Predicting the best place to start your new business

Augusto Barbosa

June 15, 2021

1. Introduction

    a. **Business problem**
Our objective is to have one system able to give directions to our clients about where is the best city in Portugal to start a new business based on its category/classification.

Besides many other variables that might induct the person to decide where to start this new business, we will try to help him to decide where to start by creating one index which will consist of the ratio between the population of the city versus the number of restaurants that this city already has.

This index aims to give to the requester which city may be better once that will have an index with more demand and less offering restaurants.

Formula:
Potential Index = Population of each city divided by the number of restaurants

b. **Target**
We can offer this service to anyone recently moved to Portugal and this person has the intention of starting a new bussiness. This person will be able to know which will be the best city to fit his business based on the category and the population of the cities.

2. **Data**

For this first version, we decided to use two main databases:
a. One spreadsheet which has been retrieved from https://simplemaps.com/data/pt-cities that consists with a list of all cities from Portugal, including the geospatial coordinates and the population.
b. We will also use the Foursquare API in order to get how many business venues we have in each city.

Also, for this simulation we've decided to use "Restaurant" as the hypotesis' validation category based on our own experience that is one category that gives good financial return and have good oportunities at Portugal once is one of the most turistic destiny on Europe.

3. **Methodology**

The methodology used is mainly Python extracting informations from the Foursquare API using the SOAP connection, transforming JSON files into a dataframe which turns possible to connect and relate this data with the dataframe that contains the city's information.

During our exploratory analysis, we discovered the following points:
- Once that Foursquare is a community maintened platform, we can

have some missing bussiness that may not be regitered inside the platform and this could direct our client to a different place than the most fertile city. Also, in some smaller cities, we discovered that we have no venues and restaurants registered.

- We've also discovered that the restaurant's categories are splitted by the different kind of food that they may offer. So, to overcome this problem, we've decided to get all categories that includes the string "Restaurants" registered.

After collecting the two data sets, we've applied clustering algorithms in order to better understand the categories, the top 10 categories for each city and how they are distributed on the country.

city	admin_name	capital	population	population_proper	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
Évora	Évora	admin	55620.0	55620.0	2	Portuguese Restaurant	Café	Coffee Shop	Accessories Store	Multiplex	Museum
Ponte de Lima	Viana do Castelo	minor	43498.0	43498.0	2	Portuguese Restaurant	Restaurant	Hotel	Coffee Shop	Pizza Place	Café
Chaves	Vila Real	minor	41243.0	41243.0	2	Portuguese Restaurant	Hotel	Bar	Brewery	Hot Spring	Restaurant
Bragança	Bragança	admin	35341.0	35341.0	2	Museum	Portuguese Restaurant	Bar	Castle	Bed & Breakfast	Taverna
Tavira	Faro	minor	26167.0	26167.0	2	Portuguese Restaurant	Italian Restaurant	Mediterranean Restaurant	Gastropub	Bar	Café
Almeirim	Santarém	minor	23376.0	23376.0	2	Portuguese Restaurant	Accessories Store	Other Nightlife	Mountain	Movie Theater	Multiplex
Elvas	Portalegre	minor	20706.0	20706.0	2	Portuguese Restaurant	Plaza	BBQ Joint	Dessert Shop	Castle	Restaurant
Macedo de Cavaleiros	Bragança	minor	15776.0	15776.0	2	Convenience Store	Portuguese Restaurant	History Museum	Big Box Store	Accessories Store	Other Nightlife
Moura	Beja	minor	15167.0	15167.0	2	Hotel	Portuguese Restaurant	Supermarket	Castle	Bakery	Other Nightlife
Falagueira	Lisboa	NaN	14530.0	14530.0	2	Portuguese Restaurant	Restaurant	Gym	Accessories Store	Other Nightlife	Mountain

Top 10 categories generated for each city
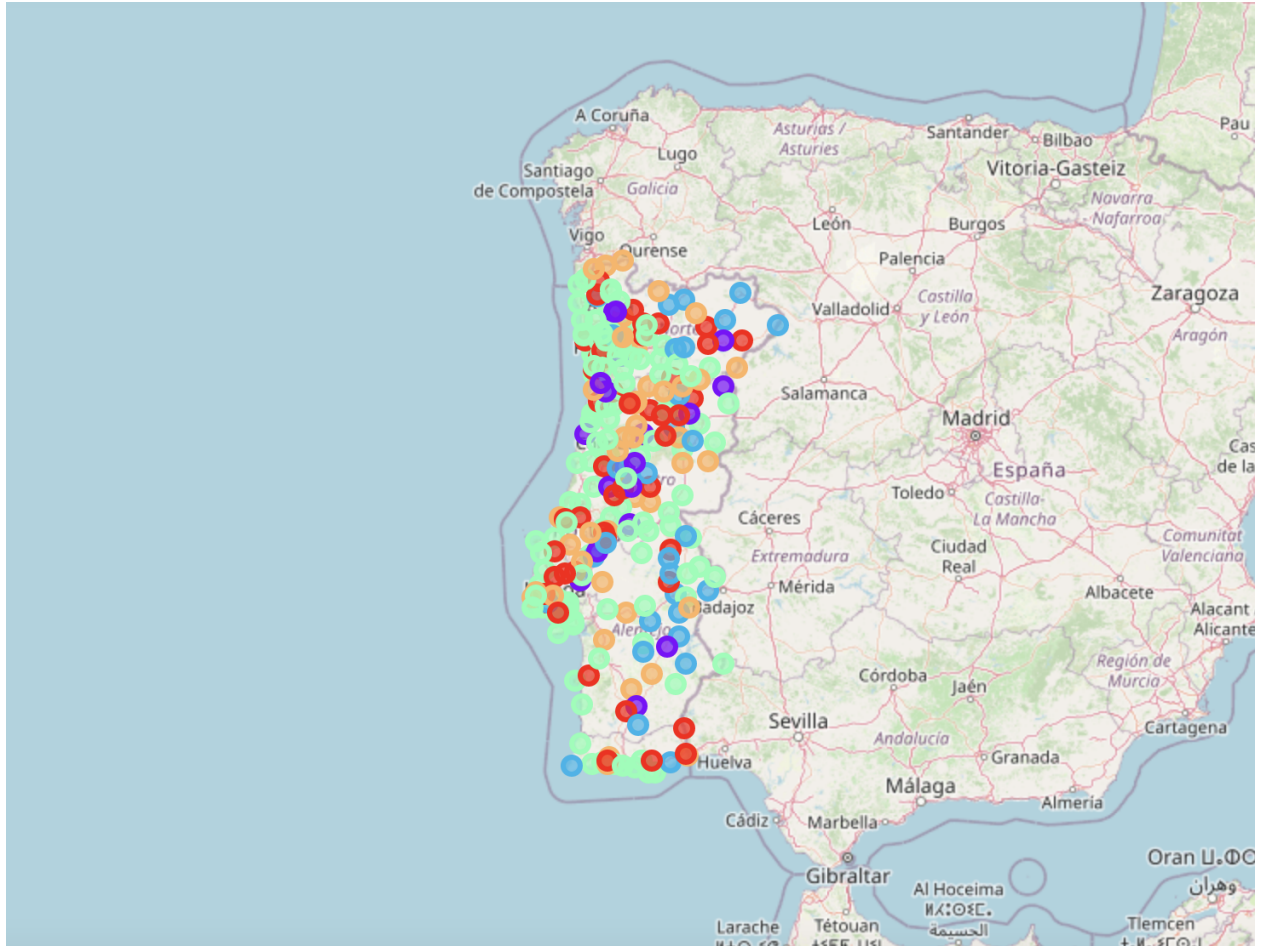
*Clustering cities and its venues categories*

4. Results

The results of our analysis has been able to show a great direction for our stakeholder about where to start his new business.
We mainly verified that the city of Lisbon, even being one of the most populous city of Portugal have good market opportunity to start the new restaurant.
This city is more digital engaged which can open different opportunities than only the physical restaurants. They can also start new fronts like Uber Eats application and take away food service.
Another good point to Lisbon is that the city is surrounded by smaller cities (e.g. Amadora) and that are not more than 20 minutes from

distance which also can have consumers for our restaurant service.

We cannot forget to mention Braga as one of the most probably city to start the business once the city has won recently prizes of best places to travel in Europe. This city is not big as Lisbon but is located on the North of the country, which is a different area from Lisbon, which is based on South, and can also have a fertile space to start new business.
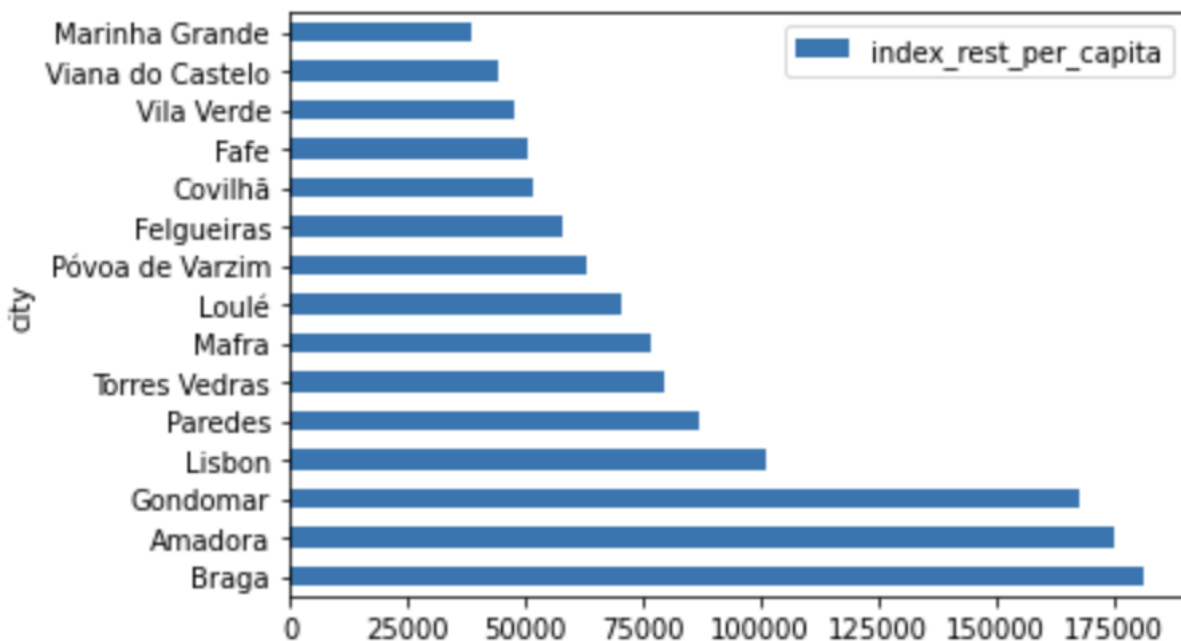
5. **Discussion**

Regarding the observations mentioned above, we have to take care with the databases that we will choose to get our Potential Index once that this information afflects totally the assertivity of our service.

We also have in mind that according to the desired category choose by the stakeholder we may need to use other datasource for the

venues and business once that Foursquare may not have all categories filled.

6. **Conclusion**

The hyphotesys that we initialy created has been validated and for us this is the main achievement. We have been able to predict two main areas where the stakeholder could start his new business. Based on

We also have been able to verify that Porto, which is one of the most known cities of Portugal is not a fertile place as we were imagining in the beggining. This city don't have a big population once is a small city in terms of geospace and have many restaurants already opened

Porto	41.1495	-8.6108	Portugal	PT	Porto	admin	237591.0	237591.0	36.0	6599.750000

7. **Jupyter notebook used on this analysis**
   Link to access the notebook

8. **Next steps**

   a. Use another API to get restaurant or any other business database
   b. we can also implement this service for other geolocations once that we only need to the cities' population and geocoordenates.