

Procesamiento y disponibilización de Datos de ADATRAP (Etapa III)

Informe N°1 – OT5

Responsables: Mauricio Zúñiga y Felipe Hernández

28/7/2023

Tabla de Contenido

RESUMEN EJECUTIVO	3
1. INTRODUCCIÓN	4
2. PROCESAMIENTO	5
2.1 DATOS DE ENTRADA	5
2.1.1 Los registros GPS	5
2.1.2 Transacciones Bip	5
2.1.3 Reporte 1.96	6
2.1.3 Información del Programa de operación	6
2.2 VALIDACIÓN DATOS DE ENTRADA	7
2.2.1 Criterio GPS	8
2.2.2 Criterio Transacciones Bip!	8
2.2.3 Criterio Reportes 1.96	9
2.3 DATOS DE SALIDA	9
2.3.1 Criterio Estimación de bajada	10
2.3.2 Bases de datos de salida ADATRAP	11
2.3.3 Velocidades del sistema	12
2.3.4 Validación general	12
3. ALMACENAMIENTO	14
4. MANTENCIÓN	16
4.1 Sistema de Visualización	16
4.1.1 Mantención visualizador ADATRAP	17
4.1.2 Mantención sistema de extracción de Datos	17
4.1.3 Mantención sistema de visualización ambiente desarrollo y producción	18
4.1.4 Actividad de usuarios	18

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 N° 1 de los resultados de la orden de trabajo número cinco 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 esta regularización del servicio 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 cubrir el procesamiento y disponibilización de los datos desde el 1 de mayo de 2023 al 30 de junio de 2023. No considera horas de desarrollo.

Esta se desarrolla en un contexto de post pandemia con un notable impacto en la movilidad de los usuarios de transporte público, lo que se tradujo en una disminución considerable en el número de etapas y viajes realizados en el sistema RED, lo que implicó una dificultad en la detección de problemas que están más allá del fenómeno del transporte, es por esto que en este informe se muestran todos los datos procesados sin considerar criterios de detección automáticas de anomalías en los datos.

Los resultados del proceso y carga de los datos en el visualizador de ADATRAP son descritos en detalle en los capítulos de este informe con el objeto de dejar constancia de todos los procedimientos y circunstancias especiales que pudieron ocurrir en este periodo.

Como conclusión general este periodo no tuvo inconvenientes relevantes.

Este informe presenta nuestro servicio, con el que esperamos proveer al Ministerio de Transporte y Telecomunicaciones con datos y una interfaz para interactuar con ellos que permita tomar decisiones a corto y mediano plazo en la operación del sistema RED Metropolitana de Movilidad.

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 fue 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”.

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 se divide en tres capítulos, en el primero se analiza la disponibilidad de datos de entrada y la factibilidad de poder procesarlos para producir los datos resultantes, el segundo capítulo tiene por objeto detallar los datos que quedaron almacenados en este periodo y el tercero describe las actividades realizadas para la mantención del sistema.

2. PROCESAMIENTO

En este capítulo se analizan los datos de entrada y salida del programa ADATRAP durante el periodo requerido, del 1 de mayo de 2023 al 30 de junio de 2023, con el fin de identificar posibles anomalías en ambas etapas.

Este análisis será separado en tres etapas, la primera consiste en una explicación y descripción de los datos, necesario para lectores que no conocen el sistema, la segunda etapa corresponde a una validación de estos datos, principalmente su existencia y cantidad en el tiempo, finalmente se realiza una validación similar para los datos de salida como lo son etapas, viajes, y perfiles de carga.

2.1 DATOS DE ENTRADA

Los datos de entrada se dividen en dos tipos, primero los datos diarios donde están los GPS, las transacciones bip y la asignación de buses a servicios (reporte 1.96), por otro parte, están los datos asociados a la programación de la operación estos datos son los datos de la red como paraderos, trazados, frecuencias y datos específicos como las zonificaciones, diccionarios de servicios, datos de estaciones de metro, metro tren entre otros.

2.1.1 Los registros GPS

Archivo diario de posicionamiento de los buses del sistema. Cada fila o registro de este archivo corresponde a un pulso GPS, sus campos deben contener latitud, longitud, fecha y hora, servicio, sentido y patente. Estos registros son emitidos cada 30 segundos por todos los buses del sistema. Estos datos son obtenidos desde un servidor directo en DTPM el cual se actualiza cada un minuto, de esta forma los datos son almacenados en archivos diarios de nombres con el formato yyyy-mm-dd.gps.

Durante este periodo se utilizó una fuente de datos GPS proveniente de un sistema de almacenamiento histórico que tiene DTPM. Se tuvo que realizar un procesamiento adicional para dar formato a estos datos al estándar de ADATRAP.

2.1.2 Transacciones Bip

La segunda fuente de datos corresponde a las transacciones Bip! del sistema. Estas transacciones son registradas para cada subida en bus o metro por los validadores respectivos. El 95% de las transacciones está disponible a partir del cuarto día de realizada la transacción. Una transacción consiste en id de la tarjeta, tiempo de la transacción y sitio donde se realizó la transacción (patente de bus, código validador o estación de metro). Estos datos son obtenidos en forma directa desde un servicio FTP que provee DTPM, cada día es recibido un conjunto de registros de transacciones de

diferentes días, de esta forma son almacenados en una única base de datos de la cual se extrae todos los días el día once días hacia atrás.

2.1.3 Reporte 1.96

La tercera fuente de datos corresponde a la tabla de asignación de servicio a las expediciones. Cada registro de esta tabla contiene la información de servicio al cual está asignada la patente del bus, además incorpora los tiempos de inicio de esa ruta y el tiempo en el cual finaliza. Con este dato se determina a qué servicio se asignan las transacciones Bip! realizadas en los buses. Este dato es entregado a través de un FTP de DTPM donde es actualizado en forma quincenal por el personal encargado.

2.1.3 Información del Programa de operación

DTPM facilita una serie de archivos de entrada para el funcionamiento de ADATRAP. La periodicidad con que este cambia es variable. Las fechas de entrega de programas de operación correspondientes al periodo de esta orden de trabajo se listan a continuación en formato aaaa/mm/dd desde el más nuevo al más antiguo: 2023/06/10, 2023/05/13, 2023/05/02 y 2023/04/22

Tabla 2.1.3.a : Días procesados por cada programa de operación

PO	Días procesados
2023/06/10	2023/06/10 al 2023/06/30
2023/05/13	2023/05/13 al 2023/06/09
2023/05/02	2023/05/02 al 2023/05/12
2023/04/22	2023/04/22 al 2023/04/30
2023/04/08	2023/04/08 al 2023/04/21
2023/04/01	2023/04/01 al 2023/04/07

Los tipos de archivos que se facilitan se dividen en 7 carpetas. A continuación, la Tabla 2.1.3.b muestra un detalle de cada una y los archivos que contienen.

Tabla 2.1.3.b: Listado de archivos provenientes del programa de operación.

Carpeta	Nombre de archivo	Descripción
Diccionarios	Diccionario-Comunas.csv	Contiene id y nombre de comunas de Santiago.
Diccionarios	Diccionario-DetalleServicioZP_fecha1_fecha2.csv	Contiene id de ZP, nombre, ubicación, tipo de ZP y servicios. Se facilita cada 15 días dentro del PO.
Diccionarios	Diccionario-EstacionesMetro.csv	Contiene id de TRX de la estación, comuna, ubicación, línea, tipo, código y color.
Diccionarios	Diccionario-EstacionesMetroTren.csv	Contiene id de TRX de la estación, comuna y ubicación.
Diccionarios	Diccionario-Patentes.csv	Contiene folio, unidad, descripción tipo de vehículo, patente para cada bus.
Diccionarios	Diccionario-PeriodosTS.csv	Contiene id, tipo de día, período, hora inicio y hora fin de periodos definidos por DTPM.

Diccionarios	Diccionario-Servicios.csv	Contiene id de la ruta, nombre, servicio para todos los servicios de la red.
Diccionarios	Diccionario-Zonificaciones.csv	Contiene descripción de distintos tipos de zonificación.
Evasión	EvasionServicioSentidoParadaMH.csv	Contiene factores de evasión por parada, servicio, período.
Evasión	Zonas777Fevasion.csv	Contiene factores de evasión por zona 777.
Frecuencias	Frecuencias.csv Capacidad.csv Distancia.csv Velocidad.csv	Contiene los parámetros por servicio, sentido, período, tipo de ruta, tipo de día, definidos en el Programa de Operación.
Metro	tiempos_metro.csv	Contiene tiempos de acceso y espera entre estaciones de metro.
Paraderos	ConsolidadoParadas.csv	Contiene descripción de paradas, nombre, ubicación, si es ZP, cruce cercano.
Rutas	ShapeRutas.csv	Contiene id, nombre de la ruta, coordenadas que la describen para cada ruta del sistema.

Estos datos son entregados a través de un FTP en la medida que se considere que el programa de operación haya tenido cambios suficientes que ameriten su actualización, esta decisión es tomada por DTPM.

2.2 VALIDACIÓN DATOS DE ENTRADA

En esta sección se determinan cuáles días procesados son factibles de subir a la plataforma de visualización de ADATRAP. Se deben cumplir los siguientes criterios para que el día sea considerado válido.

1. Que el día tenga datos de GPS, transacciones y reporte 1.96.
2. Que el día cumpla con los criterios establecidos para cada dato de entrada descritos más adelante
3. Que el día cumpla con los criterios de estimación de bajada.

Los días a revisar en este informe van desde el 1 de mayo de 2023 al 30 de junio de 2023. Como referencia se grafican los valores desde mayo del año 2022 hasta la fecha.

Durante este periodo no se encontraron días con problemas, por lo que todos fueron procesados.

Criterios por tipo de dato

En esta orden de trabajo, y en general para todas las órdenes de trabajo en periodo de pandemia, se realiza un chequeo visual de las variables y utilizando un criterio de común acuerdo con DTPM se definen los días que quedan fuera.

2.2.1 Criterio GPS

Los valores normales de pulsos totales de GPS son del orden de los 13.000.000 registros para día laboral, 10.500.000 de registros para los días sábados y 9.800.000 de registros los días domingos. Estos valores fueron estimados en base a la información de agosto 2019 hasta marzo 2020.

En la Ilustración 2.2.1 se observa la cantidad de registros de GPS entre el 1 de mayo de 2022 al 30 de junio de 2023. En el corte procesado en la presente OT, no se aprecian anomalías en los datos. Solo mencionar que la discontinuidad en junio corresponde al feriado del día 20.

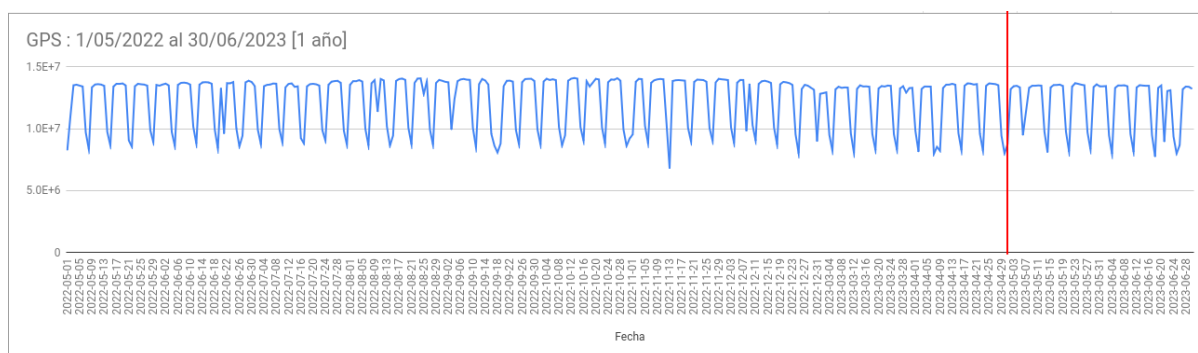


Ilustración 2.2.1: Curva de número de pulsos GPS diarios, desde el 1 de mayo de 2022 al 30 de junio de 2023.

2.2.2 Criterio Transacciones Bip!

Los valores normales totales de transacciones Bip! al día están en el orden de 6.000.000 de transacciones para día laboral, 3.000.000 de transacciones para día sábado y 1.800.000 de transacciones para día Domingo. Esto permite, al igual que en los datos GPS, determinar días anómalos cuando hay diferencias significativas en este número de registros. Estos valores fueron estimados en base a la información de agosto 2019 hasta marzo 2020.

En la Ilustración 2.2.2 se observa la cantidad de transacciones Bip! entre el 1 de mayo de 2022 al 30 de junio de 2023. El corte no presenta problemas en los datos. Solo mencionar la discontinuidad en junio por el día feriado.

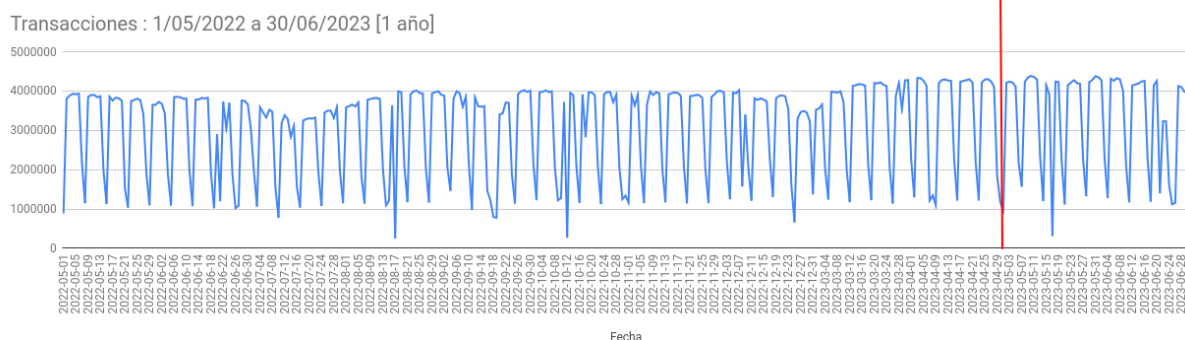


Ilustración 2.2.2: Curva de número de número de transacciones diarias, desde el 1 de mayo de 2022 al 30 de junio de 2023.

2.2.3 Criterio Reportes 1.96

De la misma manera que las fuentes de datos anteriores, la cantidad de registros es bastante estable en el tiempo. Los valores normales de expediciones al día están en el orden de 75.000 a 77.000 para día laboral, entre 57.000 a 60.000 para día sábado y entre 45.000 a 49.000 para día Domingo. Estos valores fueron estimados en base a la información de marzo 2019 hasta noviembre 2019.

En la Ilustración 2.2.3 se observa el número de expediciones entre el 1 de mayo de 2022 al 30 de junio de 2023. No se aprecia ninguna anomalía, solo una discontinuidad en junio que corresponde al día feriado (20 de junio).

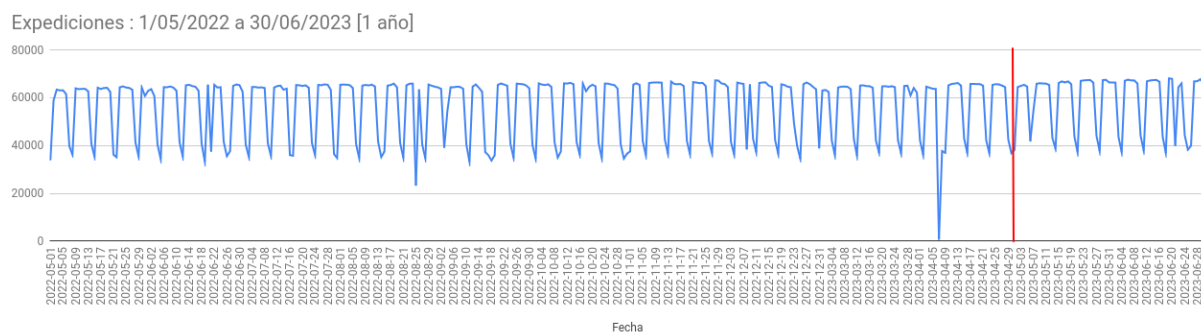


Ilustración 2.2.3: Curva de número de número de expediciones diarias, desde el 1 de mayo de 2022 al 30 de junio de 2023.

2.3 DATOS DE SALIDA

Para los días procesados, se obtienen un conjunto de resultados. Estos resultados también tienen un comportamiento similar para días del mismo tipo (laboral, sábado y domingo), por lo que también es posible realizar un análisis para identificar días anómalos revisando algunos indicadores. En este caso se revisarán los criterios de estimación de bajada y velocidades.

2.3.1 Criterio Estimación de bajada

Uno de los procesos más relevantes realizados por ADATRAP, es la estimación de bajada de los usuarios.

El sistema de transporte público en Santiago solo exige validar para el pago al subir a un bus o entrar en una zona paga, por lo que el paradero de bajada del usuario de esa tarjeta es un dato no entregado por el sistema y debe estimarse para poder construir la información de viajes. Por lo anterior, se desarrolló una metodología que estima estos paraderos de bajada que consiste a grandes rasgos en determinar, según el paradero de subida siguiente, el paradero de bajada anterior. Esta metodología tiene variantes según si el usuario viaja en Metro, Metro Tren, Bus y en caso de ser bus si la validación fue realizada directamente sobre el bus o en una zona paga, el detalle de esta metodología está descrita en Munizaga y Palma (2012).

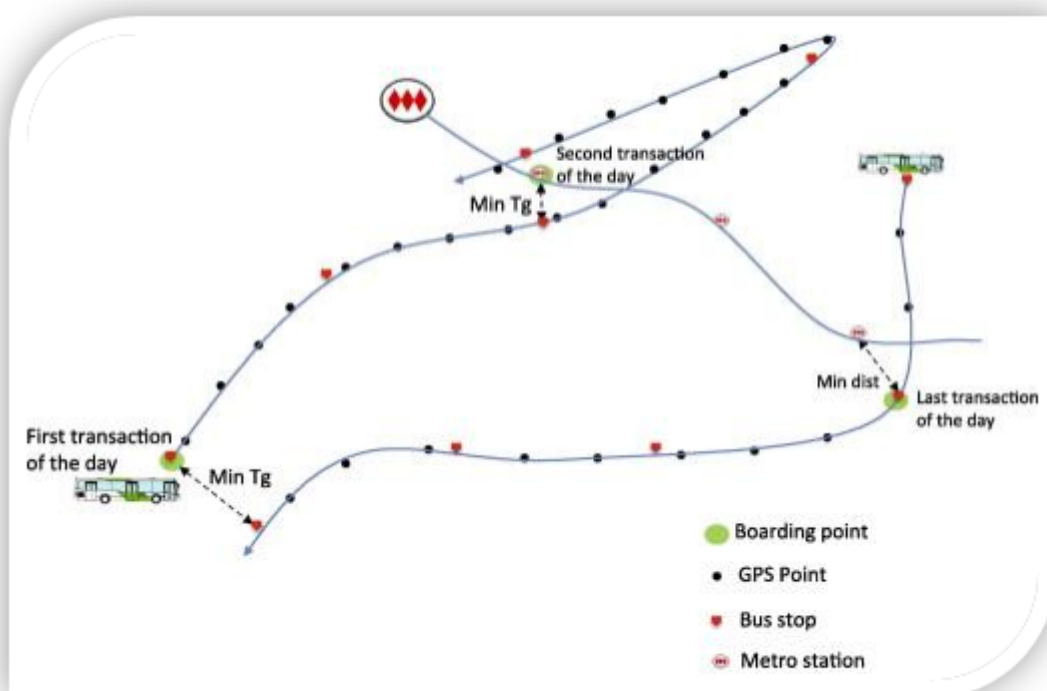


Ilustración 2.3.1.1: Esquema que describe la estimación de bajada para un pasajero.

Esta metodología asume que el pasajero realiza solo viajes en transporte público integrado al modo de pago bip!, en caso contrario la metodología dejará la etapa anterior sin bajada estimada.

A la fecha el mejor indicador de algún problema en el procesamiento corresponde al porcentaje de estimación de bajada en los distintos modos de transporte público. El primero corresponde al modo bus. Normalmente la estimación de bajada en este modo para un día laboral es del orden del 74%, y para sábado-domingo alrededor de 65%. El

segundo modo corresponde a Metro el cual para un día laboral tiene un 87% de estimación de bajada y para fines de semana entre un 79% a 81%. El tercer modo es Metro-Tren con una estimación de bajada para día laboral del 82% y un 78% a 80% para fines de semana. El último modo corresponde a validaciones hechas en zonas pagas donde la estimación de bajada en un día laboral ronda el 67% y 56% a 58% para fines de semana.

La Ilustración 2.3.1.2 muestra los desde el 1 de mayo de 2022 al 30 de junio de 2023. El corte no presenta anomalías.

Estimación de bajada : 1/5/2023 a 30/6/2023 [1 año]

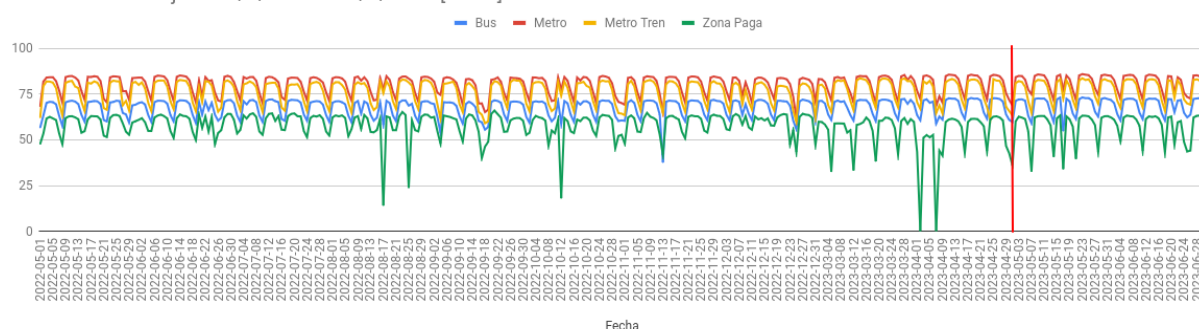


Ilustración 2.3.1.2: Curva de porcentaje de estimación de bajada, rojo para metro, amarillo para metro-tren, azul para buses y verde para zonas pagas entre el 1 de mayo de 2022 al 30 de junio de 2023

2.3.2 Bases de datos de salida ADATRAP

La salida del sistema no es única sino un conjunto de archivos, estos archivos contienen datos como las etapas de viajes, los viajes, los perfiles de carga, también sumamos como salida del sistema los datos de entrada que son procesados y estandarizados por el mismo flujo del proceso como los son las transacciones, GPS y reportes 1.96, el estándar de sus nombres y su contenido se pueden apreciar en la Tabla 2.3.2.

Tabla 2.3.2: Descripción estándar de nombre y contenido de archivos de salida.

Tabla	Archivo	Descripción
Etapas	aaaa-mm-dd.etapas	Contiene una línea por cada etapa, modo de transporte, paradero/estación de metro de subida y bajada, tiempos asociados, factor de expansión, período.
Viajes	aaaa-mm-dd.viajes	Contiene una línea por cada viaje, detalle de cada etapa, modo de transporte, tiempos asociados, distancias, paradas/estación de subida y bajada.
Perfiles	aaaa-mm-dd.profile	Contiene una línea por bus-paradero, servicio, subidas, bajadas, correlativo en la ruta, distancia en ruta.
Transacciones	aaaa-mm-dd.trx	Contiene los datos de transacciones en un formato estandarizado (id, tiempo, sitio, operador)
GPS	aaaa-mm-dd.gps	Contiene los datos de GPS en un formato estandarizado (patente, servicio, tiempo, latitud, longitud), el servicio corresponde al servicio original asignado en el dato GPS

1.96	aaaa-mm-dd.196	Contiene la información de asignación de servicio a los buses
Pasadas	aaaa-mm-dd.stoptime	Contiene los datos de tiempos de pasadas de los buses sobre las paradas en sus expediciones con servicio-sentido

2.3.3 Velocidades del sistema

Un segundo producto entregado por ADATRAP son las velocidades del sistema, estas velocidades corresponden a las velocidades medidas para cada uno de los servicios sobre tramos de 500 metros en la ruta en periodos de 30 minutos. En el caso de estos datos no existe una validación por el número de registros, los problemas que impactan en este proceso están directamente relacionados a los datos de GPS de los buses y las rutas del sistema. Ambas fuentes de datos son revisadas en las secciones anteriores.

Para el corte temporal en cuestión, se cuenta con los valores de velocidad para todos los días laborales que no son descartados en los procesos anteriores. La Tabla 2.3.3 describe el estándar de los nombres de los archivos y su contenido.

Tabla 2.3.3: Descripción de tablas asociadas a datos de velocidades.

Tabla	Archivo	Descripción
Velocidades	aaaa-mm-dd.speed	Contiene una línea por sección de ruta, fecha, tipo de día, distancia, tiempo, velocidad, operador. Un archivo por día del mes.
Velocidades	velocidadesaaaa-mmMes.velocidadesAgregadas	Velocidad, distancia y tiempo por servicio-período-tramo-día.
Velocidades	velocidadesaaaa-mmMes.velocidadesFull	Velocidad, distancia y tiempo por período-tramo-día.

2.3.4 Validación general

El proceso considera dentro de sus salidas un archivo con un conjunto de campos que contienen agregaciones de los distintos tipos de datos calculados, esto con el fin de poder evaluar si algún día en particular tiene alguna anomalía en alguna de sus líneas de cálculos. La Tabla 2.3.4 muestra dos ejemplos de valores de los campos, y una descripción del campo, el ejemplo 2 muestra el caso de un día que tuvo algún problema en su ejecución por la cual perdió toda la información, estos días son necesario revisar en particular, normalmente están asociados a alguna inexistencia en los datos base como son GPS, transacciones o reportes 1.96.

Tabla 2.3.4: Descripción de tablas de validación general

Campo	Ejemplo	Descripción
Fecha	2018-03-01	Fecha del día
Tipodia	LABORAL	Tipo de día (LABORAL, SABADO, DOMINGO)
Version	v1.0.0	Versión de software de procesamiento de validaciones
nExpediciones	73794.0	Cantidad de expediciones
minTiempoExpediciones	0.5	duración mínima de una expedición [min]
maxTiempoExpediciones	226.0	Duración máxima de una expedición [min]

mediaTiempoExpediciones	52.8	Duración promedio de expediciones [min]
nPatentes	6578.0	Número de patentes, cantidad de buses
nGPS	11809695.0	Cantidad de pulsos gps
mediaTiempoEntreGPS	47.7	Promedio de diferencia entre pulsos gps [s]
nGPSConServicio	7415077.0	Cantidad de pulsos GPS con servicio-sentido asignado
nGPSSinServicio	4264304.0	Cantidad de pulsos GPS sin servicio-sentido asignado
nTrxTotales	4921137.0	Cantidad total de transacciones Bip!
nTrxTotalesBus(%)	47.2	% de transacciones hechas en Bus
nTrxTotalesMetro(%)	43.0	% de transacciones hechas en Metro
nTrxTotalesMetroTren(%)	1.0	% de transacciones hechas en Metro-Tren
nTrxTotalesZonasPagas(%)	8.7	% de transacciones hechas en Zonas Pagas
nTarjetas	1762677.0	Cantidad total de tarjetas distintas
nTrxsConServicio	2284754.0	Cantidad de transacciones con servicio estimado
nTrxsSinServicio	36226.0	Cantidad de transacciones sin servicio estimado
nEtapasConBajadaBus(%)	73.8	% de etapas con bajada estimada en Bus
nEtapasConBajadaMetro(%)	86.0	% de etapas con bajada estimada en Metro
nEtapasConBajadaMetroTren(%)	78.8	% de etapas con bajada estimada en Metro-Tren
nEtapasConBajadaZP(%)	67.0	% de etapas con bajada estimada en Zonas Pagas
nViajes	3635268.0	Cantidad total de viajes estimados
nViajes1E(%)	69.4	% de viajes con una etapa
nViajes2E(%)	26.2	% de viajes con dos etapas
nViajes3E(%)	4.4	% de viajes con tres etapas
nViajes4E(%)	0.1	% de viajes con cuatro etapas
nViajes5E(%)	0.0	% de viajes con cinco etapas
nViajesSoloMetro(%)	36.3	% de viajes hechos solo en Metro
nViajesConAlgunaEtapaEnMetro(%)	58.2	% de viajes hechos con alguna etapa en Metro
nViajesSinBajadaFinal(%)	26.7	% de viajes sin estimación de bajada en etapa final
tViajeTotal	25.4	Promedio de tiempos de viaje
dViajeTotal	7383.8	Promedio de distancias de viaje
vViajeTotal	17.4	Promedio de velocidades de viajes
nViajeMediaPM	594823.0	Cantidad de viajes en Punta Mañana
tViajeMediaPM	30.1	Promedio de tiempos de viaje en Punta Mañana
dViajeMediaPM	8729.8	Promedio de distancias de viaje en Punta Mañana
vViajeMediaPM	17.4	Promedio de velocidades de viaje en Punta Mañana
nViajeMediaPT	872358.0	Cantidad de viajes en Punta Tarde
tViajeMediaPT	29.7	Promedio de tiempos de viaje en Punta Tarde
dViajeMediaPT	7913.7	Promedio de distancias de viaje en Punta Tarde
vViajeMediaPT	16.0	Promedio de velocidades de viaje en Punta Tarde
nTrxPM(%)	17.6	Cantidad de validaciones Bip! en Punta Mañana
nTrxPT(%)	23.9	Cantidad de validaciones Bip! en Punta Tarde
nBajadas	4927440.5	Cantidad de etapas con bajada estimada
nBajadasPM	768483.8	Cantidad de etapas con bajada estimada en Punta Mañana
nBajadasPT	1174508.0	Cantidad de etapas con bajada estimada en Punta Tarde
nParadasE	220.0	Cantidad de paraderos de Tipo E
nParadasT	3480.0	Cantidad de paraderos de Tipo T
nParadasL	7179.0	Cantidad de paraderos de Tipo L
nParadasI	5.0	Cantidad de paraderos de Tipo I
nTrxE(%)	9.1	Porcentaje de transacciones Bip! en paraderos Tipo E
nTrxT(%)	26.7	Porcentaje de transacciones Bip! en paraderos Tipo T
nTrxL(%)	18.3	Porcentaje de transacciones Bip! en paraderos Tipo L

nTrxl(%)	1.7	Porcentaje de transacciones Bip! en paraderos Tipo I
PO	PO20210101	Código del programa de operación usado para el día

3. ALMACENAMIENTO

El almacenamiento de los datos se hace en un repositorio en la nube que puede ser accedido a través de la sección de almacenamiento del visualizador de datos de ADATRAP (<https://www.adatrap.cl/>). En la Ilustración 3.1 se muestra la sección con todos los ítems disponibles.



Ilustración 3.1: menú de almacenamiento en visualizador de datos de ADATRAP

Cada ítem representa un repositorio con datos de un tipo, por ejemplo, el ítem “GPS” muestra una vista con todos los días con datos GPS disponibles, lo mismo ocurre con el ítem “Viajes” pero esta vez con los datos de viajes generados por ADATRAP. Existen algunos ítems de datos que no se asocian a días sino que a grupos de estos o descripción de datos que requieren trazabilidad en el tiempo como lo son los programas de operación. En la Tabla 3.1 se describe cada ítem del menú junto con su tipo (si se genera por día o por evento).

Tabla 3.1: Descripción de ítems disponibles en la sección de almacenamiento

Nombre	Tipo	Descripción
GPS	Por día	Pulsos GPS de todos los buses del sistema de transporte
Viajes	Por día	Viajes generados por ADATRAP

Prog. operación	Por evento	Programas de operación definidos
Semana rep.	Por evento	Semanas representativas
1.96	Por día	Archivos de despacho de buses
Perfiles de carga	Por día	Perfiles de carga generados por ADATRAP
Etapas	Por día	Etapas generadas por ADATRAP
Velocidades	Por día	Velocidades generadas por ADATRAP
Transacciones	Por día	Transacciones generadas en el sistema de transporte
Misceláneos	Por evento	Contiene dos archivos, el primero corresponde al número de viajes entre comunas por día desde el día 2019-01-01 a la actualidad. El segundo corresponde al número de transacciones por día
Subidas por parada	Por día	Número de subidas por cada paradero del sistema de transporte
Velocidades PO	Por evento	Velocidades de períodos
Cruce de buses por paraderos	Por día	Información del tiempo en que cada bus del sistema de transporte cruzó una parada de su servicio

En este almacén se guardan datos usados como entrada para ejecutar el software ADATRAP como datos de salida de este software. Los datos de entrada corresponden a GPS, reportes 1.96 y transacciones Bip!. Los datos de salida que se almacenan son las bases de datos diarias de viajes, etapas, perfiles de carga, velocidades y tiempos de pasadas por paraderos. Con respecto a los programas de operación, aún no se disponibilizan al igual que la semana representativa.

Los datos se almacenan de forma histórica y acumulativa. Aquí se encuentra información desde el año 2016 hasta la fecha, a diferencia de la visualización, donde solo se puede ver un año de datos. El repositorio de estos datos se encuentra alojado en AWS (Amazon Web Services) y puede ser accedido desde cualquier versión del visualizador (desarrollo o producción).

En la Ilustración 3.2 muestra una vista con datos de tipo “por día”, en la parte superior se muestra un calendario donde se destacan los días que contienen datos (en verde) y fechas relevantes como cambios en programas de operación o semana representativa. En la parte inferior de la vista se despliega una tabla con los días disponibles con la posibilidad de descargar cada uno de estos.

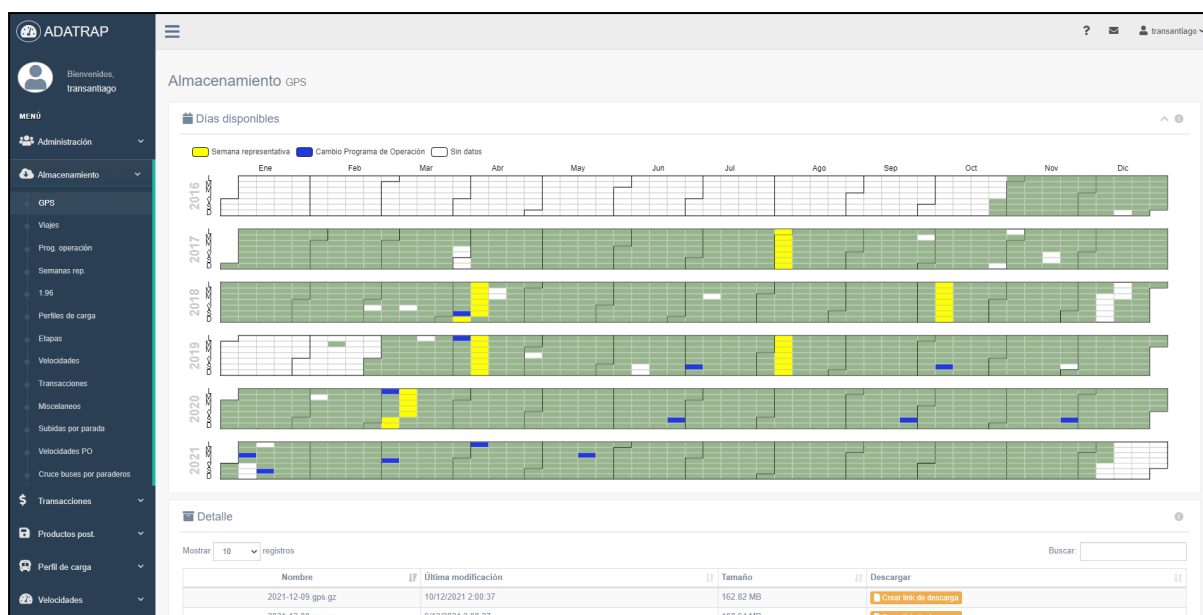


Ilustración 3:2 Vista de almacenamiento de datos de tipo “por día”

4. MANTENCIÓN

La mantención considera un constante monitoreo de la operación de los distintos sistemas, como el servidor de extracción de transacciones, el servidor de extracción de GPS y los distintos servidores involucrados para mantener el servicio operativo. Además, la mantención considera los desarrollos necesarios para adaptar el sistema a cambios en los formatos de los datos de entrada si es que existiesen.

4.1 Sistema de Visualización

El sistema de visualización consiste en una plataforma web que permite visualizar los datos generados por el procesamiento de ADATRAP, tiene restricciones de acceso por usuario, registro de actividad y además permite exportar los datos mostrados en las visualizaciones.

Los ocho datos resultantes del procesamiento de ADATRAP y que son utilizados para actualizar las vistas del visualizador se describen a continuación:

- **Perfiles de carga:** estimación de la capacidad de cada bus del sistema en cada paradero de la ciudad
- **Velocidades:** medidas estadísticas de las velocidades que presentan los servicios en tramos de 500 metros a lo largo de su ruta
- **Matriz de etapa por servicio:** estimación de las subidas y bajadas que experimenta un servicio a lo largo de su ruta
- **Viajes:** estimación de los viajes realizados en el sistema, contiene una estimación de lo que un pasajero realizó en función de los datos capturados
- **Estadísticas generales:** métricas asociadas a una ejecución de ADATRAP, que muestran estadísticas por día

- **Factor de pago:** cálculo que permite conocer el porcentaje de validaciones que se deben asociar a cada operador para realizar el pago de operación en una zona paga
- **Validaciones bip!:** contiene el número de validaciones realizadas en el sistema agrupadas por operador y períodos temporales
- **Etapas:** contiene todas las etapas detectadas por ADATRAP, un conjunto de una o más etapas representa un viaje

Cada tipo de dato descrito anteriormente tiene una o más visualizaciones en la plataforma web que permite a los usuarios resumir, visualizar y exportar millones de registros rápidamente.

La infraestructura está compuesta de un ambiente de desarrollo, alojado en la U. de Chile y otra de producción alojada en AWS (Amazon Web Service). El ambiente de producción está formado por cuatro máquinas donde tres de ellas tienen el objetivo de responder consultas de datos mientras que la cuarta es responsable del sitio web, que opera como intermediario entre los usuarios y el motor de consulta de datos.

4.1.1 Mantención visualizador ADATRAP

La Tabla 4.1.1.1 describe todas las actividades realizadas para los datos comprendidos entre el 1 de mayo de 2023 hasta el 30 de junio de 2023:

Tabla 4.1.1.1: Descripción de actividades del periodo.

Fecha	Acción	Ambiente	Datos involucrados	Periodo de datos
26/05/2023	Subir PO	Producción	Programa de operación	13/05/2023
17/07/2023	Subir datos	Producción	Perfiles de carga Viajes Velocidades Factores de pago Transacciones bip! Estadísticas globales Etapas	01/05/2023 - 31/05/2023
23/07/2023	eliminar datos	Producción	Perfiles de carga Viajes Velocidades Factores de pago Transacciones bip! Estadísticas globales Etapas	01/06/2022 - 30/06/2022
24/07/2023	Subir datos	Producción	Perfiles de carga Viajes Velocidades Factores de pago Transacciones bip! Estadísticas globales Etapas	01/06/2023 - 30/06/2023
27/07/2023	eliminar datos	Producción	Perfiles de carga Viajes Velocidades Factores de pago Transacciones bip! Estadísticas globales	01/07/2022 - 31/07/2022

			Etapas	
--	--	--	--------	--

4.1.2 Mantenimiento sistema de extracción de Datos

Las transacciones Bip fueron obtenidas a través del FTP disponibilizado por Sonda (usradatrap@200.29.15.112:21) para el sistema de subidas por bajada, las cuales fueron procesadas para estandarizar su formato al requerido por ADATRAP.

Por otra parte, los datos de GPS están siendo entregados por el DTPM de una fuente distinta a la fuente de extracción automática, esto porque después de un análisis se determinó que la extracción en línea tiene una pérdida de datos que puede afectar el rendimiento en la asignación de subida y estimación de bajada.

4.1.3 Mantenimiento sistema de visualización ambiente desarrollo y producción

Durante el periodo de medición no hubo inconvenientes en el ambiente de producción. En la Ilustración 4.1.3.1 se muestran tres máquinas del ambiente de producción (filas 2, 3 y 5) sin interrupciones (sin franjas rojas), mientras que la fila uno, que corresponde al servidor de visualización tuvo un problema de conectividad por dos minutos el día 11/07/2023 entre las 09:50 y 09:52 hrs..

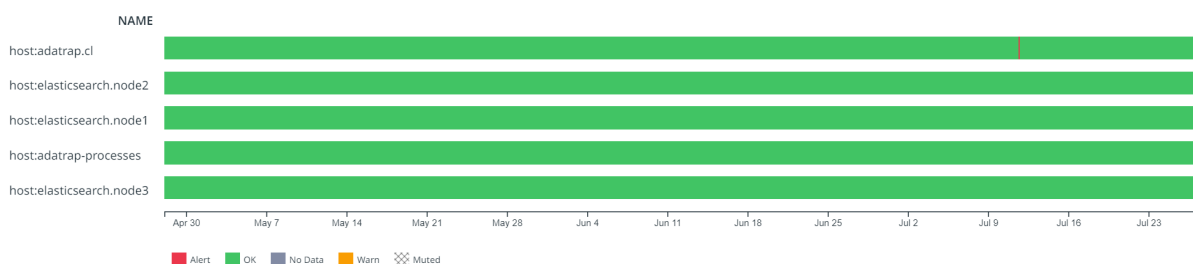


Ilustración 4.1.3.1: disponibilidad de sistema de extracción de datos

4.1.4 Actividad de usuarios

Entre el 1 de mayo de 2023 hasta el 30 de junio de 2023 se registró actividad en la plataforma de producción, en total, 45 usuarios fueron los que accedieron a la plataforma. En la Tabla 4.1.4 se muestra la lista de usuarios que han iniciado sesión durante el período indicado en el ambiente de producción. En la tabla se describe el número de sesiones en el período, última sesión iniciada, tiempo máximo de sesión, tiempo mínimo de sesión y tiempo promedio de sesión. Los usuarios que no aparecen indican que no han ingresado a la plataforma de visualización durante el período analizado.

Las sesiones de usuario se definen como un conjunto solicitudes al servidor cuya diferencia en tiempo de peticiones sucesivas no es superior a 15 minutos. Pasado 15 minutos, nuevas solicitudes se contabilizan como una nueva sesión de usuario.

Tabla 4.1.4: Usuarios con inicio de sesión durante el mes de mayo a junio de 2023 en el servidor de producción (www.adatrap.cl)

N°	usuario	n° sesiones	última sesión	tiempo máximo de sesión	tiempo mínimo de sesión	sesión promedio
1	tomas.rebolledo	217	2023-06-30 13:38:15	0:57:35	0:00:00	0:05:49
2	juan.lagos	216	2023-06-30 12:36:11	0:50:26	0:00:00	0:09:12
3	fabian.perez	149	2023-06-30 16:11:38	0:47:11	0:00:00	0:04:50
4	macarena.salazar	128	2023-06-30 16:44:18	1:53:10	0:00:00	0:06:45
5	miguel.vielma	119	2023-06-30 15:25:29	0:29:10	0:00:00	0:03:13
6	agustin.musante	96	2023-06-30 12:45:47	0:17:01	0:00:00	0:02:27
7	katherine.Garrido	87	2023-06-28 16:40:53	1:15:08	0:00:00	0:10:18
8	sebastian.tamblay	85	2023-06-30 19:45:39	1:13:40	0:00:00	0:06:00
9	andres.gonzalez	74	2023-06-30 12:04:30	0:32:34	0:00:00	0:03:20
10	patrick.oemick	66	2023-06-30 12:40:09	0:44:01	0:00:00	0:03:48
11	andres.vilches	58	2023-06-30 11:56:31	1:02:45	0:00:00	0:05:12
12	francisca.alarcon	58	2023-06-29 16:01:08	1:59:21	0:00:00	0:10:16
13	macarena.villalobos	50	2023-06-07 12:30:52	0:39:50	0:00:00	0:06:00
14	andres.avalos	42	2023-06-28 15:23:04	0:43:43	0:00:00	0:04:52
15	daniel.lopez	39	2023-06-16 22:03:30	0:58:27	0:00:00	0:11:24
16	ignacio.araya	37	2023-06-27 9:45:45	0:23:48	0:00:00	0:02:18
17	sebastian.busquets	34	2023-06-14 14:44:21	0:33:42	0:00:00	0:03:20
18	maria.viveros	33	2023-06-30 16:37:23	0:27:13	0:00:00	0:02:28
19	rafael.cozmar	29	2023-05-30 14:53:54	2:41:08	0:00:01	0:15:49
20	diego.munoz	27	2023-06-16 4:13:21	0:30:46	0:00:00	0:07:40
21	alejandra.pillajo	26	2023-06-29 15:51:14	0:21:21	0:00:00	0:03:15
22	gabriel.vargas	24	2023-06-29 11:35:57	0:25:43	0:00:01	0:03:56
23	christian.romero	19	2023-06-28 9:02:26	0:54:15	0:00:00	0:07:29
24	felipe.tapia	19	2023-05-24 17:58:30	0:20:38	0:00:00	0:04:00
25	camila.silva	17	2023-06-30 12:44:33	0:30:22	0:00:00	0:09:00
26	diego.fuentealba	14	2023-06-14 10:33:22	0:18:56	0:00:00	0:06:15
27	franco.banchero	13	2023-06-06 18:01:33	0:05:56	0:00:00	0:00:38
28	loreto.bravo	11	2023-06-29 15:21:02	0:31:30	0:00:04	0:10:15
29	mauricio.zuñiga	11	2023-06-28 8:40:39	0:15:45	0:00:00	0:03:07
30	eduardo.lizana	10	2023-06-29 12:13:20	0:21:09	0:01:05	0:07:36
31	victor.seguel	7	2023-06-13 9:21:57	0:06:22	0:00:04	0:02:47
32	elkin.chaparro	7	2023-06-15 12:18:29	0:41:49	0:00:07	0:22:29
33	javier.yevenez	6	2023-06-29 11:53:13	0:04:13	0:00:03	0:00:51

34	muriel.manriquez	5	2023-06-07 13:01:09	0:06:35	0:00:19	0:01:51
35	francisca.giraldez	4	2023-06-28 15:25:12	0:03:10	0:00:13	0:01:08
36	cristina.valenzuela	3	2023-06-12 17:31:39	0:02:13	0:00:01	0:00:53
37	gonzalo.godoy	3	2023-06-08 8:25:24	0:06:58	0:00:00	0:02:51
38	mario.gonzalez	3	2023-05-31 9:36:29	0:05:24	0:01:11	0:02:45
39	alexis.nilo	3	2023-06-05 15:46:08	0:02:34	0:00:20	0:01:05
40	juan.olea	3	2023-06-28 15:46:19	0:12:56	0:00:01	0:04:20
41	monica.munoz	2	2023-06-12 11:58:26	0:00:33	0:00:19	0:00:26
42	jordi.vanspronsen	2	2023-06-12 15:26:18	0:01:55	0:01:37	0:01:46
43	ignacio.riquelme	1	2023-05-16 18:22:13	0:09:47	0:09:47	0:09:47
44	juan.rojas	1	2023-05-10 9:24:07	0:00:03	0:00:03	0:00:03
45	jorge.carrasco	1	2023-05-09 10:58:37	0:00:02	0:00:02	0:00:02
46	cesar.labra	1	2023-05-23 10:57:04	0:03:40	0:03:40	0:03:40
47	jaime.moya	1	2023-05-08 17:18:35	0:00:05	0:00:05	0:00:05
48	antonio.gschwender	1	2023-05-02 17:58:35	0:12:09	0:12:09	0:12:09
49	diego.silva	1	2023-06-12 15:02:14	0:00:16	0:00:16	0:00:16
50	oscar.gray	1	2023-06-13 16:00:19	0:01:04	0:01:04	0:01:04
51	daniel.huencho	1	2023-06-19 10:28:16	0:03:31	0:03:31	0:03:31
52	catalina.gonzalez	1	2023-06-27 11:07:20	0:04:35	0:04:35	0:04:35