Applied Data Science Capstone Project

Exploring areas and venues in Sydney NSW, Australia

Harry Ngo

1. Introduction

New South Wales (NSW) is one of the major states in Australia, being most populous with over 8.1 million residents. There are many cities and towns within NSW, where its capital city Sydney is home to almost two-thirds of the NSW population.

The aim of the project is to explore different areas of Sydney and find optimal locations for a new resident based on the top common venues in each borough. This report will be targeted to people interested in becoming a resident in Sydney NSW, Australia, as well as homebuyers, people staying over a long period and other similar conditions. Others who are interested in learning about different boroughs of Sydney may also benefit from this.

With Australia being a multicultural country, its landscape is constantly evolving, seeing increases in different ethnical communities and venues. Each Sydney area will be discussed such that stakeholders can choose an optimal living location based on their interests.

2. Data

The following data will be used:

1. *Second-level Administrative Divisions, Australia, 2015*[1]
    - Contains second-level administrative and political divisions of Australia. This includes aboriginal councils, boroughs, cities, government and district councils, municipalities, rural cities, shires and territories. The relevant information from the exported JSON file includes the state/territory, borough/council and the multiple coordinates of each division.
2. *Foursquare API*
    - Used to obtain the common venues of each borough within Sydney such as restaurants, stores and parks.

Exploring the JSON file of Australian administrative divisions, the geometry of each location was of type *MultiPoint*. To maintain consistency and handle the data simpler, the first coordinate of each location was taken. This dataset consisted of 11 states or territories and 1395 boroughs, which was subsequently extracted to obtain only data pertaining to NSW; this contained 199 boroughs.

Exploring the data for NSW, there was no quick method to find boroughs in NSW in the Sydney region. Instead, I manually looked through the data frame to drop rows which were not in the vicinity of Sydney, which was chosen to be within an 80km radius of Sydney.

[1] *Second-level Administrative Divisions, Australia, 2015*, NYU Spatial Data Repository <https://geo.nyu.edu/catalog/stanford-fc944xn1421>

For example, the first five observations of the NSW data frame look like the following:

	State/Territory	Borough	Latitude	Longitude
0	New South Wales	Albury	-35.914963	146.971710
1	New South Wales	Armidale Dumaresq Bal	-30.589890	151.632751
2	New South Wales	Armidale Dumaresq	-30.483726	151.661728
3	New South Wales	Ashfield	-33.870983	151.119431
4	New South Wales	Auburn	-33.850048	151.081467

The boroughs of Albury, Armidale Dumaresq Bal and Armidale Dumaresq do not belong within an 80km radius of Sydney (where the centre is in the city of Sydney), so they were removed from the data frame.

3. Methodology

4. Analysis

5. Results

6. Discussion

7. Conclusion