

Proyecto de Bases de datos para la prevención y control de accidentes de tráfico.

Juan Andrés Benavides¹, Daniel Salgado², Jorfan Vargas³

¹⁻²Dpto. de Matemáticas, ³Dpto. de Matemáticas,

Universidad Uexternado

Pregrado Ciencia de Datos

Curso de Bases de Datos

Bogotá, Colombia

{¹*JuanBenavides*,²*Daniel Salgado*}, ³*Jorfan Vargas*

November 22, 2023

Contents

1	Introducción	3
2	Características del proyecto de investigación que hace uso de Bases de Datos	3
2.1	Titulo del proyecto de investigación (Max 100 Palabras) - (<i>Primera entrega</i>)	3
2.2	Objetivo general (Max 100 Palabras) - (<i>Primera entrega</i>)	3
2.2.1	Objetivos especificos (Max 100 Palabras) - (<i>Primera entrega</i>)	3
2.3	Alcance (Max 200 Palabras) - (<i>Primera entrega</i>)	4
2.4	Pregunta de investigación (Max 100 Palabras) - (<i>Primera entrega</i>)	4
2.5	Hipotesis (Max 100 Palabras) - (<i>Primera entrega</i>)	4
3	Reflexiones sobre el origen de datos e información (Max 400 Palabras) - (<i>Primera entrega</i>)	5
3.1	¿Cual es el origen de los datos e información ? (Max 100 Palabras) - (<i>Primera entrega</i>)	5
3.2	¿Cuales son las consideraciones legales o eticas del uso de la información? (Max 100 Palabras) - (<i>Primera entrega</i>)	5
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? (Max 100 Palabras) - (<i>Primera entrega</i>)	6
3.4	¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto? (Max 100 Palabras) - (<i>Primera entrega</i>)	6

4	Diseño del Modelo de Datos del SMBD (Sistema Manejador de Bases de Datos)(Primera entrega)	7
4.1	Características del SMBD (Sistema Manejador de Bases de Datos) para el proyecto (Primera entrega)	7
4.2	Diagrama modelo de datos (Primera entrega)	7
4.3	Imágen de la Base de datos(Primera entrega)	8
4.4	Código SQL - lenguaje de definición de datos (DDL) (Primera entrega)	9
4.5	Código SQL - Manipulación de datos (DML) (Primera entrega) . .	9
4.6	Código SQL + Resultados: Vistas (Primera entrega)	9
4.7	Código SQL + Resultados: Triggers (Primera entrega)	10
4.8	Código SQL + Resultados: Funciones (Primera entrega)	11
4.9	Código SQL + Resultados: procedimientos almacenados (Primera entrega)	12
5	Bases de Datos No-SQL (Segunda entrega)	14
5.1	Diagrama Bases de Datos No-SQL (Segunda entrega)	14
5.2	SMBD utilizado para la Base de Datos No-SQL (Segunda entrega)	14
6	Aplicación de ETL (Extract, Transform, Load) y Bodega de Datos (Tercera entrega)	16
6.1	Ejemplo de aplicación de ETL y Bodega de Datos (Tercera entrega)	16
7	Proximos pasos (Tercera entrega)	17
7.1	Definición de Objetivos del Análisis Predictivo	17
7.2	Preparación y Limpieza de Datos	17
7.3	Selección de Técnicas y Modelos	17
8	Lecciones aprendidas (Tercera entrega)	18
9	Bibliografía	19

1 Introducción

Los accidentes de tráfico son una de las afectaciones mas comunes que tenemos que afrontar los seres humanos todos los días, sabemos que es un problema latente como es enfocado en el estudio Gangi (2018). Buscamos a través de este proyecto de investigación usar las habilidades obtenidas en nuestro curso de bases de datos y nuestras capacidades únicas para el análisis de resultados a partir de los datos para identificar las variables que tienen mayor influencia en el desarrollo de un accidente de tránsito; de esta manera poder proponer una idea útil para reducir y prevenir estas actividades. Ahora bien conocemos las capacidades y campos de estudio que logra abarcar nuestra carrera como bien se destaca en el estudio DataScientest(2022) , los usos más comunes en los que rige la ciencia de datos son: Salud, finanzas y el sector de la ciberseguridad. Si bien estos son usos maravillosos en términos de datos, creemos que se pueden dar nuevos enfoques debido a la gran versatilidad que nos brindan programas específicos de nuestra carrera como lo es Bases de datos. Falta profundizar en nuestros objetivos, tanto generales como específicos, además de el alcance que buscamos abarcar aun así creemos que es una buena probada de lo que queremos mostrar en nuestro proyecto.

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

2.1 Título del proyecto de investigación (Max 100 Palabras) - *(Primera entrega)*

Ornare arcu dui vivamus arcu felis. Egestas integer eget aliquet nibh praesent.

2.2 Objetivo general (Max 100 Palabras) - *(Primera entrega)*

Realizar un análisis cuantitativo exhaustivo de los datos de accidentes de tráfico para identificar patrones y tendencias en la siniestralidad vial, con el propósito de proponer medidas efectivas para mejorar la seguridad en las carreteras y reducir el número de accidentes y víctimas. Este análisis incluirá el estudio de variables como la hora del día, el tipo de vehículo, la dirección del accidente y la gravedad de estos, entre otros factores relevantes.

2.2.1 Objetivos específicos (Max 100 Palabras) - *(Primera entrega)*

- Analizar la distribución temporal de los accidentes de tráfico, estudiando variables como la hora del día y el día de la semana, para identificar los momentos de mayor riesgo de accidente.
- Investigar la relación entre el tipo de vehículo y la gravedad de los accidentes, con el fin de determinar si ciertos tipos de vehículos están asociados con accidentes más graves

- Examinar la distribución geográfica de los accidentes, identificando las direcciones con mayor número de accidentes, para proponer medidas de seguridad vial específicas para estas áreas.

2.3 Alcance (Max 200 Palabras) - (*Primera entrega*)

El alcance en el que queremos enfocar nuestro estudio, teniendo en cuenta nuestro conocimiento sobre el tema, los datos e interés por nuestra parte es Bogotá. Nuestra ciudad maneja una de las mayores cantidades de movimiento en vías de transito de todo el mundo, y hemos realizado una investigación sobre los estudios generados y son casi nulos. Los principales estudios buscan un enfoque mas amplio donde recubran todos los sucesos del país, como Ruiz y Herrera (2016), Una de las investigaciones que mas nos ayudo a contextualizarnos en el sector de la capital fue el estudio realizado por Vargas, Mozo y Herrera (2012), donde se mencionan a partir de datos, las principales causas de los accidentes en Bogotá. Encontrar diferentes relaciones entre los posibles factores que causan más accidentes de tránsito y los siniestros viales. Exploración de los datos de accidentes de tráfico analizando detalladamente cada uno de los factores que pueden estar relacionados con la frecuencia o gravedad de los accidentes. Después de identificar los factores más relacionados con los accidentes de tránsito, identificar las posibles estrategias para reducir el número de accidentes o la gravedad de estos.

2.4 Pregunta de investigación (Max 100 Palabras) - (*Primera entrega*)

¿Cuáles son los factores que contribuyen a la gravedad de los accidentes de tráfico en Bogotá?

2.5 Hipotesis (Max 100 Palabras) - (*Primera entrega*)

La hora del día, el tipo de vehículo y la ubicación del accidente son factores significativos e importantes que afectan la gravedad y la frecuencia de los accidentes de tráfico. Un análisis detallado y exploratorio de los datos de accidentes revelará patrones y tendencias clave en el contexto de la siniestralidad vial, lo que permitirá proponer medidas efectivas para mejorar la seguridad en las carreteras y reducir el número de accidentes y víctimas.

3 Reflexiones sobre el origen de datos e información

(Max 400 Palabras) - (*Primera entrega*)

3.1 ¿Cual es el origen de los datos e información ? (Max 100 Palabras) - (*Primera entrega*)

Los datos son registrados a través de la plataforma de datos abiertos de Bogotá, proporcionados por la Secretaría Distrital de Movilidad. Dado que provienen de una entidad gubernamental, tenemos una gran confianza en esta fuente de datos. Como se detalla en el punto anterior, consideramos que esta base de datos es muy completa y relevante para nuestra investigación. Cada registro de datos se refiere a un incidente ocurrido en la ciudad de Bogotá y reportado a la Secretaría. Además, muchas decisiones y datos publicados por la alcaldía se basan en estos registros.

3.2 ¿Cuales son las consideraciones legales o eticas del uso de la información? (Max 100 Palabras) - (*Primera entrega*)

En este caso, si bien es un tema algo delicado, creemos que la publicación de estos datos en toda la internet es una gran decisión. Aunque un estudio exhaustivo de los datos podría prestarse para malas prácticas, usualmente las decisiones tomadas en este tipo de entidades están más que fundamentadas. Aun así, nuestro papel con respecto a los datos es generar, en gran parte, conocimiento y tal vez rectificar o modificar una hipótesis principal que tenemos acerca del desarrollo de los siniestros viales.

Ley de Protección de Datos Personales (Ley 1581 de 2012): Esta ley regula la recolección, almacenamiento, uso y circulación de datos personales. Incluso si los datos están públicamente disponibles, su uso debe respetar los principios establecidos en la ley, como la finalidad, la libertad, la veracidad o calidad, y la seguridad de los datos.

Decreto Reglamentario 1377 de 2013: Este decreto establece los términos y condiciones bajo los cuales se debe obtener la autorización para el tratamiento de datos personales. Incluso para los datos públicos, se debe informar a los titulares sobre el uso que se les dará a sus datos.

Habeas Data (artículo 15 de la Constitución Política de Colombia): Cualquier persona tiene derecho a conocer, actualizar y rectificar las informaciones que se hayan recogido sobre ellas en bases de datos o archivos.

Autoridad Nacional de Protección de Datos: Es el ente encargado de hacer cumplir las normas de protección de datos en Colombia. Cualquier investigación debe seguir las directrices y regulaciones establecidas por esta autoridad.

Anonimización de Datos: Si los datos de accidentes de tránsito contienen información personal identificable, deben ser anonimizados antes de su uso en la investigación para proteger la privacidad de los individuos.

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? (Max 100 Palabras) - (Primera entrega)

La Secretaría, tanto como nosotros, entiende que hay siniestros que no son reportados ante ellos. A pesar de esto, pensamos que la información tan detallada y completa es suficiente para hacer una generalización en la ciudad. En este sentido, creemos que quizás este sea uno de los mayores retos, no esto en sí, sino que la terminología que se usa para referirse a cada sujeto, elemento, vehículo, barrio, etc., la base de datos la registra con un número de identificación, lo que provoca que debamos tomar mayor precaución y realizar mayor investigación para no caer en errores de nomenclatura..

3.4 ¿Que espera de la utilización de un sistema de Bases de Datos para su proyecto? (Max 100 Palabras) - (Primera entrega)

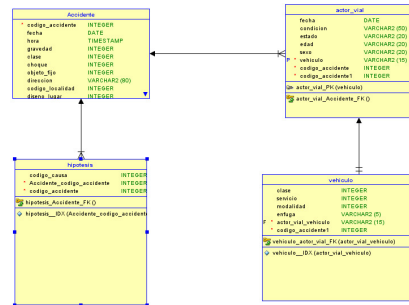
Esperamos desarrollar una base de datos mucho más entendible que la que tenemos en nuestras manos; en ese sentido, definir de una mejor manera cuáles son las entidades principales y cómo, a raíz de estas, podemos no solo responder a nuestra pregunta propuesta, sino que también, si en algún punto nuestro interés se desvía por otras ramas involucradas en la base, a través de las relaciones y vistas poder también extraer información relevante para las partes interesadas. En general, esperamos que el proceso de desarrollo con los datos y su implementación en la base sea gratificante a la hora de referirnos a la absorción de conocimiento. Sin el uso específico de un sistema de bases de datos, podríamos no tener las herramientas que facilitan la manipulación de los datos, lo cual nos costaría mayor cantidad de tiempo y recursos, no solo eso sino que además, deberíamos optar por una organización de los datos y sus variables de una manera poco ortodoxa, cuyas consecuencias podrían ser un vago entendimiento de los datos y por ende una menor capacidad para entender el problema de investigación.

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

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

Oracle se refiere comúnmente a "Oracle Database", que es un sistema de gestión de bases de datos relacional (RDBMS) desarrollado por Oracle Corporation. Oracle Database es uno de los sistemas de gestión de bases de datos más populares y ampliamente utilizados en el ámbito empresarial y en aplicaciones críticas. Hemos decidido utilizar este programa principalmente debido al conocimiento que hemos desarrollado en las últimas semanas sobre el mismo.

4.2 Diagrama modelo de datos (Primera entrega)



4.3 Imágen de la Base de datos(Primera entrega)

	CODIGO_ACCIDENTE	FECHA	HORA	GRAVEDAD	CLASE	CHOQUE	OBJETO_FIJO	DIRECCION	CODIGO_LOCALIDAD	DISEÑO_LUGAR		
1	440194	08/01/15	01/11/23	06:50:00.000000000	AM	3	1	1	(null)	AV AVENIDA BOYACA-AC 12C 02	8	1
2	440182	08/01/15	01/11/23	07:00:00.000000000	AM	3	1	1	(null)	CL 17-KR 1284 02	9	1
3	4401754	08/01/15	01/11/23	07:10:00.000000000	AM	2	2	(null)	(null)	KR R86-CL 548 S 17	7	1
4	4401779	08/01/15	01/11/23	07:10:00.000000000	AM	2	2	(null)	(null)	KR R80-CL 2 S 1	8	1
5	4401826	08/01/15	01/11/23	07:30:00.000000000	AM	3	1	1	(null)	KR R4-CL 514 S 16	6	1
6	4401869	08/01/15	01/11/23	07:30:00.000000000	AM	2	1	1	(null)	KR R3-CL 41 S 28	18	1
7	4401794	08/01/15	01/11/23	07:50:00.000000000	AM	3	1	1	(null)	AC 134-KR 53C 02	11	2
8	4401963	08/01/15	01/11/23	07:50:00.000000000	AM	3	1	1	(null)	AV AVENIDA CIUDAD DE VILLAVICENCIO-AC 70 S 02	19	1
9	4401788	08/01/15	01/11/23	07:55:00.000000000	AM	3	1	1	(null)	KR R1-CL 99 00	2	1
10	4401789	08/01/15	01/11/23	08:00:00.000000000	AM	2	1	1	(null)	KR R112-CL 68A 02	10	2
11	4401776	08/01/15	01/11/23	08:15:00.000000000	AM	3	1	1	(null)	CL 13-KR 05 02	16	1
12	4401783	08/01/15	01/11/23	08:30:00.000000000	AM	3	1	1	(null)	KR R6C-CL 79 02	10	1
13	4401790	08/01/15	01/11/23	08:40:00.000000000	AM	3	1	1	(null)	KR R5-CL 174 02	1	1
14	4401801	08/01/15	01/11/23	08:40:00.000000000	AM	3	1	1	(null)	CL 64-KR 9 32	2	2
15	4401787	08/01/15	01/11/23	09:20:00.000000000	AM	2	2	(null)	(null)	KR R2-CL 58 S 02	8	1
16	4401794	08/01/15	01/11/23	10:00:00.000000000	AM	3	1	1	(null)	KR R8D-CL 39 S 02	8	2
17	4401758	08/01/15	01/11/23	10:00:00.000000000	AM	2	1	1	(null)	AK 30-CL 22 02	14	1
18	4401752	08/01/15	01/11/23	10:25:00.000000000	AM	2	4	(null)	(null)	AV AVENIDA CARACAS-CL 56 S 02	5	1
19	4401771	08/01/15	01/11/23	10:30:00.000000000	AM	3	1	1	(null)	AK 45-CL 80 49	12	1
20	4401802	08/01/15	01/11/23	10:30:00.000000000	AM	3	1	1	(null)	CL R50-KR 80 02	7	1
21	4401809	08/01/15	01/11/23	10:30:00.000000000	AM	3	1	4	3	CL 120-KR 1 26	17	1
22	4401859	08/01/15	01/11/23	11:00:00.000000000	AM	3	1	1	(null)	CL 17F-KR 128 02	9	2
23	4401774	08/01/15	01/11/23	11:45:00.000000000	AM	2	1	1	(null)	AV AVENIDA BOYACA-AC 284 02	6	2
24	4401795	08/01/15	01/11/23	12:00:00.000000000	PM	3	1	1	(null)	AV AVENIDA BOYACA-CL 169 02	11	2
25	4401805	08/01/15	01/11/23	12:30:00.000000000	PM	3	1	1	(null)	AV AVENIDA CIUDAD DE CALI-KR 123 02	9	1
26	4401777	08/01/15	01/11/23	01:00:00.000000000	PM	3	1	1	(null)	KR 5-CL 29 02	3	1
27	4401776	08/01/15	01/11/23	01:00:00.000000000	PM	3	1	1	(null)	CL 36-KR 13 02	1	1
28	4401911	08/01/15	01/11/23	01:40:00.000000000	PM	3	1	1	(null)	CL 184-KR 30 02	14	1
29	4401807	08/01/15	01/11/23	01:55:00.000000000	PM	2	1	1	(null)	CL 244-KR 14815 SE 02	4	2
30	4401795	08/01/15	01/11/23	02:00:00.000000000	PM	3	1	1	(null)	CL 16-KR 69F 02	9	1
31	4401797	08/01/15	01/11/23	02:00:00.000000000	PM	3	1	1	(null)	CL 36-KR 75 S 02	8	2
32	4401853	08/01/15	01/11/23	02:00:00.000000000	PM	2	1	1	(null)	CL 127-KR 54 02	11	1
33	4401798	08/01/15	01/11/23	02:40:00.000000000	PM	3	1	1	(null)	KR 19-CL 18 20	15	1
34	4401786	08/01/15	01/11/23	03:10:00.000000000	PM	3	1	1	(null)	CL 129-KR 59 02	11	2
35	4401868	08/01/15	01/11/23	03:10:00.000000000	PM	3	1	1	(null)	AC 127-KR 144 44	1	1
36	4401821	08/01/15	01/11/23	03:30:00.000000000	PM	2	2	(null)	(null)	KR R1-CL 73A S 02	7	2
37	4401842	08/01/15	01/11/23	03:35:00.000000000	PM	3	1	1	(null)	AK 79-CL 74815 02	10	1
38	4401788	08/01/15	01/11/23	04:10:00.000000000	PM	3	1	1	(null)	CL R5-CL 26 02	13	1
39	4401800	08/01/15	01/11/23	04:20:00.000000000	PM	3	1	4	10	CL 604-KR 751 S 22	19	1
40	4401806	08/01/15	01/11/23	04:20:00.000000000	PM	2	1	1	(null)	KR 54-CL 44 02	13	2
41	4401781	08/01/15	01/11/23	04:30:00.000000000	PM	3	1	1	(null)	KR 128-CL 174 02	9	1
42	4401801	08/01/15	01/11/23	04:30:00.000000000	PM	3	1	1	(null)	AV AVENIDA SUBA-AC 120 02	11	1
43	4401822	08/01/15	01/11/23	04:40:00.000000000	PM	3	1	1	(null)	AV AVENIDA BOYACA-CL 26 02	10	1
44	4401796	08/01/15	01/11/23	05:00:00.000000000	PM	3	1	1	(null)	AV AVENIDA BOYACA-CL 59 S 02	19	5
45	4401780	08/01/15	01/11/23	05:30:00.000000000	PM	3	1	1	(null)	CL 22A-KR 6 SE 02	1	1
46	4401782	08/01/15	01/11/23	05:30:00.000000000	PM	3	1	1	(null)	AV AVENIDA CIUDAD DE CALI-CL 129 02	11	1
47	4401805	08/01/15	01/11/23	05:30:00.000000000	PM	2	1	1	(null)	KR R0-CL 44 S 02	8	1
48	4401784	08/01/15	01/11/23	05:33:00.000000000	PM	3	1	1	(null)	AK 79-CL 74815 02	8	1
49	4401759	08/01/15	01/11/23	05:40:00.000000000	PM	3	1	1	(null)	KR R6-CL 714 02	10	2
50	4401818	08/01/15	01/11/23	06:15:00.000000000	PM	2	1	1	(null)	AV AVENIDA DEL SUR-CL 85C S 33	7	1
51	4401833	08/01/15	01/11/23	03:00:00.000000000	PM	3	1	1	(null)	AV AVENIDA CIUDAD DE CALI-CL 151 02	11	2
52	4401750	08/01/15	01/11/23	07:50:00.000000000	PM	2	2	(null)	(null)	CL 58B-KR 28815 02	13	1
53	4401840	08/01/15	01/11/23	08:10:00.000000000	PM	3	1	1	(null)	AV AVENIDA BOYACA-CL 127 02	11	1
54	4401764	08/01/15	01/11/23	09:00:00.000000000	PM	3	1	1	(null)	CL 145-KR 115 02	11	1
55	4401770	08/01/15	01/11/23	09:00:00.000000000	PM	3	1	1	(null)	CL 38-KR 871 S 15	8	1
56	4401772	08/01/15	01/11/23	09:15:00.000000000	PM	3	1	1	(null)	KR R4-CL 9 02	16	2
57	4401751	08/01/15	01/11/23	09:20:00.000000000	PM	3	1	1	(null)	CL R80-KR 28 02	12	1
58	4401751	08/01/15	01/11/23	09:30:00.000000000	PM	2	2	(null)	(null)	CL 100-KR 59 02	12	2
59	4401755	08/01/15	01/11/23	10:30:00.000000000	PM	2	1	1	(null)	AC R80-KR 94L 02	10	2
60	4401763	08/01/15	01/11/23	10:40:00.000000000	PM	2	1	1	(null)	KR 9-CL 52 S 02	6	2
61	4401765	08/01/15	01/11/23	01:10:00.000000000	AM	3	1	4	6	KR R6-CL 162 57	1	1
62	4401788	08/01/15	01/11/23	03:00:00.000000000	AM	2	1	4	1	KR R0-CL 48A S 21	8	1
63	4401762	08/01/15	01/11/23	03:00:00.000000000	AM	3	1	1	(null)	AV AVENIDA CARACAS-AC 13A 04	6	2
64	4401804	08/01/15	01/11/23	04:05:00.000000000	AM	2	1	1	(null)	AK 68-CL 8 S 05	8	1
65	4401840	08/01/15	01/11/23	04:20:00.000000000	AM	3	1	1	(null)	KR 115-CL 142 02	11	1
66	4401815	08/01/15	01/11/23	06:30:00.000000000	AM	2	3	(null)	(null)	CL 50A-KR 33 S 25	6	1
67	4401895	08/01/15	01/11/23	06:40:00.000000000	AM	3	1	1	(null)	KR 108B-CL 12504 02	11	2
68	4401832	08/01/15	01/11/23	07:10:00.000000000	AM	3	1	1	(null)	AK 45-CL 30 02	12	1
69	4401829	08/01/15	01/11/23	07:14:00.000000000	AM	3	1	1	(null)	KR 7-CL 05 02	2	1
70	4401817	08/01/15	01/11/23	07:15:00.000000000	AM	2	6	(null)	(null)	CL 127-KR 56 02	11	1
71	4401860	08/01/15	01/11/23	07:30:00.000000000	AM	2	4	(null)	(null)	AV AVENIDA SUBA-CL 106 02	11	1
72	4401830	08/01/15	01/11/23	07:40:00.000000000	AM	3	1	1	(null)	CL 68-KR 46 S 35	19	1
73	4401838	08/01/15	01/11/23	07:40:00.000000000	AM	3	1	1	(null)	KR 7-CL 159 02	1	1
74	4401819	08/01/15	01/11/23	08:00:00.000000000	AM	3	1	1	(null)	AV AVENIDA DE LAS AMERICAS-KR 43 02	16	2
75	4401858	08/01/15	01/11/23	08:00:00.000000000	AM	3	1	1	(null)	CL 63-KR 7805 02	7	1
76	4401816	08/01/15	01/11/23	08:20:00.000000000	AM	2	1	1	(null)	AV AVENIDA CIUDAD DE CALI-CL 15A 80	8	1
77	4401822	08/01/15	01/11/23	08:20:00.000000000	AM	3	1	1	(null)	CL 13-KR 32 02	16	1
78	4401831	08/01/15	01/11/23	08:30:00.000000000	AM	3	1	1	(null)	CL 127-KR 116 02	1	1
79	4401905	08/01/15	01/11/23	08:30:00.000000000	AM	3	1	1	(null)	AC 1-KR 13A 14	14	1
80	4401825	08/01/15	01/11/23	08:50:00.000000000	AM	3	1	1	(null)	KR 30-CL 63A 02	12	1
81	4401896	08/01/15	01/11/23	08:50:00.000000000	AM	3	1	1	(null)	AK 30-CL 68 02	12	1
82	4401834	08/01/15	01/11/23	09:00:00.000000000	AM	3	1	4	2	KR 28B-CL 72A 33	12	1
83	4401850	08/01/15	01/11/23	09:10:00.000000000	AM	3	1	1	(null)	KR R84-DG 74B 02	10	2
84	4402003	08/01/15	01/11/23	09:30:00.000000000	AM	2	1	1	(null)	CL R80-KR 113 02	10	1
85	4401818	08/01/15	01/11/23	09:50:00.000000000	AM	2	2	(null)	(null)	CL 40-KR 7 S 2	1	1
86	4401852	08/01/15	01/11/23	10:00:00.000000000	AM	3	1	1	(null)	KR 17-CL 9 02	14	1
87	4401808	08/01/15	01/11/23	10:05:00.000000000	AM	2	2	(null)	(null)	KR 24-CL 62 24	13	1
88	4401825	08/01/15	01/11/23	10:50:00.000000000	AM	2	1	1	(null)	AV AVENIDA BOYACA-CL 17 02	9	1
89	4401840	08/01/15	01/11/23	11:00:00.000000000	AM	3	1	1	(null)	CL 13-KR 68B 02	9	1
90	4401851	08/01/15	01/11/23	11:00:00.000000000	AM	3	1	1	(null)	CL 106-CL 15A 02	9	1

4.4 Código SQL - lenguaje de definición de datos (DDL) *(Primera entrega)*

```
CREATE TABLE accidente (
    codigo_accidente INTEGER NOT NULL,
    fecha DATE,
    hora TIMESTAMP,
    gravedad INTEGER,
    clase INTEGER,
    choque INTEGER,
    objeto_fijo INTEGER,
    direccion VARCHAR2(80),
    codigo_localidad INTEGER,
    diseno_lugar INTEGER
);

CREATE TABLE actor_vial (
    fecha DATE,
    condicion VARCHAR2(50),
    estado VARCHAR2(20),
    edad VARCHAR2(20),
    sexo VARCHAR2(20),
    vehiculo VARCHAR2(15) NOT NULL,
    codigo_accidente INTEGER NOT NULL,
    codigo_accidentel INTEGER NOT NULL
);

ALTER TABLE actor_vial ADD CONSTRAINT actor_vial_pk PRIMARY KEY ( vehiculo );

CREATE TABLE hipotesis (
    codigo_causa INTEGER,
    accidente_codigo_accidente INTEGER NOT NULL,
    codigo_accidente INTEGER NOT NULL
);

CREATE UNIQUE INDEX hipotesis_idx ON
hipotesis (
    accidente_codigo_accidente
ASC );

CREATE TABLE vehiculo (
    clase INTEGER,
    servicio INTEGER,
    modalidad INTEGER,
    enfuga VARCHAR2(5),
    actor_vial_vehiculo VARCHAR2(15) NOT NULL,
    codigo_accidentel INTEGER NOT NULL
);

CREATE UNIQUE INDEX vehiculo_idx ON
vehiculo (
    actor_vial_vehiculo
ASC );

ALTER TABLE vehiculo
ADD CONSTRAINT vehiculo_actor_vial_fk FOREIGN KEY ( actor_vial_vehiculo )
REFERENCES actor_vial ( vehiculo );
```

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

```
INSERT INTO accidente (codigo_accidente, fecha, hora, gravedad, clase, choque, objeto_fijo, direccion, codigo_localidad, diseno_lugar) VALUES (4441445, TO_DATE('14-JUL-2018', 'DD-MM-YYYY'), TO_TIMESTAMP('11:00:00', 'HH24-MI-SS'), 2, 2, NULL, NULL, 'VIA ALA-CL 30', 4);
INSERT INTO actor_vial (fecha, condicion, estado, edad, sexo, vehiculo, codigo_accidente, codigo_accidentel) VALUES (4441445, TO_DATE('14-JUL-2018', 'DD-MM-YYYY'), TO_TIMESTAMP('05:00:00', 'HH24-MI-SS'), 2, 2, NULL, NULL, 'VIA ALA-CL 30', 4);
INSERT INTO hipotesis (codigo_causa, accidente_codigo_accidente, codigo_accidente) VALUES (4441445, TO_DATE('14-JUL-2018', 'DD-MM-YYYY'), TO_TIMESTAMP('07:15:00', 'HH24-MI-SS'), 2, 2, NULL, NULL, 'VIA ALA-CL 30', 4);
INSERT INTO vehiculo (clase, servicio, modalidad, enfuga, actor_vial_vehiculo, codigo_accidentel) VALUES (4441445, TO_DATE('14-JUL-2018', 'DD-MM-YYYY'), TO_TIMESTAMP('09:00:00', 'HH24-MI-SS'), 2, 2, NULL, NULL, 'VIA ALA-CL 30', 4);
INSERT INTO vehiculo (clase, servicio, modalidad, enfuga, actor_vial_vehiculo, codigo_accidentel) VALUES (4441445, TO_DATE('14-JUL-2018', 'DD-MM-YYYY'), TO_TIMESTAMP('12:00:00', 'HH24-MI-SS'), 2, 2, NULL, NULL, 'VIA ALA-CL 30', 4);
INSERT INTO vehiculo (clase, servicio, modalidad, enfuga, actor_vial_vehiculo, codigo_accidentel) VALUES (4441445, TO_DATE('14-JUL-2018', 'DD-MM-YYYY'), TO_TIMESTAMP('13:00:00', 'HH24-MI-SS'), 2, 2, NULL, NULL, 'VIA ALA-CL 30', 4);
INSERT INTO vehiculo (clase, servicio, modalidad, enfuga, actor_vial_vehiculo, codigo_accidentel) VALUES (4441445, TO_DATE('14-JUL-2018', 'DD-MM-YYYY'), TO_TIMESTAMP('13:00:00', 'HH24-MI-SS'), 2, 2, NULL, NULL, 'VIA ALA-CL 30', 4);
```

4.6 Código SQL + Resultados: Vistas *(Primera entrega)*

Vista 1. 5 Primeros vehiculos de la base de datos

```
CREATE OR REPLACE VIEW vista_vehiculos AS
SELECT *
FROM vehiculos
WHERE ROWNUM <= 5;
```

Resultado 1

	CODIGO_ACCIDENTE	FECHA	VEHICULO	CLASE	SERVICIO	MODALIDAD	ENFUGA
1	4401670	06/01/15	4401670-2	1	2	5	N
2	4401423	01/01/15	4401423-1	1	2	5	N
3	4401420	01/01/15	4401420-1	1	2	5	N
4	4401429	01/01/15	4401429-1	7	2	4	N
5	4401430	01/01/15	4401430-1	10	1	(null)	N

Vista 2. 10 primeros actores viales de la base de datos

```
CREATE OR REPLACE VIEW vista_actores_viales AS
SELECT *
FROM actor_vial
WHERE ROWNUM <= 10;
```

Resultado 2

	FECHA	CONDICION	ESTADO	EDAD	SEVO	VEHICULO	CODIGO_ACCIDENTE
1	04/01/15	CONDUCTOR	ILESO	42	MASCULINO	4401556-1	4401556
2	04/01/15	CONDUCTOR	ILESO	19	MASCULINO	4401556-2	4401556
3	04/01/15	CONDUCTOR	ILESO	40	MASCULINO	4401560-1	4401560
4	04/01/15	CONDUCTOR	ILESO	53	MASCULINO	4401579-1	4401579
5	04/01/15	CONDUCTOR	ILESO	26	MASCULINO	4401579-2	4401579
6	04/01/15	CONDUCTOR	ILESO	18	MASCULINO	4401631-1	4401631
7	04/01/15	CONDUCTOR	ILESO	35	MASCULINO	4401631-2	4401631
8	04/01/15	CONDUCTOR	ILESO	50	MASCULINO	4401635-1	4401635
9	04/01/15	CONDUCTOR	ILESO	62	FEMENINO	4401635-2	4401635
10	04/01/15	CONDUCTOR	ILESO	34	MASCULINO	4401643-1	4401643

Vista 3. 5 primeras direcciones de la base

```
CREATE OR REPLACE VIEW vista_direcciones AS
SELECT *
FROM direcciones
WHERE ROWNUM <= 10;
```

Resultado 3

	DIRECCION
1	AV AVENIDA BOYACA-AC 12C 02
2	CL 17-KR 126A 02
3	KR 89B-CL 54B S 17
4	KR 80-CL 2 51
5	TR 44-CL 51A S 16

4.7 Código SQL + Resultados: Triggers (Primera entrega)

Triggers 1.Registro de auditoría de cambios en los datos

```

CREATE OR REPLACE FUNCTION vehiculo_mas_recurrenente RETURN NUMBER IS
  v_vehiculo_id NUMBER;
  v_max_count NUMBER := 0;
BEGIN
  FOR rec IN (SELECT codigo_accidente, COUNT(*) AS count_vehiculo
              FROM vehiculos
              GROUP BY codigo_accidente
              ORDER BY COUNT(*) DESC)
  LOOP
    IF rec.count_vehiculo > v_max_count THEN
      v_max_count := rec.count_vehiculo;
      v_vehiculo_id := rec.codigo_accidente;
    END IF;
  END LOOP;
  RETURN v_vehiculo_id;
END vehiculo_mas_recurrenente;
/

```

Resultado 1

RESULT

4401419

Triggers 2. Validación de datos antes de la inserción

```

CREATE OR REPLACE TRIGGER validar_edad_actor_vial_try
BEFORE INSERT ON actor_vial
FOR EACH ROW
BEGIN
  IF :new.edad < 18 THEN
    RAISE_APPLICATION_ERROR(-20001, 'La edad del actor vial debe ser mayor de 18 años.');

```

4.8 Código SQL + Resultados: Funciones (*Primera entrega*)

Funcion 1. Localidad que mas aparece

```

CREATE OR REPLACE FUNCTION localidad_mas_comun RETURN VARCHAR2 IS
  v_localidad_mas_comun VARCHAR2(50);
  v_max_count NUMBER := 0;
BEGIN
  FOR rec IN (SELECT codigo_localidad, COUNT(*) AS count_localidad
              FROM siniestros
              WHERE codigo_localidad IS NOT NULL AND codigo_localidad <> '(null)'
              GROUP BY codigo_localidad
              ORDER BY COUNT(*) DESC)
  LOOP
    IF rec.count_localidad > v_max_count THEN
      v_max_count := rec.count_localidad;
      v_localidad_mas_comun := rec.codigo_localidad;
    END IF;
  END LOOP;
  RETURN v_localidad_mas_comun;
END localidad_mas_comun;
/

```

Resultado 1

	LOCALIDAD_MAS_FRECUENTE
1	(null)

Funcion 2. Vehiculo mas recurrente

```
CREATE OR REPLACE FUNCTION vehiculo_mas_recurrente RETURN NUMBER IS
v_vehiculo_id NUMBER;
v_max_count NUMBER := 0;
BEGIN
FOR rec IN (SELECT codigo_accidente, COUNT(*) AS count_vehiculo
FROM vehiculos
GROUP BY codigo_accidente
ORDER BY COUNT(*) DESC)
LOOP
IF rec.count_vehiculo > v_max_count THEN
v_max_count := rec.count_vehiculo;
v_vehiculo_id := rec.codigo_accidente;
END IF;
END LOOP;
RETURN v_vehiculo_id;
END vehiculo_mas_recurrente;
/
```

Resultado 2

	ACCIDENTE_MAS_FRECUENTE
1	4401419

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

Procedimiento 1. Procedimiento para insertar un nuevo actor vial

```
CREATE OR REPLACE PROCEDURE insertar_actor_vial(
P_fecha IN DATE,
P_condicion IN VARCHAR2,
P_estado IN VARCHAR2,
P_edad IN NUMBER,
P_sexo IN VARCHAR2,
P_vehiculo IN VARCHAR2,
P_codigo_accidente IN NUMBER
) IS
BEGIN
INSERT INTO actor_vial(
Fecha,
condicion,
estado,
edad,
sexo,
vehiculo,
codigo_accidente
)
VALUES(
P_fecha,
P_condicion,
P_estado,
P_edad,
P_sexo,
P_vehiculo,
P_codigo_accidente
);
COMMIT;
END insertar_actor_vial;
```

Resultado 1

Procedure INSERTAR_ACTOR_VIAL compilado

Procedimiento 2.Procedimiento para actualizar la condición de un actor vial

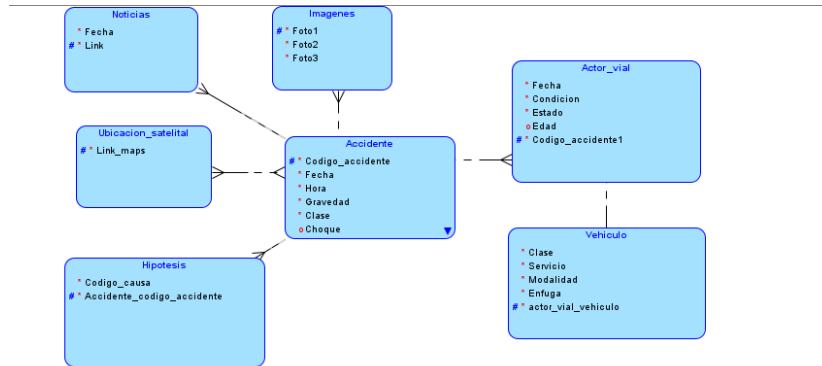
```
CREATE OR REPLACE PROCEDURE actualizar_condicion_actor_vial(  
    p_id_propietario IN INTEGER,  
    p_nueva_condicion IN VARCHAR2  
)  
IS  
BEGIN  
    UPDATE actor_vial  
    SET condicion = p_nueva_condicion  
    WHERE id_propetario = p_id_propietario;  
    COMMIT;  
END actualizar_condicion_actor_vial;
```

Resultado 2

Procedure ACTUALIZAR_CONDICION_ACTOR_VIAL compilado

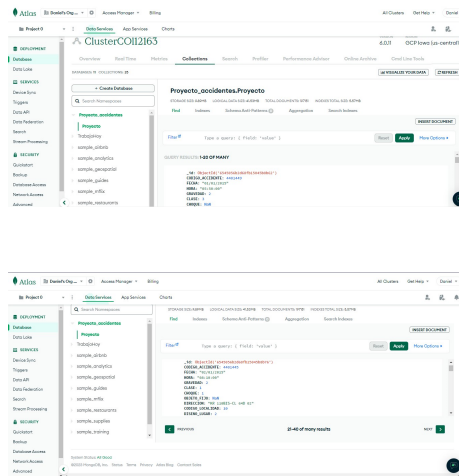
5 Bases de Datos No-SQL (*Segunda entrega*)

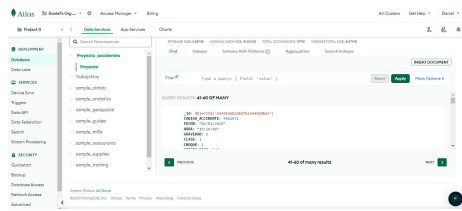
5.1 Diagrama Bases de Datos No-SQL (*Segunda entrega*)



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

Se realiza la conexión con MongoDB que es una base de datos NoSQL orientada a documentos. A diferencia de las bases de datos relacionales que utilizan tablas para almacenar datos, MongoDB utiliza colecciones de documentos. Aunque nuestras nuevas entidades aún no tienen datos, pudimos realizar la conexión exitosamente.





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)

A continuación se presentará un ejemplo de cómo se realiza el proceso de ETL para un determinado caso, pero antes de visualizar este ejemplo se explicará cada uno de los pasos del ETL.

Proceso ETL

Extracción: En la etapa de extracción se realiza la identificación de la tipología de la base de datos para posteriormente extraer los datos. Este paso es crucial para asegurar que se comprende la estructura y formato de los datos, lo que permite una extracción efectiva y precisa.

Transformación: En esta sección se realiza la transformación de la base de datos después de la extracción al formato deseado. Este proceso incluye la limpieza de valores, mapeo y otras transformaciones necesarias para preparar los datos para su análisis y uso.

Carga: Después de realizar la Extracción y la Transformación, los datos se cargan al sistema de destino deseado. Este paso finaliza el proceso ETL, asegurando que los datos estén disponibles y en el formato adecuado para su uso en el sistema de destino.

A continuación se presentará un ejemplo donde se realizó ETL:

```
import pandas as pd
from google.colab import files

def realizar_etl(archivo_entrada, archivo_salida):
    df = pd.read_excel(archivo_entrada)
    df = df.drop_duplicates()
    df.to_csv(archivo_salida, index=False)
    files.download(archivo_salida)
    print("Por favor, seleccione el archivo de entrada:")
    archivo_entrada = files.upload()
    nombre_archivo = list(archivo_entrada.keys())[0]
    archivo_salida = "datos_ETL.csv"
    realizar_etl(nombre_archivo, archivo_salida)

for favor, seleccione el archivo de entrada:
[2025/01/01] Se archiva subcarpetas (local output is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
Saving: seleccionados, seleccionados, imagen_0, click to select, seleccionados, imagen_0, click to select
```

En este código se aprecia cómo se extrae una base de datos tipo 'xlsx' y se realiza la transformación de datos a 'csv' como se aprecia en esta imagen.

```
[CODIGO,ACCIDENTE,FECHA,HORA,GRAVEDAD,CLASE,CHOQUE,OBJETO,FUO,DIRECCION,CODIGO_LOCALIDAD,DISEÑO_LUGAR]
4401418,01/01/2015,01:05:00,2,2,,KR 64A-CL 2C 02,16,2
4401449,01/01/2015,05:50:00,2,3,,AV AVENIDA DEL SUR-KR 65A 41,7,1
4401480,01/01/2015,07:15:00,2,3,,KR 19D-CL 62 5 02,19,1
4401493,01/01/2015,09:30:00,3,1,1,0,,KR 79-CI 43P 5 72,1,1
4401423,01/01/2015,09:45:00,2,1,1,0,,CL 66A-KR 76 02,10,2
4401437,01/01/2015,12:50:00,3,1,1,0,,DG 77A-AK 18D 02,19,1
4401420,01/01/2015,13:00:00,3,1,1,0,,KR 17A-CL 638 02,19,1
4401425,01/01/2015,13:25:00,3,1,1,0,,CL 143A-KR 142 02,11,1
4401447,01/01/2015,13:30:00,3,1,1,0,,CL 72-KR 91A 45,10,1
4401443,01/01/2015,14:30:00,3,1,1,0,,KR 100-CL 23 02,9,1
4401482,01/01/2015,16:15:00,2,2,,KR 1A-CL 418 03 02,4,2
4401429,01/01/2015,16:45:00,2,2,,KR 91-CI 113 20,1,1
4401433,01/01/2015,20:30:00,3,1,1,0,,AK 7-CL 43 33,2,1
4401461,02/01/2015,00:25:00,2,2,,CL 22-KR 16 02,14,1
4401427,02/01/2015,06:30:00,2,1,1,0,,AV AVENIDA CARACAS-CL 6 02,16,2
```


7 Proximos pasos *(Tercera entrega)*

[12pt]article [utf8]inputenc geometry a4paper, margin=1in times
Análisis Predictivo de Datos de Accidentes

7.1 Definición de Objetivos del Análisis Predictivo

Este análisis se centra en la utilización de técnicas avanzadas de modelado predictivo para extraer insights valiosos de los datos de accidentes. Los objetivos específicos incluyen la identificación de patrones y correlaciones ocultas dentro de los datos, la predicción de la probabilidad de ocurrencia de futuros accidentes en diferentes contextos y la evaluación de los factores de riesgo más significativos. Este enfoque permitirá no solo una comprensión más profunda de las dinámicas subyacentes de los accidentes sino también facilitará la formulación de estrategias efectivas de prevención y mitigación.

7.2 Preparación y Limpieza de Datos

Antes de proceder con el análisis, se realiza una exhaustiva preparación de los datos. Este proceso implica la limpieza de datos, la eliminación de duplicados, la corrección de inconsistencias y la imputación de valores faltantes. Posteriormente, se realiza la transformación de los datos, incluyendo la normalización y la creación de variables derivadas, para asegurar que los conjuntos de datos estén optimizados para el análisis predictivo. Se presta especial atención a la selección de características relevantes que contribuirán significativamente a la precisión y eficacia de los modelos predictivos.

7.3 Selección de Técnicas y Modelos

Para abordar los objetivos establecidos, se seleccionarán técnicas de modelado predictivo adecuadas, basándose en la naturaleza de los datos y los objetivos específicos del análisis. Entre las opciones consideradas se incluyen modelos de regresión lineal y logística, árboles de decisión, métodos de ensamble como Random Forest y Gradient Boosting, y algoritmos de aprendizaje automático avanzado. La elección de los modelos se basará en su capacidad para manejar las particularidades de los datos de accidentes y su potencial para proporcionar predicciones precisas y significativas.

En conclusion estos son los pasos a seguir que fueron definidos gracias al desarrollo tematico que se abordó en este proyecto de las distintas implementaciones de bases de datos.

8 Lecciones aprendidas *(Tercera entrega)*

A lo largo del proyecto de la base de datos de accidentes, el desarrollador adquirió conocimientos y experiencias valiosas. Desde una perspectiva técnica, se destacó la importancia de un diseño de base de datos eficiente y coherente. El manejo de datos sobre accidentes presentó desafíos únicos, especialmente en la estructuración y normalización de datos, enfatizando la necesidad de mantener la integridad y la calidad de la información. El proyecto también subrayó la relevancia del análisis y la interpretación de datos en la predicción y prevención de futuros accidentes. A nivel personal, el proyecto reforzó habilidades cruciales como el pensamiento crítico y la resolución de problemas. Las dificultades enfrentadas, desde la gestión de grandes volúmenes de datos hasta la implementación de algoritmos predictivos complejos, fueron una oportunidad invaluable para desarrollar resiliencia y adaptabilidad. Además, la colaboración efectiva con otros profesionales y la gestión eficiente del tiempo se destacaron como elementos esenciales para el éxito del proyecto. En conjunto, estas lecciones no solo ampliaron su comprensión técnica sobre bases de datos y análisis de accidentes, sino que también fortalecieron sus capacidades para abordar futuros proyectos con una base sólida y una perspectiva integral.

9 Bibliografía

- Ganji, S. M. R. E., Ganji, S. M. E. (2018). [FATAL TRAFFIC ACCIDENTS-A REVIEW ON THE CAUSES OF MORTALITY IN IRAN]. 7(49), 5323-5327. <https://doi.org/10.14260/jemds/2018/1177> - DataScientest(2022). Título del artículo. DataScientest, páginas.<https://datascientest.com/es/data-science-definicion-problematica-y-casos-de>
- Ruiz, J. I., Herrera, A. N. (2016). Accidentes de tránsito con heridos en Colombia según fuentes de información: caracterización general y tipologías de accidentes. [Revista CES Psicología], [Volumen(9 Número 1 Enero-Junio 2016)], [pp. 32-46].
- Vargas V., W. E., Mozo P., E., Herrera O., E. (2012). Análisis de los puntos más críticos de accidentes de tránsito en Bogotá. <https://revistas.udistrital.edu.co/index.php/azimut/article/view/5741/7212> - <https://www.funcionpublica.gov>