

Battle of Neighborhoods

COURSERA CAPSTONE PROJECT

ARUN NATARAJAN

Table of Contents

1. Background	3
2. Business Problem	3
Importance of improving missing data of neighborhoods/venues	3
Stakeholders/Audience who would benefit from the solution to this problem	3
3. Data and Approach	4
Data Requirements and Data Sources	4
Analytic Approach	4
References	5

1. Background

Battle of Neighborhoods, this project is aimed at solving a problem involving Foursquare location data and demonstrating data science skills.

I choose to explore the city I am born and brought up; Coimbatore located in southern state of Tamil Nadu in India. Coimbatore is second largest city in the state and very well connected by road, rail, and air transport. It has mixed cultures from all over India. Well known for engineering and manufacturing, education institutions, health care facilities and services, textile businesses as well as it is a tourism hub.

Coimbatore has 64 neighborhoods (*Neighbourhoods of Coimbatore, n.d.*) excluding the sub-urban neighborhoods.

Initial thought was to cluster neighborhoods, which could help small business start-ups in terms of assessing opportunities and existing competition. However, when I explored neighborhoods using Foursquare I figured out a problem of missing data, which might be current limitations/challenges to location-based companies like Foursquare.

2. Business Problem

My Initial exploration of Coimbatore's places using Foursquare's explore endpoint revealed that 177 venues from 51 neighborhoods out of 64 total neighborhoods in the city, while search end point resulted in 7472 venues covering all 64 neighborhoods.

Note that there are **13 neighborhoods itself did not appear in recommendations and 376 venues appeared as uncategorized**. Second tier cities like Coimbatore may lack availability of data (categories, user reviews, etc...) for recommendations to new/prospective users (especially tourists/travellers and Foursquare's own customers).

Out of all attributes of a venue, category is plays crucial role in fetching results. In rest of project I would attempt find a solution to assign categories to uncategorized venues.

Importance of improving missing data of neighborhoods/venues

This could affect variety of people and their needs,

- As stranger/traveller one could be presented with no recommendations near him or show something far, while some venue is really close by which is unreviewed but of same category.
- As Foursquare's customer / business planer, venue details with missing categories, reviews, recommendation ratings would make it difficult to analyse and make decisions related to choose of neighborhoods/venues

Stakeholders/Audience who would benefit from the solution to this problem

- Foursquare / Similar location data providers
 - Enrich their databases for better competitive advantage and improved customer experience
- App developers/owners, who offer services based on location-data (example Food Delivery, Cab Services etc...)
 - When an un-reviewed venue is found closer than a recommended venue, enable users with optional features like "Be First to Review" and reward for doing so

- Such an approach could be taken if venues are categorized otherwise it would even be exceedingly difficult.

3.Data and Approach

Data Requirements and Data Sources

- A list of neighborhood names (*Neighbourhoods of Coimbatore, n.d.*) and Neighborhood location as latitude and longitude (GeoPy , n.d.)

Region	Neighborhood	Latitude	Longitude
North	Kavundampalayam	11.0452351	76.9472197
North	Chinnavedampatti	11.0629428	76.9843304
North	Press Colony	11.178903	76.9569137
North	Vadamadurai	11.084491	76.9390361
North	Kanuvai	11.0793411	76.9415814
North	KNG Pudur	11.0533111	76.9199558

- List of venues from neighborhoods and their categories (*Foursquare Endpoints, n.d.*)

Neighborhood	Venue	Venue Category
Kavundampalayam	Kavundampalayam	Parking
Kavundampalayam	Sabari Bakery	Bakery
Kavundampalayam	Ayyappas Restaurant	
Kavundampalayam	Ayyappa's Pearl Wedding Hall	Event Space

- List of all possible categories (*Foursquare Categories, n.d.*)

Top Category	Sub Category
Arts & Entertainment	Amphitheater
Arts & Entertainment	Aquarium
Arts & Entertainment	Arcade

Analytic Approach

In short problem is to categorize a venue, which could be approached as **Supervised Multi-Class Text Classification problem**.

Supervised approach requires a **labelled dataset** to train and evaluate models, which contain minimum two columns (**text as input and category as output**)

- Input Text: Venue Names
- Output: Anyone of Foursquare Defined Categories

Neighborhood	Venue	Venue Category
Kavundampalayam	Kavundampalayam	Parking
Kavundampalayam	Sabari Bakery	Bakery
Kavundampalayam	Ayyappas Restaurant	
Kavundampalayam	Ayyappa's Pearl Wedding Hall	Event Space

Such final dataset for model training and testing could be built by analysing neighborhood, location, venue and venue category data.

References

Foursquare Endpoints. (n.d.). Retrieved from Foursquare Places API:

<https://developer.foursquare.com/docs/places-api/endpoints/>

Foursquare Categories. (n.d.). Retrieved from Foursquare Places API:

<https://developer.foursquare.com/docs/api-reference/venues/categories/>

GeoPy. (n.d.). Retrieved from GeoPy Documentation: <https://geopy.readthedocs.io/en/stable/>

Neighbourhoods of Coimbatore. (n.d.). Retrieved from Wikipedia:

https://en.wikipedia.org/wiki/Neighbourhoods_of_Coimbatore