



**Desarrollo de un modelo de gestión de datos, que permita la posibilidad de disminuir la concentración de clientes, estudio de caso: PROING CIA LTDA.**

Tipán Chamba, Ana Cristina

Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

Maestría en Gestión de Sistemas de Información e Inteligencia de Negocios

Trabajo de titulación, previo a la obtención del título de Magíster en Gestión de Sistemas de Información e Inteligencia de Negocios

PhD. Fonseca Carrera, Efraín Rodrigo

16 de diciembre del 2021

**Copyleaks**  
**Informe de Originalidad**



Identical Words	132
Words with Minor Changes	5
Paraphrased Words	144
Omitted Words	372

Firma:



EFRAIN RODRIGO  
FONSECA CARRERA

PhD. Fonseca Carrera, Efraín Rodrigo

Director

C.C.: 1710979574



**VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE  
TECNOLOGÍA  
CENTRO DE POSGRADOS**

**CERTIFICACIÓN**

Certifico que el trabajo de titulación, **"Desarrollo de un modelo de gestión de datos, que permita la posibilidad de disminuir la concentración de clientes, estudio de caso: PROING CIA LTDA"** fue realizado por la Señorita **Tipán Chamba, Ana Cristina** el mismo que ha sido revisado y analizado en su totalidad, por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 16 de diciembre del 2021

Firma:



Firmado electrónicamente por:  
**EFRAIN RODRIGO  
FONSECA CARRERA**

PhD. Fonseca Carrera, Efraín Rodrigo

Director

C.C.: 1710979574



VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE  
TECNOLOGÍA  
CENTRO DE POSGRADOS

RESPONSABILIDAD DE AUTORÍA

Yo, **Tipán Chamba, Ana Cristina**, con cédula de ciudadanía n° 1719286500, declaro que el contenido, ideas y criterios del trabajo de titulación: **"Desarrollo de un modelo de gestión de datos, que permita la posibilidad de disminuir la concentración de clientes, estudio de caso: PROING CIA LTDA"** es de mi autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 16 de diciembre del 2021

Firma (s)

Tipán Chamba, Ana Cristina

C.C.: 1719286500



VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y TRANSFERENCIA DE  
TECNOLOGÍA  
CENTRO DE POSGRADOS

**AUTORIZACIÓN DE PUBLICACIÓN**

Yo, **Tipán Chamba, Ana Cristina** con cédula de ciudadanía n° 1719286500, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: "**Desarrollo de un modelo de gestión de datos, que permita la posibilidad de disminuir la concentración de clientes, estudio de caso: PROING CIA LTDA**" en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 16 de diciembre del 2021

Firma (s)

Tipán Chamba, Ana Cristina

**C.C.: 1719286500**

## Agradecimiento

Dios en su infinito amor y grandeza, siempre me ha permitido estar en el espacio oportuno, para la enseñanza precisa, gracias por guiar mi camino, por todas sus bendiciones que día a día me acompañan, por concederme fortaleza y claridad para sobrellevar los bemoles de la vida.

Si hay algo que aporta a los logros, son la actitud positiva, el esfuerzo, el trabajo honesto y el amor, y para mí el amor más grande está representado en mi hijo Andrés y en mis padres Rosita y Hernán, gracias por ser y estar SIEMPRE. ¡Los amo infinitamente!

Quiero expresar mi más profundo agradecimiento al Ing. Pablo Trujillo, Gerente General de Proing Cia Ltda, por abrirme las puertas de su empresa para el desarrollo del presente trabajo de titulación, y sobre todo por su gran don de gente.

También quiero agradecer a la Universidad de las Fuerzas Armadas ESPE y a cada uno de los maestros que, de la manera más profesional, cálida y humana supieron impartir sus cátedras dentro de la maestría.

Un agradecimiento especial a PhD. Rodrigo Fonseca, por su acertada dirección en la elaboración del presente proyecto.

Gracias (con G) a todos a quienes la vida me ha permitido llamarlos amigos y quienes han estado pendientes de este proceso, algunos más cercanos, otros, ahora más distantes, pero en el avance de este proyecto, seguro se quedaron marcadas las palabras de aliento y buenas vibras. Gracias por los mensajes que me envía la vida, para saber que aún con el paso de los años y la distancia, en el camino siempre hay gente con el alma bonita a la que puedes regresar a ver, gracias mi querida Anita Z. por tu ayuda y tu tiempo para concluir los detalles finales de este trabajo 😊.

*Ana Cristina T.*

### **Dedicatoria**

Quiero dedicar este logro a quienes me han acompañado en el espacio y tiempo en el que he tenido la oportunidad de tocar tangencialmente a este mundo, en lo efímero de la existencia humana, a quienes me han acompañado en mis aciertos y desatinos, en quienes encuentro refugio y apoyo, a quienes me permiten mirar la vida con positivismo, energía, fortaleza y amor.

A Dios, por ser luz en mi camino.

A mi hijo Pablo Andrés, ¡¡¡eres el mejor!!! y el motivo de muchas de mis inspiraciones.

A mis padres Rosita y Hernán, ¿habrá bendición más grande que su amor?

*Ana Cristina T.*

## Índice de contenido

Copyleaks .....	2
CERTIFICACIÓN .....	3
RESPONSABILIDAD DE AUTORÍA.....	4
AUTORIZACIÓN DE PUBLICACIÓN .....	5
Agradecimiento .....	6
Dedicatoria .....	7
Índice de contenido .....	8
Índice de Tablas .....	12
Índice de Figuras.....	14
Resumen.....	16
Abstract.....	17
Introducción.....	18
Antecedentes.....	18
Problema de Investigación.....	20
Planteamiento del Problema .....	21
Objetivos.....	25
Objetivo general.....	25
Objetivos específicos .....	25
Justificación, importancia y alcance del proyecto .....	25



Preguntas de Investigación .....	26
Hipótesis de investigación.....	27
Marco Metodológico .....	28
Estudio del estado del arte.....	28
Objetivo.....	28
Preguntas de Investigación .....	28
Aproximación de la revisión de literatura preliminar .....	28
Metodología .....	40
Marco Teórico Referencial .....	44
Categorización de las variables .....	44
Fundamentación de la Variable Independiente .....	45
Tecnologías de la Información y Comunicación .....	45
Inteligencia de Negocios .....	45
Analítica de Datos .....	50
Modelo de gestión de datos .....	57
Fundamentación de la variable independiente .....	57
Empresa .....	57
Administración Comercial.....	57
Proceso de Ventas.....	58
Prospección de clientes .....	58
Diseño y Construcción de la Solución .....	59
Comprensión del Negocio.....	60

	10
Entendimiento de los datos .....	61
Ubicación Geográfica .....	61
Vertical de Negocio .....	62
Ingresos Anuales .....	63
Utilidad neta .....	63
Variables que afectan a la vertical de negocio .....	63
Preparación de los datos .....	64
Selección de los datos .....	64
Construcción del set de datos .....	66
Modelado .....	67
Carga de set de datos .....	67
Librerías para correr K-means.....	68
Cambio de variable para la identificación de cada fila .....	69
Preparación del Set de datos .....	69
Estimación del número de clústers.....	70
Creación de los clústers .....	74
Visualización gráfica de los clústers .....	75
Modelo K- Means con 4 centroides .....	78
Modelo K-means con 5 centroides .....	79
Modelo K-means con 6 centroides .....	81
Dataframe con datos clusterizados .....	82
Análisis de los datos .....	83
Validación de la Solución .....	90
Prospección por parte del equipo comercial.....	90

	11
Informe de prospección haciendo uso del modelo .....	91
Evaluación de variables del informe de prospección .....	91
Validación por expertos.....	92
Encuesta de validación .....	92
Objetivo de la encuesta.....	93
Preguntas de la encuesta.....	93
Propósito de la encuesta.....	94
Resultados de la encuesta .....	95
Conclusiones y Recomendaciones.....	102
Conclusiones .....	102
Recomendaciones .....	105
Bibliografía .....	106

## Índice de Tablas

<b>Tabla 1</b> Estudios por grupo de control .....	30
<b>Tabla 2</b> Selección de palabras clave .....	31
<b>Tabla 3</b> Cadenas de búsqueda probadas en Science Direct.....	32
<b>Tabla 4</b> Selección de estudios relevantes.....	33
<b>Tabla 5</b> Estudios primarios .....	34
<b>Tabla 6</b> Metodología Design Science .....	42
<b>Tabla 7</b> Técnicas de Investigación.....	43
<b>Tabla 8</b> Arquitectura de Inteligencia de Negocios .....	46
<b>Tabla 9</b> Comparativa metodologías para minería de datos.....	56
<b>Tabla 10</b> Variables para Ubicación Geográfica .....	62
<b>Tabla 11</b> Recesión económica por ubicación geográfica .....	63
<b>Tabla 12</b> Recesión económica por vertical de negocio .....	64
<b>Tabla 13</b> Sectores y subsectores para el modelo analítico .....	65
<b>Tabla 14</b> Variables para desarrollo del modelo de clusterización.....	66
<b>Tabla 15</b> Formato de los datos a modelarse.....	66
<b>Tabla 16</b> Variables del set de datos.....	67
<b>Tabla 17</b> Cálculo del número de clústers.....	72
<b>Tabla 18</b> Clústers seleccionados .....	84
<b>Tabla 19</b> Clientes potenciales para K-means con 4 centroides.....	85
<b>Tabla 20</b> Clientes potenciales para K-means con 5 centroides.....	85
<b>Tabla 21</b> Clientes potenciales para K-means con 6 centroides.....	87
<b>Tabla 22</b> Set de datos resultado del modelo desarrollado .....	88
<b>Tabla 23</b> Variables o criterios para evaluación de la prospección.....	91
<b>Tabla 24</b> Variables comerciales luego de aplicar el modelo.....	91

<b>Tabla 25</b> Fórmula Tasa de Conversión – Prospección.....	92
<b>Tabla 26</b> Preguntas de Investigación y cuestionario .....	93

## Índice de Figuras

<b>Figura 1</b> Clasificación de las empresas .....	18
<b>Figura 2</b> Problemática identificada y los efectos derivados .....	22
<b>Figura 3</b> Causas de la problemática identificada .....	23
<b>Figura 4</b> Árbol de Problemas.....	24
<b>Figura 5</b> Planeación de la revisión.....	29
<b>Figura 6</b> Frecuencia de publicaciones por año .....	39
<b>Figura 7</b> Ciclos de Design Science Research.....	41
<b>Figura 8</b> Categorización de las variables.....	44
<b>Figura 9</b> Arquitectura de Inteligencia de Negocios .....	46
<b>Figura 10</b> Arquitectura Kimball .....	48
<b>Figura 11</b> Arquitectura Inmon.....	49
<b>Figura 12</b> Técnicas para Modelos Predictivos .....	51
<b>Figura 13</b> Fases del proceso de KDD.....	54
<b>Figura 14</b> Fases de la metodología SEMMA .....	55
<b>Figura 15</b> Fases de la metodología CRISP DM .....	55
<b>Figura 16</b> Proceso de Ventas.....	58
<b>Figura 17</b> Fases CRISP DM aplicadas en el proyecto .....	60
<b>Figura 18</b> Verticales de negocio .....	62
<b>Figura 19</b> Set de datos para el modelo K-means.....	68
<b>Figura 20</b> Librerías instaladas .....	68
<b>Figura 21</b> Filas identificadas con el nombre de la empresa .....	69
<b>Figura 22</b> Código para normalizar los datos .....	70
<b>Figura 23</b> Dataset normalizado .....	70
<b>Figura 24</b> Número de Clústers por el método wss.....	71

<b>Figura 25</b> Código para calcular número de clústers .....	71
<b>Figura 26</b> Frecuencias del número de clústers .....	73
<b>Figura 27</b> Número de clústers por distancia inter grupos.....	73
<b>Figura 28</b> Número de clústers por el método de distancias intergrupos .....	74
<b>Figura 29</b> Clusterización.....	75
<b>Figura 30</b> Visualización inicial de los clústers .....	76
<b>Figura 31</b> Visualización clústers distancia euclidiana .....	77
<b>Figura 32</b> Visualización agrupación elíptica.....	78
<b>Figura 33</b> K-means con 4 centroides.....	78
<b>Figura 34</b> Comparación de variables para K-means con 4 centroides.....	79
<b>Figura 35</b> K-means con 5 centroides .....	80
<b>Figura 36</b> Comparación de variables para K-means con 5 centroides.....	80
<b>Figura 37</b> K-means con 6 centroides.....	81
<b>Figura 38</b> Comparación de variables para K-means con 6 centroides.....	82
<b>Figura 39</b> Agrupamiento de datos dentro de cada clúster .....	83
<b>Figura 40</b> Exportación del dataframe a excel .....	84
<b>Figura 41</b> Proceso de validación del modelo .....	90
<b>Figura 42</b> Resultado de la encuesta a la Pregunta No.1.....	95
<b>Figura 43</b> Resultado de la encuesta a la Pregunta No.2.....	96
<b>Figura 44</b> Resultado de la encuesta a la Pregunta No.3.....	97
<b>Figura 45</b> Resultado de la encuesta a la Pregunta No.4.....	98
<b>Figura 46</b> Resultado de la encuesta a la Pregunta No.5.....	99
<b>Figura 47</b> Resultado de la encuesta a la Pregunta No.6.....	100
<b>Figura 48</b> Resultado de la encuesta a la Pregunta No.7.....	101

## **Resumen**

PROING CIA. LTDA., ubicada en la ciudad de Quito, es una empresa proveedora de servicios para el diseño e implementación de sistemas de cableado estructurado, redes eléctricas y moving para data center. PROING ha concentrado el 90% de sus ingresos por facturación en un segmento reducido de tres clientes, lo cual ha desembocado en falta de liquidez. Ante estas dificultades, PROING busca reducir la concentración de su facturación, mediante la prospección de clientes, en los que se puedan identificar oportunidades de negocio, cuyas posibles ventas le permitan mantener el punto de equilibrio. PROING no cuenta con un plan de prospección que le permita disminuir la concentración de clientes, actualmente esta toma de decisiones, la realizan con base en la intuición del gerente general, lo cual no ha dado el resultado esperado. Esta problemática ha motivado el planteamiento del presente proyecto de investigación, en el que se desarrolla un modelo de gestión de datos, basado en analítica, que permite identificar clientes potenciales, con el propósito de incrementar la posibilidad de que PROING CIA LTDA disminuya la concentración de clientes. Como resultado de este estudio hemos obtenido un modelo de datos, que permite identificar clientes potenciales, que posibiliten la disminución de la concentración de clientes de PROING.

### **PALABRAS CLAVE**

- **MODELO ANALÍTICO DE DATOS**
- **MINERIA DE DATOS**
- **CONCENTRACION DE CLIENTES**
- **PROSPECCIÓN DE CLIENTES**
- **PYMES**



### **Abstract**

PROING CIA. LTDA. located in the city of Quito, is a service provider for the design and implementation of structured cabling systems, electrical networks and moving for data centers. PROING has concentrated 90% of its billing income in a reduced segment of three clients, which has led to a lack of liquidity. Faced with these difficulties, PROING seeks to reduce the concentration of its invoicing by prospecting for clients where it can identify business opportunities whose potential sales will enable it to maintain its break-even point. PROING does not have a prospecting plan that would allow it to reduce the concentration of clients; currently, decisions are made based on the general manager's intuition, which has not yielded the expected results. This problem has motivated the approach of the present research project, in which a data management model is developed, based on analytics, which allows identifying potential customers, with the purpose of increasing the possibility of PROING CIA LTDA to reduce the concentration of customers. As a result of this study, we have obtained a data model that allows identifying potential customers, which makes it possible to decrease the concentration of PROING's customers.

#### **KEY WORDS**

- ANALYTICAL DATA MODEL**
- DATA MINING**
- CUSTOMER CONCENTRATION**
- CUSTOMER PROSPECTING**
- SMES**

## Capítulo I

### Introducción

#### Antecedentes

La empresa es una organización social, cuyas actividades están enfocadas en cumplir determinados objetivos, por ejemplo, satisfacer las necesidades de un mercado específico, con la finalidad de obtener lucro económico, o no; para lo cual, se emplean recursos financieros, materiales tecnológicos, humanos, entre otros (Thompson, 2007).

Las organizaciones productiva o empresas se clasifican de acuerdo a su tamaño, ingresos y número de trabajadores. Para el caso de Ecuador, esta clasificación se acoge al Programa Estadístico Comunitario de la Comunidad Andina de Naciones, que define los aspectos básicos para elaborar las estadísticas comunitarias de las empresas en la región (CCQ, 2017), misma que se presenta en Figura 1.

#### Figura 1

##### *Clasificación de las empresas*

0 1	<b>MICROEMPRESAS</b>	Volumen de Ventas anuales: menor a \$100.000,00 Personal empleado: 1 a 9
0 2	<b>PEQUEÑA EMPRESA</b>	Volumen de Ventas anuales: Entre \$100.001,00 y \$1 Millón Personal empleado: 10 a 49
0 3	<b>MEDIANA EMPRESA "A"</b>	Volumen de Ventas anuales: Entre \$1 Millón y \$ 2 Millones Personal empleado: 50 a 99
0 4	<b>MEDIANA EMPRESA "B"</b>	Volumen de Ventas anuales: Entre \$2 Millones y \$ 5 Millones Personal empleado: 100 a 199
0 5	<b>GRANDE EMPRESA</b>	Volumen de Ventas anuales: Mayor a 5 Millones Personal empleado: Mayor a 200

*Nota.* Tomado del Artículo 3 (INEC, 2017) de la Comunidad Andina de Naciones

En Ecuador apenas el 0.46% del total de empresas corresponden a grandes empresas y el 99.54% corresponden a PYMES (INEC, 2017).

Las PYMES representan el 99.54% de las organizaciones productivas, tanto comerciales, como de servicios e industriales, las cuales generan el 60% del empleo y participan en el 50% de la producción y generación de servicios; por lo tanto, desempeñan un rol fundamental en la economía nacional. No obstante, este tipo de empresas enfrentan serias dificultades en el desarrollo de sus actividades cotidianas (Ron, 2017).

Entre las dificultades que enfrentan las PYMES está la falta de liquidez, de acuerdo con (Wyderka, 2018) las principales causas corresponden a: (1) Cuentas por cobrar, es decir tener demasiadas facturas pendientes de cobro y (2) Bajo volumen de ventas, mismas que pueden ser por mal servicio al cliente, escasa cartera de clientes.

En cuanto al bajo volumen de ventas, se ha evidenciado que una de las causas es que las PYMES centran sus ventas en pocos clientes, lo cual representa un riesgo de concentración. Una empresa que depende de un segmento reducido de clientes, en al menos 10% de sus ingresos anuales, se considera una empresa con alta concentración de clientes (Marín, 2020). La dependencia de la facturación de la compañía en un número reducido de clientes puede representar pérdida de ingresos, si sus principales clientes presentan dificultades financieras, se declaran en bancarrota, cambian de proveedor o inclusive si decidieran solventar internamente los servicios que contrata con las PYMES (Dhaliwal, 2015), es decir, que a menor número de clientes, mayor es el grado de dependencia financiera que tienen las PYMES de sus clientes (Manriquez, 2018).

Se ha evidenciado que empresas con alta concentración de clientes corren serios riesgos de tener altas pérdidas que afecten a la operación y continuidad del negocio (Flecha, 2020), por lo cual es importante considerar este riesgo y acciones que lo mitiguen.

El presente trabajo de investigación aborda un estudio de caso para la empresa PROING CIA LTDA, misma que registra una alta concentración de facturación en un segmento reducido de clientes.

### **Problema de Investigación**

PROING CIA. LTDA es una empresa ecuatoriana proveedora de servicios eléctricos, electrónicos y sistemas informáticos, que enfoca sus actividades comerciales en el diseño e implementación de redes eléctricas, sistemas de cableado estructurado, y moving de data centers<sup>1</sup>.

Esta empresa inicia sus actividades en el año 2006 en la ciudad de Quito, siendo sus principales clientes empresas integradoras de tecnología, con las cuales han podido desarrollar importantes proyectos tanto a nivel local y fuera del país.

Uno de los objetivos estratégicos de PROING es la de ampliarse en el mercado, manteniendo altos estándares de calidad que exige la industria, otorgando el mejor servicio y atención a sus clientes, para lo cual busca la optimización de sus procesos y mejora continua.

---

<sup>1</sup> Moving de data center o traslado físico del equipamiento para los centros de datos.

## Planteamiento del Problema

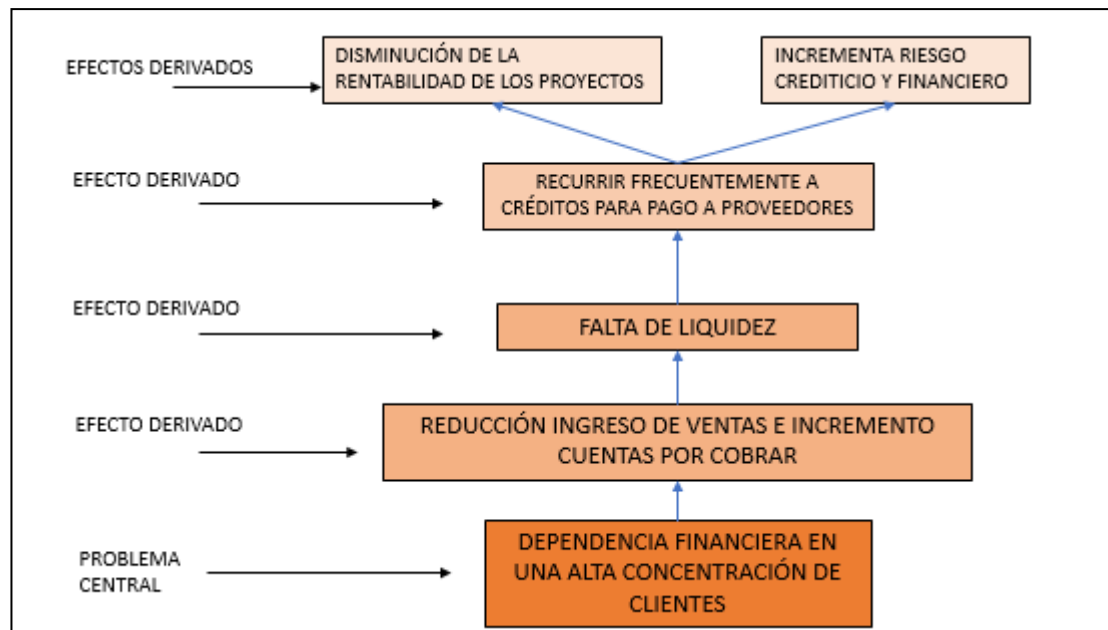
PROING CIA LTDA ha logrado mantenerse en el mercado por más de 14 años; no obstante, en los últimos 2 años ha enfrentado dificultades para hacer frente a los pagos de nómina, proveedores y gastos corrientes, debido a problemas de liquidez.

La gerencia administrativa ha realizado un estudio previo frente a esta problemática, el mismo corresponde a un análisis financiero, mediante revisión de balances y estados financieros, determinando que la principal problemática por la que PROING presenta falta de liquidez, corresponde a la *dependencia de su facturación en un número reducido de clientes*, ya que el 90% de lo que factura PROING se encuentra concentrado en tres de sus principales clientes y en los dos últimos años se ha evidenciado una reducción de dichos ingresos de ventas, además se ha identificado un importante incremento en cuentas por cobrar, que al no ser pagadas a tiempo por los pocos clientes de PROING, desemboca en una falta de liquidez. Lo cual coincide con lo encontrado en la literatura, citado previamente.

Ante la falta de liquidez, PROING CIA LTDA ha tenido que recurrir a préstamos con entidades financieras, para cubrir sus obligaciones de pagos, principalmente pago a proveedores, además la Gerencia Administrativa ha detectado, mediante la revisión de sus balances, la disminución del margen de los proyectos que se ejecutan, pues al recurrir a créditos externos, la utilidad de los proyectos se reduce, debido a que la empresa debe asumir intereses por créditos, así como también se identifica un incremento de riesgo financiero y crediticio. En la Figura 2 se puede visualizar la problemática identificada y sus efectos.

**Figura 2**

*Problemática identificada y los efectos derivados*

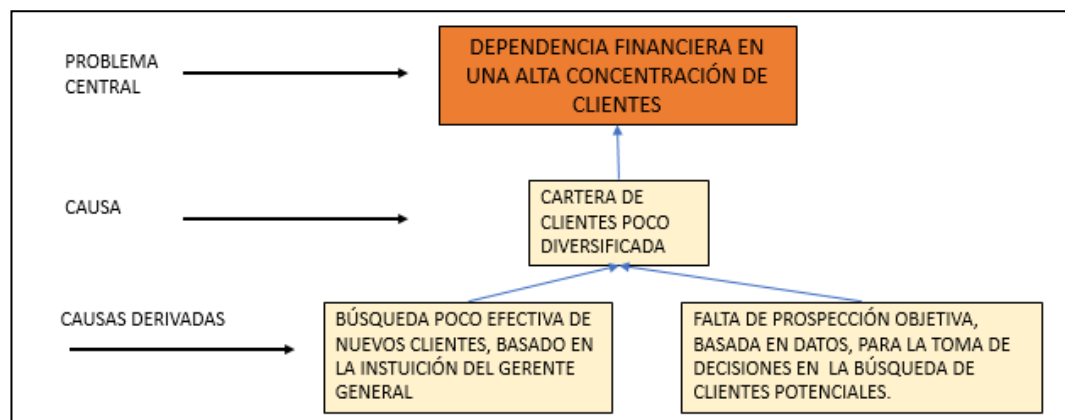


*Nota.* Muestra los efectos derivados del problema central identificado.

En el caso de PROING, la alta concentración de clientes, se debe principalmente, a que cuenta con una cartera poco diversificada, el Gerente General concluye que, la búsqueda de nuevos clientes, ha sido poco efectiva, basada en la intuición. No se ha realizado una prospección objetiva basada en datos para la tomar decisiones, que permita la búsqueda de clientes potenciales, en la Figura 3 se pueden visualizar las causas que fueron determinadas.

### Figura 3

#### *Causas de la problemática identificada*

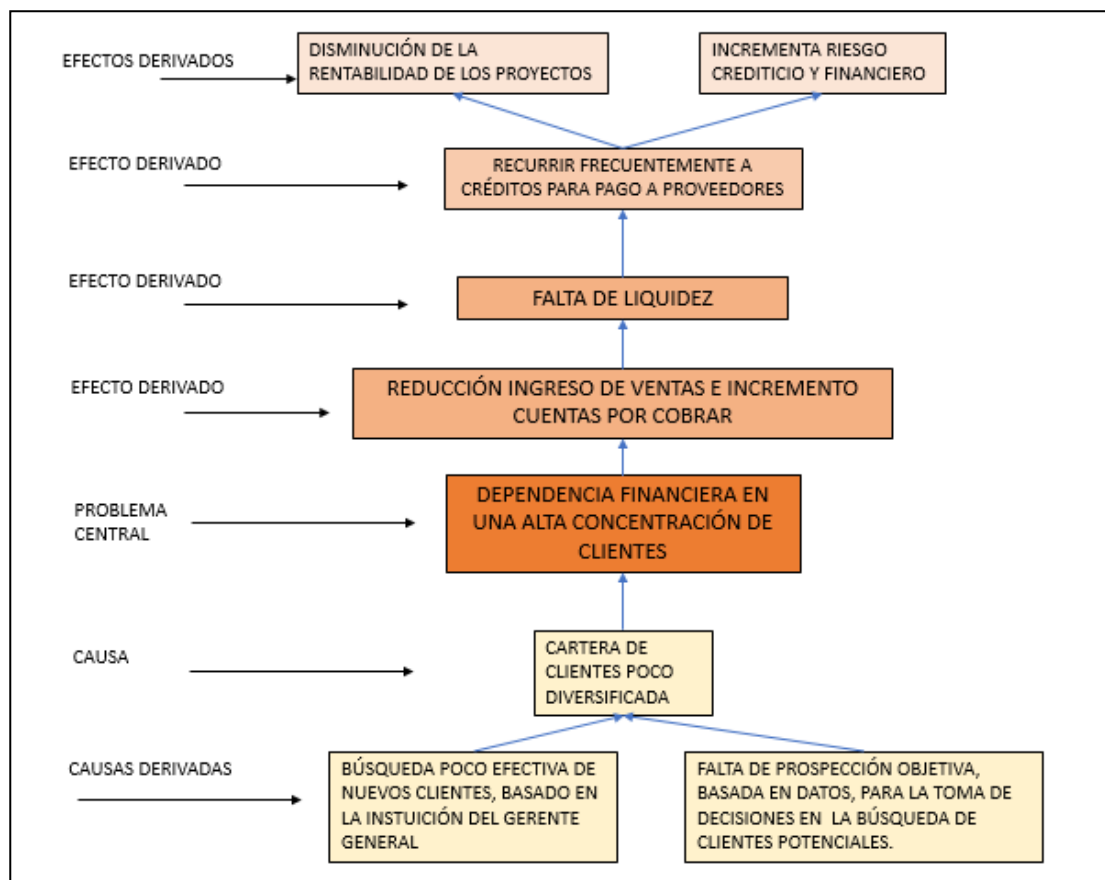


*Nota.* Muestra las causas derivadas del problema central identificado.

Con base en lo expuesto, se elabora un árbol de problemas, mostrado en la Figura 4, donde se describen la causas y efectos de la problemática identificada: Dependencia financiera en una alta concentración de clientes, en la empresa estudio de caso PROING CIA LTDA.

Figura 4

## Árbol de Problemas



*Nota.* Muestra el árbol de problemas con todas las causas y efectos identificados

El presente proyecto abarca la alta concentración de clientes, como causa de la falta de liquidez que afronta la empresa estudio de caso, y para mitigar los efectos que la misma produce, se aborda una de las causas derivadas relacionada con la falta de prospección objetiva basada en datos para tomar decisiones en la búsqueda de clientes potenciales.

Lo expuesto en este apartado, ha motivado a realizar el planteamiento del problema a través de sus objetivos, los cuales se describen a continuación.



## **Objetivos**

### ***Objetivo general***

Desarrollar un modelo de gestión de datos, basado en analítica, que permita una adecuada prospección de clientes potenciales, con el propósito de incrementar la posibilidad de que PROING CIA LTDA disminuya la concentración de clientes.

### ***Objetivos específicos***

OE1: Analizar la situación actual de PROING CIA LTDA, a través de un estudio exploratorio, con el fin de conocer estado actual del manejo de la información.

OE2: Realizar un estudio de factibilidad de una propuesta tecnológica enfocada en la disminución de concentración de clientes en las PYMES, mediante una revisión de literatura preliminar.

OE3: Implementar un modelo de gestión de datos, que permita llevar a cabo una adecuada prospección de clientes, mediante uso de analítica de datos.

OE4: Evaluar el modelo de gestión de datos, a través de un juicio de expertos.

### **Justificación, importancia y alcance del proyecto**

PROING registra el 90% de su facturación concentrada en tres de sus clientes, lo que representa un alto porcentaje de dependencia financiera, que pone en riesgo la operación y continuidad de la empresa, ante lo cual es necesario diversificar su cartera de clientes, para ello es indispensable contar con una base de prospectos potenciales, con alta probabilidad de compra o contratación de los servicios de PROING y que posibilite la disminución de la concentración de clientes.

El presente proyecto tiene como objetivo desarrollar un modelo de datos, mediante uso de analítica, con el propósito de incrementar la posibilidad de que PROING CIA LTDA, disminuya la concentración de clientes.

La realización de este proyecto se justifica plenamente, toda vez que PROING requiere tomar acciones, en un corto plazo, que permitan disminuir la concentración de clientes, caso contrario se agudizarían los efectos financieros negativos que podrían determinar el quiebre financiero de la empresa.

Como parte de este proyecto se analiza la situación actual de PROING CIA LTDA, a través de un estudio exploratorio, con el fin de conocer estado actual del manejo de la información. Se realiza un estudio de factibilidad de una propuesta tecnológica enfocada en la disminución de concentración de clientes en las PYMES, mediante una revisión de literatura preliminar. Se implementa un modelo de gestión de datos, que permita llevar a cabo una adecuada prospección de clientes, mediante uso de analítica de datos. Se evalúa el modelo de gestión de datos, a través de un juicio de expertos.

### ***Preguntas de Investigación***

Para la consecución del objetivo general del proyecto: Desarrollar un modelo de gestión de datos, basado en analítica, para la prospección de clientes, con el propósito de incrementar la posibilidad de que PROING CIA LTDA, disminuya la concentración de clientes, se responden las siguientes preguntas de investigación (RQ) para cada objetivo específico (OE), las cuales delimitan el alcance del proyecto:

**OE1-RQ1:** ¿Cuáles son las causas que han determinado que la facturación de PROING CIA LTDA dependa altamente de un segmento reducido de clientes?

**OE1-RQ2:** ¿Cuáles son los efectos que PROING ha evidenciado como consecuencia de la alta dependencia de un segmento reducido de clientes?

**OE1-RQ3:** ¿De qué fuentes de información dispone PROING para la toma de decisiones?

**OE2-RQ4:** ¿Cuáles son las propuestas tecnológicas que se han desarrollado para disminuir la concentración de clientes en las PYMES?

**OE2-RQ5:** ¿Cuál es la frecuencia con la que se presentan estudios acerca de las propuestas tecnológicas que se han desarrollado para disminuir la concentración de clientes en las PYMES?

**OE3-RQ6:** ¿Cuál es la metodología, la técnica y herramienta, más recomendada, para desarrollar un modelo basado en analítica, que apoye la prospección de clientes?

**OE3-RQ7:** ¿Cómo seleccionar el conjunto de datos para un modelo basado en analítica, que apoye a la prospección de clientes?

**OE4-RQ8:** ¿Los datos obtenidos del modelo desarrollado, permiten una adecuada prospección de clientes potenciales?

**OE4-RQ9:** ¿Cuál es el nivel de aceptación del modelo planteado para el estudio de caso, de acuerdo al juicio de expertos?

### **Hipótesis de investigación**

**H1:** La aplicación de un modelo de gestión de datos, basado en analítica, permitirá una adecuada prospección de clientes potenciales.

## Capítulo II

### Marco Metodológico

#### Estudio del estado del arte

Con el objetivo de analizar el estado del arte se realizó una revisión de literatura preliminar, utilizando la base de datos digital Science Direct, en función de los lineamientos de revisiones de literatura de Barbara Kitchenham (Kitchenham, 2015).

#### *Objetivo*

El objetivo que rigió el estudio del estado del arte corresponde al objetivo específico OE2: Realizar un estudio de factibilidad de una propuesta tecnológica enfocada en la disminución de concentración de clientes en las PYMES, mediante una revisión de literatura preliminar.

#### *Preguntas de Investigación*

El estudio del estado del arte se desarrolló con base en las preguntas de investigación (RQ) correspondientes al OE2.

**OE2-RQ4:** ¿Cuáles son las propuestas tecnológicas que se han desarrollado para disminuir la concentración de clientes en las PYMES?

**OE2-RQ5:** ¿Cuál es la frecuencia con la que se publican estudios acerca de las propuestas tecnológicas que se han desarrollado para disminuir la concentración de clientes en las PYMES?

#### *Aproximación de la revisión de literatura preliminar*

El protocolo de revisión de literatura preliminar se estructura a partir del objetivo OE2 y las respectivas preguntas de investigación citadas en el apartado anterior.

### ***Planeación de la revisión***

En la Figura 5 se muestra la estrategia de búsqueda para planear la revisión de literatura.

#### **Figura 5**

##### *Planeación de la revisión*



*Nota:* Muestra los pasos para la planeación de la revisión de literatura preliminar

### ***Criterios de Inclusión***

- Estudios actuales que se centren en soluciones propuestas para reducir la concentración de clientes en las PYMES.
- Estudios actuales cuyo contenido corresponda a soluciones que hagan uso de tecnología, para disminuir el riesgo de la concentración de clientes en las PYMES.
- Estudios actuales enfocados en soluciones para reducir la concentración de clientes, mediante el uso de los datos para la toma de decisiones.

### ***Criterios de Exclusión***

No serán considerados:

- Estudios cuyas herramientas o técnicas implementadas para mejorar la búsqueda o segmentación de clientes, no correspondan a soluciones enfocadas en reducir la concentración de clientes.
- Estudios anteriores al 2014, que sean de revistas o de congresos de poca relevancia o que estén escritos en un idioma diferente al inglés.

### ***Grupo de Control***

La conformación del grupo de control consiste en realizar una búsqueda inicial de estudios o artículos relacionados con las preguntas de investigación planteadas. Posteriormente se realiza una validación cruzada de estudios, se verifica que los artículos cumplan con los criterios de inclusión y exclusión, se ejecuta una revisión completa de los artículos seleccionados, finalmente se obtienen los documentos académicos, en los que se identifican los términos relevantes que nos permiten construir la cadena de búsqueda. En la Tabla 1 se detallan los artículos técnicos de control seleccionados, que en este caso corresponden a *5 estudios*.

**Tabla 1**

*Estudios por grupo de control*

<b>Artículo</b>	<b>Palabras Clave</b>
<b>EC1.-</b> “Digital platforms for business-to-business markets: A systematic review and future research agenda”	Business, B2B, business, marketing, digital plataform, technological, customer concentration, technology, SMEs
<b>EC2.-</b> “Technological innovations and its influence on the growth of auto component SMEs of Bangalore: A case study approach”	Technological, market, SMEs, customer, innovation, sales, business, technology

Artículo	Palabras Clave
<b>EC3.-</b> “B2B market segmentation: A systematic review and research agenda”	B2B, market segments, segmentation, customer, marketing, potential customer, concentration
<b>EC4.-</b> “Customer concentration and corporate innovation: Evidence from China”	Customer concentration, innovation, business, customers, technological, technology
<b>EC5.-</b> “Balanced scorecard in SMEs: effects on innovation and financial performance”	Small business, customer concentration, innovation, technological

*Nota.* Estudios seleccionados para el grupo de control.

### ***Cadena de búsqueda y resultado del pilotaje***

Para construir la cadena de búsqueda se identificaron los contextos que abordan el objetivo de la investigación, se ubicaron las palabras mostradas en la Tabla 2, que están relacionadas con dichos contextos que más se repiten en los estudios del grupo de control.

**Tabla 2**

*Selección de palabras clave*

Palabras clave	EC1	EC2	EC3	EC4	EC5	No. De repeticiones
<b>Business</b>	x	x	x	X	x	5
<b>B2B</b>	x		x			2
<b>Technological</b>	x	x		X	x	4
<b>Customer</b>	x	x	x	X	x	5
<b>Concentration</b>	x		x	X	x	4
<b>Innovation</b>		x		X	x	3
<b>Technology</b>	x	x		X	x	4
<b>SMEs</b>	x	x				2
<b>Marketing</b>	x		x			2

*Nota.* Palabras seleccionadas en los artículos del grupo de control para construir la cadena de búsqueda.

Con el objetivo de construir la cadena de búsqueda, se identificaron las palabras que más se repiten en cada contexto, usamos el conector OR para las palabras que están dentro del mismo contexto y el conector AND para palabras en contextos diferentes. En la Tabla 3 se muestran las cadenas que se probaron.

**Tabla 3**

*Cadenas de búsqueda probadas en Science Direct*

#	Cadena	Artículos	Grupo de Control
1	("technological" or "technology" or "innovation") and ("customer concentration") and ("business" or "B2B")	2	EC1, EC3
2	("technological") and ("customer concentration") and ("business" or "B2B")	3	EC1, EC3
3	("technology" or "innovation") and ("customer concentration") and ("business" or "B2B")	6	EC1, EC3
4	("technological" or "innovation") and ("customer concentration") and ("business")	45	EC4, EC1, EC3,
5	<b>(("technology" or "technological") and ("customer concentration") and ("business"))</b>	46	EC1, EC2, EC3, EC4

*Nota.* Cadenas construidas con base en las palabras seleccionadas en Tabla 4.

Se seleccionó la cadena No. 5, debido a que nos devuelve un número aceptable de artículos, así como también nos permite encontrar la mayor cantidad de los estudios de control utilizados en la fase inicial.

((**"technology" or "technological"**) and ("customer concentration") and ("business"))



## **Selección de Estudios**

### **Selección de estudios candidatos**

Los estudios candidatos se obtienen al aplicar la cadena de búsqueda en el repositorio académico seleccionado para el estudio, en este caso Science Direct, con la cual se obtienen **46 estudios candidatos**.

### **Selección de estudios relevantes**

Una vez encontrados los estudios candidatos, procedemos a encontrar los estudios relevantes, se realizó la lectura de los títulos y resúmenes de cada artículo candidato, con base en el objetivo, sin perder de vista los criterios de inclusión y exclusión; en este caso, se obtuvieron **21 estudios relevantes** (ver Tabla 4).

**Tabla 4**

*Selección de estudios relevantes*

<b>CantidadTotal</b>	<b>Rechazados</b>	<b>Aceptados</b>	<b>Motivo</b>
<b>46</b>	25	21	Los títulos, palabras clave y abstract no cumplen con los criterios de inclusión y exclusión

*Nota.* Detalle del número de artículos encontrados, rechazados y aceptados

### **Selección de estudios primarios**

En función de los estudios relevantes, se procedió a descargar su texto completo y a leer exhaustivamente cada uno de ellos, sin perder de vista a los criterios de inclusión y exclusión y al objetivo de búsqueda, para luego identificar los estudios primarios, en la Tabla 5 se listan los **10 estudios primarios** identificados para este caso.

**Tabla 5***Estudios primarios*

<b>Artículo</b>	<b>Motivo</b>
<b>EP1.-</b> “Digital platforms for business-to-business markets: A systematic review and future research agenda”	Aplicación de plataformas digitales para empresas B2B.
<b>EP2.-</b> “The relationship between technology, business model, and market in autonomous car and intelligent robot industries”	Solución para la concentración de mercados, optimizando modelos de negocio.
<b>EP3.-</b> “Achieving high efficiency and sustainability through new ventures exploration and exploitation strategies: Insight from well-established and emerging technology standards”	Uso de tecnología para optimizar estrategias financieras.
<b>EP4.-</b> “Technological innovations and its influence on the growth of auto component SMEs of Bangalore: A case study approach”	Innovaciones tecnológicas para el desarrollo de las Pymes.
<b>EP5.-</b> “Constructing marketing decision support systems using data diffusion technology: A case study of gas station diversification”	Sistemas de soporte a la toma de decisiones empresariales.
<b>EP6.-</b> “Reliance on major customers and product market competition”	Solución a la concentración de mercados, mediante innovación de productos.
<b>EP7.-</b> “B2B market segmentation: A systematic review and research agenda”	Diversificación en empresas B2B, mediante segmentación de mercados.
<b>EP8.-</b> “Customer concentration and corporate innovation: Evidence from China”	Concentración de clientes y su incidencia en la innovación.
<b>EP9.-</b> “The impact of the customer relationship management on the organization performance”	Sistemas CRM para la gestión empresarial de clientes
<b>EP10.-</b> “Customers and cash: How relationships affect suppliers' cash holdings”	Impacto de la concentración de clientes en la financiación de empresas

*Nota.* Estudios primarios seleccionados.

### ***Extracción de la información***

Luego de realizar la lectura completa de cada uno de los estudios primarios, se procedió con la extracción de la información, cuya síntesis se muestra a continuación.

En la publicación de (Shree, 2021) se hace una síntesis del uso de plataformas digitales en entornos de negocio Business to Business (B2B), como apoyo a las actividades inherentes a las empresas e industrias, destaca el uso de marcos metodológicos para la adopción de plataformas digitales, tales como: Difusión de la Innovación (DOI) y Tecnología – Organización y Entorno (TOE), como parte de este estudio se concluye que uno de los factores más importantes para la adopción de herramientas tecnológicas dentro de la operación empresarial, es el objetivo que buscamos alcanzar, en cuanto a la gestión de las relaciones con los clientes, como puede ser lealtad, concentración de los clientes, entre otros.

En (Yun, 2016) desarrolla un diagrama de bucle causal, basado en las relaciones dinámicas entre tecnología, modelo de negocio y mercado. Establece que una disminución de concentración de mercados (conjunto de clientes) se puede abordar desde el cambio de modelos de negocio, uso de productos novedosos y tecnologías actuales, para lo cual destaca la participación de los gobiernos y las regulaciones económicas.

En (Liu, 2021) se hace un análisis de las estrategias financieras de explotación y exploración, aplicables al desarrollo empresarial, dichas estrategias dependen del rendimiento financiero de la organización, así como de los recursos tecnológicos que utiliza. Este estudio plantea que la estrategia financiera de explotación tiene un mejor desempeño en cuanto a la reducción de concentración de clientes, para ello sugiere mejorar de forma moderada y sostenible los productos/servicios y uso de tecnologías actuales, en lugar de adoptar un cambio de comportamiento drástico.

En la publicación de (Krishnaswamy, 2014) se hace un estudio a tres PYMES de la industria automotriz de Bangalore, en las cuales se identifica que el crecimiento empresarial y diversificación de mercado, se debe a la introducción de nuevos productos e innovación tecnológica, para responder a las necesidades de sus clientes. El estudio pone de manifiesto que los empresarios toman un papel decisivo a la hora de realizar innovaciones tecnológicas internas, que se traducen en un crecimiento gradual en términos de empleo, inversión y volumen de ventas.

En (Li, 2014) se analiza un caso de estudio de diversificación comercial para una compañía petrolera, en la que solo se cuentan con pequeños conjuntos de datos. Este estudio propone la técnica Mega Trend Diffusion (MTD) para construir un Sistema de Soportes para la toma de decisiones (DSS), con lo cual se busca diversificar las operaciones de la empresa estudio de caso, lo cual a su vez permite diversificar los nichos de mercado. La técnica MTD propuesta hace uso de una red de retro propagación y una red bayesiana, el estudio también sugiere utilizar mecanismos de aprendizaje automático como clasificadores y redes neuronales.

En (Larkin, 2019) habla de la concentración de clientes a causa del bajo nivel de competencia de los productos en el mercado, hace uso del Herfindahl-Hirschman index (HHI), el cual muestra que cuando la competencia de los clientes es baja, los proveedores condensan su red de clientes vendiendo a una red menos dispersa, en este estudio se sugiere promover competitividad de productos en el mercado para reducir la concentración de clientes, mediante el uso de tecnología.

En (Cortez, 2021) se aborda la diversificación de clientes, mediante la segmentación de mercados, para ello define un proceso que consiste en: pre segmentación, segmentación, aplicación y evaluación, también sugiere el tipo de variables a utilizarse para dicha segmentación, por ejemplo, información demográfica, comportamiento de

compra, información financiera de las empresas. Propone cinco modelos estadísticos que permiten la segmentación de mercados: (1) análisis de redes, (2) variables categóricas para análisis matricial, (3) análisis de conglomerados, (4) análisis de clases latentes, (5) procedimientos de optimización por criterios definidos. Para empresas B2B se han enfocado los estudios de segmentación para necesidades de negocio como: satisfacción del cliente, rendimiento de la fuerza de ventas y rendimiento financiero. En este estudio se cita el uso de recursos digitales y tecnológicos para el acercamiento hacia nuevos clientes, mediante uso de aplicaciones web, comercio electrónico y medios sociales.

En (Pan, 2020) se investiga el impacto de la concentración de clientes en la innovación empresarial China, desde un enfoque financiero, se constata que la concentración de clientes está relacionada negativamente con la innovación empresarial en un país de mercado emergente, lo que sugiere que la concentración de clientes aumenta el riesgo operativo de un proveedor, además, se muestra que el efecto de la concentración de clientes en la innovación empresarial es más pronunciado en las empresas con una menor diversificación del negocio y en las empresas con una menor estabilidad de los clientes principales. Este estudio sugiere gestionar información mediante herramientas tecnológicas, a lo largo de la cadena de suministros, lo cual amplía la gama de conocimientos, estimulando la innovación de las empresas.

En (Soltani, 2018) se hace referencia al uso de un Customer Relation Manager (CRM) para gestionar la relación de una empresa y sus clientes actuales o futuros, se determina que el uso de esta herramienta tecnológica impacta positivamente en el crecimiento de ventas. Es importante destacar que el éxito en el uso de esta herramienta depende de la organización y su rendimiento, la gestión de la información del cliente, la tecnología de la información, la capacidad de la organización y la orientación al cliente.

En (Itzkowitz, 2014) se hace un enfoque del riesgo operativo que representa la concentración de clientes, desde un enfoque financiero de las empresas proveedoras, dicho riesgo afecta a la rentabilidad de la empresa, el tamaño, el apalancamiento, la capacidad de endeudamiento, la posición en el mercado de productos, la disponibilidad de coberturas naturales y otras características, como mecanismo para mitigar dicho riesgo, plantea el manejo de flujo de caja efectivo y alternativas financieras para el efecto.

### ***Características del estado del arte***

Luego de realizar la revisión preliminar de literatura, se contestan las preguntas de investigación planteadas, de acuerdo al de detalle que se muestra a continuación:

**OE2-RQ4:** ¿Cuáles son las propuestas tecnológicas que se han desarrollado para disminuir la concentración de clientes en las PYMES?

En la revisión de literatura preliminar, se puede evidenciar el impacto negativo de la concentración de clientes en las PYMES, varios de los estudios plantean soluciones enfocadas en instrumentos financieros (Liu, 2021), (Pan, 2020), (Itzkowitz, 2014), mejoras en la cadena de producción (Krishnaswamy, 2014), (Larkin, 2019) e innovación tecnológica (Shree, 2021), (Yun, 2016), como mecanismos para mitigar el riesgo operativo que representa dicha concentración de clientes.

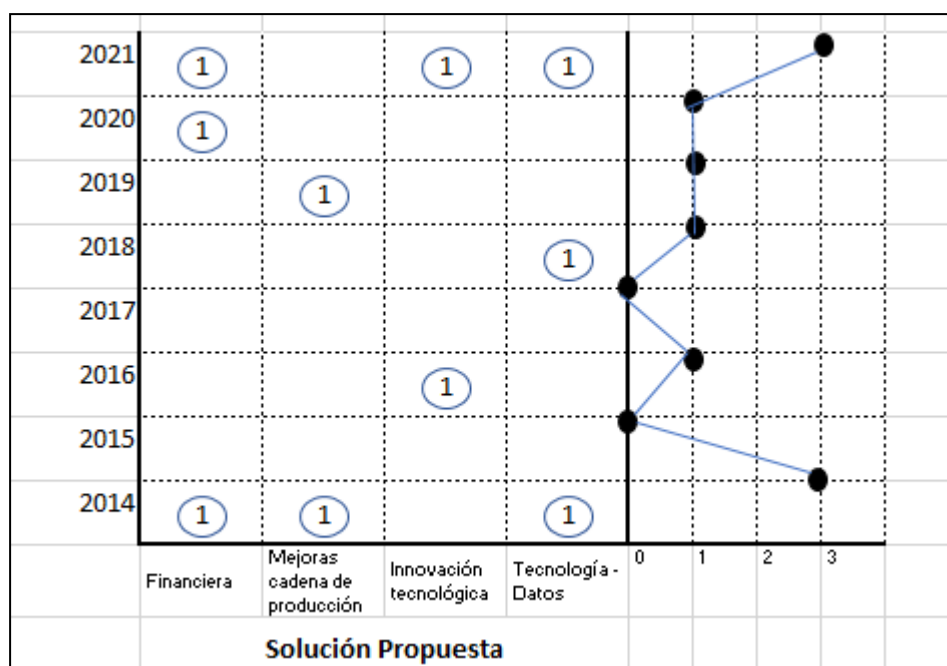
En (Cortez, 2021) se hace referencia al uso de métodos estadísticos para segmentación de mercado y diversificación de clientes. En (Li, 2014) se sugieren herramientas tecnológicas como Sistemas de Soporte a la toma de decisiones (DSS) para diversificar la operación de las empresas y sus nichos de mercado, mientras que en (Soltani, 2018) define el uso de Customer Relation Manager (CRM) para incremento de ventas y gestión de clientes.

**OE2-RQ5:** ¿Cuál es la frecuencia con la que se publican estudios acerca de las propuestas tecnológicas que se han desarrollado para disminuir la concentración de clientes en las PYMES?

En la Figura 6 se muestra la frecuencia con la que se publican estudios acerca de propuestas tecnológicas que se han desarrollado para disminuir la concentración de clientes, como podemos observar en el 2014 y 2021 existe la mayor cantidad de publicaciones realizadas. La gráfica también muestra las publicaciones por año y por tipo de solución propuesta.

**Figura 6**

*Frecuencia de publicaciones por año*



*Nota.* Frecuencia de publicaciones anuales, a partir del 2014. Fuente: Propia

### **Conclusión del estado del arte**

La concentración de clientes es un tema que cuenta con varios estudios referentes al impacto financiero y al riesgo operativo que representa para las PYMES. Se identificaron

estudios que proponen soluciones enfocadas en diversificación de producto, en alternativas de financiamiento e incluso la participación de los gobiernos locales en cuanto a políticas e incentivos para las PYMES.

En cuanto a propuestas de herramientas tecnológicas para disminuir la concentración de clientes, pudimos identificar en uno de los estudios, a los Sistemas de Soporte a la toma de decisiones (DSS) (Li, 2014), el cual hace uso de algoritmos analíticos de datos. Otra herramienta identificada es la de Customer Relation Manager (CRM) (Soltani, 2018), la cual, si bien apoya al proceso de ventas, no mitiga el impacto del riesgo de concentración de cliente, pues únicamente permite realizar un enfoque descriptivo de la situación comercial de la empresa y la gestión de los clientes.

Luego de la revisión de literatura preliminar, se evidencia que no hay muchas soluciones basadas en analítica de datos para abordar la problemática de concentración de clientes, por lo cual se valida plenamente la propuesta de solución basada en el desarrollo de un modelo de datos, haciendo uso de analítica, para la prospección de clientes.

## **Metodología**

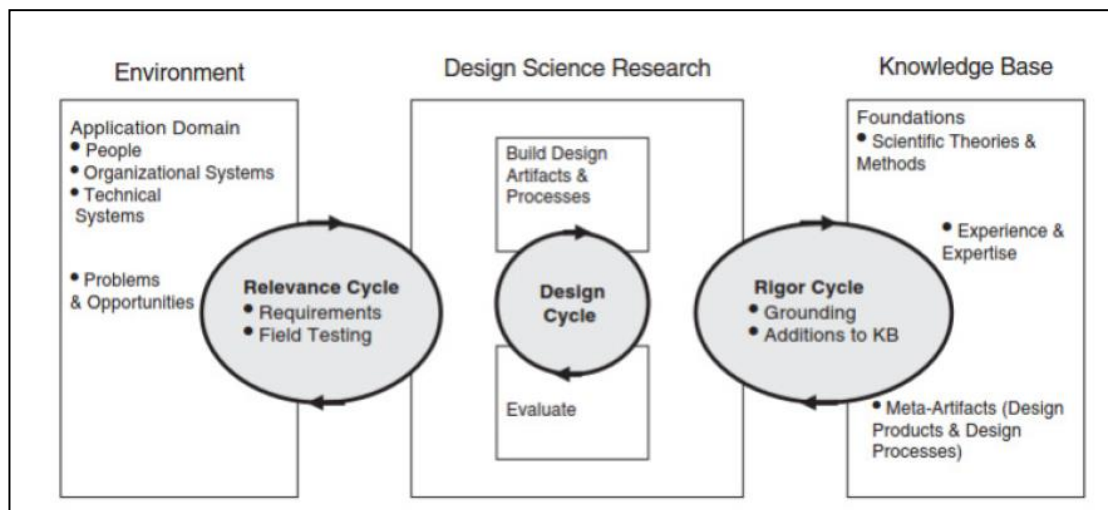
Para el presente proyecto se hizo uso de la metodología Design Science Research, “Design Science Research (DSR) que es un enfoque de investigación riguroso que propone la construcción de artefactos para brindar una solución útil y efectiva a un problema de un dominio dado. El artefacto debe ser una solución innovadora a un problema no trivial. El desarrollo del artefacto implica un ciclo de actividades de diseño, construcción, evaluación, que iteran tantas veces como sean necesarias antes que el artefacto sea finalmente validado y comunicado para su utilización” (Tebes, 2019).



En la Figura 7 se muestran los ciclos de los que se conforma la metodología Design Science Research.

**Figura 7**

*Ciclos de Design Science Research*



*Nota.* Tomado de (Tebes, 2019)

Los ciclos de la metodología de Design Science que se utilizaron para el presente proyecto son: Relevancia, Diseño y Rigor.

- **Ciclo de Relevancia**
  - Identificar el Problema
  - Identificar la Solución
- **Ciclo de Diseño**
  - Diseñar Solución
  - Desarrollar Solución
  - Validar Solución
- **Ciclo de Rigor**
  - Comunicar la investigación

En la Tabla 6 se muestra la relación directa entre las fases de la metodología Design Research, los objetivos planteados y las preguntas de investigación (RQ).

Tabla 6

*Metodología Design Science*

Metodología Design Science	Objetivos Específicos	Preguntas de Investigación
<p><b>Ciclo de Relevancia</b></p> <p><b>-Identificar el Problema</b></p> <p><b>-Identificar la Solución</b></p>	<p><b>OE1:</b> Analizar la situación actual de PROING CIA LTDA, a través de un estudio exploratorio, con el fin de conocer estado actual del manejo de la información.</p> <p><b>OE2:</b> Realizar un estudio de factibilidad de una propuesta tecnológica enfocada en la disminución de concentración de clientes en las PYMES, mediante una revisión de literatura preliminar</p>	<p>OE1-RQ1: ¿Cuáles son las causas que han determinado que la facturación de PROING CIA LTDA dependa altamente de un segmento reducido de clientes?</p> <p>OE1-RQ2: ¿Cuáles son los efectos que PROING ha determinado como ¿Cuáles son los efectos que PROING ha evidenciado como consecuencia de la alta dependencia de un segmento reducido de clientes?</p> <p>OE1-RQ3: ¿De qué fuentes de información dispone PROING para la toma de decisiones?</p> <p>OE2-RQ4: ¿Cuáles son las propuestas tecnológicas que se han desarrollado para disminuir la concentración de clientes en las PYMES?</p> <p>OE2-RQ5: ¿Cuál es la frecuencia con la que se presentan estudios acerca de las propuestas tecnológicas que se han desarrollado para disminuir la concentración de clientes en las PYMES?</p>
<p><b>Ciclo de Diseño</b></p> <p><b>- Diseñar Solución</b></p> <p><b>-Desarrollar Solución</b></p> <p><b>-Validar Solución</b></p>	<p><b>OE3:</b> Implementar un modelo de gestión de datos, que permita llevar a cabo una adecuada prospección de clientes, mediante uso de analítica de datos.</p> <p><b>OE4:</b> Evaluar el modelo de gestión de datos, a través de un juicio de expertos.</p>	<p>OE3-RQ6: ¿Cuál es la metodología, la técnica y herramienta, más recomendada, para desarrollar un modelo basado en analítica, que apoye la prospección de clientes?</p> <p>OE3-RQ7: ¿Cómo seleccionar el conjunto de datos para un modelo basado en analítica, que apoye a la prospección de clientes?</p> <p>OE4-RQ8: ¿Los datos obtenidos del modelo desarrollado, permiten una adecuada prospección de clientes potenciales?</p>

Metodología Design Science	Objetivos Específicos	Preguntas de Investigación
		OE4-RQ9: ¿Cuál es el nivel de aceptación del modelo planteado para el estudio de caso, de acuerdo al juicio de expertos?
<b>Ciclo de Rigor</b>	Este ciclo se sustenta mediante la publicación del presente proyecto de titulación.	

*Nota.* Modelo Design Science relacionado con los objetivos específicos y preguntas de investigación del presente proyecto.

En la Tabla 7 se especifican las técnicas de investigación que se utilizaron en cada ciclo de la metodología Design Science.

**Tabla 7**

*Técnicas de Investigación*

Ciclos de la Metodología Design Science	Objetivos Específicos	Técnica de Investigación
<b>Ciclo de Relevancia</b> <ul style="list-style-type: none"> <li>• Identificar el Problema.</li> <li>• Identificar la Solución</li> </ul>	<b>OE1</b>	Estudio exploratorio
	<b>OE2</b>	Revisión de literatura preliminar
<b>Ciclo de Diseño</b> <ul style="list-style-type: none"> <li>• Diseñar Solución</li> <li>• Desarrollar Solución</li> <li>• Validar Solución</li> </ul>	<b>OE3</b>	CRISP DM
	<b>OE4</b>	Juicio de expertos
<b>Ciclo de Rigor</b>	Este ciclo se sustenta mediante la publicación del presente proyecto de titulación.	

*Nota.* Métodos que se aplicarán en cada ciclo de la metodología Design Science.

## Capítulo III

### Marco Teórico Referencial

La fundamentación teórica se encuentra en concordancia con la hipótesis, para ello se realizó una categorización de variables del problema, con la finalidad de investigar cada categoría y explicar las variables dependientes e independientes del tema de estudio.

**Variable Dependiente:** Prospección de clientes.

**Variable Independiente:** Modelo de gestión de datos.

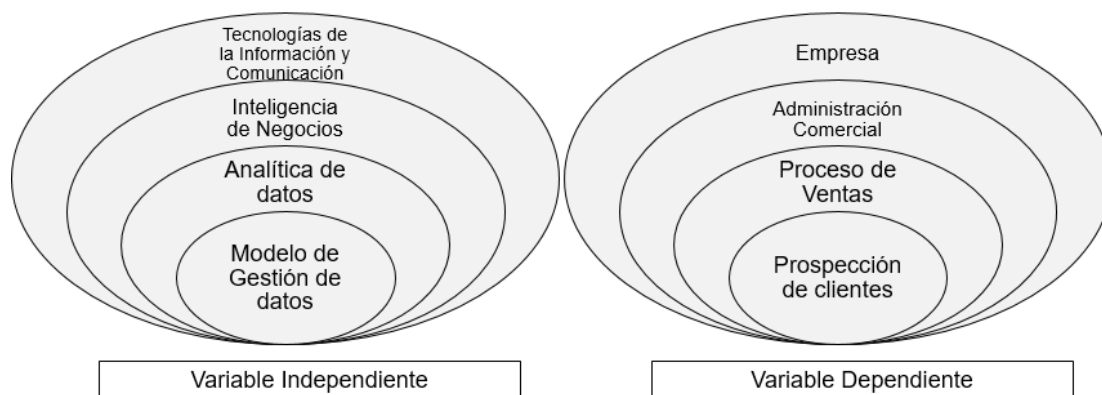
Con base a estas variables y a su categorización se detalla el marco teórico referencial necesario para la realización del presente proyecto.

#### Categorización de las variables

En la Figura 8, se representa la jerarquización de las variables dependiente e independiente.

#### Figura 8

*Categorización de las variables*



## **Fundamentación de la Variable Independiente**

### ***Tecnologías de la Información y Comunicación***

Se definen como la forma de conseguir nuevas realidades comunicativas, mediante el uso de sistemas de informáticos, la microelectrónica y las telecomunicaciones (Cabero, 1998). Otra definición que se puede citar es “Tecnologías para el almacenamiento, recuperación, proceso y comunicación de la información” (Belloch, 2012).

### ***Inteligencia de Negocios***

Es un conjunto de mecanismos para generación y gestión del conocimiento, mediante el almacenamiento, modelado y gestión de datos (Ahumada-Tello, 2011).

Estas estrategias se fundamentan en la gestión del conocimiento y permiten tomar decisiones dentro de una empresa, de modo que pueda marcar una ventaja competitiva, donde el valor agregado a los servicios y/o productos sean consecuencias de estas acciones, que desarrollan eficiencia productiva y eficacia en su funcionamiento, estas ventajas difícilmente podrán ser replicadas por aquellas empresas que no han definido estos procesos o estrategias (Ahumada, 2015).

### ***Arquitectura en Inteligencia de Negocios***

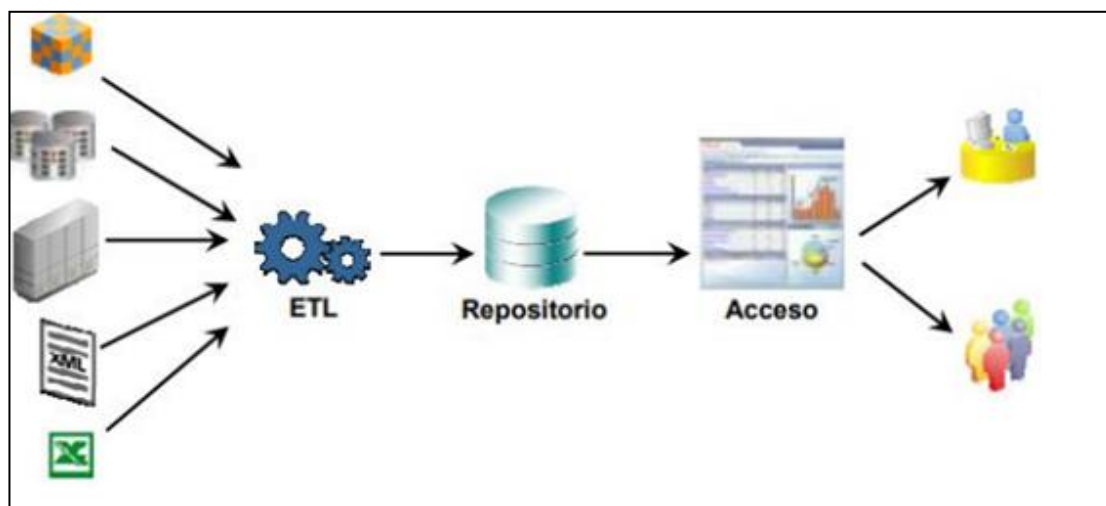
De acuerdo a (Lozada, 2018) la arquitectura de una plataforma de inteligencia de negocios, que se visualiza en la Tabla 8, está conformada por diferentes componentes en cada una de sus fases.

**Tabla 8***Arquitectura de Inteligencia de Negocios*

<b>FASE</b>	<b>COMPONENTES</b>
Origen de los datos	Cubos, bases de datos, ERP, sistemas legados, archivos planos, archivos XML.
Proceso de Extracción, Transformación y Carga (ETL)	Fuentes heterogéneas, de las cuales se escogen los campos que se van a utilizar, se realizan las modificaciones necesarias y se selecciona el sitio de almacenamiento.
Repositorio de datos	Repositorio que almacena los datos transformados, representados visualmente en modelos, tablas y dimensiones
Acceso	La interface de acceso al usuario, permite interactuar con los datos, para visualizar los datos de forma gráfica.

*Nota.* Tomado de (Lozada, 2018)

En la Figura 9 se ilustran los elementos que intervienen en la arquitectura de inteligencia de negocios.

**Figura 9***Arquitectura de Inteligencia de Negocios*

*Nota.* Tomado de (Lozada, 2018)

## ***Herramientas de Inteligencia de Negocios***

Haciendo referencia a (Lozada, 2018) se citan las siguientes herramientas, sobre las cuales se apoya la inteligencia de negocios:

### ***Sistemas de Soporte a la toma de decisiones (SSD)***

Son herramientas informáticas de inteligencia de negocios, que disponen de una serie de informes predefinidos, que permiten generar, estimar, comparar y evaluar diferentes alternativas sobre las cuales se toman decisiones. Los SSD presentan información de manera estadística y estática, ya que no es posible manipular los datos.

### ***Datamart***

El Datamart es una base de datos relacionada con un área específica del negocio, cuenta con una estructura óptima que permite analizar la información desde diferentes perspectivas que inciden en la operación de dicha área de negocio. Un datamart puede ser alimentado desde diferentes fuentes de datos, incluidas las bodegas de datos.

Los datamarts cuentan con estructuras efectivas para el análisis, sus principales ventajas son: escaso volumen de datos, consultas rápidas mediante sentencias SQL sencillas, facilidad en el acceso y validación de los datos.

### ***Datawarehouse***

Una bodega datos se caracteriza por los procesos de limpieza e integración de la información que proviene de una o varias fuentes, el manejo de la información almacenada en el datawarehouse se realizará de acuerdo a las necesidades que existan para apoyar a la toma de decisiones oportunas y eficientes.

Los componentes de un almacén de datos son:

- Fuentes de datos de producción
- Extracción de datos y conversión

- El gestor de la base de datos del datawarehouse
- La administración del almacenamiento de datos
- Herramientas de inteligencia de negocios

### ***Metodologías para Inteligencia de Negocios***

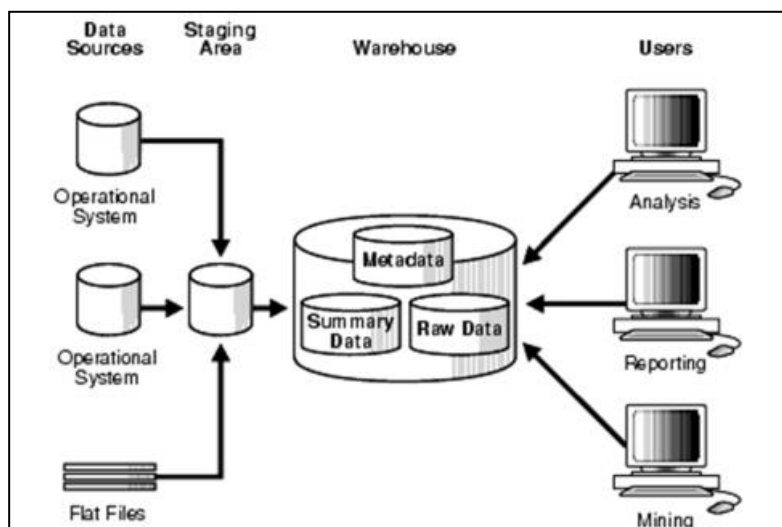
Se destacan dos metodologías para el desarrollo de un Datwarehouse, como parte de la Inteligencia de Negocios: Kimball e Inmon (Rivadeneira, 2014).

#### ***Metodología de Kimball***

Kimball es una metodología diseñada para acceso y explotación de datos de forma ágil y sencilla, de modo que no se precisa de usuarios especializados para el efecto. El diseño y construcción del datawarehouse parte de un Datamart para un área específica, haciendo una construcción Bottom- up. Trabaja con modelado de bases de datos dimensional, no normalizado (Bustos, 2013). En la Figura 10 se muestra la arquitectura Kimball.

**Figura 10**

*Arquitectura Kimball*



*Nota.* Tomado de (Bustos, 2013)



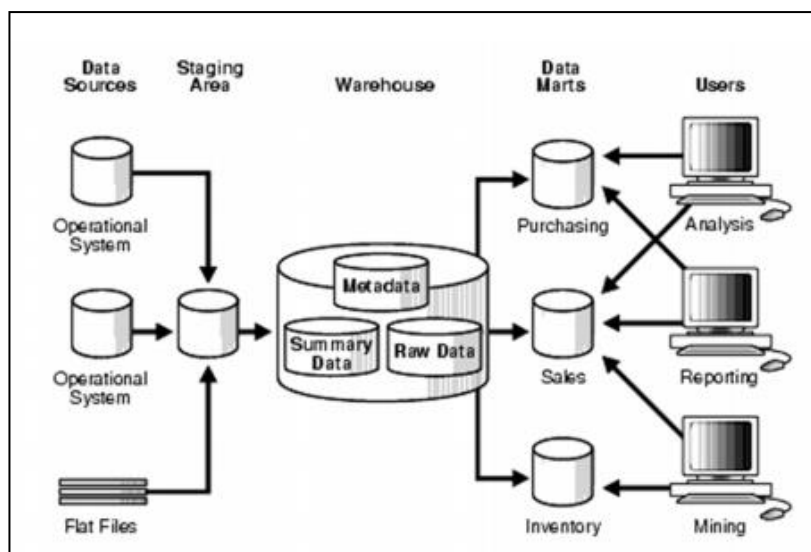
### **Metodología de Inmon**

Inmon corresponde a una metodología para diseñar una bodega de datos, que se alimenta de la información de diferentes áreas de una organización, se orienta hacia el almacenamiento de grandes volúmenes de datos, pretende evitar la redundancia, por lo cual sus datos se presentan en tercera forma normal, simplificando la consulta y mantenimiento de la información (Rivadeneira, 2014).

La arquitectura de la metodología Inmon se muestra en la Figura 11. El diseño del datawarehouse incluye a toda la organización, haciendo una construcción Top-Down, comenzando por un datawarehouse completo, esta metodología se fundamenta en bases de datos relacionales (Bustos, 2013).

**Figura 11**

#### *Arquitectura Inmon*



*Nota.* Tomado de (Bustos, 2013)

### ***Analítica de Datos***

Es una ciencia que permite analizar datos, descubrir patrones y comportamientos, que generan conocimiento para apoyar la toma de decisiones, optimizando procesos en el entorno en donde se producen dichos datos (Aguilar, 2018).

La analítica datos utiliza técnicas, herramientas y habilidades relevantes para obtener valiosa información sobre las futuras prácticas, convirtiéndose en un recurso empresarial exhaustivo y detallado, que identifica el tipo de datos a analizar, los recopila y los organiza en los conjuntos apropiados mediante técnicas estadísticas y algoritmos. También abarca la limpieza de datos para entregar el análisis exacto según el cual se pueden tomar las decisiones apropiadas. Cuando estos datos se encuentran a una forma utilizable, se aplica un proceso algorítmico o mecánico mediante procesos estadísticos o algoritmos de aprendizaje automático, comparando los distintos conjuntos de datos para obtener información clave que dé respuestas a los problemas que se buscan resolver. Luego de este proceso, se representan los datos en una forma que puede entenderse adecuadamente para el beneficio del negocio. La analítica de datos también implica la creación de modelos cuantitativos, que son construidos a partir de diferentes variables, para identificar oportunidades de negocio, apoyar a la toma de decisiones y mejorar la competencia de las empresas en un mercado saturado (Jones, 2019).

### ***Técnicas de Minería de datos***

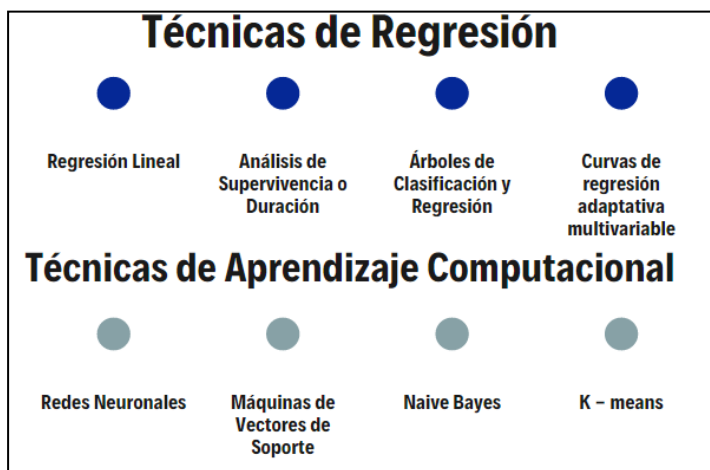
La minería de datos permite la construcción de modelos analíticos que ajusta y relaciona las variables involucradas para proporcionar conocimiento, por tanto, es importante la elección del modelo y el ajuste del mismo a los datos de interés (Riquelme J. , 2006). Entre las principales técnicas para minería de datos tenemos a los modelos predictivos y modelos descriptivos.

### **Modelos predictivos**

Son mecanismos que predicen el comportamiento de un individuo, partiendo de los datos de entrada se obtiene una predicción como salida, una calificación alta es indicativo de una alta probabilidad de que el individuo presente el comportamiento predicho. Las principales técnicas para realizar análisis predictivo se segmentan en dos grandes grupos: Técnicas de regresión y técnicas de aprendizaje computacional (Espino, 2017), en la Figura 12 se muestra dicha clasificación.

**Figura 12**

*Técnicas para Modelos Predictivos*



*Nota.* Tomado de (Espino, 2017)

### **Regresión Lineal**

La regresión lineal corresponde a una técnica estadística, que hace uso de datos aleatorios para determinar relaciones entre variables y realizar predicciones. La precisión y la exactitud del modelo, son directamente proporcionales a la cantidad de datos de la muestra (Fassler, 2018).

### ***Análisis de Supervivencia***

Un análisis de supervivencia, hace referencia al análisis del tiempo de seguimiento (T) de una observación hasta que ocurre un evento predefinido, los métodos más utilizados son: Kaplan Meier y el análisis de riesgos proporcionales de Cox (Salinas, 2008).

### ***Árboles de clasificación y regresión***

Son estructuras de datos en forma de árbol, que generan reglas para la clasificación, explicando el comportamiento de una variable en relación a otras, fácilmente traducibles a reglas de negocio. Estas técnicas son utilizadas para análisis predictivo y de clasificación (Fassler, 2018).

### ***Redes Neuronales***

Las redes neuronales artificiales (NNA) es uno de los métodos más utilizados en machine learning y puede ser implementado de varias maneras, empezando desde un esquema básico, con una red neuronal de una sola capa, es decir, cuenta con una sola capa de neuronas intermedias, que se denomina capa oculta, matemáticamente se define la función de salida como  $y(x)$ , y  $x$  son los valores de entrada.

$$y(x) = g\left(\sum_{i=0}^n w_i x_i - \theta_i\right)$$

Donde los parámetros  $w$  son los pesos, que afectan a las entradas, y  $\theta$  es el umbral o sesgo de la entrada  $i$  (Torneró, 2017).

### ***Máquina de vectores de Soporte***

Es una técnica de aprendizaje que hace uso de un espacio de hipótesis lineales, que permite entrenar máquinas de aprendizaje lineal, mediante clasificación y regresión, por

ejemplo, clasificación de imágenes, clasificación de patrones, identificación de funciones, entre otras (Resendiz, 2006).

### ***Naive Bayes***

Naive Bayes es un clasificador de probabilidades, que proporciona una mejor precisión para datos en tiempo real, opera a partir de una pequeña cantidad de datos de entrenamiento, de los cuales aprende y predice la clase de la instancia dada (Mosquera, 2018).

### ***K – means***

K – means es un algoritmo de clustering, el cual toma un valor k que corresponde al número de clústers que se formarán, el algoritmo selecciona un número k de elementos en forma aleatoria, los cuales representan el centroide o media de cada clúster, cada uno de los objetos es asignado a un clúster, luego realiza varias iteraciones hasta que el valor de la media no cambia. El algoritmo utiliza una medida de similitud basada en el error cuadrático, el cual representa la menor distancia entre los elementos y el centro del clúster, el objetivo de K – means es el de minimizar dicho error cuadrático. La medida de similitud corresponde a la distancia entre el objeto y la media de un clúster.

### ***Metodologías empleadas en minería de datos***

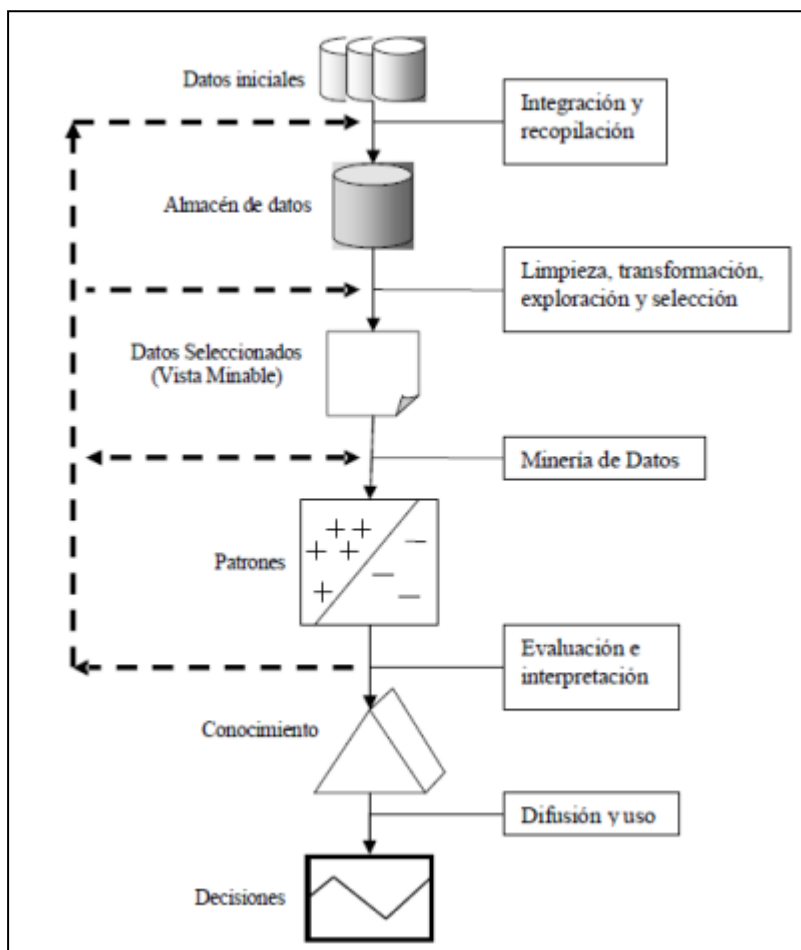
Las principales metodologías, para el desarrollo de proyectos de minería de datos son: KDD, SEMMA, CRIPS – DM y Catalyst (Moine J. , 2011).

### ***KDD***

KDD (Knowledge Discovery in Database) define las principales etapas para proyectos que buscan explotar la información (Moine J. M., 2011), en la Figura 13 se muestran las fases de esta metodología.

Figura 13

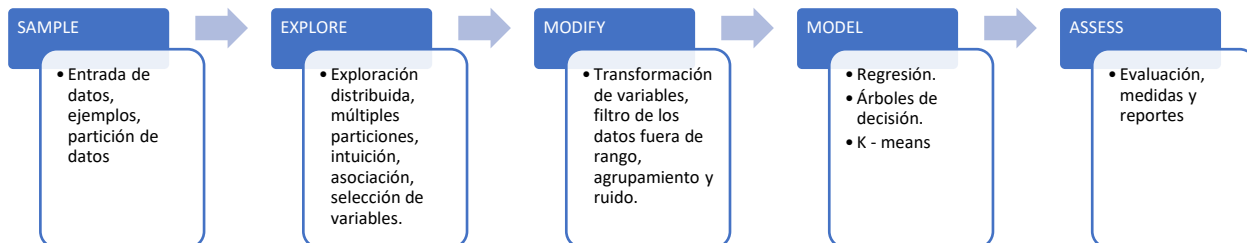
Fases del proceso de KDD



Nota. Tomado de (Eckert, 2013)

### SEMMA

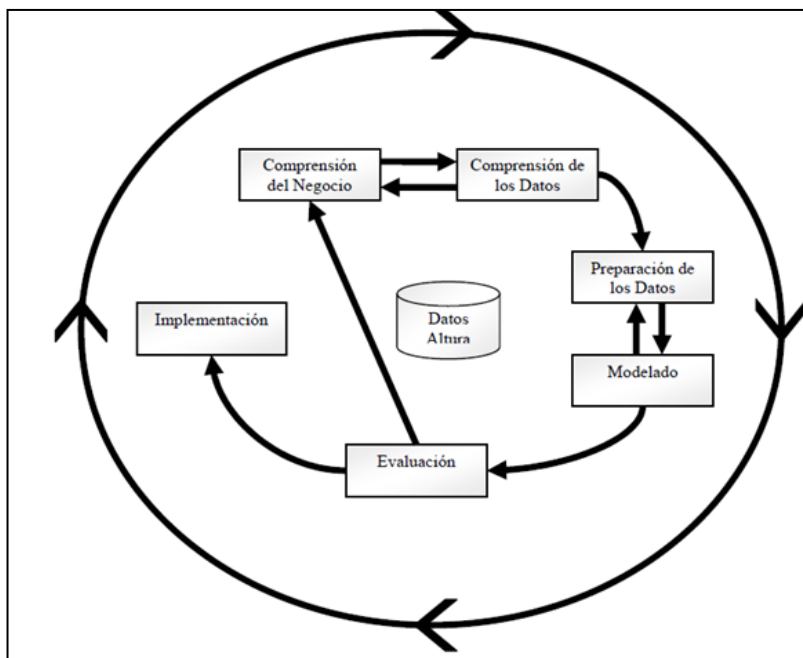
SEMMA (Sample, Explore, Modify, Model and Acces) es la metodología propuesta por SAS Institute Inc, la cual permite comprender, organizar, desarrollar y mantener proyectos de minería de datos, ayudando a solucionar los problemas y objetivos empresariales, consta de cinco fases mostradas en la Figura 14.

**Figura 14***Fases de la metodología SEMMA*

*Nota.* Tomado de (Gujrat, 2014)

**CRISP - DM**

CRISP - DM (Cross – Industry Standard Process for Data Mining) proporciona lineamientos para los proyectos de minería de datos, en la figura 15 se muestran las fases de esta metodología (Gujrat, 2014).

**Figura 15***Fases de la metodología CRISP DM*

*Nota.* Tomado de (Gujrat, 2014)

## CATALYST

Catalyst es una metodología conocida como P3TQ (Product, Place, Price, Time, Quantity) que plantea la formulación de un modelo de Negocio y un modelo de explotación de la información, en los cuales se establecen acciones, se evalúan los resultados y se definen los siguientes pasos, lo cual proporciona flexibilidad (Moine J. , 2011).

En la Tabla 9 se muestra una comparación de las diferentes metodologías citadas.

**Tabla 9**

*Comparación metodologías para minería de datos*

<b>FASES</b>	<b>KDD</b>	<b>CRISP – DM</b>	<b>SEMMA</b>	<b>CATALYST</b>
<b>Análisis y comprensión del negocio</b>	Comprensión del dominio de aplicación	Comprensión del negocio		Modelado del negocio
<b>Selección y preparación de los datos</b>	<ul style="list-style-type: none"> <li>• Crear el conjunto de datos</li> <li>• Limpieza y pre procesamiento de los datos</li> <li>• Reducción y proyección de los datos</li> </ul>	<ul style="list-style-type: none"> <li>• Entendimiento de los datos</li> <li>• Preparación de los datos</li> </ul>	<ul style="list-style-type: none"> <li>• Muestreo</li> <li>• Comprensión</li> <li>• Modificación</li> </ul>	Preparación de los datos
<b>Modelado</b>	<ul style="list-style-type: none"> <li>• Determinar la tarea de minería</li> <li>• Determinar el algoritmo de minería</li> <li>• Minería de datos</li> </ul>	Modelado	Modelado	Selección de herramientas y modelado inicial
<b>Evaluación</b>	Interpretación	Evaluación	Valoración	Refinamiento del modelo
<b>Implementación</b>	Utilización del nuevo conocimiento.	Despliegue		Comunicación

*Nota.* Tomado de (Moine J. , 2011)



### ***Modelo de gestión de datos***

“La gestión de datos se refiere a la práctica profesional de construir y mantener un framework para la ingesta, almacenamiento, extracción y archivado de los datos esenciales para la empresa moderna. La gestión de datos es la columna vertebral que conecta todos los segmentos del ciclo vital de la información” (Pearlman, 2019).

Actualmente la gestión de la información conlleva al uso de modelos y técnicas de minería de datos para obtener conocimiento, utilizando técnicas de reconocimiento de patrones, redes neuronales, lógica difusa, algoritmos genéticos, entre otras técnicas avanzadas para el análisis de datos (González, 2008).

### **Fundamentación de la variable independiente**

#### ***Empresa***

Es una organización en la cual se producen y se comercializan bienes y servicios necesarios para satisfacer necesidades, cuyo principal objetivo es la obtención de rentabilidad económica. La estructura de una empresa está compuesta por departamentos, los cuales tienen funciones específicas e interrelacionadas entre sí, los departamentos que generalmente se encuentran en una empresa son los departamentos de administración, financiero, comercial y de personal (Núñez, 2019).

#### ***Administración Comercial***

Dentro de la estructura de la administración de una empresa, encontramos la Administración Comercial, la cual gestiona los recursos disponibles para lograr un objetivo de ventas. La administración comercial utiliza herramientas y técnicas para aprovechar los recursos (humanos, financieros, materiales y tecnológicos), que permita la

organización, planificación, dirección y control de la estabilidad y el crecimiento de las organizaciones (Riquelme M. , 2018).

### **Proceso de Ventas**

Las ventas son un proceso sistemático que parte de la ubicación o identificación del cliente a ser contactado (Navarro, 2012). Dentro de las fases del proceso de ventas encontramos: prospección, presentación, estudio de necesidades, análisis de soluciones, propuesta, acuerdo y cierre, tal como se muestra en la Figura 16.

**Figura 16**

*Proceso de Ventas*



*Nota:* Fases del proceso de ventas. Tomado de (Navarro, 2012).

### **Prospección de clientes**

La prospección de clientes corresponde a la primera fase de las ventas consultivas, que consiste en la búsqueda sistémica y organizada de nuevos clientes potenciales, a los cuales se denominan prospectos. El objetivo de la prospección es encontrar un conjunto de prospectos, a los cuales se deben contactar de forma sistemática para convertirlos en clientes (Olivier, 2020). Para localizar un prospecto se recurre a fuentes de información externas como bases de datos, directorios de empresas, revistas especializadas, encuestas, sondeos. Luego de ubicar los prospectos se debe realizar una segmentación de los mismos para realizar una valoración de la necesidad de contratar el servicio, poder de compra, lealtad. Por último, dentro de esta fase, se da prioridad a aquellos que se puedan atraer a corto plazo (Navarro, 2012).

## Capítulo IV

### Diseño y Construcción de la Solución

Para dar solución a la problemática identificada en Proing Cia Ltda. se plantea un modelo basado en analítica de datos que aportará a la disminución de la concentración de clientes, la prospección de los mismos y a mediano plazo la consecuente diversificación de su cartera.

Para el desarrollo del presente modelo, se hizo uso de la técnica analítica de clustering K-means, la cual corresponde a una técnica predictiva de machine learning, no supervisada, que permite clasificar y extraer perfiles de comportamiento o clases, lo cual solventa la necesidad del presente proyecto, para identificar y clasificar a un conjunto de cuentas, como potenciales clientes de negocio.

Para el desarrollo del modelo analítico, se usó R y RStudio de acuerdo a las recomendaciones encontradas en (Bohanec, 2016).

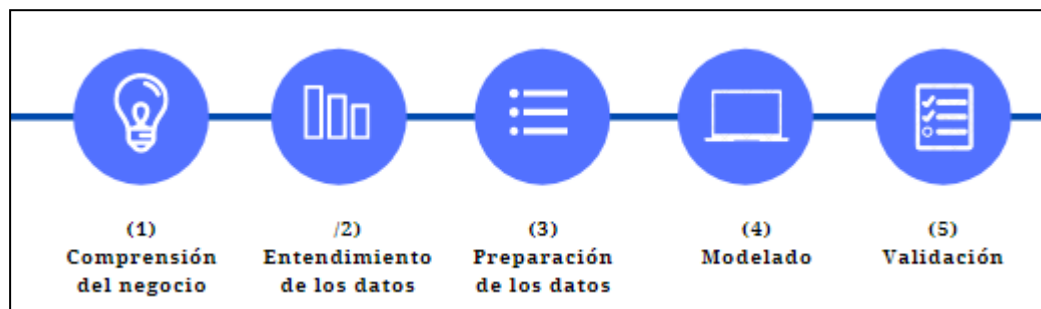
De acuerdo con (Meire, 2017), para el desarrollo de modelos comerciales de prospección, basados en analítica de datos, se recomienda el uso de fuentes externas. Para este caso se hizo uso de fuentes gubernamentales, que cuentan con información financiera de las diferentes empresas, dentro de las cuales se analizará el perfil financiero para identificar clientes potenciales, que puedan ser prospectados por Proing Cia Ltda.

Por otro lado, para la selección de la metodología de analítica de datos (Moine J. M., 2011) recomienda identificar el punto de partida del proyecto a desarrollarse, dependiendo del escenario es posible hacer uso de CRISP DM, SEMMA o Catalyst. Para la presente investigación, el punto de partida es un *Escenario donde se aborda una situación organizacional utilizando minería de datos, para buscar patrones y relaciones que puedan aportar a dicha situación.* Para este tipo de escenarios, (Moine J. , 2011)

recomienda el uso CRISP DM, la cual fue seleccionada. Las fases de la metodología CRISP DM utilizadas en el presente proyecto, se muestran en la Figura 17.

**Figura 17**

*Fases CRISP DM aplicadas en el proyecto*



La última fase de validación, se desarrolla por completo en el Capítulo V.

### **Comprensión del Negocio**

El caso de estudio corresponde a una PYME que maneja un modelo de negocio Business to Business (B2B)<sup>2</sup>; es decir, se caracteriza por contar con ciclos de venta de mediano y largo plazo y montos de venta altos en un número de clientes reducido; por tanto, la prospección representa un proceso de gran valor, de modo que se puedan ubicar clientes potenciales, desarrollar ventas consultivas y cierre de proyectos; lo que en definitiva significaría el aumento de la cartera de clientes (Flecha, 2020).

Una vez analizada la problemática, podemos concluir que, debido a la naturaleza del negocio B2B, actualmente PROING CIA LTA, cuenta con una facturación reducida en tres principales clientes, con lo cual no es posible hacer uso de estos datos para definir un modelo analítico de prospección, es necesario recurrir a fuentes externas y a la

<sup>2</sup> Business to Business (B2B) es un modelo de negocio donde una empresa vende a otras empresas.

experiencia de perfiles comerciales, con cuya información podamos definir un modelo analítico de datos formal y válido, aplicable a la prospección.

### **Entendimiento de los datos**

En empresas B2B, proveedoras de soluciones tecnológicas existen parámetros que orientan la prospección. De acuerdo a (Meire, 2017) algunas variables que se consideran para la prospección son:

- Ubicación geográfica
- Vertical de negocio
- Ingresos Anuales
- Utilidad neta
- Variables que afectan a la vertical de negocio

Normalmente la prospección de clientes se hace con base en la intuición, algunas de las variables citadas se usan de forma aislada, en el presente proyecto se desarrolla un modelo de datos K-means que interrelaciona estas variables, para identificar clientes potenciales.

### ***Ubicación Geográfica***

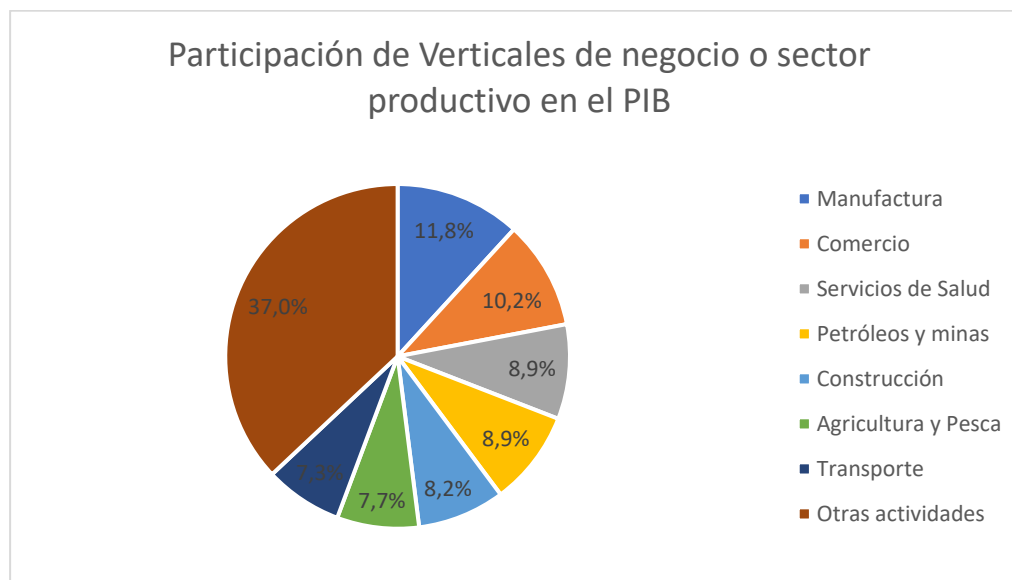
Corresponde a la ubicación de la empresa, dentro de una región del país, para el presente proyecto se consideraron tres ubicaciones geográficas, Pichincha, Guayas y resto del país (Tabla 10), ya que bajo esta segmentación se definen valores relevantes para el desarrollo del modelo.

**Tabla 10***Variables para Ubicación Geográfica*

<b>Ubicación Geográfica</b>
Pichincha
Guayas
Resto del país

*Nota.* Tomado de (EKOS, 2020)**Vertical de Negocio**

La vertical de negocio, corresponde a la actividad comercial o de producción, a la que se dedica la empresa, para este caso, hicimos uso de las verticales de negocio que contribuyen con mayor porcentaje al PIB, de acuerdo al análisis de EKOS 2020, donde se analizan a las siete principales actividades que sustentan la economía del país, de acuerdo a lo indicado en la Figura 18.

**Figura 18***Verticales de negocio**Nota.* Participación en el PIB de las verticales de negocio, tomado de (EKOS, 2020)

Como se puede observar en el Figura 18, los sectores productivos con mayor participación en la economía del Ecuador son manufactura, comercio, servicios de salud, petróleo y minas, construcción, agricultura y pesca y transporte, por tanto, son los sectores productivos seleccionados para el desarrollo del presente modelo.

### ***Ingresos Anuales***

Los ingresos anuales corresponden al volumen de facturación registrado en las diferentes empresas que se analizaron como potenciales clientes, en este caso tomamos los datos del ranking de EKOS 2020, mismo que hace uso de la información tomada de la Superintendencia de Compañías Ecuador.

### ***Utilidad neta***

Corresponde a la utilidad bruta menos los impuestos causados por la empresa en el año de análisis, este dato es importante, ya que la probabilidad de inversión en tecnología, está directamente relacionada con la utilidad neta de la empresa.

### ***Variables que afectan a la vertical de negocio***

Se toma como dato relevante la recesión económica por ubicación geográfica, así como por vertical de negocio, de acuerdo a lo indicado en las tablas 11 y 12.

**Tabla 11**

#### *Recesión económica por ubicación geográfica*

<b>Ubicación Geográfica</b>	<b>Caída de ventas</b>
Pichincha	26%
Guayas	14%
Resto del país	19%

*Nota.* Tomado de (EKOS, 2020)

**Tabla 12***Recesión económica por vertical de negocio*

<b>Vertical de negocio</b>	<b>Caída de ventas</b>
Manufactura	10%
Comercio	20%
Enseñanza y Servicios de Salud	10%
Petróleo y minas	10%
Construcción	20%
Agricultura y Pesca	10%
Transporte	10%

*Nota.* Tomado de (EKOS, 2020)

Los datos correspondientes a *Ingresos Anuales, Utilidad Neta y Variables que afectan a la vertical de Negocio*, están contenidos en el set de datos utilizados para el desarrollo del modelo analítico, cuyo detalle se especifica en la sección Modelado.

### **Preparación de los datos**

Para preparar los datos de entrada del modelo K-means procedemos con: (1) Selección de los datos, (2) Construcción del set de datos.

### **Selección de los datos**

Los datos con los cuales se construyó el modelo analítico para la prospección de clientes de Proing CIA. LTDA., corresponde a información financiera del año 2020 de (EKOS, 2020), la cual a su vez consolida la información de la Superintendencia de Compañías, se tomó esta información, ya que es la que se tiene disponible a la fecha del desarrollo del presente proyecto, y son las fuentes a las que se remiten las áreas comerciales de las empresas.



Se eligieron las verticales de negocio, mostradas en la Tabla 13, las cuales aportan mayormente al PIB, pues de acuerdo a (EKOS, 2020) en estas verticales se espera un mayor dinamismo en la economía para nuevas compras y contrataciones.

En la Tabla 13 se muestra que, para el caso de la vertical de *Manufactura y Comercio* por su diversidad, se subdivide en *Sectores*.

**Tabla 13**

*Sectores y subsectores para el modelo analítico*

<b>Vertical de negocio</b>	<b>Sector</b>
<b>Manufactura</b>	Agroindustria
	Industria de alimentos y bebidas
	Farmacéutica
	Insumos y materiales de construcción
<b>Comercio</b>	Retail
	Comercio Automotriz
	Comercio alimentos y bebidas
	Comercio productos farmacéuticos
	Venta de electrodomésticos
	Venta de combustibles
<b>Enseñanza y Servicios de Salud</b>	
<b>Petróleo y minas</b>	
<b>Construcción</b>	
<b>Agricultura y Pesca</b>	
<b>Transporte</b>	

*Nota.* Tomado de (EKOS, 2020)

En la Tabla 14 se detallan las variables que se utilizaron en el desarrollo del presente modelo.

**Tabla 14**

*Variables para desarrollo del modelo de clusterización*

<b>Variables</b>	<b>Nomenclatura</b>
<b>Vertical de negocio</b>	VERTICAL DE NEGOCIO
<b>Sector</b>	SECTOR
<b>Subsector</b>	SUBSECTOR
<b>Empresa</b>	EMPRESA
<b>Ubicación geográfica</b>	UbicGeografica
<b>Ingresos anuales</b>	IngresosAnuales
<b>Utilidad bruta</b>	UtilidadBruta
<b>Impuestos causados</b>	ImpCausados
<b>Utilidad neta</b>	UtilidadNeta
<b>Caída en ventas por ubicación geográfica</b>	CvUbicacGeogr
<b>Caída en ventas por sector</b>	CvPorSector

*Nota.* Variables para desarrollo de K-means

### **Construcción del set de datos**

Los datos se encuentran consolidados en un archivo de texto, para el uso de los mismos, se consolidó la información en un set de datos en Excel, en la Tabla 15 se detallan los campos utilizados.

**Tabla 15**

*Formato de los datos a modelarse*

<b>Campos</b>	<b>Nomenclatura</b>	<b>Tipo de dato</b>	<b>Detalle</b>
Vertical de negocio	VERTICAL	“Cadena de caracteres”	Nombre de la vertical a la que pertenece la empresa
Sector	SECTOR	“Cadena de caracteres”	Sector dentro de la vertical
Subsector	SUBSECTOR	“Cadena de caracteres”	Subsector dentro de la categoría sector
Empresa	EMPRESA	“Cadena de caracteres”	Nombre de la empresa, fuente Ekos 2020
Ubicación geográfica	UBICACIÓN GEOGRÁFICA	“Cadena de caracteres”	Ubicación de la empresa, de acuerdo a Tabla 13
Ingresos anuales	IngresosAnuales	“Numérico”	Valor en millones de dólares, fuente Ekos 2020
Utilidad bruta	UtilidadBruta	“Numérico”	Valor en millones de dólares, fuente Ekos 2020

<b>Campos</b>	<b>Nomenclatura</b>	<b>Tipo de dato</b>	<b>Detalle</b>
Impuestos causados	ImpCausados	“Numérico”	Valor en millones de dólares, fuente Ekos 2020
Utilidad neta	UtilidadNeta	“Numérico”	Valor en millones de dólares, campo calculado
Caída en ventas por ubicación geográfica	CvUbicacGeogr	“Numérico”	Valor en porcentaje, fuente Ekos 2020
Caída en vetas por sector	CvPorSector	“Numérico”	Valor en porcentaje, fuente Ekos 2020

*Nota.* Detalle de los datos utilizados en el modelo.

### **Modelado**

El modelo se desarrolló mediante la técnica de clusterización K-means.

### **Carga de set de datos**

Se cargó el set de datos, el cual está conformado por 498 observaciones y 11 variables, de las cuales 5 son variables nominales y 6 son variables numéricas, de acuerdo al detalle mostrado en la Tabla 16.

**Tabla 16**

*Variables del set de datos*

<b>Variable</b>	<b>Nomenclatura</b>	<b>Tipo de Variable</b>
Vertical de negocio	VERTICAL	Nominal
Sector	SECTOR	Nominal
Subsector	SUBSECTOR	Nominal
Empresa	EMPRESA	Nominal
Ubicación geográfica	UBICACIÓN GEOGRÁFICA	Nominal
Ingresos anuales	IngresosAnuales	Numérica
Utilidad bruta	UtilidadBruta	Numérica
Impuestos causados	ImpCausados	Numérica
Utilidad neta	UtilidadNeta	Numérica
Caída en ventas por ubicación geográfica	CvUbicacGeogr	Numérica
Caída en vetas por sector	CvPorSector	Numérica

*Nota.* Detalle de las variables definidas para el modelo de clusterización

En la Figura 19 se muestra la vista en RStudio, luego de importar el set de datos inicial.

**Figura 19**

*Set de datos para el modelo K-means*

	SECTOR...1	SECTOR...2	SUBSECTOR	EMPRESA	UBICACIÓN GEOGRÁFICA	INGRESOS ANUALES	UTILIDAD BRUTA	IMPUESTOS CAUSADOS	UTILI NETA
1	COMERCIO	RETAIL	Pescado, crustáceo...	WINREP SA	GUAYAS	57.60	1.42	0.23	
2	COMERCIO	COMERCIO ALIM...	Industria y comeri...	VINLITORAL SA	GUAYAS	11.18	0.60	0.15	
3	COMERCIO	COMERCIO ALIM...	Industria y comeri...	VINESA SA	PICHINCHA	23.49	0.96	0.27	
4	COMERCIO	Salud	Centros médicos y...	VERIS	GUAYAS	50.22	2.85	0.61	
5	MANUFACTURA	Agroindustria	Conservas Fabricac...	VECONSA	PICHINCHA	54.87	2.29	0.23	
6	COMERCIO	COMERCIO DE VEH...	Comercio de vehíc...	VALLEJO ARAUJO	PICHINCHA	162.70	0.78	0.33	
7	TRANSPORTE	TRASPORTE Y LOG	Transporte y Logíst...	UNITED PARCEL SERV...	PICHINCHA	20.96	0.47	0.10	
8	TRANSPORTE	TRASPORTE Y LOG	Transporte y Logíst...	UNITED AIRLAINS	PICHINCHA	18.29	1.71	0.09	
9	COMERCIO	Salud	Centros médicos y...	UNIOFKEN SA	GUAYAS	6.33	0.66	0.29	

*Nota.* Vista data set cargado, desde RStudio. Fuente: Propia.

### ***Librerías para correr K-means***

Para ejecutar K-means, son necesarias las librerías: tidyverse, cluster, factoextra, NbClust y tidyr, en la Figura 20 se muestran las librerías instaladas en RStudio.

**Figura 20**

*Librerías instaladas*

```
> ipak <- function(pkg){
+   new.pkg <- pkg[!(pkg %in% installed.packages()[, "Package"])]
+   if (length(new.pkg))
+     install.packages(new.pkg, dependencies = TRUE)
+   sapply(pkg, require, character.only = TRUE)
+ }
>
> packages <- c("tidyverse","cluster", "factoextra","NbClust","tidyr")
> ipak(packages)
tidyverse   cluster factoextra   NbClust   tidyr
      TRUE      TRUE      TRUE      TRUE      TRUE
```

*Nota.* Instalación de las librerías necesarias para correr K – means en RStudio

### **Cambio de variable para la identificación de cada fila**

Se hizo un cambio de identificación al nombre de cada fila, es decir, en lugar que la identificación sea una numeración consecutiva, se colocan los nombres de los posibles clientes – empresas (ver Figura 21); esto permite la trazabilidad de los nombres, ya que al normalizar el set de datos, se deben eliminar las columnas con variables nominales, toda vez que la normalización acepta únicamente variables numéricas, este cambio lo realizamos con la función `rownames(df)<-c()`

**Figura 21**

*Filas identificadas con el nombre de la empresa*



	SECTOR...1	SECTOR...2	SUBSECTOR	EMPRESA	UBICACIÓN GEOGRÁFICA	INGRESOS ANUALES	UTILIDAD BRUTA
INDUSTRIAS LÁCTEAS TO...	MANUFACTURA	Agroindustria	Productos Lácteos	INDUSTRIAS LÁCTEAS...	PICHINCHA	158.71	
LACTEOS SAN ANTONIO CA	MANUFACTURA	Agroindustria	Productos Lácteos	LACTEOS SAN ANTON...	PICHINCHA	98.10	
REYLACTEOS CL	MANUFACTURA	Agroindustria	Productos Lácteos	REYLACTEOS CL	PICHINCHA	85.76	
EMPRESA PASTEURIZADO...	MANUFACTURA	Agroindustria	Productos Lácteos	EMPRESA PASTEURIZ...	PICHINCHA	80.03	
ALPINA PRODUCTOS	MANUFACTURA	Agroindustria	Productos Lácteos	ALPINA PRODUCTOS	PICHINCHA	63.74	
CORPABE SA	MANUFACTURA	Agroindustria	Productos Lácteos	CORPABE SA	PICHINCHA	63.36	
EL ORDEÑO SA	MANUFACTURA	Agroindustria	Productos Lácteos	EL ORDEÑO SA	PICHINCHA	43.19	
NUTRITON CIA LTDA	MANUFACTURA	Agroindustria	Productos Lácteos	NUTRITON CIA LTDA	PICHINCHA	32.85	
FLOPRALP SA	MANUFACTURA	Agroindustria	Productos Lácteos	FLOPRALP SA	PICHINCHA	27.55	

*Nota.* Dataset luego de cambiar identificación de cada fila con el comando `rownames`

### **Preparación del Set de datos**

Para trabajar con el set de datos, es necesario escalar o normalizar, de modo que cada variable tenga el mismo peso, adicional se retiraron las variables nominales, correspondientes a las 5 primeras variables, es decir, el set de datos está compuesto de 6 variables numéricas. En la Figura 22, se indica el código para normalización.

**Figura 22**

*Código para normalizar los datos*

```
> #Preparación del Dataset
> df<- DATOSF
> #Escalar o Normalizar el dataset
> df.scale <- as.data.frame(scale(df[,6:11])) #normalizar datos
> df.scale
  IngresosAnuales UtilidadBruta ImpCausados UtilidadNeta CvubicacGeogr
1      0.496056667   -0.262248275  -0.313264305  -0.242520560    1.0289058
2      0.110747461    0.388474707  0.377772666  0.382727107    1.0289058
3      0.032299752   -0.300553245  -0.313264305  -0.290394690    1.0289058
```

En la Figura 23 se visualiza el set de datos normalizado en RStudio.

**Figura 23**

*Dataset normalizado*

	INGRESOS ANUALES	UTILIDAD BRUTA	IMPUESTOS CAUSADOS	UTILIDAD NETA	CV POR UBICACIÓN GEOGRÁFICA	CV POR SECTOR
INDUSTRIAS LÁCTEAS TONI SA	0.496056667	-0.26224828	-0.313956225	-0.242520560	1.0289058	-0.2129033
LACTEOS SAN ANTONIO CA	0.110747461	0.38847471	0.375825690	0.382727107	1.0289058	-0.2129033
REYLACTEOS CL	0.032299752	-0.30055324	-0.313956225	-0.290394690	1.0289058	-0.2129033
EMPRESA PASTEURIZADORA QUITO	-0.004126939	-0.18517683	-0.189118394	-0.180226512	1.0289058	-0.2129033
ALPINA PRODUCTOS	-0.107685543	-0.13671874	-0.100250785	-0.143888317	1.0289058	-0.2129033
CORPABE SA	-0.110101274	-0.21471319	-0.242015780	-0.202721585	1.0289058	-0.2129033
EL ORDEÑO SA	-0.238325770	-0.25440268	-0.275870107	-0.243097357	1.0289058	-0.2129033
NUTRITON CIA LTDA	-0.304059101	0.07788140	0.071136746	0.077601631	1.0289058	-0.2129033
FLOPRALP SA	-0.337752201	-0.29547668	-0.284333689	-0.292125080	1.0289058	-0.2129033
LECHE GLORIA ECUADOR SA	-0.341185083	-0.21655921	-0.237783989	-0.206182365	1.0289058	-0.2129033

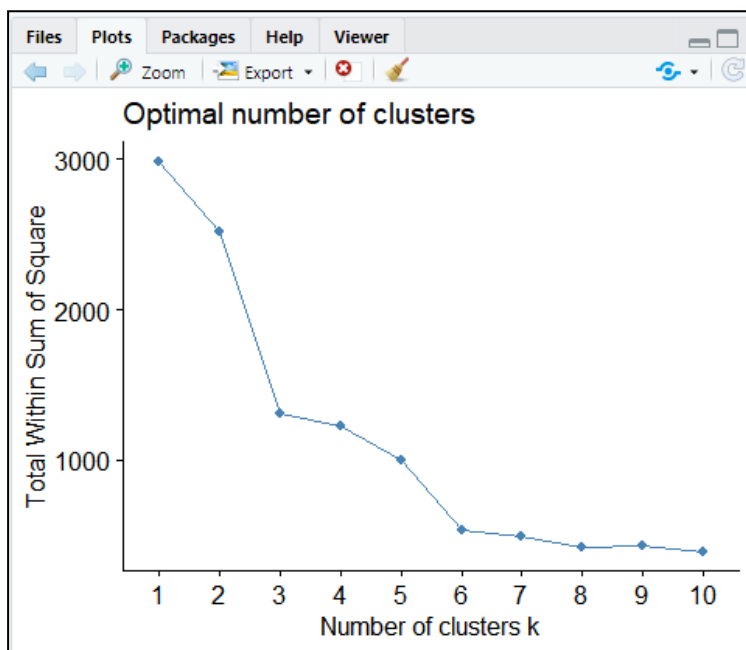
*Nota.* Vista del dataset normalizado en R

### ***Estimación del número de clústers***

Existen varios métodos para definir el número de clústers, por ejemplo: wss, silueta, cambio de estadística, entre otros. En la Figura 24 se muestra el método wss, también conocido como el método del codo, para determinar el número de clústers.

**Figura 24**

*Número de Clústers por el método wss*



*Nota.* Vista gráfica del método wss para el cálculo del número de clústers en R

En la Tabla 17, se muestran las diferentes técnicas para calcular el número de clústers, en la Figura 25 se cita el código mostrado para tal efecto.

**Figura 25**

*Código para calcular número de clústers*

```
> #cálculo del número de clústers, varias técnicas
> resnumclust<-NbClust(df.scale, distance ="euclidean", min.nc=2, max.nc=10, method =
"kmmeans", index = "kl")
> fviz_nbclust(resnumclust)
Number_clusters      value_Index
          4.0000          22.2919
```

*Nota:* En el parámetro de index se coloca el nombre del método, por ejemplo, kl, ch, cindex, etc.

**Tabla 17***Cálculo del número de clústers*

<b>Método</b>	<b># Clústers</b>	<b>Value_index</b>
<b>KI</b>	4.0000	22.2919
<b>Ch</b>	6.0000	450.3092
<b>Hartigan</b>	4.0000	232.3897
<b>Cindex</b>	10.0000	0.0254
<b>Db</b>	4.0000	0.8399
<b>Silhouette</b>	10.0000	0.6531
<b>Duda</b>	2.0000	4.8984
<b>pseudot2</b>	2.0000	-388.3753
<b>Beale</b>	2.0000	-3.0498
<b>Ratkowsky</b>	4.0000	0.4174
<b>Ball</b>	3.0000	822.0904
<b>Ptbserial</b>	4.0000	0.4316
<b>Gap</b>	2.0000	0.7907
<b>Frey</b>	1	NA
<b>Mcclain</b>	4.0000	0.3822
<b>Dunn</b>	4.0000	0.0955
<b>Sdindex</b>	2.0000	3.9355
<b>Sdbw</b>	10.0000	0.8851

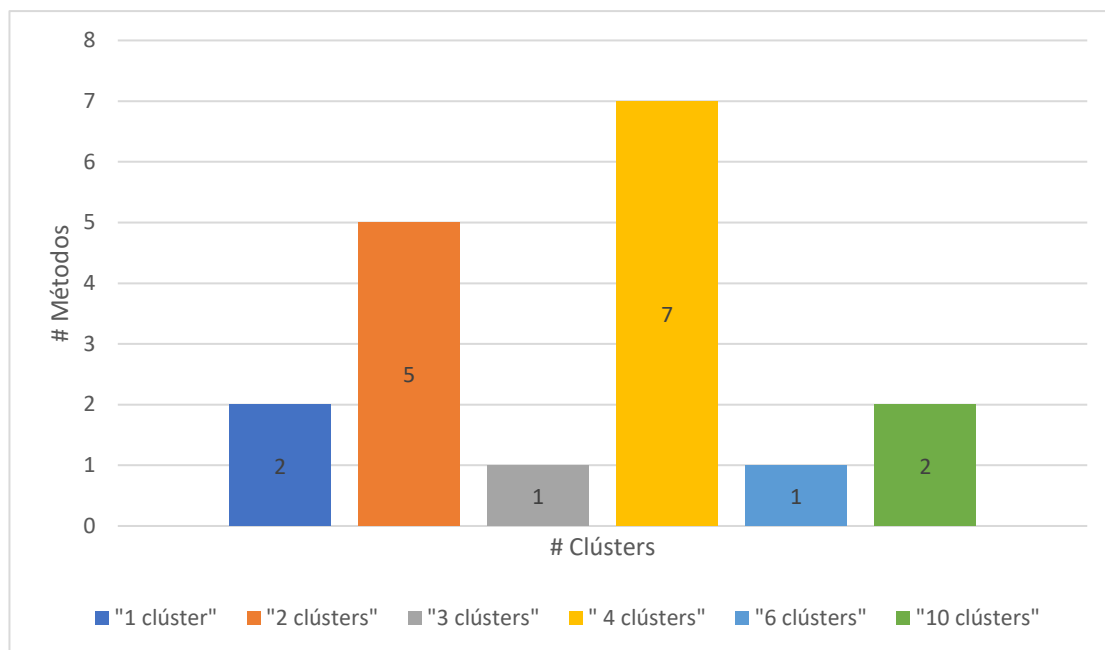
*Nota. Diferentes métodos probados para calcular el número de clústers*

En la Figura 26 se muestra un resumen con el número de clústers sugeridos para los diferentes métodos versus el número de métodos que lo sugieren.



**Figura 26**

*Frecuencias del número de clústers*



*Nota.* Resumen que muestra que la mayor cantidad de métodos, sugieren 4 clústers

Otro método utilizado para definir el número de clústers, es el de valor de inercia intergrupos, para lo cual es necesario sumar los cuadrados de las distancias inter grupos, para ello se utilizó del código indicado en la Figura 27, cuyo resultado se visualiza en la Figura 28.

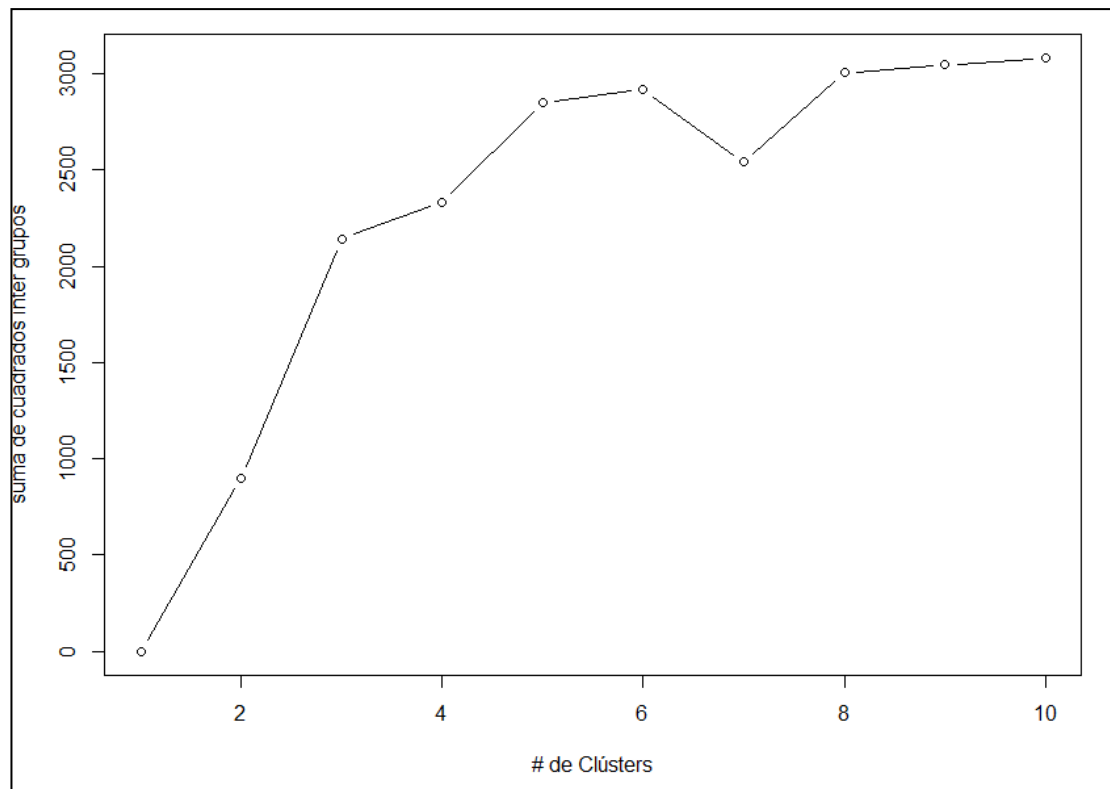
**Figura 27**

*Número de clústers por distancia inter grupos*

```
+ submt<-kmeans(df.scale,centers = 1)$betweenss
+ for (i in 2:10) submt[i] <- kmeans(df.scale, centers = i)$betweenss
+ plot (1:10, submt, type = "b", xlab = "# de Clústers", ylab = "suma de cuadrados inte
r grupos")
```

**Figura 28**

*Número de clústers por el método de distancias intergrupos*



*Nota.* Gráfico del número de clústers sugeridos por distancia inter grupos. Fuente: Propia.

De acuerdo al método de distancia intergrupos, se observa que el número de clústers apropiado pueden ser 5 o 6 clústers.

Para el análisis del presente proyecto, se desarrollaron modelos K-means para 4, 5 y 6 clústers, de acuerdo al análisis realizado con los métodos citados en la Tabla 17 y el método de las distancias intergrupos mostrada en la Figura 28.

### ***Creación de los clústers***

Una vez seleccionado el número de clústers, se crearon los segmentos con el código mostrado en la Figura 29.

Figura 29

## Clusterización

```

141 #Creamos el cluster
142 k4 <- kmeans(df.scale, centers = 4, nstart = 25)
143 k4
144 str(k4)
145
146
147 df1<-USArrests
148 df1
149
144:1 (Top Level)

```

---

```

Cluster means:
INGRESOS ANUALES UTILIDAD BRUTA IMPUESTOS CAUSADOS UTILIDAD NETA CV POR UBICACIÓN GEOGRÁFICA
1 4.007203499 6.80876840 6.74840222 6.67341657 0.06689921
2 -0.008654808 -0.09682664 -0.07753415 -0.09991784 1.02890584
3 -0.170036849 -0.14481816 -0.14061105 -0.14273472 0.50233379
4 -0.101443983 -0.11971443 -0.13313395 -0.11339427 -0.83587761
CV POR SECTOR
1 -0.2032995
2 -0.2045332
3 4.8835087
4 -0.2042056

```

---

```

Clustering vector:
INDUSTRIAS LÁCTEAS TONI SA          LACTEOS SAN ANTONIO CA
                                2                                2
                                REYLACTEOS CL          EMPRESA PASTEURIZADORA QUITO
                                2                                2
                                ALPINA PRODUCTOS          CORPABE SA
                                2                                2

```

*Nota.* Código para creación de 4 clústers y visualización en RStudio. Fuente: Propia.

### Visualización gráfica de los clústers

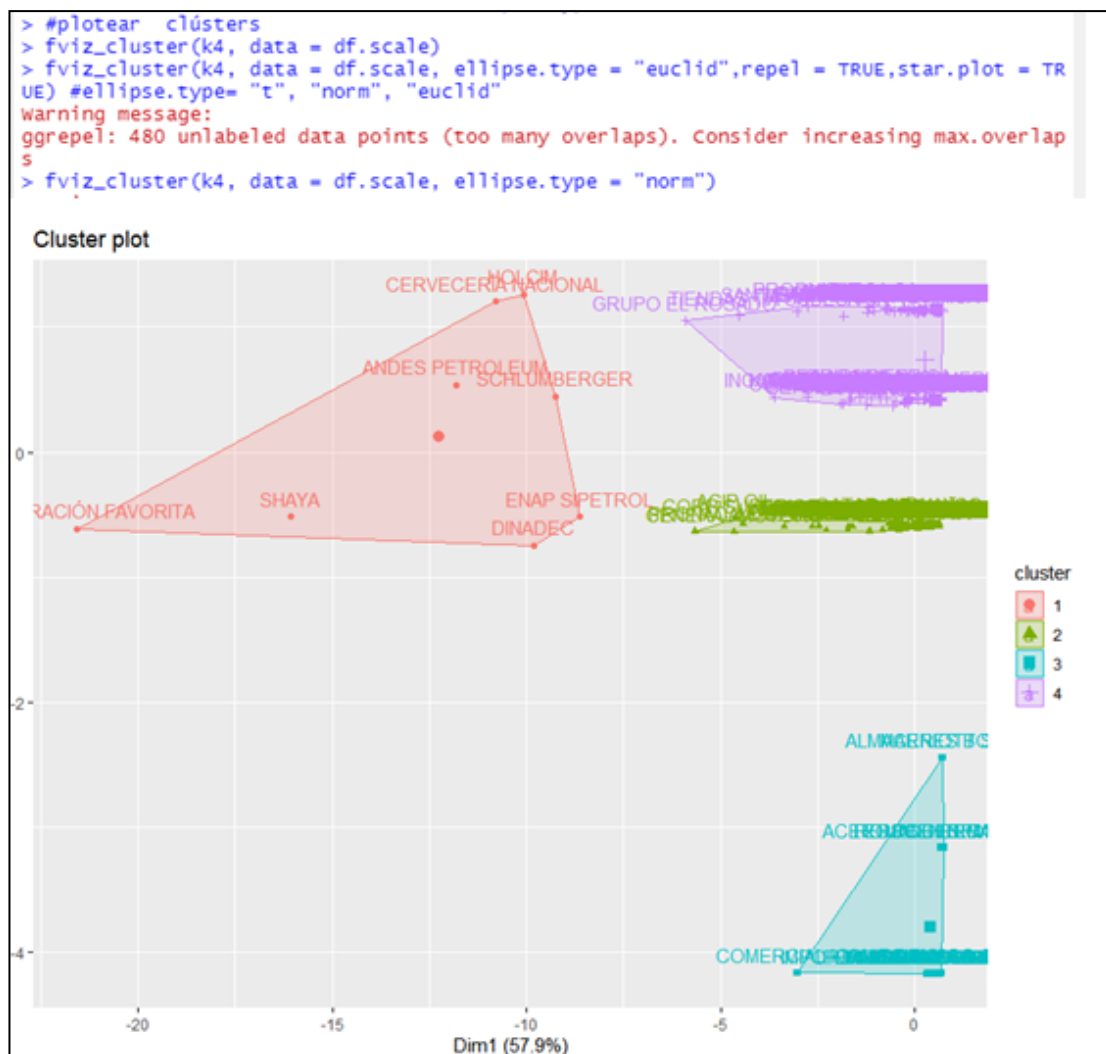
Para realizar el análisis de los clústers, se procedió a correr el código que nos permita tener una visualización gráfica.

### Visualización inicial

Para una visualización inicial de los clústers modelados, se utilizó la función: *fviz\_cluster*, en la Figura 30 se muestran los resultados de esta visualización.

Figura 30

Visualización inicial de los clústers



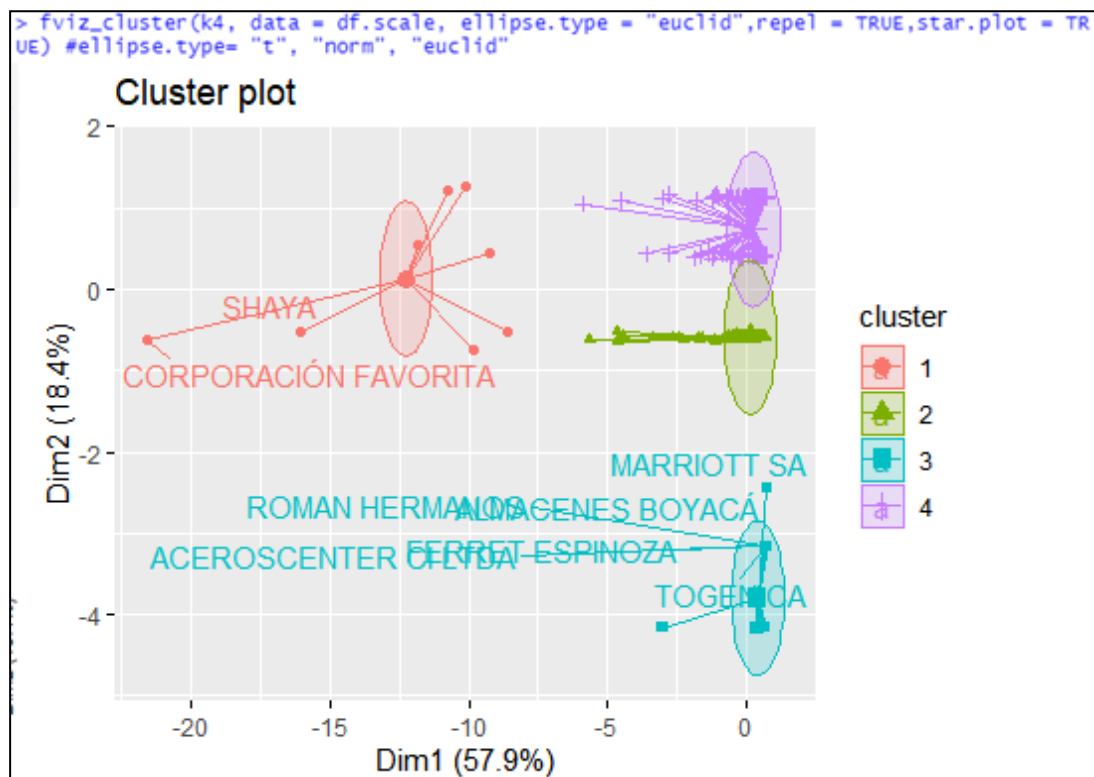
Nota. Agrupamiento 4 clústers. Fuente: Propia

### Visualización distancia euclidiana

Este tipo de visualización permite identificar los centroides de cada clúster, y la distancia de cada elemento hacia dicho centroide, para ello se utilizó el código que se observa en la Figura 31.

Figura 31

Visualización clústers distancia euclidiana



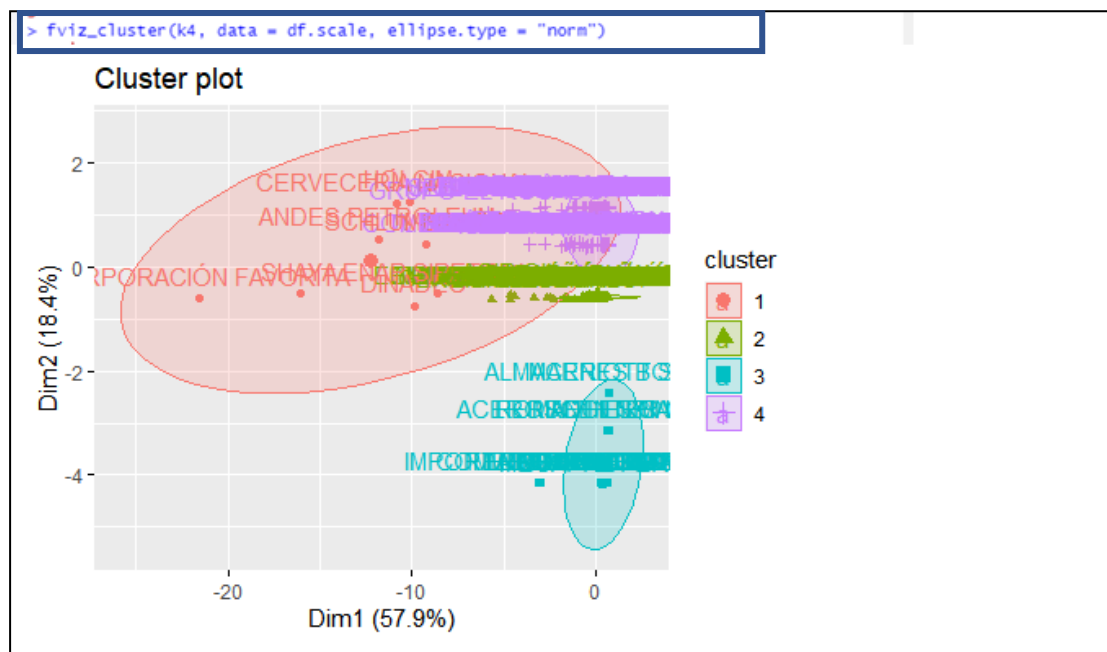
Nota. Muestra el centroide de los 4 clústers y la distancia a cada elemento. Fuente: Propia.

### Visualización agrupación elíptica

En la Figura 32 se muestra una vista de los clústers en una agrupación elíptica, este tipo de visualización define a cada clúster dentro de una elipse.

**Figura 32**

Visualización agrupación elíptica



Nota. Muestra los 4 clústers en una agrupación en forma de elipse. Fuente: Propia.

### Modelo K- Means con 4 centroides

En las Figuras 33 y 34 se muestran los resultados de los clústers, tomando cuatro centroides.

**Figura 33**

K-means con 4 centroides

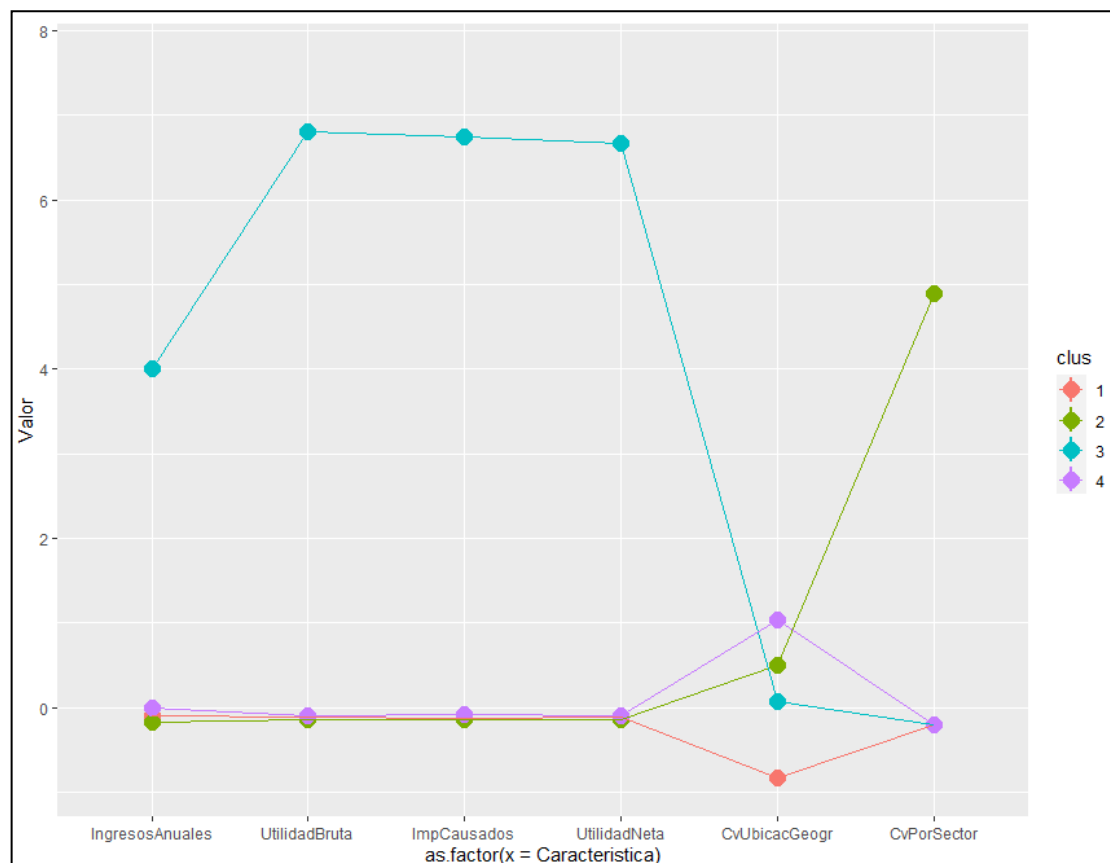
```
> k4
K-means clustering with 4 clusters of sizes 265, 20, 8, 205

Cluster means:
  IngresosAnuales  UtilidadBruta  ImpCausados  UtilidadNeta  CvubicacGeogr  CvPorSector
1  -0.101443983   -0.11971443  -0.13313395  -0.11339427  -0.83587761  -0.2042056
2  -0.170036849  -0.14481816  -0.14061105  -0.14273472   0.50233379   4.8835087
3   4.007203499   6.80876840   6.74840222   6.67341657   0.06689921  -0.2032995
4  -0.008654808  -0.09682664  -0.07753415  -0.09991784   1.02890584  -0.2045332
```

Nota: Muestra los clústers formados para agrupar a los clientes potenciales, de acuerdo a las variables indicadas. Fuente: Propia.

**Figura 34**

*Comparación de variables para K-means con 4 centroides*



*Nota:* Muestra una comparación gráfica de los clústers creados con 4 centroides. Fuente: Propia.

Para el caso del modelo desarrollado con 4 clústers, se puede visualizar que el clúster que agrupa a los clientes con una mejor posición financiera, de acuerdo a las variables indicadas en las gráficas, se encuentra en el clúster No. 3.

### **Modelo K-means con 5 centroides**

En las Figuras 35 y 36 se muestran los resultados de los clústers, tomando cinco centroides.

**Figura 35***K-means con 5 centroides*

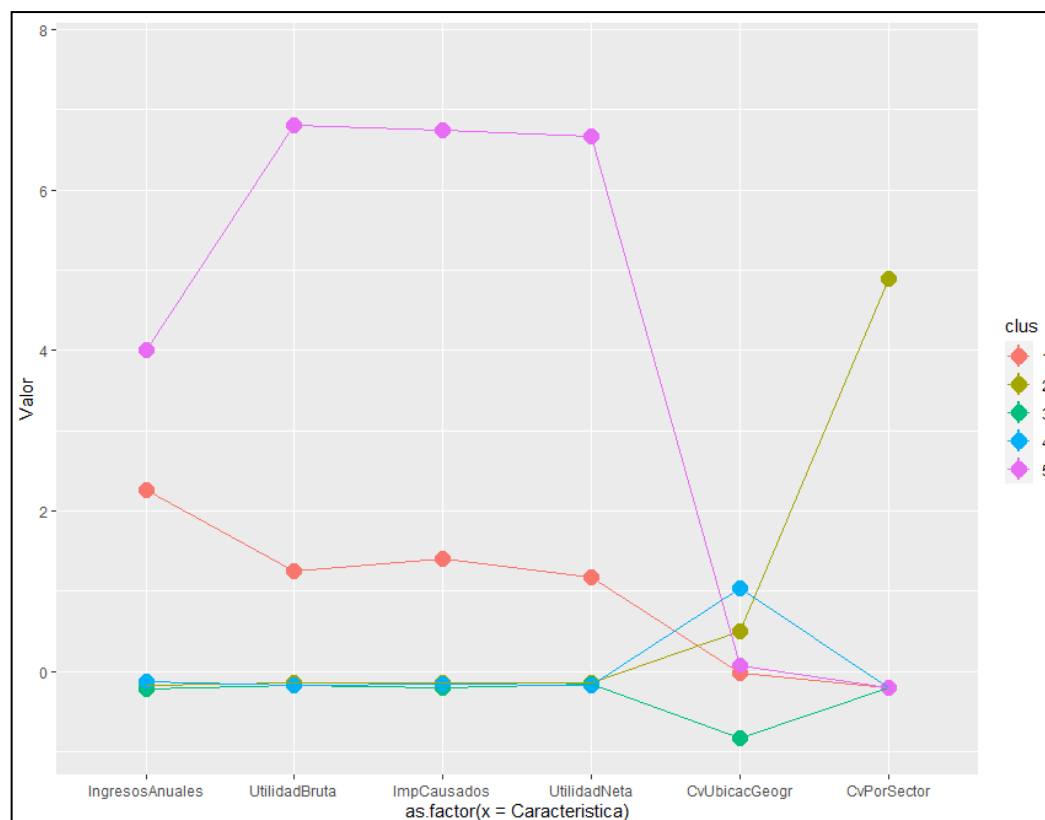
```

> k5
K-means clustering with 5 clusters of sizes 22, 20, 253, 195, 8

Cluster means:
  IngresosAnuales  UtilidadBruta  ImpCausados  UtilidadNeta  CvUbicacGeogr  CvPorSector
1      2.2614913      1.2417777      1.4051747      1.1696352     -0.02055593    -0.2012624
2     -0.1700368     -0.1448182     -0.1406110     -0.1427347      0.50233379     4.8835087
3     -0.2169237     -0.1753101     -0.1979016     -0.1652549     -0.83306829    -0.2041979
4     -0.1206565     -0.1771253     -0.1642038     -0.1766930      1.02890584    -0.2048920
5      4.0072035      6.8087684      6.7484022      6.6734166      0.06689921    -0.2032995

```

*Nota:* Muestra los clústers formados para agrupar a los clientes potenciales, de acuerdo a las variables indicadas. Fuente: Propia.

**Figura 36***Comparación de variables para K-means con 5 centroides*

*Nota:* Muestra una comparación gráfica de los clústers creados con 5 centroides. Fuente: Propia.



Para el caso del modelo desarrollado con 5 clústers, se puede visualizar que el clúster que agrupa a los clientes con una mejor posición financiera, de acuerdo a las variables indicadas en las gráficas, se encuentra en el clúster No. 5 y el clúster No. 1.

### **Modelo K-means con 6 centroides**

En las Figuras 37 y 38 se muestran los resultados de los clústers, tomando seis centroides.

#### **Figura 37**

##### *K-means con 6 centroides*

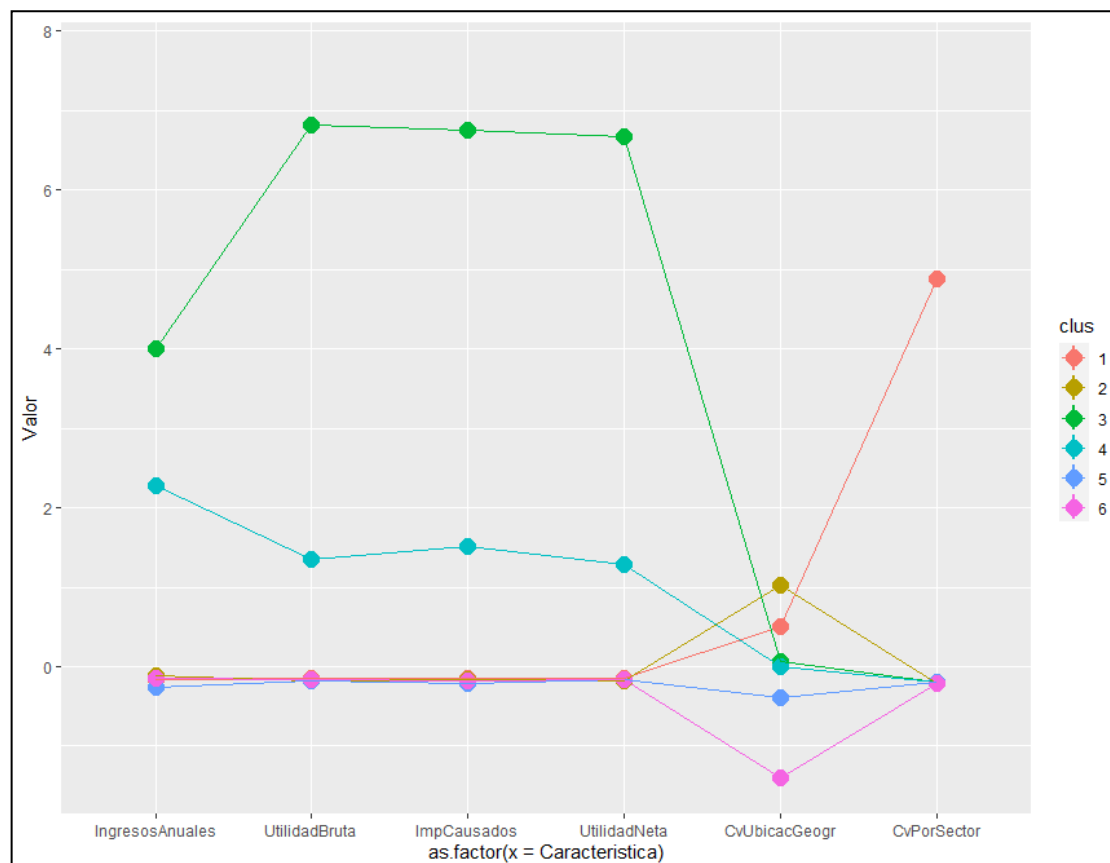
```
> k6 <- kmeans(df.scale, centers = 6, nstart = 25)
> k6
K-means clustering with 6 clusters of sizes 20, 196, 8, 20, 142, 112

Cluster means:
  IngresosAnuales  UtilidadBruta  ImpCausados  UtilidadNeta  CvUubicacGeogr  CvPorSector
1    -0.1700368    -0.1448182   -0.1406110   -0.1427347    0.502333793    4.8835087
2    -0.1089651    -0.1760150   -0.1615792   -0.1760194    1.028905843   -0.2048022
3     4.0072035     6.8087684    6.7484022    6.6734166    0.066899214   -0.2032995
4     2.2724173     1.3530446    1.5197855    1.2775119   -0.003985485   -0.2013788
5    -0.2601144    -0.1785955   -0.2151440   -0.1646692   -0.388788137   -0.2033446
6    -0.1411771    -0.1676355   -0.1727744   -0.1625003   -1.401426693   -0.2053575
```

*Nota:* Muestra los clústers formados para agrupar a los clientes potenciales, de acuerdo a las variables indicadas. Fuente: Propia.

**Figura 38**

*Comparación de variables para K-means con 6 centroides*



*Nota:* Muestra una comparación gráfica de los clústers creados con 6 centroides. Fuente: Propia.

Para el caso del modelo desarrollado con 6 clústers, se puede visualizar que el clúster que agrupa a los clientes con una mejor posición financiera, de acuerdo a las variables indicadas en las gráficas, se encuentra en el clúster No. 3 y el clúster No. 4.

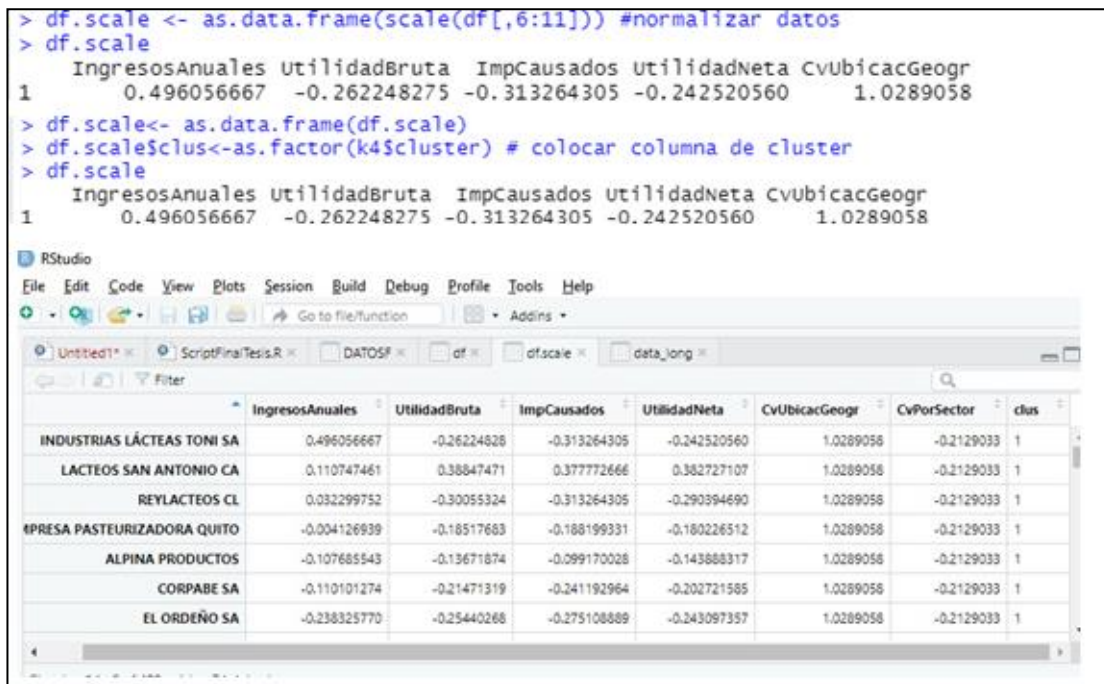
### ***Dataframe con datos clusterizados***

En la Figura 39 se muestra el código que permite identificar a cada empresa (cliente potencial) en el clúster que le corresponde.

Figura 39

*Agrupamiento de datos dentro de cada clúster*

```
> df.scale <- as.data.frame(scale(df[,6:11])) #normalizar datos
> df.scale
  IngresosAnuales UtilidadBruta ImpCausados UtilidadNeta CvUbicacGeogr
1      0.496056667  -0.262248275 -0.313264305 -0.242520560   1.0289058
> df.scale<- as.data.frame(df.scale)
> df.scale$clus<-as.factor(k4$cluster) # colocar columna de cluster
> df.scale
  IngresosAnuales UtilidadBruta ImpCausados UtilidadNeta CvUbicacGeogr
1      0.496056667  -0.262248275 -0.313264305 -0.242520560   1.0289058
```



The screenshot shows the RStudio interface with a data frame view. The data frame has the following columns: IngresosAnuales, UtilidadBruta, ImpCausados, UtilidadNeta, CvUbicacGeogr, CvPorSector, and clas. The data is as follows:

	IngresosAnuales	UtilidadBruta	ImpCausados	UtilidadNeta	CvUbicacGeogr	CvPorSector	clas
INDUSTRIAS LÁCTEAS TONI SA	0.496056667	-0.262248275	-0.313264305	-0.242520560	1.0289058	-0.2129033	1
LACTEOS SAN ANTONIO CA	0.110747461	0.38847471	0.377772666	0.382727107	1.0289058	-0.2129033	1
REYLACTEOS CL	0.032299752	-0.30055324	-0.313264305	-0.290394690	1.0289058	-0.2129033	1
MPRESA PASTEURIZADORA QUITO	-0.004126939	-0.18517683	-0.188199331	-0.180226512	1.0289058	-0.2129033	1
ALPINA PRODUCTOS	-0.107685543	-0.13671874	-0.099170028	-0.143888317	1.0289058	-0.2129033	1
CORPABE SA	-0.110101274	-0.21471319	-0.241192964	-0.202721585	1.0289058	-0.2129033	1
EL ORDEÑO SA	-0.238325770	-0.25440268	-0.275108889	-0.243097357	1.0289058	-0.2129033	1

*Nota.* Muestra el código y el dataframe, en el que se coloca el clúster al que pertenece cada empresa. Fuente: Propia.

### **Análisis de los datos**

Luego de obtener los clústers para cuatro, cinco y seis centroides, se identificaron los grupos en los cuales se encuentran los clientes con un mejor perfil financiero, de acuerdo a las siguientes variables:

- Mayores Ingresos anuales
- Mayor utilidad neta
- Baja caída en ventas por ubicación geográfica
- Baja caída en ventas por sector productivo

En la Tabla 18 se muestra el resumen de los clústers que se seleccionaron para el presente proyecto.

**Tabla 18**

*Clústers seleccionados*

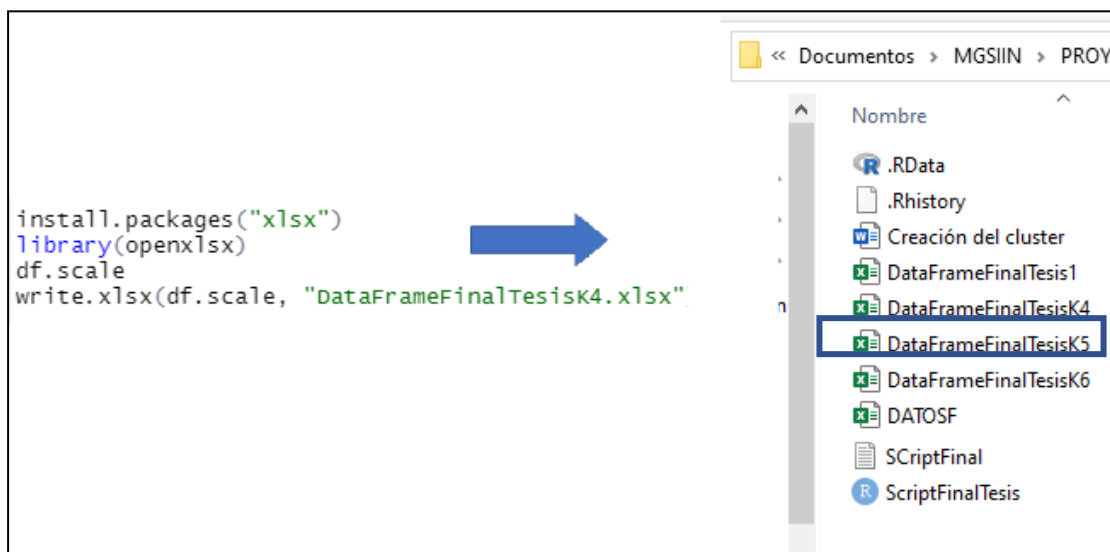
No. Centroides K- Means	Clústers seleccionados
4	3
5	5, 1
6	3, 4

*Nota:* Muestra los clústers seleccionados como clientes potenciales.

Una vez seleccionados los clústers, se exportó la información del dataframe final a un formato de Excel, de acuerdo a lo indicado en la Figura 40.

**Figura 40**

*Exportación del dataframe a excel*



*Nota.* Muestra el código para exportar el dataframe a un archivo de Excel que se almacena en la carpeta del proyecto de RStudio.

### **Datos obtenidos del modelo desarrollado**

Como entregable del modelo desarrollado, se obtuvo un conjunto de clientes potenciales, los cuales deben ser abordados por el recurso comercial de Proing Cia Ltda, para la correspondiente prospección.

En la Tabla 19 se muestran las empresas que corresponden a los clientes potenciales identificados en el modelo desarrollado para 4 centroides.

**Tabla 19** *Clientes potenciales para K-means con 4 centroides*

IngresosAnuales	UtilidadBruta	ImpCausados	UtilidadNeta	CvUbicacGeogr	CvPorSector	clustros	EMPRESA
1,94835642	6,6566446	3,91986708	7,2529530	-1,40142669	0,2129033	3	HOLCIM
12,867718	10,574366	9,46088132	10,641633	1,02890584	0,1872932	3	CORPORACIÓN FAVORITA
5,31054638	3,3845694	9,38669023	1,6759054	1,02890584	0,1872932	3	DINADEC
1,71364902	6,6714127	6,04385188	6,6934602	-1,40142669	0,1872932	3	CERVECERÍA NACIONAL
4,0174942	9,2332307	9,28282271	9,0139135	1,02890584	0,2129033	3	SHAYA
2,8698945	5,0612195	5,52239453	4,8229084	-0,38878814	0,2129033	3	SCHLUMBERGER
2,61935769	7,4928928	5,08996649	7,9797169	-0,38878814	0,2129033	3	ANDES PETROLEUM
0,71061179	5,3958111	5,28074357	5,3068408	1,02890584	0,2129033	3	ENAP SIPETROL

*Nota.* Se muestran los clientes potenciales para el modelo K-means de 4 centroides, que corresponden al clúster No. 3. Fuente: Propia.

En la Tabla 20 se pueden visualizar las empresas que corresponden a los clientes potenciales identificados en el modelo desarrollado para 5 centroides.

**Tabla 20**

*Clientes potenciales para K-means con 5 centroides*

IngresosAnuales	UtilidadBruta	ImpCausados	UtilidadNeta	CvUbicacGeogr	CvPorSector	clustros	EMPRESA
2,1336049	0,2177176	0,16791788	0,2264151	-1,40142669	-,2129033	1	EXPALSA SA
2,56538516	-0,0199578	0,38837139	0,1306219	-0,38878814	-,2129033	1	LA FABRIL
0,54348129	1,1951866	2,14352051	0,9104961	1,02890584	-,2129033	1	NESTLÉ

IngresosAnuales	UtilidadBruta	ImpCausados	UtilidadNeta	CvUbicacGeogr	CvPorSector	clust	EMPRESA
5,52828009	2,0582021	2,16047847	1,9844916	1,02890584	-,2129033	1	PRONACA
1,94835642	6,6566446	3,91986708	7,2529530	-1,40142669	-,2129033	5	HOLCIM
0,38080084	1,6825365	1,72381094	1,6337992	-0,38878814	-,2129033	1	UNACEM
1,58892097	1,5196250	2,29190268	1,2756084	1,02890584	-,2129033	1	UNILEVER
12,867718	10,574366	9,46088132	10,641633	1,02890584	-,1872932	5	CORPORACIÓN FAVORITA
7,04307101	1,8057585	1,96758165	1,7214723	-1,40142669	-,1872932	1	GRUPO EL ROSADO
3,99937621	1,8149887	1,80436126	1,7774216	-1,40142669	-,1872932	1	TIENDAS TÍA SA
1,69712033	0,6824538	0,8610746	0,6186369	-0,38878814	-,1872932	1	GERARDO
2,01326332	0,4397018	1,25746697	0,2073809	-1,40142669	-,1872932	1	ORTIZ
0,64640418	2,6724662	2,65013964	2,6189680	1,02890584	-,1872932	1	UBESA SA
4,54742945	1,7019198	1,74288865	1,6528335	1,02890584	-,1872932	1	CORPORACIÓN SUPERIOR GENERAL
2,25922295	0,4613926	0,43288604	0,4588642	1,02890584	-,1872932	1	MOTORS
5,31054638	3,3845694	9,38669023	1,6759054	1,02890584	-,1872932	5	AEKIA
2,92240487	0,6409183	0,6088249	0,6353640	1,02890584	-,1872932	1	DINADEC
1,71364902	6,6714127	6,04385188	6,6934602	-1,40142669	-,1872932	5	AC BEBIDAS DEL CV
1,2035482	1,3313307	1,41220838	1,2796460	1,02890584	-,1872932	1	CERVECERÍA NACIONAL
2,17085803	0,0404994	0,35021598	0,0446792	1,02890584	-,1872932	1	ARCADOR SA
4,0174942	9,2332307	9,28282271	9,0139135	1,02890584	-,2129033	5	LETERAGO
2,8698945	5,0612195	5,52239453	4,8229084	-0,38878814	-,2129033	5	SHAYA
2,61935769	7,4928928	5,08996649	7,9797169	-0,38878814	-,2129033	5	SCHLUMBERGER
0,71061179	5,3958111	5,28074357	5,3068408	1,02890584	-,2129033	5	ANDES
0,61919541	0,7221433	1,53303386	0,4853969	-0,38878814	-,2129033	1	PETROLEUM ENAP
0,34322985	3,0684380	2,33217784	3,2003791	1,02890584	-,2129033	1	ENAP
0,25505564	2,1038912	2,78580334	1,8714395	-0,38878814	-,2129033	1	SIPETROL HALLIBURTON
3,87159672	1,0899633	0,24634846	1,2952195	-1,40142669	-,2129033	1	AGIP OIL
3,15825657	0,3317095	0,35869496	0,3169722	-0,38878814	-,2129033	1	INGAPO SA
0,26230284	1,7582235	1,6941345	1,7364690	-1,40142669	0,2129033	1	SANTA PRISCILIA
							OMARSA
							PRODUMAR

*Nota.* Se muestran los clientes potenciales para el modelo K-means de 5 centroides, que corresponden al clúster No. 5 y No. 1. Fuente: Propia.

En la Tabla 21 se detallan las empresas que corresponden a los clientes potenciales identificados en el modelo desarrollado para 6 centroides.

Tabla 21

*Clientes potenciales para K-means con 6 centroides*

IngresosAnuales	UtilidadBruta	ImpCausados	UtilidadNeta	CvUbicacGeogr	CvPorSector	clust	EMPRESA
2,56538516	0,0199578	0,3883713	0,130621	-0,38878814	-0,212903	4	LA FABRIL
0,54348129	1,1951866	2,1435205	0,910496	1,02890584	-,2129033	4	NESTLÉ
5,52828009	2,0582021	2,1604784	1,984491	1,02890584	-,2129033	4	PRONACA
1,94835642	6,6566446	3,9198670	7,252953	-1,40142669	-,2129033	3	HOLCIM
0,38080084	1,6825365	1,7238109	1,633799	-0,38878814	-,2129033	4	UNACEM
1,58892097	1,5196250	2,2919026	1,275608	1,02890584	-,2129033	4	UNILEVER CORPORACI ÓN
12,867718	10,574366	9,4608813	10,64163	1,02890584	-,1872932	3	FAVORITA GRUPO EL ROSADO
7,04307101	1,8057585	1,9675816	1,721472	-1,40142669	-,1872932	4	TIENDAS TÍA SA
3,99937621	1,8149887	1,8043612	1,777421	-1,40142669	-,1872932	4	GERARDO
1,69712033	0,6824538	0,8610746	0,618636	-0,38878814	-,1872932	4	ORTIZ
2,01326332	0,4397018	1,2574669	0,207380	-1,40142669	-,1872932	4	UBESA SA CORPORACI ÓN
0,64640418	2,6724662	2,6501396	2,618968	1,02890584	-,1872932	4	SUPERIOR GENERAL
4,54742945	1,7019198	1,7428886	1,652833	1,02890584	-,1872932	4	MOTORS
2,25922295	0,4613926	0,4328860	0,458864	1,02890584	-,1872932	4	AEKIA
5,31054638	3,3845694	9,3866902	1,675905	1,02890584	-,1872932	3	DINADEC AC BEBIDAS
2,92240487	0,6409183	0,6088249	0,635364	1,02890584	-,1872932	4	DEL CV CERVECERÍ A NACIONAL
1,71364902	6,6714127	6,0438518	6,693460	-1,40142669	-,1872932	3	ARCADOR
1,2035482	1,3313307	1,4122083	1,279646	1,02890584	-,1872932	4	SA
4,0174942	9,2332307	9,2828227	9,013913	1,02890584	-,2129033	3	SHAYA SCHLUMBER
2,8698945	5,0612195	5,5223945	4,822908	-0,38878814	-0,212903	3	GER ANDES
2,61935769	7,4928928	5,0899664	7,979716	-0,38878814	-,2129033	3	PETROLEUM ENAP
0,71061179	5,3958111	5,2807435	5,306840	1,02890584	-0,212903	3	SIPETROL HALLIBURTO
0,61919541	0,7221433	1,5330338	0,485396	-0,38878814	-0,212903	4	N
0,34322985	3,0684380	2,3321778	3,200379	1,02890584	-0,212903	4	AGIP OIL
0,25505564	2,1038912	2,7858033	1,871439	-0,38878814	-0,212903	4	INGAPO SA SANTA
3,87159672	1,0899633	0,2463484	1,295219	-1,40142669	-0,212903	4	PRISCILIA
3,15825657	0,3317095	0,3586949	0,316972	-0,38878814	-0,212903	4	OMARSA
0,26230284	1,7582235	1,6941345	1,736469	-1,40142669	-0,212903	4	PRODUMAR

*Nota.* Se muestran los clientes potenciales para el modelo K-means de 6 centroides, que

corresponden al clúster No. 3 y No. 4. Fuente: Propia.

En el campo EMPRESA constan los nombres de los clientes de cada clúster seleccionado. Haciendo uso de Excel, se analizó cada conjunto de datos de Tabla 19, Tabla 20 y Tabla 21 para identificar duplicidad y consolidar un solo listado.

Finalmente se encontró el set de datos correspondientes a los clientes potenciales, detallado en la Tabla 22.

**Tabla 22**

*Set de datos resultado del modelo desarrollado*

<b>Sector</b>	<b>Empresa</b>
MANUFACTURA	EXPALSA SA
MANUFACTURA	LA FABRIL
MANUFACTURA	NESTLÉ
MANUFACTURA	PRONACA
MANUFACTURA	<b>HOLCIM</b>
MANUFACTURA	<b>UNACEM</b>
MANUFACTURA	UNILEVER
COMERCIO	CORPORACIÓN FAVORITA
COMERCIO	GRUPO EL ROSADO
COMERCIO	TIENDAS TÍA SA
COMERCIO	GERARDO ORTIZ
COMERCIO	UBESA SA
COMERCIO	<b>CORPORACIÓN SUPERIOR</b>
COMERCIO	GENERAL MOTORS
COMERCIO	AEKIA
COMERCIO	DINADEC
COMERCIO	AC BEBIDAS DEL CV



Sector	Empresa
COMERCIO	CERVECERÍA NACIONAL
COMERCIO	ARCADOR SA
COMERCIO	<b>LETERAGO</b>
PETRÓLEOS Y MINAS	SHAYA
PETRÓLEOS Y MINAS	SCHLUMBERGER
PETRÓLEOS Y MINAS	ANDES PETROLEUM
PETRÓLEOS Y MINAS	ENAP SIPETROL
PETRÓLEOS Y MINAS	HALLIBURTON
PETRÓLEOS Y MINAS	AGIP OIL
PETRÓLEOS Y MINAS	INGAPO SA
PESCA	<b>SANTA PRISCILIA</b>
PESCA	OMARSA
PESCA	PRODUMAR

*Nota:* Base de datos de 30 clientes potenciales, información obtenida luego de correr el modelo de clusterización.

Como podemos observar la mayor cantidad de clientes se concentran en el sector productivo de Manufactura y Comercio. Esta base de 30 clientes potenciales permitirá una prospección objetiva, ya que los clústers seleccionados están agrupando a empresas, cuyas variables financieras los perfilan como clientes potenciales, esta información fue entregada al área comercial de la empresa caso de estudio, para la prospección o búsqueda de clientes.

## Capítulo V

### Validación de la Solución

La validación del modelo desarrollado, pertenece a la fase de diseño de la metodología de investigación propuesta Design Research, en cuyo Objetivo Específico No. 4, se plantea la evaluación del modelo, mediante juicio de expertos.

En la Figura 41, se muestran los pasos definidos para validar el modelo propuesto.

**Figura 41**

*Proceso de validación del modelo*



*Nota.* Proceso de validación del modelo. Fuente: Propia.

#### Prospección por parte del equipo comercial

El equipo comercial prospecció el listado de clientes obtenido del modelo de datos desarrollado, el detalle de dicho listado se muestra en la Tabla 22.

La prospección se realizó en un lapso de 3 meses, en el período de abril a junio del 2021.

### Informe de prospección haciendo uso del modelo

Para el proceso de prospección, se definió un período de prueba de tres meses, el cual inició en abril del 2021 a junio del 2021. Se definieron las variables o criterios de evaluación mostrados en la Tabla 23, que permitieron cuantificar el desempeño de la búsqueda de nuevos clientes.

**Tabla 23**

*Variables o criterios para evaluación de la prospección*

<b>Criterios de evaluación del modelo propuesto</b>	<b>Período de evaluación</b>
No. Total de clientes prospectados	<b>abril 2021– junio 2021</b>
No. Reuniones a nuevos clientes	
No. Leads generados	
Monto de pipeline generado	

*Nota.* Se detallan las variables que se han definido para medir la prospección

### Evaluación de variables del informe de prospección

Luego de realizar la prospección por parte del recurso comercial de Proing, se obtuvo el informe que se muestra en la Tabla 24.

**Tabla 24**

*Variables comerciales luego de aplicar el modelo*

<b>Criterios de evaluación del modelo propuesto</b>	<b>Valores haciendo uso del modelo abril 2021- junio 2021</b>
No. Total de clientes prospectados	20
No. Reuniones nuevos clientes	10
No. Leads generados	5
Monto de pipeline generado	\$50.000,00

*Nota.* Variables reportadas en informe comercial, luego de utilizar el modelo

Luego de analizar los datos del informe comercial de la Tabla 27, se determinó la tasa de conversión de la prospección posterior al uso del modelo, con la fórmula mostrada en la Tabla 25.

### **Tabla 25**

*Fórmula Tasa de Conversión – Prospección*

---


$$\frac{\text{CANTIDAD DE LEADS GENERADOS}}{\text{TOTAL DE CLIENTES PROSPECTADOS}} * 100\%$$


---

*Nota.* La tabla muestra la fórmula usada para calcular la tasa de conversión del proceso de prospección luego de utilizar el modelo desarrollado.

$$\frac{\text{CANTIDAD DE LEADS GENERADOS}}{\text{TOTAL DE CLIENTES PROSPECTADOS}} * 100\% = \frac{5}{20} = 25\%$$

La tasa de conversión que ha generado el modelo propuesto es del 25%.

### **Validación por expertos**

Con la finalidad de evaluar el modelo propuesto y de acuerdo a la recomendación de CRISP DM, se evaluó el modelo en función de la utilidad o aporte del mismo al dominio de aplicación o problema de negocio, para ello se realizó una encuesta a diez asesores comerciales, la población entrevistada cuenta con el siguiente perfil:

- Profesionales del área comercial con más de 5 años de experiencia.
- Comercialización de soluciones tecnológicas en empresas B2B.
- Incluye perfiles de Gerentes Comerciales.

### **Encuesta de validación**

Se formularon siete preguntas que nos permitieron evaluar el modelo propuesto.

### **Objetivo de la encuesta**

Evaluar el modelo de gestión de datos desarrollado, que permita la posibilidad de disminuir la concentración de clientes en la empresa Proing Cia Ltda.

### **Preguntas de la encuesta**

Las preguntas se definen con base en las preguntas de investigación **OE4-RQ8** y **OE4-R9**. En la Tabla 26 se detallan las preguntas de la encuesta y su correspondencia con las preguntas de investigación.

**Tabla 26**

#### *Preguntas de Investigación y cuestionario*

<b>Preguntas de Investigación</b>	<b>Cuestionario para entrevista</b>
<b>OE4-RQ8:</b> ¿El uso de analítica de datos, aporta en procesos comerciales como lo es la prospección de clientes?	<ol style="list-style-type: none"> <li>1. ¿En una escala del 1 al 5, siendo 5 la escala más alta y 1 lo más bajo, que tan útil considera aplicar analítica de datos para encontrar clientes potenciales?</li> <li>2. ¿Qué ventajas considera que aporta el uso de analítica de datos para prospectar clientes potenciales?               <ol style="list-style-type: none"> <li>a. Prospección acertada</li> <li>b. Mayor tasa de conversión en prospección</li> <li>c. Optimización de tiempo de los recursos comerciales</li> <li>d. Ninguna</li> </ol> </li> </ol>
<b>OE4-RQ9:</b> ¿Cuál es el nivel de aceptación del modelo planteado para el estudio de caso, de acuerdo al juicio de expertos?	<ol style="list-style-type: none"> <li>3. ¿Un modelo de datos, en el cual se relacionan variables como Ingresos anuales, utilidad total, caída de ventas por ubicación geográfica, caídas de venta por sector, aporta a la prospección objetiva de nuevos clientes? SI/No</li> </ol>
	<ol style="list-style-type: none"> <li>4. ¿Considera que el listado de clientes obtenido luego de correr el modelo de datos, corresponden a clientes potenciales que apoyarán a la prospección en una empresa B2B de</li> </ol>

Preguntas de Investigación	Cuestionario para entrevista
	servicios tecnológicos (Proing Cia Ltda)? Si/No
	5. ¿Considera que la tasa de conversión en la prospección, es un valor que da un indicativo del éxito del proceso comercial para cierre de ventas? Si/No
	6. Con el modelo desarrollado hemos encontrado una tasa de conversión de prospección del 25%, considera Ud, que esta tasa de conversión incrementará la probabilidad de cierre de ventas, en un corto o mediano plazo? Si/No
	7. ¿Considera que el conocimiento obtenido del modelo desarrollado, permitirá la posibilidad de reducir la concentración de clientes en una empresa B2B de servicios tecnológicos (Proing Cia Ltda)? Si/No

***Propósito de la encuesta***

Determinar el nivel de aceptación del modelo propuesto, que permita la posibilidad de disminuir la concentración de clientes en la empresa Proing Cia Ltda mediante el uso de analítica de datos.

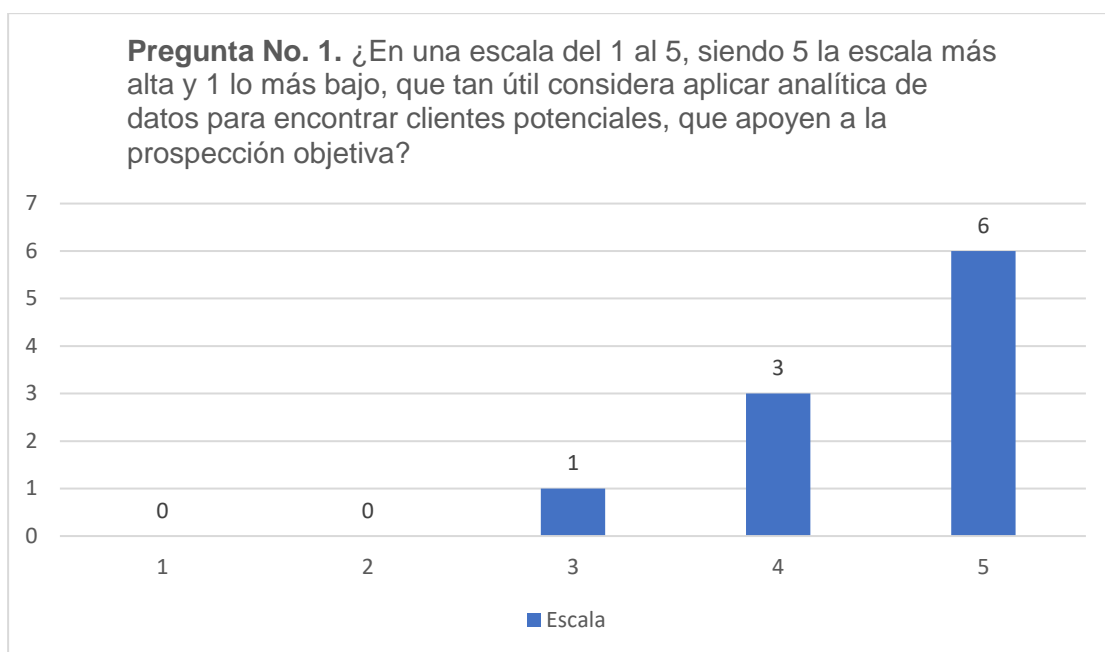
## Resultados de la encuesta

### Pregunta No. 1

Con esta pregunta buscamos evaluar la utilidad del uso de la analítica de datos en la prospección de clientes. En la Figura 42 mostramos los resultados para la pregunta No.1.

### Figura 42

Resultado de la encuesta a la Pregunta No.1



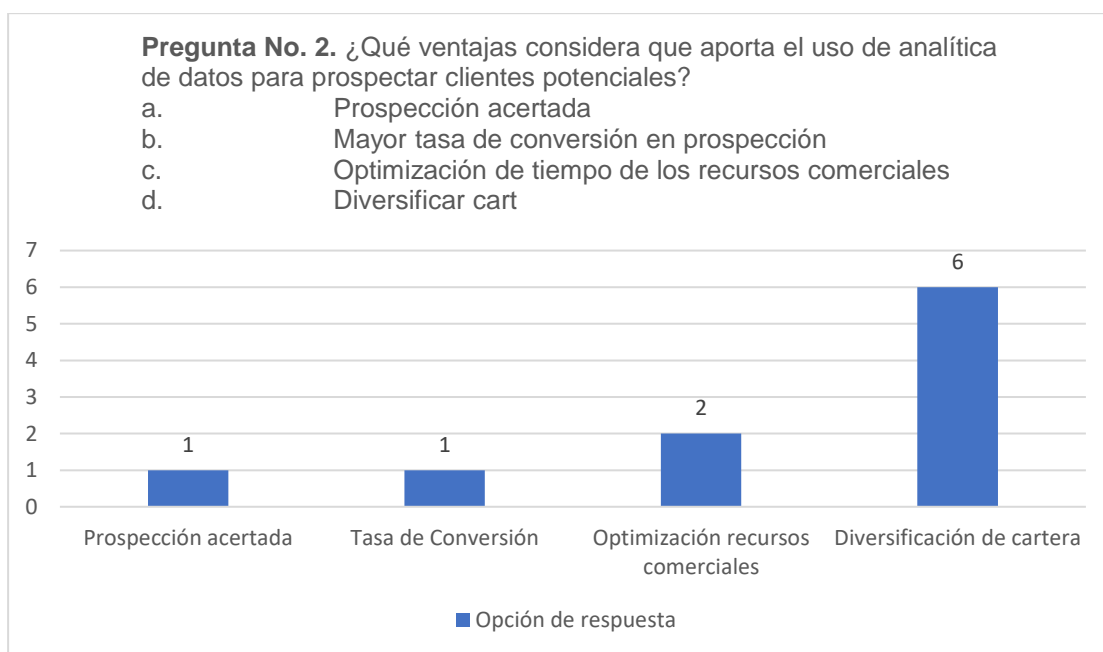
El 90% de los entrevistados considera que es útil el uso de la analítica de datos para encontrar clientes potenciales.

## Pregunta No. 2

La pregunta No 2, nos permite determinar cuáles son las ventajas más importantes, del uso de analítica de datos en la prospección de clientes. En la Figura 43 se visualizan los resultados para la pregunta No.2.

### Figura 43

*Resultado de la encuesta a la Pregunta No.2*



*Nota. El gráfico muestra los resultados a la Pregunta No. 2. Fuente: Propia.*

Los resultados obtenidos en la pregunta No.2 indican que, el 60% de los encuestados consideran a la diversificación de la cartera como la principal ventaja del uso de la analítica de datos en la prospección de clientes, seguida con un 20% la ventaja de optimización del tiempo de los recursos comerciales.

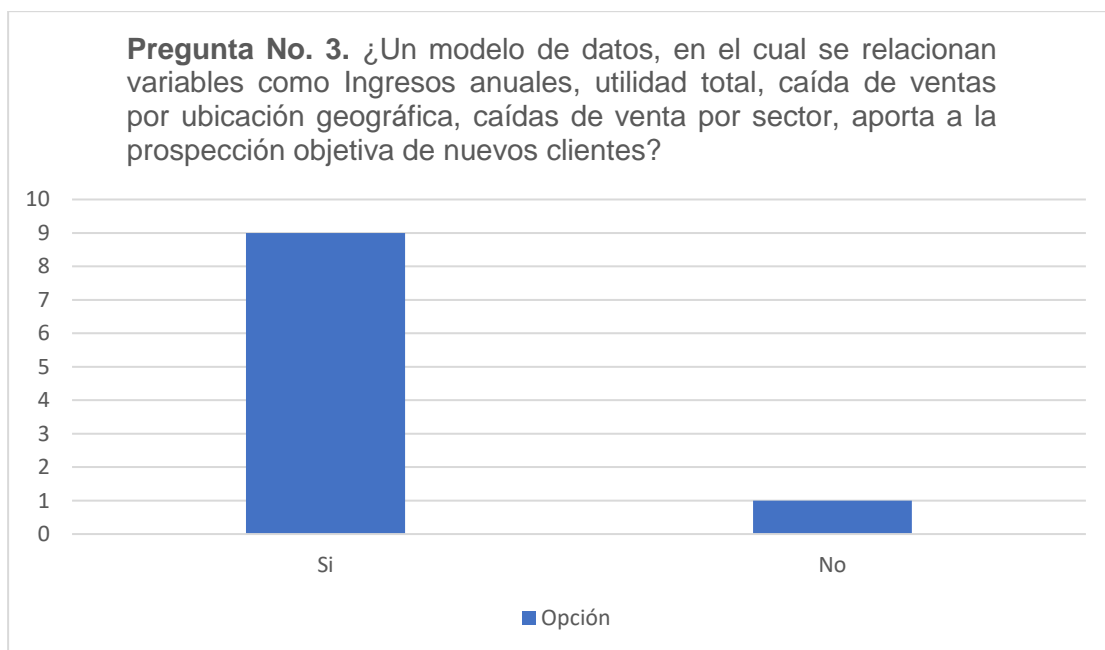


### Pregunta No. 3

Con esta pregunta buscamos evaluar la utilidad del modelo, desde el punto de vista de la relación objetiva de las variables utilizadas. En la Figura 44 se indican los resultados para la pregunta No.3.

#### Figura 44

*Resultado de la encuesta a la Pregunta No.3*



*Nota.* El gráfico muestra los resultados a la Pregunta No. 3, con las respuestas de los expertos. Fuente: Propia.

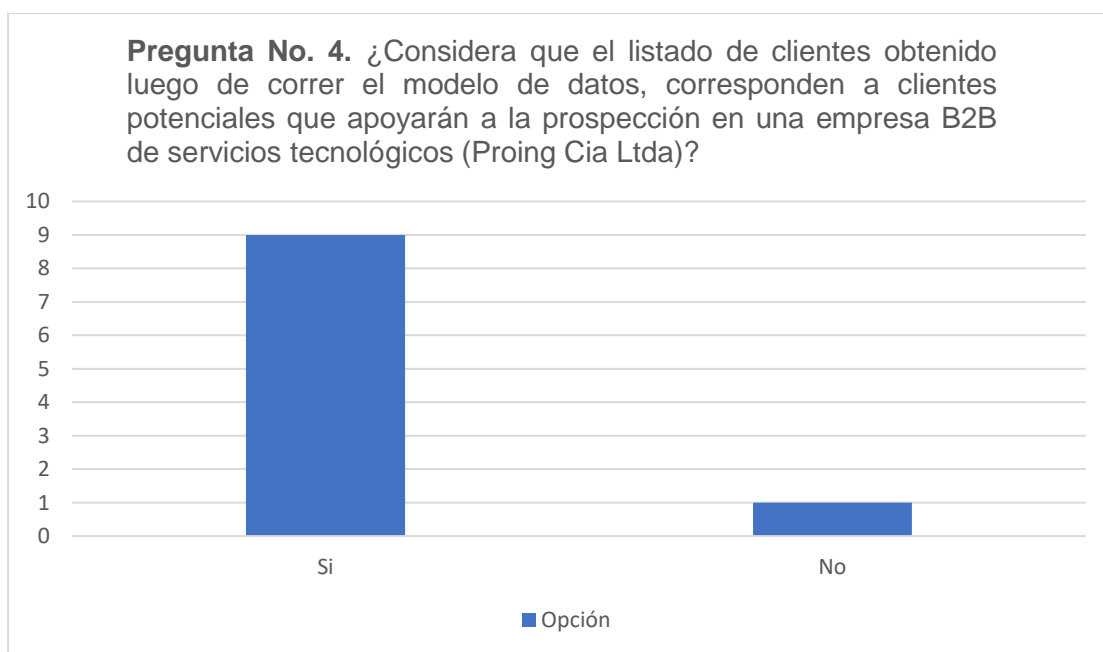
Como se puede observar, el 90% de los encuestados indica que el modelo desarrollado, haciendo uso de las variables que se interrelacionaron en el modelo K-means, si aporta al proceso de prospección objetiva.

#### Pregunta No. 4

Con la pregunta No. 4 pretendemos evaluar la utilidad de la información entregada por el modelo desarrollado, para la prospección de clientes, en la Figura 45 se observan los resultados de la encuesta para esta pregunta.

#### Figura 45

*Resultado de la encuesta a la Pregunta No.4*



*Nota. El gráfico muestra los resultados a la Pregunta No.4. Fuente: Propia.*

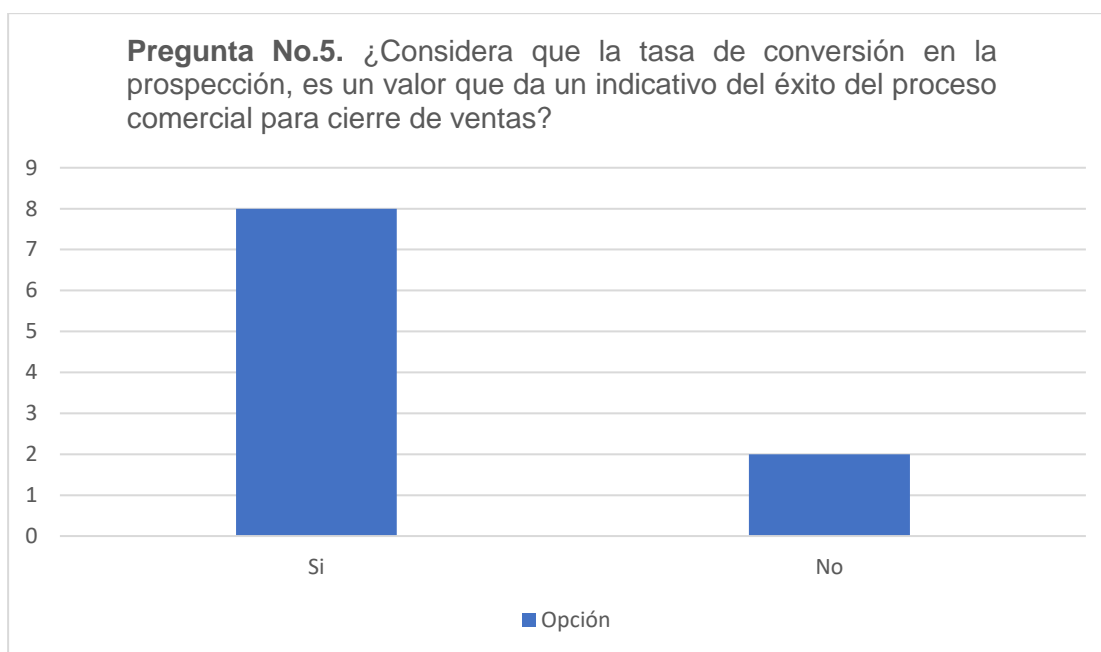
Como se puede visualizar, el 90% de los encuestados considera que el listado de clientes obtenido del modelo, apoyarán a la prospección de clientes.

### Pregunta No. 5

Con esta pregunta pretendemos determinar la importancia de la tasa de conversión en la prospección, dentro del proceso comercial para el cierre de ventas, en la Figura 46 se pueden visualizar las respuestas para la pregunta No. 5.

### Figura 46

*Resultado de la encuesta a la Pregunta No.5*



*Nota. El gráfico muestra los resultados a la Pregunta No. 5. Fuente: Propia.*

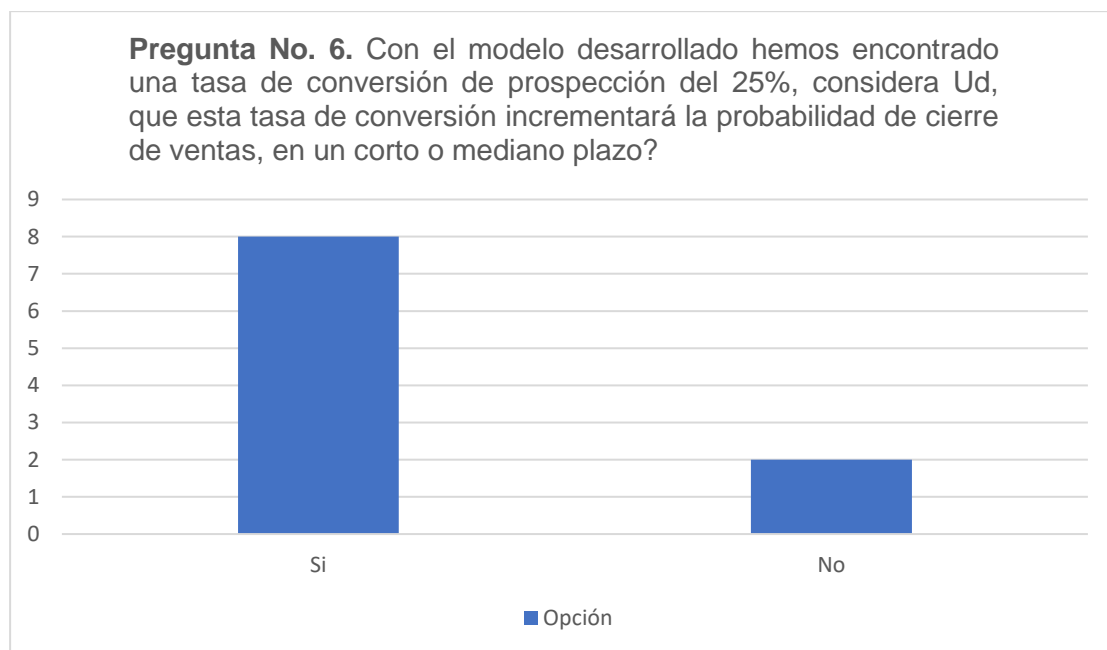
El 80% de los encuestados coinciden en que la Tasa de conversión en prospección, es un valor importante para proyectar el cierre de venta comercial.

### Pregunta No. 6

Con la pregunta No. 6 se busca validar si tasa de conversión en prospección, generada por el modelo, es útil para incrementar la posibilidad de cierre de ventas. En la figura 47 se visualiza los resultados de la pregunta No. 6

### Figura 47

*Resultado de la encuesta a la Pregunta No.6*



*Nota. El gráfico muestra los resultados a la Pregunta No. 6. Fuente: Propia.*

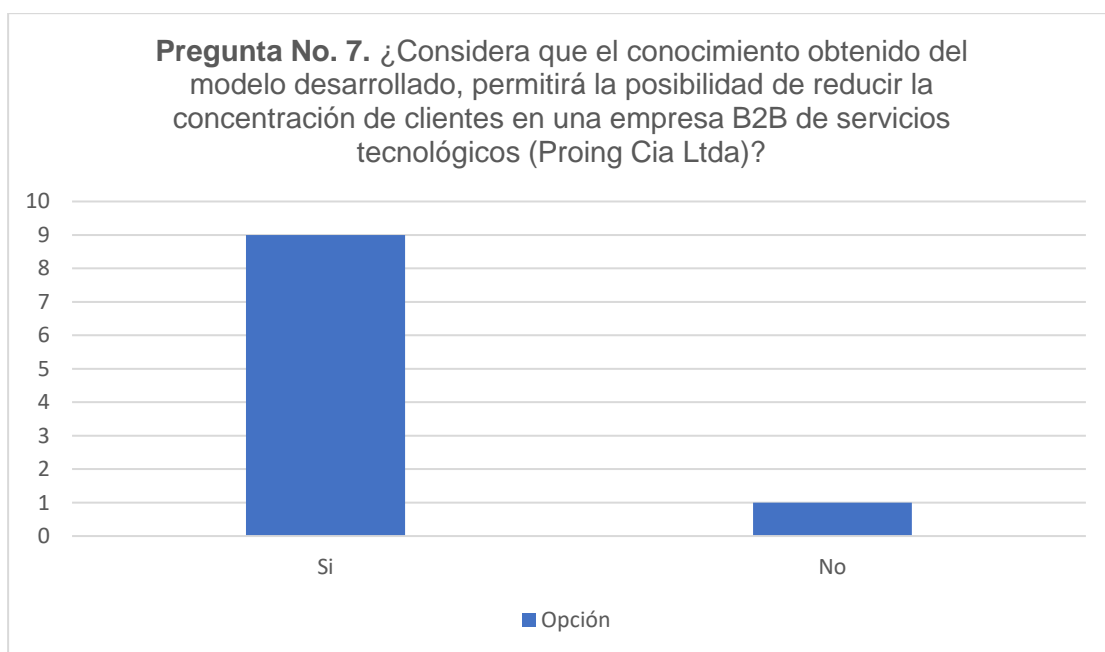
De acuerdo a las respuestas obtenidas, el 80% de los encuestados coincide en que la tasa de conversión en prospección del 25% obtenida luego de utilizar la información del modelo, incrementará la posibilidad de cierre de ventas en un corto a mediano plazo.

### Pregunta No. 7

La pregunta No.7 pretende identificar la importancia del modelo desarrollado y su aporte a la solución de la problemática identificada para el presente proyecto. En la Figura 48 se muestran los resultados.

### Figura 48

*Resultado de la encuesta a la Pregunta No.7*



*Nota. El gráfico muestra los resultados a la Pregunta No. 7. Fuente: Propia.*

Como podemos visualizar, el 90% de los encuestados coincide en que el conocimiento generado por el modelo desarrollado aportará a reducir la concentración de clientes.

De acuerdo a los resultados mostrados para cada una de las preguntas planteadas, se evidencia una aceptación del modelo propuesto, con lo cual podemos concluir que la tasa de conversión de prospección del 25% obtenida con el modelo, permitirá en un corto o mediano plazo, que la cartera de clientes de Proing Cia Ltda se diversifique, incrementando lo posibilidad de reducir la concentración.

## Capítulo VI

### Conclusiones y Recomendaciones

#### Conclusiones

Las pequeñas y medianas empresas (PYMES) son de gran importancia para las economías de los países (Ron, 2017), no obstante, enfrentan dificultades que afectan su operación y ponen en riesgo la continuidad del negocio, siendo la concentración de clientes una de sus causas (Marín, 2020) y que corresponde a la problemática identificada en la empresa estudio de caso Proing Cia Ltda.

Mediante un estudio exploratorio realizado a través de reuniones con la gerencia administrativa y financiera de Proing Cia Ltda, se pudo conocer los principales efectos que se han evidenciado como consecuencia de la *dependencia financiera en una alta concentración de clientes*, los cuales son: (1) Reducción de ingresos de ventas e incremento de cuentas por cobrar, (2) Falta de liquidez, (3) Incremento de créditos bancarios para pago a proveedores.

En cuanto a las causas identificadas, Proing cuenta con una cartera de clientes poco diversificada, lo cual es consecuencia de una prospección poco efectiva, basada en la intuición del gerente general. Proing, no cuenta con herramientas para realizar una prospección objetiva basada en datos para la toma de decisiones en la búsqueda de clientes potenciales.

Proing cuenta con un sistema de facturación, el cual maneja información correspondiente a las transacciones de venta con sus tres principales clientes, por tanto, es una fuente con escasa información que no aporta en el desarrollo del presente proyecto.

Con el propósito de identificar propuestas de soluciones para abordar la problemática identificada, se realizó una revisión preliminar de literatura, encontrando que se plantean soluciones enfocadas en instrumentos financieros (Liu, 2021), (Pan, 2020), (Itzkowitz, 2014), mejoras en la cadena de producción (Krishnaswamy, 2014), (Larkin, 2019) e innovación tecnológica (Shree, 2021), (Yun, 2016). En (Li, 2014) se plantea el uso de Sistemas de Soporte a la toma de decisiones el cual hace uso de algoritmos analíticos de datos.

Luego de la revisión de literatura preliminar, se evidencia que no hay muchas soluciones basadas en analítica de datos para abordar la problemática de concentración de clientes, por lo cual se valida plenamente la propuesta de solución basada en el desarrollo de un modelo de datos, haciendo uso de analítica, para la prospección de clientes.

En cuanto a la frecuencia de publicaciones relacionadas con propuestas de solución para la concentración de clientes, pudimos observar que en el 2014 y 2021 existe la mayor cantidad de publicaciones al respecto.

De acuerdo a (Paschen, 2021) y (Niladri, 2018), se define al modelo de aprendizaje no supervisado, como la mejor *técnica* para segmentación de clientes, dentro de ello se encuentra K-means como el modelo de analítica de datos sugerido para determinar el grupo o clúster de clientes potenciales.

Dentro de la *metodología* más apropiada para el desarrollo de modelos analíticos y minería de datos, se tiene a CRISP DM, ampliamente utilizada, dado que se basa en el entendimiento del negocio (Moine J. M., 2011).

Como *herramienta* para el desarrollo del modelo analítico hemos seleccionado el lenguaje R, de acuerdo a la recomendación citada en (Bohanec, 2016).

Para desarrollar el modelo analítico con K-means, hacemos uso de un set de datos construido a partir de la información financiera proporcionada por la Superintendencia de Compañías, cuyas variables son interrelacionadas con K-means para perfilar clientes potenciales.

Luego de analizar las diferentes técnicas para definir el número de clústers de k-means, se seleccionan cuatro, cinco y seis clústers para el desarrollo del modelo, con el cual se obtuvo un listado de 30 clientes potenciales.

El modelo desarrollado, fue probado en la empresa estudio de caso, la lista de los 30 clientes potenciales obtenidos de la clusterización bajo K-means, fue puesta a disposición de un recurso comercial de Proing Cia Ltda, quien comenzó la tarea de prospección, evidenciando una tasa de conversión para prospección del 25%.

Luego de someter el modelo a un juicio de expertos, se confirma la aceptación del modelo analítico propuesto, desde el punto de vista de su utilidad, de la pertinencia de las variables utilizadas y de la tasa de conversión de prospección obtenida, es decir, los datos obtenidos del modelo desarrollado, permitieron iniciar una adecuada prospección, lo cual permite concluir que, en un corto a mediano plazo, se podrá evidenciar la posibilidad de disminuir la concentración y diversificación de la cartera de clientes de la empresa caso de estudio.



## Recomendaciones

El uso de los datos proporciona una manera objetiva de resolver problemas de negocio, para la selección de la metodología de minería de datos, se recomienda partir del escenario que se desea abordar, considerando que es posible partir de una problemática dada, como el caso del presente proyecto, o también es posible partir de un conjunto de datos sobre los cuales se pueden analizar patrones de comportamiento para encontrar eventos innovadores, descubrir situaciones que causan gastos operativos, así como oportunidades de mejora no vistas.

La prospección o búsqueda de nuevos clientes, es una tarea cuyo paso inicial es identificar la empresa con quien se desea tener el acercamiento comercial, sin embargo, se debe aclarar que el cierre de ventas es un proceso sistemático que incluye una serie de pasos como la calificación, propuesta, acuerdos y cierre, en cuyas fases también se recomienda aplicar el uso de analítica de datos para conseguir mejores resultados, el alcance del presente proyecto abarca únicamente la fase de prospección, como punto de partida para diversificar la cartera de clientes de la empresa caso de estudio.

## Bibliografía

- Aguilar, J. (2018). Implementación de tareas de analítica de datos para mejorar la calidad de servicios en redes de comunicaciones. *Universidad de los Andes*.
- Ahumada, E. (2015). Inteligencia de negocios: estrategia para el desarrollo de competitividad en empresas de base tecnológica. *Universidad Autónoma de Baja California*.
- Ahumada-Tello. (2011). La Gestión del Conocimiento en la Generación de Competitividad Sistémica: El Sector de Tecnologías de la Información. *Universidad Autónoma de Baja California*.
- Belloch, C. (2012). Las Tecnologías de la Información y Comunicación en el aprendizaje. *Depto MIDE. Universidad de Valencia*.
- Bohanec, M. (2016). Explaining machine learning models in sales predictions. *Elsevier Ltd*.
- Bustos, S. (2013). Análisis, diseño e implementación de una solución de BI para la generación de indicadores y control de desempeño, en la empresa OTECEL S.A, utilizando la metodología HEFESTO V2.0. *Universidad de las Fuerzas Armadas ESPE*.
- Cabero, J. (1998). Impacto de las nuevas tecnologías de la información y la comunicación en las organizaciones educativas. <http://tecnologiaedu.us.es/revistaslibros/organiz.htm>.
- Campbell, C. (2019). From data to action: How marketers can leverage AI. *Business Horizons, Elsevier, vol. 63(2), 227-243*.

- CCQ, C. d. (Junio de 2017). *Clasificación de las Pymes, pequeña y mediana empresa*.  
Obtenido de [http://www.ccq.ec/wp-content/uploads/2017/06/Consulta\\_Societaria\\_Junio\\_2017.pdf](http://www.ccq.ec/wp-content/uploads/2017/06/Consulta_Societaria_Junio_2017.pdf)
- Cortez, R. M. (2021). B2B market segmentation: A systematic review and research agenda. *Journal of Business Research*.
- Dhaliwal, D. (2015). Customer Concentration Risk and the Cost of Equity Capital. *Journal of Accounting & Economics (JAE), Forthcoming*.
- Eckert, K. (2013). Aplicación de técnicas de Minería de Datos al análisis de situación y comportamiento académico de alumnos de la UGD. *Universidad Nacional de la Plata*.
- EKOS. (2020). Rankign Empresarial TOP 1000 Ecuador 2020.
- El País*. (2018). Obtenido de [https://cincodias.elpais.com/cincodias/2018/03/09/companias/1520610248\\_467686.html](https://cincodias.elpais.com/cincodias/2018/03/09/companias/1520610248_467686.html)
- Espino, C. (2017). Análisis predictivo: técnicas y modelos utilizados y aplicaciones del mismo - herramientas Open Source. *Universidad Oberta de Catalunya*.
- Fassler, M. (2018). Minería de datos para la toma de decisiones en la unidad de nivelación y admisión universitaria ecuatoriana. *Cumbres*, 55-67.
- Flecha, Á. (2020). Obtenido de <https://alvaroflecha.com/depender-de-un-cliente/>
- Flores, F. R. (2016). Inteligencia de negocios y minería de datos aplicado a la industria refresquera. *Computing Science*.
- González, D. S. (2008). *Minería de datos, técnicas y herramientas*. Paraninfo.
- Gujrat, P. (2014). A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA). *Innovative Space of Scientific Research Journals*, 217-222.

- INEC. (2017). *Tabulados del directorio de empresas y establecimientos, AÑO 2017*. Quito.
- Iitzkowitz, J. (2014). Customers and cash: How relationships affect suppliers' cash holdings. *Journal of Corporate Finance*.
- Jones, H. (2019). *Analítica de datos*. Kindle Edition.
- Kitchenham, B. (2015). *Evidence-Based Software Engineering and Systematic Reviews*. Chapman & Hall.
- Krishnaswamy, K. (2014). Technological innovations and its influence on the growth of auto component SMEs of Bangalore: A case study approach. *Technology in Society*.
- Larkin, Y. (2019). Reliance on major customers and product market competition. *Finance Research Letters*.
- Li, D.-C. (2014). Constructing marketing decision support systems using data diffusion technology: A case study of gas station diversification. *Expert Systems with Applications*.
- Liu, T. (2021). Achieving high efficiency and sustainability through new ventures exploration and exploitation strategies: Insight from well-established and emerging technology standards. *Sustainable Cities and Society*.
- Lozada, D. (2018). *Inteligencia de Negocios*. Instituto Tecnológico de Orizaba. Obtenido de <https://www.gestiopolis.com/wp-content/uploads/2018/04/inteligencia-en-los-negocios-diana-lozada.pdf>
- Mamani, Z. (2015). Aplicación de la minería de datos distribuida usando algoritmos de clustering K- means. *Universidad Nacional Mayor de San Marcos*.

- Manriquez, C. (03 de Agosto de 2018). *Los peligros a los que se expone una pyme cuando tiene un solo cliente*. Obtenido de Portalpyme: <https://pyme.emol.com/12320/peligros-tener-solo-un-cliente/>
- Marín, M. J. (2020). *Gateway Commercial Finance*. Obtenido de <https://gatewaycfs.com/bff/avoiding-high-customer-concentration>
- Meire, M. (2017). The added value of social media data in B2B customer acquisition systems: A real-life experiment. *The University of Tennessee*.
- Moine, J. (2011). Análisis comparativo de metodologías para la gestión de proyectos de minería de datos. *XVII Congreso Argentino de Ciencias de la Computación*.
- Moine, J. M. (2011). Estudio comparativo de metodologías para minería de datos. *UTN Rosario*.
- Mosquera, R. (2018). Máquinas de Soporte Vectorial, Clasificador Naïve Bayes y Algoritmos Genéticos para la predicción. *Información Tecnológica Vol.29*.
- Navarro, M. (2012). *Técnicas de ventas*. México: Red Tercer Milenio.
- Niladri, S. (2018). Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice. *Science Direct*.
- Núñez, M. (2019). *ADMINISTRACIÓN COMERCIAL*.
- Olivier, E. (2020). *Genwords*. Obtenido de <https://www.genwords.com/blog/prospeccion-de-clientes>
- Pan, J. (2020). Customer concentration and corporate innovation: Evidence from China. *North American Journal of Economics and Finance*.
- Paschen, J. (2021). Collaborative intelligence: How human and artificial intelligence create value along the B2B sales funnel. *Science Direct*.

- Pearlman, S. (2019). *TALEND REOSURCE CENTER*. Obtenido de <https://es.talend.com/resources/what-is-data-management/#:~:text=La%20gesti%C3%B3n%20de%20datos%20se,ciclo%20vit al%20de%20la%20informaci%C3%B3n>.
- Resendiz, J. (2006). Las maquinas de vectores de soporte para identificación en línea. *Instituto Politécnico Nacional de México*.
- Riquelme, J. (2006). Minería de Datos: Conceptos y Tendencias. *Revista Iberoamericana de Inteligencia Artificial*, 11 -18.
- Riquelme, M. (4 de Mayo de 2018). *Web y Empresas*. Obtenido de <https://www.webyempresas.com/administracion-comercial/>
- Rivadeneira, G. (2014). *La metodología de Kimball para el diseño de almacenes de datos (Data warehouses)*. Ucasal Argentina.
- Ron, R. (2017). Las PYMES ecuatorianas: su impacto en el empleo como contribución del PIB PYMES al PIB total. *ESPACIOS*.
- Salinas, M. (2008). Modelos de Regresión VI. Análisis de Supervivencia. *Ciencia & Trabajo*, 75-78.
- Shree, D. (2021). Digital platforms for business-to-business markets: A systematic review and future research agenda. *Journal of Business Research*.
- Soltani, Z. (2018). The impact of the customer relationship management on the organization performance. *Journal of High Technology Management Research*.
- Tebes, G. (2019). Especificación del Proceso de Design Science Research: Caso Aplicado a una Ontología de Testing de Software. *7mo Congreso Nacional de Ingeniería Informática*.
- TELÉGRAFO. (2019). Pequeñas y grandes empresas se enfrentan a tres problemas. *EL TELEGRAFO*.

- Thompson, I. (Abril de 2007). *Concepto de Empresa y como aplicarlo*. Obtenido de <https://www.promonegocios.net/empresa/concepto-empresa.html>
- Tornero, J. (2017). *Machine Learning: Modelos Ocultos de Markov (HMM) y Redes Neuronales Artificiales (ANN)*. *Universitat de Barcelona*.
- Wyderka, A. (Octubre de 2018). *Sistema Económico Latinoamericano y del Caribe*. Obtenido de <http://www.sela.org/media/3205340/las-cuentas-por-cobrar-fci.pdf>
- Yun, J. J. (2016). The relationship between technology, business model, and market in autonomous car and intelligent robot industries. *Technological Forecasting & Social Change*.