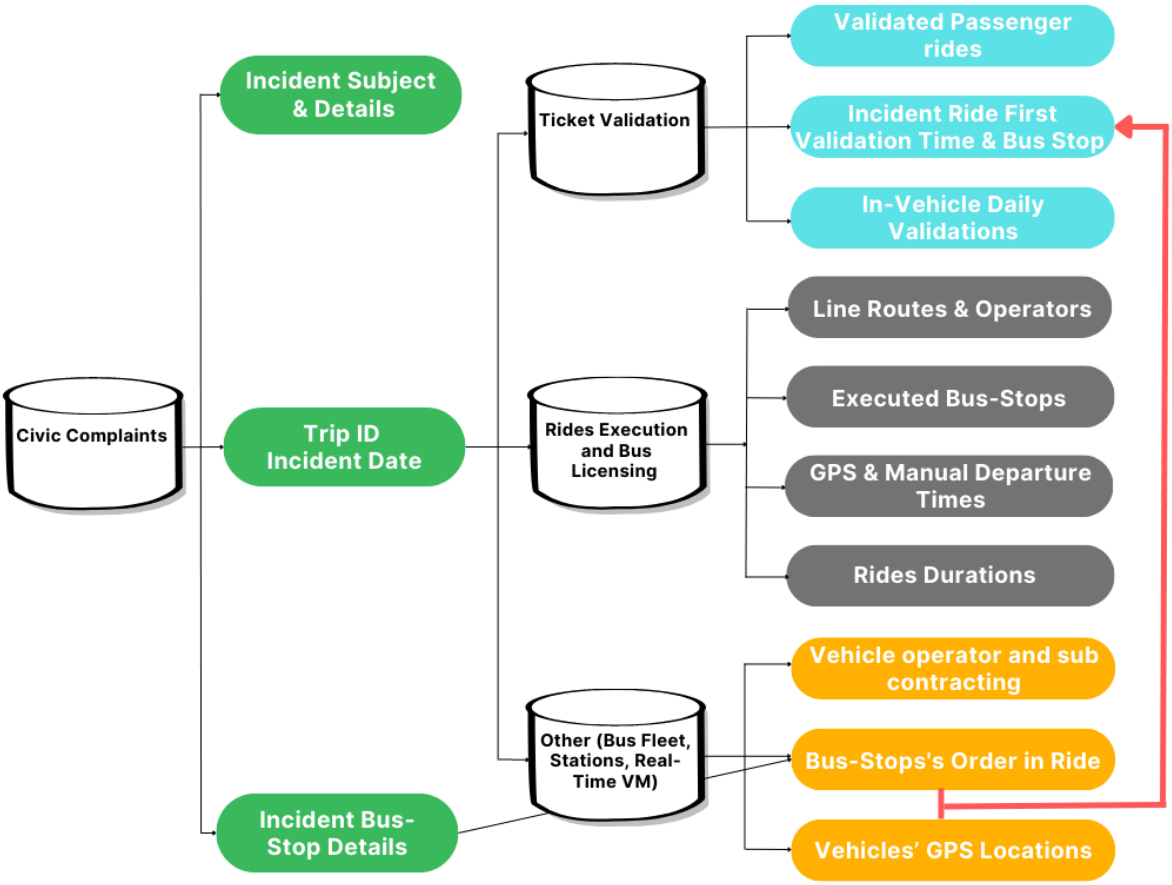
Using Big Data science and passenger complaints for debugging public transport

Yodfat Ben-Shalom, Sigal Kaplan

Transit service reliability is important for increasing passenger satisfaction and maintaining transit ridership, contributing to environmental and societal benefits. Transport authorities are increasingly concerned with monitoring transit operations to detect, locate, predict, and prevent service disruptions. Big data are available from regular transport operations, but data are underused for service disruption detection. Service complaints form a powerful indicator of locating service disruptions that were experienced by passengers, and they are becoming more abundant with digital advancements. The Israeli National Public Transport Authority (NPTA) maintains a civic complaint system with 91,368 annual complaints, of which 32,000 are about reliability issues. However, so far these complaints were checked separately only for enforcement purposes. Recent studies argue that smart handling of complaints can be used for better management of smart cities, and have shown that using data science for system-wide complaint analysis can generate important insights for improving public transport.

This study suggests a new approach to using passenger complaints to improve the automatic detection and prevention of service disruptions. Specifically, we suggest combining public complaints with big data related to transit operations, to identify structural weaknesses in the system and amend them. The study uses the NPTA's data set including 26 million annual rides on 10,200 buses along 7181 bus lines (rural, urban, and inter-urban) operated by 34 agencies. The data is compiled from several resources including GPS vehicle tracking, bus fleet data, ticket validations, and the GTFS line operations data (flow chart 1). The two main data sources, ticket validations, and the GTFS have respective storage capacities of 900 and 195 million kilobytes. The analysis includes: i) crossing complaints with their relevant service information from the operations database by using SQL server, ii) testing the validity of each complaint whether it is true or false, iii) for validated (true) complaints, classification analysis is conducted to identify "detection rules" that can be associated with each event, iv) building a 'troubleshooting model' for identifying undetected events in other trips and in the hope of 'debugging' them (flow chart 2). Preliminary findings include the following detection rules: reported trips are often characterized by absent ticket validations, exceptionally short travel durations, off-track bus locations, and manual modification of departure and arrival times.

Flow chart 1 – the matching data process.



Flow chart 2 – conceptual 'troubleshooting model'