# Comparing Neighborhoods Project

Ashraf Ul Kabir

## Introduction

Aldente is an Italian restaurant chain looking to invest in Pakistan and for this purpose they want to evaluate the best location to open their first restaurant.

There are many considerations they want to study, and one of them is definitely the social behavior and characteristics of the different neighborhoods.

They also want to understand the value and potential for opening their restaurant.

## Data

In order to fulfill this first a geographical map of the different neighborhoods in each city would be needed.

This can be easily done using Google Maps.

Once the neighborhoods are geo coded, they can be converted into csv format so they can be imported into a dataframe.

Once imported the query on Foursquare would determine the restaurant locations and the number of likes per location.

The number of likes is a direct correlation to the number of potential customers who would be visiting that location daily.

## Methodology

In order to study the potential of the neighborhood for the restaurant chain we used the following method:

We did a search query to find out all the restaurant venues in the vicinity of that neighborhood, usually a neighborhood comprised of an area of 500 meters radius. This resulted in a list of the venues.

Each venue was then further queried to find out the ratings, likes, tips etc. This can be added to the data frame as a column or dimension that will be measured.

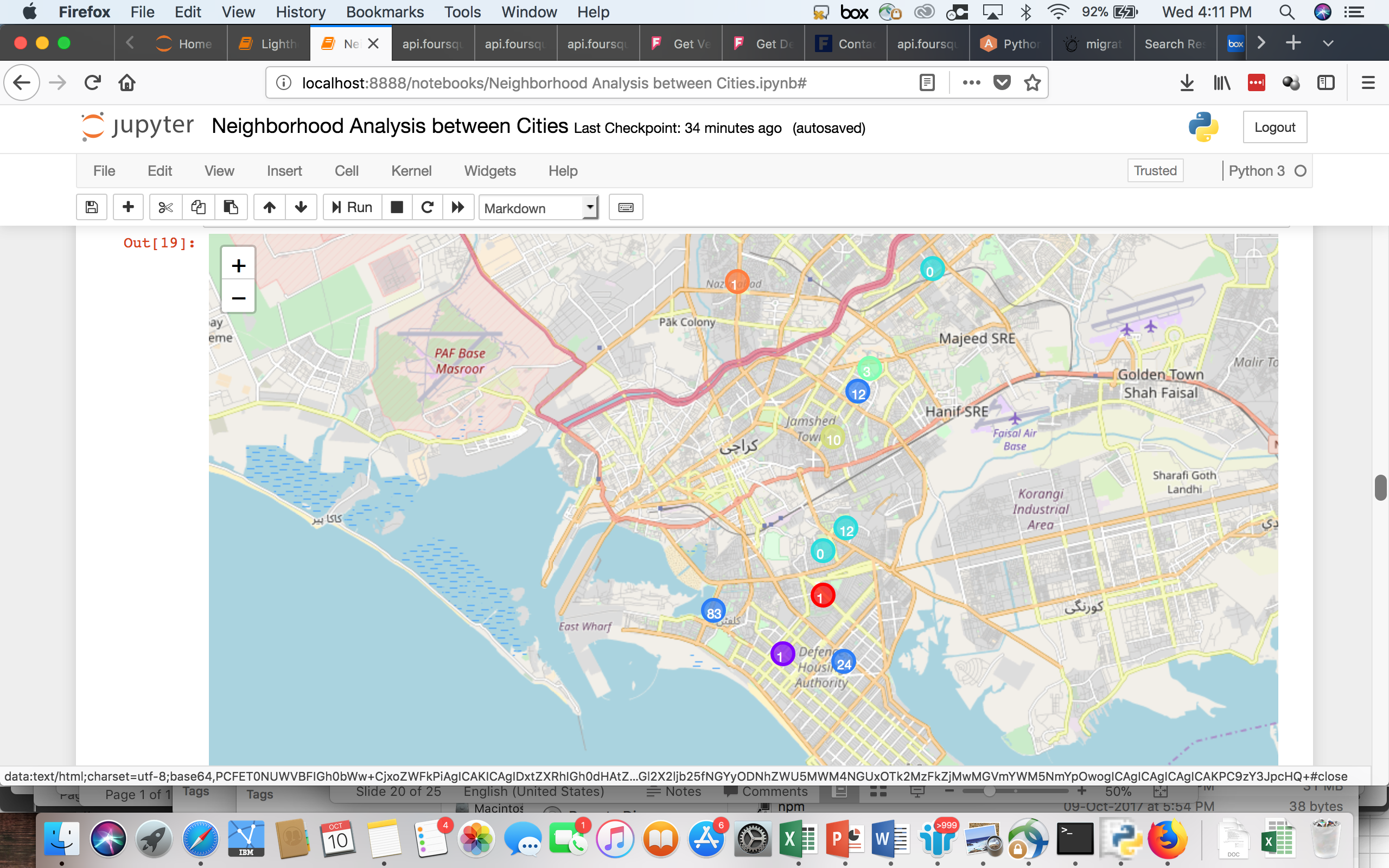
We can group these venues by neighborhoods and find a likability score as a mean of all the likes collected by restaurant venues.

Once this is established, we can further examine the clusters to find out the top most characteristics of that neighborhood.

We can use kmeans clustering to categorize the clusters and color code them so that individual characteristics can inform us better of the kind of clientele the restaurant will be getting.

Once we plot all of this information in the form of a map showing the cluster groupings as color coded cicles and the likability score as a number, it starts to become obvious which locations are more lucrative.

## Results for Karachi

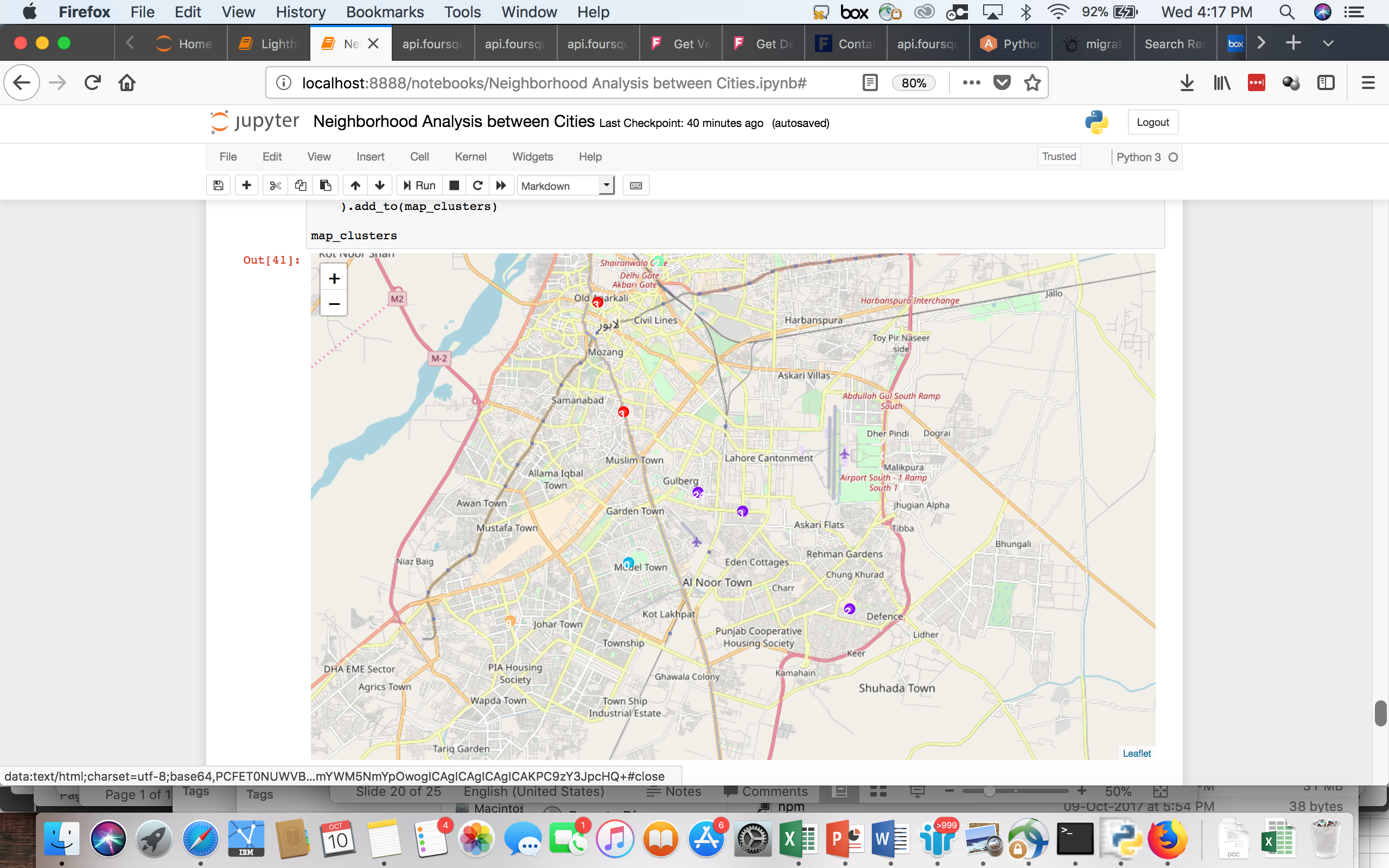


The colors give a clue to the characteristics of that neighborhood

The blue circles are more likable and more alike in nature

## Results for Lahore

We can notice that the bottom part is more alike than the upper part



The upper part denotes a more historical and old part of the city

## Results

When we ran the model to cluster the neighborhoods, the model factored in the various aspects of the top venues for each neighborhood.

This resulted in clustering the neighborhoods that were similar together.

Later we ran the query to find the top categories of each neighborhood, and we discovered that the more commercial shopping districts were clubbed together, similarly the historical locations were clustered closer, and the newer trendier locations were aligned together.

We labeled each cluster by understanding the unique characteristics of that location based on the foursquare information, as well as the geo context of that location.

## Discussion

Upon investigating, we can see that in Karachi the cluster 3 is more lucrative and stands out. There are 4 neighborhoods namely Bahadurabad, Clifton, DHA Phase 6 and North Nazimabad. Of these choices, Clifton stands out as being way ahead. This gives us a view that it is more visited by people who are savvy in technology adoption and express it using the mobile.

Similarly, when we analyze Lahore we can determine that of all the clusters, cluster 2 having DHA Phase 5, Cavalry Ground and Gulberg are most prominent in terms of likability. Of these Gulberg appears to be the trendy place for foodies, having a diverse set of restaurants, high number of likes and densely commercialized neighborhood.

The analysis of public domain information such as Foursquare can provide an insightful lense into the public behavior and allow a new entrant to determine for itself the best option for opening a restaurant.

## Conclusion

Since the number of checkins, likes determine the smartphone usage it can be directly correlated to the kind of customers.

These customers are likely to have an expensive smart phone, are literate, have a data package so they can tweet, comment or checkin and have spare money to spend on luxuries.

These are the target customers for Aldente and hence it is an important factor to determine the potential of the neighborhood for their business.

We can rate using the power of data science what would be the potential for each neighborhood for Aldente.

Hence we conclude that Clifton, having the most likability, and being closer to beach is one of the best locations for Aldente to open a restaurant in Karachi and would attract not only frequent visitors but also the ones who can spend more.

In Lahore, we noticed that the newer areas are more likable as a dining place by the segment of our choice, and hence of all the locations, Gulberg being central and accessible is more lucrative.