

Big Data Analytcs

Diplomado en Innovación Empresarial – Universidad de Celaya en México

Autor

Edison Emel Sierra Moreno

Trabajo de grado para optar al título de Profesional en Negocios Internacionales

Dirigido por

Humberto Librado Castillo

Rodrigo Castelazo Torres

Universidad Santo Tomás

Facultad de Negocios Internacionales

Bogotá

2022

Introducción

Por medio de este ensayo vamos a tener en cuenta el seminario de diplomado en innovación empresarial que cursé en el periodo 2021-2 en el mes de diciembre, donde fue muy satisfactorio ya que obtuve nuevos conocimientos tanto para nivel profesional como en la vida personal, inicialmente este seminario fue una alianza de la Universidad Santo Tomas y la universidad de Celaya en México, donde este estuvo dividido en cuatro módulos donde el aprendizaje fue bastante completo de acuerdo a cada uno de los temas cursados, sin embargo el tema de mayor importancia y gusto para mí fue el módulo número III donde el objetivo principal era conocer un poco más a fondo sobre el Big data analytics y tener un conocimiento más práctico de algunas de las plataformas que les voy a explicar brevemente más adelante, para entrar en contexto podemos decir que el big data es la forma más rápida, eficaz y mecanizada de trabajar bases de datos ayudándonos así a tener un análisis mas completo, rápido, y de gran almacenamiento en el momento de corroborar información de las bases de datos que se esta manejando, sin embargo este tema es mucho más amplio y abarca distintos sub temas que cada día se complementan más y más, cabe resaltar que este es un tema ampliamente informático donde todo está relacionado entre sí, y siempre tiene que llegar a un objetivo final, simplificando todos los datos de la mejor manera posible para tener un resultado concreto y profesional en el momento de realizar una entrega.

Las 3VS un termino muy conocido en el mundo del big data también conocida por sus términos **volumen, variedad, y velocidad** y es principalmente lo que busca este big data, ya que las compañías y grandes negocios del mundo lo que necesitan hoy en día es una gran cantidad de información que sea administrada de la mejor manera, y sea accesible en tiempo real. No dejando atrás una 4V ya que muchos expertos aseguran que existe la llamada **4V** que aseguran ellos es el

desglose de diferentes términos técnicos, dando como resultado y recibiendo el nombre de VERACIDAD a la 4V, agregando en el entorno del big data un nuevo termino, rodeado de muchos otros factores adicionales que lograríamos encontrar en el campo de la analítica de datos, por ultimo se dio a conocer una 5V que abarca muchas acciones y precisión en la búsqueda de información por lo que se le llamo VALOR y es todo aquello que busca cualquier compañía actualmente es darle un valor a su compañía haciendo de esta una gran competencia en el mercado.

Palabras clave: Big data, informático, tiempo real, analítica, plataforma.

Introduction

Through this essay we will take into account the diploma seminar in business innovation that I attended in the period 2021-2 in the month of December, where it was very satisfying as I gained new knowledge for both professional and personal life, initially this seminar was an alliance of the University Santo Tomas and the University of Celaya in Mexico, However, the most important topic for me was the module number III, where the main objective was to learn a little more about Big data analytics and to have a more practical knowledge of some of the platforms that I will explain briefly below, to enter in context we can say that big data is the fastest, most efficient and mechanized way to work databases helping us to have a more complete analysis, fast, and of great storage at the time of corroborating information from the databases that are being managed, however this topic is much broader and covers different sub topics that every day are complemented more and more, it should be noted that this is a widely computerized

topic where everything is interrelated, and always has to reach a final goal, simplifying all the data in the best possible way to have a concrete and professional result at the time of delivery.

The 3VS is a very well known term in the world of big data also known by its terms "volume, variety, and velocity" and it is mainly what this big data is looking for, since companies and big businesses around the world what they need today is a large amount of information that is managed in the best way, and is accessible in real time. Not leaving behind a 4V as many experts claim that there is the so-called "4V" that they claim is the breakdown of different technical terms, resulting and receiving the name of VERACITY to the 4V, adding in the environment of big data a new term, surrounded by many other additional factors that we would find in the field of data analytics, Finally, a 5V was unveiled that encompasses many actions and precision in the search for information, which is why it was called VALUE and is everything that any company currently seeks is to give value to your company making this a great competition in the market.

Keywords: Big data, informatics, real-time, analytics, platform.

Figura 1.**5 V del Big Data**

Nota: Las 5 variables que rodean el Big Data. Tomado de Gutta, S. (2021). The 5 v's of big data. Analytics Vidhya. <https://medium.com/analytics-vidhya/the-5-vs-of-big-data-2758bfc51d>

Contenido

Para comenzar podríamos decir que el big data es una herramienta muy importante que abarca varias dimensiones de información, teniendo en cuenta la globalización, tecnología, y demás ventajas que vivimos en el mundo actual para la posibilidad de comprar, vender, tener información en tiempo real, conocer movimientos y muchas de otras acciones todo esto nos lo da el big data, por ende la mayoría de empresas buscan hoy en día implementar esta herramienta con la cual buscan incrementar sus ventas, mejorar estrategias de marketing, mejorar la experiencia de compra de los clientes, analizar la competencia y así día a día tener un mejor posicionamiento de su marca o empresa.

Un ejemplo práctico del big data es Nike

Según, Da Silva, D. “asociada a una empresa de tecnología, Nike comenzó a ofrecer a sus clientes dispositivos wearables (del inglés, “vestibles” o “llevables”) que monitorean los datos generados mientras se hace ejercicio” (2021).

“Así, los usuarios pueden obtener diferentes datos, como la velocidad, la frecuencia cardíaca, la distancia recorrida, entre otras, e incluso compartir toda esta información en sus redes sociales” (Da Silva, 2021).

“La estrategia aumenta el interés y compromiso del cliente con la marca, su fidelización y contribuye a que la empresa conozca mejor las expectativas de los consumidores y, con ello, mejore sus productos y servicios” (Da Silva, 2021).

En nuestras sesiones nos mostraron dos diferentes tipos de plataformas sobre el manejo de bases de datos, que son PYTHON y ANACONDA una breve enseñanza de uno de los modos de trabajo de estas plataformas fue lo que logramos evidenciar que factible, rápido, y analizado era un proceso de búsqueda y análisis de una manera mucho más accesible, teniendo en cuenta los diferentes tipos de comandos que se utilizaron en el momento de entrar a navegar en estos programas, teniendo en cuenta que dichos programas se basaban según una base de datos existente ya sea que estuvieran creadas en Excel o algún otro programa de bases, cabe resaltar que era bastante complejo algunos de estos ejercicios ya que era algo nuevo para la mayoría de los estudiantes, y que los programas funcionan diferente de acuerdo al sistema operativo que tenga cada uno de nosotros, sin embargo fueron ejercicios muy gratificantes ya que fue una nueva enseñanza en cada uno de los estudiantes para la carrera profesional, basándonos en esto podemos decir que básicamente estos programas nos ayudan en la búsqueda de alguno de los datos que se necesiten específicamente en ese momento de una manera más ágil, si nos basamos

en Excel alguna de las veces la base de datos contiene demasiados archivos y es un documento muy pesado donde en el momento de empezar a realizar una búsqueda sería un poco más tedioso, demorado, y la mayoría de las veces el Excel empezaría a fallar y el análisis no sería de lo más efectivo.

Figura 2.

Ejercicio práctico en jupyter.

```

jupyter ll (unsaved changes)
File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (
+ ↻ ↺ ⬆ ⬇ ▶ Run ⏹ Code
In [6]: df2.loc[df2['Duración'].idxmax()]
Out[6]: País          Brasil
        Ciudad      Manaus
        Año          7/04/2010
        Ciudad de Destino  Brasilia
        Duración      174
        Tipo de Cliente Promedio
        Genero        Femenino
        Ingreso       32190
        Costo         7830
        Calidad      REGULAR
        Quejas        No
        Name: 52, dtype: object

In [7]: df2['Ingreso'].std()
Out[7]: 579776.8475466274

In [8]: df2['Ingreso'].max()
Out[8]: 76298242

In [1]: df2.loc[[País,Quejas,Genero]]
NameError                                Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_15396\3254022563.py in <module>
----> 1 df2.loc[[País,Quejas,Genero]]
NameError: name 'df2' is not defined

In [ ]:

```

Nota: Analizando datos del Titanic en Jupyter . Elaboración propia con información de (s.f.). Anaconda 13-12-2021 <https://www.anaconda.com/>

Dando como resultado final un ejercicio ejemplo “ver figura 2” donde buscando los datos que se investigaron de cada pasajero de un crucero internacional donde abarcaba más de 2100 personas en su tripulación, y teniendo en cuenta lo que se quería saber acerca de una de las personas que viajaban allí, era lo siguiente: según nuestros datos recopilados y realizando un análisis de lo que queríamos conocer acerca de esta información que teníamos era la duración máxima del tripulante que viajaba, obteniendo así un resultado más completo y detallado, donde

nos muestra por ejemplo datos como la ciudad de destino, sexo, costo del tiquete, y otros datos adicionales dándole así mas efectividad a la búsqueda y de una manera más rápida y completa.

Data Warehouse

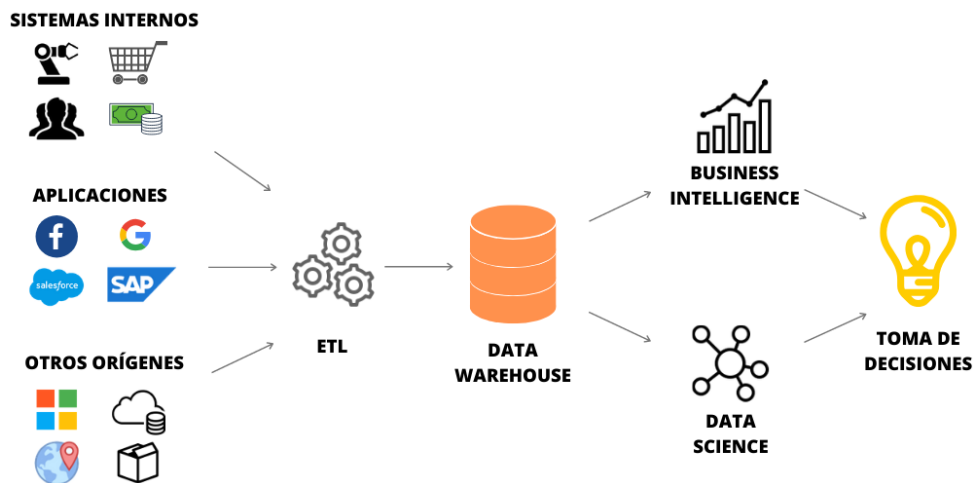
Un Data Warehouse es un repositorio electrónico donde la mayoría de las empresas u organizaciones mantiene una gran cantidad de información. Los datos deben almacenarse de forma segura, fácil de recuperar, fiable y fácil de administrar ya que maneja un sistema de gestión de datos para un análisis más rápido y concreto que ayuda a las empresas en la toma de decisiones en menos tiempo real.

Los 3 tipos de data warehouses

- “Data warehouses empresariales” (Sordo, 2020).
- “Almacén de datos operativos” (Sordo, 2020).
- “Mercado de datos” (Sordo, 2020).

Figura 3.

Proceso interno del Data Warehouse



Nota: Filtración de bases de datos y otros orígenes mediante data Warehouse para una mejor toma de decisiones. Tomado de Expertos en data warehouse. (s. f.). Mistral Business Solutions. Recuperado 13 de enero de 2022, de <https://www.mistralbs.com/expertos-data-warehouse/>

Importancia del Data Analytics en Negocios

“Data Analytics es un término empleado para describir el conjunto de procesos, tecnologías y modelos de negocio que están basados en el análisis de datos y en capturar el valor que los propios datos encierran” (Universidad de Alcalá, 2021).

La Data que generan las organizaciones constituye hoy el centro de la disrupción que viene ocurriendo en el mundo de los negocios. Las empresas se preguntan cuánto debe valer la data que almacenan. Este valor está ligado a cómo la usan. Puede ser que hoy solo exploten una fracción de ella, pero en el futuro cercano podría convertirse en la fuente para desarrollar nuevos productos y nuevos mercados. De acuerdo con la Universidad de Alcalá (Universidad de Alcalá, 2021).

Ejemplo práctico según la UNIVERSIDAD DE ALCALÁ. (2021).

Amazon

En el campo del retail, las empresas analizan la data de sus transacciones, así como el comportamiento de sus clientes para personalizar las ofertas. El objetivo principal es crear modelos predictivos, capaces de detectar las futuras necesidades del cliente o que productos se venderán mejor (5 ejemplos de uso real de Data Analytics, 2018).

Quienes compran en Amazon, pueden constatar cómo la empresa le sugiere productos que le puedan interesar como resultado de entender las preferencias del consumidor. Las empresas amplían sus datos con la información de las redes sociales, logs de navegación análisis de textos y datos de sensores para obtener una imagen completa de cliente que analizar y computar (5 ejemplos de uso real de Data Analytics, 2018).

“Mediante la aplicación de herramientas que un Máster en Data Analytics te ofrece, empresas como Target han sido capaces de detectar cuándo sus clientes esperaban un bebé” (5 ejemplos de uso real de Data Analytics, 2018).

“La razón de la importancia del Data Analytics reside en este análisis, que permite traducir estos datos en oportunidades y decisiones de negocio, permitiendo conocer en profundidad las características del mercado y los compradores a niveles analíticos hasta ahora desconocidos” (5 ejemplos de uso real de Data Analytics, 2018).

Conclusión

Los conceptos abordados juegan un rol importante en el mundo de los negocios y en la sociedad actualmente, además en el entorno en el que nos encontramos con avances cada vez más rápidos podemos deducir y evidenciar que es necesario mantener una información completa y constante, ya que esto permitirá mejorar notablemente el rendimiento de las empresas al mismo tiempo produciendo beneficios en cuanto a ahorro de dinero, análisis de su competencia, incremento de ventas y a su vez dando oportunidad a las organizaciones de crear vínculos con los clientes a los cuales se dirige, teniendo así tener un conocimiento mas amplio y personalizado de cada uno de sus consumidores con esto logrando llegar a cada uno de ellos con ofertas y promociones para incentivar y fidelizar a cada uno de sus clientes.

Sin embargo, me gustaría agregar un breve resumen acerca de lo que se hablo en una conferencia con los maestros, estudiantes y licenciados de México acerca de nuestro mundo en un futuro, haciendo una pequeña reflexión podríamos decir que el mundo actual es ahora, ya que día por día la tecnología avanza y siempre vemos nuevos emprendimientos, productos, innovaciones y demás que sorprende a la sociedad de una manera inteligente, amigable, y curiosa se hablo de que puede ser posible que en unos 20 años vayan desapareciendo los trabajos que realiza el personal humano, ya que vendría a ser remplazado por robots, que incluso en algunos países ya

podemos ver, los drones realizando el trabajo de un repartidor humano y muchas otras innovaciones que cambiarían el mundo, sin embargo no parece tan fácil como decirlo ya que tendría un proceso con muchos más fundamentos teniendo en cuenta varios aspectos negativos, positivos y demás puntos de vista en el momento de lograr realizar un trabajo como el que se menciona anteriormente ya que en el momento de ser posible, tendría que ser un proceso bastante satisfactorio, y se debatía acerca del ejemplo de los drones ya que mencionaban algunos de los compañeros que un dron puede ser robado, perdido, dañado por infraestructura, falta de control por distancia, conexión, y demás aspectos negativos que se analizaron en el momento de opinar sobre este tema, entonces ya no sería tan eficiente la entrega de un producto y podríamos decir que no podría ser incursionado un sistema de drones como opción de entrega, sin embargo cada día vemos más tecnología y no es nada descabellado que sea posible en el futuro.

Gracias al gran avance tecnológico que estamos viviendo es importante que las empresas estén al tanto de lo que sucede en nuestra sociedad, para poder llevar el ritmo del avance mundial para que puedan obtener grandes oportunidades si saben aprovechar los entornos digitales que hoy en día el mundo ofrece siendo uno de esos el big data donde es objetivo principal en el momento de interactuar con análisis, datos, rapidez tiempo real y demás acciones que abarcan este tema. Para concluir, y con lo anteriormente expuesto a lo largo del ensayo, se puede decir que el big data analytics, es la creación de valor compartido o social que identifica el contenido específico del cliente a analizar y el desarrollo de habilidades informáticas que son factores importantes a tener en cuenta en la creación y funcionamiento de empresas en el panorama actual del mundo globalizado y con la tecnología que avanza más y más, puesto que brindan diferentes herramientas que maximizan o mejoran los procesos a nivel interno de las organizaciones y así mismo con el mercado y la aceptación de su marca.

Referencias

(s.f.). Anaconda 13-12-2021 <https://www.anaconda.com/>

Da Silva, D. (2021). *¿Qué es el Big Data y para qué sirve?* 13-01-2022, de Zendesk Sitio web: <https://www.zendesk.com.mx/blog/big-data-que-es/>

Gutta, S. (2020). *Data Science: The 5 V's of Big Data*. 13-01-2022, de Analytics Vidhya Sitio web: <https://medium.com/analytics-vidhya/the-5-vs-of-big-data-2758bfcc51d>

Expertos en data warehouse. (s. f.). Mistral Business Solutions. Recuperado 26 de enero de 2022, de <https://www.mistralbs.com/expertos-data-warehouse/>

5 ejemplos de uso real de Data Analytics. (2018). Máster en Data Science. <https://www.master-data-scientist.com/ejemplos-master-en-data-analytics/>

Sordo, A. I. (2020). *Data warehouse: Guía sencilla para entender qué es un almacén de datos*. Consultado 13 de enero de 2022, de <https://blog.hubspot.es/marketing/data-warehouse>

Universidad de Alcalá. (2021). *Importancia del Data Analytics en negocios*. 13-01-2022, de Universidad de Alcalá Sitio web: <https://www.master-data-scientist.com/ejemplos-master-en-data-analytics/>

Bibliografía

Python (s.f.). Python 13-12-2021 <https://www.python.org/>