



**UNIVERSIDAD PERUANA DE CIENCIAS APLICADAS**

**FACULTAD DE NEGOCIOS**

**PROGRAMA ACADÉMICO DE ADMINISTRACIÓN DE EMPRESAS**

**PROGRAMA ACADÉMICO DE ADMINISTRACIÓN Y GERENCIA  
DEL EMPRENDIMIENTO**

“Aplicación de Data Science para el pronóstico de las devoluciones en soles de los  
cinco principales clientes de la empresa GSK”

**TRABAJO DE INVESTIGACIÓN**

Para optar el grado de bachiller en Administración de Empresas,  
Para optar el grado de bachiller en Administración y Gerencia del Emprendimiento

**AUTOR(ES)**

Ayala Campos, Mayra Alexandra (0000-0001-9472-2466)

Calderón Zavala, Jonathan Rolando (0000-0002-5688-0305)

Núñez Espinal, Fernando Javier (0000-0001-8245-1179)

Rivera Gonzáles, Tatiana Alexandra (0000-0002-5594-8443)

Rodríguez Chávez, Juan Ernesto (0000-0002-2854-2974)

**ASESOR**

Palacios Ruiz, Julio Cesar (0000-0003-1610-5205)

**Lima, 03 de diciembre de 2021**

## *DEDICATORIA*

*El presente trabajo de investigación se lo dedicamos, en primer lugar, a Dios por ser un guía constante en nuestras acciones y nos permite continuar luchando por nuestros sueños. En segundo lugar, a nuestros familias por ser siempre un apoyo emocional y de fortaleza que nos impulsan y motivan al logro de nuestros objetivos.*

## AGRADECIMIENTOS

Queremos agradecer a las personas que nos han acompañado en este proceso arduo pero satisfactorio de la carrera profesional. Asimismo, queremos agradecer a la Universidad Peruana de Ciencias Aplicadas por anteponer la continuidad de los estudios académicos frente a una situación poco favorable y, en especial, al profesor Julio César Palacios Ruiz, quien con su enseñanza y predisposición se pudo sacar adelante el presente proyecto.

## Resumen Ejecutivo

La presente investigación busca analizar el desempeño de la devolución de los productos farmacéuticos por parte de los principales clientes de la empresa GlaxoSmithKline – GSK durante el periodo enero 2019 a junio 2021, tomando en consideración las diferentes variables que involucran su comportamiento.

El objetivo del proyecto se enfoca en generar un modelado de datos que permita otorgar información oportuna a la empresa, con la finalidad que esta pueda generar conocimiento y establecer estrategias que permitan dar solución al problema. Para ello, la metodología utilizada para desarrollar la investigación es un enfoque predictivo, puesto que nos ayudó a describir, entender y proponer soluciones para el problema identificado.

Se desarrolló un marco teórico que nos ayudó a entender el negocio y el sector en el que se encuentra, así como la descripción y presentación de la empresa. Además, se procedió con la comprensión del negocio, sustentado con datos, que nos dará una base sólida para definir el problema y establecer los objetivos que se deseaban lograr.

La técnica supervisada usada para resolver el problema fue de regresión lineal múltiple, el cual nos permitió conocer las variables que dan respuesta a nuestra pregunta de ciencia de datos, que están determinadas por 10 variables independientes. Se presentaron diferentes visualizaciones para representar los hallazgos obtenidos.

Palabras clave: Ciencia de datos, técnica supervisada, regresión lineal, modelo predictivo

## Abstract

The present research seeks to analyze the performance of the return of pharmaceutical products by the main customers of the company GlaxoSmithKline - GSK during the period January 2019 to June 2021, taking into consideration the different variables that involve their behavior.

The objective of the project is focused on generating a data modeling that allows to provide timely information to the company, so that it can generate knowledge and establish strategies to solve the problem. For this, the methodology used to develop the research is a predictive approach, since it helped us to describe, understand and propose solutions to the identified problem.

A theoretical framework was developed to help us understand the business and the sector in which it is located, as well as the description and presentation of the company. In addition, we proceeded with the understanding of the business, supported with data, which will give us a solid basis to define the problem and establish the objectives to be achieved.

The supervised technique used to solve the problem was multiple linear regression, which allowed us to know the variables that answer our data science question, which are determined by 10 independent variables. Different visualizations were presented to represent the findings obtained.

Key words: data science, supervised technique, linear regression, predictive modeling

## Tabla de contenido

Resumen Ejecutivo .....	4
Abstract.....	5
<b>1. MARCO TEÓRICO.....</b>	<b>10</b>
1.1. Metodología de ciencia de datos.....	10
1.2. Entendimiento del negocio y solución del problema.....	11
1.2.1. Definición de industrias farmacéuticas.....	12
1.2.2. Ente regulador .....	12
1.2.3. Tendencias y Contextualización.....	12
1.2.4. Desafíos .....	13
<b>2. PRESENTACIÓN DE LA EMPRESA PARA EL PROYECTO.....</b>	<b>14</b>
2.1. Presentación y descripción de la empresa.....	14
2.2. Modelo de negocio.....	15
2.3. Objetivos de la empresa.....	17
2.4. Análisis externo .....	17
2.4.1. Análisis PESTEL.....	17
<b>3. COMENZANDO EL PROYECTO .....</b>	<b>19</b>
3.1. Identificación del problema – Comprendiendo el negocio y el problema .....	19
3.2. Formulación de preguntas para Data Science:.....	21
3.3. Posibles soluciones al problema, basadas en ciencia de datos.....	21
3.4. Justificación .....	22
3.5. Objetivos del proyecto .....	22
3.6. Hipótesis basadas en el modelo de ciencia de datos .....	22
<b>4. REQUISITOS Y RECOPIACIÓN DE DATOS .....</b>	<b>23</b>
4.1. Determinación de datos para responder a las preguntas de ciencias de datos .	23
4.2. Ubicación de fuentes de datos.....	23
4.3. Presentación de datos recopilados .....	24
<b>5. COMPRENSIÓN Y PREPARACIÓN DE DATOS.....</b>	<b>25</b>
5.1. Diccionario de Datos.....	25
5.2. Calidad de Datos .....	26
5.2.1. Revisión y Depuración de Datos .....	26
5.2.2. Niveles de calidad.....	27
5.3. Gobierno de datos .....	27
<b>6. PRODUCCIÓN Y ANÁLISIS EXPLORATORIO DE DATOS.....</b>	<b>28</b>
6.1. Determinación de los clientes principales para el estudio .....	28
6.2. Presentación de datos a utilizar para el estudio .....	29
6.3. Plan de muestreo y tamaño de la muestra.....	30

6.4.	Diseño del estudio de muestra .....	30
6.5.	Ejecución del muestreo .....	31
6.6.	Resumen y examen de la distribución de variables .....	32
6.7.	Resultado del análisis de relaciones.....	36
<b>7.</b>	<b>COMENZANDO LA INTERPRETACIÓN DE DATOS .....</b>	<b>43</b>
<b>8.</b>	<b>ARQUITECTURA DE DATOS .....</b>	<b>46</b>
8.1.	Arquitectura funcional de GSK .....	46
8.2.	Arquitectura tecnológica, herramientas e infraestructura .....	47
<b>9.</b>	<b>TÉCNICAS DE CIENCIA DE DATOS .....</b>	<b>48</b>
9.1.	Supervisado.....	48
9.1.1.	Regresión lineal múltiple.....	48
<b>10.</b>	<b>EL MODELO Y LA HISTORIA .....</b>	<b>51</b>
10.1	Determinación del modelo más apropiado para la descripción o predicción .....	51
10.2	Descripción de los hallazgos, ¿Qué se descubrió? .....	51
<b>11.</b>	<b>IMPLEMENTACIÓN Y AJUSTES .....</b>	<b>52</b>
11.1	Visualización de los hallazgos, grafique lo descubierto y compártalo.....	52
11.2	Estrategia de implementación.....	62
<b>12.</b>	<b>CALIBRADO E IMPLEMENTACIÓN.....</b>	<b>63</b>
12.1	Método de ajuste.....	63
12.2	Retroalimentación.....	63
<b>13.</b>	<b>FACTORES CRÍTICOS DE ÉXITO DEL PROYECTO .....</b>	<b>64</b>
<b>14.</b>	<b>CONCLUSIONES .....</b>	<b>64</b>
<b>15.</b>	<b>NUEVAS PREGUNTAS .....</b>	<b>65</b>
<b>16.</b>	<b>DECISIONES PARA TOMAR A PARTIR DE LOS DATOS ESTUDIADOS</b> 65	
<b>17.</b>	<b>ANEXO.....</b>	<b>67</b>
<b>18.</b>	<b>REFERENCIAS BIBLIOGRÁFICAS .....</b>	<b>70</b>

## Índice de imágenes

Imagen 1. Metodología de ciencia de datos	11
Imagen 2. Ranking de empresas farmacéuticas por valor de marca 2020	13
Imagen 3. Consulta RUC GlaxoSmithKline Perú SA	15
Imagen 4. Modelo Canvas GSK	16
Imagen 5. Ventas totales de GSK. Enero 2019 – junio 2021	20
Imagen 6. Devoluciones totales a GSK. Enero 2019 – junio 2021	20
Imagen 7. Proporción de devolución por clientes	21
Imagen 8. Retorno de productos	24
Imagen 9. Ventas de productos en unidades	24
Imagen 10. Ventas de GSK a clientes	28
Imagen 11. Muestra de tablón de análisis	29
Imagen 12. Propuesta de diseño de muestra	30
Imagen 13. Caja y bigotes. Ventas en soles de GSK a sus clientes	32
Imagen 14. Caja y bigotes. Ventas en unidades de GSK a sus clientes	32
Imagen 15. Caja y bigotes. Ventas en soles de los clientes de GSK a sus consumidores	33
Imagen 16. Caja y bigotes. Venta en unidades de los clientes de GSK a sus consumidores	33
Imagen 17. Caja y bigotes. Devoluciones en soles de los clientes a GSK	34
Imagen 18. Caja y bigotes. Precio promedio de venta de GSK a sus clientes	34
Imagen 19. Campana de Gauss. Sell in en soles	35
Imagen 20. Campana de Gauss. Precio promedio	35
Imagen 21. Campana de Gauss. Sell out en soles	36
Imagen 22. Campana de Gauss. Devoluciones en soles de los clientes a GSK	36
Imagen 23. Correlación entre Devoluciones vs Devoluciones Deco Aquafresh en miles de soles	38
Imagen 24. Correlación entre Devoluciones vs Devoluciones Deco Panadol en miles de soles	38
Imagen 25. Correlación entre Devoluciones vs Dev por vencimiento Deco en miles de soles	39
Imagen 26. Correlación entre Devoluciones vs Devoluciones F del Norte Aquafresh en miles de soles	39
Imagen 27. Correlación entre Devoluciones vs Devoluciones F del Norte Scott en cientos de soles	40
Imagen 28. Correlación entre Devoluciones vs Devoluciones Química Suiza Panadol en miles de soles	40
Imagen 29. Correlación entre Devoluciones vs Devoluciones por vencimiento Química Suiza en miles de soles	41
Imagen 30. Correlación entre Devoluciones vs Dev por sobre stock Química Suiza en miles de soles	41
Imagen 31. Correlación entre Devoluciones vs PBI en miles de soles	42
Imagen 32. Correlación entre Devoluciones vs Devoluciones promedio de devoluciones por fallas 6 meses en cientos de soles	42
Imagen 33. Correlación entre Devoluciones vs Sell-in Dimexa Corega en miles de soles	43
Imagen 34. Arquitectura funcional de GSK	46
Imagen 35. Arquitectura tecnológica GSK	47
Imagen 36. Clientes principales de GSK y participación.	52
Imagen 37. Tipos de devoluciones de Alfaro	53
Imagen 38. Tipos de devoluciones de Deco	53
Imagen 39. Tipo de devolución de Dimexa	54
Imagen 40. Tipo de devolución de F del Norte	54
Imagen 41. Tipo de devolución de Química Suiza	55
Imagen 42. Tabla de dispersión de las variables del modelo	56
Imagen 43. Análisis descriptivo de las variables del modelo	57
Imagen 44. Matriz de correlaciones de las variables del modelo	57
Imagen 46. Matriz de correlaciones. Variables para el modelo	58
Imagen 45. Campana de Gauss de las variables del modelo	58
Imagen 47. Diagnóstico de residuos	61

## Índice de tablas

<i>Tabla 1. Diccionario de datos</i> .....	25
<i>Tabla 2. Muestra de matriz de correlaciones</i> .....	37
<i>Tabla 3. Matriz de correlaciones – Mapa de calor</i> .....	37
<i>Tabla 4. Estadística descriptiva 1</i> .....	44
<i>Tabla 5. Estadística descriptiva 2</i> .....	44
<i>Tabla 6. Estadística descriptiva 3</i> .....	45
<i>Tabla 7. Estadística descriptiva 4</i> .....	45
<i>Tabla 8. Estadística descriptiva 5</i> .....	45
<i>Tabla 9. Estadística descriptiva 6</i> .....	46

## 1. MARCO TEÓRICO

### 1.1. Metodología de ciencia de datos

La ciencia de datos se ha convertido actualmente en una herramienta importante para las organizaciones, ya que permite la exploración y análisis de datos con la finalidad de identificar *insights* y generar conocimiento. García et al (2018, p. 5) indica que uno de los objetivos que busca la ciencia de datos es “la búsqueda de modelos que describan patrones y comportamientos a partir de los datos con el fin de tomar decisiones o hacer predicciones”.

Para poder desarrollar el presente trabajo y dar solución a nuestra pregunta de investigación basado en un estudio de datos exhaustivo, aplicaremos un marco que nos servirá como guía para obtener las respuestas esperadas, es decir, la metodología de ciencia de datos de IBM.

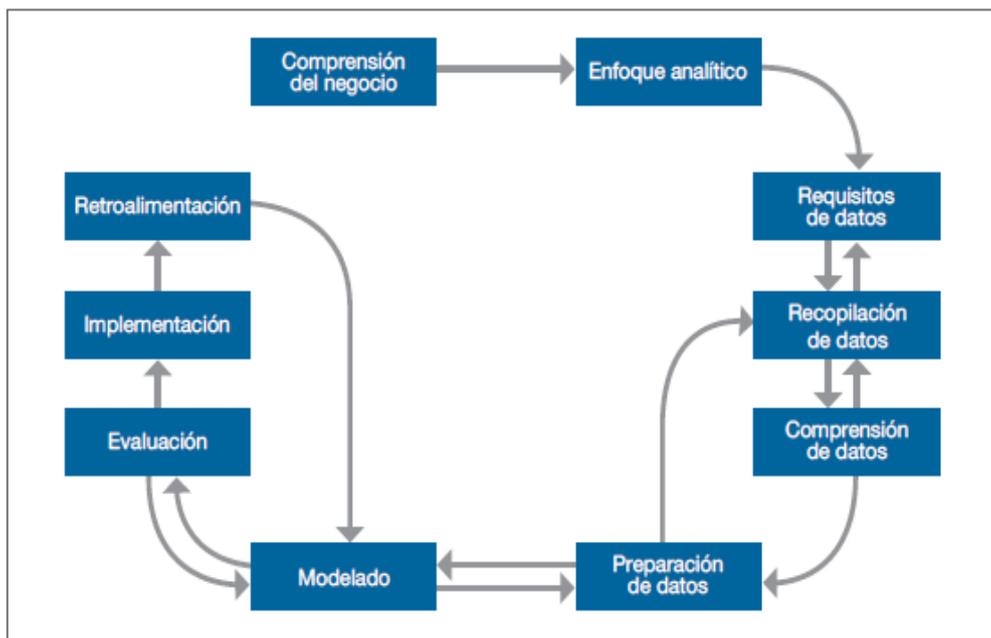
Esta metodología es un proceso iterativo y secuencial que permiten obtener de manera acertada *insights* y que nos llevará a generar soluciones mediante el análisis de datos (IBM, 2015). Este marco es una estrategia que consiste en una secuencia de pasos importantes para el proceso y desarrollo del análisis de datos que facilitan la obtención de una respuesta o resultado en una situación determinada.

Este proceso involucra 10 etapas divididos en 3 módulos, tal como se describen a continuación y de manera resumida en la Imagen 1. (IBM, 2015):

- Módulo 1: Del problema al enfoque
  - Comprensión empresarial: Permite entender el negocio con la finalidad de determinar el problema y establecer los objetivos del proyecto.
  - Enfoque analítico: Permite definir la metodología según la comprensión empresarial.
- Módulo 2: Trabajo con los datos
  - Requisitos de datos: Permite encontrar los contenidos de datos necesarios para generar conocimiento.
  - Recopilación de datos: Permite la búsqueda de fuentes y obtención de data, ya sean estructurado, no estructurado y/o semi estructurado.
  - Comprensión de datos: Permite entender los datos a través de herramientas estadísticas y visualizaciones.

- Preparación de datos: Permite la construcción de los datos, desde la limpieza, tratamiento, combinación y transformación de estos.
- Módulo 3: Obtención de la respuesta
  - Modelado: Permite desarrollar los modelos descriptivos o predictivos según el enfoque analítico establecido.
  - Evaluación: Permite confirmar la calidad del modelado y que su aplicabilidad sea de manera adecuada, a través de medidas de diagnóstico.
  - Implementación: Permite la ejecución del modelado, luego de su aprobación, de manera progresiva y transversal.
  - Comentarios: Permite la recepción de retroalimentación respecto al rendimiento e impacto del modelo, con la finalidad de realizar los ajustes posteriores para una mayor fidelidad.

*Imagen 1. Metodología de ciencia de datos*



*Fuente y elaboración: IBM, 2015*

## 1.2. Entendimiento del negocio y solución del problema

Para tener un mayor alcance de nuestra investigación, en este apartado se explica la definición, funciones, entes reguladores, tendencias y desafíos de las industrias farmacéuticas, con la finalidad de ayudarnos a tener un mejor entendimiento del tema a tratar.

### 1.2.1. Definición de industrias farmacéuticas

La industria farmacéutica, es definida como la “rama de la industria que se dedica al descubrimiento, desarrollo, fabricación y la comercialización de medicamentos para la salud” (Quiroa, s.f.). Para ello, se efectúa una serie de investigaciones, apoyados por la innovación, que permitan elaborar medicinas y vacunas para prevenir y conservar la salud humana.

Jara (2015) la describe como el sector enfocado en la fabricación, elaboración y venta de productos químicos medicinales para tratar las enfermedades y prevenirlas. A través, de actividades de investigación y desarrollo para obtener tratamientos más eficientes y cumplir con los objetivos de la empresa.

### 1.2.2. Ente regulador

Si bien es cierto, cada país mantiene sus propias regulaciones y políticas para velar por la calidad y eficacia de los medicamentos, la Organización Mundial de la Salud – OMS presenta las funciones principales que cada Autoridad Reguladora de Medicamentos debe cumplir, ya sea como una Agencia Reguladora o como un Organismo de Regulación de Medicamentos. Estas son:

- Registro de productos
- Inspección y concesión de licencias a fabricantes
- Inspección y concesión de licencias a distribuidores
- Atención post comercialización
- Reglamentación para la presentación de beneficios para la promoción comercial de los productos
- Autorización de los ensayos clínicos

En el caso de nuestro país, Perú, el ente regulador que cumple las funciones presentadas en el párrafo anterior, es la Dirección General de Medicamentos, Insumos y Drogas – DIGEMID del Ministerio de Salud – MINSA.

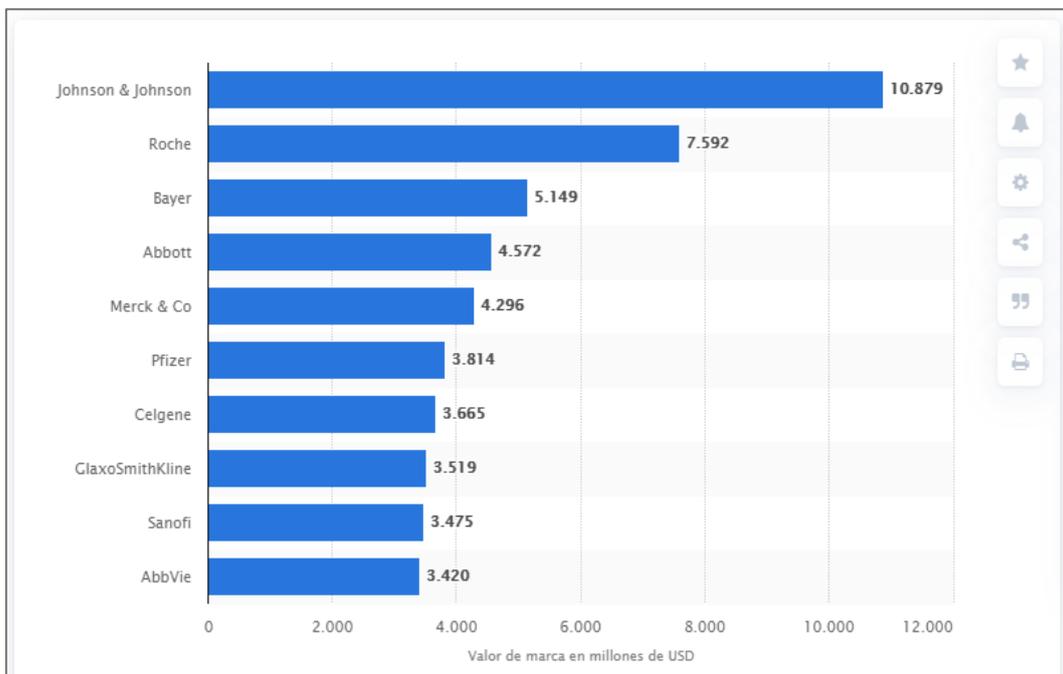
### 1.2.3. Tendencias y Contextualización

Actualmente, a medida que avanza la tecnología y las necesidades de innovación de las empresas en sus procesos y productos, la industria farmacéutica también se ha visto beneficiada con esta transformación. En consecuencia, se están presentando tendencias que este sector debería tomar en cuenta para no quedar relegados, como lo menciona Quiroa (s.f.):

- Especialización de productos de medicamentos
- Embalajes que incrementan la vida útil de producto
- Registro del conocimiento como propiedad intelectual como valor agregado
- Uso de la tecnología para el desarrollo y elaboración de medicamentos

Según Statista (2021), GlaxoSmithKline se encuentra dentro del ranking mundial de las principales empresas farmacéuticas en el año 2020, ocupando la octava posición con un valor de marca de 3,519 millones de dólares, detrás de grandes empresas como Johnson & Johnson, Roche, Merck & Co., Pfizer, y otros, tal como se muestra en la Imagen 2.

Imagen 2. Ranking de empresas farmacéuticas por valor de marca 2020



Fuente y elaboración: Statista, 2021

#### 1.2.4. Desafíos

La industria farmacéutica también presenta desafíos que las grandes compañías alrededor del mundo no deben dejar de lado, aún más en la situación que ha presentado crisis sanitaria de la Covid-19. Estos desafíos, según Quiroa (s.f.) son:

- Largos periodos requeridos para producir nuevos medicamentos
- Necesidad de implementar alta tecnología y sus costos elevados
- Ciclos de vida del productos cada vez más cortas
- Nuevos competidores

## 2. PRESENTACIÓN DE LA EMPRESA PARA EL PROYECTO

### 2.1. Presentación y descripción de la empresa

GlaxoSmithKline – GSK es una compañía transnacional farmacéutica y atención médica, con alrededor de 100,000 colaboradores a nivel global, que se dedica al cuidado de la salud de las personas, guiado por la ciencia, tecnología e innovación, desarrollándose en tres grandes áreas de negocio como son las medicinas, vacunas y productos sanitarios, con la consigna de buscar no solo el bienestar físico, sino mental. La compañía inicia su historia en el año 1715, cuyo representante Silvanus Bevan constituye una farmacia llamada “Plow Court” ubicada en la ciudad de Londres – Inglaterra, la cual, un centenar de años más tarde se haría conocida como Allen & Hanburys Ltd. Posteriormente, la farmacia fue adquirida por Glaxo Laboratories Ltd., para que finalmente nazca GSK debido a la fusión entre Glaxo Wellcome y SmithKline Beecham en el año 2000.

Entre los eventos más resaltantes y de alta relevancia que la compañía GSK ha registrado a lo largo de su historia, que involucra investigaciones, desarrollo y adquisición de conocimientos, podemos mencionar los siguientes:

- Desde ya hace más de un siglo, la empresa busca generar medicamentos y vacunas, a través de sus laboratorios de investigación y desarrollo, en contra de la malaria.
- Hace poco más de 40 años, la compañía GSK en colaboración con Pfizer Inc. y Shionogi Limited, crearon la empresa ViiV Healthcare, quienes vienen trabajando en nuevas maneras de tratar y prevenir la transmisión del Virus de Inmunodeficiencia Humana – VIH, causante del SIDA.
- En el año 2015, terminaron el proceso de adquisición del modelo de negocio de vacunas de la empresa Novartis, convirtiendo a GSK en el líder mundial respecto a la fabricación de vacunas.

Actualmente, cuenta con más de 100 sedes distribuidos en diferentes regiones alrededor del mundo, siendo algunas de ellas las ubicadas en Inglaterra, España, Alemania, Estados Unidos, Turquía, Sri Lanka, Croacia, México, Perú, entre otros.

La Imagen 3 presentada a continuación, muestra la descripción básica registrada en la Superintendencia Nacional de Aduanas y Administración Tributaria – SUNAT referida a su sede en Lima – Perú.

Imagen 3. Consulta RUC GlaxoSmithKline Perú SA

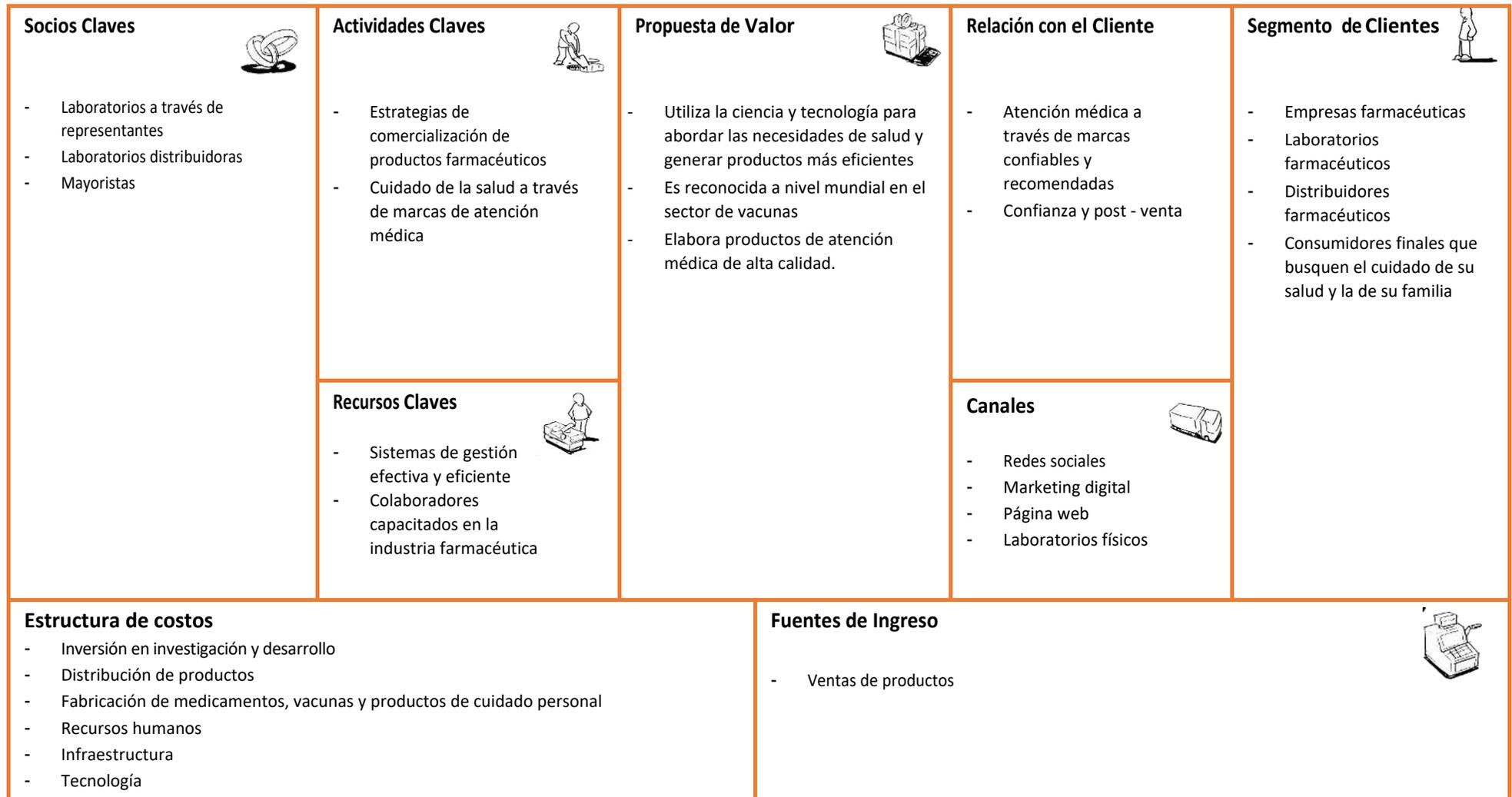
Resultado de la Búsqueda			
Número de RUC:	20100123682 - GLAXOSMITHKLINE PERU S.A.		
Tipo Contribuyente:	SOCIEDAD ANONIMA		
Nombre Comercial:	GLAXOSMITHKLINE		
Fecha de Inscripción:	09/10/1992	Fecha de Inicio de Actividades:	20/04/1967
Estado del Contribuyente:	ACTIVO		
Condición del Contribuyente:	HABIDO		
Domicilio Fiscal:	AV. JAVIER PRADO OESTE NRO. 995 URB. ORRANTIA LIMA - LIMA - SAN ISIDRO		
Sistema Emisión de Comprobante:	MECANIZADO	Actividad Comercio Exterior:	IMPORTADOR/EXPORTADOR
Sistema Contabilidad:	COMPUTARIZADO		
Actividad(es) Económica(s):	Principal - 4690 - VENTA AL POR MAYOR NO ESPECIALIZADA		
	Secundaria 1 - 4772 - VENTA AL POR MENOR DE PRODUCTOS FARMACÉUTICOS Y MÉDICOS, COSMÉTICOS Y ARTÍCULOS DE TOCADOR EN COMERCIOS ESPECIALIZADOS		

Fuente y elaboración: Superintendencia Nacional de Aduanas y Administración Tributaria – SUNAT, 2021

## 2.2. Modelo de negocio

A continuación, en la Imagen 4, se muestra el modelo de negocios de la empresa GlaxoSmithKline Perú aplicado en la herramienta Business Model Canvas, la cual nos ayudará a entender de una manera clara, precisa y contundente el desarrollo de sus operaciones en sus tres grandes áreas de negocio, como son la elaboración y comercialización de medicamentos, vacunas y productos de cuidado personal innovadores.

Imagen 4. Modelo Canvas GSK



Fuente: GSK, elaboración propia

### 2.3. Objetivos de la empresa

**MISIÓN:** “Somos una compañía de atención médica global liderada por la ciencia con un propósito especial de mejorar la calidad de vida humana al ayudar a las personas a hacer más, sentirse mejor y vivir más tiempo” (GSK, 2020).

**VISIÓN:** “Llevar productos de atención médica diferenciados, de alta calidad y necesarios a la mayor cantidad de personas posible, previniendo y tratando enfermedades y manteniendo a las personas bien con nuestro conocimiento científico y técnico y nuestra gente talentosa.” (GSK, 2020).

**VALORES:**

- Respeto
- Transparencia
- Integridad
- Enfoque al paciente
- Responsabilidad

### 2.4. Análisis externo

#### 2.4.1. Análisis PESTEL

##### **Político**

Actualmente el Perú se encuentra en una inestabilidad política ya que después de las elecciones presidenciales 2021 existe mucha incertidumbre de las nuevas políticas que se llevarán a cabo para reactivar la economía del Perú en este nuevo gobierno. Lo que si se ha estado dando son las políticas relacionadas a la COVID-19 para controlar las exportaciones y el proceso de vacunación. A raíz de esta pandemia hubo niveles de regulación en el sector salud, además de la interferencia de fabricantes de medicamentos y sobre todo el etiquetado de medicamentos principales para contrarrestar la COVID-19.

##### **Económico**

La tasa de inflación, tasa de ahorro, tasa de interés, tasa de cambio y ciclo económico determinan la demanda y la inversión agregadas. La empresa GSK puede utilizar el factor económico del país y los indicadores económicos de la industria como Fabricantes de medicamentos (Tasa de crecimiento de la industria principal, gasto del consumidor, etc.), para pronosticar la trayectoria de crecimiento no solo del sector sino también de la organización. Pese a caídas, el Perú, presenta un escenario prometedor para la

reactivación ya que el Fondo Monetario Internacional ha indicado que el Perú tendría el repunte más elevado de América según el último reporte del organismo multilateral. A la vez, la empresa tiene que considerar la tasa de desempleo, la intervención gubernamental en el mercado libre y la asistencia sanitaria relacionada, el tipo de cambio, la estabilidad de la moneda ya que actualmente hay incremento en el tipo de cambio.

### **Social**

La pandemia dejó en evidencia la jerarquía y estructura de poder en la sociedad que existe en el Perú, sobre todo en el sector de Educación, siendo los más perjudicados los niveles más bajos. Esto trajo como consecuencia más informalidad e inseguridad donde el gobierno tendrá que intervenir para reducir brechas sociales. En el aspecto cultural, en esta situación actual ha cambiado totalmente la manera de exponerse donde se involucró la capacidad de transformación y reinención a un plano digital debido al impacto que la pandemia tuvo en el país.

### **Tecnológico**

La tecnología está alterando rápidamente varias industrias en todos los ámbitos. En el Perú tanto las empresas como las personas han tenido que modernizarse en el último año. Las empresas se han tenido que adecuar y elaborar procedimientos digitales, el E-Commerce ha tenido uno de los mayores crecimientos, así como el consumo online. La empresa deberá considerar el impacto de la tecnología en la oferta de productos, el impacto en la estructura de costos en los fabricantes de medicamentos y el impacto en la estructura de la cadena de valor en el sector de la salud.

### **Ecológico**

En el aspecto ecológico se debe tener en cuenta las leyes de ambientales y leyes de responsabilidad. La conservación del entorno y la lucha contra el cambio climático se han convertido para las empresas farmacéuticas en un verdadero activo estratégico ya que su actividad industrial depende en gran medida de los recursos naturales y los servicios ecosistémicos, además es uno de los sectores que más invierte en I+D.

### **Legal**

En relación con aspecto legal, se tiene dos actores importantes en la industria farmacéutica, uno es la Dirección General de Medicamentos, Insumos y Drogas (DIGEMID) como entidad reguladora de medicamentos cuyo objetivo principal es garantizar que la población peruana acceda a medicamentos seguros, eficaces, de calidad, y el uso racional de los mismos (DIGEMID,2019). Otro actor es el Instituto Nacional de Defensa de la Competencia y de la Protección de la Propiedad Intelectual (INDECOPI)

como ente regulador de la libre competencia cuyo propósito es promover la competencia leal y honesta en el mercado peruano, protegiendo toda propiedad intelectual, cuenta con autonomía técnica, económica, presupuestal y administrativa (INCEPOPI, 2016).

Este último año se han aprobado varias leyes que buscan hacer frente a la pandemia, así como a lo que ha provocado a la economía nacional y de las familias. Entre las principales leyes del 2020 fueron las que declararon en estado de emergencia el territorio nacional, la ley que permitió el retiro de las AFP's, lo que permite al Banco de la Nación la apertura de cuentas de ahorro.

### **3. COMENZANDO EL PROYECTO**

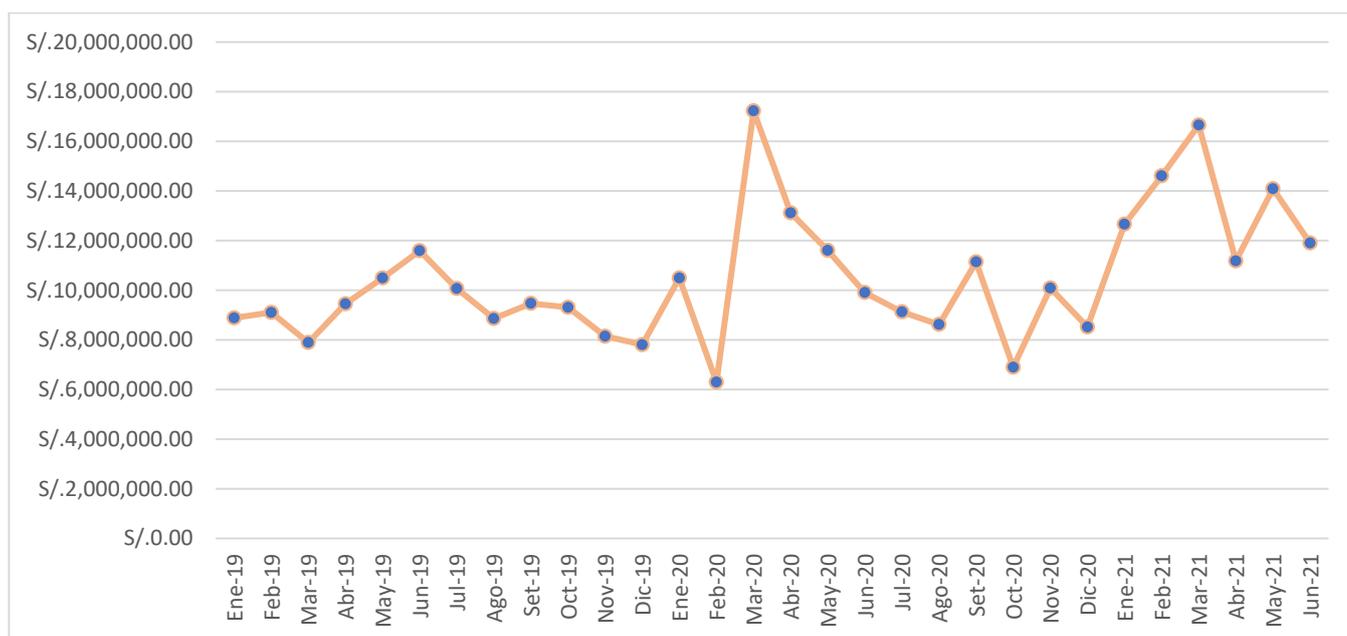
#### **3.1. Identificación del problema – Comprendiendo el negocio y el problema**

El sector al que evaluaremos es la industria farmacéutica, la cual se enfoca en productos de medicina sin prescripción médica. Por ejemplo, la venta productos en la categoría de “Dolor de Cabeza” como Panadol, “Dolor Muscular” como Voltaren, entre otros. Así también, con productos de la categoría de “Salud Oral” como Sensodyne para la sensibilidad dental o Corega y “Limpiadores” para las prótesis dentales. Y otras categorías como la “Dermales” (Lamisil) o “Vitaminas” (Scott y Centrum).

GSK ofrece estos productos a sus distintos clientes/distribuidores como Química Suiza, Alfaro, entre otros, con los que mantienen un tipo de acuerdo comercial donde éstos tienen la posibilidad de devolver aquellos productos que no son vendidos en un determinado periodo. Partiendo de esta premisa, se observó la existencia de devoluciones de productos por parte de los clientes, ocasionando que el total de los productos devueltos generen pérdidas a la empresa y un gran impacto financiero por devolución o costos de destrucción de productos que retornan al almacén.

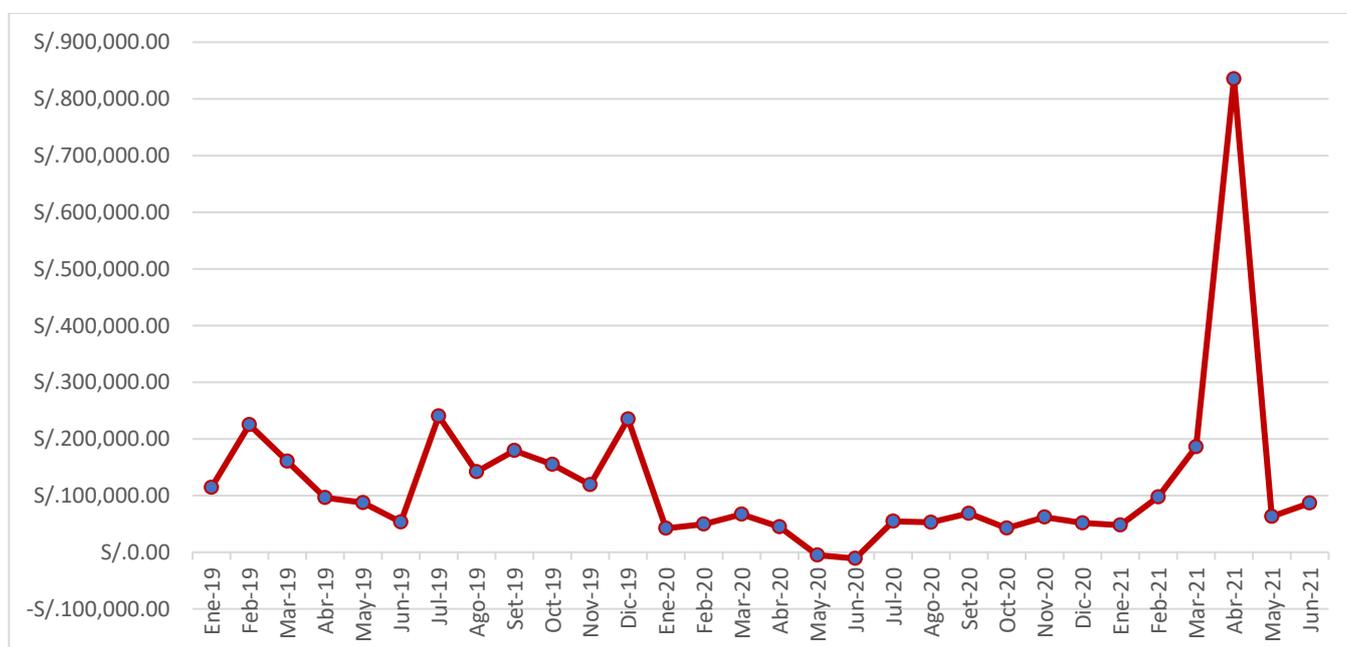
En la Imagen 5, se muestra un cuadro de productos devueltos a la compañía por parte de los diferentes clientes que mantienen este tipo de concesiones como, por ejemplo: Química Suiza, Deco, Alfaro, y otras empresas farmacéuticas que vienen acumulando este tipo de prácticas en cada periodo.

Imagen 5. Ventas totales de GSK. Enero 2019 – junio 2021



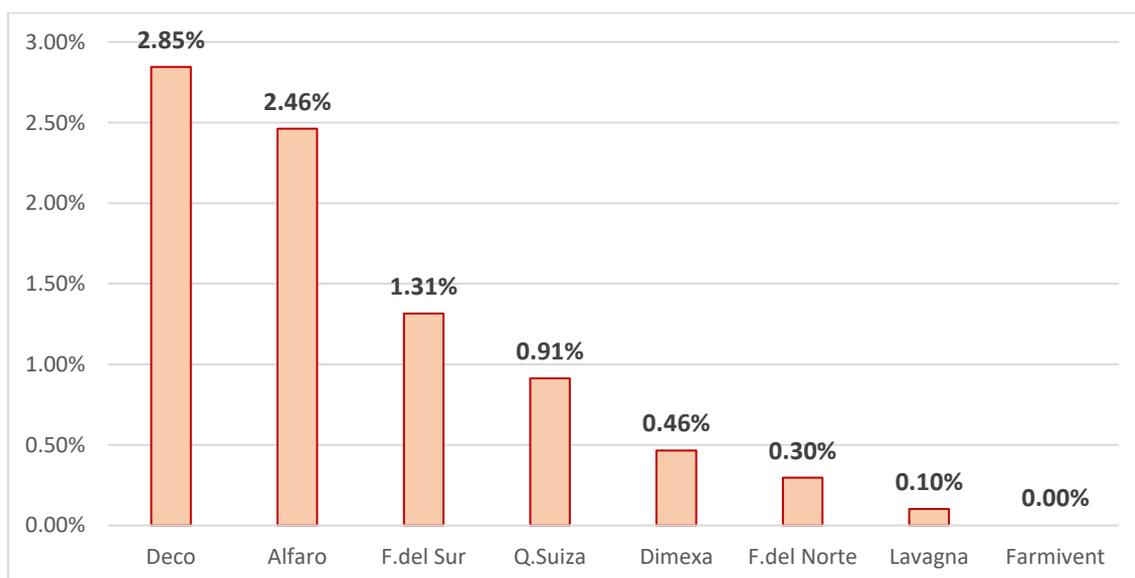
Fuente: GSK  
Elaboración propia

Imagen 6. Devoluciones totales a GSK. Enero 2019 – junio 2021



Fuente: GSK  
Elaboración propia

Imagen 7. Proporción de devolución por clientes



Fuente: GSK  
Elaboración propia

### 3.2. Formulación de preguntas para Data Science:

Pregunta Principal:

**¿Cuánto sería el valor monetario de devolución de productos de los cinco principales clientes de GSK para el año 2022?**

Preguntas Secundarias:

**¿Quiénes son los cinco principales clientes de GSK respecto a su volumen de ventas?**

**¿Cuáles son las variables que determinan las devoluciones de los productos?**

**¿Cuál es el comportamiento de las operaciones por cliente?**

### 3.3. Posibles soluciones al problema, basadas en ciencia de datos.

Se propone realizar una predicción de los valores de las devoluciones de los clientes que permita identificar quiénes son los que presenten mayores índices de devolución. Esto permitirá conocer aquellos que muestren una mayor devolución respecto a sus pedidos, así como a quienes presenten una menor devolución.

Se propone realizar predicciones para identificar qué clientes son los que tendrán mayores devoluciones de productos farmacéuticos. Esto ayudará a establecer posibles estrategias

respecto a las condiciones contractuales que tiene la empresa con sus clientes y que permitan reducir del número de devoluciones.

### 3.4. Justificación

Se han evidenciado reportes que muestran notables devoluciones de productos farmacéuticos que no logran ser vendidos por parte de nuestros clientes, la cual genera preocupación para la empresa, ya que según contrato, aquellos productos retornados deben ser destruidos, generando pérdidas para GSK. Es por ello, que este proyecto de investigación nos ayudará a identificar a los clientes con mayor índice de devolución, con la finalidad de tomar decisiones estratégicas para reducir las pérdidas monetarias.

### 3.5. Objetivos del proyecto

Los objetivos de este proyecto son los siguientes:

Objetivo general:

- Calcular el valor monetario de las devoluciones de los cinco principales clientes de GSK para el año 2022.

Objetivos específicos:

- Identificar a los cinco principales clientes de GSK.
- Identificar las variables que determinan las devoluciones de los productos
- Determinar el comportamiento de las operaciones por cliente.

### 3.6. Hipótesis basadas en el modelo de ciencia de datos

De acuerdo con nuestro proyecto de investigación, el equipo de ciencia de datos ha propuesto las siguientes hipótesis que se mencionan a continuación:

Hipótesis A:

H<sub>1</sub>: La empresa GSK no realiza un análisis de métricas a sus clientes que les permita identificar el volumen oportuno de venta de productos.

H<sub>0</sub>: La empresa GSK sí realiza un análisis de métricas a sus clientes que les permita identificar el volumen oportuno de venta de productos.

Hipótesis B:

H<sub>1</sub>: Nuevas condiciones en los contratos de los clientes permitirán a GSK reducir las pérdidas monetarias por devolución.

H<sub>0</sub>: Nuevas condiciones en los contratos de los clientes no permitirán a GSK reducir las pérdidas monetarias por devolución.

#### **4. REQUISITOS Y RECOPIACIÓN DE DATOS**

##### 4.1. Determinación de datos para responder a las preguntas de ciencias de datos

Nuestra compañera de equipo, Mayra Alexandra Ayala Campos, trabaja directamente con un representante de la empresa GSK, quien será la persona que nos brinde la base de datos requerida para el desarrollo de nuestro trabajo de investigación.

Para determinar los datos que nos ayudarán a responder las preguntas de ciencia de datos, debemos explorar y analizar diversas variables. Algunas de las variables que necesitamos revisar, son los siguientes:

- Nombre de clientes/distribuidor
- Marca de productos
- Tipo de productos
- Nombre de productos
- Ventas en unidades monetarias
- Ventas en cantidades
- Fecha de operación de ventas
- Devolución de productos en unidades monetarias
- Devolución de productos en cantidades

##### 4.2. Ubicación de fuentes de datos

Una de nuestras integrantes del equipo de investigación tiene acceso a información de primera mano debido al contacto directo que tiene con el representante de la empresa GSK. La cual, es información confiable que contribuirá a desarrollar un trabajo altamente profesional.

Nuestra base de datos se encuentra en el formato Excel y está estructurada para trabajarla con las diferentes herramientas tecnológicas. Asimismo, la data se encuentra ubicada en fuentes internas, como el programa SAP.

Cabe resaltar que en nuestra base de datos manejaremos información sensible y/o delicada como nombres de trabajadores, DNI, entre otros. Los cuales, serán modificados salvaguardando los intereses de la empresa y velando por seguridad.

#### 4.3. Presentación de datos recopilados

Adjuntamos una pequeña muestra de la base de datos que estaremos utilizando para responder y predecir la respuesta de la pregunta principal.

Imagen 8. Retorno de productos

Channel/Customer	Sub_channel	Customer Name	Customer code	Actual	%Retur	Returns	%Retur
Indirect Trade	Wholesaler	Quimica Suiza	1100572740	64,359	0.429549	489	0.8 %
Indirect Trade	Wholesaler	Deco	1100572742	8,123	0.303722	346	4.3 %
Indirect Trade	Wholesaler	Alfaro	1100572741	12,247	0.248997	284	2.3 %
Indirect Trade	Wholesaler	Dimexa	1100572739	5,757	0.006834	8	0.1 %
Indirect Trade	Wholesaler	Farmacos Del Norte	1100574011	3,621	0.00671	8	0.2 %
Indirect Trade	Wholesaler	Farmaceutica Del Sur	1100572427	485	0.00265	3	0.6 %
Direct Trade	Supermarket	Tottus	1100572820	1,176	0.001174	1	0.1 %
Direct Trade	Supermarket	Cencosud	1100572744	1,960	0.000275	0	0.0 %
Indirect Trade	Wholesaler	Lavagna	1100654861	103	8.83E-05	0	0.1 %
Direct Trade	Supermarket	Cencosud	8300004626	0	0	0	0.0 %
Direct Trade	Supermarket	Supermercados Perú	1100572745	2,201	0	0	0.0 %
Direct Trade	Supermarket	Supermercados Perú	1100601573	92	0	0	0.0 %

Fuente: GSK, 2021

Imagen 9. Ventas de productos en unidades

Cliente	Marca 2	Producto	Unidades											
			2018	2018	2018	2018	2018	2018	2018	2018	2018	2018	2018	2018
			Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Set	Oct	Nov	Dic
Q,Suiza	PANADOL Adulto	[695680034] Panadol Adulto x 100 tabs	1,120	4,463	5,861	6,831	5,317	0	7,784	5,600	6,348	7,896	2,240	3,920
Q,Suiza	PANADOL Adulto	[695680043] Panadol Efervescente x 24 tabs	1,152	1,152	1,723	2,880	2,304	0	4,800	3,456	2,688	2,496	960	2,400
Q,Suiza	PANADOL Forte	[695680048] Panadol Forte x 48 tabs		9,500	13,000	14,500	21,999	0	20,000	22,850	19,000	29,000	8,000	17,000
Q,Suiza	PANADOL Niños	[695680021] Panadol Jarabe 60 ml x 1 un	16,200	18,864	34,632	35,797	70,064	0	94,680	91,440		52,200	46,800	65,520
Q,Suiza	PANADOL Niños	[695680022] Panadol Gotas 15 ml x 1 un	7,200	26,999	44,000	30,000	73,000	0	66,000	79,048	42,000	67,000	45,000	62,000
Q,Suiza	PANADOL Niños	[695680037] Panadol Masticable x100 Tabs	500	1,600	1,900	2,400	2,100	0	3,900	2,600	800	1,600	500	1,100
Q,Suiza	PANADOL RT	[695641021] Panadol RT NF x 100 tabs	17,193	11,102	16,500	13,905	40,699	1	37,400	42,424	10,000	40,400	16,552	23,000
Q,Suiza	PANADOL RT	[695641023] Panadol Allergy x 50 tabs	500	549	700	1,900	1,550	0	1,800	1,650	403			
Q,Suiza	PANADOL RT	PANADOL ANTIGRIPIAL NF TABLETAS 104'S												
Q,Suiza	PANADOL RT	Panadol Antigripal NF x 104 tab RP												
Q,Suiza	PANADOL RT Vapor	[695641036] Panadol Gripe Vapor Mentol 10s							0					
Q,Suiza	PANADOL RT Vapor	[695641041] Panadol Gripe Vapor Mentol 30s		160				740	0	1,520	855	2,900	1,329	300

Fuente: GSK, 2021

## 5. COMPRENSIÓN Y PREPARACIÓN DE DATOS

### 5.1. Diccionario de Datos

El diccionario de datos es un conjunto de definiciones de las características de los datos que se van a utilizar en el presente trabajo de investigación, incluye nombre, descripción, tipo, valores, estructura, reglas de negocio.

A continuación, se plantea en una tabla el diccionario con las variables utilizadas en el presente trabajo de investigación.

Tabla 1. Diccionario de datos

	TIPO	LONGITUD	NULO	ÍNDICE	ÚNICO	DESCRIPCIÓN	REGLA DE NEGOCIO
<b>Año</b>	Entero	4	No	No	No	Año correspondiente	Desde enero 2019
<b>Mes</b>	Cadena	10	No	No	No	Mes del año correspondiente	Ninguno
<b>Territorio</b>	Cadena	20	No	No	No	Departamento a donde se envía el producto	Lima y Provincias del Perú
<b>Id</b>	Cadena	10	No	Si	Si	Identificador de cliente	Ninguna
<b>Nombre Cliente</b>	Cadena	20	No	No	No	Nombre de la empresa que compra en GSK	Cientes que cuenten con el tipo de contrato especial
<b>Nombre Producto</b>	Cadena	80	No	No	No	Nombre del producto que compra el cliente	Producto que se encuentre en la cartera
<b>Marca</b>	Cadena	20	No	No	No	Marca de producto que se vende a los clientes	Ninguna
<b>Precio</b>	Decimal	5 enteros 2 decimales	No	No	No	Tarifa del precio del producto	Ninguna
<b>Sell-in (Ventas de GSK a clientes)</b>	Decimal	5 enteros 2 decimales	No	No	No	Ventas brutas	Ninguna
<b>Sell-out (Ventas de clientes a usuarios finales)</b>	Decimal	5 enteros 2 decimales	No	No	No	Ventas brutas	Ninguna
<b>Devoluciones en soles</b>	Decimal	5 enteros 2 decimales	No	No	No	Pérdidas monetarias según devoluciones de producto	Ninguna
<b>Cantidad de devoluciones</b>	Entero	4	No	No	No	Número de productos devueltos	Productos vendidos por GSK en un periodo no mayor a 1 año

<b>Cantidad de productos</b>	Entero	5	No	No	No	Número de productos	Ninguna
<b>Descripción de incidencia</b>	Cadena	80	No	No	No	Detalle de la incidencia reportada, motivo de la devolución	Ninguna

## 5.2. Calidad de Datos

Es importante obtener buenos datos y evitar la propagación de datos erróneos. Para el presente trabajo utilizaremos datos internos que nos brinda la empresa GSK, es por ello que los datos recopilados que hemos obtenido cuentan con atributos importantes para la investigación los cuales son:

- **Compleitud.** – Los datos recopilados contienen poseen las características estipuladas según las reglas del negocio.
- **Exactitud.** – Los datos representan el valor real según las operaciones realizadas por la empresa.
- **Consistencia.** – Los datos recopilados muestran coherencia entre las variables.
- **Actualización.** - Los datos en los cuales se va a trabajar son actuales y vigentes para todas las variables estudiadas, teniendo el mismo periodo de análisis.
- **Accesibilidad.** – El equipo tiene acceso a la data por medio un representante de la empresa GSK.
- **Trazabilidad.** - Se tuvo que aplicar el tipo de cambio a ciertas variables que mostraban los datos en moneda extranjera (libras esterlinas).
- **Disponibilidad.** – La persona autorizada de la manipulación de la data es aquella que nos provee de la información.

### 5.2.1. Revisión y Depuración de Datos

El presente trabajo de investigación utilizará todos los datos internos de la empresa GSK desde el mes de enero del 2019 hasta el mes de junio del año 2021. Se procedió a la revisión, depuración y transcripción de los datos.

En la revisión de la base datos se realizó una tabla dinámica, la cual nos arrojó 4 principales errores que se estaban desarrollando en la base de datos:

- En algunos casos contenían un espacio de más, por lo que se tuvo que usar la técnica de Excel “reemplazar” para unificar de tal manera todas las categorías.
- Se corrigieron errores ortográficos.
- Se depuró los campos que no tenían mucha relevancia con el objetivo del trabajo.

La información obtenida fueron los reportes de las ventas netas de la empresa GSK hacia sus clientes, las devoluciones de productos por clientes, las ventas realizadas por los clientes a sus consumidores finales. Todos estos datos en el periodo de enero 2019 a junio 2021.

### 5.2.2. Niveles de calidad

En base a los atributos mencionados se procedió a medir la calidad de los datos, en la cual hemos considerado un peso representativo para cada atributo y asignado un valor según su importancia, asimismo se realizó la regla de tres simple para obtener un porcentaje de la calidad de datos.

<i><b>ATRIBUTO</b></i>	<b>PESO</b>	<b>VALOR</b>	<b>TOTAL</b>
<i>Completitud</i>	15%	15	2,25
<i>Exactitud</i>	15%	20	3
<i>Consistencia</i>	15%	18	2,7
<i>Actualización</i>	10%	20	2
<i>Accesibilidad</i>	20%	15	3
<i>Trazabilidad</i>	5%	16	0,8
<i>Disponibilidad</i>	20%	10	2
<b>TOTAL</b>	<b>100%</b>	<b>114,00</b>	<b>15,75</b>

#### **Regla de Tres Simple:**

Si    100    -    20  
       X     -    15,75

<b>X= 78,75%</b>
------------------

Según el cálculo realizado, se ha obtenido el 78,75% de la calidad de datos, esto nos permitirá realizar un mejor análisis, el cual nos ayudará en dar respuesta a nuestra pregunta de investigación.

### 5.3. Gobierno de datos

La gobernabilidad de datos y a la vez la gestión eficiente de los datos de la empresa implica analizar su situación actual y la realidad de los datos, es decir hoy en día los datos es considerado un activo de gran valor tanto a nivel operativo como para crear valor en el

mercado y convertirlo en información y conocimiento útil para la organización. Es por ello por lo que la empresa GSK considera que el control mediante reportes es importante para poder recopilarla, analizarla y poder aprovecharla de manera eficiente.

Como grupo de ciencia de datos necesitamos datos confiables, seguros, oportunos y coherentes, por ello, este trabajo estipula la normativa y políticas de la empresa en la cual dichos datos están alineadas y se basa en la transparencia y privacidad, así como el uso adecuado de la data dentro de la empresa.

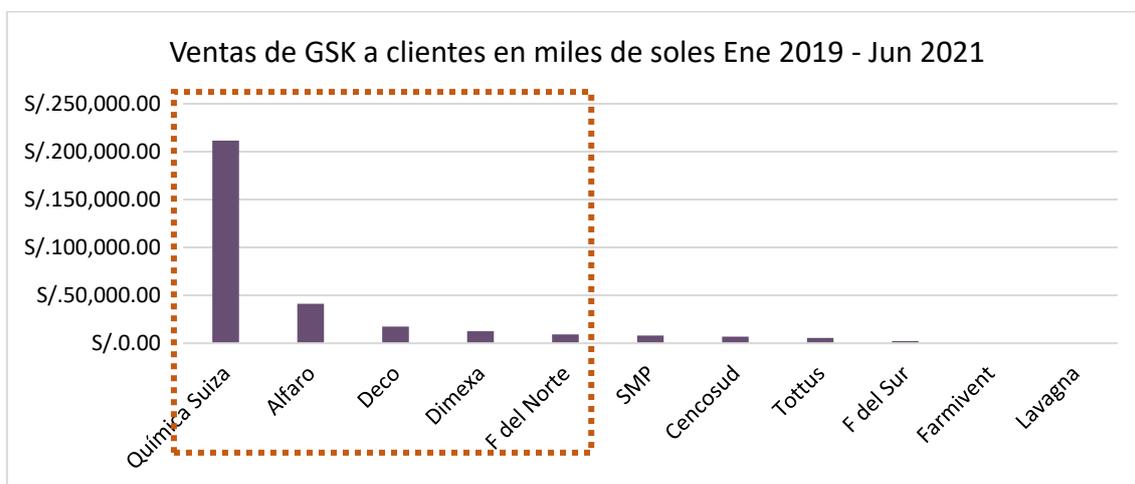
La organización determina los cargos responsables del manejo de data, sus funciones y la implementación de una plataforma técnica que facilite estas prácticas empresariales. Para el soporte técnico y distribución de datos será asignada al área de TI, quien se encargará de la gestión de la calidad de datos, ejecutará acciones necesarias para adquirir, mantener, difundir y poner a disposición de quien corresponda los activos de datos de la empresa.

## 6. PRODUCCIÓN Y ANÁLISIS EXPLORATORIO DE DATOS

### 6.1. Determinación de los clientes principales para el estudio

Es importante mencionar que, para la presente investigación, el equipo ha decidido tomar a los clientes más representativos en las ventas de la empresa GSK. Para ello, se ha realizado un análisis descriptivo sobre las ventas que ha realizado la empresa a sus clientes o distribuidores, obteniendo como resultado a los siguientes: Alfaro, Deco, Dimexa, F del Norte y Química Suiza, siendo este último el de mayor relevancia con un 67.10% de las ventas totales. Se puede revisar el siguiente gráfico:

*Imagen 10. Ventas de GSK a clientes*



*Fuente: Elaboración propia*

## 6.2. Presentación de datos a utilizar para el estudio

Las variables han sido previamente seleccionadas según el objetivo de la investigación para poder dar solución a la problemática planteada, por lo que la data seleccionada debe contribuir con nuestro objetivo principal de identificar a los clientes que tengan un mayor índice de devolución con la finalidad de anticipar las posibles pérdidas monetarias. Se presenta una muestra de nuestra data que se utilizará para el análisis.

Imagen 11. Muestra de tablón de análisis

Devoluciones en soles	Periodo	Sell-in Alfaro Aquafresh en soles	Sell-in Alfaro Corega en soles	Sell-in Alfaro Eno en soles	Sell-in Alfaro Lamisil en soles	Sell-in Alfaro Panadol en soles	Sell-in Alfaro Parodontax en soles	Sell-in Alfaro Respi en soles	Sell-in Alfaro Scott en soles	Sell-in Alfaro Sensodine en soles	Sell-in Alfaro Voltaren en soles	Sell-in Alfaro Aquafresh en unidades	Sell-in Alfaro Corega en unidades	Sell-in Alfaro Eno en unidades	Sell-in Alfaro Lamisil en unidades	Sell-in Alfaro Panadol en unidades	Sell-in Alfaro Parodontax en unidades	Sell-in Alfaro Respi en unidades
S/14,282.71	201901	S/28,537.80	S/53,138.88	S/115,617.36	S/31,033.50	S/100,455.52	S/0.00	S/0.00	S/6,123.72	S/28,491.42	S/13,798.16	6,760	2,304	7,464	1,414	50,084	0	0
S/217,305.24	201902	S/58,426.58	S/234,736.80	S/111,528.00	S/71,231.25	S/788,336.76	S/16,383.00	S/0.00	S/55,309.49	S/90,439.32	S/22,872.88	12,514	1,280	7,200	3,317	34,723	1,524	0
S/156,604.75	201903	S/20,259.60	S/76,790.40	S/0.00	S/36,561.75	S/888,749.08	S/0.00	S/0.00	S/19,033.92	S/41,735.88	S/63,233.36	4,260	3,360	0	1,739	18,360	0	0
S/95,804.25	201904	S/17,305.20	S/176,679.84	S/82,302.48	S/10,984.50	S/808,939.68	S/0.00	S/0.00	S/27,084.60	S/42,516.00	S/24,497.32	3,600	8,304	5,352	650	24,880	0	0
S/40,302.70	201905	S/3,111.46	S/0.00	S/0.00	S/7,692.75	S/968,910.46	S/4,386.00	S/114.56	S/4,062.72	S/6,157.56	S/7,242.84	804	0	0	339	22,666	408	72
S/41,289.56	201906	S/12,983.64	S/15,704.16	S/24,164.40	S/21,725.25	S/1,749,633.90	S/1,935.00	S/2,229.12	S/24,216.72	S/37,055.52	S/28,378.80	2,952	816	1,560	1,089	58,786	180	144
S/234,532.80	201907	S/24,739.23	S/173,142.24	S/0.00	S/23,182.25	S/994,385.74	S/6,321.00	S/2,229.12	S/77,857.20	S/32,071.92	S/22,126.40	5,535	8,304	0	1,053	29,560	588	144
S/138,364.27	201908	S/26,523.12	S/106,433.28	S/71,548.31	S/19,404.00	S/1,484,147.48	S/6,966.00	S/1,857.60	S/33,625.24	S/34,523.76	S/33,731.84	6,396	4,992	4,618	1,100	51,676	648	120
S/170,414.57	201909	S/14,453.88	S/87,718.08	S/57,622.80	S/23,269.50	S/1,227,798.72	S/6,837.00	S/3,343.68	S/38,917.16	S/89,901.24	S/24,958.28	2,520	4,128	3,720	1,450	49,620	636	216
S/154,988.47	201910	S/6,436.68	S/72,890.88	S/0.00	S/13,305.75	S/901,392.60	S/5,805.00	S/1,857.60	S/14,716.60	S/42,726.72	S/9,943.04	1,584	3,552	0	639	27,296	540	120
S/91,267.27	201911	S/10,037.93	S/47,647.68	S/0.00	S/44,811.75	S/504,742.96	S/6,063.00	S/2,600.64	S/7,623.00	S/47,894.20	S/27,718.56	2,079	2,400	0	1,967	28,736	564	168
S/214,400.96	201912	S/0.00	S/0.00	S/0.00	S/0.00	S/0.00	S/0.00	S/0.00	S/0.00	S/0.00	S/0.00	0	0	0	0	0	0	0
S/41,724.00	202001	S/14,549.76	S/173,326.56	S/0.00	S/97,488.75	S/1,251,057.86	S/5,805.00	S/4,066.72	S/33,916.20	S/63,473.52	S/86,261.29	3,024	8,208	0	4,295	66,502	540	264
S/49,391.94	202002	S/25,534.08	S/167,958.80	S/166,176.72	S/149,619.75	S/757,091.34	S/4,515.00	S/0.00	S/65,296.60	S/29,104.20	S/13,680.00	5,172	7,964	10,728	6,643	34,330	420	0
S/66,397.40	202003	S/6,745.08	S/66,563.52	S/163,574.40	S/31,906.50	S/1,263,428.32	S/5,418.00	S/1,114.56	S/52,336.96	S/37,939.80	S/10,260.00	1,104	2,688	10,560	1,378	50,519	504	72
S/43,090.13	202004	S/3,195.04	S/12,702.72	S/0.00	S/7,297.50	S/1,499,534.88	S/5,160.00	S/1,857.60	S/38,175.60	S/7,072.80	S/0.00	576	768	0	350	34,821	480	120
S/20,287.66	202005	S/3,460.80	S/0.00	S/0.00	S/27,105.00	S/2,253,269.80	S/2,193.00	S/4,829.76	S/92,423.80	S/1,906.92	S/5,130.00	1,680	0	0	1,300	68,052	204	312
S/56,031.34	202006	S/14,028.60	S/82,736.00	S/0.00	S/0.00	S/2,030,151.00	S/7,095.00	S/1,857.60	S/77,681.40	S/60,968.07	S/11,970.00	2,484	0	0	8106	660	120	120
S/53,174.56	202007	S/16,647.60	S/30,717.60	S/0.00	S/1,755.00	S/947,407.50	S/1,935.00	S/2,972.16	S/22,737.20	S/31,240.44	S/11,641.92	4,140	1,296	0	60	16,174	180	192
S/61,180.60	202008	S/33,115.92	S/73,967.52	S/0.00	S/0.00	S/890,713.73	S/5,805.00	S/4,829.76	S/19,983.60	S/48,951.52	S/0.00	7,596	3,216	0	0	11,827	540	312
S/68,591.72	202009	S/60,885.27	S/197,343.84	S/0.00	S/21,892.50	S/1,042,680.65	S/5,805.00	S/1,486.08	S/71,363.04	S/105,295.12	S/45,416.24	14,377	9,072	0	1,050	21,926	540	96
S/85,437.88	202010	S/10,857.80	S/74,646.26	S/0.00	S/13,507.52	S/718,883.71	S/12,771.02	S/4,829.76	S/47,207.59	S/61,390.37	S/20,611.03	1,908	3,792	0	614	12,624	1,188	312
S/43,811.98	202011	S/14,040.24	S/60,007.68	S/0.00	S/70,860.60	S/1,054,611.74	S/9,933.00	S/2,229.12	S/147,563.28	S/65,646.72	S/30,967.44	2,508	2,880	0	3,036	32,314	924	144
S/51,140.51	202012	S/5,752.32	S/31,772.20	S/0.00	S/7,297.49	S/619,544.24	S/0.00	S/371.50	S/3,817.83	S/50,087.01	S/7,478.86	1,008	1,440	0	350	27,674	0	24
S/45,713.69	202101	S/19,560.62	S/41,476.84	S/0.00	S/47,639.46	S/556,107.48	S/0.00	S/2,972.17	S/3,648.25	S/34,837.92	S/36,191.66	4,220	1,920	0	1,967	40,906	0	192
S/37,524.28	202102	S/17,399.53	S/58,770.74	S/0.00	S/52,408.14	S/1,774,686.82	S/0.00	S/1,857.59	S/29,760.51	S/23,393.80	S/11,172.46	3,876	2,832	0	2,339	56,964	0	120
S/222,896.72	202103	S/13,526.16	S/195,752.16	S/0.00	S/57,404.48	S/1,045,805.66	S/0.00	S/2,600.64	S/85,788.72	S/93,329.80	S/48,536.76	2,928	8,976	0	2,417	38,556	0	168
S/835,329.23	202104	S/24,180.84	S/80,120.64	S/0.00	S/41,123.48	S/1,280,081.24	S/0.00	S/8,173.44	S/711,693.84	S/85,692.48	S/24,173.20	4,944	3,936	0	1,667	36,796	0	528
S/62,722.11	202105	S/45,947.88	S/124,695.52	S/0.00	S/3,785.00	S/1,288,837.19	S/0.00	S/2,600.64	S/30,182.40	S/68,315.76	S/34,231.12	12,564	4,944	0	450	39,911	0	168
S/138,747.51	202106	S/10,788.49	S/114,208.34	S/0.00	S/8,679.99	S/1,140,286.29	S/0.00	S/0.00	S/42,055.18	S/40,389.00	S/33,683.94	1,944	5,088	0	400	46,236	0	0

Fuente: Elaboración propia

En total, se obtuvieron 357 variables que se han propuesto para el presente análisis, algunas de ellos son las siguientes, teniendo en cuenta que la variable independiente es “Devoluciones en soles”:

- Periodo
- Sell-in Alfaro Aquafresh en soles
- Sell-out Alfaro Voltaren en unidades
- Dev por Fallas Deco en soles
- Precio Promedio Deco Panadol
- Devoluciones Dimexa Eno en soles
- Sell-in F del Norte Productos Sanitarios en soles
- PBI
- Promedio de Sell-in Química Suiza 6 meses en unidades

### 6.3. Plan de muestreo y tamaño de la muestra

Como equipo de ciencias de datos, revisamos y validamos la calidad de nuestra base de datos, por lo que se ha estimado al final de la base ya limpia y depurada de la empresa GSK, nos hemos quedado con 30 registros que agrupan 10,740 datos contenidos en nuestra población a evaluar. Para dicha evaluación se realizará el siguiente plan de muestreo:

- Plan de muestreo probabilístico aleatorio simple
- Para una población finita

Para el cual, se presenta la siguiente fórmula:

$$n = \frac{N \times Z^2 \times p \times q}{e^2 \times (N - 1) + Z^2 \times p \times q}$$

Donde:

n: Tamaño de muestra que queremos hallar

N: Tamaño de población de registros

Z: Parámetro relacionado al nivel de confianza

e: Margen de error

p: Probabilidad de que ocurra un evento

q: Probabilidad de que no ocurra un evento

### 6.4. Diseño del estudio de muestra

Nuestro equipo de ciencia de datos identificó las variables que nos ayudarán a ejecutar nuestro diseño que guardan relación y está asociado a los objetivos que deseamos desarrollar en el transcurso de nuestro proyecto de investigación.

*Imagen 12. Propuesta de diseño de muestra*

Concepto	Definición	Propuesta Diseño Muestral
Unidad de Análisis	Unidad que es el objeto de estudio	Un cliente de GSK
Población	Conjunto de todas las unidades que forman parte del análisis	Todos los clientes de GSK
Unidad de muestreo	Conjunto de elementos que fueron seleccionados en cada etapa del muestreo	Ventas /devoluciones / clientes

Marco muestral	Insumos empleados para identificar las unidades de muestreo	Listado de clientes de GSK / Devoluciones de los clientes de GSK
Muestra	Serie de unidades seleccionadas de la población con la intención de estimar los valores que caracterizan a esta población	Conjunto de clientes de GSK escogidos por muestreo probabilístico aleatorio

Fuente: Elaboración propia

### 6.5. Ejecución del muestreo

La ejecución del muestreo se dará con un muestreo de tipo probabilístico aleatorio simple, mediante el cual una organización de investigación puede identificar a la muestra representativa del universo. En consecuencia, se presenta nuestro cálculo de la muestra según la aplicación de la fórmula:

$$n = \frac{30 \times 1.96^2 \times 0.50 \times 0.50}{0.05^2 \times (30 - 1) + 1.96^2 \times 0.50 \times 0.50}$$

Donde:

n: Tamaño de muestra que queremos hallar

N: Tamaño de población de registros: 30

Z: Parámetro relacionado al nivel de confianza: 95%  $\cong$  1.96

e: Margen de error: 5%

p: Probabilidad de que ocurra un evento: 50%

q: Probabilidad de que no ocurra un evento: 50%

$$n = 27$$

Según la ejecución de nuestro diseño de muestreo, se ha podido determinar que el número de registros que debemos considerar para nuestra muestra es de 27.

Cabe resaltar que este número de muestra se usará solo para validar el modelo. En ese sentido, se seguirá tomando a toda la población para el análisis de las variables, ya que creemos que es lo mejor para la empresa presentada.

## 6.6. Resumen y examen de la distribución de variables

Luego de realizar el respectivo análisis de las variables que hemos propuesto, se ha llegado a los resultados que se mostrarán a continuación. Para ello, se ha tomado como herramienta a la caja y bigotes y a la Campana de Gauss.

Imagen 13. Caja y bigotes. Ventas en soles de GSK a sus clientes



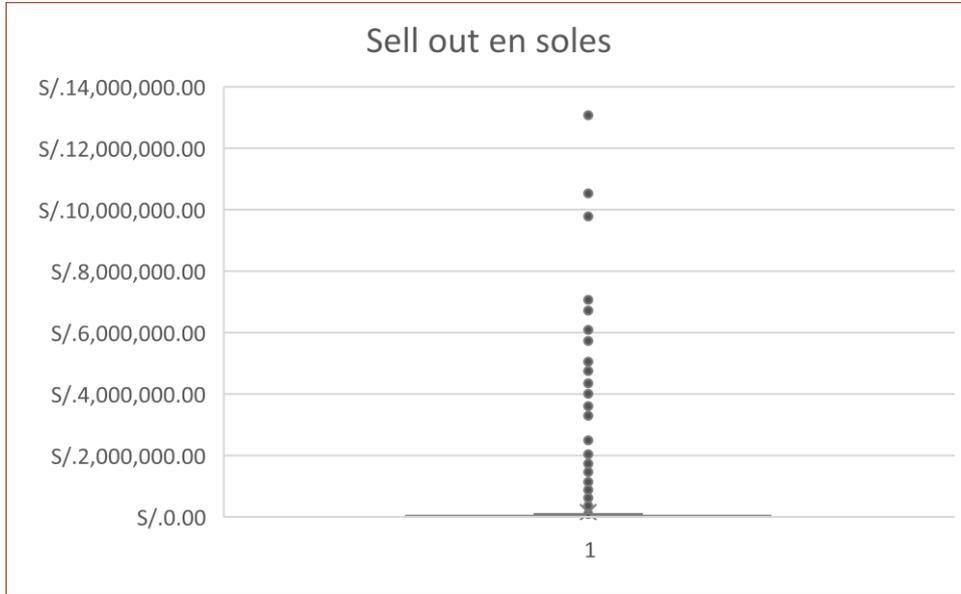
Fuente: Elaboración propia

Imagen 14. Caja y bigotes. Ventas en unidades de GSK a sus clientes



Fuente: Elaboración propia

Imagen 15. Caja y bigotes. Ventas en soles de los clientes de GSK a sus consumidores



Fuente: Elaboración propia

Imagen 16. Caja y bigotes. Venta en unidades de los clientes de GSK a sus consumidores



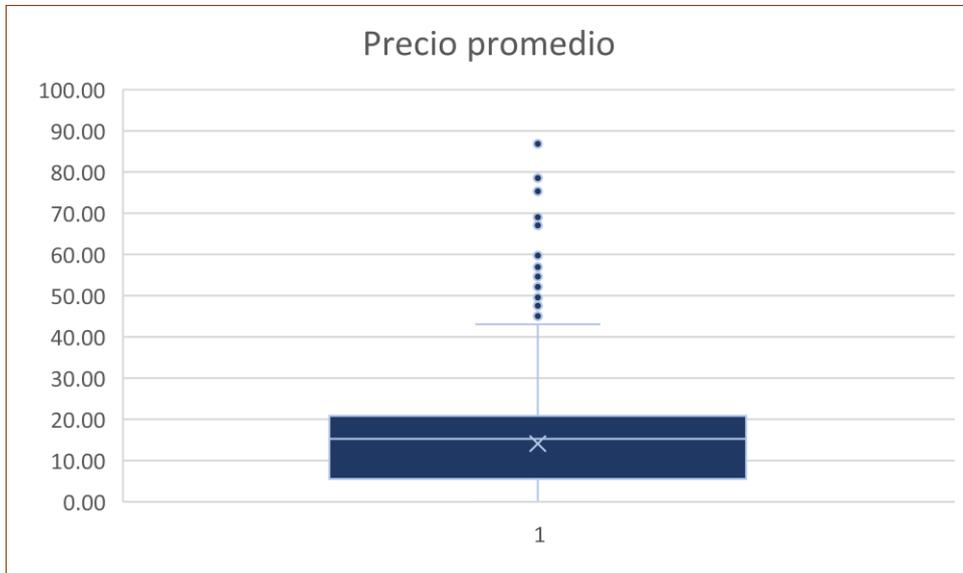
Fuente: Elaboración propia

Imagen 17. Caja y bigotes. Devoluciones en soles de los clientes a GSK



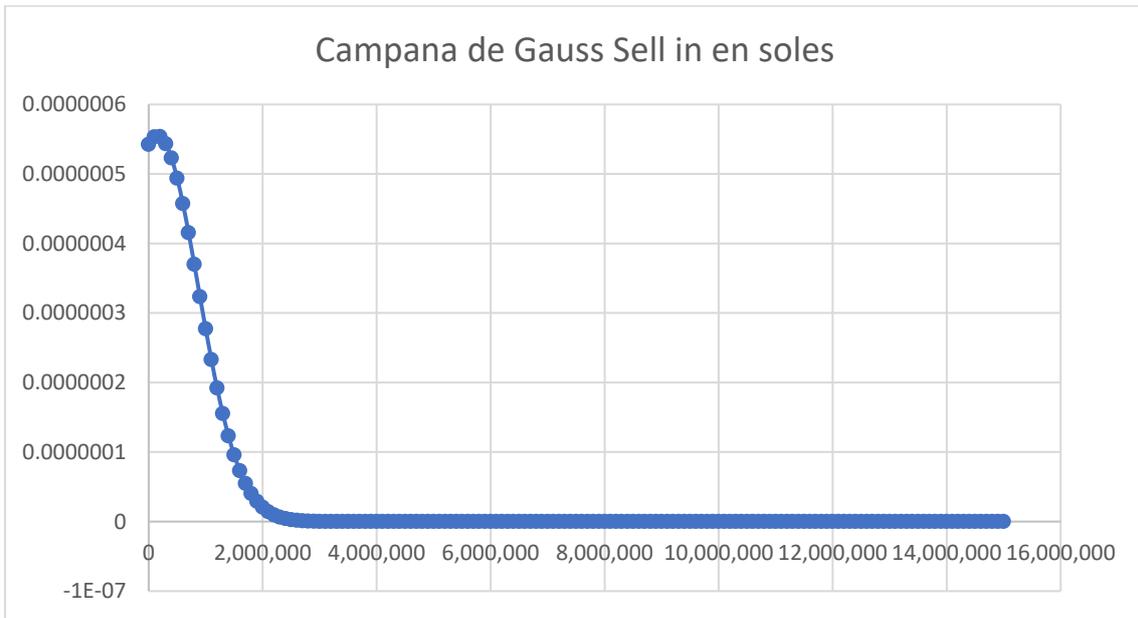
Fuente: Elaboración propia

Imagen 18. Caja y bigotes. Precio promedio de venta de GSK a sus clientes



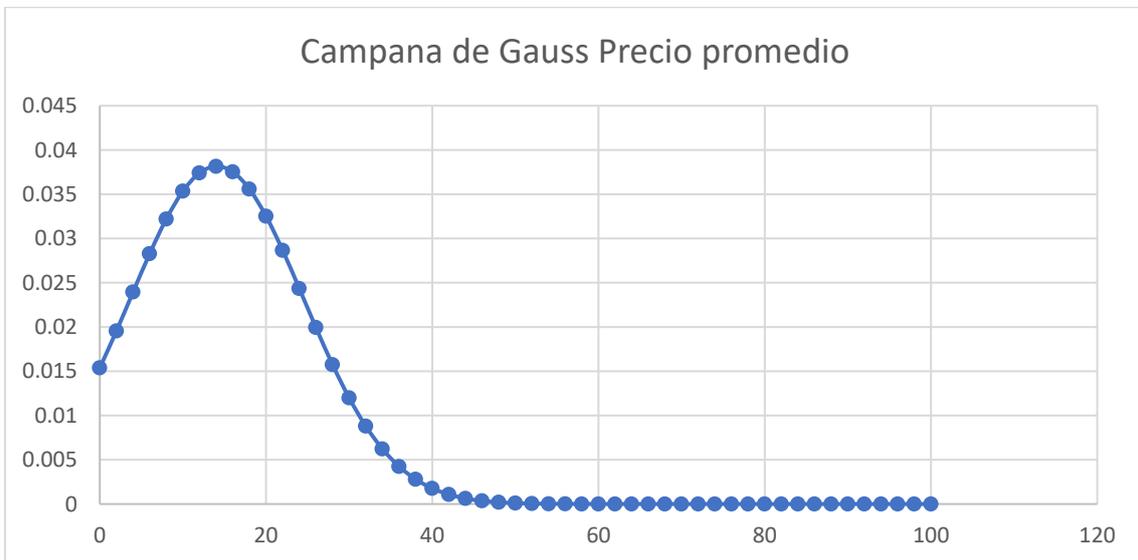
Fuente: Elaboración propia

Imagen 19. Campana de Gauss. Sell in en soles



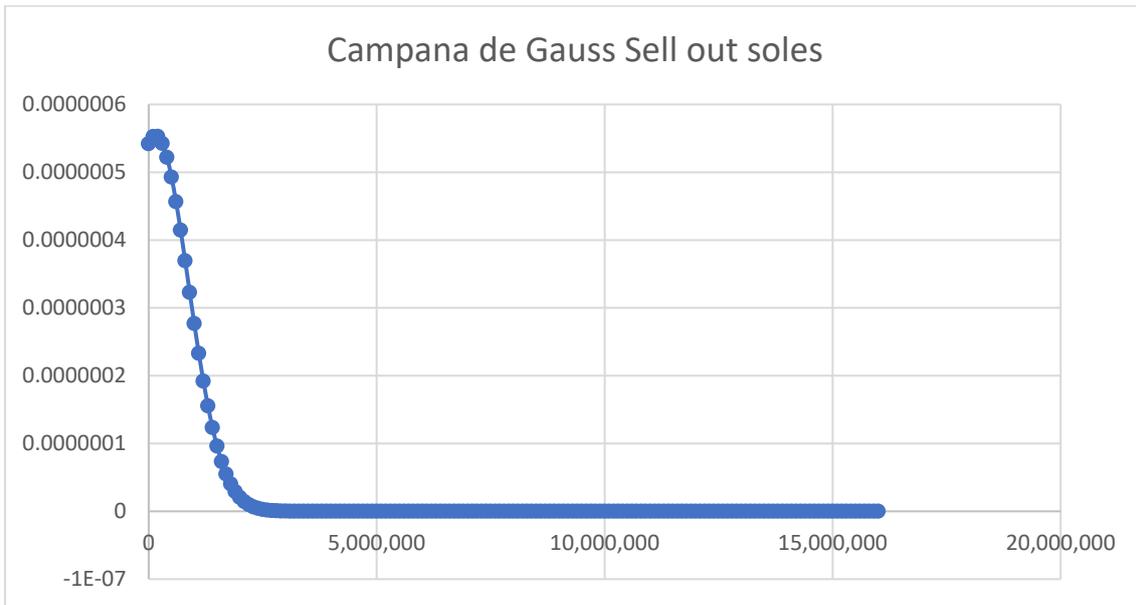
Fuente: Elaboración propia

Imagen 20. Campana de Gauss. Precio promedio



Fuente: Elaboración propia

Imagen 21. Campana de Gauss. Sell out en soles



Fuente: Elaboración propia

Imagen 22. Campana de Gauss. Devoluciones en soles de los clientes a GSK



Fuente: Elaboración propia

### 6.7. Resultado del análisis de relaciones

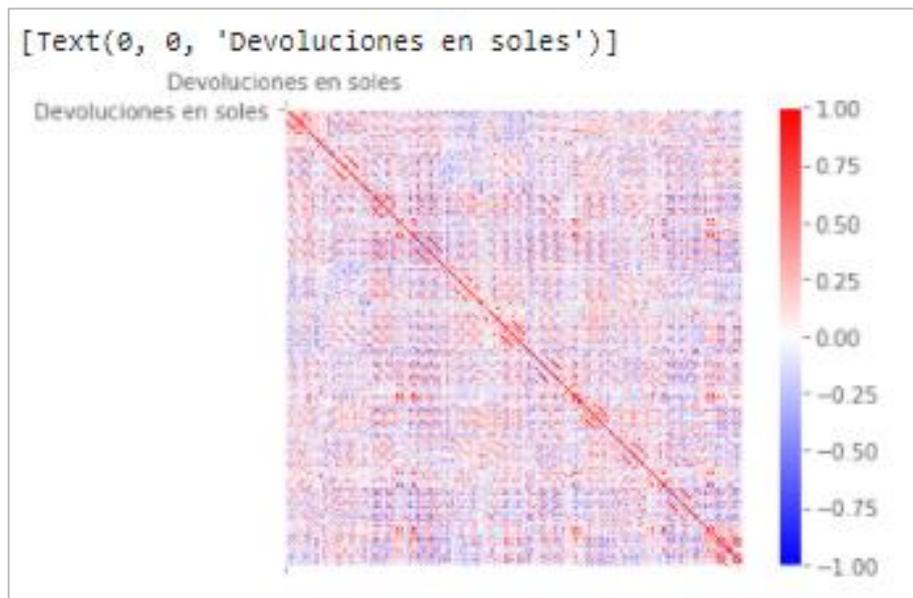
Asimismo, se ha realizado un análisis exhaustivo de correlaciones entre las variables presentadas para el proyecto. Se presenta la matriz de correlaciones:

Tabla 2. Muestra de matriz de correlaciones

	Devoluciones en soles	Periodo	Sell-in Alfaro Aquafresh en soles	Sell-in Alfaro Corega en soles	Sell-in Alfaro Eno en soles	Sell-in Alfaro Lamiill en soles	Sell-in Alfaro Panadol en soles	Sell-in Alfaro Parodontax en soles	Sell-in Alfaro Respira en soles	Sell-in Alfaro Scott en soles	Sell-in Alfaro Sensodyne en soles	Sell-in Alfaro Voltaren en soles	Sell-in Alfaro Aquafresh en unidades	Sell-in Alfaro Corega en unidades	Sell-in Alfaro Eno en unidades	Sell-in Alfaro Lamiill en unidades	Sell-in Alfaro Panadol en unidades	Sell-in Alfaro Parodontax en unidades	Sell-in Alfaro Respira en unidades	Sell-in Alfaro Scott en unidades	Sell-in Alfaro Sensodyne en unidades	Sell-in Alfaro Voltaren en unidades
Devoluciones en soles	1.00000	0.144363	0.116387	0.156698	-0.058863	0.035510	-0.074913	-0.124457	0.442926	0.147054	0.309565	0.050494	0.083403	0.174981	-0.058863	0.019629	-0.108576	-0.124457	0.442978	0.124928	0.275558	0.009377
Periodo	0.144363	1.000000	0.057413	0.054374	-0.292145	0.117545	0.245926	-0.324288	0.384765	0.281637	0.173768	0.126234	0.073900	0.026497	-0.292146	0.078201	0.193002	-0.324288	0.384763	0.234447	0.093323	0.105821
Sell-in Alfaro Aquafresh en soles	0.116387	0.057413	1.000000	0.645049	0.210333	0.190458	-0.015371	0.262156	-0.018828	0.165135	0.601651	0.252983	0.989437	0.627651	0.210332	0.207737	-0.049997	0.262155	-0.018786	0.212757	0.578184	0.231344
Sell-in Alfaro Corega en soles	0.156698	0.054374	0.645049	1.000000	0.302133	0.438442	-0.092622	0.296951	-0.140803	0.338531	0.599550	0.512270	0.601604	0.996665	0.302132	0.453027	0.036638	0.296951	-0.140781	0.381191	0.585432	0.474190
Sell-in Alfaro Eno en soles	-0.058863	-0.292145	0.210333	0.302133	1.000000	0.431357	-0.067229	0.194704	-0.404462	0.016779	-0.003447	-0.149156	0.167048	0.301313	1.000000	0.455119	0.155762	0.194704	-0.404453	0.065707	0.021862	-0.193168
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Promedio de Sell-out Alfaro 8 meses en unidades	-0.101688	-0.279546	0.146550	-0.113944	0.326827	-0.161863	0.201829	0.017872	-0.211022	-0.152788	-0.160803	-0.188274	0.142262	-0.131193	0.326828	-0.149609	0.222137	0.017872	-0.211017	-0.144784	-0.154513	-0.196723
Promedio de Sell-out Deao 8 meses en unidades	0.286463	0.734320	-0.061367	-0.012127	-0.232461	-0.143275	0.300742	-0.423760	0.263101	0.100016	0.021043	0.044439	-0.017376	-0.050794	-0.232461	-0.169971	0.232899	-0.423760	0.263090	0.054295	-0.012586	0.051724
Promedio de Sell-out Dimexa 8 meses en unidades	0.076209	0.513530	-0.276813	-0.136021	-0.352469	-0.073381	0.144340	-0.107253	0.436975	0.150293	0.023853	-0.045551	-0.239662	-0.149222	-0.352469	-0.097015	0.093470	-0.107253	0.436936	0.141324	0.015011	-0.024070
Promedio de Sell-out F del Norte 8 meses en unidades	0.103303	0.801490	-0.098703	-0.050895	-0.396744	-0.066101	0.202651	-0.160473	0.466632	0.274486	0.162201	0.009632	-0.071990	-0.079490	-0.396745	-0.100440	0.095571	-0.160473	0.466609	0.247485	0.121425	0.021303
Promedio de Sell-out Quimica Suiza 8 meses en unidades	0.050259	-0.190099	-0.092600	-0.128845	0.140183	-0.317180	0.451603	-0.157733	0.022354	-0.100881	-0.307646	-0.215666	-0.059694	-0.154962	0.140184	-0.304977	0.281679	-0.157733	0.022334	-0.066298	-0.272417	-0.213502

Fuente: Elaboración propia

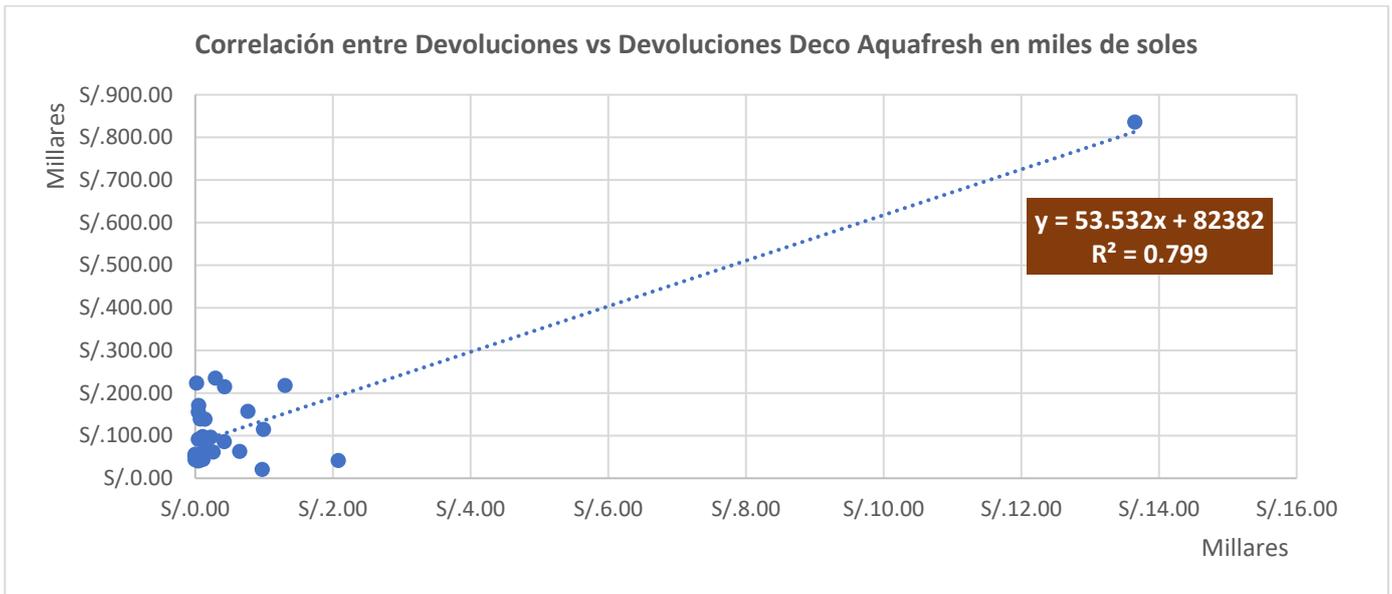
Tabla 3. Matriz de correlaciones – Mapa de calor



Fuente: Elaboración propia

Ahora se presentan, algunas variables representativas de las 358 del total y que muestran correlaciones fuertes, regulares y débiles.

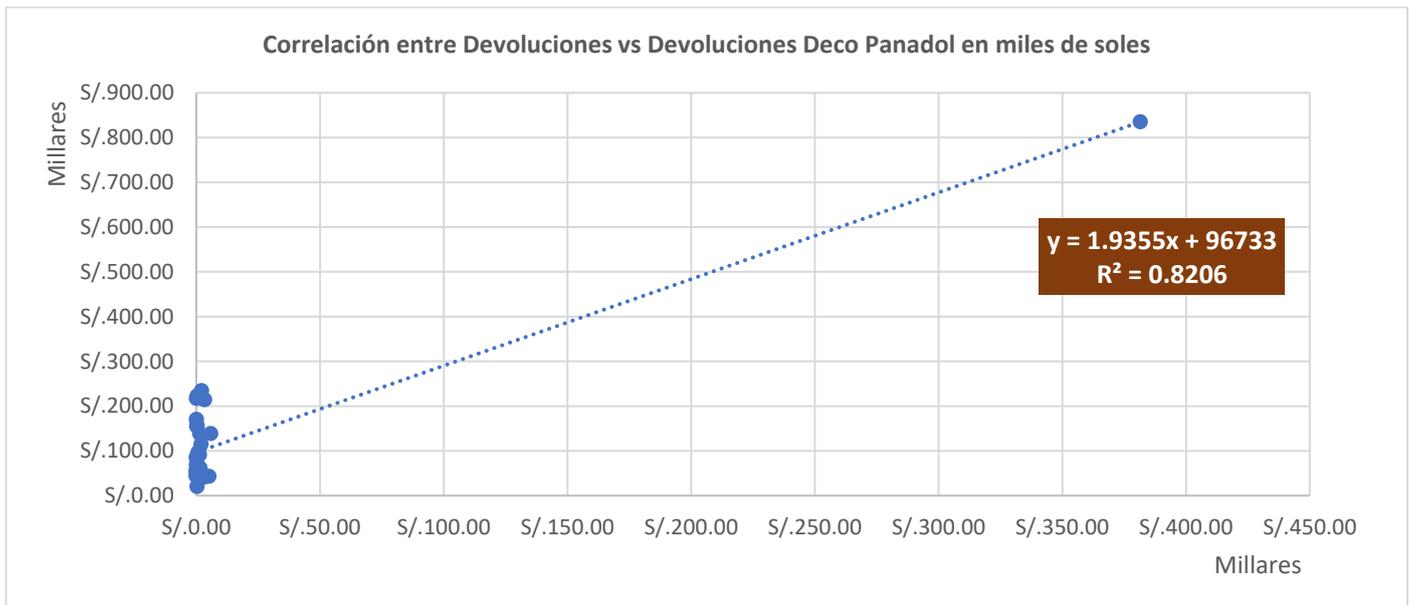
Imagen 23. Correlación entre Devoluciones vs Devoluciones Deco Aquafresh en miles de soles



Fuente: Elaboración propia

El gráfico 23 presenta una correlación positiva considerable, de 0.799 entre las variables ya mencionadas.

Imagen 24. Correlación entre Devoluciones vs Devoluciones Deco Panadol en miles de soles

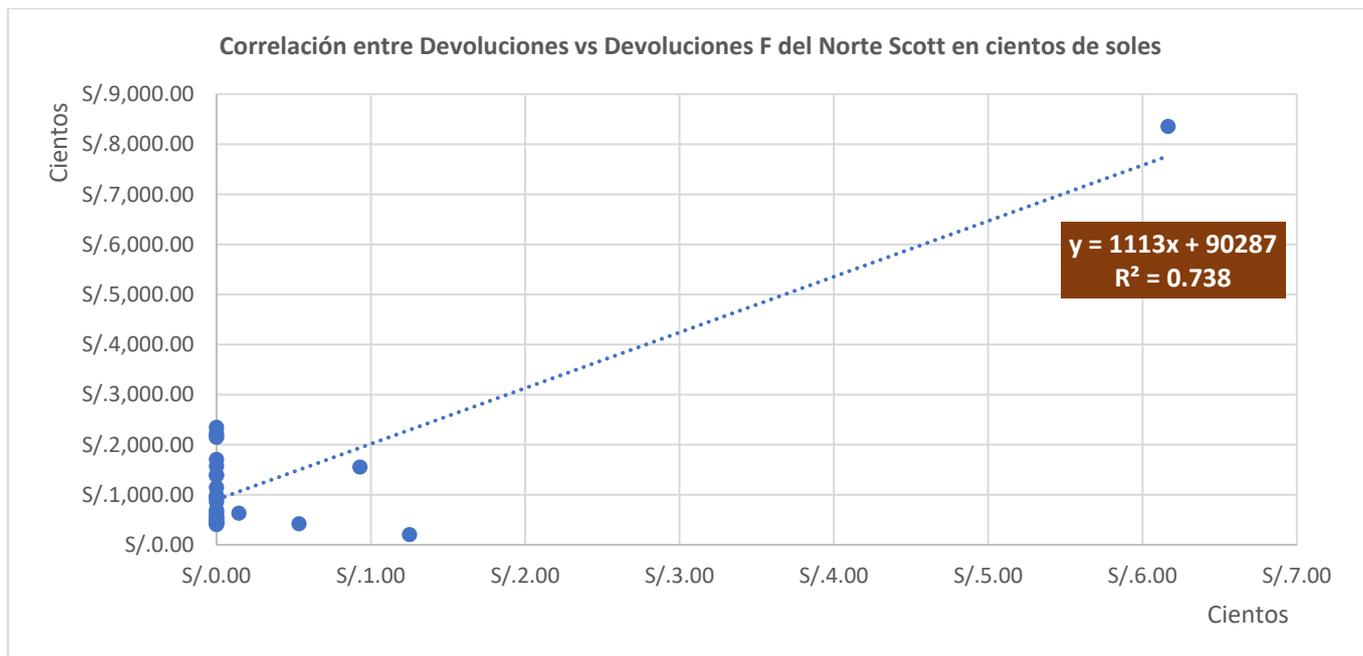


Fuente: Elaboración propia

El gráfico 24 presenta una correlación positiva considerable, de 0.8206 entre las variables ya mencionadas.



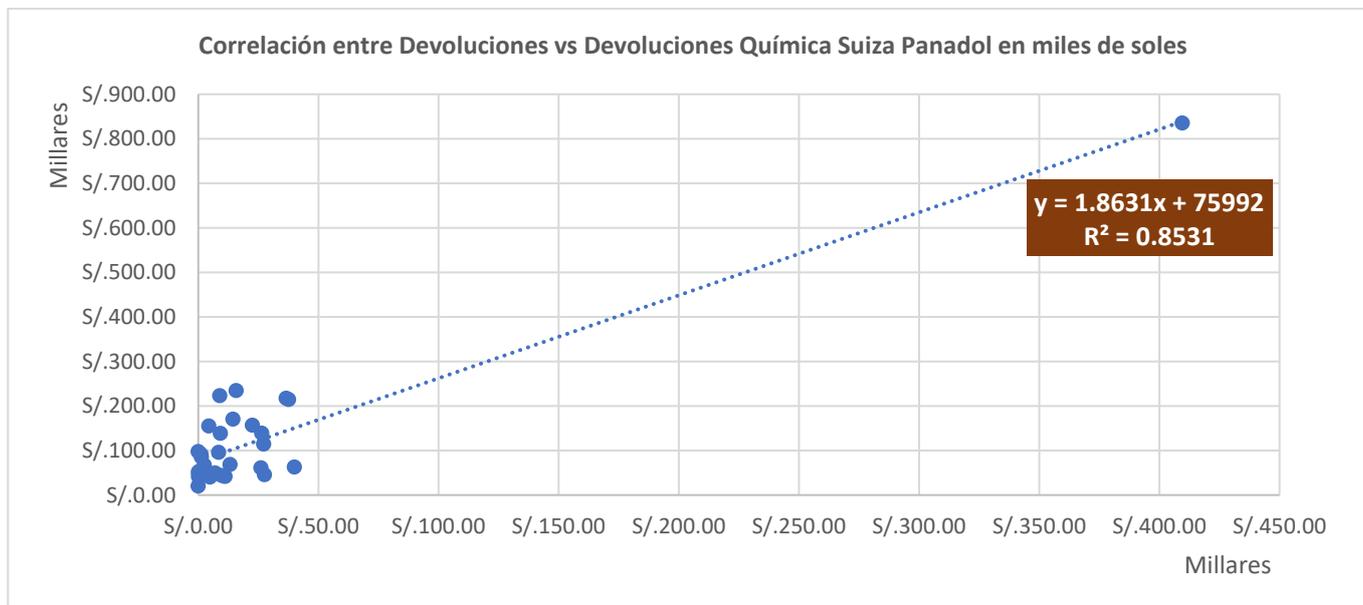
Imagen 27. Correlación entre Devoluciones vs Devoluciones F del Norte Scott en cientos de soles



Fuente: Elaboración propia

El gráfico 27 presenta una correlación positiva alta, de 0.0738 entre las variables ya mencionadas.

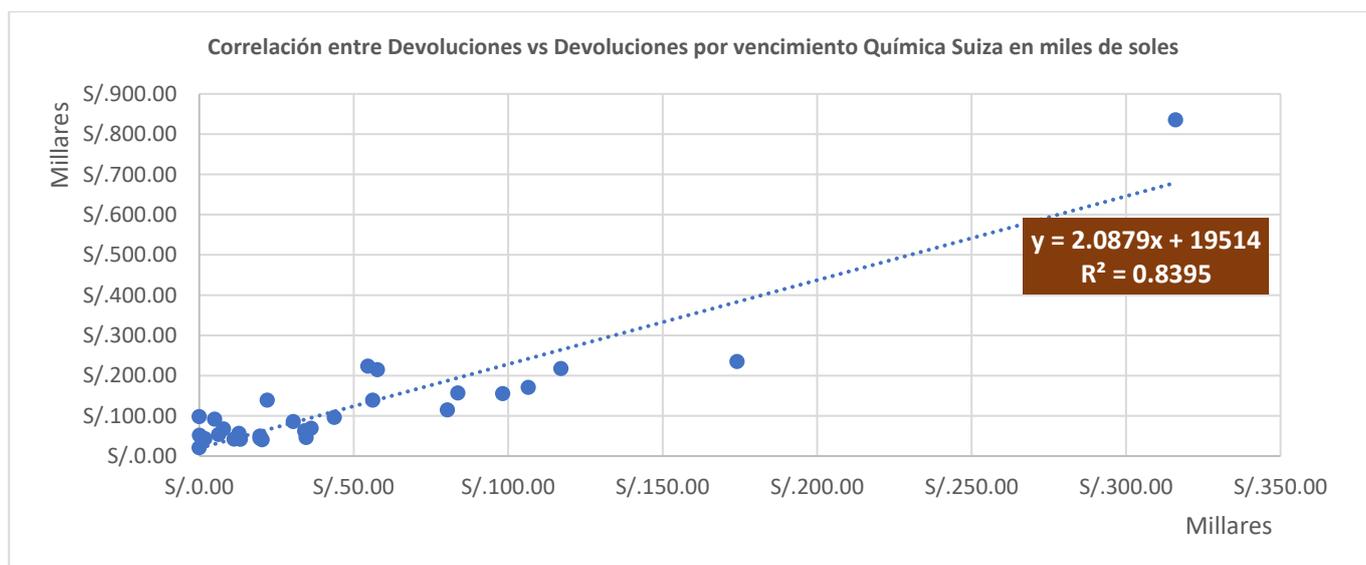
Imagen 28. Correlación entre Devoluciones vs Devoluciones Química Suiza Panadol en miles de soles



Fuente: Elaboración propia

El gráfico 28 presenta una correlación positiva considerable, de 0.8531 entre las variables ya mencionadas.

Imagen 29. Correlación entre Devoluciones vs Devoluciones por vencimiento Química Suiza en miles de soles



Fuente: Elaboración propia

El gráfico 29 presenta una correlación positiva considerable, de 0.8395 entre las variables ya mencionadas.

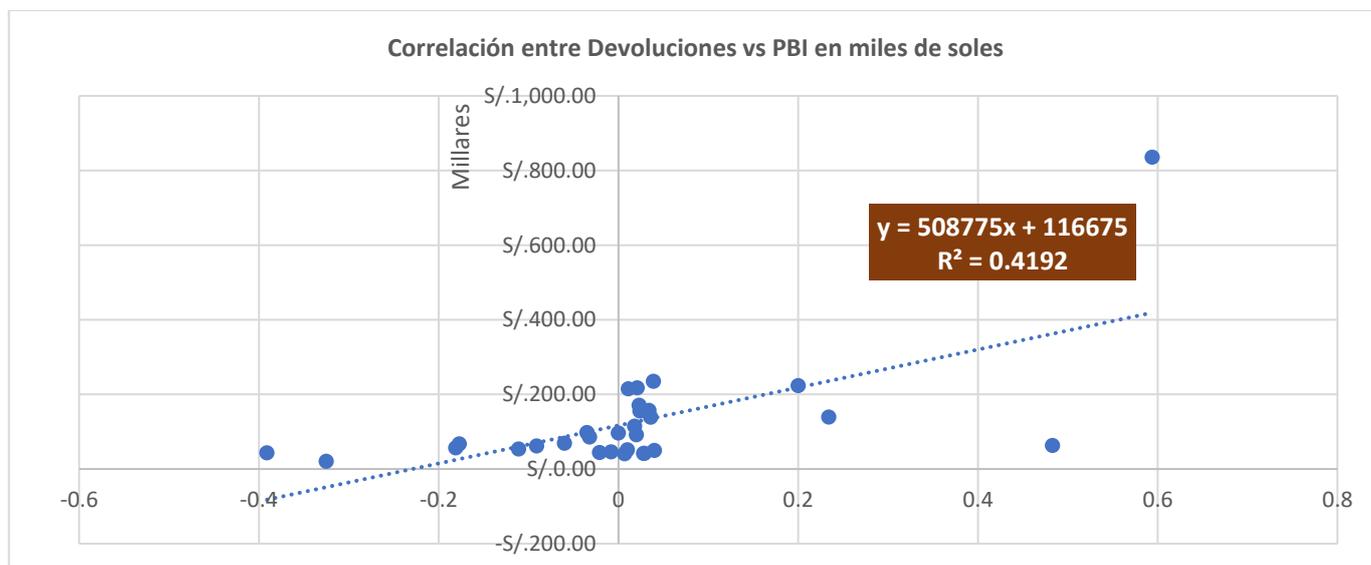
Imagen 30. Correlación entre Devoluciones vs Dev por sobre stock Química Suiza en miles de soles



Fuente: Elaboración propia

El gráfico 30 presenta una correlación positiva considerable, de 0.812 entre las variables ya mencionadas.

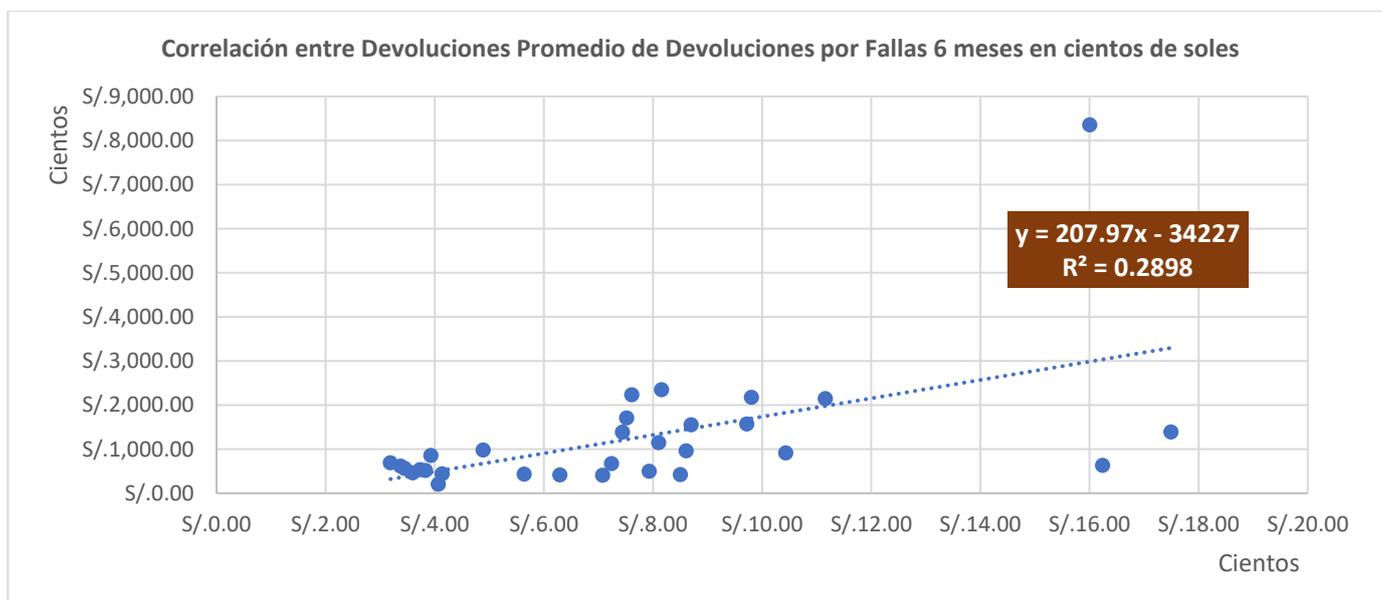
Imagen 31. Correlación entre Devoluciones vs PBI en miles de soles



Fuente: Elaboración propia

El gráfico 31 presenta una correlación positiva regular, de 0.4192 entre las variables ya mencionadas.

Imagen 32. Correlación entre Devoluciones vs Devoluciones promedio de devoluciones por fallas 6 meses en cientos de soles



Fuente: Elaboración propia

El gráfico 32 presenta una correlación positiva baja, de 0.2898 entre las variables ya mencionadas.

Imagen 33. Correlación entre Devoluciones vs Sell-in Dimexa Corega en miles de soles



Fuente: Elaboración propia

El gráfico 42 presenta una correlación positiva baja, de 0.0003 entre las variables ya mencionadas.

## 7. COMENZANDO LA INTERPRETACIÓN DE DATOS

Es importante recordar que para el análisis presentado se ha utilizado la totalidad de la población de los datos para obtener un resultado que se acerque más a la realidad. De acuerdo con ello, se obtiene las siguientes interpretaciones:

- En primer lugar, se puede apreciar en la Campana de Gauss cómo se distribuyen las variables evaluadas. Por ejemplo: se valida que la variable Sell in en soles se concentra en la media de aproximadamente 153,000 soles, sin embargo, existen datos que están muy lejos de esta media. Asimismo, como es el caso de la variable Sell out en soles que muestra la concentración o distribución de sus datos sobre la media de aproximadamente 150,000 soles, mientras que otros datos se visualizan a una distancia mucho mayor.
- En segundo lugar, las variables que tienen mayor correlación son Sell in y Sell out, tanto en soles como en unidades. De las cuales se puede deducir que ambas variables están relacionadas linealmente. Es decir, la primera como la segunda variable cambian a una tasa constante.

- En tercer lugar, existen otras variables que hemos identificado que presentan correlaciones casi inexistentes, como es el caso a las relacionadas al PBI. En este caso, el equipo de ciencia de datos tuvo el criterio de aplicar esta variable para ver alguna relación entre el consumo nacional y estas devoluciones de los clientes o distribuidores GSK. No obstante, se obtuvo el cálculo mostrado.

Por otro lado, se ha realizado un análisis descriptivo de nuestras variables que se mostrarán a continuación:

Tabla 4. Estadística descriptiva 1

	Devoluciones en soles	Periodo	Sell-in Alfaró Aquafresh en soles	Sell-in Alfaró Corega en soles	Sell-in Alfaró Eno en soles	Sell-in Alfaró Lamisil en soles	Sell-in Alfaró Panadol en soles	Sell-in Alfaró Parodontax en soles	Sell-in Alfaró Respira en soles	Sell-in Alfaró Scott en soles	Sell-in Alfaró Sensodyne en soles	Sell-in Alfaró Voltaren en soles
count	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	3.000000e+01	30.000000	30.000000	30.000000	30.000000	30.000000
mean	123780.766667	201985.900000	19000.766667	87720.400000	26437.766667	31832.400000	1.093753e+06	4037.700000	2130.200000	41639.400000	46177.333333	23863.533333
std	148396.244991	75.207002	14725.484669	65587.346566	50890.191461	32796.137841	4.674447e+05	4177.489282	1889.155228	33267.73014	25801.280541	19356.566074
min	20288.000000	201901.000000	0.000000	0.000000	0.000000	0.000000	0.000000e+00	0.000000	0.000000	0.000000	0.000000	0.000000
25%	49829.250000	201908.250000	10805.500000	43019.750000	0.000000	8951.250000	8.032548e+05	0.000000	557.750000	19271.500000	31448.000000	10488.000000
50%	76995.000000	202003.500000	14502.000000	74307.000000	0.000000	22542.500000	1.022568e+06	4450.500000	1858.000000	33770.500000	42126.000000	22499.500000
75%	150935.500000	202010.750000	24599.500000	122044.000000	18123.000000	43891.250000	1.286648e+06	5998.500000	2879.250000	62801.500000	61082.000000	33040.750000
max	835329.000000	202106.000000	60885.000000	234737.000000	166177.000000	149620.000000	2.253270e+06	16383.000000	8173.000000	147563.000000	105285.000000	86261.000000

8 rows x 358 columns

Fuente: Elaboración propia

Tabla 5. Estadística descriptiva 2

Sell-in Alfaró Aquafresh en unidades	Sell-in Alfaró Corega en unidades	Sell-in Alfaró Eno en unidades	Sell-in Alfaró Lamisil en unidades	Sell-in Alfaró Panadol en unidades	Sell-in Alfaró Parodontax en unidades	Sell-in Alfaró Respira en unidades	Sell-in Alfaró Scott en unidades	Sell-in Alfaró Sensodyne en unidades	Sell-in Alfaró Voltaren en unidades	Devoluciones Alfaró Aquafresh en soles	Devoluciones Alfaró Corega en soles	Devoluciones Alfaró Eno en soles
30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000
4168.566667	4061.133333	1706.766667	1435.766667	37130.833333	375.600000	137.600000	2544.700000	4137.933333	1304.233333	3693.566667	562.533333	1459.133333
3584.385127	3080.151348	3285.361321	1438.861228	18364.662938	388.603654	122.040581	1909.630135	2409.900870	1192.800578	9127.852109	685.865511	3011.211015
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1917.000000	2016.000000	0.000000	412.500000	25484.000000	0.000000	36.000000	1332.000000	2670.000000	421.500000	99.000000	83.500000	0.000000
2988.000000	3367.000000	0.000000	1094.500000	34772.000000	414.000000	120.000000	2154.000000	3936.000000	1163.000000	375.500000	271.500000	47.000000
5115.000000	5064.000000	1170.000000	1910.000000	49968.000000	558.000000	186.000000	3876.250000	5321.000000	1984.000000	1916.000000	836.500000	583.250000
14377.000000	11280.000000	10728.000000	6643.000000	81016.000000	1524.000000	528.000000	8064.000000	9744.000000	5027.000000	46469.000000	3010.000000	11623.000000

Fuente: Elaboración propia

Tabla 6. Estadística descriptiva 3

Devoluciones Alfaró Lamisil en soles	Devoluciones Alfaró Panadol en soles	Devoluciones Alfaró Parodontax en soles	Devoluciones Alfaró Respira en soles	Devoluciones Alfaró Scott en soles	Devoluciones Alfaró Sensodyne en soles	Devoluciones Alfaró Voltaren en soles	Sell-in Alfaró Medicamento en soles	Sell-in Alfaró Productos Sanitarios en soles	Dev por Vencimiento Alfaró en soles	Dev por Sobre Stock Alfaró en soles	Dev por Fallas Alfaró en soles
30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	3.000000e+01	30.000000	30.000000	30.000000	30.000000
248.633333	26939.300000	269.466667	7.700000	1259.433333	1932.500000	1675.166667	1.178016e+06	198575.433333	28973.933333	7586.500000	1486.933333
382.364376	32915.650237	417.057983	28.227096	1821.333060	1850.838948	2823.383298	4.683002e+05	112590.114479	28718.766910	6853.666694	1532.455409
0.000000	65.000000	0.000000	0.000000	0.000000	34.000000	0.000000	0.000000e+00	0.000000	916.000000	183.000000	46.000000
34.500000	5708.500000	13.000000	0.000000	237.750000	599.250000	74.750000	9.363622e+05	116715.250000	10811.250000	2183.250000	490.500000
119.000000	14507.500000	85.500000	0.000000	601.500000	1164.000000	449.000000	1.132912e+06	194148.500000	23349.000000	5250.500000	943.500000
242.250000	30882.250000	317.000000	0.000000	1443.750000	3578.500000	1606.500000	1.426334e+06	267723.000000	35169.000000	9949.250000	1978.750000
1565.000000	126599.000000	1723.000000	148.000000	8966.000000	5368.000000	11719.000000	2.290335e+06	455295.000000	116953.000000	26171.000000	6880.000000

Fuente: Elaboración propia

Tabla 7. Estadística descriptiva 4

Sell-out Alfaró Aquafresh en soles	Sell-out Alfaró Corega en soles	Sell-out Alfaró Eno en soles	...	Precio Promedio Química Suiza Corega	Precio Promedio Química Suiza Eno	Precio Promedio Química Suiza Lamisil	Precio Promedio Química Suiza Panadol	Precio Promedio Química Suiza Parodontax	Precio Promedio Química Suiza Respira	Precio Promedio Química Suiza Scott	Precio Promedio Química Suiza Sensodyne	Precio Promedio Química Suiza Voltaren	Sell-in Totales en soles
30.000000	30.000000	30.000000	...	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	3.000000e+01
18925.800000	82376.633333	35291.133333	...	22.933333	6.500000	21.866667	30.866667	8.066667	15.000000	17.433333	12.033333	22.000000	9.730321e+06
10804.466357	52526.894366	55523.622344	...	1.172481	7.560104	4.328999	10.503639	4.947541	0.000000	1.977169	0.614948	5.239111	2.636759e+06
817.000000	0.000000	0.000000	...	21.000000	0.000000	0.000000	18.000000	0.000000	15.000000	11.000000	10.000000	0.000000	5.541324e+06
12501.250000	39371.750000	0.000000	...	22.000000	0.000000	22.000000	23.000000	2.750000	15.000000	17.000000	12.000000	21.000000	7.944007e+06
16541.000000	84223.500000	4174.500000	...	23.000000	0.000000	22.000000	30.000000	11.000000	15.000000	17.000000	12.000000	22.000000	9.164893e+06
26130.250000	115842.500000	43600.750000	...	23.000000	15.000000	23.000000	34.000000	11.000000	15.000000	19.000000	12.000000	23.750000	1.089341e+07
48642.000000	187946.000000	188142.000000	...	26.000000	15.000000	28.000000	60.000000	11.000000	15.000000	20.000000	13.000000	29.000000	1.649232e+07

Fuente: Elaboración propia

Tabla 8. Estadística descriptiva 5

Sell-in Totales en unidades	Devoluciones por Vencim. Totales en soles	Devoluciones por Sobre Stock Totales en soles	Devoluciones por Fallas Totales en soles	Sell-out Totales en soles	Sell-out Totales en unidades	PBI	Promedio de Sell-in Alfaró 6 meses en soles	Promedio de Sell-in Deco 6 meses en soles	Promedio de Sell-in Dimexa 6 meses en soles	Promedio de Sell-in F del Norte 6 meses en soles	Promedio de Sell-in Química Suiza 6 meses en soles
30.000000	30.000000	30.000000	30.000000	3.000000e+01	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000
420433.966667	95102.033333	24455.433333	4223.300000	9.654773e+06	416138.066667	0.033333	135017.266667	51767.533333	38234.733333	27270.733333	684221.366667
110917.811123	115457.974901	28457.084254	5241.426919	2.911030e+06	114951.790145	0.182574	20230.665813	21683.043984	11758.693600	11632.460442	89163.294782
221866.000000	16262.000000	3068.000000	478.000000	4.979010e+06	250571.000000	0.000000	106083.000000	23966.000000	14356.000000	8334.000000	571836.000000
363167.000000	37858.500000	9381.750000	1602.250000	8.233657e+06	354812.750000	0.000000	121742.500000	40075.250000	33901.750000	20843.750000	613071.500000
401313.500000	56912.000000	14257.000000	2787.500000	8.604543e+06	404905.500000	0.000000	134038.000000	46940.500000	37614.500000	28056.500000	663433.000000
468002.000000	107657.500000	35374.250000	4753.750000	1.082837e+07	441133.750000	0.000000	145160.500000	52483.000000	42648.750000	33691.750000	756809.750000
786066.000000	650449.000000	156091.000000	28789.000000	1.947644e+07	884877.000000	1.000000	183729.000000	106335.000000	64330.000000	50546.000000	902246.000000

Fuente: Elaboración propia

Tabla 9. Estadística descriptiva 6

```
gsk.describe(include='all')
```

Promedio de Sell-in Alfaro 6 meses en unidades	Promedio de Sell-in Deco 6 meses en unidades	Promedio de Sell-in Dimexa 6 meses en unidades	Promedio de Sell-in F del Norte 6 meses en unidades	Promedio de Sell-in Química Suiza 6 meses en unidades	Promedio de Devoluciones por Vencim. 6 meses en soles	Promedio de Devoluciones por Sobre Stock 6 meses en soles	Promedio de Devoluciones por Fallas 6 meses en soles	Promedio de Sell-out Alfaro 6 meses en soles	Promedio de Sell-out Deco 6 meses en soles	Promedio de Sell-out Dimexa 6 meses en soles	Promedio de Sell-out F del Norte 6 meses en soles	Promedio de Sell-out Química Suiza 6 meses en soles
30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000
738.166667	2431.233333	1700.533333	1078.533333	30707.800000	17466.666667	4469.400000	759.733333	114004.433333	50823.966667	35513.000000	25722.200000	682643.466667
798.449276	676.663808	422.296328	394.048022	4005.788565	8103.968931	2191.719237	384.068056	19337.187121	19104.505909	13365.76319	11288.373969	98982.540700
048.000000	1645.000000	749.000000	406.000000	20974.000000	7208.000000	1698.000000	319.000000	91770.000000	28529.000000	3767.000000	4238.000000	544653.000000
347.750000	1903.750000	1432.500000	821.750000	29245.250000	9803.500000	2216.000000	408.750000	99647.000000	39111.500000	34294.750000	19789.750000	607281.250000
629.000000	2312.000000	1713.000000	1152.500000	31943.000000	17937.500000	4776.500000	748.000000	108116.000000	47337.000000	39582.500000	25708.500000	643423.000000
024.250000	2615.500000	2037.000000	1279.750000	33976.000000	20253.000000	6207.000000	867.750000	122046.000000	53584.250000	41970.500000	34744.250000	784663.250000
734.000000	4357.000000	2373.000000	1863.000000	35221.000000	36291.000000	8724.000000	1749.000000	153923.000000	103403.000000	58903.000000	45872.000000	864004.000000

Fuente: Elaboración propia

## 8. ARQUITECTURA DE DATOS

### 8.1. Arquitectura funcional de GSK

Según lo investigado en el transcurso del proyecto, se puede mostrar la siguiente arquitectura funcional de la organización.

Imagen 34. Arquitectura funcional de GSK

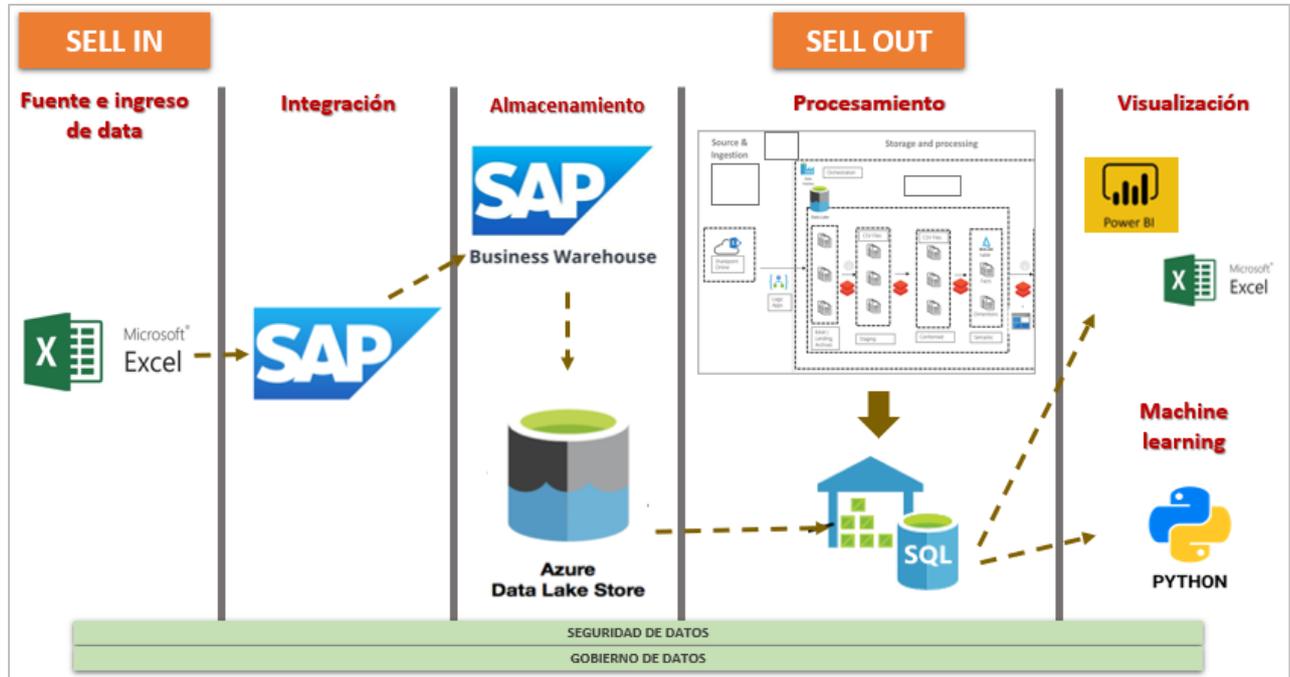


Fuente: GSK  
Elaboración propia

## 8.2. Arquitectura tecnológica, herramientas e infraestructura

Según lo investigado en el transcurso del proyecto, se puede mostrar la siguiente arquitectura funcional de la organización.

Imagen 35. Arquitectura tecnológica GSK



Fuente: GSK  
Elaboración propia

**Fuente de ingreso de data:** La data se obtiene a través de las operaciones que se realizan con los clientes o distribuidores de GSK, el cual se utiliza la herramienta Excel para su registro.

**Integración:** La data obtenida se integra en el programa SAP, el cual es un programa que permite tener el control de todas las áreas y/o unidades de la organización.

**Almacenamiento:** El primer almacenamiento de la data obtenida, se registra en SAP Business Warehouse, el cual sirve como un primer backup. Es decir, toda la información que se integra en el SAP, se “asegura” en este primer almacenamiento. Como segundo almacenamiento, se cuenta con el Azure Data Lake Store, el cual permite almacenar datos desde el SAP Business Warehouse a nivel histórico.

**Procesamiento:** En esta etapa la empresa GSK usa la herramienta SQL, el cual le permite filtrar toda la información que se requiere, en un menor tiempo y más oportuno, por un lado el Sell in (ventas a distribuidores) y/o por otro lado, el Sell out (ventas de clientes a usuarios finales).

**Visualización:** Para obtener los reportes que se requieran en el desarrollo de negocio, se utilizarán las herramientas de visualización Power BI y/o Microsoft Excel.

**Modelado:** Así como, para nuestro modelado utilizaremos la herramienta Python.

## 9. TÉCNICAS DE CIENCIA DE DATOS

### 9.1. Supervisado

En nuestro proyecto vamos a construir un modelo predictivo supervisado, puesto que se usará data histórica para entrenar los algoritmos del modelado. Además, de esta manera, podremos comprobar la viabilidad del modelo propuesto. Recordemos nuestra pregunta de ciencia de datos, el cual busca identificar el valor del monto de las devoluciones de los clientes de GSK. Por ello, según los objetivos, la técnica más oportuna para dar solución a nuestro problema es la regresión lineal.

De esta forma, podremos realizar un análisis predictivo a partir de los montos de devoluciones de los diferentes clientes de GSK, según sus variables numéricas continuas.

#### 9.1.1. Regresión lineal múltiple

Se ha realizado el análisis para nuestro modelado de regresión lineal múltiple. En este caso, se realizó el modelado a través de la plataforma Google Colab, el cual nos ayudó a poder realizar una análisis detallado de las variables. El proceso para realizar el presente modelado fue el siguiente:

- Se realizaron las correlaciones de todas las variables independientes respecto a la variable dependiente, con la finalidad de verificar cuáles de estas variables tienen mayor coeficiente de correlación con la variable dependiente “Devoluciones en soles”.
- Luego de identificar estas variables, se procedió a realizar nuestro modelo de regresión lineal múltiple con las variables seleccionadas con el que nos brindó el siguiente resultado:

## Modelo de regresión múltiple:

OLS Regression Results						
Dep. Variable:	y	R-squared:	0.994			
Model:	OLS	Adj. R-squared:	0.984			
Method:	Least Squares	F-statistic:	100.8			
Date:	Thu, 09 Dec 2021	Prob (F-statistic):	3.94e-08			
Time:	04:07:58	Log-Likelihood:	-260.94			
No. Observations:	24	AIC:	551.9			
Df Residuals:	9	BIC:	569.6			
Df Model:	14					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	-3.274e+04	1.23e+05	-0.267	0.796	-3.1e+05	2.45e+05
PBI	2.536e+07	8.59e+06	2.953	0.016	5.93e+06	4.48e+07
Dev por Fallas Química Suiza en soles	12.8599	14.183	0.907	0.388	-19.223	44.943
Dev por Sobre Stock Química Suiza en soles	12.9626	4.156	3.119	0.012	3.561	22.364
Dev por Vencimiento Química Suiza en soles	-2.9118	1.621	-1.797	0.106	-6.578	0.754
Devoluciones F del Norte Aquafresh en soles	-5942.3166	2051.116	-2.897	0.018	-1.06e+04	-1302.369
Devoluciones Química Suiza Panadol en soles	0.6661	0.827	0.805	0.441	-1.205	2.537
Dev por Sobre Stock Deco en soles	-112.3741	33.393	-3.365	0.008	-187.915	-36.833
Devoluciones F del Norte Scott en soles	-1391.8294	473.068	-2.942	0.016	-2461.985	-321.674
Dev por Vencimiento Deco en soles	26.7424	6.089	4.392	0.002	12.967	40.517
Devoluciones Deco Aquafresh en soles	149.3233	59.894	2.493	0.034	13.834	284.813
Sell-out Dimexa Eno en unidades	10.2291	7.432	1.376	0.202	-6.584	27.042
Sell-in Dimexa Productos Sanitarios en soles	0.2455	0.196	1.255	0.241	-0.197	0.688
Devoluciones Dimexa Parodontax en soles	237.6483	222.597	1.068	0.313	-265.900	741.197
Precio Promedio Dimexa Corega	1589.9046	5762.621	0.276	0.789	-1.14e+04	1.46e+04
Omnibus:	3.464	Durbin-Watson:	1.638			
Prob(Omnibus):	0.177	Jarque-Bera (JB):	2.361			
Skew:	-0.767	Prob(JB):	0.307			
Kurtosis:	3.086	Cond. No.	2.72e+08			

## Interpretación

Para resolver la pregunta principal de nuestro trabajo, estamos usando a técnica de regresión lineal múltiple debido que trataremos con un algoritmo de aprendizaje supervisado la cual es una técnica muy utilizada en machine learning y estadística.

Es decir, mediante el uso de una recta identificaremos si existe una tendencia en el conjunto de datos que estamos utilizando en nuestro modelo para la relación entre una variable escala dependiente “y” y una o más variables explicativas nombradas con “X”.

En nuestro modelo se puede observar el análisis de regresión lineal múltiple con una Variable dependiente y 14 variables independientes, para el cual se tiene la siguiente interpretación del resumen.

Estadísticas de la regresión:

- El modelo ha resultado un  $R^2$  ajustado de 0.984
- Ello indica que nuestra variables independientes responden y está explicado en un 98.4% a nuestro modelado de datos.

#### Análisis de la varianza:

- Nuestro valor crítico de F dio como resultado  $3.94e-08$
- Ello quiere decir que: Valor crítico de  $F < 0.05$ . Por lo que se puede interpretar que la hipótesis nula se rechaza y el modelo tiene significancia.

#### Coefficientes:

- Respecto al PBI, se entiende que en medida que aumente en un valor esta variable, la variable dependiente Devoluciones en soles aumentaría en 25,630,000.
- Respecto a Devoluciones por fallas Química Suiza en soles, se entiende que en medida que aumente en un valor esta variable, la variable dependiente Devoluciones en soles aumentaría en 12.8599.
- Respecto a Devoluciones por sobre stock Química Suiza en soles, se entiende que en medida que aumente en un valor esta variable, la variable dependiente Devoluciones en soles aumentaría en 12.9626.
- Respecto a Devoluciones por vencimiento Química Suiza en soles, se entiende que en medida que aumente en un valor esta variable, la variable dependiente Devoluciones en soles disminuiría en 2.9118.
- Respecto a Devoluciones F del Norte Aquafresh en soles, se entiende que en medida que aumente en un valor esta variable, la variable dependiente Devoluciones en soles disminuiría en 5942.3166.
- Respecto a Devoluciones Química Suiza Panadol en soles, se entiende que en medida que aumente en un valor esta variable, la variable dependiente Devoluciones en soles aumentaría en 0.6661.
- Respecto a Devoluciones por sobre stock Deco en soles, se entiende que en medida que aumente en un valor esta variable, la variable dependiente Devoluciones en soles disminuye en 112.3741.
- Respecto a Devoluciones F del Norte Scott en soles, se entiende que en medida que aumente en un valor esta variable, la variable dependiente Devoluciones en soles disminuye en 1391.8294.
- Respecto a Devoluciones por vencimiento Deco en soles, se entiende que en medida que aumente en un valor esta variable, la variable dependiente Devoluciones en soles aumentaría en 26.7424.

- Respecto a Devoluciones Deco Aquafresh en soles, se entiende que en medida que aumente en un valor esta variable, la variable dependiente Devoluciones en soles aumentaría en 149.3233.
- Respecto a Sell-out Dimexa Eno en unidades, se entiende que en medida que aumente en un valor esta variable, la variable dependiente Devoluciones en soles aumentaría en 10.2291.
- Respecto a Sell-in Dimexa Productos Sanitarios en soles, se entiende que en medida que aumente en un valor esta variable, la variable dependiente Devoluciones en soles aumentaría en 0.2455.
- Respecto a Devoluciones Dimexa Parodontax en soles, se entiende que en medida que aumente en un valor esta variable, la variable dependiente Devoluciones en soles aumentaría en 237.6483.
- Respecto a Precio Promedio Dimexa Corega, se entiende que en medida que aumente en un valor esta variable, la variable dependiente Devoluciones en soles aumentaría en 1589.9046.

## **10. EL MODELO Y LA HISTORIA**

### 10.1 Determinación del modelo más apropiado para la descripción o predicción

Según el análisis realizado a la data entre el periodo enero 2019 a junio 2021 y alineados a nuestro objetivo principal del proyecto que, cabe recordar, es calcular el monto de la devolución en unidades monetarias de los principales clientes de GSK, el modelo más apropiado que responde a nuestra pregunta de ciencia de datos es la de regresión lineal múltiple. Puesto que, este modelado nos permitirá conocer cuánto sería el valor de devoluciones que realizarían los principales clientes de la organización. Siento la variable independiente “Devoluciones en soles”. De esta manera, permitiría a GSK a tomar acciones preventivas y evitar pérdidas monetarios que puedan provocar estos retornos de productos.

### 10.2 Descripción de los hallazgos, ¿Qué se descubrió?

Entre los principales hallazgos podemos mencionar los siguientes:

- La empresa GSK mantiene un contrato especial con sus clientes, el cual les da la opción de devolver los productos que se les vendieron en caso lo decidan. Normalmente, estas devoluciones se dan por tres (3) razones, que son: Por vencimiento del producto, por sobre stock de los propios clientes, o por fallas en los productos. Siendo las más representativa, las devoluciones por vencimiento.
- Se identificó que los principales clientes de GSK son: Alfaro, Deco, Dimexa, F del Norte y Química suiza, en relación con su participación en las ventas de la organización.
- Las empresas con mayor proporción de devolución son Deco con 2.85% de sus compras, Alfaro con 2.46% y F del Sur con 1.31%. No obstante, si bien Química Suiza solo tiene un 0.91% de devolución de sus compras, el valor monetario que representa este genera un perjuicio para GSK. Puesto que, Química Suiza tiene un 67.10% de participación en las ventas de la empresa.
- Se debe poner mayor énfasis y cuidado a los diferentes tipos de devoluciones, ya sea por producto o tipo, mencionadas en el modelo, ya que son los que tienen un mayor impacto en las devoluciones totales.

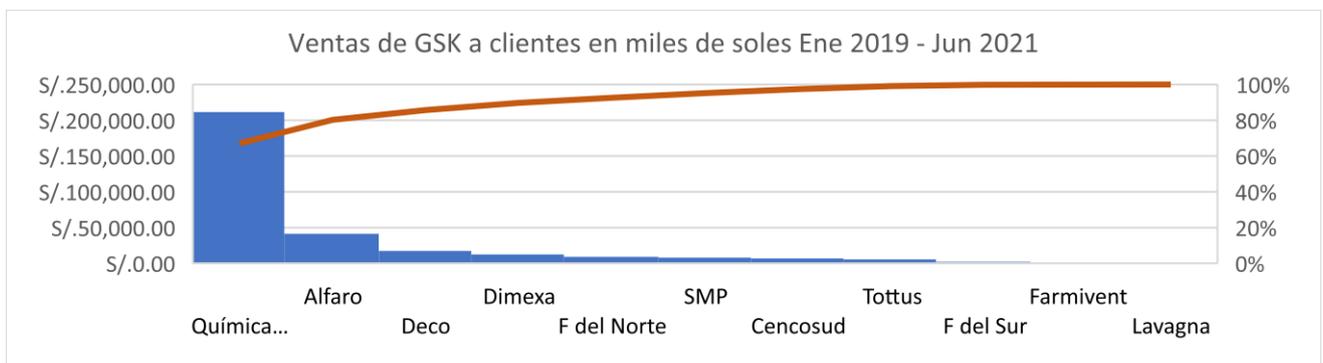
## 11. IMPLEMENTACIÓN Y AJUSTES

### 11.1 Visualización de los hallazgos, grafique lo descubierto y compártalo

Para dar respuesta a nuestras preguntas de ciencias de datos, se ha podido identificar lo siguiente:

- La empresa cuenta con los principales clientes: Química Suiza, Alfaro, Deco, Dimexa y F del Norte, tal como se muestra en la gráfica siguientes:

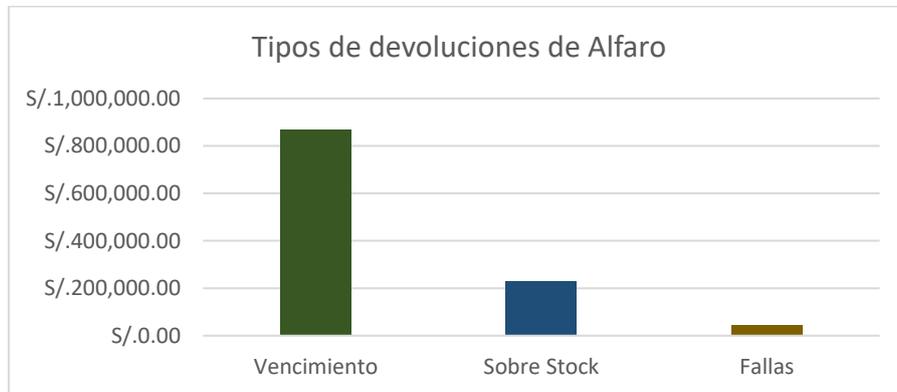
*Imagen 36. Clientes principales de GSK y participación.*



*Fuente: GSK*

- Se ha identificado los motivos de devolución: vencimiento, sobre stock y fallas. Asimismo, estas devoluciones se distribuyen por clientes de la siguiente manera:

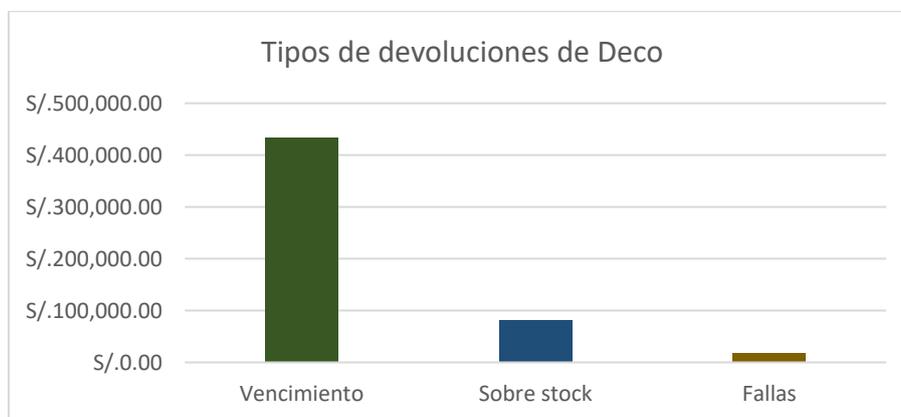
Imagen 37. Tipos de devoluciones de Alfaro



Fuente GSK

Alfaro cuenta un mayor nivel de devolución por vencimiento seguido de sobre stock y fallas.

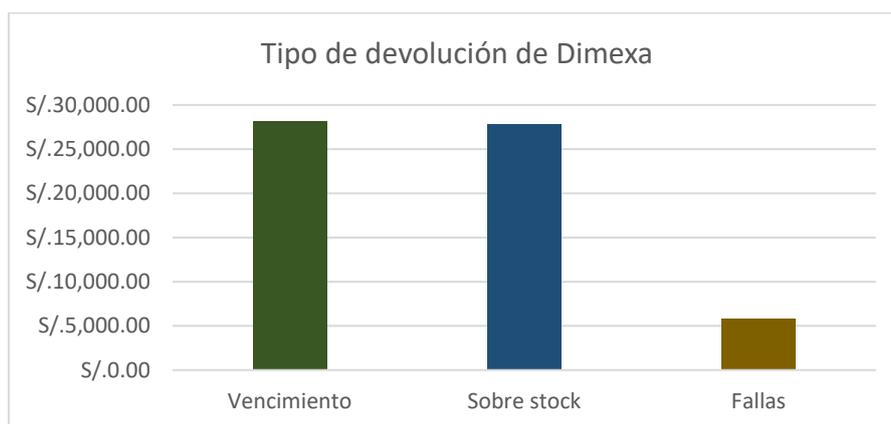
Imagen 38. Tipos de devoluciones de Deco



Fuente GSK

Deco cuenta un mayor nivel de devolución por vencimiento seguido de sobre stock y fallas.

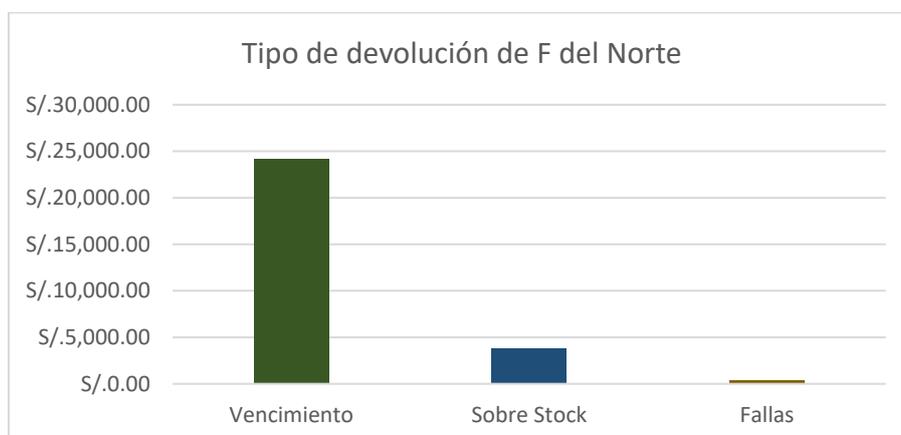
Imagen 39. Tipo de devolución de Dimexa



Fuente GSK

Dimexa cuenta un mayor nivel de devolución por vencimiento y sobre stock y seguido por la razón de fallas.

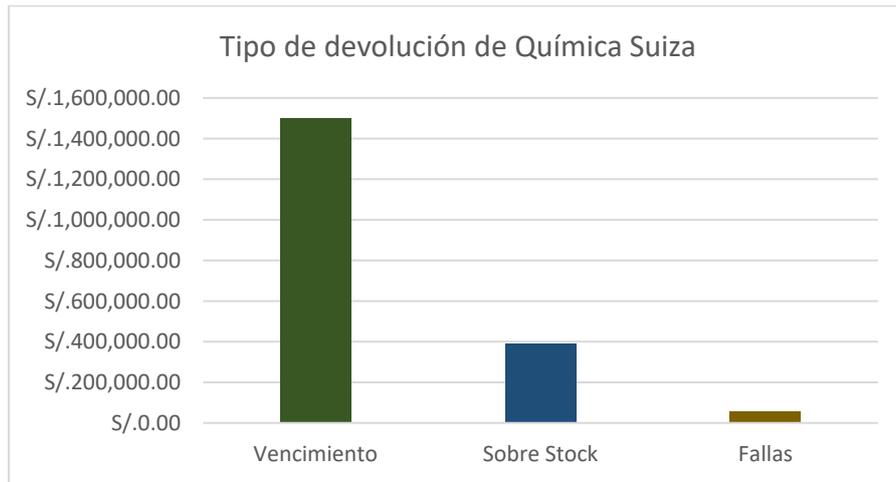
Imagen 40. Tipo de devolución de F del Norte



Fuente GSK

F del Norte cuenta un mayor nivel de devolución por vencimiento seguido de sobre stock y fallas.

Imagen 41. Tipo de devolución de Química Suiza

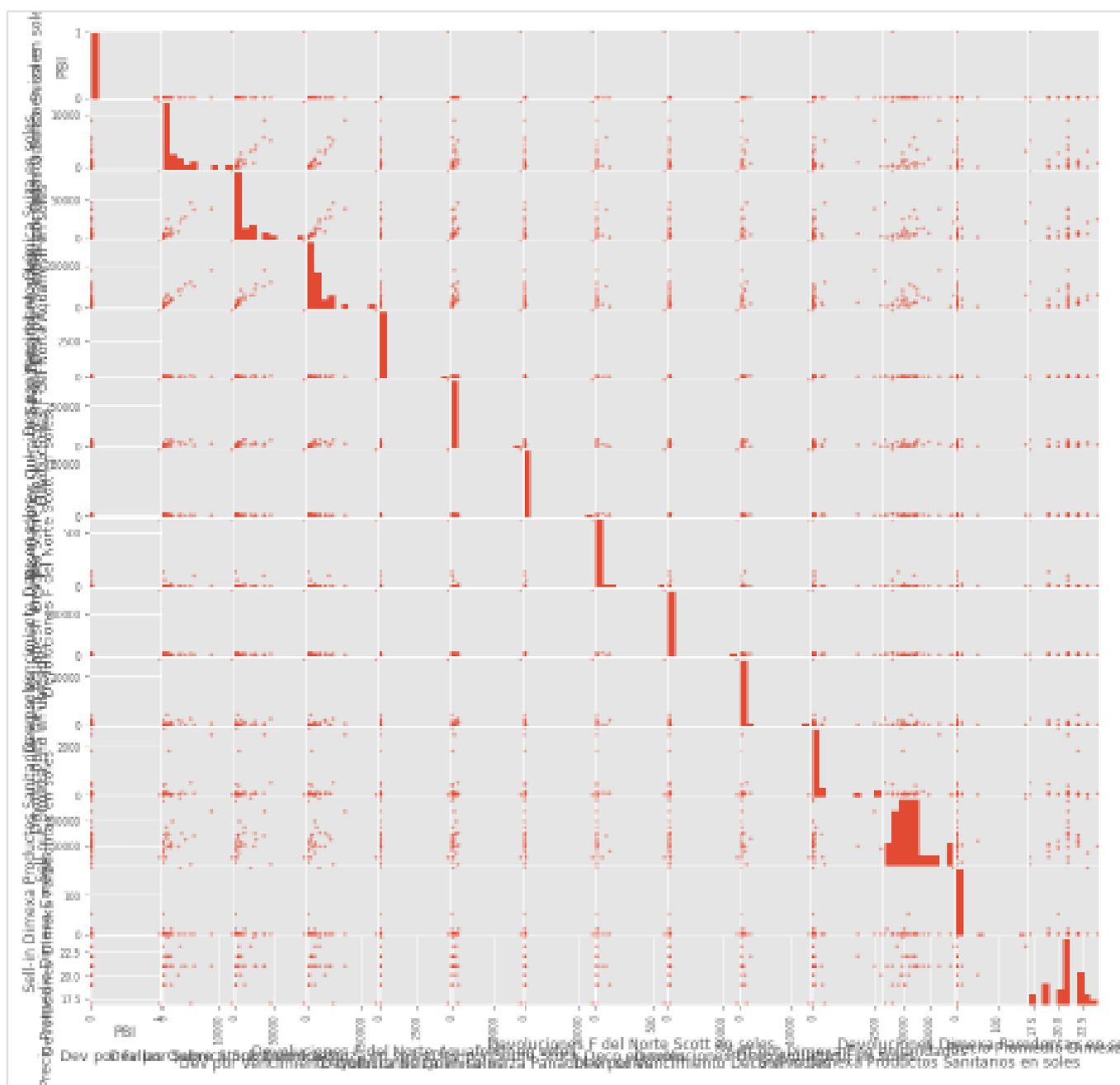


Fuente GSK

Química Suiza cuenta un mayor nivel de devolución por vencimiento seguido de sobre stock y fallas.

- Para obtener un análisis global de las variables seleccionadas en nuestro modelo, se ha generado gráficos de dispersión de todas estas variables, tanto la dependiente “Devoluciones en soles”, como las variables independientes que mayor responder a nuestro modelo. Todo ello, para lograr un análisis más complementario, como se muestra en la siguiente gráfica:

Imagen 42. Tabla de dispersión de las variables del modelo



Fuente: GSK  
Elaboración: propia

- Se ha realizado un análisis descriptivo de las variables de nuestro modelo, con la finalidad de analizarlos y validar los hallazgos encontrados, como por ejemplo: la variable que tiene un mayor impacto en las devoluciones de Química Suiza es el tipo de vencimientos, teniendo un promedio de devoluciones mensuales de S/ 49,937.60 soles. Este valor supera en más de 3 veces las devoluciones por vencimiento del cliente Deco. Además, también se puede observar que la mayoría

de los datos de las Devoluciones F del Norte Aquafresh en soles está disperso en 849.70 soles respecto a su media que es 163.60 soles, así como los datos de Dev por Sobre Stock Deco en soles que están dispersos en su mayoría en 114.59 soles respecto de su promedio 30.10 soles.

Imagen 43. Análisis descriptivo de las variables del modelo

```
[20] gskmodelo.describe(include='all')
```

	PBI	Dev por Fallas Química Suiza en soles	Dev por Sobre Stock Química Suiza en soles	Dev por Vencimiento Química Suiza en soles	Devoluciones F del Norte Aquafresh en soles	Devoluciones Química Suiza Panadol en soles	Dev por Sobre Stock Deco en soles	Devoluciones F del Norte Scott en soles	Dev por Vencimiento Deco en soles	Devoluciones Deco Aquafresh en soles	Sell-out Dimexa Eno en unidades	Sell-in Dimexa Productos Sanitarios en soles	Devoluciones Dimexa Parodontax en soles	Precio Promedio Dimexa Corega
count	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000
mean	0.033333	1907.300000	13081.666667	49937.600000	163.600000	25649.766667	2731.633333	30.100000	14446.600000	773.433333	301.166667	61319.900000	8.300000	20.800000
std	0.182574	2089.205803	18704.448832	65118.212625	849.703705	73564.773111	11459.122123	114.593239	57156.920012	2477.802442	707.967664	31838.430891	32.279117	1.584407
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	12482.000000	0.000000	17.000000
25%	0.000000	104.000000	2201.500000	11668.750000	0.000000	1458.750000	326.000000	0.000000	2491.750000	47.000000	0.000000	44148.750000	0.000000	20.000000
50%	0.000000	751.500000	5141.500000	32340.000000	0.000000	9098.500000	607.000000	0.000000	3552.500000	128.500000	47.500000	56105.000000	0.000000	21.000000
75%	0.000000	2447.500000	17395.500000	57278.500000	0.000000	25227.250000	941.500000	0.000000	5510.500000	426.000000	82.750000	71938.500000	0.000000	22.000000
max	1.000000	12475.000000	87328.000000	316044.000000	4661.000000	409551.000000	63361.000000	617.000000	316806.000000	13650.000000	2721.000000	145146.000000	171.000000	24.000000

Fuente: GSK

Elaboración: propia

- Se ha generado una nueva matriz de correlaciones que identifica a las variables que tienen una mayor relación y las que no. Por ejemplo, se muestra en la imagen 44 que la mayoría variables independientes responden a la variable dependiente Devoluciones en soles, siendo Devoluciones Química Suiza Panadol en soles el de mayor correlación con 0.923619, seguido por Devoluciones por Vencimiento Química Suiza en soles con 0.916218 de coeficiente de correlación de Pearson.

Imagen 44. Matriz de correlaciones de las variables del modelo

	Devoluciones en soles	PBI	Dev por Fallas Química Suiza en soles	Dev por Sobre Stock Química Suiza en soles	Dev por Vencimiento Química Suiza en soles	Sell-out Dimexa Eno en unidades	Devoluciones F del Norte Aquafresh en soles	Precio Promedio Dimexa Corega	Devoluciones Dimexa Parodontax en soles	Devoluciones Química Suiza Panadol en soles	Dev por Sobre Stock Deco en soles	Devoluciones F del Norte Scott en soles	Sell-in Dimexa Productos Sanitarios en soles	Dev por Vencimiento Deco en soles	Devoluciones Deco Aquafresh en soles
Devoluciones en soles	1	0.905616	0.856236	0.901106	0.916218	-0.026429	0.907838	-0.48878	0.081554	0.923619	0.893854	0.859224	-0.0204664	0.905267	-0.408878
PBI	0.905616	1	0.695635	0.749709	0.771819	-0.080344	0.99967	-0.452981	-0.048565	0.985626	0.999296	0.967315	-0.202706	0.999119	0.981513
Dev por Fallas Química Suiza en soles	0.856236	0.695635	1	0.935576	0.966594	0.174995	0.698018	-0.393619	0.026453	0.723768	0.688503	0.68521	0.055512	0.691125	0.703064
Dev por Sobre Stock Química Suiza en soles	0.901106	0.749709	0.935576	1	0.959793	0.052336	0.756865	-0.396173	0.010771	0.782172	0.743219	0.736577	-0.058651	0.746488	0.757646
Dev por Vencimiento Química Suiza en soles	0.916218	0.771819	0.966594	0.959793	1	0.096118	0.775348	-0.382345	-0.043038	0.810302	0.764948	0.740082	-0.019916	0.768089	0.778521
Sell-out Dimexa Eno en unidades	-0.026429	-0.080344	0.174995	0.052336	0.096118	1	-0.082379	-0.061575	-0.111818	-0.099627	-0.080659	-0.108526	0.045785	-0.084499	-0.101499
Devoluciones F del Norte Aquafresh en soles	0.907838	0.99967	0.698018	0.756865	0.775348	-0.082379	1	-0.450627	-0.051215	0.985701	0.998911	0.966363	-0.207513	0.998901	0.983888
Precio Promedio Dimexa Corega	-0.48878	-0.452981	-0.393619	-0.396173	-0.382345	-0.061575	-0.450627	1	0.211576	-0.470262	-0.458227	-0.487607	0.40559	-0.461037	-0.496589
Devoluciones Dimexa Parodontax en soles	0.081554	-0.048565	0.026453	0.010771	-0.043038	-0.111818	-0.051215	0.211576	1	-0.062762	-0.048482	-0.068052	0.452318	-0.052974	-0.07641
Devoluciones Química Suiza Panadol en soles	0.923619	0.985626	0.723768	0.782172	0.810302	-0.099627	0.985701	-0.470262	-0.062762	1	0.984693	0.945442	-0.204765	0.986214	0.974917
Dev por Sobre Stock Deco en soles	0.893854	0.999296	0.688503	0.743219	0.764948	-0.080659	0.998911	-0.458227	-0.048482	0.984693	1	0.9685	-0.213789	0.999816	0.981533
Devoluciones F del Norte Scott en soles	0.859224	0.967315	0.68521	0.736577	0.740082	-0.108526	0.966363	-0.487607	-0.068052	0.945442	0.9685	1	-0.237411	0.968732	0.954955
Sell-in Dimexa Productos Sanitarios en soles	-0.0204664	-0.202706	0.055512	-0.058651	-0.019916	0.045785	-0.207513	0.40559	0.452318	-0.204765	-0.213789	-0.237411	1	-0.215575	-0.235766
Dev por Vencimiento Deco en soles	0.905267	0.999119	0.691125	0.746488	0.768089	-0.084499	0.998901	-0.461037	-0.052974	0.986214	0.999816	0.968732	-0.215575	1	0.983143
Devoluciones Deco Aquafresh en soles	-0.408878	0.981513	0.703064	0.757646	0.778521	-0.101499	0.983888	-0.496589	-0.07641	0.974917	0.981533	0.954955	-0.235766	0.983143	1

Fuente: GSK

Elaboración: propia

- Para una mejor visualización de las variables con mayor correlación se muestra la siguiente matriz:

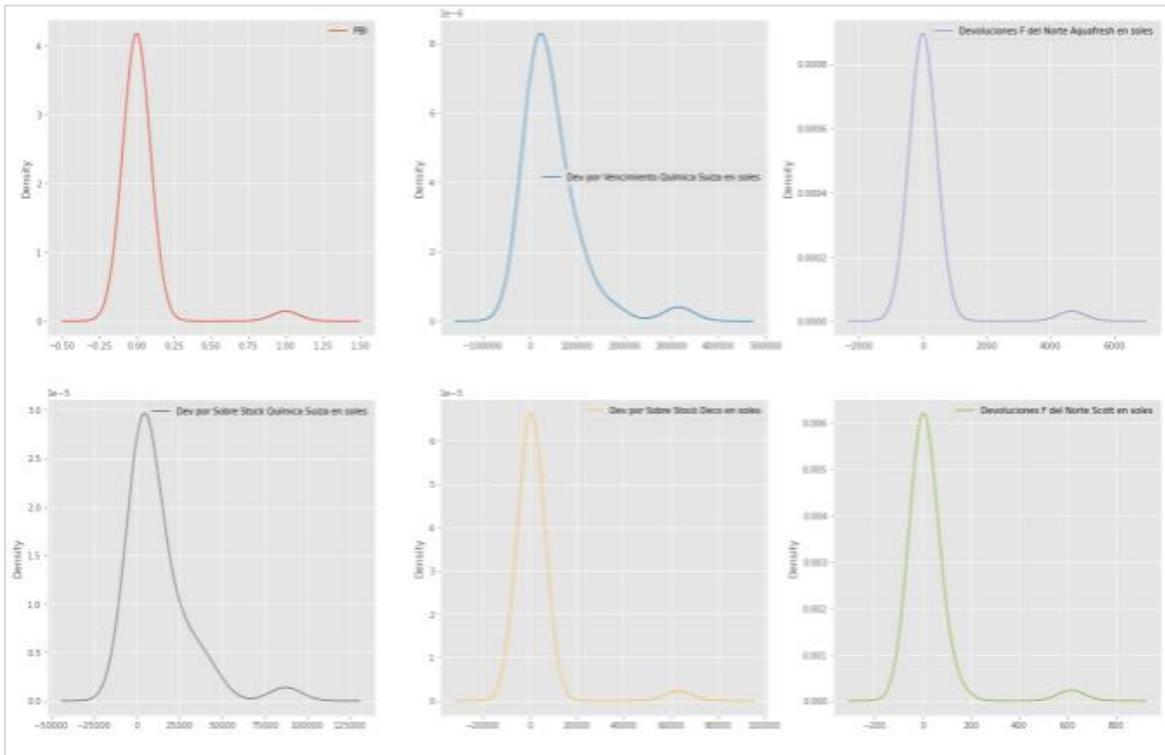
Imagen 45. Matriz de correlaciones. Variables para el modelo

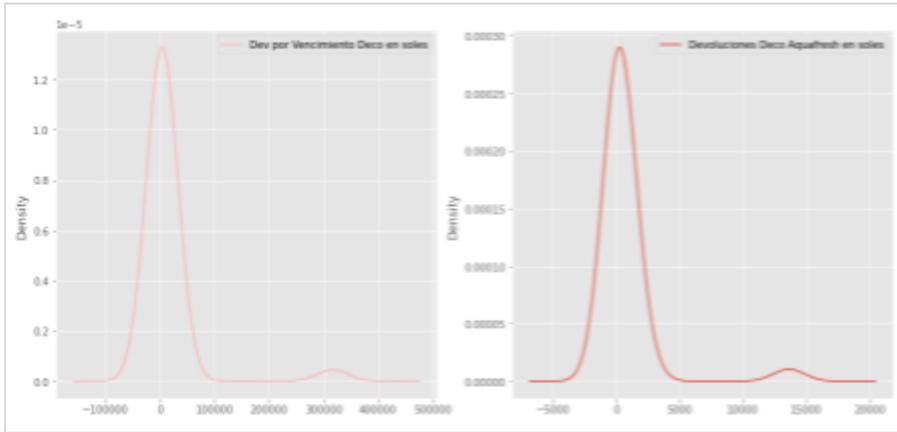
	Devoluciones en soles	PBI	Dev por Vencimiento Química Suiza en soles	Devoluciones F del Norte Aquafresh en soles	Dev por Sobre Stock Química Suiza en soles	Dev por Sobre Stock Deco en soles	Devoluciones F del Norte Scott en soles	Dev por Vencimiento Deco en soles	Devoluciones Deco Aquafresh en soles
Devoluciones en soles	1	0.905616	0.916218	0.907838	0.901106	0.902587	0.859224	0.905267	0.893854
PBI	0.905616	1	0.771819	0.99967	0.749709	0.999296	0.967315	0.999119	0.981513
Dev por Vencimiento Química Suiza en soles	0.916218	0.771819	1	0.775348	0.959793	0.764948	0.740082	0.768089	0.778521
Devoluciones F del Norte Aquafresh en soles	0.907838	0.99967	0.775348	1	0.756865	0.998911	0.966363	0.998901	0.983888
Dev por Sobre Stock Química Suiza en soles	0.901106	0.749709	0.959793	0.756865	1	0.743219	0.736577	0.746488	0.757646
Dev por Sobre Stock Deco en soles	0.902587	0.999296	0.764948	0.998911	0.743219	1	0.9685	0.999816	0.981533
Devoluciones F del Norte Scott en soles	0.859224	0.967315	0.740082	0.966363	0.736577	0.9685	1	0.968732	0.954955
Dev por Vencimiento Deco en soles	0.905267	0.999119	0.768089	0.998901	0.746488	0.999816	0.968732	1	0.983143
Devoluciones Deco Aquafresh en soles	0.893854	0.981513	0.778521	0.983888	0.757646	0.981533	0.954955	0.983143	1

Fuente: GSK

- También, se ha realizado campanas de Gauss que permiten conocer la distribución de los datos de las variables seleccionadas para nuestro modelo, como se visualiza a continuación:

Imagen 46. Campana de Gauss de las variables del modelo





*Fuente: GSK*

*Elaboración: propia*

Se puede apreciar en la Campana de Gauss cómo se distribuyen las variables evaluadas. Por ejemplo: se valida que la variable Dev vencimiento Química Suiza en soles se concentra en la media de aproximadamente 50,000 soles, sin embargo, existen datos que están muy lejos de esta media. Como se puede validar, muestra una curva de asimetría positiva lo que significaría que la mayoría de los datos se encuentran por encima del valor de su media.

Asimismo, se valida que la variable Devoluciones Deco Aquafresh en soles se concentra en la media de aproximadamente 800 soles, sin embargo, existen datos que están muy lejos de esta media. Como se puede validar, muestra una curva de asimetría positiva lo que significaría que la mayoría de los datos se encuentran por encima del valor de su media.

Un último ejemplo para mencionar es que se valida que la variable PBI se concentra en la media de aproximadamente 0.033, sin embargo, existen datos que están muy lejos de esta media. Como se puede validar, muestra una curva de asimetría positiva lo que significaría que la mayoría de los datos se encuentran por encima del valor de su media.

- Según el modelo de regresión lineal múltiple presentado en el anterior avance, el equipo de ciencia de datos ha decidido realizar un ajuste en sus variables, con la finalidad de tener un modelo que responda con mayor eficiencia a nuestra pregunta de ciencia de datos planteada. Para ello, se ha decidido tomar únicamente a las variables que tienen una mayor respuesta a la variable dependiente “Devoluciones en soles”. El criterio seleccionado es seleccionar a las variables

con un P Value > a T más cercano a cero y/o menores a 0.05, ya que son aquellas que responden de una manera más directa a la variable dependiente. Según ello, nuestro modelo de regresión lineal múltiple se representaría de la siguiente manera:

OLS Regression Results							
Dep. Variable:	y	R-squared:	0.984				
Model:	OLS	Adj. R-squared:	0.976				
Method:	Least Squares	F-statistic:	117.7				
Date:	Thu, 09 Dec 2021	Prob (F-statistic):	3.96e-12				
Time:	04:15:22	Log-Likelihood:	-271.81				
No. Observations:	24	AIC:	561.6				
Df Residuals:	15	BIC:	572.2				
Df Model:	8						
Covariance Type:	nonrobust						
	coef	std err	t	P> t	[0.025	0.975]	
const	2.485e+04	1.29e+04	1.927	0.073	-2633.023	5.23e+04	
PBI	2.663e+07	5.83e+06	4.569	0.000	1.42e+07	3.91e+07	
Dev por Vencimiento Química Suiza en soles	-2.4480	0.701	-3.490	0.003	-3.943	-0.953	
Devoluciones F del Norte Aquafresh en soles	-6134.0600	1356.660	-4.521	0.000	-9025.712	-3242.408	
Dev por Sobre Stock Química Suiza en soles	14.6242	2.797	5.228	0.000	8.662	20.586	
Dev por Sobre Stock Deco en soles	-101.3209	31.727	-3.193	0.006	-168.946	-33.695	
Devoluciones F del Norte Scott en soles	-1576.6993	306.490	-5.144	0.000	-2229.967	-923.431	
Dev por Vencimiento Deco en soles	24.0341	5.743	4.185	0.001	11.792	36.276	
Devoluciones Deco Aquafresh en soles	149.8644	39.571	3.787	0.002	65.521	234.208	
Omnibus:	4.939	Durbin-Watson:	1.386				
Prob(Omnibus):	0.085	Jarque-Bera (JB):	3.042				
Skew:	-0.594	Prob(JB):	0.218				
Kurtosis:	4.277	Cond. No.	1.17e+08				

En este sentido se ha obtenido una regresión lineal múltiple con 8 variables independientes que son: “PBI”, “Dev por Vencimiento Química Suiza en soles”, “Devoluciones F del Norte Aquafresh en soles”, “Dev por Sobre Stock Química Suiza en soles”, “Dev por Sobre Stock Deco en soles”, “Devoluciones F del Norte Scott en soles”, “Dev en Vencimiento Deco en soles” y “Devoluciones Deco Aquafresh en soles” que responden a nuestra variable dependiente “Devoluciones en soles”.

Se ha obtenido un valor de 0.976 de R<sup>2</sup> ajustado, el cual demuestra la alta correlación existente entre las variables mencionadas.

- Además, se muestra nuestra fórmula de regresión lineal múltiple el cual representa a nuestro modelo de ciencia de datos:

**Leyenda:**

Y: Variable dependiente “Devoluciones en soles”

X1: Variable independiente “PBI”

X2: Variable independiente “Dev por Vencimiento Química Suiza en soles”

X3: Variable independiente “Devoluciones F del Norte Aquafresh en soles”

X4: Variable independiente “Dev por Sobre Stock Química Suiza en soles”

X5: Variable independiente “Dev por Sobre Stock Deco en soles”

X6: Variable independiente “Devoluciones F del Norte Scott en soles”

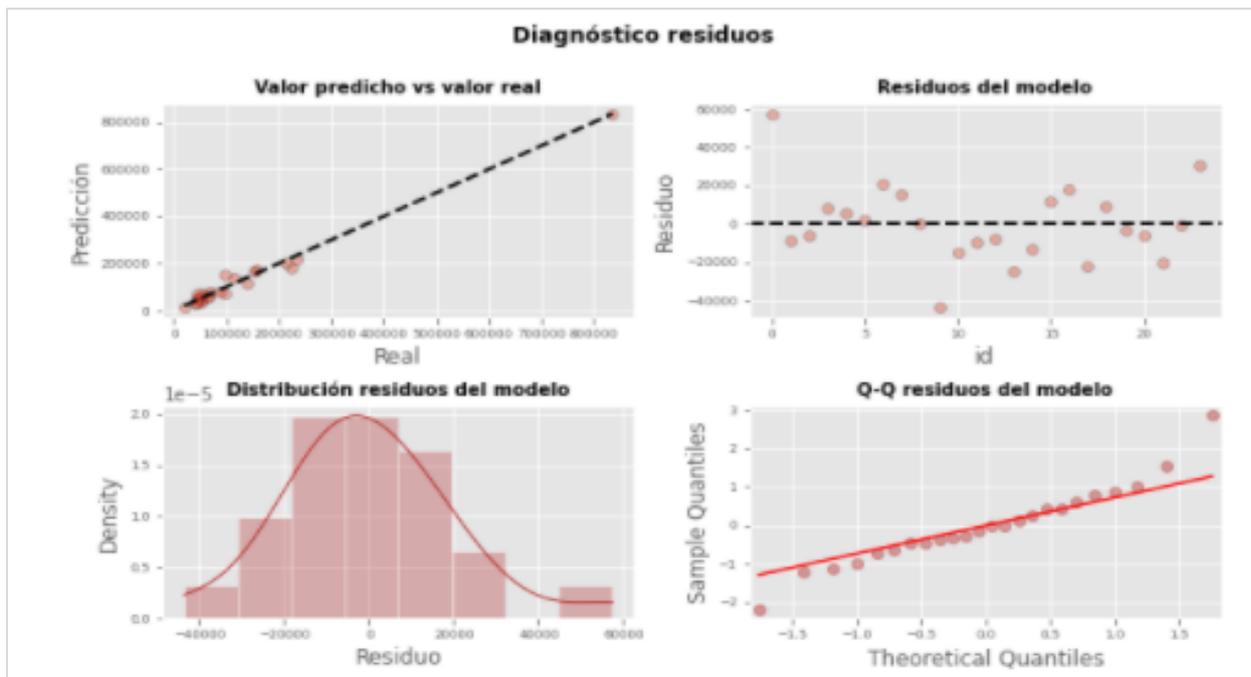
X7: Variable independiente “Dev por Vencimiento Deco en soles”

X8: Variable independiente “Devoluciones Deco Aquafresh en soles”

$$Y = 24850 + 26630000X1 - 2.4480X2 - 6134.06X3 + 14.6242X4 - 101.3209X5 - 1576.6993X6 + 24.0341X7 + 149.8644X8$$

- Por último, se muestra el diagnóstico de residuos. En el cual vemos, por ejemplo, un gráfico del valor que se predijo versus lo real, donde podemos que esta comparación entre estos valores no es tan significativa. Por otro lado, se puede ver las diferencias entre los residuos del modelo, donde se puede apreciar que estos valores se mueven entre un poco más de los -40,000 hasta los 60,000, estando la mayoría de estos entre los -20,000 hasta los 20,000. Para visualizarlos, se muestra la siguiente gráfica.

Imagen 47. Diagnóstico de residuos



Fuente: GSK  
Elaboración: propia

## 11.2 Estrategia de implementación

Luego de obtener nuestro modelo de ciencia de datos que más se acerca a nuestro objetivo, prosigue implementarlo y ponerlo a prueba de manera definitiva. Para ello, se ha decidido establecer 3 etapas de implementación que son las siguientes:

- Primera fase: Esta se ejecutará en la etapa previa a la implementación.
  - Validar que nuestra arquitectura de datos soporte nuestro modelo. La empresa GSK ya presenta una arquitectura de datos, para ello se realizará un seguimiento y optimización en caso sea requerido.
  - Generar un programa de implementación, en el que se establecerá, las etapas, las métricas de evaluación, el periodo de implementación, herramientas tecnológicas a emplear, la necesidad de capacitación y orientación por parte del equipo de ciencia de datos.
  
- Segunda fase: Esta se ejecutará durante la implementación.
  - La implementación será establecida de manera progresiva. Es decir, en primer lugar, el modelo estará disponible para los gerentes de la organización, luego se podrá desplegar para cada representante de las divisiones en la empresa, posteriormente, se ampliará el alcance hacia los colaboradores. Cada progreso tendrá un periodo de seguimiento de 3 meses y en cada uno estará el equipo de ciencia de datos otorgando soporte y capacitación en caso se requiera.
  
- Tercera fase: Esta se ejecutará en la etapa posterior a la implementación.
  - Comentarios por parte de los usuarios interesados. La implementación del modelo de datos permitirá obtener retroalimentación respecto a su viabilidad según los objetivos propuestos a inicios del proyecto. Estas opiniones permitirán al equipo de ciencia de datos realizar los ajustes necesarios, con la finalidad de optimizar la eficiencia del modelo.
  - Es importante recalcar que nuestro modelo de datos tendrá revisiones y retroalimentaciones constantes cada periodo.

## 12. CALIBRADO E IMPLEMENTACIÓN

### 12.1 Método de ajuste

El método de ajuste ayuda al equipo de ciencia de datos a analizar y validar la confiabilidad de nuestro modelo de ciencia de datos según la propuesta realizada al inicio del proyecto. Para su validación, se ha usado el método del Test de significancia de F, el cual nos ha generado el siguiente reporte:

- Significación de F-test (Validar el modelo de regresión):

Nivel de significancia: 0.05

Valor crítico de F: 3.96e-12

- R<sup>2</sup>:

Nivel de correlación: 0.976

Nuestro valor crítico de F nos confirma que se rechaza la hipótesis nula al ser un valor menor a 0.05, demostrando que nuestro modelo tiene significancia. Asimismo, nuestro R<sup>2</sup> ajustado nos indica que las variables independientes dan respuesta a nuestro modelo en un 98.3%.

### 12.2 Retroalimentación

El modelo de ciencia de datos es una representación de la realidad, es decir, nos muestra el entorno y contexto en el que se encuentra, en este caso, la empresa GSK respecto a sus operaciones. Por ello, la presentación de nuestro modelo no es estático, se tiene que ajustar y retroalimentar de manera constante, con la finalidad de que no se aleje del objetivo principal de su desarrollo.

En este sentido, el equipo de ciencia de datos ha visto necesario establecer un cronograma de retroalimentación del modelo, el cual se realizará anualmente. Puesto que, ello permitirá ajustar y recalibrar nuestro modelo según los nuevos datos históricos obtenidos por nuevas operaciones. Asimismo, esta recalibración será fundamental porque mantendría el modelo cerca de los objetivos establecidos.

Otro criterio para esta retroalimentación del modelo es debido al tipo de contrato que tienen los clientes, el que les permite devolver los productos no vendidos en un periodo no mayor a un año. Por lo que, en caso exista un cambio en las condiciones de este contrato, será motivo de recalibrar el modelo.

### **13. FACTORES CRÍTICOS DE ÉXITO DEL PROYECTO**

Según el criterio de nuestro equipo de ciencia de datos, los puntos que hay que tener en cuenta para que un proyecto funcione, son los siguientes:

- Es necesario tener data suficiente que nos permita guiar el análisis hacia el logro de los objetivos propuestos. En el caso que no se tenga la data de manera oportuna, es necesario que se ejecuten estrategias para empezar a generarlos y registrarlos.
- Es necesario que la data que se tenga esté estructurada y ordenada, el cual les facilitará a los interesados su entendimiento.
- Según la complejidad de la data, por ejemplo, el volumen, es necesario tener las herramientas tecnológicas acertadas para poder almacenarlos, transformarlos, distribuirlos y analizarlos. Así como, para generar reportes que ayuden a entender la data de manera visual.
- Es importante comprender el negocio y determinar el problema de manera eficiente, ya que ello guiará oportunamente al equipo de ciencia de datos en el análisis de este.
- Otro punto esencial, es que no debemos partir un proyecto de ciencia de datos desde los datos disponibles, sino se debe partir desde el problema del negocio.
- Además, es fundamental no solo conocer sino aplicar la iteración de la metodología de ciencia de datos, ya que en ciertas situaciones será necesario retornar procesos para volver a direccionar el proyecto según el objetivo.
- También, se debe considerar la integración y complementariedad del equipo de ciencia de datos, que puedan ofrecer no solo habilidad técnicas sino, también, habilidad blandas que contribuyan en la elaboración del proyecto.
- Por último, se tendría que determinar algunas medidas de evaluación de nuestro modelo que nos ayude a entender si tuvo el rendimiento esperado o no.

### **14. CONCLUSIONES**

El principal objetivo de este trabajo es calcular el valor monetario de las devoluciones de los cinco principales clientes de GSK, analizar las variables que determinan el motivo de las devoluciones y lo que impacta en la empresa en valor monetario, esto nos permitirá tener un mayor control de las devoluciones ya que sabríamos el problema que tendrían los clientes para devolver los productos.

El modelo de aprendizaje supervisado fue de regresión lineal lo cual nos permitió identificar el valor del monto de las devoluciones de los clientes de la empresa GSK, en este sentido nos permitió predecir valores cuantitativos según sus variables numérica continuas.

Se identificó 8 variables con alta correlación las cuales nos llevó a realizar un análisis de regresión lineal múltiple para comprender mejor la respuesta de la pregunta de ciencia de datos. Ello nos llevó a un resultado de correlación de 0.976 el cual nos indica que las variables seleccionadas responden en un 97.6% el modelo de ciencia de datos.

Haciendo un seguimiento a las variables independientes, se podrá reducir las devoluciones totales de la empresa GSK, logrando reducir las pérdidas monetarias que se incurre por retorno de ventas y el costo de destrucción de productos. También podríamos controlar el stock de la empresa y las ventas hacia nuestros cinco principales clientes según su demanda.

El producto donde mayor devolución se concentra es de Panadol. Así como el cliente que tiene mayor potencial de devolver productos es Química Suiza.

Se concluye que las devoluciones del año 2022 serán aproximadamente de S/ 1,800,000 soles.

## **15. NUEVAS PREGUNTAS**

Nuestro proyecto nos ha permitido profundizar más respecto al entorno que envuelve a GSK, ello genera nuevos caminos de investigación, la cual se puede representar en las siguientes preguntas de ciencia de datos propuestos:

- ¿Cuál es el comportamiento de las devoluciones de Química Suiza?
- ¿Qué tipo de devoluciones realizará los nuevos clientes de GSK?
- ¿Qué acciones tomar para reducir y/o mitigar las devoluciones de los clientes?

## **16. DECISIONES PARA TOMAR A PARTIR DE LOS DATOS ESTUDIADOS**

De acuerdo con el análisis detallado de los clientes de GSK y basándonos en el problema principal, a partir del uso de los datos estudiados, pudimos obtener a los cinco principales clientes que presentan el mayor índice de devolución de los productos adquiridos, que vienen a ser las empresas Alfaro, Deco, Dimexa, F. del Norte y Química Suiza.

En el presente apartado, mencionaremos las decisiones a tomar basándonos en los motivos de devolución que han presentado cada uno de los cinco principales clientes:

En primer lugar, Alfaro, cliente que presenta el mayor índice de devolución a la compañía. El motivo principal de devolución es el sobre stock, el cual acumula cada vez que compra a GSK, y que mediante el acuerdo comercial se le otorga la posibilidad de devolver la mercadería en menos de doce meses, a pesar de que los productos aún no tengan la fecha de vencimiento dentro del año. Se consideró que esta repercute en la rentabilidad de la compañía, por tal motivo basándonos en la premisa de sobre acumulación de mercadería y mediana rotación en su inventario, luego de hacer el análisis del sell out, se podría determinar, otorgarle el 30% menos en disponibilidad de compra, durante los próximos doce meses, para poder medir el resultado partiendo desde la frecuencia de compra que mantienen durante los últimos tres años.

En segundo lugar, Dimexa, basándonos en el resultado de los datos estudiados y la información recopilada, se obtuvo que el motivo de devolución que presenta este cliente es el exceso de adquisición de productos exclusivamente en las temporadas de cada uno de estos, pues pasado ello, se ha observado una caída en sus ventas según el análisis del sell out. Asimismo, cabe resaltar que su porcentaje de compra es tan elevado a diferencia de otros clientes, ya que presenta un 3.98% del total. En este sentido, la medida recomendada para este cliente es reducir su capacidad de compra en un 20%, así como otorgarle un acuerdo comercial no menor a diez meses.

En tercer lugar, Química Suiza, principal cliente de la compañía pues a junio del año 2021, ha presentado un 67.10% de participación de adquisición de mercadería, llegando a ser el mayor comprador de GSK, por lo que al mismo tiempo ha presentado porcentaje de devolución considerable en comparación con los otros clientes. Cabe mencionar, que la empresa mantiene un contrato comercial de corto periodo, pues al ser una empresa de presencia y renombre en el mercado peruano, ha visto conveniente tener productos de cercana fecha de fabricación, por lo que se ha visto reflejado en sus devoluciones constantes que no están basadas en sus ventas, sino más bien en su obligación de devolución a los tres meses de adquirida la mercadería. Por tal motivo, se concluyó como equipo en realizarle una extensión de fecha en el acuerdo comercial no menor de seis meses.

En cuarto lugar, Deco, quien ha presentado el gran índice de devolución en los últimos tres años, que lo sitúa entre los tres primeros. Luego del resultado de los datos analizados, se observó que el problema que presenta este cliente, parte de la mala estrategia de

distribución de la mercadería que ha realizado en los últimos periodos, pues menos centros de distribución han ofrecido sus productos, obteniendo así una lenta rotación de inventario, y sumándole a ello, se detectó que se presentaron constantes cambios en los pedidos de mercadería que finalmente se les entregaba más cantidades sin ser necesarias, generando así mercadería devuelta, en algunos casos el mismo día y, en otros, durante el periodo del acuerdo comercial pactado.

Finalmente, la empresa Fármacos del Norte, quienes han venido sumándose como clientes frecuentes de GSK, y se ha evidenciado en su análisis del sell in, sell out y sobre todo en las devoluciones por parte de este cliente desde el 2019 a junio 2021. Basándonos en los resultados mensuales que a su vez se complementan con las reuniones mensuales de la empresa, se logró obtener que el factor más frecuente de devolución que presenta el cliente es por el incremento de compra de manera abrupta cada mes, sin considerar las temporadas, o fechas festivas. Pues, si bien es cierto los productos que adquiere en su mayoría son de cuidado personal y el sistema respiratorio. En este sentido se recomienda que los volúmenes de venta que se pacten con este cliente vayan en función a la temporada que se presente, por lo menos en los últimos doce meses, con la finalidad de que el mismo cliente pueda medir su ritmo de compra y sin tener la necesidad de perjudicar a la compañía.

## **17. ANEXO**

Video de nuestro proyecto:

**LINK:** <https://youtu.be/tDW2bOLUfnY>



## Autorización de publicación en el Repositorio Académico UPC

### 1. Información del documento

#### Título del documento

Aplicación de Data Science para el pronóstico de las devoluciones en soles de los cinco principales clientes de la empresa GSK

#### Autores del documento

Firme e indique sus apellidos y nombres completos tal como están en su DNI y registros académicos

	Apellidos completos	Nombres completos	Código	Correo electrónico	Firma
1.	Ayala Campos	Mayra Alexandra	U201322632	U201322632@upc.edu.pe	
2.	Calderón Zapata	Jonathan Rolando	U201625155	U201625155@upc.edu.pe	
3.	Núñez Espinal	Fernando Javier	U201601677	U201601677@upc.edu.pe	
4.	Rivera Gonzáles	Tatiana Alexandra	U20181A043	U20181A043@upc.edu.pe	
5.	Rodríguez Chávez	Juan Ernesto	U201821745	U201821745@upc.edu.pe	

#### Docente(s) asesor(s)

Indique los apellidos y nombres completos de su asesor, tal como están en su DNI y registros académicos

Apellidos completos	Nombres completos	Correo electrónico
Palacios Ruiz	Julio Cesar	PCAPJPAL@UPC.EDU.PE

Facultad

Facultad de Negocios

Programa académico

Administración y Gerencia del Emprendimiento (EPI)

Modalidad

- Tesis
- Trabajo de Investigación
- Trabajo de suficiencia profesional
- Trabajo académico

Grado o título obtenido

- Bachiller
- Título profesional o licenciatura
- Maestría
- Segunda especialidad

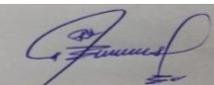
## 2. Originalidad del trabajo presentado

Con la presentación de esta ficha, el(la) autor(a) o autores(as) señalan expresamente que la obra es original, ya que sus contenidos son producto de su directa contribución intelectual. Se reconoce también que todos los datos y las referencias a materiales ya publicados están debidamente identificados con su respectivo crédito e incluidos en las notas bibliográficas y en las citas que se destacan como tal.

## 3. Autorización de publicación

La publicación de cualquier contenido requiere la autorización expresa de quien ha realizado la obra. Asimismo, en concordancia con la Ley N° 30035 que regula el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto y su Reglamento (DS N° 006-2015-PCM) todo contenido publicado en el Repositorio Académico UPC es de Acceso Abierto. Los contenidos publicados en el Repositorio Académico UPC están bajo una Licencia Creative Commons Atribución-No Comercial-Compartir Igual 4.0 Internacional. Las únicas excepciones a esta disposición se establecen en el Reglamento SICA-REG-06.

Yo, **Rodríguez Chávez Juan Ernesto** con DNI N° **70091062** en mi calidad de autor(a) y actuando en representación de mis co-autores(as), autorizo la publicación del documento indicado en el punto 1, bajo las condiciones indicadas en el punto 2 y 3, dejando constancia que el archivo digital que estoy entregando a la Universidad Peruana de Ciencias Aplicadas (UPC), como parte del proceso de obtención del título profesional o grado académico, contiene la versión final del documento sustentado y aprobado por el jurado.



Firma

09/12/2021

Fecha

## 18. REFERENCIAS BIBLIOGRÁFICAS

- García, J., Molina, J., Berlanga, A., Patricio, M., Bustamante, A. & Padilla, W. (2018). *Ciencia de datos. Técnicas analíticas y aprendizaje estadístico*. Publicaciones Altaria, S.L.
- Gartner. (2021, marzo 30). *Predictive Analytics*. <http://www.gartner.com/it-glossary/predictiveanalytics>
- GlaxoSmithKline. (2021). *About us*. Recuperado el 18 de agosto de 2021 de <https://www.gsk.com/en-gb/about-us/>
- International Business Machines Corporation. (2015). *Metodología fundamental para la ciencia de datos*. <https://www.ibm.com/downloads/cas/6RZMKDN8>
- Jara, L. (2015, setiembre 03). *Industria farmacéutica*. Observatorio Económico Social UNR. <https://observatorio.unr.edu.ar/industria-farmaceutica/>
- Organización Panamericana de la salud. (s.f.). *Acerca del Proyecto de Calidad y Regulación de Medicamentos y Tecnologías Sanitarias*. Recuperado el 18 de agosto de 2021 de [https://www3.paho.org/hq/index.php?option=com\\_content&view=article&id=2384:acerca-calidad-regulacion-medicamentos-tecnologias-sanitarias&Itemid=1179&lang=es](https://www3.paho.org/hq/index.php?option=com_content&view=article&id=2384:acerca-calidad-regulacion-medicamentos-tecnologias-sanitarias&Itemid=1179&lang=es)
- Quiroa, M. (s.f.). *Industria farmacéutica*. Economipedia. Recuperado el 18 de agosto de 2021 de <https://economipedia.com/definiciones/industria-farmaceutica.html>
- Statista. (2021). *Ranking mundial de las principales empresas farmacéuticas por valor de marca en 2020*. <https://es.statista.com/estadisticas/1027173/valor-de-marca-de-las-principales-farmaceuticas/>
- Superintendencia Nacional de Aduanas y Administración Tributaria. (2021). *Consulta Ruc*. <https://e-consultaruc.sunat.gob.pe/cl-ti-itmrconsruc/FrameCriterioBusquedaWeb.jsp>