

Análisis de clustering para entender la vulnerabilidad poblacional relacionada con el COVID-19

Clustering analysis to understand population vulnerability related to COVID-19

Jose David Sánchez Grajales



Daniel Elías Cuartas Arroyave



Víctor Andrés Bucheli Guerrero



Universidad del Valle, Colombia

OPEN ACCESS

Recibido: 21/02/2022

Aceptado: 17/05/2022

Publicado: 06/07/2022

Correspondencia de autores:

jose.d.sanchez@correounivalle.edu.co



Copyright 2020
by Investigación e
Innovación en Ingenierías

Resumen

Objetivo: El objetivo del estudio es identificar grupos poblacionales en la ciudad de Cali-Colombia, cuya vulnerabilidad poblacional dependen de sus características sociodemográficas y socioeconómicas. La hipótesis de este trabajo es que el crecimiento de los casos de COVID-19 es mayor para los grupos poblacionales más vulnerables. **Metodología:** Para este trabajo se integraron datos de casos positivos de COVID-19 para las comunas de la ciudad de Cali-Colombia y variables que representan la vulnerabilidad poblaciones en términos sociodemográficos y socioeconómicos. Así, se desarrolló un análisis de clustering jerárquico, para el periodo marzo-junio 2020. **Resultados:** Los resultados evidencian que existen grupos poblacionales más vulnerables que otros en la ciudad de Cali, debido a las dinámicas propias de la población. De esta forma, la dinámica de la pandemia por COVID-19 impacta de diferente manera a cada grupo poblacional de acuerdo con sus propias características. Además, algunos de los grupos obtenidos forman corredores geográficos, donde las comunas de un mismo grupo están ubicadas contiguamente. **Conclusiones:** Los resultados fueron corroborados con las curvas de crecimiento de los casos para cada uno de los grupos obtenidos y estos son consistentes. Estos resultados permiten diseñar estrategias diferentes de mitigación y prevención de COVID-19 por sectores, zonas o por grupos de comunas específicas y no estrategias generales para toda la ciudad. Adicionalmente, el diseño de estrategias basadas en datos que se pueden actualizar en el tiempo y permiten actualizar las estrategias de acuerdo con la dinámica de la pandemia..

Palabras clave: Clustering, CRISP-DM, vulnerabilidad, COVID-19, comunas, Cali, SIGELO.

Abstract

Objective: The objective of the study is to identify population clusters in the city of Cali, Colombia, whose population vulnerabilities depend on its sociodemographic and socioeconomic characteristics. The hypothesis of this work is that the growth of the cases of COVID-19 is greater for the most vulnerable populations. **Methodology:** For this work, it was integrated data of positive cases of COVID-19 for the communities of the city of Cali, Colombia and variables that represent the vulnerability of populations in sociodemographic and socioeconomic terms. Like this, an analysis of hierarchical clustering was developed for the period of March - June 2020. **Results:** The results show that there exist groups of populations more vulnerable than others in the city of Cali, due to its own dynamics of the population. In this form, the dynamic of the COVID-19 pandemy, impact differently each populational group based on its own characteristics. Furthermore, some of the obtained groups form geographical corridors, where the communes of the same group are located contiguously. **Conclusions:** The results were corroborated with curves of growth of the cases for each group obtained and these were consistent. These results allow to design different strategies of mitigation and prevention of COVID-19 by sectors, zones, or by groups of specific communes and no general strategies for the whole city. Additionally, the design of strategies based on data that can be actualized at the time and permit the actualization of strategies based on the dynamic of the pandemy.

Keywords: Clustering, CRISP-DM, vulnerability, COVID-19, communes, Cali, SIGELO.

Introducción

Para el contexto colombiano, el primer caso de COVID-19 se detectó el 6 de marzo de 2020 en la ciudad de Bogotá, posteriormente, el 25 de marzo del mismo año se presentó el primer caso de COVID-19 en la ciudad de Cali y a fecha de 24 de noviembre de 2020, en Cali, existían 74.281 casos positivos acumulados de COVID-19. Sin embargo, por lo general se han implementado estrategias comunes para toda la población, esto con el fin de mitigar los efectos tanto de salud, económicos, y otros, a raíz de la pandemia.

La Oficina de las Naciones Unidas para la Reducción del Riesgo de Desastres [1] define la vulnerabilidad como, las características y las circunstancias de una comunidad, sistema o bien que los hacen susceptibles a los efectos dañinos de una amenaza. Para este trabajo los niveles de vulnerabilidad de los habitantes se entienden como el efecto potencial del COVID-19 sobre el carácter histórico y biosocial, que se expresa en proceso de la fragilidad de las personas o grupos en términos de morbilidades, condiciones sociales, culturales y económicas que influyen en su capacidad para anticipar, hacer frente, resistir y recuperarse de los efectos adversos de la pandemia por COVID-19 [2].

En el ámbito nacional se propusieron algunas estrategias para distintos momentos de la pandemia, que incluyeron diferentes tipos de medidas, las cuales incluyen el aislamiento preventivo obligatorio, reapertura por sectores económicos y aislamiento selectivo. Algunas medidas fueron el aislamiento obligatorio para los adultos mayores de 60 años, el cierre de la mayoría de los sectores económicos con algunas excepciones, uso obligatorio de tapabocas, protocolos de bioseguridad en establecimientos públicos, restricciones de movilidad por pico y cédula, entre otros. Todas estas medidas fueron generalizadas para la población, sin tener en cuenta las dinámicas sociodemográficas y socioeconómicas de cada una de ellas.

Los gobiernos locales, específicamente la alcaldía de Cali, enfocaron la mayoría de las estrategias de control a la ciudadanía general, pero también, en algunas ocasiones se implementaron medidas por comunas o para ciertos sectores económicos. Por ejemplo, en junio de 2020, la plaza de mercado del sector Santa Elena fue cerrada como medida preventiva debido a que presentaba un crecimiento acelerado de casos de contagio, otra medida fue, decretar toques de queda algunos fines de semana para ciertas comunas basados principalmente en el aumento de casos de COVID-19 [3]. Como se nota en estos ejemplos, las medidas por sectores o por grupos poblacionales son alternativas para afrontar los diferentes retos de la pandemia, las cuales tuvieron en cuenta únicamente el comportamiento del número de casos de COVID-19.

En el marco de la pandemia por COVID-19, la Universidad del Valle desarrolló un proyecto denominado SIGELO, Sistema de gestión logística local en escenarios de riesgo de pandemia como el COVID-19, cuyo objetivo era crear un sistema enfocado a apoyar la toma de decisiones mediante la inteligencia artificial para manejo del riesgo en diferentes aspectos como accesibilidad, vulnerabilidad, ayudas económicas, entre otras, dirigido a gobernantes, empresas y ciudadanos. Consecuentemente, se realizó el estudio de vulnerabilidad que se describe en este artículo.

Como respuesta a la gestión de la pandemia, es necesario encontrar estrategias y medidas para aplanar la curva de contagios y la reactivación de los diferentes sectores económicos como el comercio, transporte, educación, etc [4, 5, 6, 7, 8, 9]. En este trabajo, se estudian características poblacionales de vulnerabilidad y contagios de COVID-19, para brindar soporte a estas medidas sectoriales, ya que aporta información, patrones y herramientas para diseñar e implementar estrategias inteligentes de acuerdo con las características de la población y el comportamiento de la pandemia, de esta forma se pueden crear barreras de contagio y minimizar el impacto económico en el escenario de riesgo COVID-19 en la ciudad de Cali. En resumen, se evidencia que las estrategias implementadas impactan de diferentes maneras a la

población de acuerdo con sus características intrínsecas. Así, la vulnerabilidad, tal como fue definida en líneas anteriores, permite diseñar estrategias adaptadas a las características propias de una población determinada, estas estrategias podrían tener un impacto en la economía al buscar formas inteligentes o soportadas en los datos para la toma de decisiones cierre o apertura de zonas específicas de la ciudad.

El paper se organiza de la siguiente manera: en la sección 2 describimos el estado del arte. Sección 3 explicamos la metodología implementada para el análisis de clustering. Luego, en la sección 4 se muestran los resultados obtenidos, describiendo los grupos cualitativa y cuantitativamente. En la sección 5, concluimos el paper realizando una discusión y trabajos futuros.

Estado del arte

En el estado del arte se encontraron diversos estudios de clustering de datos para análisis de casos de COVID-19, donde se identifica dos granularidades, países y regiones.

A nivel de países, [10] muestra varios análisis de clustering por países según la cantidad de casos. Ellos usan los datos publicados por John Hopkins University on COVID-19 [11], el autor construye las series de tiempo de casos activos de COVID-19 y casos activos por tamaño de la población y área, para cada uno de los países a partir el 22 de enero 2020 hasta el 4 abril de 2020. La técnica usada es clustering jerárquico. Los resultados muestran diversas dinámicas de contagio para los grupos de países, esto es, grupos de países donde se evidencia crecimiento acelerado de los contagios de virus con alto número de casos, otros países con crecimiento acelerado pero bajo número de casos comparado con el primer grupo, otros países donde el virus apareció en fechas tempranas pero tiene pocos casos de contagio y por último, un grupo de países donde el virus llevaba varios días pero en la última semana de estudio se disparó el número de contagios. Un análisis de clustering jerárquico adicional es presentado por [12]. Donde se integran diferentes conjuntos de datos públicos, que representan cifras de casos de COVID-19, población, índice de riesgo por COVID-19 y capacidad de salud de los diferentes países. Con ello, aplican preprocesamiento de datos usando PCA y SOM logrando reducir las dimensiones del dataset de 43 variables a 24, luego, muestran dos modelos de clustering aglomerativa para análisis mediante los resultados de PCA y SOM. Los resultados muestran 4 grupos para cada caso, donde existe una correspondencia uno a uno entre los grupos de ambos métodos. Así mismo, [13] presenta un análisis de clustering de casos de COVID-19 para países seleccionados (Estados Unidos, España, Italia, Alemania, Reino Unido, Francia, Iran) mediante el algoritmo Fuzzy C-means. Relacionan los casos confirmados, casos de muertes y tamaño de la población de cada país. Finalmente, los grupos obtenidos indican que España e Italia tienen comportamiento similar en la propagación de COVID-19, mientras Estados Unidos, Alemania Reino Unido, Francia e Irán pertenecen a otro grupo.

A nivel regional, estudios previos de clustering se han realizado en la India para el análisis del comportamiento de COVID-19 en algunos distritos. [14] realiza un análisis de clúster para los 50 distritos de la India con mayor número de casos confirmados por COVID-19. El estudio usa el algoritmo k-means y jerárquico para construir los diferentes grupos, y relaciona las variables de casos confirmados de COVID-19, densidad poblacional y número de hospitales para atención de COVID-19. Para caracterizar los distritos del estado de Maharashtra, [15] implementa clustering jerárquico, mediante dos modelos, uno tomando el conteo de casos confirmados y recuperados, dos, tomando casos de muertes. Para el estudio recopilamos datos de 32 distritos del estado de Maharashtra, India, desde el sitio web "COVID-19 Monitoring Dashboard by Public Health Department, Government of MH [16]. Obteniendo 3 grupos por cada clustering, los resultados permitieron identificar los distritos más impactados por la pandemia según las cifras de COVID-19 presentadas en cada uno. Otras investigaciones a nivel regional se han presentado para otros países, como Italia y Estados Unidos [17,

18, 19]. Ellos usan técnicas de clustering para encontrar características que agrupen comportamientos similares en cada región del país específico, basado en los casos confirmados de COVID-19.

Los estudios de clustering relacionados a COVID-19 encontrados en la literatura tienen granularidad a nivel de países y regiones, los cuales caracteriza el comportamiento de las cifras de casos de COVID-19 en el nivel respectivo (país o región). Por otro lado, no se encuentran estudios que lleguen a granularidad de comuna o barrio. Este nivel de detalle cobra relevancia debido a que gracias a la dinámica propia de la pandemia por COVID-19, las variables locales impactan en el comportamiento de la pandemia [19]. Así, los análisis a nivel de comuna aportan gran valor para los gobiernos locales, brindando herramientas en el apoyo a la toma de decisiones que permitan minimizar el impacto del desarrollo de las actividades económicas, de salud y demás en una ciudad. Decisiones como abrir o cerrar sectores económicos específicos, distribución de ayudas gubernamentales, abastecimiento de elementos de primera necesidad, distribución de atención médica para la población, son algunos ejemplos.

Metodología

La ciudad de Cali se encuentra ubicada en el suroccidente colombiano, constituida por 15 corregimientos en su área rural y 22 comunas en su área urbana, a su vez, cada comuna se divide por una cantidad de barrios y urbanizaciones contiguos geográficamente, clasificados en estratos socio económicos de 1 a 6. Según cifras del Censo Nacional de Población y Vivienda 2018 del DANE, Cali cuenta con 1'822.869 habitantes, de los cuales 53% son mujeres y 43% son hombres. El 16.5% de la población son adultos de la tercera edad mayores de 60 años y el 17.7% son niños y adolescentes entre 0 y 14 años. También, 272.404 mujeres son jefes de hogar, mientras que 339.472 hombres son jefes de hogar. Por último, la tasa de desempleo en Cali para el segundo trimestre de 2020 fue de 27,7% [20].

La comuna más grande de la ciudad de Cali es la comuna 17, que cuenta con 144.520 habitantes, que se distribuyen en 45% hombres y 55% mujeres. Esta comuna también es la que presenta mayor porcentaje de personas adultos mayores de 60 años, correspondiente al 9% del total de adultos mayores de 60 años en la ciudad. Por su parte, la comuna 3 es la que tiene menor población, con 26.775 habitantes, 49% hombres y 51% mujeres. La comuna con menor área es la comuna 12, ubicada en el centro de la ciudad. Además, la comuna con mayor % de empleo informal es la comuna 20 con 87%, y la de menor % empleo informal es la comuna 22 con 68% [21]. Se verifica entonces, que las comunas de Cali son heterogéneas en cuanto a sus características sociodemográficas y socioeconómicas.

Respecto a casos de COVID-19, Cali fue uno de los municipios con mayor número de casos positivos en Colombia. Al finalizar el mes de noviembre 2020, Cali contaba con 76.032 casos confirmados de COVID-19, el 4.1% de ellos estaban activos, 92.9% recuperados y un total de fallecidos de 2.261 personas. Comparando los contagios por edad y sexo se encontró que, del total de casos confirmados en Cali, 49.8% eran personas de sexo masculino y 50.2% femenino, y el rango de edad donde se presentaba mayor contagios es de 30 a 39 años tanto para hombres como mujeres [22].

En este sentido, era necesario constatar si el impacto de la pandemia por COVID-19 afecta en mayor medida a poblaciones con ciertas características sociodemográficas y socioeconómicas, esto se traduce, en definir si existían grupos poblacionales más vulnerables que otros en la ciudad de Cali. Consecuentemente, en el presente trabajo se integraron diferentes fuentes de datos y aplicaron algoritmos de clustering de datos para identificar automáticamente patrones que relacionan la vulnerabilidad poblacional de cada comuna de la ciudad de Cali con el número de casos de contagio por COVID-19 en dichas comunas.

Para el desarrollo del estudio planteado se necesitó, entre otras tareas, capturar, integrar y construir datos desde diferentes fuentes que recopilaron características poblacionales y casos de COVID-19 en Cali [21, 22]. Las características socioeconómicas de la población se tomaron de los datos del Censo Nacional de Población y Vivienda 2018 y los casos de COVID-19 en la Ciudad de Cali se obtuvieron a partir de los datos de boletines epidemiológicos semanales publicados por la Alcaldía de Cali y, por último, se calculó el acumulado de los casos positivos por comuna durante el periodo marzo-junio de 2020.

Las características socioeconómicas y sociodemográficas de una población son amplias. Por lo tanto, para desarrollar el presente trabajo se identificaron aquellas variables que aportaran más información al estudio de vulnerabilidad frente al comportamiento de la pandemia por COVID-19, esto se realizó, construyendo una matriz de correlación entre las variables. Finalmente, las variables seleccionadas son de tipo numérico y se normalizaron con el fin de obtener un conjunto de datos de magnitudes equivalentes.

A partir del conjunto de datos construido se aplicaron técnicas de aprendizaje automático, específicamente clustering de datos, para identificar grupos poblacionales semejantes, a nivel de comuna, en términos de la relación entre sus características y el comportamiento del número de casos positivos de COVID-19, lo que permitió analizar la vulnerabilidad de cada comuna. Se decide utilizar la técnica de clustering jerárquico aglomerativo. Para este estudio, la matriz de distancia se calculó mediante la métrica Euclidiana y se usó como método de optimización la máxima distancia entre elementos de dos grupos diferentes.

Para seleccionar el número de grupos se realiza un corte vertical en el dendrograma, para ello se encuentra la línea vertical con máxima distancia sin una línea horizontal que la atraviese [23], como resultado se hizo un corte en el dendrograma en la altura 3. Con ello, el resultado del modelo de clustering muestra 6 grupos diferentes, donde cada comuna pertenece a un grupo específico. Esto nos indica cuales comunas son similares entre sí en términos de su vulnerabilidad frente al comportamiento de casos positivos por COVID-19.

Para llevar a cabo el estudio se tomó como base de desarrollo la metodología *Cross Industry Standard Process for Data Mining* (CRISP-DM) [24, 25], la cual es comúnmente usada en técnicas de minería de datos como clustering de datos [26, 27, 28, 29, 30]. CRISP-DM consta de 6 fases, de las cuales se desarrollaron 5 en este estudio, desde fase 1 hasta fase 5. En cada una de ellas se ejecutaron diferentes tareas. En fase 1, entendimiento del negocio, se realizó la comprensión del problema, se definieron los objetivos de la solución y de minería de datos, adicionalmente, evaluación de recursos disponibles y necesarios. Durante la fase 2, comprensión de los datos, se identificaron las fuentes de datos necesarias y se realizó el proceso de exploración, descripción estadística e integración de los datos. En la fase 3, preparación de los datos, se ejecutaron tareas de transformación de datos, incluyendo agregaciones, agrupaciones y normalización de los datos, para obtener los datos limpios, dispuestos para el posterior análisis. En la fase 4, modelado, como primera medida se realizó el proceso de selección de variables para clustering, selección de la técnica, en la que se eligió clustering jerárquico aglomerativo y finalmente se ejecutó el modelo. Finalmente, en la fase 5, evaluación del modelo, se evaluó gráficamente el modelo resultado, por medio de un dendrograma, se comparó los grupos obtenidos y se realizó un análisis sobre los resultados.

La descripción de cada fase en la metodología implementada se organiza en esta sección como sigue. Descripción de las fuentes y construcción de los datos se presentan en la sección 3.1, llamado construcción de datos. El proceso de selección de variables para el modelo construido se muestra en la sección 3.2, selección de variables. La sección 3.3, refiere la fase de entrenamiento y construcción del modelo de clustering. Finalmente, la evaluación del modelo y resultados se exponen en la sección 4 de este artículo, resultados.

Construcción de datos

A partir de las fuentes de datos descritas anteriormente, se realizó un proceso de extracción y transformación de los datos.

- Los datos del censo de población y vivienda 2018 [21] fueron extraídos en formato csv y posteriormente se integraron en un único conjunto de datos con todas las variables de caracterización por comuna.
- Los datos sobre casos de COVID-19 en Cali publicados por la alcaldía local [22], fueron extraídos de documentos pdf y transformados a archivos csv.
- Algunos boletines presentaban cifras de COVID-19 por comuna, mientras que otros presentaban por barrio. Entonces, se integraron y se aplicaron agregaciones sobre los datos para mantener la granularidad por comuna en los conjuntos de datos construidos.
- Se mantiene un formato numérico para cada una de las variables del conjunto de datos.
- El conjunto de datos fue normalizado con el fin de tener magnitudes equivalentes para todas las variables.

En resumen, en la tabla 1 se muestra las variables consideradas para el análisis realizado, que contiene 22 registros, correspondiente a información de cada comuna de Cali.

Tabla 1. Variables para análisis de clustering de vulnerabilidad por comuna en la ciudad de Cali.

Nombre de Variable	Descripción	Valor Mínimo	Valor Máximo	Valor promedio	Desviación Estándar
Comuna	Identificación de la comuna.	NA	NA	NA	NA
Número de barrios	Cantidad de barrios que conforman la comuna.	4	33	15.32	7.24
Número de estratos	Cantidad de estratos socioeconómicos presentes en los barrios de la comuna.	1	5	2.64	0.9
Delta estratos	Diferencia entre el estrato mayor y menor presentes en la comuna.	0	5	1.77	1.19
Población	Cantidad de habitantes en la comuna.	26775	144520	80466.09	33714.04
Superficie	Área geográfica que conforma la comuna en km2.	2.33	12.55	5.47	3.02
Densidad poblacional	Cantidad de habitantes por cada km2 (población/superficie).	2888.1	27540	16776.19	7092.82
% informalidad	Proporción de los habitantes de la comuna que laboran informalmente.	0.68	0.87	0.81	0.06
% educación media	Proporción de habitantes de la comuna que su máximo nivel académico alcanzado es educación media.	0.14	0.32	0.27	0.05
% educación ninguno	Proporción de habitantes de la comuna que no alcanzaron ningún nivel académico.	0.01	0.04	0.02	0.01
% adulto mayor	Proporción de habitantes de la comuna mayores de 60 años.	0.08	0.26	0.17	0.04
casos acumulados	Cantidad de casos positivos acumulados de COVID-19 en la comuna.	50	877	293.32	183.48

Fuente: Elaboración propia

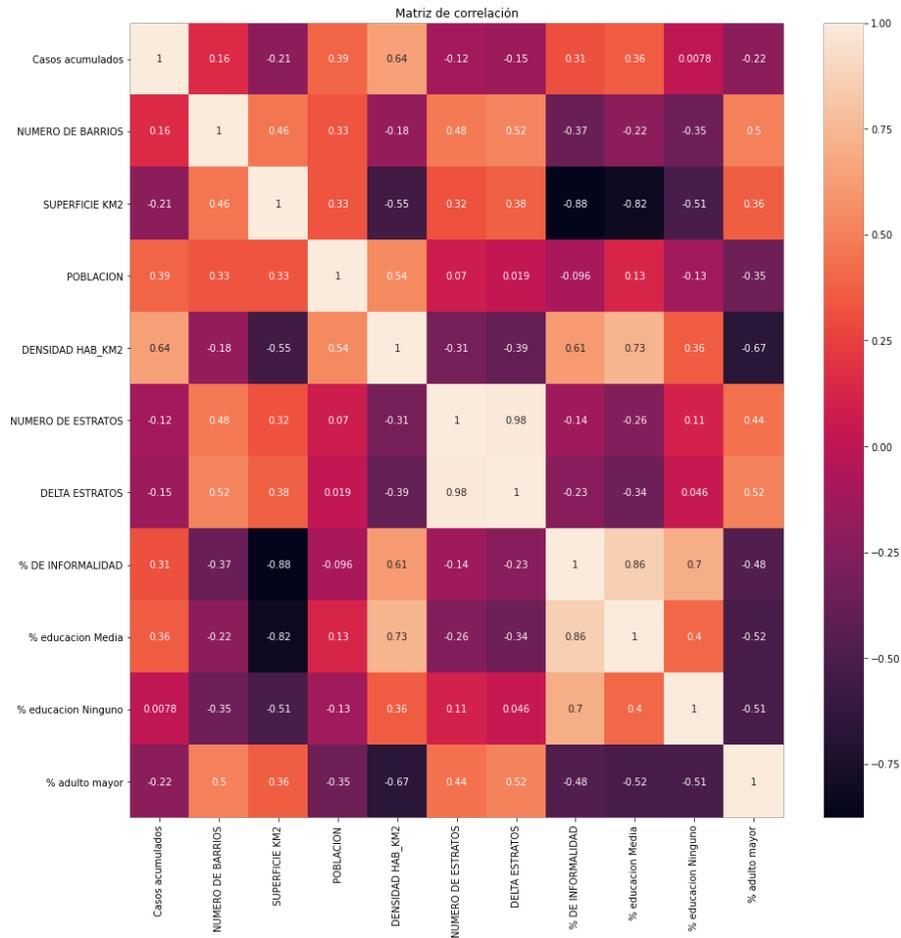
Selección de variables

Para seleccionar las variables que aportan más información para construir el modelo de clustering se calculó el coeficiente de correlación de Pearson.

En este proceso se calculó los coeficientes de correlación de cada par de variables, luego, se consolidaron en una matriz de correlación, mostrada en la figura 1. Los valores de los coeficientes se encuentran en un rango entre [-1,1], un valor cercano a 0 significa que existe poca o nula correlación entre las variables, un valor cercano a 1 significa correlación positiva, es decir, las variables son directamente proporcionales, y un valor cercano a -1 significa correlación negativa, es decir, las variables son inversamente proporcionales.

Finalmente, las variables con las cuales se construyó el modelo de clustering son: casos acumulados, densidad poblacional, delta estratos, informalidad, educación media, adulto mayor.

Figura 1. Matriz de correlación de variables de caracterización y casos acumulado de COVID-19 por comuna en Cali.



Fuente: Elaboración propia

El valor de cada intersección fila y columna es el coeficiente de correlación de Pearson entre las variables correspondientes de la fila y columna. En la matriz de correlación se estudia dos aspectos, la correlación entre cada par de variables y la correlación entre cada variable y los casos acumulados. De aquí que, se observa que la densidad poblacional tiene alta correlación con las variables superficie y población, este hecho es explicado a que la densidad es un resultado del cálculo de las variables mencionadas. También,

se encontró que las variables que representan educación tienen correlación de 0.4. Otras correlaciones encontradas a partir de la matriz son entre la variable delta estratos-número de barrios, y delta estratos-número de estratos. Entonces, para seleccionar y/o descartar las variables correlacionadas se tiene en cuenta cuales aportan más información a la variable objetivo que es casos acumulados, por lo tanto, se seleccionan densidad, delta estratos, % educación media, descartando superficie, población, número de barrios, número de estratos y % de no educación.

Construcción del modelo

Clustering de datos es una técnica de aprendizaje no supervisado que construye grupos de individuos que comparten características similares. Existen diferentes medidas de similitud y diferentes algoritmos para construir los grupos. En general, los algoritmos se pueden dividir en dos categorías, clustering jerárquico y no jerárquico. En la categoría de clustering no jerárquico el algoritmo más utilizado es k-means, mientras que en la categoría de clustering jerárquico es bien conocido el método aglomerativo. La principal diferencia entre estos dos enfoques es que el clustering no jerárquico requiere determinar el número de grupos a priori y en el clustering jerárquico el número óptimo de grupos puede ser obtenido por el modelo en sí mismo.

El modelo construido en este trabajo está basado en un algoritmo jerárquico aglomerativo, se escoge como métrica de disimilitud la distancia Euclidiana, la cual calcula la distancia entre dos puntos con base al teorema de Pitágoras. La máxima distancia es usada como método de distancia, quien básicamente maximiza la distancia entre todas las observaciones de dos grupos. Así, el modelo obtenido se gráfica mediante un dendrograma, el cual muestra diferentes grupos según el punto de corte seleccionado. Tomando la línea vertical más larga sin una línea horizontal que la atravesase en el dendrograma [23], se decide un corte en el punto 3. De aquí que, se obtienen 6 grupos homogéneos en términos del número de elementos. La tabla 2 presenta las características de cada comuna y a que grupo pertenecen (Ver tabla 2).

Tabla 2. Descripción de características por comuna en Cali.

Comuna	Casos acumulados	Densidad poblacional (hab/ Km ²)	Delta estratos	% de Informalidad	% Educación Media	% Adulto Mayor
1	83	12775.75	1	0.82	0.27	0.14
2	219	8621.14	5	0.69	0.2	0.23
3	133	7227.70	4	0.84	0.24	0.23
4	145	11724.27	1	0.81	0.29	0.19
5	181	21954.69	1	0.75	0.26	0.17
6	280	24697.21	1	0.8	0.32	0.15
7	170	11456.01	2	0.84	0.31	0.17
8	242	15264.78	1	0.82	0.29	0.2
9	158	10280.79	2	0.85	0.29	0.24
10	434	19554.44	1	0.8	0.27	0.21
11	495	22508.11	1	0.84	0.3	0.18
12	877	24106.48	1	0.85	0.3	0.17
13	487	24584.34	2	0.87	0.3	0.15
14	386	27539.52	1	0.87	0.31	0.12
15	421	25467.93	2	0.87	0.29	0.11

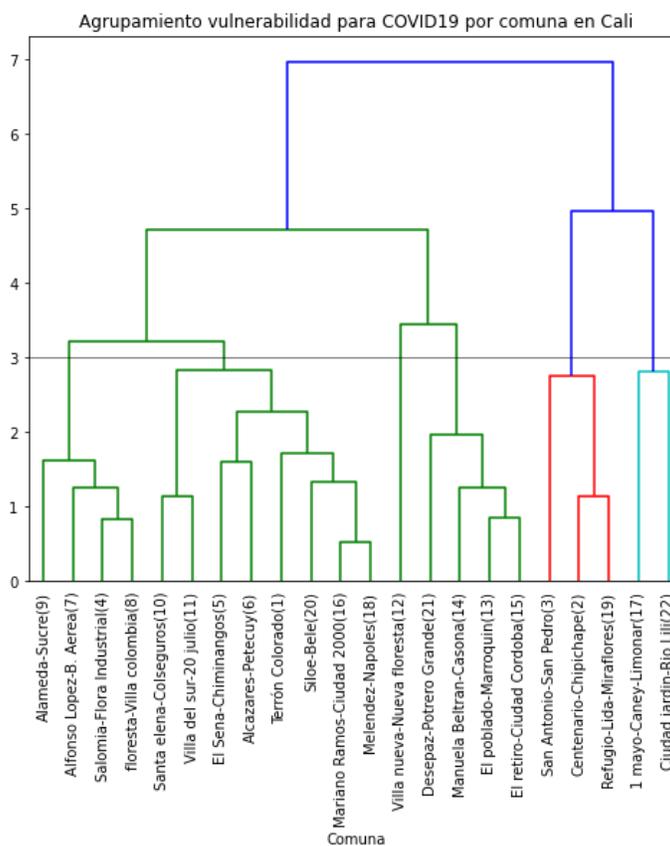
16	292	18576.24	2	0.84	0.29	0.15
17	325	11515.54	2	0.71	0.19	0.19
18	306	18142.20	2	0.83	0.29	0.13
19	269	7823.42	4	0.7	0.19	0.26
20	123	19843.82	2	0.87	0.27	0.14
21	377	22523.76	1	0.84	0.32	0.08
22	50	2888.09	0	0.68	0.14	0.17

Fuente: Elaboración propia

Resultados

En la figura 2, se observa los diferentes grupos obtenidos en el modelo. Las etiquetas de cada comuna corresponden a algunos barrios representativos de la comuna, lo cual ayuda a la identificación geográfica de ella. De manera general, se observa que la vulnerabilidad de la población en la ciudad de Cali en escenarios de riesgo de pandemia por COVID-19, está determinada por características sociodemográficas y socioeconómicas. En el dendrograma se determina un punto de corte en 3, con esto, se identifican 6 grupos homogéneos de acuerdo con la vulnerabilidad de la población de cada comuna y el crecimiento de casos positivos por COVID-19.

Figura 2. Clustering de vulnerabilidad poblacional por comuna en Cali.



Fuente: Elaboración propia

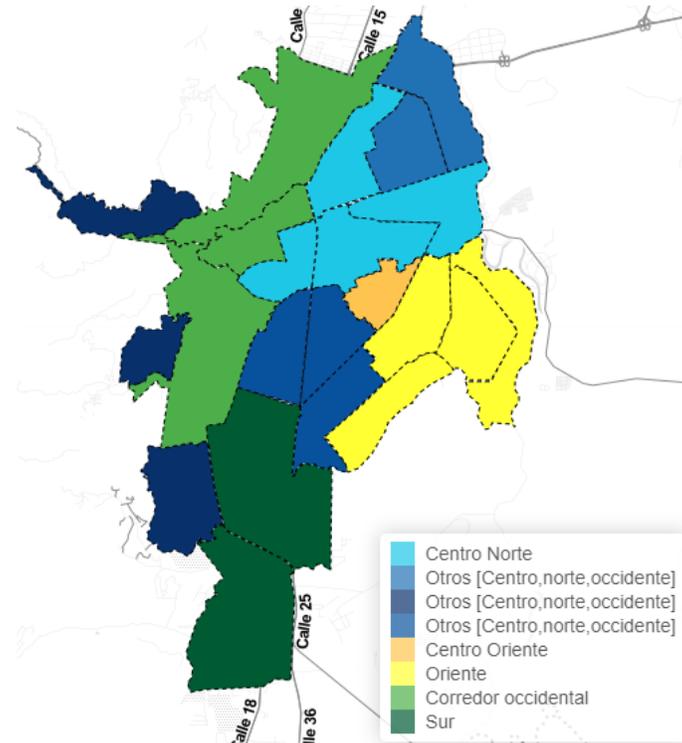
La tabla 3 muestra los grupos obtenidos y consolida los resúmenes de las variables de cada grupo, mientras que la figura 3 ubica geográficamente en un mapa los grupos obtenidos diferenciados por colores. Los conglomerados obtenidos son los siguientes: uno, algunas comunas del centro y norte de la ciudad que son de estrato bajo y medio, y presenta el 11% de los casos de COVID-19 acumulados. Dos, otras comunas del centro y la zona ladera que son estratos bajo y medio, con un 34% es el grupo que tiene el mayor número de casos acumulados por COVID-19. Tres, la comuna 12 ubicada en el centro oriente de Cali de estrato bajo y medio, cuenta con 14% de casos acumulados por COVID-19. Cuatro, comunas del oriente de la ciudad, las cuales son de estrato socioeconómico bajo y el segundo grupo con mayor porcentaje de casos acumulados por COVID-19, correspondiente al 26%. Cinco, las comunas del corredor occidental de la ciudad, pertenecientes a estrato medio y alto, tiene 10% de casos acumulados por COVID-19 siendo el segundo grupo con menos contagios. Seis, las comunas del sur de la ciudad, estrato alto y es el grupo que presenta el porcentaje más bajo de contagios de COVID-19 con 6%. Una observación particular, es de la comuna 12 que se ubica como único elemento de un grupo, está comuna es similar al grupo de oriente de la ciudad en sus características poblacionales, pero se ubica en un grupo diferente debido a que tiene gran diferencia de casos positivos de COVID-19 comparado con el grupo 4.

Tabla 3. Grupos y resumen de variables por cada grupo de vulnerabilidad poblacional por comuna en Cali.

Grupo	Color	Comunas	Casos Acumulados	% Casos	Promedio por grupo				
					Densidad Poblacional	Delta estratos	% de Informalidad	% Educación Media	% Adulto Mayor
1. Centro norte		4, 7, 8, 9	715	11%	12181.46	2	0.83	0.30	0.20
2. Otros [Centro,norte,occidente]		1, 5, 6, 10, 11, 16, 18, 20	2194	34%	19756.56	1	0.82	0.28	0.16
3. Centro oriente		12	877	14%	24106.48	1	0.85	0.30	0.17
4. Oriente		13, 14, 15, 21	1671	26%	25028.89	2	0.86	0.31	0.11
5. Corredor occidental		2, 3, 19	621	10%	7890.75	4	0.74	0.21	0.24
6. Sur		17, 22	375	6%	7201.81	1	0.70	0.16	0.18

Fuente: Elaboración propia

Figura 3. Geovisor de clúster de vulnerabilidad poblacional por comuna en Cali.



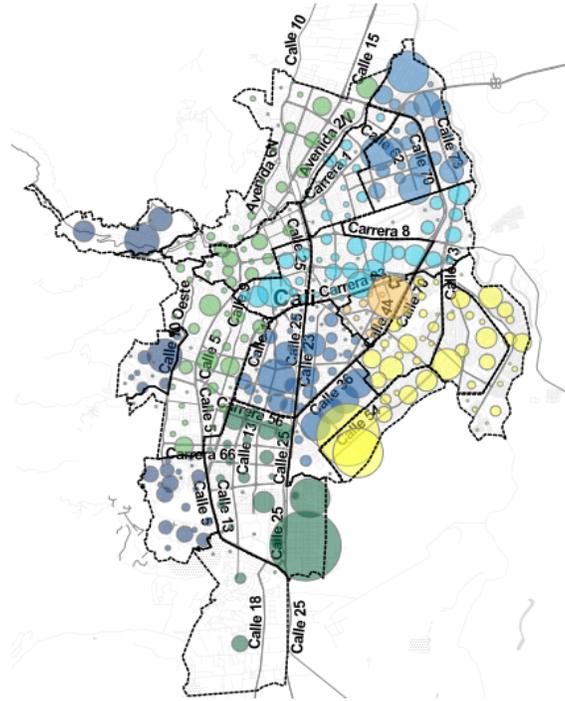
Fuente: Elaboración propia

Detallando en el consolidado de la tabla 3, se puede notar que el grupo 5 y 6 presentan densidad poblacional más baja, y su % de trabajo informal se encuentra por debajo del 80%, esto contrasta con los grupos del 1-4, que presentan alta densidad poblacional y su % de trabajo informal es superior al 80%, lo que indica que el trabajo informal de la ciudad de Cali se concentra en las comunas que pertenecen a estos grupos. Además, teniendo en cuenta el % de adulto mayor, se presenta menor población adulta mayor en los grupos de 1-4, significa que gran parte de la población de los grupos 1-4 son personas jóvenes y trabajadores informales. También, se evidencia que el grupo 5 presenta barrios con diversidad de estrato socioeconómico, ya que muestra el promedio de delta estratos mayor, con un valor de 4. El estrato de estos barrios en el grupo 5 oscilan desde el estrato 1 hasta el 6.

Respecto al % de educación media, esta variable nos indica el porcentaje de la población en cada comuna que su máximo nivel de educación es educación media. Analizando los resultados por grupo, se nota que los grupos de menor porcentaje son los grupos 5 y 6, esto puede explicarse por dos motivos, que la mayoría de población tiene nivel educativo más bajo no alcanzando la educación media, o que siguieron estudiando y han logrado niveles de profesional y postgrado. Relacionando esta variable con el % de informalidad, podemos deducir que la población con menor nivel de educación tiene mayor % de informalidad.

Los grupos se pueden visualizar mediante el mapa de la figura 3, donde se diferencia cada conglomerado por un color diferente. En este mapa, se puede evidenciar que el comportamiento del COVID-19 relacionado con las características poblacionales de la comuna en la ciudad de Cali está dado por corredores, donde se identifican 3. Los cuales están marcados por un mismo color en diferentes tonos: azul es corredor central desde norte a sur, ver corredor occidental desde norte al centro y amarillo corredor de oriente a sur. Además, en la figura 4 se muestra una vista adicional de los grupos, complementado con el número de casos de COVID-19 por cada barrio.

Figura 4. Geovisor de grupos de vulnerabilidad poblacional por comuna en Cali (Vista casos de COVID-19 por barrio).



Fuente: Elaboración propia

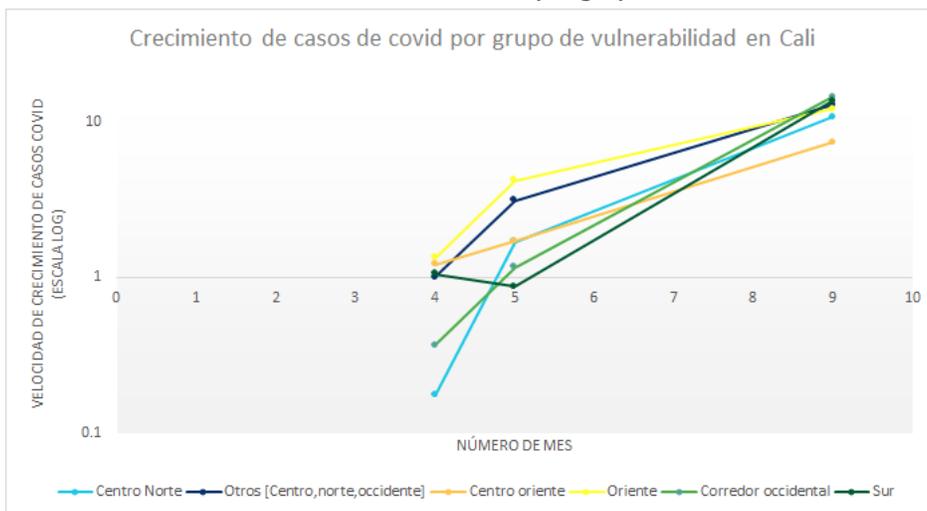
En la columna perfil estadístico (como delta del número de casos al final de un mes menos número de casos al inicio del mismo mes) de la tabla 4, se representa la velocidad de crecimiento de casos de COVID-19 para cada grupo mediante el perfil estadístico correspondiente, visualizado en la figura 5a y 5b. En ella se puede constatar que los grupos encontrados son disjuntos entre sí ya que cada uno tiene un perfil estadístico distinto. Se identifica algunos comportamientos particulares, el caso del grupo sur, donde la velocidad de crecimiento de casos descendió entre abril y mayo, y a partir de mayo creció aceleradamente el número de casos de COVID-19. Otro caso es el grupo corredor occidental, el cual tiene un leve crecimiento en la velocidad de propagación, este es un aspecto positivo ya que este corredor comprende el 12% de la población total de Cali, además, comunica la ciudad de norte a sur por algunas troncales principales. Adicionalmente se identifica, que el grupo centro oriente y centro norte tienen comportamientos diferentes entre el periodo abril y mayo, pero a partir de mayo su comportamiento es similar. Finalmente, el comportamiento del grupo oriente y las comunas pertenecientes al grupo otros, muestran un comportamiento similar respecto a la velocidad de crecimiento de casos de COVID-19 a pesar de no estar contiguos geográficamente.

Tabla 4. Perfil estadístico por grupo.

Grupo	Ecuación	R2	Perfil Estadístico (log)
1. Centro norte	$y=13.466\ln(x) - 19.121$	0.982	13.47
2. Otros [Centro,norte,occidente]	$y=14.828\ln(x) - 20.064$	0.99	14.83
3. Centro oriente	$y=8.0029\ln(x) - 10.431$	0.963	8
4. Oriente	$y=13.188\ln(x) - 16.988$	1	13.19
5. Corredor occidental	$y=18.216\ln(x) - 26.26$	0.953	18.22
6. Sur	$y=16.72\ln(x) - 23.772$	0.923	16.72

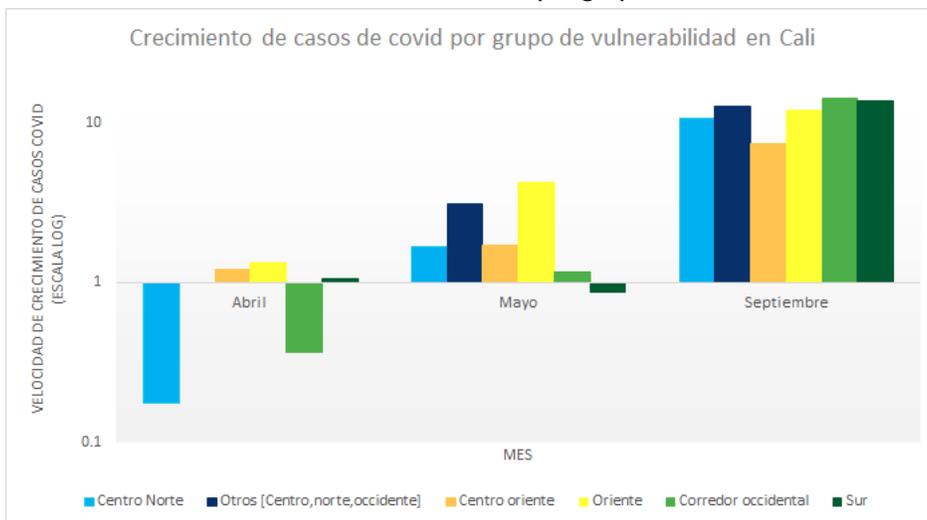
Fuente: Elaboración propia

Figura 5a. Velocidad de crecimiento de casos de COVID-19 por grupo en Cali.



Fuente: Elaboración propia

Figura 5b. Velocidad de crecimiento de casos de COVID-19 por grupo en Cali.



Fuente: Elaboración propia

Conclusiones

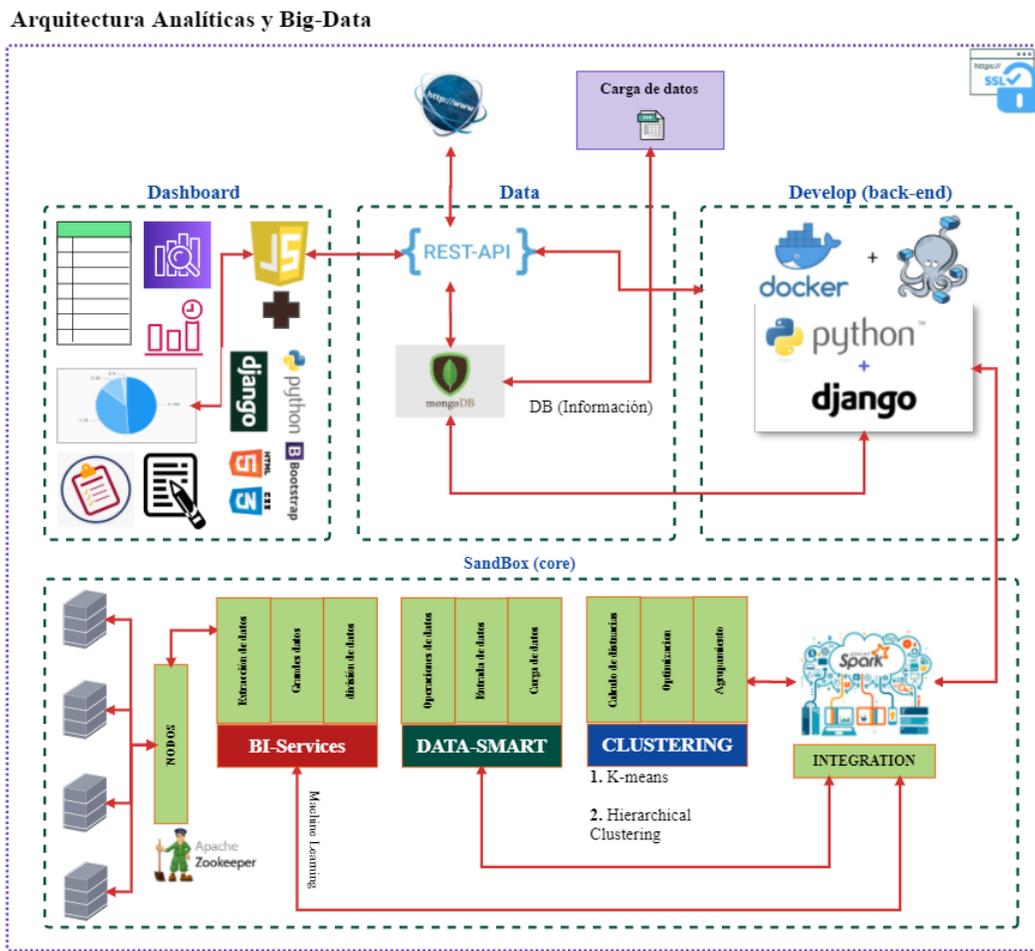
Finalmente, el análisis realizado sugiere que las estrategias generales de control de la pandemia, aplicadas a toda la ciudad, por ejemplo, restricciones de movilidad para toda la ciudad, se deben complementar con el diseño e implementación de estrategias focalizadas por comunas o barrios. Además, para el diseño de dichas estrategias las variables de vulnerabilidad deben ser tenidas en cuenta para lograr mayor efectividad en su implementación, igual como muestra un estudio en Estados Unidos donde se reportan tres grupos de factores de riesgo (comorbilidades, sociodemográficos y ambientales) en el contexto de COVID-19 para identificar condados más vulnerables que se beneficiarían de la asignación de recursos [31].

El método implementado en el estudio de vulnerabilidad permite identificar grupos poblacionales que son similares en cuanto a sus variables de vulnerabilidad y cómo se relaciona con el comportamiento del número de casos de COVID-19, permitiendo focalizar comunas que comparten las mismas dinámicas en

el escenario de pandemia de COVID-19. Además, estos grupos son homogéneos, en tanto que tienen un número de elementos equivalentes y de manera general concuerdan con otros estudios de vulnerabilidad realizados en el marco del proyecto SIGELO.

Los resultados finales de este análisis se pueden visualizar en formas de mapa geoespacial en el geovisor prototipo desarrollado para el proyecto. Además, para el desarrollo de este estudio se implementó un prototipo de una arquitectura de analíticas y Big Data, basada en REST API, ver figura 6, los detalles de la arquitectura se explican en la sección arquitectura de este informe. Esta arquitectura está dividida en cuatro componentes, un dashboard prototipo que cumple el rol de interfaz de usuario, un componente de datos que almacena los conjuntos de datos, el componente de analítica donde se implementan los diferentes algoritmos de aprendizaje automático y finalmente, un sandbox es quien tiene toda la infraestructura de hardware que soporte la integración y procesamiento de datos. En consecuencia, los análisis presentados pueden ser replicados de manera fácil y rápida ya que la arquitectura permite la automatización de estos procesos. Consecuentemente, la arquitectura también brinda apoyo a la toma de decisiones inteligentes en el tiempo, ya que estos análisis se pueden ejecutar periódicamente con datos actualizados para monitorear los cambios de las poblaciones en el comportamiento de la pandemia, haciendo uso de la arquitectura implementada.

Figura 6. Arquitectura de analíticas y Big Data.



Fuente: Elaboración propia

El análisis de los grupos mediante el perfil de la velocidad de crecimiento de casos de COVID-19 valida que los grupos son disjuntos ya que los perfiles resultantes son diferentes entre sí, por lo que, el resultado de este trabajo representa un valor importante para caracterizar las comunas que pertenecen a cada grupo. También, estos perfiles concuerdan con el crecimiento de casos de COVID-19 que se ha presentado en la ciudad, siendo esto un buen indicador de la validez de los resultados obtenidos en el trabajo.

Los resultados obtenidos verifican que las comunas tienen comportamientos diferentes que están relacionado a sus características sociodemográficas y socioeconómicas, por lo que, se plantea la necesidad de tener en cuenta las características de cada comunidad al momento de diseñar estrategias que mitiguen el contagio y el impacto económico, generando así, barreras naturales a la pandemia acordes para cada comunidad. Así, se puede implementar dichas estrategias por sectores o comunas, apoyados en la información que brinda el estudio realizado.

Se encuentra que las comunas que pertenecen a un mismo grupo están ubicadas en zonas contiguas, lo que facilita la implementación de estrategias focalizadas por comunas o sectores, tal como se ha hecho en otras ciudades del país. Es el caso de Bogotá, donde se usó un sistema tecnológico llamado Saludata [32], la alcaldía de Bogotá monitorea el comportamiento de casos de COVID-19 en cuanto a contagios, transmisión, muertes, zonas que presentan mayor población de adultos mayores. etc [33]. Los indicadores que mostraba este sistema, durante la reactivación económica, permitió identificar zonas de alto riesgo, las cuales llamaron Zonas de Cuidado Especial, y dio lugar a implementar en la ciudad de Bogotá cuarentena por localidades [34]. La cuarentena por localidades consiste en restricciones de movilidad de personas y vehículos en las localidades, tanto dentro de la localidad como su salida a cualquier otra por un determinado periodo de tiempo [35, 36, 37, 38]. Del 13 al 26 de julio el turno fue para los habitantes de: Ciudad Bolívar, San Cristóbal, Rafael Uribe, Chapinero, Santa Fe, Usme, Los Mártires y Tunjuelito. Luego, del 23 de julio al 6 de agosto en: Bosa, Antonio Nariño, Kennedy, Puente Aranda y Fontibón. Del 31 de julio al 14 de agosto en: Suba, Engativá y Barrios Unidos y por último del 16 al 30 de agosto en: Usaquén, Chapinero, Santa Fe, La Candelaria, Teusaquillo, Puente Aranda y Antonio Nariño. Con estas estrategias por localidad se logró mitigar la velocidad de contagios en Bogotá que se presentó debido a la reapertura de algunos sectores económicos, y al mismo tiempo permitió continuar con fomento de la reactivación económica.

Como trabajo futuro se propone tener un panorama más detallado de la población de Cali, es ideal tener una granularidad de casos de COVID-19 a nivel de barrios, en este mismo sentido se hace necesario romper las barreras que se tienen para el acceso a los datos, que permita acceder de manera fácil y rápida a las fuentes de información. Además, poder integrar y capturar nuevas fuentes de información que alimenten el análisis con casos nuevos, acumulados, número de muertes, recuperados, esto generaría un panorama amplio del comportamiento de la pandemia por cada barrio o comuna.

Referencias bibliográficas

1. Naciones Unidas, «2009: UNISDR terminología sobre reducción del riesgo de desastres», 2009.
2. D. Elías-Cuartas et al., «Análisis espacio-temporal del SARS-coV-2 en Cali, Colombia», *Rev. Salud Pública*, vol. 22, n.º 2, Art. n.º 2, mar. 2020, doi: <https://doi.org/10.15446/rsap.v22n2.86431>.
3. Santiago de Cali. Alcaldía, *Decreto no. 4112.010.20.0917 de 2020 (mayo 28 de 2020) «Por el cual se declara la alerta naranja en el distrito especial, deportivo, cultural, turístico, empresarial y de servicios de Santiago de Cali, se imparten instrucciones en materia de salud, orden público y reactivación económica para preservar la vida y la seguridad ciudadana, y se dictan otras disposiciones»*. 2020.
4. L. Willem et al., «The impact of contact tracing and household bubbles on deconfinement strategies for COVID-19», *Nat. Commun.*, vol. 12, n.º 1, Art. n.º 1, mar. 2021. DOI: <https://doi.org/10.1038/s41467-021-21747-7>.
5. C. J. Silva et al., «Optimal control of the COVID-19 pandemic: controlled sanitary deconfinement in Portugal», *Sci. Rep.*, vol. 11, n.º 1, Art. n.º 1, feb. 2021. DOI: <https://doi.org/10.1038/s41598-021-83075-6>.
6. B. L. Dickens et al., «Strategies at points of entry to reduce importation risk of COVID-19 cases and reopen travel», *J. Travel Med.*, vol. 27, n.º taaa141, dic. 2020. DOI: <https://doi.org/10.1093/jtm/taaa141>.
7. S.-Y. Cheng, C. J. Wang, A. C.-T. Shen, y S.-C. Chang, «How to Safely Reopen Colleges and Universities During COVID-19: Experiences From Taiwan», *Ann. Intern. Med.*, vol. 173, n.º 8, pp. 638-641, jul. 2020. DOI: <https://doi.org/10.7326/M20-2927>.
8. P. J. Goldschmidt-Clermont, «COVID-19 real-world data for the US and lessons to reopen business», *PLOS Pathog.*, vol. 16, n.º 8, p. e1008756, ago 2020. DOI: <https://doi.org/10.1371/journal.ppat.1008756>.
9. J. Kinnunen, I. Georgescu, Z. Hosseini, y A.-M. Androniceanu, «Dynamic indexing and clustering of government strategies to mitigate Covid-19», *Entrep. Bus. Econ. Rev.*, vol. 9, n.º 2, pp. 7-20, 2021. DOI: <https://doi.org/10.15678/EBER.2021.090201>.
10. V. Zarikas, S. G. Pouloupoulos, Z. Gareiou, y E. Zervas, «Clustering analysis of countries using the COVID-19 cases dataset», *Data Brief*, vol. 31, p. 105787, ago. 2020. DOI: <https://doi.org/10.1016/j.dib.2020.105787>.
11. Johns Hopkins University, «Coronavirus COVID-19 (2019-nCoV)». <https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6> (accedido feb. 28, 2021).
12. C. Lopez y N. Chhibba, «Clustering Analysis of Countries using INFORM risk and COVID-19 data», *Towards Data Science*, nov. 28, 2020. Accedido: mar. 07, 2021. [En línea]. Disponible en: <https://towardsdatascience.com/clustering-analysis-of-countries-using-inform-risk-and-covid-19-data-6dbc54e0bd80>
13. M. R. Mahmoudi, D. Baleanu, Z. Mansor, B. A. Tuan, y K.-H. Pho, «Fuzzy clustering method to compare the spread rate of Covid-19 in the high risks countries», *Chaos Solitons Fractals*, vol. 140, p. 110230, nov. 2020. DOI: <https://doi.org/10.1016/j.chaos.2020.110230>.
14. P. Sengupta, B. Ganguli, S. SenRoy, y A. Chatterjee, «An analysis of COVID-19 clusters in India», *BMC Public Health*, vol. 21, n.º 1, p. 631, mar. 2021. DOI: <https://doi.org/10.1186/s12889-021-10491-8>.
15. S. Kumar, «Use of cluster analysis to monitor novel coronavirus-19 infections in Maharashtra, India», *Indian J. Med. Sci.*, vol. 72, n.º 2, pp. 44-48, ago. 2020. DOI: https://doi.org/10.25259/IJMS_68_2020.
16. Maharashtra. Departamento de Salud Pública, «Public Health Department, Governemnt of Maharashtra». <https://phdmah.maps.arcgis.com/> (accedido feb. 28, 2021).

17. A. Maugeri, M. Barchitta, G. Basile, y A. Agodi, «Applying a hierarchical clustering on principal components approach to identify different patterns of the SARS-CoV-2 epidemic across Italian regions», *Sci. Rep.*, vol. 11, n.º 1, Art. n.º 1, mar. 2021. DOI: <https://doi.org/10.1038/s41598-021-86703-3>.
18. T. Tian *et al.*, «Risk factors associated with mortality of COVID-19 in 3125 counties of the United States», *Infect. Dis. Poverty*, vol. 10, n.º 1, p. 3, ene. 2021, doi: <https://doi.org/10.1186/s40249-020-00786-0>.
19. D. Lilleri, F. Zavaglio, E. Gabanti, G. Gerna, y E. Arbustini, «Analysis of the SARS-CoV-2 epidemic in Italy: The role of local and interventional factors in the control of the epidemic», *PLOS ONE*, vol. 15, n.º 11, p. e0242305, nov. 2020, doi: <https://doi.org/10.1371/journal.pone.0242305>.
20. M. Cruz Rendón, «Radiografía del desempleo en Cali: jóvenes y mujeres los más afectados», 2020. Accedido: nov. 26, 2020. [En línea]. Disponible en: <https://www.cali.gov.co/publicaciones/155527/radiografia-del-desempleo-en-cali-jovenes-y-mujeres-los-mas-afectados/>
21. Departamento Administrativo Nacional de Estadística, «DANE::Redatam Webserver | Procesamiento y Difusión de Microdatos», 2018. <http://systema59.dane.gov.co/bincol/RpWebEngine.exe/Portal?BASE=CNPVBASE4V2&lang=esp> (accedido nov. 25, 2020).
22. Santiago de Cali. Alcaldía, «Boletines Epidemiológicos», 2020. <https://www.cali.gov.co/publicaciones/152840/boletines-epidemiologicos/> (accedido nov. 25, 2020).
23. W. Sullivan, *Python Machine Learning Illustrated Guide For Beginners & Intermediates: The Future Is Here!* PublishDrive, 2019.
24. F. Provost y T. Fawcett, *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media, Inc., 2013.
25. R. Wirth y J. Hipp, «CRISP-DM: Towards a standard process model for data mining», *Proc. 4th Int. Conf. Pract. Appl. Knowl. Discov. Data Min.*, ene. 2000, [En línea]. Disponible en: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.198.5133>
26. N. Legowo y K. Hwibawa, «Data mining as tools to improve marketing campaign», *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, n.º 5, pp. 7283-7288, 2020. DOI: <https://doi.org/10.30534/ijatcse/2020/57952020>.
27. S. P. Bagherighadikolaei, R. Ghousi, y A. Haeri, «A Data Mining Approach for Forecasting Failure Root Causes: A Case Study in an Automated Teller Machine (ATM) Manufacturing Company», *J. Optim. Ind. Eng.*, vol. 13, n.º 2, pp. 101-121, 2020. DOI: <https://doi.org/10.22094/JOIE.2020.1863364.1630>.
28. R. Valter *et al.*, «Intelligent Epidemiological Surveillance in the Brazilian Semi-arid», en *2020 IEEE International Conference on E-health Networking, Application Services (HEALTHCOM)*, mar. 2021, pp. 1-6. DOI: <https://doi.org/10.1109/HEALTHCOM49281.2021.9399018>.
29. N. N. Fakhry, E. Asfoura, y G. Kassam, «Tracking Coronavirus Pandemic Diseases using Social Media: A Machine Learning Approach», *Int. J. Adv. Comput. Sci. Appl. IJACSA*, vol. 11, n.º 10, Art. n.º 10, 31 2020. DOI: <https://doi.org/10.14569/IJACSA.2020.0111028>.
30. I. D. Sudirman, R. Aryanto, y Mulyani, «Optimizing decision tree criteria for predicting COVID-19 mortality in South Korea dataset», *J. Theor. Appl. Inf. Technol.*, vol. 98, n.º 15, pp. 2889-2900, 2020.
31. N. Vahabi, M. Salehi, J. D. Duarte, A. Mollalo, y G. Michailidis, «County-level longitudinal clustering of COVID-19 mortality to incidence ratio in the United States», *Sci. Rep.*, vol. 11, n.º 1, Art. n.º 1, feb. 2021. DOI: <https://doi.org/10.1038/s41598-021-82384-0>.
32. Bogotá. Observatorio de Salud, «Casos confirmados de COVID-19 | SALUDATA». <https://saludata.saludcapital.gov.co/osb/index.php/datos-de-salud/enfermedades-trasmisibles/covid19/> (accedido mar. 03, 2021).

33. H. Lara, «Datos y estadísticas en tiempo real sobre COVID-19 estarán a disposición de la ciudadanía», abr. 18, 2020. Accedido: mar. 03, 2021. [En línea]. Disponible en: <https://saludata.saludcapital.gov.co/osb/index.php/2020/04/>
34. H. Lara, «Inicia intervención en zonas de cuidado especial», jun. 17, 2020. <https://saludata.saludcapital.gov.co/osb/index.php/2020/06/17/inicia-intervencion-en-zonas-de-cuidado-especial/> (accedido mar. 03, 2021).
35. Bogotá D.C. Alcaldía Mayor, *Decreto No.186 de 2020 (15 de Agosto 2020) "Por medio del cual se adoptan medidas transitorias de policía para garantizar el orden público en diferentes localidades del Distrito Capital y se toman otras determinaciones"*. 2020.
36. C. J. Obando Gamboa, "Influencia del agua en el desempeño de los pavimentos: lluvia ácida", *Investigación e Innovación en Ingenierías*, vol. 5, n.º 2, pp. 190-206, 2017. DOI:<https://doi.org/10.17081/invinno.5.2.2761>
37. G. E. Chanchí Golondrino, W. Y. Campo Muñoz, y L. M. Sierra Martínez, «Aplicación de la regresión polinomial para la caracterización de la curva del COVID-19, mediante técnicas de machine learning», *Investigación e Innovación en Ingenierías*, vol. 8, n.º 2, pp. 87 - 105, jul. 2020.
38. Bogotá D.C. Alcaldía Mayor, *Decreto No. 169 de 2020 (Julio 12 de 2020) "Por medio del cual se imparten órdenes para dar cumplimiento a la medida de aislamiento preventivo obligatorio y se adoptan medidas transitorias de policía para garantizar el orden público en las diferentes localidades del Distrito Capital"*. 2020.