



Commuters

Introduction

One of the travel attributes which Network Rail (NR) are interested in, as part of the RDIS system, is the proportion of travellers making commuter trips. Citi Logik developed a process by which commuter trips can be identified from analysing the inferred home and work locations of devices making rail trips. This note summarises the methodology used to identify the commuter trips, it then presents the file format and provides the AP access details.

Methodology

A commuting trip is deemed to be a trip made by anyone travelling from home to work, or vice versa. These are primarily in the peak working hours (i.e., 6 to 9am or 5 to 7pm respectively) but are not fixed to these. Citi Logik identify a device's home location as its predominant night-time cell location. In principle, this translates to the cell at which a device has the longest dwell time during the night; and, correspondingly, a device's work location is defined as the cell at which the device has the longest dwell time during the day period (i.e., between 10:00 and 16:00) as observed using long term trend analysis (over 3 months) on device behaviour.

NOTE: The derivation of the Home and Work cells is based on long term trend analysis of a device. Typically, the analysis is undertaken using 3 months' worth of data. Because of the current behaviour towards travelling to work, identifying work locations can be difficult as workers, especially office based, rarely travel to work 5 days a week currently. The system is designed to pick up multiple work locations to reflect on more current behaviours toward travelling to work, but if people do not travel 2 to 3 days a week to a regular place of work the system may not identify the work location. In addition, as the system undertakes analysis over a 3 month period, it will take time for the system to detect new work places as people start to travel to work again. This can explain the perceived low level of commuters in the system at the moment, but this should raise as more regular patterns are observed.

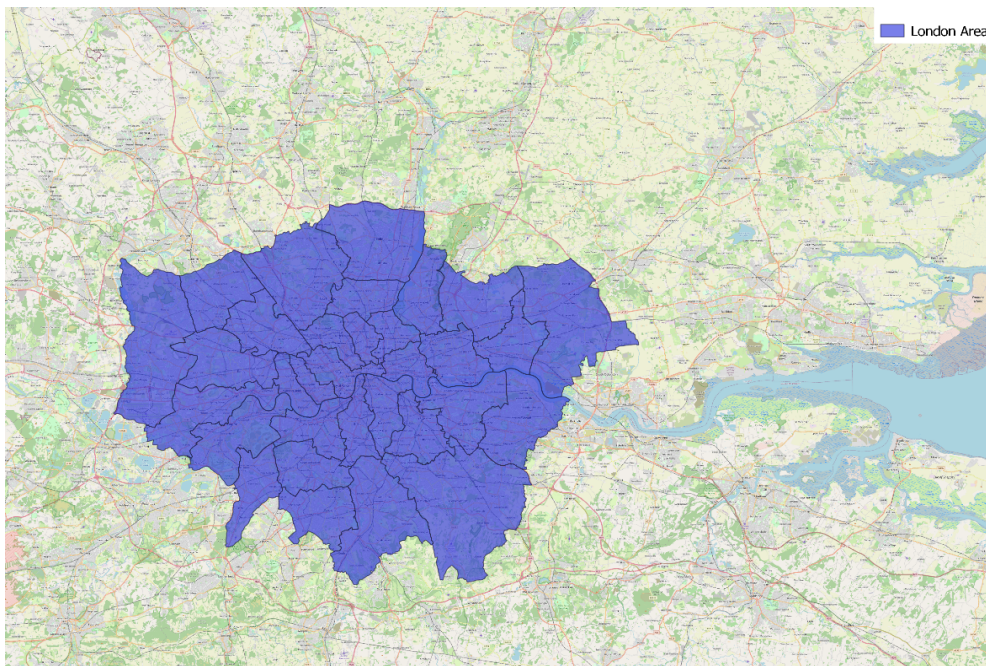
Devices' home and work locations are initially converted from cell to LSOA¹ level, which are subsequently related to trips' origin and destination stations or, especially for trips interacting with London, districts containing devices' home and work LSOAs. Hence, the key driver for defining a commuting trip is:

- Origin station in district of the Home or Work; and
- Destination station in district of the other end (Home if Work at origin or Work if Home at Origin).

Since LSOAs are very small in area, commuting trips are further constrained to be such that the LSOA containing a trip's origin station is different from the LSOA containing its destination station.

¹ Lower Layer Super Output Areas: <https://data.gov.uk/dataset/c481f2d3-91fc-4767-ae10-2efdf6d58996/lower-layer-super-output-areas-lsoas>

Figure 1. The London sector used to identify commuters



When it comes to London, a large sector was defined to include the City of London, Westminster, and surrounding districts (31 in total) and potential commuting trips (as derived based on home and work locations) originating from, or destined to, this district in the PM and AM peaks, respectively, were deemed to be commuting trips, this is to reflect the fact that commuters are likely to use an onward mode, different to rail (bus, underground...), to reach their final work destination in London. Trips that do not interact with this sector were graded by the population density of devices' home districts, into 10 classes such that:

- districts around densely populated areas such as Manchester and Birmingham, with population density in excess of 4000 persons per km², are at the top of the scale, while
- districts related to areas such as the Scottish Borders, Scottish Highlands, and Northumberland, with population densities below 50 persons per km² are towards the lower end of the scale.

Population density grades were subsequently used to guide the minimum distance between origin and destination stations for commuting trips. The idea here is that it is expected that commuters entering a standard urban city via rail are likely originate from outside the boundary of the city, at distances beyond or equal to the outer suburbs, and, correspondingly, commuters will likely travel much further when originating from more remote places.

Directionality

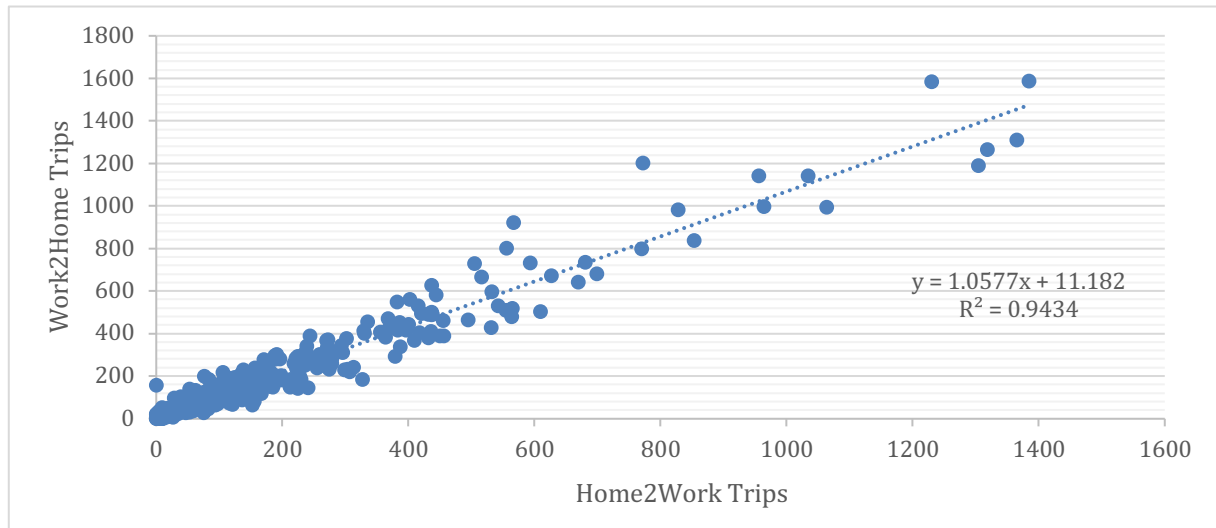
The direction of travel for all commuting trips identified based on the above methodology were flagged as follows:

- **From-Home** – these were determined to be commuting trips travelling from the Home LSOA/district to the Work LSOA/District
- **To-Home** – these were determined to be commuting trips travelling from the Work LSOA/district to the Home LSOA/District

Verifications

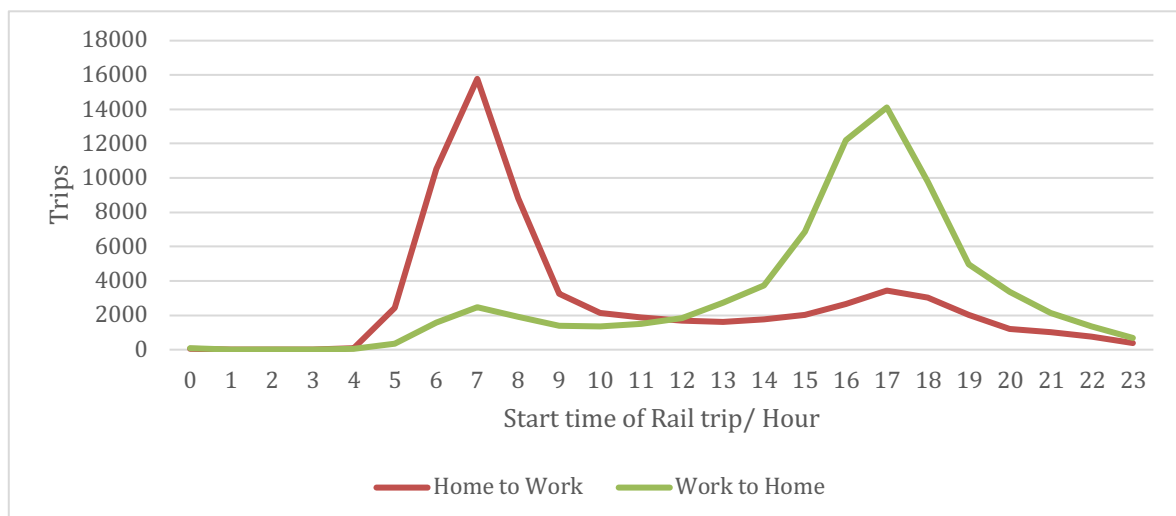
The identification of trips' directionality is useful for checking that data exhibit behaviour corresponding with that expected from commuting trips. For instance, it is expected that trips going to work from home and those going to home from work should be symmetrical. Figure 2 shows the symmetry between from-home and to-home trips as a linear regression plot. As can be seen, there is a good correlation between trips going in both directions, with a R-squared exceeding 0.9 and a slope close to 1.

Figure 2. Correlation between from-home and to-home commuting trips



While the symmetry in Figure 2 indicates that commuters generally return home from work, another expected characteristic is that the time distributions of from-home and to-home trips should show peaks in the AM and PM peak periods, when the highest volume of commuting trips are made. This comparison is shown in Figure 3. As can be seen, there are prominent peaks around 06:00 – 07:00 for from-home trips and 16:00 – 18:00 for to-home trips, indicating the typical times when rail commuters make work-bound trips from home and home-bound trips from work, respectively.

Figure 3. Trips plotted as a function of the start time of the rail trip for both from-home and to-home trips.



Data Field Format

Commuters trips report for APIs

The standard report was extended to create a commuters report with three fields to accommodate the total commuting trips as well as the direction of these commuting trips as follows:

- **TripDate** - date on which the trips were made
- **TimeSlotStart** - 15-minute timeslot at which trips departed from their origin station
- **TimeSlotEnd** 15-minute timeslot at which trips arrived at their destination station
- **Start_CRS** - station CRS code for trips starting at that station
- **End_CRS** - station CRS code for trips ending at that station
- **Trips** - total trips made between the start and end stations in the stated timeslot
- **CommutingTrips** - total commuting trips made between the start and end stations in the stated timeslot
- **CommutingFH** - total from home commuting trips made between the start and end stations in the stated timeslot
- **CommutingTH** - total to home commuting trips made between the start and end stations in the stated timeslot
- **AvgJourneyTime** – average time taken by trips between the start and end stations in the stated timeslot
- **Confidence** – the confidence level of the trip

Download Commuter Report – API Access

Returns a full daily station to station matrix broken down by commute/non commute grouped by 15-minute time slot for a single day.

Endpoint: <https://api.citianalytics.com/networkrail/od/downloadCommuters>

Method: GET

Input parameters: date(required), url (optional)

Headers: authToken (required)

Body: N/A

Commuter Data – Portal Query

The RDIS portal front end has been adjusted so that when a query is run on the portal, outputs are reporting commuting trips as well, both From Home and To Home as shown in Figure 4.

The information button has been updated to reflect all the fields that have been added to the report.

Figure 4. Portal query including commuters.

The screenshot shows the 'railanalytics' portal interface. At the top, there's a navigation bar with 'railanalytics' and 'OD Analyser'. A sidebar on the left contains 'Rail OD Analyser' and 'Report Downloads'. The main content area has a blue header with a system status message and a list of items: '1. National UK recovery in rail demand' and '2. Weekly trends and daily usage profile'. Below this is a search interface with filters for 'Reference Station' (Group: Select Group, Station: BASINGSTOKE), 'Search Type' (Whole Date Period, Time Period Only), and 'Date Range' (From: 10/10/2021 00:00, To: 16/10/2021 23:45). A 'New Search' button is present. Below the filters is a table with columns: From Station, From Station Code, To Station, To Station Code, Total Count, Confidence, Average Journey Time, Commuting Trips, Commuting From Home, and Commuting To Home. The table contains 10 rows of data for Basingstoke to various destinations. At the bottom left, there are logos for 'RDIS THE RAIL DEMAND INFORMATION SYSTEM' and 'citilogik', along with the text 'Citi Analytics v 2.102'.

The RDIS system is available with an on-going program of improvements and upgrades. Please see the System Status tab for the current information and known issues. Following Release 14, Network Rail intend to use the system for high level trend analysis and monitoring change in network usage in the following areas

1. National UK recovery in rail demand
2. Weekly trends and daily usage profile

This trending is carried out at a regional (Network Rail Regions) and National (UK). The information is being used by Analysis and Economic team to monitor trends against Lennon and PFM data. We are not using for decision making yet but putting the framework in place so that it can be used when data quality reaches maturity which is yet to be defined.

Network Rail have also integrated the live stations functionality into some NR's ACUMEN system. This information is being used by trial users to understand platform usage at Birmingham New Street station in conjunction with other data and systems.

Due to the current level of data being presented the accuracy is yet to be determined by NR independent testing and as such we are unable to recommend the full integration of this data.

Reference Station: Group: Select Group, Station: BASINGSTOKE

Search Type: Whole Date Period, Time Period Only

Date Range: From: 10/10/2021 00:00, To: 16/10/2021 23:45

Search: New Search

Engine version: R14.1

From Station	From Station Code	To Station	To Station Code	Total Count	Confidence	Average Journey Time	Commuting Trips	Commuting From Home	Commuting To Home
Basingstoke	BSK	London Waterloo	WAT	5010	1a	54.5	592	467	125
Basingstoke	BSK	Clapham Junction	CLJ	2944	1a	44.18	371	222	149
Basingstoke	BSK	London Waterloo	WAT	1715	1b	58.66	222	162	60
Basingstoke	BSK	Southampton Central	SOU	1108	1a	36.46	57	31	26
Basingstoke	BSK	Clapham Junction	CLJ	580	1b	45.39	93	61	33
Basingstoke	BSK	Bramley (Hants)	BMY	491	1b	6.67	36	13	22
Basingstoke	BSK	Bournemouth	BMH	442	1a	75.61	23	19	10
Basingstoke	BSK	London Waterloo	WAT	425	1c	54.12	82	69	13

RDIS THE RAIL DEMAND INFORMATION SYSTEM
powered by citilogik
making sense of chaos

Citi Analytics v 2.102