

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

Aplicación y uso del Data Mining y Big Data para estimar el comportamiento de compra de los usuarios en una empresa del sector retail de la Ciudad de Medellín

Nancy Milena Álzate Jiménez

Yesid Suárez Restrepo

Claudia Yanett Zabala Valencia

Programa Académico

Administración Tecnológica

Asesor del trabajo de grado
Julián Alberto Uribe Gómez
Msc Gestión Tecnológica

INSTITUTO TECNOLÓGICO METROPOLITANO

Fecha de entrega

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

RESUMEN

Actualmente los datos se consideran el petróleo del mundo digital, por lo tanto, el análisis Big Data debe permitir a las empresas generar una ventaja competitiva acercando su perspectiva de negocio desde el ámbito tecnológico, donde el aprovechamiento de la información pueda lograr establecer estrategias para anticipar los cambios que están vinculados directamente con el comportamiento de los consumidores.

Esta investigación se desarrolla por medio de la metodología CRISP – DM, el cual es un modelo utilizado para obtener mejores y más rápidos resultados en la minería de datos. La aplicación del modelo, se hace a través de Microsoft office Excel versión 2016 y por medio de sus complementos estadísticos, que permitieron obtener resultados cuya información ha permitido predecir y proyectar unidades a vender en determinado periodo de tiempo.

Después de haber implementado los modelos y las herramientas disponibles se logra identificar que el limpia pisos Citronela es el producto líder con un 28% de participación en ventas y por otro lado el producto menos vendido es el Limpia vidrios con una participación del 0,02%, también se observa que el grupo más vendido es la línea de aseo general con un 72,6%, adicional se proyectaron 22730 unidades a vender en diciembre de 2020 del producto líder y 295 unidades de suavizante, el cual presenta disminución de ventas en los periodos evaluados.

PALABRAS CLAVE.

Big Data, CRISP- DM, Tecnología, Información, Predicción, Estrategia.

	<p style="text-align: center;">INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm</p>	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

RECONOCIMIENTOS

En primera instancia agradecemos a nuestras familias por su apoyo y amor incondicional, en especial a nuestros padres que nos brindaron su apoyo y el aliento de continuar, siempre que lo necesitábamos para alcanzar este gran reto.

A la empresa SELLO GLOBAL S.A.S por brindarnos la oportunidad de desarrollar esta consultoría y poder cumplir con nuestro trabajo de grado.

Finalmente agradecemos a nuestro asesor Julián Alberto Uribe Gómez quien nos orientó durante esta investigación y todas aquellas personas que estuvieron dispuestas en brindarnos ayuda en este camino de aprendizaje.

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

ACRÓNIMOS

CRISP-DM (por sus siglas en ingles Cross Industry Standard Process for Data Mining)

BD (Base de datos)

 Institución Universitaria	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

CONTENIDO

RESUMEN.....	2
RECONOCIMIENTOS	3
ACRÓNIMOS	4
INTRODUCCIÓN	8
1. JUSTIFICACIÓN.....	10
2. PLANTEAMIENTO DEL PROBLEMA O SITUACIÓN	11
3. OBJETIVOS	12
3.1. Objetivo General	12
3.2. Objetivos Específicos.....	12
3.3 Alcance.....	12
4. MARCO REFERENCIAL Y DOCUMENTACIÓN CONSULTADA	13
4.1 Big Data	13
4.2 Minería de Datos	14
4.3 Sector Retail	19
4.4 Aplicación del Big Data en el mundo y sus herramientas.....	21
5. METODOLOGÍA.....	24
5.1 Metodología CRISP-DM.....	24
6. RESULTADOS.....	27
6.1 Estrategias.....	34
7. CONCLUSIONES Y RECOMENDACIONES	35
REFERENCIAS	36

	<p style="text-align: center;">INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm</p>	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

INDICE DE ILUSTRACIONES

Ilustración 1. Martínez (2017) Características de los datos.	14
Ilustración 2. Altamajer (2004) Químicos utilizados para la elaboración de detergentes.	15
Ilustración 3. Altamajer (2004) Agentes químicos empleados para la limpieza en la industria.	16
Ilustración 4. Altamajer (2004) Desinfectantes más empleados para la industria.	16
Ilustración 5. Berengueras (2017). Áreas que se benefician del Big Data.	18
Ilustración 6. Calvo (2016) Modelo CRISP-DM Data Mining.	25
Ilustración 7. Elaboración propia. Proyección de ventas limpia pisos citronela.	31
Ilustración 8. Elaboración propia. Proyección de ventas Suavizante.	33

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

INDICE DE TABLAS

Tabla 1. Adaptado de Flores & Villacís (2017). Herramientas para análisis Big Data.	22
Tabla 2. Adaptado de Rayo (2016). Profesión y perfil ocupacional del Analista de Big Data.	24
Tabla 3. Fuente Elaboración propia. Descripción variables base de datos.	27
Tabla 4. Fuente Elaboración propia. Análisis productos más vendidos.	29
Tabla 5. Fuente Elaboración propia. Análisis productos menos vendidos.	29
Tabla 6. Fuente Elaboración propia. Análisis grupo más vendido.	29
Tabla 7. Fuente Elaboración propia. Análisis de clientes con mayor participación en compras.	30

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

INTRODUCCIÓN

La investigación pretende realizar un análisis por medio de la metodología de Big Data, para determinar el comportamiento y tendencia de compra de productos fabricados con materias primas biodegradables como son los detergentes líquidos, con el fin de estudiar los resultados que permitan a las empresas del sector retail generar estrategias de marketing.

El Big Data es el conjunto de herramientas informáticas que permite la manipulación, gestión y análisis de grandes cantidades de datos que soluciona problemas que con sistemas tradicionales no se podría, debido al volumen, variabilidad, velocidad y valor (Zaforas, 2016). De esta manera, se puede concluir que dicha herramienta, puede ser útil en el ámbito empresarial para dar soluciones a problemáticas que se presentan en las organizaciones, debido a la falta de uso o incorrecto aprovechamiento de la información.

Teniendo en cuenta el concepto mencionado anteriormente y según Aguilar (2016), cuando se utiliza este tipo de herramienta se puede garantizar una ventaja competitiva que aporte características diferenciadoras frente a la competencia, generando un impacto social y ambiental frente a la situación actual de la sociedad. También se podría decir que un apropiado uso de la información ayuda a fidelizar un segmento determinado de clientes y para llegar a esto, es de suma importancia que se realice una apropiada gestión empresarial, como lo afirma Méndez (2016), la necesidad del consumidor, son verdades humanas, frescas y reveladoras que generan oportunidades de innovación, branding¹ y comunicación para las empresas, las cuales permiten entender la profunda relación emocional y simbólica entre un consumidor y un producto.

Actualmente, el aprovechamiento de este tipo de aplicaciones como son la minería de datos y el análisis Big Data, permiten que muchas empresas implementen su uso como una estrategia competitiva, permitiendo que las organizaciones puedan acercar su visión de negocio desde un panorama tecnológico, donde el aprovechamiento de la información, no solo la que se encuentra de manera evidente, como por ejemplo: saber qué frecuencia de compra tienen los clientes, que cantidades solicitan y su ubicación, sino también esa gran cantidad de datos que se encuentran vinculados directamente con la industria. Esta investigación pretende determinar desde la perspectiva tecnológica donde se pueda anticipar los cambios que se presentan en los gustos y tendencias tales como: la frecuencia de compra, cantidades específicas compra y los productos de

¹ Branding: construcción de marca.

	<p style="text-align: center;">INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm</p>	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

mayor rotación de determinado segmento de clientes, la aceptación que podría tener el lanzamiento de un nuevo producto que utilice materiales biodegradables en un sector del mercado y también que ayude a que las personas a generar conciencia de por qué se debería utilizar este tipo de productos.

Finalmente, uno de los principales objetivos que debería tener una empresa sería construir bases de datos completas y efectivas, que permita estructurar información actualizada y real sobre los tipos de clientes, para crear una estrategia de marketing que sea viable y concreta, que tenga como base un valor agregado, que ayude a conformar un historial importante y que arroje información relevante para la toma de decisiones que beneficie a ambas partes.

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

1. JUSTIFICACIÓN

Es de suma importancia el cuidado del medio ambiente, ya que este proporciona los recursos suficientes para subsistir, dicho esto, se considera de suma importancia analizar las tendencias actuales, tales como: la frecuencia de compra, cantidades específicas compra y los productos de mayor rotación de determinado segmento de clientes, donde las industrias elaboran proyectos de sostenibilidad y responsabilidad social, en este caso específico, el uso de insumos biodegradables para la elaboración de detergentes líquidos, que contribuyen con el cuidado del mismo y donde la premisa del negocio sea desarrollar mecanismos para ganar mercado con la comercialización de artículos que promuevan el cuidado de un entorno sostenible, esto se convierte en un valor agregado para las empresas del sector retail las cuales cada vez son más comprometidas con el debido uso y aprovechamiento de los recursos.

Ahora bien, es relevante resaltar la enorme ventaja que proporciona el debido uso y aprovechamiento de la información, la cantidad de datos que se encuentra relacionados con la industria de detergentes líquidos y sobre todo basados en la investigación, las empresas emergentes pequeñas y medianas a excepción de las grandes industrias no tienen estructuradas sus bases de datos, esto debido a que no cuentan con recursos suficientes o simplemente no lo consideran como una inversión prudente dentro su horizonte de negocio, es por esto y como lo menciona Merodio (2018), un análisis con el Big Data permite llevar a cabo posibles oportunidades de negocio, incrementar las ventas y evaluar los riesgos, con respecto a futuras inversiones, todo esto con el fin de ayudar a la toma de decisiones y lo más importante, reducir costos en la adquisición de materias primas, equipos, materiales y tiempos de producción; esto se convierte en un beneficio en el tiempo para toda la organización.

Por esto se considera importante desarrollar este proyecto, ya que servirá para identificar los factores de mayor importancia para el análisis del cliente, conocer sus necesidades y anticipar sus requerimientos, las características particulares que identifican sus gustos y preferencias por los diferentes productos ofrecidos por la empresa, esto resaltado dentro de un marco de innovación, con enfoque en las estrategias adecuadas de mejoramiento de procesos internos, como producción y el fortalecimiento del área de compras, con el fin de lograr el éxito en las ventas.

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

2. PLANTEAMIENTO DEL PROBLEMA O SITUACIÓN

Las empresas en la actualidad han apuntado tecnológicamente a dar soluciones y dinamismo a los requerimientos de los clientes, en su mayoría con estrategias de entrega de productos en los lugares requeridos por ellos y por supuesto, las compras online. En ocasiones dejando atrás frentes que siempre serán muy importantes, tales como, las expectativas que tiene el comprador al momento de buscar y elegir una marca. Los estudios de mercado han demostrado que los compradores se inclinan hacia temas de economía o reconocimiento de los productos; pero las nuevas tendencias demuestran que los consumidores quieren apoyar aquellas empresas que vayan de la de mano con la sostenibilidad ambiental, estos compradores también se conocen como consumidores ecológicos y que según Pardave (2010), citado por Arroyave-Ramírez & Arrubla-Zapata (2018) “son aquellos que evitan los productos que implican un derroche innecesario, que atenten contra el medio ambiente con materia prima que ponga en peligro al planeta”(parr.13).

Hechas las consideraciones anteriores y siendo Sello Global SAS, una empresa joven en el mercado, que se encuentra en el sector retail, dedicada a la fabricación y comercialización de productos de aseo para el hogar y la industria, que enmarca un concepto ambiental, surge la necesidad de explorar estrategias que le permitan posicionarse en el mercado a lo largo del tiempo, sin embargo, dada la globalización del mismo mercado, el reto es cada vez más fuerte, la gran oferta comercial conduce a la organización al desarrollo de la competitividad a la que debe dirigirse para estar a la vanguardia, lograr el reconocimiento de marca y del mismo modo incrementar las ventas.

Analizando el contexto de las tecnologías de la información y según menciona García (1994) “la tecnología ha modificado las áreas del marketing, se han creado canales de distribución, se han introducido nuevas características en el servicio a los clientes y se han desarrollado nuevos métodos de investigación de mercados y distribución”(p.96).

Por lo expuesto anteriormente se plantea en este proyecto responder a la pregunta ¿cómo las tecnologías de la información Big Data y Data Mining pueden ayudar a una empresa del sector retail a anticipar las tendencias de compra de los consumidores para generar estrategias que le permitan ser más fuertes en el mercado?

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

3. OBJETIVOS

3.1. Objetivo General

Implementar estrategias basadas en el estudio de análisis de Big Data en una empresa del sector retail que permitan anticipar tendencias de compra del consumidor

3.2. Objetivos Específicos

- Conocer el negocio de fabricación y comercialización de productos de aseo para el hogar y la industria, identificando las variables más importantes.
- Recolectar datos amplios y suficientes, referentes al estudio que se plantea y estructurando la información relevante, de tal manera que los datos sean verídicos con la mayor exactitud posible.

3.3 Alcance

Este proyecto de investigación se enmarca en un ámbito empresarial, y tiene como propósito la aplicación y uso de las tecnologías de la información Big Data, en el análisis de los datos sobre las compras realizadas por los clientes, en una empresa del sector retail de fabricación y comercialización de productos de aseo para el hogar y la industria, cuya información será evaluada a partir de los datos históricos de compra generados desde junio de 2018 hasta mayo de 2020, con el fin de modelar una herramienta que permita predecir y anticipar las tendencias del consumidor al momento de realizar sus compras, esta investigación se desarrollara en un periodo de tiempo comprendido entre febrero y noviembre de 2020.

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

4. MARCO REFERENCIAL Y DOCUMENTACIÓN CONSULTADA

4.1 Big Data

Con el fin de analizar de una manera más detallada la propuesta de utilización de Big Data para determinar estrategias gerenciales para la empresa en mención, es importante evaluar desde todos los puntos de vista el alcance de la propuesta, por lo tanto, los términos relacionados son de suma importancia para soportar la presente investigación.

Por lo tanto, se comenzará por abordar los términos relacionados con Big Data, según Mate (2014), en su repositorio “Un nuevo paradigma de análisis de datos” según el cual afirma que “En nuestros días, es un hecho incuestionable la ingente cantidad de información que se genera cada segundo en nuestro planeta. Dicha información puede ser estructurada, semi-estructurada o no estructurada. También puede aportar enorme valor a cualquier entidad o puede suponer un consumo excesivo de recursos humanos, informáticos, etc”. Según el autor, el hecho de tener grandes cantidades de datos hace imprescindible contar con tecnologías informáticas que nos ayuden a ordenar esta gran cantidad de datos e información de esta manera se podría clasificar datos de alta relevancia.

Como sugiere utilizar la fundación BBVA (2013), al término inglés Big Data para aludir a un conjunto de datos que, por su volumen y variedad y por la velocidad a la que necesitan ser procesados, supera las capacidades de los sistemas informáticos habituales. Teniendo en cuenta al significado de este término y al enfoque que se le da a esta investigación, en el cual se pretende relacionar el estudio de determinada cantidad de información de una empresa del sector industrial, en la cual su actividad económica es la fabricación y comercialización de productos de aseo para el hogar y la industria que utilizan materiales biodegradables, donde se hace completamente relevante relacionar técnicas que hagan parte de la cuarta revolución industrial para la formulación de estrategias propias que le permitan a esta industria tener una ventaja comparativa en este mercado.

Ahora bien, se debe enmarcar dentro de las características importantes del Big Data y es la veracidad de los datos que se van a analizar, puesto que esta indicara el grado de confianza asociado a cierto tipo de datos, como lo expone el autor Martínez (2017): “un factor de incertidumbre ante algunos datos, es decir, ante un contenido posiblemente asociado a error o mentira, la Big Data debe incluiría ese factor. Un ejemplo característico del sesgo que se puede producir se podría observar en sector de la energía renovable en el que para obtener X producción de energía es indispensable

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

viabilizar el máximo de variables como el viento, etc. para ver donde se pondría por ejemplo una estación de energía eólica.”

En la ilustración 1 se pueden observar las características de los datos y la manera de definir que característica se podría evaluar por medio del Big Data:



Ilustración 1. Martínez (2017) Características de los datos.

4.2 Minería de Datos

Por otro lado, también es importante mencionar tópicos relacionados con el análisis de grandes cantidades de datos, a partir de la minería de datos como lo define el autor Mendoza (2011): La minería de datos es un conjunto de técnicas, métodos y algoritmos que permiten extraer información útil desde fuentes de distinto tipo. A través de su historia diversas comunidades científicas han contribuido a su avance desde mundos tan diversos como el aprendizaje de máquina, las bases de datos, el procesamiento de lenguaje natural y la recuperación de información. Por tanto, es importante mencionar la calidad de los datos que se van a relacionar y su aporte en esta investigación.

Por otro lado, se tiene la enorme necesidad en encontrar soluciones que permitan avanzar en temas tan coyunturales como es la sostenibilidad ambiental, el mundo cambiante exige que las industrias implementen practicas cada vez más limpias y que vayan de la mano con el medio ambiente, por lo tanto el desafío de la industria en los próximos años, es buscar los medios, basados en nuevas tecnologías, la evolución constante en sus métodos de elaboración y planificación de sus proyectos

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

de producción.

Es así, como se crea la necesidad de enfrentar de manera directa los nuevos retos que presenta la industria y la producción limpia, basados en las políticas en el ámbito de las Naciones Unidas donde manifiestan según el autor Pujadas (2017) que el siguiente impulso se encuentra en la Cumbre Mundial sobre el Desarrollo Sostenible, celebrada en Johannesburgo en el año 2002 y donde resaltan la importancia de proponer lineamientos necesarios para cambiar los patrones insostenibles de consumo y producción, afirmando que para alcanzar un desarrollo sostenible a nivel global es indispensable generar cambios en la manera en que las sociedades producen y consumen.

Con el paso del tiempo se requieren técnicas más precisas que permitan a las industrias del sector productivo, especialmente a las del sector de fabricación y comercialización de productos de aseo para el hogar y la industria, generar soluciones para avanzar en un mercado competitivo. A continuación, en la ilustración 2, se pueden ver los agentes químicos que más se utilizan para la elaboración de jabones industriales y los cuales se utilizan en los hogares en la vida cotidiana:

JABONES	En 1994, el consumo mundial total fue de 4.5 millones Tm, de los cuales 2 millones Tm se utilizaron en detergentes. La demanda se concentró en Asia y Sudamérica. Los restantes 2.5 millones Tm se emplearon en la fabricación de jabón de tocador.
LAS Alquilbenceno sulfonatos	Son los principales componentes de la industria de detergentes. En Latino América y Asia está permitido el consumo de alquilbenceno sulfonatos de cadena ramificada, sin embargo en la mayoría de los países se sustituyen, debido a que no son biodegradables, por los alquilbenceno sulfonatos de cadena lineal.
FAS Sulfatos de alcoholes grasos	Los sulfatos de alcoholes grasos, aumentan en importancia, especialmente combinados como cosurfactantes en diferentes formulaciones. Se espera que sustituyan a los jabones en Asia, lo que incrementaría su consumo y utilización en un futuro próximo.
FAES Etersulfatos de alcoholes grasos	Están experimentando una velocidad de crecimiento media en los últimos años del 4.5%. Se utilizan fundamentalmente para la fabricación de detergentes líquidos, champús y geles de baño.
FAEO Alcoholes grasos etoxilados	Experimentan un crecimiento medio anual del 4%, la razón fundamental es la sustitución de los alquilfenoles etoxilados por motivos ecológicos.

Ilustración 2. Altamajer (2004) Químicos utilizados para la elaboración de detergentes.

Ahora bien, el reto de la industria actual es resolver los problemas que trae los detergentes tradicionales los cuales tienen productos químicos que son agresivos con el medio ambiente, mediante la elaboración de nuevas fórmulas de fabricación que sean altamente biodegradables y que

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

sean de alta eficacia para el lavado de prendas, en sí que presenten altos estándares de calidad, a continuación, las ilustraciones 3 y 4 relacionan los productos químicos que más se utilizan en esta industria.

Nombre de la sustancia		Amenaza /Peligro
Alcalis y Ácidos		
Hidróxido potásico		Irritante
Hidróxido sódico		Origina corrosión
Ácido fosfórico		Irritante/Origina corrosión
Ácido nítrico		Irritante/Necrosante/Origina corrosión
Ácido clorhídrico		Problemas de respiración/ Irritante/Corrosivo
Acido sulfámico		Irritante
Secuestrantes		
Trifosfato y difosfato sódico		Eutrófico/Secuestra metales pesados
Ácido Nitrilo triacético		Eutrófico/Secuestra metales pesados
Tensioactivos		
Aniónicos	Alquilbenceno sulfonato Alquilsulfato Alquilétersulfato	Espumas/Biodegradabilidad
Catiónicos	Haluro de trialquilbenzilamonio	Incompatibilidad con aniónicos
No iónicos	Haluro de Alquilpiridinio	Presencia de cloro/Biodegradabilidad
	Alquilfenol oxietilenado	Productos tóxicos en la biodegradación

Ilustración 3. Altamajer (2004) Agentes químicos empleados para la limpieza en la industria.

Nombre de la Sustancia	
Alcalis y ácidos	
Guanidinas	
Compuestos fenólicos	
Halógenos y sus compuestos	Cloro Dióxido de cloro Compuestos generadores de cloro Yodo, Bromo y Flúor
Agentes oxidantes productores de oxígeno	Peróxido de hidrógeno Perácidos Percompuestos inorgánicos
Aldehidos	Formaldehído
Tensioactivos	Compuestos de amonio cuaternario Tensioactivos anfóteros

Ilustración 4. Altamajer (2004) Desinfectantes más empleados para la industria.

La información anterior puede ayudar a comprender el proceso químico con el cual se fabrican normalmente los jabones, los procesos químico-reactivos y la contaminación que por ende genera una vez estos productos se hayan utilizado en casa o en otros sectores de la sociedad.

Sin dejar atrás la importancia que juega la industria química en la economía, la idea del presente proyecto es poder relacionar técnicas modernas como son la utilización de materiales

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

biodegradables para la manufactura de detergentes que reduzcan la contaminación, debido a esto se buscara utilizar el Big Data y sus herramientas para relacionar el comportamiento de consumo y poder generar estrategias para elevar el consumo de dichos productos amigables con el medio ambiente.

Teniendo en cuenta dicha información es de suma importancia relacionar el análisis de grandes cantidades de datos con el mercadeo de detergentes donde se utilicen materiales biodegradables. Martínez (2017), menciona: “en la actual época digital y de un entorno cambiante, las organizaciones deben investigar más en los potenciales clientes y en los ya consumidores y sus preferencias, deben investigar mercados y conocer la competencia para poder ofrecer productos de mayor rendimiento económico, es decir, la información cada vez adquiere mayor relevancia y las compañías que conozcan bien a sus clientes podrán ofrecer productos de mayor éxito. El principal problema de las organizaciones es buscar la manera más adecuada de analizar los datos que sus clientes aportan. Los empresarios deben buscar la manera no solo de recopilar datos sino de tratar estos adecuadamente para obtener estadísticas y tendencias que nos puedan ayudar en la dirección indicada.”

Por consiguiente y dando sustento a la importancia de utilizar el Big Data en las organizaciones podemos enumerar algunas de las ventajas e inconvenientes para su implementación, según Martínez (2017), estas son algunas de ellas:

Ventajas:

- Permite mejorar la gestión del cambio.
- Genera un mayor acierto en la búsqueda de oportunidades de negocio.
- Aumenta la efectividad en la oferta de servicios a clientes (mejora de la segmentación).
- Incrementar la cartera de clientes a través del uso compartido de las redes sociales.
- Desarrollar técnicas de marketing digital más efectivas basadas en un mayor estudio del consumidor efectivo.
- Proporciona información en tiempo real sobre indicadores de negocio.
- Permite detectar fraudes.

Desventajas:

- Elevado coste de implantación si se quiere obtener la infraestructura estándar.

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

- Falta de infraestructura en pymes en general.
- El Big Data requiere de seguridad, si se roban los datos almacenados de nada serviría, la inversión en sistemas de seguridad no está al alcance de todas las pymes.
- Gastos de formación.
- Problemas de la privacidad.
- Ventajas de retraso. En ocasiones muchas empresas esperar a que sus competidores instalen estos programas para ver que errores ocurren y así ellos puedan posteriormente implementar esta tecnología aprendiendo de los errores de los competidores.

Teniendo en cuenta lo anterior, es importante resolver desde el ámbito organizacional qué importancia tiene generar este tipo de estrategia y si implementar análisis Big Data es de necesario para la empresa.

Por consiguiente, se adopta una postura frontal que permitan por medio de esta investigación plantar las bases para sustentar la importancia de implementar estudios de Big Data en la empresa, así como adaptar las vías acordes, que conduzcan a un método para poder generar estrategias para el comportamiento de compra, es así como se puede ver en la ilustración 5 y dando razón a esta investigación, que el análisis Big Data puede ayudar a darle un impulso a la marca de detergentes que usan materiales biodegradables:



Ilustración 5. Berengueras (2017). Áreas que se benefician del Big Data.

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

4.3 Sector Retail

Después de analizar los conceptos relacionados con Big Data y la industria de detergentes que utilizan materiales biodegradables, también es importante relacionar el comportamiento del sector retail² frente a la evolución del procesamiento de información, es así como las empresas que componen este sector almacenan grandes cantidades de información, sin embargo, no la utilizan para generar estrategias de mercado que les permitan obtener mayores beneficios.

Ahora bien, la tecnología no solo se usa en sectores como educación y electrónica, sino también en áreas estratégicas como sistemas y mercadeo, como lo menciona Sotelo (2019), su utilidad va más allá del desarrollo de aplicaciones, programas y demás actividades. En el retail brinda soporte para administrar la información, que es abundante y cambia cada día, es así como el uso del Big Data para este sector permite obtener la siguiente información:

- El total de tickets de venta que una cadena de tiendas con presencia en el ámbito nacional genera por día.
- Los productos que incluye la lista de compra de cada persona.
- La hora en que se compra cada producto con mayor frecuencia.
- Los artículos adicionales que adquieren los que van a comprar productos de aseo.

La información que brinda el Big Data es muy amplia, por lo tanto, las empresas que componen el sector retail pueden recolectarla de los clientes mediante otras plataformas distintas a las tiendas físicas, estos medios pueden ser digitales, como los Smartphone. El sector retail suele almacenar cantidades de datos que no utilizan, es por esto que el Big Data brinda soporte para utilizar estos datos de una manera adecuada, lo cual conlleva a innovar el negocio, interactuar con el cliente y mejorar su experiencia de compra.

Por otro lado, es importante enmarcar el auge de la tecnología en Colombia y como la transformación digital se ha convertido en un impulsor de nuevas formas de relacionarse con los clientes, poco a poco sectores tan diversos como la salud, los negocios, la educación y en este caso

² Sector retail: Sector económico que engloba a las empresas especializadas en la comercialización masiva de productos o servicios uniformes a grandes cantidades de clientes.

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

específico en el sector retail forjan caminos cruciales para el crecimiento de las empresas que utilizan el análisis de grandes cantidades de datos enfocados en los negocios, puesto que entre más interacción tengan las personas con la tecnología más fácil será la obtención de datos actualizados, que sean confiables para analizar, esto con el fin de tomar decisiones oportunas.

Dicho esto, se puede relacionar empresas que marcan una pauta importante en el sector, debido a que utilizan las nuevas tecnologías propias de la cuarta revolución industrial y que según Sánchez (2018), en su repositorio “Propuesta de un plan estructurado de información a las empresas y supermercados del retail, como redes de negocios desde el enfoque de las smartcities³ para Perú y Colombia”, algunas de las empresas que están marcando una pauta importante dentro del uso y aplicación del análisis Big Data son:

- Oster

Para el año 2016 la plataforma digital de Oster se convierte en una de sus prioridades, a través de esta los consumidores pueden encontrar desde un computador o un teléfono inteligente hasta una licuadora o lo último en tecnología para su cocina. Brinda una experiencia cómoda y segura para sus clientes donde tienen la posibilidad de reservar cualquier artículo en su e-Commerce⁴ retail, para después adquirirlo en cualquiera de sus puntos de venta, generando una presencia omnicanal⁵.

- Linio

En Colombia Linio es uno de los retailers e-Commerce que cuentan con más visitas por parte de los consumidores, ofreciendo en su plataforma cientos de ofertas en artículos de diversas categorías.

- Grupo Éxito

Su larga experiencia y trayectoria lo han posesionado como uno de los mejores e-Commerce del retail en Colombia. El retail opera bajo las marcas Éxito, Carulla, Surtimax, Superinter y Surtimayorista.

- Falabella

Otro de los grandes e-Commerce retail en Latinoamérica es Falabella, cuya tienda por departamentos es una de las más importantes en Suramérica, ya que cuenta con presencia en Chile,

³ Smartcities: Ciudades inteligentes.

⁴ E-commerce: Consiste en la distribución, venta, compra, marketing y suministro de información de productos o servicios a través de Internet.

⁵ Omnicanal: Intención de unificar todos los canales en los que está presente una marca o un negocio de tal manera que el cliente no aprecie diferencias entre todos ellos.

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

Argentina, Perú y por supuesto Colombia. En su gran catálogo se pueden encontrar desde electrodomésticos y tecnología hasta moda para toda la familia.

- Tiendas Jumbo

Uno de los más grandes e-Commerce retail en América Latina es Cencosud y en Colombia su cadena Tiendas Jumbo y en su tienda online los consumidores pueden encontrar tecnología, electrodomésticos, hogar y decoración, así como deportes y artículos para tiempo libre.

En definitiva, la aplicación de estas herramientas se constituye como un avance significativo para la aplicación de estrategias organizacionales, por ejemplo, el sector financiero ha sacado un gran provecho relacionando con actividades de seguridad y privacidad, así como también evitando el fraude electrónico y la evasión de impuestos. Según la empresa Grupo Bit (2020) las cifras comprueban el crecimiento de las empresas que usan Big Data, desde 2016 el crecimiento de este tipo de empresas está muy por encima del nivel del mercado, aproximadamente el 40%. Esto demuestra que las soluciones basadas en analítica de negocios están tomando cada vez más fuerza y relevancia.

4.4 Aplicación del Big Data en el mundo y sus herramientas

El uso del Big Data se ha convertido en una herramienta fundamental en la revolución tecnológica de diferentes sectores que buscan a través de la innovación, encontrar diferentes alternativas para el empleo eficiente de los datos, de esta manera gracias a la gran cantidad de información, los datos pueden ser estructurados de tal forma que las organizaciones puedan identificar problemas que sean mucho más fáciles de comprender (Mayor-Ríos et al. 2019)

Es por esto que la recopilación de datos, junto con la búsqueda de tendencias genera enormes posibilidades para que las empresas se puedan mover mucho más rápido y de una forma eficiente, anticipando posibles problemas que puedan causar inconvenientes y por consecuencia limitar los beneficios (Maldonado-Pinto y Portilla-Barco, 2020). Por eso es importante relacionar algunos de los usos más significativos y como se puede observar en la tabla 1, el Big Data está relacionado con diversos sectores de la sociedad.

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

#	Herramientas	Descripción
1	Apache Hadoop.	Una herramienta de las más utilizadas procesa grandes volúmenes de datos, con modelos de programación simple. Es escalable, por lo que puede pasar de operar en un sólo servidor a hacerlo en múltiples.
2	Apache Storm.	Puede ser usada con cualquier lenguaje de programación. Procesa en tiempo real y de forma sencilla grandes cantidades de datos, el sistema crea topologías de los macro datos para transformarlos y analizarlos de forma continua mientras flujos de información entran al sistema constantemente.
3	MongoDB.	Base de datos NoSQL optimizada para trabajar con grupos de datos que varían con frecuencia, o que son semiestructurados o inestructurados. Se emplea para almacenar datos de aplicaciones móviles y de sistemas de gestión de contenidos,
4	Apache Spark.	Se destaca por su velocidad, siendo 100 veces más rápida que Hadoop. Spark analiza datos por lotes y también en tiempo real, y permite la creación de aplicaciones en diferentes lenguajes: Java, Python, R y Scala.
5	Python.	Python cuenta con la ventaja de que sólo hay que tener unos conocimientos mínimos de informática para poder usarla, lo hace que tenga una gran comunidad de usuarios con la opción de crear sus propias librerías.
6	Lenguaje R.	Es un entorno y lenguaje de programación enfocado mayormente al análisis estadístico, ya que es muy parecido al lenguaje matemático.

Tabla 1. Adaptado de Flores & Villacís (2017). Herramientas para análisis Big Data.

Con el auge y el continuo crecimiento de la información digital, las empresas se encuentran en la necesidad de recopilar, gestionar y analizar grandes cantidades de datos que se encuentran en la red, una vez transformados en información, se pueden utilizar para determinar el comportamiento de los clientes y generar decisiones estratégicas. Todo este proceso exige que haya una serie de profesionales con diferentes aptitudes, competencias y habilidades, que sean acordes para utilizar las herramientas y obtener la enorme cantidad de datos disponibles y convertirlos en conocimiento útil. Para el análisis de los datos se requiere tener claro cuál es el perfil profesional de las personas que intervienen en el proceso del Análisis Big data; los perfiles con mayor demanda están

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

relacionados a continuación en la tabla 2.

#	Profesión	Perfil ocupacional del analista de Big Data
1	Científico de datos.	Son los miembros clave del equipo de ciencia de datos. Permiten extraer conocimiento e información valiosa de los datos. Tienen visión general del proceso de extremo a extremo y pueden resolver problemas de ciencias datos, la construcción de modelos analíticos y algoritmos.
2	Analista de datos.	Participa en las iniciativas y proyectos de análisis de datos. Es la persona que recoge las necesidades de los usuarios de negocio para los científicos de datos y presenta resultados obtenidos.
3	Arquitecto de datos.	Son los responsables de diseñar, definir la estructura y mantener los datos, garantizando su seguridad, calidad, fiabilidad y accesibilidad.
4	Ingeniero de Datos.	Se encarga de proporcionar los datos de una manera accesible y apropiada a los usuarios. Es un perfil especializado en infraestructura Big Data. Desarrolla y explota técnicas, procesos, herramientas y métodos que deben servir para el desarrollo de aplicaciones Big Data.
5	Administrador de Base de Datos.	Es responsable de mantener la calidad, disponibilidad y seguridad de los datos. Persigue mejorar el almacenamiento y presentación de los datos en toda la empresa.
6	Estadístico.	Su labor consistirá en obtener, analizar y e interpretar datos cualitativos y cuantitativos utilizando diferentes métodos estadísticos (medias, histogramas, desviaciones, regresiones, test-X, test-Z).
7	Analista de negocio.	Son expertos responsables de crear los gráficos, infografías y otras herramientas visuales para ayudar a las diferentes personas de la organización a comprender datos complejos, además de generar la visión del negocio desde la información obtenida en proyectos Big Data.

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

Tabla 2. Adaptado de Rayo (2016). Profesión y perfil ocupacional del Analista de Big Data.

5. METODOLOGÍA

La propuesta metodológica del presente estudio se basa en la consultoría desarrollada en una empresa del sector retail, que fabrica y comercializa productos de aseo biodegradables. Con la cual se desarrollará un modelo de análisis de datos que le permita conocer las tendencias y el comportamiento del consumidor y de esta manera tomar decisiones al momento de implementar estrategias de marketing, con las que se pueda posicionar en el mercado.

La investigación desarrollada es de tipo descriptivo y exploratorio en una primera etapa donde se recolecta la información necesaria propia de la investigación, posteriormente, habrá una etapa con un enfoque cuantitativo, las técnicas que se van a realizar para la recolección de información por medio de una base de datos en Excel extraída del software PSL y su dimensión es de 16 variables y 39.026 filas con información estructurada, donde se relacionan datos correspondientes a ventas realizadas entre el 27 de junio de 2018 hasta el 31 de julio de 2020.

El enfoque de la investigación es descriptivo dado que la finalidad es generar estrategias de marketing para productos detergentes que utilicen materias primas amigables con el medio ambiente. Para esto se lleva un proceso secuencial basado en la investigación en un periodo de tiempo determinado, en el cual se hará seguimiento al problema planteado analizando la realidad de forma objetiva.

Para llevar a cabo el estudio de las de la base de datos se utilizará la metodología CRISP-DM, la cual se describirá a continuación y sustenta la base para el estudio de Big Data que se plantea realizar.

5.1 Metodología CRISP-DM

Para el análisis de las bases de datos se implementará la metodología CRISP-DM que se utiliza para obtener mejores y más rápidos resultados en la minería de datos, esta metodología presentada en la ilustración 6, se compone de seis fases, las cuales no requieren una secuencia lineal dado que es posible revisar total o parcialmente las anteriores (Gallardo, 2009).

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

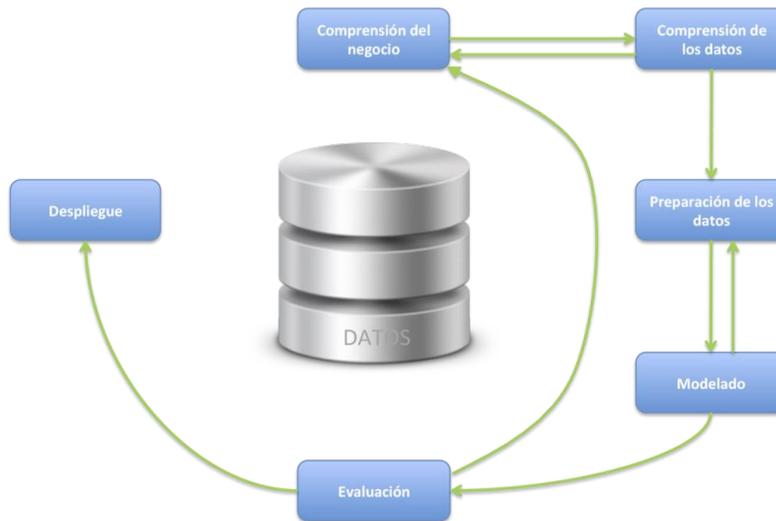


Ilustración 6. Calvo (2016) Modelo CRISP-DM Data Mining.

Según Gallardo (2009) las etapas en las que se divide el modelo son las siguientes:

1. Comprensión del negocio.
2. Comprensión de los datos.
3. Preparación de los datos.
4. Modelado.
5. Evaluación.
6. Implementación.

Teniendo en cuenta las etapas mencionadas, se sintetiza cada una de ellas ajustándolas directamente en la metodología que se pretende realizar y al objetivo propuesto para la investigación. En la etapa inicial del proyecto, el propósito general es tener claridad sobre los objetivos y requerimientos del planteamiento del problema, a partir del conocimiento de la información se procede a determinar las variables de interés y en consecuencia identificar el problema de minería de datos, por consiguiente se deben reunir una cantidad de datos para adaptarse y relacionarse, teniendo muy presente la base fundamental que son los objetivos del negocio y de los subprocesos que dependen de ella como son la recolección, descripción, exploración y verificación de los datos.

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

Después a esto vienen dos momentos cruciales, una vez realizadas las etapas anteriores, se procede a acondicionar los datos para adaptarlos a las técnicas de Data Mining que se utilicen en las etapas posteriores, este proceso comprende desde la selección y limpieza de los datos, la generación de variables adicionales, la integración de los diferentes orígenes de los datos y los posibles cambios en los formatos que pueda presentar la información, todo esto seguido del modelado donde se escogen las técnicas más convenientes, en esta fase se relacionan tareas como selección de la técnica, diseño de las pruebas, construcción y evaluación del modelo, todo esto de acuerdo a los criterios anteriormente establecidos.

Posteriormente se evaluará el modelo CRISP DM; aquí se determinan las herramientas para examinar los resultados, si son aplicables y si estos proporcionan resultados que afecten de manera positiva el objetivo principal propuesto en la investigación, seguidos de una etapa de revisión donde se calificara la evaluación, esto permite determinar si se puede avanzar a una fase final que es la implementación, donde será imprescindible que los datos previamente evaluados cumplan con la función de brindar información de conocimiento, obtenido claro está, dentro del horizonte del negocio, aquí se plantean las estrategias propias y objetivas apoyadas por un debido control y seguimiento, donde se garantice una óptima aplicación del modelo, se deben establecer parámetros y coordinación acertada entre todos los actores involucrados, basados en la organización, los recursos usados y las fuentes de la propuesta.

Para continuar con el desarrollo de la investigación y teniendo en cuenta las herramientas disponibles para hacer análisis de datos, se propone utilizar el programa Excel versión 2016, el cual dentro de sus aplicaciones brinda la opción de utilizar el método de regresión lineal⁶ que genera la ecuación ($y = mx + b$) cuyos resultados posibilitan hacer pronósticos en una línea de tiempo. Este método se puede utilizar teniendo en cuenta factores como volumen, variabilidad y velocidad de los datos, en este sentido la hoja de cálculo brinda la posibilidad de hacer Big Data con bases de datos que en general no tengan un peso significativo y que sobre pase las capacidades del programa, adicional posibilita que personas que no son expertas en manejo de Big Data puedan generar este tipo de estudios, por esto y como lo menciona IGI Global (2020), el software de hoja de cálculo proporciona un medio para que quienes no son programadores realicen análisis que antes solo podía realizar un programador o analista capacitado y agrega, la investigación demuestra que Excel (especialmente cuando se usan sus complementos) puede cumplir la misma función en el análisis de

⁶ Método de regresión lineal: Permite hallar el valor esperado de una variable aleatoria A cuando B toma un valor específico; cuando la demanda presenta un comportamiento creciente o decreciente.

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

Big Data. El análisis de hoja de cálculo puede proporcionar información valiosa sobre qué análisis avanzados son apropiados.

6. RESULTADOS

A continuación, se relaciona en la tabla 3, la información clasificada de la base datos suministrada por la empresa Sello Global, extraída del software PSL y presentada en formato Excel, la cual cuenta con una dimensión de 16 variables, 9 numéricas, 6 alfabéticas, 1 binaria y 39.026 filas, adicional se relaciona el nombre de la variable, la descripción y su correspondiente acrónimo.

Dimensión: La base de datos cuenta con 16 variables y 39.026 filas				
#	Nombre variable	Descripción variable	Acrónimo	Tipo de variable
1	Movimiento año	Movimiento del producto en un año específico	movano	Numérica
2	Movimiento mes	Movimiento del producto en un mes específico	movmes	Numérica
3	Movimiento día	Movimiento del producto en un día específico	movdia	Numérica
4	Código ítem	Código interno del producto	codiitem	Numérica
5	Descripción corta ítem	Nombre corto del ítem o producto	itedescort	Alfabética
6	Consecutivo de factura	Número consecutivo de factura	consfact	Numérica
7	Movimiento cantidad ítem	Cantidad facturada de un producto	movcantump	Numérica
8	Precio unitario	Precio unitario del producto	precunit	Numérica
9	Precio total bruto	Cantidad facturada por precio unitario= precio total bruto	prectotabrut	Binaria
10	Tipo identificación	Tipo de identificación persona natural o jurídica	tipo_iden	Alfabética
11	Identificación del tercero	Número de identificación persona natural o jurídica	idtercero	Numérica
12	Tercero nombre completo	Nombre completo del cliente	ternombcomp	Alfabética
13	Departamento	Departamento	depto	Alfabética
14	Ciudad	Ciudad	ciudad	Alfabética
15	Ítem grupo	Grupo al que pertenece el ítem	itegrupo	Numérica
16	Grupo nombre	Nombre del grupo al que pertenece ítem	gitnombre	Alfabética

Tabla 3. Fuente Elaboración propia. Descripción variables base de datos.

Luego de consolidar la información sobre el proceso de análisis, la cual se encuentra relacionada en el contenido de esta investigación, se construye una base de datos estructurada con información que

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

ha sido clasificada como relevante para determinar de esta manera los siguientes interrogantes que ayudaran a dar mayor claridad a los objetivos planteados para esta tesis.

- ¿Cuál son los productos más vendido y como ha sido su comportamiento en el periodo estudiado?
- ¿Cuáles son los productos de menor rotación?
- El portafolio de productos de la empresa se compone de tres grupos, aseo, lavandería y cuidado personal, ¿cuál de estos es el más vendido y en qué año tuvo su mayor representación en ventas?
- ¿Cuál cliente realiza más compras en un determinado periodo de tiempo?
- ¿Teniendo en cuenta la proyección de ventas de limpia pisos citronela, para cuantas unidades de producto terminado se debe aprovisionar de materias primas en el mes de diciembre de 2020?
- ¿Cuál es la proyección de venta del suavizante para el mes de enero del 2021?

Como resultado a la primera pregunta planteada, se relaciona la información correspondiente al análisis de los productos más vendidos y su comportamiento en los periodos de tiempo en los cuales se recolectaron los datos, de esta manera se destaca el top 15 de los productos con mayor representación, presentados en la tabla 4.

		Año	Año	Año	
	Descripción corta del ítem	2018	2019	2020	Total general
1	LIMPIAPISOS CITRONELA CAJA X12	0,00%	2,08%	26,19%	28,28%
2	PRELAVADO.1000ML	1,32%	2,09%	1,01%	4,41%
3	SUAVIZANTE.1000ML	1,68%	1,80%	0,92%	4,41%
4	DETERGENTE FLORAL 1000ML	1,72%	1,68%	0,73%	4,13%
5	MULTIUSOS.1000ML	1,23%	1,61%	1,17%	4,00%
6	DESINFECTANTE.1000ML	0,64%	0,78%	2,53%	3,95%
7	PRELAVADO.500ML	1,22%	1,58%	0,56%	3,36%
8	DESENGRASANTE.1000ML	0,80%	1,48%	0,92%	3,20%
9	DETERGENTE FLORAL 5000ML	0,56%	1,14%	0,89%	2,59%
10	DESMANCHADOR SUPERFICIES1000M	0,76%	1,07%	0,62%	2,45%
11	DESENGRASANTE.500ML	0,94%	1,03%	0,45%	2,42%
12	MULTIUSOS.500ML	0,95%	0,92%	0,47%	2,34%
13	SUAVIZANTE.5000ML	0,47%	0,94%	0,70%	2,11%

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

14	DESMANCHADOR SUPERFICIES.500ML	0,88%	0,88%	0,35%	2,11%
15	MULTIUSOS.5000ML	0,29%	0,78%	0,83%	1,90%

Tabla 4. Fuente Elaboración propia. Análisis productos más vendidos.

De igual manera, en la tabla 5 se relaciona el top 15 de los productos con menor rotación y su correspondiente análisis, el cual se representa en porcentaje ordenado de manera ascendente y el cual se puede observar a continuación.

	Descripción corta del ítem	Porcentaje de participación en las ventas
1	ALCOHOL ETILICO SOLU70% 5000ML	0,01%
2	LIMPIA VIDRIOS.20LTS	0,02%
3	AMBIENTADOR HERBAL.5000ML	0,02%
4	PISOS CANELA 20LTS	0,03%
5	PISOS EXPLOSION FRUTAL 20LTS	0,03%
6	AMBIENTADOR FRUTAL.5000ML	0,04%
7	PISOS PINO.20LTS	0,06%
8	DESMANCHADOR SUPERFICIES.20LTS	0,07%
9	DETERGENTE HERBAL 20LTS	0,08%
10	PISOS LAVANDA.20LTS	0,09%
11	MULTIUSOS SIN OC.500ML	0,15%
12	LIMPIA VIDRIOS.5000ML	0,16%
13	MULTIUSOS SIN OC. 20LTS	0,16%
14	JABON MANOS HERBAL.5000ML	0,21%
15	MULTIUSOS SIN OC.1000ML	0,22%

Tabla 5. Fuente Elaboración propia. Análisis productos menos vendidos.

En la tabla 6 se puede evidenciar el portafolio de productos de la empresa y su composición en porcentaje de participación en ventas en los periodos de tiempo evaluados.

		Año	Año	Año	
	Grupo	2018	2019	2020	Total general
1	ASEO	11,93%	18,93%	41,77%	72,63%
2	LAVANDERIA	7,50%	11,53%	6,31%	25,34%
3	CUIDADO PERSONAL	0,55%	0,69%	0,79%	2,03%
	Total general	19,97%	31,15%	48,88%	100,00%

Tabla 6. Fuente Elaboración propia. Análisis grupo más vendido.

Según los resultados obtenidos para los grupos de productos con mayor volumen en ventas entre el periodo 2018 a 2020 donde se evidencia un 72.6 % en la participación correspondiente al grupo de

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

aseo y donde se puede constatar que ha mantenido un aumento considerable en las ventas al periodo de 2020 en un 41.7%, lo cual constituye como el grupo de productos de mayor relevancia para la empresa, por lo tanto se define como estrategia la expansión de este grupo de productos donde se potencialice la apertura de nuevos convenios a nivel nacional como es la venta en grandes superficies y aprovechar el éxito en las ventas que ha mostrado en los periodos de tiempo previamente analizados para potencializar la distribución, generar reconocimiento de marca y aumentar porcentaje de ventas en un 20% en los próximos 2 años.

A continuación, en la tabla 7 se puede evidenciar el top 15 de los clientes que más han comprado en los periodos de tiempo a evaluar y entre los cuales están personas naturales y empresas, brindando como resultado su respectivo porcentaje de participación.

		Año	Año	Año	
	Cientes más representativos	2018	2019	2020	Total general
1	MERCADERIA SAS	0,00%	2,08%	26,19%	28,27%
2	ARANGO CRUZ CARLOS ENRIQUE	2,19%	2,76%	3,69%	8,63%
3	MARIN GUERRA EDGAR YOVANNY	1,08%	1,57%	1,19%	3,84%
4	LLANOS MARTINEZ ZULMA AIDA	1,07%	1,15%	0,80%	3,01%
5	RENDON CASTANO CIELO CATALINA	0,93%	1,79%	0,25%	2,97%
6	CAÑAS MUÑOZ LINA MARCELA	0,42%	1,31%	1,22%	2,95%
7	LOZANO CORDERO JORGE LUIS	0,87%	1,41%	0,34%	2,62%
8	HINESTROZA VELASQUEZ DANIELA	0,21%	1,37%	0,99%	2,56%
9	SANCHEZ CASTRO MAGRED	0,00%	1,27%	1,18%	2,45%
10	ECO SAN ANDRES ISLA SAS	0,00%	0,95%	0,89%	1,84%
11	MONTOYA ZAPATA JUAN PABLO	0,60%	0,78%	0,18%	1,56%
12	TABARES SUAREZ DORIS ELENA	0,39%	0,82%	0,33%	1,54%
13	VELASQUEZ ARANGO LINA CLEMENCIA	0,41%	0,60%	0,52%	1,53%
14	ECOINICIATIVAS S.A.S	0,88%	0,27%	0,29%	1,45%
15	MUÑOZ CARDONA MARTHA LUCIA	0,00%	1,32%	0,00%	1,32%

Tabla 7. Fuente Elaboración propia. Análisis de clientes con mayor participación en compras.

En la ilustración 7 se observa la tendencia del producto con mayor participación, donde se evidencia el comportamiento de las ventas en el periodo comprendido entre septiembre de 2019 el cual se denomina mes 1 y diciembre de 2020 el cual se denomina mes 11.

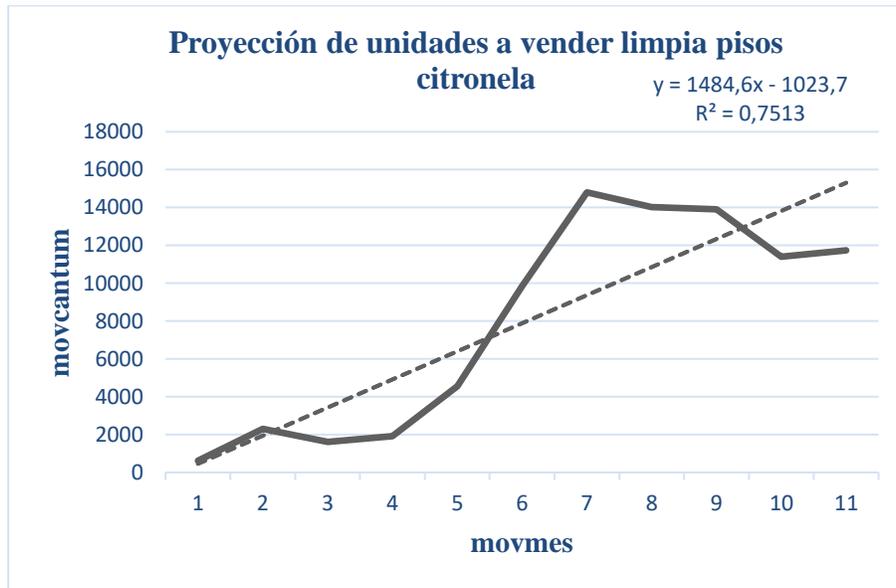


Ilustración 7. Elaboración propia. Proyección de ventas limpia pisos citronela.

La ecuación 1 sirve para determinar las cantidades vendidas en diciembre de 2020, determinado en este caso por el periodo 16.

Coefficientes

Intercepción -1023,58577
 movmes 1484,594466

Estos coeficientes resultan del análisis estadístico realizado, los cuales se utilizan para reemplazar en la ecuación que brinda el pronóstico del modelo de regresión lineal.

$$\text{Cantidades vendidas} = 1484.6 * \text{mes a proyectar} - 1023.7 \text{ Ec(1)}$$

$$\text{Cantidades vendidas} = 1484.6 * 16 - 1023.7$$

$$\text{Cantidades vendidas} = 22730 \text{ unidades}$$

Estadísticas de la regresión

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

Coefficiente de correlación múltiple 0,8667

Coefficiente de determinación R² 0,7512

Este resultado indica que existe un 86,67% de relación entre la cantidad de unidades vendidas en un periodo determinado y además existe una confianza (R²) del 75,12% en que el modelo de pronóstico es acertado, de esta manera se puede decir que según la ecuación planteada para el mes de diciembre de 2020 se aprovisionar de materia prima para un estimado de ventas de 22730 unidades.

Análisis de varianza

Regresión

F 583,5487097

Valor crítico de F 8,33559E-17

En este caso se puede evidenciar que el valor de F es mucho mayor que el valor crítico de F lo que significa que se poseen datos suficientes para decir que el modelo de regresión lineal es acertado, y que el modelo de regresión establecido es estadísticamente significativo.

En la ilustración 8 se observa la tendencia de venta del producto suavizante, este ocupa el tercer lugar entre los más vendidos y su respectivo comportamiento en el periodo comprendido entre junio de 2018 el cual se denomina mes 1 y julio de 2020 el cual se denomina mes 26.

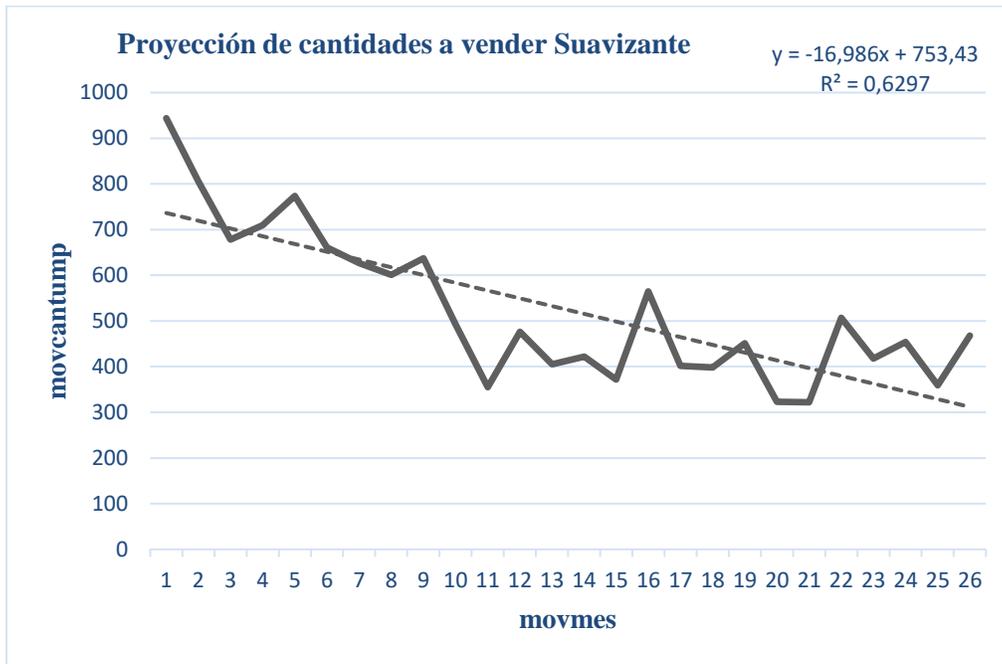


Ilustración 8. Elaboración propia. Proyección de ventas Suavizante.

La ecuación 2 sirve para determinar las cantidades vendidas en enero de 2021, determinado en este caso por el periodo 27.

Coefficientes

Intercepción 753,4307692
 movmes -16,98632479

Estos coeficientes resultan del análisis estadístico hecho en Excel, los cuales se utilizan para reemplazar en la ecuación que brinda el pronóstico del modelo de regresión lineal.

$$\text{Cantidades vendidas} = -16.986 * \text{mes a proyectar} + 753.43 \quad \text{Ec(2)}$$

$$\text{Cantidades vendidas} = -16.986 * 27 + 753.43$$

$$\text{Cantidades vendidas} = 295 \text{ unidades}$$

Estadísticas de la regresión

Coefficiente de correlación múltiple 0,7935378
 Coeficiente de determinación R² 0,62970224

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

Este resultado indica que existe un 79% de relación entre la cantidad de unidades vendidas en un periodo determinado y además existe una confianza (R^2) del 62% en que el modelo de pronóstico es acertado, de esta manera se puede decir que según la ecuación planteada para el mes de enero de 2021 se estima una cantidad de ventas de 295 unidades.

Análisis de varianza

Regresión

F 40,81270632

Valor crítico de F 1,32129E-06

En este caso se puede evidenciar que el valor de F es mucho mayor que el valor crítico de F lo que significa que se poseen datos suficientes para decir que el modelo de regresión lineal es acertado.

6.1 Estrategias.

- El top de los 15 productos más vendidos de la compañía entre el Junio del 2018 y agosto de 2020, revela que el producto líder limpia pisos citronela tiene una participación significativamente alta, superando a los demás productos en un 24% en las ventas reportadas durante el periodo mencionado, adicionalmente se evidencia una tendencia ascendente muy marcada en el consumo de este limpia pisos, por esto, y con el fin de mantener cubierta la demanda se propone como estrategia mantener un aprovisionamiento de las materias primas que asegure tener el producto disponible en caso de una de sobredemanda, como también contar con aliados estratégicos que provean los insumos de calidad y en el menor tiempo posible.
- Realizar un análisis para saber la razón por la cual el producto no es tan rotativo como los demás y desarrollar formas y medios de comercialización como promociones de venta para generar compra y recompra de parte de los clientes. Otra de las estrategias que se puede implementar es la venta on line, es una herramienta que puede aportar al incremento de las ventas y por ende abrir nuevos mercados que nos garantice posicionar la marca.

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

- Según los resultados obtenidos para los grupos de productos con mayor volumen en ventas entre el periodo 2018 a 2020 donde se evidencia un 72.6 % en la participación correspondiente al grupo de aseo y donde se puede constatar que ha mantenido un aumento considerable en las ventas al periodo de 2020 en un 41.7%, lo cual constituye como el grupo de productos de mayor relevancia para la empresa, por lo tanto se define como estrategia la expansión de este grupo de productos donde se potencialice la apertura de nuevos convenios a nivel nacional como es la venta en grandes superficies y aprovechar el éxito en las ventas que ha mostrado en los periodos de tiempo previamente analizados para potencializar la distribución, generar reconocimiento de marca y aumentar porcentaje de ventas en un 20% en los próximos 2 años.
- Según la información obtenida del análisis de ventas de suavizante, donde se evidencia que a pesar de ser el tercer producto con mayor porcentaje de participación, también se puede observar que en el periodo evaluado y proyectado hay una tendencia a bajar en su volumen de unidades vendidas pasando de 944 unidades en el mes junio de 2018 a una proyección para el mes de enero de 2021 de 295 unidades, por lo tanto se considera como estrategia potencializar este producto en el mercado fijando como meta un aumento en el porcentaje de la ventas 10% para el periodo de junio de 2021, promoviendo la venta de suavizante a través del grupo de aseo que cuenta con un gran éxito en ventas por medio de paquetes promocionales.

7. CONCLUSIONES Y RECOMENDACIONES

A partir del uso de la hoja de análisis de Excel y después de hacer una limpieza de los datos suministrados por la empresa sello global, evaluando diferentes variables donde las unidades vendidas representan el insumo principal y utilizando la herramienta de regresión lineal se puede hacer un pronóstico de unidades vendidas en determinado periodo de tiempo, con lo cual el resultado permite formular las estrategias que ayudan a anticipar el comportamiento de compra de los consumidores.

El conocimiento de la información que el negocio suminstras se constituye como un factor de suma importancia cuando se pretende hacer un análisis de datos, pues tener un conocimiento amplio sobre

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

la forma como los datos interactúan dentro de la organización y la forma como se almacenan permiten desarrollar de manera adecuada este tipo de proyectos.

La calidad de la información en referencia a la veracidad, permiten desarrollar un análisis de datos mucho más acertado, por lo tanto, es de suma importancia la selección de las variables que se estudian, el volumen de los datos y su relevancia para la investigación relacionadas directamente con los resultados que se desean obtener.

Se recomienda que las compañías que desean hacer análisis BD y no cuentan recursos económicos y técnicos para invertir en software especializado para hacer este tipo de análisis, también lo pueden hacer por medio de las herramientas de análisis de datos de la hoja de cálculo de Excel, pues en el estudio que se llevó a cabo en esta investigación se pudo evidenciar que el método de regresión lineal se puede aplicar como método de proyección para avaluar en este caso un pronóstico de ventas.

Se sugiere que el recurso humano que interviene en este tipo de estudios posea un conocimiento amplio sobre el funcionamiento de las áreas y procesos de la empresa, pues esto ayuda a tener una mayor visión de la información que se está evaluando, aportando al personal técnico especializado un análisis de los resultados con mejor claridad.

REFERENCIAS

Aguilar, A. (5 de octubre de 2016). Innovacion. Qué es Big Data y como funciona | Tecnología emprendimiento | social media. Obtenido de archivo de video: recuperado de <https://www.youtube.com/watch?v=2buYs2GVpGE>

Altmajer, D. (2004). <https://hera.ugr.es/tesisugr/15847093.pdf>. Obtenido de <https://hera.ugr.es/tesisugr/15847093.pdf>

Berengueras, J. (2017). Big Data analytics y sus aplicaciones en marketing. Obtenido de <http://www.ocstemsolutions.com/wp-content/uploads/2017/05/Big-data-analytics-y-sus-aplicaciones-en-marketing.pdf>

Calvo, D. (2016). Modelo CRISP DM – Data mining. Obtenido de <http://www.diegocalvo.es/modelo-crisp-drm-data-mining/>

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

Digital, P. (22 de marzo de 2016). ¿Qué es Big Data? Obtenido de archivo de video: recuperado de <https://www.youtube.com/watch?v=3AjaLbt6uUU>

Fundeau, BBVA. (2013). Macrodatos e inteligencia de datos, alternativas a big data. Obtenido de Fundacion Fundeau BBVA https://www.fundeau.es/recomendacion/macrodatosalternativa-abig-data-1582/?_ga=2.126823502.2128467

Gallardo, A. (2009). Metodología para la definición de requisitos en proyectos de data mining (ER-DM). pág. 317.
http://oa.upm.es/1946/1/JOSE_ALBERTO_GALLARDO_ARANCIBIA.pdf

Garcia, C. M. (1994). Tecnología de la información y ventaja competitiva: algunos ejemplos
Cristina Madorran Garcia Katrin Simon elorz. Obtenido de https://www.researchgate.net/profile/Katrin_SimonElorz/publication/28277206_Tecnologia_de_la_informacion_y_ventaja_competitiva_algunos_ejemplos/links/553c06d50cf2c415bb0b174a/Tecnologia-de-la-informacion-y-ventaja-competitiva-algunos-ejemplos.pdf

Grupo Bit. (2020). Cual es el panorama de la analítica de datos en Colombia . Obtenido de <https://business-intelligence.grupobit.net/blog/cual-es-el-panorama-de-la-analitica-de-negocios-en-colombia>

IGI Global. (2020). Uso de Excel y Excel VBA para el análisis preliminar en la investigación de Big Data. Obtenido de <https://www.igi-global.com/chapter/using-excel-and-excel-vba-for-preliminary-analysis-in-big-data-research/224335>

Martinez, R. (2017). La Big Data y el Marketing. Obtenido de <http://dspace.umh.es/bitstream/11000/4067/1/TFG%20Martinez%20Montesinos%2C%20Riveran%20Jos%C3%A9.pdf>

Maldonado-Pinto, J. E., & Portilla-Barco, L. F. (2020). Procesos de innovación en la industria manufacturera colombiana. *Revista CEA*, 6(11), 145-163.
<https://doi.org/10.22430/24223182.1395>

Mate, C. (2014). Big data. Un nuevo paradigma de análisis de datos. Obtenido de <https://www.iit.comillas.edu/docs/IIT-14-153A.pdf>

Mayor-Ríos, J. A., Pacheco-Ortiz, D. M., Patiño-Vanegas, J. C. & Ramos-y-Yovera, S. E. (2019). Análisis de la integración del Big Data en los programas de contaduría pública en universidades acreditadas en Colombia. *Revista CEA*, 5(9), 53-76.
<https://doi.org/10.22430/24223182.1257>

Mendez, J. (17 de junio de 2016). Que es Insight. Obtenido de archivo de video: recuperado de <https://www.youtube.com/watch?v=YxdLk5DF7xU>

	INFORME FINAL DE TRABAJO DE GRADO - FCEyAdm	Código	FDE 169
		Versión	02
		Fecha	30-07-2019

- Mendoza, M. (2011). Minería de datos en la web. Obtenido de https://www.researchgate.net/publication/229068670_Mineria_de_datos_en_la_Web/link/0fcfd4ffed47cb57b9000000/download.
- Merodio, J. (22 de febrero de 2018). 4 Ventajas del Big Data analytics en tiendas online. Obtenido de archivo de video: recuperado de <https://www.youtube.com/watch?v=G1zlkRnVIDc>
- Pérez, M. (2015). Big Data Técnicas, herramientas y Aplicaciones . Obtenido de http://rclibros.es/wp-content/uploads/2015/04/capitulo_9788494305559.pdf
- Pujadas, C. (2017). El Consumo Sostenible: concepto y relevancia para los países. Obtenido de <https://www.austral.edu.ar/eedu/wp-content/uploads/2017/05/Consumo-Sostenible-concepto-y-relevancia-para-LATAM.pdf>
- Rayo, A. (2016). Las habilidades y el perfil de un profesional Big Data. Obtenido de <https://www.bit.es/knowledge-center/habilidades-y-perfil-de-un-profesional-big-data/>
- Sanchez, E. (2018). Redes de negocios en la logistica del retail desde el enfoque de las smarcities en Peru y Colombia. Obtenido de <http://polux.unipiloto.edu.co:8080/00004726.pdf>
- Sotelo, V. (2019). Cómo usar el big data en el sector retail. Obtenido de <https://www.esan.edu.pe/conexion/actualidad/2019/05/13/como-usar-el-big-data-en-el-sector-retail/#:~:text=Informaci%C3%B3n%20abundante&text=Seg%C3%BAn%20Power%20Data%20el%20uso,la%20compra%20de%20cada%20persona>
- Universidad de Alcalá .(2020). 10 jemplos reales del uso del Big Data. Obtenido de <https://www.master-bigdata.com/10-usos-reales-big-data/>
- Villacís, P. F. (2017). Análisis comparativo de las herramientas de Big Data en la Facultad de Ingeniería de la Pontificia Universidad Católica del Ecuador. Obtenido de <http://repositorio.puce.edu.ec/bitstream/handle/22000/14119/DisertacionPaoAndreFinal.pdf?sequence=1&isAllowed=y>
- Zaforas, M. (22 de marzo de 2016). ¿Qué es Big Data? Obtenido de archivo de video: recuperado de <https://www.youtube.com/watch?v=3AjaLBt6uUU>