Predict Grocery Demand for Mobile Grocery Store based on Cuisine Type for Franklin County Neighborhoods in OH

Jagadeesh Sreeram

Dec 27th 2019

1. Introduction

1.1 Background

Raj my friend wants to start a Asian mobile grocery store that servers most of the Neighborhoods in Franklin County. His main target is restaurants in these neighborhoods. He procures and sells mostly 5 type of Asian grocery I.e Indian, Chinese, Thai, Japanese and Korean. He will have just one warehouse in Franklin county from where in the morning all his mobile store trucks for most of these neighborhoods will leave. So, he wants to know if a mobile grocery truck is leaving for a particular neighborhood what percent of each of above 5 type Asian grocery need to be carried in the mobile grocer store. His main target is Asian restaurants of above 5 types in these neighborhoods. He also doesn't want the percent demand for each neighborhood at this point , but just want to keep to 5 groups.

1.2 Problem

Predicting Grocery demand based on restaurant concentration in ~25 mile radius of each neighborhood in Franklin county OH. Also minimize the category of grocery percentage groups to 5.

1.3 Interest

Raj is very interested to know what type of Asian restaurants exists in his county and how much percentage quantity he need to carry in his mobile store trucks. Even though currently Raj is targeting restaurants in the neighborhood, later this also can be extended by taking different kind of population type in those areas, so that Raj can target residential areas too.

2. Data acquisition and cleaning

2.1 Data Source

The Franklin County neighborhood data is obtained from US Postal service website https://www.unitedstateszipcodes.org/zip-code-database/ . There are 67 neighborhoods in Franklin County. For restaurant data we used Foursquare venue search API with category filters to search restaurants in these neighborhood that are Indian, Chinese, Thai, Japanense and Korean. For each neighborhood we get above category restaurants within 25 mile radius
Foursquare API Ref :
https://developer.foursquare.com/docs/api/endpoints

2.2 Data Cleaning

The neighborhood information was readily available from US Postal website in csv format with all the required information for this project. We had to simple filter for Franklin county and upload the data in the project area. Below is snippet of Franklin county neighborhood data

Snapshot of just top 5:

	Borough	Neighborhood	Latitude	Longitude
0	Franklin County OH	Amlin	40.07	-83.18
1	Franklin County OH	Blacklick	40.02	-82.80
2	Franklin County OH	Dublin	40.10	-83.15
3	Franklin County OH	Dublin	40.11	-83.13
4	Franklin County OH	Hilliard	40.03	-83.14

There are actually 67 neighborhoods in Franklin county and we use all for this project.

The category filter for the 5 type of restaurant types are obtained form Foursquare category doc and used in the project:

https://developer.foursquare.com/docs/resources/categories

2.3 Feature selection

Even though we provided the category for search API, foursquare gives restaurants for other types, since this data was very minimal and doesn't impact overall model, we removed that data from the model as part of data cleanup. The data that was removed is for category types "Seafood Restaurant,Sushi Restaurant,Vietnamese Restaurant & Deli / Bodega" . since the dataset was very low for these and these causes issue for our model we removed this data. Since Foursquare only gives ~100 venues for each neighborhood, we had ~ 6700 restaurant information overall for the model.

Below are the highlighted categories that were removed.

	Neighborhood	Asian Restaurant	Chinese Restaurant	Deli / Bodega	Indian Restaurant	Japanese Restaurant	Korean Restaurant	Seafood Restaurant	Sushi Restaurant	Thai Restaurant	Vietnamese Restaurant
0	Amlin	0	1	0	0	0	0	0	0	0	0
1	Amlin	0	0	0	1	0	0	0	0	0	0
2	Amlin	0	0	0	0	1	0	0	0	0	0
3	Amlin	0	0	0	0	1	0	0	0	0	0
4	Amlin	0	0	0	0	1	0	0	0	0	0

The using neighborhood location information we retrieve the restaurant information from Foursquare for ~25 mile radius of each nighborhood and then compute what type of restaurants are more predominant in each neighborhood and rank them . And later using k-means approach , we categorize them into 5 categories and compute the average grocery demand for each of these 5 grocery type .