

**UNIVERSIDAD DIEGO PORTALES**  
**FACULTAD DE INGENIERÍA Y CIENCIAS**  
**ESCUELA DE INFORMÁTICA Y TELECOMUNICACIONES**



---

**Data Science**  
**Informe N°2: Avance de Proyecto**

---

**Integrantes:**

Lucas Almonacid  
Benjamín Fernández  
Felipe Condore  
Pamela Saldías

**Profesor:**

Karol Suchan

**Fecha de entrega:**

27 de Noviembre de 2022

# Índice

<b>1. Usuario</b>	<b>1</b>
<b>2. Descripción de la Problemática</b>	<b>2</b>
<b>3. Datos del Problema</b>	<b>3</b>
<b>4. Objetivos y medidas de desempeño</b>	<b>6</b>
4.1. Objetivo General . . . . .	6
4.2. Objetivo Específico . . . . .	6
4.3. Medidas de Desempeño . . . . .	6
<b>5. Alternativas de Solución</b>	<b>6</b>
5.1. Estructura de Datos . . . . .	6
5.2. Modelamiento y procesamiento de datos . . . . .	6
5.2.1. Indexación de buses a paraderos . . . . .	6
5.2.2. Separación de rutas . . . . .	7
5.2.3. Procesamiento de Frecuencia . . . . .	7
5.2.4. Procesamiento de Regularidad . . . . .	7
5.3. Herramientas Utilizadas . . . . .	7
5.3.1. Apache Spark . . . . .	7
5.3.2. Postgres - PostGis . . . . .	7
5.3.3. Qgis . . . . .	7

## 1. Usuario

El usuario objetivo de este proyecto corresponde al ente encargado de planificar los aspectos estratégicos referentes al transporte público en la región Metropolitana, tales como, rutas, frecuencia de micros, cantidad de micros por ruta, etc. Por ello, con el fin de comprender a quien corresponde esta función, es necesario entender al ente encargado del transporte público, el cual corresponde al Ministerio de Transportes y Telecomunicaciones de Chile actualmente dirigido por Juan Carlos Muñoz Abogabir. Tal y como es mencionado por el propio ministerio “La principal función del ministerio es proponer las políticas nacionales relacionadas con el transporte y telecomunicaciones acorde a las directrices del Gobierno y ejercer la dirección y control de su puesta en práctica; supervisar las empresas públicas y privadas que operen medios de transportes y comunicaciones en el país, y coordinar y promover el desarrollo de estas actividades y controlar el cumplimiento de las leyes, reglamentos y normas pertinentes.” [7].

El ministerio está dividido en dos entidades principales: la Subsecretaría de Telecomunicaciones y Subsecretaría de Transportes. En el caso de la Subsecretaría de Transportes, esta es dirigida por Cristóbal Pineda Andradez, y tal como es descrito por la propia subsecretaría: “La Subsecretaría es la principal encargada de promover el desarrollo de sistemas de transportes eficientes, seguros y sustentables. Esto a través de la definición de políticas, normas y del control de su cumplimiento. Para de esta manera contribuir a la integración territorial del país, favorecer el desarrollo económico y asegurar servicios de alta calidad a los usuarios” [5].

Dentro de la Subsecretaría de Transportes, el encargado de gestionar el transporte público de la región Metropolitana es el Directorio de Transporte Público Metropolitano (DTPM) dirigido actualmente por Paola Tapia Salas, siendo esta entidad el usuario objetivo de la propuesta.

El organigrama de esta entidad se puede observar en la siguiente figura:

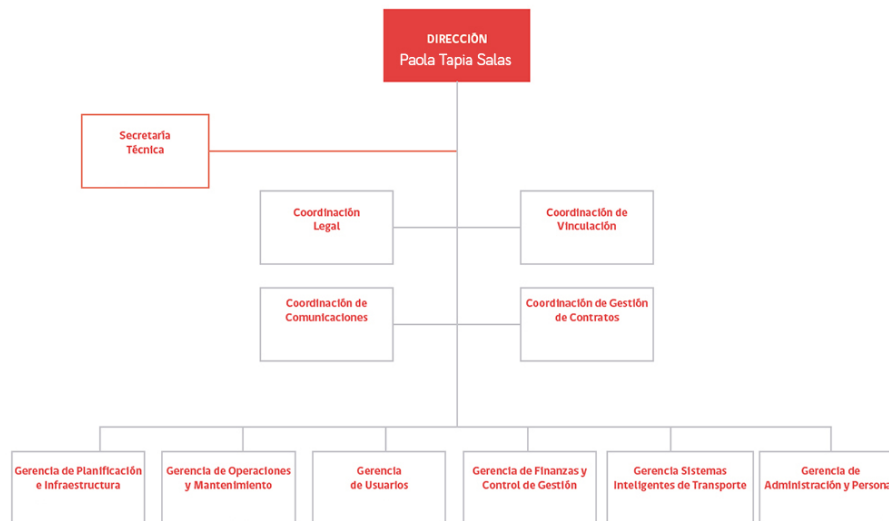


Figura 1: Organigrama DTPM del año 2022.

Las funciones generales del DTPM son descritas por este mismo en varios puntos, los cuales se pueden observar a continuación:

- “Proponer a las autoridades del Ministerio de Transportes y Telecomunicaciones estudios, planes de licitación de transporte público y las condiciones administrativas, económicas y financieras de dichas licitaciones, así como también el programa presupuestario que se requiera para el cumplimiento del Plan de Transporte Urbano de donde se deriva el Sistema de Transporte Público de Santiago” [6].
- “Coordinar los procesos de licitación de vías y la contratación de los servicios de transporte público, así como la revisión de especificaciones y contratos respecto de los servicios complementarios” [6].
- “Coordinar los procesos de negociación que pudieran requerirse en el marco de las acciones para dar cumplimiento a las misiones encomendadas” [6].

- “Supervisar los contratos, participar en instancias de estudios, análisis y mejoramiento del transporte público” [6].
- “Servir de instancia de coordinación para las autoridades y organismos involucrados en la definición y ejecución de los programas, planes y medidas aplicadas al Sistema de Transporte Público de la ciudad de Santiago” [6].
- “Revisar, actualizar y renovar el Plan Maestro de Infraestructura de Transporte Público, y coordinar la ejecución de las obras contenidas en dicho plan por parte de los organismos ejecutores, y ejecutar obras públicas menores de transporte público” [6].
- “Velar por la correcta operación del Sistema, a través del seguimiento de las metas y plazos que se definan para la ejecución de sus programas, planes y medidas” [6].
- “Establecer vínculos de coordinación y colaboración con organismos públicos y privados, nacionales, extranjeros o internacionales, que desarrollan actividades en el ámbito del transporte público de pasajeros” [6].
- “Velar por la oportuna y adecuada satisfacción de las necesidades de los usuarios y proponer los ajustes correspondientes en los lineamientos de la autoridad sectorial” [6].
- “Evaluar la normativa vigente y proponer los cambios legales y reglamentarios que resulten necesarios para la creación de una institucionalidad que vele en forma permanente por una adecuada prestación de los servicios de transporte público en la región Metropolitana” [6].

## 2. Descripción de la Problemática

La fiscalización del transporte público es un problema que Transantiago debe enfrentar día a día, junto con brindar el mejor servicio a los clientes. Esto último implica poder habilitar los buses que sean necesarios para que no haya congestión y las personas puedan llegar a su destino. En la actualidad en el transporte público existen distintas concesiones, las cuales deben de cumplir un contrato que ayuda, entre muchas otras cosas, en mantener cierta calidad de servicio brindado a las personas que lo consuman.

El problema es que en la actualidad, DTPM tiene información generalizada respecto al cumplimiento de las concesionarias sobre las métricas de frecuencia y regularidad, pero no se puede identificar cuáles son los puntos críticos en donde estas pueden estar fallando. Tener conocimiento del cumplimiento de frecuencia y regularidad, asociada a los buses pertenecientes a cada ruta, permitiría identificar con mayor facilidad los tramos que se vean más afectados. Por otra parte, brindar esta herramienta permitiría agilizar la toma de medidas correspondientes para gestionar el flujo del transporte público. Un ejemplo podría ser plantear nuevas rutas, aumentar el flujo de buses a ciertas horas, etc.

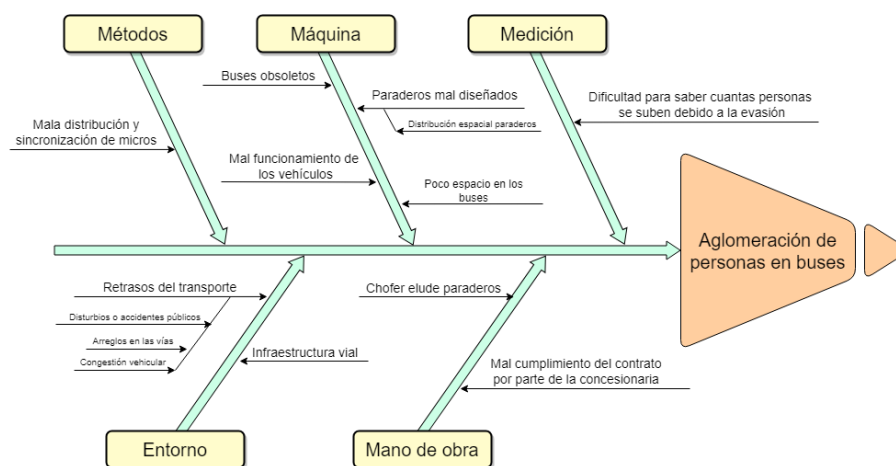


Figura 2: Diagrama Ishikawa

En la figura adjunta se puede observar un “Diagrama de Ishikawa”, este describe distintas causas que pueden generar el problema a resolver. Para este problema, se dividieron las causas en cinco puntos correspondientes a métodos,

máquina, medición, entorno y la mano de obra.

La solución propuesta para este trabajo no ayudará a las fallas dadas por el entorno debido a que estas, con el dataset con el cual se va a trabajar, no se pueden predecir ni estimar. La solución para este problema que brindara este trabajo ayudara a la mala distribución y sincronización de los buses, junto con el mal cumplimiento de contrato por parte de las concesionarias. Con respecto a la dificultad al saber cuantas partes han subido al bus, se refiere directamente a la evasión, dificultando así, obtener mediciones exactas de acumulación de personas dentro de los buses.

### 3. Datos del Problema

Para lograr medir el desempeño de las concesionarias, el Directorio de Transportes Públicos Metropolitanos (DTPM), elabora una serie de rankings presentando un resumen trimestral sobre la calidad de servicio de las diversas empresas concesionarias. Entre los principales indicadores que se mencionan están:

- **Indicadores de Frecuencia:**

Mide la cantidad real de buses que cada empresa dispuso para sus recorridos y lo compara con el número de buses planificado, según lo indicado en los Programas de Operación del Directorio de Transporte Público Metropolitanos. La frecuencia de cada empresa concesionaria mide las salidas de buses efectivamente realizadas desde el punto de inicio de cada recorrido (cabezales y terminales).

- **Indicadores de Regularidad:**

Mide el cumplimiento del intervalo existente entre buses de un mismo recorrido, y lo compara con lo indicado en los Programas de Operación sancionados por el Directorio de Transporte Público Metropolitanos. El Indicador de Cumplimiento de Regularidad tiene por objetivo prevenir que se produzcan tiempos excesivos entre buses respecto a lo planificado, lográndose así menores tiempos de espera para los usuarios.

Los datos realizados en estos rankings son obtenidos a partir de un sistema de georreferenciación de Google llamado GTFS instalado en cada uno de los buses y las bases de datos de transacciones de tarjetas Bip, los cuales en conjunto al software de ADATRAP permite realizar un seguimiento al flujo de personas, y la generación de matrices de origen destino para localizar los lugares de subida y bajada de usuarios. Los datos utilizados para el desarrollo de esta solución, se encuentran almacenados en una base de datos Postgres de Citylog (Centro de innovación en transporte y logística UDP)[1], y corresponden a un post-procesamiento de los datos provenientes de las fuentes anteriores. Los resultados de dicho procesamiento, generan una serie de tiempo, donde es señalada la cantidad de pasajeros presentes en un bus en cada momento del trayecto, la ruta asociada, su ubicación georreferenciada en dicho momento, entre otros. A partir de estos, es posible reconstruir el flujo de micros para cada uno de los paraderos en sus rutas correspondientes, y medir el grado de cada métrica asociada a tramos en específico, de manera similar a como la DTPM realiza estadísticas sobre distintos niveles de rendimiento del servicio de transporte público que evidencian la problemática planteada anteriormente. Los indicadores utilizados a continuación, muestran la información sobre los datos correspondientes al ranking de calidad de servicio de 2018, y al informe de gestión anual de 2018.[2][3][4]

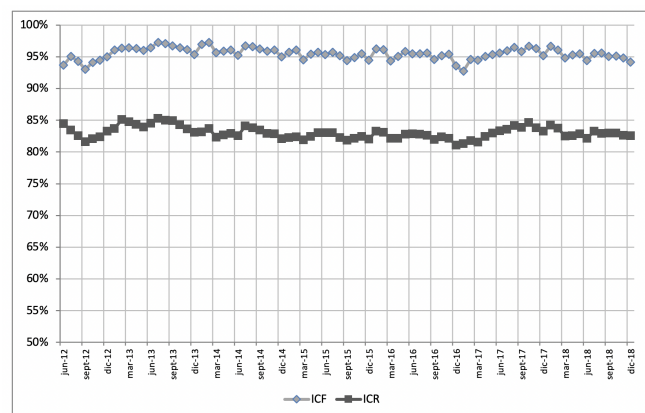


Figura 3: Evolución de los indicadores de frecuencia (ICF) y regularidad (ICR) 2012 – 2018

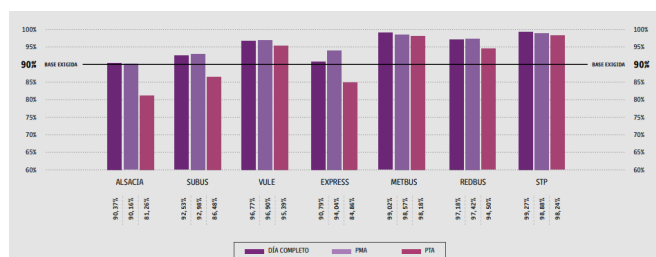


Figura 4: Indicadores de frecuencia de Abril a Junio 2018

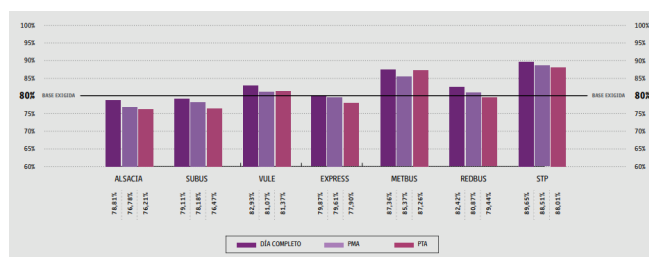


Figura 5: Indicadores de regularidad de Abril a Junio 2018

En las figuras 4 y 5, se observa que no todas las concesionarias cumplen con la base mínima exigida, la cual se encuentra establecida en los contratos de estas mismas. Esto puede deberse tanto a factores como la evasión, los tramos o recorridos asignados a cada concesionaria y la congestión de paraderos ubicados en estas mismas.

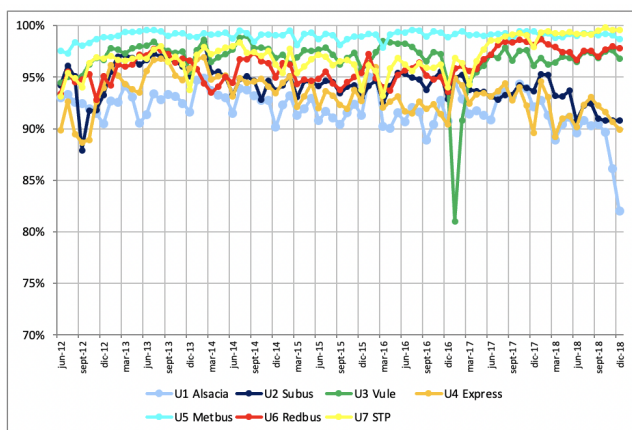


Figura 6: Evolución mensual del indicador de frecuencia, por Unidad de Negocio 2012 - 2018

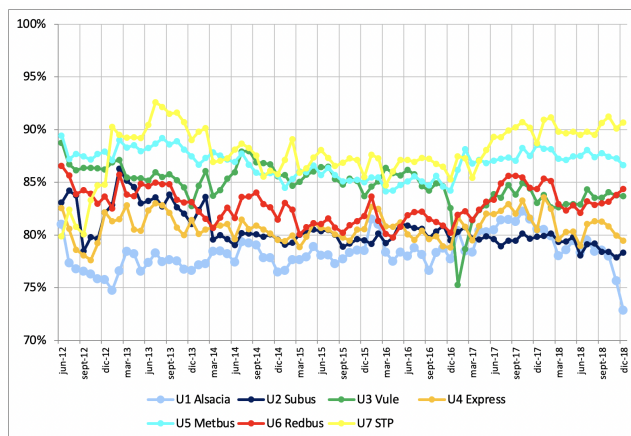


Figura 7: Evolución mensual del indicador de regularidad, por Unidad de Negocio 2012 - 2018

Por otra parte, también si se extiende el periodo de evaluación de estos indicadores desde el año 2012 a 2018, como se observa en las figuras 6 y 7, se mantiene el mismo comportamiento donde no se satisface la base mínima exigida.

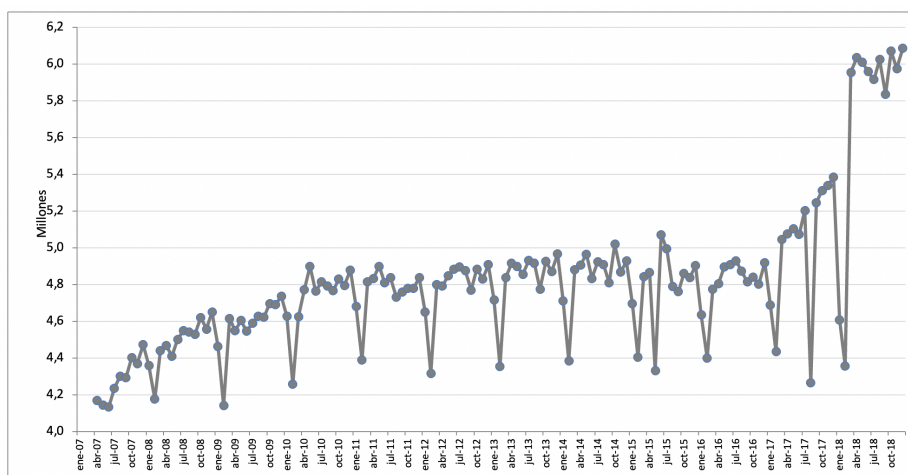


Figura 8: Evolución del promedio mensual de usuarios del sistema 2007 - 2018

En la figura 8 se puede observar que a partir de los datos, es posible determinar la cantidad de usuarios que se encuentran en el sistema. En particular, para el año 2018 se incrementa en gran medida esta cantidad, cuyo

periodo corresponde al intervalo donde se encuentran los datos obtenidos. En el informe de gestión de la DTPM del mismo año, se atribuye este incremento a la implantación de ciertas medidas para disminuir los índices de evasión. Estos aspectos de medición resultan importantes, puesto que el grado de evasión, ya sea en periodos o en zonas, va a contribuir en gran medida a los resultados obtenidos sobre la congestión de paraderos. Asimismo, muestra una variación notable a escala mensual, por lo que se podría atribuir de manera similar esta variación a la congestión expuesta en paraderos.

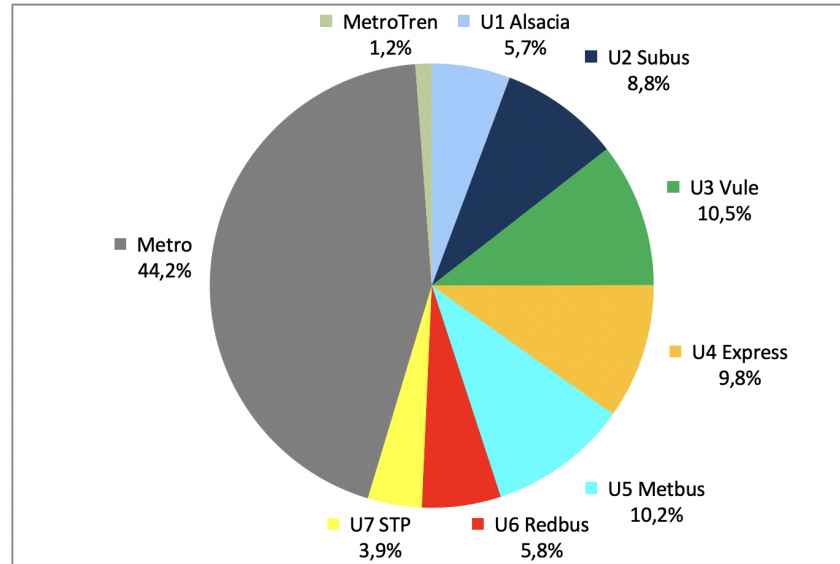


Figura 9: Participación por empresa en el total de transacciones en 2018

El porcentaje de transacciones por empresas concesionarias de la figura 9, muestra la capacidad de poder asignar las coordenadas de georreferencia a cada empresa, y consecuentemente también a los recorridos respectivos de cada ruta o tramo del servicio.

En la figura 10 se observa la evolución mensual de la evasión al pasar los años. Resulta esencial poder analizar e inferir el grado de evasión no solo a nivel temporal, si no también determinar las zonas y rutas afectadas por esto, puesto que una incorrecta asignación de esta evasión en una ruta, puede invisibilizar el grado de congestión real de algún paradero y por ende afecta la calidad del resultado final.

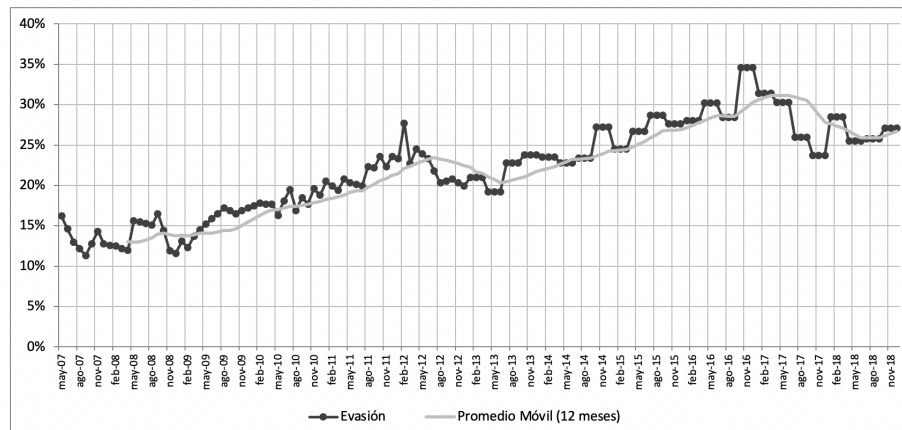


Figura 10: Evolución de la evasión 2007-2018

## 4. Objetivos y medidas de desempeño

### 4.1. Objetivo General

El objetivo principal de este proyecto es otorgar una herramienta a la DTPM que permita facilitar la toma de decisión sobre la gestión de rutas. Esto por medio de una interfaz que permita identificar si se está o no cumplimiento los indicadores de frecuencia y regularidad de los buses en las rutas de la red de Transantiago.

### 4.2. Objetivo Específico

El objetivo del proyecto se logra si se cumplen los siguientes puntos:

#### 1. Identificar como calcular las métricas de frecuencia y regularidad

- Identificar los factores que involucran al cálculo de frecuencia.
- Identificar los factores que involucran al cálculo de regularidad.
- Modelar el cálculo de frecuencia.
- Modelar el cálculo de regularidad.

#### 2. Procesar los datos respecto a los modelos de frecuencia y regularidad

- Separar los datos respecto a ruta y buses dentro de las rutas.
- Dentro de cada caso de ruta separar los datos por intervalos de tiempo de una hora.
- Procesar los datos de frecuencia respecto al modelamiento del punto 1.
- Procesar los datos de regularidad respecto al modelamiento del punto 1.

#### 3. Visualizar los resultados en una interfaz de mapas

- Mostrar los indicadores por cada ruta

### 4.3. Medidas de Desempeño

## 5. Alternativas de Solución

### 5.1. Estructura de Datos

En este apartado se describen los datos que se utilizarán para la creación de la solución. Estos datos se obtienen respecto a los datos recopilados por CityLog (explicado en mayor detalle en Datos del Problema) y son de interés su uso respecto a la información obtenida sobre buses y paraderos. Sus estructuras son de la siguiente manera:

- **Buses:** en el caso de los buses se identifica que es de interés tener información sobre a qué ruta corresponde el bus, cuál es la patente del bus (identificador único) y la posición del bus en el tiempo. (este valor se obtiene cada 30 segundos)
- **Paraderos:** en el caso de los paraderos se identifica que es de interés tener información de a qué rutas se encuentra asociado y la posición geográfica en la que se encuentra en el mapa.

Para identificar el impacto de los indicadores en los tramos de una ruta, resulta necesario unir ambas tablas de datos en una vista en la base de datos, para lograr realizar los cálculos de distancias respectivos.

### 5.2. Modelamiento y procesamiento de datos

#### 5.2.1. Indexación de buses a paraderos

Lo primero que se debe realizar es filtrar los datos en donde los buses pasan por los paraderos. Para ello se calcula la distancia entre cada muestra de un bus con respecto a cada paradero de su ruta correspondiente. Para cada agrupación se determinan las muestras más importantes, es decir, aquellas que correspondan al paradero más cercano en dicho momento. Luego se indexan los valores según el tiempo y las distancias mínimas, para así solo tener las muestras de los buses correspondientes al paradero al cual se encuentre más cercano.



### 5.2.2. Separación de rutas

Después de haber obtenido solo las muestras en donde los buses se encuentran en un paradero, se deben de crear dos tipos de agrupaciones: uno dado por la separación de datos por rutas y otro en donde a la agrupación de rutas se debe de poder identificar por separado cada uno de los buses.

### 5.2.3. Procesamiento de Frecuencia

Para empezar es necesario agrupar los datos con respecto a los paraderos de la ruta correspondiente que se esté analizando. Luego se segmentan los datos por cada hora del día, y se calcula el número de diferentes micros que han pasado por dicho paradero en el bloque horario. Finalmente, para cada paradero se calcula la media, y los valores del vector resultante se ingresan a una función de clasificación que normalice el cumplimiento de la frecuencia.

### 5.2.4. Procesamiento de Regularidad

El objetivo consiste en obtener un vector que contenga la diferencia de tiempo entre cada micro que pase por paradero. Para ello, primero se agrupan los datos según la matrícula/id del bus con la finalidad de obtener la trayectoria del recorrido de cada máquina, y consecutivamente se segmentan dichos datos por cada hora del día. Dado que para determinar la regularidad la marca de tiempo es importante, se debe agrupar este conjunto acotado de datos por paraderos para así obtener la muestra con la distancia mínima, donde dicho tiempo ha de ser cuando tiene la posibilidad de recoger pasajeros en un paradero.

Con todas las muestras exactas del instante que un bus atraviesa un paradero, se procede a agrupar nuevamente los datos por estos y ordenarlos según sus marcas de tiempo. Con esto es posible determinar los instantes que una micro pasa por una parada, y calcular la diferencia con el tiempo cuando llegue la siguiente. A partir de ello se calcula el promedio de estas diferencias por cada paradero, considerando los bloques horarios del sistema de transporte público, o sea, horario bajo, valle y punta.

## 5.3. Herramientas Utilizadas

Para la solución implementada se pensaron distintas herramientas que facilitan lograr el objetivo planteado para este proyecto.

### 5.3.1. Apache Spark

Spark es una herramienta que permite el procesamiento de datos por medio de la división de los datos en lotes. Se utilizará esta herramienta para poder agrupar los datos respecto a los puntos mencionados en el apartado 5.2. y también para poder procesar de forma más rápida los grandes volúmenes de datos, ya que se está trabajando con una muestra de 2.552.400.068 datos.

### 5.3.2. Postgres - PostGis

Postgres es un sistema orientado a la administración de bases de datos tipo SQL. Se pensó en esta herramienta principalmente por la ubicación de los datos utilizados para el proyecto, los cuales se ubican en un servidor dentro del sistema Postgres y, aparte, los datos tienen atributos en ciertas tablas que describen variables geométricas creadas con ubicaciones en el mapa con latitud, longitud y tipo de figura específicas. Las variables geométricas solo pueden ser especificadas como atributo tipo *geometry* que solo pueden ser usados al instalar el *plugin* de PostGis y añadiendo las extensiones necesarias dentro de la base de datos a utilizar. Utilizar variables geométricas es de mucha importancia para el proyecto, ya que, mediante una herramienta de visualización que permita este tipo de atributo, se podrán observar mejor los datos en un mapa y crear vistas que permitan la obtención de los distintos resultados de los análisis realizados en este trabajo.

### 5.3.3. Qgis

Qgis es una herramienta que permite usar información geográfica para visualizarla dentro de un mapa. Esta aplicación *open source* se puede integrar con Postgres para lograr visualizar los datos *geometry* que puede tener una tabla.

Para este proyecto, el uso de esta herramienta es vital, debido a que este será el apartado visual de todo el análisis y trabajo que se haya hecho con los datos empleados.

## Referencias

- [1] Centro de Innovación de Transporte y Logística. *Pasajeros*. [http://citylog.cl/cantidad\\_de\\_pasajeros/](http://citylog.cl/cantidad_de_pasajeros/). Sep. de 2022.
- [2] Fiscalización de Transporte. *Evasión*. <http://www.fiscalizacion.cl/indice-de-evasion-de-pago-de-tarifa-en-transantiago/>. Sep. de 2022.
- [3] Fiscalización de Transporte. *Informes*. <https://www.dtpm.cl/index.php/documentos/informes-de-gestion>. Sep. de 2022.
- [4] Fiscalización de Transporte. *Ranking de Calidad*. <https://www.dtpm.cl/index.php/documentos/ranking-calidad-de-servicio>. Sep. de 2022.
- [5] Subsecretaria de Transporte. *Historia*. <https://www.subtrans.gob.cl/nosotros/>. Sep. de 2022.
- [6] Directorio de Transporte Público Metropolitano. *Funciones*. <https://www.dtpm.cl/index.php/homepage/directorio-de-transporte-publico>. Sep. de 2022.
- [7] Ministerio de Transportes y Telecomunicaciones. *Reseña institucional*. <https://www.mtt.gob.cl/resenainstitucional>. Sep. de 2022.