

UNIVERSIDAD DIEGO PORTALES

FACULTAD DE INGENIERÍA Y CIENCIAS
ESCUELA DE INFORMÁTICA Y TELECOMUNICACIONES



Data Science

Informe N°3: Entrega Final del Proyecto

Integrantes:

Lucas Almonacid
Benjamín Fernández
Felipe Condore
Pamela Saldías

Profesor:

Karol Suchan

Fecha de entrega:

18 de Diciembre de 2022

Índice

1. Usuario	1
2. Descripción de la Problemática	2
3. Datos del Problema	3
4. Objetivos	6
4.1. Objetivo General	6
4.2. Objetivo Específico	6
5. Alternativas de Solución	7
5.1. Estructura de Datos	7
5.2. Modelamiento y procesamiento de datos	7
5.2.1. Indexación de buses a paraderos	7
5.2.2. Separación de rutas	7
5.2.3. Procesamiento de Frecuencia	7
5.2.4. Procesamiento de Regularidad	7
5.2.5. Estándares de calidad	8
5.3. Herramientas Utilizadas	8
5.3.1. Apache Spark	8
5.3.2. Postgres - PostGIS	8
5.3.3. QGIS	8
6. Solución aplicada	8
6.1. Implementación del modelo propuesto	9
6.2. Visualización del proyecto	9
7. Resultados y análisis del modelo aplicado	13
7.1. Análisis de sensibilidad	15
8. Conclusión	16
9. Recomendaciones	16

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.”[8].

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”[6].

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:

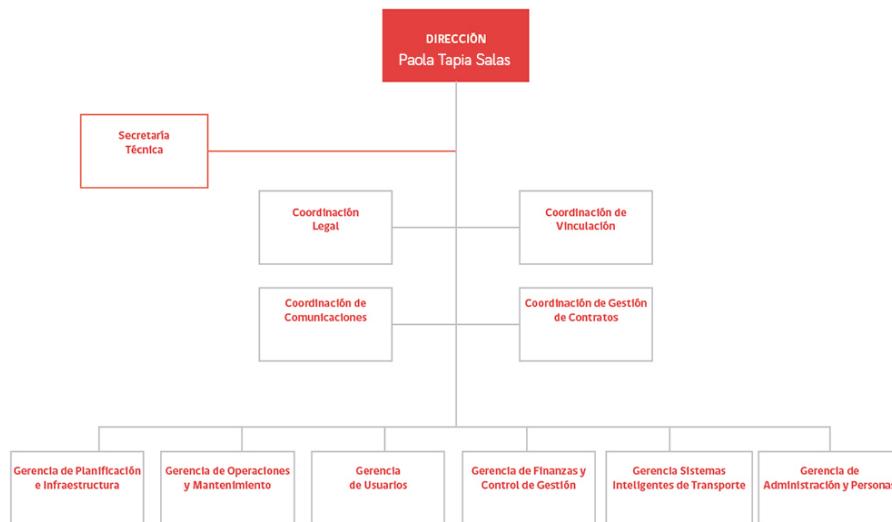


Figura 1: Organigrama DTPM del año 2022.

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”[7].
- “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”[7].
- “Coordinar los procesos de negociación que pudieran requerirse en el marco de las acciones para dar cumplimiento a las misiones encomendadas”[7].

- “Supervisar los contratos, participar en instancias de estudios, análisis y mejoramiento del transporte público”[7].
- “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”[7].
- “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”[7].
- “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”[7].
- “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”[7].
- “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”[7].
- “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”[7].

En la actualidad la DTPM a través del departamento de Operaciones y Mantenimiento, el cual es el encargado de controlar y monitorear la operación diaria de los buses, además de supervisar el buen desempeño del transporte público, genera trimestralmente rankings de calidad de servicio brindado por las concesionarias respecto a los estándares esperados, para ello la DTPM describe su método utilizado, como un cálculo mostrando el indicador de frecuencia y regularidad para un trimestre, en donde los resultados de cumplimiento del indicador de frecuencia, corresponden a la razón entre el número de salidas ejecutadas y salidas programadas en el periodo trimestral completo, mientras que los resultados de cumplimiento del indicador de regularidad se obtiene de la razón entre el número de intervalos entre buses que cumplen lo establecido en el Programa de Operación y el número de intervalos observados en el periodo trimestral completo. Estos resultados son dados en un porcentaje del servicio completo de cada una de las concesionarias y debe tener un cumplimiento de 90 % en el caso de la frecuencia y un 80 % en el caso de la regularidad. Por lo que, en la actualidad solo existe una visión global del cumplimiento de estas métricas y no se puede tener una visión más cercana de cuáles son los puntos en específico de cada ruta en donde puede haber fallas.

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 ven 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.

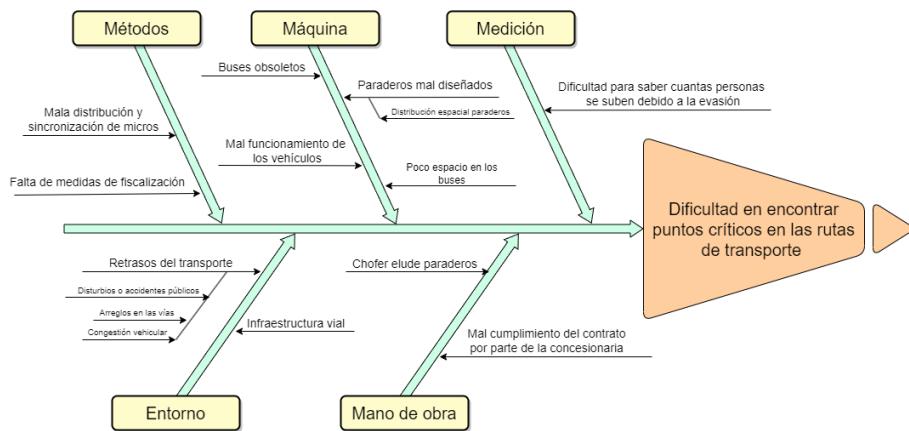


Figura 2: Diagrama Ishikawa

En la figura adjunta se puede observar un “Diagrama de Ishikawa”, este describe las causas que inciden en el problema. En este contexto 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 Pùblicos Metropolitano (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 Metropolitano. La frecuencia de cada empresa concesionaria mide las salidas de buses efectivamente realizadas desde el punto de inicio de cada recorrido.

- **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 Metropolitano. 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.

De esta manera, se observa que DTPM tiene una visión mas global a la hora de evaluar la calidad de servicio que se brinda. Ya que no se especifica en su calculo información sobre rutas, paraderos o buses.

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, permiten 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)[2], 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. Las gráficas utilizadas 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.[3][4][5]

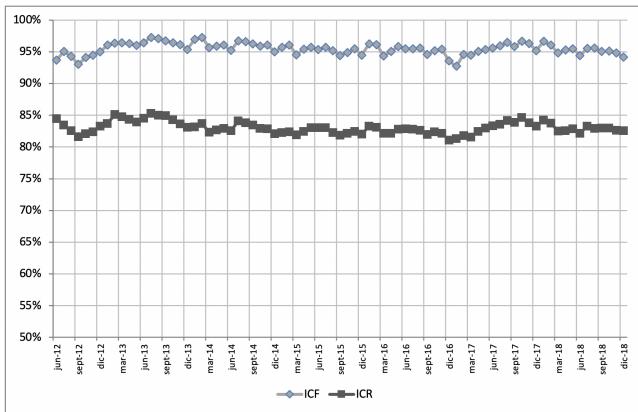


Figura 3: Evolución de los indicadores de frecuencia (ICF) y regularidad (ICR) 2012 – 2018

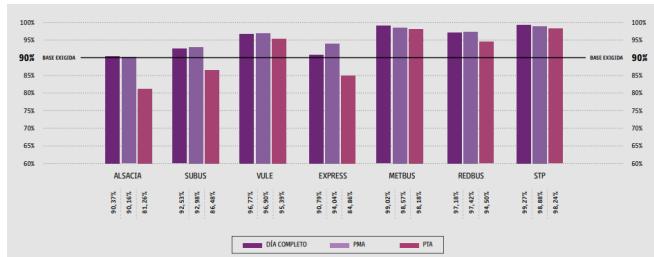


Figura 4: Indicadores de frecuencia de Abril a Junio 2018

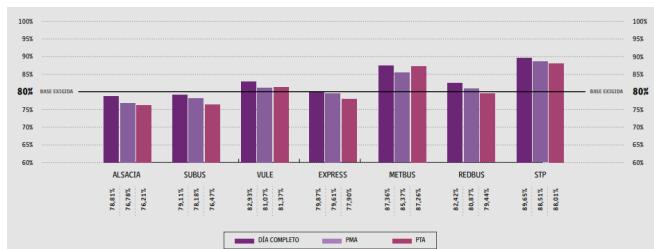


Figura 5: Indicadores de regularidad de Abril a Junio 2018

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 los tramos o recorridos asignados a cada concesionaria y la congestión de paraderos ubicados en estas mismas, siendo este problema posiblemente causado por ubicaciones específicas como colegios, universidades, zona de oficinas, zonas en mal estado o calles por la cual es difícil circular que pueden producir dificultades para afrontar correctamente esas rutas a horas específicas.

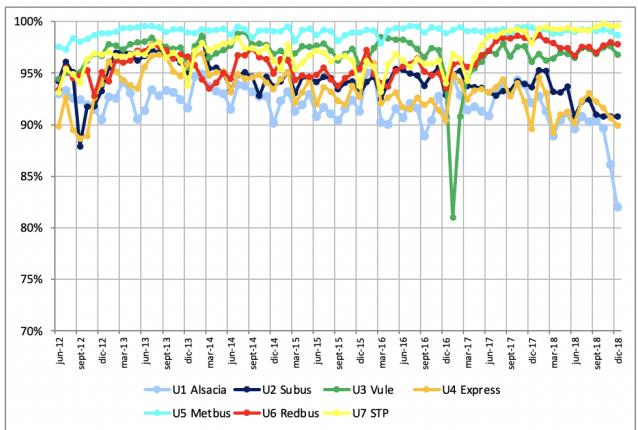


Figura 6: Evolución mensual del indicador de frecuencia, por Unidad de Negocio 2012 - 2018

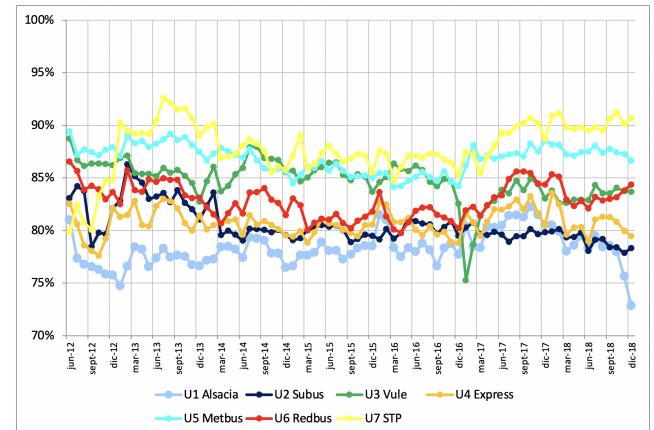


Figura 7: Evolución mensual del indicador de regularidad, por Unidad de Negocio 2012 - 2018

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. Con estas gráficas también se puede denotar comportamientos de cada una de las concesionarias, como el

hecho de que *STP* va mejorando su frecuencia en el paso del tiempo, junto con mantener niveles más altos que el promedio de concesionarias cuando se ve el tópico de la regularidad, en contraste con *Alsacia*, que en las dos figuras se puede denotar que suele tener estas métricas bajas, indicando menor cumplimiento en el servicio que está brindando.

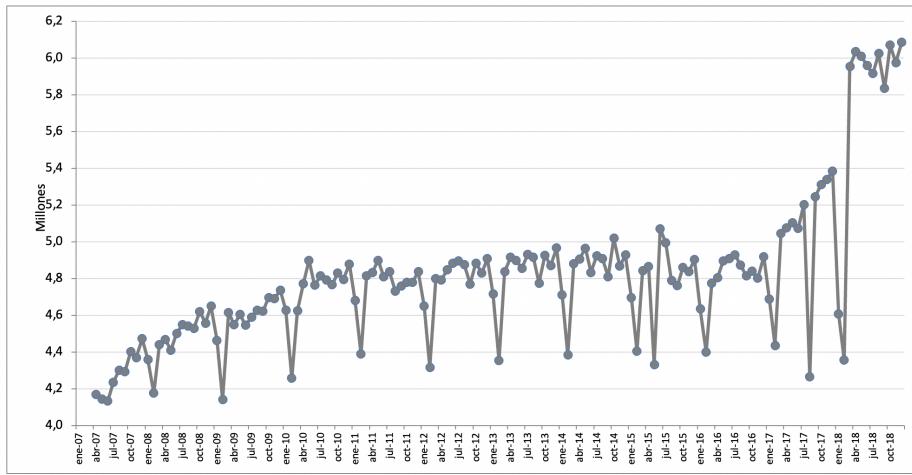


Figura 8: Evolución del promedio mensual de usuarios del sistema 2007 - 2018

En la figura ?? 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.

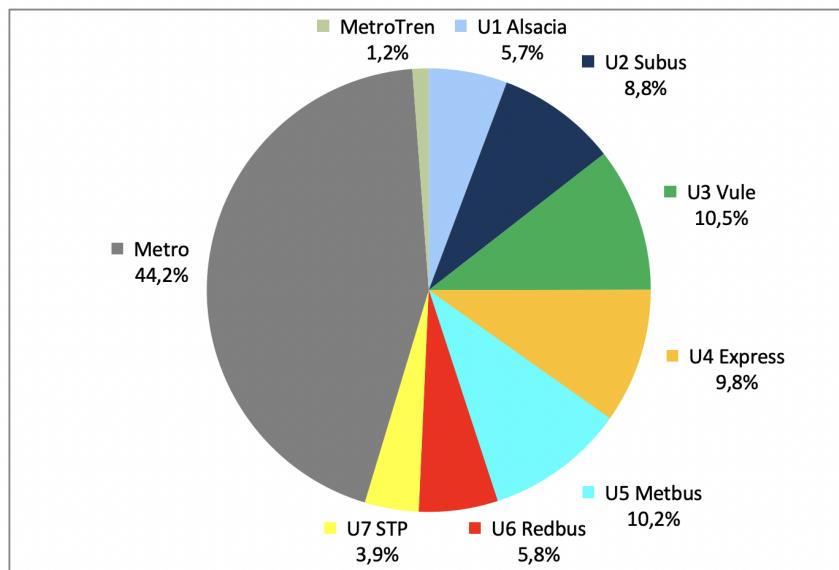


Figura 9: Participación por empresa en el total de transacciones en 2018

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, extraída de [3], se observa la evolución mensual de la evasión al pasar los años. Resulta esencial poder analizar e inferir el grado de evasión, puesto que este problema altera la cantidad de usuarios real que ocupa el sistema y, por ende, invisibiliza una de las posibles razones por la cual un bus puede demorar tiempo extra en algún paradero.

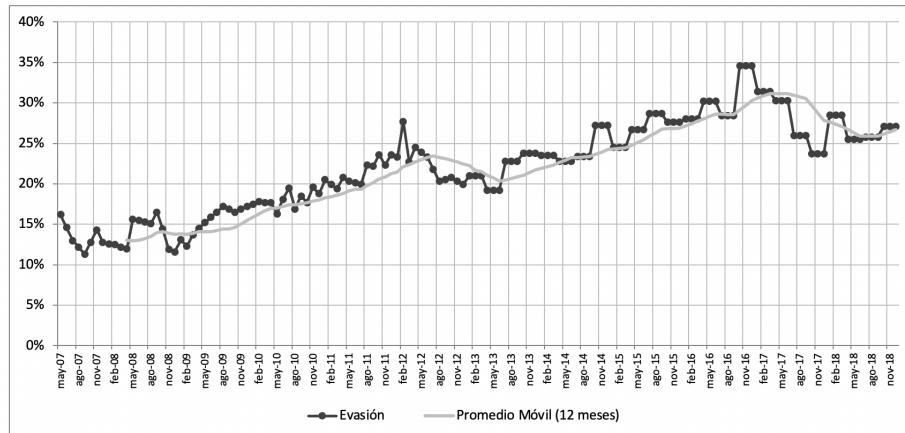


Figura 10: Evolución de la evasión 2007-2018

4. Objetivos

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

5. Alternativas de Solución

5.1. Estructura de Datos

En este apartado se describen los datos que se utilizaran 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 cada bloque horario. Finalmente, para cada uno de los paraderos se saca la media de los resultados obtenidos en todas las franjas de tiempo evaluadas y respecto a estos resultados se hace una evaluación en cada paradero de si esta cumpliendo o no con los estándares esperados de 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 un paradero. Para ello, primero se agrupan los datos según la matrícula 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 para cada paradero el promedio de estas diferencias, considerando los bloques horarios del sistema de transporte público, o sea, horario bajo, valle y punta. De esta manera se puede evaluar para cada paradero si están cumplimiento con los estándares esperados de regularidad.

5.2.5. Estándares de calidad

La medición del cumplimiento de frecuencia y regularidad estarán dados por el cumplimiento de una cantidad promedio mínima de 6 a 7 buses que deben pasar por un paradero en una hora para el caso de la frecuencia y una cantidad promedio de 600-700 segundos máximos que puede pasar entre un bus y otro en un paradero. De esta manera, se evalúa respecto a cada paradero de la ruta si cumplen o no con este estándar dándole una calificación de cumple ($x \geq 7$), cerca de cumplir ($x = 6$) y no cumple ($x \leq 5$) en el caso de la frecuencia y de cumplir ($x < 700$) y no cumplir en el caso de la regularidad ($x > 700$).

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.

6. Solución aplicada

La plataforma que se ocupa para visualizar los resultados es QGIS, como se puede ver en la figura 11, esta herramienta está conectada a una base de datos Postgres, con el *plugin* de PostGIS, donde consigue todos los datos de las rutas y paraderos existentes. Se realizó un archivo de QGIS donde ya se tienen creadas las capas de rutas, paraderos y algunos factores de la zona metropolitana como son las comunas de Santiago y los colegios dentro de la región para visibilizar mejor razones de posibles congestiones dentro de la región.

Los resultados obtenidos por parte de Spark deben ser solicitados por el cliente y este debe de brindar los archivos que se deben de analizar a esta herramienta, como los archivos que se deben de procesar y las rutas que quiere visualizar. Cabe decir que el usuario del sistema no debe instalarse Spark o tener la base de datos en su equipo, estos componentes estarán almacenados en un servidor aparte, de tal forma que el usuario solo debe de tener instalado QGIS y tener las credenciales necesarias para obtener los datos de las capas ya insertadas por parte de Postgres.

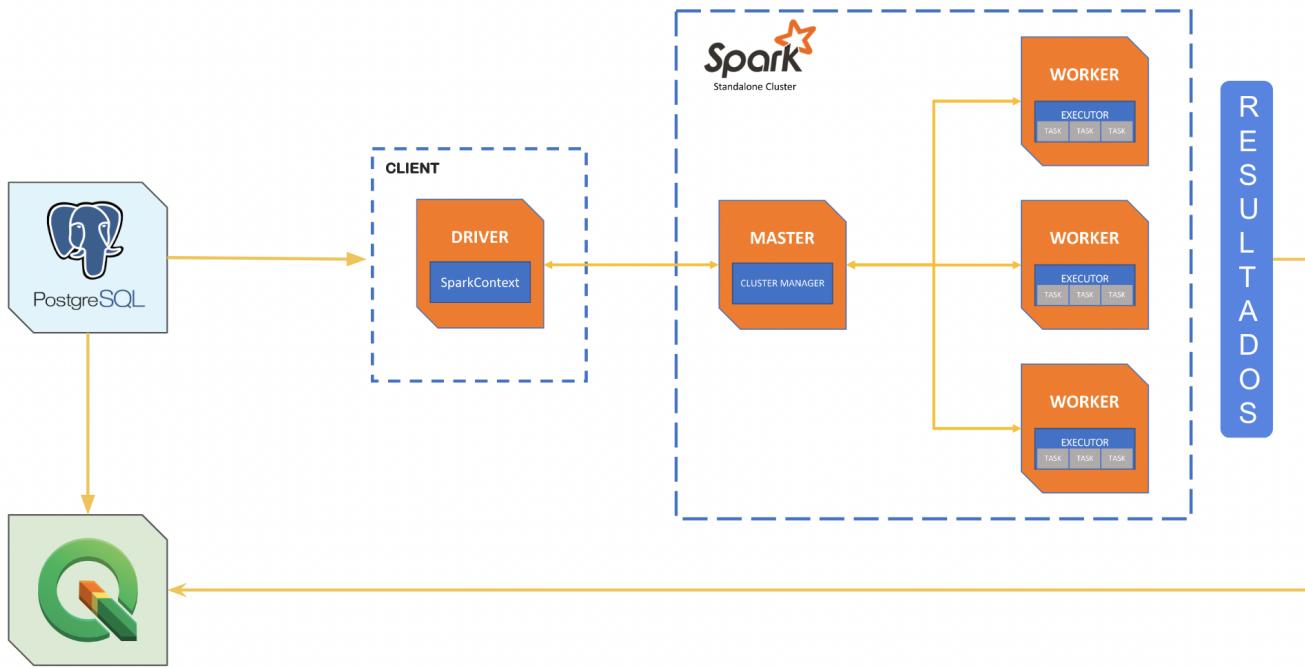


Figura 11: Arquitectura con las tecnologías

6.1. Implementación del modelo propuesto

Para implementar el modelo se lleva a cabo la generación de la arquitectura presente en la figura ??, como una serie de procesos por lotes según cada una de las rutas correspondientes. La primera parte del modelo correspondiente a la generación de datos relevantes, se logra a través de la indexación de los valores georreferenciados de los buses con los paraderos más cercanos. Para lograr esto el modelo recibe como entrada documentos CSV, luego las API de pandas de PySpark permite indexar la información en dataframes que pueden ser agrupados y tratados en lotes por separado según su ruta.

La segunda parte del modelo propuesto simplemente aplica los procesamiento descritos anteriormente para los índices de regularidad y frecuencia respectivamente sobre cada lote obtenido del paso anterior. Cada una de estas iteraciones tiene la tarea de aplicar una calificación sobre las métricas correspondiente a todos los paraderos presentes en el dataframe que se está procesando, a la vez que para no exceder la capacidad de memoria utilizada por el sistema procede a guardar la información de cada ruta correspondiente en el disco de la máquina, antes de continuar con el siguiente lote.

6.2. Visualización del proyecto

Con lo que respecta a la visualización, el cliente debe de descargar un código escrito en Python, el cual, mediante decir la ruta del archivo, al ejecutarse, dentro de QGIS, debe crear las capas que indiquen la regularidad y frecuencia de los datos que el usuario le entregó a Spark.

En primera instancia, el cliente al iniciar QGIS debe exportar el proyecto, de tal forma que siempre en el inicio del programa este pueda acceder a una visualización predeterminada. En la figura 12 se puede ver la pantalla de inicio de QGIS, la visualización que se tiene preparada para el cliente es el proyecto llamado “Proyecto”.

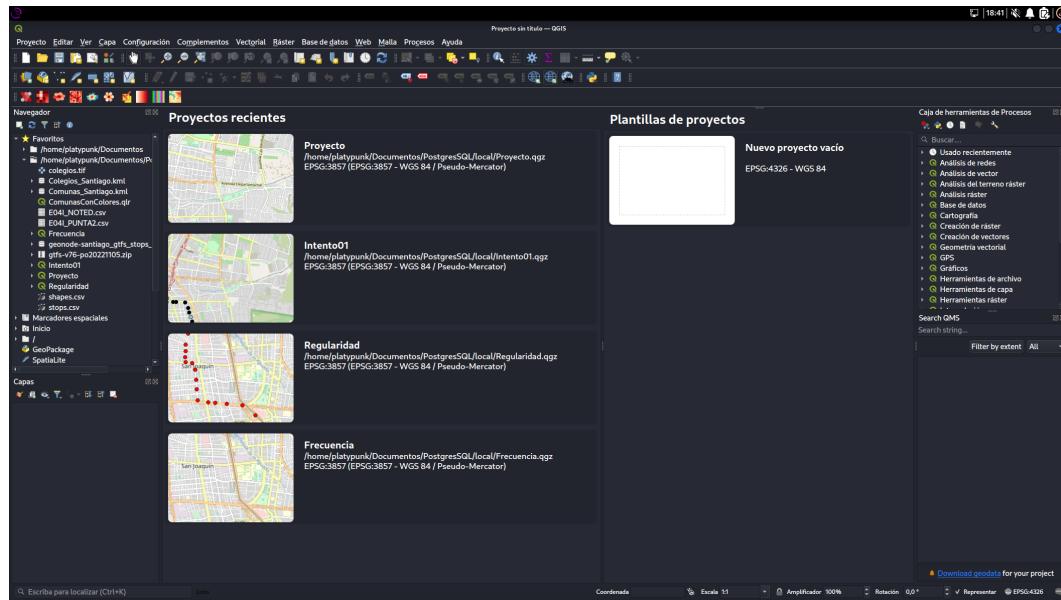


Figura 12: Vista inicio de QGIS

Cuando el usuario esté en el proyecto, este deberá dar las credenciales de la base de datos, estas se piden mediante una ventana al seleccionar el proyecto, la cual debe ser similar a la que se muestra en la figura 13

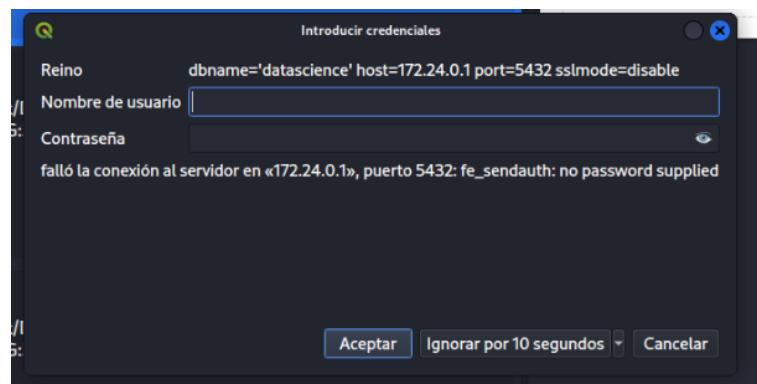


Figura 13: Credenciales de la base de datos

La base de datos en el proyecto debe estar en un servidor externo, para que así el usuario no tengo que tener almacenado las rutas y los paraderos en su equipo. La conexión con este elemento se logra mediante la configuración de todos los datos necesarios, lo cual se puede ver en la figura 14. Ya que el usuario estaría utilizando un proyecto ya creado de QGIS, la conexión ya debería de ir establecida, por lo tanto, el usuario no debería de cambiar la IP, el puerto u otros datos dados para establecer la conexión.

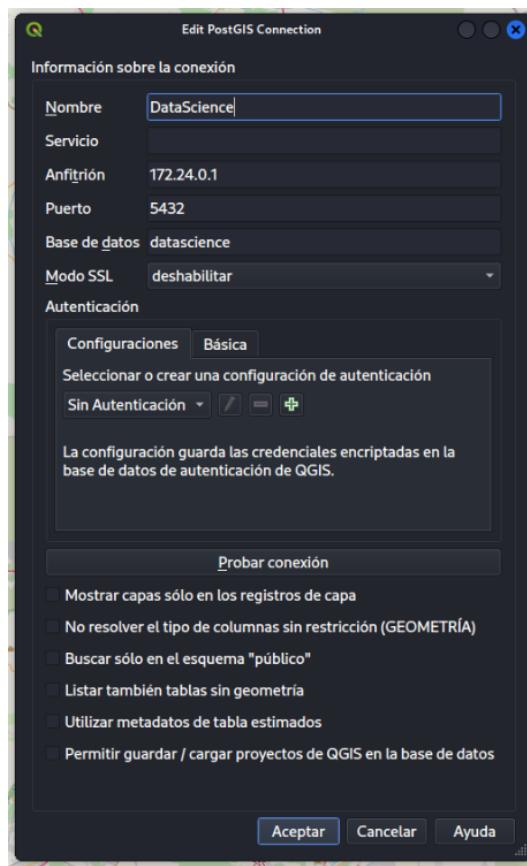


Figura 14: Parámetros necesarios para la conexión de la base de datos

Cuando el usuario ya ingrese al proyecto, podrá visibilizar un mapa extraído de la OSM (*Open Street Map*). El mapa será la base donde se muestren el resto de capas, ya que muestra todas las calles de santiago, ver figura 15.

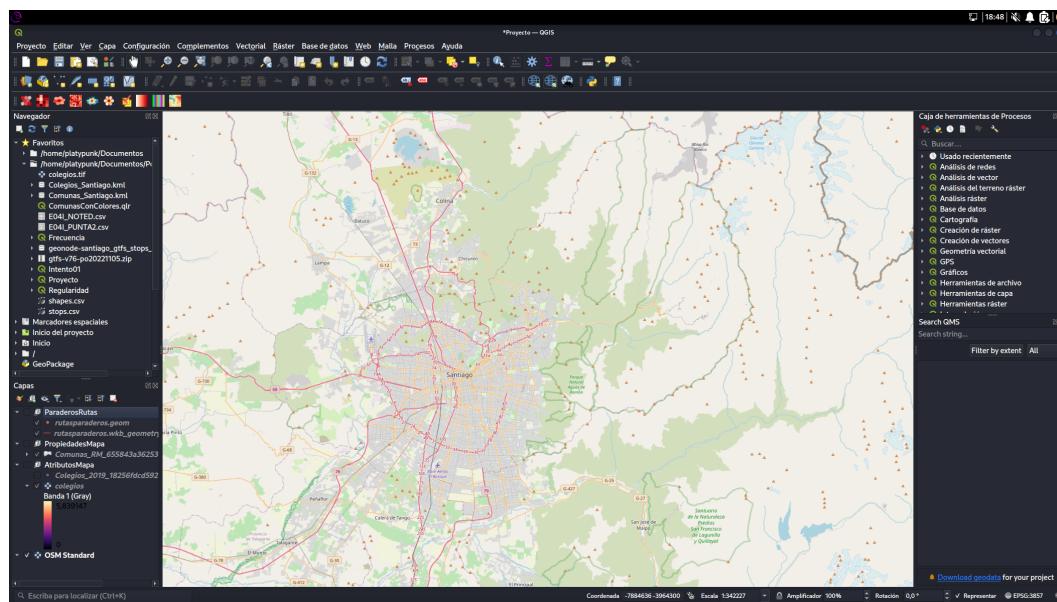


Figura 15: Mapa base utilizando para el proyecto

Las capas que vengan por defecto están ordenadas dentro de carpetas, como se puede ver en la figura 16

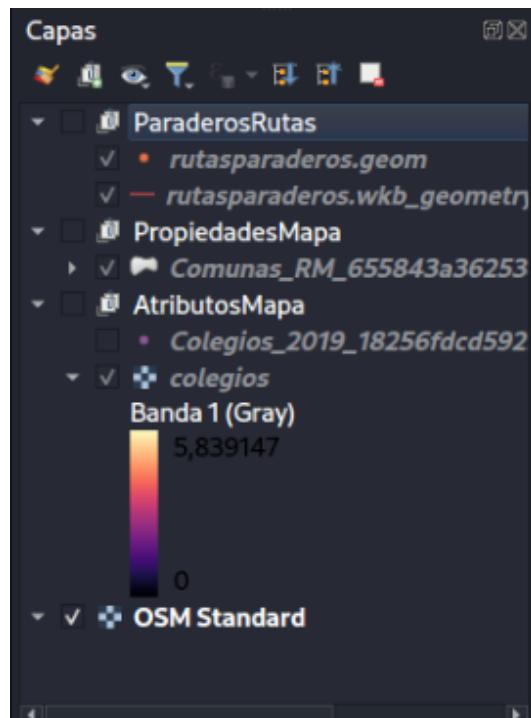


Figura 16: Orden de las capas ya estandarizadas

Algunas de las capas que están por defecto se pueden ver en las figuras 17, 18 y 19.

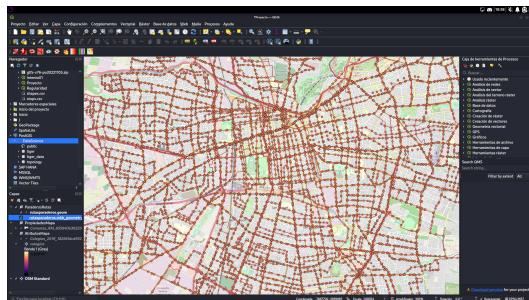


Figura 17: Capa que muestra los paraderos y rutas de santiago

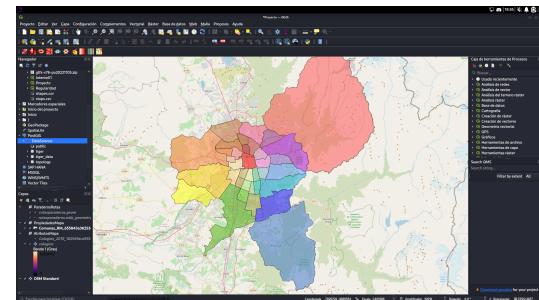


Figura 18: Capa de las comunas de santiago

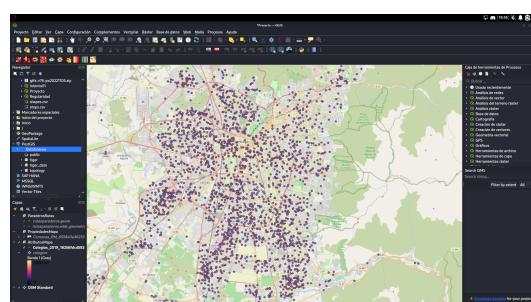


Figura 19: Capa que muestra los colegios de santiago

Para que el usuario pueda visualizar los resultados obtenidos después de aplicar el modelo, debe ejecutar el código Python dentro de QGIS, este le generara una capa con propiedades estándar que le permitan visualizar las

métricas de frecuencia y regularidad según horario.

Con respecto a la visualización de los resultados, al importar los CSV mediante el código de python se generarán 3 capas. La capa que indica frecuencia varía los puntos en tres colores, siendo el rojo el no cumplimiento, el amarillo estar cerca de cumplir la frecuencia y el verde como que en ese punto se logra cumplir la frecuencia. Para el caso de la regularidad se generan dos capas que varían sus puntos entre azul y blanco, siendo el color azul la demostración de que en ese punto no se cumple y el color blanco que si hay cumplimiento en ese punto. Los puntos marcados en las capas son paraderos de las rutas, ya que estos son los puntos de interés donde se evalúan las métricas y los puntos de regularidad y frecuencia poseen distintos tamaños para lograr visualizar estos dos parámetros al mismo tiempo. Con respecto a las dos capas que se generan de regularidad, en una se toma en cuenta todo el tiempo del día para hacer el cálculo, mientras que la otra capa, considera distintos rangos de horarios diferenciados por las horas bajo, punta y valle en sus dos etapas en el día. En la figura 20 se pueden ver los horarios correspondientes a los rangos antes dichos además de sus tarifas, la información de esta tabla fue extraída del sitio [1]

bip! portador	bip! personalizada	Bloqueo y desbloqueo bip! personalizada
Centros bip! y estaciones de Metro \$1.550 (Sólo la tarjeta)	Oficinas de Servicio al Cliente \$2.800 (Sólo la tarjeta)	Oficinas de Servicio al Cliente \$300
Lunes a Viernes Metro y Tren		
Horario Bajo \$640 06:00 a 06:59:59 20:45 a 23:00 hrs.	Horario Valle \$720 09:00 a 17:59:59 20:00 a 20:44:59 hrs.	Horario Punta \$800 07:00 a 08:59:59 18:00 a 19:59:59 hrs.
Sábado, Domingo y festivos Metro y Tren		
Horario Valle \$720 Todo horario	Todo Horario \$700	Todo Horario + Buses, Metro y Tren \$230

Figura 20: Tarifas transporte publico

Mediante filtros que se pueden aplicar en QGIS el usuario podrá filtrar las zonas horarias, las rutas que quiera visualizar y algún paradero específico dentro de las capas que se le entregaran como resultado.

7. Resultados y análisis del modelo aplicado

Si se aplica el modelo en una ruta se podrá visualizar el siguiente resultado en QGIS (ver figura 21)

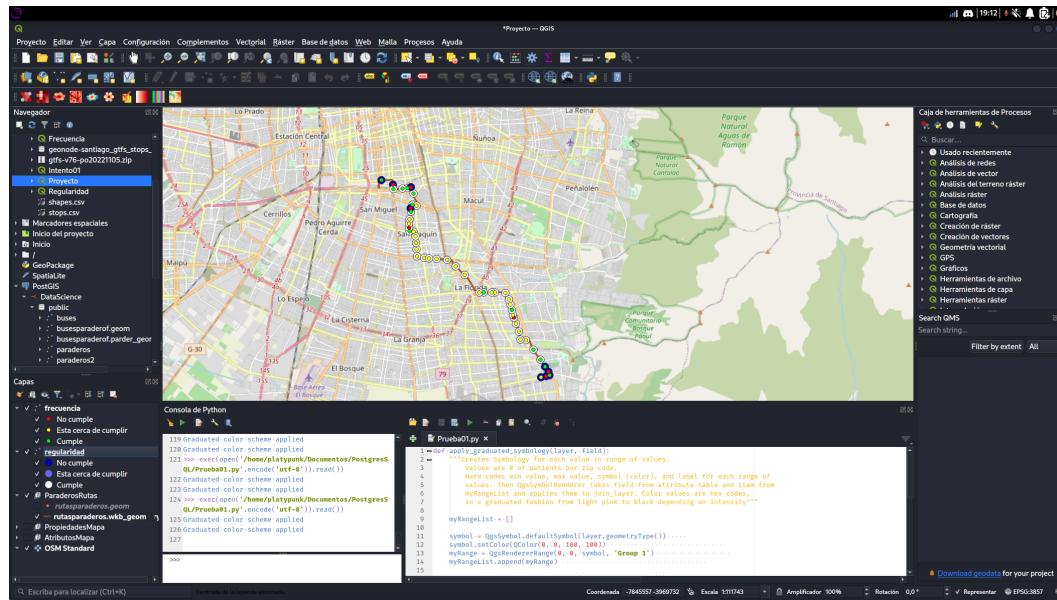


Figura 21: Resultado obtenido de la ruta E04 de ida

En la figura 21 se puede visualizar como el recorrido de la E04 de ida suele cumplir, en promedio, con su

regularidad en su trayecto de ida. Sí se observa la frecuencia, esta es más variable, visualizando así más puntos indicativos de estar cerca de cumplir. Podemos observar en la figura 22 que da la coincidencia que los puntos donde se ve peor regularidad son cercanos a los límites de las comunas de santiago.

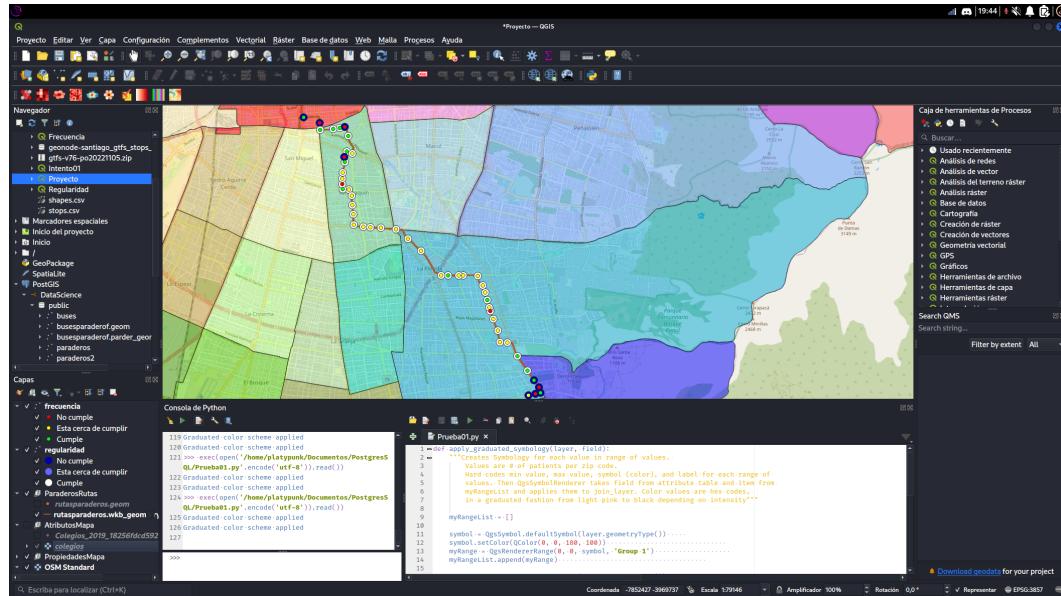


Figura 22: Resultado obtenido de la ruta E04 de ida

Otro caso que se puede analizar es cuando el usuario solicita la información de más de una ruta, para ello, cuando el usuario ya tengo los resultados de *Spark* y haber creado las capas, debe filtrar en las capas según las rutas que quiere visualizar.

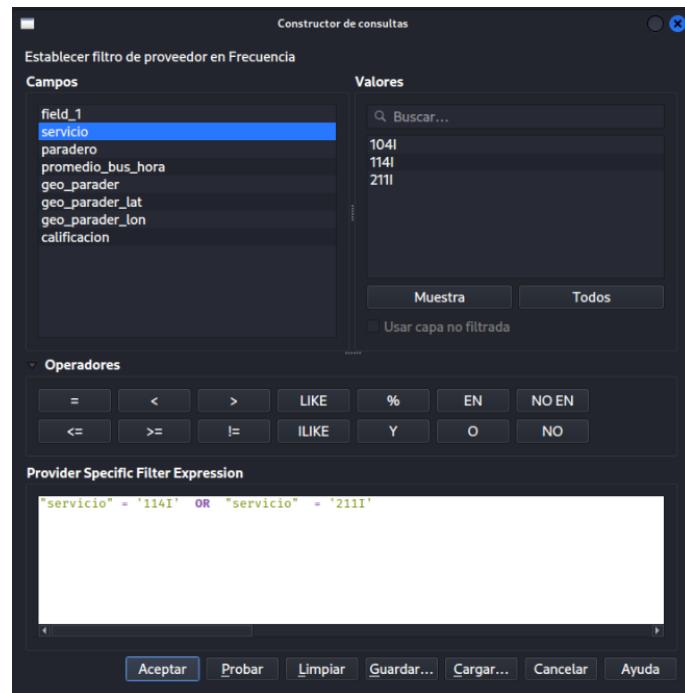


Figura 23: Filtro aplicado

En la figura 23 se puede ver como se generan los filtros para visualizar mejor la información que el usuario pide en las capas. Se puede ver que se tiene a disposición 3 rutas para visualizar y mediante el filtro se dice que rutas se querrán ver, en este caso, la 114 y la 211 en su trayecto de ida.

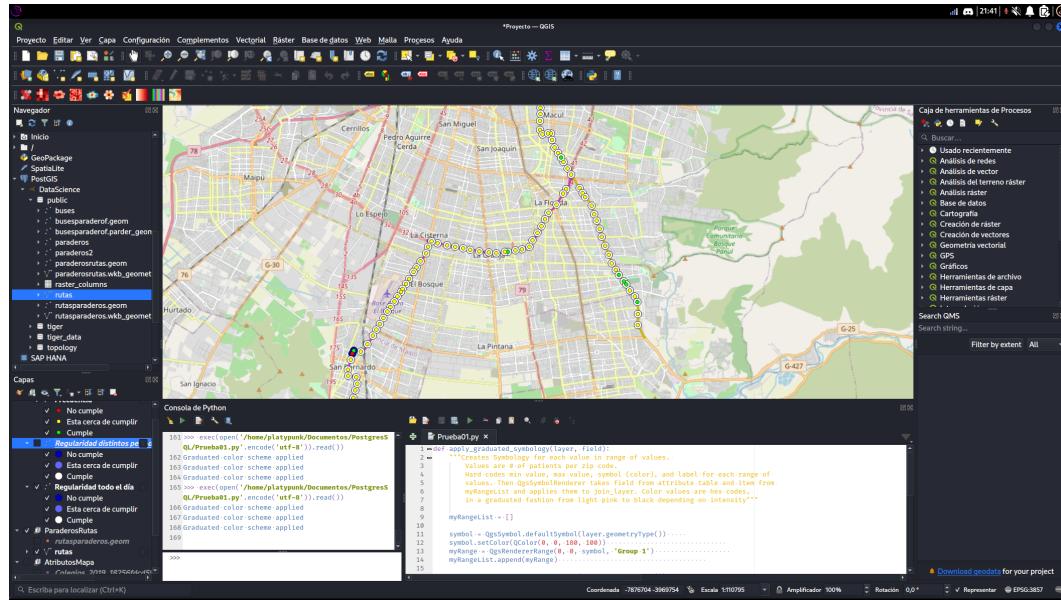


Figura 24: Ruta 114 y 211 tomando en cuenta todas las horas del día para calcular la regularidad

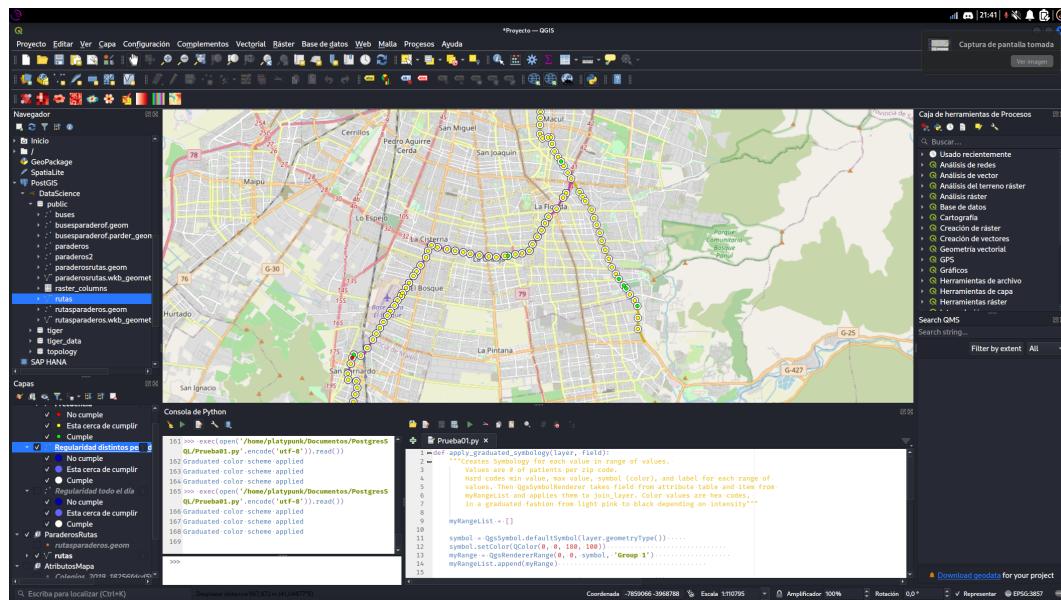


Figura 25: Ruta 114 y 211 tomando en la segunda hora punta del día para calcular la regularidad

En las figuras 24 y 25 se pueden ver las mismas dos rutas, pero tomando en cuenta distintas formas de calcular la regularidad. En la figura 24 se calcula la regularidad tomando en cuenta todas horas del día, mientras que en la figura 25 solo considera el horario punta que se genera en la tarde. Se puede observar que el rendimiento de las dos rutas es bastante bueno tomando, pero, cuando se calcula la regularidad de las rutas tomando en cuenta todo el día, la ruta de la 211 tiene problemas en un paradero, siendo este un paradero que también no cumple con su regularidad, por lo tanto, se puede decir que ese es un punto crítico de la ruta. El hecho de que en la ruta de la 211 no se vea ni un punto que no cumpla con su regularidad, no implica que existan otras horas donde ese punto se vea afectado.

7.1. Análisis de sensibilidad

Dado las muestras anteriores, se puede determinar que uno de los principales factores que influyen o muestran un cambio de tendencia en los resultados corresponden al análisis con respecto al tiempo. Para las métricas como

la regularidad por ejemplos se aplican distintas funciones para su estimación en el modelo, según el horario que le corresponda, sea horario valle, punto o bajo. Esto evidentemente marca distintas tendencias como se observa previamente en las figuras 25 y 24, donde bajo ciertos horarios el incumplimiento de los servicios puede verse mayormente comprometido, y al agrupar los resultados según ventanas de tiempo más grandes puede ignorarse ciertos puntos críticos.

8. Conclusión

En este trabajo se describe un problema de la DTPM, que es la poca visibilidad que tienen las medidas de desempeño, tales como frecuencia y regularidad, con respecto a cada concesionaria y las rutas asociadas. Se propuso un modelo que logra calcular y visibilizar cada ruta y las falencias con respecto a las métricas antes mencionadas, para el análisis de estos puntos y, al finalizar el sistema propuesto, se puede concluir que el programa logra visualizar las falencias de desempeño de cada ruta enfocándose en los paraderos como puntos críticos a observar y analizar. En los resultados del modelo, gracias a como se presentan los puntos, se puede ver los puntos que más problemas causan a las rutas y, mediante la inserción de capas como comunas o colegios, se puede buscar razones que evoquen esta condición.

9. Recomendaciones

Si bien, se logró crear un modelo que pudiera visualizar y poder calcular factores de interés para cada ruta en específico, se pueden destacar ciertos puntos que hubieran facilitado la aplicación del modelo y el análisis de resultado.

- En primer lugar, debido a la cantidad de información con la que se trabajó, el procesamiento del modelo fue uno de los mayores desafíos, teniendo que llegar al punto de realizar los cálculos de cada ruta en una plataforma en la nube. Por lo tanto, se recomienda dejar el servicio, que realice el procesamiento, en un equipo potente que permita un buen uso del modelo o pensar en la optimización de este.
- Otra complicación fue la aplicación de filtros con lo que respecta a la visualización de QGIS. Esto es debido a que los datos utilizados, para hacer asociaciones que permitieran un mejor manejo de los filtros, se necesitaban hacer muchas uniones entre tablas con abundantes filas que complicaban la unión de datos. Un ejemplo de esto es la tabla de rutas y transacciones, la única forma de saber las entidades de cada ruta era uniendo estas dos tablas, esto implicaba unir buses con transacciones, para después unirlo a los paraderos y, por último, unir las rutas. Este enlace terminaba creando tablas del tamaño de millones de filas, complicando así poder unir estos atributos y, por ende, se terminó por establecer límites a la visualización al no lograr unir ciertos atributos.
- Es ideal añadir más capas dentro de la aplicación de QGIS que permitan visibilizar más parámetros en el mapa, aparte de los colegios. Esto es muy recomendable debido a que permitiría un mejor análisis con respecto a las zonas problemáticas, añadiendo puntos como zona de oficinas, estaciones de metro, zonas conflictivas, etc.

Referencias

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