Finding the Best Location to Live

for new migrant families

coursera APPLIED DATA SCIENCE

Capstone Project

Kamin Arifin

6 June 2019

Agenda

- Introduction
- Data Sources
- Data Analysis

Introduction

As an immigrant myself, I still remember very clear the struggle that I was experiencing when I migrated to Australia. My family and I chose Melbourne as a new place we call home. Of course, the first mission which I needed to accomplish was to find a roof for my family and I to live under.

As new comer, most people will like to live in an area that they feel more familiar or comfortable, eg: in the area where they have the same background, either etnics, religion or language.

In this project, I will use k-Means Clustering to provide a better picture as suburbs in Melbourne.

Data Source

The main sources of data that I will use for this project will be from Australian Bureau Statistics in Excel Format:

- Population and People, ASGS, 2011 to 2018
- Education and Employment, ASGS, 2011 to 2018
- List of Venues (Using Foursquare)

For Geocoders I will use GeoPy. It has simple interface and easy to be used. It's able to find the coordinates of addresses, cities, or even to measure distance of one location to the other.

Data Analysis

ABS provides a very rich statistic data of Australia population. From those data we will be able to review their ages, educations, religions, country of origins of each suburbs (areas).

By using Foursquare, we will able to see the most popular venues in those suburbs.

By combining all those data, we will have a complete picture of suburb.

Data Analysis (Continue)

Our main goal is to see the which suburbs that will suit most to the new immigrant that's trying to find a new place to live.

By clustering the suburbs based on the interest of that family I hope they can see very clear the profile of each suburbs and make the correct decision.

For Geocoders I will use GeoPy. It has a simple interface and easy to be used. It will help me to find the coordinates of addresses, cities, also to measure distance of one location to the other.

Sample Data

CODE	Suburb	YEAR	Median Age	Born_in_Oceania_Ex_OZ	Born_in_North_West_Europe	Born_in_Southern_Eastern_Europe	Born_in_North_Africa_Middle_East	Born_in_South_East_Asia	Bo
201011001	Alfredton	2016	34.1	1.1	3.1	0.8	0.4	1.4	
201011002	Ballarat	2016	41.5	1.2	4	0.8	0.3	1	
201011003	Ballarat - North	2016	38.7	0.7	3.6	0.8	0.2	0.8	
201011004	Ballarat - South	2016	35.3	0.9	3.7	0.7	0.5	0.9	
201011005	Buninyong	2016	38	0.9	5.5	0.7	0.3	0.5	
201011006	Delacombe	2016	35	0.8	3	0.8	0.4	0.7	
201011007	Smythes Creek	2016	39.9	0.8	4.5	1.2	0.1	0.3	
201011008	Wendouree - Miners Rest	2016	37.9	0.7	3.1	0.5	0.2	1.2	
201021009	Bacchus Marsh	2016	42.8	1.1	7.4	2.1	0.2	0.4	