

## **Extracción de datos de redes sociales para monitorear el servicio de transporte de la Ciudad de Buenos Aires**

### **Resumen**

En esta ponencia se presenta un estudio sobre el funcionamiento del subterráneo de la Ciudad de Buenos Aires realizado mediante el uso de la herramienta R. El trabajo se basa en el registro y procesamiento de la información pública brindada por la empresa Metrovías S.A, responsable del servicio de transporte subterráneo de la Ciudad desde hace 25 años, a través de su cuenta oficial de Twitter. Los datos de esa red social se procesaron mediante una codificación original elaborada *ad hoc* para su posterior sistematización y análisis.

El transporte es un derecho esencial de los habitantes de una ciudad dado que afecta el bienestar cotidiano de quienes deben utilizarlo para trabajar, estudiar o realizar diversas actividades. En este sentido, el transporte público cobra una importancia aún mayor ya que permite la movilización de grandes masas de personas a un costo menor que el transporte privado y descongestiona las ciudades.

En la Ciudad de Buenos Aires circulan seis líneas de transporte subterráneo (A, B, C, D, E y H) gestionadas por la empresa Metrovías S.A. desde el 1ro de enero de 1994 cuando se privatizó la provisión de este servicio previamente gestionado por la estatal SBASE. El recorrido total de la red es de 60,8 km, de los cuales 54,7 km los son de tramo comercial, en los que se distribuyen 87 estaciones. En 2016, el conjunto de las líneas transportó a 303.859.273 pasajeros, según informó el Ministerio de Desarrollo Urbano y Transporte de la Ciudad.

Pese a la relevancia del conocimiento sobre el funcionamiento del transporte público, tanto del saber especializado como de aquel que afecta el uso cotidiano de los habitantes, la Ciudad de Buenos Aires actualmente no presenta estadísticas de acceso público y rápido sobre las problemáticas presentadas en la prestación del servicio de subterráneos. Esta falta de información obstaculiza los debates públicos sobre la calidad del servicio.

Este estudio representa la única fuente de información de acceso público y gratuito actualizada mes a mes con los datos brindados en forma virtual y desagregada por la empresa que presta el servicio. A su vez, abre la posibilidad de replicar esta aplicación en cada ciudad con sistemas de información vía Twitter sobre las problemáticas presentes en los sistemas de transporte.

Con respecto a la metodología utilizada, se relevaron los tweets de la empresa que se encuentran formateados de la siguiente manera (Imagen 1): línea que presenta un problema, la característica del problema y la hora de publicación del tweet.

(Imagen 1)



Cada tweet se extrajo con el paquete “twitter”, a través de la función userTimeline. Una vez realizada la extracción mediante las funciones de “R base” y del paquete “tidyverse” (fundamentalmente se utilizaron los paquetes “dplyr”, “lubridate” y “stringr”), se procedió a desagregar el mensaje de Twitter según diferentes dimensiones de forma de poder analizar cada aspecto en particular como se observa en la siguiente tabla a modo de ejemplo.

Texto	Creación	Línea	Fecha	Hora	Mensaje	Justificación
#LineaE - Circula con demora. 08:18	2018-10-31 11:18:42	#LineaE	2018-10-31	11:18	demora	no justificado
#LineaA - Por obras en zona de vías. 05:17	2018-10-31 8:17:40	#LineaA	2018-10-31	8:17	limitado	justificado
#LineaD - Circula con demora. 18:22	2018-10-30 21:22:35	#LineaD	2018-10-30	21:22	demora	no justificado
#LineaB - Servicio limitado. 14:57	2018-10-30 17:57:33	#LineaB	2018-10-30	17:57	limitado	no justificado
#LineaB - Circula con demora. 14:50	2018-10-30 17:50:32	#LineaB	2018-10-30	17:50	demora	no justificado

El reporte de los datos es accesible de manera libre y gratuita a través de un link de la plataforma Data Studio. Se espera que el estudio constituya un aporte en el acceso a información sobre servicios públicos y amplíe las posibilidades de debate ciudadano.

## Demoras totales por mes y línea

