

**TRABAJO DE GRADO**  
**Opción Seminario-Diplomado.**

**Análisis de Datos de Consumidores con CRISP-DM: Estrategias para la Toma de Decisiones y el Crecimiento de Ventas.**

Corporación Universitaria Remington.  
Ciencias Empresariales.  
Administración de Empresas.

Sergio Sánchez Ríos.<sup>1</sup>  
Verónica Jhoanna Cárdenas Henao.<sup>2</sup>  
Ivonne Castaño Osorio.<sup>3</sup>  
Opción de Trabajo de grado Seminario-Diplomado.  
2025.

---

<sup>1</sup> Estudiante de noveno semestre de Administración de Empresas CUR Pereira. E-mail: sergio.sanchez.1096@miremington.edu.co

<sup>2</sup> Estudiante de noveno semestre de Administración de Empresas CUR Pereira. E-mail: veronica.cardenas.9558@miremington.edu.co

<sup>3</sup> Asesora Temática Seminario en Big Data y Ciencia de Datos sede Pereira.

## Tabla de Contenido

Resumen.....	3
Palabra clave.....	3
Pregunta orientadora de la búsqueda.....	6
Metodología de búsqueda de la información.....	7
Sustentación teórica de la pregunta.....	8
Conclusiones.....	11
Referencias.....	172
Anexos.....	¡Error!
<b>Marcador no definido.</b>	<b>3</b>

## Resumen

Este informe tiene como objetivo ayudar al mejoramiento de las estrategias con la ayuda óptima de la ciencia de datos, buscando optimizar la toma de decisiones y fomentar el crecimiento de ventas en empresas mediante la aplicación de la metodología CRISP-DM. En este caso se realizarán los diferentes análisis a la empresa EasyTech la cual presenta un estancamiento en sus ventas, buscando diferentes estrategias basadas en el análisis de datos de consumidores. Se puede evidenciar que, a través de la ciencia de datos, la cual combina estadística y programación, se llegan a abordar mediante el análisis de datos diferentes problemáticas que se presentan en las empresas, por lo que, por medio de esto, podemos utilizar los datos para lograr el mejoramiento en la segmentación de clientes y en la personalización de estrategias de venta, incrementando potencialmente los ingresos de la empresa. Mediante el desarrollo de modelos descriptivos, este informe se enfocará en buscar soluciones óptimas a las problemáticas presentes en la empresa, logrando de manera eficiente el desarrollo de nuevas estrategias que permitan mejorar las ventas y ayudar en la toma de decisiones.

**Palabras clave:** Ciencia de Datos, CRISP-DM, Toma de Decisiones, Estrategias de Ventas, Modelos Descriptivos.

### **Pregunta orientadora de la búsqueda**

La ciencia de datos es una disciplina que apoya a las empresas con la toma de decisiones. Esto lo hace tomando una gran cantidad de datos y convirtiéndolos en información valiosa que podrá ser analizada y estudiada por la empresa o el analista de datos para de esta forma poder tener una toma de decisiones basada en información, en datos históricos, en predicciones, tendencias y proyecciones. Gracias a la ciencia de datos al analizar las variables relevantes en la información que tiene cada empresa sobre ventas, clientes, productos y otras variables se pueden tomar mejores decisiones y se puede tener una ventaja competitiva sobre la competencia, además se pueden aprovechar oportunidades que de otro modo podrían pasar desapercibidas.

Según Provost & Fawcett, (2013), la ciencia de datos combina técnicas estadísticas y computacionales para transformar datos en acciones estratégicas, permitiendo a las organizaciones tomar decisiones basadas en evidencia (p. 15). Así que, por medio de la información obtenida de grandes volúmenes de datos, se puedan abordar diferentes problemas que se presentan en las empresas, como la disminución de ventas, la falta de retención de clientes, o poca conversión de leads; con la toma de decisiones acertadas mediante la transformación de datos en información útil y pertinente

Todo esto lleva al planteamiento de la siguiente pregunta:

¿Cómo puede la ciencia de datos ser utilizada para encontrar patrones en los datos de los consumidores que ayuden a la creación de estrategias para la toma de decisiones en las empresas?

## Metodología de búsqueda de la información

La metodología de ciencias de datos que se utilizó en este proyecto es la CRISP-DM (Cross-Industry Standard Process for Data Mining). Esta metodología es una de las más utilizadas en el mundo de la ciencia de datos. Esta proporciona un marco estructurado que sirve como guía para desarrollar proyectos de análisis de datos de manera eficiente. Esta metodología es muy flexible y adaptable, por lo que es ideal para aplicar en cualquier industria.

En este informe se aplicarán ocho pasos para el correcto desarrollo de los diferentes análisis, los cuales serán mencionados brevemente a continuación:

- El primero paso consiste en identificar la iniciativa del negocio, en donde su objetivo es asegurarse de la correcta comprensión del reto y del impacto que este tendrá en la empresa.
- El paso número dos será identificar los stakeholders donde se pretende entender las diferentes necesidades de todas las personas interesadas y como cada uno de estos desde sus diferentes funciones puede apoyar o verse afectado
- En el paso tres se identifican las entidades analíticas en donde se logra entender que análisis se pueden hacer y que preguntas pueden responderse con estos datos.
- En el paso cuatro se priorizan las decisiones enfocando así todos los esfuerzos analíticos en las decisiones con mayor impacto para la iniciativa de negocio
- Como paso número cinco se encuentra el Brainstorm de variables y métricas en donde se trabajará la metodología CRISP-DM la cual será profundizada más adelante, en este punto el objetivo es tener un listado amplio de las posibles variables, para seleccionar así las más relevantes

- En el paso seis encontramos la creación de métricas y visualizaciones en donde se pretende facilitar la interpretación que los stakeholders les den a los resultados
- En el séptimo paso se identifican las recomendaciones para lograr traducir los datos en acciones que la empresa pueda tomar
- Por último, se pretende mapear los scores a acciones, lo que se llamaría plan de acción en donde se asegura la implementación práctica de las recomendaciones y establecer KPIS de seguimiento.

En este proyecto se seguirán diez fases que son fundamentales para desarrollar la metodología CRISP-DM: comprensión del negocio, enfoque analítico, requisitos de datos, recopilación de datos, comprensión de datos, preparación de datos, modelado, evaluación de modelo, implementación y retroalimentación. Cada una de estas fases se detallará a continuación:

1. **Comprensión del negocio:** La primera fase de la metodología CRISP-DM es la comprensión del negocio, Este primer paso es fundamental en la ciencia de datos, ya que en esta etapa se busca entender el problema a resolver. Según Provost & Fawcett, (2013) uno de los errores más comunes en proyectos de ciencia de datos es saltarse esta fase y pasar directamente al análisis de datos. Sin embargo, tener una comprensión profunda del problema asegura que los resultados sean relevantes y útiles para las partes interesadas.

El principal objetivo de esta fase es definir la pregunta central bajo la cual se llevará a cabo todo el proyecto, para ello, se debe identificar los objetivos del negocio y comprender las limitaciones y oportunidades existentes. Además, durante esta primera fase es importante involucrar a todas las partes que puedan estar interesadas, tales como: gerentes, expertos en el dominio y usuarios finales. Esto es importante, ya que

así se asegurará tener una visión más amplia del problema y de la situación de la empresa y se asegurará que todas las necesidades sean tenidas en cuenta.

2. **Enfoque analítico:** Al tener una buena comprensión del negocio, el paso siguiente es seleccionar el enfoque analítico adecuado. Según Provost & Fawcett, (2013) elegir el enfoque correcto es crucial porque determina como se utilizarán los datos para abordar el problema de manera efectiva. El enfoque va a depender del tipo de pregunta que se esté buscando resolver y de los objetivos planteados en el proyecto. Existen diferentes tipos de enfoques analíticos, los cuales se mencionarán a continuación:
  - **Modelos predictivos:** Si la pregunta se encuentra orientada a determinar una probabilidad de alguna acción futura, el modelo predictivo es el modelo adecuado.
  - **Análisis descriptivo:** Este enfoque es utilizado cuando hay una gran cantidad de datos que se necesitan analizar con la finalidad de encontrar patrones o tendencias.
  - **Análisis estadístico:** Cuando hay un problema que requiere de una respuesta binaria (si/no) dentro del análisis estadístico se puede aplicar un enfoque de clasificación
  - **Aprendizaje automático (Machine learning):** Esta técnica es ideal cuando se necesitan identificar relaciones y tendencias complejas que no serían tan fácil de identificar mediante los métodos tradicionales
  - **Clustering y asociación:** cuando la pregunta está orientada a aprender sobre el comportamiento humano este enfoque es el que se debe utilizar
3. **Requisitos de Datos:** En esta fase se establecen las bases necesarias para identificar cuáles datos son necesarios. Esto se hace teniendo en cuenta que los datos se encuentren alineados al enfoque analítico definido en la fase anterior. Ya que la falta de

alineación entre los datos y el enfoque analítico puede llevar a resultados inexactos o modelos poco confiables (Field, 2018). Para ello, es necesario identificar el contenido, el formato de los datos y las fuentes de donde han sido extraídos.

4. **Recopilación de Datos:** Inicialmente se realiza una recopilación de datos, se evalúan los datos y se verifican los requisitos de datos, con esto se toma la decisión de si es necesario recopilar más datos. Al tener todos los datos se puede evaluar la calidad, el contenido de los datos y se logrará tener unos conocimientos iniciales sobre los datos que se utilizarán, para esto, se pueden aplicar técnicas como la estadística descriptiva y la visualización al conjunto de datos.
5. **Comprensión de Datos:** El objetivo principal de esta fase es determinar si los datos que se han recopilado son útiles o representativos para responder a la pregunta que se ha planteado inicialmente. Para esto, se pueden aplicar diferentes métodos como lo son las estadísticas descriptivas, correlación por pares, se pueden examinar los histogramas.
6. **Preparación de datos:** La preparación de los datos es una de las etapas más demoradas, generalmente suele tomar más del 70% del tiempo total del proyecto, pero también es una de las más importantes para garantizar el éxito del proyecto. El objetivo principal es transformar los datos al formato que sea el más adecuado para el modelo que se aplicará, en este proceso se eliminan los datos duplicados, se corrigen los valores faltantes y se asegura que todos los datos tengan el formato adecuado
7. **Modelado:** En esta etapa se realiza el desarrollo del modelo predictivo o descriptivo. Esto se hace basado en el enfoque analítico seleccionado, y se desarrolla con aprendizaje automático en el caso del modelo predictivo y en el caso del modelo descriptivo se realiza estadísticamente. Es muy importante que el modelo esté alineado

son el enfoque analítico. Según Field (2018), un modelo mal alineado con el enfoque analítico generará resultados irrelevantes, incluso si técnicamente es “preciso”.

8. **Evaluación del Modelo:** La evaluación del modelo es una etapa que va de la mano con la creación del modelo y esta se hace antes de que el modelo sea implementado. Según Provost y Fawcett, (2013), esta etapa no solo mide la precisión técnica del modelo, sino también su relevancia práctica para el negocio. Con la evaluación del modelo se puede evaluar la calidad de este y además ver si cumple con la solicitud inicial.
9. **Implementación:** Al haber terminado con la evaluación del modelo el paso siguiente es la implementación, este es muy importante ya que según Larose y Larose (2015)., muchos proyectos de ciencia de datos fracasan no por errores técnicos, sino por una mala implementación (p. 15). Para esto se puede poner a prueba con un grupo pequeño de usuarios, para probar su efectividad y posteriormente ser aplicado en todos los ámbitos que sea necesario.
10. **Retroalimentación:** Un modelo de ciencia de datos no es un producto estático, sino un sistema dinámico que requiere monitoreo y ajustes continuos para mantener su utilidad (Provost & Fawcett, 2013). Al tener los comentarios de los usuarios se puede evaluar el rendimiento de este y refinar el modelo de ser necesario y aunque esto es algo que se debe hacer de manera continua es importante definir hasta que punto se refinará para evitar seguir haciendo ajustes de manera indefinida.

## Sustentación teórica de la pregunta

### Identificar la Iniciativa de Negocio

- **Problema de la empresa:** EasyTech es una empresa que se dedica a la venta de productos electrónicos y de tecnología. En esta empresa se ha detectado un problema y es que actualmente, enfrenta un estancamiento en sus ventas, y un muy bajo nivel de recurrencia entre sus clientes, afectando de esa manera sus ingresos.
- **Objetivo principal:** Identificar estrategias de marketing para de esta forma lograr incrementar las ventas de la compañía al menos un 20 % durante el próximo semestre.

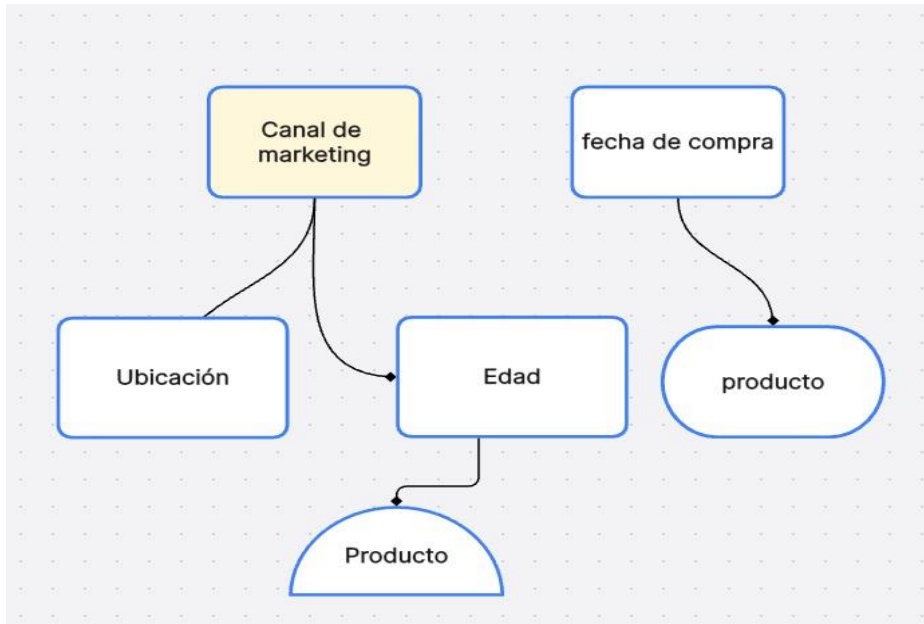
### Identificar a los Stakeholders

- **Director de Ventas:** tiene un nivel de influencia alto. Es el responsable de optimizar el proceso de ventas y la conversión de clientes
- **Director de Marketing:** Nivel de influencia alto. Crea estrategias publicitarias y de marketing para aumentar la retención y conversión de clientes.
- **Gerente general:** Nivel de influencia medio-alto. Está interesado en mejorar el crecimiento y los ingresos de la empresa.

### Identificar las entidades analíticas

A continuación, se presenta un flujograma con las entidades analíticas sobre las cuales se realizarán los análisis de los canales de marketing de la empresa EasyTech.

**Figura 1**  
*Entidades analíticas*



**Fuente:** elaboración propia. **Nota.** La figura 1 muestra el flujograma de entidades analíticas.

### Priorizar las Decisiones

- ¿El canal de marketing utilizado tiene alguna relación con los productos vendidos considerando la edad del cliente? Para abordar esta pregunta se realizará un análisis a la relación que existe entre el canal de Marketing, la edad de los clientes y el producto. Eso se realizará debido a que se cree que puede haber factores como el tipo de publicidad que se usa para llegar a las personas de diferentes edades, dado que no es lo mismo montar publicidad vía correo electrónico o redes sociales para personas con edades diferentes. Además, se tendrá en cuenta que personas de diferentes edades suelen comprar productos diferentes. Por ello, se cree que esta es una buena hipótesis por desarrollar para evaluar cuál es el alcance que estos factores tienen en las ventas.

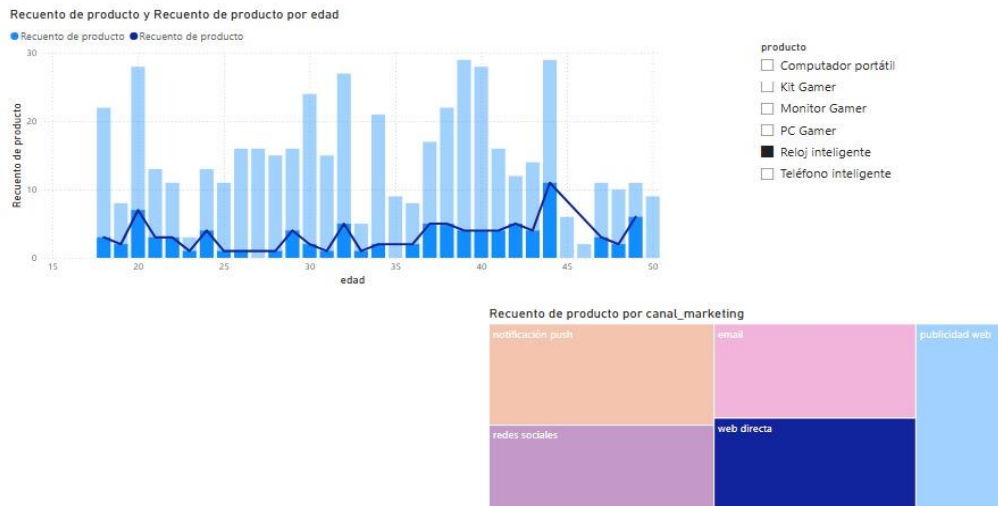
- ¿Existen temporadas en las que las ventas de algún producto sean mayores? Buscando responder esta pregunta se analizará la fecha de compra con los productos comprados para determinar si hay temporadas o fechas en las que se puede incrementar el número de ventas de ciertos productos
- ¿El mismo canal de marketing utilizado es igual de eficiente en las diferentes ubicaciones en que se encuentran los clientes? Finalmente, para responder a esta pregunta se realizará un análisis para determinar si el canal de marketing que se está usando para llegar a los clientes es igual de eficiente en las diferentes ubicaciones donde se encuentran ubicados los clientes, esto debido a las diferencias culturales que existen en las diferentes regiones lo que puede hacer que el mismo método de comunicación no sea tan efectivo en todos los lugares.

### **Creación de Métricas y Visualizaciones**

Al realizar un análisis entre las variables de producto, canal de marketing y edad de los clientes se ha logrado identificar que las personas de 44 años son las que tienen una mayor adquisición de estos productos, en las gráficas se pueden evidenciar la compra de dos productos diferentes, en edades iguales, pero con diferentes canales de marketing, obteniendo así diferentes resultados en las ventas.

#### **Gráfico 1.**

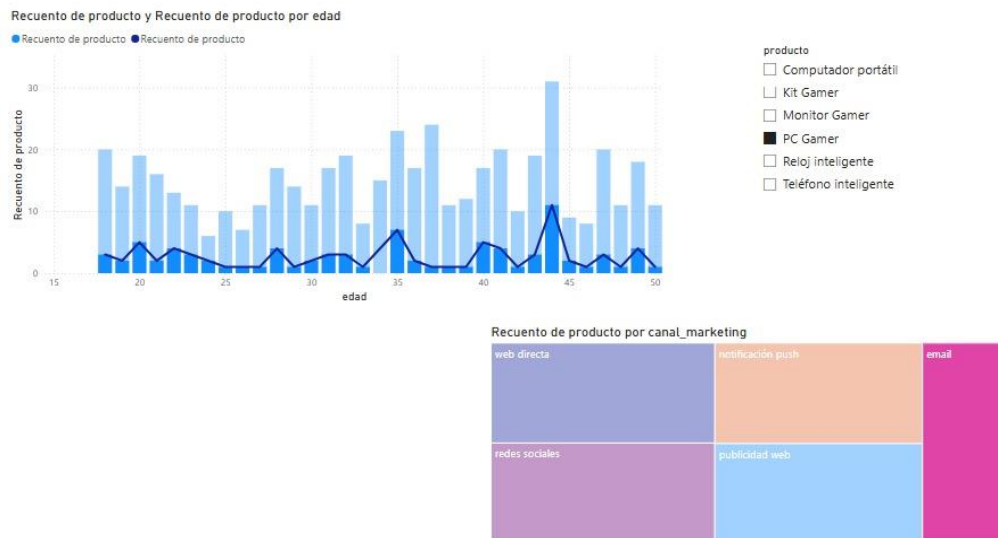
Recuento de productos por edad, producto 1 (Reloj inteligente).



**Fuente:** elaboración propia. **Nota.** El gráfico 1 muestra el recuento del producto 1 (reloj inteligente) por edad.

**Gráfico 2**

Recuento de productos por edad, producto 2 (PC Gamer).

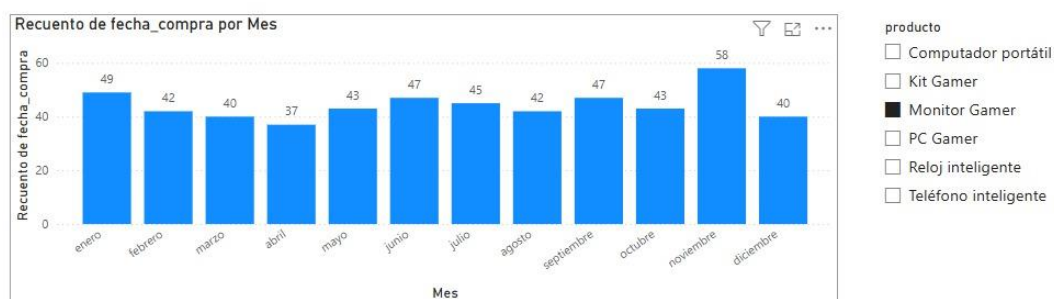


**Fuente:** elaboración propia. **Nota.** El gráfico 2 muestra el recuento del producto 2 (PC Gamer) por edad.

Después de realizar el análisis podemos observar la relación del producto con la fecha de compra, en la grafica se puede ver que este producto generalmente se vende constantemente en el año, sin embargo, el mes con mayores ventas es el de noviembre.

### Gráfico 3

*Recuento de productos por fecha de compra*



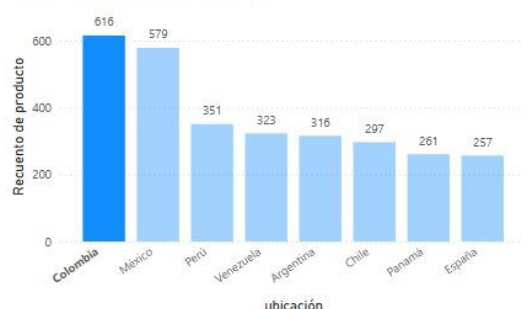
**Fuente:** elaboración propia. *Nota.* El gráfico 3 muestra el recuento de productos por la fecha de compra por mes.

En esta gráfica se evidencia que los países con mayores ventas son Colombia, seguido por México, en ambos los canales de marketing que más se utilizan son notificación push y publicidad web.

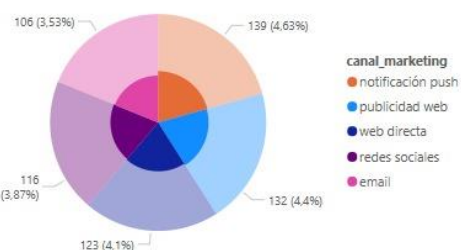
### Gráfico 4

*Recuento de productos por ubicación*

Recuento de producto por ubicación



Recuento de ubicación por canal\_marketing



**Fuente:** elaboración propia. **Nota.** El gráfico 4 muestra el recuento de productos por

## Recomendaciones

Tras realizar el análisis de las dos primeras gráficas se recomienda explorar diversos canales de marketing en diferentes productos para las personas de la misma edad, para así lograr el mejoramiento de sus estrategias y una mayor eficiencia en las ventas.

En el análisis de las siguientes graficas se evidencian mayores ventas durante un mes en específico por lo que se recomienda aumentar el inventario de dicho producto durante este mes, para lograr cumplir con un stock adecuado para el cumplimiento de las ventas.

En la gráfica de recuento por ubicación se logran evidenciar dos países en el que las ventas son más elevadas utilizando principalmente dos canales de marketing, por lo cual se le recomienda a la empresa implementar en mayor medida estos canales en otros países, monitoreando su desempeño y el nivel de ventas que se puedan alcanzar por dichos canales.

## Conclusiones

Al desarrollar todo el proceso de ciencia de datos bajo la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), se ha podido observar que la ciencia de datos es una herramienta muy importante para el mundo empresarial. Ya que permite usar los datos que tienen las empresas y convertirlos en información que podrá ser utilizada por las organizaciones para la toma estratégica de decisiones en las diferentes áreas de la compañía

A través de la ciencia de datos, podemos desarrollar modelos descriptivos de análisis de datos que pueden darnos mucha información relevante sobre las empresas, podemos observar y estudiar patrones y relaciones de compras en los clientes para determinar si hay algún factor que impulse más las ventas de la empresa, o si hay algún producto que sea el que represente la mayoría de las ventas de la compañía.

También podemos desarrollar modelos predictivos, con los cuales se pueden predecir resultados futuros con el análisis de datos históricos al utilizar machine learning, el cual le permite al modelo aprender de los datos y hacer predicciones como las posibilidades de generar una venta, predecir la demanda de productos, riesgos crediticios o fraudes

Por lo tanto, la ciencia de datos es una disciplina de gran impacto y valor para las empresas, su adecuado manejo ayuda al mejoramiento continuo de los diferentes problemas que se presentan y de la forma en que pueden ser abordados, todo mediante el correcto uso de los diferentes volúmenes de datos, buscando siempre estrategias de mejora y dándole utilidad a los diferentes datos que a simple vista no muestran mucha información.

## Referencias

- Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics (5ta edición)*. Los Angeles: Sage Publications.
- Larose, D. T., & Larose, C. D. (2015). *Data Mining and Predictive Analytics (2da edición)*. Hoboken: John Wiley & sons, Inc.
- Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know about Data-Analytic Thinking*. California: O'Reilly Media.