Arquitectura web con Big data para mejorar la toma de decisiones y reducción de contagios de COVID en Kennedy, Bogotá, Colombia.

Darin J. Mosquera Palacios[1], Juan Eduardo Moreno Villamil[1] y David Gutierrez Alvis[1]
(1) Facultad Tecnológica, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia. (correo-e: djmosquerap@correo.udistrital.edu.co, jemorenov@correo.udistrital.edu.co, dsgutierreza@correo.udistrital.edu.co)

Resumen

En este artículo se desarrolla una arquitectura web con Big data clasificando la información para la toma de decisiones buscando la reducción de contagios de COVID en la localidad de Kennedy, Bogotá - Colombia, se realiza un acercamiento de la afectación que ha dejado la pandemia (COVID -19) aplicando Big data. El método utilizado consiste en que por medio de la Big data y acoplado a una arquitectura está analice los datos arrojados por la alcaldía de Bogotá en la cual se brinda a los ciudadanos la posibilidad de acceder a mapas de calor que muestran la concentración de casos de COVID-19 en diferentes zonas de la ciudad, además de ofrecer información actualizada sobre el número diario de pruebas tomadas y procesadas, el porcentaje de ocupación de los servicios de hospitalización general, unidad de cuidado intermedio y UCI, así como la disponibilidad de camas en las áreas de UCI y hospitalización en Bogotá y también que proporciona un archivo con los datos de cada localidad, incluida Kennedy, buscando como objetivo analizar todos estos datos brindando herramientas por medio de la tecnología como Big data en la toma de decisiones ante el COVID o una nueva pandemia. Como conclusión y muy importante en tener en cuenta es que la información no es exacta, ya que solo es la alcaldía la entidad la encargada de recopilar la información de toda la ciudad teniendo margen de error.

Palabras clave: Arquitectura Web, Big Data, Covid 19, Contagios, Pandemia

Abstract

This article develops a web architecture with big data classifying information for decision making seeking the reduction of covid infections in the town of Kennedy, Bogotá - Colombia, an approach of the impact that has left the pandemic (covid -19) applying big data is made. the method used consists of using big data and coupled to an architecture that analyzes the data provided by the mayor's office of Bogota, which offers citizens the possibility of accessing heat maps showing the concentration of COVID-19 cases in different areas of the city, in addition to providing updated information on the daily number of tests taken and processed, the percentage of occupancy of general hospitalization services, intermediate care unit and ICU, as well as the availability of beds in ICU and hospitalization areas in Bogota and also that provides an archive with data from each locality, including Kennedy, seeking as an objective to analyze all these data providing tools through technology such as big data in decision making in the face of covid or a new pandemic. As a conclusion and very important to keep in mind is that the information is not accurate, since only the mayor's office is the entity in charge

of collecting the information of the entire city having a margin of error.

Keywords: Web Architecture, Big Data, Covid 19, Infections, Pandemic, Pandemic.

INTRODUCCIÓN

Hace un siglo, la humanidad sufrió una de las mayores catástrofes sanitarias, la pandemia de influenza, que cobró millones de vidas y dejó una profunda huella en la historia. Sin embargo, nos enfrentamos nuevamente a una crisis sin precedentes: la pandemia de COVID-19. Esta emergencia de salud global ha sacudido al mundo entero, impactando a todos los niveles y afectando significativamente a Colombia. Las consecuencias económicas y sociales han sido devastadoras, con una interrupción masiva de actividades comerciales y laborales, así como la implementación de restricciones para contener la propagación del virus, [1]. Posteriormente el 6 de marzo de 2020, Colombia reportó su primer caso de Covid-19, y hasta el momento,[2] en los tres picos de la pandemia se han registrado más de 3 millones de contagios confirmados y más de 80 mil fallecimientos.

Cada uno de los momentos de mayor incidencia de la enfermedad presenta estadísticas que reflejan el número de personas que han contraído el virus, las víctimas fatales y el total de afectados. Es de suma importancia analizar la magnitud del impacto que cada pico ha tenido en el país. En la ciudad de Bogotá de las 19 localidades con las que cuenta, las más afectadas fueron las de estrato 1,2,3, en donde sobresale Kennedy, según la alcaldía de Bogotá (2022) la localidad con mayores habitantes, esta población es una de las más afectadas ya qué en pandemia debieron desplazarse a sus lugares de trabajo presencial, lo cual generó agudizar el número de contagios en la localidad, siendo una de las localidades que más habitantes tiene en total son 1.034.838 personas siendo la segunda en Bogotá, por debajo de Suba.

Este proyecto busca proponer la arquitectura para un sitio web que se alimentará de datos extraídos de fuentes gubernamentales, asegurando que la información sea verídica y de alta certeza, para organizar y mostrar lo más relevante. La arquitectura permitirá que los usuarios consulten acerca de la pandemia del COVID – 19, más exactamente de la ciudad de Bogotá y específicamente de la afectación que esta tuvo en la localidad de Kennedy, mostrando una herramienta tecnológica buscando la toma de decisiones de acuerdo a los datos analizados por la Big Data, el articulo está organizado de la siguiente manera introducción , metodología, diagrama de flujo, diagrama de la arquitectura, resultados y discusión, conclusiones y bibliografía.

METODOLOGÍA

Los sistemas más usados para el almacenamiento de la información son las bases de datos, particularmente las bases de datos relacionales de [3]. Pero hoy día se han valorado mucho las de tipo NoSQL, especialmente en Big Data, ya que no son tan restrictivas como SQL [4] en donde es necesario procesar grandes datos y tener flexibilidad a la hora de anexar nuevos tipos de datos y estructuras.
Al igual que existe NoSQL, el mercado ha presentado otro tipo de alternativas para almacenar los datos, como Hadoop, que usa HDFS, que resulta ser un sistema de archivos altamente tolerante a fallos, escalable y con una arquitectura distribuida [8]. La mayor ventaja de usar este sistema en Big Data es que se puede distribuir la información en diferentes servidores con facilidad, algo que resulta conveniente cuando el volumen de información crece constantemente.

Considerando lo anterior, se puede determinar con facilidad que el sistema más adecuado para llevar a cabo procesos de almacenamiento de Big Data sobre la evolución del COVID-19 en la Localidad de Kennedy en Bogotá, es Hadoop HDFS. Esto debido al hecho que existen cada vez más datos, lo que implica que se debe tener una plataforma fácilmente escalable para volúmenes de datos cada vez mayores.

Si bien existen varias alternativas para realizar procesamiento de datos, se ha tomado la iniciativa de elegir Hadoop Mapreduce que es un marco de software creado con el fin de hacer aplicaciones que puedan procesar grandes cantidades de datos de forma paralela, en un mismo hardware según [5]. Al igual que el sistema de almacenamiento elegido (HDFS) [6] Mapreduce viene integrado en Hadoop, lo que proporciona un ambiente integral y uniforme bajo la misma plataforma, lo que reduce la cantidad de errores proveniente a incompatibilidades o configuraciones ocasionados por sistemas que unen diferentes tecnologías, ya que ambos forman parte de Hadoop.

La infraestructura utilizada es importante para que el sistema de Big Data se desempeñe eficientemente. Si bien existen las alternativas on-premise y Cloud, la primera no es fácilmente escalable en el tiempo, por lo tanto, lo mejor es elegir un excelente servicio en la nube, [7] 3 que las lideran, Microsoft con su servicio de Azure, Google con Google Cloud y Amazon con AWS, las tres empresas ofrecen servicios similares y de alta calidad, por ello la decisión se ha definido según quien lidera el mercado. En este caso, vemos que AWS es el líder actual, y que constantemente está mejorando sus servicios, a pesar de que Microsoft y Google hacen su mejor intento por ser los mejores por lo cual la mejor elección es Amazon con AWS.”

HTML es un sistema de etiquetas para el desarrollo de la parte más básica de una web, definiendo el significado y la estructura del contenido en una web [8]**.**. CSS, por otro lado, es un lenguaje de estilos utilizado para definir la apariencia de documentos HTML [9] y JavaScript es reconocido principalmente como el lenguaje utilizado para

programar páginas web según [10]. Considerando lo anterior, se ha determinado que el desarrollo de la web se lleve a cabo usando estas tres tecnologías.

Considerando los puntos planteados de los autores [11] en donde se refleja la desproporcionada relación entre fuentes de noticias falsas y verídicas y, teniendo en cuenta que para los procesos de Big Data es importante tener datos confiables [12], se ha optado por tomar información proveniente de fuentes oficiales gubernamentales que proporcionen archivos con datos sobre la evolución del COVID-19 en la localidad de Kennedy en Bogotá.

Debido a que la página web Saludata (https://saludata.saludcapital.gov.co), proporcionada por la Secretaría de Salud de Bogotá, brinda a los ciudadanos la posibilidad de acceder a mapas de calor que muestran la concentración de casos de COVID-19 en diferentes zonas de la ciudad, además de ofrecer información actualizada sobre el número diario de pruebas tomadas y procesadas, el porcentaje de ocupación de los servicios de hospitalización general, unidad de cuidado intermedio y UCI, así como la disponibilidad de camas en las áreas de UCI y hospitalización en Bogotá y también que proporciona un archivo con los datos de cada localidad, incluida Kennedy y con un tiempo de actualización bastante aceptable se consideró la candidata idónea como la fuente oficial a usar en este proyecto, pero al solo proporcionar un archivo con los datos, se descartó por no ser óptimo para la automatización del proceso.

Datos Abiertos (https://www.datos.gov.co), una plataforma encargada de publicar datos abiertos, acogiendo los lineamientos establecidos por MinTIC y dando cumplimiento con lo establecido en la Ley 1712 de 2014 [13] es la opción seleccionada debido a que, proporciona los mismos datos que se pueden encontrar en Saludata, pero en este caso ofrece un sistema API con lo cual se pueden realizar consultas de manera automatizada al sistema.

Cabe destacar que según el DANE (2022) (Departamento Administrativo Nacional de Estadística), cuyo trabajo es ser la entidad que se encarga de producir y también comunicar información estadística a nivel nacional, contiene múltiples archivos que contienen datos sobre el COVID-19 a nivel Colombia, pero no datos específicos de la localidad de Kennedy en Bogotá.

En la figura 1 se muestra el diagrama de flujo con el paso a paso de cómo se trabajó la información obtenida en donde se definieron las variables iniciales para organizar la información, manejarla de forma fácil y ágil, las variables utilizadas son (sexo, edad, año, localidad, Estado (recuperado o fallecido), con ello se obtiene una

Información Inicial en donde después de obtener los datos de los contagios a nivel Bogotá por medio de varias fuentes se clasifica la información con las variables iniciales.

Clasificación que ayuda a obtener la información total de la ciudad de Bogotá por localidades, separando la data por medio de la big data y aplicando los filtros de las variables iniciales, después de aplicar big data se obtiene la información deseada de

la localidad de Kennedy, en caso de que no se aplicará dicha tecnología esta información se almacenará en una base de datos.

luego de aplicar Big data se obtiene una web siendo la forma ideal de mostrar los datos obtenidos y luego de realizar el proceso con Big data por medio de Power bi poder interactuar con gráficas, con ello tomar decisiones. toda la información se almacena en una base de datos en donde se muestran los contagios de Covid en Bogotá, teniendo un respaldo en caso de necesitar nuevas consultas de información.

DIAGRAMA DE FLUJO

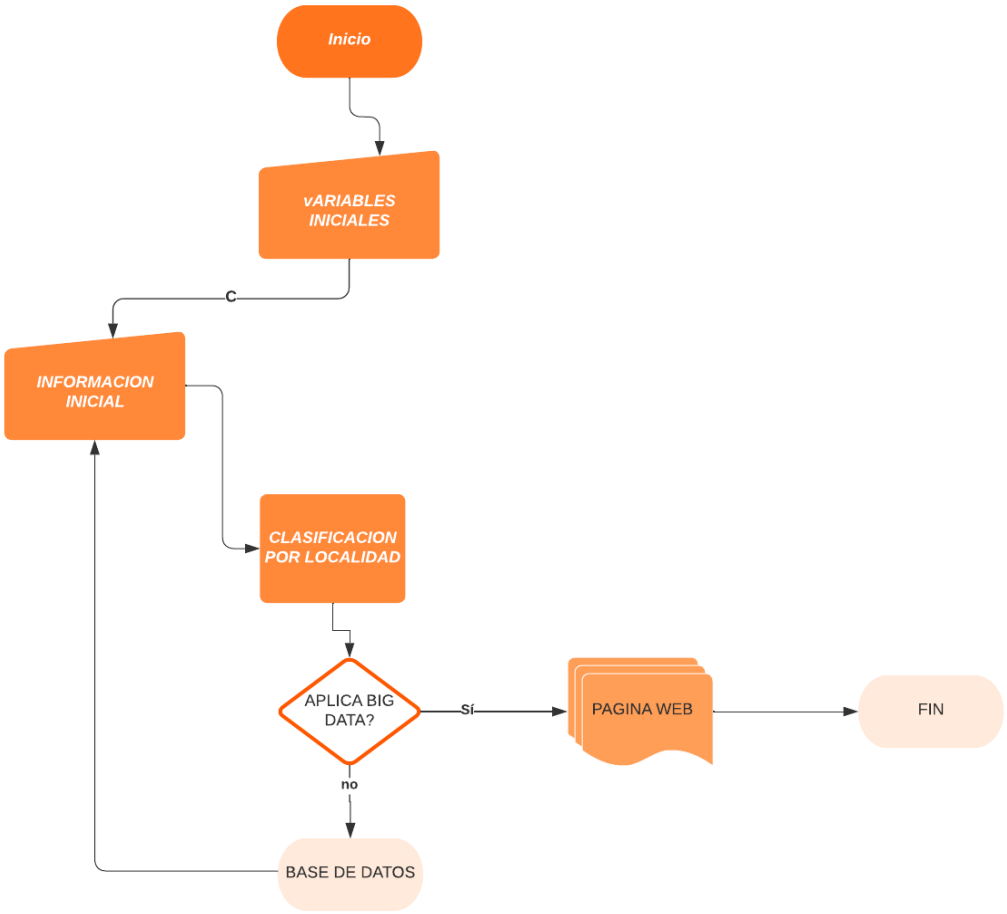


FIGURA 1 : DIAGRAMA DE FLUJO DE LA METODOLOGÍA PROPUESTA

En la Figura 2 se muestra a los usuarios, la información se almacena y se organiza en un servidor web y por medio de una base de datos, esta información es administrada vía internet por las entidades gubernamentales, quienes son finalmente que le dan el sentido a dicha data, luego esta información es entregada a los usuarios de forma clara y concisa, con el objetivo que las personas se encuentren informadas y al día con lo que está pasando en su comunidad. Dicha arquitectura se divide de la siguiente manera:

Usuarios: son los que utilizan la información recolectada y debidamente organizada para su buen uso y toma de decisiones.

Entidades: Son las webs usadas para obtener información verídica sobre la evolución del COVID-19 en la localidad de Kennedy en Bogotá. En este caso se usará la plataforma de Datos Abiertos que posee un sistema API para obtener los datos de forma automatizada.

Son organismos gubernamentales que se encarga de acoplar las información de los contagiados por el COVID en la ciudad de Bogotá, como por ejemplo Saludata, la Alcaldía mayor de Bogotá, DANE etc.

Nube: Es donde toda la información recolectada por las entidades se digitaliza y se sube a la internet para acceso público y fácil acceso.

Bases de datos: Corresponde al sistema que se encarga de tomar los datos, procesarlos y almacenarlos. En este proyecto se hace uso de Hadoop con el sistema HDFS para almacenamiento y Mapreduce para procesamiento de datos, debido a que es un sistema homogéneo donde se pueden realizar ambos pasos, evitando problemas de compatibilidad al unir diferentes tecnologías. Adicional, es un sistema que proporciona fácil escalabilidad, lo que es ideal cuando los datos aumentan su volumen.

Posteriormente al ser una gran cantidad de datos se almacenan en bases de datos para salvaguardar la integridad de la información a posteriori.

Servidor web: Es aquí donde los usuarios pueden hacer las solicitudes de consulta de la información obteniendo como respuesta por parte del servidor los contagiados en la ciudad de Bogotá, página web encargada de mostrar al usuario el resultado obtenido al realizar el procesamiento de datos. Dicha data es extraída en este caso de HDFS, ya que JavaScript posee una librería que lo permite.

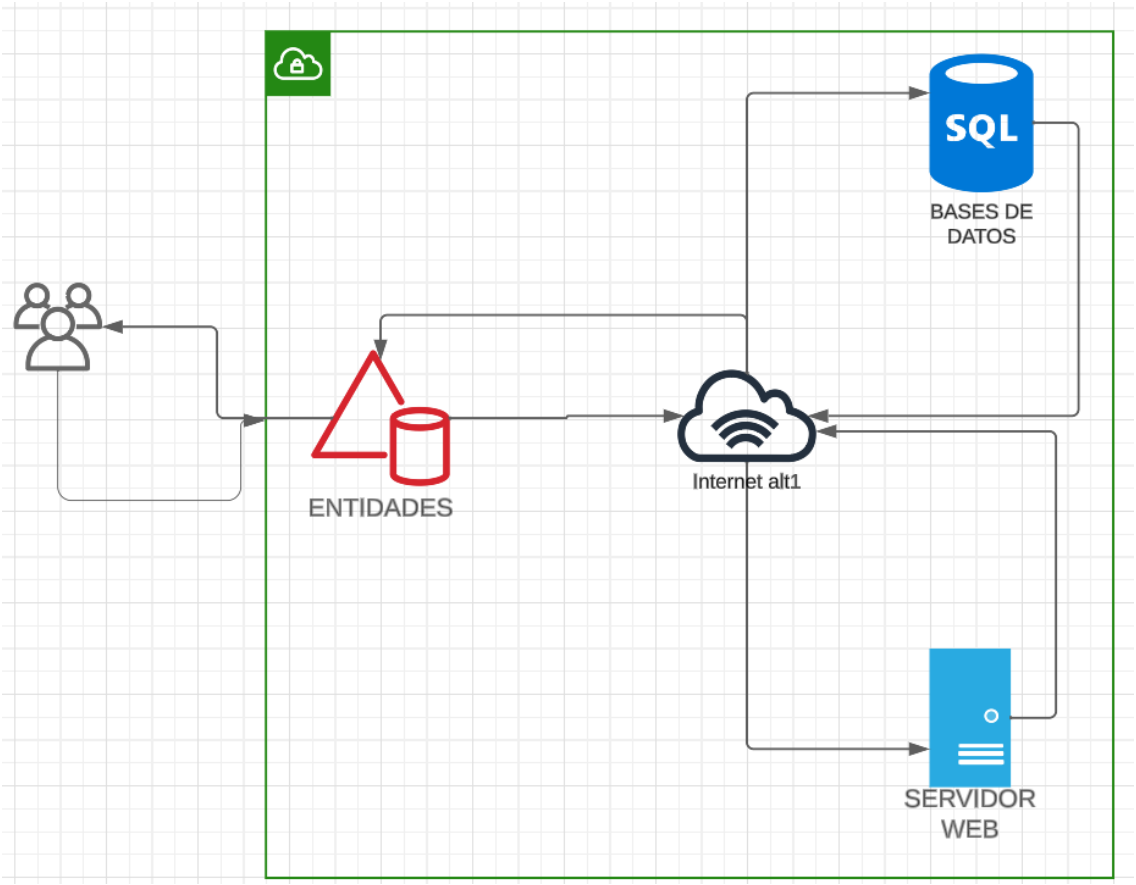


FIGURA 2 : DIAGRAMA DE ARQUITECTURA PROPUESTA

RESULTADOS

En Colombia el número de contagiados por COVID asciende a los 6.369.916 millones, en donde la capital Bogotá cuenta con 1.883.181 siendo Kennedy unas de las localidades más grandes por longitud y por habitantes, donde el número de contagiados por COVID asciende 225,513, en la figura 3 se describe de forma detallada año por año, el número de contagios, los recuperados, fallecidos, discriminados por sexo y por edad, desde 0 años en adelante.

COVID EN KENNEDY																	
AÑO	CONTAGIOS	RECUPERADOS	0-20 AÑOS	MASCULINO	FEMENINO	21-40 AÑOS	MASCULINO	FEMENINO	41-60 AÑOS	MASCULINO	FEMENINO	61-80 AÑOS	MASCULINO	FEMENINO	81 EN ADELANTE	MASCULINO	FEMENINO
2020	63753	62035	8333	4085	4248	28583	13256	15327	18618	8642	9976	5862	2698	3164	639	248	391
2021	126974	124299	18585	8915	9670	56221	25356	30865	37391	16539	20852	11050	4722	6328	1052	400	652
2022	34085	33501	4332	2123	2209	14232	5509	8723	9942	3930	6012	4374	1810	2564	621	276	345
2023	701	672	153	79	74	194	68	126	167	59	108	121	57	64	37	13	24
AÑO	CONTAGIOS	FALLECIDOS	0-20 AÑOS	MASCULINO	FEMENINO	21-40 AÑOS	MASCULINO	FEMENINO	41-60 AÑOS	MASCULINO	FEMENINO	61-80 AÑOS	MASCULINO	FEMENINO	81 EN ADELANTE	MASCULINO	FEMENINO
2020	63753	1451	2	1	1	52	20	32	325	244	81	758	518	240	314	142	172
2021	126974	2316	3	1	2	146	40	106	703	231	472	1051	625	426	413	214	199
2022	34085	249	1	0	1	8	5	3	19	12	7	83	51	32	138	68	70
2023	701	7	0	0	0	0	0	0	1	1	0	2	1	1	4	2	2

FIGURA 3 : TABLA DATOS COVID KENNEDY FUENTE PROPIA

Las figuras 4,5 y 6 que se muestran en a continuación son desarrolladas en Power bi, analizadas con Big data y recopiladas con Web Scraping, en donde se definieron unas variables como lo son año, estado (recuperados, fallecidos) edad, sexo y localidades y por medio de filtros se puede jugar con la información teniendo datos exactos de la localidad de Kennedy.

En la figura 4 Se filtra por la localidad de Kennedy en donde se muestra el total de casos contagiados en los años 2020,2021,2022 y 2023, el cual asciende a los 225.513 contagiados de los cuales el 45% son hombres y el 55 % son mujeres, de los cuales más del 90 % de los contagiados están recuperados y un poco más del 8 % fallecieron, se evidencia que el mayor número de contagiados están entre 21 y 40 años de edad, posteriormente le sigue las personas de 41 a 60 años de edad, siendo los adultos mayores de 81 en adelante lo menos contagiados.

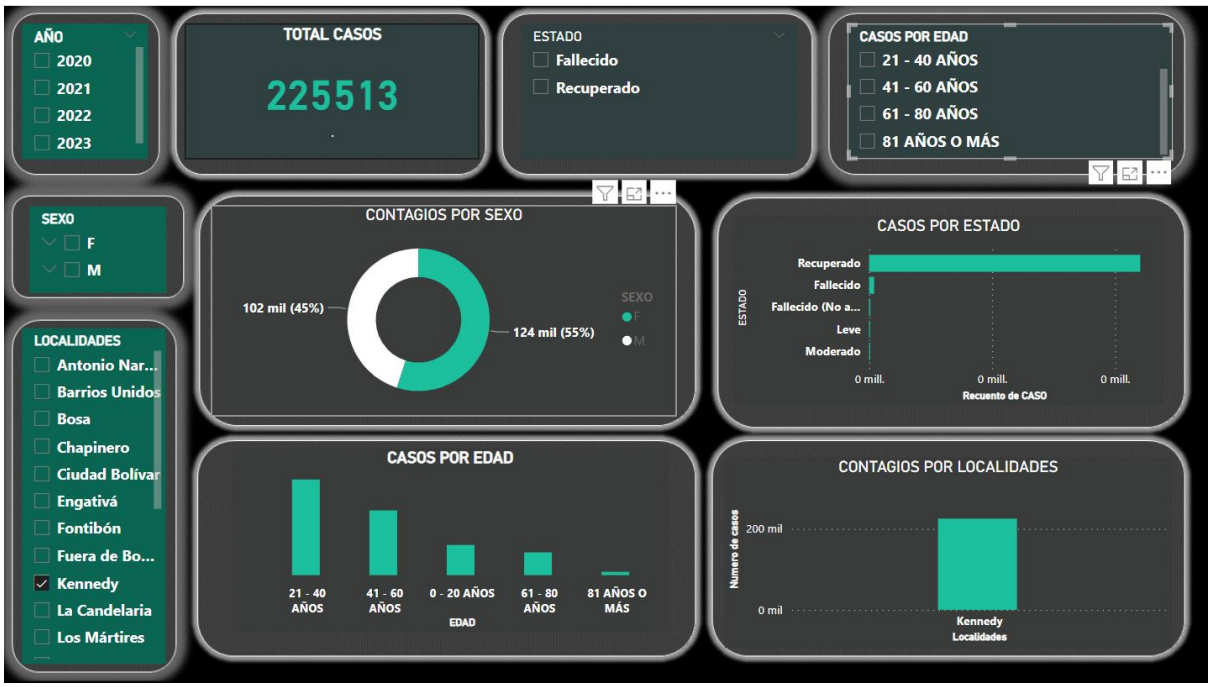


FIGURA 4 : POWER BI FUENTE PROPIA

En la figura 5 se describe por estado (recuperado o fallecido) en donde se selecciona la opción de fallecidos mostrando que en la localidad de Kennedy en todos los años de pandemia desde el 2020 hasta el 2023 en donde se evidencia que fallecieron 4023 personas de las cuales el 37 % son mujeres y el 63 % son hombres siendo el rango de edad de 61 a 80 años el más afectado, le sigue las personas de 41 a 60 años de edad siendo los 0 a 20 años con menos porcentaje de personas fallecidas a causa del COVID en la localidad de Kennedy.

.



FIGURA 5 : POWER BI FUENTE PROPIA

Por último en la figura 6 se filtra las gráfica por estado se selecciona la opción de recuperados, mostrando el número de personas recuperadas en la localidad de Kennedy desde el 2020 a 2023 en donde se muestra que el total de recuperados asciende a los 220.507 personas de las cuales el 45 % son Hombres y el 55 % son mujeres, siendo el rango de edad de 20 a 41 años con más personas recuperadas, luego le sigue las personas que están dentro del rango de edad de 42 a 60 años, siendo las personas de 81 años en adelante las que menos se han recuperado.

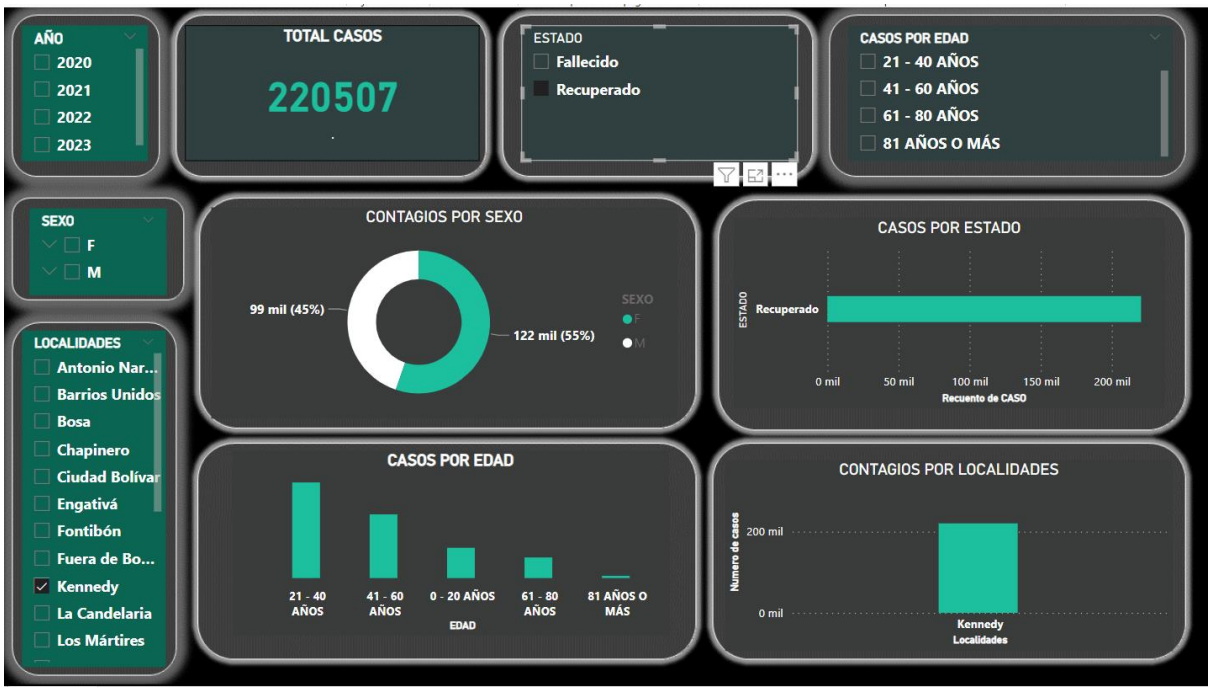


FIGURA 6 : POWER BI FUENTE PROPIA

DISCUSIÓN

¿Por qué la localidad de Kennedy y no otra localidad de la ciudad?

Se escoge la localidad de Kennedy por varias razones la principal es por su estrato social ya que cuenta con estratos 1, 2 y 3, siendo la mayoría de sus habitantes trabajadores de empresas o trabajo informal, teniendo que salir a la calle en la pandemia, además de eso por ser una de las localidades más grandes y habitadas de Bogotá con 1.034.838 personas.

con esta investigación se busca que las entidades gubernamentales o cualquier persona al leer el artículo llegue a una serie de conclusiones analizando las gráficas y lo ocurrido en esta localización por ejemplo el mayor porcentaje de recuperados en la localidad de Kennedy son mujeres y los que más falleció eran los hombres siendo ellos la cabeza de hogar y la gran mayoría únicos trabajadores en el hogar, en donde el mayor número de contagiados están en la edad de 20 a 40 años pero el rango de edad donde más falleció era de 41 a 60 años.

por esta deducción y muchas otras era necesaria esta investigación por si existe una nueva pandemia implementar o fortalecer las políticas de salud por parte de la alcaldía de la localidad y de la ciudad para que el número de fallecidos y contagiados sea un porcentaje mucho menor a lo ocurrido con el COVID 19, una de esas políticas es acordar con las empresas el 100 del teletrabajo antes de que la pandemia coja fuerza, con eso evitar que las personas salgan a la calle y se contagien.

Comparación de la localidad de Kennedy con las demás

En la figura 7 se muestra el porcentaje de contagiados por localidad en donde se muestra en las localidades más golpeadas por la pandemia son Suba, Kennedy y Engativá, siendo unas de las localidades más grandes y con más habitantes, en algo se parecen y es que su mayoría son estratos 1,2 y 3, trabajadores de empresa o informales la gran mayoría.

siendo las localidades de chapinero, Barrios unidos, Teusaquillo las localidades menos golpeadas por el COVID 19, siendo las localidades de mejor economía, los barrios más ricos de la ciudad se encuentran en estas localidades.

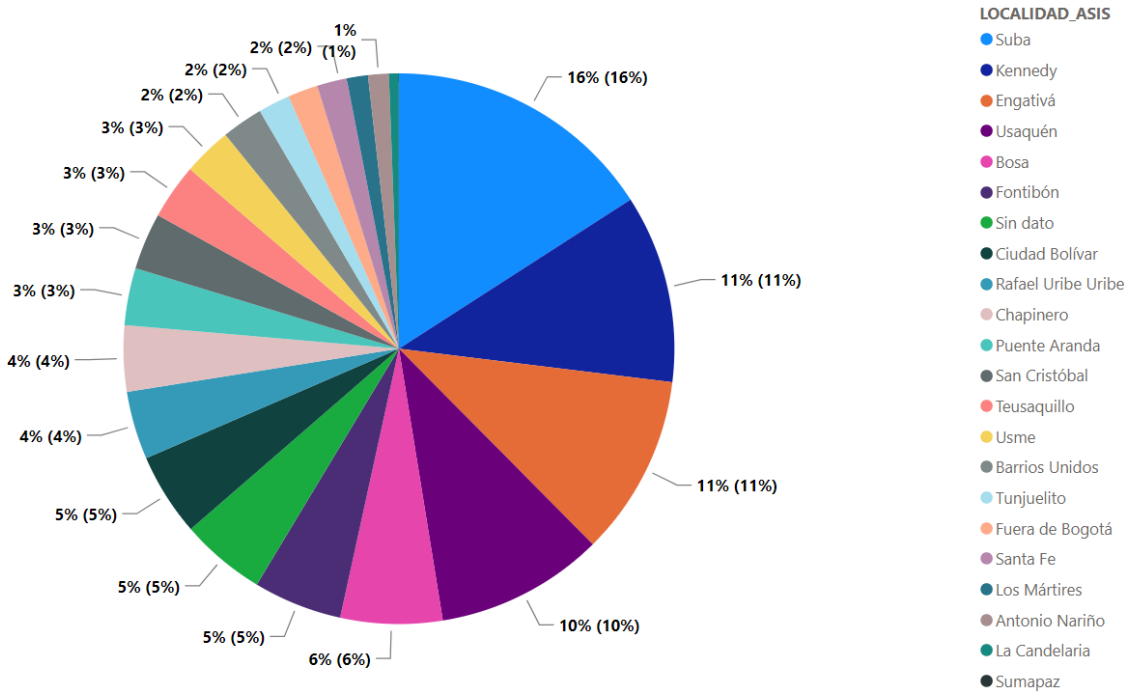


Figura 7 **POWER BI FUENTE PROPIA**

CONCLUSIONES

Después de la investigación se llega a las siguientes conclusiones para atender este tipo de problemática de manera eficiente, se ha hecho uso de la tecnología más reciente, por ejemplo aplicaciones web y móviles que muestran datos sobre la evolución de COVID-19 en el mundo y a nivel local. Pero adicional, hay otra tecnología que ayuda en estos casos de contingencia, y es el Big Data**.**

Sin embargo, aplicar Big Data se convierte en todo un reto debido a que se necesitan datos fidedignos, y lastimosamente la pandemia de COVID-19 trajo consigo otra, la de desinformación proveniente de noticias falsas. Considerando lo anterior, para aplicar de manera efectiva un sistema de Big Data, lo ideal es obtener datos confiables, de calidad.

Por esa misma razón, se ha aprovechado los sistemas oficiales de datos que ha creado el gobierno colombiano para poder dar solución a estos problemas. La arquitectura de Big Data que se ha propuesto en este proyecto busca brindar

información veraz a las personas de la localidad de Kennedy, en Bogotá, para que se puedan tomar decisiones objetivas basadas en datos, y con ellos mejorar las condiciones de vida de las personas a pesar de estar en una contingencia.

BIBLIOGRAFÍA

[1] H. J. de Matos, “A próxima pandemia: estamos preparados?”, *Rev Panamazonica Saude*, vol. 9, núm. 3, sep. 2018, doi: 10.5123/s2176-62232018000300001.

[2] M. X. Arias-Barahona y J. J. Vanegas-Giraldo, “Comparison of the probability distributions in Covid-19 deaths in three peaks of the pandemic in Colombia”, *Rev Acad Colomb Cienc Exactas Fis Nat*, vol. 45, núm. 177, pp. 971–979, oct. 2021, doi: 10.18257/raccefyn.1474.

[3] D. Adams *et al.*, “Oracle® Database Database Concepts 19c Oracle Database Database Concepts, 19c”, 2021.

[4] J. V. Craig S. Mullins, “What is NoSQL and How do NoSQL Databases Work_”, *https://www.techtarget.com/searchdatamanagement/definition/NoSQL-Not-Only-SQL*.

[5] Paula Cuéllar de la Vega y D. Kennedy, “Los retos de la informalidad en el marco de la emergencia sanitaria por COVID-19 en”, 2021.

[6] A. Hernández Dominguez y A. H. Yeja, “Acerca de la aplicación de MapReduce + Hadoop en el tratamiento de Big Data About MapReduce + Hadoop application in the treatment of Big Data”, *Revista Cubana de Ciencias Informáticas*, vol. 9, núm. 3, 2015, [En línea]. Disponible en: http://rcci.uci.cuPág.49-62

[7] ARANTXA HERRANZ, “Así se reparte el mercado cloud_ los tres grandes tienen el 63”, *xakata*, 2021.

[8] Mdn web docs, “HTML_ Lenguaje de etiquetas de hipertexto _ MDN”, 2023.

[9] Mdm web docs y 2023, “CSS_ Cascading Style Sheets _ MDN”, 2023.

[10] Mdm web docs, “Acerca de JavaScript _ MDN”, 2023.

[11] L. M. Leal Villamizar y S. M. Torres Quiroga, “21(42) • Enero-junio de 2023 •a12”.

[12] R. Rawat y R. Yadav, “Big Data: Big data analysis, issues and challenges and technologies”, en *IOP Conference Series: Materials Science and Engineering*, IOP Publishing Ltd, ene. 2021. doi: 10.1088/1757-899X/1022/1/012014.

[13] J. Daniel *et al.*, “DANE-Información para todos 2 3 DEPARTAMENTO ADMINISTRATIVO NACIONAL DE ESTADÍSTICA-DANE Coordinación de la serie notas estadísticas Elaboración y revisión de la nota Natalia Borrás Torres Sandra Arenas Pérez Jhon Quinchua (coordinador)”.