# ANÁLSIS DE UBICACIÓN, DEMOGRAFÍA Y ESTILO DE VIDA EN ENFERMEDADES CARDIOVASCULARES EN BOGOTÁ

Santiago Pardo Felipe Rueda Luis Plazas Félix Duarte

23/05/2024

Inteligencia de Negocios – Proyecto 2

## 0. Resumen

El proyecto pretende utilizar los datos de las encuestas multipropósito de 2017 y 2021 para ayudar a los actores del sector salud en Bogotá a comprender y caracterizar los comportamientos de enfermedades en la población, para mejorar la calidad de vida y extender estas prácticas a otras ciudades de Colombia. Involucra la identificación de requerimientos analíticos, el desarrollo de tableros de control interactivos, y la automatización de procesos de datos, teniendo en cuenta recomendaciones previas para evitar errores. El equipo está compuesto por profesores y estudiantes de medicina, así como estudiantes de administración, que colaborarán para asegurar el éxito del proyecto a través de un enfoque integrado y multidisciplinario.

#### Contenido

1. Introducción	3
1.1 Objetivos del Proyecto	4
1.2 Contexto y Justificación	4
2. Metodología	4
2.1 Identificación de Requerimientos Analíticos	4
2.2 Entendimiento de los datos	6
2.3 Arquitectura de la solución construida y Diseño de Data Marts	6
2.4 Diseño e Implementación de ETL	11
3. Análisis Descriptivo	13
3.1 Tableros de Control	13
3.2 Análisis Realizados e interpretación de los resultados	15
4. Discusión y Conclusiones	15
4.1 Significancia Estadística	5
4.2 Significancia Clínica	15
4.3 Limitaciones del Estudio	5
5. Video	16
6. Anexos	5
6.1 Documentación Técnica	5
6.2 Material Complementario	5
7. Referencias	16

# 1. Introducción

## 1.1 Objetivos del Proyecto

El objetivo principal de este proyecto es desarrollar un sistema de inteligencia de negocios que utilice los datos de las encuestas multipropósito de 2017 y 2021 para monitorear métricas e indicadores de interés para expertos en salud, con el fin de guiar intervenciones que mejoren la calidad de vida de la población de Bogotá y, en el futuro, otras ciudades de Colombia.

### **Objetivos Específicos:**

- **Identificar requerimientos analíticos** que puedan ser resueltos con análisis descriptivos y desplegarlos en tableros de control.
- Proponer una arquitectura de solución de inteligencia de negocios que soporte los análisis necesarios.
- Desarrollar data marts que permitan resolver los requerimientos analíticos del cliente.
- **Diseñar e implementar procesos ETL** (extracción, transformación y carga de datos) para asegurar la integridad y actualidad de los datos.
- Implementar requerimientos analíticos basados en tableros de control que sean de interés para los expertos en salud.

# 1.2 Contexto y Justificación

El proyecto busca aprovechar los datos proporcionados por las encuestas multipropósito de 2017 y 2021 para proporcionar a los responsables del sector salud de Bogotá (como secretarios de salud y gerentes de EPS o IPS) una herramienta efectiva para entender y caracterizar los comportamientos de enfermedades en la población. Esto permitirá definir e implementar intervenciones que mejoren la calidad de vida de los habitantes. Además, la metodología y los resultados del proyecto tienen el potencial de ser replicados en otras ciudades y municipios de Colombia, extendiendo los beneficios a nivel nacional. La automatización de procesos y el desarrollo de tableros de control avanzados garantizarán la accesibilidad y el uso efectivo de los datos, tomando en cuenta recomendaciones previas para evitar errores comunes y mejorar la presentación de información.

# 2. Metodología

# 2.1 Identificación de Requerimientos Analíticos

En esta parte del proyecto, se identificaron las necesidades analíticas, las cuales se acordaron con los clienttes, y se encuentran documentados de la siguiente forma

Tema analítico	Análisis requeridos o inferidos	Categoría del análisis - <u>Tablero de</u> <u>control,</u> análisis OLAP, Minería de datos	Procesos de negocio	Fuentes de datos y datos
Análisis de la influencia de los factores de vivienda y estilo de vida	Evaluación del impacto de la proximidad de la vivienda a diversas infraestructuras (fábricas, terminales de buses, basureros, etc.) en la presencia de enfermedades cardiovasculares.	Tablero de control	Investigación y salud pública	Encuesta multipropósito, Datos de la vivenda y su entorno (Capítulo B), Salud (Capítulo F)
en la prevalencia de enfermedades cardiovasculares.	Evaluación del perfil demográfico (lugar de nacimiento, edad, sexo, orientación, donde vive) de las personas con y sin presencia de enfermedades cardiovasculares.		Investigación y características demográficas	Encuesta multipropósito, Composicion del hogar y demografia (Capítulo E), Salud (Capítulo F)
	Evaluación del impacto de la actividad física y el consumo de cigarrillos en la presencia de enfermedades cardiovasculares.	Tablero de control	Investigación y salud pública	Encuesta multipropósito, Salud (Capítulo F)

## 2.2 Entendimiento de los datos

Se hará el entendimiento para cuatro sets de datos pertenecientes al conjunto de microdatos del DANE en Colombia, estos datasets fueron el conjunto de datos A, que corresponde a las identificaciones de los usuarios, el conjunto de datos B, que corresponde a los datos de la vivienda y su entorno, al conjunto de datos E, que corresponde a los datos del hogar y demográficas y por último el conjunto de datos F, que corresponde a los datos de salud.

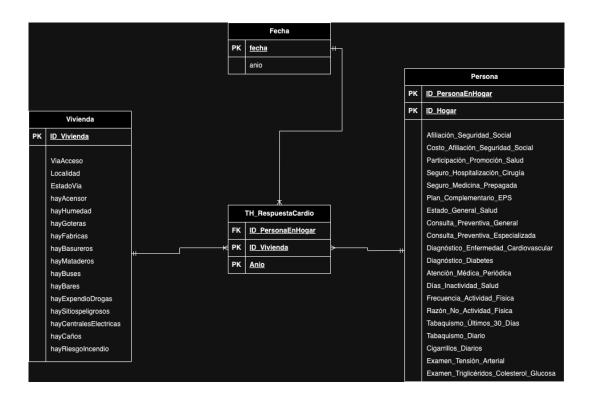
El primer conjunto de datos corresponde a las identificaciones, el cual contiene un total de 11 variables, de las cuales 5 son numéricas y 6 son categóricas, no contiene datos duplicados (haciendo el conteo de los datos de identificación) y contiene un 7.3% de datos nulos, por lo que es necesario eliminar este porcentaje de datos al no ser significativo.

El segundo conjunto de datos corresponde a los datos de vivienda y su entorno, que contiene 63 variables, de las cuales 7 son numéricas y 56 son categóricas, no contiene datos duplicados (haciendo el conteo de los datos de identificación) y un 7.3 % de datos nulos, por lo que hay que eliminar este porcentaje de datos al no ser significativo.

Basado en estos análisis, que se realizaron en el tablero de control, se sacó el siguiente análisis de los datos

# 2.3 Arquitectura de la solución construida y Diseño de Data Marts

De forma general, al tener 2 bases de datos (una de 2017 y una de 2021), se utilizó la misma estructura de modelo multidimensional para extraer las consultas de BigQuery. La cual estuvo organizada de la siguiente forma



# Dimensión 1: Vivienda

Nombre del Campo	Descripción
ID_Vivienda	Llave primaria
viaAcceso	Variable con la vía de acceso
Localidad	Variable con la localidad
EstadoVia	Estado de la vía
hayAscensor	Hay un ascensor
hayHumedad	Hay humedad en la vivienda
hayGoteras	Hay goteras en la vivienda
hayFabricas	Hay fábrias en la vivienda
hayFBasureros	Hay basureros en la vivienda
hayBuses	Hay mataderos en la vivienda
hayBares	Hay buses en la vivienda
hayExpendioDrogas	Hay expendio de drogas en la vivienda
haySitiosPeligrosos	Hay sitios peligrosos en la vivienda

hayCentralesElectricos	Hay centrales eléctricas en la vivienda
hayCanios	Hay cantos en la vivienda
hayRiesgoIncendio	Hay riesgo de incendio en la vivienda

Para determinar cuáles de los datos pueden influir en enfermedades cardiovasculares, debemos considerar factores que afectan la salud cardiovascular, tales como el ambiente físico y social, la calidad de la vivienda y los servicios disponibles. Los datos que podrían tener un impacto en las enfermedades cardiovasculares son:

- 1. NVCBP8A a NVCBP8J: Condiciones de la vivienda, como humedades, goteras, grietas, fallas en tuberías, ventilación escasa, inundaciones y peligro de derrumbe, pueden generar estrés y ansiedad, además de problemas de salud física que indirectamente afectan el corazón.
- 2. NVCBP14A a NVCBP14J: La proximidad de la vivienda a fábricas, basureros, plazas de mercado, terminales de buses, bares, expendios de droga, sitios oscuros y peligrosos, líneas de alta tensión, caños de aguas residuales y zonas de riesgo de incendio puede incrementar la exposición a contaminantes ambientales y generar estrés, ambos factores de riesgo cardiovascular.
- 3. NVCBP15A, NVCBP15C, NVCBP15D, NVCBP15E y NVCBP15H: Problemas en el sector como ruido, inseguridad, contaminación del aire, malos olores y presencia de insectos, roedores o animales pueden incrementar el estrés y la exposición a contaminantes, ambos relacionados con el riesgo cardiovascular.

Estos factores ambientales y de condiciones de vida influyen directamente en la salud y pueden aumentar el riesgo de enfermedades cardiovasculares a través de mecanismos como el estrés crónico, la exposición a contaminantes y la falta de servicios básicos adecuados.

## Dimensión 2: Fecha

Nombre del Campo	Descripción
fecha	Llave primaria
Anio	Año de los datos

Como hay dos archivos CSV de 2017 y 2021 para determinar la fecha de cada fila solo se creó un atributo llamado Anio que determina el origen de cada fila con respecto al archivo CSV.

# Dimensión 3: Persona

Persona	Descripcion
ID_PersonaEnHogar	Llave primaria
Afiliacion_Seguridad_Social	Está afiliado a la seguridad social
Costo_Afiliacion_Seguridad_Social	Costo de afiliación a seguridad social
Participacion_Promocion_Salud	Participación de promoción a la salud
Seguro_Hospitalizacion_Cirugia	Seguro de hospitalización a la cirugia
Seguro_Medicina_prepagada	Seguro de medicina prepagada
Plan_Complementario_EPS	Plan complementario de EPS
Estado_General_Salud	Estado general de la salud
Consulta_Preventiva_General	Hizo una consulta preventiva general
Consulta_Preventiva_Especializada	Hizo una consulta preventiva especializada
Diagnostico_Enfermedad_Cardiovascular	Tuvo un diagnóstico de enfermedad cardiovascular
Diagnostico_Diabetes	Tuvo un diagnóstico de diabetes
Atencion_Medica_Periodica	Tuvo atención media periódica
Dias_Inactividad_Salud	Tuvo días de inactividad de salud
Frecuencia_Actividad_Fisica	Frecuencia de actividad física
Razon_No_Actividad_Fisica	Razón de no hacer actividad física
Tabaquismo_Ultimos_30_Dias	Tabaquismo en los últimos 30 días
Cigarrillos_Diarios	Hace cigarrillos diarios
Examen_Tension_Arterial	Hace un examen de tensión arterial

## Tabla de Hechos

Para identificar qué datos tienen una relación directa o indirecta con enfermedades cardiovasculares, debemos considerar factores de acceso y calidad del servicio de salud,

diagnóstico y manejo de condiciones de salud relacionadas, hábitos de vida y comportamientos de salud. Aquí están los datos relevantes:

- 1. NPCFP1 a NPCFP5: Información sobre la afiliación a la seguridad social en salud y quién paga por la afiliación. El acceso a la seguridad social puede influir en la detección y tratamiento de enfermedades cardiovasculares.
- 2. NPCFP7: Participación en acciones de promoción de salud y prevención de enfermedades. Este dato es importante porque las campañas de prevención pueden incluir actividades relacionadas con la salud cardiovascular.
- 3. NPCFP10A a NPCFP10C: Tener planes o seguros complementarios de salud puede afectar el acceso a tratamientos y servicios de salud que previenen o manejan enfermedades cardiovasculares.
- 4. NPCFP12: Estado general de salud. Esto puede reflejar la presencia de factores de riesgo o enfermedades cardiovasculares.
- 5. NPCFP13A, NPCFP13B: Consultas preventivas a medicina general y especializada. La atención médica preventiva es crucial para la detección y manejo temprano de enfermedades cardiovasculares.
- 6. NPCFP14A, NPCFP14F: Diagnóstico de enfermedades cardiovasculares e hipertensión, y de diabetes. Ambas condiciones están directamente relacionadas con la salud cardiovascular.
- 7. NPCFP15: Atención médica periódica para tratar enfermedades. Es relevante para el manejo continuo de condiciones cardiovasculares.
- 8. NPCFP18: Días de actividad interrumpida debido a problemas de salud. La incapacidad para realizar actividades normales puede ser un indicador indirecto de enfermedades cardiovasculares.
- 9. NPCFP36: Frecuencia de práctica de deporte o actividad física. La actividad física es un factor protector contra enfermedades cardiovasculares.
- 10. NPCFP37: Razón para no practicar deporte o actividad física. La falta de actividad física es un factor de riesgo para enfermedades cardiovasculares.
- 11. NPCFP38, NPCFP39, NPCFP39A: Tabaquismo. Fumar es un factor de riesgo significativo para las enfermedades cardiovasculares.
- 12. NPCFP40D, NPCFP40F: Exámenes de toma de tensión arterial y exámenes de triglicéridos, colesterol, glucosa. Estos exámenes son cruciales para la detección de factores de riesgo cardiovasculares.

Estos datos pueden proporcionar una visión integral de los factores de riesgo, el acceso a la atención médica y los comportamientos de salud que influyen en la prevalencia y manejo de enfermedades cardiovasculares.

# Tabla de Hechos: TH RespuestaCardio

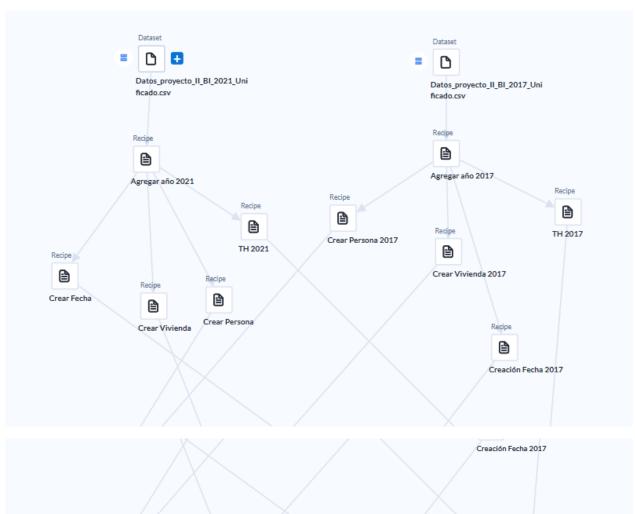
La tabla tuvo solo 3 llaves foráneas referidas a las 3 dimensiones mencionadas.

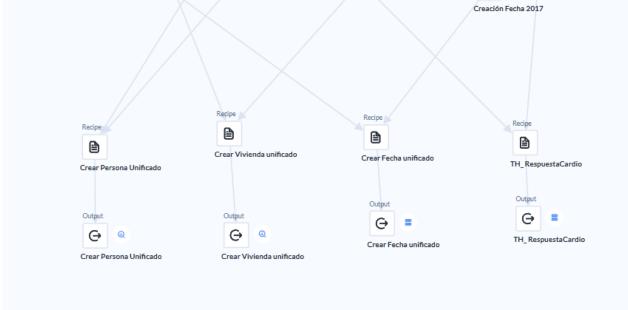
Tabla de hechos respuesta cardio	Descripción
ID_PersonaEnHogar	Llave Foreneo Persona
ID_Vivienda	Llave Foreneo Vivienda
Anio	Fecha

# **Tabla de Medidas**

Nombre de medida	Descripción
Días_inactividad_salud	Días de inactividad de salud
Costo_Afiliacion_Seguridad_Social	Costo de cada persona de afiliación en seguridad social

# 2.4 Diseño e Implementación de ETL





Para su implementación del ETL en GCP nos basamos en el Excel que ya hicimos, eliminamos los duplicados y para los nombres quitamos todos los caracteres con tilde y "ñ" para no son soportados por GCP. El archivo JSON está dentro del repositorio, se llama "Flow\_Proyecto\_2\_BI.zip".

# 3. Análisis Descriptivo

## 3.1 Tableros de Control

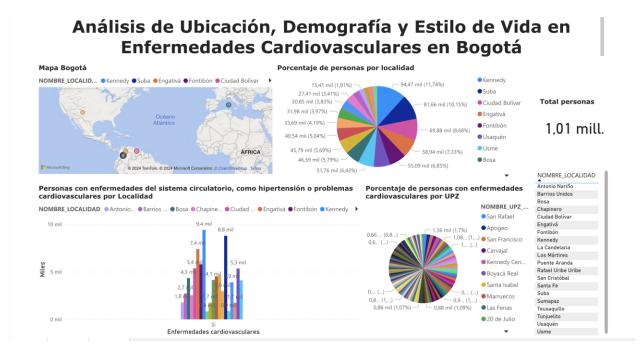
El tablero de control funcional se puede acceder mediante el siguiente link: https://uniandes-

my.sharepoint.com/personal/l\_plazasp\_uniandes\_edu\_co/\_layouts/15/onedrive.aspx?id =%2Fpersonal%2Fl%5Fplazasp%5Funiandes%5Fedu%5Fco%2FDocuments%2FProyectoBl%2FProyecto2Bl%2Epbix&parent=%2Fpersonal%2Fl%5Fplazasp%5Funiandes%5Fedu%5Fco%2FDocuments%2FProyectoBl&ga=1

Se hicieron 3 tableros, de los cuales a continuación se adjuntará captura de cada uno y se hará una breve descripción

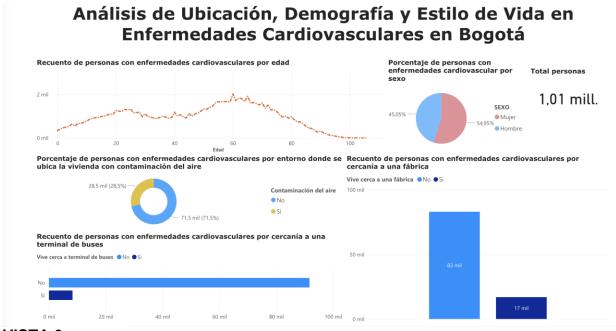
#### VISTA 1

En esta vista se puede observar un mapa de Bogotá en el cual se puso un mapa de Bogotá que redleja las localidades de todas las personas que fueron analizadas en este estudio, también hay un porcentaje de personas por cada localidad, por otro lado también se realizó para personas con enfermedades del sistema circulatorio, como hipertensión o problemas cardiovasculares por localidad y por último hay un porcentaje de personas con enfermedades cardiovasculares por UP2



#### VISTA 2:

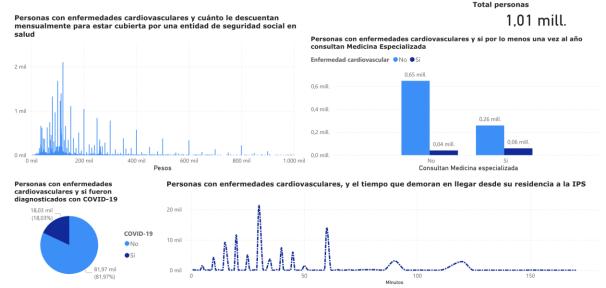
En esta vista se realizó un recuento de personas con enfermedades cardiovasculares por edad, hay un porcentaje de personas con enfermedades cardiovasculares por sexo, se sacó un porcentaje de personas con enfermedades cardiovasculares por entorno donde se ubica la vivienda con contaminación del aire, un recuento de personas con enfermedades cardiovasculares por cercanía a una fábrica y por último un recuento de personas con enfermedades cardiovasculares por cercanía a un terminal de buses



## VISTA 3:

En esta vista se intenta mirar personas con enfermedades cardiovasculares y cuánto le descuentan mensualmente para estar cubierta por una entidad de seguridad social en salud en cantidad de millones de pesos, también se intenta mirar con enfermedades cardiovasculares y si por lo menos una vez al año consultan medicina especializada en cantidad de millones de personas, también se intenta mirar personas con enfermedades cardiovasculares y si fueron diagnosticadas con COVID-19 en un gráfico de pie y por último, se observan las personas con enfermedades cardiovasculares y el tiempo que demoran en llegar desde su residencia a la IPS

# Análisis de Ubicación, Demografía y Estilo de Vida en Enfermedades Cardiovasculares en Bogotá



# 3.2 Análisis Realizados e interpretación de los resultados

Se obtuvieron las siguientes conclusiones basado en los resultados obtenidos en el tablero de control

- Kennedy es la localidad que contiene mayor cantidad de personas en el estudio, ya sea que tengan enfermedades cardiovasculaes o no
- San Rafael representa el mayor porcentaje de personas con enfermedades cardiovasculares por UP2
- Las personas de un aproximado de 60 años son aquellas más propensas a tener enfermedades cardiovasculares
- Las mujeres tienen un porcentaje mayor de riesgo de enfermedades cardiovasculares que los hombres
- Las personas con enfermedades cardiovasculares son más propensas a tener dicha enfermedad por otros factores como contaminación del aire y su cercanía a una terminal de buses
- Las personas más propensas a enfermedades cardiovasculares deben descontar un aproximado de 150000 pesos a su aseguradora
- Aquellas personas que tienen enfermedades cardiovasculares no atienden a consultas especializadas
- La gran mayoría de personas con enfermedades cardiovasculares no fueron diagnosticados con COVID-19

# 4. Discusión y Conclusiones

# 4.2 Significancia Clínica

Según (Xia, et.al, 2024) los determinantes sociales de la salud hacen parte de las condiciones del entorno donde las personas nacen, viven, aprenden, trabajan, juegan, adoran y envejecen, que afectan una amplia gama de resultados y riesgos para la salud, el funcionamiento y la calidad de vida. Se ha resaltado que los determinantes sociales de la salud son importantes para la salud cardiovascular. Esta pregunta de investigación busca destacar los determinantes sociales (ingresos, estrato y edad) de la salud a nivel individual y del área, en el cual se puedan capturar factores sociales de diferentes maneras y representar influencias distintas en los resultados cardiovasculares, permitiendo así obtener determinantes sociales específicas de cada población, como se analiza en la población de Cundinamarca

## 5. Video

El video se puede ver a través del siguiente link: (21) YouTube

## 6. Referencias

- Xia M, An J, Safford MM, et al. Cardiovascular Risk Associated with Social Determinants of Health at Individual and Area Levels. JAMA Netw Open. 2024;7(4):e248584. doi:10.1001/jamanetworkopen.2024.8
- Rosengren, A., Smyth, A., Rangarajan, S., Ramasundarahettige, C.,
- Bangdiwala, S. I., AlHabib, K. F., Avezum, A., Bengtsson Boström, K.,
- Chifamba, J., Gulec, S., Gupta, R., Igumbor, E. U., Iqbal, R., Ismail, N., Joseph,
- P., Kaur, M., Khatib, R., Kruger, I. M., Lamelas, P., ... Yusuf, S. (2019).
- Socioeconomic status and risk of cardiovascular disease in 20 low-income,
- middle-income, and high-income countries: the Prospective Urban Rural
- Epidemiologic (PURE) study. The Lancet. Global Health, 7(6), e748–e760.
- https://doi.org/10.1016/s2214-109x(19)30045-2
- Schultz, W. M., Kelli, H. M., Lisko, J. C., Varghese, T., Shen, J., Sandesara, P.,
- Quyyumi, A. A., Taylor, H. A., Gulati, M., Harold, J. G., Mieres, J. H., Ferdinand,
- K. C., Mensah, G. A., & Sperling, L. S. (2018a). Socioeconomic status and cardiovascular outcomes: Challenges and interventions. Circulation, 137(20),
- 2166–2178. https://doi.org/10.1161/circulationaha.117.029652