código de acceso a la base de datos |
| 2. Parámetros | El cliente define los parámetros para los cálculos del cubo gerencial. |
| 3. Cubo OLAP | Explora y recopila los datos, descripción de datos, tipos de valores y administra la conexión y entrega al cliente la información comercial almacenada en la base de datos relacional de la empresa para ser administrado por medio de un cubo OLAP. Se marca la primera lista donde se escoge el diseño de los informes y su tiempo de generación y el sistema crea el cubo comercial. Se genera tomando como referencia los parámetros ingresados y marcados por el usuario, permite que el cliente realice las combinaciones en el cubo comercial. |
| 4. Generar gráfica OLAP | Explora y recopila los datos, para ser administrado por medio de un gráfico dinámico. Gráfica el cubo comercial, y toma como referencia los parámetros ingresados. Permite que el cliente realice las combinaciones en el cubo comercial. |

Fuente: elaboración propia (2015)

Para la *comprensión de los datos* se creó copia de la base de datos utilizada en la empresa. Se exploraron con una vista sin campos calculados y se comprobó el rendimiento y registro con uno de los interesados.

En la *preparación de los datos* se construyeron nuevos datos, por medio de cálculos en asociación con otros para generar los de la base de datos relacional en el cubo. Ver Tabla 3.

Tabla 3. Preparación de los datos

| Campo Cubo | Resuelve | Función |
|----------------------------|-----------|--|
| Agrupado días categoría | Calculado | f_dyna_diasinvcat(f_dyna_fechadocinv(articulos.ka_nl_articulo,bodegas.ka_nl_bodega)) |
| Nit Vendedor | Calculado | f_vendedor_dyna('1', movimientos.ka_nl_movimiento) |
| Nombre sucursal encabezado | Calculado | f_dyna_sucursal('2', movimientos.ka_nl_movimiento) |
| Nombre vendedor | Calculado | f_vendedor_dyna('2', movimientos.ka_nl_movimiento) |
| Numero pedido | Calculado | f_dyna_numeroped(movimientos_items.ka_nl_movimiento_item) |
| Zona cliente | Calculado | f_zona_mvto(movimientos.ka_nl_movimiento) |
| Código sucursal | Calculado | f_dyna_sucursal('1', movimientos.ka_nl_movimiento) |
| Mes | Calculado | f_periodo(year(movimientos.d_fecha_documento),month(movimientos.d_fecha_documento)) |
| Año | Calculado | STRING(YEAR(movimientos.d_fecha_documento)) |
| Bimestre | Calculado | f_dyna_coverper(substring(movimientos.d_fecha_documento,6,2),'B') |
| Trimestre | Calculado | f_dyna_coverper(substring(movimientos.d_fecha_documento,6,2),'T') |
| Semestre | Calculado | f_dyna_coverper(substring(movimientos.d_fecha_documento,6,2),'S') |
| Hora | Calculado | hour(movimientos.dt_hora) |

Fuente: elaboración propia (2015)

El cumplimiento de la fase de *Modelado* se realizó dando cumplimiento a los requisitos del negocio graficando su caracterización mediante herramientas de análisis y diseño de software como el diagrama de clases de la Figura 1 y el diagrama de actividades de la Figura 1.

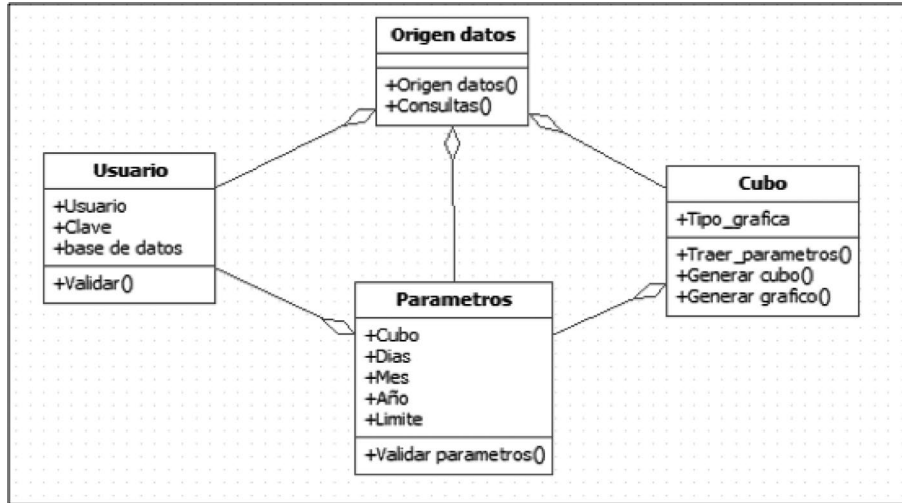


Figura 1. Diagrama de clases
Fuente: elaboración propia (2015)

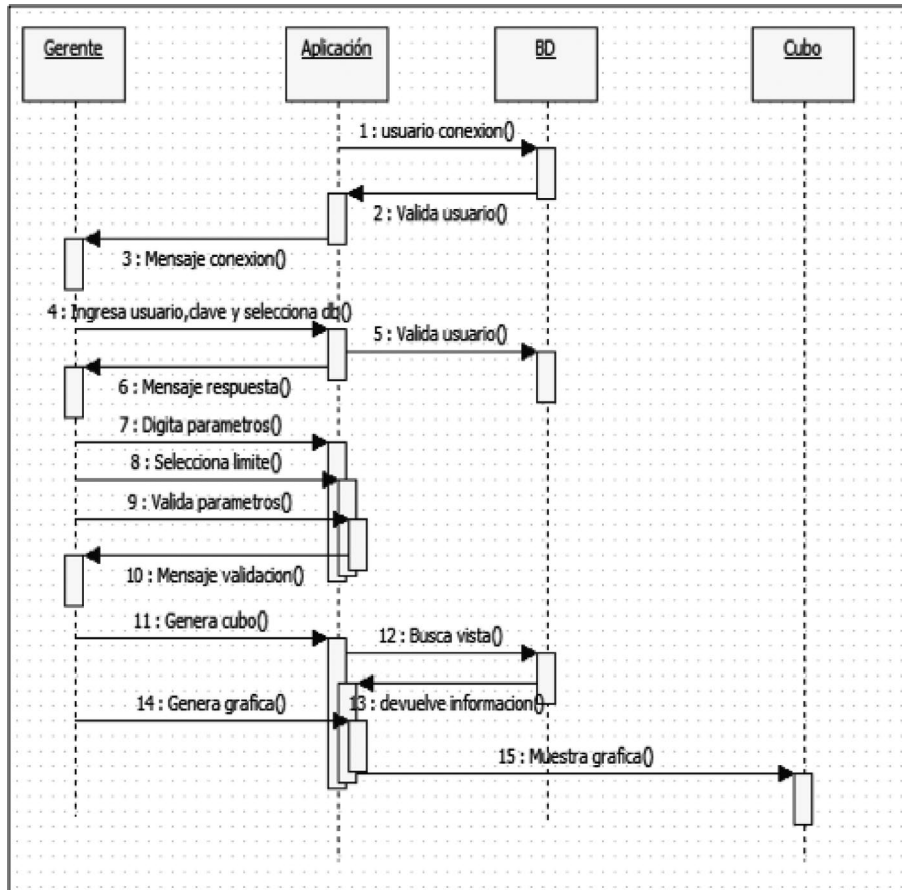


Figura 2. Diagrama de clases
Fuente: elaboración propia (2015)

Al llevar a cabo la *implementación* se integraron los diferentes atributos de medición en ventas y compras donde el instrumento de visualización permite mostrar elementos de las diferentes tablas, para que por medio de gráficos se pueda realizar el análisis de la información en las tendencias o comportamientos futuros y en un tiempo determinado. La funcionalidad de las

visualizaciones, consiste en señalar parámetros que se desean analizar en el cubo, posteriormente se pueden adicionar otros campos, filtros y agrupar valores según la necesidad. En la Figura 3, se pueden observar visualizaciones del módulo de ventas, donde se puede seleccionar las gráficas para analizar otras variables observar los campos de las tablas dinámicas.

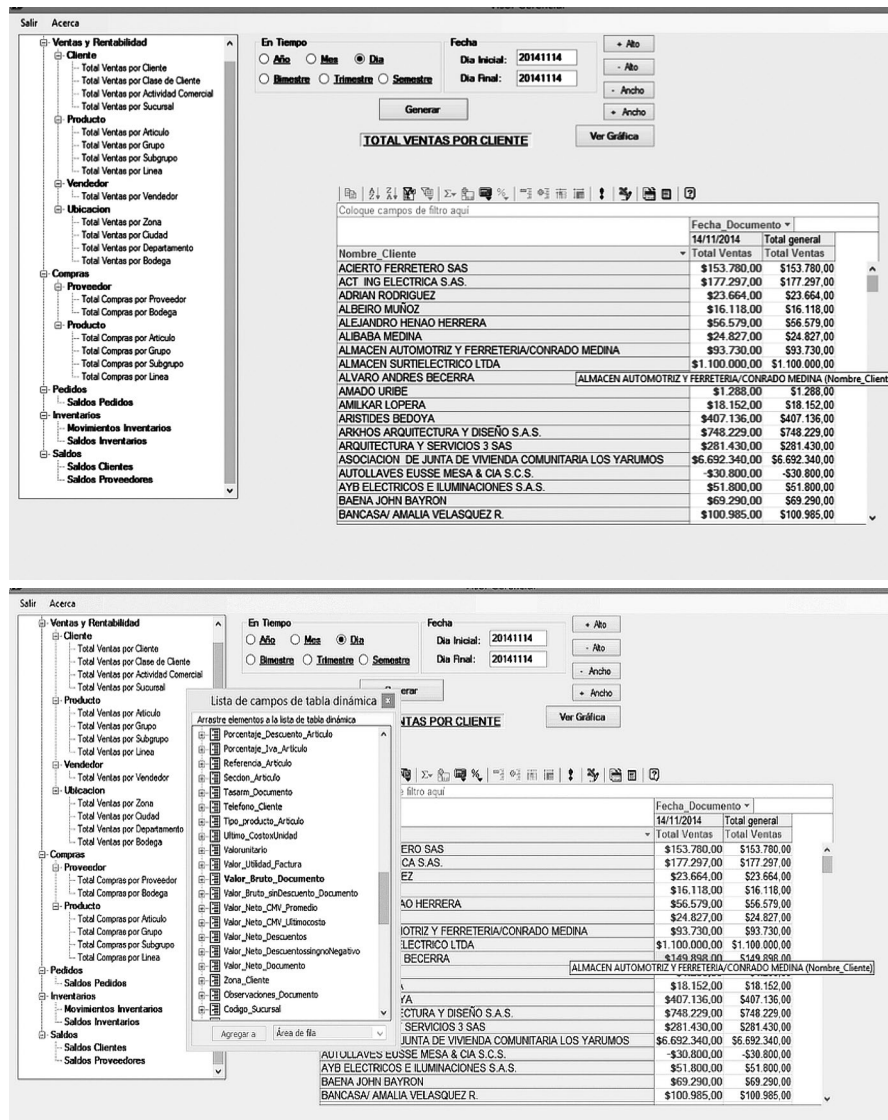


Figura 3. Funcionalidad del instrumento de visualizaciones analíticas
Fuente: elaboración propia (2015)

Finalmente para *evaluar* las visualizaciones analíticas en la herramienta se efectuó la validación dentro de la empresa Francisco Murillo S.A.S en

base al análisis de las ventas. Para entregar esta información se utilizó la herramienta, tomando la información de la base de datos que se conectó

desde el departamento comercial. Se puede notar en unidades, y porcentualmente dependiendo la en la Figura 4, las gráficas que presenta el cubo participación en el total de las ventas.

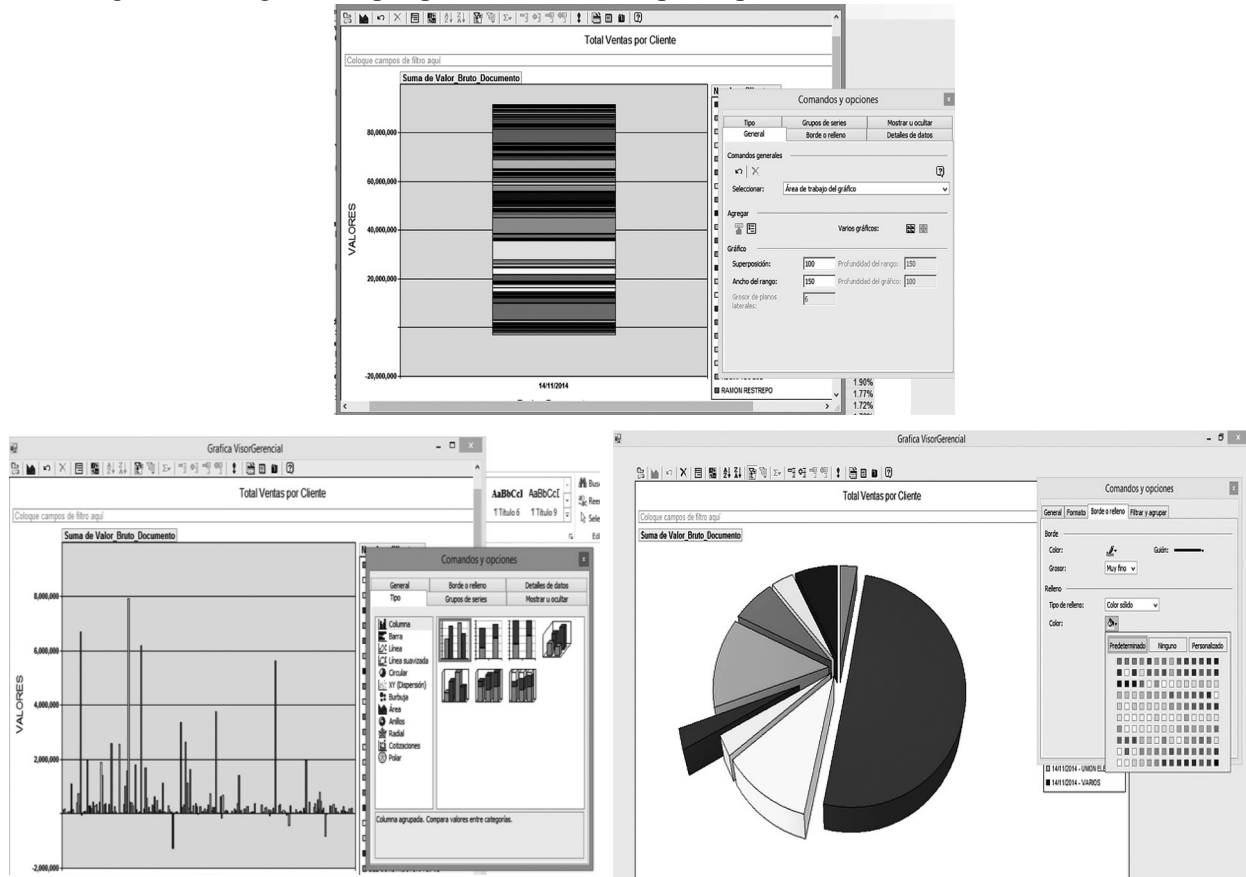


Figura 4. Generación de gráficas del Cubo
Fuente: elaboración propia (2015)

Conclusiones

La información es el recurso más importante de una empresa, sin él no se podría ejercer la gestión de sus procesos. El interés de cuidar este recurso, analizarlo y extraerlo permite profundizar en técnicas que permitan su manipulación. La aplicación de la metodología CRISP-DM es estándar ante otras en la extracción de información en minería de datos, por eso en compañía de visualizaciones analíticas se puede observar detalladamente la información, permitiendo que pequeñas y medianas empresas puedan tomar decisiones adecuadas que afecten positivamente sus productos y servicios. Además se optimizan recursos y se facilitan el análisis de su comportamiento futuro en la empresa.

Referencias

- Berry, M. y Linoff, G. (2004). *Data Mining Techniques, Second Edition*. Indianapolis: John Wiley & Sons. Inc.
- Calzada, L y Abreu, J. (2009). El impacto de las herramientas de inteligencia de negocios en la toma de decisiones de los ejecutivos. *Daena: International Journal of Good Conscience*. 4(2). 16-52. Recuperado de [http://www.spentamexico.org/v4-n2/4\(2\)%2016-52.pdf](http://www.spentamexico.org/v4-n2/4(2)%2016-52.pdf)
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. y Wirth, R.(2000). CRISP-DM 1.0 Step by step data mining gui-

- de. SPSS Inc.9 (13). Recuperado de <https://the-modeling-agency.com/crisp-dm.pdf>
- Casillas, J. y Delgado, O. (2005). Una Metodología de KDD basada en Sistemas Difusos Genéticos Multiobjetivo para Modelado Causal en Marketing. *IV Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados*, MAEB 2005. Granada, España.
- Ferreira, D. y Mira. M. (2005). *Using process mining for itil assessment: a case study with incident management. Organizational Engineering Center, INESC-INOV*. Recuperado de <http://web.tecnico.ulisboa.pt/diogo.ferreira/papers/ferreira08using.pdf>
- Giraldo, J. y Builes, J. (2013). Caracterización del Proceso de Obtención de Conocimiento y Algunas Metodologías para Crear Proyectos de Minería de Datos. *Revista Latinoamericana De Ingenieria De Software*,(1),42–44. Recuperado de <http://sistemas.unla.edu.ar/sistemas/redisla/RELAIS/relais-v1-n2-p-42-44.pdf>
- Hernández, J., Ramirez, M., Ferri, C. (2007). *Introducción a la Minería de Datos*. Madrid, España: Pearson.
- Microsoft Corporation. (2010). *Introducing Business Intelligence.Development Studio*. Recuperado de <http://technet.microsoft.com/es-es/library/ms173767.aspx>
- Moine J., Gordillo S. y Haedo, A. (2001a). Análisis comparativo de metodologías para la gestión de proyectos de Minería de datos. *Procedding VIII Workshop Bases de datos y Minería de datos*. 931-938. Recuperado de <http://sedici.unlp.edu.ar/handle/10915/18749>
- Moine J., Gordillo S. & Haedo, A. (2001b). Estudio comparativo de metodologías para minería de datos. *XIII Workshop de Investigadores en Ciencias de la computación*. 278-281. Recuperado de <http://www.scieosp.org/pdf/rsap/v11n4/v11n4a16.pdf>
- Vallejos, S (2006). *Trabajo de Adscripción en minería de datos*. Universidad Nacional del Nordeste. Argentina. Recuperado de http://exa.unne.edu.ar/informatica/SO/Mineria_Datos_Vallejos.pdf
- SAS Institute Inc. (2002). *Uncover gems of information*. Recuperado de <http://www.sas.com/products/miner/index.html>
- SAS Institute Inc. (2013). *Visualización del negocio determinada por la analítica de alto poder – para todos*. Recuperado de http://www.sas.com/content/dam/SAS/bp_es/doc/factsheet/SAS%20Visual%20Analytics_fact_sheet.pdf
- Sauter, V. (2014). *Decision Support Systems for Business Intelligence*. John Wiley & Sons: University of Missouri, Columbia, USA.
- Arisa Shollo, A. Y Galliers, R. (2015). Towards an understanding of the role of business intelligence systems in organisational knowing. *Information System Journal*
- Soroush, S. & Bameni, M. (2009). *A Proposed Data Mining Methodology and its Application to Industrial Procedures*. Recuperado de http://www.qjie.ir/pdf_31_2e-9f779810eaef02d9bcc00959616080.html.
- Soto, E. (2001). *Data Warehouse*. Antecedentes, Situación Actual y tendencias. Departamento de Economía y Dirección de Empresas. Documento de trabajo. Universidad de la Laguna. España. Recuperado de http://iude.webs.ull.es/investigacion/publicaciones/pdf_docs_trabajo/SERIE%20ESTUDIOS%200144.pdf



“El éxito
consiste en obtener
lo que se desea
la felicidad
en disfrutar lo que
se obtiene”

EMERSON.