

# Agentes salvavidas: prediciendo accidentes fatales en Bogotá D.C<sup>\*</sup>

Daniel Lasso  
Gabriela Mejía  
Matteo Rozo

## Abstract

Bogotá D.C es una de las ciudades más grandes de América Latina, cuya expansión ha traído una serie de problemáticas que afectan el bienestar de sus habitantes, y se han convertido en un reto para sus dirigentes. Entre ellas, se encuentran los siniestros viales, responsables de múltiples defunciones en la ciudad (Bogotá Cómo Vamos, 2022) y objeto de estudio de la presente investigación. En este sentido, el trabajo buscará encontrar los puntos de mayor riesgo en la ciudad a través de la aplicación de *Ordinary Kriging*, una metodología de aprendizaje de máquinas de predicción espacial. También, la investigación buscará entender los principales factores que determinan la fatalidad de un accidente en un contexto de deterioro vial agravado, como en Bogotá D.C. Para ello, se realizará un análisis de los [Datos Abiertos de la Secretaría de Movilidad](#). Teniendo esto en cuenta, se da evidencia con modelos de clasificación, como *bosques aleatorios* y *XGBoost*, que las condiciones de la vía juegan un rol muy importante a la hora de predecir la gravedad del accidente, siendo este un hallazgo que no se ha estudiado a profundidad en la literatura.

**Palabras claves:** Bogotá D.C, siniestros viales, aprendizaje de máquinas, predicción espacial, modelos de clasificación.

La información y datos necesarios para replicar los resultados de esta investigación pueden encontrarse en el siguiente [repositorio](#)

---

<sup>\*</sup>Los autores corresponden respectivamente a: Investigador Jr., Área de Minería de Datos y Modelos Económicos Quantil; Auxiliar de Proyectos de Investigación, Centro de Estudios para el Desarrollo Económico (CEDE); Asistente de Investigación, Centro de Estudios para el Desarrollo Económico (CEDE)

# 1. Introducción

En los últimos años, Bogotá D.C ha sido una de la ciudades que más se ha expandido y desarrollado en términos urbanos en toda América Latina. No obstante, su evolución ha sido enmarcada por la poca armonía que existe en su sistema transporte, hasta el punto de ser motivo de constantes críticas por la ciudadanía por las deficiencias en la calidad de sus vías (Bogotá Cómo Vamos, 2022). A pesar de los esfuerzos de la Alcaldía de Bogotá y la Secretaria Distrital de Movilidad, todavía la situación es crítica en lo que respecta a la mortalidad en las vías. Específicamente, la figura 6 ofrece un oscuro panorama frente a la seguridad en las vías, ya que para algunos días del periodo 2015-2022 se reportaron hasta 8 muertes al día por accidentes de tránsito. Históricamente, para hacerle frente a esta problemática el gobierno ha establecido estrategias como los controles salvavidas (retenes de tránsito), las cámaras salvavidas (radares de velocidad), decretos para reducir la velocidad máxima en las vías y la asignación de personal para el patrullaje y control vial (Bogotá Cómo Vamos, 2022).

En línea con esta última estrategia, en febrero de 2021, la ciudad adoptó la [Resolución 6781 del 20 de diciembre de 2019](#) del Ministerio de Defensa, la cual estipula el gradual reemplazo de las funciones de tránsito desarrolladas por la Policía Nacional hacia agentes de tránsito civiles encargados de fungir como autoridad en vía. Esta política contará con un total de 1000 agentes, los cuales se incorporarán en tres fases que se extiende hasta mediados del 2023, de las cuales ya se han ejecutado las primeras dos. Dentro de las funciones de estos agentes, se destacan las siguientes:

- Programar, distribuir, coordinar y hacer seguimiento al grupo de control en vía en los diferentes corredores viales de la ciudad con el propósito de lograr un equilibrio entre la movilidad y la seguridad vial.
- Regular y controlar el flujo vehicular y peatonal, mediante funciones preventivas, de asistencia técnica y vigilancia con el objetivo de mejorar la movilidad y la seguridad vial en la ciudad de Bogotá.

De tal forma, una distribución eficiente de estos agentes es clave para prevenir los siniestros viales y sus costos asociados para la ciudad. Para ello, se hace necesario entender los determinantes de los accidentes de tránsito y qué elementos hacen más probables que estos accidentes resulten en víctimas fatales. En este orden de ideas, la presente investigación buscará responder a la siguiente pregunta: ¿Cómo se deben asignar los recursos de movilidad, como son los agentes de tránsito, para la prevención de accidentes mortales en Bogotá? Para dar respuesta a esta pregunta, en la presente investigación se hará uso de técnicas de aprendizaje de máquinas con el fin de; primero, realizar predicciones espaciales que permitan identificar los puntos calientes de accidentes de tránsito - fatales y no fatales- en la ciudad y otorgar recomendaciones de política acerca de la distribución espacial de los recursos físicos. Segundo, poder utilizar la información vial e individual de cada incidente para entender cuáles son los factores más determinantes para predecir los accidentes fatales y así poder determinar acciones de política pública que pueden ayudar a mitigar aquellos factores de riesgo de la fatalidad.

Antes de adentrarnos a las particularidades del estudio, es importante recalcar que hay una gran literatura sobre la predicción de la severidad de accidentes viales y su concentración en puntos de las ciudades aplicando modelos de aprendizaje de máquinas. Por un lado, trabajos como los de Delen, et al. (2017) y Mafi, et al. (2018) se enfocan en entender los factores más importantes que explican la severidad de incidentes automovilísticos. Estas investigaciones encuentran similares resultados con lo que respecta a la importancia de ciertas variables. Para mencionar algunas, está la edad y el género del conductor, las condiciones meteorológicas del día del incidente, el tipo y peso del automóvil, el sistema de seguridad del vehículo y el consumo de drogas por parte de los conductores. De tal forma, nuestro trabajo brindará una fuente de contraste de los hallazgos de otros países como también para probar otras alternativas que no se han estudiado dentro de la literatura.

En lo que respecta al análisis espacial de los accidentes viales fatales, unos primeros estudios se han llevado a cabo, dentro de los cuales se destaca el artículo de Atumo, Jiang & Chuanyun Fu. (2021). Estos autores buscan identificar y predecir los puntos calientes de accidentes de tránsito en la interestatal de Michigan en Estados Unidos. Dentro de algunos de los elementos que se destaca en su estudio es contemplar los posibles efectos de autocorrelaciones espaciales y las agrupaciones que afectan la predicción de la concentración de los incidentes en zonas de la ciudad, lo cual será clave para el desarrollo de nuestra investigación.

Por último, frente a estudios a nivel local sobre Bogotá D.C, Varela et al. (2018) utilizan datos de la Secretaría de Movilidad de Bogotá durante 2007-2016 para hacer un análisis descriptivo de los accidentes en la ciudad. De esta forma, la gran contribución de nuestro trabajo es que una primera aproximación al problema de predicción de accidentes viales fatales con modelos de aprendizaje de máquinas, además, es el primero en tratar de identificar estas predicciones de forma espacial.

Siendo así, el presente estudio encuentra una mapa de distribución de los puntos de mayor fatalidad predicha (ver Grafico 3) a partir de la metodología de *Ordinary Kriging* y realiza una comparación de diferentes modelos de clasificación (modelo de predicción lineal, *elastic net*, *bosques aleatorios* y *XGBoost*) con el fin de observar aquel que sea más preciso en predecir accidentes fatales, donde se aprecia que la condición de las vías es uno de los factores más importantes para predecir estos hechos.

## 2. Datos

Para realizar el modelo de predicción de accidentes fatales serán utilizados los datos históricos georreferenciados de siniestros viales en la ciudad de Bogotá D.C. Estos datos son obtenidos a partir del portal de [Datos Abiertos de la Secretaría de Movilidad](#). Las distintas bases de datos que se encuentran disponibles nos permiten obtener datos de cada siniestro ocurrido, su ubicación, los vehículos involucrados, las personas fallecidas, lesionadas y otros actores viales así como algunas características sobre la causa del accidente y el estado registrado de la vía dentro del formulario único del siniestro. Además buscamos aproximarnos a otros factores que puedan explicar la fatalidad de accidentes; relacionados con las condiciones climáticas de los días de los accidentes fatales; así como características espaciales asociadas a la ubicación de los accidentes. Para esto se realizó el pegue de las bases de datos a través de un número único de formulario que diligencia la autoridad y a través de la georreferenciación del accidente. De esta manera fue obtenida una base de datos que permite conocer a nivel de accidente-involucrado que contiene:

- Gravedad de los siniestros para 2021. Aquí se catalogan los siniestros según su nivel de gravedad, entre ellos, daños materiales, heridos y muertos.
- Causa, donde se especifica una de las 89 distintas causas atribuibles al accidente entre los que se encuentran: estado de embriaguez, desobedecer señales, fallas en los frenos, exceso de velocidad, huecos, entre otras.
- Tipo de causa, donde se especifica de manera más general una categorización de las distintas causas: Si están asociadas al conductor, al vehículo, a la vía, al pasajero o al peatón.
- Condiciones climáticas, asociadas a la humedad, visibilidad, precipitación, sensación térmica promedio del día en el que sucedió el accidente.

Por otra parte, para identificar el estado de las vías utilizamos como variable proxy el [puntaje de priorización de vías](#) de la Unidad de Mantenimiento Vial. En este, la alcaldía a través de un modelo que tiene en cuenta la presencia de huecos, los defectos en las losetas de concreto y fallas estructurales de las vías; establece un puntaje de priorización de la misma, donde 0 representa una vía en perfecto estado y 100 una vía que requiere fuertemente priorización. Adicional a esto, utilizando esta misma fuente de datos, recopilamos información espacial sobre otras

características observables como la ubicación de [paraderos del SITP](#) , [cámaras de control vial](#), [semáforos de la ciudad](#), y [ciclorrutas](#). De la misma forma se hizo uso de la plataforma Open Street Map (OSM) para extraer información sobre los restaurantes de Bogotá D.C, esto para evidenciar una proxy de actividad económica que permita identificar el flujo e interacciones de personas.

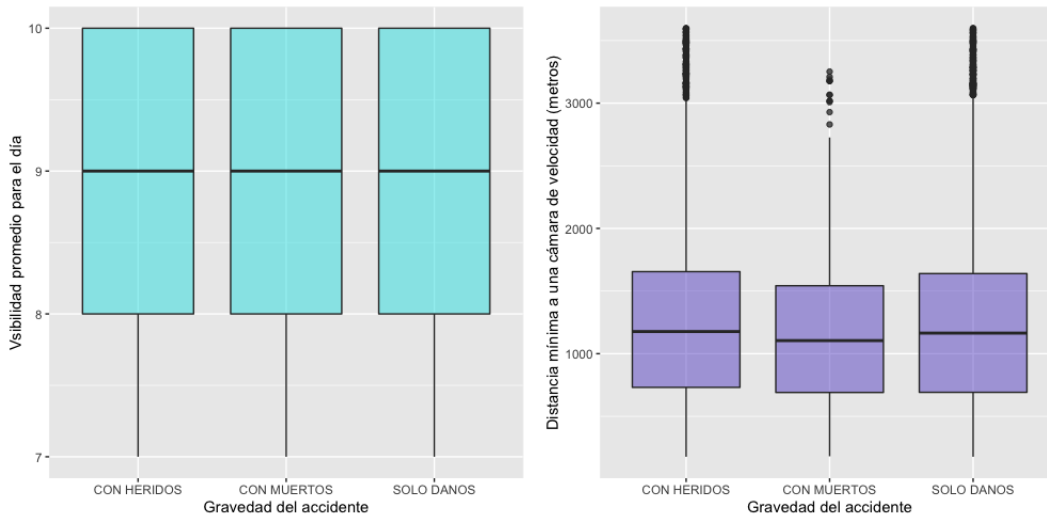
Por último, vale la pena aclarar que para obtener la información sobre las condiciones climáticas de los accidentes, se realizó un proceso de minería de datos con la [API](#) de la entidad *World Weather Online* (WWO), la cual tiene información a nivel día y hora de variables climatológicas de varias ciudades del mundo, incluyendo Bogotá D.C. Dado que la base del portal de [Datos Abiertos de la Secretaría de Movilidad](#) contaba con información sobre el momento de los accidentes, se logró unir ambas fuentes, por lo cual se comprende un rango muy completo de las condiciones que determinan la fatalidad del incidente vial.

A continuación, serán brevemente descritos algunos hechos estilizados que pueden orientar algunas determinaciones metodológicas y otorgar algunos *insights* sobre el comportamiento de los datos y los accidentes.

## 2.1. Hechos estilizados\*

- Sólo el 1.5 % de los accidentes de tránsito fueron fatales. Mientras que el 61 % tuvieron únicamente daños materiales y el 37 % tuvieron heridos (Figura 8). Esto tiene implicaciones metodológicas en materia de balanceo muestral que serán descritas en la sección 3.
- No parecen existir diferencias significativas en condiciones climáticas como la visibilidad; o en condiciones espaciales como la distancia a las cámaras salvavidas en la gravedad de los accidentes viales.

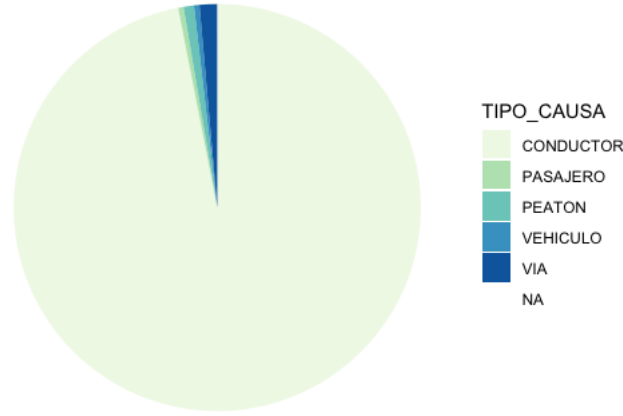
Figura 1: gravedad del accidente según visibilidad y distancia a cámaras salvavidas



- Los accidentes fatales son en su gran mayoría causados por el conductor (96 %); mientras que el 1.4 % son causados por las condiciones de la vía, 0.8 % por el peatón, y otro 0.8 % conjuntamente entre el pasajero o fallas del vehículo (Figura 2). Estas proporciones se mantienen a nivel de los accidentes no fatales, pues los accidentes fatales y no fatales se distribuyen a nivel de sus causas y ubicación de manera similar.

\*Estadísticas descriptivas detalladas pueden encontrarse anexas en los cuadros 1 y 2

Figura 2: Distribución de causas de accidentes fatales



### 3. Metodología

Como se menciona en secciones anteriores, la aproximación metodológica de este trabajo se compone de dos etapas, primero, se realizará una predicción de dónde ocurren accidentes graves (fatales o con heridos) y segundo, se determinará cuáles son las causas o determinantes mayormente asociados a los accidentes fatales. Es decir, se busca explorar si son las causas asociadas con el comportamiento de los conductores, como el exceso de velocidad o el estado de embriaguez; o si son las causas externas, por ejemplo, la calidad de la vía, las condiciones metereológicas o las fallas mecánicas de los vehículos. Para esto, se desarrollarán dos metodologías separadas que son postuladas a continuación:

#### 3.1. Predicción espacial de los accidentes graves

En primera instancia, para lograr predecir donde van a ocurrir estos accidentes graves se recurrió a la metodología llamada *Ordinary Kriging* la cual se usa para predecir el valor de una variable sobre un espacio continuo. A diferencia de otros métodos estadísticos espaciales, esta utiliza la autocorrelación espacial entre los puntos muestrados para interpolar los valores en el espacio. En general, esta metodología puede verse como una predicción en dos etapas: primero, se ajusta un variograma que determina la covarianza entre los puntos muestrales. Segundo, se usan estas covarianzas entre los datos para ajustar unos pesos que permiten interpolar valores que no se encuentran en la muestra en el espacio determinado.

Siendo así, el modelo corresponde a un método de minimización del error medio al cuadrado de las predicciones espaciales, donde las predicciones son una interpolación sujeta a una ponderación de los puntos observados. Para desarrollar más esta idea, matemáticamente, se busca minimizar la siguiente función:

$$L(Z(B), \hat{Z}(B)) = \sum_{i=1}^n (Z(B_i) - \hat{Z}(B_i))^2 \quad (1)$$

Donde  $Z(B)$  corresponde a la variable observada en el punto en la ubicación específica del espacio  $B$  y  $\hat{Z}(B)$  representa la predicción de la variable en ese espacio  $B$ . Ahora, para obtener la predicción  $\hat{Z}(B)$  esta se define como

una interpolación sujeta a algunos pesos que varían según la estimación, pero, en general está dada por:

$$\hat{Z}(B_i) = \sum_{i=1}^n \lambda_i Z(B_i) \quad t.q. \quad \sum_{i=1}^n w_\alpha = 1 \quad (2)$$

Donde  $\lambda_i = 1$  son los pesos de la interpolación lineal y el resto de las variables son las descritas antes.

Para encontrar los  $\lambda_i$  óptimos que minimizan el error de predicción, se utiliza la correlación entre las observaciones. En particular, esto corresponde a la primera etapa. Como se mencionó se ajusta un variograma, el cual establece la relación entre la covarianza (semivarianza) entre puntos muestrales (usualmente graficado en el eje Y) y los rezagos o distancia entre los mismos (graficado en el eje X), un ejemplo de esta se puede observar en la figura 7. La matriz de covarianza de este método se usa para calcular los pesos  $\lambda_i$ , con diferentes formas funcionales según el método puntual de estimación: Ordinary, Robust, Block, etc. Así mismo, La principal ventaja de utilizar este modelo es que es mucho más robusto y predice mejor en escenarios de alta correlación entre las variables, como se puede observar en la figura 7, donde los accidentes graves en Bogotá D.C están fuertemente correlacionados. Además, bajo algunos supuestos, e incluso en presencia de autocorrelación, los valores interpolados serán los mejores estimadores lineales insesgados (MELI) (Auchincloss, 2007). Del mismo modo, se puede estimar de forma lineal con una variable dicótoma, por lo que en realidad los resultados deben interpretarse como una probabilidad.

### 3.2. Predicción de accidentes fatales y sus determinantes

Para la segunda estrategia, se realizan modelos de clasificación de la presencia de víctimas fatales de un accidente de tránsito con el fin de identificar los principales factores que conlleva este tipo de eventos. Antes de entrar en detalle de los modelos utilizados, cabe recordar que existe un gran desbalanceo de la muestra, donde una reducida porción son accidentes mortales. En este orden de ideas, se procedió a realizar una comparativa de los resultados con la muestra normal y con un remuestreo que logra aumentar el número de observaciones de accidentes fatales haciendo uso de técnicas de *oversampling*. Bajo este contexto, las métricas de medición que se tendrán en cuenta para calibrar y derrimir el desempeño de los modelos será una combinación del *accuracy*, una medida del porcentaje de predicciones correctas, y el *f-score*, una métrica que busca balancear la medida de *precision* y *recall* en un solo puntaje, tal que se valore más las predicciones correctas de accidentes fatales. Por último, es clave aclarar que se realizó una separación aleatoria de la muestra en dos subconjuntos: una de entrenamiento (70%) y una de prueba (30%). Esto será útil para ver el poder predictivo de los modelos por fuera de muestra y evitar problemas de sobre-ajuste sobre la muestra de entrenamiento.

En lo que respecta al apartado de modelos, se realizó una estimación por 4 metodologías: el modelo de predicción lineal, *elastic net*, bosques aleatorios y XGBoost, utilizando como insumo las variables que se encuentran en las estadísticas descriptivas y los cuadros 1 y 2. De forma breve, se presenta cada uno de ellos.

El modelo de predicción lineal construye sus predicciones a partir de la estimación de una ecuación de parámetros lineales que minimizan el error cuadrático medio. Este tiene varias deficiencias que pueden llevar a una gran varianza en el modelo y un sobre-ajuste de los resultados a la muestra de entrenamiento.

El *elastic net* es una combinación de modelos de regularización de *ridge* y *lasso*, teniendo como objetivo la reducción de dimensionalidad de las variables independientes. Como resultado, al comparar con el modelo lineal, sus medidas de varianza son mucho más reducidas, por lo cual se espera encontrar resultados mucho menos ajustados. Aún así, sigue sufriendo del problema de linealidad en los parámetros, lo que dificulta su ajuste a funciones complejas.

Los bosques aleatorios extienden mucho más su campo de análisis, al explorar las no linealidades e interacciones que pueden existir entre las variables, por cual resulta ideal en un contexto de múltiples variables categóricas.

Para entender esta técnica, primero se debe comprender el funcionamiento de árboles de decisión, donde cada partición brindada por una variable nos genera ramificaciones o hojas de predicción, donde la impureza de cada una nos determina la capacidad predictiva del modelo. Ahora, por problemas de varianza elevado y sobre-ajuste, los bosques aleatorios realizan selecciones aleatorias de predictores para evitar la alta correlación entre diferentes árboles de decisión.

Por último. XGBoost es una metodología que refina y optimiza aquellos pasos realizados por los bosques aleatorios, donde se hace mucho más importante las ganancias en predicción que tiene cada partición generada por cada variable, reduciendo así también la dimensionalidad de los árboles. Para ello, se realiza una búsqueda de grilla de aquellos hiperparámetros que mejoran la predicción del modelo, incluyendo el número de árboles, la profundidad, entre otros.

## 4. Resultados Preliminares

### 4.1. Modelo de predicción espacial de accidentes graves

La figura 7 (ver Anexo) corresponde al variograma del modelo ajustado por *Ordinary Kriging*. En esta, se puede identificar que hay una fuerte correlación entre los accidentes sobre una región del espacio. Además, sus semivarianzas están agrupadas entre 0.20 y 0.25, lo que indica que la correlación no varía fuertemente conforme se separan espacialmente las observaciones. Por otra parte, es importante mencionar que el modelo es bueno ajustando los valores hasta una distancia de 10 kilómetros (aproximadamente), a partir del cual todavía hay correlación entre los accidentes, pero el modelo no lo captura correctamente.

Adicionalmente, la figura 3 presenta las predicciones realizadas con sus respectivas probabilidades de que ocurra un accidente grave (lesionados o muertos). Como se puede evidenciar el modelo logra interpolar correctamente hacia zonas de las cuales no presentaba información de accidentes previos. Por ejemplo, hacia la vía a La Calera, El Codito, y las salidas de Bogotá como la calle 80 y la 13. En estas predice una probabilidad moderadamente alta de que ocurra un accidente. Por otro lado, se identifican zonas neurálgicas con alta probabilidad de accidentes graves (en rojo) donde se podría priorizar la presencia de los guardas de movilidad, así como se identifican lugares con baja probabilidad (en azul) donde su permanencia puede no ser tan eficiente. Un resultado interesante es que la mayor parte de las zonas con probabilidad predicha alta se encuentra hacia el suroriente de la ciudad, donde hay zonas industriales y focos de actividad económica.

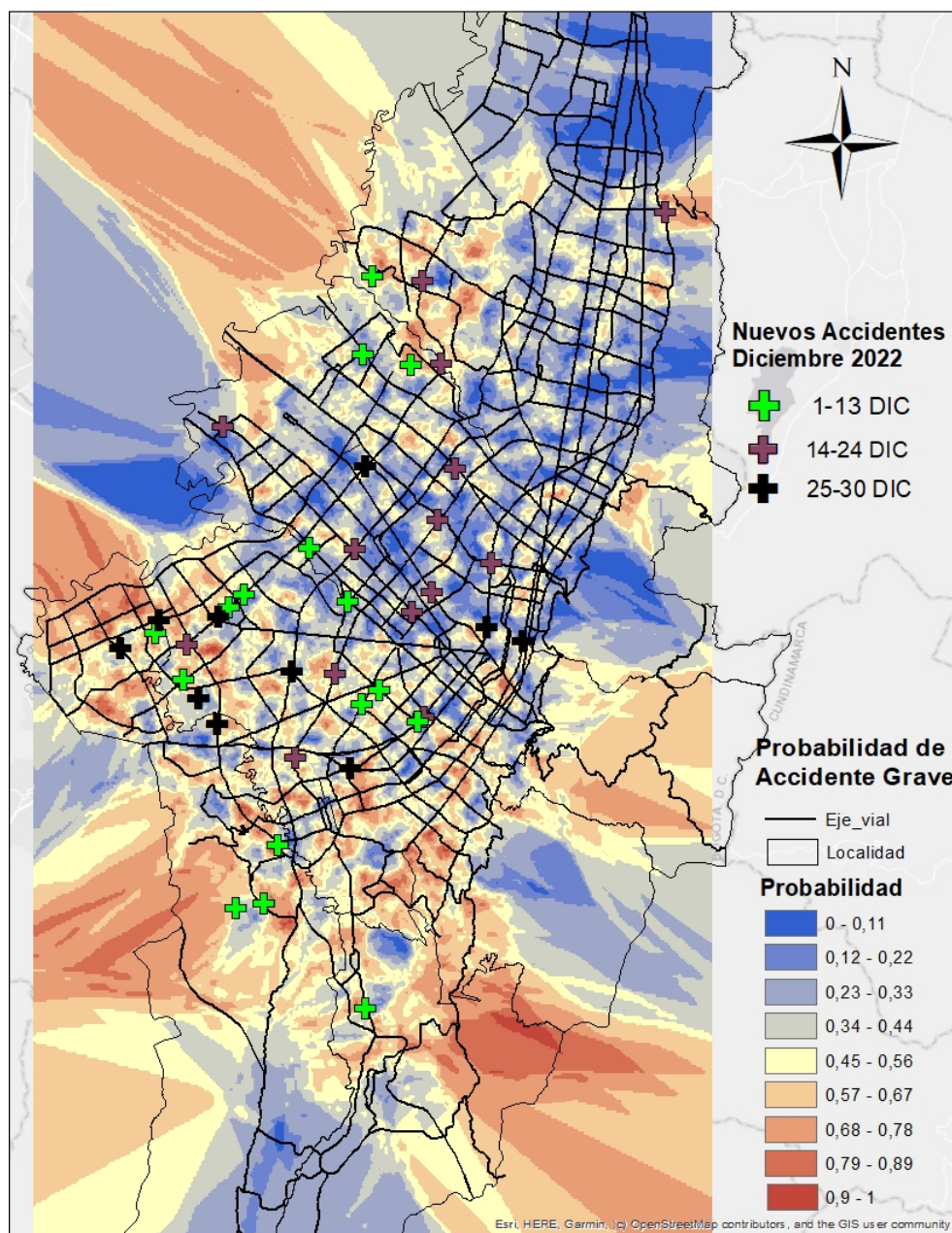
Análogamente, encontramos que la Avenida Las Américas y la Autopista Sur son las calzadas con mayor probabilidad de existencia de accidentes graves. Esto puede ser explicado por el alto flujo de peatones en estas vías, hecho que será verificado en la siguiente sección. No obstante, exceptuando estas dos rutas, se encuentra que los accidentes graves no están fuertemente correlacionados con las vías principales, sino que esta probabilidad es mayor en las secundarias o terciarias. Por ejemplo, la Autopista Norte y la Avenida el Dorado tienen una probabilidad de accidente grave muy baja, lo cual puede estar explicado por la presencia de doble calzada, la calidad de las vías y la señalización de las mismas.

Por otra parte, y para dar cuenta de que el modelo propuesto funciona, se recopiló de la página de Movilidad Bogotá la información de los accidentes de tránsito fatales ocurridos en Bogotá entre el 1 y 30 de diciembre del 2022 y se graficaron en la figura 3 con una cruz de color verde, café y negro según la semana del mes correspondiente. Como se puede observar de los 38 accidentes fatales ocurridos en esa semana, 33 de estos ocurrieron en zonas sobre las que el modelo predecía una alta probabilidad de ocurrencia de accidentes de tránsito fatales. Esto evidencia una precisión de aproximadamente el 90 % en la predicción. En suma, con el presente modelo se le ofrece a la Alcaldía



de Bogotá un criterio de priorización de los Agentes de Movilidad para lograr prevenir este desastrozo evento<sup>2</sup>.

Figura 3: Predicción espacial de los accidentes graves



#### 4.2. Modelo de predicción y determinantes de los accidentes fatales

Los resultados se resumen en la tabla 4 que contiene las principales métricas de predicción, como también el tipo de muestro realizado. Partiendo de los resultados con remuestreo tipo *oversampling*, se puede apreciar que las medidas del modelo de predicción lineal y *elastic net* tienen valores muy altos de *specificity* y valores muy bajos de *accuracy* dentro de muestra. Estos resultados se vuelven más evidentes en las predicciones fuera de muestra, por lo

<sup>2</sup>Esta estimación se realizó a través de *Ordinary Kriging* con un radio máximo de interpolación de 5 Km y con un máximo de puntos cercanos a la grilla de 100, a través del software ArcGis con la herramienta Kriging



cual esto puede estar explicado por un sobre-ajuste de la muestra de entrenamiento. Con respecto a las predicciones de bosques aleatorios y XGBoost, es clara las mejoras en las predicciones tanto dentro, como fuera de muestra, donde ambos alcanzan un accuracy del 99% por fuera de muestra. Pero, hay leves mejoras en el  $F1$ , siendo esto representativo de la optimización que realiza el XGBoost.

Frente a los resultados sin remuestreo, se puede apreciar que las medidas del modelo de predicción lineal y *elastic net* alcanzan valores de *accuracy* cercanos al 98%. También, es interesante apreciar que no se estiman algunas de las métricas. Estos comportamientos se deben a que ambos modelos predicen que todos los valores son 0, es decir, que no se producen accidentes fatales. Esto porque el desbalance de las clases genera que el modelo obtenga muy buenas predicciones solamente sesgándose hacia una alternativa, pero no es ideal para identificar los factores más importantes. En contraposición, los resultados de bosques aleatorios y XGBoost muestran una clara consistencia con lo mencionado en el remuestreo, donde todavía el  $F1$  por fuera de muestra de XGBoost indica que este es el modelo más apropiado para predecir los siniestros viales. Esto posiblemente tenga que ver con las fuertes no linealidades en los datos, pues el exceso de velocidad puede ser muy diferente con precipitaciones altas que bajas, o en una autopista vs una vía secundaria, elementos que ambos modelos manejan adecuadamente<sup>3</sup>.

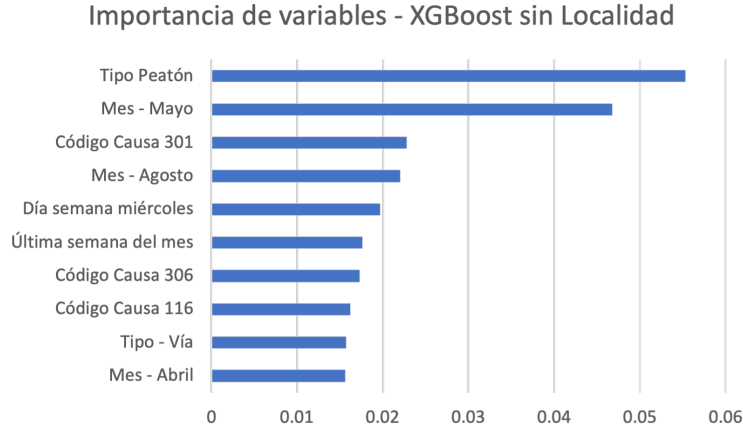
Figura 4: Comparativa de métricas entre modelos de clasificación

Modelo	Muestreo	Evaluación	Sensitivity	Specificity	FPR	Accuracy	Precision	Recall	F1
Linear Model	SMOTE - Oversampling	Dentro de muestra	0.00	1.00	0.00	0.50	0.50	1.00	0.67
Linear Model	SMOTE - Oversampling	Fuera de muestra	0.00	1.00	0.00	0.02	0.02	1.00	0.03
Elastic Net	SMOTE - Oversampling	Dentro de muestra	0.86	0.76	0.24	0.81	0.85	0.76	0.80
Elastic Net	SMOTE - Oversampling	Fuera de muestra	0.86	0.59	0.41	0.86	0.07	0.59	0.12
RF	SMOTE - Oversampling	Dentro de muestra	1.00	1.00	0.00	1.00	1.00	1.00	1.00
RF	SMOTE - Oversampling	Fuera de muestra	1.00	0.24	0.76	0.99	0.96	0.24	0.39
XGBoost	SMOTE - Oversampling	Dentro de muestra	1.00	1.00	0.00	1.00	1.00	1.00	1.00
XGBoost	SMOTE - Oversampling	Fuera de muestra	1.00	0.34	0.66	0.99	0.59	0.34	0.43
Linear Model	Normal	Dentro de muestra	1.00	0.00	1.00	0.99	NA	NA	NA
Linear Model	Normal	Fuera de muestra	1.00	0.00	1.00	0.98	NA	NA	NA
Elastic Net	Normal	Dentro de muestra	1.00	0.00	1.00	0.99	NA	NA	NA
Elastic Net	Normal	Fuera de muestra	1.00	0.00	1.00	0.98	NA	NA	NA
RF	Normal	Dentro de muestra	1.00	1.00	0.00	1.00	1.00	1.00	1.00
RF	Normal	Fuera de muestra	1.00	0.35	0.65	0.99	0.99	0.35	0.51
XGBoost	Normal	Dentro de muestra	1.00	1.00	0.00	1.00	1.00	1.00	1.00
XGBoost	Normal	Fuera de muestra	1.00	0.51	0.49	0.99	0.89	0.51	0.65

En este orden de ideas se presenta en la tabla 5 las variables más importantes para la predicción. Por fuera de aquellas medidas temporales como el mes y el día de la semana, resulta interesante apreciar que las variables sobre las condiciones del accidente son las más importantes, incluyendo algunos códigos de causa relacionados con el estado de las vías.

<sup>3</sup>Se encontró que los mejores hiperparámetros son los siguientes: una tasa de aprendizaje de 0.03, una profundidad máxima de 12 nodos y un cantidad de 300 árboles de decisión

Figura 5: 10 variables más importantes según XGBoost



\*Los códigos 301, 306 y 116 corresponden a Falta de Señales, Huecos y Exceso de velocidad respectivamente

En particular, se hace énfasis que aquellas vías que no tienen un buen señalamiento y tienen huecos son más probables a tener accidentes fatales, por lo cual las recomendaciones que se hace sobre la asignación de los agentes también debe incluir un enfoque sobre aquellos segmentos de vías que tienen estas deficiencias. Otro resultado a resaltar es que las variables de condiciones climáticas y del vehículo no son tan relevantes, por lo cual esto representa un punto de inflexión con respecto a la literatura. Para ser más preciso, los resultados encontrados por las investigaciones de Delen, et al. (2017) y Mafi, et al. (2018) tienen limitaciones de extropolación a ciudades con infraestructura del nivel de Bogotá D.C, por lo cual el presente estudio brinda un diagnóstico alternativo a la problemática de siniestros viales.

## 5. Conclusiones y Recomendaciones de Política

En este trabajo se elaboró un modelo de predicción espacial que identifica dónde ocurren accidentes graves. Esto provee a la Secretaría de Movilidad de Bogotá una herramienta que le permita hacer uso eficiente de los recursos escasos -como lo son los agentes de tránsito- para lograr prevenirlos. Por otra parte, se utilizó aprendizaje de máquinas para predecir, con base en características internas y externas de los individuos, los determinantes que establecen que los accidentes de tránsito sean fatales. Con este análisis se encuentra que los factores más importantes son la presencia de peatones, los huecos en la vía, el exceso de velocidad, la falta de señales y el día de la semana. En este sentido, la recomendación de política es clara, primero existen formas de asignación eficiente de los agentes de movilidad para salvaguardar la vida de los bogotanos, pues estos deberían ser asignados a las zonas que se establece una alta probabilidad de accidentes de tránsito graves. Segundo, es importante realizar controles y estrategias para la mejora de la malla vial y la infraestructura alrededor de las vías, donde la señalización y la reparación de huecos juega un rol esencial en la protección del los actores viales, siendo el peatón aquel que se enfrenta a los mayores riesgos.

## 6. Anexo

Figura 6: Histórico de datos sobre accidentes viales

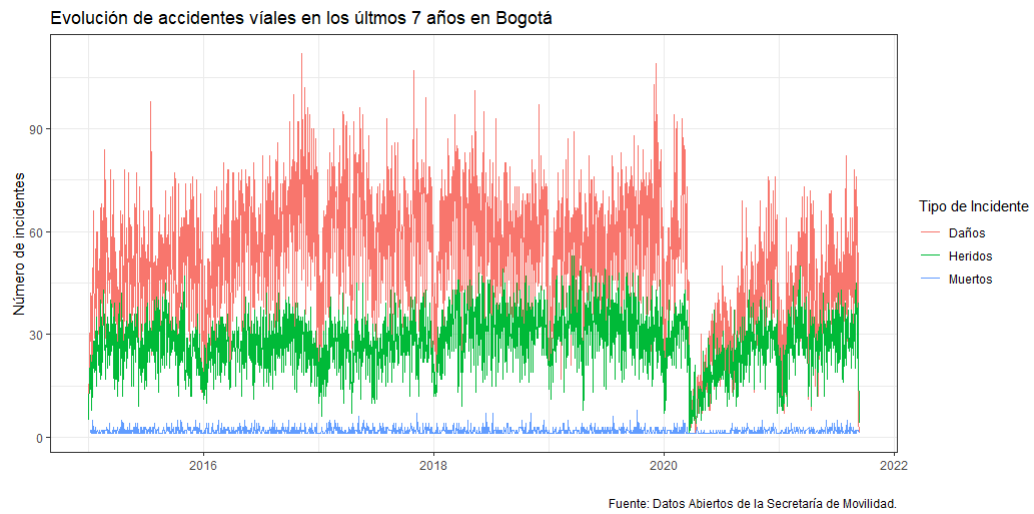


Figura 7: Variograma del modelo de predicción espacial estimado por *Ordinary Kriging*

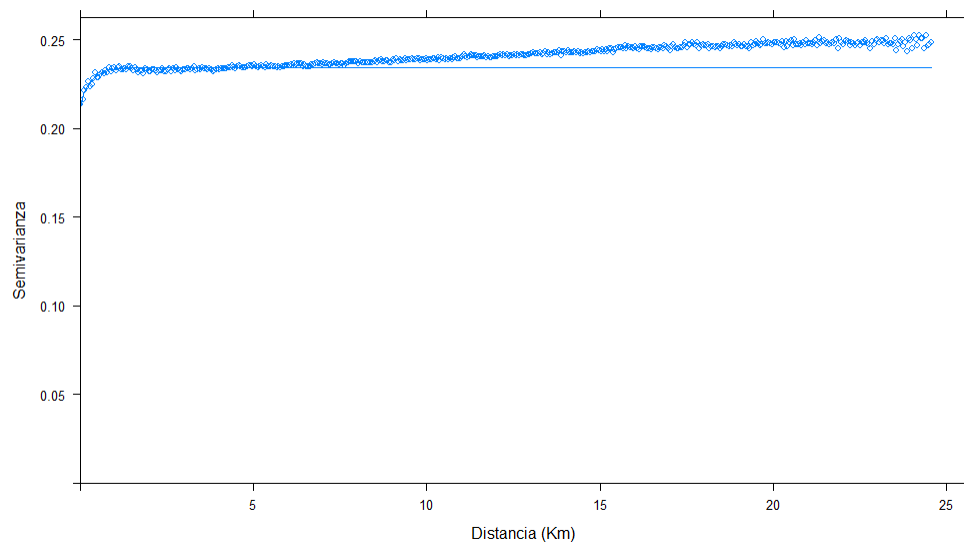
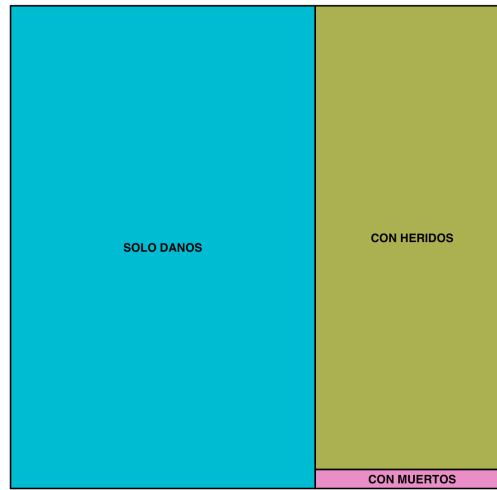


Figura 8: Distribución de siniestros según gravedad



Cuadro 1: Estadísticas Descriptivas de Base de Datos

	count	mean	Var	sd	min	max
Ancho de calzada	44320	7.495013	6.195152	2.489006	.83	51.76
Área de calzada	44320	777.3617	766603.6	875.559	14.91	20100.8
Cobertura de nubes	44320	62.8174	119.8844	10.94917	25	82
Temperatura de derretimiento (C)	44320	11.14673	1.263089	1.123872	8	13
Sensación Térmica (C)	44320	13.80566	1.453715	1.205701	10	16
Índice de calor	44320	13.90487	1.198785	1.09489	11	17
Humedad	44320	84.90957	19.30033	4.393215	69	96
Índice calidad de vías	44320	39.52593	156.5606	12.51242	12.64	100
Distanica media a restaurantes	44320	9638.744	6547074	2558.725	6236.153	53161.18
Distancia mínima a cicloruta	44320	246.615	312517.8	559.0329	.0245916	33226.2
Distancia mínima a cámara	44320	1431.959	1978041	1406.428	1.649927	42392.56
Distancia mínima a semafóro	44320	274.6566	345151.8	587.4962	.6353	34072.99
Distancia mínima a estación de SITP	44320	101.8087	73513.05	271.1329	.6309039	24182.21
Temparatura mínima (C)	44320	10.76816	2.116223	1.454724	7	13
Índice de rayos UV	44320	3.611146	.2376519	.4874955	3	4
Grado de viento	44320	144.4636	1040.858	32.26233	87	288
Ráfaga de viento (kmph)	44320	6.37157	3.026489	1.739681	3	10
Velocidad de viento (kmph)	44320	4.477956	1.556183	1.247471	2	8
Temperatura Vientos (C)	44320	13.80566	1.453715	1.205701	10	16
N	44320					

Cuadro 2: Causas de siniestros viales para 2021

	(1) Causa	
	Número	Porcentaje
ADELANTAR CERRANDO	5995	13.52662
ADELANTAR EN CURVA	59	.1331227
ADELANTAR EN ZONA PROHIBIDA	190	.4287004
ADELANTAR INVADIENDO CARRIL DEL MISMO SENTIDO EN ZIG-ZAG	261	.5888989
ADELANTAR INVADIENDO VIA	887	2.001354
ADELANTAR POR LA DERECHA	378	.8528881
ANIMALES EN LA VIA	2	.0045126
APROVISIONAMIENTO INDEBIDO	2	.0045126
ARRANCAR SIN PRECAUCION	477	1.076264
AUSENCIA O DEFICIENCIA DE DEMARCACION	32	.0722022
AUSENCIA TOTAL O PARCIAL DE SEÑALES	53	.1195848
CARGA SOBRESALIENTE SIN AUTORIZACION	7	.0157942
CARGA SOBRESALIENTE SIN SEÑALES	6	.0135379
CARGUE O DESCARGUE PROHIBIDO	9	.0203069
CIRCULAR POR CALZADAS DESTINADAS A BUSES Y BUSETAS	46	.1037906
CRUZAR EN CURVA	1	.0022563
CRUZAR EN DIAGONAL	3	.006769
CRUZAR EN ESTADO DE EMBRIAGUEZ	12	.0270758
CRUZAR SIN OBSERVAR	91	.2053249
DEFECTOS FISICOS Y PSIQUICOS	6	.0135379
DEJAR O MOVILIZAR SEMOVIENTES EN LA VIA	5	.0112816
DEJAR O RECOGER PASAJEROS EN SITIOS NO DEMARCADOS	8	.0180505
DESCENDER O SUBIR DEL VEHICULO EN MARCHA	3	.006769
DESOBEDECER AL AGENTE	14	.0315884
DESOBEDECER SEÑALES	4441	10.02031
EMBRIAGUEZ APARENTE	346	.7806859
EMBRIAGUEZ O DROGA	484	1.092058
ESTACIONAR SIN SEGURIDAD	71	.1601986
EXCESO DE VELOCIDAD	102	.2301444
EXCESO EN HORAS DE CONDUCCION	22	.049639
FALLA EN EL SISTEMA ELECTRICO	3	.006769
FALLA EN LAS PUERTAS	7	.0157942
FALLAS EN DIRECCION	15	.0338448
FALLAS EN LAS LLANTAS	36	.0812274
FALLAS EN LOS FRENOS	108	.2436823
FALLAS EN LUCES DE FRENOS	1	.0022563
FALTA DE MANTENIMIENTO MECANICO	24	.0541516
FALTA DE PRECAUCION POR NIEBLA, LLUVIA O HUMO	23	.0518953
FALTA DE SEÑALES EN VEHICULO VARADO	19	.04287
FRENAR BRUSCAMENTE	214	.482852
HUECOS	259	.5843863
IMPARTIR ENSEÑANZA AUTOMOVILISTICA SIN AUTORIZACION	5	.0112816
JUGAR EN LA VIA	1	.0022563
NIÑOS EN EL ASIENTO DELANTERO	1	.0022563
NO CAMBIAR LUCES	8	.0180505
NO HACER USO DE SEÑALES REFLECTIVAS	66	.148917
NO MANTENER DISTANCIA DE SEGURIDAD	10065	22.70984
NO PORTAR ESPEJOS	3	.006769
NO RESPETAR PRELACION DE INTERSECCIONES O GIROS	417	.9408845
NO RESPETAR PRELACIÓN	751	1.694495
NO SUJETAR LOS MANUBRIOS	9	.0203069
NO UTILIZAR - GIRAR BRUSCAMENTE	1443	3.255866
NO UTILIZAR - IMPERICIA EN EL MANEJO	468	1.055957
NO UTILIZAR - SALIRSE DE LA CALZADA	43	.0970217
OBSTACULO EN LA VIA	17	.0383574
OTRA	10033	22.6363
PARARSE SOBRE LA CALZADA	13	.0293321
PASAJERO EMBRIAGADO	1	.0022563
PASAJEROS OBSTRUYENDO EL CONDUCTOR	4	.0090253
PASAR SEMAFORO EN ROJO	14	.0315884
PONER EN MARCHA UN VEHICULO SIN PRECAUCIONES	568	1.281588
REALIZAR GIRO EN U"	32	.0722022
RECOGER O DEJAR PASAJEROS SOBRE LA CALZADA	35	.0789711
REPARAR VEHICULO EN VIA PUBLICA	2	.0045126
REVERSO IMPRUDENTE	1188	2.680505
SALIR POR DELANTE DE UN VEHICULO	10	.0225632
SEMÁFORO EN ROJO	1722	3.885379
SIN ESTABLECER	5	.0112816
SUBIRSE AL ANDEN O VIAS PEATONALES	42	.0947653
SUJETARSE A OTRO VEHÍCULO	2	.0045126
SUPERFICIE HUMEDA	92	.2075812
SUPERFICIE LISA	42	.0947653
TRANSITAR CON LAS PUERTAS ABIERTAS	34	.0767148
TRANSITAR DISTANTE DE LA ACERA U ORILLA DE LA CALZ	63	.142148
TRANSITAR EN CONTRAVIA	743	1.676444
TRANSITAR ENTRE VEHICULOS	764	1.723827
TRANSITAR POR LA CALZADA	47	.1060469
TRANSITAR POR VIAS PROHIBIDAS	68	.1534296
TRANSITAR SIN LOS DISPOSITIVOS LUMINOSOS DE DETENCION	2	.0045126
TRANSITAR SIN LUCES	5	.0112816
TRANSITAR UNO AL LADO DEL OTRO	33	.0744585
TRANSPORTAR PASAJEROS EN VEHICULOS DE CARGA	4	.0090253
TRANSPORTAR OTRA PERSONA O COSAS	48	.1083032
TRANSPORTE DE CARGA SIN SEGURIDAD	11	.0248195
VEHICULO MAL ESTACIONADO	674	1.520758
Total	44320	100
N	44320	

## Referencias

- Bivand, R., Pebesma, E., y Gómez-Rubio, V. (2008). Applied spatial data analysis with r, ch. 8, interpolation and geostatistics.
- Bogotá Cómo Vamos. (2022). Informe de calidad de vida en bogotá 2021. Descargado de <https://bogotacomovamos.org/informe-calidad-vida-2021/>
- Delen, D., Tomak, L., Topuz, K., y Eryarsoy, E. (2017). Investigating injury severity risk factors in automobile crashes with predictive analytics and sensitivity analysis methods. *Journal of Transport & Health*, 4, 118–131.
- Mafi, S., Abdelrazig, Y., y Doczy, R. (2018). Machine learning methods to analyze injury severity of drivers from different age and gender groups. *Transportation research record*, 2672(38), 171–183.
- Sameen, M. I., y Pradhan, B. (2017). Severity prediction of traffic accidents with recurrent neural networks. *Applied Sciences*, 7(6), 476.
- Varela, S. E. M., Díaz, H., Urbano, E., Sánchez, J., y Quintero, I. (2018). *Modelo de predicción de gravedad de accidentes de tránsito: un análisis de los siniestros en bogotá, colombia*. Universidad Nacional de Colombia. Descargado de <https://bit.ly/2SMOEGJ>
- Waller, L., y Gotway, C. (2008). Applied spatial data analysis with r, ch. 8, interpolation and geostatistics.
- Zheng, M., Li, T., Zhu, R., Chen, J., Ma, Z., Tang, M., . . . Wang, Z. (2019). Traffic accident’s severity prediction: A deep-learning approach-based cnn network. *IEEE Access*, 7, 39897–39910.