# Proyecto 2
# Inteligencia de negocios

Santiago Latorre 202111851 and Andrés Villota 201914885

*Departamento de Ingeniería de Sistemas y Computación, Universidad de los Andes*

2 de diciembre de 2023

## Contenido

# Identificación de necesidades analíticas

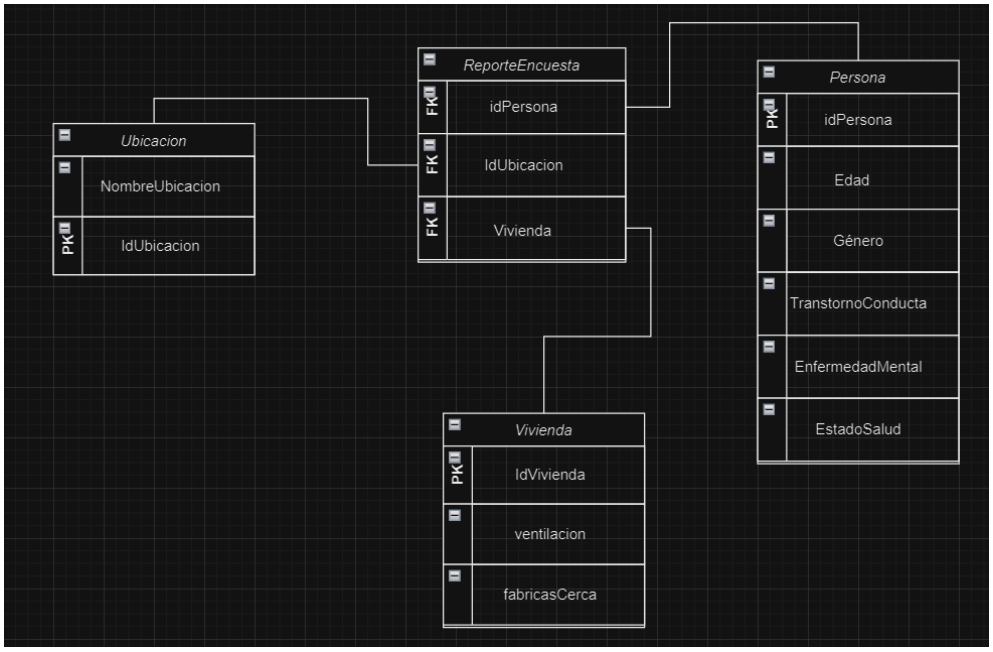
El objetivo principal del proyecto es identificar y comprender los factores que influyen en la probabilidad de que una persona desarrolle asma. A través del análisis de variables demográficas, condiciones ambientales, historial clínico, información genética y estilos de vida, el proyecto busca discernir patrones y relaciones que arrojen luz sobre los posibles factores de riesgo. Con estos resultados, se podrían formular recomendaciones para la prevención del asma, tanto a nivel individual como en el ámbito de políticas de salud pública, contribuyendo así a una mejor comprensión y gestión de esta enfermedad respiratoria.

En particular, se abordarán 4 frentes principales, todos con el fin de identificar posibles causas, factores o condiciones que puedan jugar un papel en el desarrollo de asma. En el primer enfoque, se explorará cómo factores como la exposición a alérgenos, contaminantes atmosféricos y otros elementos ambientales pueden influir en el desarrollo del asma. En cuanto a las condiciones de salud, se investigarán las conexiones entre el historial médico, y otras enfermedades respiratorias, examinando cómo estas condiciones pueden afectar la predisposición al asma. Las características demográficas, como la edad, el género y otros factores sociodemográficos, serán analizadas para comprender su impacto en la susceptibilidad al asma. Además, se evaluará cómo las condiciones geográficas, como la ubicación geográfica y el entorno local, pueden desempeñar un papel crucial en la prevalencia del asma. El objetivo general es obtener una comprensión integral de cómo estos diversos elementos interactúan y contribuyen a la probabilidad de desarrollar asma, lo que proporcionará información valiosa para la prevención y gestión efectiva de esta enfermedad respiratoria.

# Modelado de Data Marts

## Modelo Dimensional Planteado

Una vez se comprendieron los objetivos y el alcance del proyecto, se procede a construir un modelo que sea capaz de responder las consultas que pueden surgir en el contexto del ejercicio. Para lograr esto, se plantea un total de 3 dimensiones: Ubicación, Vivienda y Persona; y una tabla de hechos ReporteEncuesta que no tiene medidas. Esto debido a que, al revisar los requerimientos analíticos, notamos que no hace falta determinar medidas para encontrar resultados pertinentes.



### Justificación del Modelo

En este modelo, la tabla de hechos representa la respuesta al reporte de una única persona, es decir, hay una persona por cada fila en la tabla de hechos. En cuanto a las dimensiones, consideramos que es importante llevar un registro de los cambios que se puedan presentar, pues dada la naturaleza de los datos, no tener en cuenta el historial de la persona puede llevar a conclusiones erróneas (por ejemplo, puede que la persona haya vivido cerca a fábricas en el pasado, pero ya no lo haga). Debido a esto, y dada la distribución de los atributos, lo mejor es utilizar el Tipo II para manejo de cambios (creando una nueva fila para la misma persona, donde se registren los nuevos cambios, de esta forma se tiene la información del pasado y el presente, y así se puede realizar un análisis completo, donde se tienen en cuenta todos los factores. En un principio se evaluó utilizar mini-dimensiones para manejar estos cambios, pero no resultan útiles para las dimensiones de ubicación y vivienda, dados los pocos atributos que tienen, mientras que en la dimensión Persona casi todos los atributos son susceptibles a cambios constantes, de forma que la mini dimensión tomaría casi todos los atributos, restando la utilidad de la dimensión inicial.

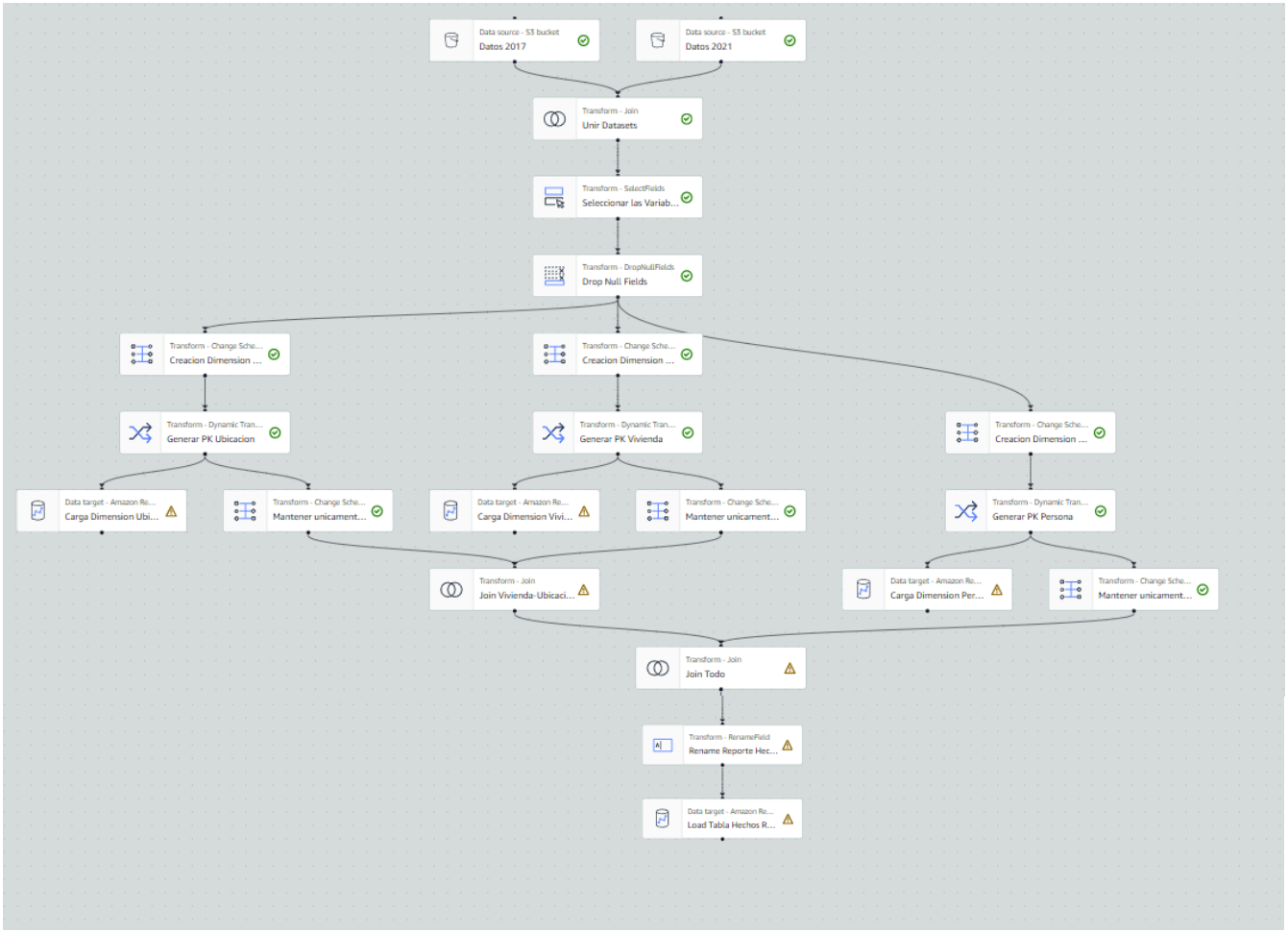
## Entendimiento de los datos y Proceso ETL

### Entendimiento de las Fuentes de Datos

Las encuestas multipropósito de 2017 y 2021 realizadas por el DANE nos dan un entendimiento de las condiciones sociales, económicas, y de salud de las personas de Bogotá y sus alrededores. Para nuestro modelo es importante conocer estos datos para poder caracterizar las personas que responden y su entorno tanto de vivienda como de condiciones físicas. Para esto hemos creado 3 tablas de dimensiones:

- Vivienda: Esta dimensión nos permite caracterizar la vivienda de la persona, principalmente con dos características, la cercanía a fábricas y la presencia de ventilación. Estas dos características son importantes para entender posibles enfermedades respiratorias como lo pueden ser el asma, ya que la cercanía a fábricas y la falta de ventilación pueden conllevar a este tipo de enfermedades.
- Ubicación: Un aspecto importante del modelo también es identificar las zonas con mayores problemas respiratorios para poder realizar proyectos e iniciativas que busquen ayudar estas zonas más afectadas. Por lo tanto, esta dimensión busca localizar la vivienda del encuestado para poder después agrupar los casos por localidades
- Persona: Una de las cosas más importantes para el proyecto es conocer las características del encuestado como lo son la edad, género y estado de salud. Además, para nuestro proyecto y nuestra pregunta de investigación es importante conocer la relación entre enfermedades respiratorias y enfermedades mentales, por lo tanto, es importante conocer si el encuestado presenta o no enfermedades mentales para luego relacionarlo con las respiratorias.

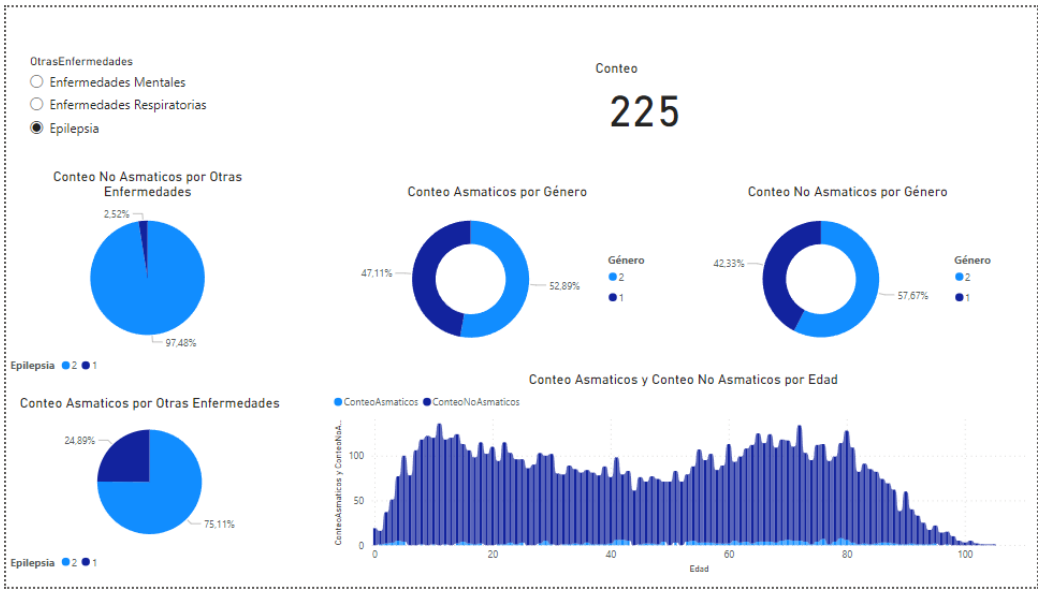
### Diseño de Proceso ETL

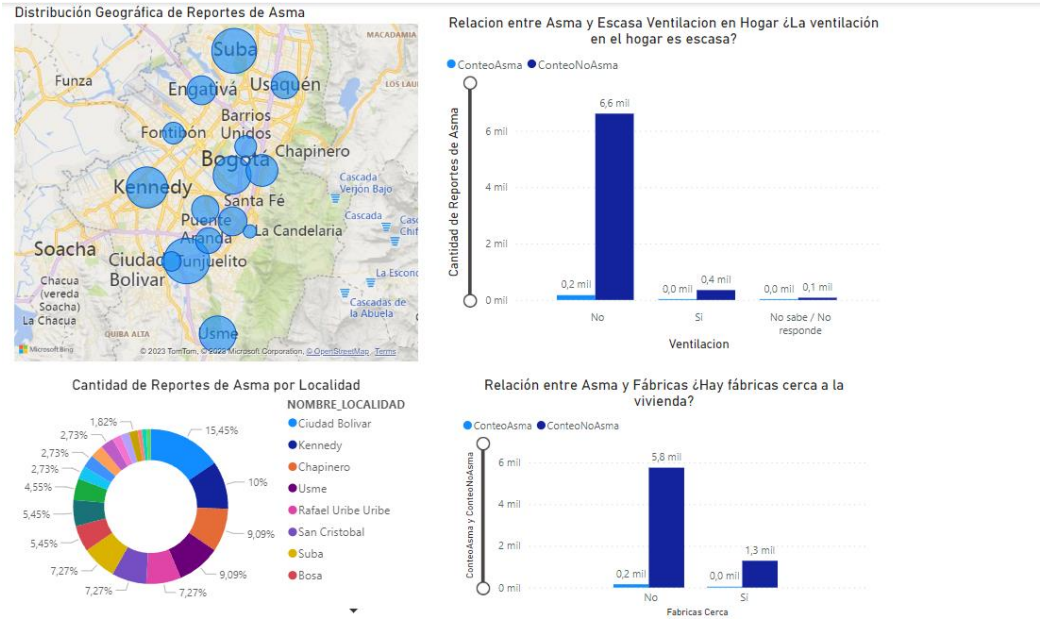



Una vez se definió el modelo y se logró una comprensión adecuada de los datos, se procede a prepararlos y transformarlos para que cumplan con lo que se definió. La descripción detallada del diseño se encuentra en el archivo Excel adjunto, pero la lógica general seguida en la construcción del ETL es juntar ambos archivos, para luego eliminar las columnas que no se utilizarán. Posteriormente se derivan 3 transformaciones, cada una para una dimensión, y finalmente se realiza otra transformación por cada dimensión, para obtener tablas con únicamente los identificadores que serán incluidos en la tabla de hechos. Finalmente se hace un join de estas tablas, y se guarda el resultado como dicha tabla de hechos, definida como ReporteEncuesta.

## Propuesta de Arquitectura de la Solución

Con el fin de satisfacer los requerimientos definidos, se plantean los siguientes tableros de control, un primero enfocado en analizar la relación entre características de salud y características demográficas de las personas, con respecto a si tienen asma o no, y un segundo enfocado en analizar la relación entre el asma y las condiciones tanto geográficas como ambientales en las que las personas viven.




## Conclusiones

- Se puede afirmar que, si existe una relación entre el asma y otras afectaciones de la salud como enfermedades respiratorias o mentales, pues en los gráficos se evidencia que la proporción de personas que sufre tanto de asma como de una de estas enfermedades es mucho mayor a aquella que solo sufre de otras enfermedades. Pese a que esto no necesariamente signifique que el asma sea un factor causante de enfermedades respiratorias o mentales, o viceversa, si se puede decir que estas enfermedades pueden ser causadas o evitadas mediante los mismos protocolos.

- Se puede afirmar que a nivel geográfico existen variaciones con respecto a la proporción de casos de asma que se presentan en toda la ciudad. Pues como se evidencia en los gráficos, zonas como Ciudad Bolívar o Kennedy presentan, con respecto a su población, un porcentaje de reportes de asma mucho más elevado que otras como Puente Aranda o Teusaquillo. Esto implica que las condiciones de vida que se presentan en las zonas difieren entre sí, y esas diferencias pueden llevar a que se presenten afectaciones en la salud, como lo es el asma.

## Evaluación Trabajo en Equipo

A nivel general, el trabajo desarrollado fue positivo, pues pese a que hubo algunas complicaciones, sobre todo en el manejo del ambiente en la nube, se logró extraer conclusiones interesantes de los datos.

Al reflexionar sobre la calidad y el aporte al proyecto entregado, estamos satisfechos con el nivel de dedicación y esfuerzo que cada miembro ha aportado. Hemos colaborado de manera eficiente, asegurándonos de que cada parte del proyecto sea de alta calidad y coherente con los objetivos establecidos.