What and where to open a new food business

Business Problem

Investors and stakeholders of a new restaurant that will be opening in São Paulo, Brazil, are willing to take data-driven decisions on where and what type of restaurant to open.

Their main goal is to find out where the best place is, what type of restaurant or establishment and why.

What data will be used to answer the question

Two data sources will be used:

- Subway survey
    - An extensive survey had been carried out by São Paulo's subway in 2017
    - It has a lot of information about the city population, such as income, trips, from where to where and many other
    - It includes the coordinates of each neighborhood, and in each one many surveys were carried out
    - For this project, we will be using the 'centroid' of each zone with its average income, previously calculated
- Foursquare data
    - We will be using the centroids of each zone together with foursquare data using the foursquare API
    - We will acquire data from foursquare database to each zone, such as number and types of restaurants and other types of establishments relevant to our analysis

Methodology

Using the zone centroids of the city and the foursquare API, all establishments from each neighborhood were extracted from Foursquare's database.

After that, K-means algorithm was used to group similar neighborhoods in 5 different clusters.

Then, the API was used again, but this time, to extract only food venues from each neighborhood. With this, the most common food venues categories for each neighborhood had been identified.

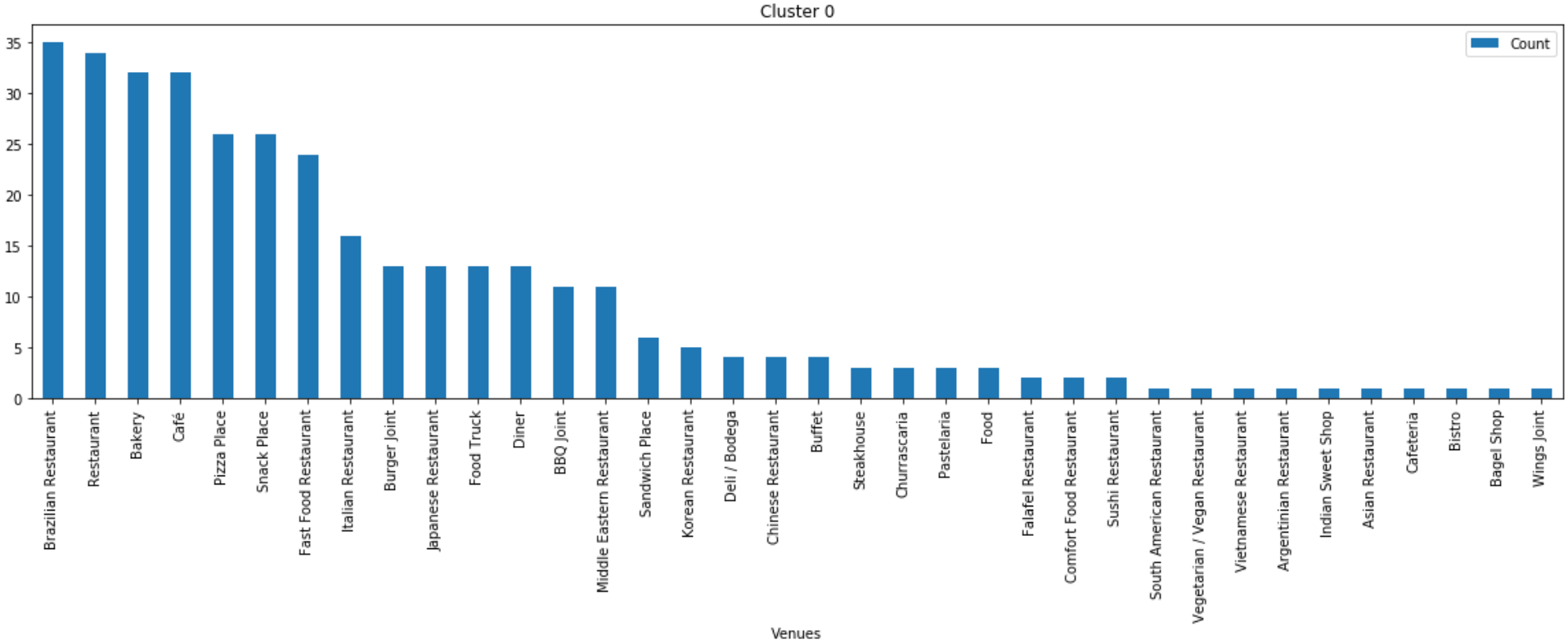
Crossing this dataset with the identified clusters from the previous step, we can then identify where and which kind of business to open, considering each neighborhood in a cluster is similar and therefore have the same interests.

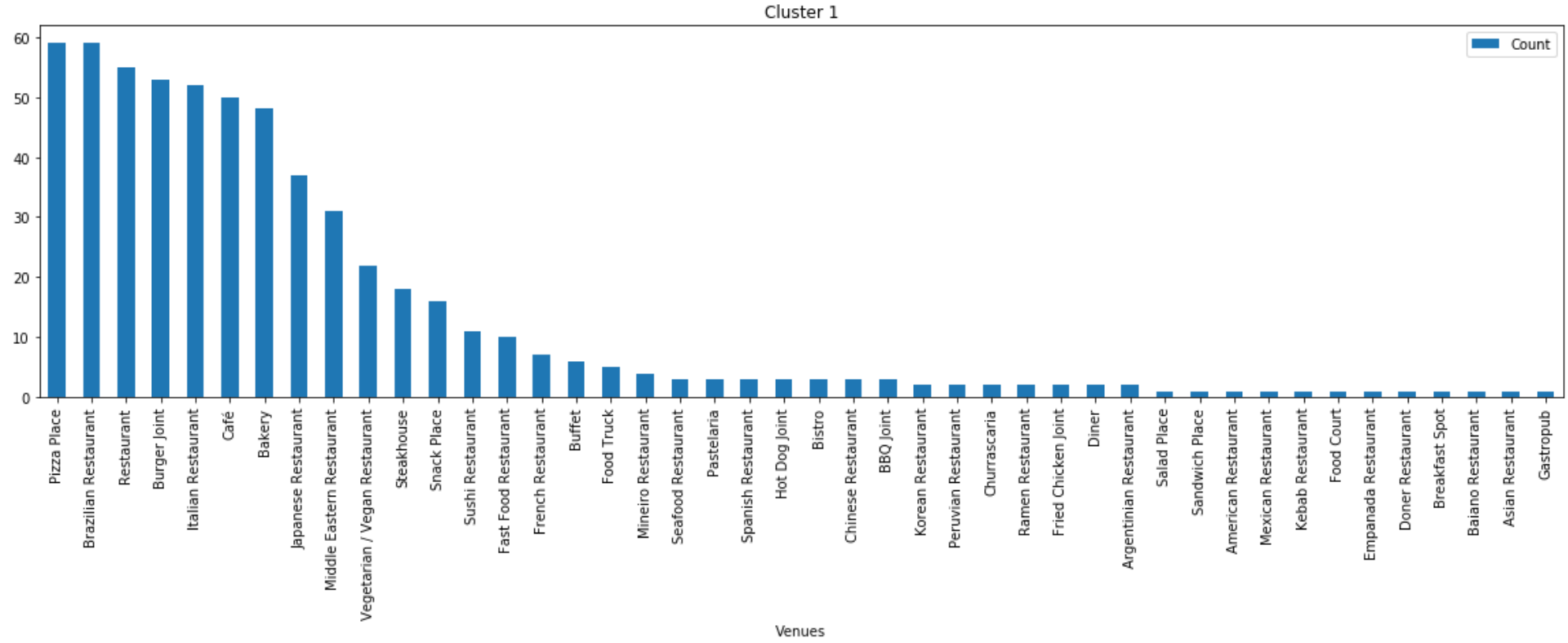
With this in hand, there is a plentiful of options yet. So, the top neighborhoods on income were examined so we can maximize our revenue by opening our establishment on the highest average income neighborhoods.

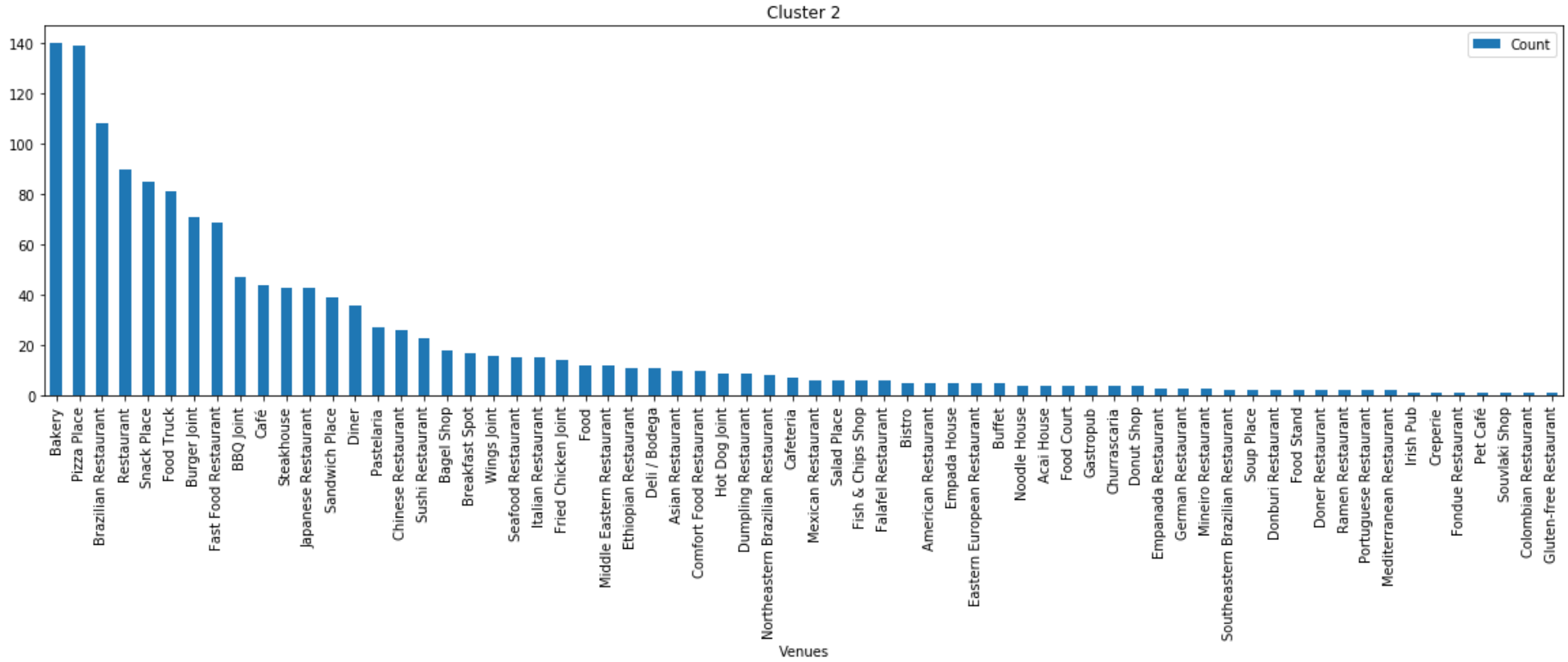
Results

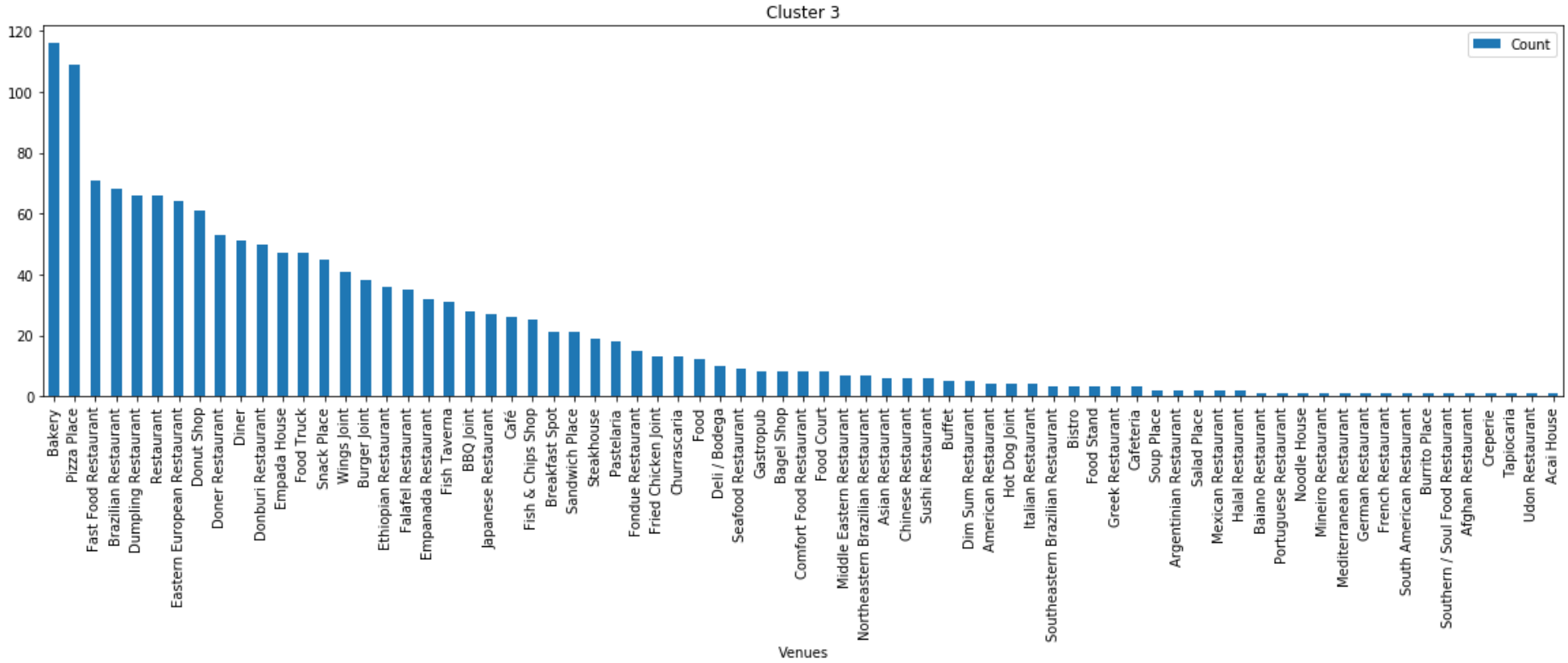
The following graphs shows the quantity of each venue category on each cluster.

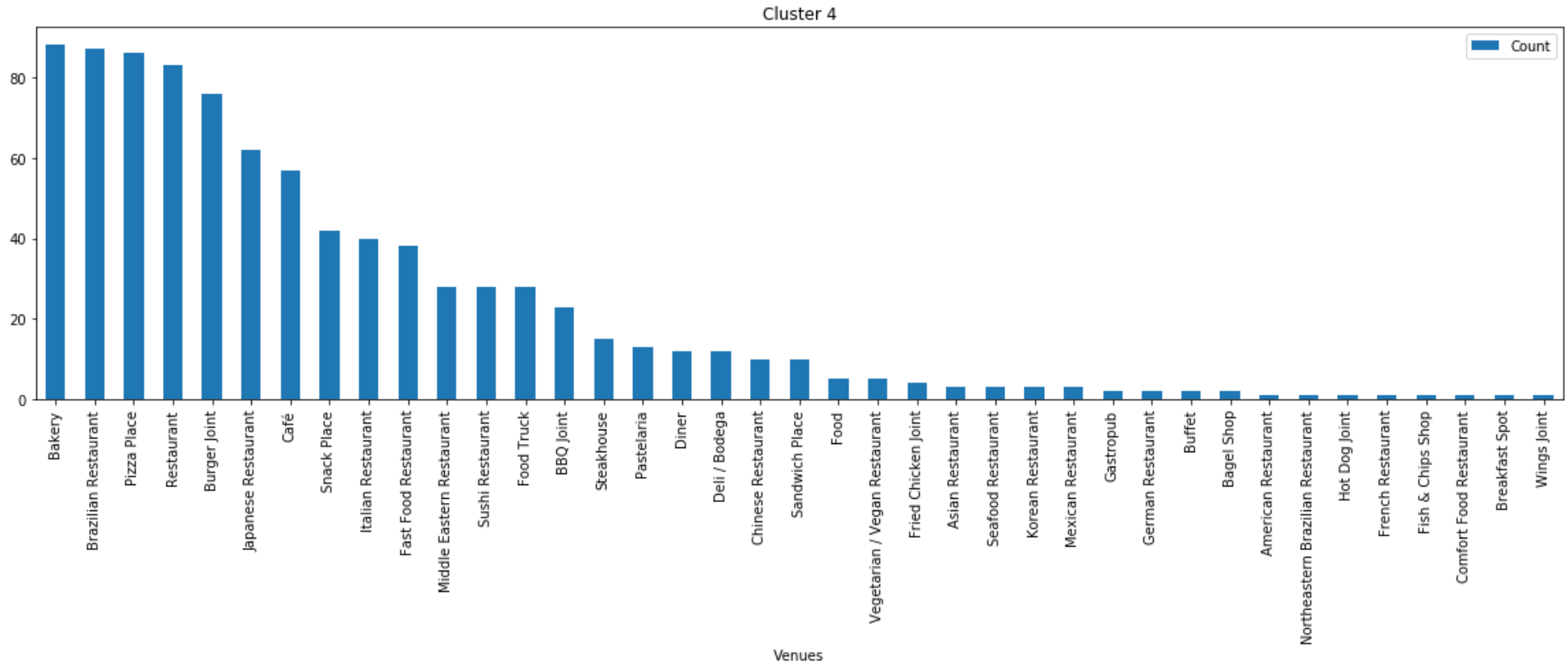
We can already see some opportunities: for example, in a neighborhood in cluster 0, a wings joint or bagel shop are good options. A Brazilian restaurant however can be too much competition.


Cluster 0
Count
35
30
25
20
15
10
5
0
Brazilian Restaurant
Restaurant
Bakery
Café
Pizza Place
Snack Place
Fast Food Restaurant
Italian Restaurant
Burger Joint
Japanese Restaurant
Food Truck
Diner
BBQ Joint
Middle Eastern Restaurant
Sandwich Place
Korean Restaurant
Deli / Bodega
Chinese Restaurant
Buffet
Steakhouse
Churrascaria
Pastelaria
Food
Falafel Restaurant
Comfort Food Restaurant
Sushi Restaurant
South American Restaurant
Vegetarian / Vegan Restaurant
Vietnamese Restaurant
Argentinian Restaurant
Indian Sweet Shop
Asian Restaurant
Cafeteria
Bistro
Bagel Shop
Wings Joint
Venues


Cluster 1
Count
60
50
40
30
20
10
0
Pizza Place
Brazilian Restaurant
Restaurant
Burger Joint
Italian Restaurant
Café
Bakery
Japanese Restaurant
Middle Eastern Restaurant
Vegetarian / Vegan Restaurant
Steakhouse
Snack Place
Sushi Restaurant
Fast Food Restaurant
French Restaurant
Buffet
Food Truck
Mineiro Restaurant
Seafood Restaurant
Pastelaria
Spanish Restaurant
Hot Dog Joint
Bistro
Chinese Restaurant
BBQ Joint
Korean Restaurant
Peruvian Restaurant
Churrascaria
Ramen Restaurant
Fried Chicken Joint
Diner
Argentinian Restaurant
Salad Place
Sandwich Place
American Restaurant
Mexican Restaurant
Kebab Restaurant
Food Court
Empanada Restaurant
Doner Restaurant
Breakfast Spot
Baiano Restaurant
Asian Restaurant
Gastropub
Venues


Cluster 2
Count
140
120
100
80
60
40
20
0
Bakery
Pizza Place
Brazilian Restaurant
Restaurant
Snack Place
Food Truck
Burger Joint
Fast Food Restaurant
BBQ Joint
Café
Steakhouse
Japanese Restaurant
Sandwich Place
Diner
Pastelaria
Chinese Restaurant
Sushi Restaurant
Bagel Shop
Breakfast Spot
Wings Joint
Seafood Restaurant
Italian Restaurant
Fried Chicken Joint
Food
Middle Eastern Restaurant
Ethiopian Restaurant
Deli / Bodega
Asian Restaurant
Comfort Food Restaurant
Hot Dog Joint
Dumpling Restaurant
Northeastern Brazilian Restaurant
Cafeteria
Mexican Restaurant
Salad Place
Fish & Chips Shop
Falafel Restaurant
Bistro
American Restaurant
Empada House
Eastern European Restaurant
Buffet
Noodle House
Acai House
Food Court
Gastropub
Churrascaria
Donut Shop
Empanada Restaurant
German Restaurant
Mineiro Restaurant
Southeastern Brazilian Restaurant
Soup Place
Donburi Restaurant
Food Stand
Doner Restaurant
Ramen Restaurant
Portuguese Restaurant
Mediterranean Restaurant
Irish Pub
Creperie
Fondue Restaurant
Pet Café
Souvlaki Shop
Colombian Restaurant
Gluten-free Restaurant
Venues


Cluster 3
Count
120
100
80
60
40
20
0
Bakery
Pizza Place
Fast Food Restaurant
Brazilian Restaurant
Dumpling Restaurant
Restaurant
Eastern European Restaurant
Donut Shop
Doner Restaurant
Diner
Donburi Restaurant
Empada House
Food Truck
Snack Place
Wings Joint
Burger Joint
Ethiopian Restaurant
Falafel Restaurant
Empanada Restaurant
Fish Taverna
BBQ Joint
Japanese Restaurant
Café
Fish & Chips Shop
Breakfast Spot
Sandwich Place
Steakhouse
Pastelaria
Fondue Restaurant
Fried Chicken Joint
Churrascaria
Food
Deli / Bodega
Seafood Restaurant
Gastropub
Bagel Shop
Comfort Food Restaurant
Food Court
Middle Eastern Restaurant
Northeastern Brazilian Restaurant
Asian Restaurant
Chinese Restaurant
Sushi Restaurant
Buffet
Dim Sum Restaurant
American Restaurant
Hot Dog Joint
Italian Restaurant
Southeastern Brazilian Restaurant
Bistro
Food Stand
Greek Restaurant
Cafeteria
Soup Place
Argentinian Restaurant
Salad Place
Mexican Restaurant
Halal Restaurant
Baiano Restaurant
Portuguese Restaurant
Noodle House
Mineiro Restaurant
Mediterranean Restaurant
German Restaurant
French Restaurant
South American Restaurant
Burrito Place
Southern / Soul Food Restaurant
Afghan Restaurant
Creperie
Tapiocaria
Udon Restaurant
Acai House
Venues


Cluster 4
Count
0
20
40
60
80
Bakery
Brazilian Restaurant
Pizza Place
Restaurant
Burger Joint
Japanese Restaurant
Café
Snack Place
Italian Restaurant
Fast Food Restaurant
Middle Eastern Restaurant
Sushi Restaurant
Food Truck
BBQ Joint
Steakhouse
Pastelaria
Diner
Deli / Bodega
Chinese Restaurant
Sandwich Place
Food
Vegetarian / Vegan Restaurant
Fried Chicken Joint
Asian Restaurant
Seafood Restaurant
Korean Restaurant
Mexican Restaurant
Gastropub
German Restaurant
Buffet
Bagel Shop
American Restaurant
Northeastern Brazilian Restaurant
Hot Dog Joint
French Restaurant
Fish & Chips Shop
Comfort Food Restaurant
Breakfast Spot
Wings Joint
Venues

Now, crossing the previous results with the average income of the top neighborhoods in each cluster:

For cluster 0:

Neighborhood	AvgIncome	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Butantã	11909	0	Restaurant	Brazilian Restaurant	Pizza Place	Food Truck	Bakery	Burger Joint	Café	Buffet	Fast Food Restaurant	Diner

Neighborhood	Wings Joint	Bagel Shop
Butantã	0	0

We can see that there are no wings joints or bagel shops on Butantã, a high-income neighborhood in the same cluster of other neighborhoods where people like wings and bagels. So, these are two opportunities for our new business.

We can do the same for the other clusters and then obtain the following “best opportunities” with this methodology.

Discussion

With this simple methodology, we already identified some opportunities out there for our new business. Relevant examples are:

- Wings joints and bagel shops on Butantã;
- Gastropubs and Asian restaurants around FAAP;
- Irish Pubs and Fondue Restaurants on Alto de Pinheiros;
- Açai House or Tapiocaria around Jóquei Clube;
- Wings Joint and Fish & Chips Shop on Jardim Vitória Régia.

Conclusion

We have successfully identified new food businesses around the city, using a clusterization technique to group similar neighborhoods and analyzing popular venues within each group but not in specific neighborhoods, which they are missing, so that is an opportunity.

However, we have used a relatively shallow approach to a much complex problem. The foursquare API only allows to extract up to 100 venues per call with a limit of calls per day and hour.

We can always further refine our analysis by adding relevant data and focusing our analysis on certain regions or neighborhoods, bringing up new analysis for a more robust result.