



Escuela de Posgrados

**Identificación de patrones asociados a la rotación laboral en el Grupo Éxito  
mediante aprendizaje no supervisado**

Esteban Restrepo Zapata

Yeison Andrés Vargas Ángel

Trabajo de Grado presentado como requisito para optar al título de:

Especialista en Big Data e Inteligencia de Negocios

Asesor: Juan Sebastián Parra Sánchez

Universidad Católica Luis Amigó

Facultad de Ingenierías y Arquitectura

Especialización en Big Data e Inteligencia de

Negocios

2023

## **Dedicatoria**

Dedicamos este proyecto a nuestras familias, quienes siempre han estado a nuestro lado, apoyándonos y motivándonos en todo momento. Su amor y aliento han sido la clave de nuestro éxito.

A nuestras parejas, quienes han sido nuestra fuente de apoyo, motivación y amor incondicional durante todo el proceso de realización de este trabajo. Gracias por estar a nuestro lado en las buenas y en las malas, por ser nuestra roca y confidentes.

## **Agradecimientos**

Nosotros, Esteban Restrepo y Yeison Vargas, queremos expresar nuestra gratitud por la oportunidad de presentar nuestro proyecto de trabajo de grado.

Este proyecto ha sido el resultado de meses de arduo trabajo y dedicación, y no habría sido posible sin el apoyo y la guía de muchas personas que nos ayudaron en cada etapa del proceso.

En primer lugar, nos gustaría agradecer a nuestros asesores, y profesores, por su invaluable orientación y asesoramiento a lo largo del proyecto y la especialización. Su experiencia y conocimiento nos permitieron avanzar en nuestro trabajo de manera eficiente y con confianza.

También queremos agradecer a nuestras parejas y padres por su constante apoyo y motivación durante este tiempo. Sus palabras de aliento y sus mensajes de ánimo nos mantuvieron enfocados y motivados cuando nos sentíamos abrumados.

Finalmente, estamos profundamente agradecidos por la oportunidad de presentar nuestro trabajo de grado, y por el apoyo y la guía de todas las personas que nos ayudaron en el camino. Esperamos que nuestro trabajo sea útil y valioso para la comunidad académica, y estamos emocionados por las oportunidades futuras que se nos presentan.

## Resumen

El presente trabajo de grados se centra en la identificación de los patrones asociados a la rotación de personal en el Grupo Éxito, una de las compañías más grandes de *retail* en Colombia, a través del uso de técnicas de aprendizaje no supervisado. La rotación de personal representa un desafío significativo para las organizaciones, debido a sus costos económicos y su impacto en la estabilidad y productividad del equipo de trabajo.

Con el objetivo de abordar esta problemática, se plantean varios objetivos específicos. En primer lugar, se busca integrar y preparar la información necesaria para el estudio de la rotación de personal en el Grupo Éxito. A continuación, se emplearán técnicas de aprendizaje no supervisado para detectar patrones o perfiles asociados con la rotación de personal en la organización. Por último, se evaluará y desplegará el modelo de aprendizaje no supervisado desarrollado, con el fin de determinar los factores que influyen en la rotación de personal.

La aplicación de técnicas de aprendizaje no supervisado permite descubrir patrones y estructuras ocultas en los datos sin la necesidad de etiquetas previas. Esto proporciona una oportunidad única para explorar las características subyacentes que podrían estar influyendo en la rotación de personal en el Grupo Éxito.

Los resultados obtenidos en este estudio ofrecerán información valiosa para comprender los factores que influyen en la rotación de personal y contribuirán a la implementación de estrategias efectivas de retención de talento en la organización.

## Tabla de contenido

<b>1. Introducción</b>	<b>9</b>
<b>2. Planteamiento del Problema</b>	<b>11</b>
<b>3. Justificación</b>	<b>13</b>
<b>4. Marco de Referencias</b>	<b>14</b>
<b>5. Antecedentes</b>	<b>17</b>
<b>6. Objetivos</b>	<b>20</b>
6.1. Objetivo General	20
6.2. Objetivos Específicos	20
<b>7. Viabilidad</b>	<b>21</b>
<b>8. Metodología</b>	<b>22</b>
8.1. Preparación de los datos	23
8.2. Modelado	24
8.3. Evaluación	24
8.4. Despliegue	24
<b>9. Resultados</b>	<b>26</b>
9.1. Entendimiento del negocio	26
9.2. Comprensión de los datos	26
9.2.1. Descripción de variables	27
9.2.2. Análisis exploratorio	30
9.2.3. Variables irrelevantes	33
9.2.4. Identificación de nulos, datos atípicos y duplicados	34
9.3. Preparación de los datos	39
9.3.1. Transformación de datos	39
9.3.2. Renombramiento de variables	48
9.3.3. Generación de variables Dummies	48
9.3.4. Normalización de los Datos	49
9.4. Modelamiento	50
9.4.1. Identificación del valor de K	50
9.5. Implementación del Modelo	54
9.7. Evaluación	56
9.7.1. K-means	56
<b>10. Recomendaciones</b>	<b>69</b>
<b>11. Referencias</b>	<b>71</b>

## Lista de Graficas

<b>Figura 1 Fases de la Metodología CRISP-DM .....</b>	<b>23</b>
<b>Figura 2 Distribución de Géneros en los Retiros 2022.....</b>	<b>31</b>
<b>Figura 3 Distribución de Rango de edades en Retiros en el 2022.....</b>	<b>31</b>
<b>Figura 4 Distribución de tipos de contratos en Retiros en el 2022 .....</b>	<b>32</b>
<b>Figura 5 Distribución de cargos en Retiros en el 2022 .....</b>	<b>32</b>
<b>Figura 6 Conjunto de datos.....</b>	<b>35</b>
<b>Figura 7 Gráfica de tipo de contrato.....</b>	<b>36</b>
<b>Figura 9 Gráfica de tipo de contrato sin aprendiz .....</b>	<b>37</b>
<b>Figura 10 Validación de datos nulos .....</b>	<b>38</b>
<b>Figura 11 Validación sin datos nulos .....</b>	<b>38</b>
<b>Figura 12 Agrupación de causas de retiros .....</b>	<b>39</b>
<b>Figura 13 Eliminación de la variable.....</b>	<b>40</b>
<b>Figura 14 Grafica tipo de retiros .....</b>	<b>40</b>
<b>Figura 15 Cálculo de días laborados .....</b>	<b>41</b>
<b>Figura 16 Eliminación de fechas.....</b>	<b>41</b>
<b>Figura 17 Creación de rangos laborales .....</b>	<b>42</b>
<b>Figura 18 Grafica de rango laboral .....</b>	<b>43</b>
<b>Figura 19 Cálculo de rango de edades .....</b>	<b>44</b>
<b>Figura 20 Gráfica de rango de edades .....</b>	<b>45</b>
<b>Figura 21 Agrupación de variables categóricas .....</b>	<b>45</b>
<b>Figura 22 Validación de datos por sociedades.....</b>	<b>46</b>
<b>Figura 23 Eliminación de datos de sociedades .....</b>	<b>46</b>
<b>Figura 24 Validación de datos de agrupación de cargos.....</b>	<b>47</b>
<b>Figura 25 Eliminación de datos de agrupación de cargos.....</b>	<b>47</b>
<b>Figura 26 Reemplazo de datos de 'Cargo Líder' .....</b>	<b>47</b>

<b>Figura 27 Renombramiento de variables .....</b>	<b>48</b>
<b>Figura 28 Generación de variables Dummies .....</b>	<b>49</b>
<b>Figura 29 Normalización de los Datos.....</b>	<b>50</b>
<b>Figura 30 Identificación de clústeres .....</b>	<b>51</b>
<b>Figura 31 Método de Silhouette .....</b>	<b>52</b>
<b>Figura 32 Grafica de varianza .....</b>	<b>53</b>
<b>Figura 33 Análisis método Silhouette .....</b>	<b>54</b>
<b>Figura 34 Visualización de Clústeres .....</b>	<b>55</b>
<b>Figura 35 Código DBSCAN .....</b>	<b>56</b>

## Lista de tablas

<b>Tabla 1 Descripción de variables.....</b>	<b>27</b>
<b>Tabla 2 Descripción de Clúster 0.....</b>	<b>56</b>
<b>Tabla 3 Descripción de Clúster 1.....</b>	<b>57</b>
<b>Tabla 4 Descripción de Clúster 2.....</b>	<b>59</b>
<b>Tabla 5 Descripción de Clúster 3.....</b>	<b>60</b>
<b>Tabla 6 Descripción de Clúster 4.....</b>	<b>61</b>
<b>Tabla 7 Descripción de Clúster 5.....</b>	<b>62</b>
<b>Tabla 8 Descripción de Clúster 6.....</b>	<b>63</b>
<b>Tabla 9 Descripción de Clúster 7.....</b>	<b>64</b>
<b>Tabla 10 Descripción de Clúster 8.....</b>	<b>66</b>
<b>Tabla 11 Descripción de Clúster 9.....</b>	<b>67</b>

## 1. Introducción

En el entorno empresarial actual, la rotación de personal representa un desafío significativo para las organizaciones, ya que conlleva costos económicos, pérdida de conocimiento y experiencia, así como un impacto en la moral y la estabilidad del equipo de trabajo. Para abordar este problema, es fundamental comprender los factores que influyen en la rotación de personal y buscar patrones o perfiles que puedan ayudar a predecir y mitigar este fenómeno.

En el caso particular del Grupo Éxito, una de las principales compañías más grandes del *retail* en Colombia, la rotación de personal es un aspecto crítico que requiere una atención especial. El Grupo Éxito se enfrenta a diversos desafíos relacionados con la retención de talento, debido a la naturaleza dinámica y competitiva del sector *retail*, así como a la constante evolución del mercado laboral.

En este contexto, el presente trabajo de grados de especialización tiene como objetivo general identificar los patrones asociados a la rotación de personal en el Grupo Éxito, utilizando técnicas de aprendizaje no supervisado. El aprendizaje no supervisado es una rama del aprendizaje automático que permite descubrir estructuras ocultas o patrones inherentes en los datos sin la necesidad de contar con etiquetas previas. Esta metodología ofrece una oportunidad única para explorar y comprender las características subyacentes que pueden estar influyendo en la rotación de personal en la organización.

Para lograr este objetivo general, se plantean los siguientes objetivos específicos: en primer lugar, integrar y preparar la información requerida para el estudio de la rotación de personal en el Grupo Éxito. A continuación, se buscarán detectar patrones o perfiles asociados con la rotación del personal en la organización, utilizando técnicas de aprendizaje no supervisado. Por último, se evaluará y desplegará el modelo de aprendizaje no supervisado desarrollado, con el fin de determinar los factores que influyen en la rotación de personal y

proporcionar información relevante para la toma de decisiones estratégicas en el Grupo Éxito.

Este trabajo representa un aporte significativo para el Grupo Éxito, ya que permitirá obtener conocimientos profundos sobre los factores que influyen en la rotación de personal, lo que a su vez contribuirá a la implementación de estrategias efectivas de retención de talento y mejora del clima laboral. Asimismo, los resultados obtenidos pueden ser extrapolables a otras organizaciones del sector *retail* y servir como base para futuras investigaciones relacionadas con la gestión del talento humano en un entorno empresarial dinámico y competitivo.

En resumen, este trabajo busca aportar conocimientos valiosos sobre la rotación de personal en el Grupo Éxito, mediante la aplicación de técnicas de aprendizaje no supervisado. La identificación de patrones asociados a la rotación de personal permitirá comprender mejor los factores que influyen en este fenómeno y contribuirá a la implementación de acciones estratégicas para su mitigación. A continuación, se detallarán la metodología utilizada en este estudio, así como los resultados y recomendaciones obtenidas.

## 2. Planteamiento del Problema

La rotación de personal es uno de los problemas más comunes en el mundo empresarial y puede tener un impacto significativo en la productividad y el éxito de una organización. En este sentido, el Grupo Éxito, una de las principales compañías del *retail* en Colombia no es una excepción. La rotación de personal en el Grupo Éxito ha sido un desafío importante, lo que ha llevado a la necesidad de investigar los patrones y las causas subyacentes que la provocan.

La rotación laboral puede tener un impacto significativo en la eficiencia, la productividad y los costos de la empresa y puede ser causada por una variedad de factores, incluyendo el ambiente laboral, la remuneración, la falta de oportunidades de crecimiento y desarrollo, entre otros.

La vicepresidencia de Recursos Humanos del Grupo Éxito ha entendido la rotación de personal como el índice o la tasa de rotación refiriéndose al número de empleados que abandonan la empresa durante un periodo de tiempo determinado en relación con el número total de empleados, dicha rotación se puede dar de forma voluntaria o involuntaria, entendiendo la voluntaria cuando los empleados deciden dejar la compañía por diversas razones, las cuales al día de hoy el grupo éxito está realizando un esfuerzo importante para tener una identificación clara e impartir acciones que ayuden a mejorar la deserción, por otra parte, la involuntaria ocurre cuando los empleados son despedidos o desvinculados de la compañía debido a razones de bajo desempeño, reducción de costos o reorganización.

En el año 2022 se registró un índice de rotación de personal del 14.7%, lo que traduce en un impacto considerable para la compañía debido a que afecta directamente la productividad y rentabilidad de la empresa, cuando un empleado abandona la compañía existe una pérdida de conocimiento, habilidades y experiencia que afectan negativamente la productividad en los diferentes procesos, eso sin contar el impacto moral y motivacional de los empleados que permanecen en la empresa, lo que puede afectar aún más la productividad.

Cuando existe un índice elevado en rotación de personal incrementa de manera proporcional la inversión que se debe realizar en reclutamiento, selección y capacitación de nuevos empleados.

En este sentido, surge la necesidad de identificar los patrones asociados a la rotación de personal en el Grupo Éxito, con el objetivo de entender las causas y desarrollar estrategias efectivas para mitigarla. En particular, se plantea la siguiente pregunta de investigación:

¿Cuáles son los patrones asociados a la rotación de personal en el Grupo Éxito, y cómo pueden ser identificados mediante técnicas de aprendizaje no supervisado?

Para responder a esta pregunta, se utilizarán técnicas de aprendizaje no supervisado sobre un conjunto de datos que incluye información de los empleados que han abandonado la compañía. El análisis se centrará en la identificación de patrones que permitan comprender los factores que influyen en la decisión de los empleados de abandonar la organización.

### **3. Justificación**

La rotación de personal es un problema que afecta a todas las organizaciones, independientemente de su tamaño o sector. Sin embargo, en el caso del Grupo Éxito, la rotación laboral se ha convertido en un desafío significativo, lo que ha llevado a la necesidad de investigar los factores subyacentes que la provocan.

La identificación de patrones asociados a la rotación de personal en el Grupo Éxito es crucial para el desarrollo de estrategias de retención de talentos y para la mejora de la productividad de la compañía. En este sentido, el aprendizaje no supervisado se ha convertido en una herramienta valiosa para analizar grandes conjuntos de datos y encontrar patrones ocultos.

El análisis de los patrones de rotación de personal en el Grupo Éxito permitirá entender mejor los factores que influyen en la decisión de los empleados de abandonar la organización y, por lo tanto, desarrollar políticas de recursos humanos más efectivas. Además, el trabajo permitirá a la empresa tomar decisiones basadas en datos y mejorar la eficiencia en la toma de decisiones.

Por lo tanto, la identificación de patrones asociados a la rotación de personal en el Grupo Éxito mediante técnicas de aprendizaje no supervisado es un tema de investigación relevante y de gran importancia práctica, que puede ayudar a la empresa a mejorar la retención de talentos y la productividad, y a enfrentar los desafíos en un mercado cada vez más competitivo.

#### 4. Marco de Referencias

El marco referencial de una investigación consiste en una compilación breve y precisa de conceptos, teorías y reglamentos que están directamente ligados con el tema y el problema de la investigación. Para el caso concreto de esta investigación, los constructos conceptuales que se utilizarán son los siguientes:

- **Rotación de personal:** “La rotación de personal define la relación entre el número de empleados que ingresan y los que dejan de trabajar en la organización. La tasa de rotación de personal puede ser un indicador importante en la estabilidad del personal.” (Treviño, 2000, pág. 227)

- **Recursos Humanos:** (Dessler, 2001) define a la Administración de Recursos Humanos como “las políticas y las prácticas que se requieren para llevar a cabo los aspectos relativos a las personas o al personal del puesto administrativo que se ocupa”.

- **Reclutamiento:** “El reclutamiento es un conjunto de técnicas y procedimientos orientados a atraer candidatos potencialmente calificados y capaces de ocupar cargos dentro de la organización” (Chiavenato, Administración de Recursos Humanos, 2001, pág. 208).

- **Contratación:** De acuerdo con (Alles, 2006), es una actividad de clasificación donde se escoge a aquellos que presentan mayor posibilidad de adaptarse al cargo ofrecido para satisfacer las necesidades de la organización. Los candidatos pueden ser personas desempleadas o, por el contrario, con empleo.

- **Atracción:** Es una actividad que se lleva a cabo desde el departamento de recursos humanos, aunque pueda estar enmarcado dentro de las funciones del marketing interno (Blasco López et al., 2014) y va dirigida tanto a los empleados actuales de una compañía como a los empleados potenciales para que la marca les resulte

atractiva y poder además retener el talento que atraen (Edwards, 2010) Su fin es alinear, motivar y otorgar de poder a sus empleados para que sus vivencias y experiencias en el entorno laboral sea no sólo positiva sino mejor que en las empresas competidoras y así conseguir con mayor facilidad los objetivos de la empresa (Asif & Sargeant, 2000).

- **Selección:** La selección o entrevista por competencias es un proceso que debe abarcarse por etapas sucesivas. Si se divide en distintas fases, hay más probabilidad de aproximación al candidato idóneo para la cobertura de una vacante (Alles, 2003).

- **Capacitación:** Según Chiavenato (2008) “La capacitación es el proceso de desarrollar cualidades en los recursos humanos, preparándolos para que sean más productivos y contribuyan mejor al logro de los objetivos de la organización” (p. 386).

- **Big Data:** volúmenes de datos disponibles en diversos grados de complejidad, generados a diferentes velocidades y diversos grados de ambigüedad, que no pueden procesarse utilizando tecnologías tradicionales, métodos de procesamiento, algoritmos o cualquier solución comercial lista para usar (Krishnan, 2013).

- **People Analytics:** *People Analytics* tiene el potencial de transformar la forma en que las organizaciones identifican, desarrollan, gestionan y controlan su fuerza laboral (Chamorro-Premuzic et al., 2017; Huselid, 2018).

- **Machine Learning:** Es el área de la inteligencia artificial que engloba un conjunto de técnicas que hacen posible el aprendizaje automático a través del entrenamiento con grandes volúmenes de datos (Russo et al., 2016).

- **Minería de datos:** Permitirá estudiar métodos y algoritmos para la extracción automática de información sintetizada sobre deserción escolar que permita caracterizar las relaciones escondidas en la gran cantidad de datos que se tienen (Beltrán Martínez, 2001).

- **Metodología CRISP DM:** Esta metodología integra todas las tareas necesarias para el desarrollo de este estudio, desde la fase inicial de comprensión del problema hasta el despliegue del modelo predictivo propuesto como solución a través del uso de técnicas de *Machine Learning*, incluye descripciones de las fases normales de un proyecto, las tareas necesarias en cada fase y una explicación de las relaciones entre las tareas. (IBM, 2021)
- **Google Colab:** Es un documento ejecutable que permitirá escribir, ejecutar y compartir código arbitrario de *Python* en el navegador, para realizar las tareas de aprendizaje automático y análisis de los datos disponibles para este trabajo. (Agrega, s.f.)
- **Python:** Lenguaje de programación utilizado para el desarrollo de este trabajo, ampliamente utilizado en la ciencia de datos y *Machine Learning* (ML). Es eficiente y fácil de aprender, además se puede ejecutar en muchas plataformas diferentes. (AWS, s.f.)

De acuerdo con las definiciones presentadas anteriormente se identifican que los siguientes términos son de gran importancia para el proyecto dado que permiten identificar conceptualmente la necesidad organizacional y permiten tener una proyección de solución ante las tecnologías que en la actualidad están surgiendo para ayudar a la identificación de los factores que más inciden en la deserción de personal en las compañías y de esta forma concentrar esfuerzos en estrategias que permitan la reducción del índice de rotación de empleados.

## 5. Antecedentes

En el estudio realizado por (Liu, Li, Zhang, 2017), se utilizó un conjunto de datos que incluía información de 200 empleados de una empresa, como su edad, antigüedad, salario, educación, estado civil, entre otros. Los autores aplicaron la técnica de *clustering* para agrupar a los empleados en diferentes grupos según sus características y comportamientos relacionados con el retiro. A continuación, utilizaron la técnica de árboles de decisión para identificar los factores que influían en el retiro de los empleados.

Los resultados del estudio mostraron que la técnica de *clustering* permitió identificar diferentes grupos de empleados con características similares en términos de retiro, como los que se retiraron a una edad temprana y los que se retiraron después de una larga carrera. Además, los árboles de decisión identificaron los factores más importantes que influyen en la decisión de retirarse, como la edad, el salario y el estado civil.

Según el artículo de Krause, Seymour, & Pandya, (2019), En un trabajo de investigación que utiliza técnicas de aprendizaje automático para predecir la rotación de los empleados y para identificar los factores que contribuyen a la rotación.

En este estudio, los autores utilizaron un conjunto de datos que incluía información de empleados de una empresa, como su salario, antigüedad, evaluaciones de desempeño, entre otros. Aplicaron diferentes algoritmos de aprendizaje automático, incluyendo regresión logística, árboles de decisión, SVM (máquinas de vectores de soporte) y redes neuronales, para predecir la rotación de los empleados.

Los resultados del estudio mostraron que los modelos de aprendizaje automático pudieron predecir la rotación de los empleados con una precisión del 80-90%. Además, los modelos identificaron los factores más importantes que contribuyen a la rotación, como la satisfacción laboral, el compromiso, la remuneración, entre otros.

Sodhi, & Sonnenburg, (2017). utilizan técnicas de aprendizaje automático para identificar y comprender los factores que contribuyen a la rotación de los empleados.

En este estudio, los autores utilizaron un conjunto de datos de una empresa que incluía información sobre los empleados, como su antigüedad, salario, calificaciones de desempeño, entre otros. Utilizaron técnicas de aprendizaje automático, específicamente bosques aleatorios y árboles de decisión, para analizar los datos y predecir la rotación de los empleados.

Los resultados del estudio mostraron que los modelos de aprendizaje automático pudieron predecir la rotación de los empleados con una precisión del 86-92%. Además, los modelos identificaron los factores más importantes que contribuyen a la rotación, como la satisfacción laboral, la equidad salarial y el equilibrio entre la vida laboral y personal.

Chakraborty, & Chatterjee, (2018) utilizan técnicas de aprendizaje automático para predecir la rotación de los empleados en la industria de la tecnología de la información en India.

En este estudio, los autores utilizaron un conjunto de datos que incluía información de empleados de una empresa de TI en India, como su edad, género, antigüedad, salario, desempeño, entre otros. Utilizaron diferentes técnicas de aprendizaje automático, incluyendo SVM (máquinas de vectores de soporte), árboles de decisión y regresión logística, para predecir la rotación de los empleados.

Los resultados del estudio mostraron que los modelos de aprendizaje automático pudieron predecir la rotación de los empleados con una precisión del 80-90%. Además, los modelos identificaron los factores más importantes que contribuyen a la rotación, como la satisfacción laboral, la equidad salarial y el equilibrio entre la vida laboral y personal.

Los estudios previamente mencionados comparten la importante tarea de identificar los factores que están directamente relacionados con los retiros de las compañías. Sin embargo, a diferencia del presente trabajo, estos estudios también se enfocaron en predecir los retiros de los empleados mediante el uso de técnicas de aprendizaje automático. En este trabajo, por el contrario, se limitará a la identificación de los factores de retiro, sin incluir una predicción específica.

## **6. Objetivos**

### **6.1. Objetivo General**

Identificar los patrones asociados a la rotación de personal en el Grupo Éxito, a través del uso de técnicas de aprendizaje no supervisado.

### **6.2. Objetivos Específicos**

- Integrar y preparar la información requerida en el estudio de la rotación de personal en el Grupo Éxito.
- Detectar patrones o perfiles asociados con la rotación del personal en el grupo éxito utilizando técnicas de aprendizaje no supervisado.
- Evaluar y desplegar el modelo de aprendizaje no supervisado para determinar los factores que influyen en la rotación de personal.

## 7. Viabilidad

Basados en el objetivo principal del proyecto “Identificar los patrones asociados a la rotación de personal en el Grupo Éxito, a través del uso de técnicas de aprendizaje no supervisado”, los aspectos que ayudaron a tener éxito en este son:

- **Recopilación de datos:** La información que se utilizara en este caso, está relacionado con campos como el tiempo de trabajo, la antigüedad, el puesto que ocupaba, genero, edad, y demás información referente a su estructura organizacional. El segmento de datos corresponde la información recolectada de retiros realizados en el año 2022 la cual tiene aproximadamente 19.000 registros, esta información será extraída del ERP de recursos humana (*SAP SuccessFactors*) conocido internamente como RedHumana.
- **Selección del algoritmo de aprendizaje no supervisado:** existen varios algoritmos de aprendizaje no supervisado, como el de clústeres o la reducción de dimensionalidad. Para los objetivos del proyecto se utilizará (*KMeans*) y *DBSCAN* los cuales permitirán cumplir con los objetivos propuestos.
- **Análisis de resultados:** En este punto se llevará a cabo un análisis detallado de los resultados obtenidos para determinar la relevancia de los patrones identificados. Los patrones deben ser coherentes con los objetivos del proyecto y deben tener sentido desde un punto de vista práctico.

## 8. Metodología

Para el desarrollo de este proyecto y de todas las actividades propuestas, se inició con el proceso de extracción de datos desde el ERP de Recursos humanos del Grupo Éxito el cual es *SAP SuccessFactors*, conocido internamente como RedHumana. En la información extraída desde el ERP se logró identificar la base de datos de los retiros de empleados, con una muestra de aproximada de 19.000 registros, durante el año 2022 en el Grupo Éxito, con sus respectivas variables principales como “ID de usuario”, “Fecha de ingreso”, “Fecha de terminación de contrato”, “Motivo de terminación”, entre otros.

Los permisos fueron gestionados con la Líder del equipo de Herramientas y Servicios y a su vez con el Director de Organización y soluciones, con la salvedad de que dicho proyecto debe ser presentado a ellos y si es viable implementarlo en el Grupo Éxito, sobre los datos de los empleados no se utilizaron variables que identifiquen quién es el empleado, es por esto que no fue necesario utilizar autorización para el manejo de información sensible, sin embargo, todos los empleados al momento de firmar un contrato con la organización autorizan el uso de datos.

Utilizando la minería de datos, se llevó a cabo el ciclo de vida, el cual es un proceso iterativo que comprende la preparación, el modelado y la evaluación de los datos. Posteriormente, se procedió a identificar las variables presentes en la base de datos, encontrando entre ellas variables cualitativas (numéricas), cuantitativas (categóricas), cadenas de caracteres y fechas. Finalmente, se realizó la preparación de los datos con el objetivo de lograr un análisis y una modelización más sólidos. Para ello, se emplearon los 8 pasos mínimos estipulados en la metodología CRISP-DM (*Cross-Industry Standard Process for Data Mining*).

**Figura 1.** Fases de la Metodología CRISP-DM



Nota: Este es el ejemplo de las Fases de la Metodología *CRISP-DM*, Daniel Álvarez Gil, 2021, Adictos al Trabajo.

El objetivo y alcance del proyecto de grado estuvo enfocado a la aplicación de técnicas no supervisadas (*KMeans*) y *DBSCAN* teniendo en cuenta que lo que se pretende identificar son los patrones asociados a la rotación laboral en los empleados que abandonan la compañía. Con ayuda del análisis descriptivo, se logró identificar los perfiles de los empleados de la compañía, la selección de factores y detección de anomalías que se presentaron en la implementación del modelo seleccionado para el alcance del proyecto.

La implementación de los modelos anteriormente descritos se llevó a cabo mediante la metodología *CRISP - DM* (*Cross- Industry Standard Process for Data Mining*), que es una metodología en la que sus componentes giran alrededor del foco principal del estudio que son los datos y tiene los siguientes pasos para su implementación:

### **8.1. Preparación de los datos**

Durante esta fase, se llevó a cabo una serie de tareas fundamentales para garantizar la

calidad y la coherencia de la información que sería utilizada en este análisis. Entre estas actividades se incluyó la identificación y manejo de nulos, datos atípicos y duplicados. Además de estas tareas de limpieza y corrección de datos, también se realizó transformaciones de estos para facilitar su interpretación y análisis. Estas transformaciones incluyeron la normalización de variables, escalando o ajustando los datos a una escala común, y el renombramiento de variables, asignando nombres más descriptivos y claros que reflejaran su contenido o significado.

Por último, enriquecimos este conjunto de datos mediante la generación de variables *Dummies*.

## **8.2. Modelado**

Durante la fase de modelado, se empleó técnicas de minería de datos para descubrir de manera automática información de valor en los datos. Esta etapa implica la utilización de algoritmos y herramientas especializadas para explorar, analizar y extraer patrones, tendencias y relaciones ocultas en los datos.

Esta fase de modelado es crucial para descubrir información valiosa y oculta en los datos utilizando técnicas de minería de datos. Implica la selección, construcción, evaluación y validación de modelos, así como la interpretación de los resultados obtenidos.

## **8.3. Evaluación**

Esta fase de evaluación fue enfocada en analizar y examinar los resultados obtenidos durante la etapa de modelado. Esta etapa es esencial para determinar la calidad y la eficacia del modelo construido, así como para evaluar su rendimiento y su capacidad para cumplir con los objetivos establecidos.

## **8.4. Despliegue**

Esta fase de despliegue fue enfocada en distribuir de manera efectiva los análisis de resultados obtenidos durante el proceso de modelado. El objetivo principal es aprovechar estos resultados para mejorar los procesos del negocio y tomar decisiones informadas.

Adicionalmente, para la implementación de los modelos con los que se desarrollarán los

objetivos propuestos y con los que se busca responder la pregunta de investigación, se utilizaron herramientas como *PowerBI*, *Google Colab* y *Python*, las cuales son utilizadas para el aprendizaje automático y la minería de datos.

## **9. Resultados**

### **9.1. Entendimiento del negocio**

En esta fase de la metodología se realizó un entendimiento detallado del negocio, para lo que inicialmente se identificó que el grupo éxito es una empresa colombiana dedicada al sector minorista y de distribución de alimentos y productos para el hogar. Es considerada la cadena de supermercados más grande del país y una de las más importantes en toda América Latina.

El grupo éxito al igual que una gran cantidad de empresas en el país sufren por el alto índice de rotación de personal a tal medida que sus procesos productivos se están viendo afectados, dado que los conocimientos, experiencia y habilidades obtenidas en cada cargo se están yendo con las personas que deciden abandonar la compañía por diferentes motivos, esta deserción impacta en la productividad directamente, dado que ante el reemplazo de un cargo sobre un proceso se debe garantizar una curva de aprendizaje que en ciertas ocasiones es costoso para la compañía en términos de productividad. Además de verse afectada la productividad, la rentabilidad de la compañía con el índice alto de rotación se ve afectada, dado que se generan nuevos gastos en procesos de selección, reclutamiento, capacitación que se incrementa de forma proporcional al aumentar el índice de rotación, por consiguiente, para el grupo éxito es importante poder identificar los factores que están generando la deserción de sus empleados, de esta forma podrán desplegar estrategias que ayuden a disminuir la rotación de empleados.

### **9.2. Comprensión de los datos**

Entendiendo la problemática actual que presenta el Grupo Éxito, se decidió profundizar sobre la información de retiros que se generaron para la compañía durante el año 2022, en dicha información se realizó la identificación de 45 variables y un total de 19.294 registros con los cuales se realizó el procesamiento de la información dado las técnicas no supervisadas.

### 9.2.1. Descripción de variables

Descripción detallada de cada variable utilizada en el estudio, basándose en la base de datos que se empleó. Esta tabla se creó con el propósito de tener una referencia clara y concisa de las características y el significado de cada variable en el contexto del estudio realizado.

**Tabla 1. Descripción de variables.**

Nombre Variable	Tipo de variable	Descripción
ID de usuario	Cadena de caracteres.	Identificador único del usuario.
Número de identificación	Cadena de caracteres.	Número de identificación del empleado.
Nombre formal	Cadena de caracteres.	Nombre completo del empleado.
Fecha de ingreso	Fecha.	Fecha en la que el empleado comenzó a trabajar.
Fecha de terminación de contrato	Fecha.	Fecha en la que el contrato del empleado finalizó.
Lugar de trabajo (Lugar de trabajo)	Cadena de caracteres.	Nombre del lugar de trabajo donde el empleado está ubicado.
<i>Event Reason Icode</i>	Cadena de caracteres.	Código del evento que causó el cambio en el estado del empleado.
<i>Event Reason Icode (Label)</i>	Cadena de caracteres.	Descripción del código del evento que causó el cambio en el estado del empleado.
Régimen Laboral ( <i>Picklist Label</i> )	Cadena de caracteres.	Tipo de régimen laboral del empleado.
Evento ( <i>Picklist Label</i> )	Cadena de caracteres.	Tipo de evento que causó el cambio en el estado del empleado.

Tipo de contrato ( <i>Picklist Label</i> )	Cadena de caracteres.	Tipo de contrato del empleado.
Zona (Zona)	Cadena de caracteres.	Nombre de la zona donde el empleado está ubicado.
Número de teléfono	Cadena de caracteres.	Número de teléfono del empleado.
Dirección de correo electrónico	Cadena de caracteres.	Dirección de correo electrónico del empleado.
Empleado que realiza la baja	Cadena de caracteres.	Código del empleado que realizó la baja del empleado en cuestión.
Empleado que realiza la baja (Primer nombre)	Cadena de caracteres.	Primer nombre del empleado que realizó la baja del empleado en cuestión.
Empleado que realiza la baja (Primer apellido)	Cadena de caracteres.	Primer apellido del empleado que realizó la baja del empleado en cuestión.
Generación ( <i>Picklist Label</i> )	Cadena de caracteres.	Generación a la que pertenece el empleado.
Número de días del contrato	Entero.	Número de días que dura el contrato del empleado.
Sociedad ( <i>Label</i> )	Cadena de caracteres.	Nombre de la sociedad a la que pertenece el empleado.
Formato ( <i>Picklist Label</i> )	Cadena de caracteres.	Formato del lugar de trabajo del empleado.
Marca ( <i>Picklist Label</i> )	Cadena de	Marca a la que pertenece el

	caracteres.	empleado.
Ciudad de Notificación <i>(Picklist Label)</i>	Cadena de caracteres.	Ciudad del lugar de trabajo del empleado.
Código de lugar de trabajo	Cadena de caracteres.	Código único del lugar de trabajo donde el empleado está ubicado.
Lugar de trabajo	Cadena de caracteres.	Nombre del lugar de trabajo donde el empleado está ubicado.
Departamento <i>(Label)</i>	Cadena de caracteres.	Nombre del departamento al que pertenece el empleado.
Departamento 3 (principal) <i>(Label)</i>	Cadena de caracteres.	Nombre del departamento principal al que pertenece el empleado.
Departamento 3 (secundario) <i>(Label)</i>	Cadena de caracteres.	Nombre del departamento secundario al que pertenece el empleado.
Departamento secundario <i>(Label)</i>	Cadena de caracteres.	Nombre del departamento secundario al que pertenece el empleado.
División <i>(Label)</i>	Cadena de caracteres.	Nombre de la división a la que pertenece el empleado.
División secundaria <i>(Label)</i>	Cadena de caracteres.	Nombre de la división secundaria a la que pertenece el empleado.
Horario laboral <i>(Label)</i>	Cadena de caracteres.	Horario laboral del empleado.
Unidad de negocio <i>(Label)</i>	Cadena de caracteres.	Nombre de la unidad de negocio a la que pertenece el empleado.

ETC - Empleado a tiempo completo	Cadena de caracteres.	Indica si el empleado es a tiempo completo o no.
Agrupación ( <i>Picklist Label</i> )	Cadena de caracteres.	Agrupación a la que pertenece el cargo del empleado.
Cargo líder	Cadena de caracteres.	Indica si el empleado es un líder o no.
Código de puesto	Cadena de caracteres.	Código único del cargo del empleado.
Función de puesto ( <i>Label</i> )	Cadena de caracteres.	Función a la que pertenece el cargo del empleado.
Nivel del puesto ( <i>Picklist Label</i> )	Cadena de caracteres.	Nivel del cargo al que pertenece el empleado
Título de puesto ( <i>Label</i> )	Cadena de caracteres.	Nombre del cargo del empleado.
Centro de costos (ID Centro de Costo)	Cadena de caracteres.	Código del centro de costo al que pertenece el cargo del empleado.
Centro de costos ( <i>Label</i> )	Cadena de caracteres.	Nombre del centro de costo al que pertenece el cargo del empleado.
Fecha de nacimiento	Fecha.	Fecha de nacimiento del empleado
Edad Calculada	Entero.	Edad del empleado.
Género	Cadena de caracteres.	Genero del empleado.

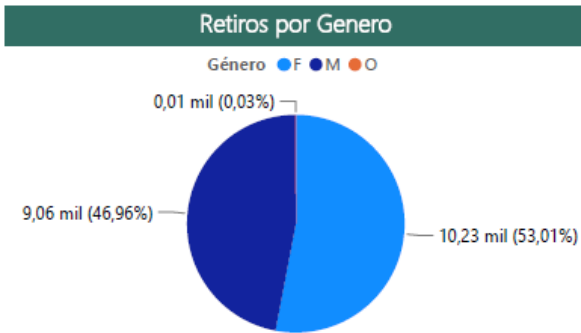
### 9.2.2. Análisis exploratorio

Se realizó un análisis exploratorio de los datos utilizando la herramienta *POWER BI* la cual permitió identificar la forma como está distribuida la información, adicionalmente ayudó a

determinar los valores atípicos, problemas en la calidad de datos y de esta forma tener una claridad para emplear las diferentes técnicas en la preparación de datos.

**Figura 2**

*Distribución de Géneros en los Retiros 2022.*

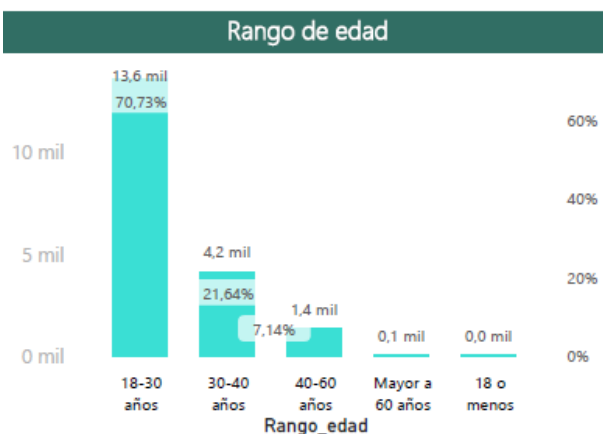


En esta Figura se evidenció la distribución que existe a nivel de género en los retiros generados en el grupo Éxito para el año 2022, dado los datos se logró identificar que el mayor porcentaje de empleados retirados para el año 2022 corresponde a Mujeres.

Nota: Grafica de distribución de Géneros en los Retiros 2022.

**Figura 3**

*Distribución de Rango de edades en Retiros en el 2022.*

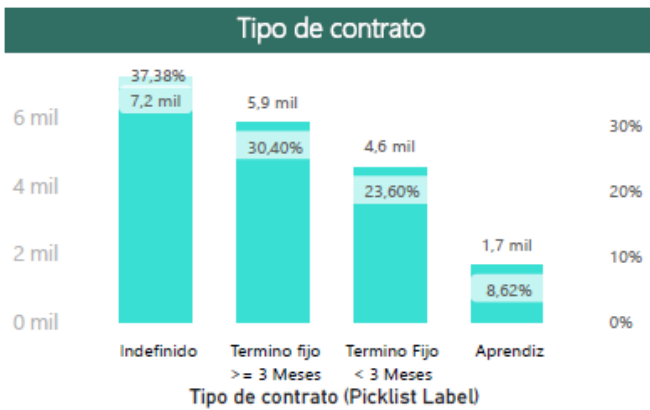


En esta Figura se identificó la distribución en rangos de edad de los empleados retirados en el año 2022, donde se logró identificar que el 70.73% de los retiros correspondieron a empleados a con un rango de edad entre los 18 y 30 años, siendo este el de mayor frecuencia.

Nota: Grafica de distribución de Rango de edades en Retiros en el 2022.

**Figura 4**

*Distribución de tipos de contratos en Retiros en el 2022.*

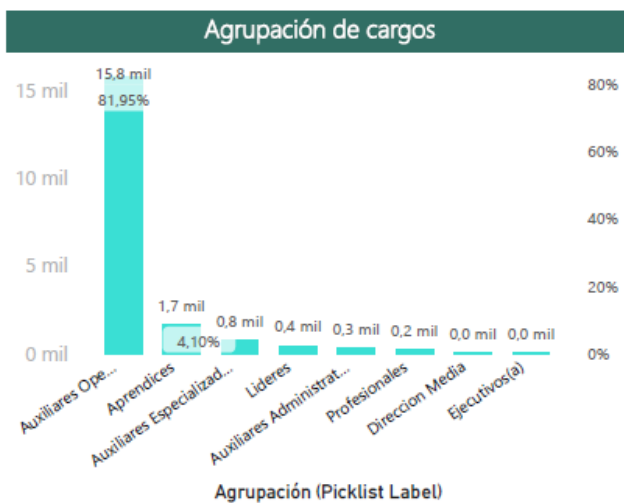


En esta Figura que logró identificar que la mayor cantidad de los empleados retirados en el año 2022 correspondieron a empleados que tenían una contratación a término Indefinido.

Nota: Grafica de distribución de tipos de contratos en Retiros en el 2022.

**Figura 5**

*Distribución de cargos en Retiros en el 2022.*



Sobre a Figura se identificó que el cargo de “Auxiliares Operativos” tuvo una alta rotación para el grupo Éxito dado la información obtenida de los retiros para el año 2022.

Nota: Grafica de distribución de cargos en Retiros.

### **9.2.3. Variables irrelevantes**

A continuación, se enumeraron las variables que fueron previamente identificadas como irrelevantes para el análisis no supervisado que se llevó a cabo, cuyo objetivo principal fue identificar patrones asociados a la rotación de personal. Por lo tanto, su eliminación permitió una mayor claridad en el análisis de los datos y garantizó que los resultados obtenidos sean más precisos y útiles para la toma de decisiones.

1. ID de usuario
2. Número de identificación
3. Nombre formal
4. Event Reason lcode
5. Número de teléfono
6. Dirección de correo electrónico
7. Empleado que realiza la baja
8. Empleado que realiza la baja (Primer nombre)
9. Empleado que realiza la baja (Primer apellido)
10. Fecha de nacimiento
11. Evento (*Picklist Label*)
12. Código de lugar de trabajo
13. Código de puesto
14. Centro de costos (ID Centro de Costo)
15. Centro de costos (*Label*)
16. Lugar de trabajo
17. Ciudad de Notificación (*Picklist Label*)
18. División secundaria (*Label*)

19. Horario laboral (*Label*)
20. Función de puesto (*Label*)
21. Título de puesto (*Label*)
22. Lugar de trabajo (Lugar de trabajo)
23. Formato (*Picklist Label*)
24. Marca (*Picklist Label*)
25. Departamento (*Label*)
26. Departamento 3 (principal) (*Label*)
27. Departamento 3 (secundario) (*Label*)
28. Departamento secundario (*Label*)
29. Generación (*Picklist Label*)
30. Zona (Zona)
31. Título de puesto (*Label*)
32. Régimen Laboral (*Picklist Label*)
33. Número de días del contrato: Se identificó que este dato está mal calculado desde la fuente, por consiguiente, esta columna se debe reemplazar por una nueva
34. División (*Label*)
35. Nivel del puesto (*Picklist Label*)

#### **9.2.4. Identificación de nullos, datos atípicos y duplicados**

En la etapa de preprocesamiento de datos, se realizó una eliminación de valores duplicados. Este proceso permitió asegurar la integridad de los datos, eliminando registros repetidos que podrían introducir sesgos en el análisis posterior. La eliminación de valores duplicados se realizó mediante una comparación de todas las filas del conjunto de datos, y la eliminación de aquellas que presenten una correspondencia exacta en todos sus atributos. De

esta manera, se aseguró que los resultados del análisis de datos sean confiables y precisos.

Entendiendo lo anterior para el escenario preciso se identificaron 2.237 valores duplicados mediante la función “*duplicated*” la cual devolvió una serie booleana que indica si una fila es duplicada o no, es decir, si hay otra fila en el conjunto de datos con los mismos valores. Esta serie puede ser utilizada para filtrar y mostrar solo los registros duplicados en el conjunto de datos.

**Figura 6**

*Conjunto de datos.*

```
#Identificando valores duplicados
data.loc[data.duplicated()]
```

	Fecha de Ingreso	Fecha de terminación de contrato	Event Reason Icode (Label)	Tipo de contrato (Picklist Label)	Sociedad (Label)	Unidad de negocio (Label)	ETC - Empleado a tiempo completo	Agrupación (Picklist Label)	Cargo Líder	Edad Calculada	Género
1153	2022-09-22	2022-10-09	Expiración plazo fijo pactado Por Expiración D...	Termino Fijo < 3 Meses	Almacenes Exito S.A.	Retail	0.88	Auxiliares Operativos	False	22	F
1207	2022-06-16	2022-07-05	Expiración plazo fijo pactado Por Expiración D...	Termino Fijo < 3 Meses	Almacenes Exito S.A.	Retail	1.00	Auxiliares Operativos	False	21	M
1363	2022-06-22	2022-07-17	Expiración plazo fijo pactado Por Expiración D...	Termino Fijo < 3 Meses	Almacenes Exito S.A.	Retail	0.88	Auxiliares Operativos	False	23	M
1367	2022-06-22	2022-07-17	Expiración plazo fijo pactado Por Expiración D...	Termino Fijo < 3 Meses	Almacenes Exito S.A.	Retail	0.88	Auxiliares Operativos	False	22	M
1378	2022-06-22	2022-07-17	Expiración plazo fijo pactado Por Expiración D...	Termino Fijo < 3 Meses	Almacenes Exito S.A.	Retail	0.88	Auxiliares Operativos	False	22	F
...	...	...	...	...	...	...	...	...	...	...	...
19160	2017-09-04	2022-06-15	Por Decisión unilateral del empleado Renuncia ...	Indefinido	Almacenes Exito S.A.	Retail	1.00	Profesionales	False	31	F
19177	2022-11-24	2022-12-02	Expiración plazo fijo pactado Por Expiración D...	Termino Fijo < 3 Meses	Almacenes Exito S.A.	Retail	0.88	Auxiliares Operativos	False	22	M
19184	2022-07-01	2022-07-17	Expiración plazo fijo pactado Por Expiración D...	Termino Fijo < 3 Meses	Almacenes Exito S.A.	Retail	0.88	Auxiliares Operativos	False	20	F
19251	2018-11-26	2022-09-19	Por Justa Causa Omisión De Procedimientos	Indefinido	Almacenes Exito S.A.	Retail	1.00	Auxiliares Operativos	False	42	F
19290	2022-11-24	2022-12-02	Expiración plazo fijo pactado Por Expiración D...	Termino Fijo < 3 Meses	Almacenes Exito S.A.	Retail	0.88	Auxiliares Operativos	False	30	F

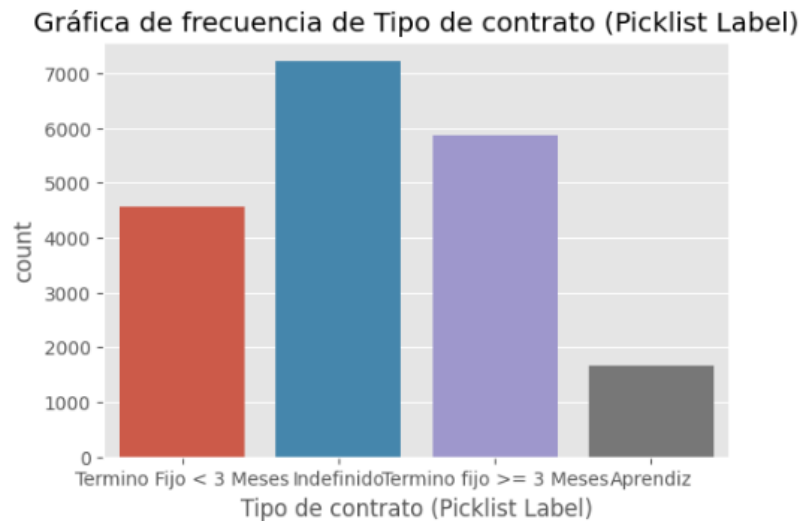
2237 rows x 11 columns

Nota: Grafica para la visualización del conjunto de datos.

En el análisis exploratorio se identificaron diferentes tipos de contratos entre ellos esta los contratos de tipo aprendiz, de acuerdo con el objetivo de identificar los patrones asociados a la rotación del personal en el Grupo Éxito, los retiros vinculados a este tipo de contrato no son relevantes dado que su causa de retiro está vinculada con la finalización de sus prácticas, por tanto, se decide realizar la eliminación de los registros que contiene tipo de contrato “Aprendiz”.

## Figura 7

Gráfica de tipo de contrato.



Nota: Gráfica para la visualización de datos según el tipo de contrato.

Para la eliminación de los registros se utilizó una función de la librería Pandas que sirve para seleccionar todas las filas en el conjunto de datos donde el valor de la columna “Tipo de contrato (*Picklist Label*)” sea diferente de “Aprendiz”.

## Figura 8

Eliminación de datos de “Aprendiz”.

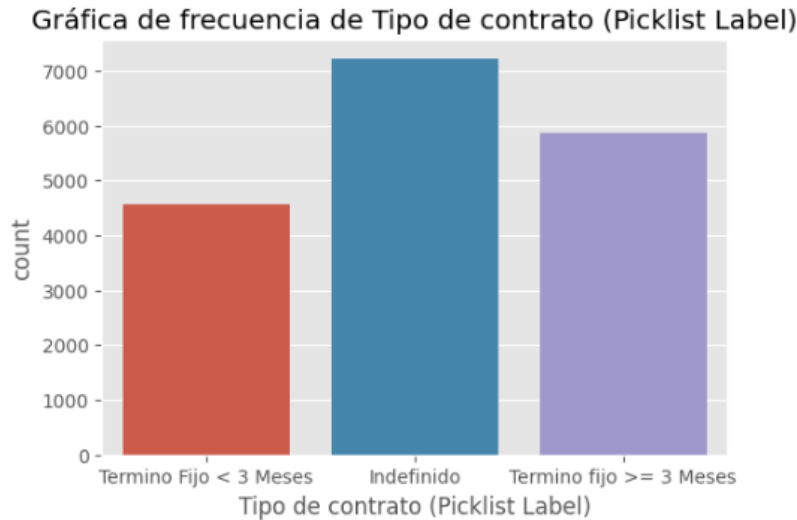
```
[16] #Se realiza eliminación de los registros que contienen tipo de contrato "Aprendiz"  
data = data.loc[data['Tipo de contrato (Picklist Label)'] != "Aprendiz"]
```

Nota: en esta grafica se evidencia el código para la eliminación de datos de “Aprendiz”

Luego, el resultado se guardó en el mismo objeto, sobrescribiendo el conjunto de datos original, como se observa en la Figura 8.

## Figura 9

Gráfica de tipo de contrato sin aprendiz.



Nota: Visualización de los datos de tipo de contrato sin aprendiz.

Luego de haber realizado la función anterior, se logró identificar que la información sobre el tipo de contrato, fueran las necesarias para utilizar en el análisis necesario, tal como lo podemos ver en la Figura 9.

Para asegurar la calidad de los datos, se llevó a cabo una eliminación de valores duplicados. Este proceso se justifica en el hecho de que los datos repetidos pudieron haber tenido un impacto negativo en el análisis y el modelado de datos. Al estar presentes valores duplicados, los resultados del análisis pueden ser sesgados y esto puede llevar a decisiones erróneas. Por lo tanto, eliminar estos valores duplicados es crucial para asegurar una correcta interpretación de los datos y obtener resultados precisos.

En el ejercicio preciso se realizó la utilización de un ciclo que permitió iterar a través de todas las columnas en el conjunto de datos y contar la cantidad de valores que estaban nulos en

cada columna, como se pudo ver en la figura 10.

## Figura 10

*Validación de datos nulos.*

```
[22] for col in data.columns:
      print (col, np.sum(pd.isnull(data[col])))

Fecha de ingreso 0
Fecha de terminación de contrato 0
Event Reason Icode (Label) 0
Tipo de contrato (Picklist Label) 0
Sociedad (Label) 0
Unidad de negocio (Label) 3
ETC - Empleado a tiempo completo 0
Agrupación (Picklist Label) 0
Cargo líder 0
Edad Calculada 0
Género 0
```

Nota: *código para la validación y visualización de datos nulos.*

Al realizar la identificación de valores nulos con los datos, se identificó que la variable “*unidad de negocio (label)*” tenía 3 registros vacíos, dada esta cantidad la cual no representó un porcentaje considerable frente al conjunto de datos y se procedió a realizar la eliminación de dichos registros utilizando el método de la librería Pandas que su función básica es eliminar todas las filas del conjunto de datos que contengan valores nulos y esto se pudo evidenciar en la figura 11.

## Figura 11

*Validación sin datos nulos.*

```
[23] #Eliminación de los registros que estan vacios sobre la Unidad de negocio (Label)
      data = data.dropna(subset=['Unidad de negocio (Label)'])

#verificacion de valores null
for col in data.columns:
    print (col, np.sum(pd.isnull(data[col])))

Fecha de ingreso 0
Fecha de terminación de contrato 0
Event Reason Icode (Label) 0
Tipo de contrato (Picklist Label) 0
Sociedad (Label) 0
Unidad de negocio (Label) 0
ETC - Empleado a tiempo completo 0
Agrupación (Picklist Label) 0
Cargo líder 0
Edad Calculada 0
Género 0
```

Nota: *Código para la Validación y visualización sin datos nulos.*

## 9.3. Preparación de los datos

### 9.3.1. Transformación de datos

Debido a la extensa lista de causales de retiro en el Grupo Éxito, se tomó la decisión de agruparlas de manera más discreta. Esto implicó la creación de nuevos grupos que engloben las causales de retiro similares, a las que se les asignaron un nombre representativo. El objetivo principal de esta agrupación era permitir que los métodos de análisis a utilizar funcionaran de manera óptima y puedan identificar patrones asociados a la rotación de personal. Al agrupar las causales de retiro de esta manera, se facilitó el proceso de análisis y se mejoró la capacidad de obtener resultados más precisos y útiles para la toma de decisiones, tal y como se observa en la figura 12.

#### Figura 12

*Agrupación de causas de retiros.*

```
[25] # Reemplazando Valores

# Se define una función que toma como entrada una fila (como una serie de Pandas)
# y devuelve el valor correspondiente para la columna "Tipo_Retiro"
def get_tipo_retiro(row):
    if "Decisión unilateral de la compañía" in row["Event Reason Icode (Label)"]:
        return "Decisión Compañía"
    elif "Expiración De Plazo Fijo" in row["Event Reason Icode (Label)"]:
        return "Expiración De Plazo Fijo Pactado"
    elif "Justa Causa" in row["Event Reason Icode (Label)"]:
        return "Justa Causa"
    elif "Mutuo acuerdo" in row["Event Reason Icode (Label)"]:
        return "Mutuo Acuerdo"
    elif "Decisión unilateral del empleado" in row["Event Reason Icode (Label)"]:
        return "Renuncia Voluntaria"
    else:
        return "No se presentó a firma de contrato"

data["Tipo_Retiro"] = data.apply(get_tipo_retiro, axis=1)
```

Nota: Grafica con el código para realizar la *agrupación de causas de retiros*.

La función mencionada anteriormente. utiliza una serie de condicionales para determinar el tipo de retiro basándose en la información proporcionada en la *columna "Event Reason Icode (Label)"*.

Dado que la función descrita anteriormente genera una nueva columna llamada

“Tipo\_Retiro” la cual realizó la agrupación a emplear sobre el modelo se vio necesario realizar eliminación de la variable “Event Reason Icode (Label)”. Este paso se puede evidenciar en la figura 13 que muestra el código y las funciones que se utilizaron y en la figura 14 donde se muestra la gráfica de frecuencias de la nueva categoría.

### Figura 13

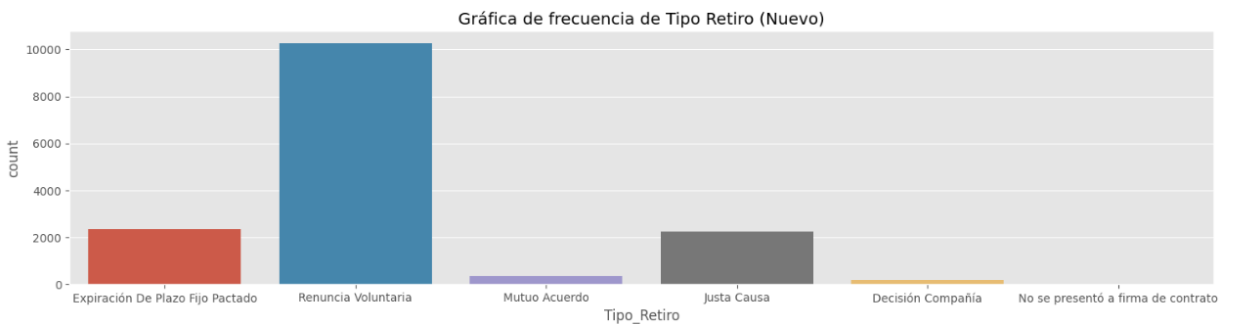
*Eliminación de la variable "Event Reason Icode (Label)".*

```
[ ] data = data.drop(['Event Reason Icode (Label)'],axis=1)
data.info()
```

Nota: Código para la eliminación de la variable "Event Reason Icode (Label)".

### Figura 14

*Gráfica tipo de retiros*



Nota: Gráfica para visualizar los datos según el tipo de retiros

Para identificar la cantidad de días laborados para cada uno de los registros del de los datos se realizó un cálculo que permitió identificar la diferencia en días entre las fechas de ingreso y la fecha de retiro. Este cálculo permitió obtener información relevante sobre la duración de la relación laboral de cada empleado y su contribución a la empresa. Para dicho calculo se empleó una función llamada “calcular\_dias\_laboras” la cual realizó una operación de resta entre fechas, y devuelve un objeto de tiempo en días, esto se puede observar en la figura 15.

## Figura 15

### Cálculo de días laborados

```
[ ] # Definir función que calcula la cantidad de días laborados
def calcular_dias_laborados(fecha_ingreso, fecha_terminacion):
    #fecha_ingreso = datetime.strptime(fecha_ingreso, '%d/%m/%Y')
    #fecha_terminacion = datetime.strptime(fecha_terminacion, '%d/%m/%Y')
    dias_laborados = (fecha_terminacion - fecha_ingreso).days + 1
    return dias_laborados

[ ] data['dias_laborados']=data.apply(lambda row: calcular_dias_laborados(row['Fecha de ingreso'],
                                                                    row['Fecha de terminación de contrato']),
                                   axis=1)
```

*Nota: Código para realizar el cálculo de días laborados*

Dada la función descrita en la imagen anterior se identificó que genera una nueva columna llamada “días\_laborados”, se decidió eliminar las columnas “Fecha de Ingreso” y “Fecha de terminación de contrato”.

## Figura 16

### Eliminación de fechas.

```
[ ] data.drop(['Fecha de ingreso'], axis='columns', inplace=True)

[ ] data.drop(['Fecha de terminación de contrato'], axis='columns', inplace=True)
```

*Nota: Código para la eliminación de fechas.*

Con la generación de los días laborados se vio necesario realizar una categorización de los datos, por lo tanto, se realizó la generación de los siguientes grupos:

- **Entre 1 y 3 meses:** Para los registros cuyos días laborados sean menores a 90 días

- **Entre 3 y 6 meses:** Para los registros cuyos días laborados sean mayor a 90 días y menor o igual a 180 días.
- **Mayor de 6 meses:** Para los registros mayores a 180 días.

Dado los grupos mencionados anteriormente se emplea la función “calcular\_rango\_laboral” la cual permite asignar el grupo correspondiente de acuerdo con la condición aplicada. Esto se puede observar en la figura 17.

## Figura 17

### Creación de rangos laborales

```
[ ] def calcular_rango_laboral(dias_laborables):
    if dias_laborables < 90:
        return "De 1 a 3 Meses"
    elif dias_laborables <= 180:
        return "De 3 a 6 Meses"
    else:
        return "Mayor a 6 Meses"

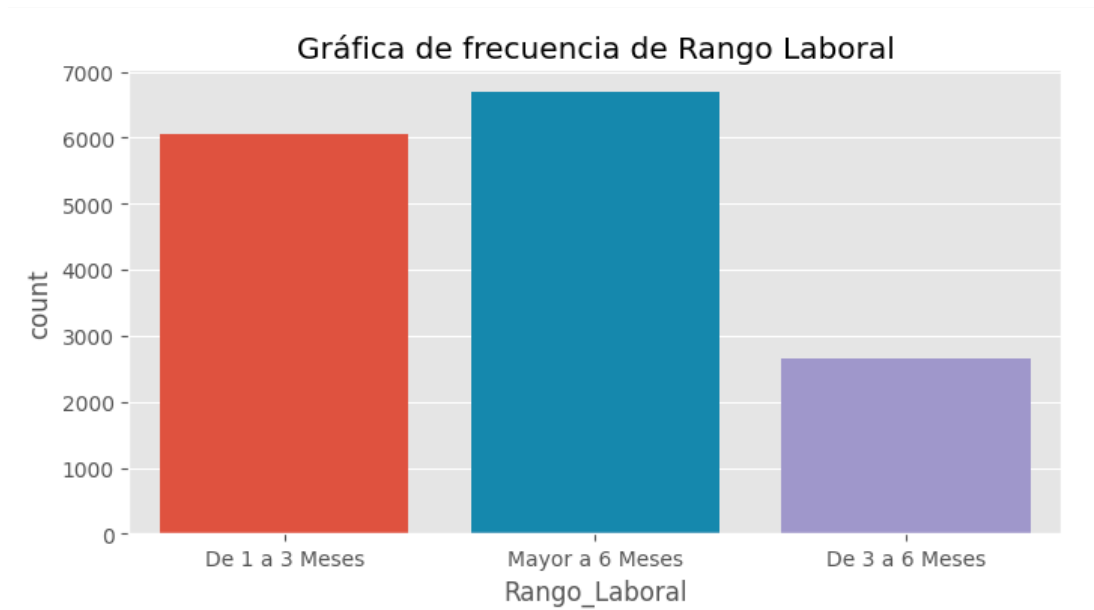
data["Rango_Laboral"] = data['dias_laborados'].apply(calcular_rango_laboral)
```

Nota: *Código para la creación de los rangos laborales*

Al emplear la función descrita en la gráfica anterior se generó una nueva columna llamada “Rango\_Laboral”, por tanto, la columna generada anteriormente llamada “dias\_laborados” se eliminó dado que la información de esta columna ya está agrupada de acuerdo con las categorías definidas. Y se graficó la variable que permitió ver la frecuencia según el rango laboral como se puede observar en la gráfica 18.

**Figura 18**

*Grafica de rango laboral.*



*Nota: Grafica para la visualización de datos según el rango laboral.*

Se definió categorizar las edades dado que permite agrupar a los individuos en categorías más generales, lo que facilitó la interpretación de los resultados. En lugar de tener un gran número de categorías individuales para cada edad, agruparlas en rangos puede ayudar a identificar patrones y tendencias comunes dentro de cada grupo. Además, la categorización de edades en rangos puede ser útil cuando se trabaja con grandes volúmenes de datos, ya que puede reducir el número de características o variables que se deben considerar en el análisis, lo que puede mejorar la eficiencia y la calidad de los resultados obtenidos. Para ello se empleó una función llamada “calcular\_rango\_edad” la cual permitió asignar una agrupación según diferentes condiciones.

## Figura 19

### *Cálculo de rango de edades*

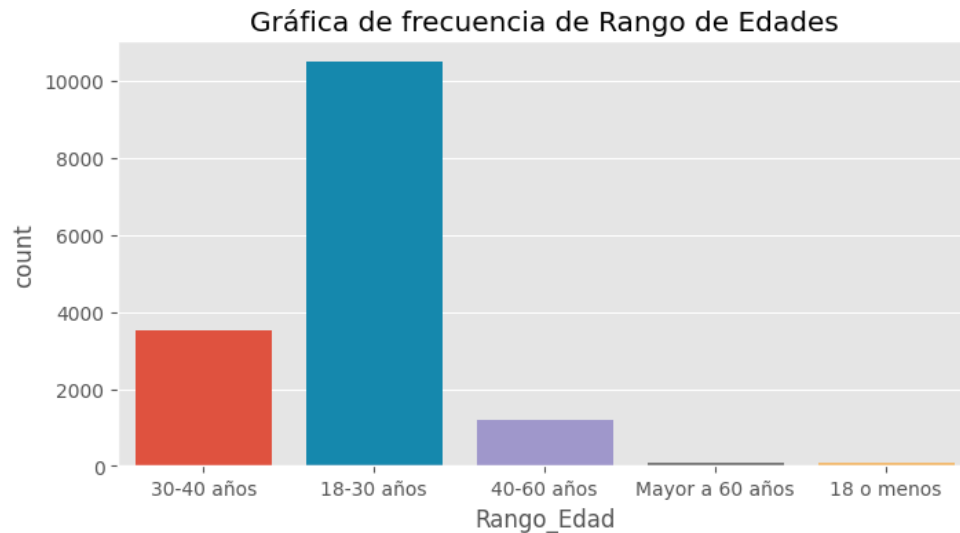
```
[ ] def calcular_rango_edad(edad):  
    if edad <= 18:  
        return "18 o menos"  
    elif edad <= 30:  
        return "18-30 años"  
    elif edad <= 40:  
        return "30-40 años"  
    elif edad <= 60:  
        return "40-60 años"  
    else:  
        return "Mayor a 60 años"  
  
data["Rango_Edad"] = data['Edad Calculada'].apply(calcular_rango_edad)
```

Nota: Código para la creación del *cálculo de rango de edades*.

Dado que al emplear la función descrita en la imagen anterior generó una nueva columna llamada "Rango\_Edad", se realizó la eliminación de la variable "Edad Calculada". Y en la figura 20 podemos ver como es el comportamiento de los datos según los rangos de edades.

## Figura 20

Gráfica de rango de edades



```
[ ] #Dada la definición de los rangos de las edades se realiza eliminación de la variables Edad Calculada
data.drop(['Edad Calculada'], axis='columns', inplace=True)
```

Nota: Gráfica para la visualización de los datos según el rango de edades.

Se realizó una identificación de los valores de las variables categóricas, esto con el fin de identificar la distribución que existe dadas sus categorías.

## Figura 21

Agrupación de variables categóricas

```
[130] catCols=data.select_dtypes(include = ["object"]).columns.tolist()
for ind, col in enumerate(data[catCols].columns):
    print ("-----" + col + "-----")
    print(data.groupby(col).size())
```

Nota: Código para la agrupación de variables categóricas

Según el código descrito en la imagen anterior, permite realizar una agrupación por cada una de las categorías sobre las variables nominales o categóricas.

## Figura 22

### Validación de datos por sociedades

```
===== Sociedad (Label) =====  
Sociedad (Label)  
Almacenes Exito S.A.                15324  
Exito Industrias SAS                10  
Exito Viajes y Turismo SAS          29  
Logistica Transporte Y Servicios Asociados SAS 26  
Marketplace Internacional Exito Y Servicios S.A.S. 1  
TRANSACCIONES ENERGÉTICAS S.A.S    1
```

Nota: Validación y visualización de cantidad de datos por sociedades

Dada la distribución que se muestra en la imagen anterior donde se describe las diferentes categorías sobre la variable “Sociedad (*label*)” se identifica que la mayor cantidad de registros pertenecen a la sociedad de Almacenes Éxito S.A. por tanto se decidió realizar una segmentación y eliminar los registros que no pertenezcan a esta sociedad. Para ello se emplea el método de la librería de pandas que permite realizar esta eliminación, como lo podemos ver en la figura 23

## Figura 23

### Eliminación de datos de sociedades

```
[43] #Eliminando Registros diferentes a la sociedad = Almacenes Exito  
data = data.drop(data[data['Sociedad (Label)'] != "Almacenes Exito S.A."].index)
```

Nota: Código para la *eliminación de datos de sociedades diferentes a Almacenes Éxito S.A.*

Con base a la distribución de las variables categóricas, se identificó que la variable “Agrupación (*Picklist Label*)” existe una categoría “Aprendices” la cual de acuerdo con el objetivo del estudio no aporta valor relevante para determinar los patrones que inciden en los retiros, por tanto, se realiza eliminación de dichos registros.

## Figura 24

*Validación de datos de agrupación de cargos*

```
===== Agrupación (Picklist Label) =====
Agrupación (Picklist Label)
Aprendices                11
Auxiliares Administrativos 340
Auxiliares Especializados  792
Auxiliares Operativos     13579
Direccion Media           27
Ejecutivos(a)             4
Lideres                   446
Profesionales             192
dtype: int64
```

Nota: *Validación de cantidad de datos de agrupación por cargos*

## Figura 25

*Eliminación de datos de agrupación de cargos*

```
[131] #Eliminand Registros con agrupación = aprendices
      data = data.loc[data['Agrupación (Picklist Label)']!= "Aprendices"]
```

Nota: *Eliminación de datos de aprendices de la variable agrupación de cargos*

La variable “cargo líder” está definida de tipo booleano, por tanto, se realiza un reemplazo de sus datos y conversión a valor entero, esto lo podemos ver en la figura 26.

## Figura 26

*Reemplazo de datos de ‘Cargo Líder’*

```
[50] data["Cargo_lider"].replace({"FALSO": 0, "VERDADERO": 1}, inplace=True)
```

```
[51] data['Cargo_lider'] = data['Cargo_lider'].values.astype(int)
```

Nota: *Código para el reemplazo de datos de ‘Cargo Líder’*

### 9.3.2. Renombramiento de variables

Se realizó un renombramiento de variables con el objetivo de proporcionar una nomenclatura más clara y coherente en el conjunto de datos. Este proceso de renombramiento ayuda a mejorar la comprensión de los datos y facilita su posterior análisis.

**Figura 27**

*Renombramiento de variables*

```
[46] data = data.rename(columns={'Tipo de contrato (Picklist Label)': 'Tipo_Contrato',
                                'Unidad de negocio (Label)': 'Unidad_negocio',
                                'ETC - Empleado a tiempo completo':'ETC',
                                'Agrupación (Picklist Label)':'Agrupacion_cargo',
                                'Cargo líder':'Cargo_lider'})
```

```
[47] data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 15314 entries, 0 to 19293
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   Tipo_Contrato         15314 non-null  object
1   Unidad_negocio       15314 non-null  object
2   ETC                   15314 non-null  float64
3   Agrupacion_cargo     15314 non-null  object
4   Cargo_lider          15314 non-null  bool
5   Género               15314 non-null  object
6   Tipo_Retiro          15314 non-null  object
7   Rango_Laboral        15314 non-null  object
8   Rango_Edad           15314 non-null  object
dtypes: bool(1), float64(1), object(7)
memory usage: 1.1+ MB
```

Nota: código para el renombramiento de variables.

### 9.3.3. Generación de variables *Dummies*

Se realizó la generación de variables *dummies* para convertir las variables categóricas en variables numéricas binarias, lo que facilita su uso en modelos estadísticos. Estas nuevas variables *dummies* se utilizaron en el análisis posterior para proporcionar una representación más completa y precisa de los datos.

**Figura 28**

*Generación de variables Dummies*

#	Column	Non-Null	Count	Dtype
0	ETC	15314	non-null	float64
1	Cargo_lider	15314	non-null	int64
2	Tipo_Contrato_Termino Fijo < 3 Meses	15314	non-null	uint8
3	Tipo_Contrato_Termino Fijo >= 3 Meses	15314	non-null	uint8
4	Unidad_negocio_Industria	15314	non-null	uint8
5	Unidad_negocio_Inmobiliaria	15314	non-null	uint8
6	Unidad_negocio_Retail	15314	non-null	uint8
7	Unidad_negocio_Servicios	15314	non-null	uint8
8	Agrupacion_cargo_Auxiliares Especializados	15314	non-null	uint8
9	Agrupacion_cargo_Auxiliares Operativos	15314	non-null	uint8
10	Agrupacion_cargo_Direccion Media	15314	non-null	uint8
11	Agrupacion_cargo_Ejecutivos(a)	15314	non-null	uint8
12	Agrupacion_cargo_Lideres	15314	non-null	uint8
13	Agrupacion_cargo_Profesionales	15314	non-null	uint8
14	Género_M	15314	non-null	uint8
15	Género_O	15314	non-null	uint8
16	Tipo_Retiro_Expiración De Plazo Fijo Pactado	15314	non-null	uint8
17	Tipo_Retiro_Justa Causa	15314	non-null	uint8
18	Tipo_Retiro_Mutuo Acuerdo	15314	non-null	uint8
19	Tipo_Retiro_No se presentó a firma de contrato	15314	non-null	uint8
20	Tipo_Retiro_Renuncia Voluntaria	15314	non-null	uint8
21	Rango_Laboral_De 3 a 6 Meses	15314	non-null	uint8
22	Rango_Laboral_Mayor a 6 Meses	15314	non-null	uint8
23	Rango_Edad_18-30 años	15314	non-null	uint8
24	Rango_Edad_30-40 años	15314	non-null	uint8
25	Rango_Edad_40-60 años	15314	non-null	uint8
26	Rango_Edad_Mayor a 60 años	15314	non-null	uint8

Nota: Código para la generación de variables *Dummies*

En la figura 28, se identificaron 26 variables dummies adicionales generadas a partir de variables categóricas, lo que permite trabajar con estas variables en modelos estadísticos y análisis posteriores. Estas variables binarias proporcionan información valiosa sobre la presencia o ausencia de una categoría en particular y mejorarán la precisión del análisis.

**9.3.4. Normalización de los Datos**

la normalización de datos es una técnica de preprocesamiento de datos que se utiliza para reducir la variación en los datos y garantizar que todas las variables se traten de manera uniforme. Esto puede mejorar la eficacia de los modelos de aprendizaje automático al reducir el impacto de las variables con valores extremos.

## Figura 29

### Normalización de los Datos

```
data_norm = (dataDum - dataDum.min()) / (dataDum.max() - dataDum.min())
data_norm.head()
```

	ETC	Cargo_lider	Tipo_Contrato_Termino Fijo < 3 Meses	Tipo_Contrato_Termino fijo >= 3 Meses	Unidad_negocio_Industria	Unidad_negocio_Inmobiliaria	Unidad_negocio_Retail	Unidad_negocio_Servicios	Agrupacion_cargo_Auxiliares Especializados	Agri
0	0.820896	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
1	1.000000	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
2	1.000000	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0
3	1.000000	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
4	0.000000	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0

5 rows × 27 columns

Nota: Visualización para la *normalización de los Datos*.

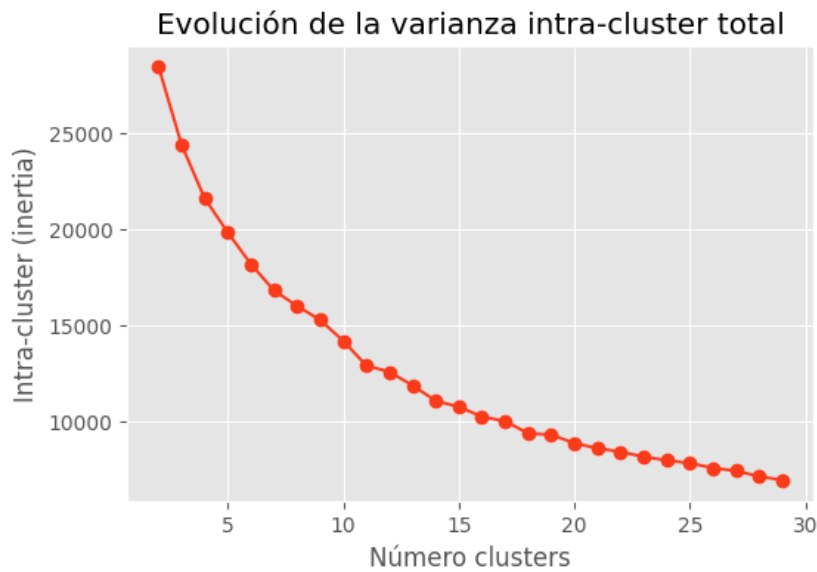
## 9.4. Modelamiento

### 9.4.1. Identificación del valor de K

**El método del codo (*Elbow Method*):** es una técnica utilizada en el análisis de *clustering* para determinar el número óptimo de clústeres que deberían ser utilizados para segmentar un conjunto de datos. Consiste en trazar la suma de las distancias cuadradas dentro del clúster (SSW) en función del número de clústeres, y elegir el punto en el que la disminución de la suma de distancias cuadradas comienza a disminuir significativamente menos a medida que aumenta el número de clúster. Este punto suele ser representado en un gráfico como un "codo", de ahí el nombre del método. El método del codo es una herramienta útil para ayudar a los analistas de datos a seleccionar un número óptimo de clústeres para una determinada tarea de *clustering*.

**Figura 30**

*Identificación de clústeres*



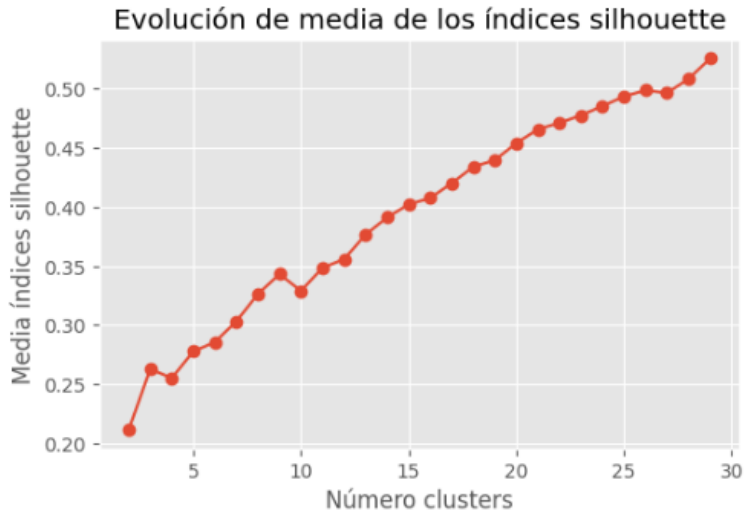
Nota: Grafica en la que podemos identificar la cantidad de clústeres.

De acuerdo con la figura 30 se identificó que el número de clúster óptimos según el método del codo serian 10 clúster.

**El método *Silhouette*:** es una técnica de *clustering* utilizada para evaluar la calidad de la agrupación obtenida a partir de un algoritmo de *clustering*. Para cada punto de datos, se calcula la medida de *Silhouette*, que es una puntuación que indica cuán bien se ajusta el punto a su propio clúster en comparación con otros clústeres. La medida de *Silhouette* varía de -1 a 1, donde los valores cercanos a 1 indican que el punto se ajusta bien a su propio clúster y no tan bien a otros, mientras que los valores cercanos a -1 indican lo contrario. Una puntuación media alta indica una buena calidad de agrupación. El método *Silhouette* se utiliza para seleccionar el número óptimo de clústeres en una solución de *clustering* y para comparar diferentes soluciones de *clustering*.

**Figura 31**

*Método de Silhouette*



Nota: Grafica en la que se muestra la evolución de medida de los *índices de Silhouette*

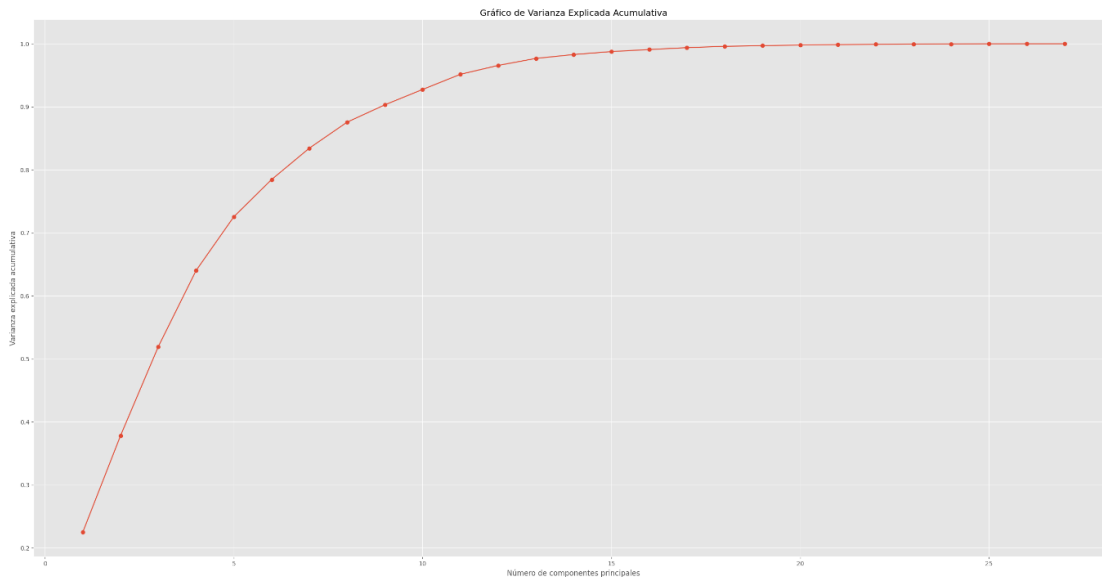
Según la figura 31 se identificó que el número óptimo de clústeres está cercano a 30, dado que es el valor más cercano a 1.

**Técnica de Análisis de Componentes Principales:** es una técnica estadística utilizada para reducir la dimensionalidad de un conjunto de datos manteniendo la mayor cantidad posible de información relevante. PCA busca transformar un conjunto de variables correlacionadas en un nuevo conjunto de variables no correlacionadas llamadas componentes principales. Dada la cantidad de variables que se tiene en el modelo se decidió implementar la técnica PCA para reducir la dimensionalidad y facilitar el análisis en la generación de los clústeres.

Para implementar esta técnica se debe indicar el valor de componente principales, esta elección se realiza a través de la varianza explicada la cual a través de un gráfico traza una línea en función del número de componentes principales y allí identificar el punto de inflexión.

**Figura 32**

*Grafica de varianza*



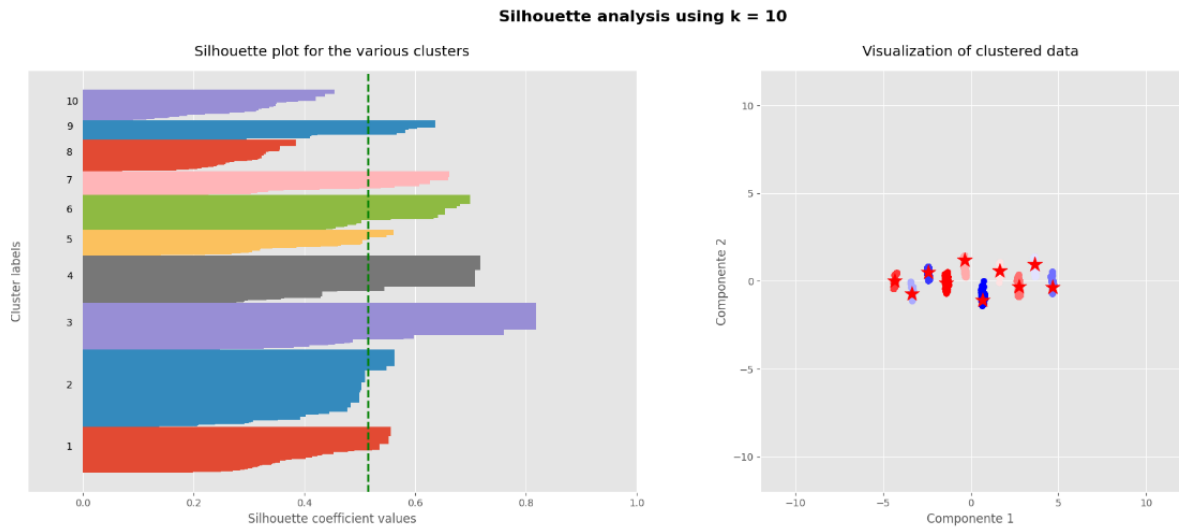
Nota: Grafica en la que se muestra la varianza explicada acumulativa.

Al ejecutar la varianza explicada se genera una grafica como la evidencia en la Figura 32, en ella se identifica como se comporta la varianza explicada acumulada frente al número de componente principales, con base a la Figura 32 se identifica que el punto de inflexión (donde el incremento en la varianza explicada se vuelve menos significativo) se da en 19 componentes principales por tanto se estableció que la cantidad de componentes principales para ejecutar la técnica PCA es de 19.

Una vez generado los componentes principales se genera el método *Silhouette* para visualizar la conformación de los 10 clústeres identificados. En la siguiente imagen se logra identificar como están conformados los clústeres donde se evidencia una buena separabilidad y cohesión entre ellos.

**Figura 33**

Análisis método *Silhouette*



Nota: Grafica en el que se muestra el Análisis del método *Silhouette*

### 9.5. Implementación del Modelo

***K-means***: Es un algoritmo de *clustering* que agrupa un conjunto de datos en K clústeres, donde K es un número previamente definido por el usuario. Se opta por este algoritmo dado que se ajusta al objetivo principal de este estudio, además es muy efectivo en la identificación de grupos o clústeres en grandes conjuntos de datos. Es útil para identificar patrones ocultos o estructuras desconocidas en los datos, lo que puede proporcionar información valiosa para la toma de decisiones en diferentes campos.

Para la utilización de este algoritmo se tomó la decisión de utilizar 10 clústeres, producto de la identificación de K de los métodos implementados anteriormente. Aunque el método del codo sugería una cantidad de 15 clúster, una vez implementado el algoritmo con este valor establecido, se identificó una distribución no muy precisa en cada uno de los clústeres, por lo tanto, se determina implementar el algoritmo para 10 clúster.

Una vez implementado el método de *K-means* se identifica los siguientes centroides.

**Figura 34**

*Visualización de Clústeres*

Kmeans_Clusters	Cantidad	ETC	Cargo_lider	Tipo_Contrato_Termino Fijo < 3 Meses	Tipo_Contrato_Termino fijo >= 3 Meses	Unidad_negocio_Industria	Unidad_negocio_Inmobiliaria	Unidad_negocio_Retail	Unidad_negocio_Servicios	...
0	0	944	0.808421	0.001059	0.047669	0.863347	0.011653	0.000000	0.988347	0.000000 ...
1	1	3281	0.807676	0.019201	0.000000	0.003657	0.010058	0.000610	0.974093	0.008534 ...
2	2	3101	0.772919	0.002257	0.000000	0.954531	0.000645	0.000000	0.997420	0.001935 ...
3	3	1565	0.930189	0.115016	0.000000	0.003834	0.031310	0.005751	0.911821	0.030671 ...
4	4	944	0.834524	0.000000	0.844280	0.000000	0.010593	0.000000	0.969407	0.000000 ...
5	5	941	0.762780	0.003188	0.000000	0.986185	0.000000	0.000000	0.995749	0.004251 ...
6	6	1184	0.890493	0.072635	0.011824	0.000845	0.007601	0.000845	0.976351	0.005068 ...
7	7	1008	0.888963	0.004960	0.000000	0.932540	0.004960	0.000000	0.992063	0.002976 ...
8	8	847	0.944316	0.153483	0.000000	0.182999	0.041322	0.001181	0.924439	0.020071 ...
9	9	1499	0.838858	0.000000	1.000000	0.000000	0.000000	0.000000	1.000000	0.000000 ...

10 rows x 29 columns

Nota: grafica en la que se puede realizar la visualización de los Clústeres

### 9.6. Implementación del Modelo *DBSCAN*

***DBSCAN***: Es un algoritmo de clustering basado en la densidad de los puntos en el espacio. A diferencia de otros algoritmos de *clustering* que asumen una forma de clúster predefinida o requieren que especifiques el número de clústeres de antemano, *DBSCAN* puede descubrir clústeres de forma irregular y no requiere conocimiento previo del número de clústeres.

Para la utilización del método se definió el parámetro eps (el cual determina la proximidad requerida entre los puntos para considerarlos vecinos), en un valor de 1.42, con el cual se identificó un número razonable de clústeres según la definición del método del codo, al correr el método con este parámetro se generaron 7 clústeres con 21 puntos estimados de ruido como se identifica en la siguiente imagen.

**Figura 35**

Código DBSCAN

```
db = DBSCAN(eps=1.42, min_samples=10)
db.fit(data_norm)

labels_db = db.labels_

n_clusters_db = len(set(labels_db)) - (1 if -1 in labels_db else 0)
n_noise_db = list(labels_db).count(-1)

print('Número estimado de clusters: %d' % n_clusters_db)
print('Número estimado de puntos de ruido: %d' % n_noise_db)

Número estimado de clusters: 7
Número estimado de puntos de ruido: 21
```

Nota: Grafica en la que se puede ver el código aplicado del método *DBSCAN*

## 9.7. Evaluación

### 9.7.1. *K-means*

Después de aplicar el algoritmo *K-Means*, se identificó 10 clústeres relevantes y coherentes dentro del conjunto de datos. Cada uno de estos clústeres representa una categoría específica o un patrón distinto en los datos, lo que resulta en una mejor comprensión de la información.

A continuación, se denominarán los clústeres de manera significativa. Este proceso de denominación permitirá una interpretación más intuitiva y una comunicación efectiva de los resultados obtenidos.

**Tabla 2. Descripción de Clúster 0.**

Clúster	0
Cantidad	944
ETC	0,808
Cargo_lider	0,001
Tipo_Contrato_Termino Fijo < 3 Meses	0,048
Tipo_Contrato_Termino fijo >= 3 Meses	0,863

Unidad_negocio_Industria	0,012
Unidad_negocio_Inmobiliaria	0,000
Unidad_negocio_Retail	0,988
Unidad_negocio_Servicios	0,000
Agrupacion_cargo_Auxiliares Especializados	0,033
Agrupacion_cargo_Auxiliares Operativos	0,961
Agrupacion_cargo_Direccion Media	0,000
Agrupacion_cargo_Ejecutivos(a)	0,000
Agrupacion_cargo_Lideres	0,001
Agrupacion_cargo_Profesionales	0,000
Género_M	0,648
Género_O	0,000
Tipo_Retiro_Expiración De Plazo Fijo Pactado	0,000
Tipo_Retiro_Justa Causa	0,912
Tipo_Retiro_Mutuo Acuerdo	0,001
Tipo_Retiro_No se presentó a firma de contrato	0,000
Tipo_Retiro_Renuncia Voluntaria	0,000
Rango_Laboral_De 3 a 6 Meses	0,268
Rango_Laboral_Mayor a 6 Meses	0,001
Rango_Edad_18-30 años	0,978
Rango_Edad_30-40 años	0,000
Rango_Edad_40-60 años	0,020
Rango_Edad_Mayor a 60 años	0,000

Este clúster tiene una cantidad de 944 empleados y se caracteriza por tener una alta proporción de empleados en la unidad de negocio "*Retail*" (98.83%). La mayoría de los empleados en este clúster tienen contratos a término fijo de más de 3 meses (86.34%) y se encuentran en el rango de edad de 18 a 30 años (97.77%). La proporción de género muestra una inclinación hacia los hombres, que representan el 64.83% de los empleados. La principal causa de retiro en este clúster es la "Justa Causa" (91.21%). En términos de cargos, los "Auxiliares Operativos" son la agrupación más representada (96.08%), seguidos por los "Auxiliares Especializados" (3.28%).

**Tabla 3. Descripción de Clúster 1.**

<b>Clúster</b>	<b>1</b>
Cantidad	3281

ETC	0,808
Cargo_lider	0,019
Tipo_Contrato_Termino Fijo < 3 Meses	0,000
Tipo_Contrato_Termino fijo >= 3 Meses	0,004
Unidad_negocio_Industria	0,010
Unidad_negocio_Inmobiliaria	0,001
Unidad_negocio_Retail	0,974
Unidad_negocio_Servicios	0,009
Agrupacion_cargo_Auxiliares Especializados	0,038
Agrupacion_cargo_Auxiliares Operativos	0,893
Agrupacion_cargo_Direccion Media	0,000
Agrupacion_cargo_Ejecutivos(a)	0,000
Agrupacion_cargo_Lideres	0,019
Agrupacion_cargo_Profesionales	0,019
Género_M	0,520
Género_O	0,000
Tipo_Retiro_Expiración De Plazo Fijo Pactado	0,000
Tipo_Retiro_Justa Causa	0,000
Tipo_Retiro_Mutuo Acuerdo	0,012
Tipo_Retiro_No se presentó a firma de contrato	0,000
Tipo_Retiro_Renuncia Voluntaria	0,986
Rango_Laboral_De 3 a 6 Meses	0,000
Rango_Laboral_Mayor a 6 Meses	0,996
Rango_Edad_18-30 años	0,991
Rango_Edad_30-40 años	0,000
Rango_Edad_40-60 años	0,000
Rango_Edad_Mayor a 60 años	0,009

Este clúster tiene una cantidad de 3,281 empleados y se caracteriza por tener una alta proporción de empleados en la unidad de negocio "Retail" (97.41%) y en la agrupación de cargos "Auxiliares Operativos" (89.27%). La mayoría de los empleados en este clúster tienen contratos a término fijo de más de 3 meses (0.37%) y se encuentran en el rango de edad de 18 a 30 años (99.12%). La proporción de género muestra una inclinación hacia los hombres, que representan el 52.03% de los empleados. La principal causa de retiro en este clúster es la "Renuncia Voluntaria" (98.63%). En términos de cargos, los "Profesionales" y los "Líderes" también están presentes, representando el 1.92% y el 1.86% respectivamente.

**Tabla 4. Descripción de Clúster 2.**

<b>Clúster</b>	<b>2</b>
Cantidad	3101
ETC	0,773
Cargo_lider	0,002
Tipo_Contrato_Termino Fijo < 3 Meses	0,000
Tipo_Contrato_Termino fijo >= 3 Meses	0,955
Unidad_negocio_Industria	0,001
Unidad_negocio_Inmobiliaria	0,000
Unidad_negocio_Retail	0,997
Unidad_negocio_Servicios	0,002
Agrupacion_cargo_Auxiliares Especializados	0,032
Agrupacion_cargo_Auxiliares Operativos	0,955
Agrupacion_cargo_Direccion Media	0,000
Agrupacion_cargo_Ejecutivos(a)	0,000
Agrupacion_cargo_Lideres	0,003
Agrupacion_cargo_Profesionales	0,001
Género_M	0,500
Género_O	0,000
Tipo_Retiro_Expiración De Plazo Fijo Pactado	0,000
Tipo_Retiro_Justa Causa	0,000
Tipo_Retiro_Mutuo Acuerdo	0,000
Tipo_Retiro_No se presentó a firma de contrato	0,000
Tipo_Retiro_Renuncia Voluntaria	1,000
Rango_Laboral_De 3 a 6 Meses	0,367
Rango_Laboral_Mayor a 6 Meses	0,000
Rango_Edad_18-30 años	0,989
Rango_Edad_30-40 años	0,000
Rango_Edad_40-60 años	0,004
Rango_Edad_Mayor a 60 años	0,001

Este clúster tiene una cantidad de 3,101 empleados y se caracteriza por tener una alta proporción de empleados en la unidad de negocio "Retail" (99.74%) y en la agrupación de cargos "Auxiliares Operativos" (95.45%). La mayoría de los empleados en este clúster tienen contratos a término fijo de más de 3 meses (95.45%). En cuanto al género, la proporción está equilibrada con un 49.98% de hombres y un 50.02% de mujeres. El rango de edad predominante en este clúster es de 18 a 30 años (98.87%). El tipo de retiro más común en este clúster es la "Renuncia

Voluntaria" (100%). Además, se observa que una proporción significativa de empleados tiene una experiencia laboral en el rango de 3 a 6 meses (36.67%).

**Tabla 5. Descripción de Clúster 3.**

<b>Clúster</b>	<b>3</b>
Cantidad	1565
ETC	0,930
Cargo_lider	0,115
Tipo_Contrato_Termino Fijo < 3 Meses	0,000
Tipo_Contrato_Termino fijo >= 3 Meses	0,004
Unidad_negocio_Industria	0,031
Unidad_negocio_Inmobiliaria	0,006
Unidad_negocio_Retail	0,912
Unidad_negocio_Servicios	0,031
Agrupacion_cargo_Auxiliares Especializados	0,104
Agrupacion_cargo_Auxiliares Operativos	0,672
Agrupacion_cargo_Direccion Media	0,006
Agrupacion_cargo_Ejecutivos(a)	0,001
Agrupacion_cargo_Lideres	0,105
Agrupacion_cargo_Profesionales	0,057
Género_M	0,500
Género_O	0,000
Tipo_Retiro_Expiración De Plazo Fijo Pactado	0,001
Tipo_Retiro_Justa Causa	0,000
Tipo_Retiro_Mutuo Acuerdo	0,084
Tipo_Retiro_No se presentó a firma de contrato	0,000
Tipo_Retiro_Renuncia Voluntaria	0,902
Rango_Laboral_De 3 a 6 Meses	0,003
Rango_Laboral_Mayor a 6 Meses	0,981
Rango_Edad_18-30 años	0,000
Rango_Edad_30-40 años	0,992
Rango_Edad_40-60 años	0,000
Rango_Edad_Mayor a 60 años	0,008

Este clúster contiene 1,565 empleados y se caracteriza por tener una alta proporción de empleados en la unidad de negocio "Retail" (91.18%) y en la agrupación de cargos "Auxiliares Operativos" (67.16%). La mayoría de los empleados en este clúster tienen contratos a término

fijo de más de 3 meses (99.62%). En cuanto al género, la proporción está equilibrada con un 50.03% de hombres y un 49.97% de mujeres. El tipo de retiro más común en este clúster es la "Renuncia Voluntaria" (90.16%). Además, se observa que una proporción significativa de empleados tiene una experiencia laboral mayor a 6 meses (98.15%). En términos de edad, este clúster muestra una mayor presencia de empleados en el rango de edad de 30 a 40 años (99.17%).

**Tabla 6. Descripción de Clúster 4.**

<b>Clúster</b>	<b>4</b>
Cantidad	944
ETC	0,835
Cargo_lider	0,000
Tipo_Contrato_Termino Fijo < 3 Meses	0,844
Tipo_Contrato_Termino fijo >= 3 Meses	0,000
Unidad_negocio_Industria	0,011
Unidad_negocio_Inmobiliaria	0,000
Unidad_negocio_Retail	0,989
Unidad_negocio_Servicios	0,000
Agrupacion_cargo_Auxiliares Especializados	0,000
Agrupacion_cargo_Auxiliares Operativos	0,996
Agrupacion_cargo_Direccion Media	0,000
Agrupacion_cargo_Ejecutivos(a)	0,000
Agrupacion_cargo_Lideres	0,000
Agrupacion_cargo_Profesionales	0,001
Género_M	0,374
Género_O	0,001
Tipo_Retiro_Expiración De Plazo Fijo Pactado	0,000
Tipo_Retiro_Justa Causa	0,000
Tipo_Retiro_Mutuo Acuerdo	0,000
Tipo_Retiro_No se presentó a firma de contrato	0,000
Tipo_Retiro_Renuncia Voluntaria	0,992
Rango_Laboral_De 3 a 6 Meses	0,001
Rango_Laboral_Mayor a 6 Meses	0,000
Rango_Edad_18-30 años	0,768
Rango_Edad_30-40 años	0,169
Rango_Edad_40-60 años	0,041
Rango_Edad_Mayor a 60 años	0,003

Este clúster contiene 944 empleados y se caracteriza por tener una alta proporción de empleados en la unidad de negocio "Retail" (98.94%) y en la agrupación de cargos "Auxiliares Operativos" (99.58%). La mayoría de los empleados en este clúster tienen contratos a término fijo de menos de 3 meses (84.43%). En cuanto al género, la proporción es mayoritariamente masculina, con un 37.39% de hombres y un 0.11% de empleados de otro género. El tipo de retiro más común en este clúster es la "Renuncia Voluntaria" (99.15%). En términos de experiencia laboral, la mayoría de los empleados han estado en la empresa entre 3 y 6 meses (0.11%). En cuanto a la edad, se observa una mayor presencia de empleados en el rango de edad de 18 a 30 años (76.80%).

**Tabla 7. Descripción de Clúster 5.**

<b>Clúster</b>	<b>5</b>
Cantidad	941
ETC	0,763
Cargo_lider	0,003
Tipo_Contrato_Termino Fijo < 3 Meses	0,000
Tipo_Contrato_Termino fijo >= 3 Meses	0,986
Unidad_negocio_Industria	0,000
Unidad_negocio_Inmobiliaria	0,000
Unidad_negocio_Retail	0,996
Unidad_negocio_Servicios	0,004
Agrupacion_cargo_Auxiliares Especializados	0,060
Agrupacion_cargo_Auxiliares Operativos	0,926
Agrupacion_cargo_Direccion Media	0,000
Agrupacion_cargo_Ejecutivos(a)	0,000
Agrupacion_cargo_Lideres	0,003
Agrupacion_cargo_Profesionales	0,003
Género_M	0,573
Género_O	0,001
Tipo_Retiro_Expiración De Plazo Fijo Pactado	0,994
Tipo_Retiro_Justa Causa	0,000
Tipo_Retiro_Mutuo Acuerdo	0,005
Tipo_Retiro_No se presentó a firma de contrato	0,000
Tipo_Retiro_Renuncia Voluntaria	0,000

Rango_Laboral_De 3 a 6 Meses	0,983
Rango_Laboral_Mayor a 6 Meses	0,014
Rango_Edad_18-30 años	0,793
Rango_Edad_30-40 años	0,151
Rango_Edad_40-60 años	0,054
Rango_Edad_Mayor a 60 años	0,001

Este clúster contiene 941 empleados y se caracteriza por tener una alta proporción de empleados en la unidad de negocio "*Retail*" (99.57%) y en la agrupación de cargos "Auxiliares Operativos" (92.56%). La mayoría de los empleados en este clúster tienen contratos a término fijo de más de 3 meses (98.62%). En cuanto al género, la proporción es mayoritariamente femenina, con un 57.28% de mujeres y un 0.11% de empleados de otro género. El tipo de retiro más común en este clúster es la "Expiración De Plazo Fijo Pactado" (99.36%). En términos de experiencia laboral, la mayoría de los empleados han estado en la empresa entre 3 y 6 meses (98.30%). En cuanto a la edad, se observa una mayor presencia de empleados en el rango de edad de 18 a 30 años (79.28%).

**Tabla 8. Descripción de Clúster 6.**

<b>Clúster</b>	<b>6</b>
Cantidad	1184
ETC	0,890
Cargo_lider	0,073
Tipo_Contrato_Termino Fijo < 3 Meses	0,012
Tipo_Contrato_Termino fijo >= 3 Meses	0,001
Unidad_negocio_Industria	0,008
Unidad_negocio_Inmobiliaria	0,001
Unidad_negocio_ <i>Retail</i>	0,976
Unidad_negocio_Servicios	0,005
Agrupacion_cargo_Auxiliares Especializados	0,084
Agrupacion_cargo_Auxiliares Operativos	0,812
Agrupacion_cargo_Direccion Media	0,000
Agrupacion_cargo_Ejecutivos(a)	0,000
Agrupacion_cargo_Lideres	0,072
Agrupacion_cargo_Profesionales	0,001

Género_M	0,661
Género_O	0,000
Tipo_Retiro_Expiración De Plazo Fijo Pactado	0,000
Tipo_Retiro_Justa Causa	0,941
Tipo_Retiro_Mutuo Acuerdo	0,057
Tipo_Retiro_No se presentó a firma de contrato	0,000
Tipo_Retiro_Renuncia Voluntaria	0,000
Rango_Laboral_De 3 a 6 Meses	0,007
Rango_Laboral_Mayor a 6 Meses	0,977
Rango_Edad_18-30 años	0,590
Rango_Edad_30-40 años	0,287
Rango_Edad_40-60 años	0,104
Rango_Edad_Mayor a 60 años	0,017

Este clúster contiene 1184 empleados y se caracteriza por tener una alta proporción de empleados en la unidad de negocio "Retail" (97.64%) y en la agrupación de cargos "Auxiliares Operativos" (81.17%). La mayoría de los empleados en este clúster tienen contratos a término fijo de menos de 3 meses (98.82%). En cuanto al género, la proporción es mayoritariamente masculina, con un 66.13% de hombres y un 0% de empleados de otro género. El tipo de retiro más común en este clúster es la "Justa Causa" (94.09%). En términos de experiencia laboral, la mayoría de los empleados han estado en la empresa por más de 6 meses (97.72%). En cuanto a la edad, se observa una mayor presencia de empleados en el rango de edad de 18 a 30 años (59.04%)

**Tabla 9. Descripción de Clúster 7.**

<b>Clúster</b>	<b>7</b>
Cantidad	1008
ETC	0,889
Cargo_lider	0,005
Tipo_Contrato_Termino Fijo < 3 Meses	0,000
Tipo_Contrato_Termino fijo >= 3 Meses	0,933
Unidad_negocio_Industria	0,005
Unidad_negocio_Inmobiliaria	0,000
Unidad_negocio_Retail	0,992

Unidad_negocio_Servicios	0,003
Agrupacion_cargo_Auxiliares Especializados	0,094
Agrupacion_cargo_Auxiliares Operativos	0,896
Agrupacion_cargo_Direccion Media	0,000
Agrupacion_cargo_Ejecutivos(a)	0,000
Agrupacion_cargo_Lideres	0,005
Agrupacion_cargo_Profesionales	0,000
Género_M	0,469
Género_O	0,001
Tipo_Retiro_Expiración De Plazo Fijo Pactado	0,003
Tipo_Retiro_Justa Causa	0,148
Tipo_Retiro_Mutuo Acuerdo	0,001
Tipo_Retiro_No se presentó a firma de contrato	0,000
Tipo_Retiro_Renuncia Voluntaria	0,821
Rango_Laboral_De 3 a 6 Meses	0,278
Rango_Laboral_Mayor a 6 Meses	0,005
Rango_Edad_18-30 años	0,000
Rango_Edad_30-40 años	0,998
Rango_Edad_40-60 años	0,000
Rango_Edad_Mayor a 60 años	0,001

Este clúster contiene 1008 empleados y se caracteriza por tener una alta proporción de empleados en la unidad de negocio "*Retail*" (99.21%) y en la agrupación de cargos "Auxiliares Operativos" (89.58%). La mayoría de los empleados en este clúster tienen contratos a término fijo de al menos 3 meses (93.25%). En términos de género, hay una proporción equilibrada, con un 46.92% de hombres y un 0.1% de empleados de otro género. El tipo de retiro más común en este clúster es la "Renuncia Voluntaria" (82.14%). En cuanto a la experiencia laboral, se observa que un número significativo de empleados ha estado en la empresa entre 3 y 6 meses (27.78%). En cuanto a la edad, la mayoría de los empleados se encuentran en el rango de 30 a 40 años (99.8%)

**Tabla 10. Descripción de Clúster 8.**

<b>Clúster</b>	<b>8</b>
Cantidad	847
ETC	0,944
Cargo_lider	0,153
Tipo_Contrato_Termino Fijo < 3 Meses	0,000
Tipo_Contrato_Termino fijo >= 3 Meses	0,183
Unidad_negocio_Industria	0,041
Unidad_negocio_Inmobiliaria	0,001
Unidad_negocio_Retail	0,924
Unidad_negocio_Servicios	0,020
Agrupacion_cargo_Auxiliares Especializados	0,125
Agrupacion_cargo_Auxiliares Operativos	0,651
Agrupacion_cargo_Direccion Media	0,013
Agrupacion_cargo_Ejecutivos(a)	0,001
Agrupacion_cargo_Lideres	0,133
Agrupacion_cargo_Profesionales	0,021
Género_M	0,384
Género_O	0,000
Tipo_Retiro_Expiración De Plazo Fijo Pactado	0,000
Tipo_Retiro_Justa Causa	0,019
Tipo_Retiro_Mutuo Acuerdo	0,118
Tipo_Retiro_No se presentó a firma de contrato	0,001
Tipo_Retiro_Renuncia Voluntaria	0,844
Rango_Laboral_De 3 a 6 Meses	0,043
Rango_Laboral_Mayor a 6 Meses	0,770
Rango_Edad_18-30 años	0,000
Rango_Edad_30-40 años	0,000
Rango_Edad_40-60 años	0,999
Rango_Edad_Mayor a 60 años	0,001

Este clúster contiene 847 empleados y se caracteriza por tener una alta proporción de empleados en la unidad de negocio "*Retail*" (92.44%) y en la agrupación de cargos "Auxiliares Operativos" (65.05%). Algunas características distintivas de este clúster incluyen la presencia de empleados en cargos de liderazgo (13.34%) y profesionales (2.13%). En cuanto al tipo de contrato, una proporción significativa de empleados tiene contratos a término fijo de al menos 3 meses (18.3%). En términos de género, el 38.37% de los empleados son hombres. El tipo de

retiro más común en este clúster es la "Renuncia Voluntaria" (84.42%). En cuanto a la experiencia laboral, se observa que un número considerable de empleados ha estado en la empresa durante más de 6 meses (76.98%). En términos de edad, la mayoría de los empleados se encuentran en el rango de 40 a 60 años (99.88%).

**Tabla 11. Descripción de Clúster 9.**

<b>Clúster</b>	<b>9</b>
Cantidad	1499
ETC	0,839
Cargo_lider	0,000
Tipo_Contrato_Termino Fijo < 3 Meses	1,000
Tipo_Contrato_Termino fijo >= 3 Meses	0,000
Unidad_negocio_Industria	0,000
Unidad_negocio_Inmobiliaria	0,000
Unidad_negocio_Retail	1,000
Unidad_negocio_Servicios	0,000
Agrupacion_cargo_Auxiliares Especializados	0,000
Agrupacion_cargo_Auxiliares Operativos	1,000
Agrupacion_cargo_Direccion Media	0,000
Agrupacion_cargo_Ejecutivos(a)	0,000
Agrupacion_cargo_Lideres	0,000
Agrupacion_cargo_Profesionales	0,000
Género_M	0,378
Género_O	0,002
Tipo_Retiro_Expiración De Plazo Fijo Pactado	0,929
Tipo_Retiro_Justa Causa	0,065
Tipo_Retiro_Mutuo Acuerdo	0,000
Tipo_Retiro_No se presentó a firma de contrato	0,000
Tipo_Retiro_Renuncia Voluntaria	0,000
Rango_Laboral_De 3 a 6 Meses	0,003
Rango_Laboral_Mayor a 6 Meses	0,003
Rango_Edad_18-30 años	0,698
Rango_Edad_30-40 años	0,207
Rango_Edad_40-60 años	0,065
Rango_Edad_Mayor a 60 años	0,005

Este clúster contiene 1,499 empleados y se caracteriza por tener una alta proporción de empleados en la unidad de negocio "*Retail*" (100%) y en la agrupación de cargos "Auxiliares Operativos" (100%). Todos los empleados en este clúster tienen contratos a término fijo de menos de 3 meses (100%). En cuanto al género, el 37.76% de los empleados son hombres y hay una pequeña proporción de empleados con género "Otro" (0.2%). El tipo de retiro más común en este clúster es la "Expiración De Plazo Fijo Pactado" (92.93%). En términos de experiencia laboral, la mayoría de los empleados tienen una experiencia laboral de 3 a 6 meses (0.33%) o mayor a 6 meses (0.27%). En cuanto a la edad, la mayoría de los empleados se encuentran en el rango de 18 a 30 años (69.78%).

## 10. Recomendaciones

- Al analizar la gran cantidad de datos recopilados para lograr el objetivo propuesto, se observó que el estudio arroja resultados muy generales en cuanto a los perfiles o patrones asociados a la rotación de los empleados en el Grupo Éxito. Por este motivo, se recomienda llevar a cabo este modelo de manera más controlada, ya sea en un lugar de trabajo específico, una zona geográfica determinada o una unidad de negocio con el fin de obtener datos más concretos y precisos.
- Desarrollar un modelo de aprendizaje supervisado con el objetivo de predecir las posibles causas de retiro de los empleados basándose en las condiciones y características individuales de cada empleado.

En esta recomendación, se sugiere la implementación de un modelo de aprendizaje supervisado, una técnica de *Machine Learning* que utiliza un conjunto de datos etiquetados para entrenar y aprender patrones, con el fin de predecir las causas de retiro de los empleados. Al utilizar este enfoque, el modelo puede analizar las diferentes condiciones y características de los empleados, como el tipo de contrato, la unidad de negocio, la posición jerárquica, el género, el rango laboral y el rango de edad, entre otros factores relevantes.

Al desarrollar este modelo, se puede obtener un sistema capaz de predecir con cierta precisión las posibles causas de retiro de los empleados, lo que puede ser de gran utilidad para la gestión de recursos humanos y la toma de decisiones estratégicas. Por ejemplo, la organización podría identificar patrones y factores de riesgo asociados con ciertas causas de retiro, lo que permitiría implementar medidas preventivas o estrategias de retención de talento más efectivas.

Además, este modelo también podría servir como una herramienta de apoyo en la toma de decisiones relacionadas con la asignación de recursos, la planificación de

sucesiones y la identificación de áreas de mejora dentro de la organización.

- Llevando a cabo un exhaustivo análisis de los patrones asociados a la rotación del personal en el Grupo Éxito, y luego analizar detalladamente los datos, se logró identificar que la mayor cantidad de empleados dentro de la organización se encuentra en la unidad de "Retail". Además, se destaca que muchos de estos empleados tienen contratos a término "Fijos". Este hallazgo es relevante, ya que el tipo de contrato puede tener un impacto significativo en la retención y la satisfacción laboral de los colaboradores.

Con base en los resultados obtenidos, se recomienda a la organización que revise y considere la posibilidad de implementar un cambio en los tipos de contratos utilizados para contratar a los empleados de la unidad de "Retail".

En particular, se sugiere que evalúen la viabilidad de ofrecer contratos a término indefinido a los empleados. El contrato a término indefinido brinda mayor estabilidad laboral y seguridad a los trabajadores, lo que a su vez puede tener un impacto positivo en su compromiso y lealtad hacia la compañía. Además, este tipo de contrato puede ser percibido como un beneficio atractivo por parte de los empleados, lo que les ayudaría a atraer y retener talento de calidad en la organización.

## 11. Referencias

- Alles, M. A. (2003). *Elija la mejor. Como entrevistar por competencias*. Ediciones Granica S.A.
- Alles, M. A. (2006). *Selección por competencias*. Buenos Aires: Ediciones Granica SA.
- Asif, S., & Sargeant, A. (2000). *Modelling internal communications in the financial services sector* (Vol. 34). *European Journal of Marketing*. <https://doi.org/10.1108/03090560010311867>
- AWS. (s.f.). *¿Qué es Python? - Explicación del lenguaje Python - AWS*. ¿Qué es Python? - Explicación del lenguaje Python - AWS: <https://aws.amazon.com/es/what-is/python/>
- Beltrán Martínez, B. (2001). *Minería de datos* (Vol. 14). Cómo hallar una aguja en un pajar. Ingenierías.
- Blasco López, M. F., Rodríguez Tarodo, A., & Fernández Lores, S. (2014). *Employer branding: estudio multinacional sobre la construcción de la marca del empleador*. Madrid: Portal Universia S.A.
- Chakraborty, S., & Chatterjee, S. (2018). *Employee Turnover Prediction using Machine Learning: A Case Study on Indian IT Industry*. *IEEE 3rd International Conference on Computing, Communication and Security (ICCCS)*. <https://doi.org/10.1109/CCCS.2018.8664235>
- Chamorro-Premuzic, T., Akhtar, R., Winsborough, D., & Sherman, R. A. (2017). *The datafication of talent: how technology is advancing the science of human potential at work* (Vol. 18). *Current Opinion in Behavioral Sciences*. <https://doi.org/10.1016/j.cobeha.2017.04.007>
- Chiavenato, I. (2001). *Administración de Recursos Humanos*. Madrid: McGraw-Hill.
- Chiavenato, I. (2008). *Gestión del talento humano*. Mc-Graw Hill.
- Dessler, G. (2001). *Administración de Personal (8ª ED.)*. Mexico: Alhambra Mexicana.
- Edwards, M. R. (2010). *An integrative review of employer branding and OB theory* (Vol. 39). *Personnel Review*.
- Gil, D. Á. (14 de 01 de 2021). *Adictos al Trabajo*. Adictos al Trabajo: <https://www.adictosaltrabajo.com/2021/01/14/metodologia-crisp-dm/>

- Huselid, M. A. (2018). *The science and practice of workforce analytics: Introduction to the HRM special issue* (Vol. 57). Human Resource Management.
- IBM. (17 de 08 de 2021). *Conceptos básicos de ayuda de CRISP-DM*. Conceptos básicos de ayuda de CRISP-DM: <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>
- Krause, T., Seymour, B., & Pandya, R. (2019). *Prediction of Employee Turnover in Organizations* (Vol. 104). Journal of Business Research. <https://doi.org/10.1016/j.jbusres.2019.06.034>
- Krishnan, K. (2013). *Data warehousing in the age of big data*. Morgan Kaufmann Series on Business Intelligence. Morgan Kaufmann Series on Business Intelligence. <https://doi.org/10.1016/C2012-02737-8>
- Lin, H., Wang, Q., Guo, P., Guan, T., & Xu, L. (s.f.).
- Python en la nube con Google Colab – Agrega*. (21 de 07 de 2021). Python en la nube con Google Colab – Agrega: <https://www.agrega.com/blog/general/python-en-la-nube-con-google-colab/>
- Russo, C. C., Ramón, H. D., Alonso, N., Cicerchia, B., Esnaola, L., & Tessore, J. P. (2016). *Tratamiento masivo de datos utilizando técnicas de Machine Learning*. Computer Science.
- Sodhi, J. S., & Sonnenburg, S. (2017). *Identifying and Understanding Employee Turnover Drivers using Random Forest and Gradient Boosted Trees*. Proceedings of the 50th Hawaii International Conference on System Sciences. <https://doi.org/10.24251/HICSS.2017.435>
- Treviño, J. G. (2000). *Administración Contemporánea*. México: Mc Graw Hill.