**Understanding the market in Canberra, Australia**

*Jacob Siddall*

April 2019

1. **Introduction**
2. *Background*

Canberra is the capital city of Australia and has a population of approximately 400,000. By 2058, the city’s population is projected by local government officials to reach 700,000. As a result, business interest in the region is growing in the form of thriving apartment developments and a growing food and entertainment scene.

1. *Problem*

I have been approached by a group of businesses, from overseas, interested in understanding (1) where younger adult/medium-high income Canberrans cluster and (2) existing competition in the area. Analysing these two issues together will provide business intelligence to inform the selection of optimal locations for establishing new business enterprises in the Canberra area.

1. **Data collection and cleaning**
2. *Data sources*

Two data sources were used for this analysis:

1. [2011 Australian Bureau of Statistics (ABS) census data](http://www.abs.gov.au/)

The ABS conducts a national census of the Australian population approximately every five years. Unfortunately, the ABS does not yet have a functioning Automated Programming Interface (API) for the dataset of interest. Instead, data was outputted from the ABS website into two Excel CSV files using the ABS ‘TableBuilder’ product – namely ‘proportion (%) of population by Canberra postcode and income bracket’ and ‘proportion (%) of population by Canberra postcode and age bracket’.

1. [Foursquare location data](https://developer.foursquare.com/)

In order to identify already existing local business competition within the food and entertainment industry, the Foursquare API was used to retrieve relevant location data by Canberra postcode.

1. *Data cleaning and feature selection*

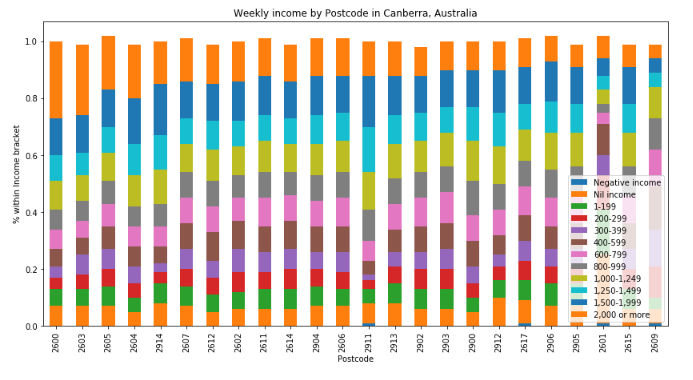
Throughout the notebook for this project, there is many variations of the same data described above. Data cleaning and manipulation was performed on both income and age ABS datasets to prepare them for various forms of analysis. This included:

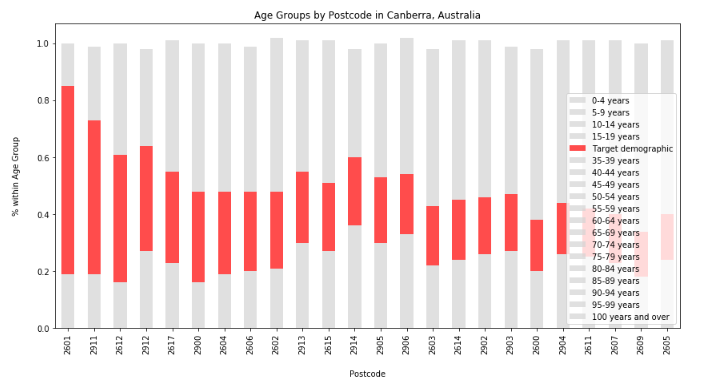
* Transforming data types
* Pivoting/stacking data tables
* Renaming and reordering columns
* Applying basic calculations across cases (e.g. sum)
* Dropping redundant features
* Computing new variables required for the analysis

1. **Exploratory Data Analysis**

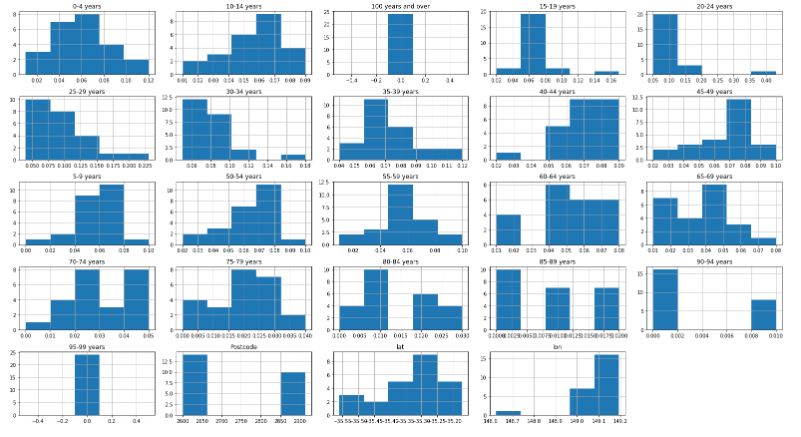
A number of exploratory techniques were applied to the prepared datasets, including:

1. *Graph representation*

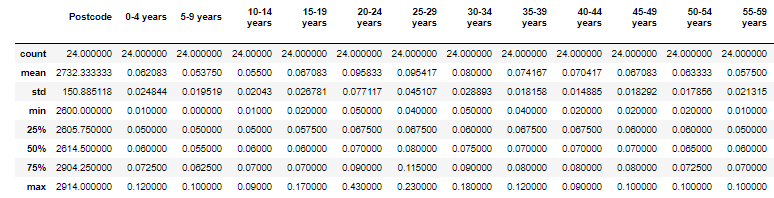


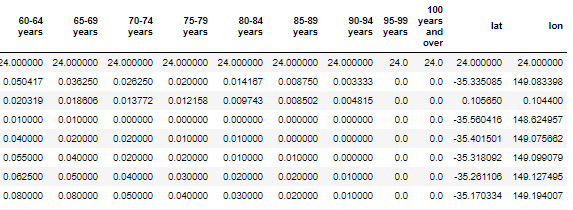


1. *Frequency distributions*



1. *Data descriptions*



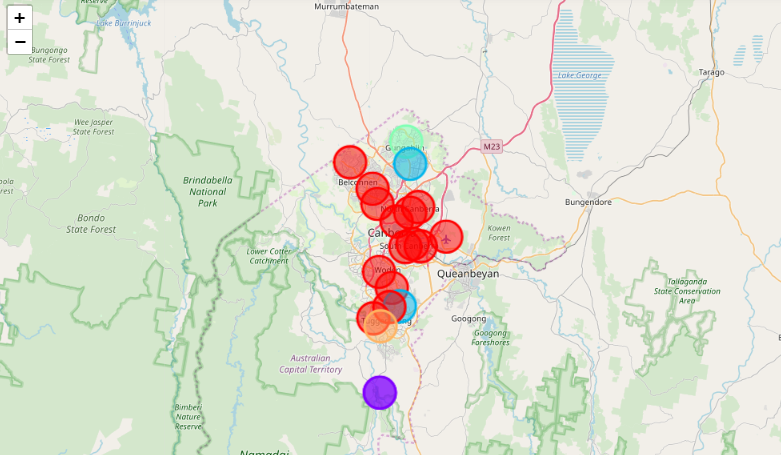


1. **Segmentation and Clustering**

Following the exploration of the datasets, segmentation and k-means clustering was applied to cluster the postcodes of Canberra into alike groups.

1. *Mapping the market competition*

First, market competition in the region was discovered through the use of Foursquare data and geographic mapping:

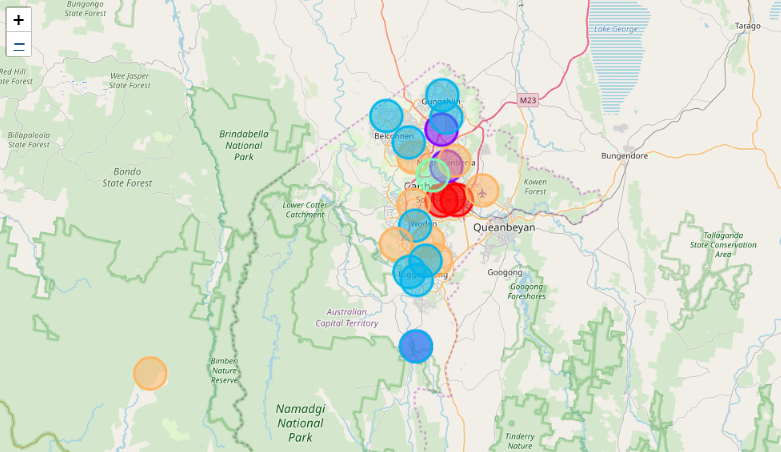


What was interesting in this analysis was the five clusters that emerged:

* + Red seemed to indicate a proximity (within 2,000 metres) to places like hotels, cafes and restaurants.
  + Light blue indicated proximity to lakes and fast food restaurants.
  + Green indicated a single unique postcode more proximate to places such as athletics and sports venues, rather than cafes/hotels.
  + Purple indicated an outlier in the facilities available, probably due to remoteness (being located far away from the city’s centre).
  + Orange was an oddity, with the most frequent venue being cupcake shops and wine bars. As a local who is familiar with the southern Tuggeranong area, this makes sense.

1. *Clustering and mapping by demographic features (income and age)*

Then, the demographic data collected from the ABS website was run through the clustering algorithm and mapped accordingly.



1. **Recommendations**

My recommendations to the interested business, based on the above analysis (the details of which can be found throughout the Jupyter Notebook) are to:

1. ***Explore opportunities in the Gungahlin region of Canberra.***

Gungahlin market competition in the food and entertainment industry if comparatively lower than other Canberra regions. At the same time, the demographic makeup of Gungahlin is like Belconnen, in that it is mostly middle income older and younger adults.

1. ***Avoid the market saturated areas of Civic and Kingston.***

Although these areas attract high-income cohorts there is already an established food and beverage industry here.

1. ***Avoid areas marked in orange on the second map, if your target market is high-income earners.***

While business opportunities can be found across income brackets, if your target group is high-income earners the orange clusters have identified low-middle income areas.