Finding an optimal housing location for a student in Amiens

Avgitidis Konstantinos

May 8, 2020

1. Introduction

1.1 Background

Housing is one of the most important things to consider when you make a decision to study abroad. Studying in a European Youth Capital winner can be very appealing to students who are not sure about where they want to study. With a cultural events programme combining music, cinema and theatre festivals, some of which happen in the streets, Amiens fully deserves its place as European Youth Capital, Although it is a historical city with the Cathedrale Notre Dame, classified by UNESCO (the 800th anniversary in 2020) there is also a wonderful youth dynamic in Amiens thanks to the university “La Citadelle”, which is located in the center of the town [1]. Also, University of Picardy Jules Verne is located just 4.5km from the city center and the Citadel of Amiens. For the purpose of this report we are going to use University of Picardy Jules Verne as our main university since “La Citadelle” is located right at the city’s center.

1.2 Problem

There are a lot of things a student should consider when searching for an optimal housing location. Data that might contribute to determining such location might include specific kind of venues that are nearby as well as the distances from the University’s campus and the City’s center. Since there are lots of venues in Amiens we will try to detect locations crowded with venues that appeal more towards students. We are also particularly interested in areas that are as close to the university as possible. So an area that satisfies both of the above statements is going to be considered as an optimal location for a student to live in. The main focus of this report is to find an optimal location, in Amiens, that fulfills each student’s needs with the data discussed above.

1.3 Interest

It is obvious that students would be very interested in an automated search engine that helps them decide where to stay in a city they want to study in, whilst having little knowledge on the city’s venues and locations. We will use our data science powers to generate a few most promissing neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by the students interested in living in Amiens.

2. Data acquisition and cleaning

2.1 Data requirements

Our main data requirements are the addresses of every neighborhood in Amiens and the coordinates of our two reference points (University Campus and City Centre). Furthermore, as previously discussed we are going to need the venues that exist in each neighborhood.

2.2 Data sources

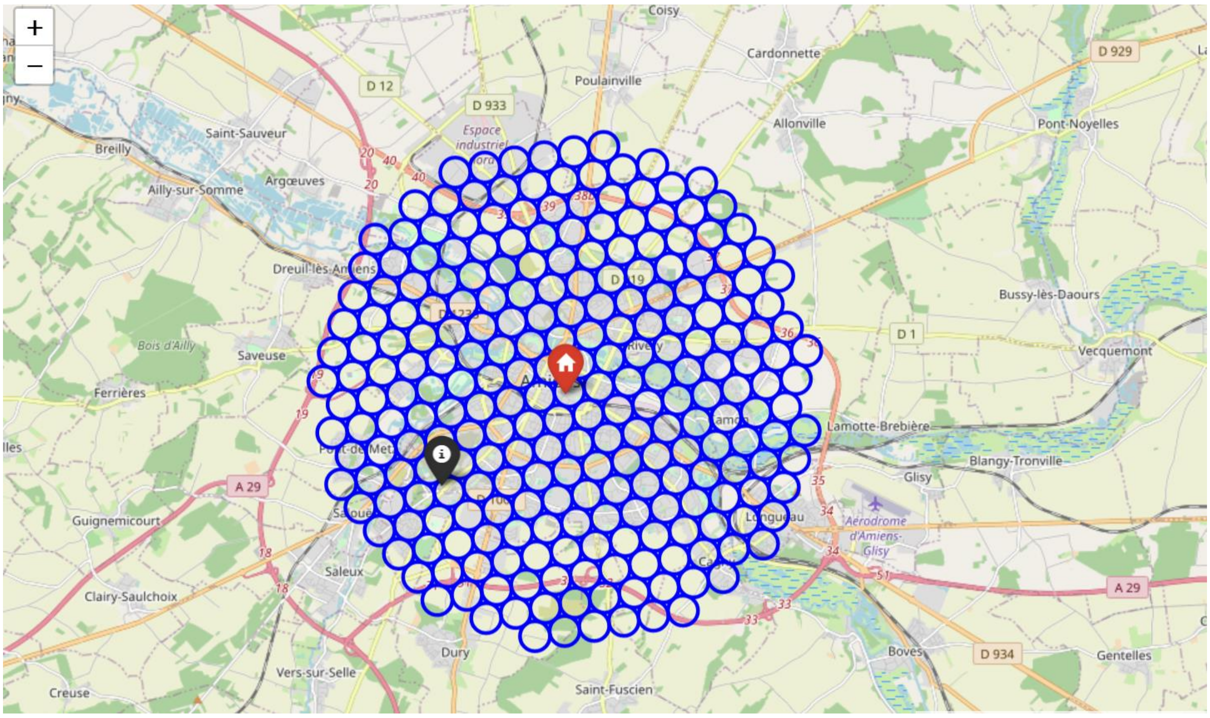
Most of the data we need can be extracted from APIs. To be more specific, we will use Google's Reverse Geocoding API to get the addresses for each neighborhood in Amiens and Google's Geocoding API to get the coordinates of our two reference points.

Centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Google Maps API reverse geocoding. We will create a grid of cells covering our area of interest which is approximately 5x5 kilometers centered around the City Center. Then venues around each candidate neighborhood can be found using Foursquare's Places API.

2.3 Data understanding and cleaning

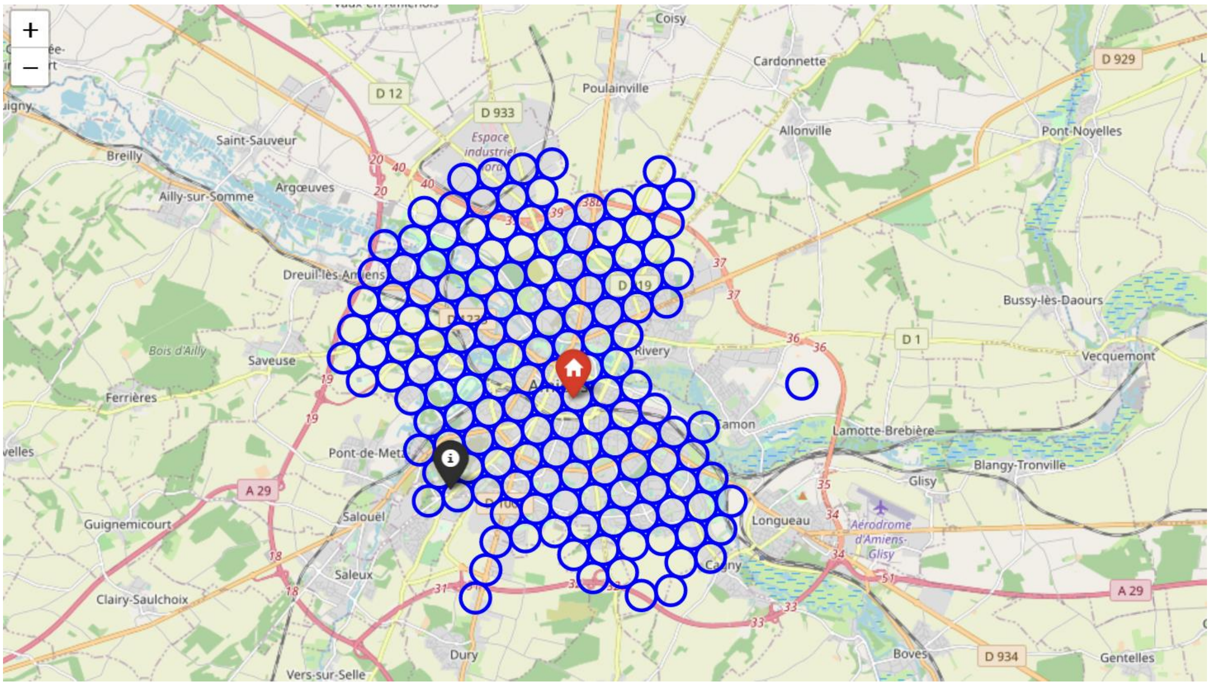
After collecting all the data that we have mentioned previously let us try to understand more about it. We have to check the type of each data and to learn more about the attributes and their names. Our main data consists of latitudes, longitudes, x and y coordinates on the Cartesian 2D system, Addresses, distances between venues and our 2 reference points and number of venues around each location. Apart from the Addresses, nothing had to be cleaned in order for us to use it. Our first data come from the Google Geocoding API and they are the coordinates of University Campus and City Centre, which we are going to use as reference points later on. We are using the City's Centre to create a grid of candidate areas, equally spaced, centered around city center and within 5km from Ctre Ville. Our neighborhoods will be defined as circular areas with a radius of 300 meters, so our neighborhood centers will be 600 meters apart. The total amount of generated neighborhoods is 251.

Image 1: Candidate neighborhoods 5km around City Centre.



Next thing we did, is to use Google's Reverse Geocoding API to get each candidate's neighborhood address and drop any address that does not contain the word Amiens in it (Since we only want addresses inside the City of Amiens).

Image 2: Candidate neighborhoods left after removing Addresses without the word Amiens.



We then proceed to combine these addresses, along with their coordinates in the Cartesian 2D system, their geographic coordinates and their distance from our two reference points into a single Pandas DataFrame. We want the coordinates in both systems in order to accurately calculate the distances we need in meters (not in latitude/longitude degrees).

Table 1: Dataframe of candidate neighborhood addresses and distances

	Address	Latitude	Longitude	X	Y	Distance from center	Distance from University
0	A29, 80090 Amiens	49.856752	2.318649	-410227.17343	5.600288e+06	4215.447782	4429.634107
1	Rue Wasse, 80090 Amiens	49.857657	2.326792	-409627.17343	5.600288e+06	4355.456348	4909.000315
2	31A Route d'Amiens, 80480 Dury	49.856321	2.272649	-413527.17343	5.600808e+06	4471.017781	2286.465879
3	Le montjoie, Rue Saint-Fuscien, 80090 Amiens	49.859952	2.305220	-411127.17343	5.600808e+06	3642.801120	3407.994908
4	D7, 80680 Amiens	49.860858	2.313363	-410527.17343	5.600808e+06	3659.234893	3874.410286
5	40 Rue Wasse, 80090 Amiens	49.862669	2.329652	-409327.17343	5.600808e+06	3973.663297	4895.335188
6	9 Rue de Montréal, 80090 Amiens	49.863574	2.337797	-408727.17343	5.600808e+06	4250.882261	5433.316885
7	91 Route d'Amiens, 80480 Dury	49.861334	2.275503	-413227.17343	5.601328e+06	3874.274126	1814.923891
8	23 Clos des Châtaigniers, 80090 Amiens	49.864058	2.299933	-411427.17343	5.601328e+06	3157.530681	2843.207000
9	3 Allée du Montjoie, 80090 Amiens	49.864964	2.308077	-410827.17343	5.601328e+06	3119.294792	3334.734144

After creating the above DataFrame, we used the Foursquare Places API along with each candidate neighborhood's coordinates to get info on existing venues around each area. For this particular report we have chosen these four categories as they are the ones that appeal more to students:

- Arts & Entertainment
- Food
- Nightlife Spot
- Outdoors & Recreation

We have selected the 50 nearest venues, including only the 4 categories mentioned before, on a radius of 400 meters around the center of each candidate neighborhood. Those venues were saved in a single matrix containing every venue on our hexagonal grid area. If a venue was closer than 300 meters from the center of a candidate neighborhood, it was added into a “local x venues” matrix, where x is the name of the venue’s category, in order to keep track of how many venues of each category are close to each neighborhood. So in total 4 “local x venues” matrices were created. Then we visualized each different category of venue with a colored dot on a folium map and add the 4 matrices into the dataframe we have previously created.

Image 3: Different categories of venues around the neighborhoods of Amiens.

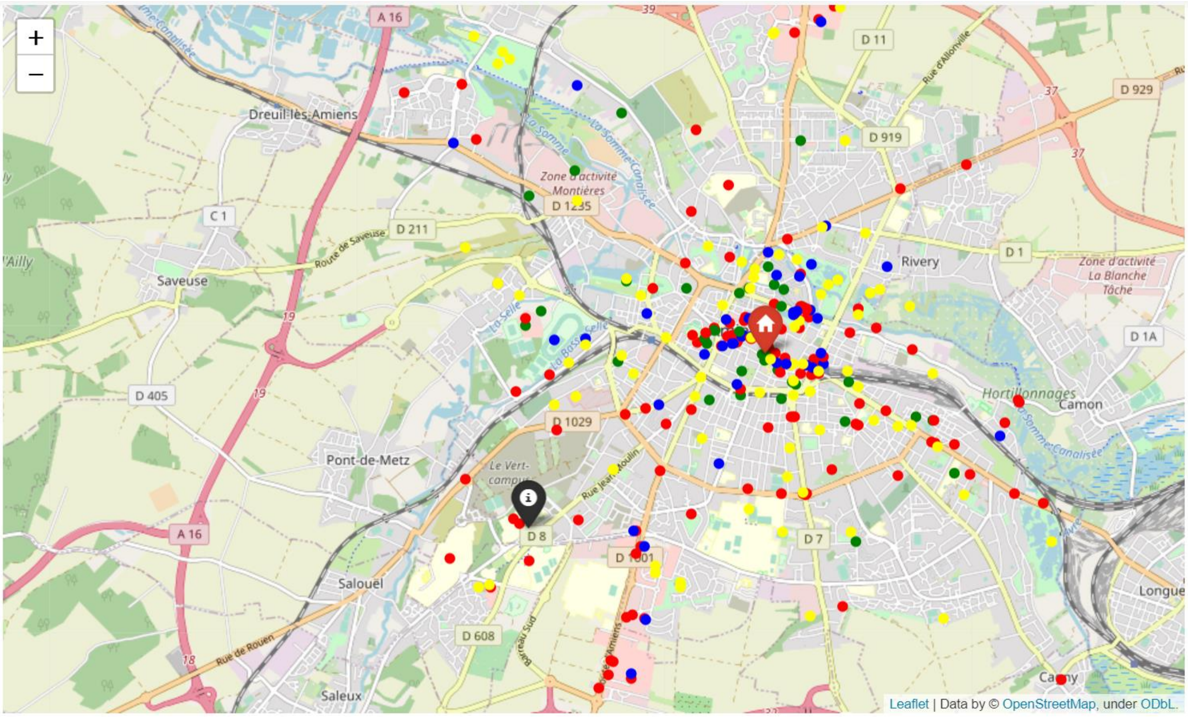
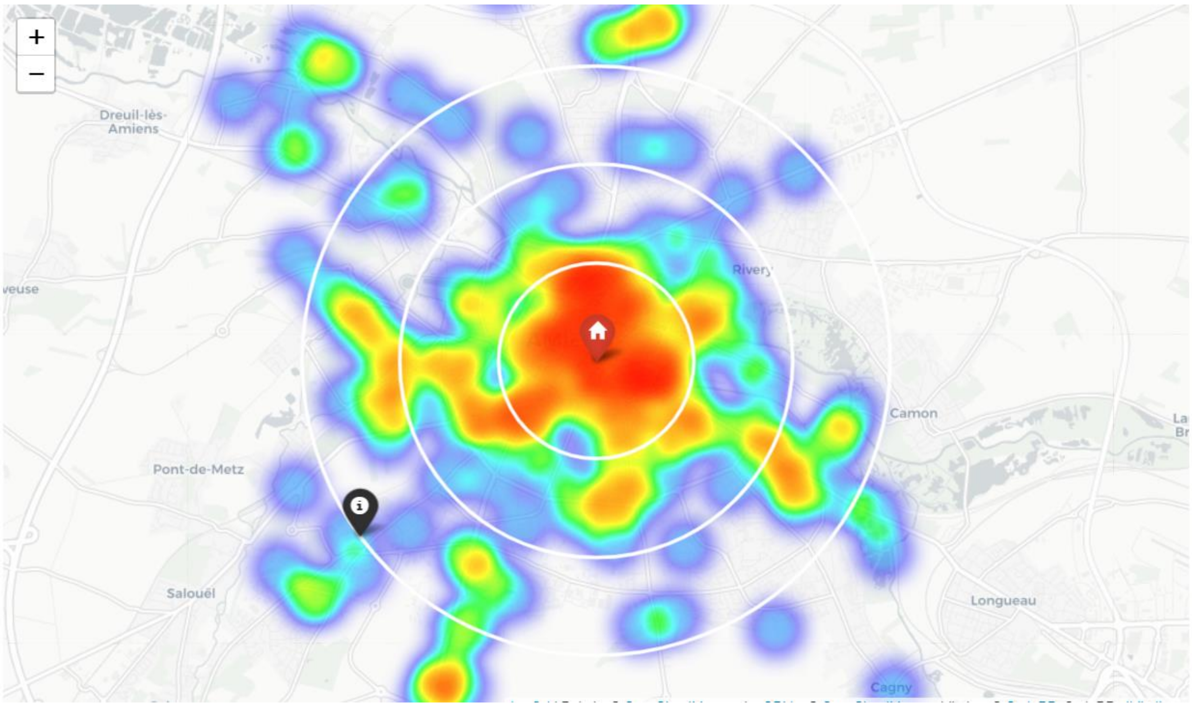


Table 2: Previous Dataframe along with number of venues for each category.

	Address	Latitude	Longitude	X	Y	Distance from center	Distance from University	Food Places in Area	Arts & Entertainment in Area	Nightlife Spots in Area	Outdoors & Recreation in Area
10	Espace mistral, 13 Rue de Redon, 80090 Amiens	49.865871	2.316222	-410227.17343	5.601328e+06	3195.309062	3857.081485	0	0	0	0
11	6 Rue Soufflot, 80090 Amiens	49.866776	2.324367	-409627.17343	5.601328e+06	3377.869151	4399.284414	0	0	0	1
12	722 Rue de Cagny, 80090 Amiens	49.867681	2.332512	-409027.17343	5.601328e+06	3651.027253	4954.828869	0	0	0	0
13	105 Marais de Cagny, 80090 Amiens	49.868586	2.340658	-408427.17343	5.601328e+06	3996.248241	5519.687935	0	0	0	0
14	116 Route d'Amiens, 80480 Dury	49.866347	2.278357	-412927.17343	5.601847e+06	3278.719262	1442.210543	4	0	1	0

We also created a folium heatmap, with 3 rings 1km each from the city center, in order to see the areas with the most venues around them. As we can see from the heatmap the City’s center is the where most venues are located, so we expect to see more neighborhoods towards the City center in our final results.

Image 4: Heatmap based on the amount of venues around each location.



After that, we generated more neighborhoods in-between all of the existing candidate neighborhoods in order to cover most of the addresses around the center of Amiens. In total we were able to generate 2261 neighborhood centers. We also filled out 6 columns, Food Venues nearby, Arts venues nearby, Nightlife venues nearby, Outdoor & Recreation venues nearby, 'Distance to the University and Distance to the Center, for every generated neighborhood center and added them into a new Pandas Dataframe.

Table 3: Dataframe containing all 2261 generated neighborhoods.

	Latitude	Longitude	X	Y	Food Venues nearby	Arts venues nearby	Nightlife venues nearby	Outdoor & Recreation venues nearby	Distance to the University	Distance to the Center
0	49.857751	2.314889	-410477.17343	5.600445e+06	0	0	0	0	4136.909576	4025.232913
1	49.857902	2.316246	-410377.17343	5.600445e+06	0	0	0	0	4214.276347	4037.635447
2	49.857680	2.307222	-411027.17343	5.600532e+06	0	0	0	0	3669.774539	3914.674913
3	49.857832	2.308579	-410927.17343	5.600532e+06	0	0	0	0	3742.104950	3913.397460
4	49.857983	2.309937	-410827.17343	5.600532e+06	0	0	0	0	3815.685227	3914.674913
5	49.858134	2.311294	-410727.17343	5.600532e+06	0	0	0	0	3890.444454	3918.504776
6	49.858285	2.312651	-410627.17343	5.600532e+06	0	0	0	0	3966.315966	3924.879575
7	49.858436	2.314008	-410527.17343	5.600532e+06	0	0	0	0	4043.237149	3933.786938
8	49.858587	2.315366	-410427.17343	5.600532e+06	0	0	0	0	4121.149225	3945.209713
9	49.858738	2.316723	-410327.17343	5.600532e+06	0	0	0	0	4199.997051	3959.126125

And if we sort them in ascending order by ‘Distance to the Center’ we can see that the areas closer to the city centered are packed with venues.

Table 4: Dataframe sorted by Distance to the University.

	Latitude	Longitude	X	Y	Food Venues nearby	Arts venues nearby	Nightlife venues nearby	Outdoor & Recreation venues nearby	Distance to the University	Distance to the Center
0	49.892101	2.300143	-410877.17343	5.604429e+06	15	5	6	3	3085.823928	52.584605
1	49.891950	2.298785	-410977.17343	5.604429e+06	15	5	6	3	2996.031545	52.584605
2	49.892785	2.299262	-410927.17343	5.604516e+06	15	5	6	3	3079.944012	70.319398
3	49.891265	2.299667	-410927.17343	5.604342e+06	15	5	6	3	3003.740871	102.885683
4	49.892936	2.300620	-410827.17343	5.604516e+06	15	5	6	3	3168.936637	122.248999
...	...	...	...	...	...	...	...	...	...	...
2256	49.863739	2.333560	-409027.17343	5.600878e+06	0	0	0	0	5131.912203	4041.459932
2257	49.862753	2.331726	-409177.17343	5.600792e+06	0	0	0	0	5035.927767	4051.076241
2258	49.860629	2.326701	-409577.17343	5.600619e+06	0	0	0	0	4771.011988	4057.937820
2259	49.859492	2.323509	-409827.17343	5.600532e+06	0	0	0	0	4606.571035	4065.055925
2260	49.861766	2.329892	-409327.17343	5.600705e+06	0	0	0	0	4944.149480	4068.051011

2261 rows × 10 columns

3. Data Analysis

As we stated previously, we want to find an optimal housing location for a student who wants to stay in Amiens considering 4 main factors: Number and type of venues around the area and distances from our 2 reference points (University and City Center).

3.1 Visualizing candidate neighborhoods

At first we decided to keep every candidate neighborhood which had at least one for every venue nearby and is less than 4km from both the University and City Center. So, we discarded every other candidate neighborhood and we were left with 160 areas down from 2261. We visualized these areas on a heatmap and marked them with a blue dot in order to be more distinct. The results were not shocking since we previously stated that we expect more optimal locations towards the city center.

Image 5: Areas with at least one of each venue nearby.

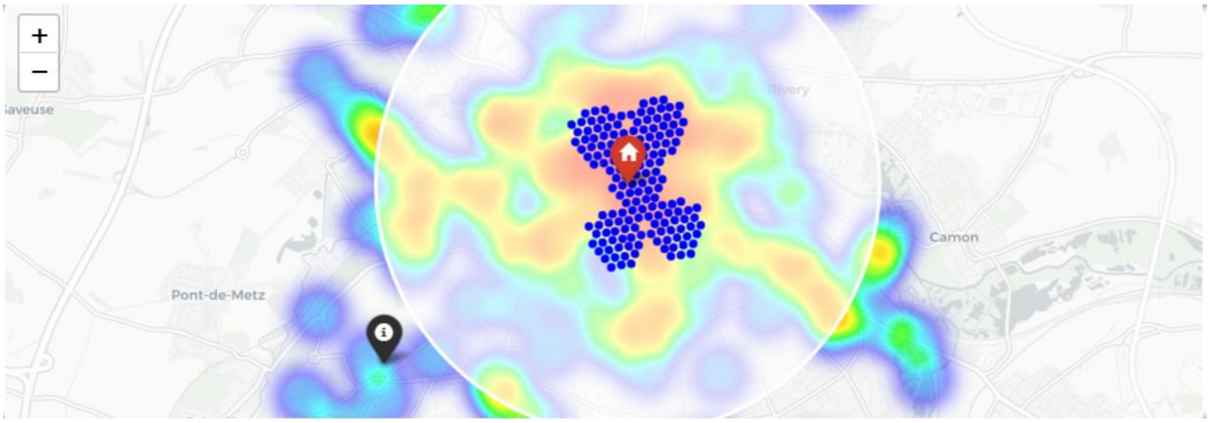
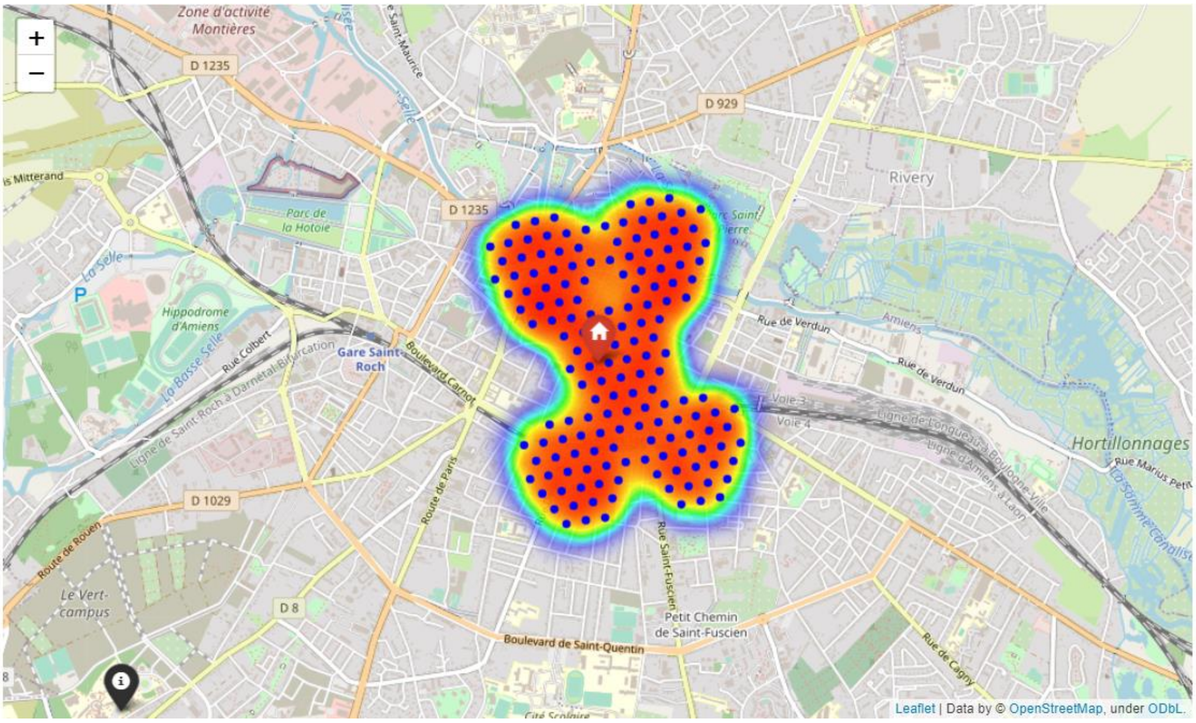


Image 6: Areas with more than one of each venue nearby portrait with a heatmap.



3.2 Clustering

Secondly, we decided to cluster these 161 neighborhoods onto 8 clusters. K-means clustering was used because it is one of the simplest and popular unsupervised machine learning algorithms. Also, since our data is unlabeled and we have the coordinates on the 2D Cartesian system, the choice for the algorithm was fairly easy.

Number 8 was chosen through trial and error for the number of clusters since we want to propose quite a few optimal housing locations, not very close to each other since we do not want them to overlap with each other too much. All the clustering was done automatically using scikit-learn's Kmeans class. We later on proceed to visualize all the data with 2 folium maps.

Image 7: Clusters generated around candidate neighborhoods.

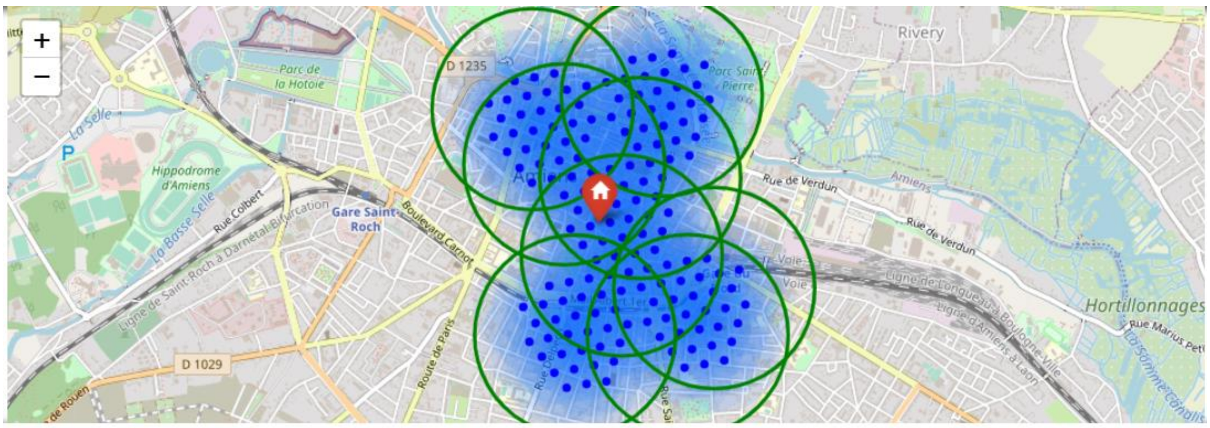
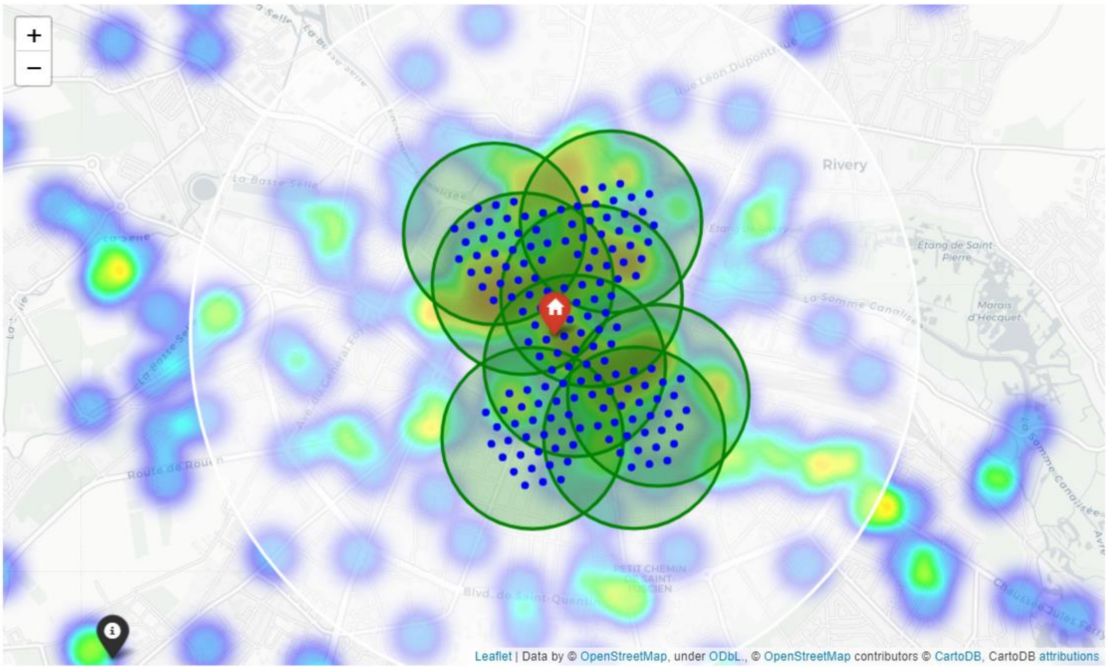


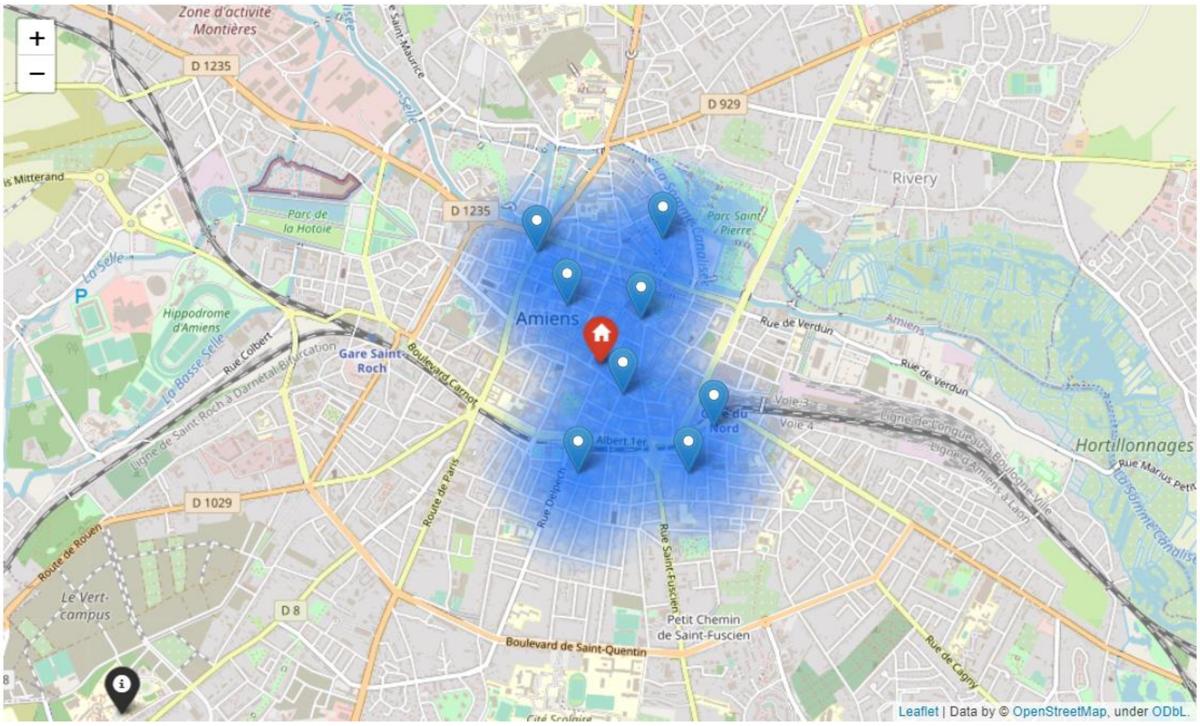
Image 8: Clusters around each neighborhoods portrait with a heatmap.



3.3 Visualizing optimal housing locations

Finally, we visualized the clusters' centroids on a single folium map and printed out all the addresses of our proposed optimal housing locations. We have marked each address with a blue marker and added markers for both our two reference points, with red being city center and black being the university.

Image 9: Optimal housing locations.

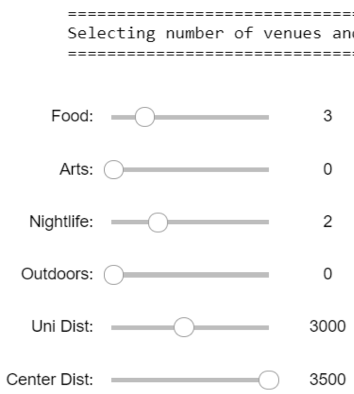


3.4 Personalization

After doing all things mentioned above, we wanted to personalize our notebook a little bit more in order for the proposed housing locations to be more fitted towards each student. That would give the student more options to play with and find his "perfect" housing location according to his needs without having to search manually for different venues around each location.

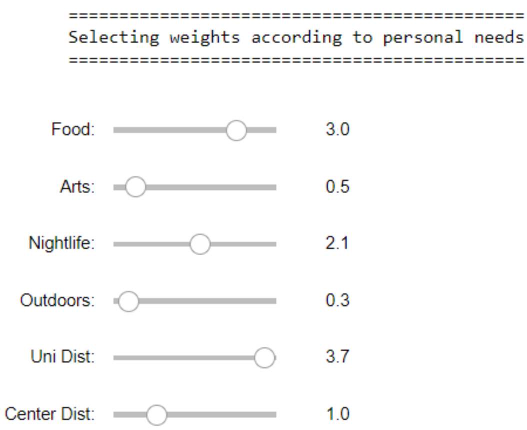
3.4.1 Selecting number of venues and distances

So the first thing we did is to let each students choose the number of each categorical venue around the area as well as the distances from the University and City center. We did that by adding some interactive sliders to help him choose the options that best suit his/her needs. For the sake of this report, we tested our final results with the values set below. Like the previous example, we excluded every neighborhood that did not meet all of the conditions selected above and were left with 45 neighborhoods out of 2261.



3.4.2 Weighted values

Secondly, we wanted to give students the option to prioritize one factor from another. Weighted values seemed to be the best practice for our case. In order to do that, we normalized each of the 6 final columns (Number of each venue and distances from the 2 reference points) to be between 0 and 1. The process was to divide each column with its max existing value so: column/max(column). In the case of distances we wanted for smaller distances to be leaning towards 1 and not the other way around so the formula was: 1 – column/max(column). The student was then given 6 sliders scaling from 0.0 up to 4.0, depending on how important each factor was to him/her. As we stated previously some values were chosen at random for the sake of this report.



3.5 Putting everything together

In order to get each weighted sum, we multiplied each of the 6 columns with their respective weight and sum each row into a single column. We then sorted the table by their weighted sums. Consequently, we were left with a dataframe where the top rows are the best housing locations in regards to each student's needs.

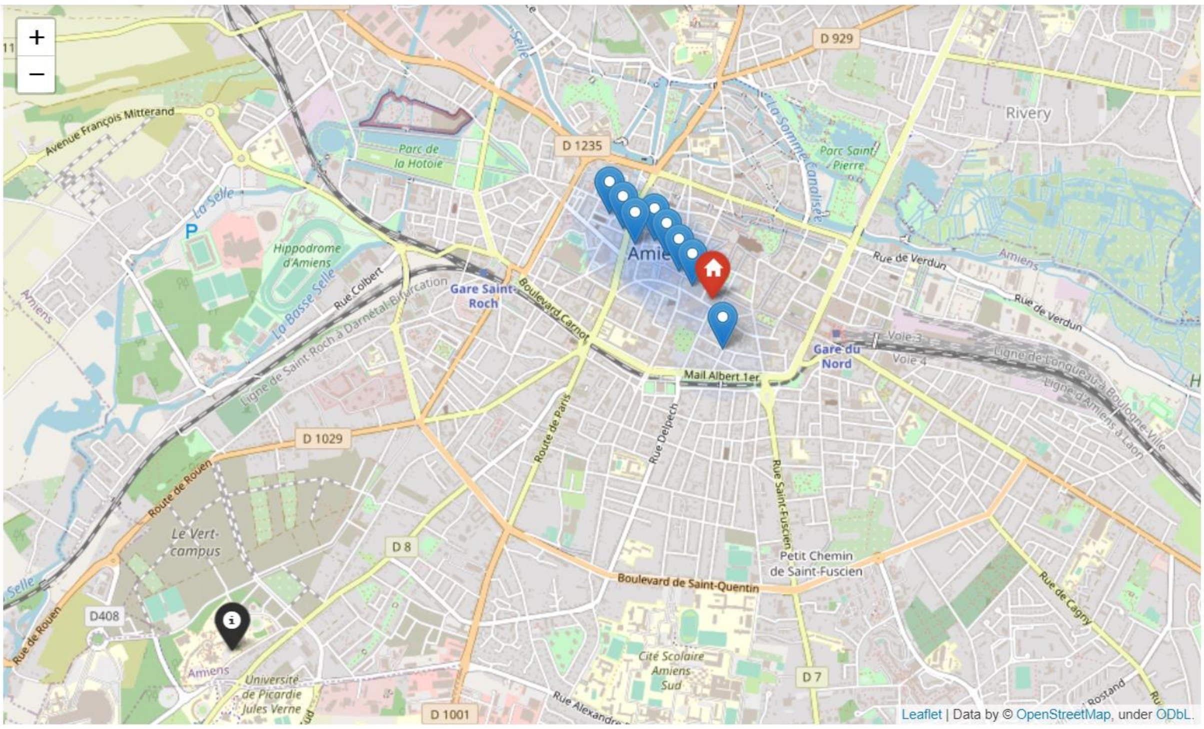
Table 5: Dataframe sorted by weighted sums

	Latitude	Longitude	X	Y	Food Venues nearby	Arts venues nearby	Nightlife venues nearby	Outdoor & Recreation venues nearby	Distance to the University	Distance to the Center	Weighted Sums
0	49.895903	2.292136	-411377.17343	5.604949e+06	3	3	2	1	2941.526055	675.161612	11573.808017
1	49.894686	2.295258	-411177.17343	5.604775e+06	3	3	2	1	2998.531011	414.106084	11523.670826
2	49.895219	2.293018	-411327.17343	5.604862e+06	3	3	2	1	2927.256349	577.636153	11423.484644
3	49.894002	2.296140	-411127.17343	5.604689e+06	3	3	2	1	2992.898666	315.125644	11403.850708
4	49.893318	2.297022	-411077.17343	5.604602e+06	15	5	6	3	2990.601381	217.081770	11343.306881
5	49.894535	2.293900	-411277.17343	5.604775e+06	3	3	2	1	2916.347704	481.127684	11286.614189
6	49.892634	2.297904	-411027.17343	5.604516e+06	15	5	6	3	2991.646841	122.248999	11252.342311
7	49.889746	2.300073	-410927.17343	5.604169e+06	15	5	6	3	2935.790067	276.090764	11199.514013
8	49.891950	2.298785	-410977.17343	5.604429e+06	15	5	6	3	2996.031545	52.584605	11198.901323
9	49.893851	2.294782	-411227.17343	5.604689e+06	3	3	2	1	2908.837935	386.398980	11164.099338

3.6 Visualizing final addresses

Finally, we took the addresses of the first 8 areas, since we want to propose 8 locations to the student and visualized them on a folium map, with markers pointing at each neighborhood and our 2 reference points.

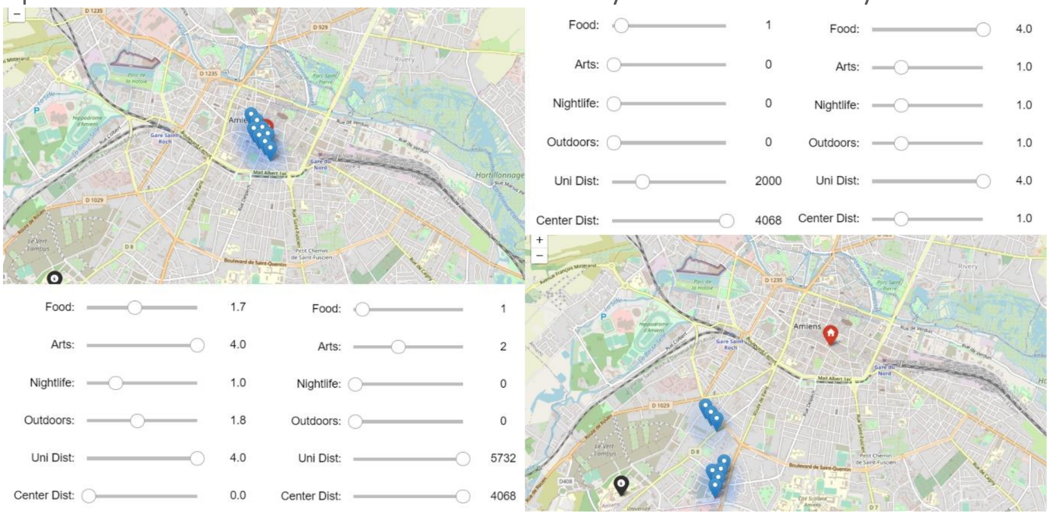
Image 10: Final optimal housing locations generated by the second model.



5. Results and conclusion

We can clearly see from the graphs above that, most of the housing locations are located towards the city center and they are close to each other. This is bound to be happens since in most towns the city center is always packed with more venues than the rest of the areas around it. Furthermore, trying different combination of weights and number of venues/distances generates interesting results. In the end, it all comes down to each student's priorities and needs. It is certain though, that if a student's requirements lean towards more venues around their housing location, the area is going to be located near the city center. If nonetheless, venues are not a point of reference for a student and the distance from the university is the main factor the housing locations can be more flexible.

Image 11: (Left Map Bottom Sliders): Results of second model while interested mainly in Art and Outdoor venues and not interested in distances. (Right Map Top Sliders): Optimal Areas under 2km from the university interested mainly in food venues.



6. Future directions

We were able to achieve interesting results with both clustering and normalized weight values. Both our models were based in the distances to our 2 reference points and the number of venues around each candidate neighborhood. However, there is a huge variance in things that may fulfil a student's needs. For example, one of the biggest factor in our opinion that was not taken into account is the student's budget. Using a specific ceiling price as a budget can eliminate a lot of candidate locations. In addition to that, distances may not be as important when you take in mind the number of and distance to bus stops around each housing location. Furthermore, we can deepen the models even more by taking into consideration each student's hobbies or activities. We can also use metadata using Google Forms so as to adapt very specific needs in our models. To sum up, all these data are obviously more difficult to extract and quantify, but if optimized, could bring significant improvements to our final model.

[1]: https://www.youthforum.org/european-youth-capital-winner-amiens-2020