

Análisis Espacial de la Efectividad de la vacunación por COVID-19 y su asociación con factores socioeconómicos en Santiago de Cali

German Avila Rodriguez

Oscar Eduardo Muñoz Pedroza

Javier Alejandro Valencia Estrada

Nota de Aceptación

Certificamos que el presente Trabajo de Grado Satisface, en alcances y calidad, todos los requisitos que demanda un Trabajo de Grado de Maestría.

*David Arango Londoño*

David Arango Londoño

*Mario Mora*

Mario Mora

*Jabían Tobar*

Jabían Tobar

Aprobado en cumplimiento de los requisitos exigidos por la Pontificia Universidad Javeriana Cali, para optar el título de Magister en Ciencia de Datos.

*Camilo Rocha*

HERNÁN CAMILO ROCHA NIÑO Ph. D.  
Decano Facultad de Ingeniería y Ciencias

*Juan Carlos Martínez Arias*

JUAN CARLOS MARTÍNEZ ARIAS  
Director Posgrados de Ingeniería y Ciencias

Santiago de Cali, 19/07/2024



Al contestar por favor cite estos datos:

Radicado No.: 202341450100074191

Fecha: 2023-10-09

TRD: 4145.010.14.12.187.007419

Rad. Padre: 202341450100074191

GERMAN AVILA RODRIGUEZ  
Investigador  
Universidad Javeriana

Asunto: Respuesta Proyecto de Investigación titulado Análisis Espacial de la Efectividad de la vacunación por COVID-19 y su asociación con factores socioeconómicos en Santiago de Cali.

Cordial saludo

El comité de Ética en investigación de la Secretaría de Salud del Distrito Especial de Santiago de Cali, revisó el proyecto de investigación titulado "Análisis Espacial de la Efectividad de la vacunación por COVID-19 y su asociación con factores socioeconómicos en Santiago de Cali", de la investigador German Avila Rodriguez, Alejandro Valencia y Oscar Muñoz.

El comité, luego de la evaluación de los documentos enviados por los investigadores, considera que el proyecto es aprobado para su realización.

El nivel de riesgo del proyecto es Mínimo.

Cualquier modificación al proyecto debe ser revisada y aprobada por el comité, los investigadores también informarán problemas relacionados a la recolección de datos o imprevistos que pudiesen resultar en riesgo para los sujetos. De igual manera cada se debe enviar reporte de avance cada 3 meses.

El comité podrá revocar la aprobación a un proyecto de investigación en desarrollo, cuando se compruebe que hay suministro de datos falsos, existencia de riesgos adicionales no mencionados en la presentación inicial, deficiencias en la realización del estudio. La revocatoria del aval tendrá efectos inmediatos y será notificada vía correo electrónico.



SC-CER437981



Calle 4B N° 36-00 San Fernando  
[www.cali.gov.co](http://www.cali.gov.co)



ALCALDIA DE  
SANTIAGO DE CALI  
SECRETARIA DE SALUD PÚBLICA

El aval del proyecto tiene una vigencia de un año a partir de la fecha de este oficio; en caso de requerirlo, el investigador solicitará una ampliación del tiempo del aval. Igualmente, en caso de requerir una enmienda por cambios sustanciales en el proyecto, este tendrá que ser remitido al presidente del comité quien decidirá si avala la enmienda o si somete la decisión al comité en pleno. *En pleno.*

Con toda atención

FIRMA

◀ **CONCHA EASTMAN**  
Presidente  
Comité de Ética en Investigación  
Secretaría de Salud del Distrito Especial de Santiago de Cali

Elaboró: Ángela María Toro Duque - Contratista  
Revisó: Alberto Concha Eastman - contratista



SC-CER437981



Calle 4B N° 36-00 San Fernando  
[www.cali.gov.co](http://www.cali.gov.co)

Santiago de Cali, 07 de 06 de 2024

**Ingeniero:**  
**Juan Carlos Martínez Arias**  
**Director Posgrados de Ingeniería**  
**Facultad de Ingeniería y Ciencias**  
**Pontificia Universidad Javeriana - Cali**

*Con el fin de cumplir con los requisitos exigidos por la Universidad para llevar a cabo el Trabajo de Grado y posteriormente optar por el título de Magíster en Ciencia de Datos, nos permitimos presentar a su consideración el proyecto de Trabajo de Grado denominado Análisis Espacial de la Efectividad de la vacunación por COVID-19 y su asociación con factores socioeconómicos en Santiago de Cali, el cual será realizado por los estudiantes German Avila Rodríguez con código 8974445, Oscar Eduardo Muñoz Pedroza con código 8975504, Javier Alejandro Valencia Estrada con código 8975356 bajo la dirección del profesor David Arango Londoño.*

*El suscrito director del Trabajo de Grado autoriza para que se proceda a hacer la evaluación de este Proyecto ante el Tribunal que para el efecto se designe, toda vez que ha revisado cuidadosamente el documento y avala que ya se encuentra listo para ser presentado oficialmente.*

Atentamente,



Firma  
German Avila Rodríguez

C.C. 1144132002 de Cali.



Firma  
David Arango Londoño

C.C. 1130586950 de Cali.



Firma  
Oscar Eduardo Muñoz Pedroza

C.C. 1113686291 de Palmira.



Firma  
Javier Alejandro Valencia Estrada

C.C. 1061795114 de Popayan.



**Maestría en Ciencia de Datos  
Facultad de Ingeniería y Ciencias**

**FICHA RESUMEN  
TRABAJO DE GRADO DE MAESTRÍA**

**TITULO:** “Análisis Espacial de la Efectividad de la vacunación por COVID-19 y su asociación con factores socioeconómicos en Santiago de Cali ”

1. ÉNFASIS: N/A
2. TIPO DE PROYECTO: Investigación
3. ÁREA DE TRABAJO: Análisis estadístico, aprendizaje espacial y salud pública.
4. ESTUDIANTE (S): German Avila Rodriguez - Oscar Eduardo Muñoz Pedroza - Javier Alejandro Valencia Estrada.
5. CORREO ELECTRÓNICO: germanavila09@javerianacali.edu.co; 1113686291@javerianacali.edu.co; alejovalencia9@javerianacali.edu.co.
6. DIRECCIÓN Y TELÉFONO: Calle 59c #2c 03 Cali, 3057159452 - Carrera 19 # 29 - 12 Palmira, 3107000147 - Calle 3 # 3 -40 Popayán, 3117759482.
7. DIRECTOR: David Arango Londoño
8. VINCULACIÓN DEL DIRECTOR (en la universidad): Planta
9. CORREO ELECTRÓNICO DEL DIRECTOR: david.arango@javerianacali.edu.co
10. CO-DIRECTOR(ES) (Si aplica): NA
11. GRUPO O EMPRESA QUE LO AVALA (Si aplica): Secretaría de Salud Distrital de Santiago de Cali.
12. OTROS GRUPOS O EMPRESAS: NA.
13. PALABRAS CLAVE (al menos 5): COVID 19, Efectividad de la Vacunación, Modelos de Análisis Estadísticos, Factores Socioeconómicos, Análisis Espacial.
14. ODS QUE APLICA EL PROYECTO (Agenda 2030): NA
15. FECHA DE INICIO (Desarrollo del proyecto): 1/03/2023
16. RESUMEN (máximo 400 palabras): La pandemia de COVID-19 ha destacado la importancia crítica de las estrategias de vacunación para proteger la salud pública. En Santiago de Cali, la falta de información específica sobre las variaciones espaciales en la efectividad de la vacunación y la influencia de factores socioeconómicos ha generado preguntas significativas. Este estudio tiene como objetivo principal realizar un análisis espacial de la efectividad de la vacunación contra el COVID-19 y modelar su asociación con factores socioeconómicos en Santiago de Cali, utilizando modelos de aprendizaje estadístico.

La investigación sigue la metodología CRISP-DM, que comprende seis etapas. En la etapa de modelado, se implementaron modelos de análisis de autocorrelación espacial y de supervivencia a escala de barrio. Estos modelos permitieron evaluar la efectividad de la vacunación en diferentes áreas de la ciudad. Además, se desarrollaron modelos estadísticos multivariados que consideran la dependencia espacial para analizar la relación entre la efectividad de la vacunación y las características socioeconómicas.



Pontificia Universidad  
**JAVERIANA**  
Cali

# **Análisis Espacial de la Efectividad de la vacunación por COVID-19 y su asociación con factores socioeconómicos en Santiago de Cali**

German Avila Rodriguez  
Oscar Eduardo Muñoz Pedroza  
Javier Alejandro Valencia Estrada

Proyecto de investigación para optar al título de Magíster en Ciencia de Datos

Director(a) Mg. David Arango Londoño

FACULTAD DE INGENIERÍA Y CIENCIAS MAESTRÍA EN CIENCIA DE DATOS  
SANTIAGO DE CALI, JUNIO DE 2024

## TABLA DE CONTENIDO

INTRODUCCIÓN .....	5
1. DEFINICIÓN DEL PROBLEMA .....	6
1.1 PLANTEAMIENTO DEL PROBLEMA .....	6
1.2 FORMULACIÓN DEL PROBLEMA .....	7
2. OBJETIVOS DEL PROYECTO.....	9
2.1 OBJETIVO GENERAL .....	9
2.2 OBJETIVOS ESPECÍFICOS .....	9
3. MARCO DE REFERENCIA.....	10
3.1 MARCO TEÓRICO .....	10
3.2 ANTECEDENTES.....	17
4. METODOLOGÍA .....	19
5. ENTENDIMIENTO DEL NEGOCIO.....	20
6. ENTENDIMIENTO DE LOS DATOS.....	22
7. PREPARACION DE LOS DATOS.....	26
8. MODELADO .....	31
9. ANALISIS DE RESULTADOS.....	35
10. EVALUACIÓN.....	55
11. TRABAJOS FUTUROS .....	63
12. CONCLUSIONES .....	64
13. REFERENCIAS BIBLIOGRAFICAS.....	65

## LISTA DE TABLAS

Tabla 1 Dimensiones y variables del IPM para Colombia.....	13
Tabla 2 Variables covid_tg.....	22
Tabla 3 Variables DANE.....	24
Tabla 4 Barrios con mayor impacto.....	32
Tabla 5 Barrios con menor impacto.....	32
Tabla 6 Cuadro comparativo modelos de regresión espacial.....	60
Tabla 7 Cuadro comparativo de los modelos de regresión espacial.....	61
Tabla 8 Cuadro comparativo probabilidad de los modelos de regresión.....	61

## LISTA DE ILUSTRACIONES

Ilustración 1 Geovisor de Medida de Pobreza Multidimensional.....	23
Ilustración 2 Geovisor de Medida de niveles de vulnerabilidad.....	24
Ilustración 3 Migración de datos a Qgis.....	26
Ilustración 4 Puntos Casos COVID-19.....	27
Ilustración 5 Unidades poligonales.....	27
Ilustración 6 casos_positivos_fallecidos_g.....	28
Ilustración 7 Casos_positivos_g.....	28
Ilustración 8 IPM Media.....	29
Ilustración 9 IPM Mediana.....	29
Ilustración 10 IPM Moda.....	29
Ilustración 11 Impacto por Barrio.....	32
Ilustración 12 Graficas del modelo barrios de mayor impacto.....	33
Ilustración 13 Graficas del modelo para barrios de mínimo impacto.....	34
Ilustración 14 Reporte de autocorrelación espacial.....	36
Ilustración 15 Resultados simulación de Montecarlo.....	37
Ilustración 16 Histograma de simulación de Montecarlo.....	38
Ilustración 17 Diagrama de dispersión índice de Moran.....	38
Ilustración 18 Resultados del índice local de Moran.....	40
ilustración 19 Distribución de los resultados del índice local de Moran.....	41
Ilustración 20 Distribución espacial del índice local de Moran.....	42
Ilustración 21 Resultados de la prueba de dos colas.....	43
Ilustración 22 Gráfico de dispersión prueba de dos colas en RStudio.....	44
Ilustración 23 Grafico de dispersión prueba de dos colas en Geoda.....	44
Ilustración 24 Análisis de autocorrelación espacial Local.....	45
Ilustración 25 Grafico de dispersión de Moran estrato socioeconómico.....	46
Ilustración 26 Cluster BILISA impacto / estrato socioeconómico.....	47
Ilustración 27 Distribución espacial del estrato socioeconómico.....	47

Ilustración 28 Grafico de dispersión de Moran densidad poblacional.....	48
Ilustración 29 Cluster BILISA impacto / densidad poblacional.....	49
Ilustración 30 Distribución espacial densidad poblacional.....	49
Ilustración 31 Grafico de dispersión de Moran de IPM.....	50
Ilustración 32 Cluster BILISA impacto / IPM.....	51
Ilustración 33 Distribución espacial IPM.....	51
Ilustración 34 Grafico de dispersión de Moran vulnerabilidad.....	52
Ilustración 35 Cluster BILISA impacto / vulnerabilidad.....	53
Ilustración 36 Distribución espacial vulnerabilidad.....	53
Ilustración 37 Cluster Geary Local.....	55
Ilustración 38 Resumen del modelo espacial básico.....	56
Ilustración 39 Resumen del modelo de regresión de retardo espacial.....	58
Ilustración 40 Resumen del modelo de regresión con error espacial.....	59

## INTRODUCCIÓN

La pandemia de COVID-19 puso de manifiesto la importancia de las estrategias de vacunación para salvaguardar la salud pública. Sin embargo, en Santiago de Cali no se contaba con información relevante para determinar si la efectividad de la vacunación contra el COVID-19 presentaba variaciones espaciales, y si los factores socioeconómicos influían en la misma, Este desconocimiento generó interrogantes relevantes acerca del comportamiento espacial de la efectividad de la vacunación, y el papel de las condiciones socioeconómicas en dicha efectividad.

El objetivo de esta investigación es exponer los resultados del análisis espacial para evaluar la efectividad de la vacunación contra el COVID-19 en Santiago de Cali y su relación con los factores socioeconómicos. Para ello, se ejecutaron las seis etapas establecidas en la metodología CRISP-DM.

En la etapa de modelado se implementaron los modelos de análisis de autocorrelación espacial y de supervivencia a escala de barrio, con el fin de establecer la efectividad de la vacunación donde también se desarrollaron los modelos estadísticos multivariados que consideran la dependencia espacial, para estudiar la asociación entre la efectividad de la vacunación y las características socioeconómicas.

Este estudio utilizó métodos de aprendizaje estadístico y análisis espacial avanzados. Que permitieron obtener resultados que facilitan una comprensión más profunda de los factores que influyen en la efectividad de la vacunación. Al identificar patrones espaciales y socioeconómicos frente a la efectividad, contribuyendo al conocimiento científico en el área de la salud pública, específicamente en el contexto de la pandemia y la vacunación contra el COVID-19.

## **1. DEFINICIÓN DEL PROBLEMA**

Actualmente, las técnicas de análisis de datos están en constante evolución, desarrollando modelos y metodologías que permiten hacer uso oportuno de la información en la toma de decisiones del sector salud, tanto en aspectos clínicos como en las políticas y acciones de salud pública. Por ello, es de considerar los aportes de la ciencia de datos en la toma de decisiones frente al manejo de los eventos de interés en salud pública, desde el aprovechamiento de datos concernientes a las distintas patologías, comportamientos y patrones [1]. La falta de esta información afecta el desarrollo y la toma de decisiones tanto en salud pública, como también la red hospitalaria y clínica para un mejor desarrollo y atención en los diferentes ámbitos de la salud.

En este sentido, cobra importancia cómo la ciencia de datos puede aportar al sector de la salud mediante métodos de análisis estadísticos como: Modelos de regresión espacial, clustering o Análisis de conglomerados, análisis de autocorrelación, LISA (local indicators of spatial association), Identificación de cluster, kriging, entre otros. Estos métodos se aplican en diferentes aspectos como el manejo de endemias, epidemias y pandemias, como el COVID-19 que afectó al mundo en 2019-2023, y virus reemergentes como la viruela símica [2]. Por ello, es importante plantear cuestionamientos como ¿Cuál es el comportamiento espacial de la efectividad de la vacunación por COVID-19 y su asociación con factores socioeconómicos en Santiago de Cali?

Las situaciones de salud pública mencionadas han llevado a desarrollar nuevos métodos y herramientas que permiten mitigarlas y controlarlas para mejorar la calidad de vida de la sociedad actual. En este contexto, se entiende cómo la distribución espacial de la efectividad de la vacunación se relaciona con variables socioeconómicas como condiciones de la niñez, nivel de ingresos, trabajo, nivel educativo, acceso a servicios de salud y condiciones de la vivienda que son las dimensiones principales del índice de pobreza multidimensional (IPM) que permite cuantificar las disparidades, y otorgar un valor porcentual evaluando las variables condicionales que nos permite proponer estrategias de intervención efectivas.

### **1.1 PLANTEAMIENTO DEL PROBLEMA**

La pandemia de COVID-19 que afectó al mundo en 2019 tuvo un gran impacto en Colombia, donde dejó un saldo de 6.366.777 casos reportados, con un número de 6.188.459 recuperados y 142.741 muertes por el virus de SARS-CoV-2 (COVID-19) [3]. Resaltando a Santiago de Cali con 8.930 muertes, 406.751 casos confirmados y la habilitación de 850 camas de UCI para COVID-19. Ante este escenario, en la ciudad se implementaron más de 37 estrategias para cumplir con las coberturas poblacionales del Programa Ampliado de Inmunización PAI COVID-19, logrando la aplicación de 4.249.789 dosis de la vacuna hasta diciembre de 2022 [4].

Para desarrollar este programa, se adoptó la normatividad dictada por el Ministerio Nacional de Salud en el documento “Lineamientos para la gestión y administración del programa ampliado de inmunizaciones PAI – 2022”, según la disposición de los biológicos adquiridos por el Ministerio [5]. Además, fue muy importante la articulación de los actores en salud pública como los prestadores de salud, IPS, EPS y las unidades notificadoras de datos UPGD, para realizar intervenciones oportunas en el territorio nacional con las entidades territoriales. Resaltando que las UPGD se emplean para la construcción de indicadores en salud, los cuales son diversos e implican tanto fuente primaria, levantada por los equipos territoriales, como fuentes secundarias como el sistema de vigilancia en salud pública SIVIGILA.

Sin embargo, en Cali se desconoce la distribución espacial de la efectividad de la vacunación en el territorio y cómo se evidencian las inequidades relacionadas con factores socioeconómicos y barreras de acceso a salud. No se dispone de un análisis detallado sobre el porcentaje de la población que fue vacunada contra el SARS-CoV-2 a escala de barrio, ni se ha determinado si la vacuna llegó efectivamente a la población objetivo mediante las diferentes estrategias implementadas. Hace falta reconocer las características del territorio y su heterogeneidad para conocer la pertinencia y eficacia de dichas estrategias.

Es en este punto donde la ciencia de datos, a través de los modelos de aprendizaje estadístico, permite la integración con técnicas de análisis espacial para estimar la efectividad de la vacunación por COVID-19 y su asociación con factores socioeconómicos en Santiago de Cali lo que permite identificar campos de mejora en las políticas de vacunación y promover la equidad en el acceso a la protección contra el virus, al reconocer las áreas de mayor riesgo. La evaluación de la efectividad de la vacunación puede ayudar a las autoridades de salud pública a identificar los grupos de población y áreas de la ciudad que tienen mayor o menor efectividad en la vacunación, determinar las desigualdades en la distribución y ajustar sus estrategias en los planes de mejora de los programas de vacunación y vigilancia epidemiológica. También, esta evaluación proporciona información valiosa para mejorar la toma de decisiones sobre la distribución de recursos, el despliegue de personal y la estrategia de vacunación en los diferentes sectores. Así, es posible garantizar que los recursos y el personal estén enfocados en las áreas de mayor necesidad donde puedan tener el mayor impacto.

## **1.2 FORMULACIÓN DEL PROBLEMA**

Las estrategias de vacunación tienen gran relevancia para la protección de la salud de las personas, como lo dejó en evidencia la pandemia de COVID-19. No obstante, en Cali, al igual que en muchas otras ciudades, se desconoce si existe una variación espacial en la efectividad de la vacunación contra el COVID-19 y cómo los factores socioeconómicos la pueden afectar.

Este desconocimiento lleva al planteamiento de las siguientes preguntas de investigación: ¿Cuál es el comportamiento espacial de la efectividad de la vacunación por COVID-19 y su asociación con factores socioeconómicos en Santiago de Cali?, ¿Cómo establecer la efectividad de la vacunación por medio de modelos de supervivencia a escala espacial de barrio en la ciudad de Santiago de Cali?, ¿Cuáles son los requerimientos para construir modelos estadísticos multivariados explicativos sobre la efectividad en función de los factores socioeconómicos y con dependencia espacial?, ¿Qué parámetros se deben evaluar para medir el rendimiento de los diferentes modelos de análisis estadístico, en términos de su ajuste a los datos?

Esta investigación brinda una respuesta a estas preguntas mediante el análisis espacial de la efectividad de la vacunación por COVID-19 y su relación con factores socioeconómicos en Cali, utilizando modelos de aprendizaje estadístico. Los resultados de esta investigación podrán brindar información valiosa para orientar las políticas de salud pública y las intervenciones en Santiago de Cali y en otras regiones con desafíos similares.

## **2. OBJETIVOS DEL PROYECTO**

### **2.1 OBJETIVO GENERAL**

Realizar un Análisis Espacial para la Efectividad de la Vacunación por COVID-19 y modelar su asociación con factores socioeconómicos en Santiago de Cali, a través de modelos de aprendizaje estadístico.

### **2.2 OBJETIVOS ESPECÍFICOS**

1. Establecer la efectividad de la vacunación por medio de modelos de supervivencia a escala espacial de barrio en la ciudad de Santiago de Cali.
2. Construir modelos estadísticos multivariados explicativos para la efectividad en función de los factores socioeconómicos y con dependencia espacial.
3. Realizar la evaluación del rendimiento de los diferentes modelos de análisis estadístico en términos de su ajuste a los datos, utilizando métricas de validación específicas.

### **3. MARCO DE REFERENCIA**

#### **3.1 MARCO TEÓRICO**

La vacunación contra el COVID-19 es una estrategia fundamental para prevenir la propagación del virus que ha afectado al mundo. En este contexto, es relevante realizar un análisis espacial de la efectividad de la vacunación y su asociación con factores socioeconómicos. A continuación, se presentan algunos conceptos clave y referencias pertinentes en este campo.

##### **3.1.1 COVID-19**

Los Coronavirus (CoV) son una gran familia de virus que causan enfermedades que van desde el resfriado común hasta enfermedades más graves. El COVID-19 es la enfermedad infecciosa causada por el virus SARS-CoV-2. Las personas infectadas con este virus experimentan una enfermedad respiratoria leve a moderada en la mayoría de los casos; sin embargo, algunas pueden desarrollar síntomas más graves que requieren atención médica urgente. Los grupos de personas que tienen más probabilidad de desarrollar mayores complicaciones con la enfermedad de COVID-19 son los adultos mayores y las personas con enfermedades cardiovasculares, diabetes, enfermedades respiratorias crónicas o cáncer. La Organización Mundial de la Salud (OMS) resalta el riesgo de enfermedad grave o muerte al contraer la enfermedad de COVID-19, sin importar la edad o el estado de salud [6].

##### **3.1.2 EFECTIVIDAD DE LA VACUNACIÓN**

La efectividad de la vacunación es un tema de gran relevancia para la salud pública y ha sido objeto de numerosos estudios a nivel mundial. En el contexto de Colombia, el Ministerio de Salud y Protección Social ha implementado el Plan Nacional de Vacunación, que tiene como objetivo garantizar la administración de dosis de vacunas necesarias para inmunizar a la población, con especial énfasis en los grupos de riesgo [7]. Es importante destacar que este plan ha contemplado estrategias específicas y prioridades claras para su ejecución.

En relación con la evaluación de la efectividad de la vacunación, se han considerado diversos indicadores, entre ellos la prevención de hospitalizaciones, la reducción de la mortalidad en pacientes hospitalizados y la disminución de la mortalidad en casos sin hospitalización previa. Un estudio realizado por el Ministerio de Salud y Protección Social en adultos mayores de 60 años encontró que la vacunación tiene una efectividad del 69% en la prevención de hospitalizaciones, un 79,4% en la reducción de la mortalidad posterior a la hospitalización y un 74,5% en la prevención de la mortalidad en casos sin hospitalización previa [8]. Estos resultados evidencian la importancia y el impacto positivo de las vacunas en la prevención de enfermedades graves y en la disminución de la propagación del virus.

En el ámbito internacional, se ha investigado ampliamente la efectividad de las vacunas, demostrando que constituyen una herramienta crucial para la prevención de enfermedades infecciosas. En el caso de Colombia, se ha evaluado específicamente la efectividad de las vacunas contra el COVID-19 y se ha asegurado la adquisición de las dosis necesarias para llevar a cabo un proceso de vacunación en dos fases y cinco etapas.

### **3.1.3 MODELOS DE ANÁLISIS ESTADÍSTICOS**

Los modelos de análisis estadístico son herramientas matemáticas que se utilizan para describir la variabilidad en los datos y las relaciones entre las variables. Estos modelos varían en su complejidad, desde análisis univariados simples hasta modelos multivariados más complejos. Los modelos de análisis estadístico pueden ser descriptivos, inferenciales o predictivos. Los modelos descriptivos se utilizan para resumir y describir los datos, los modelos inferenciales se utilizan para hacer inferencias, y los modelos predictivos se utilizan para predecir futuros resultados [9].

En el contexto de la efectividad de la vacunación, los modelos de análisis estadístico pueden ayudar a entender los factores que influyen en la efectividad. Por ejemplo, los modelos de supervivencia, que son una clase de modelos de análisis estadístico, se utilizan comúnmente en la investigación de la salud para analizar el tiempo hasta la ocurrencia de un evento de interés. Estos modelos pueden utilizarse para examinar la efectividad de la vacunación en términos de tiempo hasta la muerte después de la vacunación y tiempo hasta la infección antes y después de la vacunación [10].

### **3.1.4 ANÁLISIS ESPACIAL**

El análisis espacial y, en general, el manejo de los Sistemas de Información Geográfica permite la integración de datos acerca de múltiples tipos de características (censos, análisis de zonas de tráfico, calles, hogares, lugares de trabajo, etc.) y sus relaciones. Se emplea el análisis espacial para obtener información nueva y tomar decisiones informadas. Las organizaciones que utilizan el análisis espacial para su trabajo son muy diversas: gobiernos locales y estatales, agencias nacionales, negocios de todo tipo, empresas de suministros, colegios y universidades, organizaciones no gubernamentales, entre otras organizaciones [11].

Mediante el análisis espacial, es posible combinar información de muchas fuentes independientes y obtener nuevos conjuntos de información, aplicando una serie de herramientas espaciales. Este conjunto de herramientas responde a preguntas espaciales complejas, haciendo posible analizar varias capas con el fin de calcular la adecuación de un lugar para una actividad específica. Estas herramientas, y muchas otras que forman parte de los Sistemas de Información Geográfica, permiten abordar cuestiones y tomar decisiones cruciales que van más allá del alcance del simple análisis visual [12].

Según Buzai, las técnicas de análisis espacial representan un compendio de métodos

matemáticos y estadísticos que se utilizan para examinar datos que se distribuyen a lo largo de un área geográfica. Dependiendo del tipo de análisis espacial que se esté realizando, se deben emplear técnicas específicas. Este autor identifica cinco categorías de análisis espacial: ubicación, distribución, asociación, interacción y evolución. Entre las técnicas de análisis espacial que se podrían usar en este estudio son: Densidad de Kernel, análisis de cluster, Índice de autocorrelación espacial de Moran, Indicador local de autocorrelación espacial Lisa y el Índice de cercanías.

De acuerdo con lo anterior, se opta por llevar a cabo un sistema que ejecute diversos procedimientos para lograr un análisis a profundidad de la situación, de los impactos, ventajas, medidas y acciones a desarrollar.

### **3.1.5 FACTORES SOCIOECONÓMICOS**

Los factores socioeconómicos, también conocidos como indicadores socioeconómicos, son las condiciones en los entornos donde las personas nacen, viven, aprenden, trabajan, juegan, adoran y envejecen, y que afectan una amplia gama de resultados de salud, funcionamiento y calidad de vida. Estos factores pueden incluir, pero no están limitados a, la estabilidad económica, el acceso y la calidad de la educación y de la atención médica, el vecindario y el entorno construido, y el contexto social y comunitario. Por ejemplo, se incluyen la vivienda segura, el transporte, los barrios, la educación, las oportunidades laborales, el ingreso, el acceso a alimentos nutritivos y oportunidades de actividad física, y las habilidades de lenguaje y alfabetización [13].

El estatus socioeconómico (SES, por sus siglas en inglés) es una medida que se utiliza para clasificar a las personas o comunidades en términos de su acceso a recursos económicos y sociales. Este indicador es relevante porque influye en muchos aspectos de la vida de las personas, desde su salud y bienestar hasta su capacidad para acceder a oportunidades y servicios. De acuerdo con la Organización Mundial de la Salud, el SES incluye, entre otros factores, la educación, el empleo y los ingresos, y su importancia radica en el hecho de que es un determinante importante del acceso a una variedad de recursos de salud, incluida la posibilidad de obtener una atención sanitaria adecuada, justa y asequible.

En cuanto a la relación de los factores socioeconómicos con la efectividad de la vacunación, varios estudios han encontrado correlaciones. Un estudio en Israel determinó que el porcentaje de vacunación contra el COVID-19 estaba significativamente correlacionado con el estatus socioeconómico municipal. Las áreas con menor SES y mayor carga de enfermedad activa tenían un menor porcentaje de vacunación [14]. En los Estados Unidos, los condados con tasas de iniciación de vacunación inferiores al 50% tenían porcentajes significativamente más altos de adultos mayores con vulnerabilidades sociales que los condados con tasas de iniciación del 75% o más. Estas vulnerabilidades sociales pueden incluir la falta de acceso a un ordenador o a Internet, vivir en la pobreza o vivir solo [15].

### 3.1.6 ÍNDICE DE POBREZA MULTIDIMENSIONAL

El Índice de Pobreza Multidimensional (IPM) emerge como un instrumento esencial en la evaluación y comprensión de la pobreza, trascendiendo la simple medición basada únicamente en ingresos. En el contexto colombiano, su implementación representa un progreso significativo en la comprensión de las múltiples dimensiones de la pobreza.

El IPM en Colombia, basado en la metodología desarrollada por Alkire y Foster, adopta una perspectiva multidimensional que integra aspectos económicos, sociales y de calidad de vida. Esta metodología fue adaptada por el Departamento Nacional de Planeación (DNP) y transferida al Departamento Administrativo Nacional de Estadística (DANE) en 2012, consolidándose como un instrumento oficial para la medición de la pobreza en el país [16].

Las cinco dimensiones y variables que conforman el IPM en Colombia se presentan en la tabla 1:

Tabla 1 “Dimensiones y variables del IPM para Colombia.”

Dimensión	Variable		Puntos de corte
	Privación	Indicador	
Condiciones educativas del hogar (0.2)	Bajo logro educativo (0.1)	Escolaridad promedio de las personas de 15 años y más del hogar	9 años
	Analfabetismo (0.1)	Porcentaje de personas del hogar de 15 años y más que saben leer y escribir	100%
Condiciones de la niñez y la juventud (0.2)	Inasistencia escolar (0.05)	Proporción de niños entre 6 y 16 años en el hogar que asisten al colegio	100%
	Rezago escolar (0.05)	Proporción de niños y jóvenes (7-17 años) dentro del hogar sin rezago escolar (según la norma nacional)	100%
	Barreras de acceso a servicios para el cuidado de la primera infancia (0.05)	Proporción de niños de cero a cinco años en el hogar con acceso simultáneo a salud, nutrición y educación inicial	100%
Trabajo (0.2)	Trabajo infantil (0.05)	Proporción de niños entre 12 y 17 años en el hogar que se encuentra por fuera del mercado laboral	100%
	Desempleo de larga duración (0.1)	Proporción de la PEA del hogar que no se encuentra en desempleo de larga duración (más de 12 meses)	100%
	Empleo informal (0.1)	Proporción de la PEA del hogar que son ocupados con afiliación a pensiones (proxy de informalidad)	100%
Salud (0.2)	Sin aseguramiento en salud (0.1)	Proporción de miembros del hogar, mayores de cinco años, asegurados a Seguridad Social en Salud	100%
	Barreras de acceso a servicio de salud (0.1)	Proporción de personas del hogar que acceden a servicio institucional de salud ante una necesidad sentida	100%
Acceso a servicios públicos domiciliarios y condiciones de la vivienda (0.2)	Sin acceso a fuente de agua mejorada (0.04)	Hogar urbano: Se considera como privado si no tiene servicio público de acueducto en la vivienda. Hogar Rural: Se considera privado cuando obtiene el agua para preparar los alimentos de pozo sin bomba, agua lluvia, río, manantial, carro tanque, aguatero u otra fuente.	1
	Inadecuada eliminación de excretas (0.04)	Hogar urbano: Se considera como privado si no tiene servicio público de alcantarillado. Hogar rural: Se considera privado si tiene inodoro sin conexión, bajamar o no tiene servicio sanitario	1
	Pisos inadecuados (0.04)	Se consideran en privación los hogares que tienen pisos en tierra.	1
	Paredes exteriores inadecuadas (0.04)	Hogar urbano: Se considera privado si el material de las paredes exteriores es madera burda, tabla, tablón, guadua, otro vegetal, zinc, tela, cartón, deshechos o sin paredes. Hogar rural: Se considera privado si el material de las paredes exteriores es guadua, otro vegetal, zinc, tela, cartón, deshechos o sin paredes.	1
	Hacinamiento crítico (0.04)	Número de personas por cuarto para dormir excluyendo cocina, baño y garaje e incluyendo sala y comedor	*Urbano: 3 o más personas por cuarto *Rural: Más de 3 personas por cuarto

Fuente: Departamento nacional de planeación, Archivos de economía (DNP, DDS, SPSCV), 2011.

Estas dimensiones abarcan 15 variables cruciales de la vida cotidiana de las personas y familias en situación de pobreza, desde las condiciones educativas del hogar hasta el acceso a servicios públicos domiciliarios. Ofrecen una visión integral de las privaciones que enfrentan las comunidades vulnerables. En esta medición en Colombia, se emplea una estructura de ponderación anidada, donde cada dimensión y cada indicador tienen igual peso (20%) en el índice. De esta manera, un hogar, y todas las personas que lo componen, se consideran multidimensionalmente pobres cuando presentan un IPM de al menos 33.3%, según el punto de corte definido para cada indicador.

### 3.1.7 MODELO ECOLÓGICO DE POISSON

El Modelo Ecológico de Poisson es una herramienta estadística ampliamente utilizada para analizar datos de conteo, particularmente en el contexto de incidencia de enfermedades y la evaluación de intervenciones de salud pública como la vacunación [17]. Este modelo se basa en la distribución de Poisson, que es adecuada para datos de conteo que cumplen con ciertas suposiciones, incluida la aleatoriedad de los eventos y una tasa constante de ocurrencia a lo largo del tiempo.

En la investigación actual, se ha aplicado el Modelo Ecológico de Poisson para estimar la efectividad de la vacunación por medio del impacto a escala de barrio. Mediante el análisis de datos de conteo de casos de fallecimiento por COVID-19 en relación con la cobertura de vacunación, se pretende identificar y cuantificar la asociación entre la vacunación y la reducción de la incidencia de la enfermedad en la población estudiada.

El modelo se define matemáticamente de la siguiente manera:

$$\log(\lambda_t) = \beta_0 + \beta_1 \times \text{plan\_vacuna}_t$$

- $\lambda_t$  es la tasa de incidencia de la enfermedad en el periodo de tiempo  $t$
- $\beta_0$  es el intercepto del modelo.
- $\beta_1$  es el coeficiente de regresión que representa el impacto de la vacunación en la tasa de incidencia de la enfermedad.
- $\text{plan\_vacuna}_t$  es la variable indicadora de la implementación del plan de vacunación en el periodo de tiempo  $t$

### 3.1.8 ÍNDICE DE MORAN

El Índice de Moran se utiliza en el análisis espacial para medir la autocorrelación de datos geográficos. El Índice de Moran Global evalúa la autocorrelación espacial general de un conjunto de datos, proporcionando un valor que varía entre -1 y 1, donde valores cercanos a 1 indican autocorrelación positiva y cercanos a -1 indican autocorrelación negativa. Por otro lado, el Índice de Moran Local (LISA) analiza la autocorrelación a nivel local, identificando clústeres espaciales de valores similares en áreas específicas. Los

LISA se representan comúnmente en mapas para visualizar áreas de alta-baja y baja-alta autocorrelación, lo que permite identificar patrones espaciales localizados y áreas de interés en el análisis. Ambos índices son herramientas clave para comprender la distribución y relación de los datos en el espacio geográfico [18].

### **3.1.9 SIMULACIÓN DE MONTECARLO**

La Simulación de Montecarlo es una técnica estadística que utiliza múltiples simulaciones con valores aleatorios para las variables de entrada de un modelo, permitiendo estimar la distribución de probabilidad de un resultado de interés y entender la variabilidad e incertidumbre asociada [19].

### **3.1.10 MODELO MULTIVARIADO DE GEARY LOCAL**

El Modelo Multivariado de Geary Local es una extensión del Índice de Geary que evalúa la autocorrelación espacial de múltiples variables en un conjunto de datos geográficos. Calcula estadísticas de autocorrelación local para cada variable, comparando los valores de una unidad espacial con los de sus vecinos ponderados. Los resultados varían entre 0 y 2: valores cercanos a 0 indican autocorrelación positiva localizada, cerca de 2 indican autocorrelación negativa, y cerca de 1 sugiere una distribución espacial aleatoria [20].

### **3.1.11 MODELO DE REGRESION ESPACIAL**

El Modelo de Regresión Espacial es una técnica estadística que incorpora la autocorrelación espacial en el análisis de regresión, permitiendo capturar y modelar la dependencia espacial entre las observaciones [21].

**Modelo Básico:** En el Modelo de Regresión Espacial Básico, se incluye un término de dependencia espacial, usualmente a través de una matriz de pesos espaciales, para considerar la influencia de las observaciones vecinas en la variable dependiente. Este modelo asume que todos los errores son idénticamente distribuidos y no considera la autocorrelación en los errores [21].

**Modelo de Retardo Espacial:** El Modelo de Retardo Espacial añade un término de retardo espacial de la variable dependiente al modelo, lo que implica que el valor actual de la variable dependiente está influenciado por los valores pasados de la misma variable en las unidades espaciales vecinas. Este modelo captura la autocorrelación espacial en la variable dependiente [21].

**Modelo de Error Espacial:** El Modelo de Error Espacial incorpora la autocorrelación espacial en los errores del modelo, permitiendo que los errores de observaciones vecinas estén correlacionados. Este modelo es útil cuando la autocorrelación espacial no se captura completamente en la variable dependiente o cuando se sospecha que hay una estructura espacial en los errores del modelo [21].

### **3.1.12 ESTRATO SOCIOECONÓMICO**

El estrato socioeconómico se define como una categorización multidimensional utilizada para clasificar a individuos, hogares o áreas geográficas según su nivel de ingresos, educación, ocupación y acceso a servicios básicos y recursos socioeconómicos. En el contexto colombiano, esta clasificación juega un papel significativo en la comprensión de las disparidades socioeconómicas y en la formulación de políticas públicas destinadas a abordarlas.

Este concepto se fundamenta en la capacidad de pago de los usuarios de servicios públicos domiciliarios, y su aplicación se ha consolidado a lo largo del tiempo como una herramienta administrativa y de planificación. Sin embargo, es importante tener en cuenta que los estratos socioeconómicos no solo reflejan las diferencias objetivas en los niveles de ingresos y condiciones de vida, sino que también pueden influir en la percepción y construcción de identidades sociales, así como en la distribución desigual de recursos y oportunidades en la sociedad.

En este sentido, el análisis del estrato socioeconómico implica una comprensión profunda de las dinámicas sociales, económicas y políticas que subyacen a las desigualdades estructurales en Colombia. Además, su estudio proporciona una base teórica sólida para investigaciones interdisciplinarias que aborden temas relacionados con la equidad, la movilidad social y el desarrollo humano en el país. (Ley 732 del 25 de enero de 2002) [22].

### **3.1.13 DENSIDAD POBLACIONAL**

La densidad poblacional es un indicador demográfico fundamental que describe el número medio de habitantes por unidad de superficie en una región específica. En el contexto colombiano, se calcula dividiendo la población total de un área geográfica determinada, como un departamento o municipio, por su extensión territorial en kilómetros cuadrados.

Este indicador proporciona información crucial sobre la distribución espacial de la población, destacando áreas con alta concentración de habitantes, como zonas urbanas, y áreas con menor densidad poblacional, como zonas rurales. Su análisis es esencial para la planificación urbana y territorial, así como para la formulación de políticas públicas relacionadas con infraestructura, servicios básicos y desarrollo socioeconómico.

Asimismo, la densidad poblacional puede tener repercusiones significativas en la dinámica social, económica y ambiental de un área, influyendo en aspectos como la disponibilidad de empleo, la calidad de vida y la sostenibilidad ambiental. Por lo tanto, comprender este indicador es fundamental para abordar desafíos relacionados con el crecimiento urbano, la equidad territorial y la gestión de recursos en Colombia [23].

### **3.1.14 INDICE DE VULNERABILIDAD**

El nivel de vulnerabilidad se define como la medida en la que individuos, hogares o comunidades enfrentan riesgos y carencias que amenazan su bienestar y capacidad de recuperación ante eventos adversos. En Colombia, el Departamento Administrativo Nacional de Estadística (DANE) ha desarrollado un Índice de Vulnerabilidad que emplea diversos indicadores socioeconómicos para evaluar el grado de vulnerabilidad en

diferentes niveles geográficos.

Este índice considera una amplia gama de factores, como ingresos, acceso a servicios básicos, calidad de la vivienda, educación, salud y empleo, entre otros. Al combinar estos indicadores, se obtiene una medida compuesta que facilita la identificación y caracterización de las poblaciones más vulnerables, así como la comprensión de los principales factores de riesgo que enfrentan.

Es importante tener en cuenta que el nivel de vulnerabilidad es un concepto dinámico que puede variar según el contexto socioeconómico, geográfico y temporal. Por lo tanto, su análisis y abordaje son fundamentales para el diseño e implementación de políticas y programas de desarrollo que promuevan la resiliencia y la inclusión social, especialmente para aquellos grupos y comunidades más afectados por la pobreza, la marginalización y la exclusión [24].

### **3.2 ANTECEDENTES**

En el campo de la ciencia de datos se están desarrollando constantemente nuevas técnicas de análisis y modelos metodológicos que permiten utilizar la información de manera oportuna en la toma de decisiones en el sector de la salud. Estos avances no solo se aplican en aspectos clínicos, sino también en políticas y acciones de salud pública. Por lo tanto, es valioso considerar las contribuciones que ha hecho la ciencia de datos y el análisis espacial a la toma de decisiones adecuadas en el manejo de virus emergentes y reemergentes, utilizando datos relacionados con diversas enfermedades, comportamientos y patrones.

En la ciudad de Nueva York se llevó a cabo un estudio denominado “Spatial analysis of COVID-19 clusters and contextual factors in New York City”. En él, se resalta la necesidad de identificar áreas con limitado acceso a pruebas y una alta carga de casos para comprender el riesgo y asignar recursos de manera adecuada durante la pandemia. Mediante el análisis de datos a nivel de código postal, encontraron grupos con tasas altas y bajas de pruebas, tasas de positividad elevadas y una proporción positiva significativa. Se observó una asociación entre estos resultados, los grupos identificados y los factores contextuales [25].

Este estudio demuestra cómo las técnicas de análisis estadísticos y el análisis exploratorio de datos espaciales permiten identificar asociaciones en las que sobresale la importancia de considerar los factores socioeconómicos y el análisis espacial al abordar las inequidades en la atención y el acceso a las pruebas. Sin embargo, su alcance se centra en el análisis espacial de los conglomerados de COVID-19, mientras que la presente investigación busca profundizar en el análisis espacial de la efectividad de la vacunación contra COVID-19.

En el estudio de “Análisis espacio-temporal del SARS-COV-2 en Cali, Colombia”, utilizaron técnicas de análisis espacial para identificar agrupaciones, predecir áreas potenciales de concentración de casos, modelar las fuerzas que atraen a los individuos, identificar centralidades y evaluar el potencial de contacto y transmisión

del virus. Allí se plantea que es importante seguir los pasos del análisis espacial, que incluyen definir el área de estudio, realizar un análisis exploratorio de datos espaciales, llevar a cabo el proceso de geocodificación y realizar el análisis de puntos para comprender las dinámicas del virus [26].

Su aporte reside en la comprensión de la dinámica espacial del virus en Santiago de Cali, determinando la escala de barrio como un buen método de distribución para este tipo de estudios. No obstante, se diferencia de esta investigación en cuanto a su enfoque principal en la propagación del virus en general, en lugar de centrarse en la efectividad de la vacunación.

Por su parte, en el proyecto denominado "Spatial modeling and socioeconomic inequities of Covid 19 in the urban area of the city of Cali", se plantea que el COVID-19 no solo ha impactado la salud a nivel global, sino que también ha exacerbado las desigualdades socioeconómicas existentes. Los grupos con mayores desventajas sociales tienen más probabilidades de estar expuestos a condiciones que aumentan el riesgo de enfermedad y muerte. Factores como el sexo, la educación, los ingresos, la etnia, la posición socioeconómica, la ocupación y las condiciones de vivienda han demostrado tener una relación significativa con numerosos resultados de salud [27].

La contribución de este proyecto se basa en el uso de modelos espaciales y en la influencia de factores socioeconómicos en la pandemia de COVID-19 en el contexto de Cali. Este estudio difiere con respecto a que su enfoque principal es el análisis espacial del COVID-19, y no la efectividad de la vacunación.

A nivel nacional, el Ministerio de Salud desarrolló un estudio denominado "Efectividad de las vacunas contra el COVID-19 en Colombia estudio de cohorte de base poblacional en adultos de 60 años y más", el cual busca estimar la efectividad de las vacunas contra el COVID-19 para prevenir la hospitalización y la muerte de adultos mayores con esquemas completos de vacunación. Tomó como población objetivo la de mayores de 60 años y desarrolló análisis estadísticos para evaluar la efectividad de la vacunación en esta población [28].

Este estudio muestra un panorama sobre los modelos de efectividad aplicados que pueden orientar el desarrollo de esta investigación a realizar en Cali, la cual se hará con toda su población e incluyendo variables socioespaciales.

#### 4. METODOLOGÍA

Para determinar las etapas de desarrollo, esta investigación se fundamenta en la metodología CRISP-DM (Cross Industry Standard Process for Data Mining). Es un proceso estándar y ampliamente aceptado para la minería de datos, que proporciona un enfoque estructurado y sistemático para abordar problemas de análisis de datos y descubrimiento de conocimiento. Esta metodología consta de seis etapas principales: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue [29].

Inicialmente, la etapa de comprensión del negocio implica la definición clara de los objetivos de investigación y la comprensión de los requisitos y contexto del problema. Se establecerán las bases de datos y variables que permitan analizar la efectividad de la vacunación contra el COVID-19 y su asociación con factores socioeconómicos en Santiago de Cali.

Dentro de la etapa de comprensión de los datos, se recopilarán los conjuntos de datos relevantes relacionados con la vacunación contra el COVID-19 y los factores socioeconómicos en Santiago de Cali. Se realizará una exploración inicial de los datos para comprender su estructura y calidad.

La etapa de preparación de los datos será fundamental para garantizar que los datos sean adecuados y limpios para el análisis. Se llevarán a cabo tareas como el filtrado de variables irrelevantes, el manejo de valores faltantes y la normalización de los datos. Esto asegurará la calidad y consistencia de los datos para el análisis posterior.

Posteriormente, se procederá a la etapa de modelado, donde se aplicarán técnicas y algoritmos de minería de datos para construir modelos que permitan analizar la efectividad de la vacunación y su relación con los factores socioeconómicos. Se seleccionarán las técnicas más apropiadas según los objetivos de la investigación y se ajustarán los parámetros de los modelos.

Una vez construidos los modelos, se pasará a la etapa de evaluación, donde se medirá el rendimiento de los modelos y se determinará su capacidad para cumplir con los objetivos de la investigación. Se utilizarán métricas y técnicas de validación para evaluar el rendimiento de los modelos y realizar ajustes en caso de ser necesario.

Finalmente, en la etapa de despliegue, se presentarán los resultados y hallazgos obtenidos a partir de los modelos desarrollados. Se elaborarán informes y se comunicarán los resultados de manera clara y comprensible, para su utilización en la toma de decisiones relacionadas con la efectividad de la vacunación contra el COVID-19 en Santiago de Cali y su asociación con los factores socioeconómicos.

De esta forma, la metodología CRISP-DM proporciona un marco sólido y estructurado para llevar a cabo esta investigación, asegurando un enfoque riguroso y sistemático en cada una de las etapas del proceso de análisis de datos.

## **5. ENTENDIMIENTO DEL NEGOCIO**

La pandemia de COVID-19 transformo el panorama mundial, afectando a la salud pública, la economía y el bienestar social de las comunidades. La vacunación se presenta como la principal herramienta para combatir la propagación del virus y proteger a la población. En Santiago de Cali, una ciudad con una diversidad socioeconómica y geográfica significativa, se han implementado campañas de vacunación con el objetivo de inmunizar a la mayor cantidad posible de habitantes. Sin embargo, es esencial comprender cómo se llevó a cabo esta vacunación y cuál es su efectividad real, especialmente en relación con factores socioeconómicos y geográficos, para diseñar estrategias más efectivas y equitativas.

### **5.1 IMPORTANCIA DEL ESTUDIO**

El análisis de la efectividad de la vacunación y su asociación con factores socioeconómicos en Santiago de Cali es crucial por varias razones:

**Optimización de Recursos:** Entender qué áreas o grupos socioeconómicos tienen menor cobertura de vacunación permite dirigir recursos de manera más eficiente.

**Equidad en la Salud:** Identificar desigualdades en la distribución de la vacunación ayuda a garantizar que todos los grupos poblacionales tengan acceso igualitario a la protección contra el virus.

**Toma de Decisiones:** Los resultados del estudio proporciona información valiosa para los responsables políticos y las autoridades de salud pública para tomar decisiones informadas y basadas en evidencia.

El objetivo principal de este estudio es realizar un Análisis Espacial para la Efectividad de la Vacunación por COVID-19 en Santiago de Cali y modelar su asociación con factores socioeconómicos a través de modelos de aprendizaje estadístico. Este análisis permitirá entender mejor la distribución geográfica de la efectividad de la vacunación, identificar áreas de mejora y determinar cómo diferentes factores socioeconómicos pueden influir en la efectividad de las campañas de vacunación.

Establecer la efectividad de la vacunación por medio de modelos de supervivencia a escala espacial de barrio en la ciudad de Santiago de Cali. Permite entender cómo la efectividad de la vacunación varía geográficamente en diferentes barrios identificando áreas con menor cobertura y evaluar los factores asociados.

Construir modelos estadísticos multivariados explicativos para la efectividad en función de los factores socioeconómicos y con dependencia espacial. Analizando la relación entre la efectividad de la vacunación y variables socioeconómicas como el nivel de ingresos, el acceso a servicios de salud y la densidad de población, considerando la dependencia espacial de los datos.

## 5. 2 METRICAS DE EVALUACION

Para evaluar el éxito del estudio, se utilizarán las siguientes métricas:

**Precisión de los modelos:** Medir la capacidad de los modelos para predecir la efectividad de la vacunación y su asociación con factores socioeconómicos con un alto grado de precisión.

**Interpretabilidad de los resultados:** Asegurar que los resultados sean claros, comprensibles y relevantes para los tomadores de decisiones y los interesados del proyecto.

**Relevancia de las recomendaciones:** Confirmar que las recomendaciones propuestas sean prácticas, aplicables y alineadas con las necesidades y contextos específicos de Santiago de Cali.

## 5.3 RESTRICCIONES Y LIMITACIONES

**Variabilidad en la implementación de la vacunación:** Las diferencias en la estrategia de vacunación entre diferentes áreas o períodos de tiempo pueden introducir ruido o sesgos en los datos.

**Factores externos:** Cambios en las políticas de vacunación, la aparición de nuevas variantes del virus o la evolución de la situación epidemiológica pueden influir en los resultados y conclusiones del estudio.

## 6. ENTENDIMIENTO DE LOS DATOS

### 6.1 SOLICITUD FORMAL DE LOS DATOS

Se evaluó la necesidad de contar con un conjunto de datos que tuviera información de casos de COVID-19 a escala espacial de barrio y contara con fechas de inicio de síntomas y resultados definitivos para la Ciudad Santiago de Cali. Se realizó la solicitud formal de los datos a la secretaria de salud obteniendo la aprobación y de esta forma las copias y accesos necesarios para descargar las bases de datos. Ver anexo 1.

Otras fuentes de datos relacionadas a el índice de pobreza multidimensional (IPM), nivel de vulnerabilidad fueron obtenidas de bases de datos publicas dispuestas por entidades gubernamentales como el DANE.

Los datos obtenidos se organizaron en dos tablas distintas representadas en la tabla 2:

*Tabla 2 Variables covid\_tg*

covid_tg	
Variable	Descripción
Documento	Documento de identificación
Ubicación	Ubicación de la persona, puede ser casa, hospitalización, uci
Fecha_muestra	Fecha de toma de la muestra
fecha_resultado	Fecha de resultado de la muestra
iniciodeSíntomas_	Fecha de inicio de síntomas
No	Número de registro en la base de datos
tip_ide_	Tipo de identificación
edad_	Edad de la persona
Sexo	Sexo de la persona
Comuna	Comuna a la que pertenece
Ubicacióngeo	Ubicación geográfica
bar_ver_	Nombre del barrio o vereda
dir_res_	Dirección de residencia
Aseguradora	Nombre de la aseguradora
tip_ss_	Tipo de aseguramiento
fec_consulta	Fecha de la consulta
Estado	Estado de la afección
Fecha de hospitalización	Fecha de hospitalización
IPSHosp	IPS de la hospitalización
Fecha de alta	Fecha dada de alta
Fecha de muerte	Fecha de muerte
Comorbilidades	Consulta de presencia de comorbilidades
fec_inguci	Fecha de ingreso a unidad de cuidados intensivos (UCI)
nom_eve	Nombre del evento
nom_upgd	Nombre de la unidad primaria generadora de datos
Fecha_Nacimiento	Fecha de nacimiento de la persona
Prueba_PCR negativa	Resultado negativo de la prueba PCR
Fecha_muestra_recuperado	Valor recuperado de la fecha de muestra
Fecha_Resultado_pcr_negativo	Fecha resultado negativo de la prueba PCR
No_pruebas	Número de pruebas realizadas
Fecha_notifi	Fecha de notificación
Fecha de captacion	Fecha de captación
Probable_Trabajador_Salud	Consulta sobre posible trabajador de la salud
Fecha recuperado por certificacion EPS-IPS	Fecha recuperado certificada por la EPS-IPS
Tipo de recuperado	Tipo de recuperación
Fecha recuperado	Fecha de recuperación
Recuperado definitivo	Resultado definitivo de la recuperación, puede ser hospitalizado, recuperado o fallecido
Paciente ha estado hospitalizado	Consulta si el paciente ha estado hospitalizado
sintomatico	Persona sintomática
etnia	Etnia de la persona
Fecha de hospitalizacion SIVIGILA	Fecha de hospitalización registrada en el Sistema Nacional de Vigilancia de en la Salud Pública
Latitud - X	Valor de latitud X
Longitud Y	Valor de longitud Y
fec_vac	Fecha de vacunación
num_dos	Número de dosis
nom_vac	Nombre de la vacuna

*fuentes: Adaptado de: SISTEMA NACIONAL DE VIGILANCIA EN SALUD PÚBLICA Subsistema de información SIVIGILA Ficha de notificación individual – Datos complementarios Infeción respiratoria aguda por virus nuevo. Ficha Cod INS 346*

La base de datos "covid\_tg" constituye un recurso fundamental en el estudio dado que alberga registros detallados de casos confirmados de COVID-19, recopilando información esencial para comprender la evolución de la pandemia en esta región específica.

Uno de los aspectos más destacados de esta base de datos es el campo "Recuperado definitivo", el cual brinda una visión integral sobre el estado de salud de los pacientes afectados. Este campo presenta tres categorías principales: Hospitalizado, recuperado y fallecido.

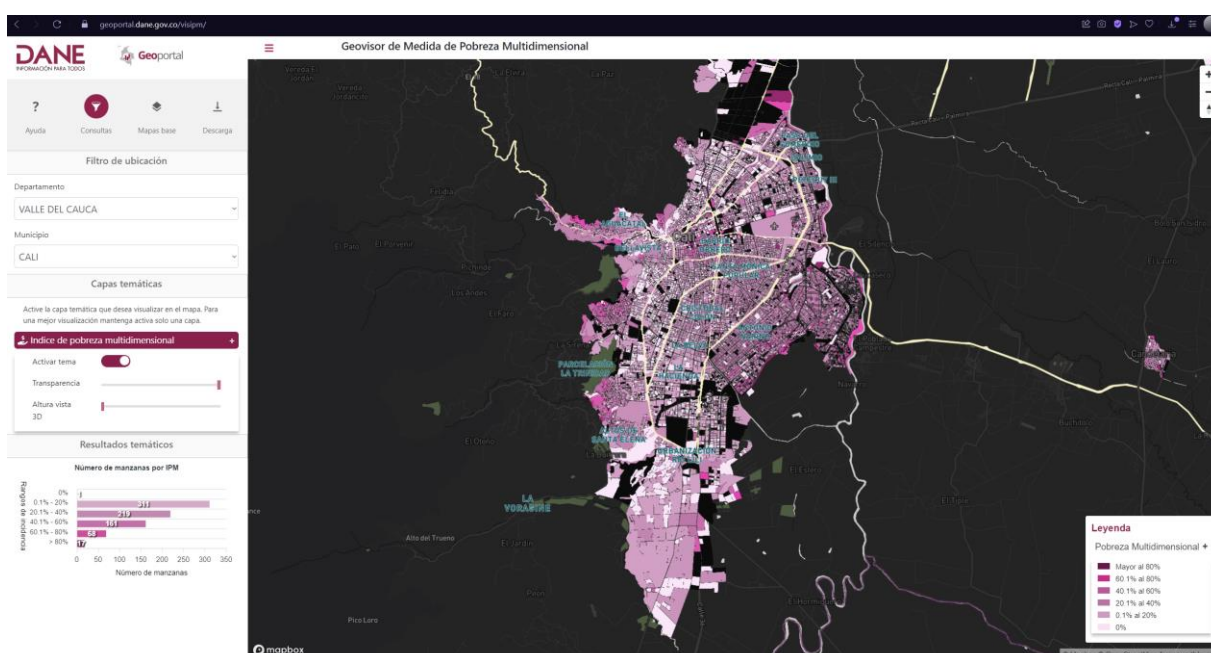
Otro dato de relevancia es el campo "Inicio de Síntomas", el cual registra la fecha en la que los pacientes experimentaron por primera vez síntomas atribuibles al COVID-19. Este dato permite analizar la evolución temporal de la persona y su impacto en la comunidad a lo largo del tiempo. Los valores de este campo oscilan desde marzo de 2020 hasta febrero de 2023, abarcando así un amplio espectro temporal.

Además, para facilitar el análisis espacial de los casos, la base de datos incluye los campos "Latitud" (X) y "Longitud" (Y), los cuales proporcionan las coordenadas geográficas de cada caso registrado. Estos datos permiten la georreferenciación de los casos, lo que posibilita el mapeo y la visualización de la distribución geográfica de la enfermedad en la ciudad de Santiago de Cali.

## 6.2 DATOS DANE

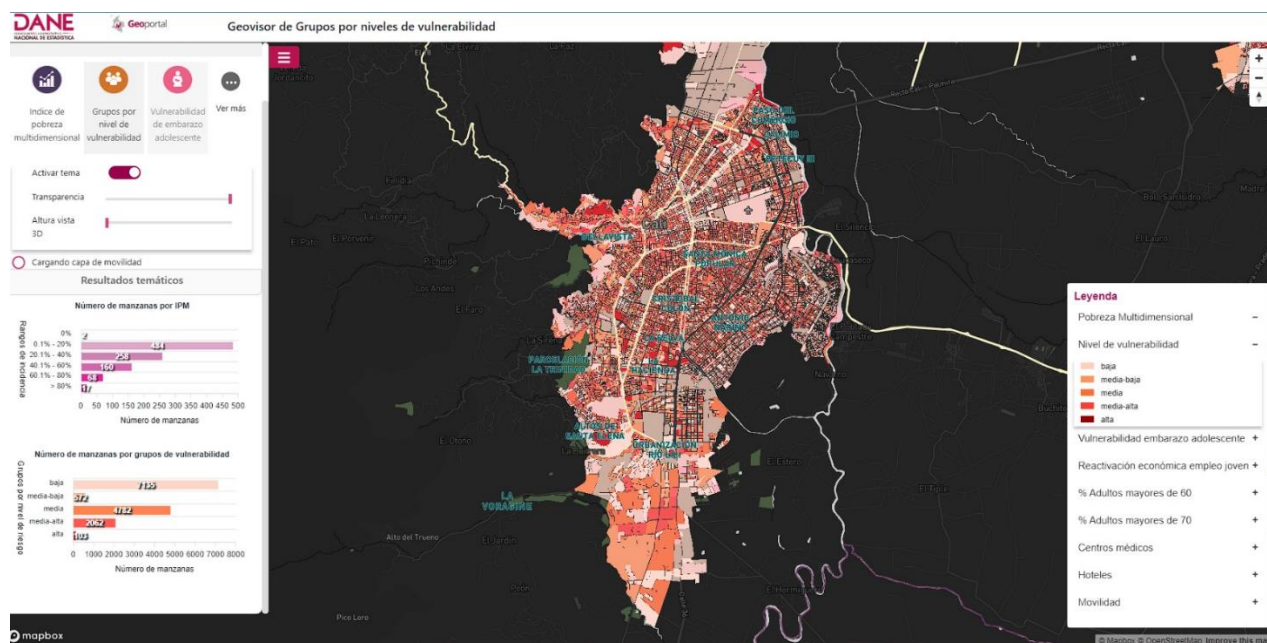
La base de datos utilizada en este proyecto fue obtenida del Departamento Administrativo Nacional de Estadística (DANE) de Colombia, a través de su geoportal disponible en la web Geoporta. El cual proporciona acceso al Geovisor de Medida de Pobreza Multidimensional (IPM) del DANE, ver ilustración 1 y medida de niveles de vulnerabilidad, ver ilustración 2, una herramienta fundamental para la recopilación y análisis de datos geospaciales relacionados con la pobreza multidimensional en el país.

Ilustración 1 Geovisor de Medida de Pobreza Multidimensional



Fuente: <https://geoportal.dane.gov.co/visor-vulnerabilidad/>

Ilustración 2 Geovisor de Medida de niveles de vulnerabilidad



Fuente: <https://geoportal.dane.gov.co/visor-vulnerabilidad/>

El Geovisor de IPM ofrece una interfaz interactiva que permite visualizar y explorar datos geográficos sobre la pobreza multidimensional, facilitando así el entendimiento de su distribución espacial y su evolución temporal. Esta plataforma fue esencial para el desarrollo del presente estudio, proporcionando los datos necesarios para llevar a cabo el análisis y la generación de resultados significativos.

La base de datos del Departamento Administrativo Nacional de Estadística (DANE) ofrece una panorámica detallada de la dinámica socioeconómica en Colombia. Destaca por incluir variables cruciales que abordan la vulnerabilidad y las condiciones de vida de la población, entre las que se encuentra el Índice de Pobreza Multidimensional (IPM). Este índice, al evaluar la pobreza desde diversas dimensiones, proporciona una visión integral de las condiciones socioeconómicas en el país. Además, la base de datos del DANE analiza la vulnerabilidad relacionada con el embarazo en la adolescencia y la reactivación económica en términos de empleo para la población joven. En resumen, la inclusión del IPM y otras variables relevantes en la base de datos del DANE ofrece una base sólida para el análisis de los factores socioeconómicos. Ver tabla 3.

Tabla 3 Variables DANE

DANE	
Variable	Descripción
COD_DPTO	Código del departamento
COD_MPIO	Código del municipio
COD_DANE	Código DANE
CATEGORIA	Categoría del registro
LABEL	Nivel de vulnerabilidad
ipm	Índice de pobreza multidimensional
embarazo_a	Vulnerabilidad embarazo adolescente
reactivaci	Reactivación económica empleo joven

Fuente: Adaptado de <https://geoportal.dane.gov.co/visor-vulnerabilidad/>

## 6.3 CARACTERÍSTICAS SOFTWARE SELECCIONADO

Para abordar esta tarea de manera efectiva, se requiere el uso de herramientas de software especializadas. En este contexto, se destacan varias plataformas que desempeñan un papel fundamental en la preparación de datos para este estudio. A continuación, se presenta una caracterización de estas herramientas seleccionadas por su relevancia en la investigación.

### 6.3.1 PostgreSQL con extensión de QGIS:

- **PostgreSQL:** Reconocido como uno de los sistemas de gestión de bases de datos relacionales más robustos y confiables, PostgreSQL ofrece una plataforma sólida para el almacenamiento y manipulación de conjuntos de datos a gran escala. Su capacidad para gestionar datos geoespaciales lo hace especialmente relevante para la investigación que involucra análisis espaciales [30].
- **QGIS:** Como sistema de información geográfica de código abierto, QGIS proporciona las herramientas necesarias para visualizar, analizar y comprender datos geoespaciales. La integración de QGIS con PostgreSQL permite una exploración profunda de la distribución geográfica de la vacunación y su correlación con factores socioeconómicos [31].
- **GeoDa:** Es una herramienta especializada en análisis espacial que es beneficiosa para el proyecto debido a que ofrece una amplia gama de herramientas para el análisis exploratorio y modelado espacial, permitiéndote identificar patrones espaciales significativos, construir modelos espaciales de regresión y visualizar datos geográficos en mapas interactivos. Su interfaz amigable y su integración con otras herramientas facilitan su uso [32].

### 6.3.2 PYTHON

Python se ha establecido como un lenguaje de programación versátil y potente en el campo de la ciencia de datos. Sus bibliotecas especializadas, como Pandas, NumPy y Scikit-learn, ofrecen una amplia gama de funcionalidades para el procesamiento, análisis y modelado de datos. Python es una elección ideal para la preparación de datos debido a su sintaxis clara, capacidades de manipulación de datos y soporte comunitario activo [33].

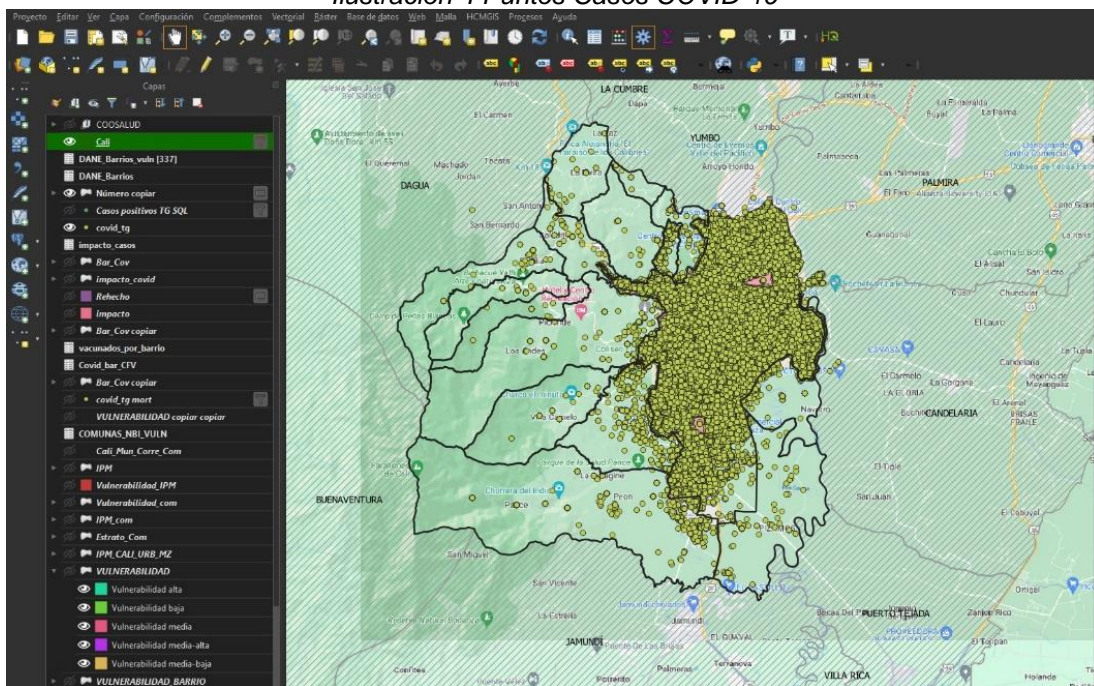
### 6.3.3 RSTUDIO

RStudio sirve como un entorno de desarrollo integral para el lenguaje de programación R, ampliamente reconocido en la comunidad estadística y de análisis de datos. Su interfaz intuitiva y las numerosas bibliotecas disponibles hacen que RStudio sea ideal para la exploración de datos, análisis estadístico y generación de informes reproducibles. En el contexto de la preparación de datos, RStudio facilita la limpieza, transformación y manipulación de datos de manera eficiente [34].



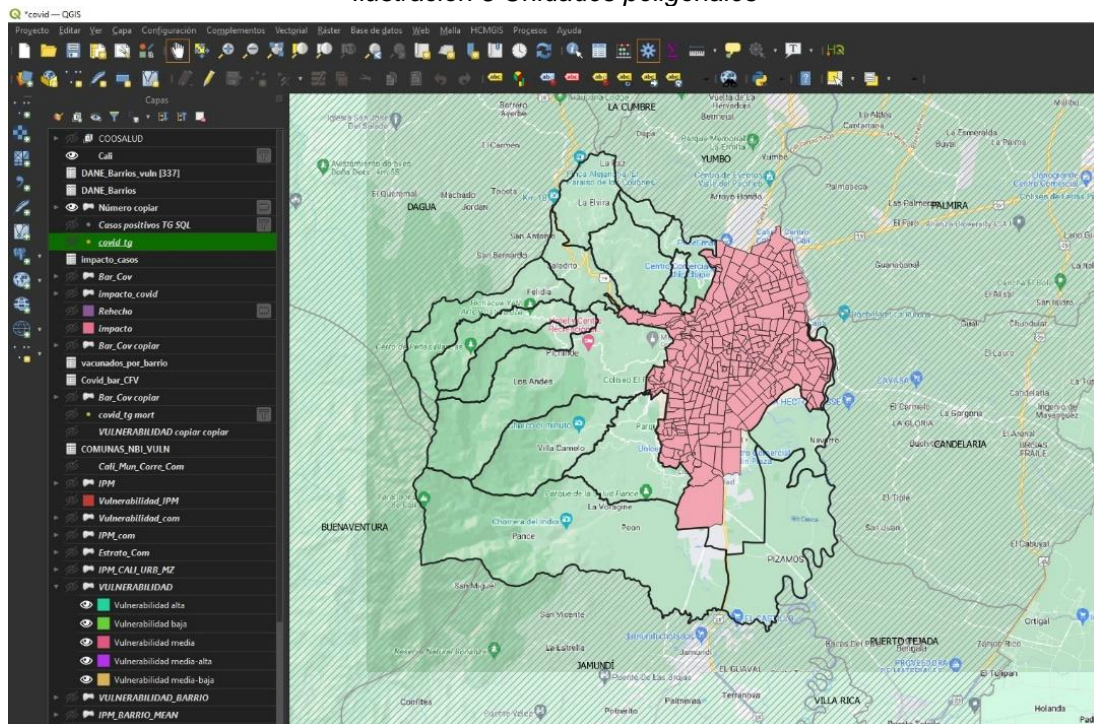
ilustración 5 en el cual se hizo un conteo de puntos por unidades poligonales las cuales fueron convertidos en polígonos a través de geo procesos. Esta conversión permitió representar los datos en forma de áreas geográficas definidas, lo que facilita la identificación de patrones espaciales y la realización de análisis más avanzados sobre la propagación del virus y su impacto en diferentes áreas de la ciudad.

Ilustración 4 Puntos Casos COVID-19



Fuente: Propia

Ilustración 5 Unidades poligonales



Fuente: Propia

## 7.1 PREPARACIÓN COVID\_TG

Una vez obtenida la georreferenciación la base de datos de " covid\_tg " a escala de barrio, se realiza un análisis detallado de los casos de COVID-19 en la ciudad de Cali. Para ello, se define un dataframe denominado "casos\_positivos", el cual incluye variables clave como la fecha de inicio de síntomas "iniciodeSíntomas\_", el estado de recuperación definitivo "Recuperado definitivo", las coordenadas geográficas "Latitud – X" y "Longitud Y", así como los códigos y nombres de los barrios y comunas correspondientes "COD\_BARRIO", "COD\_COMUNA", "NOMBRE".

Posteriormente, el dataframe dividido en dos partes: "casos\_positivos\_fallecidos\_g" ver ilustración 6, que agrupa los casos fallecidos por barrio y presenta la distribución espacial de los casos, y "casos\_positivos\_g" ver ilustración 7, que contiene el total de casos por barrio. Con el objetivo de facilitar el análisis temporal de los datos, creamos una nueva columna denominada "año\_mes", que representa el año y mes de la fecha de inicio de síntomas, estos mapas presentan la distribución espacial de los casos para cada uno de los nuevos daframes:

Ilustración 6 casos\_positivos\_fallecidos\_g

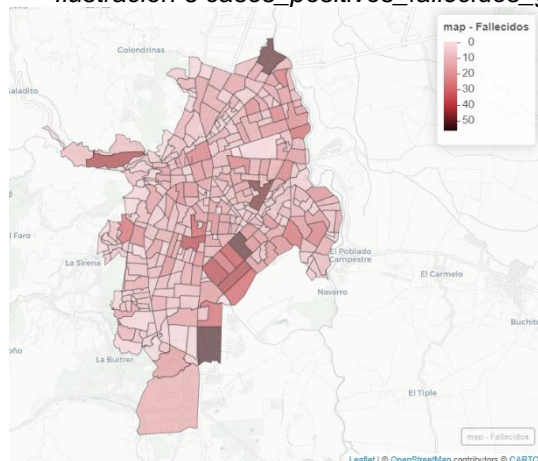
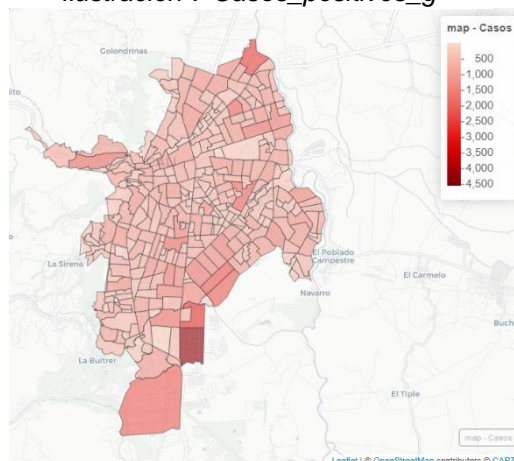


Ilustración 7 Casos\_positivos\_g



Fuente: Propia

Luego, se aplica una técnica de pivote (pívor) a ambas tablas. mediante el código de barrio ("COD\_BARRIO") como índice y los valores de la columna "año\_mes" como columnas, contabilizando el número de casos para cada combinación de barrio y mes. Estas tablas resultantes se denominaron "casos\_fallecidos\_pivot" y "casos\_positivos\_total\_pivot", respectivamente.

Finalmente, con la combinación de las tablas pivót resultantes en un único dataframe denominado "Covid\_Casos\_Fallecidos", podemos observar la información consolidada sobre los casos fallecidos por COVID-19, organizada por barrio y mes, listo para su análisis y visualización. Ver anexo 2.

Esta fase de procesamiento y análisis de datos nos permite obtener una comprensión

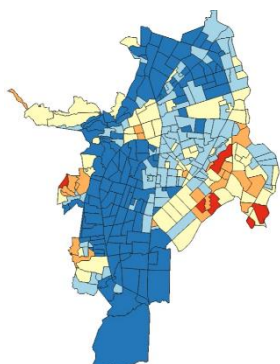
más profunda de la distribución espacial y temporal de los casos de COVID-19 en la ciudad de Cali, lo que a su vez facilita la identificación de patrones y la formulación del modelo.

## 7.2 PREPARACIÓN DANE

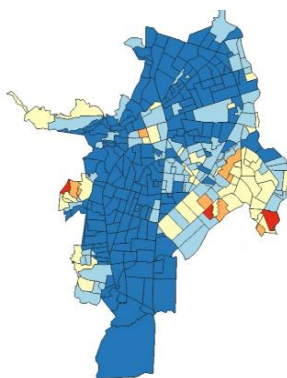
Una vez georreferenciamos la base de datos de Índice de Pobreza Multidimensional del DANE, observamos que esta proporcionaba datos a una escala más detallada que la de barrio, operando a nivel de manzana, una unidad espacial más pequeña. Ante esta situación, se optó por adaptar los datos a una escala homologada a la del análisis, lo que llevó a crear un nuevo Shape (archivo de SIG) denominado "DANE\_Barrio".

En este nuevo Shape, se realiza un procesamiento específico para las variables numéricas, como el "índice de pobreza multidimensional (IPM)", redondeando los valores a cero decimales. Además, con el fin de reducir la escala espacial a nivel de barrio, con la creación de tres nuevas variables: "mean\_imp", "median\_imp" y "mode\_imp", que representan respectivamente el promedio, la mediana y la moda de los valores del IPM en cada barrio. Ver ilustraciones 8, 9 y 10.

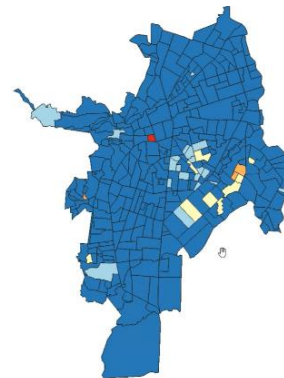
*Ilustración 8 IPM Media*



*Ilustración 9 IPM Mediana*



*Ilustración 10 IPM Moda*



*Fuente: Propia*

Por otro lado, para las variables categóricas como "LABEL", "embarazo\_a" y "reactivaci", se emplea el concepto de moda, es decir, seleccionando el valor más frecuente en cada barrio como representativo de la categoría en cuestión.

Una vez completado este proceso de ajuste de escala y transformación de variables, evaluando visualmente los resultados en un mapa de la ciudad de Santiago de Cali. Este paso fue crucial para asegurar que el preprocesamiento de los datos mantuviera la distribución espacial de la información original y que los datos adaptados reflejaran con precisión la realidad socioeconómica de los distintos barrios de la ciudad.

Tras el análisis y procesamiento de los datos georreferenciados del DANE a escala de

barrio, se ha llegado a la conclusión de que el IPM Media es la variable que mejor mantiene la distribución espacial de manera representativa en la ciudad de Santiago de Cali. Esta conclusión se basa en la observación de que el IPM Media ofrece una distribución equilibrada de los niveles de pobreza multidimensional en todos los barrios, preservando así la variabilidad y riqueza de la realidad socioeconómica de la ciudad. Por consiguiente, se ha determinado que el IPM Media será utilizado como variable clave en los modelos multivariados que se desarrollarán en las siguientes secciones de esta investigación.

## 8. MODELADO

Se llevó a cabo una cuidadosa deliberación sobre qué tipo de modelo sería más adecuado para calcular o aproximar la efectividad del programa de vacunación en la reducción de casos de fallecimiento por COVID-19. Con base en el estudio de tipo ecológico el modelo adoptado es el modelo de Poisson se determinó como el más apropiado para nuestros propósitos [17].

El modelo ecológico de Poisson se seleccionó debido a su capacidad para analizar la relación entre una variable dependiente, en este caso el número de casos de fallecimiento, y una variable independiente, que representa la implementación del plan de vacunación. Este modelo permite estimar el impacto del plan de vacunación sobre los casos de fallecimiento, teniendo en cuenta las fluctuaciones temporales y las diferencias entre los distintos barrios de la ciudad.

Una vez establecido el modelo a utilizar, se procedió a cargar los datos recopilados en un entorno de desarrollo de RStudio para llevar a cabo el análisis. El código comienza leyendo los datos desde un archivo Excel que contiene información detallada sobre los casos de fallecidos por COVID-19 en diversos barrios de la ciudad.

A continuación, se define una función llamada "calc\_impacto()" diseñada para calcular el impacto del plan de vacunación en los casos de fallecidos para un barrio específico. Dentro de esta función, se preparan los datos extrayendo las fechas y el número de casos de fallecidos para el barrio dado. Estos datos se organizan en un formato adecuado para su análisis.

Una vez organizados los datos, se crea un gráfico que muestra la evolución de los casos de fallecidos a lo largo del tiempo, diferenciando entre períodos en los que el plan de vacunación estaba activo y períodos en los que no lo estaba. Este paso nos permite visualizar la relación entre la implementación del plan de vacunación y la incidencia de casos de fallecimiento.

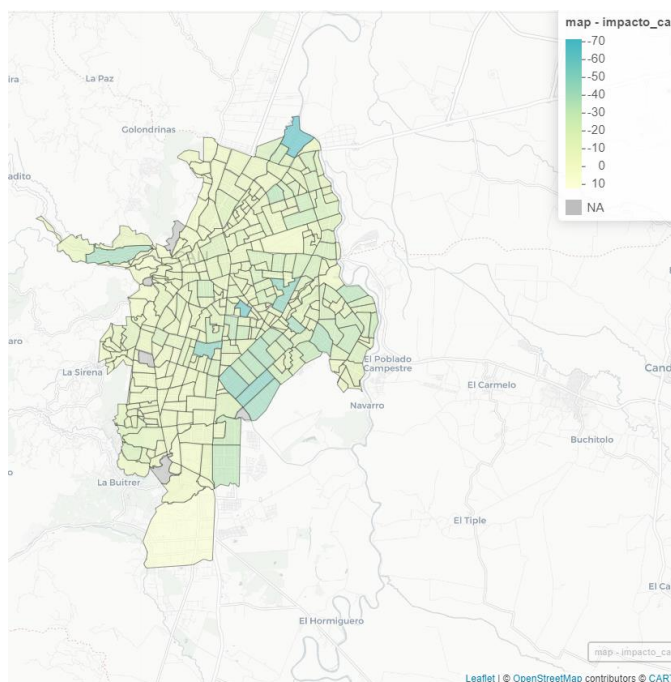
Posteriormente, se ajusta un modelo de regresión de Poisson a los datos para evaluar el impacto del plan de vacunación. Este modelo relaciona el número de casos de fallecidos con la implementación del plan de vacunación. Se extrae el coeficiente asociado al plan de vacunación del modelo ajustado, que representa el impacto del plan de vacunación en los casos de fallecimiento.

Tras realizar el análisis del impacto del plan de vacunación sobre los casos de fallecimiento por COVID-19 en los diferentes barrios de la ciudad, se almacena los resultados en un nuevo dataframe denominado "impacto\_casos". Este dataframe contiene dos variables principales: "barrio\_id", que identifica de manera única a cada barrio, y "impacto\_casos", que representa el impacto del plan de vacunación en la incidencia de casos de fallecimiento en cada barrio.

Es importante destacar que la variable "impacto\_casos" adopta valores tanto positivos como negativos. En este contexto, cuanto más positivo sea el valor de "impacto\_casos", menor será el impacto del plan de vacunación, ya que indica una reducción en el número de casos de fallecimiento. Por otro lado, cuanto más negativo sea el valor de "impacto\_casos", mayor será el impacto de la vacunación, lo que sugiere un aumento en la incidencia de casos de fallecimiento.

La ilustración 11 presenta el mapa de la distribución espacial de los barrios con mayor efectividad de la vacunación:

*Ilustración 11 Impacto de la efectividad de la vacunación por covid 19 por barrios de Cali.*



*Fuente: Propia*

Tras el análisis, se identifica los barrios que presentaron el mayor y menor impacto de la efectividad del plan de vacunación por covid 19. Los barrios con mayor impacto, es decir, aquellos en los que se observó una disminución significativa en la incidencia de casos de fallecimiento, se detallan en la tabla 4:

*Tabla 4 Barrios con mayor impacto de la efectividad de vacunación por covid 19*

Mayor impacto		
barrio_id	Nombre	Valor
610	Ciudadela Floralia	-70.86
1112	Prados de Oriente	-64.55
1596	Ciudad Cordoba	-54.73
1002	El Guabal	-52.32

*Fuente: Propia*

El impacto de la efectividad de la vacunación se presenta a través del modelo de Poisson en el cual se toma como referencia cuatro barrios diferentes donde se visualiza cambios significativos. En general, se observa una tendencia a la baja en el número de casos a lo largo del tiempo en todos los barrios. Sin embargo, la magnitud de esta disminución y la claridad del efecto de la vacunación varían entre los barrios. Algunos barrios muestran una disminución más pronunciada en los casos, con un efecto de vacunación más claro, mientras que otros tienen fluctuaciones más grandes y un efecto de vacunación menos distintivo. Ver ilustración 12.

Ilustración 12 Graficas del modelo barrios de mayor impacto



Fuente: Propia

Por otro lado, se identifica los barrios con menor impacto, donde el plan de vacunación tuvo un efecto menos pronunciado en casos de fallecimiento detallados en la tabla 5:

tabla 5 Barrios con el menor impacto

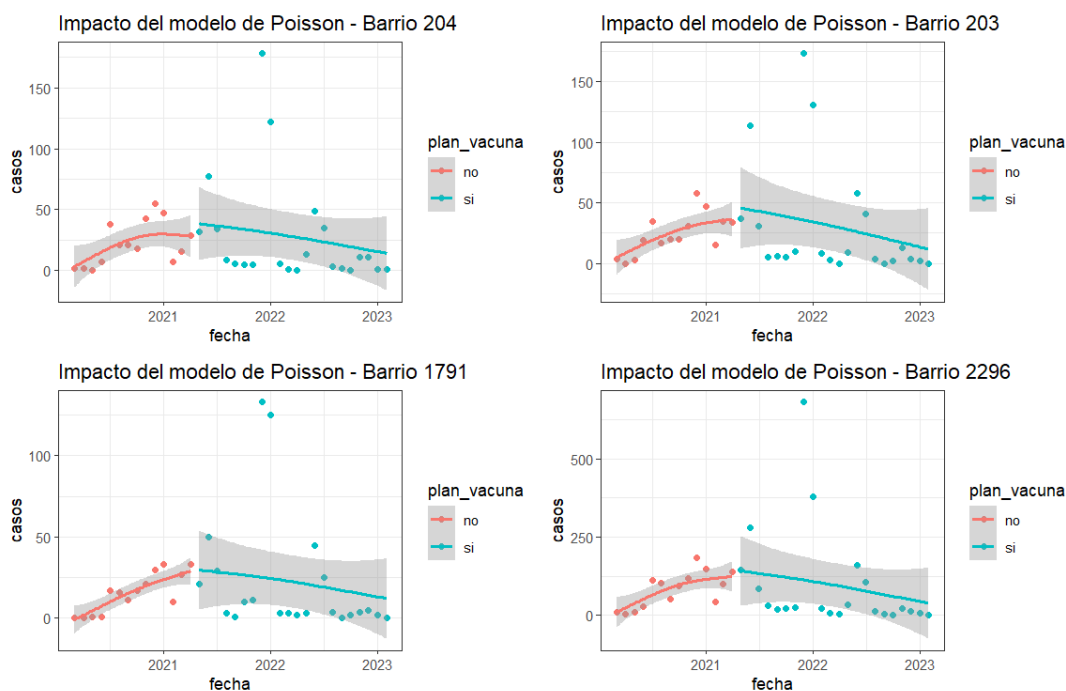
Mínimo impacto		
barrio_id	Nombre	Valor
204	Normandia	5.46
203	Arboledas	5.68
1791	Unicentro Cali	6.36
2296	Parcelaciones Pance	12.02

Fuente: Propia

Los gráficos generados muestran el impacto del modelo de Poisson en cuatro barrios diferentes. En conjunto, se observa una tendencia general al aumento en el número de

casos en todos los barrios analizados, aunque se presentan ciertas fluctuaciones a lo largo del tiempo. Ver ilustración 13.

Ilustración 13 Graficas del modelo para barrios de mínimo impacto



Fuente: Propia

Es importante destacar que el efecto de la vacunación varía entre los barrios estudiados. Sin embargo, se evidencia una superposición significativa en la distribución de casos entre individuos vacunados y no vacunados. Este hallazgo sugiere que, aunque la vacunación puede tener un impacto en la reducción de casos, este efecto puede ser limitado en estos contextos específicos.

## **9. ANALISIS DE RESULTADOS**

### **9.1 MODELADO DE ANÁLISIS ESPACIAL**

El impacto de la vacunación fue modelado y evaluado mediante el Índice de Moran global y local, así como el Índice de Geary local, con el objetivo de identificar los barrios con mayor influencia de la vacunación. Este análisis de impacto se realizó considerando las variables relacionadas con las tres dimensiones socioeconómico, poblacional y estrato socioeconómico.

En primer lugar, se utilizó el Índice de Moran global para examinar la autocorrelación espacial de los datos a nivel global, lo que proporcionó una visión general del patrón de distribución del impacto de la vacunación en el área de estudio.

Posteriormente, se aplicó el Índice de Moran local y el Índice de Geary local para identificar los barrios con agrupaciones espaciales significativas de alto impacto de la vacunación. Esta técnica permitió detectar áreas específicas donde la vacunación ha tenido un efecto notable en la población, lo que puede ser crucial reconocer las áreas con mayor impacto y orientar acciones en la planificación y la implementación de políticas de salud pública.

Una vez identificados estos barrios, se realiza un análisis detallado de la efectividad de la vacunación por covid 19 y el impacto de esta en relación con diferentes variables que abarcan las tres dimensiones pertinentes, sociales, poblacionales y de salud pública. Este análisis proporcionó una comprensión más profunda de cómo la efectividad de la vacunación y su impacto se relaciona con diferentes aspectos intrínsecos en el territorio como lo son socioeconómicos, demográficos y de salud pública en los diferentes barrios del Distrito de Santiago de Cali, lo que puede contribuir a mejorar las estrategias de intervención y promoción de la salud en la comunidad.

### **9.2 ÍNDICE GLOBAL DE MORAN**

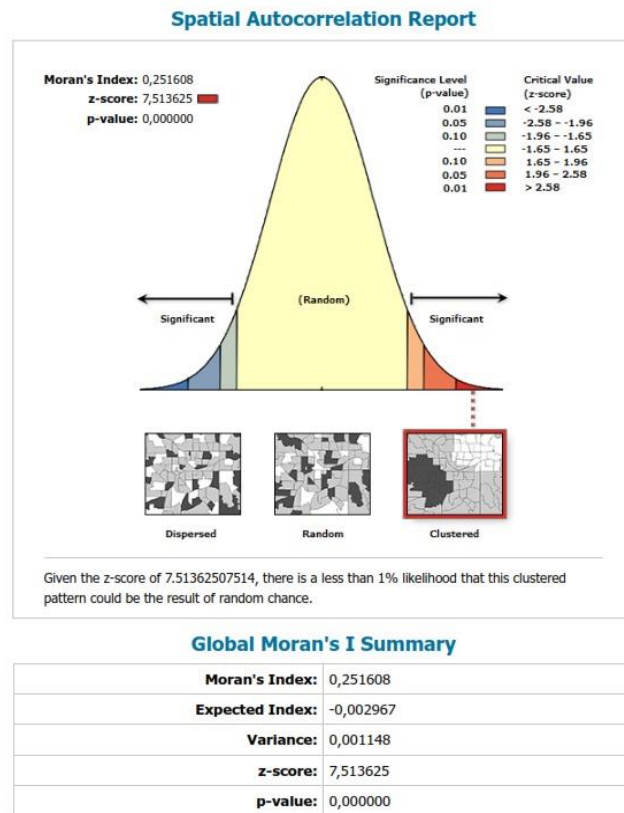
Se evaluó el impacto de la vacunación en el territorio de Santiago de Cali mediante el índice de autocorrelación espacial de Moran, con el propósito de identificar las áreas con mayor y menor impacto de la vacunación. Este análisis se realizó empleando tanto el índice global como el índice local de Moran.

Inicialmente, el índice de autocorrelación espacial de Moran proporcionó una visión general del patrón de distribución del impacto de la vacunación en toda la ciudad. Luego, se utilizó el índice de Moran local para identificar zonas específicas con agrupaciones espaciales significativas de alto y bajo impacto de la vacunación.

Este enfoque permitió detectar áreas con altos niveles de impacto de la vacunación, donde la cobertura vacunal ha sido especialmente efectiva, así como áreas con bajos niveles de impacto, que pueden requerir una mayor atención en términos de programas de vacunación y salud pública.

Se presenta en la ilustración 14, la interpretación del índice global y local de Moran a través de la campana de Gaus donde se evidencia que existe una significativa autocorrelación espacial en la cual se generan patrones espaciales de clusterización espacial:

Ilustración 14 Reporte de autocorrelación espacial a través de campana de Gauss



Fuente: Propia

El gráfico de campana de Gauss que se muestra en el informe de autocorrelación espacial es una representación visual de la distribución de los valores del índice de Moran's I.

Eje X (X-axis): Este eje representa los valores del índice de Moran's I. Este índice es una medida de autocorrelación espacial global que varía entre -1 y 1. Los valores cercanos a 1 indican una autocorrelación espacial positiva (clustering), valores cercanos a -1 indican una autocorrelación espacial negativa (dispersión), y valores cercanos a 0 indican aleatoriedad espacial.

Eje Y (Y-axis): Representa la densidad de frecuencia de los valores del índice de Moran's I.

Campana de Gauss (Gaussian Bell): La forma de campana de Gauss representa la distribución de los valores del índice de Moran's I. En este caso, la campana está sesgada hacia la derecha, lo que indica una mayor concentración de valores positivos de Moran's I.

Áreas sombreadas: Las áreas sombreadas del gráfico indican los niveles de significancia estadística. Las áreas sombreadas en rojo indican valores significativos a un nivel de

confianza del 99%, mientras que las áreas sombreadas en naranja indican valores significativos a un nivel de confianza del 95%.

En este caso específico, el índice de Moran's I tiene un valor de 0.251608, lo que indica una autocorrelación espacial positiva. Además, el valor p es muy pequeño ( $p\text{-value}=0.0000$ ), lo que sugiere que esta autocorrelación espacial es estadísticamente significativa. El gráfico de campana de Gauss confirma esta interpretación al mostrar una mayor concentración de valores positivos de Moran's I y una cola sesgada hacia la derecha.

### 9.3 SIMULACIÓN DE MONTECARLO

En la simulación de Monte Carlo realizada para evaluar la autocorrelación espacial utilizando el índice de Moran I, los siguientes resultados fueron obtenidos, Ver ilustración 15.

*Ilustración 15 Resultados simulación de Montecarlo*

```
Monte-Carlo simulation of Moran I
data: map$vb1e
weights: nbw
number of simulations + 1: 1000
statistic = 0.25371, observed rank = 1000, p-value = 0.001
alternative hypothesis: greater
```

*Fuente: Propia*

**Estadístico:** El valor del estadístico de autocorrelación espacial, en este caso, Moran I, es 0.25371.

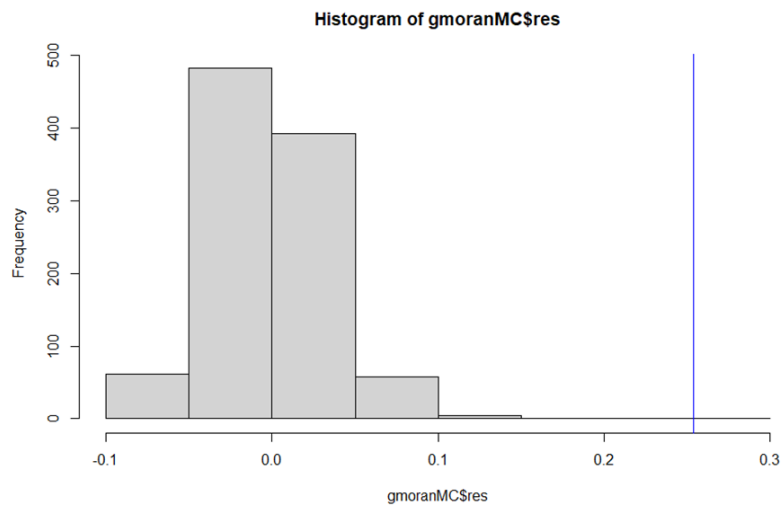
**Rango observado:** El rango observado del estadístico de Moran I es 1000. Esto indica la posición del estadístico observado en relación con los estadísticos generados en las simulaciones de Monte Carlo. Un rango más alto indica que el estadístico observado es relativamente alto en comparación con los valores generados en las simulaciones.

**Valor p:** El valor p asociado con el estadístico observado es 0.001. Este valor p representa la probabilidad de observar un estadístico de Moran I igual o mayor al valor observado, bajo la hipótesis nula de ausencia de autocorrelación espacial. Dado que este valor p es menor que un nivel de significancia común, como 0.05, se rechaza la hipótesis nula en favor de la hipótesis alternativa.

**Hipótesis alternativa:** La hipótesis alternativa indica que hay autocorrelación espacial positiva, y se está probando si el estadístico de Moran I observado es significativamente mayor que los valores generados por el azar en las simulaciones de Monte Carlo.

Los valores de la simulación de Monte Carlo se distribuyen alrededor de 0, ya que el centro del histograma se encuentra cerca de este valor. Ver ilustración 16.

Ilustración 16 Histograma de simulación de Montecarlo



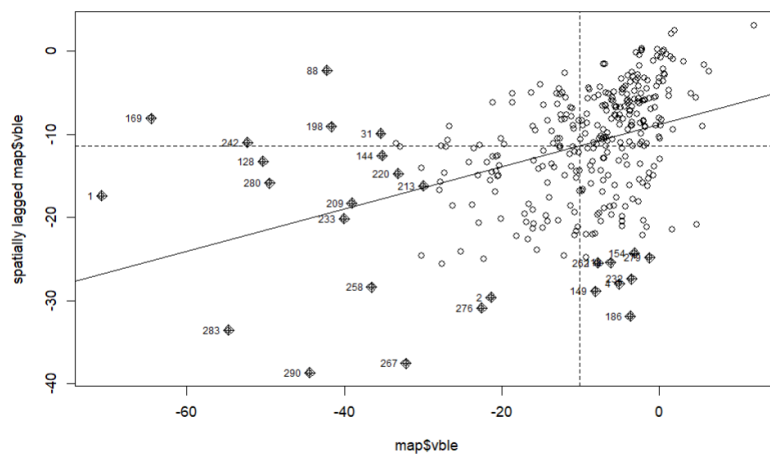
Fuente: Propia

La cola derecha del histograma indica que hay una proporción de simulaciones que resultaron en valores de Moran I mayores que 0.

La presencia de valores positivos indica que hay autocorrelación espacial positiva en los datos.

La barra azul vertical indica el valor del estadístico Moran I observado en los datos reales. Si este valor cae en la cola derecha del histograma, sugiere que el valor observado es significativamente mayor que los valores generados por las simulaciones de Monte Carlo, lo que apoya la conclusión de que hay una autocorrelación espacial significativa en los datos. Ver ilustración 17.

Ilustración 17 Diagrama de dispersión índice de Moran



Fuente: Propia

## 9.4 ÍNDICE LOCAL DE MORAN

Los Valores obtenidos se especifican en la ilustración 18.

Ilustración 18 Resultados del índice local de Moran

	$I_i$	$E.I_i$	$Var.I_i$	$Z.I_i$	$Pr(z > E(I_i))$
1	3.4953589	-0.0882994834	6.62097989	1.392725	0.0818514333
2	1.7403684	-0.0030431624	0.24952461	3.490143	0.0002413811
3	0.8481679	-0.0027101389	0.29729642	1.560531	0.0593171877
4	-0.7122424	-0.0006060027	0.03972675	-3.570398	0.9998217804
5	2.1311820	-0.0072755983	0.47377174	3.106820	0.0009455582
6	0.6626453	-0.0020929555	0.22973482	1.386875	0.0827399575

Fuente: Propia

Que representan:

**$I_i$ :** Este es el valor del índice de Moran local para cada observación. Representa la autocorrelación espacial local en torno a cada unidad espacial en tus datos. Valores positivos indican autocorrelación espacial positiva (las unidades similares están agrupadas juntas), mientras que valores negativos indican autocorrelación espacial negativa (las unidades similares están dispersas).

**$E.I_i$ :** Este es el valor esperado del índice de Moran local bajo la hipótesis nula de aleatoriedad espacial. Es la esperanza del índice de Moran local y representa el valor promedio que se esperaría si no hubiera autocorrelación espacial.

**$Var.I_i$ :** Esta es la varianza del índice de Moran local bajo la hipótesis nula. Representa la dispersión de los valores del índice de Moran local alrededor del valor esperado.

**$Z.I_i$ :** Este es el estadístico de Moran local estandarizado. Se calcula como el cociente entre el valor del índice de Moran local y la raíz cuadrada de la varianza del índice de Moran local. Indica cuántas desviaciones estándar está el valor del índice de Moran local del valor esperado bajo la hipótesis nula.

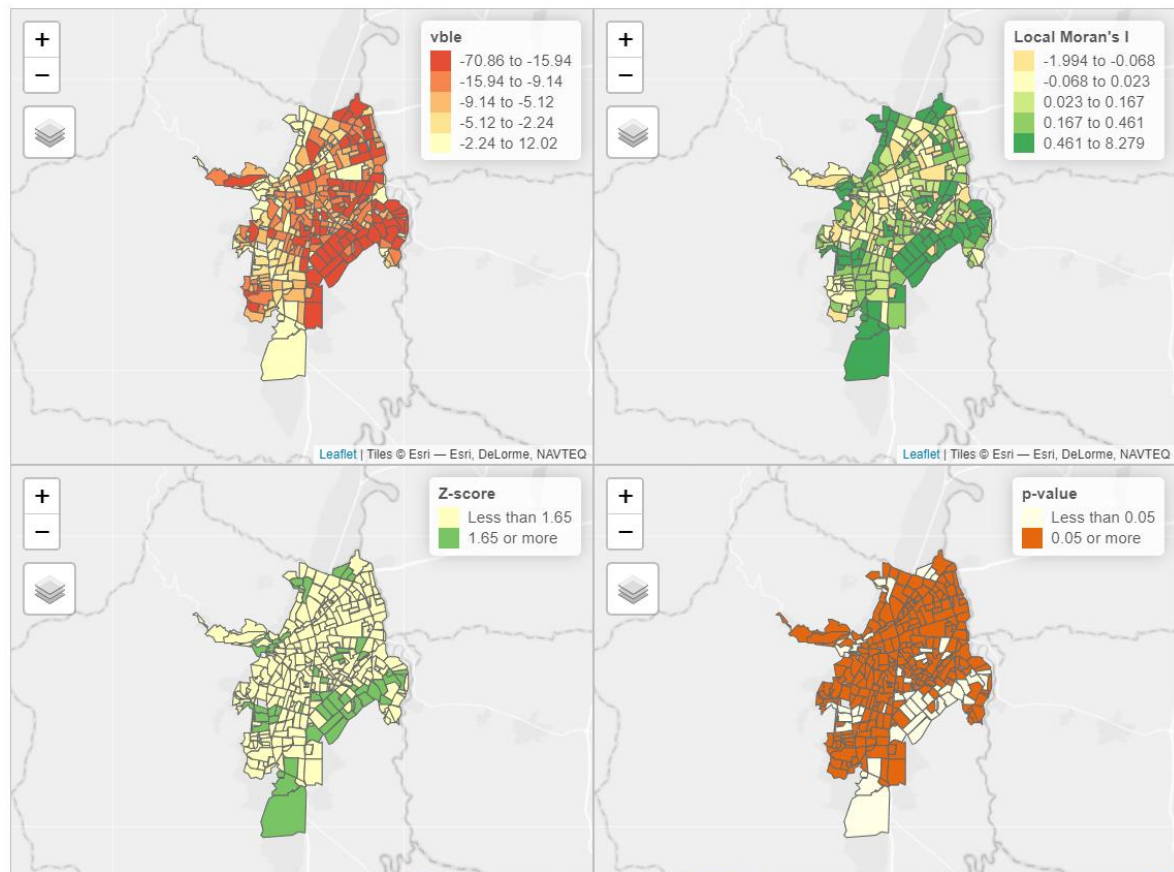
**$Pr(z > E(I_i))$ :** Esta es la probabilidad de que el valor del estadístico de Moran local estandarizado sea mayor que el valor esperado bajo la hipótesis nula. En otras palabras, es la probabilidad de observar un valor de Moran local tan extremo o más extremo que el observado, bajo la hipótesis nula de ausencia de autocorrelación espacial. Es una medida de la significancia estadística del índice de Moran local.

usamos alternativa = " two.sided" en lugar de alternativa = " greater", estaríamos probando

$H_0$ : sin autocorrelación espacial vs.

$H_1$ : autocorrelación espacial positiva o negativa. En esta prueba bilateral, los valores de puntuación z inferiores a  $-1.65$  indican autocorrelación espacial negativa, y los valores de puntuación z superiores a  $1.65$  indican autocorrelación espacial positiva. El siguiente mapa muestra las áreas con autocorrelación espacial negativa, nula y positiva, obtenidas al romper la leyenda de acuerdo con estos valores de puntuación z. Ver ilustración 19.

ilustración 19 Distribución de los resultados del índice local de Moran



Fuente: Propia

**Mapa de vble (Impacto):** Este mapa muestra la distribución espacial de la variable "vble". Los colores representan los diferentes valores de la variable, con una leyenda que indica los rangos de valores y sus correspondientes colores. Las áreas en rojo oscuro tienen valores más altos, mientras que las áreas en amarillo claro tienen valores más bajos.

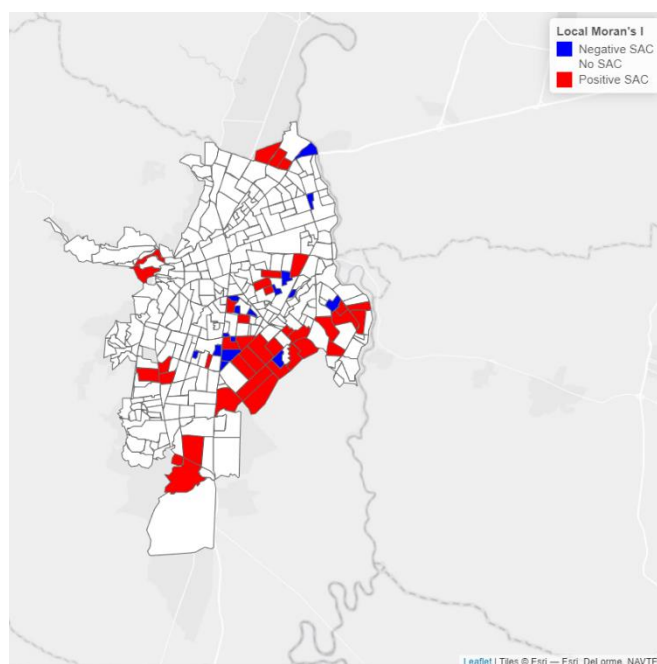
**Mapa del Índice de Moran Local:** Este mapa muestra el Índice de Moran Local para la variable "vble" en diferentes unidades espaciales. El Índice de Moran Local indica la autocorrelación espacial en torno a cada unidad espacial. Los colores representan los diferentes rangos de valores del Índice de Moran Local, con áreas verdes que indican autocorrelación espacial positiva y áreas amarillas que indican autocorrelación espacial negativa. Los valores más altos (más oscuros) indican una autocorrelación espacial más fuerte.

**Mapa de Z-score:** Este mapa muestra el Z-score asociado al Índice de Moran Local para la variable "vble". El Z-score indica cuántas desviaciones estándar está el valor del Índice de Moran Local del valor esperado bajo la hipótesis nula de ausencia de autocorrelación espacial. Los colores representan los diferentes rangos de valores del Z-score, con áreas en verde que indican valores por debajo de 1.65 (menos significativos) y áreas en verde oscuro que indican valores iguales o superiores a 1.65 (más significativos).

**Mapa de Valor p:** Este mapa muestra el valor p asociado al Índice de Moran Local para la variable "vble". El valor p indica la significancia estadística del Índice de Moran Local, es decir, la probabilidad de observar un valor de Moran I local tan extremo o más extremo que el valor observado bajo la hipótesis nula de ausencia de autocorrelación espacial. Los colores representan los diferentes rangos de valores del valor p, con áreas en naranja que indican valores p menores que 0.05 (significativos) y áreas en amarillo que indican valores p iguales o superiores a 0.05 (menos significativos).

## 9.5 AUTOCORRELACIÓN ESPACIAL (SAC)

*Ilustración 20 Distribución espacial del índice local de Moran*



*Fuente: Propia*

En la ilustración 20 se encuentra un mapa que permite identificar la autocorrelación espacial local (Local Moran's I) Spatial AutoCorrelation (SAC) de impacto de la vacunación en diferentes unidades espaciales. Aquí está la interpretación:

Las áreas coloreadas en rojo indican una autocorrelación espacial positiva significativa (SAC), lo que significa que los valores altos están rodeados principalmente por otros valores altos y los valores bajos están rodeados principalmente por otros valores bajos.

Las áreas coloreadas en azul indican una autocorrelación espacial negativa significativa (SAC), lo que significa que los valores altos están rodeados principalmente por valores bajos y viceversa.

Las áreas blancas indican que no hay una autocorrelación espacial significativa (No SAC), lo que significa que los valores altos y bajos no están significativamente agrupados espacialmente.

Este mapa proporciona una visualización de la distribución espacial de la autocorrelación local de la variable en estudio (Impacto de Vacunación), lo que puede ser útil para identificar patrones espaciales y áreas de interés para un análisis más detallado.

## I Moran local

Permite identificar clusters de los siguientes tipos:

Alto-Alto: áreas de altos valores con vecinos de altos valores,

Alto-Bajo: áreas de valores altos con vecinos de valores bajos,

Bajo-Alto: áreas de valores bajos con vecinos de valores altos,

Bajo-Bajo: áreas de valores bajos con vecinos de valores bajos.

Para detectar clusters, usando la función `localmoran()` para calcular el valor de Moran local.

### Prueba a dos colas (two.sided)

El parámetro `alternative = "two.sided"` en la función `localmoran` especifica la hipótesis alternativa que se está probando en el cálculo del índice de Moran local. En este caso, "two.sided" significa que se está considerando una prueba de dos colas.

Cuando se utiliza "two.sided", la hipótesis alternativa es que hay autocorrelación espacial, pero no especifica si es positiva o negativa. Esto significa que la prueba busca detectar cualquier tipo de autocorrelación espacial, ya sea positiva o negativa.

En resumen, al especificar `alternative = "two.sided"`, se está evaluando si hay alguna forma de autocorrelación espacial, independientemente de si es positiva o negativa. Ver ilustración 21.

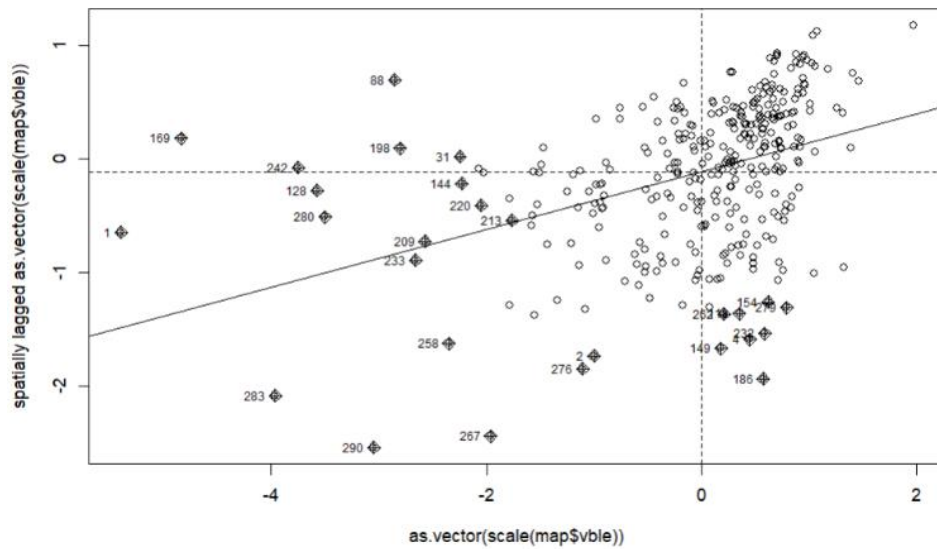
*Ilustración 21 Resultados de la prueba de dos colas*

	$I_i$	$E(I_i)$	$Var(I_i)$	$Z(I_i)$	$Pr(z \neq E(I_i))$
1	3.4953589	-0.0882994834	6.62097989	1.392725	0.1637028666
2	1.7403684	-0.0030431624	0.24952461	3.490143	0.0004827621
3	0.8481679	-0.0027101389	0.29729642	1.560531	0.1186343755
4	-0.7122424	-0.0006060027	0.03972675	-3.570398	0.0003564391
5	2.1311820	-0.0072755983	0.47377174	3.106820	0.0018911165
6	0.6626453	-0.0020929555	0.22973482	1.386875	0.1654799150

*Fuente: Propia*

Luego, se identifica los clusters de cada tipo utilizando la información proporcionada por el método de Moran I diagrama de dispersión obtenido con la función `moran.plot()`. Ver ilustración 22.

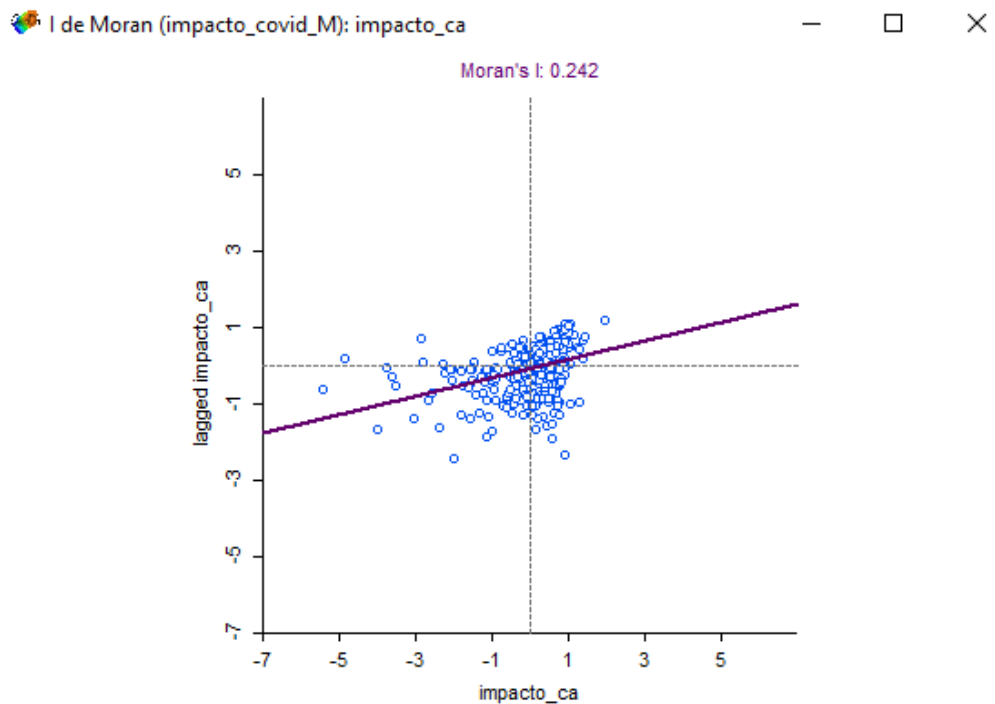
Ilustración 22 Gráfico de dispersión prueba de dos colas en RStudio



Fuente: Propia

Como se observa en la ilustración 23, el valor de Moran's I es 0.242, lo que indica una autocorrelación espacial positiva, es decir, áreas con valores altos de impacto\_ca tienden a estar rodeadas por áreas con valores altos, y áreas con valores bajos tienden a estar rodeadas por áreas con valores bajos.

Ilustración 23 Gráfico de dispersión prueba de dos colas en Geoda



Fuente: Propia

En la ilustración 24 podemos observar el análisis de autocorrelación espacial donde se obtiene:

$\text{map}[(\text{mp}\$x \geq 0 \ \& \ \text{mp}\$wx \geq 0) \ \& \ (\text{map}\$Imp \leq 0.05), \text{"quadrant"}] \leftarrow 1$ : Las observaciones que caen en el cuadrante de "alta-alta" (alto valor de la variable y alto valor del vecino) y que tienen un p-value (Imp) menor o igual a 0.05 se asignan al cuadrante 1.

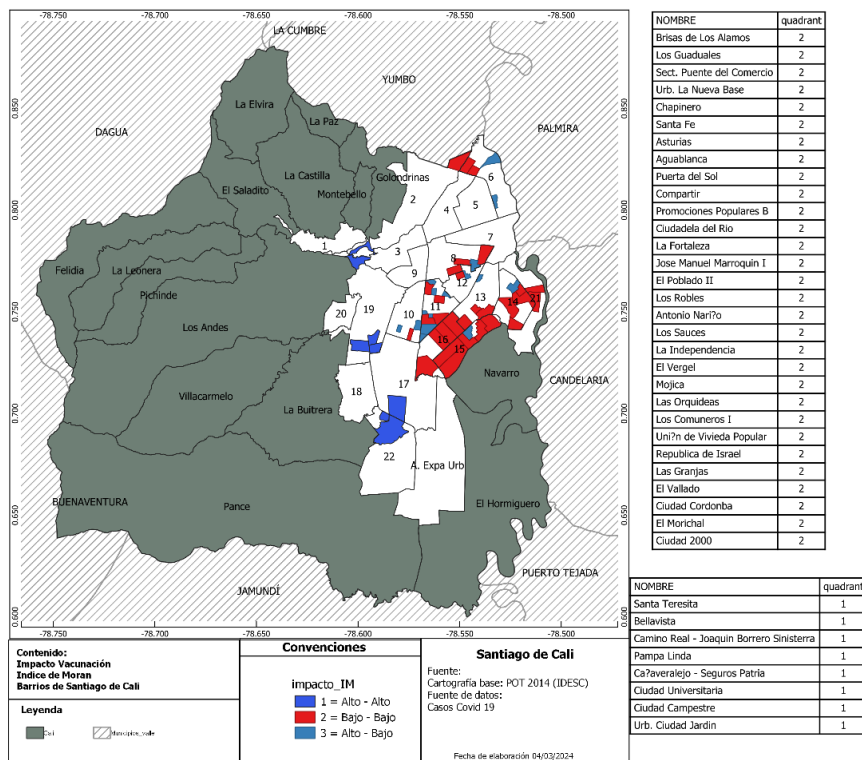
$\text{map}[(\text{mp}\$x \leq 0 \ \& \ \text{mp}\$wx \leq 0) \ \& \ (\text{map}\$Imp \leq 0.05), \text{"quadrant"}] \leftarrow 2$ : Las observaciones que caen en el cuadrante de "baja-baja" (bajo valor de la variable y bajo valor del vecino) y que tienen un p-value (Imp) menor o igual a 0.05 se asignan al cuadrante 2.

$\text{map}[(\text{mp}\$x \geq 0 \ \& \ \text{mp}\$wx \leq 0) \ \& \ (\text{map}\$Imp \leq 0.05), \text{"quadrant"}] \leftarrow 3$ : Las observaciones que caen en el cuadrante de "alta-baja" (alto valor de la variable y bajo valor del vecino) y que tienen un p-value (Imp) menor o igual a 0.05 se asignan al cuadrante 3.

$\text{map}[(\text{mp}\$x \leq 0 \ \& \ \text{mp}\$wx \geq 0) \ \& \ (\text{map}\$Imp \leq 0.05), \text{"quadrant"}] \leftarrow 4$ : Las observaciones que caen en el cuadrante de "baja-alta" (bajo valor de la variable y alto valor del vecino) y que tienen un p-value (Imp) menor o igual a 0.05 se asignan al cuadrante 4.

$\text{map}[(\text{map}\$Imp > 0.05), \text{"quadrant"}] \leftarrow 5$ : Las observaciones que tienen un p-value (Imp) mayor que 0.05 (no son significativas) se asignan al cuadrante 5.

*Ilustración 24 Análisis de autocorrelación espacial Local*



Fuente: Propia

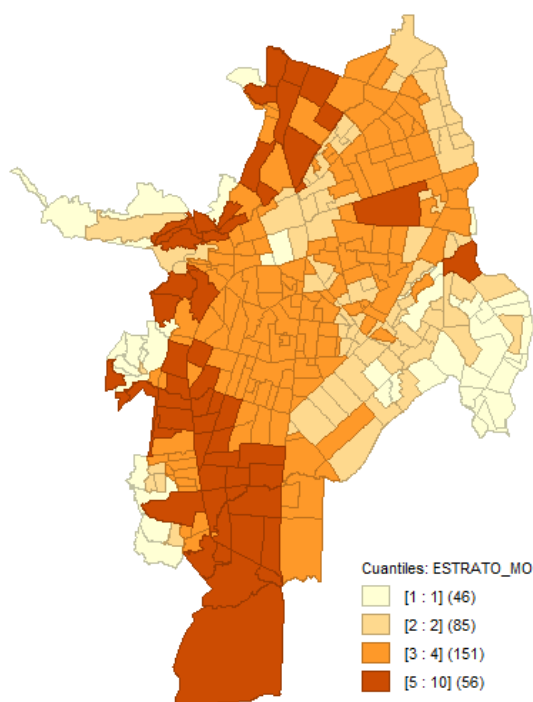
La ilustración 24 presenta los barrios con un mayor impacto de la vacunación en la ciudad de Santiago de Cali representado por cuadrantes donde 1 de color azul son los barrios con menor impacto de la vacunación entendiéndose que el impacto de la vacunación se presentan con valores negativos donde los datos con menor valor presentan un mayor impacto en la vacunación ver ilustración 11, estos se visualizan de color rojo con un indicador de impacto de 2, estos barrios se encuentran ubicados en las comunas 6, 13, 14, 15 y 16 zona nor oriente y oriente de Cali

Nota aclaratoria: para evitar una ambigüedad en la visualización de los datos presentados, se realiza para los análisis posteriores una multiplicación x -1 el valor del impacto de vacunación por covid 19 para tener datos positivos donde el impacto ya no sería tomado con valores negativos si no como valores positivos.

## 9.6 MODELO DE BILISA: IMPACTO VACUNACIÓN VS ESTRATO SOCIOECONÓMICO.

Inicialmente se plantea tener una visión clara de la relación entre el impacto de la vacunación y el estrato socioeconómico, Se presenta la distribución espacial del estrato socioeconómico donde los valores más altos se encuentran en tonos más oscuros. Ver ilustración 25.

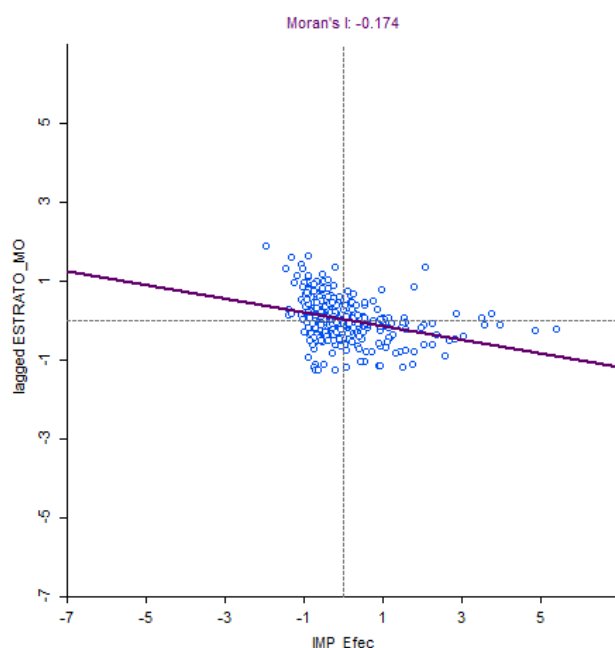
*Ilustración 25 Distribución espacial del estrato socioeconómico*



*Fuente: Propia*

La ilustración 26, muestra un gráfico de dispersión donde se evidencia la relación entre dos variables clave: el impacto en la vacunación (MP Efec) y el estrato socioeconómico (estrato). La línea de regresión en el diagrama de dispersión indica la dirección y fuerza de la relación entre estas variables. Se observa una tendencia negativa, lo que sugiere que a medida que aumenta el estrato socioeconómico, el impacto en la vacunación tiende a disminuir.

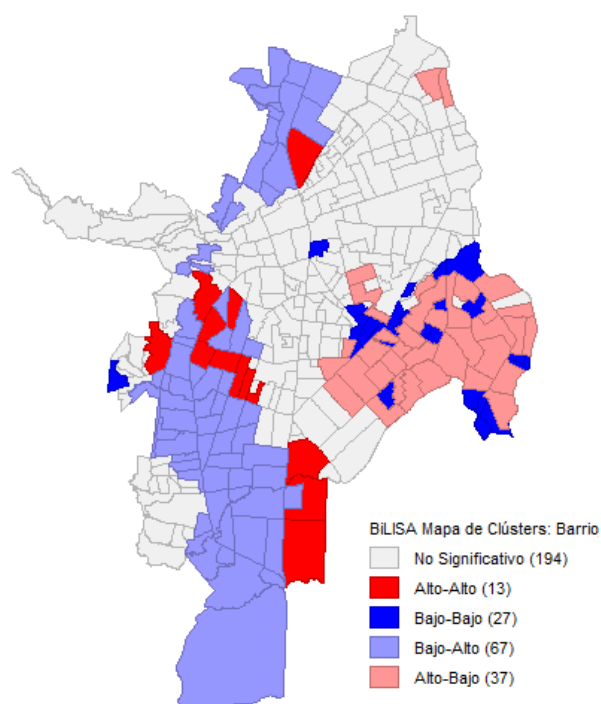
Ilustración 26 Grafico de dispersión de Moran estrato socioeconómico



Fuente: Propia

Por otro lado, en la ilustración 27, se encuentra el mapa de clusters el cual representa la distribución espacial de la agrupación identificada durante el análisis. Estos clusters se dividen en cinco categorías: alto-alto, bajo-bajo, alto-bajo, bajo-alto y no significativo. Los colores rojos en el mapa indican la presencia de clusters donde tanto el impacto en la vacunación como el estrato socioeconómico son altos o bajos simultáneamente. Por el contrario, los colores azules representan clusters donde una variable es alta mientras que la otra es baja y el cluster con color terracota presenta que mientras la variable de impacto de vacunación es alta, el estrato socioeconómico es bajo.

Ilustración 27 Cluster BILISA impacto / estrato socioeconómico

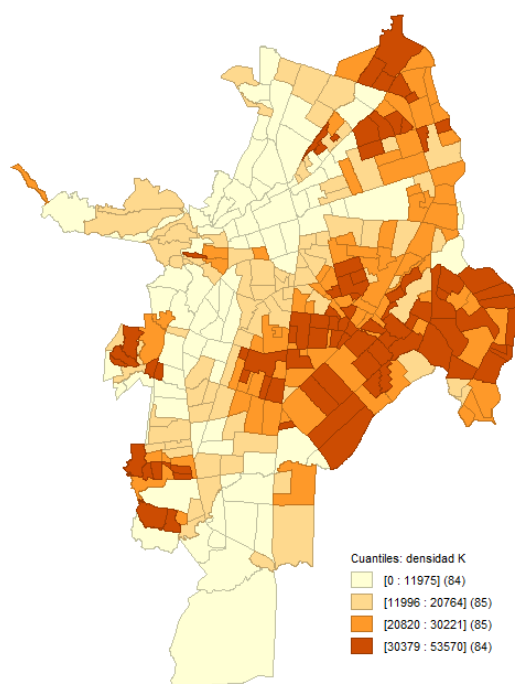


Fuente: Propia

### 9.7 MODELO DE BILISA: IMPACTO VACUNACIÓN VSDENSIDAD POBLACIONAL

Para tener una visión clara de la relación entre el impacto de la vacunación y la densidad poblacional, Se realiza la distribución espacial de la densidad poblacional donde los valores más altos se presentan en tonos más oscuros, Ver ilustración 28.

Ilustración 28 Distribución espacial densidad poblacional

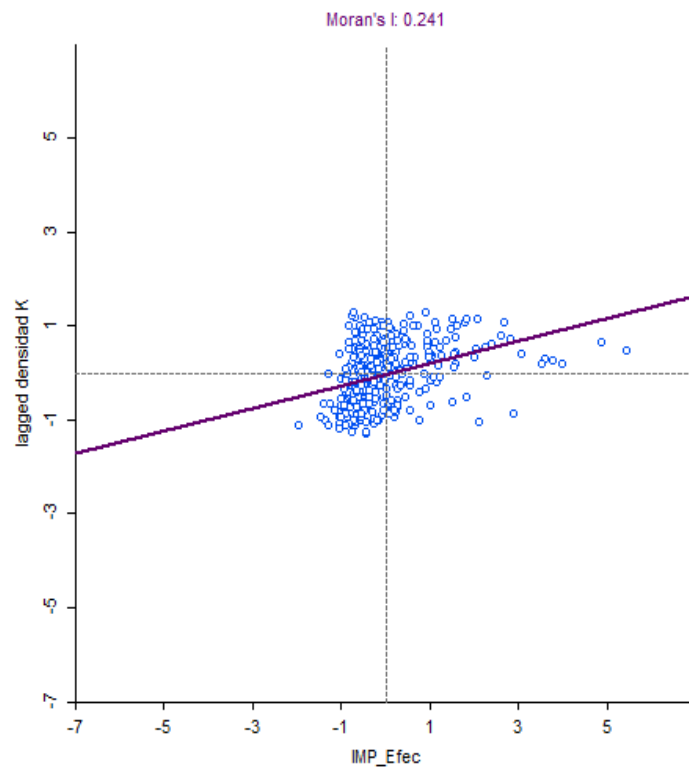


Fuente: Propia

Mediante el análisis bivariado LISA, se examinó la relación entre dos variables clave: el impacto de la vacunación (MP Efec) y la densidad poblacional. La gráfica de dispersión en la ilustración 29, presenta esta relación. La línea de regresión en el gráfico proporciona información sobre la dirección y la fuerza de la relación entre estas variables. Se observa una tendencia positiva, lo que sugiere que a medida que aumenta la densidad poblacional aumenta el impacto de la vacunación

En un segundo análisis bivariado LISA, se exploró la relación entre el impacto de la vacunación y la densidad poblacional. Se realizaron 99999 permutaciones para determinar la significancia estadística de los clusters identificados, con un grado de significancia de 0.01, lo que implica un nivel de confianza del 99%. La gráfica de dispersión muestra la relación entre estas dos variables. En este caso, se observa una tendencia positiva, lo que sugiere que a medida que aumenta la densidad poblacional, también aumenta el impacto de la vacunación.

Ilustración 29 Grafico de dispersión de Moran densidad poblacional

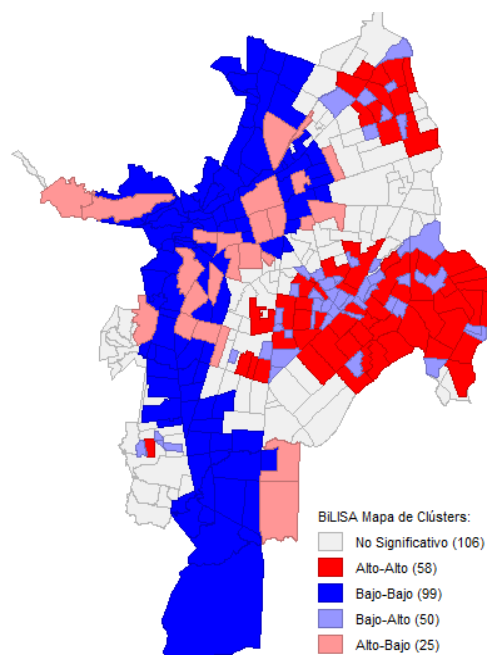


Fuente: Propia

Por otro lado, en la ilustración 30 se presenta el mapa de clusters que muestra la distribución espacial de los clusters identificados durante el análisis. Estos clusters se han clasificado en cinco categorías: alto-alto, bajo-bajo, alto-bajo, bajo-alto y no significativo.

El color rojo representa los barrios con valores altos en densidad poblacional y con valores altos en impacto de la vacunación, evidenciándose dos cluster uno en la zona nororiente de Cali y otro en el oriente, mientras que los valores presentados de color azul presenta valores bajos en densidad poblacional y valores bajos en impacto de la vacunación, esto evidencia que los barrios con una mayor densidad poblacional oriente y nororiente de Cali presentaron un mayor impacto en la vacunación, pero también es se resalta que los barrios visualizados con color terracota presentaron un bajo valor con densidad poblacional pero con un alto impacto en la vacunación, apreciando que el impacto de la vacunación tuvo un comportamiento clusterizado pero con un amplio alcance en la ciudad.

Ilustración 30 Cluster BILISA impacto / densidad poblacional

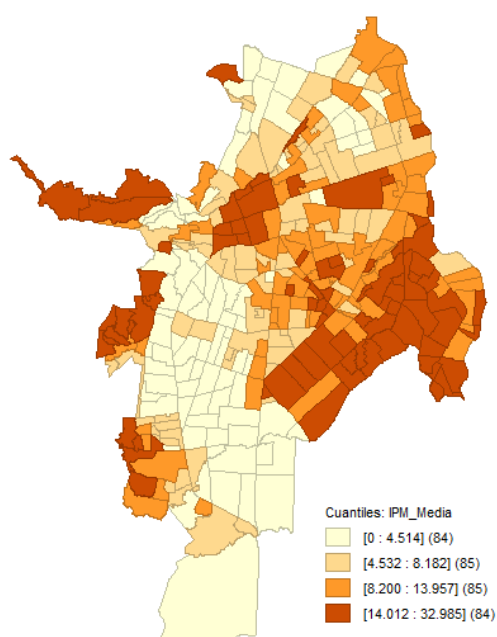


Fuente: Propia

## 9.8 MODELO DE BILISA: IMPACTO VACUNACIÓN VS ÍNDICE DE POBREZA MULTIDIMENSIONAL (IPM).

En la ilustración 31 se observa la distribución espacial del índice de pobreza multidimensional representada en cuantiles donde los valores más altos se evidencian de un tono más oscuro.

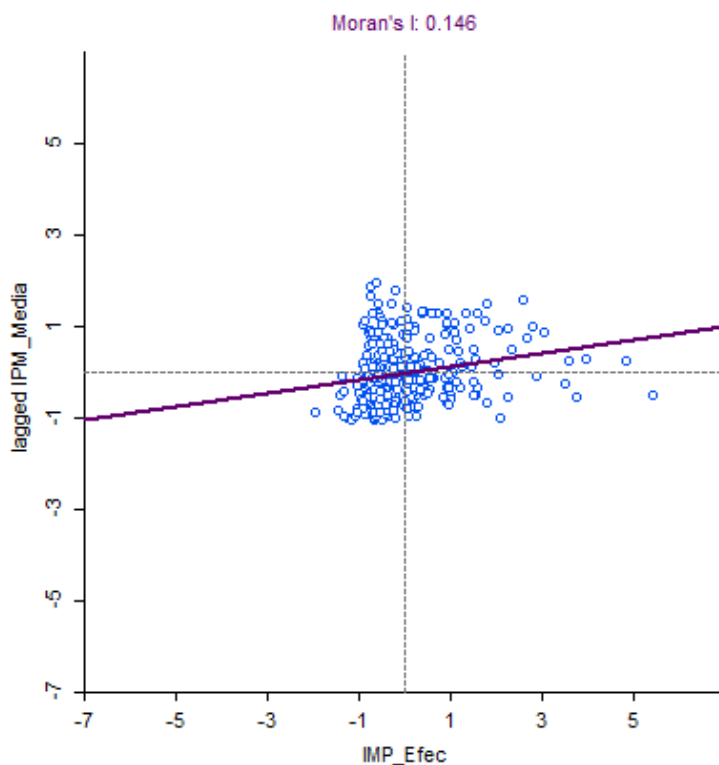
Ilustración 31 Distribución espacial IPM



Fuente: Propia

Como se observa en la ilustración 32 y con un índice de moran de 0,146 presentando una tendencia positiva, sugiere que a medida que aumenta el Índice de Pobreza Multidimensional, el impacto de la vacunación tiende a aumentar.

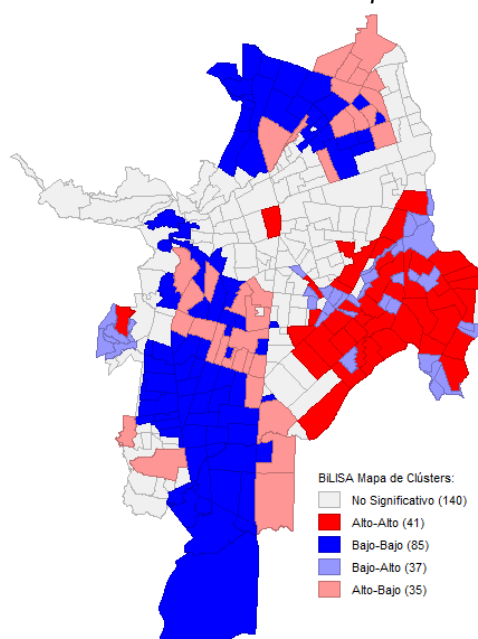
*Ilustración 32 Grafico de dispersión de Moran de IPM*



*Fuente: Propia*

El mapa de clusters representado en la ilustración 33 presenta la distribución espacial de los clusters identificados durante el análisis. Estos clusters se clasifican en cinco categorías: alto-alto, bajo-bajo, alto-bajo, bajo-alto y no significativo. Los colores rojos en el mapa indican clusters donde tanto el impacto de la vacunación como el Índice de Pobreza Multidimensional son altos o bajos simultáneamente, mientras que los colores azules representan clusters donde una variable es alta mientras que la otra es baja.

Ilustración 33 Cluster BILISA impacto / IPM

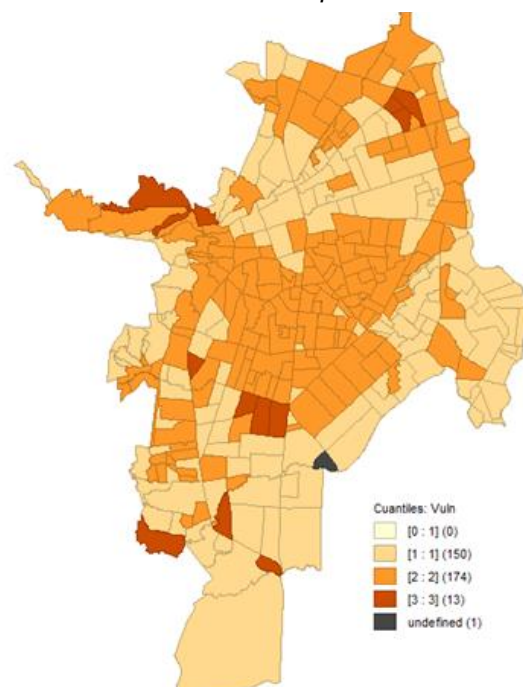


Fuente: Propia

## 9.9 MODELO DE BILISA: IMPACTO VACUNACIÓN VS VULNERABILIDAD

En la ilustración 34 se observa la distribución espacial del índice de la vulnerabilidad representada en cuartiles.

Ilustración 34 Distribución espacial vulnerabilidad

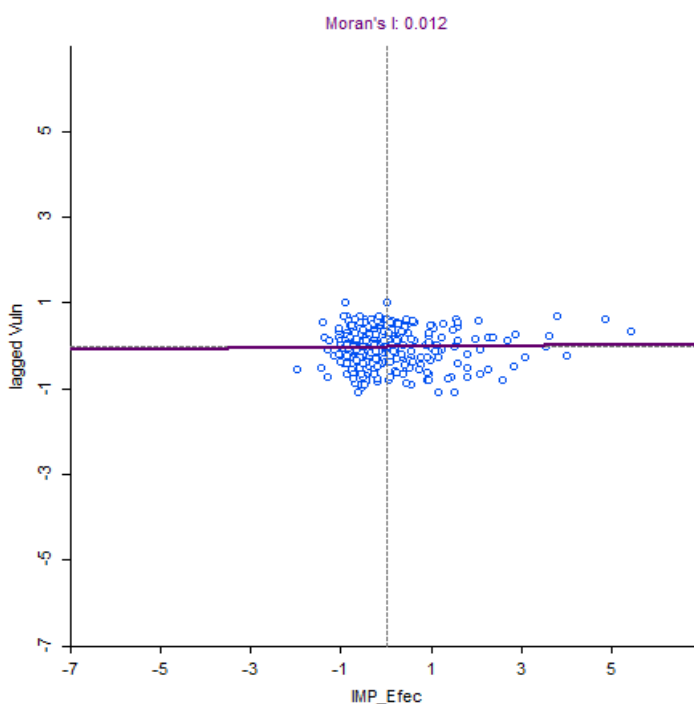


Fuente: Propia

En esta instancia, como se muestra en la ilustración 35, la línea de regresión exhibe una pendiente casi horizontal, lo que sugiere que no existe una relación evidente entre el impacto de la vacunación y la vulnerabilidad en las áreas geográficas analizadas. Esto implica que el impacto de la vacunación no parece estar directamente vinculado con el nivel de vulnerabilidad en dichos lugares.

Además, el valor de Moran's I, proporciona información sobre la autocorrelación espacial de los datos. Un valor de Moran's I cercano a 0 sugiere una distribución espacial aleatoria de los datos, mientras que valores positivos indican autocorrelación espacial positiva (lugares similares tienden a agruparse) y valores negativos indican autocorrelación espacial negativa (lugares similares tienden a estar dispersos). En este caso, el valor de Moran's I es cercano a 0 (0.012), lo que corrobora la observación de que no existe una estructura espacial clara en la relación entre el impacto de la vacunación y la vulnerabilidad.

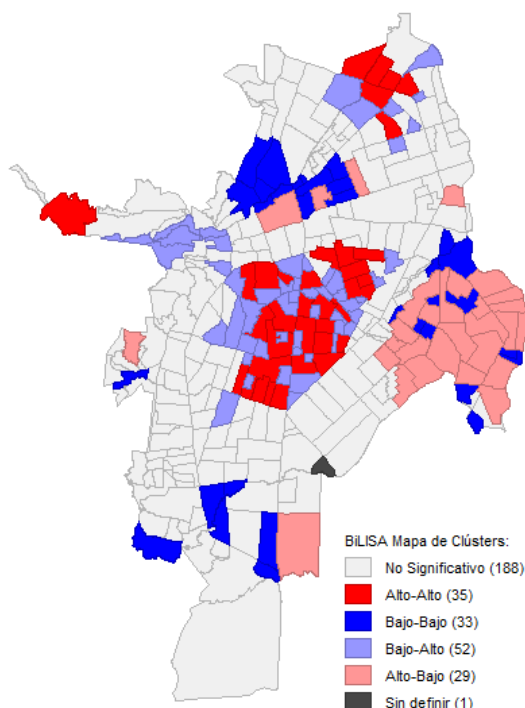
*Ilustración 35 Grafico de dispersión de Moran vulnerabilidad*



*Fuente: Propia*

El mapa de clusters BILISA representado en la ilustración 36, presenta la distribución espacial de los clusters identificados durante el análisis. Estos clusters se clasifican en cinco categorías: alto-alto, bajo-bajo, alto-bajo, bajo-alto y no significativo. Los colores representan una distribución dispersa de los valores en el mapa, donde no hay una clusterización entre valores alto-alto (rojos) y bajo-bajo (Azules). Sin embargo se presenta un impacto alto y vulnerabilidad baja en la zona oriente de Santiago de Cali.

Ilustración 36 Cluster BILISA impacto / vulnerabilidad



Fuente: Propia

## 9.10 GEARY LOCAL

El Mapa de Clúster Geary Local se ha generado utilizando múltiples variables para analizar la autocorrelación espacial local en un área geográfica específica. Las variables consideradas son las siguientes:

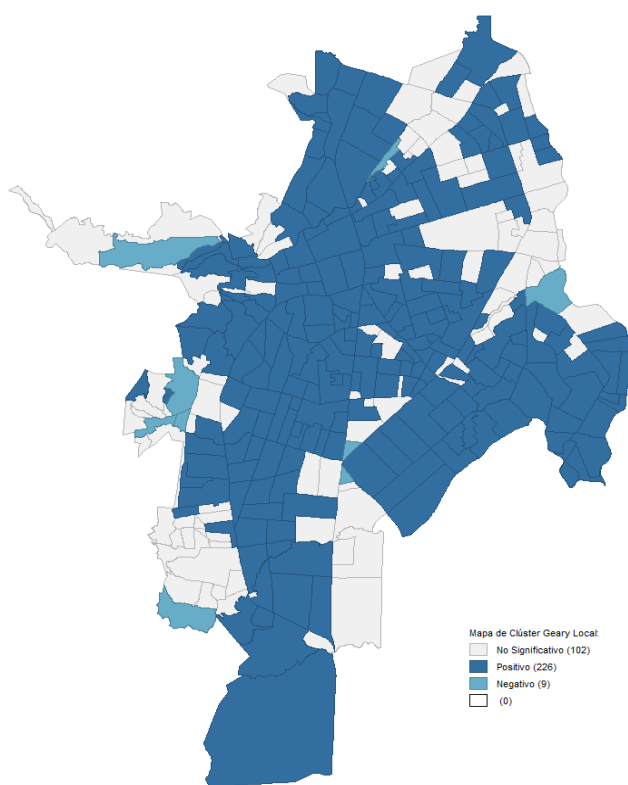
1. **Impacto de la vacunación:** Esta variable principal indica el efecto o la influencia de la vacunación en la población del área geográfica analizada.
2. **Densidad poblacional:** Representa la cantidad de personas por unidad de área en el mismo lugar geográfico.
3. **Estrato socioeconómico:** Indica el nivel socioeconómico de la población en el área, considerando factores como ingresos, educación y ocupación.
4. **Vulnerabilidad:** Esta variable refleja la susceptibilidad o fragilidad de la población ante ciertos riesgos o eventos, como desastres naturales o crisis de salud.

- 5. Índice de pobreza multidimensional:** Es una medida compuesta que evalúa diversos aspectos de la pobreza, como acceso a servicios básicos, vivienda adecuada, salud y educación.

La ilustración 37 presenta un mapa que se ha generado utilizando un enfoque de clúster de Geary local, con 99999 permutaciones y un grado de significancia de 0.01. Esto implica que se realizaron numerosas permutaciones para evaluar la significancia estadística de los resultados, y se estableció un nivel de confianza del 99% (grado de significancia de 0.01).

La interpretación del mapa implica analizar la autocorrelación espacial local de cada una de estas variables en el área geográfica representada. Los diferentes colores en el mapa indican diferentes tipos de autocorrelación espacial local para cada variable, lo que puede proporcionar información sobre cómo estas variables se relacionan en el área y cómo están distribuidas espacialmente.

*Ilustración 37 Cluster Geary Local*



*Fuente: Propia*

Se ha desarrollado un nuevo conjunto de datos denominado "multivariado", que incorpora tanto variables de impacto como socioeconómicas a nivel de barrio. Estas variables abarcan tres dimensiones principales: socioeconómica, poblacional y de salud.

En este contexto, se llevó a cabo un análisis bivariado de LISA para cada dimensión, con el propósito de explorar su relación con el impacto de la vacunación.

Para visualizar la distribución espacial, se creó un shapefile utilizando el software QGIS, el cual posteriormente se empleó como entrada en el software Geoda. Además, se generó una matriz de pesos basada en la contigüidad reina para los barrios, estableciendo un nivel de contigüidad en 1. A continuación, se procedió con el análisis bivariado de LISA, donde se contrastó la variable del impacto de la vacunación con las demás variables consideradas.

En un análisis subsiguiente, se incluyeron las variables de Estrato Socioeconómico, la Densidad Poblacional, el Índice de Pobreza Multidimensional (IPM) y la Vulnerabilidad. Se realizó el análisis respectivo, generando mapas de clusters y diagramas de dispersión de Moran para explorar las posibles relaciones entre estas variables y el impacto de la vacunación.

## 10. EVALUACIÓN

Para abordar el objetivo de evaluar el rendimiento de los diferentes modelos de análisis estadístico en términos de su ajuste a los datos, utilizando métricas de validación específicas, se llevó a cabo una serie de acciones metodológicas. Inicialmente, se procedió a realizar un análisis de regresión espacial con el propósito de investigar el impacto de la efectividad de la vacunación.

Este análisis comenzó con la implementación de un modelo de regresión básico, que sirvió como punto de partida para comprender la relación entre la efectividad de la vacunación y las variables explicativas seleccionadas. El resumen se encuentra en la ilustración 38.

*Ilustración 38 Resumen del modelo espacial básico*

REGRESSION				
-----				
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION				
Data set	: Multivariado_efectividad			
Dependent Variable	: IMP_Efec	Number of Observations	: 337	
Mean dependent var	: 9,96249	Number of Variables	: 5	
S.D. dependent var	: 11,2197	Degrees of Freedom	: 332	
R-squared	: 0,159786	F-statistic	: 15,7844	
Adjusted R-squared	: 0,149663	Prob(F-statistic)	: 7,73456e-12	
Sum squared residual	: 35644	Log likelihood	: -1263,6	
Sigma-square	: 107,361	Akaike info criterion	: 2537,21	
S.E. of regression	: 10,3615	Schwarz criterion	: 2556,31	
Sigma-square ML	: 105,768			
S.E of regression ML	: 10,2844			
-----				
Variable	Coefficient	Std. Error	t-Statistic	Probability
-----				
CONSTANT	4,04024	3,21055	1,25843	0,20912
ESTRATO_MO	-0,718388	0,452715	-1,58684	0,11350
densidad K	0,0002846	5,52006e-05	5,15574	0,00000
Vuln	0,764363	1,05577	0,723985	0,46958
IPM_Media	0,0710958	0,104372	0,681174	0,49623
-----				

*Fuente: Propia*

### 10.1 ANÁLISIS DE REGRESIÓN ESPACIAL BÁSICO

- 1. R-cuadrado:** El R-cuadrado es 0,159786, lo que indica que aproximadamente el 15,98% de la variabilidad en la variable dependiente (IMP\_Efec) puede ser explicada por las variables independientes incluidas en el modelo.
- 2. F-Estadístico:** El valor del F-Estadístico es 15,7844, con un valor p muy pequeño (7,73456e-12), lo que sugiere que al menos una de las variables independientes tiene un efecto significativo en la variable dependiente.
- 3. Variables Independientes:**
  - La densidad poblacional (densidad K) tiene un coeficiente positivo significativo (0,0002846) con un valor p muy pequeño (0,00000), lo que sugiere que a medida que aumenta la densidad poblacional, la variable dependiente también tiende a aumentar.

- El nivel de vulnerabilidad (Vuln), el índice de pobreza multidimensional (IPM\_Media) y el estrato socioeconómico (ESTRATO\_MO) no parecen tener un efecto significativo en la variable dependiente, dado que sus coeficientes tienen valores p altos (por encima de 0,05).

### 10.1.1 Diagnósticos:

- **Multicolinealidad:** El número de condición de multicolinealidad es bajo (13,410967), lo que sugiere que no presenta alta multicolinealidad entre las variables independientes.
- **Normalidad de Errores:** El test de Jarque-Bera indica que hay desviaciones significativas de la normalidad en los errores del modelo (valor p muy bajo).
- **Heterocedasticidad:** Los tests de Breusch-Pagan y Koenker-Bassett sugieren la presencia de heterocedasticidad en los errores del modelo.
- **Dependencia Espacial:** Los tests de Moran's I y Lagrange Multiplier indican la presencia de dependencia espacial en los errores del modelo, lo que sugiere que los errores no están distribuidos de manera aleatoria en el espacio.

En el marco del modelo básico de regresión espacial, utilizado para investigar el impacto de la efectividad de la vacunación, se observó que la densidad poblacional emergió como la variable más significativa, con un valor de probabilidad (p-value) notablemente bajo de  $p = 0,00000$ . Este resultado sugiere una relación altamente significativa entre la densidad poblacional y la efectividad de la vacunación, lo que indica que la densidad poblacional ejerce una influencia considerable en la variable dependiente en este contexto.

Además, se identificó que la segunda variable más explicativa fue el estrato socioeconómico, con un valor de probabilidad de  $p = 0,11350$ . Aunque ligeramente menos significativo que la densidad poblacional, este hallazgo sugiere que el estrato socioeconómico también desempeña un papel importante en la explicación de la efectividad de la vacunación en el modelo básico de regresión espacial.

Posteriormente se realizó el análisis de regresión de retardo espacial y de error espacial

## 10.2 ANÁLISIS DE REGRESIÓN DE RETARDO ESPACIAL:

La ilustración 39 presenta el resumen del modelo de regresión de retardo espacial.

*Ilustración 39 Resumen del modelo de regresión de retardo espacial*

```
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set      : Multivariado_efectividad
Spatial Weight : Barrio_Cali_r
Dependent Variable : IMP_Efec  Number of Observations: 337
Mean dependent var : 9,96249  Number of Variables : 6
S.D. dependent var : 11,2197  Degrees of Freedom : 331
Lag coeff. (Rho) : 0,381396

R-squared      : 0,197363  Log likelihood      : -1257,5
Sq. Correlation : -        Akaike info criterion : 2526,99
Sigma-square   : 101,038  Schwarz criterion   : 2549,91
S.E of regression : 10,0518
```

Variable	Coefficient	Std.Error	z-value	Probability
W_IMP_Efec	0,381396	0,109548	3,48153	0,00050
CONSTANT	1,84263	3,15834	0,583419	0,55961
ESTRATO_MO	-0,627588	0,439286	-1,42865	0,15310
densidad K	0,000210573	5,64694e-05	3,72898	0,00019
Vuln	0,792073	1,02423	0,773337	0,43932
IPM_Media	0,0133948	0,102073	0,131228	0,89559

*Fuente: Propia*

- R-cuadrado:** El R-cuadrado es 0,197363, lo que indica que aproximadamente el 19,74% de la variabilidad en la variable dependiente (IMP\_Efec) puede ser explicada por el modelo de retardo espacial.
- Coefficiente de Retardo Espacial (Rho):** El coeficiente de retardo espacial es 0,381396, lo que indica la presencia de autocorrelación espacial positiva. Esto sugiere que los valores de IMP\_Efec en una ubicación están correlacionados con los valores de IMP\_Efec en ubicaciones vecinas.
- Variables Independientes:**
  - El coeficiente de W\_IMP\_Efec (retardo espacial de IMP\_Efec) es positivo y significativo (0,381396,  $p=0,00050$ ), lo que indica que hay un efecto significativo de la influencia espacial en la variable dependiente. Esto significa que el valor de IMP\_Efec en una ubicación está influenciado por el valor de IMP\_Efec en ubicaciones vecinas.
  - Las otras variables independientes (CONSTANTE, ESTRATO\_MO, densidad K, Vuln, IPM\_Media) no parecen tener efectos significativos en la variable dependiente, dado que sus coeficientes tienen valores p altos (por encima de 0,05).

### 10.2.1 Diagnósticos:

- Heterocedasticidad:** El test de Breusch-Pagan indica la presencia de heterocedasticidad en los errores del modelo.

- **Dependencia Espacial:** El test de Dependencia Espacial sugiere la presencia de autocorrelación espacial en los errores del modelo.

En el análisis de regresión de retardo espacial, se examinó el impacto de la efectividad de la vacunación, donde la variable más significativa fue la densidad poblacional (densidad K), con un valor de probabilidad (p-value) notablemente bajo de  $p = 0,00019$ . Este resultado sugiere una relación altamente significativa entre la densidad poblacional y la efectividad de la vacunación, indicando que este factor tiene una influencia sustancial en la variable dependiente en este contexto.

Además, se identificó que la segunda variable más explicativa fue el peso espacial (W\_IMP\_Efec), con un valor de probabilidad de  $p = 0,00050$ . Aunque ligeramente menos significativo que la densidad poblacional, este hallazgo sugiere que el peso espacial también desempeña un papel importante en la explicación de la efectividad de la vacunación en el análisis de regresión de retardo espacial.

### 10.3 ANÁLISIS DE REGRESIÓN ERROR ESPACIAL:

La ilustración 40 presenta el resumen del modelo de regresión con error espacial.

*Ilustración 40 Resumen del modelo de regresión con error espacial*

```
SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set          : Multivariado_efectividad
Spatial Weight    : Barrio_Cali_r
Dependent Variable : IMP_Efec  Number of Observations: 337
Mean dependent var : 9,962493  Number of Variables   : 5
S.D. dependent var : 11,219747  Degrees of Freedom    : 332
Lag coeff. (Lambda) : 0,391457

R-squared         : 0,189375  R-squared (BUSE)      : -
Sq. Correlation   : -          Log likelihood         : -1259,258975
Sigma-square      : 102,044   Akaike info criterion : 2528,52
S.E of regression : 10,1017   Schwarz criterion     : 2547,62
```

Variable	Coefficient	Std.Error	z-value	Probability
CONSTANT	4,16192	3,33283	1,24876	0,21175
ESTRATO_MO	-0,606917	0,448113	-1,35438	0,17561
densidad K	0,000239882	6,11333e-05	3,92392	0,00009
Vuln	0,862675	1,06433	0,810534	0,41763
IPM_Media	0,0529969	0,11316	0,468336	0,63954
LAMBDA	0,391457	0,120423	3,2507	0,00115

*Fuente: Propia*

1. **R-cuadrado:** El R-cuadrado es 0,189375, lo que indica que aproximadamente el 18,94% de la variabilidad en la variable dependiente (IMP\_Efec) puede ser explicada por el modelo de error espacial.
2. **Coefficiente de Error Espacial (Lambda):** El coeficiente de error espacial es 0,391457, lo que indica la presencia de autocorrelación espacial en los errores del modelo. Esto sugiere que los errores en una ubicación están correlacionados con los errores en ubicaciones vecinas.

### 3. Variables Independientes:

- Las variables independientes incluidas en el modelo son CONSTANT (constante), ESTRATO\_MO, densidad K, Vuln e IPM\_Media.
- Ninguna de las variables independientes parece tener un efecto significativo en la variable dependiente (IMP\_Efec) debido a que los valores p son altos (por encima de 0,05).

#### 10.3.1 Diagnósticos:

- **Heterocedasticidad:** El test de Breusch-Pagan indica la presencia de heterocedasticidad en los errores del modelo.
- **Dependencia Espacial:** El test de Dependencia de Error Espacial sugiere la presencia de autocorrelación espacial en los errores del modelo.

En el análisis de regresión espacial con error espacial, se investigó el impacto de la efectividad de la vacunación, donde la variable más significativa fue la densidad poblacional (densidad K), con un valor de probabilidad (p-value) de  $p = 0,00009$ . Este resultado sugiere una relación altamente significativa entre la densidad poblacional y la efectividad de la vacunación, indicando que este factor tiene una influencia sustancial en la variable dependiente en este contexto.

Además, se identificó que la segunda variable más explicativa fue el estrato socioeconómico (ESTRATO\_MO), con un valor de probabilidad de  $p = 0,17561$ . Aunque ligeramente menos significativo que la densidad poblacional, este hallazgo sugiere que el estrato socioeconómico también desempeña un papel importante en la explicación de la efectividad de la vacunación en el análisis de regresión espacial con error espacial.

Los comparativos de las tablas 6, 7 y 8 presentan una visión detallada de los resultados obtenidos para cada modelo, incluyendo el coeficiente de determinación (R-cuadrado), el logaritmo de la verosimilitud, los criterios de información de Akaike y Schwarz, así como el error estándar de la regresión. Estas métricas nos permiten evaluar la calidad de ajuste y la complejidad de cada modelo, facilitando la selección del enfoque más apropiado para el análisis de nuestros datos espaciales.

Tabla 6 Cuadro comparativo modelos de regresión espacial

Variable	Modelo Básico	Retardo Espacial	Error Espacial
R-cuadrado	0,159786	0,197363	0,189375
R-cuadrado Ajustado	0,149663	-	-
Sigma-square	107361	101038	102044
S.E. of regression	103615	100518	101017
Log Likelihood	-1263,6	-1257,5	-1259,3
Criterio de Información de Akaike (AIC)	2537,21	2526,99	2528,52
Criterio de Información de Schwarz (BIC)	2556,31	2549,91	2553,02

*Tabla 7 Cuadro comparativo de los modelos de regresión espacial*

Variable	Modelo Básico	Retardo Espacial	Error Espacial
densidad K	0,0002846	0,000210573	0,000239882
Vuln	0,764363	0,792073	0,862675
IPM_Media	0,0710958	0,0133948	0,0529969
ESTRATO_MO	-0,718388	-0,627588	-0,606917
W_IMP_Efec	-	0,381396	-

*Fuente: Propia*

*Tabla 8 Cuadro comparativo probabilidad de los modelos de regresión*

Variable	Modelo Básico	Retardo Espacial	Error Espacial
densidad K	0,00000	0,00019	0,00009
Vuln	0,46958	0,43932	0,41763
IPM_Media	0,49623	0,89559	0,63954
ESTRATO_MO	0,11350	0,15310	0,17561
W_IMP_Efec	-	0,00050	-

*Fuente: Propia*

En el Modelo Básico y en el Modelo de Error Espacial, la variable más significativa fue la densidad poblacional (densidad K), con un valor de probabilidad extremadamente bajo ( $p = 0,00000$  y  $p = 0,00009$  respectivamente). Esto sugiere que la densidad poblacional tiene un impacto altamente significativo en la variable dependiente en estos dos modelos. La segunda variable más significativa en ambos modelos fue el nivel de vulnerabilidad (Vuln), con valores de probabilidad cercanos a los de la densidad poblacional, aunque ligeramente más altos ( $p = 0,46958$  y  $p = 0,41763$  respectivamente).

Por otro lado, en el Modelo de Retardo Espacial, la variable más significativa fue el peso del impacto efectivo (W\_IMP\_Efec), con un valor de probabilidad de 0,00050. Sin embargo, la densidad poblacional nuevamente mostró un alto grado de significancia, siendo la segunda variable más significativa en este modelo, con un valor de probabilidad de 0,00019.

Estos hallazgos destacan la importancia de la densidad poblacional como un factor clave en la explicación de la variable dependiente en los modelos de regresión espacial considerados. Además, señalan la relevancia del nivel de vulnerabilidad y el peso del impacto efectivo como variables secundarias que también influyen en la variable dependiente en diferentes contextos espaciales.

## 11. TRABAJOS FUTUROS

Se propone los siguientes trabajos para nuevas propuestas y dirección de investigaciones adicionales y destacar áreas que requieren mayor estudio o desarrollo:

- **Validación de modelos:** Realizar un estudio longitudinal para validar los modelos desarrollados en la investigación, comparando las predicciones con datos reales a lo largo del tiempo para evaluar su robustez y precisión.
- **Incorporación de nuevos datos:** Integrar nuevos datos demográficos, de salud pública o socioeconómicos que puedan estar disponibles en el futuro para enriquecer el análisis y mejorar la precisión de los modelos.
- **Análisis de factores temporales:** Investigar cómo factores temporales, como las campañas de vacunación específicas, las variaciones estacionales o la evolución de la pandemia, pueden influir en la efectividad de la vacunación a lo largo del tiempo.
- **Estudio de intervenciones específicas:** Evaluar el impacto de intervenciones específicas o políticas de salud pública en la efectividad de la vacunación, utilizando modelos de simulación o análisis de series temporales.
- **Análisis de variabilidad geográfica:** Explorar la variabilidad geográfica de otros indicadores de salud relacionados con la vacunación, como la incidencia de casos, la mortalidad o la utilización de servicios de salud, para comprender mejor el contexto de la efectividad de la vacunación.
- **Estudios Cualitativos:** Realizar estudios cualitativos, como entrevistas o grupos focales, con diferentes grupos de la población para entender las percepciones, actitudes y barreras relacionadas con la vacunación y su efectividad.
- **Inclusión de variables culturales y sociales:** Explorar cómo factores culturales, sociales y de comportamiento pueden influir en la efectividad de la vacunación y desarrollar modelos que integren estas variables en el análisis.
- **Extensión a otras áreas geográficas:** Adaptar y aplicar la metodología desarrollada a otras ciudades o regiones para comparar y contrastar los resultados, identificando patrones comunes y diferencias contextuales.
- **Inclusión de computación en la nube:** Añadir tecnologías de computación en la nube que permita actualizar la información en tiempo real y tener informes o tableros que muestren los resultados en línea.

## 12. CONCLUSIONES

- Los resultados de este estudio muestran una distribución espacial significativa de los casos de COVID-19 en Santiago de Cali, evidenciando el impacto del plan nacional de vacunación en la efectividad observada en distintos barrios de la ciudad. Esta efectividad se refleja claramente en la identificación de clusters, indicando una tendencia espacial marcada. Los análisis de correlación de Moran, tanto a nivel global como local, han revelado agrupaciones geográficas, siendo las comunas del oriente de Cali, específicamente las comunas 12, 13, 14, 15, 16 y 21, las más afectadas.
- Las comunas, con un alto número de casos y fallecidos por COVID-19, muestran condiciones socioeconómicas, poblacionales y de salud con una marcada tendencia espacial. Variables como estrato socioeconómico, densidad poblacional, Índice de Pobreza Multidimensional y vulnerabilidad, han surgido como factores significativos. Específicamente, la densidad poblacional, seguida de la contigüidad espacial y el estrato socioeconómico, son determinantes que explican las variaciones en la efectividad de la vacunación entre zonas.
- Es relevante señalar que la efectividad de la vacunación ha sido mayor en áreas con un alto índice de pobreza multidimensional, lo que sugiere que las condiciones socioeconómicas y las barreras de acceso a la salud han influido en la salud pública de estas áreas.
- Es esencial reconocer las limitaciones de este estudio. Aunque se ha identificado una asociación entre factores socioeconómicos y efectividad de la vacunación, no se puede establecer una relación causal directa. Además, la disponibilidad y calidad de los datos podrían haber afectado los resultados.
- Estos hallazgos subrayan la necesidad de considerar factores socioeconómicos y geográficos al planificar y ejecutar programas de vacunación. Este análisis ofrece una base robusta para investigaciones futuras y el diseño de intervenciones dirigidas a abordar las disparidades en la efectividad de la vacunación en Santiago de Cali y áreas similares.
- Un estudio que emplea modelos espaciales avanzados, como la regresión espacial y el análisis de correlación de Moran, aporta significativamente a la ciencia de datos al mejorar la precisión de los modelos, identificar patrones espaciales y factores influyentes, validar la robustez de los análisis, y generalizar hallazgos a diferentes contextos geográficos. Estos enfoques permiten una comprensión más profunda de los fenómenos estudiados, fomentan la interdisciplinariedad y colaboración entre áreas, y contribuyen al desarrollo de soluciones aplicables a problemas complejos en salud pública y epidemiología.

### 13. REFERENCIAS BIBLIOGRÁFICAS

- [1] J. M. Rosa and E. L. Frutos, “Ciencia de datos en salud: desafíos y oportunidades en América Latina,” *Revista Médica Clínica Las Condes*, vol. 33, no. 6, pp. 591–597, Nov. 2022, doi: 10.1016/J.RMCLC.2022.09.007.
- [2] A. Mollalo, A. Mohammadi, S. Mavaddati, and B. Kiani, “Spatial Analysis of COVID-19 Vaccination: A Scoping Review,” *Int J Environ Res Public Health*, vol. 18, no. 22, Nov. 2021, doi: 10.3390/IJERPH182212024.
- [3] “Colombia: WHO Coronavirus Disease (COVID-19) Dashboard With Vaccination Data | WHO Coronavirus (COVID-19) Dashboard With Vaccination Data.” <https://covid19.who.int/region/amro/country/co> (accessed Jun. 10, 2023).
- [4] Instituto Nacional de Salud, “Indicadores COVID-19 Colombia.” <https://app.powerbi.com/view?r=eyJrIjoiN2ZIM2NmOTgtOWMwYS00NDNILTikMWItOGMyNWYyNmQ5NmM5IiwidCI6ImE2MmQ2YzdiLTImNTktNDQ2OS05MzU5LTM1MzcxNDc1OTRiYiIsImMiOiJlR9&pageName=ReportSection6d23b54da26648396033> (accessed Jun. 10, 2023).
- [5] G. E. Morales, A. E. Hurtado, and S. General, “FERNANDO RUÍZ GÓMEZ Ministro de Salud y Protección Social MARIA ANDREA GODOY CASADIEGO Viceministra de Protección Social CLAUDIA MILENA CUELLAR SEGURA Subdirectora de Enfermedades Transmisibles”.
- [6] World Health Organization, “Coronavirus disease (COVID-19).” [https://www.who.int/health-topics/coronavirus#tab=tab\\_1](https://www.who.int/health-topics/coronavirus#tab=tab_1) (accessed Jun. 10, 2023).
- [7] Ministerio Nacional de Salud, “Vacunación contra COVID-19.” <https://www.minsalud.gov.co/salud/publica/Vacunacion/Paginas/Vacunacion-covid-19.aspx> (accessed Jun. 10, 2023).
- [8] F. Ruiz Gómez *et al.*, “Ministerio de Salud y Protección Social Equipo Técnico y Coinvestigadores Asesora externa y coinvestigadora”.
- [9] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, “Applied Logistic Regression: Third Edition,” *Applied Logistic Regression: Third Edition*, pp. 1–510, Aug. 2013, doi: 10.1002/9781118548387.
- [10] D. G. Kleinbaum and Mitchel. Klein, “Survival analysis : a self-learning text,” p. 700, 2012.
- [11] E. Goodchild, “Investigaciones Regionales Asociación Española de Ciencia Regional”, Accessed: Jun. 10, 2023. [Online]. Available:

<http://www.redalyc.org/articulo.oa?id=28900609>

- [12] “Análisis espacial para aplicaciones | Relaciones y patrones espaciales.” <https://www.esri.co/es-co/productos/arcgis-platform/servicios/analisis-espacial> (accessed Jun. 10, 2023).
- [13] “Social Determinants of Health - Healthy People 2030 | health.gov.” <https://health.gov/healthypeople/priority-areas/social-determinants-health> (accessed Jun. 10, 2023).
- [14] G. Caspi *et al.*, “Socioeconomic disparities and COVID-19 vaccination acceptance: a nationwide ecologic study,” *Clin Microbiol Infect*, vol. 27, no. 10, pp. 1502–1506, Oct. 2021, doi: 10.1016/J.CMI.2021.05.030.
- [15] A. Whiteman *et al.*, “Demographic and Social Factors Associated with COVID-19 Vaccination Initiation Among Adults Aged  $\geq 65$  Years — United States, December 14, 2020–April 10, 2021,” *MMWR Morb Mortal Wkly Rep*, vol. 70, no. 19, pp. 725–730, 2021, doi: 10.15585/MMWR.MM7019E4.
- [16] [https://www.dane.gov.co/files/investigaciones/condiciones\\_vida/pobreza/2022/bol-pobreza-multidimensional-2022.pdf](https://www.dane.gov.co/files/investigaciones/condiciones_vida/pobreza/2022/bol-pobreza-multidimensional-2022.pdf)
- [17] Gustavo D. Buzai, “(PDF) Análisis Espacial con Sistemas de Información Geográfica. Sus cinco conceptos fundamentales.” Accessed: Jun. 06, 2024. [Online]. Available: [https://www.researchgate.net/publication/298420203\\_Analisis\\_Espacial\\_con\\_Sistemas\\_de\\_Informacion\\_Geografica\\_Sus\\_cinco\\_conceptos\\_fundamentales](https://www.researchgate.net/publication/298420203_Analisis_Espacial_con_Sistemas_de_Informacion_Geografica_Sus_cinco_conceptos_fundamentales).
- [18] J. Pablo Celemín, “Autocorrelación espacial e indicadores locales de asociación Autocorrelación espacial e indicadores locales de asociación espacial. Importancia, estructura y aplicación Spatial Autocorrelation and Local Indicators of Spatial Association. Importance, Structure and Application”.
- [19] “¿Qué es la simulación Montecarlo? | IBM.” Accessed: Feb. 02, 2024. [Online]. Available: <https://www.ibm.com/es-es/topics/monte-carlo-simulation>
- [20] J. G.-M. Willington Siabato, “La autocorrelación espacial y el desarrollo de la geografía cuantitativa,” *Cuadernos de Geografía*, vol. 28, pp. 2256–5442, 2019, doi: 10.15446/rcdg.v28n1.76919.
- [21] C. Chasco and F. A. López, “Modelos de regresión espacio-temporales en la estimación de la renta municipal: el caso de la Región de Murcia 1 Estimation of Municipal Income with Space-Time Regression Models: a Case for the Region of Murcia”.
- [22] Congreso de Colombia, “LEY 732,” 2002.

- [23] DANE, “Densidad de la población en Colombia.” Distribución de la población en Colombia. Accessed: Jun. 06, 2024. [Online]. Available: [https://geoportal.dane.gov.co/servicios/atlas-estadistico/src/Tomo\\_I\\_Demografico/2.2.3.-densidad-de-la-poblaci%C3%B3n-en-colombia.html](https://geoportal.dane.gov.co/servicios/atlas-estadistico/src/Tomo_I_Demografico/2.2.3.-densidad-de-la-poblaci%C3%B3n-en-colombia.html)
- [24] DANE, “Nota metodológica Índice de vulnerabilidad por manzana”.
- [25] J. Cordes and M. C. Castro, “Spatial analysis of COVID-19 clusters and contextual factors in New York City,” *Spat Spatiotemporal Epidemiol*, vol. 34, Aug. 2020, doi: 10.1016/J.SSTE.2020.100355.
- [26] D. E. Cuartas *et al.*, “Análisis espacio-temporal del SARS-coV-2 en Cali, Colombia,” *Revista de Salud Pública*, vol. 22, no. 2, pp. 138–143, Mar. 2020, doi: 10.15446/rsap.v22n2.86431.
- [27] D. Arango-Londoño, D. Ortega-Lenis, P. Moraga, M. Torres, and F. J. Rodríguez-Cortés, “Spatial modeling and socioeconomic inequities of COVID-19 in the urban area of the city of Cali, Colombia,” *Spat Spatiotemporal Epidemiol*, vol. 44, p. 100561, Feb. 2023, doi: 10.1016/J.SSTE.2022.100561.
- [28] F. Ruiz Gómez *et al.*, “Ministerio de Salud y Protección Social Equipo Técnico y Coinvestigadores Asesora externa y coinvestigadora”.
- [29] “Guía de CRISP-DM de IBM SPSS Modeler | IBM.” Accessed: Feb. 08, 2024. [Online]. Available: [https://www.ibm.com/docs/es/SS3RA7\\_18.4.0/pdf/ModelerCRISPDM.pdf](https://www.ibm.com/docs/es/SS3RA7_18.4.0/pdf/ModelerCRISPDM.pdf)
- [30] “PostgreSQL: The world’s most advanced open source database.” Accessed: Jun. 06, 2024. [Online]. Available: <https://www.postgresql.org/>
- [31] “Bienvenido al proyecto QGIS!” Accessed: Jun. 09, 2023. [Online]. Available: <https://www.qgis.org/es/site/>
- [32] “GeoDa on Github.” Accessed: Nov. 02, 2023. [Online]. Available: <https://geodacenter.github.io/>
- [33] “Welcome to Python.org.” Accessed: Dec. 26, 2023. [Online]. Available: <https://www.python.org/>
- [34] “R: The R Project for Statistical Computing.” Accessed: Oct. 21, 2023. [Online]. Available: <https://www.r-project.org/>

## 14. ANEXOS

### ANEXO 1: CARTA SECRETARIA RESPUESTA PROYECTO INVESTIGACIÓN



Al contestar por favor cite estos datos:

Radicado No.: 202341450100074191

Fecha: 2023-10-09

TRD: 4145.010.14.12.187.007419

Rad. Padre: 202341450100074191

GERMAN AVILA RODRIGUEZ  
Investigador  
Universidad Javeriana

Asunto: Respuesta Proyecto de Investigación titulado Análisis Espacial de la Efectividad de la vacunación por COVID-19 y su asociación con factores socioeconómicos en Santiago de Cali.

Cordial saludo

El comité de Ética en investigación de la Secretaría de Salud del Distrito Especial de Santiago de Cali, revisó el proyecto de investigación titulado "Análisis Espacial de la Efectividad de la vacunación por COVID-19 y su asociación con factores socioeconómicos en Santiago de Cali", de la investigador German Avila Rodriguez, Alejandro Valencia y Oscar Muñoz.

El comité, luego de la evaluación de los documentos enviados por los investigadores, considera que el proyecto es aprobado para su realización.

El nivel de riesgo del proyecto es Mínimo.

Cualquier modificación al proyecto debe ser revisada y aprobada por el comité, los investigadores también informarán problemas relacionados a la recolección de datos o imprevistos que pudiesen resultar en riesgo para los sujetos. De igual manera cada se debe enviar reporte de avance cada 3 meses.

El comité podrá revocar la aprobación a un proyecto de investigación en desarrollo, cuando se compruebe que hay suministro de datos falsos, existencia de riesgos adicionales no mencionados en la presentación inicial, deficiencias en la realización del estudio. La revocatoria del aval tendrá efectos inmediatos y será notificada vía correo electrónico.



Calle 4B N° 36-00 San Fernando  
[www.cali.gov.co](http://www.cali.gov.co)



El aval del proyecto tiene una vigencia de un año a partir de la fecha de este oficio; en caso de requerirlo, el investigador solicitará una ampliación del tiempo del aval. Igualmente, en caso de requerir una enmienda por cambios sustanciales en el proyecto, este tendrá que ser remitido al presidente del comité quien decidirá si avala la enmienda o si somete la decisión al comité en pleno.

Con toda atención

FIRMA

CONCHA EASTMAN

Presidente

Comité de Ética en Investigación

Secretaría de Salud del Distrito Especial de Santiago de Cali

Elaboró: Angela María Toro Duque - Contratista  
Revisó: Alberto Concha Eastman - contratista



SC-C33847981



Calle 4B N° 36-00 San Fernando  
[www.cali.gov.co](http://www.cali.gov.co)

## ANEXO 2 CODIGO: PREPARACIÓN COVID\_TG

```
import pandas as pd
import numpy as np
import seaborn as sns
from datetime import datetime
# Graficos
import seaborn as sns
import matplotlib.pyplot as plt
import statsmodels.api as sm

pd.set_option('display.max_column', None)
pd.reset_option('display.max_row')

data = pd.read_excel("covid_bar_geo.xlsx")
casos_positivos = pd.DataFrame(data)
casos_positivos.head()

# Contar los valores únicos en la columna 'Recuperado definitivo'
conteo_recuperado = casos_positivos['Recuperado definitivo'].value_counts()

# Crear el gráfico de barras
plt.bar(conteo_recuperado.index, conteo_recuperado.values)

# Añadir etiquetas de datos encima de cada barra
for i, valor in enumerate(conteo_recuperado):
    plt.text(i, valor, str(valor), ha='center', va='bottom')

# Añadir etiquetas al gráfico
plt.xlabel('Recuperado definitivo')
plt.ylabel('Recuento')
plt.title('Recuento de casos por Recuperado definitivo')

# Mostrar el gráfico
plt.show()

casos_positivos['Recuperado definitivo'].value_counts()

# convertimos a tipo fecha las variables Fechadehospitalización, Fechade alta, Fecha de
hospitalizacion SIVIGILA
casos_positivos['iniciodeSintomas_'] = pd.to_datetime(casos_positivos['iniciodeSintomas_'])

# Crear una nueva columna para el año y mes de 'order_created_time'
casos_positivos['año_mes'] = casos_positivos['iniciodeSintomas_'].dt.to_period('M')

casos_positivos_fallecidos = casos_positivos[casos_positivos['Recuperado definitivo'] ==
'fallecido']
casos_positivos_fallecidos.head()

# Agrupar por año_mes, ID_BARRIO, COD_BARRIO, COD_COMUNA y NOMBRE
casos_positivos_fallecidos_g = casos_positivos_fallecidos.groupby(['año_mes',
'COD_BARRIO', 'COD_COMUNA', 'NOMBRE']).size().reset_index(name='conteo')

casos_positivos_fallecidos_g = casos_positivos_fallecidos_g.sort_values(by='conteo',
ascending=False)
```

```

# Mostrar el resultado
casos_positivos_fallecidos_g.head(10)

# Crear una tabla pivotante con recuento distintivo de casos
casos_fallecidos_pivot = pd.pivot_table(casos_positivos_fallecidos_g, values='conteo',
index='COD_BARRIO', columns='año_mes', aggfunc=sum, fill_value=0)
casos_fallecidos_pivot

convertir el indice en un atributo
casos_fallecidos_pivot.reset_index(inplace=True)
casos_fallecidos_pivot.rename(columns={'index': 'COD_BARRIO'}, inplace=True)

casos_fallecidos_pivot.to_excel('casos_fallecidos.xlsx', index=True)

Fallecidos totales

# Agrupar por año_mes, ID_BARRIO, COD_BARRIO, COD_COMUNA y NOMBRE
casos_positivos_g = casos_positivos.groupby(['año_mes', 'COD_BARRIO', 'COD_COMUNA',
'NOMBRE']).size().reset_index(name='conteo')

casos_positivos_g = casos_positivos_g.sort_values(by='conteo', ascending=False)

# Mostrar el resultado
casos_positivos_g.head(10)

# Crear una tabla pivotante con recuento distintivo de 'dbname'
casos_positivos_total_pivot = pd.pivot_table(casos_positivos_g, values='conteo',
index='COD_BARRIO', columns='año_mes', aggfunc=sum, fill_value=0)

casos_positivos_total_pivot

convertir el indice en un atributo
casos_positivos_total_pivot.reset_index(inplace=True)
casos_positivos_total_pivot.rename(columns={'index': 'COD_BARRIO'}, inplace=True)

casos_positivos_total_pivot.columns

casos_positivos_total_pivot.to_excel('casos_positivos_total.xlsx', index=True)

Renombrar las columnas de casos_fallecidos_pivot
casos_fallecidos_pivot = casos_fallecidos_pivot.add_suffix('-F')

Renombrar las columnas de casos_positivos_total_pivot
casos_positivos_total_pivot = casos_positivos_total_pivot.add_suffix('-C')

Unir los DataFrames en base al índice
Covid_Casos_Fallecidos = casos_fallecidos_pivot.join(casos_positivos_total_pivot)

Covid_Casos_Fallecidos.head()

Covid_Casos_Fallecidos.to_excel('Covid_Casos_Fallecidos.xlsx', index=True)

```

### **ANEXO 3: CODIGO PREPARACIÓN DANE**

```

import pandas as pd
import numpy as np
import seaborn as sns
# Graficos
import seaborn as sns
import matplotlib.pyplot as plt

DANE = pd.read_excel("DANE.xlsx")
DANE.head()

pd.set_option('display.max_column', None)

# Redondear la variable 'ipm' a 0 decimales
DANE['ipm'] = DANE['ipm'].round(0)

# Reemplazar los valores en la columna "LABEL"
DANE['LABEL'] = DANE['LABEL'].replace({'Vulnerabilidad media-baja': 'Vulnerabilidad
baja', 'Vulnerabilidad media-alta': 'Vulnerabilidad media'})

# Crear un DataFrame agrupado por ID_BARRIO y calcular la media, mediana y moda de ipm
agrupado =
DANE.groupby(['ID_BARRIO', 'COD_BARRIO', 'COD_COMUNA', 'NOMBRE', 'POB_2017', 'COD_
DPTO', 'COD_MPIO'])['ipm'].agg(['mean', 'median']).reset_index()

# Calcular la moda de LABEL, embarazo_a, reactivaci por ID_BARRIO
moda_label = DANE.groupby('ID_BARRIO')['LABEL'].apply(lambda x: x.mode().iloc[0] if not
x.mode().empty else None).reset_index()
moda_embarazo_a = DANE.groupby('ID_BARRIO')['embarazo_a'].apply(lambda x:
x.mode().iloc[0] if not x.mode().empty else None).reset_index()
moda_reactivaci = DANE.groupby('ID_BARRIO')['reactivaci'].apply(lambda x: x.mode().iloc[0] if
not x.mode().empty else None).reset_index()
moda_ipm = DANE.groupby('ID_BARRIO')['ipm'].apply(lambda x: x.mode().iloc[0] if not
x.mode().empty else None).reset_index()

# Fusionar el DataFrame de moda del IPM con el DataFrame agrupado por ipm
DANE_Barrio = pd.merge(agrupado, moda_ipm, on='ID_BARRIO')
DANE_Barrio = pd.merge(DANE_Barrio, moda_label, on='ID_BARRIO')
DANE_Barrio = pd.merge(DANE_Barrio, moda_embarazo_a, on='ID_BARRIO')
DANE_Barrio = pd.merge(DANE_Barrio, moda_reactivaci, on='ID_BARRIO')

DANE_Barrio.rename(columns={'ipm': 'mode_ipm'}, inplace=True)
DANE_Barrio.rename(columns={'median': 'median_ipm'}, inplace=True)
DANE_Barrio.rename(columns={'mean': 'mean_ipm'}, inplace=True)
DANE_Barrio.head(20)

DANE_Barrio['LABEL'] = DANE_Barrio['LABEL'].replace({'Vulnerabilidad baja': 1,
'Vulnerabilidad media': 2})

multi = pd.read_excel("Multivariado.xlsx")
multi.head()

# Merge para agregar la columna LABEL a multi basada en la columna ID_BARRIO
multi = pd.merge(multi, DANE_Barrio[['ID_BARRIO', 'LABEL', 'mean_ipm']], on='ID_BARRIO',
how='left')

# Ahora el DataFrame multi tiene la columna LABEL

```

```

multi.head()

# Eliminar la columna mean_IPM
multi.drop('mean_IPM', axis=1, inplace=True)

# Multiplicar por -1 el impacto y renombrar la columna
multi['impacto_1'] = multi['impacto'] * -1

# Eliminar la columna original de impacto
multi.drop('impacto', axis=1, inplace=True)

#multi.to_excel('Multivariado_L.xlsx', index=True)

```

## ANEXO 4: MODELADO

```

install.packages("readxl")
install.packages("ggplot2")
install.packages("gridExtra")
library(readxl)
library(gridExtra)

##calculo de impacto

fallecidos = read_excel("C:/Users/Racso/Desktop/Maestria/Trabajo de
Grado/Covid_Casos_Fallecidos.xlsx")
fallecidos=data.frame(fallecidos)

# Función para calcular el impacto y graficar
calc_impacto<- function(barrio) {
  require(stringr)

  # Organiza los datos
  fecha <- names(fallecidos[barrio, 36:71])
  fecha <- paste("1.", substr(fecha, 2, 8), sep="")
  fecha <- as.Date(fecha, tryFormats = c("%d.%Y.%m"))
  casos <- as.numeric(fallecidos[barrio, 36:71])
  plan_vacuna <- c(rep("no", 14), rep("si", 22))
  codigo_barrio <- fallecidos$COD_BARRIO[barrio]
  df <- data.frame(fecha, casos, plan_vacuna)

  require(ggplot2)
  p <- ggplot(df, aes(x = fecha, y = casos, colour = plan_vacuna)) +
    geom_point() +
    geom_smooth(method = "gam") +
    theme_bw() +
    ggtitle(paste("Impacto del modelo de Poisson - Barrio", codigo_barrio))

  # Modelo poisson
  t <- 1:36
  df2 <- data.frame(df, t)
  mod1 <- glm(casos ~ plan_vacuna, data = df2, family = "poisson")
  impacto <- mod1$coefficients[2]

  return(list(plot = p, impacto = impacto))
}

```

```
}
```

```
# Llama la función para los cuatro barrios y almacena las gráficas e impactos  
resultados <- lapply(c(8, 7,249,330), calc_impacto)
```

```
# Extrae las gráficas y los impactos  
plots <- lapply(resultados, function(x) x$plot)
```

```
# Combina las gráficas  
grid.arrange(grobs = plots, ncol = 2)
```