



## Autores

Miletti, Mabel

Aquel, Sandra

Berri, Ana María

Díaz, Daniel

Doffo, Gonzalo L.

Ingrassia, Ramiro

Marchese, Alicia

*Instituto de Investigaciones Teóricas y Aplicadas*

## **Multidimensionalidad de los Datos en la Información Contable. Su posible aplicación**

### **1. Introducción**

Nuestra inquietud al encarar el trabajo de investigación estuvo centrada en determinar un proceso de relación entre los datos que captan el sistema contable, tal que permita extraer conclusiones para pensar en el futuro, naturalmente todo ello contando con la inestimable ayuda de las herramientas que la tecnología nos brinda hoy en día.

El recorrido Información-Tecnología que ya ha demostrado sus ventajas y beneficios ha sido y es nuestro gran aliado. Es así como se pudo afirmar cómo las herramientas de procesos de análisis de datos en línea (OLAP) que actúan sobre los almacenes de datos nos muestran que un mismo dato puede ser visto desde distintas perspectivas del negocio, siendo el dato originario "único", hace al dato reutilizable para distintos fines.

Lo nuestro intenta lograr que utilizando la misma base, puedan desagregarse datos y utilizando como herramienta fundamental la tecnología se pueda implementar, junto al sistema contable, un sistema que permita concomitantemente una información que ayude al empresario a tomar decisiones para el futuro, aplicando lo que denominamos la "multifuncionalidad de datos". Donde trabajamos con el siguiente esquema primario:

*estados contables + información adicional digitalizada a través de datos multifuncionales*

Obviamente, lo digital permitirá una apertura de la información y el ingreso de información desde otros sistemas.

### **2. Información Contable. Normas y Medios tecnológicos**

La Contabilidad en sentido genérico, el Sistema de Información Contable para ser más preciso, intenta en la emisión de informes que constituyen la "salida" del sistema, informes que están diseñados en base a Normas específicas, las llamadas Normas Contables Profesionales. Estas normas, y no desde hace mucho tiempo, en cierto modo han intentado "abrir" los datos que mostraba el sistema, con el objetivo de hacerlos más precisos y comprensibles al lector (inversor, administradores, etc) mostrando información más detallada, que permite tomar otro tipo de decisiones. Nos estamos refiriendo específicamente a la norma RT18 de la FACPCE en su parte referida a Información por Segmentos, información que es requerida a las sociedades que coticen en Bolsa o hayan solicitado autorización para hacerlo.

La norma considera dos tipos de segmentos:



**A. Segmento de un negocio:** Lo identifica como las partes distinguibles de una empresa que proveen productos o servicios relacionados pero sujetos a riesgos y rentabilidades distintas a los de otros segmentos del mismo negocio, y para determinar cuándo los productos o servicios están relacionados determina que deben considerarse:

- Su naturaleza
- Sus procesos productivos
- Tipo o clase de clientes que adquieren los productos o servicios
- Métodos seguidos para la distribución de los productos o servicios

Las normas diferenciadas que pudieren regular alguna de las actividades efectuadas (por ejemplo la bancaria o la aseguradora)

**B. Segmentos geográficos:** considera en este caso, a aquél que proporciona productos o servicios en un contexto económico particular y que está sujeto a riesgos y rentabilidades distintos a la de otros segmentos geográficos (dentro de un país o región, o de un grupo de países). En este caso deben considerarse los siguientes factores para determinar los segmentos:

- Las similitudes entre las condiciones económicas y políticas
- Las relaciones existentes entre las operaciones en diversas áreas geográficas
- El grado de proximidad de las operaciones
- Los riesgos asociados con las operaciones efectuadas en cada área
- Las regulaciones establecidas en materia de control de cambios
- Los riesgos subyacentes en la realización de transacciones en determinadas monedas

Estos segmentos pueden basarse en la localización de los activos que producen los bienes o servicios (Base Activos), o bien en los mercados donde son destinados, es decir, según donde se realicen las ventas (Base Clientes).

La norma, asimismo, establece las condiciones cuantitativas para determinar el segmento e indica también, con el objetivo de definir la información a proporcionar, la división de los segmentos en primarios y secundarios, y el cómo asignar los activos, pasivos, ingresos y gastos en cada sector o segmento.

Como puede observarse, a nivel contable se viene reconociendo que si bien la información global sobre el patrimonio de la empresa y sus resultados reviste una importancia trascendental, que permite conocer el potencial económico total de la empresa, el disponer de información desagregada en diferentes segmentos de negocio o geográficos otorga una herramienta muy valiosa para una mejor comprensión de los Estados Contables en su conjunto y para la toma de decisiones.

### **3. Aplicando conceptos.**

Ahora bien, este ejemplo que nos sirve para visualizar el cómo la desagregación de datos constituye un instrumento valioso que ayuda a la gerencia a tomar decisiones que hacen al futuro del negocio, ayudándonos con la tecnología queremos trasladarlo a un nivel mucho más amplio.

#### **3.1 Caso: Comercialización de productos de marroquinería**

##### **3.1.1 Objetivo del caso**

Como objetivo general se intenta demostrar la relación entre el la información por segmentos y el datawarehouse.



En consecuencia, en el caso se planteará:

- 1) Diseñar un modelo de minería de datos en base a la información del área de ventas,
- 2) La aplicación de herramientas informáticas que permitan desarrollar e implementar el modelo anterior,
- 3) Diseñar el datamart para el área de ventas.

### 3.1.2. Metodología

La implementación del data mart (DM) se realizará sin la existencia de un datawarehouse. En consecuencia, se adopta la arquitectura bottom-up, en la cual primero se definen los DM y luego se integran a un datawarehouse centralizado.

### 3.1.3 Análisis de requerimientos

Una empresa que comercializa productos de marroquinería a nivel local y nacional necesita desagregar la información sobre sus ventas que obtiene del sistema transaccional respectivo. La **estructura de datos** de las ventas que realiza la empresa es la siguiente:

- ID
- CodArticulo
- Medida
- NroFactura
- Fecha
- Cliente
- Cantidad
- Importe
- Costo
- Rubro
- CodProveedor
- Prov\_cliente
- Prov\_prov

El campo (o variable) 'Importe' es aquel que cuantifica la transacción, mientras que el resto 'describe' o da significado a dicho campo<sup>1</sup>. Por consiguiente, al campo 'Importe' se lo considera el campo medida (measure) y la tabla en dónde se encuentre dicho campo se denomina tabla de "hecho".

A continuación, se debe establecer la dimensión o la descripción para dicha tabla que en el lenguaje contable se denomina "segmento". Dado que se utiliza una herramienta informática, podría establecerse n dimensiones o segmentos y de forma intuitiva e iterativa escoger aquellos que mejor describan las ventas. Sin embargo, se estaría vulnerando el principio de eficiencia y eficacia. Por tal motivo, se propone el diseño de un modelo de minería de datos, a través del cuál, se establecerían las dimensiones correctas.

### 3.1.4 Diseño de la minería de datos

Como punto de partida del diseño anterior, debe preguntarse qué aspecto de la venta tiene

---

<sup>1</sup>Más allá que puedan ser un código y la descripción completa esté en otra tabla. Pues la vinculación se dará a través del referido campo.



relevancia para quién debe decidir sobre esta temática. En el presente caso, la inquietud que se plantea es sobre la procedencia de los clientes. La empresa tiene clientes en la mayoría de las provincias y quiere conocer las características particulares de estos. De esta manera, el campo que tiene el origen de los clientes es: Prov\_cliente. Dicho campo puede asumir los siguientes valores:

- Buenos Aires
- Corrientes
- Córdoba
- Mendoza
- Neuquén
- San Luis
- Santa Fe
- Santiago
- Ausente (se perdió el dato sobre la procedencia del cliente)

Luego, para el presente caso, se estipuló que las características particulares que se desean saber son:

1. El momento en el cual los clientes efectúan la compra,
2. La mercadería que se compran dichos clientes, y,
3. El monto de la venta que se le realizó a dicho cliente.

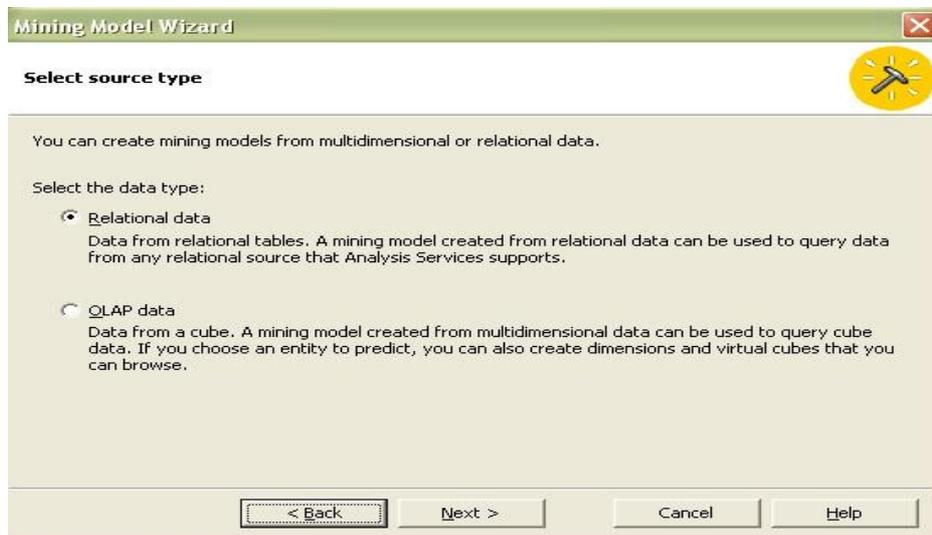
Las características anteriores deben traducirse o relacionarse en función de la estructura de datos antes mencionada. Es decir, el **momento de la compra** está dado por el campo '**Fecha**'; la **mercadería** está dada por el '**Rubro**'; y por último, el **monto de la venta** está dado por el campo '**Importe**'.

Como resultado, en función de la combinación de los distintos valores que asumen los campos: fecha, rubro e importe, se obtendrá la caracterización de la procedencia de los clientes.

### 3.1.5. Implementación

Para realizar el modelo de minería de datos se utilizó el módulo Analysis Manager de SQL Server 2000 (versión educativa)

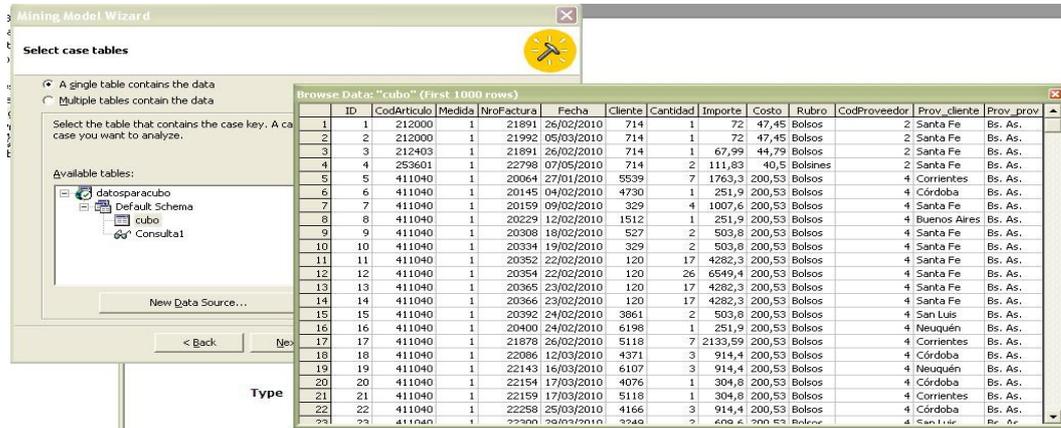
El primer paso es seleccionar la fuente de datos que utilizará la herramienta de minería de datos, la cual corresponde a la base de datos relacional 'datosparacubo'.



**Ilustración 1:** Asistente del modelo de minería de datos. En este cuadro de diálogo se debe seleccionar la fuente de datos, las opciones son: Datos relacionales (por ej.: base de datos) y Datos OLAP (cubos). Fuente: SQL Server 2000



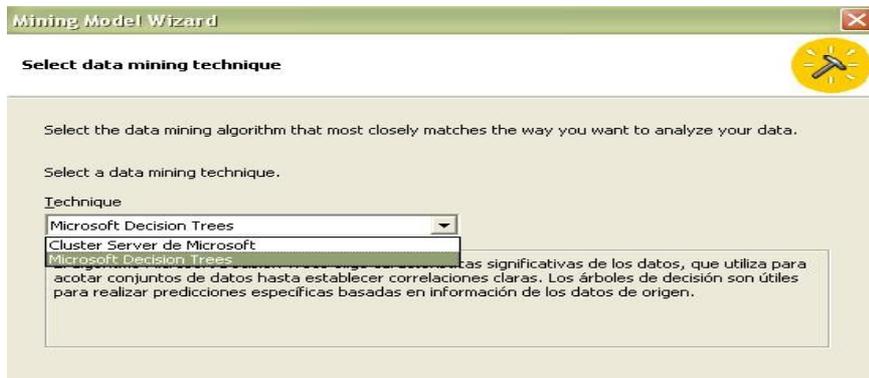
Posteriormente, se debe indicar la tabla que contiene a los campos a operar.



**Ilustración 3:** Cuadro de diálogo del Asistente para el modelo de minería de datos. Con ambas figuras ('Select case tables' y 'Browse Data: "cubo" (First 1000 rows)') se visualiza la selección del origen de datos a utilizar y su contenido. Fuente: SQL Server 2000.

En la figura anterior, en el cuadro de diálogo 'Select case tables' puede visualizarse que la información sobre las ventas está en una sola tabla denominada Cubo. Y al lado, muestra los datos y los valores que asume ('Browse Data: "cubo" (First 1000 rows)')

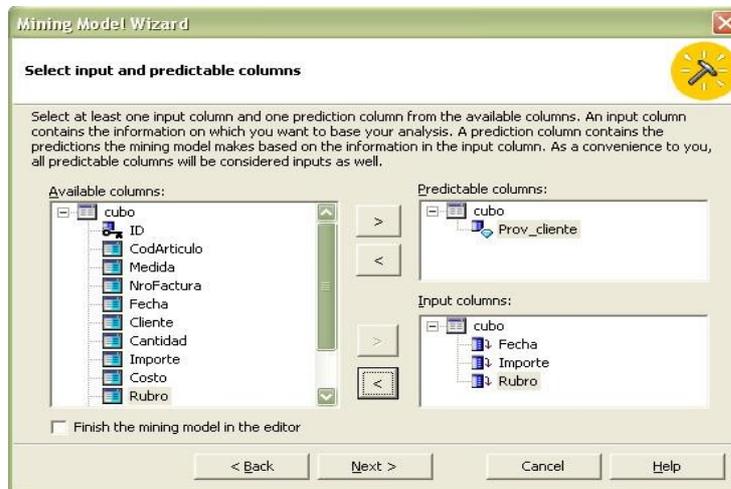
A continuación, se debe seleccionar el algoritmo de minería de datos. El referido software ofrece dos alternativas: árbol de decisión y clusters. Para este caso se selecciona el algoritmo 'árbol de decisión' porque se trabaja con el campo 'Rubro' el cual es un campo discreto y es en definitiva aquel primordial para saber qué producto venderle a cada uno de los clientes. Los árboles de decisión son mejores que otras técnicas de clasificación al trabajar con campos 'disjuntos'.



**Ilustración 4:** Cuadro de diálogo del Asistente para el modelo de minería de datos. En esta instancia se puede seleccionar los algoritmos para la minería de datos. En la versión 2000 del SQL Server sólo se tiene: árboles de decisión (Microsoft Decision Trees) y conglomerados (Cluster Server de Microsoft). Fuente: SQL Server 2000.



Luego, se establece que el campo 'ID' es la clave principal (evita duplicaciones en las transacciones comerciales bajo análisis) Y después, se debe especificar la variable a 'predecir' y las variables de 'entrada', como se indica a continuación.



**Ilustración 5:** Cuadro de diálogo del Asistente para el modelo de minería de datos. En esta instancia se seleccionan las variables a predecir (Predictable columns) y las variables de entrada (Input columns). *Fuente:* SQL Server 2000.

La variable a predecir es aquella que refleja la procedencia de los clientes, dado que es esta última a la cual se quiere encontrar su caracterización. Y las variables que se utilizan para esto últimos, son la fecha, el importe y el rubro, que son las variables de entrada.

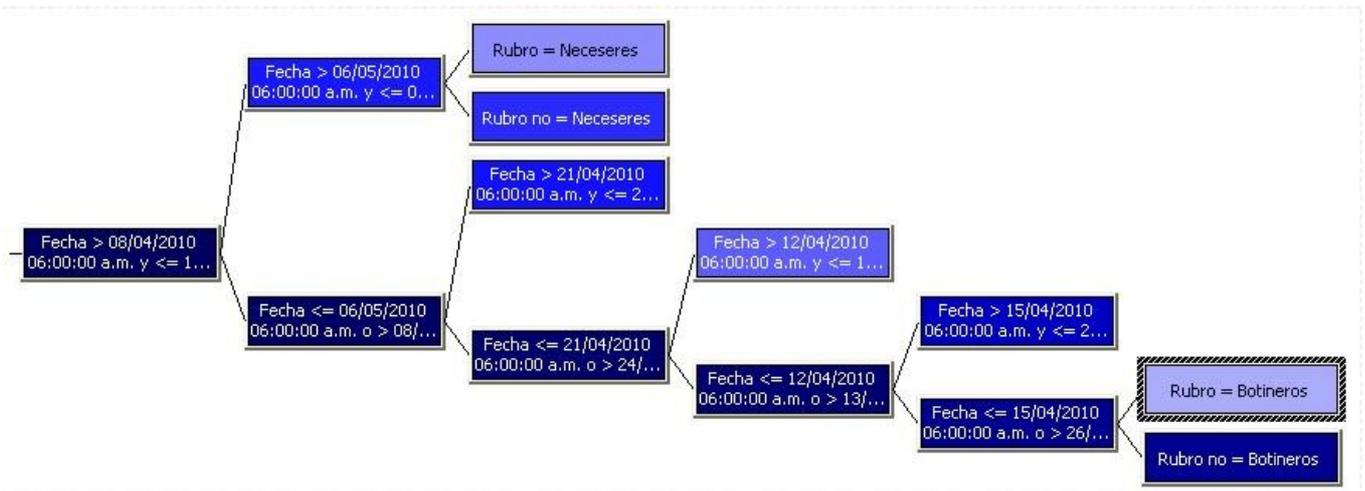
Como resultado se obtienen los siguientes gráficos:

Attributes		
Totals   Histogram		
Value	Cases	Probability
<b>(Node Total)</b>	<b>28</b>	<b>100.00%</b>
Buenos Aires	0	2,70%
Corrientes	18	51,35%
Córdoba	0	2,70%
Mendoza	0	2,70%
Neuquén	0	2,70%
San Luis	0	2,70%
Santa Fe	10	29,73%
Santiago	0	2,70%
ausente	0	2,70%

**Ilustración 5:** Se visualiza la variable a predecir, el origen provincial del cliente. En el ejemplo, toda la descripción del árbol es para aquellos clientes provenientes de Corrientes (en un 51,35% de probabilidad) y en menor medida para los de Santa fe (29,73% de probabilidad). *Fuente:* SQL Server 2000.

Node Path
Fecha ( > 08/04/2010 06:00:00 a.m. y <= 12/04/2010 06:00:00 a.m. ) o ( > 13/04/2010 06:00:00 a.m. y <= 15/04/2010 06:00:00 a.m. ) o ( > 26/04/2010 06:00:00 a.m. y <= 06/05/2010 06:00:00 a.m. ) o ( > 08/05/2010 06:00:00 a.m. y <= 10/05/2010 06:00:00 a.m. ) y
Rubro = Botineros

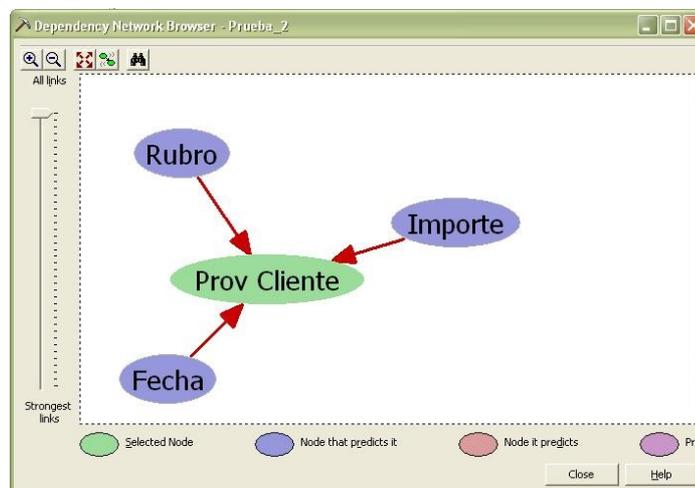
**Ilustración 6:** Visualización analítica del árbol de decisión. **SI** la venta se produjo en los períodos descriptos en la regla (o en un árbol como en la Ilustración 5) y el rubro objeto de la venta son 'Botineros' **ENTONCES** la procedencia del cliente es de Corrientes en un 51,35% **SINO** es de Santa Fe en un 29,73%. *Fuente:* SQL Server 2000.



**Ilustración 7:** Árbol de decisión. Las distintas 'ramas' indican el detalle de las características de la variable a predecir (Provincia Clientes). El detalle se basa en las variables de entrada: Fecha, Importe y rubro. Fuente: SQL Server 2000.

En el ejemplo planteado, en el cuadro 'Attributes' se visualiza la variable a predecir, en este caso, la provincia de los clientes. En el 'node path' se muestra las condiciones para las características que se muestran en 'Attributes', y en forma gráfica en el árbol de decisión de la parte superior. Entonces, con una probabilidad del 51,35% para los clientes de la provincia de Corrientes y del 29,73% para los clientes de la provincia de Santa fe; compran 'Botineros' en los períodos: del 8 al 12 de abril, del 13 al 15 de abril, del 26 de abril al 16 de mayo, y, del 8 al 10 de mayo.

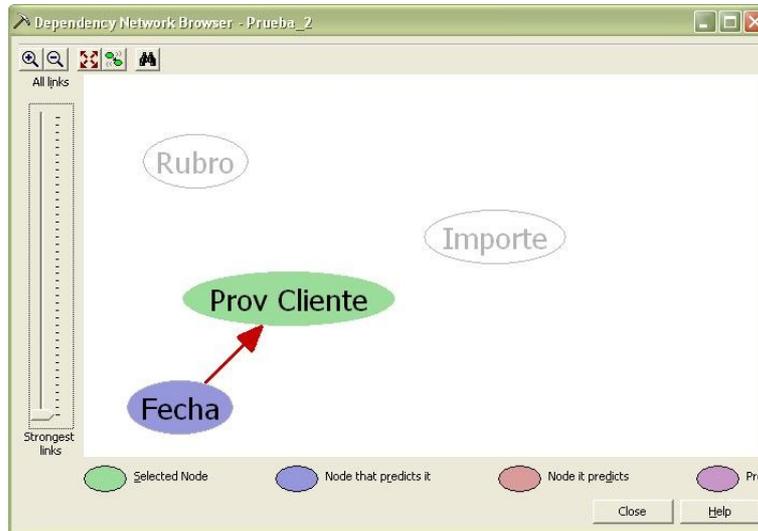
Sin embargo, la finalidad del análisis anterior no sería establecer "reglas de comportamiento" sino los segmentos a utilizar para emitir la información. Para ello, se presenta la siguiente red de dependencia entre las variables utilizadas. En este gráfico se visualiza el grado de fuerza de relación entre las anteriores.



**Ilustración 8:** Red de relaciones en donde se visualizan las dependencias entre variables sin exigir ninguna 'robustez' en la relación. Fuente: SQL Server 2000.



En esta instancia se presentan todas las variables del modelo. A continuación, se exigirá la mayor fuerza de relación entre estas.



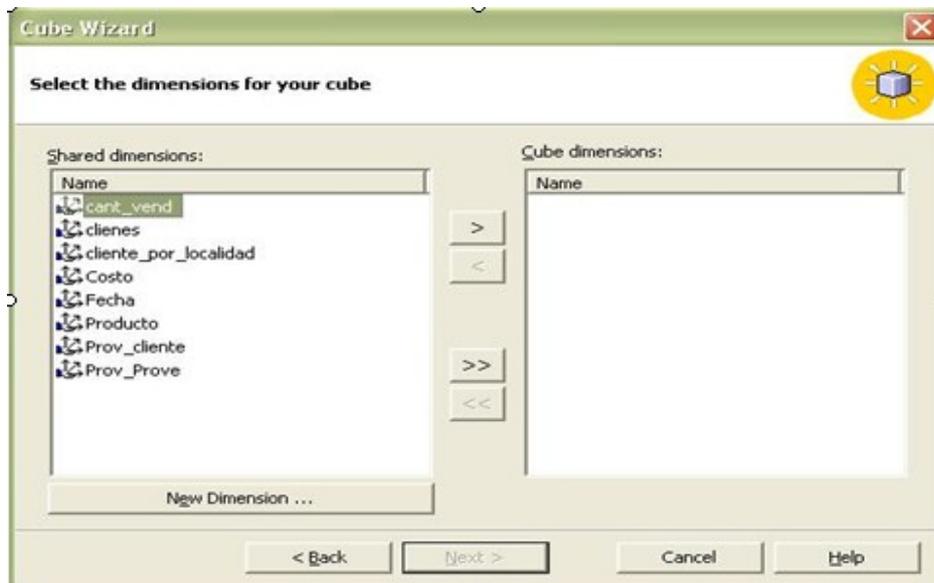
**Ilustración 9:** Al exigirse la mayor fuerza de relación entre las variables, la fecha es un atributo que predice el con mayor probabilidad el origen del Cliente (Prov Cliente). A partir de esta conclusión, se puede estimar que las ventas responden a cierta 'estacionalidad' con respecto a la procedencia geográfica del cliente. Fuente: SQL Server 2000.

En esta segunda instancia, la relación que se sigue sosteniendo es entre la fecha y la provincia de origen del cliente. Por tal motivo, las dimensiones que mejor explican las ventas según el origen de procedencia del cliente es la variable fecha.

### 3.2 Datamart

El desarrollo del datamart es relativamente sencillo porque los datos están en un único formato y en una única base de datos. El diseño del datamart es similar al de minería de datos.

Primero, se indicará la fuente de datos a utilizar y cuál es el campo que se utilizará como "medida" (measure). Luego, se seleccionan las dimensiones apropiadas:



**Ilustración 10:** Cuadro de diálogo del Asistente para el modelo de minería de datos. En esta instancia se debe seleccionar la variable que mejor describe las ventas por origen del cliente. En este caso, es la variable 'fecha'. Dicho atributo se encuentra en 5º lugar en el listado 'Shared dimensions'. Fuente: SQL Server 2000.



Las ventas deberán "describirse" a partir de la procedencia de los clientes (objetivo), la fecha, producto. Al final del proceso, la información se presentará de la siguiente manera:

Prov_cliente			All Prov_cliente
<ul style="list-style-type: none"> <li><input checked="" type="radio"/> All Prov_cliente</li> <li><input type="radio"/> Buenos Aires</li> <li><input type="radio"/> Córdoba</li> <li><input type="radio"/> Corrientes</li> <li><input type="radio"/> Mendoza</li> <li><input type="radio"/> Neuquén</li> <li><input type="radio"/> San Luis</li> <li><input type="radio"/> Santa Fe</li> <li><input type="radio"/> Santiago</li> </ul>			

			MeasuresLevel	
- Year	- Quarter	+ Month	Importe	
All Fecha	All Fecha Total		6.640.075,80	
	2010 Total		6.640.075,80	
- 2010	- Quarter 1	Quarter 1 Total	3.624.927,34	
		+ Enero	468.314,81	
		+ Febrero	1.380.584,53	
			+ Marzo	1.776.028,00
	- Quarter 2	Quarter 2 Total	3.015.148,46	
		+ Abril	1.560.433,73	
+ Mayo		1.454.714,73		

**Ilustración 11:** Tabla en donde se desagrega la información sobre las ventas, primeramente por origen del cliente (menú desplegable) y posteriormente por fecha (cuadro). El tiempo se expresa a través de diferentes unidades de tiempo: año (year), cuatrimestre (quarter) y mes (month). La secuencia anterior no es aleatoria, se va de lo general a lo particular para 'refinar' el análisis. Este procedimiento se denomina 'drill-down'. Fuente: SQL Server 2000.

Con la barra desplegable 'Prov\_cliente' se selecciona las provincias, y de esta forma se filtran las ventas de la tabla inferior. Y en la tabla inferior se combina los importes de venta con la fecha. También podría colocarse como lista desplegable la fecha y en el cuadro mostrarse las ventas por provincia. La importancia radica en la facilidad de clasificar la información como una "tabla dinámica".

#### 4. Conclusión

La presente experiencia, a partir de datos contables ingresados en dos dimensiones, nos permite extraer información particionada (segmentada) para ayudar a la toma de decisiones gerenciales.

En realidad, no es más que la automatización del proceso que haría un ser humano leyendo los distintos informes y extrayendo conclusiones u obteniendo "imágenes" de la realidad. El aporte de las herramientas utilizadas consiste en la obtención automática de la información numérica, exacta y oportuna.

Párrafo aparte merecería el entorno requerido para que estos elementos se utilicen de manera eficiente, ya que se requiere de cierta madurez organizacional e informática. Esto es, no se trata sólo de adquirirlos, sino que además, se debe contar con un buen sistema de procesamiento de transacciones que le brinde soporte y, sobre todo, de un usuario capacitado para su formulación y aplicación a la estrategia de la organización: objetivos explícitos y capacidad decisional.

No obstante, consideramos que la formulación de los requerimientos debe ser parte sustancial de nuestra tarea como profesionales.



## 5. Bibliografía Consultada

HAX, A., WILDE, D., El modelo Delta: Una nueva perspectiva para los negocios. En su: El proyecto Delta. Bogotá, Grupo Editorial Norma, 2003. Pág. 5

<http://h20000.www2.hp.com/bizsupport/TechSupport/Home.jsp>

<http://www.bbiz.com/partners.htm>

MINTZBERG, H. Diseño de Organizaciones Eficientes. 2º ed. Bs. As. El ATENEO. 2003. 365p.

MITCHEL, Tom, "Machine Learning", McGraw Hill, New York, March 1997,

NUÑEZ PAULA, Israel. La gestión de la información, el conocimiento, la inteligencia y el aprendizaje organizacional desde una perspectiva socio-psicológica. ACIMED [online]. 2004, vol. 12, no. 3 [citado 2008-05-28], pp. 1-1. Disponible en: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1024-94352004000300004&lng=es&nrm=iso](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352004000300004&lng=es&nrm=iso). ISSN 1024-9435.

OCDE, EUROSTAT, Manual de Oslo. Guía para la recogida e interpretación de datos sobre innovación, Traducción española, 2005.

REVERTÉ, Ma., Propuesta para la Formulación de una Estrategia de Gestión del Conocimiento. Gestión del conocimiento. Com [en línea] Febrero 2001. [fecha de consulta: Diciembre del 2006] Disponible en: <  
<http://www.gestiondelconocimiento.com/documentos2/mgradillas/estrat.htm>>

Revista Expandit, Argentina, Año 2, Nro. 14, pág. 53