

Data Science Capstone Assignment

1. Description of the problem and a discussion of the background

My client Ernesto owns an Italian restaurant Fabia located in Melbourne (Australia) and is interested in opening another restaurant in a different suburb of Melbourne.

Fabia restaurant has become increasingly popular over the recent years and is well known for its authentic Italian dishes at reasonable prices and great service. The restaurant is located in a southeast suburb of Melbourne called Glen Iris.

Ernesto (owner) is keen to open another Fabia restaurant in the Northern suburbs of Melbourne but is unsure which suburb would be the most appropriate or be the most similar to Glen Iris.

Ernesto believes that suburbs with similar characteristics (population density, socio economic, demographic attributes as well as similar types of venues) to Glen Iris should be considered first.

The problem we are trying to solve is:

identify suburbs (North Melbourne) with similar population and location characteristics to Glen Iris, understand what makes these suburbs similar to Glen Iris and finally make a recommendation to Ernesto which suburbs he should consider first and why with regards to opening his second restaurant.

2. A description of the data and how it will be used to solve the problem

Data and data sources:

- Postal Codes/Suburbs location data (latitude and longitude) – Australian Post
- Socio Economic & Demographic data - Australian Bureau of Statistics Census 2016 data
- Location and Venue Data - Foursquare

Analytic approach that we will be using is k-means clustering to find similarities and dissimilarities between suburbs in Melbourne Australia.

The output of clustering will then be analysed, the insights and recommendations will be documented in the final report that will be discussed and provided to my client Ernesto.