

Proyecto de Bases de datos para el análisis de la calidad de vida de los habitantes de Bogotá D.C.

Cristian David Carvajal Parra¹, Alejandro Salazar Molina²

¹Facultad de ingeniería y ciencias básicas

Universidad Central

Maestría en Analítica de Datos

Curso de Bases de Datos

Bogotá, Colombia

²ccarvajalp2@ucentral.edu.co, ³esalazarm@ucentral.edu.co

May 14, 2023

Contents

1	Introducción (Max 250 Palabras) - (<i>Primera entrega</i>)	3
2	Características del proyecto de investigación que hace uso de Bases de Datos (Max 500 Palabras) - (<i>Primera entrega</i>)	6
2.1	Titulo del proyecto de investigación (Max 100 Palabras) - (<i>Primera entrega</i>)	6
2.2	Objetivo general (Max 100 Palabras) - (<i>Primera entrega</i>)	7
2.2.1	Objetivos especificos (Max 100 Palabras) - (<i>Primera entrega</i>)	7
2.3	Alcance (Max 200 Palabras) - (<i>Primera entrega</i>)	7
2.4	Pregunta de investigación (Max 100 Palabras) - (<i>Primera entrega</i>) .	7
2.5	Hipotesis (Max 100 Palabras) - (<i>Primera entrega</i>)	7
3	Reflexiones sobre el origen de datos e información (Max 400 Palabras) - (<i>Primera entrega</i>)	9
3.1	¿Cual es el origen de los datos e información ? (Max 100 Palabras) - (<i>Primera entrega</i>)	9
3.2	¿Cuales son las consideraciones legales o eticas del uso de la información? (Max 100 Palabras) - (<i>Primera entrega</i>)	9
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? (Max 100 Palabras) - (<i>Primera entrega</i>)	10
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

4	Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)(Primera entrega)	12
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>)	17
4.5	Código SQL - Manipulación de datos (DML) (<i>Primera entrega</i>) . .	19
4.6	Código SQL + Resultados: Vistas (<i>Primera entrega</i>)	20
4.7	Código SQL + Resultados: Triggers (<i>Primera entrega</i>)	22
4.8	Código SQL + Resultados: Funciones (<i>Primera entrega</i>)	23
4.9	Código SQL + Resultados: procedimientos almacenados (<i>Primera entrega</i>)	26
5	Bases de Datos No-SQL (<i>Segunda entrega</i>)	28
5.1	Diagrama Bases de Datos No-SQL (<i>Segunda entrega</i>)	29
5.2	SMBD utilizado para la Base de Datos No-SQL (<i>Segunda entrega</i>)	31
6	Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos (<i>Tercera entrega</i>)	33
6.1	Ejemplo de aplicación de ETL y Bodega de Datos (<i>Tercera entrega</i>)	33
7	Proximos pasos (<i>Tercera entrega</i>)	34
8	Lecciones aprendidas (<i>Tercera entrega</i>)	35
9	Bibliografía	36

1 Introducción (Max 250 Palabras) - (*Primera entrega*)

La calidad de vida es un tema recurrente entre la sociedad, haciendo alusión a su nivel de bienestar en sus vivencias diarias, de esta manera una persona puede ver influenciado este indicador por condiciones externas como el estado de las vías de su barrio, o factores internos como la salud mental. Este tópico ha sido estudiado por diferentes organizaciones a nivel mundial, llevando a múltiples discusiones y consensos para llegar a puntos en común que puedan compartirse desde los niveles locales, regionales e internacionales (Naciones Unidas, 1961)

En algunos casos, expertos han tratado de verificar cuáles son las variables que influyen en la calidad de vida de las personas, llegando a hallazgos importantes como la relevancia de variables internas y externas. Entre las variables internas se pueden encontrar la edad, el estado civil, los ingresos, el nivel educativo y de endeudamiento, entre otros; por su parte, entre los factores externos se pueden encontrar: la discriminación, la infraestructura, la urbanización, el uso de medicamentos, el acceso a internet y muchos más (Huerta-Riveros, Patricia, Yáñez-Alvarado, Miguel, Leyton-Pavez, Carolina, 2022).

Lo anterior también coincide con organizaciones internacionales como la ONU, la cual en 1961 aseguraba la importancia de tener en cuenta factores como la salud, el consumo de alimentos, la educación, el empleo, las condiciones de trabajo, la vivienda, la seguridad social, el vestido, el esparcimiento y las libertades humanas, como variables significativas para medir la calidad de vida (Naciones Unidas, 1961).

Sin embargo, dependiendo del campo del conocimiento desde el que se abarque el concepto de calidad de vida se pueden tener en cuenta diferentes aspectos para su definición. En economía, pueden ser relevantes medidas como el PIB, los ingresos de una nación, el consumo de una persona y su capacidad para adquirir bienes (Hurtado Briceño, Alberto Zerpa de Hurtado, Sadcidi, 2011). Por otra parte, en el caso de la salud son protagonistas las dimensiones como la movilidad, el dolor, la movilidad física, el nivel de esfuerzo y la preocupación frente al estado de salud, asimismo, en el campo de la psicología son tenidos en cuenta la repercusión del funcionamiento cognitivo, uso de la memoria y la capacidad de pensar con claridad (Ruiz, Miguel Pardo, Antonio, 2005).

Por otra parte, a nivel mundial se han desarrollado estudios para analizar la Calidad de Vida, los cuales nacen después de los años 60, momento en el que la ONU decide motivar a los países a analizar la calidad de vida bajo 12 variables. Con el paso del tiempo, empiezan a surgir nuevos ejercicios como el “The quality of american life: Perceptions, Evaluations and Satisfaction” realizado por el Institute for Social Research de la Universidad de Michigan, en el que se evaluó la percepción de la calidad de vida en un barrio, preguntándole

a más de 2.000 personas sobre sus relaciones personales, su trabajo, educación y nivel de satisfacción con la vida en general (Campbell A, Converse P, Rodgers E., 1976).

Significativamente, conocer y cuestionar la calidad de vida se ha convertido en una necesidad mundial, por lo que entidades como la Organización para la Cooperación y el Desarrollo Económicos - OCDE han ejecutado 2 estudios How's life? con los que busca comparar la calidad de vida entre los países miembros. La organización ha ejecutado 2 reportes, uno en el 2011 y otro en el 2020, siendo este último relevante para Colombia. Llegando a hallazgos significativos como (OCDE, 2020):

- Casi el 40% de hogares de la OCDE son financieramente inseguros, lo que significa que pueden caer en pobreza si se abstienen de 3 meses de ingresos. Si bien el 12% de la población de la OCDE vive en una pobreza relativa de ingresos, la proporción de quienes informan tener dificultades para llegar a fin de mes en los países europeos de la OCDE es casi el doble, con un 21%.
- La mediana de la riqueza de los hogares disminuyó un 4, en promedio, desde alrededor de 2010, en aquellos países donde existen datos. Uno de cada cinco hogares de bajos ingresos gasta más del 40 de su ingreso disponible en costos de vivienda, dejando poco para otros elementos esenciales de la vida.
- Las experiencias de seguridad también contrastan fuertemente entre mujeres y hombres: por un lado, los hombres en los países de la OCDE tienen 4,5 veces más probabilidades de morir debido a una agresión, lo que refleja principalmente los altos valores observados en Colombia (donde los hombres tienen más de diez veces más probabilidades que mujeres a ser víctimas de homicidio) y México (donde la misma proporción es superior a ocho).
- Por otro lado, en promedio, ocho de cada diez hombres, pero solo seis de cada diez mujeres, informan que se sienten seguras cuando caminan solas por la noche, lo que posiblemente refleja el mayor riesgo de las mujeres de sufrir delitos de contacto y agresión sexual.
- La biodiversidad se ha perdido constantemente en muchos países de la OCDE (23) desde 2010. Las mayores disminuciones en el Índice de la Lista Roja para especies amenazadas generalmente han ocurrido en aquellos países con tasas de riesgo ya altas, incluidos Nueva Zelanda, México, Corea, Colombia, Chile, Reino Unido, Japón y Australia, además de Francia.

En el caso colombiano, entes gubernamentales como los Ministerios y otras entidades que buscan mejorar la calidad de vida de la población, necesitan contar con información veraz y actualizada de las condiciones de sus habitantes, siendo la fuente principal del análisis y toma de decisiones (DANE, 2021). Uno de los mecanismos principales para la obtención de esta información, es la aplicación de encuestas y entrevistas a la población objetivo, con las que se busca la participación de grupos poblacionales con diversas características o condiciones como tipos de territorios, razas, edades, etnias, principales ciudades e ideologías, entre otros (Cámara de Comercio de Medellín, 2021).

Teniendo en cuenta lo anterior, el Departamento Administrativo Nacional de Estadística – DANE, aplica las encuestas de calidad de vida anualmente “como una respuesta a la necesidad de caracterizar la población en los diferentes aspectos involucrados en el bienestar de los hogares”. Esta encuesta permite identificar en la población colombiana factores de pobreza desde dos enfoques: Necesidades Básicas Insatisfechas (NBI) y Línea de Pobreza, sin embargo, con el paso del tiempo se amplió el alcance de esta encuesta, incluyendo nuevas variables que permiten conocer más a profundidad la calidad de vida de los colombianos (DANE 2021).

En casos más específicos, grandes ciudades como Bogotá y Medellín, tienen condiciones que pueden afectar la calidad de vida como el tráfico, la inseguridad, la calidad del aire, y la falta de acceso a internet. En el 2021, se aplicó el Estudio de Calidad de Vida, desarrollado por la Pontificia Universidad Javeriana, la Cámara de Comercio de Bogotá, la Fundación Corona y el diario El Tiempo, en el que se analizaron 94 indicadores en factores en salud, empresa, pobreza, educación, ambiente, seguridad, movilidad y hábitat para conocer cómo están los bogotanos (Cámara de Comercio de Bogotá et al., 2021).

Este estudio dio a conocer grandes resultados como por ejemplo que en la capital colombiana mejoró el acceso a vivienda, basados en su calidad, en el 2020 el déficit denotaba un 12,7%, mientras que, en el 2021, la cifra se redujo llegando al 10,5%. Asimismo, el reporte informa que para el mismo año Bogotá mejoró económicamente llegando a un PIB del 9% y una reducción de la pobreza del 1,8%, a pesar estos valores positivos, el informe también señala que la ciudad es la más desigual del país (Cámara de Comercio de Bogotá et al., 2021).

Para el presente trabajo se usarán los datos de la Encuesta de Calidad de Vida del 2021, realizada por el DANE, empleando específicamente los registros obtenidos en la ciudad de Bogotá D.C. Con el fin de comparar los resultados obtenidos para conocer cuáles son las variables o factores que inciden en la calidad de vida de los habitantes de la capital colombiana, usando técnicas de machine learning que permitan este proceso. El machine learning es estratégico para este cometido porque su aplicabilidad permite el análisis de datos para la

toma de decisiones y para el estudio social de acontecimientos como la calidad de vida. Asimismo, este campo permite que los computadores aprendan por ellos mismos, cada vez teniendo un mejor desempeño, precisando sus habilidades para obtener resultados valiosos (Umaquina, Ana Suárez Zambrano, Luis Oña, Omar, 2018).

La aplicabilidad del machine learning es ideal para este proyecto porque permite analizar una cantidad de datos muy grande o pequeña, además, para su funcionamiento no se requiere una hardware de alto rendimiento. Asimismo, para su aplicabilidad divide las tareas en pequeñas actividades y luego combina los resultados para llegar a una conclusión. Por último, el tiempo de entrenamiento no es tan amplio, en comparación con otros métodos como el Deep Learning.

La calidad de vida es un tema recurrente entre la sociedad, haciendo alusión a su nivel de bienestar en sus vivencias diarias, de esta manera una persona puede ver influenciado este indicador por condiciones externas como el estado de las vías de su barrio, o factores internos como la salud mental.

Grandes ciudades como Bogotá, tienen condiciones que pueden afectar la calidad de vida como el tráfico, la inseguridad, la calidad del aire, y la falta de acceso a internet. De acuerdo con el Estudio de Calidad de Vida del 2021, desarrollado por la Pontificia Universidad Javeriana, la Cámara de Comercio de Bogotá, la Fundación Corona y el diario El Tiempo, en el que se analizaron 94 indicadores en factores en salud, empresa, pobreza, educación, ambiente, seguridad, movilidad y hábitat para conocer cómo están los bogotanos.

Este estudio dio a conocer grandes resultados como por ejemplo que en la capital colombiana mejoró el acceso a vivienda, basados en su calidad, en el 2020 el déficit denotaba un 12,7%, mientras que, en el 2021, la cifra se redujo llegando al 10,5%. Asimismo, el reporte informa que para el mismo año Bogotá mejoró económicamente llegando a un PIB del 9% y una reducción de la pobreza del 1,8%, a pesar estos valores positivos, el informe también señala que la ciudad es la más desigual del país.

2 Características del proyecto de investigación que hace uso de Bases de Datos (Max 500 Palabras) *- (Primera entrega)*

2.1 Título del proyecto de investigación (Max 100 Palabras) - *(Primera entrega)*

Analítica de datos aplicada para el mejoramiento de la calidad de vida de los habitantes de Bogotá D.C.

2.2 Objetivo general (Max 100 Palabras) - (*Primera entrega*)

Aplicar herramientas y procedimientos de análisis de datos que permitan identificar los puntos claves para mejorar el nivel de calidad de vida de los habitantes de Bogotá D.C.

2.2.1 Objetivos específicos (Max 100 Palabras) - (*Primera entrega*)

- Identificar el conjunto de factores más representativos que inciden en la calidad de vida de los habitantes de Bogotá D.C, mediante la aplicación de herramientas y procedimientos de análisis a los datos obtenidos por el DANE en el año 2021.
- Construir y emplear un modelo de analítica de datos que permita simular el adecuado manejo de diversos factores en búsqueda del mejoramiento de la calidad de vida de los habitantes de Bogotá D.C.
- Construir y emplear un modelo de analítica de datos que permita actualizar y predecir la calidad de vida de cada nuevo habitante de Bogotá D.C, y su implicación para los demás habitantes.

2.3 Alcance (Max 200 Palabras) - (*Primera entrega*)

Si bien la "Encuesta nacional de calidad de vida" realizada por el DANE en el año 2021, proporciona información relevante y diversa de las condiciones de la población colombiana a nivel nacional, el presente trabajo tiene como enfoque principal los habitantes de Bogotá D.C, los cuales habitan una de las principales ciudades y la capital del país que a su vez, es elegida como destino habitacional por algunas personas que provienen de otras ciudades y países vecinos en búsqueda de mejores oportunidades y calidad de vida.

Así mismo, el modelo que se pretende desarrollar en el presente trabajo tiene como objetivo desarrollar y aplicar un modelo que permita predecir la calidad de vida de los habitantes de Bogotá D.C. y a su vez, categorizar las condiciones y factores que influyen significativamente en la misma.

2.4 Pregunta de investigación (Max 100 Palabras) - (*Primera entrega*)

¿Cómo mejorar la calidad de vida de los habitantes de Bogotá D.C?

2.5 Hipotesis (Max 100 Palabras) - (*Primera entrega*)

Al abordar un estudio de calidad de vida en los habitantes de Bogotá D.C. permite plantear diferentes hipótesis que pueden ser analizadas en el desarrollo del presente trabajo, tales como:

- Los habitantes de Bogotá D.C. que no provienen de otras ciudades, tienen un alto nivel de calidad de vida.

- Los habitantes de Bogotá D.C. que tienen vivienda propia, tienen un alto nivel de calidad de vida.
- Los factores de ingreso y nivel educativo influyen la calidad de vida de los habitantes de Bogotá D.C.
- Los habitantes de Bogotá D.C. que tienen acceso al trabajo formal, tienen un alto nivel de calidad de vida.
- Los habitantes de Bogotá D.C. que presentan complicaciones de salud, tienen un bajo calidad de vida.

3 Reflexiones sobre el origen de datos e información

(Max 400 Palabras) - (*Primera entrega*)

A lo largo del tiempo, la calidad de vida de la población ha sido un tema de gran interés, el cuál ha tenido diversos aportes desde la perspectiva económica, política, social, psicológica, médica, entre otras; así mismo, se han implementado diferentes mecanismos en búsqueda de su constante mejoramiento.

En el caso Colombiano, los entes gubernamentales y la entidades que se encuentran dentro de su competencia, se encuentran constantemente en búsqueda del mejoramiento de la calidad de vida de su población, para lo que es necesario contar con información veraz y actualizada de las condiciones con las que cuentan sus habitantes, sirviendo de fuente principal del análisis y toma de decisiones en torno a los objetivos planteados.

Uno de los mecanismos principales para la obtención de esta información, es la elaboración y aplicación de encuestas a la población objetivo, en las cuales es importante asegurar su implementación en grupos poblacionales con diversas características o condiciones como tipos de territorios, razas, edades, etnias, principales ciudades, ideologías, entre otros.

3.1 ¿Cual es el origen de los datos e información ? (Max 100 Palabras) - (*Primera entrega*)

El Departamento Administrativo Nacional de Estadística – DANE, aplica las encuestas de calidad de vida anualmente “como una respuesta a la necesidad de caracterizar la población en los diferentes aspectos involucrados en el bienestar de los hogares”. Esta encuesta permite identificar en la población colombiana factores de pobreza desde dos enfoques Necesidades Básicas Insatisfechas (NBI) y Línea de Pobreza, sin embargo, con el paso del tiempo se amplió el alcance de esta encuesta, incluyendo nuevas variables que permiten conocer más a profundidad la calidad de vida de los colombianos.

De acuerdo con el DANE, el principal objetivo de la encuesta es “obtener información que permita analizar y realizar comparaciones de las condiciones socioeconómicas de los hogares que habitan en el país, las cuales posibiliten hacer seguimiento a las variables necesarias para el diseño e implementación de políticas públicas.” (DANE, 2021).

3.2 ¿Cuales son las consideraciones legales o éticas del uso de la información? (Max 100 Palabras) - (*Primera entrega*)

La información resultante de la “Encuesta Nacional de Calidad de Vida 2021” corresponde a datos abiertos que se encuentran publicados en la página web del DANE, por lo cual, su acceso es de libre consulta y su empleabilidad no puede incurrir en incumplimientos legales, partiendo del hecho que se encuentran

anonimizados con el principal objetivo de cumplir con la protección de datos personales contemplados en la ley de habeas data 1581 de 2012.

Así mismo, las características de la información orienta su uso explícitamente para la exploración y análisis de las condiciones de la calidad de vida de los colombianos, así como también a proporcionar el entorno real para la toma de decisiones e implementaciones de políticas públicas en beneficio de la sociedad.

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? (Max 100 Palabras) - (Primera entrega)

La encuesta está conformada por once capítulos que han sido aplicados de manera continua durante los últimos años. Además, existen capítulos especiales que se aplican periódicamente para satisfacer las necesidades específicas de los usuarios, los cuales son:

- Cada capítulo cuenta con una base de datos en formato .csv agrupada por un número de identificación del hogar encuestado.
- Cada hogar encuestado cuenta con un número de identificación de la persona encuestada.
- Se encuestaron 88.723 hogares a nivel nacional.
- Se encuestaron 257.589 habitantes a nivel nacional.
- El promedio de habitantes por hogar a nivel nacional es de 2,9.
- Se encuestaron 2.777 hogares en Bogotá.
- Se encuestaron 7.692 habitantes en Bogotá.
- El promedio de habitantes por hogar en Bogotá es de 2,7.
- La mayoría de las variables de la base de datos son categóricas.
- La base de datos cuenta con variables numéricas y categóricas.

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

Los principales propósitos del uso de bases de datos para la implementación del proyecto son:

- Contar con un repositorio en donde se encuentre la información almacenada y disponible para cuando se requiera procesar. Permitir realizar las consultas necesarias a la información cuando se requiera.

- Contar con la información organizada y clasificada de la manera en que se requiera para el análisis de los datos, construcción e implementación de los modelos.
- Hacer uso de una herramienta que permita la conexión con diferentes procesadores de código empleado para la construcción e implementación de los modelos.
- Determinar las relaciones entre los datos y las variables.
- Realizar operaciones entre las variables con el fin de reducir columnas para mejorar el procesamiento de los 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 este proyecto se decide emplear el Sistema Manejador de Bases de Datos MySQL porque es una herramienta de gestión de bases de datos muy completa y versátil que permite diseñar, modelar, generar y mantener bases de datos MySQL de una manera fácil y rápida.

Además, proporciona una interfaz gráfica de usuario (GUI) intuitiva que facilita el trabajo con bases de datos. Cuenta con múltiples funcionalidades que permiten manipular y visualizar los datos de forma eficiente, como la edición de tablas, la creación de consultas y la ejecución de scripts.

Es una herramienta multiplataforma, lo que significa que se utilizar en diferentes sistemas operativos, como Windows, macOS y Linux, lo que la hace muy accesible para diferentes tipos de usuarios.

Se destaca por ser gratuita y de código abierto, lo que significa que no se necesita invertir grandes cantidades de dinero para utilizarla y además se tiene la libertad de modificar y personalizar la herramienta según las necesidades del administrador.

Por último, cuenta con una amplia comunidad de usuarios que proporciona soporte técnico, documentación, tutoriales y recursos adicionales que pueden ayudar a resolver problemas.

4.2 Diagrama modelo de datos (Primera entrega)

Para este ejercicio se emplearon 2 tablas de la encuesta de Calidad de Vida del Dane del 2021, las cuales contienen información por habitante y hogar encuestado. Adicionalmente, se encuentran diferentes llaves auxiliares que identifican las variables categóricas.

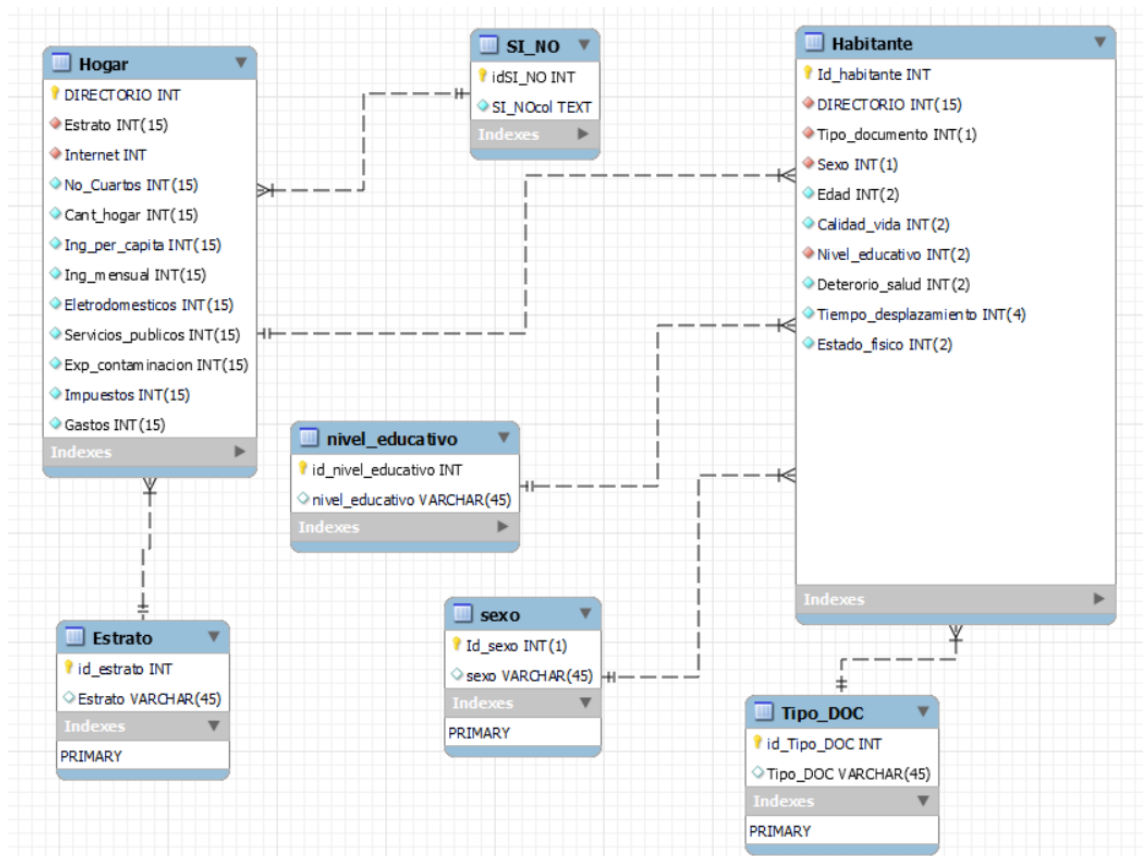


Figura 1

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

Se relacionan las visualizaciones de la base de datos la cual contiene 8 tablas; 3 vistas; 3 procedimientos almacenados y 4 funciones.

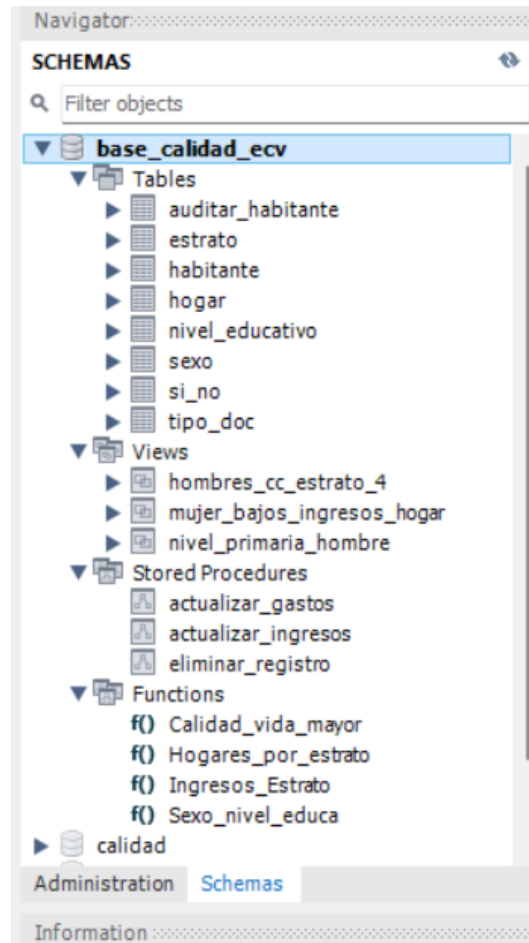


Figura 2: Base de datos ECV

```
1 • SELECT * FROM base_calidad_ecv.habitante;
```

Result Grid										
Filter Rows:										
Edit: Export/Import: Wrap Cell Content:										
	Id_habitante	DIRECTORIO	Tipo_documento	Sexo	Edad	Calidad_vida	Nivel_educativo	Deterioro_salud	Tiempo_desplazamiento	Estado_fisico
▶	1	1	3	1	30	7	1	66	157	32
	2	1	3	1	68	9	3	81	173	28
	3	2	3	2	66	9	3	53	173	28
	4	2	3	2	28	9	3	73	173	28
	5	3	2	1	12	9	3	0	173	28
	6	3	4	2	81	11	1	65	75	32
	7	4	3	1	38	13	1	66	230	32
	8	4	3	2	33	13	1	88	230	32
	9	5	1	2	2	13	1	0	230	32
	11	6	3	2	26	11	1	40	202	32
	12	6	1	1	6	11	1	0	202	32
	13	7	3	2	59	13	1	86	155	32
	14	7	3	2	55	5	3	35	226	30
	15	8	3	2	24	5	3	69	226	30
	16	8	2	1	17	5	3	65	226	30
	17	9	3	2	87	5	3	57	226	30
	18	9	3	1	48	5	3	72	226	30
	19	10	3	1	49	10	2	71	153	31
	20	10	3	2	46	10	2	77	153	31
	21	11	2	2	16	10	2	69	153	31
	22	11	2	2	16	10	2	68	153	31
	23	12	3	2	73	10	2	84	153	31
	24	12	3	2	69	3	3	66	90	29

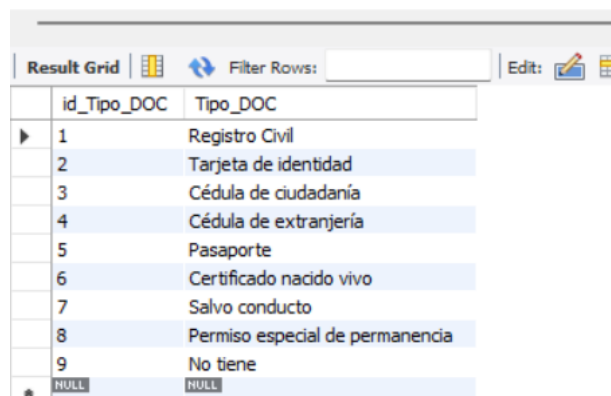
Figura 3

```
1 • SELECT * FROM base_calidad_ecv.hogar;
```

Result Grid												
Filter Rows:					Edit:			Export/Import:		Wrap Cell Content:		
DIRECTORIO	Estrato	Internet	No_Cuartos	Cant_hogar	Ing_per_capita	Ing_mensual	Eletrdomesticos	Servicios_publicos	Exp_contaminacion	Impuestos	Gastos	
1	3	1	3	1	925192	850000	33	4	1	0	850000	
2	3	1	4	4	1508333	6033333	24	5	9	1923098	835000	
3	3	1	4	4	1508333	6033333	24	5	9	1923098	835000	
4	3	1	4	4	1508333	6033333	24	5	9	1923098	850000	
5	3	1	4	4	1508333	6033333	24	5	9	1923098	835000	
6	6	1	3	1	5500000	5500000	27	5	0	9500000	1260000	
7	6	1	4	3	7767361	23302083	27	5	1	0	4164225	
8	6	1	4	3	7767361	23302083	27	5	1	0	4164225	
9	6	1	4	3	7767361	23302083	27	5	1	0	4164225	
10	3	1	2	3	906666	2720000	35	4	3	0	640099	
11	3	1	2	3	906666	2720000	35	4	3	0	640099	
12	3	1	2	3	906666	2720000	35	4	3	0	640099	
13	6	1	3	1	18000000	18000000	32	5	0	0	3850098	
14	3	1	3	5	211266	1056333	32	5	4	0	1402000	
15	3	1	3	5	211266	1056333	32	5	4	0	1402000	
16	3	1	3	5	211266	1056333	32	5	4	0	1402000	
17	3	1	3	5	211266	1056333	32	5	4	0	1402000	
18	3	1	3	5	211266	1056333	32	5	4	0	1402000	
19	3	1	3	5	840000	4200000	29	5	4	4000099	430400	
20	3	1	3	5	840000	4200000	29	5	4	4000099	430400	
21	3	1	3	5	840000	4200000	29	5	4	4000099	430400	
22	3	1	3	5	840000	4200000	29	5	4	4000099	430400	

Figura 4

```
1 • SELECT * FROM base_calidad_ecv.tipo_doc;
```



The screenshot shows a database query result grid. The grid has two columns: 'id_Tipo_DOC' and 'Tipo_DOC'. The rows are numbered 1 through 9, and there is a final row with 'NULL' values. The grid is titled 'Result Grid' and includes a 'Filter Rows' button and an 'Edit' button.

	id_Tipo_DOC	Tipo_DOC
▶	1	Registro Civil
	2	Tarjeta de identidad
	3	Cédula de ciudadanía
	4	Cédula de extranjería
	5	Pasaporte
	6	Certificado nacido vivo
	7	Salvo conducto
	8	Permiso especial de permanencia
	9	No tiene
▲	NULL	NULL

Figura 5

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

Se comparte el código con el que se crearon las tablas, la asignación de las llaves primarias, el tipo de datos que irá por columna. Los datos se importaron directamente sobre cada tabla desde un archivo CSV.

```
CREATE TABLE IF NOT EXISTS `base_calidad_ecv`.`Estrato` (  
  `id_estrato` INT NOT NULL,  
  `Estrato` VARCHAR(45) NULL,  
  PRIMARY KEY (`id_estrato`))  
ENGINE = InnoDB  
  
CREATE TABLE IF NOT EXISTS `base_calidad_ecv`.`Habitante` (  
  `Id_habitante` INT NOT NULL AUTOINCREMENT,  
  `DIRECTORIO` INT(15) NOT NULL,  
  `Tipo_documento` INT(1) NOT NULL,  
  `Sexo` INT(1) NOT NULL,  
  `Edad` INT(2) NOT NULL,  
  `Calidad_vida` INT(2) NOT NULL,  
  `Nivel_educativo` INT(2) NOT NULL,  
  `Deterioro_salud` INT(2) NOT NULL,  
  `Tiempo_desplazamiento` INT(4) NOT NULL,  
  `Estado_fisico` INT(2) NOT NULL,  
  INDEX `Id_sexo_idx` (`Sexo` ASC) VISIBLE,  
  INDEX `Id_nivel_educativo_idx` (`Nivel_educativo` ASC) VISIBLE,  
  INDEX `id_Tipo_DOC_idx` (`Tipo_documento` ASC) VISIBLE,  
  PRIMARY KEY (`Id_habitante`),  
  INDEX `Directorio_idx` (`DIRECTORIO` ASC) VISIBLE,  
  CONSTRAINT `Id_sexo`  
    FOREIGN KEY (`Sexo`)  
    REFERENCES `base_calidad_ecv`.`sexo` (`Id_sexo`)  
    ON DELETE NO ACTION  
    ON UPDATE NO ACTION,  
  CONSTRAINT `Id_nivel_educativo`  
    FOREIGN KEY (`Nivel_educativo`)  
    REFERENCES `base_calidad_ecv`.`nivel_educativo` (`id_nivel_educativo`)  
    ON DELETE NO ACTION  
    ON UPDATE NO ACTION,  
  CONSTRAINT `id_Tipo_DOC`  
    FOREIGN KEY (`Tipo_documento`)  
    REFERENCES `base_calidad_ecv`.`Tipo_DOC` (`id_Tipo_DOC`)  
    ON DELETE NO ACTION  
    ON UPDATE NO ACTION,  
  CONSTRAINT `Directorio`  
    FOREIGN KEY (`DIRECTORIO`)  
    REFERENCES `base_calidad_ecv`.`Hogar` (`DIRECTORIO`)
```

```

        ON DELETE NO ACTION
        ON UPDATE NO ACTION)
ENGINE = InnoDB

```

```

CREATE TABLE IF NOT EXISTS `base_calidad_ecv`.`Hogar` (
  `DIRECTORIO` INT NOT NULL,
  `Estrato` INT(15) NOT NULL,
  `Internet` INT NOT NULL,
  `No_Cuartos` INT(15) NOT NULL,
  `Cant_hogar` INT(15) NOT NULL,
  `Ing_per_capita` INT(15) NOT NULL,
  `Ing_mensual` INT(15) NOT NULL,
  `Eletrodomesticos` INT(15) NOT NULL,
  `Servicios_publicos` INT(15) NOT NULL,
  `Exp_contaminacion` INT(15) NOT NULL,
  `Impuestos` INT(15) NOT NULL,
  `Gastos` INT(15) NOT NULL,
  PRIMARY KEY (`DIRECTORIO`),
  INDEX `id_estrato_idx` (`Estrato` ASC) VISIBLE,
  INDEX `idSI_NO_idx` (`Internet` ASC) VISIBLE,
  CONSTRAINT `id_estrato`
    FOREIGN KEY (`Estrato`)
    REFERENCES `base_calidad_ecv`.`Estrato` (`id_estrato`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION,
  CONSTRAINT `idSI_NO`
    FOREIGN KEY (`Internet`)
    REFERENCES `base_calidad_ecv`.`SI_NO` (`idSI_NO`)
    ON DELETE NO ACTION
    ON UPDATE NO ACTION)
ENGINE = InnoDB

```

```

CREATE TABLE IF NOT EXISTS `base_calidad_ecv`.`nivel_educativo` (
  `id_nivel_educativo` INT NOT NULL,
  `nivel_educativo` VARCHAR(45) NULL,
  PRIMARY KEY (`id_nivel_educativo`))
ENGINE = InnoDB

```

```

CREATE TABLE IF NOT EXISTS `base_calidad_ecv`.`sexo` (
  `Id_sexo` INT(1) NOT NULL,
  `sexo` VARCHAR(45) NULL,
  PRIMARY KEY (`Id_sexo`))
ENGINE = InnoDB

```

```

CREATE TABLE IF NOT EXISTS `base_calidad_ecv`.`sexo` (
  `Id_sexo` INT(1) NOT NULL,

```

```

        'sexo' VARCHAR(45) NULL,
        PRIMARY KEY ('Id_sexo'))
ENGINE = InnoDB

```

```

CREATE TABLE IF NOT EXISTS 'base_calidad_ecv' (
    'id_Tipo_DOC' INT NOT NULL,
    'Tipo_DOC' VARCHAR(45) NULL,
    PRIMARY KEY ('id_Tipo_DOC'))
ENGINE = InnoDB

```

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

Se hace uso de los comandos SELECT, para indicar las columnas a las que se añadirá un registro a través de la función INSERT TO.

The screenshot displays a database management interface. On the left, the 'SCHEMAS' pane shows a tree view of the database structure. Under the 'base_calidad_ecv' schema, several tables are listed: 'auditar_habitante', 'estrato', 'habitante', 'hogar', 'nivel_educativo', 'sexo', 'si_no', and 'tipo_doc'. The 'tipo_doc' table is selected, and its details are shown at the bottom: 'id_Tipo_DOC' is an integer primary key, and 'Tipo_DOC' is a varchar(45). On the right, the 'Codigo SQL ECV' pane shows a series of SQL commands for inserting data into the 'base_calidad_ecv' database. The commands are numbered 1 through 27. They include 'USE base_calidad_ecv;', 'SELECT * FROM base_calidad_ecv.sexo;', 'INSERT INTO sexo (Id_sexo,sexo) VALUES ('1','hombre'), ('2','mujer');', 'SELECT * FROM base_calidad_ecv.si_no;', 'INSERT INTO si_no (idSI_NO,SI_NOcol) VALUES ('1','si'), ('2','no');', 'SELECT * FROM base_calidad_ecv.estrato;', 'INSERT INTO estrato (id_estrato,Estrato) VALUES ('0','Pirata'), ('1','Muy Bajo'), ('2','Bajo'), ('3','Medio Bajo'), ('4','Medio'), ('5','Medio Alto'), ('6','Alto');', 'SELECT * FROM base_calidad_ecv.estrato;', 'INSERT INTO estrato (id_estrato,Estrato) VALUES ('8','Planta electrica'), ('9','No sabe');'.

Figura 6: Manipulación de datos

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

Se generaron las siguientes vistas, las cuales actúan como filtros que facilitan el acceso a la información.

The screenshot shows a database management interface. On the left, the 'SCHEMAS' pane lists various database objects. Under the 'Views' section, 'hombres_cc_estrato_4' is selected. The main SQL editor displays the query: `SELECT * FROM base_calidad_ecv.hombres_cc_estrato_4;`. Below the editor, the 'Result Grid' shows the following data:

	Tipo_documento	Estrato	sexo
▶	3	4	1
	3	4	1
	3	4	1
	3	4	1
	3	4	1
	3	4	1
	3	4	1
	3	4	1
	3	4	1
	3	4	1

Figura 7: se evidencian 3 vistas.

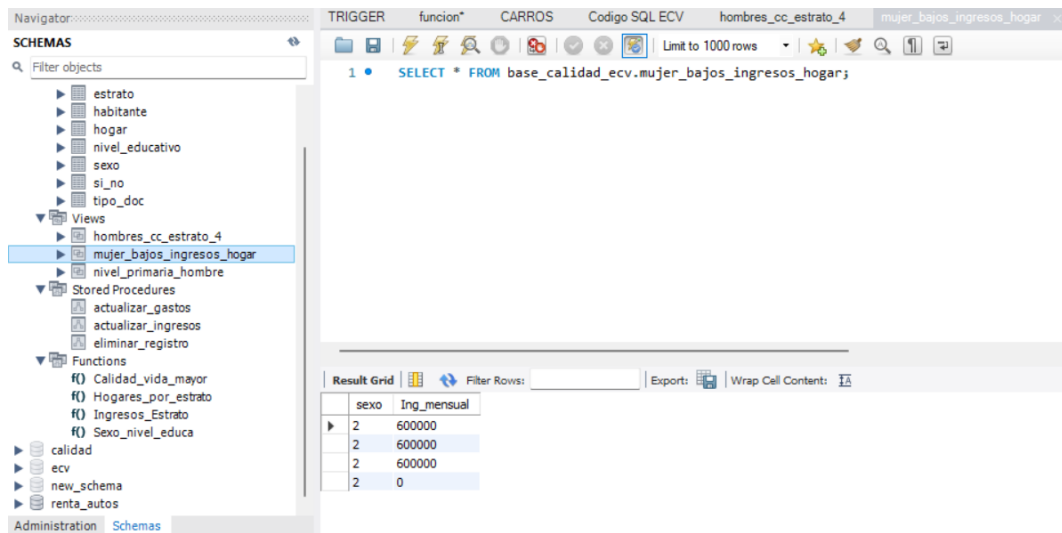


Figura 8: Vista 1

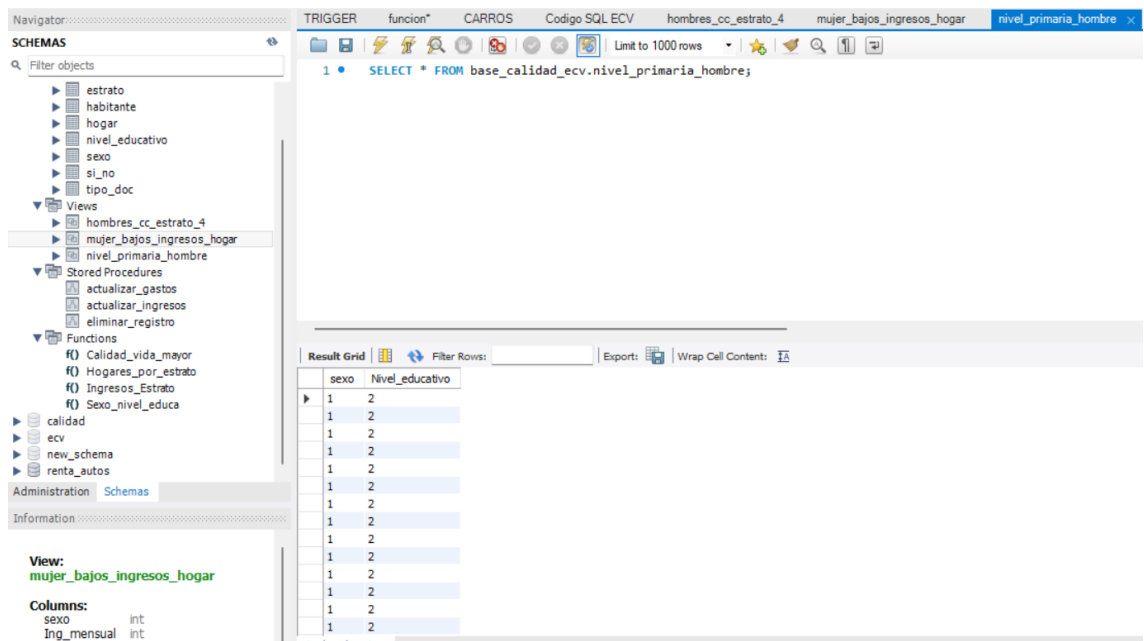


Figura 9: Vista 2

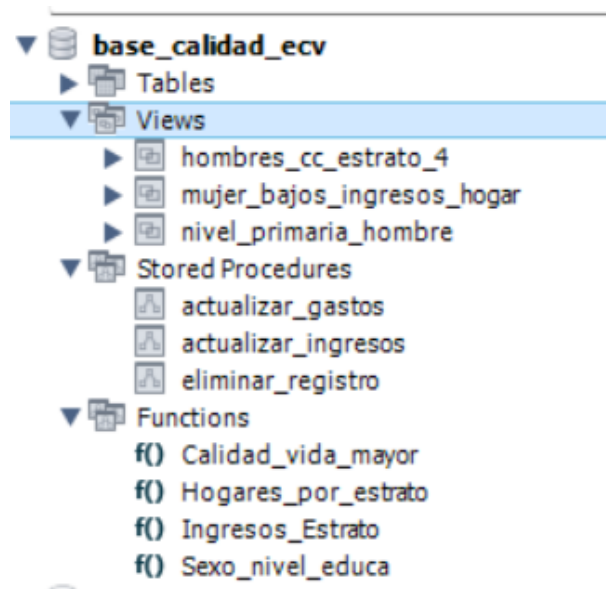


Figura 10: Vista 3

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

Diseño de trigger para auditar los datos que ingresan a la tabla habitante, los datos que se ingresen se almacenarán en una tabla aparte llamada auditar habitante, lo que permitirá analizar cómo se ha hecho la modificación de la tabla.

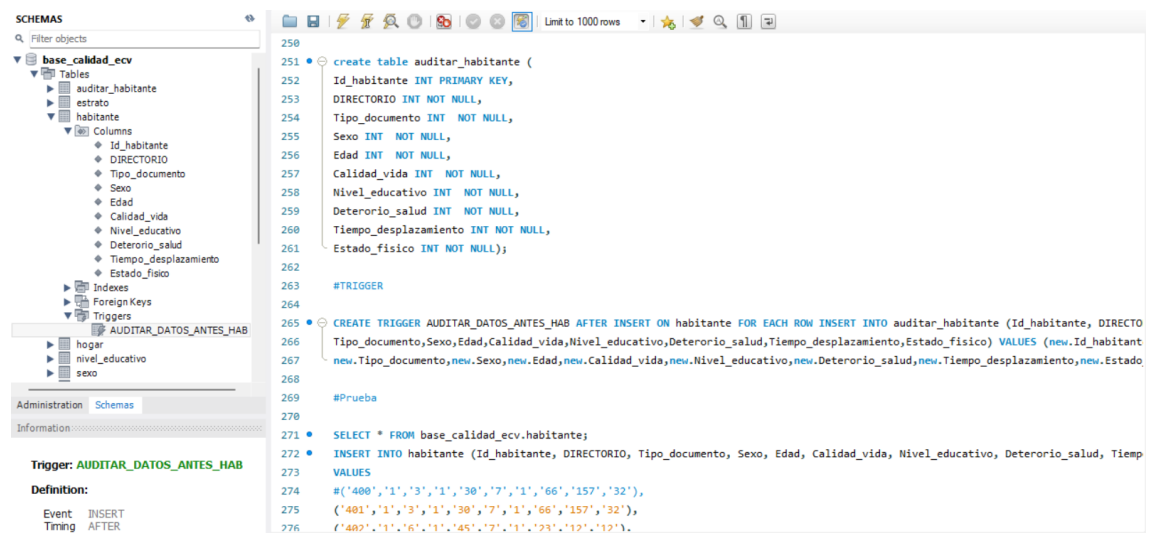


Figura 11: Trigger 1

Navigation pane (left):

- base_calidad_ecv
 - auditar_habitante (Table)
 - Columns: Id_habitante, DIRECTORIO, Tipo_documento, Sexo, Edad, Calidad_vida, Nivel_educativo, Deterioro_salud, Tiempo_desplazamiento, Estado_fisico
 - estrato (Table)
 - habitante (Table)
 - Columns: Id_habitante, DIRECTORIO, Tipo_documento, Sexo, Edad, Calidad_vida, Nivel_educativo, Deterioro_salud, Tiempo_desplazamiento, Estado_fisico
 - Indexes
 - Foreign Keys
 - Triggers
 - AUDITAR_DATOS_ANTES_HAB (Table)
 - hogar (Table)
 - nivel_educativo (Table)
 - sexo (Table)

SQL Editor (right):

```
SELECT * FROM base_calidad_ecv.auditar_habitante;
```

Result Grid (bottom right):

Id_habitante	DIRECTORIO	Tipo_documento	Sexo	Edad	Calidad_vida	Nivel_educativo	Deterioro_salud	Tiempo_desplazamiento	Estado_fisico
400	1	3	1	30	7	1	66	157	32
401	1	3	1	30	7	1	66	157	32
402	1	6	1	45	7	1	23	12	12
403	1	5	1	54	7	1	66	157	32
404	1	4	1	93	7	1	81	90	32
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Figura 12: Trigger 2

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

Functions:

- f() Calidad_vida_mayor
- f() Hogares_por_estrato
- f() Ingresos_Estrato
- f() Sexo_nivel_educa

Figura 13: Funciones 1


- #Permite consultar la cantidad de hogares por estrato ingresado
- `USE base_calidad_ecv;`
`delimiter //`
 - `CREATE FUNCTION Hogares_por_estrato (numero INT) RETURNS INT`
`DETERMINISTIC`
`BEGIN`

`DECLARE Hogares INT;`
`SELECT COUNT(*) INTO Hogares FROM hogar where Estrato = numero;`
`RETURN Hogares;`
`END; //`
`delimiter ;`
 - `SELECT Hogares_por_estrato(3);`

Figura 14: Funciones 2


- #Permite consultar la cantidad de habitantes con nivel de calidad de vida mayor al ingresado:
- `USE base_calidad_ecv;`
`delimiter //`
 - `CREATE FUNCTION Calidad_vida_mayor (numero INT) RETURNS INT`
`DETERMINISTIC`

`BEGIN`
`DECLARE Calidad INT;`
`SELECT COUNT(*) INTO Calidad FROM habitante where Calidad_vida > numero;`
`RETURN Calidad;`
`END; //`
`delimiter ;`
 - `SELECT Calidad_vida_mayor(6);`

Figura 15: Funciones 3


- #Permite calcular el promedio de ingresos mensuales por estrato:
- `USE base_calidad_ecv;`
`delimiter //`
 - `CREATE FUNCTION Ingresos_Estrato (numero INT) RETURNS INT`
`DETERMINISTIC`
`BEGIN`

`DECLARE Promedio INT;`
`SELECT AVG(Ing_mensual) INTO Promedio FROM hogar where Estrato = numero;`
`RETURN Promedio;`
`END; //`
`delimiter ;`
 - `SELECT Ingresos_Estrato(2);`

Figura 16: Funciones 4


- #Permite consultar la cantidad de habitantes por Sexo y Nivel educativo:
- `USE base_calidad_ecv;`
`delimiter //`
 - `CREATE FUNCTION Sexo_nivel_educa (Sexo_ INT, educa INT) RETURNS INT`
`DETERMINISTIC`
`BEGIN`

`DECLARE S_N INT;`
`SELECT COUNT(*) INTO S_N FROM habitante where Sexo = Sexo_ and Nivel_educativo = educa;`
`RETURN S_N;`
`END; //`
`delimiter ;`
 - `SELECT Sexo_nivel_educa(2,3);`

Figura 17: Funciones 5

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

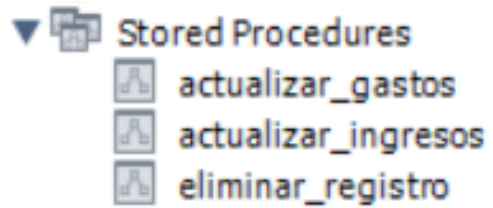


Figura 18: Procedimientos almacenados 1

```
#Permite actualizar el valor de los Gastos de un hogar en específico:

DELIMITER //
• CREATE PROCEDURE actualizar_gastos(IN id_hogar INT, IN gastos_act INT)
  BEGIN
    UPDATE hogar SET Gastos = gastos_act WHERE DIRECTORIO = id_hogar;
  END //
DELIMITER ;

• SELECT Gastos FROM hogar where DIRECTORIO = 4;
• CALL actualizar_gastos(4, 850000);
• SELECT Gastos FROM hogar where DIRECTORIO = 4;
```

Figura 19: Procedimientos almacenados 2

```
#Permite actualizar el valor de los Ingresos de un hogar en específico:
```

```
DELIMITER //
```

- `CREATE PROCEDURE actualizar_ingresos(IN id_hogar INT, IN ingreso_act INT)`

```
BEGIN
    UPDATE hogar SET Ing_mensual = ingreso_act WHERE DIRECTORIO = id_hogar;
END //
```

```
DELIMITER ;
```

- `SELECT Ing_mensual FROM hogar where DIRECTORIO = 1;`
- `CALL actualizar_ingresos(1, 850000);`
- `SELECT Ing_mensual FROM hogar where DIRECTORIO = 1;`

Figura 20: Procedimientos almacenados 3

```
#Permite eliminar una fila en específico de la tabla principal por Id de Habitante:
```

```
DELIMITER //
```

- `CREATE PROCEDURE eliminar_registro(IN id INT)`

```
BEGIN
    DELETE FROM habitante WHERE Id_habitante = id;
END //
```

```
DELIMITER ;
```

- `SELECT * FROM habitante where Id_habitante = 10;`
- `CALL eliminar_registro(10);`
- `SELECT * FROM habitante where Id_habitante = 10;`

Figura 21: Procedimientos almacenados 4

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

La base de datos NoSQL que se usó para este ejercicio fueron las estadísticas de las variables del DataFrame Encuesta Calidad de Vida 2021. Estas cifras ayudan a conocer las características de desviación estándar, coeficientes y P-Valor, entre otros. Estos datos ayudan a conocer si las variables son estadísticamente significativas o no.

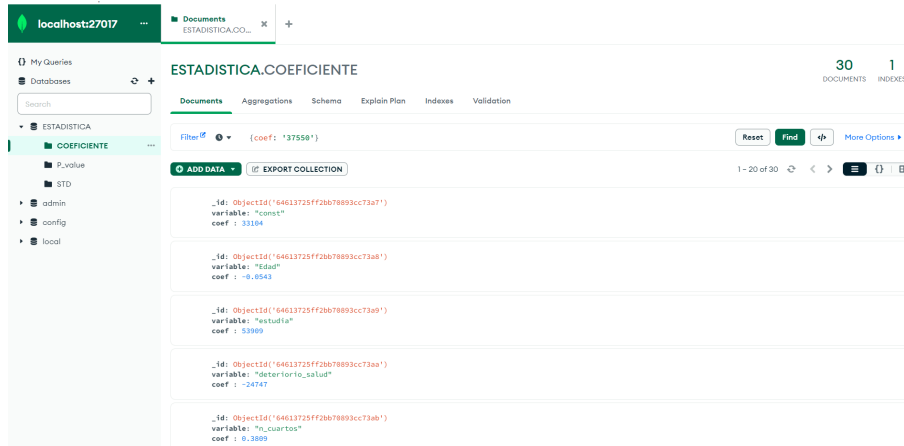


Figura 22: Base de datos NOSQL

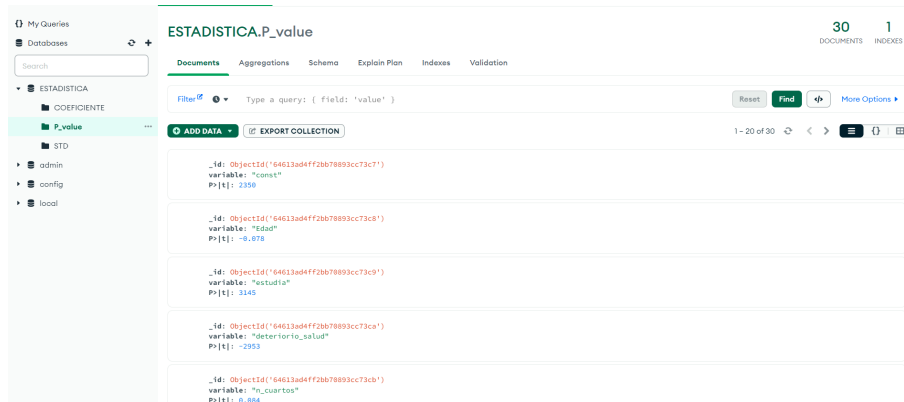


Figura 23: Base de datos NOSQL

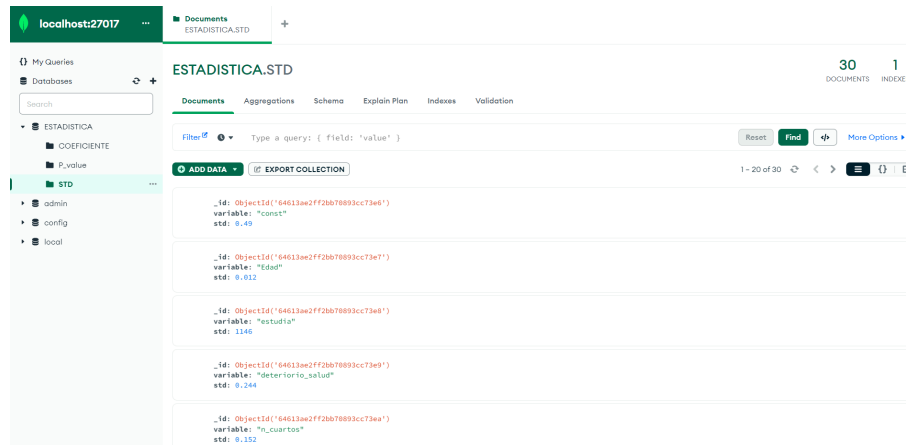


Figura 24: Base de datos NOSQL

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

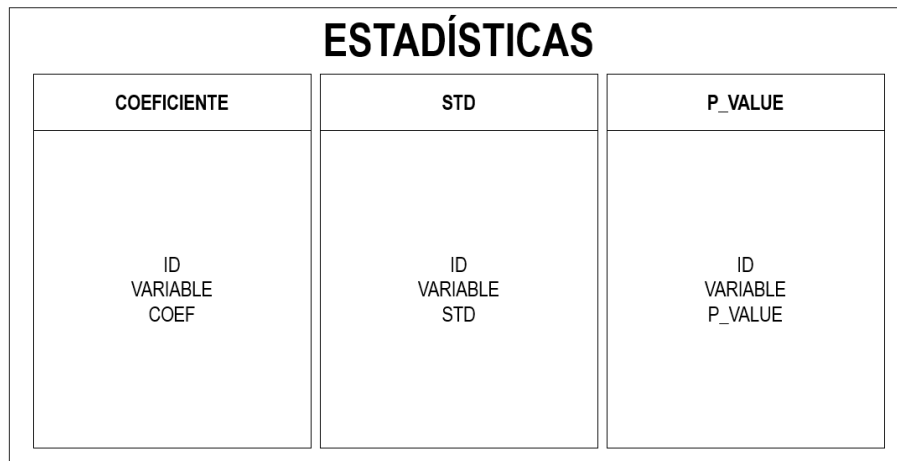


Figura 25: Diagrama base de datos NOSQL

En las bases de datos NoSQL, como MongoDB, no se utilizan diagramas de entidad-relación (ER) tradicionales como en las bases de datos relacionales. Esto se debe a que las bases de datos NoSQL, como MongoDB, tienen un enfoque de almacenamiento de datos diferente al modelo relacional.

En lugar de utilizar tablas, filas y columnas como en las bases de datos relacionales, las bases de datos NoSQL almacenan los datos en estructuras flexibles, como documentos o grafos. Esto significa que no hay una estructura de esquema rígida predefinida.

En MongoDB, los datos se almacenan en documentos BSON (JSON binario) dentro de colecciones. Cada documento puede tener su propio esquema y no hay una restricción estricta sobre la estructura de los documentos dentro de una colección.

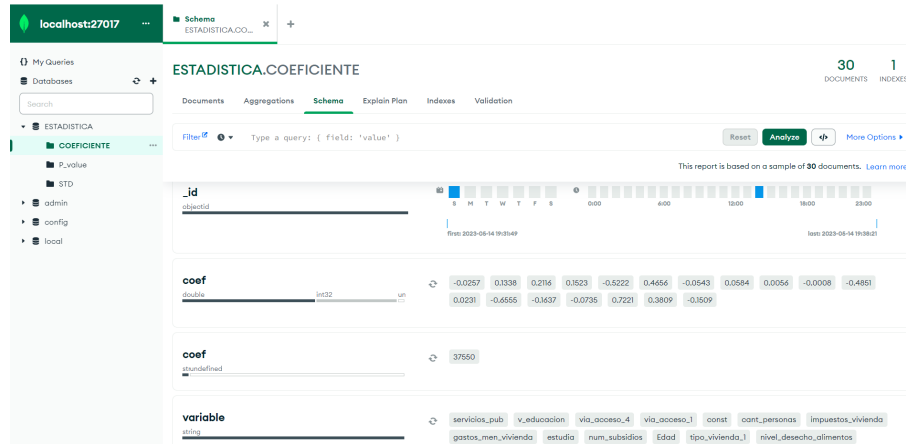


Figura 26: Esquema 1 - Coeficiente

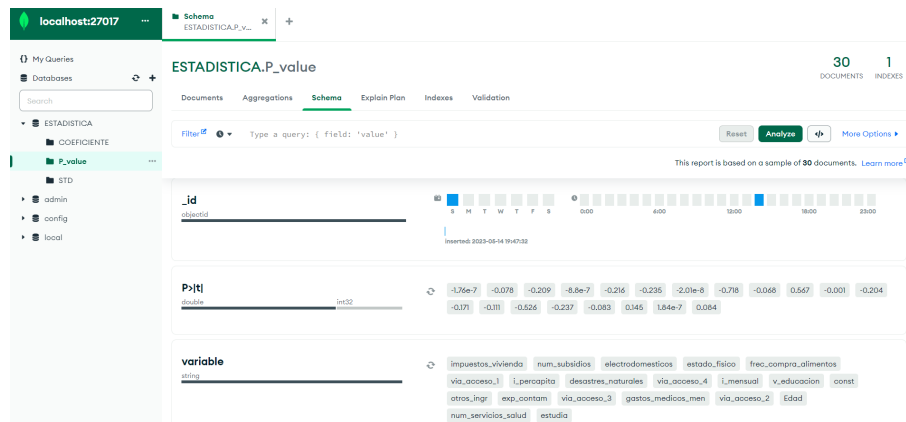


Figura 27: Esquema 2 - P Value

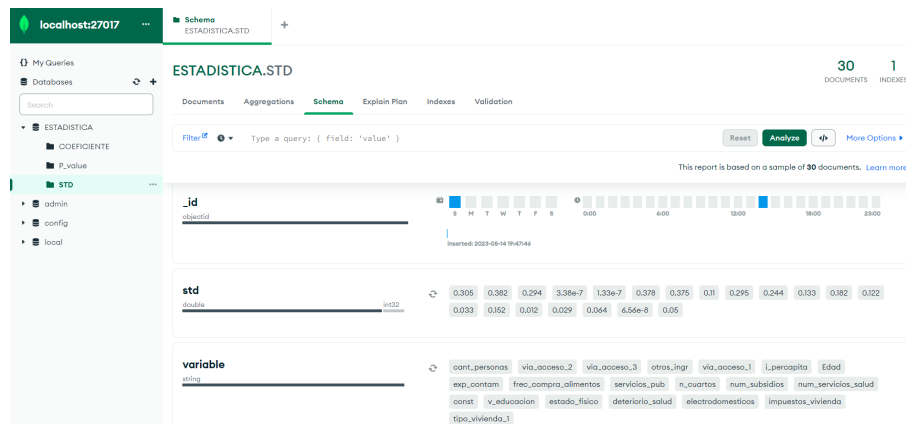


Figura 28: Esquema 3 - STD

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

MongoDB es una base de datos NoSQL, orientada a documentos, que se ha vuelto ampliamente popular debido a sus características distintivas. Estas son algunas de las características clave de MongoDB:

- **Modelo de datos flexible:** MongoDB utiliza un modelo de datos flexible y no impone una estructura rígida para los documentos almacenados. Puedes almacenar documentos JSON (BSON, en realidad) con diferentes estructuras dentro de una misma colección, lo que permite adaptar fácilmente el esquema de datos a medida que evoluciona la aplicación.
- **Escalabilidad horizontal:** MongoDB está diseñado para la escalabilidad horizontal, lo que significa que puede distribuir los datos a través de múltiples servidores o clústeres. Esto permite manejar grandes volúmenes de datos y alto rendimiento al escalar horizontalmente añadiendo más servidores.
- **Alta disponibilidad y tolerancia a fallos:** MongoDB es capaz de mantener una alta disponibilidad de datos mediante la replicación. Puedes configurar réplicas de datos en diferentes servidores, lo que garantiza que los datos estén disponibles incluso en caso de fallo de un servidor.
- **Consultas potentes:** MongoDB proporciona un lenguaje de consultas rico que permite realizar consultas flexibles y complejas. Admite consultas por campos, consultas geoespaciales, consultas de texto completo y más. Además, MongoDB incluye índices para mejorar el rendimiento de las consultas.
- **Operaciones de agregación:** MongoDB ofrece potentes operaciones de agregación que permiten realizar transformaciones y análisis de datos.

complejos en la base de datos. Puedes utilizar la agregación para realizar operaciones como agrupación, filtrado, proyección, cálculos y más.

- Integración con lenguajes de programación populares: MongoDB cuenta con controladores y bibliotecas para una amplia variedad de lenguajes de programación, lo que facilita la integración con diferentes aplicaciones y entornos de desarrollo.
- Soporte para almacenamiento de datos geoespaciales: MongoDB incluye características geoespaciales integradas que permiten almacenar y consultar datos basados en la ubicación geográfica. Esto hace que sea especialmente adecuado para aplicaciones que trabajan con información geoespacial, como sistemas de seguimiento de ubicación o análisis geográficos.

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 Proximos pasos (*Tercera entrega*)

8 Lecciones aprendidas (*Tercera entrega*)

9 Bibliografía

References

- [1] DANE., *Encuesta Nacional de Calidad de Vida 2018 [dataset]*, <https://microdatos.dane.gov.co/index.php> (2021).
- [2] Bogotá Cómo Vamos., *Informe de calidad de vida 2021*, <https://bogotacomovamos.org/> (2021).
- [3] Cámara de comercio de Bogotá, Fundación Corona, Pontificia Universidad Javeriana, El Tiempo., *Informe de calidad de vida en Bogotá, 2021: Una mirada al derecho a la ciudad*, <https://s3.documentcloud.org/documents/22123000/icv-2021-ii.pdf> (2022).
- [4] United Nations. *Definición y medición internacional del nivel de vida. Department of Economic and Social Affairs, United Nations.*, https://unstats.un.org/unsd/publication/seriee/seriee_cn3270_evs.pdf (1961).
- [5] Huerta-Riveros, Patricia, Yáñez-Alvarado, Miguel, Leyton-Pavez, Carolina., *Factores que impactan en la calidad de vida y su incidencia en las directrices sanitarias. Hacia la Promoción de la Salud*, <https://doi.org/10.17151/hpsal.2022.27.2.10> (2022).
- [6] Campbell A, Converse P, Rodgers E., *The Quality of american life: Perceptions, Evaluations and Satisfactions* (2022).
- [7] Hurtado Briceño, Alberto Zerpa de Hurtado, Sadcidi., *Perspectivas teóricas del estudio de la calidad de vida en economía*, (2011).
- [8] Ruiz, Miguel Pardo, Antonio., *Calidad de vida relacionada con la salud: Definición y utilización en la práctica médica*, (2005)
- [9] OECD., *How's Life? 2020: Measuring Well-being*, <https://doi.org/10.1787/9870c393-en> (2020)
- [10] Umaquinga, Ana Suárez Zambrano, Luis Oña, Omar., *Machine learning: Importance, advances, techniques and applications*, (2018)