Market Research
Bakery

Figure 1 - Bakery

Introduction:

The bakery is a type of commercial enterprise in which its target audience is mainly formed by people who live in the surroundings of the place, given the nature of the business. Thus, in order to plan the installation of a new bakery, it is essential to analyze places where there are deficiencies in serving the public in your region.

Methodology:

We will be using the data provided by the Foursquare API along with the population density and IDH data that was extracted from Wikipedia Page, whose source of this data is the Rio de Janeiro city hall website.

Neighborhoods analyzed:

The Neighborhoods of this study model will be those of the South Zone of the city of Rio de Janeiro. Neighborhoods included:

Gávea	Laranjeiras
Leblon	Jardim Botânico
Ipanema	Copacabana
Lagoa	Leme
Flamengo	Botafogo
Catete	Urca
Humaitá	

Data Source and Analysis

- Foursquare Data:

We used the API provided by the Foursquare to get information about the bakeries spreaded of the zone of interest. At this API, we have two option of querys: by search of a word and exploring, by looking everything within the radius.

Thus, to improve the amount of data acquired, I used two words in the query: "Padaria" and "Panificação", along with the exploritation of data, by limiting the radius to 6000m.

Them, we put everything together into a dataframe and we started to work on this data.

- Wikipedia Data:

By using web scraping, we were able to get the data of the population density (calculated through the population info and the area of the region) and the IDH information. This data has as source the Rio de Janeiro's city hall.
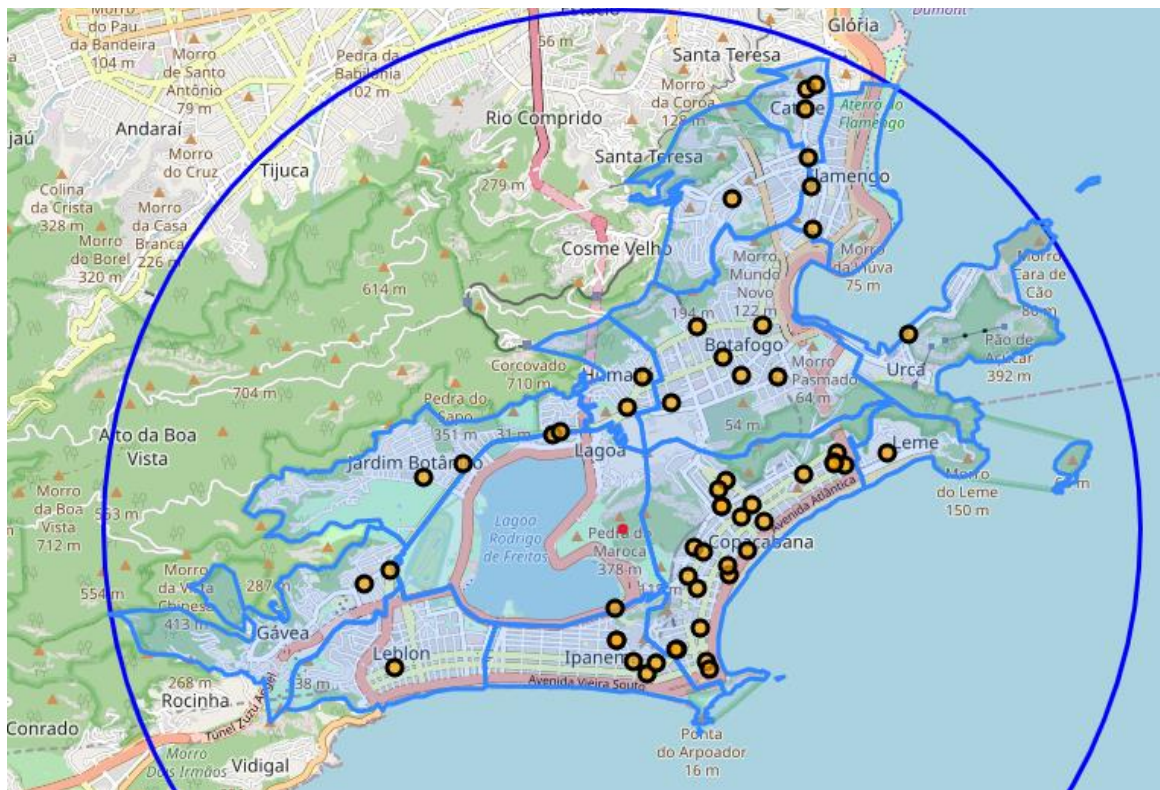
Exploring the Data:

Figure 2 - South Zone Bakeries

After plotting the points into the area of analyzes, we started to work on it. We used the information of IDH to get people's income information, along with the population density, that was used to estimate how crowded a given zone is.

After applying the population density into the map, we can see how concentrated the bakeries are, taking into account the local population. This resulted in the map below:
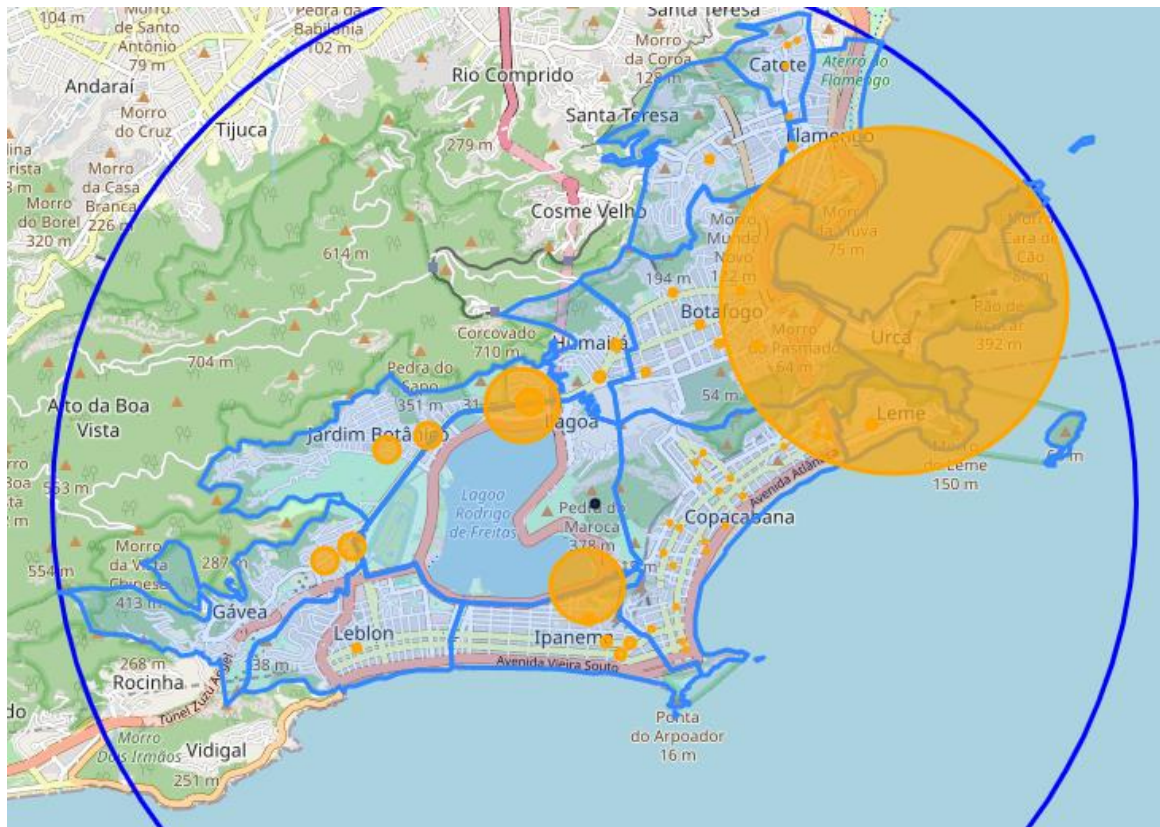
Figure 3 - South Zone Bakeries weighted by population density

Cluster Analysis:

By analyzing the Elbow method, along with the silhouette, we can come to the conclusion that k=3 is the optimal value.
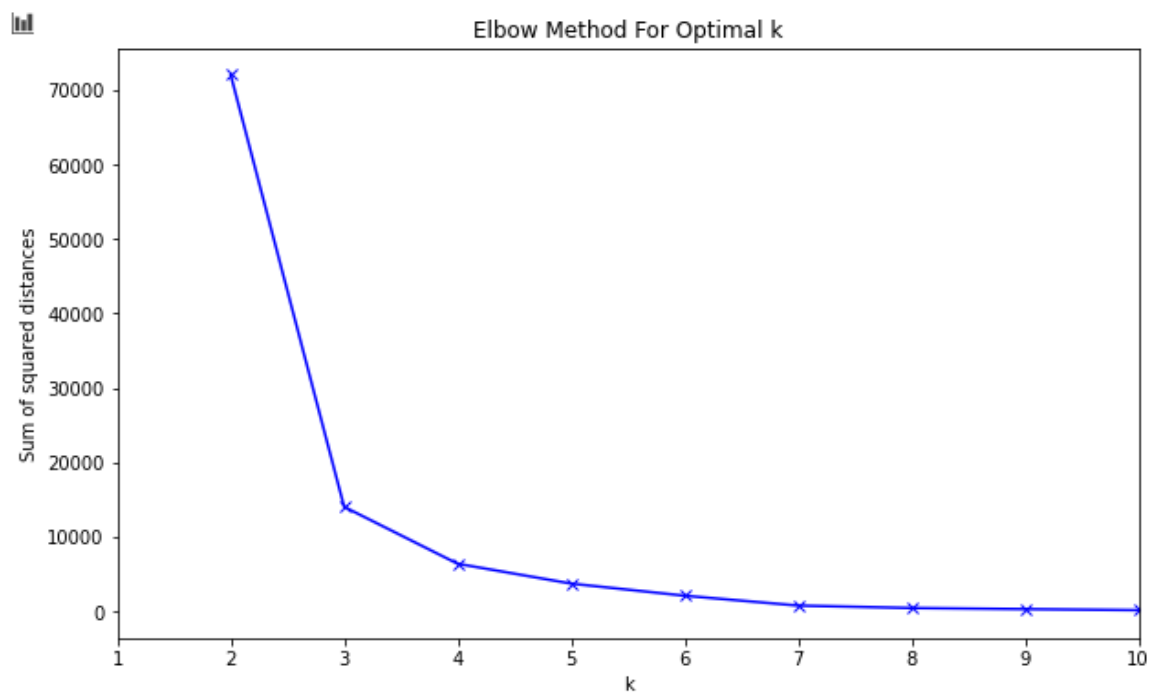

Figure 4 - Elbow method
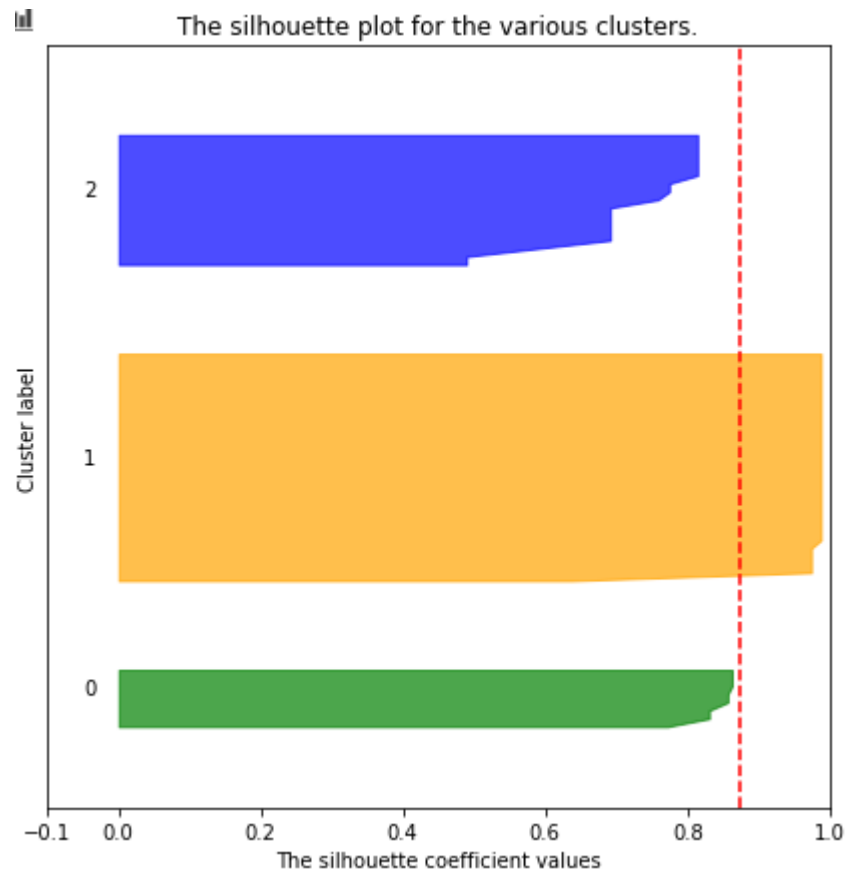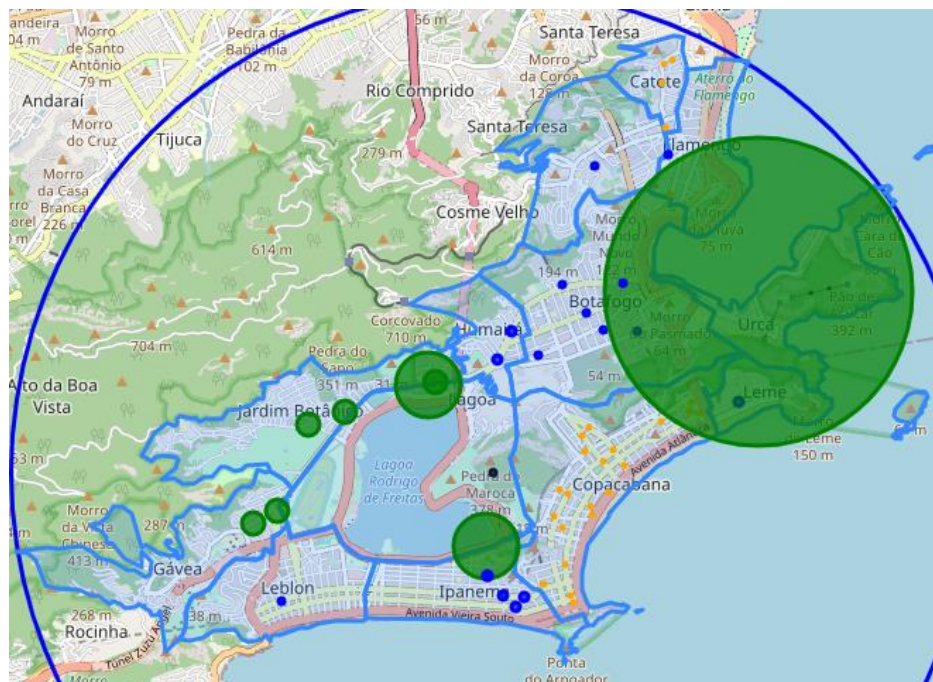
Figure 5 – Silhouette coefficient values

Figure 6 - Map after applying k-means

Conclusion

By analyzing the information provided by the