

Chinese Restaurants in Sydney

Table of contents

- * Introduction
- * Data
- * Methodology
- * Analysis
- * Results and Discussion
- * Conclusion

Introduction

Business Problem

Sydney is a gateway to Australia for many international visitors. It has hosted over 2.8 million international visitors in 2013, or nearly half of all international visits to Australia. These visitors spent 59 million nights in the city and a total of \$5.9 billion. (Wikipedia)

In this project we will try to analyse the distribution of Chinese restaurants in Sydney and try to find a possible optimal location for a new Chinese restaurant.

This report is expected to provide information on Chinese restaurant in Sydney, Australia. Sydney is the largest and the most populous city of Australia, and also the capital of New South Wales. The population of today's Sydney is approaching to 5 million people. Since there are lots of Chinese restaurants in Sydney, we will only focus on the area with postal code from 1000 to 2249 in Sydney. We will use Foursquare to detect Chinese restaurant and its locations.

Data

Based on definition of our problem, factors that will influence our decision are:

- * number of existing Chinese restaurants in the neighborhood
- * distance to Chinese restaurants in the neighborhood
- * distribution of Chinese restaurants in Sydney

We downloaded Australian postcodes from Github.

<https://github.com/matthewproctor/australianpostcodes>

We This csv file contains

Table 1 Australian postcodes file description

Fieldname	Description
postcode	The postcode in numerical format - 0000 to 9999
locality	The locality of the postcode - typically the city/suburb or postal distribution centre
state	The Australian state in which the locality is situated
long	The longitude of the locality - defaults to 0 when not available
lat	The latitude of the locality - defaults to 0 when not available
dc1	The Australia Post distribution Centre servicing this postcode - defaults to blank when not available
type1	The type of locality, such as a delivery area, post office or a "Large Volume Recipient" such as a GPO, defaults to blank when not available
status	A note indicating whether the data is new, removed or updated - new column Nov 2018

We will only focus on postcodes from 1000 to 2249. Sydney, NSW, Australia Latitude and longitude coordinates are: -33.865143, 151.209900.

Following data sources will be needed to extract/generate the required information:
* longitude and latitude of candidate areas will be used to get Chinese restaurant list from Foursquare number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API.

Methodology

Review our sources of data We will review our data source and see what data preparation process is needed. We have two main sources of data:

- a. one CSV file gathering all zip codes in Australia and each corresponding latitude, longitude coordinates. This database is coming from
<https://github.com/matthewproctor/australianpostcodes>
- b. we will gather information by Foursquare API to get neighbourhood information for each (latitude, longitude) postal area in Sydney.

* Collect Data

As aforementioned, we can download Australian postcode file from github. Then we can import it into pandas dataframe. Fig 1 shows the raw data.

	postcode	locality	State	long	lat	id	dc	type	status
0	6532	CARRARANG	WA	115.004595	-28.440886	10861	GERALDTON DC	Delivery Area	NaN
1	6532	COBURN	WA	115.004595	-28.440886	10862	GERALDTON DC	Delivery Area	NaN
2	6532	COOLCALALAYA	WA	115.004595	-28.440886	10863	GERALDTON DC	Delivery Area	NaN
3	6532	DARTMOOR	WA	115.004595	-28.440886	10864	GERALDTON DC	Delivery Area	NaN
4	6532	DEEPDALE	WA	115.004595	-28.440886	10865	GERALDTON DC	Delivery Area	NaN

Fig 1 Australian Postcode

Then we need clear data and only keep the data with postcode from 1000 to 2249. Fig 2 shows the result.

	postcode	locality	long	lat	dc
14279	2205	ARNCLIFFE	151.147956	-33.937551	ROCKDALE DC
14280	2205	TURRELLA	151.147956	-33.937551	ROCKDALE DC
14281	2205	WOLLI CREEK	151.147956	-33.937551	ROCKDALE DC
14282	2206	CLEMTON PARK	151.122881	-33.926056	KINGSGROVE DC
14283	2206	EARLWOOD	151.122881	-33.926056	KINGSGROVE DC

Fig 2 Sydney Postcode

* Explore and Understand Data

In this part we use folium to visualize our data. Fig 3 shows all the check points we will use.

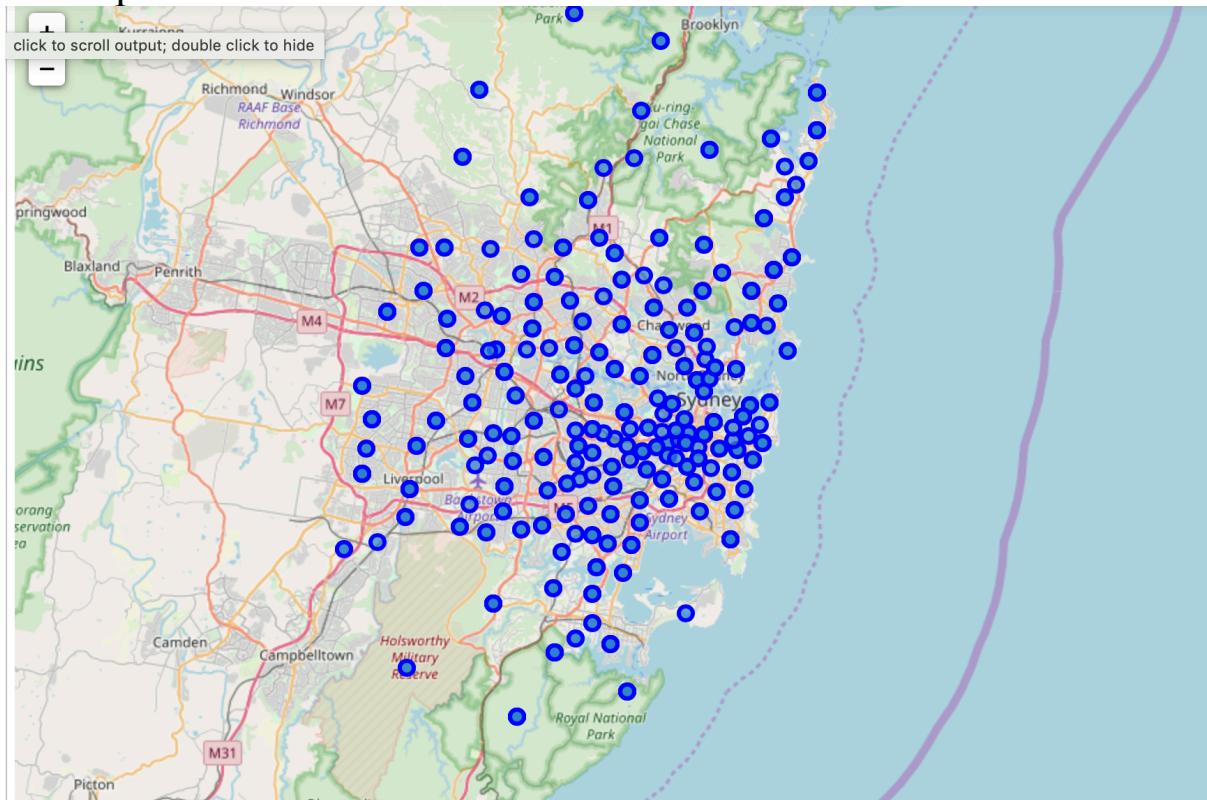


Fig 3 Check Points from Sydney Postcode

* Data Preparation and Preprocessing

Next, we are going to start utilizing the Foursquare API to explore Chinese restaurant and segment them. The total number of restaurants we get is 922. The total number of Chinese restaurants we get is 78. Percentage of Chinese restaurants is 8.46%. Average number of restaurants in each neighbourhood is 3.8. Fig 4 shows the restaurant locations. Note that blue point indicates a restaurant, whereas the red point indicates a Chinese restaurant.

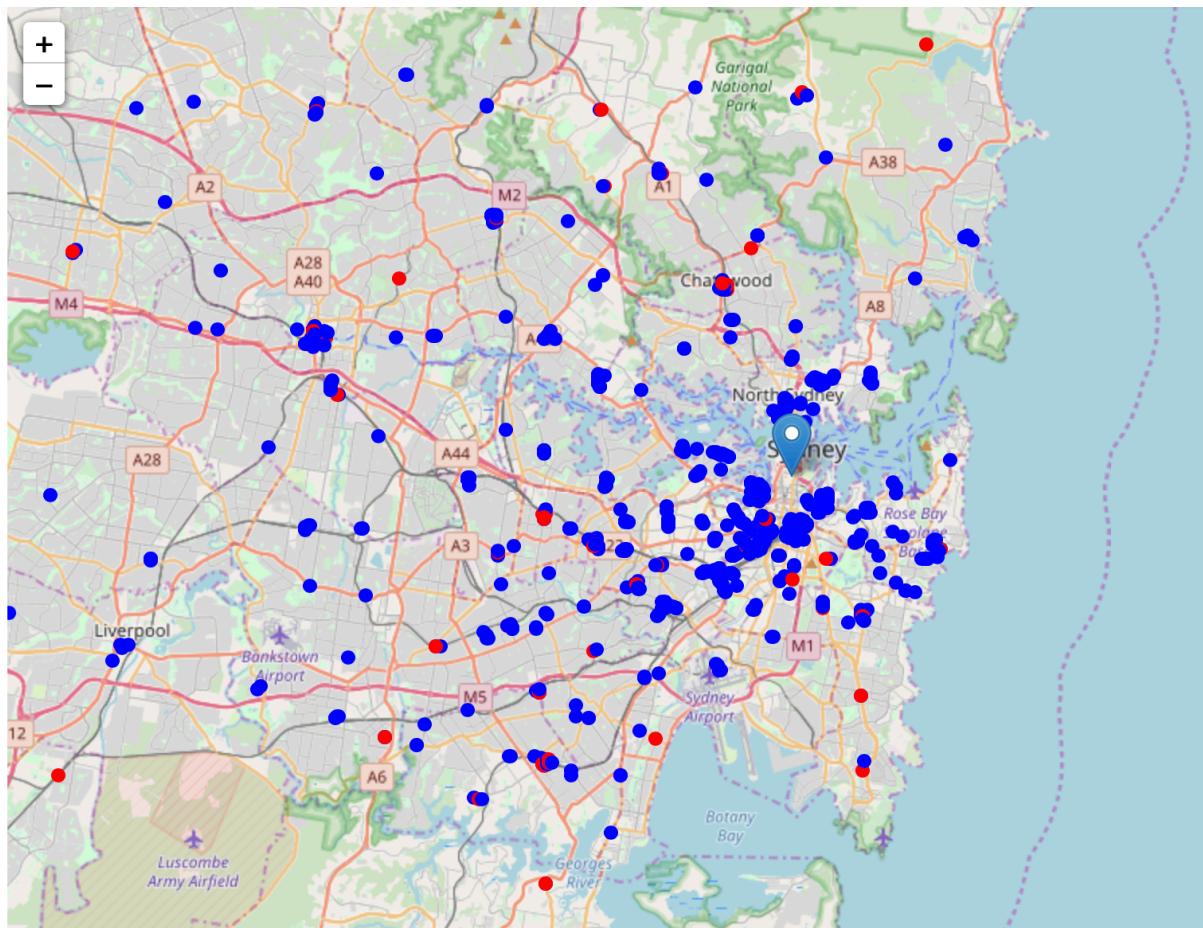


Fig 4 Restaurant data in Sydney

* Modeling

Analysis

Let's perform some basic explanatory data analysis and derive some additional info from our raw data. First let's count the number of restaurants in every area candidate and distance to nearest Chinese restaurant from every area candidate centre. Average distance to closest Chinese restaurant from each area centre: 3.7KM. Fig 5 shows the new data. Heatmap of density of all restaurants and Chinese Restaurant are shown in Fig 6 and 7.

postcode	locality	long	lat	dc	Restaurants in area	Distance to Chinese restaurant
14279	2205 ARNCLIFFE	151.147956	-33.937551	ROCKDALE DC	2	2.344885
14280	2205 TURRELLA	151.147956	-33.937551	ROCKDALE DC	2	2.344885
14281	2205 WOLLI CREEK	151.147956	-33.937551	ROCKDALE DC	2	2.344885
14282	2206 CLEMTON PARK	151.122881	-33.926056	KINGSGROVE DC	2	0.190975
14283	2206 EARLWOOD	151.122881	-33.926056	KINGSGROVE DC	2	0.190975

Fig 5 Calculated New DataFrame

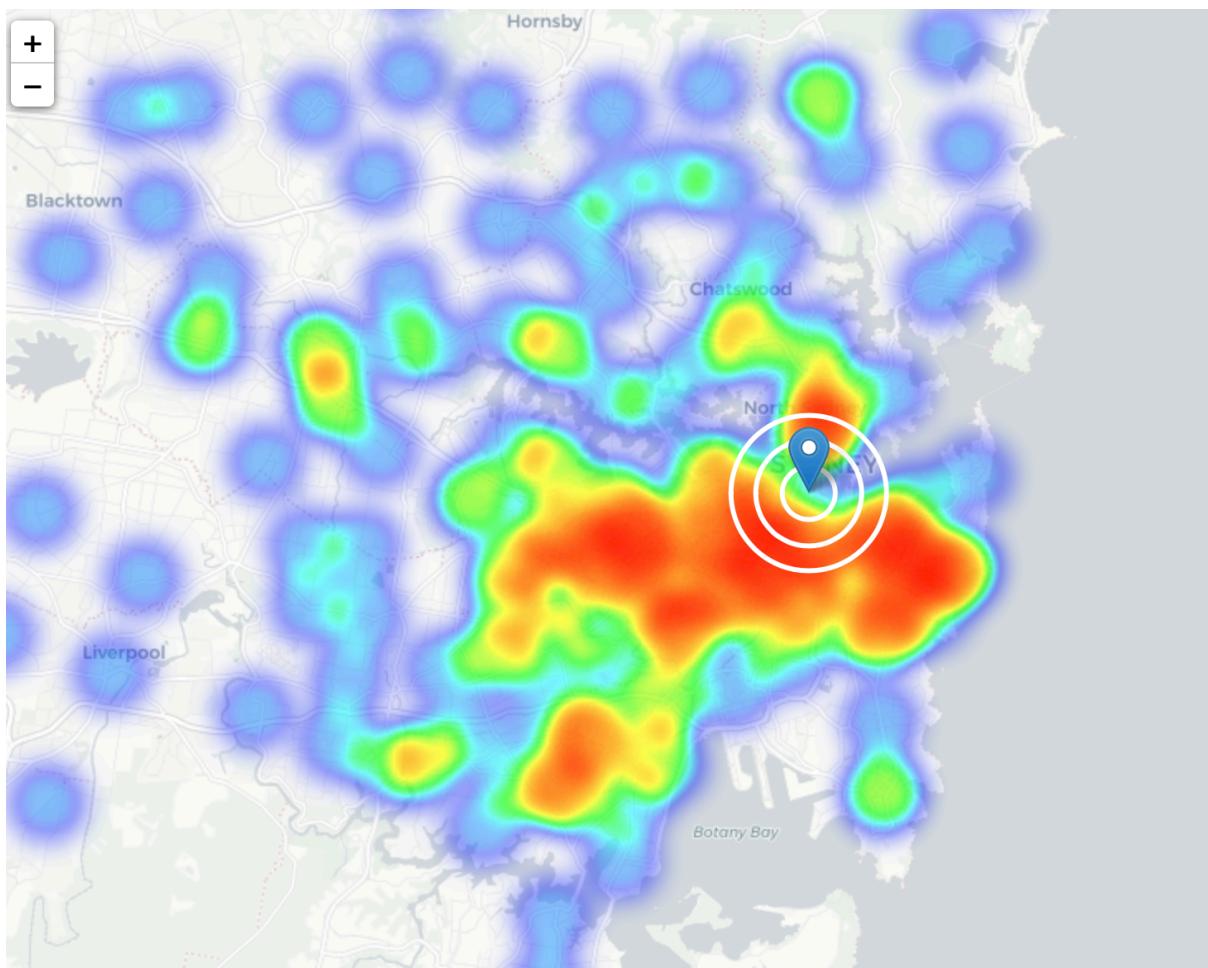


Fig 6 Heatmap of Restaurant Density in Sydney

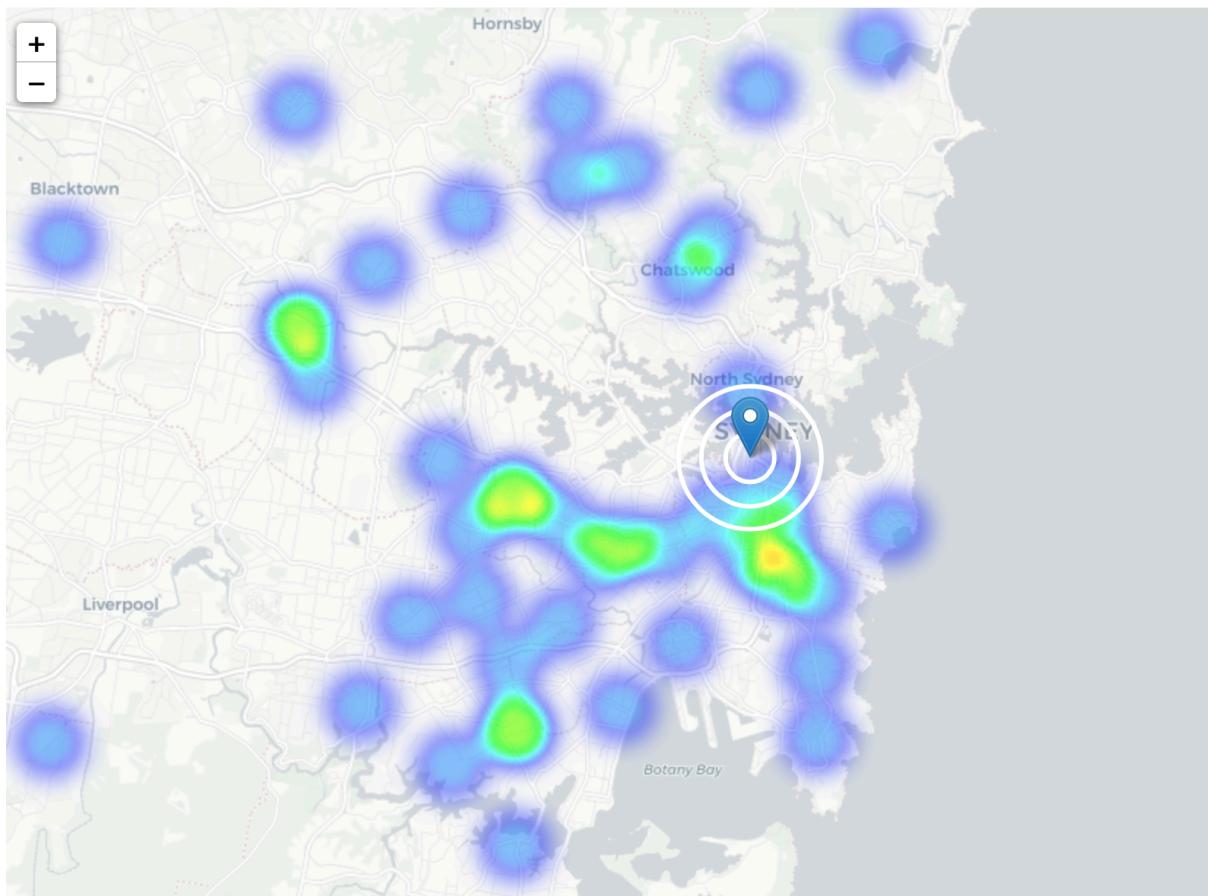


Fig 7 Heatmap of Chinese Restaurant Density in Sydney

From comparison, obviously we can find that Fig 7 map is not so 'hot'. This means that Chinese restaurant business is still underdeveloped in Sydney, especially in the skirt of city. Note that Chinese restaurants only represent a subset of ~8% of all restaurants in Sydney. In the meantime, Fig 7 also shows higher density of existing Chinese restaurants around Sydney centre, with low Chinese restaurant density positioned skirt of city.

Notice that we only have limited data in this analysis. If we have more data on population in different suburbs, we can consider the influence of different population density, especially for those suburbs which have higher ratio of Asian populations.

Also, if we have tourists data in Sydney, we can also consider the influence of the Asian tourists on Chinese restaurant business.

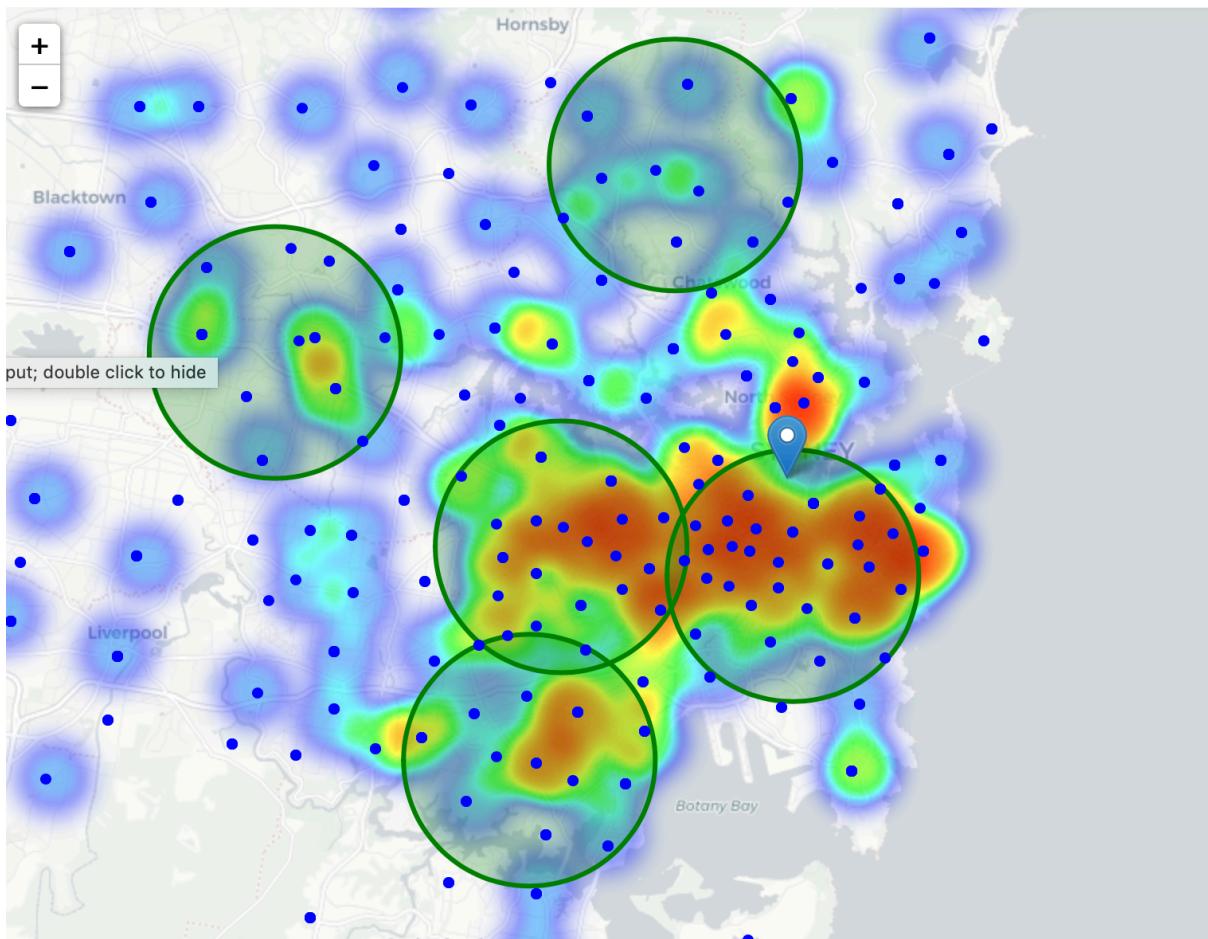


Fig 8 Clusters of Chinese Restaurants

We can use Kmeans method to cluster our data. Fig 8 shows the 5 clusters of Chinese Restaurants. Note that our clusters represent groupings of most of the Chinese restaurant locations and cluster centres are placed in the middle of the zones.

Results and Discussion

Our analysis shows that the number of Chinese restaurant in the skirt of Sydney is much lower than that in the centre of Sydney. Although a great number of restaurants in Sydney have been established, Chinese restaurant is still an underdeveloped business. We use the location information to group Chinese restaurant. There are mainly five groups.

We think the skirt of Sydney should be locations deserve more attention if you want start a new business. Due to limit data and time,

we can not consider more factors in this project. However, if possible, the following analysis should provide more valuable information for your decision.

First, restaurant's customers are humans. If we have more data on the population living in each suburb, we can take the number of potential customers as a very important factor in our analysis. Especially, Asian customers are key roles in Asian restaurant business.

Second, income is another important factor which we did not take into consideration in this project. We can analysis customer's behaviour to decide the menu of new restaurant.

Conclusion

Purpose of this project was to find out the distribution of Chinese restaurant in Sydney and provide information for people who want to start a new business in Sydney. Our analysis shows that the skirt of Sydney have less Chinese restaurant and should be pay more attention.

This project is just an initial analysis and, if interested, more data and more analysis should be done to provide more insightful conclusion.