# Comparison of Food Choices of Cochin and Calicut cities

Ajaz Hameed

June 19, 2021

## Introduction

### Background

Cochin and Calicut are two of the main cities in the Indian State of Kerala. Geographically these places are located within 200Km, but we can see a big difference in culture, cuisine, language etc. between the people of these cities. These cities are also two of the biggest economic hubs of the state. In Calicut, service sector dominates the economy followed by industries. The city also has a strong mercantile aspect. Investments in IT is still in the beginning stage.[¶](https://render.githubusercontent.com/view/ipynb?color_mode=auto&commit=e6693e98eb01b4779ee7495f41486fc65fc5340e&enc_url=68747470733a2f2f7261772e67697468756275736572636f6e74656e742e636f6d2f616a617a48616d6565642f436f7572736572615f43617073746f6e652f653636393365393865623031623437373965653734393566343134383666633635666335333430652f636f7572736572615f63617073746f6e652e6970796e62&nwo=ajazHameed%2FCoursera_Capstone&path=coursera_capstone.ipynb&repository_id=352623976&repository_type=Repository#Cochin-and-Calicut-are-two-of-the-main-cities-in-the-Inidan-State-of-Kerala.-Geographically-these-places-are-located-within-a-distance-of-200Km-but-we-can-see-a-big-difference-in-culture,-cuisine,-language-etc-between-the-people-of-these-cities.-These-cities-are-also-two-of-the-biggest-economic-hubs-of-the-state.-In-Calicut,-service-sector-dominates-the-economy-followed-by-industries.-The-city-also-has-a-strong-mercantile-aspect.-Investments-in-IT-is-still-in-the-beginning-stage.)

Kochi is often referred as the commercial and financial capital of Kerala. Availability of electricity, fresh water, long coastline, backwaters, good banking facilities, presence of a major port, container trans-shipment terminal, harbor terminal and an international air terminal are some of the factors which accelerated the industrial growth in the city and its adjoining district. In recent years the city has witnessed heavy investment, thus making it one of the fastest-growing second-tier metro cities in India. The district contributes 15.1% of the state GDP. Construction and manufacturing industries combined contributes 37% of the district's total GDP, and trade, tourism and hospitality industries together provide another 20%. Major business sectors include construction, manufacturing, shipbuilding, transportation/shipping, seafood and spices exports, chemical industries, information technology (IT), tourism, health services, and banking.

### Problem

This project is a comparative study of food choices of Kochi and Calicut to understand how much the economic and cultural differences can affect the food preferences of people.

### Interest

This project is useful for restaurant chains who want to start their outlets in different locations to customize their offerings based on the interest of people of the respective locations.

## Data Acquisition and Cleaning

### Data Sources

The first step in data acquisition is to get the details of places which come under the city limit of both Kochi and Calicut. Web scraping technique is used to collect the postal code details of each neighborhood. BeautifulSoup and Python Requests are used for this purpose. Then coordinate information of neighborhoods are obtained using geocoder package.

Once the coordinate information of neighborhoods is obtained, Foursquare API is used to get the restaurant details. Using Foursquare API, from each location, the top 100 venues of each neighborhood are identified. By using category ID, the search is limited to restaurants.

### Data Cleaning

The data about neighborhoods of both cities are obtained from web scraping and python geocoder package. The postal code details neighborhoods were collected and used as input for geocoder. With geocoder, the latitude and longitude details of neighborhoods are obtained. We couldn’t get the location details of all the neighborhoods. Those neighborhoods are removed from the table. With this, the input for Foursquare API is ready. The radius of search is limited to 3Km and the results are limited to 100. The category ID for restaurants is obtained from Foursquare Developer website. This category ID is used to filter out restaurants. Same procedure applied for both locations.

### Feature Selection

The restaurants are first grouped based on neighborhoods to understand the densest areas. After that the unique categories of returned venues were curated from the dataframe. There were 73 and 65 unique categories respectively at Kochi and Calicut. Later one-hot encoding is done on venue data grouped by neighborhood and the frequency of occurrence of each category of restaurants were identified. Then we obtained the list of top 10 category of restaurants from each neighborhood.

## Exploratory Data Analysis

The different categories of restaurants are plotted as a word cloud to identify the most common restaurant categories and different unique restaurants present at the location.

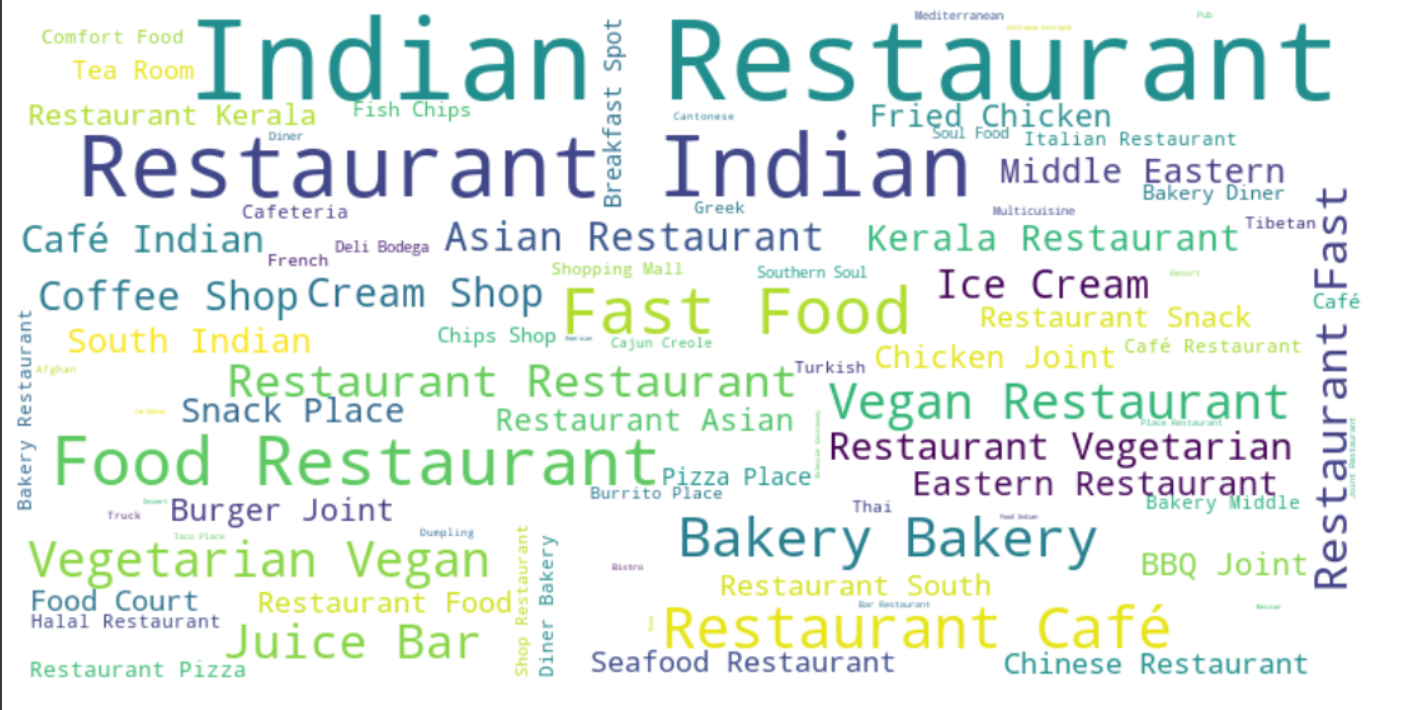


Figure 1: Word Cloud Kochi

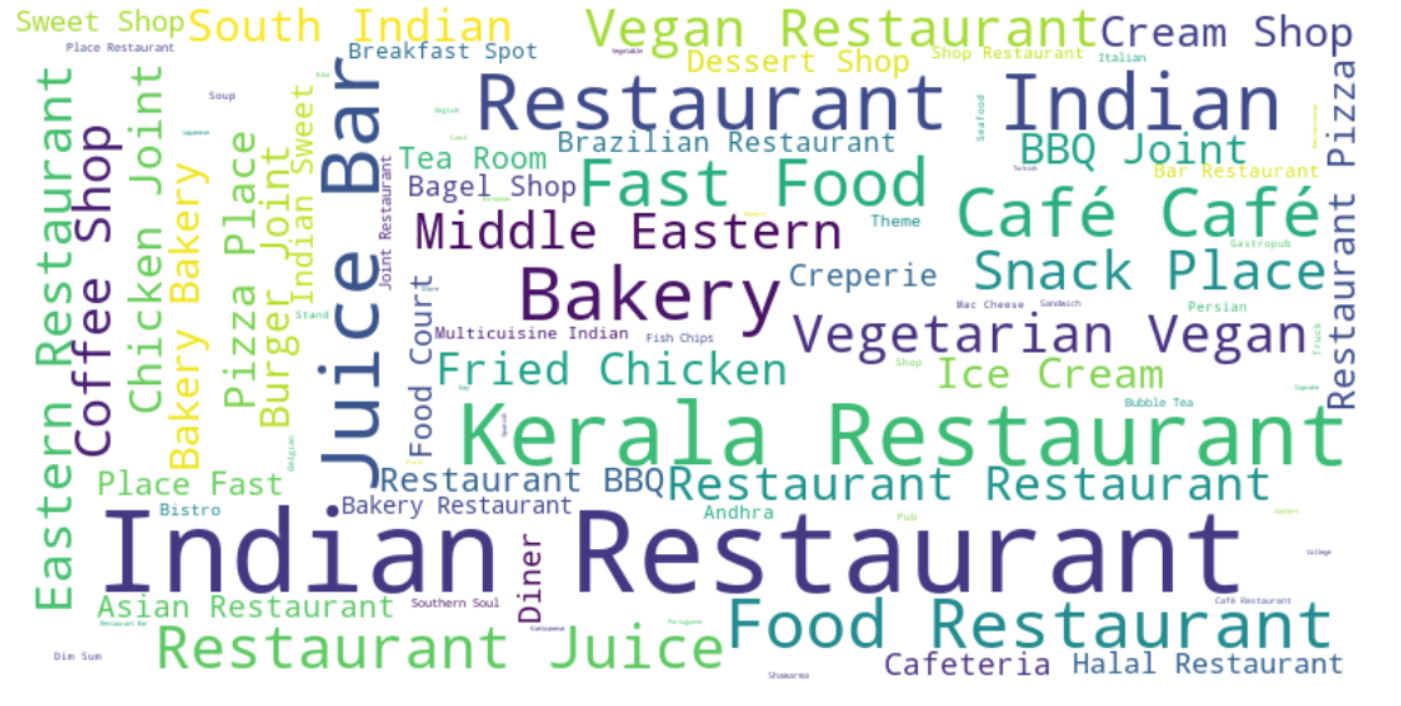


Figure 2: Word Cloud Calicut

From the word cloud of both cities, it is understood that the types of restaurants are mostly similar. Some differences are also observed. The percentage of Middle Eastern restaurant is more in Calicut as compared to Kochi. The reason can be because many Keralites are working in the Middle Eastern Countries. There were also trade relations from centuries. Unlike Calicut we could see a lot of Restaurants which serve continental dishes in Kochi.

## Clustering Neighborhoods

The restaurant details are grouped by Neighborhood. The frequency of occurrence of each category is determined for both the cities.

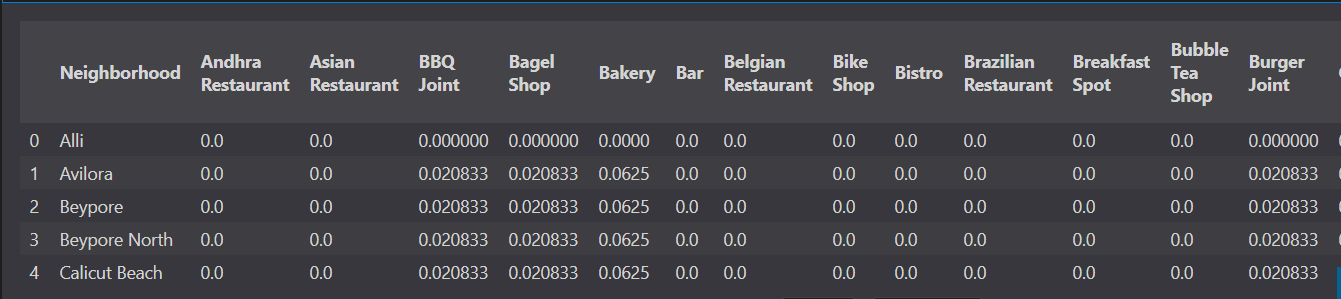


Figure 3: Calicut Restaurant details grouped



Figure 4: Kochi Restaurant details grouped

Then 10 most common venues of each location are sorted out.



Figure 5: Most common restaurant types in Kochi



Figure 6: Most common restaurant types in Calicut

The neighborhoods are classified into 5 clusters using K-Means clustering technique. From scikit learn documentation, the K-Means algorithm clusters data by trying to separate samples in n groups of equal variances, minimizing a criterion known as the inertia or within-cluster sum-of-squares. The neighborhood is divided into 5 clusters.

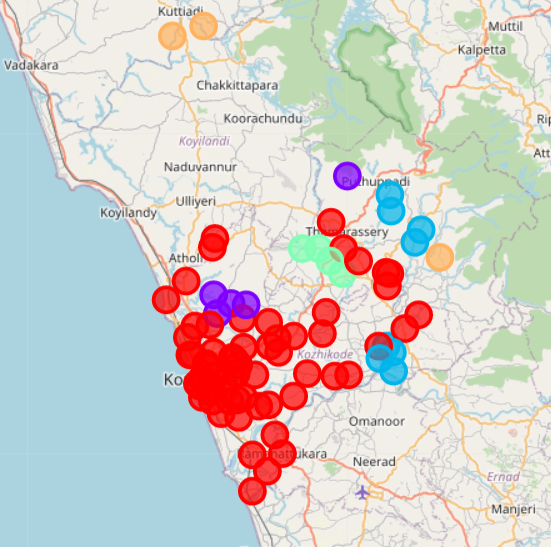


Figure 7: Calicut Clusters

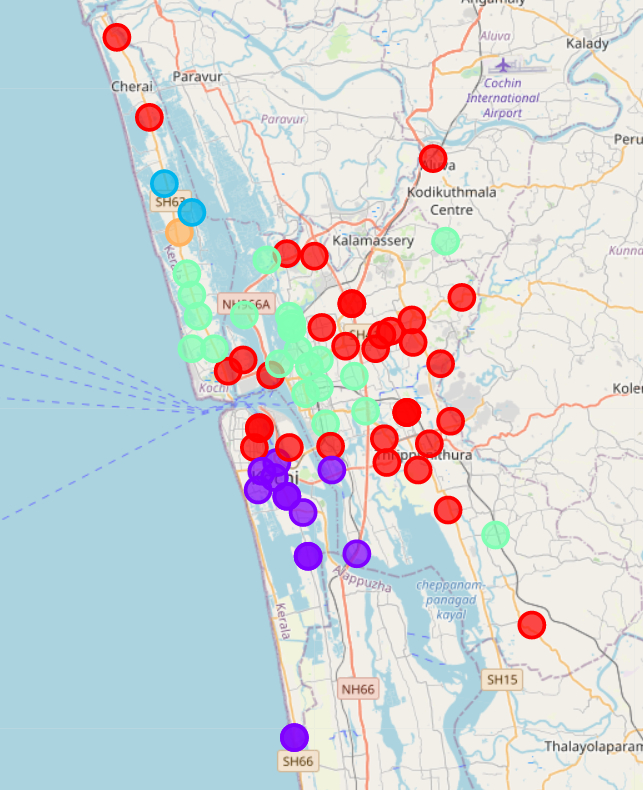


Figure 8: Kochi Clusters

## Conclusions

In this stud I have done a comparison of food choices between the cities of Calicut and Kochi. The type of restaurants present in these cities will be a reflection of its food choices. So, the study was done on the restaurants present in these cities. Though the restaurants types were similar, I could spot some differences. More restaurants which serves continental dishes were found in Kochi. It can be since Kochi is the financial capital of Kerala state and is an IT hub. Calicut has more interest in native recipes as well as middle eastern. Seeing the distribution of clusters, Calicut has one major cluster and other minor ones. But for Kochi, the cluster size is almost distributed. So, in Kochi you can expect foodies who like more variety of dishes in comparison with Calicut.

## Future Directions

The project is done on a wider scope. Further analysis can be done on specific categories in depth. The analysis can be also done neighborhood specific or incorporating lesser number of neighborhoods.