# Foodbank Australia

# Demand Forecasting Methodology
## Food Insecurity Model

**Objective**
To develop a model which identifies Food Insecurity across Australia, and thus provides insights as to how Foodbank can meet the demand for Food Relief.

**Overview**
Food security and insecurity is used to describe whether a person or household can access food in the quantity and of the quality they (and their dependents) need to live an active and healthy life.

Food insecurity is extremely diverse and complex and is not measured at a population level regularly or consistently in Australia. This is further complicated by the complexity of the supply chain and the large number of organisations involved in responding and distributing to areas of need and ultimately onto the end user.

Foodbank is the largest supplier of food to people in need in Australia and are in a prime position to help provide insights and more understanding in this area. Whilst their existing supply of food to their network of agencies is not in itself a reliable indicator of what drives demand, it provides a unique and practical understanding of the key variables that do, which can be used to project demand across Australia.

A wide number of data sources were introduced to the analysis to understand the diversity of demand indicators. The main sources were Census 2016, DSS Support Services and The Vulnerable Australia Model (developed by The Art of More) which has assessed and integrated over 70 data sources to develop an understanding of vulnerable communities.

Food insecurity can be broadly split between systemic and transient. Systemic are often people or households that cannot afford to feed themselves and/or are unable to access food. This group's needs can fluctuate substantially in terms of food demand, frequency and the length of time each is reliant on support.

Transient food insecurity is typically due to an event or natural disaster (drought, bush fires, floods etc) where immediate assistance is required for broad audiences often not typically exposed to food insecurity. These are usually more short term in need but can leave pockets of the population with systemic food insecurity.

This project has set out to understand what key variables or indicators drive demand for food and identify Australians that are more vulnerable to food insecurity. The focus has been, initially, on systemic food insecurity as it is more of a constant and more reliable to predict.

A number of modelling techniques were undertaken to take the various data sources, to identify and forecast food insecurity across Australia by location, volume and an estimate of impacted Australians. A gravity model was utilised to allocate Foodbanks kilogram of supply data to each SA1 within its vicinity. Decision tree analysis (CHAID) was utilised to then identify the key variables driving the demand allocated by the gravity model, and the associated volume of food shortage. K means cluster analysis was then utilised to define the unique relationships between different demographics of food insecure people, and their key demand drivers, providing a more nuanced view of the varying circumstances of the food insecure and hence their drivers of food demand. With this understanding, the team were able to map out these segments, their drivers and their associated demand at an SA2 level, to provide a view of the volume and degree of food insecurity across Australia.

Below outlines a step by step process undertaken in developing the model. This is very much an evolving model and it has been designed so it can continually integrate refreshed and new data sources as they become available to enrich our understanding and insights.

**Step by Step Methodology**

**Stage 1: Identify Demographic, Socio-economic and Vulnerability Drivers of Demand**

**Mapping the distribution of agencies and the volume of food that they distribute as an overlay on regional demographic, socio-economic and vulnerability indicators to identify what drives variations in demand. I.e demand changes with variations in demography, DSS etc.**

| | Description | Process | Commentary |
|---|---|---|---|
| 1 | Agency Distribution | Geo-code Agencies and tag them with the volume of food that they are supplied with (by Foodbank) over time (monthly, by food category for 2019 and 2020). | Agencies are the organisations that Foodbank supply to that in turn provide food to the people in need. |

| | Description | Process | Commentary |
|---|---|---|---|
| 2 | Outlier Agencies | Identify agencies that supply significantly larger in volume than the mean (and more likely not to be standard).<br>Reduced these agency volumes to mean plus 2 standard deviations. | It was required to clearly understand the existing distribution footprint of Foodbanks agencies. Some agencies (95 of 3786 Agencies in total, 2.5%) were disproportionately large when compared to others. These are more likely to be secondary distribution warehouses, agency centres or centres used for transient food insecurity and if included could adversely bias the analysis of volume distributions to populations (and the demand drivers - see later). To normalise, rather than deleting these agencies, their influence has been softened by reducing their volumes within two standard deviations from the mean volume across all agencies. This is a common statistical practice and in effect means that 5% of agencies had their volume reduced to fit within the volume bands of 95% of agencies . |

| | Description | Process | Commentary |
|---|---|---|---|
| 3 | Distribute food to a small area level | Applied a gravity model within a range of 5km from each agency to each SA1. This allowed us to clearly map and quantify the actual geographic footprint of Foodbanks agencies, the density of agencies and the concentration of food volume. | Little at present is known about where and who the food is distributed to (and where people come from to acquire it) once it reaches the agency.The food supplied was apportioned to SA1s within the agency's 5km vicinity to determine the geographic footprint of each agency. A gravity model enabled these volumes to be distributed to populations more accurately by applying a rule that the further away from an agency a population is the less share of that volume could be attributed to that agency allowing for an unbiased view of which populations are actually receiving food. 5kms was used because research into regular movement of people across a region showed that 80% of activity occurred within 5kms of their dwelling. This included retail shopping, restaurant attendances and regular entertainment. Gravity modelling also gave us a clear picture of the degree of influence/impact that Foodbank has across regions.<br><br>To enhance the model's accuracy, further research will be required to understand the agency's distribution to their client base. |

| | Description | Process | Commentary |
|---|---|---|---|
| 4 | Population and demographic overlay | Overlaid the population for each SA1 to determine kgs per person and aggregated this up to SA2 to iron out any discrepancies. 125 sub-set demographic and household variables were used in the attribute library for analysis - these variables represented cross sections of age, ethnicity, indicators of health and well-being and general socio-economic status | Develop an attribute library at the SA1 level which will allow us to take into account population and demographic and socio-economic variation to understand what indicated need in kgs per person |

| | Description | Process | Commentary |
|---|---|---|---|
| 5 | Determining the key demand drivers | Decision Tree analysis (Chaid) to identify the key variables that drive demand at the SA1 level. Aggregate those indicators to the SA2 level and append additional information to the SA2 attribute library including DSS, Demographic Time Series . | Analysis to determine what key variables are most responsible for influencing demand and how best to describe them (i.e. the demand drivers).<br>Chi-square automatic interaction detection (CHAID) is a decision tree technique and can be used for prediction as well as classification, and for detection of interaction between variables. Having apportioned the food from agencies to the population in a 5km radius, CHAID determined the relationship of a tested (120) key variables to these food volumes. This allowed us to determine the quantifiable degree to which each variable was predictive of food demand (in kgs). |

| | Description | Process | Commentary |
|---|---|---|---|
| 6 | Develop Segmentation Model | K Means Cluster Analysis to develop “hunger segments” based on the demand drivers and calculating the relationship between unique segments, their demand drivers and food requirements at the SA2 level | A number of variables were shown to influence demand (demand drivers) that differed from area to area (especially when comparing rural and metro areas).<br><br>Having now understood the statistical relationship of variables to food, a cluster analysis was run to identify unique clusters / segments of the population and their relationship to Food Insecurity, so a calculation of demand could be applied to each SA2 in Australia that would be a statistically accurate representation of demand.<br>10 core segments were identified (6 x metro and 4 x rural). Each segment has its own distinct set of influences on food demand. The segments allowed each group to be described (i.e. contextualised/humanised) along with the key demand drivers specific to that segment, and in turn effectively communicate them both for internal and external purposes. |

| | | | |
|---|---|---|---|
| **Stage 2: Estimate Demand x Volume and Population**<br><br>**Developing robust demand estimates to understand need in terms of KGs and apply a people metric** | | | |
| 7 | Develop a demand model | Two methodologies were developed to calculate demand:<br><br>1. **Calculating Foodbank's influence and aggregating demand for each segment** Utilising the gravity model (step 3), a gravity score was calculated for each SA1 based on Huff's probability which states that every agency has an influence on every SA1 around it. That influence is described as 1/Distance squared and the sum for all agencies by SA1s is a quantifiable measure of the total influence of Foodbank in each SA1 by all agencies. All SA1 scores in each segment were aggregated to calculate demand as KG per capita multiplied by the population within each SA1 to provide a demand estimate that could be projected to every SA2 and thereby to every segment.<br><br>2. **Verifying demand based on existing consumer surveys** Geo-code participants of recipient surveys from 2019 and 2020 undertaken by Foodbank that had confirmed if an individual was food insecure and tagged to SA2 and to the segments | Apply a calculation of demand to each community in Australia that would be a statistically accurate representation of demand.<br><br>1. In developing a gravity score, each SA1 was rated based on the density of agencies and the volume of food supplied within each SA1. 8 different scores were developed ranging from high 80% dominance score to low 20% dominance score) in areas were Foodbank were currently supplying. Relationship between dominance score, demand drivers (as defined in Step 5) and volumes were defined for the primary dominance score in terms of KG per capita and that relationship was projected to the balance of SA1s.<br><br>Aggregated KGs in total up to the SA2 and divided by total population of SA2 providing a KG per capita per SA2. This was used to validate the demand drivers for the segmentation model and consistency of KGs per capita per segment across Australia.<br><br>2. The surveys provided an additional source of data to verify the demand estimates by geography as above. The survey supplied us key demographics that were weighted against the segments themselves and that in turn verified the demand estimates within our segments. |

| | Description | Process | Commentary |
|---|---|---|---|
| 8 | Quantifying the amount of food insecure people | Leveraging the recipient surveys from 2018,2019 & 2020 to understand the frequency of food insecurity (e.g. everyday, fortnightly, monthly etc) by segment. NB the 2018 Survey incorporated the USDA model for measuring frequency of need. | Evidence from Foodbank suggested that a meal was approximately .555kgs and that a respondent required 1 meal per time. This was weighted by frequency of need (i.e Daily, weekly etc sourced from survey) per segment and that allowed us to quantify the number of people in need by area.<br><br>The above metrics were applied to further understand demand estimates both as individuals and as total volumes.l. |

| | Description | Process | Commentary |
|---|---|---|---|
| 9 | Identify gaps between demand and supply | Calculate the difference between current Foodbank supply to each SA2 and the estimated demand for food in each SA2/segment - expressed in KGs and % of population experiencing food insecurity in KGs and % of population experiencing food insecurity | This will identify where Foodbank supply is reaching suitable levels to match demand, where their supply is short (and by how much) and areas of need/demand currently not serviced by Foodbank. |

## Strengths and Weaknesses in the Methodology

Key strengths include the integration and use of multiple data sources to the model. DSS data for example provides us a currency to understand changes in small geographic areas that impact demand. This information is supplied by the government each quarter and provides valuable insights into the economic status of each SA2. Regular analysis and comparison to agency demand trends means that Foodbank can respond to variations in need over time.

The segmentation should be reviewed regularly (maybe annually) based on changes to DSS, population patterns, survey data, new data sources (as yet unused-unknown) and changes to demand/requests by individual agencies. This will also have an impact on Foodbanks Area of Influence and therefore the predictive value of the model.

At the moment the survey data did not cover all segments with an adequate sample size so the survey average was applied. This is easily resolved with modifications to the sample frame design in future surveys, but does raise some concerns as to the accuracy of estimates of demand/need. The model will allow us to test assumptions and test logistical strategies over time.

An additional strength is that the methodology is open to integration and enhancement with additional small area data bases.

The total database library at the SA2 level should/could include

- Population estimates x age x sex @ SA2 level
- Population forecasts base sourced through AIHW @SA2 level
- Census and census  time series data to show the overall demographic trends.(this includes data modelled from census such as SEIFA)  at SA1 and SA2 level
- DSS data at SA2 level
- Geocoded agency distribution and gravity model output as an attribute to define areas of influence and degree of regional dominance.
- Distance Matrix from Agencies to SA2
- Dwelling approvals at SA2 level to define economic activity in the small area
- Counts of Australian Businesses x number of employees and turnover at SA2