

UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA



**ANÁLISIS DE LOS SINIESTROS VEHICULARES EN
ECUADOR EN EL 2023**

DOCENTE:

ING. DANILO JARAMILLO

POR:

JUAN DIEGO GUERRERO CAMARGO

GABRIELA ELIZABETH LOYOLA ROMERO

LUIS ALONSO SARANGO ROMERO

FECHA: 25/07/2025

ASIGNATURA: PROGRAMACIÓN AVANZADA

Contenido

Contenido.....	2
Introducción.....	5
Repositorio GitHub:.....	5
Datos Base:	5
Dataset Principal	5
Vehiculos Matriculados 2023.....	5
Dataset secundarios.....	8
Sinistralidad vial por el mes del año 2023.....	8
Victimas por Provincia del 2023	9
Catalogo de Mes.....	10
Catalogo de Provincia del Ecuador	11
Datos Complementarios.....	12
Diseño lógico relacional	12
Estructura:	13
Herramientas utilizadas.....	15
Sistema operativo Ubuntu.....	15
Virtual Box (Máquina virtual)	16
Mysql.....	16
Zeppelin.....	16

Spark	16
Mysql Workbench.....	16
Configuración del entorno	16
Instalación de la Máquina Virtual	16
Instalación de Mysql.....	17
Instalación de MySQL Workbench.....	17
Instalación de Zeppelin.....	18
Instalación de Spark	18
Interprete Zeppelin configuración de Mysql, Spark	19
Configuración de Mysql en Interprete de Zepellin.....	19
Análisis	19
Visualización de consultas	19
Consulta 1 - Total de siniestros por provincia.....	19
Consulta 2 - Siniestros por mes (nacional).....	20
Consulta 3 - Índice de víctimas por siniestro por provincia.....	21
Consulta 4 - Tipos de vehículos más frecuentes por provincia	22
Consulta 5 - Relación entre número de vehículos y número de siniestros por provincia.....	23
Consulta 6 - ¿Qué tipo de vehículos predominan en provincias con muchos siniestros?.....	25
Consulta 7 - ¿Hay relación entre el tipo de combustible y accidentes?	26

Consulta 8 - ¿Qué meses tienen más siniestros? **¡Error! Marcador no definido.**

Consulta 9 - Provincias con muchos vehículos, pero pocos siniestros.....	27
Interpretaciones	28
Consulta 1 - Total de siniestros por provincia.....	28
Posibles soluciones.....	35
Consulta 1 - Total de siniestros por provincia.....	35
Consulta 2 - Siniestros por mes (nacional).....	36
Consulta 3 - Índice de víctimas por siniestro por provincia.....	37
Consulta 4 - Tipos de vehículos más frecuentes por provincia	38
Consulta 5 - Relación entre número de vehículos y número de siniestros por provincia.....	39
Consulta 6 - ¿Qué tipo de vehículos predominan en provincias con muchos siniestros?	40
Consulta 7 - ¿Hay relación entre el tipo de combustible y accidentes?	40
Consulta 8 - ¿Qué meses tienen más siniestros? ¡Error! Marcador no definido.	
Consulta 9 - Provincias con muchos vehículos, pero pocos siniestros.....	41
Conclusión	42
Bibliografía	45

Introducción

El presente informe trata sobre un análisis de la siniestralidad de los vehículos a nivel nacional del país Ecuador en el año 2023 a partir de una muestra de 15.000 vehículos matriculados. Realizar análisis de datos sobre algún tema permite generar puntos de introspección para solucionar o prevenir problemas a futuro, por ello, es relevante realizar este tipo de proyectos. Este informe se justifica en la necesidad de aplicar habilidades prácticas en el área de análisis de datos y con ello llevar a cabo conclusiones y soluciones que mejoren la vialidad en Ecuador. Se abordaron las diferentes aplicaciones usadas para la construcción del proyecto y su instalación, también el diagrama relacional que se creó para trabajar con las tablas, el uso de Zeppelin para analizar y extraer los datos, con su posterior interpretación.

Repositorio GitHub:

Repositorio: <https://github.com/Lxuxs/PROYENCTO2B.git>

Datos Base:

Los datos base contienen información detallada sobre accidentes de tránsito, víctimas y vehículos matriculados en Ecuador durante 2023, desglosada por provincia, mes y características de los vehículos. A continuación, se presenta una descripción de las variables presentes en los datos:

Dataset Principal

Vehiculos Matriculados 2023

1	id	PROVINCIA	MARCA	CLASE	PASAJEROS	TONELAJE	COMBUSTIBLE	MODELO	SERVICIO	ESTRATONE	ESTRAPASAJERO	MES
2	1	1	HYUNDAI	1	5	0,75	2	2011	3	1	1	7
3	2	17	JEEP	6	5	0,75	2	1111	3	1	1	11
4	3	9	CHEVROLET	1	5	0,75	2	2010	3	1	1	7
5	4	13	CHEVROLET	6	5	1	2	1111	3	1	1	9
6	5	9	SHINERAY	7	2	0,25	2	2022	3	1	1	9
7	6	24	IGM	7	2	0,25	2	2020	3	1	1	7
8	7	17	TOYOTA	6	5	0,75	2	2024	3	1	1	9
9	8	17	CHEVROLET	1	5	0,5	2	2012	3	1	1	9
10	9	3	CHEVROLET	4	2	1,5	1	1111	3	1	1	3
11	10	2	HONDA	7	2	0,25999999	2	2020	3	1	1	8
12	11	9	KIA	1	5	0,75	2	2020	3	1	1	12
13	12	9	HONDA	7	2	0,25	2	2019	3	1	1	4
14	13	17	HYUNDAI	5	11	0,829999983	1	2024	3	1	2	12
15	14	13	SUZUKI	1	5	0,75	2	1111	3	1	1	3
16	15	6	CHEVROLET	1	5	0,75	2	2013	3	1	1	11
17	16	23	LINGKEN	7	2	0,25	2	2015	3	1	1	8
18	17	17	HINO	3	2	8	1	1111	2	3	1	7
19	18	19	CHEVROLET	1	5	0,75	2	1111	3	1	1	12
20	19	1	FORD	4	2	1	2	1111	3	1	1	12
21	20	5	RANGER	7	2	0,25	2	2021	3	1	1	8
22	21	8	NISSAN	4	5	0,75	2	2011	3	1	1	7
23	22	23	CHEVROLET	4	2	0,75	1	2010	3	1	1	1
24	23	24	CHEVROLET	1	5	0,5	2	1111	3	1	1	12
25	24	17	CHEVROLET	1	5	0,5	2	2010	3	1	1	2
26	25	9	CHEVROLET	1	5	0,75	2	2015	2	1	1	3
27	26	9	DFSK	6	7	0,75	2	2023	3	1	1	2

Columna	Definición
id	Identificador único de cada registro en la tabla.
PROVINCIA	Código numérico de la provincia donde está registrado el vehículo.
MARCA	Marca del vehículo (por ejemplo: HYUNDAI, CHEVROLET, TOYOTA, etc.).
CLASE	Código de clase del vehículo (tipo de vehículo según clasificación vehicular).

PASAJEROS	Número de pasajeros que el vehículo puede transportar según su diseño.
TONELAJE	Capacidad de carga en toneladas del vehículo.
COMBUSTIBLE	Código del tipo de combustible utilizado (por ejemplo: 1 = gasolina, 2 = diésel, etc.).
MODELO	Año del modelo del vehículo (o 1111 si no está registrado).
SERVICIO	Código del tipo de servicio (por ejemplo: 3 = servicio particular).
ESTRATONE	Código de estrato del tonelaje (puede ser usado para clasificación en análisis).
ESTRAPASAJERO	Código de estrato del número de pasajeros (puede ser usado para clasificación en análisis).

Dataset secundarios

Siniestralidad vial por el mes del año 2023

1	provincia_id	anio	mes_id	num_siniestros
2	1	2023	1	82
3	1	2023	2	93
4	1	2023	3	80
5	1	2023	4	85
6	1	2023	5	107
7	1	2023	6	70
8	1	2023	7	101
9	1	2023	8	102
10	1	2023	9	101
11	1	2023	10	123
12	1	2023	11	144
13	1	2023	12	109
14	2	2023	1	11
15	2	2023	2	14
16	2	2023	3	8
17	2	2023	4	12
18	2	2023	5	15
19	2	2023	6	13
20	2	2023	7	12
21	2	2023	8	7
22	2	2023	9	15
23	2	2023	10	13
24	2	2023	11	17
25	2	2023	12	16
26	3	2023	1	3
27	3	2023	2	9

Columna	Definición
provincia_id	Identificador de la provincia donde ocurrió el siniestro. Es un valor numérico que representa cada provincia de forma única.

anio	Año en el que se registró el siniestro (en este caso, 2023).
mes_id	Número del mes en el que ocurrió el siniestro (1 = enero, 2 = febrero, etc.).
num_siniestros	Número total de siniestros registrados en esa provincia durante ese mes y año.

Victimas por Provincia del 2023

	provincia_id	anio	total_victimas	fallecidos	lesionados
1					
2	1	2023	1286	82	1204
3	2	2023	204	33	171
4	3	2023	107	44	63
5	4	2023	43	20	23
6	5	2023	256	89	167
7	6	2023	366	101	265
8	7	2023	644	100	544
9	8	2023	245	81	164
10	9	2023	8499	583	7916
11	10	2023	328	61	267
12	11	2023	483	41	442
13	12	2023	1332	168	1164
14	13	2023	1322	179	1143
15	14	2023	98	20	78
16	15	2023	29	13	16
17	16	2023	50	11	39
18	17	2023	3134	438	2696
19	18	2023	726	83	643
20	19	2023	82	23	59
21	20	2023	13	2	11
22	21	2023	33	16	17
23	22	2023	134	40	94
24	23	2023	940	113	827
25	24	2023	624	32	592

Columna	Definición
provincia_id	Identificador numérico de la provincia donde ocurrieron los siniestros.
anio	Año en el que se registraron los datos (en este caso, 2023).
total_victimas	Número total de víctimas en siniestros de tránsito en esa provincia durante el año.
fallecidos	Número de personas fallecidas en siniestros de tránsito en esa provincia durante el año.
lesionados	Número de personas lesionadas en siniestros de tránsito en esa provincia durante el año.

Catalogo de Mes

id_mes	mes
1	Enero
2	Febrero
3	Marzo
4	Abril
5	Mayo
6	Junio
7	Julio
8	Agosto
9	Septiembre
10	Octubre
11	Noviembre
12	Diciembre

Columna	Definición
id_mes	Identificador numérico del mes (1 = enero, 2 = febrero, ..., 12 = diciembre).
mes	Nombre del mes correspondiente al id_mes.

Catalogo de Provincia del Ecuador

provincia_id	Provincia
1	Azuay
2	Bolivar
3	Canar
4	Carchi
5	Cotopaxi
6	Chimborazo
7	El Oro
8	Esmeraldas
9	Guayas
10	Imbabura
11	Loja
12	Los Rios
13	Manabi
14	Morona Santiago
15	Napo
16	Pastaza
17	Pichincha
18	Tungurahua
19	Zamora Chinchipe
20	Galapagos
21	Sucumbios
22	Orellana
23	Santo Domingo de los Tsachilas
24	Santa Elena

Columna	Definición
provincia_id	Identificador numérico único de cada provincia.

Provincia	Nombre de la provincia correspondiente al provincia_id.
-----------	--

Estos datos, junto con los análisis que se realicen, serán almacenados en un repositorio de GitHub para facilitar su revisión de aquellos interesados en el tema. El repositorio permitirá mantener un registro histórico de los cambios realizados, fomentando así la transparencia y la reproducción de los resultados obtenidos.

Datos Complementarios

Vehículos Matriculados 2023

Datos obtenidos de <https://www.ecuadorencifras.gob.ec/vehiculos-matriculados/>

Siniestralidad vial por el mes del año 2023

Datos obtenidos de <https://www.ecuadorencifras.gob.ec/transporte/>

Victimas por Provincia del 2023

Datos obtenidos de <https://www.ecuadorencifras.gob.ec/siniestros-de-transito-i-trimestre-2023/>

Catálogo de códigos de provincias de Ecuador

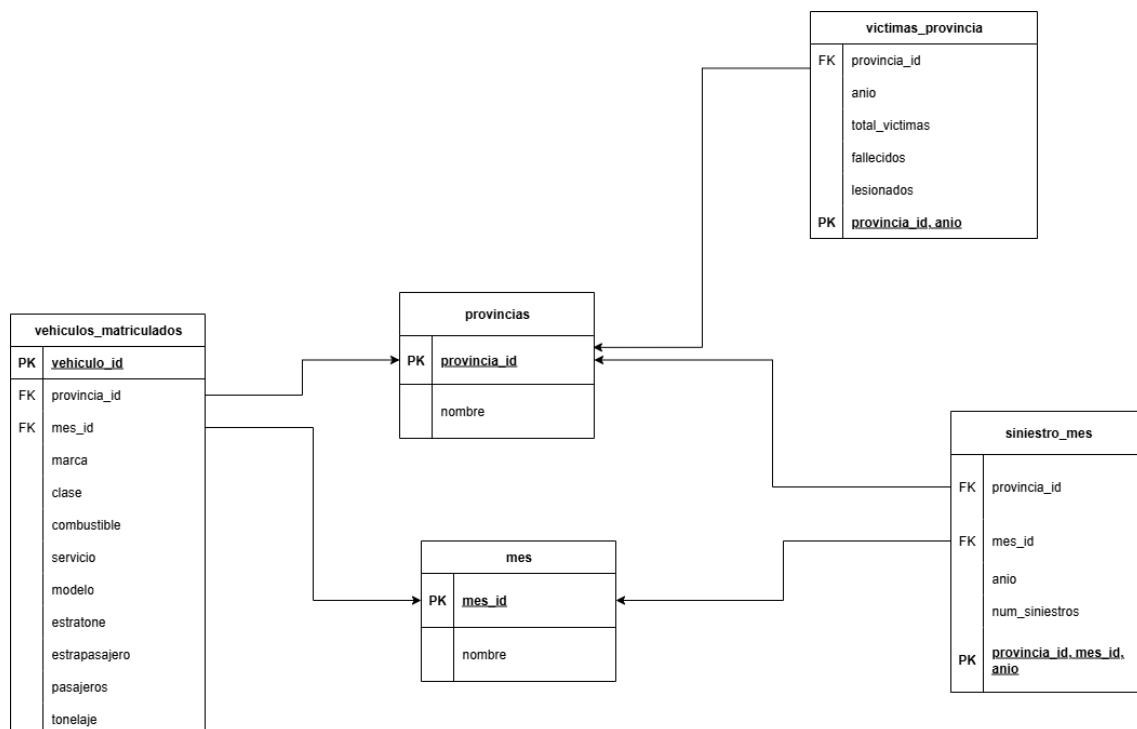
Datos obtenidos de <https://www.ecuadorlegalonline.com/consultas/registro-civil/codigos-provincia-cedula-identidad/>

Diseño lógico relacional

Este diagrama relacional muestra la estructura de un sistema para gestionar información sobre siniestros viales y víctimas a nivel de provincia y mes.

Estructura:

- Tablas principales:
 - provincia: contiene el listado de provincias con su identificador y nombre.
 - mes: almacena los 12 meses del año.
- Tablas relacionadas:
 - victimas_provincia: registra el total de víctimas, fallecidos y lesionados en accidentes por provincia y año.
 - siniestro_mes: detalla el número de siniestros registrados por provincia, mes y año.
 - Vehiculos_matriculados: almacena información sobre las características de vehículos registrados (marca, clase, combustible, etc.) vinculados a provincia y opcionalmente a mes.



Diccionario de datos

ENTIDAD	ATRIBUTOS	DESCRIPCIÓN
provincias	<ul style="list-style-type: none"> provincia_id nombre 	<ul style="list-style-type: none"> Identificador de la provincia. Nombre de la provincia.
mes	<ul style="list-style-type: none"> mes_id nombre 	<ul style="list-style-type: none"> Identificador del mes. Nombre del mes.
victimas_provincia	<ul style="list-style-type: none"> provincia_id año total_victimas fallecidos lesionados 	<ul style="list-style-type: none"> Identificador de la provincia. Año de siniestro que dejó víctimas Total de víctimas en la provincia por siniestros de tránsito. Total de fallecidos en la provincia por siniestros de tránsito. Total de lesionados en la provincia por siniestros de tránsito.
siniestro_mes	<ul style="list-style-type: none"> provincia_id mes_id Año Num_siniestros 	<ul style="list-style-type: none"> Identificador de la provincia. Identificador del mes. Año de siniestro.

		<ul style="list-style-type: none"> • Numero de siniestros.
Vehiculos_matriculados	<ul style="list-style-type: none"> • vehiculo_id • provincia_id • mes_id • marca • clase • combustible • servicio • modelo • estratone • estrapasajero • pasajeros • tonelaje 	<ul style="list-style-type: none"> • Identificador del vehiculo. • Identificador de la provincia. • Identificador del mes de matriculación. • Marca del vehiculo. • Clase del vehiculo. • Tipo de combustible del vehiculo. • Tipo de modelo del vehiculo. • Grupos de capacidad de carga. • Grupos de capacidad de pasajeros. • Capacidad de carga.

Herramientas utilizadas

Sistema operativo Ubuntu

Ubuntu es una distribución Linux ampliamente adoptada, valorada por su seguridad, estabilidad y facilidad de uso. Como explica Vidal (2023), su diseño enfocado

en el usuario y su filosofía de código abierto la convierten en una opción atractiva para millones de personas en todo el mundo.

Virtual Box (Máquina virtual)

VirtualBox permite ejecutar sistemas operativos dentro de tu sistema como máquinas virtuales.

Mysql

MySQL Server es un servidor de bases de datos que te permite almacenar, organizar y gestionar datos.

Zeppelin

Zeppelin es un cuaderno interactivo para análisis de datos y visualización, usado con Spark y SQL.

Spark

Apache Spark es una herramienta de procesamiento de datos a gran velocidad, muy usada en análisis de Big Data.

Mysql Workbench

MySQL Workbench es una interfaz gráfica para gestionar bases de datos MySQL, permitiéndote crear tablas y realizar consultas de forma visual.

Configuración del entorno

Instalación de la Máquina Virtual

- Descarga VirtualBox desde su página oficial.

- Instala con “Siguiente → Instalar → Finalizar”.
- Descarga el Extension Pack e instálalo para habilitar soporte USB.
- Descarga la imagen ISO de Ubuntu desde su página oficial.
- Abre VirtualBox y crea una máquina virtual:
- Selecciona tipo “Linux” y versión “Ubuntu (64-bit)”.
- Asigna entre 2 a 4 GB de memoria RAM.
- Asigna 25 GB de espacio en disco.
- Inicia la máquina virtual y realiza la instalación de Ubuntu en ella.

Instalación de Mysql

- Actualiza los paquetes del sistema:
`sudo apt update`
- Instala el servidor de bases de datos MySQL:
`sudo apt install mysql-server -y`
- Configura la seguridad de MySQL:
`sudo mysql_secure_installation`
- Verifica que MySQL funcione correctamente para gestionar bases de datos.

Instalación de MySQL Workbench

- Actualiza los paquetes del sistema:
`sudo apt update`
- Instala MySQL Workbench:
`sudo apt install mysql-workbench -y`

- Abre MySQL Workbench desde el menú de aplicaciones para crear tablas y consultas de forma gráfica.

Instalación de Zeppelin

- Instala Java, necesario para ejecutar Zeppelin:
`sudo apt install openjdk-17-jdk -y`
- Descarga Zeppelin desde su página oficial.
- Descomprime el archivo descargado:
`tar -xvf zeppelin-0.11.0-bin-all.tgz`
- Ingresa a la carpeta descomprimida.
- Inicia Zeppelin:
`bin/zeppelin-daemon.sh start`
- Abre en el navegador: <http://localhost:8080> para usar Zeppelin y visualizar datos.

Instalación de Spark

- Instala Java, necesario para Spark:
`sudo apt install openjdk-17-jdk -y`
- Descarga Spark desde su página oficial.
- Descomprime el archivo descargado:
`tar -xvf spark-3.5.1-bin-hadoop3.tgz`
- Ingresa a la carpeta descomprimida.
- Inicia Spark en consola:
`./bin/spark-shell`

- Spark está listo para procesar y analizar datos en Ubuntu.

Interprete Zeppelin configuración de Mysql, Spark

Configuración de Mysql en Interprete de Zepellin

- Permite consultar bases de datos MySQL directamente desde Zeppelin.
- Se usa el interprete JDBC de Zeppelin para conectar con MySQL.
- Pasos generales:
 1. Ingresar a Zeppelin en <http://localhost:8080>.
 2. Ir a Interpreter.
 3. Buscar jdbc y hacer clic en edit.
 4. Configurar:
 - URL: `jdbc:mysql://localhost:3306/nombre_base`
 - Usuario: `root` u otro usuario MySQL.
 - Contraseña: la configurada en MySQL.
 5. Guardar y reiniciar el intérprete.
 6. Ahora se pueden escribir consultas SQL directamente en Zeppelin para ver tablas y gráficas.

Análisis

Visualización de consultas

Consulta 1 - Total de siniestros por provincia

Esta consulta mide el volumen total de siniestros viales registrados en el año 2023 por provincia.

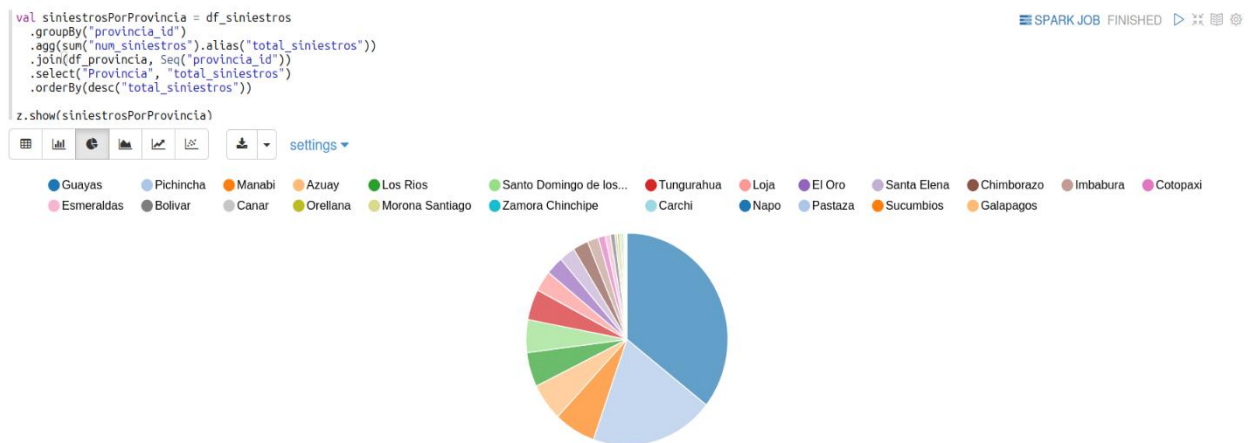
Código utilizado:

```

val siniestrosPorProvincia = df_siniestros
  .groupBy("provincia_id")
  .agg(sum("num_siniestros").alias("total_siniestros"
))
  .join(df_provincia, Seq("provincia_id"))
  .select("Provincia", "total_siniestros")
  .orderBy(desc("total_siniestros"))
  z.show(siniestrosPorProvincia)

```

Grafico:



Consulta 2 - Siniestros por mes (nacional)

Esta consulta muestra cuántos siniestros ocurrieron en cada mes durante a lo largo del año 2023 en todo el país.

Código utilizado:

```

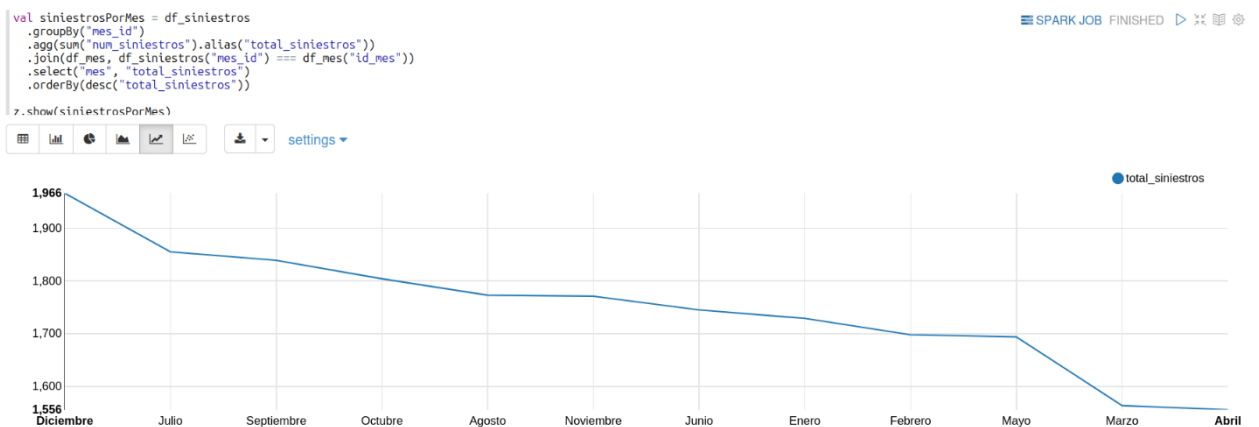
val siniestrosPorMes = df_siniestros .groupBy("mes_id")
    .agg(sum("num_siniestros").alias("total_siniestros"))
    .join(df_mes,          df_siniestros("mes_id")          ===
df_mes("id_mes"))

    .select("mes", "total_siniestros")

    .orderBy(desc("total_siniestros"))

z.show(siniestrosPorMes)

```



Consulta 3 - Índice de víctimas por siniestro por provincia

Esta consulta dice cuántas víctimas (lesionados + fallecidos) hay, en promedio, por cada siniestro, en las diferentes provincias.

Código utilizado:

```

    val siniestrosPorProvinciaAgg = df_siniestros
      .groupBy("provincia_id")
      .agg(sum("num_siniestros").alias("siniestros_total"))

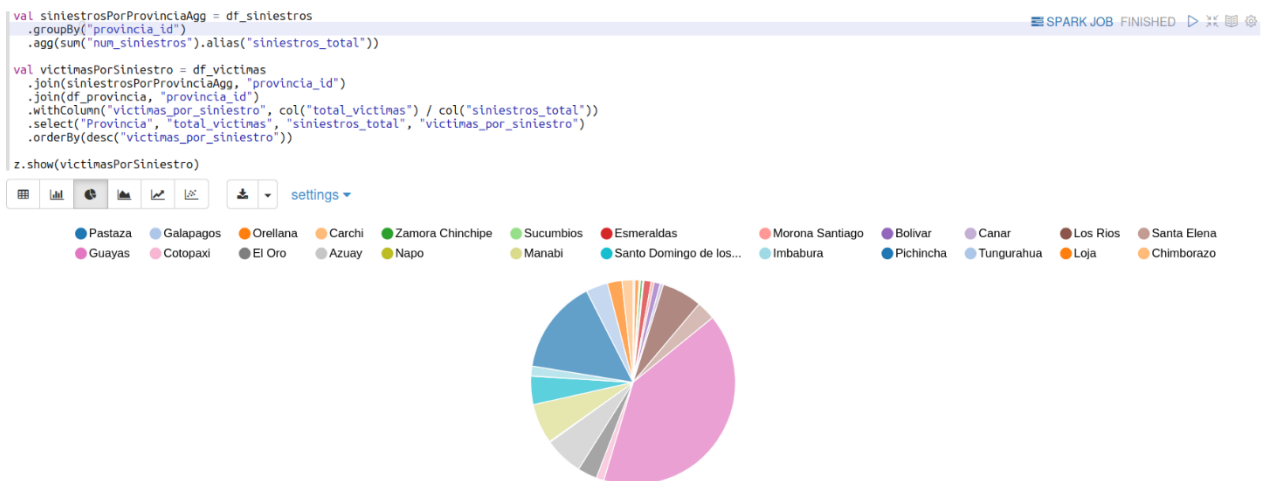
    val      victimasPorSiniestro      =      df_victimas
      .join(siniestrosPorProvinciaAgg, "provincia_id")

      .join(df_provincia,
        "provincia_id").withColumn("victimas_por_siniestro",
          col("total_victimas") / col("siniestros_total"))
      .select("Provincia", "total_victimas", "siniestros_total",
        "victimas_por_siniestro")
      .orderBy(desc("victimas_por_siniestro"))

    z.show(victimasPorSiniestro)

```

Grafico:



Consulta 4 - Tipos de vehículos más frecuentes por provincia

Esta consulta muestra, por provincia, qué tipo de vehículo es más común.

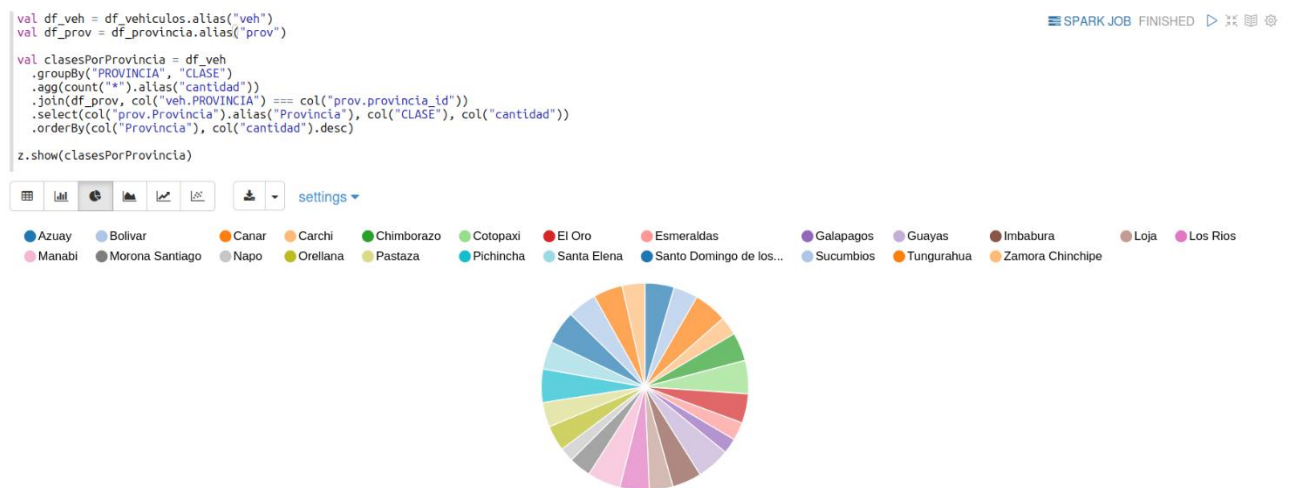
Código utilizado:

```

val df_veh = df_vehiculos.alias("veh")
val df_prov = df_provincia.alias("prov")
val clasesPorProvincia = df_veh
    .groupBy("PROVINCIA", "CLASE")
    .agg(count("*").alias("cantidad"))
    .join(df_prov, col("veh.PROVINCIA") ===
col("prov.provincia_id"))
    .select(col("prov.Provincia").alias("Provincia"),
col("CLASE"),
    col("cantidad"))
    .orderBy(col("Provincia"), col("cantidad").desc)
z.show(clasesPorProvincia)

```

Grafico:



Consulta 5 - Relación entre número de vehículos y número de siniestros por provincia

Esta consulta muestra, por provincia:

¿Cuántos siniestros ocurren por cada vehículo registrado?

¿Dónde los vehículos son proporcionalmente más propensos a verse involucrados en un accidente?

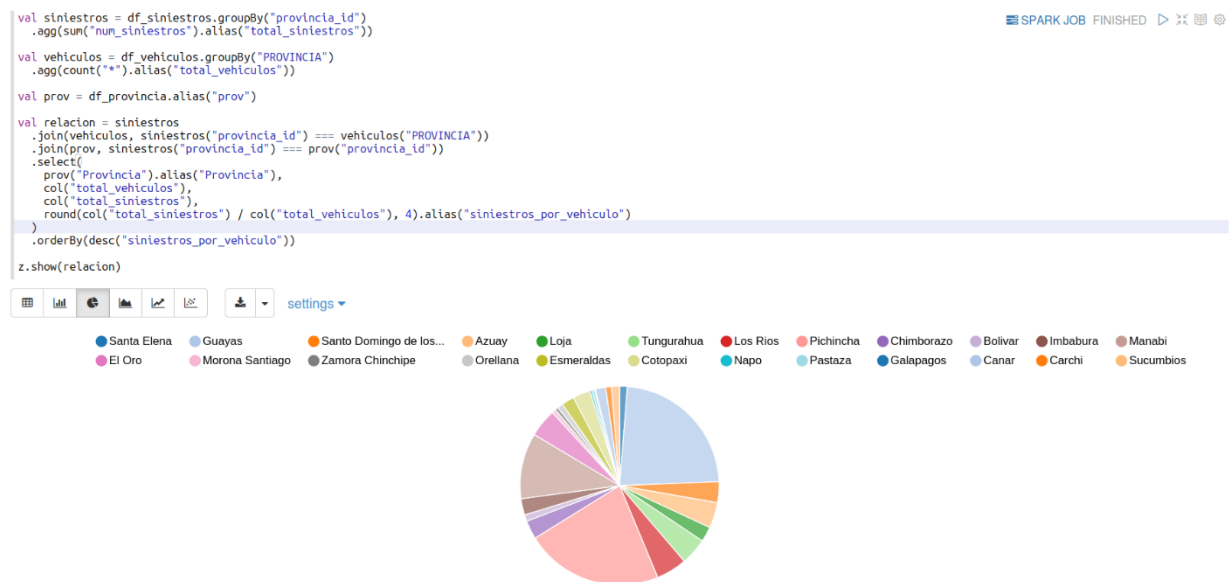
Código utilizado:

```
val siniestros =
df_siniestros.groupBy("provincia_id")
    .agg(sum("num_siniestros").alias("total_siniestros"))

val vehiculos =
df_vehiculos.groupBy("PROVINCIA")
    .agg(count("*").alias("total_vehiculos"))

val prov = df_provincia.alias("prov")
val relacion = siniestros
    .join(vehiculos, siniestros("provincia_id") ===
        vehiculos("PROVINCIA"))
    .join(prov, siniestros("provincia_id") ===
        prov("provincia_id"))
    .select( prov("Provincia").alias("Provincia"),
        col("total_vehiculos"),
        col("total_siniestros"),
        round(col("total_siniestros") /
            col("total_vehiculos"),
            4).alias("siniestros_por_vehiculo")
        )
    .orderBy(desc("siniestros_por_vehiculo"))
z.show(relacion)
```

Grafico:



Consulta 6 - ¿Qué tipo de vehículos predominan en provincias con muchos siniestros?

Esta consulta Permite ver, cuál es el tipo de vehículo más frecuente en las provincias con muchos siniestros.

Código utilizado:

```

val tiposPorProvincia = df_vehiculos
  .groupBy(col("PROVINCIA"), col("CLASE"))
  .agg(count("*").alias("cantidad")) .join(df_provincia,
    df_vehiculos("PROVINCIA") === df_provincia("provincia_id"))
  .select(df_provincia("Provincia"), df_vehiculos("CLASE"),
    col("cantidad"))
  .orderBy(df_provincia("Provincia"), col("cantidad").desc)
  z.show(tiposPorProvincia)

```

Grafico:



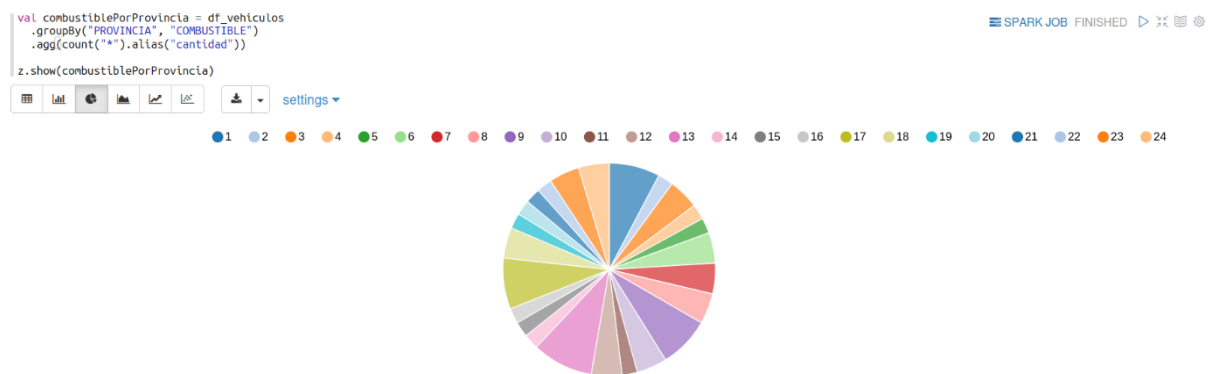
Consulta 7 - ¿Hay relación entre el tipo de combustible y accidentes?

Esta consulta cruza el tipo de combustible con la tasa de siniestros por vehículo.

Código utilizado:

```
val combustiblePorProvincia = df_vehiculos
  .groupBy("PROVINCIA", "COMBUSTIBLE")
  .agg(count("*").alias("cantidad"))
  z.show(combustiblePorProvincia)
```

Grafico:



Consulta 8 - Provincias con muchos vehículos, pero pocos siniestros

Esta consulta refleja las provincias con gran cantidad de vehículos, pero que tienen una tasa baja de siniestros por cada vehículo.

Código utilizado:

```
val buenasPracticas = relacion
  .filter(col("siniestros_por_vehiculo") < 0.50)
z.show(buenasPracticas)
```

Grafico:

```
val buenasPracticas = relacion.filter(col("siniestros_por_vehiculo") < 0.50)
z.show(buenasPracticas)
```

SPARK JOB FINISHED

Provincia	total_vehiculos	total_siniestros	siniestros_por_vehiculo
Pastaza	69	26	0.3768
Galapagos	19	7	0.3684
Canar	251	81	0.3227
Carchi	140	29	0.2071
Sucumbios	196	23	0.1173

Interpretaciones

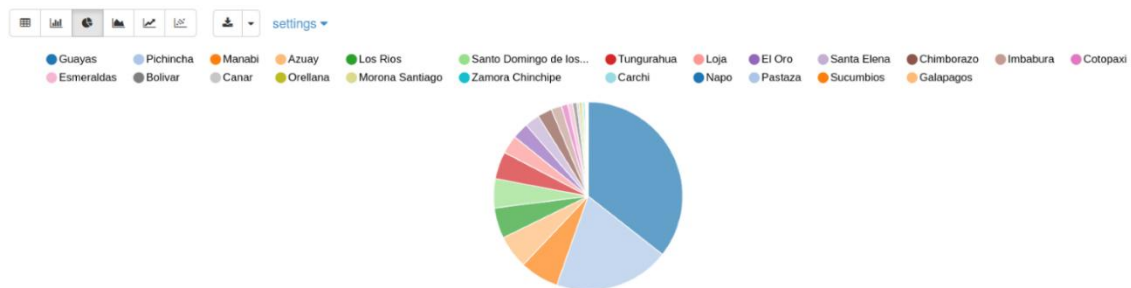
Consulta 1 - Total de siniestros por provincia

La consulta realizada muestra los siniestros viales del año 2023 por provincia. A partir del gráfico generado, se observa que las provincias con mayor cantidad de accidentes son las provincias de Guayas y Pichincha, lo cual resulta coherente considerando que allí se ubican ciudades como Guayaquil y Quito, capitales del país y que se caracterizan por su alta densidad poblacional y mayor cantidad de vehículos.

Por otro lado, provincias como Manabí, Azuay y Santo Domingo también presentan cifras elevadas, mientras que territorios con menor población o tránsito, como Galápagos o Zamora Chinchipe, registran pocos siniestros.

Si bien esta consulta no profundiza en las causas, sí permite identificar las zonas más afectadas por la siniestralidad vial, y permite dar a conocer los puntos más conflictivos a nivel nacional.

Grafico:



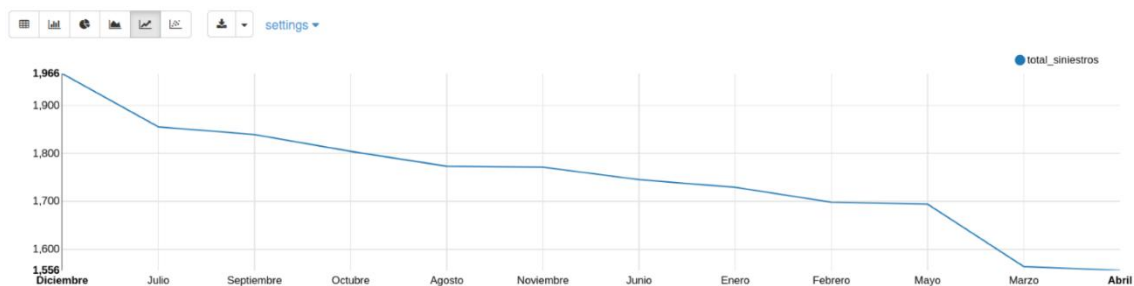
Consulta 2 - Siniestros por mes (nacional)

La consulta realizada refleja los siniestros viales registrados en el año 2023 divididos por meses. Los resultados se ven en un gráfico de líneas que permite identificar una tendencia general ascendente en el paso del tiempo en el año.

Diciembre fue el mes con mayor número de siniestros, seguido por julio y septiembre. Esto podría estar relacionado con el incremento del tráfico debido a periodos de feriados nacionales, vacaciones o eventos de fin de año. En contraste, marzo y abril presentan las cifras más bajas, posiblemente por factores como condiciones climáticas adversas o una mayor eficacia en los controles viales.

Los resultados permiten detectar los periodos de mayor riesgo, lo que puede servir de base para planificar campañas preventivas o reforzar medidas de seguridad vial en los meses más críticos.

Grafico:



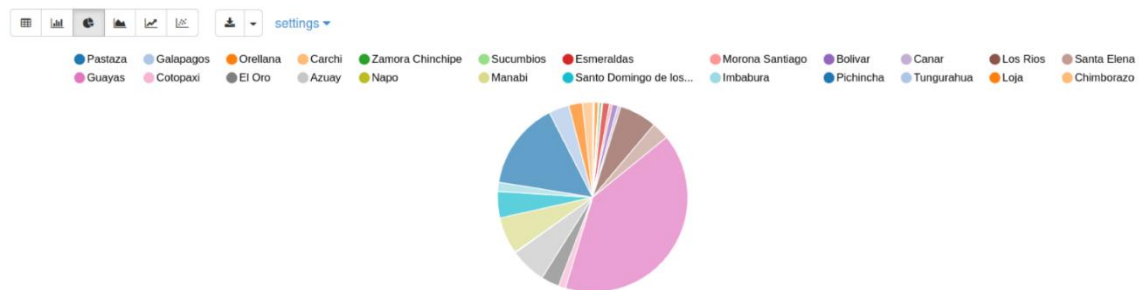
Consulta 3 - Índice de víctimas por siniestro por provincia

Esta consulta cruza el total de víctimas registradas con el total de siniestros por provincia, calcula un indicador llamado “víctimas por siniestro”, que muestra un promedio de cuántas personas resultan afectadas por accidente en cada provincia.

El gráfico revela que existe una relación entre las provincias con mayor cantidad de siniestros y su gravedad, como es el caso de Guayas y Pichincha. Esto indica que en esas zonas los accidentes suelen ser más graves o afectan a más personas.

Este tipo de análisis ayuda a identificar regiones donde no solo hay siniestros frecuentes, sino también más severos. Las causas podrían estar relacionadas con el tipo de vía, la velocidad con la que se circula, una señalización deficiente o acceso limitado a servicios de emergencia.

Grafico:



Consulta 4 - Tipos de vehículos más frecuentes por provincia

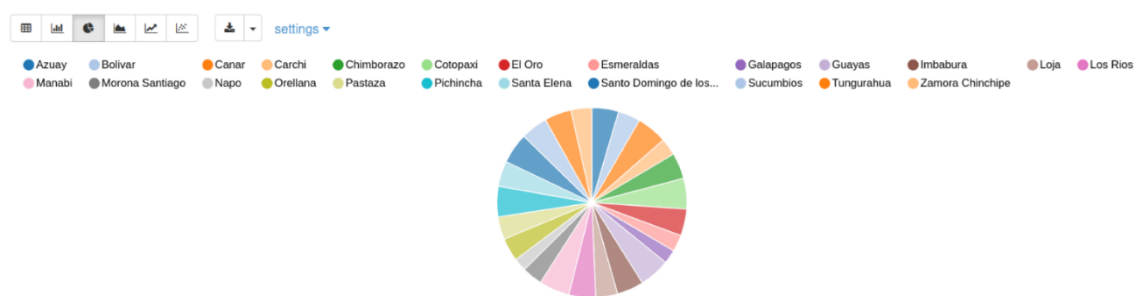
La consulta muestra cuántos vehículos están matriculados en cada una de las provincias, agrupados por diferentes clases (moto, auto particular, camión, etc.). El gráfico permite identificar qué tipo de vehículo predomina en cada provincia.

Las motos tienen una alta presencia en varias provincias, lo que refleja su uso como medio de movilidad económica, aunque también representan mayor vulnerabilidad. Los camiones destacan en zonas con fuerte actividad comercial o rutas logísticas, como Guayas o El Oro, lo que indica su relación con el transporte de carga. Por otro lado, los

autos particulares predominan en provincias urbanizadas como Pichincha y Azuay, donde hay mayor densidad poblacional.

Esta distribución permite entender el perfil de movilidad de cada provincia y puede relacionarse con el tipo y frecuencia de los siniestros analizados en consultas anteriores.

Grafico:



Consulta 5 - Relación entre número de vehículos y número de siniestros por provincia

Esta consulta cruza el número total de siniestros con el total de vehículos registrados por provincia. El objetivo es observar si existe alguna correlación entre la cantidad de vehículos en circulación y la ocurrencia de accidentes.

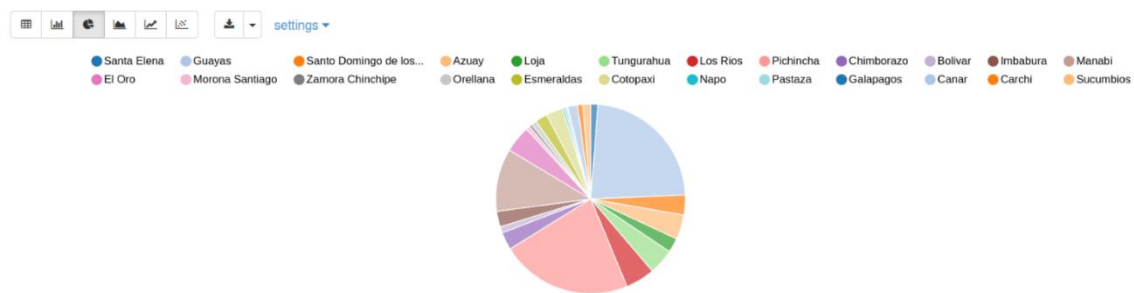
El gráfico muestra que provincias como Guayas y Pichincha concentran tanto una alta cantidad de vehículos como una alta cantidad de siniestros, lo cual es acorde a la realidad considerando su densidad poblacional, nivel de urbanización y volumen de tránsito diario.

Sin embargo, también se evidencian provincias donde, a pesar de tener un número relativamente bajo de vehículos, los siniestros son proporcionalmente altos, como lo son

Cotopaxi y Esmeraldas. Esto puede sugerir problemas en infraestructura vial, señalización deficiente o menor control del tránsito.

En general, esta relación ayuda a entender si la frecuencia de accidentes está asociada simplemente al número de vehículos o si hay otros factores estructurales o sociales involucrados.

Grafico:



Consulta 6 - ¿Qué tipo de vehículos predominan en provincias con muchos siniestros?

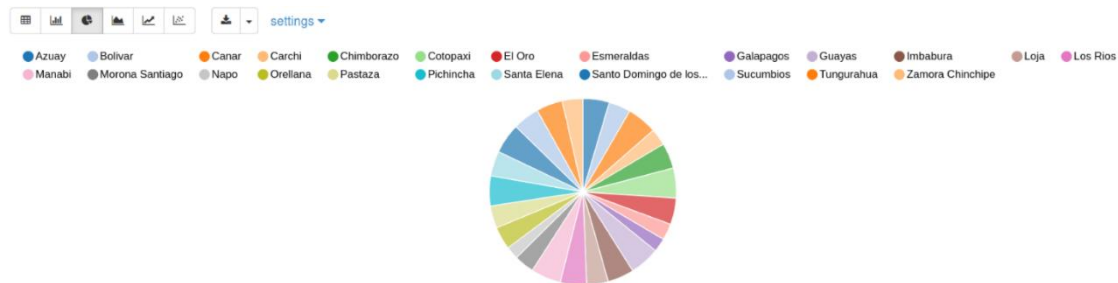
La consulta agrupa los vehículos por tipo y por provincia, mostrando cuántos hay de cada tipo en cada provincia. El gráfico permite comparar cómo varía el parque vehicular según la región.

Se observa que provincias como Guayas, Pichincha y Manabí presentan una mayor diversidad y cantidad de vehículos, lo cual está relacionado con su nivel de urbanización y actividad económica.

Por otro lado, provincias con menor densidad poblacional como Galápagos o Zamora Chinchipe tienen menos presencia vehicular, lo que también implica menos tipos de vehículos en circulación, considerando que tienen menor población.

Esta información es útil para entender el perfil de movilidad de cada provincia, ya que el tipo de vehículo predominante (motos, camiones, autos particulares, etc.) puede estar vinculado con el uso cotidiano, la economía local o la infraestructura vial disponible.

Grafico:



Consulta 7 - ¿Hay relación entre el tipo de combustible y accidentes?

La consulta agrupa cuántos vehículos hay por tipo de combustible en cada provincia, permitiendo observar si existe alguna relación indirecta con la cantidad de accidentes registrados.

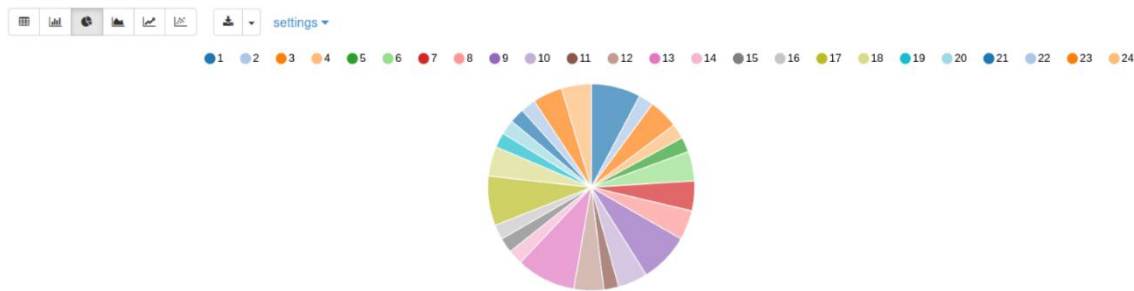
Aunque el gráfico no muestra directamente la conexión entre combustible y siniestros, se puede hacer una interpretación basada en el contexto:

- Vehículos a diésel, que suelen ser camiones o transporte pesado, están más presentes en provincias con actividad comercial o logística como Guayas, El Oro o Manabí, lo que podría influir en el tipo y gravedad de los siniestros registrados allí.
- Vehículos a gasolina, más comunes en zonas urbanas como Pichincha o Azuay, suelen ser autos particulares con alta circulación diaria, lo que incrementa la exposición al riesgo de estar en un siniestro.

- En cambio, vehículos eléctricos o híbridos aún tienen baja presencia, por lo que no se puede establecer un patrón claro con respecto a su relación con los accidentes.

En conclusión, aunque el tipo de combustible no causa directamente un accidente, sí puede estar relacionado con el tipo de vehículo y su función, lo cual influye en el comportamiento vial y en la dinámica de los siniestros por provincia.

Grafico:



Consulta 8 - Provincias con muchos vehículos, pero pocos siniestros

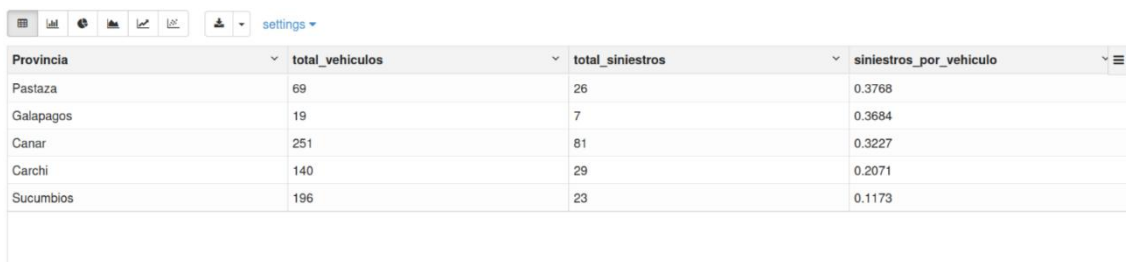
Esta consulta identifica provincias donde, a pesar de tener una cantidad considerable de vehículos matriculados, el número de siniestros registrados es bajo en proporción. Se utiliza como criterio que el promedio de siniestros por vehículo sea menor al 50%, ya que este porcentaje es óptimo para poder obtener algunas provincias y poder desarrollar una interpretación.

Provincias como Cañar, Sucumbíos y Carchi destacan porque, aunque cuentan con cientos de vehículos registrados, logran mantener los niveles de accidentes relativamente bajos. Esto podría estar relacionado con mejores prácticas viales, menor congestión vehicular, o una conducción más responsable.

Es importante resaltar que el porcentaje de vehículos involucrados en un siniestro de tránsito en Ecuador es bastante alto, teniendo en cuenta que Galápagos, el cual no se caracteriza por tener una gran cantidad de vehículos, tiene un porcentaje del 36%, y así se pueden encontrar más provincias que presentan el mismo caso.

Este análisis permite identificar zonas que, por sus resultados, podrían tomarse como ejemplo para otras provincias más conflictivas en términos de siniestralidad.

Grafico:



The image shows a screenshot of a web application interface. At the top, there is a toolbar with various icons (grid, bar chart, pie chart, line chart, etc.) and a 'settings' dropdown menu. Below the toolbar is a table with the following data:

Provincia	total_vehiculos	total_siniestros	siniestros_por_vehiculo
Pastaza	69	26	0.3768
Galapagos	19	7	0.3684
Canar	251	81	0.3227
Carchi	140	29	0.2071
Sucumbios	196	23	0.1173

Posibles soluciones

Consulta 1 - Total de siniestros por provincia

Teniendo en cuenta el problema de alta concentración de siniestros viales en ciertas provincias, estas serían algunas posibles soluciones, junto con los resultados esperados:

- Implementar campañas de concienciación vial en las zonas más afectadas, con el objetivo de educar a la población sobre normas de tránsito y riesgos. Esto puede reducir conductas imprudentes y mejorar la cultura vial.

- Reforzar los controles de tránsito y las sanciones en puntos críticos, lo que busca disuadir infracciones como el exceso de velocidad, el uso del celular al conducir o el consumo de alcohol, disminuyendo así la ocurrencia de accidentes.
- Mejorar la infraestructura vial y la señalización en áreas de alto riesgo, como cruces peligrosos o vías con mala visibilidad. Esto contribuiría a reducir errores y generar una circulación más segura.
- Instalar radares y sistemas inteligentes de monitoreo, que permiten detectar patrones de riesgo en tiempo real y actuar de forma preventiva. Se espera una reducción tanto en la frecuencia como en la gravedad de los siniestros.

En conjunto, estas medidas buscan no solo disminuir el número de accidentes, sino también mitigar su impacto, salvando vidas y mejorando la seguridad vial en las provincias más afectadas.

Consulta 2 - Siniestros por mes (nacional)

Teniendo en cuenta el incremento de siniestros en ciertos meses del año, como diciembre y julio, estas serían algunas posibles soluciones y los efectos esperados:

- Reforzar los operativos de tránsito y las acciones de prevención durante las épocas festivas y vacacionales. Esto permitiría aumentar la presencia en las vías, reducir infracciones y prevenir accidentes en momentos de alta circulación.

- Ajustar los horarios de circulación y mejorar la gestión del tráfico en temporada alta. Con una mejor organización del flujo vehicular, se espera disminuir la congestión y los puntos críticos que suelen derivar en siniestros.
- Promover campañas de conducción segura previas a los meses críticos. Estas campañas pueden generar mayor conciencia en la población sobre los riesgos de estas fechas, fomentando una conducción más responsable.

En conjunto, estas medidas buscan anticiparse a los periodos de mayor riesgo, reduciendo tanto la cantidad como la gravedad de los siniestros viales en los meses más críticos del año.

Consulta 3 - Índice de víctimas por siniestro por provincia

Teniendo en cuenta que algunas provincias presentan siniestros más graves, con un mayor número de víctimas por evento, se proponen las siguientes soluciones junto con sus efectos esperados:

- Reforzar el uso obligatorio de cinturón de seguridad y casco, mediante controles y campañas educativas. Esto puede reducir significativamente la gravedad de las lesiones en caso de accidente.
- Aumentar la capacidad de respuesta de los servicios de emergencia en zonas rurales o de difícil acceso. Una atención más rápida y eficiente puede disminuir la mortalidad y las secuelas tras un siniestro.

- Evaluar las condiciones de seguridad de las vías donde ocurren los accidentes más graves, con el fin de identificar factores de riesgo como curvas peligrosas, falta de iluminación o señalización deficiente. Mejorar estos puntos ayudaría a prevenir eventos severos.

Estas acciones buscan no solo reducir la cantidad de siniestros, sino también mitigar su impacto en las provincias más afectadas.

Consulta 4 - Tipos de vehículos más frecuentes por provincia

Teniendo en cuenta la alta presencia de motos y camiones en algunas zonas, lo cual incrementa el riesgo de siniestros, se plantean las siguientes soluciones con sus posibles efectos:

- Regular el uso de motocicletas en áreas con alta siniestralidad, mediante restricciones, controles y normas específicas. Esto podría disminuir la exposición al riesgo en zonas urbanas o de alta circulación.
- Establecer rutas específicas para camiones y controlar su circulación en horarios clave, especialmente en zonas escolares o en horas pico. Con ello se busca reducir conflictos con vehículos livianos y mejorar la fluidez del tránsito.
- Promover campañas de educación vial dirigidas específicamente a motociclistas, enfocadas en el uso de casco, visibilidad, respeto a las normas y conducción defensiva. Esto puede ayudar a disminuir tanto la frecuencia como la gravedad de los accidentes que involucran motos.

Estas medidas apuntan a gestionar mejor los vehículos de mayor riesgo, protegiendo tanto a sus conductores como al resto de los usuarios de la vía.

Consulta 5 - Relación entre número de vehículos y número de siniestros por provincia

Teniendo en cuenta que las provincias con mayor cantidad de vehículos tienden a concentrar también más siniestros viales, se proponen las siguientes soluciones junto con sus posibles efectos:

- Mejorar la planificación vial y la semaforización en ciudades densamente pobladas. Esto puede optimizar el flujo vehicular, reducir puntos de conflicto y mejorar la seguridad en intersecciones.
- Estimular el uso del transporte público mediante mejoras en la cobertura, frecuencia y comodidad del servicio. Al reducir la cantidad de vehículos particulares en circulación, se disminuye la probabilidad de accidentes y la congestión.
- Fomentar el uso de aplicaciones y tecnologías de movilidad que permitan gestionar rutas, evitar zonas con tráfico pesado y distribuir mejor el flujo de vehículos. Esto contribuye a una movilidad más eficiente y segura.

Estas medidas están orientadas a reducir el volumen de tráfico en zonas críticas y, con ello, disminuir tanto la frecuencia como la severidad de los siniestros en provincias con alta densidad vehicular.

Consulta 6 - ¿Qué tipo de vehículos predominan en provincias con muchos siniestros?

Teniendo en cuenta que las zonas con predominio de vehículos diésel (de carga) o a gasolina (particulares) pueden enfrentar diferentes tipos de riesgo vial y ambiental, se proponen las siguientes soluciones:

- Promover la transición hacia vehículos eléctricos o híbridos, especialmente en entornos urbanos. Esto no solo reduce las emisiones contaminantes, sino que también puede mejorar la eficiencia y seguridad en la circulación.
- Establecer planes de mantenimiento obligatorio para vehículos diésel de carga, con el fin de garantizar su buen estado mecánico y reducir el riesgo de fallos en ruta que puedan provocar siniestros.
- Aplicar incentivos tributarios, financieros o mecánicos para fomentar la renovación del parque vehicular. Con vehículos más modernos y seguros, se espera disminuir tanto la siniestralidad como el impacto ambiental.

Estas medidas buscan adaptar las políticas viales al tipo de vehículo predominante en cada zona, mejorando la seguridad y la sostenibilidad del sistema de transporte.

Consulta 7 - ¿Hay relación entre el tipo de combustible y accidentes?

Teniendo en cuenta que algunas zonas presentan un predominio de vehículos diésel (de carga) o a gasolina (particulares), es importante considerar que cada tipo implica riesgos distintos, tanto en términos de seguridad como de impacto ambiental. A partir de esto, se proponen las siguientes soluciones:

- Promover la transición hacia vehículos eléctricos o híbridos en entornos urbanos. Esto contribuiría a reducir la contaminación, el ruido y, en algunos casos, mejorar el rendimiento y la seguridad vial.
- Crear planes de mantenimiento obligatorio para vehículos diésel de carga, con controles técnicos regulares que aseguren su buen estado. Esto puede disminuir el riesgo de fallas mecánicas y accidentes asociados al mal estado del vehículo.
- Aplicar incentivos tributarios, financieros o técnicos para renovar el parque vehicular, priorizando la salida de circulación de unidades antiguas y menos seguras. Esto favorecería una flota más moderna, eficiente y menos riesgosa.

Estas medidas buscan ajustar las políticas de movilidad según el perfil vehicular de cada zona, apuntando a una circulación más segura, limpia y sostenible.

Consulta 8 - Provincias con muchos vehículos, pero pocos siniestros

Existen provincias que demuestran que es posible mantener altos niveles de movilidad con baja siniestralidad, lo que representa un modelo positivo a seguir.

A partir de esta observación, se proponen las siguientes acciones:

- Documentar y analizar las políticas públicas exitosas implementadas en estas provincias, como educación vial, buena señalización, fiscalización efectiva y gestión del tránsito.

- Utilizar estos casos como referencia en capacitaciones interprovinciales, fomentando el intercambio de experiencias y aprendizajes entre gobiernos locales.
- Establecer un reconocimiento nacional para las provincias que destaquen por sus buenas prácticas viales, con el fin de motivar la mejora continua y la replicación de modelos exitosos.
- Incentivar la adopción de estrategias similares en otras regiones mediante programas piloto de movilidad segura, adaptados al contexto local.

Estas medidas permiten aprovechar las experiencias positivas ya existentes como punto de partida para mejorar la seguridad vial a nivel nacional.

Conclusión

En este proyecto grupal, el equipo trabajó con una muestra de 15.000 vehículos matriculados y los siniestros ocurridos en Ecuador durante el año 2023. Para el análisis se usaron archivos CSV que fueron importados a una base de datos mysql, donde posteriormente se usó Apache Zeppelin como herramienta principal para el análisis. Con consultas en Spark, los datos se organizaron, relacionaron y visualizaron desde distintos enfoques: por provincia, por mes, por tipo de vehículo, tipo de combustible y sus diferentes cruces para la mayor extracción de información.

Durante el desarrollo del proyecto, el grupo aplicó los conocimientos aprendidos en clases, como el uso de MySQL y las funciones de Zeppelin. Gracias a esto, se pudo

obtener una visión más clara sobre los accidentes de tránsito en el país e identificar patrones importantes.

Uno de los hallazgos más interesantes fue que las provincias con más población y actividad urbana, como Guayas y Pichincha, concentran la mayor cantidad de accidentes debido al gran tráfico y población que tienen. Sin embargo, también se encontraron provincias con muchos vehículos registrados pero con menos accidentes de lo esperado, lo que podría indicar que en esa zona hay buenas prácticas de conducción que valdría la pena replicar en otras zonas.

Otro punto que llamó la atención es que el porcentaje de vehículos involucrados en accidentes en Ecuador es bastante alto. Un ejemplo es la provincia de Galápagos: aunque tiene pocos vehículos, presenta una tasa de siniestralidad del 36%. Esto demuestra que no siempre hay una relación directa entre la cantidad de autos y los accidentes, ya que también influyen factores como la educación vial, la infraestructura y el control de tránsito.

El análisis por meses mostró que los accidentes no ocurren de manera uniforme a lo largo del año. Diciembre, julio y septiembre fueron los meses con más siniestros, posiblemente por feriados o festividades. En cambio, marzo y abril tuvieron la menor cantidad de casos, que podría estar vinculado a cambios en el clima. Esto sugiere que se podrían implementar estrategias de prevención en los meses donde históricamente aumentan los accidentes.

También se identificaron provincias donde los siniestros tienden a ser más graves al comparar los accidentes con el número de víctimas. Además, el análisis por tipo de

vehículo y tipo de combustible reflejó diferencias que tienen relación con el perfil económico y urbano de cada región, dando una visión más completa de la movilidad en el país.

La comparación entre el total de accidentes y la cantidad de vehículos registrados permitió encontrar provincias con tasas de siniestros bajas. Estas podrían servir como ejemplo para crear modelos de movilidad segura que se apliquen en otras partes de Ecuador.

En conclusión, este proyecto mostró que el análisis de datos puede ser una herramienta muy útil para analizar problemas reales como la seguridad vial y ayudar a tomar decisiones basadas en gráficos e interpretaciones. Además, fue una buena oportunidad para que el equipo pusiera en práctica los conocimientos aprendidos en un análisis con impacto social.

Bibliografía

Oracle. (2023). *MySQL documentation*. <https://dev.mysql.com/doc/>

Apache Software Foundation. (2024). *Apache Spark – Lightning-fast unified analytics engine*. <https://spark.apache.org/>

Oracle. (2024). *Oracle VM VirtualBox user manual*. <https://www.virtualbox.org/manual/UserManual.html>

Canonical Ltd. (2023). *Ubuntu desktop guide*. <https://help.ubuntu.com/stable/ubuntu-help/index.html>

Vidal, S. (2023, 25 de agosto). *¿Qué es el sistema operativo Ubuntu?* Tecnobits – Guías del Campus. <https://tecnobits.com/que-es-el-sistema-operativo-ubuntu/>

Secretaría Técnica Planifica Ecuador. (s.f.). *Datos Abiertos Ecuador*. <https://datosabiertos.gob.ec/>

Oracle Corporation. (2025). *Oracle VM VirtualBox*. <https://www.virtualbox.org/>

Canonical Ltd. (s.f.). *Ubuntu*. Recuperado de <https://ubuntu.com/>