

## **User Guide for Intermediate Zone Level Origin-Destination Matrix Data** **(Glasgow City Region)**

### **Introduction**

This guide provides essential information on how to interpret and utilize the Origin-Destination (OD) matrix data for the Glasgow City Region (GCR) derived from GPS data generated from smartphone apps. Since 2019, Huq.io, an independent research company based in London specializing in location intelligence, has been aggregating this data from various UK cities through smartphone apps. The firm collaborates with a wide array of partners, including more than 100 government agencies, consultancies, retailers, and universities. This GPS data provided by Huq.io is used to estimate the number of trips between each intermediate zones.

This dataset includes records from 10,626 users totaling 22,716,662 observations in 2019, expanded to 30,436 users with 169,529,841 observations in 2020, 30,268 users contributing 356,306,418 observations in 2021, and 8,814 users generating 105,852,747 observations in 2022. These estimated trips are then converted into percentages, representing their share of the total number of trips in the region.

The estimated trips are weighted based on the activity level of each individual, which is determined by their total number of observations within the dataset. This approach helps to standardize the data, accounting for variations across different years. For instance, if a user is active for 183 days and makes 500 trips during that period, we extrapolate this to 1000 trips for the full year. Furthermore, to achieve more accurate and representative counts, the estimated trips are weighted according to the estimated home location of each user. This adjustment ensures that the data better reflects variations across different Scottish Index of Multiple Deprivation<sup>1</sup>(SIMD) categories and council areas.

The Origin-Destination (OD) Matrix is calculated for each of the four years at the intermediate zone level, encompassing 417 zones within the Glasgow City Region (GCR).

The final output i.e. the 2022 Origin-Destination (OD) matrix aims to reduce the instances of missing OD pair combinations by presenting the data as an average percentage of trips over both 2021 and 2022. This approach is taken to mitigate the impact of the pandemic on travel patterns, ensuring that the final output more accurately reflects typical travel behaviors by averaging out the anomalies caused by the pandemic's effects on transportation.

### **Understanding the OD Matrix**

The OD matrix gives the percentages of total number of trips originating in one intermediate zone and ending in another within the Glasgow city region. The data consists of the following variables explained in Table 1 below.

---

<sup>1</sup> <https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020/>

Table 1: Data Description

Variable	Description
origin_iz_id	ID of origin intermediate zone
destination_iz_id	ID of destination intermediate zone
origin_iz_name	Name of origin intermediate zone
destination_iz_name	Name of destination intermediate zone
trips_percentage	Percentage of the total number of trips

These percentages are relative to the total number of trips in the Glasgow City Region. A higher percentage in a cell signifies a more frequently travelled route from that particular origin to the destination. To convert these percentages back into the actual number of trips at the individual intermediate zone level, knowledge of the total number of trips in the Glasgow City Region is necessary. One potential source for estimating the total number of trips is the Scottish Household Survey<sup>2</sup> data.

To facilitate the analysis of travel trips during peak and non-peak hours, four distinct types of Origin-Destination (OD) matrix files are made available, each catering to specific travel patterns and times. Here's a detailed explanation of each type:

- Type 1 - AM Peak Weekdays (7 am-10 am):** This matrix captures travel patterns during the morning rush hours on weekdays.
- Type 2 - PM Peak Weekdays (4 pm-7 pm):** Similar to the AM peak matrix but for the evening rush hours.
- Type 3 - Everything:** This comprehensive matrix encompasses all recorded travel trips without any time restrictions. It's ideal for getting an overall sense of travel behavior, including both peak and off-peak hours.
- Type 4 - Off-Peak Hours (Type 3 - Type 1 - Type 2):** Unlike the other three, this matrix is derived by subtracting the trips recorded in the AM and PM peak matrices from the 'Everything' matrix. It offers a focused look at travel during off-peak hours, excluding the morning and evening rush.

Each of these OD matrices serves as a valuable tool for transportation planning, traffic management, and policymaking, enabling a nuanced analysis of travel patterns during different times of the day and for various purposes.

## Data Considerations

When using the Origin-Destination (OD) matrix derived from mobile app data, it's important to consider several key factors, such as:

- **Coverage Bias:** Mobile app data may not be representative of the entire population, as it is dependent on smartphone ownership and the usage of specific apps. This can result in

<sup>2</sup> <https://www.gov.scot/collections/scottish-household-survey/>

\*\* This data product is currently in its beta phase and may undergo changes as methodologies improve, thus it should be used with caution \*\*

underrepresentation of certain demographic groups, such as older adults or lower-income individuals who may have lower rates of smartphone ownership.

- **Sampling Bias:** The dataset is derived from users who have opted into location tracking, which may not be a random sample of the population. This self-selection bias can affect the generalizability of the findings.
- **Data Processing and Interpretation:** The methodology for converting raw GPS pings into meaningful OD trips involves assumptions and modelling choices that can influence the results. Errors in trip detection, or inaccuracies in estimating home locations can affect the quality of the OD matrix. Given the innovative nature of the data source and methodology, it's advisable to approach the OD output with caution.
- **Data Scaling Concerns:** The recorded trips between certain intermediate zone OD combinations may be based on data from a limited number of individuals, which could result in inaccuracies or insights that don't fully represent typical trip behavior. To address this, the final output is an averaged OD matrix, aiming to reduce potential biases.
- **Data Access Restrictions:** Access to the OD output data is restricted and available only to select applicants who have secured the appropriate license with the Urban Big Data Centre.
- **Beta Product:** This data product is currently in its beta phase and may undergo changes as methodologies improve, thus it should be used with caution.

\*\* This data product is currently in its beta phase and may undergo changes as methodologies improve, thus it should be used with caution \*\*