

Proyecto de Bases de datos para la Predicción de la probabilidad que tienen los niños de los hogares colombianos menores de 12 años de la GEIH, de acceso a la educación superior.

Diana Isabel Arévalo Durán

Universidad Central

Maestría en Analítica de Datos

Curso de Bases de Datos

Bogotá, Colombia

darevalod@ucentral.edu.co

November 20, 2022

Contents

1	Introducción	3
2	Características del proyecto	3
2.1	Titulo del proyecto	3
2.2	Objetivo general	3
2.2.1	Objetivos especificos	3
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 terminos de la calidad y la consolidación?	6
3.4	¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto?	7
4	Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)	8

4.1	Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto	8
4.2	Diagrama modelo de datos	8
4.3	Imágenes de la Base de Datos	9
4.4	Código SQL - lenguaje de definición de datos (DDL)	12
4.5	Código SQL - Manipulación de datos (DML)	15
4.6	Código SQL + Resultados: Vistas	18
4.7	Código SQL + Resultados: Triggers	18
4.8	Código SQL + Resultados: Funciones	19
4.9	Código SQL + Resultados: procedimientos almacenados	19
5	Bases de Datos No-SQL	20
5.1	Diagrama Bases de Datos No-SQL	20
5.2	SMBD utilizado para la Base de Datos No-SQL	22
6	Lecciones aprendidas	24

1 Introducción

El presente trabajo se realiza con base en los datos recolectados en la gran encuesta de los hogares Colombianos 2020, con el objetivo de predecir la probabilidad de que los niños de Colombia, menores de 12 años al 2020, en los próximos 10 años accedan a educación superior o universitaria.

Para esto, inicialmente se realiza normalización de la base de datos obtenida del Archivo Nacional de Datos (ANDA) que permite la consulta por temática, operación estadística y año, de los diferentes microdatos anonimizados dispuestos al público en la página web del DANE, haciendo uso del sistema manejador de base de datos MySQL para crear un modelo de base de datos que permita el acceso eficiente a la información.

Posteriormente se aplican métodos de estadística exploratoria multidimensional, resumiendo el conjunto de variables en unas pocas nuevas variables, buscando agrupaciones en los datos y clasificando para identificar relaciones entre variables. Finalmente se contrastan los resultados contra la hipótesis planteada.

4.3 Imágenes de la Base de Datos Figure 2: Características hogar Figure 3: Educación de integrantes hogar Figure 4: Hogar 9 Figure 5: Otros ingresos Figure 6: Ocupación Figure 7: Seguridad Social 10 Figure 8: Integrantes hogar 11 13 14 4.5 Código SQL - Manipulación de datos (DML) Figure 9: Sentencia 1 15 Figure 10: Sentencia 2 Figure 11: Sentencia 3 16 Figure 12: Sentencia 4 Figure 13: Sentencia 5 17 Figure 14: Sentencia 6 4.6 Código SQL + Resultados: Vistas Figure 15: Vista 1 18

2 Características del proyecto

2.1 Título del proyecto

Predicción de la probabilidad que tienen los niños de los hogares colombianos menores de 12 años de la GEIH, de acceso a la educación superior.

2.2 Objetivo general

Predecir la probabilidad de que los niños de Colombia, menores de 12 años al 2020, en los próximos 10 años accedan a educación superior o universitaria.

2.2.1 Objetivos específicos

- Selección de variables de interés de la base obtenida.
- Diseño de modelo de base de datos en el SMBD Mysql.
- Carga de la base de datos al SMBD Mysql
- Generación de scripts para la visualización y análisis de los datos cargados.

- A través del uso de métodos de reducción de dimensionalidad generar visualizaciones de datos que permitan la fácil comprensión de los datos de la encuesta general de hogares colombianos.
- Por medio de la clasificación y regresión, analizar comportamientos para predecir la probabilidad de acceso a educación superior de los niños menores de 12 años.

2.3 Alcance

El alcance de este proyecto será la predicción probabilidad de que los niños de Colombia, menores de 12 años al 2020, en el futuro accedan a educación superior o universitaria.

2.4 Pregunta de investigación

Entre 0 y 12 años ¿Qué probabilidad tienen los niños de los hogares colombianos de la Gran encuesta integrada de hogares 2020, de acceder a la educación superior o universitaria de acuerdo con condiciones de empleo, fuentes de ingreso y características generales de los hogares?.

2.5 Hipotesis

La probabilidad de acceso a la educación superior o universitaria de los niños en Colombia está determinada por el nivel educativo de los padres, condiciones de empleo e ingresos.

3 Reflexiones sobre el origen de datos e información

La Gran encuesta integrada de hogares (GEIH) es una encuesta que se realiza en Colombia desde el año 2006, mediante la cual se solicita información sobre las condiciones de empleo de las personas, además de las características generales de la población como sexo, edad, estado civil y nivel educativo, se pregunta sobre sus fuentes de ingresos. Esta encuesta proporciona indicadores a nivel nacional del mercado laboral en Colombia que permiten conocer entre otros aspectos: la tasa de ocupación, la tasa de desocupación, la rama de actividad en que se desempeñan los colombianos y su remuneración, así como, el comportamiento del mercado laboral para jóvenes, mujeres y otros grupos poblacionales específicos.

Tiene una muestra total anual de aproximadamente 232.000 hogares, lo que la hace la de mayor cobertura a nivel nacional, a través de los años ha ido incrementando la cobertura y actualmente la información se recolecta en 13 ciudades y sus áreas metropolitanas, 11 ciudades intermedias, 8 capitales de nuevos departamentos; cabeceras y centros poblados y rural disperso en más de 443 municipios del país.

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

Los datos corresponden a la Gran Encuesta Integrada de Hogares Colombianos 2020 realizada por el Departamento Administrativo Nacional de Estadística - DANE, a la población civil no institucional residente en hogares particulares. Esta población se estima con base en los censos de población, las estadísticas vitales y de migración.

Estos datos fueron obtenidos del Archivo Nacional de Datos (ANDA) que permite la consulta por temática, operación estadística y año, de los diferentes microdatos anonimizados dispuestos al público en la página web del DANE.

En la información publicada se puede encontrar la ficha metodológica de la encuesta, diccionarios y documentación de información general con el diseño de la encuesta, la cual se encuentra dividida en 18 capítulos con 224 preguntas:

1. Identificación, 9 preguntas
2. Vivienda, 5 preguntas
3. Datos del hogar, 15 preguntas
4. Registro de personas, 4 preguntas
5. Características generales, 8 preguntas
6. Seguridad social en salud, 7 preguntas
7. Educación, 5 preguntas
8. Fuerza de trabajo, 14 preguntas
9. Ocupados, 65 preguntas
10. Desocupados, 14 preguntas
11. Inactivos, 10 preguntas

- 12.Otras actividades, 1 pregunta
- 13.Ingresos no laborales, 4 preguntas
- 14.Módulo de Micronegocios, 6 preguntas
- 15.Módulo Formación para el trabajo, 16 preguntas
- 16.Módulo de Fecundidad, 11 preguntas
- 17.Módulo de Migración, 7 preguntas
- 18.Módulo de Trabajo Infantil, 23 preguntas

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

Toda la información recolectada para los censos y encuestas de los procesos estadísticos del DANE está protegida por la Ley 79 de 1993 o Ley de Reserva Estadística.

Según esta ley los datos suministrados al DANE a través de censos, encuestas u operaciones estadísticas ”no podrán darse a conocer al público ni a las entidades u organismos oficiales, ni a las autoridades públicas, sino únicamente en resúmenes numéricos”.

En consonancia con la Ley 1581 de 2012, las bases de datos y archivos regulados por la Ley 79 de 1993 están exceptuadas del régimen de protección de datos personales. Sin embargo, deben aplicarse los principios establecidos para su protección como son: el principio de legalidad en materia de tratamiento de datos, de finalidad, de libertad, de veracidad o calidad, de transparencia, de acceso y circulación restringida, de seguridad y de confidencialidad.

Por lo anterior, las bases anonimizadas y disponibles en el ANDA han surtido un proceso de modificación y transformación de los datos originales, con el objetivo de garantizar la confidencialidad de las unidades de análisis, por lo cual es necesario tener en cuenta estas modificaciones para el uso de la información.

3.3 ¿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?

Los datos obtenidos están representados con variables categóricas y el nombre de estas es dado con códigos, por lo que se debe realizar listas de homologación tanto para los nombres como para las categorías, con el objetivo de poder generar información sencilla y entendible.

Debido a que las bases publicadas por el DANE se encuentran anonimizadas y han surtido un proceso de modificación y transformación de los datos originales, la estructura de la base, entregada en el documento técnico de la DIAN, no

corresponde con la publicada en la página, es por esto que se debe trabajar en la identificación de variables y homologaciones disponibles.

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

Se hace uso del sistema de base de datos con el objetivo de organizar los datos obtenidos para poder acceder a búsquedas eficientes y consultas simplificadas que puedan ser reutilizadas.

Del sistema se espera extraer bases de datos unificadas con las variables más importantes de la operación estadística para realizar análisis de correspondencia.

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

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

El Sistema Manejador de Bases de Datos utilizado para el proyecto es MySQL y dentro de sus principales características están:

- Sistema de base de datos relacional de uso libre y gratuito.
- Soporta gran cantidad de datos, incluso con más de 50 millones de registros.
- Ejecución de transacciones y uso de claves foráneas.
- Conectividad segura.

4.2 Diagrama modelo de datos

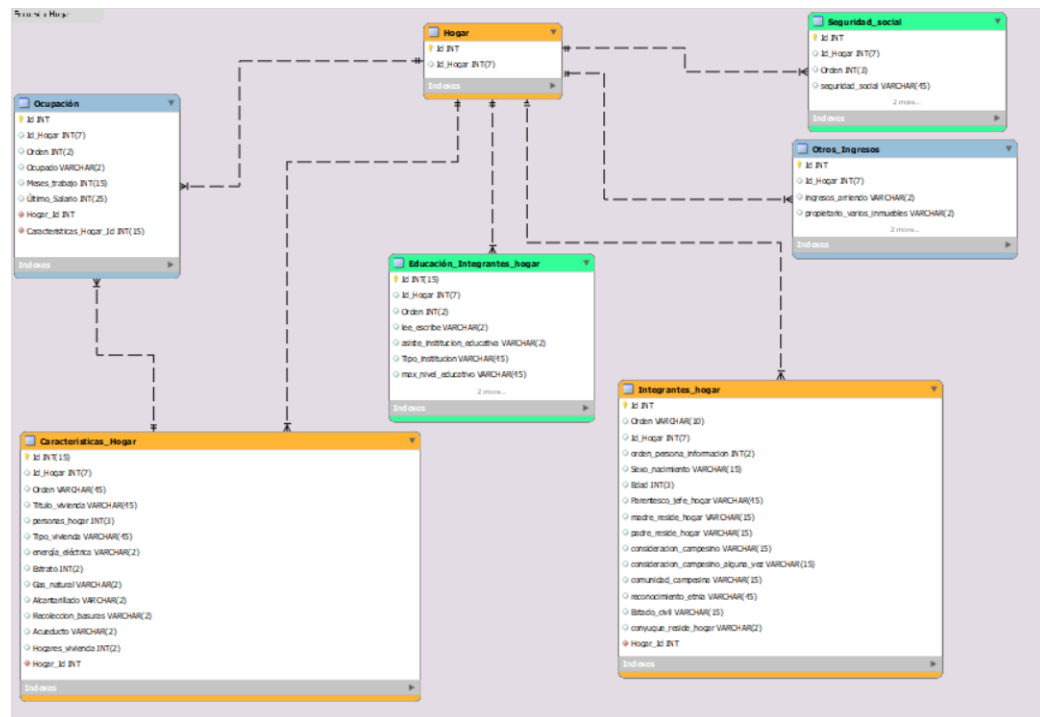
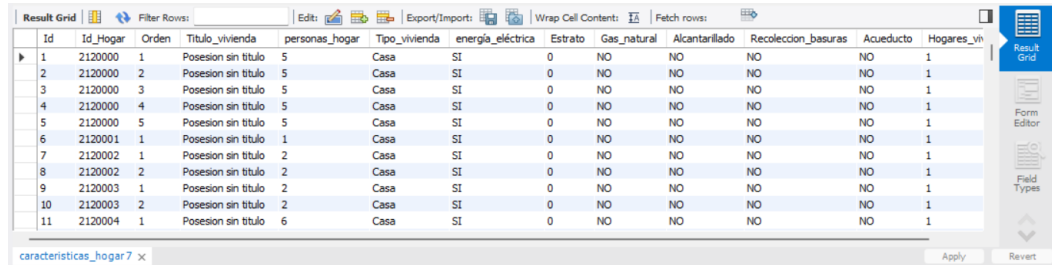


Figure 1: Diagrama Entidad Relación

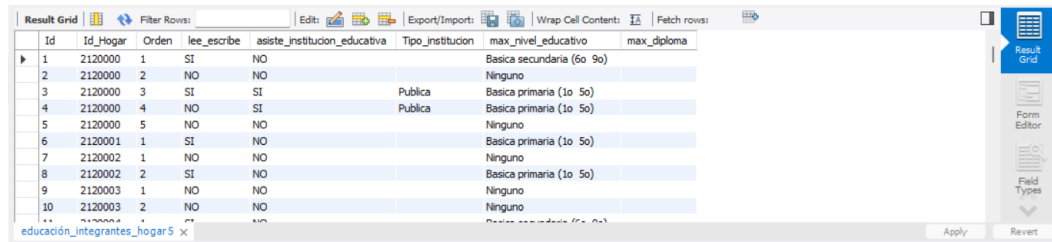
4.3 Imágenes de la Base de Datos



The screenshot shows a database interface with a table titled 'caracteristicas_hogar 7'. The table has 12 columns: Id, Id_Hogar, Orden, Titulo_vivienda, personas_hogar, Tipo_vivienda, energia_eléctrica, Estrato, Gas_natural, Alcantarillado, Recolección_basuras, Acueducto, and Hogares_vh. The data is organized into 11 rows, each representing a different household.

Id	Id_Hogar	Orden	Titulo_vivienda	personas_hogar	Tipo_vivienda	energia_eléctrica	Estrato	Gas_natural	Alcantarillado	Recolección_basuras	Acueducto	Hogares_vh
1	2120000	1	Posesion sin título	5	Casa	SI	0	NO	NO	NO	NO	1
2	2120000	2	Posesion sin título	5	Casa	SI	0	NO	NO	NO	NO	1
3	2120000	3	Posesion sin título	5	Casa	SI	0	NO	NO	NO	NO	1
4	2120000	4	Posesion sin título	5	Casa	SI	0	NO	NO	NO	NO	1
5	2120000	5	Posesion sin título	5	Casa	SI	0	NO	NO	NO	NO	1
6	2120001	1	Posesion sin título	1	Casa	SI	0	NO	NO	NO	NO	1
7	2120002	1	Posesion sin título	2	Casa	SI	0	NO	NO	NO	NO	1
8	2120002	2	Posesion sin título	2	Casa	SI	0	NO	NO	NO	NO	1
9	2120003	1	Posesion sin título	2	Casa	SI	0	NO	NO	NO	NO	1
10	2120003	2	Posesion sin título	2	Casa	SI	0	NO	NO	NO	NO	1
11	2120004	1	Posesion sin título	6	Casa	SI	0	NO	NO	NO	NO	1

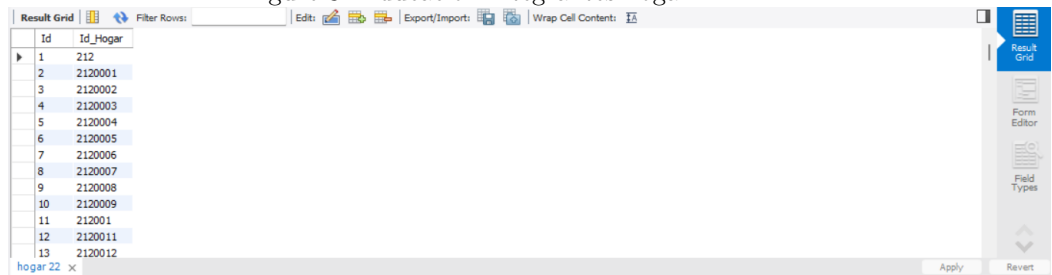
Figure 2: Características hogar



The screenshot shows a database interface with a table titled 'educación_integrantes_hogar 5'. The table has 7 columns: Id, Id_Hogar, Orden, lee_escribe, asiste_institucion_educativa, Tipo_institucion, max_nivel_educativo, and max_diploma. The data is organized into 10 rows, each representing a different household member.

Id	Id_Hogar	Orden	lee_escribe	asiste_institucion_educativa	Tipo_institucion	max_nivel_educativo	max_diploma
1	2120000	1	SI	NO		Basica secundaria (6o 9o)	
2	2120000	2	NO	NO		Ninguno	
3	2120000	3	SI	SI	Publica	Basica primaria (1o 5o)	
4	2120000	4	NO	SI	Publica	Basica primaria (1o 5o)	
5	2120000	5	NO	NO		Ninguno	
6	2120001	1	SI	NO		Basica primaria (1o 5o)	
7	2120002	1	NO	NO		Ninguno	
8	2120002	2	SI	NO		Basica primaria (1o 5o)	
9	2120003	1	NO	NO		Ninguno	
10	2120003	2	NO	NO		Ninguno	

Figure 3: Educación integrantes hogar



The screenshot shows a database interface with a table titled 'hogar 22'. The table has 2 columns: Id and Id_Hogar. The data is organized into 13 rows, each representing a different household member.

Id	Id_Hogar
1	212
2	2120001
3	2120002
4	2120003
5	2120004
6	2120005
7	2120006
8	2120007
9	2120008
10	2120009
11	212001
12	2120011
13	2120012

Figure 4: Hogar

Result Grid					
Filter Rows:					
Edit: Export/Import: Wrap Cell Content: Fetch rows:					
Id	Id_Hogar	Orden	ingresos_arriendo	propietario_varios_inmuebles	ayudas_gobierno
1	2120000	1	NO	NO	NO
2	2120000	2	NO	NO	NO
3	2120000	3			NO
4	2120000	4			NO
5	2120000	5			NO
6	2120001	1	NO	NO	NO
7	2120002	1	NO	NO	NO
8	2120002	2	SI	NO	NO
9	2120003	1	NO	NO	NO
10	2120003	2	NO	NO	NO
11	2120004	1	NO	NO	SI
12	2120004	2			NO

otros_ingresos 13 x

Apply Revert

Figure 5: Otros ingresos

Result Grid						
Filter Rows:						
Edit: Export/Import: Wrap Cell Content: Fetch rows:						
Id	Id_Hogar	Orden	Ocupado	Ocupacion	Meses_trabajo	Último_Salario
9	2120003	1	SI	Asalariado	240	0
17	2120005	1	SI	Asalariado	8	900000
32	2120008	1	SI	Asalariado	10	908526
38	2120009	2	SI	Asalariado	2	600000
45	2120010	3	SI	Asalariado	24	0
79	2120020	2	SI	Asalariado	24	908526
86	2120021	3	SI	Asalariado	3	0
100	2120024	2	SI	Asalariado	24	0
122	2120037	3	SI	Asalariado	8	0
125	2120039	1	SI	Asalariado	492	2400000
133	2120042	4	SI	Asalariado	72	1200000
137	2120043	3	SI	Asalariado	12	908526

ocupación 11 x

Apply Revert

Figure 6: Ocupación

Result Grid				
Filter Rows:				
Edit: Export/Import: Wrap Cell Content: Fetch rows:				
Id	Id_Hogar	Orden	seguridad_social	Regimen
1	2120000	1	SI	Subsidiado
2	2120000	2	SI	Subsidiado
3	2120000	3	SI	Subsidiado
4	2120000	4	SI	Subsidiado
5	2120000	5	SI	Subsidiado
6	2120001	1	SI	Subsidiado
7	2120002	1	SI	Subsidiado
8	2120002	2	SI	Subsidiado
9	2120003	1	SI	Subsidiado
10	2120003	2	SI	Subsidiado
11	2120004	1	SI	Subsidiado
12	2120004	2	SI	Subsidiado

seguridad_social 15 x

Apply Revert

Figure 7: Seguridad Social

Result Grid											
Filter Rows:											
Edit: Export/Import: Wrap Cell Contents: Fetch rows:											
	Id	Orden	Id_Hogar	orden_persona_informacion	Sexo_nacimiento	Edad	Parentesco_jefe_hogar	madre_reside_hogar	padre_reside_hogar	consideracion_campesino?	conside
▶	3	3	2120000	2	Masculino	7	Hijo hijastro	SI	SI	NO	NO
	4	4	2120000	2	Masculino	6	Hijo hijastro	SI	SI	NO	NO
	5	5	2120000	2	Masculino	3	Hijo hijastro	SI	SI	NO	NO
	12	2	2120004	2	Femenino	12	Hijo hijastro	SI	NO	NO	NO
	13	3	2120004	1	Femenino	9	Hijo hijastro	SI	NO	NO	NO
	14	4	2120004	1	Femenino	6	Hijo hijastro	SI	NO	NO	NO
	15	5	2120004	1	Masculino	5	Hijo hijastro	SI	NO	NO	NO
	16	6	2120004	1	Masculino	1	Hijo hijastro	SI	NO	NO	NO
	19	3	2120005	2	Femenino	1	Hijo hijastro	SI	SI	NO	NO
	20	4	2120005	2	Masculino	0	Hijo hijastro	SI	SI	NO	NO
	23	3	2120006	2	Femenino	7	Hijo hijastro	SI	SI	NO	NO

Integrantes_hogar 9 x

Apply Revert

Figure 8: Integrantes hogar

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

```
24 • CREATE TABLE IF NOT EXISTS `Integrantes_hogar` (  
25     `Id` INT NOT NULL AUTO_INCREMENT,  
26     `Orden` VARCHAR(10) NULL,  
27     `Id_Hogar` INT(7) NULL,  
28     `orden_persona_informacion` INT(2) NULL,  
29     `Sexo_nacimiento` VARCHAR(15) NULL,  
30     `Edad` INT(3) NULL,  
31     `Parentesco_jefe_hogar` VARCHAR(45) NULL,  
32     `madre_reside_hogar` VARCHAR(15) NULL,  
33     `padre_reside_hogar` VARCHAR(15) NULL,  
34     `consideracion_campesino` VARCHAR(15) NULL,  
35     `consideracion_campesino_alguna_vez` VARCHAR(15) NULL,  
36     `comunidad_campesina` VARCHAR(15) NULL,  
37     `reconocimiento_etnia` VARCHAR(45) NULL,  
38     `Estado_civil` VARCHAR(15) NULL,  
39     `conyugue_reside_hogar` VARCHAR(2) NULL,  
40     PRIMARY KEY (`Id`))  
41 ENGINE = InnoDB;  
44 -----  
45 -- Table `Caracteristicas_Hogar`  
46 -----  
47 • CREATE TABLE IF NOT EXISTS `Caracteristicas_Hogar` (  
48     `Id` INT(15) NOT NULL AUTO_INCREMENT,  
49     `Id_Hogar` INT(7) NULL,  
50     `Orden` VARCHAR(45) NULL,  
51     `Titulo_vivienda` VARCHAR(45) NULL,  
52     `personas_hogar` INT(3) NULL,  
53     `Tipo_vivienda` VARCHAR(45) NULL,  
54     `energía_eléctrica` VARCHAR(2) NULL,  
55     `Estrato` INT(2) NULL,  
56     `Gas_natural` VARCHAR(2) NULL,  
57     `Alcantarillado` VARCHAR(2) NULL,  
58     `Recoleccion_basuras` VARCHAR(2) NULL,  
59     `Acueducto` VARCHAR(2) NULL,  
60     `Hogares_vivienda` INT(2) NULL,  
61     PRIMARY KEY (`Id`))  
62 ENGINE = InnoDB;
```

```

65 -----
66 -- Table `Seguridad_social`
67 -----
68 ● CREATE TABLE IF NOT EXISTS `Seguridad_social` (
69     `Id` INT NOT NULL AUTO_INCREMENT,
70     `Id_Hogar` INT(7) NULL,
71     `Orden` INT(3) NULL,
72     `seguridad_social` VARCHAR(45) NULL,
73     `Regimen` VARCHAR(45) NULL,
74     PRIMARY KEY (`Id`))
75 ENGINE = InnoDB;
76
77
78 -----
79 -- Table `Hogar`
80 -----
81 ● CREATE TABLE IF NOT EXISTS `Hogar` (
82     `Id` INT NOT NULL AUTO_INCREMENT,
83     `Id_Hogar` INT(7) NULL,
84     `Fecha_actualizacion` date NULL,
85     PRIMARY KEY (`Id`))
86 ENGINE = InnoDB;
87
88 -----
89
90 -----
91 -- Table `Educación_Integrantes_hogar`
92 -----
93 ● CREATE TABLE IF NOT EXISTS `Educación_Integrantes_hogar` (
94     `Id` INT(15) NOT NULL AUTO_INCREMENT,
95     `Id_Hogar` INT(7) NULL,
96     `Orden` INT(2) NULL,
97     `lee_escribe` VARCHAR(2) NULL,
98     `asiste_institucion_educativa` VARCHAR(2) NULL,
99     `Tipo_institucion` VARCHAR(45) NULL,
100     `max_nivel_educativo` VARCHAR(45) NULL,
101     `max_diploma` VARCHAR(45) NULL,
102     PRIMARY KEY (`Id`))
103 ENGINE = InnoDB;

```

```

115  -----
116  -- Table `Ocupación`
117  -----
118  ● CREATE TABLE IF NOT EXISTS `Ocupación` (
119      `Id` INT NOT NULL,
120      `Id_Hogar` INT(7) NULL,
121      `Orden` INT(2) NULL,
122      `Ocupado` VARCHAR(2) NULL,
123      `Ocupacion` VARCHAR(25) NULL,
124      `Meses_trabajo` INT(15) NULL,
125      `Último_Salario` INT(25) NULL,
126      PRIMARY KEY (`Id`))
127  ENGINE = InnoDB;
128  -----
129  -- Table `Otros_Ingresos`
130  -----
131  ● CREATE TABLE IF NOT EXISTS `Otros_Ingresos` (
132      `Id` INT NOT NULL,
133      `Id_Hogar` INT(7) NULL,
134      `Orden` INT NOT NULL,
135      `ingresos_arriendo` VARCHAR(2) NULL,
136      `propietario_varios_inmuebles` VARCHAR(2) NULL,
137      `ayudas_gobierno` VARCHAR(2) NULL,
138      PRIMARY KEY (`Id`))
139  ENGINE = InnoDB;

```

4.5 Código SQL - Manipulación de datos (DML)

```
1 • SELECT hogar.Id_Hogar, integrantes_hogar.Orden, integrantes_hogar.Sexo_nacimiento, integrantes_hogar.Edad,
2 educación_integrantes_hogar.lee_escribe FROM encuesta_hogares.hogar, encuesta_hogares.integrantes_hogar,
3 encuesta_hogares.educación_integrantes_hogar
4 where integrantes_hogar.Id_Hogar=hogar.Id_Hogar and hogar.Id_Hogar=educación_integrantes_hogar.Id_Hogar
5 and educación_integrantes_hogar.lee_escribe ='SI';
```

Id_Hogar	Orden	Sexo_nacimiento	Edad	lee_escribe
2120000	3	Masculino	7	SI
2120000	3	Masculino	7	SI
2120000	4	Masculino	6	SI
2120000	4	Masculino	6	SI
2120000	5	Masculino	3	SI
2120000	5	Masculino	3	SI
2120004	2	Femenino	12	SI
2120004	2	Femenino	12	SI

Figure 9: Sentencia 1

```
20 • select Id_Hogar, sum(último_Salario) from encuesta_hogares.ocupación
21 group by Id_Hogar ;
22
23
24
25
```

Id_Hogar	sum(último_Salario)
2120055	900000
2120057	908526
2120058	600000
2120059	1300000
2120060	910000
2120064	910000
2120065	1020000

Figure 10: Sentencia 2

```

23 • select Titulo_vivienda, sum(personas_hogar) from encuesta_hogares.caracteristicas_hogar
24 group by Titulo_vivienda;
25

```

Titulo_vivienda	sum(personas_hogar)
En arriendo o subarriendo	2038
En usufructo	2558
Posesion sin titulo	1597
Propia	2573
Propiedad colectiva	195

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

Result Grid
Form Editor

Figure 11: Sentencia 3

```

26 • select count(distinct Id_Hogar) from encuesta_hogares.integrantes_hogar where Edad<12;
27

```

count(distinct Id_Hogar)
6948

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

Result Grid
Form Editor

Result 29 x Read Only

Figure 12: Sentencia 4

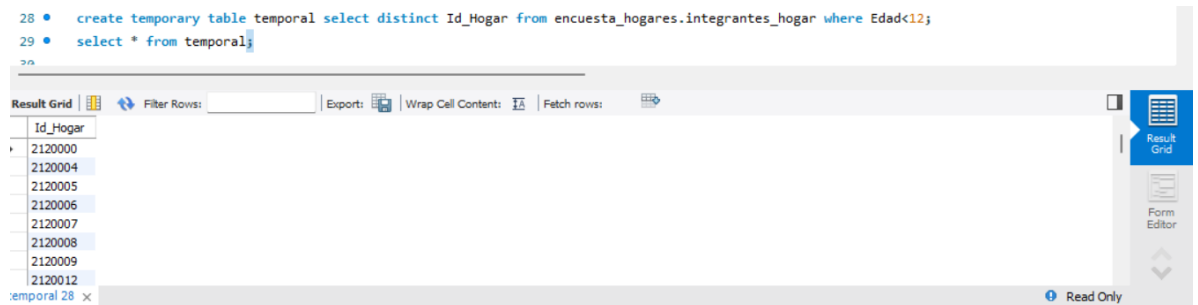


Figure 13: Sentencia 5

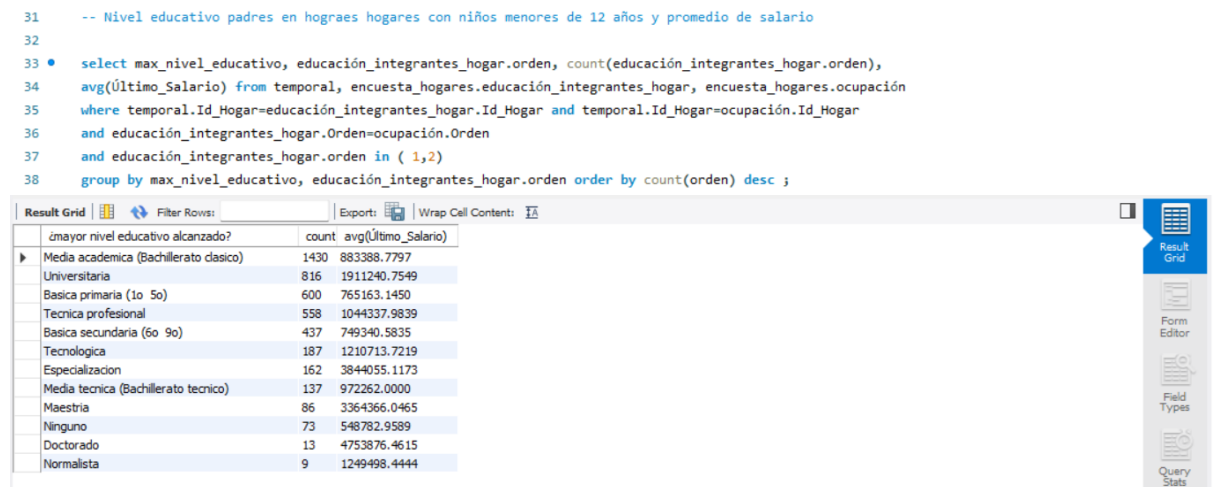


Figure 14: Sentencia 6

4.6 Código SQL + Resultados: Vistas

```
7 • CREATE VIEW ESCRITURA AS
8   SELECT count(*) FROM encuesta_hogares.hogar, encuesta_hogares.integrantes_hogar,
9   encuesta_hogares.educación_integrantes_hogar
10  where integrantes_hogar.Id_Hogar=hogar.Id_Hogar and hogar.Id_Hogar=educación_integrantes_hogar.Id_Hogar
11  and educación_integrantes_hogar.lee_escribe='NO' and integrantes_hogar.Edad>5;
12
13 • SELECT * FROM ESCRITURA;
```

Result Grid

count(*)
18

Figure 15: Vista 1

4.7 Código SQL + Resultados: Triggers

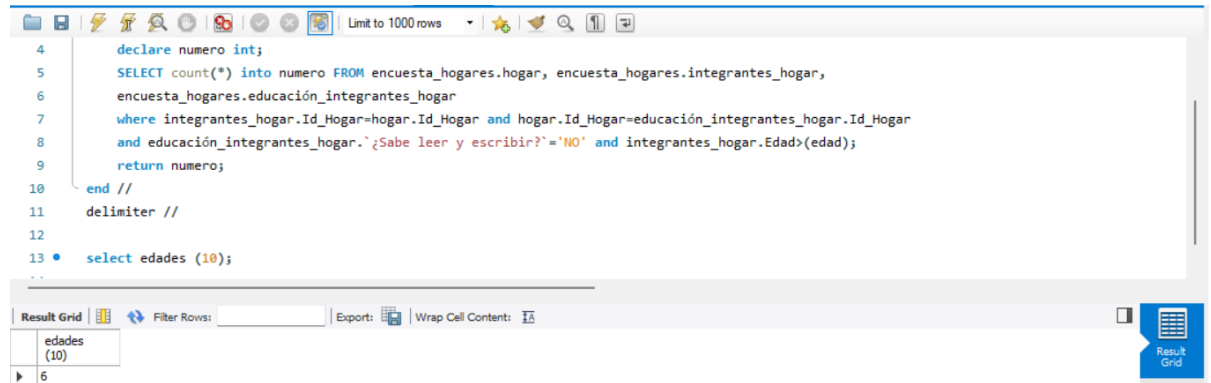
```
1 • CREATE trigger Ingreso_nuevo_hogar3 BEFORE insert on hogar FOR EACH ROW INSERT INTO nuevos_hogares (Id, Id_Hogar, Fecha_actualizacion)
2   values (new.Id, new.Id_Hogar, NOW());
18 • INSERT INTO encuesta_hogares.hogar (Id, Id_Hogar) values (25000, 3456783);
15 • select * FROM encuesta_hogares.nuevos_hogares;
```

Result Grid

Id	Id_Hogar	Fecha_actualizacion
25000	3456783	2022-10-21
NULL	NULL	NULL

Figure 16: Trigger 1

4.8 Código SQL + Resultados: Funciones



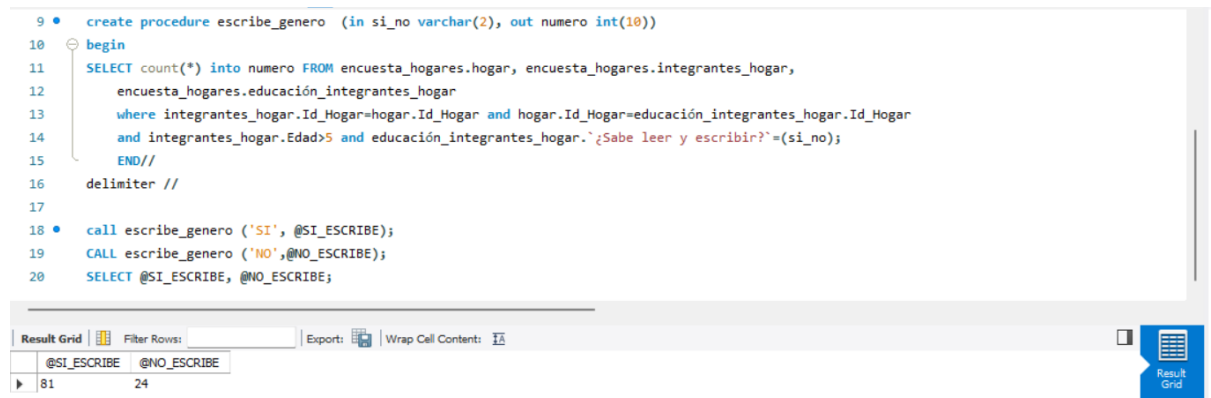
```
4 declare numero int;
5 SELECT count(*) into numero FROM encuesta_hogares.hogar, encuesta_hogares.integrantes_hogar,
6 encuesta_hogares.educación_integrantes_hogar
7 where integrantes_hogar.Id_Hogar=hogar.Id_Hogar and hogar.Id_Hogar=educación_integrantes_hogar.Id_Hogar
8 and educación_integrantes_hogar.`¿Sabe leer y escribir?`='NO' and integrantes_hogar.Edad>(edad);
9 return numero;
10 end //
11 delimiter //
12
13 • select edades (10);
14
```

Result Grid

edades (10)
6

Figure 17: Función 1

4.9 Código SQL + Resultados: procedimientos almacenados



```
9 • create procedure escribe_genero (in si_no varchar(2), out numero int(10))
10 begin
11 SELECT count(*) into numero FROM encuesta_hogares.hogar, encuesta_hogares.integrantes_hogar,
12 encuesta_hogares.educación_integrantes_hogar
13 where integrantes_hogar.Id_Hogar=hogar.Id_Hogar and hogar.Id_Hogar=educación_integrantes_hogar.Id_Hogar
14 and integrantes_hogar.Edad>5 and educación_integrantes_hogar.`¿Sabe leer y escribir?`=(si_no);
15 END//
16 delimiter //
17
18 • call escribe_genero ('SI', @SI_ESCRIBE);
19 CALL escribe_genero ('NO',@NO_ESCRIBE);
20 SELECT @SI_ESCRIBE, @NO_ESCRIBE;
```

Result Grid

@SI_ESCRIBE	@NO_ESCRIBE
81	24

Figure 18: Procedimiento 1

5 Bases de Datos No-SQL

5.1 Diagrama Bases de Datos No-SQL

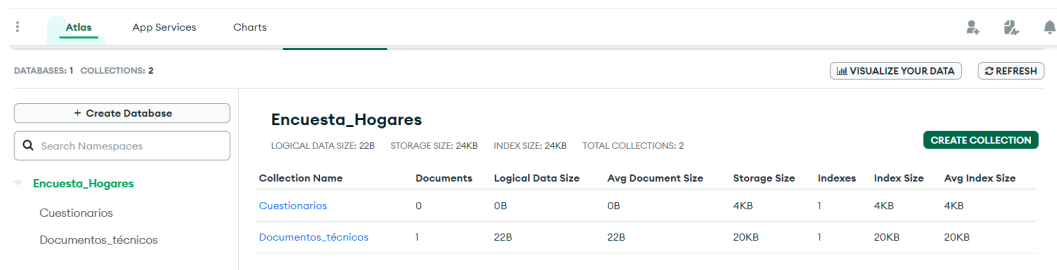


Figure 19: Base de datos y colecciones

Documents

Encuesta_Hogare...

+

Encuesta_Hogares.Cuestionarios

41

DOCUMENTSINDEXES

DocumentsAggregationsSchemaExplain PlanIndexesValidation

FILTER { field: 'value' }

OPTIONS

FIND

RESET

...

ADD DATA

VIEW

...

...

Displaying documents 1 - 4 of 4

1

_id: ObjectId('636c5020290a48dbf1a1b7')

Código: "PES-GEIH-MOT-003-r-001,"

Versión: 5

Año: 2019

Etapas: "1901-1902-1903 enero-marzo 2019,"

Nombre: "Cuestionario_GEIH_I TRIM_2019,"

ObjectId

String

Int64

Int64

String

String

CANCELUPDATE

1

_id: ObjectId('636c50d0290a48dbf1a1b8')

Código: "PES-GEIH-MOT-003-r-001,"

Versión: 6

Año: 2019

Etapas: "1904-1905-1906/Abril - junio 2019,"

Nombre: "Cuestionario_GEIH_II TRIM_2019,"

ObjectId

String

Int64

Int64

String

String

CANCELUPDATE

_id: ObjectId('636c5114290a48dbf1a1b9')

Código: "PES-GEIH-MOT-003-r-001"

Versión: 7

Año: 2019

Etapas: "1907 -1908 -1909 Julio - septiembre 2019"

Nombre: "Cuestionario_GEIH_III TRIM_2019"

ObjectId

String

Int64

Int64

String

String

CANCELUPDATE

1

_id: ObjectId('636c5114290a48dbf1a1b9')

Código: "PES-GEIH-MOT-003-r-001,"

Versión: 7

Año: 2019

Etapas: "1907 -1908 -1909 Julio - septiembre 2019,"

Nombre: "Cuestionario_GEIH_III TRIM_2019,"

ObjectId

String

Int64

Int64

String

String

CANCELUPDATE

1

_id: ObjectId('636c5161290a48dbf1a1ba')

Código: "PES-GEIH-MOT-003-r-001,"

Versión: 8

Año: 2019

Etapas: "1910 -1911 -1912 Octubre - Diciembre 2019,"

Nombre: "Cuestionario_GEIH_IV_TRIM_2019,"

ObjectId

String

Int64

Int64

String

String

CANCELUPDATE

Figure 20: Cuestionarios

21

#	Documentos_técnicos	_id ObjectId	Código String	Fecha Date	Nombre documento String	Proceso String	Versión Int64
1	ObjectId('636c419a298a48df1f...')	"DSO-GEHFE-001"	2016-04-08T00:00:00.000+00:00	"FICHA METODOLÓGICA GRAN ENCL...	"OPERACIÓN O INVESTIGACIÓN ES...	10	
2	ObjectId('636c4d79298a48df1f...')	"PES-GEIH-GU-001"	2017-11-01T00:00:00.000+00:00	"GUÍA DELIGENCIAMIENTO DEL FO...	"OPERACIÓN O INVESTIGACIÓN ES...	6	
3	ObjectId('636c4df2298a48df1f...')	"DSO-GEIH-EIN-002"	2016-09-26T00:00:00.000+00:00	"ESPECIFICACIONES DE LA ORTEN...	"OPERACIÓN O INVESTIGACIÓN ES...	1	
4	ObjectId('636c4e99298a48df1f...')	"DSO-GEIH-EIN-001"	2016-09-27T00:00:00.000+00:00	"ESPECIFICACIONES DE LA ORTEN...	"OPERACIÓN O INVESTIGACIÓN ES...	1	
5	ObjectId('636c4ecc298a48df1f...')	"PES-GEIH-MCC-001"	2017-11-10T00:00:00.000+00:00	"MANUAL DE COORDINACIÓN OPERA...	"OPERACIÓN O INVESTIGACIÓN ES...	8	
6	ObjectId('636c4fa298a48df1f...')	"PES-GEIH-MDI-001"	2017-11-16T00:00:00.000+00:00	"MANUAL DE DELIGENCIAMIENTO D...	"OPERACIÓN O INVESTIGACIÓN ES...	15	

Figure 21: Documentos técnicos

5.2 SMBD utilizado para la Base de Datos No-SQL

El sistema manejador de base de datos no relacionales utilizado es MongoDB, este sistema está orientado a documentos que son almacenados en formato BSON que es una representación binaria de JSON.

La arquitectura de MongoDB se basa en varios componentes principales. En primer lugar, id es un campo obligatorio para cada documento. Representa un valor único y puede considerarse como la clave principal del documento para identificarlo dentro de la colección.

Un documento es el equivalente a un registro en una base de datos tradicional. Se compone de campos de nombre y valor. Cada campo es una asociación entre un nombre y un valor y es similar a una columna en una base de datos relacional.

Una base de datos es un contenedor de colecciones. Cada uno tiene su propio conjunto de archivos en el sistema de archivos. Un servidor MongoDB puede almacenar múltiples bases de datos.

Ventajas:

Cualquier servidor o cualquier ordenador personal sirve para montar MongoDB y tener un servidor para tus proyectos.

Posee una documentación muy buena, muy amplia y detallada en comparación con otras bases de datos NoSQL.

Sintaxis amplia que permite obtener todo tipo de información pertinente sobre un tema en particular.

Soporta los principales lenguajes de programación: Python, Java y JavaScript.

Desventajas:

A pesar de estar ampliamente usada en la actualidad, sigue siendo una tecnología joven.

Poco fiable para aplicaciones con transacciones complejas.

6 Lecciones aprendidas

- Los sistemas de gestión de base de datos son necesarios y muy importantes en la creación y gestión de los datos de una organización. Permiten almacenar la información de manera organizada y permiten acceder a la información de forma ágil.
- El modelo relacional de bases de datos con sus relaciones normalizadas es una solución para satisfacer diversas condiciones de consulta y extracción de datos e información.

References

LAZARO (2019) Cómo crear y usar una función almacenada MySQL Obtenido de <https://www.neoguias.com/funciones-almacenadas-mysql/>: :text=Las

SANCHEZ 2022 Triggers, procedimientos y funciones en MySQL Obtenido de <https://josejuansanchez.org/bd/unidad-12-teoria/index.htmlfunciones>

DANE (2021) Gran Encuesta Integrada de Hogares Obtenido de http://microdatos.dane.gov.co/index.php/catalog/MICRODATOS/about_collection/23/1

DANE Consideraciones para el uso de la información Obtenido de <https://microdatos.dane.gov.co/home>