

## BIG DATA PARA LA SEGMENTACIÓN DE MERCADOS EN REDES SOCIALES EN ACCESORIOS DE MODA EMERGENTE

Recibido: 02/02/2017 Aceptado: 15/05/2017

**Chirinos, Roberth**

Universidad Privada Dr. Rafael Beloso Chacín, URBE, Venezuela  
[roberthchirinos87@gmail.com](mailto:roberthchirinos87@gmail.com)

**Villalobos, Melissa**

Universidad Privada Dr. Rafael Beloso Chacín, URBE, Venezuela  
[mcvillalobos1@urbe.edu.ve](mailto:mcvillalobos1@urbe.edu.ve)

### RESUMEN

El objetivo principal de esta investigación es analizar el Big Data para la segmentación de mercados en redes sociales en accesorios de moda emergente. La investigación se desarrolló en torno a la variable Big Data según las teorías de Zikopoulos y otros (2012), Provost y Fawcett (2013) y en cuanto a la segmentación en redes sociales, se consideró el enfoque de Lamb, Hair y McDaniel (2011), Kotler y Keller (2014). El tipo de investigación fue descriptiva, con un diseño no experimental, transeccional, de campo. Para la recolección de datos se trabajó con tres unidades informantes: clientes reales y potenciales de las marcas de accesorios de moda emergente, Community Managers de comunidades online de Facebook orientadas a la gestión de redes, ambas poblaciones infinitas para las cuales se tomó una muestra de 384 sujetos y una última población finita de expertos en computación compuesta por tres profesores de la Universidad Rafael Urdaneta. La técnica de recolección de datos fue la encuesta, se diseñaron dos instrumentos que poseen uno y dos cuestionarios que fueron validados por 6 expertos en mercadeo y computación. El primero, de escala tipo Likert de 9 ítems de 4 alternativas, el segundo de selección simple y consta de 23 ítems; el otro cuestionario es de tipo Likert y consta de 25 ítems con 5 opciones. Al realizar la prueba piloto se obtuvo la confiabilidad de 0,89 de Alfa de Cronbach; 0,83 en el segundo instrumento por medio del coeficiente de interclase mediante la prueba de test-retest, obteniendo una alta confiabilidad. Luego se desarrolló un programa para la extracción de datos de Instagram, analizando 3017 cuentas de usuarios. Concluyendo que existen tres segmentos offline: cautelosos-dinámicos, explorador-experimentador y práctico-equilibrado mientras que online existen cuatro: Sociables muy influyentes, Sociables altamente influyentes, Sociables de poca actividad y Sociables activos.

**Palabras Clave:** big data, segmentación de mercados, redes sociales, moda, data mining.

## BIG DATA FOR MARKET SEGMENTATION IN SOCIAL MEDIA FOR THE EMERGING JEWELRY INDUSTRY

### ABSTRACT

The main purpose of this research is to analyze how Big Data is related to market segmentation in social media in emerging jewelry industry. For which it was consulted, authors such as Zikopoulos (2012), Provost and Fawcett (2013) for Big Data and Lamb, Hair and McDaniel (2011); Kotler and Keller (2014) in social media and market segmentation. The research is considered descriptive, non-experimental and Cross-sectional. For collecting data, the population consisted of Community managers, clients and potential clients of emerging jewelry brands, both infinite, divided in samples of 384 representatives each. And also a third population consisting of computer experts, which was a census of 3 representatives from Rafael Urdaneta University. Two instruments were designed, the first one consisted of one poll of 9 items and 4 options based on Likert scale, and the second one consisted of two surveys the first one consisting of 23 items and the second one of 25 items and 5 options based on Likert scale. They all were validated by 6 experts in marketing and computer engineering. After the pilot testing was done, the results were 0,89 of Alpha Cronbach reliability, and 0.83 for the second instrument after a test-retest. Which means, they were both high? Afterwards, a program was designed in order to extract data from instagram and 3017 user accounts were analyzed. So it was possible to conclude there were three offline segments: careful-dynamic, explorer-experimenter, practical-balanced; and four online segments: Sociable highly influential, Sociable with low activity, Sociable and very influential and Sociable and active.

**Keywords:** big data, market segmentation, social media, fashion, data mining.

### INTRODUCCIÓN

El emprendimiento siempre ha sido un elemento característico de la cultura venezolana. Sin embargo, muchos emprenden sin tener conocimiento alguno sobre el sector en el cual se quieren desenvolver. Esto ha propiciado que muchos negocios fracasen dentro de sus primeros cinco años de vida. Aun así, el venezolano es emprendedor por vocación y sigue intentándolo a pesar de las vicisitudes. La crisis actual del país, ha impulsado con mayor fuerza la motivación emprendedora, generando nuevas y creativas formas de comercializar y promocionar los negocios.

En este sentido, muchas personas han decidido desarrollar sus proyectos personales, especialmente en el sector de moda. Cada vez son más las marcas que confeccionan vestuario o crean accesorios de algún tipo. A pesar de la escasez, limitaciones económicas o sociales, los emprendedores de este sector siempre trabajan en impulsar sus marcas por distintas vías. Por ello, las redes sociales representan una de las alternativas más económicas y de mayor alcance para la promoción. Muchas de estas marcas, utilizan estos medios y han logrado comercializar sus productos por allí.

No obstante, las posibilidades de las redes sociales son más extensas, y muchos las desconocen; es aquí donde una nueva variable hace presencia: El Big Data. Cada

dispositivo electrónico conectado a internet genera algún tipo de rastro digital que en la actualidad, puede ser utilizado para tomar decisiones de negocios. Las redes sociales son una fuente rica en datos y cada clic, share o post, representa una oportunidad para conocer mejor el mercado meta.

En la actualidad existen herramientas que permiten conocer las métricas de las redes sociales, pero no son muy conocidas o simplemente no se adaptan a las necesidades locales. Por esta razón, analizar las redes sociales de marcas locales, contribuye con la comprensión del mercado meta y por ende a una mejor proyección de las estrategias de una marca. El principal propósito de esta investigación es analizar el Big Data para la segmentación en redes sociales en accesorios de moda emergente. Para ello, es necesario teorizar sobre las distintas técnicas, variables y metodologías que intervienen en el proceso, de manera que puedan determinarse los segmentos para este sector.

De esta manera, la investigación se fundamentó en tres etapas: Inicialmente, se realizó un análisis de la problemática y luego se establecieron las bases teóricas que dan sustento a los planteamientos de este artículo; de manera que posibilita la estructura basada en un sistema de variables con sus respectivas dimensiones, sub-dimensiones e indicadores.

Del mismo modo, fue preciso definir el marco metodológico de la investigación, explicando su tipo, diseño, muestra, población y los distintos instrumentos utilizados para recolectar datos, así como las variantes necesarias por las características particulares de ella. Asimismo, se exponen los métodos para calcular la confiabilidad y la técnica de análisis utilizada. Finalmente, se describen los resultados obtenidos contrastándolos con la teoría, de manera que pueda realizarse un aporte significativo al dar respuesta a los objetivos planteados.

## **FUNDAMENTACIÓN TEÓRICA**

Para poder comprender la naturaleza de esta investigación, es necesario analizar el rol que tiene la tecnología y su relación con el análisis de datos en el mundo de los negocios, es preciso tener claro el concepto de Big Data y cómo es la extracción y análisis de la misma. La tecnología tiene distintas aplicaciones en varias áreas del saber, permite no solo optimizar procesos, sino que también es fundamental al momento de explorar, extraer y analizar datos que generen información útil para distintos sectores ahorrando tiempo y costos.

A este respecto, es preciso mencionar que la Big Data en la actualidad es considerada el nuevo "petróleo" (Haupt, 2006) por todas las posibilidades que brinda en el mundo empresarial. Por esto mismo, al momento de aplicar las técnicas de minería de datos que son necesarias para este tipo de investigación, y antes de la elaboración de cualquier proyecto destinado a ello; es indispensable tener claro el concepto de Big Data para poder entender las distintas fuentes de donde provienen los datos que se analicen.

En este orden de ideas, Zikopoulos y otros (2012) definen la Big Data como "... aquella información que no puede ser procesada o analizada, utilizando herramientas

tradicionales”. (p.1) Posteriormente, los autores concatenan esta idea haciendo referencia al hecho de que las empresas constantemente obtienen información de la cual no sacan su verdadero valor y que gracias al Big Data pueden hacerlo. A propósito de esto comentan “...las organizaciones de hoy en día enfrentan cada vez más retos de Big Data. Tienen acceso a información valiosa, pero no saben cómo obtener valor de ella”. (p.1).

Por otra parte, Feinleib (2014) sugiere la siguiente definición:

“...Big Data son los datos que exceden la capacidad de un sistema de base de datos convencional. La data es demasiado grande, se mueve demasiado rápido o no cabe en la estructura de la arquitectura de las bases de datos.” (p.1).

En ambos enfoques, es posible percibir puntos en común en cuanto a la definición de Big Data como aquel proceso que maneja grandes cantidades de datos. Sin embargo, Zikopoulos y otros (2012), definen la variable relacionándola con su uso en las organizaciones y enfocándose más en la técnica de procesamiento que en las grandes cantidades de datos. Asimismo, Feinleib (2014) plantea el concepto desde la perspectiva del tamaño de su impacto y no por su espacio de almacenamiento o complejidad de procesamiento.

De este modo, a pesar de la variedad de definiciones en torno al concepto, para efectos de esta investigación, se entiende Big Data como un conjunto de datos que al ser analizados poseen un valor agregado para una organización. Esta acepción, permite una aproximación cercana al principal aspecto del Big Data a considerar: el valor. De esta forma, es posible que empresas del sector moda sin mucho capital, puedan aprovechar los beneficios generados por estos datos.

En este orden de ideas, es preciso mencionar que el valor de los datos se obtiene a partir del proceso de extracción y análisis, para ello, se requiere un modelado cuyo propósito sea obtener la información valiosa para la empresa o sector. En el caso de esta investigación, se trata de la obtención de distintos segmentos de mercado para poder comprender el desenvolvimiento del mismo.

Ahora bien, la segmentación es un concepto ampliamente usado en el mundo del mercadeo y según el enfoque de cada autor, se define de una manera diferente, atendiendo a factores distintos. Para Kotler y Keller (2012) “la segmentación de mercado consiste en dividir en partes bien homogéneas según sus gustos y necesidades. Un *segmento de mercado* consiste en un grupo de clientes que comparten un conjunto similar de necesidades y deseos” (p. 214). A este respecto, los autores añaden que la tarea del mercadólogo debe ser identificar el número y la naturaleza de los segmentos que conforman el mercado para decidir el mercado meta hacia el cual se dirigirá. Asimismo, Lamb y otros (2011) definen que: “podemos referirnos a los segmentos de mercado como un punto intermedio entre ambos extremos. El proceso de dividir un mercado en segmentos o grupos significativos, relativamente similares e identificables se conoce como segmentación de mercados.” (p. 261).

Por otra parte, Mullins y otros (2007) coinciden con esta definición al establecer que:

“... la segmentación del mercado es el proceso por el cual un mercado se divide en subconjuntos distintos de clientes, con necesidades y características similares, que los llevan a responder de manera semejante a la oferta de un producto particular...” (p. 182).

Los autores resaltan la importancia de la definición de un mercado objetivo, al cual relacionan con el posicionamiento de un producto, pues, con una estrategia de mercadeo orientada hacia un mercado meta definido, se podrá obtener una ventaja competitiva y crear una imagen en la mente del cliente.

No obstante, ninguno de estos autores hace mención explícitamente acerca de la segmentación en redes sociales, por lo cual es necesario indagar sobre qué son estas y cómo es posible relacionar ambos conceptos. Si bien las redes sociales no son nuevas, en la actualidad se han hecho populares gracias al internet y los avances tecnológicos que permiten la interacción entre las personas sin importar la ubicación. El surgimiento de redes como Facebook, Twitter, Instagram o LinkedIn han tenido gran influencia en distintas industrias y en consecuencia, han incidido en la forma de entender los mercados. A propósito de esto, ONTSI citado por INTECO (2011) menciona que las redes sociales son:

“...los servicios prestados a través de Internet que permiten a los usuarios generar un perfil público, en el que plasmar datos personales e información de uno mismo, disponiendo de herramientas que permiten interactuar con el resto de usuarios afines o no al perfil publicado.” (p.12).

Asimismo, Neumann (2011) vincula el concepto de redes sociales con el Internet al aludir que estas son *plataformas en línea* y dependiendo de la orientación de ellas, la información manejada será distinta. A este respecto, es preciso traer a colación su reflexión sobre el término; para él:

“Las redes sociales son plataformas en línea, en las cuales el usuario puede registrarse, permanecer en contacto con la gente de los alrededores o conocer nuevas personas. Según la orientación de la plataforma, serán amigos, conocidos, parientes o contactos de negocio. Las redes sociales tienen en común, que el usuario crea su propio sitio, el perfil, como una especie de perfil con foto. Este tiene información sobre los intereses, trabajo, estudios, libros favoritos o preferencias musicales.” (p. 5).

Ahora bien, una de las aproximaciones más utilizadas hacia este concepto, es la definición de Danah y Ellison (2007). Los cuales introducen el término *servicios de la web* para hacer referencia a las redes sociales y a él añaden unas características particulares. Estos autores, así como Neumann (2011), consideran que las redes sociales permiten la interacción y generación de conexiones con otros individuos a través de la creación de perfiles públicos o semi-públicos. A este respecto los autores comentan lo siguiente:



“Definimos sitios de redes sociales como servicios de la web que permiten a los individuos 1) construir un perfil público o semi-público dentro de un sistema limitado, 2) articular a una lista de otros usuarios con quienes comparte conexiones, y 3) ver y rastrear sus listas de contactos y aquellas hechas por otros dentro del sistema. La naturaleza y nomenclatura de estas conexiones puede variar de una red social a otra.” (p. 2).

De esta manera, luego de analizar las definiciones previamente mencionadas es posible elaborar un concepto de segmentación de mercados en redes sociales, orientado hacia el objeto de estudio, para relacionar el Big Data con la segmentación de mercados. Así, es posible convenir que la segmentación de mercados en redes sociales es el proceso en el cual se detectan las características comunes de usuarios de las redes sociales, mediante la utilización de una serie de algoritmos que permitan hallar patrones de coincidencia para agruparlos en grupos homogéneos.

## **METODOLOGÍA**

Para la realización de este artículo fue preciso establecer un marco que sirvió como referencia para todas las fases de investigación, atendiendo el problema estudiado de una manera objetiva. A este respecto debe entenderse el método científico como un conjunto de reglas, y normas que permiten el estudio y resolución de un problema (Bernal, 2010).

Asimismo, Arias (2006) establece que en este tipo de metodología se incluyen el diseño de la investigación, la población, muestra, muestreo, técnicas e instrumentos con los cuales se recolectan los datos y los procedimientos utilizados para ello. De esta forma, se toman en consideración estos criterios en el estudio del Big Data para la segmentación de mercados en redes sociales en accesorios de moda emergente.

Esta investigación se realizó con un enfoque cuantitativo, el cual se fundamenta en la recolección de datos para la medición de los fenómenos sociales. En esta investigación es preciso cuantificar los datos para poder alcanzar los objetivos planteados, los cuales derivan de un marco teórico. Asimismo, se trata de una investigación no experimental, de campo de tipo descriptiva.

A este respecto, es preciso mencionar que en esta investigación se consideraron tres poblaciones basadas en los datos necesarios para dar respuesta a las interrogantes relacionadas con las dos variables estudiadas, a saber, Big Data y segmentación de mercados en redes sociales. Estas se dividen de la siguiente manera:

Población “a”: Usuarios de redes sociales que sean clientes reales y clientes potenciales de diseñadores de accesorios de moda emergente que participen en los bazares de moda de los hoteles cinco estrellas de Maracaibo entre los meses junio y septiembre de 2016.

Población “b”: Profesores de las escuelas de computación de la Universidad Rafael Urdaneta y la Universidad Privada Dr. Rafael Bellosó Chacín con experiencia docente entre uno y tres años que hayan cursado la materia “minería de datos” durante su ciclo de

formación universitaria. En la cual se aprenden las técnicas necesarias para el análisis de grandes cantidades de datos.

**Tabla 1. Distribución de la población de profesores de las escuelas de computación**

Universidad	Cantidad
Universidad Rafael Urdaneta	3
Universidad Privada Dr. Rafael Bellosó Chacín	0
<b>Total</b>	<b>3</b>

**Fuente:** Suministrado por las universidades Dr. Rafael Bellosó Chacín y Rafael Urdaneta (2016).

Población “c”: *Community Managers* de habla hispana que son miembros de comunidades activas orientadas a la gestión de redes sociales dentro de Facebook y que actualmente desempeñen sus labores como administradores de redes sociales para una o más empresas.

Para efectos de la presente investigación, la población “a” es infinita, en vista de que se trata de los clientes reales y potenciales de las marcas de accesorios de moda emergente, los cuales son todas aquellas personas que hayan comprado en algún momento cualquier tipo de accesorio de marcas pertenecientes a bazares o no. Esto resulta difícil de determinar, pues la población, además de ser actualmente grande, constantemente aumenta de tamaño, por lo cual se toma tal posición. Asimismo, se debe mencionar que esta población tendrá una edad mayor de quince años, residente de Maracaibo y será usuaria de la red social Instagram.

Del mismo modo, la población “c” es infinita y consta de los *Community Managers* activos de habla hispana que son miembros de grupos o comunidades de esta área dentro de Facebook. Se considera que la población tiene estas características, pues no existe registro alguno sobre su tamaño y así como la anterior, esta aumenta constantemente. Por otra parte, la población “b” es finita y accesible por lo cual se realizó un censo poblacional.

Ahora bien, la población de los clientes y clientes potenciales de los diseñadores de accesorios de moda emergente es infinita debido a que los datos provienen de la red social Instagram, una red social en constante crecimiento y las encuestas realizadas (Forbes, 2016). Además, frecuentemente nuevos usuarios se interesan por este tipo de productos, haciendo difícil la obtención de la totalidad de la población, pues esta varía constantemente. Asimismo, la población “c” es infinita pues, a diario se crean nuevas comunidades orientadas hacia *Community Managers*, y nuevos miembros se unen a ellas,

haciendo difícil determinar un número exacto. Por lo tanto, para calcular estas muestras se utilizó la ecuación para poblaciones infinitas sugerida por Arias (2006):

$$N = \frac{Zc^2 * S^2}{e^2}$$

Dónde:

N: es el tamaño de la muestra

Zc: nivel de confianza

S: Desviación estándar

e: error muestral

$$N = \frac{1.96^2 * 0.5^2}{0.05 * 0.05} = 384 \text{ Para las poblaciones "a" y "c"}$$

Para esta investigación se trabajó con 95% de nivel confianza, estimación del error muestral  $e = 0.05$  y una desviación estándar  $S = 0.5$ .

Luego de aplicar la ecuación para el cálculo de la muestra de una población infinita, se obtuvo como resultado una cantidad de 384 clientes reales y clientes potenciales de los diseñadores de accesorios de moda emergente, así como 384 *Community Managers*. En vista de que se utilizó un muestreo probabilístico por conglomerados para la población "a", fue preciso dividir a la población en distintos grupos con características similares para luego analizarlos.

A este respecto, se construyeron dos instrumentos que poseen uno y dos cuestionarios. El primero consta de un cuestionario de veintitrés (23) *ítems* de selección simple y está dirigido a la población "a", siendo este rellenado vía electrónica a través de una aplicación. Por otra parte, el resto de los cuestionarios está dirigido hacia las poblaciones "b" y "c"; estos son de carácter físico y electrónico, teniendo nueve (9) y veinticinco (25) *ítems* respectivamente, ambos con una escala tipo Likert, el primero de cuatro (4) y el segundo de cinco (5) opciones. Posteriormente, al realizar la prueba piloto se obtuvo la confiabilidad de 0,89 de Alfa de Cronbach en el primer instrumento y 0,83 en el segundo por medio del coeficiente de inter-clase mediante la prueba de test-retest.

**Tabla 2. Baremo para el instrumento dirigido a los profesores expertos en computación**

ESCALA	CATEGORÍAS
1-1,75	Totalmente en desacuerdo
1,76-2,50	En desacuerdo
2,51-3,25	De acuerdo



3,26-4	Totalmente de acuerdo
--------	-----------------------

**Fuente:** elaboración propia (2016).

**Tabla 3. Baremo para el instrumento dirigido a los *Community Managers***

ESCALA	CATEGORÍAS
1-1,80	Nunca
1,81-2,60	Casi nunca
2,61-3,40	Algunas veces
3,41-4,20	Casi siempre
4,21-5	Siempre

**Fuente:** elaboración propia (2016).

## LA INVESTIGACIÓN

Para la realización de los cálculos extraídos directamente de las redes sociales se utilizó un programa desarrollado en JAVA. Los datos recolectados de las encuestas realizadas a la población “a” y los extraídos de sus respectivos perfiles de Instagram se codificaron, agrupando las respuestas de las preguntas con sus números correspondientes. Asimismo, se utilizó la clasificación de los datos de Instagram para analizar los distintos segmentos en función de su comportamiento en la red.

En este orden de ideas, se debe mencionar que el principal objetivo del análisis de clusters es generar subgrupos homogéneos partiendo de una población heterogénea, para lo cual se precisan técnicas estadísticas que permitan dilucidar los patrones existentes entre los distintos elementos de la población estudiada. Para ello se implementó el algoritmo para el análisis de conglomerados de dos etapas del *Spss*. Por medio de él, fue posible establecer los clusters. Luego de obtener los conglomerados, se realizó un análisis de las características comunes encontradas y se interpretaron los resultados para posteriormente determinar los distintos segmentos del mercado en el sector dentro de la red social.

Asimismo, se debe mencionar que esta investigación se llevó a cabo de una manera sistemática, basada en distintas fases con el propósito de alcanzar el objetivo principal. Del mismo modo, debido a las características de la misma fue preciso hacer uso de una metodología orientada especialmente hacia los proyectos de minería de datos, la cual es de naturaleza cíclica. Por esta razón, es preciso mencionar que para el análisis y modelado de algunos datos se utilizó la metodología *CRISP-DM*.

Esta metodología incluye las descripciones de las fases típicas de un proyecto de ingeniería de software, es decir, en ella se especifican las tareas que pertenecen a cada fase y una explicación de las relaciones existentes entre estas tareas. A nivel general esta metodología está orientada hacia el cumplimiento del ciclo de vida de la minería de datos. Según lo descrito por ella, la minería de datos tiene un ciclo de seis fases que no precisamente deben tener un orden específico. Es una metodología flexible que permite saltar de una fase a otra a medida que se desarrolla el proyecto. A este respecto las fases aquí expuestas según Gettler y otros (2010) son:

(a) Entendimiento del negocio: Es la fase inicial y se concentra en entender los objetivos del proyecto y sus requerimientos desde la perspectiva del negocio, para luego convertir el conocimiento en un problema de minería de datos. De esta manera puede diseñarse un plan para alcanzar los objetivos establecidos.

(b) Entendimiento de la data: Esta fase comienza con una recolección inicial de la data para posteriormente realizar actividades que ayuden al investigador a familiarizarse con ella, identificando problemas en la calidad de los datos, tener las primeras percepciones de la data y generar hipótesis.

(c) Preparación de la data: En esta fase se realizan todas las actividades pertinentes para la construcción del conjunto de datos que posteriormente serán introducidos en la herramienta destinada al modelado.

(d) Modelado: En esta etapa se aplican las técnicas de modelado, y los parámetros se calibran para obtener los valores óptimos. A este respecto, se debe mencionar que usualmente hay varias técnicas que pueden ser aplicadas a un mismo problema.

(e) Evaluación: En esta fase, los modelos construidos aparentemente permitirán obtener resultados de alta calidad. Sin embargo, antes de la implementación es preciso evaluar los modelos y los distintos pasos realizados para la creación de los mismos para garantizar que los objetivos planteados se cumplan.

(f) Implementación: Usualmente la creación de un modelo no representa el final del proyecto. La data generada y el conocimiento adquirido, debe ser presentado de manera apropiada para el análisis del cliente o dueño del negocio. Dependiendo de cada caso, esta fase puede ser más o menos compleja, dependiendo de los requerimientos del cliente.

Como es posible observar, muchas de estas fases se cumplen en proyectos de investigación de otras ciencias, sin embargo; en los proyectos de minería de datos es preciso considerar que existen ciertas variaciones sobre todo al momento de preparar, modelar y evaluar la data, pues el mismo procedimiento requiere el test y retest de los modelos para evaluar la consistencia de los datos. Asimismo, dicha evaluación requiere de la extracción de datos durante un tiempo que dependerá de si se tienen o no los datos, pues, es necesario tener una base de datos lo suficientemente grande como para un análisis efectivo.

En primera instancia, lo que pretende esta investigación es poder dilucidar los patrones existentes dentro de la población estudiada para poder realizar una segmentación de mercado. Asimismo, esta brinda la oportunidad de poder contrastar (y complementar) la data obtenida *online* y *offline*, es decir; aquella cedida directamente por los sujetos y la extraída desde las redes sociales. Vale la pena destacar, que este es un criterio del investigador y es un eslabón fundamental en el proceso de modelado y análisis de los resultados.

Por esta razón, el hecho de recolectar datos de ambas maneras, representa un aporte significativo dentro del mismo proceso inherente al análisis de resultados finales. Por ende, fue necesario dividir esta investigación en dos fases. (a) Una fase exploratoria de análisis de datos concretos en menor escala y (b) Una fase concluyente de extracción y análisis de datos por medio del uso de la plataforma tecnológica diseñada para procesar mayores volúmenes de información.

### **PRIMERA FASE: ANÁLISIS DE DATOS A MENOR ESCALA**

En esta fase, luego de validar los instrumentos, se procedió a aplicarlos a sus respectivas unidades informantes. Como esta investigación requirió de la aprobación o verificación de expertos en dos áreas específicas, a saber; las ciencias de la computación y el mercadeo; fue necesario validar y aplicar los instrumentos de las poblaciones “b” y “c” para posteriormente proceder a la aplicación del instrumento dirigido a la población “a”, así como la selección de la red social con la cual se iba a trabajar para la extracción de datos, pues, estos resultados ayudaron a determinarla.

Asimismo, dentro de la primera fase, se obtiene el primer grupo de resultados que si bien, fueron estudiados desde una perspectiva más tradicional (es decir, sin considerar las técnicas de la minería de datos), permitió tener mayor claridad respecto al fenómeno estudiado y así poder tomar decisiones pertinentes a la fase de exploración y modelado de la metodología CRISP-DM. Esta puede asociarse a la primera fase de dicha metodología, referente al entendimiento de la data.

Es preciso mencionar que debido a la vasta cantidad de datos que pueden extraerse de una red social, un análisis previo (en este caso) genera un gran aporte al momento de filtrar la data con la cual se va a trabajar. De la misma manera, un estudio *offline* complementa el análisis de redes sociales de una forma significativa, pues hay datos que no pueden extraerse directamente de las mismas; por ejemplo, el ingreso mensual.

### **SEGUNDA FASE: EXTRACCIÓN Y ANÁLISIS DE DATOS DE LAS REDES SOCIALES**

Luego de realizar un análisis de la data de la fase previa y establecer los segmentos en función de los 384 encuestados pertenecientes a la población “a”, se definió la red social de la cual se extrajeron los datos en función de los resultados arrojados. Se trabajó con los datos que se consideraron pertinentes para la investigación luego de estudiar los segmentos obtenidos. A pesar de que en un inicio la investigación estuvo planificada para trabajar con la red social Facebook, luego del análisis de los resultados se determinó que

Instagram era la más adecuada para la extracción de la data. Ya que, si bien, la mayor parte de los clientes reales y potenciales utilizan ambas redes, se pudo determinar que en la segunda, la interacción con las marcas de accesorios de moda emergente es mucho mayor, teniendo más *followers* y generando mayor contenido.

Debido a que se trata de una red social directa, en la cual la información tiene una clasificación basada en la interacción de sus usuarios dentro de la red, es posible utilizar esta data para segmentar en función de lo que provee la misma API de Instagram. Sin embargo, esta no es de libre acceso, pues por políticas de privacidad, la red social restringe los datos de sus usuarios. Para poder obtener dicha información, es necesario que el usuario conceda los permisos a través de una aplicación desarrollada para tal fin. Por esta razón, se creó una aplicación cuyo objetivo es extraer la data directamente de la web. Se trata del desarrollo de un *Web Crawler*, el cual es un programa que analiza páginas web de manera iterativa en la búsqueda de información.

En esta ocasión, se trata de un *crawler* orientado hacia los perfiles públicos de Instagram que estén directamente relacionados con las marcas de accesorios de moda analizadas. Asimismo, los usuarios estudiados, son aquellos que tienen algún tipo de interacción con las marcas, es decir, que generen *likes* o comenten algún post.

## RESULTADOS

Luego de analizar los datos recolectados con los distintos instrumentos, se obtuvieron los resultados. Estos pertenecen a cada una de las fases previamente mencionadas, para ello se utilizaron los algoritmos predefinidos por las librerías de los *Software* usados. En la primera fase, se analizaron los datos provenientes de las encuestas y en la segunda, la totalidad de la data recolectada de las redes sociales mediante el uso de técnicas de minería de datos.

### PRIMERA FASE

Al aplicar las encuestas y obtener los resultados para responder a los objetivos referentes a las características demográficas y psicográficas, pudieron deducirse los distintos conglomerados que se forman de manera natural. A continuación, se explica el procedimiento y los conglomerados hallados en la primera fase.

### CONGLOMERADOS

El propósito principal de la segmentación de mercados para Lamb y otros (2011) es la creación de grupos significativos relativamente similares y fáciles de identificar. Por esta razón, es preciso utilizar los datos generados por el instrumento dirigido a los clientes y clientes potenciales para generar conglomerados, es decir, grupos de sujetos con características similares que puedan representar distintos segmentos de mercado.

Para ello, en esta primera fase, se utilizó inicialmente un algoritmo exploratorio para la clasificación de conglomerados con el *software SPSS Statistics* versión 19, conocido como “análisis de conglomerados bietápico o de dos fases”. En este, los reactivos se sometieron a análisis y luego de varios tests, se midió el grado de cohesión y separación

entre los distintos ítems para poder determinar qué tan bien definido estaba cada conglomerado.

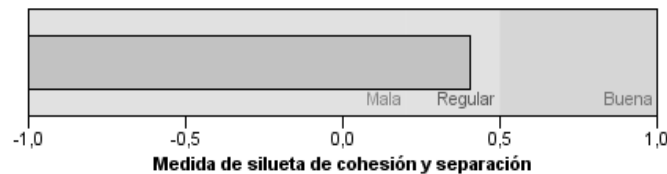
A pesar de la variedad de respuestas y sujetos entrevistados, la población aún permanecía muy dispersa como para establecer conglomerados con todos los ítems, por lo cual, luego de varios *tests* y análisis se fueron descartando algunos hasta conseguir un grado de cohesión “regular” cercano a bueno. Para ello, fue preciso seleccionar los nueve predictores con mayor grado de importancia según la información provista por el programa. De este modo, fue posible definir tres conglomerados. Los cuales dependen de las nueve preguntas seleccionadas que tienen una mayor ponderación asignada por el sistema, luego de su análisis estadístico.

**Gráfico 1. Resumen de modelo y calidad de conglomerado**

**Resumen de modelo**

<b>Algoritmo</b>	Bietápico
<b>Entradas</b>	9
<b>Conglomerados</b>	3

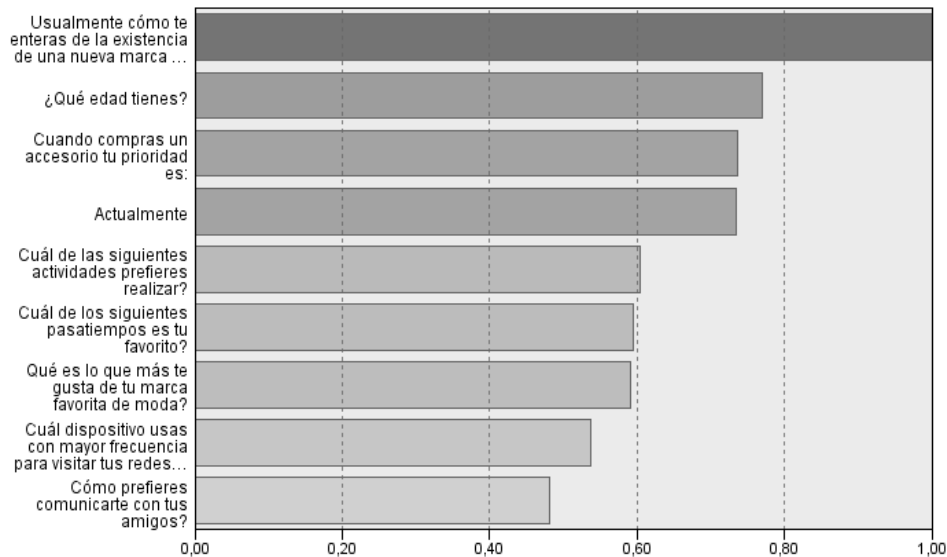
**Calidad de conglomerado**



**Fuente:** elaboración propia (2016).



**Gráfico 2. Importancia del predictor**

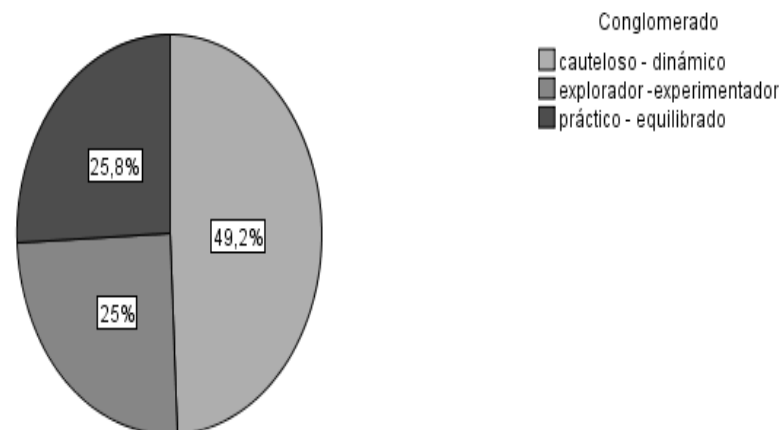


**Fuente:** Elaboración propia (2016).

Al analizar las distintas características de los conglomerados, fue posible crear un perfil para cada uno de estos segmentos. Luego de estudiar las características demográficas y psicográficas, se crearon tres denominaciones según segmento: (a) cauteloso – dinámico (b) práctico – equilibrado y (c) explorador – experimentador. Asimismo, estos se dividieron en 49.2%, 25.8% y 25% respectivamente. Siendo los predictores más relevantes, según el programa, la edad y la manera en que se enteran de las nuevas marcas.

**Gráfico 3. Tamaños de conglomerados**

**Tamaños de conglomerados**



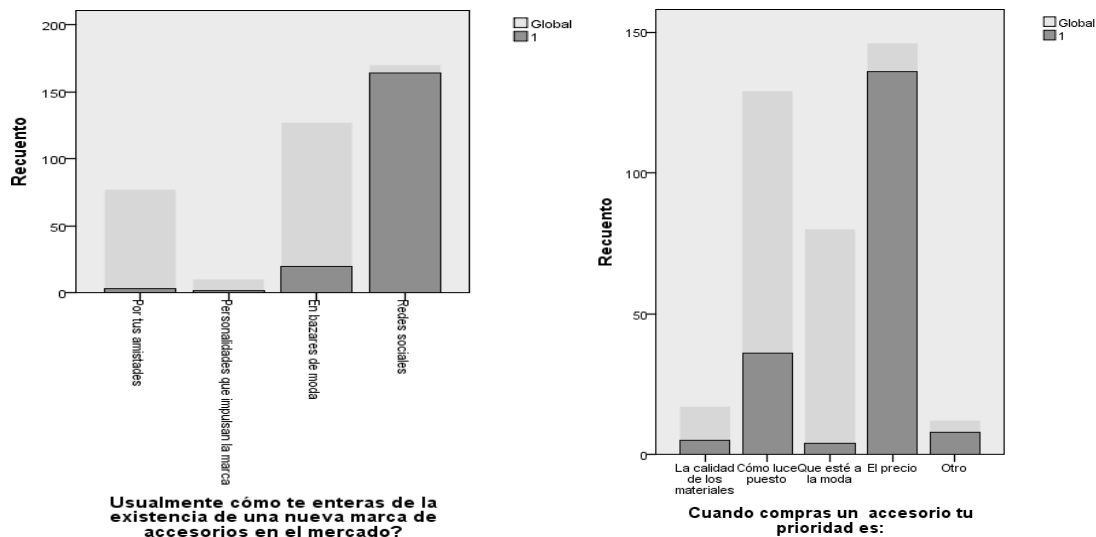
Fuente: Elaboración propia (2016).

### SEGMENTO: CAUTELOSO – DINÁMICO

Este segmento consta de 49.2% de la población, se caracteriza por buscar practicidad y calidad en los productos. La mayoría de ellos tienen edades comprendidas entre 25 y 30 años, siendo la mayoría profesionales independientes, son cautelosos con sus inversiones y gastos. También tienen un estilo de vida activo y variado, hacen ejercicio, visitan cafés y son sociables. La prioridad de ellos al momento de comprar un producto es la calidad de los materiales.

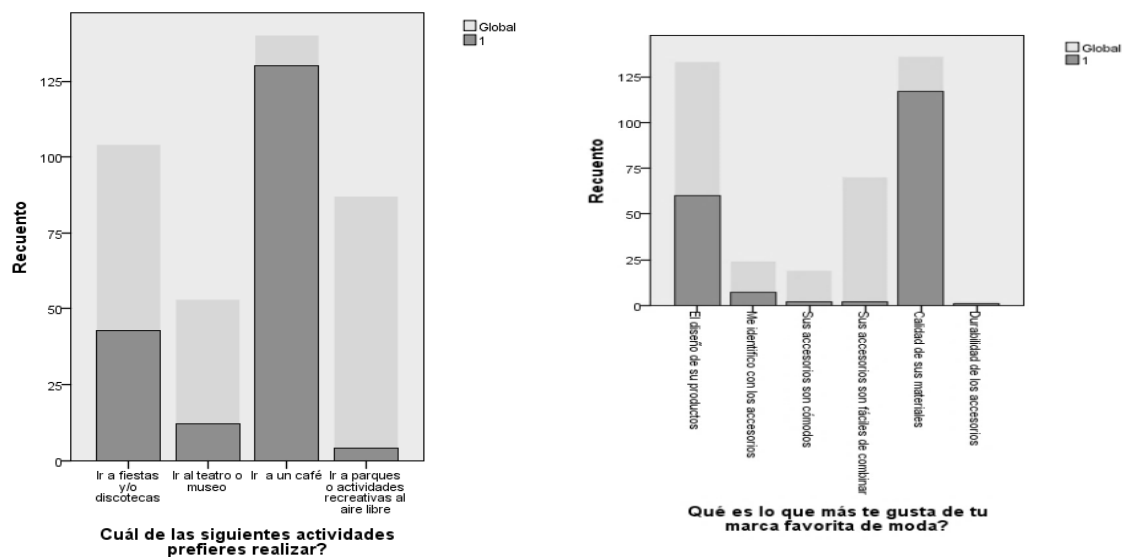
Asimismo, la mayor parte frecuenta las redes sociales en su teléfono móvil y descubre nuevas marcas de moda a través de las mismas. Otro de los factores importantes a considerar por este segmento es el precio, ellos buscan una relación calidad precio, donde se sientan conformes con el producto por el cual están pagando.

**Gráfico 4. Distribución de casillas – existencia de nuevas marcas y prioridad de compra de accesorios**



Fuente: elaboración propia (2016).

**Gráfico 5. Distribución de casillas – prioridad de compra de accesorios y gustos en marca de moda favorita**



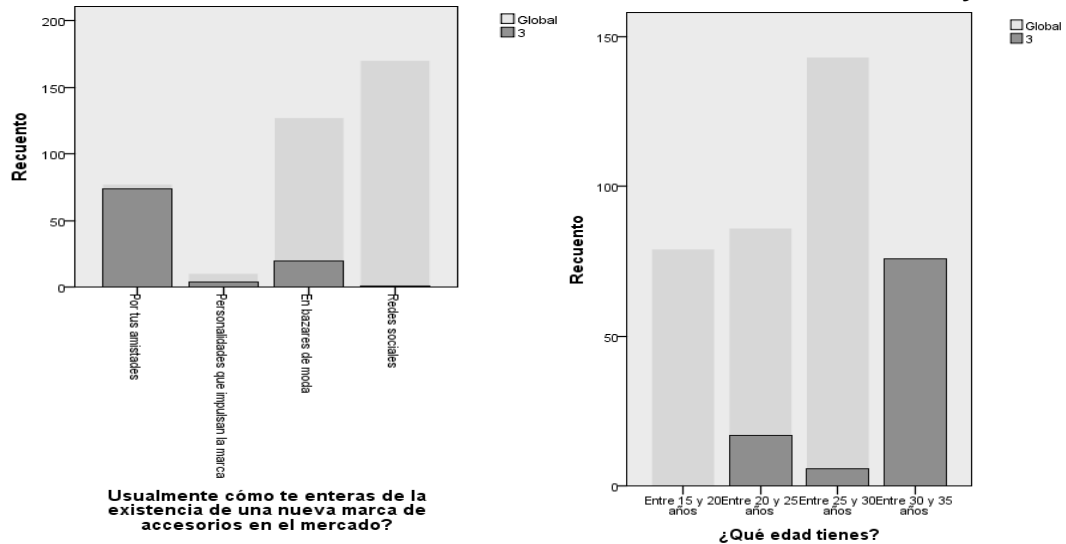
**Fuente:** elaboración propia (2016).

**SEGMENTO: PRÁCTICO – EQUILIBRADO**

Este segmento de mercado se preocupa por la practicidad y el estilo de los productos, se caracteriza por llevar un tipo de vida tranquilo donde hay un equilibrio entre recreación, trabajo y estudio. Está representado en su mayoría por personas de edades comprendidas entre 30 y 35 años. Al momento de adquirir accesorios, su prioridad es que la prenda esté a la moda o se ajuste a su ideal de vestimenta, además de ser fáciles de combinar, es decir; deben ser accesorios prácticos.

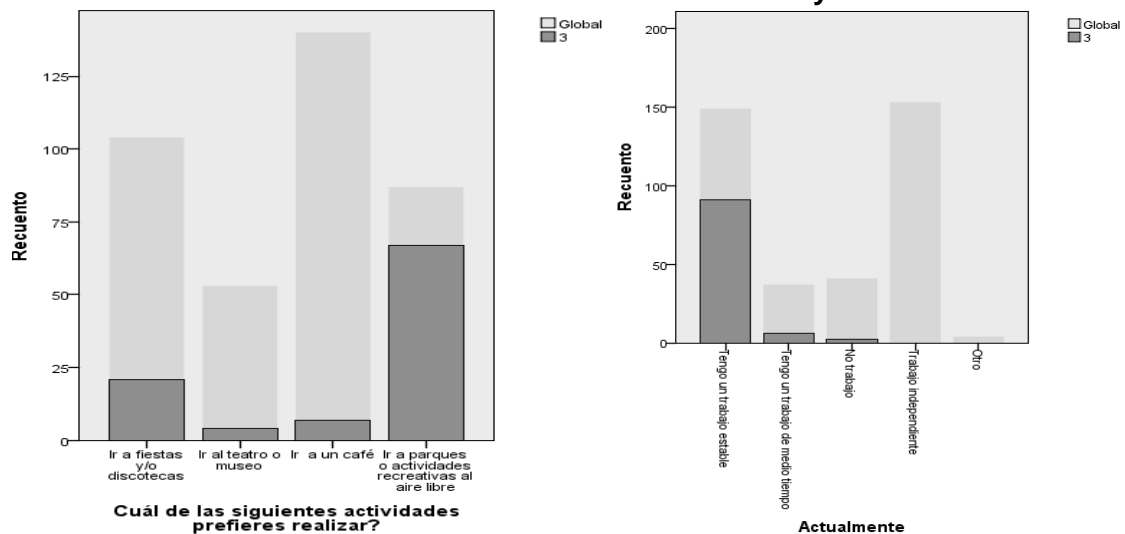
Tienen un trabajo estable y en sus tiempos libres prefieren realizar alguna actividad recreacional o asistir a alguna fiesta. Asimismo, la mayor parte continúa sus estudios o tiene como pasatiempo leer. También, utiliza la laptop con frecuencia para visitar las redes sociales. Sin embargo, cuando los sujetos de este segmento conocen alguna marca nueva, por lo general lo hace a través de recomendaciones de amistades. Si bien, se adaptan a la tecnología, ellos, así como los cautelosos – dinámicos, prefieren interactuar en persona. Del mismo modo, es posible observar, que el precio no suele ser un factor decisivo para ellos, se ven influenciados sobre todo por el círculo social y el estilo o concepto de moda que manejen.

**Gráfico 6. Distribución de casillas – existencia de las marcas favoritas y edad**



Fuente: elaboración propia (2016).

**Gráfico 7. Distribución de casillas – actividades favoritas y situación laboral**



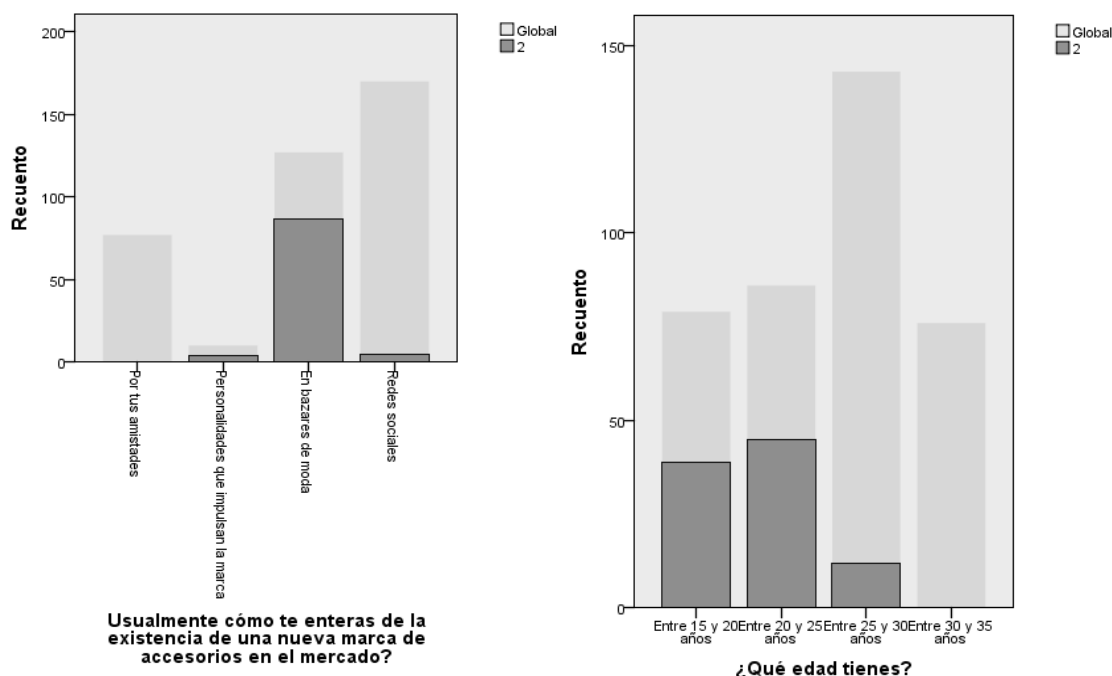
Fuente: elaboración propia (2016).

## SEGMENTO: EXPLORADOR – EXPERIMENTADOR

Este segmento de mercado se encuentra en fase de autodescubrimiento, varía en tendencias y estilos. Le preocupa principalmente adquirir los productos que estén de moda, que posean diseños acorde a sus expectativas. Tienen un estilo variado, dinámico, experimentan con variedad de actividades y para ellos es muy importante la apariencia.

A diferencia de los dos segmentos anteriores, las edades de los sujetos de este conglomerado se agrupan en dos rangos, uno de 15 a 20 años y otro de 20 a 25 años. Una parte de ellos tiene trabajo estable y otra no trabaja o tiene un trabajo de medio tiempo. Por otra parte, las actividades de preferencia también son variadas, las que más destacan son ir al teatro o museo e ir a fiestas o discotecas. La mayor parte de ellos, utiliza el teléfono móvil y prefiere la interacción vía redes sociales. No obstante, de los tres segmentos son los que con mayor frecuencia descubren nuevas marcas de moda en bazares; de ello se puede inferir que son los que más visitan estos eventos.

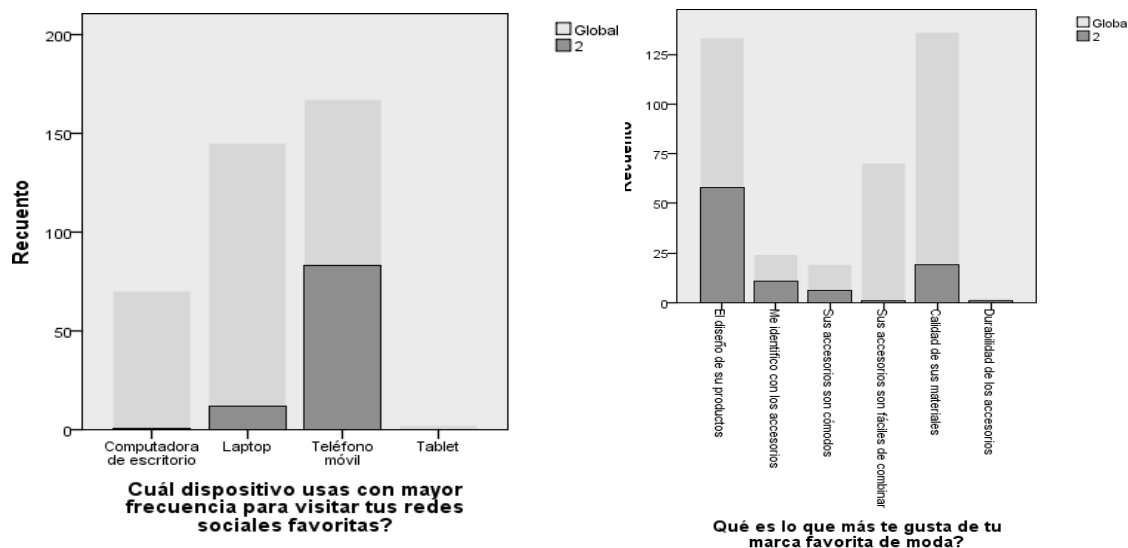
**Gráfico 8. Distribución de casillas – existencia de las marcas favoritas y edad**



Fuente: elaboración propia (2016).



**Gráfico 9. Distribución de casillas – dispositivos utilizados y gustos por marca favorita**



**Fuente:** elaboración propia (2016).

Luego de evaluar los conglomerados generados por el algoritmo de clasificación de dos etapas, fue posible determinar que, si bien los grupos no están totalmente cohesionados, expresan una tendencia regular de cómo se distribuye este mercado. Gracias a esto, es posible determinar qué elementos pueden considerarse al momento de hacer un análisis más profundo en redes sociales. A este respecto, pudieron sugerirse las características de cada segmento. Con la fase dos se obtendrán resultados más específicos orientados hacia las redes sociales que podrán complementar la información obtenida.

## SEGUNDA FASE

En la fase previa se pudo determinar que existen tres segmentos *offline* que poseen cierto grado de dispersión, es decir, los conglomerados no están totalmente definidos. Sin embargo, la aplicación de las encuestas permitió discernir sobre qué red social era más conveniente para la presente fase. De los resultados obtenidos, se pudo deducir que la mayor parte de la población (46,1%) prefiere la red social Facebook, y en segundo lugar se encuentra Instagram con 40,9%; por lo cual la primera sería la red elegida para la extracción de datos.

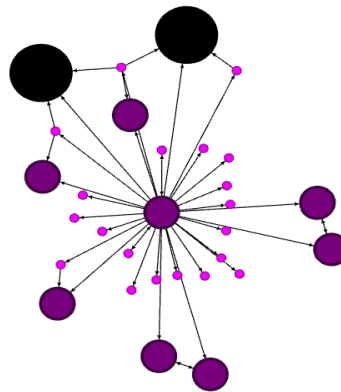
Para ello, se seleccionaron ocho cuentas de marcas del sector cuyos clientes fuesen reales o potenciales de aquellas marcas que participasen en los bazares previamente mencionados. Estas fueron: kapauzna, noresoficial, kilchis, grazia\_hombres, soffitta\_, dralucrochet y maricrisaccesorios. Dentro de los criterios más importantes de selección,

se encuentran: la actividad de la marca en la red y el uso de hashtags que denotasen su participación en bazares como bazar más, bululú bazar, entre otros.

Ahora bien, el principal objetivo de esta fase es extraer los datos pertinentes para poder determinar el comportamiento de los usuarios dentro de la red y así establecer conglomerados en función del mismo, por lo cual Facebook, parecía la red social ideal en vista de que esta provee una clasificación en función de los intereses de los usuarios. Por lo cual, inicialmente la aplicación de extracción de datos estaba dirigida hacia Facebook, pretendiendo realizar una segmentación psicográfica, basada en intereses.

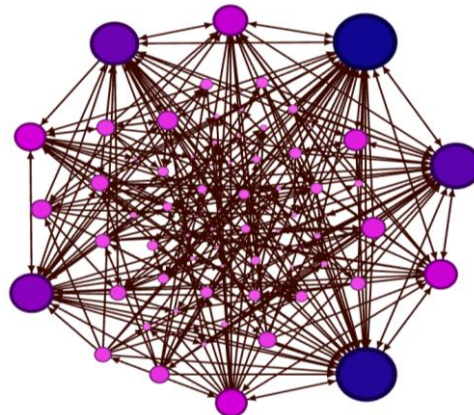
No obstante, luego de los análisis preliminares de las dos redes más utilizadas por los usuarios, fue posible observar que la actividad de las marcas en Facebook era muy baja en comparación con la de Instagram. Sin mencionar que solo dos de las marcas seleccionadas poseen perfil en esta red. Por esta razón, la red social seleccionada fue la segunda.

#### **Gráfico 10. Grafo de Kapauzna Accesorios de la red social Facebook**



**Fuente:** elaboración propia (2016).

#### **Gráfico 11. Grafo de Kapauzna Accesorios de la red social Instagram**



**Fuente:** elaboración propia (2016).

Kapauzna Accesorios es una de las dos marcas que poseen cuenta en ambas redes sociales. Sin embargo, en Facebook no existe actividad desde septiembre de 2016 y cuenta con 163 seguidores. Mientras que en Instagram, la misma posee más de 14.000 seguidores y es actualizada a diario. En estos grafos se visualizan las reacciones de los distintos usuarios con los posts publicados en la cuenta y las relaciones que estos tienen con otros usuarios.

El tamaño y color de los distintos nodos refleja el grado de conexiones que entran y salen del mismo; siendo los más grandes, aquellos que tienen mayor número de conexiones o peso. Con este gráfico, puede ejemplificarse la situación y justificarse por qué se selecciona la segunda por encima de la primera. Demostrando así, que a pesar de que Facebook es la red social preferida; las marcas y los usuarios prefieren interactuar en Instagram para temas relacionados con este sector.

Ahora bien, el programa de extracción de datos (*web crawler*) se diseñó para que tomase como parámetro principal la cuenta de una marca y que además de recopilar sus datos principales (*posts, likes, descripciones, textos, comentarios e imágenes y videos*); extrajera los perfiles de cada uno de los usuarios que tuvieron algún tipo de interacción con los posts de la marca. A este respecto, se debe mencionar que se analizó un total de 3017 cuentas de usuarios, de las cuales solo se tuvo acceso a 1236 ya que el resto eran perfiles privados. De estos, se analizaron 603932 publicaciones y 4081819 comentarios.

Luego de extraer toda la información, esta se almacena en una base de datos y posteriormente se realizan los cálculos preliminares para luego introducirlos en el *SPSS Statistics* versión 19, y aplicar el método bietápico de generación de conglomerados. Las características de esta investigación ameritan este tipo de métodos para la recolección de datos; pues se trata de abundantes cantidades de información que sería muy difícil de procesar manualmente.

Si bien los resultados de la primera fase representan una segmentación que es de utilidad para cualquiera que desee entender el mercado, el análisis en las redes sociales puede profundizar en ella para conseguir nuevos patrones o datos que sean más precisos, pues como pudo observarse, la población aún en la segmentación de la fase anterior, aún no se comporta de manera totalmente homogénea.

### **ANÁLISIS LA INFORMACIÓN OBTENIDA**

Provost y Fawcett (2013) hacen mención de las distintas tareas y técnicas utilizadas en el análisis de datos, dando a entender que no existe una manera única de interpretar los datos. Por lo cual, esto depende de la intuición y experticia del científico. A este respecto, debe mencionarse que, la interpretación y análisis de los datos de esta investigación son una propuesta del investigador y tal como pudo corroborarse previamente por los expertos; cada proyecto amerita un modo distinto de aproximarse al problema.

Por ello, al momento de segmentar en la red social Instagram se toma en consideración lo teorizado en el segundo capítulo a propósito de Aimia (2015) en cuanto a

la búsqueda de grupos homogéneos en las redes sociales en función del comportamiento de la audiencia dentro de la plataforma. De esta forma, se estudiaron como principales métricas, el *engagement*, la cantidad de *posts*, *followers*, hora de publicación y *comentarios*.

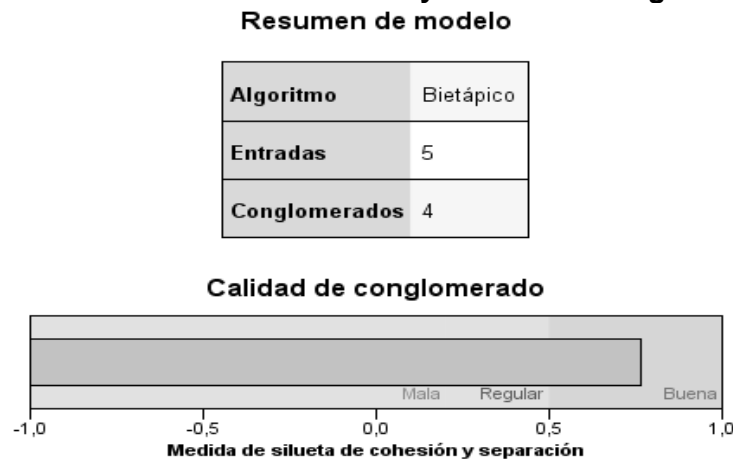
Cada red social posee sus propios elementos a considerar al momento de calcular el *engagement* (Hootsuite, 2016). Asimismo, existen distintas fórmulas propuestas. No obstante, para esta investigación se consideró la propuesta por Hootsuite (2016), en vista de que es la recomendada por la empresa para los casos donde no se tiene acceso de administrador a la cuenta, esto puede proveer un estimado, sin embargo, es útil para el propósito de este estudio. Asimismo, debe mencionarse que esta métrica puede calcularse en períodos de tiempo determinados, por posts o a través de una media general relacionada con todos los posts realizados.

$$\text{engagement} = \frac{\text{likes} + \text{comments}}{\text{followers}}$$

## SEGMENTOS EN REDES SOCIALES

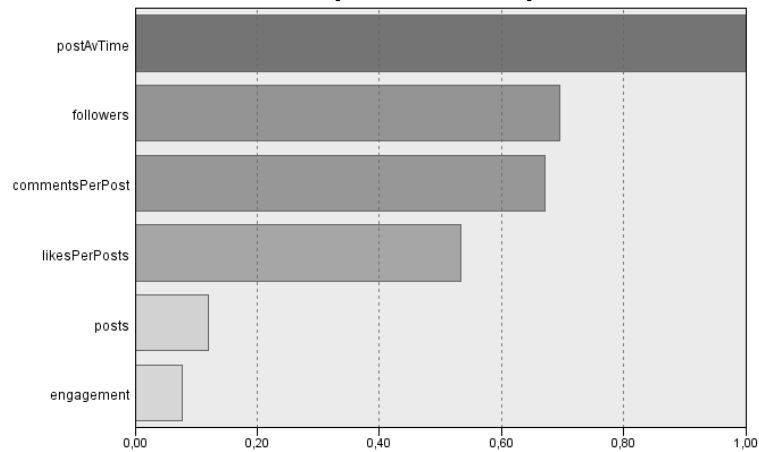
Luego de realizar el análisis de conglomerados bietápico, se observaron cuatro segmentos bien definidos, basados en el promedio de tiempo de publicación, la cantidad de seguidores, los comentarios que obtienen los usuarios por post, los likes por post, cantidad de posts y el *engagement*. A este respecto, el predictor más sobresaliente es el relacionado con el tiempo, permitiendo así definir los *clusters* en función de los momentos en que más interactúan los distintos tipos de usuarios.

**Gráfico 12. Resumen de modelo y calidad de conglomerado**



**Fuente:** elaboración propia (2016).

**Gráfico 13. Importancia del predictor**

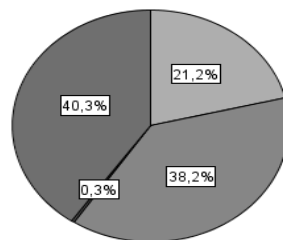


**Fuente:** elaboración propia (2016).

A este respecto, los conglomerados arrojados por el estudio demuestran la alta actividad que existe en esta red, especialmente en los perfiles públicos que cuentan con una gran exposición y reciben muchas interacciones. A pesar de que en algunos de ellos, los números son similares, fue posible segmentar a la audiencia en función de la actividad y el nivel de reacción recibido en cada cuenta, determinando así cuatro segmentos.

**Gráfico 14. Tamaños de conglomerados**

Tamaños de conglomerados



Conglomerado

- Sociables muy influyentes
- Sociables activos
- Sociables altamente influyentes
- Sociables de poca actividad

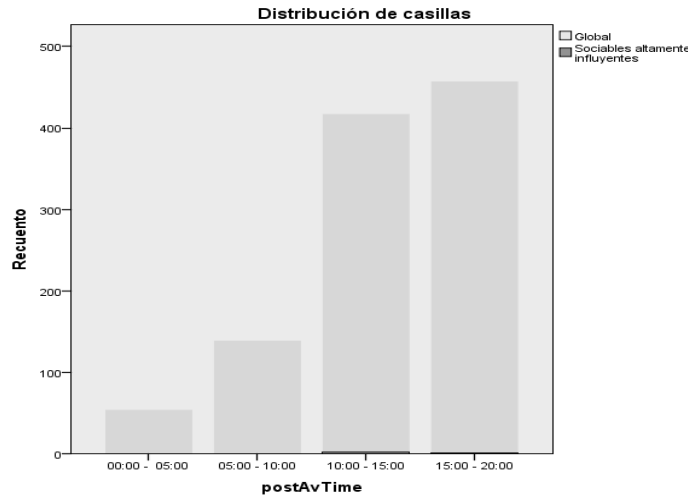
Tamaño de conglomerado más pequeño	3 (0.3%)
Tamaño de conglomerado más grande	430 (40.3%)
Cociente de tamaños: Conglomerado más grande a conglomerado más pequeño	143.33

**Fuente:** elaboración propia (2016).



## SOCIABLES ALTAMENTE INFLUYENTES

Gráfico 15. Tamaños de conglomerados

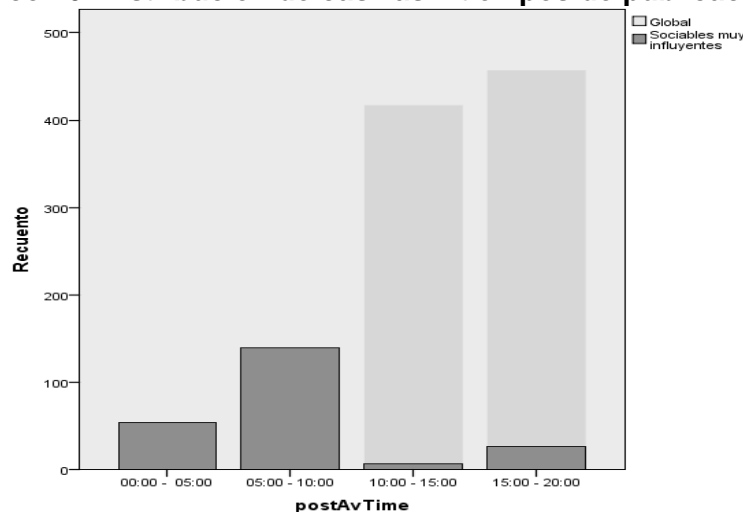


Fuente: elaboración propia (2016).

Este segmento representa 0,3% de la población estudiada. Son usuarios diurnos cuyo momento de mayor actividad se encuentra entre las 10:00 y 15:00 hrs. Tienen gran influencia sobre sus seguidores debido a su alto engagement. Se caracterizan por su alta actividad, teniendo una media de 5228 posts en su cuenta, superando los 2 millones de *followers*. Se trata de usuarios que muy rara vez se encuentran en los comentarios de otras cuentas y usualmente son celebridades o personas con mucha exposición a los medios.

## SOCIABLES MUY INFLUYENTES

Gráfico 16. Distribución de casillas – tiempos de publicación

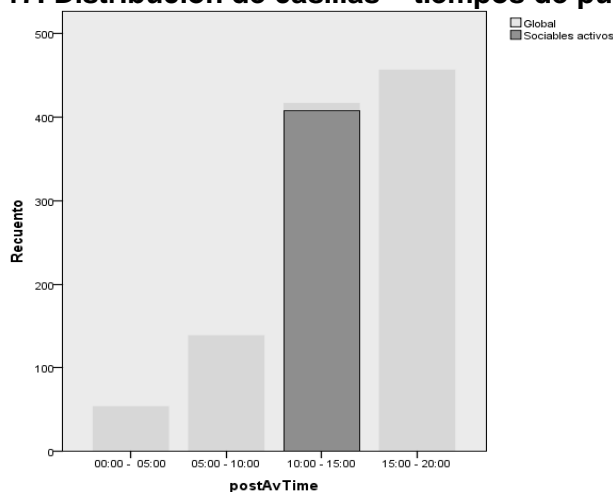


**Fuente:** elaboración propia (2016).

Representan 21.2% de la población y tienen mucha actividad dentro de la red. 61.5% de este grupo es matutino, publica entre las 05:00 y 10:00 hrs. Mientras que el segundo mayor porcentaje lo hace entre las 00:00 y 05:00 hrs. Este grupo tiene una media de 10482.40 followers y una influencia alta, con aproximadamente 386 comentarios por post. Asimismo, tienen una media de 916.78 posts, recibiendo un estimado de 385 comentarios por post. Estos suelen ser marcas en crecimiento que se promocionan en redes sociales o personalidades de las redes, como por ejemplo vloggers.

### SOCIABLES ACTIVOS

**Gráfico 17. Distribución de casillas – tiempos de publicación**

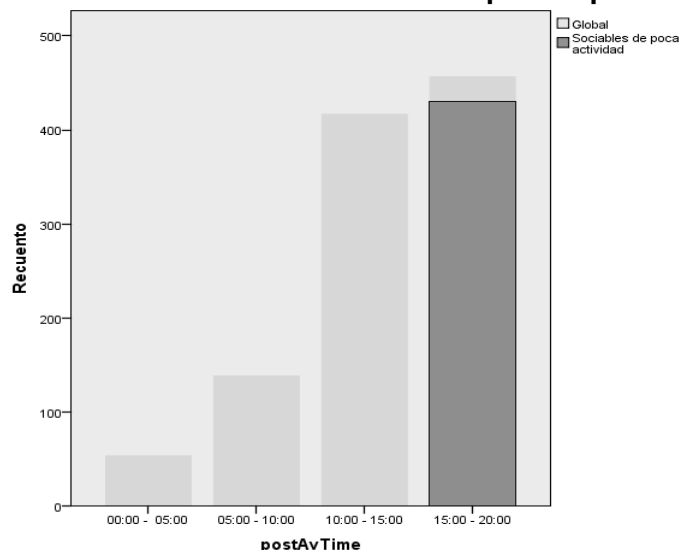


**Fuente:** elaboración propia (2016).

Este segmento representa 38,2% de la población y posee una media de 6881,1 followers. La hora de mayor actividad es entre las 10:00 y 15:00 hrs. Son sociables, pues publican con regularidad y tienen un promedio de 523,58 posts. Sin embargo, el nivel de influencia no es tan alto en comparación con los otros segmentos. Estos son usuarios que usan regularmente las redes sociales, pero su nivel de exposición no es tan alto. Parte de ellos son cuentas de usuarios públicas o de empresas que tienen una actividad regular en las redes sociales.

## SOCIABLES DE POCA ACTIVIDAD

Gráfico 18. Distribución de casillas – tiempos de publicación



**Fuente:** elaboración propia (2016).

Representan 40,3% de la población de estudio y tienen en promedio de 5303,23 posts en sus cuentas, con 389,37 publicaciones. A pesar de que el número de publicaciones es relativamente bajo en comparación con los segmentos anteriores la influencia es alta, pues, en promedio obtiene 296,24 comentarios por post y 7369,5 likes por post. En general, se trata de personas de baja actividad que cuando publica, lo hace entre las 15:00 y 20:00 hrs. Suelen ser usuarios cuyo perfil es público, pero con una interacción especialmente dentro de su red de amistades.

## CONSIDERACIONES FINALES

Para poder utilizar el Big Data en la segmentación de mercados en redes sociales para el sector de accesorios de moda emergente, fue necesario atravesar un proceso de investigación en el cual participaron varios expertos de distintas áreas. Luego de aplicar los instrumentos y la consulta a los expertos, se pudo verificar la teoría y la viabilidad de la investigación. Posteriormente, se dio inicio a la primera fase de la investigación, donde se aplicó el instrumento validado por los expertos a la población de estudio. En esta fase, luego de estudiar una muestra de 384 sujetos, fue posible identificar que la población no se comporta de una manera totalmente homogénea; y luego de varios análisis fue posible agrupar los predictores más significativos al momento de realizar los conglomerados, obteniendo un grado de cohesión, según el software *SPSS Statistics*, de tipo regular.

A este respecto, el algoritmo utilizado por el software fue el análisis bietápico, sugiriendo la existencia de tres (3) *clusters* o conglomerados, los cuales sirvieron como base para el análisis y determinación de los distintos segmentos. En esta ocasión, se

dividieron los segmentos de mercado en tres: (a) cauteloso – dinámico, (b) práctico – equilibrado y (c) explorador – experimentador.

A propósito de ellos, el primero denominado cauteloso-dinámico está representado por 49.2% de la población, la mayoría de ellos tiene edades comprendidas entre 25 y 30 años, se caracterizan por buscar practicidad y calidad en los productos. Asimismo, son profesionales independientes o emprendedores, siendo cautelosos con sus inversiones y gastos. Descubren nuevas marcas especialmente a través de las redes sociales y el dispositivo más común para hacerlo es el teléfono móvil. Tienen un estilo de vida dinámico y variado, son sociables y realizan actividades como ir a cafés o ejercitarse. Al momento de comprar, buscan una relación calidad-precio.

El segundo segmento representa 25.8% de la población, se denominaron prácticos-equilibrados, porque al momento de elegir un producto buscan que sean versátiles, fáciles de combinar y que estén a la moda; además en cuanto a su estilo de vida, existe un equilibrio entre lo laboral, académico y recreativo. Es un segmento cuya mayoría se encuentra en edades comprendidas entre 30 y 35 años, tienen una visión clara sobre su estilo, dando prioridad al diseño de los productos.

La mayor parte de ellos utiliza la laptop para visitar las redes sociales, sin embargo, descubren nuevas marcas a partir de la interacción con sus amistades. Asimismo, otro factor caracterizador es el hecho de que tienen trabajo estable y les gusta leer o realizar algún tipo de actividad académica. El otro segmento, representa 25% de la población, fue denominado exploradores - experimentadores; tienen un estilo de vida variado y dinámico, es decir; practican diversas actividades como asistir a teatros y/o fiestas, les gusta la interacción en redes sociales y con frecuencia descubren nuevas marcas de moda en bazares.

Del mismo modo, se debe mencionar que la mayoría se preocupa porque los productos estén a la moda y que posean diseños acorde a sus expectativas. Por esta razón pudo deducirse que no existe un estilo determinado que los defina, pues están en constante cambio detrás de las nuevas tendencias. Hay que destacar que en este segmento los rangos de edades fueron más amplios y se dividió especialmente entre los rangos 15-20 años y 20-25 años; dividiéndose también el estatus laboral, entre aquellos que no tienen empleo y los que tienen un trabajo estable.

Posteriormente, se dio inicio a la fase dos de la investigación, la cual se orientó hacia las redes sociales. Luego de analizar los resultados de la fase previa, se obtuvo como resultado que la red social a analizar debía ser Instagram, puesto que allí se encuentra la mayor parte de las marcas e interacciones con las mismas. Para poder extraer los datos, se diseñó un programa que recibiría como parámetros el nombre de la cuenta y extraería los datos de cada uno de los seguidores o usuarios que interactuaran con ella. Para ello, se seleccionaron ocho marcas y se estudiaron 3017 usuarios, 603932 posts y 408189 comentarios.

Después de la extracción, otro programa diseñado para ello realizó los cálculos preliminares que seguidamente se introdujeron en el *SPSS Statistics*, donde se utilizó

nuevamente el algoritmo de análisis de conglomerados bietápico. Para esto, se tomaron como principales variables los datos recolectados de la red social, a saber: cantidad de posts, *likes*, horas de publicación, *followers* y *following*; para luego realizar los cálculos pertinentes, (promedio de comentarios por post, promedio de likes por post, engagement, entre otros).

Los resultados arrojaron cuatro segmentos bien definidos, basados principalmente en la hora de mayor actividad, las interacciones recibidas y la cantidad de *followers* de los grupos de usuarios. A este respecto, se denominó a estos conglomerados de la siguiente manera: (a) Sociables altamente influyentes (b) Sociables muy influyentes (c) Sociables activos y (d) Sociables de poca actividad.

El segmento de los sociables altamente influyentes es el más pequeño de la población estudiada, está representado por 0,3% de la población y son diurnos. Tienen mayor actividad entre las 10:00 y 15:00 hrs. Son muy influyentes y se caracterizan por tener alta actividad de publicación (cerca de 5000) y más de dos millones de seguidores. Suelen ser celebridades o personas con frecuente exposición a los medios.

Los sociables muy influyentes representan el 21.1% de la población y tienen mucha actividad dentro de la red. 61.5% de este grupo es matutino, publica entre las 05:00 y 10:00 hrs. Tienen una media de 10482 *followers*, recibiendo aproximadamente 386 comentarios por post. Se caracterizan por ser marcas en crecimiento que se promocionan en redes sociales o personalidades de las redes, como vloggers.

Por otra parte, los sociables activos representan 38,2% de la población y poseen una media de 6881,1 *followers*. La hora de mayor actividad es entre las 10:00 y 15:00 hrs. Son sociables, pues publican con regularidad y tienen un promedio de 523,58 posts. El nivel de influencia no es tan alto en comparación con los otros segmentos. Estos son usuarios que usan regularmente las redes sociales, pero su nivel de exposición no es tan alto. Parte de ellos son cuentas de usuarios públicas o de empresas que tienen una actividad regular en las redes sociales.

El último segmento son los sociables de poca actividad. Representan 40,3% de la población y tienen en promedio 5303,23 posts en sus cuentas, con 389,37 publicaciones. En general, se trata de personas de baja actividad, que cuando publican lo hace entre las 15:00 y 20:00 hrs. Suelen ser usuarios cuyo perfil es público, pero que interactúan especialmente dentro de su red de amistades.

A pesar de que se hallaron cuatro conglomerados, no debe ignorarse el hecho de que se estudiaron 3017 sujetos de los cuales 1781 poseían cuentas privadas, por lo cual no pudieron ser analizados. Pudiendo considerar a este último grupo aparte como usuarios de poca exposición. Estos por lo general tienen mucho menos seguidores e interacciones por parte de otras cuentas. Sin embargo, representan más de la mitad de los usuarios estudiados, indicando que ellos interactúan frecuentemente con las marcas del sector.

## REFERENCIAS BIBLIOGRÁFICAS

- Aimia. (2015). Aimia. Documento en línea. Disponible en: <http://www.aimia.com> Consulta: 3/07/2015.
- Arias, F. (2012). Proyecto de investigación: Introducción a la investigación científica. México. Editorial Episteme.
- Bernal, C. (2010). Metodología de la investigación. México. Editorial Pearson.
- Dannah, B. y Ellison, N. (2007). Social Network Sites: Definition, History, and Scholarship. Documento en línea. Disponible en: <http://www.danah.org/papers/JCMCIntro.pdf>. Consulta: 10/07/2015.
- Feinleib, D. (2014). Big Data. Bootcamp. Estados Unidos. Editorial Apress.
- FORBES. (2016) How Facebook Turned its Gratest weakness into its Greatest Strength. Documento en línea. Disponible en: <http://www.forbes.com/sites/briansolomon/2016/01/28/how-facebook-turned-its-greatest-weakness-into-its-greatest-strength/#458cc97c4a06>. Consulta: 15/05/2017.
- Gettler, M. Bottou, L. Goldfarb, B. Murtagh, F. Pardoux, C. y Touati, M. (2010). Statistical learning and Data Science. United States. Chapman & Hall/CRC.
- Hootsuite, (2016). Hootsuite. Documento en línea. Disponible en: <http://hootsuite.com>. Consulta: 26/10/2016.
- Haupt, M. (2006). Data is the new Oil. Documento en línea. Disponible en: <https://medium.com/twenty-one-hundred/data-is-the-new-oil-a-ludicrous-proposition-1d91bba4f294> Consulta: 06/11/2014.
- Kotler, P. Keller, K (2012). Dirección de marketing. Edición 14, (Pp. 212-239).
- Lamb, C. Hair, J. y McDaniel, C. (2011). Marketing. 11va ed. México.
- Mullins, J. Walker, O. Boyd, H. y Larréché, J. (2007). Administración de marketing. México. Editorial McGraw-Hill.
- Neumann, K. (2011). Social Media als Marketing-Instrument für Unternehmen. Master en gerencia de la información. Hannover, Fachhochschule
- ONTSI (2011). Las redes sociales en internet. Documento en línea. Disponible en: [http://www.osimga.gal/export/sites/osimga/gl/documentos/d/20111201\\_ontsi\\_redes\\_sociais.pdf](http://www.osimga.gal/export/sites/osimga/gl/documentos/d/20111201_ontsi_redes_sociais.pdf). Consulta: 2/07/2015.
- Provost, F. y Fawcett, T. (2013) Data Science for Business. Estados Unidos. Editorial O'Reilly.





Zikopoulos, P. deRoos, D. Eaton, C. Lapis, G. Deustsch, T. (2012) Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. United States. McGraw-Hill.