Restaurant Recommender System- Sydney

IBM Capstone Project

Introduction

- Sydney is the state capital of NSW and the most populous city in Australia. It has very diverse population as people migrated from different part of world. Food is an integral part of any culture and with migration Sydney is enriched with variety of cuisines.

- As there is variety of cuisines available here, it becomes very important to find an authentic place to eat which is reasonable. Restaurant recommendation seems quite helpful here.

- Target audiences for this project does not limit to a person who loves to try food from their own place but everyone. People could simply decide to look for a similar restaurant all the time because they are addicted to a specific category of food.

- With restaurants evolving, new food categories emerge, hybrid food starts to become more popular, we need a system that could help us access vast number of food varieties.

Data Requirement

- To find a solution to the questions and build a recommender model, we need data and lots of data. Data can answer question which are unimaginable and non answerable by humans because humans do not have the tendency to analyze such large dataset and produce analytics to find a solutions.
  Let's consider the base scenario :

Suppose i want to find a restaurant, then logically, i need 3 things :

1)Its geographical coordinates(latitude and longitude) to find our where exactly it is located.

2)Population of the neighborhood where the restaurant is located.

3)Average income of neighborhood to know how much is the restaurant worth.

Data Collection

Collecting geographical coordinates is not difficult but after googling for few days, it was not available on open source data websites such as wikipedia, census report websites etc. Hence I created my own dataset, where I selected all the suburbs in Sydney city and googled their latitude and longitude. Here is the link of the website from where I got the suburbs data : https://www.cityofsydney.nsw.gov.au/learn/research-and-statistics/the-city-at-a-glance This same link provided the information about population and income. This is how the data frame looks like:

	Borough	Neighborhoods	Latitude	Longitude	Population	City	AverageIncome
0	South	Alexandria	-33.9080	151.1903	8262	Sydney	2421
1	West	Annandale	-33.8814	151.1707	9451	Sydney	2337
2	North	Barangaroo	-33.8638	151.2022	189	Sydney	3661
3	West	Beaconsfield	-33.9110	151.1999	987	Sydney	2330
4	West	Camperdown	-33.8862	151.1791	10341	Sydney	1796

# Data Collection
Four Square API

- Use of foursquare is focused to fetch nearest venue locations so that we can use them to form a cluster. Foursquare api leverages the power of finding nearest venues in a radius(in my case : 500 mts) and also corresponding coordinates, venue location and names. After calling, the following data frame is created:

|	Neighborhood	Borough	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Alexandria	South	-33.908	151.1903	Pino's Vino e Cucina	-33.905480	151.190950	Italian Restaurant
1	Alexandria	South	-33.908	151.1903	La Cachette	-33.904849	151.189727	Café
2	Alexandria	South	-33.908	151.1903	Blackbird & Co	-33.906612	151.187861	Café
3	Alexandria	South	-33.908	151.1903	The Copper Mill	-33.906068	151.188426	Café
4	Alexandria	South	-33.908	151.1903	The Grounds of Alexandria	-33.910774	151.194406	Café

Methodology
Exploratory Analysis

Scrapping the data from different sources and then combining it to form a single-ton dataset is a difficult task. To do so, we need to explore the current state of dataset and then list up all the features needed to be fetched.
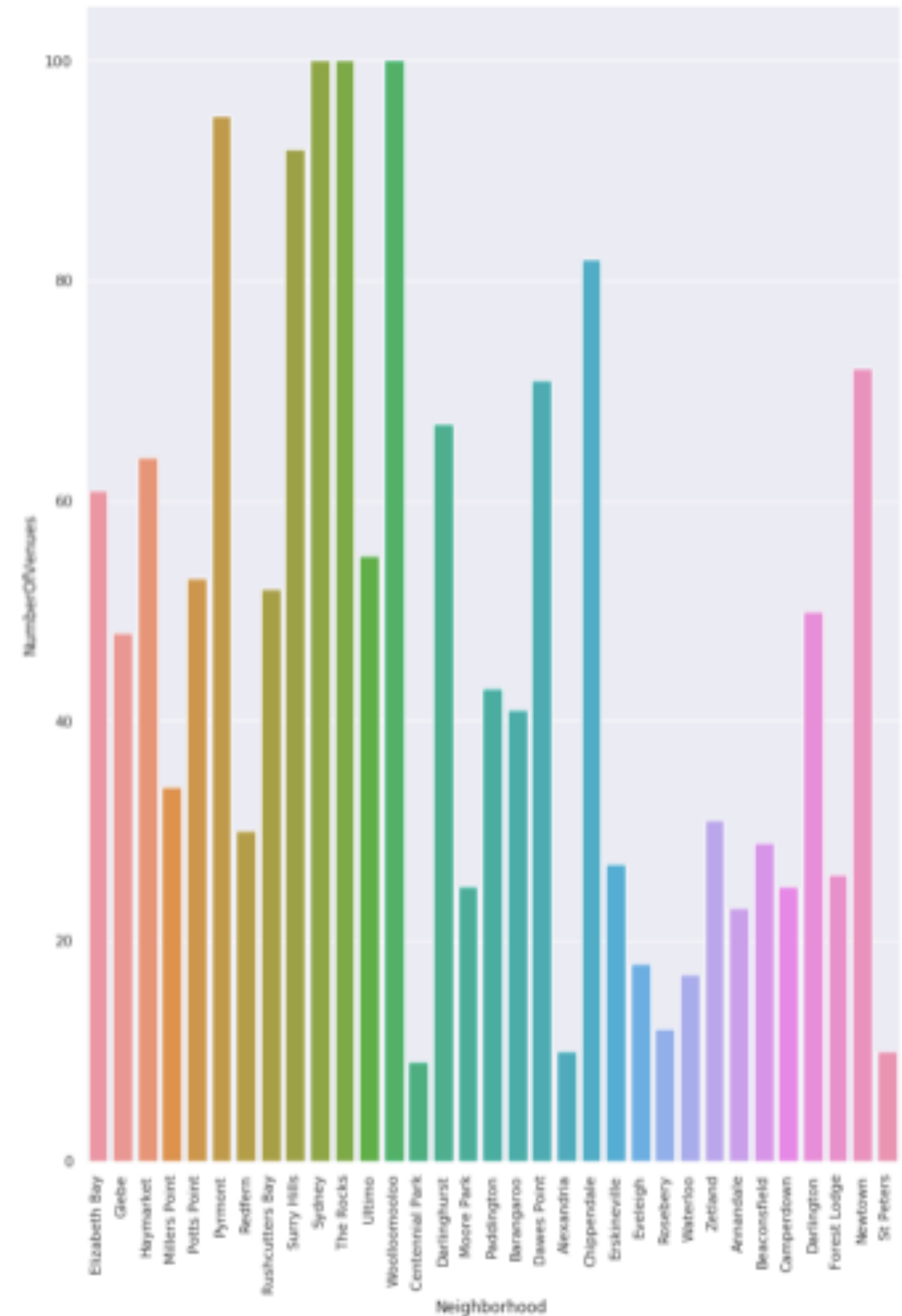
Methodology
Inferential

Most important factors while building the recommender system were population and income. They are the most import factor because they have a nonlinear relationship according to our dataset.
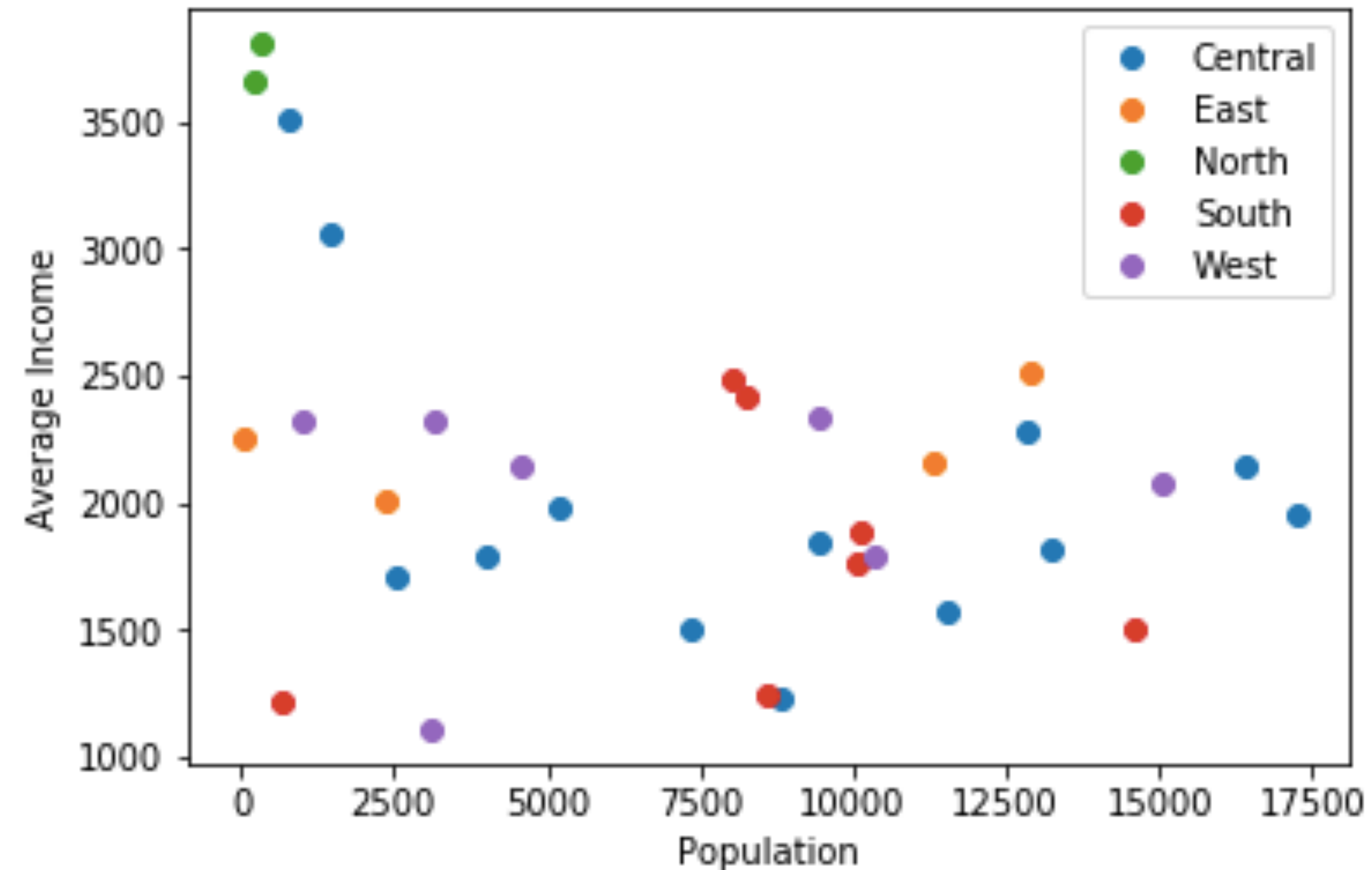
Mapping & Clustering

Since there was a nonlinear relationship between income and population, it can be concluded that we must always perform inferential approach to find relationship among different set of features. Also during clustering, similar neighborhoods must be dumped into the right cluster.
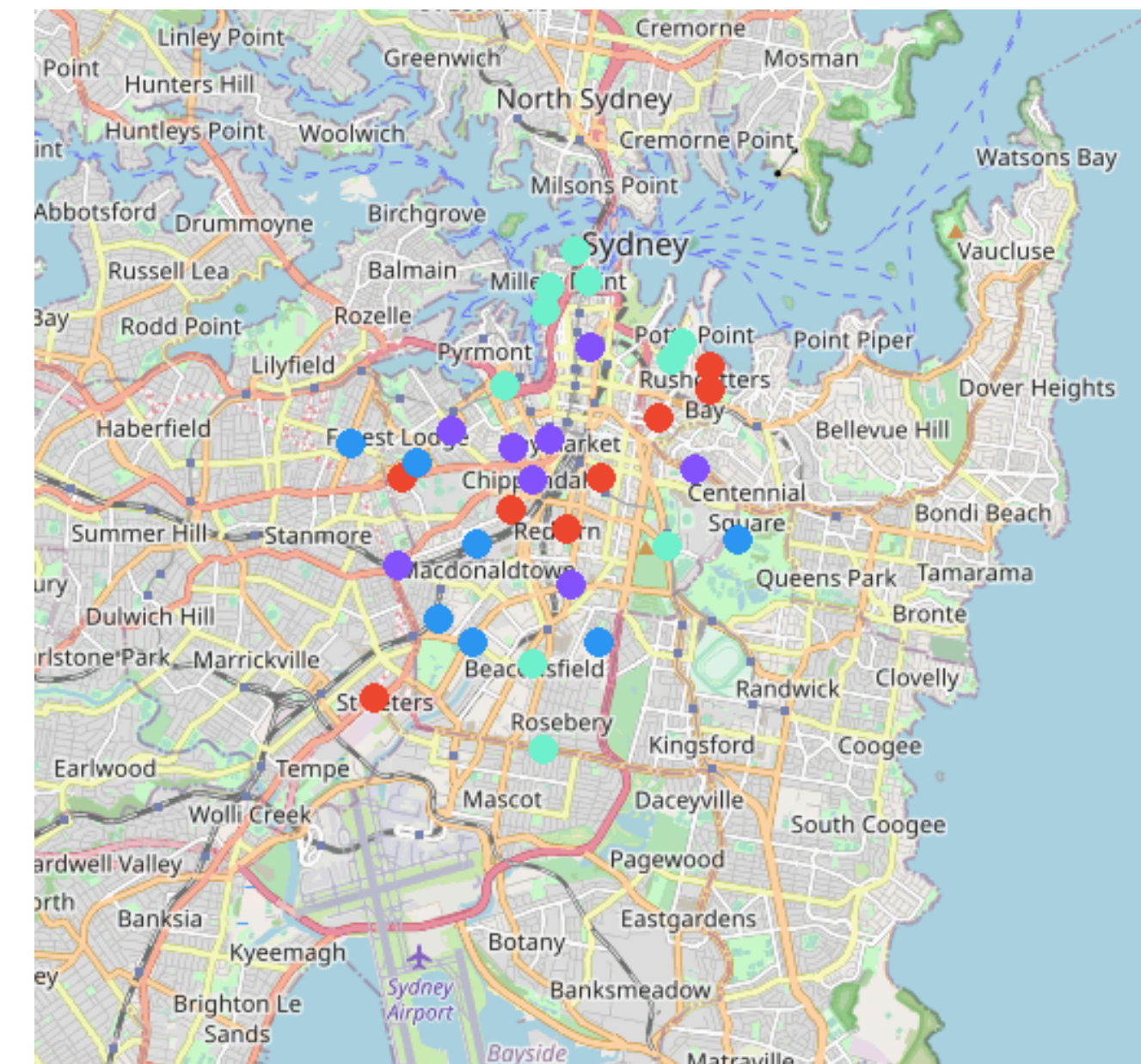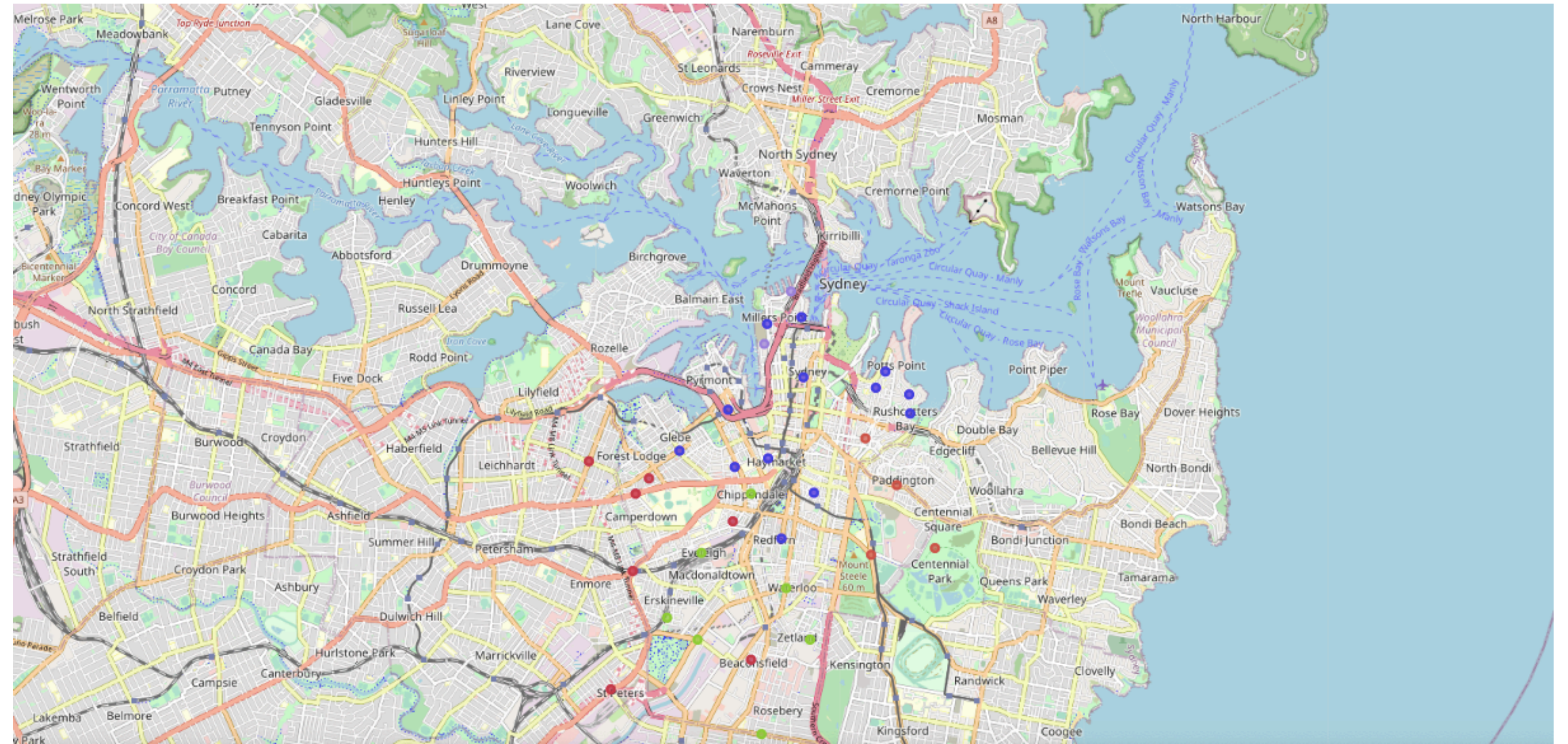
Results

The result of the recommender system is that it produces a list of top restaurants and the most common venue item that the user can enjoy.

Out[103]:

	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Ranking
0	Barangaroo	Venue Category_Café	Venue Category_Hotel	Venue Category_Bar	[0.4421434166390603]
1	Beaconsfield	Venue Category_Café	Venue Category_Coffee Shop	Venue Category_Furniture / Home Store	[0.3428723260123892]
2	Dawes Point	Venue Category_Café	Venue Category_Australian Restaurant	Venue Category_Pub	[0.46034662647808944]

```
--------- Alexandria ---------
                                 Venue  Frequency
0                  Venue Category_Café        0.5
1     Venue Category_Basketball Stadium        0.1
2       Venue Category_Italian Restaurant      0.1
3               Venue Category_Brewery        0.1
4    Venue Category_Australian Restaurant      0.1


--------- Annandale ---------
                                 Venue  Frequency
0                  Venue Category_Café       0.26
1                  Venue Category_Park       0.09
2         Venue Category_Grocery Store       0.09
3           Venue Category_Pizza Place       0.04
4         Venue Category_Garden Center       0.04


--------- Barangaroo ---------
                                 Venue  Frequency
0                  Venue Category_Café       0.10
1                 Venue Category_Hotel       0.07
2           Venue Category_Coffee Shop       0.05
3          Venue Category_Burger Joint       0.05
4                   Venue Category_Bar       0.05
```

Conclusion

The recommender system is a system that considers factors such as population, income and makes use of Foursquare API to determine nearby venues. It is a powerful data driven model whose efficiency may decrease with more data but accuracy will increase. It will help users to finish their hunger by providing the best recommendation to fulfill all their needs.