UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA

La Universidad Católica de Loja



ESTUDIANTES:

ÁNGEL DAVID PARDO CORREA

JOSÉ ADRIÁN CRIOLLO JIMÉNEZ

MARÍA ALEJANDRA PAUTE ERAS

JOEL SEBASTIAN ROMERO CAÑAR

ÁREA TÉCNICA

INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN

Proyecto Integrador (Componente de Base de Datos): INFORME FINAL

DOCENTE:

Mgtr. ENCALADA ENCALADA ANGEL EDUARDO

ABRIL 2022-AGOSTO 2022

Contenido

1.	Introducción	4
2.	Repositorio Github	4
3.	Datos base	4
4.	Datos complementarios	5
5.	Posibles análisis a realizar	5
6.	Diseño conceptual	6
7.	Diseño lógico relacional	7
8.	Implementación y carga	8
9.	Script SQL	9
10.	Descarga en formato CSV	10
11.	Zeppelin	10
12.	Bibliografía	20

1. Introducción

Se busca desarrollará una solución para el análisis de datos reales sobre un problema de interés social, en este caso información turística del Ecuador.

Se usaron distintas herramientas para la exploración, visualización y manejo de la información, por ejemplo: MySQL Community Server, MySQL Workbench, XAMPP, Excel.

El recurso usado más destacable de los ya mencionados es MySQL Community Server, o simplemente conocido como MySQL. Es una base de datos muy popular entre los desarrolladores, su velocidad y pequeño tamaño la hacen ideal para un proyecto de desarrollo, si a esto se sumamos el hecho de que es de código abierto, es decir gratis.

Para realizar de manera rápida la creación de conexiones, workspaces, tablas y ejecución de consultas sql, se utilizó la herramienta MySQL Workbench.

Algo que mencionar es que es muy importante el proceso de población de la base de datos, ya que esto permitirá tener datos íntegros y no redundantes.

Este proyecto se realizó desde la semana 9 hasta la semana 14 con la guía y ayuda del Mtr. Encalada Encalada Angel Eduardo.

2. Repositorio Github

https://github.com/adpardo1/PROYECTO-INTEGRADOR-2022

3. Datos base

El archivo denominado "Catastro turístico de establecimientos a nivel nacional" se obtuvo del ministerio de turismo, con el siguiente enlace

https://servicios.turismo.gob.ec/descargas/Catastro-turistico/Consolidado-Nacional-2022-p ublico-1-web.xlsx, en el cual se encontraba con 22209 registros, y con campos como: nombre comercial, actividad, clasificación, categoría, provincia, cantón, parroquia, referencia dirección, dirección, teléfono principal, teléfono secundario, email, dirección web, población, luego se lo convirtió en .csv. Se analizo y limpio el archivo csv para evitar errores al ingresar datos, utilizamos herramientas como excel, sublimetxt, para corregir errores, se cambió la separación de ";" a "|", ya que en algunos campos los datos tenían ";" dentro del dato.

4. Datos complementarios

Al archivo principal se le añadieron 5 nuevos campos en los que se pueda realizar análisis estadísticos como: número de cantones, horario de atención, estimación de usuarios mensuales, idioma, número de población de cada provincia.

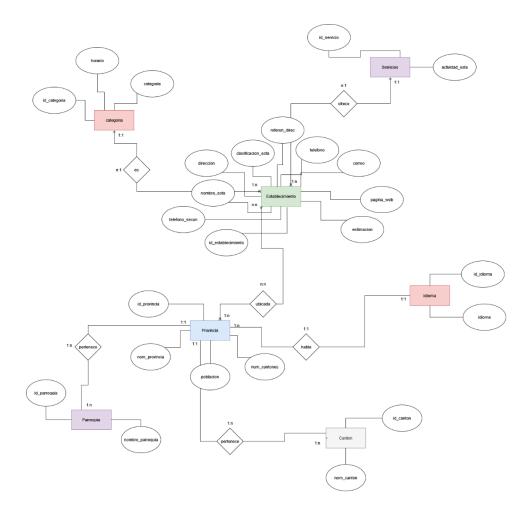
5. Posibles análisis a realizar

- Analizar qué comerciales, según su categoría, son los que tienen una asignación alta.
- Determinar el porcentaje de cada provincia donde los comerciales atienden las 24 horas.
- Predecir el crecimiento o decrecimiento en 5 años, de clientes de los comerciales,
 según la estimación de comensales mensual.
- Determinar el comercial más visitado basándose en la estimación de usuarios mensuales.
- Según el número de habitantes por provincia determinar que comercial sería el más visitado.

- Del número de cantones que existe por provincia determinar s existe mayor número de comerciales.
- Determinar qué porcentaje resulta de los comerciales que tienen su propio sitio web.
- Conocer el porcentaje de comerciales de primera categoría en un determinado cantón.

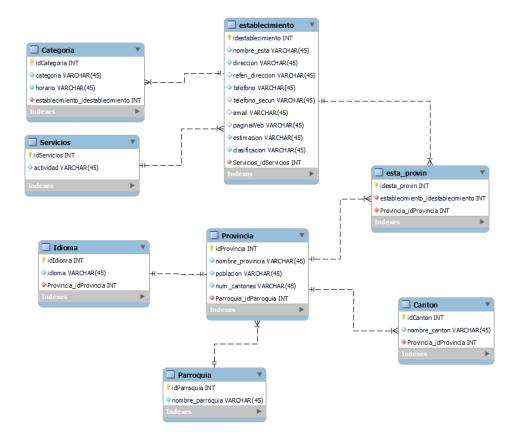
6. Diseño conceptual

Analizado el archivo database con los datos complementarios se determinaron 7 entidades como, establecimiento, provincia, cantón, parroquia, idioma, servicios, categoría, determinando también sus atributos y la relación entre tablas.



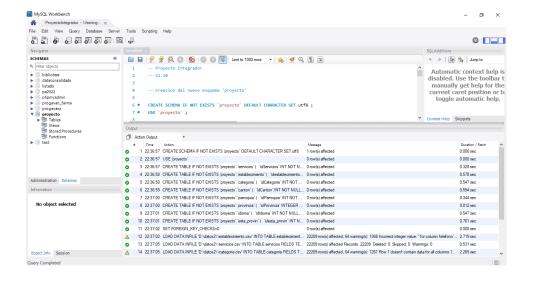
7. Diseño lógico relacional

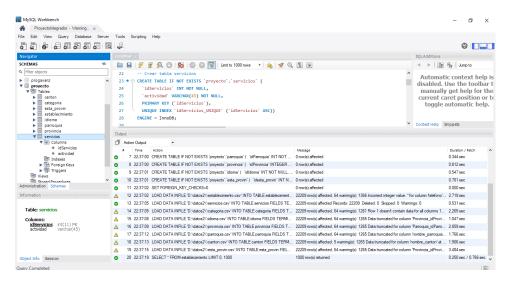
Como resultado de la multiplicidad del modelo logico se creó una nueva tabla, se determinó los tipos de datos de los atributos y su domino, así también claves primarias, foráneas, not null, campos obligatorios y únicos.

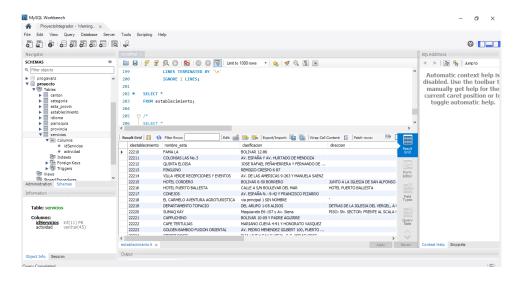


8. Implementación y carga

Para la carga de los datos, al csv general se lo divido en tablas creando nuevos csv para poder ingresar fácilmente los datos añadiendo las claves primarias y foráneas.







9. Script SQL

https://github.com/adpardo1/PROYECTO-INTEGRADOR-2022/tree/main/BASE%20DE %20DATOS/Script%20SQL/scripfinal.sql

10. Descarga en formato CSV

https://github.com/adpardo1/PROYECTO-INTEGRADOR-2022/tree/main/BASE%20DE%20DATOS/DATOS%20EN%20ARCHIVOS%20CSV

11. Zeppelin

Primeramente, se levanta el servidor de zeppelin usando el comando zeppelin-daemon.sh start

A la par se instala y se inicia el servidor de base de datos MySQL.

```
sudo apt-get install mysql-server mysql-client -y
mysql --version
sudo service mysql start
sudo mysql
```

Se ejecuta el archivo sql que contiene el código de creación (tablas, columnas, restricciones) y llenado de la base de datos.

```
mysql> source database.sql;
Query OK, 0 rows affected (0.01 sec)

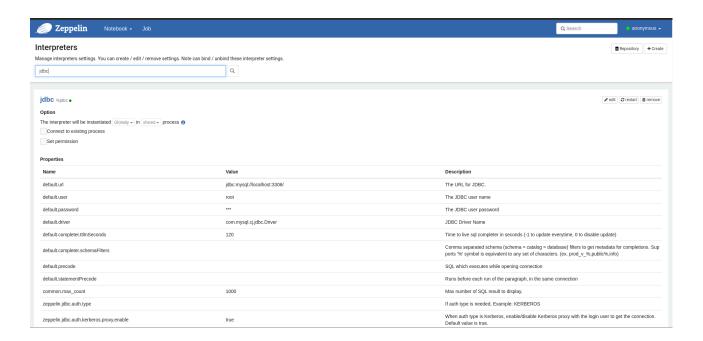
Query OK, 0 rows affected (0.00 sec)

Query OK, 1 row affected (0.00 sec)

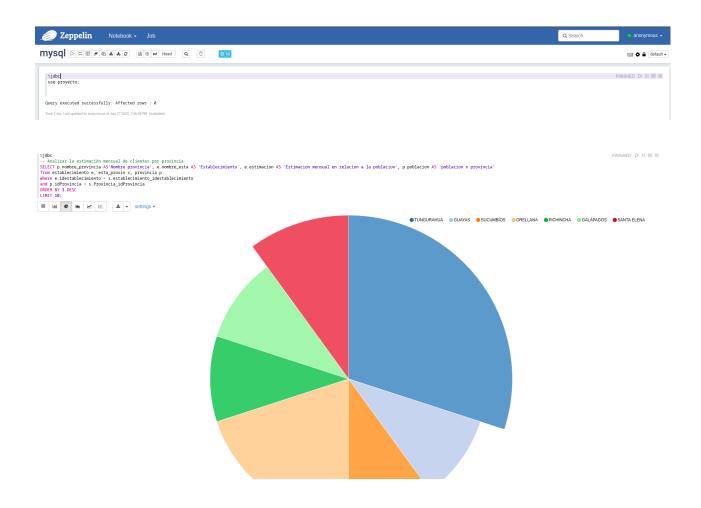
Database changed
```

```
| Comparison | Com
```

Se configura el intérprete jdbc para poder ejecutar consultas sql en el notebook de zeppelin.

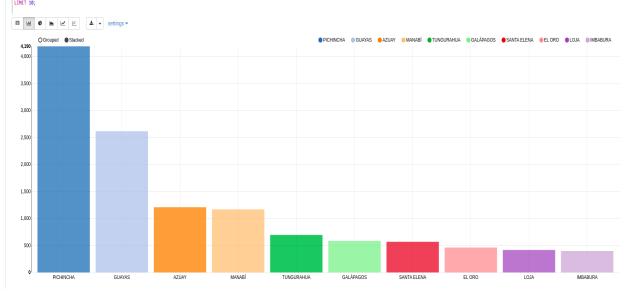


Ahora se puede ejecutar cualquier sentencia en lenguaje sql, anteponiendo la clave %jdbc antes de cada sentencia.

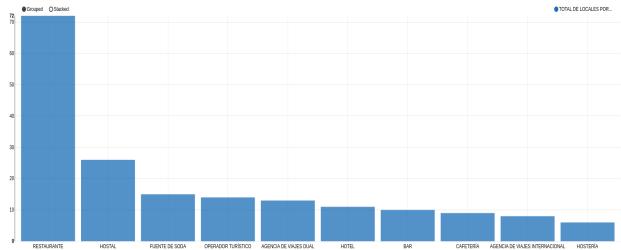












FINISHED ▷ X 10 @

**Sjöbc -- Mailzar qué comerciales, según su categoría, son los que tienen uma asignación de 5 estrellas.

***SELECT DISTINCT(p.nombre_provincia) "nombre provincia", p.num_cantones "número de cantones", count(t.nombre_esta) "numero de locales",p.población "población"

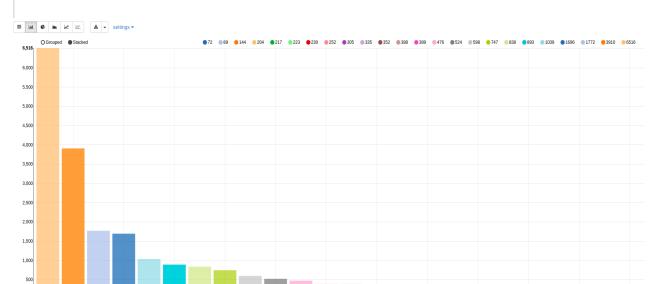
**FROM esta_provin e, establecimiento - t.idestablecimiento

**AU e.Provincia_idProvincia - p.idProvincia

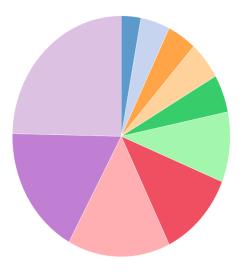
**GROUP 8" p.nombre_provincia

**GROUP 8" p.nombre_provincia

**ORDER 8" count(t.nombre_esta) DESC;



⊞ W © W W Settings ▼



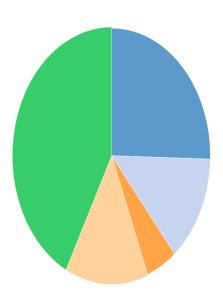
●BOLÍVAR ●LOS RÍOS ●PASTAZA ●ZAMORA CHINCHIPE ●CARCHI ●CAÑAR ●MORONA SANTIAGO ●SUCUMBÍOS ●ORELLANA ●EL ORO

FINISHED ▷ X ■ ◎

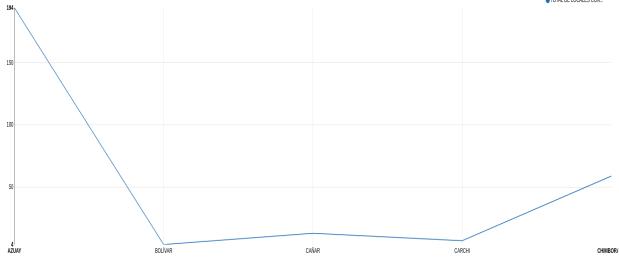
FINISHED ▷ X 10 @

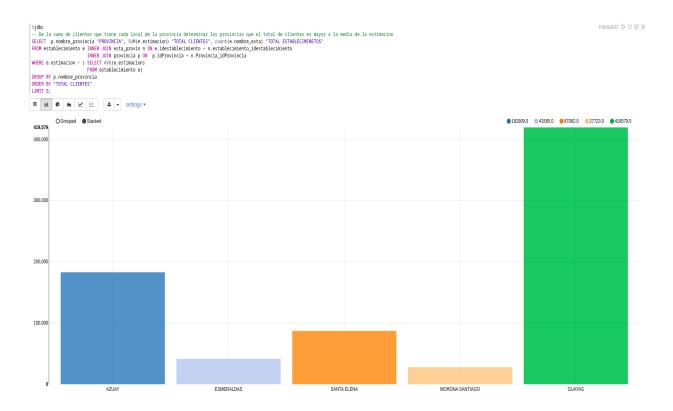
■ M C M M M settings ▼

●AZUAY ●ESMERALDAS ●SANTA ELENA ●PICHINCHA ●GUAYAS

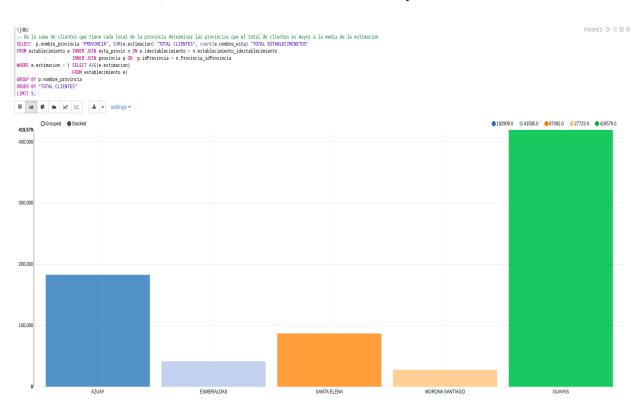


Njóbc
--- Determinar los comerciales que tienen su propio sitio web, amalizar si son los que tienen mayor numero de estimación de clientes.
SELECT p.nombre_provincia "PROVINCIA", count(e.paginakeb)"**iOTAL DE LOCALES CON PÁGINAMEB", e.estimación "ESTIMACIÓN USUARIOS"
FROM esta_provin v INNER JOIN establecimiento e ON v.establecimiento idestablecimiento e e.idestablecimiento
INNER JOIN provincia p ON v.Provincia_idProvincia = p.idProvincia
WHERE e.paginaMeb like "%.com%" FINISHED ▷ ※ 🗐 🐵 GROUP BY p.nombre_provincia
ORDER BY p.nombre_provincia
LIMIT 5; ■ W C W E settings ▼ ● TOTAL DE LOCALES CON... 194





En cuanto a el sitio web, se codificó un sencillo servidor node.js



Para mostrar los párrafos ejecutados del notebook de zeppelin, se codificó esté script que crea iframes con cada parrafo, de manera dinámica.

```
let iframesInfo = (
    "paragraph_1658435484758_1497429471"; "Analizar la estimación mensual de clientes por provincia.",
    "paragraph_1658265887899_13539352937"; "Conocer el porcentaje de comerciales que utilizan mas de un idioma por provincia.",
    "paragraph_16582654646_1248903284"; "Determinar el porcentaje de cada provincia donde los comerciales atienden las 24 horas.",
    "paragraph_165826428912892_1239361956"; "Proyección de la provincia com más número de personas en 1 año",
    "paragraph_1658263389916_9290452691"; "Analizar qué comerciales, según su categoria, son los que tienen una asignación de 5 estrellas.",
    "paragraph_16582631692_1220737041"; "Analizar qué comerciales, según su categoria, son los que tienen una asignación de 5 estrellas.",
    "paragraph_1658265260663_538832777": "Determinar el numero de locales que tienen pagina web y el numero de clientes mensuales."

);

window.addEventListener('load', function () {
    Object.entries(iframesInfo).fotEach(([k, v]] => {addIframe(k, v);})
    document.body.removeChild(document.body.getElementsByTagName("div")[0]);
));

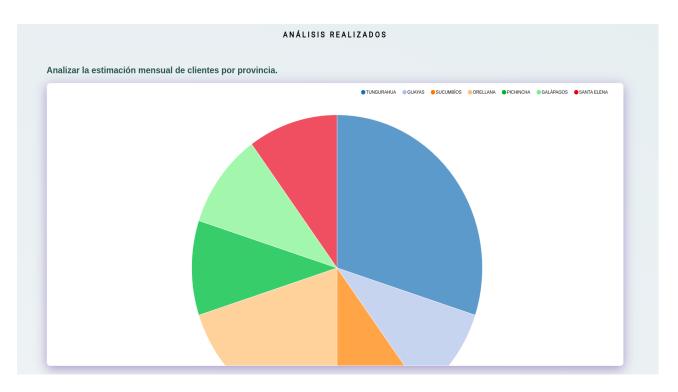
function addIframe(paragraph, text_title) {
    var title = document.createElement("hz")
    let text = title;
    title.appendChild(document.createElement("farame")
    iframe.scrolling = "no";
    iframe.scrolling iframe");

div.appendChild(title);
    div.appendChild(ititle);
    div.appendChild(div)
```

Por último, se levanta el servidor de node.

```
o gitpod /workspace/PA2022_scala/server (main) $ npm start
> server@1.0.0 start
> nodemon index.js

[nodemon] 2.0.19
[nodemon] to restart at any time, enter `rs`
[nodemon] watching path(s): *.*
[nodemon] watching extensions: js,mjs,json
[nodemon] starting `node index.js`
Example app listening on port 3000
```





12. Bibliografía

Hn, C. (2021, 6 abril). Las Ventajas de MySQL por Sobre Otras Bases de Datos.

Hosting Rápido con cPanel y certificado SSL en Chile | HN Datacenter. Recuperado

https://www.hn.cl/blog/las-ventajas-de-mysql-por-sobre-otras-bases-de-datos/#:%7

E:text=MySQL%20es%20una%20base%20de,queda%20muy%20clara%20su%20p
opularidad.

(n.d.). Ministerio de Gobierno – Responsable de velar por la gobernabilidad de la nación, la paz y seguridad ciudadana, el mantenimiento del régimen democrático y el irrestricto respeto a la Constitución y a los Derechos Humanos. Retrieved July 31, 2022, from https://www.ministeriodegobierno.gob.ec