

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	2
Modelado de Data Marts.....	2
Modelo Dimensional Planteado	2
Justificación del Modelo.....	3
Entendimiento de los datos y Proceso ETL	3
Entendimiento de las Fuentes de Datos	3
Diseño de Proceso ETL	3
Propuesta de Arquitectura de la Solución.....	4
Evaluación Trabajo en Equipo	5

Identificación de necesidades analíticas

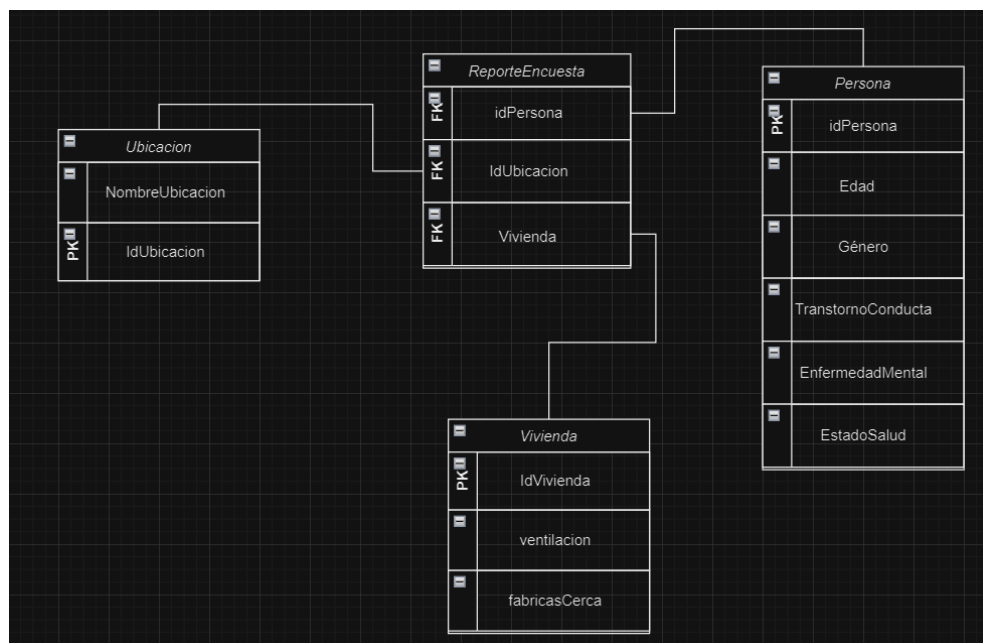
El objetivo principal del proyecto es identificar y comprender los factores que influyen en la probabilidad de que una persona desarrolle asma. A través del análisis de variables demográficas, condiciones ambientales, historial clínico, información genética y estilos de vida, el proyecto busca discernir 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.



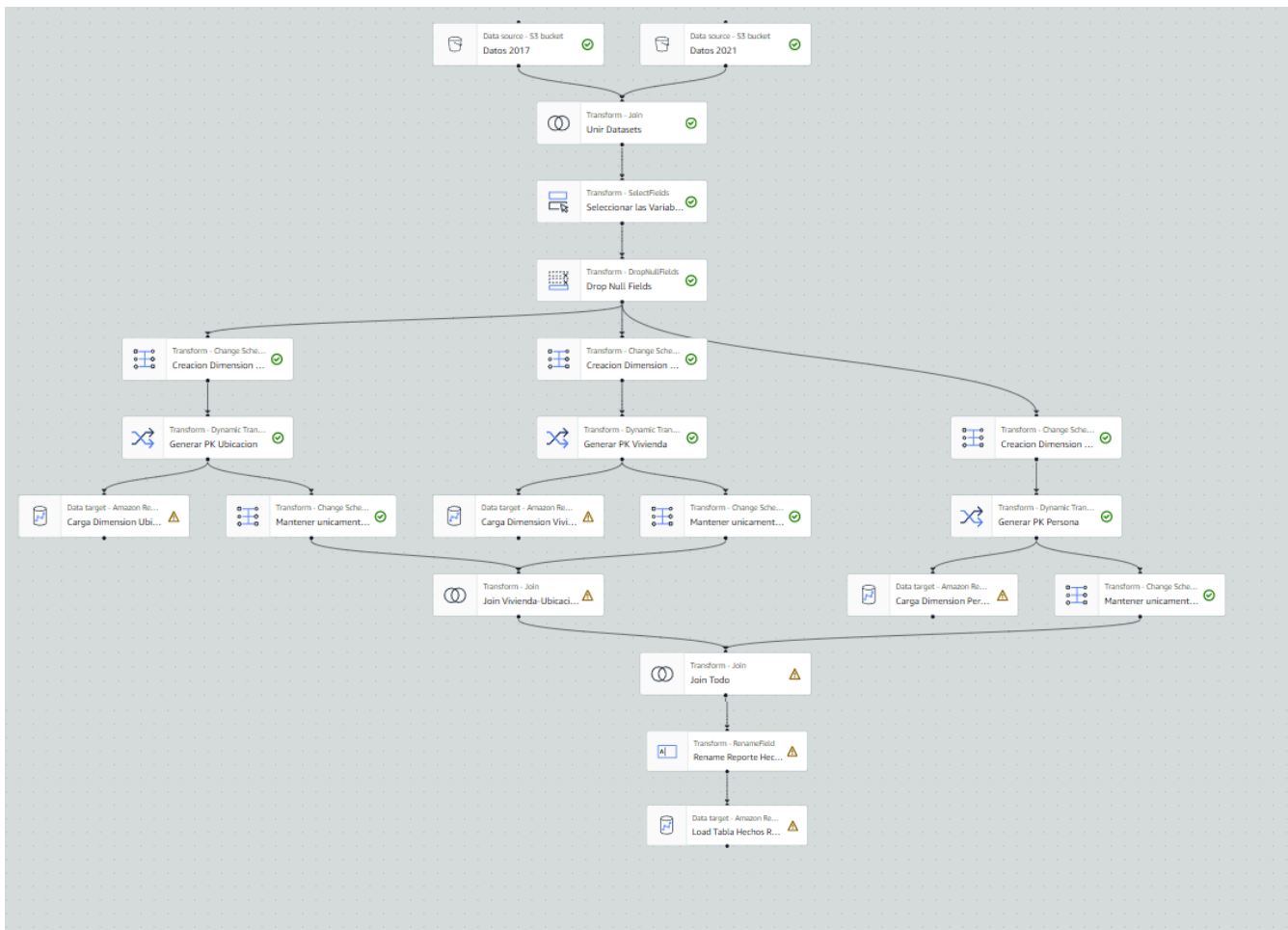
Justificación del Modelo

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

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

Evaluación Trabajo en Equipo