

THE BATTLES OF NEIGHBORHOODS

EXPLORING FOOD TYPES AROUND
TRAIN STATIONS IN WEST ZONE -
RIO DE JANEIRO



INTRODUCTION

The propose of this project is research what are the most common types of restaurants around three train stations on the west zone of Rio de Janeiro (Bangu, Campo Grande and Santa Cruz). These train stations are located in very populous neighborhoods, near to their commercial area with: boardwalks, business centers, shopping malls and schools. These train stations are integrate with bus lines and BRT Terminal (Campo Grande and Santa Cruz) making these areas have a large flow of people not only inside the stations, but also in their surroundings. With this, it is expected to draw food consumption profile of each analyzed group.

DATA DESCRIPTION

- **Data Requirements for this project:**

- Neighborhoods Information (name and coordinates).
- Venue informations (name, category and coordinates)

- **Data Sources for this project:**

- Wikipedia Pages;
- Foursquare API.

- **Data Processing for this project:**

- Data cleaning is required;
- Data needs to be in a structured format.

	id	name	categories	lat	lng	distance	Bairro
123	528aa04211d2eff2b3e0a5a2	Gago's Burguers	Snack Place	-22.915258	-43.687668	502	Santa Cruz
124	4d2c76fe888af04d1aa3e8af	Ponto Com Sabor	Brazilian Restaurant	-22.920622	-43.681520	390	Santa Cruz
125	51918ab4498e63587e264797	Pastel Crokante	Snack Place	-22.913197	-43.687312	629	Santa Cruz
126	51d6eb2be4b076c0db605e54	Sabores da Tia Ellen	Restaurant	-22.916426	-43.683418	133	Santa Cruz
127	5730aacc498e0ee5fcb92a0f	Lanchonete Nan Chen	Snack Place	-22.918447	-43.682493	137	Santa Cruz
128	56c307d5cd10bb6461a46328	Iaris' bela confeitaria	Snack Place	-22.917876	-43.684325	89	Santa Cruz
129	514c47e5e4b0852c37337e22	Loris' Bella	Juice Bar	-22.917945	-43.680866	271	Santa Cruz
130	4edf7babe5faee99926441fb	Casa do Kilo	Restaurant	-22.916342	-43.681952	212	Santa Cruz
131	52200aec11d2780596b2b98d	Pimenta Carioca Sopas	Food Truck	-22.916086	-43.684372	193	Santa Cruz
132	501aa84be4b07abf182f9eb2	Restaurante Copo De Leite	Restaurant	-22.918580	-43.682050	182	Santa Cruz
133	502995a2e4b0372608df2eb9	Pastel Com Tudo Dentro	Food Truck	-22.915947	-43.683534	186	Santa Cruz
134	53223f6b498ed923eed7e79c	Ravi's Sucos & Lanches	Snack Place	-22.917706	-43.681458	208	Santa Cruz
135	5910e3fcd4cc981e837b9340	Pastelone	Snack Place	-22.917809	-43.681610	193	Santa Cruz
136	5227a8e911d21c6b6232cf33	L@nches.com	Snack Place	-22.918371	-43.681644	206	Santa Cruz
137	527ee4e611d27152869afb1c	Dan Pizzas Express	Pizza Place	-22.922729	-43.682720	573	Santa Cruz
138	521e6e9e498eb042b4cb0858	Folheados e Cia	Snack Place	-22.918325	-43.681432	224	Santa Cruz
139	4f3d1e9ee4b0a92482b7c0d7	pastelaria dragao de ouro	Chinese Restaurant	-22.918812	-43.680954	291	Santa Cruz
140	50992ef9498e20f6977406f	Sônia's lanches	Brazilian Restaurant	-22.919576	-43.678511	554	Santa Cruz
141	5491f04c498e40fd4db61dad	Moderninho	Snack Place	-22.918707	-43.680710	309	Santa Cruz
142	5740c470498e6896a498fae9	Pimenta Maluca	Pie Shop	-22.920914	-43.681620	413	Santa Cruz

METHODOLOGY

- **Data web Scraping**

- To scrap the neighborhoods from Wikipedia;
- To make data into a structured format.

```
1 html_data = requests.get('https://pt.wikipedia.org/wiki/Lista_de_bairros_do_Rio_de_Janeiro_')
```

```
1 soup = BeautifulSoup(html_data, 'html.parser')
```

```
1 rio_table = soup.find_all(name = 'table')
```

```
1 table_str = str(rio_table)
2 df_rio_bairro = pd.read_html(table_str)[0]
3 df_rio_bairro.head()
```

	Nº	Bairro	R.A	Região	IDS
0	1	Lagoa	Lagoa	Zona Sul	0.854
1	2	Leblon	Lagoa	Zona Sul	809.000
2	3	Ipanema	Lagoa	Zona Sul	801.000
3	4	Humaitá	Botafogo	Zona Sul	798.000
4	5	Urca	Botafogo	Zona Sul	795.000

METHODOLOGY

- **Foursquare**

- Retrieving venues informations from API.

```
1 df_bg_venues = dataframe_filtered_bg[['id','name','categories','lat','lng','distance']]
2 df_bg_venues['Bairro'] = "Bangu" #let's create a column with the neighborhood
3 df_bg_venues.tail(2)
```

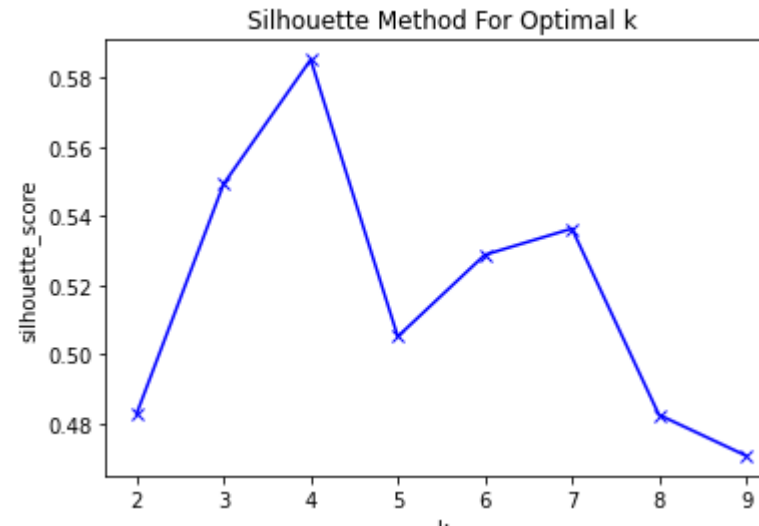
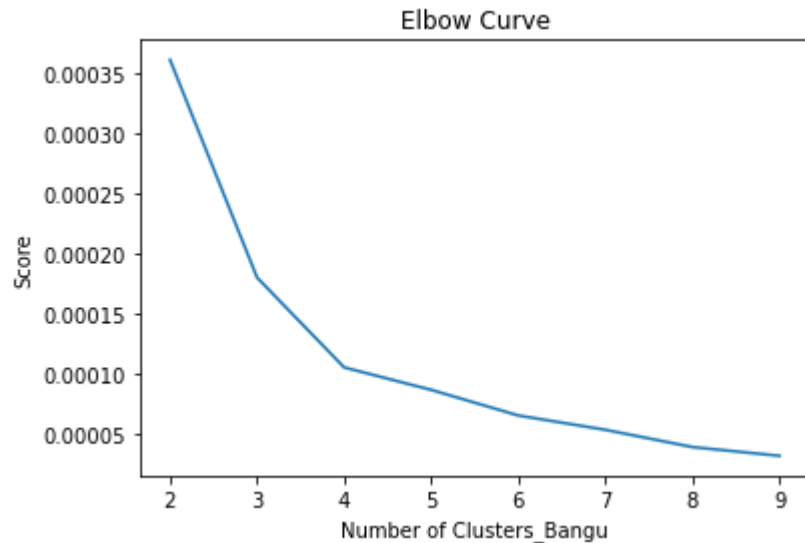
	id	name	categories	lat	lng	distance	Bairro
44	528d4d29498e6c0940b9da91	Ravis Sucos e Lanches	Snack Place	-22.879053	-43.459799	667	Bangu
45	5671cab2498ea992ff4f84ae	Resenha Bar & Grill	Bar	-22.878174	-43.467165	396	Bangu

```
1 print('{} venues were returned by Foursquare.'.format(df_bg_venues.shape[0]))
```

46 venues were returned by Foursquare.

METHODOLOGY

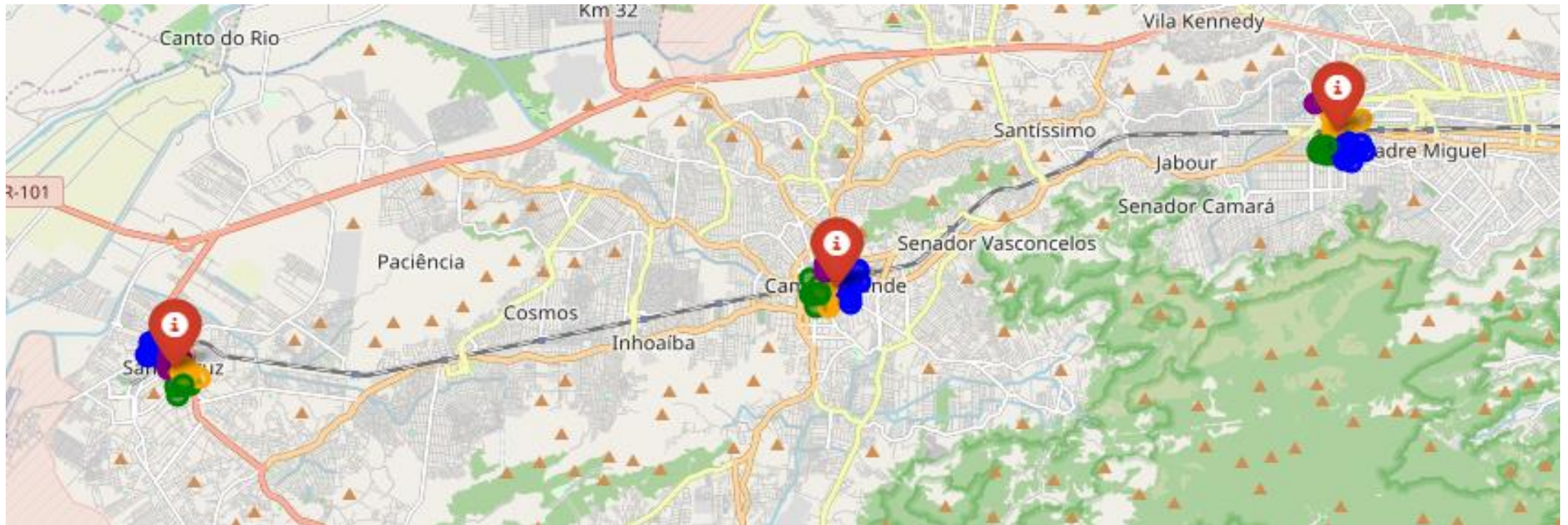
- **K-Means Unsupervised Machine Learning**
 - Segment and cluster similar venues based on the geolocation.



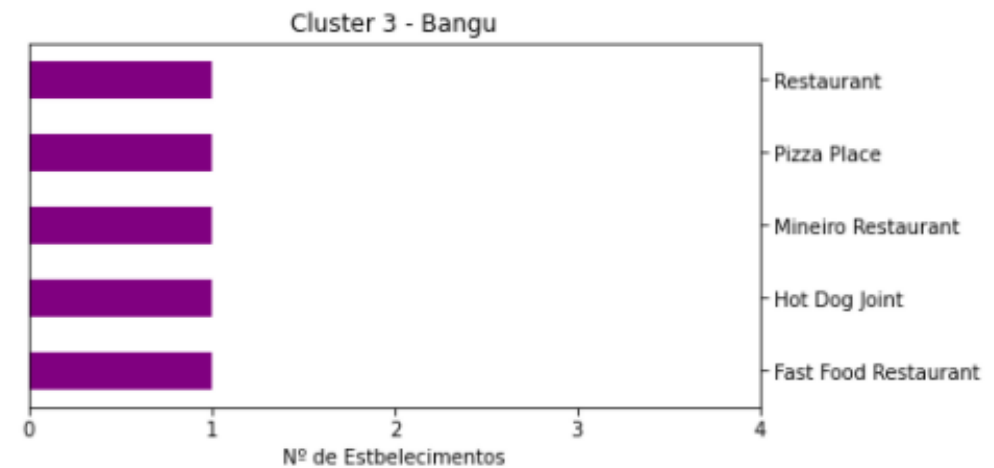
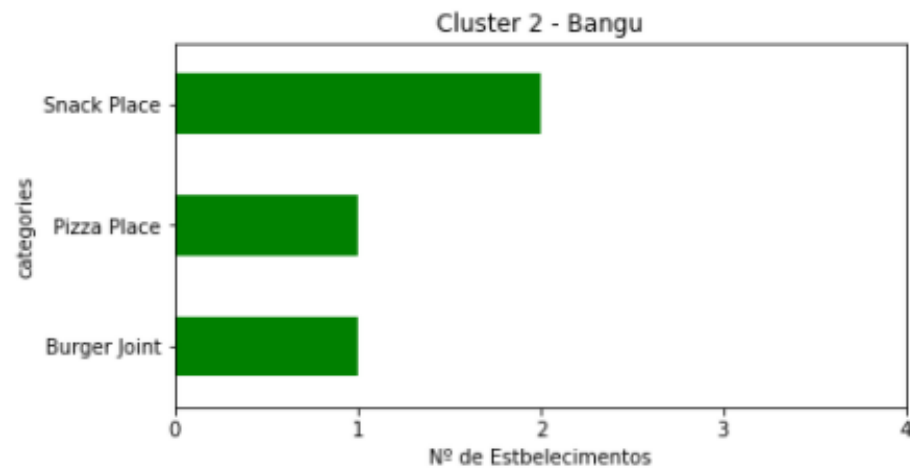
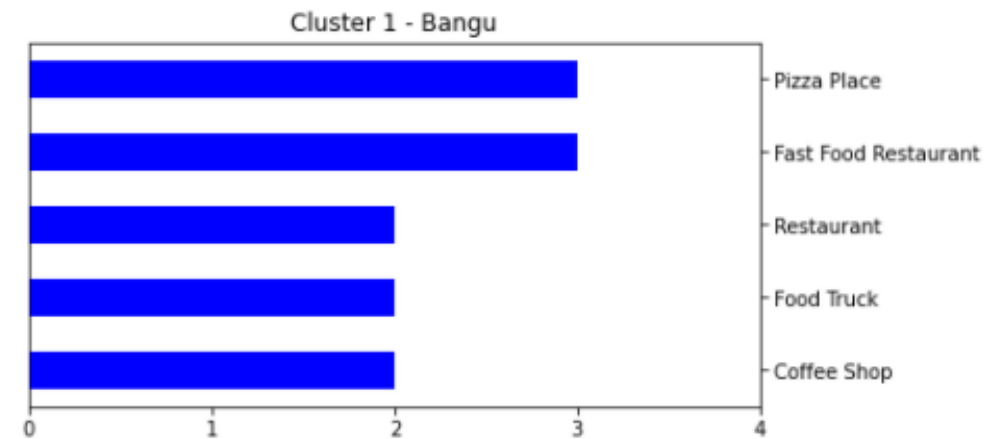
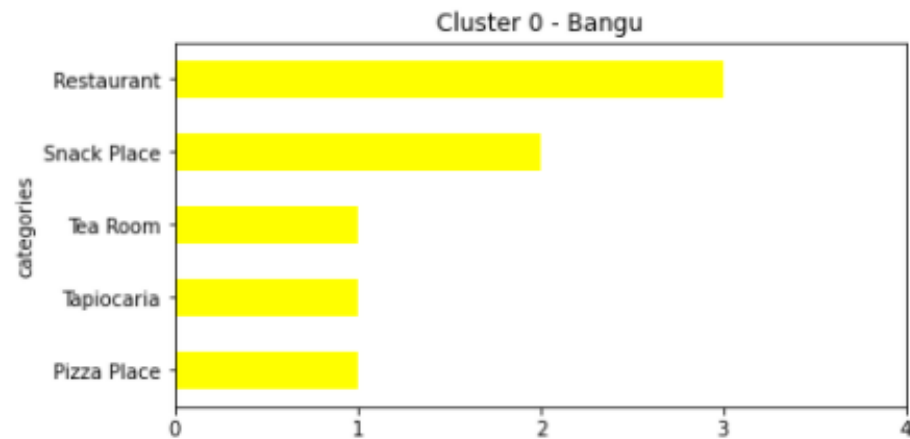
```
N_cluster: 2, score: 0.48281474825934917
N_cluster: 3, score: 0.5493154591108179
N_cluster: 4, score: 0.5852903506831614
N_cluster: 5, score: 0.5183728642217134
N_cluster: 6, score: 0.537585766452794
N_cluster: 7, score: 0.48946458493743483
N_cluster: 8, score: 0.4997127525554461
```


METHODOLOGY

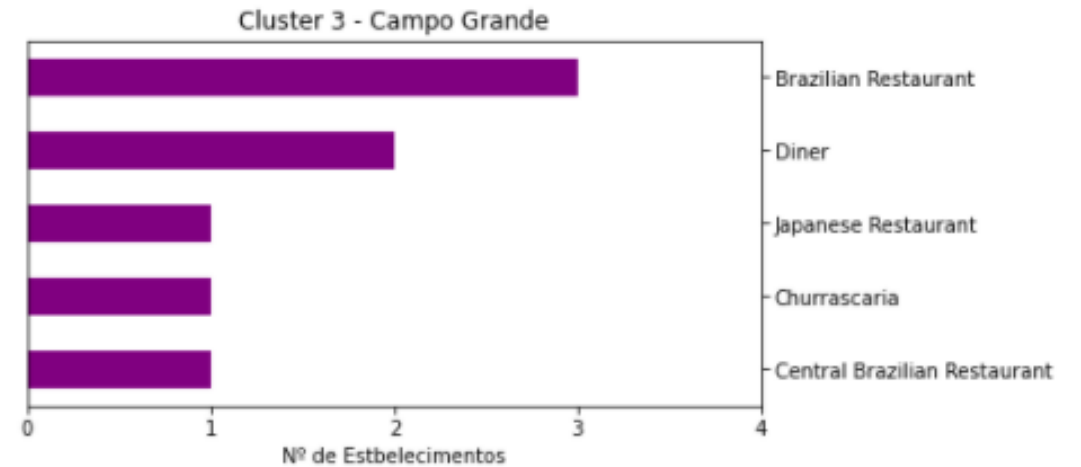
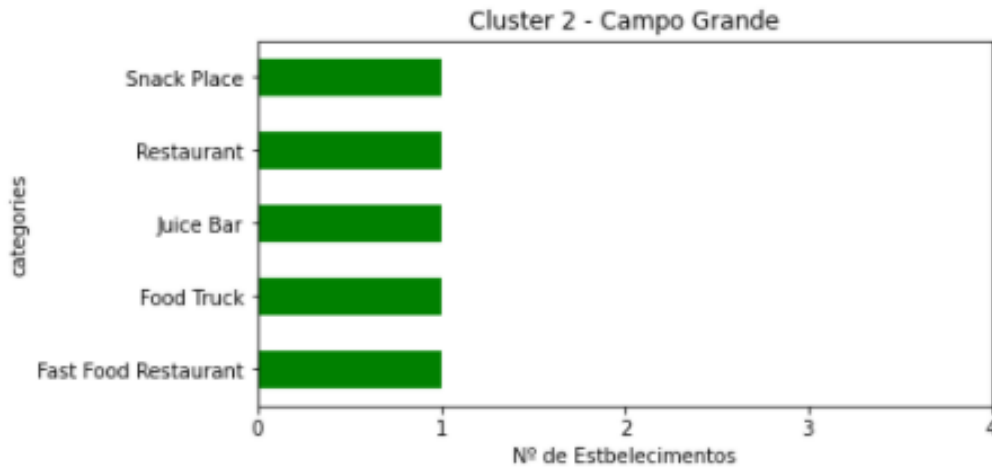
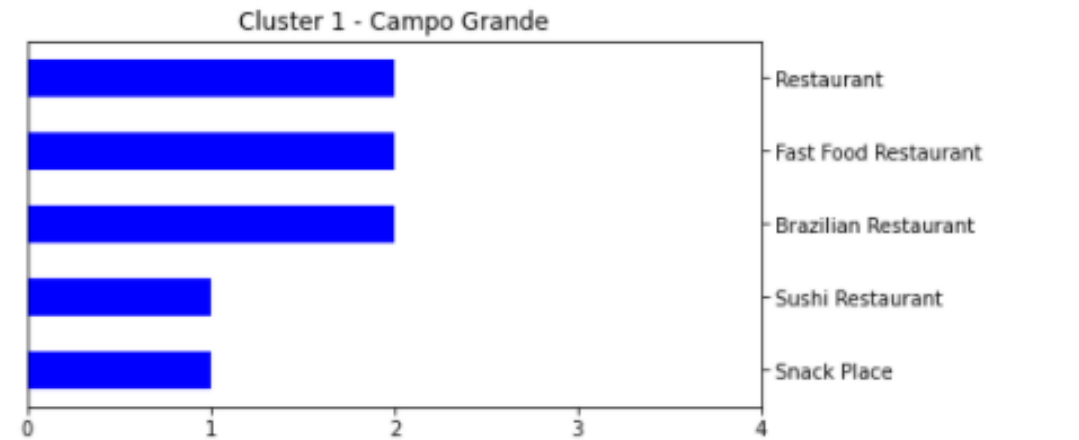
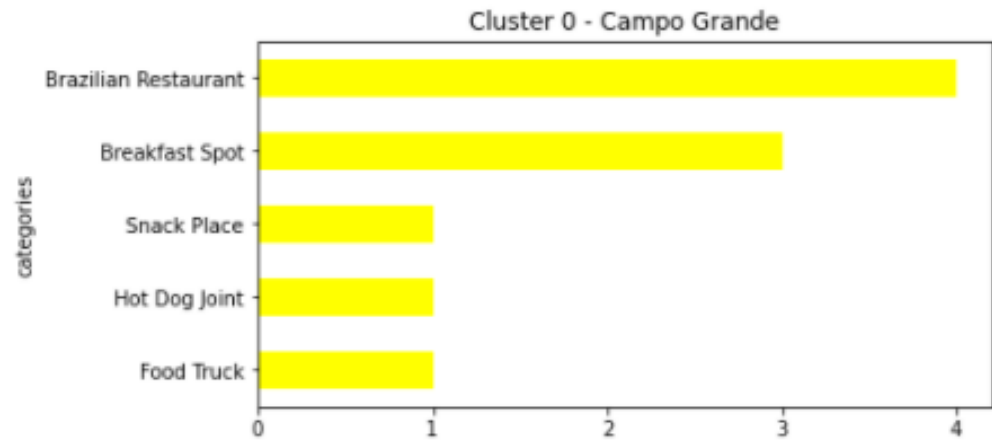
- **K-Means Unsupervised Machine Learning**
 - Segment and cluster similar venues based on the geolocation.



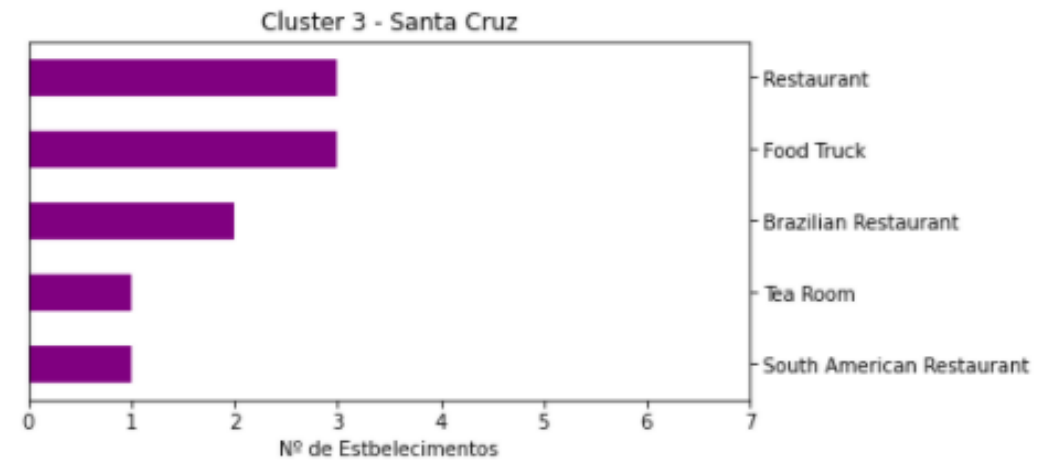
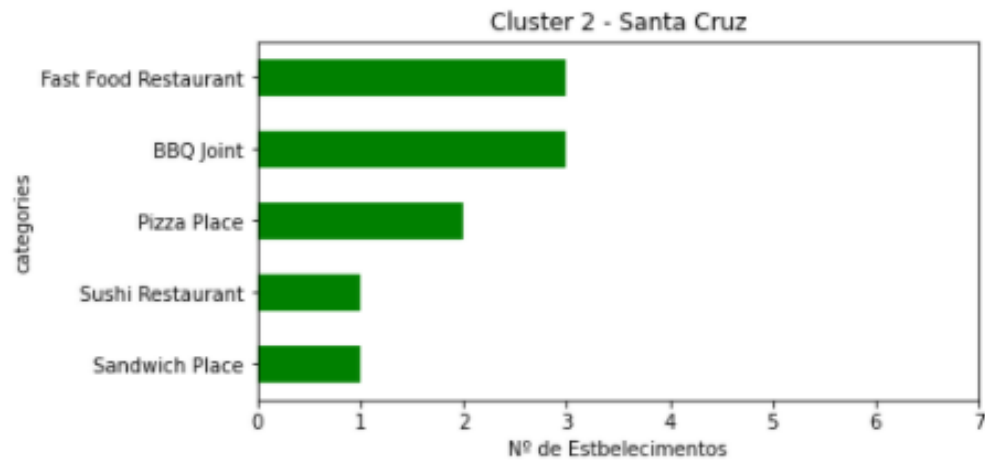
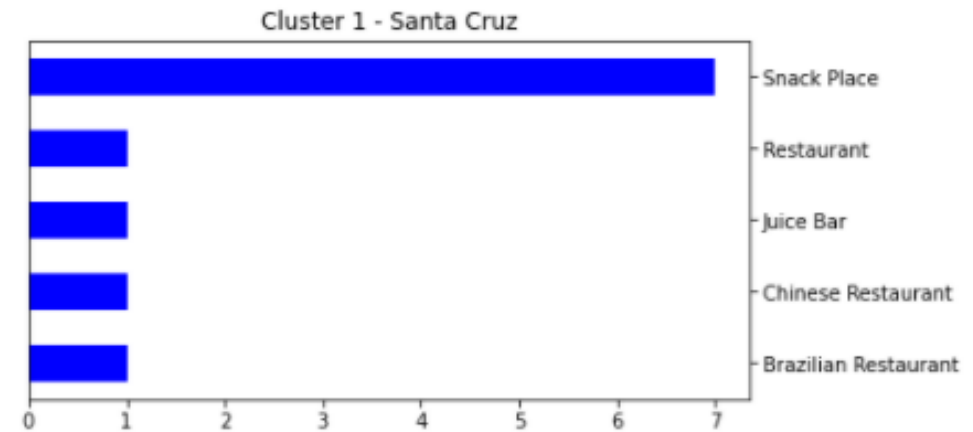
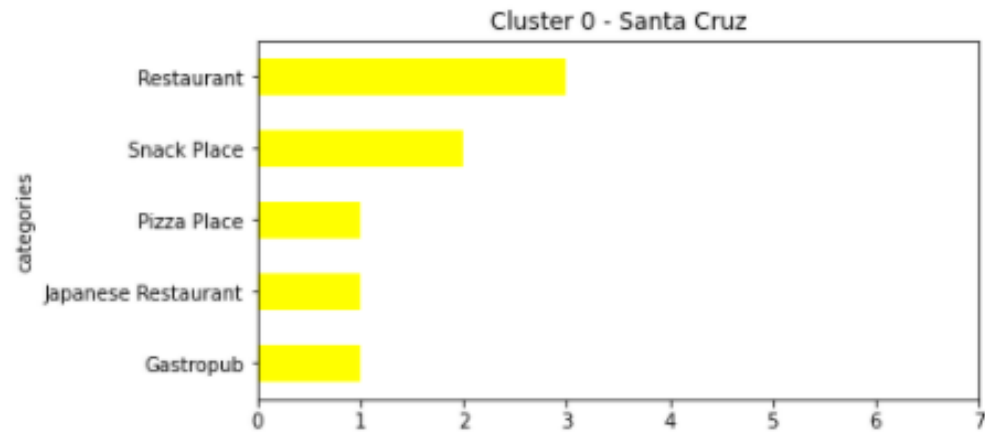
RESULT



RESULT



RESULT



DISCUSSION

Bangu

The **Cluster 1** has a wide variety of restaurants, but analyzing the data on the map, I verified that this all locations in cluster 02 are located inside Shopping Bangu, which is next to the train station. That's why it has more places.

Campo Grande

In the **Cluster 0** we can see that this cluster has more restaurants when compared to the others and this cluster is the closest to the train station.

Santa Cruz

In the **Cluster 1** we can see that this cluster has more sneack place when compared to the others clusters.

CONCLUSION

When comparing the three locations, we can see that the profile of restaurants is very similar to the regulars of Iacais. Campo Grande for having more business buildings, has a larger number of restaurants in the region. Bangu and Santa Cruz, on the other hand, have a slightly different profile than Campo Grande, but are more similar to each other.

From this first analysis, it can be concluded that the public in Campo Grande has a greater variety of places with healthier foods and in Bangu and Santa Cruz have a greater number of snack bars.

The initial proposal of the project was to analyze these regions and verify what types of restaurants we have around the stations and from there, for another future project, correlate the analysis with the quality of food of the people who use them and transit around of the locations.