**Capstone Project**

1. **Introduction and Business Understanding**

**In this project we will look for the price range for a new restaurant.** We are going to check the **food venues** in the center of the major cities of **Morocco**, and build a model on their foursquare data. We will use their rating, food type and position of the city and within the city to create a model that will predict the best price range for your new restaurant in this area.

1. **Data Understanding**

**Moroccan cities data**

The table of the major Moroccan cities is imported from wikipedia <https://en.wikipedia.org/wiki/List_of_cities_in_Morocco>, then sorted, polished and finally all the city with more than 200.000 inhabitants are selected.



**Food Venues Data Imported from Foursquare**

With this function we are going to import the 33 food venues closest to the center of the city. Unfortunately we cannot put up the maximum number possible (50) for limit in foursquare accout premium calls (500 max per day).

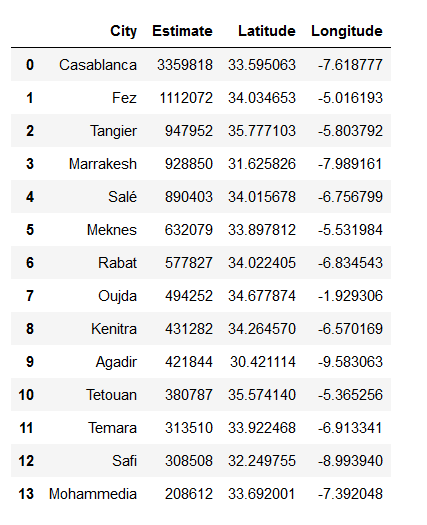
1. **Data Preparation :**

**Wikipedia Data :**

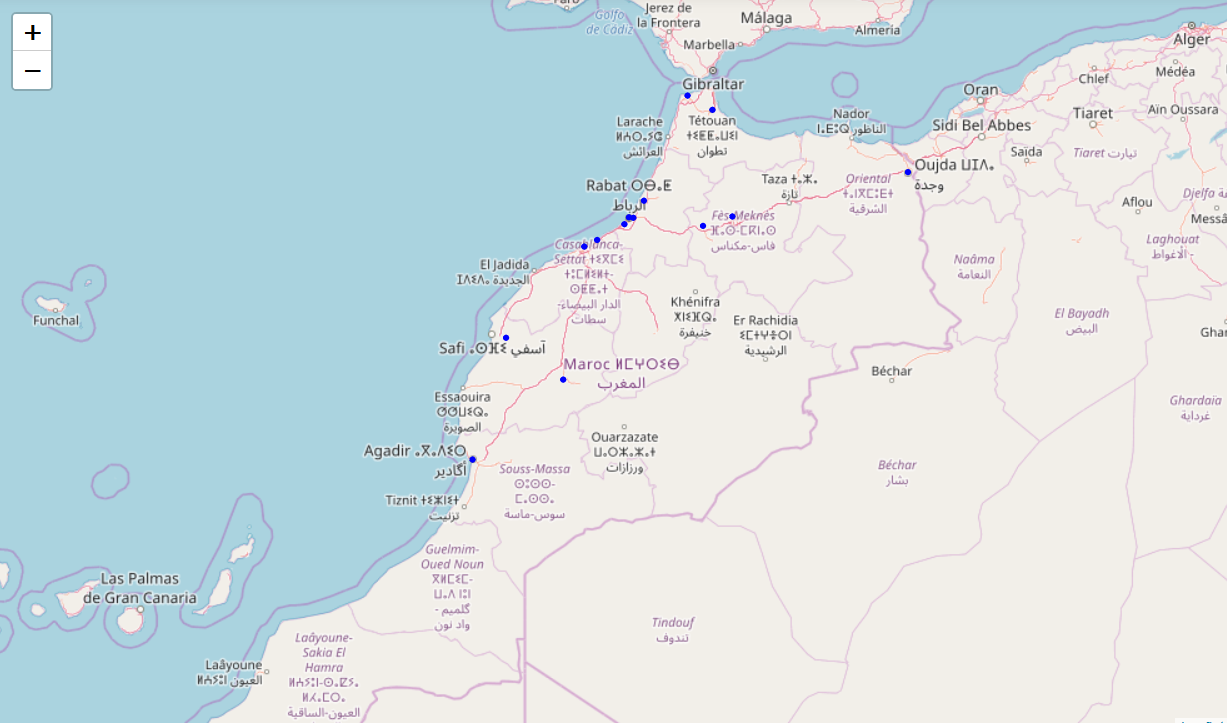




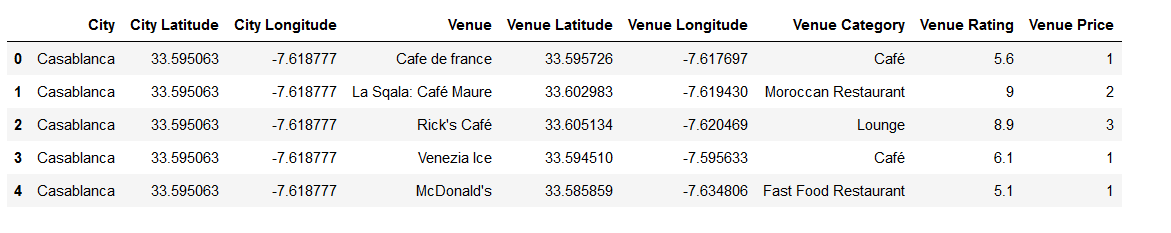
The final Geospatial data about the cities should be :



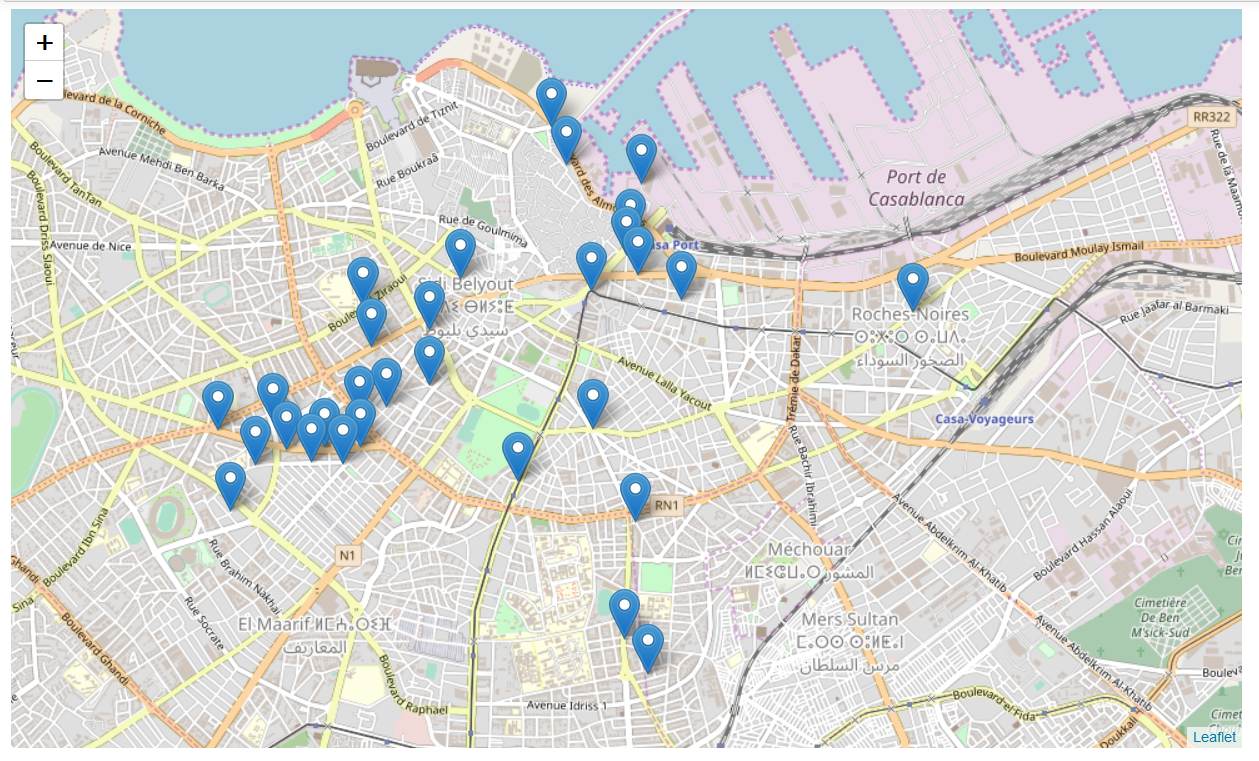
If we plot the data using Foluim



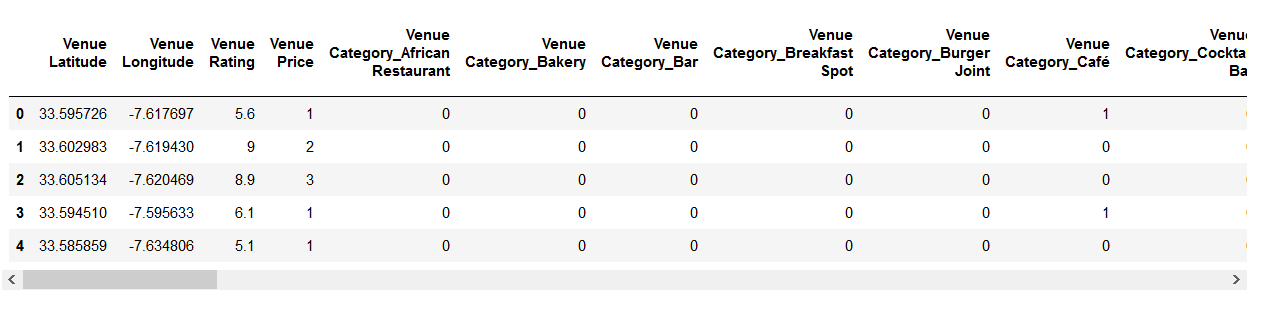
Adding Foursquare Data  we should get:



If we zoom into for example Casablanca city, we should see :



We then transform the feature values into binaries, using Dummies features and drop the coordination data

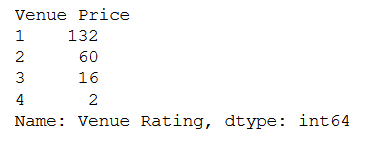


1. **Analytics Approach and modeling**

The machine learning algorithm I used is K nearest Neighbor (KNN), and I used 2 evaluation test (Jaccard and MSE)

I removed the venue latitude and longitude because, in the susequent machine learning algorithm they were needlessly adding complexity.

The Lat Long values are surely more useful in a project regarding only one city instead this one involving the entire nation. As an exploratory analysis I checked the price unique values count and I discovered that only 2 are at tier 4 (the most expensive),

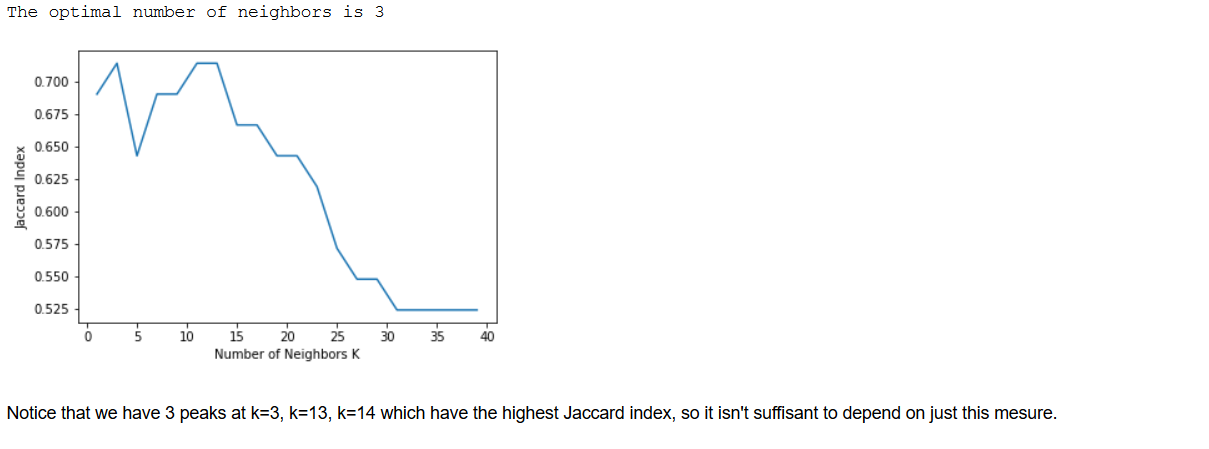


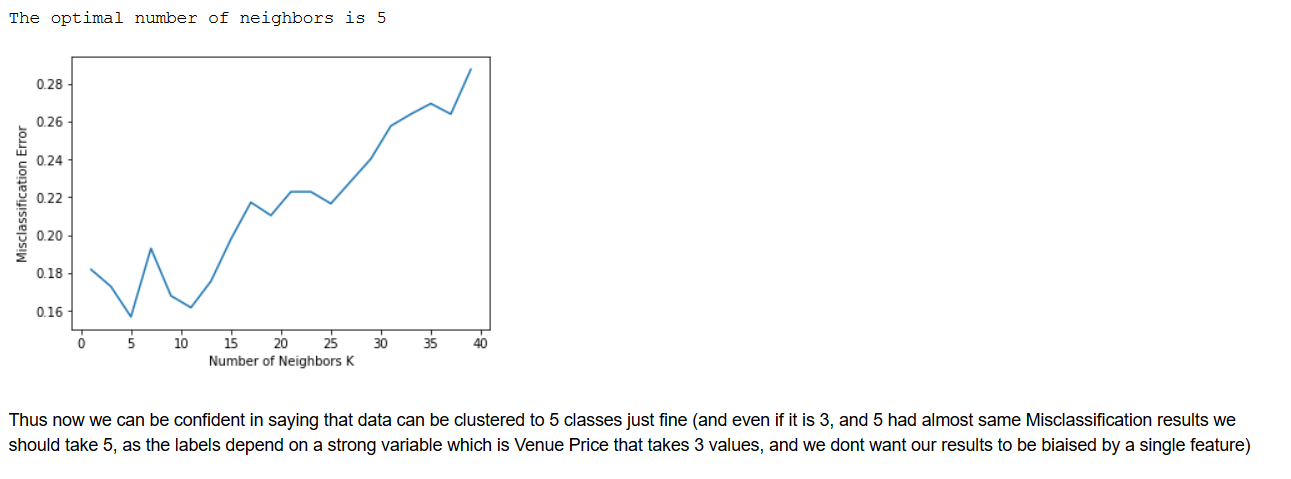
so I decided to treat that data as outliers and I removed them. Moreover, the presence of only two values skew the results of the K fold cross validation.

As for the machine Learning algorithms I choose two. KNN and Kfold Cross Validation. and I used 2 evaluation test (Jaccard and MSE)

They both are classification algorithms, made for the kind of task needed here. The first is of course bound to the randomness of the train test split, the Kfold method reduce much of this fitting issue, you can see the hyperparameter K dropping from 14 or 11 to 5.

1. Evaluation :



****

1. Conclusion

The biggest shortcoming of this project are pretty evident, unless a 400$ monthly subscription you have 500 premium calls per day to the Foursquare API, thus the dataset dimension is not impressive. Plus the definition of price range is pretty vague. Lastly Foursquare is not at all common in Morocco.

The best source for data is to do web scrapping but that would take so much time.