



# UNA MIRADA AL BIG DATA Y AL ANÁLISIS DE CONJUNTOS

---

Manuela Linares Barbero  
Escuela de Ingeniería de la Universidad de Lima. Lima, Perú

Recibido: 28 de setiembre de 2015 / Aprobado: 16 de octubre de 2015

## Resumen

*En el presente artículo se busca exponer la utilización de big data en análisis de conjuntos (conjoint analysis). Las empresas deben desarrollar nuevos productos de forma constante, para lo cual usan investigaciones de mercados, identificando así los gustos y las preferencias de los clientes. En la actualidad, el análisis de conjuntos es una de las herramientas más empleadas en la investigación de mercados, debido a que ayuda a reducir costos en el proceso. A través de la investigación descriptiva se revisan diferentes estudios que han aplicado el análisis de conjunto para levantar información y simularla con las herramientas de software como el Sawtooth, aprovechando la tecnología de big data. Se puede concluir que la información transaccional y la información social que se maneja con big data nos ayudan, de una forma más eficiente, en la utilización de análisis de conjuntos.*

*Palabras clave: big data / análisis de conjuntos / software Sawtooth*

## A Look at Big Data and the Conjoint Analysis

### Summary

*This article aims to expose the use of Big Data in Conjoint Analysis or analysis of sets (trade-off analysis). Companies are always developing new products, which is aided by the use of market data in order to research and identifying the preferences of their targeted customers. Currently, the cluster analysis is one of the most used tools within market research. This is because it helps cut costs in the process. Through descriptive research, various investigations using conjoint analysis can be reviewed to gather information and simulate software tools like Sawtooth, taking advantage of Big Data technology. It can be concluded that transactional and social information used with Big Data can efficiently aid researchers if used in conjunction with the conjoint analysis.*

*Key word: big data / conjoint analysis / software Sawtooth*

## Introducción

En el mundo de hoy la competencia es cada vez más intensa, por lo que para las empresas es clave mantener la innovación y el desarrollo de nuevos productos. En el pasado se ha visto cómo varias compañías, después de haber sido líderes en el mercado, han fracasado y hasta desaparecido por dejar de innovar.

Uno de los instrumentos más efectivos para el desarrollo de nuevos productos, y garantizar que estos contengan las preferencias y los gustos de los clientes, es el análisis de conjuntos (*conjoint analysis*), técnica estadística que se utiliza en la investigación de mercados. Esta metodología ha sido utilizada por las empresas desde los años setenta, pero no es hasta que la tecnología de *big data* aparece, hace tres décadas, que se explota su máximo potencial.

Más que determinar los gustos y preferencias de los clientes en su intención de compra, tal como lo hacen otras herramientas de la investigación de mercados, el análisis de conjuntos busca identificar los criterios con las prioridades que utiliza el consumidor para tomar la decisión de compra. Es decir, conocer cuáles son los gustos y las preferencias en los criterios de decisión, de forma que se modifique el producto para incluir en este los atributos relacionados con sus criterios de manera prioritaria.

### 1. *Big data*

Cuando se habla de *big data*, inmediatamente se piensa en una gran cantidad de información, pero ¿de qué información se está hablando? Se observa que compañías como Facebook y Amazon manejan grandes cantidades de información, pero ¿con qué fin? Y es aquí donde se aprecia una gran diferencia entre los datos obtenidos por ambas empresas. Facebook recopila datos como mensajes, fotos, conexiones de personas, etc., es decir, información de índole social y se basa en la creación y desarrollo de las redes sociales. Existen varias empresas que utilizan este tipo de datos, entre las cuales están Twitter, LinkedIn, Orkut, A Small World, etc.

En el caso de Amazon, recopila datos para la elaboración de predicciones: compras realizadas, listas de deseos, clics de búsqueda, etc. Esta información muestra los intereses y las preferencias de los consumidores, qué libros compran y cuáles quieren adquirir en un futuro (lista de deseos). Estos datos son analizados por Amazon para desarrollar un perfil de consumidor de manera casi personalizada, con el fin de ofrecerle otros productos adicionales, buscando incrementar sus ventas. En la actualidad, muchas empresas utilizan esta información para desarrollar perfiles de consumidor e intentar prever sus acciones futuras. En este ámbito, el análisis de la *data* puede ser muy provechoso para enriquecer la recopilación de los datos que necesita la investigación de mercados.



Desde el inicio del tercer milenio, el término *big data* ha alcanzado gran popularidad y es usado por personas de diferentes sectores industriales y profesiones, especialmente por quienes trabajan en el ámbito de la tecnología. No obstante, existen varias opiniones en lo que se refiere a su definición exacta. La tabla 1 muestra definiciones formales relacionadas al *big data*, brindadas por dos grandes empresas del ramo tecnológico.

**Tabla 1**  
**Definiciones de *big data***

Entidades	Definición
Oracle	Es la derivación de valor a partir de la base de datos relacional tradicional. Es impulsada por la toma de decisiones de negocio y aumentada con nuevas fuentes de datos no estructurados. Estas nuevas fuentes incluyen: blogs, redes sociales, redes de sensores, datos de imagen y diversas formas de datos que varían en tamaño, estructura, formato y otros factores.
Microsoft	Es el término que más se usa para describir el proceso de utilizar un fuerte poder de computación, lo último en aprendizaje automático e inteligencia artificial, así como una cantidad masiva y muchas veces complejos conjuntos de información.

Fuente: Ward y Barker (2014)

De estas definiciones se aprecia que cuando se habla de *big data* se hace referencia a una combinación de análisis de los datos, tecnologías relacionadas (inteligencia artificial) y las organizaciones que generan o incentivan la creación de enormes cantidades de datos, ya sea para analizarlos ellas mismas, venderlos o compartirlos de forma privada o pública. Como Ward y Barker (2014) mencionan, las empresas que manejan *big data* crean aproximadamente 300 *terabytes* de datos por semana. Para poder almacenar esta cantidad de datos digitales con la velocidad de creación mencionada, el avance tecnológico es de vital importancia, especialmente en la reducción de costos de almacenamiento. En las figuras 1 y 2 se puede notar cómo el costo del almacenamiento digital, ya sea en discos duros o cualquier otro medio, se ha reducido dramáticamente y continuará bajando.

Sin esta drástica reducción de los costos de almacenamiento, de más de 100 000 dólares por *gigabyte* en 1980 a menos de 10 centavos de dólar en 2015, *big data* habría sido imposible de solventar por las empresas: demasiado costoso para ser útil. "Hoy día, el coste de una unidad en la que se pueda almacenar toda la música

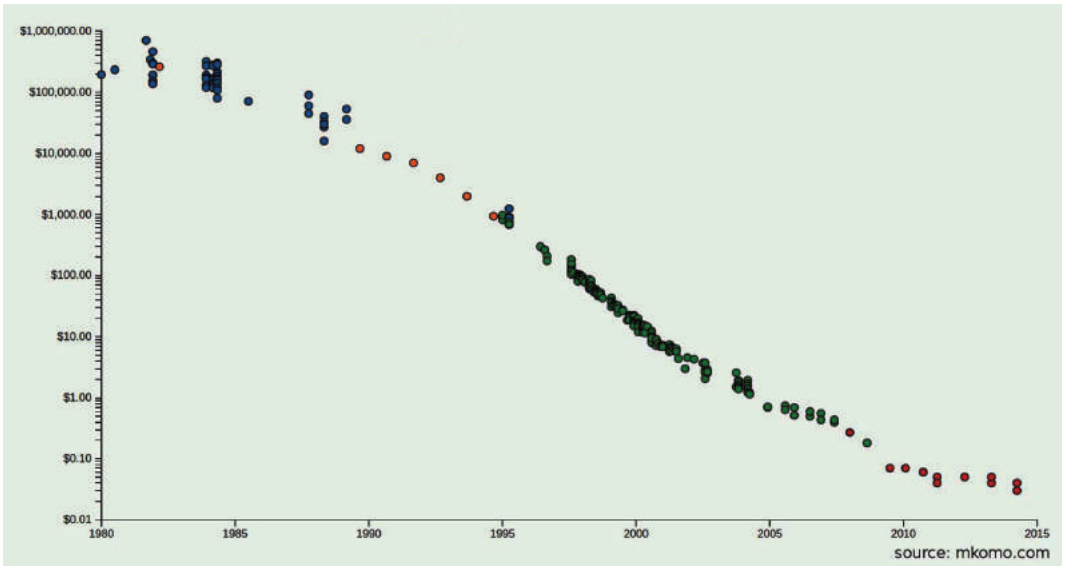


Figura 1. Costo por gigabyte en los discos duros (US\$)  
Fuente: Komorowski (2014)

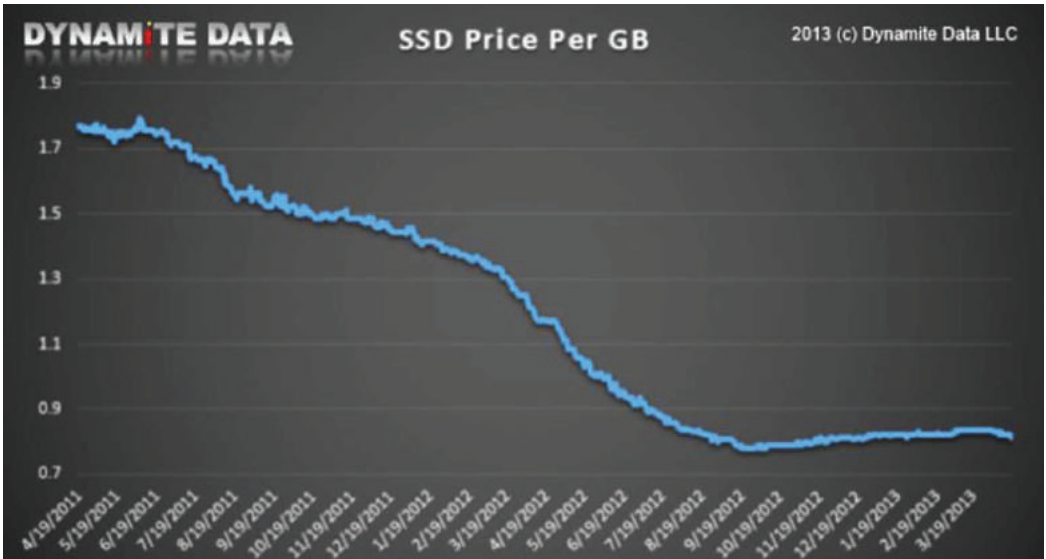


Figura 2. Costos por gigabyte de discos duros SSD (discos de estado sólido) (US\$)  
Fuente: Hruska (2013)

del mundo tiene un precio aproximado de solo 600 dólares” (Salvador, 2014). Esta información muestra el gran avance en almacenamiento de datos, así como en el *software* para manejarlo como algoritmos de búsqueda. Al medir el análisis de datos, con frecuencia se utilizan las siguientes variables: volumen, velocidad, variedad y complejidad. Estas variables nos dan una mejor aproximación a la información que se requiere gestionar. La evolución en la gestión de los datos se aprecia en la figura 3.

La carrera en el aprovechamiento de los datos y su análisis en información relevante y útil para facilitar la toma de decisiones ha tenido varias etapas. El avance tecnológico permite hoy en día a las empresas reducir sus costos en almacenamiento de las enormes cantidades de datos transaccionales, guardándolos en la nube con tecnologías *infrastructure as a service* (IaaS) o infraestructura como servicio (Polyviou, Pouloudi y Rizou, 2014), con proveedores (entre ellos Amazon) que ofrecen este tipo de servicios.

Para tener una idea referida al volumen, variedad y velocidad de datos, cuando se habla de *big data*, en la figura 4 se muestran los datos correspondientes al año 2014.

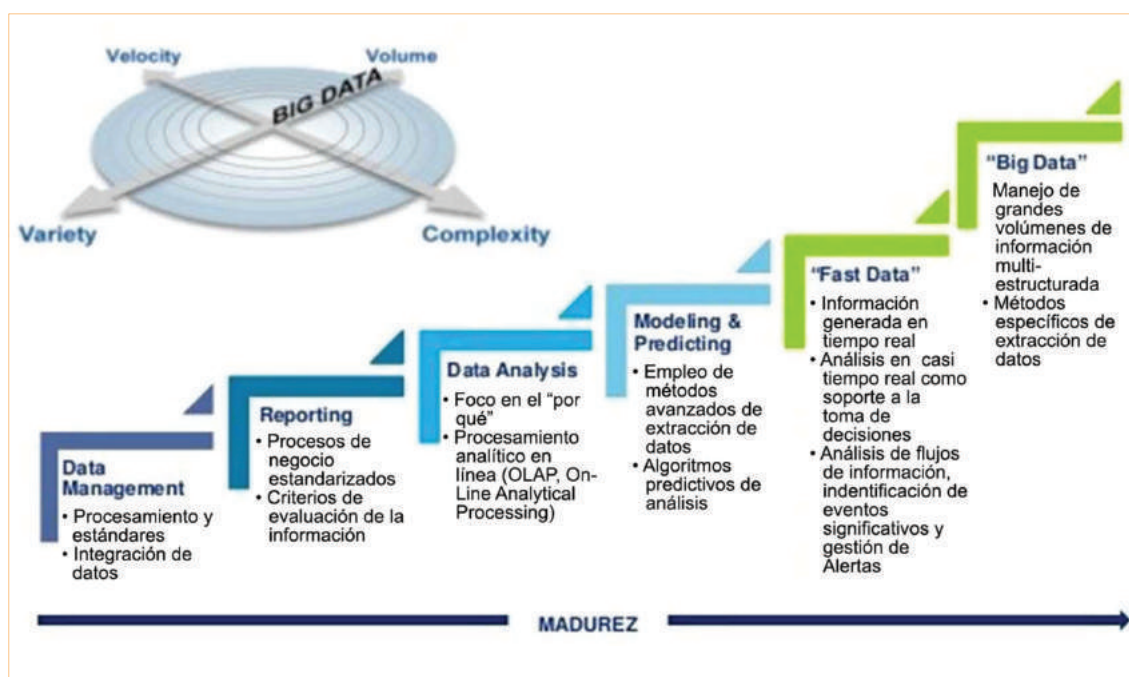


Figura 3. La evolución en la gestión de los datos  
Fuente: Machado (2014)

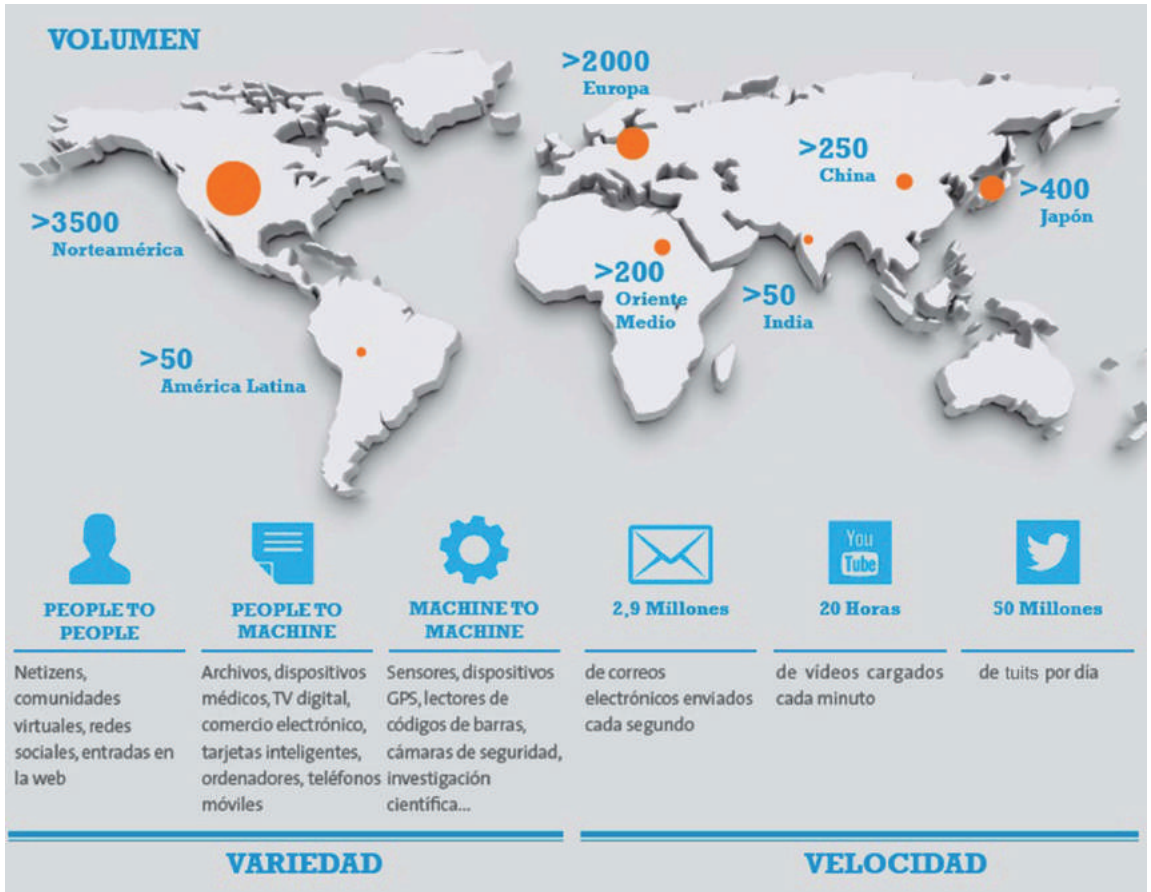


Figura 4. Volumen, variedad y velocidad de *big data*, 2014 (en *pegabytes*, PB)

Fuente: Salvador (2014)

Este volumen de datos solo es posible, como se ha visto, porque el costo de almacenamiento se ha reducido. Esto también ha dado cabida a la gran variedad de los datos no estructurados, como voz, video, etc., los cuales se manejan en la nube y se pueden acceder, que requieren una mayor velocidad de conexión y disponibilidad con la nube.

Las empresas tratan de obtener provecho de esta enorme cantidad de información que *big data* dispone, para lo cual se han diseñado aplicaciones de *software* como Customer Relation Manager (CRM), que permiten generar perfiles de clientes y facilitar la labor del *marketing* en segmentación de clientes, al igual que identificación de oportunidades de negocio y reducción de costos publicitarios. Sin embargo, *big data* ofrece una cantidad mucho mayor de información que se puede utilizar con otras herramientas del *marketing*, como el análisis de conjuntos (*conjoint analysis*).



La tendencia actual del *marketing* está basada en mantener el foco de atención en el cliente o consumidor, por lo que es indispensable conocer sus gustos y preferencias en cuanto a los atributos de los productos, así como de toda la mezcla de *marketing*. En la actualidad el manejo de información de clientes y la utilización de *software* CRM son de gran ayuda para maximizar las ventas y, de esta forma, las utilidades. Entre las herramientas de *marketing* se encuentra el análisis de conjuntos, técnica que necesita un ingreso de datos de clientes, los cuales tradicionalmente son obtenidos mediante encuestas (Ringel y Skiera, 2014); pero con el manejo de datos de clientes se pueden cargar en forma de simulaciones, en función a la información que se maneja en *big data* de los gustos y preferencias de los consumidores; por ejemplo, por medio de algoritmos de texto mínimo, para analizar en línea los comentarios sobre los productos (Ringel y Skiera, 2014).

## 2. Análisis de conjuntos

El análisis de conjuntos o *conjoint analysis (trade-off analysis)* es una herramienta desarrollada para realizar investigación de mercados e identificar las preferencias de los consumidores de los diferentes atributos de un producto. Es decir, cuál es la valoración o prioridad que le da el consumidor a un atributo con respecto a otro atributo del producto (bien o servicio). Para identificar el valor percibido por el cliente de un atributo u otro en un producto, se realiza una encuesta. Estas encuestas no son simples sino complejas, donde no solo se obtiene información sobre la intención de compra del cliente sino también sobre los criterios de intención de compra del cliente (Talledo, 2011).

La obtención de los datos para el análisis conjunto se basa en encuestas donde se le muestra al cliente, de dos en dos, las opciones del producto con características completas; cada par con características completas en variadas opciones o diferentes atributos, de forma tal que escoge la opción que más le gusta. Luego, con métodos estadísticos de regresión y análisis de varianza (ANOVA), se analizan los datos para asignar un valor a cada atributo en cada opción del producto.

Tal como cita Radler (1993), el análisis de conjuntos fue desarrollado en los años sesenta por el psicólogo matemático Duncan Luce y el químico John Tukey, y sus algoritmos fueron perfeccionados, a fines de los años setenta, por Kruskal y Carmone. Las empresas han utilizado desde entonces esta técnica estadística, principalmente en el desarrollo de nuevos productos, en especial en la asignación de precios o en nuevas versiones de productos existentes o con nuevos servicios o atributos.

Para entender mejor el análisis de conjuntos, véase la figura 5, la cual nos muestra una cartilla donde aparecen cinco de todos los posibles atributos que tiene un helado (sabor, envase, bajo en grasas, artesanal o de fábrica y precio), así como las diferentes opciones de cada uno de estos atributos.



Figura 5. Cartilla de análisis conjunto de helados

Fuente: Ajjan Associates (s. f.)

Con la información de esta cartilla se elabora la encuesta, poniendo en cada pregunta dos opciones de helado diferentes; cada opción con el detalle de los cinco atributos. Por ejemplo, se pregunta al cliente qué prefiere: helado de chocolate en cono artesanal, regular, a \$2,00; o un helado de vainilla en vasito, bajo en grasas, de fábrica, a \$2,50. De esta manera se continúa con los diferentes atributos y sus opciones, hasta agotar todas las combinaciones y así obtener los datos de los clientes. Luego de analizar los datos con métodos estadísticos, por ejemplo el ANOVA, se muestran los resultados, como se ve en la figura 6.

En los resultados se observa que el precio es el atributo más importante para el cliente en la toma de decisión en la compra de helados.

En la figura 7 se puede ver el ejemplo de una pregunta en una encuesta de análisis de conjuntos para televisores, donde se agrupan de tres en tres los televisores con diferentes atributos y opciones.



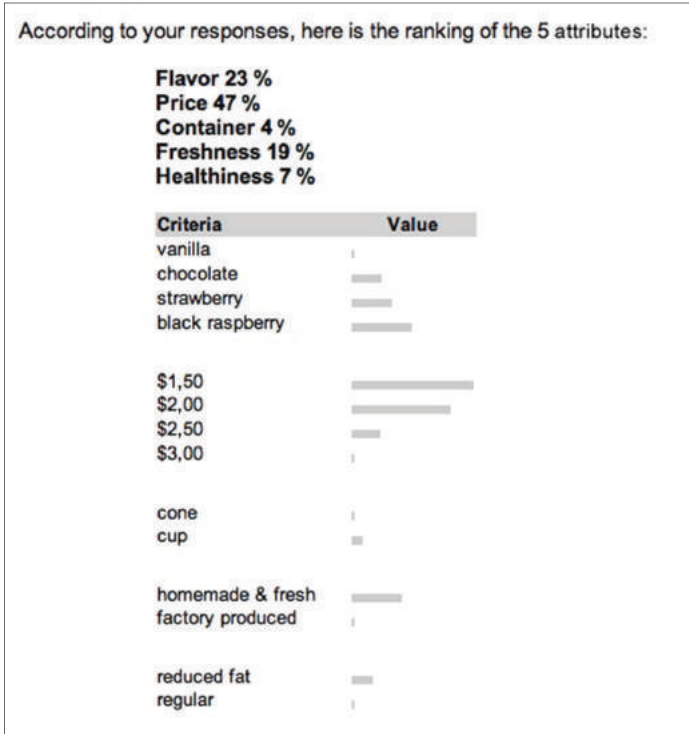


Figura 6. Resultados de análisis de conjuntos de helados  
 Fuente: Ajjan Associates

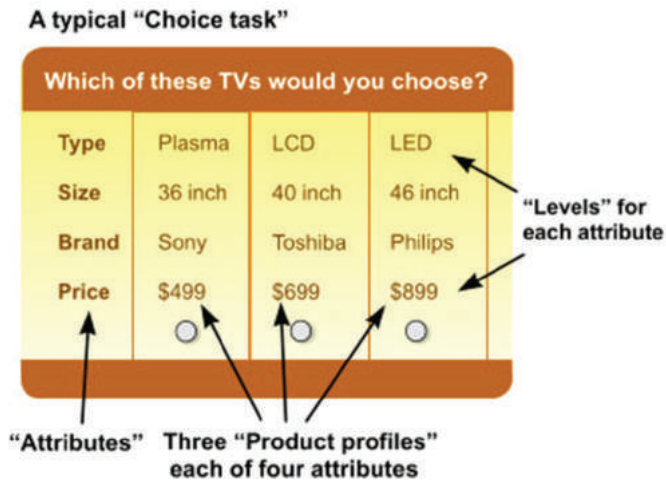


Figura 7. Pregunta de encuesta de análisis de conjuntos de televisores  
 Fuente: Dobney (s. f.)

Si bien la elaboración y el desarrollo de las encuestas de análisis de conjuntos –como en los ejemplos vistos– son más complejos que las encuestas simples y, por lo mismo, más costosos, son de mayor utilidad al momento de desarrollar nuevos productos que tienen múltiples atributos y distintas opciones de cada atributo, así como para definir su precio.

De igual forma, se han realizado investigaciones de mercado para mejorar servicios, como la que se presentó en la VII Reunión Nacional de Currículo. I Congreso Internacional de la Calidad e Innovación en Educación Superior, en Caracas, Venezuela, en abril de 2007. Dicha investigación se elaboró con la metodología de análisis de conjuntos del valor, en la cual se especificaron atributos al servicio de la educación virtual industrial. Su finalidad fue identificar los atributos que los estudiantes veían como de mayor valor, para, de acuerdo con ello, diseñar el curso virtual ideal. Entre los resultados de este estudio se advirtió que uno de los atributos más importantes era el tiempo de duración, lo mismo que el sistema de evaluación. Además, que el contenido no era tan relevante, ya que los estudiantes consideran que pueden obtener la información buscando en diferentes fuentes.

En palabras de Harrar de Dienes y Alcaide (2007), “La metodología del análisis de conjuntos representa un avance importante en las herramientas de investigación de mercados utilizadas hasta hoy para el diseño instruccional en el Tercer Nivel” (p. 13).

El análisis de conjuntos es utilizado no solo por las empresas para desarrollar nuevos productos, sino también por investigaciones de universidades, tales como la Universidad de St. Gallen (Suiza), institución que realizó un estudio de las preferencias de los clientes sobre el servicio de electricidad utilizando el método de Choice Based Conjoint (CBC) del análisis de conjuntos. Se identificaron, de esta manera, las preferencias de los clientes en cuanto al servicio de electricidad, para así desarrollar un nuevo servicio con los atributos que estos prefieren y lanzar de forma exitosa el nuevo servicio de electricidad. En la tabla 2 se presenta la cartilla con los atributos y las opciones o características de cada atributo.

Las grandes empresas almacenan una vasta cantidad de datos de sus propios clientes y sus productos, de forma que pueden definir sus preferencias y gustos basados en patrones de compra, cantidades, métodos de pago, etc. Esta información histórica de la empresa se combina con la de redes sociales, clics de *likes*, intenciones de compra, clics en búsquedas de su página web o de otras páginas, listas de deseos, etc. Con esto se puede diseñar un análisis de conjuntos y así es posible prever respuestas de los clientes en cuanto a aceptación de productos, atributos particulares más apreciados del producto y precio esperado.

Creemos que la integración de múltiples fuentes de datos en formas innovadoras puede aumentar la fiabilidad, validez y generalización de los análisis de conjuntos en el futuro. La integración de los aspectos cualitativos y las reacciones emocionales de los consumidores con los datos de preferencias declaradas en la formación de las preferencias y elecciones es una importante vía de investigación. (Agarwal, DeSarbo, Malhotra y Rao, 2015, p. 33)

Estudiar los datos históricos de compra y las cifras del ingreso medio por usuario, nos permite diseñar análisis de conjuntos que definen y guían respuestas hacia al set de conjuntos (*trade-offs*) más relevantes para cada situación (Forsyth y Boucher, 2015). Esta combinación de información hace posible tener un panorama completo y realizar una proyección fundamentada no solo en información histórica de nuestra empresa sino en información integral del entorno. La promesa del futuro es manejar de forma más completa los datos sociales y emocionales dentro de criterios cualitativos, proyectándolos con mayor exactitud.

**Tabla 2**  
*Atributos del servicio de electricidad*

Dimensión	Atributo	Opciones
Producto	Domicilio del proveedor de servicios	Local
		Suiza
		Territorio vecino
		Este de Europa
	Combinación energética	Libre de CO <sub>2</sub> (nuclear, recursos renovables)
		Renovables (hidráulica, solar, etc.)
		Bajo costo (nuclear, fósil)
		Libre de energía nuclear (fósil, recursos renovables)
Servicio	Canal de servicio (servicio central)	Internet, teléfono y centro de servicio
		Internet y teléfono
		Internet
	Presentación del consumo de electricidad (servicio adicional)	Cada segundo
		Cada mes
		Cada 3 meses
	Servicios que permitan el ahorro de energía (servicio adicional)	Sin servicio
		Eventos, consultoría
		Portal de eficiencia
Precio	Precio por persona y por mes	30 francos
		40 francos
		50 francos
		60 francos
		70 francos

Fuente: Adaptado de Domigall, Albani y Winter (2014, p. 211, mi traducción)

### 3. Conclusiones

Para realizar una investigación de mercados, muchas empresas siguen utilizando la recolección de datos tradicional como observación, *focus groups*, pruebas de producto, etc. A pesar de ello, cada vez más empresas complementan esta recolección tradicional de datos con los recopilados a través de *big data*. Hoy en día no solo las grandes empresas tienen acceso a manejar grandes cantidades de información, y verse en el reto de analizar, de forma coherente, esta información para la toma de decisiones en el momento preciso. Una de las herramientas de *software* que puede ayudar en la gestión de la información transaccional de la web de la empresa es el Google Analytics, el cual tiene, en su versión gratuita, suficientes funciones para que una microempresa analice mejor a su cliente, creando perfiles y segmentándolo. El Google Analytics, así como otros *softwares* de análisis de datos, utiliza la estadística para manipular los datos y ofrecer informes.

En el mundo actual, la innovación es una constante que deben mantener todas las empresas. La empresa que no innova caerá en la obsolescencia, como le ha sucedido a muchas en el pasado. El análisis conjunto ayuda al desarrollo de nuevos productos, especialmente cuando estos son complejos y poseen gran cantidad de atributos que se le pueden agregar. Empresas transnacionales como General Motors, General Electric, Procter & Gamble, Google, etc., utilizan el *software* Sawtooth para el análisis de conjuntos y para prever los atributos más deseados por los consumidores, antes del lanzamiento de nuevos productos (Sawtooth Software, 2015). Se estima que se realizan más de 8000 análisis de conjuntos al año de los diferentes clientes de la empresa Sawtooth Software.

Herramientas de *software*, como el Sawtooth, nos ayudan a la simulación de diferentes escenarios, reduciendo así la cantidad de encuestas con clientes potenciales. Este *software* es capaz de simular las encuestas sobre la base de un valor designado a cada atributo y su opción del producto (Orme, 2010). Este valor proviene de los gustos y preferencias de los clientes y puede ser obtenido de forma tradicional con encuestas, o aprovechando la información transaccional y social de *big data*, donde no solo se puede identificar las preferencias de compra de los clientes, sino que se puede identificar la prioridad de los criterios de decisión del cliente con la información transaccional combinada con la social. Este valor que se busca asignar para realizar el análisis de conjuntos y sus simulaciones de escenarios radica justamente en estas prioridades en los criterios de decisión de compra.

El análisis de conjuntos ayuda a desarrollar nuevos productos y definir sus precios, especialmente a los productos (bienes o servicios) que tienen varios atributos y opciones. De tal manera, se puede definir qué atributos y qué opción incluir en el producto. Conocer el valor percibido por el cliente acerca de cada atributo y opción ayuda a establecer los precios de la línea del nuevo producto base al producto

básico, e incrementar el precio de acuerdo a los atributos que se incluyen. Las empresas pueden aprovechar estas herramientas y tecnologías para reducir costos en el desarrollo de productos, aminorar el grado de incertidumbre al definir los atributos en el lanzamiento de un nuevo producto, e identificar el precio que el cliente está dispuesto a pagar por él.

## Referencias

- Agarwal, J., DeSarbo, W., Malhotra, N., y Rao, V. (2015). An interdisciplinary review of research in conjoint analysis: Recent developments and directions for future research. *Customer Needs and Solutions*, 2(1), 19-40. DOI: 10.1007/s40547-014-0029-5
- Ajjan Associates. (s. f.). Ajjan Associates. Conjoint analysis specialists. Recuperado de <http://consulting.ajjan.com/conjoint.htm>
- Dobney. (s. f.). Dobney - Insight and Intelligence. Conjoint analysis. Recuperado de [http://www.dobney.com/Conjoint/Conjoint\\_analysis.htm](http://www.dobney.com/Conjoint/Conjoint_analysis.htm)
- Domigall, Y., Albani, A., y Winter, R. (2014). Identification of customer preferences for new service development in the electricity domain. *2014 IEEE 16th Conference on Business Informatics* (pp. 207-214). DOI: 10.1109/CBI.2014.34
- Forsyth, J., y Boucher, L. (2015). Why big data is not enough. *Research World* 50, 26-27. DOI: 10.1002/rwm3.20187
- Harrar de Dienes, A., y Alcaide, J. (2007). *Uso del análisis conjunto para la evaluación de un curso virtual de principios de procesos industriales*. VII Reunión Nacional de Currículo. I Congreso Internacional de Calidad e Innovación en Educación Superior, Universidad Metropolitana de Venezuela y Universidad Politécnica de Valencia, Caracas.
- Hruska, J. (19 de abril de 2013). Storage pricewatch: HDDs back to pre-flood prices, SSDs grow as \$/GB holds steady. Recuperado de <http://www.extremetech.com/computing/153879-storage-pricewatch-hdds-back-to-pre-flood-prices-ssds-grow-as-gb-holds-steady>
- Komorowski, M. (9 de marzo de 2014). *A history of storage cost*. Recuperado de <http://www.mkomo.com/cost-per-gigabyte-update>
- Machado, M. (2014). *Big data* y servicios financieros. Ponencia presentada en la Jornada "Big data: de la investigación científica a la gestión empresarial" [en SlideShare]. Recuperado de <http://es.slideshare.net/FundacionAreces/manuel-machado-big-data-de-la-investigacion-cientifica-a-la-gestion-empresarial>

- Orme, B. (2010). *Getting started with conjoint analysis: strategies for product design and pricing research*. (2.ª ed.). Wisconsin: Research Publishes LLC.
- Polyviou, A., Pouloudi, N., y Rizou, S. (2014). Which factors affect software-as-a-service selection the most? A study from the customer's and the vendor's perspective. *2014 47th Hawaii International Conference on System Sciences* (pp. 5059-5068). IEEE. DOI: 10.1109/HICSS.2014.621
- Radler, B. (1993). *A case study of conjoint analysis: New approaches to product line decisions*. Tesis. Masster of Arts. (Psychology). Cleveland State University. United States of America.
- Ringel, D., y Skiera, B. (2014). Understanding competition using big consumer search data. *2014 47th Hawaii International Conference on System Sciences* (pp. 3129-3138). DOI: 10.1109/HICSS.2014.388
- Salvador, F. (2014). *Big data: ¿la ruta o el destino?* Recuperado de [http://www.ie.edu/fundacion\\_ie/Comun/Publicaciones/Publicaciones/Big%20Data%20ESP%207.pdf](http://www.ie.edu/fundacion_ie/Comun/Publicaciones/Publicaciones/Big%20Data%20ESP%207.pdf)
- Sawtooth Software. (2015). Conjoint Analysis. Sawtooth Software, the survey software of choice. Recuperado de <http://www.sawtoothsoftware.com/products/conjoint-choice-analysis>
- Talledo, H. (2011). Más allá del *conjoint analysis*. *ANDA News* (Asociación Nacional de Anunciantes), 108, 37-41.
- Ward, J. S., y Barker, A. (2014). *Undefined by data: A survey of big data definitions*. Recuperado de: [http://www.adambarker.org/papers/bigdata\\_definition.pdf](http://www.adambarker.org/papers/bigdata_definition.pdf)