

# Universidad Técnica Particular de Loja



## Asignatura:

Prácticum 1.2

## Integrantes:

Arianna Marikrys Ramón

Jandry Josué Jaramillo

José Fabián Montoya

Jandri Giovanni Villavicencio

## Docentes:

Ángel Eduardo Encalada

Jorge A. López Vargas

Ciclo: 4

Abril/Agosto 2022

## 1. Introducción

El proyecto integrador de saberes se orienta a la realización de análisis de datos complementando el componente de base de datos para el procesamiento de la data y creación de la base de datos y el componente de programación para realizar análisis estadísticos y extracción de información desde la base de datos creada.

Específicamente para el desarrollo del componente de base de datos, se utiliza como data base un dataset extraído del Ministerio de Turismo que contiene información de establecimientos turísticos registrados en Ecuador. Con estos datos base se realiza un análisis para conocer su estructura y determinar si se tendrá que agregar, corregir o eliminar; posterior a esto se da el procesamiento de la Data, dando paso al modelado con el cual se procede a la implementación de la base de datos para la posterior carga de la data y subsiguiente uso en el componente de programación.

En componente de programación se realiza limpieza de los datos de algunas de las columnas de la data que contenían errores, también se realizaron reportes estadísticos sobre la data con las nuevas variables que se incluyeron en el componente de base de datos, se desarrolla una página web con los reportes estadísticos más relevantes que se desarrolló mediante el lenguaje sql.

## 2. Planificación del trabajo

El proyecto como tal se lo desarrollo en distintas fases, independientemente a los componentes de Bases de Datos y Programación Avanzada, las mismas que se detallan a continuación:

Planificación del trabajo	
Etapas	Descripción
Lluvia de ideas	Tras recibir el documento Excel con la data necesaria para el proyecto se procedió a realizar una lluvia de ideas sobre los posibles análisis que a futuro podríamos realizar con la misma.
Definición de datos complementarios	Tras el análisis del documento Excel <a href="#">Consolidado-Nacional-2022-publico-1-web</a> se identificó que algunas variables se podrían agregar y quedaría una documentación más completa y rica en cuanto a análisis de data se refiere, por la misma razón mediante investigación se decidieron añadir siete variables las cuales se describen más adelante.
Diseño conceptual y lógico de la base de datos.	Una vez teniendo identificada toda la estructura del documento Excel se procedió a modelar el diseño conceptual que tendría la base de datos, posterior a esto se lo tradujo a un diseño lógico relacional.
Preparación de la data	En esta etapa se incorporó la información de las variables extra previamente definidas, estas misma se colocaron en una hoja de Excel cada una, aparte se realizó una pequeña limpieza de datos a nivel de Excel sobre el documento Consolidado-Nacional-2022-publico-1-web.
Generación de Script DDL	En MySQL se generó un script donde posteriormente se cargaría toda la data propuesta.
Preprocesamiento de la data	A nivel de Excel se dio una limpieza de datos profunda, lo más que para ese momento se nos era posible.
Implementación de la base de datos y carga de data.	Se implementó al completo la base de datos en MySQL y se procedió a la respectiva carga de datos, cabe recalcar que para este último paso previamente se habían dividido cada variable dentro del documento Excel en un archivo tipo CSV distinto,

	para de esta manera el proceso de la carga sea más sencillo, limpio y organizado.
Creación de repositorio GITHUB	Se creó un repositorio GitHub donde se le agregó a todos los integrantes del grupo además del docente a cargo de Base de datos Avanzada, aparte se crearon dos carpetas donde se depositaria la documentación correspondiente a Base de datos Avanzada y Programación Avanzada respectivamente, el cual se puede acceder con el siguiente enlace: <a href="https://github.com/f4biaan/Info-Turistica-Ecuador---Proyecto-Integrador">https://github.com/f4biaan/Info-Turistica-Ecuador---Proyecto-Integrador</a> .
Informe final del componente de Base de datos Avanzada	Una vez culminando el componente de Base de datos Avanzada se procedió a realizar un informe donde se detallan todas las actividades realizadas hasta ese momento.
Instalación y capacitación en Apache Zeppelin	Con la ayuda del docente de Programación Avanzada instalamos, configuramos e instalamos Apache Zeppelin, aparte que durante unas semanas recibimos capacitación en la misma, etapa que describe de mejor manera más adelante en la documentación.
Carga y limpieza de datos en Zeppelin	Se procedió con la carga y posterior limpieza de toda la data a nivel de Zeppelin es decir mediante sus comandos y sentencias, desde borrar datos duplicados hasta eliminar espacios innecesarios.
Análisis de datos tanto usando la data original como la data modificada con variables extra.	Una vez la data se encontraba limpia y sin errores se procedió a realizar un extenuante análisis tanto de la data original como de la data final. Se sacó a flote los análisis más importantes y que más podrían aportar para un análisis de datos real.
Análisis de datos a nivel de MySQL	Solo con la data final se procedió a realizar un análisis de datos mediante sentencias MySQL donde también se dio especial atención a los análisis más relevantes.
Análisis de resultados	En esta etapa final tras se realizó un complejo y bien pensado análisis sobre todo el proceso que habíamos realizado, dando más énfasis claro en el análisis de datos donde resaltamos los más impactantes tanto a nivel grafico como a nivel de sentencias.

### 3. Fuentes de datos

#### a. Fuentes base

Al iniciar el proyecto se nos hizo entrega de un documento Excel denominado "[Consolidado-Nacional-2022-publico-1-web](#)", este mismo contaba con cierta información acerca del catastro turístico de nuestro país, contando con 13 columnas las cuales fueron nuestros datos base.

Datos Base	
Datos	Descripción
Nombre Comercial	Proporciona el nombre del local comercial en cuestión, muy útil pues funciona como un identificador.
Actividad	Clasifica la actividad que realiza un cierto local comercial, en este caso se subdivide en siete, siendo las mismas “Alimentos y bebidas, Alojamiento, Centro de turismo comunitario, Intermediación, Operación e intermediación, Parques de atracción estable, Transporte turístico”.

Casificación	Se subdivide de actividad siendo de esa manera un dato más específico y ofreciendo más información, por ejemplo, un establecimiento que como actividad ofrezca Alimentos y bebidas puede ser una cafetería, un restaurante o incluso una fuente de soda.
Categoría	Categoriza el servicio que ofrece un establecimiento, el caso más claro vendría siendo el de los hoteles que pueden ser desde una hasta cinco estrellas.
Provincia	Proporciona a nivel de provincia la ubicación de cierto establecimiento siendo está bastante general.
Cantón	Proporciona a nivel cantonal la ubicación de cierto establecimiento siendo esta poco general.
Parroquia	Proporciona a nivel parroquial la ubicación de cierto establecimiento, siendo esta más específica.
Referencia de Dirección	Proporciona una referencia de ubicación del local en cuestión.
Dirección	Proporciona la dirección exacta del establecimiento.
Teléfono Principal	Nos da el medio de contacto principal con el establecimiento.
Teléfono Secundario	Un teléfono de respaldo que sirva como contacto con el establecimiento.
Correo Electrónico	Medio de contacto extra, generalmente se lo suele usar para asuntos más formales.
Dirección Web	Proporciona la dirección web de los establecimientos que cuentan con su propio dominio en internet.

#### b. Fuentes complementarias

En el documento Excel denominado "[Consolidado-Nacional-2022-publico-1-web](#)" se le han agregado nuevas variables, algunas fueron ingresadas manualmente y en otras se logró encontrar información en documentos Excel o también en páginas web, todo este proceso se lo almaceno en un archivo al que lo llamamos "[Preprocesamiento-Consolidado-Nacional-2022-publico-1-web.xlsx](#)".

Las siguientes variables que fueron agregadas son:

Datos Complementarios	
Dato	Descripción
Guías turísticas	Información que fue encontrada en la página web del <a href="#">Catastro nacional de Establecimientos Turísticos y Guías de Turismo Continente</a> , nos permite descargar un documento Excel con toda la información como nombre, clasificación, provincia, cantón, nacionalidad, fecha emisión licencia, idiomas, email, área protegida y modalidad aventura que pertenece a cada guía.
Climas	Información que se la implemento por cada cantón investigando en tres diferentes páginas web: <ul style="list-style-type: none"> <li>• <a href="https://bioweb.bio/fungiweb/GeografiaClima/">https://bioweb.bio/fungiweb/GeografiaClima/</a></li> <li>• <a href="https://es.wikipedia.org/wiki/Clima_del_Ecuador">https://es.wikipedia.org/wiki/Clima_del_Ecuador</a></li> <li>• <a href="https://es.db-city.com/search?q=ecuador&amp;sa.x=0&amp;sa.y=21">https://es.db-city.com/search?q=ecuador&amp;sa.x=0&amp;sa.y=21</a></li> </ul>
Capacidad	Información sobre el número de camas y habitaciones que contiene cada cantón y a la subactividad a la que pertenecen, esta información se encontró en la página web del <a href="#">Catastro de Servicios turísticos</a> .

Población	Información sobre la población que se encuentra en cada cantón y provincia, para obtener la población de las provincias se realizó la suma de cada cantón que pertenece a una provincia en particular. Para obtener los datos de la población de cada cantón se encontró en la página web de la <a href="#">INEC</a> .
Puntuación	Información que se investigó de forma manual cada link de la página web de los locales comerciales del documento de Excel <a href="#">Consolidado-Nacional-2022-publico-1-web</a> , analizando cada calificación del público al local comercial. En caso de que la pagina no cuente con el apartado de calificación, se usó la puntuación que los usuarios le otorgaron en Google maps.
Web funcional	Información que se investigó de forma manual cada link de la página web de los locales comerciales ya que algunos contenían links inválidos.
Pet_Friendly	Información que se investigó de forma manual cada link de la página web de los locales comerciales para ver si en algunos restaurantes, alojamientos, parques, etc. Se permitía el ingreso de las mascotas a los establecimientos.

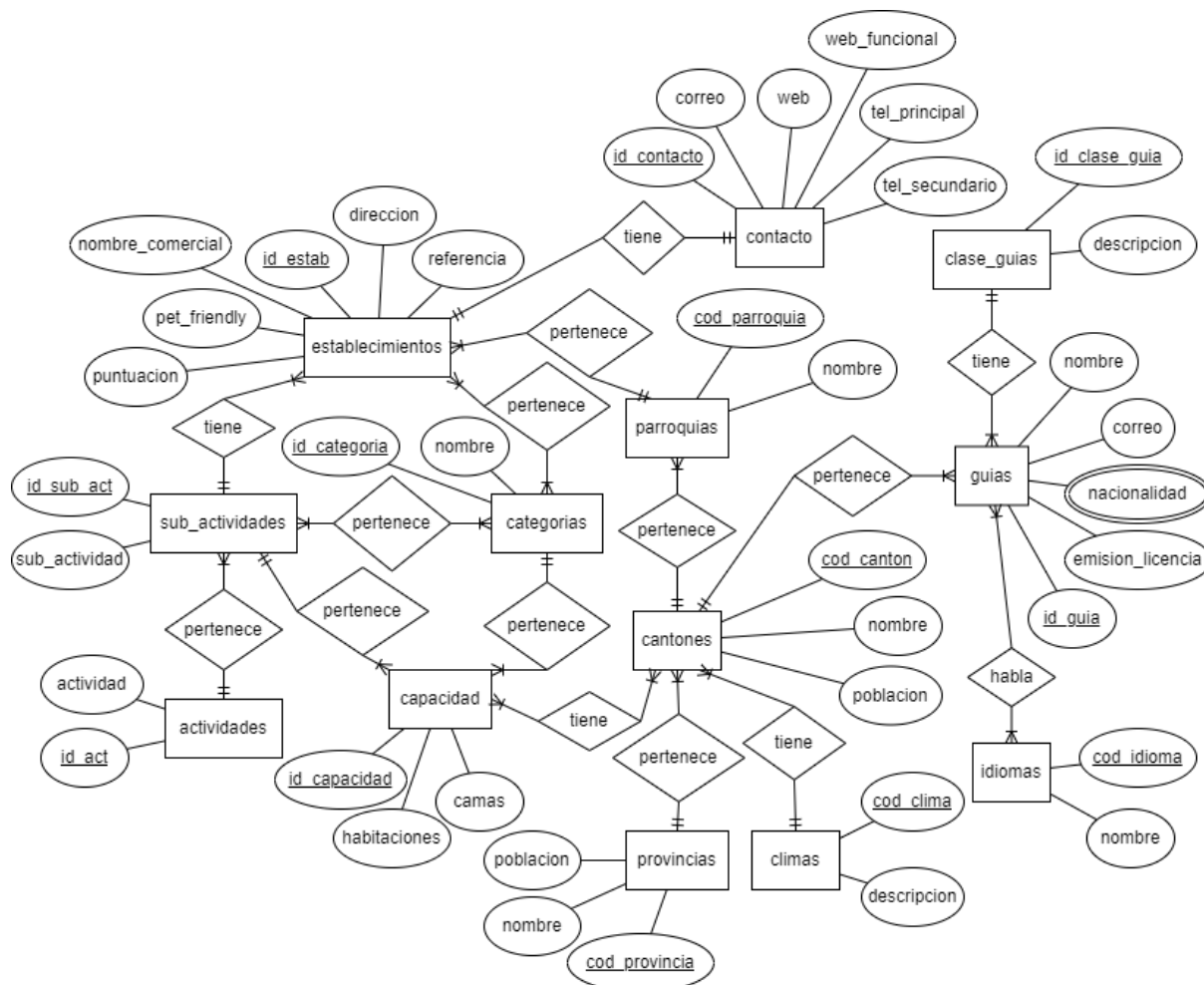
#### 4. Componente Base de datos

##### a. Explicación de preprocesamiento de datos realizado

##### i. Diseño conceptual

Antes de modelar la base de datos, necesitamos estructurar los datos que vamos a almacenar. Como los vamos a relacionar para que tengan coherencia entre sí, y para esto realizamos el Modelo Entidad/Relación que se puede visualizar en la siguiente imagen donde se muestran las entidades y las

relaciones que tienen entre sí y sus atributos, además sus restricciones y requerimientos:

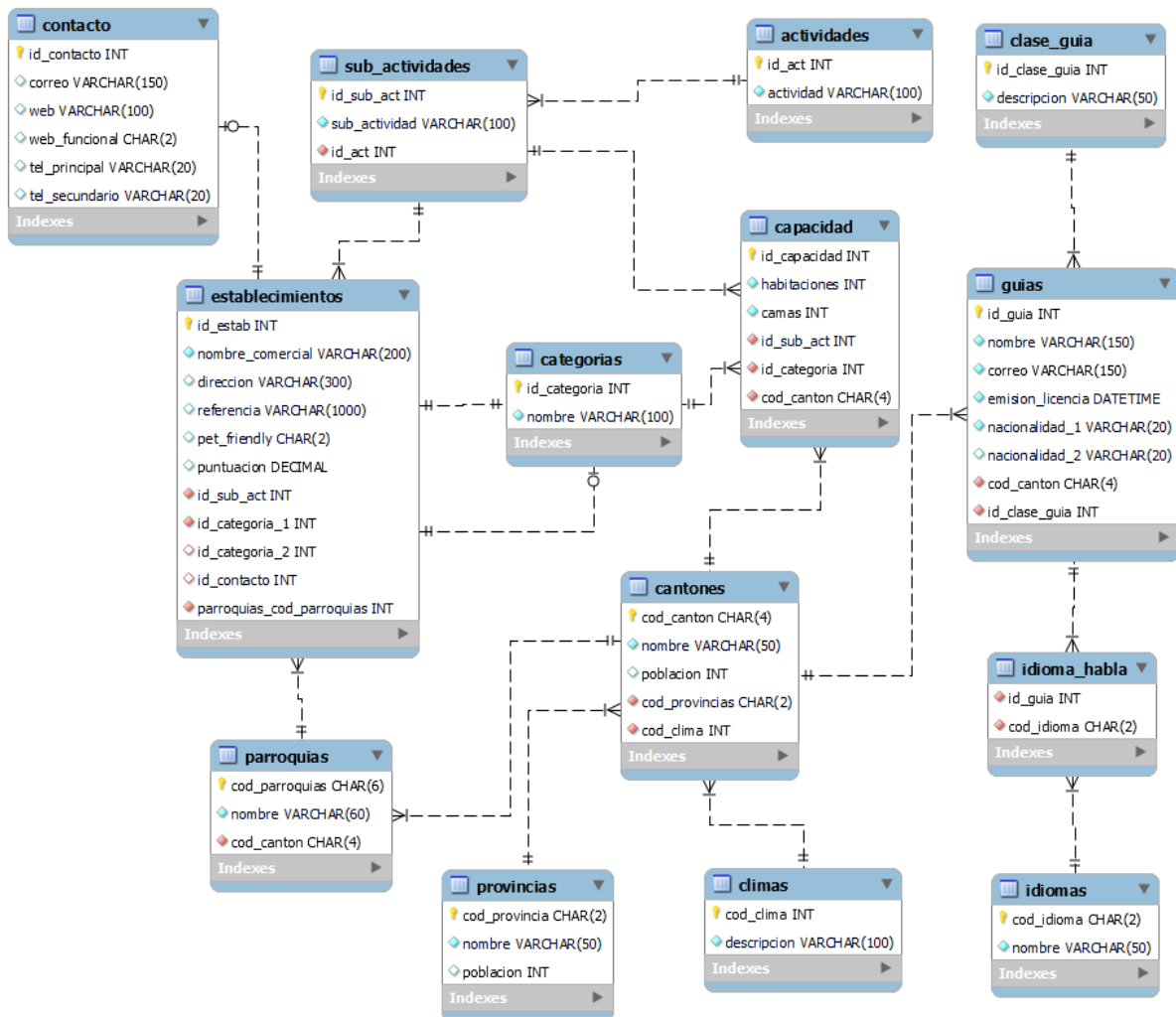


#### Restricciones y requerimientos:

- Se necesita por lo menos un método de contacto para que un establecimiento pueda registrar su forma de contacto.
- El atributo pet\_friendly solo acepta char(2), “SI” o “NO”.
- El campo nacionalidad de la entidad guías es obligatorio, pero solo puede tener como máximo 2 nacionalidades.
- El atributo web\_funcional solo acepta char(2), “SI” o “NO”.
- La entidad establecimientos debe tener como mínimo 1 categoría y máximo 2 categorías.
- El campo de población en la entidad Provincias debe ser calculado, es decir, se deben sumar las poblaciones de los cantones que pertenecen a la misma provincia, dando así el resultado de la población de la provincia.

#### ii. Diseño lógico

Una vez estructurados los datos, se procede a realizar el [Modelo Lógico](#) en este caso para la base de datos relacional que se va a implementar en el motor de base de datos MySQL, la cual nos permitirá diseñar el modelo físico de la base de datos:



#### Restricciones y requerimientos:

- Para registrar el contacto de un establecimiento se debe tener al menos una forma de contacto.
- El campo de población de una provincia es calculado, se deben sumar las poblaciones de los cantones de una provincia para dar como resultado la población de esa misma provincia.
- El campo pet\_friendly de la entidad establecimientos tiene el dominio: SI, NO
- El campo web\_funcional de la entidad contacto tiene dominio: Si, NO.

En el siguiente link “[Modelado de base de datos](#)”, se encontrará como se realizó el diseño conceptual y el diseño lógico relacional del documento Excel modificado que se lo denominó “[Preprocesamiento-Consolidado-Nacional-2022-publico-1-web.xlsx](#)” en el cual se implementó las nuevas variables.

#### iii. Implementación de la base de datos y carga de datos

El proceso que realizamos previo a la implementación de la base de datos sería el siguiente:



1. Extracción de la data a utilizar: desde la página oficial del Ministerio de Turismo del Ecuador se obtuvo un documento Excel(\*.xlsx) [Consolidado-Nacional-2022-publico-1-web.xlsx](#) el cual contiene datos referentes a los distintos establecimientos que han sido registrados.
2. Análisis de la data: ahora con los datos accesibles para nuestra visualización, procedemos a realizar un análisis de la estructura, la utilidad y relación que tienen los datos entre sí.
3. Luego realizamos un proceso de limpieza de datos y también agregación de datos que podrían ser de ayuda para generar más reportes estadísticos los cuales los agregamos al archivo de la data base.
4. Ahora conociendo la [estructura del dataset](#), se procede a realizar el modelado de la base de datos:
  - Primero se realiza el [Modelo Entidad/Relación](#) que abarcan todos los atributos agrupados con la entidad que mejor los define, y así nos ayudamos a dar una estructura formal a los datos.
  - Luego con base al modelo relacional procedemos a realizar el [Modelo lógico](#) el cual nos sirve como base para realizar el modelo físico de la base de datos, pero con ayuda de la herramienta Workbench al realizar el [Modelo Relacional](#) nos permitía generar el [Script DDL](#) para la creación de la base de datos.
5. Después de generar el [ScriptDDL](#), lo revisamos para constatar que este correcto, además también para agregar, la inserción de los datos a partir de [archivos CSV](#) que generamos al momento de procesar la data, además de la creación de un usuario para que se puedan realizar consultas, sin necesidad de conectarnos como usuario ROOT, todo este proceso se automatizó en un solo [ScriptDDL](#).

El [script SQL](#) se lo implementó en MySQL, el cual se realizó el proceso de creación de la estructura para las tablas que se encuentran en el documento de Excel modificado además provee maneras de borrar las tablas en caso de que ya exista una o que contenga un error en algún tipo de dato de los atributos, en la creación de la estructura de las tablas se define cual es la Primary Key así como también cuales son las Foreign Key esto aparte también se incluyó la creación de los índices y se dio paso a la implementación de restricciones. Cabe recalcar que la carga de datos se realizó mediante la creación de csv y se los subió a la base de datos MySQL mediante el uso de los comandos SQL.

Después de implementar y cargar la base de datos en MySQL se podía realizar consultas a partir de los datos existentes, esto se puede realizar a partir de un [ScriptSQL](#), el cual guarda la data en [archivos CSV](#) dependiendo la conveniencia y coherencia de los atributos entre tablas, en este caso hemos podido agrupar en 3 grupo de datos:

1. [Establecimientos](#)
2. [Guías](#)



### 3. [Capacidad de alojamientos](#)

## 5. Componente Programación

### a. Explicación de herramientas utilizadas

- Gitpod: Una vez teniendo una base de datos funcional y disponible al uso necesitábamos hacer análisis de datos, ahí es donde entra Apache Zeppelin, un cuadernillo basado en la web, ciertamente una herramienta muy útil, pero con un pequeño problema, no es compatible para el sistema operativo de Windows razón por la cual nuestro docente guía en programación avanzada nos dio a conocer algunas alternativas, de entre ellas la que más nos llamó la atención fue Gitpod, una plataforma IDE en la nube de código abierto, haciendo uso de la misma y siguiendo las instrucciones del docente logramos instalar y configurar Zeppelin de manera exitosa, el proceso que siempre debíamos seguir era el siguiente: abrir Gitpod en nuestro navegador de confianza, instalar la versión de Java 8.0.302-open y la versión de Scala 2.15.16 así nuestro Zeppelin estaría 100% funcional.
- Zeppelin: Para el respectivo análisis de datos se trabajó con Apache Zeppelin, como ya se mencionó, es un cuadernillo que permite realizar análisis de datos interactivos además de que en él se pueden trabajar con varios lenguajes de véase Scala o MySQL, pero no solo se queda ahí ya que también se puede hacer adquisición de datos, descubrimiento de datos, análisis de datos y la que más nos ha llamado la atención la visualización de datos y colaboración, para esta última Zeppelin ya trae algunas herramientas por defecto, pero también se permite ampliar el abanico de visualizaciones añadiendo pluggins, para este proyecto en concreto hemos hecho uso tanto de [ultimate-pie-chart](#) como de [zeppelin-number](#) dos herramientas que creemos ayudaran a comprender mucho mejor los resultados obtenidos representados por gráficos. Por parte de la colaboración nos ha sido muy útil ya que hemos podido trabajar todos los integrantes del grupo en un solo notebook de Zeppelin haciendo nuestro trabajo mucho más ameno y organizado. En el periodo de tiempo que abarca desde la semana nueve hasta la semana quince el docente tutor nos instruyó en el uso de esta herramienta, conociendo así sobre sus comandos, configuraciones e intérpretes de los cuales podríamos hacer uso. Realizamos análisis de diversos tipos de data, desde archivos tipo csv hasta tipo JSON, como también archivos cargados directamente desde la web así haciendo uso de una de sus propiedades, la adquisición de datos.
- MySQL: Como se presentó la necesidad de conectar una base de datos con Zeppelin, fue conveniente recurrir al uso del motor de base de datos MySQL, pero para que esto sea posible era necesario configurar Apache Zeppelin, esto mediante la agregación de una dependencia y por supuesto también un driver, una vez esto listo, se nos es posible realizar consultas mediante sentencias SQL directamente desde Zeppelin.
- SQL: Como se especificó, Zeppelin soporta varios tipos de lenguaje, para el análisis final de los datos se decidió que se ha de usar SQL, he aquí donde hacemos uso del lenguaje y el motor de base de datos

MySQL pues en el esta almacenada toda la información que recuperamos y con la cual realizamos análisis de datos que creemos son los más relevantes e impactantes, todo esto llevado a cabo mediante sentencias SQL claro está.

- Página web: Específicamente en este componente se desarrolló lo que se puede llamar una página web por supuesto esta solo se llevó a cabo a nivel local, lo importante de esta herramienta es que en la misma se proyectaron los análisis de datos más relevantes que se pudieron encontrar, en total se presentaron cinco análisis de lo más interesantes, cabe recalcar que se mostró tanto una parte grafica como escrita, la parte grafica estaba directamente vinculada con Zeppelin razón por la cual si se hacía un cambio en las sentencias dentro de Zeppelin las mismas se veían reflejadas de manera automática en la página web por último se debe aclarar que algunos análisis eran de tipo cuestionario , es así que algunas imágenes dentro de la página web también eran interactivas.

## b. Explicación de comandos y sentencias usadas para consultar y visualizar datos a nivel del archivo fuente

### a) Lectura de archivos:

#### a. xlsx:

```
val dfData = spark
  .read
  .format("com.crealytics.spark.excel")
  .option("header", "true")
  .option("inferSchema", "true")
  .load("/path/file.xlsx")
```

Este comando escrito en lenguaje Scala sirve para realizar la extracción de datos contenidos en un documento Excel que lo podemos leer desde la web o de forma local y lo almacena en un dataframe, además nos permite inferir el tipo de dato de cada columna o generar un schema personalizado con el tipo de datos que nosotros requiramos.

#### b. csv:

```
val dataCapacidad = spark
  .read
  .option("inferSchema", "true")
  .option("header", "true")
  .option("delimiter", ";")
  .csv("/path/file.csv")
```

Por otro lado, también tenemos la lectura de archivos CSV que nos permite personalizar el separador de cada dato y al igual que en archivos XLSX también permite inferir o personalizar el tipo de dato de las columnas también permite personalizar el schema y es más también podemos especificar si las columnas tienen encabezado o no.

### b) Limpieza de datos:

Todavía seguimos utilizando el lenguaje Scala para trabajar con los datos y para su limpieza utilizamos las siguientes sentencias para corregir los datos y más:

- a. Para renombrar las columnas de un dataframe se utilizó el comando:

```
dfData.withColumnRenamed("Nombre Comercial", "nombre")
```

- b. Comando trim se utiliza para eliminar los espacios en blanco al inicio o final de la cadena, y upper para convertir todos los caracteres en mayúsculas:

```
dfData.withColumn("nombre", trim(upper($"nombre")))
```

- c. Corrección de cadenas de columnas para corregir cadenas con casos en los que queremos reemplazar cadenas:

```
dfData.withColumn("actividad",  
  when($"actividad" === "INTERMEDIACIÓN", "INTERMEDIACION")  
  .when($"actividad" === "OPERACIÓN E INTERMEDIACIÓN", "OPERACION E INTERMEDIACION")  
  .when($"actividad" === "PARQUES DE ATRACCIÓN ESTABLE", "PARQUES DE ATRACCION ESTABLE")  
  .when($"actividad" === "TRANSPORTE TURÍSTICO", "TRANSPORTE TURISTICO")  
  .otherwise($"actividad"))
```

- c) Imprimir esquemas: para conocer la estructura de los dataframes y los tipos de datos de las columnas.

```
dfData.printSchema
```

- d) Combinar dataframes: con el método join nos permite combinar dataframes por medio de un valor para que los datos conserven su integridad:

```
dfData.printSchema; dfDataClean.join(dataCapacidad, Seq("provincia"), "inner")
```

- e) Convertir Dataframe en vista temporal en SQL:

```
dfAllData.createOrReplaceTempView("datos")
```

- f) Consultas SQL: Por medio de sentencias sql, teniendo como encabezado %sql y sentencias SQL:

```
%sql  
SELECT provincia, actividad, habitaciones, camas  
FROM datos  
WHERE actividad = "ALOJAMIENTO"  
GROUP BY provincia, actividad, habitaciones, camas;
```

- g) Formularios:

Para crear gráficas interactivas en las que se puedan elegir varias opciones se utilizó los siguientes comandos:

○ ○ ○

consulta

```
val dfDesarrollo = dfDataClean
.select("provincia", "actividad", "categoria")
.groupBy("actividad", "categoria")
.count() GROUP BY provincia, actividad, habitaciones, camas;
```

○ ○ ○

window

```
import org.apache.spark.sql.expressions.Window

val w = Window.partitionBy($"actividad").orderBy(asc("count"))

val dfRespuesta = dfDesarrollo.withColumn("rank", row_number().over(w))
dfRespuesta.show(200, false)
```

○ ○ ○

filtros

```
dfRespuesta
.select("actividad", "categoria")
.where($"rank" == 15 || $"rank" == 11 || $"rank" == 1 || $"rank" == 4 || $"rank" == 3 || $"rank" ==
14).show(false)
```

○ ○ ○

creación de formulario de opciones

```
val valoresCombo = Seq(
  ("AZUAY", "AZUAY"),
  ("BOLIVAR", "BOLIVAR"),
  ("CAÑAR", "CAÑAR"),
  ("CARCHI", "CARCHI"),
  ("COTOPAXI", "COTOPAXI"),
  ("CHIMBORAZO", "CHIMBORAZO"),
  ("EL ORO", "EL ORO"),
  ("ESMERALDAS", "ESMERALDAS"),
  ("GUAYAS", "GUAYAS"),
  ("IMBABURA", "IMBABURA"),
  ("LOJA", "LOJA"),
  ("LOS RIOS", "LOS RIOS"),
  ("MANABI", "MANABI"),
  ("MORONA SANTIAGO", "MORONA SANTIAGO"),
  ("NAPO", "NAPO"),
  ("PASTAZA", "PASTAZA"),
  ("PICHINCHA", "PICHINCHA"),
  ("TUNGURAHUA", "TUNGURAHUA"),
  ("ZAMORA CHINCHIPE", "ZAMORA CHINCHIPE"),
  ("GALAPAGOS", "GALAPAGOS"),
  ("SUCUMBIOS", "SUCUMBIOS"),
  ("ORELLANA", "ORELLANA"),
  ("SANTO DOMINGO DE LOS TSACHILAS", "SANTO DOMINGO DE LOS TSACHILAS"),
  ("SANTA ELENA", "SANTA ELENA")
)

val valorDefecto = valoresCombo(0)
z.put("PROVINICA", z.select("PROVINICA", valoresCombo, valorDefecto))

val valoresCombo1 = Seq(
  ("ALIMENTOS Y BEBIDAS", "ALIMENTOS Y BEBIDAS"),
  ("ALOJAMIENTO", "ALOJAMIENTO"),
  ("INTERMEDIACIÓN", "INTERMEDIACIÓN"),
  ("OPERACIÓN E INTERMEDIACIÓN", "OPERACIÓN E INTERMEDIACIÓN"),
  ("TRANSPORTE TURÍSTICO", "TRANSPORTE TURÍSTICO")
)

val valorDefecto1 = valoresCombo1(0)
z.put("ACTIVIDAD", z.select("ACTIVIDAD", valoresCombo1, valorDefecto1))
z.run("paragraph_1658848961680_178835406")
```

○ ○ ○

generación de gráfica con las opciones

```
z.show(dfRespuesta.where($"ACTIVIDAD" === z.get("ACTIVIDAD")))
```

c. Explicación de comandos y sentencias usadas para consultar y visualizar datos a nivel de base de datos MySQL

a) Creación de formularios en consulta en lenguaje SQL en la sentencia WHERE:

```

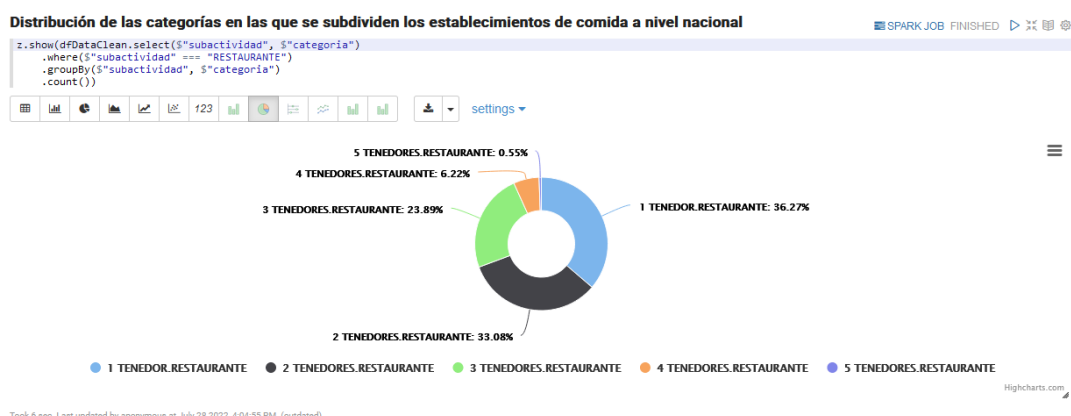
Consulta en la base de datos MySQL tij

%mysql
SELECT COUNT(*) Total, sact.sub_actividad 'SUBACTIVIDAD ECONOMICA'
FROM establecimientos e
  INNER JOIN parroquias parr ON e.cod_parroquias = parr.cod_parroquias
  INNER JOIN cantones can ON parr.cod_canton = can.cod_canton
  INNER JOIN provincias prov ON can.cod_provincias = prov.cod_provincia
  INNER JOIN sub_actividades sact ON e.id_sub_act = sact.id_sub_act
  INNER JOIN actividades act ON sact.id_act = act.id_act
WHERE prov.nombre = "${PROVINCIA=LOJA,LOJA|ZAMORA|GALAPAGOS|AZUAY|SANTO DOMINGO DE LOS
TSACHILAS|CARCHI|LOS RIOS|NAPO|EL ORO|SANTA ELENA|PICHINCHA|PASTAZA|MORONA
SANTIAGO|ORELLANA|SUCUMBIOS|CHIMBORAZO|MANABI|BOLIVAR}"
  AND act.actividad = "${ACTIVIDAD ECONOMICA=ALIMENTOS Y BEBIDAS,PARQUES DE ATRACCION
ESTABLE|OPERACION E INTERMEDIACION|ALOJAMIENTO|TRANSPORTE TURISTICO|ALIMENTOS Y
BEBIDAS|INTERMEDIACION|CENTRO DE TURISMO COMUNITARIO}"
GROUP BY sact.sub_actividad;

```

- a. SELECT: se filtran solo las columnas o cálculos que se necesitan.
- b. FROM: se seleccionan las tablas donde se van a realizar las consultas.
  - i. INNER JOIN: combinación de tablas para aplicar filtros.
- c. WHERE: filtros necesarios que se deben aplicar para realizar análisis.
- d. GROUP BY: agrupa los registros dependiendo de atributos de una o varias columnas.

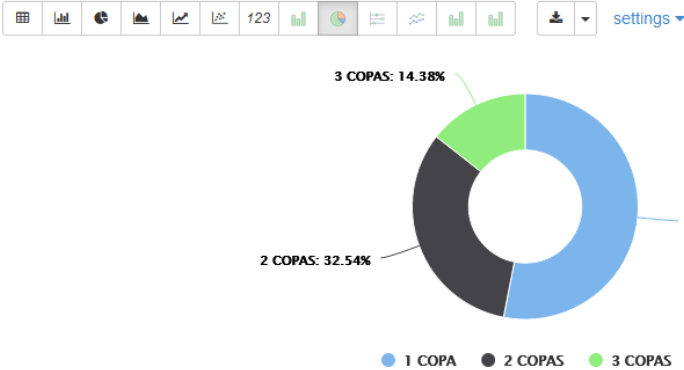
## 6. Resultados obtenidos



## Distribución a nivel nacional de las categorías en las que se subdividen los establecimientos de de bar y discoteca

SPARK JOB FINISHED

```
z.show(dfDataClean.select($"subactividad", $"categoria")
  .where($"categoria" === "1 COPA" || $"categoria" === "2 COPAS" || $"categoria" === "3 COPAS")
  .groupBy($"subactividad", $"categoria")
  .count())
```



Highcharts.com

Took 5 sec. Last updated by anonymous at July 28 2022, 4:07:07 PM. (outdated)

## Mayor presencia de restaurantes de alta cocina a nivel nacional

```
z.show(dfDataClean.select($"categoria", $"provincia", $"canton")
  .where($"categoria" === "5 TENDONES")
  .groupBy($"categoria", $"provincia", $"canton")
  .count())
```

SQL SELECT provincia, actividad, habitaciones, camas PROP. datos WHERE actividad = "ALOJAMIENTO" GROUP BY provincia, actividad, habitaciones, camas

25k 20k 15k 10k 5k 0

camas(suon) habitaciones(suon)

AMBA TO AZOGUES BABAHYO CUENCA DAULE GUAYAZUIL MANTA MEJÍA QUITO RI OBAMBA SAMBORONDÓN SANTA ELENA

SQL SELECT provincia, actividad, habitaciones, camas PROP. datos WHERE actividad = "ALOJAMIENTO" GROUP BY provincia, actividad, habitaciones, camas

25k 20k 15k 10k 5k 0

camas(suon) habitaciones(suon)

AMBA TO AZOGUES BABAHYO CUENCA DAULE GUAYAZUIL MANTA MEJÍA QUITO RI OBAMBA SAMBORONDÓN SANTA ELENA

SQL SELECT provincia, actividad, habitaciones, camas PROP. datos WHERE actividad = "ALOJAMIENTO" GROUP BY provincia, actividad, habitaciones, camas

25k 20k 15k 10k 5k 0

camas(suon) habitaciones(suon)

AMBA TO AZOGUES BABAHYO CUENCA DAULE GUAYAZUIL MANTA MEJÍA QUITO RI OBAMBA SAMBORONDÓN SANTA ELENA

SQL SELECT provincia, actividad, habitaciones, camas PROP. datos WHERE actividad = "ALOJAMIENTO" GROUP BY provincia, actividad, habitaciones, camas

25k 20k 15k 10k 5k 0

camas(suon) habitaciones(suon)

AMBA TO AZOGUES BABAHYO CUENCA DAULE GUAYAZUIL MANTA MEJÍA QUITO RI OBAMBA SAMBORONDÓN SANTA ELENA

SQL SELECT provincia, actividad, habitaciones, camas PROP. datos WHERE actividad = "ALOJAMIENTO" GROUP BY provincia, actividad, habitaciones, camas

25k 20k 15k 10k 5k 0

camas(suon) habitaciones(suon)

AMBA TO AZOGUES BABAHYO CUENCA DAULE GUAYAZUIL MANTA MEJÍA QUITO RI OBAMBA SAMBORONDÓN SANTA ELENA

SQL SELECT provincia, actividad, habitaciones, camas PROP. datos WHERE actividad = "ALOJAMIENTO" GROUP BY provincia, actividad, habitaciones, camas

25k 20k 15k 10k 5k 0

camas(suon) habitaciones(suon)

AMBA TO AZOGUES BABAHYO CUENCA DAULE GUAYAZUIL MANTA MEJÍA QUITO RI OBAMBA SAMBORONDÓN SANTA ELENA

SQL SELECT provincia, actividad, habitaciones, camas PROP. datos WHERE actividad = "ALOJAMIENTO" GROUP BY provincia, actividad, habitaciones, camas

25k 20k 15k 10k 5k 0

camas(suon) habitaciones(suon)

AMBA TO AZOGUES BABAHYO CUENCA DAULE GUAYAZUIL MANTA MEJÍA QUITO RI OBAMBA SAMBORONDÓN SANTA ELENA

SQL SELECT provincia, actividad, habitaciones, camas PROP. datos WHERE actividad = "ALOJAMIENTO" GROUP BY provincia, actividad, habitaciones, camas

25k 20k 15k 10k 5k 0

camas(suon) habitaciones(suon)

AMBA TO AZOGUES BABAHYO CUENCA DAULE GUAYAZUIL MANTA MEJÍA QUITO RI OBAMBA SAMBORONDÓN SANTA ELENA

SQL SELECT provincia, actividad, habitaciones, camas PROP. datos WHERE actividad = "ALOJAMIENTO" GROUP BY provincia, actividad, habitaciones, camas

25k 20k 15k 10k 5k 0

camas(suon) habitaciones(suon)

AMBA TO AZOGUES BABAHYO CUENCA DAULE GUAYAZUIL MANTA MEJÍA QUITO RI OBAMBA SAMBORONDÓN SANTA ELENA

SQL SELECT provincia, actividad, habitaciones, camas PROP. datos WHERE actividad = "ALOJAMIENTO" GROUP BY provincia, actividad, habitaciones, camas

25k 20k 15k 10k 5k 0

camas(suon) habitaciones(suon)

AMBA TO AZOGUES BABAHYO CUENCA DAULE GUAYAZUIL MANTA MEJÍA QUITO RI OBAMBA SAMBORONDÓN SANTA ELENA

SQL SELECT provincia, actividad, habitaciones, camas PROP. datos WHERE actividad = "ALOJAMIENTO" GROUP BY provincia, actividad, habitaciones, camas

25k 20k 15k 10k 5k 0

camas(suon) habitaciones(suon)

AMBA TO AZOGUES BABAHYO CUENCA DAULE GUAYAZUIL MANTA MEJÍA QUITO RI OBAMBA SAMBORONDÓN SANTA ELENA

SQL SELECT provincia, actividad, habitaciones, camas PROP. datos WHERE actividad = "ALOJAMIENTO" GROUP BY provincia, actividad, habitaciones, camas

25k 20k 15k 10k 5k 0

camas(suon) habitaciones(suon)

AMBA TO AZOGUES BABAHYO CUENCA DAULE GUAYAZUIL MANTA MEJÍA QUITO RI OBAMBA SAMBORONDÓN SANTA ELENA

SQL SELECT provincia, actividad, habitaciones, camas PROP. datos WHERE actividad = "ALOJAMIENTO" GROUP BY provincia, actividad, habitaciones, camas

25k 20k 15k 10k 5k 0

camas(suon) habitaciones(suon)

AMBA TO AZOGUES BABAHYO CUENCA DAULE GUAYAZUIL MANTA MEJÍA QUITO RI OBAMBA SAMBORONDÓN SANTA ELENA

SQL SELECT provincia, actividad, habitaciones, camas PROP. datos WHERE actividad = "ALOJAMIENTO" GROUP BY provincia, actividad, habitaciones, camas

25k 20k 15k 10k 5k 0

camas(suon) habitaciones(suon)

AMBA TO AZOGUES BABAHYO CUENCA DAULE GUAYAZUIL MANTA MEJÍA QUITO RI OBAMBA SAMBORONDÓN SANTA ELENA

SQL SELECT provincia, actividad, habitaciones, camas PROP. datos WHERE actividad = "ALOJAMIENTO" GROUP BY provincia, actividad, habitaciones, camas

25k 20k 15k 10k 5k 0

camas(suon) habitaciones(suon)

AMBA TO AZOGUES BABAHYO CUENCA DAULE GUAYAZUIL MANTA MEJÍA QUITO RI OBAMBA SAMBORONDÓN SANTA ELENA

SQL SELECT provincia, actividad, habitaciones, camas PROP. datos WHERE actividad = "ALOJAMIENTO" GROUP BY provincia, actividad, habitaciones, camas

25k 20k 15k 10k 5k 0

camas(suon) habitaciones(suon)

¿Cuáles son las potenciales subactividades económicas que se podría realizar en cada provincia?

Ecuador es un país en continuo desarrollo y que brinda muchas oportunidades de trabajo o emprendimiento si la persona sabe explotar las carencias de negocios o las necesidades del lugar, si alguna persona quiere invertir en un campo en específico en el Ecuador tenemos la información que necesita, ya que aquí se puede observar las



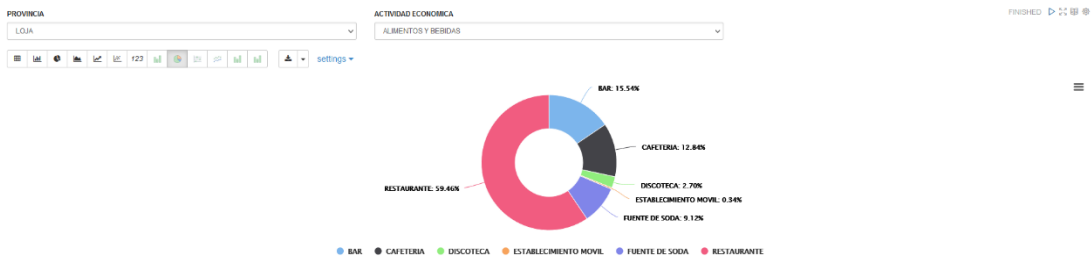
actividades económicas de cada provincia y la cantidad de estas para así buscar una opción atractiva para desarrollar.

¿CUÁLES SON LAS POTENCIALES SUB ACTIVIDADES ECONOMICAS QUE SE PODRIAN REALIZAR EN CADA PROVINCIA?

FINISHED

El resultado que se obtiene como se puede apreciar en la grafica es muy variado, por ejemplo si yo soy un ciudadano que reside en la provincia de pichincha y quiero invertir en un local de alimentos y bebidas me doy cuenta que lo que mas me conviene sería un bar puesto que es el que con menos presencia cuenta.

Took 2 sec. Last updated by anonymous at July 31 2022, 9:57:07 PM.

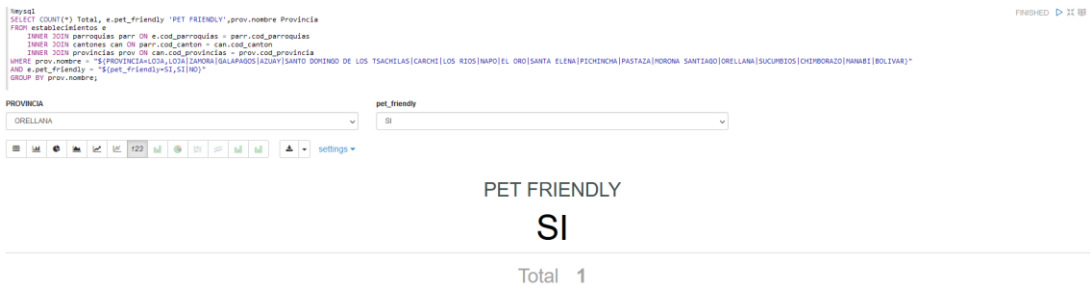


Cuantos establecimientos admiten mascotas por provincia y cuantos no admiten?

FINISHED

Se recolectó esta información de los negocios que si tenían página web válida ademas de consultar tambien en otras fuentes, debido a que muchos establecimientos no especifican esta información muchos negocios no pudieron ser registrados en la base de datos.

Took 0 sec. Last updated by anonymous at July 31 2022, 9:57:40 PM.



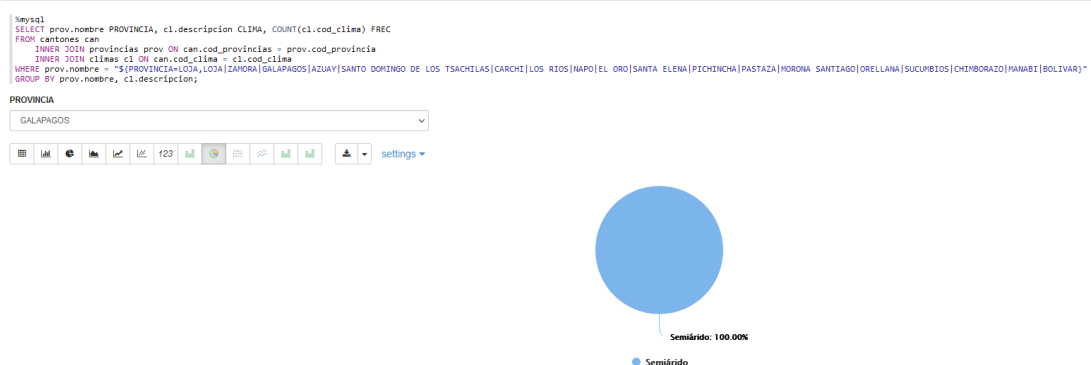
¿Está listo para invertir en una actividad turística en la intemperie? nosotros le informamos en qué provincia le convendría iniciar su negocio basándonos en el clima

Ecuador es un país megadiverso por lo cual es importante conocer sus climas, dependiendo si vamos a viajar a la Costa, Sierra, Oriente o Región Insular, todas las personas sean turistas o no, necesitan saber el clima para llevar toda la ropa y equipo necesario para disfrutar de mejor manera su estadía en cualquier provincia del país, basados en los climas de cada provincia de la siguiente página web <https://es.db-city.com/Ecuador>, para tener mucha información del tema.

Ecuador es un país megadiverso por lo cual es importante conocer sus climas, dependiendo si vamos a viajar a la Costa, Sierra, Oriente o Región Insular, todas las personas sean turistas o no, necesitan saber el clima para llevar toda la ropa y equipo necesario para disfrutar de mejor manera su estadía en cualquier provincia del país.

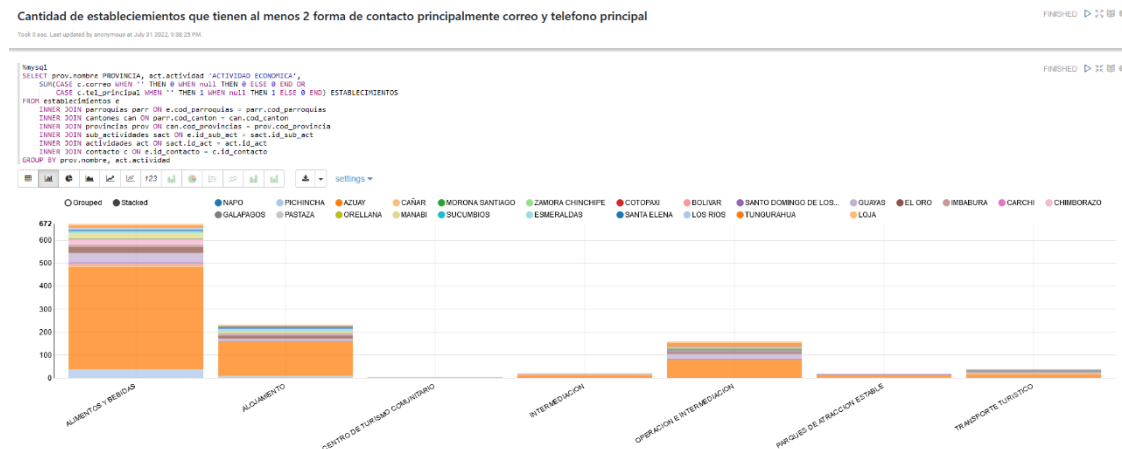
¿Está listo para invertir en una actividad turística en la intemperie? nosotros le informamos en que provincia le convendría iniciar su negocio basándonos en el clima

Took 0 sec. Last updated by anonymous at July 31 2022, 9:58:06 PM.



## Cantidad de establecimientos que tienen al menos 2 forma de contacto principalmente correo y teléfono principal

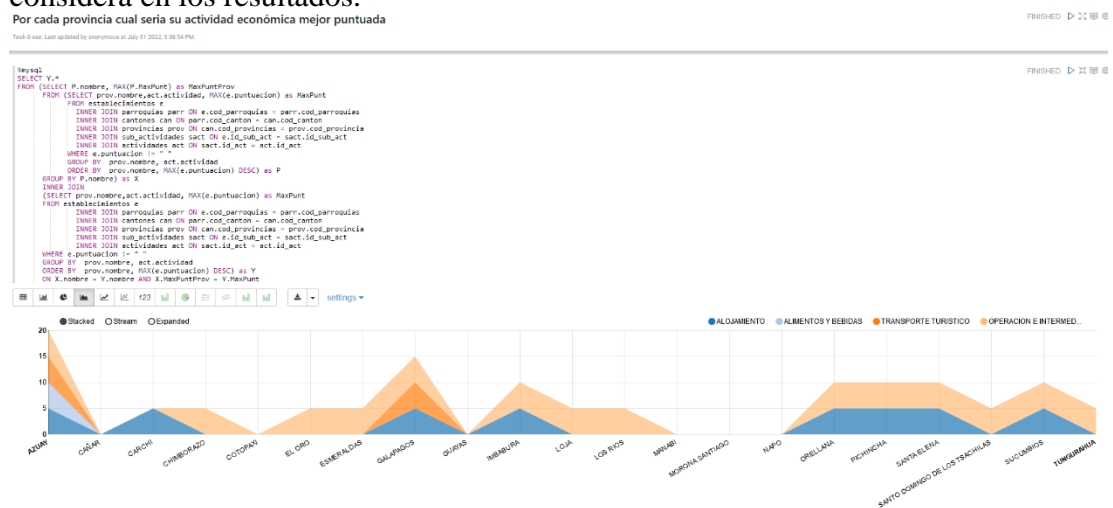
En esta gráfica podemos visualizar que la Actividad Económica con al menos un correo y un teléfono de contacto es la Categoría de Alimentos y Bebidas, específicamente la Provincia del Azuay es la que más establecimientos con escasa información.



## Actividades mejor aceptadas por los turistas a nivel provincial

Ecuador país donde los ingresos económicos son variados gracias a la mega diversidad que posee, si usted son de aquellos que disfrutan de los mejores servicios que cada provincia tiene por ofrecer, les presentamos a continuación cuáles son las actividades que más resaltan en cuanto a su aceptación por los turistas en cada provincia el Ecuador, cabe resaltar que estos resultados se basan en los establecimientos que poseen página web, por lo que si no posee página web, no se considera en los resultados.

Por cada provincia cual sería su actividad económica mejor puntuada



## ¿Para realizar un viaje a Ecuador? Entérese cuales son las lenguas más habladas por los guías turísticos de nuestro país

Al momento de viajar a un lugar nuevo lo primero en lo que piensas es en que tiene el lugar, donde se encuentra los lugares de interés y que actividades puedes realizar para conocer y aprender de la cultura de ahí, es por eso que es importante saber si hay

¿Planea realizar un viaje a Ecuador? Entérese de cuales son las lenguas más habladas por los guías turísticos de nuestro país

FINISHED ▶ 20 88

Took 0 sec. Last updated by anonymous at July 31 2022, 9:39:14 PM



Y la evidencia que presentamos anteriormente demuestra que en algunos datos contenían errores en las columnas de categorías que contenían dos datos en una sola celda que se lograron solucionar mediante el componente de programación con el uso de comandos de spark en Zeppelin, además se incluyeron variables nuevas las cuales no contenía la data inicial, estas variables nos permitieron realizar nuevos reportes estadísticos sobre el consolidado nacional 2022.