IBM Applied Data Science Capstone

recommending a suburb to start an indian resturant

Aneesh Raj

2021

# Introduction

Hospitality industry in Sydney is one of the highly competitive industries. Often it is daunting to make a decision on the location to start a restaurant without considering all the facts. Most of the time these decisions are made subjectively. With the advancement in data science technology, these type of business decisions can be made objectively through data thereby improving the chance of success considerably for the business.

## Business Problem

The objective of this project is to analyse the two local government areas (LGA) in Sydney (Parramatta and Sutherland Shire) and recommend the best suburb in one of these LGAs to start an Indian restaurant. The target audience for this project is entrepreneurs who is thinking about opening an Indian restaurant in any of the two LGAs in Sydney.

# Data

In order to find a solution for the business problem, the following datasets are required.

* List of suburbs in the two LGAs
* Latitude and Longitude of all the suburbs
* Population density of the two LGAs to gauge potential customer traffic
* Median household income of the two LGAs
* Number of business in accommodation and food services in the two LGAs
  + Business entry vs business exit
* Number of people employed in accommodation and food services in the two LGAs

We can use Australian bureau of statistics website to retrieve the population density, median household income, number of business in accommodation and food services and its entry vs exit overtime, number of people employed in the accommodation and food services industry to gauge availability of skilled workforce. Based on this information, we can select the LGA that is best suited to open the Indian restaurant. This can be done via the read\_html function in pandas.

The Wiki page 'https://en.wikipedia.org/wiki/Template:Sydney\_Sutherland\_suburbs' and 'https://en.wikipedia.org/wiki/Template:Sydney\_Parramatta\_suburbs' provides the list of suburbs in the two LGAs. We can then use read\_html function in pandas to extract the suburb tables in the wiki and convert it to pandas dataframe. Then we can use Python’s geopy package to obtain the latitude and longitude of all the suburbs present in the dataframe. With this we should be able to retrieve the available venue details of the suburbs in the selected LGA using the Foursquare API. This can be used to select the suburbs suited to open an Indian restaurant. The example below shows the 10 common venues in suburbs of Parramatta LGA.

