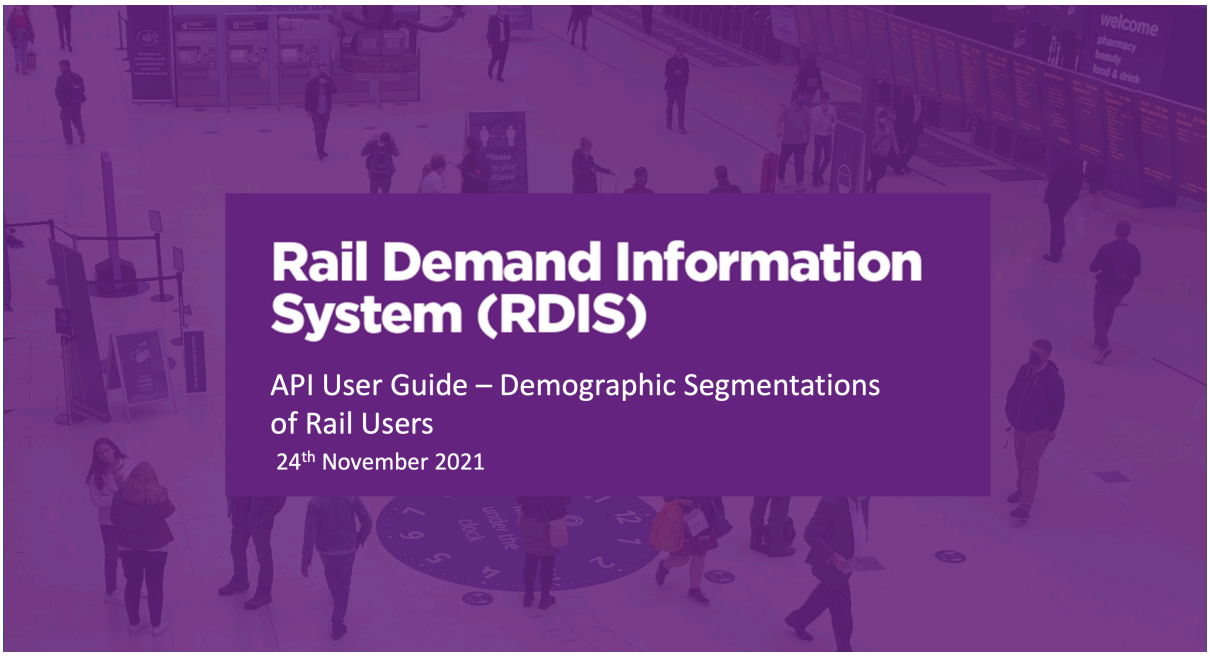
THE RAIL & STATION INNOVATION COMPANY



Rail Demand Information System (RDIS)

API User Guide – Demographic Segmentations of Rail Users

24th November 2021

Demographic Segmentations of Rail Travellers

Introduction

This document describes the derivation and presentation of the data set with demographic characteristics of rail travellers against specifications provided by RASIC. The process to derive traveller demographics is calculated weekly, typically on a Monday, with the resulting data giving demographic information for the preceding week. The data is presented in the file format described in the stated section that follows. The first section presents an overview of the methodology used to segment the data while the second presents the file structure and descriptions of headings.

Identifying Travellers' Demographic Segments

The demographic data provided by RASIC was obtained from P^2 (People & Places), Beacon Dodsworth,[1] which defines 15 categories and 44 sub-categories, termed Trees and Branches, respectively, at LSOA level. This enabled proportions within each category to be defined, which largely corresponded with the desired demographic segmentations. For instance, Table 1 below shows an example age segments and the corresponding proportions in each segment for LSOAs.

Table 1. Age segments in the P^2 data from a given LSOA.

Age	Population Proportion
Zero To Four	5%
Five To Fourteen	14%
Fifteen To Twenty-Four	11%
Twenty-Five To Forty-Four	20%
Forty-Five To Sixty-Four	32%
Sixty-Five To Seventy-Four	10%
Seventy-Five Plus	8%

Similar segments were derived for employment grade and Industry, and these were presented, separately for each origin/destination movement, in a given time-period, where the time-period segments were defined as:

- Morning – 02:00 to 07:00
- AM Peak – 07:00 to 10:00
- Inter Peak – 10:00 to 16:00
- PM Peak – 16:00 to 19:00
- Evening – 19:00 to 02:00

The proportions in Table 1 are combined with the number of travellers (based on devices' inferred home locations) resident in each LSOA to estimate trip making by demographic segments. This is done by identifying the home locations of devices making trips. The home locations will belong to a LSOA with one or several of the 44 branches. If it belongs to a LSOA with more than one branch, then the trips would be split into the number of relevant branches using the split branch proportions worked out previously, renormalised across relevant branches.

[1] https://beacon-dodsworth.co.uk/landing-pages/p2-people-and-places/

Each device 'produces' a number of trips expressed by the device expansion factor. The number of trips is broken down, if need be, into the different branches that make up the home location LSOA of the device. Once a number of trips have a unique branch, by multiplying these trips by the demographic segment's probability, the demographic characteristics of the trips can be defined.

Including NTS with Bayes' theorem

A refinement to the methodology is possible by incorporating NTS (DfT's National Travel Survey[2]) data into the process. It is known that people with higher income tend to make more trips per day, especially in the case of rail trips. People between 30 and 39 years old make considerably more trips than people over 70. The NTS data could be leveraged to update the trips probabilities using Bayes' theorem.

$$P(P_i|T) = \frac{P(T|P_i) * P(P_i)}{P(T)}$$

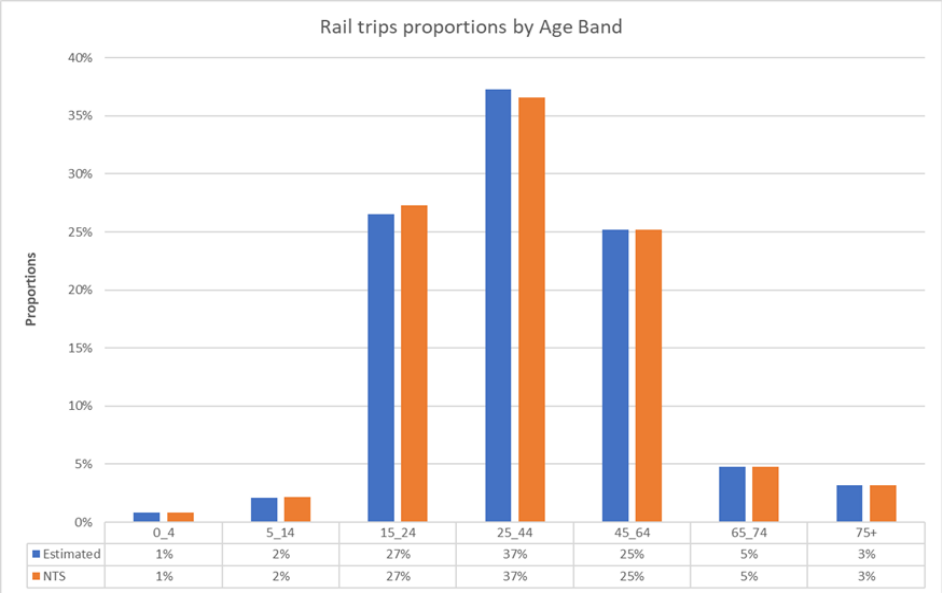
where T represent the event of observing a rail trip, Pi represents the event of belonging to the population segment x are events and $P(T) \neq 0$

- $P(Pi \mid T)$ represents the conditional probability of belonging to the population segment Pi knowing that a rail trip was observed.
- $P(T \mid Pi)$ represents the conditional probability of observing a trip knowing that the individual belongs to the population segment Pi;
- $P(Pi)$ represents the probability of belonging to the population segment Pi; and
- $P(T)$ represent the probability of observing a rail trip.

$P(T \mid Pi)$ is essentially the demographic rail trip rate for each of the segments. $P(Pi)$ is the population proportion provided RASIC and P(T) is the $\sum[P(T|P_i) * P(P_i)]$.

Figure 1 shows rail trip distributions by age segments. It can be observed that the resulting rail trip distribution compares very well with the converted NTS proportions presented above.

Figure 1. Comparison of adjusted rail trips with NTS.



[2] https://www.gov.uk/government/collections/national-travel-survey-statistics

Data Field Format

Full OD Trip rate

The dataset shall contain the following fields:

- **Orig_Station**
  Station CRS code for trips starting at that station.
- **Dest_Station**
  Station CRS code for trips Ending at that station.
- **TimePeriod**
  The time period in which the trips were made. Integer based. Key:

  1 - Morning 0200-0700
  2 - AM Peak 0700-1000
  3 - Inter Peak 1000-1600
  4 - PM Peak 1600-1900
  5 - Evening 1900-0200
- **Trips**
  Number of people who travelled between the stated origin and destination stations.
- **Age_0_4**
  Number of people in the 0-4 age band making trips between the stated origin and destination stations.
- **Age_5_14**
  Number of people in the 5-14 age band making trips between the stated origin and destination stations.
- **Age_15_24**
  Number of people in the 15-24 age band making trips between the stated origin and destination stations.
- **Age_25_44**
  Number of people in the 25-44 age band making trips between the stated origin and destination stations.
- **Age_45_64**
  Number of people in the 45-64 age band making trips between the stated origin and destination stations.
- **Age_65_74**
  Number of people in the 65-74 age band making trips between the stated origin and destination stations.
- **Age_75Plus**
  Number of people in the 75+ age band making trips between the stated origin and destination stations.
- **UnEmployed**
  Number of unemployed people making trips between the stated origin and destination stations.
- **Professional**
  Number of people in the professional employment segment making trips between the stated origin and destination stations.

- **OtherWhiteCollar**
  Number of people in the “other white collar” employment segment making trips between the stated origin and destination stations.
- **SkilledManual**
  Number of people in the “skilled manual” employment segment making trips between the stated origin and destination stations.
- **UnskilledManual**
  Number of people in the “unskilled” employment segment making trips between the stated origin and destination stations.
- **Industry_Agriculture**
  Number of people in working in agricultural industries making trips between the stated origin and destination stations.
- **Industry_ KnowledgeIntensive**
  Number of people working in the knowledge intensive industries segment making trips between the stated origin and destination stations.
- **Industry_ Manufacturing**
  Number of people who work in manufacturing industries making trips between the stated origin and destination stations.
- **Industry_ Service**
  Number of people who work in service industries making trips between the stated origin and destination stations.
- **Industry_ Tourism**
  Number of people who work in tourism industries making trips between the stated origin and destination stations.
- **Industry_ Other**
  Number of people who work in other industries making trips between the stated origin and destination stations.

Download Demographics Report – API Access

Returns a full 1 week sum time period based station to station demographic report. Date to be based on Monday.

Endpoint: https://api.citianalytics.com/networkrail/od/downloadWeeklyDemographic
Method: GET
Input parameters: date(required), url (optional)
Headers: authToken (required)
Body: N/A