Modelo para el análisis de correlación de variables frente a la percepción de seguridad ciudadana en Colombia

Edwin Alberto Nieto Fagua

María Camila Yepes Guzmán

Directores

Elio H. Cables Pérez, Ph.D.

Edison Leonardo Neira Espitia, Esp.

Especialización en Gobierno de Datos

Facultad de ingeniería, Universidad Antonio Nariño

2021

Resumen

En la actualidad múltiples organizaciones gubernamentales gestionan en sus labores del día a día grandes volúmenes de datos, mismos que según su criticidad se exponen o no, a la comunidad con el fin de que exista mayor participación por parte de los ciudadanos y de generar transparencia en su gestión.

Así mismo para la ciudadanía, los temas de seguridad van tomando cada vez mayor relevancia; ya que la percepción de seguridad en los últimos años ha ido en declive; según lo que presentan múltiples estadísticas de instituciones como la Cámara de Comercio de Bogotá, entre otras; pese a que en la última década se presentaron cambios socio políticos como la firma de un acuerdo de paz con uno de los grupos alzados en armas más grande del país. Sin embargo, se encuentran otros indicadores como los índices de pobreza, que según el DANE, Departamento Administrativo Nacional de Estadística han tenido una fuerte tendencia al alza, a su vez también se viven otros fenómenos socio demográficos que han venido siendo significativos en los últimos años, tales como el cambio en las dinámicas migratorias; puesto que se observa un gran éxodo por parte de los ciudadanos venezolanos al territorio colombiano.

Por lo tanto, se plantea un modelo que permita el análisis de correlación de variables de diversas fuentes relacionadas con la percepción de seguridad ciudadana en Colombia, por medio del procesamiento de conjuntos de datos de fuentes abiertas diversas, el cual involucre un proceso de transformación que permita generar valor a través de la visualización de estas variables de manera dinámica, que a su vez apoye la toma de decisiones del ejecutivo y demás instituciones.

Palabras Claves: Datos abiertos, Toma de decisiones, Inteligencia de Negocio, Percepción de Seguridad.

Tabla De Contenido

Lista De Tablas

Lista De Figuras

Introducción

Actualmente en Colombia la ciudadanía tiene enfocada su atención principalmente en algunos temas que atañen a la sociedad por su gran impacto, uno de ellos es la pandemia del COVID-19 que llegó para cambiar muchos aspectos de la vida cotidiana, así mismo pero con una historia un poco más larga se encuentra la percepción de la seguridad ciudadana, los migrantes en el país, las condiciones socioeconómicas de la población, la eficiencia del pie de fuerza y la eficacia en la impartición de justicia, la cual toma mayor relevancia para la sociedad.

Hoy por hoy la percepción de seguridad, como el sentir de la ciudadanía frente a su conocimiento del nivel de seguridad de su entorno con el paso de los años ha ido en fuerte declive, por lo que, no es extraño encontrar titulares de prensa, tales como:

- "Mala percepción de seguridad: una bicicleta estática" [1].
- "Anuncian nuevas acciones para hacerle frente a la inseguridad en Sucre" [2].
- " ¿Quién controla a la policía?" [3].
- " Percepción de inseguridad en Bogotá está en el nivel más alto en los últimos cinco años según la CCB" [4].
- "El reto de la percepción en seguridad" [5].

Entre otros muchos que se relacionan a esta temática de diferentes editoriales a lo largo y ancho del país. Así mismo han surgido diferentes estudios como el denominado "Global Peace Index" realizado por el Institute for Economics & Peace, donde Colombia ha ido escalando poco a poco en el ranking, puesto que en el 2014 ocupaba el número 150 de 162 países evaluados pasando al puesto número 140 de 163 en 2020 [6][7]. Sin embargo, sigue siendo uno de los países menos pacíficos de la región. Igualmente, la percepción de inseguridad de la ciudadanía según estadísticas del DANE para algunas de las principales ciudades presenta aumentos entre el 2017 y 2019[8].

Por ello, mediante el uso de la analítica descriptiva, la cual comprende el entendimiento de datos históricos a fin de descubrir patrones y la correlación de diversas variables, mediante el procesamiento de grandes cantidades de datos de múltiples orígenes, empleando métodos estadísticos y matemáticos además de tecnología; ello da lugar a agilizar la generación de conocimiento y descubrimiento de información poco perceptible [9]. Por otra parte, el aprovechamiento de las múltiples fuentes de datos abiertas existentes, permite plantear el desarrollo de un estudio que correlacione las variables de migración, pobreza, procedimiento penal y percepción de seguridad, tal que faciliten el entendimiento de estos indicadores en el territorio nacional para los entes públicos y la ciudadanía, coadyuvando a entender el aumento en la percepción de inseguridad en el país y a su vez facilitar a los entes públicos la identificación de puntos de mejora que deriven en la construcción de programas sociales y otras medidas que contribuyan en la mejora los indicadores actuales.

Este trabajo se organiza de la forma siguiente: primero se presenta el marco referencial sobre el cual se soporta este documento, en él se abordan temáticas como analítica descriptiva, análisis de correlación y productos similares. Luego, se procede a abordar la metodología junto con las herramientas a emplear, para posteriormente presentar los resultados respecto a la pregunta de investigación y finalmente las conclusiones del trabajo realizado.

PLANTEAMIENTO DEL PROBLEMA

DESCRIPCIÓN DEL PROBLEMA

En la actualidad existen múltiples fuentes de datos oficiales, que contienen el registro de diversas variables, una de ellas es el DANE, que facilita el estudio de diversas variables en el ámbito nacional. En este caso, algunas de las que se tendrán en cuenta son la pobreza y la percepción de seguridad. Por otra parte, se encuentra otra fuente oficial en lo referente a los delitos que son los datos que otorga la policía nacional, así mismo se encuentran datos que entidades como la Fiscalía General de la nación y Migración Colombia han cargado a la plataforma de Datos Abiertos. A pesar de ello, en la actualidad en Colombia cada una de estas variables son evaluadas por separado; puesto que cada entidad se encarga de las cifras que se relacionan directamente con sus funciones, por ello no se ha identificado la correlación entre ellas. Sin esto, las diversas instituciones de orden gubernamental no tienen una visión global del problema, lo cual les permita una mejor toma de decisiones respecto a los programas sociales que se deben implementar, las necesidades de la comunidad, la eficiencia en la impartición de justicia, entre otras medidas que contribuyan a cambiar la percepción de seguridad por parte de la ciudadanía; puesto que en gran parte de los casos esta percepción se ve sesgada a causa de la información suministrada por los diversos medios de comunicación.

FORMULACIÓN DEL PROBLEMA

Lo anteriormente analizado, permite identificar la problemática siguiente:

Cómo unificar fuentes de datos abiertas con diferentes estándares que son ofrecidos por instituciones tales como el DANE, la Policía, Migración Colombia y Fiscalía, tal que permita el análisis de la incidencia de las variables de pobreza, migración y procesamiento penal, en la percepción de la seguridad ciudadana de los departamentos colombianos.

OBJETIVOS

Objetivo General

Generar un modelo que permita unir datos con diferentes estándares de las fuentes de datos abiertas del DANE, Policía, Migración Colombia y Fiscalía para el análisis de la incidencia de las variables pobreza, migración, y procesamiento penal en la percepción de la seguridad ciudadana de los departamentos colombianos.

Objetivos Específicos

1. Identificar los métodos, procedimientos y/o modelos empleados para el análisis de datos a partir de la exploración de diversas fuentes de datos.
2. Diseñar la arquitectura de datos necesaria para la extracción, limpieza y almacenamiento de los conjuntos de datos.
3. Aplicar las técnicas sobre la extracción, limpieza y almacenamiento de los conjuntos de datos relacionados con las variables pobreza, migración, y procesamiento penal, de las fuentes (DANE, Policía, Migración Colombia y Fiscalía) entre 2017 y 2019.
4. Aplicar procesos de analítica descriptiva para la identificación de patrones y/o correlaciones de las variables pobreza, migración y procesamiento penal en la percepción de seguridad ciudadana.

MARCO REFERENCIAL

MARCO TEÓRICO

Seguridad Ciudadana

La ***Seguridad Ciudadana*** se centra en la protección de los individuos frente a actos delictivos en el ámbito público; favoreciendo sus derechos y libertades, en miras a una convivencia pacífica. En el marco de esta se busca la reducción del delito, mejora de la seguridad en general, además, de prevenir la violencia; por ello como apoyo a esta surgen proyectos, actividades e ideas enfocados en la generación de empleo, formación de la población, procesos de mediación, acceso a la justicia, entre otros que contribuyan a la cohesión social y el cumplimiento de los derechos de la población [10][11].

Cuando se habla de seguridad ciudadana se pueden ver varias aristas una de ellas es la "***Percepción de la Seguridad Ciudadana***".

Percepción Seguridad Ciudadana

Esta corresponde al sentir de la ciudadanía frente a su idea o conocimiento en cuanto al nivel de seguridad que se da en el contexto social de su comunidad [12]. Esta se ve influenciada por sus ideas frente a las tendencias de delincuencia y violencia, sus preferencias respecto a la seguridad, el cumplimiento de normatividad por parte de los individuos de su comunidad, el accionar de los gobernantes frente a las medidas que se toman para velar por la seguridad y coadyuvar en la correcta impartición de justicia y velar por los derechos de la población, así mismo, por la información divulgada por los medios de comunicación, igualmente influye la respuesta del pie de fuerza; que se define como el conjunto de individuos que tienen como fin de velar por los derechos y libertades de los ciudadanos, coadyuvando al cumplimiento de la normatividad [13]. En el caso de Colombia está conformado por los efectivos de las siguientes instituciones, el ejército, la armada, la fuerza aérea y la policía.

Desde el pie de fuerza, instituciones como la Policía Nacional contribuyen junto con la rama judicial y la fiscalía en el Procedimiento Penal, que es la respuesta punitiva frente a los hechos delictivos de los individuos que da como resultado la absolución o la imposición de la restricción de derechos, a cargo del sistema judicial, siguiendo el correspondiente proceso penal que para el sistema colombiano comprende las etapas de indagación, investigación y juicio [14].

Procedimiento Penal

El procedimiento penal Colombiano, según los parámetros establecidos en la ley 906 de 2004 [15], establece los lineamientos sobre los cuales se surten las diferentes etapas que se llevan a cabo dentro de un proceso penal desde su inicio, con la presunta comisión de una conducta punible, en donde se entra a determinar la realización de un hecho ilícito que se encuentre tipificado en el código penal colombiano; posteriormente los hechos materia de investigación son puestos a disposición de la fiscalía general de la nación una vez la víctima establece la correspondiente denuncia, la cual será investigada por este ente con ayuda de la policía judicial. Luego, se procede a la captura del presunto delincuente donde se da paso a los actos urgentes, legalización de la captura, formulación e imputación, solicitud de la medida de aseguramiento, dicha etapa se surte ante el juez de control de garantías quien se encargará de imprimir la legalidad en aras de garantizar los derechos del presunto delincuente o infractor.

A continuación, se llega a la formulación de acusación que se surte ante un juez de conocimiento donde se procede a dar a conocer los hechos que dieron lugar a la investigación de manera adicional en este escenario las partes enuncian los elementos probatorios y/o evidencia física que poseen y pretendan hacer valer dentro del juicio oral, dando continuidad al proceso se instaura la audiencia preparatoria. En dicha audiencia los actores enuncian los elementos probatorios y/o evidencia física pertinente (Estipulaciones probatorias), con ello se da lugar a la fecha y hora de la audiencia de juicio oral. Finalmente, en esta la fiscalía y defensa presentan la teoría del caso con sus evidencias y correspondientes alegatos de conclusión, dando, así como resultado el fallo condenatorio o absolutorio por parte del juez [16].

Así mismo se bordará la migración ya que es una de las variables a analizar.

Migración

Por otro lado, se pretenden analizar la ***Migración***, entendiéndose como el movimiento espacial de la residencia habitual de forma permanente de un individuo, en el que se involucra el paso de un límite geográfico, este puede darse de forma voluntaria o no, y se impulsa principalmente por las condiciones de inseguridad, privaciones económicas, discriminación, persecución, enfermedades entre otras y tienen como fin la búsqueda de mejores condiciones educativas, de seguridad, etc. [17] [18]. Igualmente eludir la ***Pobreza*** como condición socioeconómica donde se carece de recursos para suplir las necesidades básicas (Vivienda, alimentación, educación, salud, etc); esta puede darse por nivel bajo de ingresos, segregación social, desempleo, desplazamiento forzado, etc. [19].

Por otro lado, es necesario referirnos a los Datos abiertos, ya que serán empleados como fuente de información.

Datos abiertos

Los ***Datos abiertos*** han surgido como un movimiento, en el cual se publican datos gubernamentales en formatos reutilizables, a los que se les atribuyen características como disponibilidad, no restrictivos frente a su uso y distribución, a fin de aumentar la transparencia, eficiencia y participación algunos ejemplos de ellos son datos geográficos, meteorológicos, entre otros [20] [21].

Igualmente es necesario abordar el concepto de análisis de correlación de variables.

Análisis de correlación

Así mismo, en el marco de este proyecto surge al ***análisis de correlación*** de estas variables como técnica estadística cuyo fin es comprobar la relación entre dos o más variables, demostrando si los cambios en una variable repercuten en la otra variable de forma positiva o negativa; o simplemente no se relaciona. Para llevar a cabo el análisis de correlación se surten diferentes etapas, entre ellas el proceso de identificación en el cual se emplean técnicas que soportan procesos como la extracción de conocimiento a partir de los datos, que han de atravesar una etapa de preparación donde se realizan las tareas de selección, limpieza y transformación los datos, para su posterior exploración y auditoría que a su vez da paso al desarrollo de modelos y análisis de datos, y finalmente la evaluación, difusión y utilización de los modelos [22].

Para este caso se apoya a través del análisis de datos.

Analítica de datos

Por lo anterior se hace necesario definir la ***Analítica de datos,*** es decir la examinación de los datos a fin de llegar a conclusiones a partir de ellos; la cual conlleve a la toma de decisiones; así mismo la analítica puede clasificarse en descriptiva, predictiva y prescriptiva. La analítica descriptiva se encarga de identificar patrones de comportamiento en variables desde una perspectiva histórica, en este escenario generalmente se realizan regresiones, análisis de correlación. Por otro lado, la analítica predictiva la cual busca el entender las causas del comportamiento histórico, y finalmente, la analítica prescriptiva que busca predecir el comportamiento futuro [23].

Igualmente, la ***analítica descriptiva*** realiza un estudio de los datos de un periodo pasado a fin de responder la pregunta ¿Qué pasó? y con ello facilitar el entendimiento de estos; lo cual facilita dar un enfoque que permita influir en los resultados a futuro; esta se enfoca en proporcionar conocimientos obtenidos a través de la historia como el comportamiento de la producción, consumo, ventas, inventarios entre otros, que se alinean con la inteligencia de negocio [24].

Dentro de la analítica de datos podemos segmentar un poco llegando a la inteligencia de negocio.

Inteligencia de negocio

Así mismo el término ***inteligencia de negocio (business intelligence, en inglés)*** comprende las metodologías, estrategias, prácticas, aplicaciones y tecnologías usadas por las compañías para la creación y gestión de la información que cimienta y facilita la toma de decisiones mucho más eficientes y efectivas para el negocio; todo esto a partir de datos de calidad los cuales cumplan con criterios como precisión, oportunidad, exactitud; etc. [25].

Igualmente, basados en la definición de (Gartner, Inc., s.f.) hablamos de un “término sombrilla que incluye las aplicaciones, infraestructura, herramientas y mejores prácticas que permitan el acceso y análisis de la información para mejorar y optimizar el desempeño de las organizaciones” esto enmarcado dentro de un proceso interactivo que comprende exploración y análisis sobre un tema, con el fin de evidenciar patrones y tendencias de los cuales se genere un nuevo entendimiento que sirva para el cumplimiento de objetivos y estrategias del negocio. [26] .

Anteriormente se habló de la obtención de conocimiento a partir de los datos, algunos ejemplos claros de los temas que se pueden comprender mediante la inteligencia de negocio son patrones de compra, comportamientos de los clientes, entre otros. Lo cual se deriva en ventajas competitivas, ya que, al tener patrones de compra y comportamiento de los clientes, se pueden crear estrategias publicitarias más efectivas. Por otro lado, al conocer los patrones de compra se puede dar una mejor administración a los inventarios, en fin, el conocimiento generado permite tomar mejores decisiones entorno a estrategias y a su vez se puede dar una mejor gestión de los recursos [27].

Mismo que en sus procesos de ingestión emplean diversas fuentes tales como archivos, bases de datos transaccionales OLTP, entre otros.

OLTP (Online Transaccional Processing)

Toda está inteligencia de negocio es soportada por una arquitectura tecnológica, donde se involucran unas fuentes de datos que pueden ser archivos planos, fuentes externas o bases de datos ***OLTP (Online Transaccional Processing)***. Estas últimas han de entenderse como sistemas gestores de bases de datos orientados a suplir la transaccionalidad de las organizaciones, es decir el sistema en el que se realizan consultas, actualizaciones e inserciones en tiempo real y de forma detallada en el caso de bases de datos relacionales adopta estructuras como modelo entidad relación para el manejo de los datos actuales del sistema [28].

Estos se usan herramientas para carga en la bodega tales como ETL.

ETL (Extract, Transform, and Load)

Posterior a ello, se encuentran componentes que atienden los procesos transformación y carga de los datos denominados ***ETL (Extract, Transform, and Load)*** donde se toman las diferentes fuentes de datos, se aplican reglas de negocio y realizan transformaciones y limpieza necesarias para cumplir con métricas de calidad y continuar con la carga en el componente siguiente el cual consiste en un área de almacenamiento; que bien puede ser un lago de datos, un data mart o una bodega de datos [29].

Como parte de los procesos de ETL se pasan los datos a las bodegas de datos.

Bodegas de datos

Las ***bodegas de datos*** son soluciones donde se consolidan datos de diversas fuentes, esta debe estructurarse de forma tal que contribuya en la estrategia de la compañía y tienen como objetivo soportar la toma de decisiones y centralizar los datos que atienden una o varias áreas de negocio; en ella pueden integrarse uno o más data marts [30] [31].

Así mismo las bodegas de datos responden a diseños lógico y físicos dados por modelos dimensionales donde se localizan componentes principales como ***tablas de hechos*** y dimensiones; las primeras representan lo que va a ser analizado por medio de medidas y a través de llaves foráneas se relacionan con las ***dimensiones***, mismas que tienen como finalidad reducir la redundancia y contribuyen a describir el valor de las medidas en los hechos [32]. Lo anterior bajo modelos físicos y lógicos que se categorizan en ***Start Flake*** que consiste en una tabla de hechos relacionada con unas dimensiones a un solo nivel, es decir, que estas dimensiones solamente tendrán relación con la tabla de hechos [33] y ***Snow Flake*** el cual tiene como punto de partida el modelo anterior. Sin embargo, en busca de normalizar la data de las dimensiones se generan nuevas dimensiones a un nivel mayor, estas cambian las relaciones en las dimensiones, ya que se han relacionar con la tabla de hechos y la nueva dimensión, lo cual da paso a la existencia de llaves foráneas en las dimensiones [34].

Estas son soportadas por herramientas OLAP.

OLAP (Online Analytical Processing)

Subsiguiente al área de almacenamiento se halla la zona de agregación que se realiza por medio de cubos que responden soluciones ***OLAP (Online Analytical Processing),*** herramientas de análisis rápido de grandes volúmenes datos para respaldar la toma de decisiones mediante, la presentación de una vista multidimensional de los mismos; este permite operaciones de agrupamiento, desglose, entre otras, que buscan facilitar su análisis.[35] Finalmente, se halla el área de visualización donde se pueden encontrar reportes, cubos o dashboards, los cuales representan la data procesada.

Transversal a los componentes anteriormente mencionados, están presentes algunos módulos como la zona de ***metadatos*** que corresponden a la información entorno a la bodega de datos tales como su estructura los cuales buscan facilitar la navegación de los usuarios, identificación de orígenes, procesos de transformación entre otros que se dividen en comerciales que van orientados al usuario final y los técnicos que explican la ejecución y administración del modelo [36].

Calidad de datos

De igual manera como se mencionó con anterioridad en todo el proceso, ha de involucrarse la ***calidad de datos,*** entendida como la capacidad de usar los datos por parte de un consumidor, para ello se han establecido algunos criterios como son disponibilidad, usabilidad, fiabilidad, relevancia y calidad de presentación [37];cada uno de ellos con elementos a evaluar, por ejemplo, la ***disponibilidad,*** entendida como la facilidad de acceso que tienen los usuarios autorizados para acceder a los datos, en esta se valoran elementos como la accesibilidad, autorización y oportunidad. Otro criterio es la ***usabilidad*** que está orientada a evaluar si los datos son útiles y satisfacen a los usuarios; de igual manera la ***fiabilidad*** evalúa si se puede confiar o no en ellos allí surgen medidas como la precisión, integridad, consistencia y auditabilidad. Por otra parte, la ***relevancia*** donde se valida la aplicabilidad y usabilidad de los datos. Otro de los criterios es la ***calidad de presentación*** el cual comprende la legibilidad de los datos, es decir, si son fáciles de entender y poseen claridad [38][39].

Así mismo el estándar ***ISO25012,*** dicta métricas sobre los criterios a evaluar en la calidad de datos, los cuales son similares y los divide en características que a su vez se subdividen en tres grupos por un lado se encuentran las ***inherentes*** tales como exactitud, completitud, consistencia, credibilidad y actualidad. Por otra parte, se encuentran los ***dependientes del sistema*** como disponibilidad, portabilidad y recuperabilidad, así mismo en el intermedio se hallan características por ejemplo la accesibilidad, conformidad, confidencialidad, eficiencia, precisión, trazabilidad y comprensibilidad [40] [41].

De igual manera, es necesario resaltar algunos de los problemas de calidad de datos frecuentes, los cuales pueden ser clasificados en ***valores atípicos,*** que comprenden valores con una gran variación respecto a los demás del mismo grupo de datos. Además, se hallan catalogados como ***valores faltantes*** mismos que son necesarios de identificar para dar un tratamiento especial, ya que de no realizarse pueden distorsionar los resultados del análisis. Igualmente se pueden hallar otros problemas cuando se tienen diferentes fuentes de datos tales como conflictos entre estructuras, diferentes niveles de agregación entre otros [42].

Python

Ha de señalarse que las técnicas mencionadas se soportan a nivel tecnológico en diversas herramientas, entre ellas ***Python*** un lenguaje de programación multiparadigma, puesto que soporta la programación funcional, orientada a objetos, entre otros; así mismo es estructurado e interpretado, versátil; ya que se puede correr en muchas plataformas, en él se obvian reglas sintácticas y mediante el uso de sus librerías hace mucho más compacto y limpio el código que otros lenguajes, lo cual conlleva que sea más amigable la escritura y lectura de su código [43].

Jupyter Notebooks

Así mismo, como herramienta para la construcción de código Python se encuentran los ***Jupyter Notebooks***, que desde un navegador estándar permiten describir el proceso llevado a cabo en el análisis de datos, e incluye elementos como el código ejecutable en uno o varios lenguajes de programación, textos, gráficos etc. [44].

PostgreSQL

Otra de las herramientas tecnológicas que contribuye es un Sistema Gestor de Base de datos para este caso ***PostgreSQL,*** el cual tienen un enfoque Relacional, es de código abierto, además de emplear un modelo cliente/servidor y multiprocesos, cumple al 100% con ACID (Atomicidad, Consistencia, Aislamiento, Persistencia), es decir no deja transacciones sin completar, maneja constraints, una transacción no interfiere con otra y no se pierde la información, y soporta lenguajes como PHP, Java y C++ [45].

Con lo anteriormente expuesto se permite por un lado dar contexto de las variables que se plantean analizar en el proyecto, a su vez, se identifica el marco conceptual de las herramientas que se emplearan, así mismo se puede identificar las características de las herramientas de software y lenguajes a usar; por un lado se elegio Python debido a que el lenguaje facilita los procesos de analítica ya que posee diferentes librerías que tienen funcionalidades preestablecidas; igualmente como IDE se emplearon los Jupyter NoteBooks que permiten describir el código, además de soportar el lenguaje seleccionado; en el marco de la persistencia de datos se seleccionó PostgreSQL debido a su robustez, soporte; además que la data a procesar es una data estructurada; es decir que posee un unos campos inmutables en cuanto a estructura.

ESTADO DEL ARTE

Teniendo en cuenta que el análisis de correlación es una herramienta la cual permite determinar cómo la diversificación de una variable impacta una o más de las misma, lo cual conlleva a que constantemente sea adoptada por diversas instituciones en múltiples ámbitos, tales como gobierno, comercio, salud, seguridad, etc. Para este último ámbito el cual se enlaza con las variables que se plantea relacionar en este documento. A continuación, se describen algunas de ellas.

Inicialmente se hallan trabajos de grado como el denominado “Un índice dinámico para la seguridad ciudadana en Colombia: Un acercamiento bayesiano”, el cual realiza un análisis desde una perspectiva netamente estadística de la correlación entre delitos como el homicidio, las lesiones, los delitos sexuales, las extorsiones, el secuestro y el hurto en sus categorías de personas, residencias, vehículos, comercio, desde una óptica nacional y otra enfocada a cinco de las principales ciudades del país tales como: Bogotá, Medellín, Cali, Barranquilla y Bucaramanga.; el cual busca servir como herramienta de diagnóstico que permita la priorización de acciones en materia de seguridad [46].

Por otra parte se encuentran los múltiples tableros generados por la organización Esri (Environmental Systems Research Institute) Colombia, los cuales abordan mapas y gráficos producto del análisis de delitos como homicidio, hurto, así como también de las denuncias ciudadanas, a partir de datos generados por entidades tales como alcaldías y Policía nacional, y a partir de esto genera la solución autodenominada observatorio y análisis del delito, que en la revisión realizada se identifica como herramienta publicada de forma abierta, cuya visualización se limita a la ciudad de Barranquilla y tiene su fundamento en facilitar a la ciudadanía y a las autoridades locales una visión general de la seguridad en materia de delitos y convivencia ciudadana, mediante esta solución tecnológica la cual contribuye a la posibilidad de realizar analítica descriptiva para facilitar la toma de decisiones encaminada a lograr la disminución y el control de las distintas actividades delictivas [47].

Por otra parte, la cámara de comercio de Bogotá, desde su programa de Open Data, de la misma forma que Esri Colombia, publica abiertamente algunos gráficos en los cuales se presentan datos subjetivos y objetivos con base en los distintos comportamientos de los delitos, lo cual conlleva a identificar los niveles de seguridad en Bogotá desde una óptica descriptiva, Con el objetivo de facilitar la generación de iniciativas y propuestas para hacer entornos más seguros y competitivos; lo cual favorezcan la actividad mercantil, mediante el desarrollo de iniciativas público-privadas. De esta herramienta se destaca el aprovechamiento de las fuentes abiertas de datos para la generación de reportes y el enfoque en los temas de seguridad relacionados con los comercios [48].

Así mismo, la Universidad del Norte, con apoyo de la cámara de comercio de Barranquilla y otras organizaciones, construye por su parte, una solución local para el departamento del Atlántico, denominada “Observatorio de seguridad ciudadana”, cuyo fin es procesar datos relacionados con los índices de violencia y criminalidad en la ciudad de Barranquilla y su área metropolitana, cuya fuente es la base de datos del Sistema de Información Estadístico, Delincuencial Contravencional y Operativo de la Policía Nacional – SIEDCO con el acompañamiento del Observatorio del Delito de la DIJIN (Dirección de Investigación Criminal e INTERPOL de la Policía Nacional). Esta solución pretende facilitar la toma de decisiones mediante la evaluación de las políticas públicas encaminadas a la mitigación de las problemáticas relacionados con la seguridad ciudadana, además de la georreferenciación de zonas estratégicas de incidencia [49].

También se halla otra solución tecnológica llamada IBM Intelligent Operations Center for Emergency Management (IBM), herramienta tecnológica que permite la integración de múltiples fuentes de datos dispares, near real-time; que permite la gestión de emergencias y seguridad, la cual cuenta con análisis de datos, relación con redes sociales, mapas geodinámicos, tableros administrativos, alertamiento, etc. con el fin de disminuir los tiempos de respuesta y facilitar la toma de decisiones mediante la identificación de condiciones, impacto de los diferentes eventos etc. [50].

Además, se identifica la tesis denominada “Modelo para la Caracterización del Delito en la Ciudad de Bogotá, Aplicando Técnicas de Minería de Datos Espaciales” misma en la que se tomaron como fuentes de datos Estadísticos de los estudios de percepción ciudadana. Resultados publicados por parte de la DIJIN, Boletines de seguridad de la Secretaría Distrital de Planeación ,Observatorio Nacional de Seguridad de Bogotá de la Cámara de Comercio, Informes de Seguridad de la Veeduría Distrital, Revista Criminalidad de la Policía

Metropolitana de Bogotá; las cuales fueron procesadas aplicando un modelo descriptivo y empleando algoritmos como K-means y DBscan se pudieron resolver los siguientes interrogantes ¿Cuál es la ubicación del delito? , ¿Cuál es la modalidad más frecuente?, ¿Qué armas son las más utilizadas?, ¿Cuál día de la semana en la que se presentan más hechos delictivos?, ¿Cuál es el rango horario en el que se presentan más hechos delictivos?, ¿Cuál es el mes de mayor ocurrencia de delitos?; esto en términos estadísticos, que se ven reflejados en los gráficos del mismo documento [51].

De igual forma se localiza el trabajo de grado titulado "Caracterización de los delitos en Cartagena mediante la aplicación de minería de datos", que fue desarrollado teniendo como orígenes de datos el portal web del gobierno: Data.gov.co en los ámbitos de hurto a vehículos, entidades financieras, comercio, residencias que posterior al uso de técnicas de clusterización permite la creación de un portal virtual, donde se pueden apreciar los resultados de las consultas sobre las variables relacionadas con hurtos (tipo de arma, género de la víctima, día, zona, hora, etc) en la ciudad. Desarrollado como herramienta de consulta para los habitantes [52].

Igualmente, en un ámbito distinto se encuentra la tesis doctoral denominada "Metamodelo para integración de datos abiertos aplicado a inteligencia de negocios", la cual generó un modelo para el tratamiento de los datos abiertos a fin de servir de apoyo en la toma de decisiones; así mismo concibió herramientas como dashboards para la visualización de datos, en esta se tuvo como objetos de estudio, portales de Estados Unidos, Europa, Gran Bretaña y a nivel de Latinoamérica se tuvo en cuenta México, Uruguay, Brasil y Colombia. Este último fue tomado como referente con datos del ámbito agropecuario [53].

De lo anterior se logro establecer, que en la actualidad existen múltiples herramientas las cuales permiten el análisis de variables relacionadas con la seguridad, tanto a nivel nacional como internacional; así mismo se establece que existen modelos para la explotación de datos de fuentes abiertas ofrecidos en los marcos de transparencia y gobierno en línea, al igual que se identifican herramientas a modo de tableros que han sido generados en contextos locales, generalmente en el marco de observatorios de seguridad; sin embargo se logra identificar que no se realizan correlaciones entre variables; lo cual fundamenta este proyecto.

IMPACTO

El proyecto a través de la creación de un modelo para el análisis de correlación de las variables de pobreza, procesamiento judicial, migración en Colombia; frente a la percepción de seguridad ciudadana, enfoca su interés en otorgar una herramienta útil para diversas instituciones donde se identifique la incidencia de las variables de pobreza, migración, y procesamiento judicial en la percepción de seguridad ciudadana de los departamentos colombianos y de esta manera contribuya en la toma de decisiones frente a la planeación de los programas sociales, migratorios, educativos entre otros ofrecidos a la ciudadanía, con fines de mitigación y reducción del índice de percepción de inseguridad en el país.

COMPONENTE DE INNOVACIÓN

Con el establecimiento del estado del arte, se identifica la existencia de varias herramientas entorno a la seguridad ciudadana en contextos locales; es decir de ámbito municipal y distrital, sin embargo, ninguno de estos nos permite correlacionar las variables pobreza, migración y procesamiento judicial.

METODOLOGÍA

Para el desarrollo de este proyecto, se adaptó una versión de la metodología SEMMA (Sample, explore, modify, model, assess)., la misma fue desarrollada por el instituto SAS(Statistical Analysis Systems), y se enfoca en el uso de las herramientas de esta organización, dicha metodología contempla las etapas de muestreo, exploración, modificación, modelado y evaluación. En esta estrategia de trabajo, se establece el proceso de muestreo, el cual permite identificar características comunes de los datos; seguidamente se indica el proceso de selección, que comprende la aplicación de técnicas estadísticas para detectar datos anómalos y con ello se concluye la fase de exploración. Posteriormente, se procede con la etapa denominada modificación, en la cual se llevan a cabo la selección y transformaciones de los datos para continuar con el modelado, donde se aplican las técnicas de minería de datos, con el fin de descubrir las relaciones entre los datos. Finalmente, la etapa de evaluación donde se procede a la verificación del modelo, todo esto con el principal objetivo del descubrimiento de patrones [54][55].

Por lo anterior y debido a la compatibilidad de este proyecto, se propone un esquema de trabajo en el cual se desarrollen las siguientes fases que se describen a continuación. (Figura 1.)

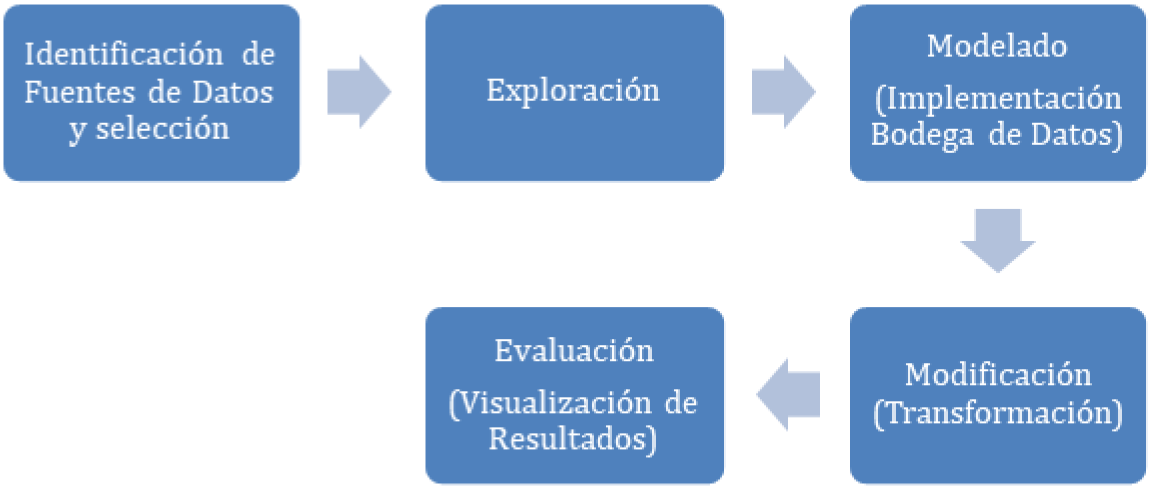

Figura 1.Etapas de la Metodología. Fuente Propia

Descripción de las Fases de trabajo

Identificación de las fuentes de datos y selección

Comprender el proceso de búsqueda de las variables que se plantea estudiar, partiendo de fuentes de datos abiertas.

Exploración

En esta fase se procede al cargue de los diferentes conjuntos de datos identificados en una base de datos PostgreSQL mediante herramienta de extracción y carga (EL) realizada en Python, a la información cargada se aplicarán diferentes procesos a fin de realizar la identificación de falencias en calidad tales como datos anómalos que se deban tratar; así mismo identificar el comportamiento de las fuentes de forma individual. Lo anterior haciendo uso de notebooks en Jupyter y power BI.

Modelado

En esta fase llevarán a cabo tareas como la construcción de una estrategia de datos, posterior a ello la definición arquitectónica de la solución orientada a una bodega de datos, e igualmente el diseño lógico de la capa de persistencia de los datos, que posteriormente será consumida por la herramienta de visualización.

Modificación

Esta etapa procede a consumir los conjuntos de datos cargados en bases de datos PostgreSQL, a fin de prepararlos para posteriormente cargarlos un modelo dimensional por medio de procesos de limpieza, que se han de realizar a través de procedimientos almacenados dentro de PostgreSQL [54].

Evaluación

En esta fase se han de normalizar los indicadores empleando técnicas como los porcentajes de variación, la categorización de valores por encima o debajo del promedio, ponderar cada indicador y cada perspectiva, definición de escalas de color y numéricas a manejar, tipos de visualizaciones, el cual facilite la comprensión e identificación de patrones [55].

DESARROLLO PROPUESTA

Identificación de las fuentes de datos y selección

Partiendo de las variables planteadas, se procede a buscar en las fuentes de datos abiertos como el DANE, Policía nacional, Datos Abiertos Colombia, fiscalía general de la nación, migración Colombia; sitios donde se encuentran datos oficiales de las diversas variables que se contemplan en este proyecto, estos fueron encontrados en formatos csv y manejan historia de aproximadamente 5 años; algunos de ellos están divididos por años y para el caso de las fuentes de Policía Nacional se encuentran divisiones por tipo de delito.

En las fuentes de datos del DANE, se halla información sobre las estadísticas nacionales en términos de pobreza, pobreza extrema y percepción de seguridad, desde la óptica de la confianza que tiene la ciudadanía en las instituciones encargadas de la seguridad, en este caso Policía Nacional y Ejercito todo lo anterior disgregado a nivel de departamentos; con base en ella se gestionaran las variables de percepción de seguridad y pobreza que se plantean analizar.

Por otro lado, se encuentran los conjuntos de datos de la Policía Nacional, en los cuales se listan algunas características de los hechos que han sido de su conocimiento, tales como la fecha de los hechos, lugar del hecho, armas y/o medios empleados entre otros, Estos conjuntos de datos a su vez se encuentran divididos por anualidades y tipo de delito como abigeato, hurto a personas, hurto a comercios, hurto a automotores, lesiones personales, extorsión, secuestro, entre otros, este conjunto se identifican los comportamientos de los delitos.

Asimismo, en la fuente de datos abiertos Colombia se hallaron conjuntos con información perteneciente a migración Colombia donde se encuentran los ingresos de extranjeros al país en términos de años, meses, nacionalidades y punto de ingreso al país, igualmente se hallaron datos generados por la Fiscalía general de nación en los relativo a los delitos del sistema penal oral acusatorio, donde se aprecian las etapas de los procesos, fecha de los hechos, lugar de los hechos a una granularidad de municipio o centro poblado, seccional encargadas, si genero captura, imputaciones o no, entre otras características relevantes en términos jurídicos; mismo con los que se evaluará el avance en la impartición de la justicia.

Exploración

En esta fase se realizó el cargue de los diferentes conjuntos de datos identificados en una base de datos PostgreSQL mediante herramienta de extracción y carga (EL) realizada en Python, a la información una vez cargada en la base de datos se procedió a la generación de notebooks que consumían los datos desde allí, en estos de revisaron los tipos de datos, identificación de valores; además se realizaron diferentes gráficas a fin de comprender los datos, lo cual permitió identificar que todas las fuentes de información no contienen información para los últimos cinco años, por lo tanto, se generó una restricción donde se toman los últimos 3 años. Así mismo se realizaron notebooks en Jupyter los cuales permitieron hacer otra exploración de los datos, allí se aplicaron cálculos estadísticos a las fuentes seleccionadas en la fase anterior, que consiste en 5 conjuntos de datos obtenidos de las fuentes abiertas, con el objetivo de identificar datos anómalos y darles su respectivo tratamiento [56].

En este proceso se hace uso de algunas librerías de Python, tales como, Pandas que es un desarrollo Open source iniciado por Applied Quantitative Research (AQR) Capital Management en 2008, y actualmente soportado por NumFocus, cuyo objetivo es hacer de forma rápida y eficiente la manipulación de datos en memoria y en diferentes formatos, lo cual permite el ajuste, transformación, unión y reordenamiento de los conjuntos de datos [57]; Así mismo ConfigParser que permite la implementación de una configuración básica, la cual facilite la personalización del código [58]; por otro lado se encuentra psycopg2, que se comporta como adaptador de base de datos y tiene como funcionalidad la conexión entre Python y el motor de base de datos de PostgreSQL, permitiendo así la inserción de los conjuntos de datos de la etapa anterior [59]. Finalmente, SQLAlchemy, herramienta con licencia MIT, cuya finalidad consiste en brindar una herramienta para el manejo de la sintaxis del motor de base de datos seleccionado; en este caso el de PostgreSQL [60].

De igual manera se empleó la Herramienta Power BI en su versión Desktop, herramienta de Microsoft, la cual permite la visualización de datos, facilitando la creación de informes dinámicos, con visual interactiva [61].

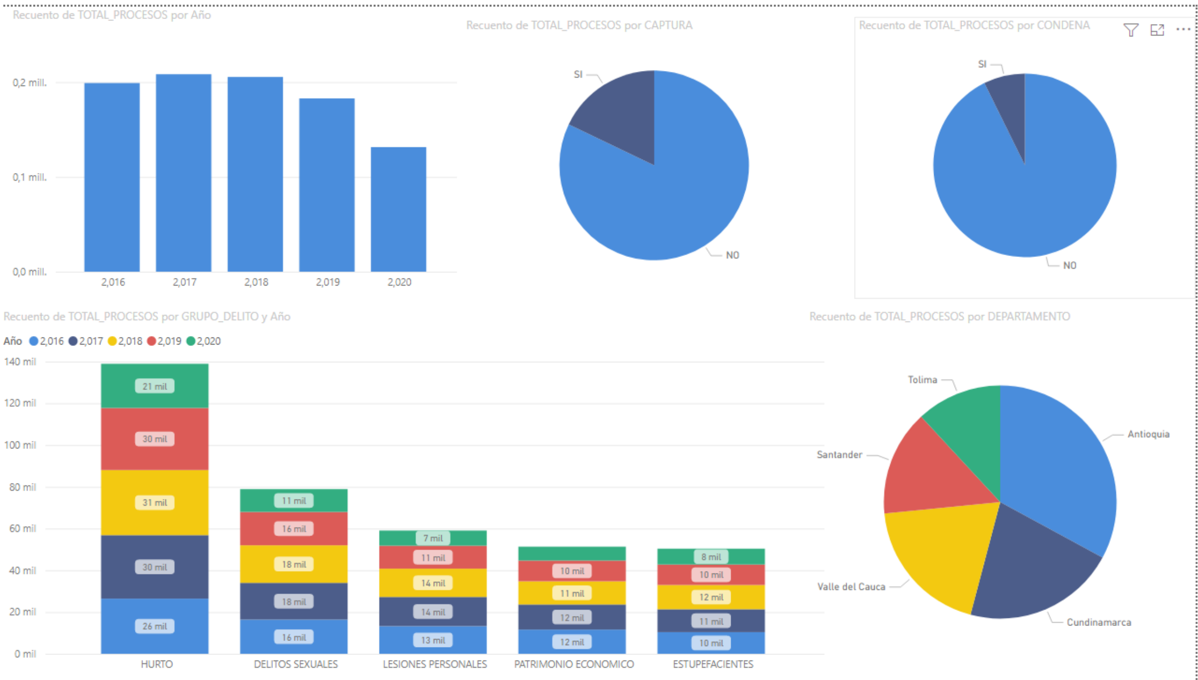

Figura 2. Navegación Datos Procesamiento Judicial (Fiscalía General de la Nación). Fuente Elaboración Propia

En la Figura 2 se identificó un comportamiento con una variación significativa para el año 2020, mismo en el que a nivel país se llevaron a cabo diversos eventos desde el mes de marzo, como cuarentenas obligatorias, toques de queda; cuyo fin fue la disminución de la movilidad de la población como medida para evitar altas tasa de contagio de la pandemia Covid-19 en Colombia. Así mismo se identificaron múltiples grupos de delitos (ver Tabla 1) de los cuales no todos afectan directamente la percepción de seguridad ciudadana, por lo cual algunos fueron excluidos. Igualmente se identificó que existían diferentes atributos para la gestión de la fecha (ANIO_DENUNCIA, ANIO_ENTRADA y ANIO_HECHO) por lo cual se decidió realizar las asociaciones por año de los hechos. Además, se identificaron los delitos que mayor participación tienen, al igual que los departamentos en los que más delitos se reportan.

Tabla 1. Acumulado de Procesos Judiciales por Grupo Delito.

Grupo delito	Total acumulado	Grupo delito	Total acumulado
Hurto	1903480.0	Lesiones personales	601775.0
Violencia intrafamiliar	524047.0	Estupefacientes	372111.0
Patrimonio económico	249382.0	Lesiones personales culposas	248274.0
Amenazas	234244.0	Contra la familia	223677.0
Estafa	221824.0	Delitos sexuales	205335.0
Injuria y calumnia	196377.0	Homicidio doloso	159593.0
Falsedad en documento	157009.0	Impartición de justicia	146785.0
Fe pública	137933.0	Delitos informáticos	116294.0
Fabricación, tráfico y porte de armas	113195.0	Administración pública	63243.0
Corrupción judicial	57656.0	Constreñimiento	54907.0
Extorsión	48720.0	Homicidio culposo	42011.0
Corrupción administrativa	33557.0	Corrupción tributaria	29203.0
Orden económico social	26829.0	Concierto para delinquir	25158.0
Desplazamiento	22940.0	Delitos ambientales	21435.0
Libertad individual y otras garantías	18184.0	Seguridad pública	17663.0
Desaparición forzada	11066.0	Corrupción privada	10517.0
Violación medidas sanitarias	7109.0	Vida e integridad personal	7006.0
Personas y bienes protegidos por el dih	6479.0	Secuestro simple	6248.0
Maltrato animal	5867.0	Uso de menores de edad	4430.0
Corrupción electoral	4138.0	Régimen constitucional y legal	3992.0
Salud pública	3079.0	Feminicidio	2371.0

Secuestro extorsivo	2190.0	Aborto	1849.0
Actos de discriminación	1776.0	Derechos de autor	1530.0
Reclutamiento ilícito	1209.0	Derechos de reunión y asocia.	1194.0
Trata de personas	973.0	Lesiones perso agentes químicos	215.0
Otros delitos	97.0	Seguridad del estado	38.0

Nota: Fuente Elaboración Propia

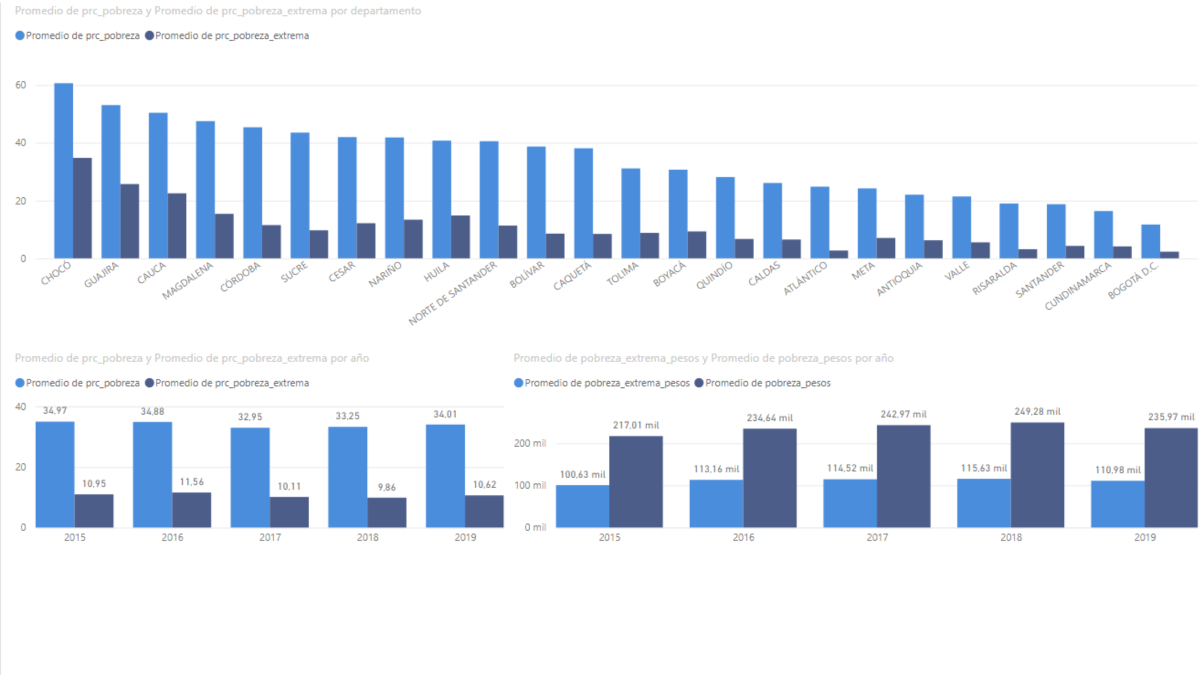

Figura 3. Navegación Datos Pobreza (DANE). Fuente Elaboración Propia

De la misma forma, se llevó a cabo el proceso de exploración con la fuente de pobreza Figura3, donde se logró apreciar que los datos se encuentran consolidados por año, adicionalmente los atributos de medición de la población están en términos de valores porcentuales e igualmente se aprecia el comportamiento para los diferentes departamentos, adicionalmente se identificó inexistencia de registros para la anualidad 2020. Igualmente se aprecia como entre 2015 y 2018 poco a poco subió el valor en pesos que se usaba para catalogar si una persona era pobre o pobre externo en el país; sin embargo, para el 2019 se evidencia que estos valores tuvieron una variación negativa; es decir que en 2018 con 249.280 al mes una persona se consideraría pobre mientras que para el 2019 con 235.970 al mes era considerado pobre. Así mismo se evidencia como los Departamentos de Choco y Guajira albergan los índices de pobreza más altos mientras que Bogotá y Cundinamarca albergan menos población en condición de pobreza.

Modelado

La etapa de modelado comprende la estandarización de los datos a fin facilitar su consumo, para ello se requiere llevar a cabo unas subetapas que son: identificación de la estrategia de datos definición de arquitectura y modelo de datos. A continuación, se abordan cada una de ellas:

Estrategia de datos

El proceso de diseño de los datos experimenta una etapa clave que es la construcción de un modelo que nos permita entender la información desde un punto de vista estratégico, y a su vez da paso a plantear una solución que simplifique el análisis correspondiente a las distintas variables que involucran el análisis de percepción de seguridad ciudadana.

Figura 4. Estrategia de datos planteada Fuente: Elaboración Propia

La estrategia de datos (Ver Figura 4) se plantea inicialmente con unas iniciativas que pueden dar respuesta a los distintos interrogantes planteados por las entidades gubernamentales. A partir de esto, se identificaron estrategias que dan lugar a objetivos misionales que son los que pretenden ayudar a solucionar, la correlación de las variables, a partir del uso de la analítica descriptiva.

Definición de arquitectura

A continuación, se muestra el diseño de arquitectura implementado para la solución planteada, la cual está compuesto por una zona de fuentes de información, que representa el origen de la información, que reposa en archivos planos, a partir de esta zona se ejecuta el primer proceso de carga de datos. (Ver Figura 5)

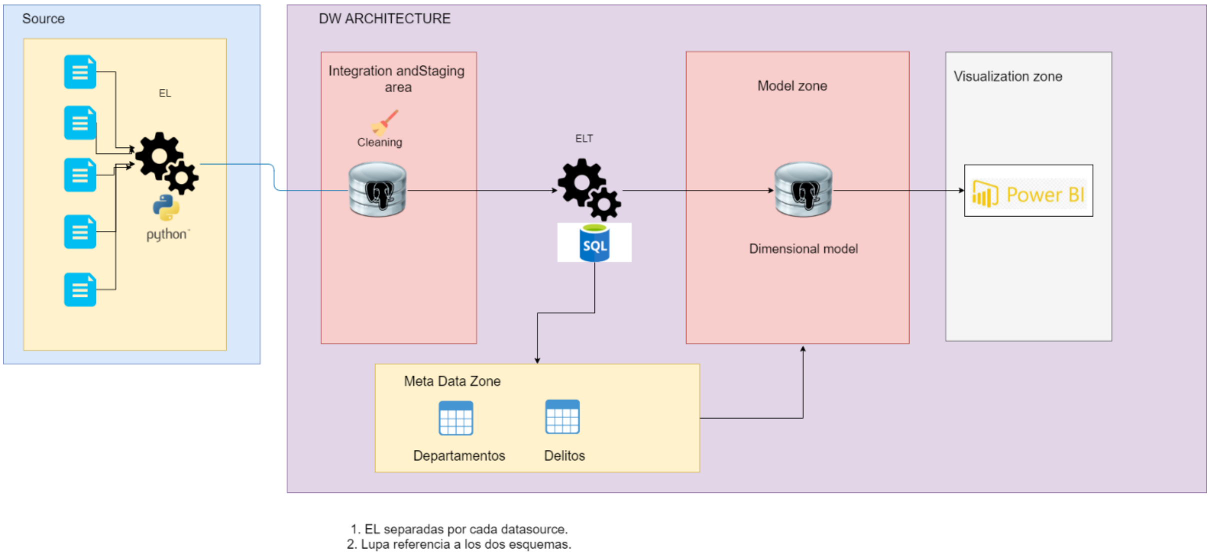

Figura 5. Arquitectura Planteada Fuente: Elaboración Propia

Seguido de la primera carga de información, los datos extraídos deben pasar por un proceso de homologación, en donde se identificaron problemas de calidad de información, por esta razón se hace necesaria aplicar reglas de calidad de datos, que hacen alusión a temáticas relacionadas al correcto gobierno de la información. Una vez aplicadas estas reglas de calidad, los datos son retornados al modelo dimensional que se encargará de almacenar la información de manera que puedan ser consumidos desde Power BI.

Modelo de datos

Se diseñó un modelo dimensional, con el fin de atender las necesidades de correlación de variables, mediante la captura de las métricas de forma periódica, tomando como fuente, los diversos archivos obtenidos desde cada una de entidades públicas Figura 6.

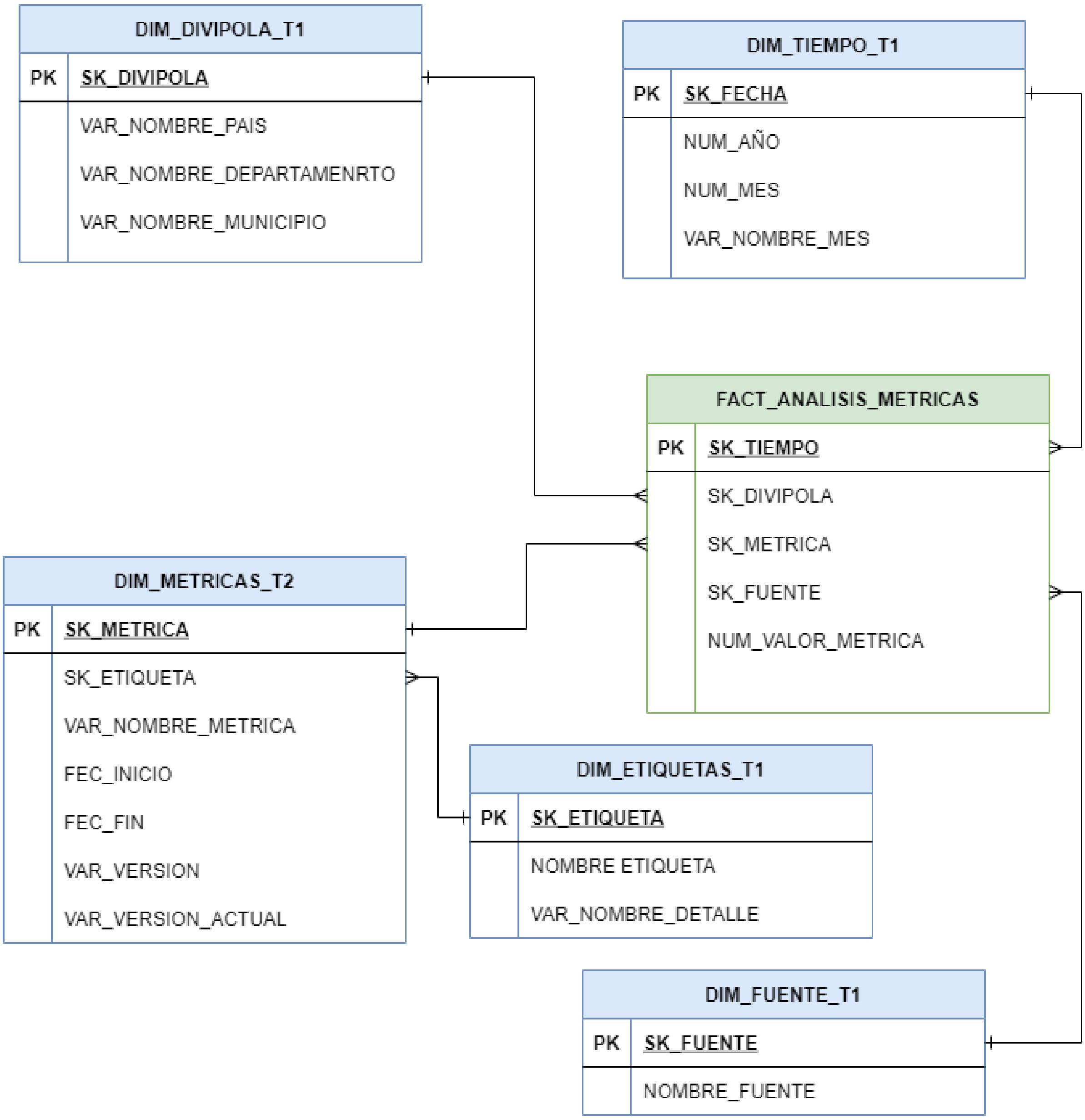

Figura 6. Modelo Físico Planteado. Fuente: Elaboración Propia

A continuación, en la Tabla 2 se describe a alto nivel de modelo dimensional planteado en la Figura 6; en esta tabla se describe lo que se ha de almacenar en la tabla de hechos y las distintas dimensiones, así como la clasificación de las dimensiones según su velocidad de cambio; donde las dimensiones tipo1 son dimensiones que se sobre escriben; por otro lado están las tipo2 que almacena históricos y se establece una tabla de hechos de tipo Periodic Snapshot; ya que representa una visualización acumulativa en un instante de tiempo.

Tabla 2. Descripción de las Dimensiones y Tablas de Hechos Planteadas

Nombre de objeto	Descripción	Tipo de objeto
DIM_DIVIPOLA_T1	Corresponde a la división político-administrativa, cuya granularidad es de municipio en el caso de Colombia y para los demás países se deja su capital.	Dimensión tipo 1.
DIM_TIEMPO_T1	Dimensión que contiene los distintos atributos de tiempo, cuya granularidad es de mes.	Dimensión tipo 1.
DIM_METRICAS_T2	Dimensión que contiene las distintas métricas a evaluar, según la fuente de datos.	dimensión tipo 2.
DIM_FUENTE_T1	Corresponde a las diferentes organizaciones generadoras de datos que gestionan en el modelo.	Dimensión tipo 1.
DIM_ETIQUETAS_T1	Corresponde a las diferentes características que se podrán evaluar.	Dimensión tipo 1.

FACT_ANALISIS_METRICAS	Tabla que contiene los valores de las métricas con su respectiva agregación de tiempo y ubicación geoespacial.	Periodic Snapshot.

Nota: Fuente Elaboración Propia

Modificación

En esta etapa se consumen los conjuntos de datos cargados en bases de datos PostgreSQL, a fin de prepararlos para posteriormente cargarlos un modelo dimensional a través de procesos de limpieza, los cuales se realizan en procedimientos almacenados dentro de PostgreSQL, en este proceso se incluye el tratamiento de datos irrelevantes, nulos y ajustes de los diferentes tipos de datos que se encuentran en los conjuntos.[59]

En el Anexo 1 reposan las transformaciones realizadas a fin de contribuir con la calidad de los datos.

Evaluación

En esta fase se han de normalizar los indicadores empleando técnicas como los porcentajes de variación, la categorización de valores por encima o debajo del promedio, ponderar cada indicador y cada perspectiva, definición de escalas de color y numéricas a manejar, tipos de visualizaciones, el cual facilite la comprensión e identificación de patrones.

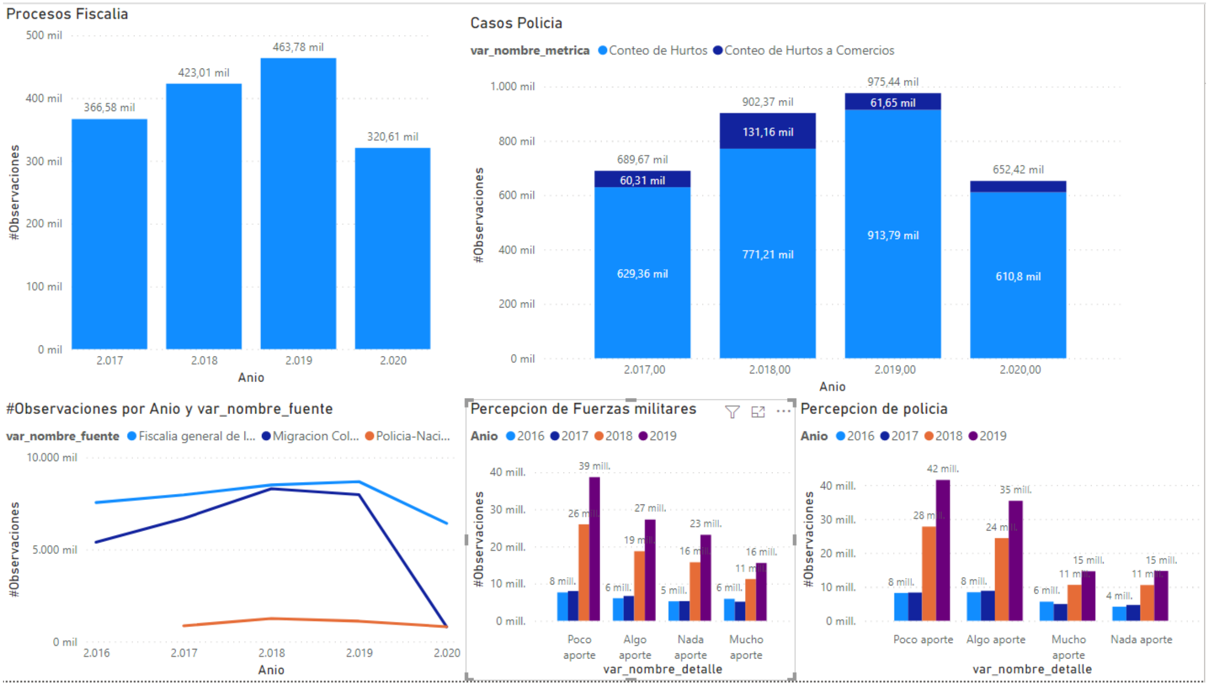

Figura 7. Tablero Integrado de Variables. Fuente: Elaboración Propia

En la Figura7 se presentan gráficos que fueron generados a partir de las variables en el modelo dimensional planteado en las etapas anteriores, donde se aprecia el comportamiento de las variables en los años; de allí se puede apreciar lo siguiente:

- Los casos reportados en la policía difieren considerablemente de los reportados por la fiscalía el delito de Hurto uno de los más representativos, lo cual podría percibirse como una desconfianza de la ciudadanía en las instituciones que imparten justicia o dificultad para acceder a esta.
- Así mismo cuando se analizan variaciones a nivel global, se evidencia un campo de acción más limitado de la policía frente al de la fiscalía cuyas cifras son muy superiores, esto se presenta por el hecho de que no todo delito genera un caso en la fiscalía y viceversa. Sin embargo, se puede apreciar que tienen un comportamiento similar en términos de tendencia con los datos de migración.

- Por otra parte, se percibe que el estudio de percepción de seguridad ciudadana a ampliado su población muestral, sin que esto modifique la tendencia de los datos donde se aprecia que la percepción de la ciudadanía frente al aporte que hacen instituciones como la policía y fuerzas militares tienen poco impacto en la población.

CONCLUSIONES

- En el proceso de implementación de esta solución se evidenciaron múltiples problemas de calidad de datos en las fuentes de datos abiertos, ya que se hallaron tanto valores faltantes como atípicos en diferentes conjuntos de datos abiertos.
- Se comprobaron claras dificultades a la hora de relacionar las diferentes fuentes de datos abiertas en el país, ya que los niveles de granularidad en términos temporales y geográficos son bastante dispares.
- Igualmente se identifica que los reportes generados por migración Colombia frente a la información de emigrantes venezolanos cuenta con algunas falencias, ya que los reportes generados no incluyen todos los departamentos colombianos y pueden variar en cantidad de registros de un año a otro.

- Existen múltiples herramientas que permiten el análisis de datos, tales como la inteligencia de negocio, inteligencia artificial, minería de datos cada uno de ellos enfocado a resolver diferentes tipos de preguntas de analítica, algunas prescriptivas y otras predictivas; estas por lo general hacen uso de diversas fuentes de datos que durante el proceso se limpian, estandarizan y transforman para procesarlos según las necesidades que se tengan alineadas con la estrategia.

- Todo proceso de analítica requiere una arquitectura la cual soporte que se realiza para la explotación de datos; está compuesta por diferentes componentes, los cuales respondan a las necesidades de la solución, en este caso una bodega de datos que requiere de herramientas para el cargue de los datos desde las diferentes fuentes requeridas, así mismo, los procesos de transformación y el almacenamiento que agilice el proceso de consumo que los datos; que bien puede realizarse por herramientas visuales o por reportes.

- Así mismo se evidencio que debido a la disparidad existente en los datos no se puede realizar un estudio de correlación; ya que no se evidencia relaciones en términos de proporcionalidad entre las variables estudiadas.

REFERENCIAS

[1] Castillo, J. (04 de febrero de 2021). *Mala percepción de seguridad: una bicicleta estática*. Obtenido de https://www.eltiempo.com/bogota/opinion-564823

[2] Bustamante Fernandez, M. V. (05 de marzo de 2021). *Anuncian nuevas acciones para hacerle frente a la inseguridad en Sucre*. Obtenido de https://www.elheraldo.co/sucre/anuncian-nuevas-acciones-para-hacerle-frente-la-inseguridad-en-sucre-799353

[3] Oróstegui, Ó. (28 de enero de 2021). *¿Quién controla a la policía?* Obtenido de https://www.eltiempo.com/bogota/quien-controla-a-la-policia-opinion-563358

[4] Pérez Díaz, V. (01 de febrero de 2021). *Percepción de inseguridad en Bogotá está en el nivel más alto en los últimos cinco años según la CCB*. Obtenido de https://www.asuntoslegales.com.co/actualidad/percepcion-de-inseguridad-en-bogota-esta-en-el-nivel-mas-alto-en-los-ultimos-cinco-anos-3120141

[5] Bogotá, F. (21 de febrero de 2021). *El reto de la percepción en seguridad*. Obtenido de https://www.eltiempo.com/bogota/el-reto-de-la-percepcion-en-seguridad-opinion-568429

[6] Institute for Economics & Peace. (Junio de 2014). *Global Peace Index 2014: Measuring Peace in a Complex World*. Obtenido de

https://www.economicsandpeace.org/wp-content/uploads/2015/06/2014-Global-Peace-Index-REPORT_0-1.pdf

[7] Institute for Economics & Peace. (Junio de 2018). *Global Peace Index 2018: Measuring Peace in a Complex World*. Obtenido de https://www.economicsandpeace.org/wp-content/uploads/2020/08/Global-Peace-Index-2018-2-1.pdf

[8] DANE. (24 de febrero de 2021). *Encuesta de Convivencia y Seguridad Ciudadana ECSC 2020*. Obtenido de https://www.dane.gov.co/files/investigaciones/poblacion/convivencia/2019/Presentacion_v_corta_ECSC_2019.pdf

[9] Riquelme, J. C., Ruiz, R., & Gilbert, K. (2006). Minería de Datos: Conceptos y Tendencias. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial*, 11-18.

[10] Quintero Cordero, S. P. (2020). Seguridad ciudadana y participación de las comunidades en América Latina. *Revista Científica General José María Córdova*, 5-24.

[11] Rincón, A. (2018). Abordajes teóricos sobre la relación entre seguridad ciudadana y violencia urbana en Colombia: una lectura crítica. *URVIO Revista Latinoamericana de Estudios de Seguridad*, 86-100.

[12] Huamani Cahua, J., Lazo Manrique de Vargas, M., & Calizaya López, J. (2019). Percepción de la seguridad ciudadana en pobladores de un distrito de la ciudad de Arequipa. Revista De Investigación En Psicología, 95-110.

[13] REAL ACADEMIA ESPAÑOLA. (s.f.). *Diccionario de la lengua española, 23.ª ed.* Obtenido de https://dle.rae.es

[14] Moya Vargas, M. F., & Bernal Castro, C. A. (2015). *Los menores en el sistema penal colombiano.* Bogotá D.C - Colombia: Universidad Católica de Colombia.

[15] *Código de Procedimiento Penal [CPP]. Ley 906 de 2004.* (31 de Agosto de 2004). Obtenido de http://www.secretariasenado.gov.co/senado/basedoc/ley_0906_2004.html

[16] Fiscalía General de la Nación. (14 de enero de 2008). *Estructura del Proceso Penal Acusatorio.* Obtenido de https://www.fiscalia.gov.co/colombia/wp-content/uploads/2012/01/EstructuradelProcesoPenalAcusatorio.pdf

[17] Polo, S., Serrano, E., & Jiménez, S. (2019). Las migraciones colombianas hacia Perú: la invariabilidad de los flujos migratorios en un periodo de auge de la diáspora (2005-2015). *Ciencia Política*, 143-174.

[18] Pan American Health Organization. (s.f.). *Migración nacional e internacional*. Obtenido de https://www.paho.org/salud-en-las-americas-2017/?post_type=post_t_es&p=313&lang=es

[19] Stezano, F. (2021). *Enfoques, definiciones y estimaciones de pobreza y desigualdad en América Latina y el Caribe: un análisis crítico de la literatura.* Ciudad de México: Comisión Económica para América Latina y el Caribe (CEPAL).

[20] Huijboom, N., & Van den Broek, T. (2011). Open data: an international comparison of strategies. European journal of ePractice, 12(1), 4-16. Obtenido de https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.465.2613&rep=rep1&type=pdf

[21] Hidalgo Delgado, Y., Mariño Molerio, A. J., Amoroso Fernández, Y., & Leiva Mederos, A. A. (2018). Some Reflections on Linked Open Data in Cuba. Revista Cubana de Información en Ciencias de la Salud, 29(4), 1-9. Obtenido de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2307-21132018000400009&lng=es&tlng=en

[22] Universidad Tecnológica De Pereira. (s.f.). *ANÁLISIS DE CORRELACIONES.* Obtenido de https://academia.utp.edu.co/seminario-investigacion-II/files/2017/03/06a.An%C3%A1lisisDeCorrelaciones.pdf

[23] Coronado Medina, L. A. (2019). Analítica de datos : un estudio de caso de su uso para identificar riesgos estratégicos en grandes compañías de Medellín. Obtenido de https://repository.eafit.edu.co/bitstream/handle/10784/14347/LuisAlejandro_CoronadoMedina_2019.pdf?sequence=2&isAllowed=y

[24] Gartner, Inc. (s.f.). Information Technology Gartner Glossary. Obtenido de Descriptive Analytics: https://www.gartner.com/en/information-technology/glossary/descriptive-analytics

[25] Rosado Gómez, A. &. (2010). Inteligencia de negocios: Estado del arte . Scientia Et Technica, 1(44), 321-326. Obtenido de https://doi.org/10.22517/23447214.1803

[26] Gartner, Inc. (s.f.). Information Technology Gartner Glossary. Obtenido de Analytics and Business Intelligence (ABI): https://www.gartner.com/en/information-technology/glossary/business-intelligence-bi

[27] Ferreira, T., Pedrosa, I., & Bernardino, J. (2019). Integration of Business Intelligence with e-commerce. 2019 14th Iberian Conference on Information Systems and Technologies (CISTI), 1-7. Obtenido de 10.23919/CISTI.2019.8760992

[28] Reddy, G., Srinivasu, R., Rao M, P., & Reddy Rikkula, S. (2010). Data warehousing, data mining, OLAP and OLTP technologies are essential elements to support decision-making process in industries. (IJCSE) International Journal on Computer Science and Engineering, 2865-2873.

[29] Kakish, K., & Kraft, T. A. (2012). ETL evolution for real-time data warehousing. Proceedings of the Conference on Information Systems Applied Research, 1508.

[30] Cobos, C., Muñoz, J., Mendoza, M., Acosta Muñoz, L., & Gómez, L. (2006). Bodega de datos y olap en unicauca virtual. Revista UIS Ingenierías, 99-109.

[31] Roldan Pinzón, D. E. (2017). Diseño de una guía general para construir una bodega de datos del área de ventas de una empresa. Obtenido de https://repository.unilibre.edu.co/handle/10901/11042

[32] Fernández Morales, M., & Bonilla Carrión, R. (2020). Bibliominería, datos y el proceso de toma de decisiones. Revista Interamericana de Bibliotecología. Obtenido de https://doi.org/10.17533/udea.rib.v43n2ei8

[33] Bimonte, S., Hifdi, Y., Maliari, M., Patrick, M., & Rizzi, S. (17 de Noviembre de 2020). To Each His Own: Accommodating Data Variety by a Multimodel Star Schema. Obtenido de the 22nd International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data co-located with EDBT/ICDT 2020 Joint Conference: https://hal.archives-ouvertes.fr/hal-03009808/document

[34] Podaras, A. (s.f.). Data - Based Agricultural Business Continuity Management. Obtenido de https://www.researchgate.net/profile/Athanasios-Podaras/publication/348806019_Data_-Based_Agricultural_Business_Continuity_Management_Policies/links/60112a1b45851517ef1a345d/Data-Based-Agricultural-Business-Continuity-Management-Policies.pdf

[35] Han, J. (1998). Olap mining: an Integration of olap with data mining. Data Mining and Reverse Engineering, 3-20.

[36] Heinonen, J. (2020). From Classical DW to Cloud Data Warehouse. Obtenido de https://helda.helsinki.fi/bitstream/handle/10138/322467/JyrkiHeinonen_Masters_Thesis_V1.0.pdf?sequence=2&isAllowed=y

[37] Cai, L., & Zhu, Y. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. Data Science Journal, 14, 2. Obtenido de http://doi.org/10.5334/dsj-2015-002

[38] Pipino, L., Lee, Y., & Wang, R. (2002). Data Quality Assessment. Communications of the ACM Vol45, 211-218. Obtenido de http://web.mit.edu/tdqm/www/tdqmpub/PipinoLeeWangCACMApr02.pdf

[39] Rafique, I., Lew, P., Abbasi , M. Q., & Zhang, L. (2012). Information Quality Evaluation Framework: Extending ISO 25012 Data Quality Model. World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering Vol:6, No:5, 568-573.

[40] iso25000.com. (s.f.). ISO25000. Obtenido de ISO/IEC 25012: https://iso25000.com/index.php/normas-iso-25000/iso-25012?start=0

[41] Calabrese, J., Esponda, S., Pasini, A., Boracchia, M., & Pesado, P. (2019). Guía para evaluar calidad de datos basada en ISO/IEC 25012. XXV Congreso Argentino de Ciencias de la Computación (CACIC), (págs. 694-706). Río Cuarto. Obtenido de http://sedici.unlp.edu.ar/bitstream/handle/10915/91086/Documento_completo.pdf-PDFA.pdf?sequence=1&isAllowed=y

[42] Gómez Quintero, D., & López Murillo, L. Y. (2018). Guía metodológica para la depuración de datos del sistema de información y registro cinematográfico – SIREC, del ministerio de cultura. Universidad Nacional Abierta y a Distancia UNAD. Obtenido de https://repository.unad.edu.co/handle/10596/18280

[43] Álvarez Sarmiento, K. L. (2020). *Investigación y análisis de herramientas para extracción de Tweets sobre COVID19 focalizadas en RStudio y Python que permitan crear una base de datos relacional.* Guayaquil: Universidad de Guayaquil. Facultad de Ciencias Matemáticas y Físicas. Carrera de Ingeniería en Networking y Telecomunicaciones.

[44] Bilheux, J.-C., Bilheux, H., Lin, J., Lumsden, I., & Zhang, Y. (2019). Neutron imaging analysis using jupyter Python notebook. Journal of Physics Communications, 3(8), 083001. Obtenido de https://iopscience.iop.org/article/10.1088/2399-6528/ab3bea/meta

[45] León Soberón, J. J. (2020). *Análisis comparativo de sistemas gestores de bases de datos postgresql y mysql en procesos crud.* Pimentel: Universidad Señor de Sipán.

[46] Guerrero Fonseca, J. M. (Julio de 2018). *Un índice dinámico para la seguridad ciudadana en Colombia: Un acercamiento bayesiano.* Obtenido de https://repository.usta.edu.co/bitstream/handle/11634/12493/2018juanguerrero.pdf?sequence=1&isAllowed=y

[47] Esri Colombia. (s.f.). *Solución Observatorio y Análisis del delito.* Obtenido de https://seguridad-gov-esri-co.hub.arcgis.com/

[48] Cámara de Comercio de Bogotá (CCB). (2020). *Observatorio de la regón Bogotá - Cundinamarca.* Obtenido de https://www.ccb.org.co/observatorio/Entorno/Entorno-favorable-para-los-negocios/Seguridad-y-convivencia

[49] Universidad Del Norte. (s.f.). *Observatorio de Seguridad ciudadana.* Obtenido de https://www.uninorte.edu.co/web/departamento-de-ciencia-politica-y-relaciones-internacionales/observatorio-de-seguridad-ciudadana

[50] IBM. (s.f.). Incident and Emergency Management. Obtenido de https://www.ibm.com/analytics/nz/en/industry/government/emergency-management.html

[51] Peña Suarez, A. (Julio de 2017). *Modelo para la Caracterización del Delito en la Ciudad de Bogotá, Aplicando Técnicas de Minería de Datos Espaciales.* Obtenido de https://repository.udistrital.edu.co/bitstream/handle/11349/6519/Pe%c3%b1aSuarezAlfonso2017.pdf?sequence=1&isAllowed=y#page=33&zoom=100,92,354

[52] Aguilar, A. L., & Contreras B., M. C. (Septiembre de 2018). *Caracterización de los delitos en Cartagena mediante la aplicación de minería de datos.* Obtenido de https://biblioteca.utb.edu.co/notas/tesis/0074619.pdf

[53] Rodríguez Rojas, L. A. (13 de Junio de 2017). Metamodelo para integración de datos abiertos aplicado a inteligencia de negocios. Obtenido de http://hdl.handle.net/10651/44552

[54] León Guzmán, E. (s.f.). *Metodologías aplicadas al proceso de Minería de Datos.* Obtenido de https://disi.unal.edu.co/~eleonguz/cursos/md/presentaciones/Sesion5_Metodologias.pdf

[55] Hernández, C. L., & Dueñas, M. X. (2009). Hacia una metodología de gestión del conocimiento basada en minería de datos. *COMTEL*, 79-95.

[56] Bisong, E. (2019). Introduction to Scikit-learn. En E. Bisong, *Building Machine Learning and Deep Learning Models on Google Cloud Platform.* Berkeley: Apress.

[57] Pandas. (s.f.). *About pandas*. Obtenido de https://pandas.pydata.org/about/

[58] Python Software Foundation. (s.f.). *configparser — Configuration file parser*. Obtenido de https://docs.python.org/3/library/configparser.html

[59] Python Software Foundation. (s.f.). *psycopg2 - Python-PostgreSQL Database Adapter*. Obtenido de https://pypi.org/project/psycopg2/

[60] Python Software Foundation. (s.f.). *Database Abstraction Library*. Obtenido de https://pypi.org/project/SQLAlchemy/

[61] Torres Álvarez, L. (27 de 07 de 2020). Power BI Desktop: en qué consiste y cuáles son sus ventajas. Obtenido de https://www.cice.es/noticia/power-bi-desktop/

[62] Bisong, E. (2019). Introduction to Scikit-learn. En E. Bisong, *Building Machine Learning and Deep Learning Models on Google Cloud Platform.* Berkeley: Apress.