

Proyecto de Bases de datos: Encuesta de Carga y Educación Financiera de los Hogares

Nubia Suárez Torres¹

¹Facultad de ingeniería y ciencias básicas

Universidad Central

Maestría en Analítica de Datos

Curso de Bases de Datos

Bogotá, Colombia

{¹Nubia } nsuarez@ucentral.edu.co

December 2, 2022

Contents

1	Introducción	3
2	Características del proyecto de investigación	4
2.1	Titulo del proyecto de investigación	4
2.2	Objetivo general	4
2.2.1	Objetivos específicos	4
2.3	Alcance	4
2.4	Pregunta de investigación	4
2.5	Hipotesis	4
3	Reflexiones sobre el origen de datos e información	5
3.1	¿Cual es el origen de los datos e información ?	5
3.2	¿Cuales son las consideraciones legales o eticas del uso de la información?	6
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?	6
3.4	¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto?	6
4	Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)	7
4.1	Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto	7

4.2	Diagrama modelo de datos	7
4.3	Imágenes de la Base de Datos	8
4.4	Código SQL - lenguaje de definición de datos (DDL)	10
4.5	Código SQL - Manipulación de datos (DML)	12
4.6	Código SQL + Resultados: Vistas	14
4.7	Código SQL + Resultados: Triggers	15
4.8	Código SQL + Resultados: Funciones	16
5	Bases de Datos No-SQL (<i>Segunda entrega</i>)	18
5.1	Diagrama Bases de Datos No-SQL (<i>Segunda entrega</i>)	18
5.2	SMBD utilizado para la Base de Datos No-SQL (<i>Segunda entrega</i>)	19
6	Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos (<i>Tercera entrega</i>)	21
6.1	Ejemplo de aplicación de ETL y Bodega de Datos (<i>Tercera entrega</i>)	21
7	Lecciones aprendidas (<i>Tercera entrega</i>)	22
8	Bibliografía	24

1 Introducción

La encuesta de información financiera y educación de hogares (IEFIC) que se abordara en el presente trabajo es la realizada en el 2018, la cual surge a partir de la pregunta 15 de la Gran Encuesta Integrada de Hogares (GEIH) que realizó el DANE en el 2018, a partir de esta pregunta seleccionaron la muestra para realizar la encuesta IEFIC.

Teniendo en cuenta la relevancia que tiene el factor financiero en cada uno de los hogares para poder suplir sus necesidades básicas y secundarias, y como logran cada uno de los hogares en suplirlas de acorde a sus capacidades de adquisición de dinero o endeudamiento, entre otros medios. El Banco de la República ha venido implementado la IEFIC, la cual contiene información acerca de los pagos financieros que realizan los hogares, así como de su comportamiento por modalidad de crédito.

En este sentido, la encuesta busca identificar variables financieras, de consumo y que hacen las personas normalmente para cubrir los déficit financieros, a quien recurren y si conocen las condiciones de negociación al momento de tomar un crédito, cuánto pagan a capital, interés, valor de la cuota y valor total de la deuda, buscando realizar un diagnóstico de los hogares que cuentan en la actualidad con crédito de vivienda, distinguiendo entre quienes lo han utilizado para financiar su vivienda principal, y quienes han adquirido vivienda secundaria y si los créditos que hace los utiliza para compras diferentes a las que se tenían proyectadas, también las personas y entidades con las cuales realizaron estos créditos.

2 Características del proyecto de investigación

2.1 Título del proyecto de investigación

Capacidades de endeudamiento y pago de los hogares colombianos de acuerdo a sus ingresos monetarios y sus deudas adquiridas.

2.2 Objetivo general

Agrupar los hogares de acuerdo a su oportunidad en el pago, capacidades de endeudamiento y otras formas de suplir los déficit financieros, con el fin de analizar si existe algunas relaciones notorias entre los grupos formados.

2.2.1 Objetivos específicos

- Analizar y entender la información de la encuesta y su estructura.
- Ordenar la información de tal forma que nos permita dar respuesta, sobre el comportamiento financiero de los hogares.
- Entender el concepto de educación financiera y como aplicarla puede impactar en las finanzas de las personas.
- Segmentar los hogares de acuerdo a sus capacidades monetarias y de endeudamiento, por medio del algoritmo de K-Means.

2.3 Alcance

Teniendo en cuenta que la información que se va a emplear es de una encuesta en concreto, no se podría expandir a periodos futuros, ya que se debe tener en cuenta los cambios que pueden haber dentro del hogar, como por ejemplo quedarse sin empleo, conseguir empleo, cambiar de empleo, la fluctuación de los precios de los bienes y servicios, entre otros, lo cual puede cambiar los agrupamientos que se pueden obtener.

2.4 Pregunta de investigación

¿Como la educación financiera ayudaría a fomentar buenas practicas en el manejo del dinero para suplir las necesidades básicas y secundarias en los hogares colombianos, teniendo en cuenta sus ingresos y capacidades de endeudamiento?

2.5 Hipotesis

Si los colombianos acceden a información sobre educación financiera efectiva impactaría en la reducción de la pobreza y la desigualdad, además de contribuir al desarrollo del sistema financiero.

3 Reflexiones sobre el origen de datos e información

En este mundo digital, donde los datos de las personas se han vuelto de carácter público y de manejo descontrolado, para obtener información de los mismos que les permitan a las empresas tomar decisiones de acuerdo a los resultados obtenidos, y todo esto en muchas ocasiones, los dueños de la información no se percatan o saben que sus datos andan siendo analizados.

Como se menciono anteriormente nos encontramos en un mundo digital, donde los aparatos electrónicos, las redes sociales y demás, se han vuelto parte de nuestro que hacer diario, se han vuelto indispensables en la vida social de las personas, y las empresas pioneras de estos proyectos y aplicaciones aprovechan para obtener la mayor cantidad de datos posibles, se consideraría abuso al contar con esta información, que muchas veces aceptamos que la usen, por los mismos condicionamientos que nos imponen el uso de las mismas tecnologías.

Sabemos, sin embargo, que las redes sociales, las aplicaciones de mensajería instantánea o las de geolocalización, nos fueron ofrecidas como inocentes juegos que hemos adaptado a nuestra vida cotidiana con la ilusión de estar mejor informados y conectados, tanto en la esfera personal como profesional. Algunos incluso parece que fueron diseñados para hacernos la vida más fácil, aunque estemos descubriendo rápidamente que en realidad nos están generando problemas que no habíamos imaginado. (Colmenarejo, 2017)

Orientados en los datos que se emplearan, los datos son anonimizados y son compartidos por el DANE de forma informativa, para poder dar luz a preguntas orientadas en términos nacionales y con ellas poder dar sus respuestas y orientarlas al beneficio colectivo.

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

Las fuentes de origen de la información pueden ser múltiples, verdaderamente es el propósito que se quiere lograr con esos datos lo que orienta el origen, es decir, en la antigüedad el dato nace con el fin de no perder las cuentas de un suceso en particular, en la actualidad los propósitos son muy diversos, y con ello también se diversifica el origen de ellos, ya sea de formas legales o ilegales, de dominio público o no.

Si nos orientamos a los datos que se emplearan en el presente trabajo son descargados de la pagina del DANE y son de dominio público.

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

La palabra ética ha tomado gran relevancia, se podría pensar en una redefinición de la ética para este mundo informático, ya que se debe cuestionar los alcances de la información que se obtiene en cierta manera legal pero que del todo no consensuada con el dueño de la misma. Si bien es cierto que gracias a la información las avances tecnológicos han logrado salvar vidas si lo vemos en el ámbito de la medicina, y todo esto gracias a la recopilación de la información, pero no toda información se usa para el bien común, vulnerando la privacidad del ser humano, y muchas veces violando su propia intimidad.

Al orientarse en los datos que se emplearan, se usaran solo con fines académicos.

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?

Se realizará un preprocesamiento de los datos buscando, orientados en la búsqueda de posibles correlaciones que nos permita reducir el numero de variables ya que la base cuenta con 331 variables, además se quitaran aquellas variables que no aporten a los objetivos del trabajo o que no cuenten con mas de un 70% de la informacion.

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

Tener la información de forma compacta, organizada y que me permita obtener información clave de forma rápida y clara.

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

MySQL con la interfaz Workbench

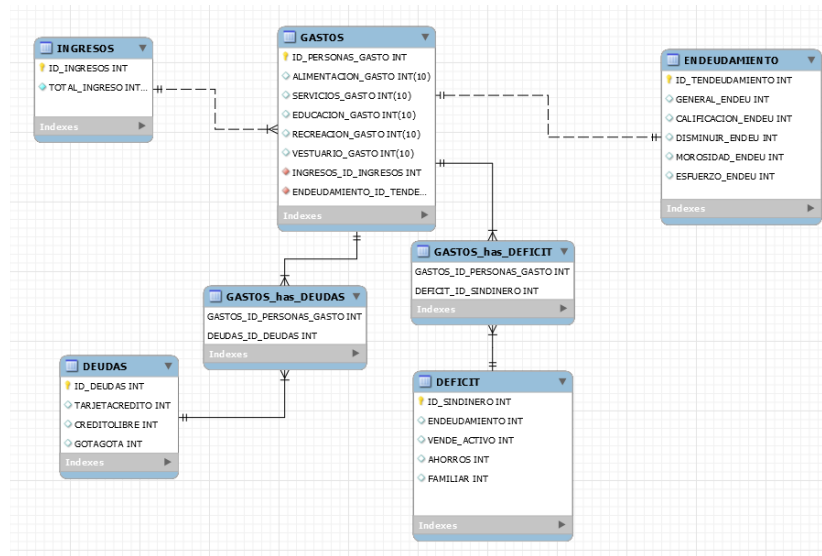
4.1 Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto

MySQL es un sistema de gestión de bases de datos relacional es de código abierto. Algunas características:

1. Arquitectura Cliente y Servidor
2. Compatibilidad con SQL
3. Vistas
4. Procedimientos almacenado
5. Desencadenantes.
6. Transacciones.

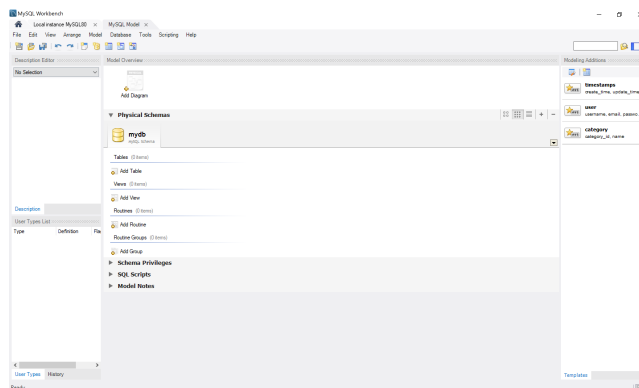
MySQL al estar basado en código abierto permite a desarrolladores disponer de una solución fiable y estandarizada para sus aplicaciones.

4.2 Diagrama modelo de datos

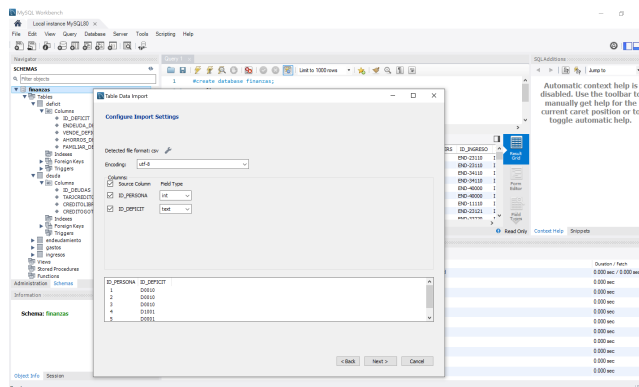


Modelo Relacional

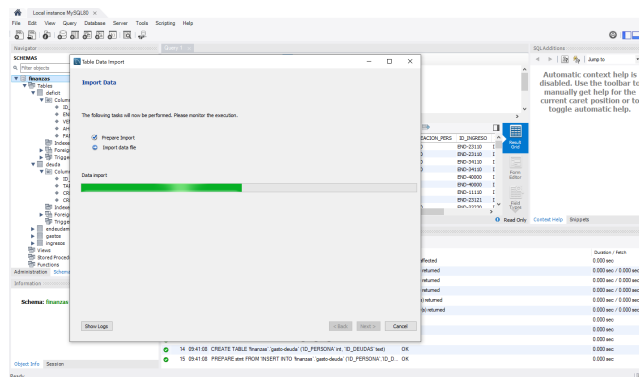
4.3 Imágenes de la Base de Datos



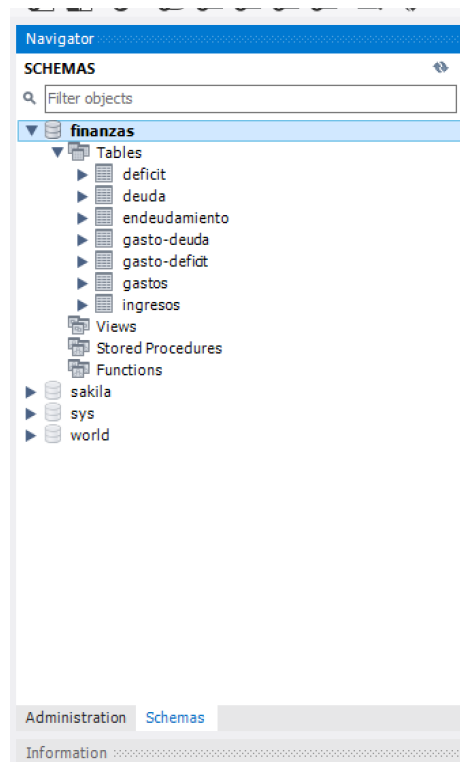
Workbench MySQL interfax



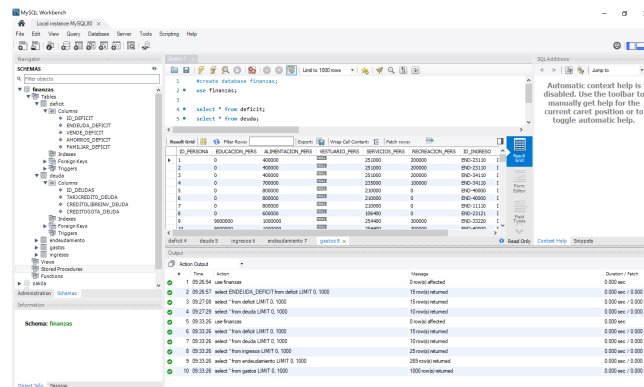
Formato tabla auxiliar Gasto-Deficit



Exportación de tabla

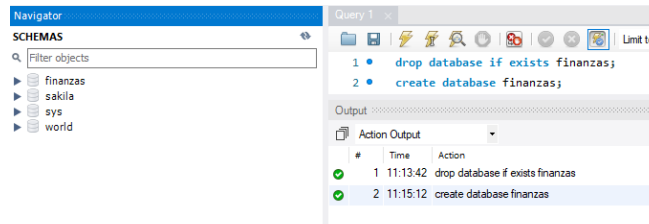


Tablas



Primera consulta de la base

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



Creado de la base

13 • `select * from deuda;`

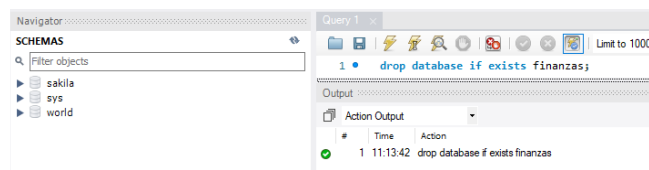
ID_DEUDAS	TARJCREDITO_DEUDA	CREDITOLIBREINV_DEUDA	CREDITOGOTA_DEUDA
DEU-000	0	0	0
DEU-100	1	0	0
DEU-111	1	1	1
DEU-112	1	1	2
DEU-121	1	2	1
DEU-122	1	2	2
DEU-211	2	1	1
DEU-212	2	1	2
DEU-221	2	2	1
DEU-222	2	2	2

Consulta antes de alter

13 • `select * from deuda;`
14 • `ALTER TABLE deuda ADD prueba2 int;`

ID_DEUDAS	TARJCREDITO_DEUDA	CREDITOLIBREINV_DEUDA	CREDITOGOTA_DEUDA	prueba2
DEU-000	0	0	0	NULL
DEU-100	1	0	0	NULL
DEU-111	1	1	1	NULL
DEU-112	1	1	2	NULL
DEU-121	1	2	1	NULL
DEU-122	1	2	2	NULL
DEU-211	2	1	1	NULL
DEU-212	2	1	2	NULL
DEU-221	2	2	1	NULL
DEU-222	2	2	2	NULL

Insertando columna (ALTER)



Borrando la base, si existe

```

15 • ALTER TABLE deuda drop prueba2;
16 • select * from deuda;

```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: [IA](#)

	ID_DEUDAS	TARJCREDITO_DEUDA	CREDITOLIBREINV_DEUDA	CREDITOGOTA_DEUDA
▶	DEU-000	0	0	0
	DEU-100	1	0	0
	DEU-111	1	1	1
	DEU-112	1	1	2
	DEU-121	1	2	1
	DEU-122	1	2	2
	DEU-211	2	1	1
	DEU-212	2	1	2
	DEU-221	2	2	1
	DEU-222	2	2	2

deuda 3 x

Output

Action Output

#	Time	Action
54	12:53:02	ALTER TABLE deuda drop prueba2
55	12:53:27	select * from deuda LIMIT 0, 1000

Borrando la columna que se creo con el ALTER

4.5 Código SQL - Manipulación de datos (DML)

```
18 • select * from gastos WHERE ID_PERSONA='20';
```

ID_PERSONA	EDUCACION_PERS	ALIMENTACION_PERS	VESTUARIO_PERS	SERVICIOS_PERS	RECREACION_PERS	ID_INGRESO	ID_ENDEUDAMIENTO
20	99	99	99	494	99	1-0	END-00000

Consulta

```
25 • UPDATE gastos
26 SET EDUCACION_PERS= '98', VESTUARIO_PERS = '99'
27 WHERE ID_PERSONA = 20;
28 • select * from gastos WHERE ID_PERSONA='20';
```

ID_PERSONA	EDUCACION_PERS	ALIMENTACION_PERS	VESTUARIO_PERS	SERVICIOS_PERS	RECREACION_PERS	ID_INGRESO	ID_ENDEUDAMIENTO
20	98	99	99	494	99	1-0	END-00000

gastos 9 x

#	Time	Action	Message
58	14:16:48	DELETE FROM gastos WHERE ID_PERSONA=20'	1 row(s) affected
59	14:17:14	select *from gastos LIMIT 0, 1000	1000 row(s) returned
60	14:19:24	INSERT INTO gastos (ID_PERSONA,EDUCACION_PERS,ALIMENTACION_PERS,VESTUARIO_PERS,SERVICIOS_PERS,RECREACION_PERS,ID_INGRESO,ID_ENDEUDAMIENTO) VALUES (20,98,99,99,494,99,1-0,END-00000)	1 row(s) affected
61	14:19:58	select *from gastos LIMIT 0, 1000	1000 row(s) returned
62	14:20:35	select *from gastos WHERE ID_PERSONA=20' LIMIT 0, 1000	1 row(s) returned
63	14:21:55	UPDATE gastos SET EDUCACION_PERS= '98', VESTUARIO_PERS = '99' WHERE ID_PERSONA=20'	1 row(s) affected Rows matched: 1 Changed: 1 Warnings: 0
64	14:22:21	select ID_INGRESO from gastos WHERE ID_PERSONA=20' LIMIT 0, 1000	1 row(s) returned
65	14:22:37	select *from gastos WHERE ID_PERSONA=20' LIMIT 0, 1000	1 row(s) returned

UPDATE

```
19 • DELETE FROM gastos WHERE ID_PERSONA='20';
```

ID_PERSONA	EDUCACION_PERS	ALIMENTACION_PERS	VESTUARIO_PERS	SERVICIOS_PERS	RECREACION_PERS	ID_INGRESO	ID_ENDEUDAMIENTO
16	0	800000	81300	171000	0	1-0	END-40000
17	0	800000	81300	171000	0	1-1201-1400	END-40000
18	1200000	300000	41600	221900	0	1-1001-1200	END-40000
19	0	250000	41600	156000	50000	1-0	END-40000
21	0	700000	98	232000	0	1-0	END-40000
22	0	700000	98	232000	0	1-801-900	END-40000
23	0	300000	60000	211000	0	1-1001-1200	END-40000
24	0	350000	98	233000	0	1-801-900	END-11210
25	0	300000	0	64900	0	1-0	END-13120
26	0	300000	0	64900	0	1-801-1000	END-13120
27	1000000	400000	46600	90000	0	1-401-500	END-40000
28	1000000	400000	46600	90000	0	1-0	END-40000

gastos 5 x

#	Time	Action	Message
50	12:49:59	use finanzas	0 row(s) affected
51	12:50:03	select *from deuda LIMIT 0, 1000	10 row(s) returned
52	12:51:14	ALTER TABLE deuda ADD prueba2 int	0 row(s) affected Records: 0 Duplicates: 0 Warnings: 0
53	12:51:25	select *from deuda LIMIT 0, 1000	10 row(s) returned
54	12:53:02	ALTER TABLE deuda drop prueba2	0 row(s) affected Records: 0 Duplicates: 0 Warnings: 0
55	12:53:27	select *from deuda LIMIT 0, 1000	10 row(s) returned
56	12:58:57	SET SQL_SAFE_UPDATES = 0	0 row(s) affected
57	13:00:06	select *from gastos WHERE ID_PERSONA=20' LIMIT 0, 1000	1 row(s) returned
58	14:16:48	DELETE FROM gastos WHERE ID_PERSONA=20'	1 row(s) affected
59	14:17:14	select *from gastos LIMIT 0, 1000	1000 row(s) returned

DELETE

```

21 • INSERT INTO gastos (ID_PERSONA, EDUCACION_PERS, ALIMENTACION_PERS, VESTUARIO_PERS,
22   SERVICIOS_PERS, RECREACION_PERS, ID_INGRESO, ID_ENDEUDAMIENTO)
23   VALUES (20,99,99,99,494,99,'I-0','END-000000');
24 • select * from gastos WHERE ID_PERSONA=20;

```

ID_PERSONA	EDUCACION_PERS	ALIMENTACION_PERS	VESTUARIO_PERS	SERVICIOS_PERS	RECREACION_PERS	ID_INGRESO	ID_ENDEUDAMIENTO
20	99	99	99	494	99	I-0	END-000000

gastos 7 x

#	Time	Action	Message
55	12:53:27	select * from deuda LIMIT 0, 1000	10 row(s) returned
56	12:58:57	SET SQL_SAFE_UPDATES = 0	0 row(s) affected
57	13:00:06	select * from gastos WHERE ID_PERSONA=20 LIMIT 0, 1000	1 row(s) returned
58	14:16:48	DELETE FROM gastos WHERE ID_PERSONA=20	1 row(s) affected
59	14:17:14	select * from gastos LIMIT 0, 1000	1000 row(s) returned
60	14:19:24	INSERT INTO gastos (ID_PERSONA, EDUCACION_PERS, ALIMENTACION_PERS, VESTUARIO_PERS, SERVICIOS_PERS, RECREACION_PERS, ID_INGRESO, ID_ENDEUDAMIENTO) VALUES (20,99,99,99,494,99,'I-0','END-000000');	1 row(s) affected
61	14:19:58	select * from gastos LIMIT 0, 1000	1000 row(s) returned
62	14:20:35	select * from gastos WHERE ID_PERSONA=20 LIMIT 0, 1000	1 row(s) returned

INSERT

4.6 Código SQL + Resultados: Vistas

The screenshot shows a SQL IDE with a database schema on the left and a query editor on the right. The query editor contains the following SQL code:

```

32 CREATE VIEW deudas AS
33 SELECT TARCREDITO_DEUDA, CREDITOLIBREINV_DEUDA, CREDITOGOTA_DEUDA
34 FROM deudas
35 WHERE TARCREDITO_DEUDA=1;
36 SELECT * FROM deudas;

```

The results pane shows the following data:

TARCREDITO_DEUDA	CREDITOLIBREINV_DEUDA	CREDITOGOTA_DEUDA
1	0	0
1	1	1
1	1	2
1	2	1
1	2	2

The bottom pane shows the execution log with the following messages:

```

61 14:15:58 select * from gastos LIMIT 0, 1000 1000 rows returned
62 14:20:35 select * from gastos WHERE ID_PERSONA=20 LIMIT 0, 1000 1 row(s) returned
63 14:21:55 UPDATE gastos SET EDUCACION_PERSONA=99, VESTUARIO_PERSONA=99 WHERE ID_PERSONA=20 1 row(s) affected Rows matched: 1 Changed: 1 Warnings: 0
64 14:22:21 select ID_INGRESO from gastos WHERE ID_PERSONA=20 LIMIT 0, 1000 1 row(s) returned
65 14:22:37 select * from gastos WHERE ID_PERSONA=20 LIMIT 0, 1000 1 row(s) returned
66 14:23:13 SET SQL_SAFE_UPDATES=1 0 row(s) affected
67 14:33:43 CREATE VIEW deudas AS SELECT TARCREDITO_DEUDA, CREDITOLIBREINV_D... 0 row(s) affected
68 14:38:04 SELECT * FROM deudas LIMIT 0, 1000 5 row(s) returned

```

VISTA 1

The screenshot shows a SQL IDE with a database schema on the left and a query editor on the right. The query editor contains the following SQL code:

```

45 CREATE VIEW alimentacion AS
46 SELECT ALIMENTACION_PERSONAS
47 FROM gastos
48 WHERE ALIMENTACION_PERSONAS < 1000000;
49 SELECT * FROM alimentacion;

```

The results pane shows the following data:

ALIMENTACION_PERSONAS
250000
240000
200000
200000
200000
200000
99
99
99
200000
230000
100000

The bottom pane shows the execution log with the following messages:

```

89 15:02:53 CREATE VIEW alimentacion AS SELECT ALIMENTACION_PERSONAS FROM gastos WHERE... 0 row(s) affected
90 15:03:52 SELECT * FROM Ingresoconida LIMIT 0, 1000 Error Code: 1146. Table 'finanzas.ingresoconida' doesn't exist
91 15:04:37 SELECT * FROM alimentacion LIMIT 0, 1000 1000 row(s) returned

```

VISTA 2

4.7 Código SQL + Resultados: Triggers

```
drop trigger if exists nuevos_hogares
DELIMITER $$
• create trigger nuevos_hogares after insert on gastos for each row
begin
  insert into nuevos_hogares (educacion,alimentacion,vestuario,servicios,recreacion)
  values (new.gasto_EDUCACION_PERS,new.gasto_ALIMENTACION_PERS,new.gasto_VESTUARIO_PERS,
  new.gasto_SERVICIOS_PERS,new.gasto_RECREACION_PERS);
end $$
DELIMITER ;
```

TRIGGER

4.8 Código SQL + Resultados: Funciones

```
70
71 • SELECT AVG(ALIMENTACION_PERS) AS Averagealimentacion FROM gastos;
72
```

Averagealimentacion
564082.1556

Promedio de gastos de comida

```
76
77 • SELECT ID_INGRESO, AVG(ALIMENTACION_PERS), COUNT(*) AS cuentaALIMENTACION FROM gastos
78 GROUP BY ID_INGRESO;
79
```

ID_INGRESO	AVG(ALIMENTACION_PERS)	cuentaALIMENTACION
I-1501-1800	579242.9748	2066
I-1401-1500	563627.7968	1417
I-1801-2000	607938.3937	1359
I-401-500	484039.1791	1513
I-0	600150.0009	12219
I-701-800	515296.6185	3114
I-3001-4000	706954.4218	1183
I-5001-8000	804827.9027	781
I-2001-2500	627566.5363	1486
I-1201-1400	538206.0990	2636
I-1001-1200	514500.6816	4341
I-801-900	502996.5441	2757
I-901-1000	511946.6622	4251
I-201-300	486785.7624	1490
I-100-150	483008.6397	1260
I-601-700	533338.4160	2238
I-2501-3000	670911.1940	1098
I-151-200	495993.6553	998
I-501-600	492582.7254	1373
I-4001-5000	780785.5763	701
I-12000-20...	1079528.3649	148
I-301-400	481679.8776	1462

Result 26 x

Output

#	Time	Action	Message
105	16:17:08	SELECT ID_INGRESO, ALIMENTACION_PERS, COUNT(*) AS cuentaALIMENTACION F...	25 row(s) returned
106	16:17:48	SELECT ID_INGRESO, AVG(ALIMENTACION_PERS), COUNT(*) AS cuentaALIMENTAC...	25 row(s) returned

Promedio de alimentación teniendo en cuenta el ingreso

```
72 • SELECT AVG(SERVICIOS_PERS) AS Average servicios FROM gastos;
73
```

Averageservicios
243714.1622

Promedio de gastos de servicios

74 •

SELECT ID_INGRESO, COUNT(*) AS cuentaingreso FROM gastos

75

GROUP BY ID_INGRESO;

Result Grid

Filter Rows:

Export:

Wrap Cell Content:

ID_INGRESO	cuentaingreso
I-1501-1800	2066
I-1401-1500	1417
I-1801-2000	1359
I-401-500	1513
I-0	12219
I-701-800	3114
I-3001-4000	1183
I-5001-8000	781
I-2001-2500	1486
I-1201-1400	2636
I-1001-1200	4341
I-801-900	2757
I-901-1000	4251
I-201-300	1490
I-100-150	1260
I-601-700	2238
I-2501-3000	1098
I-151-200	998
I-501-600	1373
I-4001-5000	701
I-12000-20...	148
I-301-400	1462
I-8001-12000	325

Result 24 x

Output

Action Output

#	Time	Action	Message
102	16:11:14	SELECT COUNT(*) AS cuentaingreso FROM gastos GROUP BY ID_INGRESO LIMIT 0...	25 row(s) returned
103	16:11:41	SELECT ID_INGRESO, COUNT(*) AS cuentaingreso FROM gastos GROUP BY ID_ING...	25 row(s) returned

Cuenta de hogares segun sus ingresos

82 •

SELECT ID_DEFICIT,count(*),ENDEUDA_DEFICIT FROM deficit WHERE ENDEUDA_DEFICIT LIKE '1%'

83

Result Grid

Filter Rows:

Export:

Wrap Cell Content:

ID_DEFICIT	count(*)	ENDEUDA_DEFICIT
D1000	8	1

Cuenta de cuantas personas acuden a endeudarse

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

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

U CENTRAL > PROJECT 0 > DATABASES

Cluster0

VERSION: 5.0.13 REGION: AWS N. Virginia (us-east-1)

Overview Real Time Metrics Collections Search Profiler Performance Advisor Online Archive Cmd Line Tools

DATABASES: 11 COLLECTIONS: 24

+ Create Database

Search Namespaces

- EducacionFinanciera
 - EducacionFinanciera
- librosbases
- sample_airbnb

EducacionFinanciera.EducacionFinanciera

STORAGE SIZE: 4KB LOGICAL DATA SIZE: 0B TOTAL DOCUMENTS: 0 INDEXES TOTAL SIZE: 4KB

Find Indexes Schema Anti-Patterns Aggregation Search Indexes

INSERT DOCUMENT

FILTER { field: 'value' } OPTIONS Apply Reset

STORAGE SIZE: 4KB LOGICAL DATA SIZE: 0B TOTAL DOCUMENTS: 0 INDEXES TOTAL SIZE: 4KB

Find Indexes Schema Anti-Patterns Aggregation Search Indexes

FILTER { field: 'value' }

```
_id: ObjectId('6372f4a84bbb68b2f62e85ae')
Autores: "Ministerio de educación nacional"
Año: "2021"
Nombre: "Nueva Pangea"
Tipo: "Artículo con vinculo de video"
URL: "https://www.mineducacion.gov.co/portal/salaprensa/Comunicados/408010:E_"

_id: ObjectId('6372f6d04bbb68b2f62e85af')
Autores: "Carlos Cano, Maria Esguerra, Nidia García, Leonardo Rueda"
Año: "2014"
Nombre: "Inclusión financiera en Colombia"
Tipo: "Artículo"
URL: "https://www.banrep.gov.co/sites/default/files/eventos/archivos/sem_357_"
```

EducacionFinanciera.EducacionFinanciera

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

Mongo es una base de datos documental (NOSQL). El cual trabaja con documentos los cuales forman colecciones; comparando con las bases de datos relacionales, las colecciones son tablas. A diferencia de las tablas estas permiten guardar documentos con diferentes formatos, no están sometidos a un esquema fijo y los documentos son registros.

Mongo está escrito en C++, lo que hace que sea bastante rápida a la hora de ejecutar tareas. Actualmente lidera las bases de datos NOSQL.

Ventajas al usar Mongo

- Velocidad: Las bases de datos pueden ser muy rápidas con respecto a las bases de datos SQL, un error común es tratar de que una base de datos NoSQL funcione como una base de datos relacional, mientras uses el motor de bases de datos para su propósito inicial vas a obtener las mejores características del motor.
- Esquemas de bases de datos flexibles: Podrías guardar diferentes tipos de información bajo la misma tabla, diferentes registros con diferentes campos en la misma colección.
- Escalabilidad: Es la forma en la que la base de datos se expande a la hora

de almacenar información, las bases de datos NoSQL están diseñadas para escalar fácilmente de manera horizontal, generando beneficios como alta disponibilidad.

Desventajas:

- No son transaccionales: Las transacciones son consultas generalmente de escritura, donde si ocurre un error, se hace rollback a todo lo que se hizo en la base de datos. NoSQL no garantiza que eso suceda, es desventaja si tu aplicación requiere que la información se guarde completa y se requiere volver a estados anteriores.
- No tiene joins: Porque de hecho no tiene relaciones. Con MongoDB muchas personas tratan de simular relaciones con los índices pero en ese punto Mongo pierde sentido, si estás usando una base de datos NoSQL debes evitar usar joins.

6 Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos *(Tercera entrega)*

6.1 Ejemplo de aplicación de ETL y Bodega de Datos *(Tercera entrega)*

7 Lecciones aprendidas (*Tercera entrega*)

- Cada uno de los procesos que conllevan formar una bases de datos es muy importante para garantizar el buen funcionamiento de la bases de datos.
- La solvencia, versatilidad y consistencia de las bases de datos relacionales, hacen de este lenguaje uno de los más importantes útiles y demandados para analizar datos.
- Entender el problema a trabajar es de suma importancia ya que nos brinda orientación de como plantear el proyecto de forma eficiente, y poder dar respuesta a los objetivos.
- La calidad de los datos es importante, por ello se debe realizar un proceso de limpieza y transformación donde lo requiera.

Anexo

¿Qué es la capacidad de carga?	hoja nueva
¿Cómo se realiza la relación de los ingresos y nivel de endeudamiento de los hogares con la educación financiera?	objetivos
Revisar un poco el objetivo de la segmentación, en el objetivo general hablas solo de segmentar, pero en lo que nos interesa habla de un nivel de educación financiera y como los hogares suplen deficit, sería chever ver ese tema reflejado en los objetivos y en los específicos de pronto plantear un poco la metodología a emplear en esta segmentación.	implementar
Descriptivos y segmentación, recomiendo para investigación saber si se mejora o no las finanzas personales de los hogares y si estos cambios tiene una relación directa con una mejora en su educación financiera.	
Este trabajo se integraría perfectamente con el trabajo de investigación de los compañeros Oscar Mauricio y Lizeth.	
¿Cómo resuelve al análisis la respuesta de investigación?	pregunta
Soportar las cifras planteadas como parte de la investigación	implementar
¿Con la data disponible cómo se puede evaluar la hipótesis?	hoja nueva
que porcentaje de participación tienen las variables seleccionadas en la base de datos.	hoja nueva
Como se presentara la información final, por familia, estrato, ubicación?	hoja nueva
Se realizaran en Paralelo modelos de Segmentación y clasificación?	hoja nueva
Tener en cuenta contextos económicos de los hogares en relación con los ingresos.	implementar
Revisar la hipótesis para no tener una visión sesgada y ser mas objetivo.	hipotesis

Observaciones

- Que es capacidad de carga.
RTA Hace referencia a la cuantía a la que tiene que hacer frente un individuo o personalidad jurídica para devolver el principal y los intereses de los préstamos.
- Relación de ingresos y nivel de endeudamiento con la educación financiera.
RTA La idea es establecer si existe esa relación, si las personas que han obtenido algún tipo de educación financiera verdaderamente les resulta útil y aplicable.
- ¿Con la data disponible cómo se puede evaluar la hipótesis?
RTA En las variables de la encuesta realizan preguntas de como manejan el dinero, si realizan inversiones, proponen situaciones hipotéticas con respecto al dinero, es decir, por ejemplo, si tuvieran un ingreso adicional en el mes, en que lo invertirían, a partir de esas preguntas asumen si la persona tiene o no esa educación financiera. No se tuvieron en cuenta en el modelo de bases de datos por que eran preguntas abiertas las cuales necesito procesar.
- Que porcentaje de participación tienen las variables seleccionadas en la base de datos.
RTA 80% de participación, la encuesta cuenta con 62839 registros y al quitar los registros con muy poca información quedaron 50273.
- Como se presentara la información final, por familia, estrato, ubicación?
RTA Pienso que por estrato suena interesante, mas sin embargo creo que esperaría a ver el comportamiento de los cluster.
- Se realizaran en paralelo modelos de Segmentación y clasificación?
RTA Me gustaría realizar una segmentación inicial y luego poder ingresar a cada segmento y analizarlo y si se pudiera realizar una clasificación. Por el momento la idea es realizar una segmentación.

8 Bibliografía

1. Departamento Administrativo Nacional de Estadística. (2022). Datos. COLOMBIA - Encuesta de Carga Financiera y Educación Financiera de los Hogares - IEFIC-2017 -2018. Disponible en:
https://microdatos.dane.gov.co/catalog/626/get_microdata
2. Departamento Administrativo Nacional de Estadística. (2013).Diccionario de datos. COLOMBIA - Encuesta de Carga Financiera y Educación Financiera de los Hogares - IEFIC-2017 -2018 Disponible en:
https://microdatos.dane.gov.co/catalog/626/data_dictionary
3. Rosa Colmenarejo Fernández. Una ética para Big data: Introducción a la gestión ética de datos masivos. Editorial UOC, 2017.
4. Saray Córdoba González. La ética de la información. Bibliotecas, 23(1), 2005.