



Trabajo Final de Grado

Estudio sobre uso de Big Data en pymes

Autor: Augusto Sirera Martínez

Plan de Estudios: Grado de Ingeniería Informática

Consultor: Xavier Martínez Fontes

15 de junio de 2015



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

FICHA DEL TRABAJO FINAL

Título del trabajo:	Estudio sobre uso de Big Data en pymes
Nombre del autor:	Augusto Sirera Martínez
Nombre del consultor:	Xavier Martínez Fontes
Fecha de entrega (mm/aaaa):	06/2015
Área del Trabajo Final:	Bussiness Intelligence
Titulación:	Grado de Ingeniería Informática
Resumen del Trabajo:	
<p>El análisis basado en Big Data es una realidad actual y es usado por muchas empresas como fuente de información para mejorar sus procesos productivos, conocer mejor el entorno del negocio y establecer mejores relaciones con sus clientes. Es conocido que grandes empresas como Facebook, Netflix o Amazon apoyan buena parte de su gestión empresarial en el uso de Big Data, pero ¿puede una pyme hacer uso de Big Data para su beneficio?</p> <p>Una pyme no tiene la capacidad necesaria para abordar un proyecto de Big Data tradicional, pero hay gran cantidad de empresas que están presentando servicios basados en Big Data que pueden ser útiles a pymes. El coste de estos servicios es asumible por muchas pymes y únicamente deben preocuparse de que tipo de servicio necesitan y que información es relevante para su negocio.</p>	

El cuerpo central del proyecto presenta un catálogo de servicios Big Data que pueden ser útiles a una pyme. Entre los servicios presentados hay servicios orientados a superficies comerciales, marketing digital, conocimiento de opinión de clientes y valor en redes sociales.

Abstract (in English, 250 words or less):

The analysis based on Big Data is a current reality and is used by many companies as a source of information to improve their production processes, better understand the business environment and build better relationships with their customers. It is known that large companies such as Facebook, Netflix or Amazon support much of its business management in the use of Big Data. Can an SME take benefit of the use of Big Data?

An SME does not have the capacity to undertake a traditional Big Data project, but there are many companies that are creating Big Data based services that may be useful to SMEs. The cost of these services can be assumed by many SMEs and they should only worry about what kind of service they need and what information is relevant to their business.

The main body of the project presents a catalog of Big Data services that may be useful to SMEs. Among the services presented there are services oriented to stores, digital marketing, knowledge of customer 's opinion and value in social networks.

Palabras clave (entre 4 y 8):

Big Data, pyme, servicios en la nube, analítica de datos, relación con clientes, marketing,

Índice

1. Introducción.....	1
1.1 Motivación	1
1.2 Objetivos	1
1.3 Enfoque y método seguido.....	2
1.4 Planificación del Trabajo	3
1.5 Breve resumen de resultados obtenidos	4
1.6 Breve descripción de los otros capítulos de la memoria.....	4
2. Big Data.....	5
2.1 ¿Qué es Big Data?.....	5
2.2 Realidad de una pyme al acercarse a Big Data	7
3. Servicios Big Data	8
3.1 Google Analytics	8
3.2 IBM Watson Analytics.....	9
3.3 Kompyte	10
3.4 SumAll.....	11
3.5 Canopy Labs	13
3.6 Smetrica	14
3.7 Movintracks	16
3.8 Metrikea.....	17
3.9 TC Store	18
3.10 InsightSquared	20
3.11 Followthehashtag	21
4. Resultados del estudio	23
5. Conclusiones.....	24
6. Bibliografía	25

Lista de figuras

Ilustración 1 - Planificación TFG.....	3
Ilustración 2. Panel de Google Analytics	8
Ilustración 3. Watson analytics. Panel de presentación de datos.....	10
Ilustración 4. Kompyte. Pantalla de actualizaciones de competidores	11
Ilustración 5. SumAll. Pantalla de selección de servicios.	12
Ilustración 6. SumAll. Ejemplos de informes de actividad	12
Ilustración 7. Canopy Labs. Panel de generación de acciones de venta.	13
Ilustración 8. Página de resultados de Smétrica.	15
Ilustración 9. Movintracks. Panel de cupones comerciales.	16
Ilustración 10. Metrikea. Gráfica de flujo de clientes en planta.....	17
Ilustración 11.Metrikea. Panel con proyección de ventas y afluencia de clientes.	18
Ilustración 12. TC-Store. Gráfica de predicción de afluencia de clientes.	19
Ilustración 13. InsightSquared. Panel de proyección de ventas.	20
Ilustración 14. InSightSquared. Panel de influencia de campañas de marketing.	21
Ilustración 15. Followthehashtag. Panel principal.....	22
Ilustración 16. Followthehashtag. Panel de influencia.....	22

1. Introducción

1.1 Motivación

Big Data se ha convertido en la actualidad en un término muy usado y asociado a la gran avalancha de datos que se generan cada día, bien localmente en las empresas o disponibles en Internet.

La captura y posterior tratamiento de estos datos puede revelar información valiosa para la organización relacionada con sus productos, el interés que genera en el mercado, descubrimiento de clientes potenciales, grado de satisfacción de sus clientes, etc.

Debido a la gran cantidad de datos que hay que almacenar y tratar las tecnologías Big Data han estado aparentemente monopolizadas por grandes corporaciones, pero ¿qué soluciones existen para aprovechar esta fuente de información en una pyme?

Las posibilidades de inversión en tecnología de una pyme no son comparables a las cantidades de las que disponen las grandes empresas y dado que las inversiones requeridas para implantar soluciones de Big Data son elevadas, un error en el planteamiento del proyecto o la solución escogida puede suponer un grave problema para la empresa.

1.2 Objetivos

El objetivo de este TFG es hacer un estudio sobre las tecnologías Big Data y su aplicación en las pymes. A la conclusión del estudio deberíamos ser capaces de contestar qué enfoque se le puede dar al fenómeno Big Data de forma que sea útil a las pymes.

Para ello deberíamos ser capaces de responder a preguntas como:

¿Qué beneficios puede esperar una pyme de un proyecto Big data y que no puede esperar?

¿Qué herramientas puede usar una pyme según el tipo de datos de los que dispone y la información que desea obtener?

¿Qué condiciones debe cumplir una pyme para que pueda abordar un proyecto Big Data con expectativas de éxito teniendo en cuenta sus objetivos, estrategia digital, procesos empresariales, etc.?

1.3 Enfoque y método seguido

En una primera fase se hará una presentación de las tecnologías Big data como método de descubrimiento de información relevante para el negocio a partir del tratamiento de datos de diversas fuentes y que pueden ser estructurados o no. Esta información puede venir de datos de los sistemas de información de la empresa como un ERP, servidores web en el caso de empresas con una fuerte presencia en internet, redes sociales donde los clientes tanto actuales como potenciales pueden reflejar opiniones y experiencias que debidamente tratadas pueden ser una fuente de conocimiento para la gestión de una empresa.

Partiendo de la presentación más habitual de Big Data, que es la propia de las grandes corporaciones como Facebook, Google, Amazon, etc. se va a intentar particularizar el concepto a las necesidades y posibilidades de las pymes.

En este apartado se presentará una relación de servicios de Big Data susceptibles de ser usados por pymes, bien por una facilidad de manejo que no haga necesario disponer de especialistas en la materia como programadores o científicos de datos, bien por ser servicios que dan la posibilidad de acceder a una solución personalizada por el proveedor sin tener un coste muy elevado.

Se presentarán también casos de pymes que hacen uso de Big Data, pero no como herramienta para mejorar sus procesos y servicios, sino empresas tecnológicas que hacen uso de Big Data para extraer información y es esta información la que ofrecen como producto.

A partir de toda la información de los apartados anteriores se desarrollaran unas conclusiones que deben dar respuesta a las preguntas plasmadas en los objetivos del proyecto.

1.4 Planificación del Trabajo

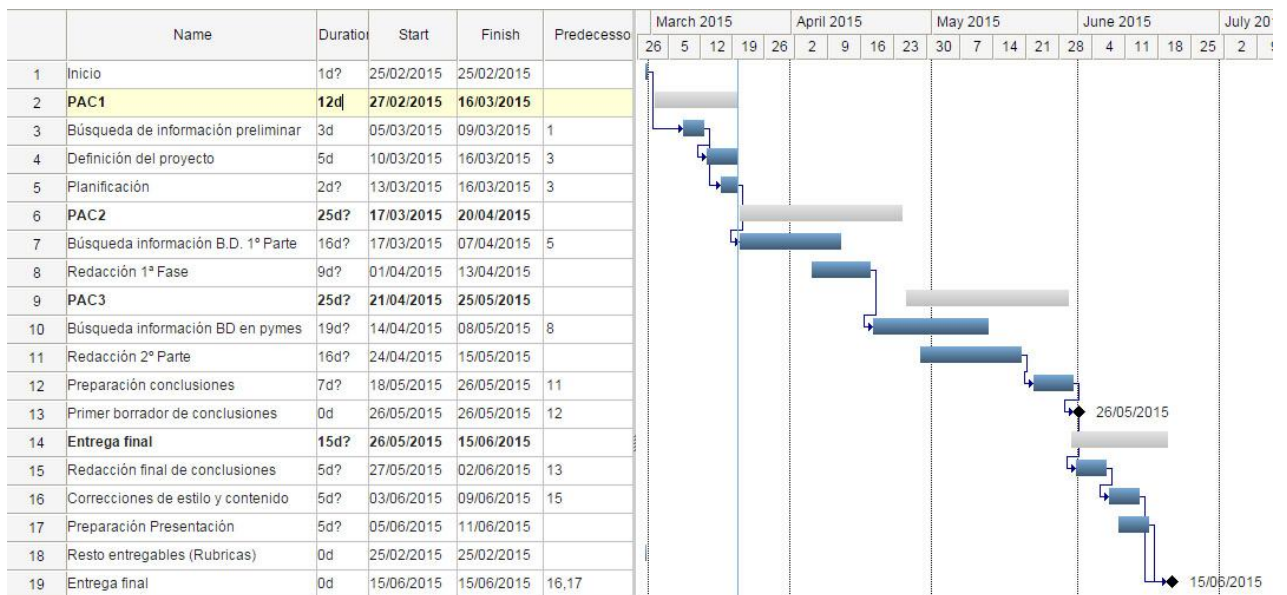


Ilustración 1 - Planificación TFG

PAC1 (16/03/2015)

Esta PAC consiste en desarrollar la definición y planificación del trabajo fin de grado.

PAC2 (20/04/2015)

Es una PAC de control deberá estar finalizada la primera parte del trabajo que habla sobre Big Data en general. Además se debería tener ya información localizada sobre Big data en pymes.

PAC3 (25/05/2015)

Es una PAC de control y debería estar ya hecha la segunda parte y un primer borrador de conclusiones. Es necesario haber llegado a este punto con el grueso del TFG hecho para disponer de tiempo antes de la entrega final para hacer las lecturas previas y correcciones necesarias.

ENTREGA FINAL (15/06/2015)

Para esta fecha hay que haber corregido la memoria del TFG, preparado la presentación para la defensa del TFG y cumplimentado el resto de entregables como las rúbricas.

1.5 Breve resumen de resultados obtenidos

La opción más viable para una pyme que quiera acercarse a Big Data es hacerlo mediante servicios en la nube. Lo más habitual es que estos servicios se nutran de una combinación de datos externos e internos a la empresa para extraer información.

El cuerpo central del trabajo es una recopilación de servicios con distintas funcionalidades que abarcan un abanico de posibilidades lo suficientemente amplio como para que muchas empresas puedan encontrar un servicio útil y viable económicamente.

1.6 Breve descripción de los otros capítulos de la memoria

El capítulo 2, Big Data, resume que es Big Data y hace una reflexión de cómo se pueden incorporar los conocimientos que brinda la analítica de datos Big Data a una pyme.

El capítulo 3 es una selección de servicios Big Data pensando en las necesidades y posibilidades de uso de una pyme.

Los capítulos 4 y 5 son los resultados y las conclusiones del trabajo respectivamente.

2. Big Data

2.1 ¿Qué es Big Data?

La primera opción para definir Big Data sería en relación al volumen de datos a procesar, pero esta medida es relativa pues el volumen de datos que puede ser considerado como Big Data dependerá del negocio que se esté considerando. Una mediana empresa productora de bienes de consumo puede considerar como Big Data un volumen de 10 Terabytes de datos. En cambio una multinacional farmacéutica consideraría un volumen de alrededor de 500 Terabytes como Big Data. En realidad Big Data no es realmente una tecnología sino más bien una combinación de nuevas tecnologías junto a tecnologías ya consolidadas que posibilitan el análisis de grandes cantidades de datos a una velocidad suficiente para poder interactuar con los datos en tiempo real. [1]

La industria ha desarrollado el concepto Big Data basándose en tres dimensiones, volumen, variedad y velocidad, las tres V que definen un juego de datos susceptible de ser considerado dentro de la categoría de Big Data.

Volumen: El volumen de un juego de datos se define por dos variables, el número de ocurrencias de la observación, que vendría a ser el número de filas en una base de datos tradicional, y la cantidad de información de cada ocurrencia, que equivale al número de campos o columnas de una base de datos. Históricamente, debido a las limitaciones para obtener datos y procesarlos, se ha trabajado con muestras de datos para poder hacer análisis predictivos, pero la proliferación de datos actual y el aumento de capacidad de proceso los actuales computadores han hecho posible la recolección y análisis de juegos de datos virtualmente completos, lo que da nuevas posibilidades en la búsqueda de relaciones entre los distintos sucesos.

Variedad; Actualmente los datos no vienen únicamente de bases de datos estructuradas donde se han recolectado datos de transacciones de negocio, encuestas a muestras de población, etc. En la actualidad muchos de los datos provienen de fuentes mucho más diversas y no estructuradas informaciones como datos de usuarios de redes sociales como Facebook, Twitter o LinkedIn,

correos electrónicos, datos de sensores, de geolocalización, logs de servidores web, etc.

Velocidad: Actualmente la agilidad en la toma de decisiones empresariales es una necesidad y el flujo de datos es constante, lo que crea la necesidad de recoger estos datos y procesarlos a tiempo casi real para acelerar la toma de decisiones, pero basándose en el conocimiento extraído de los datos.

La norma habitual en la industria es que de las tres uves un sistema debe contemplar al menos dos de ellas para ser calificado como procesos Big Data.

Los sistemas de análisis de Big Data se asocian a las tecnologías Mapreduce, Big Table y Hadoop.

Big Table es un sistema de gestión de almacenamiento distribuido de archivos desarrollado por Google pensado para almacenar grandes cantidades de datos, del orden de petabytes, repartidos por cientos de servidores simples de coste moderado.

Mapreduce fue diseñado por Google como solución a ejecutar un número elevado de operaciones sobre grandes colecciones de datos distribuidos. El componente "Map" se encarga de distribuir las operaciones entre los servidores de forma que se balancee la carga y sea tolerante a fallos y después de haber hecho las operaciones de forma distribuida el componente reduce se encarga de aglutinar los resultados para obtener un resultado final.

Hadoop es un framework desarrollado por Apache derivado de MapReduce y BigTable creado para trabajar con grandes clusters de servidores económicos. Uno de los mayores contribuidores al proyecto es Yahoo! y buena parte de la infraestructura empresarial de sistemas de Yahoo! Hace uso de Hadoop.

Hadoop ha sido diseñado para paralelizar el proceso de datos por múltiples nodos dividiendo las tareas en tareas más pequeñas y repartiendo la carga entre los distintos servidores de forma transparente, minimizando la latencia y con alta tolerancia a fallos, fallos que se gestionan automáticamente desde el sistema. Para conseguirlo Hadoop usa el algoritmo MapReduce y un sistema de archivos distribuido inspirado en BigTable.

2.2 Realidad de una pyme al acercarse a Big Data

En el anterior apartado se ha hecho un acercamiento a las tecnologías propias de Big Data y como Wolfram Rozas, ejecutivo de desarrollo Big Data de IBM, recalca en una mesa de trabajo en el 4º congreso nacional CENTAC (Centro Nacional de Tecnologías de la Accesibilidad) celebrado el 17 y 18 de octubre de 2013: “Las soluciones Big Data para mediana empresa tradicionales se mueven en unos coste de cinco cifras, es decir más de 100.000€, pero hay que tener en cuenta qué significa el termino mediana empresa para una corporación como IBM.” [2]

Estos costes no son asumibles por una pyme tal y como se entiende el termino pyme en España por lo que es necesario buscar una acercamiento a la tecnología Big Data desde una óptica diferente a la tradicional, entendida ésta como el desarrollo de un sistema de captura de grandes volúmenes de información, proceso y extracción de información con las tecnología habituales.

La mejor solución para un acercamiento a la analítica de datos para las pymes son los servicios en la nube pues son servicios que no precisan de grandes inversiones, la mayoría de ellos ningún desembolso inicial, y unas tarifas que no suponen un gran gasto para la empresa. Llegados a esta conclusión se va a realizar una enumeración de servicios basados en Big Data útiles para una pyme. Entre estos servicios hay tanto propuestas realizadas por grandes empresas como IBM y su producto Watson Analytics, como unos servicios ofrecidos por pequeñas y medianas empresas que se han especializado en el proceso de datos y usan el Big Data para ofrecer sus servicios a otras empresas.

3. Servicios Big Data

3.1 Google Analytics

https://www.google.com/intl/es_ALL/analytics/index.html

Google Analytics es una herramienta de analítica web creada por Google y gratuita en su versión estándar, suficiente para una pyme pues es capaz de procesar 10 millones de visitas al mes con una actualización diaria. Google Analytics nos ofrece diferentes métricas que permiten conocer el comportamiento de los usuarios en la página web de la empresa.

Google Analytics tiene diferentes secciones que analizan desde diferentes ópticas el tráfico de la página web. Estas secciones son:

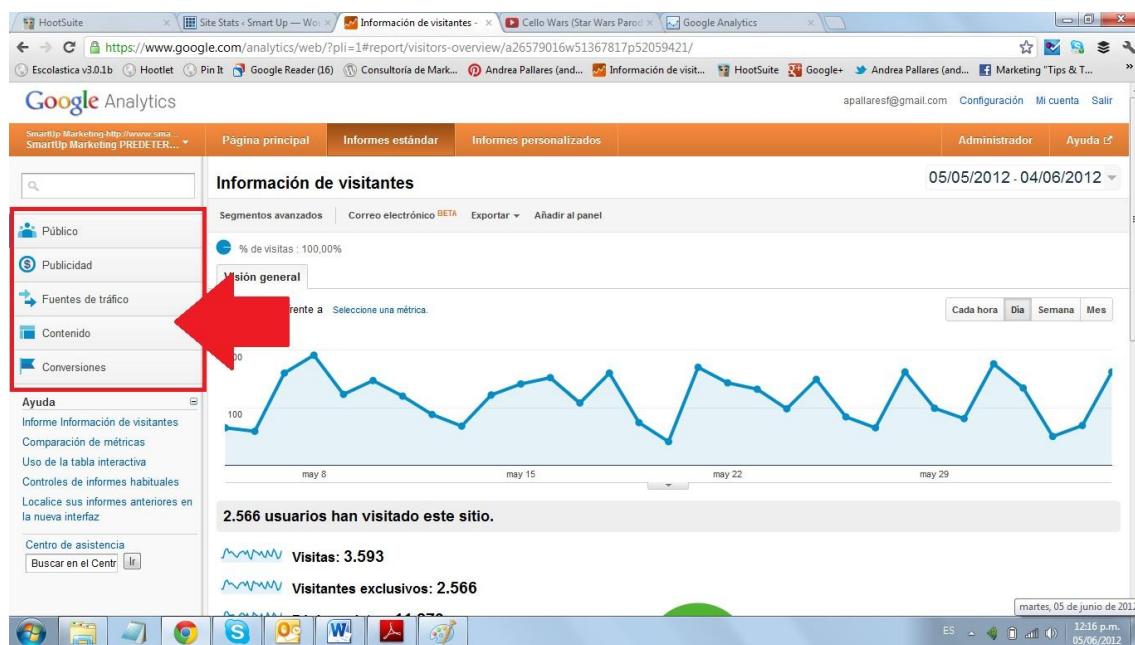


Ilustración 2. Panel de Google Analytics

- Público, esta sección permite analizar la cantidad de visitas que recibe la página web, usuarios únicos, total de páginas vistas, promedio de páginas por visita, duración media de la visita, porcentaje de rebote, porcentaje de visitas nuevas, datos demográficos como país y ciudad de los visitantes de la web e información del sistema que usan para acceder a la web, diferenciando por sistema operativo, si es fijo o móvil o resolución de pantalla.

- Publicidad, esta sección permite seguir la campaña de AdWords en el caso de que hayamos contratado este servicio de publicidad de Google.
- Fuentes de tráfico, es una sección donde podemos analizar como llegan los visitantes a la web de la empresa. Básicamente hay tres categorías de tráfico, directo, orgánico, social y derivado o referido.
- Contenido, es la sección en la que permite analizar el comportamiento de los usuarios dentro de la página web. El análisis de los contenidos más vistos y la interacción del usuario con la página permite analizar el interés que suscitan los productos de la empresa en los usuarios.
- Conversiones. Esta sección permite indicar donde queremos que llegue el usuario en la página para su análisis. En una página de comercio electrónico puede ser la página que indica la finalización de una venta., pero en una empresa que usa la web como escaparate de sus productos y no vende por ese canal puede ser la visita de la página donde explica un nuevo producto, o donde envía un mensaje pidiendo más información.

[3] Imagen cedida por la autora e información.

3.2 IBM Watson Analytics

Watson Analytics de IBM es un servicio de analítica de datos en la nube que permite una visualización de datos en diferentes formatos facilitando el análisis y la interacción con los datos y se puede usar desde un navegador web o apps en dispositivos móviles. Entre las características de Watson Analytics están el Análisis predictivo guiado, útil para revelar los hechos más relevantes y descubrir correlaciones imprevistas, y el dialogo en lenguaje natural, que permite interactuar mediante preguntas con los datos, aunque esta característica sólo está disponible en inglés actualmente. También dispone de una opción que permite revisar los datos y limpiarlos y refinarlos para obtener unos análisis de mejor calidad.

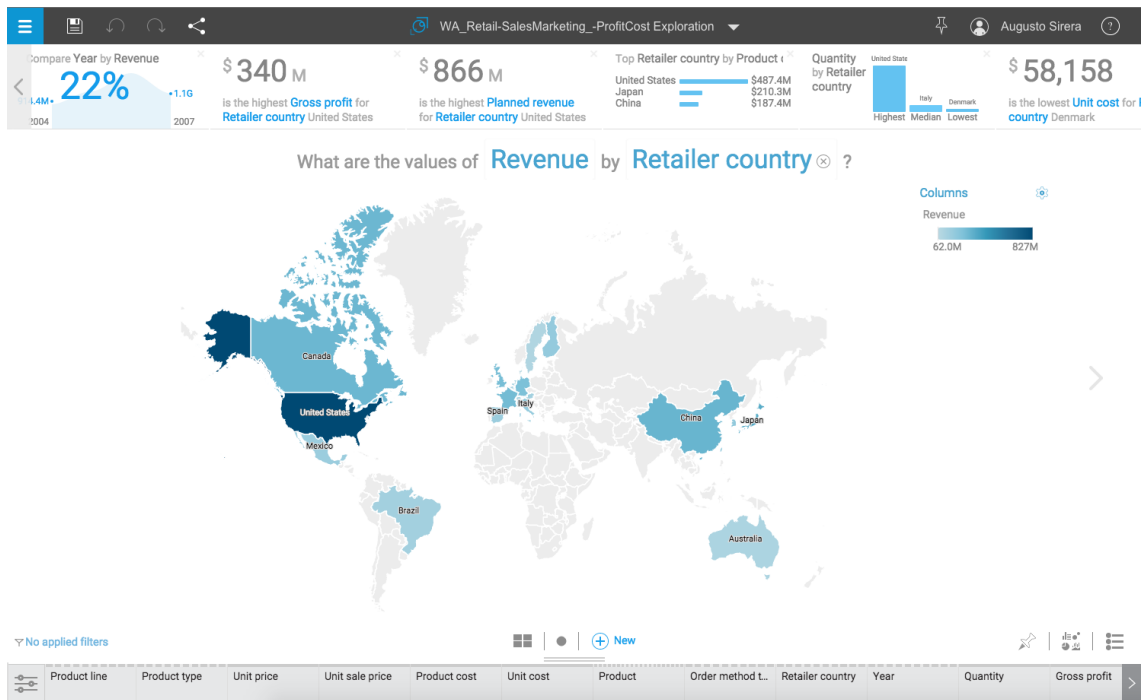


Ilustración 3. Watson analytics. Panel de presentación de datos.

La pregunta ¿Cuáles son los ingresos por país del comercio minorista?, la ha planteado el propio programa junto a muchas otras, por lo que en muchas ocasiones el propio servicio plantea las preguntas interesantes para el negocio y sólo hay que aceptarlas para ver los resultados.

Watson Analytics tiene una versión gratuita que permite 100.000 filas y 50 columnas por set de datos. La versión personal tiene un precio de 30\$ mensuales y permite hasta 1.000.000 de filas y 256 columnas y añade el acceso a datos de la red social Twitter en los análisis.

3.3 Kompyte

<http://www.kompyte.com/>

Kompyte es un servicio creado por una startup de Barcelona que consiste en una monitorización de competidores en tiempo real. Kompyte notifica los cambios en la página web, nuevas publicaciones de anuncios de búsqueda en Google, estadísticas en tiempo real de las menciones e intervenciones en redes sociales como Facebook, Twitter, LinkedIn, Youtube, Instagram y Google+ y

proporciona informes de tráfico de la página web de la competencia, fuentes de tráfico.

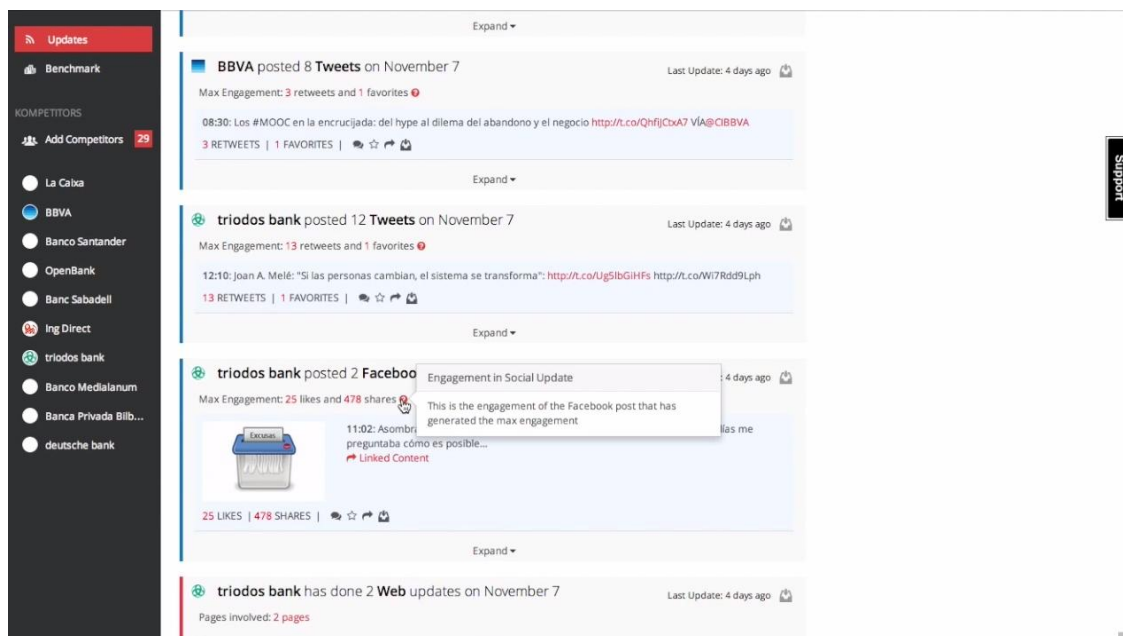


Ilustración 4. Kompyte. Pantalla de actualizaciones de competidores

Además de monitorizar los competidores que le indiquemos, Kompyte nos sugiere nuevos competidores de entre los datos que va procesando, posibilitando a la empresa conocer mucho mejor a su competencia. Kompyte es una herramienta especialmente indicada para empresas cuyo canal de venta principal sea Internet, pues al ser un canal muy dinámico permite adaptar la estrategia de marketing y venta en tiempo real a los cambios que hace la competencia.

El servicio más económico cuesta 59€ mensuales y permite monitorizar cinco competidores, por 159€ mensuales se puede monitorizar quince competidores y para un número mayor de competidores se puede pedir presupuesto personalizado.

3.4 SumAll

<https://sumall.com/>

SUMALL es un servicio en la nube que permite el análisis de datos de redes sociales con capacidad de análisis de múltiples servicios como twitter, facebook, google+, tumblr, YouTube, Instagram, Wordpress, Google Blogger, Bing Ads,

Google Adwords o Twitter Search entre otros. Permite hacer un seguimiento de los datos y descubrir relaciones entre los diferentes servicios que usa la empresa, identificación de usuarios relevantes para la empresa en las redes sociales, seguimiento de términos y etiquetas. SUMALL da la opción de hacer análisis de datos en tiempo real y recibir informes diarios personalizables.

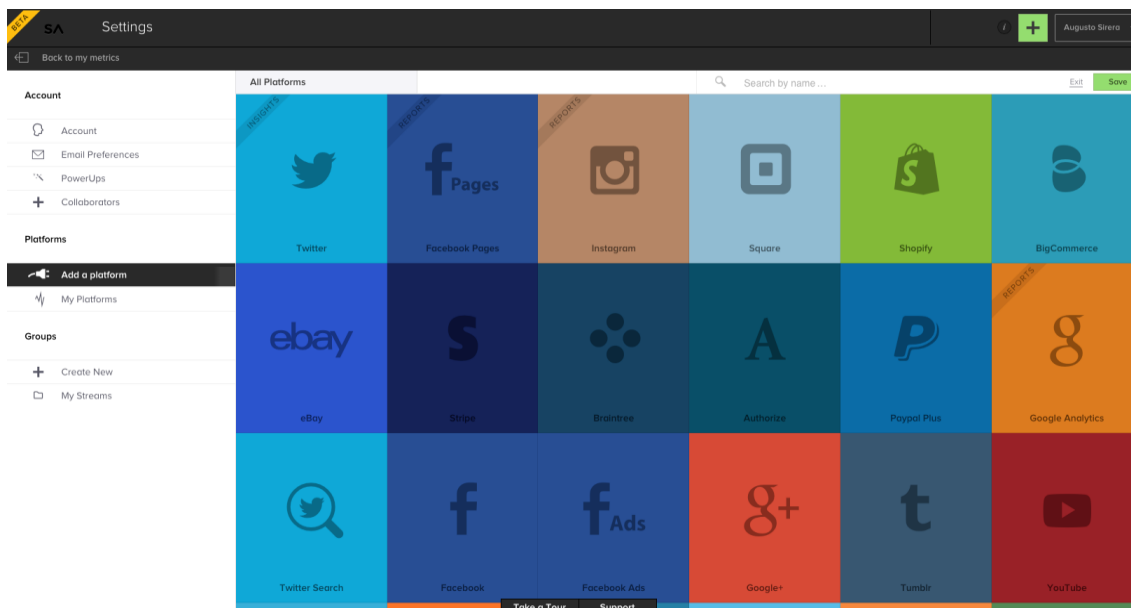


Ilustración 5. SumAll. Pantalla de selección de servicios.

El servicio de SumAll es adecuado para cualquier empresa que necesite tener información de su actividad en redes sociales para desarrollar una estrategia de marketing digital y evaluar los resultados con datos de una gran variedad de redes y las relaciones entre los datos de las distintas redes,



Ilustración 6. SumAll. Ejemplos de informes de actividad

Hay tres opciones de suscripción, SA Tracking que es gratuita, SA Reports con un coste de 59\$ mensuales y SA Insights con un coste de 99\$ mensuales. Dependiendo de las plataformas que se quieran usar en el análisis habrá que escoger una suscripción siendo necesaria como mínimo SA Reports para recibir informes periódicos automáticos y usar datos de Facebook, Twitter, Instagram y Google Adwords en los análisis.

3.5 Canopy Labs

Canopy Labs es un servicio de analítica de marketing enfocado a la monitorización y optimización de los canales de venta. El servicio evalúa el impacto de cada canal de venta en las decisiones que toman los clientes, como las visitas a la página web, las visitas desde correos electrónicos o las ventas en la tienda online. Con todo esta información hace recomendaciones personalizadas en la página web o envíos de correo electrónico personalizados basándose en la actividad del cliente y su análisis predictivo.

Toda la información puede consultarse en un panel con herramientas de reporte que se actualizan automáticamente y permite evaluar desde la globalidad de los clientes hasta clientes individuales, pasando por segmentaciones automáticas basadas en los datos adquiridos.

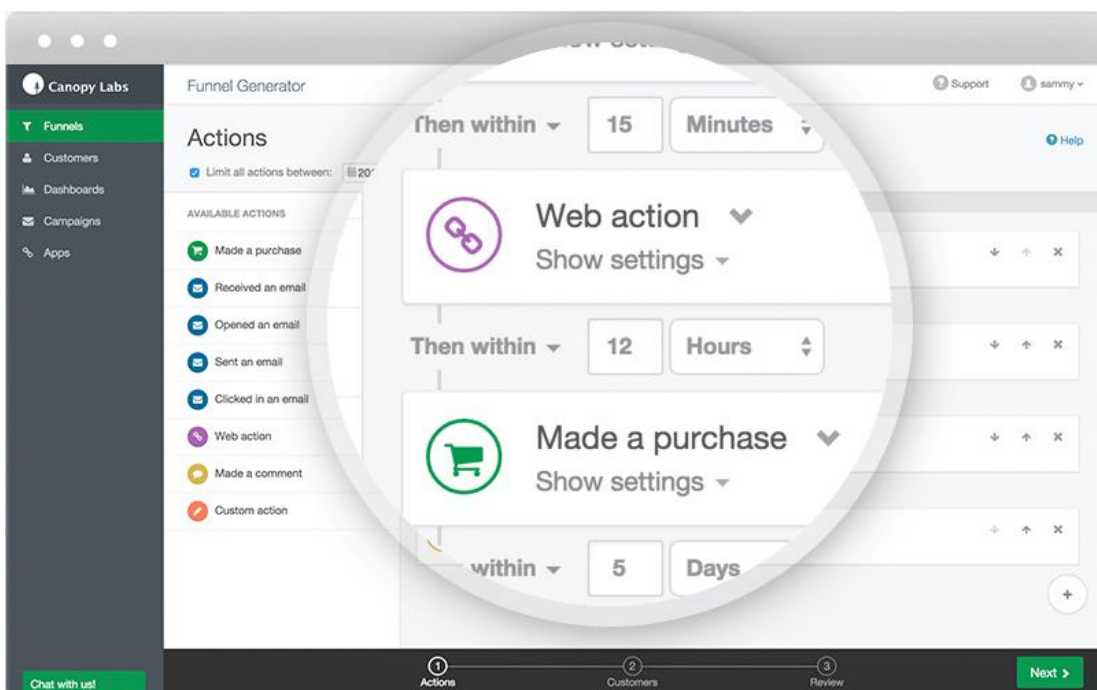


Ilustración 7. Canopy Labs. Panel de generación de acciones de venta.

3.6 Smetrica

Smétrica es una herramienta gratuita de analítica de Facebook que permite extraer estadísticas de páginas de Facebook. En un primer apartado se ve la variación en el tiempo de fans y usuarios “Hablando de esto” y en un segundo apartado el número de publicaciones, comentarios recibidos, “me gusta” recibidos y publicaciones compartidas, tanto de la marca como de los usuarios de Facebook.

Estas estadísticas sirven para ver el comportamiento de la página de Facebook corporativa así como la de posibles competidores. Para ver los resultados hay que buscar la página de Facebook en su buscador y si no se está siguiendo se hace la petición de seguimiento. A partir de ese momento Smétrica recopilará datos de la página de Facebook y los mostrará las próximas veces que se soliciten.

No es una herramienta muy sofisticada, pero si se tiene en cuenta que es gratuita y su facilidad de uso es muy recomendable para toda pyme que tenga página de Facebook.

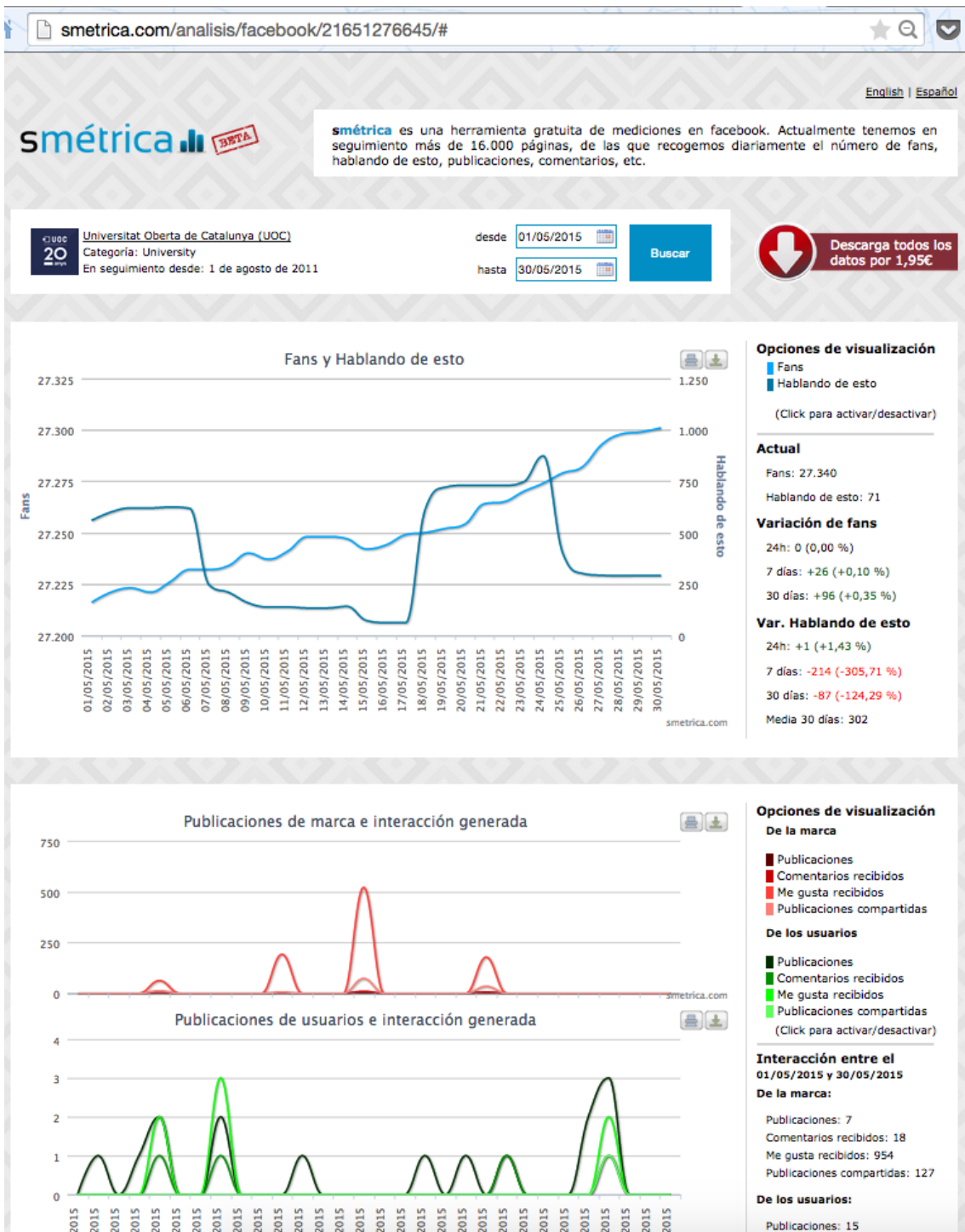


Ilustración 8. Página de resultados de Smétrica.

3.7 Movintracks

<http://movintracks.io/>

Movintracks es una compañía que hace uso de beacons y NFC para desarrollar campañas de marketing personalizadas con interacción con los clientes. Los beacons son dispositivos de unos tres centímetros de diámetro que funcionan con bluetooth 4.0 de baja energía y que se comunican con los dispositivos móviles de los clientes permitiendo reconocer sus preferencias, hábitos y analizar los movimientos en el espacio analizado.

Una de las utilidades es mediante una app de fidelización enviar mensajes interactivos a los clientes para ofrecerles promociones, ampliar información sobre productos, ofrecer cupones de descuento o cualquier otro tipo de acción de marketing.

Toda la programación de los dispositivos se realiza desde un panel que incorpora también las herramientas de analítica para poder extraer información de los datos de interacción con los clientes.

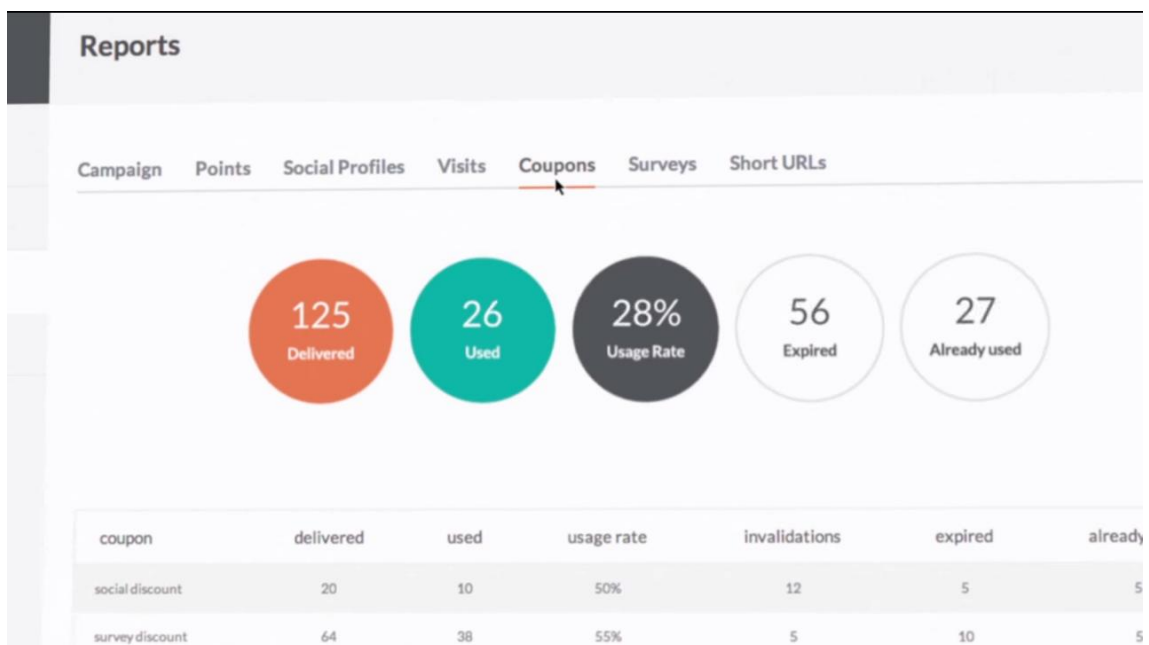


Ilustración 9. Movintracks. Panel de cupones comerciales.

3.8 Metrikea

<http://metrikea.com/>

Metrikea se define a sí misma como el Google Analytics del mundo físico. El servicio permite analizar el comportamiento de los clientes de una determinada área, típicamente un local comercial, y presentar informes de recorridos de los clientes por la tienda, zonas de afluencia, donde se paran y cuanto tiempo y a partir de estos datos se puede analizar la conversión de campañas de marketing, el grado de servicio ofrecido a los clientes o optimizar la distribución de los productos de la tienda según las necesidades del negocio.

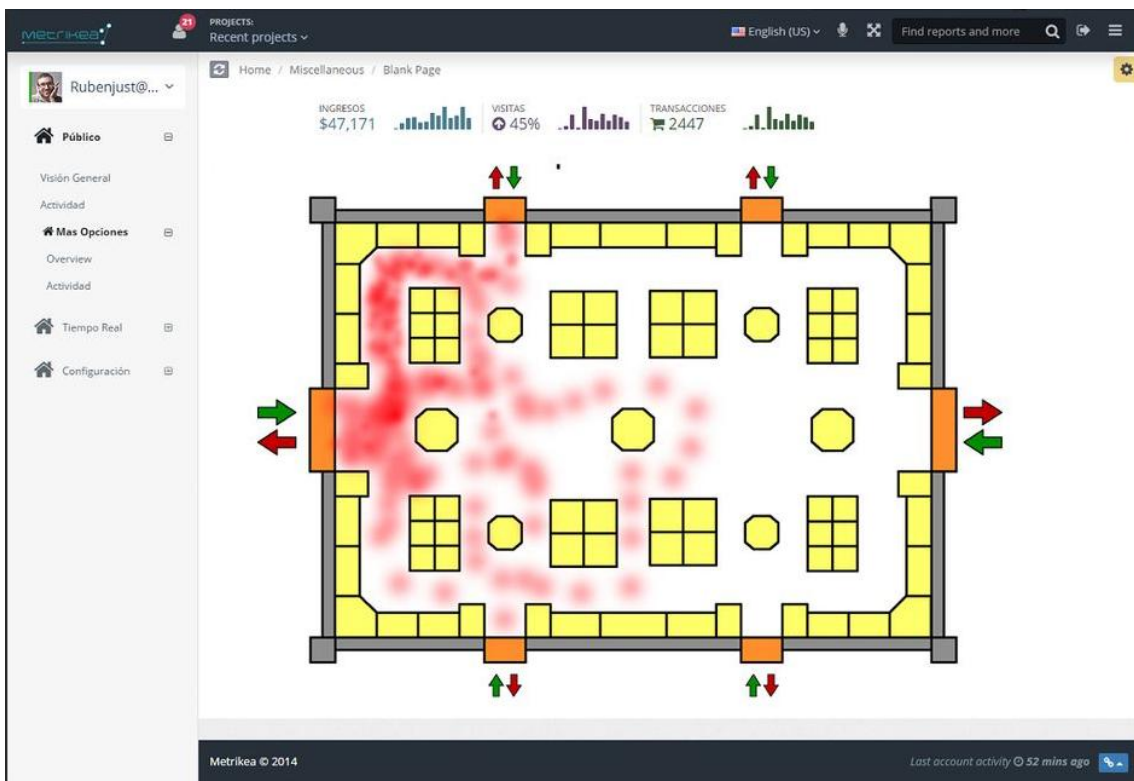


Ilustración 10. Metrikea. Gráfica de flujo de clientes en planta.

Las medidas se basan en la recolección y tratamiento de datos extraídos de las señales emitidas periódicamente por los móviles y esta información se usa para identificar usuarios únicos y sus movimientos por la tienda.

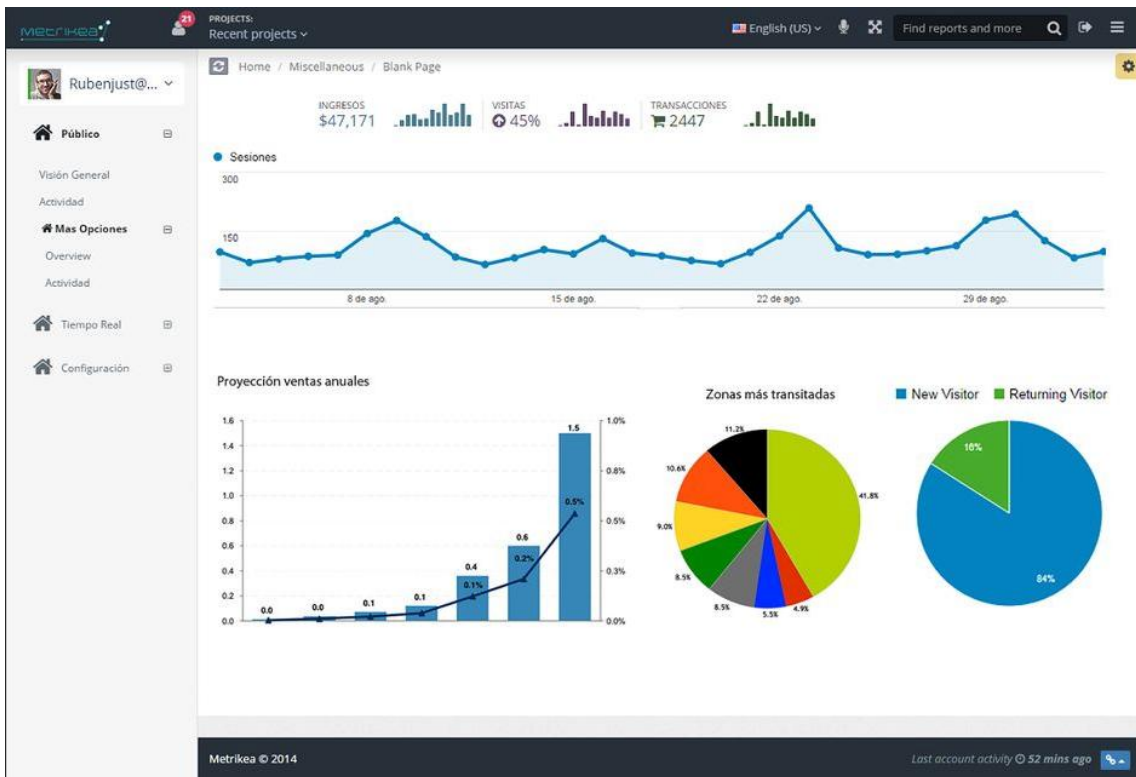


Ilustración 11. Metrikea. Panel con proyección de ventas y afluencia de clientes.

3.9 TC Store

<http://www.t-cuento.com/es/>

TC-Store es un servicio orientado a comercios proporcionado por TC-Group solutions que proporciona diferentes estadísticas como :

- Datos de flujo de personas en el interior del local comercial, incluyendo el número de potenciales clientes en el punto de venta en cada momento, el tiempo de estancia media del cliente, zonas más y menos transitadas de la tienda y estadísticas de paso frente a promociones puntuales, así como la tasa de transformación de las promociones. También se va a poder analizar cuantos de los clientes ya habían entrado anteriormente en la tienda y obtener ratios de fidelización, frecuencia de repetición y porcentaje de nuevos clientes.
- Datos del tráfico de personas exterior al local comercial, útil para evaluar la cantidad de clientes potenciales, la capacidad de atracción de los

escaparates o la publicidad exterior y el cálculo de la tasa de transformación de clientes potenciales a clientes reales.

- Reconocimiento facial mediante cámaras ópticas capaz de identificar el sexo, la franja de edad, el estado de animo y la dirección de la mirada de los clientes que pasan por delante de la cámara, así como el tiempo que permanecen frente a los escaparates o lineales, permitiendo medir incluso la efectividad de las acciones de marketing según la reacción emocional de los clientes.

Para obtener estas estadísticas se emplean lectores de entradas y salidas, lectores de tráfico exterior, sensores de movimiento, cámaras ópticas para el reconocimiento facial y dispositivos Wi-Fi para identificar los terminales móviles de los clientes. Adicionalmente se puede instalar un dispositivo para que el cliente pueda valorar su experiencia de compra o hacerlo mediante una interfaz en el teléfono móvil del cliente.

Con estos datos el software incluido es capaz de mediante el análisis de los KPIs seleccionados calcular el máximo potencial del punto de venta y ayudar al análisis de los puntos fuertes y débiles para mejorarlos. Así mismo permite calcular el personal óptimo necesario en distintos periodos de tiempo y hacer previsiones de afluencia de clientes teniendo en cuenta el histórico de la tienda, los flujos de peatones en su exterior y las predicciones meteorológicas.



Ilustración 12. TC-Store. Gráfica de predicción de afluencia de clientes.

3.10 InsightSquared

<http://www.insightsquared.com/>

InsightSquared es un servicios de analítica de datos de Salesforce, uno de los servicios líderes de CRM. Si el equipo comercial de la empresa usa Salesforce, InsightSquared permite un análisis de los datos de para realizar predicciones de ventas por empleado o periodo de tiempo, medidas de campañas de marketing o análisis de datos erróneos entre otros. La suscripción oscila entre los 35\$ y los 95\$ mensuales, dependiendo del perfil del usuario, vendedor o de staff y las aplicaciones que se vayan a usar.

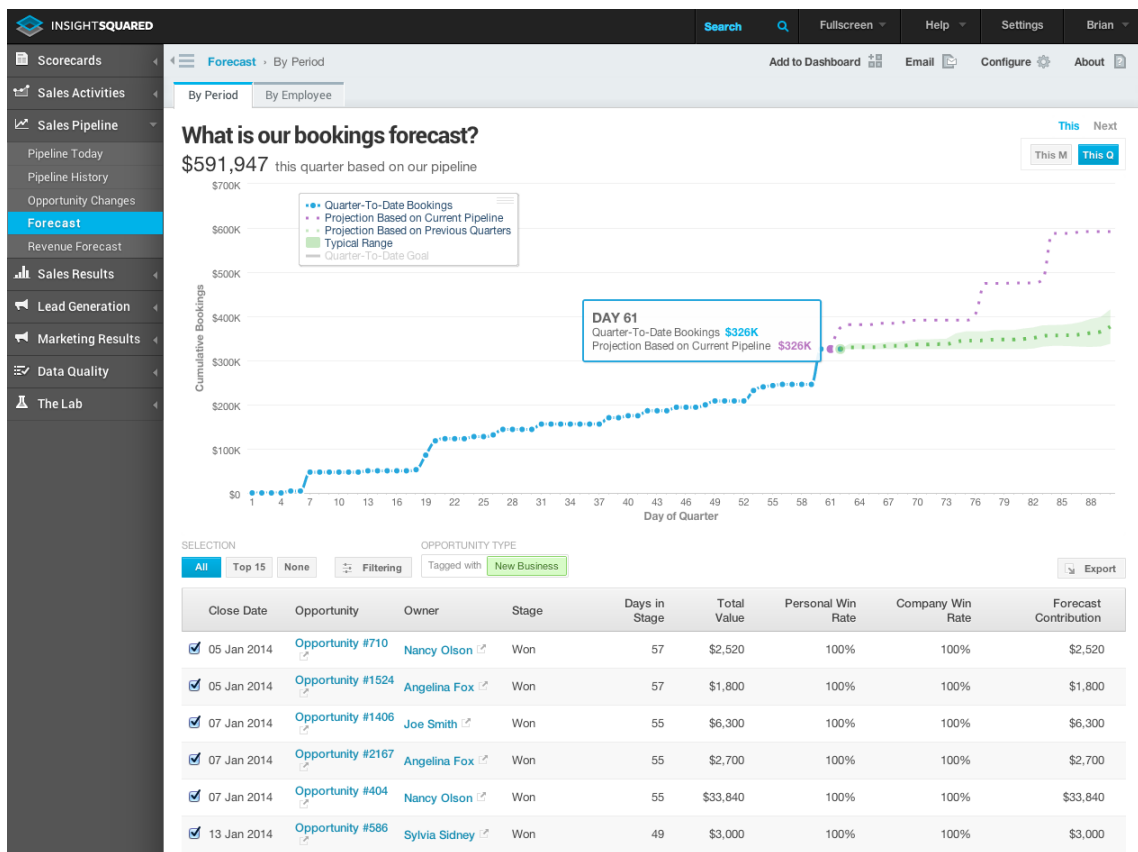


Ilustración 13. InsightSquared. Panel de proyección de ventas.

Tiene analíticas encaminadas a cubrir las necesidades de diferentes perfiles de la empresa como responsables de ventas, de marketing, financieros y gerencia, proporcionando un aspecto visual a los datos del CRM y evitando la exportación y manipulación en hojas de cálculo, procedimiento mucho más costoso en tiempo, en el que se pueden cometer errores y de mucha complejidad si se quiere elaborar estudios que combinen muchos tipos de datos distintos.

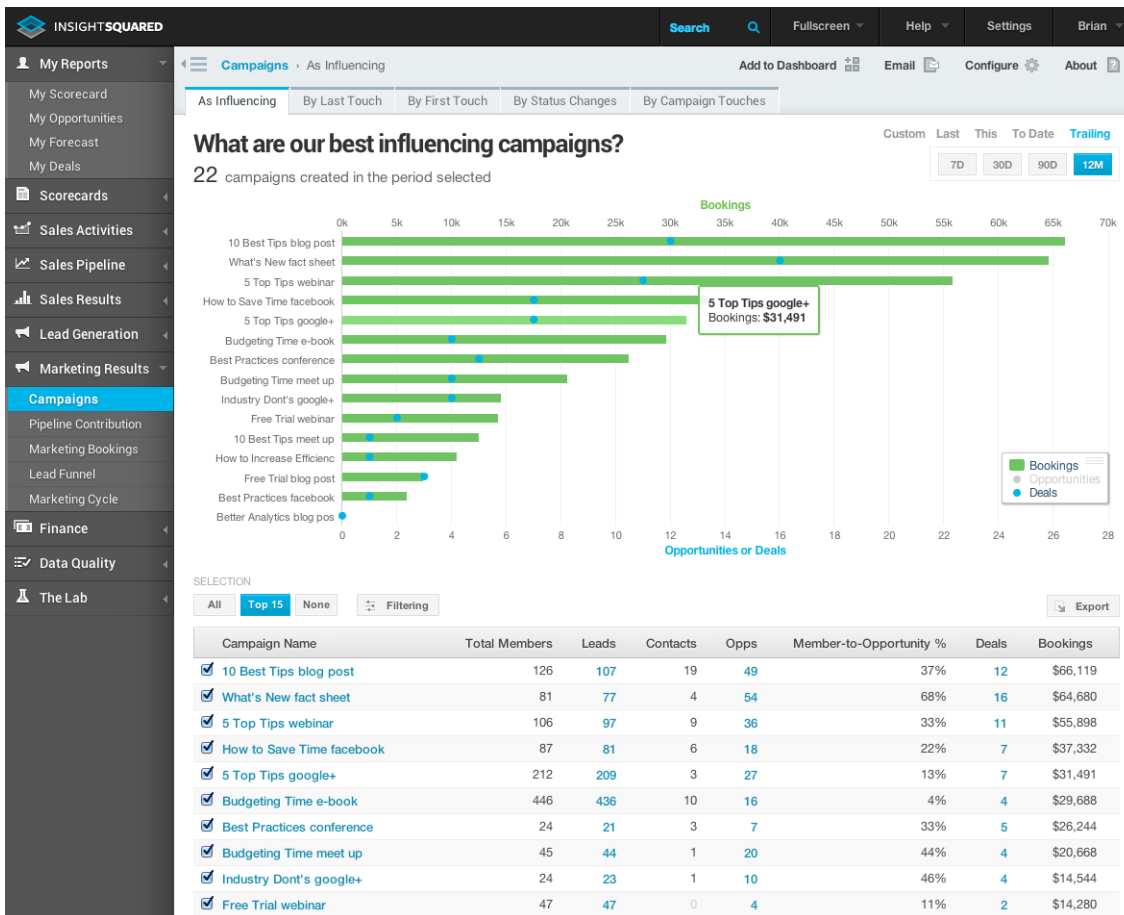


Ilustración 14. InSightSquared. Panel de influencia de campañas de marketing.

3.11 Followthehashtag

<http://www.followthehashtag.com/>

Followthehashtag es un servicio gratuito en fase beta de analítica de contenidos de twitter realizado por la agencia de comunicación DNoise. Followthehashtag permite monitorizar por palabras clave qué se está escribiendo en Twitter sobre un tema y presenta los datos de forma visual, aunque también permite exportar los datos en una hoja de cálculo. Adicionalmente a las búsquedas manuales también se puede programar un robot de búsqueda durante quince días, que permite al usuario recopilar información y hacer un seguimiento de ésta. Esta información puede ayudar a una pyme a hacer un seguimiento de sus campañas de marketing, conocer la opinión sobre sus productos y reaccionar rápidamente ante cualquier variación, sea esta positiva o negativa.

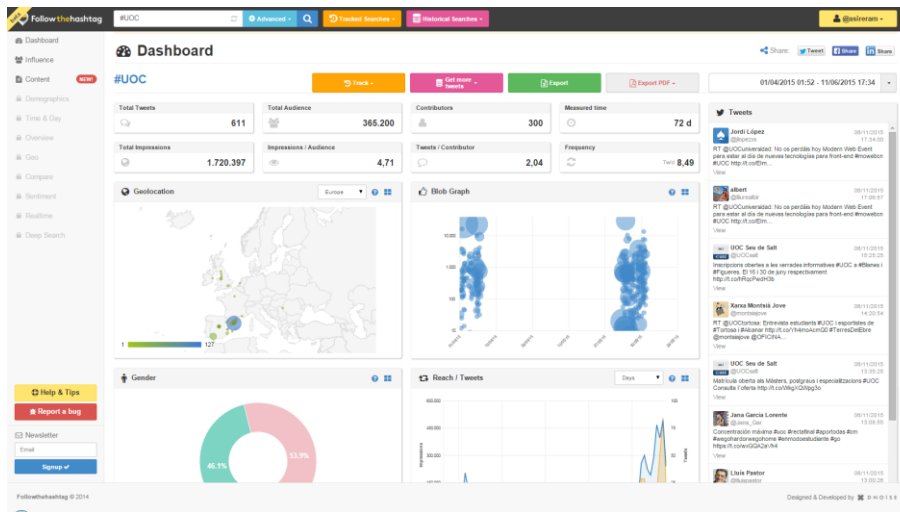


Ilustración 15. Followthehashtag. Panel principal.

La herramienta presenta datos como número total de tweets de la búsqueda, audiencia, impresiones, geolocalización, un gráfico de burbujas, donde cada burbuja es un usuario y su tamaño es proporcional al número de tweets que ha escrito, así como datos de influencia y contenido, por ejemplo cuantos seguidores tienen los usuarios que han tuiteado o el número de veces que se ha respondido o retuiteado los tweets implicados en la búsqueda.



Ilustración 16. Followthehashtag. Panel de influencia.

4. Resultados del estudio

Como se ha apuntado anteriormente en una pyme no será habitual el uso de Big Data empleando las herramientas más conocidas como Hadoop, bases de datos NOSQL y MapReduce debido al alto requerimiento de inversión y capital humano necesarios para poder acometer este tipo de proyectos. Esto no quiere decir que una pyme no pueda aprovechar las oportunidades que brinda el análisis de datos masivos acometiendo esta tarea desde otra perspectiva. En el trabajo se han reflejado algunos de los servicios más interesantes para pymes a fecha de junio de 2015, pero el análisis de datos es un mercado en expansión y no dejan de aparecer nuevos servicios.

¿Quiere decir esto que cualquier pyme puede encontrar una solución apta a sus necesidades? Seguramente sí podrá encontrarla, pero también debe cumplir unos requisitos para obtener beneficios de estas tecnologías. La analítica de datos consiste en extraer información de los datos disponibles, pero para esto lo más importante es tener datos y cuanto más fiables mejor. Por lo tanto la pyme deberá registrar y almacenar los datos de sus operaciones para poder analizarlos junto a los datos que se puedan obtener de fuentes externas como redes sociales, sensores, etc. De nada sirve llevar un control de la reputación de la empresa y sus productos en las redes sociales o de los cambios en las tendencias de ventas si no se han registrado las acciones que pueden haber sido causantes de los cambios en estos parámetros. Si no se es capaz de relacionar las causas con los efectos no se puede extraer un conocimiento válido que permita mejorar los procesos y las relaciones con los clientes y evitar volver a cometer errores ya cometidos con anterioridad. Esta es una labor costosa y no presente en la cultura actual de muchas pyme, pero lo que no se puede medir difícilmente se puede analizar y sin análisis no hay posibilidad de extracción de conocimiento, por lo que se debe apoyar la gestión en herramientas que faciliten esta tarea como SCM, ERP o CRM dependiendo de las necesidades productivas de la empresa.

5. Conclusiones

La información que tenía de Big Data antes de iniciar el TFG era muy básica y se reducía prácticamente a saber que Big Data consiste en el análisis de grandes cantidades de datos, siendo en la actualidad una de sus fuentes las redes sociales y algunos repositorios de datos abiertos ofrecidos por organismos dependientes de gobiernos, ayuntamientos, etc.

Al comenzar la búsqueda de información empecé a tomar conciencia de las herramientas usadas para procesar estos datos y la dificultad de implantación de estas soluciones en una pyme, lo que me llevo a pensar que si una pyme quiere aprovechar las ventajas de Big Data debe ser haciendo uso de servicios en la nube. Esto me hizo replantear la estructura del proyecto pues en un inicio no había pensado en un catálogo de servicios y la intención era que estuviera basado en herramientas software.

Se ha logrado desarrollar un catálogo de servicios Big Data aptos para pymes con los que adquirir conocimiento de datos internos y externos a al empresa para mejorar tanto sus productos como la relación con sus clientes. Algunos servicios como Metrikea, Movintracks o TC Store están pensados casi exclusivamente para comercios, pero el resto son de propósito más general y pueden usarse en empresas de cualquier tipo.

El cambio de visión del proyecto ha hecho que la planificación inicial no haya podido ser seguida estrictamente y la fase de búsqueda de información ha sido mucho más larga de lo planteado inicialmente, pero al final se ha logrado el objetivo y el trabajo se ha desarrollado en el tiempo establecido.

Como se ha expuesto en el anterior apartado el mercado de soluciones de analítica de datos como servicio es un mercado en expansión y continuamente están apareciendo nuevas empresas que dan servicios parecidos a ya existentes así como empresas que innovan y ofrecen nuevos servicios, por lo que este trabajo se debe considerar como un punto de partida y una foto de la situación en la fecha actual, junio de 2015.

6. Bibliografía

- [1] Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses. Michael Minelli, Michelle Chambers and Ambiga Dhiraj. Published 2013 by Jon Wiley & Sons
- [2] <http://www.centac.es/es/4congreso/mesa11> Consultado el 10/03/2015
- [3] <http://www.smartupmarketing.com/como-interpretar-las-estadisticas-de-google-analytics/> Consultado el 20/05/2015
- No se han incluido las webs de los servicios presentados pues su referencia está en la presentación de cada servicio.