



Análisis de la Movilidad Humana en el Transporte Aéreo y Terrestre para la Monitoreo de la Propagación de Enfermedades Infecciosas

Luciana Julissa Huaman Coaquira

Orientador: Prof Dr./Mag./Ing. Nombre del Asesor

Plan de Tesis presentado la Escuela Profesional Ciencia de la Computación como paso previo a la elaboración de la Tesis Profesional.

**UNSA - Universidad Nacional de San Agustín de Arequipa
Junio de 2025**

1. Motivación y Contexto

Las enfermedades infecciosas representan una amenaza constante para la salud pública mundial. La pandemia de COVID-19 evidenció cómo la rápida propagación de agentes patógenos puede sobrecargar los sistemas sanitarios, impactar la economía global y alterar significativamente la vida cotidiana de las personas (World Health Organization, 2024).

Estudios recientes han demostrado que la movilidad humana, especialmente a través de redes de transporte aéreo y terrestre, es un factor crucial para la propagación de enfermedades infecciosas, ya que conecta de manera dinámica distintas poblaciones y regiones (Cao et al., 2023; Lessani et al., 2024). En particular, Oliveira et al. (2023) propusieron un modelo híbrido que combina bases de datos gráficas con datos reales de movilidad y modelos epidemiológicos para rastrear la propagación en Brasil, destacando la efectividad de esta integración para capturar dinámicas complejas de contagio.

Además, Feng et al. (2024) y Wan et al. (2024) evidenciaron la utilidad de incorporar modelos basados en redes neuronales y aprendizaje automático sobre datos de movilidad para realizar simulaciones descriptivas que permitan visualizar la evolución epidémica y detectar patrones clave. Estas investigaciones utilizan datos masivos de transporte y aplican técnicas avanzadas de visual analytics para facilitar la comprensión y toma de decisiones en tiempo real (Mason et al., 2023; Peng et al., 2025).

Sin embargo, Cao et al. (2023) y Zhang et al. (2022) señalan que existen retos importantes relacionados con la calidad y cobertura de datos, la heterogeneidad espacial y temporal, y la validación de modelos híbridos, lo cual limita su aplicación práctica. Por ello, este proyecto propone abordar estas brechas específicamente utilizando bases de datos de transporte aéreo y terrestre, desarrollando modelos metapoblacionales basados en grafos y simulaciones basadas en ecuaciones diferenciales para describir la propagación, y diseñando visualizaciones interactivas para apoyar el análisis y seguimiento epidemiológico.

2. Definición del Problema

Kang et al. (2020) explica que la propagación de enfermedades infecciosas, como el COVID-19, representa una amenaza global significativa para la salud pública y las sociedades. Ha causado millones de muertes y pérdidas económicas sustanciales, interrumpiendo gravemente los sistemas sociales y económicos. Wan et al. (2024) da entender que la previsión eficaz de epidemias es crucial para las estrategias de salud pública y la asignación eficiente de recursos médicos.

Kang et al. (2020) afirma que existe un consenso significativo en que la movilidad humana es un factor primordial en la propagación espacial de enfermedades infecciosas. Además en su trabajo se rastrea los cambios dinámicos en la movilidad y los patrones de interacción espacial llegando a la conclusión de que es un requisito previo para medir los

efectos de la movilidad y las intervenciones en la predicción de la propagación del virus

Oliveira et al. (2023) expone que los modelos como el metapoblacional son enfoques comunes utilizados para describir cómo se propaga una enfermedad en el espacio. Cao et al. (2023) explica que este modelo metapoblacional divide a la población en subpoblaciones y modela el movimiento de personas entre ellas como un gráfico dirigido. Esto permite modelar la propagación de epidemias entre regiones

En esencia, el problema radica en la necesidad de una plataforma avanzada de análisis visual y simulación que combine el poder predictivo de los modelos basados en movilidad con la claridad interpretativa de la visualización, para informar y mejorar la toma de decisiones en el control de epidemias a nivel detallado.

3. Justificación

La movilidad humana es un factor clave en la propagación espacial de enfermedades, explica López Olmedo et al. (2021). Yang et al. (2023) en su trabajo afirma que el sistema que integra datos detallados de movilidad, como trayectorias a nivel urbano, con modelos epidemiológicos espaciales puede simular la transmisión de manera más precisa al considerar la heterogeneidad espacial y temporal y las interacciones humanas complejas. Esto es particularmente valioso cuando los datos de contacto directos son limitados

Yang et al. (2023) también indica que estos sistemas permiten a los usuarios configurar y simular el impacto de diversas intervenciones, como confinamientos regionales, teletrabajo o puntos de detección. Al ajustar parámetros y observar los resultados simulados, se puede evaluar la efectividad de diferentes políticas antes de implementarlas. En su trabajo EpiMob, simula de forma interactiva los cambios en la movilidad humana y el estado de infección en respuesta a la implementación de políticas de restricción, como confinamientos regionales, teletrabajo y detección. Li et al. (2024) en su trabajo desarrollado VIVIAN permite rastrear fuentes de infección, investigar contactos cercanos y evaluar la efectividad de las políticas de prevención y control simulando su implementación mediante el ajuste de parámetros del modelo

Estos ejemplos demuestran que la integración de datos de movilidad, modelos epidemiológicos y herramientas de análisis visual es una línea de investigación activa y valorada, con sistemas que ya abordan la necesidad de simular, visualizar y evaluar la propagación de enfermedades para mejorar la toma de decisiones.

4. Contexto

Este proyecto se desarrolla en el ámbito de la ciencia de datos aplicada a salud pública, integrando epidemiología computacional, modelado basado en grafos y visual analytics. Basado en la propuesta metodológica de Oliveira et al. (2023), se aplicarán modelos meta-

poblacionales con grafos para estructurar la movilidad humana y ecuaciones diferenciales ordinarias para simular la dinámica de contagio. Se aprovecharán bases de datos aéreas y terrestres para alimentar estos modelos y se utilizarán técnicas de visualización avanzada para explorar y comunicar los resultados (Cao et al., 2023; Feng et al., 2024; Peng et al., 2025; Wan et al., 2024).

5. Objetivos

5.1. Objetivo General

Diseñar, desarrollar e implementar una herramienta de visualización interactiva que integre datos heterogéneos de movilidad humana para rastrear, analizar y comunicar la propagación espacial y temporal de enfermedades infecciosas, facilitando la toma de decisiones en salud pública.

5.2. Objetivos Específicos

- Integrar y preparar datos de movilidad humana terrestre, aérea, fluvial y urbana, para que sean compatibles con el modelado y visualización.
- Modelar la red de movilidad y flujos epidemiológicos mediante bases de datos gráficas, enfocándose en representar las relaciones y probabilidades de propagación entre localidades.
- Diseñar y desarrollar una interfaz visual interactiva que permita explorar dinámicas de propagación, incluyendo mapas de riesgo, rutas de transmisión y evolución temporal, con herramientas intuitivas para la interacción y análisis.
- Validar la herramienta visual con datos epidemiológicos reales y mediante pruebas de usuario para evaluar su eficacia en comunicar información relevante y apoyar decisiones en salud pública.

Referencias

- Cao, Q., Jiang, R., Yang, C., Fan, Z., Song, X., and Shibasaki, R. (2023). *MepoGNN: Meta-population Epidemic Forecasting with Graph Neural Networks*, pages 453–468. Springer Nature Switzerland.
- Feng, Z., Zhu, F., Wang, H., Hao, J., Yang, S., Zeng, W., and Qu, H. (2024). Holens: A visual analytics design for higher-order movement modeling and visualization.
- Kang, Y., Gao, S., Liang, Y., Li, M., Rao, J., and Kruse, J. (2020). Multiscale dynamic human mobility flow dataset in the u.s. during the covid-19 epidemic. *Scientific Data*, 7(1):390.

- Lessani, M. N., Li, Z., Jing, F., Qiao, S., Zhang, J., Olatosi, B., and Li, X. (2024). Human mobility and the infectious disease transmission: A systematic review. *Geo-spatial Information Science*, 27(6):1824–1851.
- Li, G., Chang, B., Zhao, J., Wang, J., He, F., Wang, Y., Xu, T., and Zhou, Z. (2024). Vivian: virtual simulation and visual analysis of epidemic spread data. *Journal of Visualization*, 27:677–694.
- López Olmedo, N., Stern, D., Pérez Ferrer, C., González Morales, R., Canto Osorio, F., and Barrientos Gutiérrez, T. (2021). Revisión rápida: probabilidad de contagio por infecciones respiratorias agudas en el transporte público colectivo. *Salud Pública de México*, 63(2):225–231.
- Mason, L., Hicks, B., and Almeida, J. S. (2023). Epivecs: exploring spatiotemporal epidemiological data using cluster embedding and interactive visualization. *Scientific Reports*, 13:21193.
- Oliveira, M., Alencar, A., Oliveira, N., Sales, L., Cunha, A., and Ramos, P. (2023). Epiflow: a hybrid approach to track infectious disease spread in brazil based on travel data and graph databases. In *Anais do XXXVIII Simpósio Brasileiro de Bancos de Dados*, pages 218–230, Porto Alegre, RS, Brasil. SBC.
- Peng, R., Dong, Y., Li, G., Tian, D., and Shan, G. (2025). Textlens: large language models-powered visual analytics enhancing text clustering. *Journal of Visualization*, 28:625–643.
- Wan, G., Liu, Z., Lau, M. S. Y., Prakash, B. A., and Jin, W. (2024). Epidemiology-aware neural ode with continuous disease transmission graph.
- World Health Organization (2024). Coronavirus disease (covid-19) dashboard. <https://data.who.int/dashboards/covid19/cases>.
- Yang, C., Zhang, Z., Fan, Z., Jiang, R., Chen, Q., Song, X., and Shibasaki, R. (2023). Epi-mob: Interactive visual analytics of citywide human mobility restrictions for epidemic control. *IEEE Transactions on Visualization and Computer Graphics*, 29(8):3586–3601.
- Zhang, H., Li, P., Zhang, Z., Li, W., Chen, J., Song, X., Shibasaki, R., and Yan, J. (2022). Epidemic versus economic performances of the covid-19 lockdown: A big data driven analysis. *Cities*, 120:103502.

Contexto:

Este estudio se centra en la utilización de datos de movilidad humana para comprender y monitorear la propagación de enfermedades infecciosas en Brasil. La movilidad de las personas es un factor crítico que influye en la velocidad y el alcance de los brotes epidémicos, y por ello, analizar patrones de desplazamiento es fundamental para la salud pública.

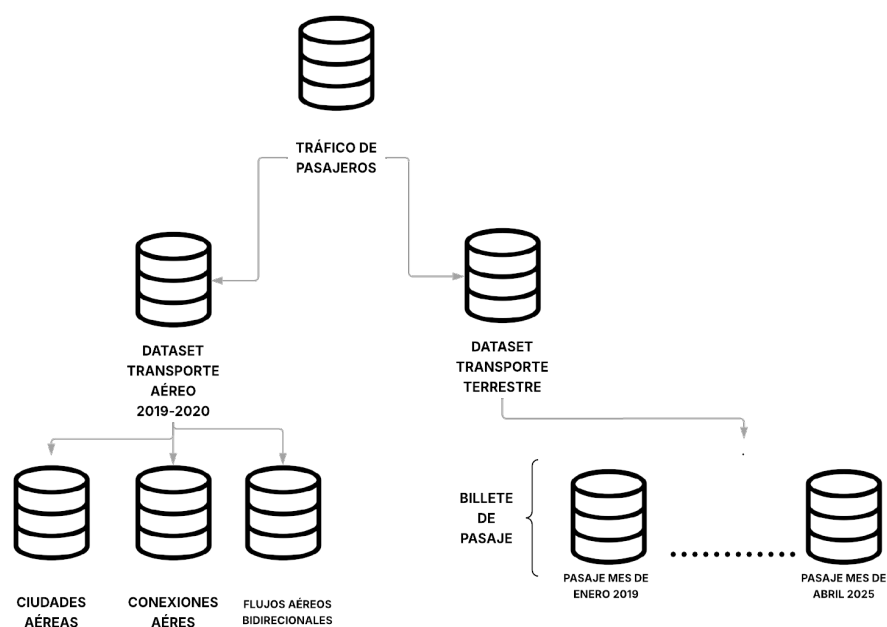
Los datos analizados provienen de fuentes oficiales que registran el transporte aéreo y terrestre, incluyendo conexiones entre ciudades, flujos de pasajeros y ventas de billetes. Estos datos, combinados con bases de datos demográficas y sanitarias, permiten construir modelos detallados de movilidad que ayudan a identificar rutas y nodos clave en la propagación de infecciones.

El análisis del comportamiento de estos datos revela la complejidad y heterogeneidad de los movimientos interurbanos, destacando la importancia de integrar distintas modalidades de transporte para obtener una visión completa. Este enfoque fue utilizado en la herramienta Epiflow[1].

Análisis del Comportamiento de Datos

El conjunto de datos está organizado en dos grandes grupos principales: transporte aéreo y transporte terrestre.

Dentro del transporte aéreo, se incluyen tres datasets: ciudades, conexiones y flujos origen-destino, que contienen información sobre la movilización de pasajeros y cargas entre las diferentes ciudades. Por otro lado, el transporte terrestre se compone de múltiples archivos con registros detallados de billetes de pasaje por usuario desde el 2019 hasta el 2025. Estos datasets se integran y procesan de forma estructurada para facilitar el análisis conjunto de la movilidad y sus impactos.



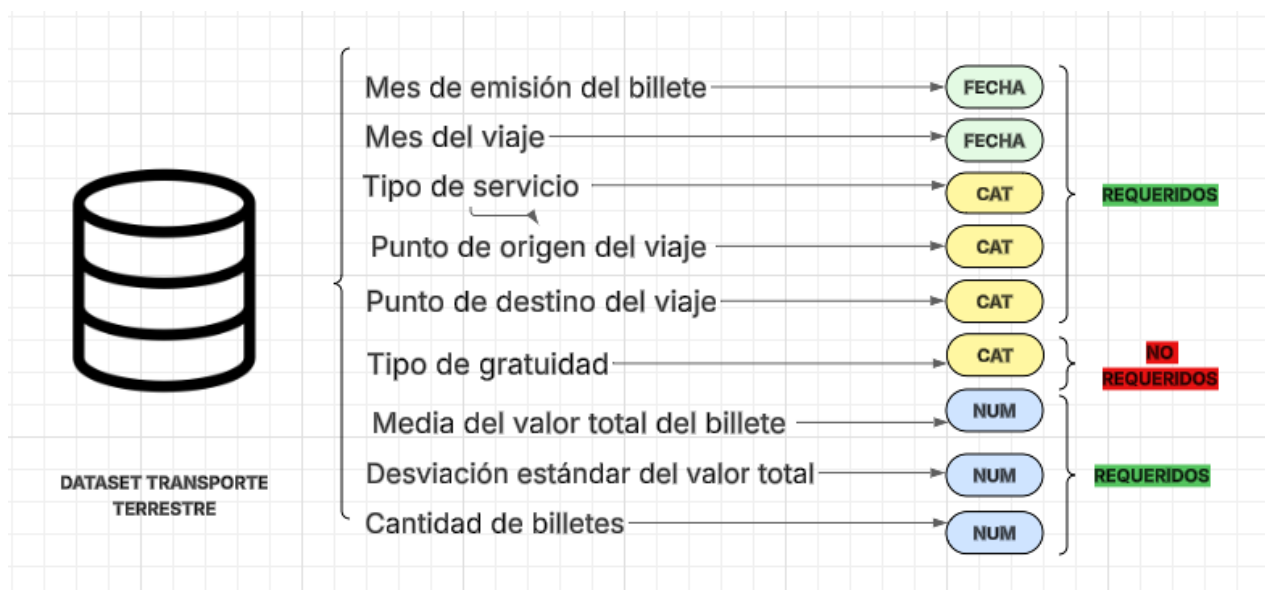
1.1 Descripción de los Datos

DATOS TERRESTRES

El dataset utilizado contiene registros detallados de billetes emitidos para el transporte terrestre interestadual e internacional de pasajeros en Brasil, correspondientes al período 2019-2025. Este conjunto incluye múltiples archivos CSV organizados por meses y años, donde cada archivo representa las ventas de billetes en distintas rutas y fechas.

Las variables principales de estos archivos incluyen:

- **mes_emissao_bilhete:** Mes y año en que se emitió el billete.
- **mes_viagem:** Mes y año en que se realiza el viaje.
- **ponto_origem_viagem:** Ciudad o terminal de origen del viaje.
- **ponto_destino_viagem:** Ciudad o terminal destino del viaje.
- **tipo_servico:** Modalidad del servicio de transporte (convencional, ejecutivo, leito, etc.).
- **tipo_gratuidade:** Categoría del descuento o gratuidad aplicada al billete (promocional, idoso, jovem, etc.).
- **media_valor_total:** Valor promedio pagado por el billete.
- **dp_valor_total:** Desviación estándar del valor del billete.
- **quantidade_bilhetes:** Cantidad de billetes vendidos en el registro.



tipo_gratuidade es solo útil si tu análisis incluye subsidios o gratuidades sociales; si no, no aporta al análisis general de viajes y volúmenes.

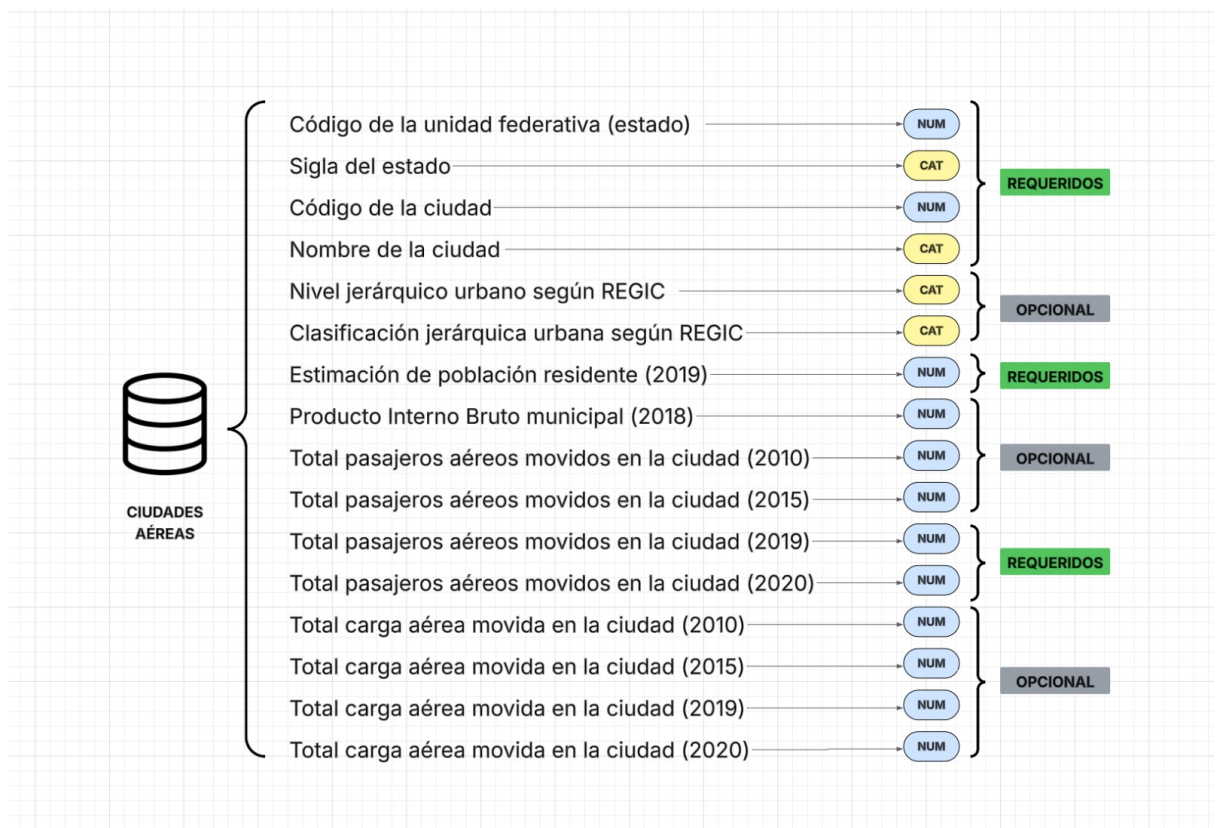
El resto de variables son clave para entender cuándo (temporal), dónde (origen/destino), qué tipo de viaje (servicio) y cuánto (cantidad y valor) viaja la gente.

Datos Aéreos

Los datos aéreos provienen de diferentes bases relacionadas con la movilidad aérea entre ciudades. Estas incluyen tres conjuntos de datos clave que reflejan la conexión entre ciudades, el movimiento de pasajeros, y los flujos de tráfico aéreo.

1. Ciudades Aéreas (LIG_AEREAS_2019-2020_cidades)

Este dataset contiene información sobre las ciudades en Brasil, incluyendo su jerarquía urbana, estimación de población, PIB y volumen de pasajeros y carga aérea movidos a lo largo de varios años.



Requeridas:

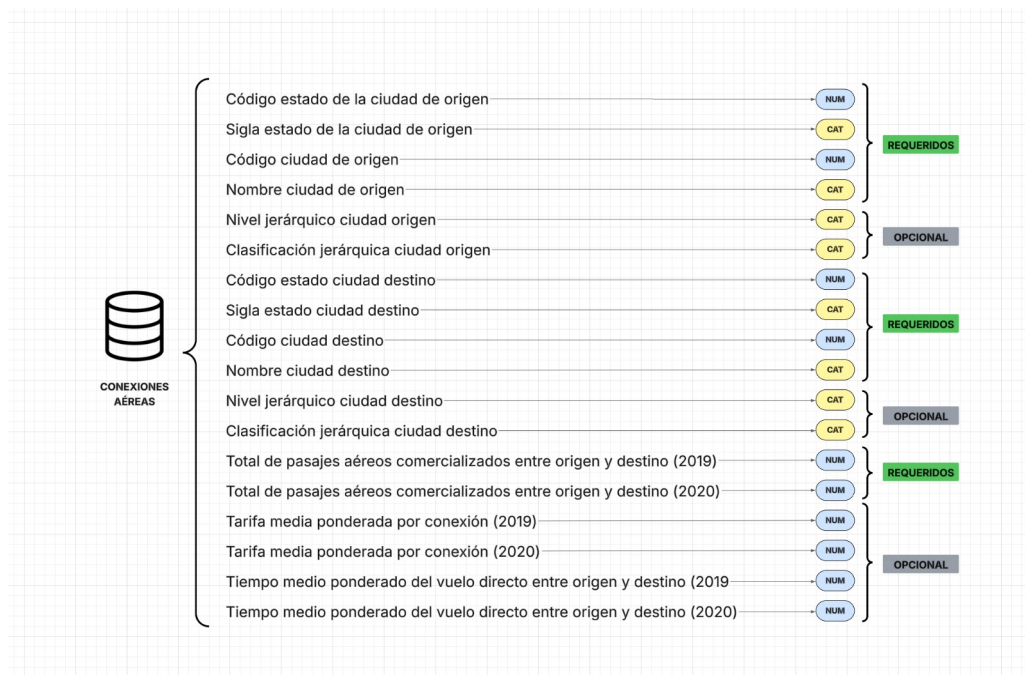
- **COD_UF** (Código de la unidad federativa): Es necesario para identificar el estado al que pertenece cada ciudad. Es crucial para cualquier análisis geográfico o de conectividad.
- **UF** (Sigla del estado): Similar al código de unidad federativa, pero más directo para asociar las ciudades con los estados en una representación más fácil de leer.
- **COD_CIDADE** (Código de la ciudad): Identificador único para cada ciudad, fundamental para diferenciar entre las ciudades en los análisis.
NOME_CIDADE (Nombre de la ciudad): Permite referirse a las ciudades de forma legible. Es útil para visualizaciones o análisis descriptivos.

- **VAR01** (Estimación de población residente en 2019): La población de una ciudad es un factor clave para entender la demanda potencial de transporte aéreo.
- **VAR05** (Total de pasajeros aéreos movidos en la ciudad en 2019): Es la variable más importante para conocer la magnitud del tráfico aéreo de una ciudad en un periodo reciente.
- **VAR06** (Total de pasajeros aéreos movidos en la ciudad en 2020): Similar a la anterior, proporciona información actualizada y comparativa para medir tendencias recientes.

Opcionales:

- **NIVEL_CID** (Nivel jerárquico urbano según REGIC) y **CLASS_CID** (Clasificación jerárquica urbana según REGIC): Aunque son útiles para clasificar y entender la importancia de las ciudades en términos de conectividad, no son estrictamente necesarias para realizar análisis de tráfico aéreo básico.
- **VAR02** (PIB municipal 2018), **VAR03** (Pasajeros movidos en 2010), **VAR04** (Pasajeros movidos en 2015), **VAR07** (Carga aérea movida en 2010), **VAR08** (Carga aérea movida en 2015), **VAR09** (Carga aérea movida en 2019), **VAR10** (Carga aérea movida en 2020): Aunque estos datos son útiles para análisis longitudinales y comparativos, no son necesarios si el enfoque está únicamente en el tráfico de pasajeros más reciente (2019 y 2020).
-

2. Conexiones Aéreas (LIG_AEREAS_2019-2020_ligacoes)



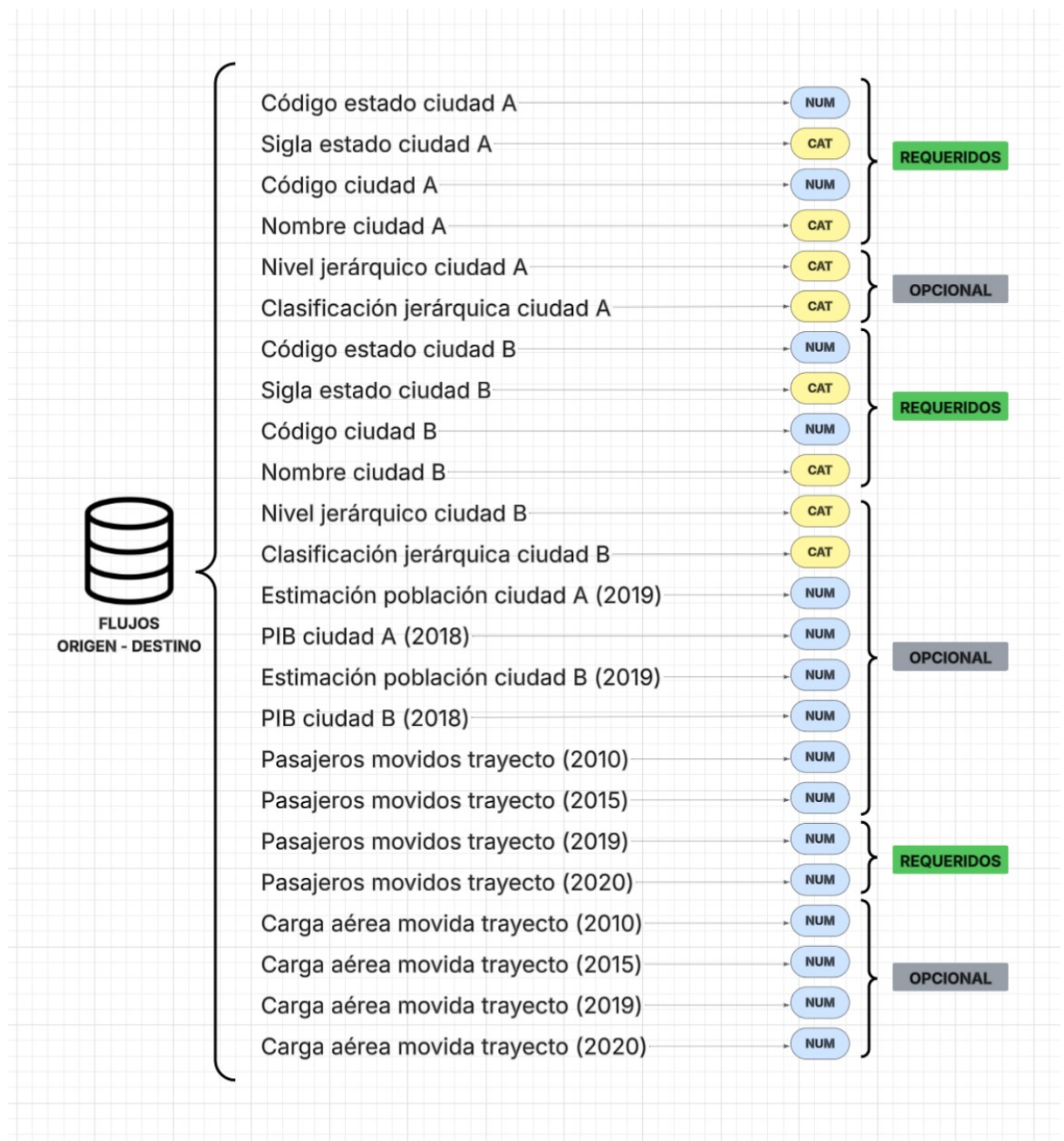
Requeridas:

- **COD_UF_O** y **UF_O** (Código y sigla del estado de la ciudad de origen): Necesarios para identificar el origen del vuelo y asociarlo con un estado específico.
- **COD_CID_O** y **NOME_CID_O** (Código y nombre de la ciudad de origen): Los códigos de las ciudades de origen son claves para diferenciar entre vuelos y destinos.
- **COD_UF_D** y **UF_D** (Código y sigla del estado de la ciudad de destino): Iguales a las variables de origen, pero para el destino del vuelo.
- **COD_CID_D** y **NOME_CID_D** (Código y nombre de la ciudad de destino): Permiten identificar la ciudad de destino, lo cual es necesario para analizar la conectividad entre ciudades.
- **VAR01** y **VAR02** (Total de pasajes comercializados en 2019 y 2020): Los datos de pasajes vendidos son esenciales para medir la demanda de vuelos entre ciudades.

OPcionales:

- **NIVEL_O**, **CLASS_O** (Nivel y clasificación de la ciudad de origen) y **NIVEL_D**, **CLASS_D** (Nivel y clasificación de la ciudad de destino): Aunque pueden proporcionar información valiosa sobre la jerarquía de las ciudades en la red aérea, no son estrictamente necesarios para la análisis de la cantidad de pasajes comercializados o la conectividad básica entre ciudades.
- **VAR03**, **VAR04** (Tarifa media ponderada por conexión en 2019 y 2020) y **VAR05**, **VAR06** (Tiempo medio ponderado del vuelo directo entre origen y destino en 2019 y 2020): Estas variables son útiles si el análisis busca comparar tarifas o tiempos de vuelo, pero no son esenciales para estudiar la conectividad y el volumen de pasajes

3. Flujos Origen-Destino(LIG_AEREAS_2019-2020_fluxos_od)



Requeridas:

- COD_UF_A, UF_A, COD_CID_A, NOME_CID_A** (Código y nombre del estado y ciudad A): Son necesarios para identificar el origen de los flujos de pasajeros y carga.
COD_UF_B, UF_B, COD_CID_B, NOME_CID_B (Código y nombre del estado y ciudad B): Iguales a las variables de origen, pero para la ciudad destino. Son claves para cualquier análisis de flujos entre ciudades.
- VAR07** (Pasajeros movidos entre las ciudades en 2019) y **VAR09** (Pasajeros movidos entre las ciudades en 2020): Estas variables permiten medir la cantidad de pasajeros que viajan entre las ciudades, lo cual es esencial para el análisis de movilidad.