Estimation of travel time savings for toll road customers

[Outline of presentation]

Introduction

Research[1] commissioned by Transurban revealed that only 26% of customers see value in tolls. Through this and other surveys[2], it was also revealed that customers would value information and metrics that would help demonstrate the value of using toll roads. Transurban has provided toll calculators that show the time savings and cost of a single trip for many years, however, we recently chose to actively communicate the time savings that customers are receiving proactively.

Whilst Transurban has detailed information on the number and location of trips on the toll road network, we have no visibility of the true origin and ultimate destination of these trips. Customers also value privacy and are not willing to trade detailed geolocation data to enable the calculation of personalised travel time savings metrics.

To overcome this, the customer insights and transport modelling teams at Transurban developed an innovative approach that estimates the travel time savings for customers based on fusing data from our toll roads with TomTom network-wide traffic statistics. The result of this work is currently being reported to customers in all three Australian regions that we operate in and has been well received.

Whilst this paper documents the estimation of travel time savings, the framework is highly extensible to enable the calculation and reporting of other travel metrics. In addition, a key part of this process can also be easily amended to provide information that would be of further interest to transport planners, modellers and demographers.

[1] Survey of 400 Linkt Customers, May 2023, Bastion, McKinsey Community Survey

[2] Brand Health Report 2023, Community Engagement Report 2023, Community Engagement Report 2022, Rewards & Recognition customer research, Voice of Customer program

SOME SUPPORTING MATERIAL (NOT YET COMPLETED)

1. Background

// information on why we set out to do this

2. General methodology

The Transurban customer and asset trip databases contain information on the true number of gantry-gantry movements and/or passages of toll road users. This information is adequate for the purposes of charging tolls but does not provide additional insights such as the catchment areas that the trips are sourced from or where they are ultimately heading to once leaving the toll road.

Through the TomTom Move platform, and specifically the O/D Analysis product, we can undertake selected link analysis that provides a sample number of the true origin and destination of users that use a particular section of the road network. Undertaking this analysis for every toll point in the tolled network and processing it in the right way will allow us to understand the distribution of toll road customers.

Using the weighted start and end points for customer trips in combination with TomTom Traffic Insights data allows the capturing of key metrics for the traversal of the road network between the origin and destination via the fastest possible tolled route and the fastest possible non-tolled route. In this way, we can estimate and report on the differences in the key metrics for what toll road users would have experienced at the time of travel.

This process is highly reliant on data from TomTom subscription-based services which collects historic traffic information from anonymous GPS data from various sources including data from personal navigation devices, in-dash systems, and mobile applications. The data is then processed by TomTom to provide insights on road speeds, travel times, and traffic density, ensuring a comprehensive view of traffic conditions while maintaining user privacy.

The full process to estimate customer travel time savings can therefore be broken down into the following components:

(1) Customer data network trip model (see Section 3);
(2) TomTom origin-destination analysis (see Section 4);
(3) TomTom network-wide traffic statistics (see Section 5); and,
(4) Computation of savings (see Section 6).

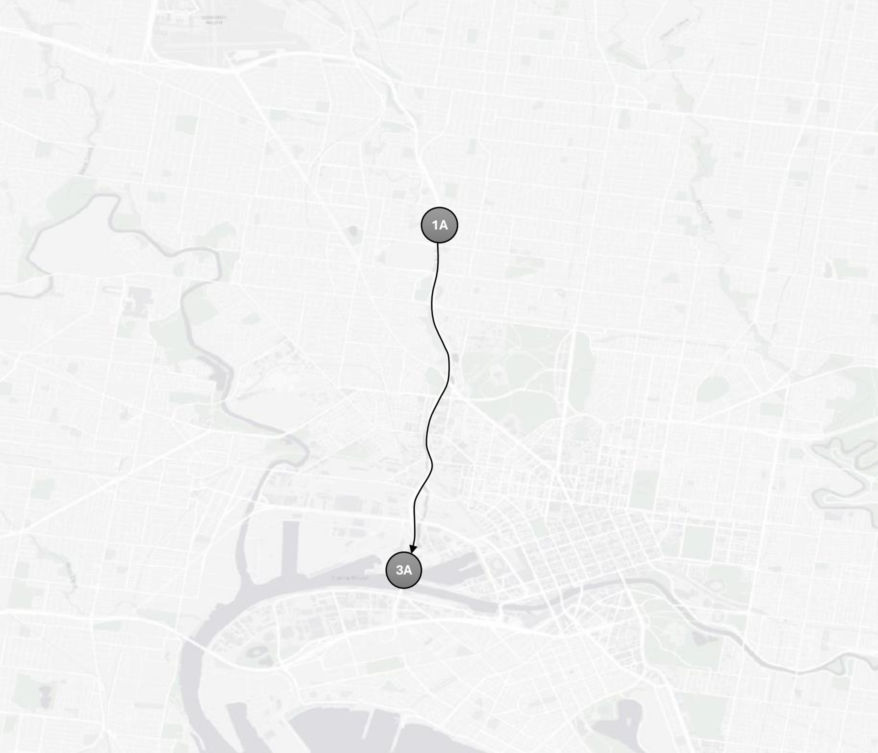
An overview of the methodology along with ownership of each component is given in Table 1.

Table 1 – Methodology overview

Component	Description
Network trip model	Converts asset-specific trip record data into a *first gantry* to *last gantry* logical trip irrespective of the combination or umber of assets used.
Origin and destination analysis	Estimates the distribution of customer trips *to the first gantry (origin)* and *from the last gantry (destination)* used as part of a logical trip for every gantry in the toll network.
Network-wide traffic statistics	Reports on the key metrics (e.g. travel time) for every *origin* and *destination* combination that uses a tolled route along with the next-best free-route alternative.
Computation of savings	Uses the data generated from the network trip model, origin and destination analysis and reporting of network-wide traffic statistics to compute the savings for each logical trip combination.

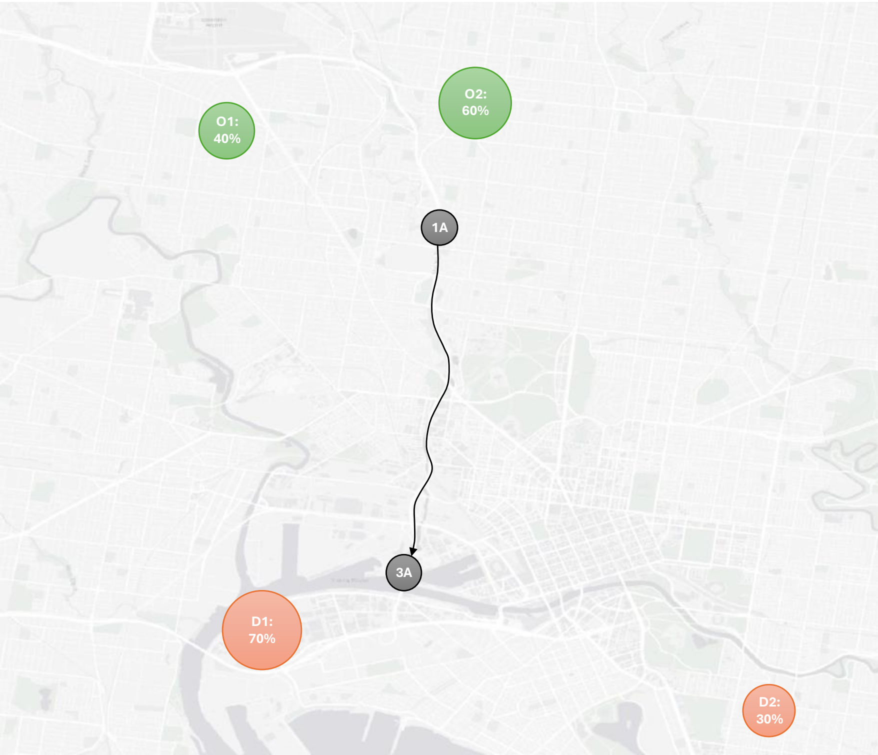
A step-by-step representation of this process is also given in Figure 1
.

Figure 1 – Example of the full process



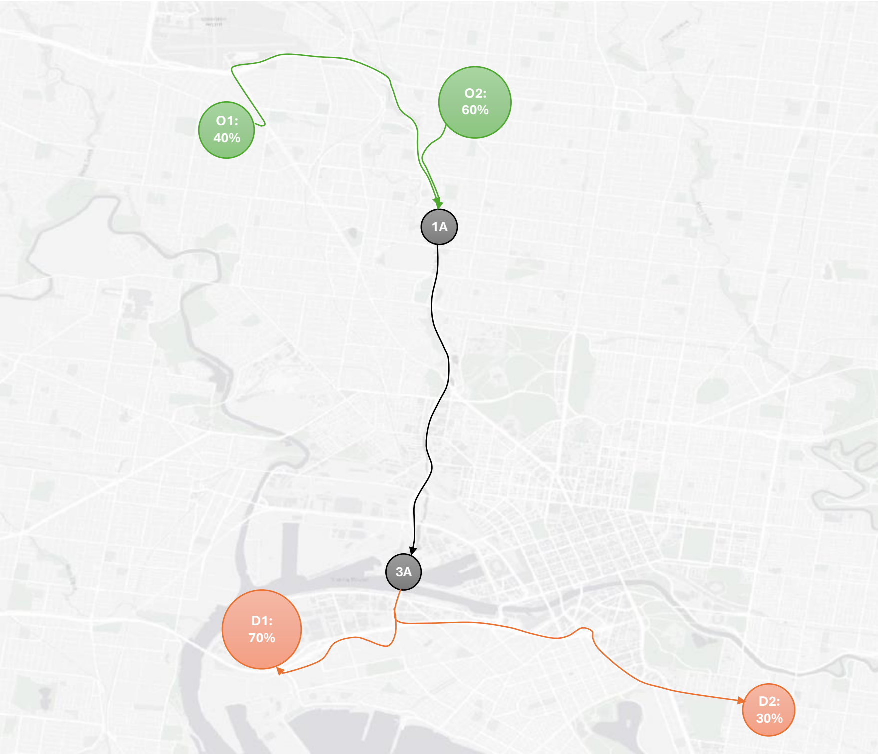
The connected trip model produces a set of all possible gantry-gantry trip combinations for a given date range and time period.

In this example, we assume toll point 1A to toll point 3A trips.



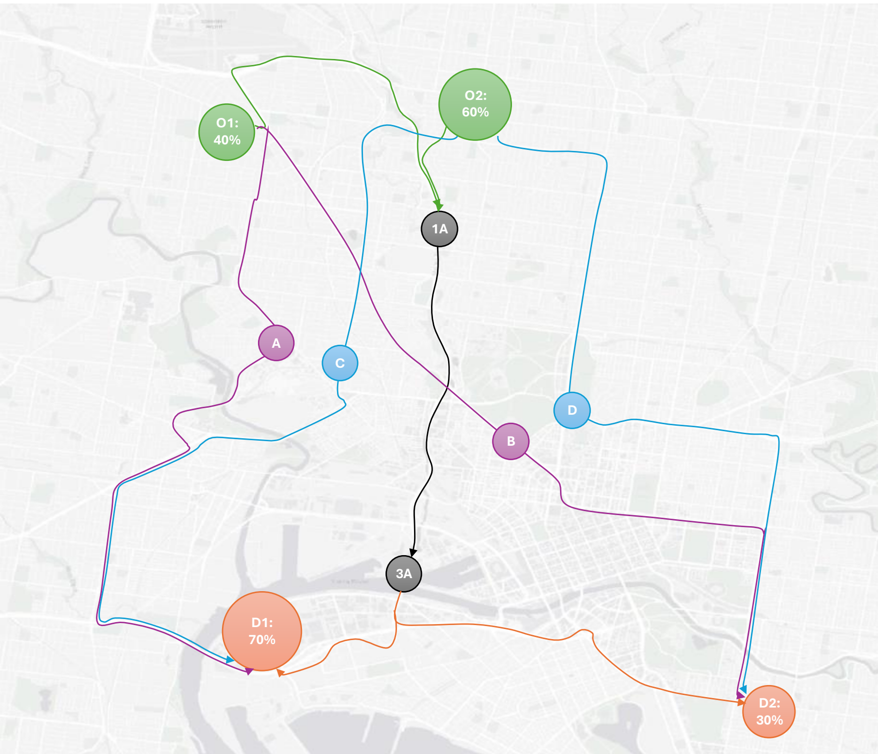
From the TomTom origin and destination analysis we obtain a set of origin weights for areas that trips going to toll point 1A emanated from and a set of destination weights for areas that trips leaving toll point 3A are destined for.

In this example, we assume that the tolled trips starting at 1A came from two zones (O1 = 40% and O2 = 60%) and the trips from toll point 3A were destined for two other zones (D1 = 70% and D2 = 30%).



The TomTom network-wide traffic statistics allows the extraction of sampled average travel times between the origin and destination zones via the tolled routes based on the fastest possible travel time path.

Orig	Dest	Time (mins)
1	1	20
2	1	16
1	2	38
2	2	33



And via the next-best non-tolled routes based on the fastest possible travel time path..

Orig	Dest	Time (mins)
1	1	35 (via A)
2	1	40 (via B)
1	2	48 (via C)
2	2	42 (via D)

Zone		Proportion *(P)*			Travel Time in Minutes *(T)*		
Orig *(i)*	Dest *(j)*	Orig *(i)*	Dest *(j)*	Orig x Dest	Tolled	Non-tolled	Saving
1	1	0.4	0.7	0.28	20	35	15
2	1	0.6	0.7	0.42	16	40	24
1	2	0.4	0.3	0.12	38	48	10
2	2	0.6	0.3	0.18	33	42	9

Savings between the tolled and non-tolled routes can then be computed and weighted based on the origin and destination weightings of the trips. The weighted average saving can then be reported for any toll point 1A to toll point 3A trips.

$$WTTS = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left(P_i \cdot P_j \cdot \Delta T_{ij}\right)}{\sum_{i=1}^{n}\sum_{j=1}^{n}\left(P_i \cdot P_j\right)} = 17.1$$

In this example, the weighted average saving will be reported as 17.1 minutes.

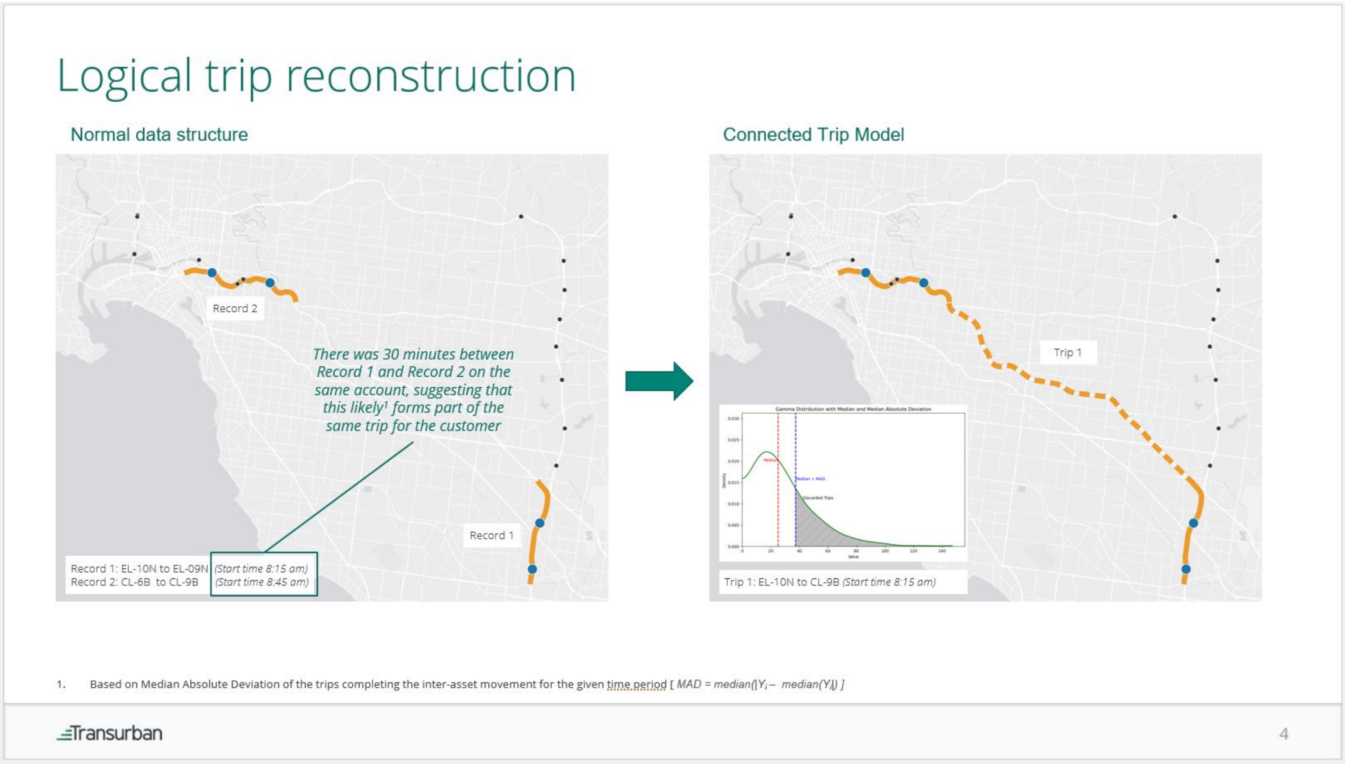
This process is repeated for every gantry-gantry combination from the network trip model.

3. Connected trip model

// summary here

3.1 Processing steps

// a little more info here



4. Origin and destination analysis

The TomTom Move OD (Origin-Destination) Select Link product is a powerful subscription-based tool designed for traffic analysis and planning. It allows users to perform detailed analyses of traffic flow and trip dynamics by focusing on specific road segments or ‘links’. The tool is particularly useful for traffic planners looking to understand and manage observed traffic patterns more effectively.

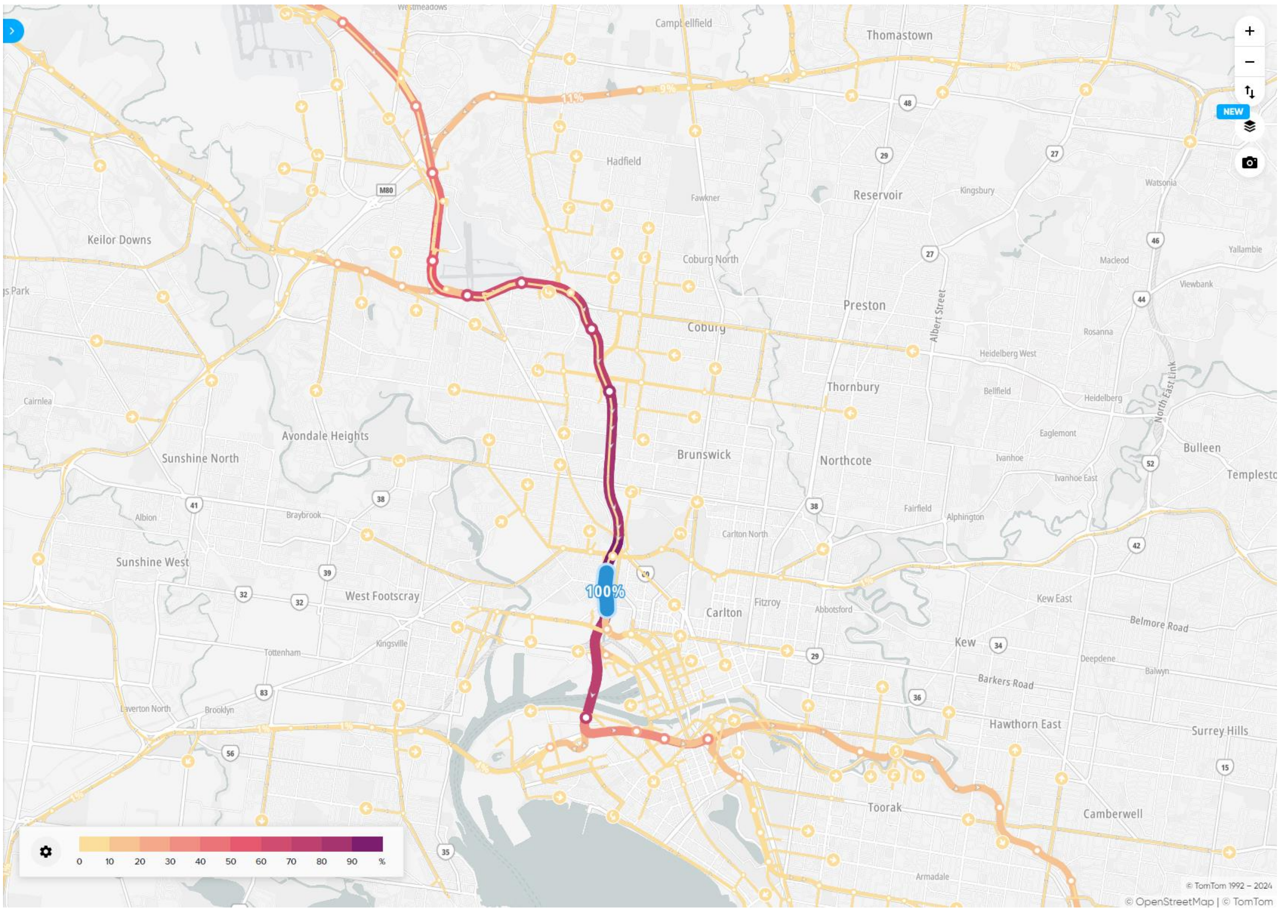
The tool collects data from various sources, including passenger vehicles, fleet management vehicles, and other traffic data sources and includes information about trip origins, destinations, and routes taken. The analysis can be customized with various parameters, including the pertinent ones below:

- Date Ranges: Define the date ranges for which the analysis should be conducted (e.g. weekdays in November)
- Time Ranges: Specify the time ranges for the analysis (e.g. between 8am and 9am)
- Buffer Radius: Set the radius around the selected link to include in the analysis (maximum of 50km).

Whilst TomTom does enable customization by ‘fleet’ or ‘passenger’ vehicles, at present, it is not possible to specify the class of the vehicles as a segmentation measure. The data that is extracted therefore represents all vehicle classes (See Table 2 – Key parameters ‘adj_factors’ for further details on the Heavy Vehicles factor assumptions applied for this work).

An example output report from the tool can be seen in Figure 2. The example shows the relative intensities of the source locations/routes of trips coming into toll point 2A and the relative intensities of the destination locations/routes of trips leaving toll point 2A for a particular date and time range.

Figure 2 – Example select link output at toll point 2A on CityLink from the TomTom OD application



The report outputs are broken down into incoming (coming to the link of interest) and outgoing (going away from the link of interest) proportions and by the date range and time range segmentations requested.

It is important to note that the analysis does not report on the tracking of individual users from true origin to true destination. The selected link location (in this instance toll point 2A) acts like a 'washing machine' jumbling up the trips coming into toll point 2A before distributing to ultimate destination. However, the overall weightings of the origins and destination still hold true.

4.1 Processing steps

The origin and destination analysis can be broken down into three main components:

(1) Running, accepting and extracting TomTom data (see Section 4.1.1);
(2) Pre-processing TomTom data (see Section 4.1.2); and,
(3) Zonal allocation of TomTom data (see Section 4.1.3).

All components are run in R Studio using bespoke scripted operations to maximise processing efficiency.

4.1.1 Running, accepting and extracting TomTom data

Running, accepting and extracting TomTom data can either be done manually via the TomTom Move platform or can be done via an API. Given the large number of reports required to be run and processed (201 toll point locations) and the large number of segments[3] that need to be extracted (between 17,000 and 25,000 depending on the month) it is recommended that the API is used with manual checking via the portal.

4.1.2 Pre-processing TomTom data

As shown in Figure 2, the TomTom data for a given data segmentation captures all of the sampled trips that pass through the specified toll point. This includes both trips that only use the specified toll point (in our example toll point 2A) and also trips that use multiple toll points (e.g. toll point 1A to 6A uses 5 different toll points).

Pre-processing the TomTom data prior to application requires the removal of the information that passes through multiple toll points and only keeping that information that either used the nominated toll point as a first ‘on’ point to the tolled network or the last ‘off’ point to the tolled network. In addition to this, any TomTom data with Functional Road Class (FRC) equal to zero (a hierarchical measure of the road type which in this instance represents motorways/highways/freeways) is also removed (the reasons for this are covered in Section 4.1.3).

An example of the outcome from pre-processing the TomTom data at toll point 2A is given in Figure 3 and shows:

- The location of toll point 2A with an orange marker;
- The location of all other CityLink toll points with a blue marker;
- The original TomTom data in red; and,
- The pre-processed TomTom data in green – all data passing through other toll points and situated on motorways/highways/freeways has been removed

[3] Calculated by [NUMBER OF TOLL POINTS] * [NUMBER OF DIRECTIONS] * [NUMBER OF DATE RANGES] * [NUMBER OF TIME RANGES]

Figure 3 – Example of outcome of pre-processing (red=original; green=pre-processed)

Repeating this process at every toll point in the tolled network will leave us with the catchment areas for trips entering the toll system ('incoming' catchment) and leaving the toll system ('outgoing' catchment) for every toll point location.

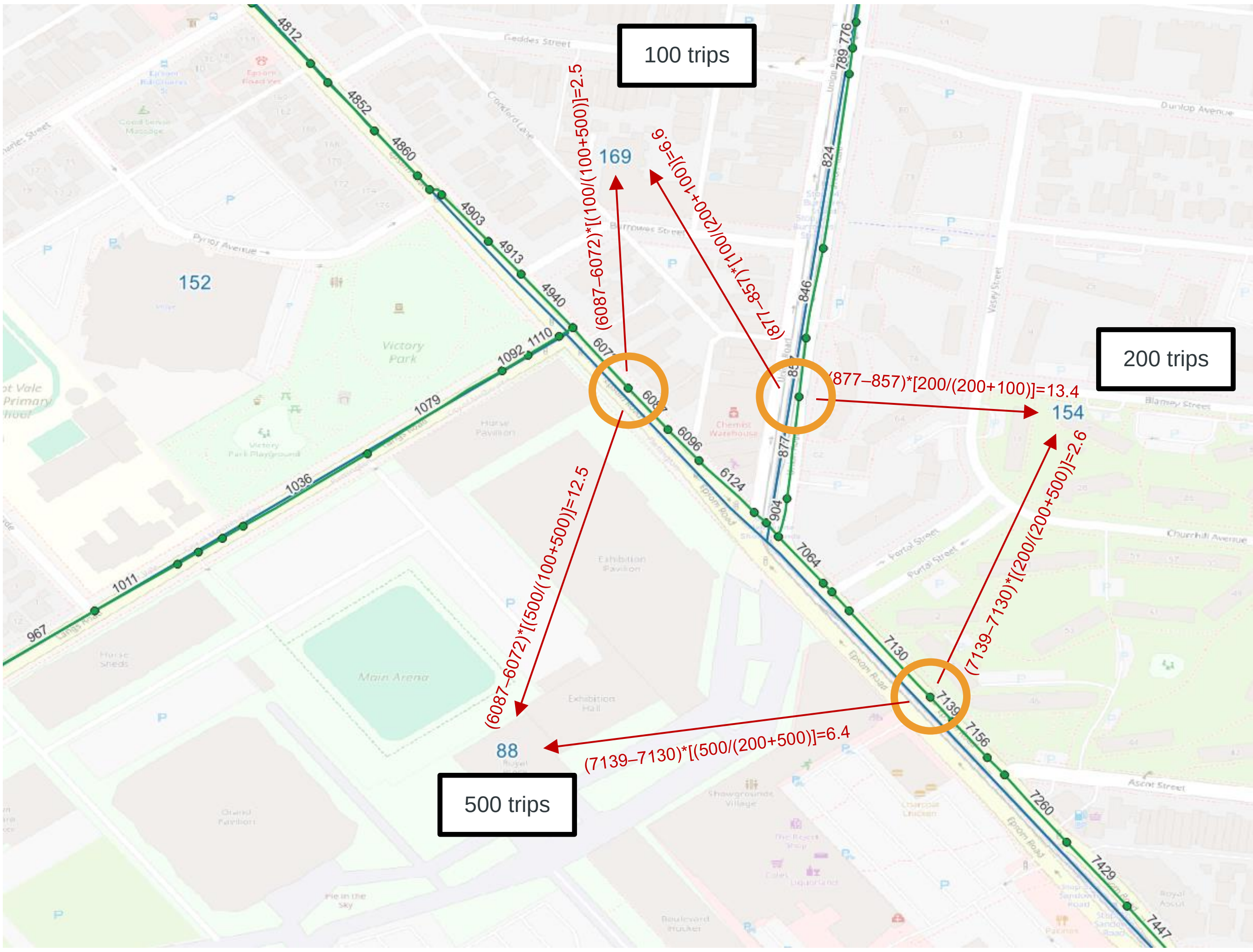
4.1.3 Zonal allocation of TomTom data

Zonal allocation is concerned with transferring the differences in the sampled number of users between each subsequent TomTom link to the underlying area or zone where this difference occurs +/- a 50m buffer. An example of this process is shown in Figure 4, with pertinent details given below:

- The zone areas are represented by blue lines and the zone numbers labelled in blue
- The TomTom pre-processed link samples are shown with green lines – the green markers denote the start/end point of each link.
- The number next to each link represents the number of sampled TomTom users for the reported data stratification.
- The orange circle shows the 50m buffer at each link start/end point. Sampled trips will be allocated to the zones that fall within the buffer.
- The sampled trip difference between each subsequent link is allocated based on the weighting of the modelled number of vehicle trips for the zones in the buffer (taken from strategic models).
- Red text notation and arrows shows the calculations, and which zones the sampled trips are allocated to.

Whilst only three examples are marked out on the Figure, it is important to note that this process occurs at every boundary between subsequent links for every sampled location and reported data stratification.

Figure 4 – Zonal allocation detail



Whilst the approach yields reasonable results, it does have limitations:

a) The accuracy of both the alignment of the zonal boundaries and the TomTom GPS coordinates that sometimes TomTom samples may be attributed to a zone on one side of the road versus another. For this reason, smaller geographic boundaries work better than larger ones, however there is a trade-off to be aware of as smaller zoning systems will increase processing times and data storage requirements.

b) The approach assumes that every difference in the number of sampled trips on subsequent TomTom links is due to travellers terminating their trip. This is not always the case. In some instances, a logical drop in sampled trips should not be possible (i.e. there is no driveway or side street) but they still occur. This is due to accuracy limitations of the GPS tracking from the TomTom service. This is the main reason we remove the FRC=0 links (motorways/highways/freeways) from the dataset as vehicles should not be able to terminate their trip on these types of roads.

Once the TomTom samples have been allocated to zones, there are two additional processes that are run to ensure the sample data aligns with the customer data and application methodology:

1) Application of class factors (see Table 2 – Key parameters 'adj_factors' for further details)
2) Normalisation – due to removing parts of the TomTom sample (see Section 4.1.2) the remaining data cannot be used directly and must be normalised. This is done by dividing the sample in each zone by the sum of the samples in every zone for the data segmentation at the location of interest. This approach ensures that for each data segmentation the proportions add up to 100% whilst still maintaining the relative pattern of interest.

The resultant data set allows for the proportional distribution of light and heavy classified vehicle trips in the customer data base that is from a given toll point to another toll point from/to the zones in its origin and destination catchment areas respectively.

As described in Section 0, whilst this is not the true OD for a single user, it gives a representation of the pattern of trips that, when combined with a skim of tolled and non-tolled path times from the network wide traffic statistics process, can be used to proportionally apply as the origin and destination points to calculate weighted metrics such as travel time or fuel use savings.

4.1.4 Key parameters

Required parameters are set at the start of each key stage of the process. Not all parameters are required at each stage, however, a description of all parameters are provided below to allow the user to understand the purpose and allowable values of each parameter.

Table 2 – Key parameters

Variable	Description
region	Sets the region that the data is being processed for. Possible values are: - "Melbourne" - "Sydney" - "Brisbane" This is important as some input files include the region name and hence enables the correct input to be selected for the region. In addition, it also works as a filter for some data inputs (e.g. the toll gantry GIS file) to minimise running time during computation.
dateRange	Must match the settings in the source .json files that are passed to the TomTom API for processing. Default setting is: - 0 (weekdays excl. school and public holidays – WD) - 1 (weekend days including public holidays – WEPH) - 2 (weekdays during school holidays – SH) A maximum of 4 date ranges are allowed.
timeRange	Must match the settings in the source .json files that are passed to the TomTom API for processing. Default setting is: - 0 (0000-0400 hours) - 1 (0400-0500 hour) - 2 (0500-0600 hour) - 3 (0600-0700 hour) - 4 (0700-0800 hour) - 5 (0800-0900 hour) - 6 (0900-1000 hour) - 7 (1000-1100 hour) - 8 (1100-1200 hour) - 9 (1200-1300 hour) - 10 (1300-1400 hour) - 11 (1400-1500 hour) - 12 (1500-1600 hour) - 13 (1600-1700 hour) - 14 (1700-1800 hour) - 15 (1800-1900 hour) - 16 (1900-2000 hour) - 17 (2000-2100 hour) - 18 (2100-2200 hour) - 19 (2200-2300 hour) - 20 (2300-0000 hour)

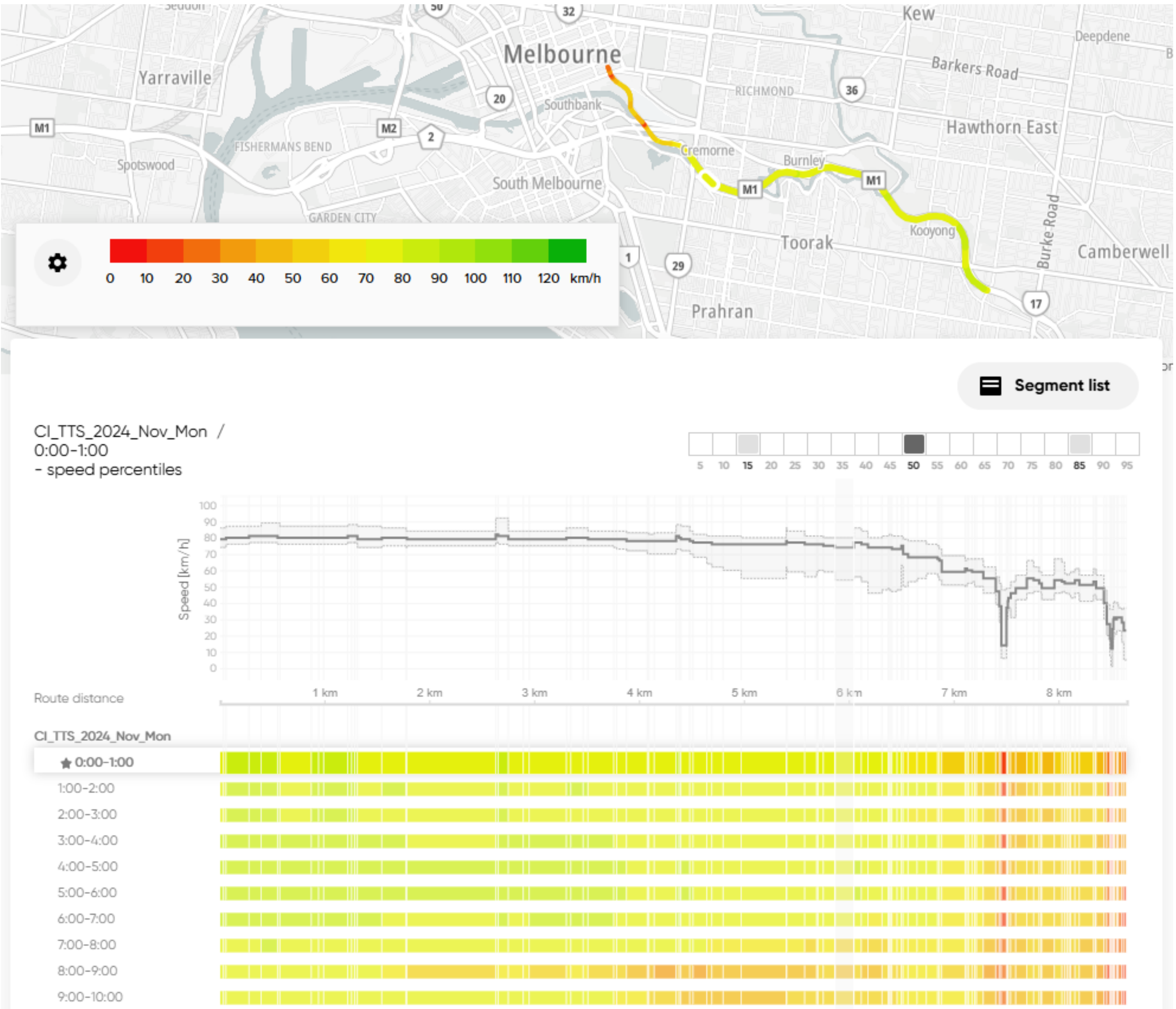
Variable	Description
	A maximum of 24 time ranges are allowed.
zonal_boundary_name	The name of the GIS file containing the spatial boundary information. There must be a file available for the region that is being run, however, the scripting selects the required file based on the condition of the 'region' variable. Defaults are as follows: - Melbourne (vic_taz_3236_v4.0.shp) - Sydney (nsw_taz_2562_v5.0_wSA2.shp) - Brisbane (qld_taz_2468_v4.0_wSA2.shp) The default files are consistent with the zoning system and numbering that is used by the strategic modelling team, however, in some instances this was expanded to capture additional detail. The TomTom select link data extends 50km to/from the point that it is being collected. In some instances, this is beyond the strategic modelling region boundaries. Coverage for the Sydney and Brisbane regions has therefore been extended to capture all possible areas within a 50km distance of the point that the data is being collected. Any new boundary system will therefore need to ensure that coverage is consistent with the default files listed above.
toll_gantry_file	The name of the GIS file containing the polygons representing the toll gantry locations. As a minimum, this file must contain four attributes as follows: - 'ASSET_NAME' (must match the asset names in the Snowflake customer database) - 'TOLL_POINT' (must match the toll point names in the Snowflake customer database) - 'REGION' (must match the 'region' variable possible values) - 'INCLUDE' (Boolean indicator whether to include the toll point in the matching process)
unique_combos	File containing a concatenation of all 'ASSET_NAME' & 'TOLL_POINT' combinations. Note that the 'ASSET_NAME' and 'TOLL_POINT' values must match those in the `toll_gantry_file` GIS file. This file is stored under `.\project_execution\data` and is unlikely to change unless new toll points are added to the network.
adj_factors	File for each region containing the modelled information originating from or destined to each geographic zone by modelled time period (see `time_periods` variable below). Allocation of TomTom link data to zones is initially based on the relative weights of the modelled number of all class vehicle trips for the zones within a 50m catchment of the sampled trip change. The accumulated all class samples in each zone are then broken down into Light Vehicle (LV) and Heavy Vehicle (HV) proportions using the modelled class factors. The adjusted values are then normalised to ensure the sample is kept whole.
time_periods	The sampled TomTom data has 21 time periods segments (see `timeRange` variable) whilst the strategic model where we obtain the HV proportions only has 7. The `time_periods` file contains the equivalence between the TomTom time periods and the modelled time periods to enable the application of the `HV_factors`. A modelled time period as defined in the `HV_factors` file must be defined for every `timeRange` specified.
multisource_grp	Due to complexities in the Sydney region (e.g. in tunnels and/or where there are multiple feeder routes to the same toll point) some locations require multiple select link extractions to capture the full distribution of traffic to/from the one toll point. The `multisource_grp` input file specifies which multiple TomTom reports need to be merged into one. The file contains the following fields that are all required: - ReportName (full list of report names that were run for the region) - Group (if grouping number is the same, the TomTom report names for that group will be merged into one) - NewName (for reports that are going to be grouped, a new name needs to be specified) - ResetJobID (for reports that are going to be grouped, some data e.g. JobID is no longer relevant and hence gets flagged to be overwritten) This file required editing every time the Syndey region origin-destination analysis is run as the filenames in this file need to match the names of the TomTom reports.

5. Network-wide traffic statistics

Network-wide traffic statistics are obtained from the TomTom Move Traffic Stats application using an ‘area analysis’ covering the Greater Capital City Area (GCCA) of the cities that have Transurban toll roads. TomTom collects data for the MOVE Traffic Stats application using aggregated, anonymous GPS data from various sources. This includes data from personal navigation devices, in-dash systems, and mobile applications. The data is then processed to provide insights on road speeds, travel times, and traffic density, ensuring a comprehensive view of traffic conditions while maintaining user privacy.

Like the OD product reported in Section 0, the Traffic Stats analysis can be customized with date range and time range parameters. An example output report from the tool for a section of Southern Link can be seen in Figure 5.

Figure 5 – Example output from TomTom Traffic Stats application



5.1 Processing steps

The network-wide traffic statistics analysis can be broken down into four main components:

(1) Running, accepting and extracting TomTom data – undertaken using TomTom and Python
(2) Merging TomTom data to Cube routable network – undertaken using Cube and Python
(3) TomTom data adjustment and calculation of other metrics – undertaken using Cube
(4) Skimming and reporting of key metrics – undertaken using Cube

Component (1) is a combination of manual setup in TomTom and running a python scripted process. Components (2) to (4) inclusive are run in Cube Voyager (proprietary transport modelling software package) which also includes some python scripted processes.

5.1.1 Running, accepting and extracting TomTom data

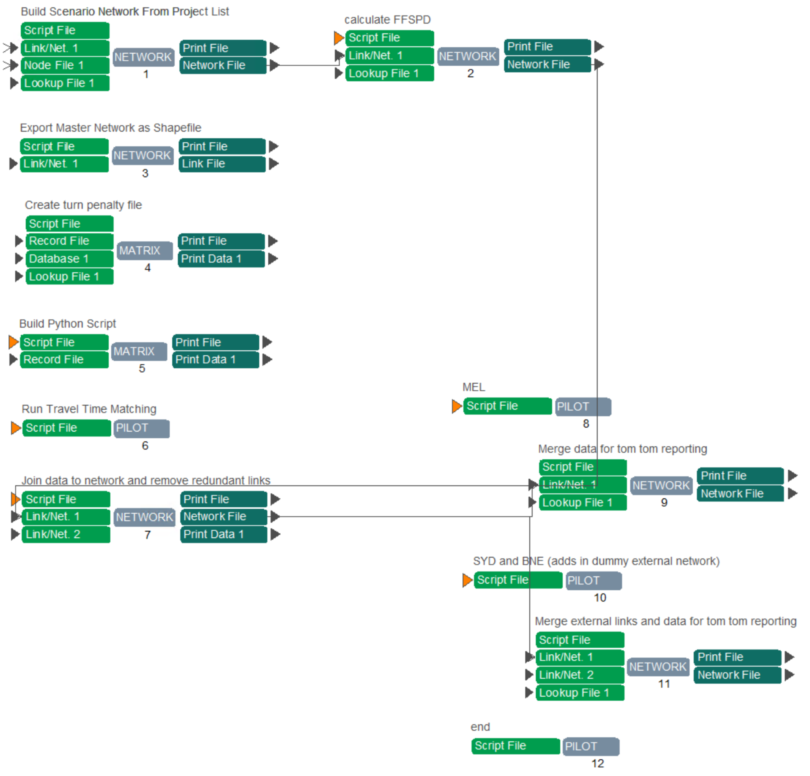
// detail here

5.1.2 Merging TomTom data to Cube routable network

The downloaded and extracted TomTom data resulting from completion of Section 5.1.1 is then merged onto a routable network so that it can be used in Cube transport modelling software. Figure 6 shows the programmatic flow for this part of the process, in particular:

- Programs 1 to 4 build the base network and populates with required attributes
- Programs 5 & 6 prepares and runs the Python process to create the TomTom data table that will be merged onto the base network. This is based on the outputs form Section 5.1.1.
- Program 7 merges the TomTom data on to the base network and removes network links that are not required for analysis (e.g. future year projects and public transport only links)
- Programs 8 to 12 adds additional external network to ensure full coverage of TomTom data beyond the base network. This is only required for Sydney and Brisbane.

Figure 6 – Programmatic flow diagram for merging TomTom data to Cube routable network



5.1.3 TomTom data adjustment and calculation of other metrics

The base network resulting from the completion of Section 5.1.2 must undergo an adjustment as follows.

Where sampled TomTom data does not exist (or is unusual) or could not be matched to the Cube network:

- Travel times are calculated based on ‘free-flow’ speeds. Taken as posted speed multiplied by a road hierarchy and area density factor that simulates the effects of ‘friction’ on the road networks (e.g. effects of traffic lights, roundabouts, parking, driveways etc…). The factors are taken from strategic modelling approaches to better estimate speeds in uncongested traffic conditions.

Where sampled TomTom average speeds are outside of a sensible range:

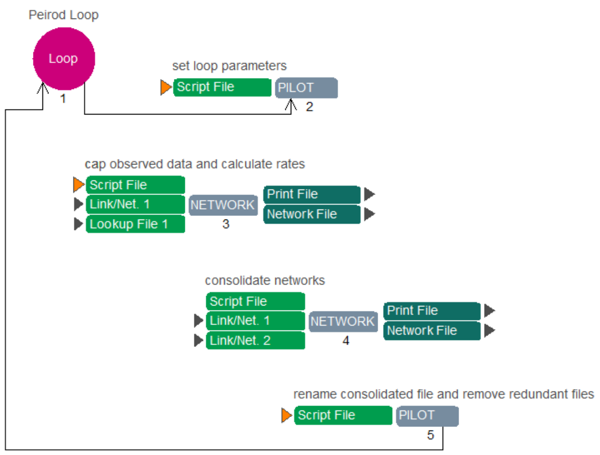
- Speeds are capped to be no slower than 5km/h and no greater than 110km/h. Average travel times are recalculated based on the adjusted speed.

This network can then be used as the basis for additional calculation of key metrics required for reporting. At present no other metrics are calculated but could include fuel consumption, GHG emissions, crash/incident rates, travel time reliability metrics, number of traffic lights etc…

The process is structured as follows:

- Programs 1, 2 & 5 controls the process to loop through each time period
- Program 3 contains the calculations required to adjust odd or missing TomTom travel time data and the application calculations for any other key metrics
- Program 4 undertakes the transfer of the data for the current period onto one consolidated network

Figure 7 – Programmatic flow diagram for calculation of key metrics

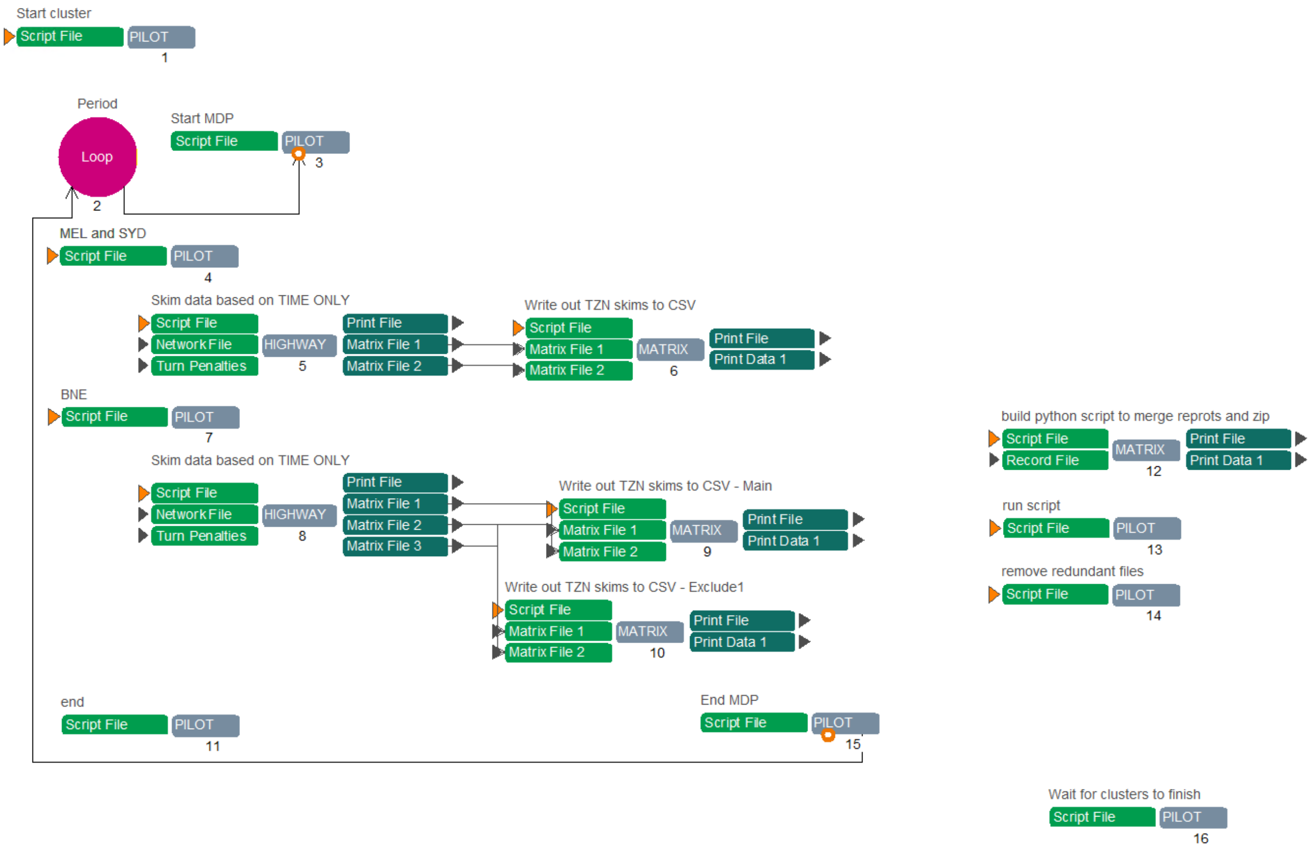


5.1.4 Skimming and reporting of key metrics

The skimming process accumulates the values calculated for each link (detailed in Section 5.1.3) on a given path between every origin and destination zone for every time range (period) of interest. Figure 8 shows the programmatic flow for this part of the process, in particular:

- Programs 1 to 3 & 15 to 16 controls the process to loop through each time period and multi-thread the processes to distribute the network skimming and reporting workload.
- Programs 5 & 6 undertake the skim and reporting for Melbourne and Sydney networks. These are separated as Brisbane has a complication that needs to be considered.
- Programs 8 to 10 undertake the skim and reporting for the Brisbane network. Brisbane is complicated by the need to simulate movements to/from toll points BS13 and BS24 for which there is no TomTom data available. This is handled by an additional path skim in the highway program and an additional reporting step.
- Programs 12 to 14 formats the reports and compresses the output .csv files to save on disk space. This is done via a call to a python script.

Figure 8 – Programmatic flow diagram for skimming and reporting of key metrics



The skimming process accumulates and saves data for two possible fastest travel time paths for every origin-destination zone combination. The first accumulation is based on a path where all transport network links are available for the skim (including toll road links). This is the tolled path skim. The second is based on a path where the toll road links are removed from the transport network. This approach forces the program to choose paths that do not use toll roads to satisfy the travel between each origin and destination pair. This is the non-tolled path skim. This allows us to compare the accumulated metrics between a theoretical toll road user and a non-toll road user.

It should be noted that for some origin-destination pairs the fastest travel time path will always be the same regardless of whether toll road links are available or not. This implies that the fastest travel time path is always on the non-tolled route. In this instance, no difference can be calculated so these records are discarded.

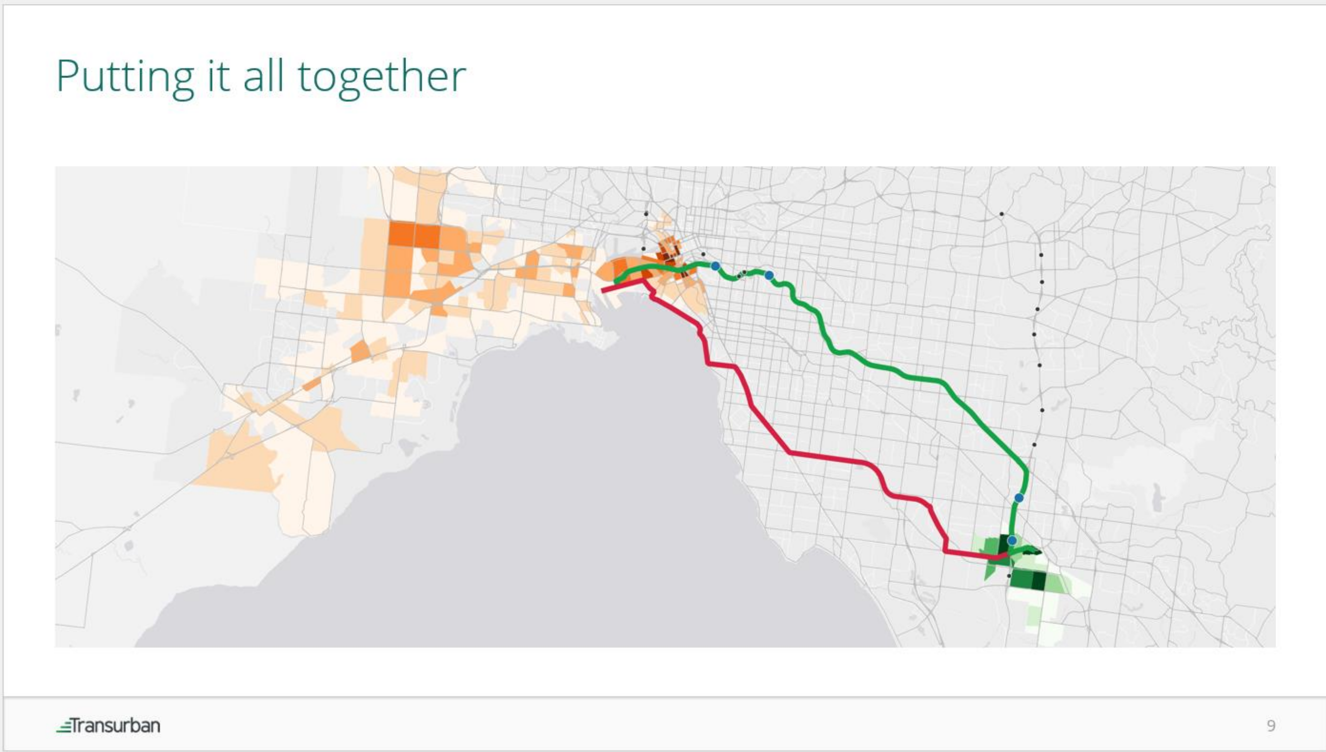
//customer database has some movements that are slower than non-tolled – report how we treat these

The data will therefore only be reported for any origin and destination zone that has a tolled route on the fastest path, and, where this is true, the next-best travel time for an alternate free-route is also reported. Data is not reported for origin and destination combinations that do not have a tolled route on the fastest path.

6. Computation of savings

// customer insights team to provide summary documentation

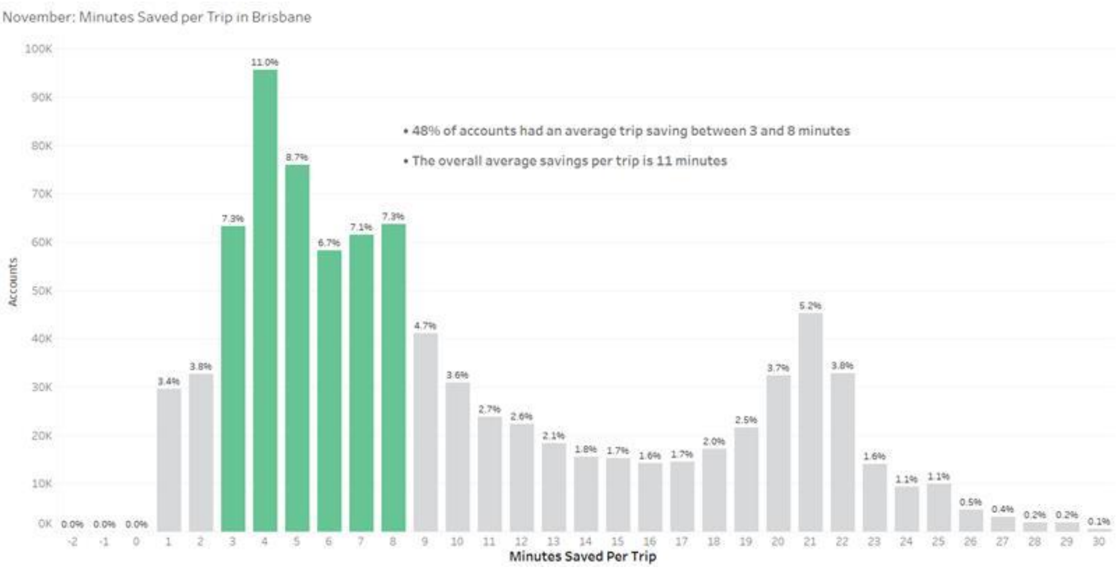
6.1 Processing steps



We can then start to query and accumulate travel times for origin-destination pairs within the catchment areas. The example path in green shows the tolled route whilst the path in red shows the next best free route alternative for a journey between Dandenong and Swanson Dock.
Note that in this example alone there are over 20,000 combinations but when we expand that out for all of the gantry-gantry combinations for all of the time periods we assess, it escalates to over 450 million combinations.

// a little more detail here

// example of high-level summary stats



7. Extensibility

// detail how other metrics can be added

// detail other areas that distribution proportion data may be useful