

Capstone Project

Insights on best commercial
location for chocolate shop franchise in Paris





Agenda

1. Introduction & Business problem
2. Business Problem
3. Data
4. Methodology
5. Results
6. Discussions and Recommendations
7. Conclusions

Introduction & Business Problem

- Usage of analytics and machine learning method are one of the major assets for Investors as a decisions helper
- Main goal is to find best commercial location using data collected from different sources
- Our focus is to answer the following question :

**Where could be located the next
chocolate shop in Paris ?**





Data

To answer our question we need to transform it to a Data Science problem and use machine learning algorithms to solve it. Data used are:

- 1- List of neighbourhoods of Paris, the city contains 80 in total

Source: https://opendata.paris.fr/explore/dataset/quartier_paris/table/

- 2- Latitude and Longitude of each neighbourhoods

Source: https://opendata.paris.fr/explore/dataset/quartier_paris/table/

- 3- Population of each of the neighbourhoods for 2012

Source : <https://public.opendatasoft.com/explore/dataset/iris-demographie/table/>

- 4- Popular Venues with Category « Chocolate Shop » near to each one

Source: Foursquare API

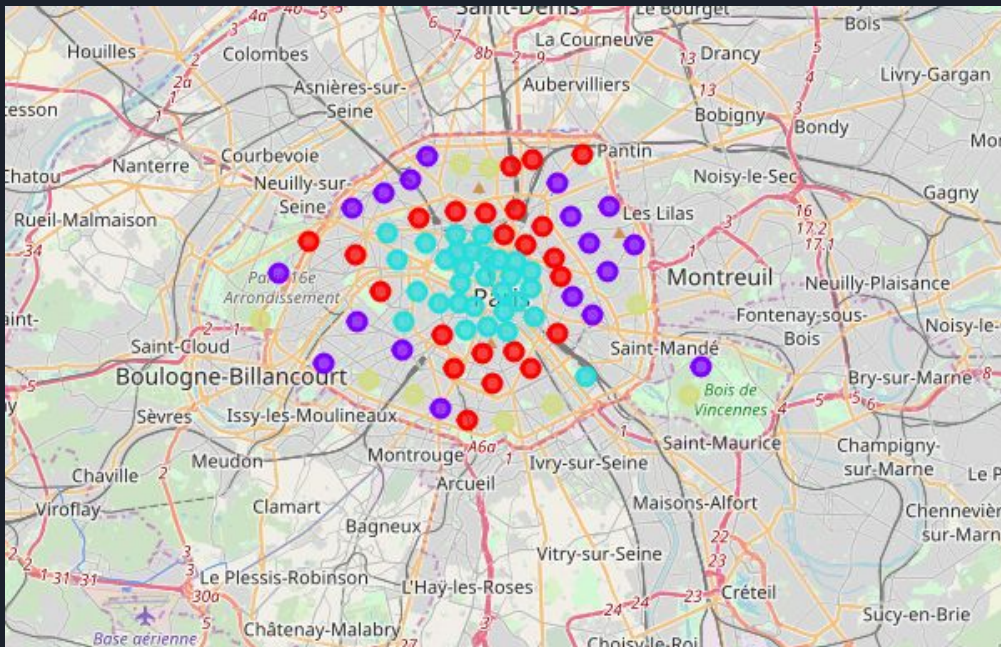


Methodology

- Collect and clean the data from the sources listed before
 - Longitude and Latitude of Paris Neighbourhoods
 - Population of each neighbourhood
- Exploring and correlating needed inputs
 - Building a single data frame containing : Neighbourhood, Longitude, Latitude, Population
 - Getting List of Venues using foursquare API within radius of 500m and filtering on category chocolate shop
- Using clustering algorithm ie: kmeans to cluster chocolate shops in Paris on each neighbourhood
 - 4 classes used
- Visualizing the results

Results (1/2)

Clustering our data on 4 clusters gave the following result on Paris Map:





Results (2/2)

Cluster characteristics are :

	Chocolate Shop	Latitude	Longitude	Population
Cluster Labels				
0	0.001304	48.863419	2.345024	24335.193599
1	0.001111	48.865161	2.346459	45427.428363
2	0.006043	48.860417	2.341183	7388.539355
3	0.000000	48.848361	2.345104	71284.628451



Discussions

- Cluster 2 (light blue) around the center of Paris is the area with the biggest number of existing chocolate shop but fewer population on it
- Cluster 0 (the red) is surrounding it with less number of chocolate shops and more density of population
- Cluster 1 (the blue) is surrounding the red one with higher population and less shops
- Cluster 3 (light yellow) is the area with the highest population but fewer competitors

⇒ **Recommendation** : chocolate shop to be located on cluster 3

⇒ **Further tuning** could be done by adding to our datasets :

- KPI about touristic flows on each area
- Risk index about seine flood on each neighborhood



Conclusions

- The current capstone project has enabled us to transform a business requirement and to a data science problem
- First data is collected then cleaned and correlated. Then we merged it with venues collected from foursquare and we prepared a final data frame to be feeded to the kmeans clustering algorithm
- Results shown that the center of Paris is the zone the most competitors concentrated and we recommended that the new chocolate shop to be opened outside this zone i.e.: cluster 3 as show in the map above.
- More insights for such use case could be found if we enrich our data with other sources.



Thank you