

# [ ANÁLISIS DE LOS SINIESTROS VEHICULARES EN ECUADOR EN EL 2023 FINISHED

Took 3 sec. Last updated by anonymous at July 25 2025, 7:20:03 AM.

## Integrantes:

FINISHED

- Juan Diego Guerrero Camargo
- Gabriela Elizabeth Loyola Romero
- Luis Alonso Sarango Romero

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:40 AM.

## Parametros de conexión con la base de datos

FINISHED

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:40 AM.

```
val jdbcUrl = "jdbc:mysql://localhost:3306/proyecto"
val connectionProperties = new java.util.Properties()
connectionProperties.setProperty("user", "root")
connectionProperties.setProperty("password", "Salamanca_0214")
connectionProperties.setProperty("driver", "com.mysql.cj.jdbc.Driver")
```

FINISHED

```
jdbcUrl: String = jdbc:mysql://localhost:3306/proyecto
connectionProperties: java.util.Properties = {user=root, password=Salamanca_0214, driver=com.mysql.cj.jdbc.Driver}
res57: Object = null
```

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:41 AM.

## Carga de la tabla vehiculos\_matriculados como DataFrame

FINISHED

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:41 AM.

```
val df_vehiculos = spark.read
  .jdbc(jdbcUrl, "vehiculos_matriculados", connectionProperties)

df_vehiculos.createOrReplaceTempView("vehiculos_matriculados")
```

FINISHED

```
df_vehiculos: org.apache.spark.sql.DataFrame = [id: int, PROVINCIA: int ... 10 more fields]
```

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:41 AM.

## Carga de la tabla provincia como DataFrame

FINISHED

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:41 AM.

```
val df_provincia = spark.read.jdbc(jdbcUrl, "provincia", connectionProperties)
df_provincia.createOrReplaceTempView("provincia")
```

FINISHED

```
df_provincia: org.apache.spark.sql.DataFrame = [provincia_id: int, Provincia: string]
```

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:42 AM.

## Carga de la tabla mes como DataFrame

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:42 AM.

```
val df_mes = spark.read.jdbc(jdbcUrl, "mes", connectionProperties)
df_mes.createOrReplaceTempView("mes")
```

FINISHED

df\_mes: org.apache.spark.sql.DataFrame = [id\_mes: int, mes: string]

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:42 AM.

## Carga de la tabla siniestro\_mes como DataFrame

FINISHED

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:42 AM.

```
val df_siniestros = spark.read.jdbc(jdbcUrl, "siniestro_mes", connectionProperties)
df_siniestros.createOrReplaceTempView("siniestro_mes")
```

FINISHED

df\_siniestros: org.apache.spark.sql.DataFrame = [provincia\_id: int, anio: int ... 2 more fields]

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:42 AM.

## Carga de la tabla victimas\_provincia como DataFrame

FINISHED

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:42 AM.

```
val df_victimas = spark.read.jdbc(jdbcUrl, "victimas_provincia", connectionProperties)
df_victimas.createOrReplaceTempView("victimas_provincia")
```

FINISHED

df\_victimas: org.apache.spark.sql.DataFrame = [provincia\_id: int, anio: int ... 3 more fields]

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:43 AM.

# Consultas

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:43 AM.

FINISHED

## Total de siniestros por provincia

FINISHED

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:43 AM.

```
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)
```

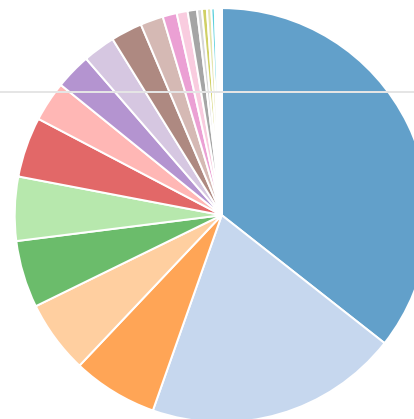
SPARK JOB FINISHED





settings ▼

- Guayas
- Pichincha
- Manabi
- Azuay
- Los Rios
- Santo Domingo de los...
- Tungurahua
- Loja
- El Oro
- Santa Elena
- Esmeraldas
- Bolivar
- Canar
- Orellana
- Morona Santiago
- Zamora Chinchipe
- Carchi
- Napo
- Pastaza
- Sucumbios



siniestrosPorProvincia: org.apache.spark.sql.Dataset[org.apache.spark.sql.Row] = [Provincia: string, total\_siniestros: bigint]

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:43 AM.

## Siniestros por mes (nacional)

FINISHED

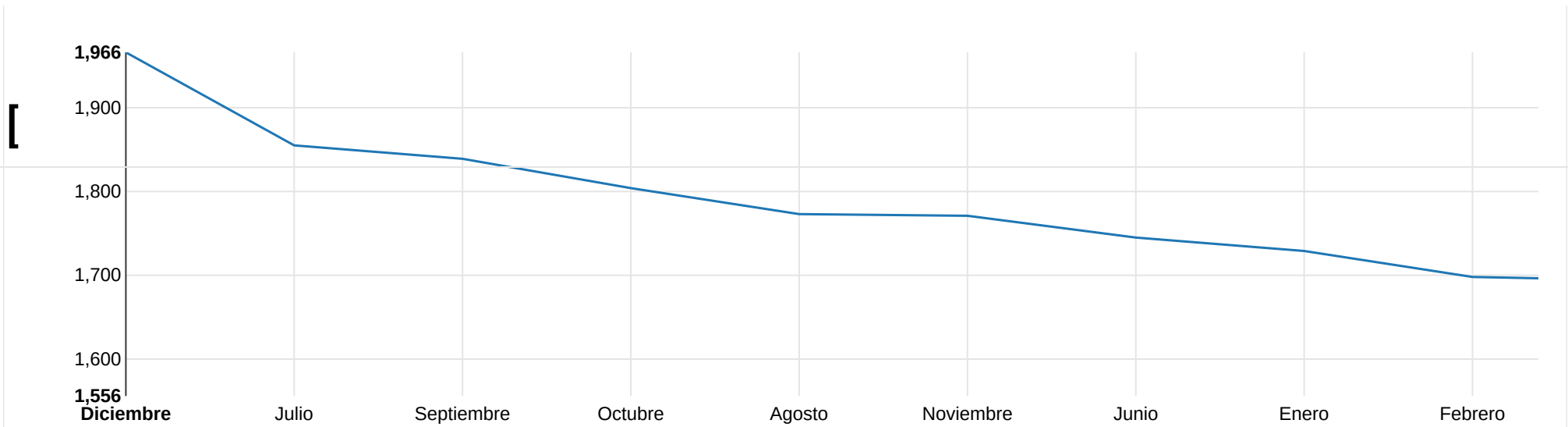
Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:43 AM.

```
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)
```

SPARK JOB FINISHED

settings



siniestrosPorMes: org.apache.spark.sql.Dataset[org.apache.spark.sql.Row] = [mes: string, total\_siniestros: bigint]

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:44 AM.

## Índice de víctimas por siniestro por provincia

FINISHED

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:44 AM.

```
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)
```

SPARK JOB FINISHED

settings ▾

Pastaza

Galapagos

Orellana

Carchi

Zamora Chinchipe

Sucumbios

Esmeraldas

Morona Santiago

Bolivar

Guayas

Cotopaxi

El Oro

Azuay

Napo

Manabi

Santo Domingo de los...

Imbabura

Pichincha

siniestrosPorProvinciaAgg: org.apache.spark.sql.DataFrame = [provincia\_id: int, siniestros\_total: bigint]  
 victimasPorSiniestro: org.apache.spark.sql.Dataset[org.apache.spark.sql.Row] = [Provincia: string, total\_victimas: int ... 2 more fields]

Took 1 sec. Last updated by anonymous at July 25 2025, 7:13:45 AM.

Tipos de vehículos más frecuentes por provincia

FINISHED

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:45 AM.

```

val df_veh = df_vehiculos.alias("veh")
val df_prov = df_provincia.alias("prov")

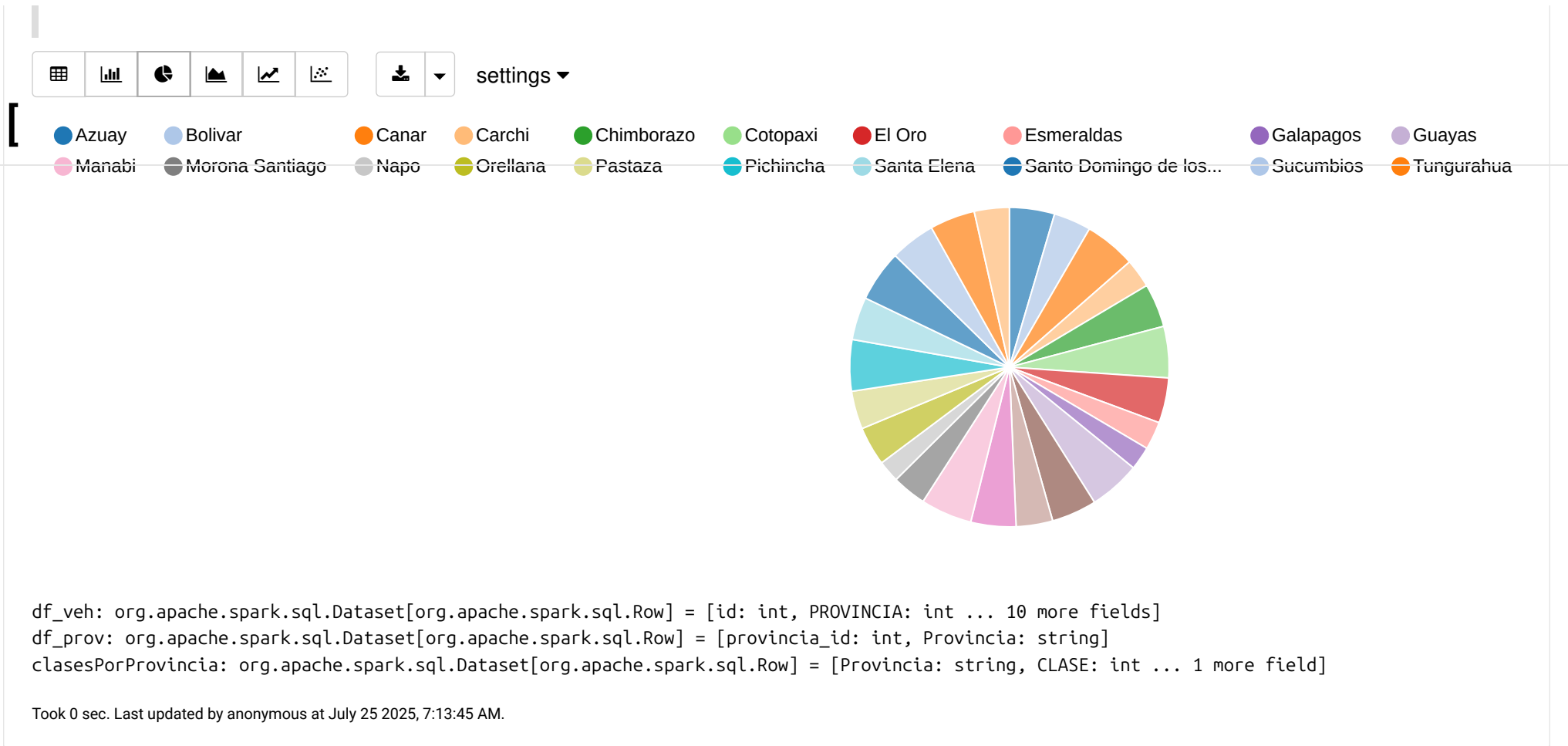
val clasesPorProvincia = df_veh
  .groupBy("PROVINCIA", "CLASE")
  .agg(count("*").alias("cantidad"))
  .join(df_prov, col("veh.PROVINCIA") === col("prov.provincia_id"))
  .select(col("prov.Provincia").alias("Provincia"), col("CLASE"), col("cantidad"))
  .orderBy(col("Provincia"), col("cantidad").desc)

z.show(clasesPorProvincia)
  
```

SPARK JOB FINISHED

7 of 17

7/25/25, 07:21



## Relación entre número de vehículos y número de siniestros por provincia

FINISHED

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:45 AM.

```
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")
```

SPARK JOB FINISHED



[

settings ▼

- Santa Elena

Guayas

Santo Domingo de los...

Azuay

Loja

Tungurahua

Los Rios

Pichincha

Chimboraz

El Oro

Morona Santiago

Zamora Chinchipe

Orellana

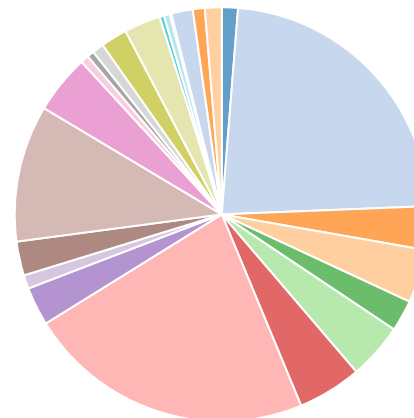
Esmeraldas

Cotopaxi

Napo

Pastaza

Galapagos



```
siniestros: org.apache.spark.sql.DataFrame = [provincia_id: int, total_siniestros: bigint]
vehiculos: org.apache.spark.sql.DataFrame = [PROVINCIA: int, total_vehiculos: bigint]
prov: org.apache.spark.sql.Dataset[org.apache.spark.sql.Row] = [provincia_id: int, Provincia: string]
relacion: org.apache.spark.sql.Dataset[org.apache.spark.sql.Row] = [Provincia: string, total_vehiculos: bigint ... 2 more fields]
```

Took 1 sec. Last updated by anonymous at July 25 2025, 7:13:46 AM.

## ¿Qué tipo de vehículos predominan en provincias con muchos siniestros?

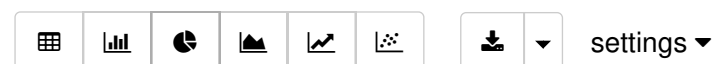
Took 0 sec. Last updated by anonymous at July 25 2025, 7:14:00 AM.

```
val dfv = df_vehiculos.alias("veh")
val dfp = df_provincia.alias("prov")

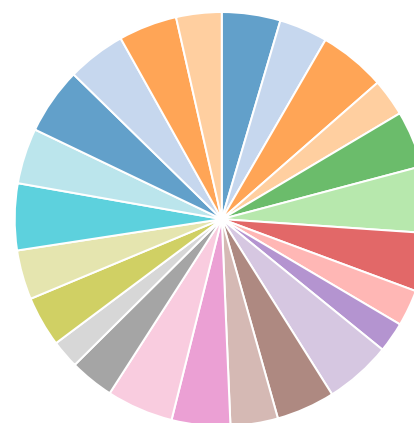
val tiposPorProvincia = dfv
  .groupBy(col("veh.PROVINCIA"), col("veh.CLASE"))
  .agg(count("*").alias("cantidad"))
  .join(dfp, col("veh.PROVINCIA") === col("prov.provincia_id"))
  .select(col("prov.Provincia").alias("Provincia"), col("veh.CLASE").alias("Clase"), col("cantidad"))
  .orderBy(col("Provincia"), col("cantidad").desc)

z.show(tiposPorProvincia)
```

SPARK JOB FINISHED



● Azuay   ● Bolivar   ● Canar   ● Carchi   ● Chimborazo   ● Cotopaxi   ● El Oro   ● Esmeraldas   ● Galapagos   ● Guayas  
 ● Manabi   ● Morona Santiago   ● Napo   ● Orellana   ● Pastaza   ● Pichincha   ● Santa Elena   ● Santo Domingo de los...   ● Sucumbios   ● Tungurahua



```
dfv: org.apache.spark.sql.Dataset[org.apache.spark.sql.Row] = [id: int, PROVINCIA: int ... 10 more fields]
dfp: org.apache.spark.sql.Dataset[org.apache.spark.sql.Row] = [provincia_id: int, Provincia: string]
tiposPorProvincia: org.apache.spark.sql.Dataset[org.apache.spark.sql.Row] = [Provincia: string, Clase: int ... 1 more field]
```

Took 1 sec. Last updated by anonymous at July 25 2025, 7:13:47 AM.

## ¿Hay relación entre el tipo de combustible y accidentes?

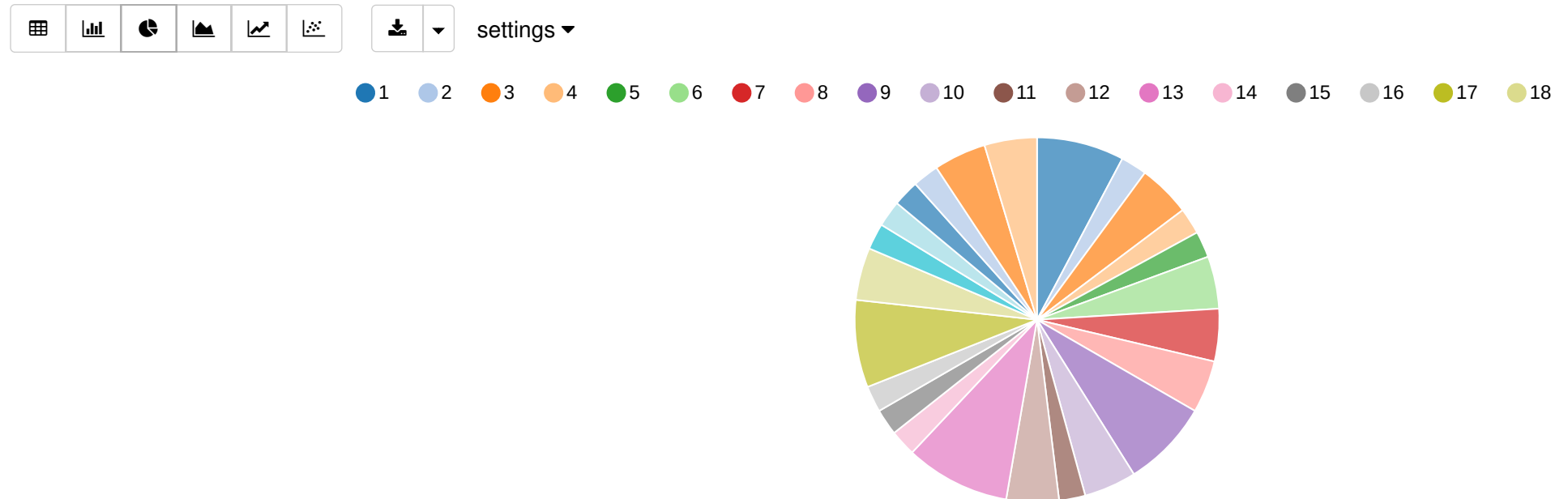
FINISHED

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:47 AM.

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

z.show(combustiblePorProvincia)
```

SPARK JOB FINISHED



combustiblePorProvincia: org.apache.spark.sql.DataFrame = [PROVINCIA: int, COMBUSTIBLE: int ... 1 more field]

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:47 AM.

## Provincias con muchos vehículos pero pocos siniestros

FINISHED

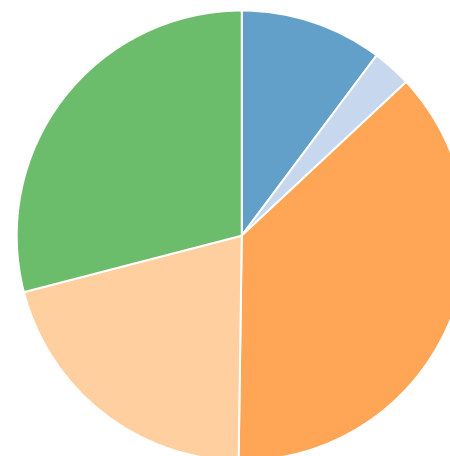
Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:47 AM.

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

SPARK JOB FINISHED

settings ▼

Pastaza Gale



buenasPracticas: org.apache.spark.sql.Dataset[org.apache.spark.sql.Row] = [Provincia: string, total\_vehiculos: bigint ... 2 more fields]

Took 1 sec. Last updated by anonymous at July 25 2025, 7:13:48 AM.

## Interpretaciones

FINISHED

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:48 AM.

FINISHED

## 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.

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:48 AM.

## Consulta 2 - Siniestros por mes (nacional)

FINISHED

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.

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:48 AM.

FINISHED

## 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.

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:48 AM.

## Consulta 4 - Tipos de vehículos más frecuentes por provincia

FINISHED

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.

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:48 AM.

FINISHED

## 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.

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:48 AM.

## 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.

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:48 AM.

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

FINISHED

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.

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:49 AM.

## Consulta 8 - Provincias con muchos vehículos, pero pocos siniestros

FINISHED

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 Galapagos, 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.

Took 0 sec. Last updated by anonymous at July 25 2025, 7:13:49 AM.