EXPLORING NEIGHBORHOODS IN CURITIBA (BRAZIL)

Coursera IBM Data Science Capstone Project

José H. K. Larcher

Background

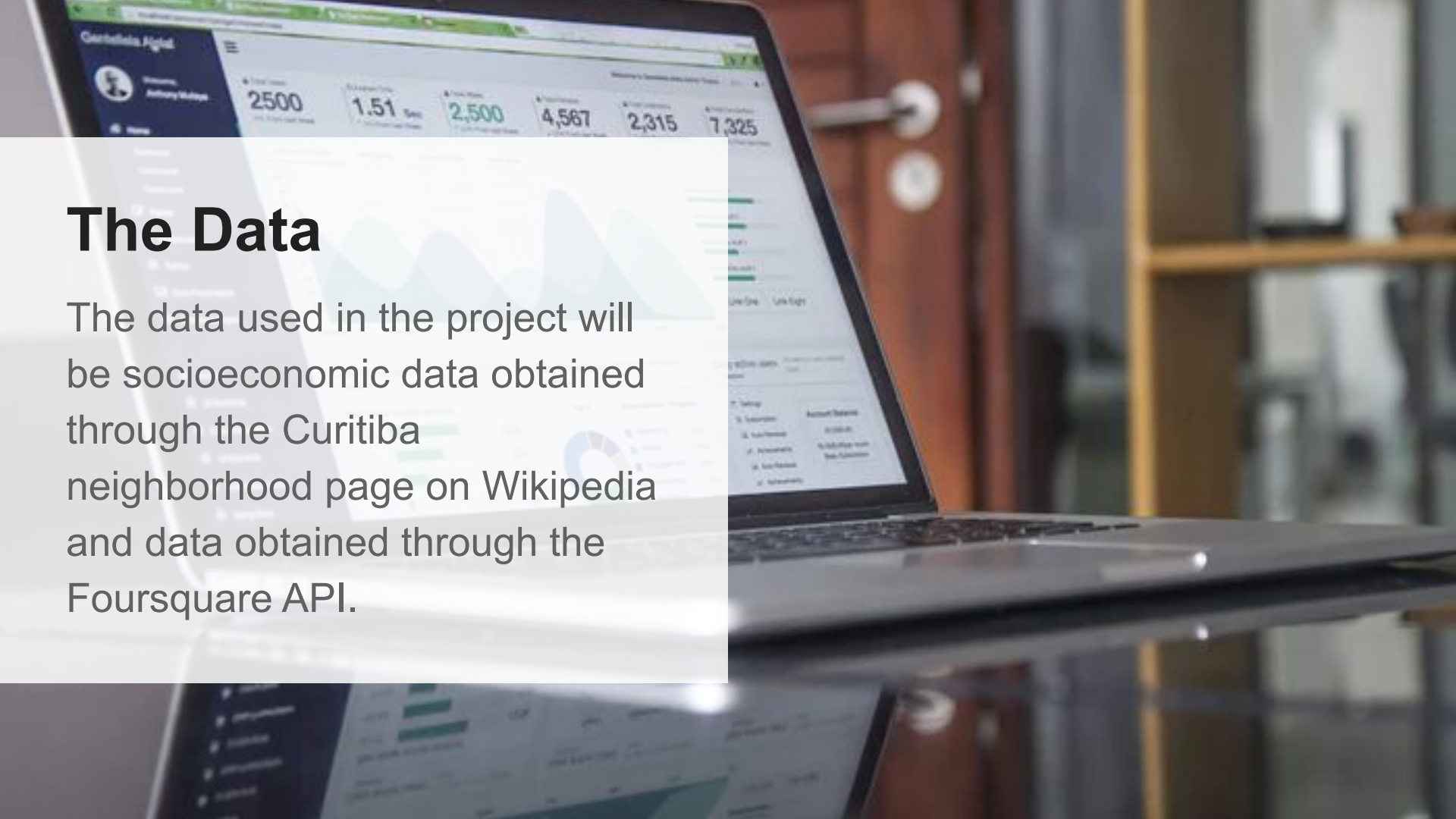
Curitiba is a Brazilian city, capital of the state of Paraná. The city of Curitiba is divided into a total of 75 neighborhoods, grouped into ten administrative regions.

Problem

This project aims to serve people looking for a place to open a business by answering where that business should be opened.

Interest

Entrepreneurs who are looking for places to open their businesses and wondering in which neighborhoods their business will be most appropriate and prosperous

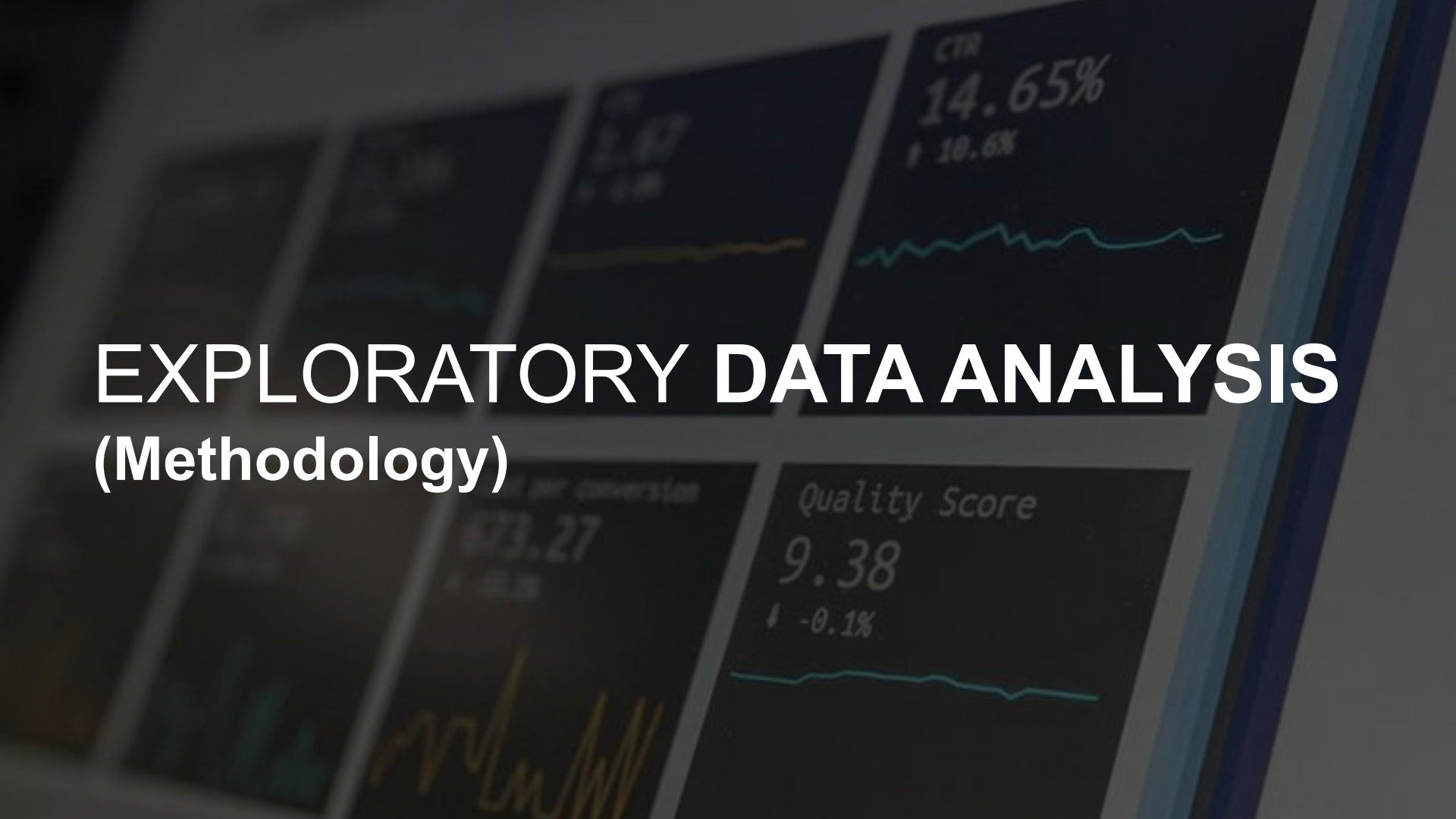


The Data

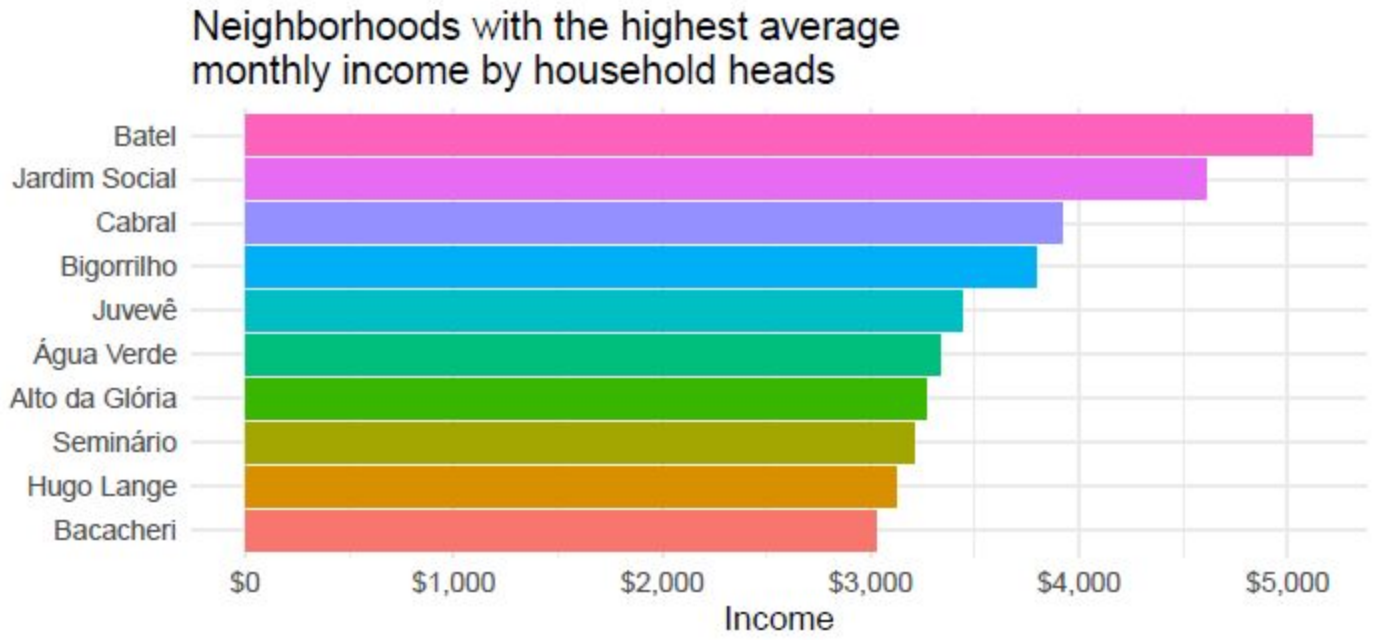
The data used in the project will be socioeconomic data obtained through the Curitiba neighborhood page on Wikipedia and data obtained through the Foursquare API.

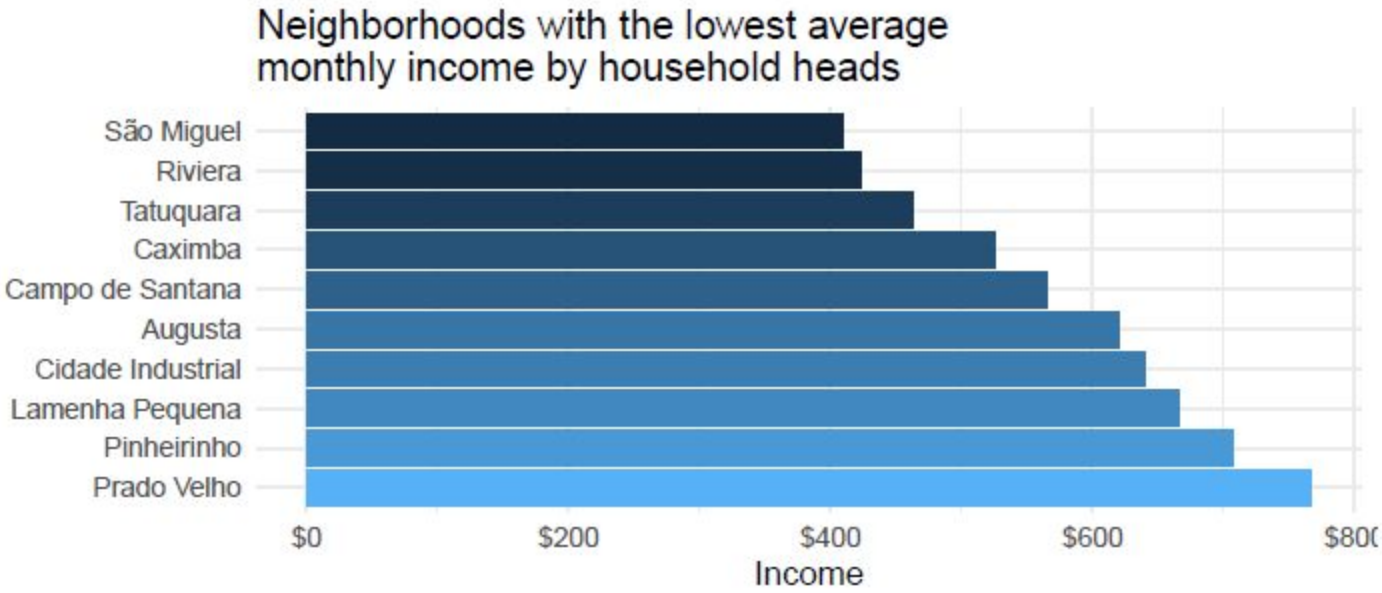
Data cleaning

- Neighborhood socioeconomic data was scraped from the Wikipedia page using the pandas HTML page reading function.
- The required columns were found and regular expressions were used.
- Some fields were normalized using z-score.
- The neighborhood coordinates were obtained using the geopy library.
- For the venues and business data of Curitiba, the Foursquare API was used.

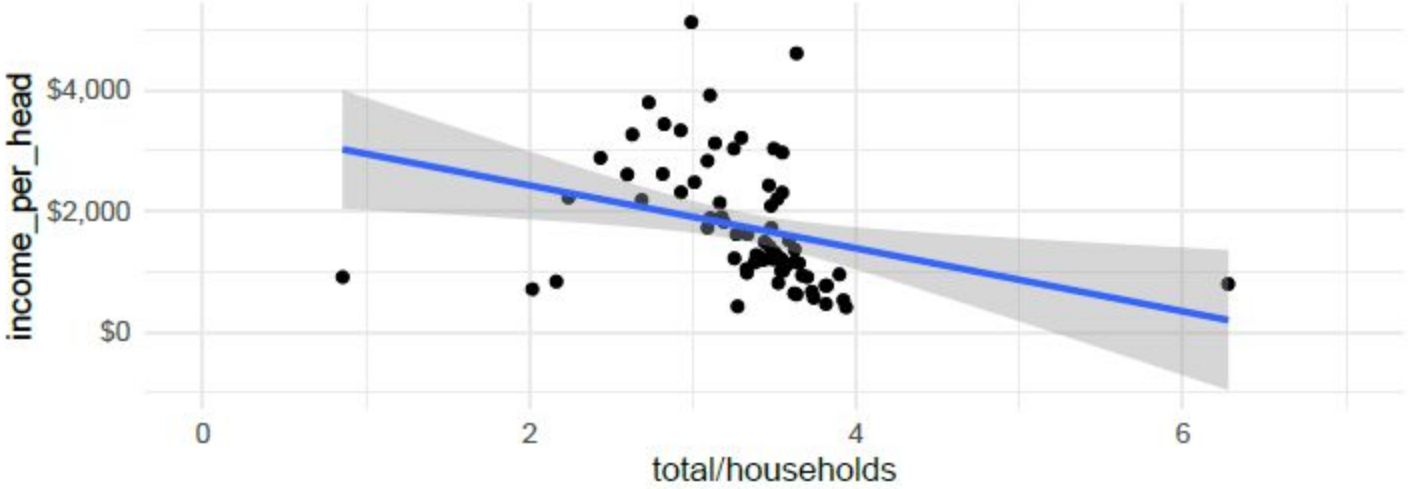


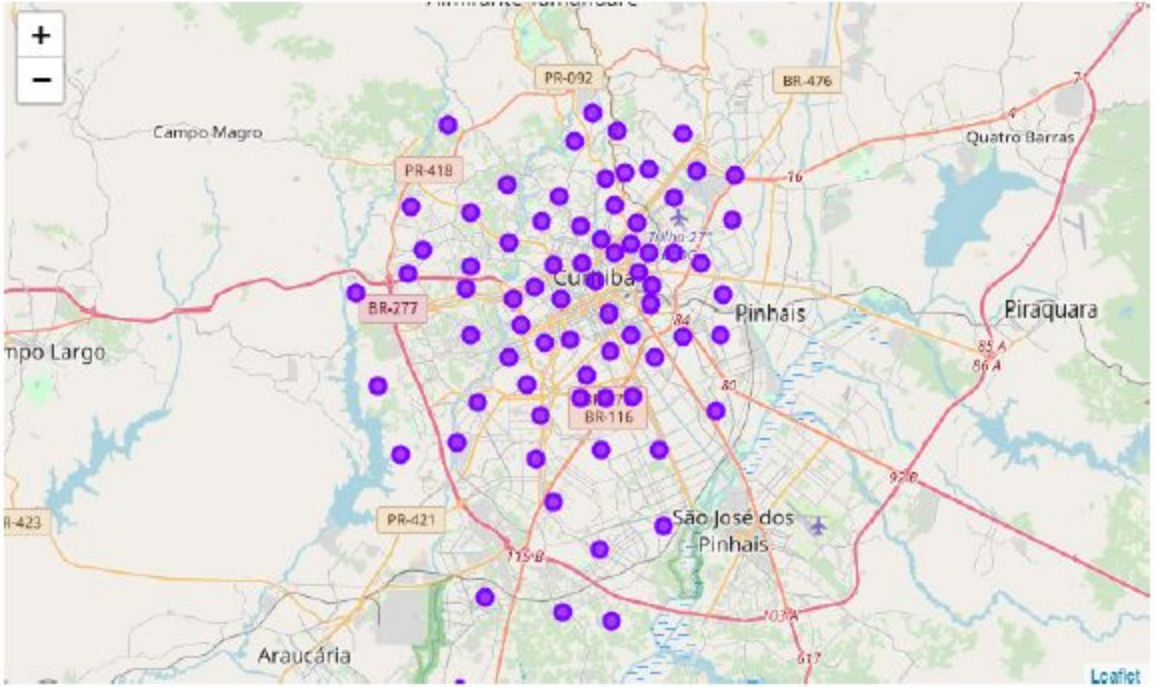
EXPLORATORY **DATA ANALYSIS**
(Methodology)


Neighborhoods with the highest average monthly income by household heads
Batel
Jardim Social
Cabral
Bigorrilho
Juvevê
Água Verde
Alto da Glória
Seminário
Hugo Lange
Bacacheri
$0
$1,000
$2,000
$3,000
$4,000
$5,000
Income


Neighborhoods with the lowest average monthly income by household heads
São Miguel
Riviera
Tatuquara
Caximba
Campo de Santana
Augusta
Cidade Industrial
Lamenha Pequena
Pinheirinho
Prado Velho
$0
$200
$400
$600
$800
Income

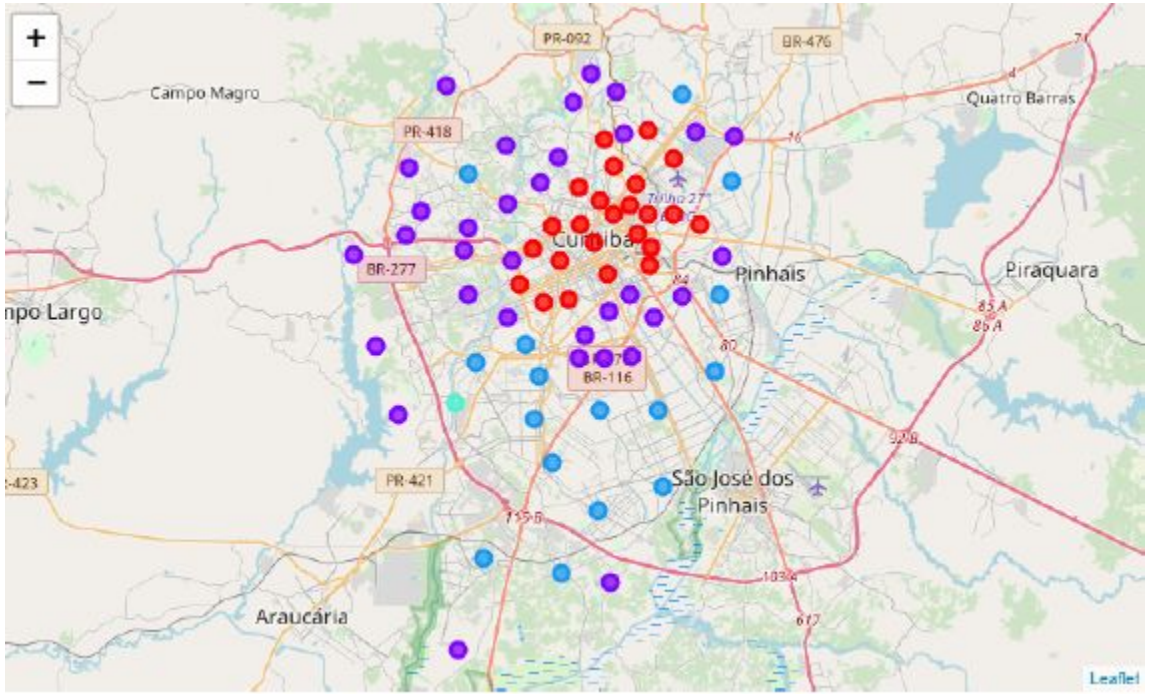
Is the average number of people in the residence indicative of the average monthly income per household head?



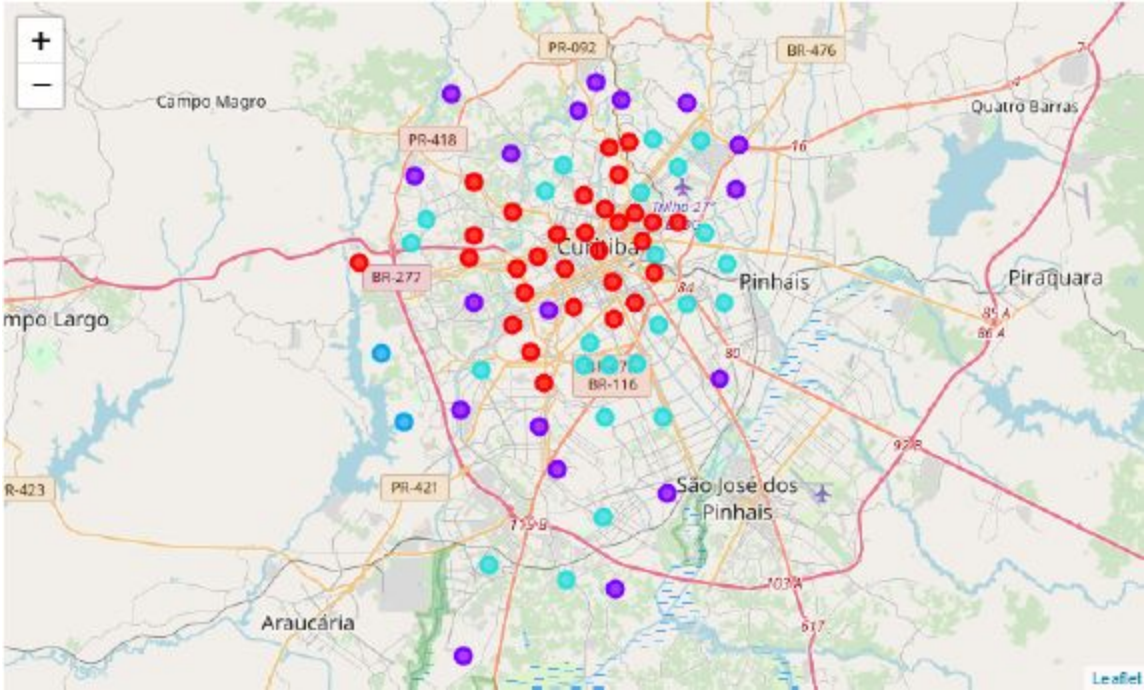

+
−
PR-092
BR-476
Campo Magro
Quatro Barras
PR-418
Curitiba
BR-277
Pinhais
Piraquara
npo Largo
BR-116
PR-421
São José dos
Pinhais
Araucária
Leaflet


MODEL
(Results)

Clustering based on socioeconomic data

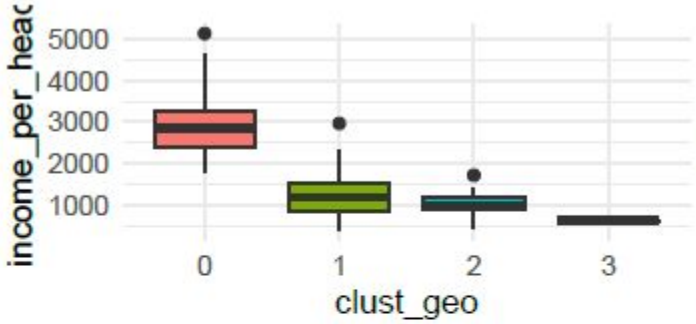


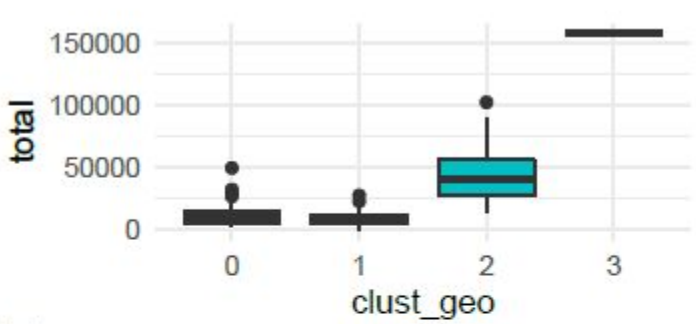
Clustering based on socioeconomic data

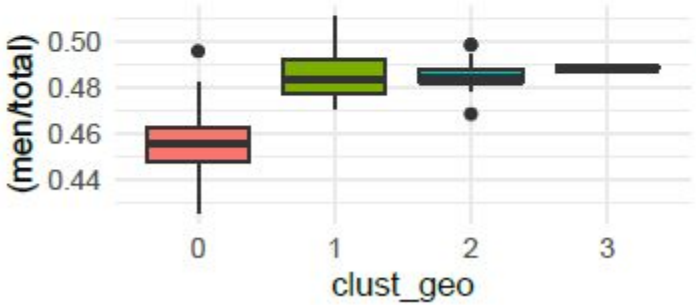


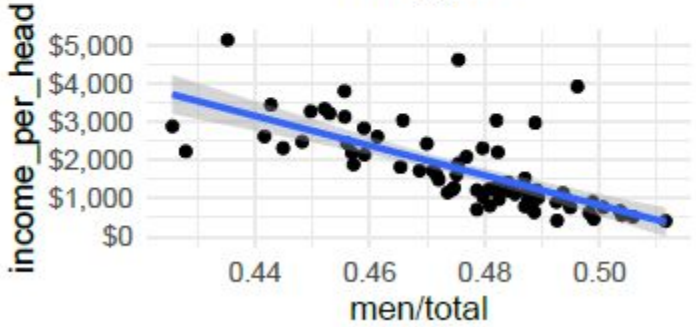

DISCUSSION

How varibles influenced clusters

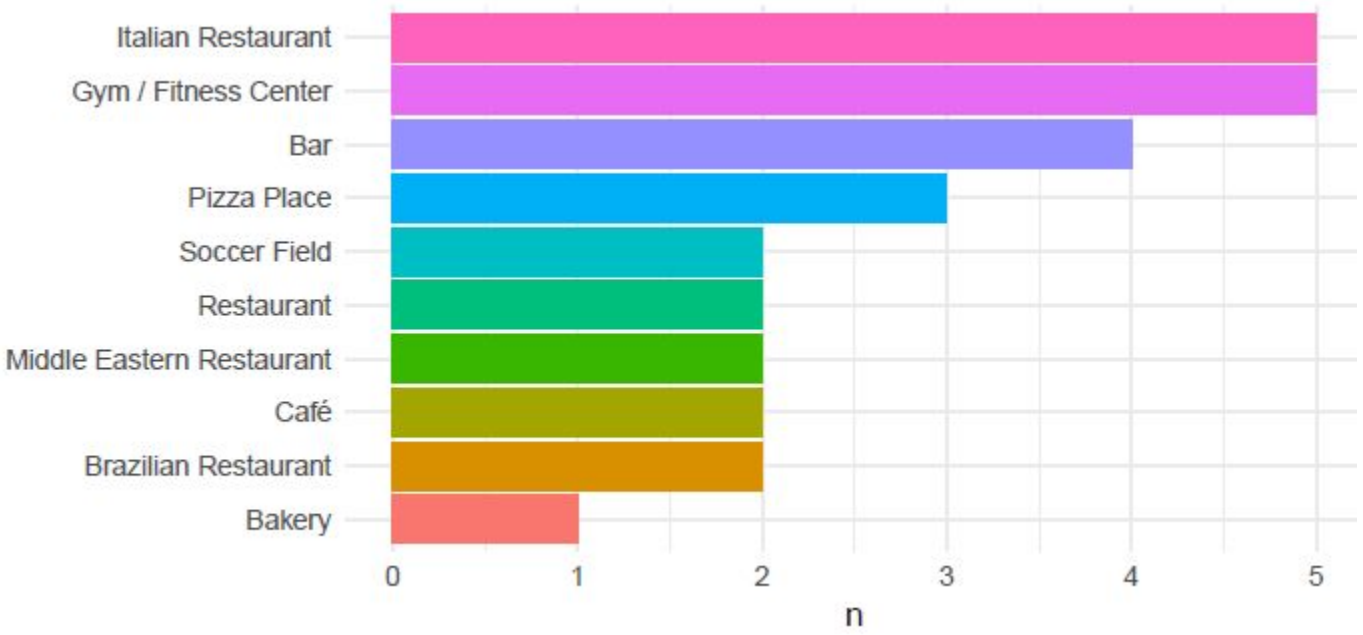








Top venues in central cluster

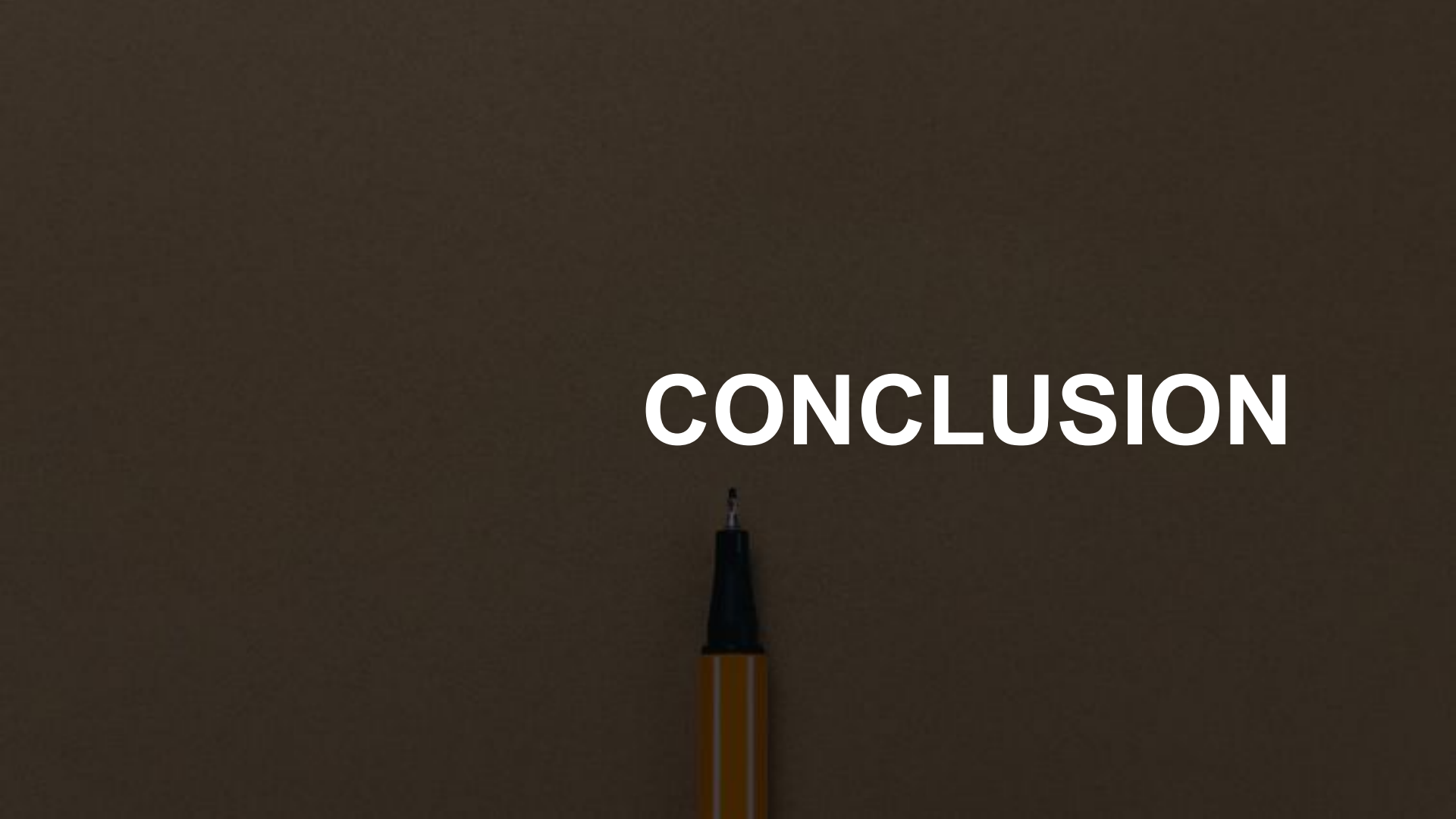


Discussion

- The central area might be good for Italian restaurants and gyms

- On the other side, these kinds of venues might be saturated in the region.

CONCLUSION



Taking the intersection of the central clusters of both model as good neighborhoods to open a business, these neighborhoods would be:

- Batel
- Jardim Social
- Cabral
- Bigorrilho
- Juvevê
- Água Verde
- Alto da Glória
- Seminário
- Hugo Lange
- Bacacheri

References

- https://pt.wikipedia.org/wiki/Lista_de_bairros_de_Curitiba
- https://developer.foursquare.com/docs/api/
- https://pypi.org/project/geopy/