Desarrollo de una plataforma de análisis de datos para el estudio de la brecha digital aprovechando el ecosistema de datos abiertos de Colombia.

Javier Mauricio Bedoya González

Universidad Nacional Abierta y a Distancia – UNAD
Escuela de Ciencias Básicas, Tecnologías e Ingeniería
Ingeniería de Sistemas
Octubre 2021

Desarrollo de una plataforma de análisis de datos para el estudio de la brecha digital aprovechando el ecosistema de datos abiertos de Colombia.

Javier Mauricio Bedoya González

Trabajo de grado presentado para optar al título de Ingeniero de Sistemas

Asesor de Proyecto:

PhD. Jheimer Julián Sepúlveda López

Universidad Nacional Abierta y a Distancia – UNAD
Escuela de Ciencias Básicas, Tecnologías e Ingeniería
Ingeniería de Sistemas
Octubre 2021

Nota de Aceptación

_____

_____

_____

_____

Presidente del Jurado

_____

Jurado

_____

Jurado

Dosquebradas, agosto de 2021.

Dedicatoria

A Dios que ha labrado mi camino y me ha dirigido por el sendero correcto el cual me permite recolectar los frutos, a mis padres Javier Bedoya y María Nohemy González que con su esfuerzo y abnegación edificaron en mí las bases sólidas para un buen desempeño personal y profesional, a mi familia y en especial a mi tía María Luz Dary González por su generosidad, amor y apoyo incondicional que hicieron posible alcanzar esta meta trazada.

Por último y no menos importante a mis grandes amigos que a lo largo de mi vida han aportado y de los cuales he tomado siempre lo mejor para seguir construyendo mi proyecto de vida.

Agradecimientos

Agradezco la oportunidad que me brindó la Universidad Nacional abierta y a distancia, al permitirme por medio de la modalidad de educación virtual continuar con mi formación profesional, colmándome de conocimientos y experiencias dignas de mi reconocimiento.

Sin olvidar ni dejar atrás la institución a la cual siento que le debo mi vida al abrirme las puertas cuando consideraba que no tenía ninguna oportunidad, mi amado SENA - Servicio Nacional de Aprendizaje el cual me permitió escalar el primer peldaño, no solo en el ámbito académico sino también en el laboral y profesional.

Mi gran admiración y respeto a mi asesor PhD Jheimer Julián Sepúlveda López
que con su entrega y profesionalismo hizo posible que mi sueño se hiciera realidad.

A mis compañeros de trabajo y amigos, en especial a la magister Blanca Inés Martínez López que en su debido momento intervinieron con sus aportes y sugerencias, enriqueciendo mi trabajo; mi inmensa gratitud que quedará plasmada en mi mente y corazón por siempre.

¡¡¡ Gracias infinitas !!!

Resumen

El presente documento muestra los resultados de la investigación, la cual describe los diversos lenguajes de programación, técnicas y plataformas tecnológicas empleadas en el análisis de datos. Aprovechando el ecosistema de datos abiertos en Colombia se seleccionó la temática de la brecha o inclusión digitales y se tomó como base el insumo que entrega el ministerio de las TIC con la gran primera encuesta que unifica información o investigaciones de diversas fuentes sobre el acceso y utilización de las herramientas tic por parte de la población colombiana.

Se implementa la herramienta tecnológica Tableau y se ilustra paso a paso de sus requerimientos y funcionalidades con el fin de darle un valor agregado a la información lo que redunda en un aporte en la disminución del fenómeno antes mencionado; que ayudará a definir estrategias para combatirlo. Entregando como producto final las gráficas y reportes de la herramienta con el procedimiento de uso de esta, donde se evidencia el aprendizaje y la información generada como aporte significativo a otros procesos de investigación.

Abstract

This document shows the results of the research, which describes the various programming languages, techniques and technological platforms used in data analysis. Taking advantage of the open data ecosystem in Colombia, the theme of the digital divide or digital inclusion was selected and the input provided by the ICT ministry with the great first survey that unifies information or research from various sources on access and Use of ICT tools by the Colombian population.

The Tableau technological tool is implemented and its requirements and functionalities are illustrated step by step in order to give added value to the information, which results in a contribution to the reduction of the aforementioned phenomenon; which will help define strategies to combat it. Delivering as a final product the graphs and reports of the tool with the procedure for using it, where the learning and information generated as a significant contribution to other research processes is evidenced.

Tabla de contenido

Lista de ilustraciones

Lista de tablas

Introducción

La brecha digital es la diferencia que existe entre las personas respecto al acceso a las herramientas tecnológicas, su capacidad para utilizarlas y cumplir sus objetivos personales. Para reducir esta brecha, los procesos se emprenden en el marco del fenómeno conocido como inclusión digital. Los gobiernos, organizaciones e instituciones relacionadas con las tecnologías reconocen la naturaleza social de la brecha digital y la necesidad de involucrar a las comunidades en el proceso de gestión de programas y proyectos asociados; sin embargo, estas comunidades no son consideradas en las propuestas que se hacen. Además, se ha identificado que existen diferentes características individuales y grupales que hacen posible la apropiación de las herramientas tecnológicas y su uso para lograr objetivos y mejorar la participación en el trabajo, la educación y las actividades políticas. Con base en este postulado, en este documento se hace una propuesta de proyecto para realizar el análisis del fenómeno de la brecha digital en Colombia mediante una plataforma que permita el análisis y el uso de datos abiertos para comprender las configuraciones de las comunidades.

En la actualidad se generan por las personas, empresas u organizaciones grandes volúmenes de datos, los cuales requieren ser administrados, gestionados y representados de manera visual para que permitan la toma efectiva de decisiones, claridad en los conceptos, explorar opciones y descubrir nuevas oportunidades. A estos grandes volúmenes de información se les denominó macrodatos o big data, ya que por la complejidad de su análisis requieren de aplicaciones informáticas especializadas que permitan el procesamiento de estos para poderlos tratar de una manera óptima.

De ahí la importancia de involucrar el análisis de datos para que las organizaciones puedan tomar decisiones acertadas de forma razonada que impacten la productividad. Con el fin de

resolver esta dificultad identificada se han venido desarrollando por parte de los profesionales en informática y otras áreas lenguajes de programación, plataformas, herramientas informáticas y técnicas de análisis de datos inteligentes, como las que se describen en este proyecto de investigación.

Después de especificar las diferentes plataformas y analizarlas a detalle se decide implementar la herramienta TABLEAU, por ser una plataforma de complejidad baja, lo cual facilita su uso y por no requerir que los usuarios tengan una vasta experiencia en programación haciéndola muy intuitiva, permitiendo al usuario la posibilidad de profundizar y explorar datos. Esta plataforma proporciona mayor amplitud y flexibilidad analítica, sus funcionalidades y características la convierten en una herramienta poderosa y adaptable para la toma de decisiones.

Se tomará un archivo plano resultado de una encuesta, test o formulario aplicado a los usuarios, posteriormente se estructurará la información utilizando el concepto de bodega de datos para alimentar a Tableau y así poder filtrar la información y generar gráficos que sean relevantes y conduzcan a informes que permitan mitigar o minimizar la brecha digital.

Es también importante incursionar en la minería de datos o big data, ya que se ha convertido en una herramienta indispensable, cada vez las empresas, organizaciones o personas en investigación generan grandes volúmenes de información y está por si sola sin ser gestionada no tiene ninguna utilidad; por ello el ingeniero de sistemas debe estar en la capacidad de construir un flujo de datos, combinar tecnologías y herramientas para así comprender e interactuar con todos estos conceptos lo cual lo formará como un profesional integral que aporta valor a las organizaciones o procesos.

Capítulo I: Generalidades

Planteamiento del problema

Diversos autores han escrito sobre la dificultad que afrontan todas las organizaciones, en donde se requiere dar un buen uso a los grandes volúmenes de información, que le permitan a las organizaciones mejorar en la toma de decisiones. Por lo considerado anteriormente se requiere determinar la mejor herramienta que permita realizar dichos análisis, en particular de la temática la brecha digital. De igual manera se pretende buscar una herramienta rápida, amigable, flexible e intuitiva que permita al usuario por medio de una interfaz simple y sin conocimientos avanzados de lenguajes de programación el cargue de los datos y el análisis de estos los cuales generarán resultados gráficos y estructurados que redundan en la generación de nuevas ideas, para mejorar los procesos de la organización. “La analítica de datos implica los procesos y actividades diseñadas para obtener y evaluar datos para extraer información útil”, Data analytics, ISACA (2011). En ese mismo sentido se pretende que dichos datos se puedan utilizar para: Reconocer puntos claves de riesgo, errores, para mejorar los procesos de las organizaciones o incluso influir en las decisiones.

Es precisamente la eficacia de los logros del análisis de los datos de la brecha digital lo que se pretende fomentar con esta investigación impactando de manera positiva la inclusión digital en las comunidades analizadas en el territorio nacional.

La inclusión digital es el resultado de procesos que emprende un gobierno, una entidad privada o pública, con o sin ánimo de lucro, con el propósito de cerrar la brecha digital existente en una comunidad determinada. Para poder comprender el alcance de la inclusión digital, se hace necesario definir, a su vez, lo que es la brecha digital, entendida como “la brecha entre individuos, hogares, empresas y áreas geográficas en los diferentes niveles

socioeconómicos en lo que respecta tanto a las oportunidades de acceso a las tecnologías de información y la comunicación (TIC) y el uso de Internet para una amplia variedad de actividades" (OECD, 2001).

La brecha digital es un "problema social complejo" que involucra aspectos diversos y que distintas organizaciones y entidades buscan solucionar por medio del desarrollo del talento humano que hace parte de una región. La brecha digital es un fenómeno dinámico ya que, a pesar de realizar una serie de procesos con el propósito de acercar las TIC a esas comunidades y personas desfavorecidas, no se está logrando el objetivo esperado, esto lo expresa el Banco Mundial en su libro "Dividendos Digitales":

"nos encontramos en medio de la mayor revolución de la información y las comunicaciones de la historia de la humanidad. Más del 40 % de la población mundial tiene acceso a Internet, y todos los días se suman nuevos usuarios. Asimismo, en casi 7 de cada 10 hogares ubicados en el 20% más pobre hay un teléfono celular. Es más probable que los hogares más pobres tengan acceso a la telefonía celular que a un inodoro o al agua potable" (Banco Mundial, 2016).

Lo anterior, muestra la volatilidad del fenómeno y como, en ocasiones, la tecnología no cumple sus objetivos con relación al mejoramiento o apoyo a los procesos de desigualdad social. En este mismo informe, el Banco Mundial indica que "aún hay 4000 millones de personas que no tienen acceso a Internet" (Banco Mundial, 2016), esto permite concluir que la brecha digital continúa aumentando. Con base en lo anterior, se resalta la importancia de abordar los programas y proyectos para reducir la brecha digital haciendo explícitos los aspectos sociales, de lo que se configura como un fenómeno con una perspectiva social. Inicialmente, las cifras presentadas, muestran, en primera medida que, desde la práctica, no

se hace explícito el abordaje de los programas/proyectos y del fenómeno de la brecha digital como social y complejo. La propuesta que se desarrolla en este documento plantea abordar el fenómeno para entender la perspectiva social del mismo por medio de una plataforma que permita el análisis y el uso de bases de datos abiertas en Colombia.

Objetivos

Objetivo general

Desarrollar una plataforma de análisis de datos para estudiar la brecha digital aprovechando el ecosistema de datos abiertos de Colombia.

Objetivos específicos

Analizar las diferentes técnicas y plataformas para el análisis de datos.

Realizar análisis, diseño y desarrollo de la plataforma.

Realizar un diagnóstico de la brecha digital en Colombia por medio de la plataforma.

Justificación

Con base en la problemática identificada, existe un fenómeno que se identifica como brecha digital, el cual analiza la diferencia que existe entre las personas que tienen y no tienen acceso a las Tecnologías de la Información y la Comunicación y el uso dado a estas. De igual forma, en el país existe una tendencia para el uso y aprovechamiento de los datos abiertos.; esto se basa en el CONPES 3920 denominado " Política nacional de explotación de datos (Big Data)", documento en el cual se establecen los lineamientos que rigen el ecosistema de datos de Colombia. En el marco de esta estrategia, se creó la plataforma datos.gov.co; plataforma en la cual las instituciones públicas y empresas del país deben publicar sus datos. Al revisar esta gran bodega de datos, se ha identificado la Primera Gran Encuesta TIC 2017, encuesta que permite consolidar históricamente datos dispersos en los diferentes indicadores sectoriales y diversas variables relacionadas con el acceso y uso de las TIC.

Aunque se evidencia que existe gran necesidad de estudiar este fenómeno que afecta a la sociedad del país y que se cuenta con los datos para hacerlo, se debe anotar que no se está ejecutando una herramienta tecnológica que permita analizar estos grandes volúmenes de datos que a la fecha existen. Con base en lo anterior, se hace necesario implementar una herramienta de análisis de datos rápida, amigable, flexible, intuitiva con una interfaz simple que permita estudiar los datos disponibles y comprender el comportamiento de los fenómenos para este caso particular, diagnosticar la brecha digital en Colombia.

El proyecto es viable por que la analítica de datos se ha convertido en una herramienta fundamental para mejorar el rendimiento de las organizaciones soportada en la toma de decisiones que afectan su estabilidad y productividad en diferentes ámbitos y campos.

La novedad de este estudio está dada en que a pesar de que se han hecho algunas investigaciones sobre esta temática, hasta ahora en los artículos revisados no se han utilizado herramientas de análisis de datos que le permitan generar gráficas o estadísticas para analizar el fenómeno de la brecha digital.

Capitulo II: Marco de referencia

A continuación, se mencionan las bases teóricas que sustentan la investigación con el fin de referenciar antecedentes históricos e investigativos.

Antecedentes

La gran cantidad de información presente en las organizaciones hace necesario poder determinar herramientas versátiles que permitan el análisis de los datos con el fin de apoyar en la toma de decisiones,

En algunas investigaciones relacionadas con Big Data los autores Méndez y Romero (2020), se preguntaron sobre cómo identificar la percepción de los estudiantes de la Universidad de Pamplona y de usuarios de Twitter sobre la violencia digital en redes y el proceso de paz en Colombia, para concretar una propuesta educativa mediada por las TIC.

Los resultados de esta investigación surgen de los datos obtenidos de la encuesta realizada a la población estudiantil de la Universidad de Pamplona, la encuesta Gallup 130 de mayo de 2019 y las nubes de conceptos extraídos a través de herramientas de minería de texto a partir de la información descargada de Twitter, con los cuales se identifican percepciones negativas, sentimientos de desesperanza e incertidumbre sobre el proceso de paz, evidenciando que, en el escenario de post-acuerdo, el Ministerio de Educación y las instituciones educativas en todos los niveles deben reformular o replantear las políticas y contenidos educativos de dicha cátedra, debido a que según los resultados obtenidos los contenidos no han sido suficientemente orientadores y eficientes para contrarrestar prácticas y sentimientos asociados a la violencia. Las fuentes de recolección de datos son complementarias y ninguna excluye a la otra, lo que permite constatar que en la opinión pública aún no hay consenso sobre el proceso de paz en Colombia y si, por el contrario, demuestra una gran

polarización en la población. Los datos recolectados y comparados demuestran que la violencia sigue vigente en algunas partes del territorio colombiano, principalmente en las zonas rurales y la periferia de las zonas urbanas, demuestran que en los imaginarios de las personas aún se mantienen sentimientos de venganza, frustración y miedo, además de marcadas dicotomías morales sobre los grupos que representan "el bien y el mal", lo cual se puede visualizar en los textos escritos oralizados o de una manera menos engolada los comentarios o tweets rastreados.

En síntesis, se puede concluir que las herramientas de minería de datos permiten identificar información relevante o muy importante en grandes volúmenes de datos. De forma ágil se detectan reglas, comportamientos, patrones o tendencias que exponen el proceder de la información o los datos en un contexto específico. Aplicado a nuestro proceso de investigación identificar o analizar el fenómeno de la brecha o inclusión digital y su comportamiento tomando como base las investigaciones e información ya recolectadas por el ministerio de las TIC.

En comparativo la investigación que se cita hace uso también de lenguajes de programación como R y herramientas tecnológicas como Tableau y Power BI para ilustrar los resultados y afianzar asertivamente la toma de decisiones o demostrar la importancia de sus planteamientos.

El propósito del artículo realizado por Piedra y Ponjuán (2021) fue analizar los patrones de colaboración del programa de formación doctoral en Bibliotecología y Documentación científica desarrollado entre la Universidad de La Habana y la Universidad de Granada en

el período 2007-2017. Para esto se creó una base de datos en EndNote® x.9, con 396 documentos. Se crearon listados de frecuencia de acuerdo con los indicadores analizados, los cuales se procesaron con los programas Excel y Tableau Public 2020.3 para generar tablas y gráficos. Se utilizó Bibexcel (Olle Persson, Universidad de Umeå, Suecia) para realizar los

conteos de frecuencia generales, la generación de matrices y el análisis de las redes de coautoría, cotutoría y de colaboración entre instituciones, en aras de procesarlas con

UCINET 6.175. Para su representación reticular se utilizaron NetDraw 2.38 y VOSviewer 1.6.16. La colaboración fue analizada de manera global, por grupos y por tipología documental. Se valoraron las relaciones establecidas para el desarrollo de las investigaciones y para la dirección de las tesis. Para los artículos se analizaron las redes de coautoría y los nexos interinstitucionales.

En las tesis se analizaron las relaciones establecidas para la tutoría. Se identificó un predominio de autoría múltiple, mayoritariamente en los artículos científicos. Se aprecian nexos relativamente importantes en la tutoría a partir del establecimiento de relaciones entre los tutores más productivos del programa. A nivel institucional se aprecia un protagonismo de la Universidad de Granada y la Universidad de La Habana por ser las coordinadoras del programa. No obstante, se aprecia una amplia gama de instituciones nacionales. Se reflejan los participantes y tutores más representados.

Los resultados obtenidos en la presente contribución reafirman la enorme relevancia del estudio de las tesis doctorales como objeto de análisis. Se corrobora lo planteado por Repiso, Torres Salinas y Delgado López-Cózar(8) cuando afirmaron que estas constituyen uno de los mejores espejos donde se reflejan las líneas, las tendencias y las potencialidades de la investigación de las universidades. Es importante basados en lo anteriormente mencionado resaltar la relevancia de la utilización de herramientas tecnológicas o plataformas de análisis de datos. En la gran mayoría de investigaciones consultadas estas realizaron aportes determinantes, pero también se pudo evidenciar que gracias a la percepción errónea o el desconocimiento de su uso no se implementan con mayor frecuencia.

Marco teórico

Este trabajo de investigación estará enmarcado en varios aspectos fundamentales como: la brecha digital, las técnicas de análisis de datos, minería de datos, big data, plataformas tecnológicas que facilitan análisis de datos, la ciencia de datos y el machine learning.

Conceptualización brecha digital

La brecha digital es comúnmente definida como la diferencia que existe entre aquellas personas que tienen acceso a las herramientas TIC y aquellas que no. En el texto de María del Carmen Agustín Lacruz y Manuel Clavero Galofré (2009); los autores realizan un proceso de análisis de la evolución histórica de la brecha digital desde la definición de su origen: "Una de las definiciones más sencillas y tempranas fue formulada por la Agencia Nacional de Telecomunicaciones e Información (NTIA).

Esta institución gubernamental estadounidense dependiente del Departamento de Comercio —equivalente a un Ministerio— enunció en 1995 el término digital divide para referirse a 'la desigualdad entre los que tienen un ordenador y los que no lo tienen'. El interés por el fenómeno estaba relacionado, en ese momento, con las repercusiones que podía suponer para la sociedad norteamericana en el corto y en el medio plazo el hecho constatable de que algunos de sus ciudadanos tuviesen ordenadores y pudiesen, por tanto, acceder a Internet y a sus servicios y contenidos, mientras que otros no podían disponer de ellos, ni acceder a sus beneficios" (Aguntín Lacruz y Clavero Galofré 2009) Esta primera mención hace referencia a la ya nombrada primera brecha o brecha de acceso, la cual se centra en estudiar la diferencia que existe entre los que tienen y no tienen una herramienta tecnológica. Lo anterior,

"explica que los primeros estudios y los indicadores más antiguos se ocupasen de analizar la

distribución de equipos y las opciones de acceso de los diferentes colectivos sociales (según su origen étnico y cultural: población blanca, de color, asiática o hispana; de género: mujeres y hombres; según su entorno de procedencia: rural o urbano; según los niveles educativos y de renta de los usuarios; según su edad, etc.). Esta misma orientación, netamente positivista y empírica, está presente en algunas de las más tempranas conceptualizaciones de Manuel Castells, quien en 2001 definió la divisoria digital como 'la disparidad entre los que tienen y los que no tienen Internet'" (Aguntín Lacruz y Clavero Galofré 2009).

Este aspecto se relaciona con la tendencia generalizada que existe en el fenómeno de realizar aproximaciones netamente cuantitativas y con un enfoque especialmente en las características demográficas, aspecto descrito como una aproximación simple al mismo. Sin embargo, "el énfasis en la disponibilidad de los equipos necesarios fue desplazándose poco a poco y de forma simultánea fue ampliándose la perspectiva desde la que se enfocaba el concepto" (Aguntín Lacruz y Clavero Galofré 2009).

La brecha digital, en este punto, tuvo un cambio en la descripción conceptual que tiene repercusiones teóricas y prácticas. Por otro lado, en el año 2011,"una institución dedicada a la cooperación y coordinación internacional de las políticas económicas y sociales de los estados miembros, como la Organización para la Cooperación y el Desarrollo Económico (OCDE), definió la brecha digital como 'el desfase o división entre individuos, hogares, áreas económicas y geográficas con diferentes niveles socioeconómicos con relación tanto a sus oportunidades de acceso a las tecnologías de la información y la comunicación, como al uso de Internet para una amplia variedad de actividades'" (Aguntín Lacruz y Clavero Galofré 2009).

Este cambio en la definición hace referencia la segunda brecha o brecha de uso, la cual se

centra en analizar la diferencia existente entre quienes tienen acceso a las TIC, pero no son usuarios.

"En la actualidad, el concepto de brecha, divisoria o fractura digital se ha extendido y popularizado ampliamente, adquiriendo una notable y constante presencia en los medios de comunicación. Quizá como consecuencia de ello, se relaciona cada vez más, no solo con la posibilidad de acceso a Internet, sino con la posesión de las habilidades y competencias necesarias para saber usarla. La brecha es considerada, desde esta perspectiva, en su dimensión política y educativa y, por ello, relacionada con una de las más antiguas desigualdades sociales: la relacionada con el acceso a la educación. Por ello, cada vez más insistentemente se vincula la brecha digital con las dinámicas sociales de inclusión/exclusión y de participación social y con la necesidad de poner en práctica políticas de alfabetización digital" (Aguntín Lacruz y Clavero Galofré 2009).

Este aspecto atiende a la tercera brecha o brecha de calidad de uso. En este recorrido por la evolución conceptual es posible identificar el dinamismo de la brecha digital, y de su conceptualización; cuando se habla de este término se hace referencia "a las desigualdades de acceso a Internet, el alcance de uso, el conocimiento de las estrategias de búsqueda, la calidad de las conexiones técnicas y el apoyo social, y la habilidad para evaluar la calidad de la información y los diversos usos" (DiMaggio, Hargittai, Russell Neuman, & Robinson, 2001).

Ahora bien, Chen, Yender y Jen (2011), realizan un análisis de las principales e importantes publicaciones en el área específica de la brecha digital, este estudio se realizó haciendo análisis de citación, co-citación (Science Citation Index – Social Sciences Citación Index) y análisis de redes sociales sobre más de 852 revistas indexadas.

Entre los aspectos a destacar de este estudio se menciona como la investigación en el área de la brecha digital se ha concentrado en la difusión de la tecnología, diferencia de prácticas culturales, difusión de la innovación y adopción de la tecnología (Cheng-Hua, Yender, & Jen-Hwa Kuo, 2011). Lo anterior indica y refuerza la evolución del fenómeno, es decir, en una primera observación, la brecha digital ha pasado de ser solo una visión binaria entre tener y no tener (Computador - Internet), a involucrar aspectos de carácter social y cultural.

Nuevamente se resalta la importancia de las herramientas TIC: estas permiten extender el poder de los especialistas urbanos (personas con influencia dentro de la comunidad), el mercado y el control de zonas de influencia regional, nacionales, internacionales e incluso globales cada vez más distantes.

Analítica de datos

Es necesario tomando como referencia el término mencionado resaltar la importancia de las plataformas y herramientas tecnológicas en el análisis de datos. Muchas de las investigaciones consultadas por no decir la gran mayoría, realizan aportes significativos a los procesos y afianzan los postulados, pero se recalca que muy pocos investigadores las utilizan por desconocimiento de su existencia o manejo y esto dificulta que se implementen de manera repetitiva, usando métodos tradicionales y desaprovechando el uso de las TIC que surgen para facilitar el desarrollo de nuestras tareas cotidianas.

El fortalecimiento de las capacidades de gestión de la información en las empresas de economías emergentes es fundamental para el desarrollo económico. La necesidad se hace más fuerte en un contexto en el que las tecnologías de la información permiten disponer de enormes cantidades de datos estructurados y no estructurados que las empresas deben aprovechar como

ventaja competitiva. Cusi y Bernal (2017).

Las organizaciones, empresas y personas han identificado la importancia de involucrar el análisis de datos o la mejor llamada ciencias de datos en el aumento de la productividad y toma de decisiones, análisis de dificultades e interpretación de la información de manera rápida y efectiva.

Big Data

Son grandes volúmenes de datos con la necesidad de ser capturados y ser analizados Joyanes (2014).

Tal como lo expresa el autor el Big Data poco a poco se ha ido convirtiendo en una de las inversiones más importantes para las organizaciones. ya que ayuda a obtener grandes beneficios para volverlas más competitivas. El gran reto está en transformar la big data en conocimiento, dado que los datos guardados no representan nada por sí solos, por lo que las organizaciones deben buscar herramientas que les permita generar valor a los datos que redunde en una buena toma de decisiones.

Plataformas de análisis de datos

Actualmente, casi la totalidad de las organizaciones cuenta con un sistema de información que soporta sus actividades diarias propias del sector de sus negocios, este sistema puede ser sencillo o robusto todo depende de las exigencias del mismo y los niveles de información gubernamental que se deba entregar; “con el tiempo las aplicaciones llegan a tener la historia de la organización y los datos almacenados en las bases de datos, pueden ser utilizados para argumentar la decisión que se quiera tomar ante cualquier aspecto para mejora en la empresa”. (Rosado Gomez, 2010).

Para tomar esas decisiones en busca de mejores oportunidades para las organizaciones se requiere del uso de una plataforma consultando en diferentes fuentes cuál era la herramienta o plataforma tecnológica más usada para gestionar la información o realizar un análisis de la misma, se determinó que la utilizada en mayor medida es Tableau software la cual pertenece a una empresa que se dedica al desarrollo de herramientas para la visualización interactiva de datos encaminada a la inteligencia de los negocios. Esta plataforma hace uso de un método estructurado de búsqueda en grandes bases de datos con un lenguaje descriptivo para la generación de gráficos, que le permiten administrar grandes cantidades de información en bases de datos relacionales, hojas de cálculo y bases de datos en la nube, generando así por medio de un tablero o interfaz dashboard un sin número de gráficos que facilitan la comprensión de la información y le brindan estructura para presentarlos en informes.

Capítulo III: Metodologías

En la actualidad, siempre se ha hecho uso de la informática como apoyo a todas las funciones operativas de las organizaciones o comunidades, con el fin de reducir los costos y aumentar la productividad, sin embargo se ha identificado que en el desarrollo de dichas tareas y a lo largo del tiempo se va generando y acumulando gran cantidad de información la cual corre el riesgo de perderse, dañarse o no suministrar la utilidad esperada a las comunidades u organizaciones por su difícil acceso, consulta y análisis. Por tal motivo el análisis de datos o la big data se ha convertido en un elemento importante, que facilita el aprovechamiento de los datos, la toma de decisiones, la identificación de oportunidades, el aumento de la eficiencia y la promoción de un uso efectivo de la tecnología predictiva.

Con la implementación de este proyecto que pretende investigar y evaluar diferentes lenguajes de programación y plataformas tecnológicas utilizadas en la ciencia de datos; se dará aplicabilidad a la información recolectada gracias al sistema de datos abiertos colombiano, "La Gran encuesta de TIC´S 2017 - COLOMBIA". Esto permite realizar un análisis profundo de la brecha digital de las comunidades encuestadas, lo que redunda en la toma de decisiones estratégicas adecuadas y eficaces a los entes gubernamentales o privados que estén interesados en mejorar el aprovechamiento de las tecnologías de la información y la comunicación, lo que repercute también en la calidad de vida de estas.

Disminuyendo así cada vez más, la gran brecha que existe en Colombia en cuanto al acceso a los dispositivos tecnológicos y la internet como herramientas que se vuelven indispensables en el desarrollo de un país que está exigiendo en el contexto actual más y nuevas oportunidades.

Para llevar a cabo la preparación de los datos y producir los resultados precisos se describirán

las diferentes técnicas y plataformas líderes en el mercado en la línea de analítica de datos que permiten procesar la información y obtener resultados de esta. De esta gran variedad de herramientas se seleccionará la que en requisitos funcionales cumpla con los requerimientos determinados por el grupo de investigación.

Búsqueda de lenguajes para la ciencia de datos y plataformas.

Lenguajes de programación

Inicialmente sólo existía un lenguaje de programación por cada modelo de computadora que había, las cuales eran en su momento muy pocas. Era el lenguaje “máquina”, representado en binario, y se programaba usando cables que se conectaban en varias configuraciones, o después usando tarjetas perforadas que se leían por medio de un aparato que las interpretaba y trasladaba el programa a la memoria para que lo ejecutara la computadora.

Pronto los profesionales en la línea identificaron que para hacer programas más complejos era necesario simplificar el proceso de programación. El primer paso fue la creación del lenguaje ensamblador, que consiste en una serie de mnemónicos que representan instrucciones del procesador, así como los operandos sobre los que operan. La traducción de ensamblador a lenguaje máquina era relativamente básica, y se podía hacer manual, pero pronto se buscó la manera de que la misma computadora lo hiciera.

Posteriormente, surgieron lenguajes de programación que permitían al profesional utilizar subrutinas o funciones que sintetizaban más el proceso. Por ejemplo, en lenguaje máquina o ensamblador no existe una función para calcular la raíz cuadrada, así que el programador tenía que escribir el código que lo hiciera cada vez que requería de esa función. Cuando surgieron los primeros lenguajes de bajo nivel o nivel medio, como Fortran, Cobol y C, se incorporó la función de raíz cuadrada para que el programador no tuviera que realizarla de nuevo. Esto agilizo en gran medida el proceso de programación, y facilitó el desarrollo de sistemas mucho más complejos y útiles que los anteriores. Lo que llevó a requerir un nuevo programa intermedio llamado compilador, que verifica y traduce el código “fuente” a código “máquina” que la computadora

puede ejecutar.

Con el paso del tiempo los lenguajes de programación evolucionaron y fueron especializándose en diferentes campos, dependiendo del tipo de problema que se quisiera resolver. Los problemas relacionados con administración de negocios, contabilidad, etc. se resolvían programando en Cobol (Common Business Oriented Language), por ejemplo. Los procesos orientados a la ciencia y las matemáticas se atacaban mediante programas en Fortran (Formula Translator); lo que tenía que ver más con lógica proposicional se abordaba mediante el uso de Prolog (Programación Lógica), etc.

En la actualidad, los lenguajes se han orientado y especializado todavía más, para enfocarse por ejemplo en los procesos distribuidos, programación de servidores (back end) y de scripts para navegadores de internet (front end), comunicaciones, cómputo paralelo, seguridad, acceso concurrente, bases de datos grandes, medianas y pequeñas, y un largo etcétera. Está en tendencia lo referente a aprendizaje máquina (machine learning) para lo cual se han aplicado lenguajes como Python, R, e incluso Java, que es quizás el lenguaje actual de aplicación general más usado. Hay lenguajes para niños, que de manera muy sencilla y visual les permiten incursionar y aprender a diseñar algoritmos que resuelvan problemas simples a su nivel.

Teniendo en cuenta que un lenguaje de programación es considerado como un conjunto de instrucciones y relaciones lógicas que guían al computador sobre la secuencia en que debe realizar una tarea permitiendo la comunicación entre la máquina y el usuario a continuación, se describen los lenguajes más usados en la ciencia de datos, big data o machine learning.

Lenguaje de programación R:

Es un entorno de software libre para gráficos y computación estadística. Se compila y se ejecuta en una amplia variedad de plataformas el cual utiliza (licencia GNU/GLP) y lenguaje de programación interpretado, es decir, ejecuta las instrucciones directamente, sin una previa compilación del programa a instrucciones en lenguaje máquina. El término entorno, en R, se refiere a un sistema totalmente planificado y coherente, en lugar de una acumulación de herramientas específicas e inflexibles, como suele ser el caso en otros softwares de análisis de datos. Funciona en plataformas UNIX y sistemas similares (incluidos FreeBSD y Linux), Windows y MacOS. Se enfoca en los siguientes tipos de análisis de información: Descriptiva, Predictiva, Prescriptiva.

Ventajas: Manejo y almacenamiento efectivo de los datos, un conjunto de operadores para la realización de cálculos con matrices, una gran colección de herramientas para el análisis de datos, utilidades gráficas para la visualización de datos, un lenguaje de programación bien desarrollado que incluye saltos condicionales, bucles, funciones recursivas, utilidades para la entrada y salida de datos, etc. Tiene un formato de documentación basado en LaTeX, que se utiliza para proporcionar documentación completa tanto en formato físico como digital.

Desventajas: las características y diferentes aplicaciones de R lo convierten en una herramienta básica para los analistas de datos.

Lenguaje de programación Python:

Es un lenguaje de programación interpretado cuya filosofía hace hincapié en la legibilidad de su código. Se trata de un lenguaje de programación multiparadigma, ya que soporta parcialmente la orientación a objetos, programación imperativa y, en menor medida, programación funcional.

Está orientado a Objetos, Multiplataforma, Interpretado, Ampliable (Puedes combinar fragmentos con otros lenguajes), Incrustarle (Scripting) y tiene un compendio considerable de librerías. Puede ejecutarse en diferentes sistemas operativos como Unix, Linux, macOS y Windows. Se enfoca en los siguientes tipos de análisis de información: Descriptiva, Predictiva, Prescriptiva. Hoy en día son muchos los desarrolladores que utilizan el lenguaje de programación Python para realizar tareas de diversa índole en campos tan distintos como el desarrollo web o el procesado de grandes volúmenes de datos. Esto está motivado por la gran cantidad de librerías, la facilidad de aprendizaje y la velocidad a la hora de crear prototipos frente a otros lenguajes.

Una de las ventajas principales de aprender Python es la posibilidad de crear un código con gran legibilidad, que ahorra tiempo y recursos, lo que facilita su comprensión e implementación. Es uno de los idiomas de programación más utilizados. Desde aplicaciones web hasta la inteligencia artificial, los usos de Python son infinitos: Python en la Inteligencia Artificial (AI), Python en Big Data, Python en Data Science, Python en Frameworks de Pruebas.

Ventajas: Estilo flexible, ordenado y limpio, open source, simplificado y rápido, estilo sano de programación, multiplataforma.

Desventajas: la principal es la velocidad. Al ser un lenguaje dinámico e interpretado, su performance no puede competir con lenguajes más tradicionales que son estáticos y compilados. Problemas con hosting, librerías incorporadas, lentitud al ejecutar múltiples hilos, no dispone de buena documentación, curva de aprendizaje, no tiene identificadores protegidos, simulaciones.

Lenguaje de programación Octava GNU:

Octave es un lenguaje de alto nivel, destinado principalmente a cálculos numéricos. Este lenguaje proporciona capacidades para la resolución numérica de problemas lineales y no lineales, y para realizar otras pruebas numéricas. También proporciona capacidades gráficas extensas para la visualización y manipulación de datos.

Octave se utiliza normalmente a través de su interfaz de línea de comandos interactiva, pero también se puede utilizar para escribir programas no interactivos. El lenguaje de programación Octave es bastante similar a Matlab, por lo que la mayoría de los programas son reutilizables en este lenguaje. Software gratuito, se ejecuta en GNU / Linux, macOS, BSD y Microsoft Windows compatible con muchos scripts de Matlab. Se enfoca en los siguientes tipos de análisis de información: Descriptiva, Predictiva, Prescriptiva.

Ventajas: GNU Octave dispone de potentes funciones matemáticas integradas (ecuaciones diferenciales, álgebra lineal, cálculo con matrices) y pueden ampliarse con la incorporación de librerías, como Scientific Library, Dionysus o Bc. También dispone de un paquete index con numerosas extensiones que enriquecen la funcionalidad de la herramienta. GNU Octave es un lenguaje de programación diseñado principalmente para resolver tareas de álgebra computacional. Es la alternativa más conocida a la solución comercial MATLAB, pero de carácter libre y gratuito. Además, no dispone de una interfaz gráfica.

Desventajas: Software de Media - alta - complejidad, octave es una aplicación informática que permite la programación en alto nivel para el cálculo numérico.

Lenguaje de programación Scala:

Scala se ejecuta en la Máquina Virtual Java (JVM). Es un lenguaje multiparadigmático, que permite tanto enfoques orientados a objetos como funcionales. El framework de computación de cluster Apache Spark está escrito en Scala. La implementación actual corre en la máquina virtual de Java y es compatible con las aplicaciones Java existentes.

Ventajas: Scala + Spark = Computación en clúster de alto rendimiento. Scala es un lenguaje ideal para quienes trabajan con conjuntos de datos de gran volumen. Multiparadigmático: los programadores de Scala pueden tener lo mejor de ambos mundos. Tanto la programación orientada a objetos como funcional. Scala se compila en el bytecode de Java y se ejecuta en una JVM. Esto permite la interoperabilidad con el lenguaje Java en sí, haciendo de Scala un lenguaje de propósito general muy poderoso, además de ser adecuado para la ciencia de datos.

Desventajas: Scala no es un lenguaje sencillo para comenzar a utilizar si está empezando. Lo mejor es descargar sbt y configurar un IDE como Eclipse o IntelliJ con un complemento específico de Scala. La sintaxis y el sistema de tipos se describen con frecuencia como complejos. Esto hace que la curva de aprendizaje sea pronunciada para aquellos que vienen de lenguajes dinámicos como Python."

Lenguaje de programación Julia:

Julia es un lenguaje de programación homoicónico, multiplataforma y multiparadigma de tipado dinámico de alto nivel y desempeño para la computación genérica, técnica y científica, con una sintaxis similar a la de otros entornos de computación similares. El nombre del lenguaje Julia fue una ocurrencia de los creadores. Es compatible con los sistemas operativos Linux, macOS,

Windows.

Ventajas: Julia es un lenguaje compilado JIT (‘just-in-time‘), que le permite ofrecer un buen rendimiento. También ofrece las capacidades de simplicidad, tipado dinámico y scripting de un lenguaje interpretado como Python. Julia fue diseñada específicamente para el análisis numérico. Pero también ofrece programación de propósitos generales. Legibilidad, Muchos usuarios del lenguaje mencionan esto como una ventaja clave.

Desventajas: Poca madurez al constituirse como nuevo idioma, algunos usuarios de Julia han experimentado inestabilidad al usar paquetes complementarios. Pero el núcleo del lenguaje es, al parecer, lo suficientemente estable para usar en producción. Los paquetes limitados son otra consecuencia de la juventud del lenguaje y de la pequeña comunidad de desarrollo. A diferencia de R y Python, Julia no tiene la posibilidad de disponer de paquetes (todavía).

Lenguaje de programación Matlab:

Es un lenguaje de computación numérica que se utiliza en el mundo académico y en la industria. Desarrollado y licenciado por MathWorks, una compañía establecida en 1984 para comercializar el software. Es compatible con plataformas SO (64-bit) Linux®, Apple Mac (64-bit), (64-bit) Windows®.

Ventajas: Diseñado para la computación numérica. MATLAB es adecuado para aplicaciones cuantitativas con requisitos matemáticos sofisticados, como procesamiento de señales, transformaciones Fourier, álgebra matricial y procesamiento de imágenes. Visualización de datos. MATLAB tiene incorporadas grandes capacidades de ploteado.

MATLAB se enseña con frecuencia como parte de cursos de pregrado en asignaturas cuantitativas como Física, Ingeniería y Matemáticas Aplicadas. Como consecuencia, es ampliamente utilizado en estos campos.

Desventajas: Licencia propietaria. Dependiendo del caso (uso académico, personal o empresarial) es posible que tengan que desembolsar una gran cantidad de dinero. Existen alternativas gratuitas disponibles como Octave. MATLAB no es una opción obvia para programación de propósito general.

Sobre estos lenguajes se han elaborado diferentes plataformas tecnológicas para un desarrollo de la ciencia de datos o minería de datos más intuitivo utilizando la inteligencia artificial y sacando el mayor potencial posible por medio de los algoritmos inteligentes en el análisis de grandes volúmenes de datos, por consiguiente, listo algunas de ellas:

- IBM Cloud.
- Google Collaboratory.
- Databricks.
- Machine Learning on AWS.
- ORANGE.
- Google Analytics.
- Tableau.
- Power BI.
- QLIK.

Relacionar las técnicas y plataformas de análisis de datos identificadas.

Definición minería de datos y técnicas

La minería de datos pretende reconocer e investigar diversas técnicas con el objetivo primordial de detectar patrones repetitivos que clarifiquen el comportamiento de los datos, que fueron reunidos a lo largo de varios años.

Explicando de una manera sencilla, el análisis de datos, big data o minería de datos se implementa con el objetivo de ofrecer una herramienta para analizar, comprender y administrar grandes repositorios de datos y en estos aplicar técnicas estadísticas, o de algoritmos de búsqueda cercanos a la inteligencia artificial y el machine learning.

Básicamente los datos son la materia prima base. El usuario o encargado de los mismos deberá atribuirles un significado especial para que se conviertan en información útil y dejen de estar inertes o muertos. Cuando se analiza detenidamente los datos y se elaboran o encuentran en ellos un modelo o patrón y este representa a su vez un valor agregado estos se convierten en conocimiento.

Para llevar a cabo una buena exploración de datos es importante tener en cuenta las siguientes etapas:

Delimitar los objetivos: Lo que desea el cliente o usuario de los datos que se van a analizar, asesorado por un especialista o ingeniero de datos.

Analizar los datos: Se relaciona con seleccionar, realizar limpieza, enriquecer, reducir y transformar la base de datos. Esta es una de las etapas más importantes, ya que si se realiza una excelente base de datos esto permitirá una administración de estos más sencilla.

Definir un modelo: Se inicia ejecutando un análisis estadístico de los datos, para posteriormente poder llevar a cabo una vista de gráfico de estos y así tener una aproximación inicial. Enmarcada en los objetivos planteados y la tarea que se desea llevar a cabo y para la cual pueden en este momento utilizarse algoritmos desarrollados en diferentes áreas de la inteligencia artificial.

Estudiar los resultados: Permite corroborar si los resultados encontrados son consistentes o congruentes en comparación con los resultados obtenidos del análisis estadístico y de visualización gráfica. El usuario debe determinar si los datos son innovadores y si proporcionan un nuevo conocimiento que va a redundar en la toma de buenas decisiones.

Técnicas de la minería de datos o análisis de datos:

Las técnicas de la minería de datos derivan de la estadística y la inteligencia artificial; utilizando algoritmos que permiten extraer la información sobre grandes volúmenes de datos o la bien llamada Big Data para obtener resultados más precisos.

Algunas de las técnicas de minería de datos o análisis de datos más utilizadas o representativas son:

Las redes neuronales utilizan ejemplos de aprendizajes y de procesamientos automatizados los cuales se inspiran en la manera cómo funciona un sistema nervioso animal. Este sistema facilita interconectar las neuronas de una red las cuales colaboran en la producción de estímulos de salida, algunos de los ejemplos de redes neuronales que vale la pena mencionar son:

El perceptrón (tipo de red neuronal artificial).

El perceptrón multicapas.

Los mapas autoorganizados.

La regresión lineal: Es una de las técnicas más usadas para determinar o formar relaciones entre los datos. Conforma entonces un sistema ágil y eficaz, pero con algunas deficiencias en

algunos espacios multidisciplinarios donde se necesiten asociar más de 2 variables a la vez.

Los árboles de decisión son un modelo que se utiliza en el campo de la inteligencia artificial aplicable a las bases de datos en donde intervienen las construcciones lógicas. Es un sistema que se asemeja a la predicción, la cual se basa en reglas o parámetros que son utilizados en la representación de serie condiciones que se relacionan de una manera repetitiva en la resolución de problemas.

Los modelos estadísticos, se utilizan como una herramienta de expresión simbólica en forma de igualdad y se emplean en la elaboración de diseños iniciales o experimentales y en la regresión, con el fin de reconocer o identificar los factores que intervienen o cambian la variable de respuesta.

El agrupamiento, se fundamenta en asociar una serie de vectores según unos criterios determinados que por lo general son a distancia. Se trata de darle orden a los vectores iniciales o de entrada de tal manera que estén cercanos a los que compartan con ellos características comunes.

Dependiendo de los objetivos que se delimiten en la realización del análisis los algoritmos se pueden clasificar como algoritmos supervisados, los cuales predicen datos desconocidos inicialmente a partir de otros datos que se tienen en conocimiento previo. Por otro lado, los algoritmos no supervisados, son los que descubren patrones y tendencias que se presentan en los datos.

Figura 1.

Interpretación grafica de las técnicas aplicadas a la minería de datos.

Nota*:* El grafico menciona el origen de las técnicas de minería de datos, la definición y uso.
Fuente: El autor.

Plataformas o herramientas que facilitan el análisis de datos:

Una plataforma de administración de datos, es un sistema centralizado para recolectar y analizar grandes volúmenes de datos provenientes de diferentes fuentes, crea un entorno combinado de desarrollo y resultados, que proporciona a los usuarios datos coherentes, concisos y acertados. En su forma más elemental, una plataforma de gestión de datos podría ser un sistema de gestión de bases de datos NoSQL o SQL que importa datos de muchos sistemas y permite ver los datos de forma coherente. Una DMP de alta gama podría combinar tecnologías de administración de datos y herramientas de análisis de datos en un solo aplicativo de software

con una consola intuitiva y sencilla de navegar, el usuario no debería tener gran conocimiento en programación de software para utilizarlo.

Una función decisiva de una plataforma de gestión de datos es reunir datos estructurados y no estructurados de una serie de fuentes internas y externas, y luego integrar y guardar esos datos. Estas plataformas también analizan y regulan los datos para proporcionar una visión de las partes de la organización impulsadas por ellos.

En la mayoría de los casos los datos que se agregan en las plataformas DMP, son datos que se originan en aplicativos, sistemas de información, páginas web y productos propios de una entidad, así como datos de externos y algunos asociados. Además, los DMP utilizan datos de terceros para completar las falencias en los datos de la empresa o colaboradores. Una buena gestión de datos le va a facilitar a la organización o empresa potencializar e interactuar con otros en una buena toma de decisiones.

Los datos que se adicionan en una plataforma DMP pueden ser datos propios derivados u originados de sistemas, aplicaciones, sitios web y productos propios de una entidad u organización, así como datos de terceros y otros asociados. Además, los DMP utilizan datos de terceros para completar las falencias en los datos de la empresa o colaboradores. Una buena gestión de datos le va a facilitar a la organización o empresa potencializar e interactuar con otros en una buena toma de decisiones.

En la evolución se encuentra que ya las organizaciones no se preocupan solamente por la protección de datos, sino que se volcó a la gestión de estos, los proveedores de salvaguarda de datos comenzaron a ampliar sus ofertas con nuevas utilidades de administración de datos más amplias. La empresa estadounidense Veritas diseñó un plan para construir una plataforma de administración de datos que combine su Plataforma de Resiliencia de Veritas, la cual orquesta la

recuperación de equipos virtuales en nubes híbridas de varios proveedores; junto con el mapa de información, que adiciona datos y los revela visualmente.

Otras empresas como Commvault incursionando en la administración de datos han decidido crear motores de búsqueda de protección de datos, los cuales cuentan con un sistema de registro integrado, estas compañías planean abrir sus plataformas a desarrolladores de aplicaciones para consolidar un ecosistema de capacidades de administración de datos en su software de respaldo, estas también le apuestan fuertemente a la administración de datos en la nube. Otras analizan el almacenamiento de archivos y usan la automatización soportada en políticas para mover los datos de archivos inactivos a objetivos de almacenamiento basados en objetos que se ejecutan localmente o en la nube.

Algunas de las características que dichas empresas u organizaciones plantean desarrollar son; plataforma de administración de datos en la nube a gran escala para informes, copia de seguridad, replicación, recuperación, análisis, búsqueda, archivo y administración de datos de copia en una única plataforma.

Las plataformas de administración de datos operan con la información de clientes y diversas fuentes y luego la analizan, organizan y segmentan para determinar factores relevantes como ubicación, ingresos, comportamiento de navegación, entre otros, según las necesidades de los usuarios que utilizan esta información como insumo.

En efecto, una plataforma de gestión de datos requiere una gran cantidad de datos lo que la hace procesable, proporcionando a las empresas una comprensión más profunda de sus datos para brindar orientación en la mejor toma de decisiones, en especial las referentes al bienestar de sus clientes o usuarios, la adquisición de nuevas tecnologías o la optimización de sus procesos internos.

El avance de las TIC o las nuevas tecnologías están dinamizando el panorama de la plataforma de administración de datos. Las herramientas o aplicaciones de internet de las cosas (IoT) y los sensores de bajo costo están generando grandes volúmenes de datos a las plataformas de administración de los DMP. La tecnología de aprendizaje automático o inteligencia artificial se está incorporando a los sistemas de gestión de datos y se utiliza para mover los datos de las plataformas de gestión a su destino más rápido que nunca. La arquitectura de "datos rápidos" permitirá procesar velozmente extensos volúmenes de datos, y al mismo tiempo la administración de datos está evolucionando para permitir que las empresas analicen la información al instante en tiempo real y en el momento que la necesiten. Se están diseñando nuevas funcionalidades de transmisión de datos para administrar grandes cantidades de datos complejos, diversos y, a menudo, no estructurados. Estas plataformas de transmisión funcionan con datos en movimiento, evaluándose en el instante en que llegan. Si bien estos sistemas aún no están financieramente al alcance de la mayoría de las empresas o usuarios, señalan el futuro de la tecnología de la plataforma de gestión de datos y muchas de estas por la aceptación o necesidad masiva han impulsado la capacitación y demanda en el mercado laboral, algunas de las plataformas que permiten el manejo e interpretación de los datos inactivos para convertirlos en información valiosa aprovechando el ecosistema de datos abiertos en Colombia son:

Tabla 1.

Plataforma de análisis de datos - IBM Cloud.

Nombre	IBM Cloud
Enlace de acceso	https://cloud.ibm.com
Licencia	Libre para estudiantes

Concepto /Definición	Es una plataforma que abarca diferentes tipos de servicios de Cloud creando soluciones para empresas que buscan transformar y modernizar su flujo de trabajo. Aparte de una alta seguridad y de ofrecer todo lo necesario para el trabajo de Devs y TI, ofrece también recursos de datos avanzados e IA.
Compatibilidad S.O.	IBM® Db2 para LUW (Linux, UNIX y Windows) versión 10.5 o posteriores Sistemas operativos admitidos por IBM Cloud con XenServer, CentOS 6. x, 7. x, Cloud Linux 6. x, CoreOS Stable, RedHat Enterprise Linux 6. x, 7. x, Debian 7. x a 9. x, Ubuntu 14.04 LTS a 16.04 LTS, Microsoft Windows Server 2012, 2012 R2, 2016, 2019.
Tipo de análisis de datos (Descriptiva, Predictiva, Prescriptiva)	Predictiva, Prescriptiva
Beneficios para el desarrollo propio	Nube Distribuida, Servicios financieros, VMware, OpenShift, Telecomunicaciones, SAP on Cloud, Power Virtual Server on Cloud, Object Storage
Dificultad de uso	Los problemas conocidos y las limitaciones incluyen no poder restringir el acceso a algunos productos del catálogo de IBM Cloud®, los límites máximos para crear recursos de IBM Cloud Identity and Access Management (IAM) y no poder suprimir usuarios de cuentas que tienen demasiadas organizaciones de Cloud Foundry.

Nota: Características de plataforma de análisis de datos. Fuente: El autor.

Tabla 2.

Plataforma de análisis de datos - Google Colaboratory.

Nombre	Google Colaboratory
Enlace de acceso	https://colab.research.google.com/notebooks/intro.ipynb
Licencia	Libre
Concepto /Definición	Es un servicio en la nube, ofrecido por Google de forma gratuita. Se basa en el entorno Jupyter Notebook y está destinado a la capacitación e investigación en aprendizaje automático. Esta plataforma permite entrenar modelos de machine learning de manera directa en la nube y de forma gratuita. Facilita la operación con grandes conjuntos de datos, creación de modelos complejos e incluso compartir nuestro trabajo sin problemas con otros.
Compatibilidad S.O.	Sistema Operativo: macOS Mojave 10.14.2, Microsoft Windows, Mac OS X y Linux.
Tipo de análisis de datos (Descriptiva, Predictiva, Prescriptiva)	Predictiva

Beneficios para el desarrollo propio	Permite ejecutar y programar en Python en tu navegador con las siguientes ventajas: No requiere configuración, da acceso gratuito a GPUs, permite compartir contenido fácilmente, agiliza tu trabajo, ya seas estudiante científico de datos o investigador de IA. Colab es una herramienta muy utilizada en la comunidad de aprendizaje automático que facilita aprovechar toda la potencia de las bibliotecas más populares de Python para analizar y visualizar datos. La celda de código. La principal ventaja que ofrece esta herramienta es que libera a nuestra máquina de tener que llevar a cabo un trabajo demasiado costoso tanto en tiempo como en potencia o incluso permite realizar ese trabajo si nuestra máquina no cuenta con recursos suficientemente potentes. Y todo de forma gratuita. Otro de los beneficios que tiene lo indica el propio nombre, «Colaboratory», es decir, colaborativo, permite realizar tareas en la nube y compartir los cuadernos si necesita trabajar en equipo.
Dificultad de uso	La desventaja es que el hardware que se estaría utilizando es el de nuestra máquina local y se perdería la capacidad de ejecutar el código.

Nota: Características de plataforma de análisis de datos. Fuente: El autor.

Tabla 3.

Plataforma de análisis de datos – Databricks.

Nombre	Databricks
Enlace de acceso	https://databricks.com/
Licencia	Libre
Concepto /Definición	Es un cluster Open Source de computación distribuida pensado para ejecutar consultas de análisis de datos y algoritmos de Machine Learning.
Compatibilidad S.O.	"Funcional en todas las plataformas. Servicio cloud en Microsoft Azure y Amazon Web Services (AWS)"
Tipo de análisis de datos (Descriptiva, Predictiva, Prescriptiva)	Predictiva.
Beneficios para el desarrollo propio	Consultas de análisis de datos y algoritmos de Machine Learning
Dificultad de uso	Alto costo de operación o funcionamiento. La evolución constante de la tecnología dificulta su uso por el cambio constante de versiones y adición de funciones.

Nota: Características de plataforma de análisis de datos. Fuente: El autor.

Tabla 4.

Plataforma de análisis de datos - Machine Learning on AWS.

Nombre	Machine Learning on AWS
Enlace de acceso	https://aws.amazon.com/machine-learning/
Licencia	Libre para pruebas
Concepto /Definición	servicio que le ayuda a crear potentes modelos de aprendizaje automático.
Compatibilidad S.O.	Funcional en todas las plataformas
Tipo de análisis de datos (Descriptiva, Predictiva, Prescriptiva)	Descriptiva, Predictiva, Prescriptiva.
Beneficios para el desarrollo propio	Crear modelos de aprendizaje automático y generar predicciones, lo que facilita el desarrollo de aplicaciones inteligentes
Dificultad de uso	Media, se requieren conocimientos específicos para su implementación poco intuitiva.

Nota: Características de plataforma de análisis de datos. Fuente: El autor.

Tabla 5.

Plataforma de análisis de datos – ORANGE.

Nombre	ORANGE
Enlace de acceso	https://orangedatamining.com/
Licencia	Libre - Código abierto
Concepto /Definición	Entre las herramientas de Data Mining se puede destacar Orange, desarrollado por el Laboratorio de Bioinformática de la Facultad de Informática y Ciencias de la Información de la Universidad de Ljubljana, Eslovenia. Es un software libre de aprendizaje automático y data mining. Sus características principales residen en sus funcionalidades como la programación visual front-end para explorar datos y la visualización de resultados. Aunque también puede usarse como una biblioteca Python.
Compatibilidad S.O.	Microsoft Windows, Mac OS X y Linux.
Tipo de análisis de datos (Descriptiva, Predictiva, Prescriptiva)	Realizar minería de datos y análisis predictivo

Beneficios para el desarrollo propio	Sus características principales residen en sus funcionalidades como la programación visual front-end para explorar datos y la visualización de resultados. Aunque también puede usarse como una biblioteca Python. Favorece a los usuarios en la creación de sus propios flujos de trabajo interactivos con el objetivo de analizar y visualizar los datos con mayor amplitud. Permite rediseñar y adaptar la herramienta a las necesidades del usuario y/o de la empresa. La visualización de la información puede realizarse en distintos formatos, diagramas de dispersión, gráficos de barras, árboles o redes y mapas de color, lo cual permite mostrar con mayor claridad los resultados para interpretar de mejor forma la información.
Dificultad de uso	Baja - Media, aplicativo con interfaz gráfica sencilla.

Nota: Características de plataforma de análisis de datos. Fuente: El autor.

Tabla 6.

Plataforma de análisis de datos - Google Analytics.

Nombre	Google Analytics
Enlace de acceso	https://analytics.google.com/analytics/web/provision/#/provision
Licencia	Libre/ Licenciada
Concepto /Definición	Google Analytics es una herramienta y plataforma online desarrollada por Google para medir y analizar lo que ocurre en un sitio Web o en una aplicación móvil. Usada por millones de empresas y webmasters en todo el mundo, dispone de una versión sin coste con una funcionalidad más que suficiente para la mayoría de los negocios, y de otra versión comercial, más potente (Google Analytics 360) para grandes organizaciones.
Compatibilidad S.O.	Analytics se puede utilizar con las dos versiones principales más recientes de Chrome, Firefox, Edge y Safari (progresivamente). / Microsoft Windows, Mac OS X y Linux.
Tipo de análisis de datos (Descriptiva, Predictiva, Prescriptiva)	Descriptiva, Predictiva, Prescriptiva.

Beneficios para el desarrollo propio	Es gratuito: Sólo necesita tener una cuenta en Google (Gmail), es fácil de instalar: Sólo es necesario crear una cuenta en Google Analytics para usarlo, permite medir todo tipo de campañas de captación de tráfico: Ayudando a tener una visión del tráfico en el sitio web, dando la información que necesita la empresa para usar una estrategia de acuerdo con sus intereses. Los informes son personalizables y programables, permitiendo así un óptimo servicio, crea informes incluyendo análisis en tiempo real (con gráficas): permitiendo observar la cantidad de tráfico que hay en el sitio web y haciendo este monitoreo más cómodo para la empresa a través de gráficas.
Dificultad de uso	Bajo muy intuitiva.

Nota: Características de plataforma de análisis de datos. Fuente: El autor.

Tabla 7.

Plataforma de análisis de datos – Tableau.

Nombre	Tableau
Enlace de acceso	https://www.tableau.com/trial/tableau-software?utm_campaign_id=2017049&utm_campaign=Prospecting-CORE-ALL-ALL-ALL-ALL&utm_medium=Paid+Search&utm_source=Google+Search&utm_language=EN&utm_country=RoLAC&kw=tableau&adgroup=CTX-Brand-Priority-Core--EN-E&adused=451455447751&matchtype=e&placement=&gclid=Cj0KCQjw2NyFBhDoARIsAMtHtZ79uU3vByjOSnV9Pyz9YKU2x1R3fU1IgDO2J89NY7ELlg5dXz7QW7gaAtw_EALw_wcB&gclsrc=aw.ds
Licencia	Permite licencia para docentes y estudiantes de 1 año libre.
Concepto /Definición	Tableau es una herramienta de Inteligencia de Negocios que permite analizar, visualizar y compartir grandes volúmenes de información en forma rápida, flexible y amigable. El usuario tan solo tiene que arrastrar los campos de su interés para lograr el cruce de información y obtener una atractiva visualización. Esta otra herramienta BI que sirve para la visualización interactiva de los datos, con los que los usuarios pueden interactuar de varias maneras: comparando datos, filtrándolos o creando una conexión entre unas variables y otras.
Compatibilidad S.O.	Compatible con entornos web y dispositivos Android, Windows, Linux, iPhone/iPad y Mac.
Tipo de análisis de datos (Descriptiva, Predictiva, Prescriptiva)	Predictivo

Beneficios para el desarrollo propio	Las ventajas que ofrece Tableau son: Los usuarios pueden profundizar y explorar datos sin ninguna experiencia de programación, tienen la capacidad de conectarse a una multitud de fuentes de datos, pueden utilizar la API de esta herramienta para la extracción sistemática de datos. Si un usuario tiene experiencia en programación, Tableau puede funcionar como front-end de herramientas que permite realizar inmersiones estadísticas profundas y análisis avanzados.
Dificultad de uso	La principal desventaja de Tableau al momento es que solo está verificado sobre sistemas operativos Windows y Mac. No soporta conexiones Mondrian y no tiene herramientas ETL.

Nota: Características de plataforma de análisis de datos. Fuente: El autor.

Tabla 8.

Plataforma de análisis de datos - Power BI.

Nombre	Power BI
Enlace de acceso	https://powerbi.microsoft.com/es-es/
Licencia	Propietario – los precios varían dependiendo del caso.
Concepto /Definición	Es una solución de análisis empresarial basado en la nube, que permite unir diferentes fuentes de datos, analizarlos y presentar un análisis de estos a través de informes y paneles. Con Power BI se tiene de manera fácil acceso a datos dentro y fuera de la organización casi en cualquier dispositivo. Herramienta de Microsoft con servicio en la nube con fácil implementación; que permite subir, compartir y tener acceso a informes desde cualquier dispositivo, ya sea un ordenador, una Tablet o un smartphone.
Compatibilidad S.O.	Funciona de manera compatible con los principales sistemas operativos: Windows, iOS y Android.
Tipo de análisis de datos (Descriptiva, Predictiva, Prescriptiva)	Descriptiva, Predictiva, Prescriptiva.

Beneficios para el desarrollo propio	Algunas de las ventajas de utilizar Power BI son: Incremento de la eficiencia en las compañías para extraer informes de manera autónoma cuando lo requieran sin tener un conocimiento específico de la misma. Power BI presenta una herramienta Quick Insights que determina las correlaciones y patrones dentro de sus datos, produciendo gráficos y gráficos personalizados. Integración del análisis avanzado a través de scripts y objetos visuales de R, Microsoft Azure Machine Learning y Azure Stream Analytics.
Dificultad de uso	Entre sus limitaciones destaca que no se pueden compartir informes con otros usuarios, no se permite el análisis en Excel dentro de la propia herramienta o no permite suscripciones por email.

Nota: Características de plataforma de análisis de datos. Fuente: El autor.

Tabla 9.

Plataforma de análisis de datos – QLIK.

Nombre	QLIK
Enlace de acceso	https://www.qlik.com/es-es/
Licencia	Propietario – los precios varían dependiendo del caso.
Concepto /Definición	Qlik Sense Desktop es una aplicación de Windows que permite a los usuarios crear visualizaciones, gráficos, cuadros de mando interactivos y aplicaciones de analítica para uso local y sin conexión. Qlick es una plataforma enfocada al análisis visual de datos y aplicaciones interactivas que tiene por objetivo mejorar el proceso de acceso a los datos de cara al usuario. Como, por ejemplo, acceder a ciertas visualizaciones 'limpias' y fáciles de comprender, diseños de gráficos llamativos, entre otros.
Compatibilidad S.O.	Herramienta compatible con entornos web y dispositivos Windows, iPhone/iPad y Mac.
Tipo de análisis de datos (Descriptiva, Predictiva, Prescriptiva)	Predictivo y descriptivo

Beneficios para el desarrollo propio	Algunas de las ventajas de utilizar Qlik para analizar datos son: Los conocimientos de datos se pueden generar rápidamente a partir de un usuario competente, ya que las capacidades de análisis de datos están limitadas sólo por sus capacidades de creación de scripts. El motor asociativo subyacente realiza uniones naturales en tiempo real en función de las selecciones del usuario, destaca las relaciones entre las entidades para el usuario.
Dificultad de uso	El usuario final está ciertamente limitado a la hora de generar nuevas visualizaciones. Resistencia al cambio por parte de los usuarios. Es recomendable proceder a una implantación progresiva que sea asimilada con facilidad por los empleados de la empresa y el departamento de IT.

Nota: Características de plataforma de análisis de datos. Fuente: El autor.

Identificar las necesidades requeridas para la plataforma a desarrollar.

Luego de la revisión de las diferentes plataformas, a continuación, se enuncian las necesidades de la plataforma a implementar para el análisis de datos sobre brecha digital.

Como un punto fundamental se debe mencionar que se espera que la plataforma que se utilice o se implemente permita hacer uso del ecosistema de datos abiertos existente en el país, esto quiere decir, que la plataforma, como una de las principales características, debe permitir la carga de datos sobre brecha digital que se encuentren almacenados en la plataforma datos.gov.co de Colombia.

De manera adicional, la plataforma debe permitir a los usuarios navegar por cada una de las gráficas sobre brecha digital generadas y hacer uso de estas, es decir, el usuario debe poder descargar estas graficas como imagen y, de ser posible, interactuar con las gráficas.

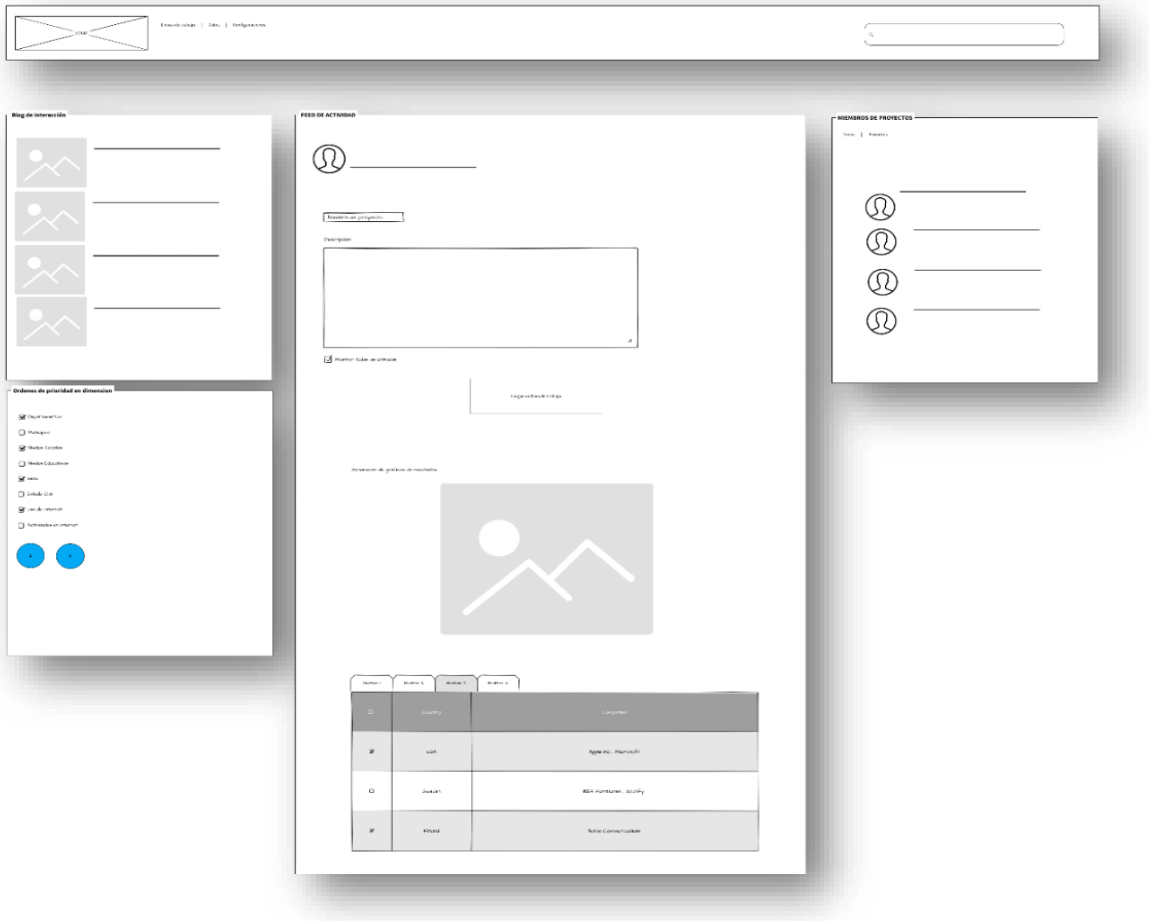
Definir requisitos funcionales y no funcionales de la plataforma.

De manera adicional, la plataforma debe permitir a los usuarios navegar por cada una de las gráficas sobre brecha digital generadas y hacer uso de estas, es decir, el usuario debe poder descargar estas graficas como imagen y, de ser posible, interactuar con las gráficas.

Como parte del proceso de desarrollo e implementación de la plataforma, se relaciona a continuación un acercamiento a la idea inicial planteada de diseño de esta; en esta primera propuesta se identificó que sería posible establecer un inicio de sesión para que cada uno de los usuarios pudiera guardar allí sus propios proyectos.

Figura 2.

Interfaz de plataforma usuarios

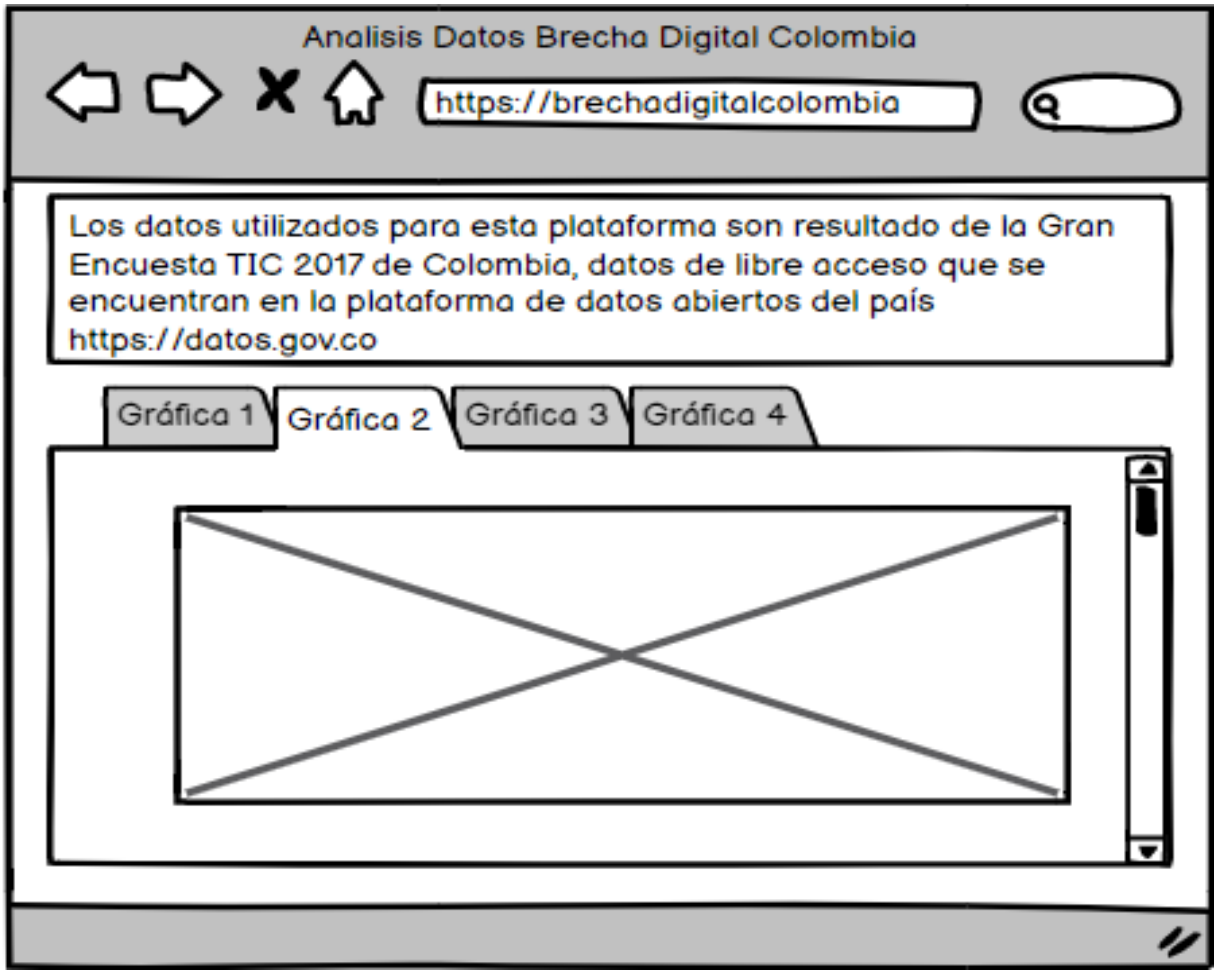


Nota: El grafico muestra un boceto de la posible interfaz de acceso a la plataforma para la interacción de los usuarios. Fuente: El autor.

Sin embargo, al revisar el alcance del proyecto de grado y las diferentes plataformas identificadas y caracterizadas, se definió que era posible, para este proyecto, la implementación de la solución de análisis de datos sobre brecha digital sobre una de las plataformas existentes, con base en estas características, a continuación, se muestra el prototipo de pantalla definido:

Figura 3.

Prototipo definido de pantalla.



Nota: El grafico muestra un boceto de la posible interfaz de acceso a la plataforma para la interacción de los usuarios. Fuente: El autor.

Organizar los datos a utilizar para el análisis.

No todas las herramientas cuentan con técnicas para organizar los datos, son los investigadores los que ordenan las bases de datos de acuerdo con los intereses de la investigación; existen aplicaciones o plataformas que ayudan a organizar la información de tal manera que facilite el análisis de esta. Estas aplicaciones o plataformas permiten por medio de funciones de emparejamiento, unión, similitudes, filtro u operaciones matemáticas unir diferentes bases de datos de distintas fuentes de información en una sola, recibiendo el nombre de bodega de datos. Las bodegas de datos permiten a las empresas o usuarios consolidar la información, administrarla y posteriormente interpretarla o analizarla para la toma de decisiones; en beneficio del negocio o desarrollo del tema del cual se quiere obtener información aplicando la estadística con el fin de darle significado a los datos.

Con la finalidad de identificar herramientas o aplicaciones que apoyan la tarea compleja de organizar y clasificar la información que será objetivo del análisis; se realizaron consultas sobre diferentes plataformas o aplicaciones que permiten preparar los datos y consolidarlos para ser analizados encontrando que la empresa desarrolladora de software Tableau además de tener aplicaciones que permiten analizar la data como Tableau Public o desktop; cuentan con desarrollos que facilitan de manera visual, directa e inteligente la limpieza de los datos y organización para realizar un análisis rápido y confiable como Tableau Prep, librando al administrador de la información de la tarea tediosa y que demanda demasiado tiempo de organizar los datos de manera manual.

Tableau Pre es una herramienta diseñada para la preparación de los datos, en proyectos de análisis de información el tema de la preparación de los datos juega un rol importantísimo, se estima que el 80% del tiempo es dedicado a la preparación de los datos mientras que solo el 20%

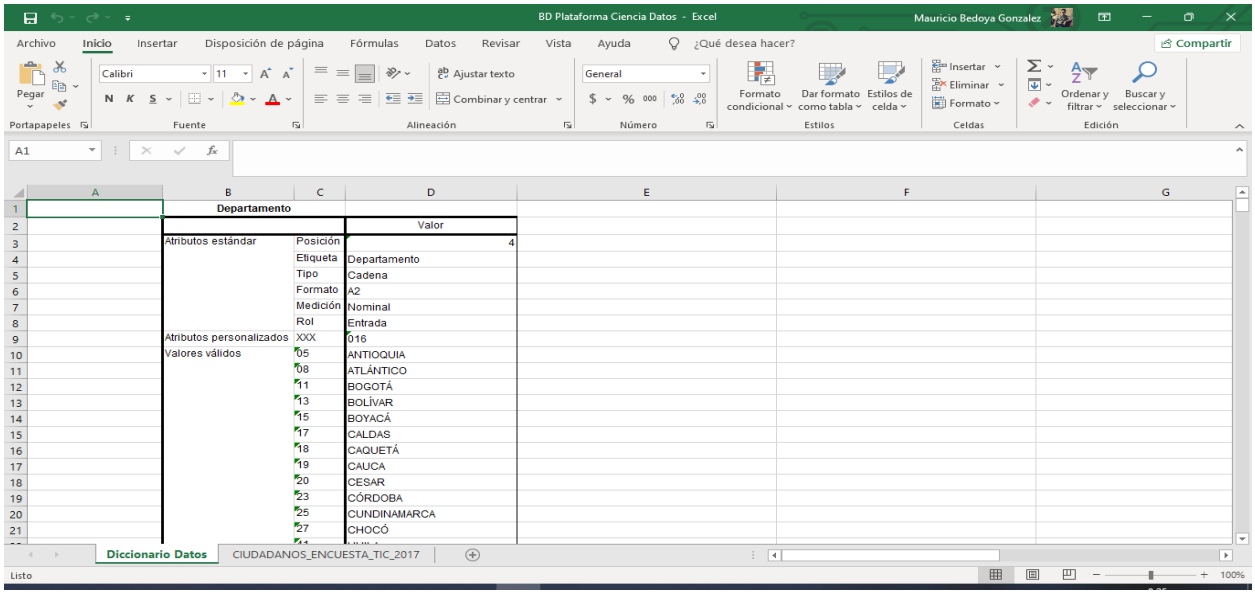
restante es dedicado al análisis de estos. Lo ideal sería poder revertir esta estadística o tiempos dedicándole más tiempo al análisis de los datos que a prepararlos, es por ello que se implementan herramientas como Tableau Prep la cual permite minimizar el proceso de la preparación de los datos a procesar gracias a su filosofía de trabajar en forma de flujo de trabajo; en Tableau Pre todas las tareas que se realicen se ejecutan de una manera visual y directa también es importante destacar la integración que tiene la herramienta de poder escribir en algunas bases de datos tradicionales como SQL server al igual que bases de datos alojadas en la nube.

Utilizando como insumo la base de datos abierta generada por el ministerio de las tecnologías de la información y la comunicación en Colombia para el análisis de la brecha digital se exportó un archivo de excel en formato CSV, el cual consolida la muestra de las respuestas del formulario aplicado que permiten analizar a profundidad la temática antes mencionada.

Este archivo de valores separados por comas será nuestra fuente de datos el cual en cada columna tiene los ítems de las preguntas aplicadas en la encuesta y en las filas las respuestas obtenidas; dentro de las herramientas de Tableau Prep utilizaremos la funcionalidad de transponer filas y columnas para organizar nuestros datos, además la consolidación de tablas para unificar valores con similitudes e interpretar los mismos utilizando el libro diccionario de datos. Todo esto nos permite realizar un join entre las tablas de nuestra base de datos. En las gráficas que se relacionan a continuación se puede observar la base de datos original en formato CSV y el orden de sus filas:

Figura 4.

Diccionario de datos para interpretación:

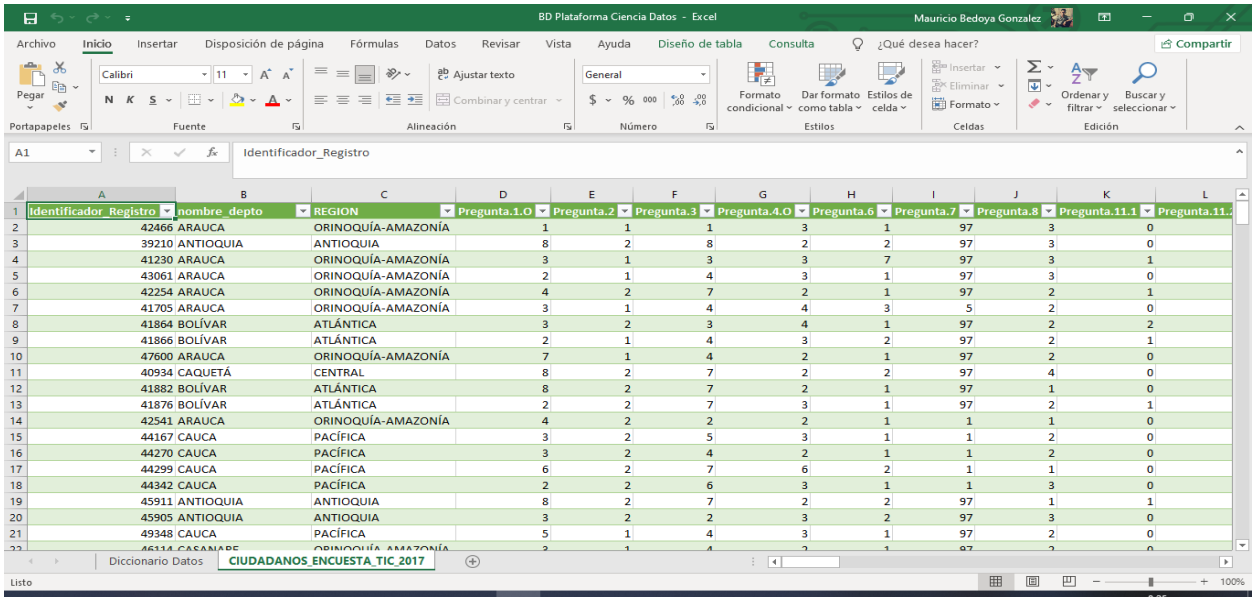


1	Departamento		
2			Valor
3	Atributos estándar	Posición	4
4		Etiqueta	Departamento
5		Tipo	Cadena
6		Formato	A2
7		Medición	Nominal
8		Rol	Entrada
9	Atributos personalizados	XXX	016
10	Valores válidos	05	ANTIOQUIA
11		08	ATLÁNTICO
12		11	BOGOTÁ
13		13	BOLÍVAR
14		15	BOYACÁ
15		17	CALDAS
16		18	CAQUETÁ
17		19	CAUCA
18		20	CESAR
19		23	CÓRDOBA
20		25	CUNDINAMARCA
21		27	CHOCÓ

Nota: Base de datos original libro para realizar interpretación y tabulación de la informacion. Fuente: El autor.

Figura 5.

Registros recolectados:



Identificador_Registro	nombre_depto	REGION	Pregunta.1.O	Pregunta.2	Pregunta.3	Pregunta.4.O	Pregunta.6	Pregunta.7	Pregunta.8	Pregunta.11.1
42466	ARAUCA	ORINOQUÍA-AMAZONÍA	1	1	1	3	1	97	3	0
39210	ANTIOQUIA	ANTIOQUIA	8	2	8	2	2	97	3	0
41230	ARAUCA	ORINOQUÍA-AMAZONÍA	3	1	3	3	7	97	3	1
43061	ARAUCA	ORINOQUÍA-AMAZONÍA	2	1	4	3	1	97	3	0
42254	ARAUCA	ORINOQUÍA-AMAZONÍA	4	2	7	2	1	97	2	1
41705	ARAUCA	ORINOQUÍA-AMAZONÍA	3	1	4	4	3	5	2	0
41864	BOLÍVAR	ATLÁNTICA	3	2	3	4	1	97	2	2
41866	BOLÍVAR	ATLÁNTICA	2	1	4	3	2	97	2	1
47600	ARAUCA	ORINOQUÍA-AMAZONÍA	7	1	4	2	1	97	2	0
40934	CAQUETÁ	CENTRAL	8	2	7	2	2	97	4	0
41882	BOLÍVAR	ATLÁNTICA	8	2	7	2	1	97	1	0
41876	BOLÍVAR	ATLÁNTICA	2	2	7	3	1	97	2	1
42541	ARAUCA	ORINOQUÍA-AMAZONÍA	4	2	2	2	1	1	1	0
44167	CAUCA	PACÍFICA	3	2	5	3	1	1	2	0
44270	CAUCA	PACÍFICA	3	2	4	2	1	1	2	0
44299	CAUCA	PACÍFICA	6	2	7	6	2	1	1	0
44342	CAUCA	PACÍFICA	2	2	6	3	1	1	3	0
45911	ANTIOQUIA	ANTIOQUIA	8	2	7	2	2	97	1	1
45905	ANTIOQUIA	ANTIOQUIA	3	2	2	3	2	97	3	0
49348	CAUCA	PACÍFICA	5	1	4	3	1	97	2	0

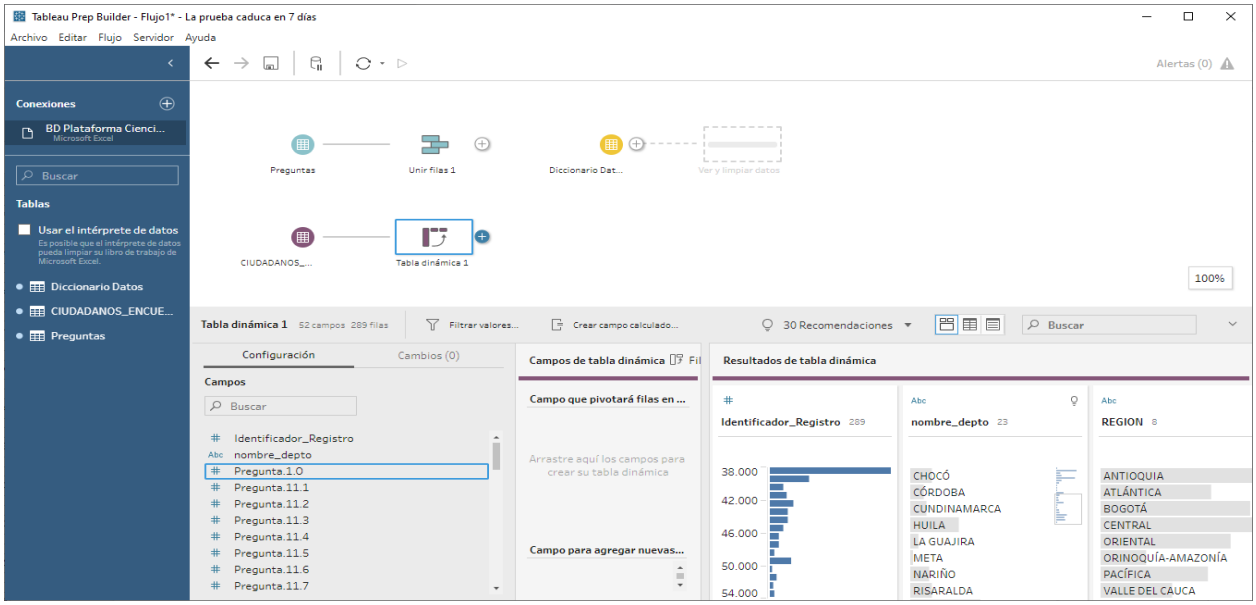
Nota: Base de datos original con registros recolectados por Min TIC y publicados. Fuente: El autor.

Como se mencionó anteriormente una de la tareas más dispendiosas o que demanda demasiado tiempo además de recolectar los datos es depurar estos, lo cual requiere de tres procesos de un análisis a detalle de la información, su reconocimiento e interpretación, ejecutados los procesos mencionados se encontraron las siguientes novedades a subsanar, la base de datos en el libro de diccionario de datos no guardaba un orden que facilitara la interpretación por tal motivo debía ser reorganizada, algunas de las preguntas en la base de datos en el libro de respuestas de la encuesta guardan estricta relación con otras, ya que son la continuación o complemento por ello se realizó consulta de la fuente directa de los datos abiertos para realizar las correcciones y complementar los faltantes, algunas preguntas al validarlas en el diccionario de datos tienen valores que están fuera de los rangos preestablecidos para su interpretación, todas estas novedades demandaron el cambio de algunos valores para facilitar la tarea de realizar la analítica de los datos. Cabe resaltar la importancia en el buen diseño del instrumento de recolección de datos , de tal manera que no dé lugar a respuestas ambiguas o campos sin información , que dificulten el proceso de preparación de la data para su análisis. Los tipos de preguntas también juegan un rol fundamental al momento de la recolección de la información, se recomienda realizar preguntas cerradas y no abiertas ya que estas dificultan su agrupamiento.

Por medio de la aplicación Tableau Prep se pudieron ejecutar todos los cambios mencionados y reorganizar la data o depurar para su análisis, con la ventaja de que Tableau Prep facilita que esta data pueda ser analizada en cualquier otra aplicación que cumpla con este fin, no específicamente en Tableau Public o Tableau desktop que pertenecen a la misma compañía.

Figura 6.

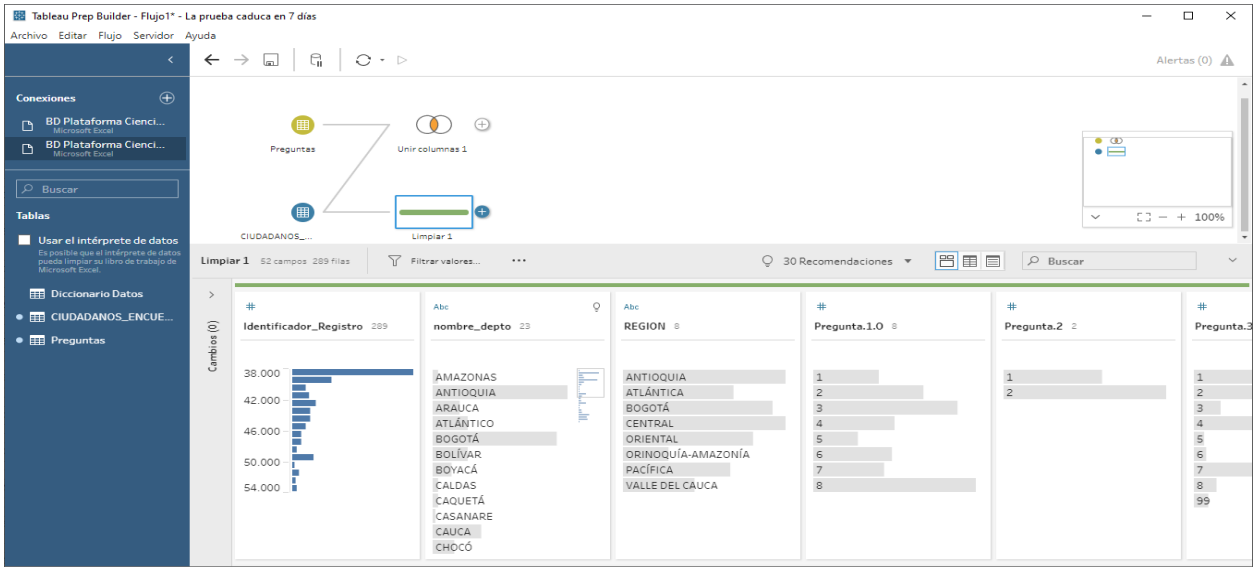
Relaciones entre tablas base de datos:



Nota: Relaciones creadas entre las tablas del archivo CSV datos abiertos Min TIC.
Fuente: El autor.

Figura 7.

Limpieza de los datos con Tableau Prep:

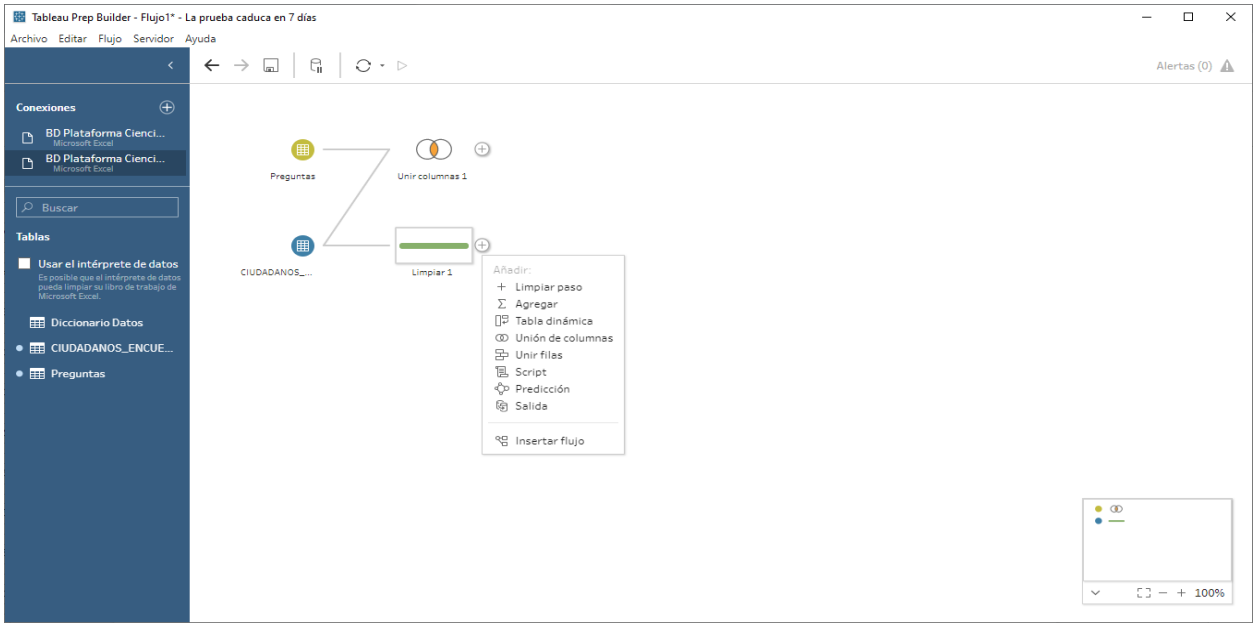


Nota: Fuente de datos filtrada y administrada con la herramienta Tableau Prep con valores previos. Fuente: El autor.

Todas estas relaciones se pueden realizar utilizando la metodología de arrastrar y soltar “drag and drop” y las funciones que vienen predeterminadas en la aplicación las cuales permiten la limpieza de los datos, agregar o quitar columnas y filas, crear tablas dinámicas, realizar script, predicciones, insertar flujos de datos o unir diferentes tablas.

Figura 8.

Funcionalidades Tableau Prep:

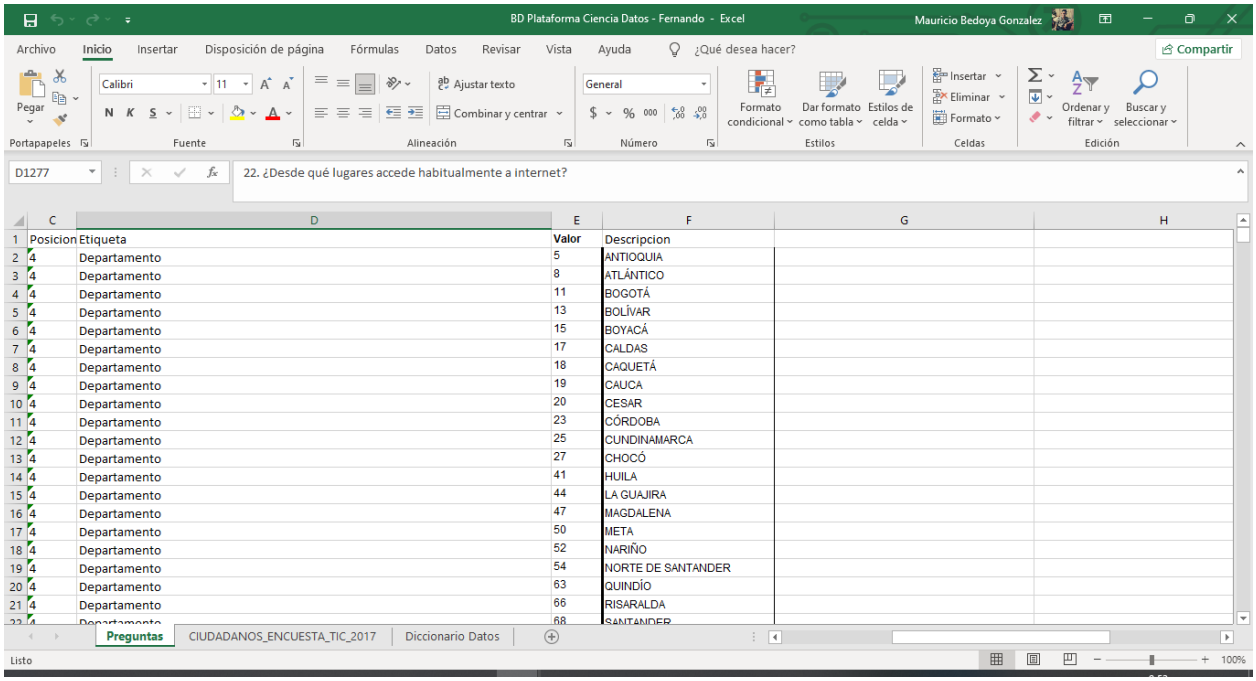


Nota: El grafico ilustra la interfaz de usuario de la herramienta Tableau Prep y sus funcionalidades. Fuente: El autor.

Después de procesar la información con dicha aplicación esto permite generar un nuevo archivo base en formato CSV el cual se utilizara en la aplicación destinada para el análisis de los datos, este nuevo archivo sigue conservando una relación estrecha con el anterior, el cambio que a simple vista se percibe es que el nuevo archivo cuenta con un nuevo libro de datos el cual permite establecer un orden y relación directa entre las entradas de datos o resultados de la encuesta y el diccionario de datos; este nuevo libro se le asignó por nombre preguntas como se muestra en la nueva gráfica:

Figura 9.

Nueva Base de datos elaborada.



	C	D	E	F
1	Posicion	Etiqueta	Valor	Descripcion
2	4	Departamento	5	ANTIOQUIA
3	4	Departamento	8	ATLÁNTICO
4	4	Departamento	11	BOGOTÁ
5	4	Departamento	13	BOLÍVAR
6	4	Departamento	15	BOYACÁ
7	4	Departamento	17	CALDAS
8	4	Departamento	18	CAQUETÁ
9	4	Departamento	19	CAUCA
10	4	Departamento	20	CESAR
11	4	Departamento	23	CÓRDOBA
12	4	Departamento	25	CUNDINAMARCA
13	4	Departamento	27	CHOCÓ
14	4	Departamento	41	HUILA
15	4	Departamento	44	LA GUAJIRA
16	4	Departamento	47	MAGDALENA
17	4	Departamento	50	META
18	4	Departamento	52	NARIÑO
19	4	Departamento	54	NORTE DE SANTANDER
20	4	Departamento	63	QUINDÍO
21	4	Departamento	66	RISARALDA

Nota: Nueva base de datos creada con la aplicación Tableau Prep para ser analizada.
Fuente: El autor.

Esta cuenta con las columnas de posición, etiqueta la cual corresponde a la pregunta en el libro de respuestas, valor asignado a los datos para facilitar su tabulación, y la columna descripción donde podemos interpretar según el diccionario de datos como se debe leer cada valor asignado en las respuestas. Se espera que con este nuevo orden dado a los datos se pueda de una manera más intuitiva obtener resultados que beneficien la toma de decisiones o la generación de resultados en el caso de nuestro proceso de investigación.

Reiterando la importancia de la preparación de los datos o el también llamado preprocesamiento de la información, que, aunque no es una tarea bien valorada dentro de los procesos de analítica de datos es clave para su correcto funcionamiento. Ya que realizar un buen proceso de validación, depuración y aumento de los datos en la unificación es vital para obtener

conocimiento o las llamadas insights que sean significativas y precisas para poder partir de ellas y darle validez y poder a cualquier proceso de analítica con la eficacia en la preparación de los datos en su etapa inicial. Todo ello encaminado en afianzar y dar solidez por medio de la data a la futura toma de decisiones o soportar las acciones de los líderes de proceso.

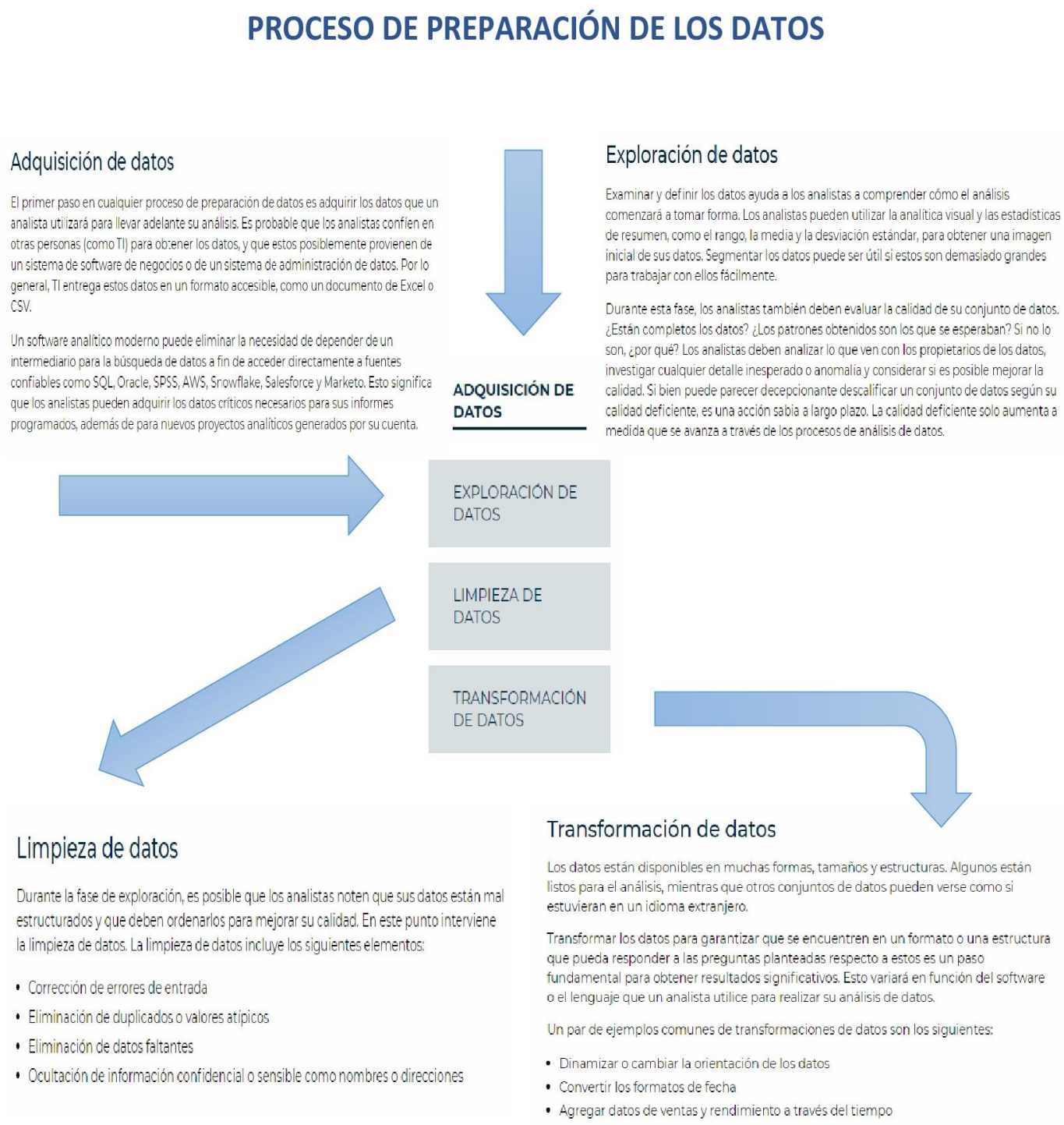
Las decisiones deben estar alineadas con las metas, los objetivos y las iniciativas de la organización, todas las personas que pertenecen a la organización tienen la capacidad para la toma de decisiones, sin embargo, los líderes son los responsables de tomar decisiones soportadas en los datos. El hecho de poder preparar los datos de manera íntegra, asegura una mayor comprensión a los analistas de la información, lo que repercute en análisis precisos y significativos, que conllevan a mejores insights y por consiguiente mejores resultados.

Con la intención de realizar un análisis más profundo de los Insights en las organizaciones, los equipos deben trazar una estrategia para preparar la data priorizando la accesibilidad, la transparencia y la capacidad de repetición. Es decir, cualquier miembro de la organización debe poder acceder con facilidad a los datos, de una manera segura desde una fuente confiable, debe ser capaz de ver, perfeccionar y verificar cualquier paso del proceso de preparación de los datos y a su vez permitir que los datos integren soluciones diseñadas para ampliar la capacidad de repetición.

La preparación de los datos según las necesidades de la industria o los líderes del proceso de investigación tiende a tornarse diferente según las prioridades, pero guardando o siguiendo las etapas que relaciono a continuación:

Figura 10.

Imagen de etapas de la preparación de los datos.



Nota: La grafica explica los elementos para tener en cuenta para la preparación de los datos.
Fuente: El autor.

La disposición de los datos puede variar según la disponibilidad de los datos y la formulación de las preguntas. Es habitual tener que revisar información previamente analizada ya que se generan a medida que se avanza nuevos insights basados en información adicional. El proceso completo de preparación de los datos como se ha evidenciado puede ser lento, complejo y repetitivo. Por este motivo se debe garantizar la ejecución de los pasos de manera individual fácilmente para que los encargados del análisis de la data inviertan menos tiempo en la preparación y más en el análisis de esta.

Implementar la plataforma de análisis de datos.

La disposición de los datos puede variar según la disponibilidad de los datos y la formulación de las preguntas. Es habitual tener que revisar información previamente analizada ya que se generan a medida que se avanza nuevos insights basados en información adicional. El proceso completo de preparación de los datos como se ha evidenciado puede ser lento, complejo y repetitivo. Por este motivo se debe garantizar la ejecución de los pasos de manera individual fácilmente para que los encargados del análisis de la data inviertan menos tiempo en la preparación y más en el análisis de esta.

Después de organizar la data uno de los objetivos fundamentales del proyecto es evidenciar el análisis de los datos abiertos por medio de las herramientas seleccionadas para tal fin, que nos permita adquirir experiencia y analizar las funcionalidades de las aplicaciones de análisis de datos para basarnos en estas y poder ejecutar nuestro desarrollo. Es importante entonces para la construcción de una base de datos Multidimensional o Data Warehouse en donde se consolide la información, se cumpla con los procesos de análisis e históricos de información requeridos, centralizar una gran variedad de datos e información, interpretar dicha información y darle un valor agregado para beneficio del negocio o el proceso de investigación, con un fácil acceso y visualización por parte de los usuarios que soporte la toma de decisiones del negocio.

El datawarehouse o almacén de bodega de datos estará conformado por distintas datamart o bases de datos centradas en información específica que permiten integrar la encuesta de ciencia de datos, con la herramienta analítica Tableau, facilitando el seguimiento y evaluación de indicadores de interés para los investigadores. Estos se construirán con los parámetros abiertos con el objetivo de que pueda seguir creciendo de acuerdo con las necesidades de adición de más encuestas, aprovechando la flexibilidad de su estructura. Se hace énfasis en que para futuras

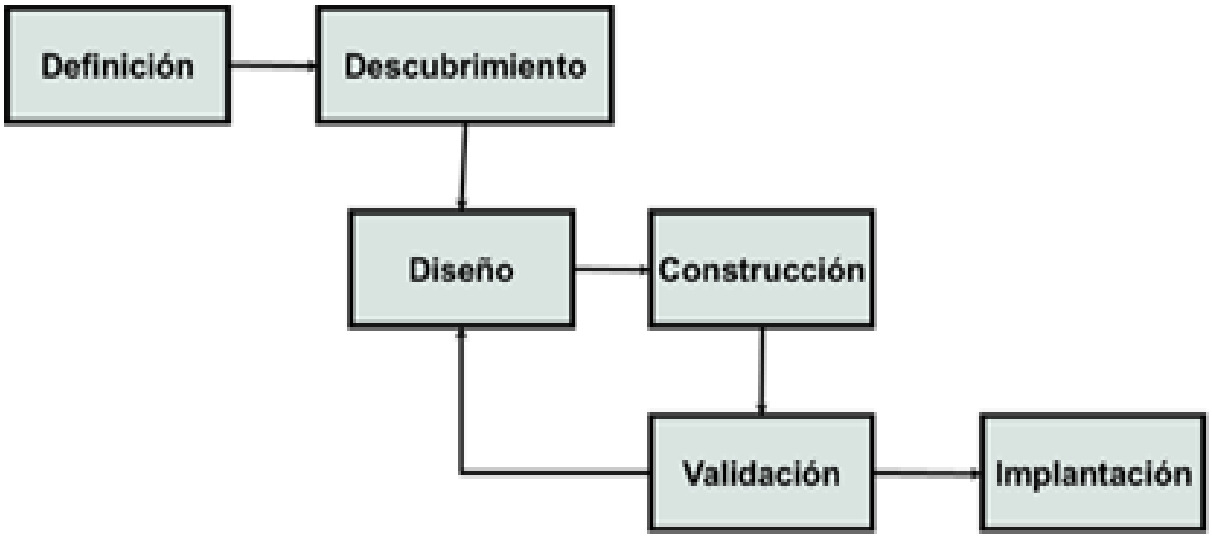
interacciones subsecuentes en el desarrollo de la bodega de datos todos los elementos desarrollados pueden ser reutilizables, lo que agilizará futuras construcciones y disminuirá no solo tiempos de ejecución de los nuevos proyectos si no también posibles costos.

Metodología de la implementación:

Como lo mencione en el punto del alistamiento de la información, la implementación de la inteligencia de negocios o la inteligencia en el análisis de los datos tiende a generar un reproceso, porque a medida que se adiciona o se modifica información en las tablas de datos se generan nuevos elementos para analizar o para tener en cuenta cuando se realizan interacciones en la construcción de la bodega de datos como lo muestra la imagen.

Figura 11.

Diagrama flujo bodega de datos.



Nota: La grafica explica los elementos para tener en cuenta para la preparación de los datos.
Fuente: El autor.

Los elementos que se muestran en la figura anterior son importantes ya que si se quiere aportar valor al negocio u organización se tienen que ejecutar paso a paso para la realización de una buena bodega de datos y tener en cuenta todo lo que esta necesita o demanda:

En la definición es importante identificar el alcance del proyecto, cada uno de los procesos de negocio del mismo o su relevancia en caso de las temáticas abordar, además de los recursos que se requieren para su implementación, es por ello importante delimitar un documento que consigne los requerimientos del negocio y otro donde se evidencie su matriz de procesos, aunque en este caso no se hace necesario ya que se está realizando solamente el análisis de una información que ya fue recolectada y tiene una finalidad delimitada con anterioridad que es el análisis de la brecha digital, pero es importante tener en cuenta el proceso general.

Viene entonces otro momento el cual es el descubrimiento donde el objetivo es obtener las necesidades de la información, y los reportes que se deben generar, así como la plataforma tecnológica que mediara en el análisis del proyecto o la temática abordada, donde se resaltan algunos entregables necesarios como lo son los requerimientos de datos, la especificación de los reportes a generar, la infraestructura técnica o tecnológica con la que se cuenta y un modelo de seguridad de la información.

En la fase de diseño se intentará cumplir con las especificaciones para el repositorio o lugar donde se van a alojar los metadatos, para cumplir con los ETL (Extract, Transform, Load) que consiste en el proceso que permite a las agrupaciones o usuarios mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, además de adicionar los elementos o estructuras necesarias que soporten las gráficas de analítica de datos o dashboard que se generen. Para esta fase es importante tener en cuenta la matriz de dimensiones y las tablas de hecho, el diseño del esquema multidimensional, el diseño del ETL y el diseño de la estrategia de pruebas.

En la construcción se centrará solamente en el desarrollo de los componentes que integran el

modelo multidimensional, en la validación se realizarán pruebas de los requerimientos de datos por medio de la generación de reportes y la construcción de ilustraciones que permitan analizar la información. Como punto final en la implantación se verá el sistema multidimensional puesto en marcha y producido según las necesidades de la organización o usuarios que esté encargada de su administración.

Generalidades de los ETL (Compilación de los datos)

Es de vital relevancia cuando se quiere tener un buen proceso al mover datos desde múltiples fuentes, reformatearlos, limpiarlos, y cargarlos en otra base de datos para generar nuevos datamart o complementar los existentes, configurar el ETL para que permita programaciones, pedidos de cargue y seleccionar las tablas que serán utilizadas durante el proceso, establecer un buen control de transmisión y una estructura organizada de los datos; el cargue de archivo temporales nos permite validar la información y su estructura hablando del tema de tipos de datos y longitud, en la validación es muy importante garantizar la integridad referencial en el ODS ya que se está hablando de datos abiertos. En la fase de transformación o de cargue final de las tablas definidas, se realizarán cambios utilizando funciones propias del software que permitan la preparación de los datos, o las que los encargados de la preparación y análisis de los datos consideren necesarios, al final vienen las sumarias donde se generan tablas de resumen que permiten agilizar la validación de la información y adicionar algunos registros que no se consideran fijos, los cuales requieren cambios frecuentes. En el caso de este análisis, aunque es importante conocer todo este proceso que ayuda a la consolidación de un buen proceso de generación de bodega de datos lo que redunda en una analítica más confiable, no se centrará en su puesta en marcha, ya que las herramientas antes mencionadas de preparación de los datos como Tableau Prep permiten la depuración de la información y la consolidación de esta para un

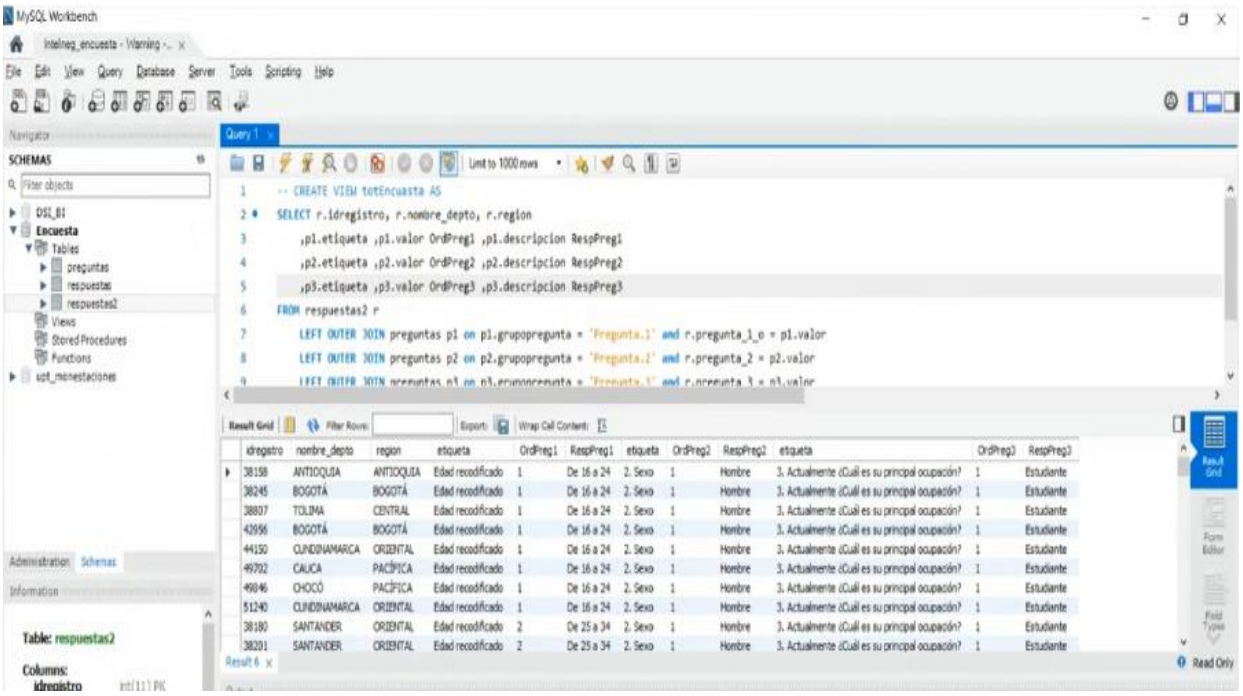
análisis más rápido e intuitivo. Aun así, es importante mencionar que la información que se tiene consolidada en el archivo CSV puede ser gestionada con herramientas más avanzadas como lo son MySQL, pero para ello la persona debe tener previamente conocimientos acerca de programación básica.

Esta base de datos se puede alojar en un servidor para garantizar el acceso a la información y su administración en este caso tenemos a disposición un servidor con las siguientes características sistema operativo LINUX UBUNTU 20.04, características de hardware procesador de 8 núcleos, Memoria RAM de 30 GB y disco duro SSD de 800 GB. En excel se normaliza la data de la encuesta, donde se organiza una base de datos con la información de la pregunta, y la respuesta válida para cada caso.

Esta información normalizada sirve para alimentar la bodega de datos en MySQL, que contiene las tablas de respuestas y preguntas. Es necesario organizarla de esta manera con el fin de dar interpretación a la data y lograr que el usuario final que diseña los informes no tenga que realizar este proceso desde la herramienta ya sea Tableau o Power BI y así con el motor de la base de datos se realiza la limpieza y depuración de la información. Como se ilustra en la imagen se estuvo analizando este método.

Figura 12.

Depuración de datos MySQL



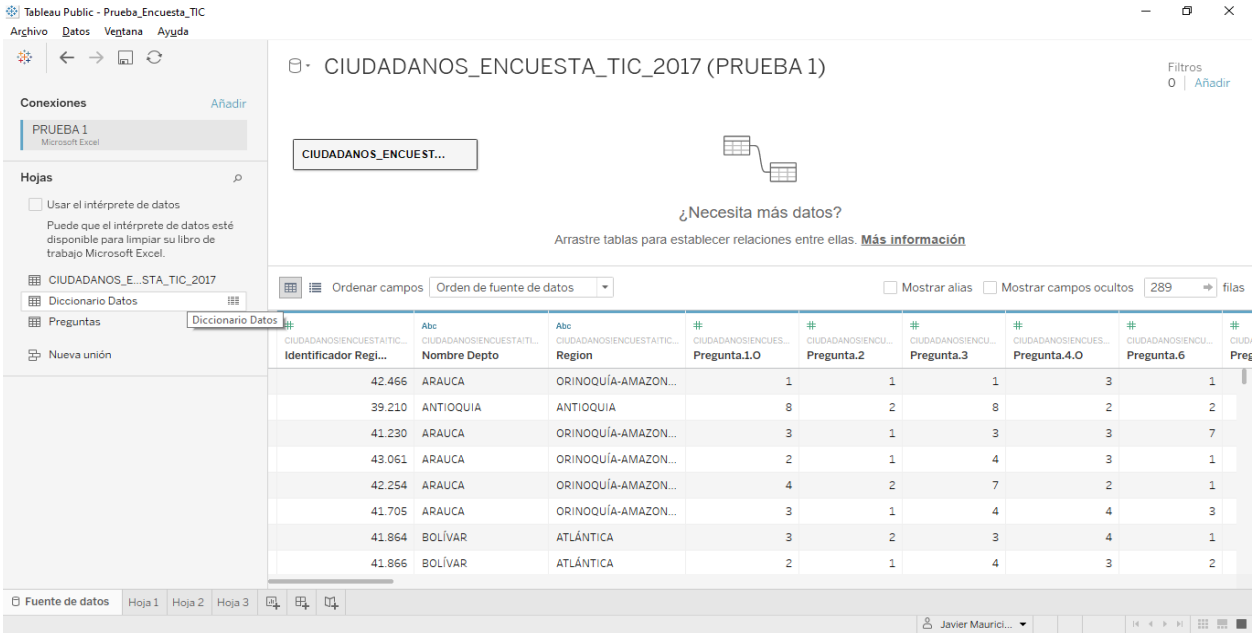
Nota: Prueba de depuración de los datos, análisis y comparación de herramientas.
Fuente: El autor.

Nuevamente se hace aclaración de que esta es una forma en cómo se puede administrar la información y prepararla para su análisis pero no fue el elegido, ya que genera el acarreo de costos para el usuario final y además de eso requiere que la persona que realiza la preparación de los datos y el análisis de los mismos tenga conocimientos básicos y avanzados en la utilización de herramientas informáticas, servidores y lenguajes de programación lo cual se ahorra con la herramienta Tableau Prep y Tableau Public, la primera permite utilizar su versión de prueba e interfaz intuitiva para preparar la data y la segunda permite realizar el análisis de manera gratuita con la única condición de que se tiene que trabajar con datos abiertos y que los gráficos generados no se guardarán de manera local sino que estarán publicados en su servidor para que otros usuarios puedan hacer uso de ellos.

La herramienta Tableau Public permite al usuario o encargado del análisis de los datos realizar el cargue de los mismos desde distintas fuentes como se muestra en la siguiente ilustración:

Figura 13.

Cargue a la fuente de datos Tableau Public.



Nota: Interfaz para cargar los datos de la herramienta Tableau Public, fuente base de datos. Fuente: El autor.

La herramienta Tableau Public le permite al usuario por medio de la metodología de arrastrar y soltar crear gráficas, dashboard e historias, también posibilita modificar y crear agrupaciones de información, etiquetas, filtros, rangos, en general administrar los datos de una manera fácil e intuitiva, la cual no necesita conocimientos especializados en lenguajes de programación, aun así, cuenta con funcionalidades programables que favorecen una interfaz amigable con el usuario.

En la fuente de datos el usuario puede crear relaciones entre las tablas y sacar el máximo provecho de los datos, utilizando cálculos avanzados y estadísticos, reorganizando información,

filtrando búsquedas o editando valores dentro de las tablas, además de conectar bases de datos de diferentes fuentes de información ya sea locales o en la nube, herramienta que es bastante novedosa e importante hoy en día, ya que la mayoría de la información debe garantizar su accesibilidad y no hay herramienta más propicia que la nube para poder compartir.

Tableau Public cuenta con licenciamiento libre, solo se debe crear un usuario y contraseña para poder publicar información en el servidor; la condición especial que la plataforma y compañía imponen a los usuarios para su uso, es que la información objeto del análisis debe ser de carácter abierto, interés público o ser disponible de forma libre y las gráficas, dashboard, historias o cualquier elemento que se cree en dicha aplicación o plataforma puede ser copiado, reutilizado, redistribuido, compartido o editado por cualquier persona que haga parte de la comunidad colaborativa de Tableau Public.

Cuenta también con acceso al servidor en el cual se pueden publicar todas las gráficas y esto permite compartirlas con otros usuarios, descargarlas en diferentes formatos o generar enlaces de visualización. Lo cual les ahorraría a los usuarios incurrir en gastos adicionales de alojamiento en la nube.

Procedimiento y elaboración de gráficas.

La fuente de datos generada previamente en la aplicación Tableau Prep de la cual se hace mención en el punto anterior, después de ser depurada y organizada para el análisis de los datos se generó archivo en formato CSV el cual se cargó a la herramienta Tableau Public y de este se generaron utilizando las herramientas antes mencionadas las gráficas que se relacionan a continuación:

El libro de gráficos generado con la herramienta se puede consultar en el servidor desde el siguiente enlace:

https://public.tableau.com/views/Anlisis_Brecha_Digital_Colombia/Hoja1?:language=es-ES&:display_count=n&:origin=viz_share_link

Figura 14.

Interfaz de visualización de graficas.

Nota: Interfaz para visualizar graficas desde el servidor de Tableau Public.
Fuente: El autor.

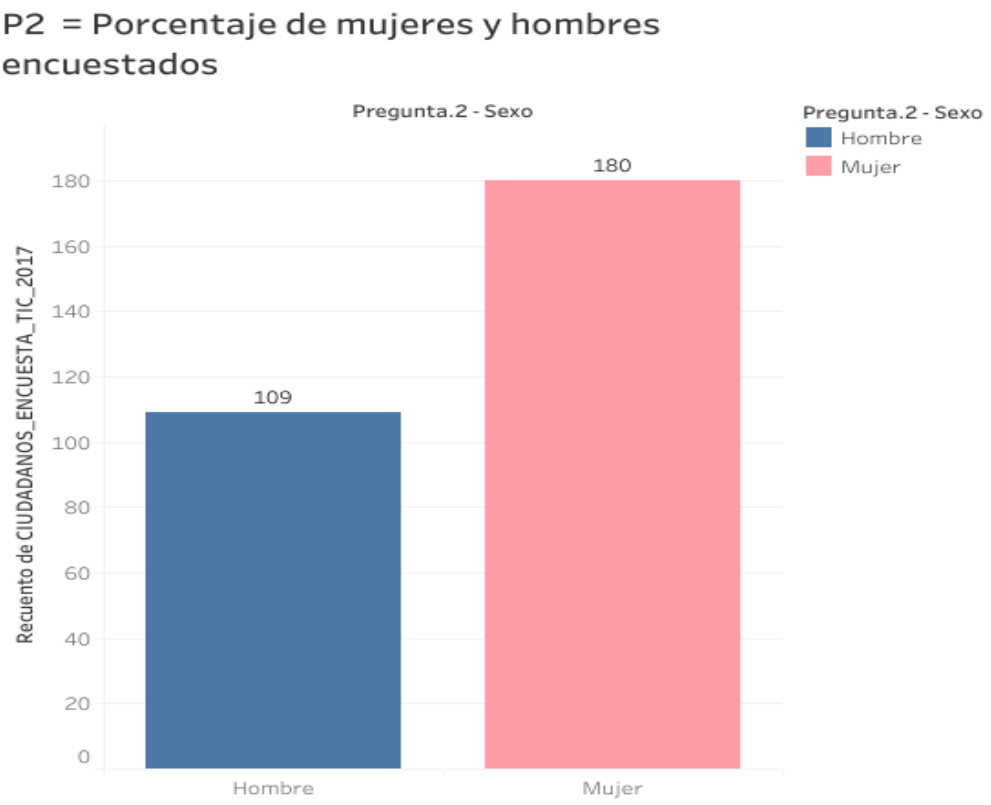
Estas mismas fueron exportadas en formato de imagen para adjuntar en el presente documento como evidencia del manejo de la plataforma y sus funcionalidades y como soporte para realizar posteriormente el estudio del comportamiento del fenómeno de la brecha digital.

Estudiar el comportamiento del fenómeno por medio de la plataforma.

En esta sección se presenta el análisis de los resultados obtenidos después de graficar en la aplicación Tableau Public la informacion seleccionada por medio de los datos abiertos de la encuesta aplicada por el ministerio de las TIC´s, lo cual permite determinar y analizar el impacto o la brecha digital que existe entre el acceso y el uso de herramientas tecnológicas e internet por parte de la ciudadanía colombiana. A continuación, se presentan las algunas de las gráficas más representativas:

Figura 15.

Población de hombres y mujeres encuestados.

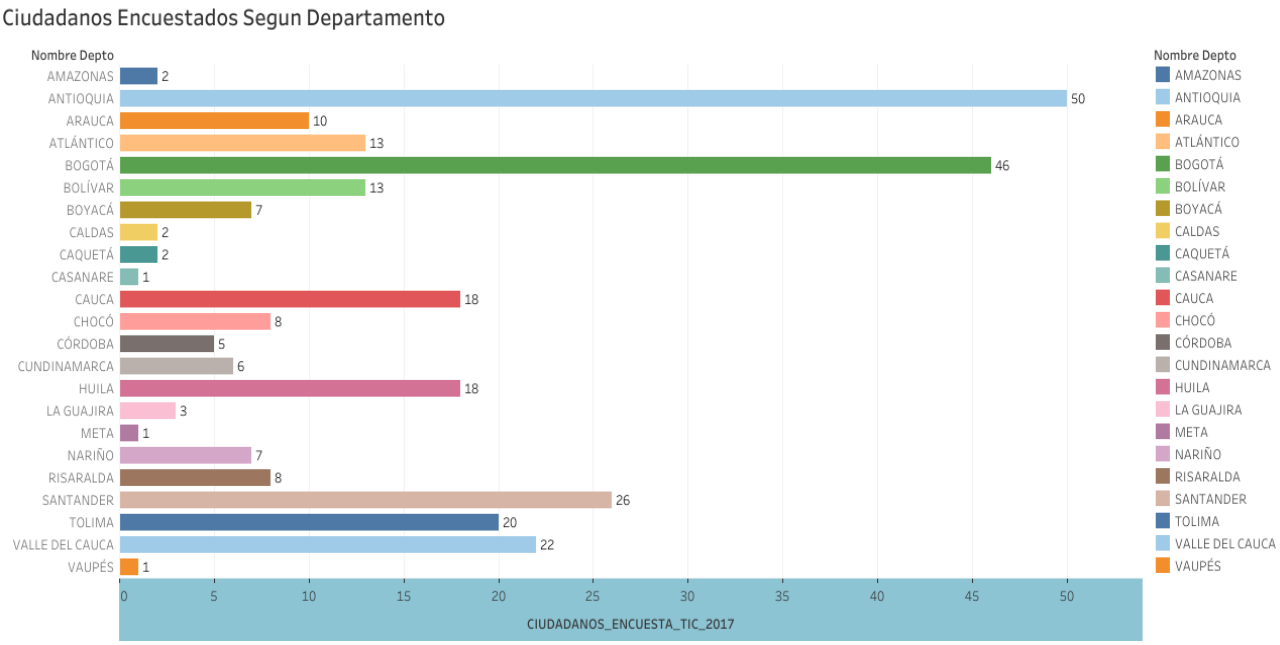


Nota: Clasificación por sexo de personas encuestadas.
Fuente: El autor.

De los encuestados, en la plataforma, en la Hoja 1 se encuentra que el mayor porcentaje de personas encuestadas corresponde al sexo femenino (180 de 289). Esta situación indica, que en los datos que se analizaron prima la opinión o percepción del sexo femenino sobre el análisis de fenómeno.

Figura 16.

Muestra por departamento de ciudadanos encuestados

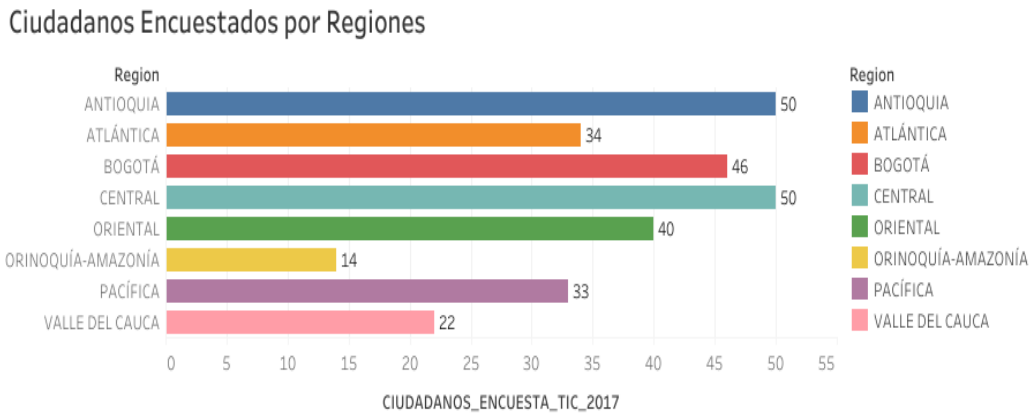


Nota: Clasificación por sexo de personas encuestadas.
Fuente: El autor.

De los encuestados, en la plataforma, en la Hoja 2 se encuentra que el mayor número de personas encuestadas corresponde a los departamentos de Antioquia (50), Bogotá (46). Esta situación indica, que el número de encuestados en los demás departamentos es relativamente bajo como es el caso de Risaralda (8), Caldas (2). La baja participación de encuestados en algunos departamentos limita la toma de decisiones objetivas en cuanto al acceso a herramientas tecnológicas e internet.

Figura 17.

Encuestados por Regiones según clasificación DANE.

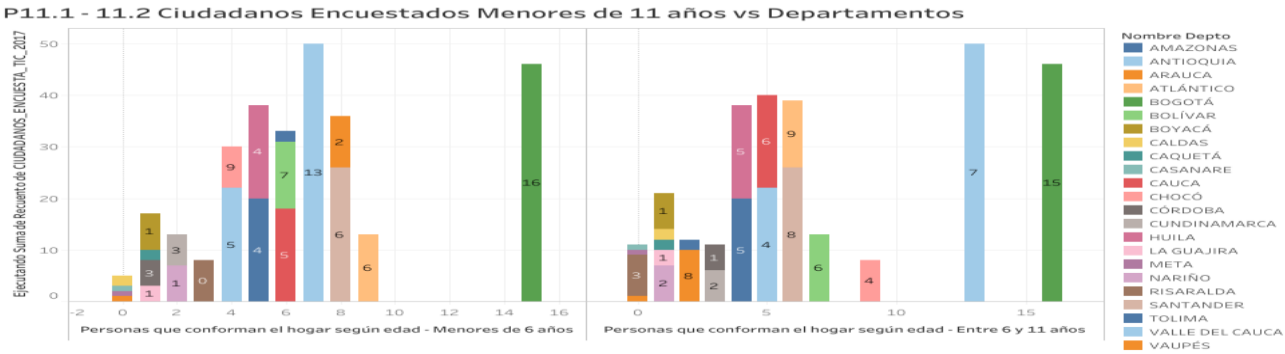


Nota: Clasificación por sexo de personas encuestadas.
Fuente: El autor.

De los encuestados, en la plataforma, en la Hoja 3 se encuentra que el mayor número de personas encuestadas corresponde a las Regiones de Antioquia (50), Bogotá (46), Central (50), Oriental (40). Esta situación indica, que al momento de desarrollar un proyecto de inversión en las tecnologías de la informacion y la comunicación que beneficie a una región determinada, se puede utilizar esta data para seleccionar la que impactaría a mayor número de población.

Figura 18.

Ciudadanos encuestados según edades.

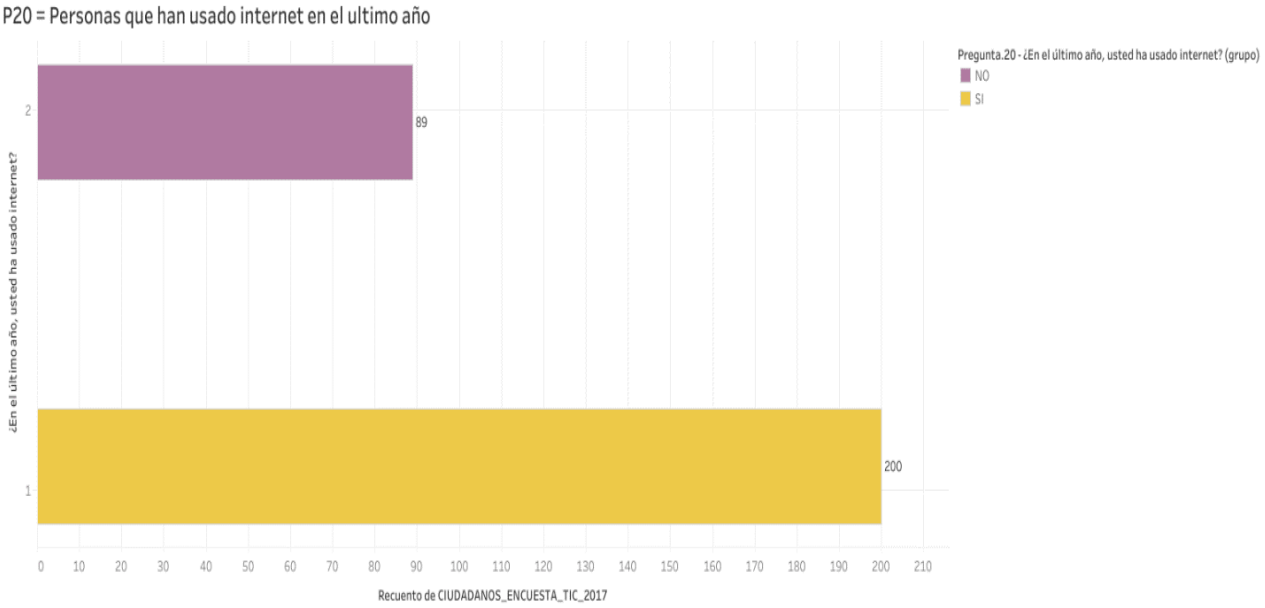


Nota: Clasificación por departamentos de ciudadanos encuestados según rango de edad seleccionado. Fuente: El autor.

De los encuestados, en la plataforma, en la Hoja 4 se encuentra que un gran porcentaje de los encuestados se clasifica dentro de un rango de edad de menores de 6 años y menores de 11 años significativo en cada región. Esta situación indica, que en los departamentos de Bogotá (31/48) y Antioquia (20/50) gran número de los encuestados corresponden a este rango de edad.

Figura 19.

Personas que han tenido acceso a internet en el último año.

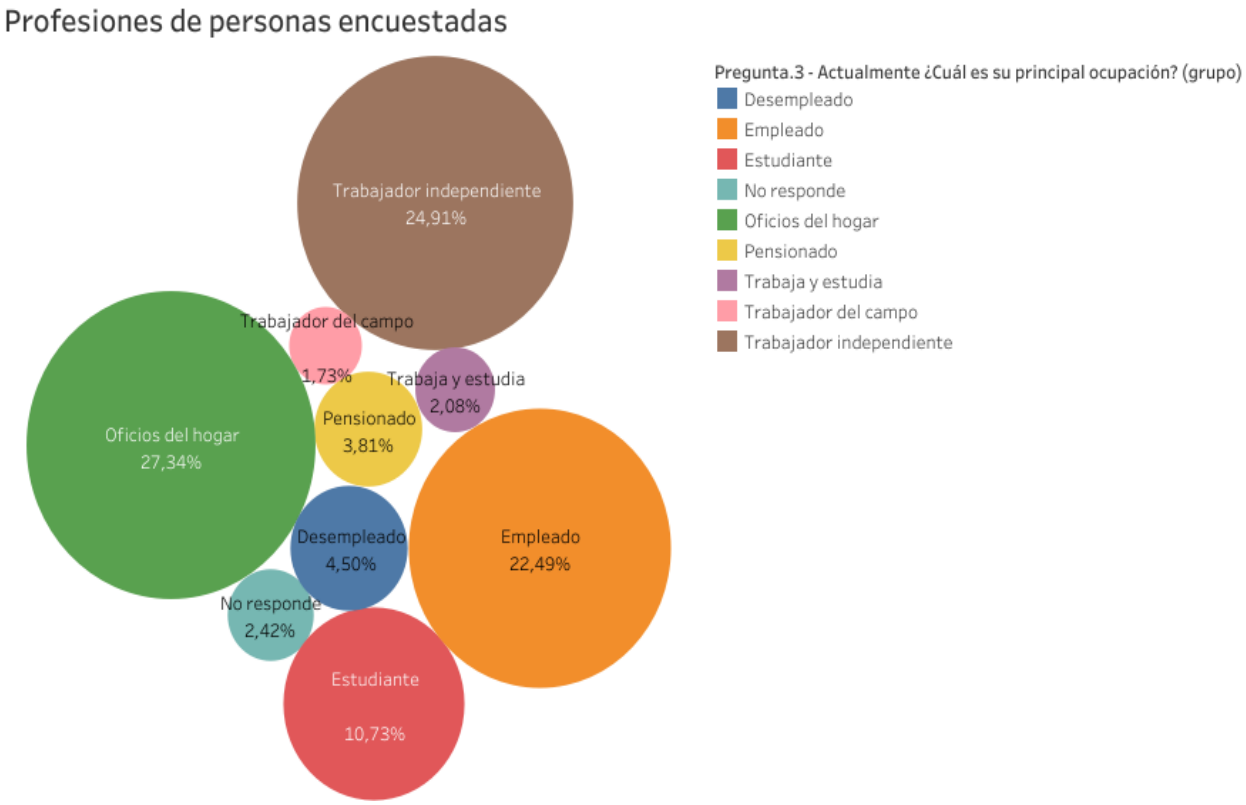


Nota: Muestra de personas que han utilizado internet en el último año según encuesta.
Fuente: El autor.

De los encuestados, en la plataforma, en la Hoja 5 se encuentra que un alto porcentaje de personas han hecho uso de Internet durante el último año (200 de 289). Esta situación indica, que, aunque existe ausencia en temas de dispositivos, las personas, buscan el acceso a este servicio que les permite participar en actividades como educación, salud y entretenimiento.

Figura 20.

Personas encuestadas según su ocupación.

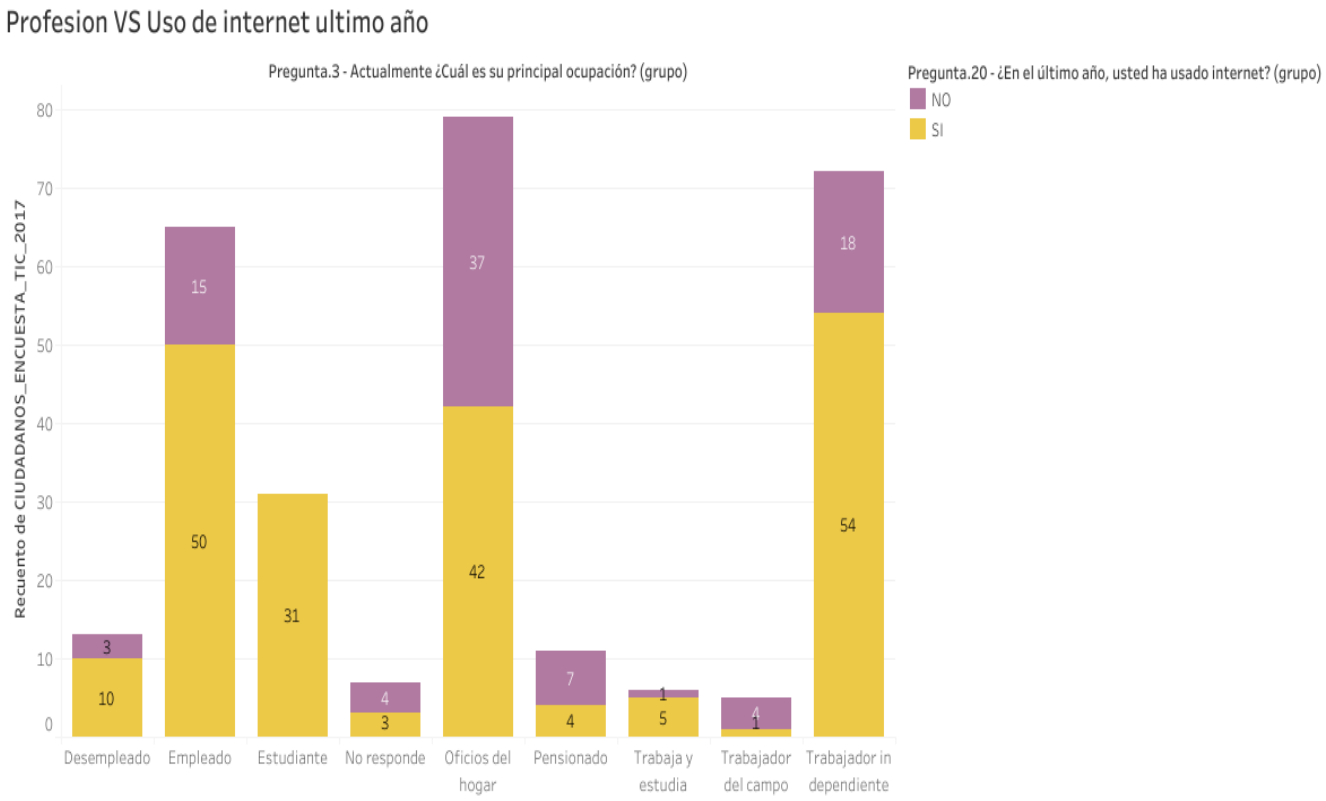


Nota: Muestra de personas encuestadas según la ocupación que desempeñan.
Fuente: El autor.

De los encuestados, en la plataforma, en la Hoja 6 se encuentra que un alto porcentaje de personas desarrolla las siguientes actividades Empleado (22,49%), Trabajadores independientes (24,91%), oficios del hogar (27,34 %). Esta situación indica, que, para las actividades de trabajador del campo (1,73%) y personas que trabajan y estudian (2,08%) se recolectaron pocos datos.

Figura 21.

Profesión Vs Uso de internet en el ultimo año

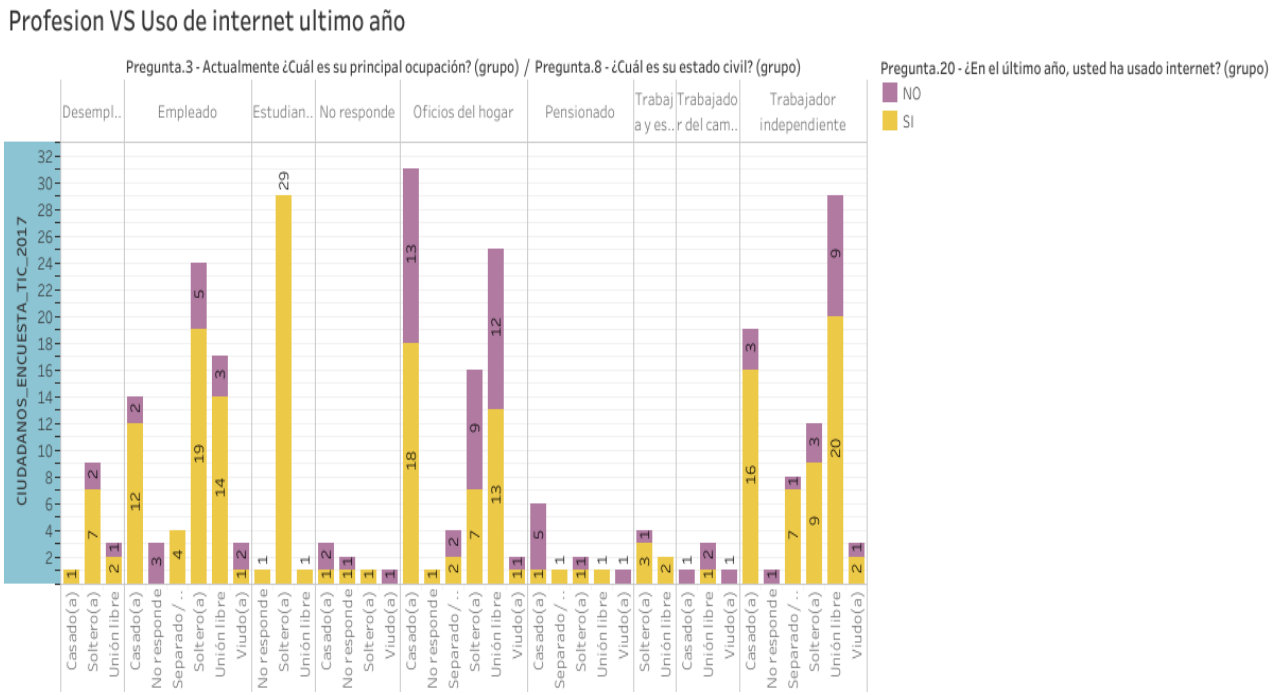


Nota: Personas por ocupación, que usaron internet en el último año según la encuesta.
Fuente: El autor.

De los encuestados, en la plataforma, en la Hoja 7 se encuentra que según las ocupaciones de 65 empleados encuestados 50 de ellos utilizan internet, en los oficios del hogar de 79 encuestados 42 utilizan internet, de igual manera de 72 Trabajadores independientes 54 utilizan internet, y en cuanto a los 31 estudiantes encuestados el 100% utilizan internet. Esta situación permite determinar que la mayoría de las personas que ejercen una profesión u oficio cuentan con acceso a internet y que las personas que no cuentan con empleo o pertenecen al sector rural no tienen internet o no han accedido a él en el último año.

Figura 22.

Profesión vs ocupación y estado civil de la población encuestada.

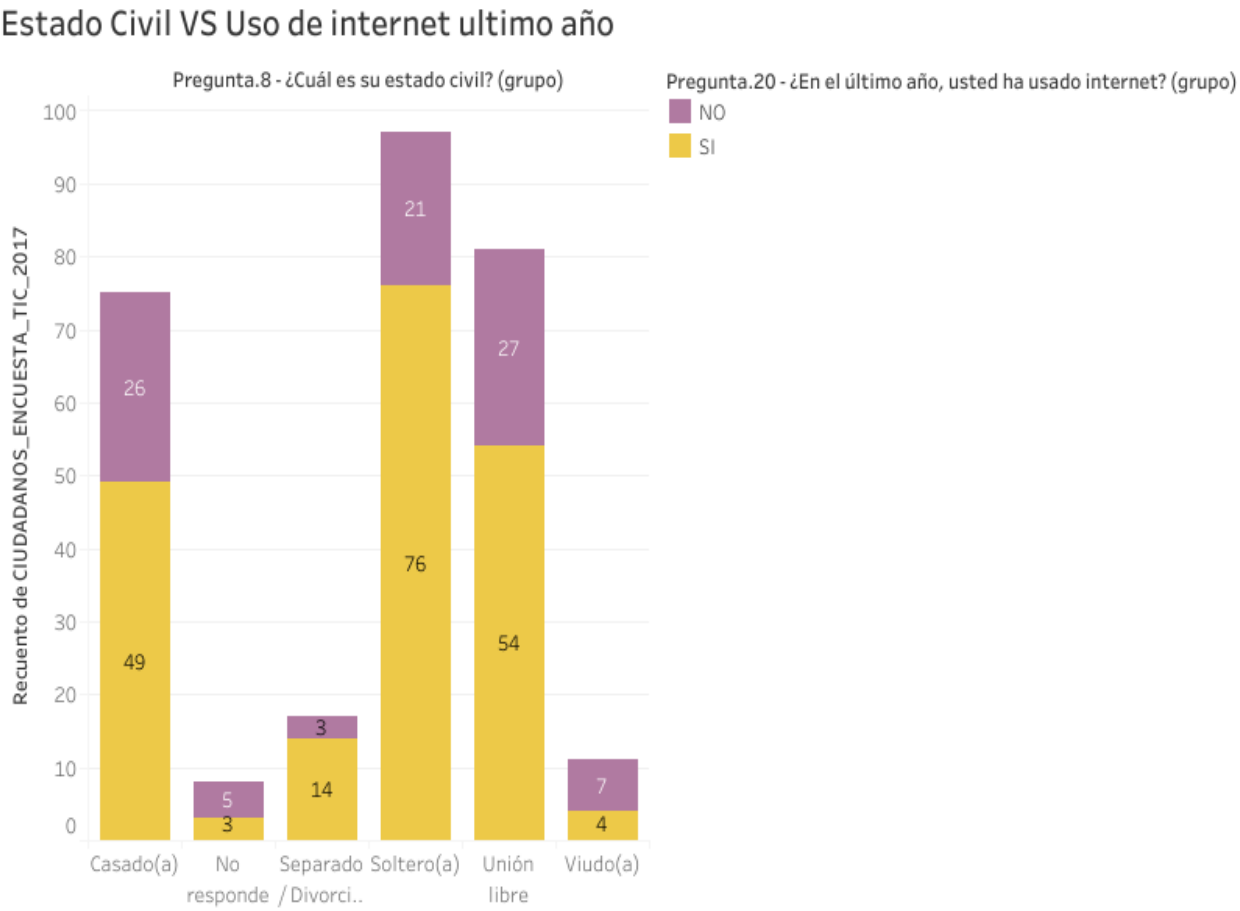


Nota: Personas por ocupación, que usaron internet en el último año clasificados por estado civil.
Fuente: El autor.

De los encuestados, en la plataforma, en la Hoja 8 se encuentra que según su estado civil de 65 empleados, 19 de estado civil solteros son los que hacen mayor uso del internet, en los oficios del hogar de 79 encuestados 18 con estado civil casados son los que hacen mayor uso del internet, de igual manera de 72 Trabajadores independientes 20 con estado civil unión libre son los que hacen mayor uso del internet, y en cuanto a los 31 estudiantes 29 con estado civil solteros son los que hacen mayor uso del internet. Esta situación permite determinar que la mayoría de las personas con estado civil solteros y casados son lo que hacen mayor uso del internet y los que menos uso hicieron del internet en el último año están dentro del oficios del hogar.

Figura 23.

Estado civil de la población vs uso del internet

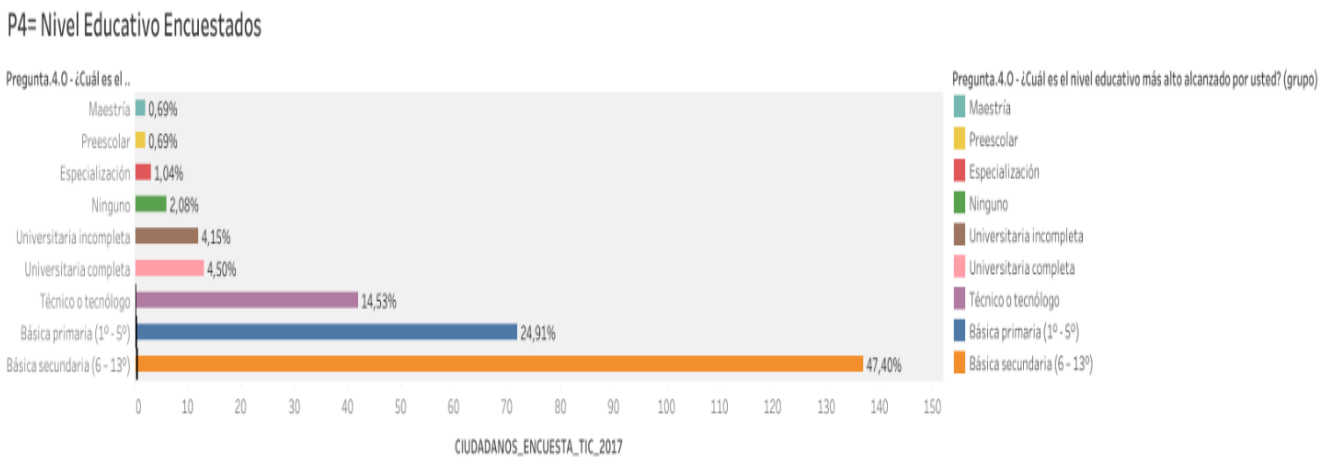


Nota: Personas encuestadas según su estado civil y el uso o acceso al internet en el último año. Fuente: El autor.

De los encuestados, en la plataforma, en la Hoja 9 se encuentra que según su estado civil las personas solteras son las que hacen mayor uso del internet y tienen acceso a las herramientas tecnológicas que se encuentras disponibles en este y por lo tanto también acceso a los dispositivos que les permiten la conexión.

Figura 24.

Nivel educativo de las personas encuestadas.

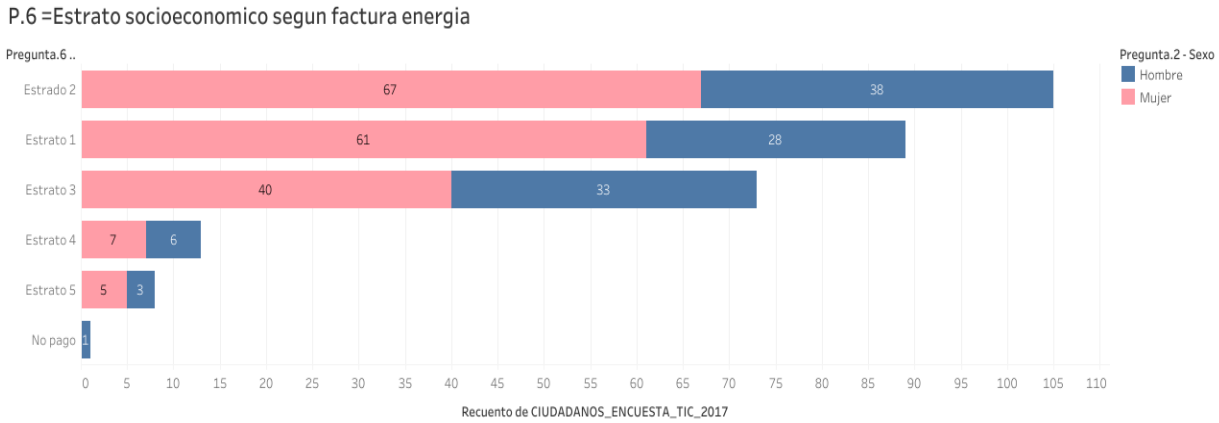


Nota: Personas encuestadas según su nivel educativo. Fuente: El autor.

De los encuestados, en la plataforma, en la Hoja 10 se encuentra que el 47,40% tienen nivel educativo de básica secundaria, el 24,91% tienen nivel educativo de básica primaria y el 14,53% tienen nivel educativo técnico o tecnológico. Esta situación permite determinar que la mayoría de las personas encuestados alcanzaron el nivel educativo media.

Figura 25.

Personas encuestadas según su estrato social y género.

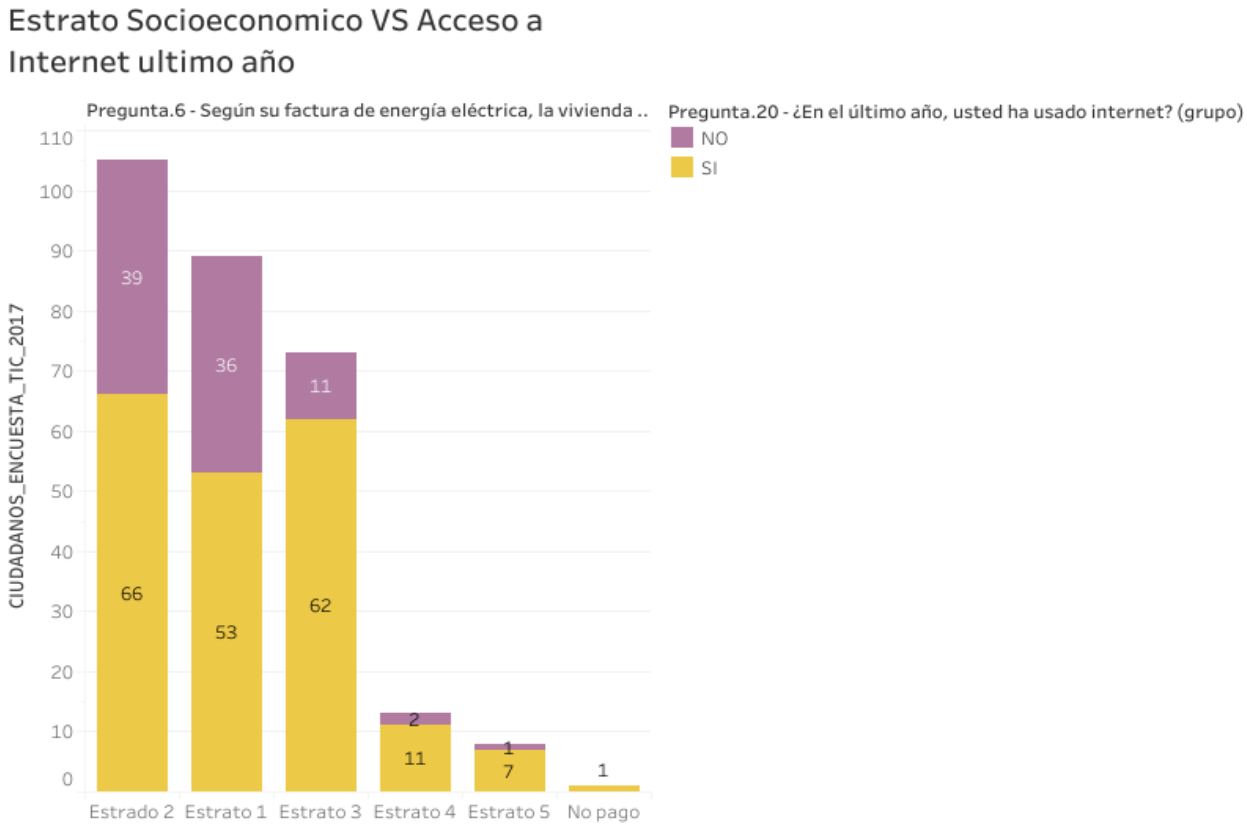


Nota: Personas encuestadas según clasificadas según su género y estrato socioeconómico percibido en la factura de la energía eléctrica. Fuente: El autor.

De los encuestados, en la plataforma, en la Hoja 11 se encuentra que la gran mayoría de las personas encuestadas son de genero femenino y pertenecen a los estratos socioeconómicos 1, 2 y 3.

Figura 26.

Estrato socioeconómico vs el acceso a internet en el último año.

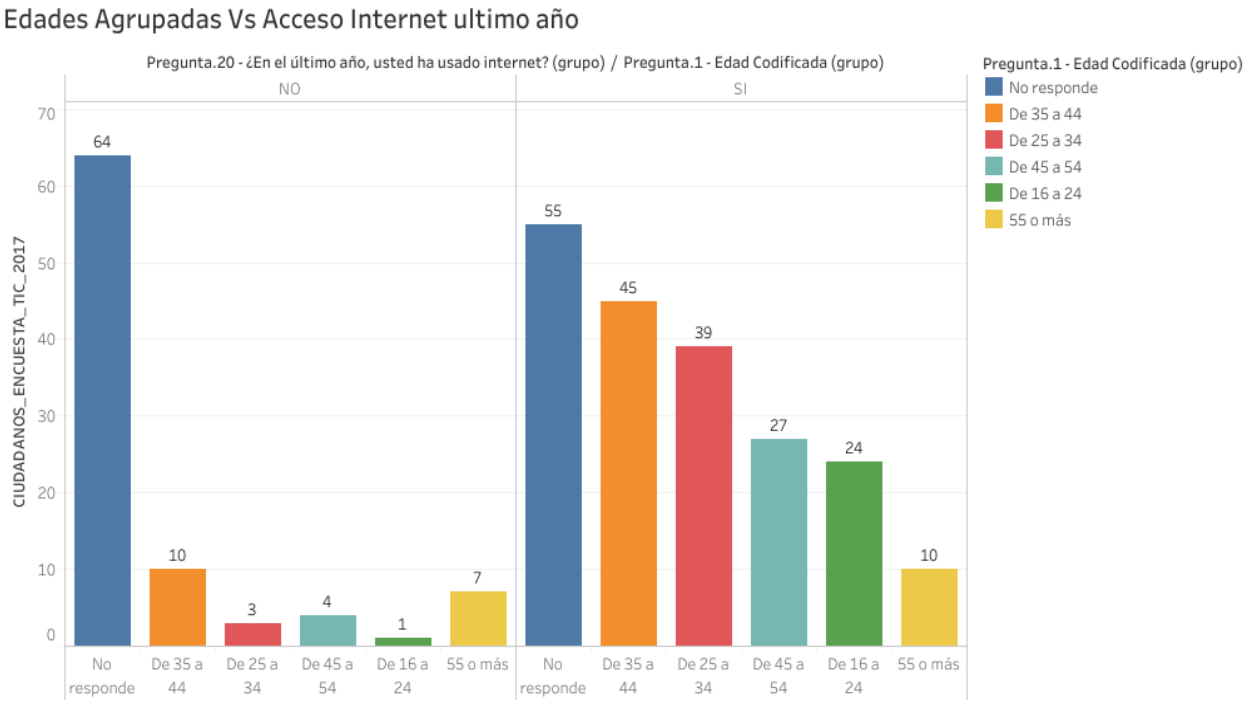


Nota: Personas encuestadas según su estrato socioeconómico, el acceso a internet y herramientas tecnológicas en el último año. Fuente: El autor.

De los encuestados, en la plataforma, en la Hoja 12 se encuentra que los estratos socioeconómicos con mayor acceso a internet y las herramientas que este tiene a disposición fueron estratos 1, 2 y 3. Esta situación permite determinar que las políticas que ha implementado el estado colombiano para beneficiar los estratos antes mencionados en cuenta a subsidios de costos de acceso a internet y las herramientas tecnológicas han surtido efecto.

Figura 27.

Edades agrupadas vs acceso a internet.



Nota: Personas encuestadas según sus edades y el acceso a internet en el último año.
Fuente: El autor.

De los encuestados, en la plataforma, en la Hoja 13 se encuentra que las personas en el rango de edad entre 35 a 44 clasificados como adultos son los que tienen mayor acceso a internet y sus herramientas. Esta situación permite determinar que las personas en el rango de edades de 55 en adelante hacen poco uso del internet.

Capítulo IV: Recomendaciones y conclusiones.

Analizada la informacion de la gran encuesta de TIC´s que nos permite evaluar la brecha digital en Colombia se observó una distribución desigual en el acceso, en el uso, o en el impacto de las Tecnologías de la Información y la Comunicación (TIC) en el sector rural.

La muestra realizada por cada uno de los departamentos y regiones en cuento a la utilización de las TIC es inequitativa, lo que amerita en una futura encuesta tratar de definir el mismo número de personas encuestadas por región, para ser más justos en la toma de decisiones para mitigar o minimizar la reducción de la brecha digital.

Se puede evidenciar que las políticas implementadas por el estado para mejorar el uso y el acceso a internet están demostrando un avance significativo en la disminución de la brecha digital, pero se deben implementar mas esfuerzos en fortalecer el acceso a dichas herramientas tecnológicas en el sector rural.

Se ha demostrado que la gran encuesta es una herramienta útil para la toma de decisiones en cuanto a proyectos y planes de acción que quieran implementar los entes gubernamentales para mitigar o disminuir la brecha digital. Por lo cual se recomienda que se realice de manera periódica.

Se deben implementar estrategias que permitan disminuir el analfabetismo digital en la población mayor a 55 años, ya que la transformación digital ha cambiado la manera en cómo se venían realizando muchos de los tramites o procesos, lo cual les ha generado dificultades en el entorno actual.

Es importante que para una futura encuesta se generen campos obligatorios en las respuestas para todas las preguntas y así disminuir la dificultad presentada en la depuración de esta por la cantidad de valores sin registros. Lo cual repercute en los resultados efectivos a representar en

las gráficas.

El diccionario de datos que permite interpretar los registros debe guardar una equivalencia con la hoja de respuestas. Ya que algunas de las interpretaciones estaban fuera de los rangos de respuesta.

Referencias bibliografía

Aguntín Lacruz, M. d., & Clavero Galofré, M. (2009). Indicadores Sociales de Inclusión Digital: Brecha y Participación Ciudadana. Derecho, gobernanza y tecnologías de la información en la sociedad del conocimiento, 143-165.

Armenta, A., Serrano, A., Cabrera, M., & Conte, R. (2012). The new digital divide: the confluence of broadband penetration, sustainable development, technology adoption and community participation. Information Technology for Development, 345-353.

Banco Mundial. (2016). Informe sobre el desarrollo mundial 2016: Dividendos digitales, cuadernillo del "Panorama general". Washington DC.

Bider, I., & Otto, H. (2015). Modeling a Global Software Development Project as a Complex Socio-Technical System to Facilitate Risk Management and Improve the Project Structure. In Proceedings - 2015 IEEE 10th International Conference on Global Software Engineering, ICGSE 2015 (pp. 1–12). Institute of Electrical and Electronics Engineers Inc. http://doi.org/10.1109/ICGSE.2015.13

Caruso, S. (2014). Creating Digital Communities: A resource to Digital Inclusion. New York: Nova Science Publishers.

Cheng-Hua, W., Yender, M., & Jen-Hwa Kuo. (2011). Mapping the Intellectual Structure of Digital Divide. International Journal of Social Science and Humanity, 49-54.

Cusi, M. L. A., & Bernal, L. D. P. Programa de fortalecimiento de capacidades en DATA ANALYTICS en empresas colombianas.

Davidsson, P. (2000). Multi Agent Based Simulation: Beyond Social Simulation. 2nd International Workshop on Multi-Agent-Based Simulation-Revised and Additional Papers (MABS '00), 97–107.

DiMaggio, P. ., Hargittai, E. ., Russell Neuman, W. ., & Robinson, J. P. . (2001). Social implications of the internet. Annual Review of Sociology, 27, 307–336. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-0035641610&partnerID=40&md5=9865948dfd4eda4fed70adb2e7d05f13

Liarte Muñoz, J. J. (2019). Análisis de datos de las organizaciones. Big data.

Méndez, C. A. P., Méndez, D. J. S., & Romero, M. M. P. (2020). Big Data, Educación y Post-acuerdo. Cultura de Paz en Redes Sociales. Publicaciones e Investigación, 14(3).

Núñez Jover, J. (1999). La ciencia y la Tecnología como Procesos Sociales. Lo que la educación científica no debería olvidar. La Habana.

OECD. (2001). Understanding The Digital Divide. Obtenido de http://www.oecd.org/sti/1888451.pdf

Piedra Salomón, Y., & Ponjuán Dante, G. (2021). Análisis de los patrones de colaboración del Programa de Doctorado en Documentación e Información Científica. Revista Cubana de Información en Ciencias de la Salud, 32(1).

Rodríguez Zoya, L., & Roggero, P. (2014). La modelización y simulación computacional como metodología de investigación social. Polis, Revista Latinoamericana, 417-440.

Rosado Gomez, A. A. (01 de 04 de 2010). Inteligencia de Negocios. (U. T. Pereira, Ed.) Scientict Technica AÑO.VVI. No. 44, 44, 321-326. Obtenido de http://www.redalyc.org/html/849/84917316060/ Salas, R. (2004). Redes Neuronales Artificiales. Valparaiso, Argentina: Departamento de Computación.

Sepúlveda López, J. J. (2018). Perspectiva social del fenómeno de la inclusión digital: una aproximación desde la complejidad. Manizales: Universidad Nacional de Colombia.

Thompson, K., Jaeger, P., Greene, N., Subramaniam, M., & Bertot, J. (2014). Digital Literacy and Digital Inclusion - Information Policy and the Public Library. Rowman & Littkefield.