

Proyecto de Bases de datos para el análisis de la ENUT

Cindy Catalina Vásquez Guerrero¹

Facultad de Ingeniería y Ciencias Básicas,

Universidad Central

Maestría en Analítica de Datos

Curso de Bases de Datos

Bogotá, Colombia

¹cvasquezg1@ucentral.edu.co,

October 8, 2022

Contents

1	Introducción (Max 250 Palabras) - (<i>Primera entrega</i>)	3
2	Características del proyecto de investigación (Max 500 Palabras) - (<i>Primera entrega</i>)	3
2.1	Titulo del proyecto de investigación (Max 100 Palabras) - (<i>Primera entrega</i>)	4
2.2	Objetivo general (Max 100 Palabras) - (<i>Primera entrega</i>)	4
2.2.1	Objetivos específicos (Max 100 Palabras) - (<i>Primera entrega</i>)	4
2.3	Alcance (Max 200 Palabras) - (<i>Primera entrega</i>)	4
2.4	Pregunta de investigación (Max 100 Palabras) - (<i>Primera entrega</i>)	4
2.5	Hipótesis (Max 100 Palabras) - (<i>Primera entrega</i>)	4
3	Reflexiones sobre el origen de datos e información (Max 400 Palabras) - (<i>Primera entrega</i>)	5
3.1	¿Cuál es el origen de los datos e información ? (Max 100 Palabras) - (<i>Primera entrega</i>)	5
3.2	¿Cuáles son las consideraciones legales o éticas del uso de la información? (Max 100 Palabras) - (<i>Primera entrega</i>)	5
3.3	¿Cuáles 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>)	5
3.4	¿Qué espera de la utilización de un sistema de Bases de Datos para su proyecto? (Max 100 Palabras) - (<i>Primera entrega</i>)	6

4	Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)(Primera entrega)	7
4.1	Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto (Primera entrega)	7
4.2	Diagrama modelo de datos (Primera entrega)	8
4.3	Imágenes de la Base de Datos (Primera entrega)	9
4.4	Código SQL - lenguaje de definición de datos (DDL) (Primera entrega)	9
4.5	Código SQL - Manipulación de datos (DML) (Primera entrega) . .	11
4.6	Código SQL + Resultados: Vistas (Primera entrega)	11
4.7	Código SQL + Resultados: Triggers (Primera entrega)	11
4.8	Código SQL + Resultados: Funciones (Primera entrega)	11
4.9	Código SQL + Resultados: procedimientos almacenados (Primera entrega)	11
5	Bases de Datos No-SQL (Segunda entrega)	12
5.1	Diagrama Bases de Datos No-SQL (Segunda entrega)	12
5.2	SMBD utilizado para la Base de Datos No-SQL (Segunda entrega)	12
6	Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos (Tercera entrega)	13
6.1	Ejemplo de aplicación de ETL y Bodega de Datos (Tercera entrega)	13
7	Lecciones aprendidas (Tercera entrega)	14
8	Bibliografía	15

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

La Encuesta Nacional del Uso del Tiempo ENUT es una investigación que busca generar información sobre el tiempo dedicado por las personas de 10 años o más a actividades de trabajo remunerado, no remunerado y actividades personales. En la edición 2016-2017 la encuesta se estructuró alrededor de 733 preguntas y recopiló la información de 44.999 hogares (DANE, 2018). Mientras que para la recolección 2020-2021 se definieron 767 preguntas y se recopilaron datos de 49.519 hogares (DANE, 2021). Ambas encuestas cuentan con representatividad en la mayoría del territorio nacional, pues se toman datos en los centros poblados, cabeceras municipales y rural disperso en seis regiones del país Bogotá, San Andrés, Caribe, Oriental, Central, y Pacífica. La cantidad de datos recopilados en ambas encuestas requiere de un manejo especial con el objetivo de permitir el uso eficiente de la información de los hogares durante las fases de Extracción, Transformación y Carga. Teniendo en cuenta este propósito, es necesario estructurar una base de datos que permita la organización de la información a través de tablas y de relaciones entre estas. De esta forma se puede mejorar la capacidad de modificación de los datos y su consulta, para futuros ejercicios estadísticos que permitan una mejor comprensión de los datos.

2 Características del proyecto de investigación (Max 500 Palabras) - (*Primera entrega*)

El tiempo dedicado al trabajo no remunerado es una de las desigualdades con mayor impacto en la vida cotidiana de las mujeres. Tradicionalmente, se piensan como tareas que deben desarrollar las mujeres de forma complementaria a su jornada laboral o incluso dedicándose por completo a ellas sin recibir ninguna compensación. La ENUT permite realizar la caracterización del tiempo dedicado por la población a actividades de trabajo no remunerado teniendo en cuenta características particulares de los hogares como quien es el jefe de hogar, sexo, estado civil, edad, número de hijos entre otras. Es por esta razón que el análisis de la encuesta permitirá describir mejor el problema de interés. Caracterizar el uso desigual del tiempo al trabajo no remunerado entre hombres y mujeres permite representar mejor las condiciones inequitativas asociadas a la dedicación a tareas del hogar relacionadas con el cuidado. La economía del cuidado propone reconocer de manera formal estas actividades. Reconociendo que “invertir en la economía del cuidado contribuye a la diversificación de la estructura productiva, sin trasgredir los límites ecológicos para la reproducción, lo que a su vez permite gestionar de forma más eficiente el patrimonio natural y garantizar la sostenibilidad de la vida” (CEPAL, 2020).

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

BASE DE DATOS PARA EL ANÁLISIS COMPARATIVO DEL USO DEL TIEMPO EN TRABAJO NO REMUNERADO A PARTIR DE LA ENCUESTA NACIONAL DEL USO DEL TIEMPO ENUT 2016-2017 Y 2020-2021.

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

- Estructurar una base de datos que permita el manejo de los resultados de la ENUT 2016-2017 y 2020-2021 sobre el tiempo dedicado al trabajo no remunerado.

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

- 1. Elegir un sistema manejador de base de datos para realizar la estructuración de los datos y definir el esquema relación-entidad.
- 2. Garantizar el cumplimiento de las reglas de normalización y realizar el cargue de la información que sustenta la base de datos.
- 3. Definir a partir de un lenguaje de programación para bases de datos los protocolos de creación, lectura, actualización y eliminación de los datos.

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

El alcance de la propuesta es definir un sistema de bases de datos que permita el análisis de los resultados de la ENUT en las versiones 2016-2017 y 2020-2021, para la estructuración de la base de datos se utiliza la información relacionada con el trabajo no remunerado con el fin de segmentar el análisis y construir una base de datos que de soporte a un usuario que corre en un solo computador, es decir que se clasifica como una “single-user desktop database”.

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

¿Al comparar las dos encuestas se pueden observar cambios en la cantidad de tiempo dedicado al trabajo no remunerado distribuido por sexo?

2.5 Hipótesis (Max 100 Palabras) - (Primera entrega)

La pandemia intensifico la carga de trabajo no remunerado en las mujeres.

3 Reflexiones sobre el origen de datos e información (Max 400 Palabras) - (*Primera entrega*)

La Encuesta Nacional del Uso del Tiempo es un instrumento que permite la recolección de información valiosa a cerca de la composición del hogar, el ámbito de salud y educación, el cuidado de los niños y niñas, la fuerza de trabajo y el uso del tiempo. Esta encuesta se recopila en la mayoría del territorio nacional, lo que brinda un marco de referencia importante en materia de bienestar de los hogares. Particularmente, con respecto al tiempo destinado a las labores no remuneradas como las tareas domésticas, el cuidado de los hijos o familiares, los tramites familiares como citas médicas o pago de facturas. Adicionalmente, se analiza el tiempo dedicado al descanso y a las actividades de ocio. Estas características de la encuesta la convierten en un insumo fundamental para medir el bienestar de las familias comparativamente antes y después de la pandemia.

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

El Departamento Administrativo Nacional de Estadística DANE a través de su sitio de micro-datos ha brindado acceso a la ciudadanía a los resultados de encuestas que se desarrollan para la caracterización del país en diferentes áreas como Economía, Sociedad y Territorio. En el ámbito de Sociedad en la sección de Género y relacionado con Economía del Cuidado se publican los resultados de la ENUT, ya que previamente han sido limpiados y organizados, en la página se puede encontrar el diccionario de datos y cada uno de los capítulos de la encuesta. Los capítulos definidos son: • Capitulo b. Condiciones de la vivienda • Capitulo c. Datos del hogar • Capitulo d. Composición del hogar • Capitulo e. Salud • Capitulo f. Cuidado de los niños y niñas menores de 5 • Capitulo g. Educación • Capitulo h. Fuerza de trabajo – clasificación, uso del tiempo e ingreso • Capitulo i. Uso del tiempo

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

La información no presenta ninguna consideración legal o ética restrictiva, corresponde más con el compromiso del investigador con informar a los lectores de la investigación el origen de la información y el uso que se le ha dado a la información publicada.

3.3 ¿Cuáles 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*)

Al ser datos publicados por el DANE en su portal web ya han sido revisados y avalados en cuanto a la calidad y consolidación de los datos, cada capítulo contiene un diccionario de datos, junto con los datos en formato CSV y además se encuentra publicada la documentación de la encuesta, como su metodología y resumen de los principales resultados encontrados.

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

Al implementar un sistema de bases de datos se espera que se mejore la eficiencia en la consulta de los datos contenidos en la encuesta, dado que al manejar archivos independiente para cada capítulo puede resultar poco eficiente la consulta y modificación d de los datos, temas que se podrían resolver a través de un sistema de base de datos que organice la información en tablas y permita el acceso a los datos a través de consultas personalizadas.

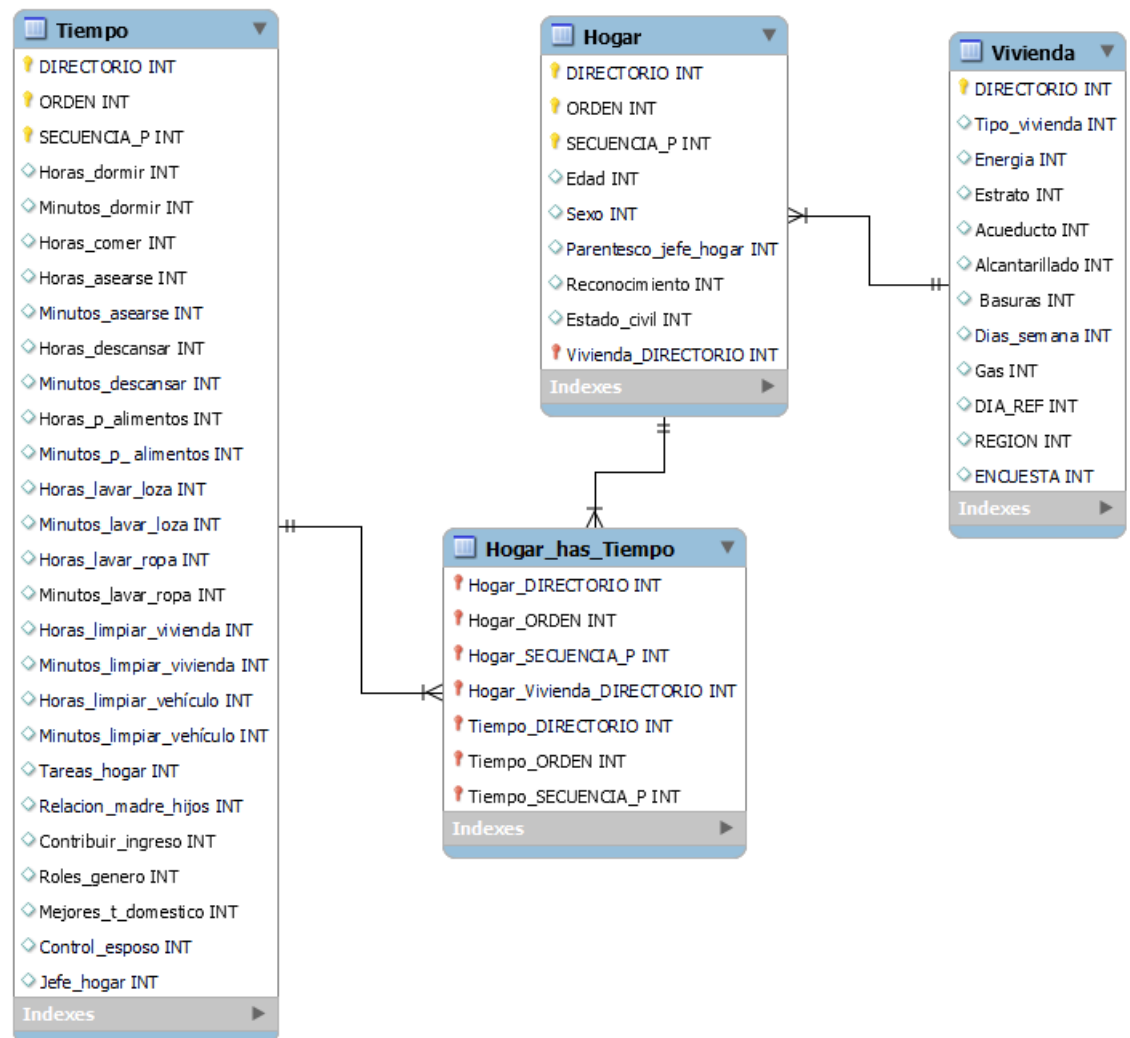
4 Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos) *(Primera entrega)*

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

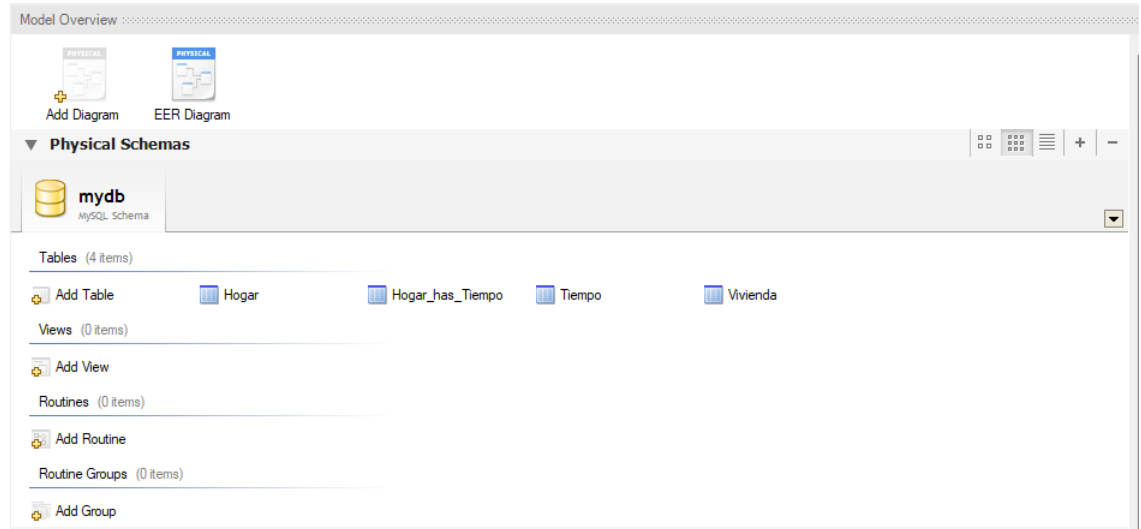
Como sistema manejador de base de datos se ha seleccionado Mysql Workbench que permite recolectar datos de manera eficiente (Silberschatz et al., 2005). Algunos de los beneficios del uso de este SMBD son:

- Estructurar un esquema de datos que permite su manipulación
- Definir la estructura de almacenamiento y metodología de acceso
- Compartir la base de datos de manera organizada
- Delimitar rutinas de mantenimiento para garantizar la eficiencia de la base de datos
- Permitir el acceso y la modificación de las tablas y los registros

4.2 Diagrama modelo de datos (Primera entrega)



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



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

(*Tabla 1*) Tabla Vivienda

```
DROP TABLE IF EXISTS `Vivienda`;  
  
CREATE TABLE `Vivienda` (  
  `DIRECTORIO` int NOT NULL,  
  `Tipo_vivienda` int NOT NULL,  
  `Energia` int NOT NULL,  
  `Estrato` int,  
  `Acueducto` int NOT NULL,  
  `Alcantarillado` int NOT NULL,  
  `Basuras` int NOT NULL,  
  `Gas` int NOT NULL,  
  `REGION` int NOT NULL,  
  `ENCUESTA` varchar(15) NOT NULL,  
  PRIMARY KEY (`DIRECTORIO`)  
) ENGINE=InnoDB DEFAULT CHARSET=latin1;
```

(Tabla 2) Tabla Hogar

```
DROP TABLE IF EXISTS `Hogar`;  
  
CREATE TABLE `Hogar` (  
  `ID` int NOT NULL,  
  `DIRECTORIO` int NOT NULL,  
  `ORDEN` int NOT NULL,  
  `SECUENCIA_P` int NOT NULL,  
  `Edad` int NOT NULL,  
  `Sexo` int NOT NULL,  
  `Parentesco_jefe_hogar` int NOT NULL,  
  `Reconocimiento` int NOT NULL,  
  `Estado_civil` int,  
  PRIMARY KEY (`ID`),  
  KEY `DIRECTORIO` (`DIRECTORIO`),  
  KEY `ORDEN` (`ORDEN`),  
  KEY `SECUENCIA_P` (`SECUENCIA_P`)  
) ENGINE=InnoDB DEFAULT CHARSET=latin1;
```

(Tabla 3) Tabla Tiempo

```
CREATE TABLE `Tiempo` (  
  `ID` int NOT NULL,  
  `DIRECTORIO` int NOT NULL,  
  `ORDEN` int NOT NULL,  
  `SECUENCIA_P` int NOT NULL,  
  `Horas_dormir` int,  
  `Minutos_dormir` int,  
  `Horas_comer` int,  
  `Minutos_comer` int,  
  `Horas_asearse` int,  
  `Minutos_asearse` int,  
  `Horas_descansar` int,  
  `Minutos_descansar` int,  
  `Horas_p_alimentos` int,  
  `Minutos_p_alimentos` int,  
  `Horas_lavar_loza` int,  
  `Minutos_lavar_loza` int,  
  `Horas_lavar_ropa` int,  
  `Minutos_lavar_ropa` int,  
  `Horas_limpiar_vivienda` int,  
  `Minutos_limpiar_vivienda` int,  
  `Horas_limpiar_vehículo` int,  
  `Minutos_limpiar_vehículo` int,  
  `Tareas_hogar` int,  
  `Relacion_madre_hijos` int,  
  `Contribuir_ingreso` int,  
  `Roles_genero` int,  
  `Mejores_t_domestico` int,  
  `Control_esposo` int,  
  `Jefe_hogar` int,  
  PRIMARY KEY (`ID`),  
  KEY `DIRECTORIO` (`DIRECTORIO`),  
  KEY `ORDEN` (`ORDEN`),
```

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

(*Manipulación de datos*)

```
USE ENUT;

select * from vivienda;
select * from hogar;
select * from tiempo;

# 1 ES CARIBE
SELECT * FROM vivienda
WHERE REGION like "1";

# Seleccionar region, estrato y tipo de vivienda de la tabla vivienda
SELECT REGION, Estrato, Tipo_vivienda FROM vivienda;

#Selecciona las personas que duermen mas de seis horas
SELECT ID, Horas_dormir FROM tiempo
WHERE Horas_dormir > '6';

#Ver las opciones de region
SELECT DISTINCT REGION FROM vivienda
ORDER BY REGION ASC;
```

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

(*Vistas*)

```
#Crear una vista
CREATE VIEW Hogares AS
SELECT * FROM hogar;
```

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

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

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

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

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

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

- 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*)

8 Bibliografía

- CEPAL. (2020). La economía del cuidado como acelerador del cambio estructural con igualdad. Notas Para La Igualdad, 30(Observatorio de Igualdad de Género de America Latina y el Caribe). https://oig.cepal.org/sites/default/files/no30_es-economia_del_cuidado.pdf
- Cristina Benavente Alejandra Valdés B, M. R. (2014). Desarrollo Social Políticas públicas para la igualdad de género Un aporte a la autonomía de las mujeres. www.cepal.org/es/suscripciones
- DANE. (2018). Boletín técnico Encuesta Nacional de Uso del Tiempo (ENUT) 2016-2017. 28. https://www.dane.gov.co/files/investigaciones/boletines/ENUT/Bol_ENUT_2016_2017.pdf
- DANE, D. A. N. E. (2021). Boletín Técnico Encuesta Nacional del Uso del Tiempo (ENUT) 2020-2021. 1–59.
- Silberschatz, A., Korth, H., Sudarshan S. (2005). Database management. In Disability and Rehabilitation (Vol. 27, Issue 16). <https://doi.org/10.1080/09638280500030605>