

Procesamiento y disponibilización de Datos de ADATRAP

Informe Final – OT04

Responsables: Mauricio Zúñiga, Felipe Hernández

26/09/2023

Tabla de Contenido

RESUMEN EJECUTIVO	4
1. Introducción	5
2. Descripción de tareas	6
2.1 Análisis de inconsistencias	6
2.1.1 Tabla de viajes	6
2.1.1.1 Variables nulas	6
2.1.1.2 Discordancia en tiempos de viaje	7
2.1.1.3 Problema tiempo fin viajes de más de 4 etapas	7
2.1.1.4 viajes sin bajada con tiempo de fin	7
2.1.1.5 Viajes sin servicio de subida	8
2.1.1.6 Comunas nulas	8
2.1.1.7 Comunas y zonas	8
2.1.1.8 Sin paradero inicial	8
2.1.1.9 Tipificación de modos	8
2.1.2 Tabla de etapas	9
2.1.2.1 Etapas sin servicio y con factor de expansión positivo	9
2.1.2.2 Servicios distintos	10
2.1.2.3 Servicios con pocas etapas	10
2.1.2.4 Paradas que no corresponden a servicio	10
2.1.2.5 Inconsistencia de suma de factor de expansión	10
2.2 Revisión de circunvalares	11
2.2.1 Situación anterior	11
2.2.2 Inspección situación actual	12
2.2.3 Tipificación de problemas	12
2.2.4 Trabajo futuro	16
2.3 Revisión de datos de perfiles	16
2.3.1 Análisis de carga final	16
2.3.2 Análisis de carga máxima	17
2.3.3 Análisis de subidas	18
2.3.4 Análisis de bajadas	19
2.3.5 Análisis de subidas y bajadas	20
2.3.6 Análisis de subidas y bajadas corregidas	20
2.3.7 Análisis de paraderos iniciales	21
2.3.8 Análisis de paraderos finales	22
2.4 Nuevo archivo gps offline	24
2.5 Nuevas funcionalidades en ficha de servicios	26
2.5.1 Incluir unidad de servicio asociada a cada ruta desplegada en el mapa	26
2.5.1.1 Implementación	26
2.5.2 Definir regla de colores para rutas y paradas	27
2.5.2.1 Implementación	27

2.5.3 Asignar orden de prioridad cuando se selecciona más de una ruta	28
2.5.3.1 Implementación	29
2.5.4 Agregar capas de salud y educación en el mapa	29
2.5.4.1 Implementación	29
2.5.5 Incluir trazados de variantes aunque no tengan paradas	33
2.5.5.1 Implementación	33
2.5.6 Agregar nombre de servicio/variante a cada ruta seleccionada	34
2.5.6.1 Implementación	34
2.6 Mejora filtro Perfil de carga	34
2.6.1 Implementación	35

Señores

Subsecretaría de Transporte

Presente

Srta. Loreto Bravo

De nuestra consideración:

Por medio de la presente, el Instituto de Sistemas Complejos de Ingeniería tiene el agrado de saludar y presentar el informe final de los resultados de la orden de trabajo número cuatro de la contratación del servicio de "Procesamiento y disponibilización de Datos de ADATRAP, Etapa III" por medio de la Resolución Ex. 459.

Quedamos a su disposición y atentos para sus consultas, la saluda cordialmente,

Mauricio Zúñiga G.

Responsable de la operación proyecto ADATRAP
ISCI



RESUMEN EJECUTIVO

En el contexto del servicio de "Procesamiento y disponibilización de Datos de ADATRAP" por medio de la Resolución Ex. 459, se ejecuta la orden de trabajo número cuatro por el Instituto de Sistemas Complejos de Ingeniería, centro de investigación y desarrollo de metodologías vinculadas a la integración de ingeniería en problemas reales y complejos.

Esta orden de trabajo tiene por objetivo el desarrollo de nuevas herramientas y el mejoramiento de las existentes, el periodo de trabajo que cubre esta orden dio inicio el 3 de julio de 2022 y finaliza con la entrega del informe final el 26 de septiembre de 2023. La orden de trabajo se realizó en dos etapas, este informe entrega los resultados de ambas etapas.

Durante esta orden de trabajo se realizaron un conjunto de análisis, mejoras y nuevos desarrollos, entre ellos están el análisis de diferentes casos de inconsistencias encontrados por DTPM tanto para la tabla de etapas como de viajes, la revisión de los casos de proyecciones anómalos, la revisión de datos de perfiles de carga, la integración de un nuevo espacio de almacenamiento para los datos gps offline y mejoras realizadas sobre el módulo de ficha de servicio.

1. Introducción

El ISCI es un centro de investigación y desarrollo de metodologías vinculadas a la integración de ingeniería en problemas reales y complejos, en sus diversas especialidades y en disciplinas afines y complementarias; tales como infraestructura, comportamiento humano, energía, medio ambiente, recursos naturales, transporte, ciudades y gestión.

El software de análisis de datos de transporte público ADATRAP fue desarrollado en el marco del proyecto FONDEF de Interés Público “Tecnología Avanzada para Ciudades del Futuro”. Este proyecto, liderado por académicos de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile, tiene como contrapartes beneficiarias de sus desarrollos a la Subsecretaría de Transportes, el Ministerio de Vivienda y Urbanismo, y el Cuerpo de Bomberos de Santiago. ADATRAP se basa en los desarrollos metodológicos de Cortés et al. (2011), Munizaga y Palma (2012), Devillaine et al. (2012), Gschwender et al. (2012), Amaya y Munizaga (2013), Núñez et al. (2013), Munizaga et al. (2014). La programación fue desarrollada por Mauricio Zúñiga, Richard Ibarra, Raúl Espinoza y Ramón Cruzat. El proyecto FONDEF es dirigido por Marcela Munizaga, investigadora encargada del área de análisis de datos de transporte público.

La Secretaría Ejecutiva del DTPM ha solicitado al Instituto de Sistema Complejos de Ingeniería, la contratación del servicio de “Procesamiento y disponibilización de Datos de ADATRAP, etapa III” por medio de la Resolución Ex. 459.

El servicio consiste en realizar el procesamiento, almacenamiento y disponibilización de información relevante para el análisis de datos del sistema de transporte de Santiago. Además, el servicio contempla horas para la investigación y desarrollo de mejoras o modificaciones en los datos de entrada del software que pueden ocurrir por cambios en el sistema, como, por ejemplo, nuevas líneas de metro o cambios en los formatos de los archivos.

Este informe contiene la descripción detallada de las nuevas herramientas desarrolladas y las mejoras realizadas a las herramientas existentes.

2. Descripción de tareas

Durante esta orden de trabajo se realizaron actividades de investigación y desarrollo enfocados a mejorar herramientas y construir nuevas. La descripción de cada tarea se describe a continuación.

2.1 Análisis de inconsistencias

2.1.1 Tabla de viajes

En base a un análisis realizado por el equipo de DTPM, se encontró un conjunto de posibles inconsistencias dentro de la tabla de viajes, en esta sección se realizará una revisión de cada una de ellas.

2.1.1.1 Variables nulas

Se encontraron los siguientes campos con valores nulos, y se acompaña una descripción de cada caso en la Tabla 2.1.1.1. Es importante adelantar que los campos que son eliminados por no uso o reemplazo por un nuevo campo se mantienen dentro de la tabla con valor nulo, esto porque modificar el orden de las columnas implica un gran impacto en el flujo de procesamiento posteriores en el visualizador. En caso que se considere relevante eliminar estos campos nulos se debe evaluar un desarrollo nuevo para ajustar este cambio.

Campo	Descripción
tviaje	Este campo fue eliminado, puesto que el nuevo campo tviaje2 cumple el mismo objetivo pero con una nueva metodología, mejorada.
mediahora_bajada_1, mediahora_bajada_2, mediahora_bajada_3, mediahora_bajada_4	Eliminados por no uso
periodo_bajada_1 periodo_bajada_2 periodo_bajada_3 periodo_bajada_4	Eliminados por no uso
op_1era_etapa op_2da_etapa op_3era_etapa	Eliminados por no uso

op_4ta_etapa	
contrato	Este campo desde abril 2023 si tiene valor, actualmente contiene el código de contrato de la tarjeta

Tabla 2.1.1.1 : Campos nulos en tabla viajes.

2.1.1.2 Discordancia en tiempos de viaje

Caso 1 : “Discordancia entre tiempos de viaje y tiempo fin viaje : 802.590 viajes no tienen tiempo fin , mientras que 853.163 no tienen tiempo de viaje.”

El tiempo de viaje calculado tiene un conjunto de filtros adicionales que permiten que el viaje pueda tener un tiempo de fin [tiempo_fin_viaje], pero no así un tiempo de viaje, puesto que los viajes que no cumplen, por ejemplo, una condición de velocidad lógica, se les anula el campo tiempo de viaje [tviaje2] para no generar ruido con análisis de tiempo de viajes posteriores.

Dicho lo anterior, no es un error que exista un tiempo de fin pero no así un tiempo de viaje.

2.1.1.3 Problema tiempo fin viajes de más de 4 etapas

Caso 2 : “73 viajes sin tiempo fin, pero con tiempo de viaje.”

Se identificó que todos los viajes que tienen más de 4 etapas tienen el campo [tiempo_fin_viaje] nulo. Se encontró que la razón de este problema radica en que la tabla de viajes está pensada para viajes de máximo 4 etapas, por tanto la última etapa, la cual contiene el tiempo fin de viaje no es considerada, se corrigió este problema para nuevos procesamientos.

2.1.1.4 viajes sin bajada con tiempo de fin

Caso 3 : “1.900 Viajes sin bajadas pero con tiempos de fin: ninguno posee tiempo de viaje, pero sí tiempos de inicio y fin. ”

Se revisó para el día 2023-05-02 y se encontraron 4.046 casos, esto no es un error puesto que la herramienta si puede identificar el GPS de bajada de un viaje pero no encontrar un paradero cercano, esto porque existe un umbral máximo de 500 metros para considerar un paradero de bajada, esto puede darse, o se han visto casos donde existe algún error en la posición del GPS o algún caso de borde no considerado. Este caso es filtrado para calcular el tiempo de viaje dado que es conocido que podría existir un error en los datos GPS.

2.1.1.5 Viajes sin servicio de subida

Caso 4 : “Existen viajes sin servicio de subida que tienen bajada, 39 casos de 1 etapa.”

El servicio es una variable que puede o no tener valor, esto porque se puede estimar la bajada de una transacción sobre un bus independiente si se conoce el servicio-sentido, esto solo impacta en la precisión de la estimación del paradero de bajada, si no se conoce el servicio se busca el paradero más cercano y no necesariamente podría ser el del servicio que si estaba realizando el bus.

2.1.1.6 Comunas nulas

Caso 5 : “Discordancia entre comunas nulas y paraderos (inicio y fin)”

Solo se encontró un caso para el día 2023-05-02, el cual corresponde al paradero I26228SN25, el problema es que ese código no existe como paradero en los datos de entrada y por lo tanto no está asociado a una comuna.

2.1.1.7 Comunas y zonas

Caso 6 : “Discordancia entre comunas y zonas (inicio y fin): 3.073 sin comuna origen, 2.755 sin zona origen; 806.854 sin comuna fin, 806.839 sin zona fin (zona 62 -> Parada E-5-38-SN-5).”

No se encontraron casos con comuna y sin zona para el día revisado 2023-05-02.

Respecto al 806.854 sin comuna fin no es un problema puesto que corresponde a los viajes que no tienen estimación de bajada y por tanto no tienen paradero de bajada, lo mismo para las zonas.

2.1.1.8 Sin paradero inicial

Caso 7 : “Dos viajes con paradero de subida servicio 1 y sin paradero inicial”

Se encontró un caso para el día en revisión (2023-05-02), el problema radica en una condición particular del proceso de asignación de etapas de un viaje, el error fue corregido para futuros procesos.

2.1.1.9 Tipificación de modos

Caso 8 : “Modos del 5 al 7 sin diccionario”

La notación o tipificación de la combinación de modos de un viaje depende de un algoritmo interno de la herramienta, estos números o ids representan combinaciones de uso de los modos en el viaje, a continuación se describe este algoritmo.

Inicialmente se tiene tres variables con valor cero :

modo_metro = 0

modo_bus = 0

modo_metrotren = 0

Por cada etapa del viaje se realizan las asignaciones que se indican en la tabla 2.1.1.9.a.

Condición	Variable y valor asignado
si la etapa es de tipo "BUS"	modo_bus = 1
si la etapa es de tipo "METRO"	modo_metro = 2
si la etapa es de tipo "ZONA PAGA"	modo_bus = 1
si la etapa es de tipo "METRO TREN"	modo_metrotren = 4

Tabla 2.1.1.9.a : Asignación de valores por modo de etapa

Finalmente el campo **modos** de la tabla de viajes se construye con la suma de estas variables:

$\text{modos} \leftarrow \text{modo_bus} + \text{modo_metro} + \text{modo_metrotren}$

Y los valores resultantes se pueden ver en la tabla 2.1.1.9.b .

Valor campo "modos"	Variable y valor asignado
1	viaje con etapas solo en bus
2	viaje con etapas solo en metro
3	viaje con etapas en bus y metro
4	viaje con etapas solo en metro tren
5	viaje con etapas en metro tren y bus
6	viaje con etapas en metro y metro tren
7	viaje con etapas en bus, metro y metro tren

Tabla 2.1.1.9.b : Valores posibles y su significado

2.1.2 Tabla de etapas

2.1.2.1 Etapas sin servicio y con factor de expansión positivo

Caso 1 : "126.955 etapas sin servicio de subida, ni paradero de subida ni bajada, pero la suma de f_exp es 6.845 y no 0"

No se encontraron casos con esta condición.

2.1.2.2 Servicios distintos

Caso 2 : "274.288 corresponden a algún servicio de bus con 696 servicios de subida (código sonda) distintos"

Se conversó del caso nuevamente con la contraparte de DTPM y se definió que este caso no es un problema sino un comentario.

2.1.2.3 Servicios con pocas etapas

Caso 3 : "27 servicios tienen 13 o menos etapas"

La metodología cruza la información del bus donde se realiza la transacción con la información de expediciones 1.96, esto permite asignar el servicio-sentido que estaba asignado al bus hacia la transacción. Si existiese alguna inconsistencia en base a lo esperable en algún servicio en particular es necesario definir qué servicios son para poder revisar el número de transacciones asignadas en conjunto a las expediciones y GPS, esta es la única manera de poder encontrar algún problema en esta línea.

2.1.2.4 Paradas que no corresponden a servicio

Caso 4 : "77 servicios tienen paradas de subidas que no corresponden según su anexo paradas (4% error aprox) o no poseen paradas porque es un cod_sonda que "no existe"

Esta condición no es un error, en la metodología cuando se da el caso de un bus que su trayectoria se aleja en más de 100 metros del trazado de su servicio, las transacciones son asignadas al paradero más cercano de la red en el sentido de avance del bus, esto significa que esta transacción probablemente quedará asignada a un paradero que no pertenece a la secuencia de paradas del servicio del bus.

2.1.2.5 Inconsistencia de suma de factor de expansión

Caso 5 : "222 servicios tienen inconsistencias con la suma de f_exp (en una etapa o más) con diferencias hasta 457 etapas"

La suma del factor de expansión es un tema complejo, principalmente porque existen dos tipos de expansión:

fExpansionZonaPeriodoTS: Esta expansión corresponde a expandir todas las etapas que no tienen paradero de bajada sobre los que sí tienen, agrupados en la misma zona 777 y periodo de Transantiago, es importante entender que se expanden todas las validaciones en ese grupo sin separar los modos, por lo que esta expansión no tendrá sentido para sumar los factores por servicio.

fExpansionServicioPeriodoTS: Esta expansión expande todas las validaciones realizadas en un mismo servicio-sentido-paradero-media hora, esta expansión tiene por objetivo ser utilizada en la construcción de perfiles de carga. Es importante mencionar que pueden darse casos donde en una media hora-servicio-sentido-paradero solo existen validaciones sin estimación de bajada, en tal caso se distribuyen estas validaciones utilizando una distribución uniforme sobre las validaciones aguas abajo.

En base a estas definiciones las evaluaciones de cuadratura solo pueden ser realizadas en base a este nivel de agregación: servicio-sentido-media hora. Queda pendiente una reevaluación del problema utilizando estas definiciones, es muy probable que haya existido un mal entendido puesto que el nombre del campo es `f_exp`, nombre del campo de una versión anterior de la tabla de etapas y no queda claro cuál expansión es la que se está revisando.

Se entregan algunos ejemplos de valores para el día 2023-08-16.

Paradero	Media hora	ServicioSentido	sum(f_exp)	#Subidas
T-21-205-SN-95	18:30:00	T205 00R	5.0312	5
T-17-140-OP-17	18:30:00	T557 00R	6.004	6
L-31-12-10-SN	06:00:00	T1216 00I	3.0192	3

2.2 Revisión de circunvalares

La tarea de revisión de datos de servicios circunvalares consta de cuatro etapas, la primera se enfoca en revisar cuál fue la situación en la última orden de trabajo donde se realizaron mejoras, la segunda consiste en realizar una evaluación para cada uno de los servicios circunvalares con datos más actualizados y verificar que no han aparecido cambios, la tercera etapa corresponde a la revisión de los casos bordes encontrados en esta inspección y la cuarta etapa se enfoca en entregar una lista de sugerencias para mejorar los puntos anteriormente encontrados.

El objetivo de esta tarea es poder confirmar que se mantienen los resultados de las mejoras desarrolladas con datos más actuales, identificar los casos donde no y poder tipificarlos para sugerir mejoras futuras.

2.2.1 Situación anterior

Con el objeto de poder mejorar los datos de los servicios circunvalares y de servicios con trazados con condiciones especiales, por ejemplo, que un bus pase por ambos sentidos de la misma calle, en la orden de trabajo número 2 del actual contrato se realizaron dos modificaciones en las proyecciones a ruta, una de ellas en los GPS y la otra en los paraderos.

En esa ocasión se pudo establecer que del conjunto de servicios con estas condiciones se pudieron mejorar el 98% de las proyecciones erradas y para el caso de los paraderos existían 91 servicios que tenían algún desorden de proyección de los cuales fueron resueltos 63% de los casos, sin considerar que se encontraron un conjunto de paraderos que tenían un error en su georeferenciación, por lo tanto los servicios que no pudieron ser corregidos con los nuevos algoritmos de proyección fueron 12%.

2.2.2 Inspección situación actual

El objetivo de esta tarea se enfoca en realizar un nuevo análisis, en particular con los últimos datos procesados, junio 2023, y poder revisar si se mantiene los mismos resultados de la corrección anterior, y como segundo objetivo lograr identificar con más detalle cuales son los casos que persisten en tener algún problema.

Se realizó una revisión sobre un día de datos, donde se buscó la condición de anomalía de proyección, es decir, más de 200 km/h o menos de -200 km/h de velocidad entre dos pulsos gps, el gráfico 2.2.2 muestra el porcentaje que se da estas anomalías antes y después del algoritmo de corrección. Como se aprecia, la curva azul representa los porcentajes por cada servicio antes de aplicar el algoritmo de corrección y se ven varios servicios con cantidades superiores al 5%, es importante notar que tener hasta un 1% de estas anomalías parece ser un problema normal en los datos, esto puede estar asociado a problemas en los datos de GPS, que se han visto, por ejemplo, saltos o problemas de señal.

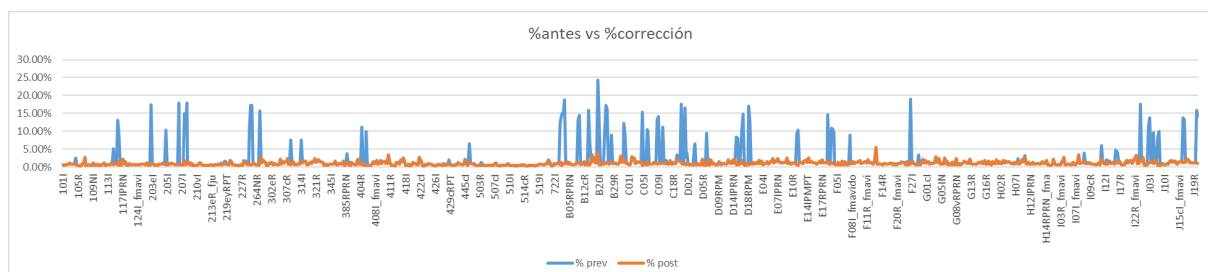


Gráfico 2.2.2 : Comparación entre el porcentaje de anomalías antes y después de la aplicación del algoritmo de corrección.

2.2.3 Tipificación de problemas

Para poder tipificar el primer paso es agrupar los casos con la mayor cantidad de problemas, el gráfico 2.2.3 muestra solo los servicios que presentan más de un 3% de sus gps con estas anomalías en su estado original. Se aprecia como casi todos los servicios fueron llevados al mismo nivel del resto de servicios considerados normales en base a sus trazados sin inconvenientes o situaciones complejas de proyección, como lo son pasos en dos sentidos por la misma calle y cruces entre el mismo trazado de un mismo sentido.

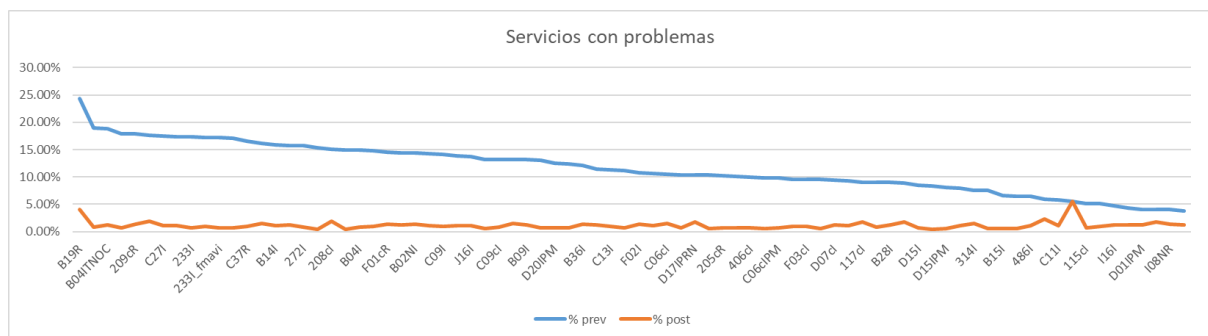


Gráfico 2.2.3: Selección de los servicios con mayores cantidades de anomalías antes y después de la corrección.

Es importante notar que un servicio parece no haber sido afectado por la corrección, este es el caso del servicio F13cR, pero se aprecia que solo existen 800 pulsos GPS para el día sobre 45 expediciones, por otro lado se aprecian 45 casos de anomalías lo que significa que existe una anomalía por expedición, el trazado se puede observar en la Ilustración 2.2.3.a.

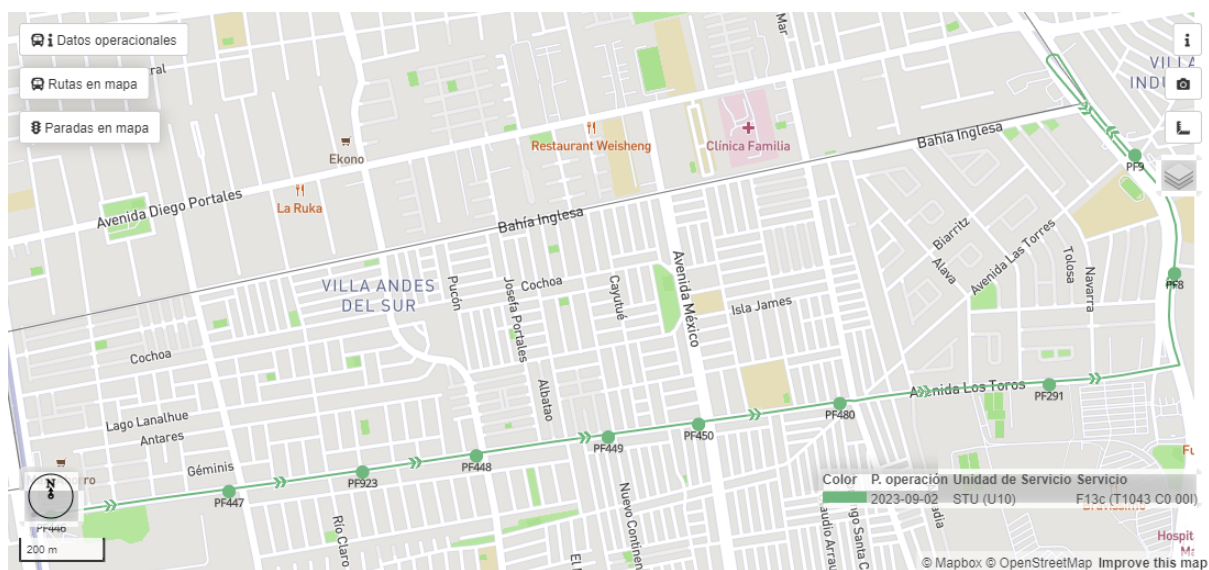


Ilustración 2.2.3.a : Trazado del servicio F13cR, no parece tener algún tipo de problema.

Al revisar los datos GPS proyectados del servicio F13cR, se observó una irregularidad muy fuera de lo normal, se observa que las 45 anomalías están relacionadas a una distancia en ruta fuera de rango, los valores de estas proyecciones son:

[987020593, 1503351254, 1956954776, 1030726180, 1731774794, 2575649969, 3626783085, 1265636840, 2502807323, 4229865213, 205707630, 2223036590, 3121832370, 3752897068, 206862118, 138967275, 1443262943, 3279172484, 168049357, 2607810684, 3149869916, 3453610003, 577342561, 3945916335, 2814958674, 2659377779, 2010391041, 1378611659, 3839098780, 4122772780, 3691819197, 2097692264, 1807915974, 1321547006, 1858823363, 1368826006, 668272172, 2460641052, 3614633481, 1577183459, 3703803874, 2559267245, 3379997700, 4038715116, 1730840209]

Con estos datos se realizó un análisis en mayor profundidad de el resto de anomalías en los otros servicios y **se observó un error en la proyección de las rutas para el último GPS de cada expedición**, este error fue corregido en el software, el resultado de esta corrección se puede observar en el histograma de la Ilustración 2.2.3.b, como se aprecia, solo quedaron algunos servicios con porcentajes en su mayoría menores al 0.1%, eliminando prácticamente la mayoría de las anomalías.

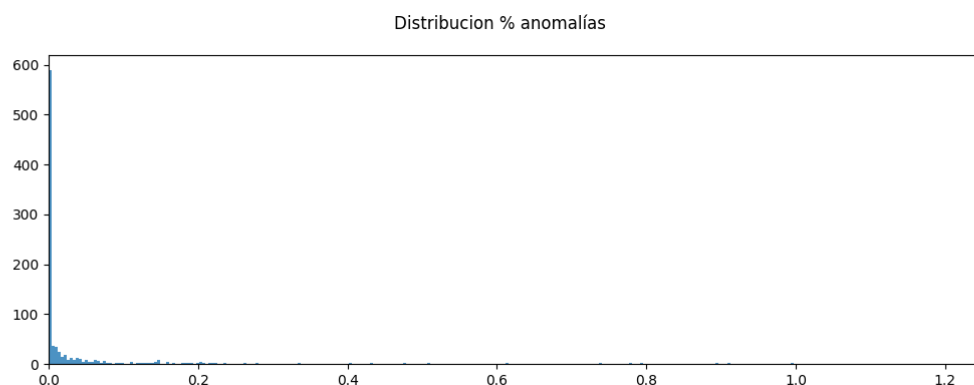


Ilustración 2.2.3.b : Histograma de porcentajes de anomalías en los servicios.

Si se observan los casos más complejos, es decir, los que no han podido ser solucionados con las correcciones algorítmicas (4%), nos quedan principalmente los servicios de la Tabla 2.2.3.t, a modo global el número de anomalías es corregido en un 96%, donde en el día de análisis se pasa de 161.750 casos a 6.751. Es importante mencionar que el total de GPS corresponden a 8.970.484, lo que significa que las anomalías a nivel global corresponden a un 0.07% de los datos, que está principalmente concentrado en estos trazados complejos.

servicio	anomalias	total	porcentaje
B19R	191	4749	4.022%
117cl	92	5907	1.557%
417el	45	2971	1.515%
B16IPRN	2	141	1.418%
I25I	231	16531	1.397%
410yl	3	226	1.327%
J16I_fmasa	110	8580	1.282%
303eR	53	4606	1.151%
I03cl_fmavi	2	175	1.143%
314I	317	28454	1.114%
415el	25	2510	0.996%
B04ITNOC	4	402	0.995%
D07IPRN	8	827	0.967%
F33I_fvi	58	6015	0.964%
B02NI	19	2065	0.920%
J20I_fmavi	42	4602	0.913%
D17IPRN	3	330	0.909%
409I	182	20125	0.904%
J16I	42	4691	0.895%
D07RPRN	10	1119	0.894%
J02I	120	13711	0.875%

Tabla 2.2.3.t : Tabla con servicios con mayor cantidad de anomalías a nivel porcentual.

Se aprecia que el caso con mayor porcentaje de anomalías corresponde al B19R, desde los datos se puede apreciar que tiene un conjunto de problemas de proyección, esto se debe al inicio y fin que están dentro de la misma calle y la existencia de orejas o intersecciones de la ruta con sigla misma.

Se realizó un nuevo análisis para entender si existe algún sesgo de estas anomalías que no han podido ser eliminadas respecto a la distancia en ruta, en el cual se aprecian dos grandes grupos de anomalías, un conjunto asociado a problemas de proyección en el inicio/fin en rutas circunvalares que usan la misma calle para iniciar y terminar el trazado y otro grupo que tienen saltos de distancia entre GPS superiores a 2 km los que generan una velocidad superior a 200 km/hr.

Como conclusión podemos ver que aún existen casos que no han podido ser resueltos pero dependen de casos muy particulares respecto al trazado de la ruta, por lo que se debería buscar otro método de proyección solo para estos casos excepcionales.

2.2.4 Trabajo futuro

Del estudio podemos entender que a pesar de la fuerte disminución de las anomalías de proyección (96%) aún queda un 4% agrupado principalmente en trazados de alta complejidad para una proyección, para esto se debe realizar un nuevo estudio que diseñe un nuevo algoritmo específico para estos casos, en primera instancia se aprecia que existen patrones como por ejemplo el caso de inicio/fin del trazado que podría ser manejado como un caso especial.

2.3 Revisión de datos de perfiles

En esta sección se mostrarán las revisiones realizadas para poder determinar la salud de los perfiles de carga, para esto se realizaron análisis desde distintos puntos de vista los cuales se presentan a continuación.

2.3.1 Análisis de carga final

Se realizó un análisis respecto al cantidad final de la carga en los buses de todas las expediciones de un día en particular (2023-08-22), la Ilustración 2.3.1a muestra como es la distribución de esta variable.

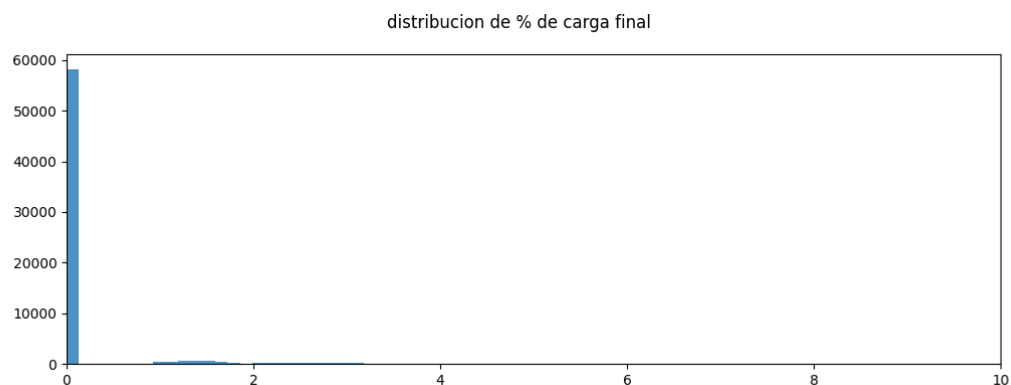


Ilustración 2.3.1.a : Histograma con la distribución de carga final de todas las expediciones.

Algunas estadísticas se muestran en la Tabla 2.3.1.b.

Item	Valor
# Expediciones totales	65.516
Media	0.5655
Moda	0.0

Mínimo	-0.00031
Máximo	132.64
% expediciones con carga > 10	1.3%

Tabla 2.3.1.b : Estadísticas respecto a la carga final de las expediciones.

Se puede apreciar que respecto a esta variable podría ser recomendable eliminar de los análisis agregados las expediciones con cargas finales mayores a 10.

2.3.2 Análisis de carga máxima

Se realizó un análisis sobre las cargas máxima de cada expedición, donde se encontraron los siguientes resultados. La Ilustración 2.3.2.a muestra la distribución de esta variable .

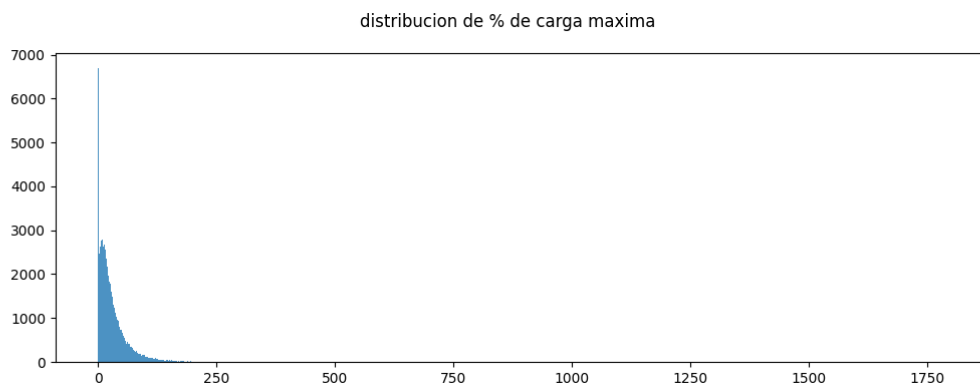


Ilustración 2.3.2.a : Distribución total de la carga máxima de todas las expediciones.

Por otra parte, las estadísticas básicas de esta variable se pueden observar en la tabla 2.3.2.b.

Item	Valor
# Expediciones totales	65.516
Media	29.52
Moda	0.0
Mínimo	0.0
Máximo	1791.28

Tabla 2.3.2.b : Estadísticas respecto a la carga máxima de las expediciones.

Considerando 250 como un límite máximo para el gráfico del histograma se puede apreciar mejor la distribución de esta variable en la Ilustración 2.3.2.c.

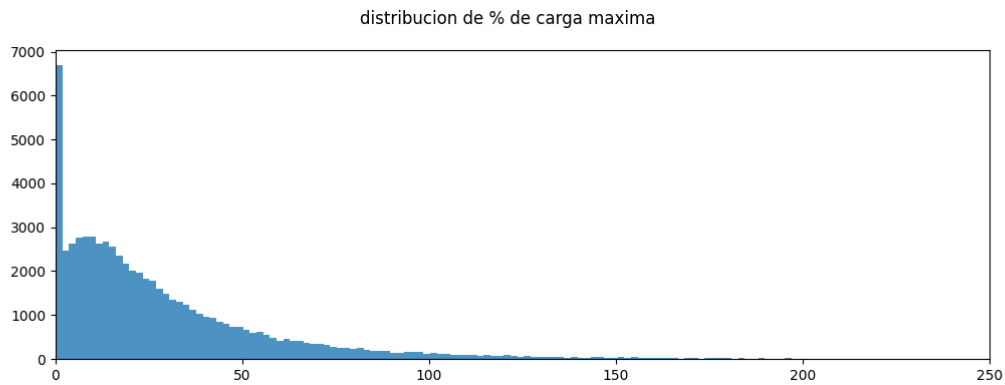


Ilustración 2.3.2.c : Distribución de carga máxima

De este análisis podemos observar que un 8.52% de las expediciones tienen como carga máxima 0, lo que podría ser considerado como una expedición inválida para usos de análisis, por otra parte existe un 1.09% de las expediciones que tienen una carga máxima mayor a 150 pasajeros, lo cual también parece un valor fuera de rango. Finalmente, un 90% de las expediciones parecen normales respecto a esta variable.

2.3.3 Análisis de subidas

Este análisis corresponde a la revisión del número de subidas totales de todas las expediciones, para esto un primer análisis corresponde a revisar su distribución, tal como muestra la Ilustración 2.3.3.

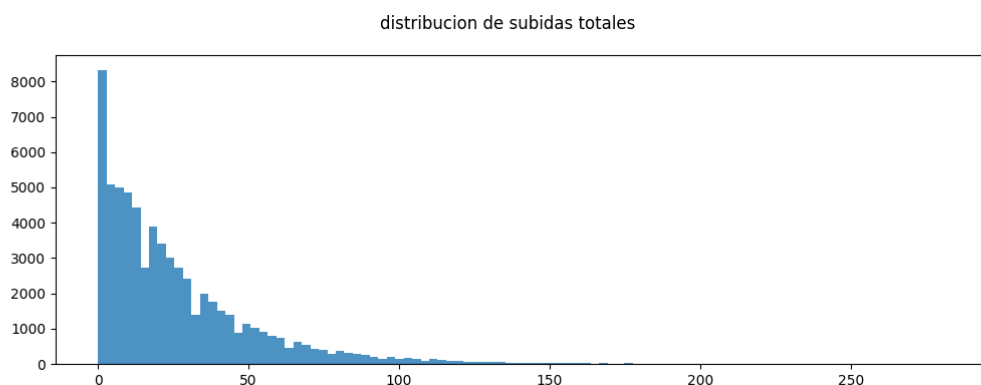


Ilustración 2.3.3.a : Distribución de subidas totales

También se revisaron las estadísticas básicas, las cuales se detallan en la tabla 2.3.3.b.

Item	Valor
# Expediciones totales	65.516
Media	26.10
Moda	0.0
Mínimo	0.0
Máximo	282

Tabla 2.3.3.b : Tabla de estadísticas de subidas totales.

Se aprecia nuevamente casos con expediciones con una demanda demasiado baja, en este caso se puede ver que un 6.6% de las expediciones tienen cero subidas.

El siguiente gráfico muestra cómo están distribuidas estas expediciones con cero subidas durante el día, donde se aprecia su acumulación en las horas de menor demanda lo que podría entregar un grado de coherencia a este problema, para los periodos donde no debiese existir este problema es necesario realizar un análisis más profundo para determinar cuál es su razón.

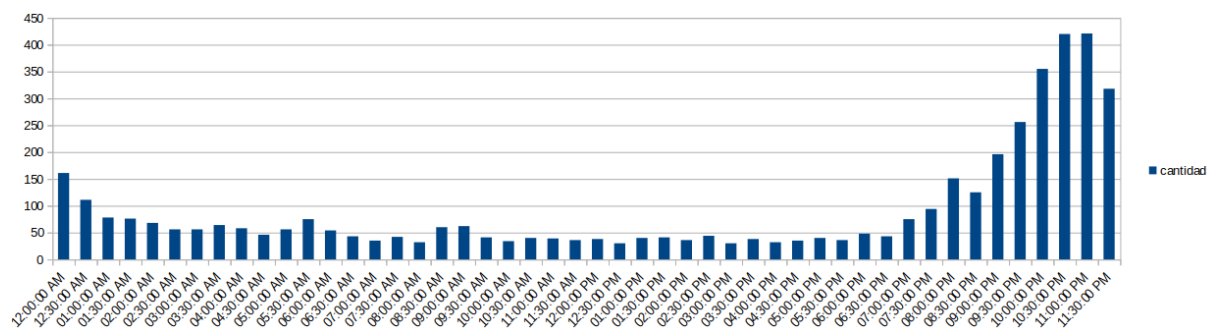


Ilustración 2.3.3.c : Distribución de expediciones con cero subidas durante el día.

2.3.4 Análisis de bajadas

Este análisis corresponde a analizar el número de subidas totales de todas las expediciones, para esto un primer análisis corresponde a revisar su distribución, tal como muestra la Ilustración 2.3.4.a.

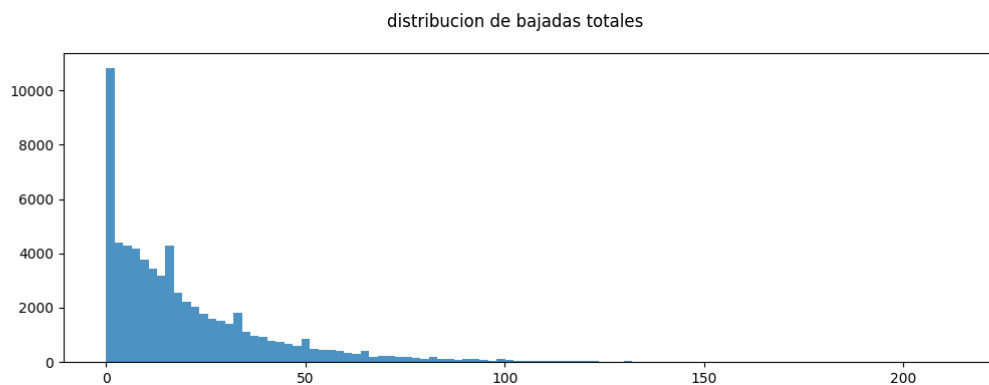


Ilustración 2.3.4.a : Distribución de bajadas totales

Tal como las subidas se aprecia un alto número de expediciones con una cantidad de bajadas igual a cero, en este caso un 8.4%.

2.3.5 Análisis de subidas y bajadas

Se realizó una comparación entre las subidas y bajadas (sin evasión, sin expansión y sin correcciones), esto con el objeto de analizar los datos más puros producidos por la metodología, es importante que desde este punto las subidas sean coherentes puestos que son directos de los datos de transacciones de pago, para esto se utilizó un gráfico tipo scatter, el cual se muestra en la Ilustración 2.3.5.a. Se puede observar un sesgo, el cual es esperado porque las bajadas solo existen para un 70% de las etapas.

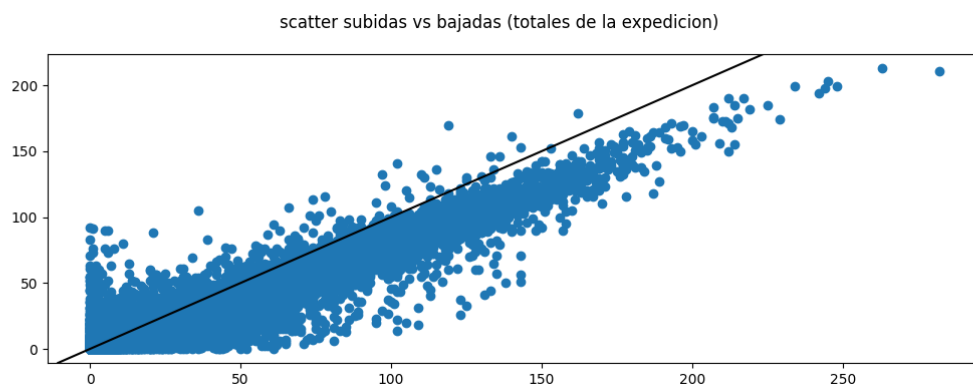


Ilustración 2.3.5.a : Comparación entre subidas y bajadas

2.3.6 Análisis de subidas y bajadas corregidas

Para un análisis con datos más acabados se realizó una comparación entre las subidas y bajadas corregidas, es decir, posterior a los procesos de expansión, integración de

evasión y correcciones adicionales, el resultado de esta comparación se puede apreciar en la Ilustración 2.3.6.a, la cual muestra lo esperado, se equiparan subidas y bajadas.

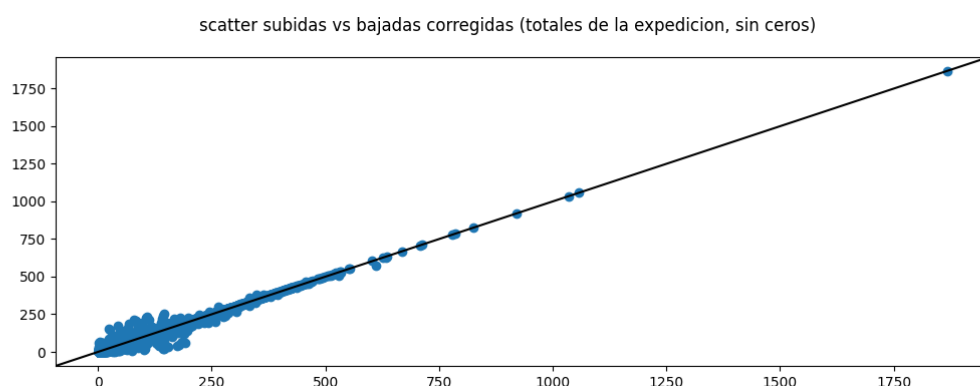


Ilustración 2.3.6.a : Comparación entre subidas y bajadas

2.3.7 Análisis de paraderos iniciales

Se realizó una revisión de los datos de subidas y bajadas en los paraderos iniciales, para esto se calcularon las estadísticas básicas de todas las subidas y bajadas realizadas en el paradero inicial de todas las expediciones, los resultados se muestran en la Tabla 2.3.7.a.

Item	Subidas	Bajadas
# Expediciones totales	65.516	65.516
Media	3.79	0.36
Moda	0.0	0.0
Mínimo	0.0	0.0
Máximo	1021	67.76
Valores negativos	0	0
% rango normal subidas = [0,50], bajadas = 0	98.7%	93%
% outliers	1.3%	7%

Tabla 2.3.7.a : Tabla de estadísticas de subidas totales.

A continuación, una tabla con los servicios más representativos de los valores outliers.

Servicio	Expediciones	Porcentaje
C14	45	5.32
B16	37	4.37
C11	35	4.14
G08	35	4.14
J04	35	4.14
301	34	4.02
C02	26	3.07
C15	25	2.96
C10	23	2.72
411	22	2.60
207	21	2.48
C20	21	2.48
104	20	2.36
G11	20	2.36
307	19	2.25
D09	19	2.25

Tabla 2.3.7.b : Servicios con más expediciones con subidas en primer paradero fuera de rango (subidas > 50).

Servicio	Expediciones	Porcentaje
211	171	3.70
204	119	2.58
E02	102	2.21
G22	94	2.04
408	89	1.93
G08	87	1.88
301	85	1.84
G13	76	1.65
228	67	1.45
B19	66	1.43
E03	64	1.39
J12	64	1.39
555	63	1.36
G05	63	1.36

Tabla 2.3.7.c : Servicios con más expediciones con bajadas en primer paradero fuera de rango (bajadas > 0).

2.3.8 Análisis de paraderos finales

Se realizó una revisión de los datos de subidas y bajadas en los paraderos finales, para esto se calcularon las estadísticas básicas de todas las subidas y bajadas realizadas en el paradero final de todas las expediciones, los resultados se muestran en la Tabla 2.3.8.a.

Item	Subidas	Bajadas
# Expediciones totales	65.516	65.516
Media	0.062	4.06
Moda	0.0	0.0
Mínimo	0.0	0.0
Máximo	39.88	437.9
Valores negativos	0	0
% rango normal subidas = 0, bajadas = [0,100]	99.88%	99.7%
% outliers	0.12%	0.3%

Tabla 2.3.8.a : Tabla de estadísticas de subidas totales.

A continuación, una tabla con los servicios más representativos de los valores outliers.

Servicio	Expediciones	Porcentaje
I16	29	38.16
C27	7	9.21
D01	6	7.89
F03	6	7.89
B15	5	6.58
B32	5	6.58
233	3	3.95
208	2	2.63
C11	2	2.63
116	1	1.32
209	1	1.32

Tabla 2.3.7.b : Servicios con más expediciones con subidas en el último paradero fuera de rango (subidas != 0), no deberían haber subidas en el último paradero.

Se aprecia que el servicio con subidas distintas de cero en el último paradero más significativo es el servicio I16.

Servicio	Expediciones	Porcentaje
J13	21	10.55
J18	20	10.05
J04	15	7.54
E09	11	5.53
410	9	4.52
201	7	3.52
H08	7	3.52
204	6	3.02
G23	6	3.02
J01	6	3.02
C10	5	2.51

Tabla 2.3.7.c : Servicios con más expediciones con bajadas en último paradero fuera de rango (bajadas > 100).

2.4 Nuevo archivo gps offline

Los archivos GPS son obtenidos de dos fuentes actualmente, una de ellas es el rescate recurrente, y en tiempo real, de los pulsos generados por el sistema. Este mecanismo es rápido pero no preciso, puesto que hay datos que son enviados con un desfase que esta forma no integra, por lo que son pulsos que no forman parte de esta fuente. La segunda fuente corresponde a un archivo que se entrega posterior a la ventana de tiempo de generación, y que contiene todos los pulsos GPS generados durante ese periodo.

En el visualizador de los datos de ADATRAP, existe una vista que permite ver todos los pulsos GPS recurrentes que se encuentran en el almacenamiento en la nube, En la Ilustración 2.5.1 se muestra la vista. En la parte superior tiene un calendario que muestra los días con datos en color verde, mientras que aquellos días sin datos se muestran en color blanco. Los días en color azul corresponde a días en que cambió el programa operacional de sistema de transporte y los días en color amarillo son aquellos que se identifican como semanas representativas del periodo.



Ilustración 2.5.1: vista para descargar pulsos GPS recurrentes

Se solicitó al consultor implementar una vista similar a la mostrada en la Ilustración 2.5.1 pero para los datos GPS que se reciben posterior al periodo de generación. En la Ilustración 2.5.2 se muestra la vista resultante. La nueva vista se muestra cuando el usuario presiona en la sección “GPS Offline” de la sección “Almacenamiento” del menú lateral. Al igual que la Ilustración 2.5.1, contiene un calendario para mostrar los días disponibles, y en la parte inferior, una tabla para generar links temporales de descarga, lo que permite que los usuarios puedan descargar los datos de forma local.

Es importante mencionar que en ambos casos, los datos son agrupados en archivos diarios, es decir, un archivo por día.

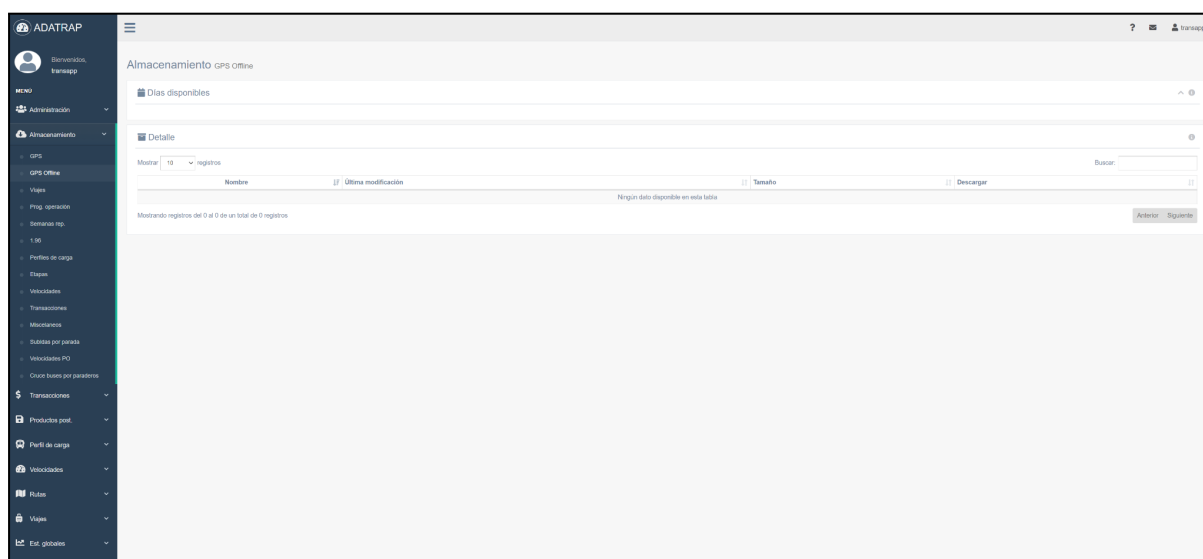


Ilustración 2.5.2: vista para descargar pulsos GPS obtenidos posterior a su generación

2.5 Nuevas funcionalidades en ficha de servicios

La ficha de servicios es una sección de la herramienta de visualización de ADATRAP disponible en el sitio web de <https://www.adatrap.cl>. Esta herramienta muestra los datos de uno o más programas operacionales, pudiendo elegir entre tres tipos de datos:

- Paradas
- Recorridos con su respectiva secuencia de paradas
- Datos operacionales de cada recorrido (frecuencia, capacidad, distancia, velocidad)

Adicional a lo anterior, se incluyen características que dan contexto a los datos operacionales, estos son:

- Capa de líneas de metro junto a sus estaciones. El usuario tiene la posibilidad de mostrar u ocultarlas
- Capa de comunas que destaca los límites territoriales de cada una. El usuario también tiene la posibilidad de mostrar u ocultar estas geometrías
- Brújula que muestra la ubicación del norte geográfico, esto es necesario puesto que el mapa puede rotar a voluntad del usuario
- Control para tomar imágenes de los datos visibles en el mapa
- Control para medir distancias
- Control que muestra la escala del mapa

La ficha de servicio ha permitido realizar análisis complejos y útiles para distintos perfiles profesionales en el equipo de DTPM, lo que ha a su vez ha generado nuevos requerimientos. A continuación se describen los nuevos requerimientos y los resultados obtenidos de su implementación.

2.5.1 Incluir unidad de servicio asociada a cada ruta desplegada en el mapa

Se solicita al consultor poder incluir como parte de los filtros de una ruta, un selector que muestre las unidades de servicios existentes para un programa de operación seleccionado. Este, a su vez, debe funcionar como filtro para el selector de servicios.

2.5.1.1 Implementación

Se incorporó al selector de rutas un campo adicional para seleccionar la unidad de servicio. Este nuevo campo utiliza la información del programa de operación seleccionado previamente para mostrar solo aquellas unidades de servicio que son válidas.

En la Ilustración 2.6.1.1.1 se muestra el resultado de la implementación. A la derecha del selector de programa de operación, se agregó un selector que muestra las unidades de servicio disponibles. Una vez que el usuario selecciona un valor, el selector de servicio es limitado al conjunto de recorridos que son responsabilidad de la unidad de servicio seleccionada.

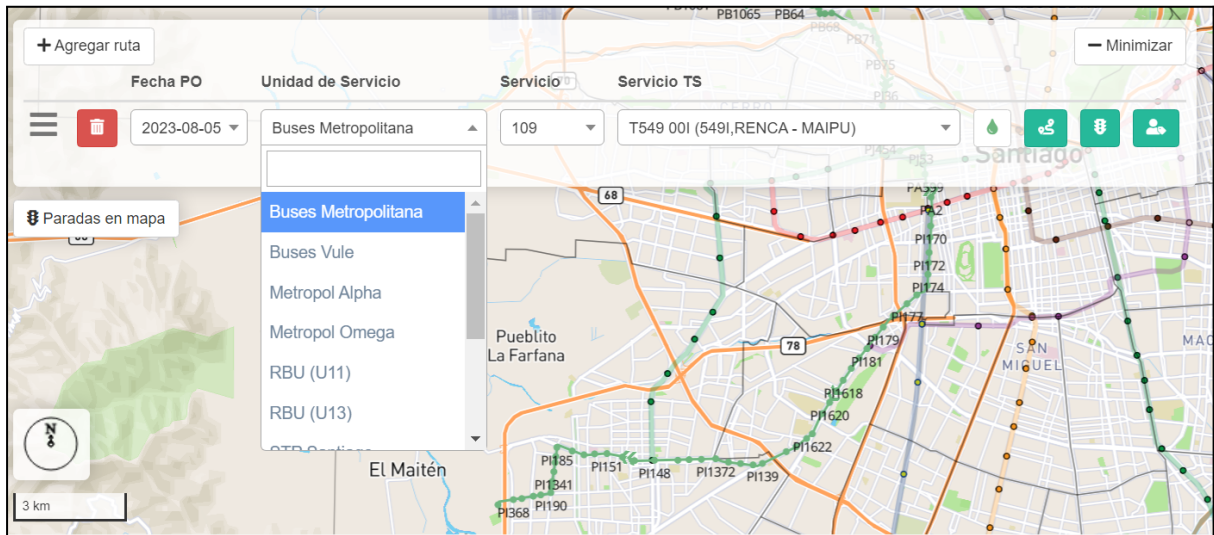


Ilustración 2.6.1.1.1: selector de unidad de servicio para agregar una ruta al mapa

2.5.2 Definir regla de colores para rutas y paradas

Se solicita al consultar establecer una secuencia fija de colores para las rutas. Actualmente, cada vez que se modifica una ruta o se agrega una nueva se asigna un color aleatorio, lo que dificulta los análisis.

2.5.2.1 Implementación

Previo a esta solicitud de cambio, la selección de colores para cada recorrido era asignada de forma aleatoria, y además, frente a cada cambio en alguno de los selectores, el color era actualizado. Con el cambio realizado, ahora los primeros diez colores son fijos e independientes de los cambios que se hagan en los selectores. La lista de colores se muestra en la Ilustración 2.6.2.1.1, luego del décimo elemento, los colores se generan de manera aleatoria.

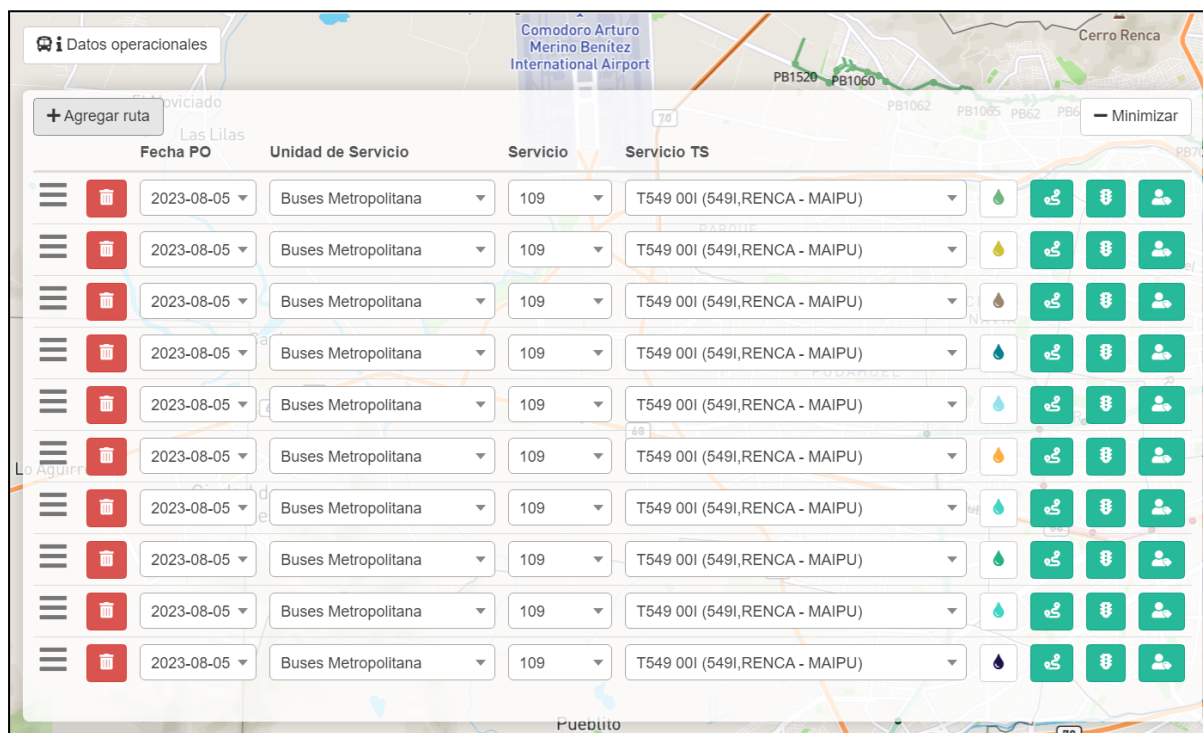


Ilustración 2.6.2.1.1: asignación fija de colores

La lista de colores, según su definición hexadecimal, es la siguiente:

- 1) #6FB67F
- 2) #CCC138
- 3) #9A856D
- 4) #07818D
- 5) #91E2EC
- 6) #F13E22
- 7) #FEAD42
- 8) #26B380
- 9) #41D5C3
- 10) #1E114C

Es importante mencionar que los usuarios aún tienen la posibilidad de cambiar el color asignado por defecto. Además, su asignación no se verá afectada por cambios en los selectores.

2.5.3 Asignar orden de prioridad cuando se selecciona más de una ruta

Actualmente, la secuencia de rutas visible en el mapa es dependiente del orden en que se van agregando los recorridos en el selector, no permitiendo que el usuario pueda alterarla con posterioridad, esto genera inconvenientes cuando las rutas se traslapan o se quiere mostrar una lista en un orden distinto al generado inicialmente. Se solicita al consultor permitir un reordenamiento arbitrario en el listado de recorridos, y a su vez, que esta modificación altere el orden de las capas en el mapa.

2.5.3.1 Implementación

El reordenamiento del listado se implementó mediante la técnica “Drag and drop”, en donde el usuario tiene la posibilidad de arrastrar elementos de una lista a otra ubicación. Cuando eso se confirma, se realizan dos actualizaciones en el mapa: se actualiza el orden de la leyenda de los recorridos, y segundo, se reordenan las capas para que el nuevo recorrido esté sobre los recorridos asignados en las filas inferiores del listado.

En la Ilustración 2.6.3.1.1 se muestra el icono de reordenamiento, ubicado en la sección izquierda de cada fila (a la izquierda del icono de eliminar).

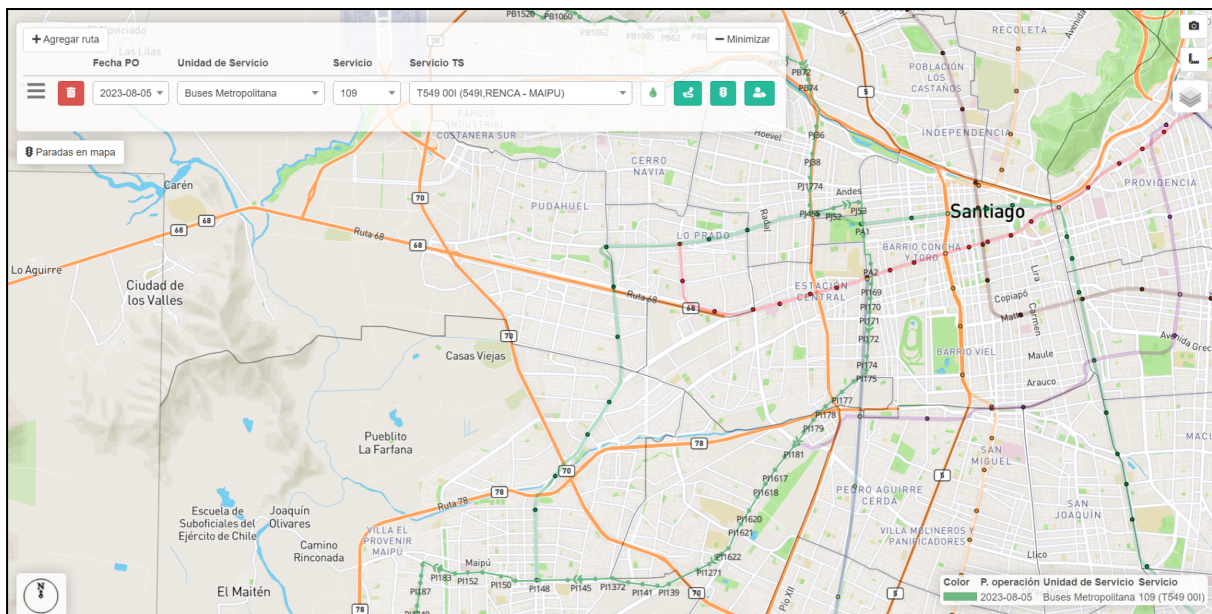


Ilustración 2.6.3.1.1: Reordenar listado de recorridos

2.5.4 Agregar capas de salud y educación en el mapa

Parte de los análisis que realiza DTPM consisten en incorporar distintas variables espaciales en los servicios de transporte, según lo anterior, es necesario que la ficha de servicios incluya una capa con los centros de salud existentes y también, otra capa con los centros de educación. La información de ambas capas fue dada por DTPM, y será una fuente de datos estática.

2.5.4.1 Implementación

Las capas de datos fueron entregadas en formato shp, formato que no es compatible con la librería de visualización espacial, por lo que un paso previo fue realizar la conversión de este formato al formato geojson.

Una vez realizada la conversión espacial, el siguiente paso fue incorporar las fuentes de datos al proyecto e incorporar los selectores de estos datos al control de capas del mapa. Ambas capas muestran sus elementos como un círculo (verde para los centros de salud y amarillo para los centros de educación) con borde negro, y además, abren un popup con el nombre de cada centro al pasar el mouse por encima de alguno de los puntos.

En la Ilustración 2.6.4.1.1 se muestra la capa de centros de educación existentes en los datos dados por DTPM. En la Ilustración 2.6.4.1.2 se muestra la burbuja informativa de un centro de educación, en ella se muestra el nombre del establecimiento.

En la Ilustración 2.6.4.1.3 se muestra la capa de centros de salud existentes en los datos. Además, en la Ilustración 2.6.4.1.4 se muestra la burbuja informativa de un centro de salud, en ella se muestra el nombre del establecimiento según está declarado en los datos.

Por último, en la Ilustración 2.6.4.1.5 se muestra el control que permite mostrar u ocultar la capa de centros de educación y la capa de centros de salud. Por defecto, ambas capas están ocultas y el usuario debe presionar cada botón para hacer visible los datos en el mapa.

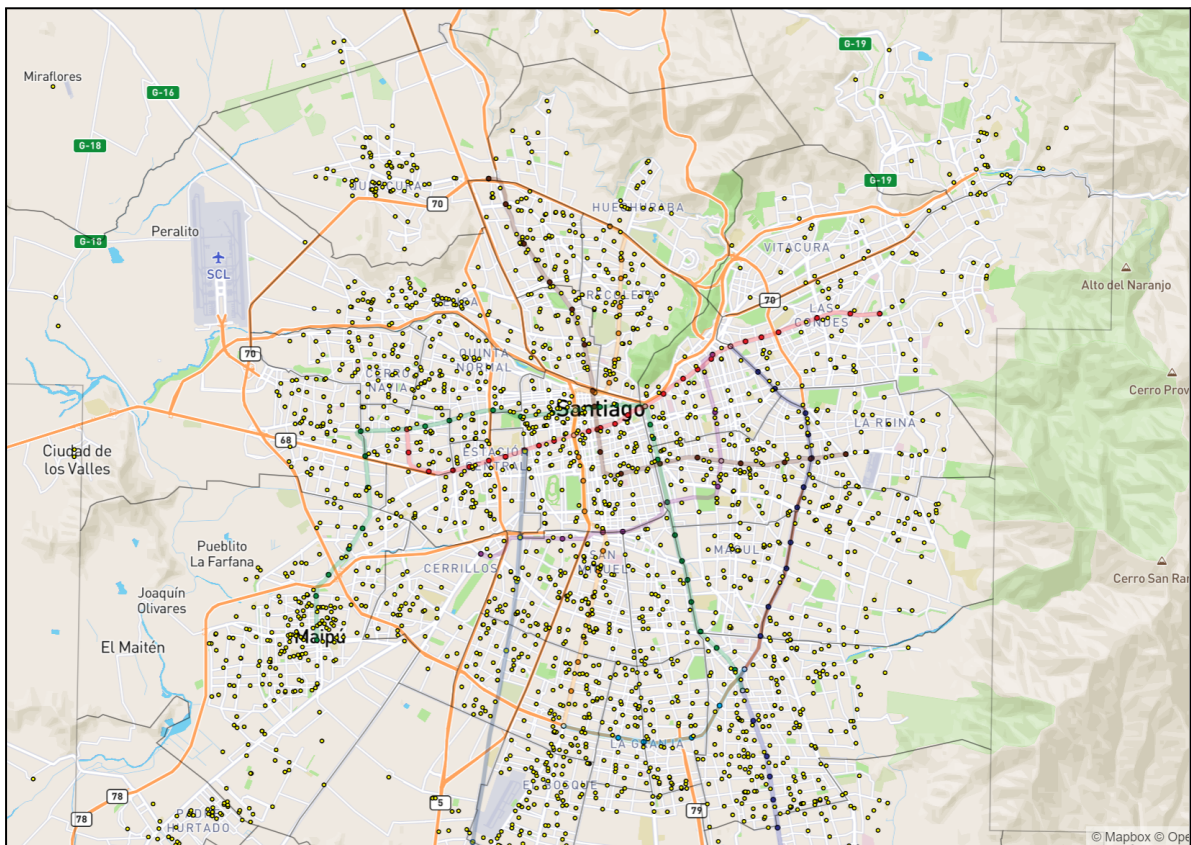
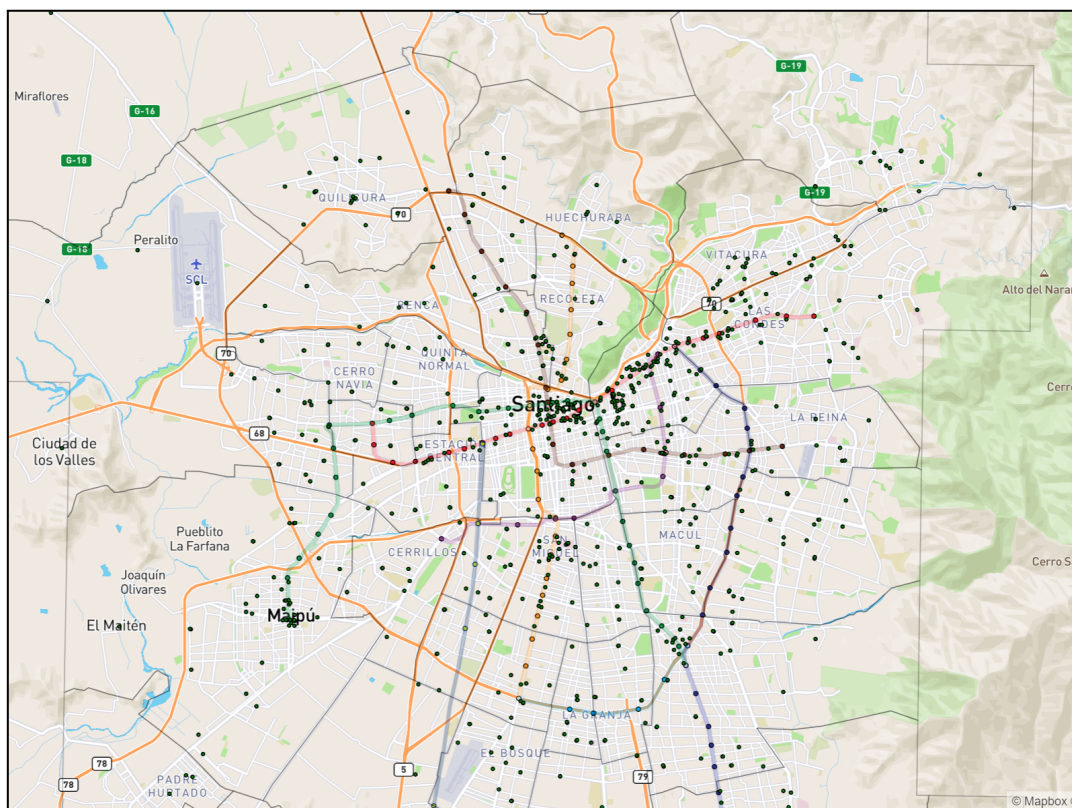
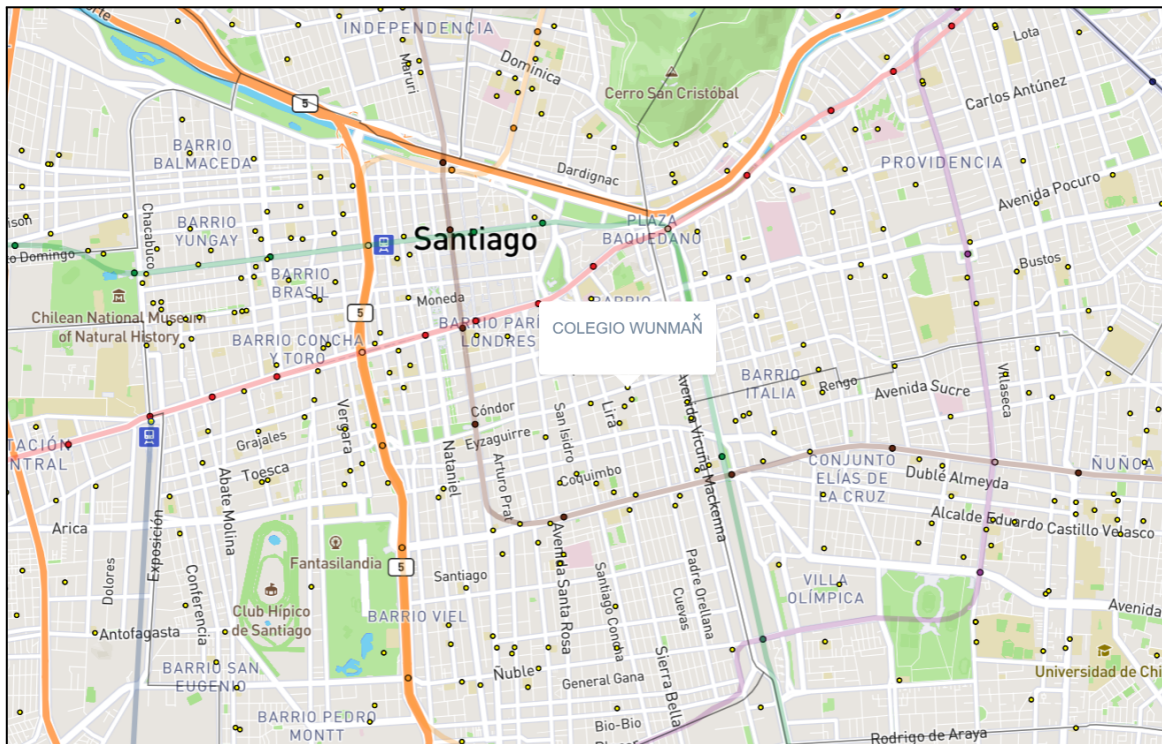


Ilustración 2.6.4.1.1: Capa de centros de educación



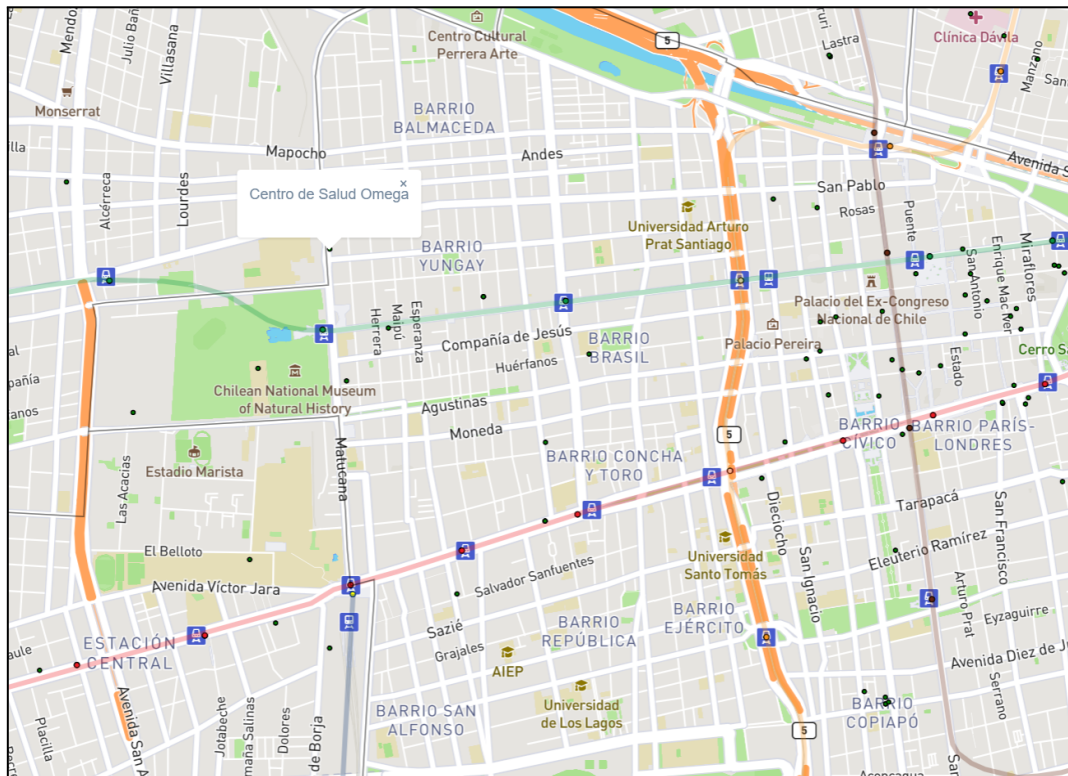


Ilustración 2.6.4.1.4: Burbuja de información de un centro de salud

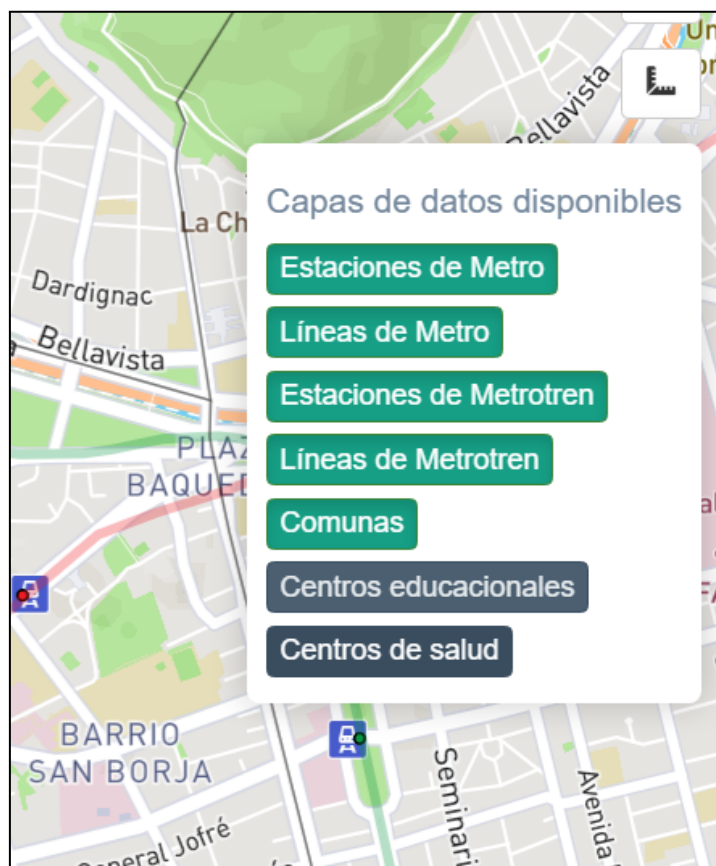


Ilustración 2.6.4.1.5: Control para mostrar u ocultar las nuevas capas de datos. Por defecto, están ocultas

2.5.5 Incluir trazados de variantes aunque no tengan paradas

Actualmente, cuando los recorridos no tienen definida una secuencia de paradas, sus trazados no son visibles en el mapa, esto es un error puesto que a pesar de ese dato faltante, el trazado puede ser mostrado. Según lo anterior, se solicita al consultor que pueda permitir visibilizar los recorridos en esta situación.

2.5.5.1 Implementación

Cuando un servicio seleccionado por el usuario no contiene información de su secuencia de paradas, se muestra un mensaje flotante informando la situación. En el comportamiento previo a este desarrollo, también genera que ninguna información se muestre en el mapa, a pesar que existe su trazado.

El nuevo comportamiento mantiene el mensaje flotante e incorpora el dato de trazado al mapa. En la Ilustración 2.6.5.1.1 se muestra el nuevo resultado: un mensaje flotante con un trazado (de color morado) en el mapa pero sin información de su secuencia de paradas.

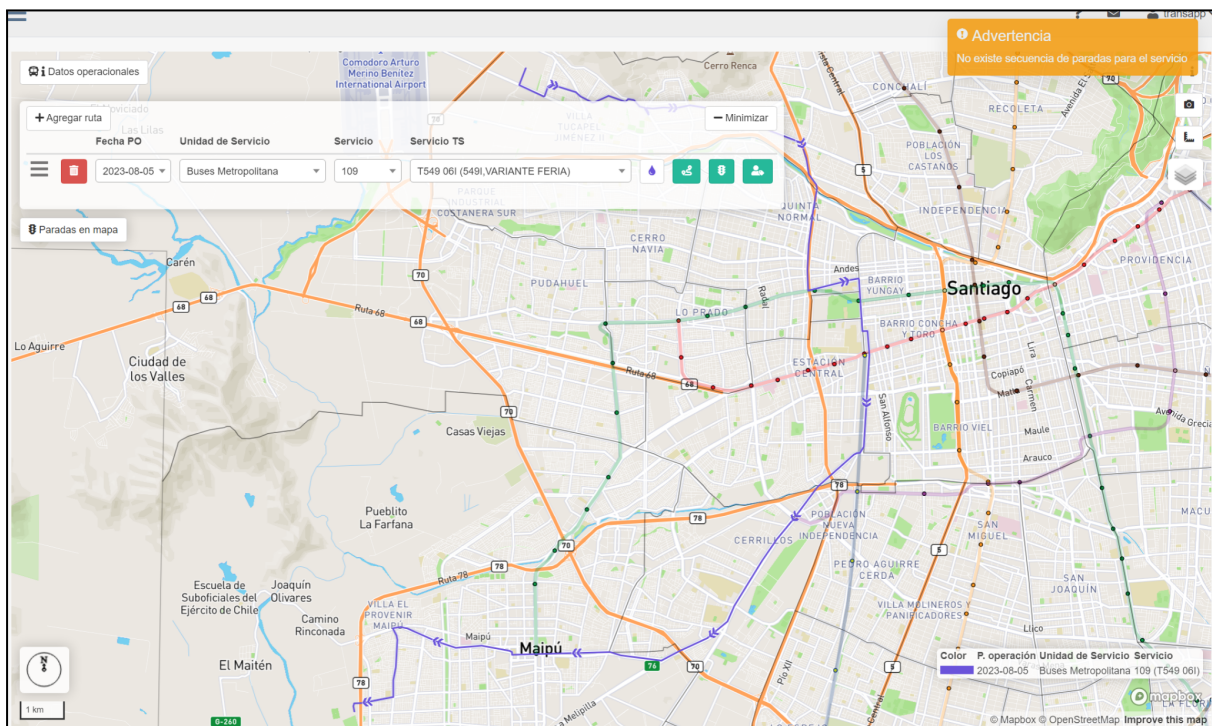


Ilustración 2.6.5.1.1: Trazado en mapa sin secuencia de paradas más mensaje informativo

2.5.6 Agregar nombre de servicio/variante a cada ruta seleccionada

En el diccionario de servicios existe el campo “service_NA” que permite reconocer el variante asociado al recorrido descrito en cada registro del diccionario. Se solicita al consultor incluir esta información como un campo informativo en el control de selección de recorridos.

2.5.6.1 Implementación

El primer cambio realizado fue modificar el proceso de carga del diccionario de servicio de los programas de operación. La modificación fue incorporar el nuevo campo (“service_NA”) al modelo de datos de la herramienta de visualización y que este valor sea actualizado cuando se cargue un nuevo archivo.

El segundo cambio fue incorporar el nuevo campo a la vista “Ficha de servicios”. En este lugar se agregó la información al selector “Servicio TS”, por lo que cada elemento disponible en el selector muestra el valor del campo “Service_NA” entre paréntesis.

En la Ilustración 2.6.6.1.1 se muestra el selector “Servicio TS” con el valor “Service_NA” como parte de su valor. En el ejemplo, el servicio T549 06I tiene el valor “VARIANTE FERIA” en el campo “Service_NA”.

Por último, fue necesario volver a cargar los diccionarios de servicios para todos los programas de operación para que se inserten los nuevos datos.

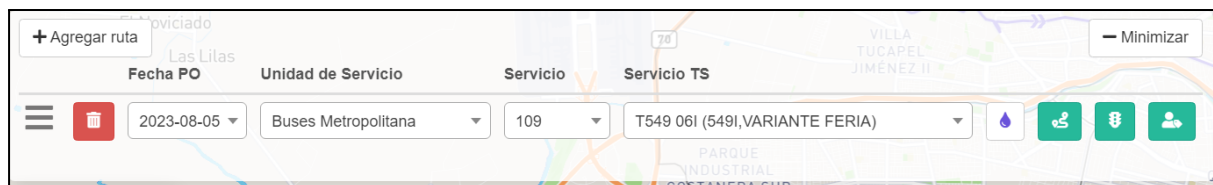


Ilustración 2.6.6.1.1: Incorporación de campo “Service_NA” en selector “Servicio TS”

2.6 Mejora filtro Perfil de carga

La vista de perfil de carga disponible en visualizador web de ADATRAP, muestra las cargas que experimenta un conjunto de expediciones que realizan el mismo servicio-sentido. La agrupación de expediciones se realiza a través de un filtro que permite indicar los siguientes parámetros:

- Periodo de días
- Tipo de día
- Periodo de despacho
- Media hora de despacho
- Operador
- Servicio usuario
- Servicio TS (código Sonda)

Lo anterior se complementa con una visualización que permite ver un gráfico de líneas junto con una representación espacial del perfil de carga a lo largo de su ruta junto a sus paradas.

Se solicitó al consultor mejorar el filtro de operador puesto que en algunas ocasiones no se conoce, en concreto se solicita incorporar una opción de nombre “Todos” que mostrará en el siguiente filtro “Servicio usuario” todos los servicios usuarios disponibles en el periodo. Adicionalmente, se solicita ordenar alfabéticamente la lista de servicios.

2.6.1 Implementación

Se modificó el endpoint que entrega el listado de servicios disponibles incorporando una opción adicional (“Todos”) en el selector de operador que agrupa todos los servicios. Además, esta opción fue definida como la opción por defecto, por lo que la vista comienza configurada en ese estado.

La opción de filtrar los servicios por operador sigue disponible, y el usuario puede hacer uso de ella si lo estima conveniente.

En la Ilustración 2.7.1.1 se muestra el resultado del selector, en donde aparece la opción “Todos” como el primer elemento, lo que habilita todos los servicios en el selector “Selector usuario”.

The image shows a web application filter interface titled 'Filtro'. It contains several input fields: 'Días' with '1 día seleccionado', 'Tipo de día' with 'Todos', 'Periodo de despacho' with 'Todos', and 'Media hora de despacho' with 'Todos'. Below these are three dropdown menus: 'Operador', 'Servicio usuario', and 'Servicio TS'. The 'Operador' dropdown is open, showing a list of options: 'Todos' (highlighted in blue), 'Su Bus', 'Buses Vule', 'Voy Santiago', 'Buses Metropolitana', 'STP Santiago', and 'Metropolitano'. The 'Servicio usuario' dropdown shows 'G12', and the 'Servicio TS' dropdown shows 'T472 00I (G12I)'. An orange button labeled 'Descargar datos crudos' is located to the right of the dropdowns. The interface has a clean, modern design with a light gray background and white input fields.

Ilustración 2.7.1.1: Selector de operador con opción “Todos”