

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

1

**ANÁLISIS DE DATOS PARA DETERMINAR EL DÉFICIT DE VENTAS EN UNA  
COMPAÑÍA**

Corporación Universitaria Remington sede Pereira

Facultad de Ingenierías.

Ingeniería de Sistemas.

Daniel David Vanegas Garavito

Ivonne Castaño Osorio

Seminario de Grado.

2025

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

2

**Tabla de contenido**

1.	Resumen.....	3
1.1.	Pregunta orientadora de la búsqueda.....	3
2.	Metodología de búsqueda de la información.....	4
3.	Sustentación teórica de la pregunta.....	6
3.1.	Identificar la iniciativa de negocio:.....	6
3.1.1.	Problema:.....	6
3.1.2.	Objetivo:.....	6
3.2.	Identificar a los <i>Stakeholders</i> :.....	7
3.3.	Identificar las entidades analíticas:.....	8
3.4.	Priorizar las decisiones:.....	8
3.5.	<i>Brainstorm</i> de Variables y Métricas:.....	9
3.6.	Creación de métricas y visualizaciones (scores)- identificar recomendaciones:.....	10
4.	Conclusiones.....	17
	Lista de referencias.....	18

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

3

## 1. Resumen

En este trabajo estará enfocado en Ciencia de Datos donde se abordará una estructura basada en una variación de la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) desarrollada por IBM. El objetivo principal, es guiar el proceso de resolución de problemas mediante el análisis de datos alineados al paso a paso de la metodología que reúne desde la comprensión del problema, la recopilación y preparación de datos, evaluación e implementación de soluciones. Esta metodología, se enfoca en tres aspectos: entender el problema, tener claro ¿qué se quiere solucionar? y definir ¿cómo analizar los datos? Elegir las herramientas correctas para estudiar la información y trabajar con los datos, permite crear modelos que ayuden a predecir resultados o entender mejor el problema. Todo esto con el fin de obtener respuestas precisas y soluciones eficientes.

**Palabras clave:** Ciencia de Datos, Metodología CRISP-DM, Análisis de Datos, Preparación de Datos, Evaluación de Modelos.

### 1.1. Pregunta orientadora de la búsqueda

Según (Google, s.f.), Big Data comprende datos demasiados grandes y una gran variedad de datos estructurados, semiestructurados y sin estructurar, que a lo largo del tiempo crecen y

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

4

seguirán en crecimiento cada vez más, estos volúmenes de datos se vuelven cada vez más complejos de manipular debido al volumen, la velocidad y la variedad de los sistemas tradicionales que existen y no pueden soportar el almacenamiento, procesamiento.

La Ciencia de datos combina diferentes herramientas métodos y tecnología lo que la hace muy importante a la hora de generar resultados a partir de los datos. Las compañías de la actualidad manejan gran cantidad de datos, existen gran cantidad de tecnología que puede almacenar, y recopilar la información de forma automática, como también los sistemas en línea, portales de pago que capturan datos del comercio electrónico, según (Amazon, s.f.)

Según IBM (2021), CRISP-DM, que son las siglas de Cross-Industry Standard Process for Data Mining, es una metodología que incluye descripciones de las fases para desarrollar un proyecto, con las tareas imprescindibles en cada uno de los pasos.

En este orden de ideas, se plantea la siguiente pregunta orientadora de la búsqueda ¿Cómo puede la metodología CRISP-DM, guiar el proceso de análisis de datos para resolver problemas específicos en el contexto de Big Data y Ciencia de Datos?

## **2. Metodología de búsqueda de la información**

La metodología de búsqueda de información, se ha diseñado siguiendo las fases de la metodología CRISP-DM, según USP/ESALQ (2022):

Es una sigla para el inglés Cross Industry Standard Process for Data Mining, que en español significa algo como “Proceso Estándar Inter-Industrias para Minería de Datos”. El

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

5

objetivo de esa metodología es desarrollar modelos a partir del análisis de informaciones y datos de un negocio para prever futuras fallas y soluciones.

En este orden de ideas, se garantiza un proceso estructurado y alineado con los objetivos del análisis de datos con sus siguientes pasos:

- 1- Identificar la idea de negocio, asegurarse de que todos comprendan el reto y el impacto que este tendrá en la empresa.
- 2- Identificar a los STAKEHOLDERS, entender las necesidades de los interesados y cómo cada uno puede apoyar o verse impactado.
- 3- Identificar las entidades Analíticas, saber qué se va a analizar y qué preguntas se pueden responder con esos datos.
- 4- Priorizar las decisiones, enfocar los esfuerzos analíticos en las decisiones con mayor impacto para la iniciativa de negocio.
- 5- Brainstorm de variables y métricas, tener un listado amplio de posibles variables, luego seleccionar las más relevantes.
- 6- Creación de métricas y visualizaciones (scores), facilitar la interpretación que los stakeholders les den a los resultados.
- 7- Identificar recomendaciones, traducir los datos en acciones que la empresa pueda tomar.
- 8- Mapear los “scores” a acciones (plan de acción), asegurar la implementación práctica de las recomendaciones y establecer KPIs de seguimiento.

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

6

**3. Sustentación teórica de la pregunta**

La empresa de comercio electrónico EasyTech.com dedicada a la venta de productos electrónicos y de tecnología está experimentando un estancamiento en sus ventas. El directorio de la compañía desea aumentar los ingresos y mejorar la fidelización de los clientes. Para ello, el equipo de Ciencia de Datos tiene la misión de identificar oportunidades de mejora en las estrategias de marketing, retención y ofertas personalizadas.

**3.1. Identificar la iniciativa de negocio:**

“El cometido general de esta fase es entender los objetivos y requerimientos del proyecto desde una perspectiva de negocio y convertir este conocimiento en la definición de un problema de minería de datos y un plan preliminar para alcanzar los objetivos” (Niño, 2016).

**3.1.1. Problema:**

EasyTech.com, una empresa de comercio electrónico especializada en la venta de productos electrónicos y tecnológicos ha experimentado una disminución en sus ventas en los últimos meses. Para abordar esta situación, la compañía ha decidido contratar un equipo de Ciencia de Datos con el objetivo de analizar a fondo las causas de este déficit.

**3.1.2. Objetivo:**

Identificar las causas de la disminución en las ventas a través del análisis de los datos proporcionados por la empresa para proponer estrategias que permitan aumentar las ventas en un lapso de 3 meses.

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

7

**3.2. Identificar a los Stakeholders:**

Según la ESAN(2021):

Los stakeholders, traducido al español como "grupos de interés" o "partes interesadas", son las personas, grupos u organizaciones que se ven afectadas por las acciones y decisiones de una empresa. Para identificarlos, es fundamental la realización de un mapeo de stakeholders.

Stakeholder's clave de EasyTech.com:

**Director General:**

**Nivel de Influencia:** Alto

**Interés en el Proyecto:**

Busca decisiones estratégicas basadas en datos para mejorar las ventas y rentabilidad, cumpliendo con los lineamientos de la compañía.

**Gerente de Ventas:**

**Nivel de Influencia:** Alto

**Interés en el Proyecto:**

Comprender qué factores están influyendo en la disminución de las ventas y encontrar soluciones efectivas para mejorar el rendimiento comercial, implementando estrategias y recomendaciones basadas en resultados del análisis.

**Gerente de Marketing:**

**Nivel de Influencia:** Alto

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

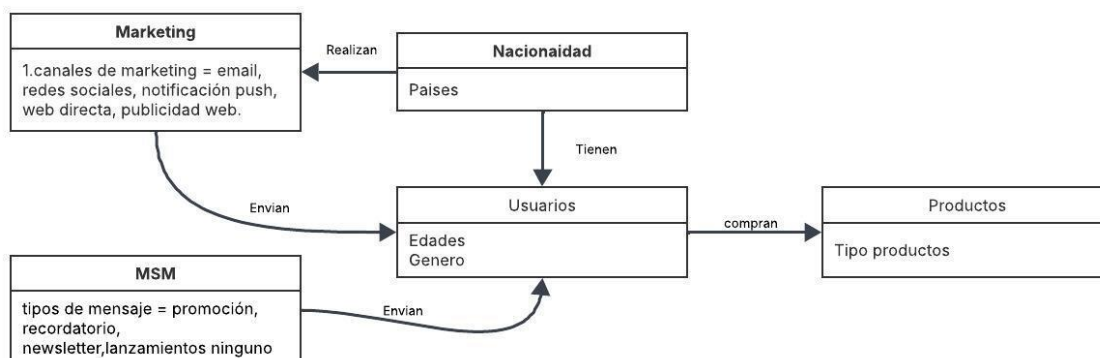
**Interés en el Proyecto:**

Entender la importancia de analizar el comportamiento de las ventas y clientes para diseñar campañas más efectivas creando estrategias enfocadas a mejorar las ventas.

**3.3. Identificar las entidades analíticas:**

Diagrama de las entidades analíticas y la relación entre ellas.

Gráfico 1. Entidades Analíticas.



Fuente: Elaboración propia. Nota. Se observa la relación entre las entidades analíticas que influyen en el análisis.

**3.4. Priorizar las decisiones:**

¿Cuál segmentación (por edad, género o nacionalidad) tiene mayor venta de ciertos productos?

Justificación:

Identificar los segmentos más rentables permite enfocar las campañas y aumentar las ventas.



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

9

¿Existe un canal de marketing (email, redes sociales, Push) que haya perdido efectividad en generar visitas a la web y ventas?

justificación:

Si un canal clave dejó de funcionar, es fundamental reasignar recursos a plataformas más efectivas.

¿Qué producto ha tenido menores ventas y cuáles son las características de los clientes que solían comprarlos?

justificación:

Saber qué producto está reduciendo ventas permitirá decidir si es un problema de demanda, precio o promoción.

**3.5. *Brainstorm* de Variables y Métricas:**

¿Cuál segmentación (por edad, género o nacionalidad) tiene mayor venta de ciertos productos?

Variables utilizadas: Genero, ubicación, producto.

¿Existe un canal de marketing (email, redes sociales, Push) que haya perdido efectividad en generar visitas a la web y ventas?

Variables utilizadas: Genero, canal de marketing, ubicación.

¿Qué producto ha tenido menores ventas y cuáles son las características de los clientes que solían comprarlos?

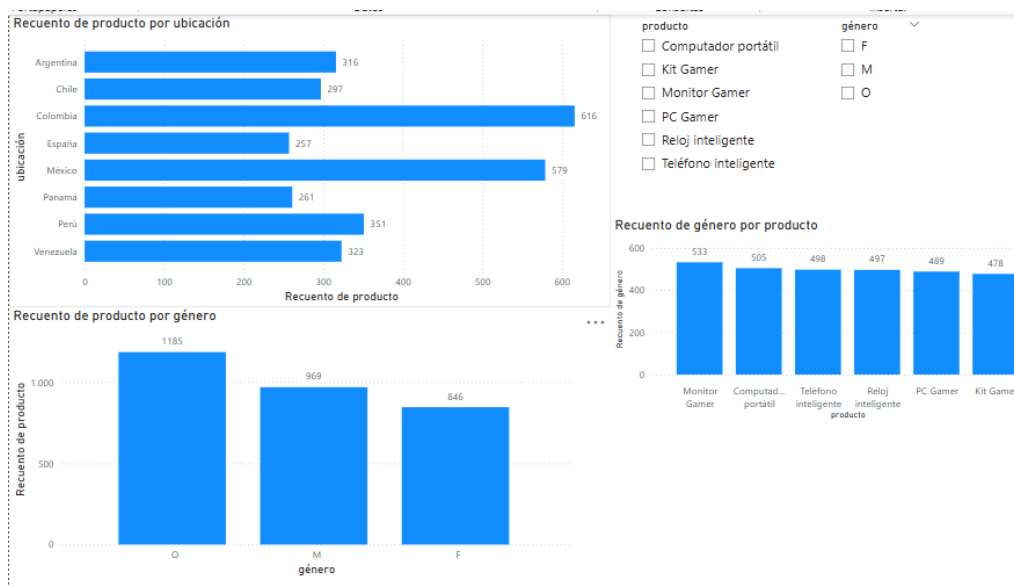
Variables utilizadas: Producto, año, género.

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

**3.6. Creación de métricas y visualizaciones (scores)- identificar recomendaciones:**

¿Cuál segmentación (por edad, género o nacionalidad) tiene mayor venta de ciertos productos?

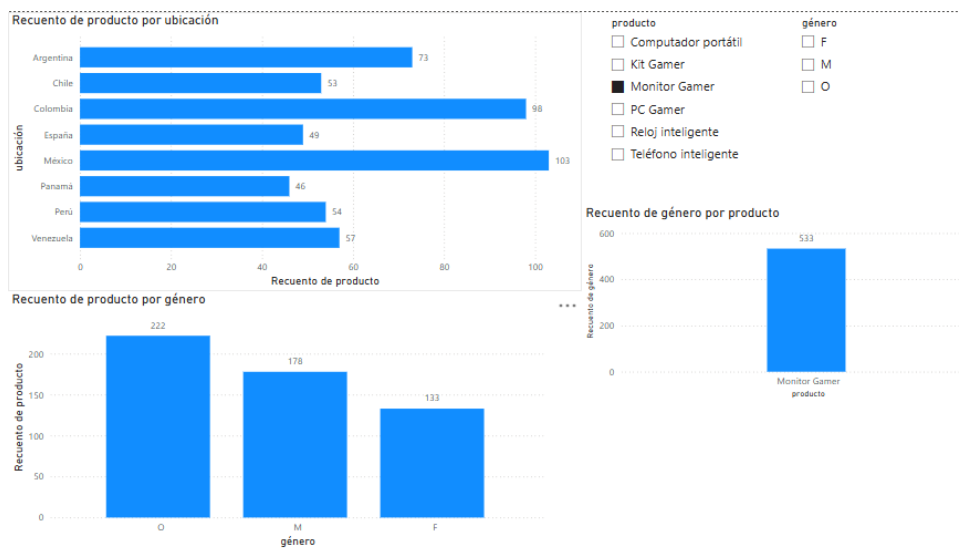
Gráfico 2. Relaciones entre producto, genero, ubicación.



Fuente: Elaboración propia. Nota: En este gráfico se observa el producto más vendido, en dónde y que género es el más activo en las compras.

Gráfico 3. Relación de producto más vendido.

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



Fuente: Elaboración propia. Nota: En este gráfico se observa las variables precisas del el artículo más vendido en donde y por qué género fue comprado.

Se puede evidenciar el género que más adquiere el producto con mayor venta (Monitor Gamer), es el género "Otro", evidenciando las preferencias tecnológicas, dejando evidencia de una gran participación de las comunidades no binarias o personas eligen no identificarse dentro de los géneros tradicionales.

Los países donde se evidencia la mayor venta de este producto son Colombia y México, donde se evidencia una cultura Gamer muy desarrollada, que han tenido un gran crecimiento e interés todo este tipo de accesorios de alta gama, y esto se puede estar dando a una expansión del mercado Gamer, donde los videojuegos en estos países han experimentado un crecimiento significativo aumentando la demanda de productos para esta comunidad como son los Monitores con mejores resoluciones y alta definición dela imagen.

- Se recomienda realizar una Investigación de mercado al género "Otro" por qué adquiere este producto según su inclinación Gamers profesionales, streamers o jugadores

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

12

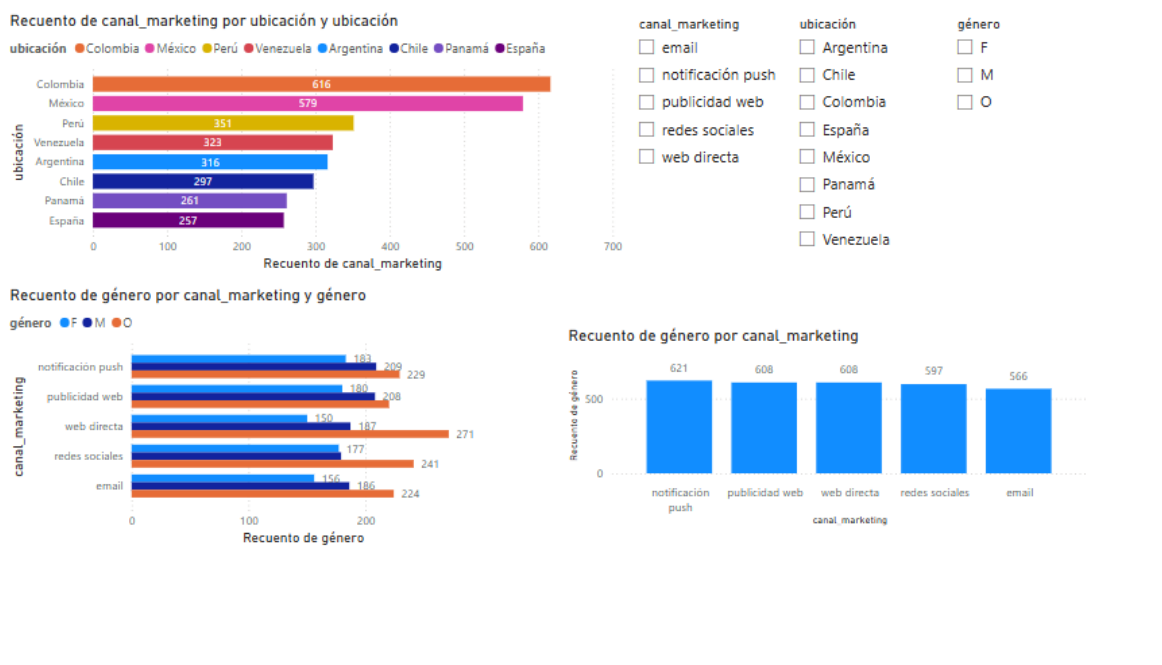
normales. entender su motivación. De esta forma podemos entender sus necesidades de los productos adicionales para poder generarles ofertas y un marketing más dirigido.

- Aunque en este análisis no solo se evidencia una preferencia clara por los monitores Gamer, se evidencia una oportunidad para enfocar estrategias de marketing que estén enfocadas en la personalización e inclusión que se alineen con las tendencias sociales y tecnológicas de la actualidad.

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

¿Existe un canal de marketing (email, redes sociales, Push) que haya perdido efectividad en generar visitas a la web y ventas?

Gráfico 4. Relación de marketing utilizados.



Fuente: Elaboración propia. Nota: En este gráfico se observa la efectividad de cada canal de marketing en las ventas.

Este análisis nos muestra que la campaña marketing “notificación Push”, ha tenido una gran participación y un gran impacto positivo sobre las ventas en los países de Colombia y México dejando evidencia que es muy bien recibida por parte de los clientes, particularmente en el genere “Otro”, con una gran aceptación de los mensajes inmediatos.

- Ya que este canal es el más efectivo, se podría seguir usando, utilizando mensajes más personalizados, y enfocados al historial de compras y posibles productos a comprar según ese historial.

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

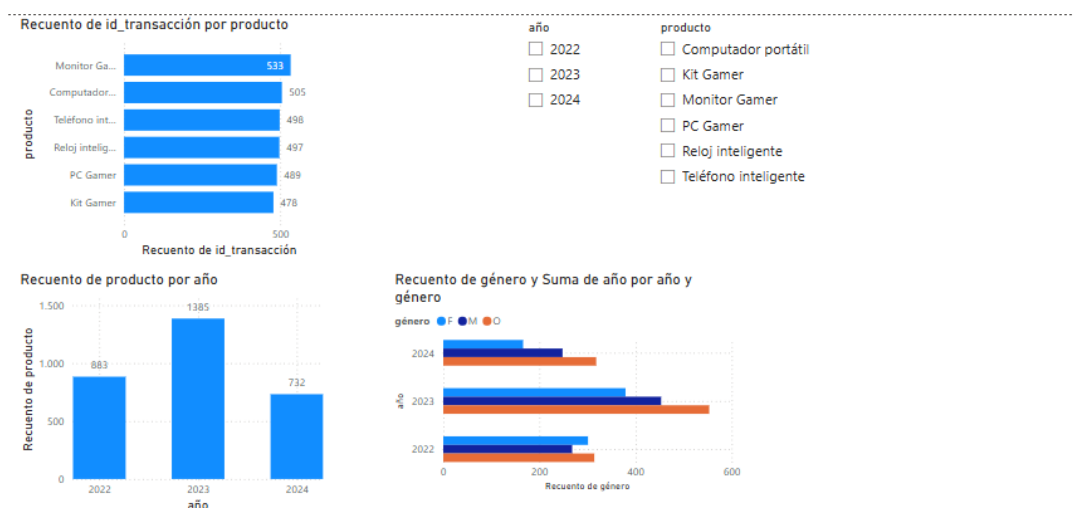
14

- Validar la creación de promociones en combos (monitor Gamer + kit Gamer), validar la posibilidad de descuentos en fechas importantes o días especiales.
- Es importante ver si hay una relación entre las notificaciones push y las compras, como en qué horarios las personas ingresan y realizan sus compras con más frecuencia. Esto con el fin de poder enfocarnos a enviar los mensajes en los momentos justos, para que más gente los vea y tenga más ganas de comprar.
- Fortalecer las campañas inclusivas que resalten la variedad, especialmente en los países con mayor impacto.

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

¿Qué producto ha tenido menor ventas y cuáles son las características de los clientes que solían comprarlos?

Gráfico 5. Relación de productos por ventas.



Fuente: Elaboración propia. Nota: En este gráfico se observa las ventas de cada producto a lo largo del tiempo.

El Kit Gamer aparte de mostrar caída en ventas, es el producto menos vendido. Esto señala una pérdida de interés, falta de promoción, calidad de los productos, estrategias de marketing débiles.

- El género "Otro" sigue liderando las compras para este producto. Es importante reforzar las estrategias dirigidas al género "Otro", pero también explorar formas de captar la atención de los demás géneros de forma inclusiva con mensajes más personalizados y con valores agregados de las descripciones del producto

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

16

- Se recomienda aprovechar las notificaciones Push que son el canal más efectivo, para los productos que presentan más caída de ventas promoviendo el Kit Gamer, resaltando los beneficios y cómo mejora la experiencia de juego.
- Crear tipos de kits agrupados para cada tipo de clientes como lo serian básicos, intermedios o avanzados, con la variedad de los periféricos, con el fin de poder garantizar una mejor experiencia a la hora de comprar y poner en función el producto.



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

17

**4. Conclusiones.**

La metodología CRISP-DM aporta un paso a paso claro y organizado para analizar datos, que facilita la toma de decisiones basadas en información real entregada por las compañías. Cuando se aplica al Big Data, se debe tener en cuenta que las herramientas utilizadas puedan asegurar el manejo de grandes cantidades de información sin perder tiempo y precisión. Esta metodología nos enseña que, es una de las formas más fáciles de entender los datos, para encontrar respuestas y tomar buenas decisiones para las compañías que se aplique.

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

18

**Lista de referencias**

Amazon. (s.f.). *AWS*. Obtenido de AWS: <https://aws.amazon.com/es/what-is/data-science/>

ESAN. (17 de 06 de 2021). Obtenido de ESAN Graduate School of Business:

<https://www.esan.edu.pe/conexion-esan/como-identificar-a-los-stakeholders-de-tu-organizacion>

Google. (s.f.). *Google Cloud*. Obtenido de Google Cloud: <https://cloud.google.com/learn/what-is-big-data>

IBM. (17 de 8 de 2021). *IBM Documentacion*. Obtenido de SPSS Modeler:

<https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>

Niño, M. (18 de 11 de 2016). *El blog de Mikel Niño*. Obtenido de

<https://www.mikelnino.com/2016/11/crisp-dm-metodologia-data-mining-comprension-negocio-business-understanding.html>

USP/ESALQ, M. (31 de 05 de 2022). *MBA USP/ESALQ*. Obtenido de

<https://blog.mbauspesalq.com/es/2022/05/31/crisp-dm-las-6-etapas-de-la-metodologia-del-futuro/>