

Data Analysis y Clustering para el análisis de crímenes de alto impacto en Barranquilla

Daniel Jose Plata Vecino
Departamento de Ingenierías
Universidad del Norte
Barranquilla, Colombia
dplataj@uninorte.edu.co

Rodolfo Enrique Torres Caballero
Departamento de Ingenierías
Universidad del Norte
Barranquilla, Colombia
etorresr@uninorte.edu.co

Erick Alberto Solano De La Hoz
Departamento de Ingenierías
Universidad del Norte
Barranquilla, Colombia
solanoae@uninorte.edu.co

Abstract— Continuously violence events have kept occurring around the world in the last couple of years, North America, Europe, South America all of them involving massive shootings, political disturbances, sexual assaults, war, gangs and even narcotrafficking. Having that in mind, is important to start analyzing why these events continue happening in our society bringing the sensation we are not safe. The state should start paying more attention to these events to prevent an increment moving forward. In order to accomplish that is important that we take the data, which thanks to the technological advances, is registered on digitally bringing the chance to have analyze it through computer using the Data Analysis tools like the pandas library developed on the programming language Python, that combined with another library like Plotly allows the development of a tool for efficient visualization of this data through graphics and heat maps to identify hotspots for violence so projects for urban violence prevention or social improvement could be developed and addressed by the State therefore, bringing safety to people.

Keywords: *Data Analysis, Machine Learning, Violence, Python, Crime, Criminality, Urban violence, Clustering, Visualization.*

Resumen— Continuamente siguen ocurriendo hechos de violencia alrededor del mundo en los últimos años, Norteamérica, Europa, Sudamérica todos ellos involucrando tiroteos masivos, disturbios políticos, agresiones sexuales, guerra, pandillas e incluso narcotráfico. Teniendo eso en cuenta, es importante comenzar a analizar por qué estos hechos continúan ocurriendo en nuestra sociedad, trayendo la sensación de que no estamos seguros. El estado debería comenzar a prestar más atención a estos eventos para evitar que avance un incremento. Para lograrlo es importante que tomemos los datos, que gracias a los avances tecnológicos, se registran digitalmente brindando la posibilidad de analizarlos a través de la computadora utilizando las herramientas de Análisis de Datos como la biblioteca pandas desarrollada en el lenguaje de programación Python, que combinado con otra biblioteca como Plotly permite el desarrollo de una herramienta para la visualización eficiente de estos datos a través de gráficos y mapas de calor para identificar puntos críticos de violencia para que el Estado pueda desarrollar y abordar proyectos de prevención de violencia urbana o mejora social, por lo tanto, brindando seguridad a la gente.

Palabras clave: *Análisis de datos, Visualización, Violencia, Python, Crimen, Criminalidad, Violencia urbana*

I. INTRODUCCION

Este proyecto se implementará un prototipo para el análisis y la predicción del estado de los crímenes de alto impacto como pueden ser, hurtos de distinta índole, homicidios, casos de abuso sexual, casos de violencia domestica o intrafamiliar, entre otros, en la ciudad de Barranquilla que se han visto en aumento en los últimos años, un ejemplo de esto es la nota informativa titulada como “Balance semestral de seguridad ciudadana en Barranquilla y su Área Metropolitana” realizada por J. Melamed (2022) la cual resalta que en el 2021 que fue el año hasta ese entonces con una tasa más alta de violencia de los últimos 4 años en la ciudad de Barranquilla, llega a ser superado en el primer semestre del 2022 donde, en el cual hubo un aumento del 10.5% en la tasa de homicidios. A pesar de que la finalidad de este proyecto no abarca la

solución de estas problemáticas, se pretende, desde el área de ingeniería de sistemas y software, el desarrollo de una herramienta la cual pretende que, en primera instancia, a través de un dashboard, como el propuesto por P. Kumar y B. Balamurugan (2022), en el cual se muestren estadísticas descriptivas y representaciones de los crímenes de alto impacto en la ciudad de Barranquilla, que, entre otras cosas, permitirá la caracterización de los distintos casos de violencia anteriormente con el fin de identificar áreas críticas o “hotspots” como se expone en “The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime” (S. Chainey, 2008) aplicado a la región de la ciudad de Barranquilla. Este proyecto contará con 3 (tres) componentes teóricos, en el primero se revisará el aspecto social del proyecto, a que se refieren los índices de criminalidad y accidentalidad, que miden, definiciones de los factores que lo componen y su impacto en la sociedad y como estos han sido afrontados por países del primer mundo como U.S.A. En el componente número dos se hablará de Data Analysis y Clustering breve historia, métodos y aplicaciones, algoritmos de clasificación. Finalmente, el tercer componente ondearemos en la arquitectura del proyecto, implementación, objetivos del proyecto y resultados esperados.

II. PROBLEMA

En los últimos años ha habido un incremento en la violencia, criminalidad e intolerancia que no solo son notables en la ciudad de Barranquilla y sus alrededores sino, en toda Colombia y en el mundo. Ya se hace costumbre ver en los medios informativos un nuevo caso de masacre en Estados Unidos, o como las manifestaciones políticas terminan en disturbios y en Colombia, el creciente número de casos de asesinatos, hurtos y masacres que se dan tanto a nivel urbano, como rural. A comienzo del presente año, Alicia Méndez en su nota para el diario el tiempo en su portal electrónico describe como ha sido el comportamiento en los últimos años de los homicidios en Colombia, analizando los datos presentados por la Policía Nacional llegando al resultado de un incremento del 8.2% entre el año 2019 y el 2021 (A. Méndez, 2022). Dos décadas atrás, V. Zubillaga y R. (2001) Nos hablan sobre las dimensiones y construcciones de la violencia en América Latina y nos dicen que para en ese entonces ya estaba en incremento la violencia urbana, por lo cual se obliga a dirigir la mirada hacia múltiples frentes y hacia una diversidad de actores, escenarios y lógicas que intervienen simultáneamente y le dan forma al estado actual de las violencias que se vive a diario en nuestras colectividades. Con lo anterior, se entiende que la violencia ya no solo se limita a los asesinatos o masacres, sino a un grupo de situaciones que impactan no solo físicamente a una sociedad, sino psicológicamente lo que afecta a la sensación de seguridad que se percibe en un espacio geográfico determinado. Es por eso por lo que en este proyecto se plantea el desarrollo de un prototipo de software, el cual nos permita mapear y visualizar las violencias en la ciudad de Barranquilla-Colombia con el fin de identificar zonas focos de violencia y criminalidad sirviendo de base para propuestas de planes de mejoramiento social o de prevención haciendo uso de tecnologías para el Análisis y manipulación de datos, así como de visualización que ofrece Python y Google Colab.

III. JUSTIFICACIÓN

A lo largo de su historia, Colombia ha sufrido por distintos conflictos internos pero los años 1930 y 1948 muestran un cambio considerable en estos datos puesto que se agudizó esta problemática (G. Guzmán, O. Fals, E. Umaña, 2019). La crisis por los casos de violencia es algo contra lo que todo el país ha estado luchando durante mucho tiempo. Sin embargo, si se revisan periódicos o noticieros, se puede notar que los casos de violencia más agresivos se reportan al interior del país o de las zonas paramilitarizadas, por lo tanto, es lógico que el mayor esfuerzo por disminuir la violencia se de en estos lugares. Aun así, los casos que no representan un peligro global para el pueblo colombiano son igual de importantes de tratar. Barranquilla, una ciudad de la costa atlántica y que no se encuentra paramilitarizada, para el año 2022 mostro un incremento significativo en sus casos de violencia. La gravedad de la problemática fue tal que el alcalde Pumarejo, pronunció que se trabajaría en conjunto con el gobierno del presidente Petro para controlar esta situación (ContentEngine, 2022). Los casos de violencia, según los índices que menciona el defensor del pueblo Carlos Camargo en una entrevista con Radio Nacional de Colombia, muestran un incremento preocupante, es que solo por poner dos casos de ejemplo, los homicidios aumentaron en un 25% y las extorsiones en un 128% (S. Salas, 2022), mientras que, en datos actuales y más puntuales, el mes de Enero del presente año, termino con 74 homicidios teniendo más de 20 casos, en comparación con Enero del año pasado, según lo reportado por D. Ortega (2023) para el diario El Tiempo. Tomando como punto de partida los aumentos de los casos de violencia que se han presentado en la ciudad de Barranquilla en el presente año, se ha optado por desarrollar un proyecto que muestre de forma estadística y mediante un mapa de calor de la ciudad los casos que aquejan más a los ciudadanos barranquilleros. El avance tecnológico en temas como la violencia urbana como se discute en la investigación realizada por T. Cheng & T. Chen (2021) la cual menciona el progreso de las tecnologías de Big Data en este ámbito, violencia intrafamiliar y accidentalidad vial se encuentran limitados en la ciudad de Barranquilla. Por tal motivo se busca desarrollar una herramienta que permita procesar los datos con el fin de generar un dashboard para la generación de gráficos descriptivos con lo cual se pueda hacer un análisis a la información de los crímenes de alto impacto como el realizado por G. García (2020) con su herramienta “Mirante”, en los distintos barrios de la ciudad de Barranquilla con base en métodos de Data Analysis y Clustering. Esta herramienta permitirá tratar esta información como base informativa y visual para el desarrollo otros proyectos ya sean de tipo académicos, civiles, educativos etc. que busquen favorecer y/o estudiar el comportamiento de los casos de violencia en la ciudad para que así puedan disminuir de una forma que beneficie a todos los miembros de la sociedad.

IV. OBJETIVO GENERAL

Diseñar y desarrollar un mecanismo de visualización de datos que represente de forma gráfica los indicadores de inseguridad en los distintos sectores de la ciudad de Barranquilla

V. OBJETIVOS ESPECIFICOS

- Identificar los componentes claves para el análisis y la visualización, desde las ciencias de la computación, de la violencia, violencia urbana y criminalidad a partir de una revisión sistemática de la literatura relacionada.
- Diseñar la arquitectura de solución para la implementación de un prototipo software que permita el análisis de un dataset contenedor de los casos reportados de crímenes y manifestaciones de violencia en la ciudad de barranquilla.
- Desarrollar un prototipo de la solución y un dashboard con el cual se pueda visualizar mediante gráficos de barra, gráficos de dispersión y tablas, la distribución y los clústeres asignados, para los casos de violencias presentadas en la región de Barranquilla.
- Validar que el prototipo de la solución genere los gráficos y tablas adecuadamente para la visualización de violencia con lo que se pueda dictar una aproximación a de sectores críticos en la región de la ciudad de Barranquilla.

VI. METODOLOGÍA

Para llevar a cabo este Proyecto se utilizará una metodología agile tipo Scrum, la cual consta de dividir el proyecto por sprint y, que al final del proyecto nos permita desarrollar un tablero digital utilizando Data Analysis.

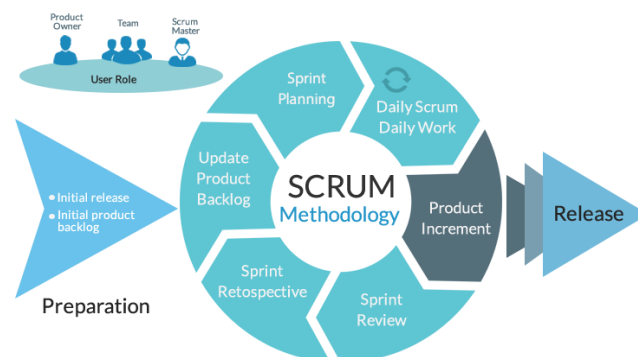


Figura 1. Metodología Scrum

Fase de investigación.

En esta fase tuvimos en cuenta la problemática que se consideró como tema central de nuestro proyecto posteriormente, mediante investigaciones exhaustiva y consultas sobre el tema se llegó a la conclusión de desarrollar un prototipo que muestre un mapa de calor con las zonas más afectadas por los distintos tipos de delitos, posterior a esto se empezó a consultar artículos o soluciones similares a lo que se busca desarrollar para luego conseguir un elemento fundamental en el diseño de la solución, los datos. Los cuales se solicitaron a la Policía Nacional de Colombia.

Fase de diseño de arquitecturas.

Con la información obtenida en la fase anterior se empezó a desarrollar diagramas y esquemas de lo que sería nuestra solución a esta problemática. Aquí se tomó decisiones sobre las tecnologías que se debían implementar, los objetivos y requerimientos que se querían cumplir. Una vez se obtuvo toda esta información se diseñó una arquitectura lógica y los mockups que sirvieron como guía para nuestro prototipo.

Fase de desarrollo e implementación.

Usando lo planteado en fases anteriores y referenciándonos en la arquitectura lógica, la cual facilitó el desarrollo. Con todo lo planteado, finalmente se comenzó una fase de programación del prototipo mediante varios Sprint semanales, los cuales fueron definidos mediante la metodología.

Fase de evaluación y resultados.

El prototipo fue evaluado por el equipo, con esta evaluación se buscó que cumpliera con lo planteado inicialmente. Esta evaluación registrada por el equipo facilitó la corrección y modificación del prototipo durante la fase de desarrollo, puesto que cuando los criterios no se cumplían se reprogramaba el aplicativo. Finalmente, se realizó una evaluación entre pares la cual fue más detallada y se realizó mediante diferentes criterios para así buscar que se cumplieran los objetivos del Proyecto de manera óptima.

VII. MARCO TEORICO

Ondeando en la historia de Colombia, luego de que nuestro proceso independentista se concluyó, el país ha brillado por sus múltiples enfoques, políticos y económicos a través de los años como se refleja en (R. Junguito & H. Rincón, 2004). De acuerdo con (M. Medina, 1989) en su texto: "Bases Urbanas de la violencia en Colombia", a finales de los años 40 había en total 126.297 muertos por la violencia en Colombia a razón de los distintos fenómenos sociopolíticos como las concentraciones gaitanistas y socioeconómicos como manifestaciones obreras que eran reprimidas a través de hechos violentos. Históricamente la violencia en Colombia nace como una serie de confrontaciones a nivel político entre dos principales posturas y que, hasta mitad del siglo XX, solo tenía impacto a nivel rural, pero con el paso del tiempo y en la prolongación del clima del conflicto, terminaría repercutiendo al sector urbano del país. Ya habiendo mencionado el primer gran periodo de auge de la violencia en Colombia (1948-1966), el segundo se presentaría entre 1980 y 1993 en donde llega al punto máximo en el cual había 79 homicidios por cada 100.000 habitantes según las cifras presentadas en "A Comparative Statistical Note on Homicide Rates in Colombia" (A. Villaveces, 2021). Algunos autores como: (M. Llorente, 2001), sugieren que este comportamiento no solo se debió a la violencia política (primer aumento) y el narcotráfico junto con las crecientes redes criminales (segundo aumento), sino también a la violencia familiar y al abuso infantil. Existe una tercera de auge en la violencia en Colombia, expuesta en: "La Violencia en Colombia: Análisis Histórico del homicidio en la Segunda Mitad del Siglo XX" (C. Bello Montes, 2008) que va desde mediados de la década de los 90s hasta 2006, periodos en donde se comienzan a realizar una presión internacional e interna, por parte del estado para la erradicación del narcotráfico, enfrentamiento a las guerrillas y

aparición de grupos paramilitares. Colombia, que desde el primer periodo de tiempo anteriormente mencionado (mitad del siglo XX), era un país predominantemente rural y en el cual, el 70% de la población pasa a asentarse en el territorio urbano por lo que se acentúa la responsabilidad del Estado frente al tratamiento de las problemáticas, anteriormente mencionadas.

Países como Estados Unidos, una de las principales potencias mundiales, no es ajeno a la violencia, particularmente, la violencia urbana asociada al uso y porte de armas por parte de la población civil. El número de asesinatos en el país creció cerca de un 30% en el 2020 con respecto al año anterior siendo este el aumento más pronunciado de este crimen, desde que el FBI comenzó a recopilar datos en la década de 1960 como se expone en el artículo: "¿Cómo se explica el histórico aumento de los homicidios en estados unidos?" presentado por la BBC (British Broadcasting Corporation) en su sección de noticias. Cifras preocupantes también se muestra en el artículo "There have been nearly 600 mass shootings so far in 2022" escrito por (J. Ledur, 2022) para el The Washington Post, que reporta cerca de 600 tiroteos masivos para la fecha del 14 de noviembre del presente año. Expandiendo un poco más el horizonte, el bienestar civil no solo se limita a casos de asesinato, existen muchas más formas en las que se puede presentar la violencia, y una de la cual año tras año se habla más, y de la que se crea conciencia es la violencia de género, maltrato intrafamiliar, violencia física o psicología a pareja íntima, de acuerdo con la encuesta Nacional de violencia íntima y sexual realizada por el Centro Nacional para la prevención y el control de daños (CDC) en inglés, alrededor de 24 personas por minuto son víctimas de violación sexual, violencia física o acoso por un compañero íntimo. Otra manifestación de violencia en esta área es la violencia a los niños y adolescentes, en el 2020 se reportaron 3.9 millones de casos de maltrato infantil de los cuales 90% involucraban uno o ambos padres, todo esto generando en última instancia unas consecuencias y factores de riesgo que terminan impactando en la sociedad en la que estos individuos se terminaran desenvolviendo como se expone en el artículo "Alcohol abuse as a risk factor for and consequence of child abuse" (C.S. Widom, 2022). Adicionalmente existen muchas más causales que impactan y nos dicen el estado de un determinado sector como lo son la presencia de bandas criminales, bandas/individuos dedicados al microtráfico, muertes por accidentes de tránsito, casos de intolerancia etc.

Después de dos guerras mundiales y una guerra fría, aun vemos como en el mundo la violencia continúa siendo algo del diario vivir a tal punto que no nos sorprende en alta medida cuando vemos que en uno u otro lado estalla una guerra, cuando se presenta un caso de abuso sexual, o un nuevo caso de acoso o abuso de autoridad y que finaliza en la terminación de una, varias o muchas vidas, como los múltiples hechos violentos, que, en 2020 se reportaron en Los Estados Unidos, 11 Norte Americanos murieron mientras participaban en manifestaciones políticas, y otros 14 en otros incidentes relacionados a agitaciones políticas presentados en el portal The Guardian, en una nota publicada por Lois Beckett, donde hace recuerdo de cada uno de las muertes. En Colombia en 2021, se reportaron al menos 24 fallecidos y 87 desaparecidos en protestas y disturbios según se cuenta en el artículo "Colombia police respond to protests with bullets, and death toll mounts" en el New York Times (J. Turkewitz, 2021). De acuerdo con la Organización Mundial de la Salud (OMS), la violencia es un problema de salud pública, así como es

mencionado por (L. Larizgoitia, 2009), por lo tanto es pertinente e importante que el estado, no solo la atienda, sino que la prevenga y esto se consigue mediante la implementación de programas preventivos que en última medida buscan un mejoramiento social actuando en espacios específicos, como pueden ser las escuelas primarias, como sugiere (A. Farrel, 2021) en su artículo: “Development and evaluation of school-based violence prevention programs”, barrios/localidades, como se menciona en: “Neighborhood interventions to reduce violence” (M. Kondo, 2018) que se han determinado como focos de violencia. Un ejemplo de estos es la iniciativa “One Vision, One life”, la cual se realizó en determinados sectores de la ciudad de Pittsburgh, Pensilvania, en donde los llamados “barrios bajos” se notaba casos de alienación, racismo y discriminación para la población afroamericana que terminaron generando un pensamiento en el que, por ejemplo, el respeto se le concede a quienes roban o responden a situaciones conflictivas con violencia. El objetivo de la iniciativa fue abordar, entender y encontrar formas de cambiar este contexto (J. Wilson, 2011). Entendiendo lo anteriormente mencionado, sale a la luz la importancia de que en la sociedad se mantenga un registro de todos estos casos de crimen y violencia, permitiendo así que estos puedan ser analizados computacionalmente aprovechando el avance tecnológico y el desarrollo que han tenido las ciencias de la computación hasta el día de hoy, con lo que se puede generar como resultados aplicativos para mostrar tendencias, visualización, como se pretende en “Crime mapping in Nigeria using GIS” (J. Wilson, 2011), detección, mostrado por (M. Baba, 2019) etc. que ayuden al entendimiento de este fenómeno social y se puedan aplicar medidas para corregirlo a través de programas de prevención, como los mencionados anteriormente.

Ahora, para que haya una intervención, una realización de programas preventivos y que en base a ello se espere un mejoramiento social, es imperativo que se reconozcan que sectores deben ser atendidos, que situaciones se están dando en ese determinado espacio y claramente que se determine y es en este aspecto en donde las ciencias de la computación tienen altas posibilidades de impacto. Desde estas se abordan áreas asociadas a, sin lugar a duda, la salud, la psicología como él se muestra en el artículo “Development of software that supports the improvement of the empathy in children with autism spectrum disorder”, una propuesta que pretende atender a la población diagnosticada con autismo (R. Muñoz, 2012). Incluso en negocios desde micro a macroempresas, por ejemplo, el acercamiento presentado por: (A. Wannes, 2020) en su artículo “KPI-based approach for Business Process Improvement” en el que se propone un desarrollo de software para el mejoramiento del proceso del negocio. Con el avance de la tecnología se ha a la capacidad de mapear (Mapping) y por ende de visualizar algún tipo de fenómeno siempre y cuando, tengamos los datos que lo representan, por lo tanto, se puede emplear en múltiples propósitos como, por ejemplo, el mapeo de la presencia de actores armados en Colombia (J. Osorio, 2019), otro ejemplo es el mapeo de incidencias de COVID-19 en los Estados Unidos continental (A. Mollalo, 2020). De la mano de la visualización, usualmente se encuentra que los datos representan de una u otra forma con el fin de mostrar o hablar sobre una característica por lo que conviene hablar de Data Analysis, que, según se expone en “Transforming Unstructured Data into Useful Information” el Data Analysis es el proceso de inspeccionar, depurar, transformar y modelar datos, con el objetivo de informar

conclusiones y proveer soporte a la toma de decisiones (M. Brown, 2014), esta tiene múltiples facetas y acercamientos, enmarcando diversas técnicas bajo una variedad de nombres y es usada el área de negocios, ciencia y ciencias sociales (C. Proneau, 2017). Existen lenguajes de programación como Python, que cuentan con librerías como Pandas, que nos permiten la manipulación y el análisis de datos que combinadas con otras como Plotly que contiene un gran número de características para el graficado, si se le agrega un dataset a esta combinación se crea una herramienta de visualización en donde se muestren datos caracterizados con el fin de informar una conclusión que para propósitos de este proyecto, en qué punto se encuentra sensación de seguridad en la ciudad de Barranquilla basándonos en los casos de criminalidad, violencia y accidentalidad reportados.

VIII. MARCO CONCEPTUAL

Sensación de seguridad

Este proyecto conceptualmente cuenta con dos énfasis, el primero es el énfasis social y el otro es el énfasis tecnológico del mismo. En el énfasis social, este proyecto abarca a lo que se conoce como la sensación de seguridad la cual se refiere a la interacción de los individuos con su entorno es constante y en ella territorio y colectividad se modifican mutuamente. En esta dinámica, la sensación de seguridad es una pieza clave para entender el uso del espacio público, además de un indicador muy importante de la calidad de vida de las personas (P. Fraile, 2007).

Violencia

Así mismo, es importante tener el concepto de violencia el cual es definido por Krug, Mercy, Dahlberg y Zwi (2002) como el uso intencional de fuerza o poder físico hacia uno mismo, otra persona o hacia un grupo o comunidad que resulte en o tenga altas probabilidades de resultar en lesiones, muerte, daño psicológico, mal desarrollo o depravación.

Crimen

Teniendo en cuenta que, en la mayoría de las constituciones o leyes de los distintos países del mundo, la mayoría de actos violentos son considerados un crimen, es conveniente definirlo, por lo cual de acuerdo con Martín (2009) este es acto que es dañino no solo a un individuo sino, a una Comunidad, Sociedad o Estado que por lo tanto es prohibido o castigable por la ley.

Análisis de datos (Data Analysis)

Ahora, en énfasis tecnológico del proyecto se dirige hacia el área de las ciencias de la computación, conocida como Análisis de Datos o mejor referenciado como Data Analysis, el cual es un proceso por el cual se obtienen datos brutos, para que, subsecuentemente convertirse en información útil para un usuario a la hora de toma de decisiones. (C. Judd, M. McClelland, C. Ryan, 2017)

Análisis de datos Exploratorio (Exploratory Data Analysis)

El Data Analysis se divide en dos conceptos de acuerdo con la finalidad que se le quiera dar al mismo. Tenemos por un lado el Análisis de datos Exploratorio cuyo objetivo de su

aplicación consiste en el descubrimiento de nuevas características en los datos (J. Tukey, 1977)

Análisis de datos confirmatorio (Confirmatory Data Analysis)

Por otro lado, tenemos el análisis de datos confirmatorio cuyo propósito se centra en confirmar o negar hipótesis existentes. (E. Karageorgiou, 2011).

Visualización de Datos

En la implementación de este proyecto se requiere que los datos que se manejan a través del Data Analysis, sean capaz de visualizarse, es por lo que se incluye la visualización de datos o Data Visualization el cual es un campo interdisciplinario que se encargara de la representación gráfica de los datos y la información. Es una particularmente eficiente forma de comunicación cuando de datos o información numerosa se trata (C. Knafllic, 2015).

Python

Siguiendo con la implementación, Python que es un lenguaje de programación de alto nivel y de propósitos generales, su diseño y filosofía enfatizan en la legibilidad del código y su indexación significativa (D. Kuhlman, 2009), el cual es comúnmente usado, por sus características anteriormente mencionadas para desarrollos que optan por aplicar técnicas de DA y ML

Pandas (Librería Python)

Por el lado del Data Analysis, Python cuenta con la librería Pandas, la cual es una herramienta poderosa, rápida, flexible y de fácil uso de código abierto con el propósito de poder analizar y manipular de datos (W. McKinney, 2015).

Plotly (Librería Python)

En cuanto al tema de la visualización de los datos, Python también cuenta con una librería para cumplir con este propósito, la cual se le conoce como Plotly, la cual es de código abierto que permite crear gráficos de alta calidad. (Python Documentation, 2022)

Clustering

En el mundo del Machine Learning, existe lo que se conoce como Clustering la cual es la tarea de agrupar un grupo de objetos en un mismo grupo a lo que llamaremos “Cluster” de tal manera que estos objetos que están en em mismo subgrupo compartan características similares. (P. Tan, 2013)

KMeans

Para el Clustering existen muchos algoritmos y formas de hacerlo efectivo, en este ejercicio, nos vamos por el lado no supervisado del machine learning, con el método K-Means o K-medias. KMeans, consiste en encontrar puntos, valores en nuestros datos a los que se les conocerá como centros o centroides, serán ellos los que representarán sobre qué punto un determinado grupo de objetos se concentra. El concepto se presenta inicialmente por James MacQueen in 1967 en “Classification and analysis of multivariate observations”.

IX. REVISION SISTEMATICA DE LA LITERATURA

Bases de datos digitales: IEEE Xplore, ACM Digital Library

Criterios de exclusión: últimos 5 años, Journals

Fuente de datos/elementos de búsqueda	Violence	Violence & Data Analysis	Crime & Forecasting
IEEE Xplore	959	231	824
ACM Digital Library	6656	6234	1,208
Last 5 years & Journals	-	-	-
IEEE Xplore	59	23	75
ACM Digital Library	668	664	138

Tabla 1. Criterios de Búsqueda R.S.L

X. ARQUITECTURA LÓGICA DE LA SOLUCIÓN

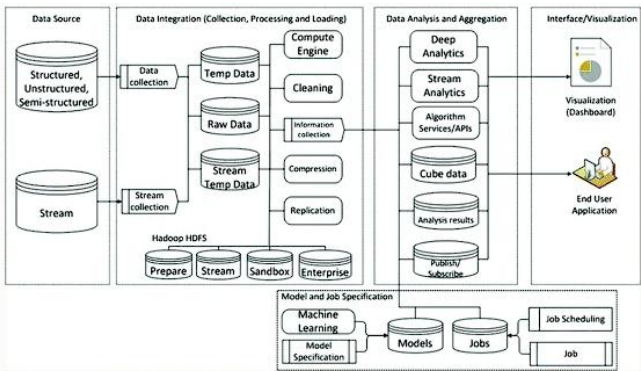


Figura 2. Arquitectura de referencia para proyectos de BD o DA

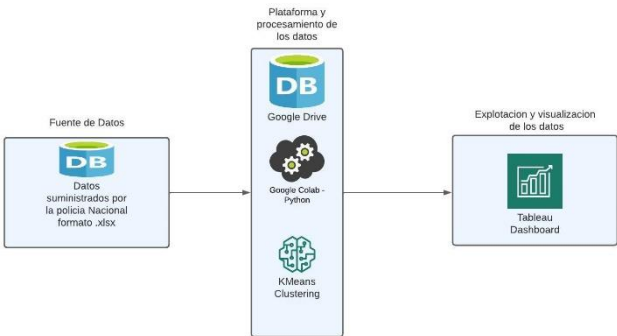


Figura 3. Arquitectura lógica de la solución

XI. PROTOTIPO

```
#Definición de categorías
categoria_dias = ['LUNES', 'MARTES', 'MIÉRCOLES', 'JUEVES', 'VIERNES', 'SÁBADO', 'DOMINGO']
categoria_meses = ['ENERO', 'FEBRERO', 'MARZO', 'ABRIL', 'MAYO', 'JUNIO', 'JULIO', 'AGOSTO', 'SEPTIEMBRE', 'OCTUBRE', 'NOVIEMBRE', 'DICIEMBRE']
categoria_año = [2018, 2019, 2020, 2021, 2022, 2023]

#Se transforman la columna día
encoder = OrdinalEncoder(categories = [categoria_dias])
features[['DIA']] = encoder.fit_transform(features[['DIA']])

#Se transforman la columna mes
encoder = OrdinalEncoder(categories = [categoria_meses])
features[['MES']] = encoder.fit_transform(features[['MES']])

#Se transforman la columna año
encoder = OrdinalEncoder(categories = [categoria_año])
features[['AÑO']] = encoder.fit_transform(features[['AÑO']])

#Para las variables nominales se usa el LabelEncoder
nominales = ['DELITOS', 'BARRIDO_HECHO', 'GENERO']
nominales = ['GENERO']
label_encoder = LabelEncoder()
```

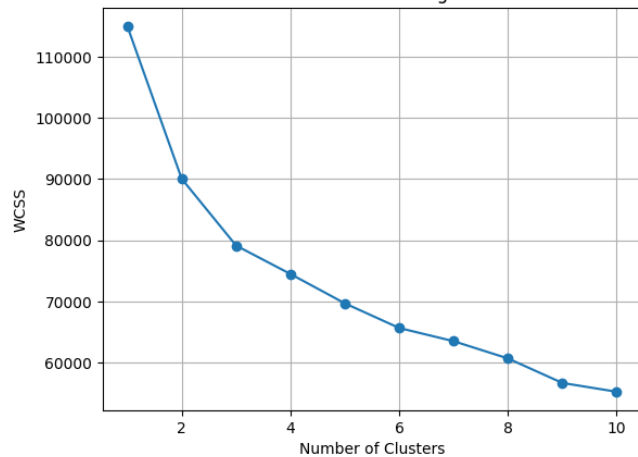
#Min-Max Escalation

```
features_norm = (features - features.min()) / (features.max() - features.min())
features_norm.describe()
```

```
WCSS = []

for i in range(1, 11):
    kmeans = KMeans(n_clusters = i, max_iter = 300)
    kmeans.fit(features_norm)
    wcss.append(kmeans.inertia_)
```

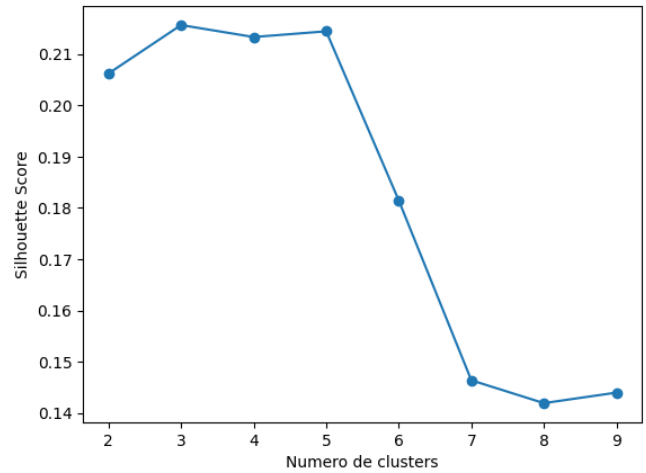
Elbow Method for determining # of clusters



```
from sklearn.metrics import silhouette_score
silhouette_scores = []

for i in range(2, 10):
    kmeans = KMeans(n_clusters=i, init='k-means++', max_iter=300, n_init=10, random_state=42)
    kmeans.fit(features_norm)
    silhouette_scores.append(silhouette_score(features_norm, kmeans.labels_))
```

Silhouette Method



```
#Iniciamos el algoritmo y se lo asignamos en variables, numero de clusters y maximo de iteraciones del algoritmo
clustering = KMeans(n_clusters = 3, max_iter = 300)

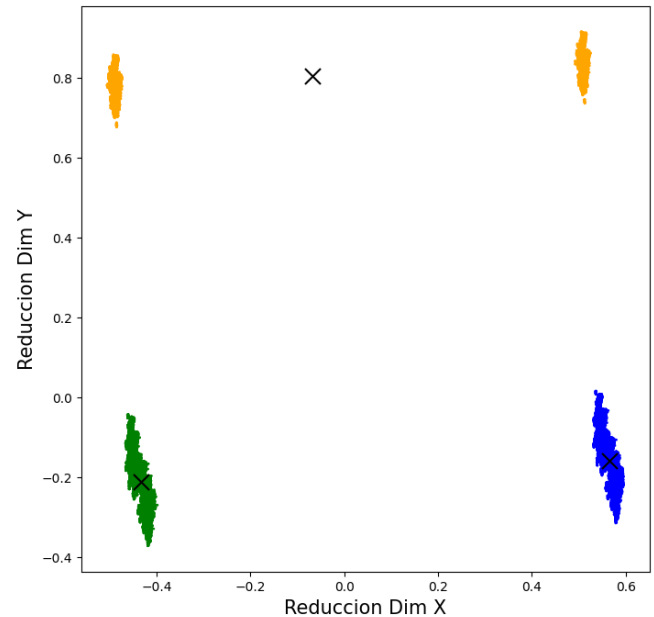
#Ajustamos el algoritmo k means a los datos normalizados/escalados/transformados
clustering.fit(features_norm)

#Guardamos los centroides resultados del algoritmo
centroids = clustering.cluster_centers_
```

```
from sklearn.decomposition import PCA

pca = PCA(n_components=2)
pca_features = pca.fit_transform(features_norm)
pca_centroids = pca.transform(centroids)
pca_features_df = pd.DataFrame(data = pca_features, columns = ['Reduccion_Dim_X', 'Reduccion_Dim_Y'])
pca_features_clusters = pd.concat([pca_features_df, features[['KMeans_clusters']], axis=1)
```

Reduccion a dimension = 2



XII. CONCLUSIONES Y RESULTADOS

Basandonos en proceso realizado en este Proyecto Podemos decir que se cumplieron los objetivos planteados para con la investigación y el desarrollo. Revisando los objetivos

planteados, se logra identificar los componentes claves para un análisis de datos exploratorio y las formas existentes para la visualización de los mismos utilizando el lenguaje Python y sus distintas librerías como pandas y matplotlib, así mismo, se diseña una arquitectura de la solución basados en arquitecturas de referencia para distintos proyectos de analítica de datos y big data que muestre el proceso de un proyecto de analítica de datos. El desarrollo de la solución se lleva a cabo usando la herramienta Google Colab y Python, herramienta la cual entre otras cosas, nos dirá como se distribuyen nuestros datos y la clusterización generada de los mismos, así como una descripción de estos, zonas críticas en la ciudad, gráficos y estadísticas para comprender los mismos así mismo, se valida que estas salidas de la herramienta, representen adecuadamente los datos que se usaron para este proyecto y se pueda obtener una comprensión de los mismos. De este proyecto se espera ayude a futuras investigaciones relacionadas con la violencia en la ciudad de Barranquilla y deseen identificar aquellas zonas críticas de la ciudad, para así proponer ideas de mejora para aquellos sectores focos de la criminalidad en la ciudad.

XIII. REFERENCIAS

- Melamed, J. (2022, August 10). Balance semestral de Seguridad Ciudadana en barranquilla y su área Metropolitana. Retrieved April 8, 2023, from <https://manglar.uninorte.edu.co/handle/10584/10844>
- Kumar, P., Jha, A. K., & Balamurugan, B. (2022, April). Dashboard for Crime Analytics Using R & Shiny. In 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (pp. 783-789). IEEE.
- Chainey, S., Tompson, L., & Uhlig, S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime. *Security journal*, 21, 4-28.
- Junguito, R., & Rincón, H. (2004). La política fiscal en el siglo XX en Colombia. *Borradores de economía*, 318.
- Medina, M. (1989). BASES URBANAS DE LA VIOLENCIA EN COLOMBIA: 1945-1950 1984-1988. *Historia crítica*, (1), 20-32.
- Villaveces, A. (1990). A comparative statistical note on homicide rates in Colombia. BERGQUIST et al (eds.) *Violence in Colombia*, 2000.
- Llorente, M. V., Echandía, C., Escobedo, R., & Rubio, M. (2001). Violencia homicida y estructuras criminales en Bogotá. *Análisis Político*, (44), 17-38.
- Bello Montes, C. (2008). La violencia en Colombia: Análisis histórico del homicidio en la segunda mitad del Siglo XX. *Revista criminalidad*, 50(1), 73-84.
- Posada, M. S. (1999). Violencia política, conflicto social y su impacto en la violencia urbana. *Reflexión Política*, 1(1).
- BBC. (n.d.). ¿Cómo se explica el histórico aumento de los homicidios en estados unidos? BBC News Mundo. Retrieved November 21, 2022, from <https://www.bbc.com/mundo/noticias-internacional-58714578>
- Abt, T. (2021, March 16). Opinion | to stop the spike in urban violence, engage those most at risk. *The Washington Post*. Retrieved November 21, 2022, from <https://www.washingtonpost.com/opinions/2021/03/16/young-men-violence-cities/>
- Ledur, J., & Rabinowitz, K. (2022, November 20). There have been more than 600 mass shootings so far in 2022. *The Washington Post*. Retrieved November 21, 2022, from <https://www.washingtonpost.com/nation/2022/06/02/mass-shootings-in-2022/>
- Statistics, D. V. (2013). Domestic violence statistics.
- Domestic Violence Statistics. The Hotline. (2022, October 31). Retrieved November 21, 2022, from <https://www.thehotline.org/stakeholders/domestic-violence-statistics/>
- Centers for Disease Control and Prevention. (n.d.). Retrieved November 22, 2022, from https://www.cdc.gov/ViolencePrevention/pdf/NISVS_Report2010-a.pdf
- Child maltreatment & neglect statistics: American SPCC - effects of parental alcohol abuse. American SPCC. (2022, June 13). Retrieved November 21, 2022, from [https://americanspcc.org/child-maltreatment-statistics/#:~:text=57%2C963%20victims%20\(9.3%25\)%20are,abuse%20and%20neglect%20in%202020.%20\(Accessed:%20November%2014,%202022\).](https://americanspcc.org/child-maltreatment-statistics/#:~:text=57%2C963%20victims%20(9.3%25)%20are,abuse%20and%20neglect%20in%202020.%20(Accessed:%20November%2014,%202022).)
- Widom, C. S., & Hiller-Sturmhöfel, S. (2001). Alcohol abuse as a risk factor for and consequence of child abuse. *Alcohol Research & Health*, 25(1), 52.
- Turkewitz, J., & Villamil, S. (2021, May 5). Colombia police respond to protests with bullets, and death toll mounts. *The New York Times*. Retrieved November 21, 2022, from <https://www.nytimes.com/2021/05/05/world/americas/colombia-covid-protests-duque.html>
- Larizgoitia, I. (2006). La violencia también es un problema de salud pública. *Gaceta Sanitaria*, 20, 63-70.
- Dusenbury, L., Falco, M., Lakem, A., Brannigan, R., & Bosworth, K. (1997). Nine critical elements of promising violence prevention programs. *Journal of School Health*, 67(10), 409-414.
- Farrell, A. D., Meyer, A. L., Kung, E. M., & Sullivan, T. N. (2001). Development and evaluation of school-based violence prevention programs. *Journal of clinical child psychology*, 30(2), 207-220.
- Kondo, M. C., Andreyeva, E., South, E. C., MacDonald, J. M., & Branas, C. C. (2018). Neighborhood interventions to reduce violence. *Annual review of public health*, 39, 253-271.
- Cerdá, M., Tracy, M., & Keyes, K. M. (2018). Reducing urban violence. *Epidemiology*, 29(1), 142-150.
- Wilson, J. M., Chermak, S., & McGarrell, E. F. (2010). *Community-Based Violence Prevention*. Santa Monica, CA: RAND Corporation.
- Balogun, T. F., Okeke, H., & Chukwukere, C. I. (2014). Crime mapping in Nigeria using GIS. *Journal of Geographic Information System*, 6(05), 453.
- Baba, M., Gui, V., Cernazanu, C., & Pescaru, D. (2019). A sensor network approach for violence detection in smart cities using deep learning. *Sensors*, 19(7), 1676.
- Muñoz, R., Barcelos, T., Noël, R., & Kreisel, S. (2012, November). Development of software that supports the improvement of the empathy in children with autism spectrum disorder. In 2012 31st International Conference of the Chilean Computer Science Society (pp. 223-228). IEEE.
- Apte, C., Liu, B., Pednault, E. P., & Smyth, P. (2002). Business applications of data mining. *Communications of the ACM*, 45(8), 49-53.
- Ur, A., Rahman, S., Naseer, I., Majeed, R., Musleh, D., Gollapalli, M. A. S., ... & Khan, M. A. (2021). Supervised machine learning-based prediction of covid-19. *Computers, Materials and Continua*, 21-34.
- Osorio, J., Mohamed, M., Pavon, V., & Brewer-Osorio, S. (2019). Mapping violent presence of armed actors in colombia. *Advances of Cartography and GIScience of the International Cartographic Association*, 16(1), 1-9.
- Mollalo, A., Vahedi, B., & Rivera, K. M. (2020). GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Science of the total environment*, 728, 138884.
- Brown, M. S. (2014). Transforming unstructured data into useful information. In *Big Data, Mining, and Analytics* (pp. 227-246). Auerbach Publications.

- Pruneau, C. A. (2017). *Data analysis techniques for physical scientists*. Cambridge University Press.
- Takatsuka, M., & Gahegan, M. (2002). GeoVISTA Studio: A codeless visual programming environment for geoscientific data analysis and visualization. *Computers & Geosciences*, 28(10), 1131-1144.
- Judd, C. M., McClelland, G. H., & Ryan, C. S. (2017). *Data analysis: A model comparison approach to regression, ANOVA, and beyond*. Routledge.
- Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2, pp. 131-160).
- Karageorgiou, E. (2011). The logic of exploratory and confirmatory data analysis. *Cognitive Critique*, 35-48.
- Kuhlman, D. (2009). *A python book: Beginning python, advanced python, and python exercises* (pp. 1-227). Lutz: Dave Kuhlman.
- McKinney, W., & Team, P. D. (2015). *Pandas-Powerful python data analysis toolkit*. Pandas—Powerful Python Data Analysis Toolkit, 1625.
- Prakash, T. G. (2016). Data visualization and communication by big data. *Int. Res. J. Eng. Technology*, 2395-0056.
- Krug, E. G., Mercy, J. A., Dahlberg, L. L., & Zwi, A. B. (2002). The world report on violence and health. *The lancet*, 360(9339), 1083-1088.
- Martin, E. A. (2009). *A dictionary of law*. OUP Oxford.
- Fraile, P. (2007). La percepción de seguridad: entre el delito, el conflicto y la organización del espacio. *Scripta nova. Revista electrònica de geografia i ciencias sociales*, 2007, vol. XI, núm. 245 (62).
- Salas, S. (n.d.). Defensoría del Pueblo alertó sobre aumento de la violencia en Barranquilla. Alerta por aumento de la violencia en Barranquilla. Retrieved November 24, 2022, from <https://www.radionacional.co/noticias-colombia/alerta-por-aumento-de-la-violencia-en-barranquilla>
- Cheng, T., & Chen, T. (2021). Urban Crime and Security. *Urban Informatics*, 213-228.
- Méndez, A. (2022, January 9). En 8,2 % Se Incrementaron Los homicidios Entre 2019 y 2021. *El Tiempo*. Retrieved November 24, 2022, from <https://www.eltiempo.com/justicia/delitos/aumento-de-los-homicidios-en-colombia-en-2021-643833>
- Briceño-León, R., & Zubillaga, V. (2001). Dimensiones y construcciones de la violencia en América Latina. *Acta científica venezolana*, 52(2).
- Garcia-Zanabria, G., Gomez-Nieto, E., Silveira, J., Poco, J., Nery, M., Adorno, S., & Nonato, L. G. (2020, November). Mirante: A visualization tool for analyzing urban crimes. In *2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)* (pp. 148-155). IEEE.
- Ortega, D.J.L. (2023) Preocupación en barranquilla y atlántico por el incremento de muertes violentas, *El Tiempo*. *El Tiempo*. Available at <https://www.eltiempo.com/colombia/barranquilla/inseguridad-en-barranquilla-muertes-violentas-en-enero-de-2023-738786> (Accessed: April 8, 2023).
- Tan, P. N., Steinbach, M., & Kumar, V. (2013). Data mining cluster analysis: basic concepts and algorithms. *Introduction to data mining*, 487, 533.
- MacQueen, J. (1967, June). Classification and analysis of multivariate observations. In *5th Berkeley Symp. Math. Statist. Probability* (pp. 281-297). Los Angeles LA USA: University of California.