

FUSION PROJECT: SUMMARY REPORT

Delivered by SYSTRA for
Department for Transport

SYSTRA



Department
for Transport

Publication Date:
30/10/2025

Author:
Caroline Broughton-Smith

Reviewer:
Eifion Jenkins

Client:
Department for Transport

Photograph © GettyImages-frankreporter 2016



WWW.SYSTRA.COM/UK/

EXECUTIVE SUMMARY

The Department for Transport (DfT) commissioned SYSTRA alongside project partners TravelAi and Yonder Data Solutions (YDS) to pilot a novel approach of generating travel insights by combining smartphone sensor-derived data ('big' data) and online surveys ('deep' data). The Fusion of User Surveys and Integrated Online Navigation (FUSION) project aimed to assess if this integrated method can enhance understanding of how different demographic groups respond to disruption on the transport network, or changes in policy.

YDS recruited 3,500 nationally and regionally representative participants across England through a recruitment survey, allowing each participant to be classified into their transport user segment. Five online surveys were then delivered to capture information about travel movements, attitudes towards transport, and life circumstances.

Concurrently, participants installed TravelAi's MyWays app, which automatically generated end-to-end multimodal travel diaries on their smartphones throughout the project.

Data collected from the surveys and MyWays app were combined to create 3 aggregated Open Datasets on travel movements and behaviours, categorised by transport user segments. These datasets and supporting documentation will soon be publicly available via the Geographic Data Service. A Proof-of-Concept Analysis (PoC) Tool, also publicly available, was developed to showcase data visualisation, facilitate insights, and bring the data to life.

The Open Datasets and PoC Tool were evaluated for their ability to enable enhanced demographic analysis; support strategic modelling; and provide deeper behavioural insights to inform targeted policy development. Comparisons were drawn with the National Travel Survey (NTS), and analyses of responses to a policy shift and a disruptive weather event. The potential to generate synthetic populations and the risk of bias in the method were also assessed.

Findings indicated under- and over-representation of some travel modes, trip lengths, and demographic groups within the participant panel. The Open Datasets and PoC Tool proved effective for comparing journey patterns and metrics before, during and after events of interest, and insights were enhanced by aggregated survey responses, but analyses at granular spatial and temporal levels can be limited by sample size. In future, exploring more advanced anonymisation techniques could further enhance the value of the data captured, whilst safeguarding participant privacy.



GLOSSARY

CATI	Computer-Aided Telephone Interviews
DfT	Department for Transport
EDIF	Economic Data Innovation Fund
EEH	England's Economic Heartland
FUSION	Fusion of User Surveys and Integrated Online Navigation
GNSS	Global Navigation Satellite System
INTS	Integrated National Transport Strategy
LA(D)	Local Authority (District)
LSOA	Lower Super Output Area
MSOA	Middle Super Output Area
ONS	Office for National Statistics
NS-SEC	National Statistics Socio-economic Classification
NTEM	National Trip End Model
NTS	National Travel Survey
PoC	Proof-of-Concept
QA	Quality Assurance
TfGM	Transport for Greater Manchester
TfL	Transport for London
TfWM	Transport for West Midlands
YDS	Yonder Data Solutions

CONTENTS

1	Introduction	4
2	Methodology	5
2.1	Overview	5
2.2	Data Collection	5
2.3	Open Datasets Creation	15
2.4	PoC Analysis Tool	17
3	Findings	19
3.1	Recruitment and Surveys	19
3.2	MyWays Data Collection	23
3.3	Quality Assurance	25
3.4	Potential Applications	26
4	Lessons Learned	30
5	Summary and Conclusions	31

1 INTRODUCTION

Department for Transport (DfT) commissioned SYSTRA to explore a novel approach of generating travel insights through the combination of smartphone-sensor derived data and online surveys in a project called Fusion of User Surveys and Integrated Online Navigation (FUSION).

FUSION, co-funded by HM Treasury's Economic Data Innovation Fund (EDIF) and DfT, was motivated by the potential to use smartphone sensor-derived data to capture user travel patterns with greater spatial and temporal detail than traditional approaches such as surveys. These 'big' data reveal when, how, and where people travel; when combined with 'deep' data, which is captured through online surveys and reveals why travel behaviours are exhibited, a richer understanding of travel patterns emerges. FUSION piloted this method of generating information to determine if it can:

- Provide an improved understanding of the transport behaviour of different demographic groups for strategic modelling and inclusive policy-making.
- Allow examination of the impact of disruptive events and moments of change - geographically, temporally, and demographically.
- Generate data which will allow questions to be answered such as:

How do different segments respond to policy changes or disruptive events?

How are the transport needs of segments reflected in their behaviours?

How do segment behaviours change through time?

Is such data compatible with the National Travel Survey?

Can the data be used to generate a synthetic population?

2 METHODOLOGY

2.1 Overview



- Approximately 3,500 participants were recruited by YDS to:
 - Download and enable the MyWays app (developed by TravelAi) for 6 months. MyWays automatically generates end-to-end multimodal travel diaries using the smartphone's sensors and transport mode detection software.
 - Complete 6 online surveys (including a recruitment survey) about travel movements, attitudes to transport, and life circumstances.
- The data collected from both sources were aggregated and combined to create 3 rich, longitudinal Open Datasets, classified into [DfT's transport user segments](#).
- A PoC Analysis Tool was developed to demonstrate how the Open Datasets can be visualised and analysed, bringing the data to life and making it accessible to DfT as well as the wider transport sector.
- The Open Datasets and PoC Analysis Tool were examined to understand how they can be utilised to support the aims of the project (as stated on the previous page) and can be scaled for future applications.

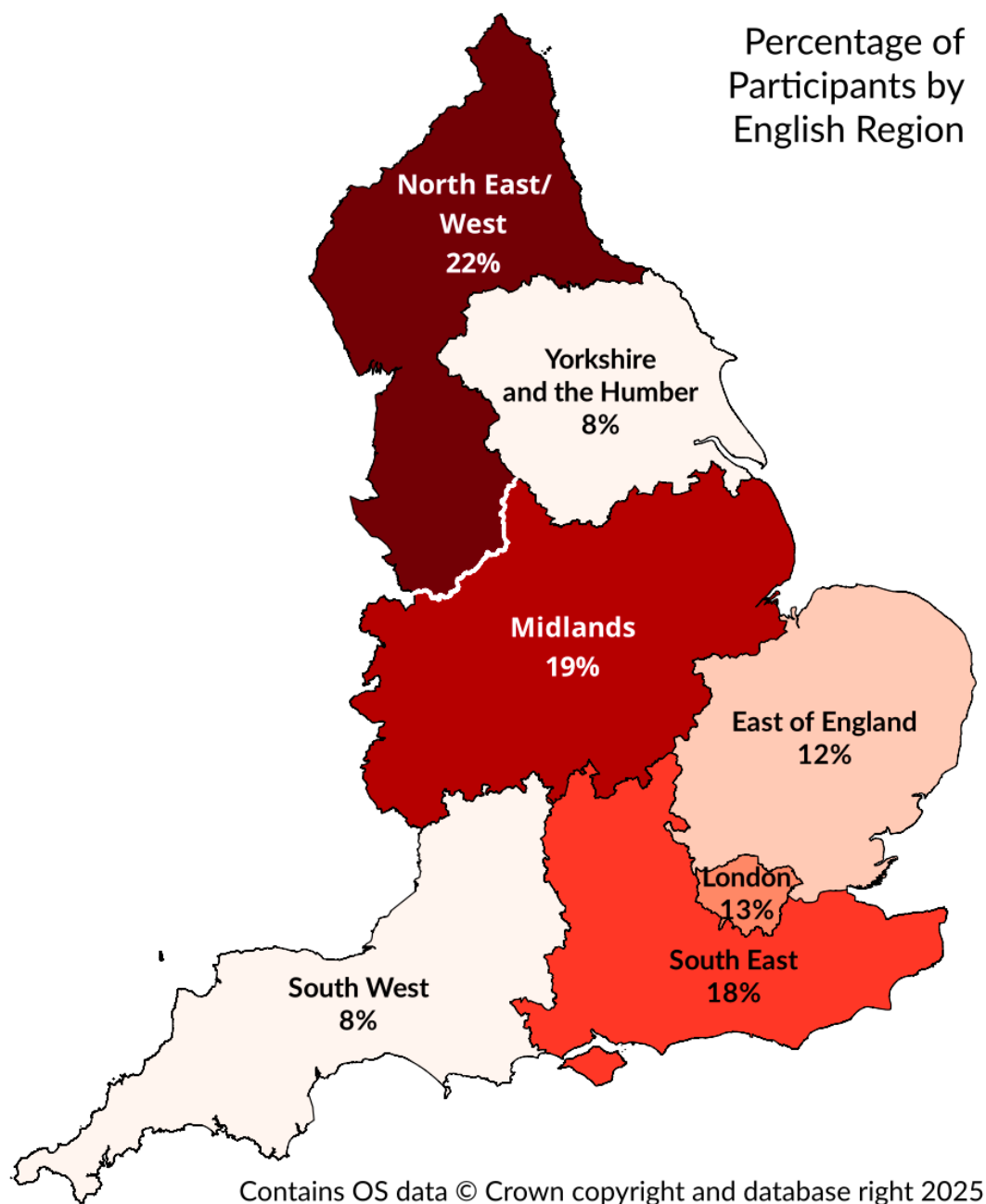
2.2 Data Collection

Data collection consisted of 3 elements:

1. Recruitment:

- A nationally and regionally representative sample of 3,500 participants were recruited by YDS from their proprietary panel across England.
 - Quotas were set based on age, gender, car ownership, and urban/rural classification using Census 2021 data.
 - Participants were paid £5 per survey.
- The recruitment survey contained the [golden questions](#) to assign participants to their [transport user segment](#), alongside other demographic questions to achieve demographic representation.
- Top-up recruitment was carried out to maintain a representative sample.

Figure 1 Percentage of participants by English Region



2. Online Surveys:

5 further surveys were delivered to participants by YDS over 5 months:

- 1 Travel diary (most recent weekday)
- 2 Travel diary (most recent weekend day)
- 3 Time use survey
- 4 Policy evaluation
- 5 Multimodal transport use

3. MyWays App:

Participants enabled the MyWays app on their mobile phones for 6 months to automatically capture end-to-end multimodal journeys.

Surveys 1 and 2: Travel Diaries

- Participants recorded information about each trip made on the previous day, including start and end time, main purpose, main mode, and answered follow up questions relating to the reasoning behind their mode choices.
- They were also asked about moments of change, and any recent disruptive events which may have caused a change in travel behaviours.
- Survey 1 captured the most recent day of weekday travel and concluded with additional questions about working from home (not specific to the weekday travel captured).
- Survey 2 captured the most recent day of weekend travel and concluded with questions about EV, e-bike and e-scooter use and ownership (not specific to the weekend travel captured).

Table 2 Examples of questions within surveys 1 and 2

T3 For your first trip, please provide the following information	
1	Your approximate trip start time: (Select using 24 hour clock)
2	Your approximate trip end time: (Select using 24 hour clock)
3	The main purpose of your trip: Single select dropdown options as follows: 1. Commuting to/from a place of work 2. Commuting to/from a place of education 3. Escorting to/from education (e.g. travelling so someone else can access education, such as giving a lift to/from school) 4. Escorting to/from another activity (e.g. travelling so someone else can access another activity, such as giving a lift to/from the train station) 5. Business travel (excluding commuting to your usual place of work) 6. Shopping (e.g. for food/grocery item, books, music, clothes, etc.) 7. Visiting friends/relatives 8. To go to/from a holiday or a day trip for leisure 9. Accessing entertainment/hospitality (e.g. cinema, theatre, gallery, museum, sporting events, music events, concerts, pubs, bars or restaurants) 10. Personal errands (e.g. medical, hospital, dentist or hairdressers appointments, visiting a place of worship, running errands, accessing services such as estate agents and banks) 11. For recreation/keeping fit (e.g. going to a place to walk/cycle, to a gym/playing sport) 12. Other (Please specify)

T4d What were your reasons for using a car for this trip? (Multi-select)	
1	I enjoy driving
2	I needed to carry bulky/heavy items (e.g. shopping)
3	Time constraints (e.g. I had other errands to run, school drop-offs)
4	Unpredictable weather
5	It was the most reliable way to get to my destination
6	It was the fastest way to get to my destination
7	It was the cheapest way to get to my destination
8	It gives me flexibility in the time I travel
9	My car is more comfortable than travelling by other modes
10	I feel safer driving compared to travelling by other modes
11	I was travelling with others
12	Other (Please specify)

T5a Did you have a car available for this trip? (Single select)	
1	Yes
2	No

T5b What were your reasons for not using a car for this trip? (Multi-select)	
1	Cost of petrol
2	Lack of parking facilities
3	Cost of parking
4	Traffic congestion/roadworks
5	Other forms of transport were available
6	Other general motoring costs
7	Environmental reasons
8	Safety concerns/ Nervous about driving
9	Health / well-being reasons
10	I enjoy using other modes instead
11	Other (Please specify)

Survey 3: Time-use Survey

- Participants recorded the purpose, time and duration of primary and secondary activities conducted throughout the previous day. The activity categories followed the Office for National Statistics' (ONS) time use survey categories.
- Follow up questions captured who the activities were carried out with, where they took place, and where relevant, how activities impacted their transport mode choice.
- It also asked about moments of change, and any recent disruptive events which may have caused a change in travel behaviours.

Figure 3 Activity categories as defined by ONS

Time Use Study - ACTIVITIES LIST

Sleeping, personal care, medication

1. **Sleeping**
2. **Washing, dressing, using the bathroom and self-grooming**
Includes getting ready, other personal hygiene (e.g. teeth cleaning), hair care, painting nails, applying make-up, shaving etc.
3. **Medication, other health related care (e.g. blood pressure, injections)**
Includes sorting medication, self- testing (e.g. blood sugar, lateral flow), injections (e.g. insulin), tending to injuries (e.g. applying plasters, dressings)

Eating, drinking, cooking

4. **Making food and drinks, cooking**
Including for other people if unpaid
5. **Eating a meal, eating out, take-away (e.g. breakfast, lunch, dinner)**
6. **Snacking**
7. **Drinking (e.g. tea, coffee, alcohol)**

Work and other activities for pay

8. **Working**
Exclude lunch and other breaks. Exclude other work-related activities below
9. **Working from home**
10. **Working from a cafe or other workspace**
11. **Providing childcare, cleaning, or doing odd jobs for pay**
Exclude if done for main job or business (use codes 8, 9, 10) or delivery services (use code 13)
12. **Leasing or renting things you own (e.g. a spare room, storage space, power tools, furniture, or clothes)**
Exclude if done for main job or business (use codes 8, 9, 10)
13. **Using your private vehicle to earn money (e.g. delivery services)**
Exclude commuting and vehicles registered for business use
14. **Showing your own house, flat, or building to potential buyers**
15. **Selling your things, apart from your home (e.g. eBay, FaceBook Marketplace, Vinted)**

Travel and getting around

16. **Travelling to or from paid work (e.g. commuting, to attend a conference, to visit clients)**
17. **Travelling to or from unpaid work (e.g. caring for others, volunteering)**

18. **Travelling to or from a shop (e.g. supermarket, garden centres, takeaways)**
19. **Travelling to or from socialising with others outside a private home (e.g. restaurant, pub, park)**
20. **Traveling to escort children to or from childcare or school**
Exclude college or university
21. **Travelling to or from another place (e.g. holiday, visiting family or friends, escorting others, medical centre, dentist, gym, appointments or errands)**
Exclude travel done for exercise purposes (e.g. use code 63 for running or jogging, 64 for cycling or 67 for walking)
22. **Packing or unpacking, preparing for journey**

Housework, pets, DIY and gardening

(includes for other people if unpaid)

23. **Cleaning, hoovering, tidying house, sorting the bins (e.g. recycling)**
24. **Using dishwasher or washing up**
25. **Arranging, sorting or unpacking household items (e.g. unpacking shopping, organising and clearing out storage or rooms)**
26. **Ironing, washing, other laundry tasks or mending clothes**
27. **Repairing, maintaining or making household goods, or vehicles**
Include car cleaning or repair, assembling furniture, or similar
28. **Lighting fire or cleaning fireplace, log burner or wood burning stove**
29. **Feeding, caring for or playing with animals (pets)**
Exclude livestock
30. **Walking the dog**
31. **DIY**
32. **Gardening**

Volunteering

33. **Volunteering as part of a group, organisation, charity, or sports club**
Include helping caring for other people if done for a voluntary organisation or a charity

Caring for and looking after children and adults

(Only if unpaid. If done for work or pay use codes 8, 9, 10, 11)

34. **Feeding, washing, dressing or preparing meals for children**

- 35. Reading with children, helping with homework, doing other educational activities with children**
Include home-schooling
- 36. Playing with children**
Include indoor and outdoor play
- 37. Attending or watching a child's event or activity**
Include sports activities, music lessons, parents' evening or baby groups
- 38. Supporting, comforting or cuddling children**
- 39. Other child care not elsewhere listed**
Include helping neighbours or friends if unpaid
- 40. Helping, caring for and looking after adults (aged 18+)**
Include caring for spouse or partner, family, neighbours or friends

Shopping, household administration tasks and appointments

- 41. Buying something, shopping**
- 42. Browsing things to buy later, or window shopping**
Include viewing a property to buy as well as browsing for other items
- 43. Household administration tasks (e.g. banking, sorting out bills)**
- 44. Attending appointments or errands (e.g. doctor, vet, bank, hospital, haircut, beautician, garage, etc).**
- 45. Queueing or waiting**
- 46. Completing a document (e.g. job or university application, passport or benefit form or similar)**
Free time, entertainment and socialising, including online
- 47. Watching TV and DVDs**
Include streaming (e.g. Sky, Netflix, Amazon Prime, Disney+, YouTube etc).
- 48. Checking phone or tablet**
Include combination of checking emails, messages, news, apps
- 49. Listening to music, podcasts, audiobooks, talk shows, radio or news**
- 50. Playing games or computer gaming**
- 51. Checking or using social media**
- 52. Browsing internet**
Exclude browsing for activities covered by other categories
- 53. Checking email**
- 54. Reading books, magazines or newspapers**
Exclude browsing internet
- 55. Socialising, spending time with friends, family, neighbours and colleagues**
Include unplanned contact in person

- 56. Having a conversation**
Include telephoning and video calling
- 57. Writing, texting or emailing**
Include writing letters, personal diaries, journalling
- 58. Visits to cinema, theatre, concerts, sporting events, museums, galleries, library etc.**
Include virtual reality (VR) and online visits
- 59. Attending a meeting or an event**
Include a place of worship, support groups or community affairs
- 60. Hobbies and other leisure activities**
- 61. Resting (doing nothing) or in bed not asleep**
Include daydreaming, lying awake or being ill in bed, intimacy

Exercise, health and being active

- 62. Gym, fitness, or exercise classes**
- 63. Running or jogging**
- 64. Cycling**
- 65. Playing team sports**
- 66. Playing other sports and exercising**
Include dancing, swimming or other activities if done for exercise
- 67. Going for a walk as exercise**
- 68. Meditating, having a massage, spa or well-being treatments**
- 69. Other health a, well-being activity**
Education and study
- 70. Attending formal education or taking a course**
- 71. Learning or teaching yourself a skill not involving taught classes**
Include cooking, coding, photography, DIY or similar
- 72. Studying, revising or doing homework**
Include group study and researching for self improvement purposes

Other computer use

- 73. Other computer or laptop use (e.g. creating or coding a website, writing online or creating content for public, assisting others online)**

Other or personal

- 74. Other activities not listed (If private time then please write 'personal')**
- 75. Praying**
- 76. Smoking or vaping**
- 77. Completing the time-use diary**
Include reading instructions and reminders, jotting down notes

Survey 4: Policy Evaluation

- Survey 4 included questions about participants' behaviour in response to a policy implemented during the study period.
- Follow up questions related to ride-sharing app usage.
- It also asked about moments of change, and any recent disruptive events which may have caused a change in travel behaviours.

Table 4 Examples of questions within survey 4

Q5	Do you own any of the following travel passes that provide you with free bus travel? Please select all that apply. (Multi-select)
1	Freedom pass
2	Disabled person's bus pass
3	Older person's bus pass
4	Young person's bus pass (e.g. Our Pass in Greater Manchester)
5	A bus pass provided by my school/college/university
6	Armed forces bus pass
7	Carer's bus pass
8	Employee card for free bus travel
9	Another type of free bus pass (Please specify)
10	I do not own a travel pass that provides me with free bus travel

A1	In the past month, have you ... (Multi-select)
1	Moved home
2	Started work for the first time
3	Changed your place of work/education
4	Had a child (by adoption, fostering or birth)
5	Had a child begin to travel to school by themselves
6	Had a child go to a new school
7	Had an adult child move out of your home
8	Began co-habiting with someone different
9	Acquired a disability, or temporary or sudden long-term illness that affects your ability to drive, walk or use other forms of transport
10	Gained a full driving license
11	Given up your driving license or had it revoked/suspended
12	Reduced the number of cars, vans or motorbikes that are available for you to use
13	Had a new e-bike, e-scooter or cycle rental hub open near your home or place of work/education
14	Had a new bus stop, bus station, or railway station open near your home or place of work/education
15	Been given a Blue Badge parking permit
16	None of these (Exclusive)
17	Other (FREE TEXT)
18	Prefer not to say

Survey 5: Multimodal Transport Use

- Survey 5 contained questions about recent weekday multimodal and/or multipurpose journeys, including journey start and end time, number of stops/changes, journey purpose(s) and mode(s) used. Follow up questions related to reasoning behind mode choice, cost of travel, and journey planning using apps and timetables.
- Follow up questions related to ride-sharing app usage.
- It also asked about moments of change, and any recent disruptive events which may have caused a change in travel behaviours.
- Survey questionnaires are provided alongside the Open Datasets.

Figure 5 Introduction to survey 5

Journeys you have recently undertaken

A **journey** is a one-way course of travel, from a starting point to a final destination, such as travelling home from a place of work. A **journey** can include multiple **legs** (stages), **purposes** (reasons), and/or **modes** (types) of transport.

As an example, you might get the bus home from work, but visit a supermarket on the way. This is:

- 1 Journey:** Travelling home from a place of work.
2 Legs: 1) Place of work → Supermarket; and 2) Supermarket → Home.
2 Purposes: 1) Commuting to/from a place of work; and 2) Shopping.
2 Modes: 1) Walking; and 2) Bus.

You might use different **modes** without realising it. For example, walking to a bus stop is one mode, and the bus ride itself is another.

For this survey, please think about your **most recent journey** on a weekday (i.e. Monday to Friday), which involved:

Travelling for multiple purposes in the same journey. For example:

- Dropping-off children at school on the way to work.
- Going to the supermarket on the way home from the gym.
- Visiting a relative before heading into town for shopping.

AND/OR

Combining multiple modes of travel in the same overall journey. For example:

- Walking to a bus stop, then boarding the bus to your destination.
- Cycling to a train station, then travelling by rail to your destination.
- Catching a taxi to a tram interchange, then boarding the tram to your destination.

MyWays App

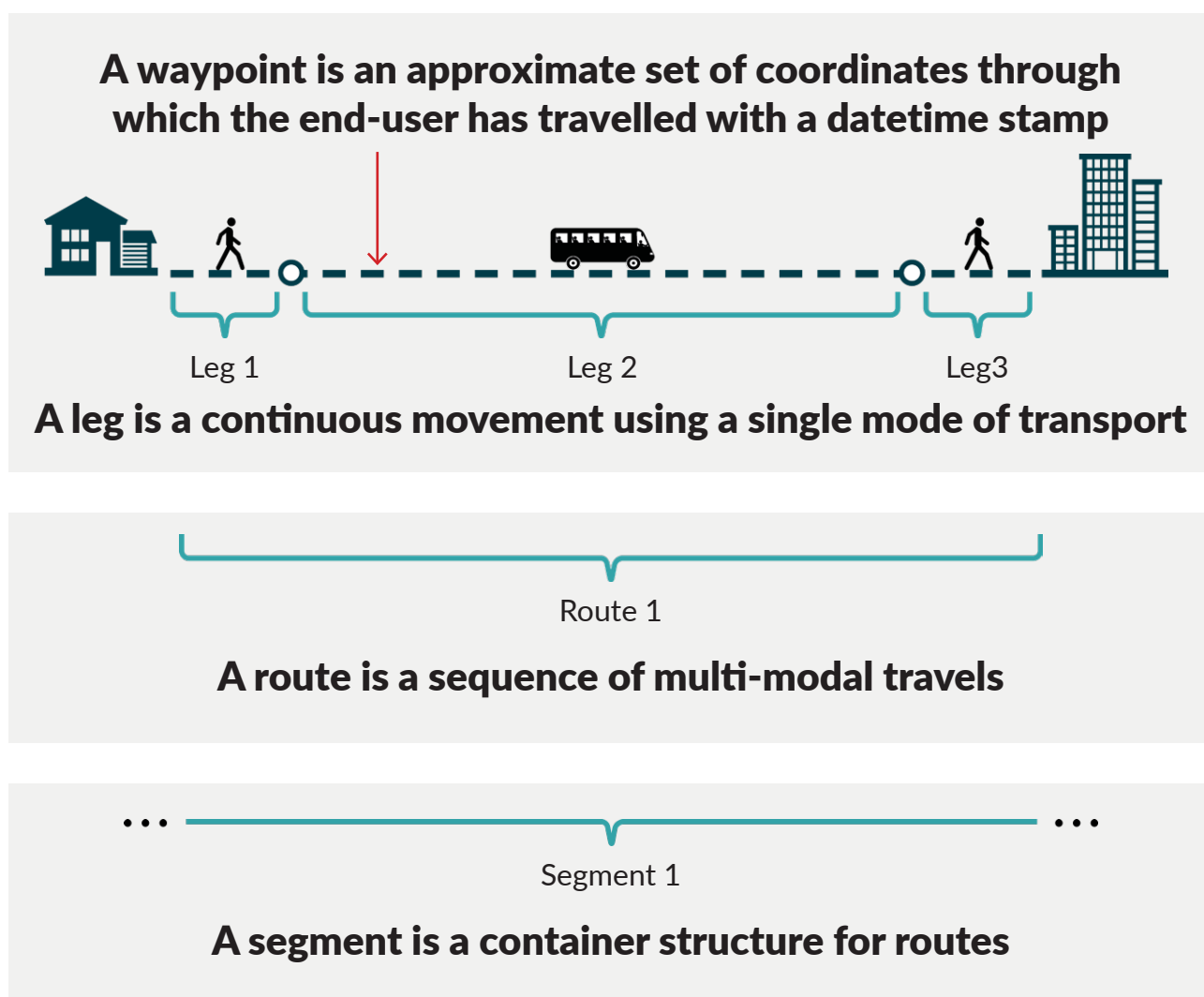
The MyWays app, developed by TravelAi, automatically collects digital multimodal travel diary information using smartphone sensors, without the need for user input.

- The app can identify 10 modes of transport: air, boat, bicycle, bus, car, metro, run, tram, train and walk.
- Users can edit and verify the data captured. They were encouraged to do so when recruited for the project.

Participants were asked to download and enable the app on their mobile phones, and to leave it running for the duration of the study (6 months). YDS and TravelAi worked together to assign participants an anonymous unique identifier which could later be used to link up MyWays data records to the online survey responses.

The data that MyWays captures is compiled into a .json file in a hierarchical nature as shown in the diagram.

Figure 6 Illustration of the hierarchical nature of MyWays data



Quality Assurance (QA) was a key element of data collection, and was carried out on the individual and combined datasets.

YDS Survey Data

- The content and routing of each survey was reviewed and tested by YDS, SYSTRA and DfT.
- YDS internal quality control measures were built into scripts to ensure clean data collection and thoughtful, genuine responses were captured.
- Surveys were soft launched with a few hundred participants for testing before full launch.
- Post-survey, YDS provided data summaries to allow SYSTRA to carry out high level data checks.
- Deeper analysis was carried out on the raw data received, including comparison with existing datasets such as the NTS.

MyWays App Data

- Internal testing was carried out and findings reported to TravelAi to facilitate software tweaks ahead of launch.
- Participants were encouraged to verify and correct trips captured by MyWays.
- A bespoke Python script was developed by SYSTRA to analyse incoming data, allowing queries to be raised and resolved.
- The capabilities and limitations of MyWays app and the data it generates were assessed.
- A comparison was carried out against existing datasets such as the NTS.

Open Datasets

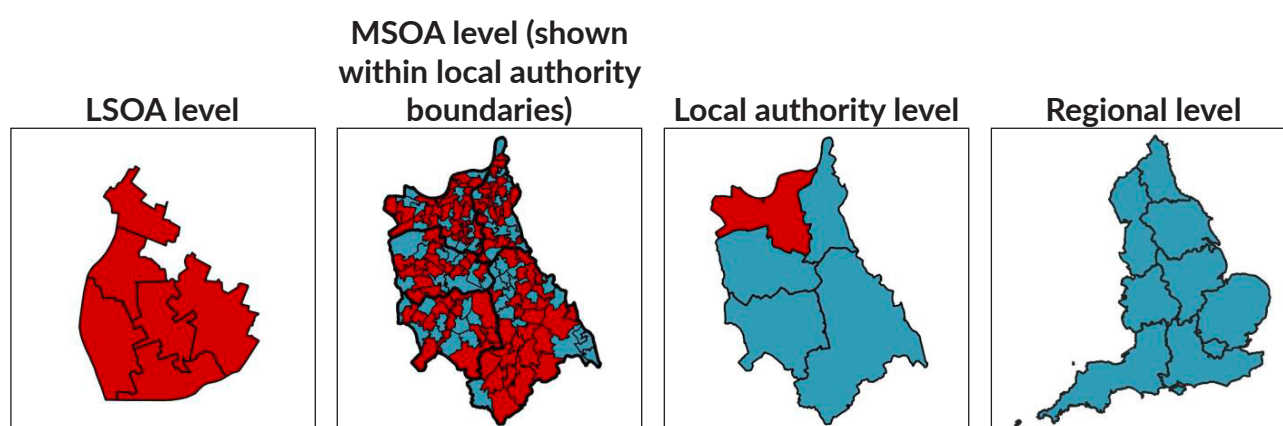
- Checks of sample size were carried out to ensure minimum participant privacy threshold was not breached.
- The Open Dataset tables retain a record of any data estimation or loss which occurred during data processing.

2.3 Open Datasets Creation

Data collected through the online surveys and the MyWays app were processed, aggregated and combined to create new Open Datasets.

Participants' MyWays data containing segment, route and leg information, with obfuscated (i.e. masked) start and end points at Lower Super Output Area (LSOA) level, was delivered from TravelAi to SYSTRA. This obfuscated MyWays data formed the core of the Open Datasets.

Data aggregation was necessary to preserve participant privacy and to create a manageable dataset for ingestion by the PoC Analysis Tool. The method of aggregation was dependent on the size of the geographic area and related sample size. Smaller administrative boundaries increase the risk of breaching participant privacy, making it necessary to aggregate to larger geographies to ensure confidentiality, as shown below.



KEY ● Breaches privacy ● Maintains privacy

The key attributes contained in the Open Datasets are: origin, destination (both at Middle Super Output Area (MSOA), local authority and region geographies), transport user segment, mode, time period (e.g. AM peak, defined as 7am to 10am), date, day of week, and place of residence. Combining all of these attributes in one data table would produce data as granular as the raw data; a set of aggregated tables generated from different combinations of these attributes were therefore generated.

Following best practices applied when working with Census data, a minimum sample threshold of 10 records was established. Data aggregation results which do not comply with this were hidden or estimated dependent on the geographical level and sample size.

To complete the Open Datasets, the data collected through the recruitment survey and 5 follow up surveys were aggregated in a similar manner and appended to the core dataset.

Table 7 An example cut of route-level data demonstrating the aggregation process

Origin MSOA	Origin LAD	Origin Region	User residence region	Transport user segment	Mode	Route Count	n MSOA Aggregated	mean LA Aggregated	n LA Aggregated	Mean Region Aggregated	UK proportions by mode	GDPR Compliance Status	Count Published
Bury 025	Bury	North West	North West	Elderly without cars	Bus	2	1	2	3	1.33	-	Estimated - region	1.33
Liverpool 026	Liverpool	North West	North West	Elderly without cars	Bus	146	-	-	-	-	-	Compliant	146
Liverpool 028	Liverpool	North West	North West	Elderly without cars	Bus	3	3	2	-	-	-	Estimated - LA	2
Liverpool 045	Liverpool	North West	North West	Elderly without cars	Bus	2	3	2	-	-	-	Estimated - LA	2
Liverpool 035	Liverpool	North West	North West	Elderly without cars	Bus	1	3	2	-	-	-	Estimated - LA	2
Oldham 012	Oldham	North West	North West	Elderly without cars	Bus	1	1	1	3	1.33	-	Estimated - region	1.33
Salford 034	Salford	North West	North West	Elderly without cars	Bus	3	2	2	-	-	-	Estimated - LA	2
Salford 036	Salford	North West	North West	Elderly without cars	Bus	1	2	2	-	-	-	Estimated - LA	2
Setton 037	Setton	North West	North West	Elderly without cars	Bus	1	1	1	3	1.33	-	Estimated - region	1.33

Aggregation Required

Compliant

A workshop with potential future users of the Open Datasets was held to establish the most useful way to aggregate and fuse the MyWays and survey data collected. The feedback from this workshop led the project team to create 3 versions of the Open Dataset:

1. MyWays data and survey data combined and aggregated to metropolitan counties/region and transport user segment. This is the version which feeds into the PoC Analysis Tool (note that all versions will soon be available for download via the Geographic Data Service).
2. MyWays data and survey data combined and aggregated nationally and by place typology based on ONS Rural/Urban Classification 2021:

1. Larger rural: further from a major town or city	4. Smaller rural: nearer to a major town or city
2. Larger rural: nearer to a major town or city	5. Urban: further from a major town or city
3. Smaller rural: further from a major town or city	6. Urban: nearer to a major town or city
3. MyWays data and survey data combined and aggregated nationally and by age group as follows:

1. 16-24	3. 30-39	5. 50-59	7. 70-79
2. 25-29	4. 40-49	6. 60-69	8. 80+

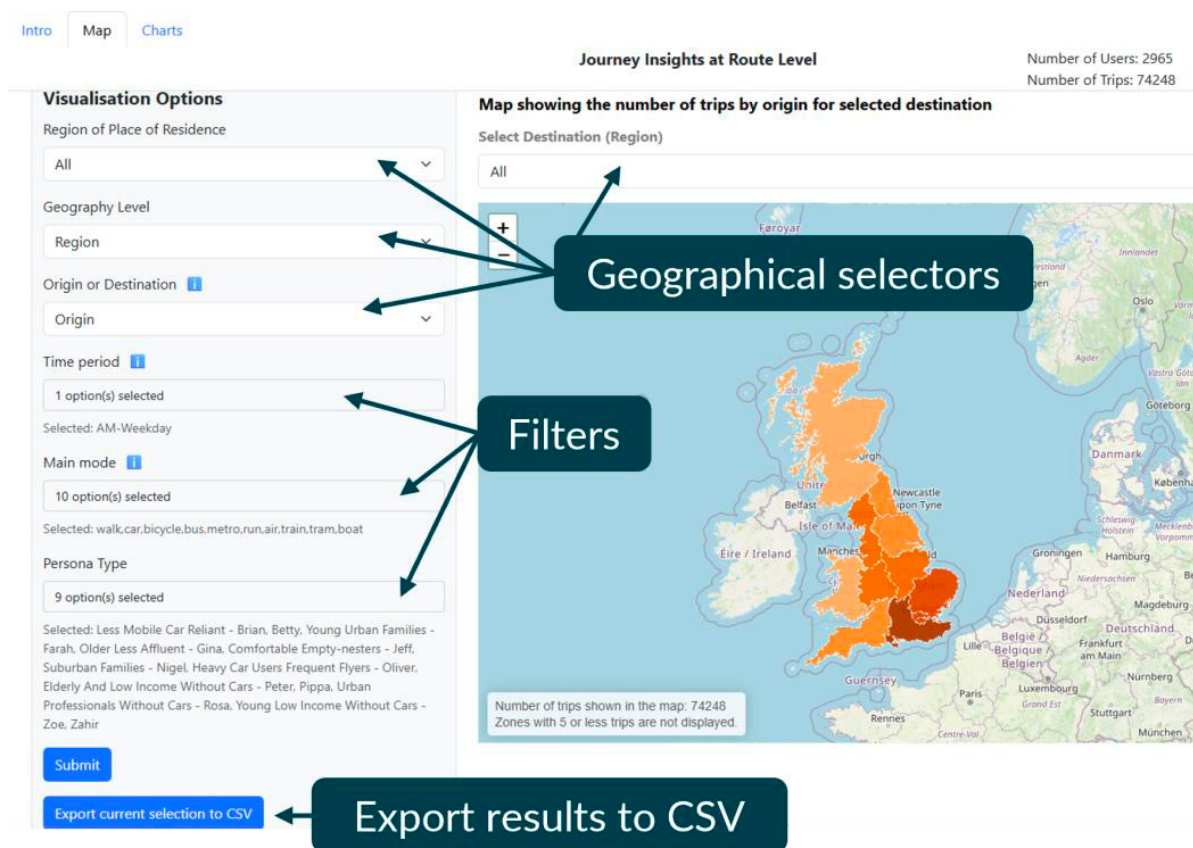
Separately to the Open Datasets, a Detailed Dataset was created for internal use by DfT only. This dataset contains pseudonymised, individual-level mobility data captured by the MyWays app with journey waypoint coordinates.

2.4 PoC Analysis Tool

A PoC Analysis Tool was built to demonstrate to a wider audience how the region and segment version of the Open Datasets can be interrogated, and the types of insights which can be unveiled. A workshop was held with potential future users of the tool within DfT to gather requirements and develop a specification. From this workshop it was determined that the main objectives of the tool would be to:

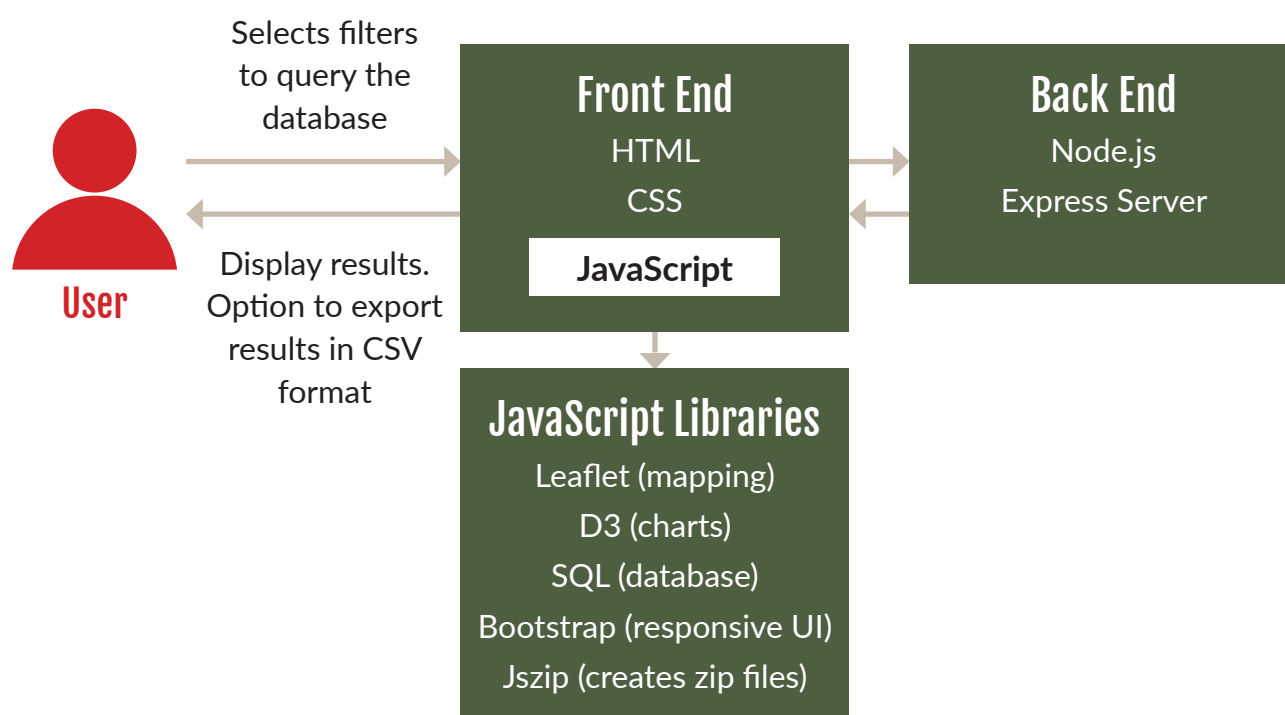
1. **Analyse:**
Support general analysis, e.g. analyse the impact of policy interventions in an area/on a segment.
2. **Inform:**
Easy access to information to advise on modelling assumptions for behaviour modelling, baselining, and more.
3. **Identify:**
Identify quick insights, patterns, and trends about travel behaviour by segment and regions.
4. **Compare:**
Compare travel behaviour across regions, segments and over time.

Figure 8 Screenshot of the tool's Map tab



- The PoC Analysis Tool was built using open-source software. The figure below illustrates its architecture.
- The underlying code will soon be available on the Geographic Data Service.

Figure 9 Illustration of the architecture of the PoC Analysis Tool



3 FINDINGS

3.1 Recruitment and Surveys

7,024

Participants
recruited in
total

4,158

MyWays app
downloads

3,577

Responses on
average per
survey

3,130

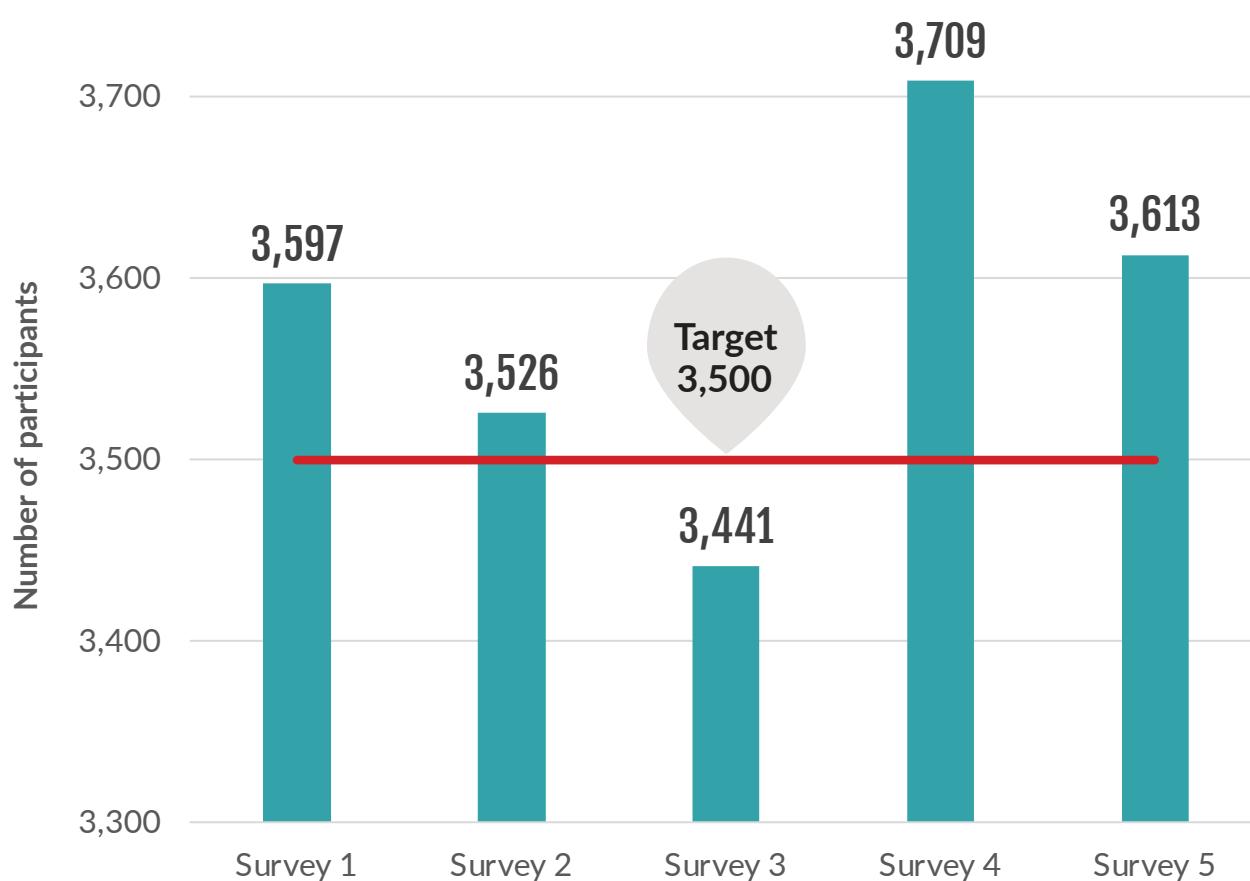
Participants
took part in all
6 surveys

- Participants were recruited from across England (see p6 for regional distribution map), with harder-to-target demographics such as the 70+ age group targeted first.
- The number of responses for each survey were closely monitored. Top-up recruitment was carried out before surveys 4 and 5 to replace drop-outs and ensure a minimum of 3,500 responses were achieved.
- The quotas set were based on region, age, gender, car/van ownership and rural/urban classification.
- Education level and income data were captured in the recruitment survey but did not count towards the quotas, as this would have made recruitment more challenging.
- The resulting sample was nationally and regionally representative by gender and rural/urban classification.
- The sample size and demographic slightly fluctuated by survey/month.
- Segment distribution changed by a maximum of +-2%, with some segments exhibiting no change at all.
- The drop-off of participants from surveys 1 to 3 (seen in figure 11) is normal and expected in this type of project. The increase in responses to survey 4 was a result of top-up recruitment.

Table 10 Summary of quotas against actual recruits per region

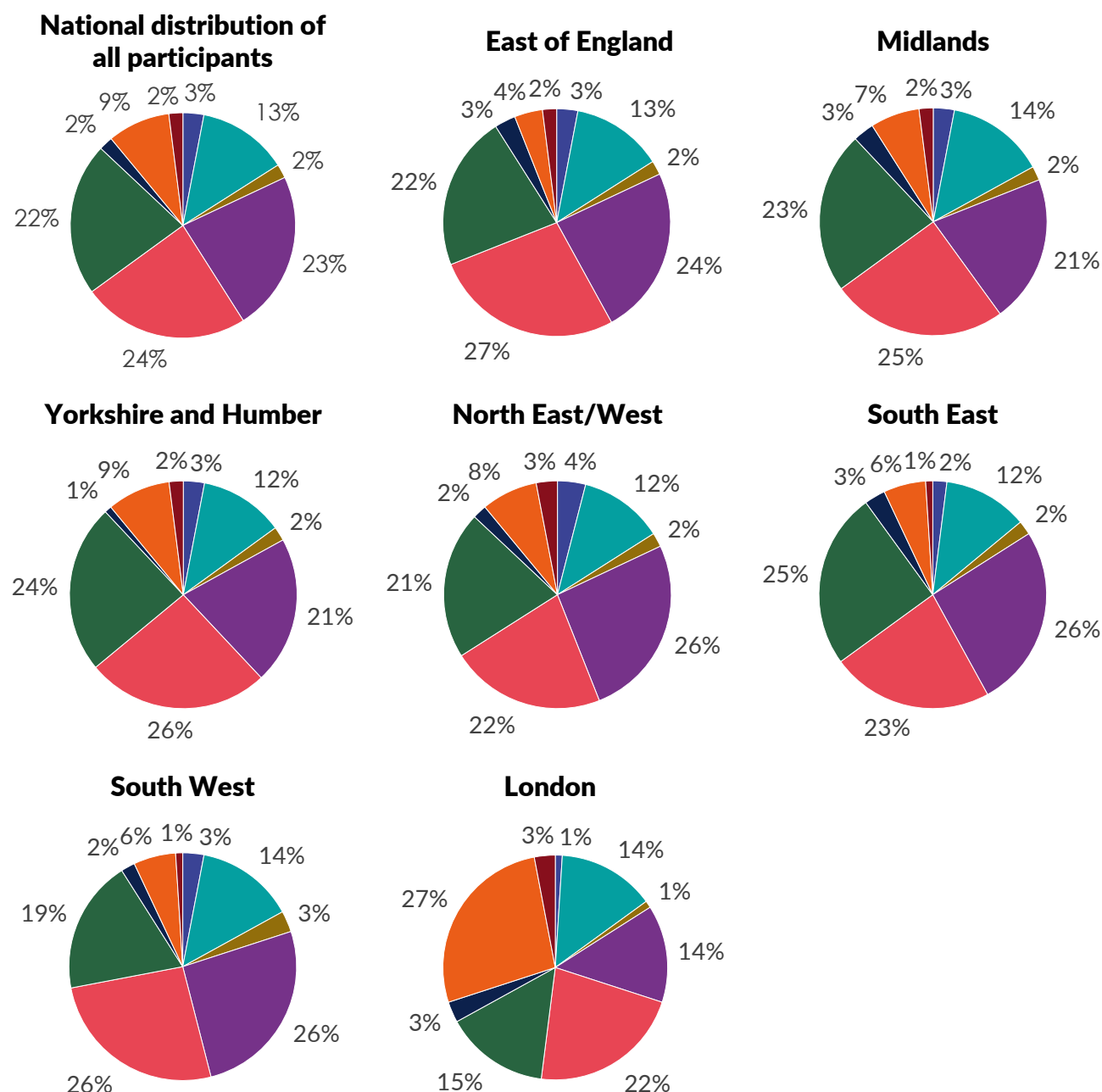
Region	Quota	S1-S5 Avg.	Diff.
London	15.4%	12.7%	-2.7%
North East & North West	17.8%	21.7%	3.9%
South East	16.4%	17.9%	1.5%
Midlands	19.1%	18.6%	-0.5%
East of England	11.2%	11.7%	0.5%
South West	10.3%	8.4%	-1.9%
Yorkshire & The Humber	9.7%	8.8%	-0.9%

Figure 11 Number of responses per survey



The below charts show the distribution of participants classified by transport user segment. The distribution was broadly similar across all regions except London.

Figure 12 National distribution of all participants



Key

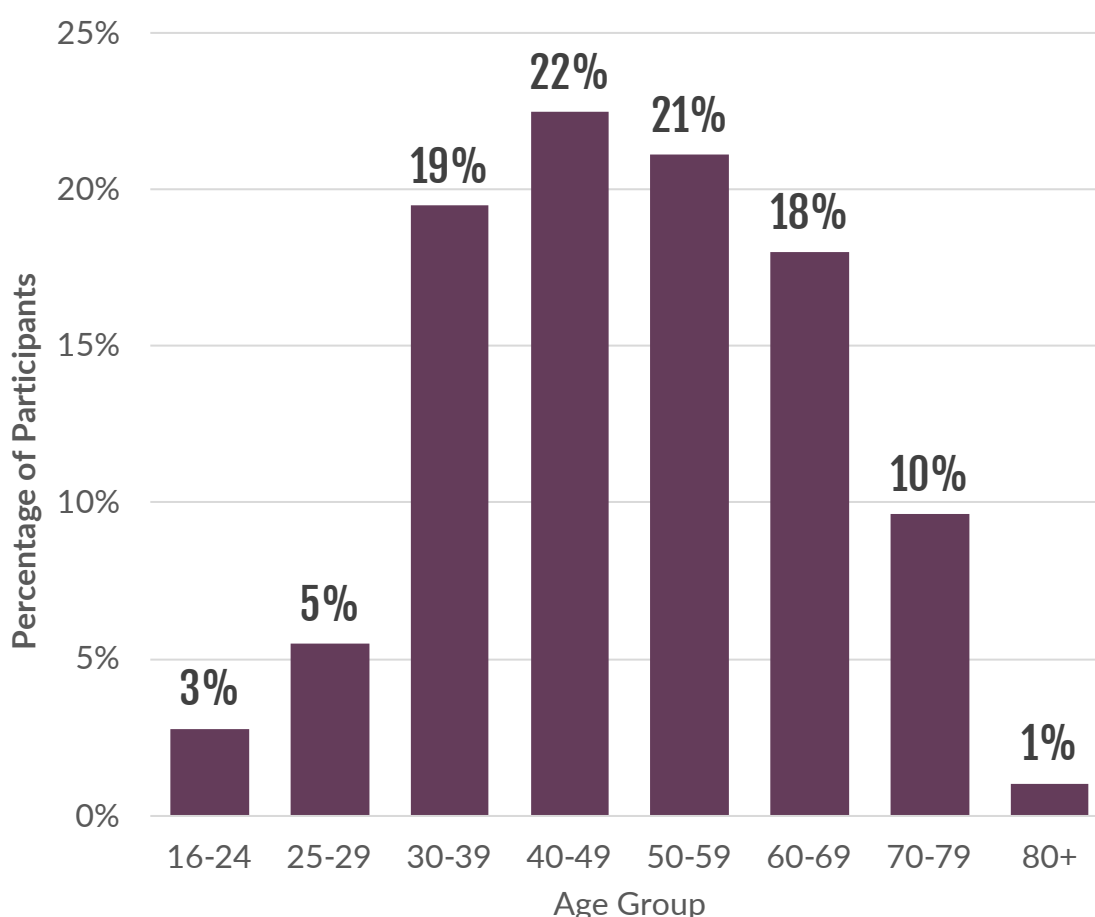
- Less mobile, car reliant
- Young urban families
- Older, less affluent
- Comfortable empty nesters
- Suburban families
- Heavy car user, frequent flyer
- Elderly and low income without a car
- Urban professionals without a car
- Young low income without a car

Comparison with Census data generally found good alignment with participant demographics. Exceptions include:

- Over-representation of ages 35-64; those with more than 1 car; level 4 or above in educational qualifications; families with no children.
- Under-representation of ages 16-34; those without a car; level 1 or below in educational qualifications; students and other economic activity statuses; higher managerial, administrative and professional occupations; single person households; households with >£100k annual income; and those living in the South West region.
- This over- and under-representation is related to the demographic make-up of YDS' survey panel, and the willingness of different demographics to take part in the study.

Comparison of the travel diary surveys with NTS data showed over-representation of car/van as a driver mode and under-representation of walk mode within the survey data. It also showed under-representation of leisure and education trips, and under-representation of trips under 5 miles in length.

Figure 13 Distribution of participants by age group



3.2 MyWays Data Collection

2.1mil

Legs
captured

970k

Routes
captured

45.6mil

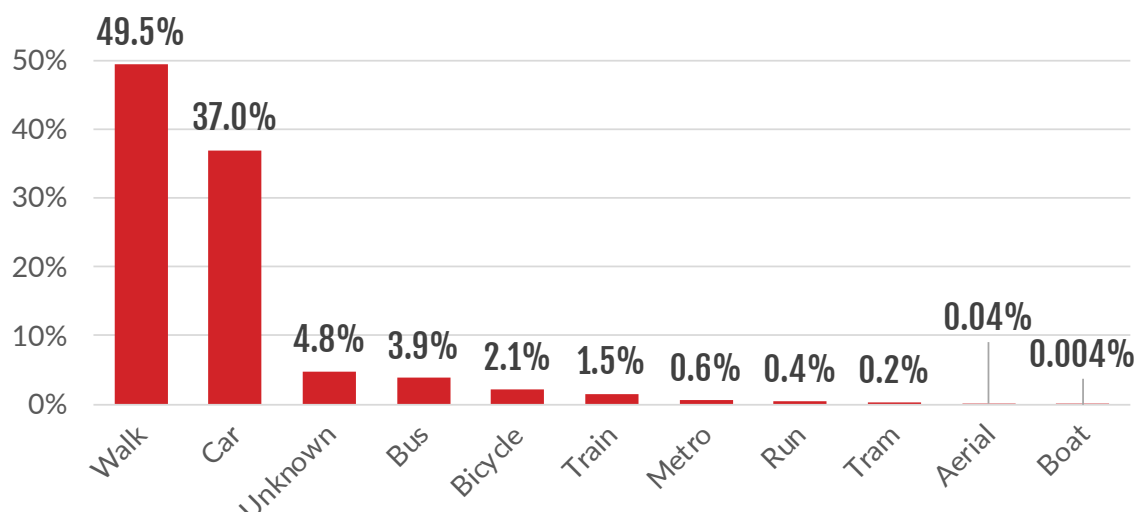
Total distance
of routes in km

132

Aggregated
tables in 3
open datasets

- 6 months of data were collected from a maximum of 3,709 participants (observed in survey 4).
- The Open Datasets generated following the procedures described on p15-17 resulted in 56 tables, aggregated to multiple levels for both routes and legs.
- Participants belonging to the transport user segment called 'Less Mobile Car Reliant' carried out the most trips by car relative to their total trips (51%); 'Urban Professionals Without Cars' carried out the least (12%).
- 'Young Low Income Without Cars participants' carried out the most trips using active modes relative to their total trips (68%); 'Less Mobile Car Reliant' carried out the least (40%).
- Car, walk, and train modes were the most accurately captured modes according to user validation records from the MyWays app.
- The data captured by the MyWays app generally shows agreement with NTS data.
 - In the MyWays data, car and walk modes are under-represented and public transport modes are over-represented.
 - Trips under 2 miles are under-represented, and trips over 10 miles are over-represented.
- Walk and car were the most frequently used modes, accounting for 86.5% of legs captured.
- Known limitations of the MyWays app are:
 - Short trips may not be detected if the phone's sensors are not functioning optimally.
 - The beginnings of journeys can be missed due to phone sensors 'waking up'.
 - Car and bus modes can be confused with each other due to their similarity.
 - Underground trips can be hard to detect due to loss of GNSS signal.
 - Unknown modes are recorded when the app's algorithm cannot detect the mode from the phone's movement.

Figure 14 Percent of legs by transport mode



- User engagement with the MyWays app was higher than expected, with up to 2,900 users engaging with the app daily; this ranges from simply opening the app, to verifying and editing trips.

44%

Legs and routes
user verified

4.4%

Legs edited by
usersNumber of views of each screen
within the app

PAGE TITLE AND SCREEN CLASS	VIEWS
MainActivity	323K
SplashScreenActivity	179K
AddJourneyActivity	139K
JourneyVC	104K
AddJourneyVC	88K
SplashScreenVC	40K
MenuVC	37K

Events happening within the app

EVENT NAME	EVENT COUNT
screen_view	1.1M
user_engagement	1M
session_start	599K
app_exception	50K
first_open	7.4K
app_update	7.4K
os_update	6.2K

Number of times
the app was
launched by
users

- A total of 658 different Apple and Android types of devices were used by participants, with the iPhone 11 being the most popular.
- The network provider, mobile phone operating system and permissions settings all influence the data which are captured.

- Most participants had Android devices (58.5% versus 41.4% iPhone users), but Android users exhibited lower data generation rates and generated greater numbers of 'unknown' transport modes.

3.3 Quality Assurance

The key findings of the QA carried out are summarised below.

YDS Survey Data

- Some survey results suggested that participants may not have fully understood the instructions, e.g. recording the start of their daily activities after midnight in the time use survey.
- Participants may not have recorded very short walking trips in their travel diaries, indicated by the under-representation of walking trips in the survey results.
- The sample size and corresponding demographics of participants fluctuated slightly each month.
- Comparison with the ONS time use survey showed that participants spent less time carrying out key activities such as sleeping, working, and eating.

MyWays App Data

- Network provider, mobile phone operating system and permissions settings, and geographical location in relation to cell towers influence the data captured.
- Short trips are not as easy to detect due to the phone's sensors taking time to 'wake up'.
- Car and bus are the modes most often confused for each other due to their similarities.
- Underground trips are hard to detect due to a lack of GNSS signal.
- A broad range of individual-level data generation rates were observed, perhaps due to MyWays app settings, geographical location or personal phone usage habits.
- According to user validation statistics, tram and metro modes were harder to identify, and cycle mode was over-identified.

Open Datasets

- Data loss caused by aggregation to meet a minimum sample size of 10 trips or participants mostly affected the region/segment set of data tables due to them having a greater number of attribute combinations.
- On average, 34% of data within the region/segment route-level tables meet the minimum sample size threshold. For region/segment leg-level tables this value is 64%, and 97% for survey data tables.
- Urban/rural classification tables saw an average data loss of 29%, and of 32% for age group, due to the minimum sample size not being met.

3.4 Potential Applications

Synthetic Population

A synthetic population is a statistically representative construct of an actual population which is used as an input to simulation models, such as agent-based models. They allow simulations to be generated which realistically reflect human behaviour without comprising personal privacy.¹

The possibility of using the Open Datasets and closed Detailed Dataset as a data source to generate synthetic population for purposes such as the National Trip End Model (NTEM) were evaluated. The assessments and conclusions are shown below.

Potential Application	Conclusion
Use of new information from the Open and Detailed Datasets as synthetic population inputs, such as transport user segment distribution and activity patterns.	This could increase the diversity of the synthetic population and deepen the understanding of their travel needs, allowing richer behaviour to be captured in downstream applications, such as agent- or activity-based modelling.
Generation of demographic insights from the responses to the golden questions, contained within the recruitment survey and used to assign participants to their transport user segment.	The survey responses could facilitate detailed understanding of the demographic representation of the sample and could be used as an alternative to simpler Census data-based scaling of the sample. It could also be used to simulate a synthetic version of the detailed dataset, allowing granularity to be retained whilst preserving privacy.
Use of the transport user segment classifications to control forecast populations to match a desired change in demographic proportions.	The detailed dataset could be used to provide new controls for variables not contained within the Census, and to enhance the granularity of input data. If it is proven feasible to associate a transport user segment with a synthetic population seed, future scenarios can be experimented with.

The individual-level data contained within the Detailed Dataset would be suitable for use as a source to generate synthetic populations at multiple levels of geographical aggregation (local, regional and national).

This cannot be done using the Open Datasets because of the aggregation carried out to create them.

The overall conclusions of the synthetic population assessment were that the individual-level data contained within the Detailed Dataset would be suitable for use as a source to generate synthetic populations at local, regional and national geographical levels. The benefit of this would be outputs which are richer in travel patterns and behavioural insights.

¹ [A Synthetic Population of Sweden: Datasets of Agents, Households and Activity-Travel Patterns](#)

Synthetic populations cannot be generated using the Open Datasets because they are aggregated, which means that individual-level patterns are lost. It would therefore be useful in future to explore new approaches of data anonymisation which can preserve both granularity and privacy.

The Open Datasets can be used to complement and enhance inputs to synthetic populations.

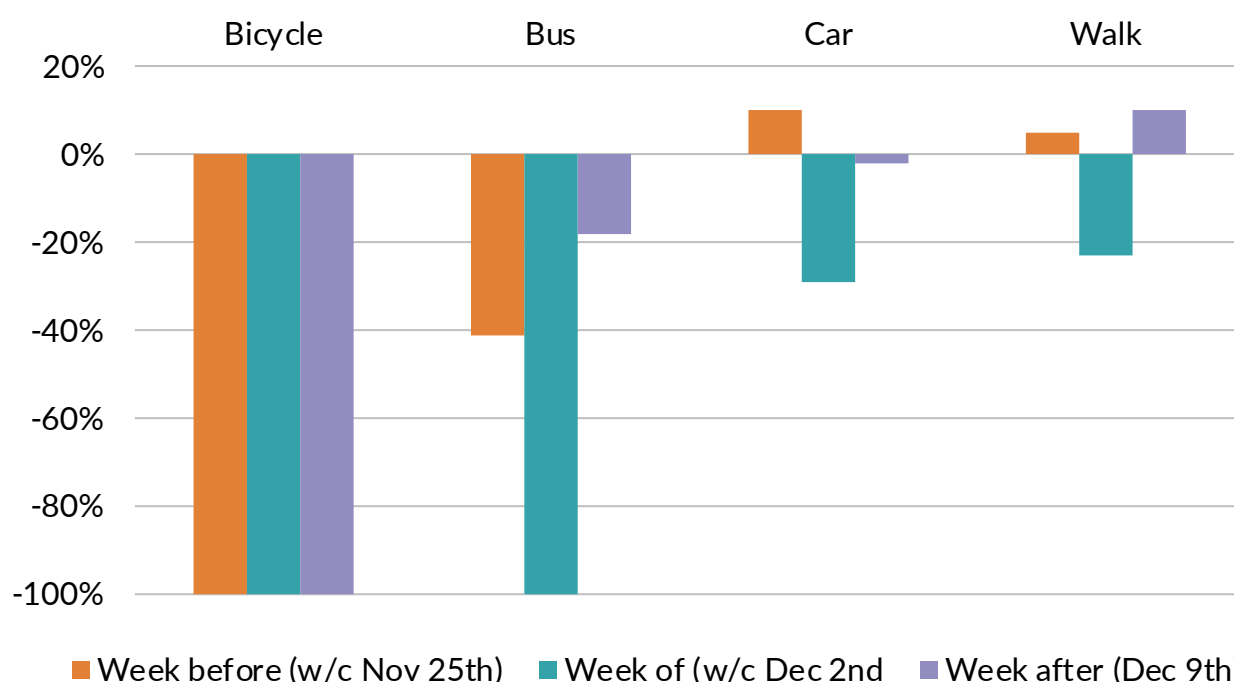
Natural Experiment and Policy Evaluation

Storm Darragh peaked in Wales and the South West of England on December 7th 2024. Although participants all resided in England, their travel throughout the UK was captured. The storm therefore provided an opportunity to assess the potential of the Open Datasets and PoC Analysis Tool to visualise changes in travel patterns in response to a disruptive event.

Using the region-segment version of the Open Dataset, average weekly journey legs by mode in Wales and South West England for the week before, during, and after Storm Darragh were compared to the 6 month period average. Bicycle and bus journeys were lower than average during all 3 weeks. Car journeys increased the week before but dropped during the storm, whilst walking legs rose before and after the storm, but decreased during the week of the storm.

Note that this analysis was limited to the number of compliant records available for the weeks of interest.

Figure 15 Percentage change in weekly average leg count compared to weekly average over project



The comparison function in the PoC Analysis Tool was used to compare journey patterns by persona over the below dates:

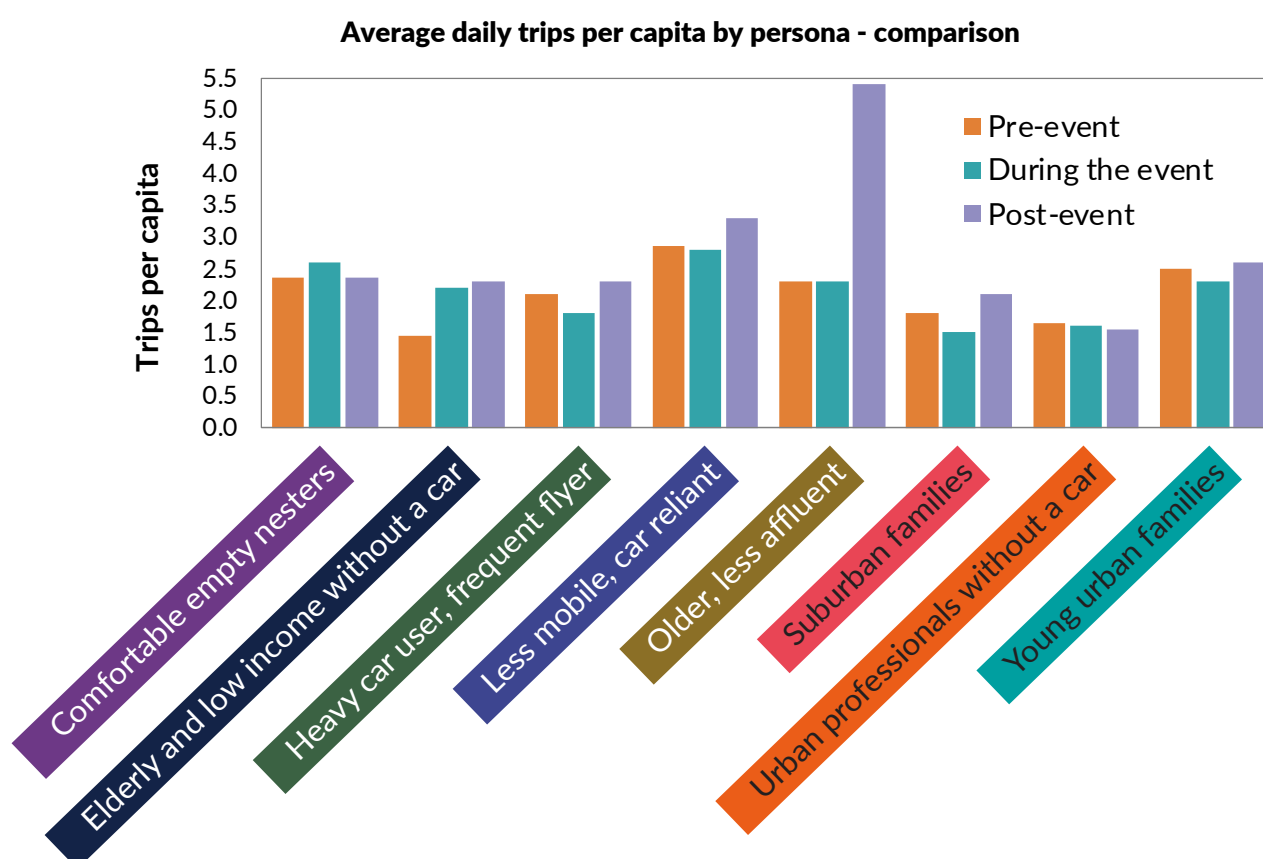
- Pre-Event: 1st to 6th December 2024 (inclusive)
- During the event: 7th December 2024
- Post-Event: 8th to 13th December 2024 (inclusive)

Figure 16 shows that some personas carried out more trips before the storm peaked; most personas carried out fewer trips during the storm; and trip numbers mostly increased post-storm. Variations across personas may be related to reliance on certain modes of transport; safety concerns; tendency to prepare for known disruptions; and need to travel.

Note however that the underlying number of trips per persona were fewer than 10, and results should therefore be treated with caution.

Figure 17 shows changes in trips per capita by mode. All modes except walk saw a comparative decrease on the 7th of December.

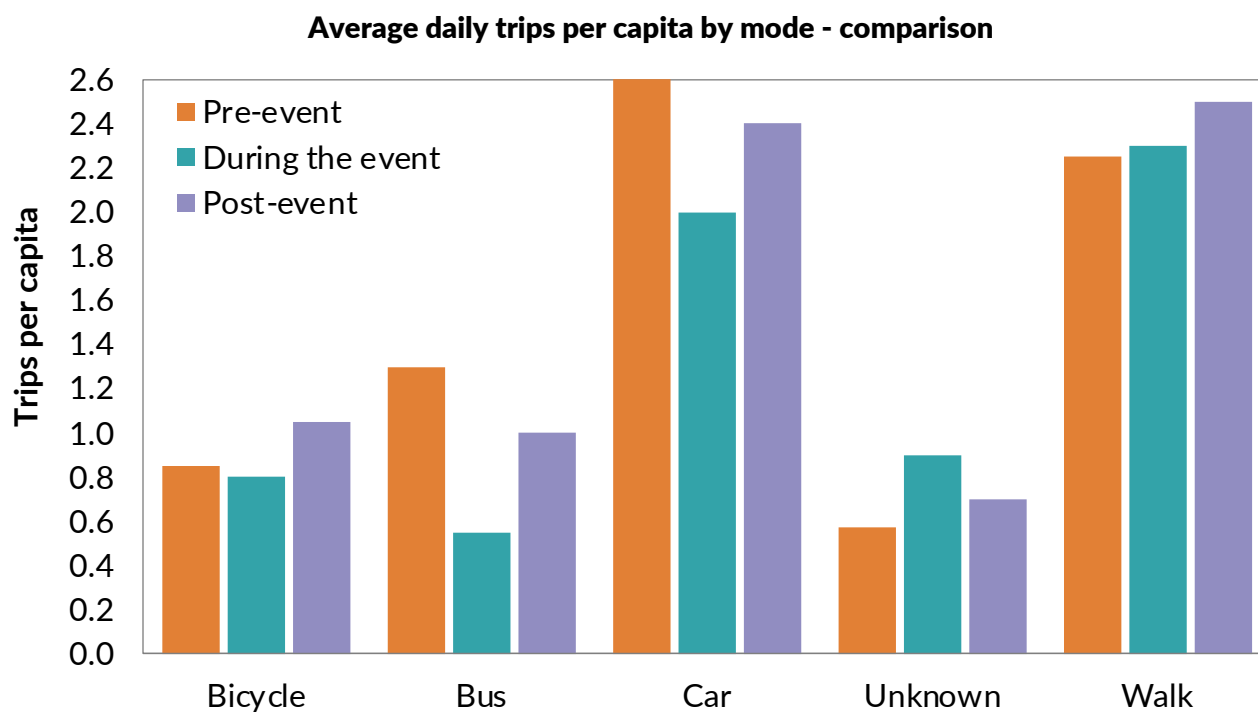
Figure 16 Comparison of trips per capita, per persona from LAs within South West England and Wales to all LAs in the UK



The conclusion of the assessment of the Open Datasets and PoC Analysis Tool was that the data and tool allow comparison of travel patterns before, during and after time periods of interest. In this case, the tool presented an insight into how different demographics respond and behave around disruptive events.

The robustness of insights drawn from the data and tool are dependent on the underlying sample sizes.

Figure 17 Comparison of trips per capita by mode from LAs within South West England and Wales to all LAs in the UK



4 LESSONS LEARNED

A lessons learned session was held with all project partners to identify ways in which the methodology could be improved for future work. The key outcomes were grouped into the following 3 categories:

Recruitment and Data Collection

- YDS found it challenging to achieve representation in the initial study areas of London, South East, and North East and North West combined, and needed to expand the study area to the entirety of England to successfully recruit 3,500 participants.
- A larger panel of potential participants, or a different survey method (such as CATI) would be required for smaller study areas.
- Targeting fewer demographic characteristics would make recruitment easier, but result in reduced representation.
- More cognitive testing is recommended for future work to ensure that participants fully understand questions, leading to more accurate data capture.
- Oversampling of demographics observed to under-generate MyWays data could be explored.

Data Governance

- Carrying out data governance was complex and time consuming, but the process is now well understood, and a template available for re-use in the future.
- Participant-level records are better for modelling purposes, but this presents privacy challenges. More advanced anonymisation techniques could be explored in future to potentially overcome this.

Technical

- The vast volumes of data generated presented processing challenges in terms of storage and computing time. Working with the data in a database such as PostgreSQL would be more efficient.
- There are currently limitations in the performance of the MyWays app which could be improved upon (see p25), such as the ability to consistently detect certain transport modes.
- The ability of the MyWays app to infer trip purpose is desired.

5 SUMMARY AND CONCLUSIONS

The overall aim of FUSION was to explore a novel approach of generating granular travel insights by combining big, mobile phone sensor-derived data and online surveys, and to understand if this method could support strategic modelling and inclusive policy-making by:

- Improving our understanding of how travel behaviours differ by demographic group.
- Examining the impact of disruptive events and moments of change geographically, temporally, and demographically.

SYSTRA, Yonder Data Solutions and TravelAi worked closely together and overcame multiple challenges to successfully explore this new approach, which captured a greater volume of more detailed travel information. 3 Open Datasets, a Detailed Dataset, and a PoC Analysis Tool were delivered.

The method, Open Datasets and PoC Analysis Tool were examined to understand their utility in supporting the aims of the project.

- The Detailed Dataset (DfT use only) and Open Datasets allow users to carry out custom analyses of transport user segments' travel behaviours temporally and spatially, and allow quantitative mobility data - the how, when, and where - to be linked to qualitative survey responses - the why - to understand related reasoning.
- The PoC Analysis Tool allows quick insights from the Open Datasets to be derived, visualised and exported, at various geographical, temporal and attributional levels. Its comparison functionality allows changes in behaviours before, during, and after an event of interest to be viewed.

The primary limitations of the approach are:

- Challenges recruiting a representative sample of multiple demographic characteristics. This is constrained by the composition and spatial distribution of the source participant panel.
- A comparison with the NTS indicated which demographics were under- and over-represented in this pilot project. This could be mitigated in the future through oversampling or weighting.
- Small sample sizes at granular geographical, temporal and attributional levels restrict the utility of the Open Datasets, because they do not meet privacy thresholds.
- Limitations of the MyWays app are related to its algorithms and the mobile phone's sensors. Examples include mode confusion and limited underground trip detection due to lack of or weak cell signal. User validation statistics can help to understand these limitations.

Valuable lessons were learned throughout the course of the project which will be taken forward to improve the outcomes of future applications.

- In the future, more time for survey development and testing would help to optimise survey data collection.
- More advanced anonymisation techniques could be explored to maximise the value of the data whilst protecting participant privacy.



WWW.SYSTRA.COM/UK/

Whilst every care has been taken to ensure that the information contained in this publication is accurate, neither SYSTRA Limited makes no representations, warranties or guarantees with respect to the content, nor accepts responsibility or liability for errors or for information which is found to be misleading. This publication is not intended to amount to advice on which you should rely and you must obtain professional or specialist advice before taking, or refraining from, any action on the basis of the content of this publication.



Copyright 2024 SYSTRA Limited. SYSTRA Limited is registered in England under number 03383212
Registered office: 3rd Floor, 1 Carey Lane, London, England EC2V 8AE. Tel: +44 20 3855 0079