



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

# **Análisis de datos de accidentalidad vial de la ciudad de Bogotá a partir de datos abiertos y datos obtenidos desde redes sociales**

**Fernando Vargas Montero**

Universidad Nacional de Colombia  
Departamento de ingeniería de Sistemas e Industrial  
Bogotá, Colombia

2022



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

# **Análisis de datos de accidentalidad vial de la ciudad de Bogotá a partir de datos abiertos y datos obtenidos desde redes sociales**

**Fernando Vargas Montero**

Tesis o trabajo de investigación presentada(o) como requisito parcial para optar al título de:

**Magister en Ingeniería de Sistemas y Computación**

Director (a):

PhD. Cesar Augusto Pedraza

Línea de Investigación:

Sistemas inteligentes

Grupo de Investigación:

PLaS

Universidad Nacional de Colombia

Departamento de ingeniería de sistemas e industrial

Bogotá, Colombia

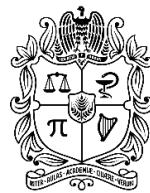
2022



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

*(Dedicatoria o lema)*

*Inicialmente agradezco a Dios. A toda mi familia por siempre estar apoyándome en mi gusto por la investigación y en la mejora profesional. A mi tutor y profesor Cesar Augusto Pedraza por estar siempre guiándome en todo momento, un excelente profesor. A mi novia que estuvo apoyándome emocionalmente para continuar de manera exitosa el proceso de desarrollo del trabajo presente.*



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

## **Declaración de obra original**

Yo declaro lo siguiente:

He leído el Acuerdo 035 de 2003 del Consejo Académico de la Universidad Nacional. «Reglamento sobre propiedad intelectual» y la Normatividad Nacional relacionada al respeto de los derechos de autor. Esta disertación representa mi trabajo original, excepto donde he reconocido las ideas, las palabras, o materiales de otros autores.

Cuando se han presentado ideas o palabras de otros autores en esta disertación, he realizado su respectivo reconocimiento aplicando correctamente los esquemas de citas y referencias bibliográficas en el estilo requerido.

He obtenido el permiso del autor o editor para incluir cualquier material con derechos de autor (por ejemplo, tablas, figuras, instrumentos de encuesta o grandes porciones de texto).

Por último, he sometido esta disertación a la herramienta de integridad académica, definida por la universidad.

Fernando Vargas Montero

Nombre

Fecha 18/03/2022



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

## **Agradecimientos**

Agradezco a Dios y a la Virgen María por siempre darme la oportunidad de ser mejor persona y asumir retos en la vida, por darme la salud y los medios para poder seguir con mis estudios y metas.

Mi familia que siempre ha estado presente apoyándome en cada momento de mi formación académica, desde primaria hasta ahora siendo candidato a Magister en la Universidad Nacional de Colombia.

Agradezco a la Universidad Nacional de Colombia por formarme como profesional. A cada profesor que conocí y aprendí de sus conocimientos, en especial a mi tutor del presente trabajo Cesar Augusto Pedraza que desde pregrado me enseñó bastantes cosas y en el proceso de la tesis dedico bastante tiempo al desarrollo de este trabajo.

## Resumen

### **Análisis de datos de accidentalidad vial de la ciudad de Bogotá a partir de datos abiertos y datos obtenidos desde redes sociales**

La accidentalidad vial es un tema que siempre se desea minimizar, pero no se han aplicado técnicas tecnológicas para encontrar y entender el comportamiento vial según las distintas variables que intervienen. Por esta necesidad se genera un motivo de poder aportar a esta comprensión de la accidentalidad vial para así mismo apoyar en la mitigación y de este modo lograr salvar vidas. En este trabajo se encontrará desde las arquitectura y técnicas de extracción de los datos viales en la ciudad de Bogotá, hasta el análisis estadístico para poder comprender el comportamiento de este.

Para poder llegar al análisis estadístico inicialmente se realizó una arquitectura autosostenible de recolección de datos viales en la ciudad de Bogotá a través de redes sociales. Al tener ya recolectada la información se procedió a su representación en una herramienta geográfica para su comprensión visual. Se procede a un análisis inicial de estos datos. Luego se aplican técnicas estadísticas geográficas para su análisis estadístico.

**Palabras clave:** Análisis espacial, geoestadística, Waze, Twitter, Minería de datos espaciales, Tráfico, Accidentes.



## **Abstract**

### **Analysis of road accident data in Bogotá city from open data and data on social networks**

Road accidents are an issue that we always want to minimize, but technological techniques have not been applied to find and understand road behavior according to the different variables involved. Due to this need, a reason is generated to be able to contribute to this understanding of road accidents to support mitigation and thus save lives. In this work you will find everything from the architecture and techniques of extracting road data in the city of Bogotá, to the statistical analysis to understand its behavior.

To reach the statistical analysis, a self-sustaining architecture for road data collection was initially carried out in the city of Bogotá through social networks. Having already collected the information, it was represented in a geographic tool for visual understanding. An initial analysis of these data is carried out, then geographical statistical techniques are then applied for statistical analysis.

**Keywords: Spatial analysis, Geostatistics, Waze, Twitter, Spatial data mining, Traffic, Accidents.**



Este Trabajo Final de maestría fue calificado en mayo de 2022 por el siguiente evaluador:

Camilo Albeiro Gutiérrez Osorio Msc.  
Facultad de Ingeniería  
Universidad Nacional de Colombia

# Contenido

	Pág.
<b>1. Antecedentes.....</b>	<b>5</b>
<b>2. Marco teórico.....</b>	<b>10</b>
2.1 Web scraping .....	10
2.2 Base de datos MongoDB.....	11
2.3 Minería de datos espacial.....	11
2.4 Sistemas de Información Geográfica.....	13
2.4.1 QGIS .....	13
<b>3. Web scraping en tráfico y accidentes viales .....</b>	<b>14</b>
3.1 Web scraping utilizando waze como fuente de información.....	15
3.1.1 Petición http para extracción de datos .....	15
3.1.2 Cuadrícula Bogotá usando Tiles de Google.....	17
3.2 Alertas viales en Waze .....	18
3.3 Segmentos de tráfico en Waze.....	19
3.4 Web scraping en red social Twitter.....	20
3.5 Arquitectura de recolección de datos.....	21
<b>4. Análisis Tráfico vial.....</b>	<b>24</b>
4.1 Cantidad de segmentos de tráfico durante días la semana .....	25
4.1.1 Tráfico durante días de la semana por mes con nivel 3 y 4 .....	25
4.1.2 Tráfico durante días de la semana por mes con nivel 5 .....	29
4.2 Tráfico a través de las horas del día.....	33
4.2.1 Tráfico por mes durante las horas del día con nivel 3 y 4 .....	34
4.2.2 Tráfico por mes durante las horas del día con nivel 5 .....	35
<b>5. Análisis accidentalidad vial.....</b>	<b>37</b>
5.1 Exploración de los datos de Waze Alert .....	38
5.1.1 Tipos de alerta .....	38
5.2 Heat maps.....	40
5.2.1 Heat maps con la totalidad de los datos .....	40
5.2.2 Heat maps con la totalidad de los datos tipo ACCIDENT_MAJOR.....	41
5.2.3 Heat maps con la totalidad de los datos tipo ACCIDENT_MINOR .....	42
5.2.4 Heat maps con la totalidad de los datos tipo ACCIDENT .....	43
5.3 Puntos de concentración de accidentes .....	44
5.4 Análisis de huecos en Bogotá .....	47
5.4.1 Huecos o defectos en la vía en los puntos de concentración de accidentes ..	49

---

5.5	Análisis accidentes por tipo de vía .....	53
5.6	Análisis accidentes por día de la semana.....	54
5.7	Análisis accidentes por hora.....	55
5.8	Alertas accidentes viales Secretaría de Movilidad en Waze .....	60
5.8.1	Total accidentes por tipo por hora.....	62
5.8.2	Duración promedio accidentes por tipo .....	62
5.9	Accidentes extraídos de Twitter.....	64
5.9.1	Heat map víctimas fatales.....	65
5.9.2	Heat map víctimas lesionadas .....	66
5.9.3	Heat map solo daños.....	67
5.10	Localidades con mayor cantidad de accidentes .....	68
5.10.1	Identificación de localidad-accidente.....	68
5.10.2	Número de accidentes por localidad.....	70
<b>6.</b>	<b>Conclusiones y recomendaciones.....</b>	<b>72</b>
6.1	Conclusiones.....	72
6.2	Recomendaciones.....	74

## Lista de figuras

	Pág.
<b>Figura 1.</b> Total accidentes con lesionados por año en Bogotá .....	8
<b>Figura 2.</b> Total accidentes con víctima fatal por año en Bogotá .....	9
<b>Figura 3.</b> Históricos accidentes por localidad en Bogotá entre el año 2015-2019 .....	9
<b>Figura 4.</b> Live map de Waze, presenta alertas y segmentos de tráfico. ....	15
<b>Figura 5.</b> Parámetros de la petición GET al endpoint de datos de alertas y tráfico de Waze.....	16
<b>Figura 6.</b> Tile Bounds o bounding box definidos por MapTiler de Google en Bogotá. ....	17
<b>Figura 7.</b> Datos en el objeto que representa una alerta de Waze.....	18
<b>Figura 8.</b> Datos en el objeto que representa un segmento de tráfico de Waze. ....	19
<b>Figura 9.</b> Datos en el objeto extraído del api de Twitter que representa un tweet. ....	21
<b>Figura 10.</b> Datos en el objeto extraído del api de Twitter que representa un tweet. ....	22
<b>Figura 11.</b> Cantidad de segmentos de tráfico a través de la semana.....	25
<b>Figura 12.</b> Cantidad de segmentos de tráfico de nivel 3-4 a través de la semana del mes de septiembre .....	26
<b>Figura 13.</b> Cantidad de segmentos de tráfico de nivel 3-4 a través de la semana del mes de octubre .....	27
<b>Figura 14.</b> Cantidad de segmentos de tráfico de nivel 3-4 a través de la semana del mes de noviembre .....	28
<b>Figura 15.</b> Cantidad de segmentos de tráfico de nivel 3-4 a través de la semana del mes de diciembre.....	29
<b>Figura 16.</b> Cantidad de segmentos de tráfico de nivel 5 a través de la semana del mes de septiembre .....	30
<b>Figura 17.</b> Cantidad de segmentos de tráfico de nivel 5 a través de la semana del mes de octubre .....	31
<b>Figura 18.</b> Cantidad de segmentos de tráfico de nivel 5 a través de la semana del mes de noviembre .....	32
<b>Figura 19.</b> Cantidad de segmentos de tráfico de nivel 5 a través de la semana del mes de diciembre.....	33
<b>Figura 20.</b> Cantidad de segmentos de tráfico a través de las horas del día .....	34
<b>Figura 21.</b> Cantidad de segmentos de tráfico de nivel 3-4 a través de las horas del día por mes .....	35

---

<b>Figura 22.</b> Cantidad de segmentos de tráfico de nivel 5 a través de las horas del día por mes .....	36
<b>Figura 23.</b> Totalidad de accidentes por mes .....	37
<b>Figura 24.</b> Total de alertas por tipo .....	38
<b>Figura 25.</b> Total de accidentes por subtipo .....	39
<b>Figura 26.</b> Heat map accidentes viales .....	40
<b>Figura 27.</b> Heat map accidentes viales de tipo ACCIDENT_MAJOR .....	41
<b>Figura 28.</b> Heat map accidentes viales de tipo ACCIDENT_MINOR .....	42
<b>Figura 29.</b> Heat map accidentes viales de tipo ACCIDENT .....	43
<b>Figura 30.</b> Heat map accidentes viales en Kennedy .....	44
<b>Figura 31.</b> Heat map accidentes viales en Salitre Mágico .....	45
<b>Figura 32.</b> Heat map accidentes viales en Autopista Norte con calle 100 .....	46
<b>Figura 33.</b> Heat map accidentes viales en Av. Boyacá con calle 80 .....	47
<b>Figura 34.</b> Heat map huecos o defectos en vías de Bogotá .....	48
<b>Figura 35.</b> Heat map huecos o defectos en Kennedy .....	49
<b>Figura 36.</b> Heat map huecos o defectos en Salitre Mágico .....	50
<b>Figura 37.</b> Heat map huecos o defectos en Autopista Norte con calle 100 .....	51
<b>Figura 38.</b> Heat map huecos o defectos en Av. Boyacá con calle 80 .....	52
<b>Figura 39.</b> Totalidad de accidentes por tipo de vía .....	54
<b>Figura 40.</b> Totalidad de accidentes por día de la semana .....	55
<b>Figura 41.</b> Totalidad de accidentes por hora para días de fin de semana .....	56
<b>Figura 42.</b> Totalidad de accidentes por hora para días de la semana laboral .....	57
<b>Figura 43.</b> Totalidad de accidentes por hora para días de la semana .....	58
<b>Figura 44.</b> Totalidad de segmentos de tráfico nivel 3-4 por hora para días de la semana .....	59
<b>Figura 45.</b> Totalidad de alertas Secretaria Movilidad por tipo .....	60
<b>Figura 46.</b> Totalidad de alertas Secretaria Movilidad por tipo por hora .....	62
<b>Figura 47.</b> Totalidad de alertas Twitter por tipo por hora .....	64
<b>Figura 48.</b> Heat map de accidentes con víctimas fatales .....	65
<b>Figura 49.</b> Heat map de accidentes con lesionados .....	66
<b>Figura 50.</b> Heat map de accidentes con sólo daños .....	67
<b>Figura 51.</b> Mapa de localidades de Bogotá .....	69
<b>Figura 52.</b> Número de accidentes por localidad .....	70
<b>Figura 53.</b> Número de accidentes fatales por localidad .....	71
<b>Figura 54.</b> Cantidad segmentos de tráfico través de las horas por día de la semana desde nuevo pico y placa 2022 .....	75

## Lista de tablas

	<b>Pág.</b>
<b>Tabla 1.</b> Tipos de vía en las alertas de Waze.....	53
<b>Tabla 2.</b> Totalidad de alertas Secretaria Movilidad por tipo y nivel de accidente .....	61
<b>Tabla 3.</b> Duración promedio accidentes por tipo .....	63

# Introducción

Los accidentes de tránsito en la ciudad de Bogotá son muy frecuentes [1]. Analizar los accidentes de tránsito de la ciudad de Bogotá es importante para dar un amplio conocimiento y una identificación de las variables que mayor aumentan los accidentes. En el presente trabajo se avanza con el objetivo de entender por qué y en qué contexto ocurren los accidentes de tránsito en Bogotá.

Los siniestros viales en la ciudad de Bogotá son muy recurrentes, según un estudio de TransitoBogota en el año 2019 un accidente ocurre cada 5,6 minutos, siendo esto una cifra importante que se debe mitigar [1]. Adicionalmente muchos de estos accidentes han sido con víctimas fatales. Según un estudio en el 2019, por cada dos homicidios que ocurren se presenta una víctima fatal por un accidente vial. Esta es una cifra realmente aterradora, esto quiere decir que son altas las muertes por accidentes viales. El análisis de este proyecto servirá como base del entendimiento de los accidentes de tránsito en tiempo, lugar, frecuencia y contexto.

Asimismo, no se cuenta con un acceso directo a la recolección de datos de accidentes viales, así como todo tipo de eventualidad vial como lo podría ser el tráfico, estado de las vías, etc. Para ello se presenta una necesidad de recolectar la información vial en la ciudad Bogotá para su pronto y constante análisis. Esta extracción de datos de accidentalidad de tránsito no se está realizando de manera rigurosa y constante [2] y por ende los análisis son cortos y pobres. En el presente trabajo se realiza una arquitectura para la dinámica de una recolección de datos vial de redes sociales de manera automatizada y autosostenible. Se extraen de manera automática los datos viales y se persisten en una base de datos para su posterior análisis.

Se entenderá por accidente vial según definido por el acuerdo al artículo 2° del código Nacional de Tránsito Ley 769 del 2002, “evento generalmente involuntario, generado al

menos por un vehículo en movimiento, que causa daños a personas y bienes involucrados en él e igualmente afecta la normal circulación de los vehículos que se movilizan por la vía o las vías comprendidas en el lugar o dentro de la zona de influencia del hecho.”.

El objetivo principal del presente trabajo es extraer y realizar un análisis estadístico de datos de accidentalidad vial en la ciudad de Bogotá, obtenidos a partir de datos abiertos y datos de redes sociales.

Para lograr este objetivo se proponen unos objetivos específicos:

1. Extraer datos de accidentalidad mediante técnicas de extracción web scraping.
2. Aplicar técnicas de limpieza y estructuración a datos extraídos.
3. Aplicar técnicas para análisis de datos geoespaciales, relacionadas con la accidentalidad vial.
4. Realizar análisis estadístico descriptivo de los datos geográficos extraídos.

En el desarrollo del trabajo y con fin de lograr estos objetivos descritos, se realiza un análisis con enfoque cuantitativo en los datos obtenidos en la accidentalidad y tráfico en la ciudad de Bogotá. Con este enfoque cuantitativo se encontrarán las variables que intervienen en el comportamiento de los accidentes de tránsito. Con herramientas estadísticas y geoestadísticas se analizarán los datos para lograr tener una mejor precisión y sustento en las conclusiones dadas.

Para lograr la obtención de los datos se aplicó una técnica de web scraping, mediante peticiones http. Se extrajeron datos del estado de tráfico, alertas y siniestros viales de la ciudad de Bogotá. Estos datos fueron extraídos de las redes sociales Waze y Twitter. Para cada uno de los datos extraídos, siempre se tiene fecha y ubicación del acontecimiento. Estos dos datos son los más importantes para lograr un análisis preciso y lograr analizar los datos mediante estas dos variables compartidas.

Se utilizó para el proceso de la investigación, formulación e implementación de los modelos estadísticos el enfoque de investigación de operaciones el cual se cumplirán todas las etapas:



1. **Orientación:** Se definieron los requerimientos para lograr cumplir cada objetivo específico.
2. **Definición del problema:** El objetivo en esta etapa fue definir aún más la descripción del problema desde la fase de orientación hasta el punto en que haya una definición clara del problema en términos de su alcance y los resultados deseados.
3. **Recolección de los datos:** Se investigó y se definió la manera y la calidad de la recolección de los datos.
4. **Formulación de los métodos:** Etapa importante donde se dio la primera aproximación de la abstracción de los métodos estadísticos del análisis.
5. **Métodos solución:** Solucionar el problema de identificar las variables en los accidentes viales mediante los métodos estadísticos formulados. Se realizó el proceso de desarrollo de estos.
6. **Validación y análisis:** Se analizarán los resultados obtenidos mediante los métodos estadísticos aplicados
7. **Implementación y monitoreo:** Se presentarán los análisis a diferentes grupos de datos para confirmar la validez de las conclusiones.

En la etapa de recolección de los datos se analizó la mejor manera para poder extraer la mayor cantidad y calidad en los datos. Se identificó la manera de poder extraer los datos de la red social Waze, en la cual extraemos diferentes tipos de datos, y es una fuente bastante confiable y amplia en el detalle de los datos recolectados. En la red social de Twitter, para poder tener datos confiables, se identificó una cuenta que publica los accidentes y estados de las vías en Bogotá. La cuenta es @BogotaTransito Realmente los datos del estado de las vías de Bogotá publicado por esta cuenta no son abundantes, así que se decidió solo recolectar los datos de accidentes o siniestros como se describe en los textos de los tweets de esta cuenta.

El presente documento organiza el proceso y desarrollo de la investigación en cinco capítulos. En el primer capítulo se exponen los antecedentes del tema a desarrollar, el cual se presentarán los estudios y trabajos realizados acerca del tema de investigación. En el segundo capítulo se definen y presentan las bases teóricas que se utilizan a lo largo de la investigación. A continuación, en el tercer capítulo se presenta la técnica y todo lo

relacionado con la recolección de los datos. En el cuarto capítulo se aplican los métodos estadísticos. En el quinto capítulo se presenta el análisis espacio temporal y las variables encontradas que intervienen en el compartimiento de la accidentalidad vial en la ciudad de Bogotá. Finalmente se exponen las conclusiones dado el análisis previo.

# 1. Antecedentes

La Plataforma Libanesa de Accidentes de Carretera (LRAP) propone una plataforma en línea en tiempo real para recopilar eventos de accidentes de las redes sociales. LRAP permite la recopilación, clasificación y visualización de datos autónomos sin intervención humana. Esta plataforma surgió por la falta de autoridad legal para recopilar y reportar datos relacionados con accidentes. LRAP es una plataforma novedosa para extraer datos relacionados con accidentes de tráfico de varias fuentes de cuentas confiables de redes sociales. La recopilación, clasificación y visualización de datos se realiza de forma autónoma y en tiempo real sin intervención humana [2].

Para una mejor clasificación e identificación de los datos geográficos es necesario la intervención humana, y para que esta sea eficiente se requiere representar los datos de manera visual. Esta visualización debe funcionar correctamente y ser representativa para una gran cantidad de datos [3].

Los datos espacio-temporales presentan muchos problemas de investigación interesantes y sin solucionar [5]. Para apoyar el objetivo de la solución de esta problemática se debe realizar una correcta visualización de los datos espacio-temporales, aplicando técnicas de big data orientadas a datos geográficos y obteniendo así un eficiente análisis y poder encontrar patrones en los datos [4]. Para poder analizar correctamente estos datos, se requiere descubrir y estructurar los parámetros de entrada de los datos geográficos recolectados y generar clústeres y poder agruparlos [6].

La recolección de datos de accidentalidad de tránsito es escasa y no se realiza de manera formal, un informe evidencia que solo se hace mediante los procedimientos de las agencias o departamentos de policía que registran la accidentalidad cuando asisten al caso del accidente [7].

El uso de big data para almacenar, manipular, analizar y extraer los conjuntos de datos de accidentes de tráfico permite la creación de conocimiento que puede ayudar a la toma de decisiones para disminuir la cantidad de accidentes. Un estudio utiliza los datos en tiempo real del tráfico de New York mediante herramientas web para la debida extracción [8]. El informe utiliza Hadoop y map-reduce para el cómputo y tratamiento de los datos.

Un informe utiliza los textos de la red social Twitter para extraer datos relacionados con el tránsito como estados de las vías, tráfico y accidentes. Este tipo de extracción puede sufrir el problema de la ambigüedad pragmática [9]. Además, se está realizando la propuesta de un sistema de enriquecimiento de datos de carretera, que es un sistema que pretende realizar la integración de varias fuentes de datos de tránsito para apoyar el funcionamiento de los sistemas de transporte inteligente (ITS). Este proyecto utiliza T-Maps.

Un estudio realizado en el año 2016 analizó los datos históricos de choques del estado de Dakota del Norte de 1996 a 2014, y se enfocó en detectar los factores más relevantes para los choques. La eficiencia de predicción del modelo es de 84.1%. Este informe hace énfasis en que los modelos de regresión tradicionales no funcionan bien con grandes cantidades de datos, y que por el contrario el árbol de decisión es más avanzado en su capacidad para manejar grandes conjuntos de datos, lidiar con valores perdidos y no requiere relaciones subyacentes predefinidas entre variables objetivo y predictores. [9]

Un informe realiza un análisis de accidentalidad para detectar factores relevantes en accidentalidad vial con un modelo de regresión de árbol de decisión (DTR), este informe se expuso en septiembre de 2019 y hace énfasis en que son los primeros en usar este tipo de modelos para datos de seguridad vial [10]. Realizaron una recolección de datos del año 2010 al 2012.

Otras investigaciones se enfocan en un análisis de escala en los datos recopilados con respecto a diferentes parámetros como el área del distrito, el tamaño de la población y la longitud de la red de carreteras. Los resultados muestran que los accidentes tienden a seguir el mismo factor de escala de la red de un organismo, haciendo que los accidentes actúen como organismos vivos en el comportamiento. El análisis temporal revela

tendencias estacionales, semanales y diarias en el agrupamiento de accidentes de tránsito [2].

Los accidentes de tránsito además generan costos adicionales, por ello es importante entenderlos y lograr minimizarlos. Un artículo de análisis de accidentes utiliza comportamiento de agrupamiento espacial y vulnerabilidad de riesgos de accidentes automovilísticos que ocurrieron en el Líbano entre 2015 y 2018. Analizan estos datos mediante densidades de puntos calientes utilizando el método Global G de autocorrelación espacial [11].

Se desea extraer y recolectar datos de accidentalidad de tránsito de la ciudad de Bogotá. Para este caso es importante poder extraer información espacio-temporal y otras características para fines de análisis. Además, se podrán analizar los tiempos de demora que causan estos accidentes en el tráfico, que es un tema poco estudiado [12]. Con estos datos se puede aplicar un correcto tratamiento de datos para aplicar técnicas de visualización que puedan representar el estado actual, el histórico y predicciones de estos [7].

La importancia de analizar los datos viales es que aproximadamente el 4% son accidentes con víctimas fatales, dando un resultado aproximado de que fallece una persona cada 23 horas [35]. Adicionalmente estos accidentes viales generan un aumento en el tráfico en las vías. Los accidentes viales vienen año tras año en crecimiento y es alarmante que siga creciendo. Esto genera adicionalmente un gasto en la economía. De acuerdo con las bases del PND 2018-2022, el impacto que esta problemática tiene sobre la economía colombiana asciende a 23,9 billones de pesos anuales, lo que equivale a 3,6% del Producto Interno Bruto [36].

En Bogotá por cada dos homicidios se presenta una muerte por accidente vial [1]. Cifra realmente alarmante y por ello es importante la analítica de los datos viales para así conocer más del comportamiento o patrones de estos accidentes viales.

La **Figura 1** expone el histórico de accidentes con lesionados por año en la ciudad de Bogotá. Las cifras presentadas son altas, y adicionalmente se evidencia un valor constante

en el total de accidentes de cada año, con excepción del año 2020, esto debido a que la pandemia por el COVID-19 promovió el aislamiento preventivo y luego obligatorio en ese año [37].

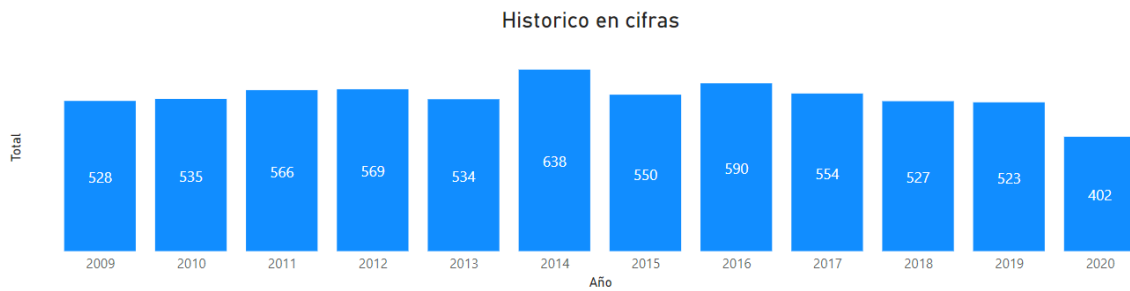
**Figura 1.** Total accidentes con lesionados por año en Bogotá



Fuente: <https://ansv.gov.co/es/observatorio/estad%C3%ADsticas/historico-victimas>

Así mismo la Figura 2 ilustra el histórico de accidentes con fatalidad en Bogotá por año. Igualmente, las cifras mantienen un valor constante o cercano para cada año, siendo el año 2020 la que levemente presenta menos víctimas fatales. Esto a pesar de que en la Figura 1 se ilustra que en el año 2020 se presentaron la mitad de los accidentes presentados en los años previos. Es decir, con una menor cantidad de accidentes viales, los accidentes con víctimas fatales bajaron muy poco. Esto es una motivante más de investigación y de la realización de este trabajo en la ciudad de Bogotá.

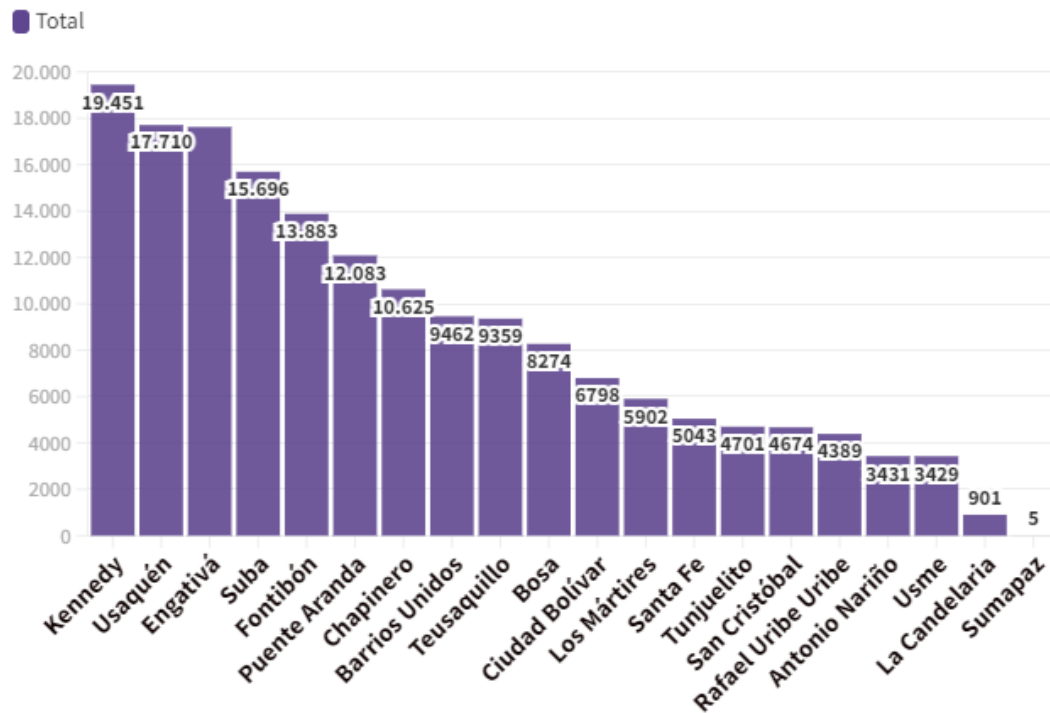
**Figura 2.** Total accidentes con víctima fatal por año en Bogotá



Fuente: <https://ansv.gov.co/es/observatorio/estad%C3%ADsticas/historico-victimas>

En el caso de estudio que es la ciudad de Bogotá, se tiene un historial de accidentes viales por localidad, que es una división geográfica de la ciudad en zonas. La localidad que mayor accidente vial presenta es Kennedy, seguido de Usaquén y Engativá.

**Figura 3.** Históricos accidentes por localidad en Bogotá entre el año 2015-2019



Fuente: <https://www.eltiempo.com/datos/cifras-de-accidentes-y-muertes-por-accidentes-de-transito-en-bogota-578915>

## 2. Marco teórico

En este capítulo se presentan las definiciones de las herramientas y métodos a utilizar para el desarrollo de la investigación. Se desarrolla acerca de *Web scraping*, el lenguaje python y su aporte en el proceso del proyecto. Se menciona acerca de la minería de datos espacial que es aplicada en el presente proyecto, así como de los sistemas de información geográficos usados para el tratamiento visual de los datos.

### 2.1 Web scraping

*Web scraping* es el proceso de extraer contenido o datos de una página web mediante métodos tecnológicos. *Web scraping* engloba una gran variedad de técnicas de programación y tecnologías, como el análisis de datos, *parseo* de lenguaje natural y la seguridad de la información [13]. Normalmente estas técnicas son aplicadas de manera periódica para una constante extracción y persistencia de los datos obtenidos para su posterior tratamiento y análisis.

Existen diferentes técnicas para la extracción de datos de páginas web, como pueden ser extracción mediante atributos html de la página para encontrar el patrón o identificar los elementos que se desean extraer en una página. El web scraping tiene bastantes usos, algunos ejemplos son:

- **Recolección de datos personales para campaña de marketing:** Se puede extraer datos de google maps o páginas amarillas para encontrar correos o números telefónicos.
- **Monitoreo de precios:** Las compañías necesitan conocer los precios que ofrecen sus competidores para lograr una mejor competitividad de mercado. Usan web scraping para extraer información de los precios de los productos ofrecidos por las demás compañías.



- **Agencias de bienes raíces:** Usan esta herramienta para la extracción de detalles y comparativa de precios para.
- **Extracción información de redes sociales:** Se usa esta herramienta para extraer información de los textos creados en redes sociales, como puede ser tweets o posts de Facebook, por ejemplo.

## 2.2 Base de datos MongoDB

MongoDB es una base de datos no relacional, basada en documentos que soporte una gran escalabilidad y flexibilidad, ofreciendo consultas e indexación avanzada. Se basa en almacenar los datos en formato similar al json. Cada objeto se representa mediante el modelo llamado documento, esto permite un trabajo simple y fácil de los datos. Esta base de datos es de uso gratuito, bajo la Licencia pública del lado del servidor (SSPL) v1. Esto permite un fácil acceso a esta herramienta a un bajo costo.

Es la quinta base de datos más usada, y la primera más usada de las bases de datos NoSQL [30]. MongoDB trabaja mucho mejor a medida que los datos crecen comparado con PostGIS, ya que a medida que aumentan los datos aumenta también la complejidad de cómputo [31].

Por estas razones se eligió usar MongoDB como base de datos para el almacenamiento y procesamiento de los datos recolectados.

## 2.3 Minería de datos espacial

Los datos extraídos para análisis cada día son más grandes, es por esto que se deben utilizar métodos para su correcto tratamiento e identificación de información relevante [14]. La minería de datos ocupa un lugar importante en el proceso del análisis de datos espaciales junto a las bases de datos espaciales. Esta se encarga de la gestión y el análisis de datos; se apoya con el uso de mapas para su mejor entendimiento y análisis [14].

La importancia de estas herramientas es el apoyo en el descubrimiento de factores o características relevantes entre los datos, pero complejas de identificar [15]. Esta se

encarga de realizar estos descubrimientos de patrones mediante la estructuración, agrupación y demás tipos de relaciones espaciales.

Se pueden agrupar los diversos métodos de la minería de datos en dos categorías que son la predictiva y descriptiva [16]. Mediante los métodos predictivos se busca poder predecir resultados mediante el conocimiento previo. Los métodos descriptivos buscan identificar el patrón o características de los datos y sus relaciones, sin la necesidad de datos previos [17].

La minería de datos generalmente realiza cuatro tareas, estas provenientes de avances en computación, estadística y visual [18]:

- **Clasificación espacial y predicción:** agrupación de datos puntuales en clases. Hacen uso de métodos de clasificación como redes neuronales, árboles de decisión, máquinas de soporte vectorial, entre otros [19].
- **Reglas de asociación espacial:** definición de las relaciones entre objetos y los predicados espaciales. Además, representan relaciones topológicas como lo son la disyunción, intersección, adyacencia, superposición, vecindad e igualdad, y la información acerca de las distancias entre los objetos.
- **Agrupación espacial, regionalización y análisis de patrones de punto:** Agrupa los datos en grupos o clusters, esto para lograr grupos por algún tipo o variable diferenciadora. Existen dos maneras de agrupar, por separación y jerárquica. El análisis de punto caliente se utiliza para lograr la agrupación de eventos inusuales en el espacio.
- **Geovisualización:** Se utiliza para lograr un mayor entendimiento de los datos geográficos. Con esto podemos comprender, comunicar y analizar de una mejor manera [20]. Se concluye que la geo visualización es una potente herramienta para la generación de conocimiento [21].

## 2.4 Sistemas de Información Geográfica

Los sistemas de información geográfica son aplicaciones de software que ofrecen todo un conjunto de herramientas para la gestión, consulta, edición, estudio y análisis de los datos espaciales. Estas herramientas de software nacieron de la necesidad de un correcto y práctico tratamiento de los datos espaciales, para de este modo tener un buen entendimiento de los datos [22].

### 2.4.1 QGIS

QGIS es un Sistema de Información Geográfica (SIG) de Código Abierto licenciado bajo GNU - General Public License . QGIS es un proyecto oficial de Open Source Geospatial Foundation (OSGeo). Corre sobre Linux, Unix, Mac OSX, Windows y Android y soporta numerosos formatos y funcionalidades de datos vector, datos ráster y bases de datos. [23].

Con esta poderosa herramienta se puede hacer exitosa la geovisualización. Dispone de un conjunto de funcionalidades como lo son seleccionar el lienzo, visualizar los datos, analizarlos, gestionarlos desde cualquier fuente, y automatizar tareas. También se podrán realizar operaciones espaciales complejas e interceptar datos geográficos.

Adicionalmente esta herramienta cuenta con la posibilidad de agregar distintos *plugins* o complementos, como es el de agregar funciones en python, así logrando automatizar herramientas de geo visualización y mejorar la aplicación de estas funciones.

### **3. Web scraping en tráfico y accidentes viales**

En el presente capítulo se presentan los procesos y las fuentes de datos a utilizar para la aplicación de *web scraping*. Se detallarán las fuentes de información, es decir las redes sociales utilizadas para la extracción de la información, así como también la definición de cada una de las variables que componen los diferentes tipos de información. Se expondrá además la técnica utilizada para la extracción de estos.

Adicionalmente para lograr una arquitectura automatizada de extracción de datos y persistencia de los datos se realizó una selección de herramientas y métodos de software para mantener una extracción continua en el tiempo.

Para la extracción de accidentalidad y alertas viales se utilizó como fuente de datos la red social Waze. Waze es una aplicación de tráfico y navegación basada en la comunidad que se creó como una herramienta de navegación social para coches particulares [24]. La aplicación ofrece el estado de las vías mediante acceso web o móvil. Expone un mapa del estado del tráfico de las vías y las diferentes alertas reportadas por los usuarios en un espacio determinado.

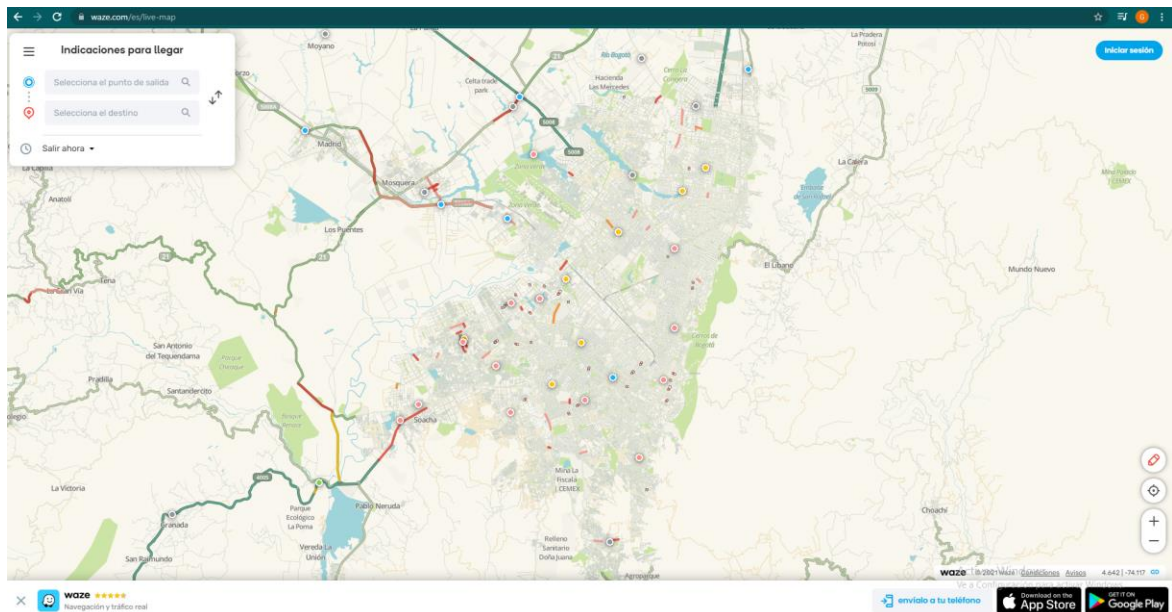
Se decidió utilizar esta importante fuente de información para el análisis aplicado en este documento. Se extraen datos de alertas reportadas por usuarios y el estado del tráfico de las vías. Ambos datos tienen en común dos datos importantes que caracterizan los datos geográficos; datos espaciales y temporales.

## 3.1 Web scraping utilizando waze como fuente de información

### 3.1.1 Petición http para extracción de datos

Al ingresar al *live map* de Waze se evidencian los segmentos de línea en el mapa. Estos segmentos representan un tráfico o *jam* en la vía o calle respectiva donde se sitúa. Estos segmentos se definen básicamente de una longitud y un color, el cual representa el nivel de tráfico en este, siendo el rojo el de mayor tráfico y el verde el de menor tráfico [25]. Para lograr extraer esta información se identifica la petición http GET que realiza el live map de Waze para presentar de manera visual y estética la información el mapa de Waze. En esta petición se encuentran los datos deseados para persistir y analizar.

**Figura 4.** Live map de Waze, presenta alertas y segmentos de tráfico.



Fuente: <https://www.waze.com/es/live-map>

La petición http GET es solicitada al servidor de Waze hacia la url <https://www.waze.com/row-rtserver/web/TGeoRSS>. Este *endPoint* retorna la información

de las alertas y los segmentos de tráfico mediante un arreglo para cada dato. Este *endPoint* requiere unas variables de petición para que sea correcta la solicitud.

**Figura 5.** Parámetros de la petición GET al endpoint de datos de alertas y tráfico de Waze.

KEY	VALUE
<input checked="" type="checkbox"/> bottom	4.678205330825949
<input checked="" type="checkbox"/> left	-74.1288414001465
<input checked="" type="checkbox"/> ma	200
<input checked="" type="checkbox"/> mj	100
<input checked="" type="checkbox"/> mu	20
<input checked="" type="checkbox"/> right	-74.00771713256837
<input checked="" type="checkbox"/> top	4.77089544019067
<input checked="" type="checkbox"/> types	alerts%2Ctrffic%2Cusers

Fuente: Elaboración propia usando Postman.

La solicitud http requiere que sea definido el espacio o zona geográfica en el cual se retorne su información. Para ello se utiliza un *bounding box* o cuadro geográfico. Un *bounding box* es una figura abstracta que se utiliza para representar y definir correctamente un objeto [34]. Normalmente se utiliza un rectángulo como figura, por lo tanto, se debe dar la información necesaria para lograr definirlo. En este caso se solicitan las coordenadas de las esquinas del rectángulo. A continuación, se definen cada uno de los parámetros necesarios en la petición GET:

- **bottom:** Coordenada de la longitud inferior del bounding box
- **left:** Coordenada de la latitud izquierda del bounding box
- **ma:** Cantidad máxima de alertas retornadas. Máximo son 200 por petición.
- **mj:** Cantidad máxima de segmentos de tráfico o jams por petición. Máximo son 200 por petición.
- **mu:** Cantidad máxima de usuarios retornados por la petición.
- **right:** Coordenada de la latitud derecha del bounding box
- **top:** Coordenada de la longitud superior del bounding box
- **types:** Tipos de eventos a obtener

### 3.1.2 Cuadrícula Bogotá usando Tiles de Google

Para lograr una mayor cantidad de detalle a la hora de realizar la petición a Waze, y además de que se limita a 200 segmentos de tráfico y 200 alertas por petición, se decidió realizar la petición con diferentes bounding box. Para esto se utilizan los tiles definidos por Google. *Google map tiler* es una representación del mundo mediante un conjunto de pequeños cuadros. Esto inicialmente creado para poder referenciar alguna zona del planeta tierra con esta definición sin necesidad de precargar todas las zonas [26]. Con estas proyecciones usadas por Google se puede trabajar sobre el mapa como si fuera un lienzo plano. Cada *tile* o cuadro es representado por coordenadas que identifican cada esquina del cuadro. Al realizar zoom en el mapa interactivo en la página web <https://www.maptiler.com/google-maps-coordinates-tile-bounds-projection/> el mapa genera los cuadros que encierra cada zona. Con esta herramienta se definieron 22 bounding box que en su totalidad delimitan a la ciudad de Bogotá. Con esto se logra recoger al mayor detalle cada eventualidad presentada.

**Figura 6.** Tile Bounds o bounding box definidos por MapTiler de Google en Bogotá.



Fuente: MapTiler de Google <https://www.maptiler.com/google-maps-coordinates-tile-bounds-projection/#3/15.00/50.00>

## 3.2 Alertas viales en Waze

Las alertas son hechos reportados por los usuarios. Cada alerta tiene su calificación para identificar la fiabilidad de este. Al realizar la petición GET obtenemos un arreglo de objetos llamado “alerts”, el cual cada objeto de este arreglo contiene los siguientes datos.

Figura 7. Datos en el objeto que representa una alerta de Waze.

```
1 {
2   "country": "CO",
3   "nThumbsUp": 8,
4   "city": "Bogota",
5   "reportRating": 0,
6   "reliability": 10,
7   "type": "ROAD_CLOSED",
8   "uuid": "4d5838b2-c9ea-451d-b386-f01f60cd0ad1",
9   "speed": 0,
10  "reportMood": 0,
11  "subtype": "ROAD_CLOSED_EVENT",
12  "street": "Calle 193",
13  "additionalInfo": "",
14  "id": "alert-591673411/4d5838b2-c9ea-451d-b386-f01f60cd0ad1",
15  "nComments": 1,
16  "reportBy": "andresparra78",
17  "inscale": false,
18  "comments": [...],
19  "isJamUnifiedAlert": false,
20  "confidence": 4,
21  "nImages": 0,
22  "roadType": 2,
23  "magvar": 0,
24  "showFacebookPic": false,
25  "wazeData": "world,-74.03452068151813,4.770486671748552,4d5838b2-c9ea-451d-b386-f01f60cd0ad1",
26  "reportDescription": "Obras en la vía",
27  "location": {
28    "x": -74.03452,
29    "y": 4.770486
30  },
31  "pubMillis": 1617323431000
32 }
```

Fuente: Elaboración propia usando Postman.

- **country:** Indica el país del cual se retorna los datos según el bounding box definido.
- **nThumbsUp:** cantidad de “me gusta” de la alerta.
- **city:** Ciudad donde se registró la alerta.
- **reportRating:** Calificación de la alerta
- **reliability:** Confiabilidad de la alerta
- **type:** Tipo de la alerta: accidente, peligro, construcción o vía cerrada.
- **uuid:** Identificador único de la alerta.
- **speed:** Velocidad con que ocurrió la alerta.
- **subtype:** Subcategoría del tipo de alerta.



- **street:** Nombre de la calle donde sucedió la alerta.
- **additionalInfo:** Información adicional de la alerta.
- **id:** Id que resulta al concatenar un identificador de alerta con su uuid.
- **nComments:** Número de comentarios realizados acerca de la alerta.
- **reportBy:** Persona que realiza el reporte o registro de la alerta.
- **inscale:** Indica si la alerta es en escala.
- **comments:** Comentarios realizados acerca de la alerta.
- **isJamUnifiedAlert:** Indica si se detecta una desaceleración.
- **confidence:** Calificación de la alerta basada en la reputación del reportador.
- **nImages:** Cantidad de imágenes relacionadas a la alerta.
- **roadType:** Tipo de vía en el que sucedió la alerta.

### 3.3 Segmentos de tráfico en Waze

Un segmento de tráfico representa una parte de una vía vehicular y el estado de congestión en ella. Cada segmento de tráfico es representado por un conjunto de variables las cuales lo definen completamente:

**Figura 8.** Datos en el objeto que representa un segmento de tráfico de Waze.

```
{
  "severity": 5,
  "country": "CO",
  "level": 3,
  "line": [ ...
],
  "speedKMH": 11.3,
  "length": 597,
  "turnType": "NONE",
  "type": "NONE",
  "uuid": 600722382,
  "endNode": "Bogotá, D.C.",
  "speed": 3.138888888888889,
  "segments": [ ...
],
  "roadType": 2,
  "delay": 106,
  "updateMillis": 1622476944501,
  "startNode": "Bogotá, D.C.",
  "street": "Calle 53 > Oriente",
  "id": 600722382,
  "pubMillis": 1622475789351
},
```

Fuente: Elaboración propia usando Postman.

- **severity:** Número que indica el grado de severidad del tráfico.
- **country:** indica el país al cual pertenece el segmento de tráfico según el bounding box definido.
- **level:** Nivel que indica la congestión del tráfico, siendo 0 libre flujo y 5 bloqueado.
- **line:** Lista de coordenadas que definen la línea que representa el segmento de tráfico en el mapa.
- **speedKMH:** Velocidad media en el segmento de tráfico vial representada en kilómetros por hora.
- **length:** Longitud del segmento en metros
- **type:** Tipo de evento
- **uuid:** Identificador único del segmento de tráfico
- **endNode:** Ciudad donde finaliza el segmento de tráfico.
- **speed:** Promedio de la velocidad en el segmento de tráfico en el momento de ocurrencia medido en metros por segundo.
- **segments:** Lista de objetos que definen un conjunto de segmentos, definidos por un nodo inicial y un nodo final cada uno.
- **roadType:** Tipo de vía.
- **delay:** Demora o tiempo de retraso representado en segundos comparado a un flujo promedio del segmento.
- **updateMillis:** Última fecha de actualización del estado del segmento.
- **startNode:** Ciudad o país donde inicia el segmento de tráfico.
- **id:** Identificador del segmento de tráfico.
- **pubMillis:** Fecha del momento de ocurrencia del segmento de tráfico.

### 3.4 Web scraping en red social Twitter

Adicionalmente a los datos de Waze se persisten datos de Twitter. Se toman datos de la cuenta de @BogotaTransito que realiza tweets acerca de información del estado de las vías en la ciudad, como tráfico y accidentes viales. Los datos se recogen con la librería de python TwitterAPI. Este es una librería de python que permite acceder a la información

alojada en twitter mediante peticiones *REST API* y *Streaming API* [27]. Esta librería solicita las *keys* y *tokens* que se generan en Twitter para poder extraer la información [28]. Para persistir solo los tweets necesarios se realiza un filtro en la petición para identificar los tweets que contengan la palabra “siniestro”, la cual es siempre utilizada por la cuenta de Twitter en cada tweet relacionado a un accidente de tránsito.

Al realizar la petición a la cuenta de Twitter mencionada se persisten los siguientes datos de cada tweet:

- **id:** Identificador único del tweet.
- **created\_at:** Fecha en la que se creó el tweet.
- **text:** Texto del tweet.
- **userId:** Identificador del usuario que realizó el tweet. En este caso es el id de la cuenta de Twitter @BogotaTransito.

**Figura 9.** Datos en el objeto extraído del api de Twitter que representa un tweet.

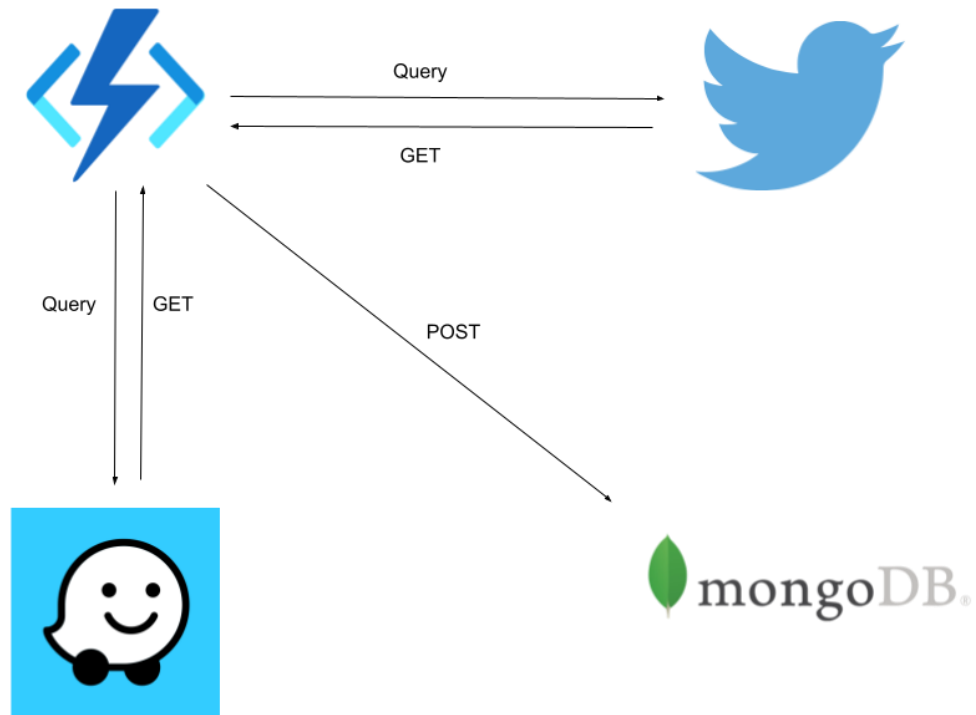
```
id: "1424490709734862850"  
created_at: "2021-08-08T22:00:25.000Z"  
text: "#GerenciaEnVía informa |Siniestro solo daños en la localidad de Kenned..."  
userId: "2470183729"
```

Fuente: Elaboración propia usando MongoDB.

### 3.5 Arquitectura de recolección de datos

Para lograr el objetivo del trabajo se decidió realizar un diseño para realizar una recolección de datos continua en el tiempo, esto con el fin de mantener un análisis con datos por cada momento del día. Se necesita un *job* o tarea que realice las peticiones cada cierto tiempo al *live map de Waze* y Twitter para persistir esos datos en una base de datos.

**Figura 10.** Datos en el objeto extraído del api de Twitter que representa un tweet.



Fuente: Elaboración propia.

Para las peticiones continuas a las redes sociales se hace uso de las *azure functions*. Las *azure functions* son un servicio de cloud que funciona bajo demanda que provee una actualización continua de la infraestructura y los recursos necesarios para lograr la ejecución de la aplicación usada [29]. Este servicio de nube funciona bajo la idea de *serverless*, es decir, la asignación de recursos la realiza el proveedor de nube según demanda.

Se crean dos *azure functions*, una para la extracción de datos de Waze y otra para la extracción de datos de Twitter. Ya que los datos más abundantes son los retornados por el api de Waze y no persisten, por el contrario, el tweet realizado se mantiene en un histórico, se realiza cada 10 minutos una petición para persistir los datos de alertas y segmentos de tráfico del live map de Waze. Estas ejecuciones se aplican para cada bounding box o *tile* de Google definido anteriormente. Para la extracción en Twitter se realiza una petición cada 6 horas.

En cada *azure function* se publica el código en python que realiza la tarea de extracción de los datos en cada API y se configura para que la función sea ejecutada cada cierto tiempo. Esta función configurada realiza en cada ejecución la petición de los datos al API de la red social respectiva y luego envía los datos a la base de datos de MongoDB para persistencia de los datos extraídos.

## 4. Análisis Tráfico vial

En este capítulo se analizarán los conjuntos de datos extraídos de las redes sociales, esto con el fin de ofrecer un análisis geoestadístico de estos. Se identificarán las variables dependientes e independientes, para así poder realizar las respectivas gráficas y analizar correctamente la gran cantidad de datos recolectados.

La congestión o tráfico vial impacta bastante en la economía y bienestar de la población. Este tráfico se evidencia en demoras en la logística o planeación de la economía, y por ende afecta a un costo o en un menor rendimiento en la generación de economía. Así mismo, el movimiento libre de las personas se ve directamente afectado por la congestión vial, ya que, en el planeamiento de su día, o en las relaciones sociales o laborales se ve obligado a realizar una estimación mayor de tiempo que debe dedicar al solo hecho de la movilidad. La mayor preocupación de las personas al realizar una reunión o actividad fuera de sus casas es el tráfico vial. Esto se traduce en un mal estilo de vida.

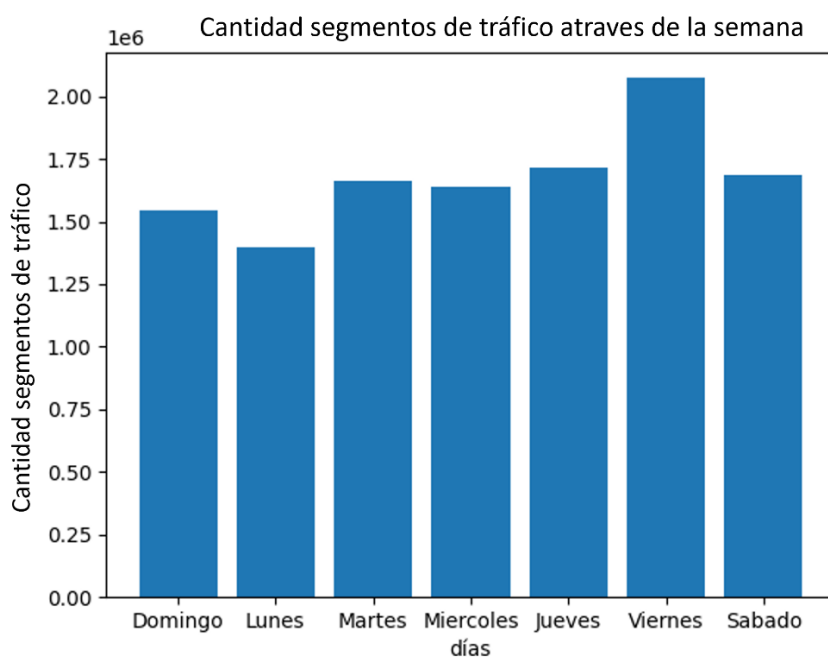
Todas estas y más problemáticas generadas por la congestión vehicular es la motivación de la realización de este trabajo. Por ello se realizará un análisis con objetivo en un buen entendimiento de los datos viales y a la identificación de los principales problemas que aportan a una mayor congestión vial, con ello poder identificar los patrones en el tráfico vial de la ciudad de Bogotá.

Los datos para analizar comprenden desde el 01 de septiembre hasta el 31 de diciembre de 2021.

## 4.1 Cantidad de segmentos de tráfico durante días la semana

Inicialmente se realizará el análisis de la cantidad de segmentos de tráfico durante los días de la semana, para evidenciar los días que mayor tráfico vial presentan.

**Figura 11.** Cantidad de segmentos de tráfico a través de la semana



Fuente: Elaboración propia.

En general el día con más tráfico vial de la semana es el viernes, seguido del jueves y el martes. Los días con menor tráfico son el lunes y el Domingo.

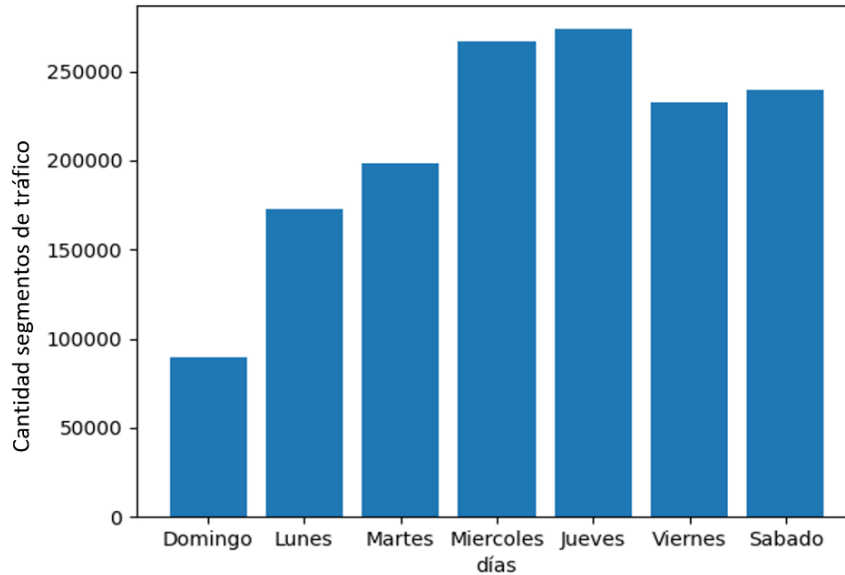
### 4.1.1 Tráfico durante días de la semana por mes con nivel 3 y 4

A continuación, se presentan las gráficas por mes dando un análisis solo con los flujos de tráfico con nivel 3 y 4, que son los niveles que representan una gran congestión vial, para

así identificar los días de la semana que presentan mayor congestión vial. Se analizaron estos tráficos de nivel 3 y 4 por cada mes.

**Figura 12.** Cantidad de segmentos de tráfico de nivel 3-4 a través de la semana del mes de septiembre

Cantidad segmentos de tráfico de nivel 3-4 a través de la semana del mes de septiembre



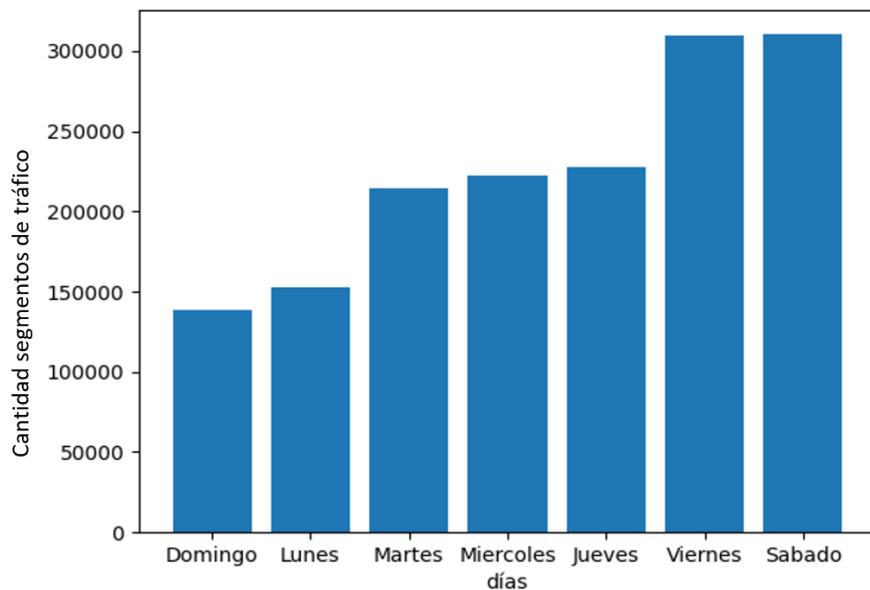
Fuente: Elaboración propia.

En el mes de septiembre se evidencia que los miércoles y jueves fueron los días con mayor congestión vial, y los días Domingo y lunes los de menor congestión.



**Figura 13.** Cantidad de segmentos de tráfico de nivel 3-4 a través de la semana del mes de octubre

Cantidad segmentos de tráfico de nivel 3-4 a través de la semana del mes de octubre

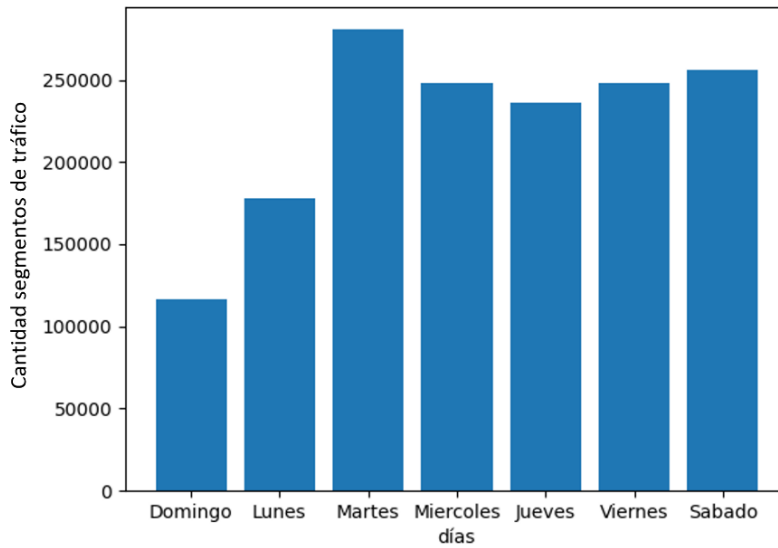


Fuente: Elaboración propia.

Para el mes de octubre el viernes y el sábado fueron los días con mayor congestión vial con una gran diferencia frente a los demás de al menos 80,000 segmentos de tráfico, y los días domingo y lunes los que menor congestión vial presentan.

**Figura 14.** Cantidad de segmentos de tráfico de nivel 3-4 a través de la semana del mes de noviembre

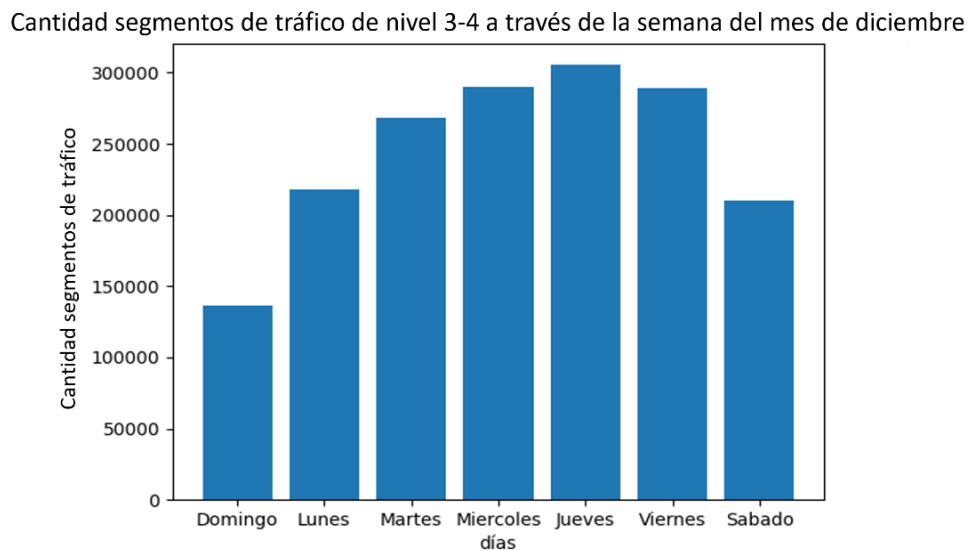
Cantidad segmentos de tráfico de nivel 3-4 a través de la semana del mes de noviembre



Fuente: Elaboración propia.

El mes de noviembre presenta que el día más congestionado fue el martes, y el domingo como el día con menor congestión vial.

**Figura 15.** Cantidad de segmentos de tráfico de nivel 3-4 a través de la semana del mes de diciembre



Fuente: Elaboración propia.

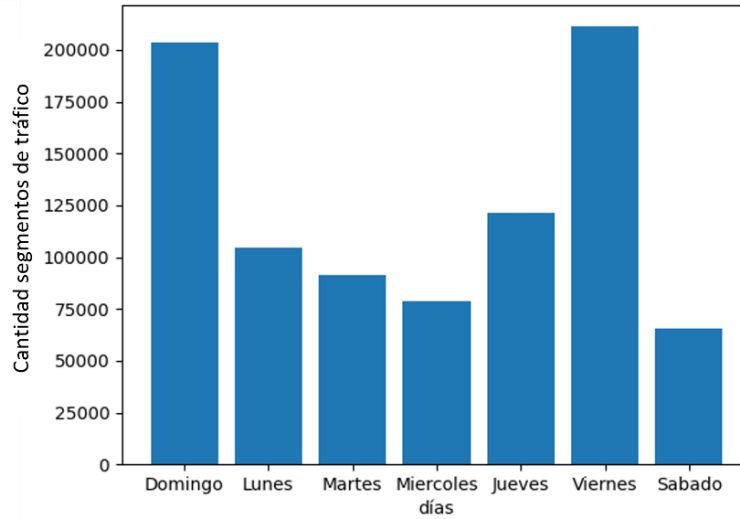
Finalmente, para el mes de Diciembre el Jueves fue el día con mayor congestión vial, y el Domingo el día con menor congestión vial. Durante cada mes fue variando el día con mayor tráfico, no siempre fue el mismo. Pero el día que menor tráfico vial tuvo siempre fue el mismo, el Domingo, que definitivamente es el día de la semana que menor tráfico vial presenta.

#### 4.1.2 Tráfico durante días de la semana por mes con nivel 5

El tráfico con nivel 5 representa bloqueos en la vía, esto debido a diferentes razones, siendo mayormente cierre por obra en la vía. Con estas gráficas podemos identificar qué día de la semana por mes tuvo mayor cierre por obras en la vía y en qué día de la semana.

**Figura 16.** Cantidad de segmentos de tráfico de nivel 5 a través de la semana del mes de septiembre

Cantidad segmentos de tráfico de nivel 5 a través de la semana del mes de septiembre

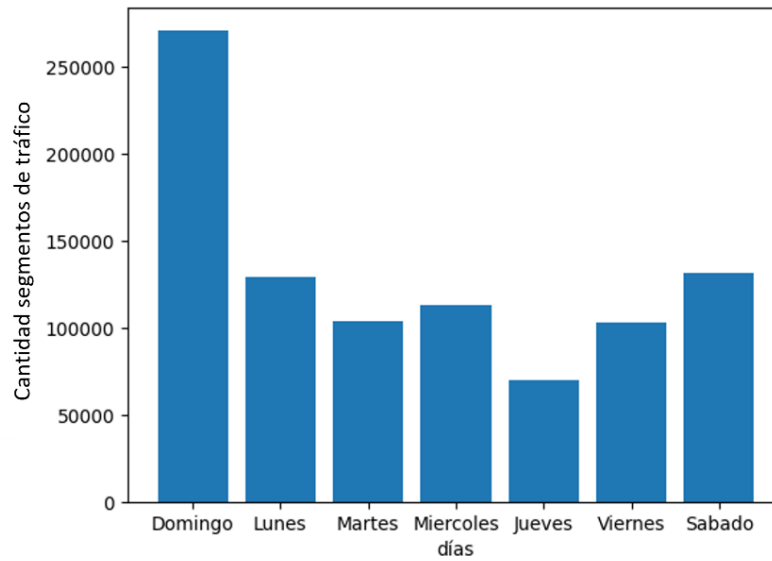


Fuente: Elaboración propia.

Para el mes de septiembre se identifica que los días con mayor cierre vial fueron los viernes y domingos.

**Figura 17.** Cantidad de segmentos de tráfico de nivel 5 a través de la semana del mes de octubre

Cantidad segmentos de tráfico de nivel 5 a través de la semana del mes de octubre

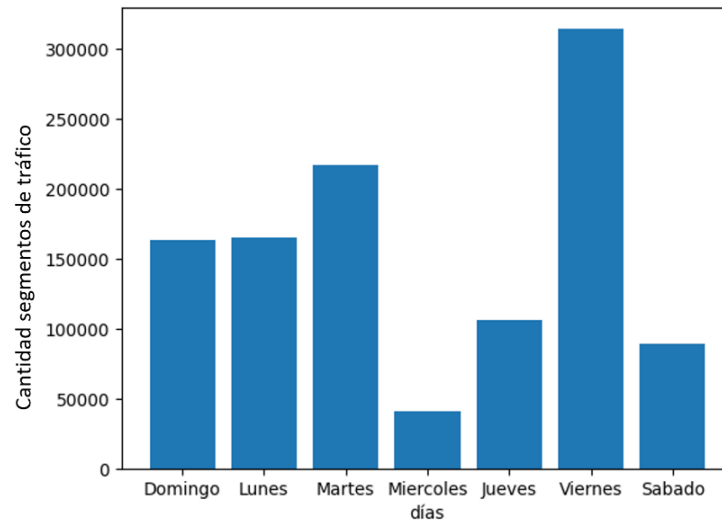


Fuente: Elaboración propia.

El domingo fue el único día que presenta mayor cierre vial debido a las obras en la vía para el mes de octubre.

**Figura 18.** Cantidad de segmentos de tráfico de nivel 5 a través de la semana del mes de noviembre

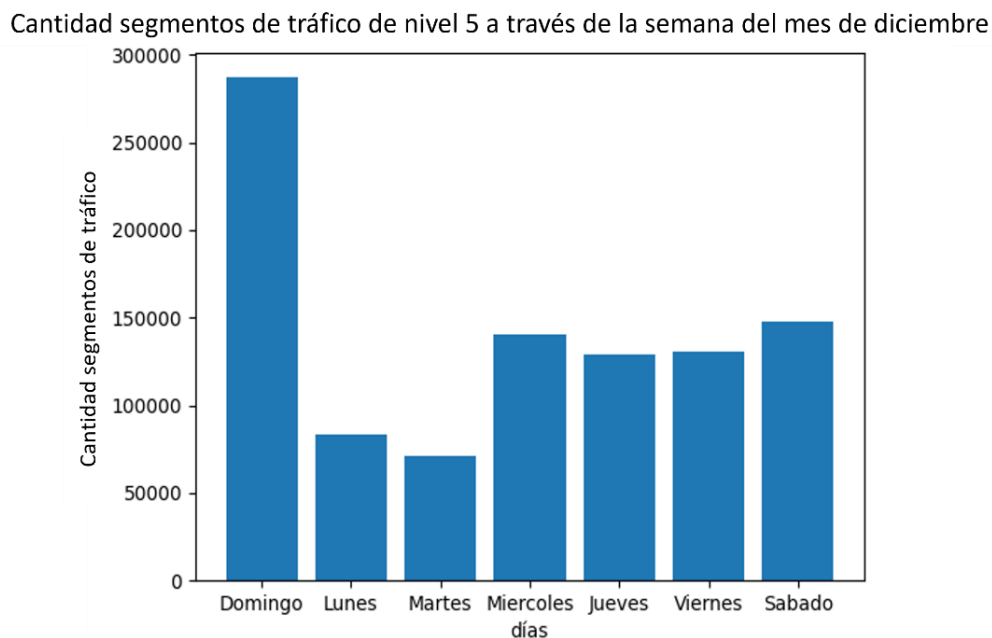
Cantidad segmentos de tráfico de nivel 5 a través de la semana del mes de noviembre



Fuente: Elaboración propia.

El mes de noviembre nos presenta un día con una gran cantidad de tráfico, siendo mayor a 300,000 segmentos de tráfico nivel 5. Seguidamente el martes también presenta gran cantidad de cierres en la vía.

**Figura 19.** Cantidad de segmentos de tráfico de nivel 5 a través de la semana del mes de diciembre

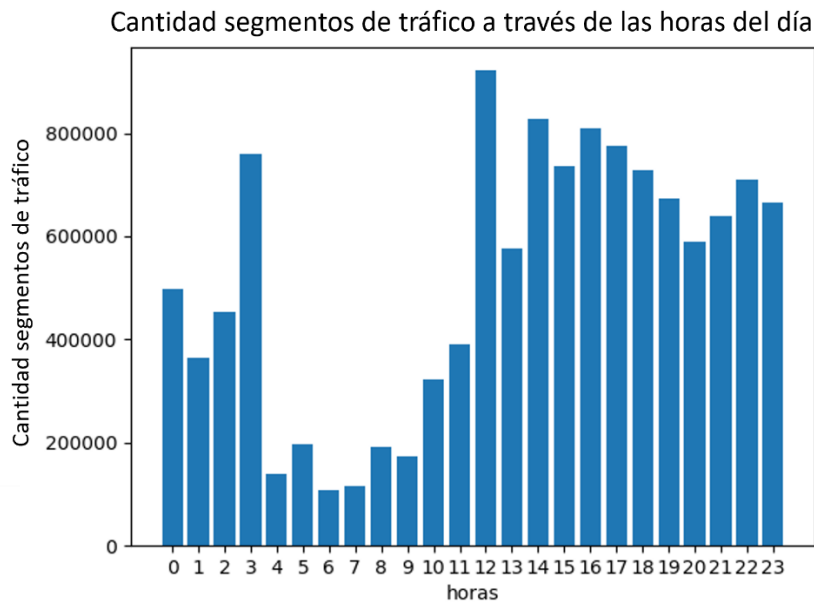


Fuente: Elaboración propia.

Finalmente, el mes de diciembre presenta el Domingo como el día con mayor cierre vial. Todos los meses menos noviembre presentaron el Domingo como el día con mayores cierres viales, por el contrario, noviembre presentó el viernes como el día con mayores cierres viales.

## 4.2 Tráfico a través de las horas del día

En la siguiente gráfica se visualiza la cantidad de segmentos de tráfico a través de las horas de manera general para todos los meses de estudio. Es decir, es la congestión o tráfico reportado en general en la ciudad de Bogotá en los cuatro meses de estudio.

**Figura 20.** Cantidad de segmentos de tráfico a través de las horas del día

Fuente: Elaboración propia.

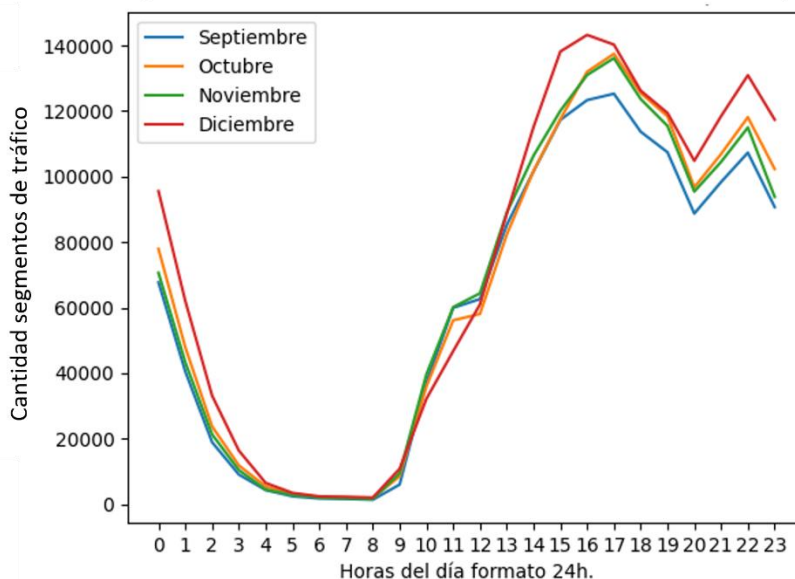
Se evidencia que la hora con mayor tráfico son a las 12, siendo alto el tráfico en las horas de la tarde, y bajo en las horas de la mañana, con excepción de las 3 am que presenta un alto tráfico vial. Para identificar el comportamiento por mes se realizará la gráfica para cada mes.

#### 4.2.1 Tráfico por mes durante las horas del día con nivel 3 y 4



**Figura 21.** Cantidad de segmentos de tráfico de nivel 3-4 a través de las horas del día por mes

Cantidad segmentos de tráfico de nivel 3-4 a través de las horas por mes

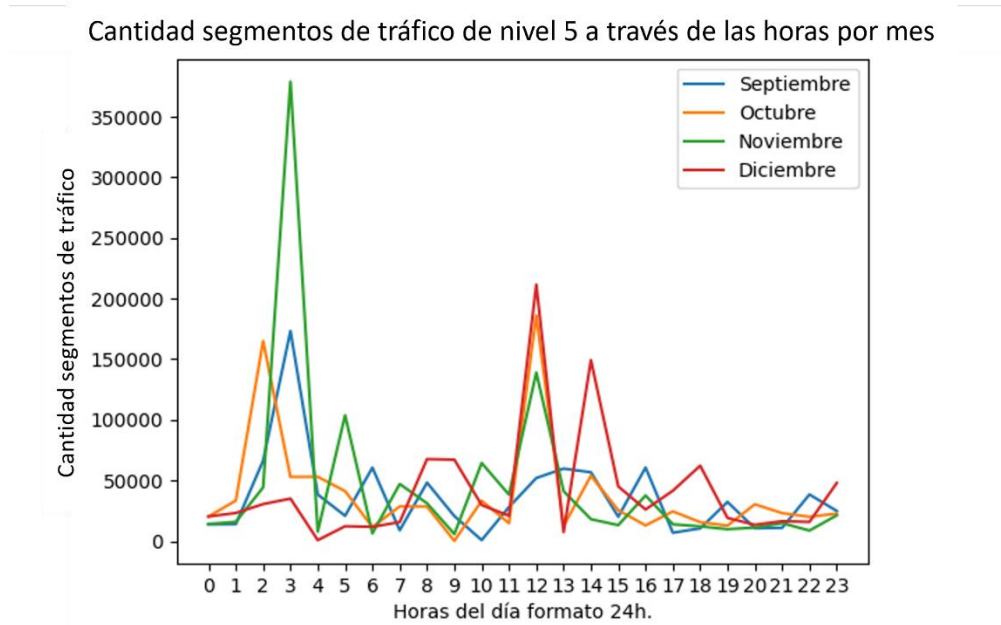


Fuente: Elaboración propia.

El tráfico de nivel 3 y 4 representan los segmentos de vía con gran congestión vial. En la gráfica se identifica que el rango de horas con mayor tráfico es entre las 15 horas y las 19 horas, teniendo otra subida nuevamente a las 22 horas. El mes que mayor congestión vial tuvo fue diciembre, y el de menor tráfico en estos cuatro meses fue septiembre. Todos los meses mantienen el mismo patrón de congestión vial en las mismas horas.

#### 4.2.2 Tráfico por mes durante las horas del día con nivel 5

**Figura 22.** Cantidad de segmentos de tráfico de nivel 5 a través de las horas del día por mes



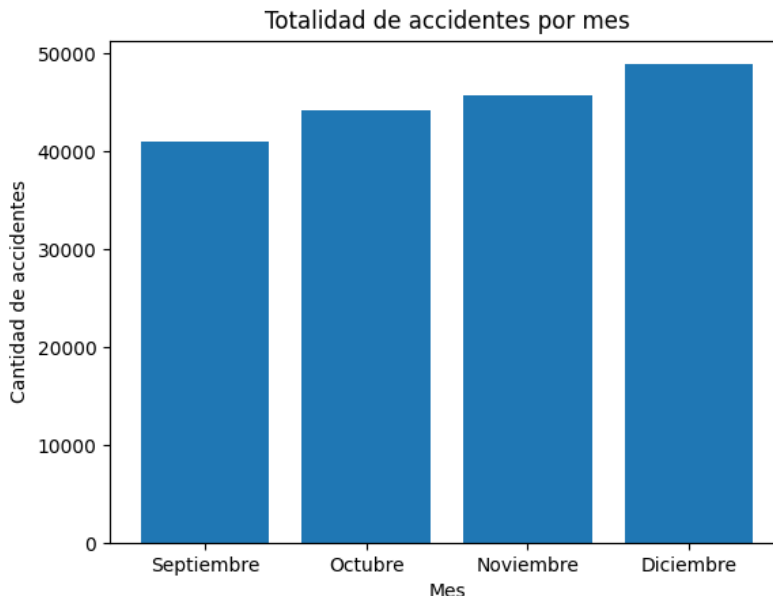
Fuente: Elaboración propia.

Este gráfico nos indica que los cierres viales en los diferentes meses si se realizaron a diferentes horas, a excepción de dos picos que se mantienen en todos los meses que son en las 12 horas y las 3 am. A las 12 horas el mes de septiembre no presentó el mismo nivel de cierre vial que los demás meses. En el caso de las 3 am se presenta que el mes de diciembre realizó un menor cierre comparado con los demás meses, pero en noviembre en ese horario se realizaron bastantes cierres comparados con los demás meses a esa misma hora.

## 5. Análisis accidentalidad vial

Los datos utilizados para el análisis de accidentalidad vial son las alertas de Waze de tipo ACCIDENT, descritos en el capítulo 3. Estos datos son extraídos del live map de Waze. En la **Figura 23** se presentan de manera general la totalidad de la accidentalidad vial en la ciudad de Bogotá entre los meses de septiembre y diciembre del año 2021.

**Figura 23.** Totalidad de accidentes por mes



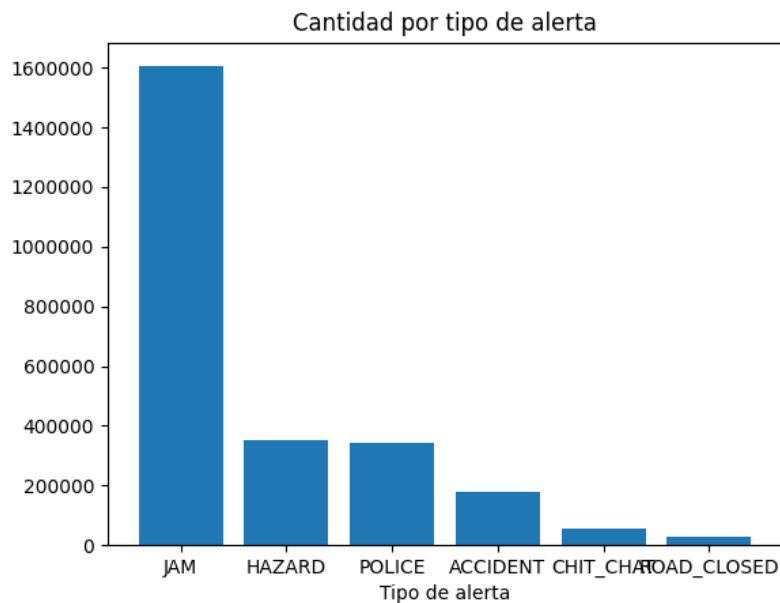
Fuente: Elaboración propia.

La gráfica nos presenta que desde el mes de septiembre se presenta un ligero aumento en la totalidad de los accidentes. El mes de diciembre fue el mes con mayor cantidad de accidentes reportados con un total de 48,901, noviembre con 45,684, octubre con 44,158 y septiembre el mes con menor cantidad de accidentes con 41,033. Analizando esta gráfica no es posible poder identificar de manera precisa los patrones que generan la accidentalidad vial. Por ello es importante el análisis espacio-temporal.

## 5.1 Exploración de los datos de Waze Alert

### 5.1.1 Tipos de alerta

Figura 24. Total de alertas por tipo



Fuente: Elaboración propia.

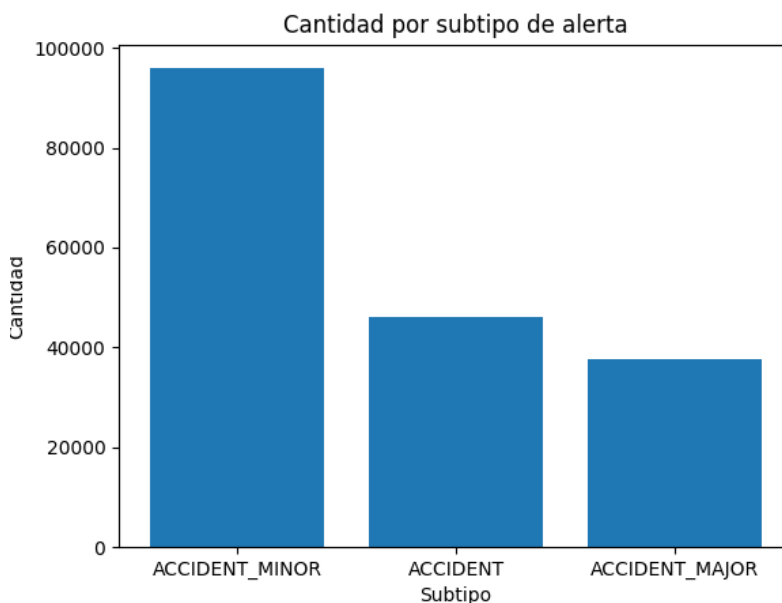
Cada tipo de alerta se define:

- **JAM:** Segmento de tráfico.
- **ROAD\_CLOSED:** Vía cerrada.
- **POLICE:** Policía en la vía.
- **HAZARD:** Peligro presentando en la vía.
- **ACCIDENT:** Accidente.
- **CHIT\_CHAT:** Mensaje acerca de un lugar o evento.

Para este caso el mayor tipo de alertas realizados en la plataforma Waze son los de tipo JAM con 1,604,336, seguidos por topografía HAZARD con 354,045. La totalidad de alertas de Waze recolectadas entre la fecha 1 de septiembre de 2021 a 31 de diciembre de 2021 fue

de 2,565,047. Para el caso de estudio solo nos enfocaremos en los tipos de accidentes ACCIDENT, debido que son los que se categorizaron y representan eventos de accidentes viales.

**Figura 25.** Total de accidentes por subtipo



Fuente: Elaboración propia.

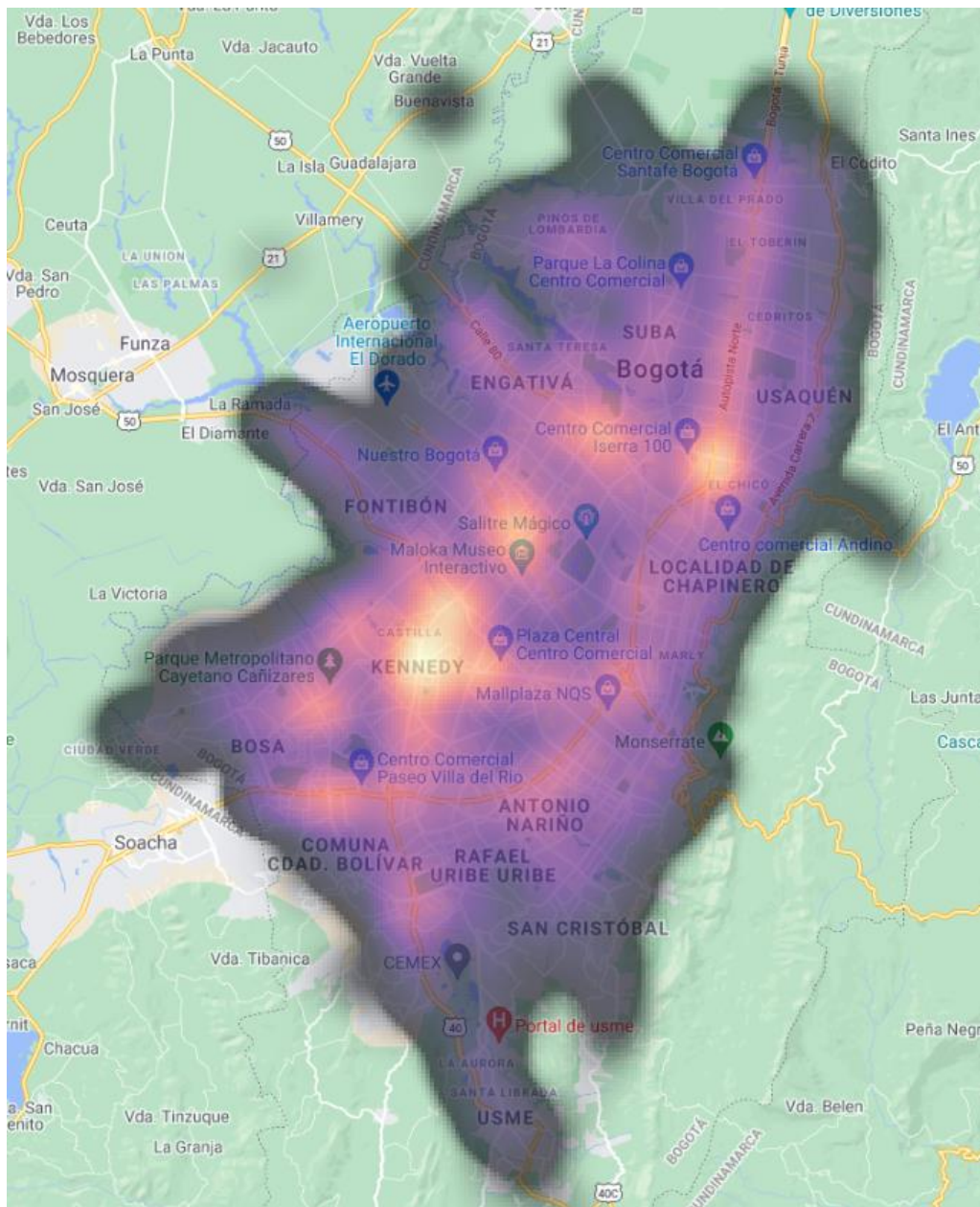
La totalidad de los datos de tipo accidente son de 179,776, siendo mayormente del subtipo ACCIDENT\_MINOR con 95,897 alertas, ACCIDENT con 46,206 y el menor ACCIDENT\_MAJOR con 37,673 alertas.

Se realizan *heat maps* para poder realizar un análisis de cluster geográfico y así representar mejor la densidad de accidentalidad por zona y analizar los sectores con mayor cantidad de accidentes viales

## 5.2 Heat maps

### 5.2.1 Heat maps con la totalidad de los datos

Figura 26. Heat map accidentes viales

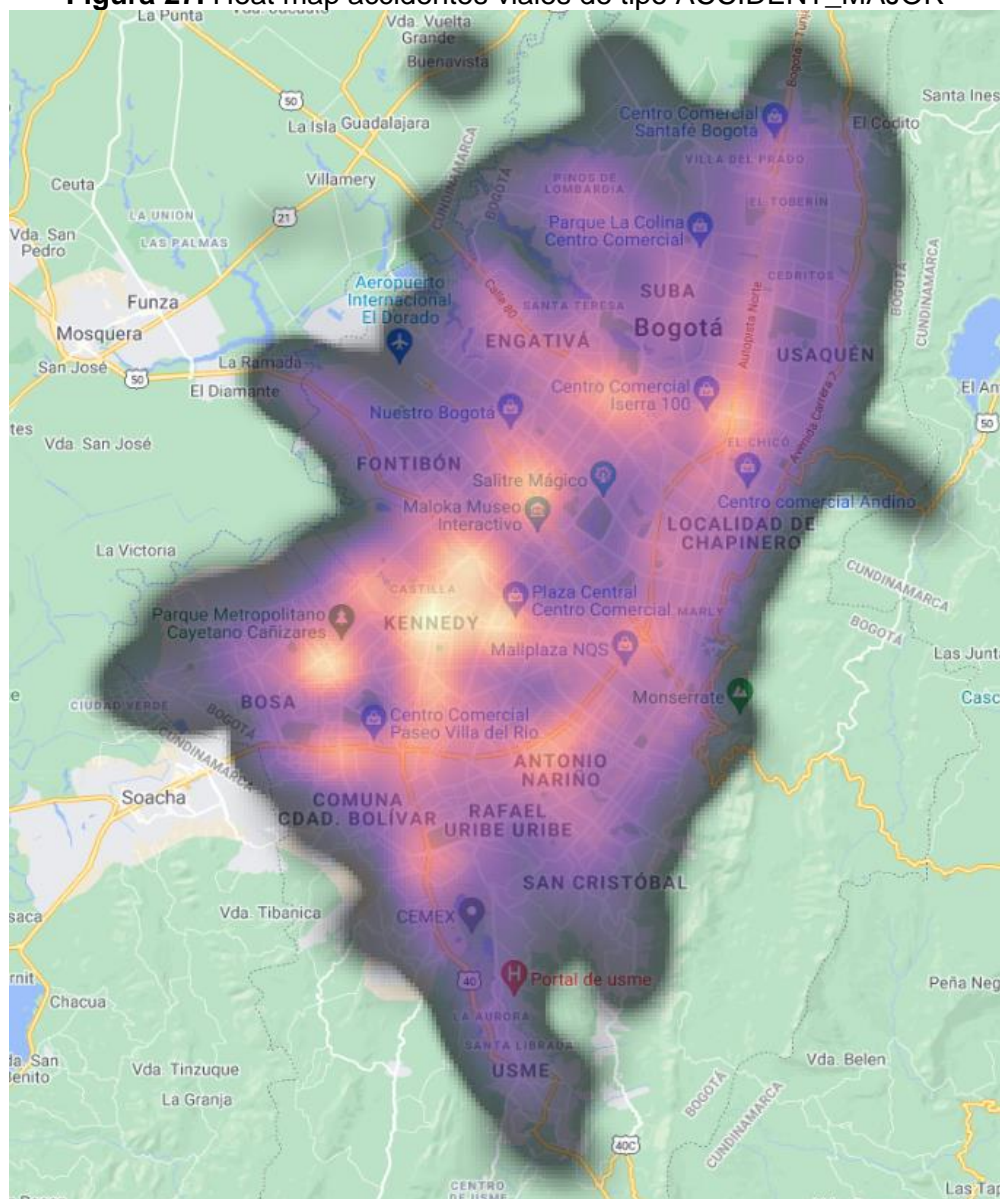


Fuente: Elaboración propia usando QGIS.

En el mapa de calor podremos identificar que la zona con mayor número de accidentes es en la localidad de Kennedy, y en la calle 100 con autopista norte. Este mapa de calor es de forma general de los tres tipos de accidentes. Seguidamente se presenta un mapa de calor por cada tipo de accidente.

## 5.2.2 Heat maps con la totalidad de los datos tipo ACCIDENT\_MAJOR

Figura 27. Heat map accidentes viales de tipo ACCIDENT\_MAJOR

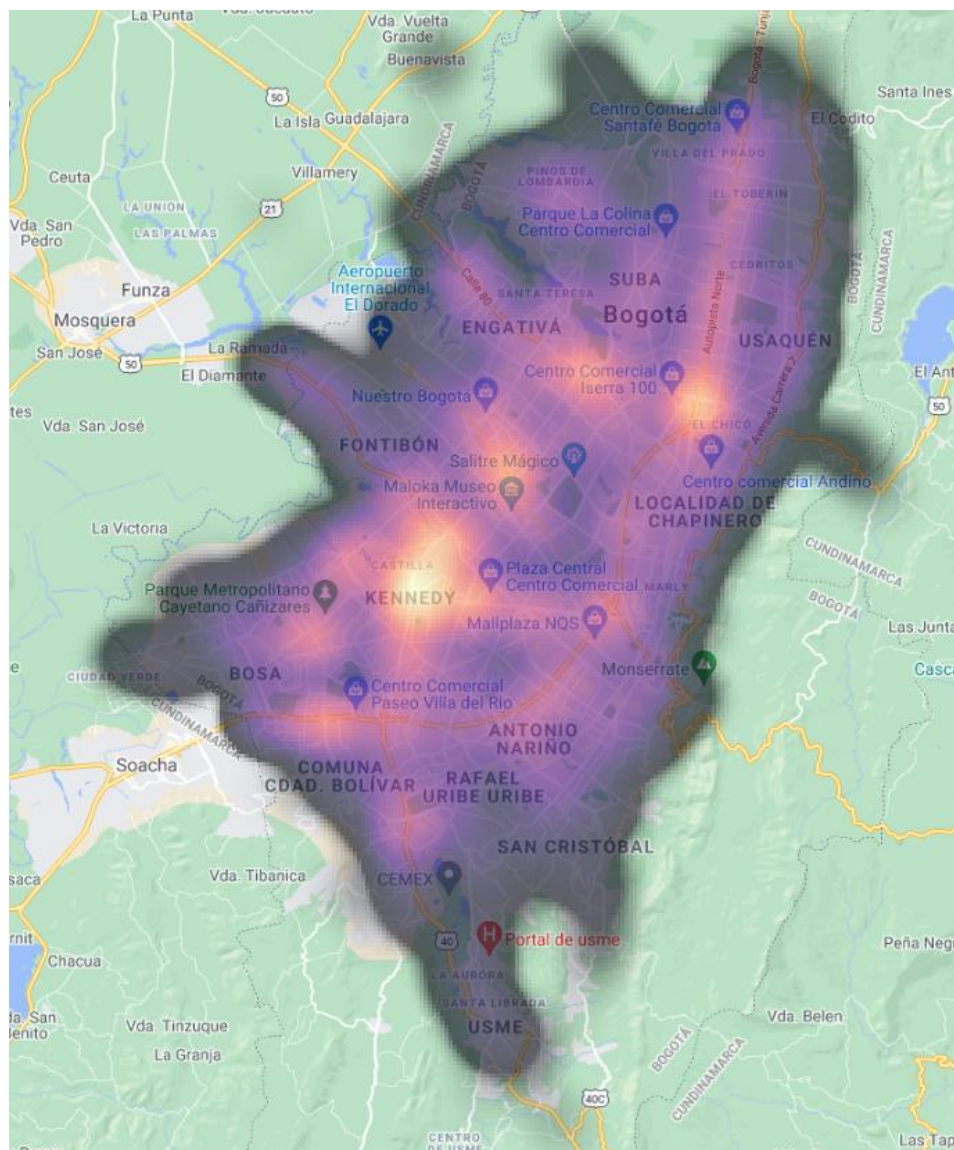


Fuente: Elaboración propia usando QGIS.

Los tipos de accidentes ACCIDENT\_MAJOR representan accidentes que en su mayoría presentan o un lesionado grave o un accidente con víctima fatal. Teniendo esto en cuenta, se observa que Kennedy tiene la mayor concentración de accidentes graves, seguida de la zona del barrio Nueva Britalia. También son puntos graves la AV. Boyacá con Calle 26, y la autopista norte con calle 100.

### 5.2.3 Heat maps con la totalidad de los datos tipo ACCIDENT\_MINOR

Figura 28. Heat map accidentes viales de tipo ACCIDENT\_MINOR



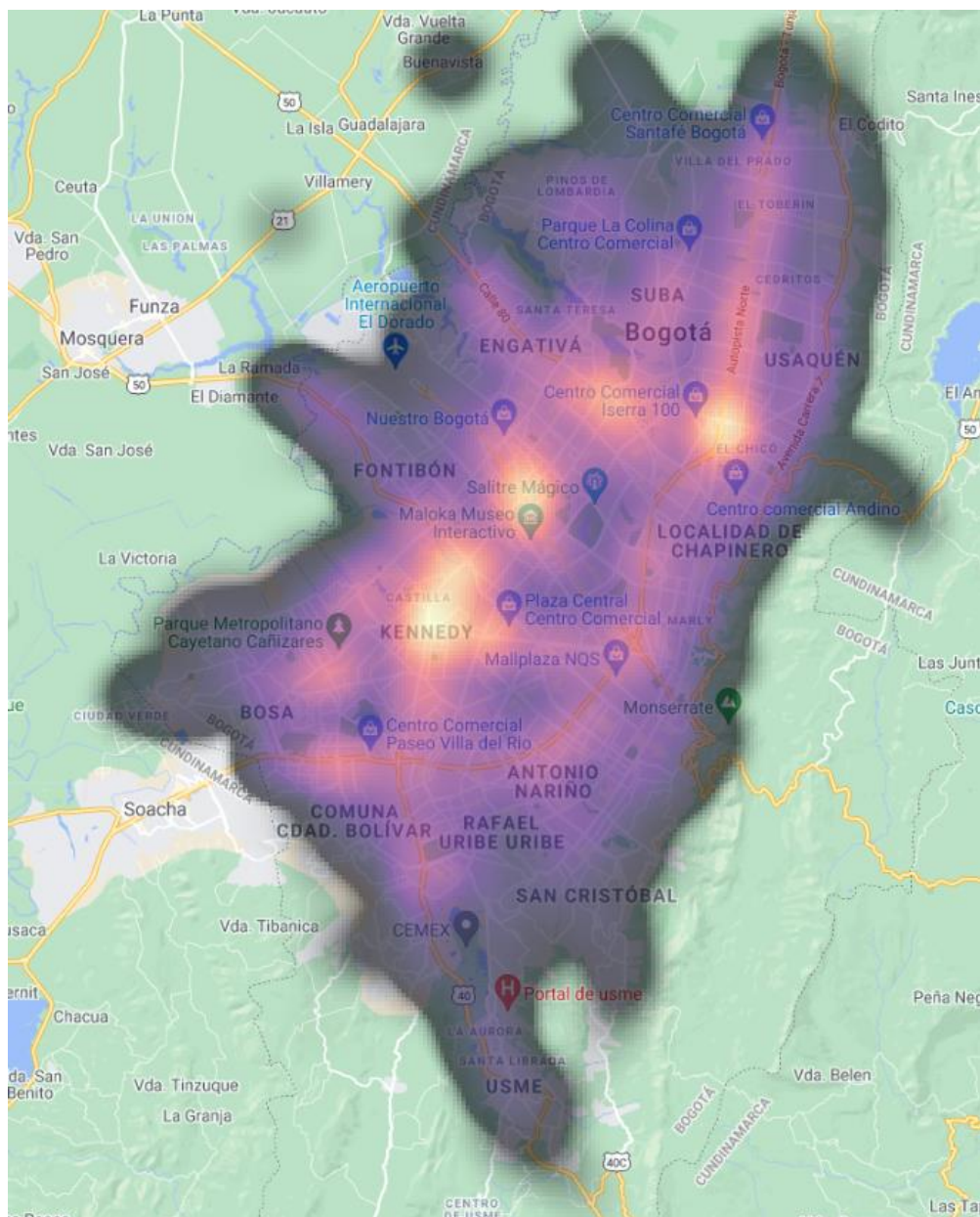
Fuente: Elaboración propia usando QGIS.



Para el caso de los accidentes tipo ACCIDENT\_MINOR persisten las mismas zonas calientes con excepción del barrio Nueva Britalia. Kennedy continúa siendo el lugar con mayor cantidad de accidentes.

## 5.2.4 Heat maps con la totalidad de los datos tipo ACCIDENT

Figura 29. Heat map accidentes viales de tipo ACCIDENT



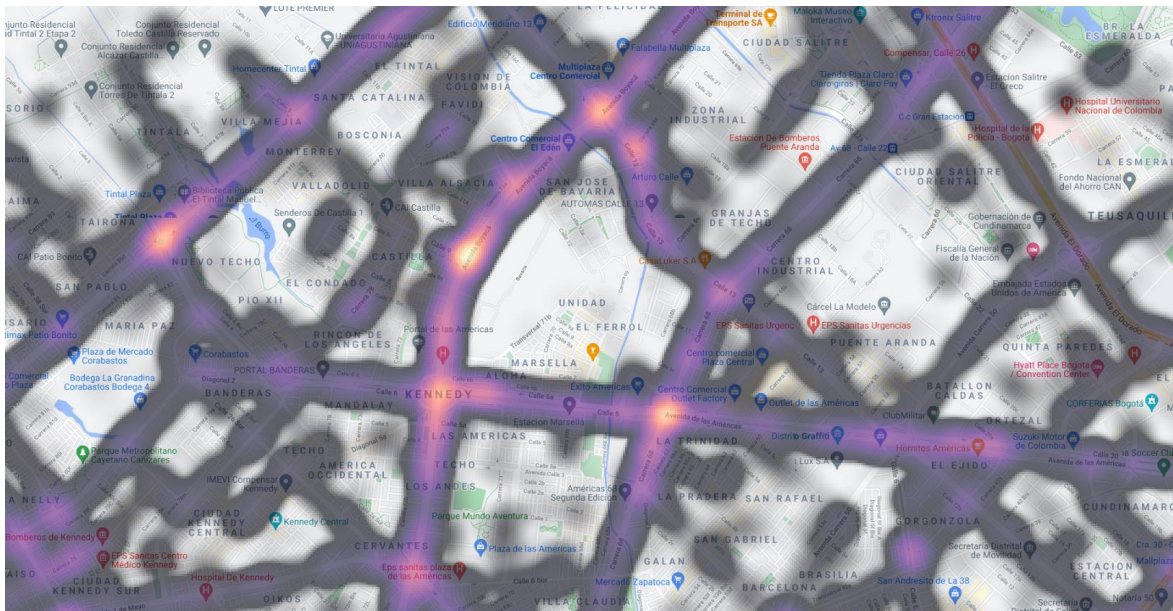
Fuente: Elaboración propia usando QGIS.

Las zonas persisten para el caso de los accidentes tipo ACCIDENT, solo que la autopista con calle 100 resalta más teniendo mayor número de accidentes. Kennedy continúa siendo un gran punto de accidentes. Se identifica que el barrio Nueva Britalia solo fue punto caliente para el caso de los accidentes tipo MAJOR, es decir, los accidentes que ocurren allí son mayormente con lesionado grave o con víctima fatal.

### 5.3 Puntos de concentración de accidentes

Se identificaron varias zonas de concentración de accidentes viales de todos los subtipos; en Kennedy, alrededores del Salitre Mágico, Autopista Norte con calle 100 y Av. Boyacá con calle 80. Estos datos concuerdan con la gráfica de mayores accidentes por localidad, siendo Kennedy la localidad que mayor número de accidentes viales presenta. Se presentará un análisis de cada zona para identificar los puntos exactos de puntos calientes.

**Figura 30.** Heat map accidentes viales en Kennedy



Fuente: Elaboración propia usando QGIS.

Se identifican los puntos calientes en la localidad de Kennedy, siendo mayormente en cruces entre avenidas:

- Av. Boyacá con 1 de Mayo
- Av. de las Américas Con Av. Boyacá
- Av. de las Américas Con Carrera 68
- Av. de las Américas Con Carrera 86
- Av. Boyacá con Calle 13
- Av. Boyacá con altura Calle 9

Cinco de estos seis puntos calientes se sitúan en intersecciones con otras vías principales, es decir, la mayoría de los accidentes ocurren en intersecciones.

**Figura 31.** Heat map accidentes viales en Salitre Mágico



Fuente: Elaboración propia usando QGIS.

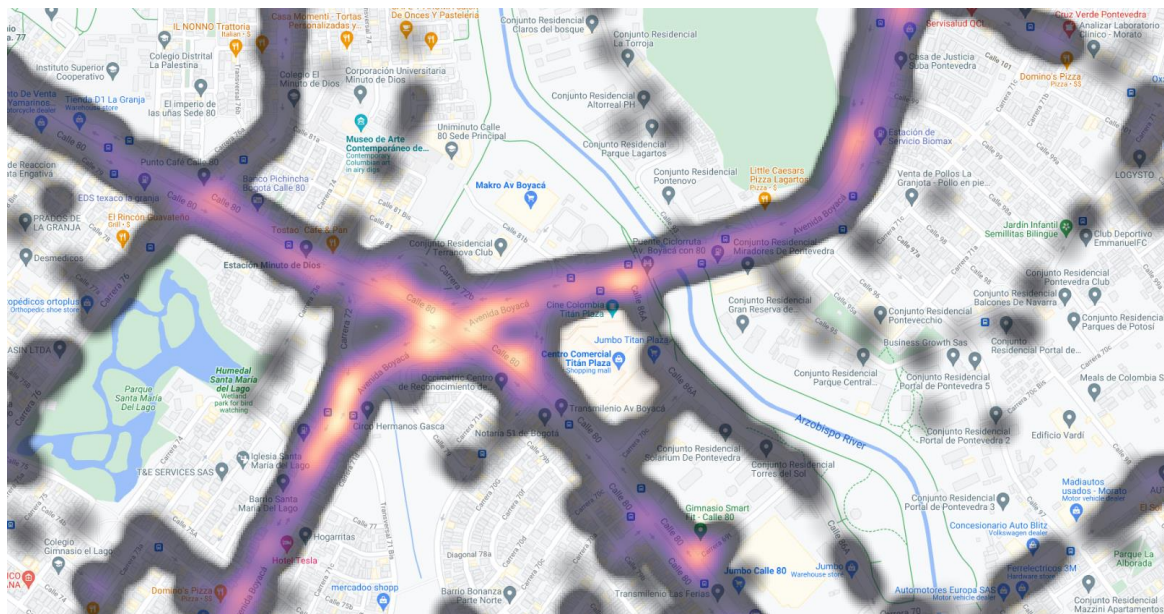
Se identifican los puntos calientes en la zona de Salitre Mágico, siendo la mayoría de estos accidentes en la Av. Boyacá desde la Avenida Calle 24 hasta la Calle 63. Adicional a los puntos calientes en la Av. Boyacá se presentan dos puntos calientes más en dos intersecciones. En la Av. Ciudad de Cali con Av. El Dorado y en la rotonda de la Calle 63 con Carrera 68.

**Figura 32.** Heat map accidentes viales en Autopista Norte con calle 100

Fuente: Elaboración propia usando QGIS.

Los puntos calientes en la Autopista Norte con calle 100 se presentan todos a lo largo de la autopista norte, y en ambos sentidos, pero mayormente en el sentido sur a norte.

**Figura 33.** Heat map accidentes viales en Av. Boyacá con calle 80

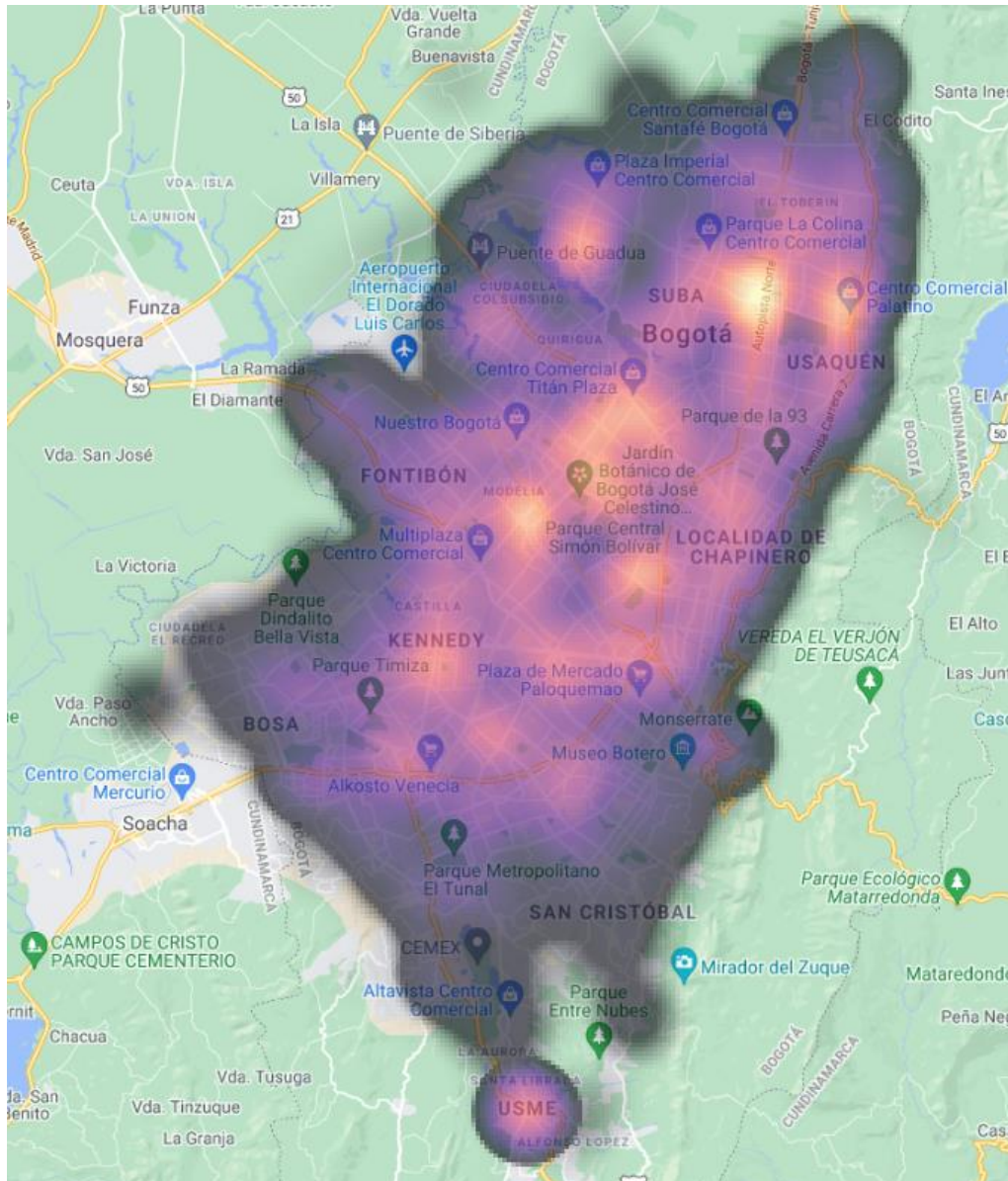


Fuente: Elaboración propia usando QGIS.

Los accidentes en la Av. Boyacá con calle 80 ocurren mayormente en la Av. Boyacá en ambos sentidos, y con un punto caliente aislado en sobre la altura de la Calle 86A. También hay puntos calientes sobre el puente de la calle 80 que cruza la Av. Boyacá.

## 5.4 Análisis de huecos en Bogotá

En esta sección se realizarán análisis de mapas de punto caliente con las alertas que reportan o identifican huecos o defectos en la vía, y así poder identificar si existe alguna relación con los puntos calientes identificados en la sección anterior. Se reportaron un total de 68,287 alertas tipo HAZARD y subtipo HAZARD\_ON\_ROAD\_POT\_HOLE, que son las alertas que identifican huecos o defectos en las vías. En el mapa de calor se identifican las zonas donde más predominan los reportes de huecos.

**Figura 34.** Heat map huecos o defectos en vías de Bogotá

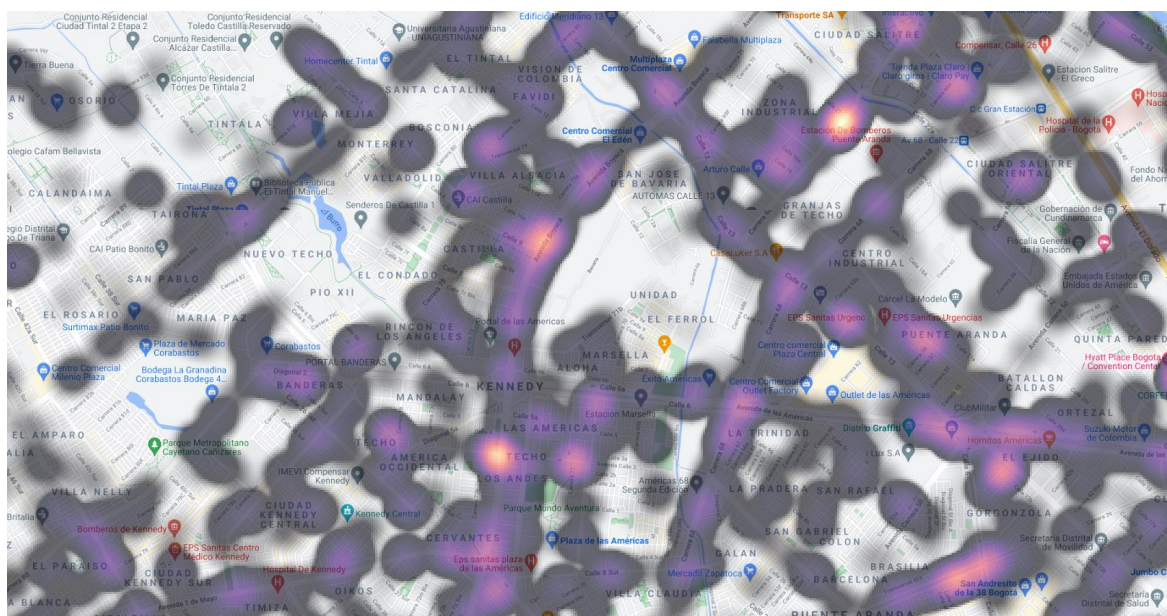
Fuente: Elaboración propia usando QGIS.

En este primer mapa de calor generado para toda la ciudad de Bogotá se identificó un punto caliente bastante predominante en la zona del barrio Spring y otro punto en los alrededores del Centro Comercial Palatino.

Se generan cuatro mapas de calor de los huecos o vía deteriorada, uno por cada zona identificada en la sección anterior, para analizar si existe alguna relación de la vía deteriorada con los accidentes.

### 5.4.1 Huecos o defectos en la vía en los puntos de concentración de accidentes

**Figura 35.** Heat map huecos o defectos en Kennedy

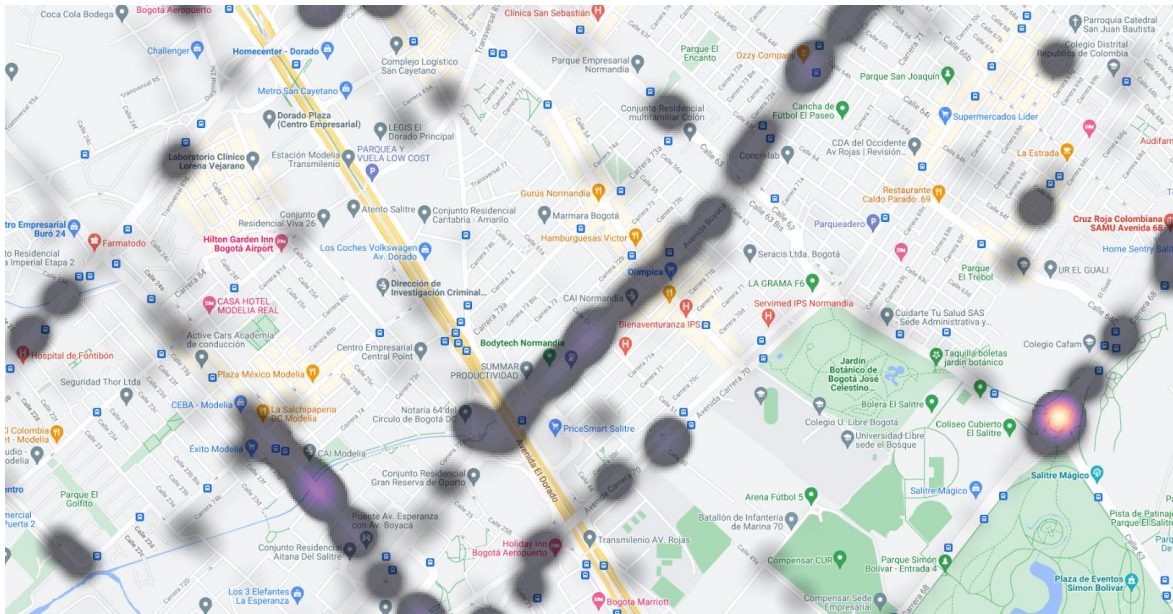


Fuente: Elaboración propia usando QGIS.

Con este mapa de calor se identifican tres puntos de calor en Kennedy:

- Av. Boyacá con Avenida Calle 3
- Av. Boyacá a la altura de la Calle 9
- Carrera 68d con calle 21

El único punto caliente en común con el mapa de calor de accidentalidad es el de la Av. Boyacá a la altura de la Calle 9. Cabe resaltar que hay reporte de vías defectuosas en las vías principales de la malla vial de esta zona de Kennedy.

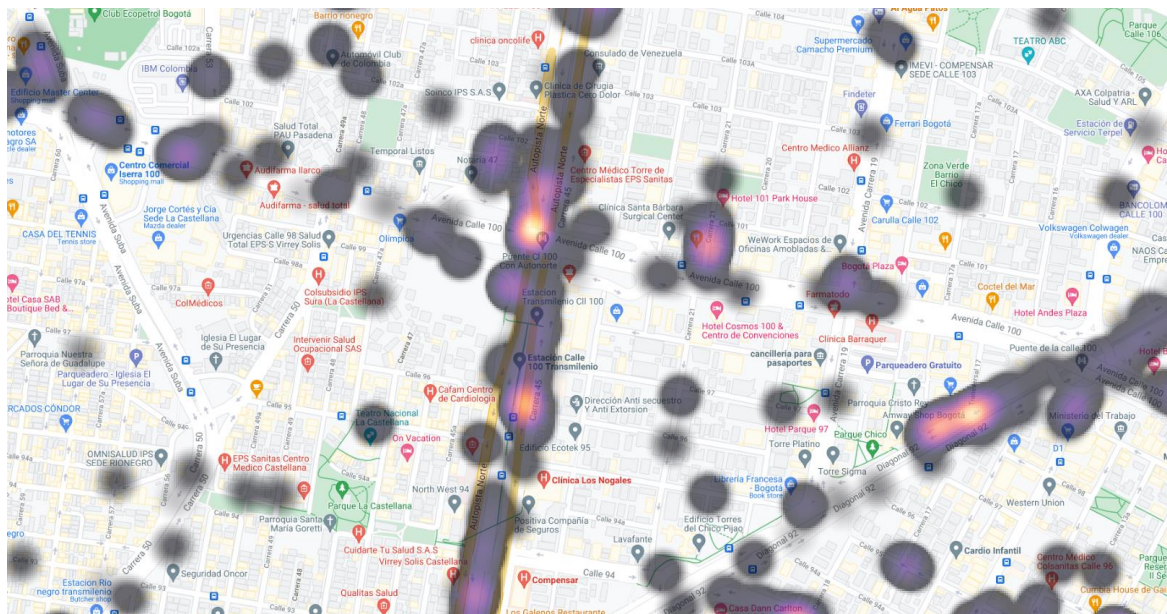
**Figura 36.** Heat map huecos o defectos en Salitre Mágico

Fuente: Elaboración propia usando QGIS.

En Salitre Mágico solo se presenta un punto caliente en la rotonda de la Calle 63 con Carrera 68. Este punto caliente es el único compartido con el mapa de calor de accidentes para la misma zona, ya que en el mapa de calor de accidentes se presentan los puntos calientes en la Av. Boyacá y este mapa no presenta puntos de calor allí.



**Figura 37.** Heat map huecos o defectos en Autopista Norte con calle 100

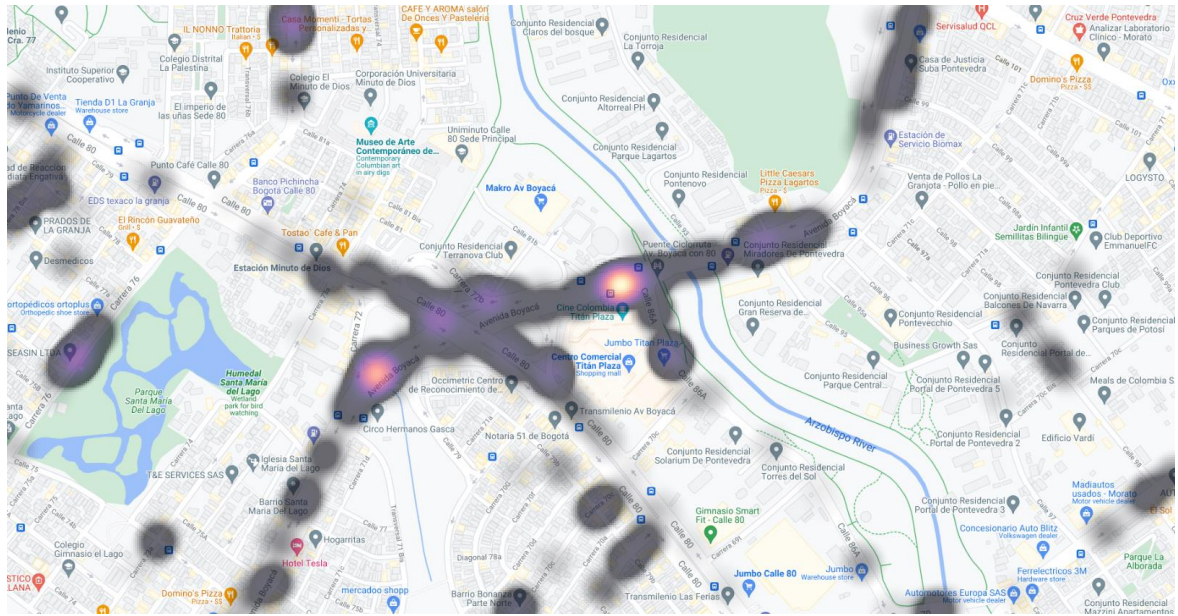


Fuente: Elaboración propia usando QGIS.

El mapa de calor de la Autopista Norte con calle 100 presenta tres puntos calientes de defectos en la vía:

- Autopista norte con calle 100, sentido norte-sur
- Autopista norte con calle 96 sentido sur-norte
- Diagonal 92 con transversal 17

Los puntos calientes situados en la autopista norte sí coinciden con algunos de los puntos calientes del mapa de calor de accidentalidad. El punto caliente a la altura de la calle 96 es cercano al punto caliente presentado en la calle 97 en el mapa de calor de accidentalidad.

**Figura 38.** Heat map huecos o defectos en Av. Boyacá con calle 80

Fuente: Elaboración propia usando QGIS.

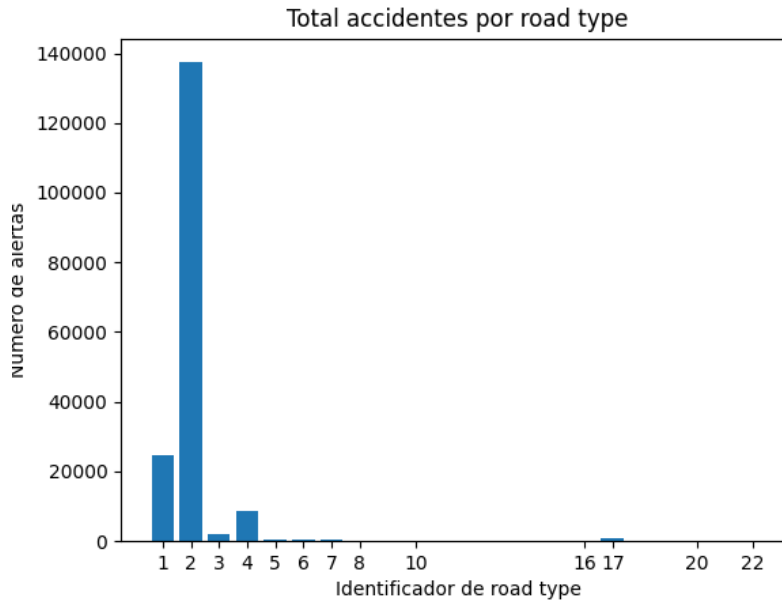
Este mapa de calor refleja dos puntos calientes, uno en la Av. Boyacá a la altura de la Calle 79 y otro en la Av. Boyacá a la altura de la Calle 86A. Estos dos puntos calientes coinciden con los puntos calientes reflejados en el mapa de calor de accidentes, pero los puntos calientes del mapa de accidentalidad presentados sobre la calle 80 no están presentes en este mapa de calor. Cabe resaltar que de las zonas presentadas es el que menos alertas de huecos presenta junto a la zona del Salitre Mágico.

## 5.5 Análisis accidentes por tipo de vía

En esta sección se realizará un análisis de accidentes con respecto al tipo de vía en que ocurre. Los identificadores road types representan a qué tipo de vía pertenece la alerta dada por su localización en coordenadas.

**Tabla 1.** Tipos de vía en las alertas de Waze

<b>roadType</b>	<b>Definición</b>
<b>1</b>	Calle en un barrio
<b>2, 3, 4</b>	Avenida
<b>5</b>	Vías compartidas carro/moto/peatonal
<b>6, 7</b>	Vías en límites de Bogotá
<b>8</b>	Vías sin pavimentar
<b>10, 16, 17</b>	Ciclorutas y peatonales
<b>20</b>	Vías sin salida y parqueaderos
<b>22</b>	Callejones

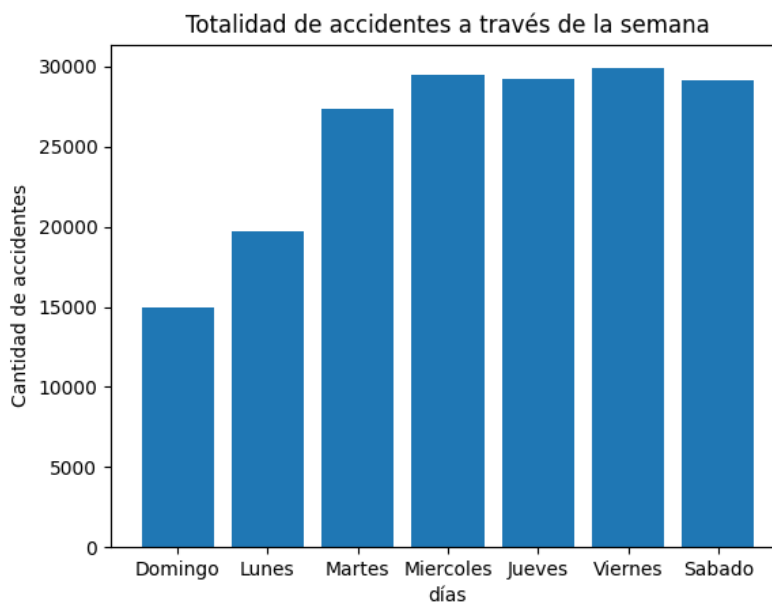
**Figura 39.** Totalidad de accidentes por tipo de vía

Fuente: Elaboración propia.

La mayoría de los accidentes viales ocurren en las avenidas, siendo el 84.6% de los accidentes. Seguidamente los accidentes ocurren en las calles de barrio, con un 14.06% de los casos. Es evidente que los accidentes tienen una relación con la alta velocidad, debido a que la mayoría de los accidentes ocurren en avenidas donde la velocidad permitida es mayor a la de una calle de un barrio. Los accidentes en ciclorrutas o peatonales son el 0.42%.

## 5.6 Análisis accidentes por día de la semana

Es evidente que la cantidad de accidentes viales aumenta con la cantidad de conductores en la vía. Se realiza una gráfica donde se evidencia la cantidad de accidentes viales por día. La figura representa el total de accidentes viales durante el día para cada día de la semana.

**Figura 40.** Totalidad de accidentes por día de la semana

Fuente: Elaboración propia.

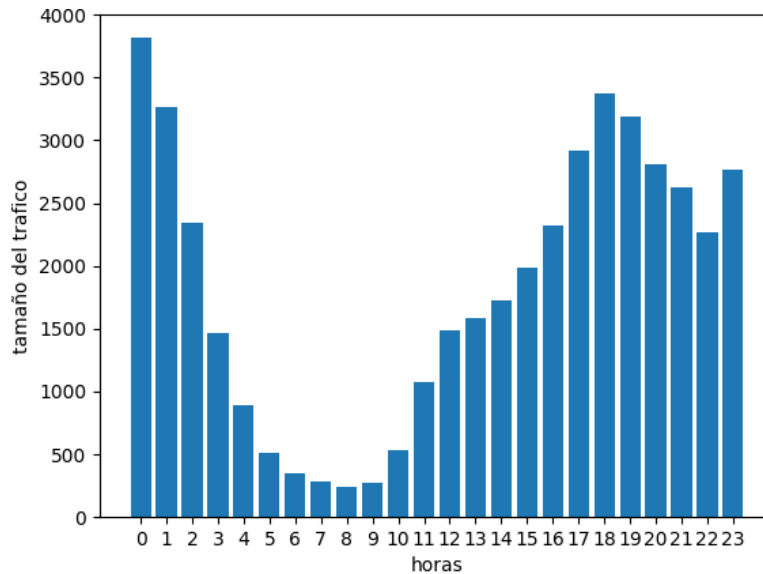
Entre los días martes a sábados ocurren la mayoría de los accidentes viales, siendo casi constante el número de accidentes en cada uno de esos días. Aparentemente el lunes tiene menor número de accidentes frente a los demás días, pero podría ser debido a que en el rango de fecha del presente estudio hubo tres días festivos y eso se refleja en una baja de los automóviles en las vías, y por ende una baja en los accidentes. El domingo presenta el menor número de accidentes viales. El viernes fue el día que mayor número de accidentes presenta con un total de 29,897 accidentes y el Domingo el día con menos accidentes, 14,931.

## 5.7 Análisis accidentes por hora

Para lograr una mayor precisión e identificar el patrón o fenómeno de los accidentes viales en la ciudad de Bogotá, se realizan gráficas separadas por días de la semana y fines de

semana o festivos, así pudiendo encontrar una diferencia significativa entre el comportamiento en el número de los accidentes de un día entre semana y un día de fin de semana.

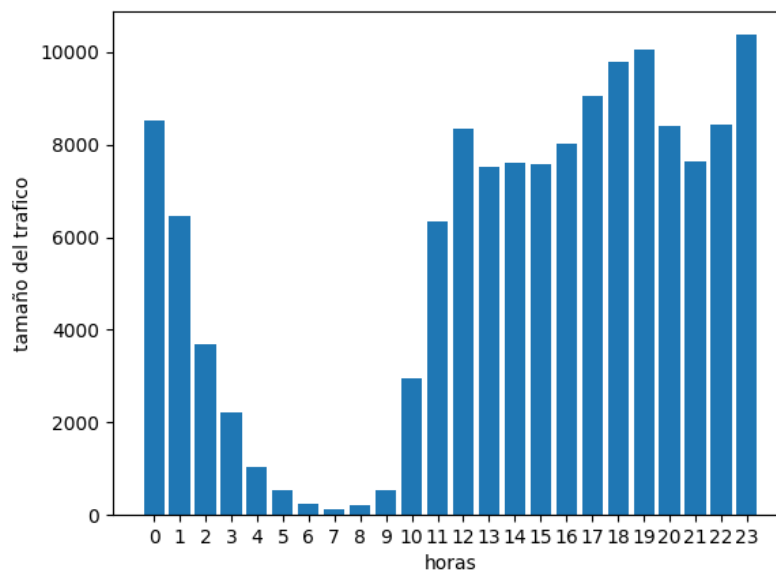
**Figura 41.** Totalidad de accidentes por hora para días de fin de semana



Fuente: Elaboración propia.

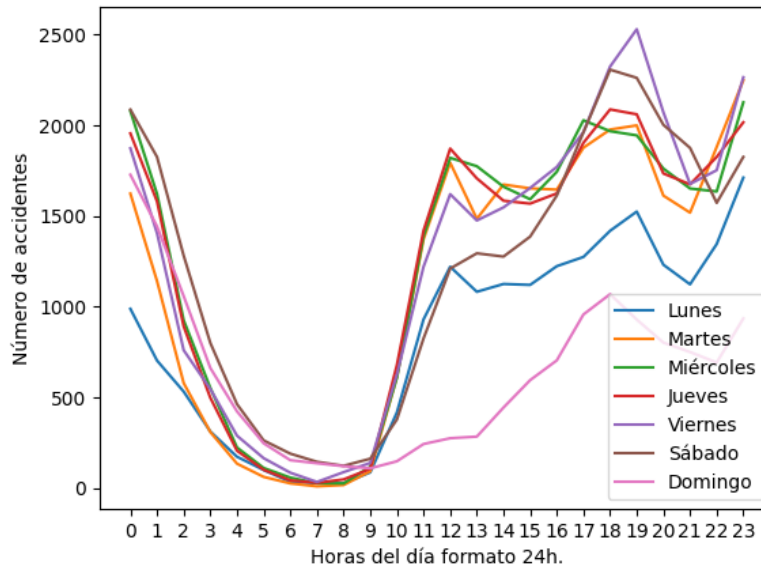
Para el caso de los días de fin de semana, la hora en que se genera mayor tasa de accidentalidad vial es entre el rango de 11 pm a 1 am. El segundo pico es entre las 5 pm y las 7 pm.

**Figura 42.** Totalidad de accidentes por hora para días de la semana laboral



Fuente: Elaboración propia.

Los accidentes en los días de la semana laboral ocurren mayormente a las 11 pm, y entre las 5 pm a 8 pm. Para un mayor detalle del comportamiento de los accidentes, se realizó la gráfica durante las horas para cada día de la semana y así poder encontrar un patrón visual.

**Figura 43.** Totalidad de accidentes por hora para días de la semana

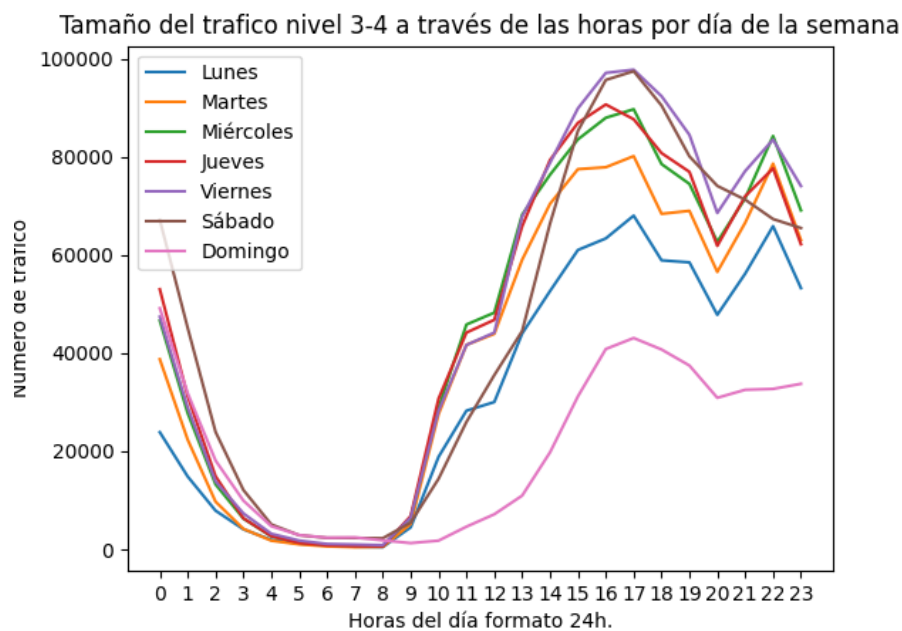
Fuente: Elaboración propia.

Esta gráfica nos confirma que los lunes y domingos presentan menor número de accidentes viales, pero desde las 11 am aproximadamente, porque entre la 1 am hasta las 11am se presenta el mismo número de accidentes viales que en los demás días de la semana. Todos los demás días de la semana presentan un comportamiento de accidentes por hora muy similar, teniendo un pico a las 12 del mediodía y a las 7pm, que es la hora donde se presenta el mayor número de accidentes viales.

Adicionalmente, comparamos esta gráfica con la gráfica del nivel de tráfico durante las horas del día para cada día de la semana y ver si tienen alguna relación accidente-tráfico. Se usaron los datos de tráfico de nivel 3 y 4, que son los niveles que representan mayor tráfico vial.



**Figura 44.** Totalidad de segmentos de tráfico nivel 3-4 por hora para días de la semana

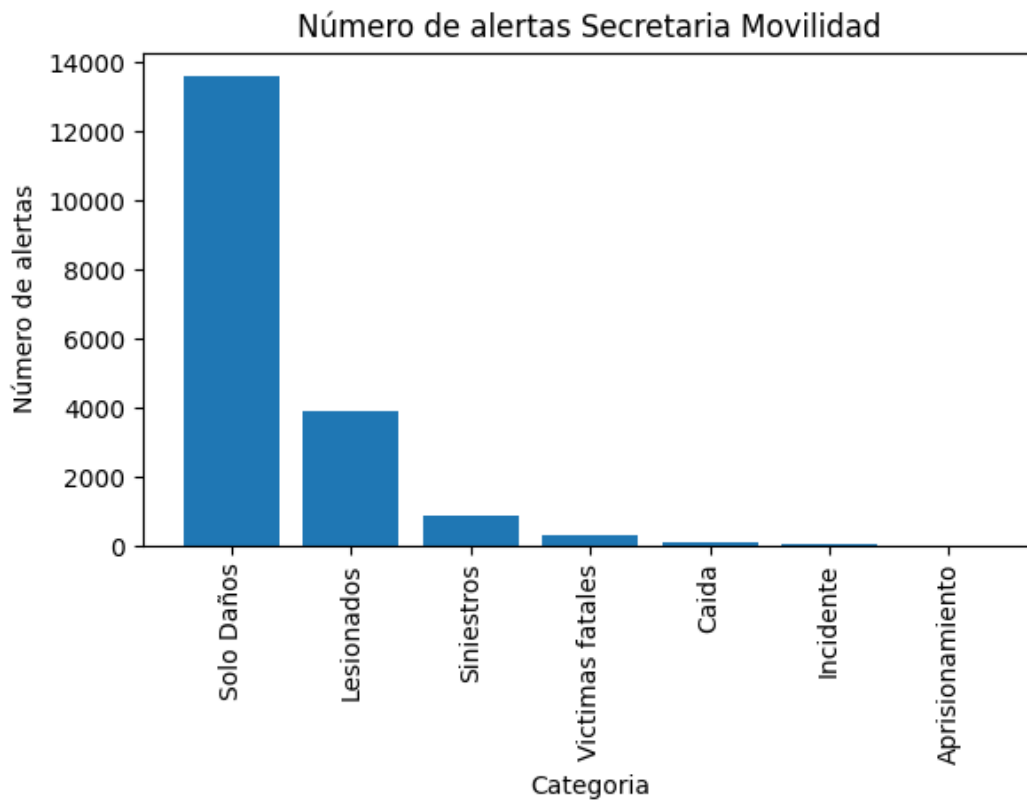


Fuente: Elaboración propia.

El pico de accidentes viales se presenta sobre las 7 pm, pero para el tráfico en esas horas es elevado, pero no el mayor, dado que el pico del tráfico se presenta sobre las 5 pm, hora en la que los accidentes viales son bastantes, pero no es el pico. El segundo pico de los accidentes viales se presenta sobre las 12 del mediodía, y a esa hora el tráfico no es tan alto. No se presenta una relación directa entre tráfico-accidentalidad. Se podría decir que un mayor tráfico vial no representa un aumento en la accidentalidad.

## 5.8 Alertas accidentes viales Secretaría de Movilidad en Waze

**Figura 45.** Totalidad de alertas Secretaria Movilidad por tipo



Fuente: Elaboración propia.

Las alertas de Waze tienen un campo llamado *'reportDescription'*, con él se logra identificar en su descripción si el accidente cuenta con víctimas fatales, lesionados o ningún afectado. Estas alertas son realizadas por la Secretaría de Movilidad

Estos datos reportados por la Secretaría de Movilidad categorizan la alerta para identificar si el accidente cuenta con lesionados, víctimas fatales o sólo daños.

**Tabla 2.** Totalidad de alertas Secretaria Movilidad por tipo y nivel de accidente

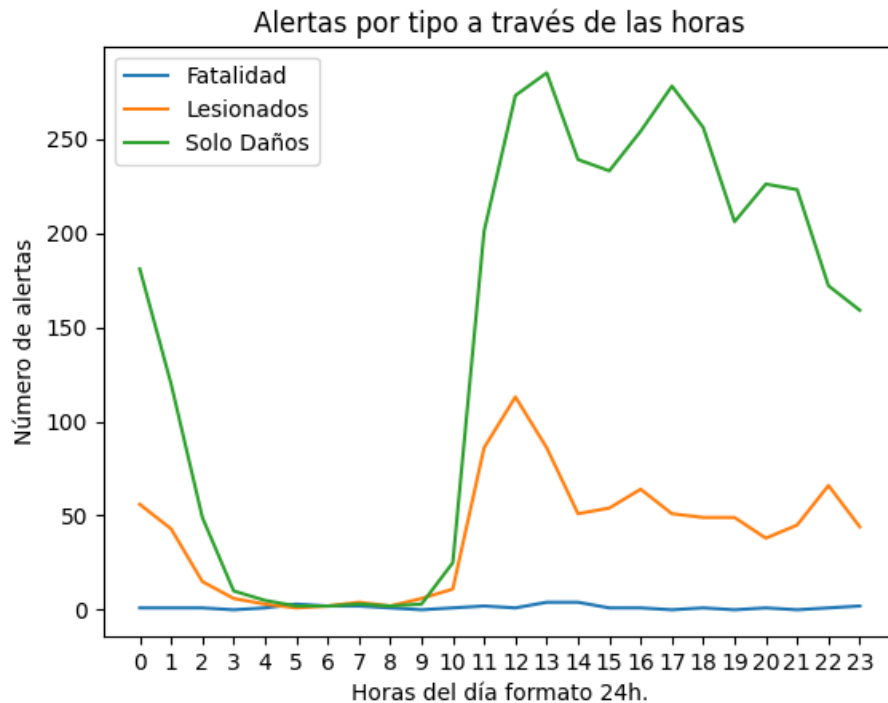
Tipo de alerta	Cantidad ACCIDENT_MINOR	Cantidad ACCIDENT_MAJOR	Total
Accidente sólo daños	3366	41	3407
Accidente con lesionados	131	814	945
Accidente con víctimas fatales	0	31	31
Caída de personas en transporte público	6	17	23
Incidentes en la vía	23	0	23
Aprisionamiento de personas	3	0	3

La Secretaría de Movilidad realizó un total de 4,432 reportes de alerta en Waze en las fechas de estudio del presente trabajo. Según las alertas realizadas por la Secretaría de Movilidad la mayoría de los accidentes con víctimas fatales, el 80,6%, ocurre en avenidas o vías principales, por ende, la posible causa de estos accidentes es debido a que estas vías tienen en común una mayor velocidad. El exceso de velocidad aumenta las posibilidades de que un accidente termine con una víctima fatal.

El 76.87% de los accidentes son accidentes que presentan sólo daños, el 21.32% de los accidentes tuvieron un lesionado en el hecho y el 0.7% fueron accidentes con víctima fatal.

### 5.8.1 Total accidentes por tipo por hora

**Figura 46.** Totalidad de alertas Secretaria Movilidad por tipo por hora



Fuente: Elaboración propia.

Como se identificó anteriormente, los accidentes empiezan a crecer desde las 10 am. Para los accidentes con lesionados presenta un pico a las 12 del mediodía, siendo la hora en la que se presenta mayor número de accidentes con lesionados. Para los accidentes que presentan solo daños tiene tres picos; a las 12 del mediodía, a las 5pm y a las 9 pm. A las 12 del mediodía es la hora donde se presentan el mayor número de accidentes con lesionados o sólo daños. Para los accidentes con víctimas fatales se presenta un leve entre la 1 pm y 2 pm.

### 5.8.2 Duración promedio accidentes por tipo

Las alertas extraídas se realizan con una periodicidad de 10 minutos. Hay alertas que permanecen aún en varias ejecuciones de extracción de los datos, esto debido a que el

hecho o acto que refleja la alerta aún permanece en la vía. Así podremos realizar un análisis acerca de la duración de las alertas de tipo fatalidad, lesionados y solo daños, obteniendo un promedio de duración de dichos eventos.

**Tabla 3.** Duración promedio accidentes por tipo

Tipo de alerta	Duración promedio (minutos)
Accidente sólo daños	39
Accidente con lesionados	41
Accidente con víctimas fatales	94

La alerta de tipo fatalidad de mayor duración fue de 18 ejecuciones de extracción de datos, es decir 180 minutos de duración. La de menor duración fue de 1 ejecución. Su tiempo promedio fue de 9.45 ejecuciones, es decir 94 minutos duran aproximadamente las alertas de este tipo.

La alerta de tipo lesionados de mayor duración fue de 42 ejecuciones de extracción de datos, es decir 420 minutos de duración. La de menor duración fue de 1 ejecución. Su tiempo promedio fue de 4.1 ejecuciones, es decir 41 minutos duran aproximadamente las alertas de este tipo.

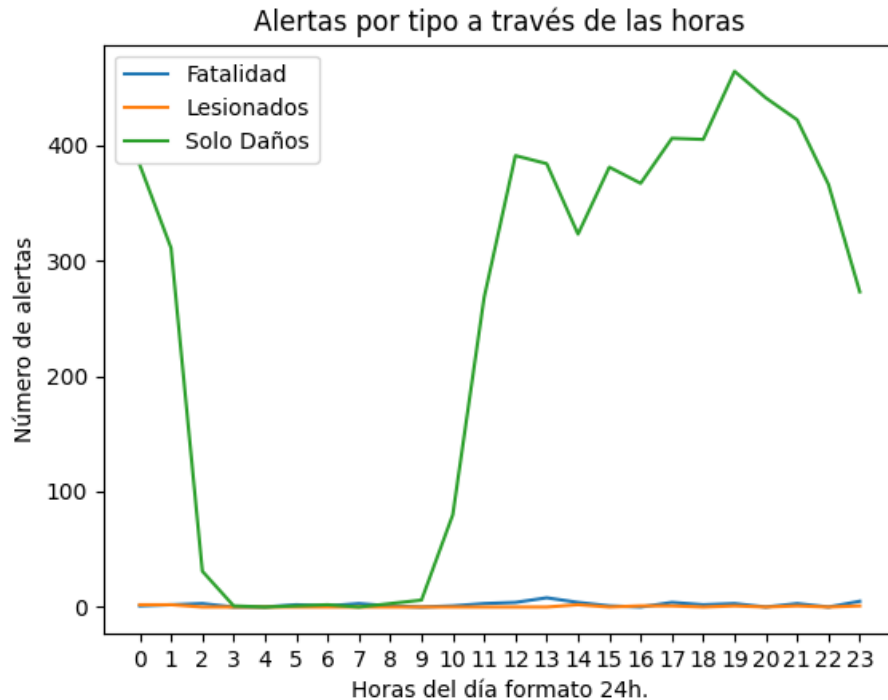
La alerta de tipo lesionados de mayor duración fue de 76 ejecuciones de extracción de datos, es decir 760 minutos de duración. La de menor duración fue de 1 ejecución. Su tiempo promedio fue de 3.99 ejecuciones, es decir 39 minutos duran aproximadamente las alertas de este tipo.

Las alertas de accidentes de mayor duración fueron las alertas que tienen implicado una víctima fatal, esto debido a la gravedad del accidente y a la necesidad de análisis de los fallecidos. Seguidamente las víctimas con lesionados y daños tienen un tiempo de duración similar, aproximadamente de 40 minutos.

## 5.9 Accidentes extraídos de Twitter

En esta sección se analizaron los datos de accidentes extraídos de twitter de la red social @BogotaTransito, tweets acerca de accidentes viales. Se extrajeron 6,652 acerca de accidentalidad con fatalidad, lesionados o sólo daños. De los tweets extraídos se analizaron y se identificaron 50 accidentes con fatalidad, de estos el 76% ocurrieron en vía principal. Se confirma que estos accidentes comparten las altas velocidades de los vehículos que transitan en ellas.

**Figura 47.** Totalidad de alertas Twitter por tipo por hora



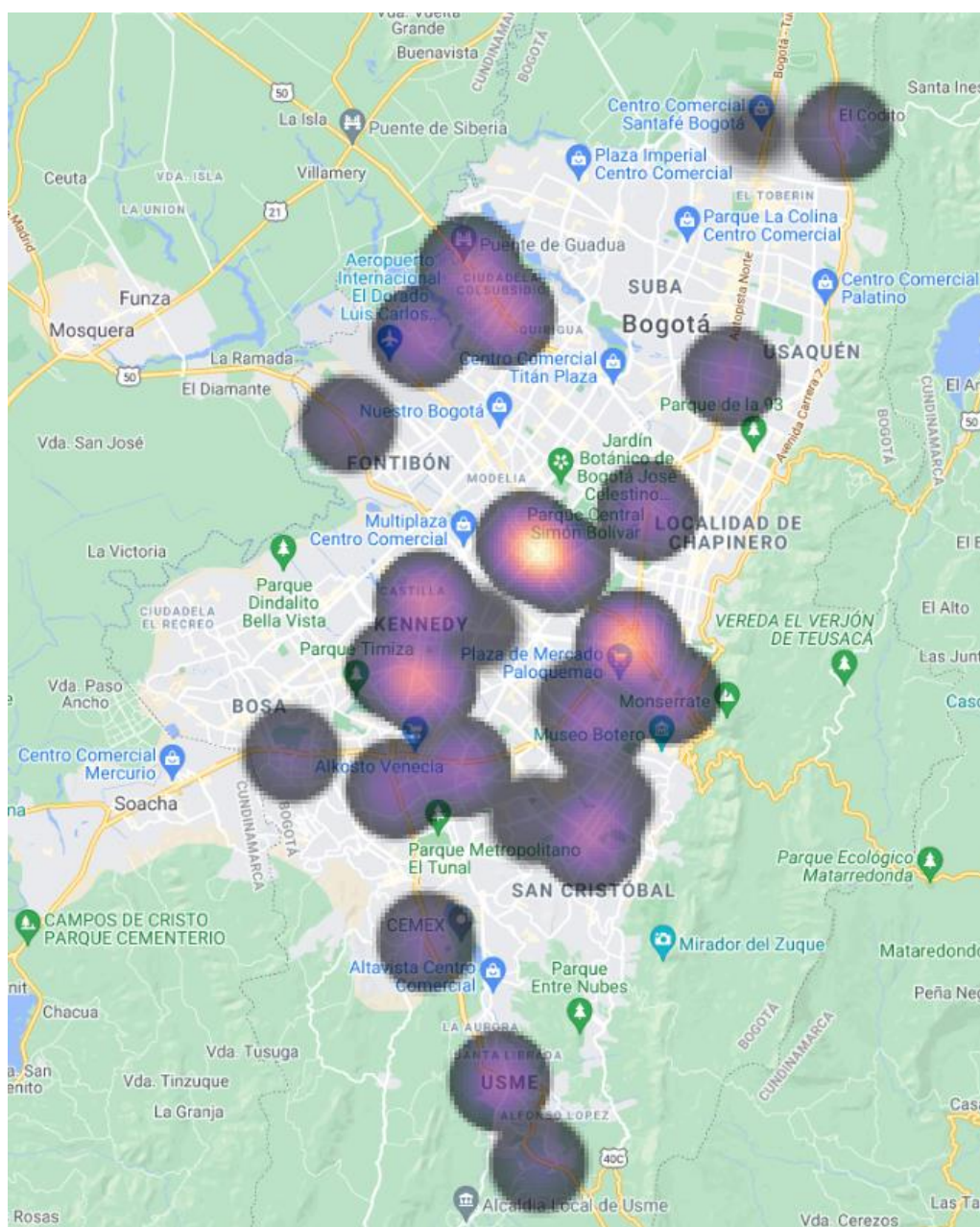
Fuente: Elaboración propia.

La mayoría de los accidentes que ocurren son de sólo daños con 6579 tweets que representan accidentes con sólo daños, siendo el 98.9% de los accidentes. Fue el caso que más reporta, ya que el de fatalidad fueron 50 casos. En la gráfica se evidencia que a partir de las 10 am se evidencia una subida de accidentes con solo daño, teniendo dos picos, uno a las 12 del mediodía y otro a las 7 pm. El pico de los accidentes con fatalidad se presenta a la 1 pm. Estos datos se comportan de la misma manera de los vistos en los

datos de la Secretaría de Movilidad en la red social Waze. Se realizan *heat maps* para poder identificar los lugares donde ocurren estos accidentes reportados.

### 5.9.1 Heat map víctimas fatales

**Figura 48.** Heat map de accidentes con víctimas fatales

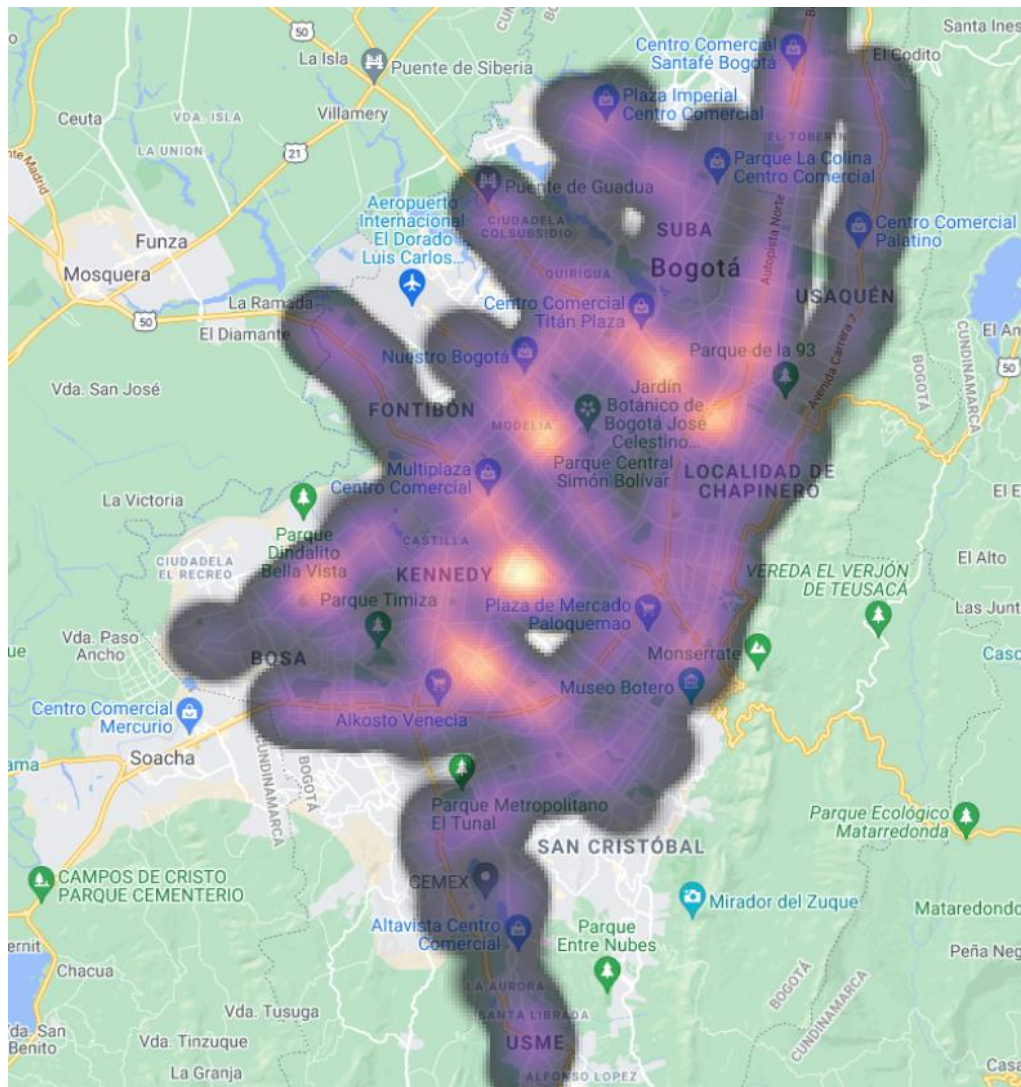


Fuente: Elaboración propia usando QGIS.

Este mapa de calor identifica dos puntos calientes de accidentes con fatalidad, uno en la Avenida Esperanza con Carrera 68a y otro en la Avenida NQS a la altura de la calle 25.

## 5.9.2 Heat map víctimas lesionadas

Figura 49. Heat map de accidentes con lesionados



Fuente: Elaboración propia usando QGIS.

Para el caso de los accidentes con lesionados se presentan seis puntos calientes:

- Avenida 1 de Mayo con Carrera 68.
- Carrera 68 entre Calle 6 y Calle 13.
- Avenida Ciudad de Villavicencio con Carrera 86.

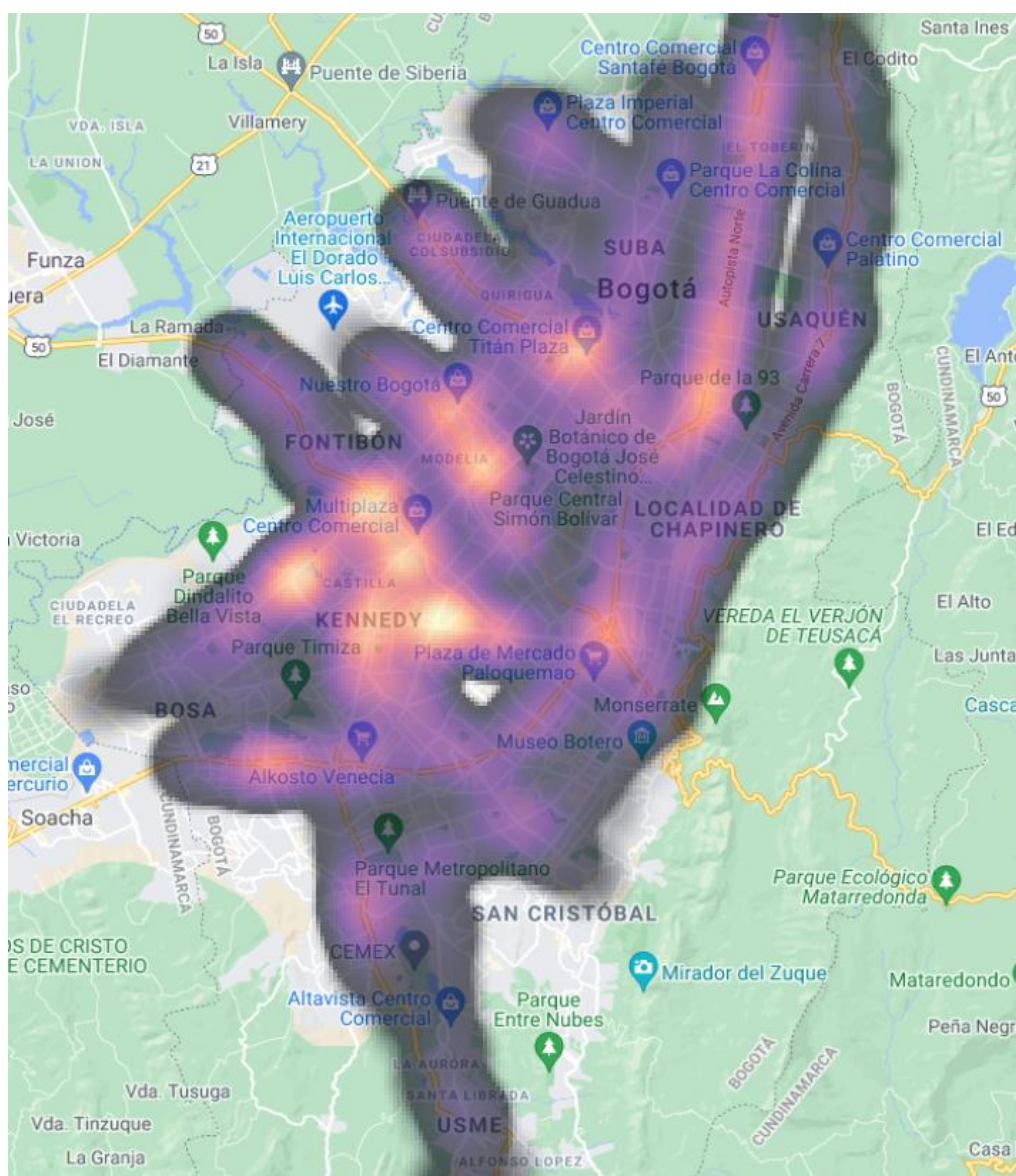


- Avenida Boyacá a la altura de la Calle 26.
- Calle 80 a la altura de la Carrera 68
- Av. NQS a la altura de la calle 80

Cabe anotar que la avenida donde se refleja el punto caliente del mapa de accidentes fatales no se refleja en este mismo mapa.

### 5.9.3 Heat map solo daños

Figura 50. Heat map de accidentes con sólo daños



Fuente: Elaboración propia usando QGIS.

Para el caso de accidentes viales con solo daños se presentan 10 puntos calientes:

- Autopista Sur a la altura de la Carrera 67.
- Avenida Boyacá con Calle 6.
- Av. de las Américas con Carrera 68.
- Av. Ciudad de Cali con Calle 6.
- Avenida Boyacá a la altura del Puente Avenida Calle 17.
- Avenida calle 17 a la altura Av. Ciudad de Cali.
- Avenida Boyacá a la altura de la Calle 19.
- Avenida el Dorado con Avenida Ciudad de Cali
- Avenida Boyacá con Calle 80.
- Autopista Norte con Calle 100.

Con el mapa de calor de accidentes con lesionados se comparten dos puntos de calor, Carrera 68 con Calle 6 y Avenida Boyacá a la altura de la Calle 19. Cabe anotar que sobre la autopista norte se mantiene gran cantidad de accidentes, y en su mayoría ocurren en las intersecciones con los puentes.

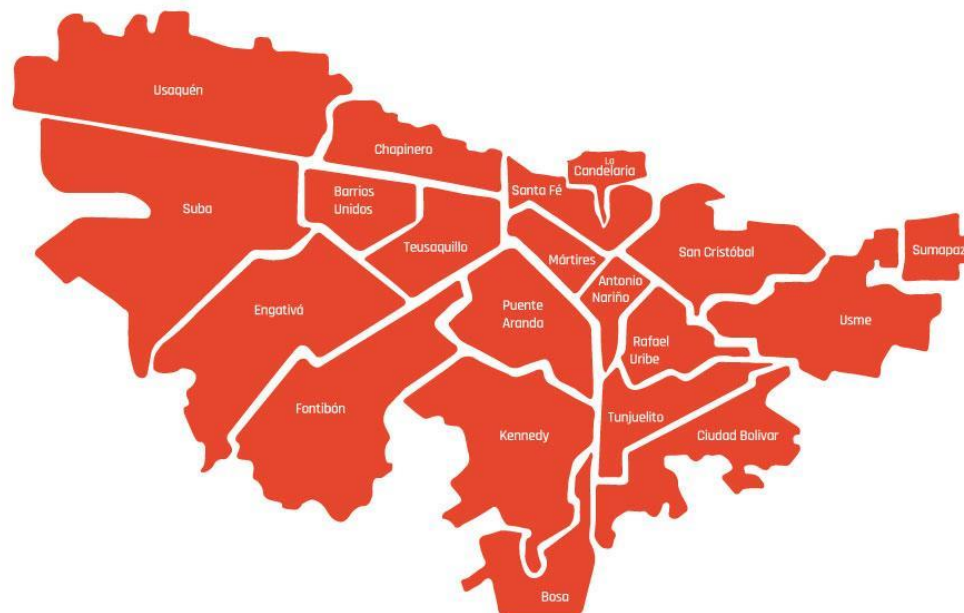
## **5.10 Localidades con mayor cantidad de accidentes**

En esta sección se realiza una clasificación y análisis de accidentes por localidad. Para ellos se utilizan los mismos datos de accidentalidad usado en las secciones anteriores y se localizan a qué localidad pertenece para así analizar la totalidad de accidentes por localidad.

### **5.10.1 Identificación de localidad-accidente**

Para lograr esta clasificación se usaron los polígonos geográficos que definen los límites de cada localidad de Bogotá. La ciudad de Bogotá se divide en 20 localidades.

**Figura 51.** Mapa de localidades de Bogotá

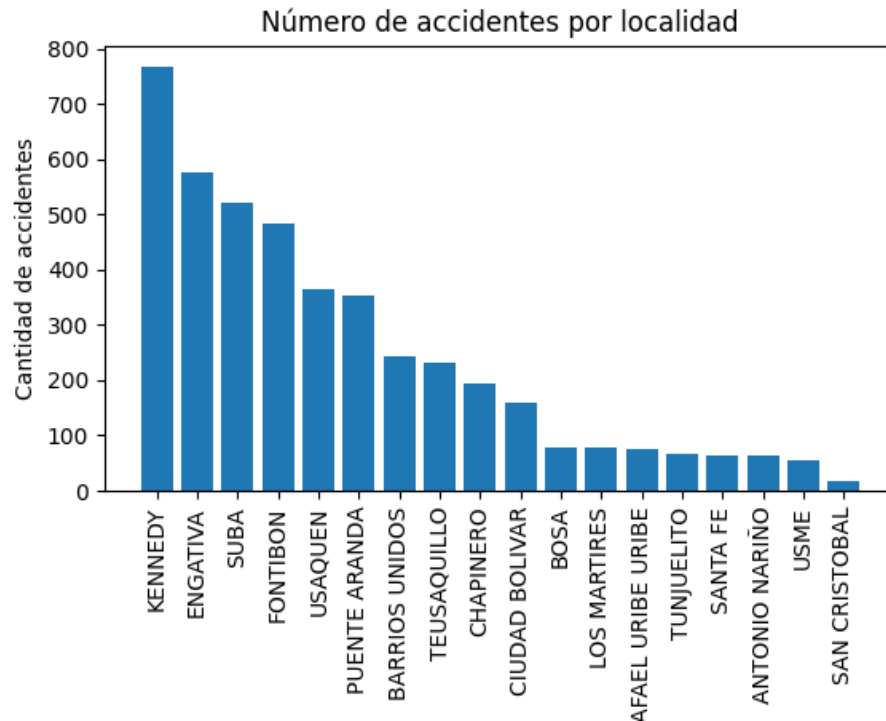


Fuente: <https://bogota.gov.co/mi-ciudad/administracion-distrital/proceso-para-eleccion-de-alcaldes-y-alcaldesas-locales-en-bogota>

Para lograr la clasificación de accidentes por localidad se tomaron los polígonos geográficos del Laboratorio Urbano de Bogotá [33], el cual se encuentra en diferentes formatos para su respectivo tratamiento. Cada polígono que representa una localidad consta de un arreglo o lista de puntos coordenadas el cual define el polígono. Se usó un algoritmo para identificar si un punto coordinado que representa un accidente pertenece a un polígono y así saber a cuál polígono pertenece. Luego de esta clasificación se obtuvieron los totales de accidentes por localidad.

## 5.10.2 Número de accidentes por localidad

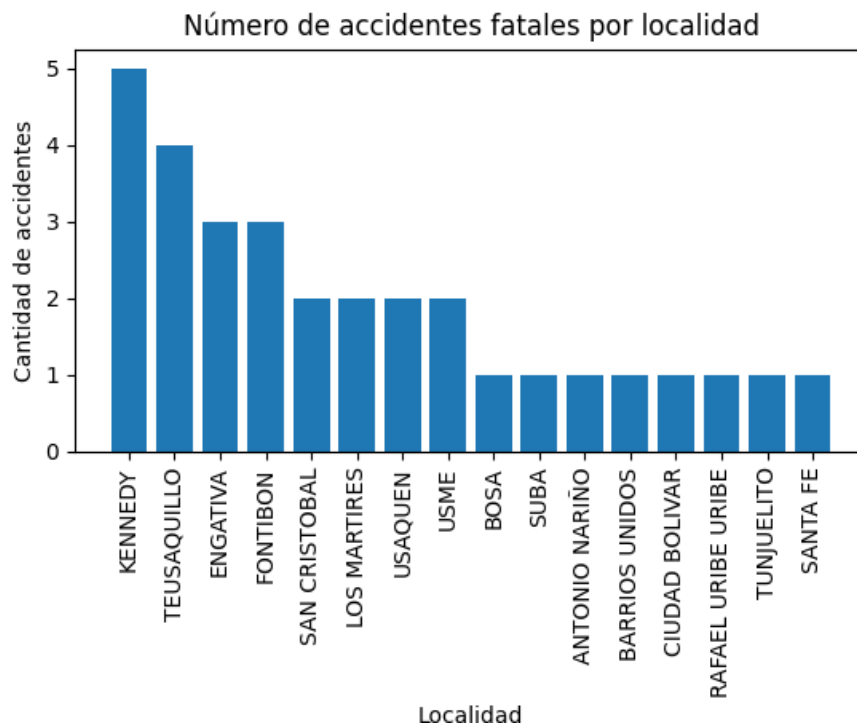
**Figura 52.** Número de accidentes por localidad



Fuente: Elaboración propia.

El 53.5% de los accidentes se encuentran solo en las primeras cuatro localidades, siendo la localidad de Kennedy la que mayor accidente vial presenta, con un 17.5%, Engativá 13.1%, Suba 11.89%, Fontibón 11%, Usaquén 8.29% y Puente Aranda 8.02%. Estas son las localidades que mayor número de accidentes viales presentan.

**Figura 53.** Número de accidentes fatales por localidad



Fuente: Elaboración propia.

Para el caso de los accidentes con fatalidad por localidad Kennedy continúa siendo la localidad con el mayor número de accidentes con un 16.13% y continúa la localidad de Teusaquillo con un 12.9%, siendo un porcentaje alto teniendo en cuenta que en la totalidad de accidentes tiene un 5.31%, no siendo la localidad más alta en accidentes.

## 6. Conclusiones y recomendaciones

### 6.1 Conclusiones

El objetivo de este trabajo es el análisis de datos de accidentalidad vial en la ciudad de Bogotá a partir de la extracción de los datos de redes sociales y datos abiertos. Inicialmente se investigó acerca del análisis y uso de los datos viales en las ciudades para identificar que han logrado o que se han planteado acerca de la accidentalidad vial. Con esta investigación previa se logró una gran motivación y un punto de partida para el desarrollo de este análisis en la ciudad de Bogotá.

Para lograr la extracción de los datos se investigó los métodos y el API necesarios para la correcta y efectiva recolección de los datos de estudio y así lograr el análisis geoespacial de estos. Se automatiza la recolección de los datos con el uso de *Azure functions* para recolectar la información cada cierto tiempo. Estos datos luego se almacenan en la base de datos de MongoDB. Para lograr la mayor cantidad de información se realizaron extracción por partes de Bogotá usando los bounding box o tiles de Google.

---

El análisis aplica un enfoque en la zona, el mes, los días de la semana y horas del día en el que ocurren los accidentes y diferentes hechos que influyen sobre los accidentes. Es decir, se aplicó un análisis espacio temporal.

En el análisis de los datos se encontraron zonas con mayor cantidad u ocurrencias de accidentes viales:

- En Kennedy fue la zona que mayor cantidad de accidentes reportó. Inicialmente se evidencia que presenta gran cantidad de accidentes viales a lo largo de su malla vial, principalmente en las avenidas. Los lugares donde ocurren mayor cantidad de accidentes son en los cruces entre avenidas. Esta localidad presentó seis puntos calientes.
- En Salitre Mágico se presentan accidentes viales sobre las avenidas principales. Principalmente la Avenida Boyacá es donde ocurren la mayoría de sus accidentes viales.
- La Autopista Norte con calle 100 es un lugar donde ocurren bastantes accidentes viales, y este ocurre en ambos sentidos de esta avenida.
- En la Av. Boyacá con calle 80 presenta accidentes tanto sobre la Avenida Boyacá y sobre la Calle 88, siendo mayor sobre la Avenida Boyacá.

La mayoría de los accidentes evidencias se reflejan sobre las avenidas principales de la ciudad, siendo la Av Boyacá la que mayor se vio implicada en los puntos calientes de accidentalidad de Bogotá. Esto refleja que los accidentes viales están relacionados con la alta velocidad de los vehículos. Se realizó un análisis de los huecos o defectos viales en la malla vial de Bogotá, y se concluye que existe una baja relación en los lugares de los accidentes frente a los lugares donde se reportan los huecos.

Con el análisis de tráfico y accidentalidad vial por hora se podría concluir que un mayor tráfico vial no representa un aumento en la accidentalidad. La hora con mayor tráfico en Bogotá es a las 11 pm y 7 pm. El viernes es el día con mayores accidentes viales. El momento en el que ocurren más accidentes viales en Bogotá son los viernes a las 7 pm y el momento donde se presentan menos accidentes viales es el miércoles a las 7 am.

La mayoría de los accidentes viales presentan solo daños, seguido de accidentes con lesionados y en menor cantidad con víctimas fatales. Adicionalmente se puede concluir que la duración promedio de los accidentes viales con víctimas fatales es de 94 minutos, solo daños de 39 minutos y lesionados de 41 minutos.

Al realizar un análisis de accidentes por localidad se obtiene que Kennedy es la localidad con mayor número de accidentes viales de la ciudad, como se encontró en la investigación realizada. Seguida de Kennedy se encuentran Engativá, Suba, Fontibón, Usaquén y Puente Aranda como las localidades con mayor número de accidentes. Usme y San Cristóbal son las localidades con menor número de accidentes viales.

## **6.2 Recomendaciones**

Se encuentran las redes sociales como una fuente grande e inmediata de datos viales en Bogotá. Es importante motivar a los ciudadanos a reportar en estas redes para lograr una información inmediata y así un análisis pronto y preciso de las vías de Bogotá.

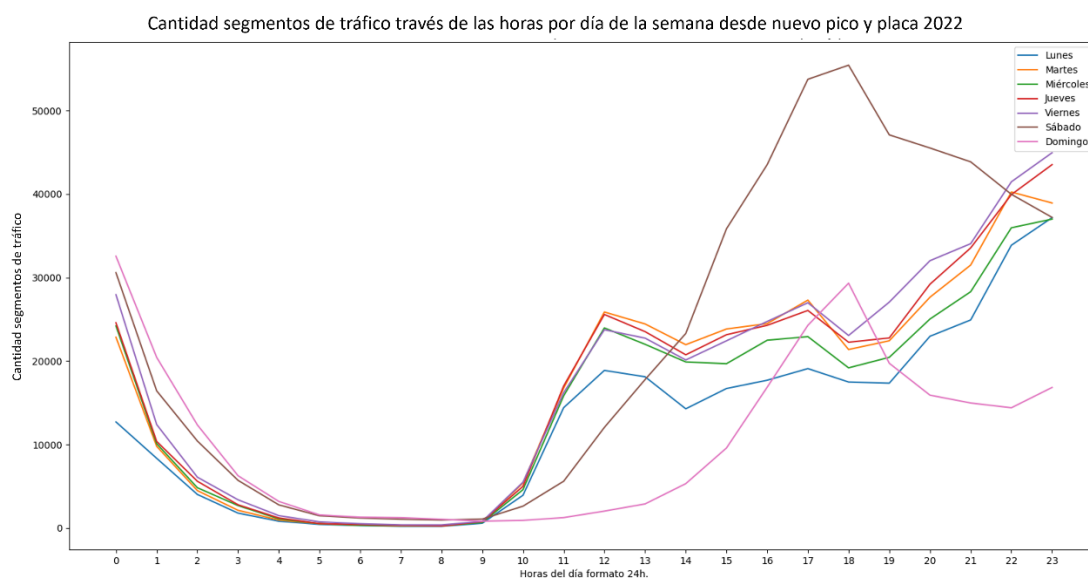
- Se resalta la importancia del uso de las redes sociales y reconocimiento como una fuente de datos importante. Por ello, mantener una extracción en tiempo real de los datos de las redes sociales nos permite realizar análisis de datos de manera oportuna.
- Generar una cultura del uso de estas redes sociales para aportar información general acerca de las vías.
- El trabajo realizado genera un aporte en la búsqueda del entendimiento del comportamiento vial, dando una base en investigaciones futuras donde se agreguen más variables de estudio que influyan en el comportamiento vial.



## A. Anexo: Análisis de tráfico con nuevo horario de pico y placa en Bogotá

Desde el 11 de enero de 2022 se reforman los horarios de pico y placa para la ciudad de Bogotá, siendo los nuevos horarios de 6:00 a.m. a 9:00 p.m [38]. Se realiza un análisis de la cantidad de segmentos de tráfico desde el 11 de enero hasta el 11 de marzo de 2022.

**Figura 54.** Cantidad segmentos de tráfico través de las horas por día de la semana desde nuevo pico y placa 2022



Fuente: Elaboración propia.

La **Figura 54** presenta que el único día con una gran cantidad de segmentos de tráfico es el sábado, teniendo los demás días de la semana un patrón similar durante las horas del día, a excepción del Domingo que mantiene menor cantidad de segmentos de tráfico. En la **Figura 44** se refleja que todos los días de la semana a excepción el Domingo mantienen un mismo patrón; pero desde que se aplicó el nuevo horario del pico y placa, que es más amplio que el anterior, se identifica que la cantidad de segmentos de tráfico disminuyeron para los días de la semana laboral.

## Bibliografía

- [1] Redacción Portal Bogotá, La movilidad también está en tus manos: cada 5,6 minutos ocurre un accidente, <https://bogota.gov.co/mi-ciudad/movilidad/analisis-de-accidentes-viales-en-bogota-en-2019>.
- [2] Ali J. Ghandour, Huda Hammoud, Mohammad Dimassi, Houssam Krayem, Jamal Haydar, Adam Issa, Allometric scaling of road accidents using social media crowd-sourced data, *Physica A: Statistical Mechanics and its Applications*, Volume 545, 2020, <http://www.sciencedirect.com/science/article/pii/S0378437119319703>
- [3] Andrienko, G., & Andrienko, N. 2008. A Visual Analytics Approach to Exploration of Large Amounts of Movement Data. *International Conference on Visual Information Systems: Web- Based Visual Information Search and Management*. [https://link.springer.com/chapter/10.1007/978-3-540-85891-1\\_1](https://link.springer.com/chapter/10.1007/978-3-540-85891-1_1)
- [4] Andrienko, G. L., Andrienko, N. V., Dykes, J., Fabrikant, S. I., and Wachowicz, M. 2008. Geovisualization of dynamics, movement and change: key issues and developing approaches in visualization research. *Information Visualization*. <http://geoanalytics.net/and/papers/ivs08a.pdf>
- [5] A Survey of Traffic Data Visualization Wei Chen, Fangzhou Guo, and Fei-Yue Wang, Fellow IEEE. [http://www.cad.zju.edu.cn/home/vagblog/VAG\\_Work/IEEEITS2015/TrafficVASurvey.pdf](http://www.cad.zju.edu.cn/home/vagblog/VAG_Work/IEEEITS2015/TrafficVASurvey.pdf)
- [6] Ester, M., peter Kriegel, H., S, J. and Xu, X., 1996. A density- based algorithm for discovering clusters in large spatial databases with noise. *International Conference on Knowledge Discovery and Data Mining*, <https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf>
- [7] Piringer, Harald & Buchetics, Matthias & Benedik, Rudolf. (2012). AIVis: Situation awareness in the surveillance of road tunnels. *IEEE Conference on Visual Analytics Science and Technology 2012, VAST 2012 - Proceedings*. 153-162.

- 10.1109/VAST.2012.6400556.  
[https://www.researchgate.net/publication/261483062\\_AIVis\\_Situation\\_awareness\\_in\\_the\\_surveillance\\_of\\_road\\_tunnels](https://www.researchgate.net/publication/261483062_AIVis_Situation_awareness_in_the_surveillance_of_road_tunnels)
- [8] Abdullah, E., Emam, A. Traffic accidents analyzer using big data (2016) Proceedings - 2015 International Conference on Computational Science and Computational Intelligence, CSCI 2015, <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84964561145&doi=10.1109%2fCSCI.2015.187&partnerID=40&md5=e1909c89f81870b9fd5c8c435eb43c5a>
- [9] Agarwal, Amit & Toshniwal, Durga. (2019). Face off: Travel habits, Road conditions and Traffic city characteristics bared using Twitter. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2917159.  
[https://www.researchgate.net/publication/333121821\\_Face\\_off\\_Travel\\_habits\\_Road\\_conditions\\_and\\_Traffic\\_city\\_characteristics\\_bared\\_using\\_Twitter](https://www.researchgate.net/publication/333121821_Face_off_Travel_habits_Road_conditions_and_Traffic_city_characteristics_bared_using_Twitter)
- [10] Md Sharikur Rahman, Mohamed Abdel-Aty, Samiul Hasan, Qing Cai. Applying machine learning approaches to analyze the vulnerable road-users' crashes at statewide traffic analysis zones. <https://doi.org/10.1016/j.jsr.2019.04.008>
- [11] Ghandour, A.J., Hammoud, H., Telesca, L. Transportation hazard spatial analysis using crowd-sourced social network data (2019) Physica A: Statistical Mechanics and its Applications, <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85060328987&doi=10.1016%2fj.physa.2019.01.025&partnerID=40&md5=b0443e086122d7c04ab4367e84e29871>
- [12] Pan, D., Zhang, W., Hamdar, S.H. From traffic analysis to real-time control: A hazard-based model to estimate postcollision recovery periods (2019) International Conference on Transportation and Development 2019: Smarter and Safer Mobility and Cities - Selected Papers from the International Conference on Transportation and Development 2019, <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85073015365&doi=10.1061%2f9780784482575.011&partnerID=40&md5=f0f6cd530e3c2bb2e8fe383236c2116f>
- [13] Ryan Mitchell, Web Scraping with Python, Second edition, [https://books.google.com.co/books?hl=es&lr=&id=TYtSDwAAQBAJ&oi=fnd&pg=PT30&dq=web+scraping&ots=y1v3AEkpfh&sig=WGuQtDVi9tsfuRBKsT1mTpIK88o&redir\\_esc=y#v=onepage&q=web%20scraping&f=false](https://books.google.com.co/books?hl=es&lr=&id=TYtSDwAAQBAJ&oi=fnd&pg=PT30&dq=web+scraping&ots=y1v3AEkpfh&sig=WGuQtDVi9tsfuRBKsT1mTpIK88o&redir_esc=y#v=onepage&q=web%20scraping&f=false)

- [14] P. Compieta, S. Di Martino, M. Bertolotto, F. Ferrucci, T. Kechadi, Exploratory spatio-temporal data mining and visualization, <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.227.7878&rep=rep1&type=pdf>
- [15] Yen-Liang Chen, Ching-Cheng Shen, Mining generalized knowledge from ordered data through attribute-oriented induction techniques, 2004, <http://ccc.inaoep.mx/~villasen/bib/mining%20generalized%20knowledge.pdf>
- [16] Andrés-Felipe Gil Torres, Aura Liliana Monroy García, Juan Sebastián González Sanabria, Minería de datos espacial en la agricultura en Latinoamérica - Una aproximación conceptual, 2020, [https://revistas.uptc.edu.co/index.php/pensamiento\\_accion/article/view/10976/9268](https://revistas.uptc.edu.co/index.php/pensamiento_accion/article/view/10976/9268).
- [17] Carlos Andrés Herrera Parra, Minería de Datos Espacial, 2006, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.130.5194&rep=rep1&type=pdf>.
- [18] Denisse Cangrejo Aljure, Juan G. Agudelo, Spatial data mining – An overview, 2011, <https://www.redalyc.org/pdf/1331/133122679009.pdf>.
- [19] E.W.T Ngai, Li Xiu, D.C.K. Chau, Application of data mining techniques in customer relationship management: A literature review and classification, 2008, <https://cmappublic3.ihmc.us/rid=1MSYC3Z3W-1B2W04K-15MY/DM-usage.pdf>.
- [20] J. M Kraak. and F. Oberling, 1996. Cartography: Visualization of Spatial Data, Longman, Essex, U.K.
- [21] Dikes, M. Kraak and M. Maceachren, 2007. Exploring Geovisualization, International Cartographic Association, Elsevier, <https://www.elsevier.com/books/exploring-geovisualization/dykes/978-0-08-044531-1>.
- [22] Meisel, Stephan & Mattfeld, Dirk. (2007). Synergies of Data Mining and Operations Research. European Journal of Operational Research - EJOR. 206. 56 - 56. 10.1109/HICSS.2007.510. [https://www.researchgate.net/publication/224686819\\_Synergies\\_of\\_Data\\_Mining\\_and\\_Operations\\_Research](https://www.researchgate.net/publication/224686819_Synergies_of_Data_Mining_and_Operations_Research).
- [23] QGIS Team, QGIS - El SIG Líder de Código Abierto para Escritorio, <https://www.qgis.org/es/site/about/index.html>.

- [24] Google Support, Cómo funciona Waze, 2020, <https://support.google.com/waze/answer/6078702?hl=es>.
- [25] Andrew Renninger, Dhruvi Kothari, Lufeng Lin, Sagari Datta, WAZE: Congestion Predictive Study, [WAZE: Congestion Predictive Study \(pennmusa.github.io\)](https://pennmusa.github.io/WAZE-Congestion-Predictive-Study/).
- [26] MapTiler, Tiles à la Google Maps Coordinates, Tile Bounds and Projection, <https://www.maptiler.com/google-maps-coordinates-tile-bounds-projection>.
- [27] geduldig, TwitterAPI, <https://github.com/geduldig/TwitterAPI>.
- [28] Twitter Developer Platform, Authentication, <https://developer.twitter.com/en/docs/authentication/oauth-1-0a/obtaining-user-access-tokens>.
- [29] Microsoft, Azure Functions documentation, <https://docs.microsoft.com/en-us/azure/azure-functions/>.
- [30] Md Kamaruzzaman, Top 10 Databases to Use in 2021, <https://towardsdatascience.com/top-10-databases-to-use-in-2021-d7e6a85402ba>
- [31] Sarthak Agarwal, KS Rajan. Performance Analysis of MongoDB Vs. PostGIS/PostgreSQL Databases For Line Intersection and Point Containment Spatial Queries, 2015. <https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1088&context=foss4g>
- [32] ArcGIS, Análisis estadístico. <https://desktop.arcgis.com/es/arcmap/10.3/analyze/commonly-used-tools/statistical-analysis.htm>
- [33] Laboratorio Urbano de Bogotá, Polígonos Localidades <https://bogota-laburbano.opendatasoft.com/explore/dataset/poligonos-localidades/export/?flg=es&location=9,4.2841,-74.21816&basemap=jawg.streets&dataChart=eyJxdWVyaWVzIjpbeyJjb25maWciOnsiZGF0YXNldCI6InBvbGlnb25vcy1sb2NhbGlkYWRIcyIsIm9wdGlbnMiOnsiZmxnljoiZXMiOisiMmNoYXN0cyI6W3siYWxpZ25Nb250aCI6dHJ1ZSwidHlwZSI6ImNvbHVtbiIsImZ1bmMiOiJDT1VOVCI6InNjaWVudGlmaWNEaXNwbGF5Ijpb0cnVILCJjb2xvcil6IiMwMDczN0MifV0sInhBeGlzIjoiTm9tYnJIIGRIIGxhIGxvY2FsaWRhZCI6Im1heHBvaW50cyI6N0TAsInNvcnQiOiIlifV0sInRpbWVzY2FsZSI6IiIsImRpc3BsYXIMZWdlbmQiOnRydWUslmFsaWduTW9udGqiOnRydWV9>

- [34] DINAS, Simena and BANON, José M. A literature review of bounding volumes hierarchy focused on collision detection.  
[http://www.scielo.org.co/scielo.php?script=sci\\_abstract&pid=S0123-30332015000100005](http://www.scielo.org.co/scielo.php?script=sci_abstract&pid=S0123-30332015000100005)
- [35] Secretaria Distrital de Movilidad, Anuario de Siniestralidad Vial de Bogotá 2020.  
<https://datos.movilidadbogota.gov.co/documents/movilidadbogota::anuario-de-siniestralidad-vial-de-bogota-2020/about?appid=62a37554ad1042e2ba86d47cb62a4a1b&edit=true>
- [36] Portafolio, Siniestros viales le cuestan al país 23,9 billones de pesos al año.  
<https://www.portafolio.co/economia/a-octubre-en-colombia-fallecieron-4-156-personas-en-siniestros-viales-546657>
- [37] Alcaldía Mayor de Bogotá, DECRETO No. 169 DE (JULIO 12 DE 2020).  
<https://secretariageneral.gov.co/sites/default/files/archivos-adjuntos/decreto-169-unificado-aislamiento-y-medidas-adicionales.pdf>
- [38] Secretaria de Movilidad de Bogotá,  
[https://www.movilidadbogota.gov.co/web/pico\\_y\\_placa\\_2022](https://www.movilidadbogota.gov.co/web/pico_y_placa_2022)