

---

Informe Final:  
Análisis Exploratorio de Datos  
Docente: Ana Maria CuadrosValdivia

## INFORME FINAL DE ANÁLISIS EXPLORATORIO DE DATOS DEL TRANSPORTE

Alumna: Luciana Julissa Huaman Coaquira

### Motivación:

La investigación tiene como objetivo analizar el comportamiento de los pasajeros tanto en el transporte aéreo como rodoviario. Las hipótesis se originaron a partir de la observación de patrones comunes en las industrias de transporte y la posibilidad de correlacionar variables como la estacionalidad y las rutas más populares.

### Hipótesis 1:

Pregunta: ¿Existen patrones estacionales en la demanda de boletos en el transporte rodoviario?

Motivación: Analizar las fluctuaciones en la cantidad de boletos vendidos por mes para determinar si hay picos estacionales.

### Hipótesis 2:

Pregunta: ¿Los destinos más populares en el transporte rodoviario tienen una correlación directa con los destinos más populares en el transporte aéreo?

Motivación: Comparar las rutas más frecuentes de ambos tipos de transporte para ver si las ciudades con mayor demanda en el transporte rodoviario coinciden con las más solicitadas en el transporte aéreo.

### Hipótesis 3:

Pregunta: ¿La cantidad de pasajeros transportados en vuelos aéreos varía según el mes del año?

Motivación: Explorar si hay un comportamiento estacional en la cantidad de pasajeros, con ciertos meses más demandados que otros.

### Plan de Análisis:

Para investigar las hipótesis planteadas se siguieron los siguientes pasos:

1. Hipótesis 1:  
Se analizaron los datos de boletos vendidos mensualmente para el transporte rodoviario. Se generaron gráficos de barras para identificar cualquier patrón estacional dado que en el dataset original brindaban datos de estos.
2. Hipótesis 2:  
Se compararon las rutas más populares de ambos tipos de transporte. Se realizó un gráfico de barras para ver la cantidad de boletos vendidos por destino tanto para el transporte aéreo como rodoviario , en temas de propagación la movilidad humana es clave.
3. Hipótesis 3:  
Se analizaron los datos de pasajeros aéreos por mes. Se utilizó un gráfico de líneas para observar cualquier fluctuación mensual en la demanda.

Fuente de Datos:

- Fuente: Los datos provienen de los archivos `aereos_final.csv` y `rodoviarios_final.csv`.
- Técnica de Recolección: Los datos fueron recolectados de fuentes oficiales, como autoridades gubernamentales y empresas de transporte, y luego fueron limpiados para enfocarse únicamente en la información más relevante que influye en el tema de la propiedad, excluyendo aquellos datos que no aportan valor al análisis.

Transporte Terrestre

<https://dados.antt.gov.br/dataset/monitriip-bilhetes-de-passagem>

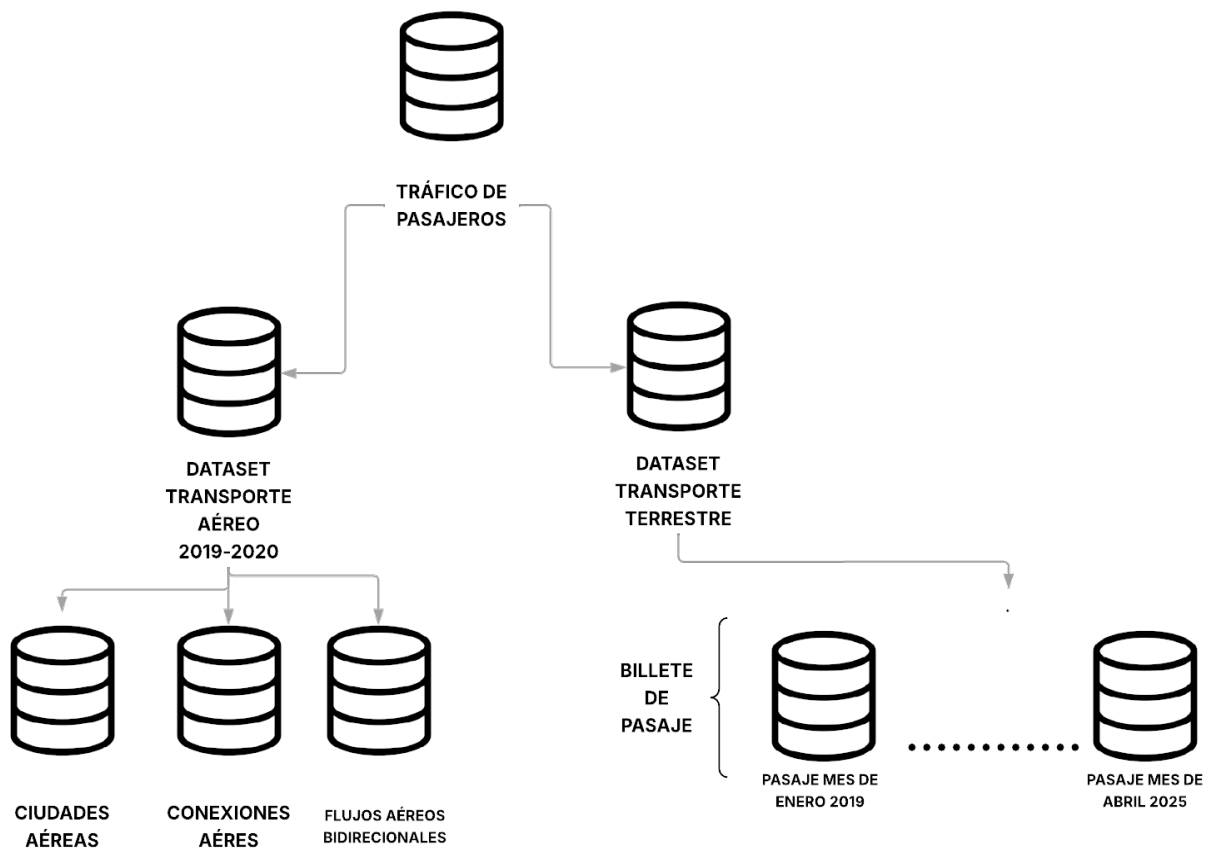
Transporte Aéreo

<https://www.anac.gov.br/acesso-a-informacao/dados-abertos/areas-de-atuacao/voos-e-operacoes-aereas/dados-estatisticos-do-transporte-aereo>

## Exploración 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.



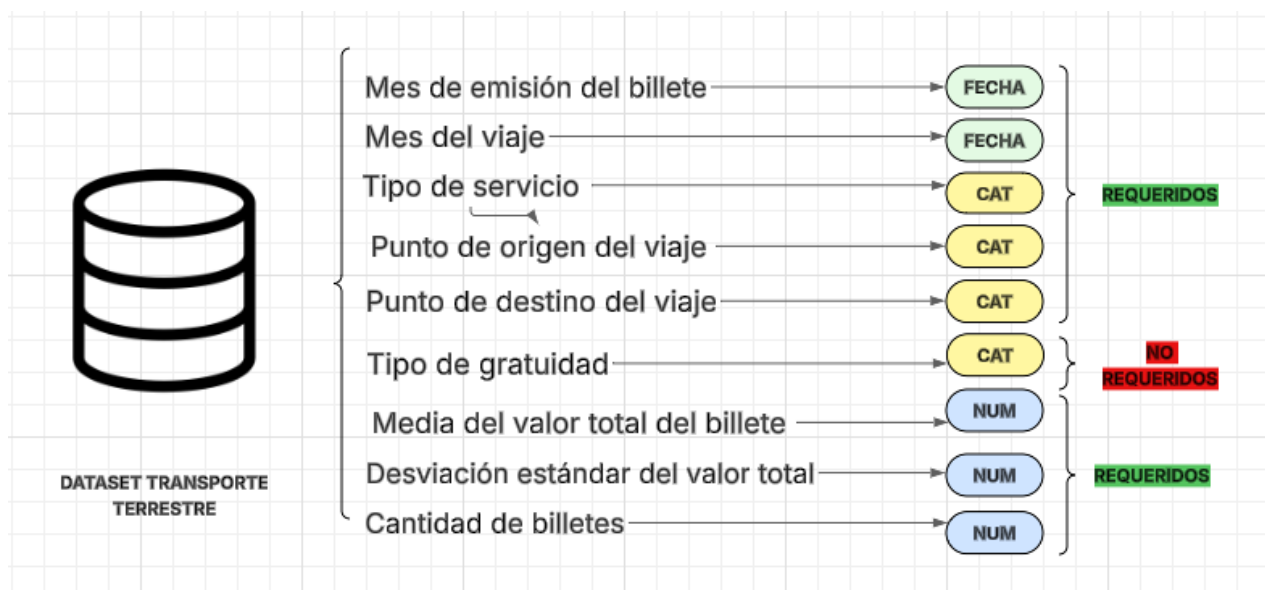
## 1.1 Descripción de los Datos

### 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.

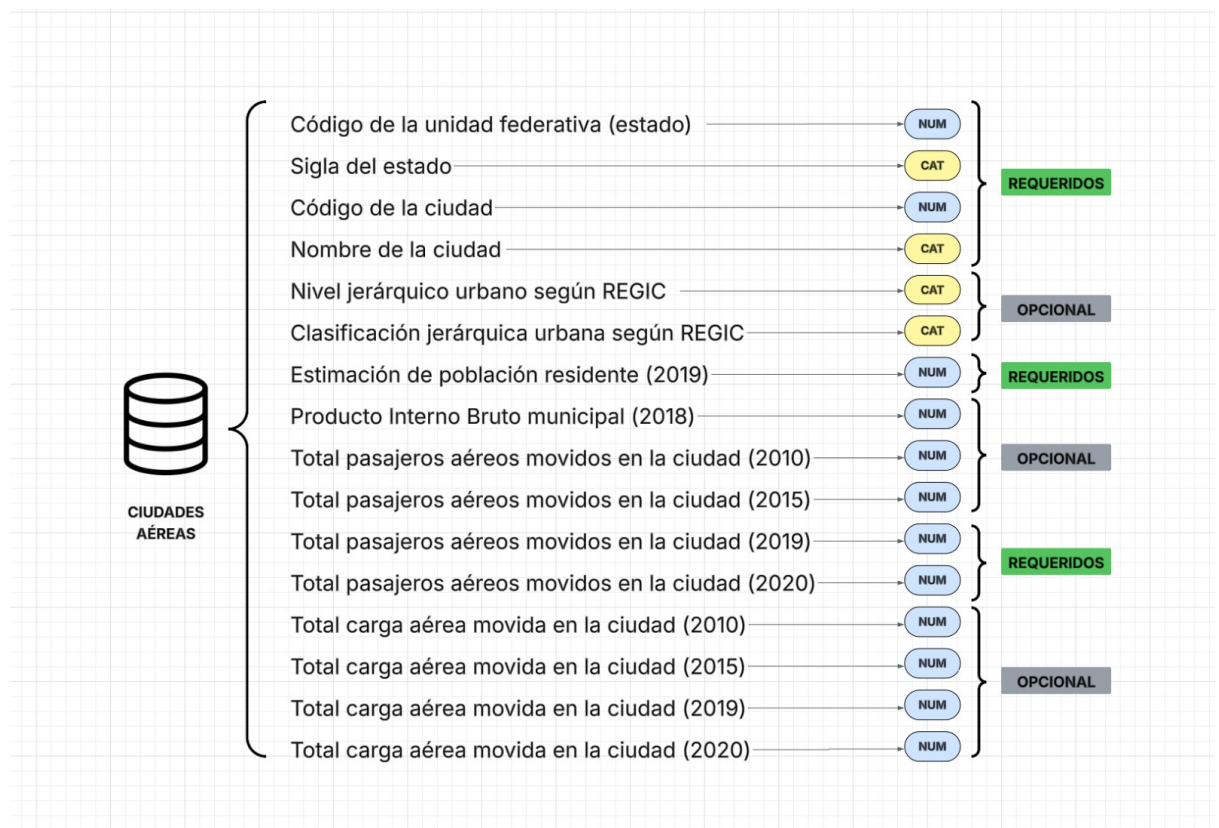
El resto de variables son clave para entender cuándo (temporal), dónde (origen/destino), qué tipo de viaje (servicio) y cuánto (cantidad y valor) viaja la gente.

# 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:

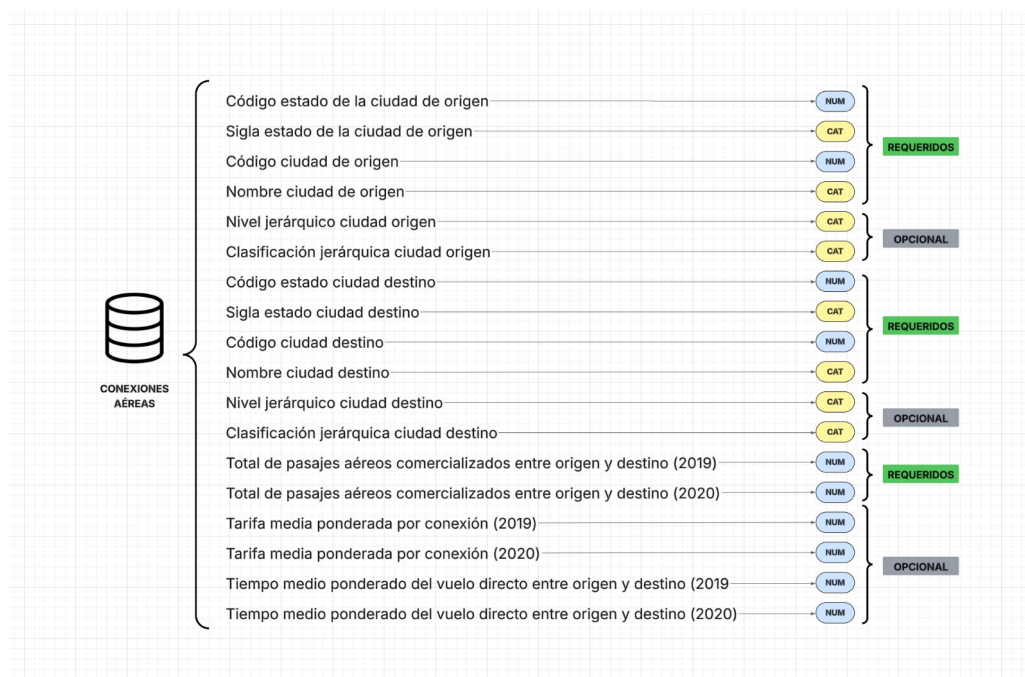
- COD\_UF (Código de la unidad federativa): Es necesario para identificar el estado al que pertenece cada ciudad. Es crucial para cualquier análisis geográfico o de conectividad.
- UF (Sigla del estado): Similar al código de unidad federativa, pero más directo para asociar las ciudades con los estados 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)



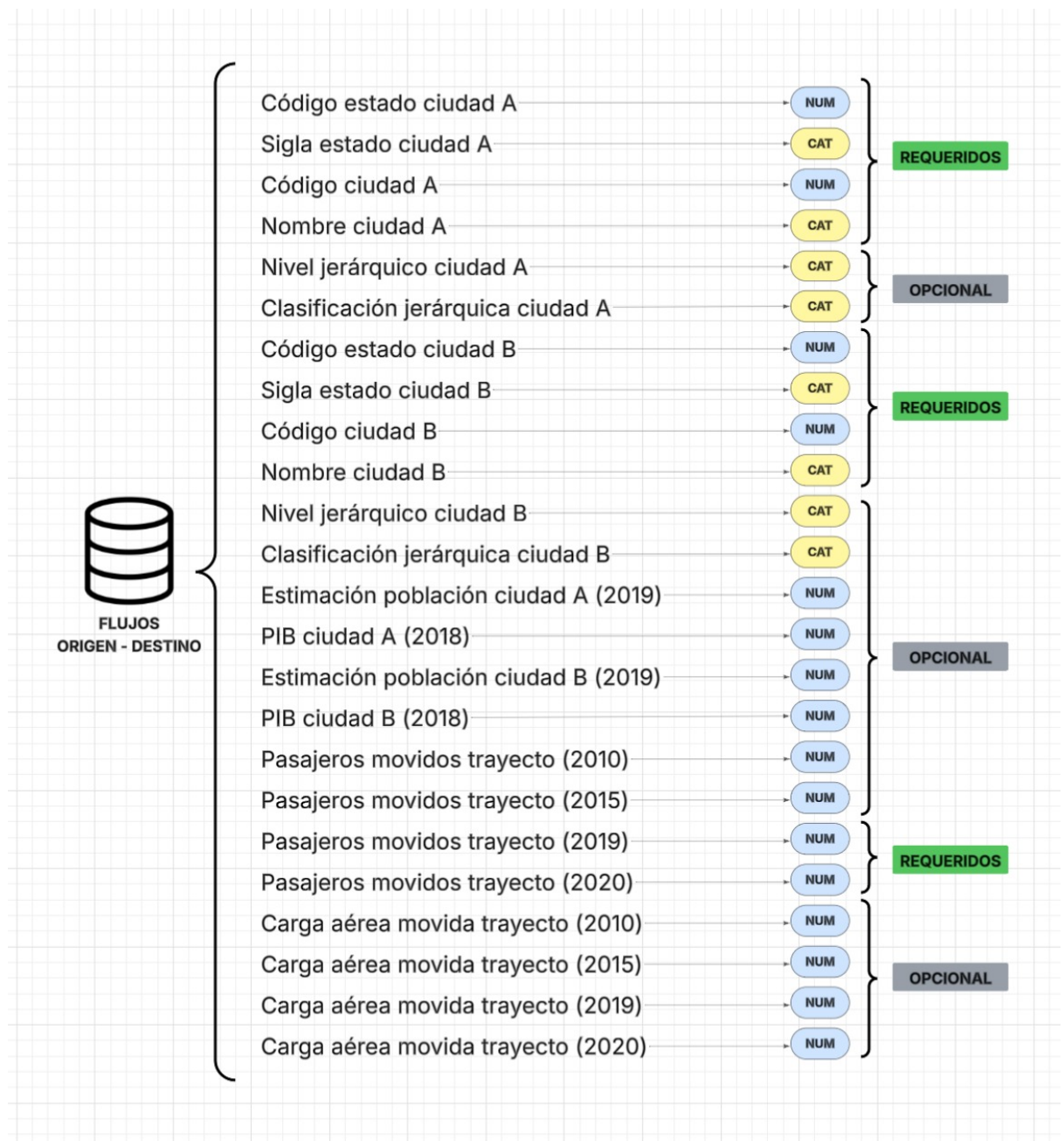
### Requeridas:

- 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 el 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.



#### Hipótesis 1: Estacionalidad en la Demanda de Boletos Rodoviaros

- Se observó que algunos meses (principalmente en los periodos festivos) presentan un incremento en la venta de boletos, con una caída pronunciada en los meses de la pandemia en 2020.

La demanda de transporte rodoviario puede verse influenciada por diversos factores estacionales, entre ellos la propagación de enfermedades. Las pandemias o brotes de enfermedades suelen afectar los patrones de desplazamiento de los pasajeros. Por ejemplo, durante un brote de enfermedad como el COVID-19, la demanda de transporte puede disminuir considerablemente debido a restricciones de movilidad, medidas de cuarentena o el miedo a la propagación del virus.

En este análisis, los datos de boletos vendidos se pueden analizar para identificar si los patrones de movilidad se ven alterados por factores relacionados con la salud pública. Durante los picos de epidemias o enfermedades, puede observarse una disminución en la cantidad de boletos vendidos debido a la menor disposición de las personas a viajar por cuestiones de salud. Esta hipótesis ayudará a entender cómo los brotes de enfermedades afectan la demanda de transporte, lo que a su vez puede ayudar a ajustar las estrategias operativas y de comercialización para las empresas de transporte.

#### Hipótesis 2: Correlación entre Rutas Aéreas y Rodoviaras

- Las ciudades con mayor demanda en el transporte rodoviario coinciden en muchos casos con las más frecuentadas en el transporte aéreo.

La movilidad humana entre diferentes medios de transporte, como el aéreo y el rodoviario, también puede estar influenciada por factores relacionados con la salud pública. Durante los brotes de enfermedades, los patrones de viaje tienden a cambiar, especialmente en rutas largas, donde las personas podrían preferir un modo de transporte más seguro o de menor contacto, como los autobuses, en lugar de los vuelos.

La comparación de las rutas más populares entre ambos modos de transporte puede ayudar a identificar si las ciudades más afectadas por brotes de enfermedades muestran una alta demanda tanto en el transporte rodoviario como aéreo. Si los destinos coinciden en ambos sistemas de transporte, esto podría indicar un patrón de movilidad de emergencia o de necesidad urgente de viajar, posiblemente debido a la propagación de una enfermedad. Este análisis puede ser valioso para las autoridades sanitarias y empresas de transporte, para que ajusten sus estrategias operativas según los cambios en la movilidad durante periodos de crisis sanitaria.

### Hipótesis 3: Estacionalidad en el Transporte Aéreo

- Se encontró una estacionalidad clara en la cantidad de pasajeros, con picos en ciertos meses, como diciembre y las vacaciones de verano.
- La cantidad de pasajeros transportados por aire tiende a variar estacionalmente, pero estas fluctuaciones pueden estar también influenciadas por el impacto de enfermedades. Durante un brote o una pandemia, las fluctuaciones en la demanda de vuelos no solo estarán relacionadas con la estacionalidad normal, sino también con el nivel de restricción de movilidad impuesto por las autoridades sanitarias. En meses de mayor propagación de enfermedades, la demanda de vuelos puede disminuir, mientras que en otros momentos, cuando las restricciones se alivian, la demanda puede aumentar.

Por ejemplo, en un contexto de brote epidémico, la demanda de vuelos puede verse completamente reducida durante los picos de la enfermedad y recuperarse después del control de la crisis, cuando las personas necesitan desplazarse para tratarse, asistir a tratamientos o regresar a sus hogares. El análisis de esta hipótesis permite entender cómo el comportamiento estacional de los pasajeros se ve modificado por factores sanitarios, proporcionando a las aerolíneas datos cruciales para ajustar la oferta de vuelos, gestionar la capacidad de los aviones y diseñar estrategias de precios adaptadas a periodos de baja demanda por cuestiones de salud.

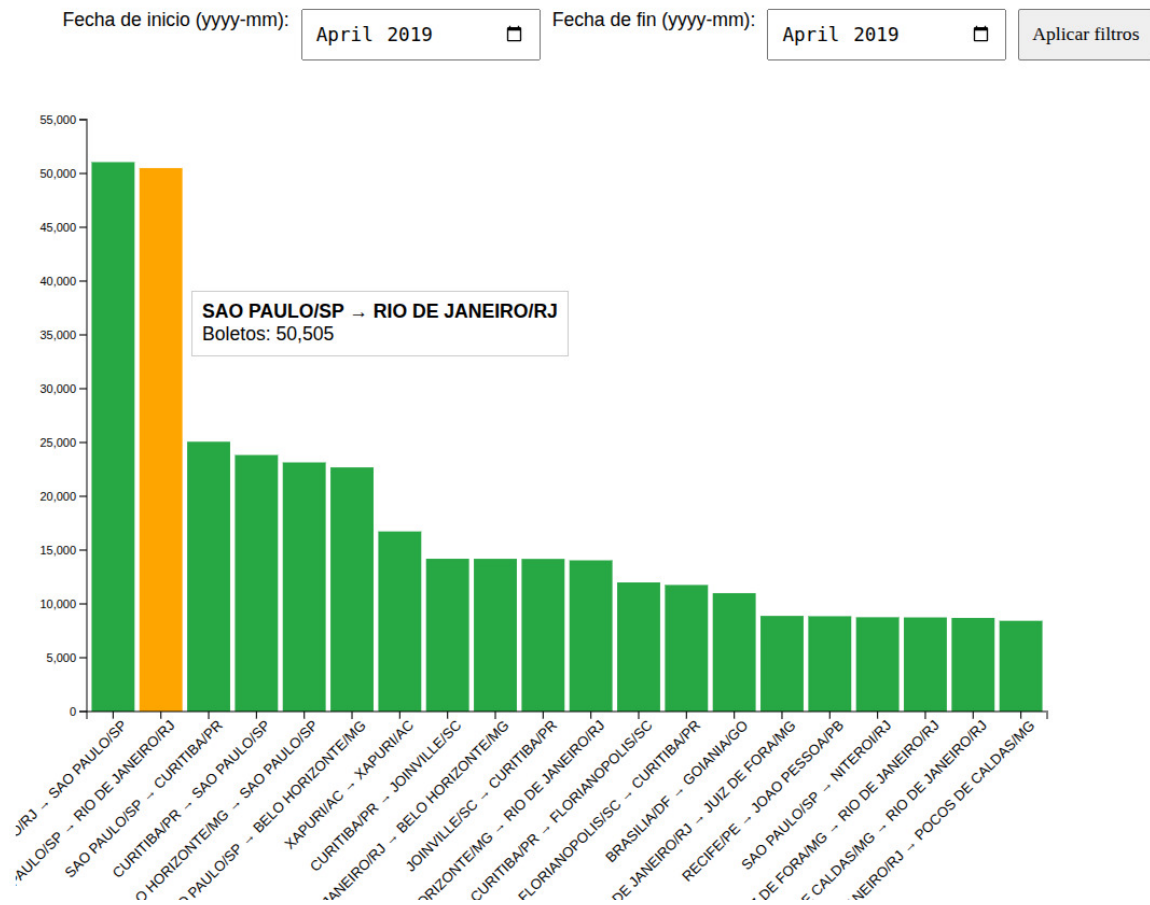
### Conclusiones:

- Hipótesis 1: La demanda de boletos en el transporte rodoviario presenta picos estacionales, especialmente en los meses de vacaciones y festividades, y una notable caída durante la pandemia de COVID-19 en 2020.
- Hipótesis 2: Hay una correlación positiva entre las rutas más populares en ambos tipos de transporte. Las ciudades con alta demanda en el transporte rodoviario, como São Paulo y Río de Janeiro, también tienen una alta demanda en el transporte aéreo.
- Hipótesis 3: El transporte aéreo muestra una clara estacionalidad en la cantidad de pasajeros, con los meses de diciembre y enero siendo los de mayor tráfico aéreo.

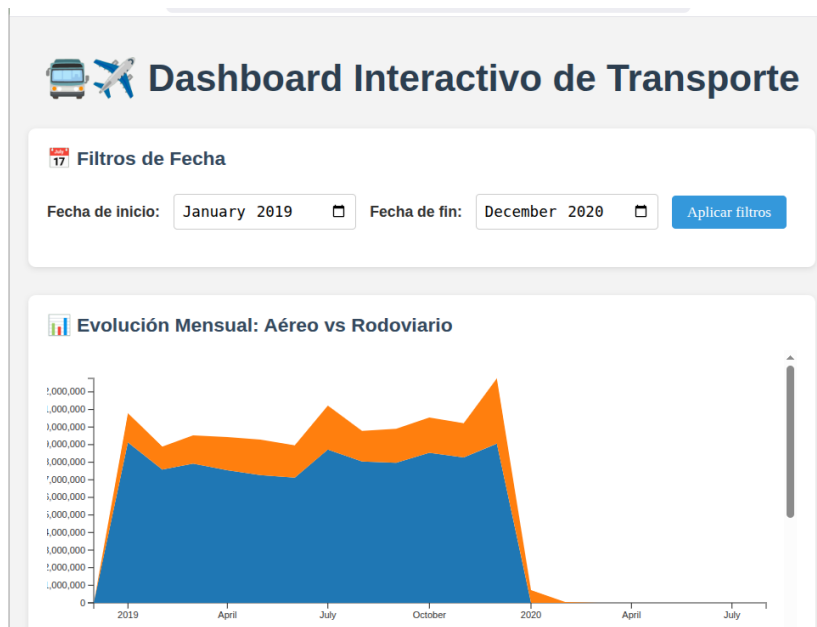
## VISUALIZACIONES:

Hipótesis 1:

### Dashboard de Transporte Rodoviario



Hipótesis 2:



Hipótesis 3:

