The Battle of Neighbourhoods

Where to open a new Korean restaurant in Melbourne

R Nakaichi

18 July 2020

1 Introduction

Melbourne where I live is one of the most culturally diverse cities in the world. There are plenty of multicultural restaurants available in Metropolitan Melbourne. However, the level of the competition between restaurants is also one of the world's highest.

I grew up in Japan and spent more than a half of my life there. I moved to Melbourne in the recent years, and I was surprised by the variety of restaurants in Melbourne. Then, Korean cuisine has become one of my favourite cuisines here. The quality of Korean restaurants in Melbourne is remarkably high due to the high level of the competition. Therefore, opening a new restaurant in Melbourne is not easy, and it is crucial where to open a new restaurant because of high competitions.

For this capstone project, I will research where to open a new Korean restaurant in Melbourne, which will help business owners to start a new Korean restaurant in Melbourne.

2 Data

2.1 Data source

Australia conduct a census every five years. It collects key characteristic data on every person in Australia, which includes the data such as age, gender, income, ancestry, education, language spoken, family composition, employment, mortgage repayment and so on. The last census was conducted in 2016, and the 2016 census data is publicly available at the following link.

Australian Bureau of Statistics – Census DataPacks
https://datapacks.censusdata.abs.gov.au/datapacks/

An interesting part of the census data is that it includes "Ancestry by Country of Birth of Parents". If I can find a suburb where a lot of Korean people live, but there are not many Korean restaurants in the suburb. That could be a good place to open a new Korean restaurant.

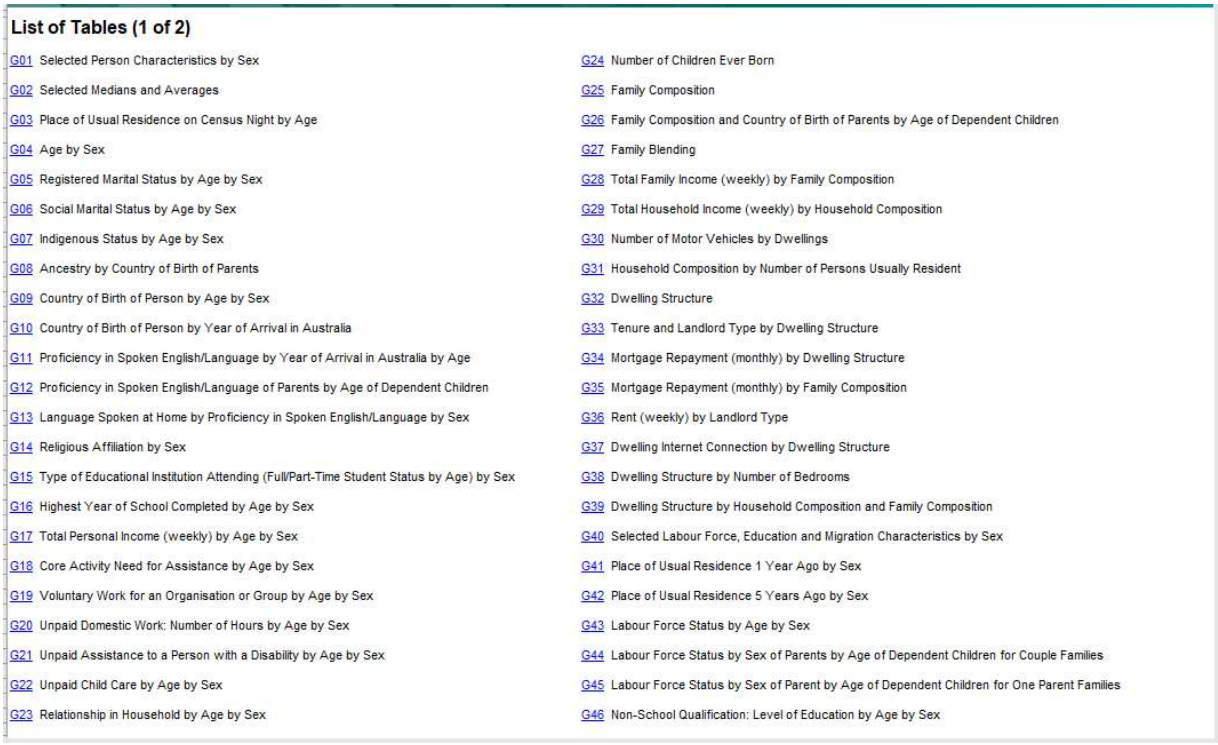
The data of restaurant locations can be obtained by Foursquare which is one of the best location data platforms. See the link below for details.

https://foursquare.com/

2.2 Data cleaning

2.2.1 Data selection from DataPacks

The census DataPacks are large and include 59 tables. Figure 1 shows a part of the tables in the DataPacks. Since the data is huge and processing everything will go beyond of this project's scope, I decided to use only the data shown in Table 1 from the DataPacks.



List of Tables (1 of 2)

G01 Selected Person Characteristics by Sex	G24 Number of Children Ever Born
G02 Selected Medians and Averages	G25 Family Composition
G03 Place of Usual Residence on Census Night by Age	G26 Family Composition and Country of Birth of Parents by Age of Dependent Children
G04 Age by Sex	G27 Family Blending
G05 Registered Marital Status by Age by Sex	G28 Total Family Income (weekly) by Family Composition
G06 Social Marital Status by Age by Sex	G29 Total Household Income (weekly) by Household Composition
G07 Indigenous Status by Age by Sex	G30 Number of Motor Vehicles by Dwellings
G08 Ancestry by Country of Birth of Parents	G31 Household Composition by Number of Persons Usually Resident
G09 Country of Birth of Person by Age by Sex	G32 Dwelling Structure
G10 Country of Birth of Person by Year of Arrival in Australia	G33 Tenure and Landlord Type by Dwelling Structure
G11 Proficiency in Spoken English/Language by Year of Arrival in Australia by Age	G34 Mortgage Repayment (monthly) by Dwelling Structure
G12 Proficiency in Spoken English/Language of Parents by Age of Dependent Children	G35 Mortgage Repayment (monthly) by Family Composition
G13 Language Spoken at Home by Proficiency in Spoken English/Language by Sex	G36 Rent (weekly) by Landlord Type
G14 Religious Affiliation by Sex	G37 Dwelling Internet Connection by Dwelling Structure
G15 Type of Educational Institution Attending (Full/Part-Time Student Status by Age) by Sex	G38 Dwelling Structure by Number of Bedrooms
G16 Highest Year of School Completed by Age by Sex	G39 Dwelling Structure by Household Composition and Family Composition
G17 Total Personal Income (weekly) by Age by Sex	G40 Selected Labour Force, Education and Migration Characteristics by Sex
G18 Core Activity Need for Assistance by Age by Sex	G41 Place of Usual Residence 1 Year Ago by Sex
G19 Voluntary Work for an Organisation or Group by Age by Sex	G42 Place of Usual Residence 5 Years Ago by Sex
G20 Unpaid Domestic Work: Number of Hours by Age by Sex	G43 Labour Force Status by Age by Sex
G21 Unpaid Assistance to a Person with a Disability by Age by Sex	G44 Labour Force Status by Sex of Parents by Age of Dependent Children for Couple Families
G22 Unpaid Child Care by Age by Sex	G45 Labour Force Status by Sex of Parent by Age of Dependent Children for One Parent Families
G23 Relationship in Household by Age by Sex	G46 Non-School Qualification: Level of Education by Age by Sex

Figure 1 Census Datapacks - List of Tables

Table 1 Data used in this project

No	Table No	Data description
1	G01	Total Population
2	G02	Median_age_of_persons
3	G02	Median_total_household_income_weekly
4	G08	Korean ancestry population
5	G08	Chinese ancestry population
6	G08	Australian ancestry population
7	G08	English ancestry population
8	G08	Vietnamese ancestry population
9	G08	Indian ancestry population

2.2.2 Data Wrangling

2.2.2.1 Reading files and merging

The necessary data was spread over the following files.

- 2016Census_G01_VIC_SSC.csv
- 2016Census_G02_VIC_SSC.csv
- 2016Census_G08_VIC_SSC.csv

Therefore, these files were imported by *pandas* individually, and then merged. But there was an issue with the csv files. The csv files only contain state suburb codes, not suburb names. The suburb codes were defined in another excel file called "2016Census_geog_desc_1st_and_2nd_release.xlsx". Therefore, this excel file was also read by *pandas*, and then merged.

2.2.2.2 Getting Latitude/Longitude

In order to use *Foursquare API*, the latitude and longitude of each suburb are required, but the census Datapacks do not include them. To get the latitude and longitude, *Nominatim* library was used. The data has 2931 suburbs, and it will take a considerable amount of time if I process all suburbs. Therefore, I excluded the suburbs less than 10,000 population. This brought down the number of the suburbs to 195.

2.2.2.3 Getting nearby Korean restaurants

Foursquare API was used to get the information of nearby Korean restaurants in each suburb. The *search* command with the category ID was used for this inquiry. The radius was set to 1,000 [m]. Then, the number of nearby Korean restaurants was calculated by using *value_counts()* function in *pandas*.

2.2.2.4 Cleaned data

The cleaned data looks like the following.

	Total Population	Median Age	Median Household Income	Korean Ancestry Population	Chinese Ancestry Population	Australian Ancestry Population	English Ancestry Population	Vietnamese Ancestry Population	Indian Ancestry Population	Suburb Name	Population Density	latitude	longitude	Number of Korean Restaurant
0	10762	42	1400	13	364	2972	3351	151	245	Altona	642.991146	-37.867206	144.830142	0.0
1	19160	38	1370	27	858	4299	4290	614	902	Altona Meadows	1817.871307	-37.881442	144.784548	0.0
2	12152	38	1240	14	500	2271	2089	551	263	Altona North	871.467194	-37.837823	144.834285	0.0
3	14750	35	1757	31	810	3567	3977	540	248	Ascot Vale	3845.050963	-37.775316	144.921849	3.0
4	11633	44	1318	18	785	1675	1521	827	231	Avondale Heights	2150.794091	-37.761454	144.862141	0.0

Figure 2 Data after data cleaning

3 Methodology

3.1 Understanding Data / Visualisation

To understand the data better and get some new ideas about the data, the data visualisation was performed. I used the *folium* library to visualise the clusters of Korean ancestry population and the number of Korean restaurants in Melbourne. I also visualised the Median Household Income and the number of Korean restaurants in Melbourne in a map.

3.1.1 Korean ancestry population and the number of Korean restaurants

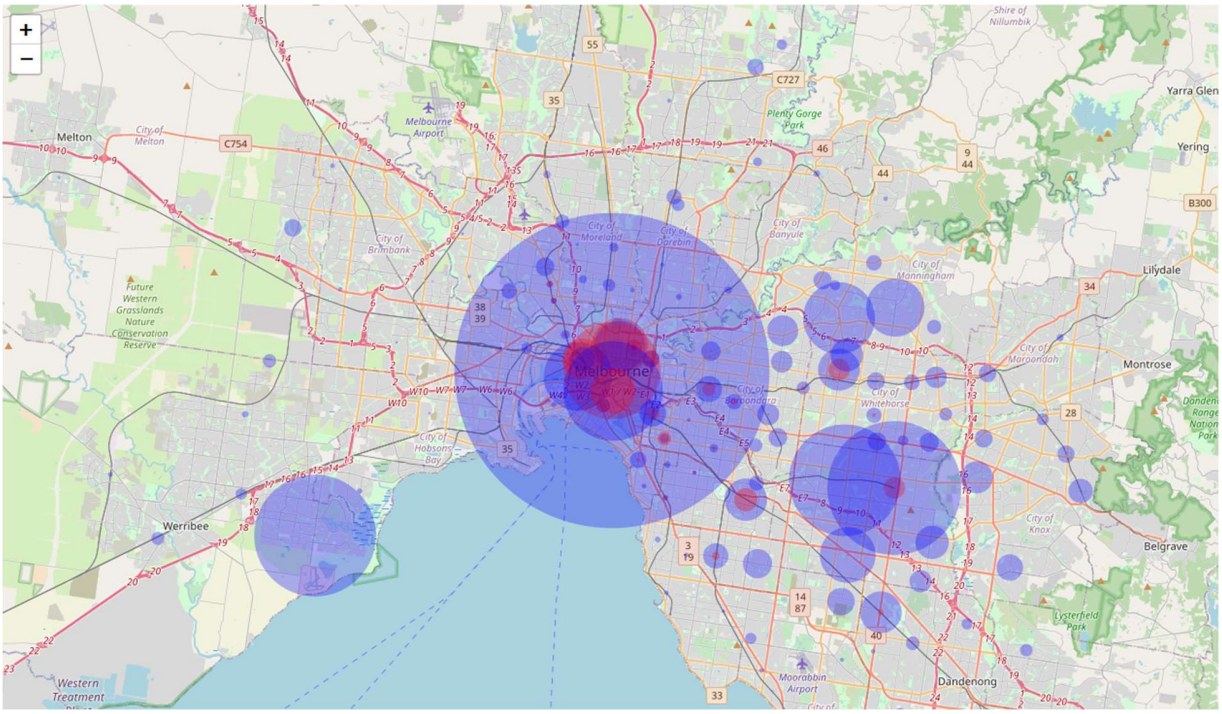


Figure 3 Korean ancestry population (Blue) and the number of Korean Restaurant (Red)

This map tells us a lot of Korean people live in the city area, and a few suburbs in the west and east also have quite a few of Korean people. However, in the city area, there are also many Korean restaurants around, and the competition around the city area would be extremely high. Those suburbs in the west or east might be a good place to open a new restaurant.

3.1.2 Median household income and the number of Korean restaurants

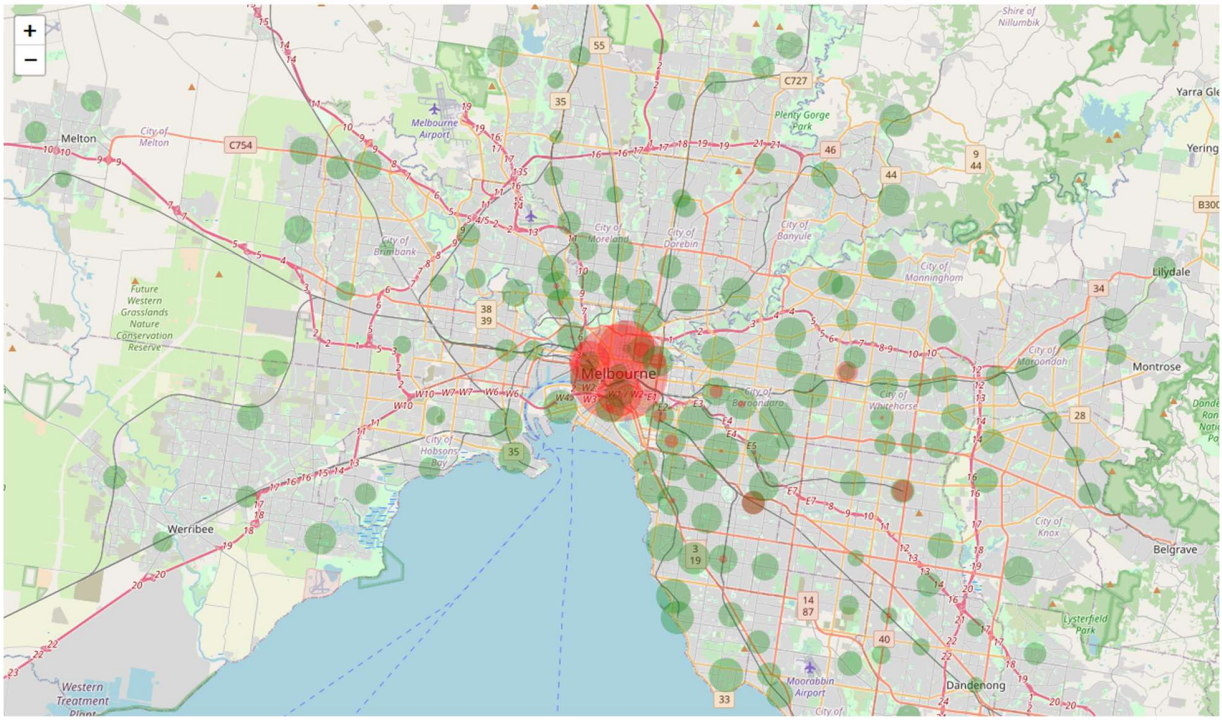


Figure 4 Median household income (Green) and the number of Korean Restaurant (Red)

This map does not tell us much clues than the first map, but it looks like people with high income live in south and south east part near the city. The suburbs where I found many Korean people live in the Figure 3 are not particularly rich suburbs, but also not poor suburbs.

3.2 Correlation analysis

To find out correlation between the data, the Pearson correlation coefficient was calculated by using corr*()* method in *pandas* and visualised as a heatmap by using *seaborn*. This is mainly to confirm my initial intuition to find a suburb where a lot of Korean people live but not many Korean restaurants in the suburb. This intuition could be totally misguided and there is no correlation between Korean restaurants and Korean people.

3.2.1 Correlation matrix

	Total Population	Median Age	Median Household Income	Korean Ancestry Population	Chinese Ancestry Population	Australian Ancestry Population	English Ancestry Population	Vietnamese Ancestry Population	Indian Ancestry Population	Population Density	Number of Korean Restaurant
Total Population	1.000000	-0.248410	-0.057841	0.397345	0.403362	0.656163	0.650710	0.205521	0.665645	-0.024063	0.154254
Median Age	-0.248410	1.000000	0.136355	-0.209042	-0.168580	0.104654	0.132578	-0.183005	-0.449924	-0.371672	-0.351008
Median Household Income	-0.057841	0.136355	1.000000	0.009196	-0.017586	-0.019256	0.059654	-0.215950	-0.021846	0.234642	-0.079569
Korean Ancestry Population	0.397345	-0.209042	0.009196	1.000000	0.885627	-0.038035	0.040593	0.049614	0.344185	0.347286	0.738716
Chinese Ancestry Population	0.403362	-0.168580	-0.017586	0.885627	1.000000	-0.128681	-0.052722	0.189576	0.362212	0.321356	0.617707
Australian Ancestry Population	0.656163	0.104654	-0.019256	-0.038035	-0.128681	1.000000	0.973171	-0.224824	0.126154	-0.343849	-0.150800
English Ancestry Population	0.650710	0.132578	0.059654	0.040593	-0.052722	0.973171	1.000000	-0.230612	0.095309	-0.228732	-0.068016
Vietnamese Ancestry Population	0.205521	-0.183005	-0.215950	0.049614	0.189576	-0.224824	-0.230612	1.000000	0.221125	0.075083	0.019926
Indian Ancestry Population	0.665645	-0.449924	-0.021846	0.344185	0.362212	0.126154	0.095309	0.221125	1.000000	-0.026988	0.107220
Population Density	-0.024063	-0.371672	0.234642	0.347286	0.321356	-0.343849	-0.228732	0.075083	-0.026988	1.000000	0.619248
Number of Korean Restaurant	0.154254	-0.351008	-0.079569	0.738716	0.617707	-0.150800	-0.068016	0.019926	0.107220	0.619248	1.000000

Figure 5 Correlation Matrix

3.2.2 Correlation heatmap

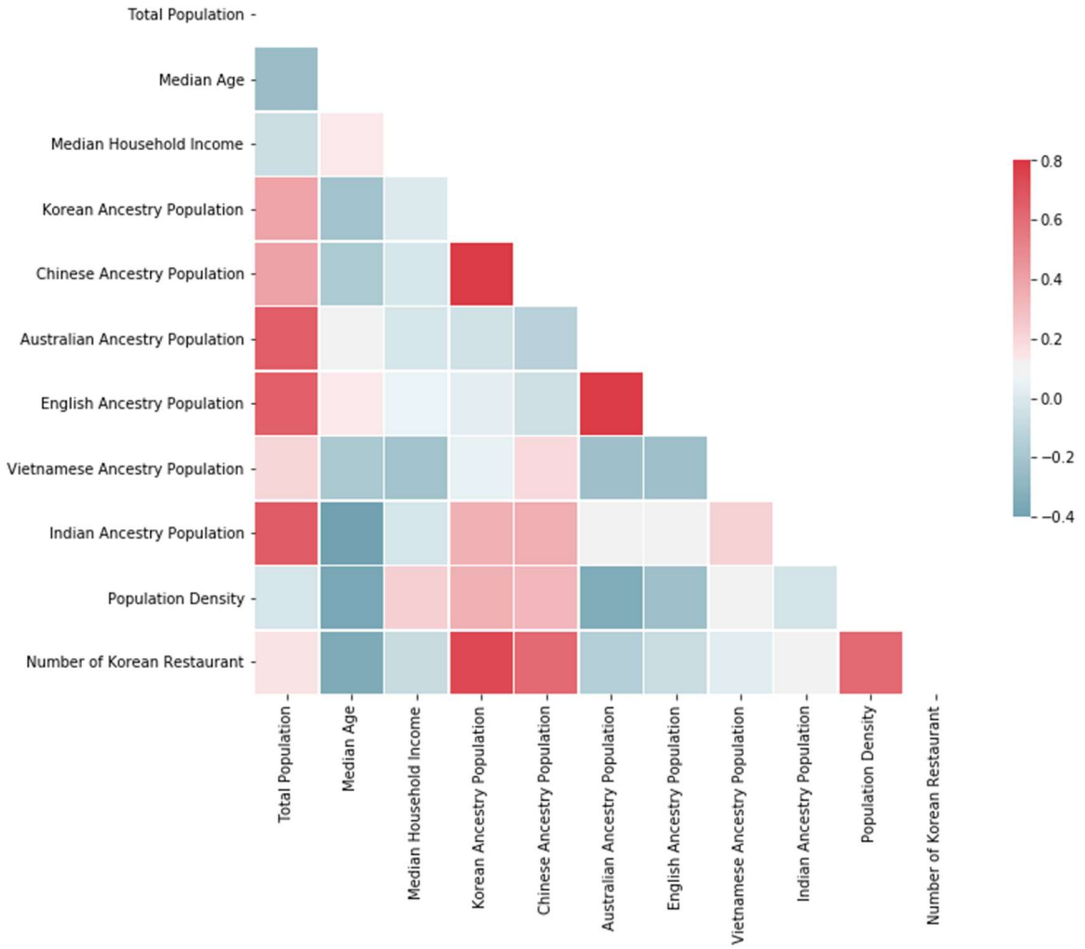


Figure 6 Correlation Heatmap

The correlation heatmap tells us intuitive correlation between data sets while the correlation matrix tells us the exact number of the coefficient.

I can see some interesting correlations through these figures.

1. Korean ancestry population and the number of Korean restaurants have a strong correlation. This means my initial guess was correct.
2. Chinese ancestry population and the number of Korean restaurants have a strong correlation too. Chinese people might like Korean food too. In fact, Korean TV drama is very popular in China.
3. The other ancestry population such as Australian, English, Vietnamese and Indian do not have a correlation with the number of Korean restaurants.
4. Korean ancestry population and Chinese ancestry population have a strong correlation. Maybe Chinese and Korean people tend to live in the similar area.
5. Median age and the number of Korean restaurants have a moderate negative correlation. It hints younger generation prefer Korean food more.
6. Population density and the number of Korean restaurants have correlation. This simply means there are more restaurants where there are more people.
7. English ancestry population and Australian ancestry population have a strong correlation. This has nothing to do with Korean restaurants, but I can confirm they have been living in the same area.

4 Results

Now I can look at the correlation between Korean ancestry population and the number of Korean restaurants closely. To be able to identify those suburbs with more Korean people but less Korean restaurants, I chose to use the scatter plot. In another way, it is possible to calculate Korean population per Korean restaurant, but the scatter plot is more visually appealing.

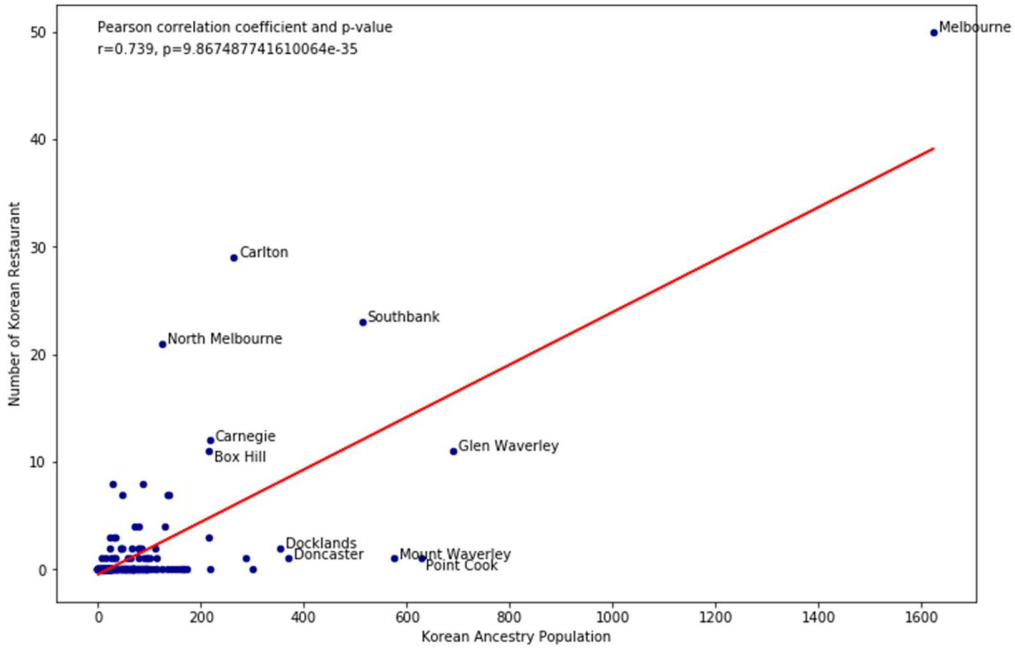


Figure 7 Scatter Plot Korean ancestry population vs the number of Korean restaurants

The blue dots indicate suburbs, and the red line is the regression line. A Pearson correlation coefficient and the p-value are calculated. The p-value is extremely small; therefore, I can confirm this correlation is genuine.

To determine where to open a new Korean restaurant, I can use the regression line. Any suburbs below the red line indicate more Korean people but less Korean restaurants. Therefore, the following suburbs are good places to open a new Korean restaurant.

- Point Cook
- Mount Waverley
- Glen Wavery
- Docklands
- Doncaster

On the other hand, the following suburbs are not good places to open a new Korean Restaurant.

- Melbourne
- Carlton
- South Bank
- North Melbourne

- Carnegie
- Box Hill

5 Discussion

I found the suburbs which could be a good place to open a new Korean restaurant in the previous section. In this section, I will discuss some observations I made during the data wrangling/processing and some future improvements.

5.1 Observations and Future improvements

5.1.1 Limitations in Foursquare API search

I can get only up to 50 results by Foursquare API search. This resulted capping the number of restaurants at 50 in each suburb. Luckily, it only happened to one suburb, Melbourne. Therefore, the numbers in Melbourne could be much higher. It seems there is no way to increase this number in Foursquare API, and if I want to get more than 50, I must use multiple latitudes/longitudes in each suburb.

5.1.2 The category of Korean restaurant

There are many different categories in Korean cuisine such as Korean traditional cuisine, Korean fried chicken, and Korean barbecue. I was not able to get this much details of restaurant category through Foursquare API. When opening a restaurant, it is probably ok to have other categories of Korean restaurant nearby but not the same category of Korean restaurant. If I can get more specific information of the category, the analysis can be improved further.

5.1.3 Suburb size and searching radius

I noticed that there were big suburbs and small suburbs during data processing. I used the radius size of 1000m when I searched Korean restaurants through Foursquare API. This method is not fair to all suburbs because I cannot find all the restaurants, particularly in a big suburb. Ideally, this searching method needs to be improved by performing multiple latitudes/longitudes search depending on the size of suburbs. The radius size needs to be optimised too.

5.1.4 Census data

The last census was conducted in 2016 which is 4 years ago. Therefore, the data is slightly outdated and might not be accurate enough. Possibly, there might be better suburbs now. The next census will be conducted in 2021 which is next year. It would be interesting to see the 2021 census data when the data is available.

6 Conclusion

In this research, I investigated where to open a new Korean restaurant in Melbourne. I used the 2016 census data and Foursquare API to find a suburb where a lot of Korean people live but not many Korean restaurants in the suburb. I found a couple of suburbs which could be good places to open a Korean restaurant, which is Point Cook, Mount Waverley, Glen Wavery etc. In the future, this research can be further improved by refining the searching method and using the 2021 census data.