



DATA VENTURES

Population density pilot

The pilot

Together we will prove:

- the details of the commercial viability of a population density product and the data
- the quality of population density data inferred from location estimates and Stats NZ population expertise
- the high value use cases

We recommend up to a 6 month pilot to prove the initial model, followed by further iterations of the product and product roadmap informed by the learnings.



What the pilot will answer?

We will have worked with you to create a business case that adds value to your organisation.

It will have answered what is the value of this product.

For government, it's about making better decisions for Aotearoa NZ.

For others, it's about learning about their community and optimising and growing revenue.

For all, it's about creating clear public benefit in doing this.



Who has been involved so far?

Over the last three months 12 government organisations have provided validation and testing towards the product being provided in this pilot.

This pilot is the next step in developing a viable product offering that will deliver value to the customers in that group.

What are the benefits of the pilot?

- Access to data at a level of frequency and resolution that is better than anything available on the market
- The expertise of Stats NZ that improves the quality of the data so you don't have to.
- Any privacy and confidentiality risks are managed by Stats NZ expertise and reviewed by the Office of the Privacy Commissioner
- Simplifying the government procurement process



What does success look like?

We will agree on success criteria for a pilot, and we expect them to be themed around:

- You develop at least one valuable workflow to incorporate the data provided by us in the pilot
- We reduce your costs relating to the acquisition and processing/handling of data



What is the commitment from you?

- People from your legal, data/insights, financial and executive areas that will be appropriate for weekly progress meetings, on-going agreement discussions, and authority to make appropriate decisions. We are expecting this will require up to 40 hours a month of time across these areas of your organisation.
- A workshop with relevant parties to understand the use cases to build success criteria.
- A commitment to work to build a business case to secure funding for the 18/19 financial year if value is identified.
- A small upfront contribution to help cover some of the data processing costs?



What to expect at the end?

Within 6 months we will have made a seamless transition from testing viability of the product to a live product in use by your organisation.



Next steps

Signing an MOU to go ahead with the pilot and confirm the commitment of resources needed.





DATA VENTURES

dataventures@stats.govt.nz @dataventuresnz

<https://medium.com/data-ventures>

Data Ventures pilot plan

Data Ventures is the commercial arm of Stats NZ. We are engaging with interested data providers on opportunities to increase the value of their data and their supporting services and functions by collaborating with Stats NZ and its unique position in the data ecosystem.

More specifically, Data Ventures adds value throughout your data product and service pipeline and the content below will describe how we would collaborate with you to achieve this value add.

Evaluation of your data	Data Ventures will work with you to understand your data and to provide an evaluation of your data. This include audits on the quality of the data and understanding the variances, and what actions could be done to improve quality.
Validating use cases	As a government organisation, Data Ventures has the trust to test and validate the use cases with government agencies and crown owned entities. We work to understand their end to end processes and how the data can be used to support their goals.
Creating a suitable offering to customers	As use cases are validated and specific needs are identified, we work with you to identify where there are opportunities to create a suitable offering. This may include applying specific models from Stats NZ expertise to the data, or combining the data with other data sources to improve the quality and fit of purpose.
Managing risks	Data Ventures will manage all the associated risk with the product. This includes the privacy and confidentiality risk of the data, and any external prespection risk. This is done through statistical methodology expertise and data management expertise within Stats NZ.
Product to market	Data Venture's being a government agency reduces the burden and overheads associated with procurement with the government. We also have better access to government customers and their data needs/use cases due to the insights from other Stats NZ product offerings.
Ongoing product management	Data ventures will provide ongoing management of the products and act as a feedback source and lead generator for additional products and services that fall outside the scope of Data Ventures.
Market analysis	Data Ventures will supply the value-added dataset back to you for your own insights and analysis e.g. understanding your market share.

Population Density

Version 1 Product Information

Released under the
Official Information Act

1. Contents

Contents	2
Version Control	3
Product Team	3
Introduction	3
Target Customers	4
Customer Use Cases	4
Customer Benefits	4
Features	5
Privacy Impact Assessment	5
Population Density From Mobile Data	6
Population Density stored in a database	6
API Connection	6
Web Interface	Error! Bookmark not defined.
Customer Interface	7
Searchable entities	7
Technical Environment	8
Product Roadmap	8
Known Issues	8

2. Version Control

Date	Details	Issued By	Version
19 Oct 18	Initial strategy draft for internal approval	JM	0.1
7 Dec 18	Pre-Release Team Revision	JM	0.2
15 Feb 19	MVP Release for Pilot	JM	0.3
May 19	Population Density - Product Release	JM	1.0

3. Product Team

Jamie Marshall	Product Owner
Robert Chiu	Business Development
Drew Broadley	Executive Director
Holly He	Operations

4. Introduction

Population Density is the first product from Data Ventures, where Stats NZ public data and methodologies meet data from private companies, to make population insights which will allow customers to make informed decisions based on recent population data.

In this first iteration, the product will allow customers to query the population density database and provide aggregated data to customers at the National, Regional and Territorial Local Authority levels through a web-hosted Application Programming Interface (API). Those

customers that are unable to utilise the API will be able to use a web interface to query the database then download the results through a web portal.

5. Target Customers

Population Density V1 will be available only to Government Departments and Crown Entities, including State-Owned Enterprises, Crown Research Institutes, Crown Financial Institutions, Crown-owned entity companies and other Crown entity companies.

6. Customer Use Cases

A number of government agencies have told us that there are problems they are trying to solve where the data just isn't available to help them answer questions to solve problems effectively.

Some of the use cases that the agencies may have could be:

- What is the population change in a region from a Rugby test match?
- Is the infrastructure sufficient for peak population demands and how can we distribute the load?
- Where do we put services after an emergency event?
- Where is a good place to put a new school?
- Is this suburb growing or shrinking?
- Where are the summer and winter hotspots and the impact of population on our national parks?

The population data from the queries are then able to be used on its own or added to the crown agencies own insights and data to assist in decision making.

7. Customer Benefits

The Product Density V1 product will allow customers to have access to data that will allow them to make the best population-based insights for their department or business. This will deliver the following benefits:

- Ability to search by day of the week allows identification of the effects of weekend travel on population density to help determine the effect that population density might have on residential areas or recreational facilities during non-work days.
- Ability to search by hour of the day allows identification of the effects of commuting on population density by seeing population density changes in a region during working hours.

- Ability to search by date allows the identification of the effects of specific events in the area of interest on population density, such as public holidays, sporting and entertainment events.
- Having the data sources updated regularly ensures that customers are making their decisions with the most current information as population density may change over time from population base information, such as the census.
- Having the data sources updated regularly allows customers to identify population density trends overtime to ensure that infrastructure is being delivered to areas that may require investment due to increasing population.
- Having the data sources updated regularly allows customers to isolate seasonal changes on population density allowing customers to make business decisions around seasonal variations in population due to cropping, seasonal work and tourist industry.
- Having the data updated regularly allows customers to respond to any time-sensitive problems

8. Features

The features associated with Population Density V1 are created to allow ease of integration into customers existing geospatial technologies through the development of a web-hosted API. Those customers without geospatial technologies need to be able to acquire and visualise the data through a graphical user interface (GUI).

8.1. Privacy Impact Assessment

The privacy impact assessment was undertaken by an independent third party, Info by Design. The key conclusion point from the Assessment is:

“There are no identified privacy risks to the proposed Population Density product. Data Ventures is not collecting or using personal information as it is defined in the Privacy Act and this analysis of Data Ventures’ proposed processes show that it is following best practice information management and the OPC’s data and analytics principles.”

The Office reviewed the Privacy Impact Assessment and a corresponding application for the Privacy Trust Mark. The office stated that:

“Population Density, therefore, is complying with its obligations under the Privacy Act by not collecting personal information where it is unnecessary”

The Privacy Trust Mark was not granted as the Population Density product does not collect any personal information or allow individuals to re-identified through the product or contributing data.

8.2. Population Density From Mobile Data

The aggregated data from the individual telcos are stored in separate bins in the AWS environment. These datasets are then aggregated further into a single dataset where Stats NZ statistical methods and demographics will take the aggregated location information and make this an indicator of the population in a geographical area.

The population density will be a count of indicated population per geographic area. The population counts will be by hourly slices, days of the week and dates. Over longer periods of time, hourly resolution of population estimates may be too granular, so summarised results at a daily, weekly, monthly or seasonal resolution will be made available to customers.

8.3. Population Density stored in a database

The aggregated Population Density will be stored in a database in a secure cloud-hosted environment. The database format is to be determined by the developer but is expected to be a natively hosted by the web hosting service.

8.4. API Connection

The predominant connection to the database is to be by way of an API. The API will detail all of the supported queries and outputs from the database.

The API is to be documented and made available publicly as part of the sales and marketing package to allow customers to fully evaluate the product before purchase.

The language of the API is GraphQL. Details of the API can be found here:

<https://graphql.org/>

8.4.1. API Documentation/Schema

The GraphQL is a self-documenting API language

8.5. Web Interface

8.5.1. Customer Interface

As there are a number of customers without the ability to implement API based interface to the database there is a requirement for a customer accessible web-based graphical user interface (GUI).

This GUI shall be accessible from the web browsers that are currently supported by the customers. Due to the environments being implemented by customer information technology departments, these may not be the latest version of web browsers.

The GUI shall include user and company authentication.

The GUI will present a graphical query/filter interface that allows the user to:

- Select the searchable entities to be used in the query
- Free searches are not available
- Return a table with the population count by area of interest
- The results can be downloaded as a comma-separated text file (CSV).
- The results shall be in accordance with the customer data dictionary

8.6. Searchable entities

The following searchable parameters will be available through the API and Web Interface.

- Search by date and time

Web Interface: Customers will be able to search the hour of the day and the date in NZDT or NZST. This will have potential risks because this is different from the API standard but the Web Interface and the outputs are designed to be human consumable. This means that there could be some inconsistencies on the days where daylight savings starts or ends. The table will return the date and time in either NZDT OR NZST, depending on the date.

API: API will accept and return the UTC DATE AND TIME

- Filter Results by location

Web Interface: Results can be filtered by either Territorial Local Authority or Region. If either a TLA or Region is selected the table will only contain those area units within the selected boundary.

API: API will accept TLA or regional filters and return results from area units within the selected boundary.

8.7. Technical Environment

The product will be hosted on the Amazon Web Service, including the database, administration and interfaces.

Security and privacy of the data, user administration is to be factored into the deployment of the product.

9. Product Roadmap

Population Density V1 will be improved upon by adding additional datasets in the future to increase the certainty of the confidence in the indications of population density.

Population Density V1 and subsequent versions will also be a base dataset for future products, such as identifying travel patterns.

10. Known Issues

No known issues with the product at this time.

Pilot: Mobile Location Data

Definition

What data do we need for the pilot using 13 months of national data?

We need two sets of data:

1. Device Counts: A per hour, per cell towers , device count across a day of 13 months from , broken down to hourly.
2. Cell Coverage Data (and historic covering the 13 months): data for telecommunications companies coverage including cell tower locations across all frequency bands for New Zealand. This would usually be three layers, one for 2G, 3G and 4G, a normalised layer of all three is fine too.

Specifications for Device Counts:

Area: National
Time period: Hourly intervals from 12:00am 01/02/2019 through to 11:59pm 28/02/2019
Data: Date/Time (hourly), Cell Tower ID, Cell Join (Coverage), MSISDNS (counts of device numbers per sector/coverage area)

Note: We have used the column names provided as per the calibration data.

Definition of a “count” is based on the first activity a device has in the hour period by cell tower/sector. The specifications are based on the output from the Feb 1st calibration data.

The Data will contain the following fields:

Date-Time (hourly intervals)	Cell Name (Cell Tower name)site ID/coverage look up ID Statistical Area 2 (suburb)	Cell Join (Mobile coverage area ID)Number of mobile devices (count)	Missions (Mobile device number count)
------------------------------	--	---	---------------------------------------

This can be provided as a CSV, or any other format that works for you.

Specifications for Cell Coverage Data:

Furthermore, the geographical boundary of each cell towers coverage, per band if easier, so that we can spatially relate the coverage to statistical boundaries.

These can be in any supported format but should contain a reference to the cell tower id to attach the device counts and date/time to. Preferences are for ESRI compatible SHP files with associated DB fields for each coverage area.

What happens during the pilot?

We develop a model that combines all telco data sets into a single view of population across 13 months.

As part of that we also convert device numbers to population counts using the Stats NZ IP around surveys and other respondent data in different areas of NZ.

Through this we may have some questions around the data.

This is the first effort to get a baseline count, to which next we can work with you to define Local, Domestic and International profiles of devices for ours and your use.

What data is required beyond the 13 months for the pilot?

On the successful completion of this phase of the pilots' data, we are looking into how we could segment data down to international, domestic and local population. We will discuss this as another phase of the pilot, as we first need industry parties (including the Office of the Privacy Commissioner) involved to workshop definitions of what is local, domestic and international population.

This will happen before we understand what is required around any data.

Questions?

Contact Robert Chiu @ Data Ventures - robert.chiu@stats.govt.nz

Population Density Data Dictionary		
Date	2019-03-05	
Version	1.10	
Name	Type	Description
Date/Time From	YYYY-MM-DD HH:MM:SS	This specifies the date that the data attributes from (start of the range). Ranges can be as small as one hour, and as large as 365 days.
Date/Time To	YYYY-MM-DD HH:MM:SS	This specifies the date that the data attributes to (end of the range). Ranges can be as small as one hour, and as large as 365 days.
Statistical Area 2 String		This specifies the geographical boundaries the data represents e.g. [Penrose] or [Ponsonby East]. We use the official Stats NZ definition of an Statistical Area 2, which is as follows: Statistical Area 2s are aggregations of meshblocks. They are non-administrative areas that are in between meshblocks and territorial authorities in size. Statistical Area 2s must either define or aggregate to define, regional councils, territorial authorities and urban areas.
Count	Integer	This represents the total number of people our models have calculated to be within the attributed time and place e.g. 9,632.
Notes		
To anticipate a likely question about people being in more than one location at once, our models calculated population counts based on aggregated data o		
Future data dictionary might be expanded to other common date filters e.g. day of the week (Tuesdays), week of the year (Week 14), seasons, months, and years.		
We will look to add other geographical boundaries such as regional councils, territorial authorities, Statistical Areas 1 and 2 in the future.		
For a visual representation here is the New Zealand Statistical Area 2s visualised on a map: https://datafinder.stats.govt.nz/layer/92212-statistical-area-2-2018-generalised/		
Delimiters - NOT space : / or -		
File Formats - I'm guessing CSV?		

Released under the Official Information Act