MURPES

Stakeholder Error Report

JUNE 2023

Background

The MURPEs data for the period November 2022 was released in early January 2023. Following this release, MBIE became aware of some unusual trends in the data. A review was undertaken immediately to investigate the decrease in domestic visitors which was contrary to patterns in accommodation and tourism spend data for the same period.

Investigation

Data Ventures immediately reviewed quality assurance diagnostics and engaged with the organisations providing data inputs to investigate.

Through this review, we identified two errors in the input data feeds:

1. Cell towers were added to one of the input datasets over time, and were not mapped to RTOs. Devices pinging these cell towers were classified as being in 'unknown RTO' rather than their correct location. This led to overestimates in national domestic visitor numbers prior to August 2022 and underestimates in both national and regional series of all segments after August 2022 **(please see appendix one for more information).**

2. Non-mobile devices with cellular capability (cellular Internet of Things, a.k.a. IoT devices) are appearing in one of the input data feeds. The number of IoT devices has increased substantially since the beginning of the series as new makes & models came on the market and older devices were updated to use more modern cellular networks such as 5G. As a result, the current method for excluding these devices is no longer fit-for purpose. Almost all of the IoT devices are domestic. Because IoT devices tend to move significantly less than cell phones (e.g., smart meters), the devices which could not be classified into domestic local and domestic visitor segments have grown over time, leading to errors in domestic local and domestic visitor estimates.

Remediation

Issue 1, involving the unknown areas, was resolved in early March.

Issue 2, ongoing when the previous report was published, has now been resolved. This issue required a more significant technical investigation. A new method to exclude IoT devices was devised by the data provider. Data Ventures quality assured the reprocessed data by comparing it with the previously supplied series and also with the data from the other data provider. The data processing pipelines of both Data Ventures and the provider were updated accordingly. The implementation of the fix required two distinct approaches: one to exclude the latest IoT devices on the market and new devices that come onto the market in the future, and two: to exclude IoT devices from the historical data of the series.

Data Ventures worked closely with the data provider to future-proof the solution, which involves maintenance of a list of TACs (Type Allocation Codes, more information on TACs can be found in Appendix two) that represent valid phone devices since this can change over time (please see appendix for more information). This list is then used to exclude the IoT devices on an ongoing basis.
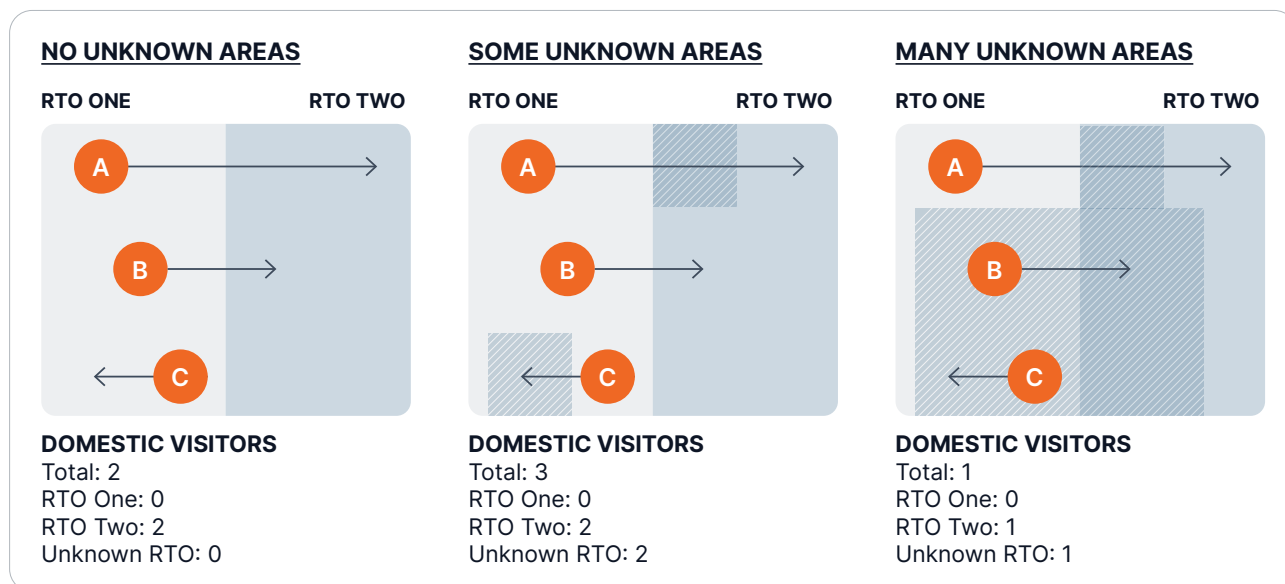
For the historical device counts in domestic segments (both local and visitor), the data provider used the current list of TACs to exclude the IoTs with the assumption that there are negligible valid TACs representing past phone devices that have vanished now. Data Ventures quality assured reprocessed past domestic device counts by comparing with the previously provided counts and also with the domestic device counts from the other data provider.

This approach could not be used for international segments as the TACs of the international devices that might have visited NZ in the past and then left would be falsely counted as IoTs. Fortunately the earlier provided international device counts from the data provider were still valid as the IoTs were appearing only in the domestic feed. For example, it is highly unlikely to have a smart meter in NZ with an external country code.

Quality checks, outlier identification, comparator analyses, and sense checking have been completed prior to the data release.

The MURPEs data series has been updated until the end of **May 2023,** has been fully quality-checked and available from **July 2023.**

Appendix one: devices in unknown areas – diagrams

NO UNKNOWN AREAS

RTO ONE   RTO TWO

DOMESTIC VISITORS
Total: 2
RTO One: 0
RTO Two: 2
Unknown RTO: 0

SOME UNKNOWN AREAS

RTO ONE   RTO TWO

DOMESTIC VISITORS
Total: 3
RTO One: 0
RTO Two: 2
Unknown RTO: 2

MANY UNKNOWN AREAS

RTO ONE   RTO TWO

DOMESTIC VISITORS
Total: 1
RTO One: 0
RTO Two: 1
Unknown RTO: 1

Three people who live in RTO One move around during the month. **Person A** and **Person B** take domestic trips to RTO Two, while **Person C** moves around within their home RTO.

When no cell towers are unmapped, all trips are identified correctly - both **Person A** and **Person B** are counted as domestic visitors, while **Person C** is not.

When a few small areas are unknown, **Person A** crosses through an unknown area on their trip but is still correctly identified as a domestic visitor in RTO Two. **Person B** is unaffected. **Person C** crosses through an unknown area and is falsely counted as a domestic visitor in the total as they seem to be outside their home RTO. The total number of domestic visitors is overestimated but the estimate for RTO Two is unaffected.

When many areas are unknown, both P**erson B** and **Person C** are counted as living in 'unknown RTO'. **Person A** is still correctly counted as a domestic visitor in RTO Two. **Person B's** domestic trip is missed as their entire movement occurs within 'unknown RTO'. **Person C's** movements also occur entirely within 'unknown RTO', so they are no longer falsely classified as a domestic visitor. Both the total number of domestic visitors and the number of domestic visitors in RTO Two are underestimated.

NOTE: These diagrams are purely for explanatory purposes. Data Ventures (including Stats NZ), only receive anonymised data and do not analyse or share information about individuals' movements, only aggregated estimates.

Appendix two: Type Allocation Code (TAC)

A Type Allocation Code (TAC) is a 8-digit number that identifies a cellular device's manufacturer and model number, and can only be allocated by the GSMA through their appointed reporting bodies.

A TAC is required by device brand owners and manufacturers to create the first 8 digits of a device's IMEI (International Mobile Equipment Identity) unique number. This code has proven to be effective in filtering out IoT devices while ensuring privacy, since it does not identify individual devices.

The data provider established a subset of TACs from their feed that covers only smartphone and mobile phone categories. This TAC list is compared against the daily data feed and excludes any records that don't appear on the TAC list. This applies to the daily extract, as well as the monthly extract.

For further information, please contact