Introduction and Business Problem

Introduction

The city of Lucknow, Uttar Pradesh,India is relatively small at ~1 square mile but it is packed with restaurants, night life and amazing people. For people that are new to Lucknow, despite its small geographic size, it can be daunting to figure out what restaurants are worth going to and where they are. For people that used to live in Lucknow or are visiting Lucknow, how do you know what the best places are to get something to eat?

Business Problem

For this project, I am going to put on my entrepreneur hat and create a simple guide on where to eat based on Foursquare likes, restaurant category and geographic location data for restaurants in Lucknow. I will then cluster these restaurants based on their similarities so that a user can easily determine what type of restaurants are best to eat at based on Foursquare user feedback.

Data Requirements

For this project, I will be utilizing the Foursquare API to pull the following location data on restaurants in Lucknow:

- Venue Name
- Venue ID
- Venue Location
- Venue Category
- Count of Likes

Data Acquisition Approach

To acquire the data mentioned above, I will need to do the following:

- Get geolocator lat and long coordinates for Lucknow,India
- Use Foursquare API to get a list of all venues in Lucknow
    - Get venue name, venue ID, location, category, and likes

Methodology

The thought process behind this is that likes are a proxy for quality. The more likes there are, the better the restaurant is. This might be incorrect but API call issues (how many I can use for free) holds me back from getting price / rating data. I will then bin this data into a quality categorical variables so we can cluster appropriately.

I am also going to create new categorical variables for the restaurants to better group them based on type of cuisine. This way you can look for good Indian food or now what type of food might be best to eat in Lucknow if you are new to the area.

I will take the gathered data (see above in Data Acquisition Approach and Data Required sections) and will create a k-means clustering algorithm that groups restaurants into clusters so that people looking to eat in Lucknow can easily see which restaurants are the best to eat at, what cuisine is available and where in Lucknow they can look to eat.

Challenges -

- Due to very less number of coordinates found using Four Squares API

  Proper Cluster are not formed .This is may be due to API's does not list all Resturants and cuisines available in Lucknow or due to lack of Digitilization of City.

Results

Running my clustering algorithm, I was able to generate Map ofclusters for restaurants.These are as follows:

	name	id	categories	lat	lng	total likes	total likes_cat	categories_new	label
0	Royal China	4caa0096d971b1f7ccca23e1	Chinese Restaurant	18.938715	72.832933	35	below avg	None	0
1	Town House Cafe	5263e1ba11d265711e8024bf	Bar	18.938550	72.833464	16	poor	bars	1
2	Chhatrapati Shivaji Maharaj Terminus	4babe4fff964a520f8d23ae3	Train Station	18.940297	72.835384	299	great	None	2
3	Cafe Excelsior	4c714f4fd7fab1f715d760c9	Café	18.937701	72.833566	31	below avg	None	0
4	Sher-E-Punjab	4b0587d9f964a52023a422e3	Indian Restaurant	18.937944	72.837853	16	poor	euro asia indian food	1

Map of Clusters for Users