Appendix E:
Data analytics methods and other technical information

Data analysis

Purpose

We analysed tram service data from 2023–24. Our analysis focused on two key attributes of accessible tram services: level-access stops and low-floor trams.

We looked to answer questions such as:

- How often did tram services offer these two accessibility features?
- Which routes offered these two accessibility features?
- Which areas of the tram network were usable for passengers with mobility restrictions?
- How did passenger experience differ for those who need a low-floor tram?

Data sources

Our analysis relied on 4 key data sources. The Department of Transport and Planning provided 3 of these datasets. Yarra Trams provided the remaining dataset.

We enriched our analysis with 3 supplementary data sources. These allowed us to:

- identify recently upgraded tram stops
- validate stop names and locations
- add extra geospatial attributes to the dataset.

Figure E1: Key data sources

Dataset	Description	Source	Period
Timetable	• Timestamped record of each tram as it travelled across the network • Each record represents a tram passing an automated vehicle monitoring (AVM) beacon, while travelling a specified route • Dataset does not include details of individual tram stops –only AVM points • Dataset does not include detail about tram vehicle used	Department of Transport and Planning	2023–24
Service runs	• Record of the vehicle used for each service run that occurred • Dataset does not include timestamps at each tram stop or AVM—only the start and end times of the service run • Used to identify which timetable records used low-floor trams	Yarra Trams	2023–24
Tram stops	• List of tram stops in the Melbourne tram network • Dataset includes detail about stop accessibility • Dataset includes detail about the AVMs that come before and after each tram stop • Used to map timetable records to individual tram stops	Department of Transport and Planning	As at December 2024
Patronage	• Estimates of passenger volumes by day of week and time of day • Used to improve accuracy of the passenger simulation	Department of Transport and Planning	2023–24

Source: VAGO.

Figure E2: Supplementary data sources

Dataset	Description	Source	Period
Upgraded stops	• List of tram stops that have been upgraded to level access, with date of upgrade • Used to identify stops that were upgraded recently	Department of Transport and Planning	February 2014–July 2024
Public transport stops	• Open geodata for train, tram and bus stops in Victoria • Used to validate tram stop names and locations	Transport Victoria	As at 28 April 2025
Boundaries	• Open geodata for suburb and postcode boundaries in Australia • Used to enrich tram stop list with suburb and postcode attributes	Australian Bureau of Statistics	As at 20 July 2021

Source: VAGO.

Methodology overview

Direct calculation

Where possible, we calculated results directly from the datasets we received.

For example, we used the tram stops dataset to determine the proportion of tram stops that offered level access.

Enriched timetable

Some calculations required information from more than one dataset.

We combined the timetable, service run and tram stop datasets to create an enriched timetable. Records in this dataset represent each time a tram departed from a tram stop in the 2023–24 financial year.

Each record in this dataset includes 6 critical attributes:

- Departure date
- Departure time
- Tram stop
- Level-access stop indicator
- Tram route and direction of travel
- Low-floor tram indicator.

We used this dataset to calculate two measures of tram network accessibility:

- Proportion of tram services that used low-floor trams
- Proportion of tram services that had both a level-access stop and a low-floor tram.

Passenger simulation

Some calculations required information that was not available in any dataset.

For example, when a passenger is unable to board a tram because it has a high floor, they may need to wait for the next low-floor tram to arrive. No record of this delay exists because their journey hasn't started yet. To represent the experience of passengers who need a low-floor tram, we ran a simulation across the enriched timetable.

We simulated more than 8 million passenger arrivals across 461 level-access tram stops. For each simulated passenger, we checked whether the first tram to arrive was low-floor or high-floor. If the first tram was high-floor, we then calculated how much longer the passenger would need to wait for a low-floor tram.

We used the simulation to calculate measures of passenger experience, such as:

- Proportion of passengers who waited no extra time for a low-floor tram
- Additional wait time for a low-floor tram on a typical day (median result)
- Additional wait time for a low-floor tram on a bad day (95th percentile result).

Enriched tram stop list

Lastly, we enriched the tram stop dataset for use in our interactive map.

To accomplish this, we combined the tram stop, upgraded stop, public transport stop and boundary datasets.

Each record in the dataset includes 7 attributes to support visualisation:

- Validated stop name
- Validated stop location (latitude/longitude)
- Suburb
- Council
- Postcode
- Level-access stop indicator
- Recently upgraded stop indicator.

Enriched timetable methodology

Dataset profiles

This analysis combines 3 datasets with varying granularity and dimensionality.

Figure E3: Datasets used to generate the enriched timetable

Dataset	Size and granularity	Key dimensions
Timetable	• 29.5 million records • One record per tram passing an AVM beacon	• Date • Time • AVM point • Tram route/direction • Status
Service runs	• 2.1 million records • One record per tram route service run	• Date • Service run start time • Service run end time • AVM starting point • AVM ending point • Tram route/direction • Tram class
Tram stops	• 1,643 records • One record per tram stop	• Tram stop type • Tram routes serviced • Direction of service • Nearest AVM point before each stop • Nearest AVM point after each stop

Source: VAGO.

Timetable data cleaning

We used R to:

- Remove non-operational services (such as vehicles moving to/from the tram depot)
- Remove out-of-scope City Circle services
- Remove services with no start and end time
- Assimilate route variations to their parent route (for example, route 57a becomes route 57)
- Determine the start and end time of each service run, to match the grain of the service run dataset.

Service run data cleaning

We used R to:

- Remove runs with no start and end time
- Remove out-of-scope City Circle service runs
- Classify each service run as using a low-floor or high-floor tram.

Tram stop data cleaning

We found that about 7 per cent of AVM point or route combinations found in the tram stop data had no matches in the timetable data. After consulting with the auditees, we applied 147 AVM point corrections to improve alignment between the datasets.

We used M to:

- Remove inactive tram stops
- Remove out-of-scope City Circle tram stops
- Classify each stop as level access or non-level access
- Change the dataset granularity to one record per stop **per route**.

Merging datasets

To determine the type of tram associated with each timetable record, we merged the timetable and service run datasets.

We used R to match records by date, route, direction and service run start/end times. More than 99.9 per cent of timetable records had a matching service run record. We removed the unmatched records.

To translate from AVM points to tram stops, we merged with the tram stop dataset.

We used R to match records by AVM point, route and direction. More than 96.5 per cent of timetable records had at least one matching tram stop record. We found that the unmatched records were from the start or end of a route, such as when a tram passes a monitoring point while changing directions. Vehicles would not be carrying passengers at these times, so we removed the unmatched records.

Records in the resulting dataset represent each time a tram departed from a tram stop in 2023–24.

Assumptions and limitations

- When we conducted this analysis, the most recent complete financial year was 1 July 2023–30 June 2024.
- We excluded route 35 (City Circle) from analysis. The City Circle is a heritage route providing primarily tourist services.
- The datasets we used reflect the actual tram services that occurred. This may differ from the scheduled tram service and vehicle. Our results include the impact of network disruptions such as major events, rallies, service obstructions, infrastructure faults and asset maintenance.
- Four tram stops were upgraded to level access during 2023–24. We modelled these stops as though they provided level access for the full year.
- Twelve tram stops on Latrobe Street were upgraded to level access in July 2024. This is outside 2023–24, so we treated these stops as non-level access. For current information about level-access tram stops, refer to the Transport Victoria public transport journey planner.

Passenger simulation methodology

Dataset profiles

This analysis relies on two datasets. The first is the result of the merge described above. The second is the patronage dataset supplied by the Department of Transport and Planning.

Figure E4: Datasets used in the passenger simulation

Dataset	Size and granularity	Key dimensions
Enriched timetable	• 74.7 million records • One record per tram departing from a tram stop	• Departure date • Departure time • Tram stop • Level-access stop indicator • Tram route and direction of travel • Low-floor tram indicator
Patronage	• 246,000 records • One record per month, route, day type, day of week and 15-minute time of day interval	• Day type (weekday, weekend, public holiday or school holiday) • Hour • Minute group (15-minute intervals)

Source: VAGO.

Patronage data cleaning

To improve the accuracy of the passenger simulation, we generated time-of-day patronage distributions for weekdays and weekends. We excluded public holidays, school holidays and Sundays from these distributions.

We also generated day-of-week patronage distributions for weekdays and weekends. We excluded public holidays from these distributions.

Simulating passengers

At each level-access tram stop on each tram route, we generated 8,000 simulated passenger arrival times.

To avoid simulating a passenger at a stop with no active tram services, we restricted the arrival times to fall between 6 am and 11 pm. Within those hours, we weighted the arrivals using the patronage distributions described above.

We also excluded tram routes with less than one per cent low-floor tram service, to avoid skewing the wait time results.

Wait time calculation

To determine the first tram that would arrive for each passenger, we merged the passenger arrival times with the enriched timetable dataset.

We used R to match records by stop, route and direction, then found the minimum duration between passenger arrival time and service departure time. This represents the first tram available to the passenger.

We then repeated the process, but with the timetable dataset restricted to low-floor tram services. The result represents the first low-floor tram available to the passenger.

When these service departure times are the same, it means that the first tram to arrive was low-floor. In this case, a passenger with restricted mobility had no additional wait time.

When these service departure times are different, it means that the first tram to arrive was high-floor. In this case, a passenger with restricted mobility was unable to board and needed to wait for a low-floor tram. We calculate their additional wait time as the difference between the two service departure times.

We removed simulated passengers with extremely long wait times (over 8 hours). This ensures that passengers who were inadvertently simulated at a stop that was temporarily closed will not skew the wait time results.

From this, we summarised the simulation results by route and stop, by route, and for the whole tram network. The passenger experience measures we calculated were:

- additional wait time for a low-floor tram on a typical day (median result)
- additional wait time for a low-floor tram on a bad day (95th percentile result)
- proportion of passengers who waited no extra time for a low-floor tram
- proportion of passengers who waited 0–10 minutes extra
- proportion of passengers who waited 10–20 minutes extra
- proportion of passengers who waited 20–30 minutes extra
- proportion of passengers who waited more than 30 minutes extra.

Assumptions and limitations

- The simulation assumes that passengers will not try to catch a low-floor tram on a route that does not offer low-floor tram service.
- The simulation assumes that a passenger will choose to wait for a low-floor tram, rather than seeking an alternative mode of transport or abandoning their journey.
- The simulation does not account for passengers who use journey planning tools or apps to identify low-floor tram services. While this information may be made available up to a day in advance, we wanted to capture the impact of needing to take a tram that is earlier or later than the passenger would prefer.
- The simulation does not account for times when a tram is too crowded for new passengers to board.
- The simulation does not differentiate between the reasons why a passenger might need to wait for a low-floor tram. In some cases, the simulated passenger might be waiting because several trams in a row were high-floor. In other cases, the simulated passenger might be waiting because there was a network disruption, and no trams can get to that stop.