



UNIVERSIDAD NACIONAL  
DE COLOMBIA

INFORME BOOTCAMP COLOMBIA TECH 2

---

## Análisis de Accidentalidad en Medellín entre los años 2015-2017

---

**Autores:**

Jerónimo Hoyos, jhoyosbo@unal.edu.co

Santiago Sosa, ssosag@unal.edu.co

Manuel D. Echeverry, mdecheve@unal.edu.co

Brayan Cumbalaza, bcumbalaza@unal.edu.co

**Ejecutor:**

Luis Ángel Córdoba

**Mentor:**

Cristian Olarte

26 de marzo de 2025

# Índice

<b>1. Introducción</b>	<b>3</b>
<b>2. Análisis Exploratorio</b>	<b>4</b>
2.1. Fechas y Horas de Mayor Accidentalidad . . . . .	7
2.2. Ubicaciones con Mayor Accidentalidad . . . . .	8
<b>3. Modelos Predictivos</b>	<b>9</b>
3.1. Red Neuronal . . . . .	9
<b>4. Conclusiones</b>	<b>10</b>

## 1. Introducción

En este trabajo se realizó un análisis del conjunto [Accidentalidad Valle de Aburrá](#), obtenido del banco de datos abiertos del **Área Metropolitana del Valle de Aburra**. En la página web de esta corporación autónoma regional pueden hallarse variedad de datos, desde movilidad, hasta análisis de riesgo por contaminación del aire.

Este conjunto de datos contiene información sobre los accidentes registrados por las Secretarías de Movilidad y Transporte durante los años 2015, 2016, 2017 y 2018 en algunos municipios del Valle de Aburrá. La tabla tiene 203450 filas y 11 columnas, que son:

- COD.MUNICIPIO: código numérico para identificar cada municipio.
- MUNICIPIO: nombre del municipio donde ocurrió el accidente de tráfico.
- FECHA: en la que ocurrió el accidente.
- HORA: del día en que ocurrió el accidente.
- DÍA DE LA SEMANA: cuando ocurrió siniestro.
- CLASE: tipo de accidente.
- DIRECCIÓN: lugar del accidente.
- GRAVEDAÑOSSADAÑOSS: tipo de fatalidad.
- BARRIO: donde pasó el accidente.
- COMUNA: donde ocurrió el accidente.
- DISEÑO: tipo de vía.

El desarrollo del problema puede encontrarse en el siguiente enlace de GitHub: [repositorio](#).

## 2. Análisis Exploratorio

A partir de un primer vistazo de los datos podemos notar que algunos no aportan información que sea relevante para nuestro análisis. Primero, el `COD_MUNICIPIO` puede ser ignorado al estar trabajando únicamente con 7 municipios, y podremos identificar a los municipios por su nombre. También se eliminarán las filas con datos *nulls* y las filas duplicadas. Esto nos deja con 203385 registros.

Podemos notar algunas inconsistencias en los datos, como ejemplificamos en la siguiente gráfica

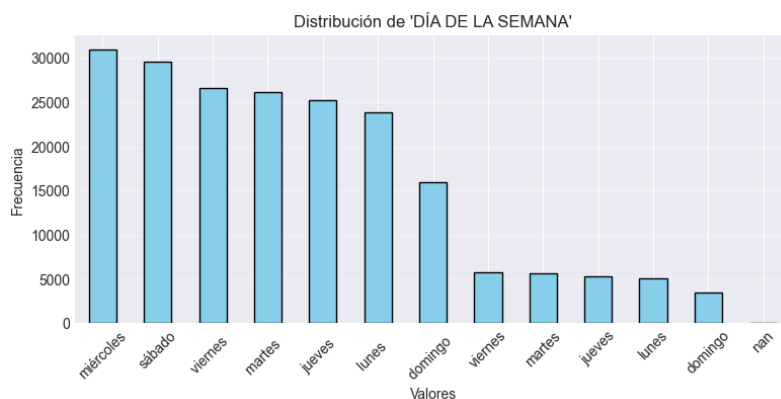


Figura 1: Histograma del Día de la Semana en que Ocurrió el Accidente.

A partir de esto, estandarizamos los datos para tener una visión adecuada de los hechos. Esto, en cada columna, para así, poder proseguir con el análisis. Finalmente, eliminaremos todas las filas en las que aparezca algún dato con alguna casilla marcada como «sin información» salvo en el caso del diseño de la vía. Los municipios de Sabaneta y Envigado no dan dicha información en la tabla.

Inicialmente, veremos cómo se distribuye la accidentalidad respecto a la variable municipio:

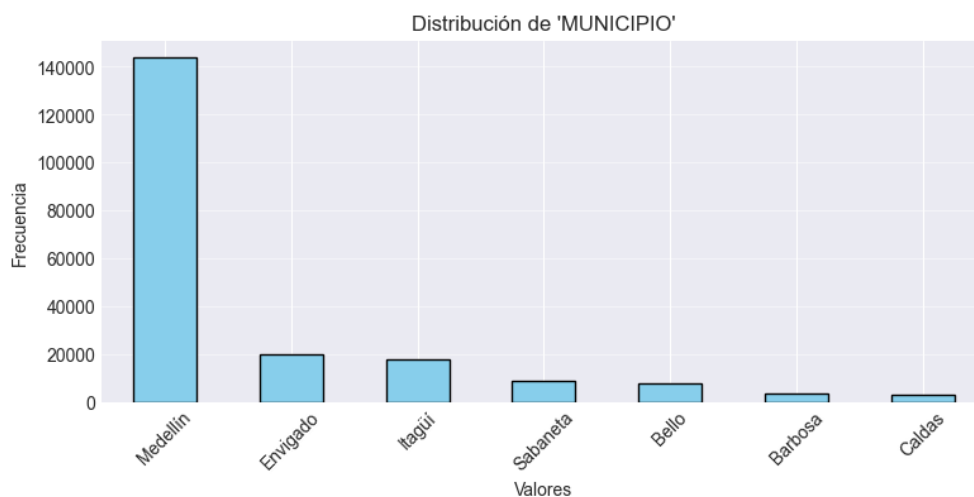
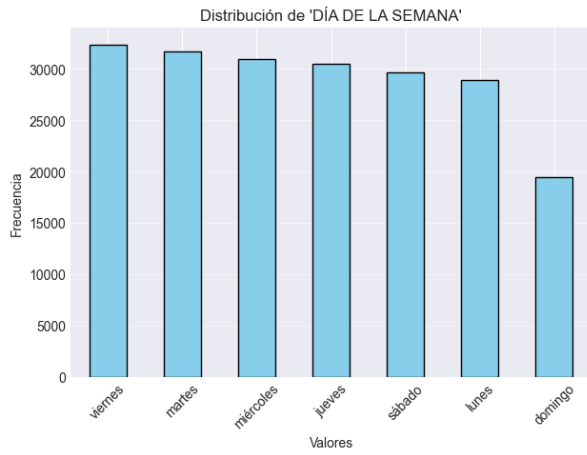


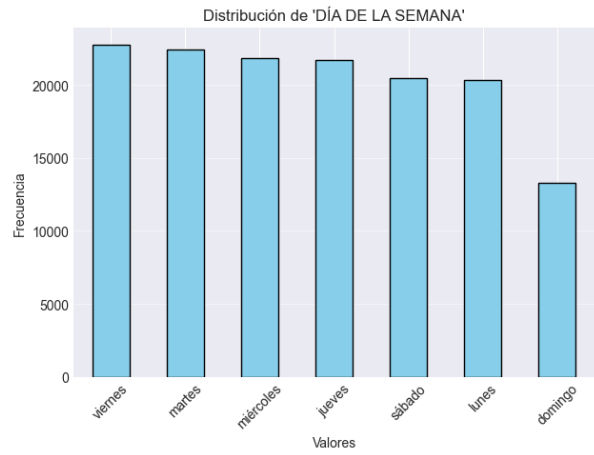
Figura 2: Distribución de la Accidentalidad por Municipio

De este primer gráfico podemos observar que el municipio de Medellín pesa sobremanera en la accidentalidad del Valle de Aburrá. Esto nos lleva a hacer una comparación entre los datos generales y los de Medellín para verificar esta apreciación.

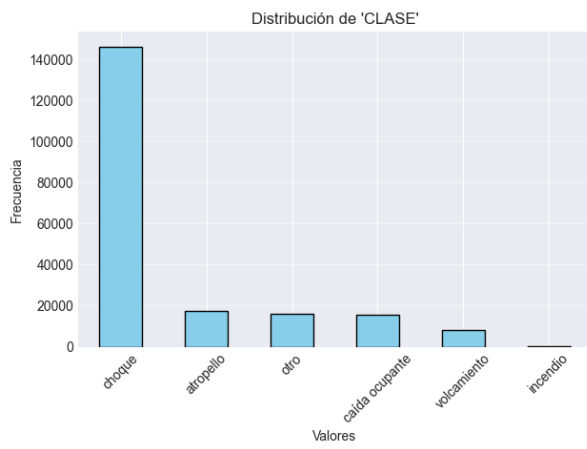
A continuación, observaremos cómo el municipio de Medellín marcaría la tendencia general de algunas de las variables. Vemos a la izquierda la distribución para el Valle de Aburrá y a la derecha para la ciudad de Medellín.



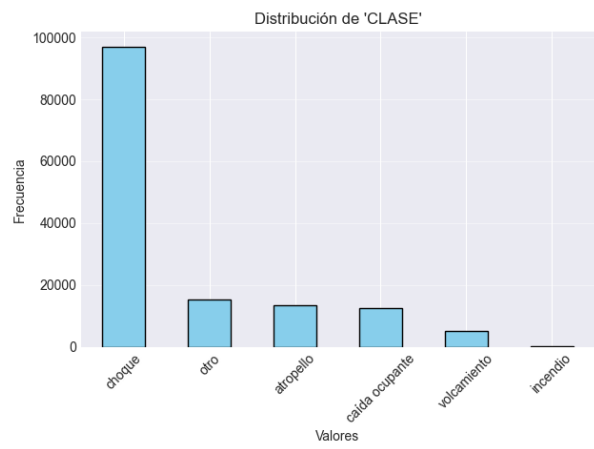
(a) Valle de Aburrá



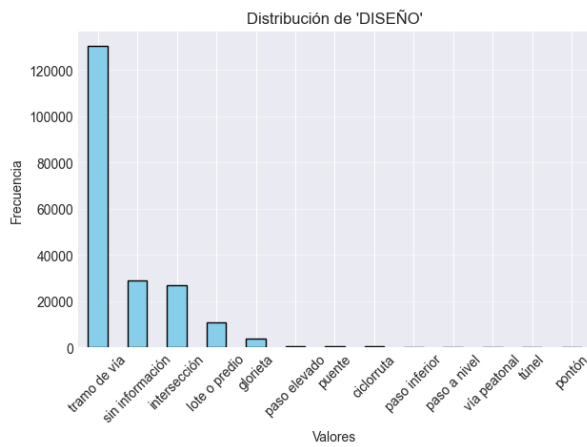
(b) Medellín



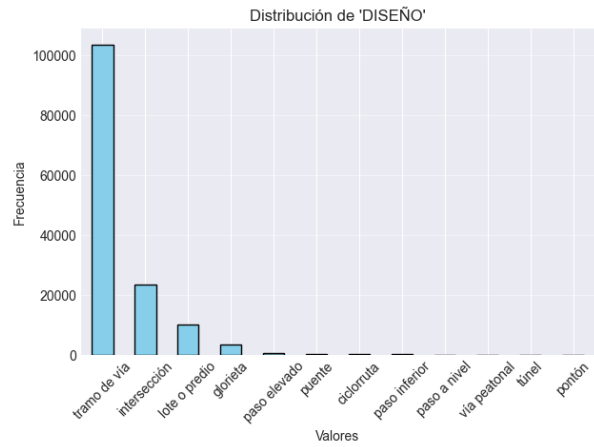
(c) Valle de Aburrá



(d) Medellín



(e) Valle de Aburrá



(f) Medellín

Figura 3: Comparación de las tendencias generales y la tendencia de la ciudad de Medellín.

Por último, podemos observar en la siguiente gráfica, un detalle del comportamiento de la accidentalidad respecto de las comunas y corregimientos de Medellín. Notemos que hay una sobre representación de La Candelaria (centro) en la accidentalidad de la ciudad.

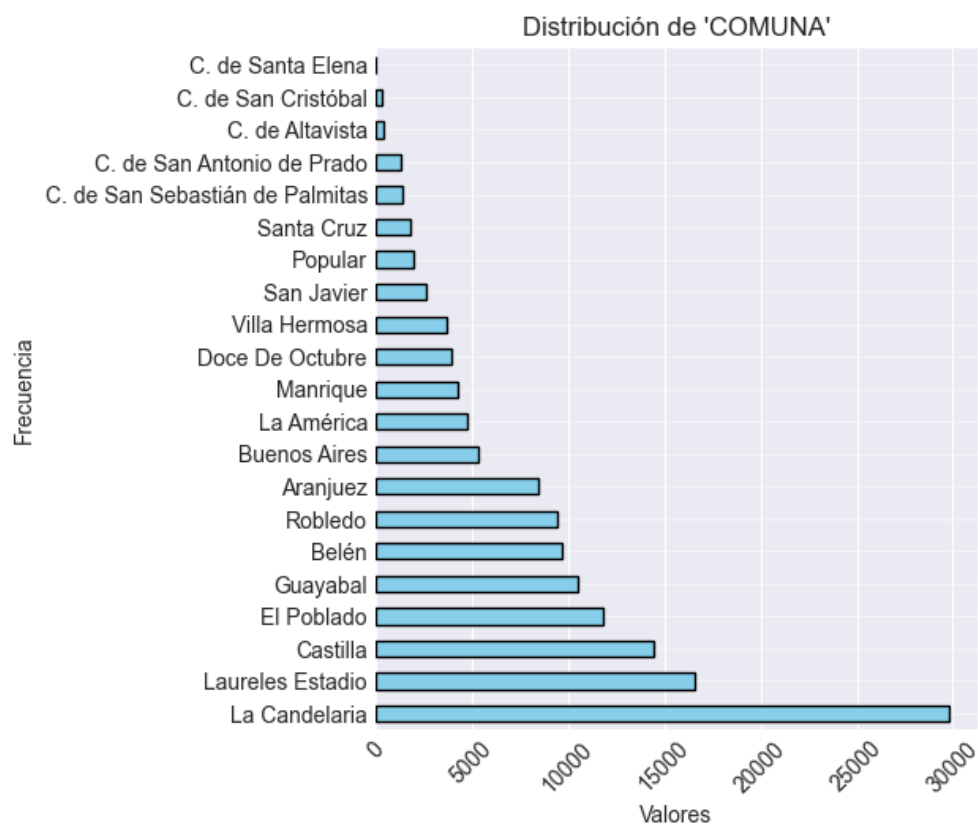


Figura 4: Distribución de la Accidentalidad por Comunas de la Ciudad de Medellín.

## 2.1. Fechas y Horas de Mayor Accidentalidad

Se notó una problemática con el formato de las fechas y las horas. Respecto a las fechas cambiaban el orden en la presentación de días, meses y año. En las horas no estaba estandarizado el uso de 24 horas, y p.m. y a.m.. El año 2018 no se tuvo en cuenta, dado que, sólo presentaba datos para los 4 primeros meses.

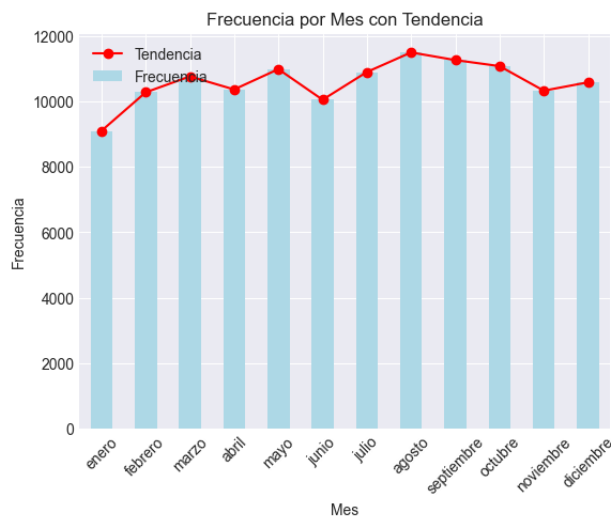


Figura 5: Distribución por Meses para los años 2015, 2016 y 2017.

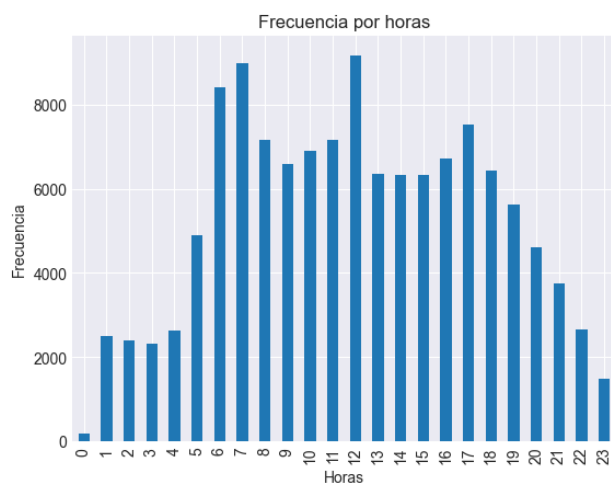


Figura 6: Distribución por horas.



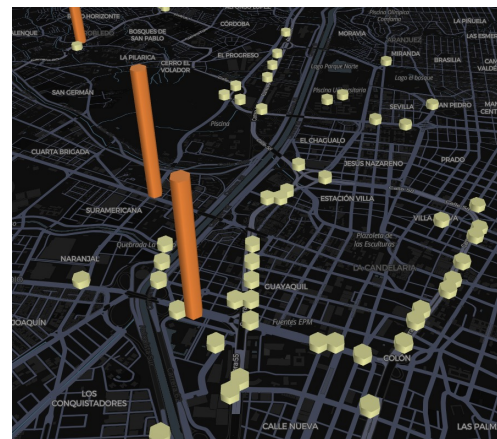
Figura 7: Visualización panorámica de la Accidentalidad en la Ciudad de Medellín.

## 2.2. Ubicaciones con Mayor Accidentalidad

Se filtraron las direcciones de las 100 ubicaciones con más accidentes de Medellín. Esto, haciendo uso de la API Google Maps, se creó un mapa interactivo en 3D mediante `streamlit`, y así, formando un mapa de calor visualizado a través de hexágonos para ilustrar las zonas.

Aquí podemos observar un detalle con algunos de dichos lugares.

En la siguiente gráfica ilustraremos el mismo hecho, pero con una vista más general de la ciudad.



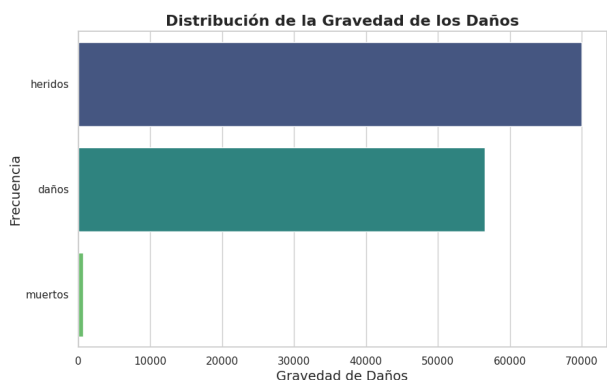


### 3. Modelos Predictivos

#### 3.1. Red Neuronal

Se entrenó una **red neuronal** con la finalidad de predecir la gravedad de los accidentes usando como variables independientes CLASE, DÍA DE LA SEMANA, MES, HORA, DISEÑO y COMUNA. Esta predice la probabilidad de que ocurra cualquiera de los tres siniestros: daños, heridos o muertos.

Como vemos en la siguiente gráfica, el número de accidentes donde ocurre un fallecimiento, es mucho menos que aquellos accidentes donde únicamente hay heridos o daños materiales.



Se tuvo en cuenta que una muerte es algo más indeseable, y por lo tanto, se le dio más peso al entrenar el modelo.

Aquí se observa la estructura de la red neuronal.

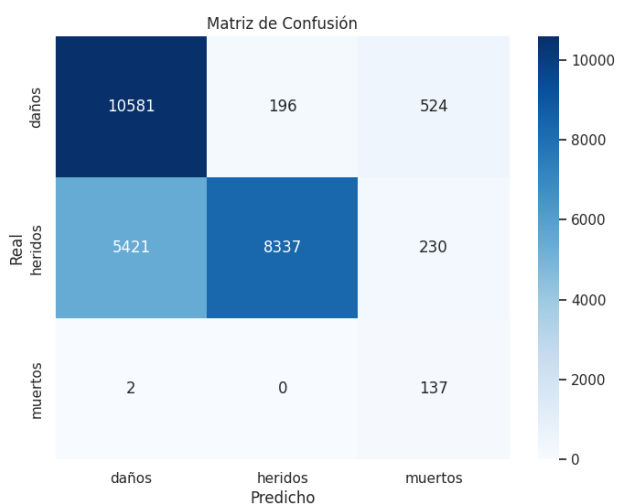
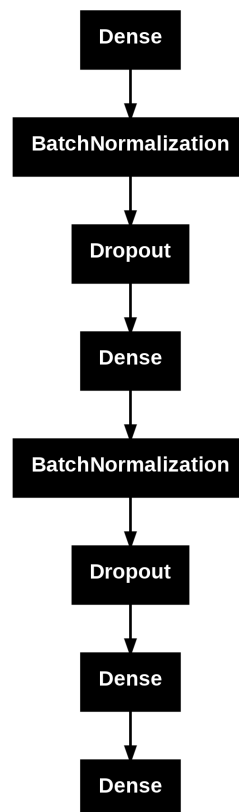


Figura 8: Matriz de Confusión.

Al haber tan pocos fallecidos respecto a los otros tipos de gravedad en los accidentes, y aún, habiéndoles dado mayor peso, el modelo no logra predecir adecuadamente la fatalidad de los siniestros.

## 4. Conclusiones

Se recomienda estandarizar la toma de los datos de la accidentalidad para mejorar la predictibilidad de los modelos que se implementen en el futuro.

La mayoría de accidentes ocurren en las horas de Sol, hay ciertas anormalidades en las horas pico, sobre todo, al mediodía.

Los meses que corresponden al verano del hemisferio norte (julio, agosto y septiembre) son los que tienen mayor accidentalidad. Esto podría deberse a la Feria de las Flores y al gran número de turistas en la ciudad.

Al parecer la accidentalidad, al ver la distribución por días, no se debe tanto a la rumba de los fines de semana, sino al afán de llegar al trabajo en los días laborables.

Es necesario construir una base de datos más robusta para poder entrenar modelos que puedan predecir las muertes, ya que, este tipo de accidentes, son los más graves. Echamos en falta datos sobre los tipos de vehículos en cada situación. Sospechamos que el uso motocicletas y bicicletas son factores a tener en cuenta para un estudio futuro.