



## UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA

Proyecto Integrador

Carrera: Computación

Integrantes:

- Jhoel Alexander Ordoñez Coronel
- Jamil Erasmo Quituizaca Jaramillo
- Gerald Yhair Jaramillo Tejada

Periodo Académico:

Abril – Agosto 2022



## Índice

1.	Introducción .....	5
2.	Planificación.....	5
3.	Componentes de Base de Datos .....	5
3.1	Modelo Conceptual .....	8
3.2	Modelo lógico.....	9
	Implementación de la base de datos en MySQL .....	10
4.	Componentes de Programación.....	11
4.1	Limpieza de Datos:.....	11
4.2	Consultas Spark:.....	15
4.3	Formulario .....	16
4.4	Unión de dataframes.....	18
4.5	Consultas SQL .....	20
4.6	Página Web .....	24
5.	Conclusiones .....	26
6.	Bibliografía.....	27

Figura 1	Modelo Conceptual .....	8
Figura 2	Modelo logico .....	9
Figura 3	Creacion de los inserts .....	10
Figura 4	Creacion del script .....	10
Figura 5	Inserción de los datos.....	11
Figura 6	Nuevo dataframe con el cambio de nombre de las columnas.....	12
Figura 7	Cosultas que se realizaron para saber los datos de la columna categoria según la subclasificación siempre y cuando sea igual a la variable .....	12
Figura 8	Creación del nuevo dataframe en el cual se realizó la limpieza de los datos y se le asignó un solo dato.....	14
Figura 9	Comprobación de que los datos de la columna categoria se han cambiado .....	14
Figura 10	Gráfica de la consulta 1 .....	15
Figura 11	Gráfica de la consulta 2.....	15
Figura 12	Gráfica de la consulta 3.....	16
Figura 13	Gráfica de la consulta 4.....	16
Figura 14	Creación de dataframe seleccionando las columnas de Provincia y subclasificación. Y realizando un conteo de las columnas agrupadas.....	17
Figura 15	Importación e implementación de Window. ....	17
Figura 16	Declaración de variables que se mostraran en el formulario y relacionamiento de las mismas, además de la consulta que va a ejecutar. ....	18
Figura 17	Imagen de ejemplo de la ejecución del formulario .....	18

Figura 18	Imagen de la creación de un nuevo dataframe que almacena los datos de la variable que vamos a unir .....	19
Figura 19	Uso del inner join para la unión de los dataframe.....	19
Figura 20	Consulta con el dataframe Unido 1 .....	19
Figura 21	Consulta con el dataframe Unido 1 .....	20
Figura 22	Consulta de Sql 1.....	21
Figura 23	Consulta de Sql 2.....	21
Figura 24	Consulta de Sql 3.....	22
Figura 25	Consulta de Sql 4.....	22
Figura 26	Consulta de Sql 5.....	23
Figura 27	Consulta de Sql 6.....	24
Figura 28	Codigo HTML de la pagina Web.....	25
Figura 29	Opcion Zeppelin.....	25
Figura 30	Pagina Web .....	26

## **1. Introducción**

En el proyecto integrador hemos aplicado conocimientos tanto de la materia de base de datos avanzada como de programación avanzada, con la realización de este proyecto buscamos realizar una práctica conjunta de los conocimientos que hemos adquirido dentro del presente semestre, para poder llevar a cabo la elaboración del mismo hemos usado diferentes herramientas como: el software de MySQL que es un sistema de gestión de base de datos es un software libre que se engloba licencias GNU GPL etc. Una de las principales ventajas es que su lectura es muy rápida. Por lo que debemos tener en cuenta la importancia de la socialización de resultados, la implementación y optimización para así poder aplicar con nuestros conocimientos adquiridos y junto a la investigación tanto de base de datos avanzada como de programación avanzada aportar al proyecto integrador.

## **2. Planificación**

Para el presente proyecto integrador aplicamos procesos y estrategias que nos permitan resolver los problemas planteados ya que ante todo tuvimos que analizar la situación de ambos componentes tanto de base de datos como de programación para el cual aplicamos objetivos a lo largo del mismo, así como planes para poder lograr detalle a detalle.

## **3 Componentes de Base de Datos**

Para iniciar con la elaboración del proyecto integrador el Ing. Ángel Eduardo Encala Encalada nos facilitó un archivo Excel, dentro del cual pudimos encontrar los datos base con los que pudimos empezar a elaborar el trabajo, detallando cada uno de los datos que pudimos encontrar en el mismo tenemos: El nombre de cada centro comercial o nombre del lugar de comercio, a los cuales les asignamos su respectivo id, la clasificación de cada uno de los mismos. Dentro de lo que sería actividades nos referimos a una clasificación de alojamiento u otros ítems más que

relacionamos con el tipo de establecimiento que tenemos, de igual manera nos encontramos con su dirección números de teléfonos tanto principal como secundario, un correo electrónico además de un sitio web para poder ser contactados que está presente en cierto locales.

Encontramos El nombre de cada una de las 24 provincias de nivel nacional que de igual manera se les asigno una id. Así mismo encontramos cada uno de los 209 cantones que conforman el territorio del Ecuador, y complementándolos con las 637 parroquias correspondientes a nivel nacional.

Al ya saber que datos nos otorgaron hicimos una lluvia de ideas la cual nos permitirá crear nuevas varias para agregarlas al Excel.

<b>Lluvia de ideas</b>
La media del costo que tiene los hoteles de cada provincia
Porcentaje de lugares turísticos en donde se pueda acampar
Numero de hoteles que hay en cuenca
Porcentaje de estrellas de los hoteles en cada ciudad
Numero de habitantes por provincia (mujeres)

Por medio de esta lluvia de ideas logramos agregar 5 variables mas las cuales serian los costos de cada uno de los hoteles que hay a nivel nacional, los cuales se obtuvo gracias a la clasificación de cada uno de los locales, dentro los cantones pudimos localizar cada uno de los aeropuertos que hay en cada cantón, dentro de la provincia pudimos sacar los lugares menos visitados y los mejores lugares para acampar y por último sacamos la población de cada cantón a nivel nacional. A estas variables nuevas les pusimos los siguientes nombres.

<b>Aeropuerto</b>	<b>Campamento</b>	<b>Lugares_m_visitados</b>	<b>Población</b>	<b>Costo_hotel</b>
Nombres de los aeropuertos que tienen los cantones a nivel nacional.	Lugares destinados para poder acampar en cada provincia.	Lugares menos visitados a nivel provincial.	Numero de población que se encuentra en cada cantón.	Costo de cada hotel.

Para obtener los datos de las nuevas variables tuvimos que consultar por internet e ir las agregando en el Excel que se nos otorgó. Con los datos recolectados de las nuevas variables podemos hacer estadísticas con la finalidad de ayudar al turista a tener una mejor experiencia dentro de nuestro país, con estas variables nuevas los turistas pueden saber cuántos aeropuertos hay que cada cantón y saber a donde deben ir para realizar su viaje, también pueden saber que hotel es más barato por cada cantón para poder ahorrar en su alojamiento.

A continuación, realizamos el modelo conceptual el cual nos permitirá poder gestionar mejor los datos, primero normalizamos los datos en el cual creamos las siguientes tablas:

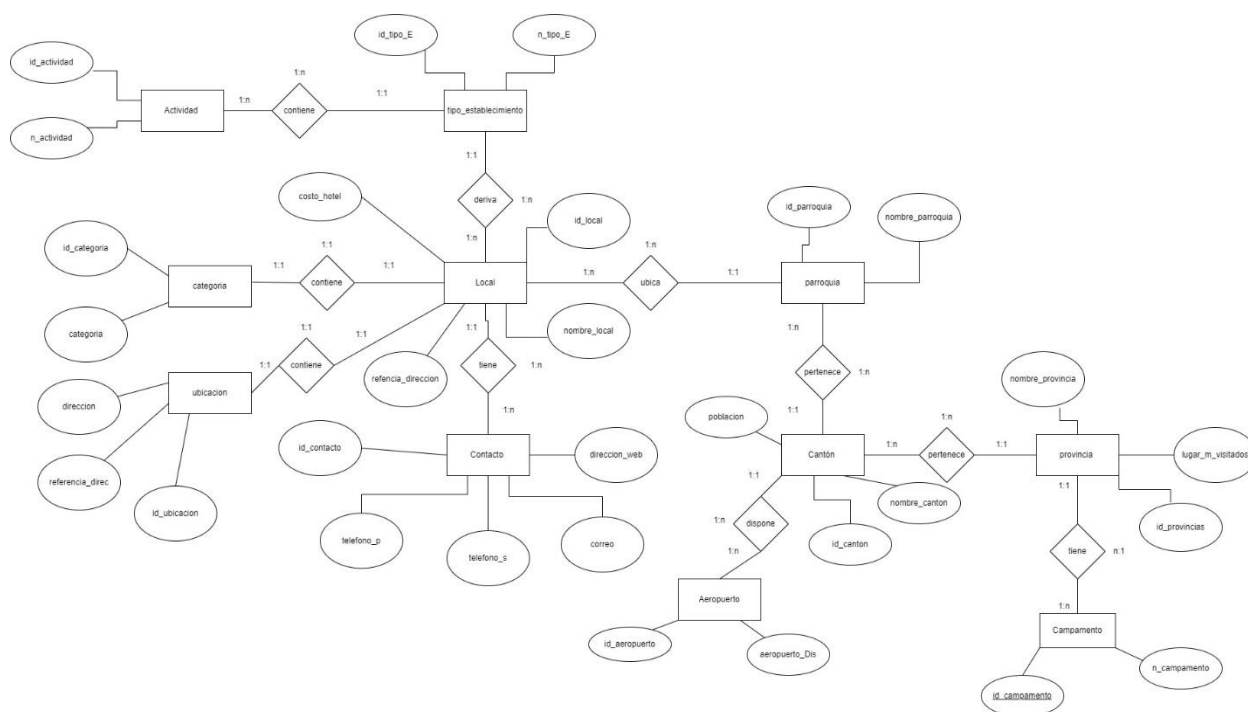
- Actividad
- Tipo\_establecimiento
- Categoria
- Ubicacion
- Contacto
- Campamento
- Provincia
- Aeropuerto

- Cantón
- Parroquia
- Local

En el caso de las nuevas variables que integramos las asignamos de la siguiente forma, aeropuerto los pusimos como una entidad que se relaciona con los cantones, igualmente con campamentos el cual está relacionado con las provincias, los lugares menos visitados lo pusimos como un atributo en la tabla provincia, la población también esta como atributo en la tabla cantón y por ultimo el costo de hoteles el cual es un atributo de la tabla local.

### 3.1 Modelo Conceptual

Figura 1 Modelo Conceptual

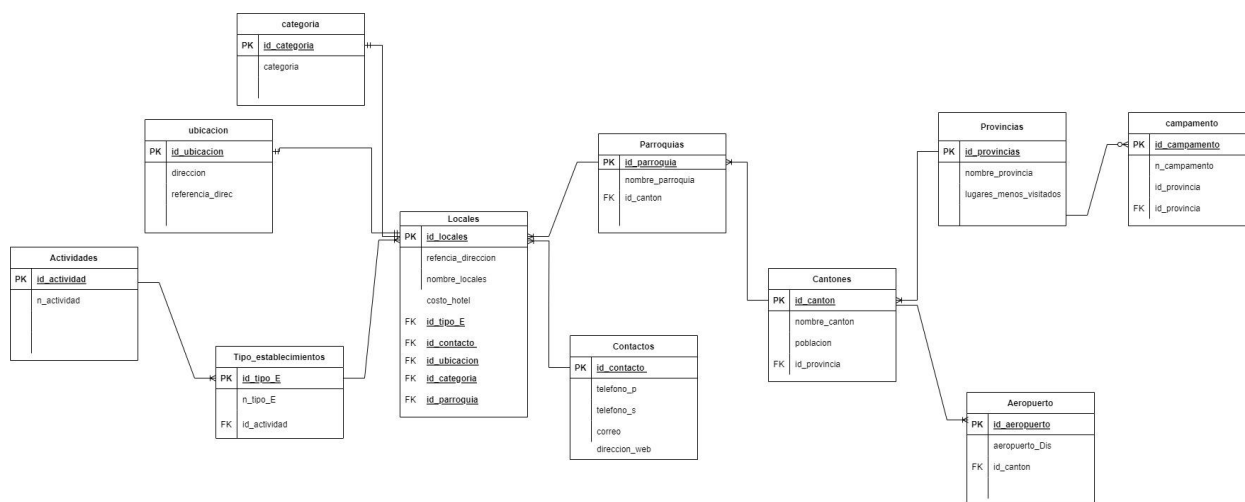


Al tener el modelo concepto procedemos a realizar el modelo lógico en el cual ya podremos las claves primarias y claves foráneas en las respectivas tablas.



### 3.2 Modelo lógico

Figura 2 Modelo logico



A continuación, empezaremos ya con la implementación y carga de los datos, dentro de algunas tablas encontramos datos que contenían algún tipo de carácter que lo hacía erróneo y no solo caracteres, sino también espacios en blanco y demás, que nos creaban conflictos en el momento de realizar la carga de estos datos dentro de la base de datos. Para poder cargar dichos datos dentro hicimos la creación de los inserts utilizando las propias herramientas o sentencias que nos brinda Excel en este caso utilizamos “=CONCAT("insert into (nombre de la tabla) values ""&A2&""", ""&B2&""", ""&C2&""");”)”. Posteriormente con los inserts creados con los datos ya limpios procedimos a insertarlos en el orden correcto por las llaves foráneas.

## Implementación de la base de datos en MySQL

Figura 3 Creacion de los inserts

	A	B	C	D
	id_aeropuerto	aeropuerto_Dis	id_caton	Columna1
1	Arp-1	Aeropuerto Coronel Edmundo Carvajal	1-Ct	=CONCAT("insert into aeropuerto value('"&A2&"', ""&B2&"', ""&C2&"',");")
2	Arp-2	Aeropuerto Internacional Mariscal Sucre - Aeropuerto Tababela	4-Ct	insert into aeropuerto value("Arp-2", "Aeropuerto Internacional Mariscal Sucre - Aeropuerto Tababela", "4-Ct");
3	Arp-3	Aeropuerto Internacional José Joaquín de Olmedo	5-Ct	insert into aeropuerto value("Arp-3", "Aeropuerto Internacional José Joaquín de Olmedo", "5-Ct");
4	Arp-4	Aeropuerto de Tena-Jumandy	7-Ct	insert into aeropuerto value("Arp-4", "Aeropuerto de Tena-Jumandy", "7-Ct");
5	Arp-5	Aeropuerto Internacional Eloy Alfaro	28-Ct	insert into aeropuerto value("Arp-5", "Aeropuerto Internacional Eloy Alfaro", "28-Ct");
6	Arp-6	Aeropuerto Regional Chachoán -Aeropuerto Izamba	42-Ct	insert into aeropuerto value("Arp-6", "Aeropuerto Regional Chachoán -Aeropuerto Izamba", "42-Ct");
7	Arp-7	Aeropuerto De Latacunga	70-Ct	insert into aeropuerto value("Arp-7", "Aeropuerto De Latacunga", "70-Ct");
8	Arp-8	Aeropuerto Los Perales	86-Ct	insert into aeropuerto value("Arp-8", "Aeropuerto Los Perales", "86-Ct");
9	Arp-9	Aeropuerto Internacional Coronel Carlos Concha Torres	93-Ct	insert into aeropuerto value("Arp-9", "Aeropuerto Internacional Coronel Carlos Concha Torres", "93-Ct");
10	Arp-10	Aeropuerto Ciudad de Catamayo	100-Ct	insert into aeropuerto value("Arp-10", "Aeropuerto Ciudad de Catamayo", "100-Ct");
11	Arp-11	Aeropuerto Regional de Santa Rosa	133-Ct	insert into aeropuerto value("Arp-11", "Aeropuerto Regional de Santa Rosa", "133-Ct");
12	Arp-12	Aeropuerto José María Velasco Ibarra	141-Ct	insert into aeropuerto value("Arp-12", "Aeropuerto José María Velasco Ibarra", "141-Ct");
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				

Figura 4 Creacion del script

**pbintegrador**

**Tables**

- actividad
- aeropuerto
- campamento
- canton
- categoria
- contacto
- local
- parroquia
- provincia
- tipo\_e

```

13
14 • CREATE TABLE actividad (
15     id_actividad VARCHAR(12) NOT NULL,
16     n_actividad VARCHAR(200)
17 );
18
19 • ALTER TABLE actividad ADD CONSTRAINT actividad_pk
20
21 • CREATE TABLE aeropuerto (
22     id_aeropuerto VARCHAR(15) NOT NULL,
23     aeropuerto_dis VARCHAR(200),
24     id_canton VARCHAR(15) NOT NULL
25 );
26
27 • ALTER TABLE aeropuerto ADD CONSTRAINT aeropuerto_f
28
29 • CREATE TABLE campamento (
30     id_campamento VARCHAR(15) NOT NULL,
31     n_campamento VARCHAR(200)
32 );

```

Figura 5 Inserción de los datos



```

1  insert into aeropuerto values ("Arp-1", "Aeropuerto Coronel Edmundo Carvajal", "1-Ct");
2  insert into aeropuerto values ("Arp-2", "Aeropuerto Internacional Mariscal Sucre - Aeropuerto Tababela", "4-Ct");
3  insert into aeropuerto values ("Arp-3", "Aeropuerto Internacional José Joaquín de Olmedo", "5-Ct");
4  insert into aeropuerto values ("Arp-4", "Aeropuerto de Tena-Jumandy", "7-Ct");
5  insert into aeropuerto values ("Arp-5", "Aeropuerto Internacional Eloy Alfaro", "28-Ct");
6  insert into aeropuerto values ("Arp-6", "Aeropuerto Regional Chachoán -Aeropuerto Izamba", "42-Ct");
7  insert into aeropuerto values ("Arp-7", "Aeropuerto De Latacunga", "70-Ct");
8  insert into aeropuerto values ("Arp-8", "Aeropuerto Los Perales", "86-Ct");
9  insert into aeropuerto values ("Arp-9", "Aeropuerto Internacional Coronel Carlos Concha Torres", "93-Ct");
10 insert into aeropuerto values ("Arp-10", "Aeropuerto Ciudad de Catamayo", "100-Ct");
11 insert into aeropuerto values ("Arp-11", "Aeropuerto Regional de Santa Rosa", "133-Ct");
12 insert into aeropuerto values ("Arp-12", "Aeropuerto José María Velasco Ibarra", "141-Ct");

```

### 3. Componentes de Programación

Al igual que en el componente de base datos para el componente de programación hemos utilizado los mismos datos bases entregados por el Ing. Angel Eduardo Encalada Encalada. Para este complemento tuvimos que utilizar herramientas de gitpod para poder utilizar zeppelin ya que en nuestro caso nuestro computador tiene sistema operativo windows y no es posible utilizar esta herramienta en nuestro sistema operativo, ya que solo está disponible para linux y Mac, para poder ser usado en gitpod utilizamos los lenguajes de programación java en su versión 8.0.302-open, scala en su versión 2.12.15 y spark 3.2.1

#### 4.1 Limpieza de Datos:

Una vez ya abierta la conexión en nuestro puerto hicimos la creación de nuestro dataframe en el cual ubicamos todos los datos que se encuentran en el archivo excel, como ya es de conocimiento y como fue mencionado anteriormente existen datos en algunas columnas en donde hay aglomeración por lo cual tuvimos que realizar algunas limpiezas, aunque antes de realizar esto fue necesario cambiar los nombres con la sentencia `.withColumnRenamed("", "")` y

en el caso de `.withColumn` (""), `trim` ("") realizamos la eliminación de espacios dentro de los datos de cada columna, al igual habían nombres de columnas que se encontraban mal escritos y además esto nos favorecía para el momento de trabajar con las tablas.

El cambio de nombre de las columnas:

Figura 6 Nuevo dataframe con el cambio de nombre de las columnas

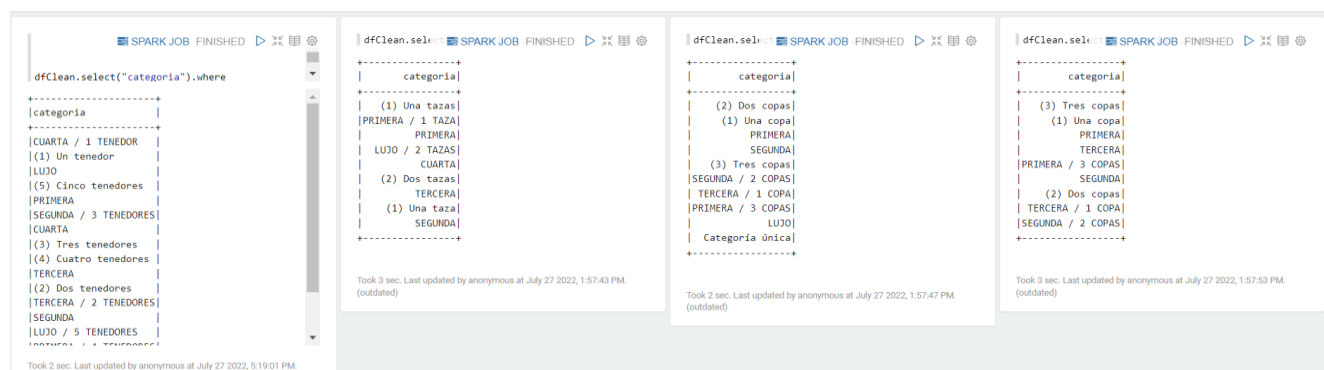
```
//Cambio de nombre a las columnas
val dfClean = consolidado
  .withColumnRenamed("Nombre Comercial", "nombre")
  .withColumnRenamed("Actividad", "clasificacion")
  .withColumnRenamed("Casificación", "subclasificacion")
  .withColumnRenamed("Categoría", "categoria")
  .withColumn("provincia", trim($"Provincia"))
  .withColumnRenamed("Cantón", "canton")
  .withColumn("canton", trim($"canton"))
  .withColumnRenamed("Parroquia", "parroquia")
  .withColumn("parroquia", trim($"parroquia"))
  .withColumnRenamed("Dirección", "direccion")
  .withColumn("direccion", trim($"direccion"))
  .withColumnRenamed("Referencia de Dirección", "dir_referencia")
  .withColumnRenamed("Teléfono Principal", "telf_principal")
  .withColumnRenamed("Teléfono Secundario", "telf_secundario")
  .withColumnRenamed("Correo Electrónico", "email")
  .withColumnRenamed("Dirección Web", "web")

dfClean: org.apache.spark.sql.DataFrame = [nombre: string, clasificacion: string ... 11 more fields]
```

Para realizar la limpieza de los datos dentro de las columnas encontramos una en particular que sería la columna de categoría relacionando la clasificación y subclasificación por lo cual tuvimos que primero hacer diferentes consultas de cada una de la subclasificación existentes para poder luego realizar los diferentes cambios dentro de los datos de esta columna.

Ejemplo de consultas realizadas antes de la limpieza:

Figura 7 Consultas que se realizaron para saber los datos de la columna categoría según la subclasificación siempre y cuando sea igual a la variable



- RESTAURANTE
- CAFETERÍA
- DISCOTECA
- BAR
- ALOJAMIENTO
- SALA DE RECEPCIONES Y BANQUETES
- FUENTE DE SODA
- HOTEL
- HOSTERÍA
- CASA DE HUÉSPEDES
- HOSTAL

De los cuales escogimos aquellos que tengan los datos con redundancia o sean repetitivos, dándoles así un único valor, para el cual creamos un nuevo dataframe en donde realizamos la limpieza de los datos seleccionados que eran: RESTAURANTE, CAFETERIA, BAR, DISCOTECA, ALOJAMIENTO, HOSTERIA, HOTEL Y HOSTAL.

Ejemplo de asignación de valores:

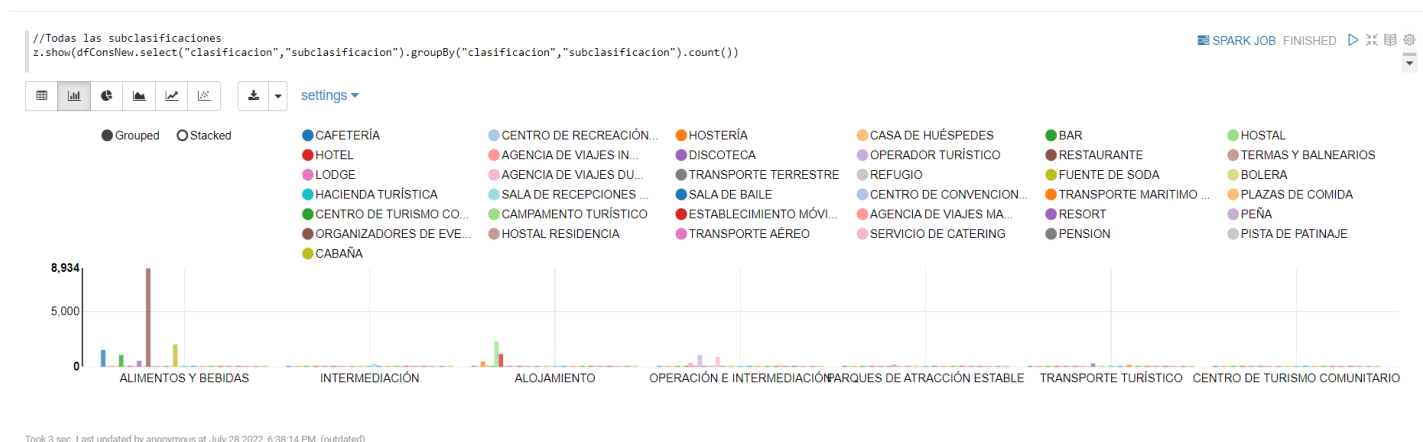


## 4.2 Consultas Spark:

Una vez que ya hemos realizado la limpieza de nuestros datos dentro del nuevo dataframe procedimos a realizar 4 consultas de nuestros datos para tener conocimiento de los mismos, a continuación, vamos a presentar cada una de las consultas.

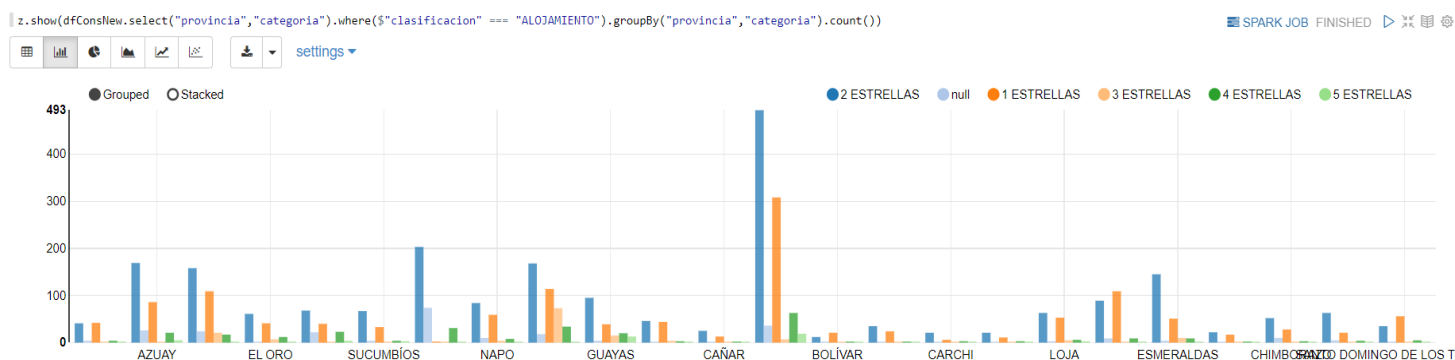
- 1) Cantidad de comercios que hay en subclasificaciones por cada clasificación.

Figura 10 Gráfica de la consulta 1



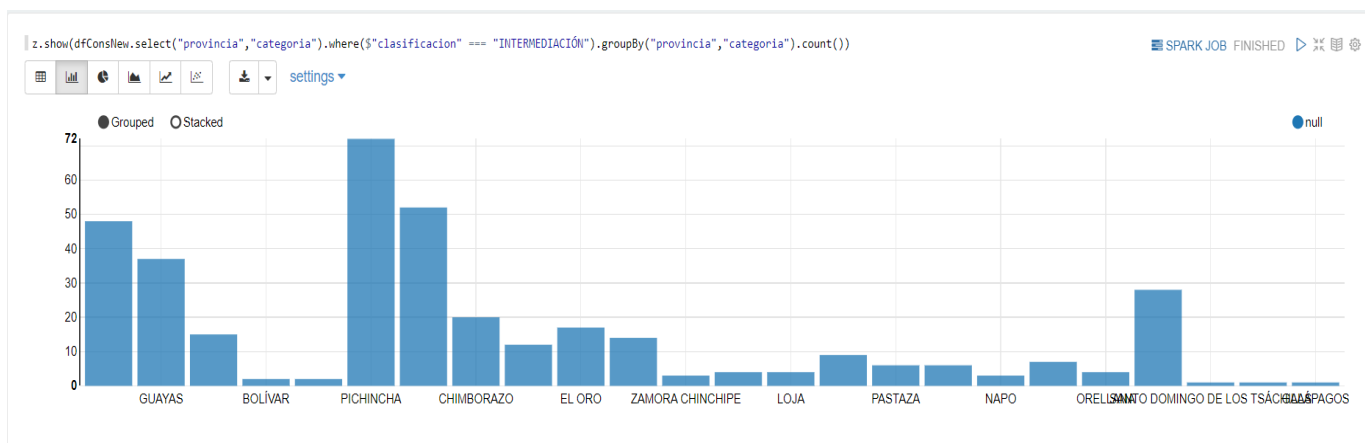
- 2) Cantidad de hoteles que hay por categoría en cada provincia.

Figura 11 Gráfica de la consulta 2



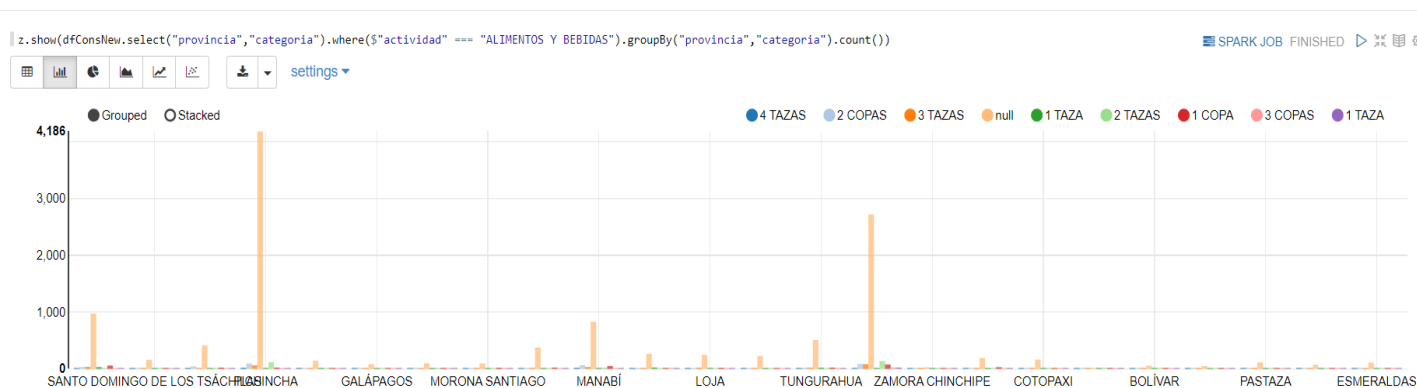
### 3) Cantidad de clasificación de la categoría de intermediación por provincias

Figura 12 Gráfica de la consulta 3



### 4) Cantidad de categorías que hay en alimentos y bebidas por cada provincia

Figura 13 Gráfica de la consulta 4



## 4.3 Formulario

Para la elaboración del formulario tuvimos que primero crear un nuevo dataframe que contenga solo dos columnas, a estas las agrupamos y realizamos un conteo de las mismas.



Figura 14 Creación de dataframe seleccionando las columnas de Provincia y subclasificación. Y realizando un conteo de las columnas agrupadas

```
val form = dfConsNew.select($"Provincia", $"subclasificacion").groupBy($"Provincia", $"subclasificacion").count()
form: org.apache.spark.sql.DataFrame = [Provincia: string, subclasificacion: string ... 1 more field]
```

FINISHED

Como segundo paso tuvimos que importar Window el cual nos va a ayudar dentro de este proceso, con el cual vamos a realizar una particion de subclasificaciones y a este lo vamos a ordenar de forma ascendente

Figura 15 Importación e implementación de Window.

```
import org.apache.spark.sql.expressions.Window
val fr = Window.partitionBy($"subclasificacion").orderBy(asc("count"))
val frR = form.withColumn("rank", row_number().over(fr))
frR.show(200, false)
```

SPARK JOB FINISHED

Provincia	subclasificacion	count	rank
BOLÍVAR	AGENCIA DE VIAJES DUAL	1	1
CARCHI	AGENCIA DE VIAJES DUAL	1	2
NAPO	AGENCIA DE VIAJES DUAL	2	3
ORELLANA	AGENCIA DE VIAJES DUAL	2	4
PASTAZA	AGENCIA DE VIAJES DUAL	3	5
SANTA ELENA	AGENCIA DE VIAJES DUAL	4	6
LOS RÍOS	AGENCIA DE VIAJES DUAL	4	7
ZAMORA CHINCHIPE	AGENCIA DE VIAJES DUAL	5	8
MORONA SANTIAGO	AGENCIA DE VIAJES DUAL	8	9
CHIMBORAZO	AGENCIA DE VIAJES DUAL	8	10
COTOPAXI	AGENCIA DE VIAJES DUAL	9	11
CAÑAR	AGENCIA DE VIAJES DUAL	11	12
ESMERALDAS	AGENCIA DE VIAJES DUAL	13	13
SANTO DOMINGO DE LOS TSÁCHILAS	AGENCIA DE VIAJES DUAL	13	14

Una vez ya realizado esto procedemos a declarar los datos que queremos mostrar y a estos los asignamos con la variable que van a representar, asignamos una consulta que en este caso va a mostrar el formulario el conteo de la subclasificacion escogida dentro de la provincia que se escogió

Figura 16 Declaración de variables que se mostraran en el formulario y relacionamiento de las mismas, además de la consulta que va a ejecutar.

```

("HOTEL" , "hotel"),
("HOSTAL" , "hostal"),
("HOSTERIA" , "hosteria"),
("ALOJAMIENTO" , "alojamiento"),
("BAR" , "bar"),
("DISCOTECA" , "discoteca"),
("RESTAURANTE" , "restaurante"),
("CAFETERIA" , "cafeteria")
)

val valorDefecto = valoresCategoria(0)
z.put("subclasificacion" , z.select("subclasificacion" ,valoresCategoria, valorDefecto ))

val ValProvincia = Seq(
("AZUAY" , "azuay"),
("SANTO DOMINGO DE LOS TSÁCHILAS" , "santo domingo de los tsachilas"),
("CARCHI" , "carchi"),
("NAPO" , "napo"),
("GALÁPAGOS" , "galapagos"),
("BOLÍVAR" , "bolivar"),
("EL ORO" , "el oro"),
("SANTA ELENA" , "santa elena"),
("SUCUMBÍOS" , "sucumbios"),
("PICHINCHA" , "pichincha"),
("PASTAZA" , "pastaza"),
("MORONA SANTIAGO" , "morona santiago"),
("MANABÍ" , "manabi"),
("ORELLANA" , "orellana"),
("LOJA" , "loja"),
("CHIMBORAZO" , "chimborazo"),
("TUNGURAHUA" , "tungurahua"),
("ZAMORA CHINCHIPE" , "zamora chinchipe"),
("GUAYAS" , "guayas"),
("LOS RÍOS" , "los rios")
)

val valorProv = ValProvincia(0)
z.put("Provincia" , z.select("Provincia" ,ValProvincia , valorProv ))
z.run("paragraph_1659041866176_4788269")

```

subclasificacion

discoteca

Provincia

pastaza

En este caso escogimos la subclasificación de discoteca en la provincia de pastaza y nos presentará lo siguiente:

Figura 17 Imagen de ejemplo de la ejecución del formulario

z.show(frR.where((\$"subclasificacion" === z.get("subclasificacion") && (\$"Provincia" === z.get("Provincia")))))

SPARK JOB FINISHED

Provincia	subclasificacion	count	rank
PASTAZA	DISCOTECA	6	4

#### 4.4 Unión de dataframes.

Se solicito dentro de las actividades realizar una unión de dos tablas, una que sería la tabla con la estamos trabajando y la siguiente seria una de las 5 variables que nosotros aumentamos en el complemento de base datos, de estas nosotros decidimos trabajar con la variable de población, para lo cual tuvimos que hacer la creación de un nuevo dataframe en donde guardaremos los datos de nuestra variable población. Para luego proceder a hacer la unión usando lo que sería el Inner Join.

Figura 18 Imagen de la creación de un nuevo dataframe que almacena los datos de la variable que vamos a unir

```
val poblacionCantones = spark
  .read.option("inferSchema", "true")
  .option("header", "true")
  .option("delimiter", ";")
  .csv("/workspace/zeppelin-paavanzada/poblacion.csv")

z.show(poblacionCantones)
```

SPARK JOB FINISHED

Id_Canton	nombre_canton	Poblacion
9-Ct	SANTA ISABEL	18393
10-Ct	GUALACEO	38587
11-Ct	PASTAZA	83.933
12-Ct	RUMIÑAHUI	109807
13-Ct	NARANJAL	53800
14-Ct	LA JOYA DE LOS SACHAS	40512
15-Ct	PASAJE	90806
16-Ct	GIRÓN	12607

Figura 19 Uso del inner join para la unión de los dataframe

```
val dataUnificada = dfConstNew.as("d1").join(poblacionCantones.as("d2"), col("d1.canton") === col("d2.nombre_canton"), "inner")
z.show(dataUnificada)
```

SPARK JOB FINISHED

nombre	clasificacion	subclasificacion	categoria	provincia	canton	parroquia	dir_referencia	direccion	telf_princ
FAMA LA	ALIMENTOS Y BEBIDAS	FUENTE DE SODA	null	AZUAY	CUENCA	CUENCA, CABECERA CANTONAL Y CAPITAL PROVINCIAL.	null	BOLIVAR 12-86	2843739
COLONIAS LAS No.3	ALIMENTOS Y BEBIDAS	RESTAURANTE	null	AZUAY	CUENCA	CUENCA, CABECERA CANTONAL Y CAPITAL PROVINCIAL.	null	AV. ESPAÑA Y AV. HURTADO DE MENDOZA	2804790
QUINTA ELOISA	INTERMEDIACIÓN	SALA DE	null	AZUAY	CUENCA	CUENCA.	null	JOSE RAFAEL	72829459

Para poder corroborar la unión y saber que estos datos pueden ser usados libremente realizamos dos consultas donde usaremos los datos incluidos y los que ya estaban antes

### 1) Cantones del Ecuador donde la poblacion es mayor a 50.000 y menor a 200.000

Figura 20 Consulta con el dataframe Unido 1

```
z.show(dataUnificada.select($"canton", $"Poblacion").where($"Poblacion">50000 && $"Poblacion"<100000).distinct.orderBy($"Poblacion"))
```

SPARK JOB FINISHED

canton	Poblacion
EL GUABO	50009
CHONE	62810
NARANJAL	53800
LA TRONCAL	54389
SALCEDO	57528
PLAYAS	59628
EL TRIUNFO	59636
SARAGURO	60000

## 2) Numero de comerciales existentes por persona en cada uno de los cantones de Loja

Figura 21 Consulta con el dataframe Unido 1

```
val comercialesCantonesLoja = dataUnificada.select($"canton").where($"provincia" === "LOJA").groupBy($"canton").count()
val poblacionLoja = dataUnificada.select($"canton" , $"Poblacion").where($"provincia" === "LOJA").distinct()
val completo = comercialesCantonesLoja.as("ccl").join(poblacionLoja.as("pl") , col("ccl.canton") === col("pl.canton") , "inner")
z.show(completo.select(col("ccl.canton") , col("count").as("NumComerciales") , col("Poblacion") , ($"count" / $"Poblacion").as("ComercialesPorPersona")))
```

SPARK JOB FINISHED

canton	NumComerciales	Poblacion	ComercialesPorPersona
LOJA	366	170280	0.002149400986610289
CATAMAYO	47	30,200	1.5562913907284768
SARAGURO	18	60000	3.0E-4
PINDAL	3	8645	3.470213996529786E-4
MACARÁ	30	25901	0.0011582564379753676
PUYANGO	4	15513	2.578482563011668E-4
ZAPOTILLO	12	16500	7.272727272727272E-4
CALVAS	19	28185	6.741174383537343E-4

## 4.5 Consultas SQL

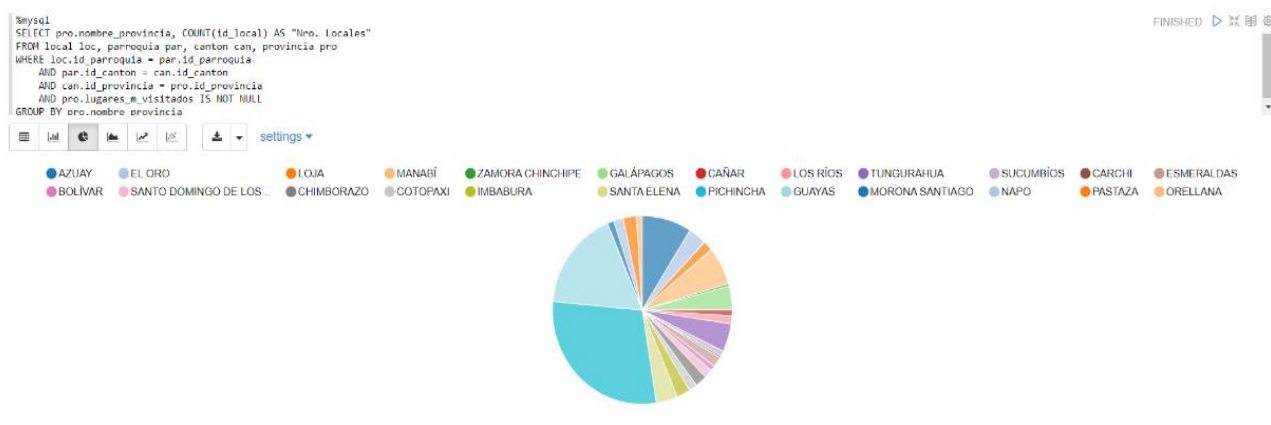
Para finalizar con el componente de programación hicimos la conexión con la base de datos de MySQL para lo cual hemos exportado el .ddl del script completo del complemento de base de datos, este script contiene la creación de la base de datos, creación del usuario, la inserción de datos que ya tienen su limpieza hecha, con este script cargado y ejecutado pudimos hacer la elaboración de las consultas.

### 1) Consultar: Mostrar el número de locales que por cada provincia que tienen lugares menos visitados

#### Explicación:

En el caso del turismo existen lugares menos visitados por lo que existe mayor numero de visitas que en otros lugares, por eso en esta consulta vamos a mostrar el numero de locales que existen en estos lugares menos visitados

Figura 22 Consulta de Sql 1

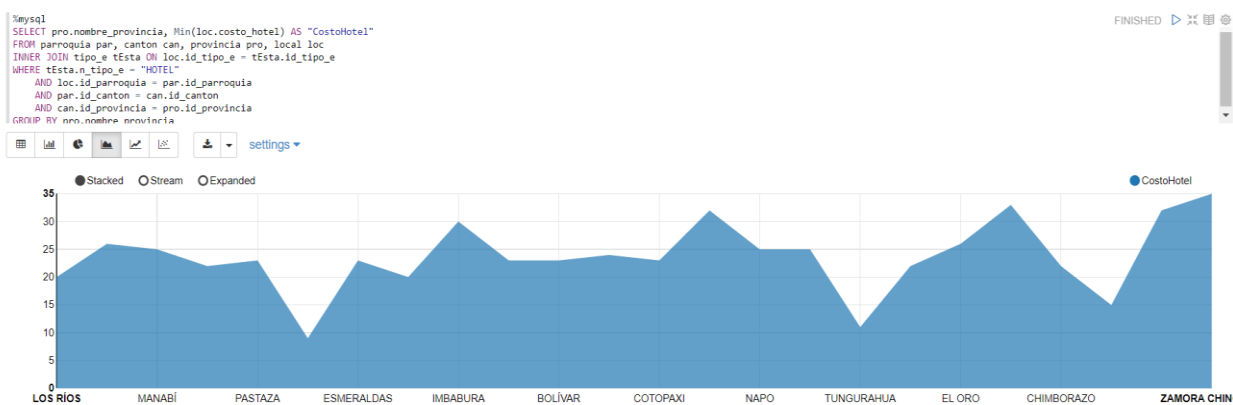


3) Consultar: ¿Cuál es el costo del hotel más barato en cada provincia?

Explicación:

Como ya es de nuestro conocimiento en nuestro país existen varios hoteles dentro de una misma provincia, por este motivo al llegar los turistas al territorio ecuatoriano o bien quieren buscar economía o placer por lo que buscan economía hemos implementado esta consulta dentro de la cual vamos a mostrar el hotel más barato dentro de cada provincia.

Figura 23 Consulta de Sql 2



#### 4) Consultar: ¿Cuántos bares hay por habitante en el cantón de Loja?

Explicacion:

En la ciudad de Loja existen varios centros de entretenimiento para personas mayores de 18 años pero siempre existe la incertidumbre de a que bares ir, y así es como pudimos determinar cuantas bares existen por persona en esta ciudad.

Figura 24 Consulta de Sql 3



#### 5) Consultar: Promedio de costos hoteles de cada provincia

Explicacion:

Los turistas y ciudadanos llevan un presupuesto que tienen destinado a hoteles por lo que implementamos esta consulta para sacar el promedio de los costos de los hoteles y asi podran saber cuanto dinero llevar para su alojamiento.

Figura 25 Consulta de Sql 4

The screenshot shows a MySQL query window with the following SQL code:

```
SELECT pro.nombre_provincia, AVG(loc.costo_hotel) AS "CostoHotel"
FROM parroquia par, canton can, provincia pro, local loc
INNER JOIN tipo_e tEst ON loc.id_tipo_e = tEst.id_tipo_e
WHERE tEst.n_tipo_e = "HOTEL"
AND loc.id_parroquia = par.id_parroquia
AND par.id_canton = can.id_canton
AND can.id_provincia = pro.id_provincia
GROUP BY pro.nombre_provincia
```

The result is displayed as a table with the following data:

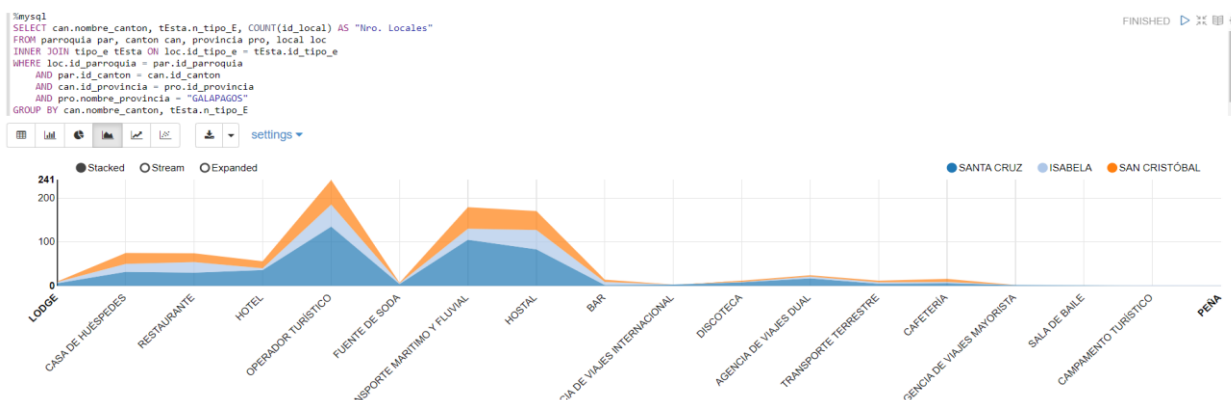
nombre_provincia	CostoHotel
TUNGURAHUA	69.8033
SANTO DOMINGO DE LOS TSÁCHILAS	56.5000
EL ORO	63.5349
GARCHI	55.7333
CHIMBORAZO	53.6970
LOJA	47.0500
CAÑAR	57.1429
ZAMORA CHINCHIPE	48.8333

- 6) Consultar: Por cada cantón de la provincia de Galápagos, y por cada tipo de Establecimiento ¿Cuántos establecimientos existen?

Explicacion:

Partiendo de la idea de mostrar a los turistas tanto nacionales como extranjeros se buscó mostrar mediante la realización de esta consulta, cuantos establecimientos existen en la provincia de Galápagos, clasificados por el tipo de establecimiento al que estos corresponden. El resultado dará un gran aporte a la comunidad de turistas, ya que al analizar la gráfica presentada podrán hacerse una idea de los tipos de establecimientos con los que se encontrarán en esta provincia.

Figura 26 Consulta de Sql 5



- 7) Cuales son hoteles menos costosos en el cantón donde hay aeropuertos

Explicacion:

Los turistas y ecuatoriano al viajar deséan ahorra dinero y tiempo por lo que buscan algún cantón se encuentre cerca de algún un aeropuerto y un algún hotel mas barato para alojarse.

Figura 27 Consulta de Sql 6



## 4.6 Página Web

Para la elaboración de la página web utilizamos la herramienta de PyCram y como base de la página usamos HTML y CSS para la distribución de esta y así poder mostrar de mejor manera las consultas.

Primeramente, empezamos dentro de un Body y con diferentes background image (se pega el url de la imagen), background size, background position, background attachment, y para la estética de la página le pusimos una imagen de fondo.



Figura 28 Codigo HTML de la pagina Web

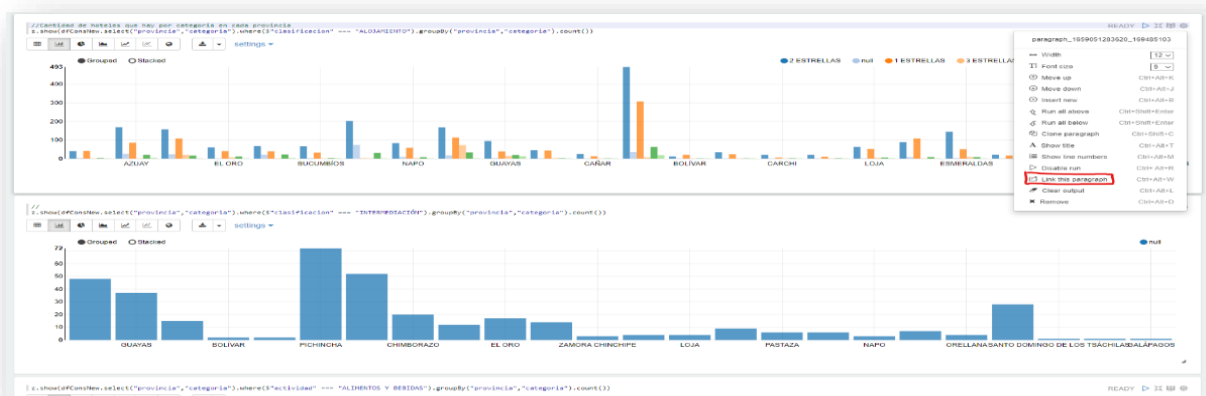
```

<!-- Datos de Ecuador.html -->
<html>
  <head>
    <title>Datos de Ecuador.html</title>
  </head>
  <body>
    <div class="container">
      <h1>Datos que no sabias del Ecuador</h1>
      <div class="contenido">
        <h2 style="text-align: center;">¿Cuántos aeropuertos hay por cada cantón?</h2>
        <p style="text-align: justify;">Para las personas que les gusta viajar rápido y no los viajes pesados de varias horas, en Ecuador hay distintos aeropuertos ubicados en diferentes cantones. En esta consulta vamos a mostrar el número de aeropuertos por cada cantón.</p>
        <iframe src="https://8080-jorgaf-zeppelinaavanza-s4vr7eydnt.ws-us54.gitpod.io/#/notebook/2H8FUMK8CR/paragraph/paragraph_1659046085406_1119504882?asiframe" height="300" width="100%"></iframe>
      </div>
      <div class="contenido">
        <h2 style="text-align: center;">¿Cuáles son los lugares menos visitados?</h2>
        <p style="text-align: justify;">En el caso de turistas existen lugares en los que existe mayor número de visitas que en otros, por eso en esta consulta vamos a mostrar el número de visitas por cada lugar.</p>
        <iframe src="https://8080-jorgaf-zeppelinaavanza-s4vr7eydnt.ws-us54.gitpod.io/#/notebook/2H8FUMK8CR/paragraph/paragraph_1659046085406_1119504882?asiframe" height="300" width="100%"></iframe>
      </div>
      <div class="contenido">
        <h2 style="text-align: center;">¿Cuál es el costo del hotel mas barato en cada provincia?</h2>
        <p style="text-align: justify;">Como ya es de nuestro conocimiento en nuestro país existen varios hoteles dentro de una misma provincia, por este motivo al llegar los turistas tenemos que elegir el hotel mas barato. En esta consulta vamos a mostrar el costo de los hoteles mas baratos por cada provincia.</p>
        <iframe src="https://8080-jorgaf-zeppelinaavanza-s4vr7eydnt.ws-us54.gitpod.io/#/notebook/2H8FUMK8CR/paragraph/paragraph_1659046085406_1119504882?asiframe" height="300" width="100%"></iframe>
      </div>
      <div class="contenido">
        <h2 style="text-align: center;">¿Cuántas bares hay por habitante en el cantón de Loja?</h2>
        <p style="text-align: justify;">En la ciudad de Loja existen varios centros de entretenimiento para personas mayores de 18 años pero siempre existe la incertidumbre de a que ciudad ir. En esta consulta vamos a mostrar el número de bares por habitante en el cantón de Loja.</p>
        <iframe src="https://8080-jorgaf-zeppelinaavanza-s4vr7eydnt.ws-us54.gitpod.io/#/notebook/2H8FUMK8CR/paragraph/paragraph_1659046085406_1119504882?asiframe" height="300" width="100%"></iframe>
      </div>
      <div class="contenido">
        <h2 style="text-align: center;">¿Por cada catño de la provincia de Salinas y por cada tipo de establecimiento cuántos establecimientos existen?</h2>
        <p style="text-align: justify;">Partiendo de la idea de mostrar a los turistas tanto nacionales como extranjeros se buscó mostrar mediante la realización de esta consulta, cuántos establecimientos existen en la provincia de Salinas y por cada tipo de establecimiento cuántos establecimientos existen.</p>
        <iframe src="https://8080-jorgaf-zeppelinaavanza-s4vr7eydnt.ws-us54.gitpod.io/#/notebook/2H8FUMK8CR/paragraph/paragraph_1659046085406_1119504882?asiframe" height="300" width="100%"></iframe>
      </div>
    </div>
    <div class="contenido">
      <h2 style="text-align: center;">Desarrolladores</h2>
      <p style="text-align: center;">Jhoni Ordóñez - Jamil Quilvazaca - Gerald Jaramillo</p>
    </div>
  </body>
</html>

```

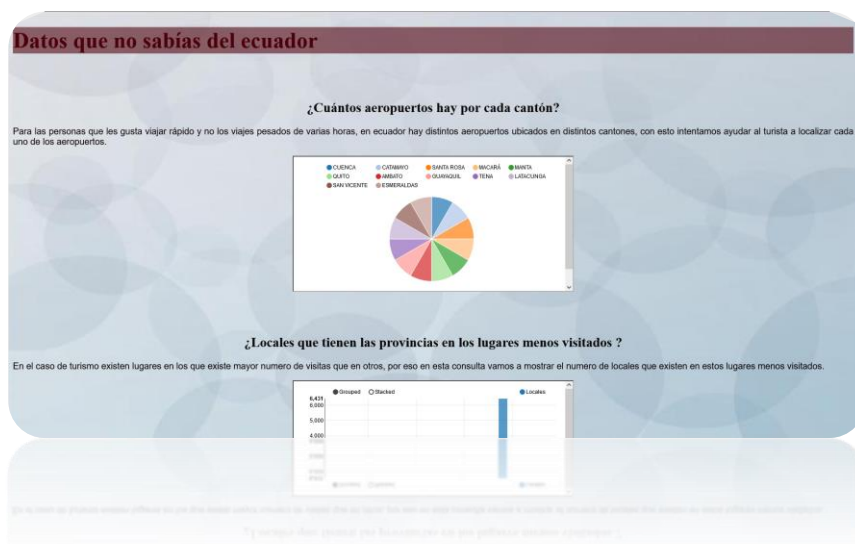
Para los títulos, subtítulos y texto en general añadimos estilos y color. Luego para continuar con la estructura puse div container y div class para ir separando las consultas e ir organizando cada texto, cada div class cuenta con estilos, justificado y centrado

Figura 29 Opcion Zeppelin



Para la carga de imágenes hubo que hacer una configuración y mediante una función llamada iframe donde en el src se coloca el url de las imágenes que se quiere traer del Zeppelin en el que previamente en las consultas del Zeppelin se obtiene los diferentes url para posteriormente colocarlos en cada función, este proceso se realizó para cada consulta. Finalmente, se le puso un footer para colocar texto y un h4 para el tamaño de letra donde coloque los integrantes.

Figura 30 Pagina Web



#### 4. Conclusiones

Al finalizar este proyecto nos dimos cuenta del refuerzo que hemos tenido de los conocimientos adquiridos durante el ciclo, al ser estas 2 materias las principales dentro de nuestra carrera, como lo son la materia de Programación avanzada y Base de datos avanzada, tuvimos que hacer diferentes actividades incluso revisión de varios temas aplicados dentro de este proyecto para así poder recordarlos en su totalidad y poder llevar este proyecto a su fin.

Temas importantes como la recolección de datos, para cada una de las variables que a nuestro parecer fueron las más importantes para ser agregadas dentro de este proyecto y poder llevar así un buen análisis en la ejecución del mismo, creación de script a del esquema principal usando

herramientas que nos han sido útiles por el tiempo que hemos cruzado el ciclo, la limpieza de los datos tanto en el complemento de programación como en el de base de datos en cada uno realizando las sentencias necesarias para la elaboración del mismo.

El saber como realizar la conexión necesaria de la base de datos con nuestra herramienta de Zeppelin para así poder realizar la ejecución del proyecto dentro de nuestra herramienta de programación. En fin tenemos muchos aspectos en donde tenemos claro la unión y relación que tienen estas dos materias, y la importancia que es que estas sean llevadas de manera pareja.

## 5. Bibliografía

Proyecto Github

[https://github.com/GeraldJT/Proyecto\\_Integrador/tree/main/Base%20de%20datos/Script/Archivo%20mySql/Creacion%20e%20inserci%C3%B3n%20de%20tablas](https://github.com/GeraldJT/Proyecto_Integrador/tree/main/Base%20de%20datos/Script/Archivo%20mySql/Creacion%20e%20inserci%C3%B3n%20de%20tablas)

Excel compartido

[https://github.com/GeraldJT/Proyecto\\_Integrador/tree/main/Base%20de%20datos/Archivos%20CSV/Excel%20compartido](https://github.com/GeraldJT/Proyecto_Integrador/tree/main/Base%20de%20datos/Archivos%20CSV/Excel%20compartido)

Csv de cada tabla

[https://github.com/GeraldJT/Proyecto\\_Integrador/tree/main/Base%20de%20datos/Archivos%20CSV/CSV](https://github.com/GeraldJT/Proyecto_Integrador/tree/main/Base%20de%20datos/Archivos%20CSV/CSV)

Apache Zeppelin

[https://github.com/GeraldJT/Proyecto\\_Integrador/tree/main/Programacion/Archivo%20de%20zeppelin](https://github.com/GeraldJT/Proyecto_Integrador/tree/main/Programacion/Archivo%20de%20zeppelin)

Página Web

[https://github.com/GeraldJT/Proyecto\\_Integrador/tree/main/Programacion/Pagina%20web](https://github.com/GeraldJT/Proyecto_Integrador/tree/main/Programacion/Pagina%20web)