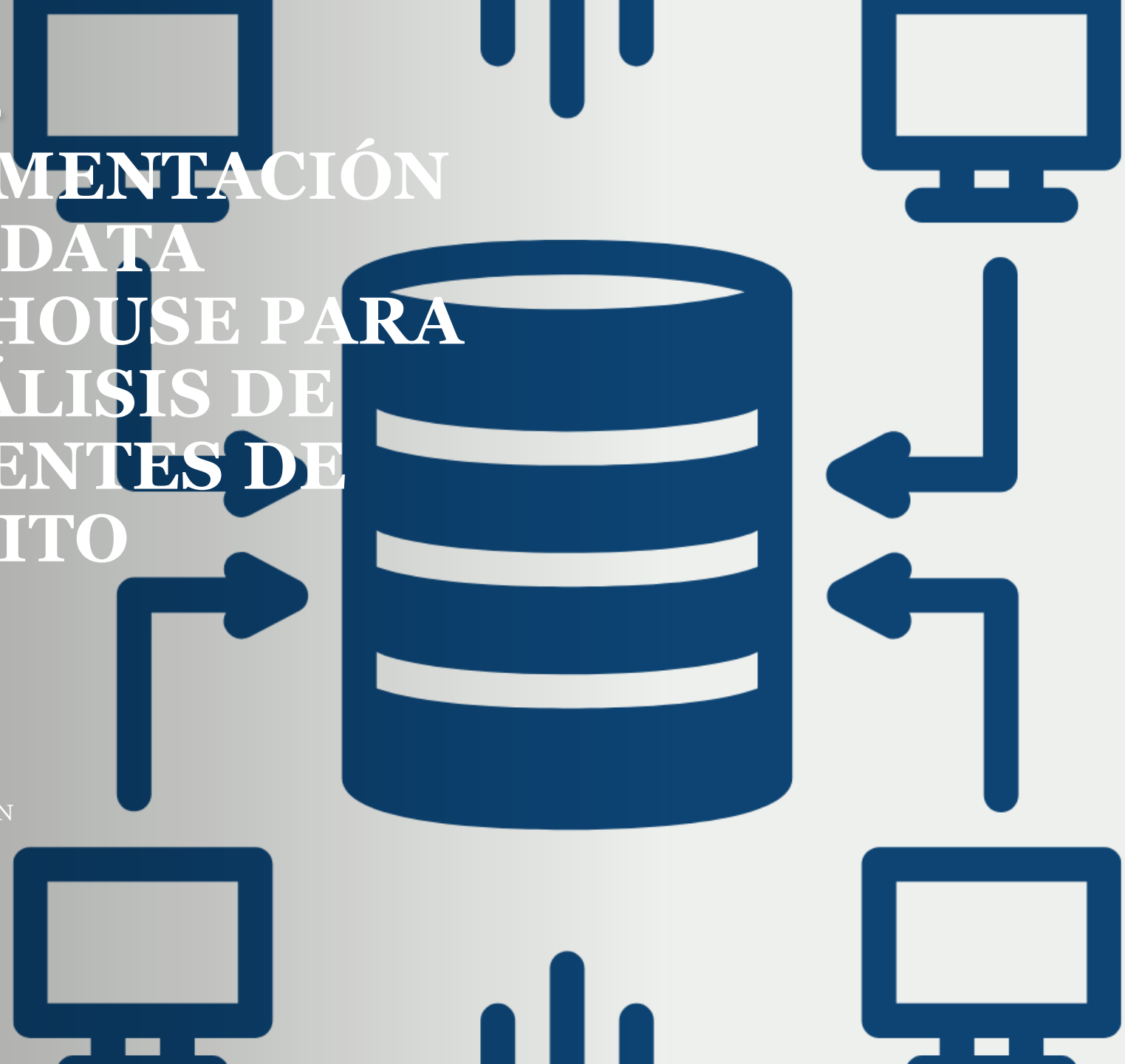
IMPLEMENTACIÓN DE UN DATA WAREHOUSE PARA EL ANÁLISIS DE ACCIDENTES DE TRÁNSITO



INTEGRANTES:

ANTHONY MONTERO ROMAN

DANIEL RAYO DÍAZ

KUN KIN ZHENG LIANG

AGENDA

DESCRIPCIÓN DATASET

MODELO DE DATA WAREHOUSE

PROCESO ETL

CONSULTAS Y RESULTADOS

CONCLUSIONES

DATASET: CRASH REPORTING - DRIVERS DATA

¿Por qué se escogió este dataset?

- Impacto social y utilidad
- Posibilidad de realizar análisis predictivo
- Diversidad de variables

DATASET: CRASH REPORTING - DRIVERS DATA

Implicaciones del dataset en el proyecto

- Impacto en la seguridad vial
- Optimización de recursos
- Desarrollo de políticas públicas basadas en datos
- Aplicación de modelos predictivos

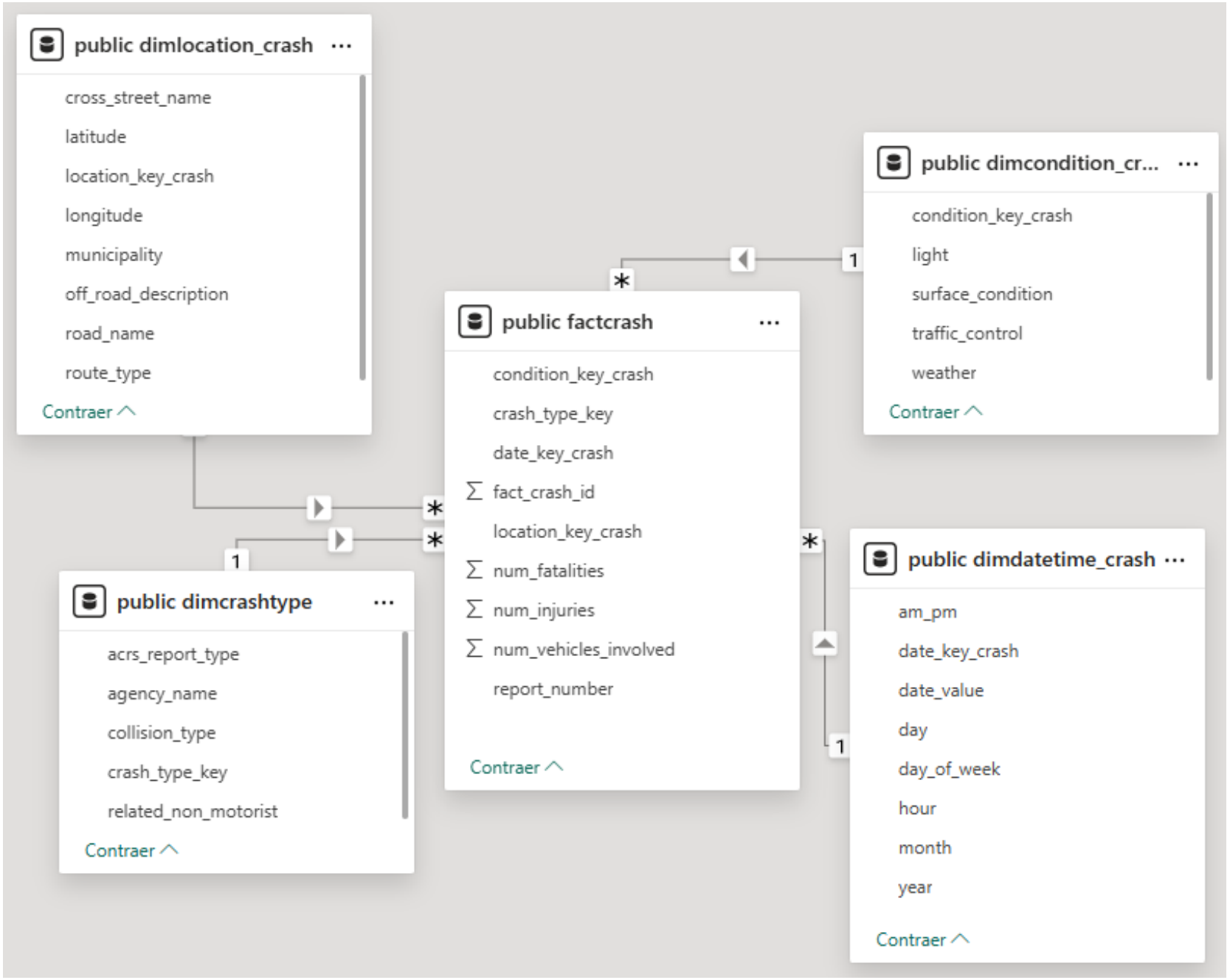
MODELO DE DATA WAREHOUSE

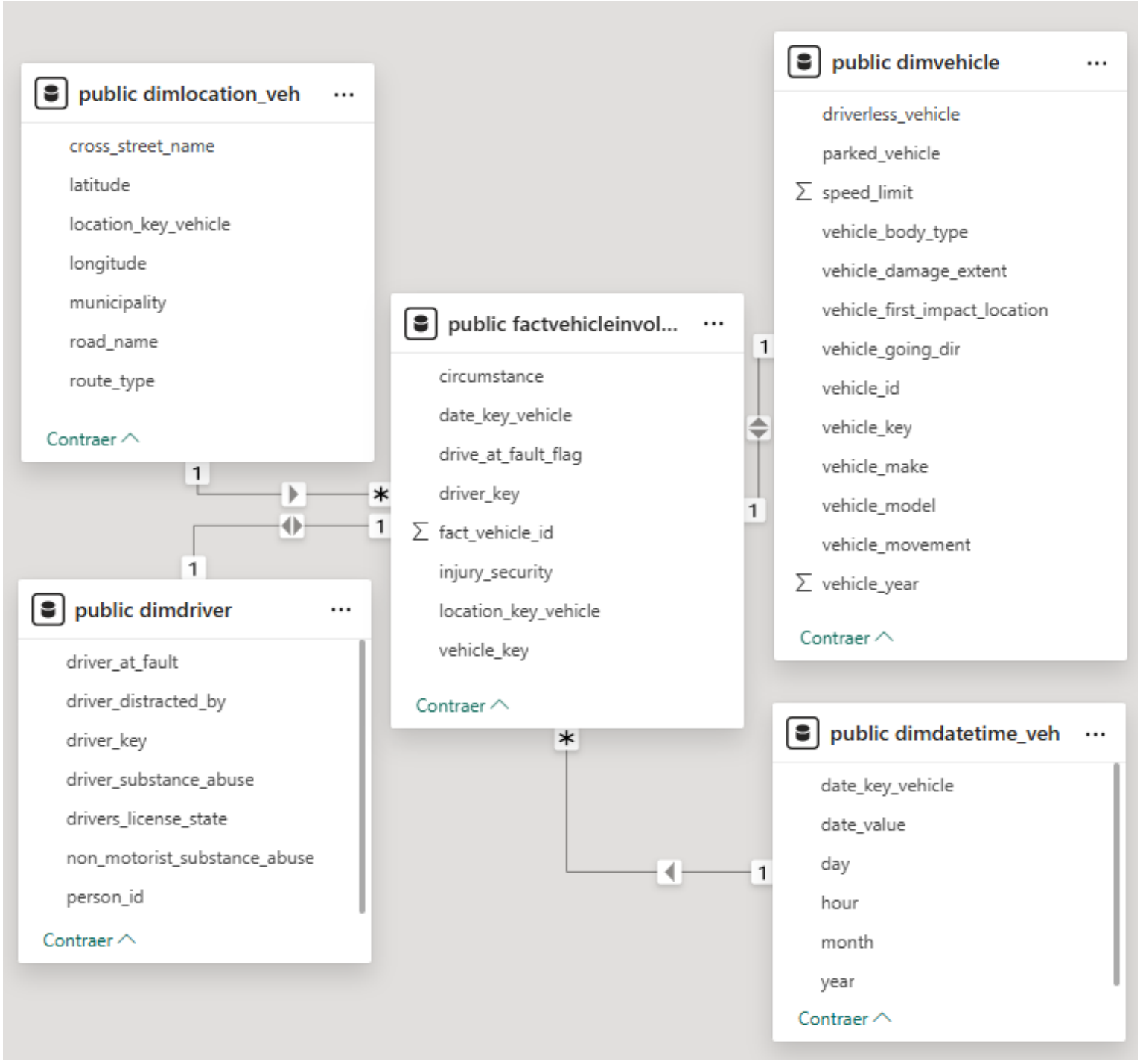
2 Esquemas estrella

A nivel de accidente en general

Nivel de participación de cada vehículo

ESQUEMA PARA ACCIDENTES





ESQUEMA PARA VEHÍCULOS

PROCESO ETL

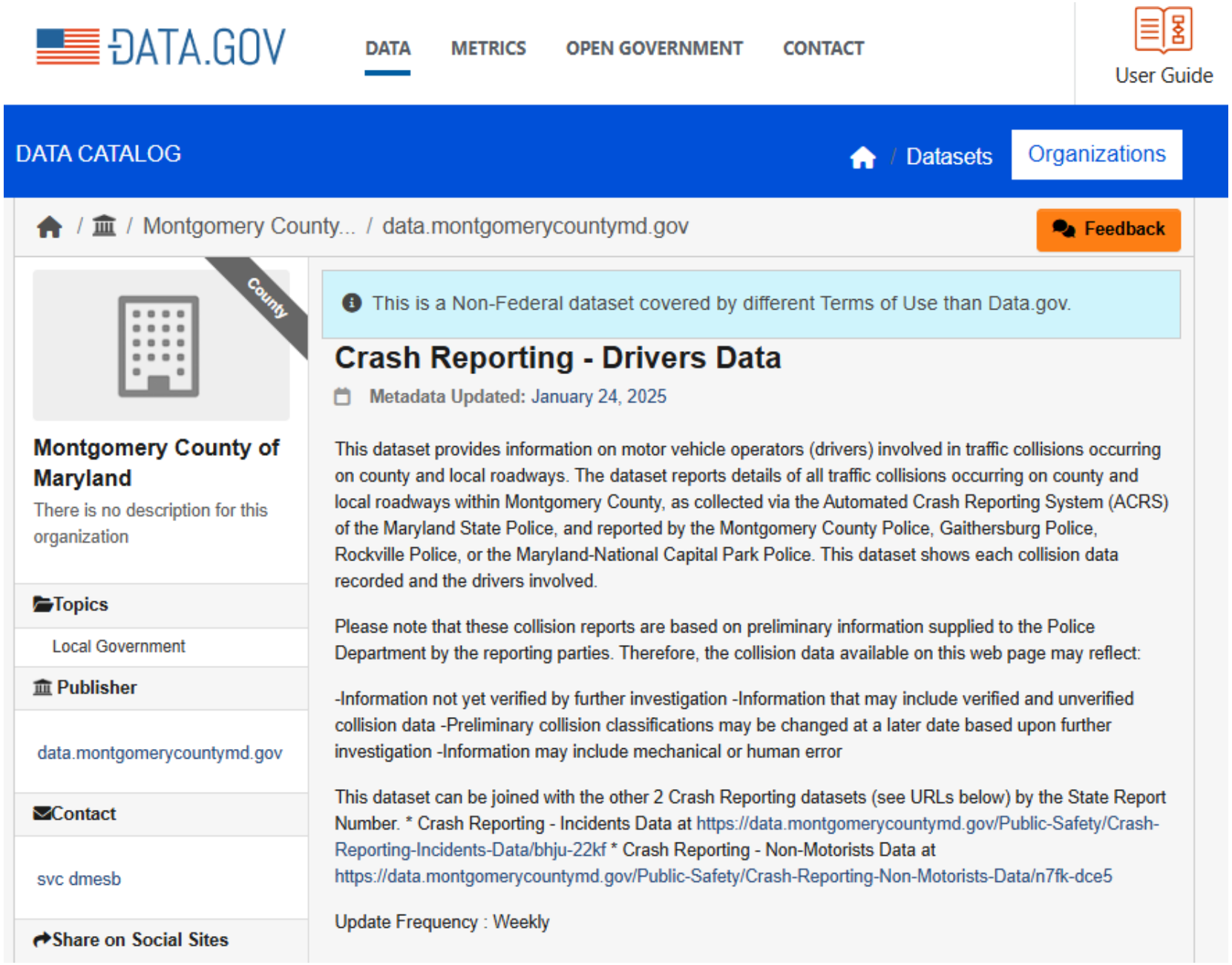
DATA.GOV

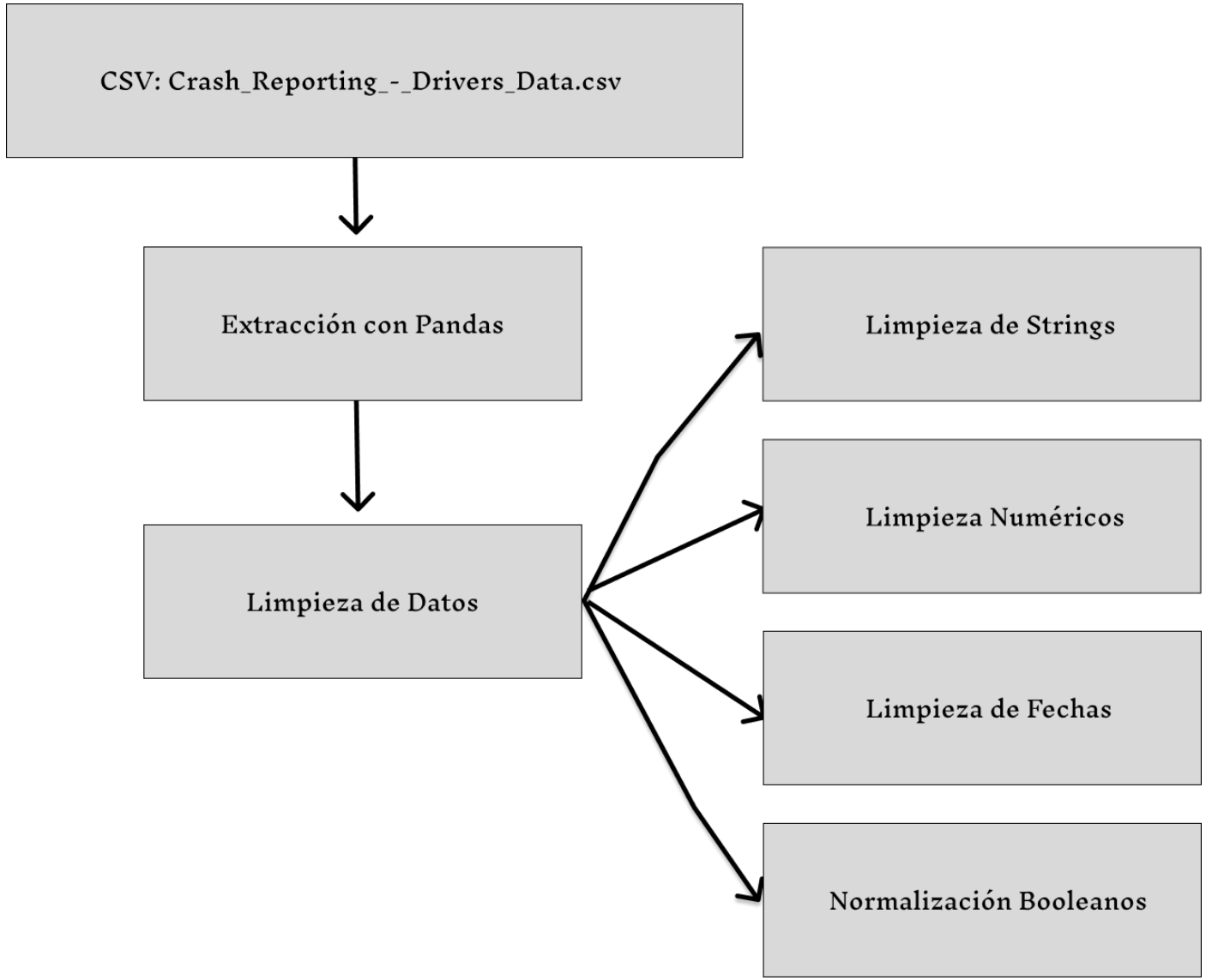
PYTHON

POSTGRESQL

EXTRACCIÓN DE DATOS

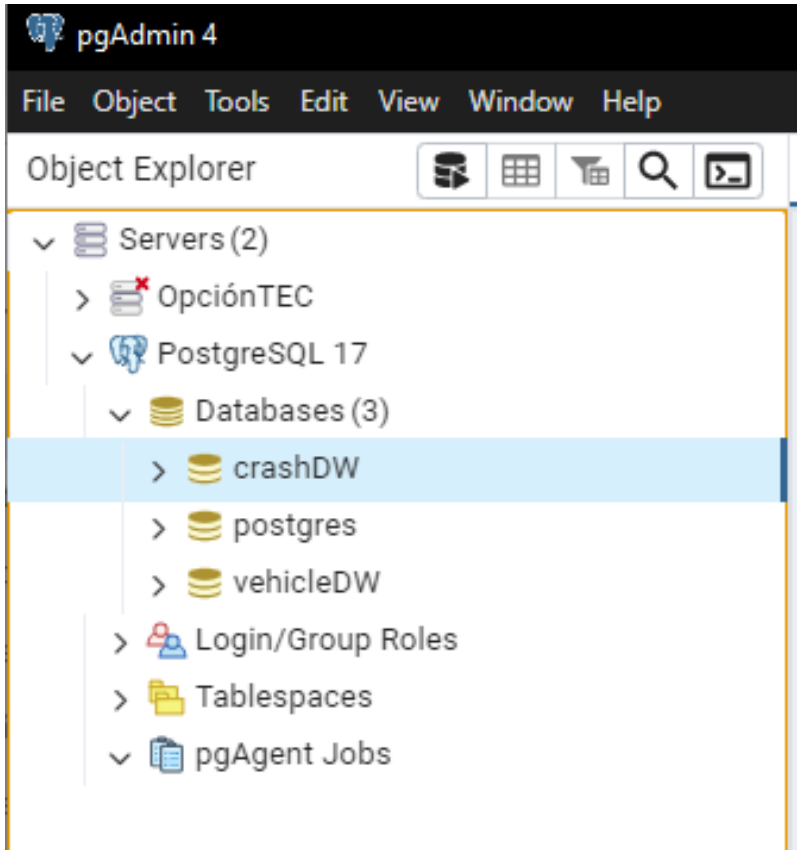
- Fecha
- Ubicación
- Clima
- Conductor
- Vehículo

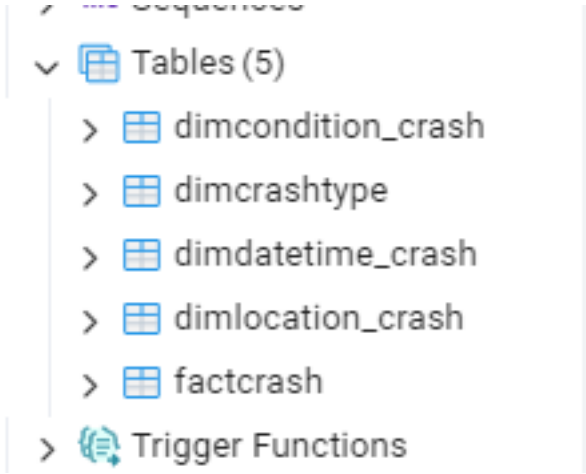




TRANSFORMACIÓN Y LIMPIEZA

CARGA EN POSTGRESQL





CONSULTAS ANALÍTICAS Y RESULTADOS

Las consultas abordan tres áreas principales:

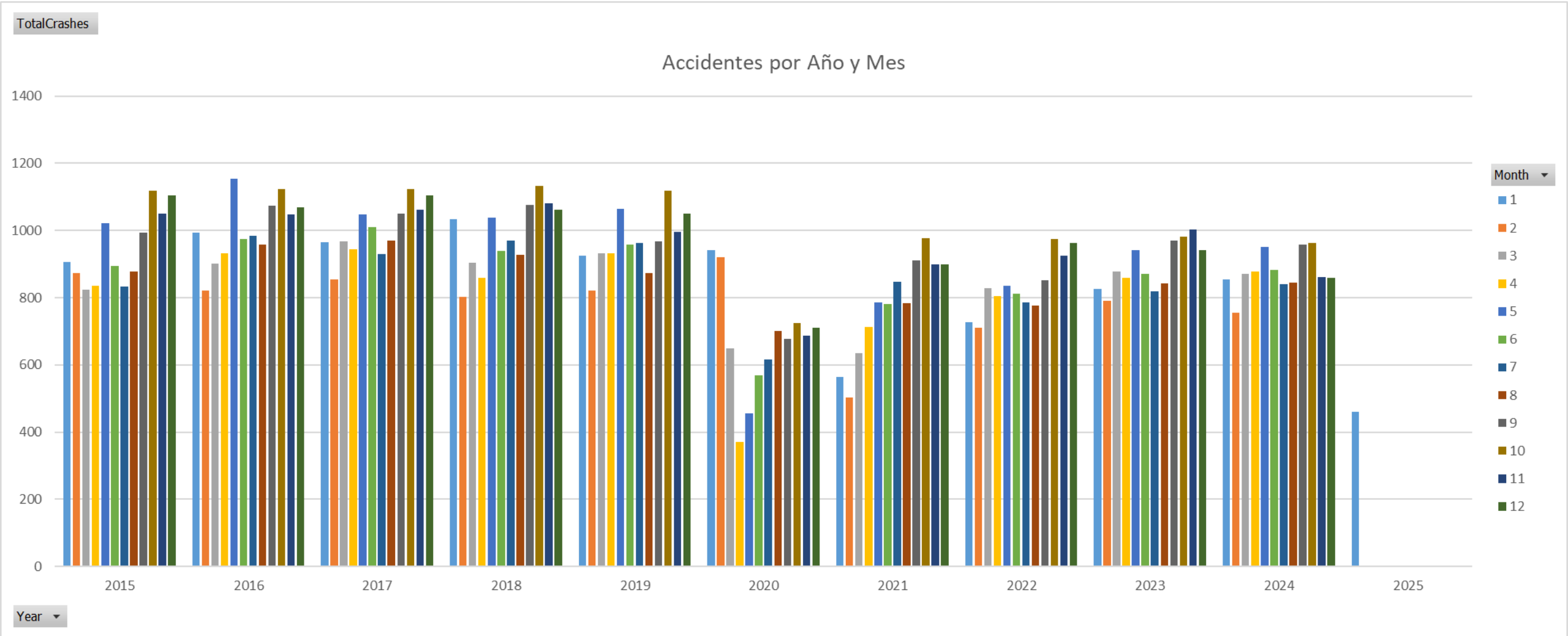
- Indicadores Clave de Rendimiento (KPIs) y Totales
- Métricas de Severidad y Promedios
- Análisis de Patrones Complejos

ANÁLISIS TEMPORAL: ACCIDENTES POR AÑO Y MES

Query | Query History

```sql
SELECT
    dt.year,
    dt.month,
    COUNT(fc.fact_crash_id) AS total_crashes
FROM FactCrash AS fc
JOIN DimDateTime_Crash AS dt
    ON fc.date_key_crash = dt.date_key_crash
GROUP BY dt.year, dt.month
ORDER BY dt.year, dt.month;
```

TotalCrashes	Mes												
Año	1	2	3	4	5	6	7	8	9	10	11	12	Total general
2015	906	873	824	835	1022	894	832	879	993	1119	1049	1104	11330
2016	993	822	901	933	1155	974	984	959	1074	1123	1047	1068	12033
2017	965	854	967	944	1047	1009	930	970	1050	1123	1061	1105	12025
2018	1034	803	904	860	1039	940	969	928	1077	1133	1080	1061	11828
2019	926	822	933	932	1064	958	963	874	967	1118	997	1049	11603
2020	942	920	649	370	455	570	617	702	677	724	686	711	8023
2021	564	504	636	714	787	781	847	783	911	977	898	899	9301
2022	728	711	829	804	836	812	785	777	853	974	925	964	9998
2023	826	790	878	860	941	870	820	843	969	981	1002	942	10722
2024	854	755	872	879	952	883	841	845	958	964	862	858	10523
2025	461												461
Total general	**9199**	**7854**	**8393**	**8131**	**9298**	**8691**	**8588**	**8560**	**9529**	**10236**	**9607**	**9761**	**107847**


TotalCrashes
Accidentes por Año y Mes
1400
1200
1000
800
600
400
200
0
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
Month
1
2
3
4
5
6
7
8
9
10
11
12
Year

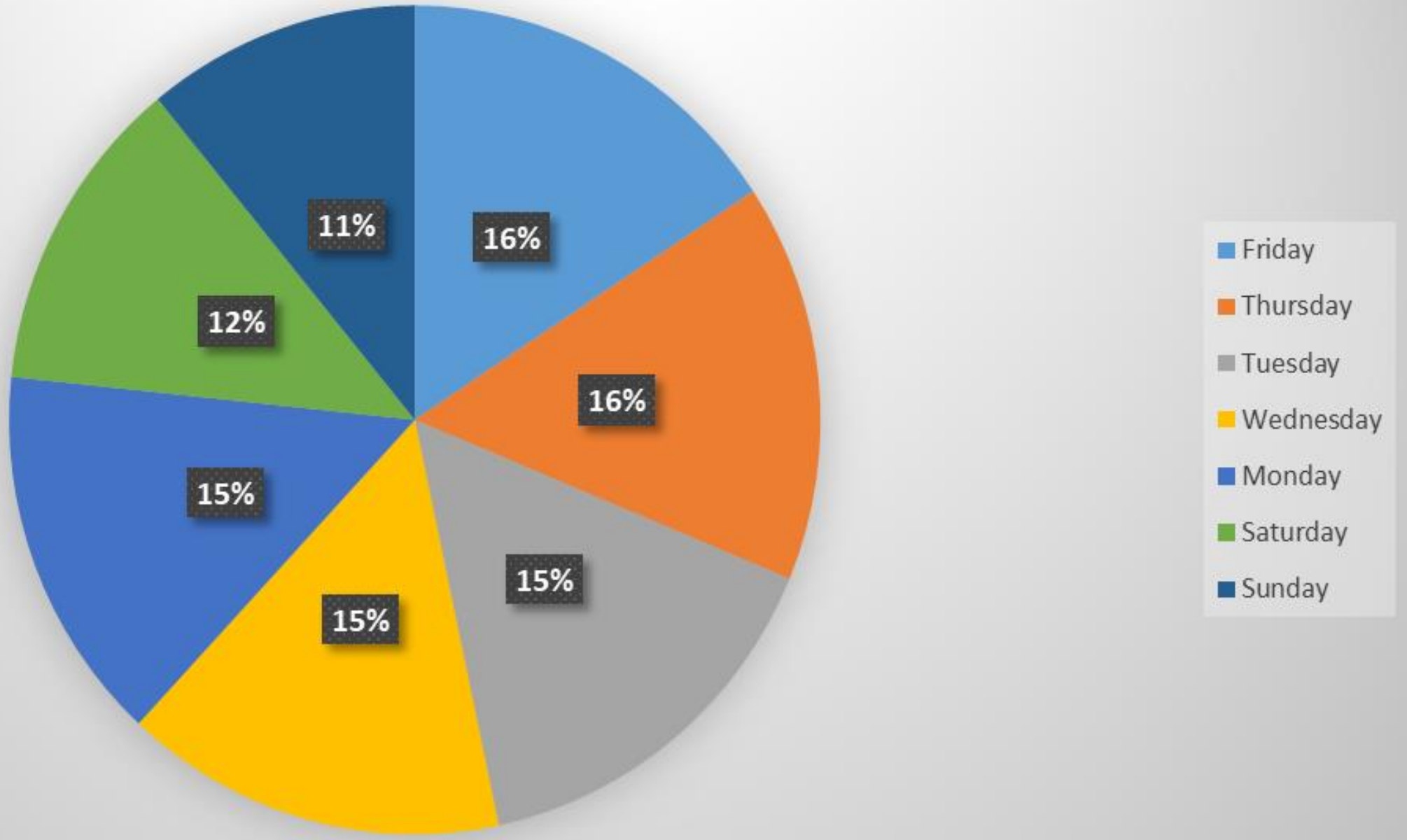
ANÁLISIS DE SEVERIDAD POR DÍA DE LA SEMANA

Query | Query History

```sql
WITH TotalInjuries AS (
    SELECT SUM(fc.num_injuries) AS total_injuries
    FROM FactCrash AS fc
)
SELECT
    dt.day_of_week,
    SUM(fc.num_injuries) AS total_injuries_day,
    ROUND((SUM(fc.num_injuries) * 100.0) /
        (SELECT total_injuries FROM TotalInjuries), 2)
        AS percentage_injuries
FROM FactCrash AS fc
JOIN DimDateTime_Crash AS dt
    ON fc.date_key_crash = dt.date_key_crash
GROUP BY dt.day_of_week
ORDER BY percentage_injuries DESC;
```

Week	Injuries	Percentage %
Friday	5533	15,7
Thursday	5484	15,56
Tuesday	5433	15,41
Wednesday	5370	15,24
Monday	5196	14,74
Saturday	4379	12,42
Sunday	3850	10,92

Análisis de severida por día de la semana



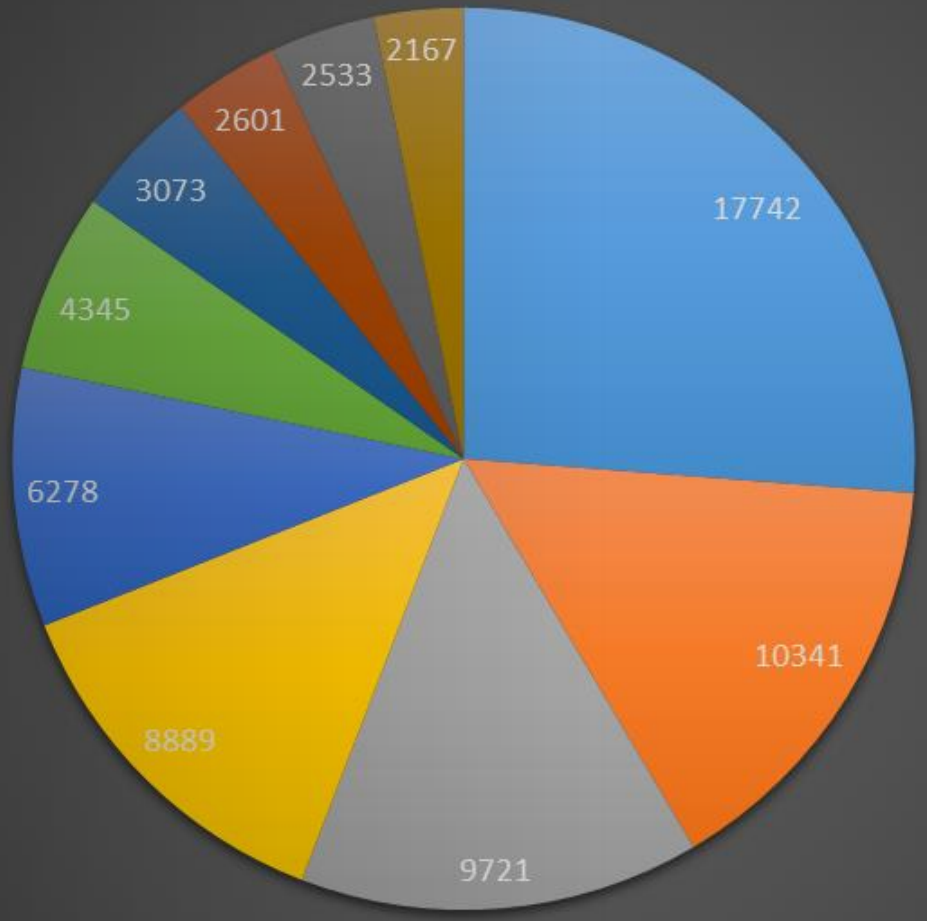
ANÁLISIS DE CLIMA Y TIPO DE COLISIÓN

Query | Query History

```sql
SELECT
    ct.collision_type,
    c.weather,
    COUNT(*) AS accident_count
FROM FactCrash f
JOIN DimCrashType ct
    ON f.crash_type_key = ct.crash_type_key
JOIN DimCondition_Crash c
    ON f.condition_key_crash = c.condition_key_crash
GROUP BY ct.collision_type, c.weather
ORDER BY accident_count DESC
LIMIT 10;
```

Colisión	Clima	# de accidente
SAME DIR REAR END	CLEAR	17742
STRAIGHT MOVEMENT ANGLE	CLEAR	10341
SINGLE VEHICLE	CLEAR	9721
OTHER	CLEAR	8889
SAME DIRECTION SIDESWIPE	CLEAR	6278
HEAD ON LEFT TURN	CLEAR	4345
SAME DIR REAR END	RAINING	3073
SINGLE VEHICLE	RAINING	2601
SAME DIR REAR END	CLOUDY	2533
Front to Rear	Clear	2167

Número de accidentes por Tipo de colisión y Clima

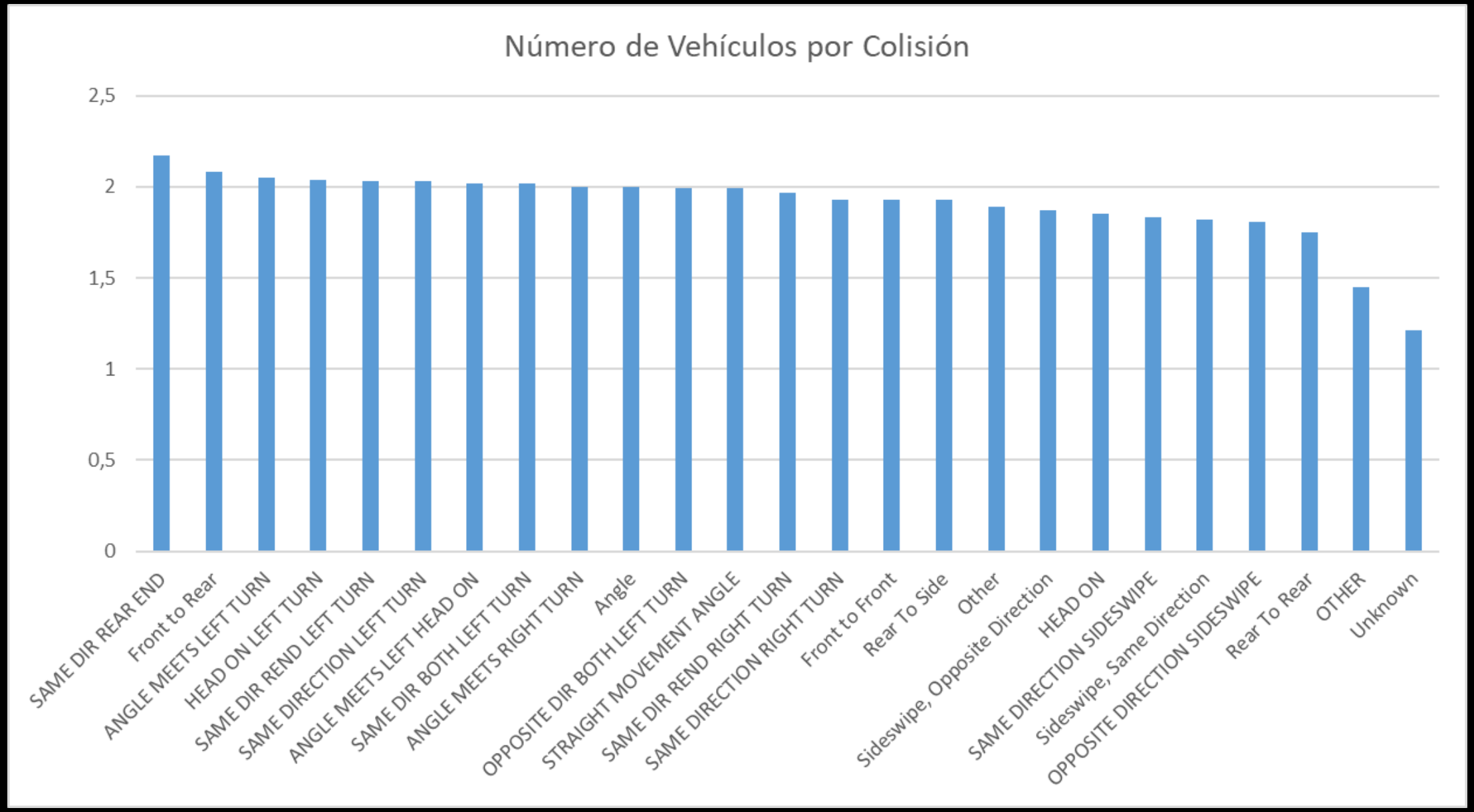


ANÁLISIS DE VEHÍCULOS POR TIPO DE COLISIÓN

Query | Query History

```sql
SELECT
    ctype.collision_type,
    ROUND(AVG(fc.num_vehicles_involved), 2)
    AS avg_vehicles_involved
FROM FactCrash AS fc
JOIN DimCrashType AS ctype
    ON fc.crash_type_key = ctype.crash_type_key
GROUP BY ctype.collision_type
ORDER BY avg_vehicles_involved DESC
LIMIT 25;
```

Colisión	# vehículo
SAME DIR REAR END	2,17
Front to Rear	2,08
ANGLE MEETS LEFT TURN	2,05
HEAD ON LEFT TURN	2,04
SAME DIR REND LEFT TURN	2,03
SAME DIRECTION LEFT TURN	2,03
ANGLE MEETS LEFT HEAD ON	2,02
SAME DIR BOTH LEFT TURN	2,02
ANGLE MEETS RIGHT TURN	2
Angle	2
OPPOSITE DIR BOTH LEFT TURN	1,99
STRAIGHT MOVEMENT ANGLE	1,99
SAME DIR REND RIGHT TURN	1,97
SAME DIRECTION RIGHT TURN	1,93
Front to Front	1,93
Rear To Side	1,93
Other	1,89
Sideswipe, Opposite Direction	1,87
HEAD ON	1,85
SAME DIRECTION SIDESWIPE	1,83
Sideswipe, Same Direction	1,82
OPPOSITE DIRECTION SIDESWIPE	1,81
Rear To Rear	1,75
OTHER	1,45
Unknown	1,21

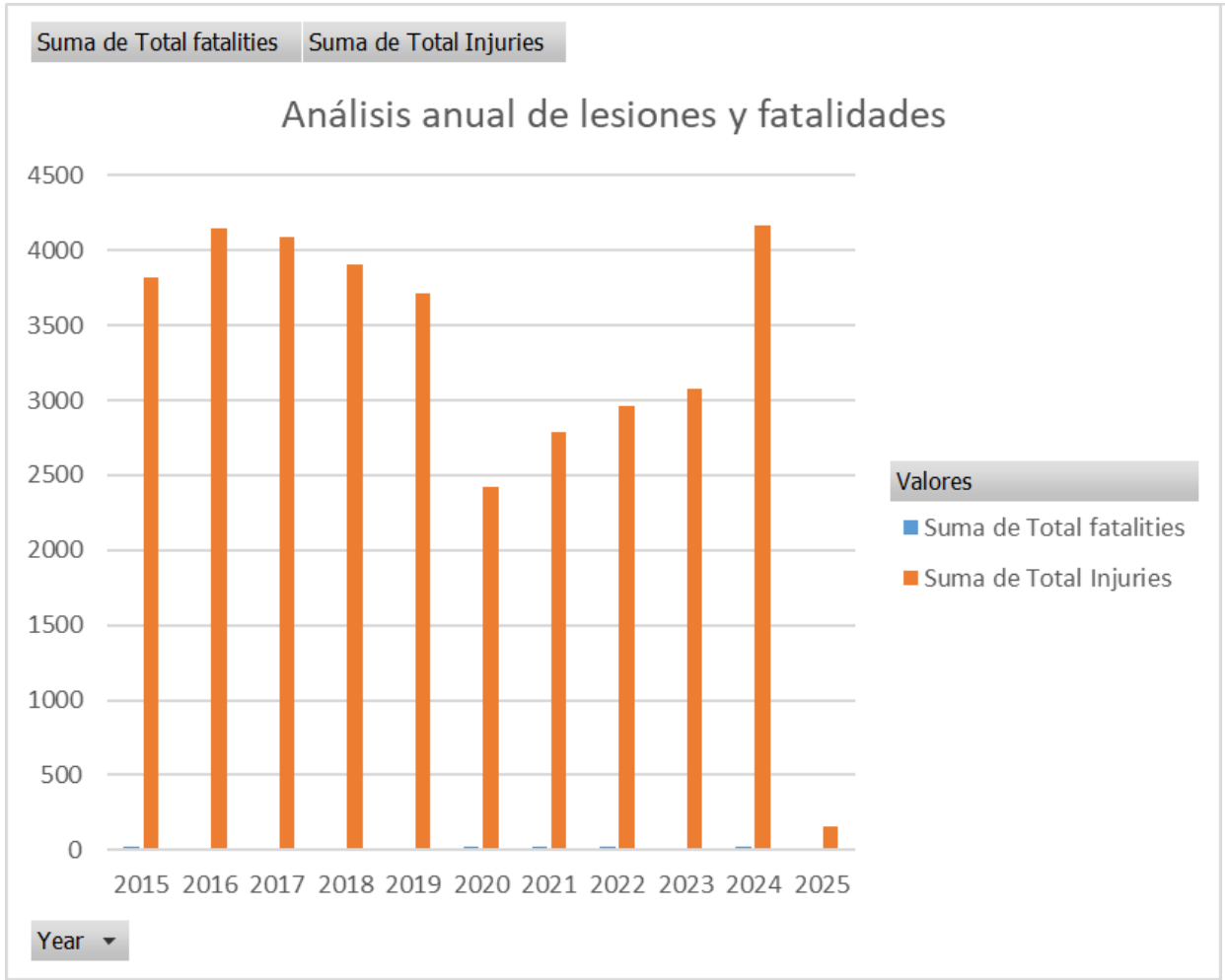

Número de Vehículos por Colisión
2,5
2
1,5
1
0,5
0
SAME DIR REAR END
Front to Rear
ANGLE MEETS LEFT TURN
HEAD ON LEFT TURN
SAME DIR REND LEFT TURN
SAME DIRECTION LEFT TURN
ANGLE MEETS LEFT HEAD ON
SAME DIR BOTH LEFT TURN
ANGLE MEETS RIGHT TURN
Angle
OPPOSITE DIR BOTH LEFT TURN
STRAIGHT MOVEMENT ANGLE
SAME DIR REND RIGHT TURN
SAME DIRECTION RIGHT TURN
Front to Front
Rear To Side
Other
Sideswipe, Opposite Direction
HEAD ON
SAME DIRECTION SIDESWIPE
Sideswipe, Same Direction
OPPOSITE DIRECTION SIDESWIPE
Rear To Rear
OTHER
Unknown

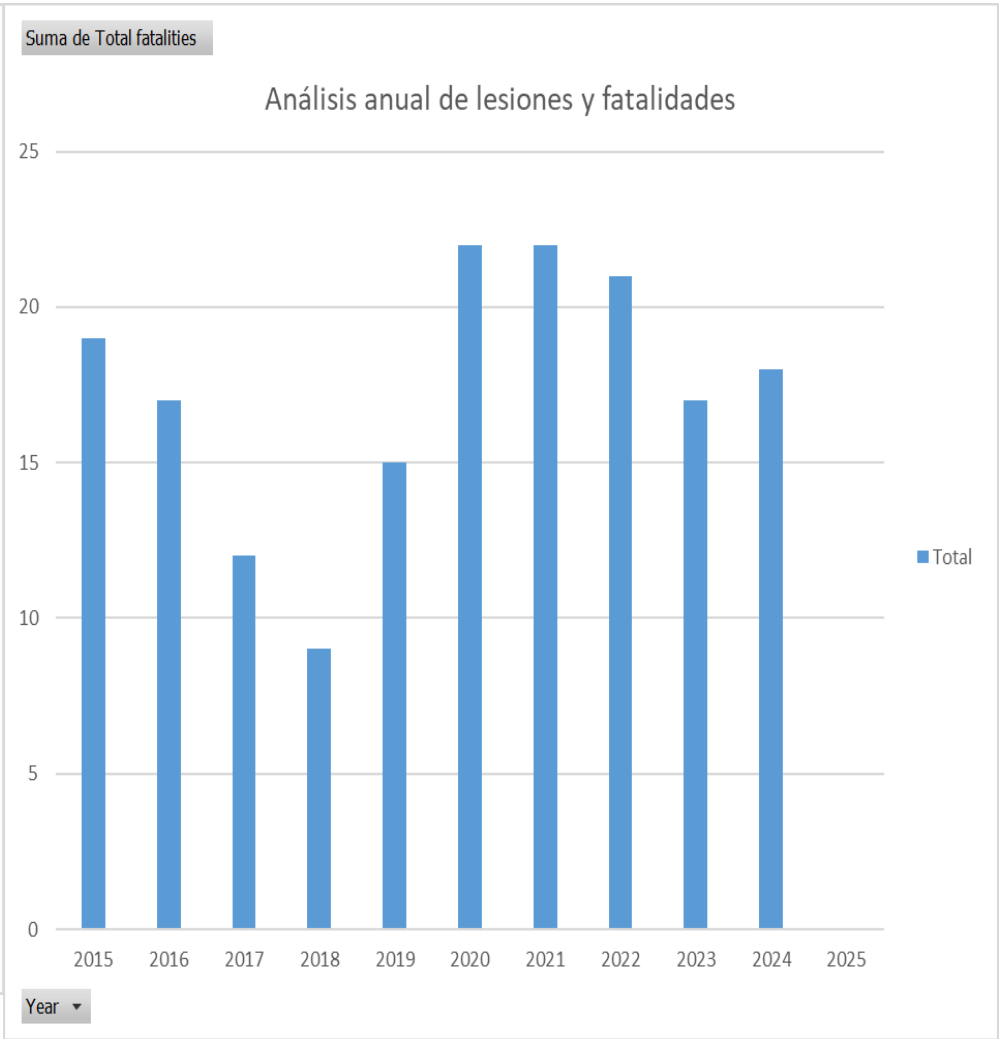
ANÁLISIS ANUAL DE LESIONES Y FATALIDADES

Query | Query History

```sql
SELECT
    dt.year,
    SUM(fc.num_injuries) AS total_injuries,
    SUM(fc.num_fatalities) AS total_fatalities
FROM DimDateTime_Crash dt
JOIN FactCrash fc
    ON dt.date_key_crash = fc.date_key_crash
GROUP BY dt.year
ORDER BY dt.year;
```

Etiquetas de fila	Suma de Total Injuries	Suma de Total fatalities
2015	3822	19
2016	4143	17
2017	4087	12
2018	3906	9
2019	3717	15
2020	2424	22
2021	2784	22
2022	2957	21
2023	3081	17
2024	4164	18
2025	160	0
Total general	**35245**	**172**


Suma de Total fatalities
Suma de Total Injuries
Análisis anual de lesiones y fatalidades
4500
4000
3500
3000
2500
2000
1500
1000
500
0
2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025
Valores
Suma de Total fatalities
Suma de Total Injuries
Year

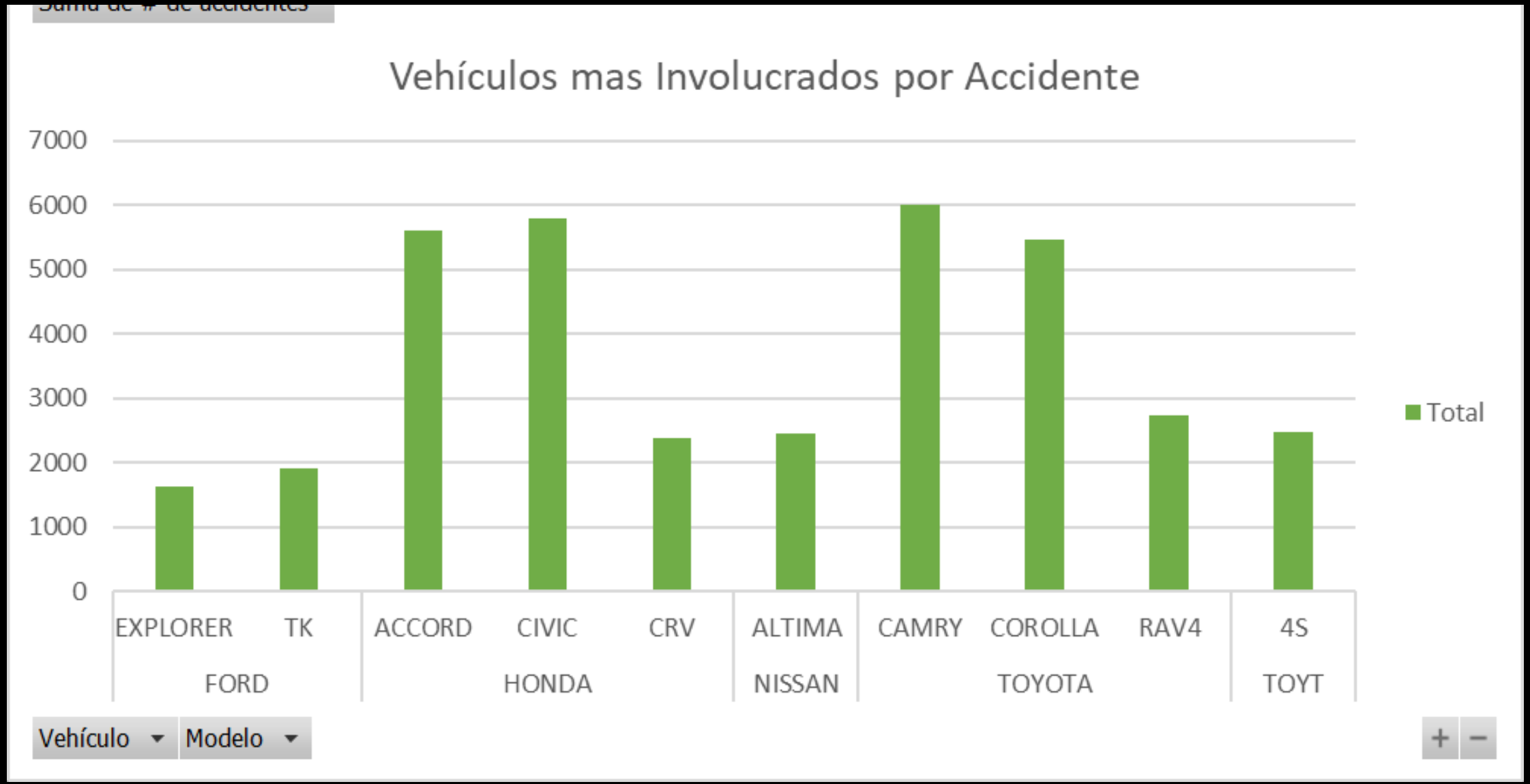

Suma de Total fatalities
Análisis anual de lesiones y fatalidades
25
20
15
10
5
0
2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025
Total
Year

ANÁLISIS DE LOS VEHÍCULOS INVOLUCRADOS

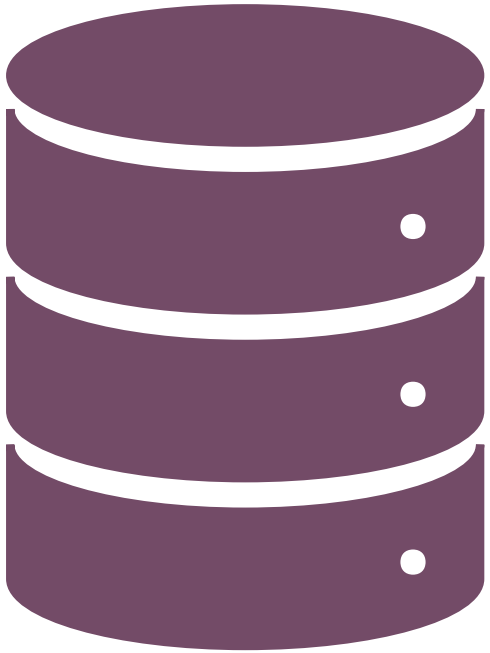
Query | Query History

```sql
SELECT
    v.vehicle_make,
    v.vehicle_model,
    COUNT(fv.fact_vehicle_id) AS total_accidentes
FROM
    DimVehicle v
JOIN
    FactVehicleInvolment fv ON v.vehicle_key = fv.vehicle_key
WHERE v.vehicle_make <> ''
GROUP BY
    v.vehicle_make, v.vehicle_model
ORDER BY
    total_accidentes DESC
LIMIT 10;
```

Etiquetas de fila	Suma de # de accidentes
FORD	**3554**
EXPLORER	1636
TK	1918
HONDA	**13783**
ACCORD	5612
CIVIC	5780
CRV	2391
NISSAN	**2457**
ALTIMA	2457
TOYOTA	**14179**
CAMRY	5998
COROLLA	5452
RAV4	2729
TOYT	**2472**
4S	2472
Total general	**36445**


Vehículos mas Involucrados por Accidente
7000
6000
5000
4000
3000
2000
1000
0
EXPLORER
TK
ACCORD
CIVIC
CRV
ALTIMA
CAMRY
COROLLA
RAV4
4S
FORD
HONDA
NISSAN
TOYOTA
TOYT
Total
Vehículo
Modelo

RETOS ENFRENTADOS Y SOLUCIONES IMPLEMENTADAS



- Manejo de datos inconsistentes y sucios
- Evitar duplicados en las tablas de dimensiones
- Conexión y manejo de múltiples bases de datos
- Normalización de campos booleanos

CONCLUSIONES

MODELO DIMENSIONAL

PROCESO ETL

INSIGHTS CLAVES

HERRAMIENTAS UTILIZADAS