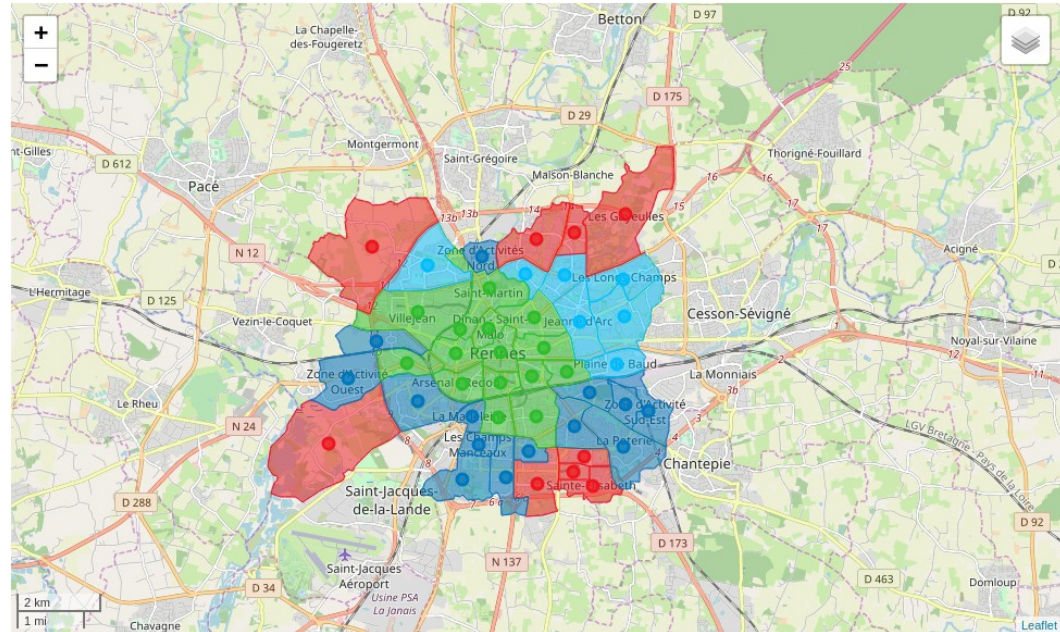


Recommender System for the City of Rennes Neighborhood's Clusters



MAHE Pierre-Antoine

March – April 2020



Recommendation systems are valuable for real-estate agents

- Searching a living place corresponding to a client expectations can be a challenging task
- Cities are bigger than ever and always changing
- Recommendation Systems could be used to narrow the search, saving time and efforts.

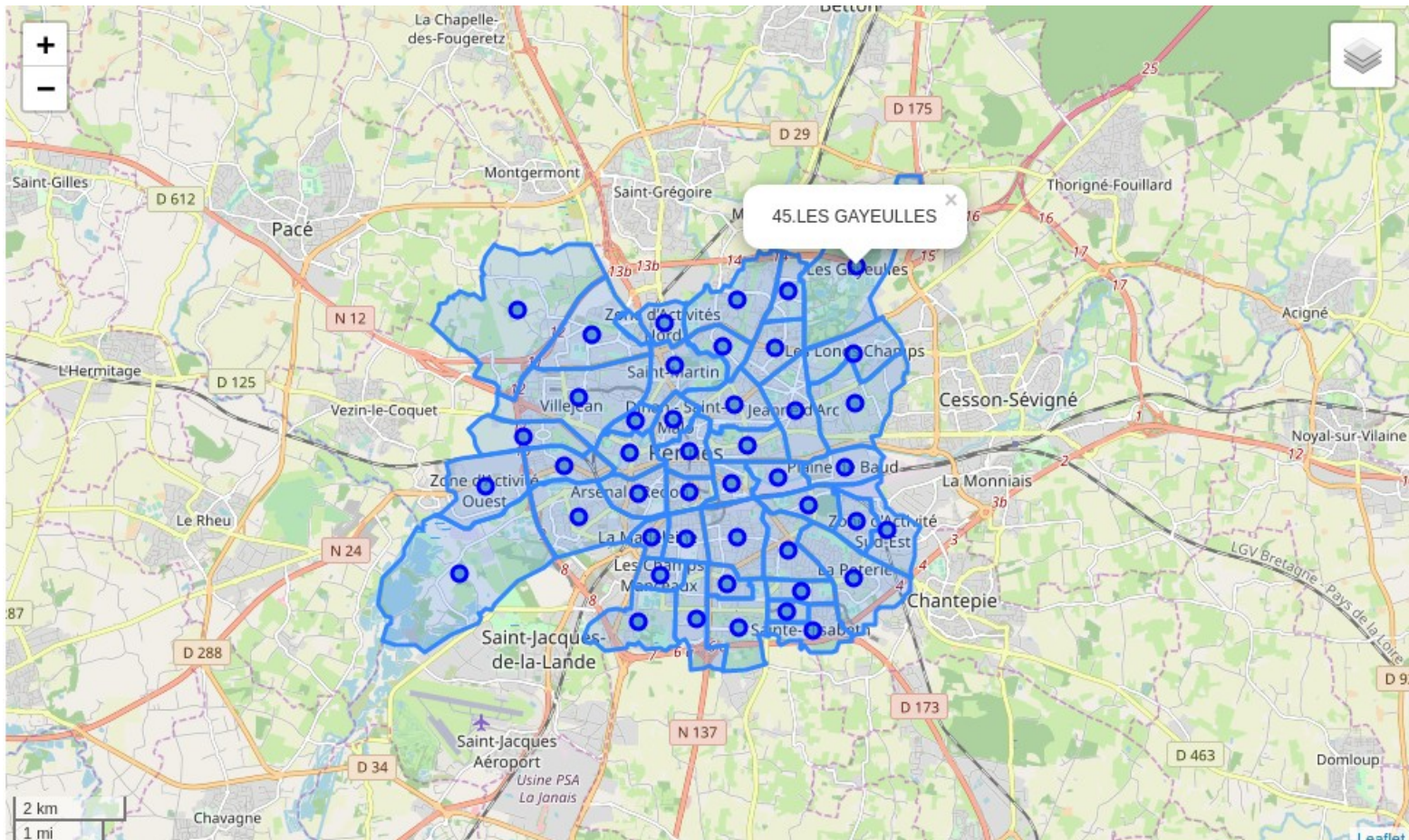


Data Acquisition

- The Neighborhoods location are extracted from the French Administration Open Data website (<https://www.opendata.gouv.fr>) and published under the Open Database License v1.0
- The data set : `perimetre-des-45-sous-quartiers-de-la-ville-de-rennes` contains all the information needed about each neighborhood (ID, Name, Latitude and Longitude, geometry)
- All other columns are either undocumented or irrelevant for our analysis and are dropped.

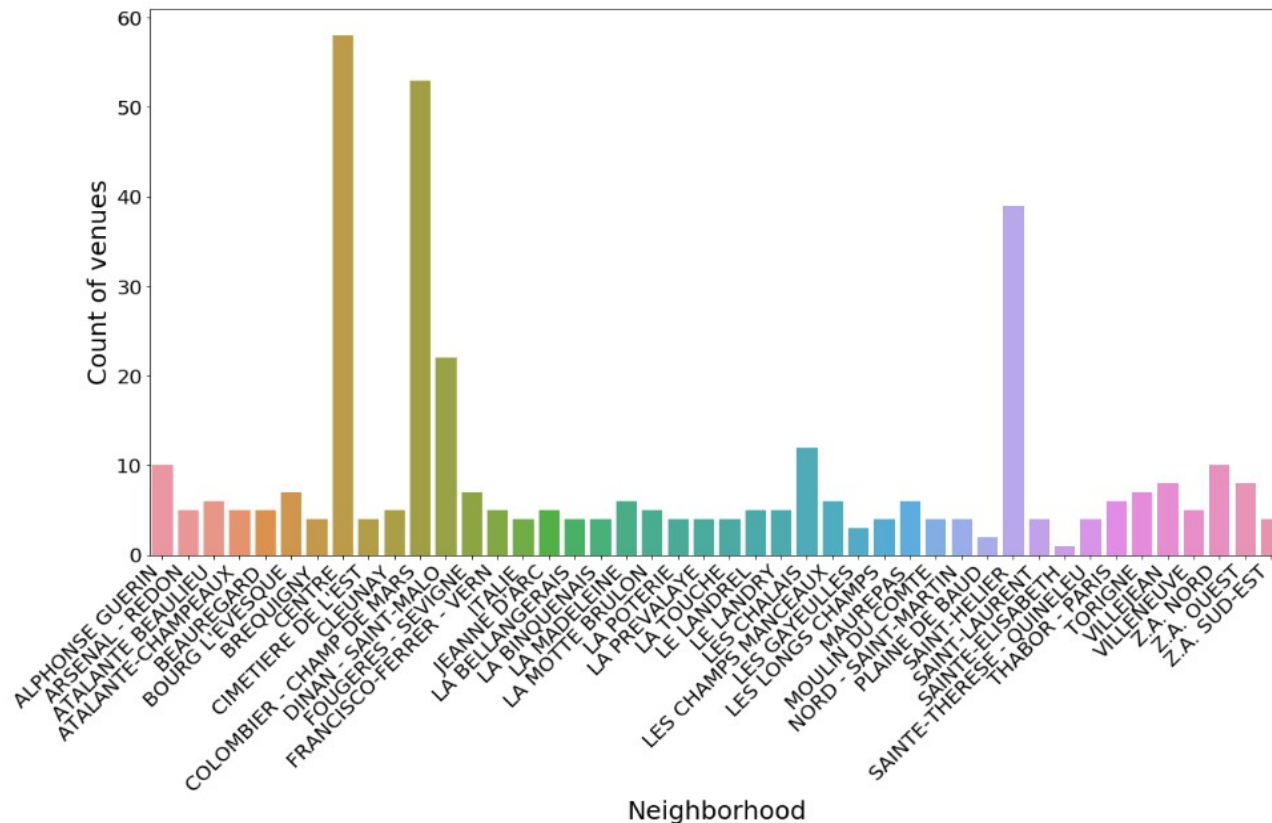
Neighborhoods First Visualization

- Neighborhoods are plotted with their geometry. The location, latitude and longitude, is used as a popup containing the neighborhood ID and name.

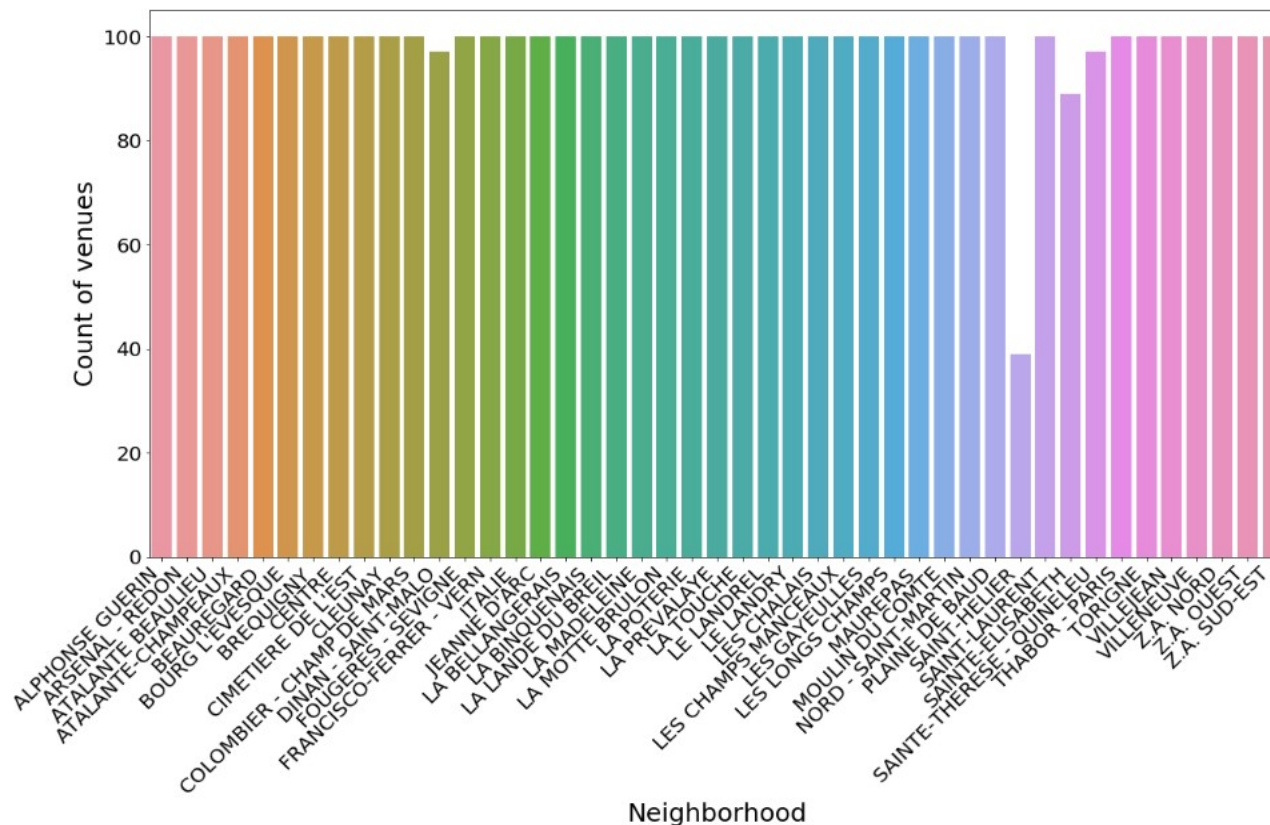


Foursquare Data

- Explore endpoint :
 - Return recommended venues around a given location, hence a limited number of venues per neighborhood

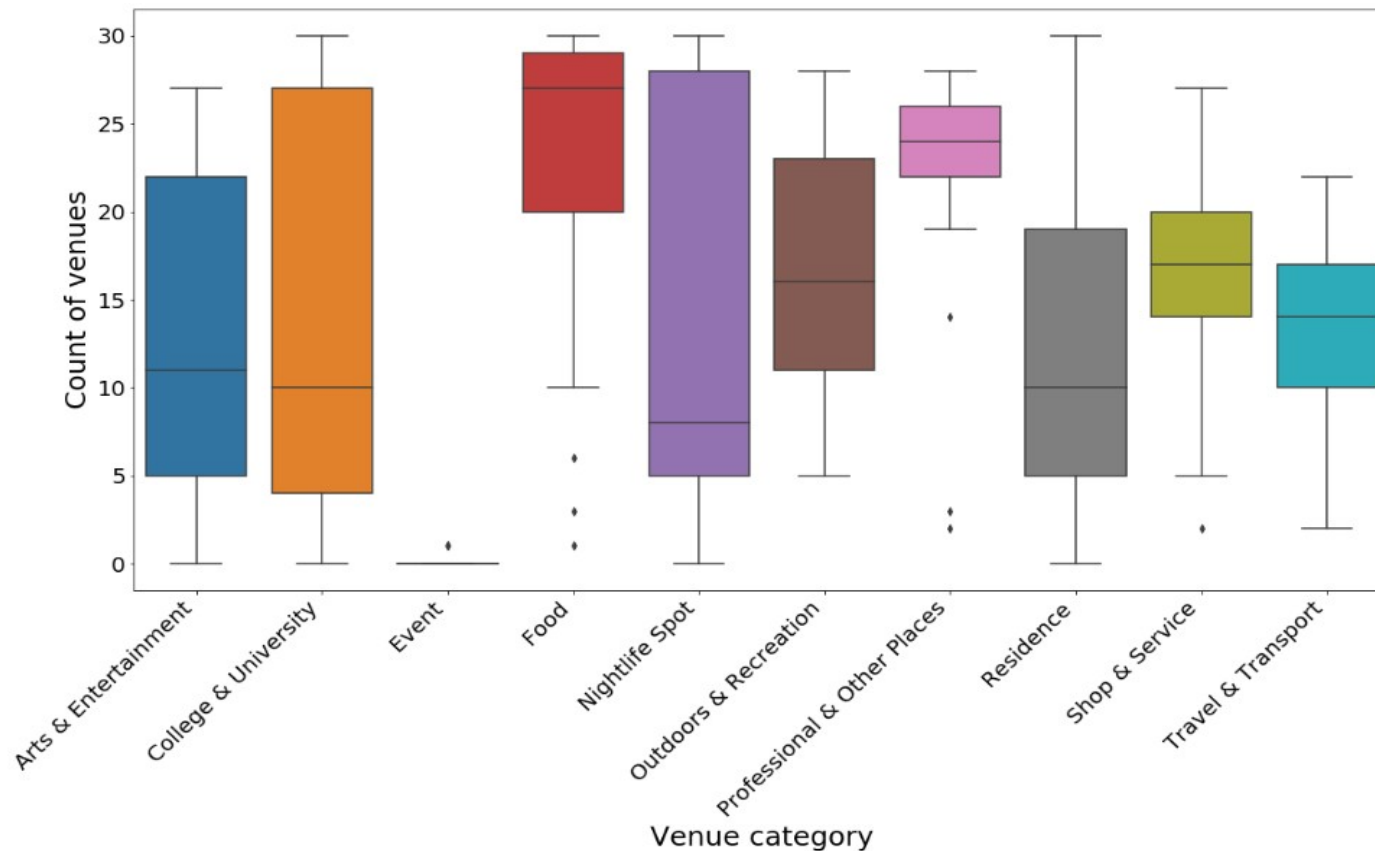


- Search endpoint :
 - Return any venues in range of a specified location. More venues per neighborhood but difficult to analyze (LIMIT = 100)



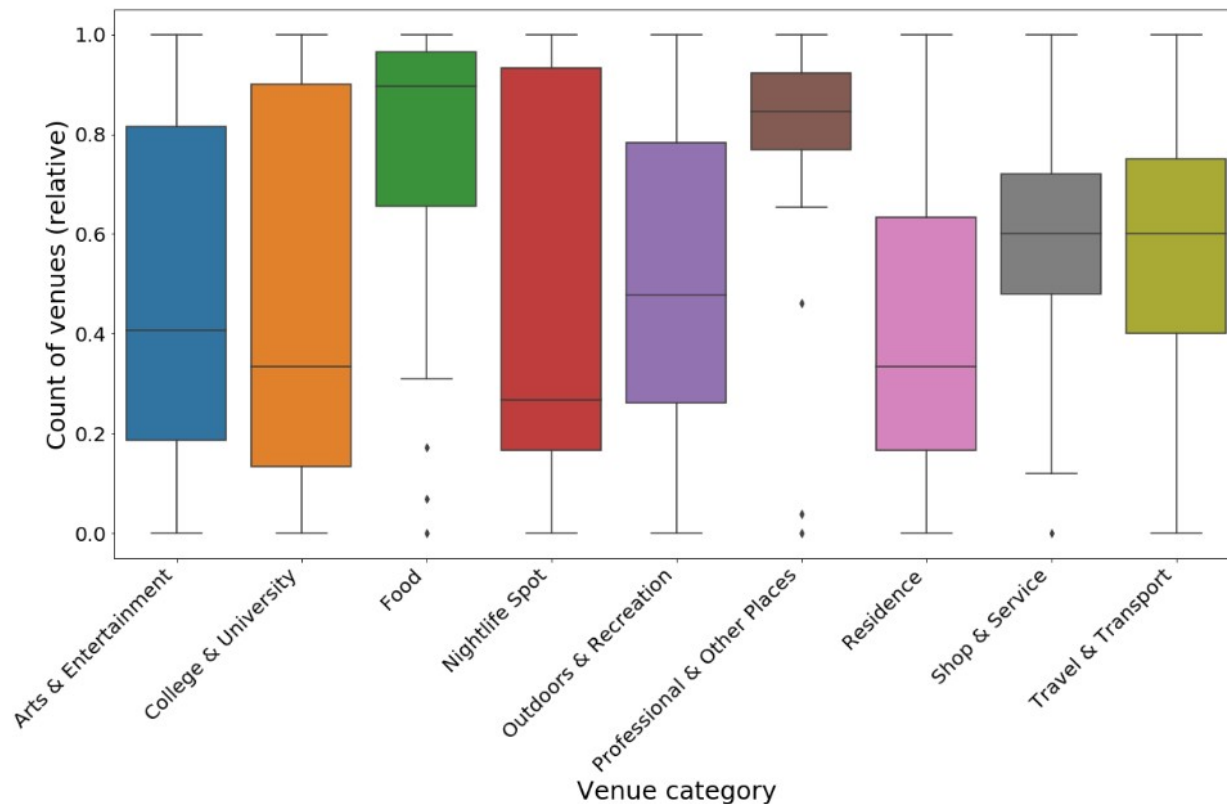
Foursquare Data

- Solution : Category Endpoint and Search Endpoint
 - Get the 10 top-level Foursquare categories to store venues returned by the Search endpoint



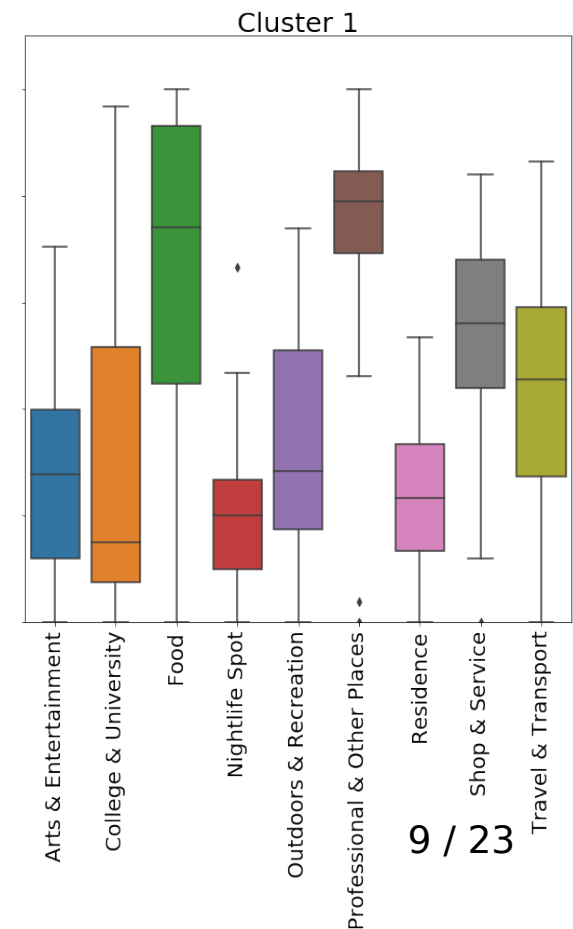
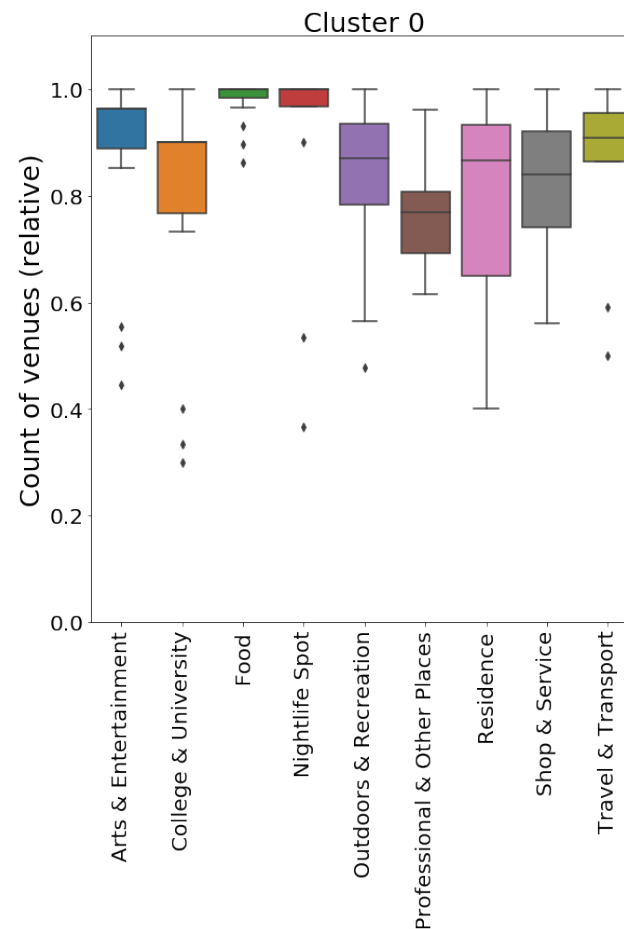
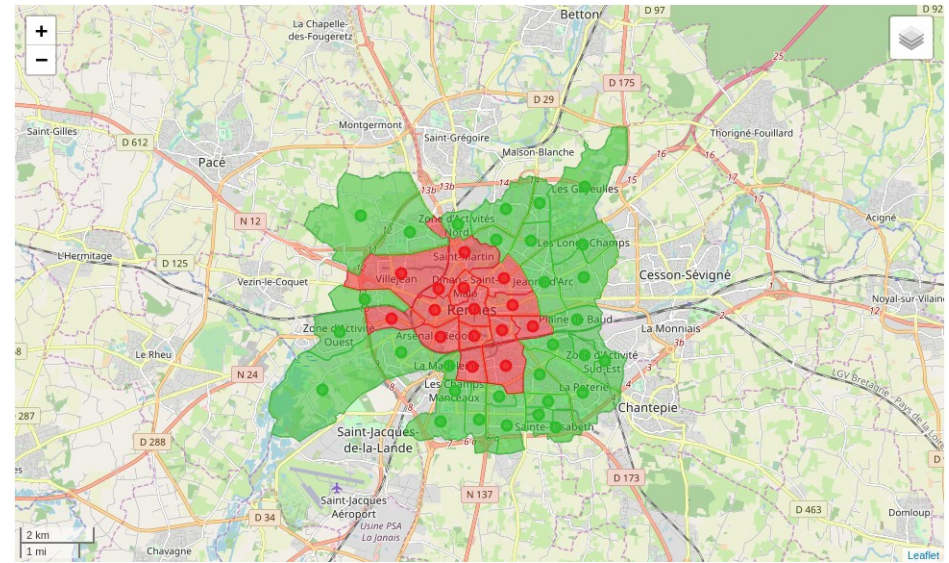
Normalization and cleaning

- 'Event' category is dropped because it is close to non-existent
- MinMaxScaler is used to improve readability between clusters, the number of venue is defined from 0 to 1



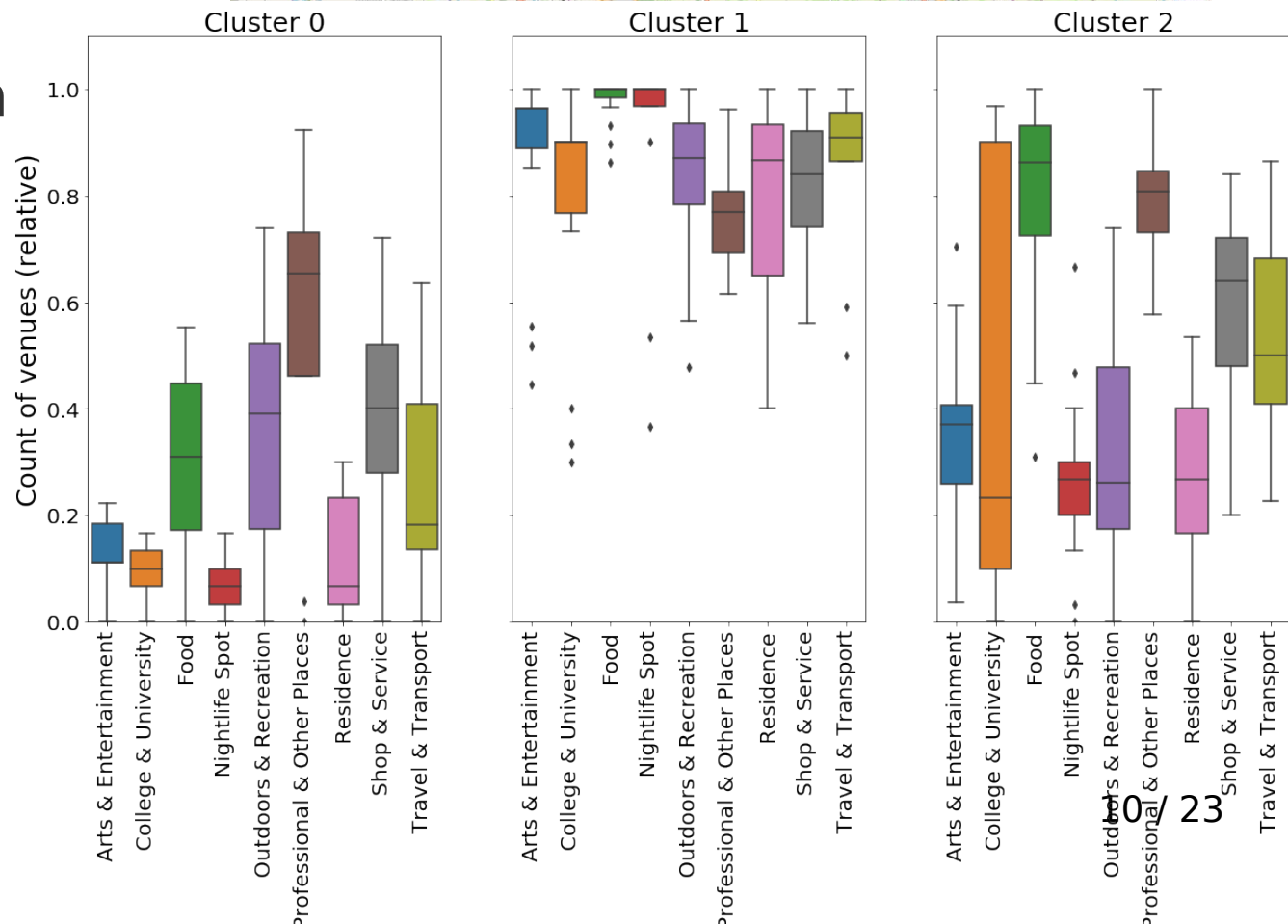
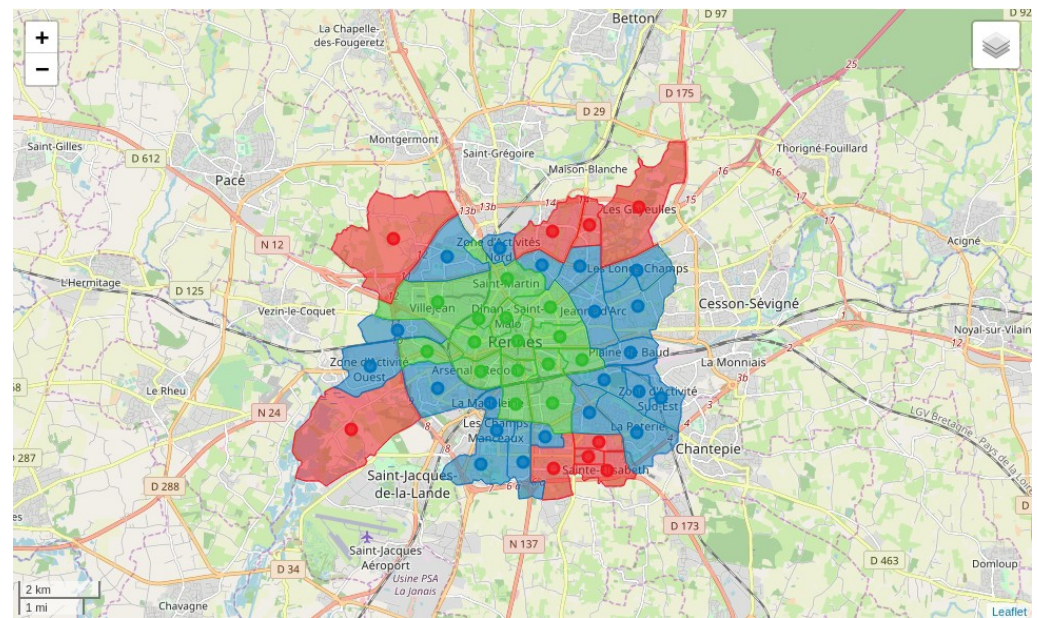
Clustering

- Tried with different number of clusters :
- 2 clusters only separate the city center and less dense areas



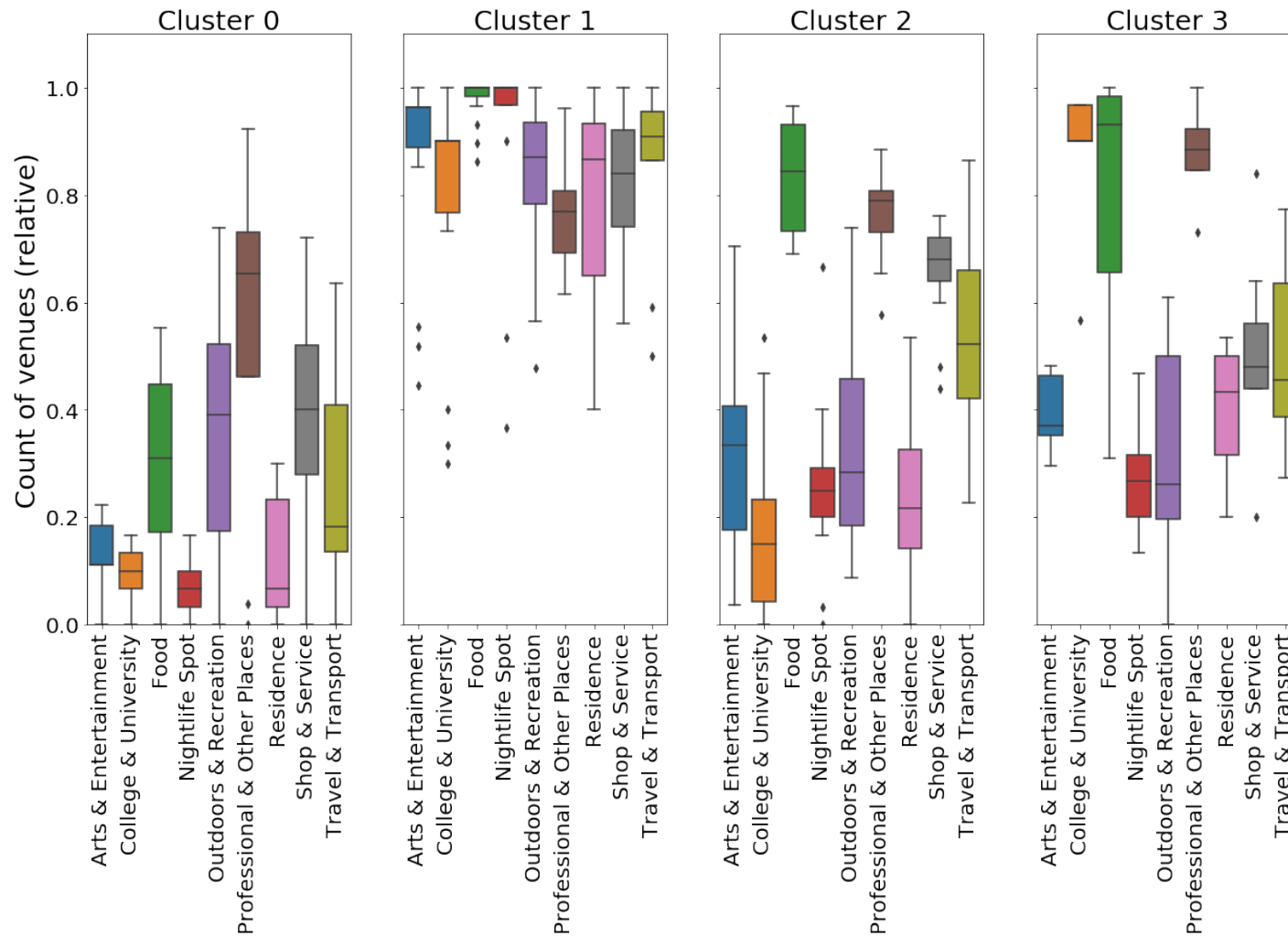
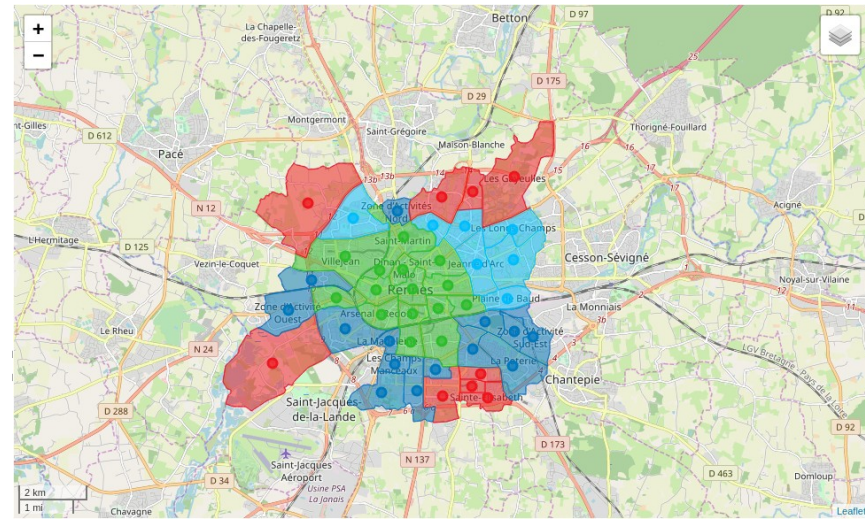
Clustering

- Tried with different number of clusters :
- 3 clusters is better but in-cluster variation is still high



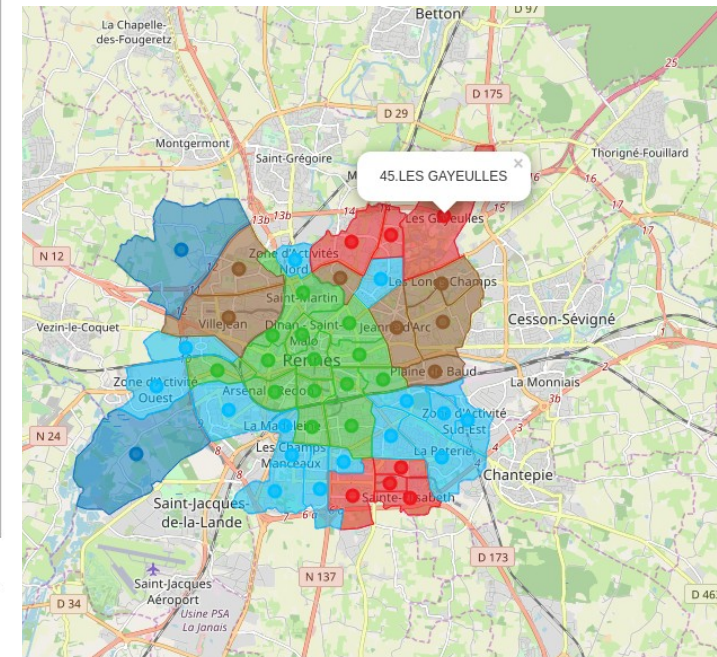
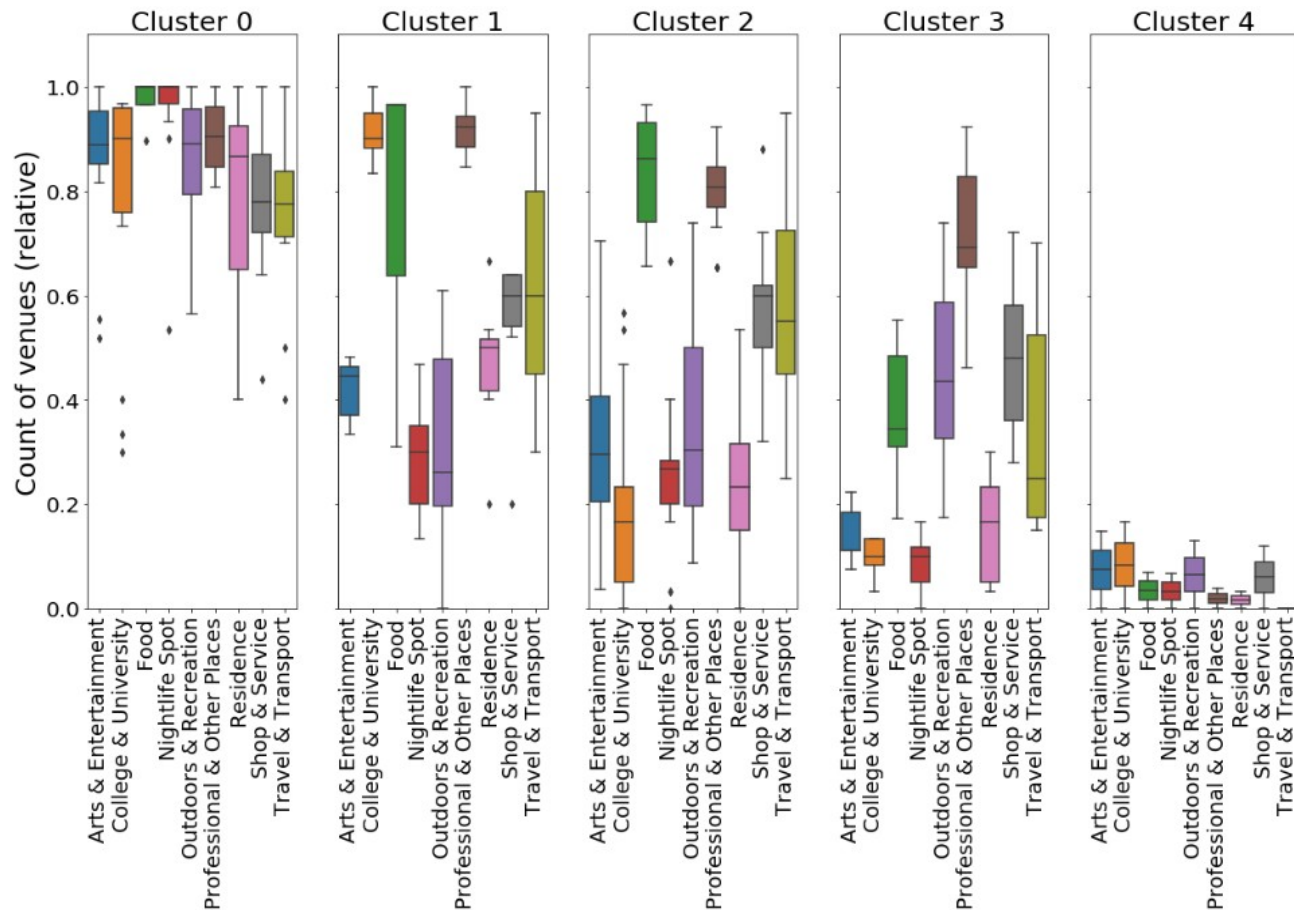
Clustering

- Tried with different number of clusters
- 4 clusters : still not enough



Clustering

- Tried with different number of clusters :
 - 5 clusters allows for a good separation of neighborhoods with low number of venues



Recommender System

- We already have all data needed to build the recommender system, all venues categories being noted from 0 to 1.
- We only need to build 'typical' user-profiles with a rating for each one of the remaining 9 categories

```
# Let's define "typical" user profile
user_profiles = {
    'student' : [0.01, 0.9, 0.01, 0.9, 0.9, 0.01, 0.01, 0.1, 0.01],
    'family' : [0.1, 0.01, 0.9, 0.01, 0.9, 0.1, 0.1, 0.9, 0.1],
    'young_adult' : [0.9, 0.01, 0.9, 0.9, 0.9, 0.9, 0.9, 0.01, 0.01, 0.01]
}
```

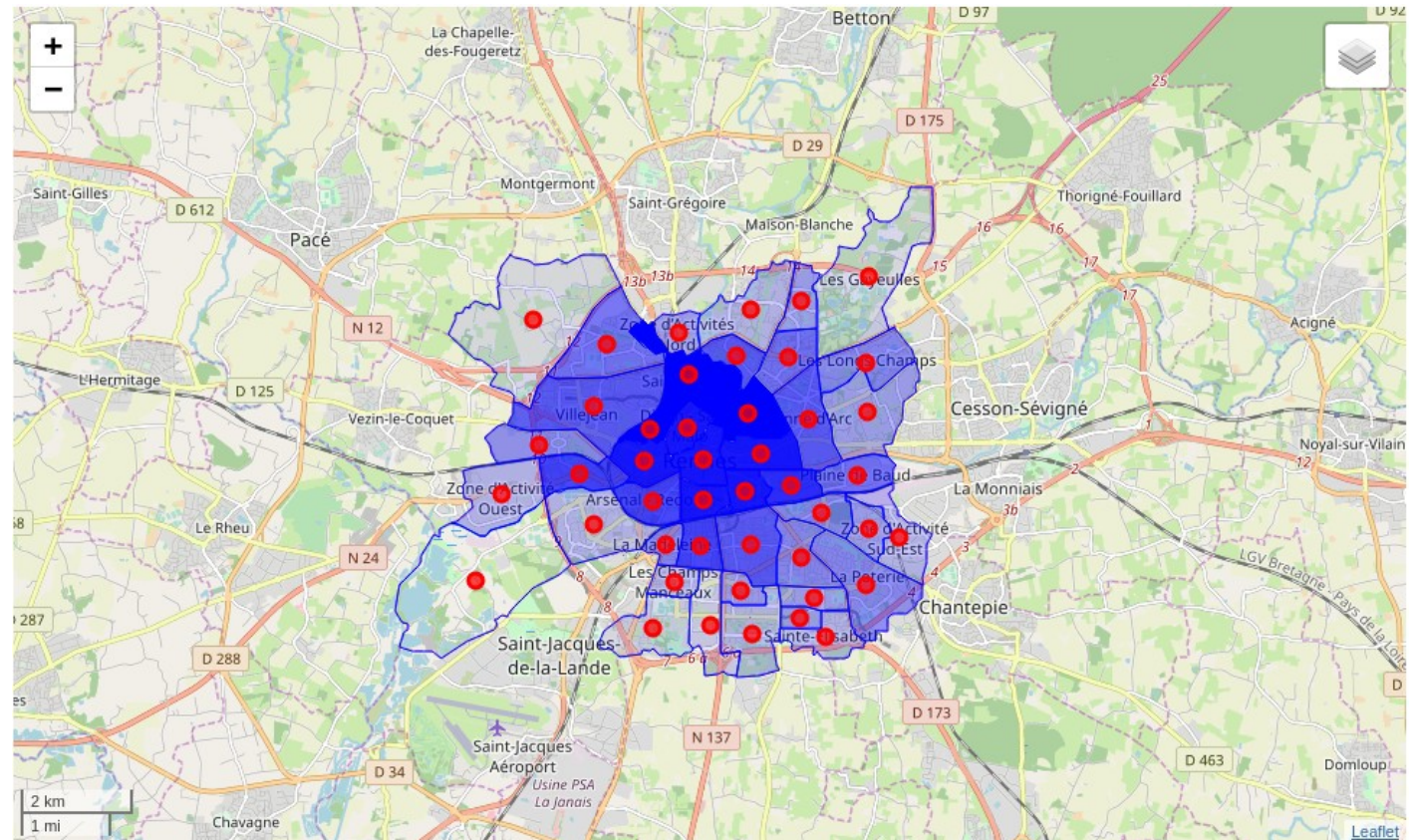

Recommender System

- We can then multiply this profile with the neighborhoods categories dataframe and extract each neighborhood rating.

	ID	Name	rating
0	1	SAINTE-ELISABETH	0.215974
1	2	TORIGNE	0.313912
2	3	LE LANDREL	0.307049
3	4	BREQUIGNY	0.115591
4	5	ITALIE	0.187698

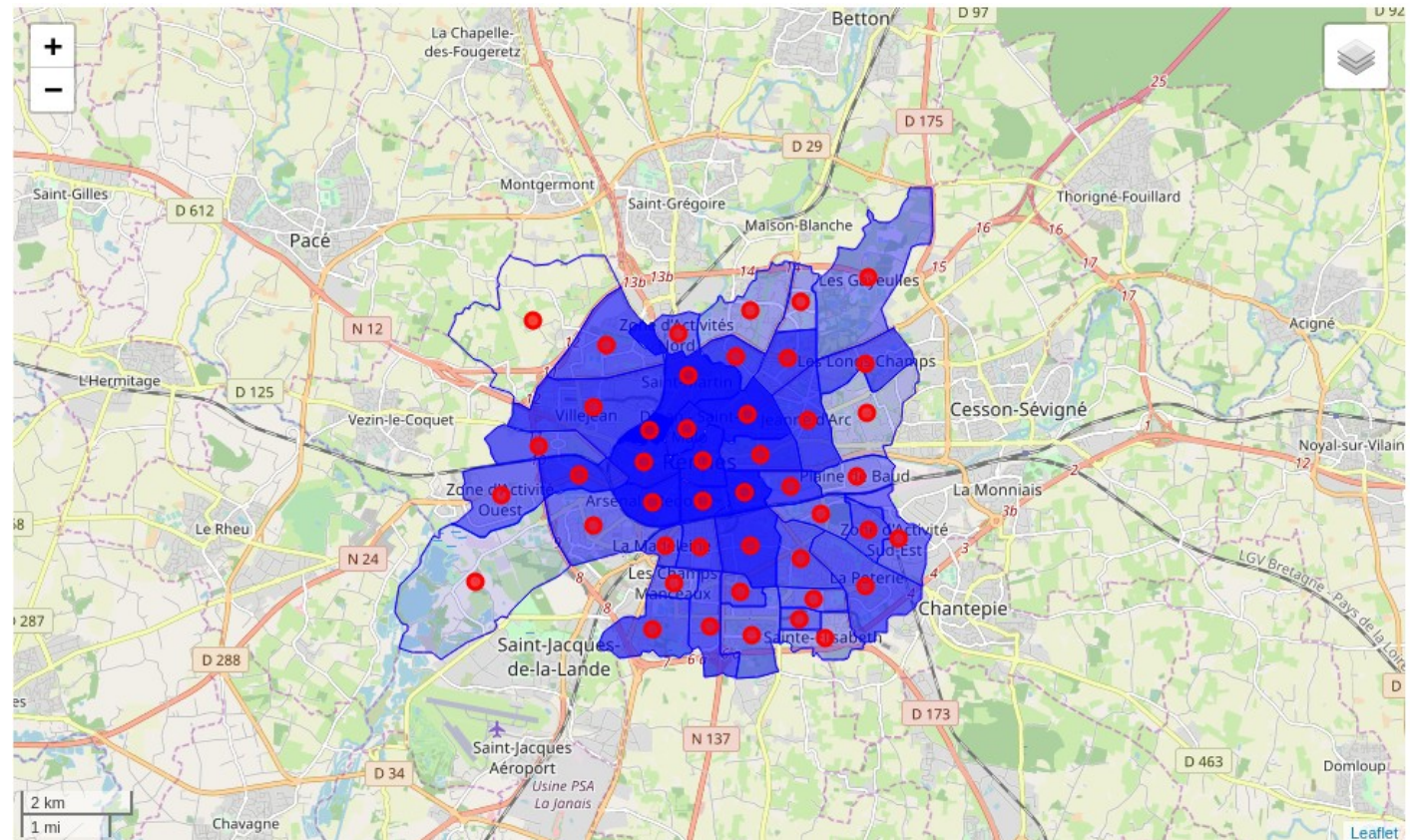
Recommender System

- We can build maps based on those ratings.
- Student :



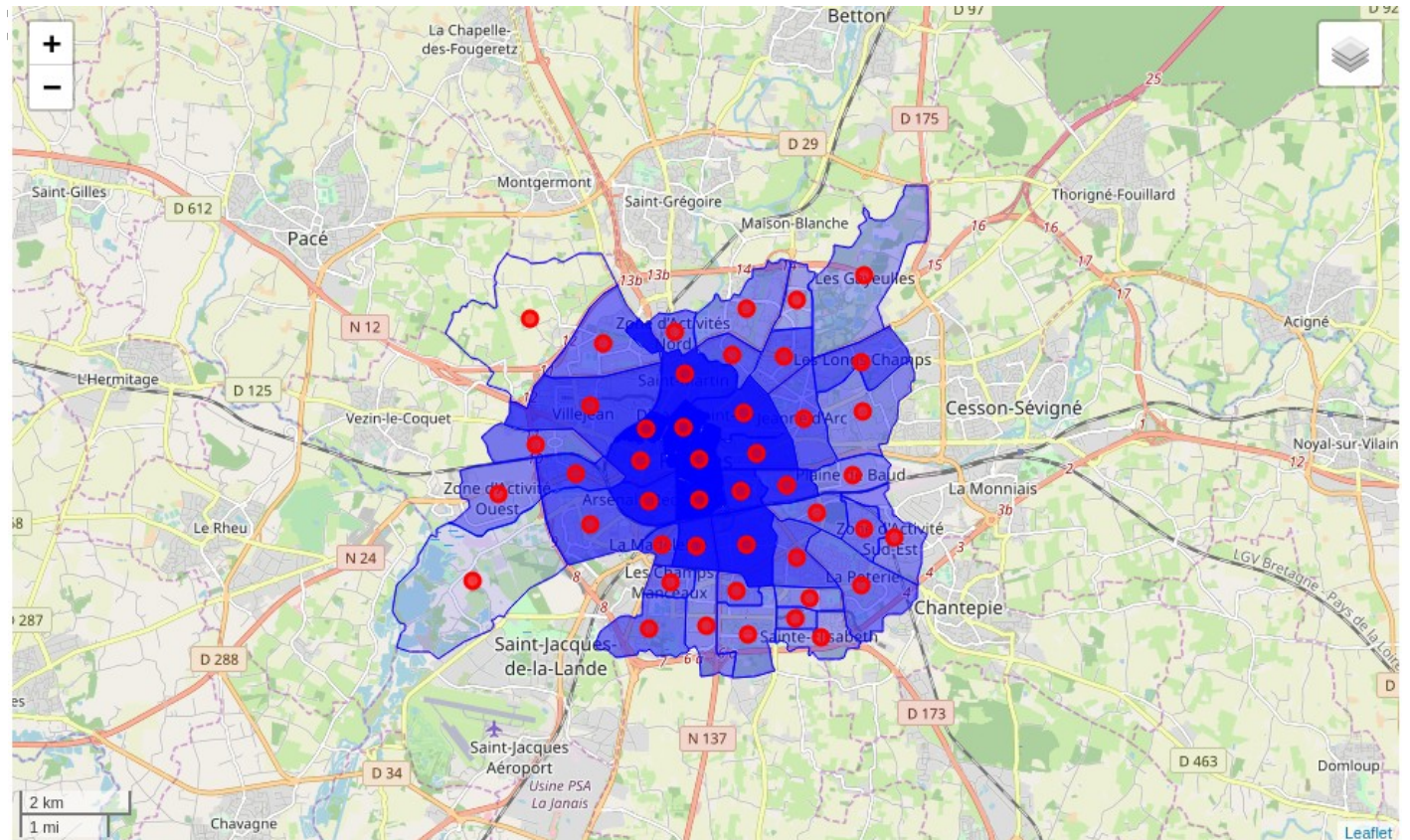
Recommender System

- We can build maps based on those ratings.
- Family :



Recommender System

- We can build maps based on those ratings.
- Young Adult



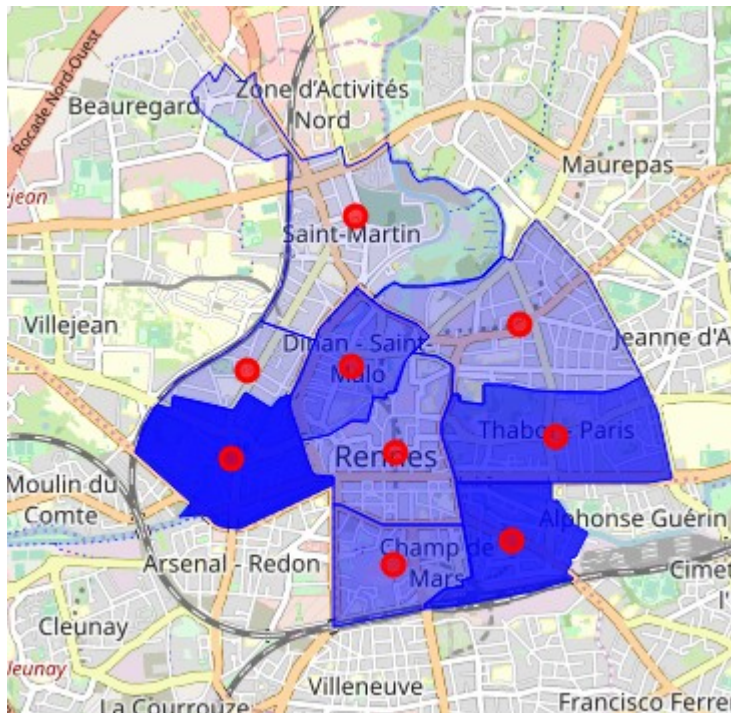


Recommender System

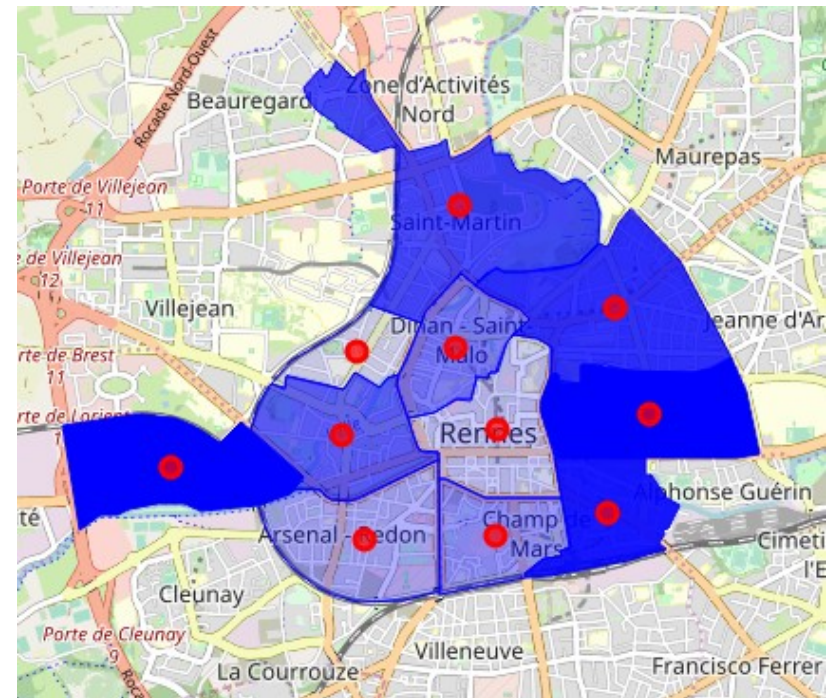
- All those maps are very similar, small differences in ratings are difficult to see since all recommended neighborhoods are close to each others.
- If we map only the neighborhoods with a rating > 0.8 we can remap for 'student' and 'family' profiles

Recommender System

- Student :



- Family :



- Differences are more visible now, and we can sort the neighborhoods based on their ratings

Recommender System

- rank column = student_rank

ID		Name	Cluster	rank	family_rank	adult_rank
38	39	NORD - SAINT-MARTIN	1	1	7	6
29	30	LA TOUCHE	1	2	1	3
31	32	FOUGERES - SEVIGNE	1	3	8	9
27	28	CENTRE	1	4	2	2
18	19	COLOMBIER - CHAMP DE MARS	1	5	4	4
30	31	DINAN - SAINT-MALO	1	6	5	1
25	26	THABOR - PARIS	1	7	10	8
20	21	SAINT-HELIER	1	8	9	10
26	27	BOURG L'EVEESQUE	1	9	6	7



Result

- City center is always recommended, no matter the profile
 - Due to the more dense venues
- Differences still exist in the neighborhoods ranking



Discussion

- Clustering and Recommendation works as expected
- But are limited by the Foursquare data :
 - Bias toward Food and Recreationnal places
 - Endpoint too limited or too wide
- Including other dataset will improve our analysis
 - Real estate past transaction
 - Venues data including rating



Conclusion

- Foursquare Data is limited
- Recommender System feasibility has been demonstrated
- More precise and relevant data would help improve the model