Victorian Integrated Survey of Travel and Activity (VISTA)

/ 2021 - 2026

Explanatory document

VICTORIA State Government | Department of Transport and Planning

Contents

1.0 Scope of this document

This document is intended to be a reference for users of survey data from the Victorian Integrated Survey of Travel and Activity (VISTA) administered from 2021 to 2026 (VISTA 2021-26). The Victorian Department of Transport and Planning (DTP) will be releasing de-identified unit record VISTA data by financial years where data is collected for the full year: 2022-23, 2023-24, 2024-25 and 2025-26. Partial data is available for FY 2021-22 upon request from DTP (Section 4.0)

This document's scope is **limited to VISTA data from the above listed periods (VISTA 2021-26).** Pre-COVD data (available as VISTA 2012-20) is largely similar; differences with the current data are covered in the following sections. Further information or guidance regarding VISTA 2012-20 can be sought by contacting DTP (Section 4.0).

2.0 What is VISTA

The Victorian Integrated Survey of Travel and Activity is a major data collection exercise conducted by DTP to understand all aspects of day-to-day travel by Victorians. VISTA is a rich dataset that investigates travel and activities people undertake as they go about their lives and informs policy, project and planning decisions made across the Victorian transport portfolio.

The collected data is appropriately weighted to represent travel and activity information and is presented at a Local Government Area (LGA) level. The weighting methodology has changed for VISTA 2021-26 and guidance on how to use weights is covered in Section 3.5.

3.0 What is different in VISTA 2021-26

This section highlights notable differences incorporated in the VISTA 2021-26 methodology.

3.1. Geographic Coverage

VISTA is conducted in the Greater Capital City (GCC) Region of Melbourne every year and a selection of regional centre(s) and / or precinct(s). Table 1 shows the coverage for VISTA 2021-26.

VISTA year	Coverage
FY 2022-23	GCC Melbourne, Bendigo SA4, Ballarat SA4
FY 2023-24	GCC Melbourne, Shepparton SA4
FY 2024-25	GCC Melbourne, Geelong SA4
FY 2025-26	GCC Melbourne, Latrobe – Gippsland SA4

Table 1 VISTA 2021-26 Geographic Coverage

Geographic coverage for pre-COVID VISTA data (VISTA 2012-20) in regional centres was by Local Government Areas (LGA). VISTA 2012-20 open data include GCC Melbourne and Geelong LGA. Care should be taken when comparing regional centres over time by taking boundary differences into consideration.

3.2. Data collection – introduction of GPS data collection

Historically VISTA data was collected using a travel diary – a booklet that every member of a selected household fills out enumerating all aspects of travel for a sampled travel day. For VISTA 2021-26, the detailed travel diary was replaced with a short memory jogger style trips log optionally accompanied by a GPS device. Where agreed, the GPS device is carried by survey participants 16 years and over.

Data collection using a GPS device has shown to improve accuracy and representation of travel patterns – for example data shows an increase in short trips which may have been considered insignificant to record in the travel diary previously.

Care should be taken when performing time-series analysis – observed changes in travel pattern can be attributed to a number of reasons including the natural unfolding of travel behaviour post-COVID and improved accuracy introduced by GPS data collection. Analysing point statistics in combination with confidence intervals (Section 3.5.2) is highly recommended.

3.3. Data Sets

The VISTA data set has four main dimensions discussed here; each dimension has its own unit record data.

3.3.1. Household data

Includes household type, composition, and location information. Data is collected at a Mesh Block level and aggregated to SA4 or LGA levels for consumption.

A VISTA 2021-26 Household data file now also includes:

1. Single household weight variable followed by a set of ten (10) jackknife group weight variables to calculate confidence intervals

3.3.2. Person data

Includes high level attributes of people living in the household – demographics, public transport usage attributes and vehicle usage attributes.

A VISTA 2021-26 Person data file now also includes:

1. Variables to capture work-from-home behaviour
2. Variables that capture vehicle travel attributes – toll, paid park and walk from parked vehicle
3. Single person weight variable followed by a set of ten (10) jackknife group weight variables to calculate confidence intervals
4. Variable to denote day type (weekday / weekend)
5. Variables to denote ABS geography – GCC Melbourne or regional SA4

3.3.3. Stop data

This data is the "core" travel data – information of all the travel taken by every person in the household on a specified travel day. This dataset enumerates every travel leg (stop) taken on the day except the ones that are outside Victoria.

A VISTA 2021-26 Stop data file now also includes:

1. Single Stop weight variable followed by a set of ten (10) jackknife group weight variables to calculate confidence intervals
2. Variable to denote day type (weekday / weekend)
3. Variables to denote ABS geography – GCC Melbourne or regional SA4
4. Variable to denote data entry type – 'Travel Diary', 'GPS' or 'both'. The value 'both' denotes that the final responses are based on travel form (referred to as Travel Diary) and attributes checked against or corrected using GPS data
5. Additional modes of travel - e-Scooter, Running/jogging

3.3.4. Trip data

This data set is a conceptual aggregation of Stops data. The legs are combined to make logical "Trips" – for example, a "work trip" may have multiple legs (stops) to change transport modes. For trips containing multiple trip legs (stops), the main mode of travel (*linkmode*) is assigned based on a mode hierarchy (Table 2).

Mode of travel	Hierarchy
Train	1
Tram	2
School Bus	3
Public Bus	4
Taxi / Rideshare Service	5
Motorcycle	6
Vehicle Driver	7
Vehicle Passenger	8
Bicycle	9
Walking	10
Plane	11
Mobility Scooter / e-Scooter / Running/jogging / Other	12

Table 2 Travel mode hierarchy to create main mode of trip

As an example, if the trip was made up of three trip legs: first leg taken in a private vehicle, the second leg by train and third leg covered by walking, then the *main mode* assigned will be 'Train'.

A VISTA 2021-26 Trip data file now also includes:

1. Single person weight variable followed by a set of ten (10) jackknife group weight variables to calculate confidence intervals
2. Variable to denote day type (weekday / weekend)
3. Variables to denote ABS geography – GCC Melbourne or regional SA4
4. Additional modes of travel - e-Scooter, Running/jogging

3.3.5. Journey to Work and Journey to Education

This data set is a conceptual aggregation of Stops data creating journeys – for journeys to work a journey represent the *last* exit from home and *first* arrival to a place of work- or work-related activity and for journey to education, a journey represents the *last* exit from home and *first* arrival to a place of education- or education-related activity.

VISTA 2021-26 Journey data is structured to include all facets of the journey.

The data file now includes:

1. Attribute of all intermediate legs (stops) of the journey – mode of transport, destination, distance and travel time
2. Attributes for the entire journey – travel time, elapsed time, distance, main mode
3. Variable to denote day type (weekday / weekend)
4. Variables to denote ABS geography – GCC Melbourne or regional SA4
5. Single journey weight variable

3.4. Weighting

VISTA 2021-26 employs a new sampling methodology where households are selected by mesh blocks and aggregate geography is constructed from this level. Consequently, the weighting methodology is multistage incorporating adjustments for non-responding households, geography, day of week, collection bias and recall bias.

Broadly the weighting process involves initialising household weights by accommodating non-responding households and then applying a multi-dimensional 'rake' to household data to match ABS[1] Census targets. This cascades to the Person data to initialise person weights where another round of multi-dimensional 'raking' is applied to match Census targets by person attributes.

The Person weights then initialise the Stop or Trip data weights which are further corrected to adjust for collection or recall bias.

The weighting process for VISTA 2021-26 is summarised in Figure 1.

[1] Australian Bureau of Statistics

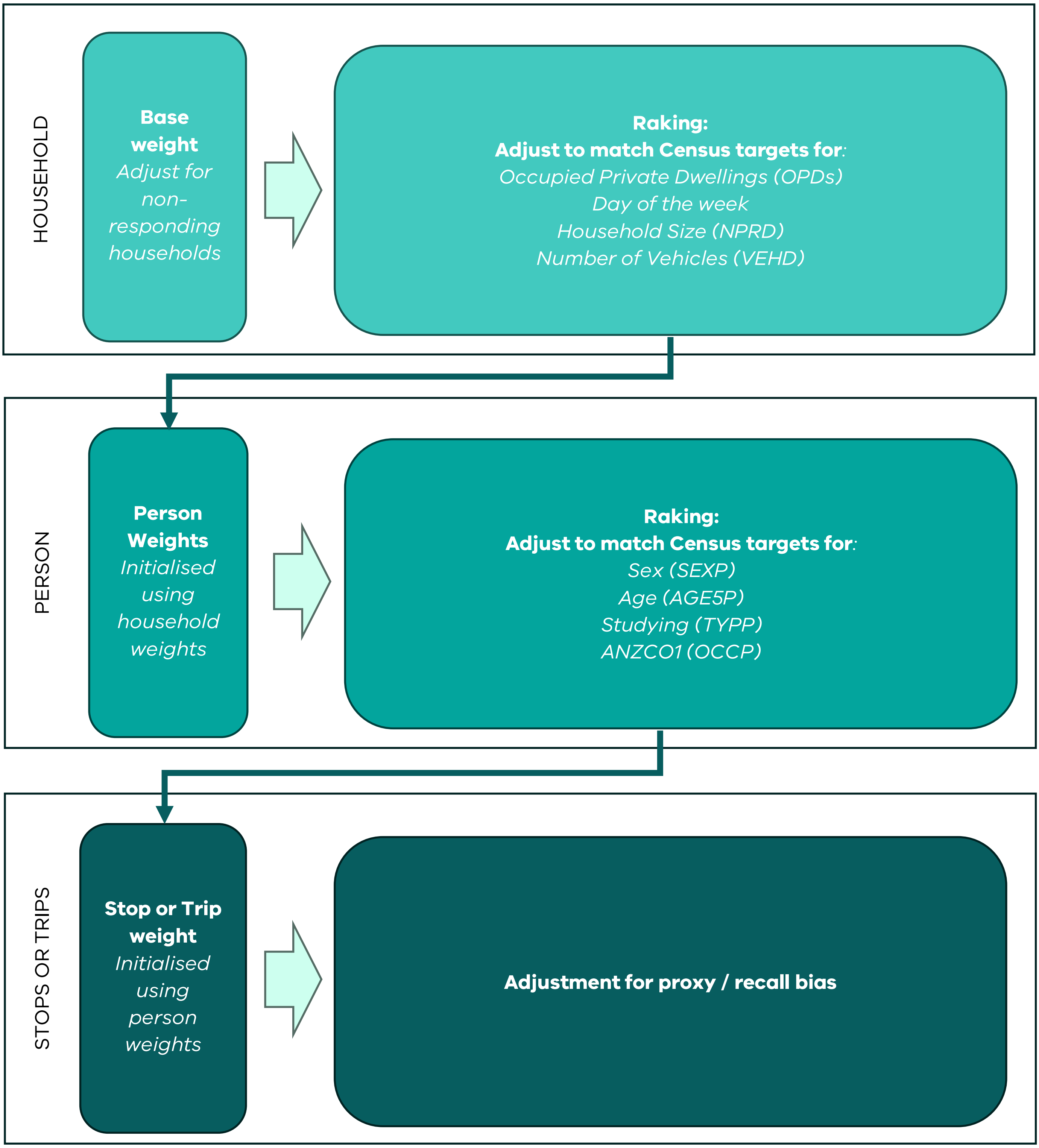

Figure 1 Overview of VISTA 2021-26 weighting process

3.5. How to use the weights

3.5.1. Adjustment for weekday vs weekend

VISTA 2021-26 employs an updated weighting methodology to adjust for responses received by day of the week. This means that weights included in each data file (mean weights and jackknife group weights) should be scaled to account for any day of week / day type filters applied while analysing data.

Weights included in each file are valid for an average day and can be aggregated as is if the dataset is not filtered by travel day of week / day type (weekday, weekend). If the data is analysed by specific day of week or day type, then the weights will need to be scaled by a factor of $\frac{7}{N}$ where N is the number of days in a week for which the data is being analysed for (filtered).

For example,

To analyse data for ***weekdays***, filter the records by *dayType = 'Weekday'* and scale the weights as follows:

$$\boldsymbol{weekday}\ weights = (supplied\ weights)\ \times \left[\frac{7}{5}\right]$$

To analyse data for ***weekends***, filter the records by *dayType = 'Weekend' and* scale the weights as follows:

$$\boldsymbol{weekend}\ weights = (supplied\ weights)\ \times \left[\frac{7}{2}\right]$$

To analyse data for any 3 days in a week, say Tuesday, Wednesday, and Thursday, filter the records by *travdow = 'Tuesday', 'Wednesday', 'Thursday'* and scale the weights as follows:

$$\boldsymbol{custom}\ weights = (supplied\ weights)\ \times \left[\frac{7}{3}\right]$$

This scaling should be applied to any data file included in the analysis.

If data is not filtered by day type or travel day of week, then no scaling is required. Below is the list of supplied weight variables by file type:

File type	Weight variable
Household file	*hhpoststratweight, hhpoststratweight_GROUP_##*
Person file	*perspoststratweight, perspoststratweight_GROUP_##*
Trips file	*trippoststratweight, trippoststratweight_GROUP_##*
Stops file	*stoppoststratweight, stoppoststratweight_GROUP_##*
Journey to work/education	*journey_weights*

3.5.2. Calculating confidence intervals (example)

Confidence intervals illustrate the range of values that the mean value is likely to be in.

Below is an example to calculate 95% confidence intervals using Jackknife weights for a given statistic. The method uses Logit CI useful for calculated intervals for proportions bounded between 0 and 1. This is a multi-step calculation as follows:

1. Calculate proportions for statistic of choice $\hat{P}$

2. Create logit transformed statistic λ of $\hat{P}$ and λ(r) for proportions $\hat{P}$(r) using Jackknife weights:

$$\lambda = LN\left(\frac{\hat{P}}{1-\hat{P}}\right)$$

3. Variance of λ using logit transformed proportions of mean value and Jackknife weights:

$$\mathrm{var}(\lambda) = \frac{R-1}{R}\sum_{r=1}^{R}\left(\lambda_{(r)} - \lambda\right)^2$$

λ	Logit transformed proportion of Statistic of choice
$\lambda_{(r)}$	Logit transformed proportion of Jackknife weight group r
R	Number of groups (10)

4. Standard error (SE) of λ:

$$\mathrm{SE} = \sqrt{\mathrm{var}(\lambda)}$$

5. Logit confidence interval (CL – 95% confidence):

$$CL = \lambda \pm 2.26 \times SE$$

6. Final confidence interval (Anti log transformation CL – 95% confidence):

$$C = \frac{e^{CL}}{1+e^{CL}}$$

Below is a working example of calculating confidence intervals for statistic of interest (S) – number of trips by mode of transport on an average *weekday*. The figures below are for illustrative purposes only and do not use real data.

i. Calculate statistic S by summing weights and applying the scaling factor (Section 3.5.1) where required:

$$\boldsymbol{S} = (tripposts tratweight) \times \left[\frac{7}{5}\right]$$

$$\boldsymbol{S}_{\mathbf{r}} = (trippoststratweight\_GROUP\_\mathrm{r}) \times \left[\frac{7}{5}\right]$$

	$\boldsymbol{S}$	$\boldsymbol{S_1}$	$\boldsymbol{S_2}$	$\boldsymbol{S_3}$	$\boldsymbol{S_4}$	$\boldsymbol{S_5}$	$\boldsymbol{S_6}$	$\boldsymbol{S_7}$	$\boldsymbol{S_8}$	$\boldsymbol{S_9}$	$\boldsymbol{S_{10}}$
Mode	Trips	GRP_1	GRP_2	GRP_3	GRP_4	GRP_5	GRP_6	GRP_7	GRP_8	GRP_9	GRP_10
Private Vehicle	7,699	7,595	7,818	7,862	7,640	7,707	7,718	7,746	7,647	7,682	7,704
Public Transport	598	590	615	583	592	606	596	615	606	573	585
Active Transport	1,658	1,635	1,643	1,644	1,669	1,630	1,681	1,648	1,704	1,687	1,647
Other	45	45	42	41	43	47	51	50	46	43	43

i. Calculate mean proportions $\hat{P}$ and group proportions $\hat{P}$(r) using Jackknife weights for statistic S:

	$\hat{P}$	$\hat{P}_1$	$\hat{P}_2$	$\hat{P}_3$	$\hat{P}_4$	$\hat{P}_5$	$\hat{P}_6$	$\hat{P}_7$	$\hat{P}_8$	$\hat{P}_9$	$\hat{P}_{10}$
Mode	Trips	GRP_1	GRP_2	GRP_3	GRP_4	GRP_5	GRP_6	GRP_7	GRP_8	GRP_9	GRP_10
Private Vehicle	77.0%	77.0%	77.3%	77.6%	76.8%	77.1%	76.8%	77.0%	76.4%	76.9%	77.2%
Public Transport	6%	6%	6%	6%	6%	6%	6%	6%	6%	6%	6%
Active Transport	17%	17%	16%	16%	17%	16%	17%	16%	17%	17%	17%
Other	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%

ii. Create logit transformed statistic λ of $\hat{P}$ and λ(r) for proportions $\hat{P}$(r) using Jackknife weights:

	$\boldsymbol{\lambda}$	$\boldsymbol{\lambda_1}$	$\boldsymbol{\lambda_2}$	$\boldsymbol{\lambda_3}$	$\boldsymbol{\lambda_4}$	$\boldsymbol{\lambda_5}$	$\boldsymbol{\lambda_6}$	$\boldsymbol{\lambda_7}$	$\boldsymbol{\lambda_8}$	$\boldsymbol{\lambda_9}$	$\boldsymbol{\lambda_{10}}$
Mode	Trips	GRP_1	GRP_2	GRP_3	GRP_4	GRP_5	GRP_6	GRP_7	GRP_8	GRP_9	GRP_10
Private Vehicle	1.21	1.21	1.22	1.24	1.20	1.22	1.20	1.21	1.18	1.20	1.22
Public Transport	-2.76	-2.76	-2.74	-2.80	-2.76	-2.74	-2.76	-2.73	-2.74	-2.80	-2.78
Active Transport	-1.62	-1.62	-1.64	-1.64	-1.60	-1.63	-1.60	-1.63	-1.58	-1.59	-1.62
Other	-5.40	-5.40	-5.47	-5.50	-5.43	-5.35	-5.28	-5.31	-5.37	-5.45	-5.44

iii. Variance of λ using logit transformed proportions of mean value and Jackknife weights:

a. Calculate $(\lambda_{(r)} - \lambda)^2$ to calculate variance where $\lambda_{(r)}$ is statistic repeated for individual group (ten in the data files):

(Figures shown below converted to percentage for illustration)

Mode	λ	$(\lambda_1 - \lambda)^2$	$(\lambda_2 - \lambda)^2$	$(\lambda_3 - \lambda)^2$	$(\lambda_4 - \lambda)^2$	$(\lambda_5 - \lambda)^2$	$(\lambda_6 - \lambda)^2$	$(\lambda_7 - \lambda)^2$	$(\lambda_8 - \lambda)^2$	$(\lambda_9 - \lambda)^2$	$(\lambda_{10} - \lambda)^2$
Private Vehicle	120.8%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%
Public Transport	-275.6%	0.0%	0.0%	0.2%	0.0%	0.0%	0.0%	0.1%	0.0%	0.2%	0.0%
Active Transport	-161.6%	0.0%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.0%
Other	-539.7%	0.0%	0.6%	1.0%	0.1%	0.2%	1.3%	0.8%	0.1%	0.3%	0.2%

b. Calculate Variance of λ:

Mode	$\mathrm{var}(\lambda) = \frac{R-1}{R} \sum_{r=1}^{R} (\lambda_{(r)} - \lambda)^2$
Private Vehicle	0.3%
Public Transport	0.5%
Active Transport	0.3%
Other	4.2%

iv. Standard error (SE) of λ and Logit confidence intervals (L and H):

Mode	$SE = \sqrt{var(\lambda)}$	$L = \lambda - 2.26 \times CL$	$H = \lambda + 2.26 \times CL$
Private Vehicle	5.0%	109.4%	132.2%
Public Transport	6.9%	-291.2%	-260.0%
Active Transport	5.8%	-174.8%	-148.4%
Other	20.4%	-585.8%	-493.6%

v. Final confidence intervals (Anti log transformation of L and H – 95% confidence):

Mode	$\hat{P}$	$Low = \frac{e^L}{1+e^L}$	$High = \frac{e^H}{1+e^H}$
Private Vehicle	77.0%	74.9%	78.9%
Public Transport	6.0%	5.2%	6.9%
Active Transport	16.6%	14.8%	18.5%
Other	0.5%	0.3%	0.7%

There are other methods of calculating confidence intervals not covered here that may be more appropriate for the situation. For example, Log-normal confidence intervals are calculated when the sample size is small which may push confidence intervals into negative values or t-distribution confidence intervals are used for variables where the distribution is assumed to be normal.

The calculation of t-distribution confidence intervals is like the illustrated example above without the log transformation:

1. Variance using Jackknife weights:

$$\mathrm{var}(S) = \frac{R-1}{R}\sum_{r=1}^{R}(S_r - S)^2$$

S Statistic of choice
S_r Statistic repeated for individual groups (ten in the data files)
R Number of groups (10)

2. Standard error (SE) for statistic of choice (S) from variance:

$$SE(S) = \sqrt{\mathrm{var}(S)}$$

3. Confidence intervals (95% confidence):

$$confidence\ interval = \left(S \pm 2.26 * SE(S)\right)$$

3.6. Privacy compliance

VISTA 2021-26 data comply with DTP's updated privacy guidelines. DTP has conducted a thorough review of the privacy implications of releasing unit record data and taken several measures to de-identify data prior to release. As a result, attributes deemed high risk are not released as part of open data.

In other instances, high risk variables have been modified to ensure privacy: a small number of survey participants indicated that they did not identify as either *Male or Female*. Releasing data with *sex* variable as collected was classified as high risk of reidentification. As a result, for these responses the *sex* variable was randomly allocated to *Male or Female* to ensure participant privacy.

Researchers or organisations requesting certain demographic or geographic attributes that are not available via open data will be required to enter into a Data Sharing Agreement (DSA) with DTP that is designed to protect all parties from any instance of privacy breach. All requests will be considered on a case-by-case basis and DTP reserves the right to decline any request for attributes that are deemed high risk.

4.0 Contact

VISTA data is available in the public domain which includes an online interactive dashboard at:

https://opendata.transport.vic.gov.au/

The Department of Transport and Planning VISTA team can be contacted via:

vista@transport.vic.gov.au

VICTORIA State Government

Department of Transport and Planning