

Recommendation of places of open a new restaurant in Canberra

Wenjing Xue

13 Mar 2020

1. Introduction

1.1 Background

Canberra is the capital of Australian, but it is not a tourist city. Canberra has the highest average wage in Australia, but it has very few good restaurants here. Thus, there is a great chance to make profit when opening a restaurant here.

1.2 Problem

There is a large area in Canberra is mountains, government office and foreign consulates, which are not suitable places for restaurant. In this report, I will try to find the population area and with less competition of many other restaurants.

1.3 Interest

People who wish to open restaurants in Australia can be highly interested in this report. Since the restaurants in modern cities in Australia such as Sydney, Melbourne are very competitive and with lower food price. There is a bigger chance to get success in Canberra.

2. Data acquisition and cleaning

2.1 Data source

The postcode and suburb data of Canberra are found here (<http://www.citypostcodes.com.au/Canberra>). The coordinate data of Canberra suburb are found here (<https://www.geonames.org/postal-codes/AU/ACT/australian-capital-territory.html>). We are going to use the suburb data as unique key to identify each area and use the coordinate data to identify the exact place and combine them with the venue source from the Foursquare API. And cluster the area, using the venue data to find which are have more restaurants, café, stores, etc. And this are the place which have more population which are more suitable for opening a new restaurant.

2.2 Data cleaning

In the postcode and suburb dataset, all the name of the suburb has the word *Postcode*, which we need to clean it before we can do any further

task. In the coordinate dataset, which has some duplicate suburb with different postcode, which we only want to keep only one of them, so we can merge them with the postcode dataset. Also, in this coordinate dataset we need to change the column name for the further merge task.

2.3 Feature selection

Since we need the unique *Suburb* name identify different areas, we are going to keep them. And the *Postcode* can help us define which area are the neighborhood, we also want to keep them. The *Latitude* and *Longitude* are the features for us to using the API get the venue data, which means that we need them. Although the *City*, *Country*, *State* are the same, since we want to show the exact place of Canberra for people, that we want to keep these features for a reminder.

3. Exploratory Data Analysis

4. Result

5. Conclusion

6. References

