

# **Procesamiento y disponibilización de Datos de ADATRAP**

## **Informe Final - Mayo 2019**

---

Cliente : Secretaría Ejecutiva del Directorio de Transporte Público Metropolitano (DTPM)

17-07-2019

Mauricio Zúñiga G.  
+56 9 94204180  
mzuniga@gmail.com

## Contenido

1. INTRODUCCIÓN.....	4
2. OBJETIVOS.....	4
3. RESUMEN TRABAJOS .....	4
4. DATOS DE ENTRADA.....	8
4.1 GPS.....	9
4.2 Transacciones .....	10
4.3 Reportes 1.96.....	11
5. DATOS DE SALIDA.....	12
5.1 Estimación de bajada .....	12
5.2 Velocidades del sistema .....	14
5.3 Metodología de estimación de viajes .....	14
5.4 Metodología zonas pagas.....	14
5.5 Metodología perfiles .....	15
6. MANTENCIÓN.....	15
6.1 Radar.....	15
6.1.1 Problemas o Caídas del sistema.....	16
6.1.2 Modificaciones .....	16
6.2 Sistema de Visualización de ADATRAP.....	17
6.2.1 Mantención .....	17
6.2.2 Problemas o Caídas del sistema.....	19
6.2.3 Usuarios .....	19
6.2.4 Almacenamiento de archivos .....	21
7. DESARROLLOS .....	23
7.1 Validación Asistida .....	23
7.2 Levantamiento de servidores en la nube .....	26
7.3 Construcción de análisis de estimación de bajada en zonas pagas.....	27
7.4 Modificación formato de entrada Zonas Pagas .....	27
7.5 Comparación de distribución con encuestas vs ADATRAP .....	27
7.5.1 Construcción de archivos de distribución .....	29
7.5.2 Comparación con mediciones 2018 (noviembre 2018).....	29
7.5.3 Comparación con mediciones 2018 (diciembre 2018).....	33
7.5.4 Salida de distribuciones en Zonas Pagas.....	34

7.6	Mantenimiento de parámetros de consulta en visualizador.....	34
7.7	Extracción de GPS .....	34
7.8	Registro de actividades de usuario .....	35
7.9	Sección de almacenamiento.....	35
7.10	Programa para carga de archivos de respaldo a la nube.....	37
7.11	Homologación de formato de GPS .....	37
7.12	Construcción de datos de Metro L3 .....	37
7.13	Actualización almacenamiento en producción .....	37
7.14	Procesamiento automático de actividad de usuarios .....	38
7.15	Vista de Accesos solicitados a datos históricos .....	39
8.	ANEXOS.....	40
8.1	Script homologación GPS.....	40
8.2	Modificaciones datos de metro L3 .....	43
8.3	Validación automática.....	45

# 1. INTRODUCCIÓN

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” por medio de la Resolución 665. El servicio consiste en realizar el procesamiento, almacenamiento y disponibilización de información relevante para el análisis de datos del sistema de transporte de Santiago, y una de las principales funciones es la construcción de fuentes de datos como matrices origen-destino, perfiles de carga, velocidades e indicadores del sistema.

El periodo a procesar corresponde a los meses entre junio del 2017 y marzo 2019, pero durante el mes de marzo 2019 se realizó una extensión del servicio por lo que se adicionan los meses de abril y mayo 2019. 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. En algunas ocasiones estas horas pueden ser destinadas a estudios sobre algún problema en particular.

El informe se divide en dos secciones, Datos de Entrada donde se hará una descripción de todos los datos utilizados como entrada por ADATRAP identificando la existencia de días con problemas que no deban ser considerados en los análisis y Desarrollo donde se detallará cada una de las modificaciones realizadas al código fuente del software.

## 2. OBJETIVOS

Este informe tiene por objetivo mostrar un resumen de todo el trabajo realizado durante el proyecto, periodo en el cual fue procesado el corte temporal entre junio 2018 y mayo 2019 que corresponde desde septiembre de 2018 a mayo de 2019. Por tanto, el contenido de este informe corresponde a un resumen de todos los informes entregados por el proyecto adicionando lo realizado en el último mes, correspondiente a mayo 2019.

## 3. RESUMEN TRABAJOS

A continuación, se describirá el trabajo realizado en cada mes durante todo el periodo del proyecto, los detalles serán descritos en las secciones posteriores según corresponda.

### **Informe septiembre 2018**

*Procesamiento:* Durante este periodo se dio paso al procesamiento de dos meses de datos mayo 2018 y junio 2018, quedando fuera del procesamiento el mes de Julio 2018 por presentar problemas en los datos de GPS y transacciones desde SONDA, quedó pendiente para el próximo periodo el procesamiento de los meses de agosto y septiembre a la espera de los nuevos datos de entrada correspondientes al programa de operación vigente que corresponda.

*Desarrollos:* Se desarrolló e integro al sistema una salida adicional por día con la información de diversos indicadores con el objetivo de poder validar el procesamiento de manera más automatizada. Se levantaron los servidores en la nube (AWS Amazon) para poner en marcha el servidor de producción del sistema. También se desarrolló un módulo de análisis para la estimación de bajada en zonas pagas con el objetivo de agilizar la revisión y mejorar los resultados en los perfiles de carga y posibilitar además un módulo para apoyar el recaudo interno de zonas pagas mixtas para DTPM.

### **Informe octubre 2018**

*Procesamiento:* Durante este periodo se procesaron dos meses de datos, agosto y septiembre 2018. Por problemas en la entrega de información de GPS y Validaciones Bip el mes de Julio 2018 no será considerado para su procesamiento.

*Desarrollos:* Se modificaron los formatos de entrada en Zonas Pagas entregando más detalle en los servicios, esto mejoró los resultados en la estimación de bajada, además, se generó una nueva salida de ADATRAP que contiene las distribuciones de transacciones por operador en zonas pagas contra encuestas hechas en terreno.

### **Noviembre 2018**

*Procesamiento:* Durante el periodo de noviembre se procesó el mes de octubre, cabe hacer notar que existió un cambio en el programa de operación por los cambios producidos con el operador Alsacia, que entregó ciertos servicios a otros operadores, por lo que el procesamiento se realizó hasta el 23 de octubre.

*Desarrollos:* Durante este periodo además se construyó un software en Python que es capaz de detectar posibles días con anomalías, utilizando una metodología basada en la desviación estándar de los datos, este software fue utilizado para detectar días con fallas en este periodo. Respecto a desarrollos se continuó realizando un estudio comparativo entre ADATRAP y encuestas para las distribuciones de transacciones en zonas pagas.

### **Informe diciembre 2018**

*Procesamiento:* En este periodo se procesaron los días posteriores al 23 de octubre hasta el 30 de noviembre, en esta oportunidad se utilizó un nuevo programa de operación correspondiente al 23 de octubre, la justificación de este cambio radica principalmente a modificaciones en los servicios de Alsacia.

*Desarrollos:* Durante el periodo de diciembre 2018 se realizaron además dos tareas correspondientes a desarrollos, estas corresponden a un estudio para comparar los datos de ADATRAP vs Encuestas realizadas en 2018 de las distribuciones de las validaciones en los distintos operadores en zonas pagas mixtas y la construcción de un archivo con esta información obtenida a través de ADATRAP, en el formato requerido por el DTPM para integrarlo a sus procesos internos.

## **Informe enero 2019**

*Procesamiento:* En este periodo se procesaron los días entre el 01 de diciembre hasta el 31 de enero. Se realizaron diversas tareas asociadas a migrar los servicios hacia los servidores finales de producción en Amazon, tareas aplazadas por los desarrollos descritos en los informes anteriores con mayor prioridad.

*Desarrollos:* Se agregó un sistema de registro de actividades de usuario, esto con el objetivo de conocer con más detalle las actividades que los usuarios realizan dentro del visualizador ADATRAP. Se agregó un módulo nuevo en el menú lateral de nombre “Almacenamiento”, este contiene la información que ha sido respaldada en la nube, específicamente en el servicio S3 de AWS.

## **Informe febrero 2019**

*Procesamiento:* En este periodo por razones de retrasos en la obtención de los datos de GPS, debidos a problemas con el servicio de entrega en línea de estos datos, se optó por utilizar el tiempo disponible procesando otros meses de 2018 con el objetivo de lograr tener dentro de la plataforma un año de corrido de información, es por esto que en este periodo se procesaron los meses de marzo, abril y diciembre de 2018, enero 2019 quedó pendiente porque se está incorporando la línea 3, entiéndase la construcción de los datos de la línea 3 para su posterior procesamiento en ADATRAP.

*Desarrollos:* Se realizó un desarrollo para poder realizar la homologación de formato GPS con una nueva fuente de datos, esto porque empezaron a aparecer problemas con estos, para esto se construyó un software en Python que transforma el nuevo formato de dato al utilizado por ADATRAP. Se realizó la construcción de los datos de entrada para integrar la línea 3 de Metro. Se construyó un script para procesar la actividad de los usuarios registrada en el servidor de producción, `adatrap.cl`. El módulo de almacenamiento fue habilitado en el ambiente de producción.

## **Informe marzo 2019**

*Procesamiento:* Durante este periodo el software de validación no fue relevante puesto que en periodo estival la gran mayoría de las variables que se utilizaron para validar quedaron fuera de rango según nuestro set de días para comparar. Por otra parte, se incrementó el número de buses en el sistema lo que impacto directamente a la cantidad de registros GPS. También se encontró un problema en la información de servicio en los datos de GPS, en el mes de marzo, se identificó que no existen servicios variantes, esto no es un problema para ADATRAP porque esa información de servicio la obtiene de otro archivo (Reporte 1.96), es por esto que el módulo de velocidad (APE + VelocidadesTS) fue desconsiderado. Este problema también impactó todos los servicios en línea, en este caso el servicio del RADAR se vio perjudicado puesto que

las asignaciones de servicio de algunos buses fueron asignadas al servicio madre cuando estaban realizando una variante generando un error en las estimaciones. Por otra parte, la información en línea con el cual se alimenta el RADAR (GPS en línea) también presentó problemas de falta de información, se realizaron estimaciones comparando con la actual fuente de datos para ADATRAP encontrándose una pérdida media del 30% de los datos, esto impacta fuertemente en la precisión de la estimación de las velocidades que alimenta el sistema. Otro ítem importante a considerar es que en los periodos de enero 2019 y febrero 2019 fueron utilizados como datos de operación el programa entregado en octubre 22, desde la contraparte quedó pendiente entregar un programa de operación especial para ese periodo para poder recalculer y mejorar las estimaciones con un programa más ajustado a la realidad del periodo.

*Desarrollos:* Se construyó un programa que cruce los datos de GPS actuales con la información del reporte 1.96 con el objetivo de generar un nuevo archivo de GPS con variantes incluidas con el cual se podrá recalculer la información de velocidades, además, en la sesión de administrador del portal web se integró un script para calcular la actividad de usuarios de manera automática cada 24 horas, también se implementaron dos vistas web, una para ver el detalle de las sesiones y otra que contiene una tabla resumen de las sesiones por cada usuario.

### **Informe abril 2019**

*Procesamiento:* Durante este periodo se procesaron los datos para el mes de marzo 2019 tomando en consideración todos los problemas que aparecieron en los datos de entrada, se crearon algunos procesos nuevos para incorporar los datos de GPS que fueron entregados desde otra fuente y hubo que estandarizarlos al formato de ADATRAP.

### **Informe mayo 2019**

*Procesamiento:* Durante este periodo se procesaron los datos del mes de abril 2019. Se detectaron problemas con los datos de transacciones, estos a partir del 24 de abril no pudieron ser rescatados por el sistema, DTPM gestionó la entrega de estos datos por otra fuente. Este problema fue resuelto desde el 2 de mayo en adelante obteniéndose las transacciones sin problemas.

*Desarrollos:* Se realizaron varios análisis adicionales para confirmar los resultados de comparar la información de distribución de transacciones en zonas pagas, se realizó una nueva comparación con los datos de octubre 2018, donde se convino una metodología de construcción de los factores de pago sobre una planilla de cálculo con datos obtenidos de ADATRAP, esta metodología fue aprobada para ser integrada al cálculo diario de ADATRAP en un futuro desarrollo.

## 4. DATOS DE ENTRADA

Como se explicó anteriormente esta sección corresponde al análisis de los datos de entrada que utiliza el sistema. El objetivo es determinar qué días procesados no cumplen uno o más de los criterios de aprobación. El no cumplimiento de alguna regla incide en que el día procesado no estará disponible en la plataforma de visualización de ADATRAP.

Durante todo el proyecto se detectaron algunos días anómalos los cuales fueron eliminados de los análisis, la Tabla 1 muestra cada uno de estos días y sus motivos, la descripción de cada motivo también puede ser revisado en la Tabla 2.

<b>FECHA</b>	<b>MOTIVO</b>
2018-03-02	Caída Adatrap
2018-03-03	% bajada muy bajo
2018-03-16	Caída Adatrap
2018-03-17	% bajada muy bajo
2018-03-21	Caída Adatrap
2018-03-30	Feriado
2018-04-02	% bajada muy bajo
2018-04-03	Caída Adatrap
2018-04-04	% bajada muy bajo
2018-04-16	% bajada muy bajo
2018-04-17	Caída Adatrap
2018-04-18	Caída Adatrap
2018-04-19	% bajada muy bajo
2018-04-30	Caída Adatrap
2018-05-12	Fuente de datos
2018-05-28	Fuente de datos
2018-06-03	Fuente de datos
2018-08-01	Fuente de datos
2018-08-02	Fuente de datos
2018-08-03	Fuente de datos
2018-08-09	Fuente de datos
2018-08-21	Fuente de datos
2018-08-23	Fuente de datos
2018-10-15	Fuente de datos
2018-11-01	Feriado
2018-11-02	Feriado
2018-11-25	Fuente de datos
2018-11-26	Fuente de datos
2018-11-30	Caída Adatrap
2018-12-18	Caída Adatrap
2018-12-25	Feriado
2019-01-01	Feriado
2019-01-17	Caída Adatrap
2019-02-07	Caída Adatrap
2019-02-18	Caída Adatrap

<b>2019-04-06</b>	Caída Adatrap
<b>2019-05-01</b>	Feriado
<b>2019-05-20</b>	Entre Feriados
<b>2019-05-21</b>	Feriado
<b>2019-05-23</b>	Caída Adatrap

*Tabla 1: Días no considerados en el sistema en producción.*

<b>MOTIVO</b>	<b>DESCRIPCIÓN</b>
<b>% bajada muy bajo</b>	Caso en que el porcentaje de bajada de las etapas procesadas sea más bajo que lo normal, los umbrales de normalidad son explicados en mayor detalle en la sección 4.2 de este informe.
<b>Caída Adatrap</b>	Este caso corresponde cuando el software de procesamiento de los datos sufre alguna falla en su ejecución, normalmente esto es debido a una inconsistencia en los datos de entrada, las razones específicas de una caída para un día particular deben ser revisadas en detalle según la prioridad de trabajo que indique el DTPM
<b>Feriado</b>	Los días feriados son eliminados del sistema en producción hasta que este tenga la capacidad de diferenciarlos en sus filtros
<b>Fuente de Datos</b>	Este caso se produce al detectarse problemas directos en las fuentes de datos, por ejemplo, una caída en la cantidad de validaciones Bip o pulsos GPS
<b>Entre Feriados</b>	Los días entre feriados y fines de semanas también tienen un comportamiento fuera de lo normal por lo que son descartados del sistema en producción

*Tabla 2: Tabla con descripción de los motivos de eliminación de días de datos en producción.*

Adicionalmente se eliminaron todos los datos del mes de julio 2018 por problemas desde la fuente (SONDA).

## 4.1 GPS

Uno de los tres datos masivos que utiliza el sistema corresponde a los GPS de los buses, estos registros son emitidos cada 30 segundos por cada bus del sistema y enviados a la empresa SONDA, proveedor tecnológico del sistema de transporte público. DTPM por su parte mantiene un *web service* de acceso público con información de posicionamiento de buses. Cada minuto son extraídos los datos en línea del sistema y almacenados en archivos diarios que son la entrada para el software ADATRAP.

Se ha visto que un día laboral normal (lunes a viernes) debiese estar en el orden de los 12.500.000 registros, para los días sábados el número es cercano a 10.000.000 y los días domingos está dentro de los 8.500.000, es por esto que un primer análisis corresponde a revisar este número de registros por día para descartar días anómalos. Estos valores fueron estimados usando información de años anteriores a 2018, cabe hacer notar que se están reevaluando estos valores con el objetivo de ajustarlos a la nueva flota del sistema, por

ejemplo, se aprecia que un valor normal bordea los 14.000.000 de registros diarios para un día laboral.

En la Ilustración 1 se observa la cantidad de registros de GPS procesados durante todo el periodo del proyecto, marzo 2018 a mayo 2019.

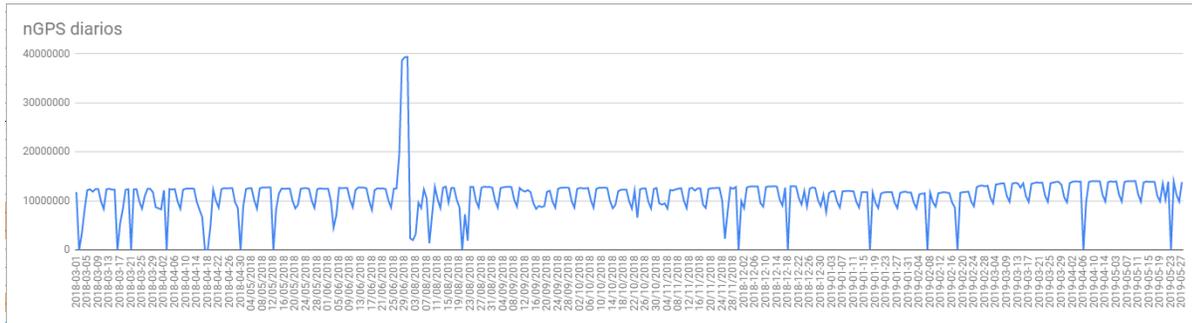


Ilustración 1: Curva de numero de pulsos GPS diarios, desde el 1 de marzo de 2018 hasta el 30 de mayo de 2019.

## 4.2 Transacciones

La segunda fuente de datos corresponde a las transacciones del sistema, estas transacciones son registradas en los validadores bip! y son cargadas a los sistemas de SONDA, nuestros sistemas recolectan diariamente estos datos desde un servicio FTP entregado por DTPM. En base a análisis realizados por el equipo se estima que es posible encontrar un 95% de los datos al cuarto día desde su generación. Estos datos contienen información relevante para el software: id de la tarjeta, tiempo de la transacción y sitio donde se realizó la transacción (patente de bus, código validador o estación de metro).

Existe una cantidad de transacciones al día relativamente constante, se aprecia en el tiempo que un día laboral normal tiene alrededor de 6.000.000 de transacciones, un día sábado del orden de 3.000.000 y los días domingos 1.800.000. Esto permite, al igual que en los datos GPS, determinar días anómalos cuando hay diferencias significativas en este número de registros.

En la Ilustración 2 a continuación se observa la cantidad de registros de validaciones Bip! durante el 1 de marzo de 2018 al 31 de mayo de 2019. En base al gráfico, se pueden apreciar los días con caídas del procesamiento donde el número de transacciones está definido como cero.

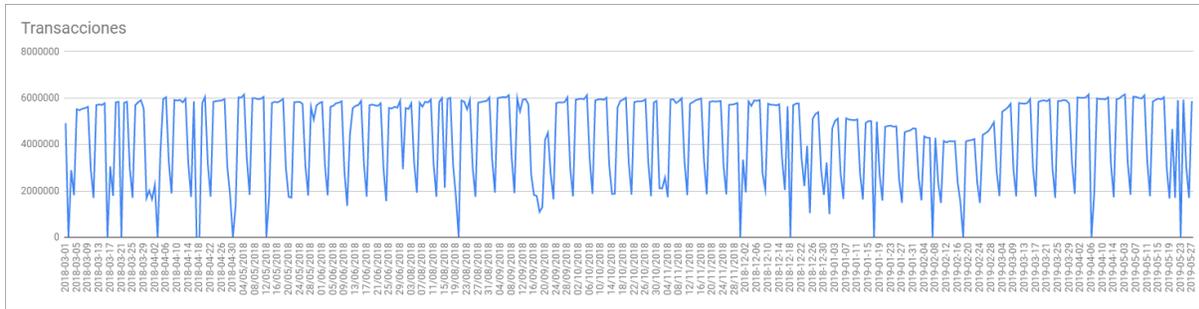


Ilustración 2: Curva de número de transacciones por día desde el 1 de marzo de 2019 al 30 de mayo 2019.

### 4.3 Reportes 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 una patente, además incorpora los tiempos de inicio de esa asignación y el tiempo en la cual finaliza. Este dato permite poder determinar finalmente a que servicio están asignadas las transacciones bip! al subir al bus, a que ruta debe asignarse la trayectoria de GPS para calcular el avance en la ruta y cuál es la secuencia de paraderos que está realizando la máquina.

De la misma manera que las fuentes de datos anteriores, la cantidad de registros es constante en el tiempo y detectar alguna anomalía en este número es indicador de algún problema en los datos. Para tener una noción de normalidad es importante saber que un día laboral tiene entre 75.000 a 77.000 expediciones, un día sábado entre 57.000 a 60.000 y un día domingo entre 45.000 a 49.000.

En la Ilustración 3 a continuación se observa la cantidad de registros de expediciones durante los días procesados en todo el proyecto.

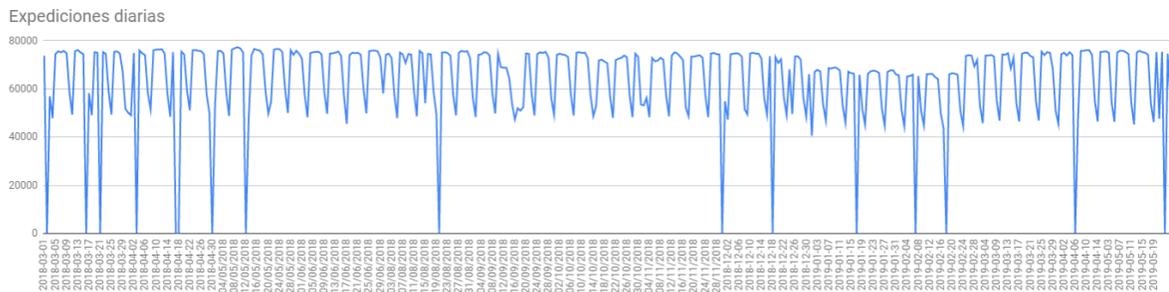


Ilustración 3: Curva de numero de expediciones en el 1.96 del 1 de marzo 2018 al 30 de mayo del 2019.

## 5. DATOS DE SALIDA

Los datos de entrada ya descritos en el capítulo 4 más otros datos, en su mayoría provenientes del programa de operación vigente, son procesados por ADATRAP con las distintas metodologías como estimación de bajada, construcción de viajes, estimación de velocidades y expansiones, resultando en un conjunto de tablas de salida (viajes, etapas, perfiles de carga, velocidades, indicadores, tablas web). Estas tablas 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. A continuación, se describen las principales metodologías de ADATRAP.

### 5.1 Estimación de bajada

Uno de los procesos realizados por ADATRAP, y que podría ser considerado como el más relevante, es la estimación de bajada de los usuarios. Es necesario entender que el sistema de transporte público en Santiago solo exige validar para el pago al subir a un bus o entrar en una zona paga, metro o metrotren, 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 del sistema, es por esto que en los proyectos de Fondef/Conicyt ya mencionados se investigó y desarrolló una metodología que estima estos paraderos de bajada. En síntesis, la metodología consiste en determinar sobre el camino del bus y servicio, al que el usuario subió, la parada óptima (tiempo generalizado) de bajada en relación a la siguiente transacción realizada por ese usuario. 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).

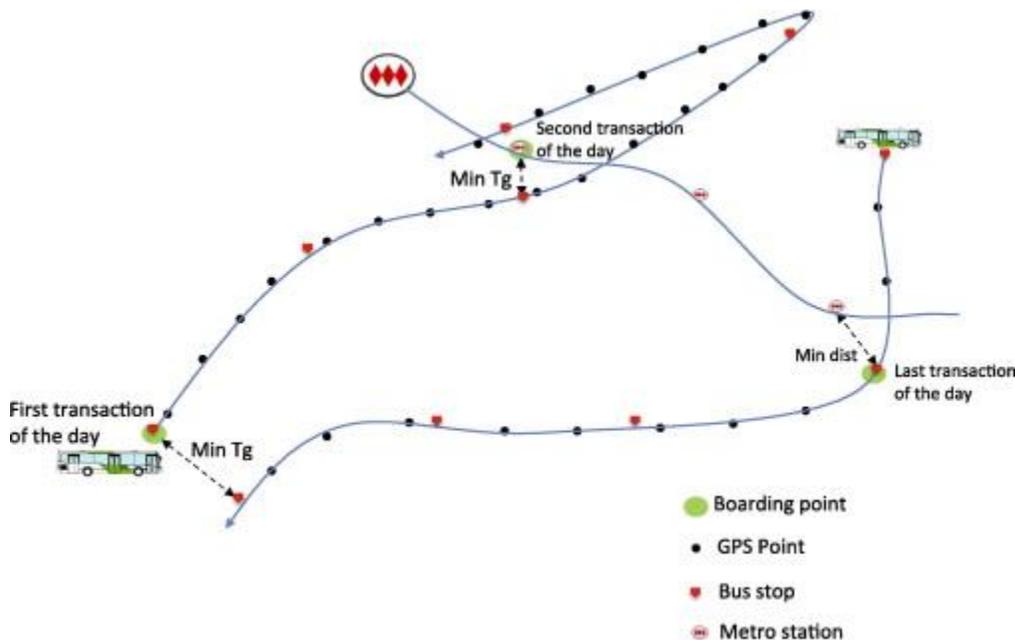


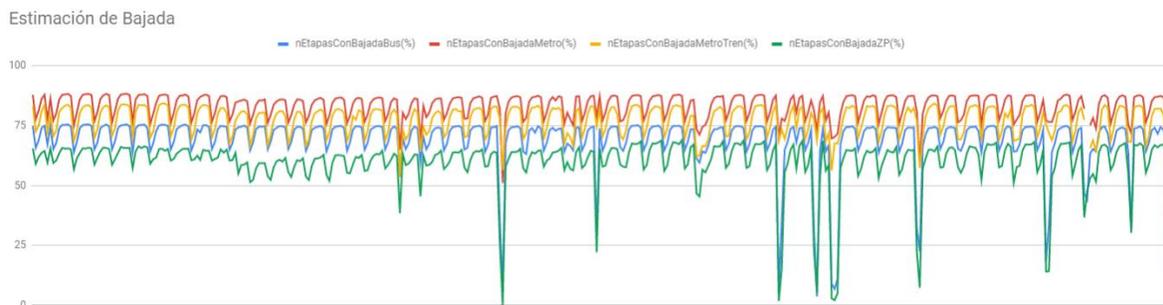
Ilustración 4: Esquema que describe la estimación de bajada para un pasajero

Es importante señalar que 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 que provee la red de transporte público, el primero corresponde al modo bus, es decir, que el usuario validó dentro del 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 ultimo 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.

A la fecha cualquier problema que exista en los datos de entrada, considerando además el Programa de Operación, impacta fuertemente en los porcentajes de estimación de bajada transformándose así en un buen indicador de errores en el proceso.

En la Ilustración 5 se puede observar el porcentaje de etapas con estimación de bajada por modo de transporte en cada día. Las caídas en la curva corresponden a días donde adtrap tuvo problemas de procesamiento, de esta manera quedan los % en cero.



*Ilustración 5: Curva de porcentaje de estimación de bajada, rojo para metro, naranja para metro-tren, azul para buses y verde para zonas pagas, del 1 de marzo 2018 al 30 de mayo del 2019.*

## 5.2 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, de todas maneras, en caso de aparecer algún problema en estos registros se traduce en la inexistencia de estos, puesto que internamente la herramienta tiene un conjunto de filtros que no permite considerar datos anómalos, dejando los resultados para los tramo-periodo con valores nulos.

Durante el proyecto se calcularon las velocidades para todos los meses entre marzo 2018 y mayo 2019, a excepción de los meses de enero 2019 y febrero 2019 donde aparecieron problemas en los datos GPS y julio 2018 que fue eliminado por completo por los problemas que presento SONDA en sus datos.

## 5.3 Metodología de estimación de viajes

Los viajes son construidos recorriendo la secuencia de etapas de una tarjeta, en este recorrido se va analizando lo ocurrido entre la bajada de una etapa y la subida de la siguiente, este espacio debe cumplir las siguientes condiciones para determinar si es un fin de viaje o un transbordo:

1. Si la etapa es la última de la tarjeta → fin de viaje
2. Si la siguiente transacciones ocurre 2 horas después → fin de viaje
3. Si en el tiempo del supuesto transbordo pasan más de 3 buses de la línea que utilizó en la siguiente etapa → fin de viaje
4. Si ambas etapas tienen el mismo servicio → fin de viaje
5. Si el tiempo de espera cuando la siguiente transacción es en Metro es mayor a 15 minutos → fin de viaje
6. En caso que la etapa genere un regreso al punto de partida → fin de viaje
7. En último caso si el tiempo de espera es superior a 40 minutos → fin de viaje

## 5.4 Metodología zonas pagas

Para la estimación de bajada en Zonas Pagas se realiza el siguiente proceso secuencial:

1. **Identificación de paradero de subida:** Esta información es un dato directo de la operación, la transacción Bip viene asociada al validador el cual a su vez está asociado a un paradero.

2. **Identificación de servicios en Zona Paga:** La información de los servicios que operan en esa zona paga es un dato entregado por DTPM.
3. **Detección de servicios factibles:** Con el listado de servicios que operan en la zona paga se realiza una búsqueda de los servicios factibles midiendo la distancia mínima entre su ruta y la siguiente transacción de la tarjeta, si esta distancia es menor a 1000 m el servicio es considerado un servicio factible.
4. **Determinación de líneas comunes:** Con la lista de servicios factibles se realiza el cálculo de líneas comunes, este cálculo permite estimar el conjunto óptimo de servicios que si utilizaría el usuario para llegar a su destino. Para calcular este conjunto se utiliza el algoritmo publicado en Cea y Fernandez (1989) donde ordenando los servicios según su tiempo de viaje son ingresados al conjunto de líneas comunes a medida que su tiempo esperado no empeore, este tiempo esperado es calculado en función del tiempo de viaje y sus frecuencias de operación.
5. **Obtención bus más probable:** Con el conjunto de servicios o líneas comunes se da paso a la estimación del bus de subida, para esto se realiza un ciclo sobre todos los buses que pasaron por la zona paga entre 3 horas antes y 1 hora después de la validación Bip!, sobre todos estos buses o pasadas sobre ese paradero se selecciona el temporalmente más cercano y posterior a la validación.
6. **Estimación de tiempo de bajada:** Finalmente con el bus estimado se realiza un seguimiento sobre los datos GPS para determinar el tiempo estimado de bajada del usuario de la tarjeta Bip!

## 5.5 Metodología perfiles

Los perfiles de carga son construidos realizando la suma de subidas y la resta de bajadas de las etapas en cada expedición de un bus, esta suma y resta se realiza sobre el peso de cada etapa, peso que proviene del cálculo de factores de expansión por cada servicio, paradero y en los periodos definidos por DTPM.

# 6. MANTENCIÓN

Además del procesamiento de los datos, el servicio también incluyó la mantención de dos servicios web, el primero corresponde al RADAR y el segundo al visualizador de ADATRAP, ambos servicios montados en *AWS Amazon*.

## 6.1 Radar

El Radar del tránsito consiste en una página web que muestra los tiempos de viajes de 52 ejes de la ciudad de Santiago, estos ejes fueron definidos en un proyecto de desarrollo en conjunto

entre académicos de la Universidad de Chile y DTPM y corresponden a ejes en dirección hacia el centro de la ciudad, por lo que esta visualización fue diseñada para monitorear el estado de las velocidades del sistema durante el período punta mañana.

### 6.1.1 Problemas o Caídas del sistema

El sistema web no presentó problemas de disponibilidad durante el periodo del proyecto pero si existieron problemas por falta de información. La falta de información se originó por la imposibilidad de acceder a los datos de velocidades producto de los altos tiempos de respuesta que experimentó el servidor desde donde se extraen los datos. Lo anterior significó pérdida de información que se tradujo en tramos sin información. Estos problemas se presentaron desde diciembre de 2018 a marzo de 2019. En la Ilustración 6 se observa la visualización del mapa de velocidades para los tramos definidos.

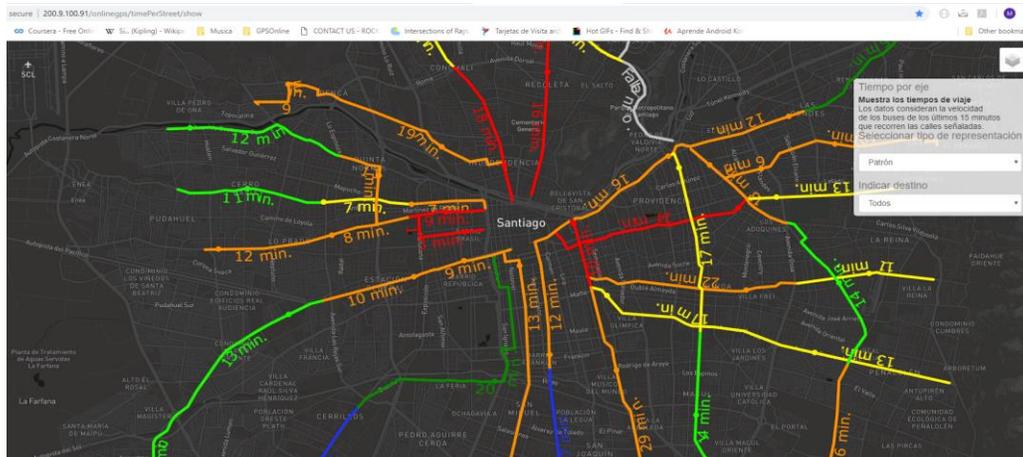


Ilustración 6: Mapa de velocidades

### 6.1.2 Modificaciones

Durante el desarrollo del proyecto se agregó una etiqueta a la leyenda de la visualización, esta consistió en indicar la métrica que está siendo coloreada en el mapa más su unidad de medida (km/hr). En la Ilustración 7 se muestra el caso previo (izquierda) y el resultado al aplicar el cambio descrito (derecha).

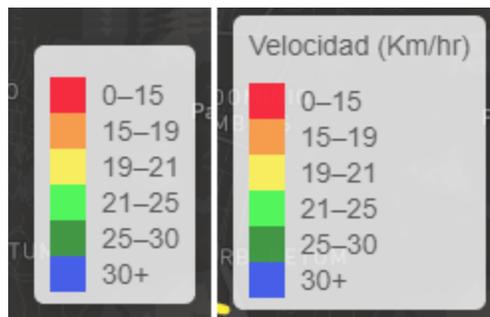


Ilustración 7: Leyenda antes del cambio (imagen izquierda) y leyenda después del cambio (imagen derecha)

## 6.2 Sistema de Visualización de ADATRAP

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 y además permite exportar los datos mostrados en las visualizaciones.

Los datos generados como resultado del procesamiento de ADATRAP y utilizados por la plataforma de visualización son cinco:

- **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, es decir, estadísticas de un día

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 (una máquina), 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 de almacenar y disponibilizar el sitio web.

### 6.2.1 Mantención

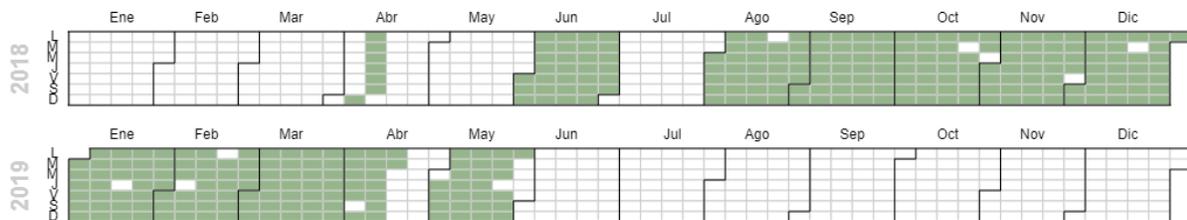
La mantención de la plataforma consiste en la realización de todas aquellas tareas necesarias para que la plataforma de visualización esté disponible en los términos acordados durante la ejecución del proyecto. Estas tareas se pueden resumir en:

- Traslado de archivos desde máquina de procesamiento a máquina de visualización en ambiente de desarrollo y ambiente de producción
- Carga de datos en ambiente de desarrollo y producción
- Monitoreo permanente de la plataforma: espacio en disco de almacenamiento, consumo de recursos, disponibilidad del sistema, entre otros
- Resolución de consultas o dudas por parte de algún usuario de la plataforma
- Soporte frente a problemas que pueda tener algún usuario en la plataforma
- Acciones puntuales que se requieren realizar para el correcto uso de la plataforma, esto puede ser la creación de un nuevo usuario, asignación de permisos, entre otros
- Elaboración de informes de funcionamiento de la plataforma

En la práctica, todos los meses se debió realizar alguna de las acciones mencionadas previamente, lo que dio como resultado (al finalizar este proyecto) la disponibilidad de un conjunto de datos para el cliente.

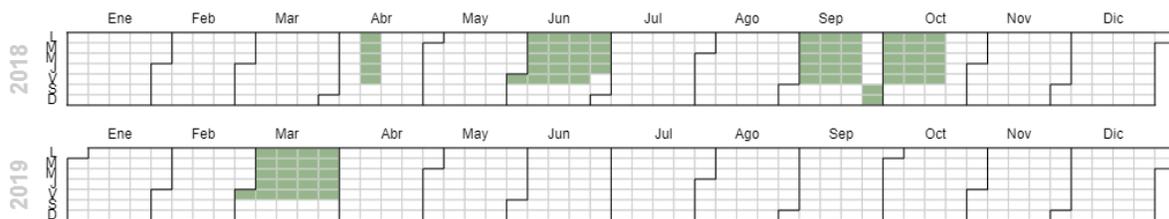
### 6.2.1.1 Días disponibles

Al finalizar el proyecto la plataforma de visualización quedó con 163 días adicionales para ser consultados de: perfiles de carga, viajes, matriz origen-destino por servicio, esto corresponde a los días comprendidos entre el 01 de septiembre de 2018 y 27 de mayo de 2019. En la Ilustración 8 se muestran los días disponibles para ser consultados, se destaca el hecho de que existen otros días disponibles en la plataforma que corresponden a ejecuciones previas al desarrollo de este proyecto, en particular, corresponde a los días de abril de 2018 y junio de 2018.



*Ilustración 8: Días disponibles de perfiles de carga, viajes y matrices o-d por servicio en plataforma de visualización al finalizar proyecto*

Por otro lado, los datos de velocidades presentan una disponibilidad menor que los datos anteriormente descritos. En la Ilustración 9 se muestran los días disponibles, estos corresponden a 53 días distribuidos de la siguiente forma: 17 días en el mes de septiembre de 2018, 15 días agrupados en el mes de octubre de 2018 y 21 días agrupados en el mes de marzo de 2019. Los días calculados corresponden a días de la semana salvo por dos que pertenecen al fin de semana, específicamente son el 29 de septiembre de 2018 y 30 de septiembre de 2018.



*Ilustración 9: Días disponibles de velocidades en plataforma de visualización al finalizar proyecto*

Otro aspecto a destacar es la presencia de datos calculados fuera del período de ejecución de este proyecto pero que aún permanecen en la plataforma de visualización por petición del cliente, estos corresponden a una semana de abril de 2018 y los días laborales del mes de junio de 2018 (ver Ilustración 9).

## 6.2.2 Problemas o Caídas del sistema

Durante la ejecución del proyecto se presentaron algunas situaciones que impidieron el libre uso de la plataforma de visualización, en esta sección se describen esos problemas y las soluciones llevadas a cabo.

Una caída del sistema se define como un desperfecto en la infraestructura o software que impide que los usuarios puedan hacer uso del servicio, en este caso el servicio corresponde a la visualización y descarga de los datos.

La arquitectura de la plataforma está compuesta por ambiente de desarrollo y producción. En el ambiente de desarrollo existen dos máquinas orientadas al ambiente de desarrollo que se encuentran físicamente en las instalaciones de la facultad de ingeniería de la Universidad de Chile. Una es usada como almacén de datos de respaldo y la segunda como una herramienta de desarrollo para el sistema de visualización, por lo que contiene una instalación de todo el sistema web y datos. Por otro lado, en el ambiente de producción existen cuatro máquinas: servidor web y almacén de datos temporales, nodo maestro de elasticsearch y dos nodos de datos de elasticsearch. La ubicación del ambiente de producción es el centro de datos de AWS ubicado en la ciudad de Sao Paulo, Brasil.

La primera caída se produjo el día 30 de diciembre de 2019 por una caída en el servidor del ambiente de desarrollo cuando se estaba realizando la carga de datos correspondiente al mes de octubre. El problema que ocasionó la caída se debió a la paralelización de la carga de archivos, lo que provocó una saturación de la memoria RAM y en consecuencia la falta de respuesta de la máquina. El incidente fue controlado y se tomaron las precauciones para futuras cargas. Una segunda caída se produjo durante el mes de enero de 2019, específicamente el día 14 de enero entre las 12:00 y 12:15 horas debido a una saturación de memoria en una de las máquinas de producción, se corrigió aplicando la misma medida usada en la máquina de desarrollo del incidente anterior. Un tercer problema ocurrió en el ambiente de desarrollo el día 28/02/2019 entre las 10:32 y 11:40 horas y se originó por un problema de acceso a internet que tuvo el edificio de ingeniería de la Universidad de Chile, lo cual imposibilitó el uso del ambiente de desarrollo, esto no afectó al ambiente de producción por lo que no se produjo una falta del servicio.

Considerando los problemas mencionados anteriormente se logra una disponibilidad superior al 99%, lo que significa que el 99% del tiempo en que se ejecutó el proyecto la plataforma estuvo disponible.

## 6.2.3 Usuarios

El sistema de visualización de ADATRAP implementa un sistema de usuarios y permisos que permite regular los accesos a distintos módulos de la aplicación y datos que pueden ser consultados. Actualmente existen dos tipos de usuarios: **lector** y **administrador**. Ambos usuarios tienen la posibilidad de consultar todas las vistas, la diferencia entre ellos radica en que el usuario administrador tiene acceso al módulo de administración de la plataforma, este módulo le permite:

- Crear, editar, desactivar o eliminar usuarios de la plataforma
- Dar o quitar permisos a cada usuario
- Ver el historial de descarga de datos de todos los usuarios
- Ver los procesos de descarga de datos de todos los usuarios que se estén realizando en el momento
- Subir datos a la plataforma de visualización
- Ver el historial de sesiones de todos los usuarios

Los usuarios administradores automáticamente tienen todos los permisos por lo que tienen acceso completo a la plataforma. Por otro lado, el usuario lector puede ver limitado su actuar de dos formas: acceso a vistas y acceso a datos. Los permisos asociados a las vistas son los siguientes:

- Viajes y estadísticas globales: usuarios que tienen este permiso tienen la posibilidad de ver todas las visualizaciones de la sección viajes y estadísticas globales
- Almacenamiento: usuarios que tienen este permiso tienen la posibilidad de ver la sección de almacenamiento, lo que les concede el derecho a descargar datos almacenados en la plataforma

Desde el punto de vista de los datos, los permisos están asociados a los operadores, es decir, un usuario lector puede tener acceso a leer todos o un subconjunto de operadores. Es importante comprender que este tipo de permisos tiene sentido cuando los filtros en las vistas contienen un selector de operador, esto ocurre en los módulos: perfil de carga y velocidades.

#### 6.2.3.1 Actividad

Para medir el uso del sistema existe un módulo interno que registra todas acciones que realiza un usuario junto con su fecha y hora, esto permite generar estadísticas de uso. Desde el mes de diciembre de 2018 existen 53 usuarios activos en la plataforma de visualización del ambiente de producción y durante el mes de abril de 2019 se incorporaron tres más. La base de este cálculo de actividad está basada en las sesiones, una sesión se define como aquel conjunto de llamadas que hace un usuario a la plataforma web que tienen una distancia temporal de 15 o más minutos con el resto de llamadas.

Durante los meses de monitoreo de la actividad de usuarios (diciembre de 2018 a mayo de 2019) se contabiliza un uso promedio de 17 usuarios distintos por mes. En la Tabla 3 se muestra la actividad de todos los usuarios de la plataforma. El usuario Jaime Miranda es el que presenta mayor número de sesiones mientras que el usuario Claudia Zarate es la que presenta la sesión promedio más larga. Es importante destacar que aquellos usuarios que no aparecen en la lista se a que nunca han registrado alguna actividad en la plataforma.

N°	Nombre usuario	N° sesiones	Tiempo máximo de sesión	Tiempo mínimo de sesión	Tiempo promedio de sesión
1	ignacio.riquelme	22	0:37:12	0:00:00	0:08:14
2	ana.leiva	1	0:00:01	0:00:01	0:00:01
3	yeissy.jara	2	0:01:06	0:00:24	0:00:45
4	pablo.herbage	4	0:02:46	0:00:02	0:01:32
5	martin.quiroz	9	0:22:27	0:00:01	0:05:55

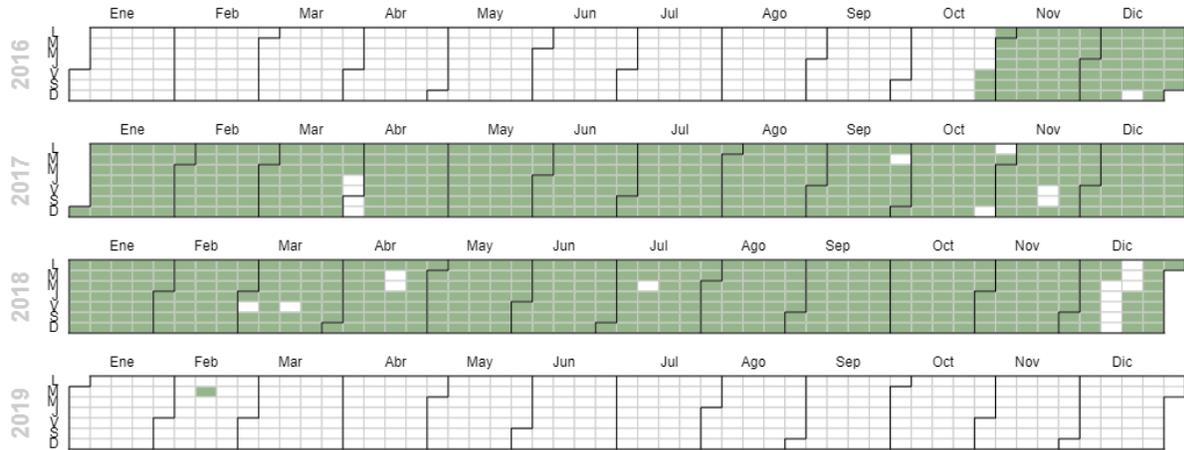
6	loreto.bravo	12	0:33:52	0:00:01	0:08:18
7	diego.cruz	2	0:01:44	0:00:52	0:01:18
8	jaime.miranda	92	1:04:42	0:00:00	0:09:50
9	celin.moreno	1	0:00:47	0:00:47	0:00:47
10	guillaume.ducroquet	4	0:19:08	0:01:09	0:06:41
11	diego.silva	37	0:49:45	0:00:00	0:05:23
12	patrick.oemick	19	0:19:54	0:00:00	0:07:36
13	muriel.manriquez	38	0:35:42	0:00:00	0:03:51
14	christian.romero	20	0:37:04	0:00:00	0:11:03
15	yessica.ortiz	5	0:22:51	0:00:30	0:07:32
16	marina.dragicevic	2	0:27:37	0:00:02	0:13:49
17	jaime.moya	10	1:35:47	0:00:08	0:12:39
18	elkin.ruiz	2	0:03:32	0:01:47	0:02:39
19	leonardo.victoriano	1	0:00:25	0:00:25	0:00:25
20	raul.cornejo	2	0:08:05	0:04:48	0:06:26
21	antonio.gschwender	1	0:11:25	0:11:25	0:11:25
22	michael.stefanic	7	0:12:34	0:00:47	0:03:38
23	nicolas.gomez	1	0:00:12	0:00:12	0:00:12
24	nathalia.maira	20	0:32:17	0:00:00	0:10:17
25	miguel.ormeno	11	0:11:45	0:00:00	0:02:36
26	sebastian.tamblay	6	0:37:08	0:00:03	0:08:33
27	claudia.zarate	1	0:19:06	0:19:06	0:19:06
28	barbara.poblete	7	0:34:59	0:00:03	0:10:57

Tabla 3: actividad de usuarios durante la ejecución del proyecto

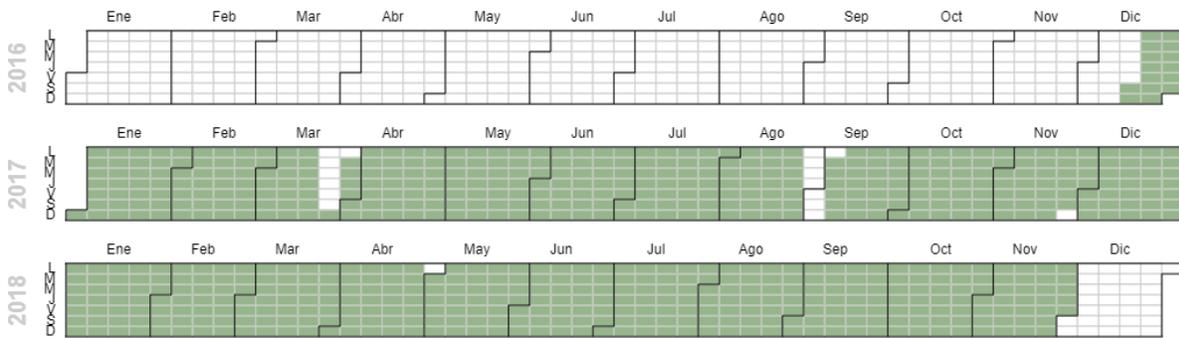
## 6.2.4 Almacenamiento de archivos

Una parte relevante del proyecto corresponde al almacenamiento de un conjunto de datos base que el área de planificación requiere periódicamente para realizar su gestión. Antes de este proyecto esta tarea era realizada a través de una cuenta FTP que no siempre tuvo el rendimiento requerido. A raíz de lo anterior se desarrolló un módulo para almacenar estos archivos en nube, sus principales características son: acceso privado, fácil acceso, simple y permite compartir links públicos temporales (durante 24 horas) a terceros.

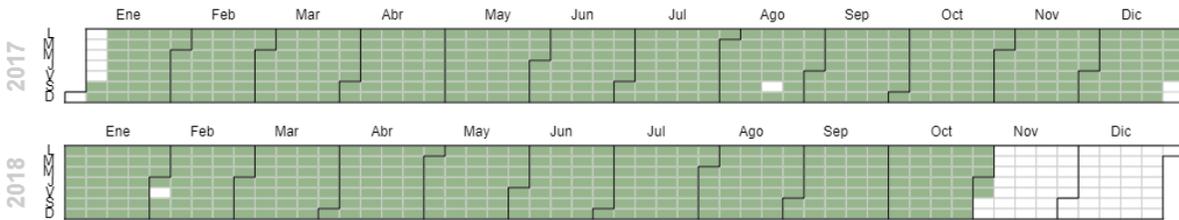
El almacenamiento de los archivos es realizado en el servicio *Simple Storage Service (S3)* de Amazon. Fue tarea del equipo durante la ejecución del proyecto subir parte de los datos almacenados localmente hasta ese momento. Durante el mes de febrero se inició el proceso de respaldo donde se almacenaron datos de GPS (Ilustración 10), transacciones (Ilustración 11) y 1.96 (Ilustración 12) de varios días.



*Ilustración 10: días de datos de GPS respaldados en S3*



*Ilustración 11: días de datos de transacciones respaldados en S3*



*Ilustración 12: días de datos de 1.96 respaldados en S3*

## 7 DESARROLLOS

A continuación, se describen todos los desarrollos realizados durante el proyecto.

### 7.1 Validación Asistida

Se modificó el programa de estimación de bajada en el módulo que contiene el detalle de variables para validar el procesamiento, esto con el fin de agilizar los tiempos de validación de los datos procesados. A continuación, se detallan los campos definidos para este archivo de validación:

CAMPO	DESCRIPCIÓN
<b>Fecha</b>	Fecha del día procesado
<b>TipoDia</b>	Tipo de día (LABORAL/SABADO/DOMINGO)
<b>version</b>	Versión del código con el que se procesaron los datos
<b>nExpediciones</b>	Cantidad de expediciones del Reporte 1.96
<b>minTiempoExpediciones</b>	Tiempo de la expedición más corta
<b>maxTiempoExpediciones</b>	Tiempo de la expedición más larga
<b>mediaTiempoExpediciones</b>	Tiempo medio de todas las expediciones
<b>nPatentes</b>	Número de patentes/buses en los datos
<b>nGPS</b>	Número de GPS del archivo de GPS del día
<b>mediaTiempoEntreGPS</b>	Tiempo medio entre GPS correlativas de un bus
<b>nGPSConServicio</b>	Número de GPS que están dentro de una expedición con servicio
<b>nGPSSinServicio</b>	Número de GPS que no están contenidas en alguna expedición
<b>nTrxTotales</b>	# total de validaciones del día
<b>nTrxTotalesBus(%)</b>	# validaciones hechas en BUS (% sobre el total)
<b>nTrxTotalesMetro(%)</b>	# validaciones hechas en METRO (% sobre el total)
<b>nTrxTotalesMetroTren(%)</b>	# validaciones hechas METRO-TREN (% sobre el total)
<b>nTrxTotalesZonasPagas(%)</b>	# validaciones hechas en Zonas Pagas (% sobre el total)
<b>nTarjetas</b>	# tarjetas distintas
<b>nTrxsConServicio</b>	# validaciones contenidas en alguna expe
<b>nTrxsSinServicio</b>	# validaciones fuera de expediciones
<b>nEtapasConBajadaBus(%)</b>	# etapas con bajada estimada en bus (% sobre etapas en bus)
<b>nEtapasConBajadaMetro(%)</b>	# etapas con bajada estimada en metro (% sobre etapas en metros)
<b>nEtapasConBajadaMetroTren(%)</b>	# etapas con bajada estimada en metro-tren (% sobre etapas metro-tren)
<b>nEtapasConBajadaZonasPagas(%)</b>	# etapas con bajada estimada en ZP (% sobre etapas zonas pagas)
<b>nViajes</b>	# Viajes estimados
<b>nViajes1E(%)</b>	# Viajes estimados de 1 Etapa

<b>nViajes2E(%)</b>	# Viajes estimados de 2 Etapas
<b>nViajes3E(%)</b>	# Viajes estimados de 3 Etapas
<b>nViajes4E(%)</b>	# Viajes estimados de 4 etapas
<b>nViajes5E(%)</b>	# Viajes estimados de 5 o más etapas
<b>nViajesSoloMetro(%)</b>	# Viajes con etapas solo en metro
<b>nViajesConAlgunaEtapaEnMetro(%)</b>	# con al menos una etapa hecha en metro
<b>nViajesSinBajadaFinal(%)</b>	# viajes sin bajada en última etapa
<b>tViajeTotal</b>	tiempo medio de todos los viaje
<b>dViajeTotal</b>	distancia media de todos los viaje
<b>vViajeTotal</b>	velocidad media de todos los viajes
<b>nViajeMediaPM</b>	# Viajes que inician en Punta Mañana
<b>tViajeMediaPM</b>	Tiempo medio viajes que inician en punta mañana
<b>dViajeMediaPM</b>	Distancia media de viajes que inician en punta mañana
<b>vViajeMediaPM</b>	Velocidad media de viajes que inician en punta mañana
<b>nViajeMediaPT</b>	# Viajes que inician en punta tarde
<b>tViajeMediaPT</b>	Tiempo medio de viaje que inician en punta tarde
<b>dViajeMediaPT</b>	Distancia media de viajes que inician en punta tarde
<b>vViajeMediaPT</b>	Velocidad media de viajes que inician en punta tarde
<b>nTrxPM(%)</b>	#Validaciones bips realizadas en punta mañana
<b>nTrxPT(%)</b>	#Validaciones bips realizadas en punta tarde
<b>nBajadas</b>	Cantidad de bajadas (suma sobre expansión zonal-periodo)
<b>nBajadasPM</b>	Cantidad de bajadas en punta mañana (suma sobre expansión)
<b>nBajadasPT</b>	Cantidad de bajadas en punta tarde (suma sobre expansión)
<b>nParadasE</b>	# Paradas de tipo E
<b>nParadasT</b>	# Paradas de tipo T
<b>nParadasL</b>	# Paradas de tipo L
<b>nParadasI</b>	# Paradas de tipo I
<b>nTrxE(%)</b>	# Validaciones en paradero de tipo E
<b>nTrxT(%)</b>	# Validaciones en paradero de tipo T
<b>nTrxL(%)</b>	# Validaciones en paraderos de tipo L
<b>nTrxI(%)</b>	# Validaciones en paraderos de tipo I
<b>par1</b>	Paradero con más validaciones en el dia
<b>par2</b>	2° paradero con más validaciones
<b>par3</b>	3° paradero con más validaciones
<b>par4</b>	4° paradero con más validaciones
<b>par5</b>	5° paradero con más validaciones
<b>par6</b>	6° paradero con más validaciones
<b>par7</b>	7° paradero con más validaciones
<b>par8</b>	8° paradero con más validaciones
<b>par9</b>	9° paradero con más validaciones
<b>par10</b>	10° paradero con más validaciones
<b>trx1</b>	# validaciones paradero con más validaciones

<b>trx2</b>	# validaciones 2° paradero con más validaciones
<b>trx3</b>	# validaciones 3° paradero con más validaciones
<b>trx4</b>	# validaciones 4° paradero con más validaciones
<b>trx5</b>	# validaciones 5° paradero con más validaciones
<b>trx6</b>	# validaciones 6° paradero con más validaciones
<b>trx7</b>	# validaciones 7° paradero con más validaciones
<b>trx8</b>	# validaciones 8° paradero con más validaciones
<b>trx9</b>	# validaciones 9° paradero con más validaciones
<b>trx10</b>	# validaciones 10° paradero con más validaciones
<b>parBus1</b>	Paradero de buses con más validaciones
<b>parBus2</b>	2° paradero de buses con más validaciones
<b>parBus3</b>	3° paradero de buses con más validaciones
<b>parBus4</b>	4° paradero de buses con más validaciones
<b>parBus5</b>	5° paradero de buses con más validaciones
<b>parBus6</b>	6° paradero de buses con más validaciones
<b>parBus7</b>	7° paradero de buses con más validaciones
<b>parBus8</b>	8° paradero de buses con más validaciones
<b>parBus9</b>	9° paradero de buses con más validaciones
<b>parBus10</b>	10° paradero de buses con más validaciones
<b>trx1</b>	# validaciones paradero de buses con más validaciones
<b>trx2</b>	# validaciones 2° paradero de buses con más validaciones
<b>trx3</b>	# validaciones 3° paradero de buses con más validaciones
<b>trx4</b>	# validaciones 4° paradero de buses con más validaciones
<b>trx5</b>	# validaciones 5° paradero de buses con más validaciones
<b>trx6</b>	# validaciones 6° paradero de buses con más validaciones
<b>trx7</b>	# validaciones 7° paradero de buses con más validaciones
<b>trx8</b>	# validaciones 8° paradero de buses con más validaciones
<b>trx9</b>	# validaciones 9° paradero de buses con más validaciones
<b>trx10</b>	# validaciones 10° paradero de buses con más validaciones

Se desarrolló un programa tipo script capaz de detectar días anómalos que no son ingresados al servidor de producción para no ingresar ruido a los análisis. Para esto se desarrolló un programa en Python que toma dos archivos de entrada, el primero corresponde a días que son seleccionados como días sin problemas y un segundo con todos los días que quieren validarse.

El programa utiliza el primer archivo para armar las estadísticas generales, en este caso, una media y una desviación estándar, se agrupan los días por tipo de día y una vez que procesado este primer archivo el programa pasa a revisar los días uno a uno del segundo archivo, donde se realiza un filtro sobre varias variables respecto a un rango dado por 4 veces su desviación estándar. El conjunto de variables que se revisan fueron seleccionadas como las variables más significativas para detectar una variación o problema relevante. La tabla a continuación muestra los nombres de las variables que fueron seleccionados para realizar el análisis, las

descripciones en detalle de cada una de ellas pueden encontrarse en el informe de avances del mes de septiembre.

Nombre Variable
<b>nExpediciones</b>
<b>nGPS</b>
<b>nTrx</b>
<b>nTrxBus</b>
<b>nTrxMetro</b>
<b>nTrxMetroTren</b>
<b>nTrxZonasPagas</b>
<b>nTarjetas</b>
<b>nEtapasConBajadaBus(%)</b>
<b>nEtapasConBajadaMetro(%)</b>
<b>nEtapasConBajadaMetroTren(%)</b>
<b>nEtapasConBajadaZP(%)</b>
<b>nViajes</b>

Tabla 4: Tabla con variables seleccionadas para el análisis.

Finalmente, el script entrega una tabla con los días que no pasan el test y una secuencia de todas las variables que no pasaron el test para este día en particular, un ejemplo de salida se muestra a continuación:

DIA	Variable1	Variable2	Variable3	Variable 4	Variable 5
<b>13/05/2018</b>	nExpediciones	nGPS	nTrx	nTrxBus	nTrxMetro
<b>10/15/18</b>	nExpediciones	nGPS	nTrx	nTrxBus	nTrxMetro
<b>14/09/2018</b>	nExpediciones	nViajeMediaPT			
<b>17/09/2018</b>	nExpediciones	nGPS	nTrx	nTrxBus	nTrxMetro
<b>27/06/2018</b>	nGPS				

Tabla 5: Ejemplo de salida del script, día con las variables que no pasaron el test.

La actual versión del script acepta como entrada los archivos de días en el formato ya utilizado para las variables de validación, el código fuente del script se encuentra en el informe de diciembre 2019.

## 7.2 Levantamiento de servidores en la nube

Se duplicó el servidor de desarrollo, levantando así el servidor de producción en la nube (amazon), se está evaluando de parte del DTPM que días de datos se subirán al sistema.

URL producción: <https://www.adatrap.cl/>

URL desarrollo: <https://desarrollo.adatrap.cl>

En una primera versión este sistema consiste en un cluster de 3 servidores con la herramienta Elastic Search para el manejo de los datos. También fueron activadas y configuradas las

herramientas de monitoreo del sistema tanto para el servidor de desarrollo como el de producción.

### 7.3 Construcción de análisis de estimación de bajada en zonas pagas

Se integró al módulo de estimación de bajada en zonas pagas un conjunto de elementos para poder trackear cualquier problema que pudiese existir con la metodología, datos de entrada o algoritmo con el objetivo de iniciar un exhaustivo análisis y mejorar toda posible falla. A la fecha el módulo es capaz de detectar inconsistencias en: determinación de servicios factibles, obtención de frecuencia de operación, obtención de velocidades, cálculo de tiempo de viaje, cálculo de líneas comunes, determinación de bus más probable y estimación de paradero y tiempo de bajada.

### 7.4 Modificación formato de entrada Zonas Pagas

Dentro de la investigación que se realizó para mejorar la estimación de bajada en zonas pagas se detectó un problema que consiste que la implementación actual no es capaz de identificar servicios variantes en las zonas pagas, lo que produce un desajuste en la distribución de las validaciones en los distintos servicios del paradero y en algunos casos impacta directamente en la no estimación de bajada. Con esto identificado se realizaron las siguientes correcciones a los datos de entrada y su correspondiente modificación en el software:

1. Fusión de tabla de validadores a tabla de servicios en zonas pagas, se eliminó el actual archivo de Diccionario-ValidadoresZP.csv y se integró esa misma información al archivo Diccionario-DetalleServicio.csv.
2. Cambio de codificación de servicios en tabla de zonas pagas, este cambio consiste en cambiar la codificación del código de servicio actual (código de usuarios) por la codificación SONDA que permite identificar servicios variantes.

Como resultado de estas modificaciones se logró mejorar la estimación de bajada de un promedio diario cercano al 63% a uno de 68%, y en conjunto con eso mejoró la distribución de validaciones entre los servicios.

### 7.5 Comparación de distribución con encuestas vs ADATRAP

Dentro del análisis realizado para buscar mejoras en la estimación de bajada y distribución de validaciones en zonas pagas se dio paso a la comparación de una encuesta realizada en abril 2017 contra las distribuciones construidas por ADATRAP. En este análisis se encontró un error medio de 6%, donde las principales diferencias radican en cambios de Programa de Operación

entre las fechas de la encuesta y los datos de ADATRAP, a continuación, se muestra un gráfico comparativo entre ambas fuentes de información:

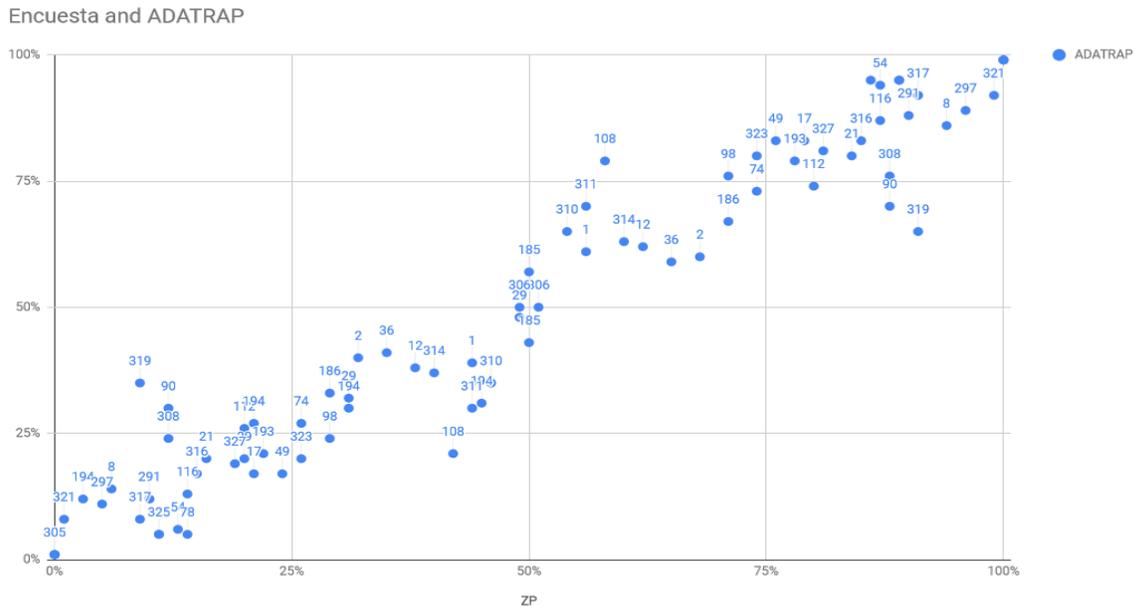


Ilustración 13: Gráfico comparativo de distribuciones de por servicio en zonas pagas, ADATRAP eje X vs Encuestas eje Y.

El gráfico nos muestra la comparación directa del porcentaje de participación de cada unidad de negocio para cada zona paga, donde el eje X representa el porcentaje en ADATRAP, el eje Y el porcentaje en la encuesta y el número sobre el punto es el id de la zona paga. Los puntos con mayores diferencias 319, 90, 108 y 308 fueron analizados con mayor profundidad y se aprecia una cierta inconsistencia en la encuesta puesto que las distribuciones están mucho mejor representadas en ADATRAP dado las frecuencias reales de los servicios en estos paraderos, se propone hacer alguna medición en esos paraderos en particular para despejar dudas. A continuación, una tabla informativa de estas zonas pagas:

Tipo	Id	Validador	Paradero	Nombre Paradero	Servicios
<b>Falta UN 4</b>	90	RM-0464	E-20-290-OP-10	Parada 5 / Estación Mapocho	T402 00R/T409 00I/T121 00I/T119 00I
<b>Falta UN 7</b>	108	RM-0481	E-32-254-SN-5	Parada 2 / (M) Macul	T104 00R/T106 00R/T114 00R/T126 00R/F52 00I
<b>Falta UN 4</b>	194	RM-0766	E-17-12-SN-35	Parada 8 / (M) Escuela Militar	F52 00I/T425 00R/B68 00I/T219 E0 00I

<b>Falta UN 6</b>	310	RM-0882	T-6-327-PO-3	Avenida Lo Marcoleta esq. / Pquia. Jesús Obrero	T314 00I/T314 E0 00I/B08 00I/B12 00R
<b>Falta UN 6</b>	311	RM-0883	L-6-24-5-PO	Avenida Lo Marcoleta esq. / Ntra. Sra. del Carmen	T307 00I/T308 00I/T315 E0 00I/B06 00R/B08 00I/B11 00R/B13 00R/B21 00R
<b>Falta UN 1</b>	319	RM-0891	E-9-75-SN-12	Parada 2 / (M) Pudahuel	T110 00R/T110 C0 00R/T534 C0 00R

*Tabla 6: Tabla con zonas pagas con datos de encuesta inconsistentes*

### 7.5.1 Construcción de archivos de distribución

Para los análisis de los puntos anteriores se integró al software la capacidad de generar 3 archivos de salida nuevos, el primero corresponde a los datos de las distribuciones por servicios en cada zona paga, el segundo contiene las distribuciones por servicio-media hora de cada zona paga y finalmente el tercero contiene la información de las distribuciones agregadas por unidad de negocio por cada zona paga.

### 7.5.2 Comparación con mediciones 2018 (noviembre 2018)

#### Comparación con encuesta 2018 vs ADATRAP

Tal como fue realizada una comparación de los resultados de las distribuciones de validaciones entre operadores en zonas pagas con una encuesta del 2017, se repitió este análisis con datos de encuestas obtenidos en 2018. La tabla 1 muestra el número de zonas pagas involucradas en las 3 fuentes de datos.

Ítem	#ZP	# ZP - UN
<b>Encuesta 2017</b>	41	81
<b>Encuesta 2018</b>	75	167
<b>Adatrap</b>	77	171

*Tabla 7: Cantidad de observaciones*

Con estos datos se realizó una comparación uno a uno con las distribuciones por operador entre las encuestas y los resultados de ADATRAP, dando como resultado los valores que muestra la tabla 2.

	Error Medio	Casos > 15% Error
<b>Encuesta 2017</b>	<b>5.9%</b>	3 zps ( 2 arregladas en 2018)
<b>Encuesta 2018</b>	<b>10.3%</b>	18 zps (4 arreglables a la fecha, falta revisar mayoría)

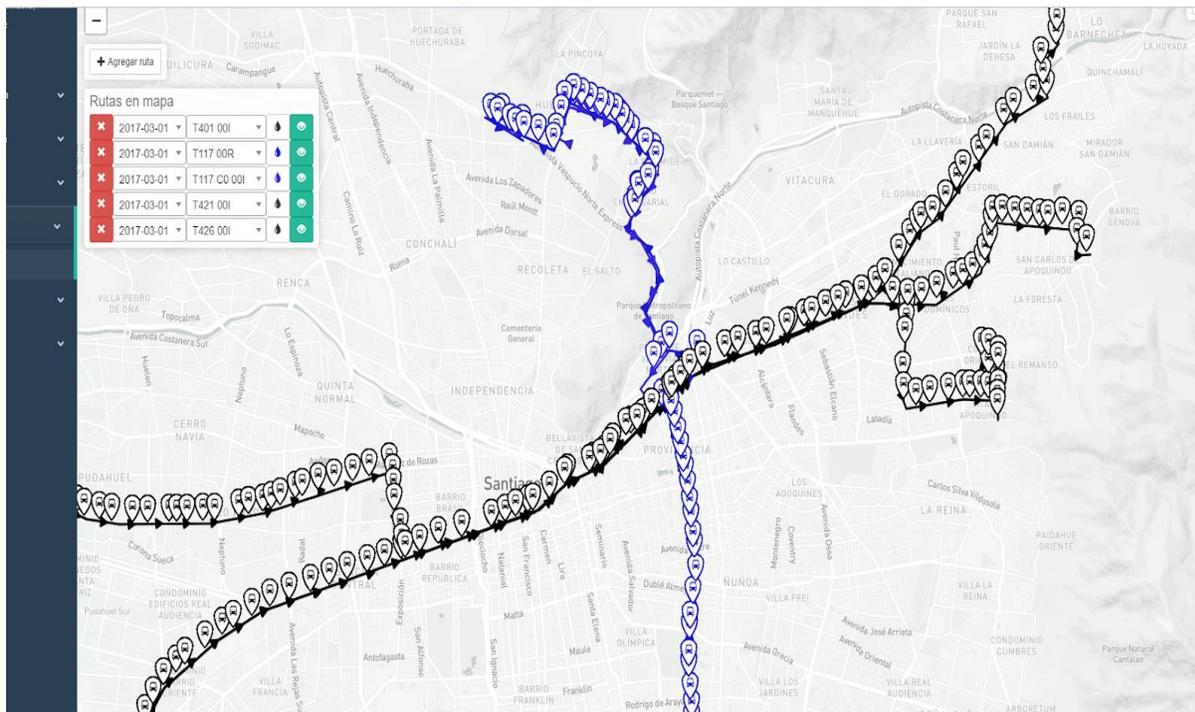
Tabla 8: Resultados de comparaciones

Cabe hacer notar que al comparar con los datos de la encuesta 2018 se nota un aumento en el error medio, al revisar estos resultados con la contraparte se identificó un problema en los datos utilizados en la comparación por lo que queda pendiente un recálculo de esta comparación en el siguiente periodo, donde dado el error mencionado es probable que este error medio disminuya.

Por otra parte, se analizaron en detalle las zonas pagas con mayor error (> 15%) y este análisis se detalla a continuación:

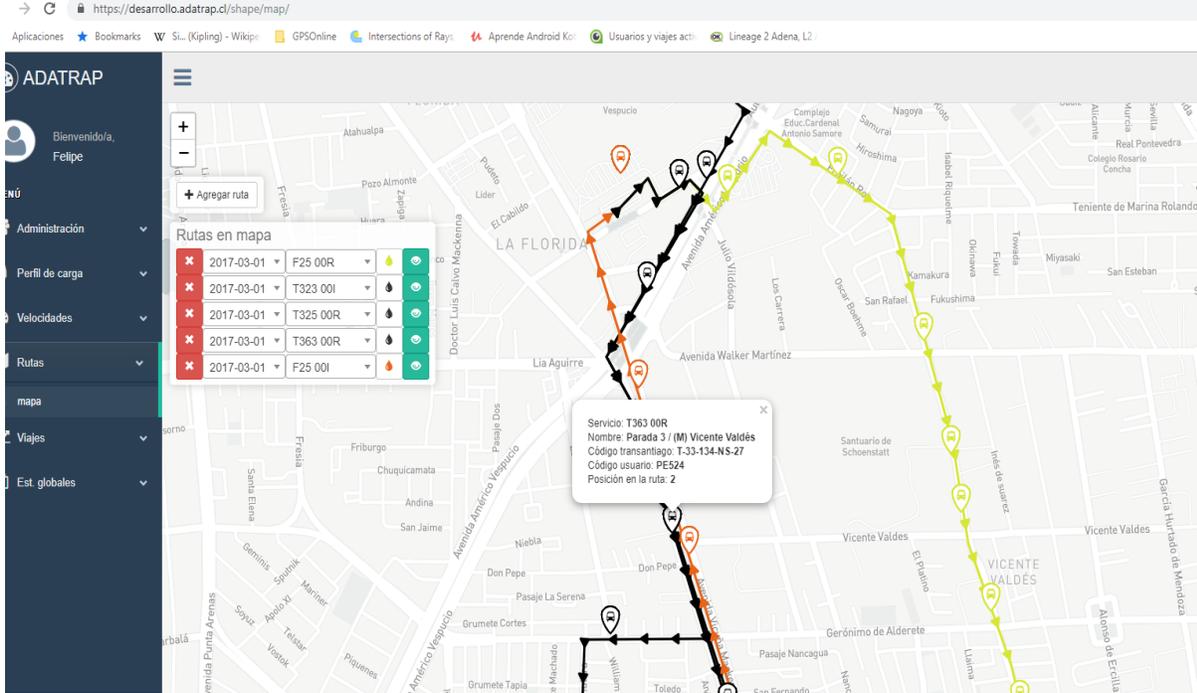
### Caso ZP (36)

Este caso presenta una diferencia muy significativa entre la encuesta 2018 (U1 9%, U4 91%) vs ADATRAP (U1 41%, U4 59%), llama la atención que los servicios que pasan por esa zona paga divergen absolutamente por operador por lo que la metodología de ADATRAP debiese tener un resultado muy cercano a la realidad, dado que está basada en los destinos de los usuarios. La figura a continuación muestra como todos los servicios del operador 1 se dirigen hacia el sector norte y los servicios del operador 4 hacia el oriente. Por lo que se propone una segunda medición para asegurar las proporciones.



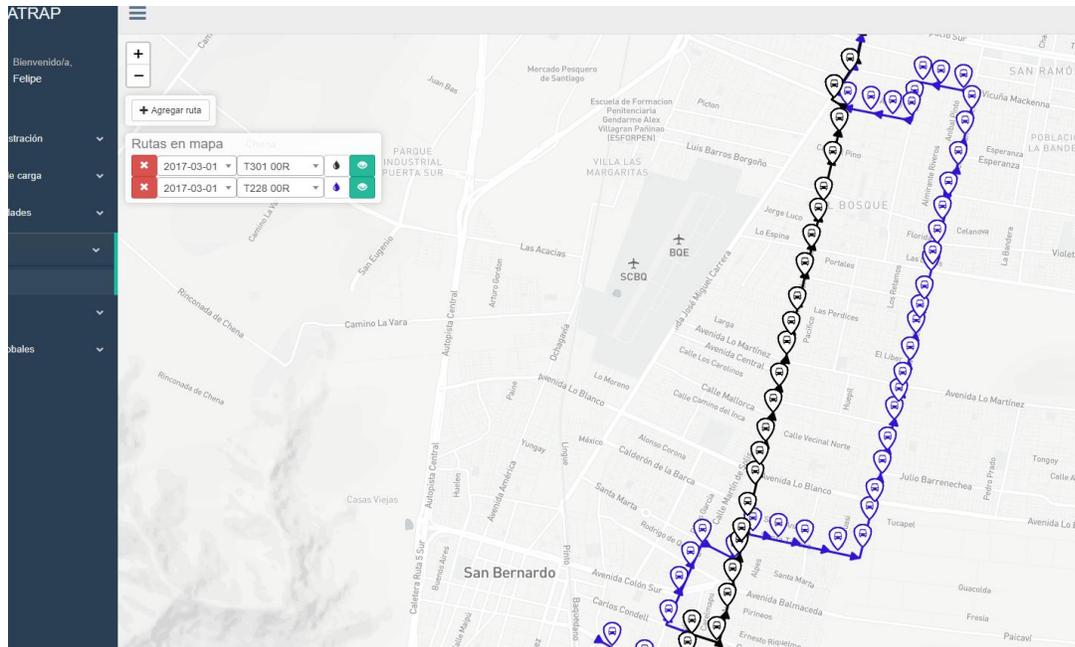
## Caso ZP ( 72 )

En este caso se detectó que la parada del servicio asignado a la zona paga ( F25R ) no corresponde, pero si en el sentido opuesto ( F25I ), por lo que se propone corregir esa información. La imagen a continuación muestra el caso en particular.



## Caso ZP (303 / 308)

Este caso sugiere una falta de ajuste en la metodología de líneas comunes puesto que todos los usuarios son asignados al servicio 301R, por otro lado, la contra parte sugiere que la encuesta podría tener algún error porque es sabido que el servicio 228R no para en esa zona paga, por lo que se sugiere una nueva medición. La figura a continuación muestra la situación en el mapa, donde la zona paga está en el principio de la imagen (al sur) y prácticamente todos los usuarios se dirigen a la estación multimodal de la Cisterna al norte.



## Caso ZP (1 y 2)

El caso más significativo en términos de error corresponde a la zona paga de Escuela Militar, donde los porcentajes de distribución están prácticamente invertidos, donde la encuesta habla de un 65% U4 y 35% U6 y por otro lado ADATRAP estima 39% U4 y 61% U6. Para esta zona paga se realizaron los siguientes análisis.

1. **Distribución de destinos:** Se analizaron los destinos utilizando la información de ADATRAP y se determinó que un 70% de los usuarios baja en donde las rutas de ambos operadores son comunes, un 20% tienen como destino lugares donde solo se puede llegar usando servicios de U6 y un 10% tiene como destino lugares donde solo se puede acceder usando servicios U4. Esto último nos podría sugerir una mayor demanda para el operador U6 dando más la razón a ADATRAP.
2. **Medición en terreno:** Se realizó una inspección en terreno por 30 minutos por parte del equipo de operación de ADATRAP, y se observó una distribución similar a la proporcionada en la encuesta 2018. Además, se detectó que los buses de la U6 son pequeños y van todos a máxima capacidad.
3. **Sobrecarga:** Se sabe que dentro de ADATRAP existe un pequeño porcentaje de sobrecarga en zona pagas, por lo que con la inspección en terreno donde se apreciaba que los buses de la U6 iban demasiado cargados se dio paso a hacer un análisis dentro de los perfiles de carga generados por ADATRAP para revisar si existe algún problema en relación a esto. Al revisar los perfiles se comprobó una sobrecarga considerable en los buses del operador U6, y al hacer la corrección repartiendo los usuarios extras de la sobrecarga hacia el otro operador se mejoró notablemente el error de la distribución. Por último, como se mencionó al principio de esta sección se encontró un error en la elección de los datos de las encuestas de 2018, por lo que se debe rehacer el análisis, pero para este caso ya se revisó y el error entre encuesta y ADATRAP quedó bajo el 5%.

### 7.5.3 Comparación con mediciones 2018 (diciembre 2018)

Durante diciembre se detectó un error en ese estudio correspondiente al uso de un campo errado para identificar el porcentaje de distribución de las encuestas el cual correspondía a la distribución solo por frecuencia de los servicios, es por esto que se volvió a realizar la comparación con el día 24 de octubre de 2018, con lo que se obtuvo un error medio de 9,94% y la distribución de este error en las distintas zonas pagas se muestra en la Tabla 9 ordenados de mayor a menor error.

ZP	Error	ZP	Error	ZP	Error	ZP	Error	ZP	Error
36	44%	74	15%	319	8%	353	4%	367	1%
2	38%	392	14%	385am	8%	193	3%	398	1%
9	35%	378	13%	403	7%	319	3%	304	1%
316	32%	351am	13%	17	7%	317	3%	305	1%
356pm	32%	388	13%	98	7%	413	3%	402	1%
186	31%	331	13%	112	7%	348	3%	354	1%
409	31%	354	12%	395	7%	108	3%	8	0%
360	26%	185	12%	317	7%	116	3%	327	0%
306	26%	307	12%	407	6%	387	2%	303	0%
308	23%	90pm	11%	385pm	5%	311	2%		
356am	23%	12	11%	291	5%	335	2%		
412	22%	21	10%	390	4%	297	2%		
351pm	20%	194	10%	375	4%	404	2%		
31	19%	78	9%	310	4%	90am	2%		
323	17%	386	9%	314	4%	300	1%		
368	17%	29	9%	26	4%	355	1%		
366	16%	313	8%	377	4%	352	1%		
1	16%	322	8%	405	4%	325	1%		
192	15%	54	8%	321	4%	406	1%		

*Tabla 9: Distribución del error en las zonas pagas al comparar ADATRAP vs Encuestas 2018*

Con esta información se dio paso al análisis de las zonas pagas con un error mayor al 10%, el detalle de cada uno de estos análisis realizados a la fecha de este informe se explica a continuación. Todo el detalle de este estudio puede encontrarse en el informe de diciembre 2018. En síntesis, se analizaron las zonas pagas con mayores diferencias porcentuales, de las cuales se encontraron casos donde la diferencia provenía de un error en los datos del programa de operación, o de problemas de definición de inicio fin de un servicio-sentido y algunos casos donde no se encontró explicación a través de los datos y donde se sugiere realizar una revisión en terreno para verificar.

## 7.5.4 Salida de distribuciones en Zonas Pagas

Con el fin de entregar un producto que logre ser lo más fácil de integrar a los procesos ya contruidos por el DTPM para resolver el pago en zona pagas mixtas se definió un formato para entregar los resultados de los datos de distribuciones por unidad de negocio en las zonas pagas. La tabla contiene 11 campos, la descripción de estos datos se muestra en la Tabla 10.

Es importante explicar que la etapa anterior y posterior a cada zona paga, se construye con la metodología de estimación de viajes de ADATRAP, esta consiste en unir las etapas de una misma tarjeta que cumplan con un conjunto de condiciones explicadas en detalle en el capítulo 5.3.

<b>Campo</b>	<b>Descripción</b>
<b>ZP</b>	Código de la zona paga
<b>TipoDia</b>	Tipo de día (LABORAL/SABADO/DOMINGO)
<b>MediaHora</b>	Media hora de la validación en la zona paga
<b>ParadaUsuario</b>	Código de usuario de la parada
<b>ParadaTS</b>	Código de Transantiago de la parada
<b>Validador</b>	Código del validador (RM-XXX)
<b>Etapa Anterior</b>	Unidad de negocio etapa anterior a validación en zona paga
<b>Etapa en ZPM</b>	Unidad de negocio de la etapa en zona paga
<b>Etapa Posterior</b>	Unidad de negocio de la etapa posterior a zonas pagas
<b>Viajes</b>	Cantidad de viajes de la triada UN1/UN2/UN3, sin expansión
<b>#Dias</b>	Cantidad de días agregados en el tipo de día correspondiente

*Tabla 10: Formato de archivo de salida para la distribución de validaciones por operador en zonas pagas mixtas*

## 7.6 Mantención de parámetros de consulta en visualizador

Se agregó a la página web la característica de recordar los parámetros utilizados en la última consulta, esto permite reutilizar datos ya completados agilizando la experiencia del usuario con la información en alguna de las vistas en la página web.

## 7.7 Extracción de GPS

Se dio paso a la migración del servicio que extrae en forma continua y tiempo real los GPS de los buses, dentro de la migración se tuvo que realizar un desarrollo adicional por la razón de incompatibilidades de las librerías con las actualizaciones del sistema operativo Linux Ubuntu,

a la fecha se utilizaba la versión 16.04 pero fue necesario actualizar a la 18.04 por temas con la base de datos PostgreSQL, esto a su vez trajo consigo problemas con las librerías de consultas lipqxx utilizadas en el software de extracción en C++, además el actual módulo de extracción que utiliza la librería Boost ha presentado recurrentemente problemas al parsear el archivo JSON entregado por el web service, es por esto que se tomó la decisión de separar el módulo de extracción y actualización de la base de datos del software en C++ para llevarlos a uno en lenguaje Python que ha mostrado ser más estable.

## 7.8 Registro de actividades de usuario

Se agregó un sistema de registro de actividades de usuario, esto con el objetivo de conocer con más detalle las actividades que los usuarios realizan dentro de la página web. Es así como en cada llamada al servidor se registran los siguientes datos:

1. Fecha y hora de la consulta
2. URL que se está consultando
3. Tiempo de respuesta de la consulta
4. Usuario que está consultando

Los resultados del análisis de este registro fueron mostrados en el capítulo 6 de este informe, o pueden ser revisados en mayor detalle en cada uno de los informes mensuales entregados por este trabajo.

## 7.9 Sección de almacenamiento

Se agregó un módulo nuevo en el menú lateral de nombre “Almacenamiento”, este contiene la información que ha sido respaldada en la nube, específicamente en el servicio S3 de AWS. Los archivos por respaldar son de nueve tipos:

1. GPS: posicionamiento de los buses durante un día
2. Viajes: viajes de un día construidos por el software ADATRAP, contiene más información que los arrojados por el visualizador
3. Programa de operación: archivos del programa de operación
4. Semana representativa: corresponde a los datos de la semana utilizada por DTPM como representativa de un año
5. 1.96: archivo que contiene datos de las expediciones realizadas
6. Perfiles de carga: perfiles de carga de un día construido por el software ADATRAP, contiene más información que los arrojados por el visualizador
7. Etapas: Datos de etapas de un día construidos por ADATRAP
8. Velocidades: velocidades de los servicios durante un día construido por el software ADATRAP, contiene más información que los arrojados por el visualizador

## 9. Transacciones: archivo con las transacciones realizadas durante un día

Los archivos almacenan la información de un día salvo los tipos “semana representativa” y “programa de operación”. El primero contiene siete días y el segundo es un archivo cuyos datos se pueden usar varios días o meses. Para los tipos de datos por día se construyó una vista con una tabla y un calendario. La vista presenta una lista de los archivos con un link de descarga, mientras que el calendario es una representación visual de los días almacenados en el respaldo. En la ilustración 17 se muestra la vista de los datos GPS respaldados, en el calendario, los cuadros verdes representan días con datos.

Por último, se crea un permiso a aquellos usuarios que se definan para que puedan visualizar la sección de almacenamiento, dejándoles descargar los archivos disponibles.

**Almacenamiento GPS**

Días disponibles

2016  
2017  
2018  
2019

Detalle

Mostrar 10 registros

Nombre	Última modificación	Tamaño	Descargar
2019-02-12.gps.gz	13/2/2019 14:37:24	195.23 MB	descargar
2018-12-31.gps.gz	13/2/2019 14:37:02	193.26 MB	descargar
2018-12-30.gps.gz	13/2/2019 14:36:40	152.14 MB	descargar
2018-12-29.gps.gz	13/2/2019 14:36:20	173.50 MB	descargar
2018-12-28.gps.gz	13/2/2019 14:35:56	217.82 MB	descargar
2018-12-27.gps.gz	13/2/2019 14:35:32	219.67 MB	descargar
2018-12-26.gps.gz	13/2/2019 14:35:09	215.83 MB	descargar
2018-12-25.gps.gz	13/2/2019 14:34:50	151.36 MB	descargar
2018-12-24.gps.gz	13/2/2019 14:34:27	205.31 MB	descargar
2018-12-23.gps.gz	13/2/2019 14:34:05	158.39 MB	descargar

Mostrando registros del 1 al 10 de un total de 773 registros

Anterior 1 2 3 4 5 ... 78 Siguiente

Ilustración 14: Vista que muestra los días con datos GPS respaldados en la nube

## 7.10 Programa para carga de archivos de respaldo a la nube

Se construyó un programa que automatiza la carga de archivos en la nube (AWS Amazon), este puede ser encontrado en el siguiente link:  
<https://github.com/SmartcitySantiagoChile/S3FileUploader>

Sus principales características son:

1. Permite subir uno o más archivos a la vez
2. Permite eliminar archivos

Este programa será utilizado para realizar los respaldos de los nueve tipos de archivos descritos anteriormente. Información de su uso, instalación y configuración puede ser encontrada en el link reportado en el párrafo anterior. Las tablas que se almacenan y la forma de acceso a ellos se describen en la sección anterior 7.9.

## 7.11 Homologación de formato de GPS

Como ya se mencionó desde enero 2019 la fuente de datos GPS presento problemas de caídas, por lo que se optó por utilizar una nueva fuente de esta información, pero esta nueva fuente (minería de datos SONDA) tenía otro formato en sus datos, para solucionar esto se construyó un script en Python el cual reformatea los archivos dejándolos con el mismo formato que requiere ADATRAP. El código fuente y mayores detalles de este desarrollo se pueden revisar en el informe de abril 2019.

## 7.12 Construcción de datos de Metro L3

Una segunda tarea consistió en la integración de datos de la nueva línea 3 de Metro, para esto se tuvieron que modificar los archivos de líneas.txt, horarios.csv, combinación.csv y diccionario de estaciones, los detalles de las modificaciones pueden ser revisadas en la sección de anexo 8.2. Y al archivo Lion.csv se le integraron expediciones ficticias con un tiempo de 2 minutos entre estaciones para poder correr ADATRAP.

## 7.13 Actualización almacenamiento en producción

El módulo de almacenamiento fue habilitado en el ambiente de producción. Para poder acceder a este módulo se requiere un permiso adicional que hoy en día solo el usuario "barbara.poblete" tiene, por lo que es la única que lo puede visualizar.

## 7.14 Procesamiento automático de actividad de usuarios

El script para calcular la actividad de usuarios se integró dentro de la plataforma de visualización para calcular de manera automática estos datos cada 24 horas, además se implementaron dos vistas web, una para ver el detalle de las sesiones y otra que contiene una tabla resumen de las sesiones por cada usuario. Esta sección está dentro del módulo de administración por lo que es visible para los usuarios que poseen ese permiso.

En la Ilustración 15 se muestra el historial de actividades, en cada registro se detalla:

- Usuario: nombre de usuario
- Tiempo de inicio: fecha y hora en que se inició la sesión
- Tiempo de fin: fecha y hora en que terminó la sesión
- Duración: tiempo que duró la sesión

User	Start time	End time	Duration
jacqueline.arriagada	12 de Abril de 2019 a las 13:50	12 de Abril de 2019 a las 13:50	0:00:09
transantiago	12 de Abril de 2019 a las 13:47	12 de Abril de 2019 a las 13:53	0:05:37
transantiago	11 de Abril de 2019 a las 21:31	11 de Abril de 2019 a las 21:31	0:00:03
transantiago	11 de Abril de 2019 a las 17:09	11 de Abril de 2019 a las 17:10	0:00:47
transantiago	9 de Abril de 2019 a las 13:51	9 de Abril de 2019 a las 13:52	0:01:29
transantiago	9 de Abril de 2019 a las 11:19	9 de Abril de 2019 a las 11:30	0:11:00
transantiago	9 de Abril de 2019 a las 10:16	9 de Abril de 2019 a las 10:17	0:00:16
transantiago	8 de Abril de 2019 a las 20:05	8 de Abril de 2019 a las 20:05	0:00:18
transantiago	8 de Abril de 2019 a las 18:13	8 de Abril de 2019 a las 18:15	0:01:16
transantiago	8 de Abril de 2019 a las 17:57	8 de Abril de 2019 a las 17:57	0:00:14
transantiago	8 de Abril de 2019 a las 17:09	8 de Abril de 2019 a las 17:09	0:00:12

*Ilustración 15: Historial de sesiones de usuarios*

Por otro lado, en la Ilustración 16 se muestra la tabla resumen de actividad, por cada fila se describe la siguiente información:

- Usuario: nombre de usuario
- N° sesiones: número de sesiones que ha realizado el usuario
- Tiempo de última sesión: fecha y hora de inicio de la última sesión
- Máxima duración: tiempo máximo que ha durado una sesión
- Mínima duración: tiempo mínimo que ha durado una sesión
- Tiempo promedio de sesión: promedio de la duración de todas las sesiones del usuario

User	Session number	Last session timestamp	Max session duration	Min session duration	Avg session duration
jacqueline.arriagada	1	12 de Abril de 2019 a las 13:50	0:00:09	0:00:09	0:00:09
alejandro.tirachini	5	26 de Marzo de 2019 a las 10:56	0:13:50	0:03:13	0:08:46
dtpm	18	18 de Febrero de 2019 a las 16:29	0:13:27	0:00:00	0:02:49
felipe.tapia	5	16 de Febrero de 2019 a las 11:07	0:11:08	0:03:05	0:05:34
memorista	2	12 de Febrero de 2019 a las 02:30	0:07:54	0:03:36	0:05:45
transantiago	79	5 de Febrero de 2019 a las 19:37	0:44:12	-7 days, 9:18:01	-1 day, 20:46:14

6 user session stats

*Ilustración 16: Resumen de actividad por usuario*

Se construyó un script para procesar la actividad de los usuarios registrada en el servidor de producción, adatrapp.cl. Este código calcula las siguientes métricas:

- Usuarios con actividad
- Número de sesiones por usuario
- Tiempo máximo de sesión
- Tiempo mínimo de sesión
- Tiempo promedio de sesión

## 7.15. Vista de Accesos solicitados a datos históricos

En el período anterior se agregó la sección de almacenamiento, esta da acceso a todos los archivos almacenados en el servicio S3 de AWS (Amazon Web Service), como transacciones, gps, viajes, perfiles de carga, entre otros. Para acceder a alguno de los archivos es necesario realizar una solicitud a través de la herramienta de visualización puesto que su acceso es restringido, esta solicitud genera una URL que permite descargar el archivo durante un período de 24 horas, luego de cumplido ese tiempo no es posible acceder al recurso (archivo de datos). Para tener control sobre las solicitudes que se han realizado se implementó una vista en la sección de administración que muestra las solicitudes de acceso que han solicitado los usuarios de la plataforma. En la Ilustración 17 se muestra la vista con tres solicitudes (filas). Cada solicitud muestra la siguiente información:

- Recurso: archivo al que se solicitó acceso
- Fecha: día y hora en que se solicitó el acceso
- Expiración: día y hora en que el acceso solicitado deja de ser válido
- Está activo: indica si la solicitud está vencida (cuando esto ocurre indica que la URL generada no permite descargar el recurso)

- Link: botón que permite descargar el recurso

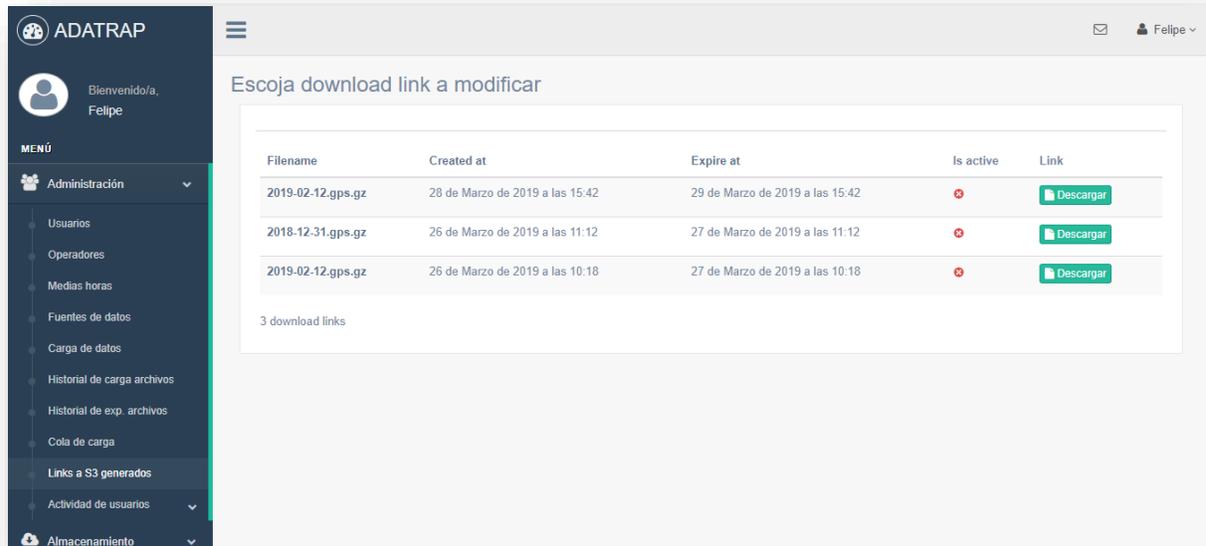


Ilustración 17: Vista de solicitudes de acceso a archivos almacenados

## 8. ANEXOS

### 8.1 Script homologación GPS

```
import sys
import csv
#SE ABRE EL ARCHIVO ORIGINAL, QUE DEBE SER CAMBIADO DE FORMATO
def creararchivo(nombre):
    archivo = csv.reader(open(nombre, 'rt'), delimiter=';')
    return archivo
#SIRVE PARA SEPARAR LAS FECHAS, DADO QUE EN EL ARCHIVO ORIGINAL VIENE
EN,EL SIGUIENTE FORMATO "AAAAMMDDHORA"
def separarfecha(archivo):
    newfecha = archivo[0]+archivo[1]+archivo[2]+archivo[3]+"-"+archivo[4]+archivo[5]+"-"+archivo[6]+archivo[7]+"
"+archivo[8]+archivo[9]+":"+archivo[10]+archivo[11]+":"+archivo[12]+archivo[13]
    return newfecha
#SE CREA UNA LISTA CON TODOS LOS NOMBRES DE LOS ARCHIVOS QUE DEBEN SER
ARREGLADOS
```

```

def crearmes(mes,año):
    meses
    ['enero','febrero','marzo','abril','mayo','junio','julio','agosto','septiembre','octubre','noviembre','diciembre']
    j=0
    for k in range(12):
        if mes == meses[k]:
            j=k+1
            if (j == 1) | (j==3) | (j==5) | (j==7) | (j==8) | (j==10) | (j==12):
                lista = [0]*31
            if j<10:
                for i in range(31):
                    i=i+1
                    if i<10:
                        lista[i-1]='Detalle_Registros_GPS_'+str(año)+'0'+str(j)+'0'+str(i)+".csv"
                    else:
                        lista[i-1]='Detalle_Registros_GPS_'+str(año)+'0'+str(j)+str(i)+".csv"
                print (lista[i-1])
            else:
                for i in range(31):
                    i=i+1
                    if i<10:
                        lista[i-1]='Detalle_Registros_GPS_'+str(año)+str(j)+'0'+str(i)+".csv"
                    else:
                        lista[i-1]='Detalle_Registros_GPS_'+str(año)+str(j)+str(i)+".csv"
                print (lista[i-1])
            elif j == 2:
                lista = [0]*28
                for i in range(28):
                    i=i+1
                    if i<10:
                        lista[i-1]='Detalle_Registros_GPS_'+str(año)+'0'+str(j)+'0'+str(i)+".csv"
                    else:
                        lista[i-1]='Detalle_Registros_GPS_'+str(año)+'0'+str(j)+str(i)+".csv"
                print (lista[i-1])
            else:
                lista = [0]*30
                if j<10:
                    for i in range(30):
                        i=i+1
                        if i<10:
                            lista[i-1]='Detalle_Registros_GPS_'+str(año)+'0'+str(j)+'0'+str(i)+".csv"
                        else:
                            lista[i-1]='Detalle_Registros_GPS_'+str(año)+'0'+str(j)+str(i)+".csv"

```

```

    print (lista[i-1])
else:
    for i in range(30):
        i=i+1
        if i<10:
            lista[i-1]='Detalle_Registros_GPS_'+str(año)+str(j)+'0'+str(i)+".csv"
        else:
            lista[i-1]='Detalle_Registros_GPS_'+str(año)+str(j)+str(i)+".csv"
        print (lista[i-1])
    return lista
#CAMBIA DE FORMATO LOS ARCHIVOS
def cambiargps(mes,año):
    j=0
    meses = ['enero', 'febrero', 'marzo', 'abril', 'mayo', 'junio', 'julio', 'agosto', 'septiembre', 'octubre',
    → 'noviembre', 'diciembre']
    for k in range(12):
        if mes == meses[k]:
            numeromes=k+1
    lista = crearmes(mes,año)
    for dia in lista:
        archivo = creararchivo(dia)
        next(archivo,None)
        if j<9:
            if numeromes<10:
                final = open(str(año)+"-0"+str(numeromes)+"-0"+str(j+1)+".gps", "w")
            else:
                final = open(str(año)+"-"+str(numeromes)+"-0"+str(j+1)+".gps", "w")
        else:
            if numeromes<10:
                final = open(str(año)+"-0"+str(numeromes)+"-"+str(j+1)+".gps", "w")
            else:
                final = open(str(año)+"-"+str(numeromes)+"-"+str(j+1)+".gps", "w")
        for row in archivo:
            final.write(row[0]+";")
            if row[4] == "":
                final.write(" ;")
            else:
                final.write(row[4]+";")
            final.write(separarfecha(row[3])+";")
            final.write(row[1]+";"+row[2]+";"+"-1;"+"-1;"+"row[5]+";"+"-1;"+"-1;\n")
        j=j+1
    cambiargps('diciembre',2018)

```

## 8.2 Modificaciones datos de metro L3

### **Lineas.txt :**

L1=LD, HM, MQ, EM, AL, GO, TB, LE, PV, MM, SA, BA, UC, SL, CH, LM, LH, RP, LA, EL, US, PG, EC, LR, P  
A, NE, SP  
L2=AV, ZA, DO, EI, CE, CB, PT, CA, AN, HE, TO, PQ, RO, FR, LL, SM, LV, DE, CN, LO, EP, LC  
**L3=LIB, CAR, VIV, CCH, PZC, HOS, CA3, PZA, CH3, PAL, MAT, IRA3, MEY, NNO3, ESP, VFR, PZE3,  
FCV**  
L4=TOB, COL, BIL, PDG, SBO, PEG, LOR, RGR, LPR, RQU, LTO, MAC, VMA, VVA, RMA, TRI, SJE, LQ  
U, ECO, HSR, PIN, LME, PPA  
L4A=VIM, SJU, LGR, SRO, SRA, LCI  
L5=PM, SB, DS, MT, LP, LS, BR, PU, SO, PR, BL, GL, QN, RC, NA, PZ, BE, BQ, PB, SI, IR, NU, RA, CV, AG,  
SJ, PE, MA, LF, VV  
L6=LEO, IDS, NNO, ENA, NUB, BIO, FRA, PAC, LVA, CER  
L2\_R=1,1,1,0,1,0,1,1,1,1,0,1,0,1,1,0,1,1,0,1  
L2\_V=1,1,0,1,0,1,0,1,1,1,1,0,1,1,0,1,0,1,0,1,1,1  
L4\_R=1,0,1,1,0,1,1,0,1,0,1,1,1,1,0,1,0,1,1,1,0,1,1  
L4\_V=1,1,1,0,1,1,0,1,0,1,0,1,1,1,1,0,1,0,1,1,1,0,1  
L5\_R=1,1,0,1,0,1,1,1,1,0,1,0,1,0,1,1,1,1,0,1,1,0,1,0,1,0,1,1,1  
L5\_V=1,0,1,0,1,1,0,1,1,1,0,1,0,1,1,1,1,1,0,1,1,1,1,0,1,0,1,0,1,1

### **Horarios.csv:**

linea;horario;dia;inicio;fin;express  
L1;ADL;LABORAL;5:30:00;23:45:00;NO  
L1;ADL;SABADO;6:30:00;23:45:00;NO  
L1;ADL;DOMINGO;8:00:00;23:45:00;NO  
L2;ADL;LABORAL;5:30:00;23:45:00;NO  
L2;ADL;SABADO;6:30:00;23:45:00;NO  
L2;ADL;DOMINGO;8:00:00;23:45:00;NO  
**L3;ADL;LABORAL;5:30:00;23:45:00;NO**  
**L3;ADL;SABADO;6:30:00;23:45:00;NO**  
**L3;ADL;DOMINGO;8:00:00;23:45:00;NO**  
L4;ADL;LABORAL;5:30:00;23:45:00;NO  
L4;ADL;SABADO;6:30:00;23:45:00;NO  
L4;ADL;DOMINGO;8:00:00;23:45:00;NO  
L4A;ADL;LABORAL;5:30:00;23:45:00;NO  
L4A;ADL;SABADO;6:30:00;23:45:00;NO  
L4A;ADL;DOMINGO;8:00:00;23:45:00;NO  
L5;ADL;LABORAL;5:30:00;23:45:00;NO  
L5;ADL;SABADO;6:30:00;23:45:00;NO  
L5;ADL;DOMINGO;8:00:00;23:45:00;NO  
L6;ADL;LABORAL;5:30:00;23:45:00;NO  
L6;ADL;SABADO;6:30:00;23:45:00;NO  
L6;ADL;DOMINGO;8:00:00;23:45:00;NO

### **Combinacion.csv**

linea\_origen;estacion\_origen;linea\_destino;estacion\_destino;entrada\_unica;nombre\_estacion  
 L1;LH;L2;HE;NO;Los Heroes  
 L1;TB;L4;TOB;NO;Tobalaba  
 L1;SP;L5;SO;NO;San Pablo  
 L1;BA;L5;BQ;NO;Baquedano  
 L1;LE;L6;LEO;NO;Los Leones  
 L2;LC;L4A;LCI;NO;La Cisterna  
 L2;AN;L5;NA;NO;Santa Ana  
 L2;FR;L6;FRA;SI;Franklin  
**L3;CA3;L2;CA;NO;Puente Cal y Canto**  
**L3;PZA;L5;PZ;NO;Plaza de Armas**  
**L3;CH3;L1;CH;NO;Universidad de Chile**  
**L3;IRA3;L5;IRA;NO;Irrazaval**  
**L3;NNO3;L6;NNO;NO;Nunoa**  
**L3;PZE3;L4;PZE;NO;Plaza Egana**  
 L4;VMA;L4A;VIM;SI;Vicuna Mackena  
 L4;VVA;L5;VV;SI;Vicente Valdes  
 L5;NU;L6;NUB;SI;Nuble

### **Diccionario de estaciones:**

Se agregaron los siguientes registros:

Los Libertadores;QUILICURA;-33.365439;-70.69244;L3;LOS LIBERTADORES;NORMAL;LOS  
 LIBERTADORES;LIB;N  
 Cardenal Caro;CONCHALI;-33.373452;-70.686104;L3;CARDENAL CARO;NORMAL;CARDENAL  
 CARO;CAR;N  
 Vivaceta;CONCHALI;-33.385877;-70.679226;L3;VIVACETA;NORMAL;VIVACETA;VIV;N  
 Conchalí;CONCHALI;-33.397985;-70.669546;L3;CONCHALI;NORMAL;CONCHALI;CCH;N  
 Plaza Chacabuco;INDEPENDENCIA;-33.406565;-70.661028;L3;PLAZA CHACABUCO;NORMAL;PLAZA  
 CHACABUCO;PZC;N  
 Hospitales;INDEPENDENCIA;-33.417663;-70.656373;L3;HOSPITALES;NORMAL;HOSPITALES;HOS;N  
 Universidad de Chile;SANTIAGO;-33.443937;-70.65039;L1;UNIVERSIDAD DE  
 CHILE;TRASBORDO;UNIVERSIDAD DE CHILE;CH;N  
 Universidad de Chile L3;SANTIAGO;-33.443809;-70.650391;L3;UNIVERSIDAD DE CHILE  
 L3;TRASBORDO;UNIVERSIDAD DE CHILE;CH3;N  
 Vicente Valdes;LA FLORIDA;-33.526338;-70.596757;L4;VICENTE VALDES;TRASBORDO;VICENTE  
 VALDES;VVA;C  
 Parque Almagro;SANTIAGO;-33.451381;-70.650491;L3;PARQUE ALMAGRO;NORMAL;PARQUE  
 ALMAGRO;PAL;N  
 Matta;SANTIAGO;-33.458237;-70.642682;L3;MATTA;NORMAL;MATTA;MAT;N  
 Vicente Valdes;LA FLORIDA;-33.526338;-70.596757;L5;VICENTE VALDES;TRASBORDO;VICENTE  
 VALDES;VV;C  
 Monseñor Eyzaguirre;NUNOA;-33.453164;-70.613835;L3;MONSEÑOR  
 EYZAGUIRRE;NORMAL;MONSEÑOR EYZAGUIRRE;MEY;N  
 Vicuna Mackena;LA FLORIDA;-33.519642;-70.596066;L4A;VICUNA  
 MACKENNA;TRASBORDO;VICUNA MACKENNA;VIM;N  
 Chile-España;NUNOA;-33.454949;-70.59748;L3;CHILE ESPANA;NORMAL;CHILE ESPANA;ESP;N  
 Villa Frei;NUNOA;-33.454618;-70.580987;L3;VILLA FREI;NORMAL;VILLA FREI;VFR;N  
 Vicuna Mackena;LA FLORIDA;-33.519642;-70.596066;L4;VICUNA MACKENNA;TRASBORDO;VICUNA  
 MACKENNA;VMA;C  
 Fernando Castillo Velasco;LA REINA;-33.452065;-70.55792;L3;FERNANDO CASTILLO  
 VELASCO;NORMAL;FERNANDO CASTILLO VELASCO;FCV;N

## 8.3 Validación automática

```

import sys
import csv
import numpy as np

#Indice de las columnas de las variables
col_fecha=0
col_tipoDia=1

#Arreglo con los dias considerados malos
diasMalos = []
diasMuyMalos = []

#Clase para almacenar estadisticos de cada variable
class Variable:
    def __init__(self,name,ncol):
        self.name = name
        self.nobs = 0.0
        self.sum = 0.0
        self.media = -1
        self.devstd = -1
        self.fechas = []
        self.ncol = ncol

def obtenerEstadisticos(variable,tipoDia):
    #Arreglo con los datos del archivo
    data = []

    # Lectura de archivo con los datos
    spamReader = csv.reader(open('seleccionados.general', 'rb'), delimiter=';', quotechar=' ')

    # Ciclo de lectura de los datos base
    i=0
    for row in spamReader:
        if i>0 :

            if row[variable.ncol] == "-nan(ind)" or row[variable.ncol] == "nan(ind)" :
                diasMuyMalos.append((row[col_fecha],tipoDia,variable.name))
            elif np.isinf(float(row[variable.ncol])) or np.isinf(float(row[variable.ncol])):
                diasMuyMalos.append((row[col_fecha],tipoDia,variable.name))
            else:
                if row[variable.ncol] != '-' \

```

```

and row[col_tipoDia]==tipoDia:
    variable.fechas.append(row[col_fecha])
    data.append(float(row[variable.ncol]))
  
```

```

    i+=1
  
```

```

# Calculo de estadisticos
variable.media = np.mean(data)
variable.devstd = np.std(data)
variable.nobs = np.alen(data)
  
```

```

#Funcion para procesar por variable
  
```

```

def chequeaVariable(variable,tipoDia):
    #Arreglo con los datos del archivo
    data = []
  
```

```

    # Lectura de archivo con los datos
    spamReader = csv.reader(open('todos.general', 'rb'), delimiter=';', quotechar=' ')
  
```

```

    # Ciclo de lectura de los datos base
    i=0
  
```

```

    for row in spamReader:
        if i>0 :
  
```

```

            if row[variable.ncol] == "-nan(ind)" or row[variable.ncol] == "nan(ind)" :
                diasMalos.append((row[col_fecha],tipoDia,variable.name))
            elif np.isinf(float(row[variable.ncol])) or np.isinf(float(row[variable.ncol])):
                diasMalos.append((row[col_fecha],tipoDia,variable.name))
            else:
  
```

```

                if row[variable.ncol] != '-' and row[col_tipoDia]==tipoDia:
                    if float(row[variable.ncol]) < variable.media -
4*variable.devstd or variable.media + 4*variable.devstd < float(row[variable.ncol]):
  
```

```

                    diasMalos.append((row[col_fecha],tipoDia,variable.name))
                    i+=1
  
```

```

#Tipo dia a revisar
tipoDias = []
  
```

```
tipoDias.append("LABORAL")  
tipoDias.append("SABADO")  
tipoDias.append("DOMINGO")
```

```
#Objetos para almacenar estadisticos  
variables = []  
variables.append(Variable("nExpediciones",3))  
variables.append(Variable("nGPS",8))  
variables.append(Variable("nTrx",12))  
variables.append(Variable("nTrxBus",13))  
variables.append(Variable("nTrxMetro",14))  
variables.append(Variable("nTrxMetroTren",15))  
variables.append(Variable("nTrxZonasPagas",16))  
variables.append(Variable("nTarjetas",17))  
variables.append(Variable("nEtapasConBajadaBus(%)",20))  
variables.append(Variable("nEtapasConBajadaMetro(%)",21))  
variables.append(Variable("nEtapasConBajadaMetroTren(%)",22))  
variables.append(Variable("nEtapasConBajadaZP(%)",23))  
variables.append(Variable("nViajes",24))  
variables.append(Variable("nViajes1E(%)",25))  
variables.append(Variable("nViajes2E(%)",26))  
variables.append(Variable("nViajes3E(%)",27))  
variables.append(Variable("nViajesSoloMetro(%)",30))  
variables.append(Variable("tViajeTotal",33))  
variables.append(Variable("nViajeMediaPM",36))  
variables.append(Variable("nViajeMediaPT",40))  
variables.append(Variable("nParadasE",49))  
variables.append(Variable("nParadasT",50))  
variables.append(Variable("nParadasL",51))  
variables.append(Variable("nParadasI",52))
```

```
#Chequeo por variable  
for tipodia in tipoDias:  
    for variable in variables:  
        obtenerEstadisticos(variable,tipodia)  
        chequeaVariable(variable,tipodia)
```

```
# Diccionario de dias con problemas  
dias = {}  
for diaMalo in diasMalos:  
    dias[diaMalo[0]] = []  
  
for diaMalo in diasMalos:  
    dias[diaMalo[0]].append(diaMalo[2])
```

```
for dia in dias:  
    print dia + " : ",  
    for variable in dias[dia]:  
        print variable+"/",  
    print ""
```