

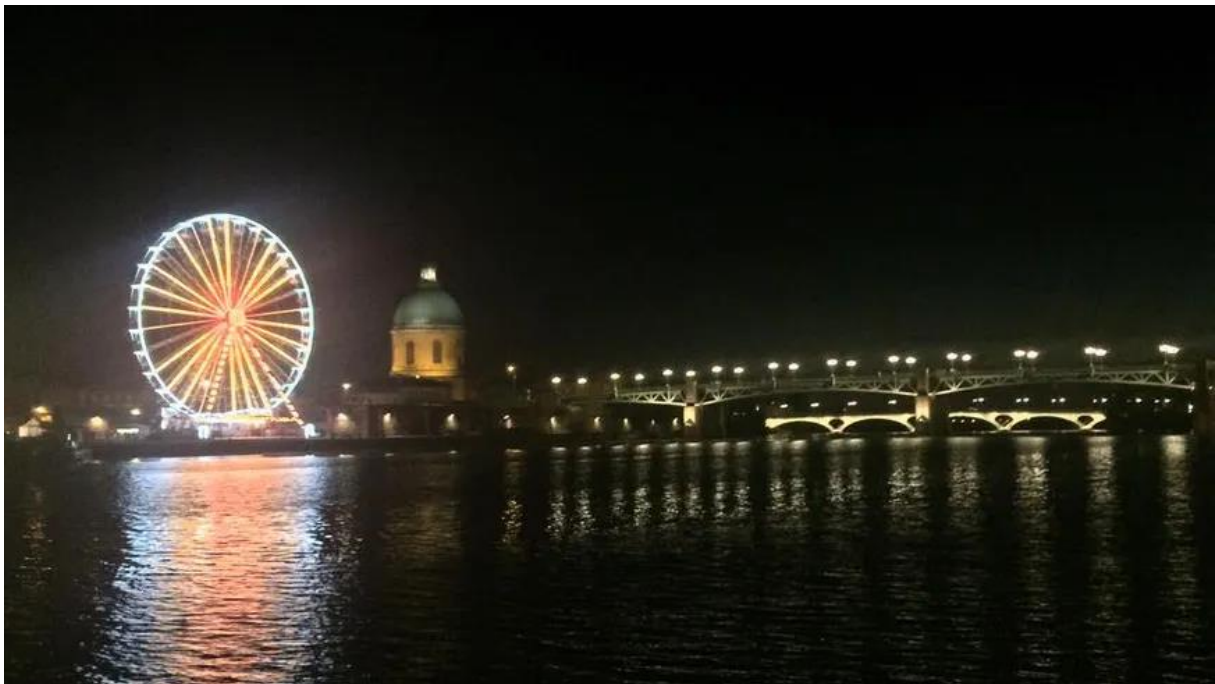
# IBM DATASCIENCE CAPSTONE

"IBM Data Science Capstone Project for Coursera"

## *Business Analysis of Toulouse*

Autor: Benoit Mateu

April 2020



## **Introduction:**

Opening a company is difficult and finding the right industry is complicated. An entrepreneur starting his business will find it difficult to function and make a profit if all the aspects are not studied. Of course, we could go blind, but this is not a solution. I therefore prepared this project, which is primarily intended for businesses or future businesses in the city of Toulouse. It is a question here of rigorously studying the data present on the city of Toulouse, restaurants, hotels, bars, various and varied sales companies, much more. All in order to make a complex but detailed analysis of the current state of the market. We will help companies here to make decisions, since this is one of the essential aspects of Data Science and Business Intelligence. Make yes, but wise decisions, taking into account the issues of the business, which I will detail in the section below.

## **Business problem:**

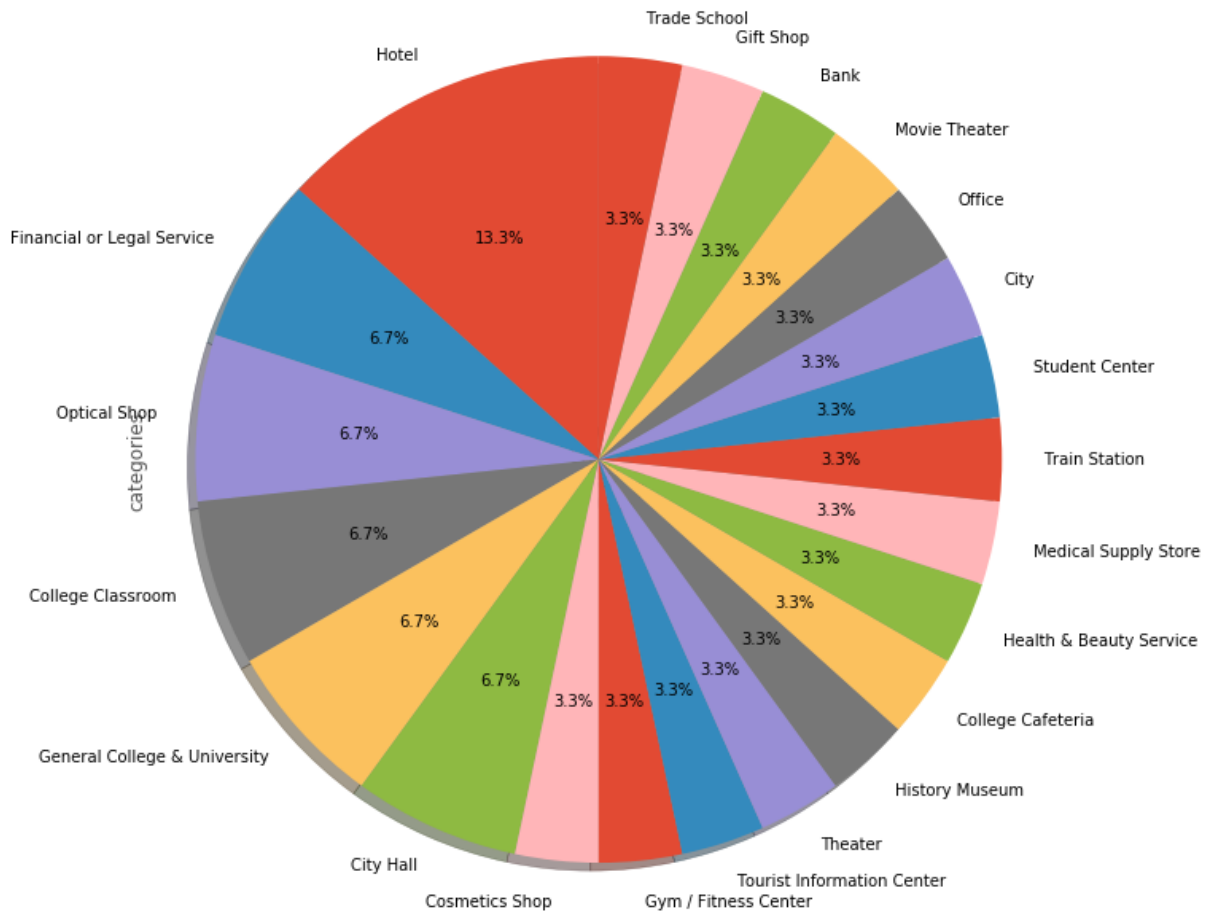
The problem here is quite basic. Opening a business requires an analysis of the market, the location of businesses, their sector of activity, the competition. Choosing a field of activity is not simple, it is necessary to study the consumption habits of the population so as not to embark on a field where there is no future. Analysis is very important and requires rigor in the choice of data. Choosing a location can also be complicated and determine the viability of a project. So, you have to look at what factors are important, one will think of the rate of passage of a street, if there are metro stations nearby, if the place is in the city center or not.

## **Data Analysis:**

Let's start the data analysis part, here we are going to look at what the data can tell us. The aim will be to analyze the objective way and make inferences on what is interpretable from this data. I will try to find the variables we wanted to observe, in order to be able to advise on opening a business and especially in what area. Let's go!

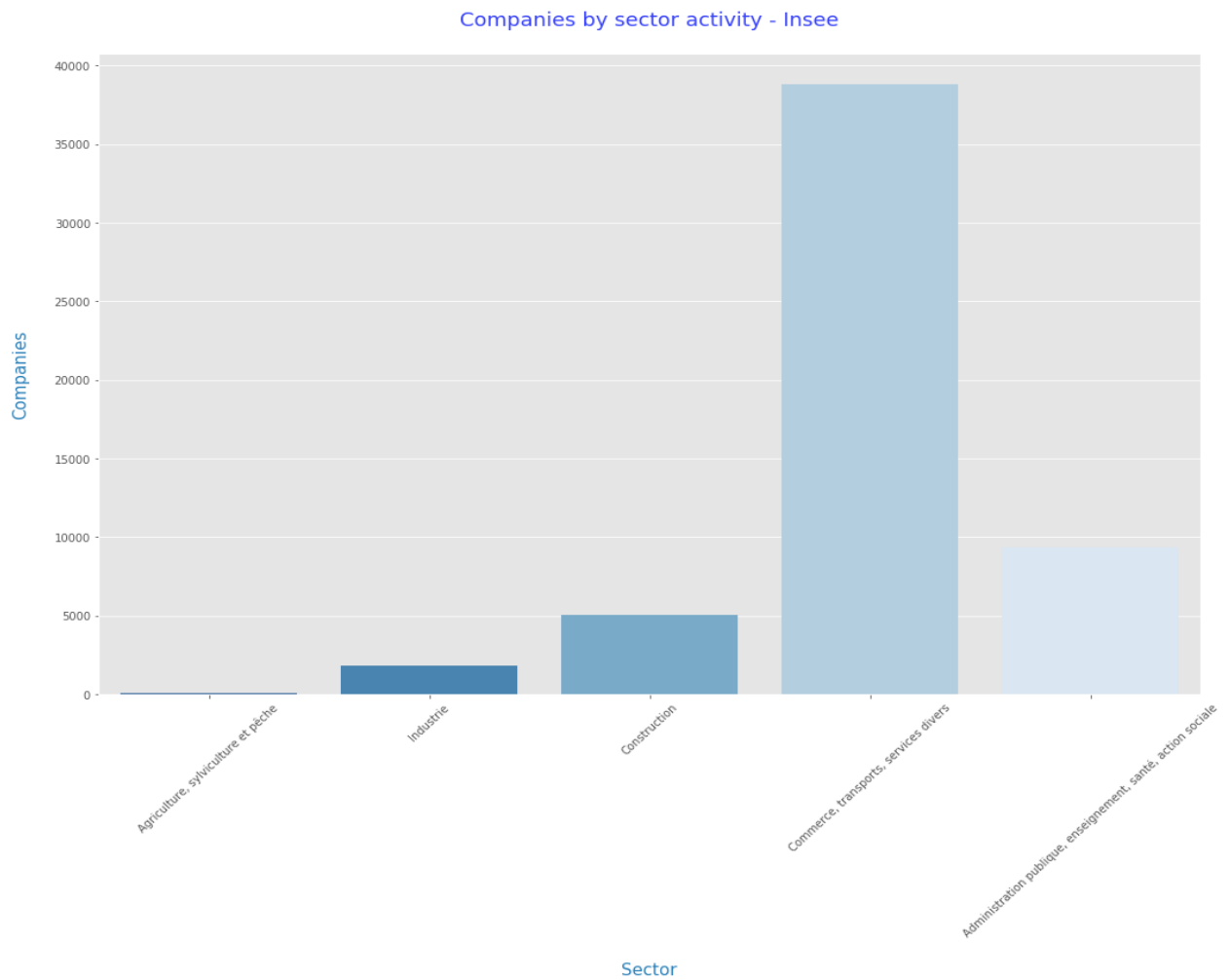
We will first observe the data obtained with Foursquare, these give us a real overview and up-to-date data. For better visualization, I have passed the data into a "pie" graph. As you can see, the hotel industry is a very active sector in Toulouse with a share of 13.3%, then we find in the 6.7% of optical stores private schools. Unfortunately, Foursquare, even with a radius of 10 km, returns us a too small data set, with only 30 companies, which is not representative, so we will have to cross this data with other relevant data

Companies by sector of activity in Toulouse - Foursquare data

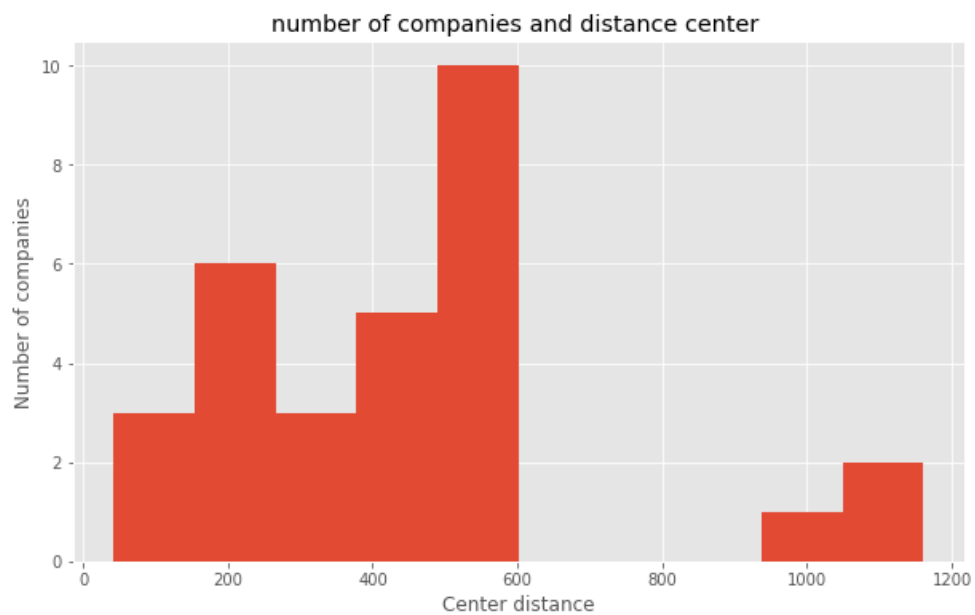


With a larger dataset, we can already see more clearly about the business sector, more generally. We have chosen to use INSEE data, which represents the limited data library open for Toulouse. We find the business sector of companies, which I put in barchar so that it is more visual.

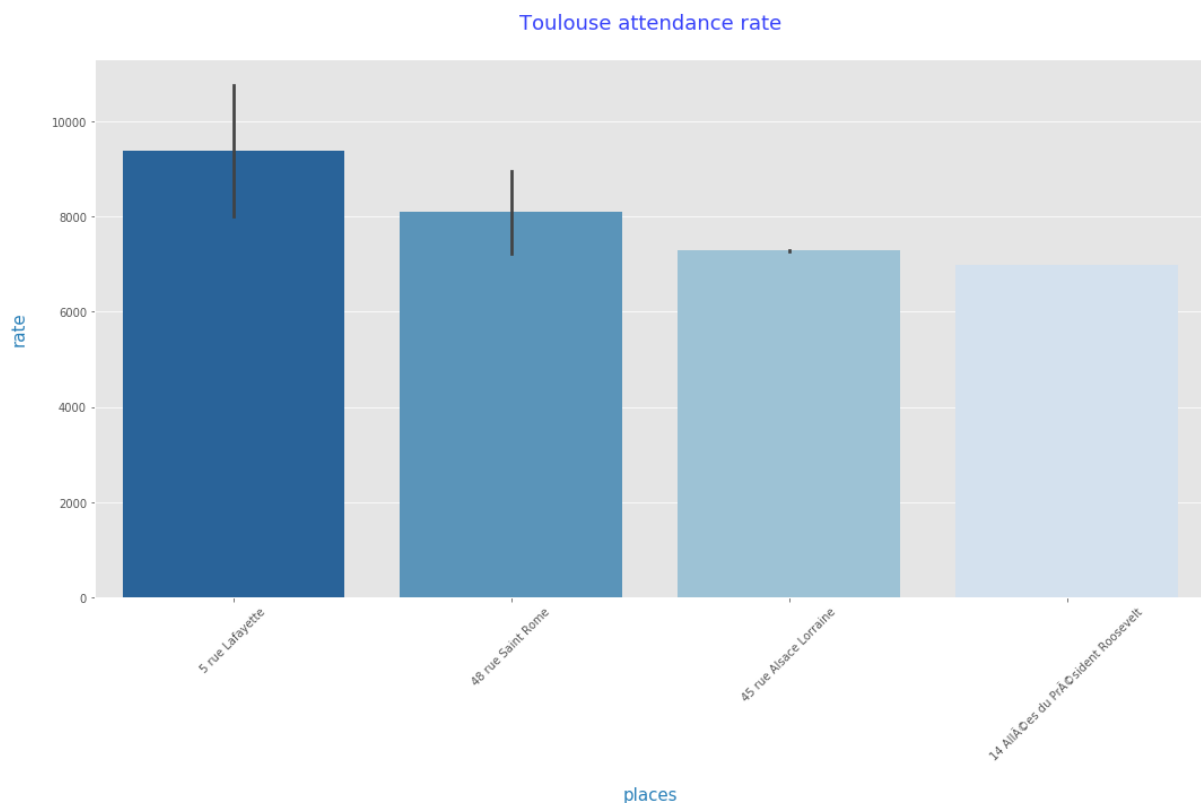
By observing the data we see that most of the companies established are in the service. The fact that there is little agricultural society makes sense since we are in a purely urban environment. The construction sector is not really active in this finding, as are training organizations or the health sector. This will allow us to qualify these Foursquare data.



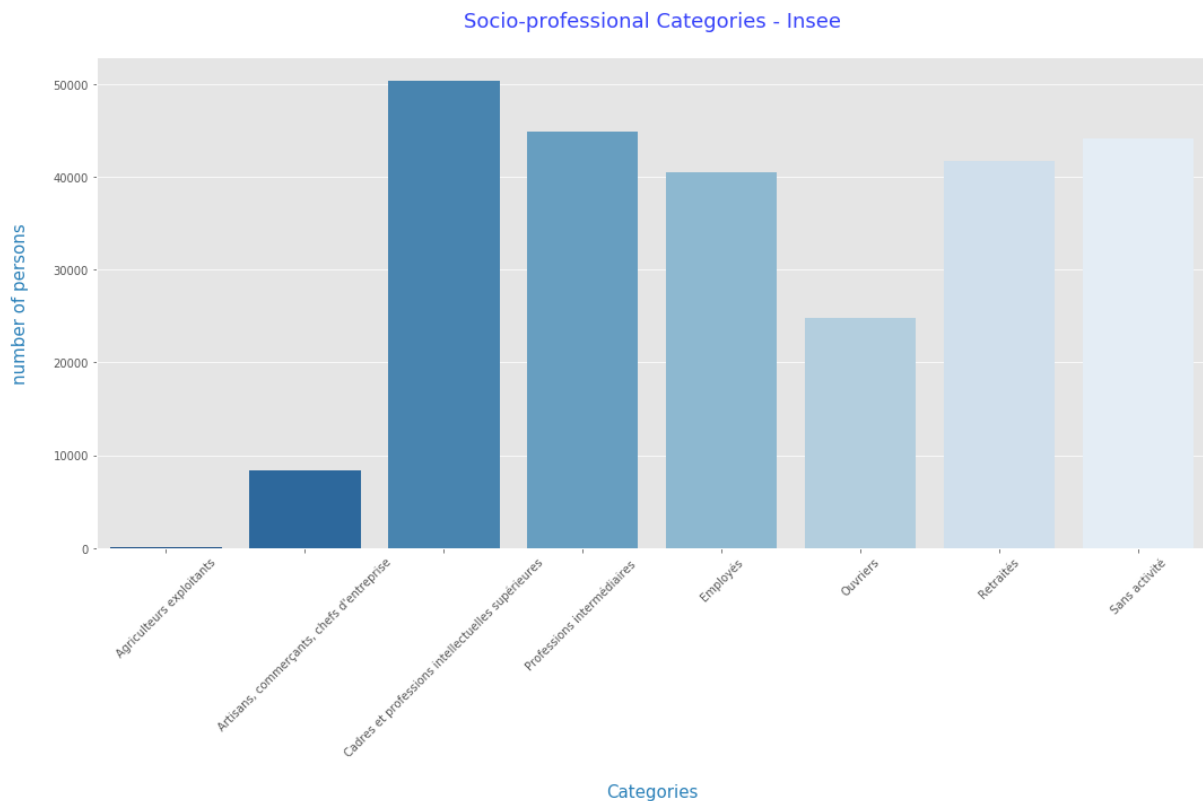
Now that we have an idea of the business lines, we can move on to other important factors. Here we will look at the location of the companies, in relation to their distance from the center. Here we are still using the Foursquare dataset.



As we can see here, the companies are located within 600m from the center, with more than 19 companies between 400 and 600m, which proves that the area is really active. It is therefore beneficial to set up within a maximum of one kilometer from the center, if we believe the Floursquare data. Of course, setting up in this department means that the competition will be tough, but there is bound to be a reason why companies are located in this sector. To help us, we will have to look at the official data of the city of Toulouse. This is pedestrian crossing data, censuses that have been made over time. Unfortunately, these are all the same approximate data, but they give us an indication of the busiest places in the city.



As we can see here, the busiest streets are clear to us. We find in first position the “Rue Lafayette”, followed by the Street “Saint Rome” and finally by the street “Alsace Lorraine”. The companies are therefore located in downtown streets, which are close to the streets shown above. The more crowds we have, the more potential customers there are. The question that remains to be asked is, what is this clientele, which group does it belong to? Are we dealing with affluent people or not, are they hardworking? What products are they interested in? Or what services? To do this we will have to call on the next dataset, still from INSEE, the socio-professional categories of the city.



Here we can see the socio-professional categories of residents of the city, of information collected, we see a small share of farmers, since we are in an urban environment There is a significant share of senior executives, intellectuals, followed by intermediate professions. This can be a pretty attractive target for a business creator or anyone who wants to start a business. You'd think it was just someone who was unemployed or not working. We rather think that this part includes students, Toulouse is a city with many schools and a lot of potential. Unfortunately, nothing was specified on the dataset, so this is pure inference.

We must therefore focus on managers and students, analyze their income and especially their consumption habits. With these points, the basic analysis will be clearly more interesting and especially to perhaps help companies that do not have access to data often paid on the net.

### **Results :**

The results show that most of the companies are located in the city center. The market in Toulouse is based on the service companies, most of the population is manager or student. We were able to determine three streets with a high density of passage or it could be interesting to be positioned. I hope this data can help.

### **Conclusion :**

Although it was interesting to do this project for datascience, I encountered data set problems and insufficient numerical quantity for the analysis of relevant data. Even when we cross with some open data, we don't really have data to do a real market analysis. We have

tried to cluster with likes in conjunction with distance, the three groups that stand out are not significant. I was clearly limited by the Foursquare API ... As you will see, we preferred to exclude this data from the result. Of course, I think to redo advanced analyzes with important dataset, since I like this exciting environment that is data. Thanks for reading.