1)

For this project i will use a fictional case:

Gilardino is an italian investor who has been pretty sucesfull in his home based pizza restaurant chain, but its facing a pretty clear bottleneck of not having enough potencial costumers in other locations of the country, where he can locate a new restaurant. Actually, he has about 20 restaurant of pizza, starting the first one in 2000, but he has failed in making earnings from his last 3 restaurants located in Italy, cause there isnt enough potential costumers to target.

He is thinking in big, and considering seriously to internationalize his bussiness. Actually, he dreams with getting as near as possible to Pizza Hut or Domino's pizza in terms of number of locations -both of them surpass the 15,000- in a distant future. Right now, his only concern is about increasing the range of his business. In order to achieve the later, he has hired a small group of analysts to evaluate his possibilities of expansion. This group, conclude, via various analysis, that the best chance of making a big impact in the international scene, was to locate a couple of restaurantes in a first developed country, with enough movement of tourists and resident people, and say "rich enough", what can provide sufficient earnings to recover the invest in soil and infraestructure requiered. Hes also considering to implement a drastic change, which is, to launch a gourmet type of pizza restaurant in the place, but that is out of the scope of this particular project.

So, the team has conclude that there are 2 places, which are quite similar and satisfy all the requiered conditions of this case -and the aditional condition of beeing located near the sea, a preference of the owner- which are Toronto and New York. Gilardino is very self driven to have its location in New York rather than Toronto, he loves the country and "USA has a bigger name" he often says. Finally, the owner has decided to hired a data scientist to study via statistical methods if possible, and by other means, if its a real option to have its location in New York, or at least, near it.

The data scientist will try to prove which is the more attractive location for setting up a restaurant, including, if possible, the ammount of locations that will optimize his earnings, explaining clearly the reasons of such decision -taking account of every form of bias or error in any analysis he made-.

2)

The data that will be used to resolve this enquiry will be:

-Geospatial information needed to capture data about the distribution of similar restaurants in several locations. For this, i will use Foursquare API queries.

-Statistial information of different conties or boroughs, such as pib per capita -if possible-, number of turists, growth, and others. For this i will use sources avialable via internet, exploiting them via Web Scrapping.

-Some other information will became necesary as the study evolves. For example, parking conditions, accessability and visibility related to the target spot. For this, if the need arise, i will use simple google maps images, cause image processing capacities are out of the scope of this course.

-Of course the study can change the scope over the course exploring the information. For example, a new city may became interesting, or maye there arent realistic chances of setting a restaurant in any of this places, and then the explorationg change the target cities.

3)

The prime source of information to capture relevant data was the geospatial output from queries to the foursquare API. Due to the clarity of output shown by folium –while using dataframes to communicate the data from the API-, even though i was very driven to apply clustering method such as KNN to explore the density of each county or borough, i refuse to do so, cause i consider it unnecesary for this particular case (a Little bit ‘pompous’). The visual results were sufficient.

Due to the harsh restraints of the 50 maximun queries from the API, it was pretty annoying to move geographically studying different places, taking consideration of all the possible bias of such constraint. There so, its pretty clear that the study is not as professional as one could expect. But, is serious enough to take consideration of the results presented.

Theres not much statistic involved in this study. There are KPI such as median income and income per capita and population that were taken on consideration, but I take the study in another direction, using the techniques and methods tought in the specialization, such as Web Scrapping and Geospatial information via API and folium

4)

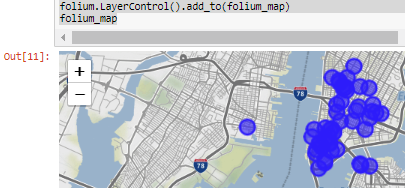
The results are pretty clear withouth the need of much statistical or AI paraphernalia. It has been found a pretty safe place to bet near New York, that its near of lots of hot spots (places which lots of pizza restaurants), near the sea, in a wealthy county and even more wealthy borough, untouched in quite some distance by pizza restaurants, which is the borough of Rumson located in the Monmouth county of New Jersey.

It remains to be seen if there is a restriction of no construction or something similar, that could explain the phenomena, but thats out of the scope of this Project.

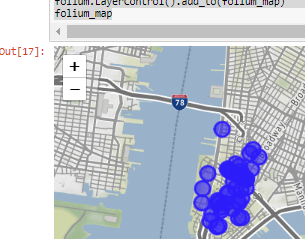
With more appropiate tools, the study would have been quite different, capturing the whole territory at ones with geospatial information of the whole country, or at least by counties. Then, clustering the whole territory identifying all of the restaurants, and other organizations such as schools, Banks and hospital would had been used to describe the whole phenomena and identifying different alternatives and lastly comparing them. But with such strong limitation of this particular API, its quite impossible.

Detailed Results

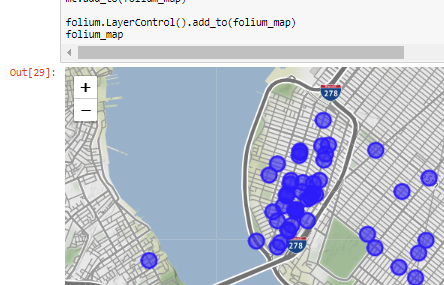
New York Pizza Restaurants (ratio 2500, results 49++). Totally Congested:



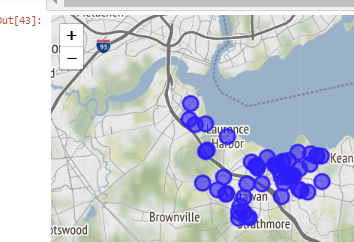
Comparable place miles to the south near Manhattan Bridge (ratio 2500, results 49++).Totally Congested:



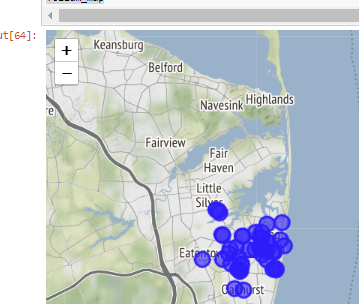
Comparable place miles to the south near 14th avenue (ratio 2500, results 49++). Totally Congested:



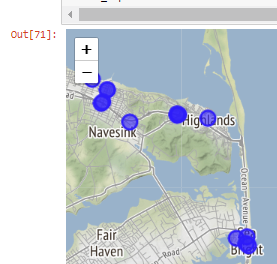
Comparable place miles to the south near Laurence Harbor same ratio/results:



Comparable place miles to the south: Long Branch. Same scenario:

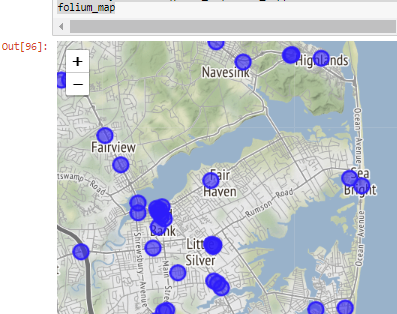


Comparable place: Highlands/Sea Bright. Same conditions, 13 results!. Uncongested!



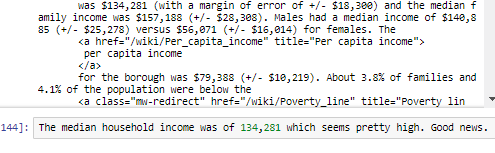
The conditions for stating that the places where comparable, were merely territorial, meaning coast, around considerable civilization and as near as possible from New York. The modus operandi for moving from one place to another was simple google earth searching, cause image recognition or other techniques are out of the scope of this course. So then, the study focus on this particular place found, and find out if its a strong enough candidate based on metrics such as population of near places, sufficient infraestructure and relevant organizations such as schools, hospital and banks, income per capita or per household, among others.

Moving toward the center of the area, inspecting the totally untouch area of Fair Heaven and Rumson, and zooming out:

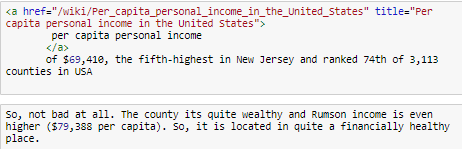
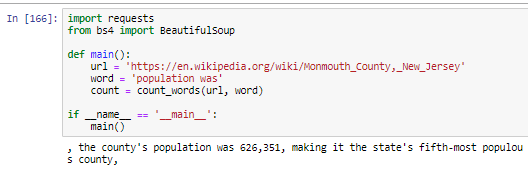


Its near lots of hot spots, but untouched and in a quite populated area. It is so?

Via scrapping and creating functions for searching particular words, i could capture the information of income and population, related to Rumson, that is near the sea and totally uncongested:

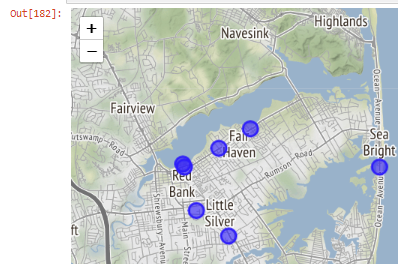


This are the results of iterating this process through the County where it belongs, to see if there are good conditions outside the borough:

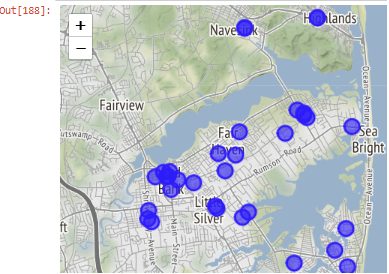


So the last stage of the study was to make sure that all the things were rolling before making the final decision of the results founded, that is, recommend to the investor to strongly consider the possibility of moving towards investing some more in inspecting this particular place and the real possibilities that offered.

Distribution of hospitals in the area (ratio:7000)



Distribution of schools in the area (ratio 7000, 49++)



5) Discusion

As said, the study suffer from quite strong external constraints, limiting considerably its scope and validity, but shows a path to reveal valid information when such cases arrive. By mere exploration of information, one can create the background for further investigation.

As also said before, my original idea was to do a much more interesting Project, related to set some background information, to add to the discusion that is taking place in chile about if its on its way to became a first world country or not. Geospatial information (when more complete) can reveal patterns of the City distribution of organizations and infraestructure, and then compare them to cities from developed countries comparable to Santiago. But the lack and inconsistence of information of Santiago via Foursquare API detected (e.g showing 0 results in an ratio of 900 for a certain result, and later, for a ratio of 950 displaying 29 results), thrown away such idea. Also the limited range of queries abort such effort.

Its remain to be discussed if this approach, with all the limitations, can be externalize, or consider as relevant for other cases.

6) Conclusions

As things are, this report shows that there is a possibility of founding interesting information via simple tools such as API and Web Scrapping, even though clear limitations of info available.

Its prooven, ones more, that Python its a tool that contains great capacity to create information via exploring and exploting sources.