

**ESTIMACIÓN DE LA PRESCRIPCIÓN DE MEDICAMENTOS EMPLEADOS PARA EL  
TRATAMIENTO DEL CÁNCER DE MAMA QUE ENTRARON AL PLAN DE BENEFICIOS  
EN SALUD**

**María Claudia Gamboa Castañeda**

**Pasantía**

**Tutor**

**Oscar Julian Perdomo Charry, PhD**



**Universidad del  
Rosario**



**ESCUELA  
COLOMBIANA  
DE INGENIERÍA  
JULIO GARAVITO**

---

**UNIVERSIDAD**

**ESCUELA COLOMBIANA DE INGENIERÍA JULIO GARAVITO  
UNIVERSIDAD DEL ROSARIO  
PROGRAMA DE INGENIERÍA BIOMÉDICA  
BOGOTÁ D.C  
2023**

## TABLA DE CONTENIDO

<b>RESUMEN</b>	<b>7</b>
<b>1. INTRODUCCIÓN</b>	<b>8</b>
<b>2. OBJETIVOS</b>	<b>12</b>
2.1. General	12
2.2. Específicos	12
<b>3. METODOLOGÍA</b>	<b>13</b>
3.1. Problema a solucionar	13
3.2. Fases del proyecto	13
3.2.1. Entendimiento del negocio	14
3.2.2. Entendimiento de los datos	14
3.2.3. Preparación de los datos	14
3.2.4. Modelamiento	14
3.2.5. Evaluación	16
3.2.6. Despliegue y culminación	16
<b>4. RESULTADOS</b>	<b>18</b>
4.1. Medicamentos entrantes al PBS	18
4.2. Modelo en Google Colaboratory	19
4.2.1. Pertuzumab	19
4.2.2. Trastuzumab emtansina	20
4.2.3. Trastuzumab	21
<b>5. DISCUSIÓN</b>	<b>23</b>
<b>6. RECOMENDACIONES Y TRABAJOS FUTUROS</b>	<b>24</b>
<b>7. CONCLUSIONES</b>	<b>25</b>
<b>REFERENCIAS</b>	<b>26</b>
<b>ANEXOS</b>	<b>27</b>
ANEXO 1. Diagrama de flujo del código	27

## LISTA DE TABLAS

Tabla 1. Comparación entre el modelo original y el adaptado.

15

## LISTA DE FIGURAS

Figura 1. Proceso CRISP-DM.	10
Figura 2. Diagrama de Gannt.	17
Figura 3. Gráficas de principios activos dentro del PBS.	18
Figura 4. Gráficas de unidades vendidas de los tres principios activos.	18
Figura 5. Gráfica de prescripciones para pertuzumab.	20
Figura 6. Gráfica de prescripciones para trastuzumab emtansina.	21
Figura 7. Gráfica de prescripciones para trastuzumab.	22

## LISTA DE ECUACIONES

Ecuación 1. Modelo original de IQVIA.	15
Ecuación 2. Modelo adaptado.	16
Ecuación 3. Error relativo.	16

## LISTA DE ANEXOS.

Anexo 1. Diagrama de flujo del código

24

## RESUMEN

**Introducción:** Roche es la empresa biotecnológica más grande del mundo la cual cuenta con medicamentos para áreas como oncología, inmunología, enfermedades infecciosas, oftalmología, neurociencias y diabetes. Para los objetivos del presente trabajo, el área terapéutica a tratar es oncología, más específicamente en cáncer de mama. Es importante anotar que para tener una mejor comprensión de este proyecto, un aspecto que influye es el relacionado con el sistema de salud colombiano, específicamente el que se refiere al Plan de Beneficios en Salud (PBS). Este plan es el conjunto de servicios, procedimientos y medicamentos garantizados por la ley para todas las personas aseguradas por una Empresa Prestadora de Salud (EPS). Dentro del sistema de salud en Colombia existe una rama llamada NO PBS, el cual hace referencia a algunos medicamentos o servicios que no están dentro del PBS, sin embargo, el médico tratante puede formularlos si lo considera necesario y la EPS está en la obligación de otorgarlos. El procedimiento para la formulación de estos medicamentos se realiza a través de la plataforma MIPRES, la cual es especialmente para prescribir servicios y tecnologías que no se encuentran dentro del PBS teniendo en cuenta las necesidades de los pacientes la cual consiste en recopilar la información de las prescripciones de los principios activos que no están dentro del PBS.

**Objetivo:** Adaptar el modelo de la empresa IQVIA como herramienta para que el área de data analytics de Roche, estime y consulte el número de prescripciones de medicamentos utilizados para el tratamiento de cáncer de mama, incluidos en el Plan de Beneficios en Salud.

**Metodología:** Para lograr estimar los datos de los medicamentos prescritos que no se encuentran dentro del Plan de Beneficios en Salud, se llevó a cabo una metodología especializada en minería de datos denominada metodología CRISP-DM (del inglés *Cross- Industry Standard Process for Data Mining*) la cual es una de las principales metodologías para el correcto desarrollo de nuevos proyectos relacionados con minería de datos o de características similares. En este apartado se llevó a cabo la adaptación del modelo presentado por la empresa IQVIA (líder mundial en el uso de datos, tecnología, análisis avanzado y experiencia humana para impulsar la industria de la salud) y se halló la nueva ecuación para la implementación del modelo adaptado a las necesidades de Roche.

**Resultados:** Los resultados del modelo adaptado de IQVIA a las necesidades de Roche para estimar y consultar los medicamentos entrantes al PBS utilizados para el tratamiento del cáncer de mama, fueron satisfactorios tanto para el equipo de data analytics de Roche como para el equipo especializado en cáncer de mama, quienes apalancados en ellos, contarán con información para crear estrategias de negocio. Roche cuenta con diferentes principios activos en sus medicamentos como son: Pertuzumab, Trastuzumab emtansina y Trastuzumab. Para el caso de Pertuzumab como para Trastuzumab emtansina, el modelo adaptado permitió contar con resultados significativos de alto impacto, pues la estimación de las prescripciones permiten tomar decisiones estratégicas, llevar a cabo acciones de marketing, realizar lanzamientos de otros medicamentos que contienen los mismos principios activos y estudios de mercado para ampliar los horizontes y desarrollo de la compañía.

**Conclusión:** El modelo presentado tuvo un error promedio del 13%, lo que lo convierte en un modelo moderado de estimación de las prescripciones de los medicamentos para el cáncer de mama entrantes al plan de beneficios (PBS). Esto quiere decir que cuenta con una precisión y exactitud media y es una excelente guía para el tipo de actividades que se van a realizar con respecto a los datos obtenidos.

## 1. INTRODUCCIÓN

Roche es la empresa biotecnológica más grande del mundo la cual cuenta con medicamentos para áreas como oncología, inmunología, enfermedades infecciosas, oftalmología, neurociencias y diabetes. Roche tiene como lema “*Hacer ahora lo que los pacientes necesitarán mañana*” [1], lo que ha permitido mejorar la calidad de vida de muchos pacientes realizando el mayor impacto posible tanto en la compañía como en la sociedad. Actualmente Roche trabaja de la mano de bacteriólogos, médicos e ingenieros para brindar soluciones a los problemas médicos no resueltos brindando medicamentos eficaces e innovadores para enfermedades como cáncer de mama, cáncer de pulmón y hemofilia y generando campos de investigación en hematología, oncología de precisión y neurociencias.

Para el desarrollo de nuevas tecnologías y diferentes estrategias de mercado, es necesario contar con un equipo de trabajo que tenga conocimientos específicos acerca de la enfermedad a tratar, manejo del mercado en el que se comercializan estos medicamentos y que estén en contacto con profesionales de la salud y especialistas en la enfermedad. A este tipo de equipos de trabajo dentro de Roche, se les llama *Squad* en donde están inmersas diferentes disciplinas como ingenieros, médicos, biólogos, bacteriólogos entre otros. Para el caso del desarrollo de este proyecto, se tuvo contacto con el Squad de mama, que es el equipo encargado de manejar los datos y proyectos relacionados con los medicamentos para el cáncer de mama, por lo cual fueron fuente directa para la recopilación de algunos datos, como se verá más adelante.

Por otro lado, el sistema de salud colombiano cuenta con el Plan de Beneficios en Salud (PBS), anteriormente llamado Plan Obligatorio de Salud (POS), el cual es un conjunto de servicios, procedimientos y medicamentos garantizados por la ley para todas las personas aseguradas por una Empresa Prestadora de Salud (EPS) [2], con ciertas características que abarcan la promoción de la salud, la prevención, el diagnóstico, la rehabilitación, el tratamiento y la paliación de una enfermedad. El PBS está obligado a garantizar la salud de los pacientes que requieren procedimientos o medicamentos por lo que existen algunos servicios que no están incluidos en el PBS. Los servicios excluidos son aquellos que tienen como finalidad un tratamiento cosmético, cuando no existe evidencia científica sobre la eficacia y efectividad del medicamento o tratamiento, cuando no cuenta con la autorización de comercialización de una entidad competente, que el tratamiento o medicamento se encuentre en fase de experimentación o que tengan que ser prestados en el exterior [3].

Se debe tener en cuenta que la prescripción de los medicamentos está dada por principios activos. Un medicamento es aquel fármaco, sustancia medicinal o asociaciones y combinaciones que se destinan a ser aplicados al hombre o a los animales para prevenir, diagnosticar, tratar, aliviar o curar enfermedades o dolencias, o para afectar a funciones corporales o al estado mental. Para que un medicamento sea efectivo y tenga un buen rendimiento se necesita de un principio activo (PA), el cual es la materia o porción de un medicamento a la que se atribuye una actividad apropiada para constituir un medicamento. [4]

Dentro de Roche existen varias divisiones de trabajo, sin embargo, el área a profundizar en este proyecto es Farma (del inglés *Pharma*) en la división de analítica de datos (del inglés *data analytics*). Esta división cuenta con acceso al almacén de



información del Ministerio de Salud y Protección Social llamado Sistema Integrado de Información de la Protección Social (SISPRO) el cual cuenta con diferentes bases de datos como lo son: el Sistema de información de precios de medicamentos (SISMED), Mi Prescripción (MIPRES), Registro individual de prestación de servicios (RIPS), entre otros.

Dentro del sistema de salud en Colombia existe una rama llamada NO PBS, el cual hace referencia a algunos medicamentos o servicios que no están dentro del PBS pero el médico tratante puede formularlos si lo considera necesario y la EPS está en la obligación de otorgarlos [5]. El procedimiento para la formulación de estos medicamentos se realiza a través de la plataforma MIPRES, la cual es especialmente para prescribir servicios y tecnologías que no se encuentran dentro del PBS teniendo en cuenta las necesidades de los pacientes. Esta plataforma guarda los datos para distintos tratamientos que se le puedan dar ya que estas son bases de datos públicas. En la base de datos MIPRES, como se ha mencionado anteriormente, se puede consultar información acerca de los medicamentos prescritos que no están dentro del plan de beneficios en salud o PBS, la cual contiene el área terapéutica a tratar, la patología específica, el principio activo que se prescribe y el número de prescripciones que se hacen de ese principio activo.

La base de datos SISMED es aquella que, como su nombre lo indica, contiene información específica de la venta de los medicamentos. Esta información corresponde a el número de unidades vendidas de cada medicamento por mes, el precio en el que se vendió por mes, qué principio activo contiene, en qué presentación viene el medicamento, vía de administración, el titular frente al invima que fabrica y distribuye el medicamento y el área terapéutica que trata el medicamento.

La base de datos RIPS es aquella que contiene información acerca de los registros de atenciones respecto a una enfermedad o patología y procedimientos realizados frente a un diagnóstico específico. Entrando más en detalle, esta base de datos corresponde al área terapéutica a tratar, año, mes, en qué región se ubica el diagnóstico, qué patología trata y el CIE\_10 que corresponde al detalle del diagnóstico. Por ejemplo, la patología es *cáncer de mama* y el CIE\_10 es *tumor maligno de la porción central de la mama*.

Por ley, en Colombia, todos los medicamentos existentes en el mercado deben estar inscritos en las bases de datos del Ministerio de Salud y Protección Social y estar avalados por una entidad competente, la cual en Colombia es el Invima, sin embargo, no todos los medicamentos están dentro del PBS para lo cual existe la plataforma MIPRES que, como se expuso anteriormente, ésta consiste en recopilar la información de las prescripciones de los principios activos que no están dentro del PBS.

El tiempo para la actualización de la información en las tres bases de datos anteriormente mencionadas es regido por trimestres o Q's lo que quiere decir que Q1 va de enero hasta marzo, Q2 va de abril a junio, Q3 va de julio a septiembre y Q4 va de octubre a diciembre. La información es subida al servidor con una frecuencia de Q vencido, esto quiere decir que a finales del Q1-2023 se sube la información del Q4-2022, a finales del Q2-2023 se sube la información del Q1-2023 y así sucesivamente por lo que para el desarrollo de este proyecto se tuvo la información actualizada hasta Q4-2022. También se debe tener en cuenta que estas bases de datos se acceden mediante el servidor de Roche, por lo tanto la información está disponible a partir del Q1-2019.

El desarrollo del proyecto gira en torno a la pérdida de varios datos de algunos medicamentos puesto que a lo largo del tiempo, diversos medicamentos ofrecidos por

Roche han pasado a ser parte del PBS por lo que el registro en MIPRES se ha visto afectado ya que se han perdido los datos respecto los profesionales de la salud que prescriben los medicamentos de Roche y al número de pacientes que reciben dichos medicamentos. Esto ha generado un vacío en el registro de los datos en cuanto a Roche respecta, ya que no se tiene el seguimiento adecuado de la prescripción de dichos medicamentos después de haber sido añadidos al PBS.

Para lograr estimar los datos perdidos de esos medicamentos se llevó a cabo una metodología especializada en minería de datos denominada metodología CRISP-DM (del inglés *Cross- Industry Standard Processfor Data Mining*) la cual es una de las principales metodologías para el correcto desarrollo de nuevos proyectos relacionados con minería de datos o características similares [6]. Esta metodología se puede apreciar en la figura 1 en donde se expone cual es el paso a paso en la cual se define una estructura para proyectos de minería de datos y suministra la orientación para su ejecución. El modelo de referencia da una visión general del ciclo de vida de un proyecto de minería de datos, contiene las fases con sus objetivos, las tareas y las relaciones entre éstas, y las instrucciones paso a paso que se deben llevar a cabo.

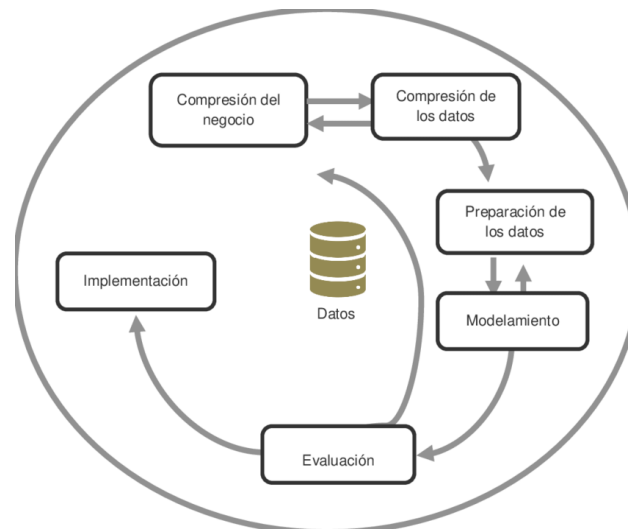


Figura 1. Proceso CRISP-DM.[7]

Iniciando por la comprensión del negocio (del inglés *Business Understanding*), como su nombre lo indica, es el paso que permite entender el negocio en el que se va a desenvolver el proyecto lo que incluye el determinar qué se quiere hacer, cuáles son los objetivos y metas a cumplir, cuanto es el tiempo de inversión y qué recursos hay disponibles para el desarrollo del mismo. La comprensión de los datos (del inglés *Data Understanding*) es el paso que permite conocer y entender los tipos de bases de datos disponibles, recopilación de datos e información que no está organizada, describir y explorar los datos y determinar la calidad de los datos según requiera el proyecto. La preparación de los datos (del inglés *Data Preparation*) es el paso que permite seleccionar los datos que se necesitan, los criterios a evaluar, limpiar, filtrar e integrar los datos y determinar el conjunto de datos final que se va a usar para iniciar el pre-análisis del modelo. El modelamiento (del inglés *Modeling*) es el paso que permite determinar la técnica de modelamiento, realizar el test design y por último la construcción y evaluación del modelo. La evaluación (del inglés *Evaluation*) es el paso que permite evaluar la calidad del modelo, evaluar los resultados y realizar un review del proceso que permita saber

cuanto toca retroceder para mejorar el modelo. El último es el despliegue (del inglés *Deployment*) e implementación del proyecto el cual es el paso que permite culminar el proyecto, generar conclusiones, terminar la documentación que se quiera dejar con respecto al proyecto.

Para entender más fácilmente los medicamentos y principios activos tratados en el documento se debe tener claro un concepto importante. El receptor 2 del factor de crecimiento epidérmico humano (Her2) es un gen que produce una proteína que se encuentra en todas las células de los senos, el cual interviene en el crecimiento normal de las células. En el caso del cáncer de mama, el Her2 sufre una mutación que otorga un ambiente propicio para la proliferación del cáncer de mama [8].

El primer medicamento con el que se va a trabajar en el presente proyecto se llama Perjeta, el cual tiene como principio activo pertuzumab. Este principio activo es un anticuerpo monoclonal que tiene como objetivo prevenir la proliferación del receptor Her2. Este medicamento actúa en conjunto con la quimioterapia para un mejor resultado [9]. El segundo medicamento se llama Herceptin, el cual tiene como principio activo trastuzumab. Este es un anticuerpo monoclonal recombinante contra el Her2 [10]. El tercer medicamento se llama Kadcyly, el cual tiene como principio activo trastuzumab emtansina, el cual es un anticuerpo conjugado para el tratamiento del cáncer de mama metastásico. Este principio activo es otorgado a personas que ya hayan sido tratadas con trastuzumab [11].

Una vez claros todos los conceptos explicados anteriormente, se expone el propósito del proyecto el cual es estudiar un modelo externo presentado dentro de Roche, mejorarlo y adaptarlo para estimar el número de prescripciones mensuales de un medicamento (principio activo) empleado para el tratamiento del cáncer de mama que entró al PBS.

## 2. OBJETIVOS

### 2.1. General

Adaptar el modelo de la empresa IQVIA como herramienta para que el área de data analytics de Roche, estime y consulte el número de prescripciones de medicamentos utilizados para el tratamiento de cáncer de mama, incluidos en el Plan de Beneficios en Salud.

### 2.2. Específicos

1. Definir las variables relacionadas con los medicamentos para el tratamiento del cáncer de mama para realizar la estimación de la prescripción de dicho medicamento.
2. Establecer la relación entre las bases de datos para la adaptación de la herramienta estimadora.
3. Entregar un modelo estadístico que otorgue el número de prescripciones mensuales que se realizan de un medicamento desde que este entró al PBS.

### **3. METODOLOGÍA**

#### **3.1. Problema a solucionar**

Los datos registrados sobre las prescripciones de los medicamentos que ofrece Roche que no están dentro del Plan de Beneficios en Salud (PBS) se encuentran en la base de datos de MIPRES, sin embargo, esto se ha visto afectado a lo largo del tiempo puesto que algunos medicamentos han pasado a hacer parte del PBS, por lo que se han perdido los datos respecto al número de prescripciones de ciertos medicamentos de Roche que realizan los profesionales de la salud y los pacientes que reciben dichos medicamentos. Esto ha generado un vacío en el registro de los datos en cuanto a Roche respecta, ya que no se tiene el seguimiento adecuado de la prescripción de dichos medicamentos después de haber sido añadidos al PBS.

El vacío visto se presenta en todas las áreas terapéuticas que trata Roche (oncología, oftalmología, diabetes, hematología, neurociencias y hemofilia), sin embargo, se quiere llevar a cabo la propuesta inicialmente para los medicamentos que tratan el cáncer de mama lo que sugiere un trabajo a futuro para las demás áreas terapéuticas.

Los datos con los que se va a llevar a cabo la herramienta de estimación del número de prescripciones van a ser tomados de las bases de datos públicas del Ministerio de Salud y Protección Social, de las cuales se destacan MIPRES, SISMED y RIPS. De dichas bases de datos se extrae la información necesaria de cada una de ellas para la adaptación de la herramienta como lo es el número de prescripciones para saber cuáles son los medicamentos que entraron en el PBS y en qué año, el número de unidades vendidas, los miligramos por presentación y el número de diagnósticos de cierta enfermedad.

Se presentó un modelo realizado por la multinacional IQVIA el cual se tomará como base para el desarrollo del propio modelo con cambios que serán explicados más adelante. IQVIA es líder mundial en el uso de datos, tecnología, análisis avanzado y experiencia humana para impulsar la industria de la salud [12]. Una vez claro el modelo presentado por dicha empresa, se procede a realizar filtros en las bases de datos para obtener la información requerida a través de Analysis Services por medio de la herramienta Excel. Posteriormente, se recurre a la programación del modelo en la herramienta Colab para leer las nuevas bases de datos realizadas y presentar distintas gráficas que permitan entender en su totalidad dichas bases de datos. Después de eso, se implementará la ecuación modificada que permitirá estimar el número de prescripciones de un medicamento mes a mes desde que este entró al PBS.

#### **3.2. Fases del proyecto**

Para el desarrollo del proyecto se implementó una metodología denominada Proceso Estándar de Inter-Industrias para Minería de Datos (CRISP-DM) expuesta anteriormente. Esta metodología se puede apreciar en la figura 1 (introducción) en donde se expone cual es el paso a paso para un buen seguimiento y progreso del proyecto.

En cada ítem de esta metodología hubo varios impactos que permitieron el desarrollo del proyecto en el cual se determinaron tareas específicas en cada paso y se obtuvo un plan estratégico acorde a las necesidades y los objetivos del mismo.

### **3.2.1. Entendimiento del negocio**

El primer paso es *Business Understanding* en el cual compete la adaptación, lo que permite estudiar la empresa y entender las necesidades de la misma dando paso al planteamiento inicial y los objetivos del nuevo proyecto los cuales fueron expuestos anteriormente.

### **3.2.2. Entendimiento de los datos**

El segundo paso es *Data Understanding* el cual permite explorar las bases de datos públicas del Ministerio de Salud y Protección Social en las que se encuentran SISMED, RIPS y MIPRES. Estas son estudiadas a través de Analysis Services realizando la conexión al servidor por la herramienta Excel. En ella se puede empezar a interactuar con la información para familiarizarse con la tabla dinámica presentada. Una vez abiertas cada una de las bases de datos, estas permiten realizar filtros, incluir una determinada columna o fila según se requiera la información y añadir los valores que se quieran visualizar.

### **3.2.3. Preparación de los datos**

El tercer paso es *Data Preparation* en el cual se consolidan las habilidades aprendidas en el paso anterior ya que se resume toda la información encontrada aplicando filtros y agregando filas y columnas según convenga. De esto se extraen las bases de datos finales que se van a emplear para el desarrollo del proyecto. De MIPRES se extrae mes, año, nombre del principio activo y el número de prescripciones para observar cuáles y cuándo entran los medicamentos al PBS. De SISMED se extrae mes, año, nombre del principio activo, titular en el invima, el producto, su presentación y el número de unidades vendidas con el fin de comparar las ventas con el número de prescripciones, ya que si no hay prescripciones pero si hay ventas significa que el medicamento sigue en circulación, pero si por otro lado las ventas han disminuido al igual que las prescripciones significa que el medicamento salió del mercado. De RIPS se extrae mes, año, diagnóstico principal y el número de atenciones, esto con el fin de tener un control de todas las atenciones realizadas ligadas a cáncer de mama.

### **3.2.4. Modelamiento**

El cuarto paso es *Modeling*, el cual está en constante cambio con el tercer paso ya que dependiendo del tipo de modelo que se escoja o se quiera probar van a cambiar los datos requeridos. Finalmente el modelo escogido para el desarrollo del proyecto fue basado en el modelo presentado por IQVIA con algunas modificaciones ya que los medicamentos de Roche para cáncer de mama no cumplen con algunos requerimientos

del modelo original de IQVIA el cual se observa en la ecuación 1. Dichos cambios se pueden observar detalladamente en la tabla 1.

$$P = \frac{U*mg*1000}{DDD_{OMS}*D*Period} \quad (1)$$

La ecuación 1 describe el modelo original presentado por IQVIA en la que la variable P representa el número de personas tratadas con x PA por cada 1000 personas, la variable U representa el número de unidades vendidas de x PA al año, la variable mg representa los miligramos por presentación de x medicamento, la variable DDDoms representa la Dosis Diaria Definida por la OMS de x PA, la variable D representa el número de atenciones para cáncer de mama y por último, la variable Period representa el periodo de evaluación el cual para este modelo es de 365 días al año.

*Tabla 1.*  
COMPARACIÓN ENTRE EL MODELO ORIGINAL Y EL ADAPTADO.

	<b>Modelo IQVIA</b>	<b>Modelo adaptado</b>	<b>Descripción</b>
U	Número de unidades vendidas de x principio activo al año	Número de unidades vendidas de x principio activo al mes	Cambia el periodo
mg	Miligramos por presentación de x medicamento	Miligramos por presentación de x medicamento	Sigue igual
DDD	Dosis Diaria Definida por la OMS.	Dosis Mensual Definida por el <i>squad</i> de mama	Cambio de variable ya que estos medicamentos son de uso altamente individualizado.
D	Número de diagnósticos para cáncer de mama	Número de diagnósticos para cáncer de mama	Sigue igual
Periodo	365 días	30 días	Ya que el modelo se va a realizar mensual se debe multiplicar por 30 días
Número de personas	1000	15000	Se multiplica de una vez por la población total de personas con cáncer de mama en Colombia

La variable que más cambios tuvo en la adaptación del modelo fue la dosis diaria definida (DDD) ya que no existe dosis en miligramos diarios definidos según OMS para medicamentos que tratan el cáncer de mama, esto se debe a que estos medicamentos son de uso altamente individualizado [13], es decir que el tratamiento de cada paciente

puede ser diferente dado que las dosis empleadas dependen del peso, sexo, entre otros y hay un amplio uso de la terapia de combinación, en otras palabras estos medicamentos pueden combinarse con otros principios activos para un mejor desarrollo del tratamiento [6].

Una vez realizado el análisis de los cambios que se necesitan para adaptar el modelo de IQVIA a las necesidades de Roche para los medicamentos que tratan el cáncer de mama, se obtiene un nuevo modelo a implementar.

$$P = \frac{U*mg*15000}{DDD_{Squad}*D*Period} * 10 \quad (2)$$

La ecuación 2 describe el modelo adaptado en la que la variable P representa el número de prescripciones de x PA para la población total de personas con cáncer de mama, la variable U representa el número de unidades vendidas de x PA al mes, la variable mg representa los miligramos por presentación de x medicamento, la variable DMDsquad representa la Dosis Mensual Definida por el squad de mama, la variable D representa el número de atenciones para cáncer de mama y por último, la variable Period representa el periodo de evaluación el cual para este modelo es de 30 días al mes.

### 3.2.5. Evaluación

El quinto paso es *Evaluation* en el cual, como su nombre lo indica, se evaluaron los resultados obtenidos y el error relativo con la siguiente fórmula:

$$\%error = \frac{|valor\ real - valor\ obtenido|}{valor\ real} \quad (3)$$

En el error relativo se pueden encontrar las siguientes métricas: del 2 al 5% se considera un modelo casi perfecto, del 5.1 al 10% se considera un modelo sobresaliente y del 10.1 al 20% se considera un modelo moderado.

### 3.2.6. Despliegue y culminación

El sexto y último paso es *Deployment* en el cual se culminaron las conclusiones del trabajo realizado, se proyecta un trabajo a futuro y se culminan las entregas finales como documentos y modelo terminado.

El desarrollo de cada uno de los pasos fue guiado por el diagrama de Gantt que se presenta en la figura 2.



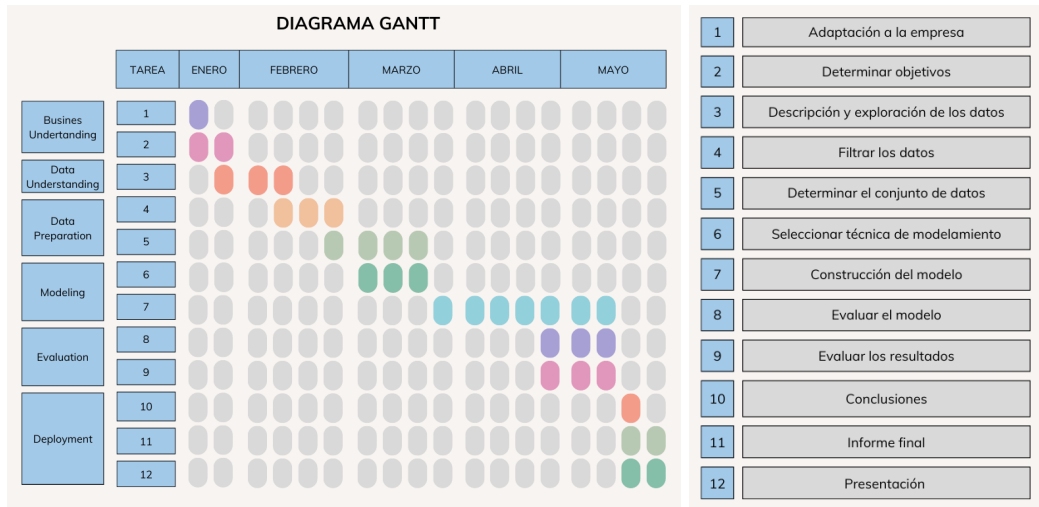


Figura 2. Diagrama de Gannt.

## 4. RESULTADOS

Todas las etapas de la metodología CRISP-DM permitieron obtener resultados satisfactorios logrando identificar cuáles son los medicamentos que entraron al PBS, en qué año y encontrar herramientas que aportarán al desarrollo del modelo.

### 4.1. Medicamentos entrantes al PBS

En el apartado de *data preparation* se pudo observar y analizar detenidamente cada medicamento de Roche empleado para el tratamiento del cáncer de mama. Inicialmente se analizó la base de datos MIPRES en la que se observó una caída evidente de las prescripciones de dos principios activos los cuales son pertuzumab y trastuzumab emtansina y valores de cero o muy bajos para trastuzumab como se muestra en la gráfica de la figura 3, en la cual se puede deducir que pertuzumab entró en enero del 2022, trastuzumab emtansina en enero del 2022 y trastuzumab entró antes del 2019.

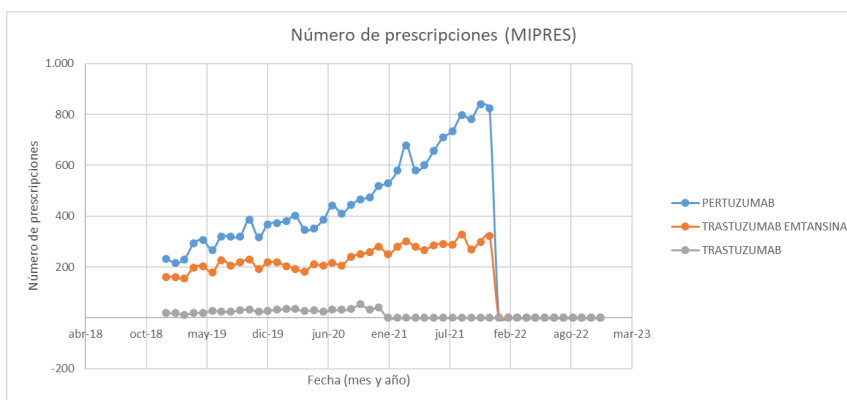


Figura 3. Gráficas de principios activos dentro del PBS.

Para validar la deducción hecha respecto los datos en MIPRES se observaron las ventas de los tres principios activos en la base de datos SISMED como se muestra en la figura 4, en la cual se observa un incremento en el número de unidades vendidas de estos principios activos, con lo cual se confirma la entrada la PBS de pertuzumab, trastuzumab emtansina y trastuzumab.

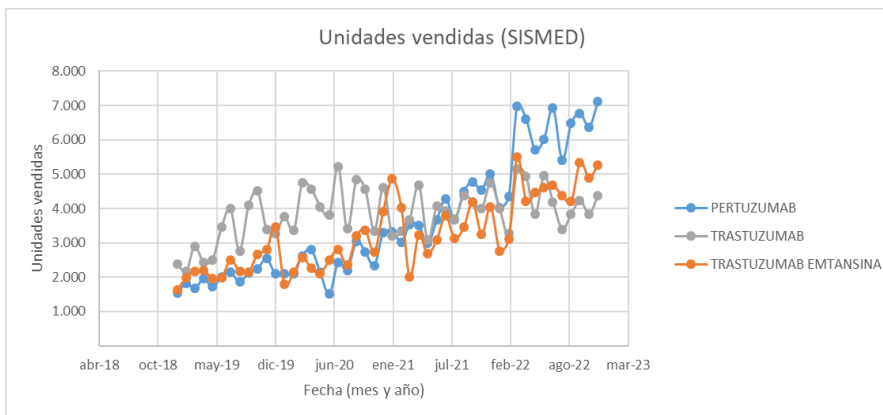


Figura 4. Gráficas de unidades vendidas de los tres principios activos.

## 4.2. Modelo en Google Colaboratory

Para el desarrollo del modelo presentado en la metodología se hizo uso de la herramienta de Google Colaboratory (Colab) ya que permite trabajar en un entorno online y en conjunto con las personas del equipo que estaban dentro del proyecto. Lo primero que se lleva a cabo en el entorno de Colab es el cargue de las tres bases de datos públicas (MIPRES, SISMED y RIPS) dando paso a la implementación de la ecuación 2.

La primera variable en determinar es la D (número de atenciones para cáncer de mama) ya que es la única variable en común para la implementación de la ecuación 2 para cada uno de los principios activos. Para hallar la variable D se recurre a la base de datos RIPS ya que en esta se encuentra el número de atenciones hechas a cáncer de mama, se suman todas estas para obtener el número total de atenciones.

### 4.2.1. Pertuzumab

La DMDsquad (dosis mensual definida por el *squad*) para pertuzumab es DMDperjeta ya que el principio activo pertuzumab pertenece al medicamento Perjeta. El *squad* de mama reportó que la dosis mensual definida para el medicamento Perjeta es de 420 mg.

Para determinar la variable Upertu (número de unidades vendidas de pertuzumab al mes) se extrajo la columna "UNIDADES\_VENDIDAS" de la base de datos SISMED para el principio activo pertuzumab.

Para determinar la variable MGpertu (miligramos por presentación de Perjeta) se extrajo la columna "PRESENTACION" de la base de datos SISMED para el principio activo pertuzumab.

Una vez se obtuvieron las variables completas, se procede a reemplazarlas en la ecuación 2 para hallar el número de prescripciones. Sin embargo, ya que pertuzumab entró al PBS en enero del 2022, se puede estimar el error con respecto al número de prescripciones reales extraídos de MIPRES de enero de 2019 a diciembre de 2021 vs el número de prescripciones estimadas de enero de 2019 a diciembre de 2021. Esto se lleva a cabo en Colab mediante la ecuación 3, lo que genera la aplicación punto a punto de cada dato de MIPRES vs el estimado como se muestra en la figura 5.

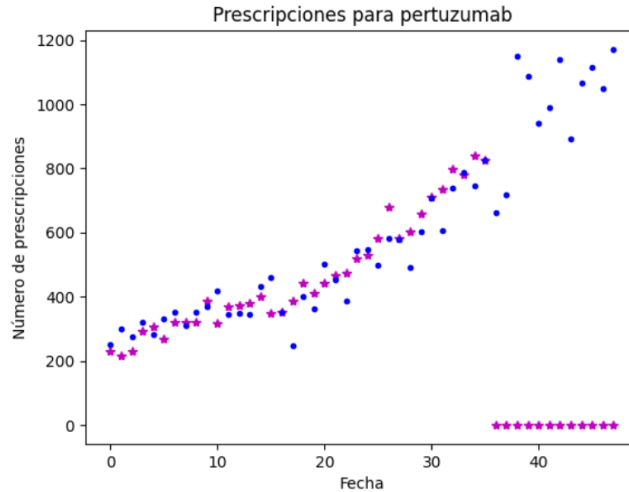


Figura 5. Gráfica de prescripciones para pertuzumab

Después de esto, se realiza un promedio del error que permita estimar el error general del modelo aplicado al principio activo pertuzumab. El error promedio para pertuzumab fue del 11.699735%.

En la figura 5 se puede observar en el recuadro verde en donde se observa mejor el ajuste el cual pertenece a los meses de abril a octubre del 2019. En el recuadro rojo se observan las prescripciones iguales a cero, esto se debe a la caída de las prescripciones que indica la entrada del medicamento al PBS.

#### 4.2.2. Trastuzumab emtansina

La DMDsquad (dosis mensual definida por el *squad*) para trastuzumab emtansina es DMDkadcyla ya que el principio activo trastuzumab emtansina pertenece al medicamento Kadcyly. El *squad* de mama reportó que la dosis mensual definida para el medicamento Kadcyly es de 3.6 mg/kg. Para llegar a un valor más general se estimó para una mujer colombiana promedio (57 - 65 kg) de 61 kg con lo cual se obtuvo que la dosis mensual promedio es de 220 mg.

Para determinar la variable MGtrastuemta (miligramos por presentación de Kadcyly) se extrajo la columna "PRESENTACION" de la base de datos SISMED para el principio activo trastuzumab emtansina. De esta variable se extrajeron dos sub-variables ya que Kadcyly tiene dos presentaciones diferentes, una de 100 mg y otra de 160 mg.

Para determinar la variable Utrastuemta (número de unidades vendidas de trastuzumab emtansina al mes) se extrajo la columna "UNIDADES\_VENDIDAS" de la base de datos SISMED para el principio activo trastuzumab emtansina. Este paso se realizó dos veces ya que, como se mencionó anteriormente, trastuzumab emtansina se encuentra en el mercado con dos presentaciones.

Una vez se obtuvieron las variables completas con las diferentes presentaciones, se procede a reemplazarlas en la ecuación 2 para hallar el número de prescripciones de

cada presentación. Ya que trastuzumab emtansina entró al PBS en enero del 2022, se puede estimar el error con respecto al número de prescripciones reales extraídas de MIPRES de enero de 2019 a diciembre de 2021 vs el número de prescripciones estimadas de enero de 2019 a diciembre de 2021. Para esto es necesario sumar el número de prescripciones para la presentación de 100 mg y para la de 160 mg.

Esto se lleva a cabo en Colab mediante la ecuación 3, lo que genera la aplicación punto a punto de cada dato de MIPRES vs la suma del estimado como se muestra en la figura 6.

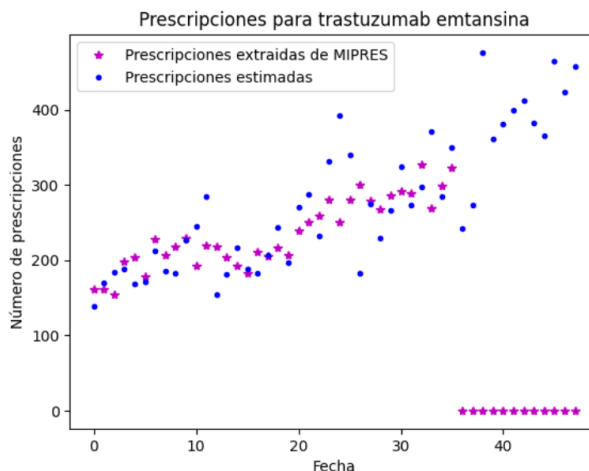


Figura 6. Gráfica de prescripciones para trastuzumab emtansina

Después de esto, se realiza un promedio del error que permita estimar el error general del modelo aplicado al principio activo trastuzumab emtansina. El error promedio para trastuzumab fue de 14.32169%.

En la figura 6 se puede observar en el recuadro verde en donde se observa mejor el ajuste el cual pertenece a los meses de enero a septiembre del 2019. En el recuadro rojo se observan las prescripciones iguales a cero, esto se debe a la caída de las prescripciones que indica la entrada del medicamento al PBS.

### 4.2.3. Trastuzumab

La DMDsquad (dosis mensual definida por el *squad*) para trastuzumab es DMDherceptin ya que el principio activo trastuzumab pertenece al medicamento Herceptin. El *squad* de mama reportó que la dosis mensual definida para el medicamento Herceptin es de 6 mg/kg, con lo cual se realizó la misma estimación para una mujer colombiana promedio (57 - 65 kg) de 61 kg con lo cual se obtuvo que la dosis mensual promedio es de 366 mg.

La variable MGtrastu (miligramos por presentación de Herceptin) fue determinada de la misma manera que Mgrastuemta, sin embargo, en el proceso se extrajeron tres sub-variables ya que Herceptin se encuentra en el mercado con dos presentaciones diferentes, una de 440 mg, otra de 600 mg y la última de 150 mg.

La variable Utrastu (número de unidades vendidas de trastuzumab al mes) fue determinada de la misma manera que Utrastuemta. Este paso se realizó tres veces ya que, como se mencionó anteriormente, trastuzumab tiene tres presentaciones.

Una vez se obtuvieron las variables completas con las diferentes presentaciones, se procede a reemplazarlas en la ecuación 2 para hallar el número de prescripciones de cada presentación. Ya que trastuzumab entró al PBS en el 2019 y las bases de datos están desde 2019, no se puede estimar el error con respecto al número de prescripciones reales extraídas de MIPRES ya que no se cuenta con la información necesaria para realizar la comparación, sin embargo, se lleva a cabo la estimación de las prescripciones para trastuzumab. Para esta es necesario sumar el número de prescripciones para la presentación de 440 mg, de 600 mg y la de 150 mg, lo que tuvo como resultado el número de prescripciones de la figura 7.

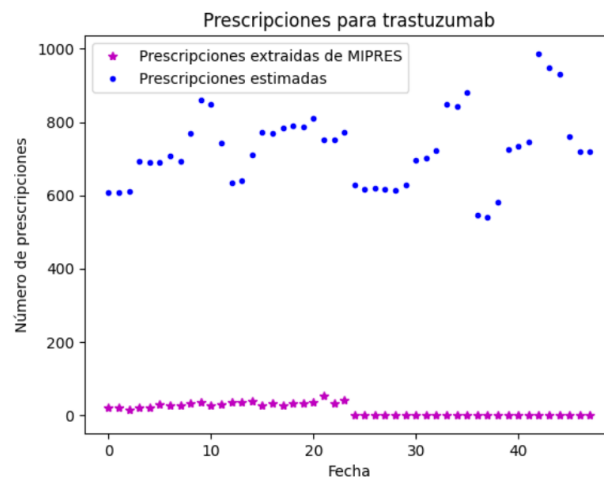


Figura 7. Gráfica de prescripciones para trastuzumab

En la figura 7 se evidencia una diferencia entre las prescripciones extraídas de MIPRES y las prescripciones estimadas ya que, como se mencionó anteriormente, Este principio activo entró al PBS antes del 2019, por lo que las prescripciones de MIPRES serán nulas o casi nulas. Por otro lado, no se observa una tendencia clara de las prescripciones estimadas, esto se debe a un error en los datos dados por el squad a los cuales se dará lugar más adelante.

## 5. DISCUSIÓN

Los resultados del modelo adaptado para los medicamentos entrantes al PBS utilizados para el tratamiento del cáncer de mama, fueron satisfactorios tanto para el equipo de trabajo como para el squad, quienes apalancados en ellos contarán con datos para crear estrategias de negocio. Vale la pena aclarar que existen oportunidades de mejora en relación con la estimación de los datos.

Para el caso de pertuzumab como para trastuzumab emtansina, el modelo adaptado permitió contar con resultados significativos de alto impacto ya que la estimación de las prescripciones resultó favorable para la toma de decisiones estratégicas, llevar a cabo acciones de marketing, realizar lanzamientos de otros medicamentos que contienen los mismos principios activos y estudios de mercado, entre otros.

Tal como se observa en la figura 7, no se evidencia una tendencia clara en la gráfica de las prescripciones de trastuzumab. Al entrar un poco más en la búsqueda de la razón por la cual dicho comportamiento estaba presente se encontró que no se tuvo en cuenta la vía de administración para Herceptin, medicamento que cuenta con tres tipos de presentaciones: 440 mg, 600 mg y 150 mg. La presentación de 440 mg y 150 mg corresponden a la administración inyectada por vía intravenosa y la de 600 mg corresponde a la administración por vía subcutánea.

Para la implementación de este modelo solo se tuvo en cuenta el uso de Herceptin de administración inyectada por vía intravenosa para las tres presentaciones, lo que generó un sesgo en la recopilación de los datos y, en este caso, se altera una de las variables más importantes del modelo la cual es la Dosis Mensual Definida por el squad (DMDsquad), ya que hubo falta de información de la dosis mensual definida para cada vía de administración por parte del squad de mama.

Los resultados de las prescripciones para trastuzumab se encuentran en la figura 7, sin embargo, a conveniencia del estudio se decidió obviar estos resultados a fin de no proporcionar datos erróneos, no obstante, es un paso importante a implementar para trabajos futuros.

Tomando en cuenta los valores de los errores del modelo para pertuzumab y trastuzumab emtansina, se puede entender que el error promedio del modelo es de 13.0107%.

## 6. RECOMENDACIONES Y TRABAJOS FUTUROS

Se pudo evidenciar el alto impacto dentro de Roche que se generó en esta área terapéutica como lo es oncología, más específicamente en cáncer de mama. Por consiguiente, el trabajo a corto plazo más importante será el ajuste de las Dosis Mensuales Definidas para el medicamento Herceptin, adquiriendo la Dosis Mensual Definida específica para cada una de las presentaciones de este medicamento teniendo en cuenta las diferentes vías de administración. De esta manera se podrán obtener unos buenos resultados respecto al medicamento Herceptin y podrán ser presentados al squad de mama.

Por otra parte, otro trabajo a futuro es llegar a implementar el modelo presentado para medicamentos de la misma área terapéutica, como puede ser para cáncer de pulmón, e incluso otras áreas terapéuticas que tengan medicamentos entrantes en el plan de beneficios. Esto puede llevarse a cabo ya que es muy sencillo adaptar dicho modelo para las Dosis Mensuales Definidas del squad que corresponda y para la correcta consulta de las unidades vendidas de dicho medicamento ya que son las variables más significativas.

Se recomienda para trabajos futuros estudiar a detalle los principios activos con los que se va a trabajar y por ende los medicamentos que contienen dichos principios activos. Dentro de este estudio se debe tener en cuenta las diferentes presentaciones, las vías de administración y por último, indagar si dicho principio activo tiene Dosis Diaria Definida por la OMS o si en dado caso se debe tener una oportuna comunicación con el squad al que pertenezca el medicamento para obtener la correcta Dosis Definida, bien sea diaria, mensual o anual.



## 7. CONCLUSIONES

El modelo presentado tuvo un error promedio del 13%, lo que lo convierte en un modelo moderado de estimación de las prescripciones de los medicamentos para el cáncer de mama entrantes al plan de beneficios (PBS). Esto quiere decir que cuenta con una precisión y exactitud media y es una excelente guía para el tipo de actividades que se van a realizar con respecto a los datos obtenidos.

Este modelo va a permitir la realización de diferentes actividades dentro de Roche que generan un alto impacto como la toma de decisiones estratégicas, llevar a cabo acciones de marketing, realizar lanzamientos de otros medicamentos que contienen los mismos principios activos y estudios de mercado, lo que garantiza una ventaja frente a otras farmacéuticas que no cuentan con la estimación de las prescripciones de sus medicamentos incluidos en el plan de beneficios (PBS).

Una desventaja de este modelo es que incluye un gran número de variables que requieren ser estudiadas minuciosamente para contar con una buena estimación de los medicamentos que se encuentran dentro del plan de beneficios (PBS), sin embargo, una vez estudiadas y adquiridas correctamente, es factible llevar a cabo el modelo de estimación.

Se pudo observar que la metodología CRISP-DM es una de las mejores metodologías existentes para llevar a cabo proyectos relacionados con la minería y la ciencia de datos ya que permite generar tareas específicas de un tema fácilmente y concretar una ruta a seguir para el cumplimiento de los objetivos planteados en el proyecto.

## REFERENCIAS

- [1] "Quienes somos". Home | Roche.  
[https://www.roche-cac.com/es/compania/\\_quienes\\_somos.html](https://www.roche-cac.com/es/compania/_quienes_somos.html)
- [2] Á. J. Riascos, A. J. y S. A. Camelo, S. A., "Una estimación del costo y cambios en el bienestar de los colombianos con el nuevo Plan de Beneficios en Salud", *Revista de Economía del Rosario*, vol. 17, n.º 24, pp. 229–314, mayo de 2015. [En línea]. Disponible:  
<https://revistas.urosario.edu.co/index.php/economia/article/view/3747/2701>
- [3] Colombia, Ministerio de salud y protección social. (2015, 16 de febrero). *Ley n.º 1751, Salud*.  
[https://www.minsalud.gov.co/Normatividad\\_Nuevo/Ley%201751%20de%202015.pdf](https://www.minsalud.gov.co/Normatividad_Nuevo/Ley%201751%20de%202015.pdf)
- [4] C. Rodríguez Moreno y C. Durán Parrondo, "Genéricos y prescripción por principio activo", Santiago de Compostela, 2008. [En línea]. Disponible:  
[https://www.agamfec.com/pdf/CADERNOS/VOL15/PDF\\_WEB\\_VOL15\\_N4/04%20Para%20Saber%20de.pdf](https://www.agamfec.com/pdf/CADERNOS/VOL15/PDF_WEB_VOL15_N4/04%20Para%20Saber%20de.pdf)
- [5] V. H. Mora. "Medicamentos que no están en el Plan de Beneficios en Salud". Ministerio de Salud y Protección Social.  
<https://www.minsalud.gov.co/Lists/FAQ/DispForm.aspx?ID=1089&ContentTypeld=0x01003F0A1BD895162D4599DC199234219AC7>
- [6] O. J. Salcedo Parra, R. M. Galeano y L. G. Rodriguez, "Metodología crisp para la implementación Data Warehouse", *scielo*, vol. 14, n.º 26, junio de 2010. [En línea]. Disponible:  
[http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S0123-921X2010000100005](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-921X2010000100005)
- [7] A. Sanchez Ortiz. El proceso de la Metodología CRISP-DM (IBM, 2020). (diciembre de 2020). [Imagen]. Disponible:  
[https://www.researchgate.net/figure/El-proceso-de-la-Metodologia-CRISP-DM-IBM-2020\\_fig3\\_347529684](https://www.researchgate.net/figure/El-proceso-de-la-Metodologia-CRISP-DM-IBM-2020_fig3_347529684)
- [8] "Prueba de cáncer de seno HER2: Prueba de laboratorio de MedlinePlus". MedlinePlus - Health Information from the National Library of Medicine.  
<https://medlineplus.gov/spanish/pruebas-de-laboratorio/prueba-de-cancer-de-seno-her2/>
- [9] A. Goodman, "Perjeta: A New Option for Patients with HER2-Positive Metastatic Breast Cancer", *Amer. Health & Drug Benefits*, vol. 6, n.º 3, abril de 2013. [En línea]. Disponible:  
<https://www.ahdbonline.com/issues/2013/april-2013-vol-6-no-3-payers-guide-to-new-fda-approvals/1386-article-1386>
- [10] W. Dean-Colomb y F. J. Esteva, "Her2-positive breast cancer: Herceptin and beyond", *Eur. J. Cancer*, vol. 44, n.º 18, pp. 2806–2812, diciembre de 2008. Accedido el 16 de mayo de 2023. [En línea]. Disponible: <https://doi.org/10.1016/j.ejca.2008.09.013>
- [11] P. Miranda Romero y R. Marín Gil, "Trastuzumab emtansine in locally advanced or metastatic HER2 positive breast cancer; GENESIS-SEFH drug evaluation report", *Sociedad Española de Farmacia Hospitalaria*, vol. 39, n.º 3, 2015. [En línea]. Disponible:  
<https://www.redalyc.org/articulo.oa?id=365961302003>
- [12] "IQVIA Colombia". Powering Healthcare with Connected Intelligence - IQVIA.  
<https://www.iqvia.com/es-co/locations/colombia>
- [13] "WHOCC - ATC/DDD Index". WHOCC - Home.  
[https://www.whocc.no/atc\\_ddd\\_index/?code=L01XY02](https://www.whocc.no/atc_ddd_index/?code=L01XY02)

## ANEXOS

### ANEXO 1. Diagrama de flujo del código

