



UTPL
La Universidad Católica de Loja

Universidad Técnica Particular de Loja

INFORME FINAL

Proyecto Integrador

ÁREA TÉCNICA

**INGENIERÍA EN CIENCIAS DE LA
COMPUTACIÓN**

Isaac Quezada
Jhandry Solorzano

Abril 2023 – Agosto 2023

Contenido

| | |
|-------------------------------------|----|
| Introducción | 3 |
| Repositorio Git Hub | 4 |
| Datos Base | 5 |
| Datos Complementarios | 6 |
| Base De Datos..... | 7 |
| Diseño E+R | 7 |
| Diseño lógico..... | 8 |
| Scripts SQL | 9 |
| Formato CSV | 9 |
| Programación Avanzada | 10 |
| Sistema operativo | 10 |
| Herramientas utilizadas | 11 |
| Instalación de Máquina Virtual..... | 11 |
| Instalación Ubuntu. | 11 |
| Instalación zeppelin..... | 11 |
| Instalación spark. | 12 |
| Instalación MySQL. | 12 |
| Análisis por Realizar | 15 |
| Consulta 1 | 16 |
| Consulta 2 | 17 |
| Consulta 3 | 18 |
| Consulta 4 | 19 |
| Consulta 5 | 20 |
| Página Web..... | 21 |
| Bibliografía: | 22 |

Introducción

En el presente documento se redactará un informe que detalla el proceso llevado a cabo para el componente del prácticum 1.2, el cual se centra en la materia de Base de Datos. Este proceso consta de varias etapas, las cuales se describen a continuación:

- **Datos base:** Se proporciona una descripción de los datos base utilizados en el proyecto. Estos datos constituyen la información primaria sobre la cual se realizarán las distintas operaciones y análisis.
- **Datos complementarios:** Se describe la naturaleza de los datos complementarios, su propósito y las fuentes de donde se obtuvieron. Estos datos complementan la información de los datos base y permiten ampliar el alcance y la calidad de los análisis realizados.
- **Posibles análisis a realizar:** Se realiza una lluvia de ideas sobre los análisis que se desearía llevar a cabo utilizando tanto los datos base como los complementarios. Esta etapa busca identificar las posibles preguntas o hipótesis que se pretenden responder mediante el análisis de los datos.
- **Elaboración del diagrama entidad-relación:** Se procede a elaborar el diagrama entidad-relación, el cual representa visualmente las entidades (objetos o conceptos) presentes en la base de datos, así como las relaciones entre ellas. Este diagrama ayuda a comprender la estructura y las interacciones de los datos.
- **Diagrama lógico:** Se realiza la construcción del diagrama lógico, el cual muestra la organización y la estructura de la base de datos de manera más detallada. En este diagrama se definen las tablas, los atributos y las restricciones necesarias para representar adecuadamente los datos.
- **Creación del script DDL (lenguaje de definición de datos):** Se crea el script DDL, el cual contiene las instrucciones necesarias para crear las tablas y definir la estructura de la base de datos. Este script se ejecutará posteriormente para implementar la base de datos.
- **Implementación física y población de la base de datos:** Se lleva a cabo la implementación física de la base de datos, lo que implica la creación de las tablas y la configuración de los índices y las claves. Además, se procede a poblar la base de datos con los datos disponibles.

Repositorio Git Hub

<https://github.com/Isaac-Quezada/ProyectoIntegradorC4.git>

Con el propósito de impulsar la transparencia, la reproducibilidad y la colaboración en la investigación, hemos decidido almacenar todos los datos recopilados y los análisis realizados en un repositorio de GitHub. GitHub es una plataforma altamente reconocida y ampliamente utilizada para la gestión de versiones y el almacenamiento de proyectos colaborativos.

La elección de GitHub como nuestro repositorio de datos y análisis se basa en varias ventajas clave. En primer lugar, GitHub permite mantener un registro histórico completo de todos los cambios realizados en los datos y análisis, lo que facilita el seguimiento y la revisión de los resultados obtenidos. Esto asegura la integridad y la trazabilidad de los datos a lo largo del tiempo, y proporciona la posibilidad de verificar y validar los procesos y métodos utilizados en cualquier etapa del estudio.

Además, GitHub ofrece una interfaz intuitiva y fácil de usar, lo que facilita la colaboración y el intercambio de conocimientos entre los miembros del equipo de investigación y otros interesados. Los colaboradores pueden acceder a los datos y análisis almacenados en el repositorio, realizar comentarios, sugerir mejoras y realizar contribuciones directas al proyecto. Esto fomenta la transparencia y la participación de la comunidad científica en la investigación, lo que a su vez puede enriquecer y fortalecer los resultados y conclusiones obtenidos.

.

Datos Base

La base de datos contiene información detallada sobre la Encuesta Nacional de Empleo, Desempleo y Subempleo realizada en febrero de 2023, enfocada en la Gestión de Estadísticas Permanentes a Hogares. Estos detalles proporcionan una visión completa de la estructura de la base de datos, permitiendo un análisis exhaustivo de los aspectos relacionados con el empleo y las características socioeconómicas de los hogares y las personas encuestadas durante ese período específico. A continuación, se presenta una descripción de las variables presentes en los datos:

- Área: Indica la ubicación geográfica del hogar encuestado dentro de la muestra.
- Ciudad: Identifica la ciudad o localidad donde se encuentra el grupo de viviendas seleccionado para la encuesta.
- Conglomerado: Número de identificación asignado al grupo de viviendas seleccionado.
- Panelm: Identificador del panel de hogares al que pertenece el hogar encuestado.
- Vivienda: Número de identificación de la vivienda seleccionada dentro del grupo.
- Hogar: Número de identificación del hogar encuestado dentro de la vivienda seleccionada.
- Variables vi01 a vi1546: Representan diferentes aspectos relacionados con el empleo, desempleo y subempleo, brindando información detallada sobre características demográficas, laborales y socioeconómicas de los hogares y las personas encuestadas.
- Estrato: Indica el nivel socioeconómico al que pertenece la vivienda encuestada.
- Fexp: Factor de expansión asignado a la vivienda encuestada para obtener estimaciones representativas a nivel nacional.
- Upm: Unidad primaria de muestreo a la que pertenece la vivienda encuestada.
- d_vivienda: Identificador único de la vivienda.
- Id_hogar: Identificador único del hogar.
- Periodo: Año y mes de la encuesta.
- Mes: Mes específico de investigación (febrero de 2023 en este caso).

Datos Complementarios

- Matrimonios y Divorcios

Enlace: [Matrimonios y Divorcios | \(ecuadorencifras.gob.ec\)](https://ecuadorencifras.gob.ec/matrimonios-y-divorcios)

Objetivo: Cuantificar los hechos vitales de matrimonios y divorcios ocurridos e inscritos en las oficinas de Registro Civil a nivel nacional. La información se obtiene a través de las inscripciones realizadas en las oficinas de la Dirección Nacional de Registro Civil, Identificación y Cedulación y la Corporación Registro Civil de Guayaquil durante el año de investigación.

- Homicidios Enero-junio 2023:

Enlace: [Homicidios Enero-junio 2023](#)

Objetivo: Obtener datos sobre homicidios a nivel de ciudad para comprender mejor la peligrosidad y el nivel de delincuencia en distintas ubicaciones, complementando así la información a nivel de cantón.

- Salud:

Enlace: [Indicadores de Salud](#)

Objetivo: Obtener el porcentaje de salud en un rango del 0 al 100 % a nivel de ciudad/parroquia para analizar los niveles de salud de los ciudadanos en diferentes áreas geográficas y calcular un promedio de salud por parroquia, complementando la información a nivel de ciudad.

- Emergencias abril 2023:

Enlace: [Base de Emergencias abril 2023](#)

Objetivo: Obtener un promedio a nivel parroquial/ciudad de los reportes de incidentes al servicio de emergencias Ecu911 para evaluar la inseguridad a nivel cantonal y su impacto en los hogares.

- Hospitalización INEC:

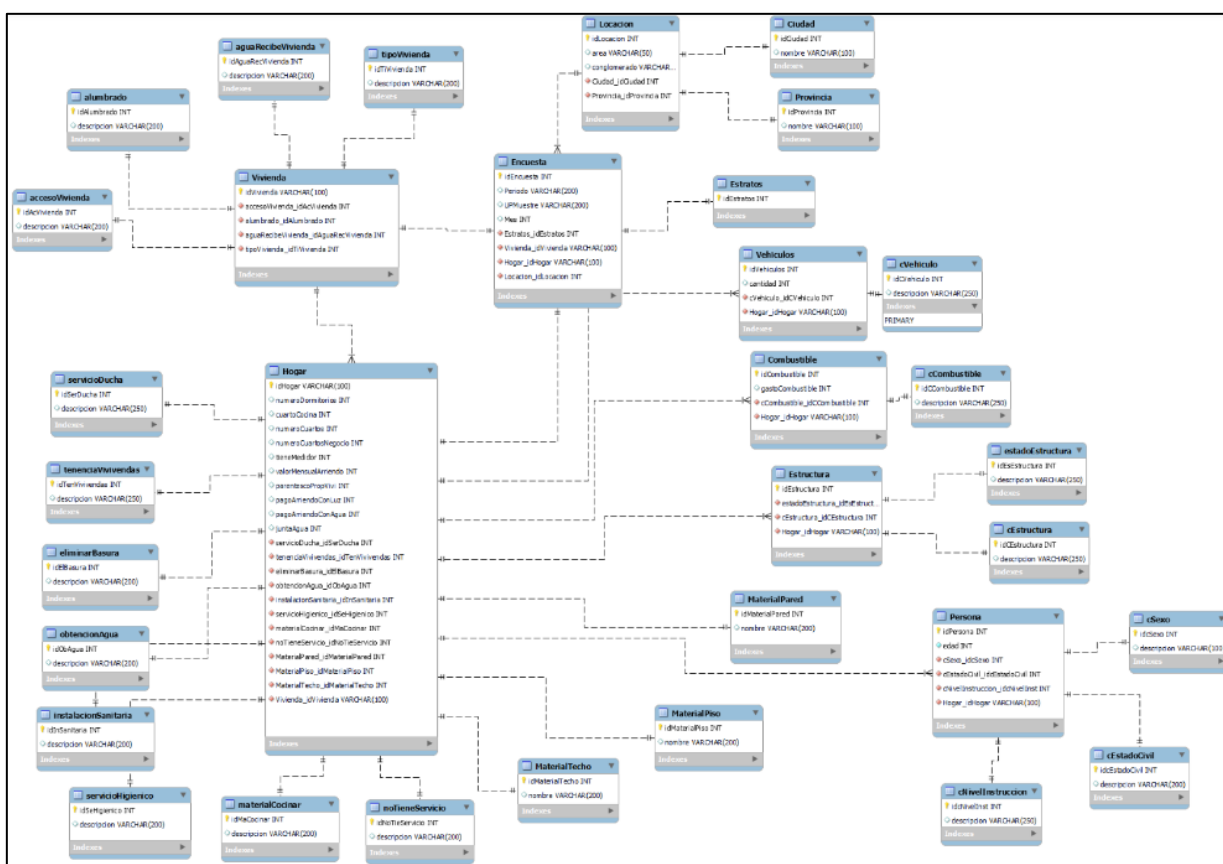
Enlace: [Camas y Egresos Hospitalarios INEC 2022](#)

Objetivo: Obtener datos sobre la disponibilidad hospitalaria a nivel provincial para evaluar el sector salud para cada hogar, complementando así la información a nivel provincial.

Diseño lógico

Un modelo de base de datos física muestra todas las estructuras de tabla, incluidos el nombre de columna, el tipo de datos de columna, las restricciones de columna, la clave principal, la clave externa y las relaciones entre las tablas (tecnologías-informacion.com, s.f.). Es por ello que el modelo físico toma como referencia el diseño conceptual.

El modelo conceptual previamente elaborado facilitó el diseño del físico ya que las tablas del archivo CSV que en un principio eran nombres en código, ya tenían un nombre establecido, además de que cada una de las columnas estaba distribuida en las tablas de manera estratégica, facilitando la inserción de datos y futuras consultas a la base de datos.



Scripts SQL

El siguiente enlace al repositorio GitHub muestra el script realizado para la implementación y carga de los datos en el esquema, tanto de los catálogos como la información de las tablas principales:

[ProyectoIntegradorC4/Base De Datos Avanzada/Script SQL at main · Isaac-Quezada/ProyectoIntegradorC4 · GitHub](#)

Formato CSV

A continuación, se presenta el link hacia el documento generado en un csv. Lo que se realizó fue, dentro de programación, la unión de todas las diferentes columnas que no fuesen catálogos. De esta forma se paso la base de datos en formato CSV:

[Isaac-Quezada/ProyectoIntegradorC4: Proyecto Integrador Realizado en el cuarto ciclo \(github.com\)](#)

Programación Avanzada

Sistema operativo

Windows Core i7

Resumen de Apache **Zeppelin**: Apache Zeppelin es un entorno de desarrollo web basado en notebooks que permite realizar análisis de datos y generar gráficos estadísticos interactivos. Con la capacidad de combinar código, visualizaciones y texto en un solo documento, Zeppelin es ideal para explorar datos, realizar análisis exploratorios y presentar resultados de manera interactiva y visualmente atractiva. Además, es compatible con múltiples lenguajes de programación, lo que brinda flexibilidad para trabajar con diferentes conjuntos de datos y realizar análisis complejos. (zeppelin.org)

Resumen de Apache **Spark**: Apache Spark es un potente motor de procesamiento de datos distribuido que ofrece un entorno rápido y escalable para realizar cálculos en grandes conjuntos de datos. Su capacidad de procesamiento en memoria permite realizar análisis de datos de manera eficiente y ejecutar operaciones complejas, como el procesamiento en tiempo real y el aprendizaje automático. Spark es ampliamente utilizado para realizar tareas de procesamiento de datos enriquecidas y complejas, lo que facilita el análisis y la generación de resultados significativos. (spark.org)

Resumen de **MySQL**: MySQL es un sistema de gestión de bases de datos relacionales ampliamente utilizado debido a su confiabilidad y eficiencia en el almacenamiento y gestión de grandes cantidades de datos estructurados. Proporciona un entorno robusto para almacenar datos obtenidos de diversas fuentes y realizar consultas complejas, asegurando la integridad y seguridad de los datos almacenados. MySQL es altamente compatible con otras herramientas y lenguajes de programación, lo que facilita la extracción y manipulación de datos para su posterior análisis y generación de informes. (Oracle.mysql-com)

Herramientas utilizadas

Instalación de Máquina Virtual.

Para instalar Oracle VirtualBox, primero accedemos al sitio web oficial <https://www.virtualbox.org/wiki/Downloads> y seleccionamos el "Platform Package" adecuado para nuestro sistema operativo. Luego, ejecutamos el instalador que hemos descargado. Una vez completada la instalación, nos encontramos con la interfaz de Oracle VirtualBox, donde podemos agregar y configurar las diversas máquinas virtuales que deseemos utilizar. Esta plataforma nos permite crear y administrar máquinas virtuales para ejecutar diferentes sistemas operativos en nuestro equipo de forma segura y aislada del sistema principal. Con Oracle VirtualBox, podemos experimentar con distintas configuraciones y entornos sin afectar el funcionamiento del sistema anfitrión.

Instalación Ubuntu.

Dentro de Oracle VirtualBox, para instalar Ubuntu como máquina virtual, procedemos a hacer clic en la sección que dice "Nueva" para crear una nueva máquina virtual. A continuación, seguimos las indicaciones del asistente para configurar nuestra máquina virtual Ubuntu.

En el asistente, seleccionamos el tipo de sistema operativo como "Linux" y la versión "Ubuntu (64-bit)" si corresponde a la versión que hemos descargado. Luego, definimos la cantidad de memoria RAM que queremos asignar a la máquina virtual.

A continuación, creamos un disco duro virtual, seleccionando la opción "Crear un disco duro virtual ahora". Durante este proceso, podemos elegir entre un disco duro virtual dinámico o fijo, y asignar el tamaño necesario para la instalación de Ubuntu.

Finalmente, antes de completar la creación de la máquina virtual, seleccionamos el archivo de instalación de Ubuntu (la imagen ISO) que hemos descargado previamente. Con esto, la máquina virtual estará lista para ejecutar el proceso de instalación de Ubuntu.

Instalación zeppelin.

Para instalar Apache Zeppelin en Ubuntu, siga estos pasos:

1. Comience actualizando el sistema mediante el comando:

```
sudo apt update
```

2. Si no tiene Curl instalado, instálelo usando:

```
sudo apt install curl
```

3. Descargue Apache Zeppelin usando el comando:

```
curl https://d1cdn.apache.org/zeppelin/zeppelin-0.10.1/zeppelin-0.10.1-bin
```

4. Descomprima el archivo descargado utilizando:

```
tar -zxvf zp.tgz
```

Instalación spark.

1. Descargue Apache Spark utilizando el siguiente comando:

```
curl https://archive.apache.org/dist/spark/spark-3.2.1/spark-3.2.1-bin-hado
```

2. Descomprima el archivo descargado:

```
tar -zxvf spark.tgz
```

Instalación MySQL.

Instalación de MySQL en Ubuntu:

Actualice el sistema:

```
sudo apt update
```

Instale MySQL Server:

```
sudo apt-get install mysql-server
```

Inicie el servicio de MySQL:

```
sudo service mysql start
```

Ejecute el cliente de MySQL:

```
sudo mysql
```

Cambie la autenticación de la cuenta root (asegúrese de reemplazar 'PA123456' con su contraseña deseada):

```
mysql> ALTER USER 'root'@'localhost' IDENTIFIED WITH mysql_native_password
mysql> exit
```

Ejecute la configuración segura de MySQL:

```
sudo mysql_secure_installation
```

Conéctese al cliente MySQL (esta vez se utilizará el método de autenticación de socket en lugar de contraseña):

```
mysql -u root -p
```

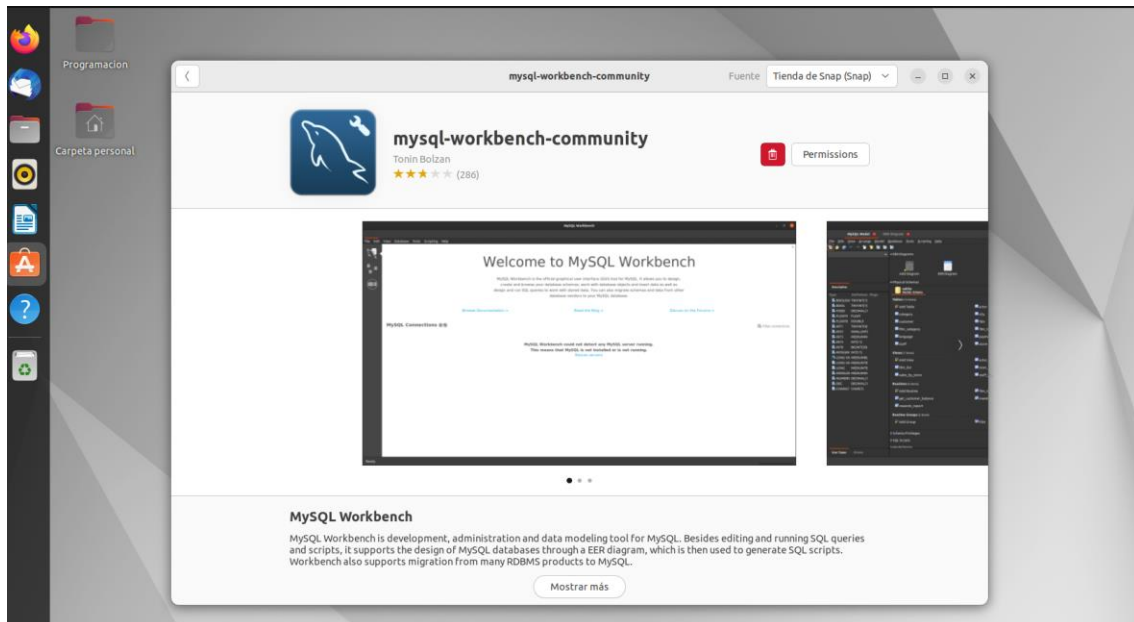
En el cliente MySQL, modifique la autenticación de nuevo (esta vez para el usuario 'root'):

```
mysql> ALTER USER 'root'@'localhost' IDENTIFIED WITH auth_socket;  
mysql> exit
```

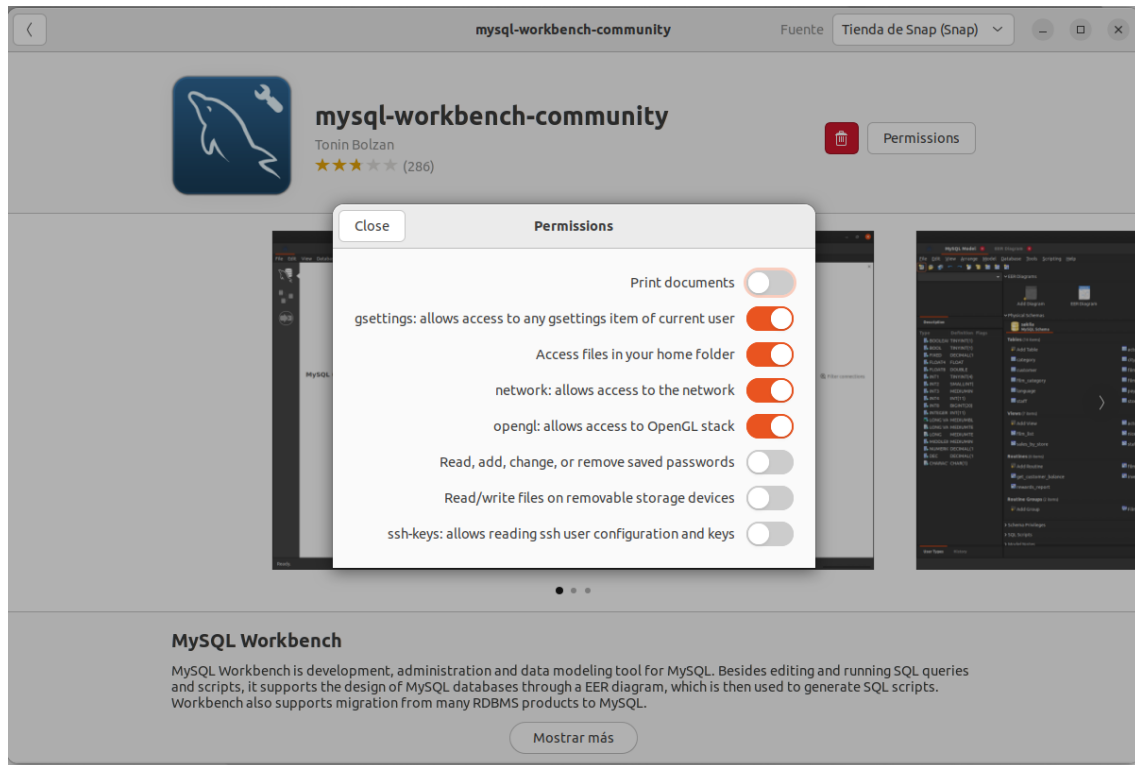
Cree un nuevo usuario de MySQL y asigne privilegios (reemplace 'nuevo_usuario' y 'contraseña' con los valores deseados):

```
sudo mysql  
mysql> CREATE USER 'nuevo_usuario'@'localhost' IDENTIFIED BY 'contraseña';  
mysql> GRANT ALL PRIVILEGES ON * . * TO 'nuevo_usuario'@'localhost';  
mysql> FLUSH PRIVILEGES;  
mysql> exit
```

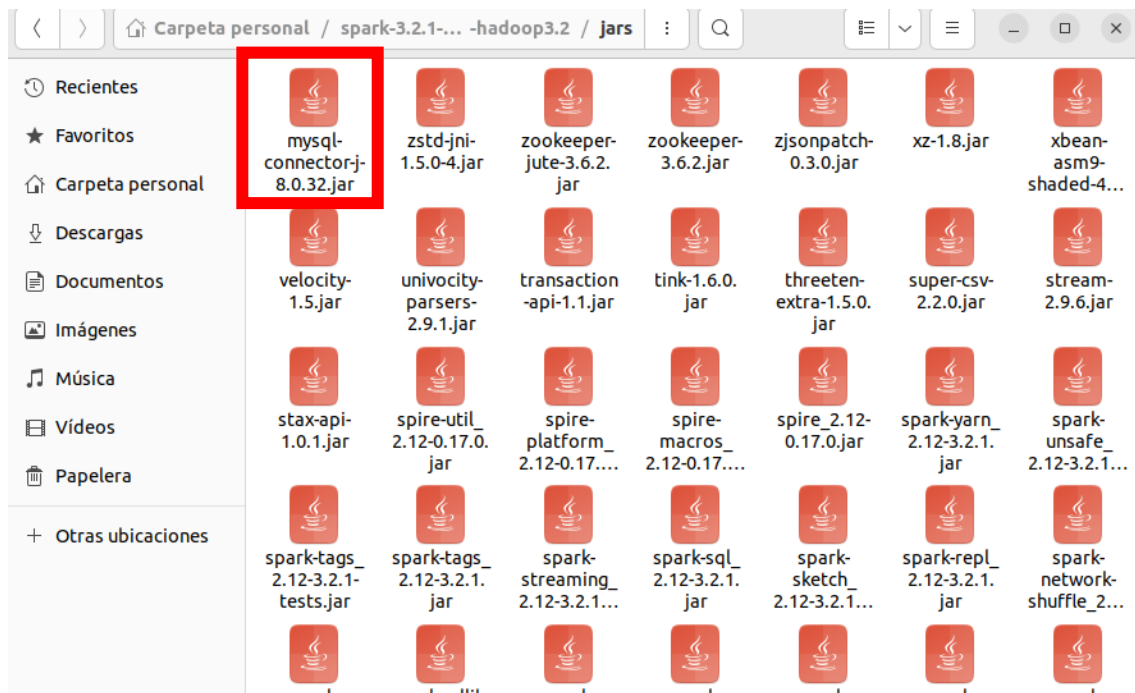
La aplicación de Mysql Workbench la descargamos desde la tienda de Ubuntu:



Luego toca habilitar un permiso necesario que es “Read, add, change, or remove saved passwords” que esto sirve para poder acceder con la contraseña que modificamos anteriormente.



Para poder conectar con la base de datos se añadió un archivo .jar:



Análisis por Realizar

Para realizar nuestro análisis, utilizaremos Zeppelin, una herramienta poderosa y versátil que nos permitirá ejecutar consultas en nuestra base de datos y visualizar los resultados de forma interactiva. Con Zeppelin, obtendremos los datos necesarios para cada uno de nuestros análisis previamente diseñados. Zeppelin destaca por su capacidad para visualizar los datos en diversos tipos de gráficos, como pie charts y otras representaciones visuales estadísticas. Estas visualizaciones facilitan la exploración y comprensión rápida de la información relevante, lo que ayuda en la toma de decisiones informadas y la identificación de patrones significativos en los datos.

1. Distribución de Homicidios por Tipo: A través de esta consulta, obtendremos una visión general de cómo se distribuyen los homicidios en distintas categorías de crímenes violentos. Esto nos ayudará a identificar los tipos de homicidios más comunes y las áreas que requieren mayor atención en términos de seguridad.

2. Provincias con Mayor Registro de Matrimonios: Mediante esta consulta, identificaremos las provincias con un mayor número de matrimonios registrados, lo que será valioso para planificar recursos y políticas relacionadas con la familia.

3. Matrimonios y Población por Provincia: Comparando el número total de matrimonios con la población por provincia, podremos analizar la relación entre el tamaño de la población y la tasa de matrimonios, ayudándonos a comprender cómo los factores demográficos influyen en las decisiones matrimoniales.

4. Tamaño Promedio de Hogares por Provincia: Con esta consulta, obtendremos información sobre el tamaño promedio de los hogares en cada provincia, lo que será útil para comprender las necesidades de vivienda y la planificación urbana en diferentes regiones.

5. Estadísticas de Viviendas por Provincia: A través de esta consulta, obtendremos diversas estadísticas sobre viviendas en cada provincia, incluyendo el número total de hogares, cuartos para dormir y cocinar, y el número máximo y mínimo de vehículos por hogar. Estos datos nos permitirán obtener una visión integral de las condiciones de vida en cada área, y detectar posibles desigualdades en el acceso a viviendas y recursos.

Con Zeppelin como herramienta para realizar estos análisis, obtendremos una comprensión más profunda de los patrones sociales y demográficos en diferentes regiones. Los resultados obtenidos servirán como base para tomar decisiones informadas y desarrollar políticas públicas que aborden las necesidades específicas de cada provincia. Además, la presentación visualmente atractiva y clara de Zeppelin nos

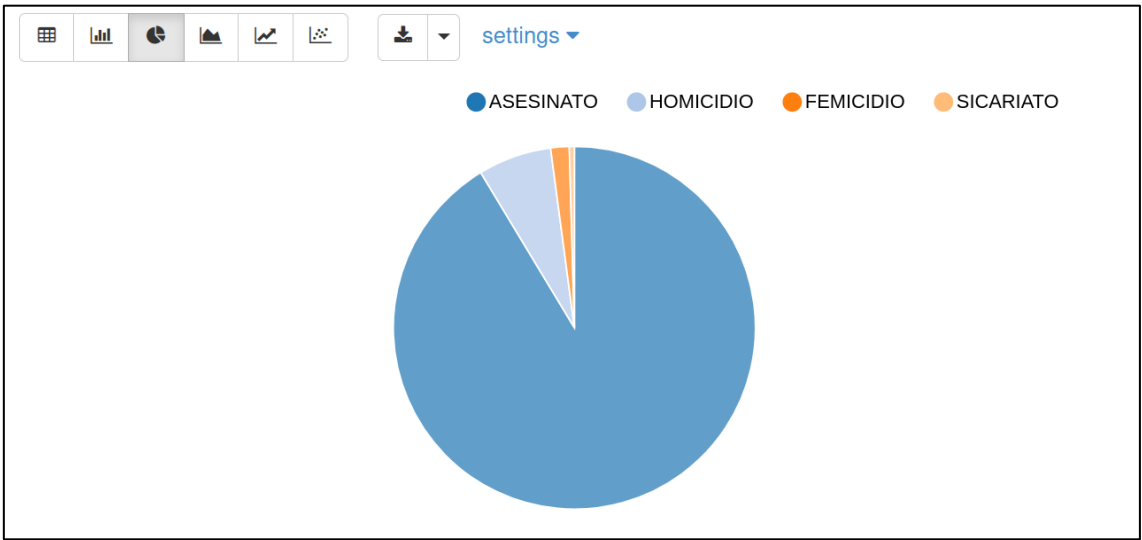
permitirá comunicar eficazmente nuestros hallazgos a otros miembros del equipo y tomadores de decisiones.

Consulta 1

Conteo de Homicidios por Tipo

Esta consulta cuenta el número de homicidios registrados, agrupados por el tipo de homicidio. Proporciona una visión general de la distribución de homicidios según las distintas categorías de crímenes violentos.

a. Grafica

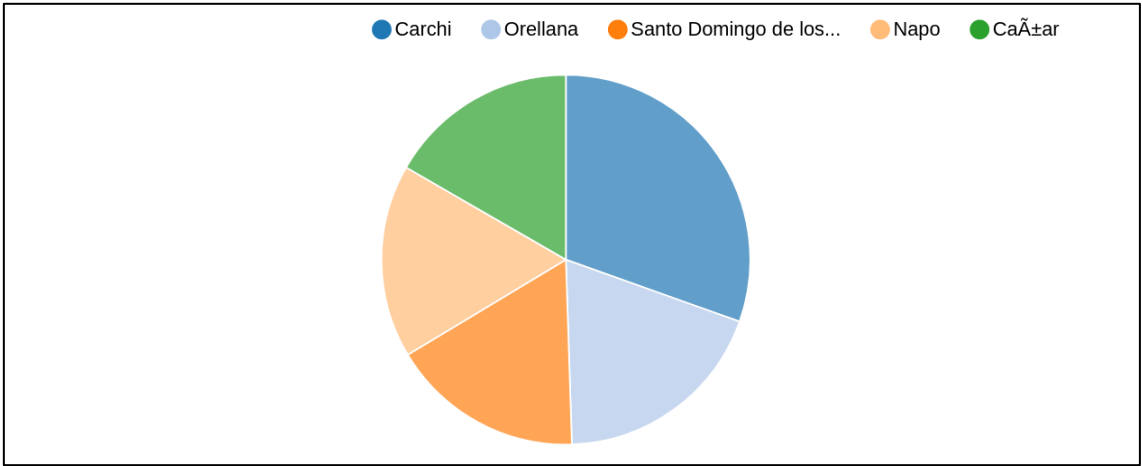


Consulta 2

Provincias con más Matrimonios

Esta consulta muestra las cinco provincias con el mayor número de matrimonios registrados. Ordena las provincias en función de la cantidad total de matrimonios, lo que permite identificar las áreas con más actividad matrimonial.

b. Grafica

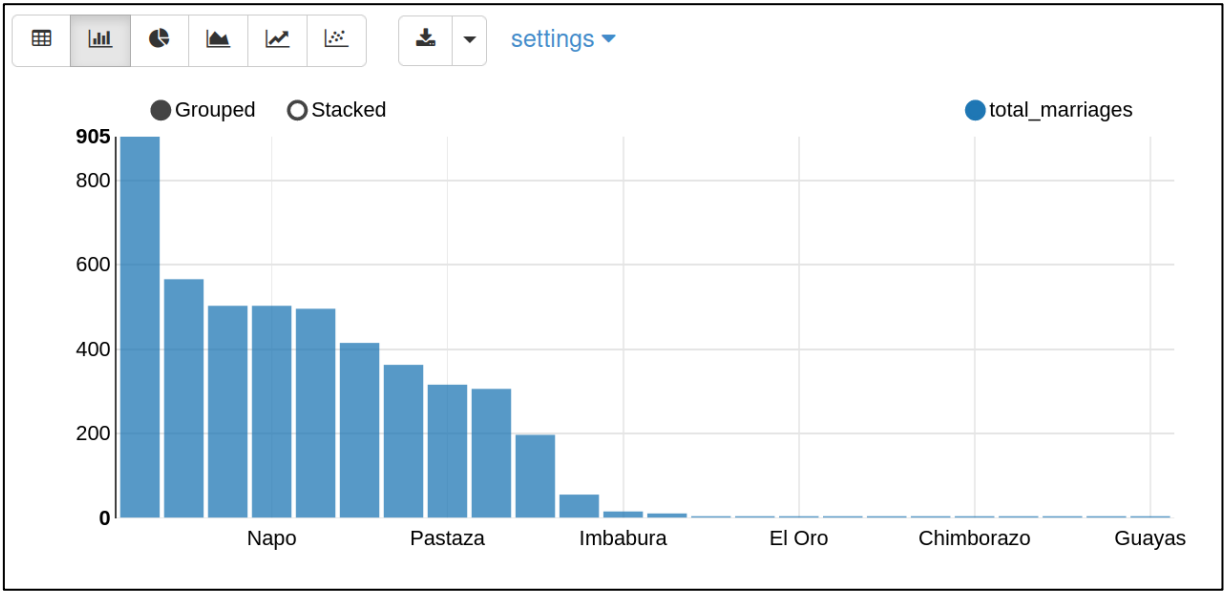


Consulta 3

Matrimonios y Población por Provincia

Esta consulta presenta el número total de matrimonios y la población total para cada provincia. Al combinar estos datos, se puede analizar la correlación entre el número de matrimonios y el tamaño de la población en cada área.

c. Grafica

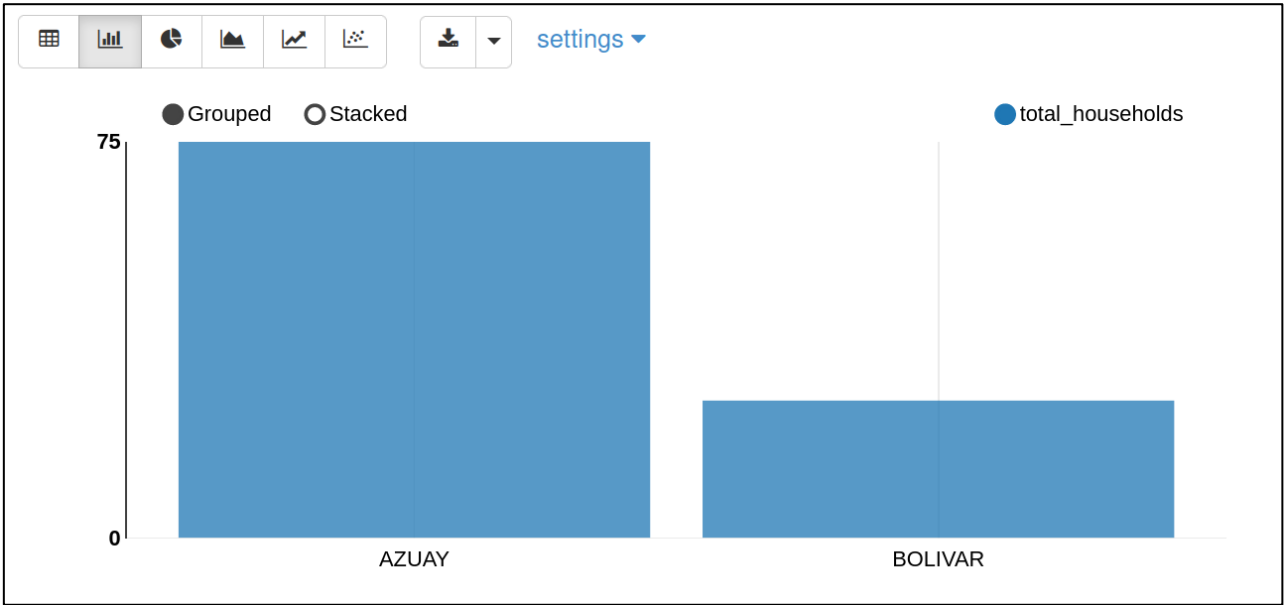


Consulta 4

Número de Hogares y Promedio de Cuartos por Hogar por Provincia

Esta consulta ofrece el número total de hogares y el promedio de cuartos por hogar para cada provincia. Esto brinda una perspectiva de la distribución del tamaño de los hogares en diferentes áreas geográficas.

d. Grafica



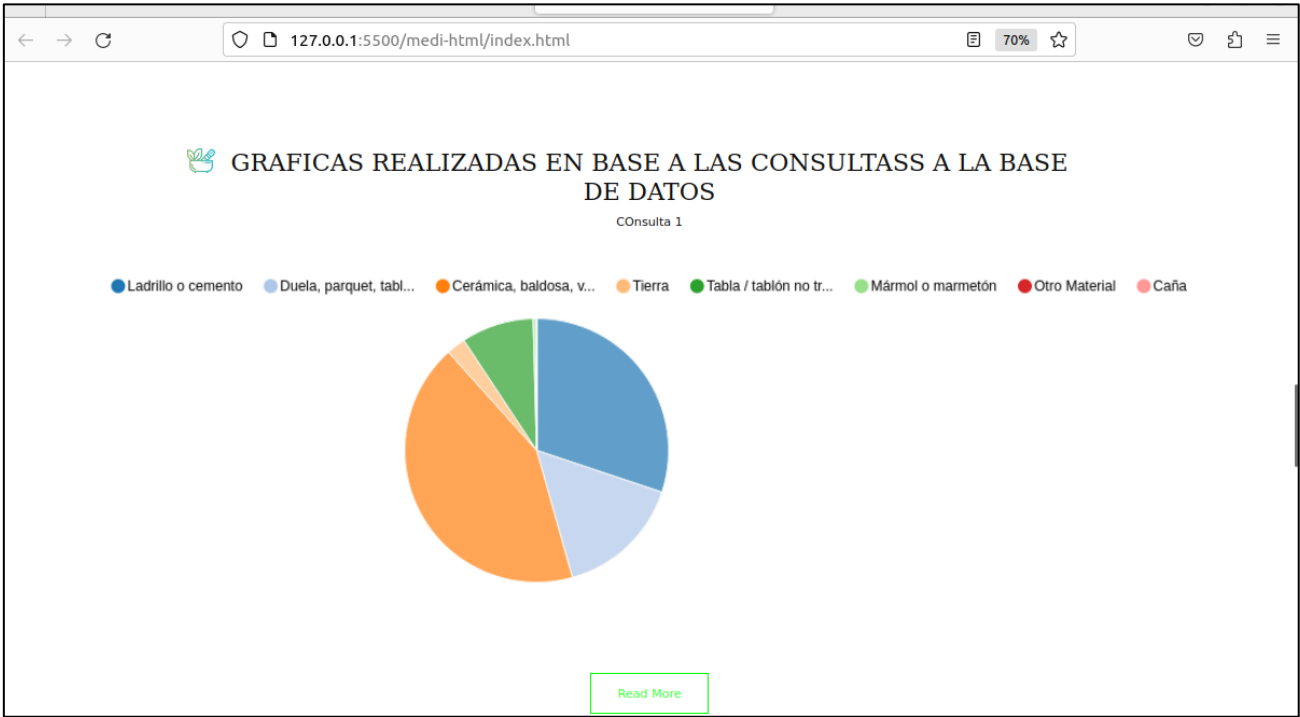
Consulta 5

Estadísticas Adicionales de Viviendas por Provincia

Esta consulta proporciona varias estadísticas sobre las viviendas en cada provincia, incluyendo el número total de hogares, la suma de cuartos para dormir y cocinar, el promedio de cuartos por hogar, así como el máximo y mínimo de vehículos por hogar. Estos datos brindan una visión general de las condiciones de vida en diferentes áreas geográficas.

e. Grafica

| provincia | total_households | total_bedrooms | total_kitchen_rooms | average_rooms_per_household | max_vehicles_per_household | min_vehicles_per_household |
|-----------|------------------|----------------|---------------------|-----------------------------|----------------------------|----------------------------|
| AZUAY | 75 | 155 | 79 | 3.0667 | 2 | 1 |
| BOLIVAR | 26 | 62 | 26 | 3.4231 | 2 | 1 |



Bibliografía:

Datos Complementarios:

1. Enlace: [Matrimonios y Divorcios | \(ecuadorencifras.gob.ec\)](https://ecuadorencifras.gob.ec/matrimonios-y-divorcios)
2. Enlace: [Homicidios Enero-junio 2023](#)
3. Enlace: [Indicadores de Salud](#)
4. Enlace: [Base de Emergencias abril 2023](#)
5. Enlace: [Camas y Egresos Hospitalarios INEC 2022](#)

Teoría Programación

1. [Cómo instalar Apache Spark en Ubuntu 20.04 - HowtoForge](#)
2. [Instalar y configurar notebooks Zeppelin para trabajar en Spark - Observatorio BI & Analytics \(spaceanalytics.blogspot.com\)](#)
3. [Cómo instalar MySQL en Ubuntu 20.04 | DigitalOcean](#)