MURPES

Stakeholder Error Report **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 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. Because IoT devices tend to move significantly less than cell phones (e.g., smart meters), the proportion of devices in the feed which are classified as domestic visitors has fallen over time, leading to errors in domestic visitor estimates.
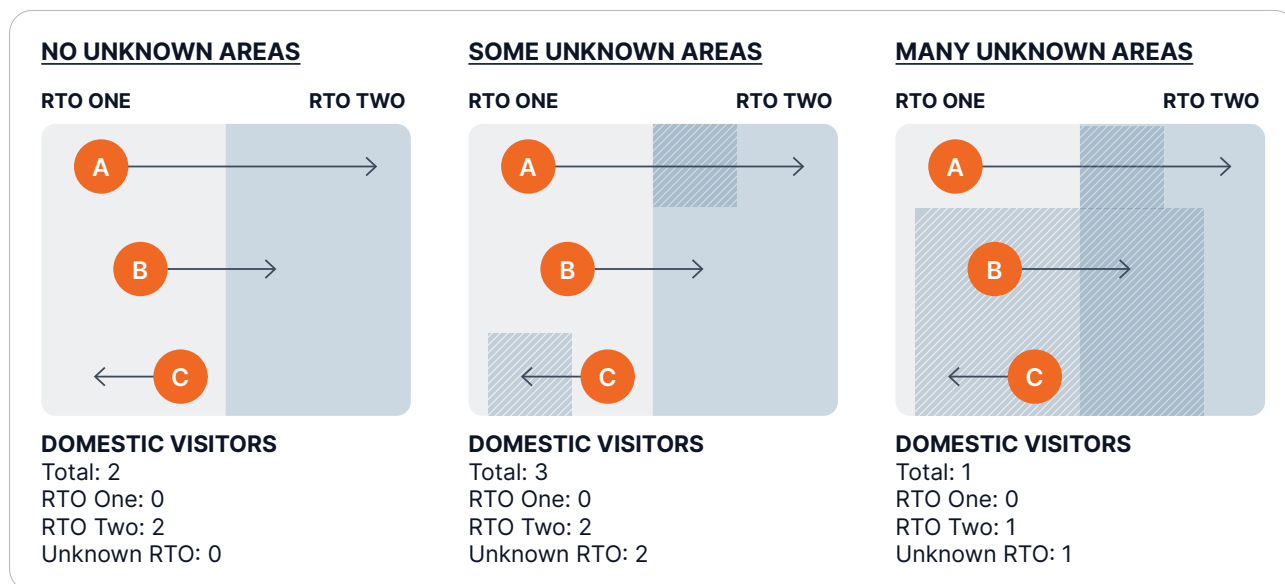
Remediation

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

Issue 2 required a more significant technical investigation and remediation is still ongoing. The fix will require the input data feed to exclude the latest IoT devices on the market and new devices that come onto the market in the future. We are working closely with the data provider to future-proof the solution. Due to the large quantity of data and the need to recreate the entire series, it will take some time to resolve this issue.

Data Ventures recognises the importance of further bolstering quality assurance processes and has also designed a Five-Gate Quality Assurance process, which will be completed before updated MURPEs are released. This QA model increases the formality of pre-release process, creating additional automated quality checks, outlier identification, comparator analyses, and sense checking. These improved checks and balances are then all compiled into a report which is reviewed by experts before release.

The MURPEs data series updated until the end of **May 2023**, will be fully quality-checked and available by **July 2023**.

Appendix – Devices in Unknown Areas

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 **Person 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.

For further information, please contact
dataventures@stats.govt.nz