**ANALYSIS OF RENT ACCOMODATION PRICES IN PORTO**

Anastasia Kondratiuk

November 6, 2020

**I Introduction**

**I. I Background**

Overlooking the Douro River, Porto is one of the most ancient European cities, stemming from the northern bank of the river during the Middle Ages. One of the most significant aspects of Porto and its historical center is its surrounding landscape, the steep river banks providing a dramatic setting for old town. Both the streets next to the river and the surrounding countryside have been classified as a World Heritage sites by UNESCO. Discovering Porto is full of surprises: beautifully cobbled streets and typical shops retain the feel of a bygone era whilst contemporary architecture jumps out in surprising places [1].

To organize outstanding trip every traveler needs an advice in booking of most comfortable and suitable apartments. It requires a lot of efforts to find ideal place in a big unfamiliar city like Porto.

Let me introduce an outstanding solution of the problem: a map with average price per apartments visualization and venues divided by clusters according to characteristics of each neighborhood.

**I. II Data description**

Airbnb statistics for 2013 – 2020 years [2]. The dataset contains price of 11 366 accommodation options provided by 2040 hosts with longitudes and latitudes of each place.

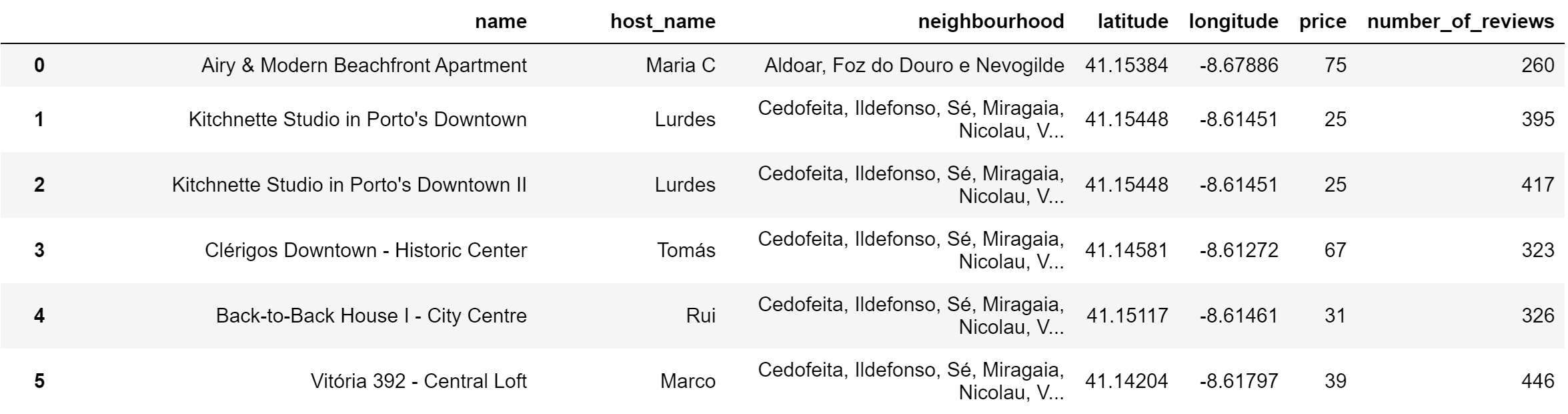
Forsquare API gives access to information about venues located within 500 meters around each position [3].

Porto geojson file [4] to create choropleth map with prices.

**II Methodology**

I used GitHub repository as a database for my project.

I cleared initial data in a way to keep only necessary columns with information about location, price and name of host only for entire home / apartments. I filtered my data by date of last review to keep only actual information of October 2020. I chose only variants with more than 130 reviews and 320 days availability.

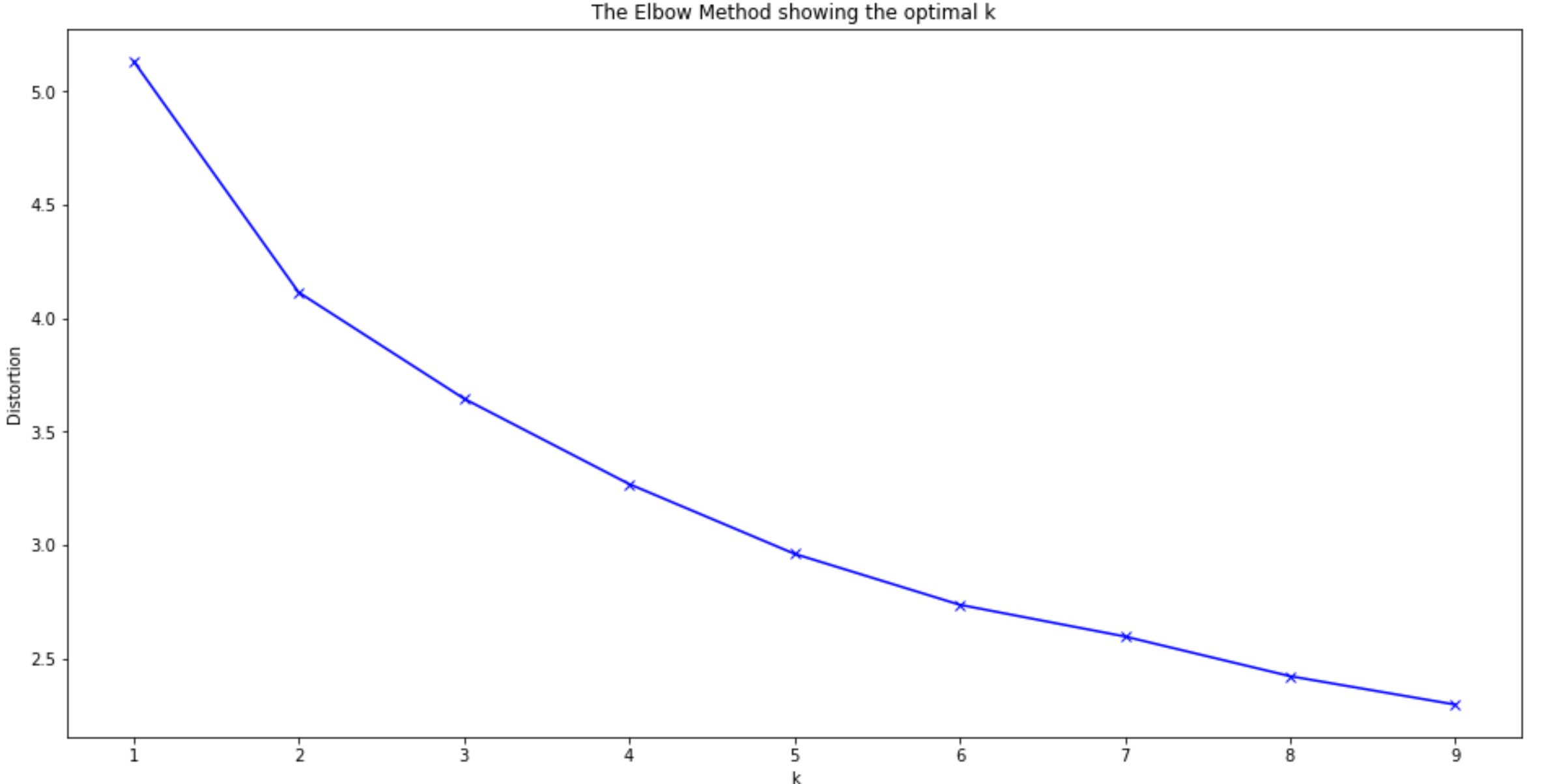


I grouped the data by host\_name and called Foursquare API to explore the hosts and segment them. 94 unique categories were returned by Foursquare, then I created a table which shows list of top 10 venue category for each borough in below table.



We have some common venue categories in hosts. In this reason I used unsupervised learning K-means algorithm to cluster the hosts. K-Means algorithm is one of the most common cluster method of unsupervised learning.

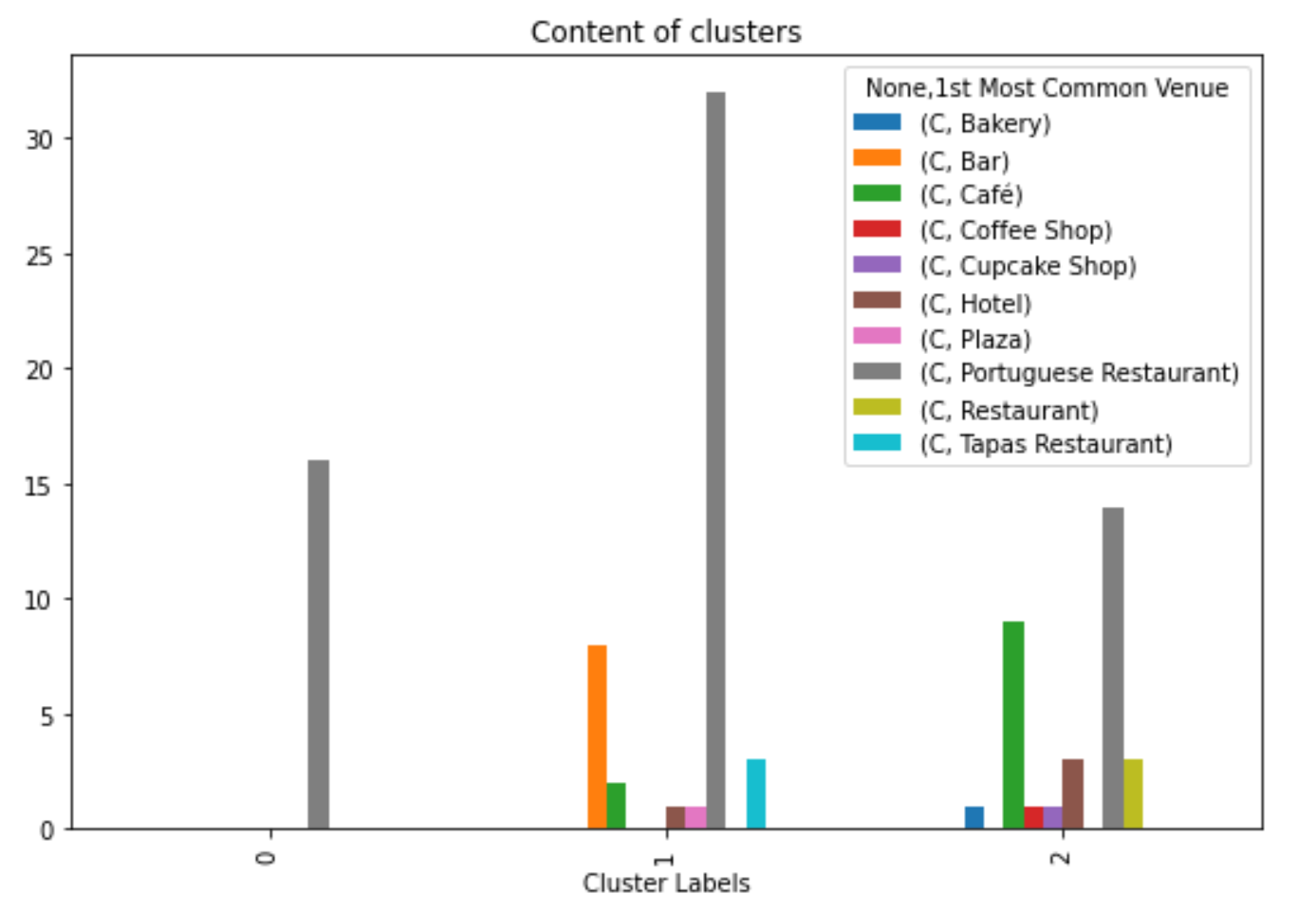
First, I ran K-Means to cluster the hosts into 3 clusters because when I analyze the K-Means with elbow method it ensured me the 3 degree for optimum k of the K-Means.



Here is my merged table with cluster labels for each borough.



Each cluster contains several types of venue categories.



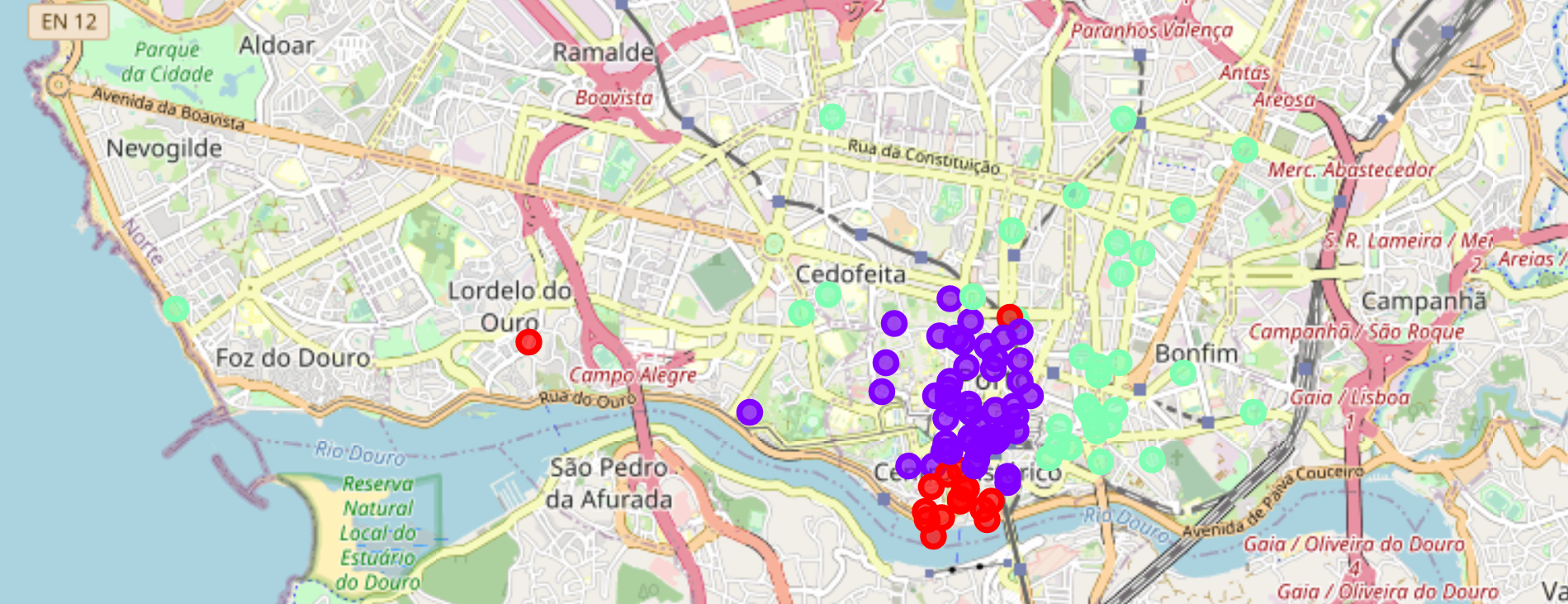
0 Cluster: Portuguese Restaurant

1 Cluster: Portuguese Restaurant, Bar, Tapas Restaurant

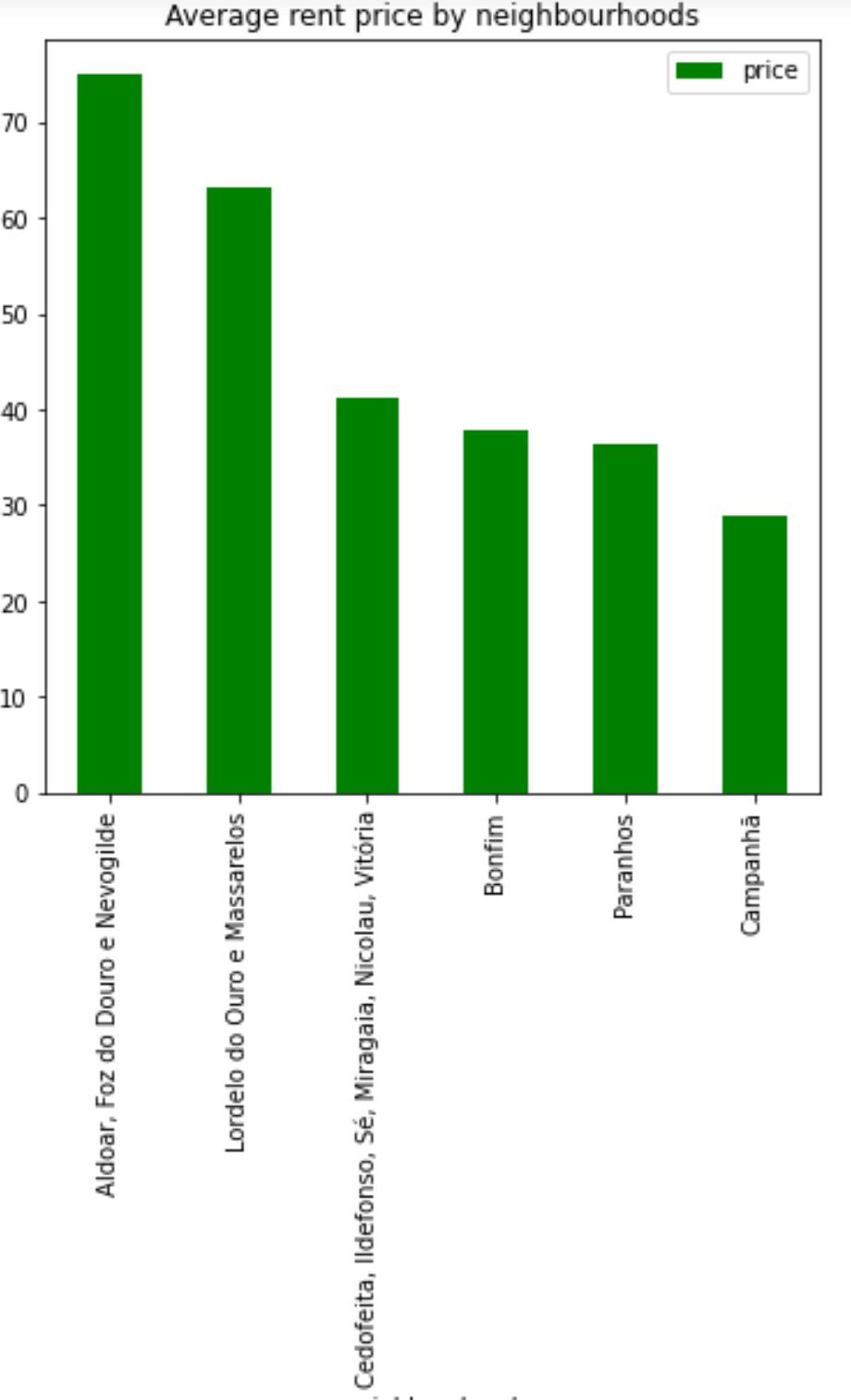
2 Cluster: Portuguese Restaurant, Café, Hotel

**III Results**

Finally clusters indicate on a map.



And it is possible to get familiar with rent prices by neighborhoods.



Thus with help of this instruments user can concludes his own value for money option.

**IV Conclusion**

Data science tools give access to individual analysis approach for users. The model is applicable for any city and is really useful for travelers because provides an opportunity to predict options of certain accommodation in unfamiliar city.

**V References**

1. <http://portoswingjam.com/about-porto/>
2. <http://insideairbnb.com/get-the-data.html>
3. <https://foursquare.com/>
4. <https://gist.github.com/magamig/4448dccd63cabe2112771d71cbd1faef#file-porto-geojson>
5. https://peteris.rocks/blog/openstreetmap-administrative-boundaries-in-geojson/