

The influence of the location of the commercial premises when installing a web business in Chile.

June 23, 2020

1 INTRODUCTION:

1.0.1 Background.

We live in an age where information and communication technologies have positioned as an element of vital importance, both for personal and professionals.

It is considered that “Chile is the Latin American country that best uses the technologies of information and communications (ICT) to promote social and economic development” (Thierry Geiger, 2015), which has been reflected in a greater demand for these technologies.

Regarding the personal sphere, “people have increased their level of consumption of ICT-related goods and services, mainly in relation to information services through the Internet and online purchases” (Achá & Bravo, 2009) .

On the other hand, with regard to the professional field, it is not only an issue at the level of large companies. SMEs have gradually incorporated these elements, hand in hand that “companies have generally increased their presence on the Internet, from those that only seek to have a presence on the Internet, to those that implement complete e-commerce portals” (Achá & Bravo, 2009). And it is not surprising, since there is evidence that “the SMEs that made the most use of IT are the ones that obtained the greatest increase in their performance, in contrast to what happened with those that made the least use of IT” (Oliva, Carvajal & Cataldo 2018), which is widely supported by the Marín econometric study, where it reaffirms that the greater the use of ICT, the higher the profitability presented by Chilean companies (2019).

The ICT industry in Chile is on the rise and “the relevance of this sector is that it provides an important part of the national economy” (Gatica, 2018). This is why it is seeking to further boost the market, which implies unsatisfied demand.

To satisfy this growing demand, the idea of forming a company offering computer services arises. It will supply part of the growing demands.

1.0.2 1.2. Description of the problem.

However, it's necessary a market study deeper. In this case a study about how to the current enterprises work will be useful. For this objective it will be used a machine learning algorithm, in specific a k-means clustering, to get the most relevant information about actual enterprises in this area.

All this will be useful to determine the influence of the location of the commercial premises when installing a web business.

2 DATA.

2.0.1 SOURCES.

About the data, first was necessary name of the states in Chile. Then, with the location of states, data downloaded were got with foursquare api. In particular, data about companies of Web services area in Chile, like:

Name.

Location: latitude and longitude. The geographical location, essential to get a conclusion about the clusters.

Category: The particular category. The project is about Web companies, therefore there are a lot of different types of company lines.

Stats: This is information about what people think of the company.

Likes: How people evaluate the company.

With this data the model is fitted, to determine a common characteristic among the companies.

2.0.2 DATA CLEANING.

The process of cleaning the data is as follow:

Even data was obtained with search method of foursquare API, there were data no related with the area of web companies. For this reason was necessary clean it, deleting the unnecessary data.

Besides, boolean data and data with labels were transformed to quantitative data.

Finally all the data was merged in a table.

3 METHODOLOGY.

To solve this problem it was used only one machine learning algorithm: K-Means Clustering. The step by step is the next:

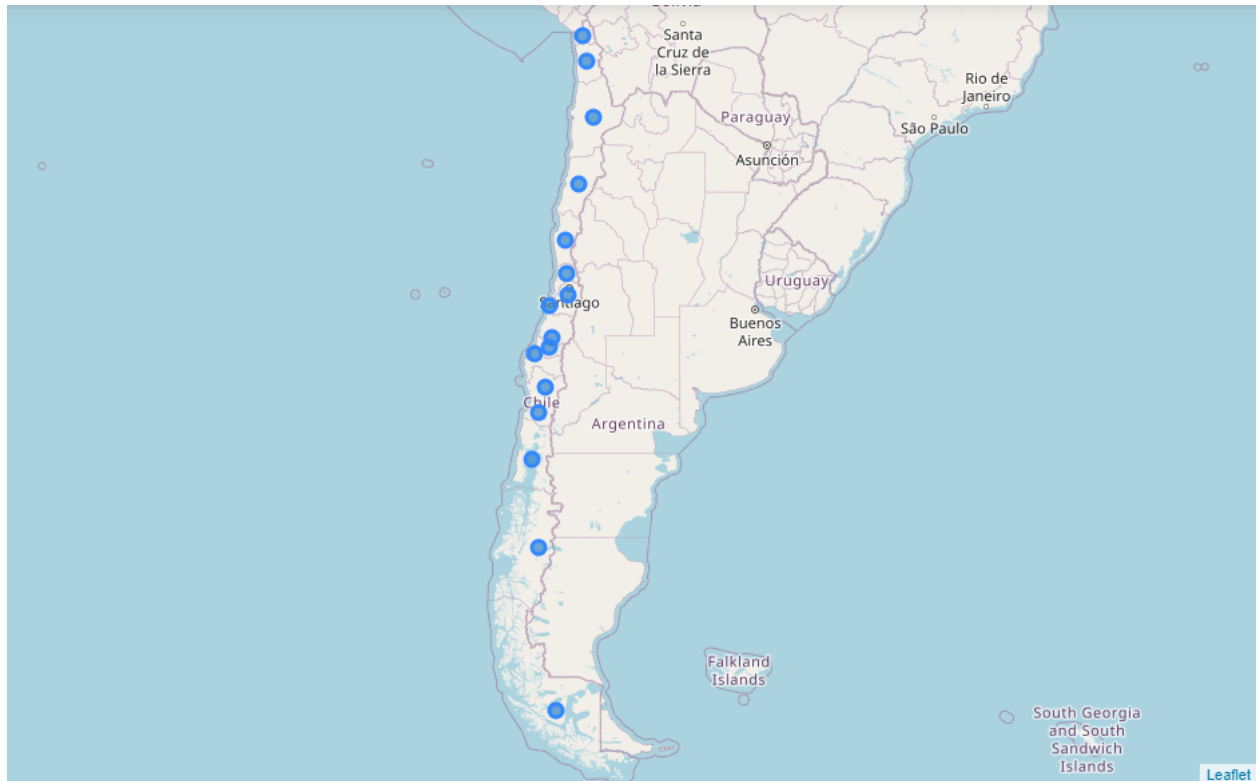
Selecting the relevant data: Not all data is important to the target of this project. Basically all related with a evaluation of the company is relevant.

All the collected data was graphed on a map, to verified if they really belonged to Chile. First just the states. Then, the points obtained with foursquare API.

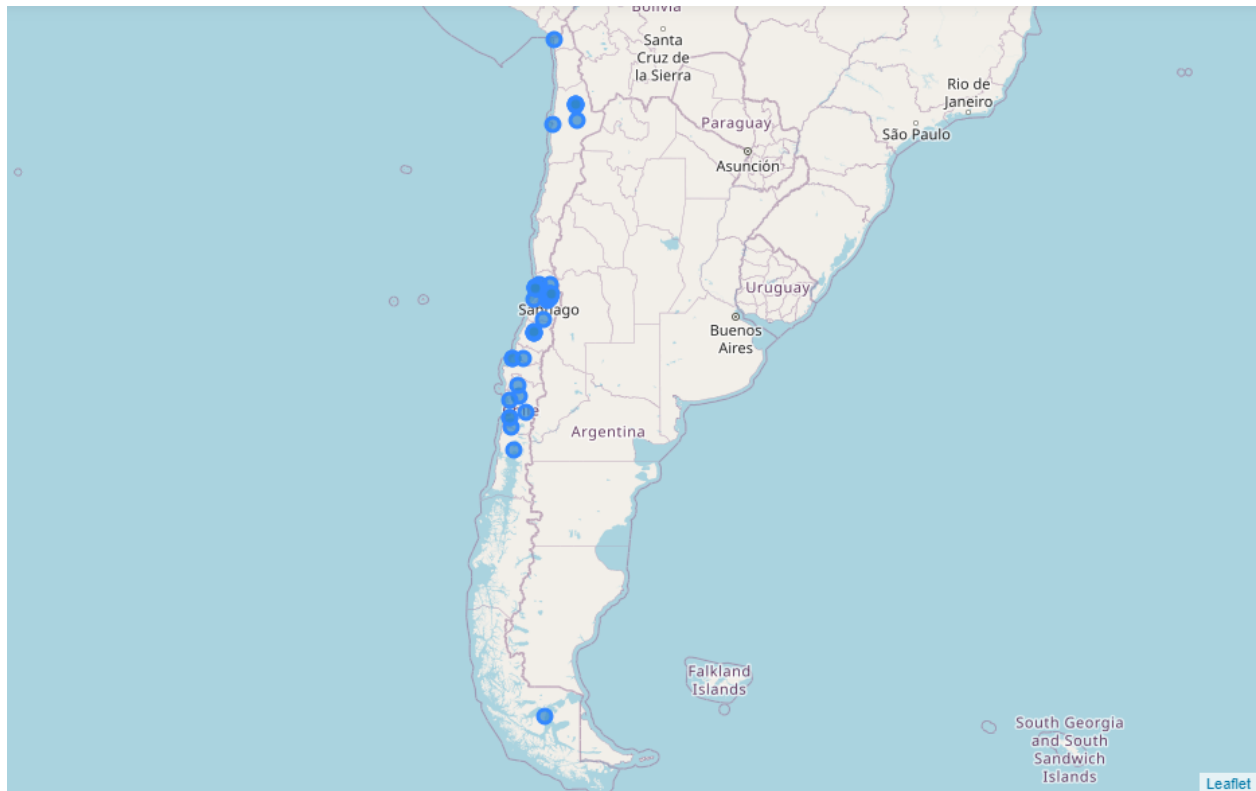
To get the model, primarily it was necessary an study to choose the correct parameters. This is solved with Elbow Method, particularly using the module yellowbrick. After this, the model was fitted with the data of any company.

4 RESULTS.

Here it's possible notice that all state locations are right.

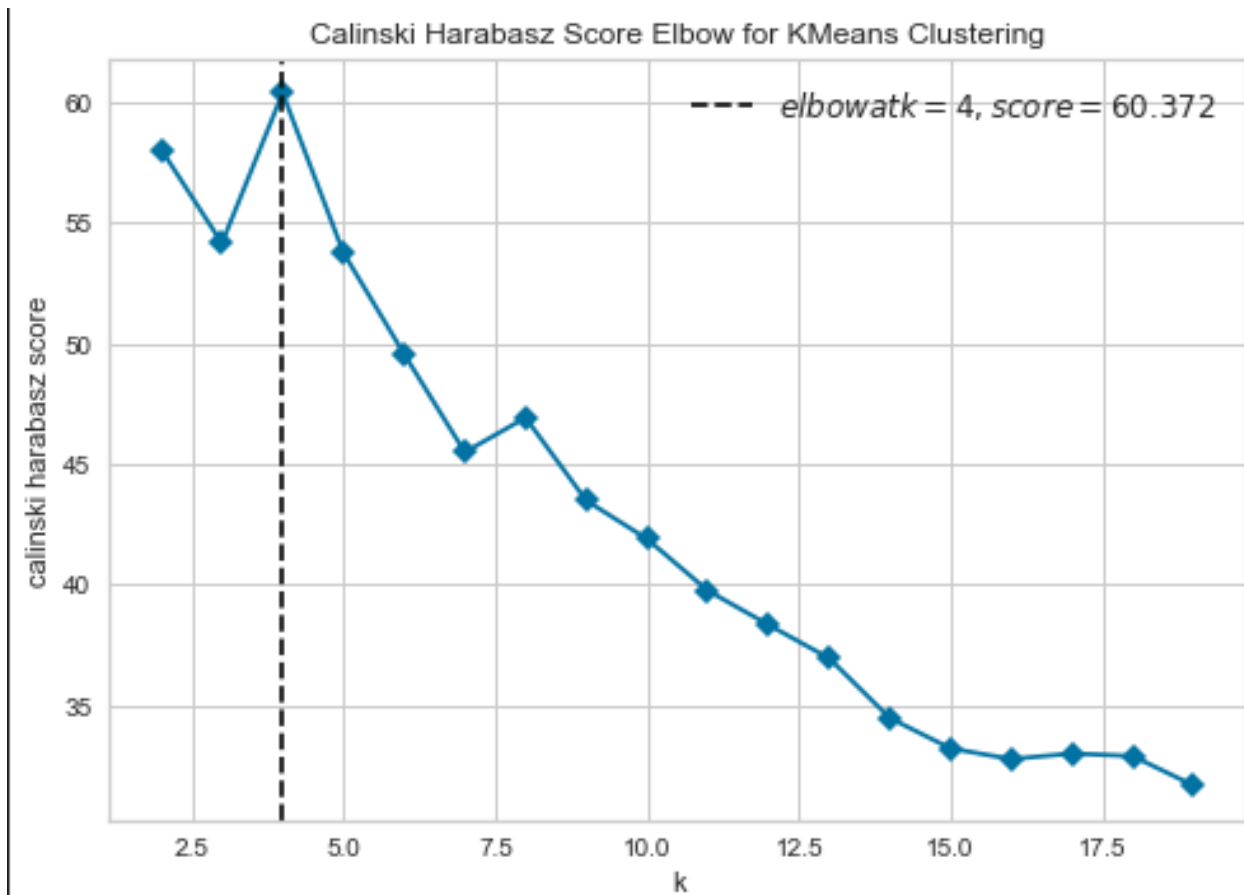


Now the company's locations are:



It's important notice that there's not information about all companies. This is a usual problem in Chile.

Here are the results to decide de quantity of clusters, using Elbow method.



For this reason the number of cluster is 4.

The results of clustering precess are below, after conclusions.

5 DISCUSSION.

Among the points to highlight are the following: - There is a tendency for companies belonging to the cluster with the highest approval to be found in the center of the country. - At the ends of the country are those clusters with the lowest scores. - In addition, there is a lack of diversification in terms of items. In general, web design companies themselves predominate, while in the center there is more diversity, with all kinds of companies in the web sector. This is particularly interesting, since one of the greatest difficulties with companies located at the extremes is the difficulty of meeting between the bidders and the plaintiffs, which means that the sector to which the company is dedicated is 100% online.

6 CONCLUSION.

A company that wants to enter this market, and that wants to use agile methodologies, that contemplate a higher level of interaction, must take into account the location. Although the option of

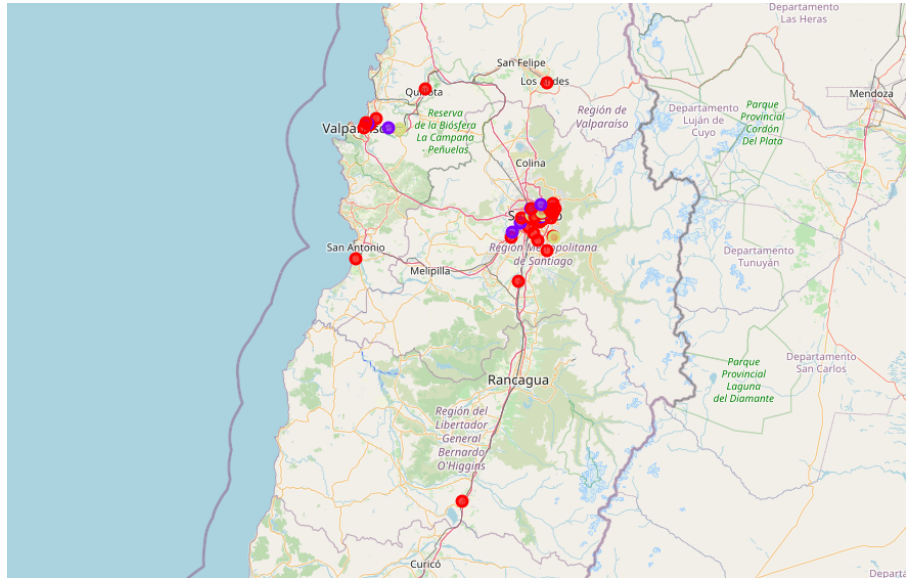
video conferencing exists, it is not apparently what the results reflect. The trend is for companies that need meetings on both sides to choose central locations.

As for the model, it could be improved by adding more data. For example, it would be interesting to see the effect of location and income. Does the business premises (legal location of the company) influence its performance? Also the volume of clients could be interesting to add. In short, it is a model that allows you to have a context regarding what it implies, but if you want to quantify the decision or not to invest in a more central business premises, it is a must.

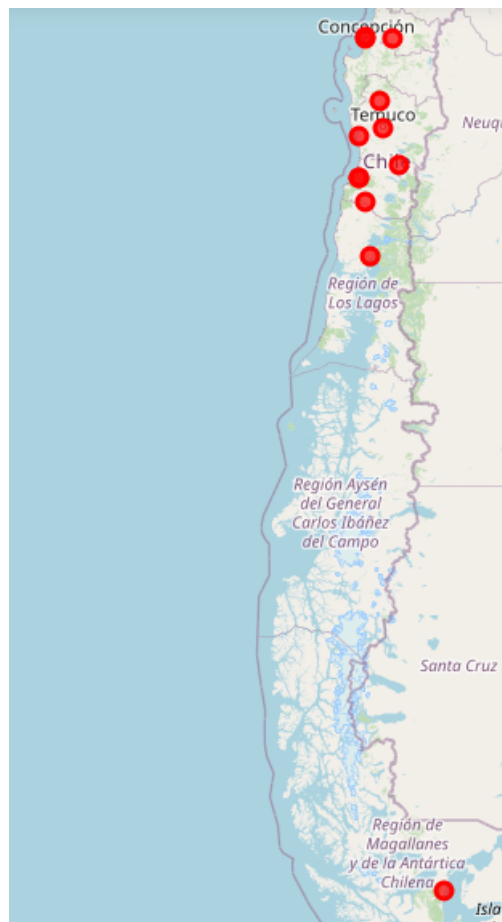
North zone



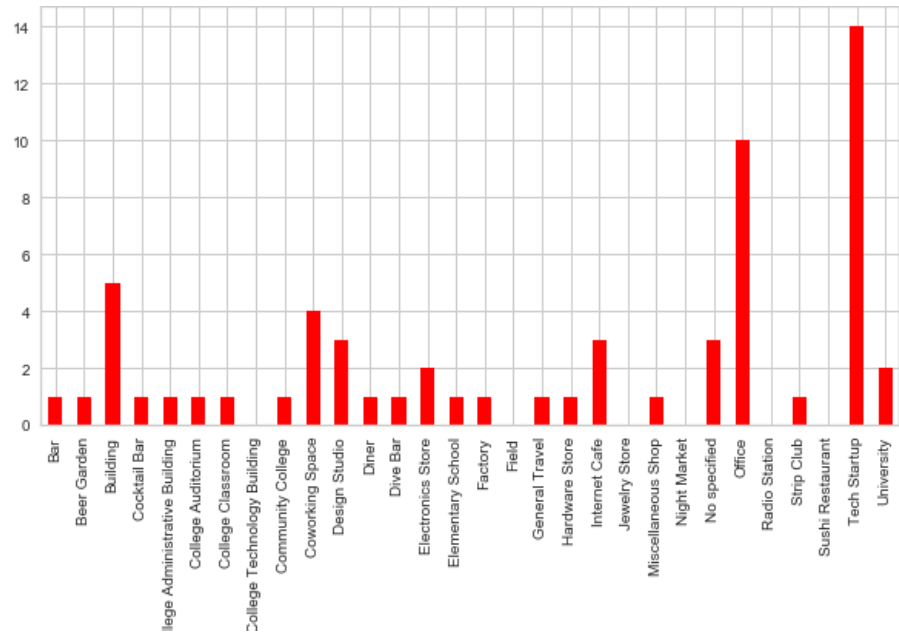
Central zone



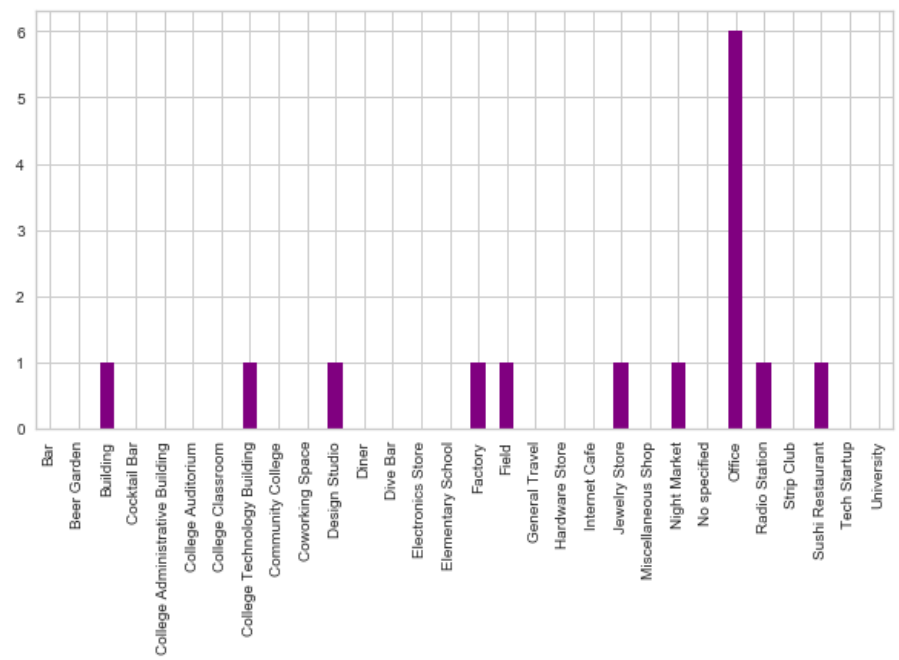
South zone



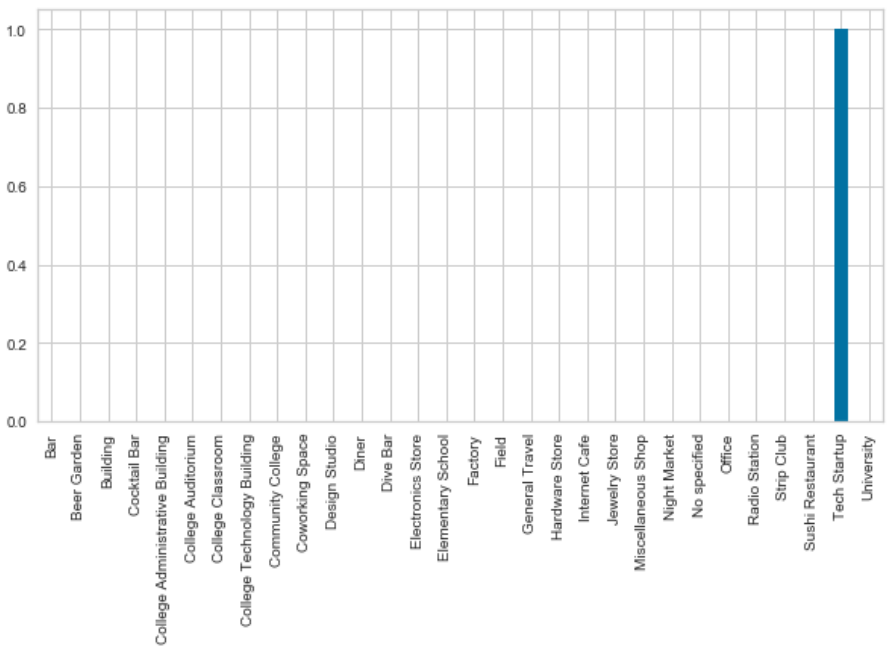
Quantity of categories, cluster 0



Quantity of categories, cluster 1



Quantity of categories, cluster 2



Quantity of categories, cluster 2

