BUSTO Dataset User Guide

Background and Purpose

Transport for London (TfL) produces the BUSTO dataset each year to provide an understanding of customer usage and travel patterns on its bus network. This detailed dataset is the best available estimate of bus boardings, alightings and loadings on a typical day. TfL uses this dataset to assess bus service provision, demand profiles and customer experience, as well as inform service planning and performance measurement.

BUSTO is based on observed passenger data from autumn each year, which is a typical period within the busiest time of year on London's public transport networks. None of the values is a count of individual passengers on any particular day; rather they are averages over the period in question. The figures are presented to high resolution (i.e., each bus stop for each 15 minute period in each day type). The figures are sometimes therefore low absolute values (e.g., if 100 people use a bus stop on a typical day then the average quarter hour value is 1.2 passengers). Non-integer values are provided to best represent the average or typical value.

This document provides detail on the 13 csv files as published on crowding.data.tfl.gov.uk, including a summary of important assumptions and exclusions. Please refer to the enclosed data dictionary for the corresponding file/s as appropriate.

Coverage

The datasets represent the travel demand on the bus network on three day types: a typical weekday (Monday-Friday), Saturday and Sunday. BUSTO is based on ticketing data recorded between November and December. This period of autumn each year is most representative of typical travel on the network as demand is usually at its highest, and is it least affected by seasonality effects including school holidays. Data is provided for every 15-minute period throughout each of the three day types.

Features

BUSTO uses ticketing data from Oyster and Contactless Payment Card (CPC) transactions generated from the ticket machines onboard buses, in addition to data from the iBus Automatic Vehicle Location system. The data also is scaled to account for users of paper tickets, non-validation and customers who are not required to touch in upon boarding. This ensures all customer trips are accounted for.

Notes on calculations and process

- The published data is representative of passenger demand for a typical weekday, Saturday and Sunday in autumn of each year. Passenger demand data is extracted for each date and averaged across a count of each of the three day types accordingly to generate "typical" demand figures.

- Some consecutive dates were excluded as input to the BUSTO dataset. This was due to significant disruptions noted on these dates, which resulted in the source data being deemed unrepresentative of typical network operations.

- Because the main data comprises of electronic taps data from the Oyster and contactless system, it needs to be scaled to account for paper ticket users and non-validation. This is done using scaling factors which are based on the Greater London Bus Passenger Survey (GLBPS). The GLBPS involves a manual survey of a sample of scheduled bus services across the network every quarter. The survey results are then scaled using data from the Oyster and contactless system to obtain figures that are representative of the whole network. As part of the survey, the following information is collected:
    1. The volume and value of total travel made by holders of Freedom Passes and National Concessionary Bus Passes. This is used to provide the basis of payments made by the London boroughs to TfL for this facility.
    2. The volume of bus travel made with all ticket types, including Travelcards, Bus Passes, Oyster Pay-As-You-Go, contactless payment cards, Freedom Passes and Free Child Passes. This includes a count of the number of instances where non-validated travel occurs for each ticket type, which would otherwise not be able to be collected by the Oyster and contactless ticketing system.
    3. The volume and reasons for ticketless travel on buses; this includes fare evasion.

  Given the granularity of the data, in compliance with GDPR regulations, data from the GLBPS is not released publicly. The scaling factors generated from the GLBPS and applied are unique to each route, day type, hour in the day, and ticket type that is used e.g., Adult Oyster, Freedom Pass, Zip Card etc.

Notes on accuracy

- The Oyster and contactless system only records the boarding stop of a bus passenger's trip. Thus, TfL uses an algorithm to infer the most likely alighting stop for that passenger trip. This is based on when the card is next used, either on another bus or a tap-in at a train station. This inference algorithm is successful for between 70 and 80 per cent of all bus journeys, with the remainder scaled up.
- As all alighting values are inferred, all loadings and occupancy values are estimated based on this inference.
- The Oyster and contactless system requires a GPS connection to always be active whilst a bus is in service to accurately record the time and location of a passenger boarding. However, the GPS connection can at times be lost which can affect the accuracy of this data and affect the ability to successfully infer an alighting point. Where possible, corrections have been made to account for these data gaps.
- Data is not included for school day only routes (route numbers 600-699), mobility route 969, temporary routes such as London Underground rail replacement routes or routes operated during special events. Also data is not included in 2023/24 BUSTO for the new Superloop routes SL1 and SL10 as these routes were introduced during the autumn period and demand profiles were not yet sufficiently established to yield representative results.
- Some stops are only used at specific times of the day or week, or only during special events such as for concerts or rail replacement services, and thus will not feature in this dataset if not served at the time by a route in regular service.
- Stops that have demand recorded some of the time will still be included. For example, if one person boards at a stop on a Friday, then given the dataset covers a weekday average from Monday to Friday, the final boarding figure will be greater than zero, but less than one.
- Data for stops not scheduled to be served by a route are excluded e.g., stops served because of temporary diversions. Only stops that were scheduled to be served by the route as recorded in iBus on the dates selected are included.
- Boarding and alighting data may not always be captured correctly in Hail & Ride sections. Some routes with Hail & Ride sections have waypoints coded in which can record demand that is specific to the general area served by the vehicle. This however is not universal across all routes with Hail & Ride sections. Where they do not feature these waypoints, the demand is instead allocated to the next fixed stop served along the route.
- It is expected that some stops and roads will be closed at any given time for maintenance and may result in temporary diversions. Therefore, some stops will have zero or reduced demand if closed on most of the selected dates.
- TfL manages a bus network comprising of approximately 675 routes serving more than 19,000 bus stops, with a bus fleet of around 9,000 vehicles operating across more than 3,000 kilometres of road. A network as large and complex as this is expected to undergo various changes throughout the year due to service changes, temporary bus stop closures, diversions, curtailments, or other operational reasons. Our ticketing and software systems can sometimes encounter difficulties in updating and incorporating the impact of these changes as they occur, especially when at short notice, which can lead to occasional data failures. Consequently, this can lead to incomplete passenger demand data on any given day, however where possible, we endeavour to supplement these gaps with data from alternative sources.

Guide to TOTAL DEMAND datasets

<u>Overview</u>

This dataset provides information on the total travel demand on the bus network, broken down by day type, route, direction of travel and stop, for every 15-minute period of a typical weekday, Saturday and Sunday in the autumn period. The data is totalled across all buses recorded within each corresponding 15-minute period.

The published datasets are grouped into four categories as follows:
1. Weekday TOTAL DEMAND BY ROUTE BY QUARTER HOUR
2. Saturday TOTAL DEMAND BY ROUTE BY QUARTER HOUR
3. Sunday TOTAL DEMAND BY ROUTE BY QUARTER HOUR
4. MAX DEMAND HOUR BY ROUTE BY TIMEBAND

For each of file groups 1-3, four files are provided, which are based on route numbers. This was done in consideration for file size and table row limits. Details on the route numbers contained within each file are provided below, and in the corresponding csv's filename:
- Routes 1 through 149.
- Routes 150 through 299.
- Routes 300 through 549.
- All letter-prefix routes, including "N" night routes. Note that data on 24-hour routes are contained within their parent day route and are not reported separately.

This data is presented at a quarter hour level, which is highly disaggregated and thus can lead to difficulties in interpreting usage for lower frequency routes, where observations can be recorded on some but not all successive quarter hours. Thus, it is recommended that all quarter hourly data is first aggregated to hourly level (i.e., sum of four consecutive quarter hours, but that need not be a 'clockface' hour) before any interpretation and/or analysis takes place.

<u>Alternative Uses</u>

As these datasets sum the demand across all buses and then averaged, it is possible to obtain total demand for routes and stops for a typical weekday, Saturday and Sunday. This can be done via pivot tables. Note that data for stops may be contained across multiple files where they are served by more than one route.

Example Data

The table below is an extract from the 2023_24 Weekday TOTAL DEMAND BY ROUTE BY QUARTER HOUR file. This shows that, on route 1 in the direction towards Canada Water and during the Weekday AM peak timeband (Timeband 2), at stopcode 37436 at Chalk Farm Station:

- An average of 21 customers boarded route 1 buses in the hour from 0800 to 0900 (broken down as approximately 8 in the quarter hour from 0800 to 0815, 5 from 0815 to 0830, 4 from 0830 to 0845, and 4 from 0845 to 0900).
- An average of 6 customers were inferred to have alighted route 1 buses in the hour from 0800 to 0900 (broken down as approximately 1 in the quarter hour from 0800 to 0815, 1 from 0815 to 0830, 2 from 0830 to 0845, and 2 from 0845 to 0900).
- An average of 104 customers were onboard route 1 buses on departure from Chalk Farm Station in the hour from 0800 to 0900 (broken down as approximately 26 in the quarter hour from 0800 to 0815, 27 from 0815 to 0830, 28 from 0830 to 0845, and 23 from 0845 to 0900).
- An average capacity of 567 customer spaces (seated + standing) was provided on route 1 buses departing from Chalk Farm Station in the hour from 0800 to 0900 (broken down as approximately 144 spaces in the quarter hour from 0800 to 0815, 148 from 0815 to 0830, 131 from 0830 to 0845, and 144 from 0845 to 0900).
- An average seated capacity of 390 customer spaces was provided on route 1 buses departing from Chalk Farm Station in the hour from 0800 to 0900 (broken down as approximately 99 spaces in the quarter hour from 0800 to 0815, 102 from 0815 to 0830, 90 from 0830 to 0845, and 99 from 0845 to 0900).
- The average occupancy – indicated by the column V/C – across all route 1 buses estimated in the hour from 0800 to 0900 was approximately 18%.

Excel Row	YEAR	DAY_TYPE	TIMEBAND	QHr	ROUTE	DIRECTION	STOPCODE	STOPNAME	STOPSEQUENCE	Boardings	Alightings	load	Capacity	Seats	V/C
228159	2023	Weekday	2	08:00:00	1	1	37436	CHALK FARM STATION	8	7.72	1.35	25.67	143.55	99	0.18
240152	2023	Weekday	2	08:15:00	1	1	37436	CHALK FARM STATION	8	5.38	1.19	27.44	147.9	102	0.19
252861	2023	Weekday	2	08:30:00	1	1	37436	CHALK FARM STATION	8	3.59	1.50	27.57	130.5	90	0.21
265219	2023	Weekday	2	08:45:00	1	1	37436	CHALK FARM STATION	8	3.64	1.77	23.47	143.55	99	0.16

TOTAL DEMAND BY ROUTE BY QUARTER HOUR – Data Definitions

Column Heading	Description
YEAR	Calendar year for which the data is extracted.
DAY_TYPE	Weekday, Saturday or Sunday.
TIMEBAND	Unique ID number from 1 to 17 that corresponds to a specific day type and time period. A lookup table is provided at the end of this guide.
QHr	Start of the quarter hour for which data is given e.g., 07:00:00 represents the quarter hour from 0700 to 0715.
ROUTE	TfL route number.
DIRECTION	Direction of travel, indicated as either 1 or 2. Directions do not have a geographic attribute and can be different for routes travelling in the same direction on a road.
STOPCODE	The TfL stopcode as recorded in the Oyster and contactless system. Not to be confused with the NAPTAN code or SMS code that is provided on stop infrastructure.
STOPNAME	The stop name as recorded in the Oyster and contactless system.
STOPSEQUENCE	The stop sequence number in relation to the route and direction of travel as recorded in the iBus schedule.
Boardings	The average number of boardings at the stop.
Alightings	The average number of estimated alightings at the stop.
Load	The average departure load at the stop.
Capacity	The average total capacity provided on departure from the stop. Note: this is based on the capacity of a bus when full and the vehicle model contracted to run on the corresponding route. The full vehicle capacity can range from 40 to 87 spaces.
Seats	The average number of seat spaces provided on departure from the stop.
V/C	The average volume-to-capacity ratio, also referred to as the occupancy. Calculated as the load divided by the capacity.

Guide to MAX DEMAND dataset

<u>Overview</u>

This dataset provides information on the peak hourly demand on the bus network, for each day type, route, direction of travel and stop, for every timeband in the autumn period.

The peak hour is defined as the hour within which the highest estimated onboard load across all buses is identified. Peak hours are identified for each timeband, route, direction of travel, and stop, and are calculated on a rolling 15-minute period basis; this means that the peak hour may not necessarily occur at the same 60-minute interval within a timeband at one stop, when compared to the peak hour at another stop, despite both stops being served by a common route.

One file is provided which covers all day types and routes.

Example Data

The table below is an extract from the 2023_24 MAX DEMAND HOUR BY ROUTE BY TIMEBAND file. This shows that, on route 1 in the direction towards Canada Water and during the Weekday AM peak timeband (Timeband 2):

- At Belsize Park Station, the highest demand is estimated to be in the hour from 0800 to 0900, with a departure load of 60 summed across all buses and averaged across the included dates (Note: this is not the departure load per bus). This is against a total capacity of 574 spaces provided in the same hour.
- At Chalk Farm Station, the highest demand is estimated to be in the hour from 0745 to 0845, with a departure load of 107 summed across all buses and averaged across the included dates (Note: this is not the departure load per bus). This is against a total capacity of 592 spaces provided in the same hour.

YEAR	DAY_TYPE	TIMEBAND	QHr	ROUTE	DIRECTION	STOPCODE	STOPNAME	STOPSEQUENCE	HourBoardings	HourAlightings	HourLoad	HourCapacity	HourSeats	V/C
2023	Weekday	2	08:00:00	1	1	29927	ROYAL FREE HOSPITAL	1	33.76	0	33.76	578.55	399	0.058
2023	Weekday	2	08:00:00	1	1	12053	ROSSLYN HILL	2	3.15	0.47	35.38	565.5	390	0.062
2023	Weekday	2	08:00:00	1	1	12058	HAVERSTOCK HILL / POND STREET	3	7.77	1.26	42.20	569.85	393	0.074
2023	Weekday	2	08:00:00	1	1	12060	BELSIZE PARK STATION	4	20.55	3.18	60.24	574.2	396	0.10
2023	Weekday	2	08:15:00	1	1	BP2006	DOWNSIDE CRESCENT	5	14.63	4.69	71.55	552.45	381	0.13
2023	Weekday	2	08:00:00	1	1	34953	UPPER PARK ROAD	6	12.50	1.83	82.03	578.55	399	0.14
2023	Weekday	2	08:15:00	1	1	37435	HAVERSTOCK HILL / STEELE'S VILLAGE	7	14.47	2.82	96.07	582.9	402	0.16
2023	Weekday	2	07:45:00	1	1	37436	CHALK FARM STATION	8	23.40	4.99	107.09	591.6	408	0.18
2023	Weekday	2	07:45:00	1	1	11994	CHALK FARM ROAD / MORRISONS	9	22.85	3.15	124.20	582.9	402	0.21
2023	Weekday	2	08:00:00	1	1	33259	HAWLEY ROAD	10	21.37	1.80	145.99	582.9	402	0.25
2023	Weekday	2	08:00:00	1	1	33096	CAMDEN GARDENS	11	26.42	10.01	164.53	582.9	402	0.28
2023	Weekday	2	08:00:00	1	1	3517	CAMDEN TOWN STATION / CAMDEN STREET	12	27.73	7.72	189.29	595.95	411	0.31

MAX DEMAND HOUR BY ROUTE BY TIMEBAND – Data Definitions

Column Heading	Description
YEAR	Calendar year for which the data is extracted.
DAY_TYPE	Weekday, Saturday or Sunday.
TIMEBAND	Unique ID number from 1 to 17 that corresponds to a specific day type and time period. A lookup table is provided at the end of this guide.
QHr	Start of the rolling hour within the corresponding timeband for which the maximum departure load was estimated e.g., 06:45:00 represents the hour from 0645 to 0745 in which the highest onboard loads were estimated.
ROUTE	TfL route number.
DIRECTION	Direction of travel, indicated as either 1 or 2. Directions do not have a geographic attribute and can be different for routes travelling in the same direction on a road.
STOPCODE	The TfL stopcode as recorded in the Oyster and contactless system. Not to be confused with the NAPTAN code or SMS code that is provided on stop infrastructure.
STOPNAME	The stop name as recorded in the Oyster and contactless system.
STOPSEQUENCE	The stop sequence number in relation to the route and direction of travel as recorded in the iBus schedule.
HourBoardings	The average number of hourly boardings at the stop.
HourAlightings	The average number of estimated hourly alightings at the stop.
HourLoad	The average hourly departure load at the stop.
HourCapacity	The average total hourly capacity provided on departure from the stop. Note: this is based on the capacity of a bus when full and the vehicle model contracted to run on the corresponding route. The full vehicle capacity can range from 40 to 87 spaces.
HourSeats	The average number of seat spaces provided on departure from the stop over an hour period.
V/C	The average volume-to-capacity ratio, also referred to as the occupancy. Calculated as the load divided by the capacity.

EVERY JOURNEY MATTERS

Timeband Definitions

TIMEBAND	Definition
1	Weekday Early 0500-0700
2	Weekday AM Peak 0700-1000
3	Weekday Interpeak 1000-1600
4	Weekday PM Peak 1600-1900
5	Weekday Evening 1900-2200
6	Weekday Late Evening 2200-0000
7	Weekday Night 0000-0500
8	Saturday Morning 0500-1000
9	Saturday Midday 1000-1900
10	Saturday Evening 1900-2200
11	Saturday Late Evening 2200-0000
12	Saturday Night 0000-0500
13	Sunday Morning 0500-1000
14	Sunday Midday 1000-1900
15	Sunday Evening 1900-2200
16	Sunday Late Evening 2200-0000
17	Sunday Night 0000-0500