

GLOSARIO

RESUMEN

0. INTRODUCCIÓN

1. PROBLEMA

1.1 PLANTEAMIENTO DEL PROBLEMA

1.2 FORMULACIÓN DEL PROBLEMA

1.3 MARCO TEÓRICO

1.3.1 Analítica web y análisis de un blog

1.3.2. Google Analytics y blogs

1.3.3. Blogger

1.3.4. Análisis estadístico de un blog

1.3.5 .Knowledge Discovery in Databases (KDD) y análisis de un blog

1.3.6. Análisis de componentes principales

2. OBJETIVOS

2.1 OBJETIVO GENERAL

2.2 OBJETIVOS ESPECÍFICOS

2.3. VARIABLES

3. MATERIALES Y MÉTODOS

3.1 Tipo de estudio

3.2 Población

3.3 Muestra

3.4 Metodología

3.4.1 Análisis estadístico

4. RESULTADOS

4.1 Análisis de estadísticas básicas

4.2 Análisis de componentes principales

5. CONCLUSIONES

6. RECOMENDACIONES

7. BIBLIOGRAFÍA

GLOSARIO

BLOG: Bitácora en tiempo real, donde uno o varios autores publican textos u otros tipos de multimedia, que pueden ser retroalimentados por los usuarios del sitio a través de comentarios. Generalmente se utiliza un blog para describir temas muy específicos o para dar opiniones propias sobre todo tipo de asuntos en general.

BLOGUERO: Persona dedicada a escribir o publicar contenido multimedia para un blog, sea este trabajo remunerado o no. Hay *bloggers* tanto independientes como contratados, como los *bloggers* de las webs de periodismo digital.

BLOGÓSFERA: Conjunto de blogs que se encuentran en Internet en su totalidad. La blogosfera se caracteriza porque los sitios en ella se relacionan por medio de hipervínculos que permiten al usuario visitar otras webs. Esta característica es inherente a los *weblogs*.

CSS: *cascade-style-sheet* es un conjunto estructurado de HTML. Se compone de hojas de estilo, las cuales pueden ser modificadas independientemente según las necesidades del usuario.

DÓLAR ÍNDICE: Es una media geométrica ponderada del valor del dólar en comparación a otras monedas internacionales. Cada visita a un blog y cada conversión poseen un índice de ganancias en centavos de dólar que puede ser relacionada con este índice para calcular las ganancias en diferentes monedas, incluyendo el peso colombiano.

FOTOBLOG: Bitácora online que publica fotografías en forma cronológica y ordenada, en forma de galerías gráficas que pueden ser de fotografías, imágenes históricas e inclusive diseños propios con el fin de servir de sustento a la información que se brinda desde un blog.

GOOGLE ANALYTICS API: Aquel componente que permite obtener las métricas básicas de un blog o weblog mediante la inserción de un código de datos de Google Analytics y que puede ser incorporado en aplicaciones y / o bases de datos con el fin de monitorear diferentes variables, según las necesidades del usuario.

HTML: Lenguaje de marcado de hipertexto para la construcción de páginas web. Como indican sus siglas, *Hyper Text Markup language*, utiliza marcas como identificadores de los diferentes segmentos en el código fuente de una página web.

NUEVO VISITANTE: Se trata de un visitante que no posee cookies de Google Analytics para realizarle seguimiento dentro de un blog. Cuando un visitante nuevo ingresa por primera vez en un blog, se le imputa un segmento de código para hacerle seguimiento desde su navegador de preferencia. Esto es llamado cookie.

METADATOS: Son datos sobre datos, es decir, información estructurada que describe a otra información y que permite encontrarla, gestionarla, controlarla, entenderla y preservarla en el tiempo, de tal manera que puede ser interpretada en cualquier momento.

MINERÍA DE DATOS: Conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto.

PÁGINAS VISTAS DIVIDIDAS ENTRE LAS VISITAS: Métrica que muestra el promedio de páginas vistas por visita, en forma de variable continua.

SEGUIMIENTO DE EVENTOS: Característica que permite realizar un seguimiento de las actividades de los visitantes (Como cuando se determina un objetivo final en una web o blog y un usuario llega allí). Esto es comúnmente utilizado para rastrear la interacción con plataformas como AJAX o el contenido gestionado en Flash.

SERVIDOR: Sistema que proporciona recursos a un número variable de usuarios de tal manera que proporciona un servicio online, ya sea de gestión de datos (mails), o de otros tipos de tareas (almacenamiento, por ejemplo). El servidor puede poseer una red interna o externa, de tal manera que a la interna solo se puede acceder a través de una línea privada y a la externa se puede acceder públicamente.

SITIOS WEB DE REFERENCIA: El tráfico para el cual se identificó un referente, y cuya referencia no es un motor de búsqueda. Además, para ella no hay variables de campaña. La URL de referencia (también conocido como la página que contiene el enlace a su sitio web) también se almacena para posteriores análisis relacionados con las principales fuentes de visitas referidas al blog o sitio web.

TRÁFICO DIRECTO: Idealmente, este es el tráfico que viene a un sitio a través de marcadores (links directos) o escribiendo directamente en la URL la dirección del blog o página web en el navegador. En realidad, es el tráfico para el que el código

de análisis de google analytics no pudo determinar una fuente. Dependiendo del sitio y el navegador, algunos enlaces no pueden mostrar un referente y en su lugar la visita se clasifica como directa.

TRAFFIC SEARCH ENGINE: Son aquellas visitas que Google Analytics clasifica automáticamente como procedentes de un motor de búsqueda, si la URL de referencia se encuentra entre su lista de motores de búsqueda conocidos y hay un término de búsqueda que sea reconocible en la URL. Tanto el tráfico de motores de búsqueda orgánica (gratuita) y de campaña (pago) se pone en este grupo.

TIEMPO DE PERMANENCIA DEL USUARIO EN EL BLOG EN SEGUNDOS: Se refiere al tiempo total obtenido desde el momento que un usuario ingresa a un blog hasta que sale de él sin importar el número de páginas visitadas en ese periodo, medido en segundos.

TIEMPO DE PERMANENCIA EN EL SITIO: Se refiere a la suma total del tiempo de permanencia de un usuario en la página para todas las páginas vistas en una visita al blog. Se ha de tener en cuenta que la visualización de páginas en diferentes pestañas en un navegador no afecta este parámetro y en esto difiere del tiempo de permanencia del usuario. Google Analytics simplemente ve una serie de páginas que se vieron en orden cronológico, sin referencia alguna a varias pestañas o ventanas, ya que toma su información a través de una IP única, proporcionada por el código API.

VIDEOBLOG: Galería online de vídeos que se publican de forma cronológica con una periodicidad determinada por el videoblogger. Con servicios gratuitos online como YouTube o Vimeo es más fácil distribuir el vídeo y crear un canal para que pueda ser visto en todas partes del mundo.

VISTA DE PÁGINA: Registro que se da cada vez que el visitante realiza un *hit* (abre una pestaña nueva en el blog o sitio web), de tal manera que se cuenta cada vez que el usuario cambia de página, en forma de variable discreta.

VISITA ÚNICA: Es el acceso de un visitante a un sitio web o blog, medido en forma de variable discreta, determinado por un tiempo de estadía máximo, generalmente 180 segundos.

VISITANTES ÚNICOS: Son los usuarios de un sitio web que se encuentran definidos por un identificador único de IP establecido en el navegador del usuario a través del API proporcionado por Google Analytics.

VISITAS ÚNICAS: Es una visita con una sola interacción de página, es decir que el visitante sólo visitó una única página una única vez dentro de su visita en el blog.

WEBMASTER: Persona responsable del mantenimiento, programación y contenido de un sitio web. En el caso de un blog, el webmaster toma la tarea de alimentar con información el sitio y responder a las dudas, inquietudes y sugerencias de los usuarios.

WEBSITE: Colección de páginas compiladas por un navegador web en formato HTML relacionadas bajo un mismo dominio (*.com, .net, .org*) y que son accesibles a través del protocolo HTTP de internet.

RESUMEN

Actualmente Internet es de las mejores fuentes de información debido al intercambio inmediato de datos en tiempo real que este sistema virtual basado en ARPANET puede ofrecer. Algunos de estos datos son sobre el uso o el movimiento de una web, un blog o cualquier otro tipo de programa, incluyendo el de las aplicaciones y los dispositivos móviles de última generación y pueden ser utilizados para realizar predicciones y dar explicaciones a comportamientos que no se verían a simple vista.

El problema es que no hay un consenso sobre qué técnica debe utilizarse para el análisis de esos datos y el mismo es subjetivo al investigador, al sitio y a los datos que se quieran investigar. Para explorar un caso se recolectaron las variables relevantes de un blog con temática paranormal entre 2008 y 2016, buscando que análisis estadístico es pertinente para su análisis.

Los datos de este espacio web fueron extraídos de Google Analytics gratuitamente y pueden ser recolectados en forma ordenada de tal manera que pueden organizarse efectivamente para la realización de un análisis estadístico. Ya que se deben estudiar múltiples variables relacionadas entre sí sobre un marco de países amplio, una forma óptima de metodología aplicable sería el Análisis de Componentes principales con el fin de reducir a componentes principales las variables analizadas y describir su comportamiento a través de las gráficas que este análisis arroja.

Es así que se busca determinar si al realizarse un ACP (Análisis de componentes principales) sobre los datos recolectados de Google Analytics de un blog independiente, se pueden ver relaciones relevantes para aumentar la cantidad de visitas obtenidas con el blog dados 85 países de los cuales se reciben más visitas.

Palabras claves: Blog, Análisis, Google Analytics, Componentes Principales, Estadística Aplicada, Marketing.

0. INTRODUCCIÓN

El valor de la información de un blog puede ser medido de varias maneras: por la asistencia de su público, su información entrópica e inclusive por medio del análisis de las palabras más utilizadas. No obstante es un problema común para los escritores de una bitácora online el poder realizar un análisis de esta información; Más todavía si se sabe de antemano que las métricas que pretenden extraerse no pueden conseguirse fácilmente dado que el blog se encuentra en un servidor privado (En el caso de blogger, los servidores de Google), además de no haber una manera online óptima y gratuita de realizar este análisis a través de alguna plataforma online, siendo las plataformas online tan específicas y caras que muy pocos sabrían cómo hacerlo.

Una solución rápida y eficaz para el problema anterior es la extracción de metadatos desde la plataforma Google Analytics, utilizada comúnmente para la recolección de la información de blogs mediante un código *Javascript*, pudiéndose importar desde un blog o web en tiempo real mediciones y datos relevantes hacia los servidores de Google inc. quienes procesan la información, devolviendo así resultados en forma de conjuntos de bases de datos llamadas métricas. De allí pueden extraerse conjuntos de estas métricas relevantes para obtener información sobre cualquier sitio del cual se posea la autoría y/o permisos de acceso.

Uno de los problemas con este método de extracción de información, es que es solo un método de recolección y extracción de información, no es una plataforma de análisis estadístico. Además representa la información recolectada para cada variable independientemente, cosa que no permite ver si existen relaciones relevantes entre las variables del sitio.

Para poder comprobar si en realidad las variables que se están analizando y que se recogen de Google Analytics influyen de alguna manera en la cantidad de visitas del blog es necesaria la aplicación de pruebas estadísticas específicas que demuestren hechos con los cuales las variables puedan ser interpretadas y aplicadas a mejorar el blog según los resultados.

El análisis de componentes principales se presenta como un candidato óptimo de metodología para el procesamiento de los datos recolectados ya que permite el visualizar cual es la relación de cada una de las variables con las visitas totales dada la variabilidad que presentan datos graficados por este método, permitiendo percibir con esto el comportamiento de los datos para los países que se estudiaron y de los cuales se reciben visitas frecuentemente y en mayor cantidad. Más allá de eso, permite apreciar las relaciones de los países con todas las variables que se

vayan a estudiar, permitiendo esto ver cómo se relacionan los países observando si estos poseen comportamientos similares.

Es por esto que se busca realizar un análisis de componentes principales a un blog de temática paranormal que pueda ser replicado con cualquier blog, con información recolectada para 85 países desde Google Analytics, entre los años 2008 y 2016, con el fin de saber cuáles son las conclusiones a las que se puede llegar mediante la aplicación de esta metodología y si las mismas son satisfactorias para determinar formas de aumentar el número de visitas en el blog.

1. PROBLEMA

1.1 PLANTEAMIENTO DEL PROBLEMA

Vivimos actualmente en la sociedad del conocimiento (García, 2003) donde el avance de las Tecnologías de la información y comunicación -TIC- ha brindado la posibilidad a las personas de expresarse libremente y compartir a través de medios digitales en la Internet (Ayala, 2015).

Uno de los espacios más populares para realizar este tipo de actividades es el Blog; investigadores señalan que el 71% de los 625 millones de usuarios activos de Internet en realidad leen blogs y que por tanto el público con que cuenta este medio rivaliza con los medios de comunicación tradicionales. (Melville p, 2009). Este medio de comunicación se define como un sitio web que incluye, a modo de diario personal de su autor o autores, contenidos de su interés, actualizados con frecuencia y a menudo comentados por los lectores (RAE, 2015). El blog funciona como una bitácora personal online prediseñada, que puede ser tanto pública como privada (San agustín, 2013)

Si bien este nuevo tipo de tecnología facilita a las personas el poder compartir su información de una manera rápida, eficaz y por sobre todo sencilla, también dificulta la extracción de metadatos directamente de la web o blog, ya que, al no poseer acceso a la información del servidor, se dificulta la tarea del *webmaster* de poder extraer información relevante de su espacio en Internet y esto presenta un primer problema ya que estos datos son importantes para el análisis estadístico de cualquier espacio web.

Para solucionar esta inconveniencia de primera mano podemos observar entre los servicios libres ofrecidos en internet, siendo el más famoso de ellos Google Analytics, plataforma online que puede ser utilizada para recolectar, segmentar y describir el tráfico de visitantes de un espacio online detalladamente. También permite supervisar todos los aspectos del tráfico de sitios web y de las referencias de palabras clave dada la actividad del motor de búsqueda Google. Al conectar este servicio con un blog o website, inmediatamente empezarán a obtenerse las estadísticas y variables más relevantes para una página web (Sugiarto, 2013).

Para medir la cantidad de visitas en un blog es importante medir las métricas relacionadas con las fuentes de tráfico (Adquisición de las visitas), la efectividad de los contenidos (Comportamiento de los usuarios), si las hay, Descargas en blog, ingresos totales del sitio, cantidad de comentarios por una unidad de tiempo

dada, número de Suscriptores a los newsletters (suscripción por correo), y últimamente, Social Share y Engagement online (resiliencia en las redes sociales)(Nuñez,2005) aunque el análisis de cada una de estas características por separado también es válido, dependiendo de las necesidades que posea el investigador (Arimetrics, 2013).

Aunque Google Analytics es una plataforma de recolección de datos muy completa, se queda corta a la hora de la realización del análisis estadísticos con las variables que pueden extraerse de ella. Muchas veces esta plataforma solo toma las variables y las analiza de tal manera que puede extraerse la estadística descriptiva a través del tiempo de cada una de las variables o de forma pareada, más no analiza en conjunto de variables que extrae para saber si estas poseen características comunes entre sí o disimilitudes notables (Rouse, 2011).

Para este caso en particular, es propicio hacer el análisis estadístico tomando solamente la característica de fuente de tráfico, con la finalidad de que el autor del blog mejore en las características de tráfico que no sean tan robustas en el blog y que quieren fortalecerse.

A su vez, se agregó al análisis la variable “ingresos netos del blog”, extraído de la plataforma hermana de Google Analytics “Google AdSense”, quien provee datos de los ingresos generados por el programa publicitario de afiliados a través del tiempo. Estos ingresos son generados por publicidad exterior puesta dentro del espacio web afiliado mediante un código HTML preestablecido.

Se tuvieron en cuenta varios análisis antes de poder dar una apreciación favorable. Entre ellos, la regresión lineal simple y la regresión lineal múltiple, los cuales fueron descartados por no profundizar lo suficiente en el análisis requerido para los datos recolectados. También se tuvo en cuenta el análisis cluster, más este tuvo inconvenientes ya que no arrojó ningún resultado que pudiera ser utilizado para aumentar las visitas del blog al no presentar ninguna prueba concluyente para mejorar el tráfico del blog (solo lo describe).

Es por esto que el análisis de componentes principales surge como una de las alternativas de análisis estadístico para agrupar las variables del presente estudio, con el fin de lograr encontrar una relación que pueda ser fácilmente reconocida a través del análisis de las gráficas de los componentes principales obtenidos a través de los datos recolectados. Esto dado que el número de variables trabajadas es mayor a 2. Así, se ve conveniente la reducción de dimensiones que ofrece este análisis para poder analizar más fácilmente las relaciones que existen y que aumentan las visitas obtenidas por el blog.

Para ello se tiene en cuenta el proceso de reducción dimensional de variables y otros procedimientos estadísticos en los cuales se ahondará más adelante, y de los cuales fácilmente se puede realizar una gráfica donde se identifica el comportamiento de los datos y sus relaciones entre sí y para con los países estudiados, enfrentando las variables de tal manera que en el resultado se aprecie una relación de la cual se pueda concluir el comportamiento de los países frente al tráfico web que llega al blog.

1.2 FORMULACIÓN DEL PROBLEMA

¿Qué información se puede obtener al realizarse un Análisis de Componentes Principales sobre los datos de un blog paranormal independiente dado que se quieren aumentar el tráfico web del sitio?

1.3 MARCO TEÓRICO

1.3.1 Analítica web y análisis de un blog

Cutroni, 2010; Maldonado, 2010; Morales, 2010; Ledford, Teixeira & Tyler, 2011; Acera, mencionado por Cabero, 2013; Sostienen que la analítica web inicialmente comenzó a ser utilizada en la parte comercial sobre los sitios web construidos por las empresas y la accesibilidad de las mismas (Cabero, 2013). Se ha definido analítica web (análisis web o web Analytics) como el acto de recolectar, medir y analizar los datos provenientes de los reportes y el análisis del uso de una web, dichos datos pueden ser usados para la interpretación del paso de tráfico a través de la web, como insumo para la investigación en marketing y medición de la efectividad de una página en internet.

Existen dos tipos de analíticas Web: *On-line* y *off-line*, la primera se refiere al conjunto de mediciones cuando el usuario se encuentra en el sitio (la duración de las visitas, promedio de tiempo en el sitio, número de impresiones de página que el usuario ve por sesión, etc.) y la segunda detalla las métricas que pueden ser recolectadas con el fin de conocer el impacto del sitio web en la blogosfera, ejemplo de ellas son: Número de menciones que tenga la página en otras web relacionadas –Hipervinculación- o la cantidad de trinos –comentarios- que posee el website en su totalidad (Quintero, 2000).

Los datos del sitio web pueden ser recolectados técnicamente de dos formas: La primera se llama *server log file análisis* (James, 2000) y estudia directamente la información que se encuentra almacenada en el *log* del sistema del servidor. El segundo método se llama *page tagging* (Herman, 2010) y analiza los tags de las páginas en las campañas publicitarias.

Sucede a menudo que un tercero posea la tecnología para el análisis, el cual realiza en el website mediante un código de seguimiento API que recolecta la información y luego la brinda al cliente de manera organizada y comprensible. La anterior información, puede ser utilizada para analizar un blog independiente porque no solo es solo el esfuerzo y la disciplina de hacerlo sino mantenerlo, generar contenido de calidad, promocionarlo, buscar links, dedicarle tiempo a la analítica para poder entender qué está pasando con el espacio web y este tipo de plataformas permiten saber si está teniendo el éxito que se espera o medir cualquier otra variable que sea de interés para el administrador del blog.

1.3.2. Google Analytics y blogs

Google Analytics es un servicio gratuito de análisis web que ofrece estadísticas y herramientas analíticas básicas para la optimización en motores de búsqueda (SEO) con fines de marketing gerencial o de análisis de datos para el webmaster. (Rouse, 2011). Está construido para sitios de pequeño y mediano tamaño (aquellos que reciben menos de un millón de visitas al mes), y para aquellos con más de un millón de visitas se transforma en un servicio de pago.

El sistema es lo suficientemente versátil como para realizar seguimiento a clientes específicos, monitoreo en motores de búsqueda, analizar el impacto del contenido de una web, así como el rendimiento y los análisis de navegación de los usuarios e inclusive el seguimiento en cualquier variable en función de una variable de tiempo requerida cualquiera, desde los segundos (en tiempo real) hasta años de ser necesario.

El sistema recolecta los datos a través de un código *JavaScript* que se inserta en la etiqueta BODY del sitio y cuya extensión es .js. Una vez instalado, este carga información hacia los servidores de google permitiendo monitorizar la web, devolviendo luego la información obtenida hacia Analytics, quien finalmente la presenta al usuario mediante una plataforma online de fácil acceso.

La plataforma virtual arroja mediciones de las variables en forma de gráficos bidimensionales, cuya información puede ser exportada en bruto fácilmente a formatos de diferentes tipos de editores de bases de datos como Excel y Access.

1.3.3. Blogger

Entre lo ofrecido por Google, existe el servicio denominado Blogger, el cual es una bitácora en línea, también llamado *Weblog*, o *Web-log*; Se define como un registro cronológico personal de los pensamientos publicados en una página Web (Dictionary, 2015). Su principal característica es que el usuario no tiene que ejecutar o escribir ninguna clase de código para poder obtener la información deseada del sitio. Así, inclusive los blogs en Blogger que no poseen un dominio propio y que se encuentran alojados en blogspot.com pueden ser partícipes del seguimiento de la plataforma.

Específicamente, el sistema de Blogger ayuda a simplificar el proceso de publicación de bitácoras en internet, omitiendo la parte en la cual el usuario debe actualizar el código HTML o CSS del sitio web a mano y simplificando el proceso de escritura mediante la opción de publicar con solo llenar un formulario preestablecido en una interfaz gráfica amigable al usuario. La plataforma es multicanal, aceptando gran variedad de contenidos multimedia, como fotografías y videos, en gran cantidad de formatos digitales, siendo posible la modificación de su posición directamente desde el HTML o por medio de múltiples aplicaciones y atajos que ofrece el mismo blogger.

La plataforma fue lanzada en 1998 y en 2003 fue adquirida por Google inc. Desde ese momento, el sitio posee características nuevas como plantillas prediseñadas con estilos CSS, *labels* para las entradas, redireccionamiento de segundo nivel a servidores locales de cada país para mejorar la moderación según la legislación local y para evitar la censura, además de otras características (Tipanta, 2010).

1.3.4. Análisis estadístico de un blog

Con el auge de las páginas Web y de los blogs, ha surgido el interés de desarrollar estudios para comparar los datos arrojados por Google Analytics y surge el interrogante de saber qué tipo de estrategia metodológica se debe utilizar para su análisis, tomando la estadística como punto de partida.

De primera mano se pensó un análisis cluster para ver en un gráfico claro cómo se relacionaban las variables en grupos de individuos diferenciados por la cantidad

de dinero obtenido y si alguno de estos segmentos temporales podía dar una idea sobre cómo aumentar las ganancias del blog. Si bien esto funcionó para visualizar cuáles eran los segmentos con mayor índice de visitas, no sirvió para determinar cómo aumentarlas así que también se descarta como opción estadística.

Finalmente se optó por el Análisis de Componentes Principales. El objetivo de realizar este análisis, fue identificar información a través del análisis de la variabilidad que presentan N variables reducidas a Y componentes principales y que son puestos gráficamente con el fin de saber, en este caso, si un blog puede generar ganancias mayor tráfico weba través del tiempo dada la variabilidad de los componentes frente a la de los ingresos netos. Para ello se utilizó el análisis de Componentes Principales (ACP) el cual es definido como un método que trata de sintetizar y dar una estructura a la información contenida en una matriz de datos (Bernal, 2004); Es decir, ante un banco de datos con 6 variables, el objetivo será reducirlas a un menor número perdiendo la menor cantidad de información posible (Gurrea, 2015).

Adicionalmente este método permitió reducir la dimensionalidad de los datos, transformando el conjunto de N variables originales en otro conjunto de Y variables correlacionadas llamadas componentes principales (González, 2014), lo cual optimiza el proceso de análisis de muchos datos al reducir su cantidad. El uso principal del ACP en este caso es el ser usado como técnica de análisis exploratorio para descubrir interrelaciones entre los componentes reducidos para las variables mensuales de promedio de impresiones de página por sesión, la duración media de la visita en segundos, el total de visitas de usuarios por mes, el promedio de páginas impresas por los usuarios por sesión única de visita y los ingresos netos por día y de acuerdo con los resultados, proponer los análisis estadísticos más apropiados para el aumento del tráfico web del blog estudiado.

1.3.5 .Knowledge Discovery in Databases (KDD) y análisis de un blog

El Knowledge Discovery in Databases (KDD) es un proceso mediante el cual se recogen, organizan y analizan datos extraídos de diferentes fuentes de información para su posterior aplicación a diferentes áreas del conocimiento. Para la realización del KDD en este caso primero se buscó una fuente de donde puedan extraerse lo datos en bruto, en este caso, Google Analytics. Luego se unificó la información en una base de datos en la plataforma excel y se organizó de manera que la información fuese de fácil acceso para su posterior ingreso en SPAD. Tras esto, y ya con los datos organizados, se procedió primero a la extracción de las estadísticas descriptivas de las variables y luego a la aplicación de la técnica

estadística del ACP. Como resultado se obtiene una gráfica donde se puede observar, en forma de nube de puntos dispersos, la variabilidad de los componentes reducidos frente al tráfico web del sitio estudiado. Tras esto, debe esperarse la retroalimentación de otros experimentos similares que se realicen con el fin de comparar los resultados y llegar a conclusiones satisfactorias (Tabasco, 2007).

1.3.6. Análisis de componentes principales

El ACP o análisis de componentes principales busca resolver, de una manera estadísticamente acertada, si es posible extraer información de un grupo de múltiples variables, reduciendo estas a otro grupo de variables menor que describa el comportamiento de las mismas sin que en este proceso se pierda demasiada información (Gómez 1993). El proceso de reducción dimensional puede realizarse mediante dos métodos: Algebraico o geométrico, de tal manera que lo que se obtiene finalmente son nuevas variables llamadas componentes principales con características favorables entre sus varianzas, o sea, su variabilidad (IBID).

Este se considera un método esencialmente descriptivo pero puede tomarse en términos de método exploratorio. De ambas maneras, el proceso es el mismo: se tienen N variables primarias que, tras la reducción dimensional, se convierten en Y variables llamadas “componentes principales”, de tal manera que $Y < N$.

Lo más interesante de este análisis es que, una vez realizada la reducción dimensional puede observarse gráficamente cómo se relacionan los datos de los individuos de las dimensiones reducidas. Esto, además de presentar gráficamente la información que se busca, ayuda a simplificar posteriores análisis que vayan a ser realizados (Grané, 2005).

El número de componentes finales que resultan puede ser determinado desde varios métodos. El más sencillo es el porcentaje explicado, en donde se fija un porcentaje de variabilidad y a partir de esto, empiezan a reducirse las dimensiones hasta que ajuste a ese porcentaje. Otra manera común de determinar un componente es el criterio de Kaiser, en donde se excluyen las componentes que tengan menos de un λ determinado por $\bar{\lambda} = \text{tr}(S)/p$. por último, se puede aplicar la modificación de Jolliffe, el cual es un arreglo al criterio de káiser para cuando la muestra es menor a 20 datos. Esta dado por la formula: $0.7\bar{\lambda} = 0.7 \text{tr}(S)/p$ (Grané, 2005).

En el caso de este trabajo, el análisis de componentes principales se aplica con la intención de saber que se puede apreciar mediante la reducción dimensional de los datos o como o porque el tráfico del blog aumenta o disminuye, dando esto a entender si hay un patrón que pueda seguirse o interpretarse para aumentar el tráfico web del sitio y de ser así, implementar un plan que ayude a aumentar el tráfico del sitio con base en la información conseguida de este modo.

El análisis ACP parte de una matriz N de dos dimensiones ($Q \times S$) donde Q representa el número de las variables observadas y S representa el número de individuos analizados. Esta matriz debe estar estandarizada con el fin de que los promedios de las variables unitarias sean nulos así como también las varianzas.

Como resultado de esta operación se obtiene una segunda matriz M , dada por la fórmula:

$$M_{ij} = \frac{y_{ij} - \bar{y}_j}{S_{jn}}$$

Donde \bar{y} es el promedio y S la desviación estándar de cada una de las variables estudiadas.

De aquí se determinan los valores y vectores propios de cada una de las variables estudiadas a partir de la matriz $U = M'M$. Los valores corresponden a la varianza de las observaciones de los componentes resultantes de la matriz M y los vectores unitarios corresponden a las coordenadas iniciales de cada uno de los componentes principales Z , siendo así la ecuación matricial principal:

$$Z = NU$$

Donde z es el resultado de la matriz de componentes principales, N es la matriz con los datos estandarizados y U es la matriz resultado de $M'M$ o matriz de correlación. Es a través de esta fórmula que el análisis cluster puede representarse gráficamente de dos maneras:

Un diagrama de variables o un círculo de correlación: Allí se representan las variables en el plano factorial donde las coordenadas son los coeficientes de correlación entre las variables originales y los componentes principales.

Un gráfico de individuos: es un gráfico donde los individuos estudiados son representados gráficamente a manera de histograma de frecuencias.

2. OBJETIVOS

2.1 OBJETIVO GENERAL

Mostrar la aplicación de una técnica multivariable, el Análisis en Componentes Principales (ACP), en el campo del análisis estadístico sobre los datos recolectados de 2008 a 2016 de un blog con temática paranormal conectado a Google Analytics.

2.2 OBJETIVOS ESPECÍFICOS

- Describir el comportamiento de los países que recogen las observaciones mensuales de promedio de impresiones de página por sesión, la duración media de la visita en segundos, el total de visitas de manera discreta, el promedio de páginas impresas por usuario por sesión única de visita, los clics totales en el mes que dieron ganancias y los ingresos netos por día para un blog con información independiente para 84 países entre los años 2008 - 2016.
- Desarrollar un ACP para los países con mayor número de visitas según el promedio de impresiones de página por sesión, la duración media de la visita en segundos, el total de visitas de manera discreta, el promedio de páginas impresas por usuario por sesión única de visita y los clics mensuales que produjeron ganancias y los ingresos netos para 84 países.
- Establecer las relaciones entre los países que generan mayores visitas para aumentar las visitas del blog.

2.3. VARIABLES

Tabla 1: Variables utilizadas en el estudio

Variable	Definición	Operacionalización	Tipo	Recolección	Acrónimo
Promedio de impresiones de página de todos los usuarios al mes	Es el total de páginas vistas por los usuarios únicos sobre el total de usuarios únicos.	N° de páginas vistas por usuarios únicos en el mes	Cuantitativa continua.	Métricas de Google Analytics	PROIMPA
Duración media de la visita en segundos	Tiempo que dura el visitante del blog en promedio en segundos	Segundos	Cuantitativa continua	Métricas de Google Analytics	DURAMEDVIS
Total de visitas de usuarios únicos por mes	Total de visitas únicas a un blog medidas a través de la IP.	N° de visitas por usuarios únicos en el mes	Cuantitativa discreta	Métricas de Google Analytics	TOTVISITAS
Total de páginas impresas	Total de páginas impresas dentro de una visita al blog	N° de páginas impresas	Cuantitativa discreta	Métricas de Google Analytics	PAGIMPRESAS
Ingresos netos del blog al mes.	Total de dinero ganado en dólares por el blog en un mes	Dólares	Cuantitativa discreta	Métricas de Google Asense	INGRESONET
Clicks totales	Número de clicks totales obtenidos por mes	No de clicks totales	Cuantitativa discreta	Métricas de Google Analytics	CLICKS

Fuente: formato de recolección de datos.

3. MATERIALES Y MÉTODOS

3.1 Tipo de estudio

El estudio es de tipo descriptivo, ya que pretende describir cuáles son las relaciones entre el tráfico web que llega al blog y los países que más tráfico web generan según las variables estudiadas.

3.2 Población

Visitantes que llegan al blog y que dejan registro de sus visitas en las variables estudiadas.

3.3 Muestra

La muestra que se selecciona para este trabajo son las bases de datos ensamblada a través de los recolectados desde 2008 a 2016 de manera mensual para las 6 variables estudiadas.

3.4 Metodología

Lo primero que se realizó para obtener la información en la base de datos del blog fue recurrir a la recopilación de datos a través del proceso KDD (*Knowledge Discovery in Databases*) (Sofia, 2002), que engloba un conjunto de etapas que guían el desarrollo de un proyecto de esta naturaleza. Este proceso de organización de datos se realizó según las siguientes fases de desarrollo: Integración y Recopilación de la información en bruto, Selección de los datos relevantes a estudiar, Limpieza y Transformación de las variables estudiadas, Aplicación de técnica de organización a las variables en bruto, Evaluación e Interpretación de los resultados obtenidos y por último Difusión y uso de la información obtenida. A continuación se describen los pasos que se realizaron para completar a plenitud cada una de las etapas anteriormente mencionadas:

En la fase de integración y recopilación lo primero en realizarse fue la descarga desde Google Analytics de los datos en bruto para el blog estudiado. Luego se extrajeron y organizaron los datos de 2008 a 2016 por países (con un total de 92 meses para 85 países) para las variables relevantes al estudio, siendo estas las observaciones mensuales de promedio de impresiones de página por sesión, la duración media de la visita en segundos, el total de visitas de manera discreta, el promedio de páginas impresas por usuario por sesión única de visita, los clicks totales y los ingresos netos por mes.

En la fase de selección y limpieza se depuró la base de datos de información que no tuviera que ver con las características seleccionadas, pero para este caso específico, con el fin de obtener datos mucho más precisos, se dejaron los datos organizados, pero intactos. Es decir, no se reemplazaron los ceros por ningún tipo

de medida de tendencia central para armonizar el estudio, de tal forma que los resultados fuesen lo más apegados a la realidad en la medida de lo posible.

Para ello, se generó una base de datos en Microsoft Excel donde se organizaron los datos en orden cronológico de tal manera que de ser la variable discreta se tomaron hasta máximo dos dígitos después de la coma y que los datos en ceros no se tomaron como información perdida sino como ceros. Para empezar la implementación del proceso de realización de las estadísticas descriptivas y luego el proceso de análisis de componentes principales se tomaron estos datos y se introdujeron al programa SPAD, finalmente se obtuvo una base de datos organizada para 85 países con variables tanto continuas como discretas, requiriendo ACP's complementarios para poder realizar la comparación entre ellas, por lo cual se requirió la elaboración de 2 bases de datos similares que se ponen a disposición en la bibliografía.

3.4.1 Análisis estadístico

Para el análisis estadístico se busca realizar inicialmente estadística descriptiva; Para cada variable se calculó el promedio anual, la desviación estándar, los histogramas de frecuencia, el mínimo y el máximo. Este análisis de estadísticas descriptivas se realizó en Microsoft Excel y se organizó de tal manera que fuese fácilmente legible e interpretable.

Tras esto, se realizaron series de tiempo para cada una de las variables estudiadas con el fin de saber cuál era su comportamiento en el tiempo a través de los años estudiados.

Lo primero a realizarse es una tabla de estadísticas básicas que contiene el número de datos y el mínimo y el máximo que se pueden hallar entre los datos estudiados.

Tras esto, se realiza una matriz de correlaciones. Esta matriz muestra cuales son las relaciones más notables entre las variables estudiadas, siendo que entre más cerca estén a 1.0 es que están directamente correlacionadas y las que están más cerca a -1.0 las mas inversamente correlacionadas. Cualquiera de estas dos relaciones puede ser estudiada, dependiendo del interés del investigador.

Luego, se realiza una tabla de valores test. Las tablas de valores test representan valores para saber cuáles variables pueden llegar a retener más información pareadamente sin perder mucha de esta en el proceso. Esta información es

retenida por cada uno de los componentes, determinando el número por el porcentaje de información retenida en cada uno. Esto con el fin de iniciar el proceso de reducción de variables y llevarlo a un punto donde la información que se gana compense la que se pierde en el proceso.

Se toman las dos matrices anteriores y a partir de ellas se determinan una serie de valores propios para cada una que determinan el número de componentes a elegir que finalmente son los que dan pie para la decisión que se debe tomar sobre el número de componentes a elegir para realizar el estudio.

Ya lo con la información de los componentes principales, el siguiente paso es definir la ubicación de estos componentes en el plano en grupos, y para eso se utilizan los intervalos de Anderson. A través de éste análisis se calculan una serie de valores propios que se utilizan como el centroide del grupo y luego se calculan los límites superiores e inferiores para saber la extensión de cada uno de los grupos. Esto es representado finalmente en una gráfica de distribución de valores que funciona como referente para la ubicación de cada uno de los datos contenidos en los componentes principales estudiados.

Tras esto, se realiza una matriz de coordenadas la cual contiene cada uno de los datos a analizar en los componentes principales tomando como componentes del eje x los valores de los componentes principales y como componentes del eje y los valores de las variables estudiadas, resultando una tabla con tres componentes: Las coordenadas de cada uno de de los puntos sobre los componentes principales, las correlaciones entre cada uno de los factores y los valores estudiados y por último los ejes unitarios donde se localizan los puntos.

Para lograr comparar los países en el gráfico, se realizó una tabla de contribuciones, cosenos y coordenadas para cada uno de los datos recolectados por países. En ella pueden apreciarse las contribuciones de cada uno de los países a cada uno de los componentes principales, las coordenadas de ubicación de cada uno de los países y los cosenos de cada uno de los países, de los cuales son relevantes aquellos con un coeficiente igual o mayor a 2.

Para poder interpretar estos datos correctamente, las ubicaciones de cada uno de los puntos y cada uno de los países se grafican en una gráfica bidimensional xy donde se grafica en centroide de donde parten los componentes principales (generalmente cero) y las variables estudiadas. Estas relaciones se pueden ver en la gráfica de la siguiente manera: los datos y/o países que estén dentro de los ejes del vector de la variable resultante de los componentes principales son los que

poseen relevancia, siendo la distancia la relación de visitas que la variable o el país posee.

4. RESULTADOS

4.1 Análisis de estadísticas básicas

Tabla 2: Resultados descriptivos totales

VARIABLES	media	error est	Moda	medi ana	1 Quar	3 Quar	Variance	Stand ar d Dev	Kurtosis	Skew	Range	Min	Max	Sum
Total de vi sitas	44586,95	14554,22	44	687,5	145,25	9617	17793345802,77	133391,70	10,11	3,33	666191	25	666216	3745304
Paginas impresas	104839,56	34574,27	86	1526	357,5	21492	100411948310,80	316878,44	10,29	3,35	1602865	63	1602928	8806523
clicks	274,88	98,17	1	3	1	45,25	809632,95	899,79	15,18	3,92	4811	0	4811	23090
ingreso netos	16,06	5,62	0,01	0,2	0,0375	4	2658,30	51,55	19,04	4,179	326,42	0,01	326,43	1349,25
Promedio impresiones pagina	1,584	0,048	NA	1,55	1,41	1,67	0,19	0,44	54,894	6,68	4,23	1	5,237	133,10
Duracion media visita en seg	91,64	6,18	NA	86,24	58,67	116,89	3211,34	56,66	17,65	3,03	445,66	0	445,667	7698,29

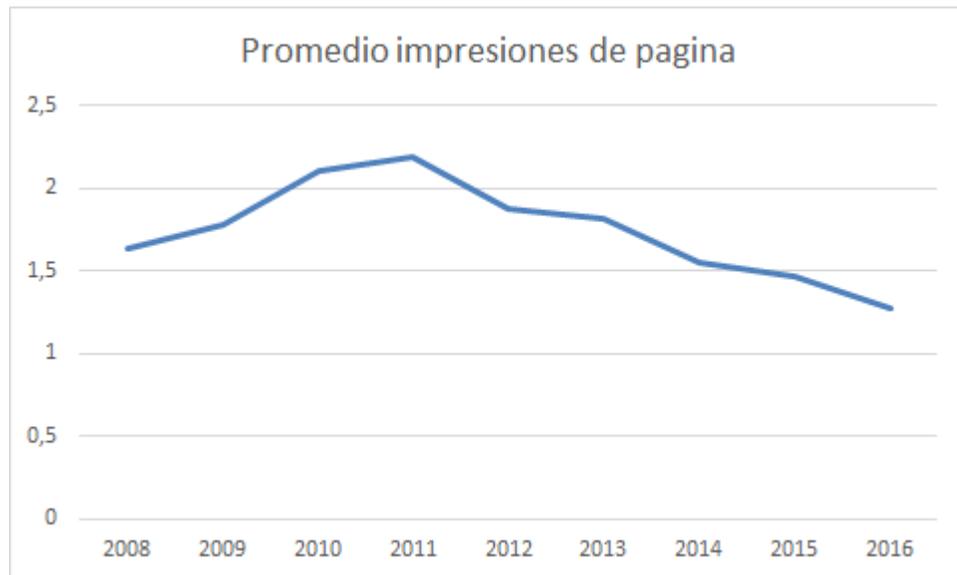
Fuente: Cálculos realizados a través de las bases de datos Excel

Para los resultados obtenidos el promedio es el valor más representativo para cada unidad de medida. Este muestra la tendencia central de los resultados de los 84 países estudiados de 2008 a 2016. A su vez, La desviación estándar muestra la variabilidad esperada. El error estándar también coincide con lo esperado dado que aumenta al aumentar la varianza. el límite inferior de los datos es de 0 y el límite superior es de 1602928. la media de las ganancias es de 16,06 USD y la

media de las visitas únicas es de 44586,95 pero debe tenerse en cuenta la alta varianza de esta variable.

El análisis descriptivo que se llevó a cabo consta, esencialmente, en la exposición de gráficos de histograma conteniendo frecuencias acumuladas (para los dólares, las impresiones de página y el total de páginas y clicks) y promedios (para los ingresos netos y el promedio de impresiones de página por visita) en función de los datos para los 85 países estudiados desde 2008 a 2016.

Fig. 1 Promedio de impresiones de página por visita única - Cambios anuales



Fuente: Análisis de estadística descriptiva realizado a los datos extraídos y compilados por año desde Google Analytics.

La figura 1 muestra que el promedio de impresiones de página por visita se ha mantenido casi constante, alrededor de 2 impresiones por visita, especialmente en 2010 y 2011. Por cada visita, en promedio, se visitaron alrededor de dos páginas en el blog. Este número tiende a disminuir con el transcurrir de los años, llegando 1,5 en 2015 y a menos de esta cifra en 2016.

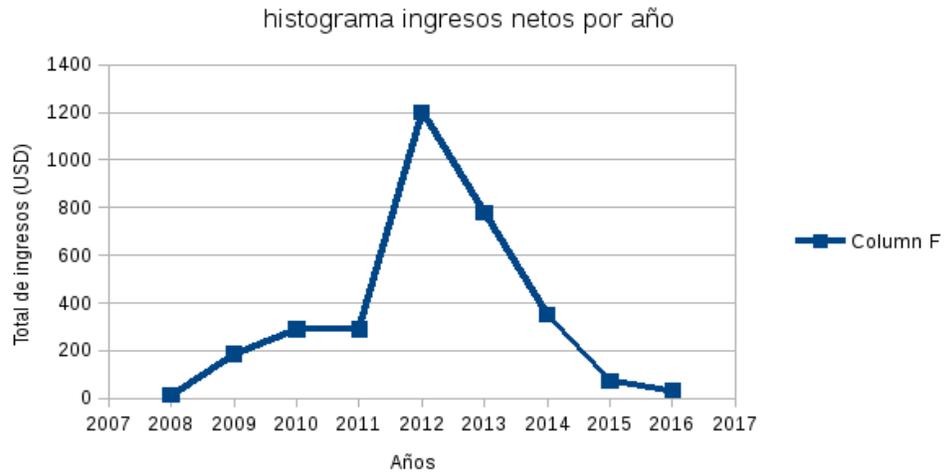
Fig. 2 Duración media de la visita en segundos para cada año



Fuente:Formato de recolección de datos.

En la figura 2, se muestra como el tiempo de duración promedio de la visita ha ido disminuyendo anualmente. Se observa que entre 2009 y 2012 bajo la duración media de la visita en 50 segundos y otros 50 segundos en el recorrido hasta 2016. Esto quiere decir que las visitas en el blog tienden a durar cada vez menos en función del tiempo transcurrido, según dicta el comportamiento de los usuarios del blog. Esto puede ocasionarse por múltiples razones, como el fraude del fin del mundo del 2012 y porque el blog no debe leerse en un orden específico, por lo cual cualquiera entra en cualquier momento a cualquier publicación y asimismo puede salir inmediatamente de ella. Posiblemente la creencia en fenómenos paranormales ha disminuído en el público estudiado.

Fig. 3 Ingreso neto promedio por año



Fuente: Formato de recolección de datos.

El ingreso neto del blog, ha tenido una oscilación importante en los años analizados, El paralelo entre 2010 y 2011 se debe a que hubo ocho meses en que se mantuvo el ingreso neto en 0,8 USC. El incremento de los ingresos entre 2011 y 2012 fue del 3104% y entre 2009 y 2012 fue del 3610 %. El incremento de los ingresos entre 2011 y 2012 es 32 veces mayor que el incremento entre 2009 y 2010. Asimismo el 2012 es el año con mayores ganancias. Las mismas van disminuyendo gradualmente en los años posteriores a 2012 hasta llegar a su punto más bajo después de 2008 en 2015. A primera vista se ve que los ingresos netos del blog por mes han tendido a aumentar entre el 2009 a 2012, siendo que en el año 2010 el porcentaje del aumento de los ingresos netos respecto del año anterior fue de 267%, mientras que en 2011 disminuyó esta cifra un en -64,4%, además de que en 2012 el incremento porcentual fue muy alto, de un 3125%.

Específicamente las ganancias se comportaron de este modo en este periodo por el fenómeno del fin del mundo del 2012. Esto hizo que el tema del calendario

maya se hiciera viral, lo que causó visitas masivas al blog a páginas que hablaban sobre este tema en específico. Tras esto, las personas simplemente perdieron interés y fueron visitando cada vez menos el blog, de tal manera que las ganancias también se vieron afectadas.

Fig. 4 Total de visitas únicas recibidas – distribución anual



Fuente: Formato de recolección de datos.

Respecto al total de visitas únicas recibidas por año, estas aumentaron en función del tiempo como se ve en la figura 4. Entre 2009 y 2012 el total de visitas únicas tuvo un incremento del 2066 % lo que equivale a 48498. En el último año (2012) el aumento respecto al año anterior fue de 30673 visitas, lo que equivale a un 152%, o sea, de 2011 a 2012 el total de visitas únicas recibidas aumentó 2,5 veces. Tras esta fecha y un 2013 que se mantiene constante, el total de visitas unicas va disminuyendo hasta llegar a su punto más bajo en 2016.

Fig. 5 Total de páginas impresas por año



Fuente: Formato de recolección de datos.

Respecto del total de páginas impresas por visita para cada año, se observa como este ha aumentado según se ve en la gráfica 5. Entre 2009 y 2012, el incremento fue de 89544 páginas impresas, lo que equivale a un aumento del 2193%. Entre 2011 y 2012 las páginas impresas aumentaron en 49960 impresiones de página lo que equivale a un aumento del 114%. Esto quiere decir que la cantidad de impresiones de página se duplicó un total de 2.14 veces entre esos años. A partir de 2012 las páginas vistas comienzan a disminuir gradualmente hasta 2016.

4.2 Análisis de componentes principales

Para la realización del ACP lo primero que se hizo fue tomar las 6 variables estudiadas y obtener el número de datos de cada variable, el promedio y los valores mínimos y máximos. A continuación se muestra la tabla con estos valores, notándose que el promedio más alto y los mínimos y máximos más extremos son los de la variable total de pagina impresas con 104839.55 (promedio) y 63;1602928 (mínimo y máximo).

Tabla 3: Estadísticas básicas del ACP con las 6 variables escogidas

NUM . IDEN - LIBELLE	EFFECTIF	POIDS	MOYENNE	ECART-TYPE	MINIMUM	MAXIMUM
1 . TOTV - TOTVISIT	84	84.00	44586.94	132595.33	25.00	666216.00
2 . PAGI - PAGIMPRE	84	84.00	104839.55	314986.63	63.00	1602928.00
3 . CLIC - CLICKS	84	84.00	274.88	894.42	0.00	4811.00
4 . ING_ - ING_NETO	84	84.00	16.06	51.25	0.01	326.43
5 . PROI - PROIMPA	84	84.00	1.58	0.44	1.00	5.24
6 . DURA - DURAMEDV	84	84.00	91.65	56.33	0.00	445.67

Fuente: Análisis ACP realizado en SPAD 5.5

A continuación se realiza una matriz de correlación entre las variables estudiadas con el fin de saber cuales son las asociaciones más robustas entre ellas, donde destaca una fuerte correlación (0.95) entre la variable clicks totales y el total de visitas recibidas. También se puede notar una relación de la misma intensidad entre los clicks totales y los ingresos netos totales del blog. De igual manera, encontramos una fuerte relación (0.93) entre los clicks totales y el total de páginas vistas en el blog.

También se ve que existe una relación un poco más baja (0.88) entre los ingresos netos totales del blog y el total de visitas recibidas, así como en el total de ingresos netos del blog y el total de páginas impresas (en el cual es de 0.83). Por último, se nota una relación de 0.86 entre el promedio de impresiones de página por visita y la duración media de la visita en segundos.

tabla 4: Matriz de correlaciones para las 6 variables escogidas

	TOTV	PAGI	CLIC	ING_	PROI	DURA
TOTV	1.00					
PAGI	1.00	1.00				
CLIC	0.95	0.93	1.00			
ING_	0.88	0.84	0.95	1.00		
PROI	0.10	0.09	0.10	0.09	1.00	
DURA	0.20	0.19	0.18	0.16	0.86	1.00

Fuente: Análisis ACP realizado en SPAD 5.5

Tras esto, se procedió a la realización de la tabla de valores test, la cual representa valores para saber cuáles variables retienen más información y cuáles de ellas pueden llegar a retener más información pareadamente sin perder mucha de esta en el proceso. Este paso determina la cantidad de variables a utilizar en el ACP. Se observan números muy grandes entre las variables total de visitas y total de páginas vistas (29.65) seguida por un número mucho más bajo entre el total de clicks y el total de visitas recibidas(16.82). También se ve una relación de esta magnitud entre los ingresos netos y el total de clicks (16.65). Seguido se encuentra la relación entre el total de clics y el total de páginas vistas (15.19).

tabla 5: Matriz de valores test

	TOTV	PAGI	CLIC	ING_	PROI	DURA
TOTV	99.99					
PAGI	29.65	99.99				
CLIC	16.82	15.19	99.99			
ING_	12.45	11.33	16.65	99.99		
PROI	0.90	0.87	0.93	0.84	99.99	
DURA	1.83	1.80	1.71	1.44	11.88	99.99

Fuente: Análisis ACP realizado en SPAD 5.5

A partir de las dos matrices anteriores se obtienen una serie de valores propios que determinan el número de componentes a elegir. En este caso, los componentes resultantes del acp fueron 6, determinado por los valores propios más altos y los porcentajes que más retuviesen información. En el histograma de la gráfica 4 se puede ver como es el primer componente resultante el que contiene más información con un porcentaje de 64,27% de la info total. Seguido el segundo componente retiene un 29,72% de información. De aquí en adelante el porcentaje de información retenida disminuye drásticamente para los componentes 3,4, 5 y 6.

tabla 6: histograma de valores propios y porcentajes retenidos por los componentes principales

NUMERO	VALEUR PROPRE	POURCENTAGE	POURCENTAGE CUMULE	
1	3.8562	64.27	64.27	*****
2	1.7832	29.72	93.99	*****
3	0.2010	3.35	97.34	*****
4	0.1300	2.17	99.51	***
5	0.0287	0.48	99.98	*
6	0.0010	0.02	100.00	*

Fuente: Análisis ACP realizado en SPAD 5.5

Seguido a esto se realizó el análisis de intervalos de confianza de Anderson para determinar el número de grupos en el cual se iban a distribuir los datos de los componentes principales. Para ello se realizó la tabla 5 , la cual contiene los límites superiores e inferiores y los valores propios (centroides) de cada uno. Además se realiza la gráfica 1, la cual da cuenta de la distribución de los intervalos en forma gráfica.

Tabla 7: Intervalos de confianza de anderson

NUMERO	BORNE INFERIEURE	VALEUR PROPRE	BORNE SUPERIEURE
1	2.6830	3.8562	5.0295
2	1.2406	1.7832	2.3257
3	0.1398	0.2010	0.2621
4	0.0904	0.1300	0.1695
5	0.0200	0.0287	0.0374

Fuente: Análisis ACP realizado en SPAD 5.5

gráfica 1: posición relativa de los intervalos



Fuente: Análisis ACP realizado en SPAD 5.5

A continuación se realizó la matriz de relación de coordenadas según la posición de los intervalos dada anteriormente. Estas matrices contienen además de las mismas coordenadas, las las correlaciones entre factores y variables y los ejes escogidos. Esta tabla puede apreciarse mucho más fácil en la gráfica 2.

Tabla 8: Variables con coordenadas y ejes unitarios (1 a 5).

VARIABLES	COORDONNEES					CORRELATIONS VARIABLE-FACTEUR					ANCIENS AXES UNITAIRES				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
IDEN - LIBELLE COURT															
TOTV - TOTVISIT	-0.98	0.12	0.17	0.04	0.03	-0.98	0.12	0.17	0.04	0.03	-0.50	0.09	0.37	0.12	0.15
PAGI - PAGIMPRE	-0.96	0.12	0.23	0.06	0.04	-0.96	0.12	0.23	0.06	0.04	-0.49	0.09	0.50	0.16	0.22
CLIC - CLICKS	-0.98	0.13	-0.09	-0.02	-0.14	-0.98	0.13	-0.09	-0.02	-0.14	-0.50	0.10	-0.21	-0.05	-0.83
ING_ - ING_NETO	-0.93	0.14	-0.32	-0.06	0.08	-0.93	0.14	-0.32	-0.06	0.08	-0.48	0.10	-0.70	-0.17	0.49
PROI - PROIMPA	-0.22	-0.94	-0.09	0.24	0.00	-0.22	-0.94	-0.09	0.24	0.00	-0.11	-0.70	-0.19	0.67	-0.01
DURA - DURAMEDV	-0.31	-0.91	0.08	-0.25	0.00	-0.31	-0.91	0.08	-0.25	0.00	-0.16	-0.68	0.18	-0.69	0.01

Fuente: Análisis ACP realizado en SPAD 5.5

Tabla 9: Variables con coordenadas y ejes unitarios (6).

VARIABLES	COORDONNEES					CORRELATIONS VARIABLE-FACTEUR					ANCIENS AXES UNITAIRES				
	6	0	0	0	0	6	0	0	0	0	6	0	0	0	0
TOTV - TOTVISIT	-0.02	0.00	0.00	0.00	0.00	-0.02	0.00	0.00	0.00	0.00	-0.76	0.00	0.00	0.00	0.00
PAGI - PAGIMPRE	0.02	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.65	0.00	0.00	0.00	0.00
CLIC - CLICKS	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00
ING_ - ING_NETO	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00
PROI - PROIMPA	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
DURA - DURAMEDV	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00

Fuente: Análisis ACP realizado en SPAD 5.5

Finalmente se realiza la tabla de coordenadas, contribuciones y cosenos de cada uno de los 84 individuos a graficar en cada uno de los 6 intervalos escogidos. estos datos pueden verse de forma completa en las tablas 8 y 9. Aquí vemos que las contribuciones de los primeros 4 países son muy altas para el primer componente, mayor a 2 en España, Colombia, México y Argentina, lo que indica que son importantes y deben ser analizados cuidadosamente en la gráfica 2. para el segundo componente la mayor contribución la realiza corea del sur por más de 50 puntos delante del que le sigue. Para el tercer componente, EEUU, Perú y venezuela fueron quienes más aportaron .Para el cuarto componente corea del sur, luxemburgo y Qatar fueron los países que más aportaron a esta variable. Para el quinto componente, fueron Colombia México y Perú los que más aportaron. Finalmente, para el sexto componente, fueron los países España, Colombia, México, Perú y Venezuela quienes más aportaron a este componente.

Tabla 10: Tabla de coordenadas, contribuciones y cosenos de los 84 puntos(1 a 5).

INDIVIDUS			COORDONNEES					CONTRIBUTIONS					COSINUS CARRES				
IDENTIFICATEUR	P.REL	DISTO	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
ESPANA	1.19	78.71	-8.41	1.22	-2.53	-0.31	0.21	21.8	1.0	37.8	0.9	1.9	0.90	0.02	0.08	0.00	0.00
COL	1.19	86.17	-9.18	1.22	0.38	0.13	-0.56	26.0	1.0	0.9	0.1	13.2	0.98	0.02	0.00	0.00	0.00
MEX	1.19	45.89	-6.76	0.03	0.13	-0.16	-0.38	14.1	0.0	0.1	0.2	5.9	1.00	0.00	0.00	0.00	0.00
ARGEN	1.19	41.61	-6.43	0.25	0.31	0.11	-0.18	12.8	0.0	0.6	0.1	1.4	1.00	0.00	0.00	0.00	0.00
EEUU	1.19	5.58	-1.44	0.56	-1.49	-0.33	0.92	0.6	0.2	13.1	1.0	35.1	0.37	0.06	0.40	0.02	0.15
CHI	1.19	9.56	-3.08	-0.19	0.11	-0.20	0.08	2.9	0.0	0.1	0.4	0.3	0.99	0.00	0.00	0.00	0.00
PER	1.19	23.39	-4.23	0.39	2.04	0.46	0.97	5.5	0.1	24.7	2.0	39.4	0.77	0.01	0.18	0.01	0.04
VEN	1.19	22.82	-4.40	0.56	1.71	0.41	0.12	6.0	0.2	17.2	1.5	0.6	0.85	0.01	0.13	0.01	0.00
ECU	1.19	0.15	-0.28	-0.13	0.20	-0.12	0.02	0.0	0.0	0.2	0.1	0.0	0.53	-0.11	0.25	0.09	0.00
CAN	1.19	0.15	0.38	-0.05	-0.06	0.01	0.03	0.0	0.0	0.0	0.0	0.0	0.95	0.02	0.03	0.00	0.01
URU	1.19	0.04	0.15	-0.09	0.08	-0.04	0.01	0.0	0.0	0.0	0.0	0.0	0.59	0.21	0.16	0.03	0.00
BRA	1.19	0.21	0.26	-0.35	0.04	-0.13	0.04	0.0	0.1	0.0	0.2	0.1	0.32	0.58	0.01	0.08	0.01
UK	1.19	0.59	0.59	0.31	-0.18	0.34	0.05	0.1	0.1	0.2	1.0	0.1	0.59	0.16	0.05	0.19	0.00
ALE	1.19	0.51	0.58	0.25	-0.14	0.30	0.04	0.1	0.0	0.1	0.8	0.1	0.66	0.12	0.04	0.18	0.00
REPDOC	1.19	0.21	0.41	0.18	0.03	-0.07	-0.04	0.1	0.0	0.0	0.0	0.1	0.80	0.17	0.01	0.02	0.01
GUAT	1.19	0.25	0.26	-0.34	0.08	-0.25	-0.07	0.0	0.1	0.0	0.6	0.2	0.26	0.45	0.02	0.24	0.02
COSRIC	1.19	0.41	0.29	-0.53	0.02	-0.19	-0.06	0.0	0.2	0.0	0.3	0.1	0.21	0.70	0.00	0.09	0.01
PAN	1.19	0.34	0.38	-0.39	0.05	-0.20	0.01	0.0	0.1	0.0	0.4	0.0	0.43	0.44	0.01	0.12	0.00
FRANC	1.19	0.47	0.58	0.30	-0.06	0.19	0.02	0.1	0.1	0.0	0.3	0.0	0.72	0.20	0.01	0.08	0.00
BOL	1.19	1.39	0.09	-1.07	0.19	-0.45	-0.01	0.0	0.8	0.2	1.8	0.0	0.01	0.82	0.03	0.15	0.00
PUERRIC	1.19	0.46	0.35	-0.56	0.06	-0.16	0.03	0.0	0.2	0.0	0.2	0.0	0.26	0.67	0.01	0.06	0.00
SALVADOR	1.19	0.68	0.25	-0.71	0.10	-0.31	-0.05	0.0	0.3	0.1	0.9	0.1	0.09	0.75	0.02	0.14	0.00
NICA	1.19	0.39	0.40	-0.46	-0.02	-0.06	-0.07	0.1	0.1	0.0	0.0	0.2	0.42	0.56	0.00	0.01	0.01
HONDU	1.19	0.28	0.45	-0.20	0.05	-0.18	-0.02	0.1	0.0	0.0	0.3	0.0	0.74	0.14	0.01	0.11	0.00
ITA	1.19	0.50	0.48	-0.47	0.04	-0.20	-0.01	0.1	0.1	0.0	0.4	0.0	0.47	0.44	0.00	0.08	0.00
PAR	1.19	0.75	0.39	-0.73	0.07	-0.24	-0.02	0.0	0.4	0.0	0.5	0.0	0.20	0.72	0.01	0.08	0.00
POR	1.19	1.78	0.33	-1.29	0.00	-0.08	0.00	0.0	1.1	0.0	0.1	0.0	0.06	0.93	0.00	0.00	0.00
SUIZA	1.19	1.56	0.41	-1.14	-0.12	0.28	0.00	0.1	0.9	0.1	0.7	0.0	0.11	0.84	0.01	0.05	0.00
AUSTRA	1.19	0.70	0.50	-0.63	0.03	-0.25	0.00	0.1	0.3	0.0	0.6	0.0	0.35	0.56	0.00	0.09	0.00
PAIBAJOS	1.19	0.41	0.64	0.08	-0.03	0.01	-0.01	0.1	0.0	0.0	0.0	0.0	0.98	0.01	0.00	0.00	0.00
SUE	1.19	0.43	0.62	-0.10	-0.08	0.18	-0.01	0.1	0.0	0.0	0.3	0.0	0.89	0.02	0.01	0.08	0.00
SUDA	1.19	0.75	0.73	0.46	-0.01	-0.03	-0.01	0.2	0.1	0.0	0.0	0.0	0.72	0.28	0.00	0.00	0.00
ANDORRA	1.19	0.44	0.61	-0.18	-0.08	0.18	-0.02	0.1	0.0	0.0	0.3	0.0	0.84	0.07	0.01	0.07	0.00
NOR	1.19	0.65	0.54	-0.45	0.08	-0.39	0.00	0.1	0.1	0.0	1.4	0.0	0.45	0.31	0.01	0.23	0.00
BUL	1.19	2.08	0.87	1.02	-0.15	0.51	-0.03	0.2	0.7	0.1	2.4	0.0	0.36	0.50	0.01	0.12	0.00
TURQ	1.19	0.60	0.64	-0.11	-0.13	0.41	-0.02	0.1	0.0	0.1	1.5	0.0	0.68	0.02	0.03	0.27	0.00
RUS	1.19	2.96	0.93	1.42	-0.07	0.28	-0.02	0.3	1.3	0.0	0.7	0.0	0.29	0.68	0.00	0.03	0.00
BELG	1.19	0.66	0.71	0.34	-0.06	0.18	-0.02	0.2	0.1	0.0	0.3	0.0	0.77	0.18	0.00	0.05	0.00
EMIRAT_ARAB	1.19	1.14	0.79	0.62	-0.11	0.36	-0.03	0.2	0.3	0.1	1.2	0.0	0.54	0.34	0.01	0.11	0.00
MARR	1.19	0.88	0.75	0.55	-0.02	0.06	-0.03	0.2	0.2	0.0	0.0	0.0	0.65	0.34	0.00	0.00	0.00
INDIA	1.19	0.89	0.50	-0.79	-0.06	0.13	-0.02	0.1	0.4	0.0	0.1	0.0	0.28	0.70	0.00	0.02	0.00
RUMA	1.19	0.54	0.70	0.25	-0.01	0.00	-0.02	0.1	0.0	0.0	0.0	0.0	0.89	0.11	0.00	0.00	0.00
POL	1.19	1.42	0.82	0.85	-0.05	0.18	-0.02	0.2	0.5	0.0	0.3	0.0	0.47	0.50	0.00	0.02	0.00
AUSTRI	1.19	0.50	0.69	0.17	-0.03	0.06	-0.02	0.1	0.0	0.0	0.0	0.0	0.94	0.05	0.00	0.01	0.00
PAK	1.19	0.49	0.66	0.09	0.04	-0.19	-0.02	0.1	0.0	0.0	0.3	0.0	0.90	0.02	0.00	0.07	0.00
IRL	1.19	1.49	0.83	0.83	-0.10	0.34	-0.03	0.2	0.5	0.1	1.1	0.0	0.46	0.46	0.01	0.08	0.00
JAP	1.19	0.52	0.56	-0.40	0.04	-0.20	-0.01	0.1	0.1	0.0	0.4	0.0	0.61	0.32	0.00	0.07	0.00
MACEDO	1.19	1.09	0.79	0.68	-0.02	0.06	-0.02	0.2	0.3	0.0	0.0	0.0	0.57	0.42	0.00	0.00	0.00
DINAM	1.19	0.48	0.67	0.10	0.03	-0.14	-0.02	0.1	0.0	0.0	0.2	0.0	0.94	0.02	0.00	0.04	0.00
FILIP	1.19	1.88	0.86	1.00	-0.10	0.36	-0.03	0.2	0.7	0.1	1.2	0.0	0.40	0.53	0.01	0.07	0.00
ARAB_SAUD	1.19	7.00	0.19	-2.08	0.39	-1.58	0.01	0.0	2.9	0.9	22.9	0.0	0.01	0.62	0.02	0.36	0.00
CURA	1.19	0.88	0.65	0.09	0.17	-0.65	-0.01	0.1	0.0	0.2	3.9	0.0	0.47	0.01	0.03	0.48	0.00
UCRA	1.19	1.00	0.78	0.60	-0.06	0.18	-0.03	0.2	0.2	0.0	0.3	0.0	0.61	0.36	0.00	0.03	0.00
ESLOV	1.19	1.23	0.80	0.76	0.00	-0.02	-0.02	0.2	0.4	0.0	0.0	0.0	0.53	0.47	0.00	0.00	0.00
IRAK	1.19	2.36	0.90	1.24	-0.02	0.09	-0.03	0.3	1.0	0.0	0.1	0.0	0.35	0.65	0.00	0.00	0.00
CHECA	1.19	0.78	0.75	0.46	-0.04	0.12	-0.02	0.2	0.1	0.0	0.1	0.0	0.71	0.27	0.00	0.02	0.00
KIRGUIS	1.19	1.65	0.84	0.96	0.03	-0.13	-0.02	0.2	0.6	0.0	0.1	0.0	0.43	0.56	0.00	0.01	0.00
INDO	1.19	2.35	0.90	1.22	-0.05	0.20	-0.03	0.3	1.0	0.0	0.4	0.0	0.35	0.64	0.00	0.02	0.00
NUEVZEL	1.19	0.42	0.64	-0.07	0.00	-0.06	-0.02	0.1	0.0	0.0	0.0	0.0	0.98	0.01	0.00	0.01	0.00
FINL	1.19	0.50	0.69	0.15	-0.02	0.03	-0.02	0.1	0.0	0.0	0.0	0.0	0.95	0.05	0.00	0.00	0.00
CUBA	1.19	0.91	0.52	-0.52	0.15	-0.59	-0.01	0.1	0.2	0.1	3.2	0.0	0.30	0.29	0.03	0.39	0.00
COR_SUR	1.19	107.57	-1.32	-10.21	-0.46	1.19	0.00	0.5	69.6	1.3	13.0	0.0	0.02	0.97	0.00	0.01	0.00
BOSNIA	1.19	0.90	0.50	-0.69	0.09	-0.41	-0.01	0.1	0.3	0.0	1.5	0.0	0.28	0.52	0.01	0.19	0.00
SING	1.19	0.80	0.52	-0.70	-0.07	0.17	-0.02	0.1	0.3	0.0	0.3	0.0	0.34	0.61	0.01	0.04	0.00
SERB	1.19	2.02	0.88	1.07	-0.09	0.31	-0.03	0.2	0.8	0.0	0.9	0.0	0.38	0.57	0.00	0.05	0.00
UKNOWN	1.19	0.50	0.55	-0.41	0.05	-0.17	-0.01	0.1	0.1	0.0	0.3	0.0	0.61	0.33	0.00	0.06	0.00
GEOR	1.19	2.09	0.89	1.12	-0.06	0.23	-0.03	0.2	0.8	0.0	0.5	0.0	0.38	0.60	0.00	0.03	0.00
ANGO	1.19	1.86	0.42	-1.09	0.16	-0.68	-0.01	0.1	0.8	0.2	4.3	0.0	0.09	0.64	0.01	0.25	0.00
ISRA	1.19	0.51	0.66	0.09	0.06	-0.24	-0.02	0.1	0.0	0.0	0.5	0.0	0.86	0.02	0.01	0.11	0.00
CROAC	1.19	0.88	0.76	0.48	-0.08	0.25	-0.03	0.2	0.2	0.0	0.6	0.0	0.66	0.26	0.01	0.07	0.00
ARUB	1.19	2.51	0.92	1.27	-0.06	0.21	-0.03	0.3	1.1	0.0	0.4	0.0	0.33	0.65	0.00	0.02	0.00
OMAN	1.19	2.11	0.88	1.07	-0.11	0.40	-0.03	0.2	0.8	0.1	1.5	0.0	0.37	0.55	0.01	0.08	0.00
NIGER	1.19	4.79	1.04	1.91	-0.05	0.21	-0.03	0.3	2.4	0.0	0.4	0.0	0.23	0.76	0.00	0.01	0.00
MAL	1.19	0.62	0.72	0.30	-0.03	0.08	-0.02	0.2	0.1	0.0	0.1	0.0	0.84	0.15	0.00	0.01	0.00
LTBAN	1.19	0.82	0.74	0.47	0.05	-0.22	-0.02	0.2	0.1	0.0	0.4	0.0	0.67	0.27	0.00	0.06	0.00
LUX	1.19	1.86	0.42	-1.08	0.17	-0.70	0.00	0.1	0.8	0.2	4.5	0.0	0.09	0.63			

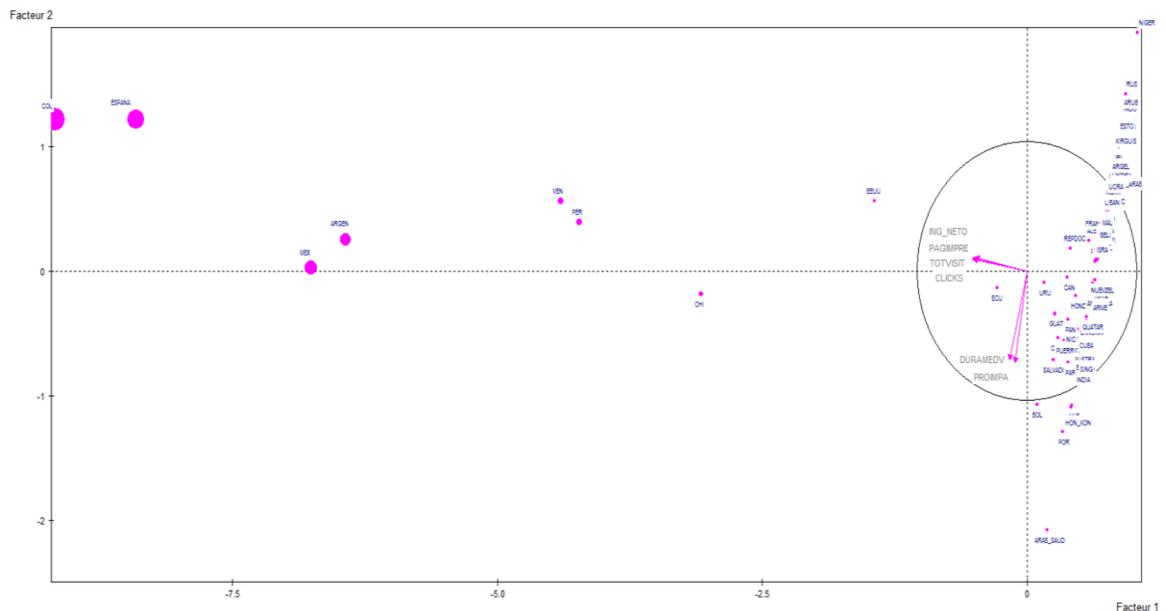
INDIVIDUS			COORDONNEES					CONTRIBUTIONS					COSINUS CARRES				
IDENTIFICATEUR	P.REL	DISTO	6	0	0	0	0	6	0	0	0	0	6	0	0	0	0
ESPANA	1.19	78.71	-0.03	0.00	0.00	0.00	0.00	1.3	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
COL	1.19	86.17	0.08	0.00	0.00	0.00	0.00	7.2	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
MEX	1.19	45.89	-0.15	0.00	0.00	0.00	0.00	27.9	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
ARGEN	1.19	41.61	0.04	0.00	0.00	0.00	0.00	1.6	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
EEUU	1.19	5.58	0.10	0.00	0.00	0.00	0.00	11.1	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
CHI	1.19	9.56	-0.01	0.00	0.00	0.00	0.00	0.1	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
PER	1.19	23.39	-0.11	0.00	0.00	0.00	0.00	15.1	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
VEN	1.19	22.82	0.16	0.00	0.00	0.00	0.00	30.9	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
ECU	1.19	0.15	-0.05	0.00	0.00	0.00	0.00	2.5	0.0	0.0	0.0	0.0	0.01	0.00	0.00	0.00	0.00
CAN	1.19	0.15	0.01	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
URU	1.19	0.04	-0.01	0.00	0.00	0.00	0.00	0.2	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
BRA	1.19	0.21	-0.01	0.00	0.00	0.00	0.00	0.1	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
UK	1.19	0.59	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
ALE	1.19	0.51	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
REPOC	1.19	0.21	-0.01	0.00	0.00	0.00	0.00	0.1	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
GUAT	1.19	0.25	-0.01	0.00	0.00	0.00	0.00	0.2	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
COSRIC	1.19	0.41	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
PAN	1.19	0.34	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
FRANC	1.19	0.47	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
BOL	1.19	1.39	-0.02	0.00	0.00	0.00	0.00	0.3	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
PUERRIC	1.19	0.46	-0.01	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
SALVADOR	1.19	0.68	-0.01	0.00	0.00	0.00	0.00	0.2	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
NICA	1.19	0.39	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
HONDU	1.19	0.28	-0.01	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
ITA	1.19	0.50	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
PAR	1.19	0.75	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
POR	1.19	1.78	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
SUIZA	1.19	1.56	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
AUSTRA	1.19	0.70	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
PAIBAJOS	1.19	0.41	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
SUE	1.19	0.43	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
SUDA	1.19	0.75	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
ANDORRA	1.19	0.44	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
NOR	1.19	0.65	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
BUL	1.19	2.08	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
TURQ	1.19	0.60	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
RUS	1.19	2.96	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
BELG	1.19	0.66	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
EMIRAT_ARAB	1.19	1.14	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
MARR	1.19	0.88	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
INDIA	1.19	0.89	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
RUMA	1.19	0.54	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
POL	1.19	1.42	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
AUSTRI	1.19	0.50	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
PAK	1.19	0.49	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
IRL	1.19	1.49	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
JAP	1.19	0.52	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
MACEDO	1.19	1.09	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
DINAM	1.19	0.48	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
FILIP	1.19	1.88	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
ARAB_SAUD	1.19	7.00	0.02	0.00	0.00	0.00	0.00	0.3	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
CURA	1.19	0.88	0.01	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
UCRA	1.19	1.00	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
ESLOV	1.19	1.23	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
IRAK	1.19	2.36	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
CHECA	1.19	0.78	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
KIRGUI	1.19	1.65	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
INDO	1.19	2.35	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
NUEVZEL	1.19	0.42	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
FINL	1.19	0.50	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
CUBA	1.19	0.91	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
COR_SUR	1.19	107.57	0.01	0.00	0.00	0.00	0.00	0.1	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
BOSNIA	1.19	0.90	0.01	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
SING	1.19	0.80	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
SERB	1.19	2.02	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
UKNOWN	1.19	0.50	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
GEOR	1.19	2.09	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
ANGO	1.19	1.86	0.01	0.00	0.00	0.00	0.00	0.1	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
ISRA	1.19	0.51	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
CROAC	1.19	0.88	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
ARUB	1.19	2.51	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
OMAN	1.19	2.11	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
NIGER	1.19	4.79	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00

MAL	1.19	0.62	0.00	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
LIBAN	1.19	0.82	0.00	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
LUX	1.19	1.86	0.01	0.00	0.00	0.00	0.00	0.00	0.1	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
HUNGR	1.19	1.32	0.00	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
HON_KON	1.19	1.48	0.00	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
GRE	1.19	0.57	0.00	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
ESTO	1.19	2.03	0.00	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
QUATAR	1.19	0.90	0.01	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
BELI	1.19	0.53	0.00	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
ARME	1.19	1.11	0.00	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
ARGEL	1.19	1.29	0.00	0.00	0.00	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00

Fuente: Análisis ACP realizado en SPAD 5.5

A continuación se presenta la gráfica del ACP donde se pueden visualizar los componentes principales y los países analizados. Para poder realizar un análisis más certero se suprimió de la gráfica el país corea del sur ya que esta quedaba muy dispersa y no era fácilmente identificable el número o nombre de las demás variables.

gráfica 2: Análisis de componentes principales para los 84 componentes



Fuente: Análisis ACP realizado en SPAD 5.5

Como actividad complementaria para la realización del primer ACP, se realizaron dos ACP más con el fin de comprobar si la información anteriormente concluida es respaldada por otro tipo de información. El primer ACP tiene que ver con la variable edad y su distribución a través de los países con más visitas y el segundo

tiene que ver con la variable sexo y su distribución a través de los diferentes países que recibieron más visitas en el periodo.

4.2.1) Análisis de componentes principales según la edad.

De nuevo, lo que se hizo para realizar el análisis de componentes principales fué tomar 4 variables: sesiones por visita, nuevos usuarios (Total de visitas) , total de páginas vistas, duración media de la visita y se le agrega la variable edad como variable suplementaria. Además de esto se observan los mínimos, máximos y la media de cada una de las variables para identificar el rango y la amplitud de los datos:

Tabla 11: Estadísticas básicas de las variables con la edad

NUM . IDEN - LIBELLE	EFFECTIF	POIDS	MOYENNE	ECART-TYPE	MINIMUM	MAXIMUM
1 . SESI - SESIONES	66	66.00	904.33	785.97	193.00	4335.00
2 . NUEV - NUEVOSUS	66	66.00	715.64	603.73	158.00	3387.00
3 . PÁG - PÁGINAS	66	66.00	1.53	0.45	1.00	3.76
4 . DURA - DURAMEDS	66	66.00	70.65	53.78	0.00	296.58

Fuente: Análisis ACP realizado en SPAD 5.5

Tras esto se realiza una tabla de correlación para mirar qué asociaciones grandes existen entre las variables estudiadas. Puede observarse cómo existe una gran correlación (0.99) entre las nuevos usuarios y el número de sesiones nuevas. También observamos una correlación un poco menor (0.80) entre el número de páginas vistas en una sesión por usuario y la duración media de la visita de un usuario al blog.

Tabla 12: Tabla de correlación de los componentes principales

	SESI	NUEV	PÁG	DURA
SESI	1.00			
NUEV	0.99	1.00		
PÁG	-0.02	0.01	1.00	
DURA	-0.01	0.01	0.80	1.00

Fuente: Análisis ACP realizado en SPAD 5.5

Luego, se realizó una tabla de valores test, para saber cuál era la paridad de los datos y si entre alguno de esos pares habría una retención de información considerable. De este paso podemos deducir el número de componentes principales a utilizar en el análisis. podemos observar fuertes correlaciones entre las nuevas visitas y las sesiones por usuario (20.09) y entre el número de páginas vistas y la duración media de la visita (8.93). También se observa la mayor correlación negativa (-0.19) entre el número de páginas vistas y el número de sesiones totales.

Tabla 13: Matriz de valores test

	SESI	NUEV	PÁG	DURA
SESI	99.99			
NUEV	20.09	99.99		
PÁG	-0.19	0.08	99.99	
DURA	-0.07	0.05	8.93	99.99

Fuente: Análisis ACP realizado en SPAD 5.5

A partir de lo anterior se obtienen valores que determinan el número de componentes principales a elegir para la realización del ACP. En este caso las variables fueron recompuestas en 4 componentes principales determinados por la cantidad de información retenida. se puede observar como el primer componente retiene un porcentaje de 49.66% de la información total, mientras que el segundo componente retiene el 45.01% de la información total. El tercer y cuarto componente retienen entre ellos poco menos del 6% de la información total.

Tabla 14: histograma de frecuencias con retención de porcentajes

NUMERO	VALEUR PROPRE	POURCENTAGE	POURCENTAGE CUMULE	
1	1.9862	49.66	49.66	*****
2	1.8003	45.01	94.66	*****
3	0.2002	5.00	99.67	*****
4	0.0133	0.33	100.00	*

Fuente: Análisis ACP realizado en SPAD 5.5

Ya con el histograma y el aporte de cada uno de los componentes, se procede a la realización del análisis de intervalos de confianza de Anderson, con lo cual se determina el número de grupos en los cuales se van a distribuir los componentes

principales. Para ellos se hallan los valores extremos de cada componente y su respectivo centroide, como se muestra en la tabla x:

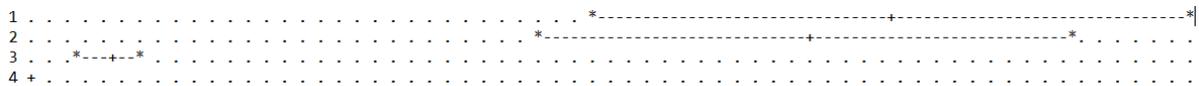
Tabla 15: Intervalos de anderson

NUMERO	BORNE INFERIEURE	VALEUR PROPRE	BORNE SUPERIEURE
1	1.3033	1.9862	2.6691
2	1.1813	1.8003	2.4193
3	0.1313	0.2002	0.2690
4	0.0087	0.0133	0.0179

Fuente: Análisis ACP realizado en SPAD 5.5

Ya que no se visualiza fácilmente la distribución de los grupos en la tabla anterior, se presenta la gráfica x, la cual muestra de manera gráfica la distribución de los intervalos localizados:

Gráfica 16: representación gráfica de los intervalos de Anderson



Fuente: Análisis ACP realizado en SPAD 5.5

Finalmente se realiza la tabla de coordenadas, contribuciones y cosenos de las variables escogidas en los grupos seleccionados. Para ellos se realiza la tabla x, la cual muestra las coordenadas, las correlaciones y los ejes unitarios para cada una de las variables estudiadas.

Tabla 17: coordenadas, correlaciones y ejes unitarios para los 4 componentes principales

VARIABLES	COORDONNEES					CORRELATIONS VARIABLE-FACTEUR					ANCIENS AXES UNITAIRES				
	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0
IDEN - LIBELLE COURT															
SESI - SESIONES	1.00	0.03	-0.02	0.08	0.00	1.00	0.03	-0.02	0.08	0.00	0.71	0.02	-0.03	0.71	0.00
NUEV - NUEVOSUS	1.00	0.05	0.02	-0.08	0.00	1.00	0.05	0.02	-0.08	0.00	0.71	0.04	0.04	-0.71	0.00
PÁG - PÁGINAS	-0.04	0.95	0.32	0.01	0.00	-0.04	0.95	0.32	0.01	0.00	-0.03	0.71	0.71	0.05	0.00
DURA - DURAMEDS	-0.04	0.95	-0.32	0.00	0.00	-0.04	0.95	-0.32	0.00	0.00	-0.03	0.71	-0.71	-0.03	0.00

Fuente: Análisis ACP realizado en SPAD 5.5

tras esto, se realiza la tabla de coordenadas, contribuciones y cosenos para cada uno de los individuos estudiados. Las contribuciones mayores a 2 son las que nos interesan en este caso.

Tabla 17: coordenadas, correlaciones y ejes unitarios para los 4 componentes principales

INDIVIDUS			COORDONNEES					CONTRIBUTIONS					COSINUS CARRES				
IDENTIFICATEUR	P.REL	DISTO	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0
Argentina	1.52	3.93	0.37	1.93	0.18	-0.14	0.00	0.1	3.1	0.2	2.4	0.0	0.03	0.95	0.01	0.01	0.00
Argentina	1.52	1.28	0.08	1.09	0.30	-0.02	0.00	0.0	1.0	0.7	0.0	0.0	0.01	0.92	0.07	0.00	0.00
Argentina	1.52	0.13	-0.26	-0.17	0.19	-0.06	0.00	0.1	0.0	0.3	0.4	0.0	0.50	0.21	0.27	0.03	0.00
Argentina	1.52	11.25	-0.47	3.32	-0.04	-0.02	0.00	0.2	9.3	0.0	0.0	0.0	0.02	0.98	0.00	0.00	0.00
Argentina	1.52	0.87	-0.84	0.27	0.30	0.01	0.00	0.5	0.1	0.7	0.0	0.0	0.81	0.09	0.10	0.00	0.00
Bolivia	1.52	0.67	0.69	-0.44	-0.04	0.03	0.00	0.4	0.2	0.0	0.1	0.0	0.70	0.29	0.00	0.00	0.00
Bolivia	1.52	5.20	-0.40	1.91	-1.18	0.00	0.00	0.1	3.1	10.6	0.0	0.0	0.03	0.70	0.27	0.00	0.00
Bolivia	1.52	1.21	-0.74	-0.78	-0.22	-0.02	0.00	0.4	0.5	0.4	0.1	0.0	0.45	0.51	0.04	0.00	0.00
Bolivia	1.52	2.78	-0.75	-1.48	0.15	-0.04	0.00	0.4	1.8	0.2	0.2	0.0	0.20	0.79	0.01	0.00	0.00
Chile	1.52	1.36	1.13	-0.12	0.27	-0.02	0.00	1.0	0.0	0.5	0.1	0.0	0.94	0.01	0.05	0.00	0.00
Chile	1.52	2.25	0.94	1.08	-0.39	-0.18	0.00	0.7	1.0	1.2	3.9	0.0	0.39	0.52	0.07	0.02	0.00
Chile	1.52	0.86	-0.48	-0.77	0.19	0.05	0.00	0.2	0.5	0.3	0.3	0.0	0.27	0.69	0.04	0.00	0.00
Chile	1.52	0.68	-0.79	0.22	-0.11	0.03	0.00	0.5	0.0	0.1	0.1	0.0	0.91	0.07	0.02	0.00	0.00
Chile	1.52	2.18	-1.18	-0.87	0.19	-0.03	0.00	1.1	0.6	0.3	0.1	0.0	0.64	0.34	0.02	0.00	0.00
Colombia	1.52	38.69	6.22	0.06	-0.13	-0.05	0.00	29.5	0.0	0.1	0.3	0.0	1.00	0.00	0.00	0.00	0.00
Colombia	1.52	10.22	3.19	-0.18	0.06	0.09	0.00	7.8	0.0	0.0	0.8	0.0	1.00	0.00	0.00	0.00	0.00
Colombia	1.52	1.81	1.32	-0.01	-0.24	0.10	0.00	1.3	0.0	0.4	1.2	0.0	0.96	0.00	0.03	0.01	0.00
Colombia	1.52	2.11	-0.31	1.01	-1.00	-0.03	0.00	0.1	0.9	7.5	0.1	0.0	0.05	0.48	0.47	0.00	0.00
Colombia	1.52	0.71	-0.63	-0.31	-0.46	0.06	0.00	0.3	0.1	1.6	0.4	0.0	0.55	0.14	0.30	0.01	0.00
Colombia	1.52	1.04	-0.71	-0.57	-0.45	-0.04	0.00	0.4	0.3	1.6	0.2	0.0	0.49	0.31	0.20	0.00	0.00
Costa Rica	1.52	1.28	-0.75	-0.81	0.02	0.22	0.00	0.4	0.6	0.0	5.7	0.0	0.44	0.52	0.00	0.04	0.00
Costa Rica	1.52	0.87	-0.86	-0.36	0.02	-0.03	0.00	0.6	0.1	0.0	0.1	0.0	0.85	0.15	0.00	0.00	0.00
Costa Rica	1.52	4.74	-1.27	0.87	-1.54	-0.08	0.00	1.2	0.6	18.0	0.7	0.0	0.34	0.16	0.50	0.00	0.00
Dominican Republic	1.52	1.80	-0.61	-1.19	-0.05	-0.11	0.00	0.3	1.2	0.0	1.3	0.0	0.20	0.79	0.00	0.01	0.00
Dominican Republic	1.52	1.57	-0.88	-0.87	0.17	-0.02	0.00	0.6	0.6	0.2	0.0	0.0	0.49	0.49	0.02	0.00	0.00
Dominican Republic	1.52	3.66	-1.18	-1.50	0.03	0.01	0.00	1.1	1.9	0.0	0.0	0.0	0.38	0.62	0.00	0.00	0.00
Ecuador	1.52	3.38	1.81	-0.11	-0.16	-0.23	0.00	2.5	0.0	0.2	5.8	0.0	0.97	0.00	0.01	0.02	0.00
Ecuador	1.52	0.91	0.27	-0.91	-0.08	-0.08	0.00	0.1	0.7	0.1	0.7	0.0	0.08	0.90	0.01	0.01	0.00
Ecuador	1.52	0.46	-0.07	-0.67	-0.01	-0.03	0.00	0.0	0.4	0.0	0.1	0.0	0.01	0.99	0.00	0.00	0.00
Ecuador	1.52	1.90	-0.93	-1.01	-0.10	-0.01	0.00	0.7	0.9	0.1	0.0	0.0	0.45	0.54	0.01	0.00	0.00
El Salvador	1.52	1.08	-0.92	-0.20	-0.42	0.10	0.00	0.7	0.0	1.4	1.2	0.0	0.79	0.04	0.16	0.01	0.00
El Salvador	1.52	1.48	-0.93	-0.78	0.01	-0.04	0.00	0.7	0.5	0.0	0.2	0.0	0.59	0.41	0.00	0.00	0.00
Guatemala	1.52	0.94	-0.70	0.66	-0.10	0.06	0.00	0.4	0.4	0.1	0.4	0.0	0.52	0.47	0.01	0.00	0.00
Guatemala	1.52	1.50	-1.14	0.03	-0.45	0.02	0.00	1.0	0.0	1.5	0.1	0.0	0.86	0.00	0.13	0.00	0.00
Honduras	1.52	3.45	-1.20	-1.40	0.25	-0.03	0.00	1.1	1.6	0.5	0.1	0.0	0.41	0.57	0.02	0.00	0.00
Mexico	1.52	3.76	1.89	0.24	0.36	0.10	0.00	2.7	0.0	1.0	1.1	0.0	0.95	0.02	0.03	0.00	0.00
Mexico	1.52	3.51	1.83	-0.11	0.24	-0.31	0.00	2.5	0.0	0.4	11.3	0.0	0.95	0.00	0.02	0.03	0.00
Mexico	1.52	0.18	0.28	0.21	-0.24	0.10	0.00	0.1	0.0	0.4	1.1	0.0	0.41	0.23	0.30	0.05	0.00
Mexico	1.52	22.06	-0.84	4.61	0.27	-0.01	0.00	0.5	17.9	0.5	0.0	0.0	0.03	0.96	0.00	0.00	0.00
Mexico	1.52	1.16	-0.81	0.30	0.64	-0.03	0.00	0.5	0.1	3.1	0.1	0.0	0.57	0.08	0.35	0.00	0.00
Nicaragua	1.52	2.41	-1.16	-1.03	0.05	-0.03	0.00	1.0	0.9	0.0	0.1	0.0	0.56	0.44	0.00	0.00	0.00

Fuente: Análisis ACP realizado en SPAD 5.5

Tras esto, se realiza una tabla de valores test y coordenadas para cada uno de los países y para cada uno de los rangos de edad escogidos con el fin de ubicarlos en la gráfica y poder identificarlos frente a los resultados del ACP.

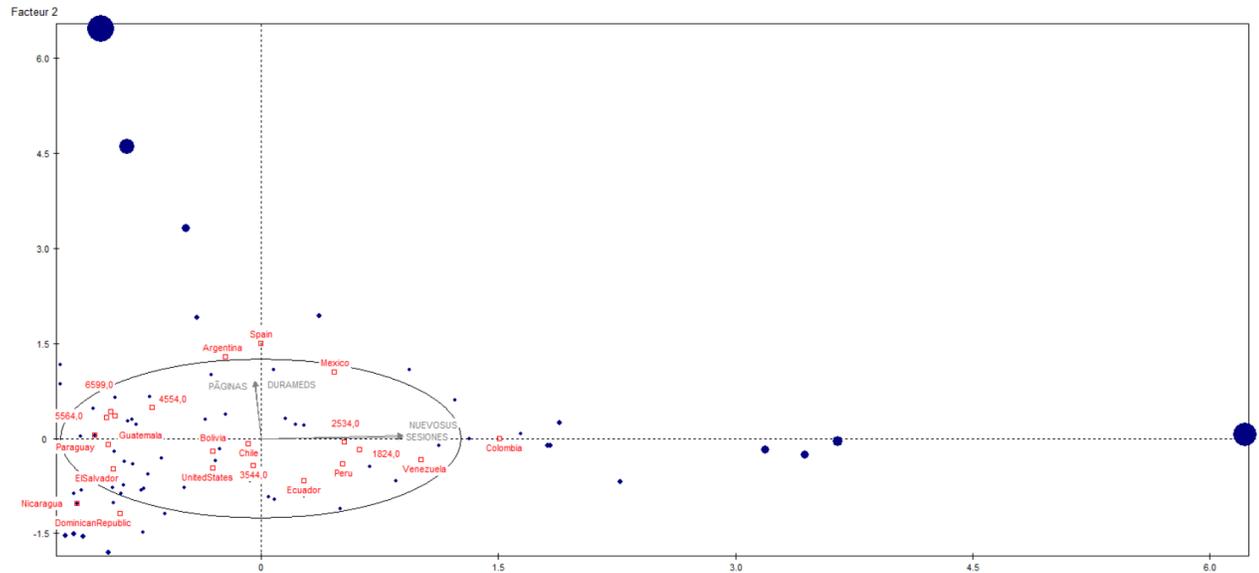
Tabla 18: tabla de coordenadas y valores test

MODALITES			VALEURS-TEST					COORDONNEES					DISTO.
IDEN - LIBELLE	EFF.	P.ABS	1	2	3	4	0	1	2	3	4	0	
5 . PAIS2													
PA01 - Argentina	5	5.00	-0.4	2.2	1.0	-0.9	0.0	-0.22	1.29	0.19	-0.05	0.00	1.75
PA02 - Bolivia	4	4.00	-0.4	-0.3	-1.5	-0.2	0.0	-0.30	-0.20	-0.32	-0.01	0.00	0.24
PA03 - Chile	5	5.00	-0.1	-0.2	0.1	-0.6	0.0	-0.08	-0.09	0.03	-0.03	0.00	0.02
PA04 - Colombia	6	6.00	2.7	0.0	-2.1	0.5	0.0	1.51	0.00	-0.37	0.02	0.00	2.43
PA05 - CostaRica	3	3.00	-1.2	-0.1	-2.0	0.6	0.0	-0.96	-0.10	-0.50	0.04	0.00	1.19
PA06 - DominicanRepublic	3	3.00	-1.1	-1.6	0.2	-0.6	0.0	-0.89	-1.19	0.05	-0.04	0.00	2.21
PA07 - Ecuador	4	4.00	0.4	-1.0	-0.4	-1.5	0.0	0.27	-0.68	-0.09	-0.09	0.00	0.55
PA08 - ElSalvador	2	2.00	-0.9	-0.5	-0.7	0.4	0.0	-0.93	-0.49	-0.21	0.03	0.00	1.15
PA09 - Guatemala	2	2.00	-0.9	0.4	-0.9	0.5	0.0	-0.92	0.35	-0.27	0.04	0.00	1.04
PA10 - Honduras	1	1.00	-0.8	-1.0	0.6	-0.3	0.0	-1.20	-1.40	0.25	-0.03	0.00	3.45
PA11 - Mexico	5	5.00	0.8	1.8	1.3	-0.7	0.0	0.47	1.05	0.25	-0.03	0.00	1.38
PA12 - Nicaragua	1	1.00	-0.8	-0.8	0.1	-0.2	0.0	-1.16	-1.03	0.05	-0.03	0.00	2.41
PA13 - Paraguay	1	1.00	-0.7	0.0	0.9	0.3	0.0	-1.05	0.05	0.41	0.04	0.00	1.27
PA14 - Peru	5	5.00	0.8	-0.7	0.6	1.0	0.0	0.52	-0.40	0.12	0.05	0.00	0.44
PA15 - Spain	6	6.00	0.0	2.9	3.5	-0.9	0.0	0.00	1.50	0.61	-0.04	0.00	2.62
PA16 - UnitedStates	5	5.00	-0.5	-0.8	-0.2	-0.7	0.0	-0.30	-0.47	-0.04	-0.03	0.00	0.31
PA17 - Uruguay	3	3.00	-1.4	-1.8	0.5	-0.5	0.0	-1.11	-1.38	0.14	-0.03	0.00	3.17
PA18 - Venezuela	5	5.00	1.7	-0.6	-1.3	3.8	0.0	1.02	-0.34	-0.26	0.19	0.00	1.25
6 . EDAD2													
ED01 - 1824,0	18	18.00	2.2	-0.6	0.2	-2.1	0.0	0.63	-0.17	0.01	-0.05	0.00	0.43
ED02 - 2534,0	14	14.00	1.6	-0.2	0.0	0.6	0.0	0.53	-0.06	0.00	0.02	0.00	0.29
ED03 - 3544,0	12	12.00	-0.1	-1.2	-0.5	1.3	0.0	-0.04	-0.44	-0.05	0.04	0.00	0.20
ED04 - 4554,0	11	11.00	-1.8	1.3	-1.6	-0.7	0.0	-0.69	0.49	-0.20	-0.02	0.00	0.76
ED05 - 5564,0	9	9.00	-2.2	0.8	1.2	0.8	0.0	-0.97	0.33	0.16	0.03	0.00	1.08
ED06 - 6599,0	2	2.00	-1.0	0.5	1.8	0.8	0.0	-0.95	0.42	0.56	0.06	0.00	1.40

Fuente: Análisis ACP realizado en SPAD 5.5

Así, finalmente se grafican las tablas anteriores en una sola gráfica que muestra la relación de los países, y los rangos de las edades con los resultados del ACP.

Gráfica 3: ACP para la edad



Fuente: Análisis ACP realizado en SPAD 5.5

Luego de esto y para complementar, se realiza una descripción de los ejes factoriales para cada uno de los factores estudiados. Para esto, primeramente se realiza la descripción de los individuos activos y sus coordenadas:

Tabla 19: Tabla descriptiva de los individuos del componente 1

COORD.	POIDS	IDENTIFICATEUR DE L'INDIVIDU	NUMERO
-1.27	1.00	Spain	1
-1.27	1.00	Costa Rica	2
-1.23	1.00	Uruguay	3
-1.20	1.00	Honduras	4
-1.18	1.00	Dominican Republic	5
-1.18	1.00	Chile	6
-1.16	1.00	Nicaragua	7
-1.14	1.00	Guatemala	8
-1.13	1.00	Uruguay	9
-1.12	1.00	UnitedStates	10
-1.06	1.00	Venezuela	11
-1.05	1.00	Paraguay	12
-1.01	1.00	Spain	13
-0.96	1.00	Uruguay	14
-0.93	1.00	El Salvador	15
-0.93	1.00	Ecuador	16
Z O N E C E N T R A L E			
0.50	1.00	Venezuela	51
0.69	1.00	Bolivia	52
0.85	1.00	Peru	53
0.94	1.00	Chile	54
1.13	1.00	Chile	55
1.23	1.00	Spain	56
1.32	1.00	Colombia	57
1.64	1.00	Spain	58
1.81	1.00	Ecuador	59
1.83	1.00	Mexico	60
1.89	1.00	Mexico	61
2.27	1.00	Venezuela	62
3.19	1.00	Colombia	63
3.44	1.00	Peru	64
3.65	1.00	Venezuela	65
6.22	1.00	Colombia	66

Fuente: Análisis ACP realizado en SPAD 5.5

A continuación se realizan dos tablas, una para las variables nominales y otra para las variables ilustrativas. Esto se realiza para cada uno de los componentes:

Tabla 20: Tabla para las variables ilustrativas del componente 1

ID.	V.TEST	LIBELLE MODALITE	LIBELLE DE LA VARIABLE	POIDS	NUMERO
ED05	-2.21	5564,0	EDAD2	9.00	1
Z O N E C E N T R A L E					
ED01	2.19	1824,0	EDAD2	18.00	23
PA04	2.74	Colombia	PAIS2	6.00	24

Fuente: Análisis ACP realizado en SPAD 5.5

Tabla 21: Tabla para las variables nominales del componente 1

COORD.	POIDS	LIBELLE DE LA VARIABLE	MOYENNE	ECART-TYPE	NUMERO
-0.04	66.00	PÁGINAS	1.53	0.45	1
Z O N E C E N T R A L E					
1.00	66.00	SESIONES	904.33	785.97	4

Fuente: Análisis ACP realizado en SPAD 5.5

Tabla 22: Tabla descriptiva de los individuos del componente 2

COORD.	POIDS	IDENTIFICATEUR DE L'INDIVIDU	NUMERO
-1.81	1.00	Uruguay	1
-1.55	1.00	UnitedStates	2
-1.53	1.00	Uruguay	3
-1.50	1.00	Dominican Republic	4
-1.48	1.00	Bolivia	5
-1.40	1.00	Honduras	6
-1.19	1.00	Dominican Republic	7
-1.11	1.00	Venezuela	8
-1.03	1.00	Nicaragua	9
-1.01	1.00	Ecuador	10
-0.97	1.00	Peru	11
-0.93	1.00	UnitedStates	12
-0.91	1.00	Ecuador	13
-0.87	1.00	Dominican Republic	14
-0.87	1.00	Chile	15
-0.82	1.00	Uruguay	16
Z O N E C E N T R A L E			
0.32	1.00	UnitedStates	51
0.38	1.00	Spain	52
0.48	1.00	Venezuela	53
0.60	1.00	Spain	54
0.64	1.00	Peru	55
0.66	1.00	Guatemala	56
0.87	1.00	Costa Rica	57
1.01	1.00	Colombia	58
1.08	1.00	Chile	59
1.09	1.00	Argentina	60
1.16	1.00	Spain	61
1.91	1.00	Bolivia	62
1.93	1.00	Argentina	63
3.32	1.00	Argentina	64
4.61	1.00	Mexico	65
6.48	1.00	Spain	66

Fuente: Análisis ACP realizado en SPAD 5.5

Tabla 23: Tabla para las variables ilustrativas del componente 2

ID.	V.TEST	LIBELLE MODALITE	LIBELLE DE LA VARIABLE	POIDS	NUMERO
Z O N E C E N T R A L E					
PA01	2.22	Argentina	PAIS2	5.00	23
PA15	2.85	Spain	PAIS2	6.00	24

Fuente: Análisis ACP realizado en SPAD 5.5

Tabla 24: Tabla para las variables ilustrativas del componente 2

COORD.	POIDS	LIBELLE DE LA VARIABLE	MOYENNE	ECART-TYPE	NUMERO
Z O N E C E N T R A L E					
0.95	66.00	DURAMEDS	70.65	53.78	4

Fuente: Análisis ACP realizado en SPAD 5.5

Tabla 25: Tabla descriptiva de los individuos del componente 3

COORD.	POIDS	IDENTIFICATEUR DE L'INDIVIDU	NUMERO
-1.54	1.00	Costa Rica	1
-1.18	1.00	Bolivia	2
-1.00	1.00	Colombia	3
-0.69	1.00	Venezuela	4
-0.52	1.00	Peru	5
-0.46	1.00	Colombia	6
-0.45	1.00	Colombia	7
-0.45	1.00	Guatemala	8
-0.42	1.00	El Salvador	9
-0.42	1.00	UnitedStates	10
-0.39	1.00	Chile	11
-0.38	1.00	Venezuela	12
-0.30	1.00	Venezuela	13
-0.24	1.00	Colombia	14
-0.24	1.00	Mexico	15
-0.24	1.00	Peru	16
Z O N E C E N T R A L E			
0.27	1.00	Mexico	51
0.27	1.00	Chile	52
0.27	1.00	UnitedStates	53
0.30	1.00	Argentina	54
0.30	1.00	Argentina	55
0.36	1.00	Mexico	56
0.40	1.00	Peru	57
0.41	1.00	Paraguay	58
0.43	1.00	Peru	59
0.51	1.00	Spain	60
0.51	1.00	Peru	61
0.52	1.00	Spain	62
0.56	1.00	Spain	63
0.63	1.00	Spain	64
0.64	1.00	Mexico	65
1.59	1.00	Spain	66

Fuente: Análisis ACP realizado en SPAD 5.5

Tabla 26: Tabla para las variables nominales del componente 3

COORD.	POIDS	LIBELLE DE LA VARIABLE	MOYENNE	ECART-TYPE	NUMERO
-0.32	66.00	DURAMEDS	70.65	53.78	1
Z O N E C E N T R A L E					
0.32	66.00	PÁGINAS	1.53	0.45	4

Fuente: Análisis ACP realizado en SPAD 5.5

Tabla 27: Tabla para las variables ilustrativas del componente 3

ID.	V.TEST	LIBELLE MODALITE	LIBELLE DE LA VARIABLE	POIDS	NUMERO
PA04	-2.12	Colombia	PAIS2	6.00	1
Z O N E C E N T R A L E					
PA15	3.47	Spain	PAIS2	6.00	24

Fuente: Análisis ACP realizado en SPAD 5.5

4.2.2. ACP - sexo

Complementariamente se realizó un tercer ACP relacionando las variables duración media de la visita, total de páginas vistas, nuevos usuarios, número de sesiones (total de visitas) y la variable sexo como variable representativa e ilustrativa. Lo primero que se realizó fue tomar las 4 variables y tomar el número de datos de cada variable, el promedio y los valores mínimos y máximos, datos que se ven reflejados en la tabla x.

Tabla 28: estadísticas básicas para las variables según sexo

NUM . IDEN - LIBELLE	EFFECTIF	POIDS	MOYENNE	ECART-TYPE	MINIMUM	MAXIMUM
3 . SESE - Sesiones	38	38.00	1972.71	1890.27	193.00	8776.00
4 . NUEV - NuevosUsuarios	38	38.00	1585.61	1409.15	193.00	6406.00
5 . PÁG - PaginasporSesion	38	38.00	1.43	0.25	1.11	2.14
6 . DURA - Duracionde laSesion	38	38.00	55.43	37.53	4.34	191.15

Fuente: Análisis ACP realizado en SPAD 5.5

Tras esto se realiza una matriz de correlaciones entre las 4 variables estudiadas buscando correlaciones altas entre las variables. podemos observar una fuerte correlación (0.99) entre las nuevas visitas y el número total de sesiones en la página. También existe una fuerte correlación entre la duración media de la visita y el total de páginas vistas (0.77). todos los demás coeficientes se encuentran por debajo de 0.34 siendo la mayor (0.33) la relación entre las nuevas visitas y la duración media de las visitas seguido de la correlación entre la duración media de las visitas y el número total de sesiones (0.31).

Tabla 29: Matriz de correlaciones

	SESI	NUEV	PÁG	DURA
SESI	1.00			
NUEV	0.99	1.00		
PÁG	0.23	0.29	1.00	
DURA	0.31	0.33	0.77	1.00

Fuente: Análisis ACP realizado en SPAD 5.5

A partir de esto se realizó una tabla de valores test que determina cual es la contribución de las variables pareándose para ver el peso que poseen entre las dos. Aquí se puede observar la relación existente entre las sesiones totales y las nuevas visitas al blog (15.94). También se observa una relación menor pero igualmente significativa entre el número de páginas vistas y la duración media de las visitas (6.31)

Tabla 30: Matriz de valores test

	SESI	NUEV	PÁG	DURA
SESI	99.99			
NUEV	15.94	99.99		
PÁG	1.48	1.81	99.99	
DURA	1.97	2.10	6.31	99.99

Fuente: Análisis ACP realizado en SPAD 5.5

Es a partir de los datos obtenidos anteriormente se realiza un histograma de frecuencias con el fin de saber en cuantos componentes principales se plantean según las variables estudiadas, siendo el resultante 4 componentes principales. En el podemos observar como el 61.75% de la información se encuentra retenida en el primer componente, mientras que el segundo componente retiene un 32.32% del total de la información:

Tabla 31: Histograma de frecuencias

NUMERO	VALEUR PROPRE	POURCENTAGE	POURCENTAGE CUMULE	
1	2.4701	61.75	61.75	*****
2	1.2929	32.32	94.08	*****
3	0.2277	5.69	99.77	*****
4	0.0093	0.23	100.00	*

Fuente: Análisis ACP realizado en SPAD 5.5

Tras esto se buscó determinar en cuántos grupos iba a ser distribuidos los componentes y esto se realizó mediante un análisis de intervalos de anderson que determinó que el número de grupos debería ser 4, como muestra la tabla x a continuación:

tabla 32: intervalos de anderson

NUMERO	BORNE INFERIEURE	VALEUR PROPRE	BORNE SUPERIEURE
1	1.3445	2.4701	3.5957
2	0.7037	1.2929	1.8820
3	0.1239	0.2277	0.3314
4	0.0051	0.0093	0.0136

Fuente: Análisis ACP realizado en SPAD 5.5

Para que los datos de esta tabla puedan visualizarse correctamente, se muestra a continuación la gráfica x, donde se muestra la posición relativa de los intervalos de anderson sobre la gráfica final, osea, el numero de grupos en el cual se dividen los datos en la gráfica:

Gráfica 33: intervalos graficados



Fuente: Análisis ACP realizado en SPAD 5.5

Ya con los intervalos claros, el siguiente paso fue encontrar las coordenadas, las correlaciones entre los factores y las variables y los ejes unitarios de cada uno de los componentes principales. Esto se ve reflejado en la tabla x, donde se pueden observar correlaciones pertinentes, como la que tiene las sesiones con el primer componente (-0.85) y la de los nuevos usuarios :

Tabla 34: Coordenadas, correlaciones y ejes unitarios

VARIABLES	COORDONNEES					CORRELATIONS VARIABLE-FACTEUR					ANCIENS AXES UNITAIRES				
	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0
IDEN - LIBELLE COURT															
SESI - Sesiones	-0.85	0.53	0.01	-0.07	0.00	-0.85	0.53	0.01	-0.07	0.00	-0.54	0.47	0.02	-0.70	0.00
NUEV - NuevosUsuarios	-0.87	0.50	-0.04	0.07	0.00	-0.87	0.50	-0.04	0.07	0.00	-0.55	0.44	-0.08	0.71	0.00
PÁG - PaginasporSesion	-0.69	-0.65	-0.33	-0.01	0.00	-0.69	-0.65	-0.33	-0.01	0.00	-0.44	-0.57	-0.69	-0.07	0.00
DURA - DuraciondeLaSesion	-0.73	-0.59	0.34	0.00	0.00	-0.73	-0.59	0.34	0.00	0.00	-0.47	-0.52	0.72	0.04	0.00

Fuente: Análisis ACP realizado en SPAD 5.5

Finalmente se calculan las coordenadas, las contribuciones y los cosenos de cada uno de los individuos (en este caso, los países) de los cuales son importantes aquellos cuyo coeficiente es mayor a 2. En la tabla x puede apreciarse uno a uno estos coeficientes, pudiendo ver contribuciones muy fuertes en el primer componente con los identificadores 1 y 4. A su vez, para el segundo componente, los identificadores 1, 2, 4 y 7 son los que más aportan. Para el tercer componente, es el identificador 4 el que más aporta. Finalmente para el cuarto componente, el 8, 14 y 37 son los identificadores que más aportan. Para una lista completa de los nombres de los identificadores, dirigirse a la tabla x :

Tabla 35: Individuos coordenadas contribuciones y cosenos

INDIVIDUS			COORDONNEES					CONTRIBUTIONS					COSINUS CARRÉS				
IDENTIFICATEUR	P.REL	DISTO	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0
1	2.63	10.59	-2.64	-1.85	-0.42	-0.02	0.00	7.4	7.0	2.0	0.1	0.0	0.66	0.32	0.02	0.00	0.00
2	2.63	5.39	-1.29	-1.00	-0.32	0.02	0.00	1.8	7.4	1.2	0.1	0.0	0.31	0.67	0.02	0.00	0.00
3	2.63	0.48	-0.11	0.52	0.44	0.02	0.00	0.0	0.6	2.2	0.1	0.0	0.03	0.57	0.40	0.00	0.00
4	2.63	14.86	-1.65	-2.97	1.82	0.05	0.00	2.9	17.9	38.3	0.8	0.0	0.18	0.59	0.22	0.00	0.00
5	2.63	2.33	-1.52	-0.03	-0.11	0.06	0.00	2.5	0.0	0.1	1.2	0.0	0.99	0.00	0.01	0.00	0.00
6	2.63	0.11	0.13	-0.21	0.23	0.02	0.00	0.0	0.1	0.6	0.1	0.0	0.14	0.38	0.47	0.00	0.00
7	2.63	25.33	-4.23	2.70	0.35	-0.08	0.00	19.1	14.8	1.4	1.8	0.0	0.71	0.29	0.00	0.00	0.00
8	2.63	2.50	-0.68	1.40	-0.13	0.22	0.00	0.5	4.0	0.2	14.1	0.0	0.19	0.79	0.01	0.02	0.00
9	2.63	1.01	0.48	-0.76	0.45	-0.02	0.00	0.2	1.2	2.3	0.1	0.0	0.22	0.58	0.20	0.00	0.00
10	2.63	3.27	1.70	0.51	0.34	-0.06	0.00	3.1	0.5	1.3	1.2	0.0	0.89	0.08	0.04	0.00	0.00
11	2.63	2.35	1.47	0.42	-0.09	0.04	0.00	2.3	0.4	0.1	0.5	0.0	0.92	0.08	0.00	0.00	0.00
12	2.63	1.43	1.08	-0.27	0.45	0.01	0.00	1.2	0.1	2.3	0.0	0.0	0.81	0.05	0.14	0.00	0.00
13	2.63	1.49	-0.43	1.13	-0.08	0.15	0.00	0.2	2.6	0.1	6.7	0.0	0.12	0.86	0.00	0.02	0.00
14	2.63	0.84	0.06	0.70	0.57	0.13	0.00	0.0	1.0	3.8	4.7	0.0	0.00	0.59	0.39	0.02	0.00
15	2.63	1.35	-0.94	0.66	0.15	0.05	0.00	0.9	0.9	0.3	0.7	0.0	0.66	0.32	0.02	0.00	0.00
16	2.63	1.40	1.06	0.31	-0.41	0.04	0.00	1.2	0.2	2.0	0.5	0.0	0.81	0.07	0.12	0.00	0.00
17	2.63	1.64	0.10	-1.06	0.70	-0.04	0.00	0.0	2.3	5.7	0.4	0.0	0.01	0.69	0.30	0.00	0.00
18	2.63	1.65	0.43	-1.17	-0.32	-0.06	0.00	0.2	2.8	1.2	1.2	0.0	0.11	0.82	0.06	0.00	0.00
19	2.63	1.79	1.19	-0.48	-0.38	-0.06	0.00	1.5	0.5	1.6	1.0	0.0	0.79	0.13	0.08	0.00	0.00
20	2.63	2.15	1.35	-0.26	-0.50	-0.10	0.00	1.9	0.1	2.9	2.7	0.0	0.85	0.03	0.12	0.00	0.00
21	2.63	8.31	-2.86	-0.10	-0.37	0.06	0.00	8.7	0.0	1.6	1.2	0.0	0.98	0.00	0.02	0.00	0.00
22	2.63	0.36	0.11	0.50	-0.31	0.09	0.00	0.0	0.5	1.1	2.4	0.0	0.03	0.68	0.26	0.02	0.00
23	2.63	1.66	1.01	-0.80	-0.06	-0.06	0.00	1.1	1.3	0.0	0.9	0.0	0.62	0.38	0.00	0.00	0.00
24	2.63	3.68	1.90	0.26	-0.01	-0.01	0.00	3.8	0.1	0.0	0.0	0.0	0.98	0.02	0.00	0.00	0.00
25	2.63	2.97	1.71	0.19	-0.01	-0.01	0.00	3.1	0.1	0.0	0.0	0.0	0.99	0.01	0.00	0.00	0.00
26	2.63	4.82	2.11	0.60	-0.09	-0.04	0.00	4.7	0.7	0.1	0.5	0.0	0.92	0.07	0.00	0.00	0.00
27	2.63	2.97	0.24	-1.58	-0.63	-0.13	0.00	0.1	5.1	4.6	4.6	0.0	0.02	0.84	0.13	0.01	0.00
28	2.63	5.65	-1.74	1.62	-0.09	-0.04	0.00	3.2	5.3	0.1	0.5	0.0	0.53	0.46	0.00	0.00	0.00
29	2.63	0.39	0.27	0.56	-0.01	0.07	0.00	0.1	0.6	0.0	1.5	0.0	0.19	0.80	0.00	0.01	0.00
30	2.63	4.55	2.09	0.37	-0.20	-0.02	0.00	4.7	0.3	0.5	0.1	0.0	0.96	0.03	0.01	0.00	0.00
31	2.63	0.06	0.03	-0.24	0.33	-0.05	0.00	0.7	0.1	1.2	0.0	0.0	0.81	0.07	0.12	0.00	0.00
32	2.63	2.98	1.68	0.07	0.38	0.02	0.00	3.0	0.0	1.6	0.1	0.0	0.95	0.00	0.05	0.00	0.00
33	2.63	4.12	-1.79	-0.18	-0.93	0.09	0.00	3.4	0.1	10.1	2.1	0.0	0.78	0.01	0.21	0.00	0.00
34	2.63	10.81	-2.22	-2.31	-0.76	-0.03	0.00	5.2	10.8	6.7	0.2	0.0	0.45	0.49	0.05	0.00	0.00
35	2.63	2.34	1.48	0.17	-0.36	-0.02	0.00	2.3	0.1	1.5	0.1	0.0	0.93	0.01	0.06	0.00	0.00
36	2.63	4.00	1.93	0.52	-0.03	-0.01	0.00	4.0	0.5	0.0	0.0	0.0	0.93	0.07	0.00	0.00	0.00
37	2.63	8.49	-2.09	1.97	0.31	-0.40	0.00	4.7	7.9	1.1	44.5	0.0	0.51	0.46	0.01	0.02	0.00
38	2.63	1.07	-0.23	1.00	0.11	0.09	0.00	0.1	2.0	0.1	2.5	0.0	0.05	0.93	0.01	0.01	0.00

Fuente: Análisis ACP realizado en SPAD 5.5

También se realizó una tabla de coordenadas y valores test, útil para la clasificación nominal de los países y para poder ubicar estos en la gráfica de manera que pudiera verse correctamente la relación entre estos y los resultados obtenidos.

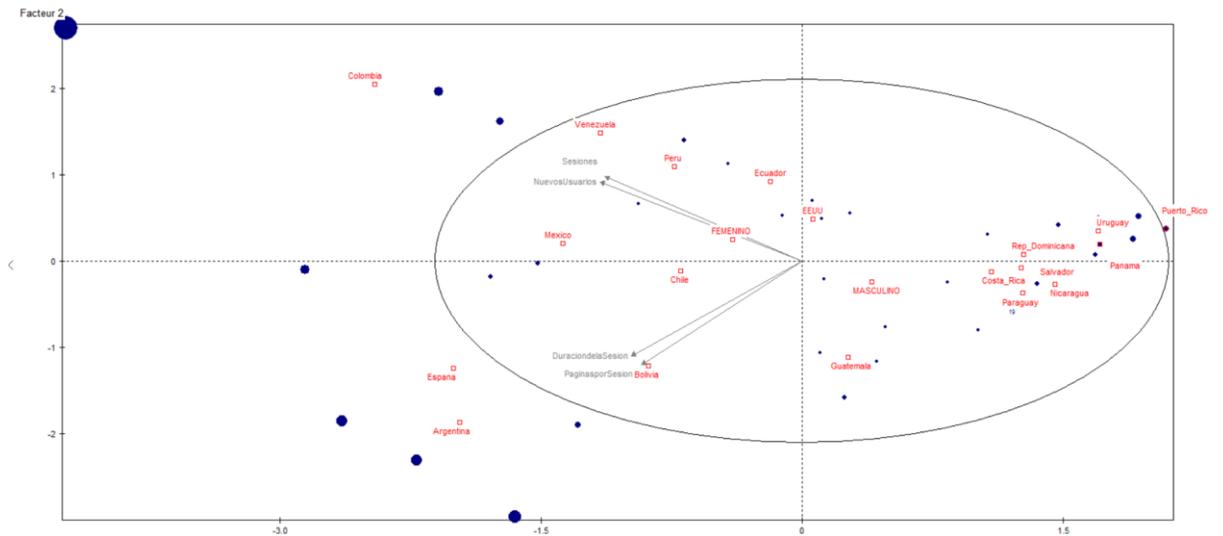
Tabla 36: Valores test y coordenadas para los países y el sexo

MODALITES				VALEURS-TEST					COORDONNEES					
IDEN	LIBELLE	EFF.	P. ABS	1	2	3	4	0	1	2	3	4	0	DISTO.
1 . Pais														
PA01	- Argentina	2	2.00	-1.8	-2.4	-1.1	0.0	0.0	-1.97	-1.88	-0.37	0.00	0.00	7.53
PA02	- Bolivia	2	2.00	-0.8	-1.5	3.4	0.5	0.0	-0.88	-1.22	1.13	0.04	0.00	3.55
PA03	- Chile	2	2.00	-0.6	-0.1	0.2	0.6	0.0	-0.70	-0.12	0.06	0.04	0.00	0.51
PA04	- Colombia	2	2.00	-2.2	2.6	0.3	1.1	0.0	-2.46	2.05	0.11	0.07	0.00	10.26
PA05	- Costa_Rica	2	2.00	1.0	-0.2	1.2	-0.6	0.0	1.09	-0.13	0.39	-0.04	0.00	1.36
PA06	- Ecuador	2	2.00	-0.2	1.2	0.7	2.1	0.0	-0.18	0.92	0.25	0.14	0.00	0.96
PA07	- EEUU	2	2.00	0.1	0.6	-0.4	0.7	0.0	0.06	0.48	-0.13	0.04	0.00	0.26
PA08	- España	2	2.00	-1.8	-1.6	-2.5	0.5	0.0	-2.00	-1.24	-0.85	0.03	0.00	6.28
PA09	- Guatemala	2	2.00	0.2	-1.4	0.6	-0.8	0.0	0.27	-1.11	0.19	-0.05	0.00	1.35
PA10	- Honduras	2	2.00	1.2	-0.5	-1.3	-1.2	0.0	1.27	-0.37	-0.44	-0.08	0.00	1.95
PA11	- Mexico	2	2.00	-1.3	0.2	-1.0	1.2	0.0	-1.37	0.20	-0.34	0.08	0.00	2.04
PA12	- Nicaragua	2	2.00	1.3	-0.3	-0.1	-0.5	0.0	1.46	-0.27	-0.04	-0.03	0.00	2.19
PA13	- Panama	1	1.00	1.1	0.2	0.0	-0.1	0.0	1.71	0.19	-0.01	-0.01	0.00	2.97
PA14	- Paraguay	2	2.00	1.1	-0.6	-1.1	-1.3	0.0	1.18	-0.49	-0.36	-0.09	0.00	1.76
PA15	- Peru	2	2.00	-0.7	1.4	-0.1	0.2	0.0	-0.73	1.09	-0.05	0.02	0.00	1.72
PA16	- Puerto_Rico	1	1.00	1.3	0.3	-0.4	-0.2	0.0	2.09	0.37	-0.20	-0.02	0.00	4.55
PA17	- Rep_Dominicana	2	2.00	1.2	0.1	0.5	0.4	0.0	1.27	0.08	0.18	0.03	0.00	1.66
PA18	- Salvador	2	2.00	1.1	-0.1	1.1	-0.3	0.0	1.26	-0.09	0.35	-0.02	0.00	1.72
PA19	- Uruguay	2	2.00	1.6	0.4	-0.6	-0.2	0.0	1.70	0.34	-0.19	-0.01	0.00	3.06
PA20	- Venezuela	2	2.00	-1.1	1.9	0.6	-2.3	0.0	-1.16	1.48	0.21	-0.15	0.00	3.61
2 . Sexo														
SE01	- MASCULINO	19	19.00	1.5	-1.3	0.3	1.2	0.0	0.40	-0.25	0.02	0.02	0.00	0.22
SE02	- FEMENINO	19	19.00	-1.5	1.3	-0.3	-1.2	0.0	-0.40	0.25	-0.02	-0.02	0.00	0.22

Fuente: Análisis ACP realizado en SPAD 5.5

Finalmente, se muestra todo en la gráfica x, la cual muestra la distribución final de los componentes principales y los países, junto a la distribución de los datos obtenidos en la tabla x:

Gráfica 4: Gráfica ACP sexo



Fuente: Análisis ACP realizado en SPAD 5.5

Para complementar el análisis realizado, se realiza una descripción de los ejes factoriales para cada uno de los factores estudiados. Para esto, primeramente se realiza la descripción de los individuos activos y sus coordenadas para cada uno de los factores estudiados:

Tabla 37: Identificadores y coordenadas de individuos activos del 1 componente

COORD.	POIDS	IDENTIFICATEUR DE L'INDIVIDU	NUMERO
-4.23	1.00	7	1
-2.86	1.00	21	2
-2.64	1.00	1	3
-2.22	1.00	34	4
-2.09	1.00	37	5
-1.79	1.00	33	6
-1.74	1.00	28	7
-1.65	1.00	4	8
-1.52	1.00	5	9
Z O N E C E N T R A L E			
1.47	1.00	11	30
1.48	1.00	35	31
1.68	1.00	32	32
1.70	1.00	10	33
1.71	1.00	25	34
1.90	1.00	24	35
1.93	1.00	36	36
2.09	1.00	30	37
2.11	1.00	26	38

Fuente: Análisis ACP realizado en SPAD 5.5

Tabla 38: variables nominales del componente 1

ID.	V.TEST	LIBELLE MODALITE	LIBELLE DE LA VARIABLE	POIDS	NUMERO
PA04	-2.24	Colombia	Pais	2.00	1
Z O N E C E N T R A L E					

Fuente: Análisis ACP realizado en SPAD 5.5

Tabla 39: variables ilustrativas del componente 1

COORD.	POIDS	LIBELLE DE LA VARIABLE	MOYENNE	ECART-TYPE	NUMERO
-0.87	38.00	NuevosUsuarios	1585.61	1409.15	1
Z O N E C E N T R A L E					

Fuente: Análisis ACP realizado en SPAD 5.5

Tabla 40: Identificadores y coordenadas de individuos activos del 2 componente

COORD.	POIDS	IDENTIFICATEUR DE L'INDIVIDU		NUMERO
-2.97	1.00	4		1
-2.31	1.00	34		2
-1.90	1.00	2		3
-1.85	1.00	1		4
-1.58	1.00	27		5
-1.17	1.00	18		6
-1.06	1.00	17		7
-0.80	1.00	23		8
-0.76	1.00	9		9
Z O N E C E N T R A L E				
0.60	1.00	26		30
0.66	1.00	15		31
0.70	1.00	14		32
1.00	1.00	38		33
1.13	1.00	13		34
1.40	1.00	8		35
1.62	1.00	28		36
1.97	1.00	37		37
2.70	1.00	7		38

Fuente: Análisis ACP realizado en SPAD 5.5

Tabla 41: Variables nominales para el componente 2

COORD.	POIDS	LIBELLE DE LA VARIABLE		MOYENNE	ECART-TYPE	NUMERO
-0.65	38.00	PaginasporSesion		1.43	0.25	1
Z O N E C E N T R A L E						
0.53	38.00	Sesiones		1972.71	1890.27	4

Fuente: Análisis ACP realizado en SPAD 5.5

Tabla 42: Variables ilustrativas para el componente 2

ID.	V.TEST	LIBELLE MODALITE	LIBELLE DE LA VARIABLE		POIDS	NUMERO
PA01	-2.37	Argentina	Pais		2.00	1
Z O N E C E N T R A L E						
PA04	2.59	Colombia	Pais		2.00	22

Fuente: Análisis ACP realizado en SPAD 5.5

Tabla 43: Identificadores y coordenadas de individuos activos del 2 componente

COORD.	POIDS	IDENTIFICATEUR DE L'INDIVIDU	NUMERO
-0.93	1.00	33	1
-0.76	1.00	34	2
-0.63	1.00	27	3
-0.50	1.00	20	4
-0.42	1.00	1	5
-0.41	1.00	16	6
-0.38	1.00	19	7
-0.37	1.00	21	8
-0.36	1.00	35	9
Z O N E C E N T R A L E			
0.34	1.00	10	30
0.35	1.00	7	31
0.38	1.00	32	32
0.44	1.00	3	33
0.45	1.00	9	34
0.45	1.00	12	35
0.57	1.00	14	36
0.70	1.00	17	37
1.82	1.00	4	38

Fuente: Análisis ACP realizado en SPAD 5.5

Tabla 44: Variables ilustrativas para el componente 3

ID.	V.TEST	LIBELLE MODALITE	LIBELLE DE LA VARIABLE	POIDS	NUMERO
PA08	-2.55	España	Pais	2.00	1
Z O N E C E N T R A L E					
PA02	3.39	Bolivia	Pais	2.00	22

Fuente: Análisis ACP realizado en SPAD 5.5

Tabla 45: Variables nominales para el componente 3

COORD.	POIDS	LIBELLE DE LA VARIABLE	MOYENNE	ECART-TYPE	NUMERO
-0.33	38.00	PaginasporSesion	1.43	0.25	1
Z O N E C E N T R A L E					
0.34	38.00	DuraciondeLaSesion	55.43	37.53	4

Fuente: Análisis ACP realizado en SPAD 5.5

5. CONCLUSIONES

Efectivamente, con un ACP pueden conseguirse análisis mucho más exactos, precisos y robustos en cuanto al origen y la distribución del tráfico web que recibe un blog se refiere. Entre esto, que países y que segmentos de edad de estos países son los que más importan y a los cuales apunta el mercado que está llegando al sitio web o al blog. Es posible que este experimento pueda ser replicado con cualquier cantidad de variables cuantitativas extraídas de Google Analytics, lo cual facilita bastante el análisis y lo agiliza.

En cuanto a los resultados específicos del trabajo, se nota como los países hispanohablantes cercanos a Colombia y Colombia como país origen del blog son los que más reciben visitas, variando bastante el rango de edad en el cual ingresan al blog, siendo las edades más tempranas (de 18 a 25) provenientes de países latinoamericanos y las edades más avanzadas (de 26 a 44) provenientes de Centroamérica. Finalmente las visitas de más edad (45 en adelante) provienen de Europa.

Colombia, España, México y Argentina son los países que más reflejan las 4 características ingresos netos, páginas impresas, total de visitas y clicks totales. Lo siguen Venezuela, Perú, Chile y Estados Unidos.

El grupo más grande de datos que se ve a la derecha son aquellos países de los cuales se ha recibido visitas esporádicas. La principal diferencia que se puede ver entre estos dos grupos es que el primero lo componen países de habla hispana y el segundo países que tienen otros idiomas que no son el español.

Los usuarios de Colombia de 18 a 24 años son los que más sesiones y usuarios aportan a la página, seguido de Venezuela y Perú en este mismo rango de edad y luego sigue otra vez Colombia, pero con un rango de edad de 25 a 34 años.

Por otro lado, los usuarios que más aportan a la duración media de las visitas y páginas visitadas son los de España en un rango de edad de 55 a 64 años seguido de México en un rango de edad de los 45 a los 54 años y Argentina de los 45 a los 54 años. En cuanto a la edad, son los usuarios de España de 69 o más años los que más aportan a la duración de las visitas. También se nota un grupo relativamente homogéneo de países en el tercer cuadrante del ACP de la edad, mostrando que las visitas de 35 a 44 años son mayormente recibidas desde Estados Unidos, Bolivia y Chile y que de 45 años para arriba las visitas se relacionan principalmente con Paraguay, El Salvador, Nicaragua y Guatemala.

Los países que más le aportan a las variables sesiones y nuevos usuarios son Colombia, seguido por México, Perú, Chile y Ecuador y que estas visitas están relacionadas directamente con el género femenino. Para las variables duración de la sesión en segundos y páginas visitadas por sesión, los países que más aportaron fueron España, Argentina y Bolivia. Hacia el primer y el cuarto cuadrante se ve un cúmulo de países de los cuales esporádicamente se han recibido visitas, pero podemos observar que estas visitas están relacionadas con el género masculino.

6. RECOMENDACIONES

Para empezar a aumentar las visitas del blog, el webmaster debe implementar una técnica de segmentación de noticias, haciendo énfasis en aquellos segmentos de edad que más visitan en blog, sobre todo aquellos que lo visitan a través de sus celulares y hacer énfasis en los países con mayor índice de visitas. Si se puede, se han de mezclar estas dos características y se debe enfocar el contenido del blog hacia los segmentos de edad por país, haciendo énfasis en el segmento de hombre de 18 a 25 años, de donde provienen la mayoría de visitas de todos los países. Tal vez una estrategia agresiva de gaming en Youtube con juegos enfocados hacia lo paranormal publicados en forma de serie ayude a aumentar significativamente las visitas en el blog dado el segmento de edad en el que se enfoca el mercado de este blog.

7. BIBLIOGRAFÍA

Ayala P., Teresa, Redes sociales, poder y participación ciudadana *Revista Austral de Ciencias Sociales* [en línea] 2014, () : [Fecha de consulta: 24 de abril de 2016] Disponible en: <http://www.redalyc.org/articulo.oa?id=45931862002> ISSN 0717-3202

ARIMETRICS. (2013). *MÉTRICAS FUNDAMENTALES PARA ANALIZAR EL EMAIL MARKETING*. Obtenido de <http://www.arimetrics.com/blog/metricas-fundamentales-para-analizar-el-email-marketing>

Andrews, H. James (2000). *Un marco para el análisis de archivos de registro*. Obtenido de https://www.google.com.co/?gws_rd=ssl#q=Andrews%2C+H.+James.+1994.++A+framework+for+log+files+analysis.+University+of+western+Ontario%2C+Department+of+Computer+Science.++

Bangbang, Sugiarto. (15 de octubre de 2015). *Who is Google analytics*. Obtenido de <http://www.minilibra.com/articles/what-is-google-analytics.html>

Bernal, Juan. (5 de Agosto de 2015). *Modelización de factores más importantes que caracterizan un sitio en la red*. Obtenido de http://www.uv.es/asepuma/XII/comunica/bernal_martinez_sanchez.pdf

Cabero Almenara JM, L. C., & V, m. D. (2013). Las analíticas Web como elemento de apoyo al análisis de sitios Web educativos. *Revista científica electrónica de educación y comunicación en la sociedad del conocimiento*.

Colomes, Xavier (2005). *Analítica Web para Blogs: métricas, objetivos y panel de control*. Obtenido de <http://www.top-rankin.com/analitica-para-blogs/>

Dictionary, G. (2015). Obtenido de <http://dictionary.reference.com/browse/blogger>

Duffy, Mary (2016). *The Internet as a research and dissemination resource*. On health promotion Journal of Oxford. Disponible en: <http://heapro.oxfordjournals.org/content/15/4/349.full>

San agustín F.E.: (2 de noviembre de 2013). Obtenido de Tu blog paso a paso: <https://books.google.com.co/books?hl=es&lr=&id=JWoFuosBflsC&oi=fnd&pg=PA18&dq=es+el+blog+publico+o+privado%3F&ots=HNaN861mEs&sig=4lqA4NtnXQPUMbF8E3aPhqpaNJg#v=onepage&q=es%20el%20blog%20publico%20o%20privado%3F&f=false>

Gallardo. (19 de octubre de 2015). *Metodos jerarquicos de análisis cluster*. Obtenido de <http://webcache.googleusercontent.com/search?q=cache:yewlRXJD7gJ:www.ugr.es/~gallardo/pdf/cluster-3.pdf+&cd=1&hl=es-419&ct=clnk>

García, R. M. (2003). La Libertad de Expresión en la Red Internet. *Revista Chilena de Derecho Informático*.

Gomez Calvo Felix (1993), "Técnicas Estadísticas Multivariantes" Universidad de

Deusto Bilbao. Disponible en: <http://www.casadellibro.com/libro-tecnicas-estadisticas-multivariantes/9788474852974/329002>

Gonzalez, P. (16 de enero de 2014). *Aplicación del análisis de componentes principales en el área educativa*. Obtenido de http://iies.faces.ula.ve/revista/articulos/revista_09/pdf/rev09gonzalez_diaz.pdf

Grané, Aurea (2005). Análisis de componentes principales. Departamento de estadística Universidad Carlos III de Madrid. Extraído de: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&ved=0ahUKEwiiyf3lpezNAhXJth4KHfmWDoUQFggkMAE&url=http%3A%2F%2Fwww.uoc.edu%2Fin3%2Femath%2Fdocs%2FComponentes_principales.pdf&usq=AFQjCNEEBaa8r9lXreYGToeUf1r4rglthw&sig2=KH-a4IIA5Eb0QmZ1qncDVQ

Gurrea, M. T. (15 de noviembre de 2015). *Análisis de componentes principales*. Obtenido de http://www.uoc.edu/in3/emath/docs/Componentes_principales.pdf

Perez, Yopis Jaume (19 de octubre de 2015). *Análisis cluster*. Obtenido de <https://estadisticaorquestainstrumento.wordpress.com/2013/01/02/tema-19-analisis-cluster/>

Rouse, margaret (2011). *Definición de Google analytics*. Obtenido de <http://searchbusinessanalytics.techtarget.com/definition/Google-Analytics>

terrarez, gudea, manuel. (11 de junio de 2003). *Análisis de conglomerados*. Obtenido de http://webcache.googleusercontent.com/search?q=cache:_-5TOjyYhSoJ:www.uoc.edu/in3/emath/docs/Cluster.pdf+&cd=4&hl=es&ct=clnk

Tipanta, Diana (2010). Obtenido de Características de un blog: <http://informaticadta.blogspot.com.co/2012/10/caracteristicas-de-blogger.html>

Melville p, S. v. (16 de octubre de 2009). Obtenido de Social media analytics: channeling the power of the blogosphere for marketing insight: <https://translate.google.com.co/translate?hl=es&sl=en&u=http://scholar.google.com/citations%3Fuser%3Dw8JK9v0AAAAJ%26hl%3Den&prev=search>

- Quintero, enric (2000). *La necesidad de medir el off line como el online*. Obtenido de <https://investigacionobs.wordpress.com/2014/04/28/la-necesidad-de-medir-el-offline-como-el-online/>
- RAE. (20 de Octubre de 2015). Obtenido de <http://www.rae.es/>
- Vallejos,Sofía. (13 de junio de 2002). *Mineria de datos*. Obtenido de http://webcache.googleusercontent.com/search?q=cache:VvtXaeC967oJ:exa.unne.edu.ar/informatica/SO/Mineria_Datos_Vallejos.pdf+&cd=12&hl=es&ct=clnk
- Truman, Herman. (2010). *Revisiting log file analysis versus page tagging*. Obtenido de <https://web.archive.org/web/20110706165119/http://web.analyticsblog.ca/2010/02/revisiting-log-file-analysis-versus-page-tagging/>
- Truman, Herman. (2010). *Web analytics blog article*. Obtenido de <https://web.archive.org/web/20110706165119/http://web.analyticsblog.ca/2010/02/revisiting-log-file-analysis-versus-page-tagging/>
- Tabasco, U. J. (2007). Avance en informatica y sistemas computacionales. En G. M. H. Mexico: Universidad Juarez Autónoma de Tabasco.
- Nuñez,Vilma. (2005). *Metricas para medir la efectividad de un blog*. Obtenido de <http://vilmanunez.com/2014/01/13/analitica-metricas-blog/>
- Villardón, Jose Luis. (19 de octubre de 2015). *Introducción al análisis cluster*. Obtenido de <http://benjamindispensa.tripod.com/spss/AC.pdf>.