Predicting Cuisines having Maximum Potential in a Neighborhood Restaurant

1. Introduction

There are thousands or restaurants in Main land USA and if one needs to set new restaurant what is the best place to do so by use of available information. By identifying best location one can maximize value and returns. Therefore it is imperative to find the best place to identify best place by use of Data science methodology.

2. Data acquisition and cleaning

For this project data gathered from foursquare about type of cuisine most served in neighborhood restaurant and US census portal about neighborhood profile through use of API's.

US Census has API's which provide information about geographies for more than 18000 variables. I plan to use variables pertaining to age, race, income etc. along with data obtained from foursquare regarding number of place, type of cuisine and then predict the type of cuisine can be served when setting a restaurant in a new neighborhood.

Based on definition of our problem, factors that will influence our decision are:

- number of existing restaurants in the neighborhood (any type of restaurant)
- Demographic profile of the neighborhood

Based on center of location I plan to get number of existing restaurants in the neighborhood in a radius of 25/50km (default foursquare location radius) and then take them as target variable based on Demographic profile.

Following data sources will be needed to extract/generate the required information:

- Number of restaurants for each cuisine type in neighborhood i.e. approximate radius of will be obtained using **Foursquare API.**
- Demographic profile of the neighborhood will be obtained using **US Census API**.
- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using **Google Maps API reverse geocoding**

1. Methodology

To predict best type of cuisine to serve in a neighborhood I followed following series of Step.

Finding Demographic profile of all US regions defined by the US Census. I would be using Income, Age and work related variables. Description of these variables is defined in US census (list of variables is given in Annexure). I would be considering variables pertaining to Male/Female age, Education profile and Income distribution profile. Once I have location of interest for mainland US I would be using Google reverse coding to find the latitude and longitude. For these locations I would find the number of restaurants for their cuisine type served in that location. Finally once data is processed and enriched I would use classification techniques to find cuisine to serve in the neighborhood location. Once I have predicted cuisine type corrected I would be using regression to predict count of restaurants each region can support. With count of existing restaurants we got from foursquare. Find difference of predicted count. The regions where the difference is maximum I can conclude that is the best location for setting the Restaurant with maximum potential.

4. Predictive Modeling

There are two types of models, regression and classification, that can be used to Best Cuisine to serve and potential of a place. Classification model can provide information about best cuisine to serve and Regression model can provide information what number of additional restaurants location of interest can sustain. Therefore, in this study, I carried out both regression and classification modeling.

4 Classification models

4.1 Applying standard algorithms and their problems

I applied standard classification models Decision Tree, SVN and RandomForsest models to find what the best cuisine to serve is. Then I used standard regression Models to find what the amount of improvement will be? I used grid search and then model recall to find best model.

4.1.3 Performances of different models

Using and applying different models I found I am able to predict 'East Asian Cuisine' as reasonably well and observed following f1 score

Decision Tree	Random Forest	SVM
0.60	.56	.59

5. Conclusions

In this study, I analyzed the relationship between demographic profile of a location and how it affects the cuisine served in a neighborhood restaurant. I used demographic variables pertaining to income, age for male and female.

6. Future directions

In this project I am constrained by limited amount of information freely available but would like to explore further about each location and then quantify the incremental revenue one can expect by setting up a restaurant in a neighborhood