

## **Mortalidad en paciente con insuficiencia renal crónica**

Johan Alexander Diaz Mayoral

Asesor

Julio Eduardo Mejia Manzano

Universidad Nacional Abierta y a Distancia UNAD

Escuela de Ciencias Básicas, Tecnología e Ingeniería ECBTI

Especialización de Ciencia de Datos y Analítica

2026

**Nota de Aceptación**

Nombre director de Trabajo de Grado

Jurado

Jurado

## Resumen

Se realiza un análisis de la mortalidad de los pacientes con insuficiencia renal crónica en el cual se toma los datos oficiales del DANE correspondientes al periodo 2019–2023 datos que se encuentra en sus microsítios, para identificar tendencias y variaciones relevantes que pueda aportar a la salud pública. De la información extraída la información consiste en salud de la mortalidad en todas sus patologías en el territorio colombiano encontrando más de 1 millón de registros con la información por lo cual se accede a realizar el ETL (Extracción Transformación y carga) desde la plataforma de DataBriks con lo cual se maneja BigData con herramienta Python y Spark para ser escalable e integrar diferentes fuentes de datos. Se filtra el diagnóstico CIE10 que contenga código N18 IRC (insuficiencia renal crónica) tomado registro de los pacientes fallecidos durante el periodo de tiempo luego se tomó información relacionada como departamento lugar, diagnostico fecha genero régimen de afiliación con el que falleció.

Esta información es porque se cuenta a nivel público y se puede utilizar para fines de estudio el cual se desarrolla con el objetivo de la posible influencia de muerte por grupos de sexo, seguridad social y ubicación geográfica. Utilizando técnicas de análisis descriptivo, tasas ajustadas y visualización gráfica. Este interés es por que IRC es costosa para el gobierno desmejora la calidad de vida y no tiene una cura lo único es por un trasplante de riñón las personas que la poseen tienen una tendencia a cambios de vida y aferrarse a tratamientos para su control para la sobrevivencia.

**Palabras clave:** Departamento administrativo nacional estadístico DANE, Extracción Transformación y carga ETL, Clasificación Internacional de Enfermedades CIE10, Insuficiencia Renal crónica IRC.

## Abstract

An analysis of mortality among patients with chronic kidney disease was conducted using official data from DANE (National Administrative Department of Statistics) for the period 2019–2023, available on their microsites. The aim was to identify relevant trends and variations that could contribute to public health. The extracted data comprises mortality information for all pathologies across Colombia, yielding over 1 million records. This information was then used to perform ETL (Extraction, Transformation, and Loading) on the DataBriks platform, which utilizes Big Data tools like Python and Spark for scalability and integration of diverse data sources. The ICD-10 diagnosis containing code N18 CKD (chronic kidney disease) was filtered from the records of patients who died during the specified period. Related information was then collected, including department, place of death, diagnosis, date of death, gender, and health insurance plan.

This information is publicly available and can be used for research purposes. The study aims to determine the potential influence of sex, social security, and geographic location on mortality rates. Descriptive analysis techniques, adjusted rates, and graphical visualization are used. This research is of interest because chronic kidney disease (CKD) is costly for the government, diminishes quality of life, and has no cure (the only option being a kidney transplant). People with CKD tend to make lifestyle changes and adhere to treatments to manage the condition and ensure their survival.

**Keywords:** National Administrative Department of Statistics DANE, Data Extraction, Transformation, and Loading ETL, International Classification of Diseases ICD10, Chronic Kidney Disease CKD.

## Tabla de Contenido

|   |    |
|---|----|
| Introducción .....  | 8  |
| Justificación .....   | 9  |
| Objetivos.....  | 10 |
| Objetivo General.....   | 10 |
| Objetivos Específicos .....   | 10 |
| Método .....  | 11 |
| Influencia de Variables de Pacientes con Insuficiencia Renal Crónica..... | 11 |
| Diseño del Estudio.....   | 11 |
| Variables .....   | 12 |
| Fuente de Datos .....   | 14 |
| Procesamiento de Datos.....   | 15 |
| Consideraciones Éticas .....  | 32 |
| Conclusiones.....   | 33 |
| Referencias Bibliográficas .....  | 35 |

### Lista de Tablas

|   |    |
|---|----|
| <b>Tabla 1</b> <i>Defunciones por Año de Pacientes con Insuficiencia Renal Crónica</i> .....              | 20 |
| <b>Tabla 2</b> <i>Defunciones Por Año de Pacientes Con Insuficiencia Renal Crónica</i> .....              | 20 |
| <b>Tabla 3</b> <i>Defunciones Nivel Educativo de Pacientes con Insuficiencia Renal Crónica</i> .....      | 21 |
| <b>Tabla 4</b> <i>Número de Fallecidos de Insuficiencia Renal Crónica Según el Departamento</i> .....     | 22 |
| <b>Tabla 5</b> <i>Número de Fallecidos con la Descripción del Tipo de Afiliación de Seguridad Social.</i> | 22 |
| <b>Tabla 6</b> <i>Descripción de Numero de Causas Básicas de la Defunción (c bas1)</i> .....              | 23 |
| <b>Tabla 7</b> <i>Defunciones Según el Área o Zona con la Descripción de Seguridad Social</i> .....       | 24 |
| <b>Tabla 8</b> <i>Defunciones con la Enfermedad de Base Frente al Genero</i> .....                        | 25 |
| <b>Tabla 9</b> <i>Cramér's V Mide la Fuerza de la Asociación Entre Variables Categóricas</i> .....        | 26 |

## Lista de Figuras

|   |    |
|---|----|
| <b>Figura 1</b> <i>Formula Correlaciones De Pearson's</i> .....   | 18 |
| <b>Figura 2</b> <i>Diagrama de Matriz de Correlación Pearson's</i> .....  | 19 |
| <b>Figura 3</b> <i>Defunciones Según el Rango de Edad de los Pacientes con Insuficiencia Renal Crónica</i><br>..... | 27 |
| <b>Figura 4</b> <i>Porcentaje de Muertes Relacionadas irc Según el Sexo con el Nivel Educativo</i> .....            | 28 |
| <b>Figura 5</b> <i>Tasas de los 5 con Mayor Población Fallecida por Municipio y Sexo</i> .....                      | 29 |
| <b>Figura 6</b> <i>Tasas Población Fallecida por Municipio</i> .....  | 29 |
| <b>Figura 7</b> <i>Número de Muertes por Año y Genero</i> .....   | 30 |
| <b>Figura 8</b> <i>Número de Muertes con Tipo de Afiliación en Salud Frente al Sexo</i> .....                       | 30 |
| <b>Figura 9</b> <i>Número de Muertes Comparadas en lo Urbano y Rural</i> .....                                      | 31 |
| <b>Figura 10</b> <i>Número de Muertes por Grupo de Edad Según el Rango</i> .....                                    | 31 |

## Introducción

En el contexto de la transformación digital, el uso de herramientas tecnológicas ha permitido que los sistemas computacionales aprendan a partir de datos, Facilitando la detección de patrones y tendencias útiles en diversas áreas, incluyendo la salud. A través del análisis de datos históricos, es posible identificar factores de riesgo y predecir eventos clínicos, como la probabilidad de mortalidad en pacientes con insuficiencia renal crónica, cuya función renal residual se encuentra comprometida o inexistente (Goldstein et al., 2017; Levey & Coresh, 2012). Dado el deterioro progresivo de estos pacientes, es fundamental evaluar su estado clínico con precisión y en el contexto de una población amplia, identificando similitudes y patrones que permitan establecer alertas tempranas. Se busca demostrar científicamente las probabilidades de muerte asociadas a diversos factores, contribuyendo a una intervención oportuna. “Analizar los factores demográficos, sociales y clínicos asociados a la mortalidad en pacientes con insuficiencia renal crónica en Colombia entre 2019 y 2023, utilizando técnicas de ciencia de datos para identificar patrones y variables de mayor impacto.” (Departamento Administrativo Nacional de Estadística [DANE], 2023).

Este trabajo tiene como objetivo establecer las especificaciones mínimas necesarias para el tratamiento médico de pacientes con insuficiencia renal crónica, con el fin de mejorar su condición clínica y por ende, su calidad de vida. (Organización Panamericana de la Salud [OPS], 2014). Un documento técnico de 2020 de la OMS discute un “marco para la atención integrada del riñón” (kidney-care) que pretende orientar estrategias desde la prevención, tratamiento conservador, trasplante o diálisis para quienes llegan a insuficiencia renal. (World Health Organization, 2020)

## Justificación

Uno de los grandes desafíos de la medicina moderna es prolongar la vida humana al mismo tiempo, garantizar una adecuada calidad de vida. Este objetivo es de gran valor para cada individuo, especialmente en el contexto de enfermedades crónicas como la insuficiencia renal crónica (IRC). A partir de esta necesidad, el presente proyecto busca aplicar el análisis de datos históricos de pacientes para identificar variables demográficas y clínicas que influyen en la defunción de pacientes evaluar su impacto en la evolución clínica del paciente. (Levey & Coresh, 2012).

La insuficiencia renal crónica es una enfermedad prevalente a nivel mundial, cuya única cura definitiva es el trasplante de riñón. Sin embargo, este procedimiento enfrenta múltiples limitaciones, entre ellas la escasez de donantes compatibles, quienes deben compartir características clínicas, químicas y biológicas específicas con el paciente receptor. Ante esta dificultad, el tratamiento clínico del riñón se convierte en la principal alternativa para mejorar la calidad de vida y reducir la progresión de la enfermedad. (Luyckx, Tonelli, & Stanifer, 2018).

Este trabajo se justifica en la posibilidad de aplicar técnicas de análisis de datos y aprendizaje automático para detectar patrones en datos históricos, que permitan predecir el deterioro de la función renal y automatizar protocolos clínicos. En la actualidad, gracias al desarrollo de tecnologías de extracción, transformación y carga de datos (ETL), junto con algoritmos predictivos avanzados, es posible transformar grandes volúmenes de información clínica en conocimiento útil para la toma de decisiones médicas. Esto permite no solo personalizar los tratamientos, sino también anticiparse a complicaciones, optimizando así los recursos del sistema de salud y mejorando los resultados en los pacientes. (Jiang et al., 2017).

## **Objetivos**

### **Objetivo General**

Analizar los factores asociados a la mortalidad en pacientes con insuficiencia renal crónica en Colombia entre 2019 y 2023, utilizando técnicas de ciencia de datos ETL, EDA

### **Objetivos Específicos**

Identificar patrones de mortalidad según variables demográficas (edad, sexo, departamento, área).

Evaluar diferencias en mortalidad según régimen de seguridad social en salud en Colombia.

Explorar la relación entre causas múltiples de defunción y la insuficiencia renal crónica.

## **Método**

### **Influencia de Variables de Pacientes con Insuficiencia Renal Crónica**

El objetivo del modelo no es realizar predicciones individuales, sino identificar los factores asociados a la mortalidad por insuficiencia renal crónica mediante técnicas de machine learning interpretables. Asimismo, busca analizar resultados de variables demográficas y clínicas en pacientes con diagnóstico de insuficiencia renal crónica.

La población de estudio comprende la información extraída de los indicadores de La Encuesta Longitudinal de Protección Social (ELPS) que recoge información sobre distintos aspectos que afectan la vida de las personas, como el acceso a la salud, las pensiones, las dificultades que enfrentan los hogares, como la pérdida de un familiar o problemas económicos. Estos datos ayudan a entender cómo viven las personas y qué necesidades tienen, lo que permite contribuir en la toma de decisiones sobre cómo se reparten los recursos en salud y bienestar y hacer un uso más eficiente de los recursos disponibles en beneficio de toda la población (Departamento Administrativo Nacional de Estadística [DANE], 2023).

### **Diseño del Estudio**

El presente proyecto se desarrolló bajo un enfoque aplicado, orientado a la utilización de técnicas de ciencia de datos para analizar los factores asociados a la mortalidad por insuficiencia renal crónica (IRC) en Colombia entre 2019 y 2023. El proyecto se estructura a partir del procesamiento, análisis y modelamiento de datos secundarios provenientes de fuentes oficiales, con el fin de generar información útil para la comprensión del fenómeno y para apoyar la toma de decisiones en salud pública. (DANE, 2023)

La ejecución del proyecto se fundamenta en la integración de datos extraídos de dos fuentes principales. La primera corresponde a los registros de defunciones por IRC identificados

mediante el código N18 de la Clasificación Internacional de Enfermedades (CIE-10), (Organización Panamericana de la Salud [OPS], 2014) disponibles en las bases de mortalidad publicadas por el Departamento Administrativo Nacional de Estadística (DANE) (DANE, 2023). Estos registros permiten identificar características demográficas, sociales y clínicas asociadas a las defunciones. La segunda fuente es la información proveniente de la Encuesta Longitudinal de Protección Social (ELPS), utilizada para contextualizar variables relacionadas con condiciones sociales, acceso a servicios de salud y características del hogar (DANE, 2023).

### **Variables**

El proyecto emplea un flujo metodológico basado en procesos de extracción, transformación y carga (ETL), seguido del análisis exploratorio y la aplicación de algoritmos orientados a identificar patrones y posibles factores asociados a la mortalidad. Las variables consideradas incluyen:

1. Código departamento: Código del dane Departamento donde ocurrió la defunción
2. Nombre departamento: Nombre del dane Departamento donde ocurrió la defunción
3. Código Area Defunción: Código Área donde ocurrió la defunción
4. Descripción Area Defunción: Descripción Área donde ocurrió la defunción
5. Código sexo: Código Sexo del fallecido
6. Descripción sexo: Descripción Sexo del fallecido
7. Código estado civil: Código del Estado conyugal del fallecido
8. Descripción estado civil: descripción Estado conyugal del fallecido
9. Código nivel educativo: Código Último nivel de estudios que aprobó el fallecido

10. Descripción nivel educativo: Descripción Último nivel de estudios que aprobó el fallecido
11. Código seguridad social: Código Régimen de seguridad social del fallecido
12. Descripción seguridad social: Descripción Régimen de seguridad social del fallecido
13. Código manera muerte: Código Probable manera de muerte
14. Descripción manera muerte: Descripción Probable manera de muerte
15. Código: Código de la causa Básica de la defunción codificación según el CIE10
16. Descripción: Descripción de Código de la causa Básica de la defunción codificación según el CIE10
17. CodigoRangoEdad1: Código Agrupación de edades, según la edad del fallecido (GRU ED1)
18. DescripcionRangoEdad1: Descripción Agrupación de edades, según la edad del fallecido (GRU ED1)
19. CodigoGrupoEdad2: Código Agrupación de edades, según la edad del fallecido (GRU ED2)
20. DescripcionGrupoEdad2: Descripción Agrupación de edades, según la edad del fallecido (GRU ED2)
21. Código lugar ocurrencia: Código Sitio donde ocurrió la defunción (SIT DEFUN)
22. Descripción lugar ocurrencia: Descripción Sitio donde ocurrió la defunción (SIT DEFUN)
23. Código grupo étnico: Código De acuerdo con la cultura, pueblo o rasgos físicos, el fallecido era o se reconocía como (IDPERTET)

24. descripción grupo étnico : Descripción De acuerdo con la cultura, pueblo o rasgos físicos, el fallecido era o se reconocía como (IDPERTET)
25. código ocupación causa defunción : Código La ocupación pudo ser causa o estar asociada con la defunción (MUERTEPORO)
26. descripción ocupación causa defunción : Descripción La ocupación pudo ser causa o estar asociada con la defunción (MUERTEPORO)
27. ANO: Descripción Año en que ocurrió la defunción (ANO)
28. MES: Descripción Mes en que ocurrió la defunción (MES)
29. CAUSA MULT: Descripción de causa múltiple para causas antecedentes (CAUSA MULT)
30. C BAS1: Descripción de la causa Básica de la defunción (C BAS1)

Departamento Administrativo Nacional de Estadística – DANE. (s.f.). Demografía y población [Base de datos]. DANE. <https://microdatos.dane.gov.co/index.php/catalog/DEM-Microdatos> (microdatos.dane.gov.co in Bing)

Este diseño aplicado permite organizar el proceso de análisis en etapas claramente definidas, garantizando que los datos se preparen, transformen y examinen de manera rigurosa, y que los resultados obtenidos contribuyan a identificar variables relevantes en la mortalidad.

### **Fuente de Datos**

Los datos utilizados en este proyecto provienen de fuentes oficiales y confiables a nivel nacional, administradas por el Departamento Administrativo Nacional de Estadística (DANE). La primera fuente corresponde a los registros de defunciones no fetales publicados en la serie de Estadísticas Vitales, los cuales consolidan la información reportada en los certificados de

defunción en todo el territorio nacional entre los años 2019 y 2023, las cuales contienen variables demográficas, sociales y clínicas relevantes para el análisis (DANE, 2023).

Estos registros permiten identificar la causa básica de muerte codificada bajo los lineamientos de la Clasificación Internacional de Enfermedades CIE-10, emitida por la Organización Panamericana de la Salud [OPS], 2014). Para este proyecto, se seleccionaron los registros cuya causa básica de muerte correspondía al código N18, asociado a la insuficiencia renal crónica.

La segunda fuente corresponde a la Encuesta Longitudinal de Protección Social (ELPS), administrada también por el DANE, la cual recopila información sobre condiciones sociales, económicas y de acceso a servicios de salud de los hogares colombianos. Esta fuente brinda un marco contextual relevante para comprender las condiciones de bienestar y las características sociodemográficas de la población (Departamento Administrativo Nacional de Estadística [DANE], 2023).

Los datos de mortalidad analizados abarcan el periodo comprendido entre 2019 y 2023, permitiendo identificar variaciones temporales y patrones asociados a factores demográficos y sociales. Estas bases integran variables como departamento de ocurrencia, área, sexo, seguridad social, manera de muerte y códigos CIE-10, las cuales resultan pertinentes para el análisis de factores relacionados con la mortalidad por insuficiencia renal crónica.

### **Procesamiento de Datos**

Extracción de datos: Para la ejecución del proyecto, los datos fueron obtenidos de fuentes oficiales y confiables, específicamente del Departamento Administrativo Nacional de Estadística (DANE). La extracción se realizó sobre dos bases de datos principales registros de mortalidad que contienen información de las defunciones ocurridas en Colombia entre 2019 y 2023. La otra

base de datos se realizó la extracción de (Organización Panamericana de la Salud [OPS], 2014) de la última publicación de actualización en donde se extrae el catálogo para categorizar “International statistical classification of diseases and related health problems, 10th revision” CIE10 OMS (2024).

Limpieza y transformación (ETL): Se realiza inicialmente unificación de toda la información contenida se tienen 5 archivos con información cada uno de 1 año luego se realiza unificación de la información con un total de Filas: 1928515 Columnas: 60 muchas columnas no son requeridas ya que no hacen parte del proyecto. Se realiza la categorización de cada una de las columnas agregando más columnas para la descripción y mejorar la información Filas: 1928515 Columnas: 92 según los catálogos se realiza una unión de columnas para determinar los valores categóricos de la descripción del código, Se realiza la selección de los datos de solo la información que en la columna CAUSA MULT: Causa agrupada en la que contiene el Código N18 por lo cual se reduce a Filas: 61340 Columnas: 30 con las que se realizara el análisis de los datos después de estar depurado transformado y cargada la información para el análisis requerido luego realizo almacenamiento en variables numéricas cada columna requerida y las categóricas almacenadas en otra variable.

Extracción de datos: los datos son extraídos del DANE según las estadísticas de defunciones no fetales se producen a partir de la información proveniente de los certificados de defunción, la cual se copia y consolida por las diferentes subdivisiones geográficas (departamentos, municipios y áreas) y por los diferentes grupos de población en el territorio nacional. Las bases de datos de defunciones, permiten obtener información por sexo, edad, causas de defunción, según departamento y municipio de ocurrencia de la defunción, el agrupamiento por edad, así como departamento y municipio de residencia del fallecido; la cual

se constituye en fuente básica para la construcción de indicadores en salud como tasa de mortalidad infantil, razón de mortalidad materna, tasas específicas por causas de muerte (enfermedades transmisibles, infecciosas, crónicas, muertes por causa externa etc.), para la formulación y evaluación de políticas, planes y programas de salud pública especialmente (Departamento Administrativo Nacional de Estadística [DANE], 2019).

Transformación de datos: Se realiza selección de datos relevantes para la obtención de las variables objetivo y las variables predictoras transformando los datos de forma que se puedan entender según su código y la descripción o la categórica para tener las variables tanto de forma categórica como numérica para poder realizar el tratamiento de datos. Identificando el número de ocurrencias por cada variable es decir cuantificando la veracidad de las predictoras para la correlación, en la variable CAUSA MULT es donde se encuentra la información objetivo para filtrar el código de diagnóstico CIE10 que contenga la palabra N18 el cual corresponde a las enfermedad de insuficiencia renal crónica obteniendo la información que falleció con alguna relación de la enfermedad luego se procede a transformar la variable ocupación por la palabra sin dato y también el estado civil sin dato, esto para no borrar la información y tener cuales y cuantos son los registros que no tendremos la información de estas variables luego el resto si borro el registro el cual se encuentra sin registro se realiza una separación de variables numéricas y categóricas para el tratamiento de datos y análisis para determinar la correlación existente entre una variable y otra realizando métricas de estadística descriptiva .

Análisis Exploratorio de Datos (EDA): Para el análisis de los datos se realiza un estudio de cada uno en cuanto a relación medición graficas correlaciones persons etc. para determinar el estudio de los datos para poder aplicar para el cumplimiento de nuestro objetivo. con las siguientes variables

Sociales: régimen de seguridad social pertenecía étnica.

Demográficas: departamento, área de residencia, sexo.

Clínicas: causa básica de defunción, causas múltiples, manera de muerte.

Temporales: año y mes de ocurrencia.

Este tamaño muestral permite realizar análisis estadísticos robustos y aplicar técnicas de ciencia de datos para identificar patrones de mortalidad y factores asociados, garantizando la validez de los resultados (Han, Kamber & Pei, 2012)

Se realizó un estudio observacional retrospectivo, basado en datos secundarios. El objetivo fue identificar patrones de mortalidad y factores asociados a la insuficiencia renal crónica (código N18 según la Clasificación Internacional de Enfermedades – CIE-10). Este tipo de diseño es adecuado para explorar tendencias poblacionales y establecer relaciones entre variables (Hernández-Sampieri & Mendoza, 2018).

Modelar los factores asociados a morir por insuficiencia renal crónica (IRC N18) frente a otras causas

## Figura 1

*Formula Correlaciones De Pearson's*

---

Fórmula >

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

$r$  = coeficiente de correlación  
 $x_i$  = valores de la variable x en una muestra  
 $\bar{x}$  = media de los valores de la variable x  
 $y_i$  = valores de la variable y en una muestra  
 $\bar{y}$  = media de los valores de la variable y

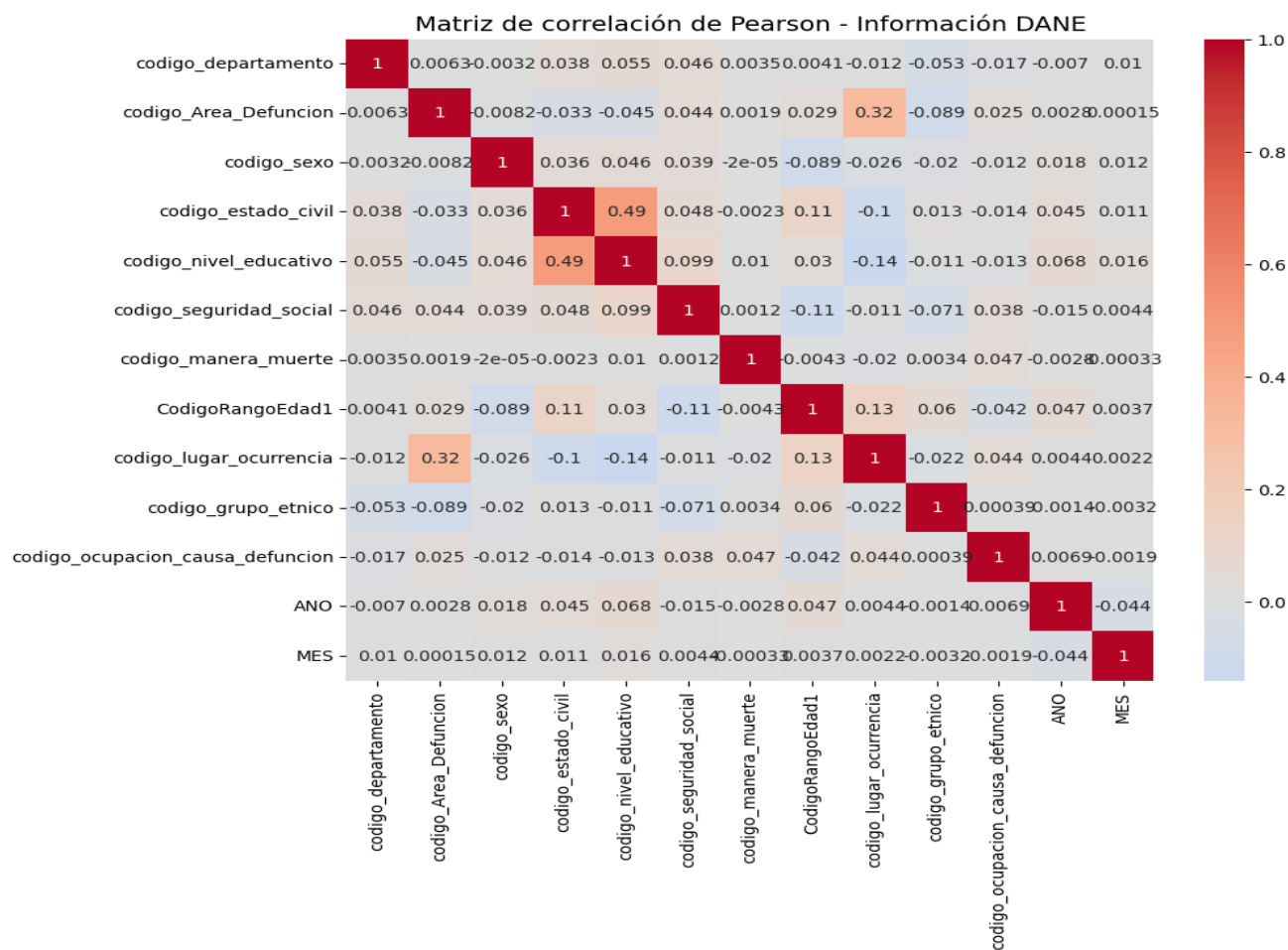
---

*Nota.* Tomado de Enciclopedia Britannica. (2026, febrero 13). *Pearson's correlation coefficient.*

Britannica. <https://www.britannica.com/topic/Pearsons-correlation-coefficient>

Figura 2

Diagrama de Matriz de Correlación Pearson's



Característica de los datos de tipo cuantificables al realizar el análisis de correlación person's, se observa la matriz muestra asociaciones débiles entre la mayoría de las variables analizadas, lo cual es esperado dado que se trata principalmente de variables categóricas codificadas numéricamente. Se observa una asociación moderada entre el estado civil y el nivel educativo, así como entre el área de defunción y el lugar de ocurrencia, lo cual es coherente desde el punto de vista sociodemográfico. No se identifican problemas de multicolinealidad severa que limiten el uso conjunto de las variables en modelos de machine learning.

**Tabla 1***Defunciones por Año de Pacientes con Insuficiencia Renal Crónica*

| Año  | Cuenta |
|------|--------|
| 2019 | 9174   |
| 2020 | 12593  |
| 2021 | 14118  |
| 2022 | 12857  |
| 2023 | 12598  |

*Nota.* La Tabla 1 muestra la información de una tendencia creciente fuerte entre 2019 y 2021 el cual coincide con la pandemia, seguida de una disminución leve y estabilización entre 2022 y 2023 se tiene una normalidad en el efecto del año es global, no específico de una sola característica sociodemográfica.

**Tabla 2***Defunciones Por Año de Pacientes Con Insuficiencia Renal Crónica*

| Descripción Sexo | Cuenta |
|------------------|--------|
| Femenino         | 22604  |
| Masculino        | 38736  |

*Nota.* La Tabla 2 muestra la información con un porcentaje de Masculino: 63,2 % Femenino: 36,8 % Esta diferencia evidencia un patrón consistente en el conjunto de datos, el cual la mortalidad y la enfermedad es más común en hombres que en mujeres. No obstante, el análisis de correlación indica que el sexo, de manera aislada, no presenta relaciones lineales fuertes con las demás variables analizadas.

**Tabla 3***Defunciones Nivel Educativo de Pacientes con Insuficiencia Renal Crónica*

| Descripción Nivel Educativo | Cuenta |
|-----------------------------|--------|
| Básica primaria             | 23206  |
| Sin información             | 18160  |
| Básica secundaria           | 6407   |
| Ninguno                     | 4872   |
| Media académica o clásica   | 3088   |
| Profesional                 | 2588   |
| Preescolar                  | 882    |
| Técnica profesional         | 728    |
| Media técnica               | 561    |
| Tecnológica                 | 399    |
| Especialización             | 234    |
| Normalista                  | 110    |
| Maestría                    | 81     |
| Doctorado                   | 24     |

*Nota.* Más del 75 % de los registros se concentran en educación básica, ausencia de escolaridad o sin información se observa que más de la mitad de la población presenta niveles educativos bajos o ausencia de escolaridad, mientras que la educación superior y de posgrado representa una proporción reducida del total de casos. Este comportamiento pertenece a que la mortalidad es en pacientes con edad avanzada los cuales no se tenía cubrimiento educativo en el país. La elevada proporción de registros sin información se reconoce como una limitación del estudio

**Tabla 4***Número de Fallecidos de Insuficiencia Renal Crónica Según el Departamento*

| Nombre Departamento | Cuenta |
|---------------------|--------|
| Bogotá              | 10244  |
| Valle del Cauca     | 7131   |
| Antioquia           | 6005   |
| Atlántico           | 4350   |
| Santander           | 3528   |
| Cordoba             | 2710   |
| Bolívar             | 2654   |
| Cundinamarca        | 2482   |
| Norte de Santander  | 2300   |

*Nota.* El número de Los cinco primeros concentran cerca del 51 % del total nacional, lo que evidencia una fuerte concentración geográfica lo cual es lógico por la densidad poblacional en cada uno de los departamentos; el análisis por departamento evidencia una marcada concentración de registros en las principales entidades territoriales del país, destacándose Bogotá, Valle del Cauca y Antioquia, las cuales concentran más de un tercio del total de los casos.

**Tabla 5***Número de Fallecidos con la Descripción del Tipo de Afiliación de Seguridad Social*

| Descripción Seguridad Social | Cuenta |
|------------------------------|--------|
| Subsidiado                   | 33398  |

| Descripción Seguridad Social | Cuenta |
|------------------------------|--------|
| Contributivo                 | 24203  |
| Excepción                    | 2697   |
| No asegurado                 | 875    |
| Especial                     | 155    |
| Sin información              | 12     |

*Nota.* Tabla 5 El número de causas de muerte de los pacientes es por infartos al miocardio por infecciones y por la diabetes mellitus los cuales son causas principales de la enfermedad y causas de muerte.

### **Tabla 6**

*Descripción de Numero de Causas Básicas de la Defunción (c bas1)*

| Descripción   | Cuenta |
|---|--------|
| Enfermedad Renal Cronica                            | 8627   |
| Infarto Agudo Del Miocardio                         | 6638   |
| Uso De Emergencia De U07                            | 5962   |
| Enfermedad Renal Hipertensiva                       | 5584   |
| Diabetes Mellitus No Especificada                   | 4004   |
| Diabetes Mellitus No Insulinodependiente            | 3422   |
| Enfermedad Cardiorrenal Hipertensiva                | 2583   |
| Otras Enfermedades Pulmonares Obstructivas Cronicas | 2171   |
| Enfermedad Isquemica Cronica Del Corazon            | 1416   |
| Neumonia, Organismo No Especificado                 | 669    |

*Nota.* El número de causas de muerte de los pacientes es por infartos al miocardio por infecciones y por la diabetes mellitus los cuales son causas principales de la enfermedad y causas de muerte.

**Tabla 7**

*Defunciones Según el Área o Zona con la Descripción de Seguridad Social*

| Descripción      | Descripción Area Defunción                              | Cuenta |
|------------------|---|--------|
| Seguridad Social |   |        |
| Subsidiado       | Cabecera Municipal                                      | 31691  |
| Contributivo     | Cabecera Municipal                                      | 23831  |
| Excepción        | Cabecera Municipal                                      | 2671   |
| Subsidiado       | Rural Disperso  | 1185   |
| No Asegurado     | Cabecera Municipal                                      | 859    |
| Subsidiado       | Centro Poblado (Inspección,<br>Corregimiento O Caserío) | 520    |
| Contributivo     | Rural Disperso  | 268    |
| Especial         | Cabecera Municipal                                      | 152    |
| Contributivo     | Centro Poblado (Inspección,<br>Corregimiento O Caserío) | 104    |

*Nota.* En este caso el número con mayor defunción son los de seguridad social corresponde al régimen subsidiado la cabecera municipal lleva el 80% de la población y es lógico ya que concentra la mayoría de población y de tratamiento para la enfermedad.

**Tabla 8***Defunciones con la Enfermedad de Base Frente al Genero*

| Descripción<br>Sexo | Descripción Diagnóstico Base                           | Cuenta |
|---------------------|--|--------|
| Masculino           | Enfermedad Renal Crónica                               | 5564   |
| Masculino           | Infarto Agudo De Miocardio                             | 4447   |
| Masculino           | Uso De Emergencia De U07                               | 4284   |
| Masculino           | Enfermedad Renal Hipertensiva                          | 3492   |
| Femenino            | Enfermedad Renal Crónica                               | 3063   |
| Masculino           | Diabetes Mellitus No Especificada                      | 2309   |
| Femenino            | Infarto Agudo De Miocardio                             | 2191   |
| Femenino            | Enfermedad Renal Hipertensiva                          | 2092   |
| Masculino           | Diabetes Mellitus No Insulinodependiente               | 1967   |
| Femenino            | Diabetes Mellitus No Especificada                      | 1695   |
| Femenino            | Uso De Emergencia De U07                               | 1678   |
| Masculino           | Enfermedad Cardiorrenal Hipertensiva                   | 1642   |
| Masculino           | Otras Enfermedades Pulmonares Obstructivas<br>Crónicas | 1470   |

*Nota.* El infarto del miocardio entonces los resultados muestran una sobrecarga de mortalidad masculina asociada principalmente a enfermedades crónicas no transmisibles, reforzando la necesidad de intervenciones preventivas diferenciadas por sexo, diagnóstico temprano y control de factores de riesgo como hipertensión y diabetes.

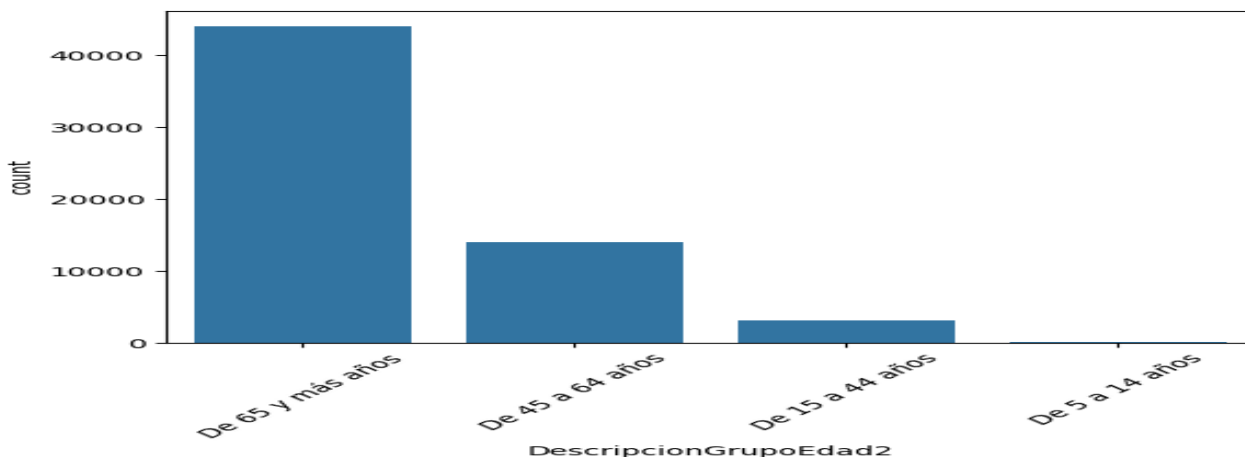
**Tabla 9***Cramér's V Mide la Fuerza de la Asociación Entre Variables Categóricas*

| Variable 1              | Variable 2              | Cramér's V | P-Valor |
|-------------------------|-------------------------|------------|---------|
| Código Area Defunción   | Código Lugar Ocurrencia | 0.19       | 0       |
| Código Area Defunción   | Código Sexo             | 0.00       | 0.18    |
| Código Area Defunción   | Código Nivel Educativo  | 0.06       | 1,21    |
| Código Area Defunción   | Código Estado Civil     | 0.03       | 1,26    |
| Código Lugar Ocurrencia | Código Sexo             | 0.02       | 1948    |
| Código Lugar Ocurrencia | Código Nivel Educativo  | 0.08       | 0       |
| Código Lugar Ocurrencia | Código Estado Civil     | 0.07       | 0       |
| Código Sexo             | Código Nivel Educativo  | 0.07       | 1,65    |
| Código Sexo             | Código Estado Civil     | 0.17       | 0       |
| Código Nivel Educativo  | Código Estado Civil     | 0.28       | 0       |

*Nota.* Los resultados evidencian que, si bien muchas asociaciones resultan estadísticamente significativas, la mayoría presenta intensidades débiles, lo cual es consistente con análisis realizados sobre los datos. Las asociaciones de nivel educativo y estado civil, y entre sexo y estado civil, podrían tener implicaciones sociales y demográficas relevantes, pero no para la mortalidad en la que estamos realizando el análisis ya que no se ve afectado ante las codificaciones de la base de la enfermedad. La asociación está por debajo de 0.2 por lo cual no se encuentra muy baja la asociación de las variables y p-valor detecta “diferencias”, pero la fuerza de asociación es prácticamente nula.

**Figura 3**

*Defunciones Según el Rango de Edad de los Pacientes con Insuficiencia Renal Crónica*

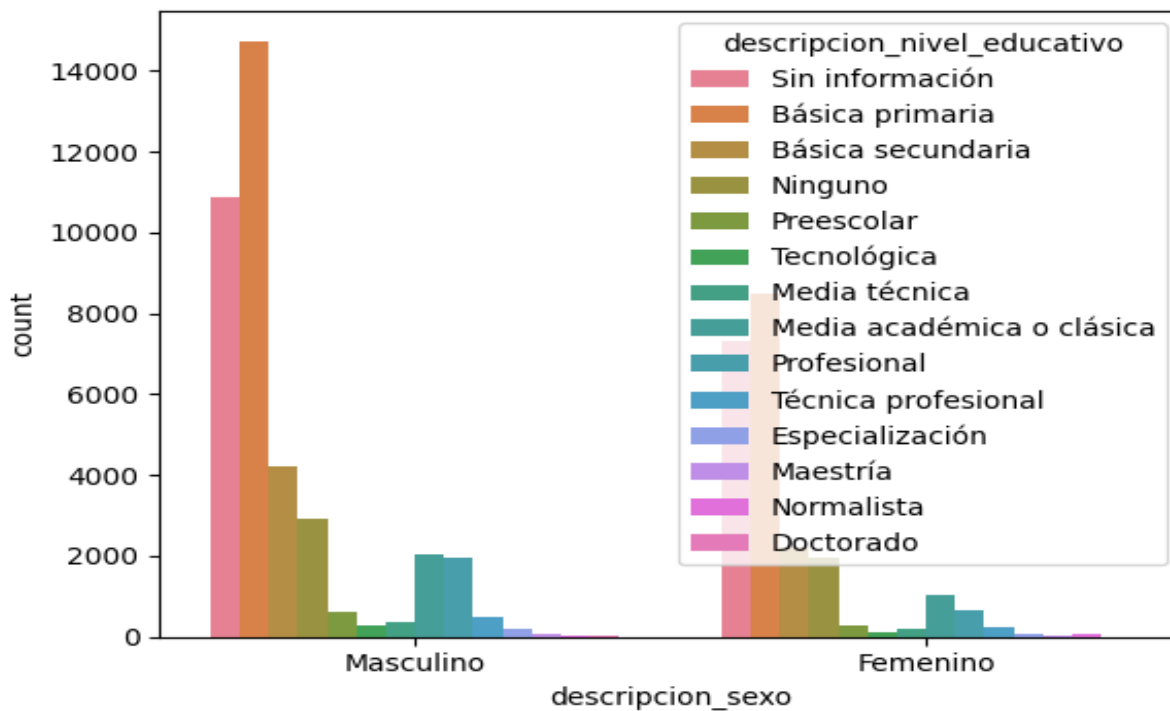


La mayoría defunciones son después de los 65 años para pacientes con insuficiencia renal crónica a población analizada presenta un perfil etario predominantemente envejecido, lo cual tiene implicaciones relevantes para la interpretación de resultados y la formulación de conclusiones, especialmente en este estudio asociados a mortalidad.

Se aplicó la prueba de chi-cuadrado de independencia para evaluar la asociación entre el diagnóstico de base del paciente (C BAS1) y los diagnósticos múltiples registrados (CAUSA MULT). Los resultados muestran un estadístico  $\chi^2 = 75,687,069.4$  con un p-valor  $< 0.001$ , lo que permite rechazar la hipótesis nula de independencia y concluir que existe una asociación estadísticamente significativa entre ambas variables

**Figura 4**

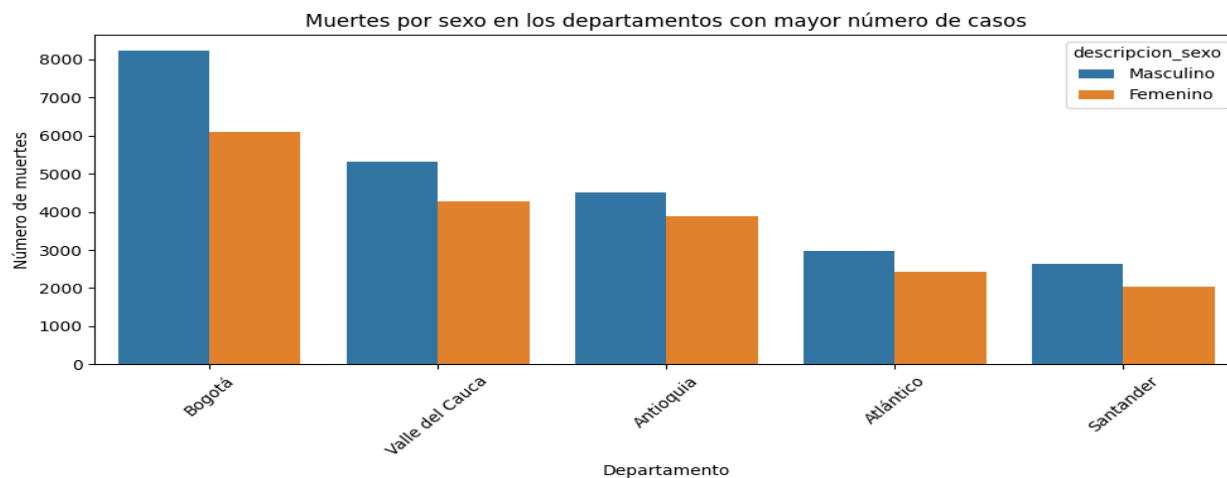
*Porcentaje de Muertes Relacionadas irc Según el Sexo con el Nivel Educativo*



En el gráfico se muestra el porcentaje más alto de muertes asociadas a la irc en hombres que en mujeres con información la gráfica muestra la distribución del nivel educativo de la población según el sexo, diferenciando entre hombres y mujeres. En términos generales, se observa que la mayor concentración de individuos, en ambos sexos, se encuentra en los niveles de educación básica, particularmente en básica primaria y básica secundaria, así como en la categoría “sin información”, lo que indica una alta proporción de registros con escolaridad básica o con ausencia de reporte del nivel educativo

**Figura 5**

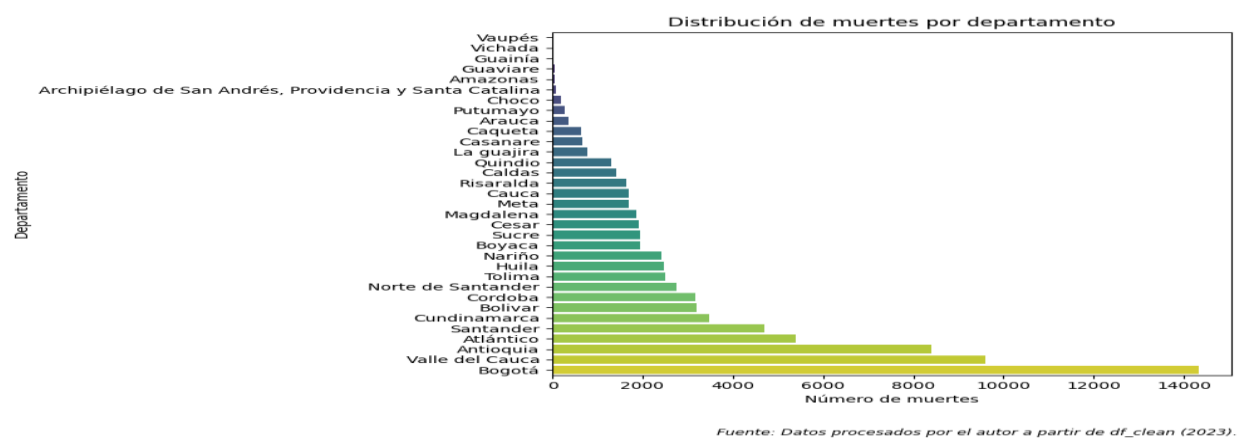
*Tasas de los 5 con Mayor Población Fallecida por Municipio y Sexo*



En Bogotá se muestra que se tiene la mayor población de muerte siendo la población masculina.

**Figura 6**

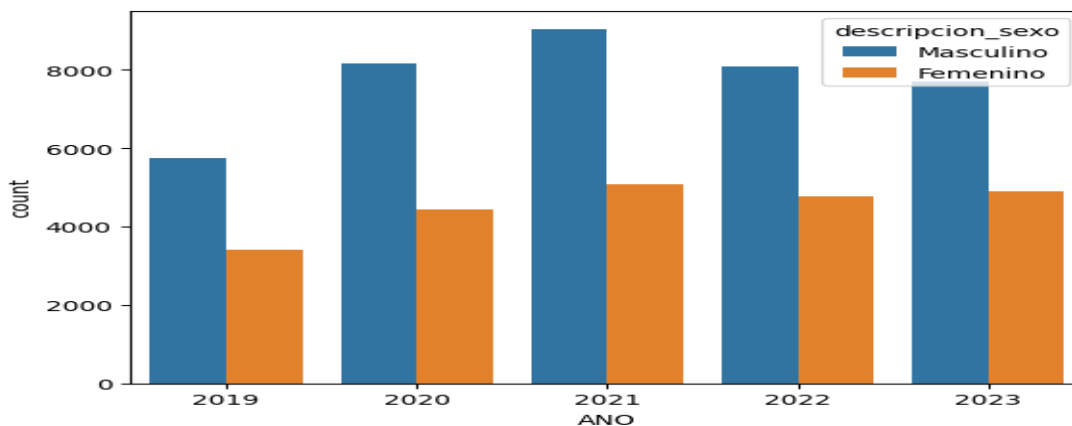
*Tasas Población Fallecida por Municipio*



Según la gráfica el mayor número de muertes por municipio se centra en Bogotá y valle del cauca

**Figura 7**

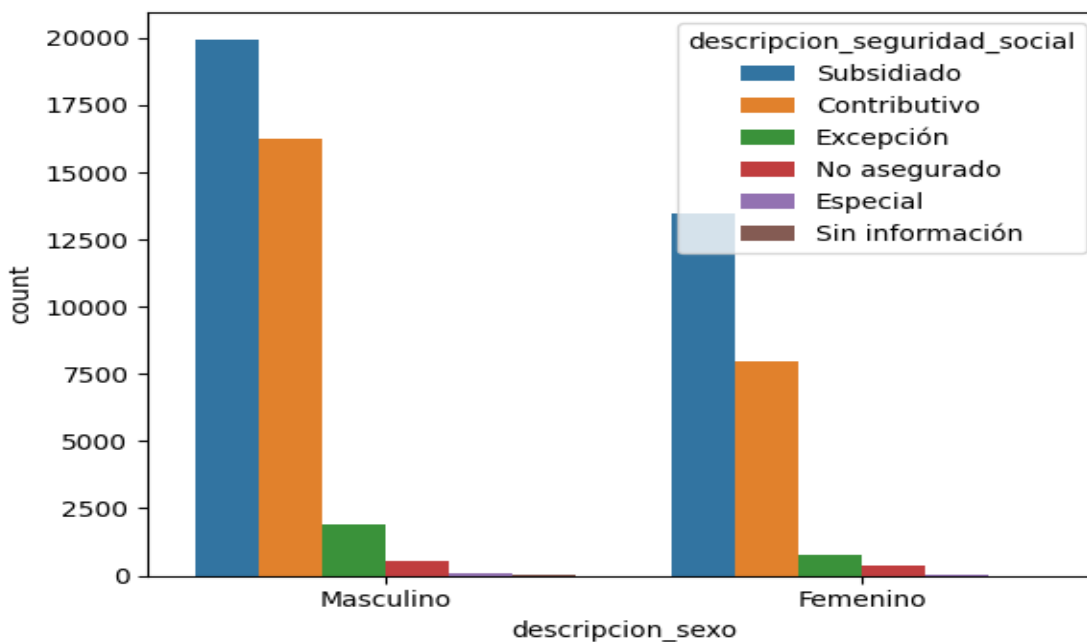
*Número de Muertes por Año y Género*



En el grafico se muestra que la tendencia de fallecimiento por año siempre prevalece el hombre por año en donde aumento en los años 2020 y 2021 siempre fue mayor el fallecimiento del género masculino.

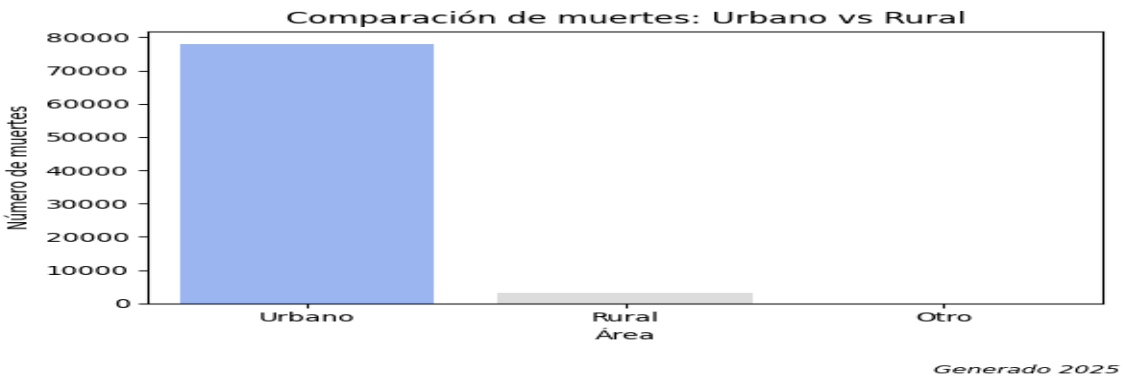
**Figura 8**

*Número de Muertes con Tipo de Afiliación en Salud Frente al Sexo*



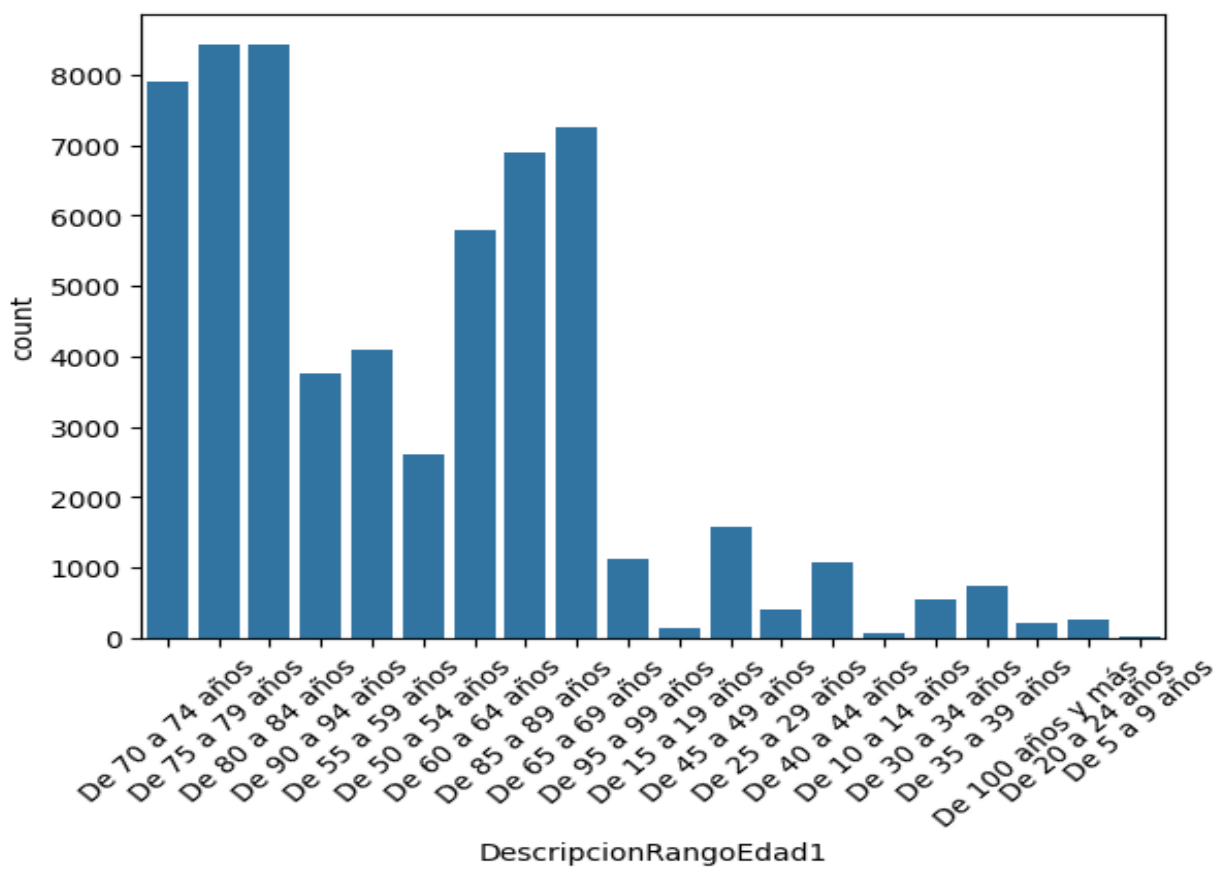
**Figura 9**

*Número de Muertes Comparadas en lo Urbano y Rural*



**Figura 10**

*Número de Muertes por Grupo de Edad Según el Rango*



En la figura encontramos que entre los 70 y 74 es el rango mayor de pacientes de fallecimiento esto para dar una diferencia a la edad con un rango mas corto para el análisis en el cual se encuentra que mayores de 75 años en donde entonces clínicamente el riñón puede empezar a fallar

Encuentro una relación categórica en la CAUSA MULT (descripción de causa múltiple para causas antecedentes) con C BAS1 (descripción de la causa Básica de la defunción) obteniendo resultados:

$$\chi^2 = 75,687,069.4$$

$$p\text{-valor} < 0.001$$

$$\text{Cramér's } V = 0.420$$

La magnitud de esta asociación fue evaluada mediante el coeficiente V de Cramér, obteniéndose un valor de 0.420, lo cual indica una asociación fuerte entre el diagnóstico de base y los diagnósticos múltiples del paciente. Este resultado sugiere que la presencia de determinados diagnósticos múltiples está estrechamente relacionada con el diagnóstico principal registrado.

Se identificó una asociación fuerte y estadísticamente significativa entre el diagnóstico de base y los diagnósticos múltiples del paciente (V de Cramér = 0.420), lo que evidencia un patrón consistente de comorbilidad clínica

### **Consideraciones Éticas**

El estudio se basó en datos públicos anonimizados, sin información personal identificable, por lo que no requirió consentimiento informado. Se respetaron los principios de confidencialidad y uso responsable de la información, siguiendo las recomendaciones de la (Organización Panamericana de la Salud [OPS], 2014) sobre el manejo de datos en salud pública

## Conclusiones

El análisis de mortalidad por insuficiencia renal crónica (IRC) en Colombia entre 2019 y 2023, basado en los registros del DANE y la Encuesta Longitudinal de Protección Social (ELPS), permitió identificar patrones demográficos, sociales y clínicos asociados a esta enfermedad. El procesamiento y depuración de más de 1.9 millones de registros, reducidos finalmente a 81.019 casos vinculados al código N18 (IRC), garantizó la calidad y pertinencia del análisis.

Incremento significativo de la mortalidad en 2021: Los hallazgos muestran un aumento notable de defunciones en 2021, lo cual coincide con el periodo posterior al pico de la pandemia por COVID-19. Esto sugiere una posible interacción entre la infección por SARS-CoV-2 y el agravamiento de enfermedades renales crónicas, así como una disminución en el acceso a servicios médicos durante la emergencia sanitaria.

Predominio de mortalidad en hombres: Los datos revelan que la mortalidad asociada a IRC es mayor en hombres que en mujeres. Este comportamiento sugiere una mayor vulnerabilidad del sexo masculino frente a las comorbilidades asociadas, o posibles diferencias en acceso, adherencia a tratamientos y hábitos de salud.

Mayor concentración de muertes en grandes centros urbanos: Bogotá, Valle del Cauca y Antioquia registran el mayor número de defunciones. Esto se relaciona tanto con el tamaño poblacional como con la mayor concentración de servicios de salud, lo que facilita el registro y la confirmación diagnóstica. También puede reflejar un mayor envejecimiento poblacional y prevalencia de enfermedades crónicas en estas regiones.

Comorbilidades cardiovasculares, infecciosas y metabólicas como principales causas asociadas: Las causas más comunes adicionales a la IRC fueron el infarto agudo de miocardio,

infecciones bacterianas resistentes, complicaciones hipertensivas y diabetes mellitus. Esto confirma que la IRC está íntimamente ligada a patologías cardiovasculares, metabólicas e infecciosas, lo cual resalta la necesidad de un abordaje integral de salud pública.

Débil asociación estadística entre variables sociodemográficas y la mortalidad por IRC: Los análisis de Cramér's V mostraron valores menores a 0.2 en todos los pares de variables evaluadas, lo que indica asociaciones muy débiles. Aunque existen diferencias significativas según la prueba chi-cuadrado, su fuerza práctica es limitada. Esto sugiere que la mortalidad por IRC no está explicada principalmente por factores como sexo, área, régimen de seguridad social o departamento, sino por variables clínicas y comorbilidades no presentes en los registros.

El régimen de seguridad social y el área muestran diferencias menores pero relevantes. Aunque la fuerza de asociación es baja, se identificó un mayor número de muertes en personas afiliadas al régimen de excepción en áreas urbanas y en el régimen subsidiado en zonas rurales. Esto podría relacionarse con diferencias en acceso, continuidad y calidad en la atención médica.

Limitada correlación entre variables numéricas: Las correlaciones de Pearson no mostraron relaciones fuertes entre los datos cuantitativos (año, mes, códigos numéricos). Esto indicó que no existía un comportamiento numérico que permitiera un modelamiento predictivo robusto con estas características, reforzando la idea de que la mortalidad por IRC responde a factores complejos y multifactoriales.

Importancia de los datos públicos y la analítica para decisiones en salud pública

Este estudio demuestra el valor de los datos anonimizados del DANE y la ELPS para comprender tendencias poblacionales. A través de técnicas de ciencia de datos, fue posible identificar patrones útiles para orientar políticas públicas en prevención de enfermedades renales, manejo de comorbilidades y fortalecimiento del sistema de salud.

### Referencias Bibliográficas

- Abat, S., Abd Rahman, R., Abdul Cader, R., Abdul Hafidz, M. I., Abdul Wahab, M. Z.,  
 Abdullah, N. K., Abdul-Samad, T., Abe, M., Abraham, N., Acheampong, S., Achiri, P.,  
 Acosta, J. A., Adeleke, A., Adell, V., Adewuyi-Dalton, R., Adnan, N., Africano, A.,  
 Agharazii, M., Aguilar, F., ... Massey, D. (2024). *Impact of primary kidney disease on  
 the effects of empagliflozin in patients with chronic kidney disease: Secondary analyses  
 of the EMPA-KIDNEY trial. The Lancet Diabetes & Endocrinology, 12(1).*  
[https://www.thelancet.com/journals/landia/article/PIIS2213-8587\(23\)00322-4/fulltext](https://www.thelancet.com/journals/landia/article/PIIS2213-8587(23)00322-4/fulltext)
- Casas Roma, J., Lozano Bagén, T., & Bosch Rué, A. (2019). *Deep learning: Principios y  
 fundamentos* (p. 24). Editorial UOC.
- Cibulka, R., & Racek, J. (2007). Metabolic disorders in patients with chronic kidney failure.  
*Physiological Research, 56(6), 697–705.*  
[https://www.biomed.cas.cz/physiolres/pdf/56/56\\_697.pdf](https://www.biomed.cas.cz/physiolres/pdf/56/56_697.pdf)
- Contributors to Wikimedia Projects. (2003). *Machine learning*. En *Wikipedia, the free  
 encyclopedia*. Recuperado el 26 de agosto de 2023 de  
[https://en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning)
- Departamento Administrativo Nacional de Estadística [DANE]. (2017). *Salud. Catálogo de  
 microdatos* [Base de datos]. DANE.  
<https://microdatos.dane.gov.co/index.php/catalog/SAL-Microdatos>
- Ehmann, M. R., Klein, E. Y., Zhao, X., Mitchell, J., Menez, S., Smith, A., Levin, S., & Hinson,  
 J. S. (2024). Epidemiology and clinical outcomes of community-acquired acute kidney  
 injury in the emergency department: A multisite retrospective cohort study. *American*

*Journal of Kidney Diseases*, 83(6). [https://www.ajkd.org/article/S0272-6386\(23\)00945-9/abstract](https://www.ajkd.org/article/S0272-6386(23)00945-9/abstract)

Fragoulis, C., Polyzos, D., Mavroudis, A., Tsioufis, P. A., Kasiakogias, A., Leontsinis, I., Mantzouranis, E., Kalos, T., Sakalidis, A., Ntalakouras, J., Andrikou, I., Dimitriadis, K., Konstantinidis, D., Thomopoulos, C., & Tsioufis, K. (2024). One-year outcomes following a hypertensive urgency or emergency. *European Journal of Internal Medicine*, 120, 1–7. <https://doi.org/10.1016/j.ejim.2023.10.020>

Galea, A. (2018). *Applied data science with Python and Jupyter: Use powerful industry-standard tools to unlock new, actionable insights from your data* (1st ed., pp. 22–28). Packt Publishing.

Kanda, E., Suzuki, A., Makino, M., Tsubota, H., Kanemata, S., Shirakawa, K., & Yajima, T. (2022). Machine learning models for prediction of HF and CKD development in early-stage type 2 diabetes patients. *Scientific Reports*, 12(1). <https://doi.org/10.1038/s41598-022>

Kane, F. (2017). *Hands-on data science and Python machine learning*. Packt Publishing.

Kane, F. (2017). *Hands-on data science and Python machine learning* (pp. 183–195). Packt Publishing.

Kyriakides, G., & Margaritis, K. G. (2019). *Hands-on ensemble learning with Python: Build highly optimized ensemble machine learning models using scikit-learn and Keras*. Packt Publishing.

Ponce, D., Nitsch, D., & Ikizler, T. A. (2023). Strategies to prevent infections in dialysis patients. *Seminars in Nephrology*, 43(5). [https://www.seminarsinnephrology.org/article/S0270-9295\(23\)00177-8/abstract](https://www.seminarsinnephrology.org/article/S0270-9295(23)00177-8/abstract)

Raschka, S., & Mirjalili, V. (2017). *Python machine learning* (2nd ed.). Packt Publishing.

Raschka, S., & Mirjalili, V. (2017). *Python machine learning* (2nd ed., pp. 90–103). Packt Publishing.

Raschka, S., & Mirjalili, V. (2018). *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow*. Packt Publishing.

Russell, S., & Norvig, P. (2009). *Inteligencia artificial: Un enfoque integral* (3.<sup>a</sup> ed., p. 229). Pearson Educación.

SAP. (s.f.). *¿Qué es machine learning?* Recuperado el 24 de agosto de 2023 de <https://www.sap.com/latinamerica/products/artificial-intelligence/what-is-machine-learning.html>

Serrano López, E., Rangel, E. L., Roa, D. R., Rodríguez, C., & Lozano Ortega, C. (s.f.). *Guía de anonimización de datos estructurados: Conceptos generales y propuesta metodológica*. Archivo General de la Nación Jorge Palacios Preciado – Colombia.

The Matplotlib development team. (s.f.). *Matplotlib documentation*. [https://matplotlib.org/stable/plot\\_types/index.html](https://matplotlib.org/stable/plot_types/index.html)

The NumPy development team. (s.f.). *NumPy documentation*. <https://numpy.org/doc/2.2/>

The pandas development team. (s.f.). *Pandas documentation*. <https://pandas.pydata.org/docs/>

The scikit-learn development team. (s.f.). *Scikit-learn documentation*. [https://scikit-learn.org/stable/user\\_guide.html](https://scikit-learn.org/stable/user_guide.html)

The SciPy development team. (2025, mayo 8). *SciPy documentation*. <https://docs.scipy.org/doc/scipy/>

The seaborn development team. (s.f.). *Seaborn documentation*. <https://seaborn.pydata.org/tutorial.html>

Wahida, A. Z., Rumahorbo, H., & Murtiningsih. (2023). The effectiveness of intradialytic exercise in ameliorating fatigue symptoms in patients with chronic kidney failure undergoing hemodialysis: A systematic literature review and meta-analysis. *Journal of Taibah University Medical Sciences*, 18(3).

<https://www.sciencedirect.com/science/article/pii/S1658361222002165>