

# Choosing the best city for food tourism

Tiago Lobo

April 4, 2020

## 1. Introduction

Suppose you are visiting a new country and wants to eat a lot of Italian food. How would you choose which cities to visit based on the quality of Italian venues and on the sheer quantity of restaurants where you can eat Italian food?

This project aims to, given some cities, provide the best cities to visit based on the quality/quantity of restaurants of a chosen kind. It will target tourists that cares about the quality of the food there are going to eat during their trips.

## 2. Data acquisition and cleaning

This project will scrap data from Foursquare to get a given type of restaurant in a list of cities. As a test case, 4 cities will be considered:

- Maceio, AL
- Aracaju, SE
- Teresina, PI
- Fortaleza, CE
- Recife, PE

All these cities are in the northeast of Brazil and are considered great vacation destinies with gorgeous beaches and high-quality services for a low cost.

The type of cuisine considered was Italian, but the analysis made here can be replicated for any type of restaurant listed on Foursquare.

### 2.1. Data cleaning

Data was scraped from foursquare using their API, where json files containing relevant information is returned based on a given query.

The data was collected in two different stages:

- All venues of Italian food for each city were stored in a given dataset.

	Name	ID	Address	Lat	Lng
0	Del Popollo	4d3ccc91d2c8f04d02f56472	Av. Menino Marcelo	-9.611041	-35.720222
1	Maria Antonieta	4b8ed2aef964a520383933e3	Av. Antônio Gomes de Barros	-9.650901	-35.701036
2	Massarella	4b8ed2d6f964a5203e3933e3	R. José Pontes de Magalhães	-9.650114	-35.701579
3	Macarronada Do Lira	4d698f19342b8cfa66d4ca2c	Av. Belo Horizonte	-9.631723	-35.740608
4	Piccola Villa	5681c1dc498e4f2ee5f56ed6	NaN	-9.663231	-35.709074

*Maceió Italian Restaurants data set*

- All information of the venues listed on the previous dataset were stored on a different dataset.

ID	Name	Likes	Rating	Tips
4bf02477d4e4d13ae9e615a7	La Tavola	57.0	7.7	25.0
4ea9e267775bf0856954a1cc	La Cantina di Giovanni	72.0	7.7	45.0
531cacfb498e321a868f82fe	Villa Macarrone	12.0	7.7	6.0
4d5ff66329ef236a7c389f59	Cantina Ferreiro	109.0	8.4	36.0
50785ad3e4b00eb86095ddbd	Macarrones	91.0	8.5	39.0

*Maceió Italian Restaurants ratings data set*

Due to a foursquare free account limitation, the second stage of data acquisition was cut short. As a daily limit of 950 queries exist, ratings for just some venues were acquired during this stage. This data was cleaned (type conversion, removal of repeated data, removal of venues with missing ratings, etc.).

### 3. Approach

- Define type of cuisine and cities one wish to visit
- Use FourSquare API to find all restaurants of the chosen kind in each city
- Find rating, tips and likes counts for each venue using FourSquare API
- Sort data using the given ratings
- Visualize the data and the results

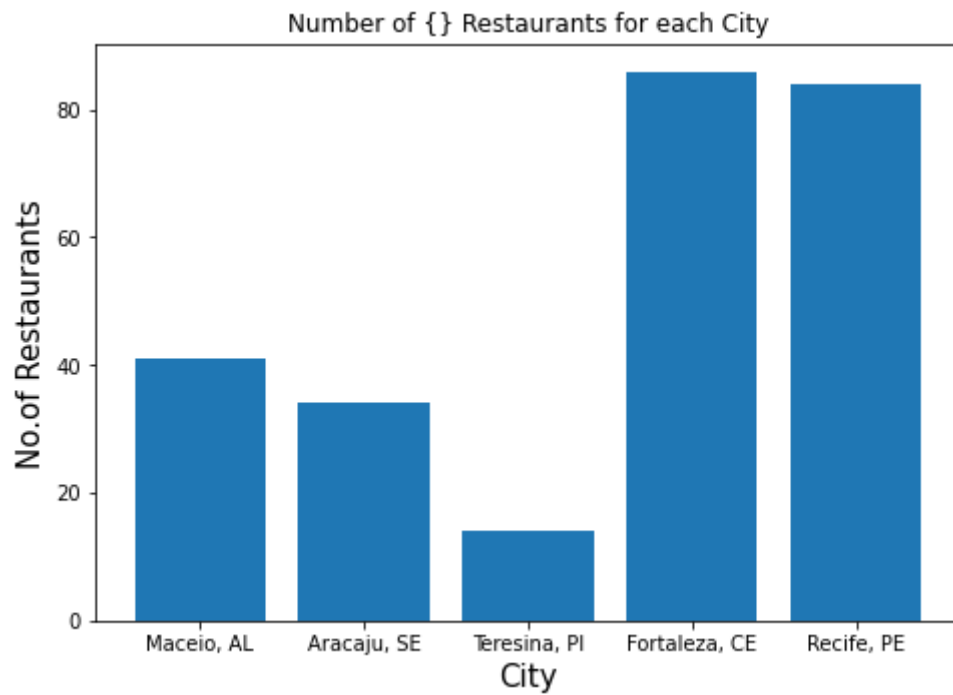
### 4. Analysis

Using python, we will retrieve venue information for each city using the foursquare api and the libraries: pandas, request, geopy, folium, matplotlib.

Pandas was used to correct treat the data set obtained from the foursquare api. Request was used to make the url calls for the foursquare api. Geopy was used to get the coordinates of the cities. Folium was used to visualize the results on a map. Matplotlib was used to plot the data.

#### 4.1. Exploratory data analysis and results

The first variable we checked was the number of Italian restaurants on each city. A bar graph was plotted.

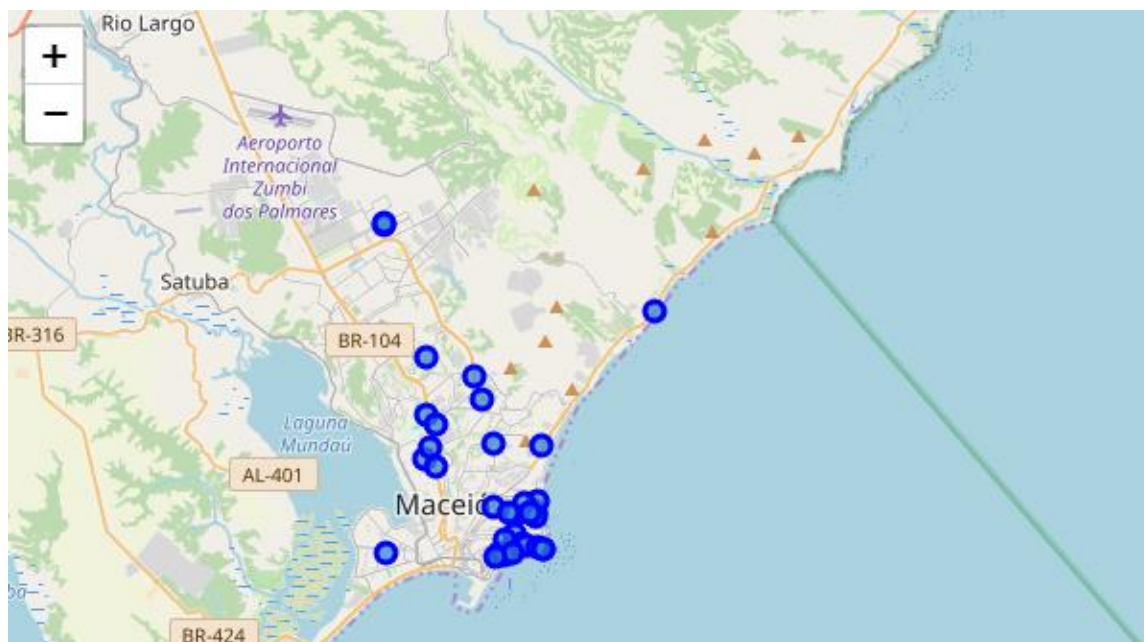


*Number of Italian Restaurants for each city*

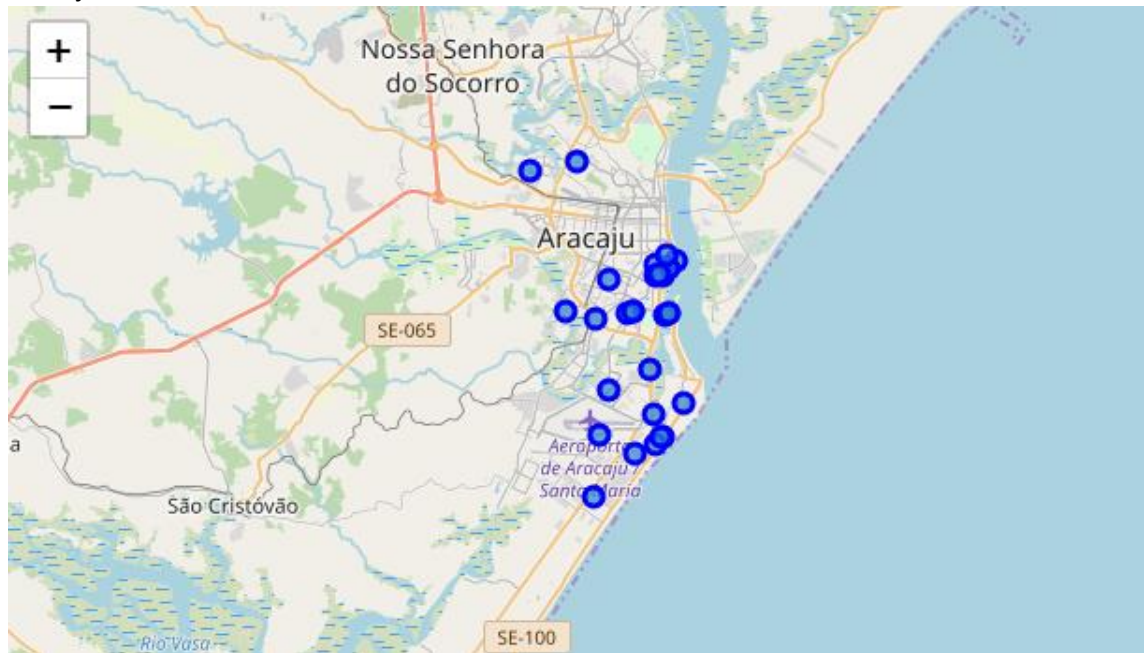
One can clearly see that Fortaleza has the highest number of Italian restaurants, closely followed by Recife. Teresina is the city with the least number of Italian restaurants.

In order to check the spatial distribution of the restaurants on each city, we plotted their locations on maps.

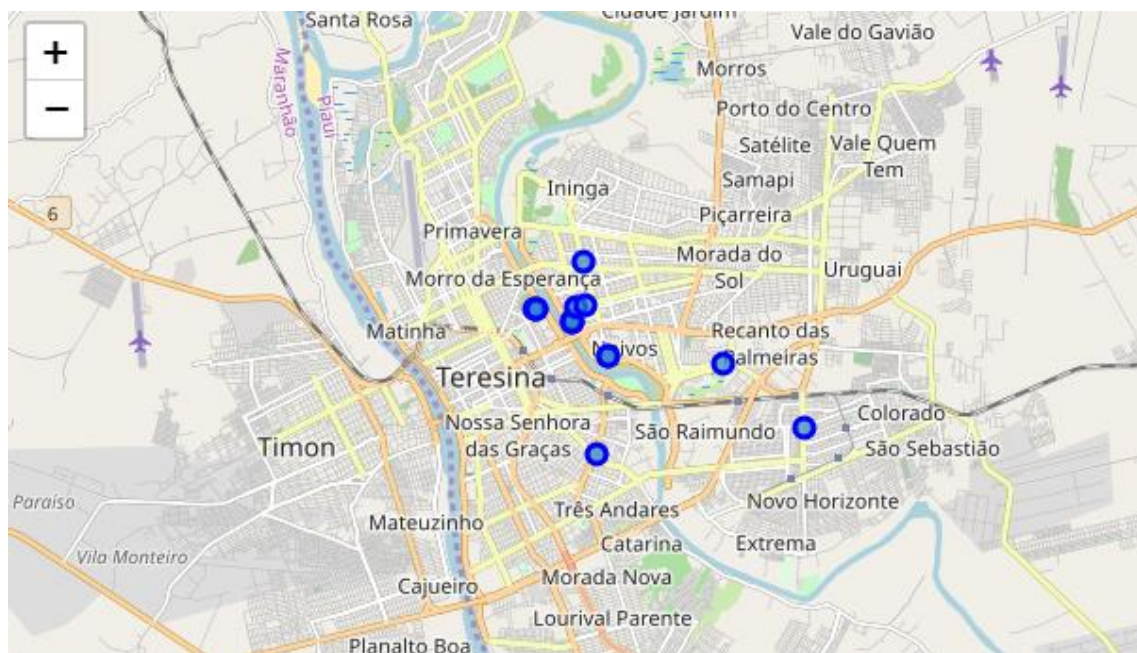
Maceió:



Aracaju:

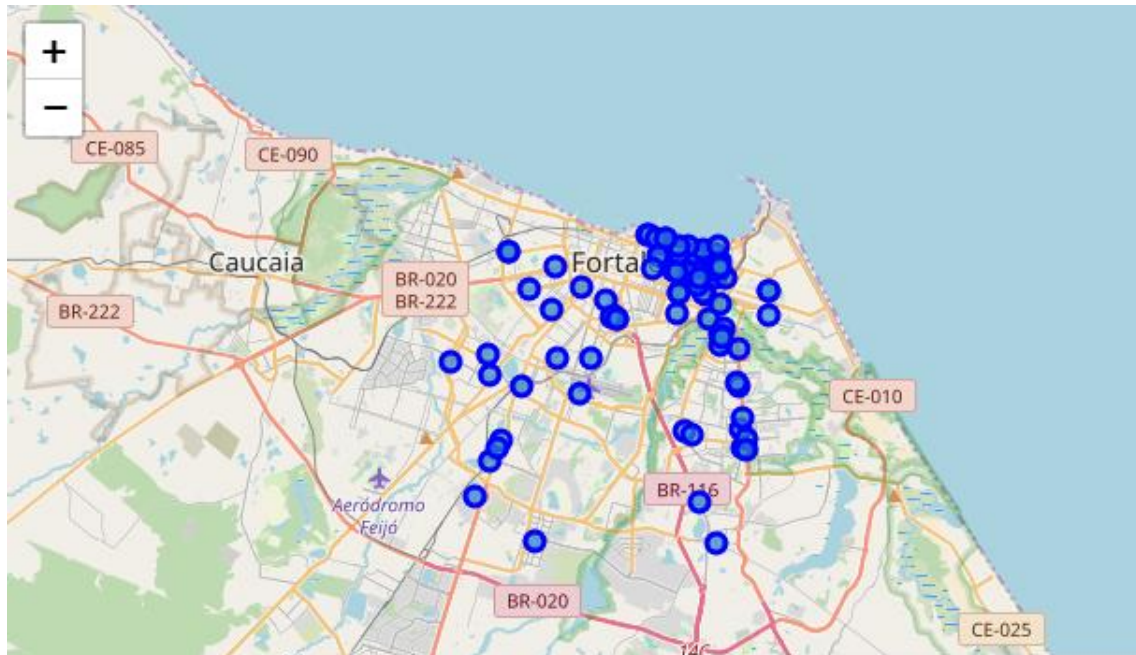


Teresina:

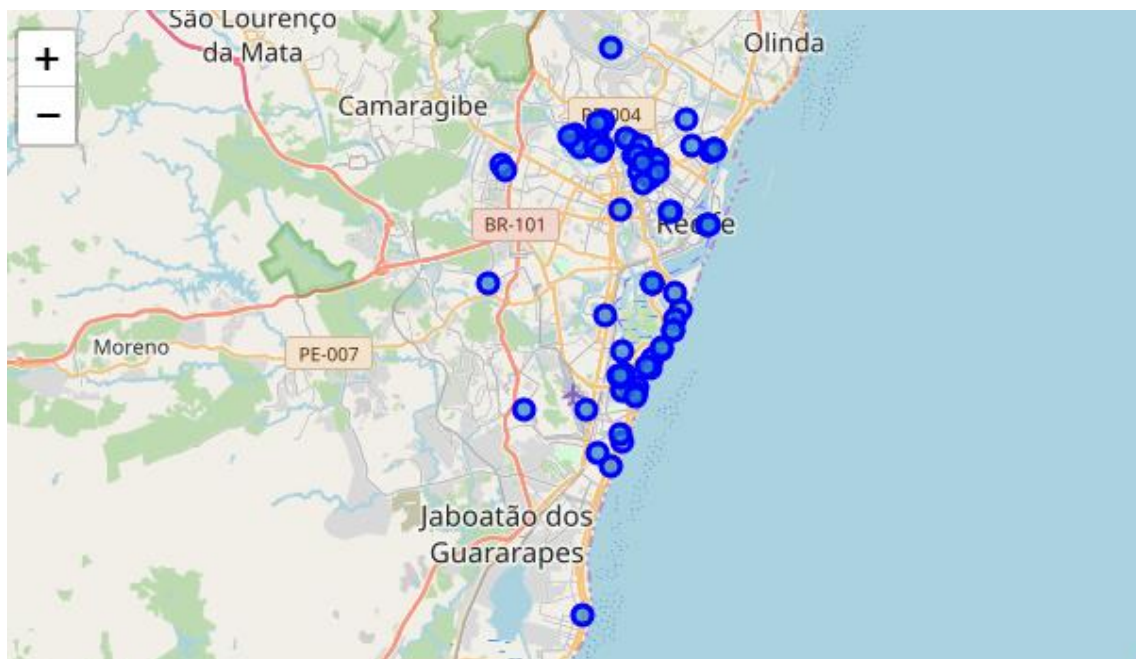




Fortaleza:



Recife



To quantitatively analysis the geolocation data we calculated the mean location of the restaurants for each city and the mean distance from this location for the restaurants.

	<i>Mean Distance from center (meters)</i>
<i>Maceio, AL</i>	0.033649453626477116
<i>Aracaju, SE</i>	0.028484988264272714
<i>Teresina, PI</i>	0.012530628699372442
<i>Fortaleza, CE</i>	0.035470805664267455
<i>Recife, PE</i>	0.042521865072031984

The restaurant with better ratings was Dell Popollo, in Maceió-AL with:

	ID	Name	Likes	Rating	Tips
29	4d3ccc91d2c8f04d02f56472	Del Popollo	252.0	9.2	133.0

## 5. Conclusions

Based on the Data Analysis and Visualization using Folium and Matplotlib we have rendered that Fortaleza is the city with most Italian restaurants, but Maceió was the city with the highest rated ones. In addition, both cities presented similar distances between Italian restaurants. This would indicate that the choice would be between the city with most restaurants and the city with the better ones. Thus, we provided numerical data that helps in the decision of the city one would visit to make Italian food tourism.