Capstone Project - The Battle of Neighborhoods

December 12, 2020

Introduction

Business problem

San Vendemiano is a town in the North of Italy. In recent years the city and the metropolitan area around it increased significantly in population, number venues and presence of tourists. Considering that this trend is going to continue in the future, a entrepreneur is contemplating to build a new restaurant in the area. We are going to help him, using a statistical analysis, by selecting the best neighborhood for his new restaurant.

Data

We will need a map of San Vendemiano and its surrounding. In order to build an index of attractiveness for each neighbourhood we can use a list of venues provided by Foursquare. In particular a list of already present restaurant.

Data Sources

- Foursquare API: list of venues and restaurants, with their coordinates (latitude and longitude) [1];
- Openstreetmap: Map of the area [2]

Methodology

Configure the notebook and Data Collection

We need the following packages in python for the job.

```
import pandas as pd
import requests   # to get urls
import numpy as np
import os
import matplotlib.cm as cm
import matplotlib.colors as colors
import matplotlib as mpl
import matplotlib.pyplot as plt
!pip install folium
import folium
!pip install osmium
import osmium as osm
from geopy.geocoders import Nominatim
```

In particular *folium* is used to plot the map, *osmium* to read and manipulate Openstreetmap files (.osm), *geopy.geocoders* is used to get the coordinates of San Vendemiano: $45.8907463°N$, $12.3338437°E$.

As there is no clear definition of neighbourhoods in the area we need to come up with our own: we will divide the area into a $10x10$ grid where each cell has a size of $\sim$ 1Km.

Figure 1: Map of San Vendemiano and surrounding.

Data Preprocessing: Creating Features

The criteria for choosing the best neighbourhood are the following:

1. good connection with the city and the population, i.e. a residential area is better than a rural area;
2. high number of venues in the neighbourhood, a signal that it can be more attractive;
3. low number of restaurants, so less competition.

Regarding the first point, we can calculate an index of inhabitatility from the number of highways in each neighbourhood; these are provided by the Openstreetmap project [2].

Figure 2: A choropleth map indicating areas with higher number of highways (darker color).

The number of venues and restaurants can be downloaded using the Foursquare API [1]. Off all 226 found venues, 47 of them are some kind of restaurants.

	Venue	Venue Category	Latitude	Longitude	neighbour
0	3Store Conegliano	Mobile Phone Shop	45.8782	12.3143	33
1	A.S.D. Palabiliardo Conegliano	Pool Hall	45.9168	12.3509	78
2	A27 - Conegliano	Toll Booth	45.8788	12.3420	63
3	Aidolà	Café	45.8829	12.2915	14
4	Al Cicchettin	Diner	45.8810	12.3805	93
...	...	...	...	...	...
221	fair play	Soccer Field	45.8809	12.2771	03
222	gaia park	Playground	45.8734	12.3587	72
223	o2 Laboratorio del Respiro	Spa	45.8917	12.3389	55
224	piscina Ranazzurra	Pool	45.8985	12.2985	16
225	pizza & pizza	Pizza Place	45.8714	12.3408	62

Table 1: All the venues

Final Model and Visualization

Our final feature table looks like the following.

	neighbour	highways count	venues count	restaurants count
0	00	2	1	0
1	01	0	0	0
2	02	6	4	2
3	03	34	3	0
4	04	3	1	0
...	...	...	...	...
95	95	0	0	0
96	96	1	2	1
97	97	2	0	0
98	98	2	4	0
99	99	7	4	1

Table 2: Feature table, all the useful data for each neighbour.

We can measure an index of attractiveness using the following formula:

$$attractiveness = highways \times venues,$$

and order the neighbours from the most attractive to the least attractive.

	neighbour	highways count	venues count	restaurants count	attractiveness
14	14	71	23	6	1633.0
24	24	90	10	5	900.0
33	**33**	**30**	**15**	**0**	**450.0**
13	13	48	9	1	432.0
23	23	34	9	2	306.0

Table 3: Most attractive neighbours.

As we can see neighbour = "33" is the third most attractive one but has zero restaurants. In conclusion, this is probabaly the best zone to build a new restaurant.

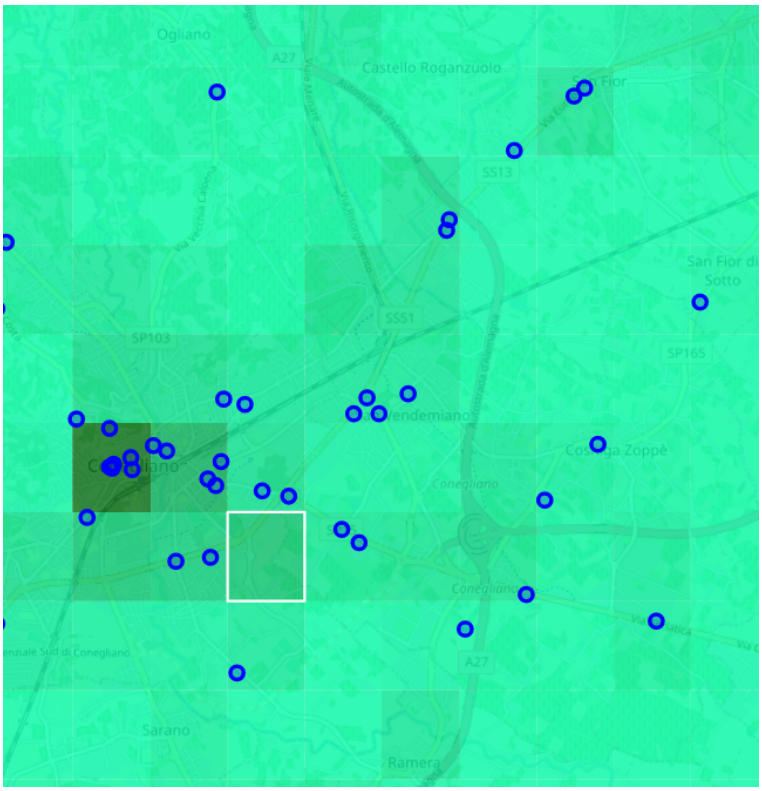

Figure 3: A choropleth map indicating the most attractive areas (darker color). Points are restaurants. White-delimited cell is the best neighbourhood.

Results and Discussion

We have found a method to classify the map into neighbourhoods of size $1km^2$. We have assigned to each neighbourhood a score, called "attractiveness", based on presence of highways and venues. By looking at table 3 or at figure 3, we can see that generally restaurants are in attractive neighbourhood, so this proves that our model works. We can see that the third most attractive one, though, it has zero restaurants.

Conclusion

We can conclude that we have found a model that correlates well between attractiveness of neighbourhoods and the presence of restaurants.

To answer our business problem, we can see that the best choice to build a new one would be in the neighbour highlighted in white in the last figure, since it doesn't already have a restaurant.

References

[1] https://developer.foursquare.com/developer

[2] https://download.geofabrik.de/europe/italy/nord-est.html

[3] My notebook: https://eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/768a875f-94fa-4d1a-9db4-48a2109159d1/view?access_token=a608a1e617b37f6bcb79002c40e3c3bfee7dc5d4442a609a985b