



Vigilada Mineducación

Rutas Seguras de transporte de Residuos Hospitalarios en Medellín
Safe Routes for the transportation of Hospital Waste in Medellín

ANA CRISTINA URCUQUI HENAO

Trabajo de Grado

Asesor, docente

Édison Valencia Díaz

UNIVERSIDAD EAFIT

ESCUELA DE ADMINISTRACIÓN

MAESTRÍA EN CIENCIAS DE LOS DATOS Y LA ANALÍTICA

Tabla de contenido

Introducción.....	1
Planteamiento del problema	4
Justificación.....	6
Objetivo.....	7
Estado del arte Marco teórico	8
Fuentes de información	10
MEData	10
DANE	12
Datos Abiertos Área Metropolitana del Valle de Aburrá (AMVA)	12
SIATA	12
Procesamiento de la información.....	12
Preparación de la información.....	13
Clustering Incidentes Viales e Intensidad vehicular Ciudad Medellín	17
Datos Georreferenciados.....	18
H3.....	18
Incidentes Viales Georreferenciados	19
Intensidad Vehicular	20
DBSCAN	21
Incidentes Viales Georreferenciados	21
Intensidad Vehicular	22
K-MEANS.....	22
Incidentes Viales Georreferenciados	23
Intensidad Vehicular	25
Caracterización de los Grupos	27
Incidentes Viales Georreferenciados	27
Intensidad Vehicular	29
Momentos de Aforo Movilidad Ciudad de Medellín	31
Modelo Optimización Ruteo de Vehículos	33
Ruta más corta	33
Modelo Optimización Programación Lineal Entera	37
Conclusiones	39
Referencias.....	40

Introducción

Los residuos hospitalarios generados durante la atención médica prestada en hospitales, clínicas y centros de atención médica, cuentan con desechos peligrosos de patogenicidad, reactividad, toxicidad, entre otras características, haciendo que su manejo y tratamiento deba ser especializado. Esta propuesta de trabajo de grado es un apoyo al diseño de rutas más seguras del transporte de residuos hospitalarios en la ciudad de Medellín, desde la fuente (hospitales, clínicas y centros de atención médica), hasta su disposición final (incineradores y recicladores especiales), a través del uso de modelos de optimización que permitan establecer rutas óptimas de recolección de residuos hospitalarios minimizando los riesgos de contagios por accidentes o incidentes y aportando al modelo de ciudades saludables.

El equipo de trabajo se conforma por la estudiante Ana Cristina Urcuqui de profesión estadística, con experiencia en modelos predictivos aplicados en los últimos 4 años en el sector financiero y como director del trabajo de grado el profesor Edison Valencia ingeniero de sistemas y doctor en física, con experiencia en el análisis y visualización de cifras y en el desarrollo de software, configurando un equipo multidisciplinario que, aportando varios puntos de vista, puede garantizar un claro entendimiento del proyecto.

Planteamiento del problema

Los residuos hospitalarios incluyen varios tipos de residuos: patológicos, radiactivos, farmacéuticos, infecciosos y químicos. Así pues, los hospitales son una de las fuentes más importantes de producción de residuos peligrosos en los países. Los estudios revelan que, de los 630 tipos de productos químicos utilizados en los hospitales, 300 son tóxicos y peligrosos [1].

El manejo integral de los residuos hospitalarios es una de las prioridades en las entidades prestadoras de servicios de salud por los efectos y riesgos que pueden generar, ya que si estos no son sometidos a tratamientos y disposiciones finales adecuados, dan lugar a una serie de impactos negativos sobre la vida humana y el medio ambiente, como la transmisión de enfermedades intrahospitalarias, el aumento de número de días de hospitalización para los enfermos, costos de tratamiento, mortalidad intrahospitalaria y contaminación del suelo o de fuentes hídricas [2].

En el informe nacional de residuos o desechos peligrosos generado por el IDEAM [3], Se observa que en Colombia se generaron 41.985 toneladas de residuos hospitalarios, en relación con el año 2014 son aproximadamente 10.000 toneladas más.



Fig 0. Generación Residuos hospitalarios en Colombia

Se puede observar la magnitud con la que estos residuos se están generando y por ello la importancia de su buen tratamiento, ya que son desechos que por sus características químicas pueden poner en riesgo las condiciones de salud de la sociedad y aumentar el deterioro del entorno en el que se desarrolla la vida.

[4] Álvarez (2016), ejecutó un estudio en las Instituciones Prestadoras de Servicios de Salud (IPS) en Neiva, Huila, en el plantea que las prácticas inadecuadas en el manejo de los residuos o desechos hospitalarios contribuyen al daño ambiental y de la salud, gracias a su relación con el incremento y desarrollo de enfermedades tales como infecciones gastroentéricas, infecciones respiratorias, paludismo, afecciones perinatales, enfermedades comunes de la infancia, el VIH/SIDA, malnutrición, asma, cáncer de pulmón e intoxicaciones, entre otras. En consecuencia, el manejo de los residuos sanitarios u hospitalarios se ha convertido en un tema vital tanto para la autoridad sanitaria y ambiental, como para la comunidad en general. Esto en la medida en que es un factor de riesgo para la salud y el ambiente, debido a las características de peligrosidad que representan los diferentes desechos.

Aprovechando el concepto de ciudades inteligentes y seguras, que se caracteriza por el uso de tecnologías de la información para la búsqueda de eficiencia en políticas, reducir al máximo los residuos y molestias, fomentar la calidad social y económica y maximizar la inclusión social, se plantea el uso de los datos

abiertos sobre la ciudad de Medellín y su análisis para contribuir a la construcción de una sociedad sostenible, a través de la problemática de gestión de los residuos hospitalarios, ya que este tipo de desechos requieren de un manejo especializado a lo largo de su proceso de generación hasta su disposición final, enfocándose en el momento del transporte desde el centro hospitalario hacia el lugar de deposición, donde debe movilizarse a través de la ciudad de Medellín generando riesgos, por lo que se propone el análisis de información abierta sobre accidentalidad, informes climáticos, aforos vehiculares y de transporte público que permita segmentar la ciudad en lugares más o menos seguros de movilidad acompañado con la construcción de un modelo de optimización pensado para minimizar los recorridos que en conjunto permita minimizar el riesgo que movilizar estos residuos con lleva.

Considerando el gran volumen de desechos hospitalarios generados diariamente, debido al crecimiento de la demanda y capacidades hospitalarias en las diferentes zonas de la ciudad de Medellín, es de suma importancia abordar el tema de las capacidades limitadas que tienen las empresas prestadoras del servicio de recolección de estos residuos, con el objetivo de plantear procesos óptimos que permitan generar ahorros económicos y un impacto positivo ambiental.

En este trabajo se abordará el transporte de sustancias infecciosas hacia el lugar de disposición final, específicamente las hospitalarias en la ciudad de Medellín, pues es motivo de gran preocupación debido al impacto y las necesidades de la salud pública, frente a una población que demanda mayor calidad de vida, y exige un tratamiento adecuado a los residuos, que garantice la preservación no solo de la salud pública sino también del medio ambiente.

Justificación

La gestión de los residuos hospitalarios es vital, este tipo de desechos requieren un manejo especializado, tanto dentro de las instituciones generadoras, como en su proceso de recolección y transporte a cargo de empresas especializadas para realizar su disposición final.

Considerando el gran volumen de desechos hospitalarios generados diariamente, gracias al crecimiento de la demanda y capacidades hospitalarias en las diferentes zonas de la ciudad de Medellín, es de suma importancia para las compañías tanto privadas como públicas que prestan servicios de recolección de residuos como por ejemplo Emvarias, Grupo ASEI, Enviaseo, entre otras, abordar el tema del transporte de estos desechos hospitalarios, desde la fuente hasta su disposición final pues al estar en movimiento en la ciudad, presenta riesgos tanto para la salud de la población como para el medio ambiente donde habitan los ciudadanos, por ellos se tiene como propósito plantear rutas óptimas que permitan minimizar riesgos en la recolección de estos residuos y transporte hacia un lugar seguro de destrucción, aportando así a la construcción de ciudades sanas con impactos positivos para el medio ambiente

Objetivo

Contribuir a la construcción de ciudades sanas, a partir del análisis de los datos accesibles.

Objetivo General.

Construir un modelo que permita optimizar la recolección de residuos hospitalarios minimizando riesgos de contagios para los ciudadanos de la ciudad de Medellín, a partir del análisis de información abierta, sobre accidentalidad, densidad urbanística, aforos vehiculares, velocidad e intensidad vehicular y ubicación geográfica de los centros generadores de estos residuos.

Objetivos Específicos

- Delimitar segmentos geográficos que permita catalogar niveles de riesgo de las rutas de transporte de residuos hospitalarios, a partir de tasas de accidentalidad históricas.
- Calcular los momentos de aforo de personas en los segmentos geográficos catalogados con niveles altos de riesgo, que intersectan con las rutas de transporte de residuos hospitalarios.
- Cuantificar la intensidad vehicular a partir de datos de movilidad, en puntos de importancia de la ciudad de Medellín, para validar posibles rutas del transporte de residuos hospitalarios.
- Diseñar rutas seguras de transporte de residuos hospitalarios de forma óptima, a partir de la combinación de los diferentes análisis

Estado del arte Marco teórico

El riesgo se define como la posibilidad de que se produzca un contratiempo, existen diferentes metodologías para abordarlos de tal manera que se gestionen e identifiquen controles para mitigar la posibilidad de ocurrencia o el impacto que pueda generar, para este caso de estudio se desea analizar el riesgo al momento de transportar residuos hospitalarios, desde la posibilidad de exposición para las personas y el medio ambiente por el que pasa el material con el objetivo de controlarlo y reducirlo combinado con un modelo de optimización que permita obtener la mejor ruta a recorrer, en la literatura se conoce como problema de ruteo de vehículo (VRP) el cual determina un conjunto óptimo de rutas que inicia y finaliza en el mismo punto, permitiendo visitar a n-clientes con una flota de k-vehículos con una capacidad determinada.

A continuación, se presenta en detalle la revisión bibliográfica realizada referente al transporte seguro de residuos hospitalarios:

Hu et al. [5] Introdujeron por primera vez el problema de diseño de redes para tratar un tipo concreto de residuos peligrosos. El modelo propuesto tiene tres objetivos: minimizar el tiempo de viaje, el riesgo de transporte y el riesgo de localización de las instalaciones de eliminación. Resolvieron su modelo con un enfoque de programación por objetivos.

Ghiani et al. [6] Propusieron un modelo de programación lineal entera para la gestión de residuos urbanos que ayuda a los responsables de la toma de decisiones a seleccionar la ubicación de los lugares de recogida de residuos, así como la capacidad de los depósitos de recogida de residuos. Analizaron el modelo para un área metropolitana del sur de Italia con el objetivo de minimizar el número de puntos de recogida entre las opciones seleccionadas.

Alshraideh y Qdais [7] desarrollaron un modelo de optimización de la recogida de residuos médicos en el norte de Jordania. El modelo de programación propuesto pretende minimizar la distancia de transporte, los costes de transporte y reducir las emisiones de gases tóxicos. Se resuelve mediante el algoritmo genético.

Zhao et al. [8] Presentaron un enfoque mejorado en el diseño de la red de sistemas regionales de gestión de residuos peligrosos. En este trabajo se aborda la gestión de residuos peligrosos, así como su reciclaje. Se propuso un modelo de programación lineal entera para minimizar los costes, como El riesgo se define como la posibilidad de que se produzca un contratiempo, existen diferentes metodologías para abordarlos de tal manera que se gestionen e identifiquen controles para mitigar la posibilidad de ocurrencia o el impacto que pueda generar, para este caso de estudio se desea analizar el riesgo al momento de transportar residuos hospitalarios, desde la posibilidad de exposición para las personas y el medio ambiente por el que pasa el material con el objetivo de controlarlo y reducirlo combinado con un modelo de optimización que permita obtener la mejor ruta a recorrer, en la literatura se conoce como problema de ruteo de vehículo (VRP) el cual determina un conjunto óptimo de rutas que inicia y finaliza en el mismo punto, permitiendo visitar a n-clientes con una flota de k-vehículos con una capacidad determinada.

Con base en los artículos mencionados se puede evidenciar que existen múltiples soluciones para problemas específicos, sin embargo, no se encontró una investigación que haya desarrollado alguna solución en la ciudad de Medellín ni alguna metodología integral que combine la optimización para minimizar los recorridos de transporte de residuos hospitalarios y la relación con el riesgo de salud pública

al transportar estos residuos desde el análisis de datos abiertos de la ciudad que permita planificar rutas seguras.

La oportunidad está en el diseño del modelo empleando fuentes de datos de movilidad, urbanismo y clima.

Fuentes de información

En el desarrollo del presente proyecto se hace uso de fuentes de datos abiertos, dispuestas de manera pública en formatos que permiten su uso y reutilización bajo licencia abierta y sin restricciones legales para su aprovechamiento, en Colombia, la Ley 1712 de 2014 de la Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional. A continuación, se presentan todas las fuentes consideradas para la realización de los objetivos del proyecto.

En Colombia existen diferentes portales que administran la información que se genera desde diferentes frentes, para el caso en estudio se usan las siguientes:

MEData

Es el portal de datos del municipio de Medellín, como lo indica su página principal en este se encuentran “los datos abiertos generados y publicados por las diferentes dependencias de la Alcaldía de Medellín, además de las visualizaciones y análisis que se han creado para generar valor y conocimiento para los ciudadanos”.

A continuación, se presenta una breve descripción proporcionada por el portal sobre la información descargada en temas de movilidad:

- *Pasajeros Movilizados*
Este indicador permite realizar seguimiento a la demanda del transporte público masivo y colectivo de la ciudad, los cuales continúan siendo los principales modos utilizados por sus habitantes. En el siguiente tablero se puede consultar la información referente a pasajeros movilizados por día laboral. El indicador de pasajeros movilizados en transporte público se consolida con periodicidad semestral, no obstante, en el caso de SITVA la desagregación se presenta de manera mensual.
- *Incidentes viales*
Incidentes de tránsito registrados por la Secretaría de Movilidad de la Alcaldía de Medellín, desde el año 2014. Esta base de datos se genera del reporte Incidentes vehículos asociados entregado por UNE. Se entiende por Incidente de tránsito: "evento, generalmente involuntario, generado al menos por un vehículo en movimiento, que causa daños a personas y bienes involucrados en él, e igualmente afecta la normal circulación de los vehículos que se movilizan por la vía o vías comprendidas en el lugar o dentro de la zona de influencia del hecho". (Ley 769 de 2002 - Código Nacional de Tránsito)
- *Velocidad y tiempo de viaje GT*
Este tablero corresponde a los datos recopilados por el Sistema Inteligente de Movilidad de Medellín - SIMM, mediante los dispositivos de Cámaras de Circuito cerrado de Televisión (CCTV) y Apoyo a la Red Semafórica (ARS), por lo que los datos corresponden a velocidades puntuales en el momento de paso por las cámaras.

- Velocidad e intensidad vehicular
Este tablero corresponde a los datos recopilados por el Sistema Inteligente de Movilidad de Medellín, SIMM, mediante los dispositivos de Cámaras de Circuito cerrado de Televisión (CCTV) y Apoyo a la Red Semafórica (ARS), por lo que los datos corresponden a velocidades puntuales en el momento de paso por las cámaras. Con este tablero se puede realizar el seguimiento de velocidad e intensidad vehicular en los principales corredores de la ciudad, los cuales pueden ser comparados mes a mes y por comuna.
- Aforos Vehiculares
Este tablero de consulta consolida los volúmenes aforados por diferentes proyectos en diversas vías de la ciudad, los cuales pueden ser consultados a manera de referencia para variables como la hora pico, composición vehicular y flujo en dichos sectores.
- Conteo Vehicular.
Información correspondiente a conteo vehicular de las cámaras ARS, CCTV, y foto detección del Simmtraffidata.
- Victimas Incidentes Viales
Víctimas en Incidentes de tránsito registrados por la Secretaría de Movilidad de la Alcaldía de Medellín y georreferenciada en el Mapa de Medellín anualmente. Esta base de datos se genera del reporte de víctimas entregado por UNE. Se entiende por Incidente de tránsito: evento, generalmente involuntario, generado al menos por un vehículo en movimiento, que causa daños a personas y bienes involucrados en él, e igualmente afecta la normal circulación de los vehículos que se movilizan por la vía o vías comprendidas en el lugar o dentro de la zona de influencia del hecho. (Ley 769 de 2002 - Código Nacional de Tránsito)
- Parque Automotor
Información general del parque automotor registrado en la Secretaría de Movilidad
- Incidentes de Motos
Incidentes de tránsito con motos registrados por la Secretaría de Movilidad de la Alcaldía de Medellín, anual. Se entiende por Incidente de tránsito: Evento, generalmente involuntario, generado al menos por un vehículo en movimiento, que causa daños a personas y bienes involucrados en él, e igualmente afecta la normal circulación de los vehículos que se movilizan por la vía o vías comprendidas en el lugar o dentro de la zona de influencia del hecho. (Ley 769 de 2002 - Código Nacional de Tránsito)
- Incidentes georreferenciados
Accidentes de tránsito registrados por la Secretaría de Movilidad de la Alcaldía de Medellín y georreferenciada en el Mapa de Medellín anualmente.

DANE

Entidad responsable de la planeación, levantamiento, procesamiento, análisis y difusión de las estadísticas oficiales de Colombia.

- *Proyecciones de Población Medellín 2020*
Reporte de proyecciones de población de Medellín correspondiente a las 16 Comunas y 5 corregimientos de la ciudad. La población se encuentra discriminada por hombres, mujeres, edades simples, grupos quinquenales, esperanza de vida y totales.

Datos Abiertos Área Metropolitana del Valle de Aburrá (AMVA)

Es el portal de Datos Abiertos del Área Metropolitana del Valle de Aburrá, como lo indica su página principal es “una de las iniciativas institucionales que acerca la ciudadanía al territorio digital. Su catálogo provee acceso a los conjuntos de datos de la entidad, así como a aquellos que son generados por los municipios socios y que son de interés metropolitano”.

- *Encuesta Origen Destino 2017 - Datos por Viajes*
Hace referencia a las variables que caracterizan los viajes en el Valle de Aburrá obtenidas a través de la Encuesta Origen Destino realizada por el Área Metropolitana del Valle de Aburrá en el año 2017.

SIATA

“Es un proyecto estratégico del Área Metropolitana del Valle de Aburrá para la gestión ambiental y de riesgos, donde uno de sus principales pilares es el monitoreo permanente y en tiempo real de variables ambientales para la entrega eficiente de información a los organismos gestores de riesgo y la ciudadanía”.

- *Humedad y Precipitación*
La red meteorológica está compuesta por sensores multiparamétricos, que proporcionan información minuto a minuto de temperatura, humedad relativa, precipitación, presión atmosférica, velocidad y dirección de vientos. Estos sensores tienen la capacidad de diferenciar si la precipitación es líquida o sólida.

Procesamiento de la información

Los datos son obtenidos de las fuentes mencionadas anteriormente, y se inicia la etapa de procesamiento de estos, a partir de su limpieza y organización, con el objetivo de que sean confiables y tengan la mejor calidad posible para los análisis posteriores. A partir de estadísticos de resumen como: mínimo, media, desviación, cuartiles, máximo y visualización de datos, se revisan las siguientes características:

- datos Ausentes.
- Coherencia.
- Sentido con el objetivo.

En este acercamiento a los datos se definen cuáles serán las bases de datos sujetos a análisis para el desarrollo de los objetivos. En la siguiente tabla se identifican las bases de datos que continúan en el análisis.

Fuente	Base de Datos	Registros	Variables	Comentario	Continúa
MEData	Pasajeros Movilizados	76	20	No permite tener detalle de la demanda del transporte público masivo, se tiene a nivel semana.	No
MEData	Incidentes viales	270.765	18	No tiene variable para cruzar con otras, y su información se encuentre en los accidentes georreferenciados	No
MEData	Velocidad y tiempo de viaje GT	682.502	20	Permite obtener por día, por hora y ubicación geográfica estadísticos de la velocidad.	Si
MEData	Velocidad e intensidad vehicular	431.491	25	Permite obtener por día, por hora y ubicación geográfica estadísticos de la aforos vehiculares	Si
MEData	Aforos Vehiculares	40.788	41	Permite obtener por día, por hora y ubicación geográfica estadísticos de la aforos vehiculares	Si
MEData	Conteo Vehicular.	3.973.596	21	solo tiene información del año 2021 y contiene gran cantidad de campos null.	No
MEData	Víctimas Incidentes Viales	234.755	19	Víctimas registradas en incidentes viales con algunas características sociodemográficas.	Si
MEData	Parque Automotor	677.491	3	Cantidad de carros registrados en Medellín	No
MEData	Incidentes de Motos	223.439	9	Contiene información georreferenciada de accidentes viales de motocicletas.	Si
MEData	Incidentes georreferenciados	160.839	24	Contiene información georreferenciada de accidentes.	Si
DANE	Proyecciones de Población Medellín 2020	20	7	Densidad poblacional por comunas.	Si
Datos Abiertos AMVA	Encuesta Origen Destino 2017 - Datos por Viajes	45.387	43	Información de momentos de movilidad de personas.	Si
SIATA	Humedad	75.600	7	Variable meteorológica de la humedad.	Si
SIATA	Precipitación	70.056	8	Variable meteorológica de la precipitación.	Si

A partir de las bases de datos que continúan en el análisis, se construye una base de datos maestra que contendrá todas las características posibles para el desarrollo de este proyecto.

Preparación de la información

En este punto se inicia con la transformación de los datos en bruto de tal manera que esta quede de manera adecuada para los posteriores análisis, este proceso se divide en 4 etapas:

Etapas 1: Consolidación información MEData

- La base origen en esta etapa es la base de los incidentes georreferenciados que tiene información entre los periodos 2018 y 2020, y es la base a la cual se le añadirán las demás variables.

- La siguiente base es la de las víctimas en los incidentes viales esta contiene información por radicado de accidente el uno a uno de las victimas resultantes en los incidentes con las siguientes características:
 - Edad: Número de años de la víctima
 - Género: Sexo de la víctima
 - Gravedad: Indicadora si la victima herida o Muerte y
 - Condición: Indica si la víctima era peatón, pasajero, conductor, entro otros.

El tratamiento que se le realiza a esta base es un proceso de agrupación con el objetivo de tener la base a nivel radicado, en donde por cada variable se tendrá un estadístico que la representa, para las variables género, gravedad y condición para cada nivel de su categoría se tendrá un conteo que indica el número de veces que sucedió el evento (nivel categoría) en cada radicado, para la variable edad se tendrá un promedio y además esta variable también viaja categorizada, por tanto también se tendrá por cada nivel el número de incidentes con esa característica. Además, se construye la variable cantidad de víctimas por radicado.

- De la base de incidentes de moto se toma solo el radicado para la construcción de la variable que indicara si el incidente fue en vehículo o en motocicleta.
- La base de velocidad e intensidad vehicular se utiliza para obtener por categoría de vehículo el número de vehículos por hora en los puntos de medición, las coordenadas geográficas de los puntos de medición se encontraban en epsg=6257 por lo que se transforma a epsg=4326, que es el formato del sistema geodésico mundial más común.
- La base velocidad y tiempo de viaje_gt se utiliza para construir la variable de velocidad promedio por corredor, día de la semana y hora del día.

Etapa 2: Consolidación información SIATA

Se descarga información mensual entre 2018 y 2021 por cada estación de medición dentro de Medellín de las variables meteorológicas: precipitación que mide la cantidad de agua que cae en un punto específico y humedad que mide la cantidad de vapor de agua que contiene el aire, con el objetivo de observar si tiene alguna relación con los incidentes viales, para este caso también se investigó desde los tableros de la página del SIATA la ubicación de las estaciones y de esta manera se crean variables del promedio de humedad y precipitación en estas ubicaciones por hora y día de la semana.

Etapa 3: Consolidación información DANE y Datos Abiertos AMVA

Se construye matriz a partir de encuesta de origen destino que permite determinar por comuna y hora del día la cantidad de personas que se estan movilizando, y se cruza con la densidad poblacional proyectada por el DANE por cada comuna.

Etapa 4: Unificación en Base de datos Maestra

En este punto se construye la base de datos maestra, que contendrá todas las variables o características que nos permitirán desarrollar el objetivo de este proyecto.

La base de datos de incidentes georreferenciados es cruzada por radicado con la base de incidentes de motos y con la base de víctimas en los incidentes registrados en un left join.

La base de datos generada con la encuesta origen destino y densidad poblacional de Medellín se cruza por hora y comuna del incidente.

Para el resto de las bases de datos es necesario definir una medida de distancia que nos permita determinar para cada incidente georreferenciado la medición más cercana según la hora y el día de la semana en el que ocurrió el evento. Esta medida de distancia se determina a partir de la distancia geodésica, dado que es la más recomendada para este tipo de ejercicio por su precisión, pues lo que interesa es tener la aproximación más cercana a lo normalmente sucedido al momento del evento.

- Distancia geodésica: [9] La Tierra es una esfera, o elipsoide, tridimensional ligeramente aplanada. La distancia geodésica se calcula en un espacio esférico 3D como la distancia a lo largo de la superficie curvada del mundo.

Al calcular esta distancia se cruza la información por hora y día de la semana, para terminar, tomando el dato que tenga la mínima distancia, esto, para las bases generadas con de velocidad e intensidad vehicular la base de velocidad y tiempo de viaje_gt y las bases con información meteorológica.

Resultado de la preparación

Luego de realizar todo este procedimiento sobre las diferentes bases obtenidas en los portales antes mencionados se obtiene la siguiente base de dato maestra, a partir de la cual se realiza todo el ejercicio propuesto en el proyecto.

Fuente	Base de datos	Campo	Descripción
MEData	incidentes georreferenciados	radicado	Número radicado del incidente.
MEData	incidentes georreferenciados	LONGITUD	Longitud del incidente
MEData	incidentes georreferenciados	LATITUD	Latitud del incidente
MEData	incidentes georreferenciados	Fecha	Fecha de ocurrencia del incidente.
MEData	incidentes georreferenciados	DIA_SEM	Día de la semana en que ocurrió el incidente.
MEData	incidentes georreferenciados	hora	Hora de ocurrencia del incidente.
MEData	incidentes georreferenciados	DIA	Día del mes en que ocurrió el incidente.
MEData	incidentes georreferenciados	MES	Mes del incidente
MEData	incidentes georreferenciados	PERIODO	Año de ocurrencia del incidente.
MEData	incidentes georreferenciados	CLASE	Tipo de accidente (choque, volcamiento, caída, atropello).
MEData	incidentes georreferenciados	DIRECCION_1	Dirección incidente.

MEData	incidentes georreferenciados	BARRIO	Barrio del incidente
MEData	incidentes georreferenciados	COMUNA	Comuna del incidente
MEData	incidentes georreferenciados	GRAVEDAD	Indicadora si en el incidente hubo heridos o muertos.
MEData	Victimas incidentes	cont.	cantidad de víctimas en el incidente.
MEData	Victimas incidentes	edad_prom	Edad promedio de la víctima.
MEData	Victimas incidentes	cond1	Número de víctimas ACOMPAÑANTE DE MOTOCICLETAS.
MEData	Victimas incidentes	cond2	Número de víctimas CICLISTAS.
MEData	Victimas incidentes	cond3	Número de víctimas CONDUCTORES.
MEData	Victimas incidentes	cond4	Número de víctimas MOTOCICLISTAS.
MEData	Victimas incidentes	cond5	Número de víctimas PASAJEROS.
MEData	Victimas incidentes	cond6	Número de víctimas PEATONES
MEData	Victimas incidentes	gen_F	Número de víctimas femenino.
MEData	Victimas incidentes	gen_M	Número de víctimas masculino.
MEData	Victimas incidentes	gen_SI	Número de víctimas sin información género.
MEData	Victimas incidentes	grav1	Número de heridos en el incidente.
MEData	Victimas incidentes	grav2	Número de muertos en el incidente.
MEData	Victimas incidentes	edad1	Número de víctimas con edad entre 0 - 9.
MEData	Victimas incidentes	edad2	Número de víctimas con edad entre 10 - 19.
MEData	Victimas incidentes	edad3	Número de víctimas con edad entre 20 - 29.
MEData	Victimas incidentes	edad4	Número de víctimas con edad entre 30 - 39.
MEData	Victimas incidentes	edad5	Número de víctimas con edad entre 40 - 49.
MEData	Victimas incidentes	edad6	Número de víctimas con edad entre 50 - 59.
MEData	Victimas incidentes	edad7	Número de víctimas con edad entre 60 - 69.
MEData	Victimas incidentes	edad8	Número de víctimas con edad entre 70 - 79.
MEData	Victimas incidentes	edad9	Número de víctimas con edad entre 80 o más
MEData	Victimas incidentes	edadSI	Número de víctimas sin información de la edad.
MEData	Incidentes de Motos	tipo_vehiculo	Indicadora si el incidente fue de vehículo o motocicleta.
MEData	Velocidad e intensidad vehicular	dist_intensidad	Mínima distancia geodésica.
MEData	Velocidad e intensidad vehicular	int_gps	Coordenadas geográficas.
MEData	Velocidad e intensidad vehicular	INTENSIDAD	Número de vehículos mixtos por hora.
MEData	Velocidad e intensidad vehicular	CATEGORIA_1	Número de vehículos categoría 1 por hora.
MEData	Velocidad e intensidad vehicular	CATEGORIA_2	Número de vehículos categoría 2 por hora.
MEData	Velocidad e intensidad vehicular	CATEGORIA_3	Número de vehículos categoría 3 por hora.

MEData	velocidad y tiempo de viaje_gt	dist_velocidad	Mínima distancia geodésica.
MEData	velocidad y tiempo de viaje_gt	velo_gps	Coordenadas geográficas.
MEData	velocidad y tiempo de viaje_gt	AVG_VELOCIDAD	Velocidad promedio.
MEData	velocidad y tiempo de viaje_gt	AVG_LONG_MEDICION	Longitud promedio de medición.
MEData	velocidad y tiempo de viaje_gt	AVG_TIEMPO_MEDICION	Tiempo promedio de medición.
SIATA	Precipitación	dist_precipitacion	Mínima distancia geodésica.
SIATA	Precipitación	precip_gps	Coordenadas geográficas.
SIATA	Precipitación	P1	Precipitación.
SIATA	Humedad	dist_humedad	Mínima distancia geodésica.
SIATA	Humedad	hum_gps	Coordenadas geográficas.
SIATA	Humedad	humedad	Humedad.
DANE-AMVA	Densidad poblacional y Encuesta origen-Destino	Personas	Número de personas promedio en movimiento.

Clustering Incidentes Viales e Intensidad vehicular Ciudad Medellín

Una de las técnicas más utilizadas en el aprendizaje de máquinas no supervisado es el clustering, usado en el análisis de datos para descubrir patrones interesantes en los datos, formando grupos basados en las características existentes en la base, donde cada grupo está caracterizado por tener ciertas propiedades que los diferencian del resto de los grupos y de esa manera se pueden plantear estrategias diferenciadoras para cada grupo.

La distribución general de los incidentes georreferenciados se presenta a continuación, se puede observar que los corregimientos son los sitios con menor cantidad de ocurrencias de incidentes, mientras que por todas las comunas existe una gran cantidad de incidentes viales, sobre este conjunto de datos se realizara la segmentación con dos objetivos:

- Encontrar zonas más o menos riesgosas en función de la ocurrencia de incidentes de tránsito en la ciudad de Medellín.
- Determinar zonas con mayor o menor aforo vehicular en la ciudad de Medellín.

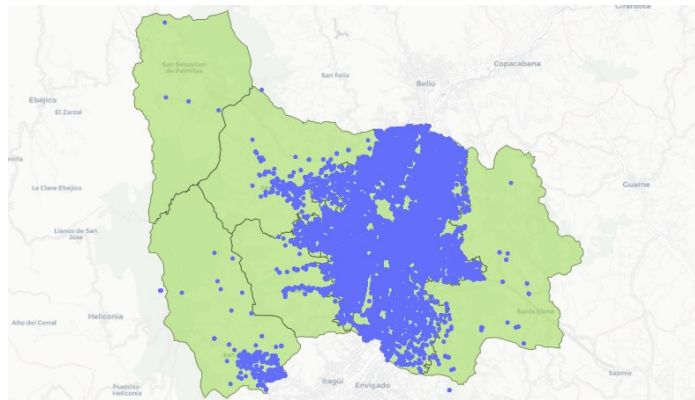


Fig 1. Distribución de los incidentes en la ciudad de Medellín

Datos Georreferenciados

En [10] definen la ciencia de la información geográfica, también conocida como GIScience, es un campo de la ciencia que explora cómo usar y comprender los datos geográficos. Los datos geoespaciales brindan una visión clara de las ubicaciones, que se pueden usar para ver cómo cambian los espacios y los atributos con el tiempo. En [11] indican que, a través del análisis, se pueden descubrir detalles, utilizados para resolver problemas y tomar decisiones, en este caso de estudio se está frente a este tipo de datos, ya que sobre un territorio se desea obtener grupos diferenciales a partir de los sucesos ocurridos.

En la actualidad existen diferentes maneras de analizar los datos georreferenciados, una de las más usadas es a partir de la traza de una lista de coordenadas y el análisis sobre las densidades u ocurrencias dentro de estos espacios, la gran desventaja es que es complejo de manejar y que es comparable entre polígonos al manejar dimensiones diferentes.

Un sistema de indexación geoespacial eficiente ayuda a superar estas desventajas, para el presente proyecto se hace uso de H3, que plantea la solución a través de un esquema hash.

H3

Eshwaran Venkat en [12] indica que H3 es un marco de código abierto desarrollado por Uber en el lenguaje de programación C. En esencia, H3 es una herramienta de análisis geoespacial que proporciona un índice espacial jerárquico hexagonal para obtener información de grandes conjuntos de datos geoespaciales. Los bloques de construcción de H3 son polígonos hexagonales regulares de diferentes tamaños. Estos polígonos están repartidos por toda la proyección del mapa terrestre de polo a polo. Esto significa que cualquier lugar del planeta se puede atribuir a un hexágono H3 con una precisión de 0,0000009 km² de área.

¿Cómo funciona H3?

João Paulo [13] explica que el algoritmo H3 divide la superficie de la Tierra en una red de hexágonos. Puede seleccionar la cantidad de detalles que contiene cada hexágono eligiendo entre los dieciséis niveles disponibles. Puede pensar en estos como niveles de "zoom" en un mapa. Cada vez

que te sumerges en un paso más, los hexágonos se vuelven más pequeños y necesitas más para cubrir la misma área.

Resoluciones H3

Eshwaran Venkat [12] menciona que una fortaleza central de H3 es que cubre el mundo entero con hexágonos de diferentes tamaños. Esto significa que la resolución de la capa se puede ajustar en función del problema que se está resolviendo, como escalar toda la cuadrícula hacia arriba y hacia abajo. H3 contiene un total de 16 resoluciones como se describe en la tabla a continuación, y cada resolución tiene una cierta cantidad de hexágonos que abarcan toda la tierra como una capa que va desde 122 hexágonos en la capa más alta y alrededor de 500 billones de hexágonos en la capa más baja. En el presente análisis se utiliza una resolución de 8.

Res	Average Hexagon Area (km ²)	Pentagon Area* (km ²)	Ratio (P/H)
0	4,357,449.416078381	2,562,182.162955496	0.5880
1	609,788.441794133	328,434.586246469	0.5386
2	86,801.780398997	44,930.898497879	0.5176
3	12,393.434655088	6,315.472267516	0.5096
4	1,770.347654491	896.582383141	0.5064
5	252.903858182	127.785583023	0.5053
6	36.129062164	18.238749548	0.5048
7	5.161293360	2.604669397	0.5047
8	0.737327598	0.372048038	0.5046
9	0.105332513	0.053147195	0.5046
10	0.015047502	0.007592318	0.5046
11	0.002149643	0.001084609	0.5046
12	0.000307092	0.000154944	0.5046
13	0.000043870	0.000022135	0.5046
14	0.000006267	0.000003162	0.5046
15	0.000000895	0.000000452	0.5046

Fig 2. Tabla Resoluciones [14]

Incidentes Viales Georreferenciados

En el siguiente gráfico se observa la ciudad de Medellín a partir de hexágonos coloreados dada la cantidad de incidentes viales ocurridos en él, se evidencia que las vías principales son las de mayor ocurrencia de incidentes viales.

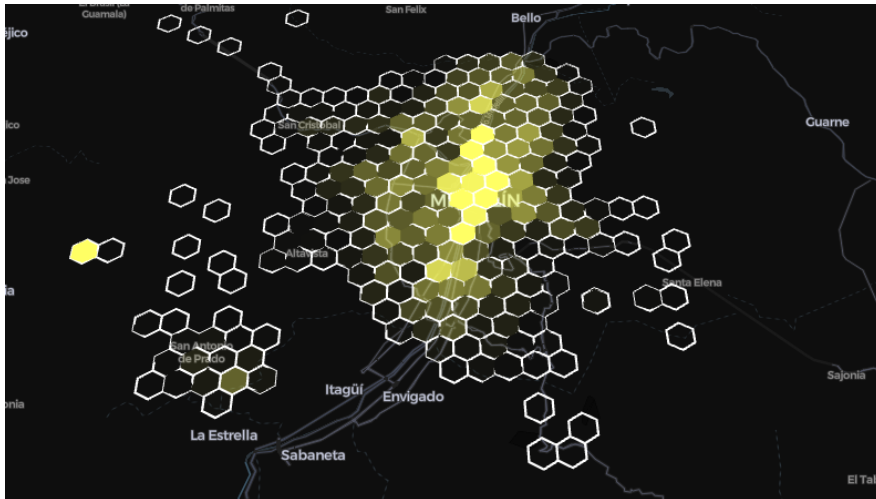


Fig 3. Distribución por Hexágonos y ocurrencia dentro de estos de incidentes viales

Intensidad Vehicular

En el siguiente gráfico se observa la ciudad de Medellín a partir de hexágonos coloreados dada la intensidad vial medida en este, se evidencia que las vías la zona norte es la que mayor intensidad vehicular maneja seguida de la zona del poblado.

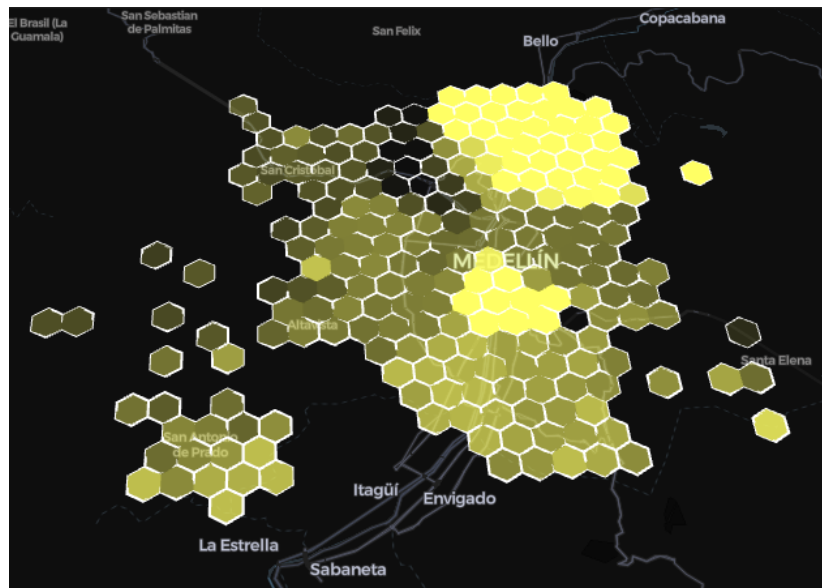


Fig 4. Distribución por Hexágonos y ocurrencia dentro de estos de la intensidad vial

DBSCAN

El agrupamiento espacial basado en densidad de aplicaciones con ruido o Density-based spatial clustering of applications with noise (DBSCAN), es un algoritmo que realiza el agrupamiento de diferentes clústeres basado en la densidad espacial de los datos. Este puede descubrir grupos de diferentes formas y tamaños a partir de una gran cantidad de datos, que contengan ruido y valores atípicos.

A diferencia de otros algoritmos de clustering no supervisado, no es necesario definir de entrada el número de clústeres a encontrar y, además tiene la capacidad de identificar aquellos puntos denominados como ruido que no hacen parte de ninguno de los clústeres definidos por el algoritmo, permitiendo enfocarse en aquellos grupos con mayor densidad de puntos y dando la posibilidad de generar clústeres con diferentes densidades.

Funcionamiento:

El DBSCAN realiza un proceso iterativo, en donde dados un radio ϵ y una cantidad mínima de puntos MinPts, inicia en un punto aleatorio y realiza el siguiente proceso para cada punto:

- Valida si el punto en un círculo con radio ϵ contiene al menos MinPts.
- Si contiene los MinPts lo asigna a un clúster y se denomina punto central.
- Si no tiene los MinPts es marcado como punto de ruido, aún que en otra posible iteración puede hacer pasar a ser un punto fronterizo.

Durante este proceso de asignación al clúster se definen puntos fronterizos que son aquellos que no tienen el mínimo de puntos para ser una región densa, pero pertenecen a la vecindad de un punto central. Adicionalmente DBSCAN conecta los puntos centrales con sus vecinos en una región densa permitiendo la conexión de puntos por densidad basado en una ruta generada por puntos centrales.

A partir de las coordenadas geográficas se realiza la implementación del algoritmo. Para la optimización de hiper-parámetros se utilizó la metodología grid search, que plantea realizar y evaluar la combinación de los diferentes valores para los hiper-parámetros y encontrar aquella combinación óptima en relación a una función de pérdida definida, en este caso el coeficiente silueta.

Incidentes Viales Georreferenciados

Los resultados obtenidos fueron la conformación de 15 grupos que no se diferencian mucho entre ellos, además dada la densidad de las coordenadas se pierde mucha información y los clústeres no se diferencian entre ellos. A continuación, se observan los grupos y el centro de cada uno de los grupos formados.

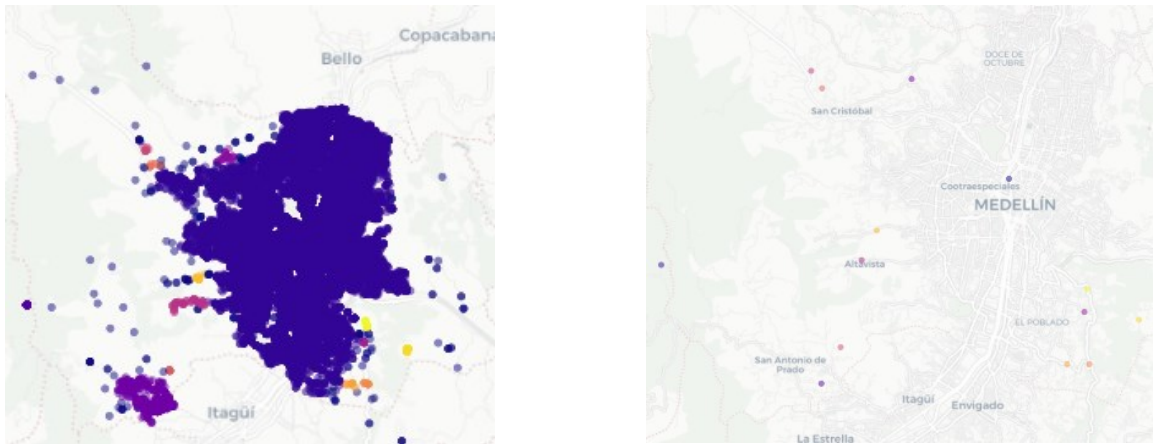


Fig 5. Segmentos entregados por DBSCAN para incidentes viales y sus centros.

Intensidad Vehicular

Los resultados obtenidos fueron la conformación de 15 grupos que no se diferencian entre ellos, además dada la densidad de las coordenadas se pierde la mayor cantidad de la información y los clústeres se traslapan entre ellos. A continuación, se observan los grupos y el centro de cada uno de los grupos formados.

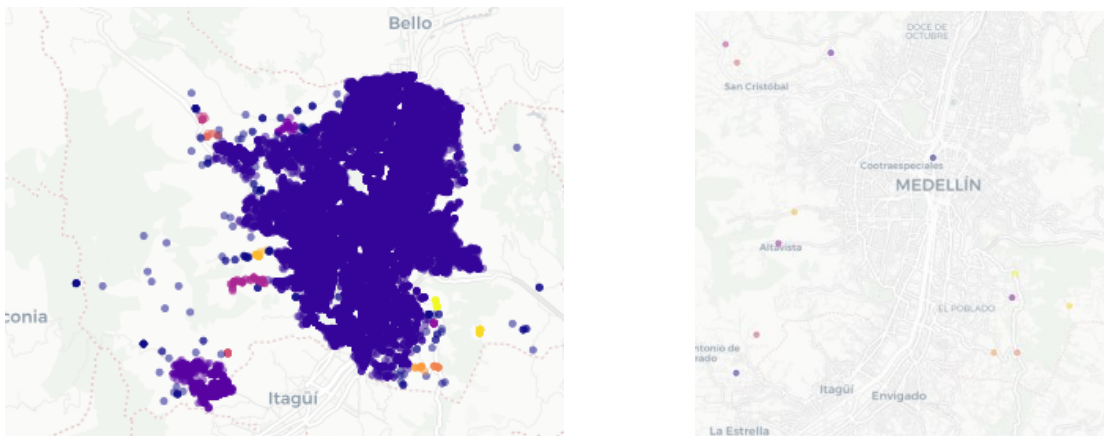


Fig 6. Segmentos entregados por DBSCAN para intensidad vehicular y sus centros.

K-MEANS

El algoritmo de K-means identifica k número de centroides y luego asigna cada punto de datos al grupo más cercano, mientras mantiene los centroides lo más pequeños posible. Las 'medias' en K-means se refieren al promedio de los datos; es decir, encontrar el centroide.

$$\min_{\mathbf{S}} E(\mu_i) = \min_{\mathbf{S}} \sum_{i=1}^k \sum_{\mathbf{x}_j \in S_i} \|\mathbf{x}_j - \mu_i\|^2$$

Funcionamiento:

Para procesar los datos de aprendizaje, el algoritmo K-means en minería de datos comienza con un primer grupo de centroides seleccionados al azar, que se utilizan como puntos de inicio para cada grupo, y luego realiza cálculos iterativos (repetitivos) para optimizar las posiciones de los centroides. Detiene la creación y optimización de clústeres cuando:

- Los centroides se han estabilizado: no hay cambios en sus valores porque la agrupación se ha realizado correctamente.
- Se ha alcanzado el número definido de iteraciones.

Selección del número de grupos usando

El método del codo:

Este método utiliza la distancia media de las observaciones a su centroide. Es decir, se fija en las distancias intra - clúster. Cuanto más grande es el número de clúster k, la varianza intra - clúster tiende a disminuir. Cuanto menor es la distancia intra - clúster mejor, ya que significa que los clústeres son más compactos. El método del codo busca el valor k que satisfaga que un incremento de k, no mejore sustancialmente la distancia media intra - clúster.

Incidentes Viales Georreferenciados

Haciendo uso de los hexágonos generados con H3, se consolida información, para los incidentes georreferenciados se construye la cantidad accidentes y la cantidad de víctimas en cada hexágono, variables insumo para la segmentación con esta metodología.

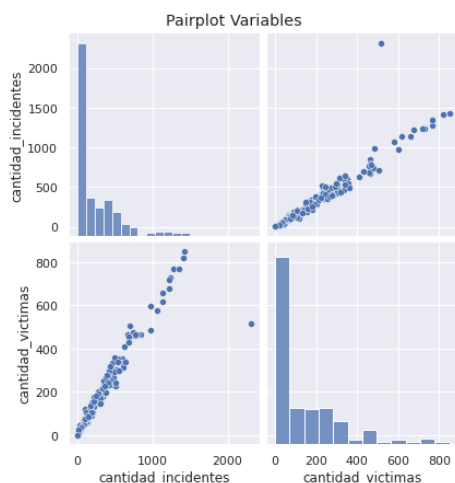


Fig 7. Pairplot de variables incidentes y victimas

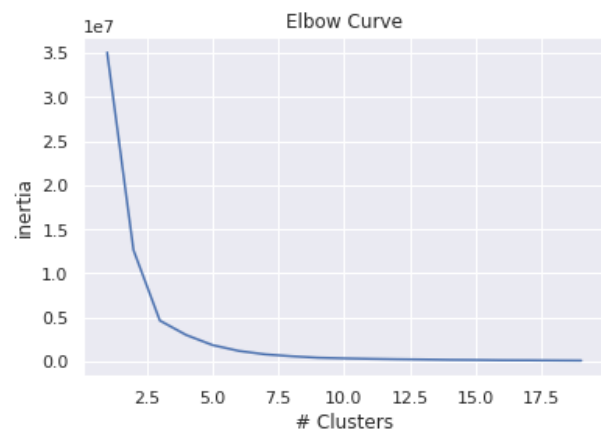


Fig 8. Método del codo selección de clústeres

Se puede observar que existe una relación lineal entre las variables propuestas para la construcción de los segmentos, además el método del codo nos sugiere 4 grupos para conformar los grupos. Los grupos se observan de la siguiente manera:



Fig 9. Distribución de los clústeres

Se conforman 4 grupos, que van desde el grupo que tiene menor cantidad de incidentes georreferenciados y menor cantidad, a los que se encuentra en medio para ambas variables y finalmente el grupo más riesgoso dado que tiene mayor cantidad de incidentes y mayor cantidad de víctimas.

En el gráfico podemos observar como los puntos en amarillo representan los puntos más riesgosos, el centro de la ciudad tiene un riesgo medio-alto y ya los alrededores se distribuyen entre riesgos medio y riesgo bajo.

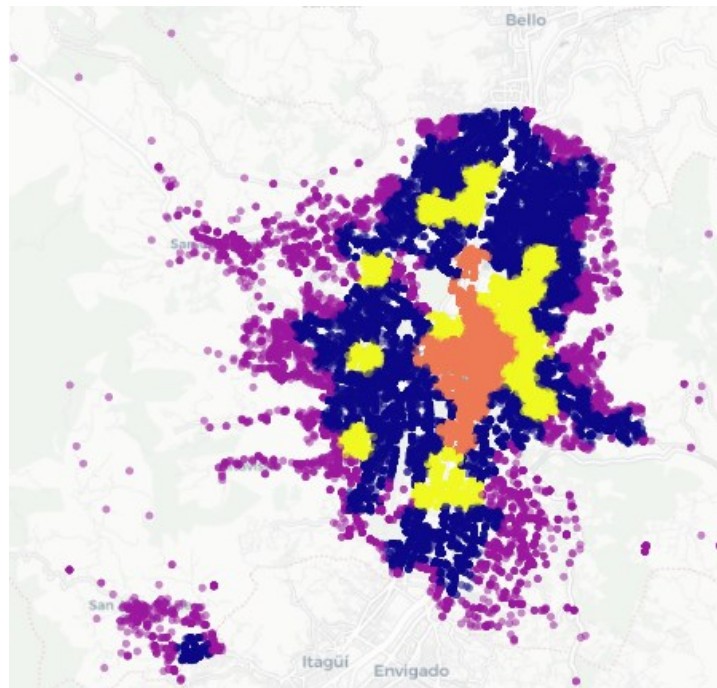


Fig 10. Distribución de los clústeres para incidentes en la ciudad en Medellín

Intensidad Vehicular

- Haciendo uso de los hexágonos generados con H3, se consolida información, Para la intensidad vehicular se realiza un promedio en cada hexágono de las variables intensidad, categoría 1, categoría 2 y categoría 3, variables insumo para la segmentación con esta metodología.

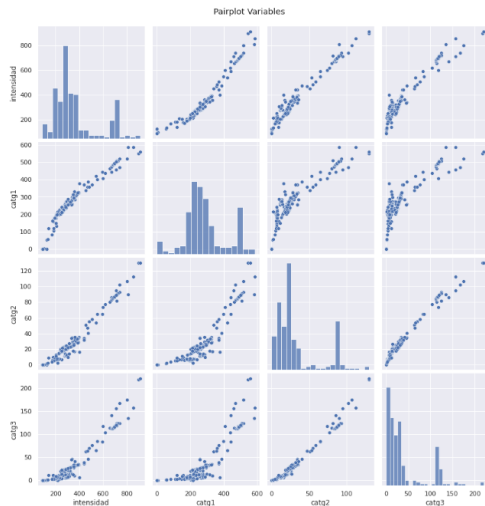


Fig 11. Pairplot de variables intensidad vehicular

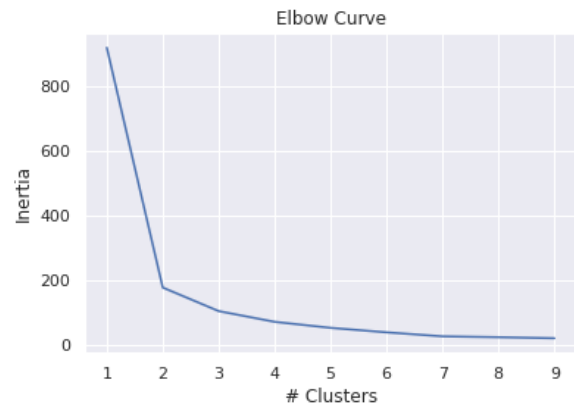


Fig 12. Método del codo selección de clústeres

Se puede observar que existe una relación lineal entre las variables propuestas para la construcción de los segmentos, además el método del codo nos sugiere 3 grupos para conformar los grupos.

Análisis de componentes principales

Es un método estadístico que permite simplificar la complejidad de espacios muestrales con muchas dimensiones a la vez que conserva su información. Supóngase que existe una muestra con n - individuos cada uno con p - variables (X_1, X_2, \dots, X_p), es decir, el espacio muestral tiene p dimensiones. PCA permite encontrar un número de factores subyacentes ($z < p$) que explican aproximadamente lo mismo que las p variables originales. Donde antes se necesitaban p valores para caracterizar a cada individuo, ahora bastan z valores. Cada una de estas z nuevas variables recibe el nombre de componente principal. Para este caso se realizó ACP sobre las variables.

Los grupos se observan de la siguiente manera:

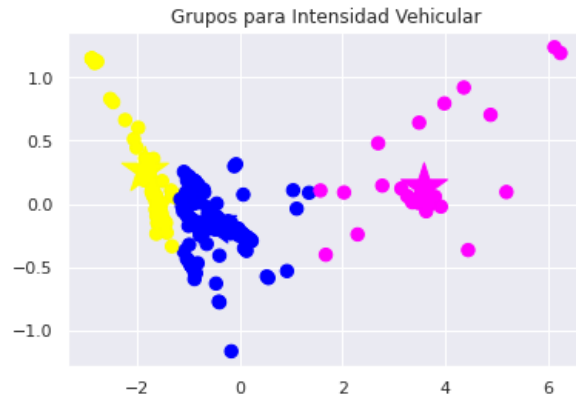


Fig 13. Distribución de los clústeres intensidad vehicular

Se conforman 3 grupos, donde el grupo 1 en azul representa mayor intensidad vehicular, el grupo 2 en rosado representa una intensidad vehicular media y el de color amarillo una intensidad vehicular baja.

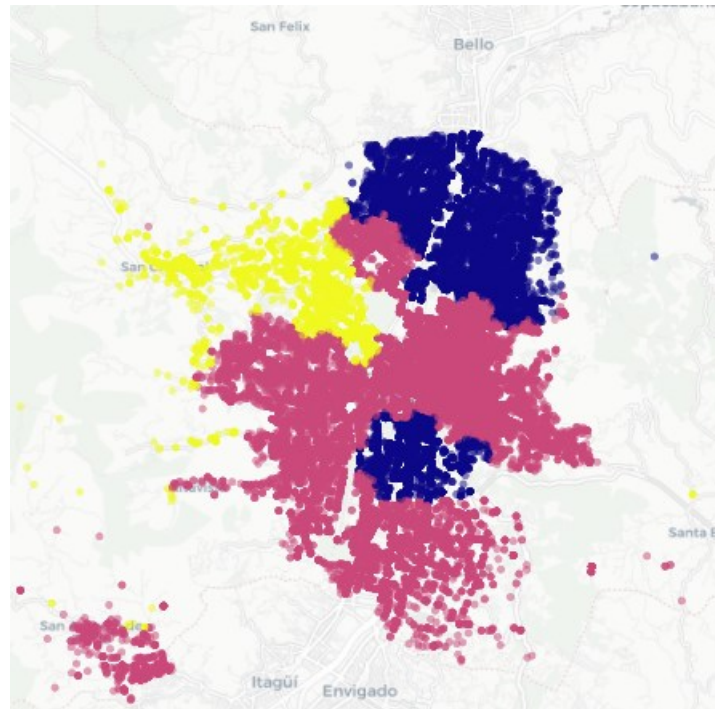


Fig 14. Distribución de los clústeres para intensidad vehicular en la ciudad en Medellín

La metodología seleccionada para definir los grupos tanto para medir el riesgo como para la intensidad vehicular es la de K-means, ya que logra obtener grupos diferenciado entre ellos, es decir, ubicaciones geográficas muy cercana entre los puntos del grupo y separadas entre los grupos, este algoritmo permite obtener grupos geográficos muy consolidados.

Caracterización de los Grupos

Para realizar la caracterización de los grupos se usa: La documentación [15] de Sweetviz indica que es una biblioteca Python de código abierto que genera hermosas visualizaciones de alta densidad para iniciar EDA (Análisis exploratorio de datos) con solo dos líneas de código. El Output es una aplicación HTML completamente autónoma.

El sistema se basa en la visualización rápida de valores objetivo y la comparación de conjuntos de datos. Su objetivo es ayudar al análisis rápido de las características del objetivo, los datos de entrenamiento frente a las pruebas y otras tareas de caracterización de datos

Incidentes Viales Georreferenciados

Clúster 1



Fig 15. Distribución diferentes variables dentro del clúster 1 para incidentes viales

Este grupo se caracteriza por estar distribuido en al noroccidente de la ciudad, la distribución de la velocidad está a lo largo, es decir tiene tanto momento de velocidad alta como baja al igual que la distribución de la intensidad vehicular tiene un gran pico en distribución baja pero otro pedazo a lo largo de una intensidad un poco más alta. Tiene humedades muy distribuías también.

Clúster 2



Fig 16. Distribución diferentes variables dentro del clúster 2 para incidentes viales

Este grupo se caracteriza por encontrarse en el centro de la ciudad, las distribuciones de las velocidades promedio tienden a ser más bajas, La distribución de la humedad tiende ser más baja al igual que la intensidad vehicular y la cantidad de víctimas.

Clúster 3

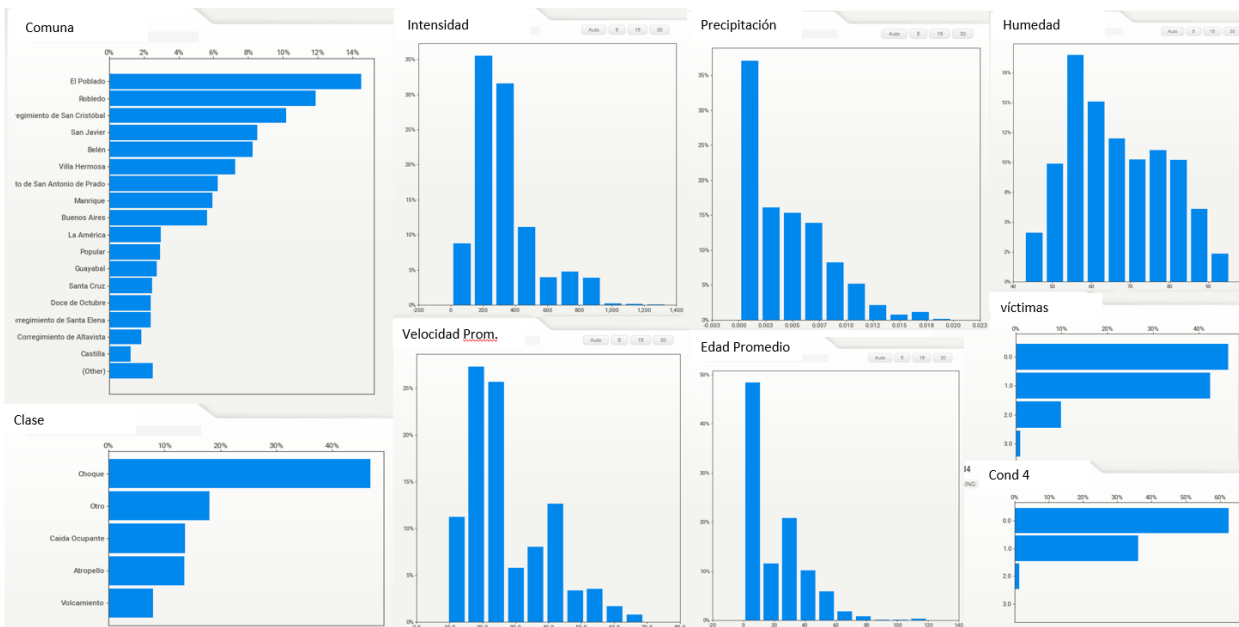


Fig 17. Distribución diferentes variables dentro del clúster 3 para incidentes viales

Este grupo se caracteriza por estar distribuido en varias zonas de la ciudad, tiende a tener velocidades más bajas, pero también tiene sectores con esta un poco más alta, la distribución de humedad esta inclinada a la derecha, es decir, es más alta y la intensidad vehicular tiene a ser media.

Clúster 4

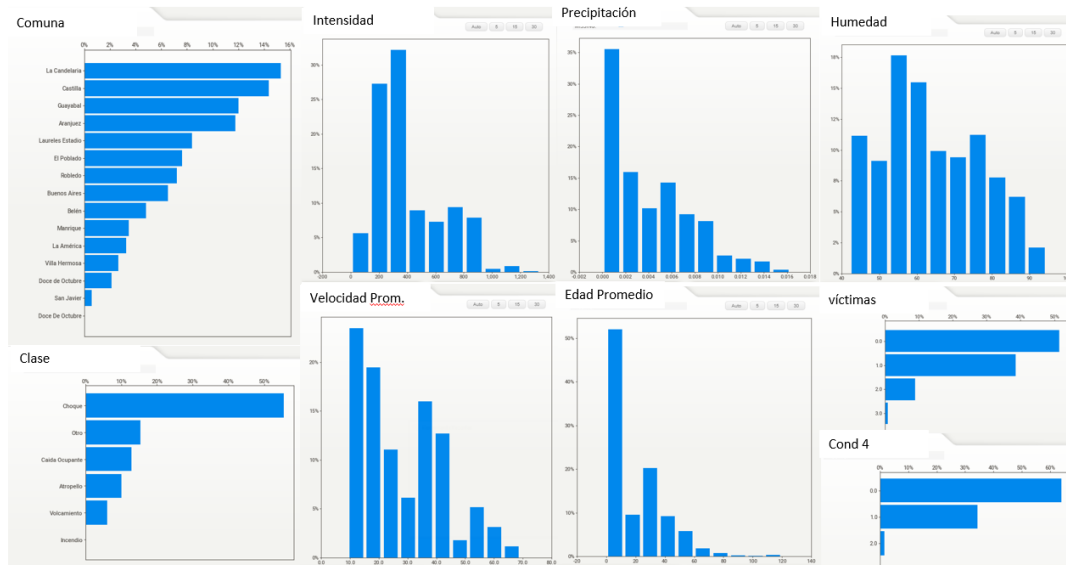


Fig 18. Distribución diferentes variables dentro del clúster 4 para intensidad vehicular

Este grupo se caracteriza por estar distribuido en varias zonas de la ciudad, tiende a tener velocidades tanto altas como bajas, la distribución de humedad esta inclinada a la derecha, es decir, es más alta y la intensidad vehicular tiene a ser media.

Intensidad Vehicular

Clúster 1

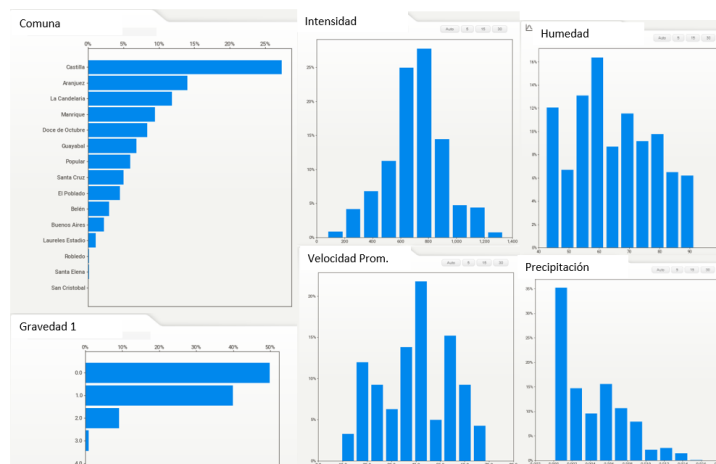


Fig 19. Distribución diferentes variables dentro del clúster 1 para intensidad vehicular

En este clúster encontramos principalmente comunas del centro norte de Medellín, además se caracteriza por ser un grupo donde la velocidad es mayor y una intensidad bien distribuida, humedades bajas y precipitación media, y la gravedad de los incidentes con heridos tiende a ser media.

Clúster 2

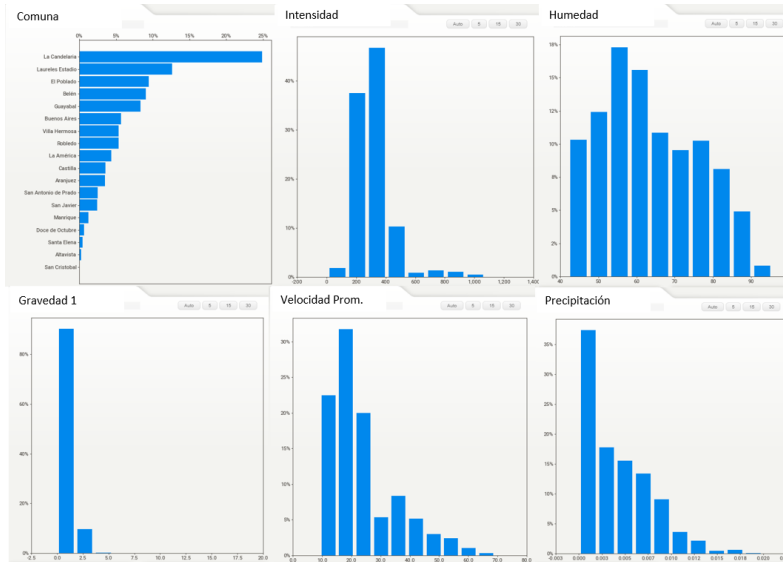


Fig 20. Distribución diferentes variables dentro del clúster 2 para intensidad vehicular

En este clúster encontramos una mayor concentración de todas las comunas de Medellín, además se caracteriza por ser un grupo donde la velocidad tiende a ser menor al igual que la intensidad vehicular, humedades bajas y precipitación baja y la gravedad de los incidentes con heridos tiende a ser baja.

Clúster 3



Fig 21. Distribución diferentes variables dentro del clúster 3 para intensidad vehicular

En este clúster encontramos principalmente comunas del occidente de Medellín, además se caracteriza por ser un grupo donde la velocidad tiende a ser menor al igual que la intensidad vehicular, humedades muy bajas y precipitación media y la gravedad de los incidentes con heridos tiende a ser baja.

Momentos de Aforo Movilidad Ciudad de Medellín

Encuesta Origen – Destino: encontramos a las variables que caracterizan los viajes en el Valle de Aburrá obtenidas a través de la Encuesta Origen Destino realizada por el Área Metropolitana del Valle de Aburrá en el año 2017.

En datos observados de la encuesta origen destino se puede concluir que existen momentos durante el día de mayor y menor aforo de personas movilizándose dentro de la ciudad de Medellín, efectivamente esta es una variable relevante a la hora de minimizar el riesgo de accidentalidad cuando se transporta residuos hospitalarios.

Para este objetivo se realiza una agrupación de datos a partir de la hora de viaje de los ciudadanos, consiste en clasificar los datos de acuerdo con una determinada regla, en este caso es agrupar aquellos momentos de mayor y menor aforo de ciudadanos movilizándose en la ciudad de Medellín.

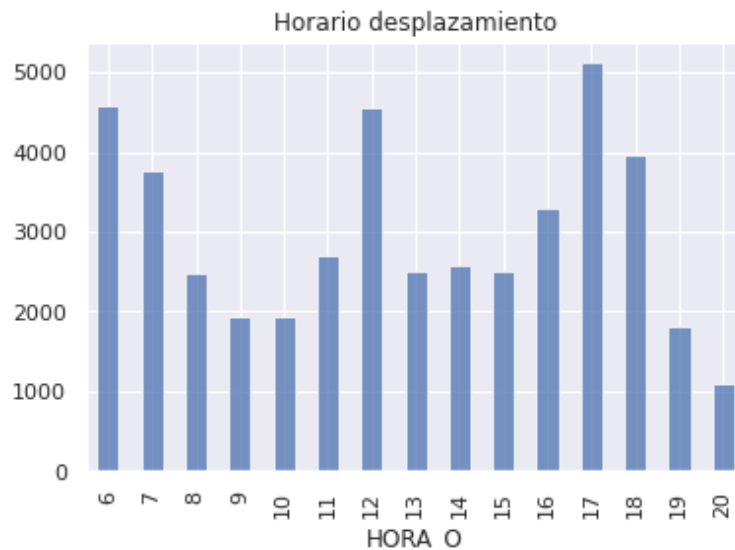


Fig 22. Distribución Horario de desplazamiento origen

Se determina que aforo alto esta dado en los horarios de 6am a 7am en las mañanas, luego al medio día se observa un incremento sustancial en la cantidad de personas desplazándose por lo que también se considera un momento de alto aforo y finalmente entre las 5pm y las 6pm también es un momento del día considerado con aforo alto, quedando de esta manera identificado los momentos de mayor cantidad de personas desplazándose en la ciudad de Medellín. El resto de franja horarias se consideran como momentos de aforo bajo.

A continuación, se observa el mapa de calor de la franja horaria y la comuna, se puede observar que por comuna el aforo de personas es muy estable, exceptuando algunos momentos entre las 5 y las 6 de la tarde que aumenta bastante.

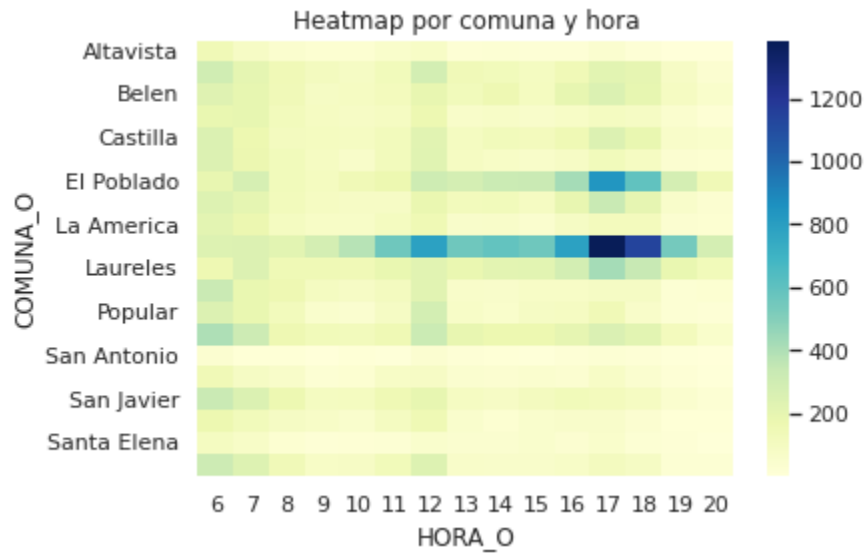


Fig 23. Mapa de calor Horario de desplazamiento origen vs comuna

Al observar los grupos de riesgo por aforos de personas se observa como la indicadora propuesta efectivamente realiza una buena separación del momento de mayor y menor aforo en la ciudad de Medellín.

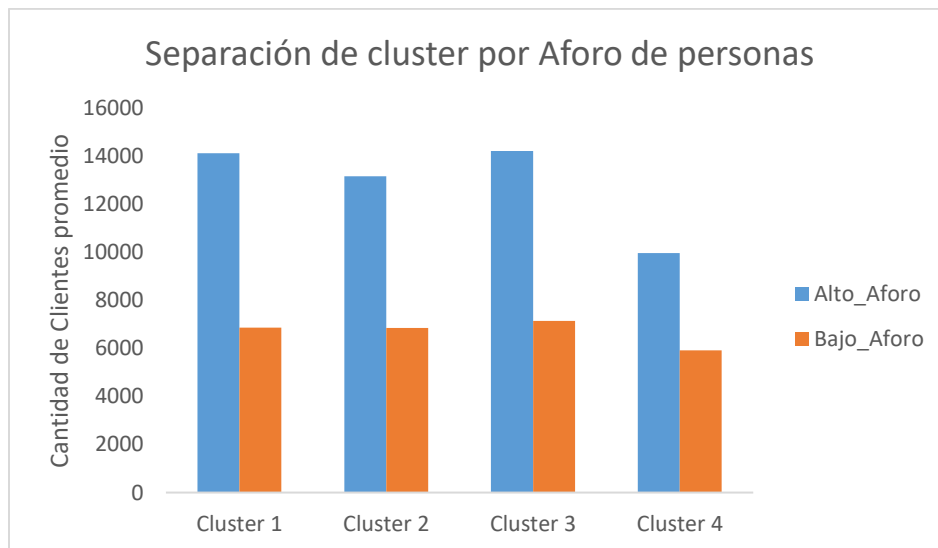


Fig 24. Distribución por clúster incidentes viales y Aforo de personas

Modelo Optimización Ruteo de Vehículos

Para este proyecto se propone combinar lo expuesto anteriormente con un modelo de optimización que permita obtener la mejor ruta a recorrer, a través de la solución de un problema de ruteo de vehículo (VRP) se determina cual es el camino óptimo más seguro por el cual deberá pasar el gestor de residuos para recolectar los residuos hospitalarios atravesando la ciudad de Medellín, para ello se realizan las siguientes definiciones:

- Gestor que realizará la recogida de los residuos hospitalarios.
El IDEAM como instituto encargado de la recopilación de la información del país sobre el manejo de residuos o desechos peligrosos [16], maneja instrumentos que permiten la consulta de la Información registrada por las Autoridades Ambientales, desde su portal se puede acceder al listado de los gestores y se selecciona:

Gestor	Latitud	Longitud
BIOCHEMICAL GROUP S.A.S.	6.377640	-75.449517

- Hospitales a los cuales deberá ir a hacer la recolección. Los hospitales seleccionados para realizar el ejercicio son:

Hospital	Latitud	Longitud
Hospital Pablo Tobón Uribe	6.281586	-75.579192
San Vicente de Paul	6.262143	-75.565859
Hospital León XIII (Hospital Universitario)	6.266538	-75.563776
Clínica Medellín sede poblado	6.209416	-75.571500
GIRARDOTA	6.377640	-75.449517

- Definir la regla que caracteriza el camino de mayor riesgo, a partir del riesgo medido con la ocurrencia de incidentes, de la intensidad vehicular y aforo de personas.

Ruta más corta

Cassingena Navone, [17] indica que el algoritmo de Dijkstra permite encontrar el camino más corto entre los nodos de un grafo. En particular, se puede encontrar el camino más corto desde un nodo (llamado "nodo origen") a todos los demás nodos del grafo, produciendo un árbol de caminos más cortos.

- El Algoritmo de Dijkstra básicamente comienza en el nodo que se elige (el nodo origen) y analiza el grafo para encontrar el camino más corto entre ese nodo y todos los demás nodos del grafo.
- El algoritmo hace un seguimiento de la distancia más corta conocida actualmente entre cada nodo y el nodo de origen y actualiza estos valores si encuentra un camino más corto.
- Una vez que el algoritmo ha encontrado el camino más corto entre el nodo de origen y otro nodo, ese nodo se marca como "visitado" y se añade a la ruta.

- El proceso continúa hasta que todos los nodos del gráfico se han añadido al camino. De este modo, tenemos un camino que conecta el nodo de origen con todos los demás nodos siguiendo el camino más corto posible para llegar a cada nodo. Cada camino mínimo corresponde a la suma de los pesos de las aristas que forman el camino para ir del nodo principal al resto de nodos, pasando únicamente por caminos especiales, es decir nodos marcados

La implementación de este algoritmo se realiza por medio de Python con la librería, [18] la documentación lo define OSMnx, como un paquete de Python que le permite descargar datos geoespaciales de OpenStreetMap y modelar, proyectar, visualizar y analizar redes de calles del mundo real y cualquier otra geometría geoespacial. Puede descargar y modelar redes urbanas transitables a pie, en automóvil o en bicicleta con una sola línea de código Python y luego analizarlas y visualizarlas fácilmente. OSMnx procesa automáticamente la topología de la red a partir de los datos originales sin procesar de OpenStreetMap, de modo que los nodos representan intersecciones/callejones sin salida y los bordes representan los segmentos de calle que los unen.

[19] La documentación indica que este paquete permite construir un mapeo de todos los nodos por donde puede pasar un vehículo y minimizando el tiempo de recorrido encuentra el camino más corto con el algoritmo Dijkstra, a continuación, se presenta en el mapa el camino de cada combinación entre las ubicaciones (en la línea de arriba se encuentra el origen y en la de abajo el destino), a partir de este camino se logra identificar los nodos que guardan las coordenadas por las cuales pasa el recorrido, estas coordenadas son indexadas en H3, y a partir de ello se puede identificar cuantas veces en la ruta cruza por un clúster más o menos riesgoso o por un clúster de mayor o menor intensidad vehicular, de esta manera se define la regla que caracteriza el camino de mayor riesgo y el camino de mayor intensidad:

- Riesgo se define como:

$$r_{ij} = \sum n_k r_k$$

donde n = cantidad hexágonos riesgo k y r = valor del riesgo k

- Intensidad vehicular se define como:

$$i_{ij} = \sum n_j c_k$$

donde n = cantidad hexágonos intensidad k y c = valor de la intensidad k

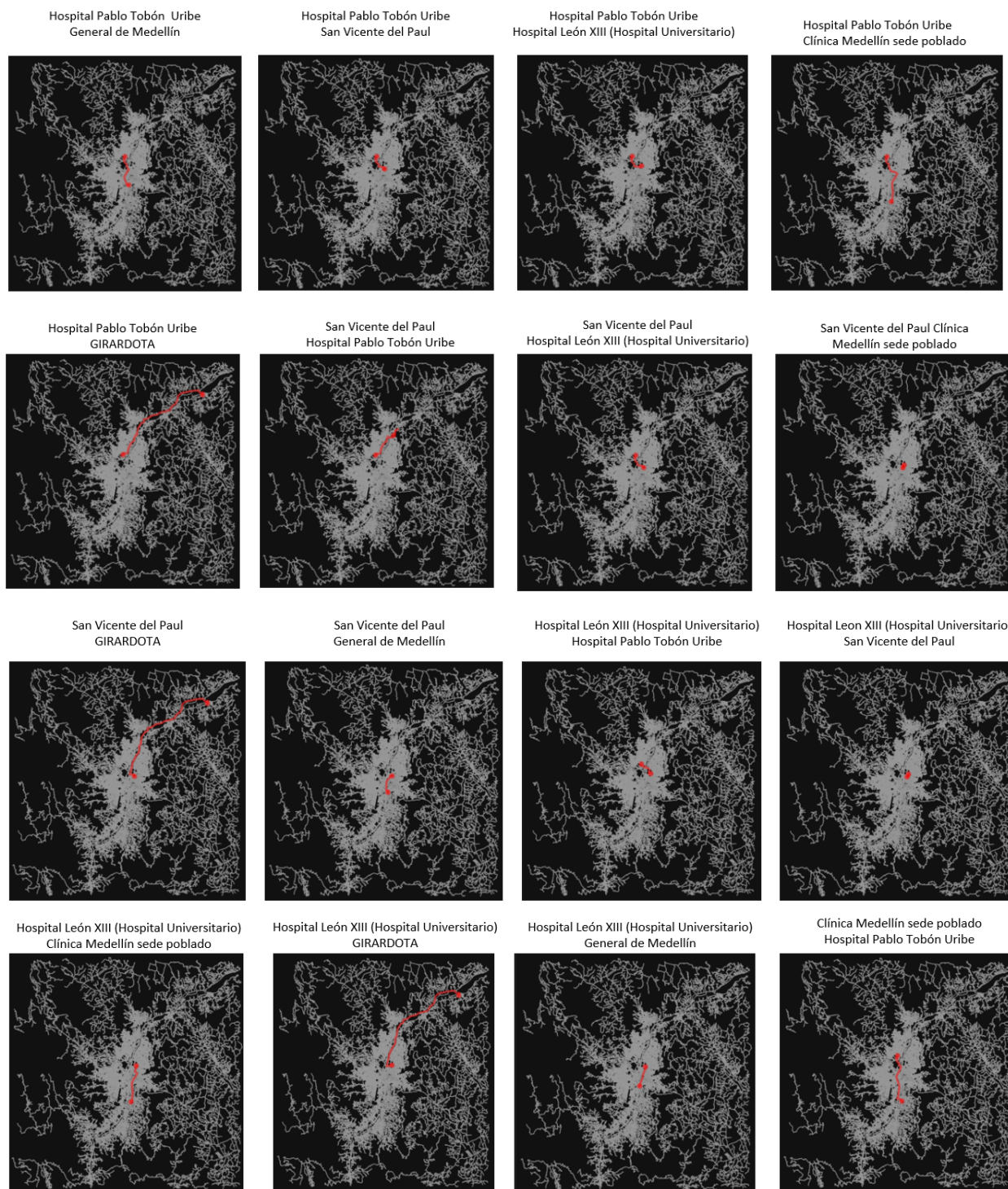


Fig 25. Mapas con camino más corto Origen-Destino

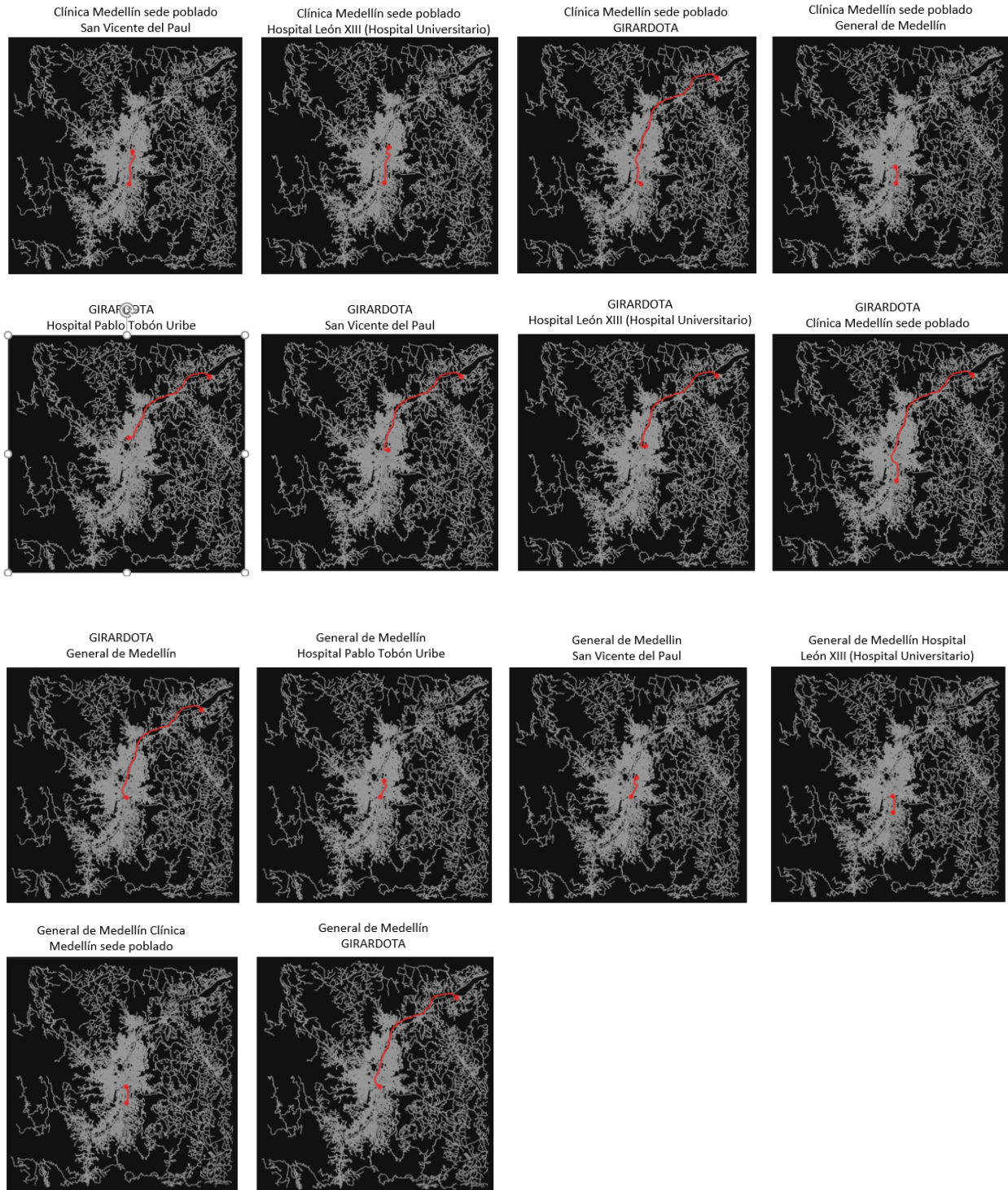


Fig 25. Mapas con camino más corto Origen-Destino

Modelo Optimización Programación Lineal Entera

TSP-Traveling Salesman Problem

Mendoza Casanova en [20] define El Problema del Agente Viajero o TSP por sus siglas en inglés (Traveling Salesman Problem), es un problema clásico y abierto de la optimización combinatoria una de las principales subdisciplinas de la investigación de operaciones. El TSP se puede formular de la siguiente manera: "Dado un conjunto de ciudades, de las cuales se conoce para cada par de ellas, la distancia que las separa, un agente viajero ha de partir de una ciudad de origen y debe visitar exactamente una vez cada ciudad del conjunto, y retornar al punto de partida". Un recorrido con estas características, es conocido dentro de este contexto, como un tour o ciclo hamiltoniano. El problema consiste en encontrar el tour para el cual la distancia total recorrida sea mínima.

Para enunciar el problema formalmente se introducen las siguientes terminologías: Sea un grafo $G = (V, A, C)$ donde V es el conjunto de vértices, A es el de aristas y C_{ij} es la matriz de costes. Esto es, C_{ij} es el coste o distancia de la arista (i, j) .

- Un camino (o cadena) es una sucesión de aristas (e_1, e_2, \dots, e_k) en donde el vértice final de cada arista coincide con el inicial de la siguiente. También puede representarse por la sucesión de vértices utilizados.
- Un camino es simple o elemental si no utiliza el mismo vértice más de una vez.
- Un ciclo es un camino (e_1, e_2, \dots, e_k) en el que el vértice final de e_k coincide con el inicial de e_1 .
- Un ciclo es simple si lo es el camino que lo define.
- Un subtour es un ciclo simple que no pasa por todos los vértices del grafo.
- Un tour o ciclo hamiltoniano es un ciclo simple que pasa por todos los vértices del grafo.

Se plante un modelo de programación lineal entera que ejecuta dependiendo del momento en el que se vaya a hacer el recorrido, haciendo uso de la segmentación por aforo alto y aforo bajo. A continuación, se presenta la formulación del modelo propuesto:

Conjuntos

D: Destinos indexados a i

Parámetros

d_{ij} : Distancia del destino i al destino j

r_{ij} : accidentalidad del destino i al destino j

i_{ij} : intensidad vehicular del destino i al destino j

Decisiones

X_{ij} : Si se va del destino i al j

Objetivo

Diseñar el recorrido para la recolección de residuos hospitalarios en el área metropolitana de tal manera que se minimice la distancia recorrida al igual del riesgo de accidentalidad, por otra parte, se transite por donde haya menos intensidad vehicular.

FORMULACIÓN:
FO (MIN)

$$Z = \sum_{i \in D} \sum_{j \in D | j \neq i} X_{ij} d_{ij} + \sum_{i \in D} \sum_{j \in D | j \neq i} X_{ij} r_{ij} + \sum_{i \in D} \sum_{j \in D | j \neq i} X_{ij} t_{ij}$$

s.a

$$\sum_{i \in D | i \neq j} X_{ij} = 1 \quad \forall j \in D$$

$$\sum_{j \in D | i \neq j} X_{ij} = 1 \quad \forall i \in D$$

$$\sum_{i \in D | i \neq j} \sum_{j \in D} X_{ij} \leq |Q| - 1 \quad \forall Q \in D, Q > 1$$

$$X_{ij} \in \{1,0\} \quad \forall i \in D, \forall j \in D$$

Usando la librería `python_tsp` en Python se resolvió la problemática a partir de los resultados del desarrollo del proyecto y se obtiene el siguiente resultado:

- Para momentos con aforos altos se recomienda seguir la siguiente ruta: 0-4-5-1-2-3
- Para momentos con aforos bajos se recomienda seguir la siguiente ruta: 0-4-5-1-2-3

Ubicación Recursos del Proyecto.

<https://drive.google.com/drive/folders/1CTP9AWCdDpIAW15hoR8eWiUOPOSyC0WM?usp=sharing>

Conclusiones

- A partir de la información de datos abiertos acerca de accidentes georreferenciados se logra segmentar zonas por nivel de riesgo de vial en la ciudad de Medellín, que aportan al diseño de estrategias de seguridad vial para rutas de transporte de residuos hospitalarios. Desde los datos observados se concluye, que las zonas centrales de la ciudad presentan un nivel de riesgo mayor debido a la cantidad de incidentes y al número víctimas en eventos de tránsito, en comparación las cifras en zonas aledañas, esto se debe a la malla vial, actividad comercial y concurrencia de peatones y vehículos.
- Con información de datos abiertos, se logra segmentar zonas de la ciudad de Medellín por nivel de intensidad vehicular, permitiendo identificar regiones de mayor congestión vehicular, es decir que tendrán mayores implicaciones de accidentes viales respecto a colapsos de las vías, heridos, etc. Dentro de las zonas que más se resaltan por intensidad vehicular se observaron la zona norte y la zona del poblado, resultado esperado debido a que el barrio el poblado es una zona turística y de recreación, mientras que la zona norte tienen mayor afluencia de camiones y vehículos transportadores.
- Es posible generar planes de acción para mitigar el riesgo debido a que se puede conocer, con datos de la encuesta de Origen-Destino, la afluencia de personas por horario y comunas con mayor cantidad de movilización dentro de la ciudad de Medellín.
- Se identifica la secuencia de hospitales a visitar considerando los tres tipos de riesgo: de accidentabilidad, intensidad vehicular y aforo de personas, por medio de un modelo de optimización TSP (*Traveling Salesman Problem*), en el cual al calcular las rutas por aforo de personas: alto o bajo, el camino es el mismo, sin embargo, los riesgos asociados a los aforos altos de personas se duplican versus los de bajo aforo de personas.
- Para trabajos futuros, la información de capacidad y tecnología de los vehículos y la generación de residuos hospitalarios por cada entidad prestadora de servicios de salud, afinan el modelo al considerar oferta-demanda, donde oferta es el transporte de residuos.

Referencias

1. Koren H, Bisesi MS (2002) Handbook of environmental health, volume I: biological, chemical, and physical agents of environmentally related disease. CRC Press, Boca Raton
2. Quinto, Y, Jaramillo, L y Cardona, J, (2013). Conocimientos y prácticas de los trabajadores de un hospital sobre el manejo de residuos hospitalarios, Chocó, Colombia, 2012. *Medicas UIS*. Vol. 26, Nº. 1. Bucaramanga, pp. 9-20.
3. IDEAM, Informe Nacional de Residuos o Desechos Peligrosos en Colombia. (2018 y 2019).
4. Álvarez, M., Tamayo, G., Lerma, P. y Castro D. (2016). Conocimientos y prácticas del manejo de los residuos hospitalarios por parte de los fisioterapeutas, Neiva. *Revista Médica de Risaralda*. Vol. 25, Nº 2.
5. Hu H, Li X, Zhang Y et al (2019) Multi-objective location-routing model for hazardous material logistics with traffic restriction constraint in inter-city roads. *Comput Ind Eng* 128:861–876
6. Ghiani G, Laganà D, Manni E, Triki C (2012) Capacitated location of collection sites in an urban waste management system. *Waste Manag* 32:1291–1296
7. Alshraideh H, Qdais HA (2017) Stochastic modeling and optimization of medical waste collection in Northern Jordan. *J Mater Cycles Waste Manag* 19:743–753
8. Zhao J, Huang L, Lee D-H, Peng Q (2016) Improved approaches to the network design problem in regional hazardous waste management systems. *Transp Res part E Logist Transp Rev* 88:52–75
9. Arcgis, Distancia geodésica frente a distancia planar
10. How Is GIScience Different From Geographic Information Systems? September 29, 2021, Recuperado el 08 de octubre de 2022 de <https://onlinedegrees.kent.edu/geography/geographic-information-science/community/how-is-giscience-different-from-geographic-information-systems>
11. Recuperado el 08 de octubre de 2022, de gisgeography.com/giscience-geographic-information-science/
12. Eshwaran Venkat, Uber H3 for Data Analysis with Python, Jan 11, 2021
13. João Paulo Figueira Geospatial Indexing with Uber's H3, Aug 11, 2019
14. Recuperado el 08 de octubre de 2022 de <https://h3geo.org/>
15. Recuperado el 08 de octubre de 2022 de <https://pypi.org/project/sweetviz/>
16. Recuperado el 08 de octubre de 2022 de <http://rua-respel.ideam.gov.co/respelpr2009/mapa.php>

17. Estefania Cassingena Navone, Dijkstra's Shortest Path Algorithm - A Detailed and Visual Introduction, 2020
18. Recuperado el 08 de octubre de 2022 de <https://osmnx.readthedocs.io/en/stable/>
19. Wei-Meng Lee, Visualization in Python: Finding Routes between Points, jan2 2022
20. Mendoza Casanova, Diseño de Algoritmos Heurísticos y Metaheurísticos eficientes para resolver el Problema del Agente Viajero, 2017