

Food Trucks and Venues in General in Campinas, SP - BR

A Study of Geographical Opportunities for the Businesses in the City, with Emphasis on Food Trucks



Problematic

And Interested Parties

In this project two problems were addressed. The initial one opened the possibility to provide a second useful resource during the studies development. Assuming there is a client interested in learning which would be the top five locations in the city to open a food truck business. Investors often depend on intel in order to determine optimal business implementation decisions, such as the tendency a given area have to generate them profit based on a business model. Food trucks have become popular in Brazil over the last few years, having already passed the peak point of the trend. The latter is a crucial point regarding the future of a business model, and those who depend solely on trend tend to fade away after the trend does, so it is imperative not only to investigate whether the food truck business model is still present in the region, but also, if so, in which specific locations; is there a concentration, or a optimum spot? Those questions are the ones that make up this study's problem. There is already an informal impression that the business model is still present and attractive to many customers after the trend has passed, although the model appears to have changed in some ways, with the observation that the food trucks tend to be more present in areas that already have the appeal for their services, such as events, parties, entertainment venues and so on, often counting with more than a few options, of varied cousines; which could be seen as a collective approach to maximizing profitability and efficiency when it comes to the use of time. Therefore, the final objective is to produce such intel for the client, specifying locations in which the food truck business model tends to succeed the most in the city of Campinas, São Paulo, Brazil. The other problem addressed in this study, is the development of analytical material regarding the neighborhoods in the city and the categories of venues associated with each one, which in turn allows the clustering of neighborhoods based upon their similarities; such material is valuable and can be used for future client's request or to indirectly help solve other problems.

Data Description

and Plan of Engagement

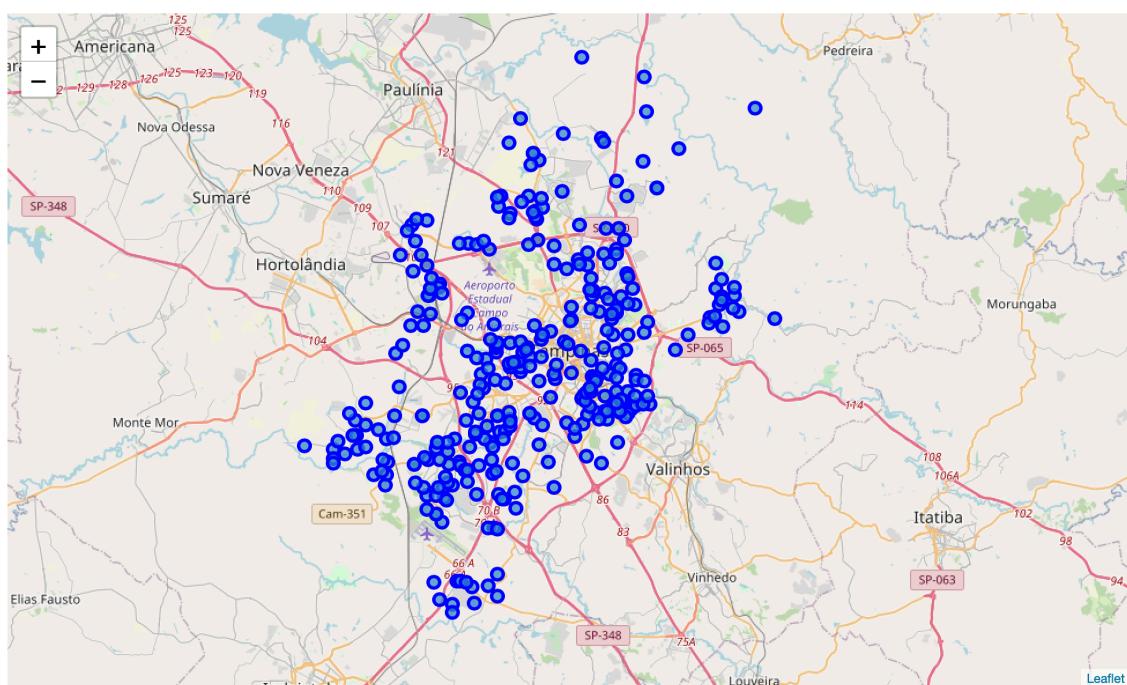
Regarding the data that is to be used in this study, the starting point to be described is the neighborhood data about Campinas. A table containing the Neighborhoods names was collected from a previous study, and with some pointers from a colleague, from my universities extension program, that studies geography, a data processing step was able to take place in the QGIS software, that based on the neighborhood names produced their respective latitudes and longitudes coordinates. With the neighborhood dataframe assembled, the next step of the study is to query the Foursquare API in order to retrieve the venues close to each neighborhood, in a radius of 500m, and a limit of 100 venues returned per neighborhood. The query from Foursquare API will then be used to create a dataframe containing each neighborhoods specific venue categories, as well as their density. The data specified will be employed in the development of a K-Means clustering algorithm, doing so based on the venues categories collected and present in each neighborhood. The intent of the study is to identify the cluster to which food trucks belong to, and then locate the top five locations that are both compatible with that business model and also minimally populated by competitors.

Methodology

The project's database was acquired in “.csv” format, which was then read into a “pandas” dataframe. In the table below, we have presented the first five lines of the database, containing the columns “Neighborhood”, “Long” (Longitude) and “Lat” (Latitude).

	Neighborhood	Long	Lat
0	Arruamento Fain José Feres	-47.084462	-22.821267
1	Arruamento Luiz Vicentin	-47.080811	-22.829629
2	Bairro das Palmeiras	-47.016018	-22.895889
3	Barão Geraldo	-47.079099	-22.845249
4	Bosque das Palmeiras	-47.041516	-22.789775

With the data loaded and ready on the environment, a map was produced, with points indicating each neighborhood location throughout the city. For the development of the map, the “geolocator” package was used to find the center point of the city, which was then fed into the “folium” package, which in turn was responsible for the generation of the map and plotting of the neighborhood points. The generated map is presented below:



Following the generation of the map, the Foursquare API was called in order to return each neighborhoods venues, with the conditions that the return should include the radius of 500m around that given neighborhood, and also only include the first 100 venues (in case there are more venues than 100 in the radius specified). Foursquare's API returned the venues, alongside their categories, which were saved in the form of a pandas dataframe, an excerpt of it can be seen below:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Arruamento Fain José Feres	-22.821267	-47.084462	Praça do Coco	-22.821576	-47.082588	Plaza
Arruamento Fain José Feres	-22.821267	-47.084462	Feirinha da Praça do Côco	-22.821189	-47.082803	Arts & Crafts Store
Arruamento Fain José Feres	-22.821267	-47.084462	Quintal De Casa Gourmet	-22.821398	-47.083359	Gastropub
Arruamento Fain José Feres	-22.821267	-47.084462	Namaste Salad	-22.821280	-47.083065	Restaurant
Arruamento Fain José Feres	-22.821267	-47.084462	Bronco Burger	-22.823686	-47.083310	Burger Joint
...
Ville Sainte Helene	-22.896746	-46.989720	Barley Club	-22.894975	-46.991331	Bar
Ville Sainte Helene	-22.896746	-46.989720	Paris 89 Restaurante	-22.893388	-46.991835	Restaurant
Ville Sainte Helene	-22.896746	-46.989720	Brakstube Padaria Alemã	-22.895352	-46.991241	Bakery
Ville Sainte Helene	-22.896746	-46.989720	Pastel Da Feira - Sousas (Kushi)	-22.895255	-46.991179	Food Truck
Ville Sainte Helene	-22.896746	-46.989720	Rapa Do Tacho	-22.895072	-46.991331	Coffee Shop

There were a total of 241 different categories returned by the query. In the next step, a K-Means algorithm is to be fitted to the data, but in order to do so it is necessary that the labels with the neighborhood and venues names are removed, alongside other features that must be dropped in order to estimate the model, such as latitudes and longitudes. After creating the dataframe that is to be used in the K-Means algorithm, the last step before estimation of the clusters is to one-hot encode the categories, transposing them as dummy variables, having the categories as column and the row values as binary, in such a way that the venues category is marked with a “1”, while the other categories are marked with a “0”. Here is a slice from the dataframe with the dummy variables, with the number of observations and categories confirmed below in a print statement:

Neighborhood	Acai House	Accessories Store	Airport	Airport Lounge	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Garage	Auto Workshop
0 Arruamento Fain José Feres	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1 Arruamento Fain José Feres	0	0	0	0	0	0	0	0	0	1	0	0	0	0
2 Arruamento Fain José Feres	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3 Arruamento Fain José Feres	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4 Arruamento Fain José Feres	0	0	0	0	0	0	0	0	0	0	0	0	0	0

```
1 print("The resulting dataframe has", cps_onehot.shape[0], "observations of venues on different neighborhoods, an
```

The resulting dataframe has 3127 observations of venues on different neighborhoods, and 241 categories of venues.

In order to understand the proportion in which the different venues categories are present in each neighborhood individually, a “groupby” operation was performed, taking the mean of the dummy variables for each neighborhood, that way we get a resulting dataframe that has only one observation for each neighborhood, with that being representative of the proportion and categories of venues present within.

Still before the development of the model, sorted frequency tables were produced to fuel future insights about specific neighborhood that may be requested on future projects. Below we have a representation of the resulting frequency table for two neighborhoods, as an example:

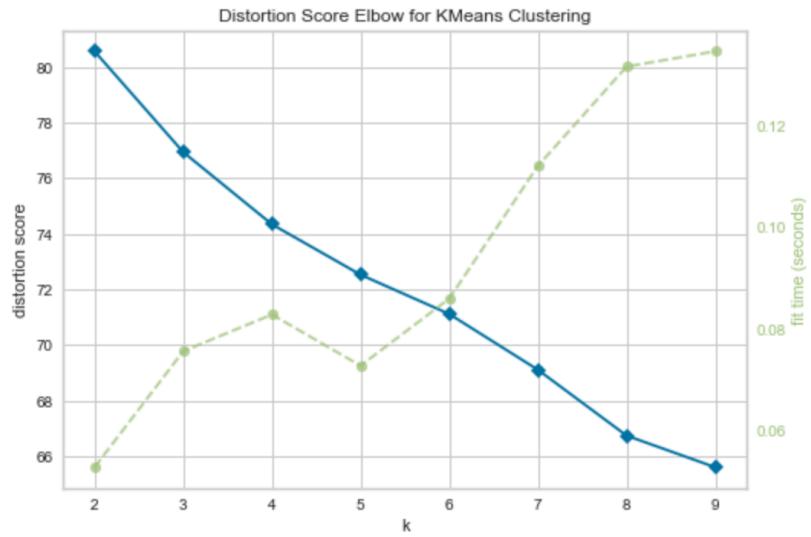
```
----Jardim Chapadão----
      venue freq
0  Brazilian Restaurant 0.17
1        Pizza Place 0.13
2            Bar 0.07
3       Bakery 0.07
4     BBQ Joint 0.03
```

```
----Jardim Cidade Satélite Íris----
      venue freq
0      Soccer Field 0.33
1    Grocery Store 0.33
2       Bakery 0.33
3     Acai House 0.00
4  Peruvian Restaurant 0.00
```

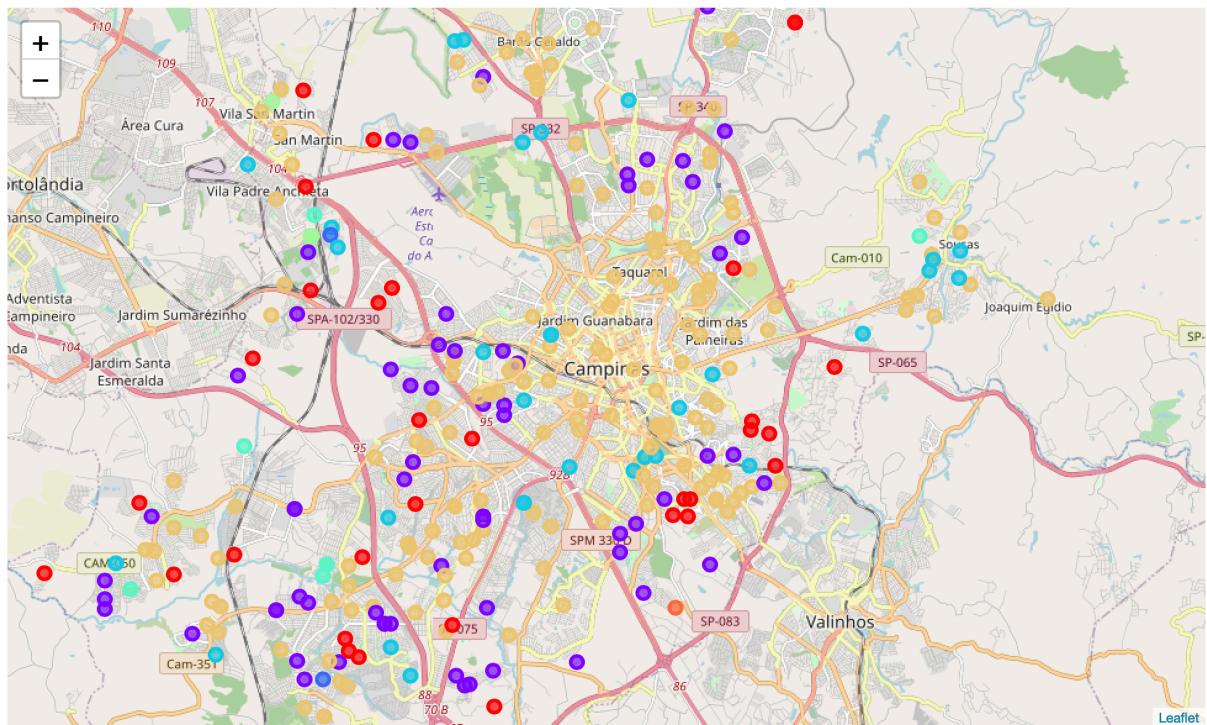
From the one-hot encoding table produced, it was possible to assemble a summary table containing the neighborhood names, and their first to tenth most common venues within. Below we have a slice of the resulting table presented:

Neighborhood	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
0 Arrumento Fain José Feres	Pizza Place	Brazilian Restaurant	Bar	Brewery	Café	Hot Dog Joint	Hotel	Gym / Fitness Center	Restaurant	Burger Joint
1 Arrumento Luiz Vicentin	Bar	Brazilian Restaurant	Pizza Place	Café	Brewery	Ice Cream Shop	Vegetarian / Vegan Restaurant	Snack Place	Supermarket	Beer Bar
2 Bairro das Palmeiras	Restaurant	Pool	Gym	Brazilian Restaurant	Tennis Court	BBQ Joint	Lake	Social Club	Volleyball Court	Diner
3 Barão Geraldo	Hardware Store	Restaurant	Bus Stop	Café	Zoo Exhibit	Fast Food Restaurant	Food Stand	Food Court	Food & Drink Shop	Food
4 Bosque das Palmeiras	Pizza Place	Restaurant	Pet Store	Cocktail Bar	Fast Food Restaurant	Field	Financial or Legal Service	Fish & Chips Shop	Zoo Exhibit	Farm
...
315 Vila Tofanello	Liquor Store	Pizza Place	Bakery	Beer Store	Gym	Electronics Store	Arts & Crafts Store	Indian Restaurant	Food	Flower Shop
316 Vila Trinta E Um de Marco	Shopping Mall	Spa	Motorcycle Shop	Comfort Food Restaurant	Bakery	Field	Food Truck	Food Stand	Food Court	Food & Drink Shop
317 Vila Ângela Marta	Bakery	Brazilian Restaurant	Farmers Market	Burger Joint	Food & Drink Shop	Fruit & Vegetable Store	Big Box Store	Market	Sushi Restaurant	Sandwich Place
318 Village Campinas	Food Truck	Brazilian Restaurant	Zoo Exhibit	Cupcake Shop	Forest	Food Stand	Food Court	Food & Drink Shop	Food	Flower Shop
319 Ville Sainte Helene	Bakery	Food Truck	Bar	Coffee Shop	Restaurant	Fishing Spot	Field	Financial or Legal Service	Fish & Chips Shop	Zoo Exhibit

Now, for the K-Means algorithm to be implemented, we first had to decide how many clusters to specify to be created. During that process, it was necessary to employ trial and error approaches to verify which value of K (amount of clusters) would be optimal. The Elbow Method for determining K was somewhat inconsistent, and for that reason that a few options had to be tested. K-values of 3; 4; 8; 12 and 13 were tested, before determining that K=8 would be the optimal parameter based on the observation of the resulting composition of each cluster. Below we have one of the many Elbow Method graph for selecting K produced, having in this one the best most coherent representation of the distortion score, and the intermediate elbow point at K=8. For higher K-values, such as some of the ones indicated by the algorithm as optimum, we have noticed distortion within the groups, similarly to the one observed with low values of K, also recommended by the algorithm.

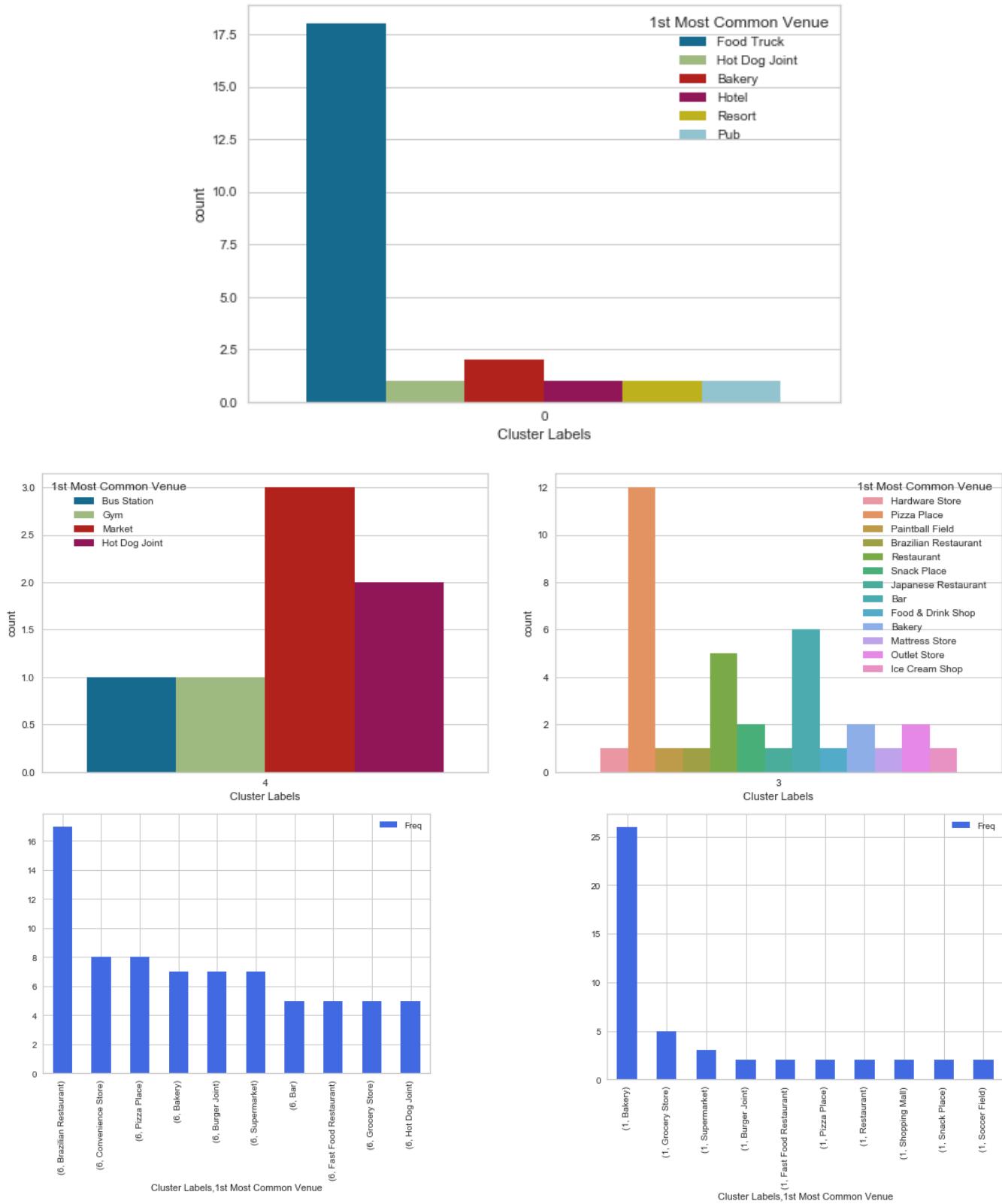


With that decided, the K-Means algorithm was implemented, resulting in the clusters labels, that were then inserted into the table containing the neighborhoods and their respective first to tenth most common venues. With that sorted out, a map was plotted including the neighborhoods and their respective cluster, the latter identified by unique colors and a label on each point, that upon a mouse click, present the information of the neighborhood name and its cluster number. The plotted map is presented below:



After the plotting of the map, the resulting data was processed and the cluster labels were merged with the dataframe containing individual venues observations per

neighborhood, in order to create a tables and bar graphs for each cluster to visualize and identify similarities in venues categories there present. The tables are available in the Jupyter Notebook associated with this study, and some of the bar graphs follow:



Finally, based upon the graphs and tables, the cluster number “0” was identified as representative of neighborhoods that are compatible with food trucks, once we can observe that there is a high concentration of that business model in the neighborhoods assigned to that cluster, being an almost exclusive cluster, having integrated a low proportion of other venues categories, that are mostly associated with social and routine supply activities, such as bakeries and pubs. Therefore, we will be investigating cluster number “0” in order to determine the best locations to implement a food truck in Campinas. To do so, it was required to produce a table containing all the neighborhood associated with that cluster, accompanied by an indicator of the proportion of food trucks present in each one. The indicators were sorted with ascending values, being the top rows the neighborhoods in that cluster that have the least amount of food trucks. A slice of the table, with relevant information, is presented below:

Neighborhood	Food Truck
Jardim Anchieta	0.142857
Jardim Andorinhas	0.166667
Nova Aparecida	0.166667
Jardim Profilurb	0.166667
Jardim Santa Lúcia	0.200000
Dic Vi	0.200000
Jardim Amazonas	0.200000
Parque Brasília	0.200000
Jardim Itayu	0.200000
Jardim Londres	0.200000
Residencial Sirius	0.200000
Vila Ipe	0.200000
	Vila Georgina
	0.222222
	Jardim Itatiaia
	0.250000
	Vila Aeroporto III
	0.250000
	Vila Boa Vista
	0.250000
	Núcleo Residencial Bairro da Vitória
	0.285714
	Parque Via Norte
	0.333333
	Conjunto Habitacional Residencial Olímpia
	0.400000
	Parque dos Pomares
	0.500000
	Jardim Novo Sol
	0.500000
	Jardim Monte Alto
	0.500000
	Parque Pomares
	0.500000
	Village Campinas
	0.500000

Discussion

Based on the study developed, we can conclude that neighborhoods pertaining to the cluster number “0” are the ones where a food truck business is most likely to succeed. In addition to that, it is the understanding of this study that an investor must prioritize neighborhoods within that cluster that are less saturated with competitors. Based on that concept the table presented at page 10 is the best indicator of where should a food truck business be implemented on. The first few rows represent the neighborhoods that pertain to the food truck cluster (“0”) that are also populated the least with that business model, indicating that the investor would have less resistance from the market, in the form of competitors, and also less resistance from the consumers, that tend not to have a favorite or frequently visited food truck already on their preferences as a customer, leaving important “leg room” from the investor to setup his business and conquer the neighborhood’s customers, those that serve as doorways for customers that don’t reside on the same neighborhood, in the form of personal indications based on the quality of both the food and service observed. With a higher tendency of neighborhood locals to visit the venue, due to the higher visibility of a business that have few competitors in the region, the amount of customers should be initially greater than it would be on a neighborhood with many already established competitors, that may have won over many customers loyalty already. Therefore, the implementation of the food truck business model in the below indicated top five locations tend to optimize and accelerate the solidification of the venues name and reputation.

Neighborhood	Food Truck
Jardim Anchieta	0.142857
Jardim Andorinhas	0.166667
Nova Aparecida	0.166667
Jardim Profilurb	0.166667
Jardim Santa Lúcia	0.200000

Conclusion

This study's conclusion is that the top five neighborhoods in which a food truck venue should be implemented in the city of Campinas, São Paulo, Brazil, in May 2020, would be, in descending order: Jardim Anchieta; Jardim Andorinhas; Nova Aparecida; Jardim Profilurb and Jardim Santa Lúcia. The top indication is Jardim Anchieta, but other variables must be taken into account, this serving as a starting point with ranked options, that should then be investigated individually, giving preference to neighborhoods that rank higher in the recommendation. For the implementation of the venue, a regional study must be performed in order to ascertain that neighborhood's compatibility with food truck business models based on variables not included in this recommendation study, such as mean level of income for that given neighborhood, the average age of residents and the magnitude of night time activities, all of which should influence the investor when choosing one of the neighborhoods recommended, being the optimal neighborhood one that is recommended and also compatible with the venues intended target audience, price levels and operation hours.

During the development of this recommendation, important data was gathered referring to the city of Campinas and its establishments. Such data may be very useful in the future, not only for the food truck case, but also to determine the ideal location in the city for many venues categories, such as, for example, bakeries, that although are understandably distributed along different clusters, are also noticeably concentrated on the cluster number "1".

The venues location is a critical point in the steps of implementing a business model, and this is where the value of this study's products reside, with enough data produced to aid in different situations and requests that may be performed by different clients in the future.