Proyecto de Bases de datos para el análisis y clustering de tasas de interes de crédito en el mercado Colombiano

Lizeth Johana Guependo - Oscar Mauricio Carvajal

Facultad de ingeniería y ciencias básicas,
Universidad Central
Maestría en Analítica de Datos
Curso de Bases de Datos
Bogotá, Colombia
lguependos@ucentral.edu.co, ocarvajals@ucentral.edu.co

October 17, 2022

Contents

	Introducción		
2 Antecedentes del proyecto			
3 C	Características del proyecto de investigación		
3.	1 Título del proyecto de investigación }	5	
3.	2 Objetivo general	5	
	3.2.1 Objetivos especificos (Max 100 Palabras) - (<i>Primera entrega</i>)	5	
3.	3 Alcance	5	
3.	4 Pregunta de investigación	5	
3.	5 Hipotesis	6	
4 R	teflexiones sobre el origen de datos e información	7	
4.	1 ¿Cual es el origen de los datos e información ?	7	
4.	2 ¿Cuales son las consideraciones legales o eticas del uso de la		
	información?	7	
4.	3 ¿Cuales son los retos de la información y los datos que utilizara		
	en la base de datos en terminos de la calidad y la consolidación?	7	
4.	4 ¿Que espera de la utilización de un sistema de Bases de Datos		
	para su proyecto?	8	

5	Diseño del Modelo de Datos del SMBD (Sistema Manejador de			
	Bas	es de Datos)	9	
	5.1	Características del SMBD (Sistema Manejador de Bases de Datos)		
		para el proyecto }	9	
	5.2	Diagrama modelo de datos	10	
	5.3	Imágenes de la Base de Datos	11	
	5.4	Código SQL - lenguaje de definición de datos (DDL)	12	
	5.5	Código SQL - Manipulación de datos (DML) (Primera entrega)	14	
	5.6	Código SQL + Resultados: Vistas (<i>Primera entrega</i>)	16	
	5.7	Código SQL + Resultados: Triggers (<i>Primera entrega</i>)	18	
	5.8	Código SQL + Resultados: Funciones (<i>Primera entrega</i>)	19	
	5.9	Código $SQL + Resultados$: procedimientos almacenados (<i>Primera</i>		
		entrega)	20	
6	Bas	es de Datos No-SQL (Segunda entrega)	22	
	6.1	Diagrama Bases de Datos No-SQL (Segunda entrega)	22	
	6.2	SMBD utilizado para la Base de Datos No-SQL (Segunda entrega)	22	
7	Apl	icación de ETL (Extract, Transform, Load) y Bodega de		
	Dat	os (Tercera entrega)	23	
	7.1	Ejemplo de aplicación de ETL y Bodega de Datos (Tercera entrega)	23	
8	Lec	ciones aprendidas (Tercera entrega)	24	
•	Du.		۵-	
9	Bib.	liografía	25	

1 Introducción

La Superintendencia Financiera de Colombia tiene por objetivo supervisar el sistema financiero colombiano con el fin de preservar su estabilidad, seguridad y confianza, así como, promover, organizar y desarrollar el mercado de valores colombiano y la protección de los inversionistas, ahorradores y asegurados. (Superfinanciera, s.f.), por esta razón la Superintendencia Financiera de Colombia expide normas de carácter general, contenidas en resoluciones y circulares, con el objeto de instruir a las entidades sobre cómo deben ejercer su actividad. (Asobancaria, s.f.) Mediante la Circular Externa 34 de 2021 la Superintendencia Financiera de Colombia creo la proforma F.1000-144 (formato 414) "Tasas de interés activas por tipo de crédito" en el anexo 1 de la Circular Básica Contable y Financiera, esto debido a la necesidad de conocer el comportamiento de las tasas de interés activas que ofrecen los Bancos, corporaciones financieras y compañías de financiamiento; clasificadas por tipo de persona, sexo, tamaño de empresa, modalidad, tipo de crédito, garantía, producto y plazo, así como el número y monto de créditos desembolsados. (Superintendencia Financiera de Colombia, Circular Externa 34 de 2021).

2 Antecedentes del proyecto

El estudio del mercado ha generado grandes oportunidades de negocio que han permito adaptar tecnologías, estrategías y recursos que operen en función de las oportunidades que se permiten al abrir y entender a la competencia. Desde sus inicios en los años sesenta, se ha dado una gran importancia al benchmarking en muchos sectores, para el sector bancario y demás gremios ha sido importante ser competitivo frente a un mercado monetario que cambia constantemente para el caso de los bancos en función de una política monetaria saludable, lo anterior se ha visto reflejado en una gran cantidad de iniciativas y desarrollos para el entendimineto y la adaptación a los cambios globales y la competencia directa. Actualmente existen diferentes mecanismos que permiten comparar indicadores de las empresas y sus directos contrincantes, en el caso de los bancos las regulaciones, tasas de usura y campañas de colocación permiten tener un mercado dinámico en la búsqueda de una mayor participación en el mercado.

3 Características del proyecto de investigación

3.1 Título del proyecto de investigación}

PROYECTO DE BASE DE DATOS PARA EL ANÁLISIS Y CLUSTERING DE TASAS DE INTERES DE CRÉDITO EN EL MERCADO COLOMBIANO

3.2 Objetivo general

Aplicar a través de herramientas tecnológicas la clasificación por clusters, para agrupar los datos reportados por las entidades financieras en relación con tasas de interés por tipo de crédito, que permitan identificar patrones y grupos de mercado.

3.2.1 Objetivos especificos (Max 100 Palabras) - (Primera entrega)

- Identificar la participación de los desembolsos y sus tasas de interés a nivel nacional en las principales líneas de crédito.
- Segmentar las diferentes carteras con sus tipologías para tener un análisis más acertado.
- Caracterizar a las entidades bancarias y su competitividad en el mercado.
- Generar estrategías de acuerdo a los resultados obtenidos en los análisis planteados.

3.3 Alcance

A través de la estructuración de la base de datos consolidada de forma ordenada, se pretende tener un histórico de cuatro meses con información relevante de tasas de interés activas de créditos hipotecarios, comercial y de consumo, para llegar a la segmentación deseada es necesario contar no únicamente con las tasas y el banco; sino discriminar por tipo de entidad, plazo, tipo de crédito, producto de crédito, genero y cantidad con el fin de lograr una mayor precisión en la construcción de modelos de clasificación. Es necesario tener una relación de la información para categorizar el segmento, preferencias y competitividad de tasas para adoptar estrategías de análisis que permitan hallar características semejantes para la toma de decisiones.

3.4 Pregunta de investigación

¿Cómo segmentar los productos y las tasas de interés aplicadas por las entidades financieras colombianas de acuerdo a sus características y variables que inciden en su colocación?

3.5 Hipotesis

De acuerdo a los lineamientos de los entes reguladores es posible que las tasas de interés activas de crédito sean homogéneas para las entidades por tipo de negocio (bancos, cooperativas, fiduciarias). Esto permite mantener un cerco reglamentado limitado por las tasas de usura y la competitividad del mercado. Es posible que el resultado sea semejante por variables como la cantidad de colocaciones, el tipo de entidad, el monto desembolsado y el producto de crédito.

4 Reflexiones sobre el origen de datos e información

Es necesario consolidar la información general de los diferentes bancos (Top 10) en una base de datos relacional que permita agrupar de forma óptima los desembolsos, con su tipología, características y cifras necesarias para el desarrollo del proyecto. La información debe ser veraz, consistente y con un rango de tiempo adecuado para poder generar las consultas necesarias para el análisis de la misma. Por otra parte, se requiere poder manipular dicha información de tal forma que sea aterrizada a los reportes requeridos para la tesis.

4.1 ¿Cual es el origen de los datos e información?

Al ser información sensible de las entidades bancarias, es necesario remitirse a los entes de control que mantienen de forma periódica al alcance esta información relevante para el proyecto. Es por esto que se decide tomar como fuente la información consolidada por la Superintendencia Financiera de Colombia (SFC) a través de su página https://www.superfinanciera.gov.co/jsp/index.jsf.

4.2 ¿Cuales son las consideraciones legales o eticas del uso de la información?

La información actualmente se encuentra en los repositorios de la Superintendencia financiera de Colombia, sin embargo, para acceder a ella las tablas son públicas a través del anexo 1 de la circular básica contable dispuesta en la página de la SFC. Para identificar las tasas de interés por modalidad de crédito se puede ingresar a los reportes semanales los cuales son de uso abierto como parte de la normativa general de la SFC y las políticas de seguridad en las cuales no se incluye información facial de los créditos desembolsados. https://www.superfinanciera.gov.co/inicio/informes-y-cifras/cifras/establecimientos-de-credito/informacion-periodica/semanal-60825.

4.3 ¿Cuales son los retos de la información y los datos que utilizara en la base de datos en terminos de la calidad y la consolidación?

Inicialmente dado que la información esta en diferentes tablas que hacen referencia a homologaciones otorgadas por la Superintendencia financiera de Colombia, el principal desafío es consolidar y clasificar de forma idónea de los datos. Por otra parte, se debe entender las combinaciones correctas que adquiere esta información, es decir, si un registro está clasificado con genero femenino, esto quiere decir que el tipo de persona es natural por lo que tiene unas particularidades de entendimiento para los diferentes tipos de clasificaciones que existen en las 7 tablas satélite.

4.4 ¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto?

Con la implementación de una base de datos relacional se pretende principalmente centralizar la información en un sistema que sea de acceso fácil con bondades pertinentes de las bases de datos en SQL, además permitir generar las vistas procedimientos disparadores funciones y diferentes consultas en un ambiente controlado para el estudio y manipulación de datos requerida por el proyecto. Para finalizar al tener diversas tablas es posible relacionarlas de tal forma que se pueda exprimir la mayor cantidad de datos y ventajas que la misma información proporciona en sí.

5 Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)

5.1 Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto }

Como sistema manejador de base de datos se trabajará con workbench la cual es una herramienta visual de diseño de bases de datos que integra desarrollo de software, administración de bases de datos, diseño de bases de datos, gestión y mantenimiento para el sistema de base de datos MySQL. Es el sucesor de DBDesigner 4 de fabFORCE.net, y reemplaza el anterior conjunto de software, MySQL GUI Tools Bundle.

Las características destacadas de MySQL Workbench 5.2 son: General:

- Conexión y gestión de instancia de base de datos.
- Asistente de elementos de acción.
- Completamente compatible con guiones escritos en Python y Lua.
- Soporte para complementos personalizados

Editor de SQL:

- Exploración de esquema de objetos.
- Resaltado de sintaxis en SQL y analizador de declaraciones.
- Conjuntos de resultados múltiples, editables.
- Colecciones de fragmentos SQL.
- Tunelización de conexión por SSH.
- Soporte Unicode.

Modelado de datos:

- Diagrama entidad relación.
- Modelado visual con arrastrar y soltar.
- Ingeniería inversa a guiones SQL y base de datos en vivo.
- Ingeniería directa a guiones SQL y base de datos en vivo.
- Sincronización de esquema.
- Impresión de modelos.
- Importación desde fabFORCE.net DBDesigner4.

Administración de base de datos:

- Iniciar y detener instancias de base de datos.
- Configuración de instancias.
- Administración de cuentas en base de datos.
- Exploración de instancias variables.
- Exploración de ficheros de registros.
- Exportación e importación masiva de datos.
- Información tomada de https://es.wikipedia.org/wiki/MySQL_Workbench.

5.2 Diagrama modelo de datos

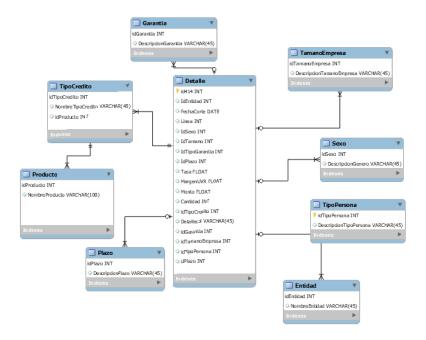


Figure 1: Modelo entidad relación de la base de datos

5.3 Imágenes de la Base de Datos

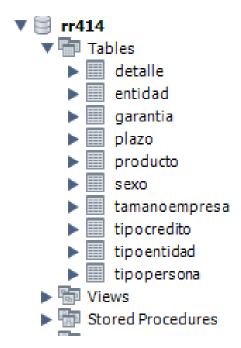


Figure 2: Base de Datos

5.4 Código SQL - lenguaje de definición de datos (DDL)

```
1 • \ominus CREATE TABLE `detalle` (
        `id414` int NOT NULL AUTO_INCREMENT,
         `idtipoentidad` int DEFAULT NULL,
         `idcodigoEntidad` int DEFAULT NULL,
 4
        `FechaCorte` date DEFAULT NULL,
 5
        `idTipoPersona` int DEFAULT NULL,
 6
 7
        `idSexo` int DEFAULT NULL,
        `idTamanoEmpresa` int DEFAULT NULL,
        `idTipoCredito` int DEFAULT NULL,
        `idGarantia` int DEFAULT NULL,
10
        `idProducto` int DEFAULT NULL,
11
        `idPlazo` int DEFAULT NULL,
12
        `Tasa` float DEFAULT NULL,
         `MargenUVR` float DEFAULT NULL,
14
         `Monto` float DEFAULT NULL,
15
         `Cantidad` int DEFAULT NULL,
16
17
        PRIMARY KEY ('id414')
     ) ENGINE=InnoDB AUTO_INCREMENT=2711 DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
   1 • ⊖ CREATE TABLE `entidad` (
           `idEntidad` int NOT NULL AUTO_INCREMENT,
    3
           `codigoEntidad` int DEFAULT NULL,
           `NombreEntidad` varchar(45) DEFAULT NULL,
           `idtipoentidad` int DEFAULT NULL,
    5
           PRIMARY KEY (`idEntidad`)
        ) ENGINE=InnoDB AUTO_INCREMENT=50 DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
   1 • ⊖ CREATE TABLE `garantia` (
           `idGarantia` int NOT NULL,
            `DescripcionGarantia` varchar(90) DEFAULT NULL
   3
        ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
   4
   1 • ⊖ CREATE TABLE `sexo` (
           `idSexo` int NOT NULL,
           `DescripcionGenero` varchar(45) DEFAULT NULL,
           PRIMARY KEY ('idSexo')
         ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
```

```
1 • \ominus CREATE TABLE `plazo` (
         `idPlazo` int NOT NULL,
 2
         `DescripcionPlazo` varchar(45) DEFAULT NULL,
 3
 4
         PRIMARY KEY (`idPlazo`)
       ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
 1 • 

CREATE TABLE `producto` (
        `idProducto` int NOT NULL,
 3
        `NombreProducto` varchar(100) DEFAULT NULL,
 4
        PRIMARY KEY ('idProducto')
    ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
1 ● ○ CREATE TABLE `tamanoempresa` (
2
        `idTamanoEmpresa` int NOT NULL,
3
        `DescripcionTamanoEmpresa` varchar(45) DEFAULT NULL,
4
        PRIMARY KEY (`idTamanoEmpresa`)
     ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
1 • ⊖ CREATE TABLE `tipocredito` (
2
        `idTipoCredito` int NOT NULL,
        `NombreTipoCredito` varchar(45) DEFAULT NULL,
        PRIMARY KEY (`idTipoCredito`)
      ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
 1 • 

CREATE TABLE `tipoentidad` (
         `idtipoentidad` int NOT NULL,
         `Nombretipoentidad` varchar(100) DEFAULT NULL,
 3
 4
        PRIMARY KEY (`idtipoentidad`)
       ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
1 • ⊖ CREATE TABLE `tipopersona` (
       `idTipoPersona` int NOT NULL,
        `DescripcionTipoPersona` varchar(45) DEFAULT NULL,
       PRIMARY KEY (`idTipoPersona`)
4
    ) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
```

5.5 Código SQL - Manipulación de datos (DML) (Primera entrega)

```
1 • INSERT INTO 'detalle' VALUES (1,1,7,'2022-08-26',1,1,1,3,1,309,100,10.17,0,166600,1),
      (2,1,7,'2022-08-12',1,1,2,3,2,302,502,16.59,0,90000000,2),
      (3,1,42,'2022-09-09',1,1,10,1,1,102,206,35.22,0,298735000,1416),
      (4,1,7,'2022-08-12',2,10,2,3,2,301,207,33.22,0,4124290000,7266);
    1 • INSERT INTO `entidad` VALUES (1,1,'Banco de Bogota',1),
          (2,1,'Bancoldex',22),
          (3,1, 'Cooperativa Financiera de Antioquia',32),
          (4,2,'Banco Popular',1),(5,2,'Findeter',22),
          (6,2,'JFK Cooperativa Financiera',32);
    1 • INSERT INTO `garantia` VALUES (1,'Sin Garantia'),
    2
          (2, 'Garantia idonea o no idonea'),
          (3, 'Garantia fondo nacional de garantias FNG o Fondo de Garantias de Antioquia FGA'),
          (4, 'Garantia del fondo agropecuario de garantias FAG');
   1 • INSERT INTO `sexo` VALUES (1, 'Femenino'),
   2
         (2, 'Masculino'),
        (3,'Trans'),
   3
   4
       (4,'No binario'),
         (10,'No aplica');
   1 • INSERT INTO `plazo` VALUES (100, 'Hasta 30 dias'),
         (101, 'Entre 31 dias y 1 anio'),
         (102, 'Mas de 1 anio y hasta 3 anios'),
       (103, 'Mas de 3 anios y hasta 5 anios'),
         (104, 'Mas de 5 anios y hasta 7 anios'),
         (105, 'A mas de 7 anios'),
         (201, 'Avances en efectivo'),
         (202, 'Consumos a un mes'),
         (203, 'Consumos entre 2 y 6 meses'),
         (204, 'Consumos entre 7 y 12 meses'),
   11
         (205, 'Consumos entre 13 y 18 meses'),
         (206, 'Consumos entre 19 y 24 meses'),
   12
         (207, 'Consumos entre 25 y 36 meses'),
  13
       (208, 'Consumos a mas de 36 meses');
```

```
1 • INSERT INTO `producto` VALUES (101,'Credito rotativo'),
       (102, 'Tarjeta de credito para ingresos hasta 2 SMMLV'),
       (103, 'Tarjeta de credito para ingresos superiores a 2 SMMLV'),
     (104, 'Libre inversion'),(105, 'Libranza salud'),(106, 'Libranza educativo'),
5
    (107,'Libranza otros'),
6
    (108, 'Vehiculo'),
    (109, 'Creditos de consumo de bajo monto');
1 • INSERT INTO `tamanoempresa` VALUES (1, 'Microempresa'),
2
       (2, 'Pequena empresa'),
3
      (3,'Mediana empresa'),
4
      (4, 'Gran empresa'),
       (10, 'No aplica');
1 • INSERT INTO `tipocredito` VALUES (1,'Consumo'),
     (2,'Vivienda'),
    (3,'Comercial ordinario'),
     (4, 'Comercial preferencial o corporativo'),
     (5,'Comercial tesoreria'),
      (6,'Comercial especial'),
      (7, 'Microcredito');
1 • INSERT INTO `tipoentidad` VALUES (1, 'BC ESTABLECIMIENTO BANCARIO'),
      (4, 'CF COMPANIA DE FINANCIAMIENTO'),
      (22, 'INSTITUCION OFICIAL ESPECIAL'),
3
      (32, 'COOPERATIVAS DE CARACTER FINANCIERO');
4
1 • INSERT INTO `tipopersona` VALUES (1,'Natural'),
       (2,'Juridica');
```

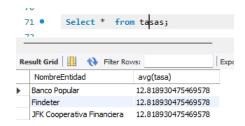
5.6 Código SQL + Resultados: Vistas (Primera entrega)

select * from Scotiabank;

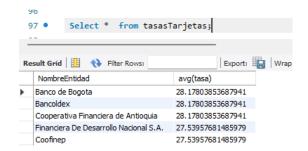
Export: Wrap Cell Content: IA id414 idtipoentidad idcodigoEntidad FechaCorte idTipoPersona idSexo idTamanoEmpresa 2022-09-09 1 2022-08-05 2 2022-09-23 2022-07-01 1 2022-07-22 2022-07-01 2 2022-08-26 2022-09-09 2 2022-07-15 2022-07-22 1

2022-09-02 1

```
#VISTA TASAS
create view tasas as
SELECT C.NombreEntidad, avg(tasa)
from rr414.detalle a join rr414.entidad b
on a.idcodigoEntidad = b.idEntidad
join rr414.entidad C
on a.idTipoCredito = c.codigoEntidad
where a.idTipoCredito = 2
group by C.NombreEntidad;
```



```
#VISTA TASAS TARJETAS
87
88 •
       create view tasasTarjetas as
89
       SELECT C.NombreEntidad, avg(tasa)
       from rr414.detalle a join rr414.entidad b
90
       on a.idcodigoEntidad = b.idEntidad
91
92
       join rr414.entidad C
93
       on a.idTipoCredito = c.codigoEntidad
94
       where a.idProducto IN (102, 103, 301)
       group by C.NombreEntidad;
95
```



5.7 Código SQL + Resultados: Triggers (Primera entrega)

```
'Accion' VARCHAR(200) NULL,
      `Fecha` DATETIME NULL DEFAULT CURRENT_TIMESTAMP,
4
      PRIMARY KEY ('idAcciones'));
5
6
         DELIMITER //
        CREATE TRIGGER logdetalle AFTER INSERT ON detalle
       3
   4
             INSERT INTO acciones(accion) VALUE ('Se creo un registro en detalle');
   5
         END //
         DELIMITER ;
           INSERT INTO rr414.detalle VALUES (2711, 1, 7, '2022/08/12', 1, 1, 2, 3, 2, 303, 503, 16.59,
           SELECT * FROM rr414.acciones
  | Edit: 🝊 🔜 🖶 | Export/Import: 🏢 🐻 | Wrap Cell Content: 🏗
      idAcciones Accion
                                   Fecha
              Se creo un registro en detalle
                                   2022-10-17 17:52:30
   NULL
```

5.8 Código SQL + Resultados: Funciones (Primera entrega)

```
135 DELIMITER //

136 CREATE FUNCTION bancos (banco INT) RETURNS INT

137 BEGIN

138 DECLARE ban INT;

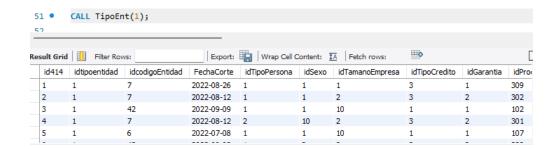
5ELECT COUNT(*) INTO ban FROM entidad WHERE idtipoentidad = banco;

140 RETURN ban;

END//

141 DELIMITER;
```

5.9 Código SQL + Resultados: procedimientos almacenados ($Primera\ entrega$)



```
#PROCEDIMIENTO ALMACENADO TIPO CREDITO

delimiter //

114 • create procedure garantia (IN garantia int)

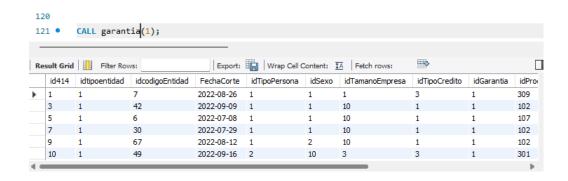
115 ⊖ BEGIN

116 SELECT * FROM rr414.detalle

where idGarantia = garantia;

END//

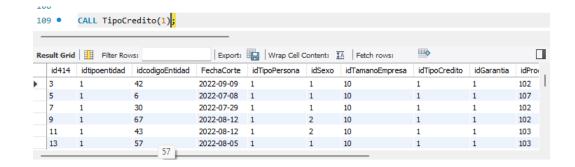
119 delimiter;
```



```
100
        #PROCEDIMIENTO ALMACENADO TIPO CREDITO
101
        delimiter //
        create procedure TipoCredito (IN tipo int)
102 •
103

⇒ BEGIN

104
        SELECT * FROM rr414.detalle
        where idTipoCredito = tipo;
105
        END//
106
        delimiter ;
107
```



- 6 Bases de Datos No-SQL (Segunda entrega)
- 6.1 Diagrama Bases de Datos No-SQL (Segunda entrega)
- 6.2 SMBD utilizado para la Base de Datos No-SQL (Segunda $_{entrega)}$

- 7 Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos (Tercera entrega)
- 7.1 Ejemplo de aplicación de ETL y Bodega de Datos $(Tercera\ entrega)$

8 Lecciones aprendidas (Tercera entrega)

9 Bibliografía

Superintendencia Financiera de Colombia (2022) Tasas de interes activas por tipo de crédito https://www.superfinanciera.gov.co/inicio/informes-y-cifras/cifras/establecimientos-de-credito/informacion-periodica/semanal-60825 el 29 de julio de 2022 Banco de la republica (2022) Tasas de interes https://www.banrep.gov.co/es el 23 de septiembre de 2022