

Un análisis estadístico sobre la accidentalidad en ciclistas en Bogotá

Ana María Cruz Pacheco¹

¹ Facultad de Ingeniería y Ciencias Básicas

Universidad Central

Maestría en Analítica de Datos

Curso de Bases de Datos

Bogotá, Colombia

¹*acruzp3@ucentral.edu.co*

Octubre, 2023

Contents

1	Introducción	3
2	Características del proyecto de investigación que hace uso de Bases de Datos	4
2.1	Titulo del proyecto de investigación	6
2.2	Objetivo general	6
2.2.1	Objetivos especificos	7
2.3	Alcance del proyecto	7
2.4	Pregunta de investigación	8
2.5	Hipotesis	8
3	Reflexiones sobre el origen de datos e información	8
3.1	¿Cual es el origen de los datos e información?	8
3.2	¿Cuales son las consideraciones legales o eticas del uso de la información?	9
3.3	¿Cuales son los retos de la información y los datos que utilizara en la base de datos en terminos de la calidad y la consolidación?	9
3.4	¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto?	10
4	Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)	10

4.1	Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto	10
4.2	Diagrama modelo de datos	11
4.2.1	Modelo Lógico:	11
4.2.2	Modelo Relacional:	11
4.2.3	Creación de tablas a partir del modelo ER:	13
4.3	Código SQL - lenguaje de definición de datos (DDL)	13
4.4	Imágenes de la Base de Datos	15
4.5	Código SQL - Manipulación de datos (DML) (<i>Primera entrega</i>) . .	16
4.6	Código SQL + Resultados: Vistas (<i>Primera entrega</i>)	16
4.7	Código SQL + Resultados: Triggers (<i>Primera entrega</i>)	16
4.8	Código SQL + Resultados: Funciones (<i>Primera entrega</i>)	16
4.9	Código SQL + Resultados: procedimientos almacenados (<i>Primera entrega</i>)	16
5	Bases de Datos No-SQL (<i>Segunda entrega</i>)	17
5.1	Diagrama Bases de Datos No-SQL (<i>Segunda entrega</i>)	17
5.2	SMBD utilizado para la Base de Datos No-SQL (<i>Segunda entrega</i>)	17
6	Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos (<i>Tercera entrega</i>)	18
6.1	Ejemplo de aplicación de ETL y Bodega de Datos (<i>Tercera entrega</i>)	18
6.2	Automatización de Datos (<i>Tercera entrega</i>)	18
6.3	Integración de Datos (<i>Tercera entrega</i>)	18
7	Proximos pasos (<i>Tercera entrega</i>)	19
8	Lecciones aprendidas (<i>Tercera entrega</i>)	20
9	Bibliografía	21

1 Introducción

La Organización Mundial de la Salud (OMS) ha catalogado la accidentalidad vial como una de las principales epidemias de la sociedad. En un estudio realizado en conjunto con el Banco Mundial, los accidentes de tránsito se posicionan como la séptima causa de morbilidad a nivel mundial. Específicamente, los accidentes de tránsito representan una de las principales causas de mortalidad en las ciudades.

En el caso de Bogotá, los accidentes de tránsito son la segunda causa de muertes violentas, siendo los homicidios la principal causa de muerte en el país, representando el 60,5% de las muertes. Estas cifras son proporcionadas por la OMS.

La Asamblea General de las Naciones Unidas, en junio de 2022, hizo un llamado a los gobiernos para garantizar la inclusión de todos los sectores de la sociedad en las acciones relacionadas con la seguridad vial. Además, instó a impulsar políticas y medidas que busquen reducir a la mitad el número de víctimas fatales y lesionadas para el año 2030. En el año 2019, se registraron un total de 6.633 accidentes de tránsito, mientras que en el año 2022 se reportaron 8.032 accidentes, lo que representa un incremento del 20.99%, según los datos proporcionados por la Agencia Nacional de Seguridad Vial (ANSV).

En Bogotá, los ciclistas son considerados actores viales especialmente vulnerables, lo que los expone a un mayor riesgo de sufrir accidentes de tránsito. De acuerdo con los datos del Observatorio de la Secretaría Distrital de Movilidad, durante el primer semestre de 2022, un total de 1.142 ciclistas resultaron lesionados en accidentes de tránsito en la capital, lo cual representa un aumento del 7% en comparación con el mismo período del año anterior, cuando se registraron 1.066 casos.

El objetivo de este proyecto de investigación es analizar las causas y consecuencias de los accidentes de tránsito que involucran a ciclistas en la ciudad de Bogotá, así como identificar las variables principales que contribuyen a su ocurrencia. Para lograrlo, se utilizará un archivo de datos abiertos que contiene información sobre los siniestros viales, y se aplicarán métodos de análisis de datos, como el análisis descriptivo, el análisis de regresión y el análisis de series de tiempo.

El análisis descriptivo implica el uso de estadísticas descriptivas para resumir y visualizar los datos, como tablas, gráficos y medidas de tendencia central y dispersión. Este método permitirá obtener información general sobre la frecuencia, gravedad, ubicación y características de los accidentes de bicicletas.

Por otro lado, el análisis de regresión se utilizará para examinar la relación entre una variable independiente y una variable dependiente. Por ejemplo, se investigará si la edad del ciclista, la hora del día o el tipo de intersección en la que ocurrió el accidente tienen algún efecto en la probabilidad de sufrir una lesión grave.

Además, se empleará el análisis de series de tiempo para estudiar los patrones temporales y las tendencias a lo largo del tiempo. Este método permitirá analizar la frecuencia de los accidentes de bicicletas en una zona específica de la ciudad y determinar si existen patrones estacionales o tendencias a lo largo

del tiempo.

Se espera que los hallazgos de esta investigación permitan identificar las variables específicas que contribuyen a la accidentalidad de los ciclistas en Bogotá, proporcionando una visión global de la situación y generando conclusiones que ayuden a abordar esta problemática de manera más eficiente y efectiva.

2 Características del proyecto de investigación que hace uso de Bases de Datos

Se realizaron búsquedas de artículos, trabajos e información relacionada con el proyecto, encontrándose los siguientes recursos:

Carvajal, G. A., Sarmiento, O. L., Medaglia, A. L., Cabrales, S., Rodríguez, D. A., Quistberg, D. A., & López, S. (2020). "Bicycle safety in Bogotá: A seven-year analysis of bicyclists' collisions and fatalities." *Accident Analysis & Prevention*, 144, 105596. Este artículo analiza las colisiones y fatalidades de ciclistas en Bogotá durante el período 2011-2017. Se muestra una disminución del 55% en las tasas de colisiones fatales y del 46% en las no fatales, ajustadas por la población total de ciclistas. También se identifican diferencias en las tasas de reducción según el sexo y se destacan los factores asociados a las colisiones, los cuales varían según el género.

Duarte, D., López, K. P., & Meneses, S. (2018). "Caracterización de riesgos en la accidentalidad de biciusuarios. Bogotá-Engativá." *Ingeniería Industrial. Actualidad y Nuevas Tendencias*, 6(21), 91-108. Este artículo evidencia el aumento del uso de la bicicleta como medio de transporte en Bogotá y la falta de estudios sobre los factores que inciden en la accidentalidad. Se presentan estadísticas de 60 ciclistas fallecidos y 715 lesionados en 2015, con un aumento del 7% en comparación con el año anterior. El artículo propone realizar análisis de los factores influyentes en los accidentes, utilizando una base de datos suministrada por la Secretaría de Movilidad de Bogotá y datos recolectados de 280 bici-usuarios frecuentes de la localidad de Engativá.

Pasquel M., J. (2017). "Sistema interactivo web desarrollado con software libre para la gestión de información requerida en el análisis de accidentalidad vial en Cali." Esta investigación se centra en el desarrollo de un sistema para la gestión de información espacial de accidentalidad vial en la ciudad de Cali. El objetivo es mejorar los procesos de captura de datos, almacenamiento en base de datos y visualización de información mediante consultas específicas. Esta herramienta resulta útil para la realización de investigaciones, generación de estadísticas y promoción de campañas de concientización sobre seguridad vial.

López Giraldo, R. A., & Sáchica Díaz, N. (2018). "Estudio de Accidentalidad en la Actividad de Mensajería en Bicicleta en Bogotá DC, un Acercamiento desde la Identificación de Factores de Riesgo Laborales." Este estudio caracteriza el sector de la bicimensajería en Bogotá y los factores de riesgo laborales asociados a esta actividad. Se utilizan encuestas, el circuito de seguridad y la matriz GTC 45 para identificar y evaluar los factores de riesgo. Los resultados contribuyen

a comprender la accidentalidad en el contexto de la mensajería en bicicleta y sirven como base para tomar medidas de prevención.

En una investigación realizada por Ruiz Ferrete, Francisco José (2019) en la Universidad de Sevilla titulada "Análisis de la accidentalidad ciclista", se lleva a cabo un análisis integral sobre la accidentalidad ciclista en la actualidad, centrándose en la ciudad de Sevilla. El estudio realiza un análisis cuantitativo de las lesiones y consecuencias de los accidentes en bicicleta utilizando una base de datos proporcionada por uno de los hospitales más prestigiosos de la ciudad. Además, se realiza una revisión actualizada del estado del arte en diversos puntos relacionados con la bicicleta y la accidentalidad ciclista. El documento está estructurado en cinco capítulos principales. El primero de ellos ofrece una introducción al uso de la bicicleta en Sevilla, proporcionando datos importantes y relacionándolos con los motivos que llevaron a elegir esta temática para el proyecto. También se presenta una breve revisión de diversos artículos y estudios científicos sobre la accidentalidad ciclista. El segundo capítulo realiza una revisión general de la accidentalidad ciclista, centrándose en los tipos de vías, entornos, causas y tipología de accidentes más comunes, así como en los principales elementos de seguridad y avances tecnológicos relevantes. Además, se resume la situación actual en diferentes marcos geográficos: Europa, España y Andalucía. A continuación, en el tercer capítulo, se analizan los datos recopilados por el Hospital Universitario Virgen del Rocío de Sevilla. Se realiza un análisis basado en diversas características, de manera similar a los estudios realizados en Europa, como la edad, el sexo, la localización geográfica y temporal, así como un estudio de los diagnósticos y el nivel de gravedad de las lesiones.

En la tesis "Análisis del comportamiento de los ciclistas de Bogotá y su relación con la accidentalidad" de Neira Medina, A. M. (2015) de la Universidad de los Andes, se muestra que los tipos de ciclistas más y menos propensos a accidentarse son: aquellos con comportamientos riesgosos, una alta disposición al riesgo y una alta exposición. El factor humano es el aspecto que muestra una mayor correlación con el número de accidentes de tránsito en ciclistas, seguido de la infraestructura. Los aspectos del factor humano más relacionados con la accidentalidad son el comportamiento riesgoso o peligroso propio y la falta de reacción oportuna. Se observa que los hombres tienden a tener más comportamientos riesgosos que las mujeres, mientras que las mujeres tienden a tener comportamientos menos riesgosos. Otros aspectos como las actitudes y la exposición también influyen en la accidentalidad. En cuanto a las actitudes, se observa que a mayor disposición al riesgo, hay una mayor relación con el número de accidentes. Los comportamientos riesgosos más comunes entre los ciclistas incluyen: conducir a una velocidad mayor que los demás ciclistas, circular cerca de vehículos adelante, utilizar los andenes para transitar, usar audífonos y adelantar por la derecha. Por otro lado, los comportamientos no riesgosos más comunes entre los ciclistas incluyen: usar casco, usar luces y reflectantes entre las 6:00 pm y las 6:00 am, utilizar las ciclorutas disponibles, realizar mantenimiento a la bicicleta y utilizar señales manuales.

En el trabajo de Galindo-Fuentes, J. D. J. (2023) titulado "Perfil epidemiológico

de los accidentes de tránsito en México, 2010-2019” publicado en Horizonte Sanitario, se presenta la problemática de los accidentes de tránsito en México. Se registra una tendencia a la baja en las cifras de fallecimientos y heridos, así como en la tasa de mortalidad y morbilidad por accidentes de tránsito en el periodo estudiado. Aunque los ocupantes de vehículos son las principales víctimas de accidentes fatales y no fatales, los peatones son el grupo más vulnerable en términos de mortalidad. El estudio contribuye a la literatura mexicana en materia de seguridad vial y busca brindar información para la toma de decisiones de las instituciones involucradas.

En una investigación de Guerre, L. E., Sadiqi, S., Leenen, L. P., Oner, C. F., & van Gaalen, S. M. (2020) titulada ”Injuries related to bicycle accidents: an epidemiological study in The Netherlands” y publicada en el European Journal of Trauma and Emergency Surgery, se presenta la situación en los Países Bajos. El ciclismo es un medio de transporte popular y una actividad de ocio con muchos beneficios para la salud y el medio ambiente. Sin embargo, junto con el aumento de la popularidad de la bicicleta, también ha crecido la preocupación por la seguridad vial. Un estudio reciente en Australia estimó que por cada 1000 km en bicicleta se producen 0,29 accidentes. Además, un estudio holandés mostró que en 2012, el 31% de los accidentes de tránsito fatales y el 59% de las víctimas de accidentes atendidas en salas de emergencia eran ciclistas. Además de las bicicletas urbanas convencionales, otros tipos de bicicletas han ganado popularidad, como las bicicletas de carretera, las bicicletas todoterreno y las bicicletas eléctricas. Cada subtipo de bicicleta es conocido por sus usuarios finales específicos y su entorno ciclista preferido, pero se sabe poco sobre las diferencias en los riesgos de lesiones con morbilidad y mortalidad específicas. El ciclismo está arraigado en la cultura holandesa y cada año se venden aproximadamente un millón de bicicletas en los Países Bajos. En comparación con otros países europeos, los Países Bajos tienen una mayor prevalencia del uso de la bicicleta como medio de transporte, pero también una mayor incidencia de lesiones graves por accidentes de bicicleta. Se estima que los costos anuales de los accidentes relacionados con bicicletas en los Países Bajos ascienden a 402 millones de euros.

Las investigaciones previas y los artículos encontrados sobre la accidentalidad vial, en particular en relación con los ciclistas, revelan que este es un problema global. Por lo tanto, es de gran importancia realizar un análisis detallado para identificar las principales variables implicadas en estos accidentes.

2.1 Título del proyecto de investigación

Un análisis estadístico sobre la accidentalidad en ciclistas en Bogotá

2.2 Objetivo general

Analizar las causas y consecuencias de los accidentes de tránsito que involucran a ciclistas en la ciudad de Bogotá, identificando las principales variables que contribuyen a su ocurrencia, utilizando métodos de analítica de datos.

2.2.1 Objetivos específicos

- Identificar y clasificar las variables más relevantes que contribuyen a la ocurrencia de accidentes de tránsito que involucran a ciclistas en la ciudad de Bogotá, considerando factores como la infraestructura vial, el comportamiento del ciclista, las condiciones ambientales y otros elementos relevantes.
- Realizar un análisis descriptivo de los datos de los siniestros viales que involucran a ciclistas en Bogotá, con el fin de obtener información detallada sobre la frecuencia, gravedad, ubicación geográfica y características específicas de los accidentes.
- Realizar un análisis de regresión para investigar la relación entre variables independientes (edad del ciclista, hora del día, tipo de intersección, entre otros) y la probabilidad de sufrir lesiones graves en un accidente de tránsito que involucre a un ciclista, utilizando datos históricos y métodos estadísticos adecuados.
- Realizar un análisis de series de tiempo para identificar patrones temporales y tendencias en la frecuencia de los accidentes de bicicletas en zonas específicas de la ciudad de Bogotá, permitiendo identificar periodos críticos y áreas de mayor riesgo.
- Generar conclusiones basadas en los análisis realizados y proponer recomendaciones concretas para mejorar la seguridad vial de los ciclistas en Bogotá, considerando acciones específicas en materia de infraestructura, educación vial y políticas de prevención.

2.3 Alcance del proyecto

El alcance de este proyecto se enfocará en analizar las causas y consecuencias de los accidentes de tránsito que involucran a ciclistas en la ciudad de Bogotá, con el objetivo central de identificar las variables más relevantes que contribuyen a su ocurrencia. Para llevar a cabo este propósito, se aplicarán métodos de análisis de datos, incluyendo el análisis descriptivo, el análisis de regresión y el análisis de series de tiempo. Las fases del proyecto comprenderán la identificación y clasificación de variables relevantes, un análisis descriptivo de datos para resumir la frecuencia y características de los accidentes, un análisis de regresión para examinar la relación entre variables independientes y la gravedad de los accidentes, y un análisis de series de tiempo para identificar patrones temporales y tendencias. Las conclusiones se basarán en los resultados de estos análisis, y se propondrán recomendaciones específicas para mejorar la seguridad vial de los ciclistas en Bogotá, abordando aspectos de infraestructura, educación vial y políticas de prevención. Es importante tener en cuenta delimitaciones, como la focalización en accidentes de los últimos ocho años y posibles limitaciones en la disponibilidad y calidad de los datos. El proyecto se ejecutará a lo largo de un periodo definido, con entregables específicos y un cronograma detallado para cada fase.

2.4 Pregunta de investigación

¿Cuáles son las variables más relevantes que intervinieron en los accidentes de tránsito que involucran a ciclistas en Bogotá durante los últimos ocho años y cómo pueden ser analizadas a través de la estadística descriptiva?

2.5 Hipotesis

Dada la creciente preocupación por la seguridad vial de los ciclistas en Bogotá y el aumento significativo en los accidentes de tránsito que involucran a esta población, se plantea la hipótesis de que existen variables multifactoriales que contribuyen a la ocurrencia de estos accidentes. Se espera que factores como la infraestructura vial, el comportamiento del ciclista, las condiciones ambientales y otros elementos relevantes desempeñen roles cruciales en la frecuencia y gravedad de los accidentes. Se predice que el análisis detallado de estos factores a lo largo de un período de ocho años, utilizando métodos estadísticos y de análisis de datos, revelará patrones, relaciones y tendencias que arrojarán luz sobre las causas subyacentes de la accidentalidad de los ciclistas en la ciudad. La confirmación de esta hipótesis proporcionaría una base sólida para el desarrollo de estrategias y políticas efectivas destinadas a mejorar la seguridad vial de los ciclistas y reducir la incidencia de accidentes en Bogotá.

3 Reflexiones sobre el origen de datos e información

3.1 ¿Cual es el origen de los datos e información?

En el marco del proyecto propuesto, la obtención de datos e información se realiza a través de fuentes confiables y pertinentes para el análisis de accidentes de tránsito que involucran a ciclistas en Bogotá. La fuente principal de información es la pagina web de datos abiertos de la Secretaria Distrital de Movilidad. Como fuentes complementarias de información, se encuentran la Agencia Nacional de Seguridad Vial (ANSV), que suministra datos estadísticos y registros oficiales de accidentes a nivel nacional, incluyendo información detallada sobre aquellos que involucran a ciclistas en Bogotá. La Secretaría Distrital de Movilidad de Bogotá por medio del Observatorio de Movilidad constituye otra fuente valiosa que , ofreciendo datos específicos sobre movilidad en la ciudad, incluyendo detalles sobre accidentes de bicicletas, ubicaciones, condiciones ambientales y otros factores relevantes. La Organización Mundial de la Salud (OMS) podría proporcionar estadísticas y datos globales sobre accidentes de tránsito, así como información sobre tendencias y mejores prácticas a nivel internacional. Además, estudios académicos, informes gubernamentales y análisis previos sobre seguridad vial y accidentes de bicicletas en Bogotá, así como encuestas y entrevistas locales con ciclistas y expertos en seguridad vial, podrían complementar la información cuantitativa. La calidad y actualidad de los datos, junto con la transparencia en su procedencia, son aspectos esenciales para respaldar la validez y fiabilidad de los hallazgos del proyecto.

3.2 ¿Cuales son las consideraciones legales o eticas del uso de la información?

El proyecto de investigación debe observar consideraciones legales y éticas en el uso de la información, especialmente al tratar datos relacionados con accidentes de tránsito y seguridad vial. Se requiere mantener la privacidad y confidencialidad de la información personal, asegurando su uso solo para fines legítimos de investigación. Cumplir con normativas de protección de datos, como el GDPR, y obtener consentimiento informado es esencial. La anonimización o desidentificación de datos personales antes de su uso ayuda a proteger la identidad de los individuos involucrados. La transparencia en la investigación, el respeto a la integridad de los datos, y la divulgación responsable de resultados son principios éticos clave. Además, se debe tener precaución al analizar datos que involucran a grupos vulnerables, como ciclistas, para evitar estigmatización. La revisión ética por parte de un comité de investigación, cuando corresponda, garantiza que el proyecto cumpla con estándares éticos reconocidos. En última instancia, la investigación debe contribuir positivamente a la sociedad, mejorando la seguridad vial y respetando los derechos de los participantes.

3.3 ¿Cuales son los retos de la información y los datos que utilizara en la base de datos en terminos de la calidad y la consolidación?

El proyecto puede enfrentar diversos desafíos en términos de calidad y consolidación de la base de datos utilizada para analizar accidentes de tránsito con ciclistas en Bogotá. La complejidad y variedad de los datos, junto con posibles inconsistencias en la clasificación de variables, demandan esfuerzos significativos para homogeneizar la información. La calidad de los registros, la falta de datos detallados y la necesidad de anonimización adecuada son consideraciones clave. Además, la identificación de tendencias temporales y estacionales, la actualización de datos y el cumplimiento de consideraciones éticas y legales representan desafíos importantes. La superación de estos retos requiere procesos rigurosos de limpieza y validación, protocolos claros de clasificación y anonimización, y colaboración estrecha con las fuentes de datos para garantizar su calidad y actualización. La transparencia en la documentación de procesos y limitaciones es esencial para la validez y confiabilidad de los resultados.

Un desafío significativo se presenta en el proceso de georreferenciación, ya que la base de datos incluye la dirección donde ocurrió el siniestro, pero el objetivo del estudio requiere información específica de las medidas de latitud y longitud. Esta necesidad se fundamenta en la intención de llevar a cabo un análisis de patrones puntuales a lo largo del tiempo y optimizar el uso de software especializado para análisis estadísticos. Sin embargo, la realización de este proceso implica el uso de software licenciado y costoso. Por consiguiente, resulta imperativo explorar alternativas más accesibles para llevar a cabo eficientemente esta etapa del estudio.

3.4 ¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto?

La implementación de un sistema de bases de datos en el proyecto se anticipa que aporte beneficios significativos para optimizar la investigación sobre accidentes de tránsito con ciclistas en Bogotá. Este enfoque posibilitará el almacenamiento eficiente de extensos conjuntos de datos, facilitando un acceso rápido y la recuperación de información esencial necesaria para análisis detallados. Además de asegurar la integridad y consistencia de los datos, el sistema proporcionará funciones avanzadas de seguridad para resguardar la información confidencial, a pesar de trabajar con datos abiertos ya anonimizados. La escalabilidad del sistema permitirá gestionar el crecimiento de datos a lo largo del proyecto, mientras que su estructura centralizada fomentará una colaboración eficiente entre los miembros del equipo de investigación. Además, respaldará específicamente los métodos de análisis propuestos, como el análisis descriptivo, de regresión y de series de tiempo, contribuyendo así a la efectividad general del estudio al simplificar el mantenimiento y la actualización continua de los datos. La conexión segura a la base de datos mediante diferentes lenguajes utilizados para los procesos asegurará que la información principal no sea modificada, proporcionando así una ventaja crucial para la protección de la información.

4 Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)

4.1 Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto

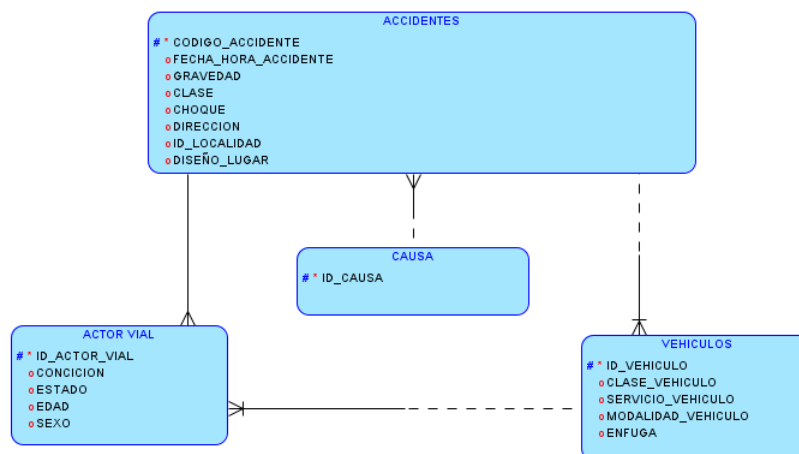
La elección de Oracle Cloud, Oracle Database Server y Oracle Data Modeler como Sistema de Gestión de Bases de Datos (SMBD) para este proyecto se basa en diversas consideraciones estratégicas. Oracle Database Server ofrece una sólida escalabilidad y rendimiento, siendo capaz de gestionar grandes volúmenes de datos y adaptarse al crecimiento del proyecto. Su robusta seguridad, con características como cifrado avanzado y control de acceso detallado, es esencial al tratar datos sensibles sobre accidentes de tránsito. Oracle Data Modeler facilita la creación y mantenimiento eficientes de la estructura de la base de datos mediante una interfaz gráfica intuitiva. La integración con Oracle Cloud proporciona flexibilidad en el despliegue y la administración eficiente de recursos. La compatibilidad con estándares de la industria, el soporte activo y las herramientas analíticas avanzadas hacen que Oracle sea una opción sólida para abordar las necesidades complejas de este proyecto, garantizando un entorno robusto y eficaz para el almacenamiento y análisis de datos relacionados con accidentes de tránsito en Bogotá. Adicionalmente, a lo largo del periodo de aprendizaje en el aula, ha sido con el sistema de gestión de bases de datos con el que se ha generado mayor interacción.

4.2 Diagrama modelo de datos

Un modelo de datos es una representación abstracta y estructurada de la información que se almacena en una base de datos o que se utiliza en un sistema de información. Sirve para describir cómo se organiza, almacena y relaciona la información dentro de una base de datos. Los modelos de datos pueden ser de varios tipos, y cada tipo se utiliza para representar diferentes aspectos de la información y las relaciones entre los datos.

4.2.1 Modelo Lógico:

El modelo lógico de datos se centra en cómo se va a organizar la información de manera específica dentro de la base de datos. Describe la estructura de la información de una manera que es comprensible para las personas y proporciona un enlace directo con el modelo conceptual. En este nivel, se definen tablas, campos y relaciones entre tablas, pero aún no se especifican detalles de implementación a nivel de sistema de gestión de bases de datos (SGBD) específico. Es una representación intermedia entre el modelo conceptual (más abstracto) y el modelo físico (más detallado). El modelo lógico para el proyecto se presenta a continuación en la siguiente imagen:



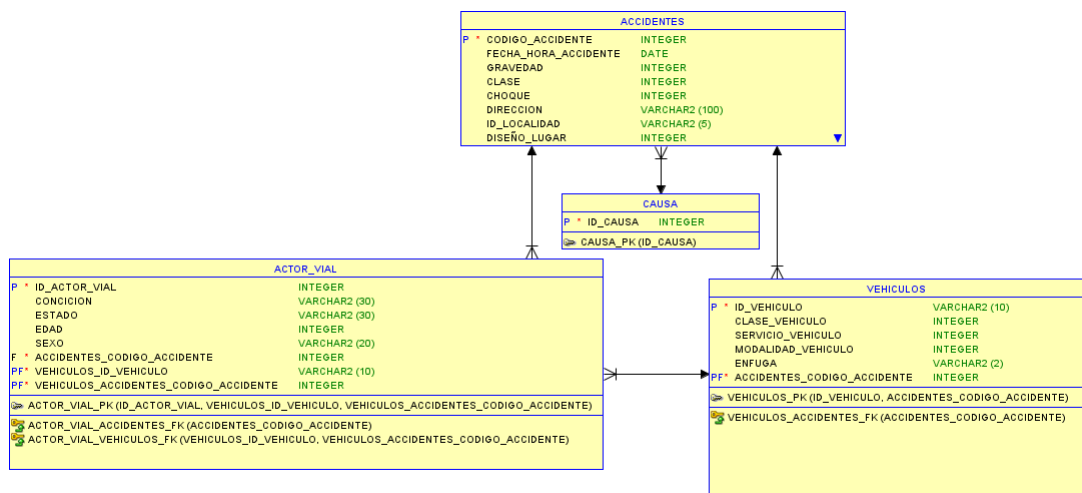
Modelo lógico de la base de datos sobre accidentes de tránsito.

4.2.2 Modelo Relacional:

El modelo relacional es un tipo de modelo lógico que organiza la información en tablas bidimensionales (relaciones) con filas y columnas. Cada tabla representa una entidad y sus atributos, y las relaciones entre entidades se establecen mediante claves primarias y foráneas. Este modelo se basa en la teoría de conjuntos y la lógica de primer orden. Aquí las relaciones creadas entre las tablas parten de las siguientes afirmaciones:

1. Un accidente puede involucrar varios actores viales.
2. Un accidente puede involucrar varios vehículos.
3. Una causa puede encontrarse en varios accidentes.
4. Un vehículo puede estar ocupado por varios actores viales involucrados en el accidente.

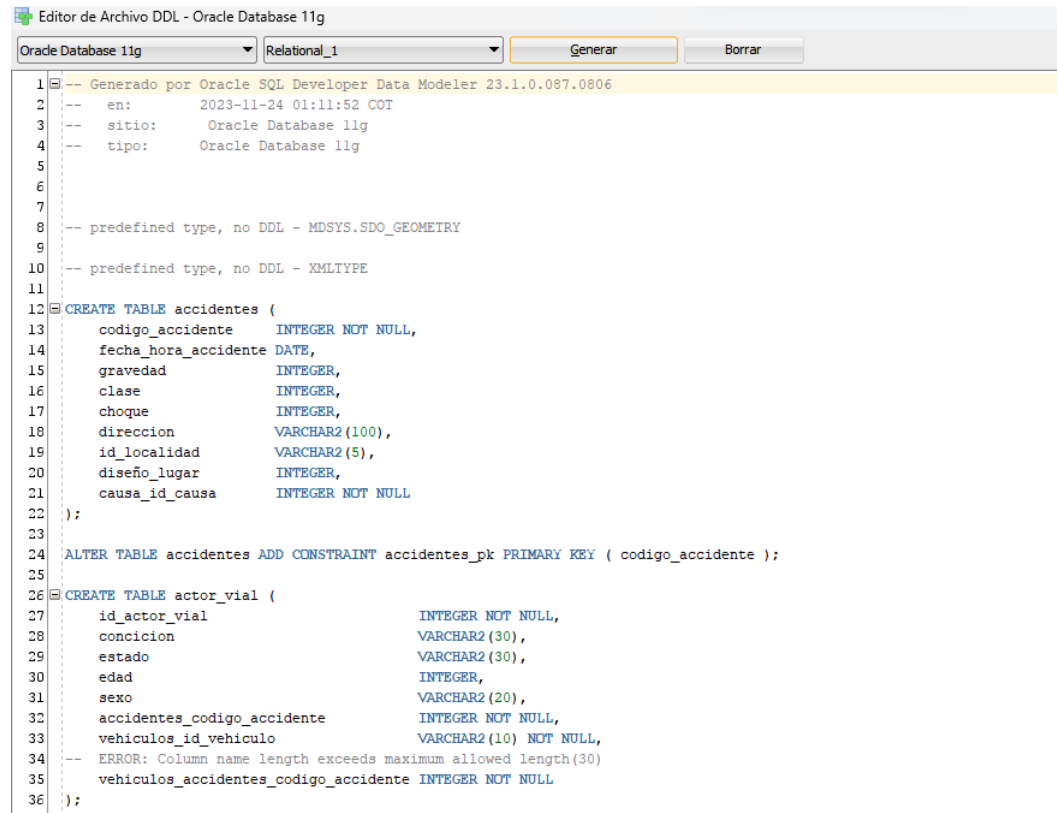
La imagen del modelo relacional creado para el proyecto se presenta a continuación:



Modelo Entidad-Relacion de la base de datos sobre accidentes de tránsito.

4.2.3 Creación de tablas a partir del modelo ER:

Luego de crear el modelo de Entidad-Relación aparecen las llaves foráneas en las distintas tablas creadas. Finalmente, se genera el código para poder realizar la creación de las tablas en el programa SQL developer. La visualización de la parte inicial del código generado se muestra en la imagen siguiente:



```
Editor de Archivo DDL - Oracle Database 11g
Oracle Database 11g Relational_1 Generar Borrar

1 -- Generado por Oracle SQL Developer Data Modeler 23.1.0.087.0806
2 -- en: 2023-11-24 01:11:52 COT
3 -- sitio: Oracle Database 11g
4 -- tipo: Oracle Database 11g
5
6
7
8 -- predefined type, no DDL - MDSYS.SDO_GEOMETRY
9
10 -- predefined type, no DDL - XMLTYPE
11
12 CREATE TABLE accidentes (
13     codigo_accidente INTEGER NOT NULL,
14     fecha_hora_accidente DATE,
15     gravedad INTEGER,
16     clase INTEGER,
17     choque INTEGER,
18     direccion VARCHAR2(100),
19     id_localidad VARCHAR2(5),
20     diseño_lugar INTEGER,
21     causa_id_causa INTEGER NOT NULL
22 );
23
24 ALTER TABLE accidentes ADD CONSTRAINT accidentes_pk PRIMARY KEY ( codigo_accidente );
25
26 CREATE TABLE actor_vial (
27     id_actor_vial INTEGER NOT NULL,
28     concision VARCHAR2(30),
29     estado VARCHAR2(30),
30     edad INTEGER,
31     sexo VARCHAR2(20),
32     accidentes_codigo_accidente INTEGER NOT NULL,
33     vehiculos_id_vehiculo VARCHAR2(10) NOT NULL,
34 -- ERROR: Column name length exceeds maximum allowed length(30)
35     vehiculos_accidentes_codigo_accidente INTEGER NOT NULL
36 );
37
```

Código para la creación de las tablas en SQL developer a partir del modelo Entidad-Relación creado en SQL Data Modeler

4.3 Código SQL - lenguaje de definición de datos (DDL)

El código mencionado en el apartado 4.2.3 se muestra en su totalidad en esta sección, junto con algunos comentarios de cómo son construidas las diferentes tablas con sus relaciones.

```
1 CREATE TABLE accidentes90 (
2     cod_accidente INTEGER NOT NULL,
3     fecha_hora_accidente DATE,
4     gravedad INTEGER,
5     clase INTEGER,
```

```

6      choque                INTEGER,
7      direccion             VARCHAR2(100),
8      id_localidad          VARCHAR2(5),
9      diseno_lugar          INTEGER,
10     causa_id_causa         INTEGER NOT NULL
11 );
12
13 ALTER TABLE accidentes90 ADD CONSTRAINT
14
15 accidentes90_pk PRIMARY KEY ( cod_accidente );

```

Listing 1: Se crea la tabla de accidentes, estableciendo su llave primaria.

```

1 CREATE TABLE actor_vial90 (
2     id_actor_vial          INTEGER NOT NULL,
3     concicion              VARCHAR2(30),
4     estado                 VARCHAR2(30),
5     edad                   INTEGER,
6     sexo                   VARCHAR2(20),
7     accidentes90_cod_accidente INTEGER NOT NULL,
8     vehiculos_id_vehiculo  VARCHAR2(10) NOT NULL,
9     vehiculos_accidentes90_cod_accidente INTEGER NOT NULL
10 );
11 ALTER TABLE actor_vial90
12     ADD CONSTRAINT actor_vial90_pk
13     PRIMARY KEY ( id_actor_vial, vehiculos_id_vehiculo,
14                 vehiculos_accidentes90_cod_accidente
15 );

```

Listing 2: Se crea la tabla de actores viales con las correspondientes llaves foráneas .

```

1 CREATE TABLE causa (
2     id_causa INTEGER NOT NULL
3 );
4 ALTER TABLE causa ADD CONSTRAINT causa_pk
5 PRIMARY KEY id_causa);

```

Listing 3: Se crea la tabla de causa con su llave primaria **id causa**.

```

1 CREATE TABLE vehiculos (
2   id_vehiculo          VARCHAR2(10) NOT NULL,
3   clase_vehiculo       INTEGER,
4   servicio_vehiculo    INTEGER,
5   modalidad_vehiculo   INTEGER,
6   enfuga               VARCHAR2(2),
7   accidentes90_cod_accidente INTEGER NOT NULL
8 );
9 ALTER TABLE vehiculos ADD CONSTRAINT vehiculos_pk
10 PRIMARY KEY ( id_vehiculo, accidentes90_cod_accidente );

```

Listing 4: Se crea la tabla de vehículos con las correspondientes llaves foráneas .

```

1 ALTER TABLE accidentes90
2   ADD CONSTRAINT accidentes90_causa_fk
3   FOREIGN KEY ( causa_id_causa )
4     REFERENCES causa ( id_causa );
5
6 ALTER TABLE actor_vial90
7   ADD CONSTRAINT actor_vial90_accidentes90_fk
8   FOREIGN KEY ( accidentes90_cod_accidente )
9     REFERENCES accidentes90 ( cod_accidente );

```

Listing 5: Se altera la tabla de accidentes y de actor-vial90 con las llaves foráneas **id causa** y **cod-accidente**.

```

1 ALTER TABLE actor_vial90
2   ADD CONSTRAINT actor_vial90_vehiculos_fk
3   FOREIGN KEY ( vehiculos_id_vehiculo,
4     vehiculos_accidentes90_cod_accidente )
5     REFERENCES vehiculos ( id_vehiculo,
6       accidentes90_cod_accidente );

```

Listing 6: Se altera la tabla de actores viales con las correspondientes llaves foráneas .

```

1 ALTER TABLE vehiculos
2   ADD CONSTRAINT vehiculos_accidentes90_fk
3   FOREIGN KEY ( accidentes90_cod_accidente )
4     REFERENCES accidentes90 ( cod_accidente );

```

Listing 7: Se altera la tabla de vehículos con las correspondientes llaves foráneas .

4.4 Imágenes de la Base de Datos

Luego de ejecutar todas las sentencias anteriores en el programa de SQL developer, se hace la consulta para casa una de las tablas. De esta manera se verifica que

todas las columnas requeridas, hayan sido creadas. Las siguientes imagenes son el resultado de los procesos de consulta. Las tablas aparecerán vacías, teniedo en cuenta que no se ha realizado el proceso de cargue de la información.

Tabla ACCIDENTES90

SELECT * FROM ACCIDENTES90;

Salida de Script x Salida de Script 1 x Resultado de la Consulta x

SQL Todas las Filas Recuperadas: 0 en 0,112 segundos

COD_ACC...	FECHA_H...	GRAVEDAD	CLASE	CHOQUE	DIRECCION	ID_LOCAL...	DISEÑO_J...	CAUSA_I...
------------	------------	----------	-------	--------	-----------	-------------	-------------	------------

Tabla ACTOR VIAL90

SELECT * FROM ACTOR_VIAL90;

Salida de Script x Salida de Script 1 x Resultado de la Consulta x

SQL Todas las Filas Recuperadas: 0 en 0,104 segundos

ID_ACTO...	CONCICION	ESTADO	EDAD	SEXO	ACCIDEN...	VEHICUL...	VEHICUL...
------------	-----------	--------	------	------	------------	------------	------------

Tabla CAUSA

SELECT * FROM CAUSA;

Salida de Script x Salida de Script 1 x Resultado de la Consulta x

SQL Todas las Filas Recuperadas: 0 en 0,111 segundos

ID_CAUSA

Tabla VEHICULOS

SELECT * FROM VEHICULOS;

Salida de Script x Salida de Script 1 x Resultado de la Consulta x

SQL Todas las Filas Recuperadas: 0 en 0,244 segundos

ID_VEHIC...	CLASE_VE...	SERVICIO...	MODALID...	ENFUGA	ACCIDEN...
-------------	-------------	-------------	------------	--------	------------

El código del proceso de cargue de la información se realiza en la sección siguiente. Se establece como ejemplo el cargue de pocos registros.

4.5 Código SQL - Manipulación de datos (DML) (Primera entrega)

4.6 Código SQL + Resultados: Vistas (Primera entrega)

4.7 Código SQL + Resultados: Triggers (Primera entrega)

4.8 Código SQL + Resultados: Funciones (Primera entrega)

4.9 Código SQL + Resultados: procedimientos almacenados (Primera entrega)

5 Bases de Datos No-SQL (*Segunda entrega*)

5.1 Diagrama Bases de Datos No-SQL (*Segunda entrega*)

5.2 SMBD utilizado para la Base de Datos No-SQL (*Segunda entrega*)

- 6 Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos** (*Tercera entrega*)
 - 6.1 Ejemplo de aplicación de ETL y Bodega de Datos** (*Tercera entrega*)
 - 6.2 Automatización de Datos** (*Tercera entrega*)
 - 6.3 Integración de Datos** (*Tercera entrega*)

7 Proximos pasos (*Tercera entrega*)

8 Lecciones aprendidas (*Tercera entrega*)

9 Bibliografía