

PROYECTO FINAL BOOTCAMP ANÁLISIS DE DATOS
NIVEL INTERMEDIO

ACCIDENTES DE TRÁNSITO EN COLOMBIA UNA REALIDAD

Presentado Por:

Julieth Andrea Bautista Tellez

Leandro Cano Castaño

Yesica Maria Montoya Puerta

Vanessa Restrepo Obando

Estiven Ospina

Ronal Ortiz - Ejecutor técnico

Gabriel Beltran - Ejecutor mentor

Javier Velasco - Monitor

CONTENIDO

Introducción	3
Objetivo del proyecto	4
Objetivo general:	4
Objetivos específicos:	4
Descripción del Proyecto:	5
1. Análisis Inicial y planteamiento del problema:	5
2. recolección de datos	5
3. Limpieza y Preprocesamiento de Datos	5
4. Análisis Exploratorio de Datos (EDA) y selección de Métodos de Análisis	7
Análisis De Resultados	9
1. Análisis de homicidios por Accidentes de Tránsito en Colombia:Insights desde el Dashboard	9
2. Análisis del modelo	11
Conclusiones	13
Documentación y enlaces del proyecto	14

INTRODUCCIÓN

En Colombia, los accidentes de tránsito representan una de las principales causas de mortalidad y afectan de manera significativa a familias y comunidades a lo largo del país. Dentro de estos eventos, aquellos que resultan en la muerte de las víctimas y que son clasificados como homicidios de tránsito deben ser un foco de atención para las políticas y campañas de seguridad vial y prevención.

Este proyecto se centra en el análisis de datos sobre homicidios en accidentes de tránsito ocurridos en Colombia durante la última década. Mediante la utilización de la base de datos suministrada por la Policía Nacional encontrado en la página datos.gov.co; en él se busca examinar factores clave como el género, el grupo etario de las víctimas, el departamento donde ocurren estos eventos y el medio de transporte involucrado. Este análisis detallado permitirá identificar patrones que ayuden a entender mejor las dinámicas y características que podrían estar asociadas con la ocurrencia de estos accidentes fatales.

El objetivo de este estudio es proporcionar una base de conocimiento clara y precisa que permita identificar perfiles de riesgo y áreas geográficas con mayor incidencia de homicidios en el tránsito. Con esta información, se pretende ofrecer recomendaciones que puedan guiar el diseño de estrategias de prevención y políticas públicas enfocadas en reducir los accidentes fatales en las vías.

OBJETIVO DEL PROYECTO

Objetivo general:

Realizar un análisis exploratorio de la base de datos de homicidios por accidentes de tránsito en Colombia, aplicando técnicas de limpieza, normalización y estadística descriptiva. Se busca identificar las tendencias y las variables más influyentes, así como construir un modelo predictivo que permita anticipar la ocurrencia de estos incidentes.

Objetivos específicos:

- Identificar las principales tendencias y correlaciones entre variables más significativas asociadas con los homicidios por transporte en Colombia.
- Construir un modelo predictivo que determine las zonas y condiciones de mayor riesgo para accidentes de tránsito en Colombia.
- Diseñar un dashboard interactivo que muestre los principales hallazgos del análisis, incluyendo al menos cinco visualizaciones clave, para facilitar tomas de decisiones a futuro.

DESCRIPCIÓN DEL PROYECTO

Este proyecto tiene como objetivo analizar los datos de homicidios en accidentes de tránsito en Colombia durante la última década, enfocándose en el uso de herramientas de análisis y visualización de datos.

Se usan datos proporcionados por la Policía Nacional de Colombia, accesibles a través de la plataforma de datos abiertos de Colombia desde el siguiente enlace: [Homicidios Accidentes de Transito](#), para examinar la relación entre variables clave y su impacto en la mortalidad vial. Este análisis no solo permitirá entender mejor las dinámicas detrás de estos eventos, sino que también servirá como una guía estructurada para procesos de toma de decisiones informadas en temas de seguridad vial.

1. Análisis inicial y planteamiento del problema:

Realizar un análisis exhaustivo de la base de datos de homicidios por accidentes de tránsito en Colombia, identificando patrones y variables más críticas que generan un mayor impacto en las vías, las posibles amenazas y vulnerabilidades, y evaluando los riesgos asociados

Documentar los resultados del análisis y la evaluación de riesgos, y establecer prioridades basadas en la criticidad de los activos y el impacto potencial de las amenazas identificadas.

2. Recolección de datos

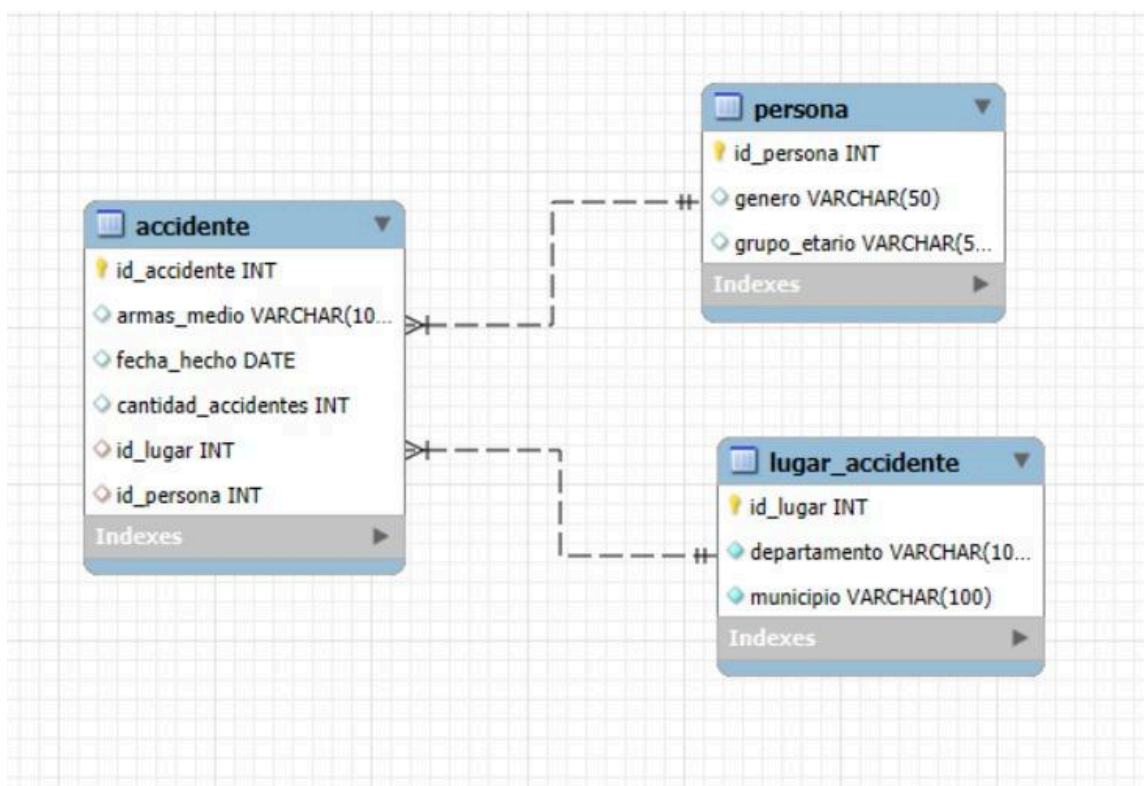
Para el análisis de homicidios en accidentes de tránsito en Colombia, la recolección de datos se basará en la base de datos pública de la Policía Nacional de Colombia, disponible en la plataforma de datos abiertos de Colombia. Esta fuente proporciona un conjunto de datos que contiene información detallada sobre los accidentes de tránsito mortales en el país, incluyendo género de la víctima, grupo etario, ubicación geográfica y medios de transporte involucrado.

3. Limpieza y preprocesamiento de Datos

Una vez descargada la base de datos desde el enlace [Homicidios Accidentes de Transito](#), se procedió a realizar una limpieza y preprocesamiento básico empleando el código que se detalla en los anexos que se encuentra en el siguiente repositorio [Repositorio Homicidios Accidentes De Transito](#)

Este proceso consistió en los siguientes pasos:

- Eliminación de duplicados: Se identificaron registros duplicados en el archivo CSV inicial, los cuales fueron eliminados para asegurar la consistencia y evitar el sesgo en los resultados.
- Normalización de nombres de columnas: Los títulos de las columnas fueron estandarizados con un formato adecuado para facilitar su transferencia a MySQL y mejorar la accesibilidad de los datos. Este formato incluyó el uso de minúsculas y la eliminación de espacios y caracteres especiales.
- Verificación de valores nulos: Se revisó la base de datos para detectar valores nulos en las variables principales. Afortunadamente, no se encontraron valores nulos, lo que permitió avanzar sin necesidad de imputación de datos.
- Carga en MySQL: Tras la limpieza y normalización, los datos fueron importados a MySQL para la estructuración final de la base de datos. La base resultante está organizada según el esquema entidad-relación que se detalla a continuación.



4. Análisis Exploratorio de Datos (EDA) y selección de Métodos de Análisis

Una vez cargados los datos en MySQL y filtrados para seleccionar la última década (desde 2014), se ajustaron las variables a analizar. Como parte de este proceso, se eliminaron dos columnas que no eran relevantes para los objetivos del proyecto, **Código DANE** y **Descripción de Conducta**, lo que simplificó el análisis y enfocó el conjunto de datos en las variables clave.

Para el análisis, se emplearon técnicas de estadística descriptiva y herramientas de visualización de datos, que permitieron explorar las tendencias y relaciones entre las variables relevantes. Este análisis incluyó:

- Análisis de tendencia en la cantidad de homicidios en accidentes de tránsito a lo largo de la década, permitiendo visualizar si existen aumentos o disminuciones en los incidentes fatales.
- Distribución de accidentes por departamento, lo que facilita la identificación de áreas geográficas con mayor incidencia de estos eventos.
- Distribución por género y categoría de edad, con el fin de detectar patrones en las víctimas en función de estos factores.
- Relación con el medio de transporte involucrado, permitiendo observar si algunos tipos de vehículos presentan un mayor riesgo de accidentes fatales.

Para llevar a cabo el EDA, se utilizaron herramientas de visualización como Tableau, que facilitó la creación de dashboards interactivos, mientras que Visual Studio Code se empleó para gestionar la carga de datos y ejecutar consultas MySQL. Este enfoque permitió identificar patrones clave y relaciones entre las variables, que ayudarán a responder los objetivos del proyecto y fundamentar futuras recomendaciones para la seguridad vial en Colombia.

Además del análisis descriptivo, se desarrolló un modelo estadístico para predecir el riesgo de accidentalidad en diferentes zonas del país. Para este fin, se empleó un Random Forest Classifier, con el cual se clasificaron las zonas en alto o bajo riesgo de accidentalidad. Este modelo predictivo se entrenó y evaluó en Google Colab, y el código se encuentra detallado en un

repositorio GitHub al cual se puede acceder mediante el siguiente enlace [Repositorio Homicidios por accidentes deTransito](#)

Durante la evaluación del modelo, se analizaron las métricas de precisión, recall y f1-score, obteniendo resultados óptimos tanto para las predicciones de alto riesgo como de bajo riesgo.

ANÁLISIS DE RESULTADOS

1. Análisis de homicidios por Accidentes de Tránsito en Colombia: Insights desde el Dashboard

Tras el análisis del dataset de accidentes de tránsito en Colombia, compuesto por un total de 22,877 registros, se destacan los siguientes hallazgos principales:

- **Departamentos con Mayor Incidencia de Accidentes Fatales:** Los departamentos con el mayor número de accidentes de tránsito que resultaron en homicidios son Cundinamarca (3,847 casos), Antioquia (3,667 casos) y Valle del Cauca (2,852 casos). Estos tres departamentos representan el 45.22% de los homicidios en accidentes de tránsito a nivel nacional.
 - Interpretación y Recomendaciones:
La elevada incidencia de accidentes en estos departamentos sugiere que son áreas críticas para la implementación de políticas públicas y campañas preventivas. Esto podría incluir medidas de seguridad vial específicas, regulación de tráfico y campañas educativas dirigidas a reducir la accidentalidad en estas zonas.
- **Tendencia Temporal de Accidentes en la Última Década:**
Al observar la evolución de los accidentes de tránsito que resultaron en homicidios en la última década, se nota una tendencia un pico en el año 2019. Sin embargo, en 2020, debido a las restricciones de movilidad impuestas por la pandemia de COVID-19, se registró una caída notable en el número de accidentes, seguida de un aumento constante en los años posteriores.
 - Interpretación y Recomendaciones:
La tendencia general hacia el incremento en accidentes fatales sugiere que se deben intensificar los esfuerzos para mejorar la seguridad vial, especialmente en un contexto de normalización de la movilidad. Es crucial continuar monitoreando estos datos para identificar picos y aplicar intervenciones oportunas.

- **Medios de Transporte Implicados en los Accidentes:** Los datos muestran que el 46.75% de los accidentes fatales involucran vehículos, seguidos por motos (27.10%) y un 17.68% de casos no reportados. Además, 7.59% corresponden a peatones, 0.78% a bicicletas y 0.10% a trenes.
 - Interpretación y Recomendaciones:
Dado que los vehículos y las motocicletas representan la mayor proporción de homicidios en accidentes, es fundamental desarrollar campañas de seguridad dirigidas a conductores de estos medios de transporte. Las iniciativas podrían incluir programas de concienciación sobre límites de velocidad y educación vial sobre el respeto a peatones y ciclistas.
- **Distribución por Categoría de Edad:** En cuanto a las edades de las víctimas, se observa que los adultos representan el 92.98% de los homicidios, seguidos por adolescentes (4.35%) y menores de edad (2.67%).
 - Interpretación y Recomendaciones:
La alta incidencia entre adultos indica que las campañas de concienciación y prevención deben enfocarse especialmente en este grupo etario. Además, podrían considerarse políticas orientadas a la seguridad laboral y al transporte en horarios nocturnos, cuando los adultos están más expuestos a situaciones de riesgo en la vía.
- **Distribución por Género:** Los datos reflejan una mayoría significativa de víctimas masculinas (80.29%), en comparación con 19.71% de víctimas femeninas.
 - Interpretación y Recomendaciones:
La mayor incidencia entre hombres sugiere que las campañas de seguridad vial podrían estar más dirigidas hacia este grupo, tomando en cuenta su exposición a factores de riesgo. Intervenciones específicas, como campañas en lugares de trabajo y educación vial en áreas con alta densidad de tránsito masculino, pueden ser beneficiosas.

Estos resultados aportan una visión integral sobre las dinámicas de homicidios en accidentes de tránsito en Colombia y destacan la necesidad de políticas públicas adaptadas a las características demográficas, geográficas y de movilidad del país. La utilización de herramientas como el

dashboard en Tableau permite una visualización clara de las tendencias y puntos críticos, lo que facilita la toma de decisiones informadas para mejorar la seguridad vial.

Para acceder al dashboard y visualizar los datos en profundidad, puede ingresar en el siguiente enlace: [Dashboard Homicidios por accidentes de transito](#)

2. Análisis Del Modelo

Inicialmente, el modelo de clasificación mostró signos de sobreajuste, manifestados en un rendimiento desequilibrado entre los datos de entrenamiento y de prueba, además de un desbalance significativo entre las clases a clasificar: 0 (bajo riesgo) y 1 (alto riesgo). Para corregir estos problemas, se implementaron ajustes con apoyo de inteligencia artificial, que incluyeron técnicas para manejar el desbalance de clases, optimizando el desempeño general y logrando un modelo más estable y aplicable.

las métricas que se usaron en la evaluación son:

- **Precisión:**
 - **Clase 0 (Bajo riesgo):** 0.99, mostrando una alta efectividad al evitar falsos positivos en esta clase.
 - **Clase 1 (Alto riesgo):** 0.79, indicando un margen de mejora, aunque el modelo sigue siendo efectivo al identificar un 79% de las predicciones de alto riesgo.
- **Exhaustividad (Recall):**
 - **Clase 0 (Bajo riesgo):** 0.92, reflejando una fuerte capacidad para detectar casos de bajo riesgo.
 - **Clase 1 (Alto riesgo):** 0.96, demostrando que el modelo identifica eficazmente los casos críticos de alto riesgo.
- **F1-Score:**
 - **Clase 0:** 0.95, manteniendo un balance adecuado entre precisión y recall.
 - **Clase 1:** 0.87, un valor aceptable que refleja el compromiso entre precisión y recall en los casos de alto riesgo.

- **Exactitud Global (Accuracy):**

La precisión general es de 0.93, lo que significa que el modelo clasifica correctamente el 93% de las instancias en el conjunto de prueba, confirmando su fiabilidad.

Estos resultados sugieren que el modelo es efectivo para clasificar zonas de riesgo, proporcionando una herramienta útil para identificar áreas de alto riesgo de accidentalidad. Su alta capacidad para detectar casos críticos lo convierte en un recurso valioso para las políticas de seguridad vial en Colombia, facilitando la identificación y priorización de zonas que requieren medidas preventivas. La reducción del sobreajuste mejoró significativamente su desempeño, aumentando su aplicabilidad práctica en el análisis de riesgos de accidentalidad. Cabe mencionar que, aunque los hiperparámetros ajustados por la IA pueden causar ligeras variaciones en las métricas de evaluación, el rendimiento general del modelo se mantiene estable, reforzando su fiabilidad para la toma de decisiones en seguridad vial.

CONCLUSIONES

En este proyecto, se cumplieron los objetivos propuestos mediante un análisis exhaustivo y estructurado de los datos sobre homicidios causados por accidentes de tránsito en Colombia. Se emplearon técnicas de análisis exploratorio y estadístico que permitieron limpiar, normalizar y organizar los datos, logrando así identificar tendencias y correlaciones entre variables relevantes. A continuación, se detallan los principales logros en relación con los objetivos específicos del estudio:

- El análisis exploratorio identificó tendencias y correlaciones clave en los homicidios por accidentes de tránsito. Los departamentos con más accidentes de tránsito son Cundinamarca, Antioquia y Valle del Cauca, que concentran el 45.22% de estos incidentes. Esto subraya la necesidad de políticas públicas y mejoras en la seguridad vial en estas regiones. Además, se evidenció que los hombres adultos, especialmente conductores de vehículos y motocicletas, son el grupo más afectado, lo que indica la importancia de estrategias preventivas dirigidas a ellos. La tendencia en el tiempo muestra un aumento constante, reforzando la urgencia de implementar medidas de intervención sostenidas para reducir la siniestralidad.
- Se construyó un modelo predictivo que permite anticipar las zonas de mayor riesgo para estos incidentes.
- Para facilitar la toma de decisiones se diseñó un dashboard interactivo con cinco visualizaciones clave del análisis: la distribución geográfica de los accidentes, la tendencia durante la última década, el tipo de transporte, el grupo etario y el género de la víctima. Esta herramienta interactiva es útil para los responsables de seguridad vial y facilita el acceso a información crítica para planificar intervenciones y medir su impacto.

DOCUMENTACIÓN Y ENLACES DEL PROYECTO

El repositorio donde se documenta y evidencia cada proceso realizado en el proyecto se encuentra disponible en el siguiente enlace: [Repositorio Homicidios por accidentes de transito](#). En dicho repositorio se pueden encontrar la base de datos en archivo CSV, el archivo con datos ya depurados, la conexión de CSV a MySQL, el proyecto y las consultas en MySQL, el dashboard elaborado en Tableau cuyo enlace directo es [Dashboard de Homicidios por accidentes de transito](#), los análisis estadísticos de homicidios en un archivo Python (.py), y el modelo de clasificación Random Forest Classifier.