



**UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA**

**La Universidad Católica De Loja**

**BASE DE DATOS AVANZADA**

**INTEGRANTES:**

Xavier Alexander Chávez Saraguro

Jhordan Kriss Bustamante Solano

Jose Cordova

**CARRERA:**

Computación

**DOCENTE:**

Ing. Eduardo Encalada

**AÑO LECTIVO:**

ABR/2022 – AGO-2022

## Tabla de contenido

<i>Introducción .....</i>	<i>3</i>
<i>Planificación del trabajo.....</i>	<i>4</i>
<i>Fuentes de datos.....</i>	<i>4</i>
<i>Componente Base de Datos .....</i>	<i>5</i>
Explicación de preprocesamiento de datos realizado .....	5
Diseño conceptual y diseño lógico de la base de datos .....	6
Implementación de la base de datos .....	7
<i>Componente Programación.....</i>	<i>8</i>
Explicación de herramientas utilizadas .....	8
Explicación de comandos y sentencias usadas para consultar y visualizar datos tanto a nivel del archivo fuente, como a nivel de base de datos MySQL. ....	9
<i>Resultados obtenidos.....</i>	<i>12</i>
Visualizaciones con su análisis e interpretación.....	12
<i>Conclusiones .....</i>	<i>15</i>
<i>Bibliografía .....</i>	<i>16</i>

## Introducción

En el presente informe se redactara sobre todo el proceso realizado para el componente del practicum 1.2 en cual es la unión de dos materias, programación avanzada y base de datos avanzada. El cual consistira en realizar un análisis de datos en los cuales se describe información turistica del Ecuador.

En donde lo primero tomamos datos del catastro turístico a nivel nacional tomando como datos un CSV brindados por nuestro profesor e implementando datos adiciones referentes al caso de estudio para esto usaremos los conocimientos adquiridos tanto de la materia base de datos avanzada y programación avanzada, así como las herramientas adecuadas para lograr el objetivo de integrar estas dos materias, por parte de base de datos usaremos MySQL la cual nos ayudara a manejar los datos adquiridos y relacionarlos entre si creando las tablas necesarias, en el caso de programación usaremos Zeppelin, entre otras herramientas, principalmente usaremos esta dado que nos permite trabajar con grandes bases de datos, pudiendo realizar consultas y modificar nuestro CSV lo cual nos es de gran ayuda al momento de realizar los distintos trabajos que requerimos, en el componente de base de datos primeramente recapitulamos los diferentes atributos, para así pensar los diferentes datos complementarios que se agregaron, para luego aplicarle las debidas reglas de normalizacion y arreglar la data de cada columna, para luego hacer el respectivo modelo conceptual y por consiguiente el modelo logico.

Seguido realizaremos el diseño fisico para el gesto de base de datos MySQL, donde se hizo el uso de la herramienta DataGrip.

Luego de dicha implementación se genera un CSV por cada tabla generada en nuestro modelo logico con los datos ya limpios.

## Planificación del trabajo

### Fuentes de datos

#### Datos base

Descripción de los datos	Proposito
Nombre Comercial	Se refiere al nombre comercial
Actividad	Se refiere a la actividad de cada comercial
Clasificación	Se refiere a la clasificación de la actividad
Categoría	Se refiere a la categoría de la clasificación
Provincia	Provincias del ecuador
Cantón	Cantones según las provincias
Parroquia	Parroquias según los cantones
Referencias de Dirección	Una referencia de la dirección del comercial
Dirección	Dirección exacta del comercial
Teléfono Principal	Teléfono principal del comercial
Teléfono Secundario	Teléfono secundario del comercial
Correo Electrónico	Correo del comercial
Dirección Web	Url de la página web del comercial

#### Datos complementarios

Descripción de los datos	Proposito	Fuente
Ventas	Este dato es con el proppsito de conocer las ventas por provincia y de todas las actividades que contiene el catastro, tambien se aprecia las recaudaciones por provincia.	<a href="https://servicios.turismo.gob.ec/visualizador-ventas">https://servicios.turismo.gob.ec/visualizador-ventas</a>

Provincias, cantones y parroquias	Este dato es con el proposito de registrar todas las zonas en donde se encuentra los catastros del Ecuador con su codigo de cada uno.	<a href="https://www.ecuadorencifras.gob.ec">https://www.ecuadorencifras.gob.ec</a>
Guías turísticas	Es dato es sobre cuantas guías turisticos abastece el ecuador por canton.	<a href="https://servicios.turismo.gob.ec/portfolio/registro-de-establecimientos-turisticos/licencia-guias-de-turismo-en-continente">https://servicios.turismo.gob.ec/portfolio/registro-de-establecimientos-turisticos/licencia-guias-de-turismo-en-continente</a>
Migrantes	Este dato es sobre todas las personas que entran y salen de nuestro país, cual es su motivo, de donde vienen y su edad.	<a href="https://www.ministeriodegobierno.gob.ec/registro-migratorio-ciudadanos-venezolanos-ecuador/">https://www.ministeriodegobierno.gob.ec/registro-migratorio-ciudadanos-venezolanos-ecuador/</a>
Países	Este dato es utilizado para ver de que país nos visitan mas con su codigo de país.	<a href="https://traduciresdescubrir.wordpress.com/2012/07/26/paises-y-abreviaturas/">https://traduciresdescubrir.wordpress.com/2012/07/26/paises-y-abreviaturas/</a>

## Componente Base de Datos

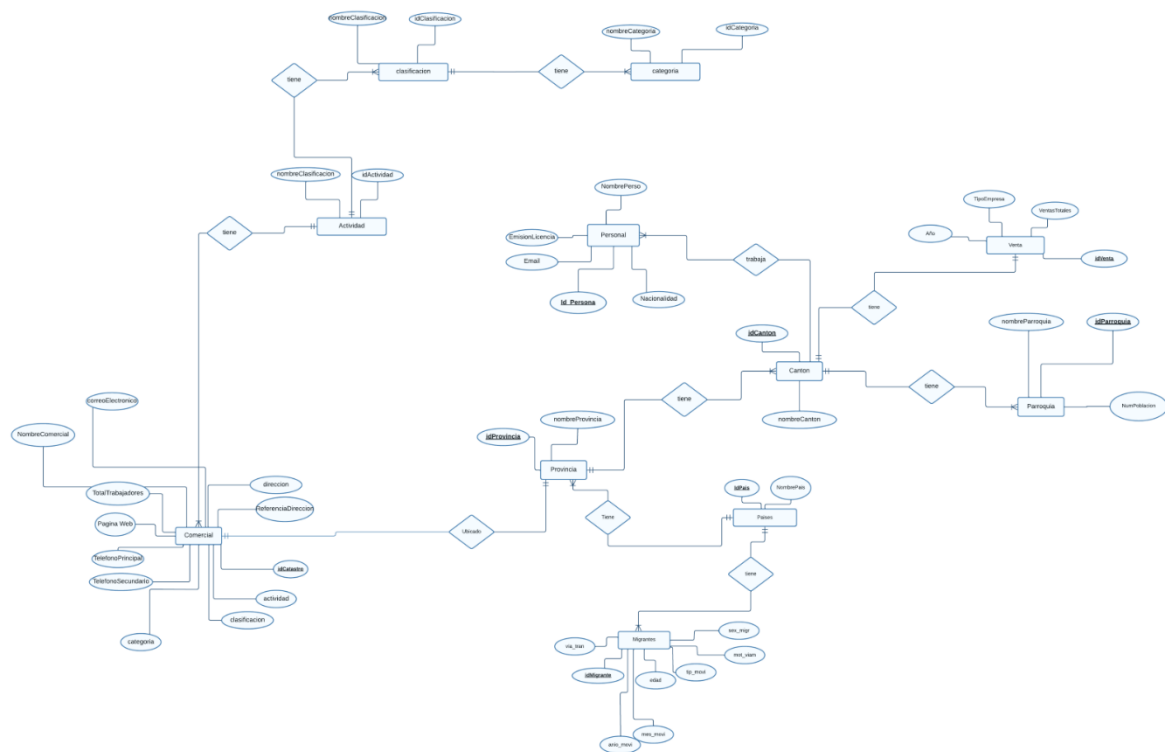
### Explicación de preprocesamiento de datos realizado

Para el preprocesamiento de datos utilizados primeramente se analizo la data enviado por el docente y los datos que nos gustaría complementar, además de esto se hizo una lluvia

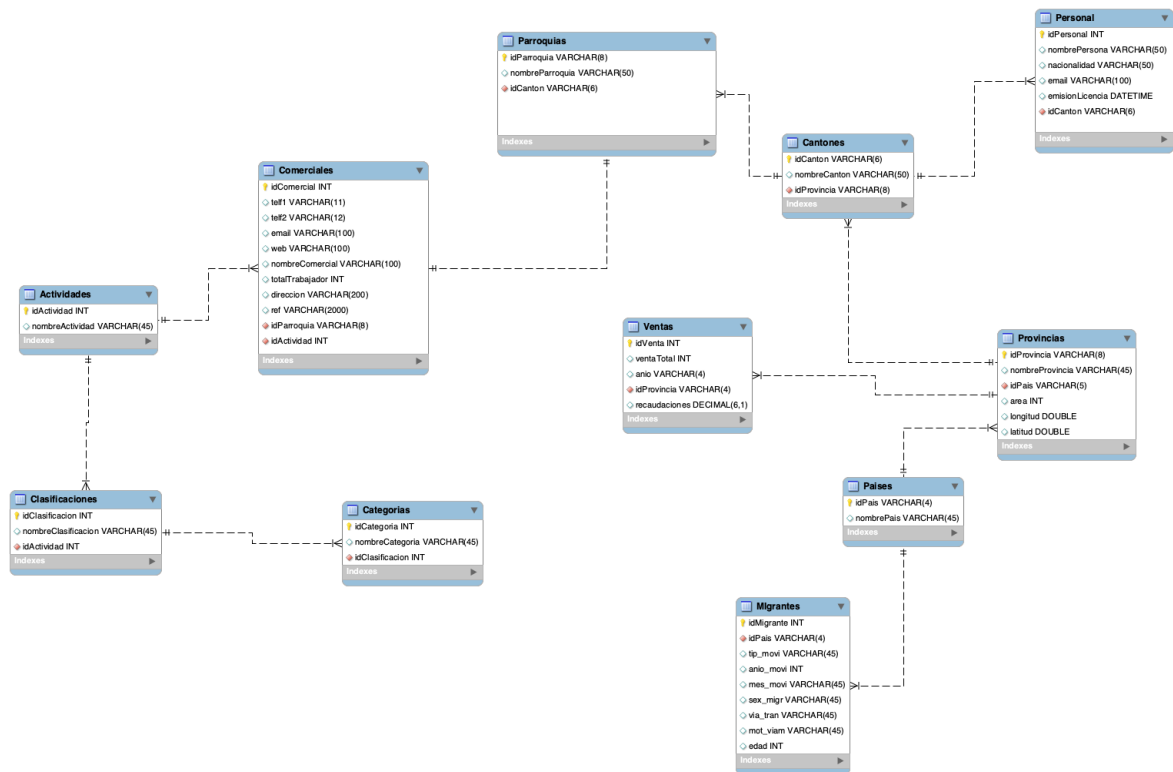
de ideas respecto a que se debe realizar con los datos relacionados al catastro turístico de establecimientos a nivel nacional, además de eso se busco diferentes datos complementarios para así poder integrarlos a la base de datos, después de obtener los datos complementarios se hizo el diseño conceptual y lógico relacional con estos y finalmente se cargo a las tablas para su preprocesamiento. Se usaron varias herramientas en este caso tablas en Excel para arreglar y extraer la data de una mejor manera.

## Diseño conceptual y diseño lógico de la base de datos

## Diseño conceptual



## Diseño lógico relacional



## Implementación de la base de datos

Para la implementación de datos se hizo mediante código utilizando la siguiente sintaxis INSERT INTO seguido del nombre de la tabla y sus atributos entre paréntesis VALUES los valores de cada atributo correspondiente igualmente entre paréntesis. En este proyecto manejamos una cantidad masiva de datos por lo tanto no podemos implementar los datos uno por uno así que con ayuda del comando load data infile buscamos el archivo CSV donde tenemos la información luego creamos las tablas necesarias y verificamos que los campos sean iguales para realizar la carga. Cabe aclarar que esto lo hicimos después de aplicar las técnicas de normalización. Una vez implementada la base de datos ya pudimos trabajar para hacer las respectivas consultas para el análisis de datos.

## Componente Programación

### Explicación de herramientas utilizadas

Para el apartado de la base de datos se usaron diferentes herramientas como lo son:

Mysql: “es un sistema de administración de bases de datos relacionales (RDBMS) desarrollado por Oracle que se basa en un lenguaje de consulta estructurado (SQL)”.

*(What Is MySQL? Everything You Need to Know | Talend, n.d.)*

Datagrip: es un entorno de gestión de bases de datos para desarrolladores. Está diseñado para consultar, crear y administrar bases de datos. Las bases de datos pueden funcionar localmente, en un servidor o en la nube. Admite MySQL, PostgreSQL, Microsoft SQL Server, Oracle y más. *(Introduction | DataGrip, n.d.)*

Excel: “Excel es un programa de hoja de cálculo de Microsoft y un componente de su grupo de productos Office para aplicaciones comerciales. Microsoft Excel permite a los usuarios formatear, organizar y calcular datos en una hoja de cálculo”. *(What Is Excel? Everything You Need to Know - Definition by WhatIs.Com, n.d.)*

Para el componente de Programacion se usaron herramientas como lo son:

Zeppelin: “El cuaderno Zeppelin brinda una manera fácil y directa de ejecutar código arbitrario en un cuaderno web. Puede ejecutar Scala, SQL e incluso programar un trabajo (a través de cron) para que se ejecute a intervalos regulares”. *(Bigdata - What Is Apache Zeppelin? - Stack Overflow, n.d.)*

Spark: Apache Spark es un marco de procesamiento de datos que puede realizar rápidamente tareas de procesamiento en conjuntos de datos muy grandes y también puede distribuir tareas de procesamiento de datos en varias computadoras, ya sea solo o en



conjunto con otras herramientas informáticas distribuidas. (*What Is Apache Spark? The Big Data Platform That Crushed Hadoop* | InfoWorld, n.d.)

CSS: CSS significa hojas de estilo en cascada. Es un lenguaje de hojas de estilo que se utiliza para describir la apariencia y el formato de un documento escrito en lenguaje de marcas. CSS fue desarrollado por W3C (World Wide Web Consortium) en 1996. (*What Is CSS* - Javatpoint, n.d.)

HTML: “es un acrónimo que significa lenguaje de marcado de hipertexto que se utiliza para crear páginas web y aplicaciones web”. (*What Is HTML* - Javatpoint, n.d.)

Explicación de comandos y sentencias usadas para consultar y visualizar datos tanto a nivel del archivo fuente, como a nivel de base de datos MySQL.

Los siguientes comandos :

1. ¿Cual es la cantidad de extranjeros que entran al país con respectó al grupo de edad al que corresponden de los diferentes años?

```
SELECT COUNT(*) as Cantidad, t.rango_edad, t.anio_movi
FROM(
  SELECT case
    when edad between 0 and 18 then '0 a 15'
    when edad between 19 and 40 then '19 a 40'
    when edad between 41 and 65 then '41 a 65'
    else 'mas de 66' end as rango_edad, edad, tip_movi, anio_movi
  FROM Migrantes
  WHERE tip_movi = 'Entrada' AND idPais != 'EC'
) as t
GROUP BY t.rango_edad, t.anio_movi
ORDER BY 3,2;
```

2. ¿Saldo migratorio turístico?: En esta consulta se puede observar la diferencia entre el número entradas y el número de salidas de personas hacia/desde un determinado

lugar, así determinando cual fue el año con mas entradas y salidas turísticas haciendo una comparación de estos

```
SELECT COUNT(*), anio_movi, tip_movi
FROM Migrantes
WHERE tip_movi = 'Entrada' AND mot_viam = 'Turismo'
GROUP BY anio_movi
UNION
SELECT COUNT(*), anio_movi, tip_movi
FROM Migrantes
WHERE tip_movi = 'Salida' AND mot_viam = 'Turismo'
GROUP BY anio_movi;
```

3. ¿Cuál es la cantidad de extranjeros que entran al país respecto a la nacionalidad a la que pertenecen?: En esta consulta se podrá observar el flujo de entrada de las diferentes personas extranjeras, comparándolas así por su lugar de procedencia.

```

(SELECT COUNT(*), anio_movi, P.nombrePais
FROM Migrantes m INNER JOIN Países P on m.idPais = P.idPais
WHERE tip_movi = 'Entrada' AND mot_viam = 'Turismo' AND m.idPais != 'EC'
AND anio_movi = 2019
GROUP BY P.nombrePais
ORDER BY 1 DESC
LIMIT 5)
UNION
(SELECT COUNT(*), anio_movi, P.nombrePais
FROM Migrantes m
INNER JOIN Países P on m.idPais = P.idPais
WHERE tip_movi = 'Entrada'
AND mot_viam = 'Turismo'
AND m.idPais != 'EC'
AND anio_movi = 2020
GROUP BY P.nombrePais
ORDER BY 1 DESC
LIMIT 5)
UNION
(SELECT COUNT(*), anio_movi, P.nombrePais
FROM Migrantes m
INNER JOIN Países P on m.idPais = P.idPais
WHERE tip_movi = 'Entrada'
AND mot_viam = 'Turismo'
AND m.idPais != 'EC'
AND anio_movi = 2021
GROUP BY P.nombrePais
ORDER BY 1 DESC
LIMIT 5);

```

4. ¿Cuál es la cantidad de guías turísticos de las diferentes provincias?: En esta consulta se hará referencia al número de guías turísticos por provincias, lo que nos dará la información referente a cuantos se encuentran ejerciendo este labor

```

SELECT COUNT(p.idPersonal) Guías, P2.nombreProvincia
FROM Personal p JOIN Cantones C on C.idCanton = p.idCanton
JOIN Provincias P2 on P2.idProvincia = C.idProvincia
GROUP BY P2.nombreProvincia
ORDER BY 1 DESC
LIMIT 5;

```

5. Comparativa de la venta total por años de las provincias con mayor incidencia de turismo: Esta consulta nos muestra el total de ventas por millones en cada provincia, comparando 3 años de ventas de las 5 provincias que reportan mayores ventas

```
SELECT ventatotal, anio, p.nombreProvincia
FROM Ventas v INNER JOIN Provincias p ON v.idProvincia = p.idProvincia
WHERE p.nombreProvincia = 'PICHINCHA' OR p.nombreProvincia =
'GUAYAS' OR p.nombreProvincia = 'GALAPAGOS' OR p.nombreProvincia
= 'AZUAY' OR p.nombreProvincia = 'TUNGURAHUA'
ORDER BY ventaTotal DESC;
```

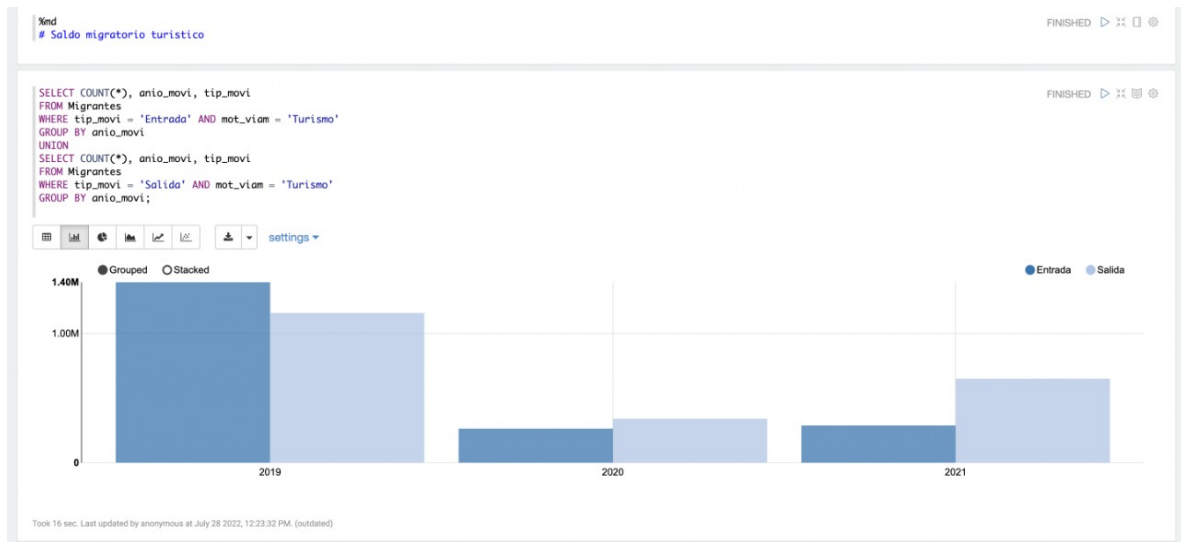
## Resultados obtenidos

### Visualizaciones con su análisis e interpretación

#### Consulta número 1:

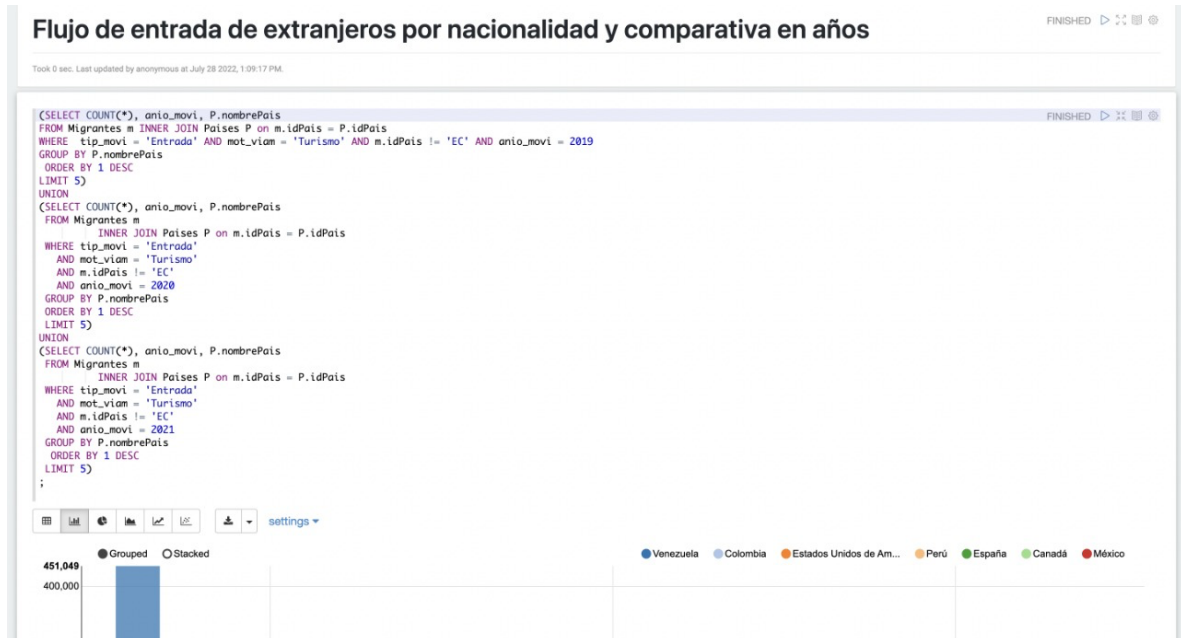


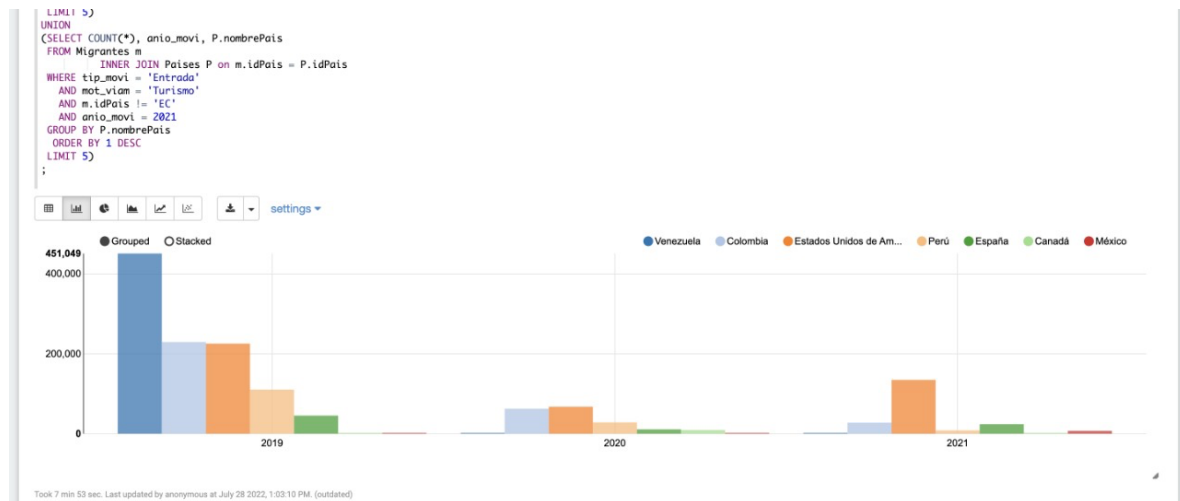
#### Consulta número 2:



Interpretación:

Consulta número 3: el saldo migratorio turístico hará referencia específicamente a

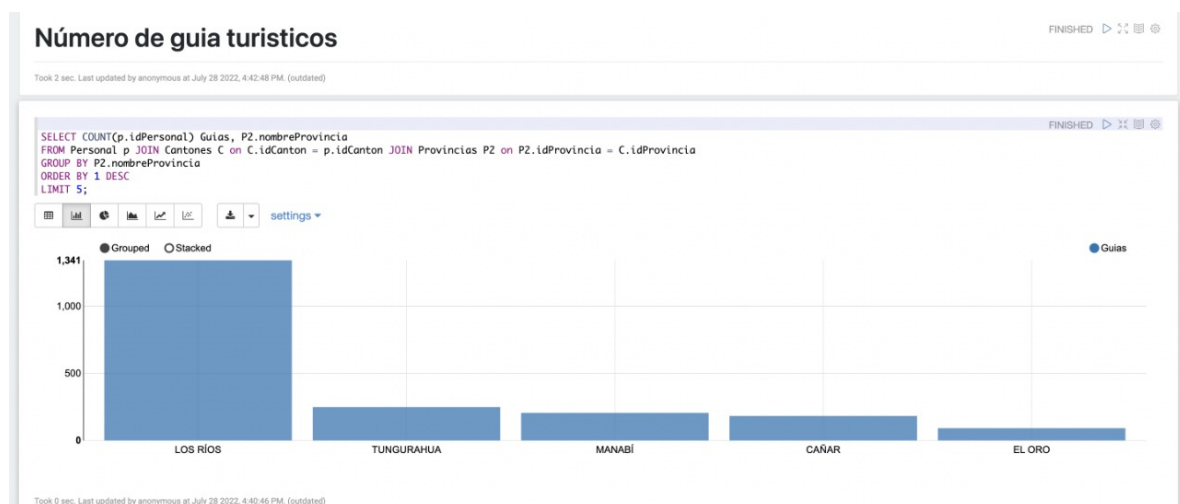




### Interpretación:

Se puede observar en esta consulta los datos de las diferentes cantidades de personas que han ingresado al país de una u otra manera, esta cantidad de personas estará especificada por la nacionalidad a la que pertenecen dándonos una idea de cuales son la mayor cantidad de personas que han ingresado y de donde provenían .

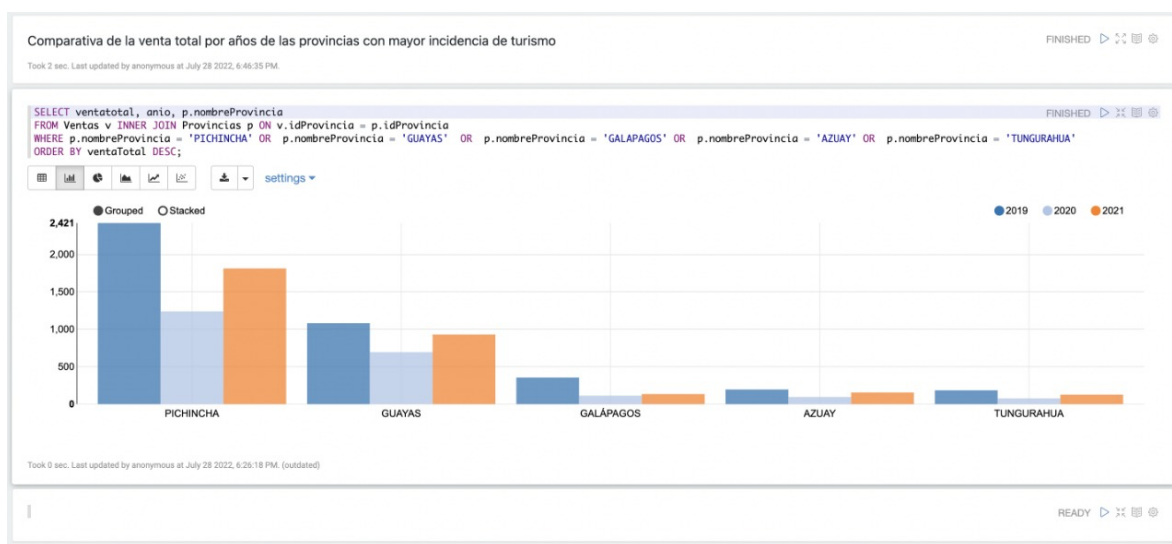
### Consulta número 4:



### Interpretación:

Se puede observar en esta consulta la cantidad de guías turísticos existentes por cada provincia dándonos una pequeña referencia de cuantas personas ejercen esta profesión de guías en los diferentes espacios de catastros turísticos dándonos como resultado los Rios con una cantidad de 1341 guías.

### Consulta número 5:



### Interpretación:

Se puede observar cómo alrededor de los diferentes años Pichincha se ha mantenido a la cabeza como una de las provincias con mayor número de incidencia de turismo llegando así a obtener más de la cuantiosa cantidad de 2.421 turistas, lo cual podría deberse a que es un provincia muy amplia con gran cantidad de habitantes, solo precedida por la provincia del guayas.

### Conclusiones

Como conclusión hemos obtenido que la diversa data ha sido de ayuda para la realización de las diferentes consultas propuestas, así sirviendo de referencia para futuras

investigaciones sobre nuestro país, pudiendo así obtener diferentes recursos con los cuales se podrían mejorar algunos aspectos, como lo son el control del turismo, poder obtener indicios sobre la población, información sobre migrantes entre otras muchas cosas, también se pudo aprender mas sobre el control y limpieza de datos como estos se relacionan de diferentes maneras además de esto, se fue manipulando la data así insertando atributos extras los cuales fueron obtenidos de fuentes confiables de datos.

## Bibliografía

*bigdata - What is apache zeppelin? - Stack Overflow.* (n.d.). Retrieved July 30, 2022, from

<https://stackoverflow.com/questions/37693383/what-is-apache-zeppelin>

*Introduction | DataGrip.* (n.d.). Retrieved July 30, 2022, from

<https://www.jetbrains.com/help/datagrip/meet-the-product.html>

*What is Apache Spark? The big data platform that crushed Hadoop | InfoWorld.* (n.d.).

Retrieved July 30, 2022, from [https://www.infoworld.com/article/3236869/what-is-](https://www.infoworld.com/article/3236869/what-is-apache-spark-the-big-data-platform-that-crushed-hadoop.html)

[apache-spark-the-big-data-platform-that-crushed-hadoop.html](https://www.infoworld.com/article/3236869/what-is-apache-spark-the-big-data-platform-that-crushed-hadoop.html)

*What is CSS - javatpoint.* (n.d.). Retrieved July 30, 2022, from

<https://www.javatpoint.com/what-is-css>

*What is Excel? Everything You Need to Know - Definition by WhatIs.com.* (n.d.). Retrieved

July 30, 2022, from

<https://www.techtarget.com/searchenterprisedesktop/definition/Excel>

*What is HTML - javatpoint.* (n.d.). Retrieved July 30, 2022, from

<https://www.javatpoint.com/what-is-html>

*What is MySQL? Everything You Need to Know | Talend.* (n.d.). Retrieved July 30, 2022,

from <https://www.talend.com/resources/what-is-mysql/>



