

# Análisis de retención de clientes en el sector financiero mediante minería de datos

Brayan David Reyes Morales<sup>1</sup>, Nicolás Alejandro Fernández Espinosa<sup>2</sup>

Universidad Central  
Maestría en Analítica de Datos  
Curso de Bases de Datos  
Bogotá, Colombia  
breyesm@ucentral.edu.co, nfernandez@ucentral.edu.co

May 15, 2023

## Contents

<b>1</b>	<b>Introducción</b>	<b>3</b>
<b>2</b>	<b>Características del proyecto de investigación que hace uso de Bases de Datos</b>	<b>4</b>
2.1	Titulo del proyecto de investigación . . . . .	5
2.2	Objetivo general . . . . .	5
2.2.1	Objetivos específicos . . . . .	5
2.3	Alcance . . . . .	5
2.4	Pregunta de investigación . . . . .	5
2.5	Hipótesis . . . . .	6
<b>3</b>	<b>Reflexiones sobre el origen de datos e información</b>	<b>7</b>
3.1	¿Cual es el origen de los datos e información? . . . . .	7
3.2	¿Cuales son las consideraciones legales o éticas del uso de la información? . . . . .	7
3.3	¿Cuales son los retos de la información y los datos que utilizara en la base de datos en términos de la calidad y la consolidación? . . . . .	8
3.4	¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto? (Max 100 Palabras) - ( <i>Primera entrega</i> ) . . . . .	10
<b>4</b>	<b>Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)(<i>Primera entrega</i>)</b>	<b>12</b>
4.1	Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto ( <i>Primera entrega</i> ) . . . . .	12

4.2	Diagrama modelo de datos ( <i>Primera entrega</i> ) . . . . .	13
4.3	Imágenes de la Base de Datos ( <i>Primera entrega</i> ) . . . . .	14
4.4	Código SQL - lenguaje de definición de datos (DDL) ( <i>Primera entrega</i> ) . . . . .	15
4.5	Código SQL - Manipulación de datos (DML) ( <i>Primera entrega</i> ) . .	17
4.6	Código SQL + Resultados: Vistas ( <i>Primera entrega</i> ) . . . . .	19
4.7	Código SQL + Resultados: Triggers ( <i>Primera entrega</i> ) . . . . .	20
4.8	Código SQL + Resultados: Funciones ( <i>Primera entrega</i> ) . . . . .	21
4.9	Código SQL + Resultados: procedimientos almacenados ( <i>Primera entrega</i> ) . . . . .	21
<b>5</b>	<b>Bases de Datos No-SQL (<i>Segunda entrega</i>)</b>	<b>22</b>
5.1	Diagrama Bases de Datos No-SQL ( <i>Segunda entrega</i> ) . . . . .	22
5.2	SMBD utilizado para la Base de Datos No-SQL ( <i>Segunda entrega</i> )	22
<b>6</b>	<b>Bibliografía</b>	<b>24</b>

# 1 Introducción

La motivación principal en la creación y mantenimiento de las empresas, (sin importar el sector a que pertenezca) ha sido el relacionamiento con los clientes ya que ellos son los generadores de ingresos para las compañías. Aunque la ecuación de ingresos a partir de clientes puede ser simple y lógica es olvidada dado a que se prioriza la obtención de clientes nuevos, pero no se prioriza los niveles de deserción en las compañías.

Actualmente por la globalización de servicios y la competencia, los productos en demanda son cada vez más similares en calidad y precio [1], [2], ante esto las compañías se vuelven dinámicas en función de los clientes, por ello que se hace indispensable conocer a los clientes, con el fin de explicar y anticiparse a sus acciones en el futuro.

Al contar con una amplia oferta de productos y servicios, los clientes tienen la facilidad de cambiar de una compañía a otra, por ello las entidades priorizan la alta competencia en cuanto la atracción de clientes nuevos, pero no es la mejor manera de solucionar la pérdida de clientes, por ello se busca un relacionamiento orientado a fidelizar los clientes correcta y responsablemente, según los productos y servicios ofrecidos por industria. Aquellos clientes que dejan de usar los productos de una empresa se denominan desertores, al identificarlos permite aplicar estrategias de retención [3], por lo que se hace indispensable conocer el entorno de las compañías y así generar propuestas de retención efectivas, traducida en una deserción menor y una masa de clientes superior.

Dado a que la obtención de clientes ha cambiado dinámicamente, ahora se piensa en la retención de una manera efectiva, es demostrado que en la actualidad las industrias implementan más recursos a la atracción, mientras que las estrategias de retención son mejores en cuanto a costos, algunas investigaciones nos muestran que el costo de obtener de un nuevo cliente es de 5 a 7 veces mayor al costo de retención de uno antiguo [4], “The loyalty effect” donde se demostró que el aumento del 5% en la tasa de retención de clientes logró aumentos del 35% y el 95% en el valor actual neto de los clientes en una empresa desarrolladora de software y una agencia de publicidad, respectivamente. [5]

Por los motivos anteriormente se propone un modelo estadístico para identificar los comportamientos de los clientes y lograr un modelo de retención que prediga el nivel de deserción y esté enfocado a usuarios del sistema financiero, apoyándose en modelos estudiados e implementados en diferentes trabajos y/o investigaciones similares.

Para abordar el problema se deben definir los criterios de evaluación (variables de dataset), ante estos criterios el más importante es cuando un cliente se marca como desertor dentro de la compañía, el concepto de clientes desertores es cambiante según los comportamientos de estos en las diferentes industrias y

tipos de mercado a los que hoy se exponen los consumidores. Para el sector bancario, un cliente desertor es aquel que cierra todas sus cuentas bancarias y cesa los negocios con el banco en estudio [6].

Se abordan los datos disponibles y se hace indispensable la minería para el caso de estudio, ya que a partir de este punto se usa la mayoría de tiempo en implementación y lograr un modelo favorable. La minería de datos es una técnica estadística considerada como una herramienta estratégica importante para las empresas, debido a que permite analizar grandes volúmenes de información desde diferentes perspectivas y sintetizarlos en información valiosa que permite la temprana y efectiva toma de decisiones de las organizaciones empresariales, sirviendo de gran utilidad para crear un perfil de cliente preciso basado en el comportamiento de este [7].

La minería de datos aporta al igual información comprimida, para entender lo que pasa o sucedió. El modelamiento de los datos es importante en las compañías ya que genera información concisa y precisa para aplicar modelos estadísticos. Los modelos estadísticos para aplicar a este caso y los estudios previamente realizados por diferentes autores, son algoritmos de programación de los modelos como: Árboles de decisión, Redes Neuronales, Reglas de Asociación, Regresión Logística, Árboles Aleatorios y SVM.

## **2 Características del proyecto de investigación que hace uso de Bases de Datos**

El estudio se realiza para una entidad del sector financiero en Colombia, aunque los modelos son aplicables a diferentes industrias y sectores, no todas las empresas son iguales y tienen detalles que las hacen únicas como a sus clientes. Como para el caso de estudio se basa en datos de una financiera, se debe definir de dónde se van obtener y el periodo de los mismos, como se evidencia en la tabla 1, se pueden tomar de una cantidad específica definiendo a los clientes retirados, una segunda opción es tomar a los clientes totales de los últimos meses, no mayor a 6 meses, e identificar los retirados. En el caso de estudio se toman los clientes de tres meses específicos, con su variable categórica (retirados), calidad, y los tipos de datos que se generen a partir de la base entregada.

A partir de las fuentes disponibles en la organización, se procede con el modelo CRISP-DM, tratando de responder la pregunta de negocios “¿Qué tipo de clientes tienen una alta probabilidad de retirarse en el futuro?”, para ello se ha generado una base con diferentes variables, que ayudarán a explicar esta pregunta.

## **2.1 Título del proyecto de investigación**

Análisis de retención de clientes en el sector financiero mediante minería de datos.

## **2.2 Objetivo general**

Identificar comportamientos de los clientes en el sistema financiero y lograr un modelo de retención exitoso.

### **2.2.1 Objetivos específicos**

- Definir los criterios de evaluación y variables del dataset, en especial la definición de un cliente desertor en el sector bancario.
- Realizar un análisis exploratorio de los datos para identificar patrones y tendencias en el comportamiento de los clientes.
- Estimar un modelo estadístico que permita predecir el nivel de deserción de clientes en una entidad financiera.
- Evaluar la efectividad del modelo propuesto y comparar el costo de obtener un nuevo cliente versus el costo de retener a uno.

## **2.3 Alcance**

Con la correcta extracción y manipulación de los datos se busca crear una base de datos solida que permita extraer información financiera y sociodemográfica de clientes retirados en una entidad financiera, mediante el uso de sistemas informáticos como SQL al tener datos estructurados.

Obtener un modelo que permita explicar y entender un cliente que se retira de una compañía. Al no haber estudios anteriores en el negocio específico se inicia con un dataset medianamente robusto, pero con variables que posiblemente ayuden a entender al cliente. Claramente este modelo se puede robustecer con la integración y salida de variables, es importante recalcar que se toman variables explicativas de negocio como saldos productos y conexiones del cliente con el entorno del negocio, pero se pueden agregar las variables de experiencias del cliente con los productos (actualmente no mapeadas), lo que ayudaría a generar la cultura de medición de los sentimientos y generación de datos a partir de cualquier comportamiento del cliente.

## **2.4 Pregunta de investigación**

¿Se puede predecir y retener a clientes desertores en una entidad financiera en Colombia mediante el relacionamiento de bases dedatos?

## **2.5 Hipótesis**

La minería de datos funciona para identificar patrones de comportamiento en los clientes de una entidad financiera, que pueden desarrollar estrategias de retención efectivas para reducir la tasa de deserción y aumentar el valor actual neto de los clientes.

### **3 Reflexiones sobre el origen de datos e información**

Se espera que una entidad tenga sus datos ubicados en un solo lugar y con poca información faltante, sin embargo, muchas veces en las compañías esto no ocurre, ya sea en tiempo real o tiempo vencido. Para este ejercicio en particular, la recopilación y acceso a la información puede estar limitada por otras áreas, ya que existe cierta precaución en el manejo de transacciones y acceso a los archivos.

Una vez superado el problema de los permisos, surge otro obstáculo al identificar las variables relevantes para generar una base de datos, ya que muchas de ellas manejan información parecida o repetida en tablas diferentes. Esto provoca molestias y un desgaste de tiempo en la minería de datos, especialmente si los datos se distribuyen en múltiples tablas. Una vez organizada la información, se puede crear una tabla de cinco variables conformadas por una variable de cada tabla, lo cual puede ser un gran obstáculo en las entidades, ya que se deja de lado el modelamiento y se requiere de un esfuerzo significativo para la creación de una base comprensible y análisis descriptivo, con la esperanza de que las variables aporten información sólida.

#### **3.1 ¿Cual es el origen de los datos e información?**

La información proviene de transacciones y creaciones de clientes que se guardan en repositorios diferentes. En general los repositorios son archivos en Excel e información recopilada a través del programa CRM.

#### **3.2 ¿Cuales son las consideraciones legales o éticas del uso de la información?**

Para el tratamiento de datos que se aplica en el presente trabajo, se hace una solicitud formal al área de gobierno de datos, solicitando el uso de las variables expuestas. Para lograr la autorización correcta se realizó el anonimato de las identificaciones de los clientes.

En cuanto al riesgo que se puede presentar sin la previa autorización se detalla gracias a la ley estatutaria 1581 del 2012, en el cual se presentan las siguientes restricciones por parte de las personas y las empresas que permiten la utilización de los datos "sensibles" que suministra una persona a una entidad.

Artículo 9°. Autorización del Titular. Sin perjuicio de las excepciones previstas en la ley, en el Tratamiento se requiere la autorización previa e informada del Titular, la cual deberá ser obtenida por cualquier medio que pueda ser objeto de consulta posterior.

Artículo 10. Casos en que no es necesaria la autorización. La autorización del Titular no será necesaria cuando se trate de:

- a) Información requerida por una entidad pública o administrativa en ejercicio de sus funciones legales o por orden judicial;
- b) Datos de naturaleza pública;
- c) Casos de urgencia médica o sanitaria;

- d) Tratamiento de información autorizado por la ley para fines históricos, estadísticos o científicos;
  - e) Datos relacionados con el Registro Civil de las Personas.
- Quien acceda a los datos personales sin que medie autorización previa deberá en todo caso cumplir con las disposiciones contenidas en la presente ley.[19]

### 3.3 ¿Cuales son los retos de la información y los datos que utilizara en la base de datos en términos de la calidad y la consolidación?

Tomando las variables identificadas en el diccionario se procede a identificar la calidad de estos, buscando que las variables cumplan con los estándares mínimos para el montaje de un modelo estadístico, se obtiene lo siguiente:

Nombre	Columna	Tipo	Cantidad de Únicos	Cantidad de Nulos	Cantidad Valores No Nulos
OFICINA	OFICINA	object	37	0	46144
TIPO	TIPO	object	2	0	46144
SEGMENTO	SEGMENTO	object	4	0	46144
RETIRARON	RETIRARON	object	2	0	46144
NICHO	NICHO	object	12	0	46144
CANAL	CANAL	object	6	0	46144
OCUPACION	OCUPACION	object	6	0	46144
SALDO_CAPTA	SALDO_CAPTA	float64	25024	0	46144
SALDO_COLOCA	SALDO_COLOCA	float64	23294	0	46144
NUM_PRODUCTOS	NUM_PRODUCTOS	int64	28	0	46144
PAGADURIA	PAGADURIA	object	291	20025	26119
ESTADO_CIVIL	ESTADO_CIVIL	object	6	5166	40978
ESTRATO	ESTRATO	int64	6	0	46144
EDAD	EDAD	float64	83	1273	44871
GENERO	GENERO	object	2	811	45333
VINCULO	VINCULO	object	3	0	46144
INGRESOS	INGRESOS	object	16005	665	45479
PRODUCTO	PRODUCTO	object	16	0	46144

Con la ayuda de Google Colab se visualiza el tipo de dato, el número de valores únicos y la cantidad de valores nulos. Se observa que el dataset tiene un tamaño de 46.144 x 18. En donde se tienen 6 variables cuantitativas y 12 variables cualitativas, de las cuales las variables de “PAGADURIA, ESTADO CIVIL, EDAD, GENERO e INGRESOS” presentan valores faltantes, es por ello por lo que la variable de “INGRESOS” no se toma ni se puede convertir a tipo numérica ya que presenta valores faltantes y esto genera un error en la identificación del tipo de dato.

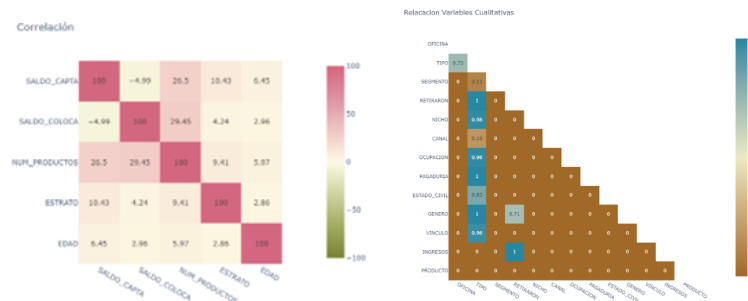


	SALDO_CAPTA	SALDO_COLOCA	NUM_PRODUCTOS	ESTRATO	EDAD
<b>count</b>	4.614400e+04	4.614400e+04	46144.00	46144.00	44871.00
<b>mean</b>	1.273259e+07	1.744456e+07	1.79	2.79	54.31
<b>std</b>	7.850941e+07	3.752638e+07	1.30	1.04	16.12
<b>min</b>	-1.214730e+05	0.000000e+00	1.00	1.00	13.00
<b>25%</b>	0.000000e+00	0.000000e+00	1.00	2.00	41.00
<b>50%</b>	1.005045e+05	2.085620e+06	1.00	3.00	56.00
<b>75%</b>	2.317506e+06	1.996970e+07	2.00	3.00	66.00
<b>max</b>	4.712371e+09	1.500000e+09	36.00	6.00	95.00

Generando un análisis de las variables numéricas se puede deducir que las variables de saldos necesitan una normalización, modificándolas con logaritmos o exponenciales, ya que sus valores al ser altos generan un poco compresión de los mismos, por otro lado los clientes de la entidad estudiada presentan una edad media de 54 años, un estrato 3 y productos adquiridos de 2. Hay que tener en cuenta que cuando se habla de saldos se debe tener presente que algunos pueden ser muy altos y otros bajos, dado el siguiente se puede estudiar la posibilidad de recortar los datos con medias recortadas, solo para dichas variables.



Con la generación de los gráficos de frecuencia se puede deducir que la base necesita ser balanceada y buscar que variables se les debe realizar una imputación de datos y otras solo perder la información, se debe tener cuidado ya que si se asume la pérdida de datos de una variable se puede estar eliminando información de los clientes retirados y así quitando la oportunidad de que los modelos no puedan aprender sobre los clientes retirados.



Se generan gráficos de correlación para las variables en donde las variables cualitativas en su mayoría se explican las unas a las otras, la variable TIPO, la cual contiene el tipo de cliente si es activo en sus saldos es la que no tiene una explicación frente a las demás, esta variable puede ser retirada ya que se tenía únicamente para generar la base con la que se genera el modelamiento, esto nos da una buena señal para plantear un modelo estadístico, en cuanto a las variables categóricas.

Por otro lado, las variables numéricas no presentan correlación es fuertes entre ellas, pero no se puede deducir que se tengan que dejar a un lado, ya que no se han realizado estudios anteriores en la compañía, al tener estos valores nos permite tener un punto inicial de como se comportan las variables dentro de la compañía y así generar o mejorar los estudios posteriores.

3.4 ¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto? (Max 100 Palabras) - (*Primera entrega*)

Con el presente ejercicio se espera obtener un modelo que permita explicar y entender un cliente que se retira de una compañía. Como no se han hecho estudios anteriores en el negocio específico se inicia con un dataset medianamente robusto, pero con variables que posiblemente ayudaran a entender al cliente. Claramente este modelo se puede robustecer con la integración y salida de variables, es importante recalcar que se están tomando variables explicativas de negocio como saldos productos y conexiones del cliente con el entorno del negocio, pero se pueden agregar las variables de experiencias del cliente con los productos (actualmente no mapeadas), ayudaría a generar la cultura de medición de los sentimientos y generación de datos a partir de cualquier comportamiento del cliente.

Con la serie de modelos mencionados en el estudio se busca dar un valor a agregado a una gestión y no tan solo un resultado. Es importante el resultado por que a partir de ello se tomaran decisiones, pero el resultado puede ser cambiante con cada modelo aplicado, el valor adicional se agrega cuando en la practica sin importar el resultado se pueden aplicar varias fórmulas “comerciales”

apoyadas de la estadística que permitan aumentar la fidelidad de un cliente generando una confianza de marca, teniendo como base principal el análisis de datos.

## 4 Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos) *(Primera entrega)*

### 4.1 Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto *(Primera entrega)*

Para el trabajo actual se utilizará MySQL el cual permite modificar y generar tablas que permiten realizar la conexión a través de llaves, para el caso se tiene una llave única que será el número de identificación del usuario.

MySQL Workbench es un sistema libre que permite trabajar con una cantidad limitada de datos, pero para el proyecto actual es suficiente ya que posee las siguientes características:

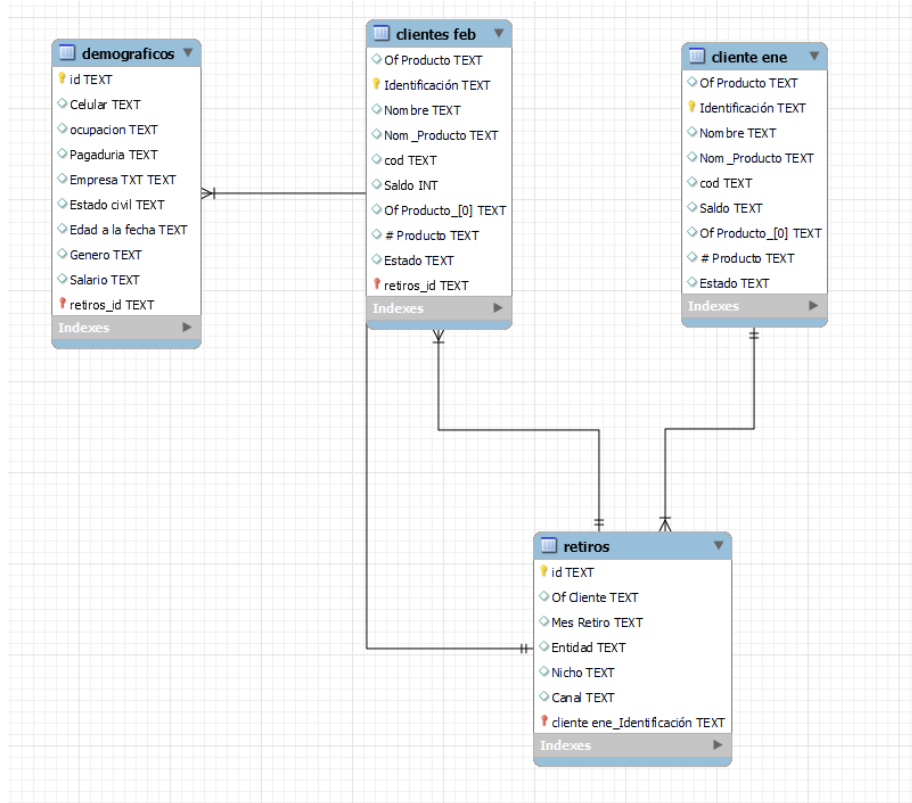
- Disponibilidad en gran cantidad de plataformas y sistemas operativos.
- Soporta gran cantidad de datos y posee diferentes opciones para su almacenamiento.
- Transacciones y claves primarias y foráneas.
- Conectividad segura.

Como segunda opción se cuenta con el programa de Oracle SQL para suplir la contingencia de datos pesados (registros mayores de 200.000) para la generación de las tablas de la base de datos.

- Oracle SQL es un sistema de gestión de bases de datos relacionales (RDBMS) que proporciona una plataforma sólida y escalable para el almacenamiento, gestión y procesamiento de grandes cantidades de datos. Algunas de las características y ventajas de Oracle SQL incluyen:

- Escalabilidad: Oracle SQL puede manejar grandes cantidades de datos y es altamente escalable, lo que significa que puede manejar el aumento de datos y usuarios sin comprometer el rendimiento.
- Fiabilidad: Oracle SQL es conocido por su alta disponibilidad y confiabilidad, lo que lo hace ideal para aplicaciones críticas de misión crítica.
- Seguridad: Oracle SQL ofrece un alto nivel de seguridad para proteger los datos almacenados, incluyendo funciones de autenticación, autorización y cifrado de datos.
- Rendimiento: Oracle SQL está optimizado para ofrecer un rendimiento rápido y eficiente, lo que lo hace ideal para aplicaciones que requieren un acceso rápido a grandes cantidades de datos.
- Funcionalidad avanzada: Oracle SQL ofrece una amplia gama de funcionalidades avanzadas, como particionamiento de tablas, indexación avanzada, gestión de transacciones y más, lo que lo hace ideal para aplicaciones empresariales complejas.
- Soporte de la comunidad: Oracle SQL tiene una gran comunidad de usuarios y desarrolladores que ofrecen soporte y recursos para ayudar a los usuarios a resolver problemas y aprender más sobre el sistema.

## 4.2 Diagrama modelo de datos (*Primera entrega*)



### 4.3 Imágenes de la Base de Datos (*Primera entrega*)

Query 1 x DML DDL

Limit to 1000 rows

```

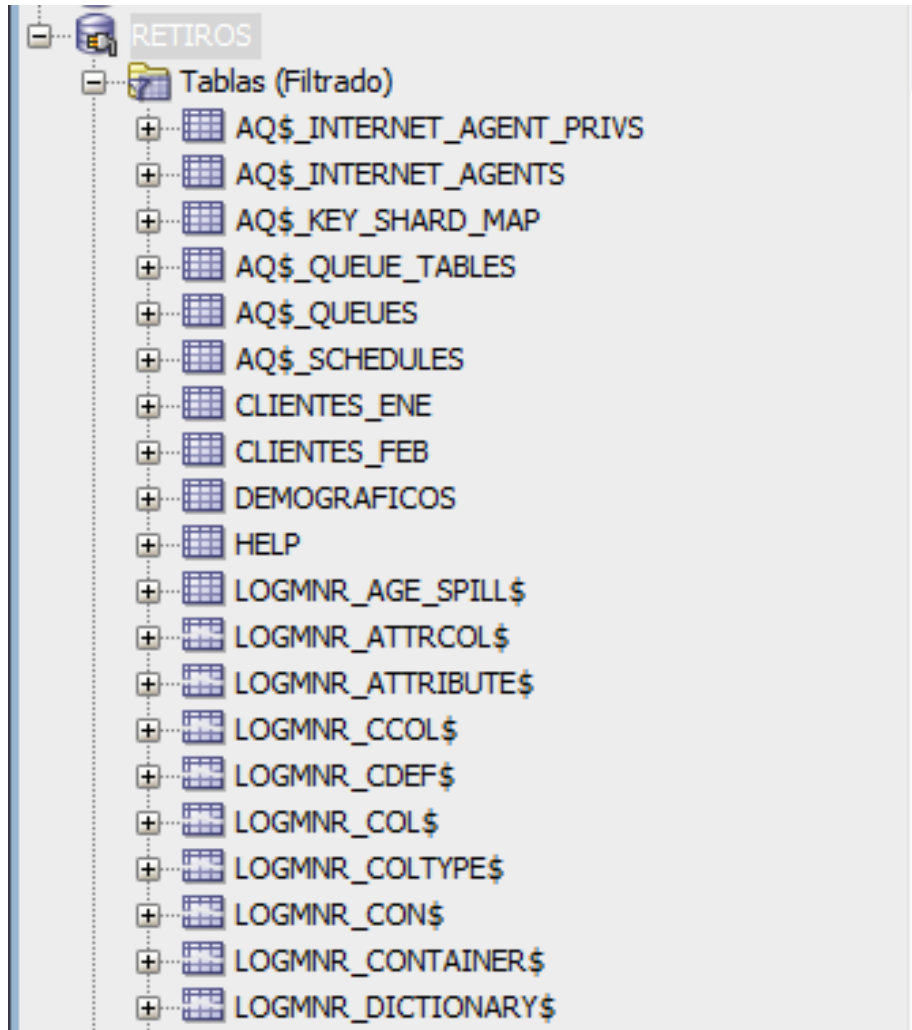
1 • select * from demograficos;
2 • select * from retiros;

```

Result Grid Filter Rows: Export: Wrap Cell Content: Fetch rows:

	id	Of Cliente	Mes Retiro	Entidad	Nicho	Canal
▶	196600	Bogota - Galerías	nov-22	Colpensiones	Pensionados	E-Credit
	588574	Duitama	nov-22	Empleado Empresa Privada	Empleado Empresa Privada	Oficinas
	1098375	Duitama	nov-22	Colpensiones	Pensionados	Oficinas
	1279325	Pereira	nov-22	Colpensiones	Pensionados	E-Credit
	2295624	Ibague	nov-22	Colpensiones	Pensionados	E-Credit
	2909718	San Gil	nov-22	Fopep	Pensionados	Oficinas
	3069502	Bogota - Galerías	nov-22	Rama Judicial	Sistema Nacional de Justicia	Oficinas
	3171460	Bogota - Centro	nov-22	Militares y Policías	Pensionados	Oficinas
	4313498	Armenia	nov-22	Otros Segmentos **	Otros Segmentos **	Oficinas
	4322090	Manizales	nov-22	Colpensiones	Pensionados	E-Credit
	4327460	Manizales	nov-22	Colpensiones	Pensionados	E-Credit
	4378988	Pereira	nov-22	Otros Pensionados	Pensionados	Oficinas
	4508852	B'bermeja	nov-22	Colpensiones	Pensionados	E-Credit
	4627172	Popayan	nov-22	Independientes	Independientes	Oficinas
	5149854	Riohacha	nov-22	Empleado Empresa Privada	Empleado Empresa Privada	Oficinas
	5202301	Bogota - Centro	nov-22	Militares y Policías	Pensionados	E-Credit
	5588376	B'lmanga	nov-22	Fiduprevisora (Magisterio)	Pensionados	E-Credit
	5765303	Socorro	nov-22	Independientes	Independientes	Oficinas
	6092561	Calli	nov-22	Colpensiones	Pensionados	E-Credit
	6243747	Pitalito	nov-22	Fiduprevisora (Magisterio)	Pensionados	Oficinas
	6759758	Tunja	nov-22	Independientes	Independientes	Oficinas
	6771669	Tunja	nov-22	Rama Judicial	Sistema Nacional de Justicia	Oficinas
	7162534	Tunja	nov-22	Empleado Empresa Privada	Empleado Empresa Privada	Oficinas
	7438550	Tunja	nov-22	Empleado Empresa Privada	Empleado Empresa Privada	Oficinas

#### 4.4 Código SQL - lenguaje de definición de datos (DDL) (Primera entrega)



```

CREATE TABLE Retiros (
    ID_Cliente NUMBER(10) PRIMARY KEY,
    Of_Cliente VARCHAR2(50),
    Mes_Retiro VARCHAR2(10),
    Entidad_Nicho VARCHAR2(50),
    Canal VARCHAR2(50)
);

alter table Retiros
add Entidad VARCHAR2(50);

SELECT * FROM Retiros

CREATE TABLE Demograficos (
    ID_Cliente NUMBER(10) PRIMARY KEY,
    Celular VARCHAR2(50),
    Ocupacion VARCHAR2(50),
    Pagaduria VARCHAR2(70),
    Empresa VARCHAR2(50),
    Estado_civil varchar2(50),
    Edad number(10),
    Genero varchar2(50),
    Salario number(20)
);

ALTER TABLE Demograficos MODIFY Pagaduria varchar2(100);
ALTER TABLE Demograficos MODIFY Empresa varchar2(100);
ALTER TABLE Demograficos MODIFY Ocupacion varchar2(50);

```



```

CREATE TABLE Clientes_ENE (
    ID_Cliente NUMBER(10) PRIMARY KEY,
    Of_producto VARCHAR2(50),
    Producto VARCHAR2(50),
    COD VARCHAR2(70),
    Saldo number(20),
    Num_producto number(20),
    Estado varchar2(10)
);

CREATE TABLE Clientes_FEB (
    ID_Cliente NUMBER(10) PRIMARY KEY,
    Of_producto VARCHAR2(50),
    Producto VARCHAR2(50),
    COD VARCHAR2(70),
    Saldo number(20),
    Num_producto number(20),
    Estado varchar2(20)
);

```

#### 4.5 Código SQL - Manipulación de datos (DML) *(Primera entrega)*

```

INSERT INTO CLIENTES_ENE (OF_PRODUCTO, ID_CLIENTE, PRODUCTO, COD, SALDO, NUM_PRODUCTO, ESTADO) VALUES ('30053',123334,'A Término','ED',20.000.000 ,9813,'Activo');
--Fila 2
INSERT INTO CLIENTES_ENE (OF_PRODUCTO, ID_CLIENTE, PRODUCTO, COD, SALDO, NUM_PRODUCTO, ESTADO) VALUES ('30053',4456454,'A Término','ED',44.388.502 ,9813,'Activo');
--Fila 3
INSERT INTO CLIENTES_ENE (OF_PRODUCTO, ID_CLIENTE, PRODUCTO, COD, SALDO, NUM_PRODUCTO, ESTADO) VALUES ('30053',77879899,'A Término','ED',18.800.000 ,10243,'Activo');
--Fila 4
INSERT INTO CLIENTES_ENE (OF_PRODUCTO, ID_CLIENTE, PRODUCTO, COD, SALDO, NUM_PRODUCTO, ESTADO) VALUES ('30053',1909090768541,'A Término','ED',5.000.000 ,10243,'Activo');
--Fila 5
INSERT INTO CLIENTES_ENE (OF_PRODUCTO, ID_CLIENTE, PRODUCTO, COD, SALDO, NUM_PRODUCTO, ESTADO) VALUES ('30053',52343450301,'A Término','ED',6.000.000 ,10243,'Activo');
--Fila 6
INSERT INTO CLIENTES_ENE (OF_PRODUCTO, ID_CLIENTE, PRODUCTO, COD, SALDO, NUM_PRODUCTO, ESTADO) VALUES ('30053',9045354317045,'A Término','ED',5.000.000 ,10243,'Activo');

```

Tabla Normalizada:

```

CREATE TABLE Norma (
    ID_Cliente NUMBER(10) PRIMARY KEY,
    Of_producto VARCHAR2(50),
    Producto VARCHAR2(50),
    COD VARCHAR2(70),
    Saldo number(20),
    Num_producto number(20),
    Estado varchar2(20),
    Mes varchar2(10)
);

select * from Norma;

INSERT INTO Norma (ID_Cliente, Of_producto, Producto, COD, Saldo, Num_producto, Estado, Mes)
SELECT ID_Cliente, Of_producto, Producto, COD, Saldo, Num_producto, Estado, 'Enero'
FROM Clientes_ENE;

INSERT INTO Norma (ID_Cliente, Of_producto, Producto, COD, Saldo, Num_producto, Estado, Mes)
SELECT ID_Cliente, Of_producto, Producto, COD, Saldo, Num_producto, Estado, 'Febrero'
FROM Clientes_FEB;

DROP TABLE Clientes_ENE;
DROP TABLE Clientes_FEB;

select * from Norma;

```

#### 4.6 Código SQL + Resultados: Vistas (*Primera entrega*)

OF_PRODUTO	PRODUCTO	COD	SALDO	NUM_PRODUCTO	ESTADO	MES
30030	A Término	CD	242	206158	Activo	Enero
30005	Programado	Programado	35	58500054652	Activo	Enero
30005	Programado	Programado	68	58500054653	Activo	Enero
30005	Programado	Programado	109	58500054654	Activo	Enero
30005	Programado	Programado	162	58500054655	Activo	Enero
30005	Programado	Programado	510	58500054656	Activo	Enero
30005	Programado	Programado	53	58500054657	Activo	Enero
30005	Programado	Programado	393	58500054658	Activo	Enero
30006	Permanente	Permanente	968	58500055676	Activo	Enero
30017	Permanente	Permanente	604	58500060019	Activo	Enero
30017	Permanente	Permanente	524	58500060020	Activo	Enero
30017	Permanente	Permanente	950	58500060021	Activo	Enero
30019	Permanente	Permanente	803	58500061100	Activo	Enero
30028	Permanente	Permanente	23	58500064362	Activo	Enero
30028	Permanente	Permanente	751	58500064363	Activo	Enero
30028	Permanente	Permanente	49	58500064364	Activo	Enero
30028	Permanente	Permanente	712	58500064365	Activo	Enero
30028	Permanente	Permanente	58	58500064366	Activo	Enero
30028	Permanente	Permanente	780	58500064367	Activo	Enero
30032	Permanente	Permanente	729	58500066575	Activo	Enero
30032	Permanente	Permanente	953	58500066576	Activo	Enero
30032	Permanente	Permanente	791	58500066577	Activo	Enero
30032	Permanente	Permanente	216	58500066578	Activo	Enero
30032	Permanente	Permanente	164	58500066579	Activo	Enero
30034	Permanente	Permanente	111	58500067655	Activo	Enero
30034	Permanente	Permanente	319	58500067656	Activo	Enero
.....	-	-	---	-----	-	-

CELULAR	OCCUPACION	PAGADURIA	EMPRESA	ESTADO_CIVIL	EDAD	GENERO	SALARIO
6910880	(null)	SEC DE EDUCACION SANTAND	(null)	(null)	(null)	(null)	(null)
3163202474	(null)	(null)	(null)	(null)	(null)	(null)	(null)
6578988	(null)	(null)	(null)	(null)	(null)	(null)	(null)
5,73144E+11	Independiente	(null)	(null)	Casado/a	26	Femenino	0
(null)	(null)	(null)	(null)	(null)	(null)	(null)	0
5,73204E+11	Empleado Público	INSTITUTO DISTRITAL PARA LA PROTECCION DE LA NIÑEZ Y LA JUVENTUD	IDIFRON	Soltero/a	27	Femenino	1700000
(null)	Empleado Privado	(null)	(null)	(null)	22	Femenino	0
5,73016E+11	Independiente	DATO DEFURABLE	(null)	Sin información	32	Femenino	0
5,73135E+11	Empleado Privado	(null)	ROYAL SEGURIDAD	Soltero/a	30	Masculino	1100000
5,73014E+11	(null)	(null)	(null)	Soltero/a	(null)	Femenino	846000
5,73215E+11	(null)	(null)	(null)	Unión libre	28	Femenino	0
5,73183E+11	Estudiante	(null)	SIN INFORMACION	Soltero/a	20	Femenino	900000
3138341716	Independiente	(null)	INDEPENDIENTE	Soltero/a	22	Masculino	1000000
(null)	(null)	(null)	(null)	(null)	37	(null)	0
5,73142E+11	(null)	(null)	(null)	(null)	(null)	(null)	(null)
5,73126E+11	(null)	(null)	(null)	(null)	(null)	(null)	(null)
5,73123E+11	Empleado Público	(null)	FISCALIA GENERAL	Soltero/a	47	Masculino	2363065
5,73174E+11	Independiente	(null)	DEPARTAMENTO ADM NACION	Soltero/a	34	Femenino	0
5,73144E+11	Empleado Privado	(null)	STAFF	Soltero/a	23	Masculino	908526
5,73133E+11	Empleado Privado	(null)	Indra Colombia	Soltero/a	22	Masculino	877000
5,73505E+11	Estudiante	(null)	NO APLICA	Soltero/a	21	Masculino	877503
5,73173E+11	Estudiante	(null)	ESTUDIANTE	Soltero/a	22	Masculino	3900000
3015447087	Empleado Privado	(null)	SERVICES Y CONSULTING	Soltero/a	19	Masculino	1100000
5,73164E+11	Estudiante	SIN PAGADURIA	NO APLICA	Soltero/a	21	Femenino	100000
5,73137E+11	Empleado Público	(null)	Sin Información	Soltero/a	47	Masculino	6810316
(null)	Independiente	(null)	SIN INFORMACIÓN	(null)	47	Masculino	0
5,73165E+11	Estudiante	(null)	NO APLICA	Soltero/a	23	Femenino	1000000
5,73144E+11	Estudiante	DATO DEFURABLE	(null)	Soltero/a	22	Masculino	670000
5,73116E+11	Estudiante	(null)	NO APLICA	Soltero/a	21	Masculino	500000
5,73156E+11	Empleado Privado	DATO DEFURABLE	(null)	Sin información	22	Masculino	828116

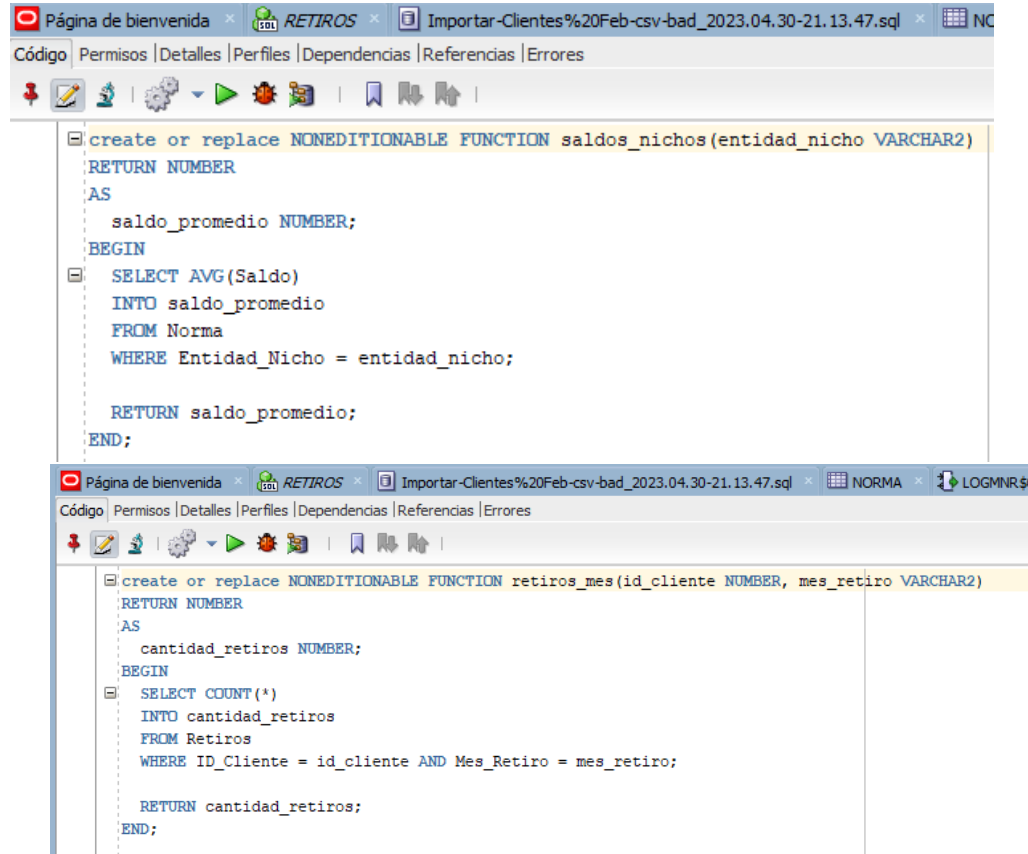
#### 4.7 Código SQL + Resultados: Triggers (*Primera entrega*)

```

## tigger
CREATE OR REPLACE TRIGGER actualiza_entidad
AFTER INSERT OR UPDATE ON Demograficos
FOR EACH ROW
BEGIN
UPDATE Retiros
SET Entidad = :new.Pagaduria
WHERE ID_Cliente = :new.ID_Cliente;
END;

```

#### 4.8 Código SQL + Resultados: Funciones (*Primera entrega*)



The image displays two screenshots of an SQL IDE interface. The top screenshot shows the creation of a function named `saldos_nichos`. The bottom screenshot shows the creation of a function named `retiros_mes`.

```
create or replace NONEDITIONABLE FUNCTION saldos_nichos(entidad_nicho VARCHAR2)
RETURN NUMBER
AS
    saldo_promedio NUMBER;
BEGIN
    SELECT AVG(Saldo)
    INTO saldo_promedio
    FROM Norma
    WHERE Entidad_Nicho = entidad_nicho;

    RETURN saldo_promedio;
END;
```

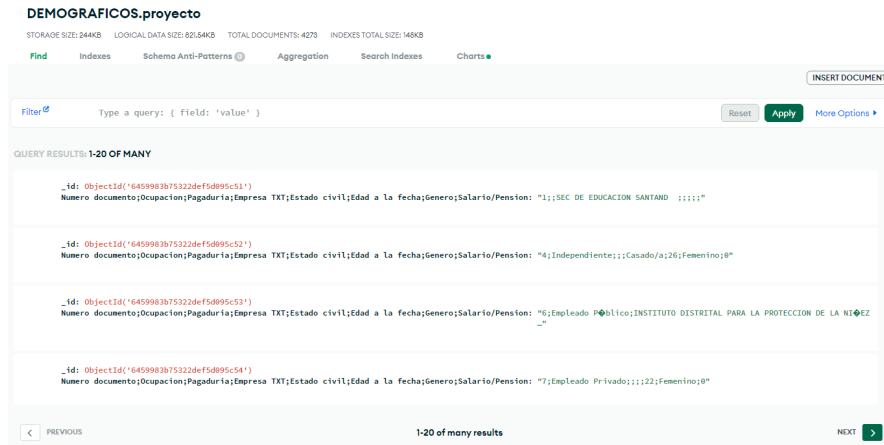
```
create or replace NONEDITIONABLE FUNCTION retiros_mes(id_cliente NUMBER, mes_retiro VARCHAR2)
RETURN NUMBER
AS
    cantidad_retiros NUMBER;
BEGIN
    SELECT COUNT(*)
    INTO cantidad_retiros
    FROM Retiros
    WHERE ID_Cliente = id_cliente AND Mes_Retiro = mes_retiro;

    RETURN cantidad_retiros;
END;
```

#### 4.9 Código SQL + Resultados: procedimientos almacenados (*Primera entrega*)

## 5 Bases de Datos No-SQL (*Segunda entrega*)

### 5.1 Diagrama Bases de Datos No-SQL (*Segunda entrega*)



```
BD proyecto NoSQL.ipynb
Archivo Editor Ver Insertar Entorno de ejecución Herramientas Ayuda Se han guardado todos los cambios

+ Código + Texto

[1] 1 !pip install pymongo
    2 !pip install dnspython

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting pymongo
  Downloading pymongo-4.3.3-cp310-cp310-manylinux_2_17_x86_64_manylinux2014_x86_64.whl (492 kB)
    492.0 kB/s | 5.4 MB/s eta 0:00:00
Collecting dnspython<3.0.0,=>1.16.0 (from pymongo)
  Downloading dnspython-2.3.0-py3-none-any.whl (283 kB)
    283.7 kB/s | 24.3 MB/s eta 0:00:00
Installing collected packages: dnspython, pymongo
Successfully installed dnspython-2.3.0 pymongo-4.3.3
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: dnspython in /usr/local/lib/python3.10/dist-packages (2.3.0)

1 import pymongo
2 from pymongo import MongoClient
3 from pymongo.server_api import ServerApi
4
5 # Conectarse a la base de datos de MongoDB Atlas
6 client = MongoClient("mongodb+srv://nfernandez@cluster0.arzkuw.mongodb.net/?retryWrites=true&majority", server_api=ServerApi('1'))

[ ] 1 !mongoimport --host <ac-m8sgqci-shard-00-01.arzkuw.mongodb.net:27017> --port <27017> --username <nfernandez> --password <Nikis1984> --db <DEMOGRAFICOS> --collection <proyecto>

[3] 1 db = client.DEMOGRAFICOS
    2 collection = db.proyecto

[4] 1 # Send a ping to confirm a successful connection
    2 try:
    3     client.admin.command('ping')
    4     print("Pinged your deployment. You successfully connected to MongoDB!")
    5 except Exception as e:
    6     print(e)

Pinged your deployment. You successfully connected to MongoDB!
```

### 5.2 SMDB utilizado para la Base de Datos No-SQL (*Segunda entrega*)

Se usa MongoDB Atlas como SMDB para la Base de Datos No-SQL al ser una solución completa para la gestión de bases de datos NoSQL, al adaptarse a las necesidades de una amplia variedad de aplicaciones y permitir escalar, asegurar y administrar los datos de manera efectiva y eficiente.

La tabla que se usa en MongoDB es la de demográficos cargando 4273 registros por importación con las herramientas de MongoDB, se conecta a python y se

trabaja con códigos desde esta aplicación.

## 6 Bibliografía

- [1] H. Arellano Díaz, La calidad en el servicio como ventaja competitiva, Dominio de las Ciencias, vol. III, pp. 72-83, 2017.
- [2] D. P. Puerto Becerra, La globalización y el crecimiento empresarial a través de estrategias de internacionalización, Pensamiento Gestión, nº 28, pp. 171-195, 2010.
- [3] Guangli, N., Lingling, Z., Xingsen, L., Yong, S. (2011). The Analysis on the Customers Churn of Charge Email Based on Data Mining Take One Internet Company for Example. Institute of Electrical and Electronics Engineers, 843-847.
- [4] J. Rozum, Defining and Understanding Software Measurement Data, Software Engineering Institute Carnegie Mellon University, 2002.
- [5] R. Feinberg y M. Trotter, Immaculate deception: the unintended negative effects of the CRM revolution: maybe we would be better off without customer relations management., Defying the Limits, pp. 26-31, 2001.
- [6] Chitra, K., Subashini, B. (2011). Customer Retention in Banking Sector using Predictive Data Mining Technique. The 5th International Conference on Information Technology.
- [7] Polo Redondo, Y., Sesé Olivan, F. J. (2009). La retención de los clientes un estudio empírico de sus determinantes. Revista Española de Investigación de Marketing, 117-137
- [8] Sarkar, D., Bali, R., Sharma, T. (2018). Practical machine learning with Python. A Problem-Solvers Guide To Building Real-World Intelligent Systems. Berkely: Apress.
- [9] Rubin, D. B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." Journal of Educational Psychology 66 (5): 688–701.
- [10] Little, R. J. A., D. B. Rubin, and S. Z. Zangeneh. 2017. "Conditions for Ignoring the Missing-Data Mechanism in Likelihood Inferences for Parameter Subsets." Journal of the American Statistical Association 112 (517): 314–20.
- [11] Breiman, L., J. Friedman, R. Olshen, and C. Stone. 1984. Classification and Regression Trees. New York: Wadsworth Publishing.
- [12] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition, O'Reilly Media, Inc., 2019.
- [13] Bruno. M., (2022). Implementación de un modelo de minería de datos para predecir la deserción de los clientes en una empresa de telecomunicaciones. Universidad católica santo toribio de Mogrovejo. 20-28.
- [14] Bohorquez. M., Torys. J., Paredes. M., (2020). MODELOS DE PREDICCIÓN DE DESERCIÓN DE CLIENTES PARA. Revista Compendium: Cuadernos de Economía y Administración., Vol 7, No 1, 1-11.
- [15] R. A. Barrueta Meza and E. J. P. Castillo Villarreal, "Modelo de análisis predictivo para determinar clientes con tendencia a la deserción en bancos peruanos," Universidad Peruana de Ciencias Aplicadas(UPC)., Lima, Perú, 2018. Doi: <http://doi.org/10.19083/tesis/626023>.
- [16] Torrado, M. y Berlanga, V. (2013). Análisis Discriminante mediante SPSS. [En línea] REIRE, Revista d'Innovació i Recerca en Educació, 6 (2), 150-166.



- [17] Hastie, T., Friedman, J., y Tibshirani, R. (2001). The Elements of Statistical Learning. Nueva York, Estados Unidos: Springer New York. DOI: 10.1007/978-0-387-21606-5.
- [18] Rossiter, D. 1994. Basic Concepts and Procedures on Land Evaluation. Cornell University course Soil, Crop Atmospheric Sciences. 'Special Topics in Soil, Crop Atmospheric Sciences: Land evaluation, with emphasis on computer applications', Spring Semester 1994.
- [19] Ley 1584 de 2012. Por la cual se dictan disposiciones generales para la protección de datos personales. Publicada en el Diario Oficial 48587 de octubre 18 de 2012.