



**ESCUELA POLITÉCNICA NACIONAL**  
**ESCUELA DE FORMACIÓN DE TECNÓLOGOS**



**INFORME DEL PROYECTO DE FIN DE  
SEMESTRE**  
**ANÁLISIS DE DATOS**

**INTEGRANTES:**

Mario Endara, Danna Lopez, Handel Manobanda, Andrew Paillacho, Paul Sandoval

**PROFESOR:**

Ing. Juan Pablo Zaldumbide

**PERIODO ACADÉMICO:**

2024-B

**FECHA DE ENTREGA:**

04 de febrero del 2025

## 1. Definición del caso de estudio.

Los casos de estudio son ilustraciones o situaciones que representan problemas específicos reales, se analizan en contexto, desarrollo y solución.

## 2. Objetivos:

### 2.1 Objetivo General

Realizar el análisis cuantitativo y cuantitativo de los dataset, a través de conocimientos sobre extracción, limpieza, transformación y visualización de datos mediante Python y Power BI.

### 2.2 Objetivos específicos.

2.2.1 Realizar un análisis detallado de un dataset que muestre información precisa y concreta, a través de las herramientas de Power BI y Python, esto permitirá que se automatice de un 60 a 70 % el tiempo en comparación a un análisis manual o hacer uso de otras herramientas.

## 3. Descripción del equipo de trabajo y actividades realizadas por cada uno.

El equipo de trabajo ha elegido cinco temáticas de las cuales, cada una cuenta con una sola fuente de datos. Cada fuente (dataset) cuenta con abundantes registros, que son suficientes para el análisis y visualización de datos.

Las actividades que realizará cada integrante son:

- **Accidentes de tránsito en diferentes períodos y regiones:** Extracción del dataset desde Kaggle en formato CSV, limpieza, transformación y visualización de datos en Power BI por: Danna Lopez.

<https://www.kaggle.com/datasets/oktayrdeki/traffic-accidents>

- **Eventos deportivos a nivel mundial:** Extracción del dataset desde Kaggle en formato CSV, limpieza, transformación y visualización de datos en Power BI por: Mario Endara.

- **Conciertos:** Extracción del dataset desde Kaggle en formato CSV, limpieza, transformación y visualización de datos en Power BI por: Handel Manobanda.

<https://www.kaggle.com/datasets/nyphil/perf-history>

- **Registro de automóviles:** Extracción del dataset desde el INEC, limpieza, transformación y visualización de datos en Power BI por: Andrew Paillacho.

<https://www.ecuadorencifras.gob.ec/estadistica-de-transporte-bases-de-datos/>

## 4. Recursos y herramientas utilizadas.

### 5.1 Herramientas Utilizadas

**5.1.1 Kaggle:** Es una plataforma de competencia de ciencia de datos que permite a los usuarios buscar y publicar conjuntos de datos, explorar y crear modelos en un entorno de ciencia de datos basado en la web. Esta herramienta será utilizada para descargar los datasets en formato CSV para su respectivo análisis.

Fuente: <https://en.wikipedia.org/wiki/Kaggle>

**5.1.2 INEC (Instituto Nacional de Estadística y Censos):** El Instituto Nacional de Estadística y Censos como institución responsable de la estadística oficial, es la entidad encargada de recopilar, procesar, elaborar, analizar, publicar y difundir las estadísticas que contribuyan a la mejor información y a la solución de los distintos problemas de orden económico, social, demográfico y ambiental, que afrontan el Estado y los particulares.

**5.1.3 Jupyter Notebook:** Es una aplicación web, especialmente popular entre los científicos de datos, analistas y desarrolladores por su capacidad para facilitar el desarrollo, la documentación y la presentación de proyectos tanto de aprendizaje como profesionales. Con esta herramienta, se utilizarán las librerías necesarias de Python para cargar los datasets y hacer la limpieza de datos de los registros.

Fuente: <https://openwebinars.net/blog/jupyter-notebook/>

**5.1.4 Power BI:** Es un servicio de análisis de datos de Microsoft orientado a proporcionar visualizaciones interactivas, que ayuda a las empresas a tomar decisiones basadas en datos. Se utilizará para crear paneles de control (dashboards) y análisis de datos de manera rápida y sencilla.

Fuente: [https://es.wikipedia.org/wiki/Power\\_BI](https://es.wikipedia.org/wiki/Power_BI)

Power BI permite conectarse a una gran variedad de fuentes de datos, tanto locales como en la nube, como bases de datos SQL, archivos Excel, servicios en la nube y muchas otras.

**5.1.5 MongoDB Compass:** Es una interfaz gráfica de usuario para interactuar con MongoDB, una base de datos NoSQL que almacena datos en documentos JSON.

Fuente:

<https://colmenaenlanube.es/informatica/programacion/que-es-mongodb-compass/>

Esta herramienta será útil para cargar los datos, cuando se haga la respectiva transformación de un dataset (CSV a MySQL, MySQL a Mongo, etc).

**5.1.6 DuckDB:** Es un gestor de base de datos relacional que ha sido creado para que se pueda ejecutar dentro de cualquier aplicación. DuckDB está concebido para hacer

tareas de análisis de datos, donde la agregación de los datos de las columnas tiene un gran peso. También crea dashboards interactivos, lo que hace esta herramienta esencial en el área de inteligencia artificial, análisis estadístico y machine learning.

Una característica muy importante de DuckDB es que es muy rápido tanto cargando los datos como ejecutando las consultas SQL, porque utiliza una tecnología que optimiza el uso del CPU. Es muy fácil de implementar en Python, se instala por medio de la terminal y luego se importa la librería para establecer la conexión.

Fuente: <https://www.dataxbi.com/blog/2023/03/21/jugando-duckdb/>

**5.1.7 Redis (REmote DIctionary Server):** Es un almacén de claves/valores NoSQL de código abierto, en memoria, que se utiliza principalmente como caché de aplicaciones o base de datos de respuesta rápida.

## 5.2 Librerías utilizadas

**5.2.1 Pandas:** Es una librería especializada en el análisis de datos que ofrece estructuras de datos y operaciones para manipular tablas numéricas y series temporales, con etiquetas que permiten identificar los datos tabulados fácilmente. Pandas será esencial para la limpieza, filtración, transformación y preparación de los datos para su análisis.

Fuente:

[https://etrp.wmo.int/pluginfile.php/66419/mod\\_resource/content/1/PIBM2023\\_Python\\_y\\_lib\\_cientificas.pdf](https://etrp.wmo.int/pluginfile.php/66419/mod_resource/content/1/PIBM2023_Python_y_lib_cientificas.pdf)

**5.2.2 Pymongo:** Es una librería que actúa como un driver oficial para interactuar con bases de datos MongoDB desde Python de manera sencilla y eficiente. Con esta librería, se podrán crear bases de datos y colecciones en MongoDB Compass para cargar los datasets desde el formato en el que estén.

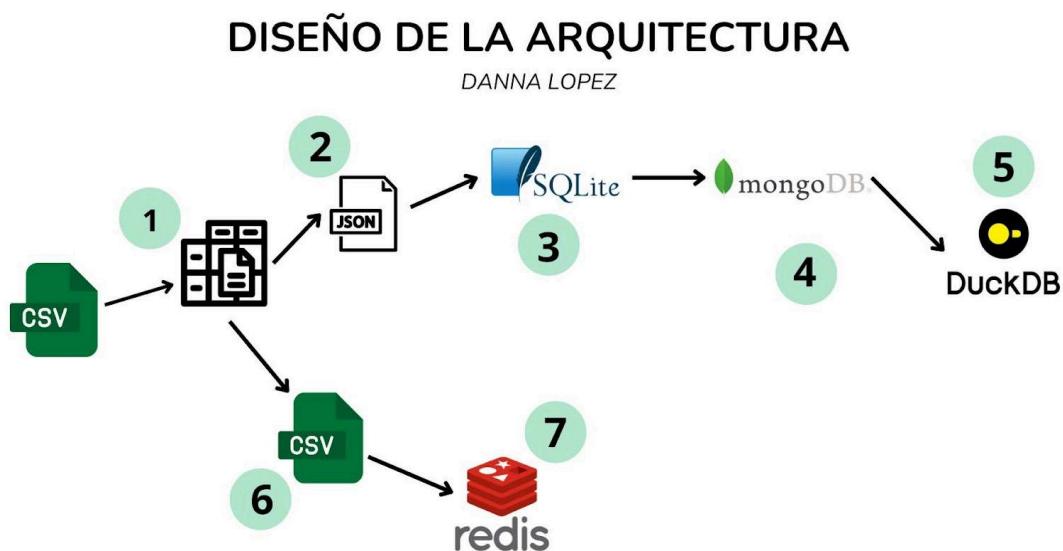
Fuente:

<https://www.ionos.com/es-us/digitalguide/paginas-web/desarrollo-web/mongodb-python/>

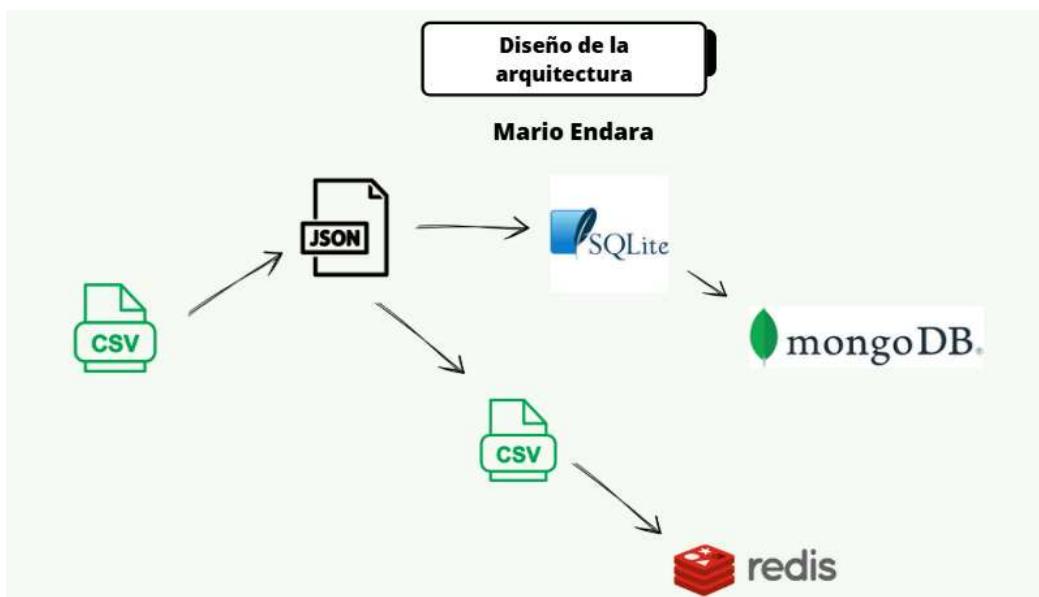
**5.2.3 Matplotlib:** Esta librería de Python será utilizada para la creación de gráficos.

## 5. Arquitectura de la solución

La arquitectura para la conversión de datos del dataset de accidentes de tráfico en diferentes períodos y regiones se la visualiza en el siguiente gráfico:



### Dataset de Olimpiadas



## Dataset de los Conciertos



## 6. Extracción de datos.

### Dataset Accidentes de tráfico en diferentes períodos:

La conversión entre los datasets, partiendo de un dataset de accidentes de tránsito, se visualiza de la siguiente manera:

accidents.duckdb	5,8 MB	Documento
accidents.db	54,9 MB	Documento
traffic_accidents_nuevo.csv	58,1 MB	Com...et (.csv)
traffic_accidents1.json	186,6 MB	JSON
traffic_accidents.csv	51,1 MB	Com...et (.csv)
traffic_accidents.json	185,1 MB	JSON
traffic_accidents.ipynb	532 KB	Documento

De la misma manera se visualiza los datos desde la interfaz gráfica en MongoDB Compass:

The screenshot shows the MongoDB Compass interface. At the top, it displays the connection information: localhost:27017. Below this is a sidebar with a tree view of databases: BDD1, accident\_db (which is expanded), admin, config, and local. The 'accidents' collection is selected and highlighted with a blue border. In the main pane, a single document from the 'accidents' collection is displayed in JSON format. The document contains numerous fields, including \_id, crash\_date, traffic\_control\_device, weather\_condition, lighting\_condition, first\_crash\_type, trafficway\_type, alignment, roadway\_surface\_cond, road\_defect, crash\_type, intersection\_related\_i, damage, prim\_contributory\_cause, num\_units, most\_severe\_injury, injuries\_total, injuries\_fatal, injuries\_incapacitating, injuries\_non\_incapacitating, injuries\_reported\_not\_evident, injuries\_no\_indication, crash\_hour, crash\_day\_of\_week, and crash\_month. A button labeled 'Show 6 more fields' is visible at the bottom of the document list.

```
_id: ObjectId('67a17a676f5149b123910242')
crash_date: "29/07/23 13:00:00"
traffic_control_device: "TRAFFIC SIGNAL"
weather_condition: "CLEAR"
lighting_condition: "DAYLIGHT"
first_crash_type: "TURNING"
trafficway_type: "NOT DIVIDED"
alignment: "STRAIGHT AND LEVEL"
roadway_surface_cond: "UNKNOWN"
road_defect: "UNKNOWN"
crash_type: "NO INJURY / DRIVE AWAY"
intersection_related_i: "Y"
damage: "$501 - $1,500"
prim_contributory_cause: "UNABLE TO DETERMINE"
num_units: 2
most_severe_injury: "NO INDICATION OF INJURY"
injuries_total: 0
injuries_fatal: 0
injuries_incapacitating: 0
injuries_non_incapacitating: 0
injuries_reported_not_evident: 0
injuries_no_indication: 3
crash_hour: 13
crash_day_of_week: 7
crash_month: 7
```

▼ Show 6 more fields

## Dataset Olimpiadas:

La conversión entre los datasets, partiendo del dataset de olimpiadas, se visualiza de la siguiente manera:

.ipynb_checkpoints	1/2/2025 20:35
archivo_convertido a	1/2/2025 20:40
archivo3	Tipo: Archivo de valores separados Tamaño: 25.9 MB 31/1/2025 9:47 Miér
convertido	1/2/2025 20:35 Fecha de modificación: 1/2/2025 20:35
olimpiadas	3/2/2025 21:06
olympics_dataset	29/8/2024 23:52

Para hacer las conexiones de SQLite a Mongo, se necesitó importar las siguientes librerías:

### SQLite

```
#El archivo3 abierto en SQLite
import sqlite3
#concatenar con la base de datos
conn = sqlite3.connect("archivo3.db")
cursor = conn.cursor()
df.head()
```

### Mongo

```
#El dataframe generado en el paso 3 se lo debe exportar a mongoDB
import pandas as pd
from pymongo import MongoClient
#Conexion a Mongo
client = MongoClient("mongodb://localhost:27017/")
db=client["olimpiadas"]
collection=db["archivo3"]
```

## Dataset de Conciertos:

La conversión entre los datasets, partiendo del dataset de conciertos, se visualiza de la siguiente manera:

📁 .ipynb_checkpoints	4/2/2025 8:06	Carpeta de archivos
📄 concerts.csv	3/2/2025 12:24	Archivo de valores... 3.193 KB
📄 concerts.sqlite	3/2/2025 18:41	Archivo SQLITE 3.520 KB
📄 perf_history_clean.csv	3/2/2025 17:52	Archivo de valores... 3.422 KB
📄 Proyecto.ipynb	3/2/2025 17:52	Archivo de origen ... 307 KB
📄 Proyecto1.ipynb	4/2/2025 8:06	Archivo de origen ... 823 KB

Para hacer las conexiones de MYSQL(WORKBENCH) a SQLITE, se necesitó importar las siguientes librerías:

```
import mysql.connector
from sqlalchemy import create_engine

# Configura la conexión a MySQL
engine_mysql = create_engine('mysql+mysqldb://root:123456@localhost/proyecto')

# Guardar el DataFrame en MySQL en una tabla llamada 'concerts'
concerts.to_sql('concerts', engine_mysql, if_exists='replace', index=False)
```

```
import sqlite3

# Conexión a una base de datos SQLite (se creará un archivo .db)
conn_sqlite = sqlite3.connect('concerts.sqlite')
concerts.to_sql('concerts', conn_sqlite, if_exists='replace', index=False)
conn_sqlite.close()
```

Y luego a MongoDB:

```
from pymongo import MongoClient

client = MongoClient('mongodb://localhost:27017/')
db = client['orquesta_db']
collection = db['concerts']
collection.insert_many(concerts_json)
```

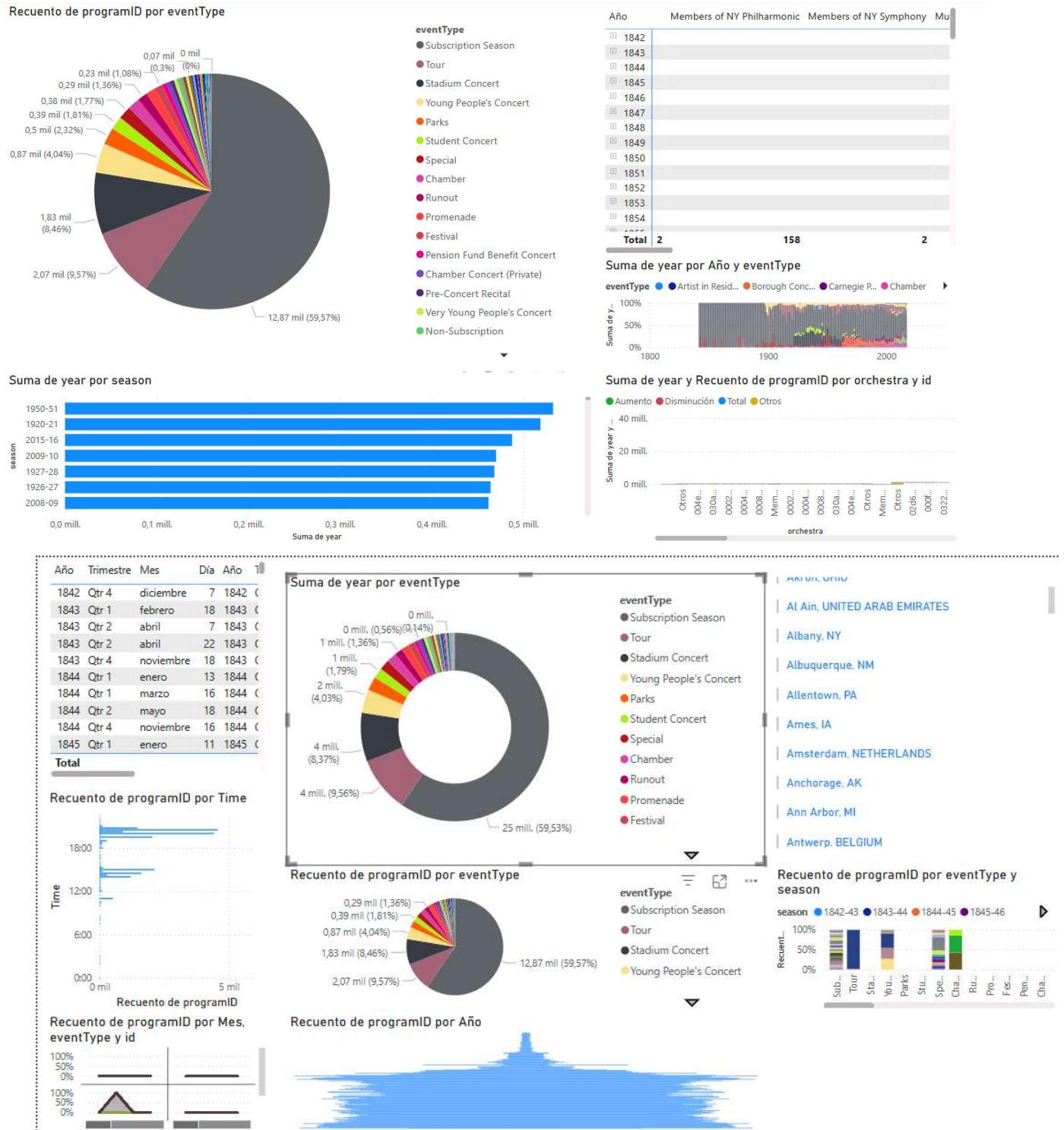
por último Redis:

```
#Instala y conecta a Redis:
#pip install redis en cmd

import redis

r = redis.Redis(host='localhost', port=6379, db=0)
# Ejemplo: Guardar el total de conciertos
total_concerts = concerts.shape[0]
r.set('total_concerts', total_concerts)
```

Análisis y Visualización de información.



**Gráfica 1:**

## Distribución de tipos de eventos

- El gráfico circular muestra que el tipo de evento predominante es Subscription Season, con un 59.57% del total de eventos registrados.

- Otros eventos con una presencia significativa son Tour (9.57%) y Stadium Concert (8.46%).
- Los demás tipos de eventos tienen una participación mucho menor, lo que indica que la mayoría de los conciertos siguen un formato de suscripción en lugar de eventos aislados o especiales.

## Tendencia histórica por temporada

- El gráfico de barras "Suma de year por season" indica que las temporadas con mayor actividad fueron las de los años 1950-51 y 1920-21.
- También se observan picos en temporadas más recientes como 2015-16 y 2009-10, lo que sugiere un repunte en la cantidad de eventos en estos períodos.
- Esto podría reflejar una mayor organización de conciertos en ciertos años debido a factores como la popularidad de la orquesta, demanda del público o apoyo financiero.

## Grafica 2:

### Distribución de los Tipos de Eventos

- El gráfico de pastel "Suma de year por eventType" muestra que el tipo de evento más común es "Subscription Season", representando más del 59% de los eventos registrados.
- Otros eventos con una presencia significativa incluyen "Tour" (9.56%) y "Stadium Concert" (8.37%), mientras que los demás tipos de eventos tienen una participación menor.
- Esto sugiere que la mayoría de los conciertos registrados en el dataset están asociados con suscripciones anuales, lo que indica una fuerte preferencia por eventos programados a largo plazo en comparación con conciertos individuales o especiales.

## Evolución Temporal de los Conciertos

- En el gráfico de barras "Recuento de programID por Año", se observa una distribución en la que hay un incremento en la cantidad de eventos a lo largo del tiempo.
- Este aumento puede estar relacionado con la creciente popularidad de los conciertos y la expansión de las orquestas en diferentes períodos.
- Además, en "Recuento de programID por Time", se nota que la mayoría de los conciertos se llevan a cabo en horarios de tarde o noche, lo que coincide con las horas en las que el público generalmente asiste a eventos culturales.

## Data sets Registro de vehículos

Es necesario primero establecer conexión con sqlite y mongoDB para poder hacer la migración de datos.

Conexión establecida en mongoDB:

```
import pandas as pd
from pymongo import MongoClient

mi_registro = "vehiculos6"

client = MongoClient("mongodb://localhost:27017/")
db=client["midb"]
collection=db["mi_registro"]
```

Conexión establecida en sqlite:

```
import sqlite3

registro_sql = "registro_sql.db"
conexion = sqlite3.connect(registro_sql)
cursor = conexion.cursor()
```

La conversión entre los datasets, partiendo del dataset de olimpiadas, se visualiza de la siguiente manera:

Archivo original:

 2014_VehículosMatriculadosBDD.csv	29/1/20
---	---------

Archivos derivados:

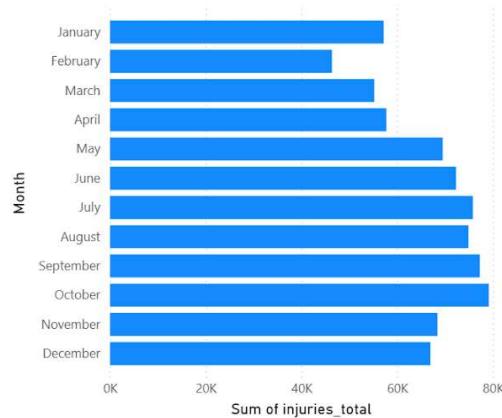
 Proyecto_Final.ipynb	3/2/2025 17:56	Archivo de origen Jupyter
 vehiculos_csv_new.csv	3/2/2025 17:55	Archivo de valores separados por ...
 vehiculos_json_new.json	3/2/2025 17:55	Archivo de origen JSON
 registro_sql.db	3/2/2025 17:53	Data Base File
 vehiculos_registro.json	3/2/2025 17:51	Archivo de origen JSON

## 7. Análisis y Visualización de información.

### 8.2 Dashboards en Accidentes de Tránsito

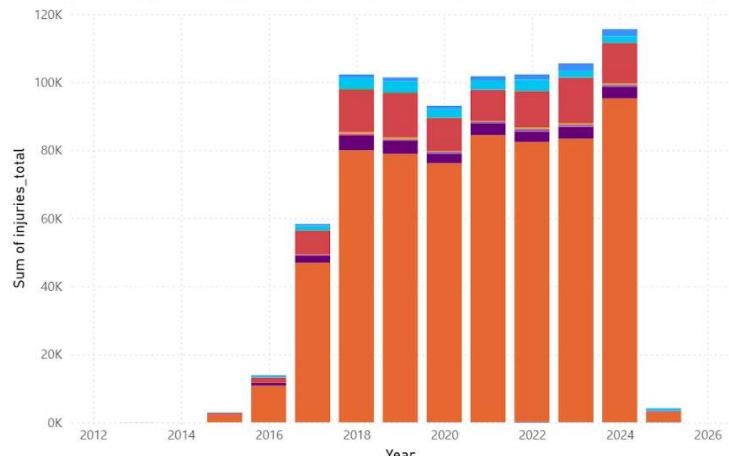
Year	Sum of injuries_total	weather_condition
2013	30	CLEAR
2015	2200	CLEAR
2015	70	CLOUDY/OVERCAST
2015	0	FOG/SMOKE/HAZE
2015	10	OTHER
2015	430	RAIN
2015	60	SLEET/HAIL
2015	20	SNOW
<b>Total</b>	<b>801050</b>	

Sum of injuries\_total by Month



#### NUMERO TOTAL DE LESIONES BY YEAR Y CONDICION CLIMATICA

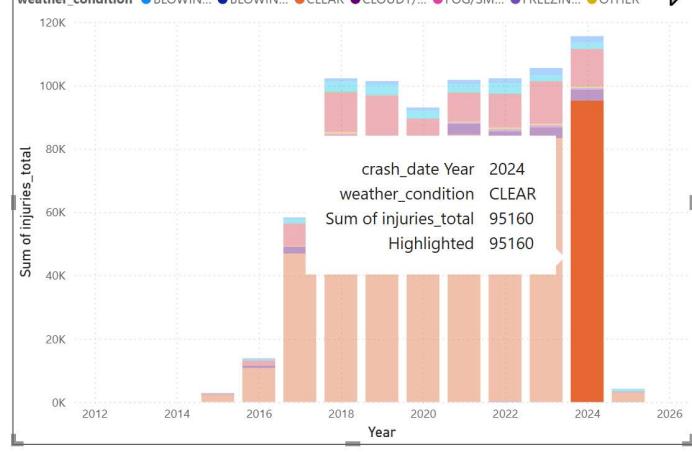
weather\_condition ● BLOWIN... ● BLOWIN... ● CLEAR ● CLOUDY/... ● FOG/SM... ● FREEZIN... ● OTHER



La categoría "CLEAR" (claro) muestra los valores más altos en casi todos los años, lo que sugiere que la mayoría de los accidentes ocurren en condiciones climáticas claras.

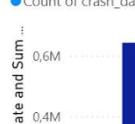
#### Sum of injuries\_total by Year and weather\_condition

weather\_condition ● BLOWIN... ● BLOWIN... ● CLEAR ● CLOUDY/... ● FOG/SM... ● FREEZIN... ● OTHER



#### Count of crash\_date and Sum of injuries\_total by weather\_condition

● Count of crash\_date ● Sum of injuries\_total



#### Distribucion del numero total de accidentes por condicion climatica

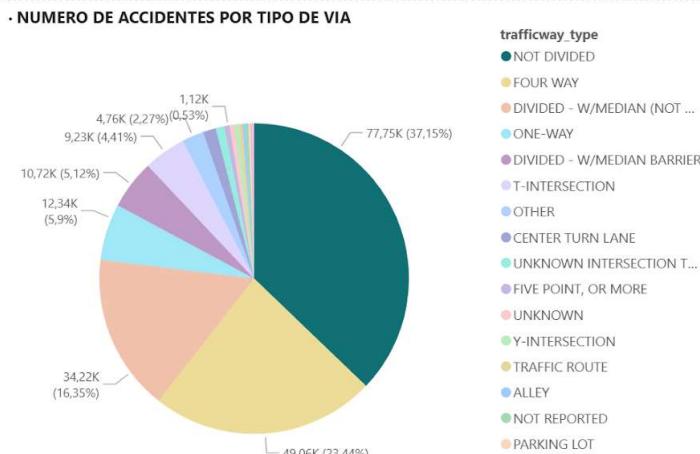
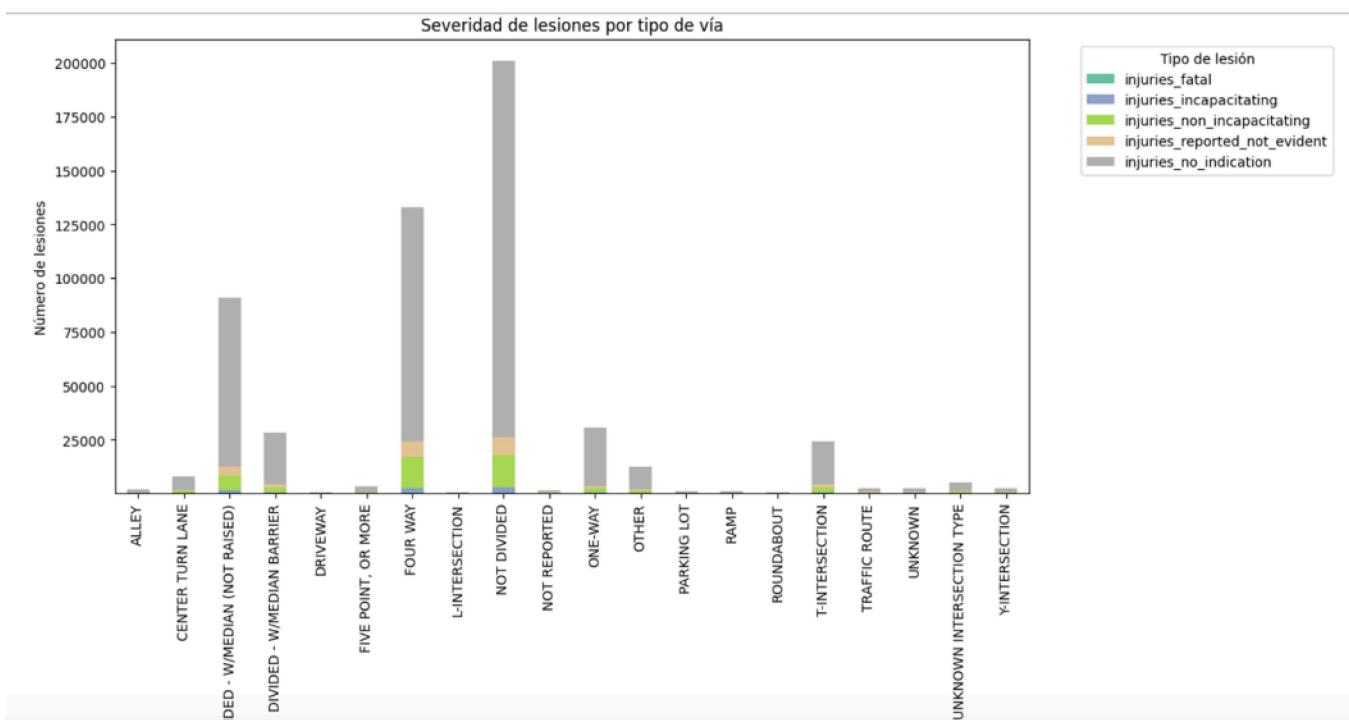
weather\_con...

● CLEAR

● RAIN

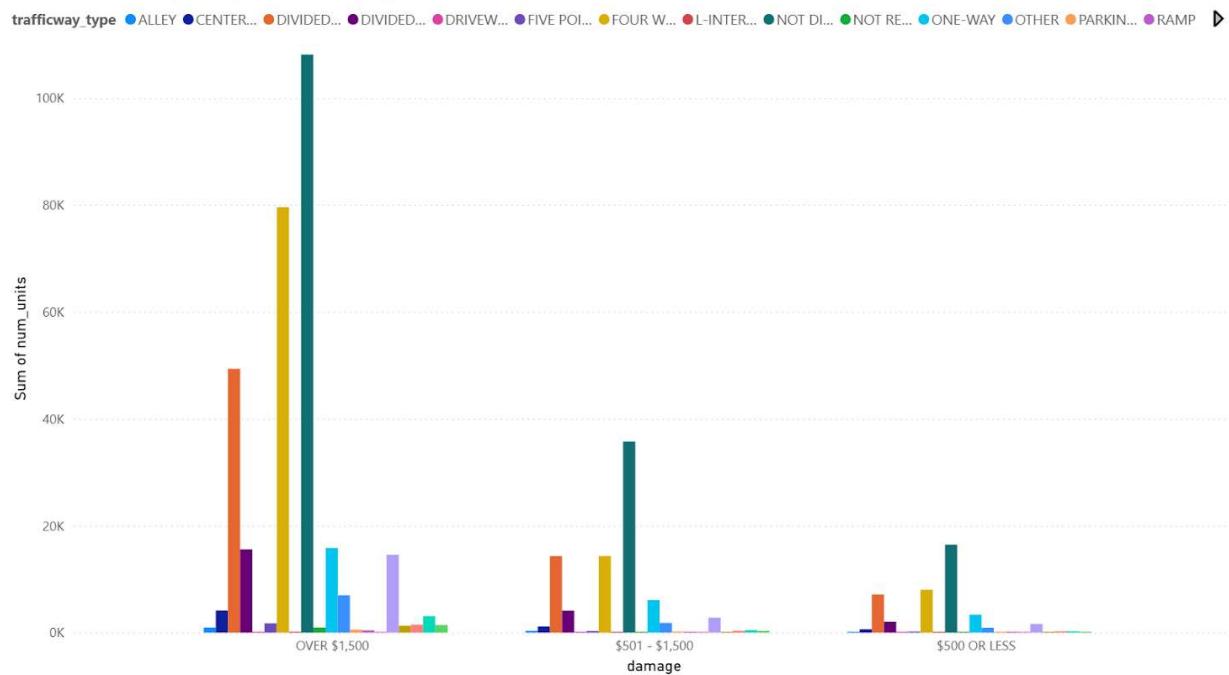
● CLOUDY/O...

6.53K (3.12%)

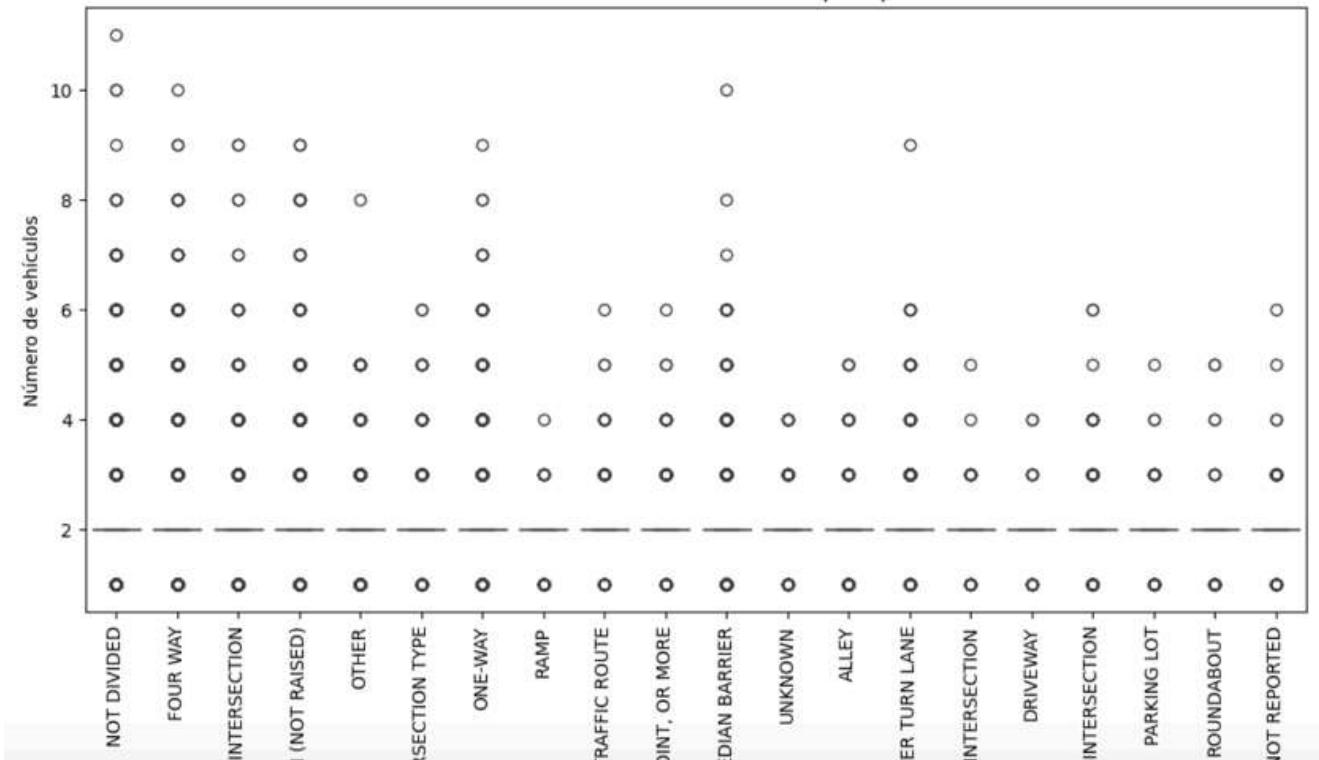


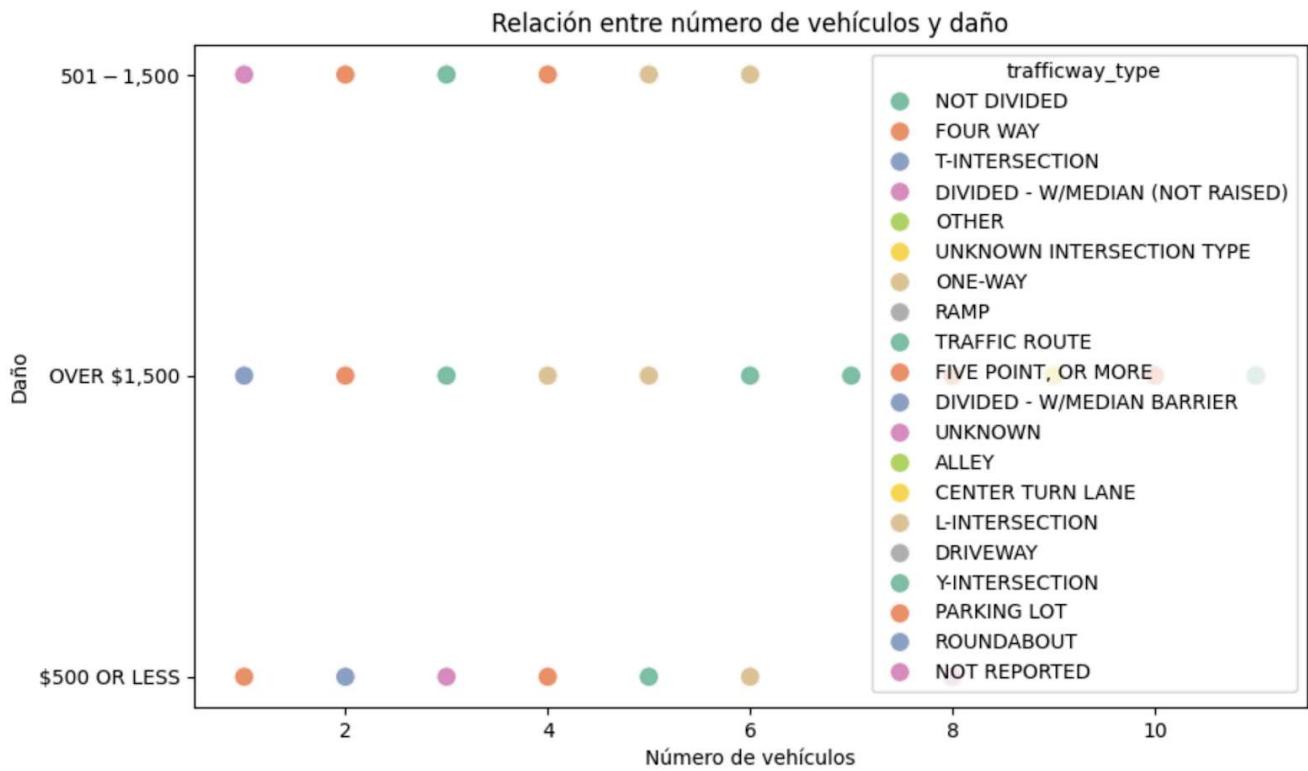


Sum of num\_units by damage and trafficway\_type



Distribución del número de vehículos por tipo de vía

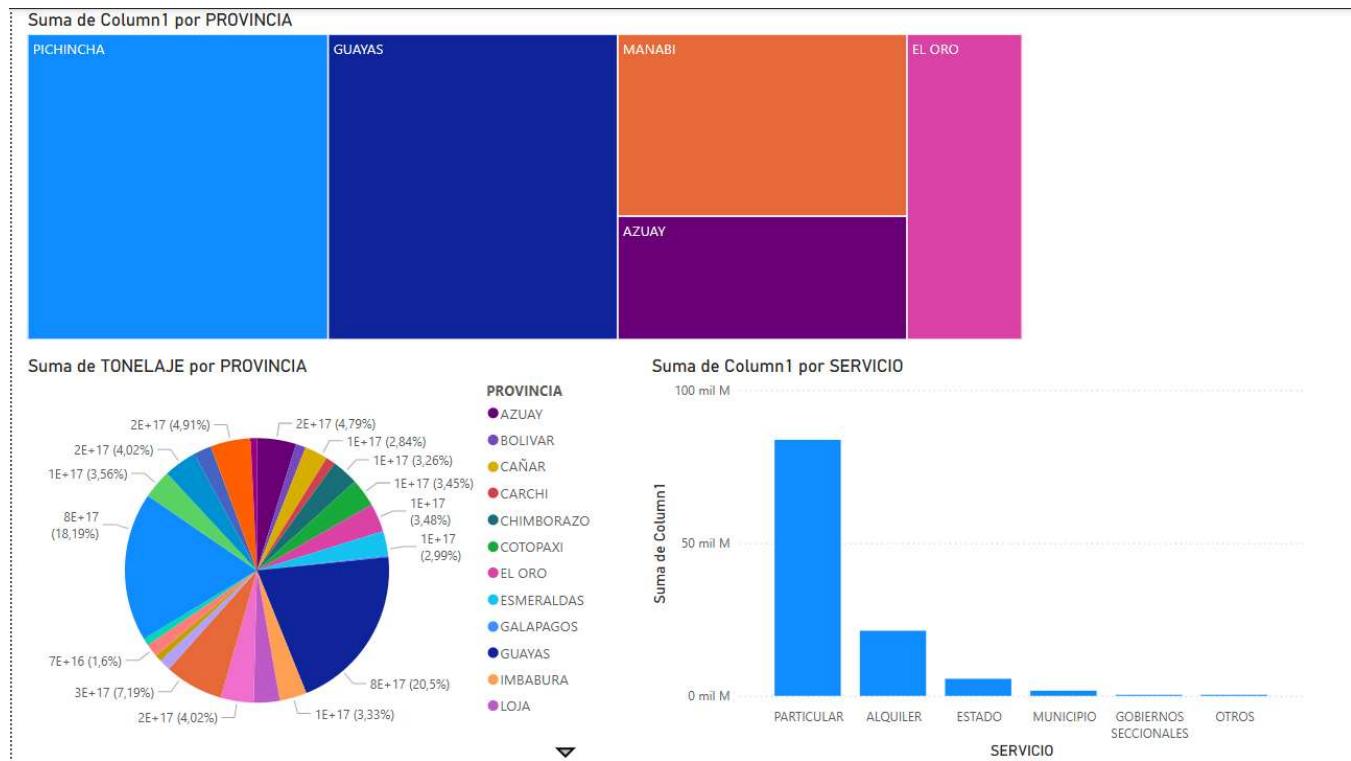
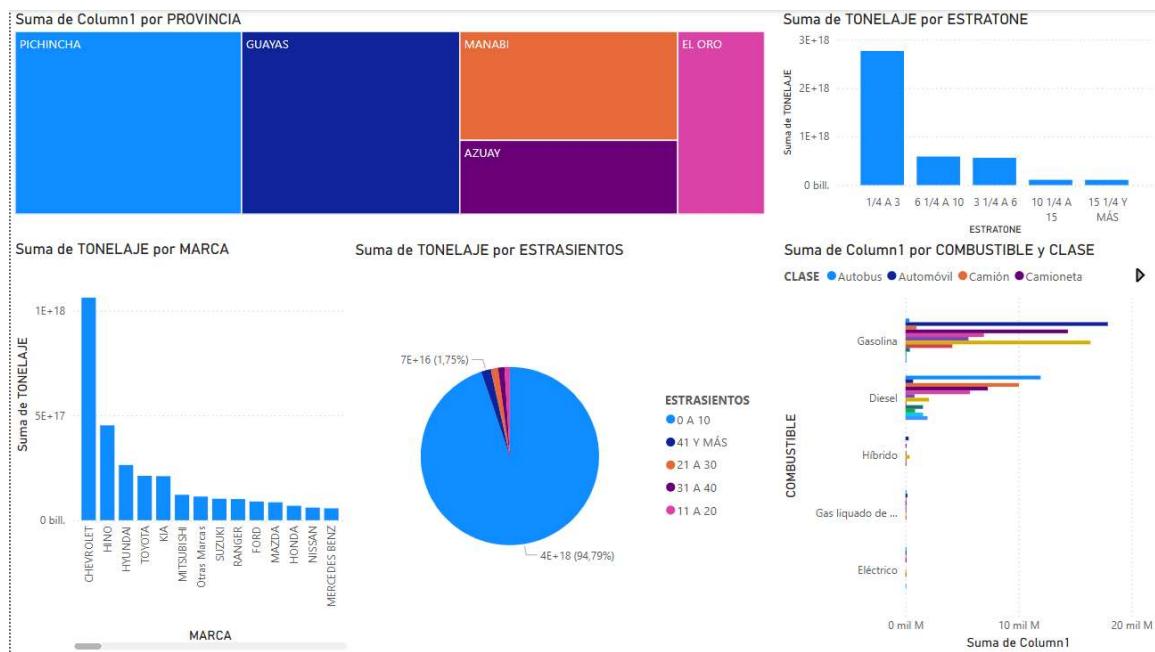


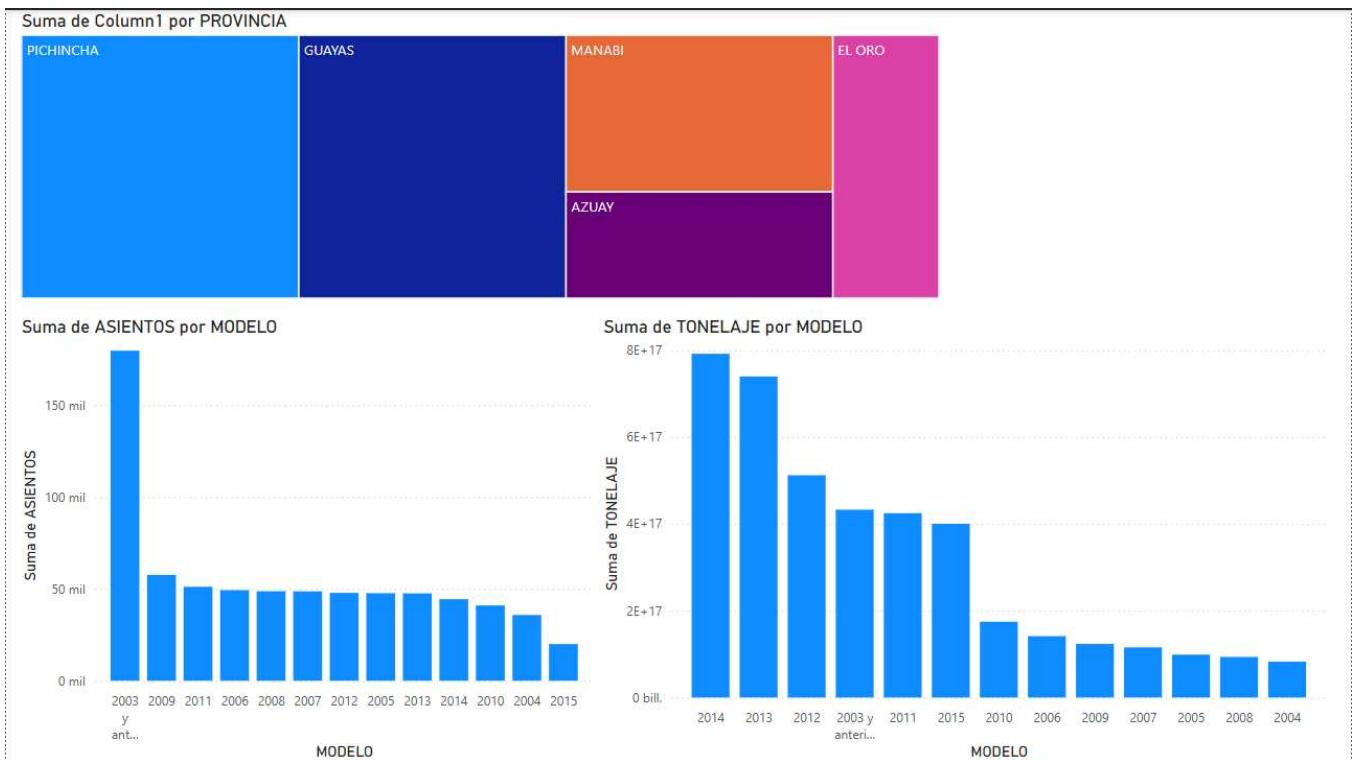


## 8.1 Dashboard en Power BI-Olimpiadas



## 8.2 Dashboards en Registro de Vehículos





## 8. Resultados obtenidos.

### Dataset Accidentes de Tránsito

#### GRÁFICA 1:

- De los resultados del primer gráfico, la categoría "CLEAR" (despejado) muestra los valores más altos en casi todos los años, lo cual sugiere que, si bien las condiciones climáticas adversas pueden aumentar el riesgo de accidentes, la mayoría de los accidentes ocurren en días despejados.
- Si bien el clima claro es el predominante, las condiciones climáticas adversas como "FOG/SM" (niebla/nieve) y "FREEZIN" (heladas) también contribuyen a un número significativo de accidentes, especialmente en ciertos años. Esto sugiere que las condiciones climáticas extremas pueden aumentar significativamente el riesgo de accidentes.

#### GRÁFICA 2:

- El hecho de que en el año 2024 se observa un pico significativo en el número total de lesiones, especialmente en condiciones climáticas claras, puede indicar varios factores como:

Cambios en la infraestructura vial en ese año (nuevas carreteras, modificaciones en intersecciones) que aumentaran el riesgo de accidentes.

Un incremento en el número de vehículos en circulación podría explicar el mayor número de accidentes.

Algún evento especial o festividad durante ese año podría haber generado un aumento en la movilidad y, por ende, en la cantidad de accidentes.

- También se pudo interpretar de la gráfica que los accidentes se dan con más frecuencia en

el mes de Octubre.

### GRÁFICA 3:

- Este gráfico de barras presenta una visión general de la frecuencia de accidentes a lo largo de diferentes horas del día y días de la semana. Cada barra representa una combinación específica de hora y día, y su altura indica el número total de lesiones reportadas en ese momento.
- Se observa una mayor concentración de accidentes en las horas pico de tráfico, tanto en la mañana como en la tarde. La mayoría de accidentes ocurren en horas de la tarde (de 2 a 5pm), en un clima despejado, y el día de la semana con mayor cantidad de accidentes es el sábado. Esto sugiere que la congestión vehicular y la mayor cantidad de vehículos en la carretera podrían estar contribuyendo a una mayor frecuencia de accidentes.
- Los accidentes más frecuentes ocurren también a medianoche. Los más frecuentes ocurren en un clima despejado, lo que da a muchas interpretaciones:

Primero, es porque los conductores carecen de visibilidad a esa hora. A los ojos les toma tiempo acostumbrarse a ver con poca iluminación, provocando que los peligros del camino no se alcancen a distinguir a simple vista

Otra posibilidad es que los conductores manejaban en un estado de somnolencia. Durante la noche aumentan los niveles de sueño y cansancio, sobre todo después de la media noche, y más aún si se ha conducido por un largo periodo de tiempo, sin descanso.

Por último, muchos de los accidentes a media noche ocurren debido al exceso de velocidad de los pasajeros, ya que a esa hora reducen completamente las horas pico.

- También vemos que si bien es cierto que los accidentes por la carretera mojada o llena de nieve son muy poco frecuentes, también hay una cantidad representativa de accidentes puesto que, los accidentes pueden ocurrir ya que la lluvia o la nieve cubren las calzadas de agua y nieve que las hace peligrosamente deslizantes.

### GRÁFICA 4

- La mayoría de las lesiones ocurren en vías clasificadas como "NOT DIVIDED" y "FOUR WAY". Estas vías tienen las barras más altas, lo que indica que la mayoría de los accidentes ocurren en ellas.
- También se observa una alta proporción de lesiones reportadas pero no evidentes y sin indicación de lesión.
- Las vías "DIVIDED - W/MEDIAN (NOT RAISED)" y "ONE-WAY" también muestran un número significativo de accidentes.
- Esto podría indicar que estas vías presentan condiciones que favorecen accidentes con lesiones.
- Las lesiones fatales (azul oscuro) y las lesiones incapacitantes (verde oscuro) son relativamente bajas en comparación con las lesiones menores. Esto sugiere que la mayoría de los accidentes no son mortales, pero aún pueden causar daños considerables.
- Algunas vías como "ROUNABOUT", "T-INTERSECTION", "RAMP" y "Y-INTERSECTION" tienen muy pocas lesiones reportadas. Esto podría indicar que estos tipos de vías son menos propensas a accidentes con lesiones graves.

### Gráfica del número de vehículos vs. el daño causado

- Mientras más vehículos colisionan, los daños son mayores (por encima de los 1500), sin embargo, los accidentes que involucran dos vehículos también han causado daños graves.

### Gráfica de tipos de vía según los vehículos involucrados

Cada punto representa un accidente, donde:

- o El eje X muestra los diferentes tipos de vía.
- o El eje Y indica el número de vehículos involucrados en el accidente.

La mayoría de los accidentes involucran 1 o 2 vehículos, lo cual es común en colisiones estándar (como choques traseros o laterales).

Algunas vías presentan accidentes con más de 10 vehículos, lo que sugiere posibles choques en cadena o accidentes en carreteras de alta velocidad.

La línea discontinua cerca del valor 2 indica la mediana o un umbral de referencia.

En vías como NOT DIVIDED, FOUR WAY e INTERSECTIONS, hay una mayor dispersión en el número de vehículos, indicando que en estas zonas pueden ocurrir accidentes más grandes.

En driveways (entradas), alleys (callejones) y parking lots (estacionamientos), los accidentes generalmente involucran pocos vehículos.

#### **De la ultima grafica:**

La mayoría de los accidentes con pocos vehículos (1-2 vehículos) pueden estar asociados con cualquier nivel de daño (\$500 o menos, 501-1500, más de 1500 dólares).

A medida que el número de vehículos aumenta (más de 6 o 8 vehículos), la mayoría de los puntos se encuentran en niveles de daño mayores (\$1,500+), lo cual tiene sentido porque colisiones con más autos suelen generar más daños.

#### **Dashboard Juegos Olímpicos**

##### **GRÁFICA 1:**

- Los resultados obtenidos en el primer gráfico, muestra que existen mayor cantidad de participantes del genero masculino que femenino esto podría deberse a razones personales de la sociedad o falta de apoyo del estado para impulsar mayor cantidad de deportistas.

##### **GRÁFICA 2:**

- En el gráfico muestra que USA es el país con mayor cantidad de medallas olímpicas y en que menos tiene es Holanda.

##### **GRÁFICA 3:**

- Muestra que la mayor cantidad de participantes no obtienen medallas

#### **Análisis de Dashboards de Registro de vehículos en Ecuador.**

##### **Página 1:**

- Ya que se deseaba obtener un análisis solo en las provincias más importantes en el

Ecuador, se procedió a hacer un filtro de los datos y luego se realizó la inserción de un treemap para solo obtener los datos de dicha provincia de forma visual.

- Primero se tomó en cuenta en la gráfica siguiente el tonelaje por estratone, el **Estratone** es una medida para la capacidad de carga en grupos de los vehículos. la explicación de este término fue muy baga a la hora de consultarla.
- Luego, se realizó un gráfico de la capacidad de carga en vehículos, catalogados por marca, este análisis nos sirve para ver cual marca de vehículo es mas adecuado para cargas pesadas.
- La gráfica de tonelaje por estrásientos nos permite calcular la cantidad de asientos en el vehículo y su capacidad de carga.
- La capacidad de combustible o su cilindraje por clase de vehículo, los principales eran los de gasolina y los de diesel, pero también hubo los híbridos los cuales estaban por debajo de los dos anteriormente mencionados, por gran diferencia

#### Página 2:

- Incluyendo al anterior gráfico de visualización de provincias, en esta página solo se encuentran tres gráficas.
- La segunda gráfica nos permite visualizar el tonelaje por provincia, un agregado para ver si el filtro de las provincias es aplicado correctamente.
- La tercera grafica es una que nos permite observar el número de vehículos por servicio ofrecido. Permite visualizar la cantidad de vehículos y el trabajo en general para el cual son utilizados.

## 9. Conclusiones y Recomendaciones

### 10.1 Accidentes de tránsito

1. La mayor cantidad de accidentes ocurren en horas de la tarde (de 14:00 a 17:00), en un clima despejado, y el día de la semana con mayor cantidad de accidentes es el sábado. También ocurren a medianoche, quizá porque los conductores no tienen muy buena visibilidad a esa hora.
2. La mayoría de accidentes involucran dos vehículos involucrados con el número de lesiones más altas.
3. Los accidentes que han causado más daños involucran 6 a 8 vehículos.

### 10.2 Juegos Olímpicos

1. Se concluye que en todos los juegos realizados hay mayor cantidad de participantes hombre que mujeres.
2. Se concluye que USA es el país con mayor cantidad de medallas seguido de Francia.
3. La mayor cantidad de participantes que concursan no obtienen medallas
4. Se concluye que las medallas como los participantes han ido en aumento a través del tiempo

5. Se concluye que los juegos son celebrados en verano y en general en la capital del país anfitrión
6. En damas se concluye que USA también lidera la cantidad de medallas.

### 10.3 Registro de vehículos

- Se pueden observar una cantidad de vehículos con una capacidad muy grande de tonelaje por lo que se puede deducir que estos vehículos podrían estar siendo usados para transporte de materiales o puede ser transporte público.
- Se puede observar una gran cantidad de vehículos chevrolet en Ecuador, y se puede vislumbrar que es debido tanto a su capacidad de carga como a su capacidad de cilindraje.
- Otro dato a observar es que, el Ecuador es un país con un número casi inexistente de vehículos híbridos o eléctricos.
- Otro dato a tener en cuenta es que la capacidad de los vehículos ha ido aumentando cada año. Seguramente esto es debido a que muchos de los vehículos usados se utilizan con la intención de transportar algo, ya sea materiales, materia prima o incluso personas en el caso de los vehículos de transporte público.

## 10. Recomendaciones

- Se recomienda instalar todas las dependencias y librerías al utilizar Jupyter notebook para evitar errores.
- En sistemas operativos como Mac OS, es mejor hacer la instalación de herramientas (Redis o MongoDB) desde Brew.
- Se recomienda realizar una limpieza de datos antes de realizar el análisis para verificar que no hayan datos nulos, ni duplicados.
- Para la visualización de datos, se recomienda utilizar gráficos de barras o histogramas para apreciar de mejor forma los datos.
- Los formatos de fecha es preferible manipularlos en dia, mes y año para que se visualicen de manera correcta en Power BI y no se cambien a otros valores incorrectos.

## 11. Desafíos y problemas encontrados.

- Un desafío encontrado fue la conexión a base de datos (Mongo, SQLite, MySQL Workbench, etc.) por lo que se tuvo que buscar soluciones como migrar a DuckDB.
- Al migrar a DuckDB, el programa se ralentizaba, puesto que almacenaba los miles de registros de manera diferente.

## 12. Problemas encontrados

Un problema encontrado es el uso adecuado de Power BI aunque es una herramienta de uso accesible es importante mencionar que se deben realizar bien las relaciones entre las tablas, para que las visualizaciones se presenten de forma oportuna y fácil de

entender.

### **13. Link de github del proyecto.**

<https://github.com/MarioEndara/Proyecto-Final-de-Analisis.git>

### **14. Videos explicativos**

#### **15.1 Extracción de datos**

Dataset de accidentes de tráfico:

<https://www.youtube.com/watch?v=mknqZuAYlIY>

Dataset de olimpiadas

<https://youtu.be/7VVLQDO2n8k>

#### **15.2 Segundo Video de YouTube sobre Conclusiones de los dashboards:**

**Dashboard de olimpiadas**

[https://www.youtube.com/watch?v=mL8H58krhVQ&ab\\_channel=MarioEndara](https://www.youtube.com/watch?v=mL8H58krhVQ&ab_channel=MarioEndara)

<https://youtu.be/R86vpW1seVY>

### **15. Presentación**

[https://www.canva.com/design/DAGeJCfxiWY/oBk934S5iVjrdT6q3mgXzA/edit?utm\\_content=DAGeJCfxiWY&utm\\_campaign=designshare&utm\\_medium=link2&utm\\_source=sharebutton](https://www.canva.com/design/DAGeJCfxiWY/oBk934S5iVjrdT6q3mgXzA/edit?utm_content=DAGeJCfxiWY&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton)