Coursera Capstone Project Report
Location of a new car park in Grenoble

Business Problem

The city of Grenoble (France) would like to build a new car park in the city. The main purpose of this new car park will be to help people to park their cars when they go shopping or dining down town. Actually, the need of this car park comes from shopkeepers and restaurants who are complaining that less and less people go shopping or dining down town because it is too difficult to find a car park.

The city of Grenoble has asked a data scientist to help them finding the best place to build an underground car park. They would like to identify the best location for the car park taking into account the density of the different neighborhoods in terms of shops, bars and restaurants. The city would like to build the car park in the venue-dense neighborhood that is currently the furthest away from all existing car parks.

Stakeholders

The main stakeholders of this study are:

- The city of Grenoble because it their duty to make sure the different venues of the city are accessible by car
- Owners of the venues because their business can be affected by the difficulty for customers to find a parking lot close to them

Data

To solve the above-described problem we are going to use 2 different datasets:

Dataset 1: Foursquare data

We are going to use Foursquare data to find all the popular venues of the city. This dataset will give us the name, category, latitude and longitude of each venue within a given radius around the city. This dataset will allow us to perform a density based clustering of Grenoble's venues.

- **Example of dataset 1:**

	name	categories	lat	lng
0	Jardin du thé	Tea Room	45.188788	5.727536
1	Place Victor Hugo	Plaza	45.188994	5.724607
2	L'Ardoise	French Restaurant	45.190558	5.725346
3	Okko Hotels Grenoble Jardin Hoche	Resort	45.184859	5.726299
4	Amorino	Dessert Shop	45.191037	5.727638

Dataset 2: City of Grenoble data

The city of Grenoble offers different datasets on their web site. One of them is particularly interesting for our project: a dataset containing the name and geolocation of all parking lots of the city. This dataset will allow us to compute the distance from each parking lot to the different density based clusters previously identified.

- **Example of dataset 2:**

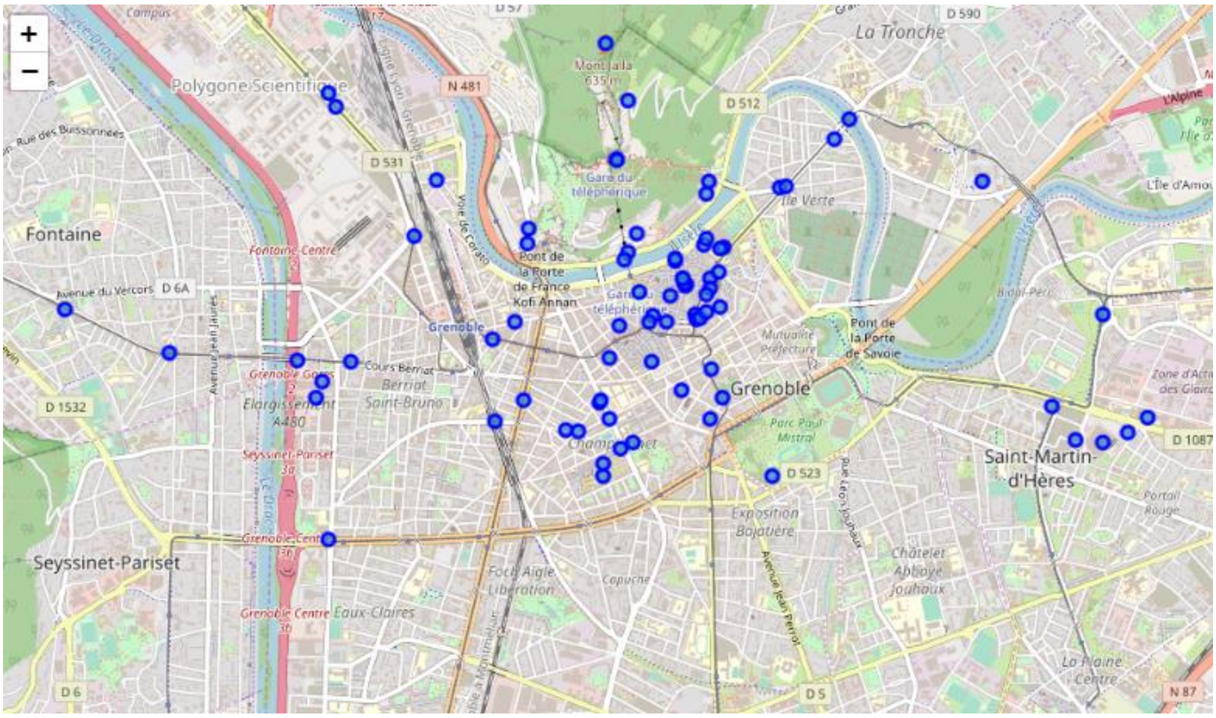
http://data.metropolegrenoble.fr/ckan/dataset/parkings-de-grenoble/resource/a6919f90-4c38-4ee0-a4ec-403db77f5a4b

_id	CODE	LIBELLE	ADRES...	TYPE	TOTAL	type	id	lon	lat
1	EFF_PK...	BERRIA...	RUE DE...	PKG	578	PKG	EFF_PK...	5.716033	45.18878
2	EFF_PK...	EUROP...	2, AVEN...	PKG	1055	PKG	EFF_PK...	5.712008	45.191959
3	PVP_PK...	HOCHE	RUE FR...	PKG	677	PKG	PVP_PK...	5.726327	45.185748
4	EFF_PK...	LAFAYE...	RUE RA...	PKG	311	PKG	EFF_PK...	5.729321	45.190778
5	PVP_PK...	MUSÉE	50, AVE...	PKG	770	PKG	PVP_PK...	5.732239	45.194399
6	PVP_PK...	PHILIPP...	PLACE ...	PKG	519	PKG	PVP_PK...	5.725235	45.191835
7	EFF_PK...	TERRAY	33 BIS, ...	PKG	106	PKG	EFF_PK...	5.712443	45.185621
8	PVP_PK...	VERDUN	PLACE ...	PKG	110	PKG	PVP_PK...	5.732018	45.18897
9	QPA_PK...	CHAVANT	17, BD ...	PKG	394	PKG	QPA_PK...	5.731463	45.185612
10	EFF_PK...	IRVOY	RUE IRV...	PKG	200	PKG	EFF_PK...	5.713873	45.181478
11	SPR_PK...	CATANE	RUE AM...	PKG	490	PKG	SPR_PK...	5.70503	45.181035
12	PVP_PK...	HÔTEL ...	VALMY (...	PKG	200	PKG	PVP_PK...	5.741289	45.188365
13	EFF_PK...	LE CÈDRE	RUE AN...	PKG	77	PKG	EFF_PK...	5.710835	45.18898
14	EFF_PK...	GARE L...	34 AVEN...	PKG	395	PKG	EFF_PK...	5.711838	45.1935
15	PVP_PK...	ENCLO...	PLACE ...	PKG	130	PKG	PVP_PK...	5.728358	45.188401
16	PVP_PK...	ENCLO...	RUE EM...	PKG	200	PKG	PVP_PK...	5.713614	45.194345

Methodology

We are trying to define the best place to build an underground parking lot in the city of Grenoble. The center of the city currently suffers from desertification because people have great difficulties to parks their cars because of a lack of parking lots. Consequently, they prefer to go shopping or partying outside of the city. The purpose of this parking lot is to allow people to park their cars when they go down town.

The first thing we are going to do is to identify the areas of the city that are the densest in terms of popular venues. These areas are where the need of parking lots is the most critical.
To do so we use the Foursquare API to get the top 100 venues within a radius of 3 km around Grenoble. We also use the Folium library to create a map of Grenoble and add markers for venues to the map.



Then we use a well-known density based clustering algorithm called DBSCAN. The purpose of this algorithm is to locate the areas that are the densest in terms of presence of popular venues and to group the venues in clusters. This technique has several advantages compared to other classical clustering algorithms like Kmeans for instance:

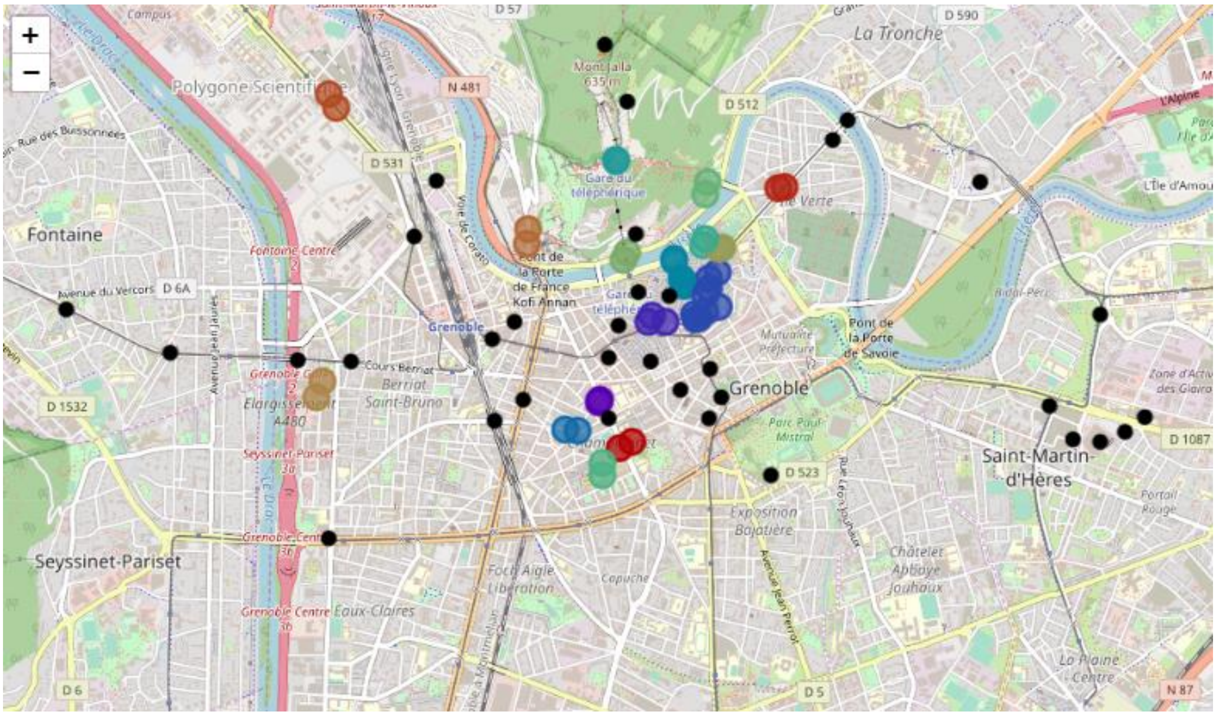
- It is able to generate clusters of arbitrary shapes
- It is not influenced by the presence of outliers

The parameters we use to run the DBSCAN algorithm are the following:

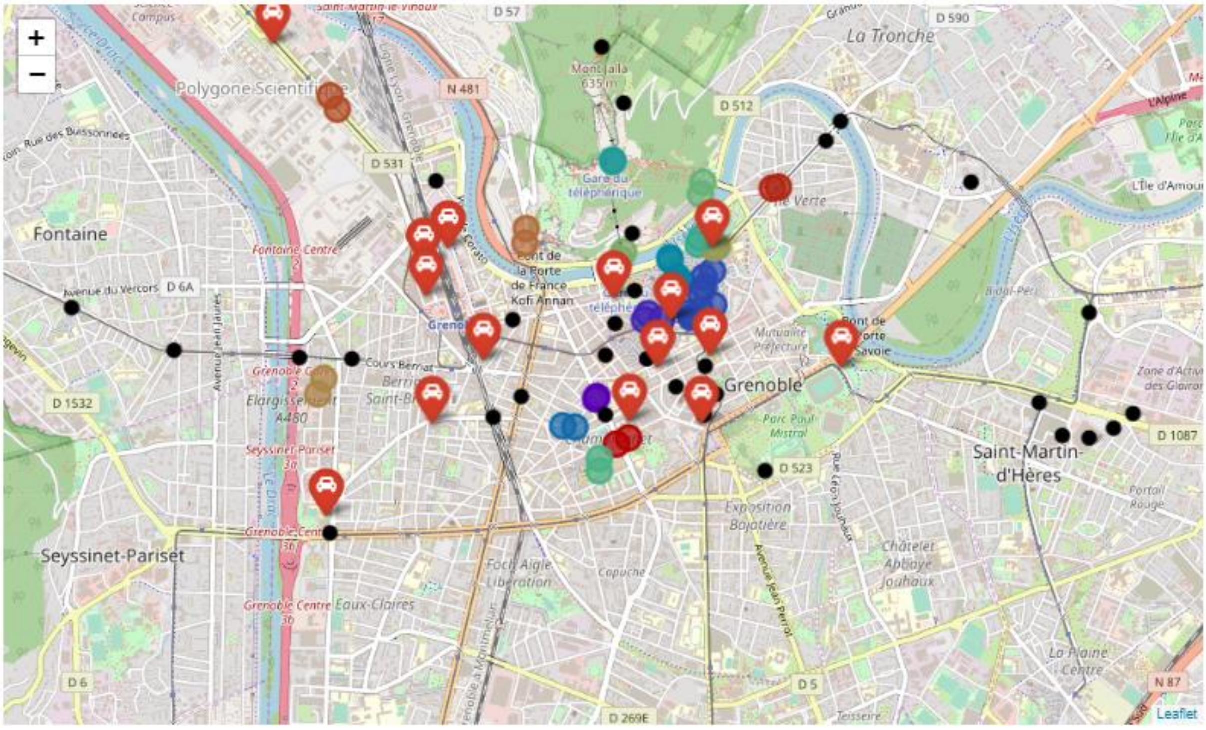
- Epsilon = 0.001
- minimumSamples=2

These parameters correspond respectively to the radius and the minimum number of venues within radius to consider a venue as a core point of a cluster.

On the below map, density-based clusters are represented by color circles whereas outliers are represented by black dots. Each different cluster of popular venues has a different color. For the rest of the study we are going to ignore the outliers as they are not relevant to determine the best place to build a new parking lot. The places where the need of parking lot is the most pregnant is close to the areas of high density in terms of popular venues.



Now this is time to use our second dataset from the city of Grenoble. In this dataset we have the location of each currently existing parking lot of the city of Grenoble. We use a different icon to represent the parking lot on the folium map in order to clearly distinguish them from the clusters and outliers.



At this stage we have gathered 2 important pieces of information:

- The location of the density-based clusters of popular venues in Grenoble
- The location of all currently existing parking lots in Grenoble

The next step will be to compute the distances between all clusters and all parking lots. Indeed, the clusters that are the furthest away from all existing parking lots will be good candidates for the construction of a new underground parking lot.

To calculate the distance between a venue and a parking lot, we naturally use the Euclidean distance. We use a triple loop over clusters, venues and parking lots to generate the below matrix. This matrix contains the list of all distances from all venues to all parking lots grouped by clusters.

	0	1	2	3	4
LIBELLE					
ESTACADE	[0.010988811909196714, 0.010256769090738317]	[0.008190057029457626, 0.00821529824562055]	[0.01182214704948849, 0.012701893328585687, 0....	[0.014684040527354189, 0.016051330807739503, 0...	[0.006520639781295689, 0.007292746094546028]
GARES - EUROPOLE	[0.01595701498417506, 0.01528398021329736]	[0.013037646612879865, 0.01303861434517998]	[0.01565691119585124, 0.01661369796391712, 0.0...	[0.01854080049910506, 0.019571259534978076, 0....	[0.011622396536212594, 0.012369489922641672]
VICTOR HUGO	[0.0008891416155753296, 0.0015209362324787783]	[0.00257050064217306, 0.0025761912001632573]	[0.0054488478551531876, 0.005499529585897249, ...	[0.006819070629408433, 0.008823957819029251, 0...	[0.004677954163752856 6, 0.003848416964458193]
LAFAYETTE	[0.006645740356982084, 0.007363580507160417]	[0.006665675849305872, 0.006538287412896047]	[0.0017030547347651076, 0.0007423400511354474,...	[0.001255626053944763, 0.0030522353573527003, ...	[0.009317354413987396, 0.00869131042475026]
NOTRE DAME - MUSÉE	[0.011238068726585922, 0.011995607723742445]	[0.011229538320205507, 0.01109501101084856]	[0.005698748087748, 0.005157548602985889, 0.00...	[0.0036995423134355988, 0.0016884094763119457,...	[0.013839745781103637, 0.013265070794577637]

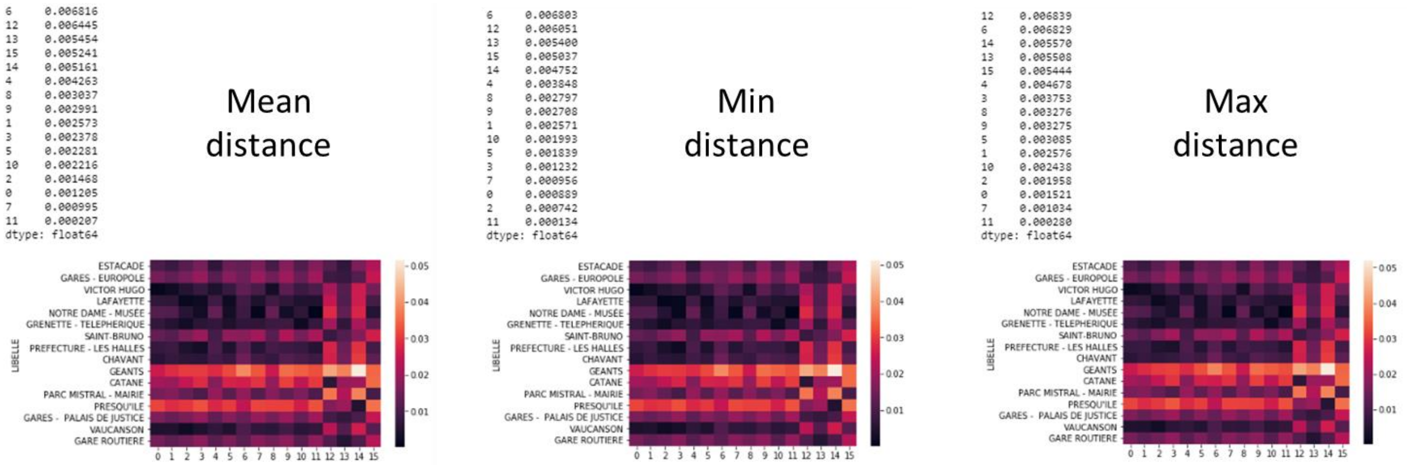
The next step is to calculate the distance between a cluster of venues and a parking lot. However, there is no obvious way to calculate this distance. We are going to use 3 different metrics to estimate it:

- Mean distance: the mean of the distances from all venues of a cluster to a given parking lot.
- Minimum distance: the minimum distance between one of the venues of the cluster and a given parking lot.
- Maximum distance: the maximum distance between one of the venues of the cluster and a given parking lot.

Technically, we use the applymap method to apply 3 different aggregation functions (mean, min and max) to each cell of the above matrix. We end up with 3 different matrixes (one for each aggregation function) like the below one.

	0	1	2	3	4	5	6	7	8	9	10	11	12
LIBELLE													
ESTACADE	0.010257	0.008190	0.011503	0.014684	0.006521	0.014029	0.013445	0.016225	0.009577	0.017388	0.010853	0.017202	0.011286
GARES - EUROPOLE	0.015284	0.013038	0.015401	0.018541	0.011622	0.017244	0.014765	0.019372	0.014678	0.020003	0.013778	0.020452	0.008302
VICTOR HUGO	0.000889	0.002571	0.005126	0.006734	0.003848	0.007655	0.012941	0.010033	0.002797	0.012300	0.008061	0.010445	0.021648
LAFAYETTE	0.006646	0.006538	0.000742	0.001256	0.008691	0.001839	0.008885	0.004185	0.008567	0.006519	0.004732	0.004628	0.024709
NOTRE DAME - MUSÉE	0.011238	0.011095	0.005158	0.001232	0.013265	0.002942	0.008237	0.000956	0.013214	0.002708	0.006277	0.000134	0.028221

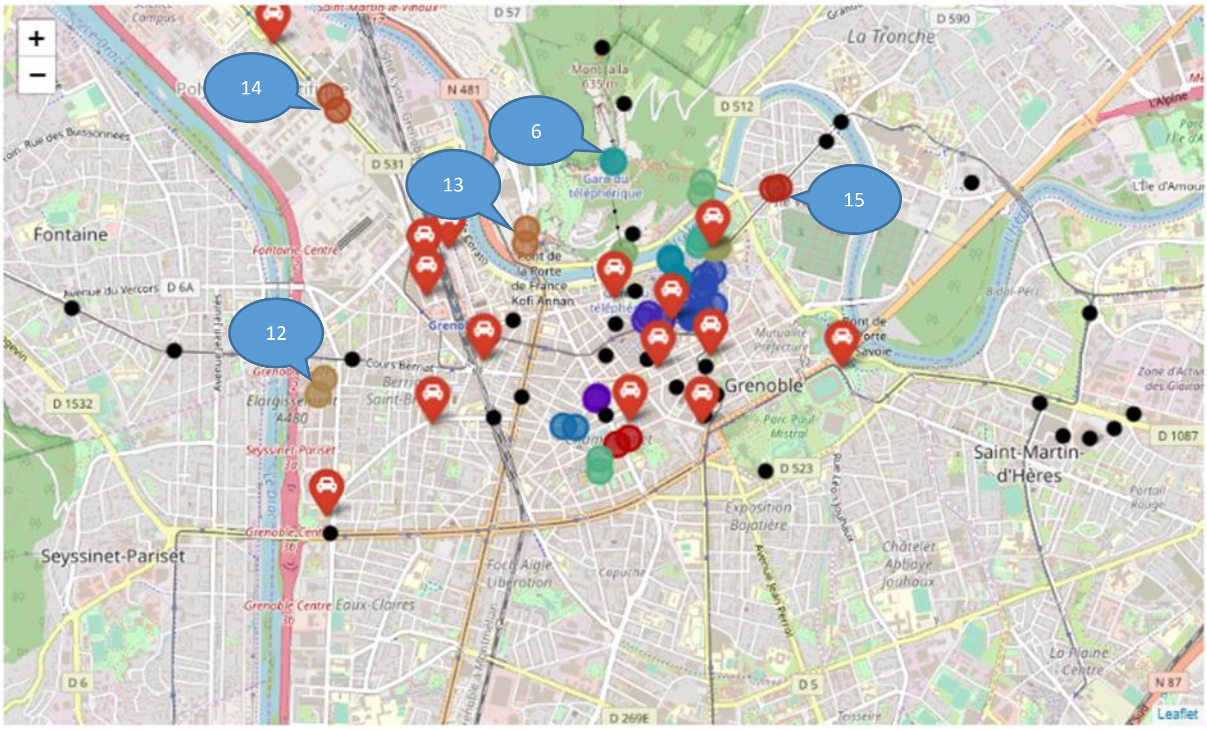
Finally, we sort the clusters by descending order according to their distances to the nearest parking lot. The clusters that are the furthest away from the nearest parking lot will be our recommendation to build a new underground parking lot. We also use the Seaborn library to visualize the distances. The visualization technique that we choose is the heatmap:



As you can see there is no spectacular difference between the 3 heatmaps as we do not have large clusters spread across the map. However if you look closely you can see that the ranking of the clusters is different depending on the distance metric used. This is particularly true when comparing the ranking of min distance and max distance.

Results

We have represented on the below map the top 5 clusters of popular venues in terms of maximum distance to the nearest existing parking lot.



- Cluster number 6 is an exception because it represents several popular venues that are located on top of a small mountain called "La Bastille" that is only accessible by cable car (so no point for a parking lot there).
- Cluster 13 is also a kind of exception because it is located at the bottom of the same mountain and it is stuck between the mountain and the river: it would probably be technically difficult to build a parking lot there.
- Clusters 12, 14 and 15 are actually very good candidates. They represent some very popular venues and they are a bit far away from the very center of the city. This is probably why they suffer from a lack of parking lots around them. These one would definitely be our recommendation to the city of Grenoble.

Discussion and conclusion

Here we would like to discuss some difficulties we have encountered during this study and make some recommendations for further improvement of the results.

- **Foursquare data:**

The number of venues returned by the Foursquare API is relatively small for the city of Grenoble. This is probably because Foursquare is a US company and Grenoble is a small French city. Foursquare is probably not yet as popular in France as it is in the USA. If there were more venues returned by the Foursquare API we could probably identify larger clusters even more representative of the city.

- **<u>Enrich the model with additional data:</u>**

One way to improve the model would be to incorporate in the model the number of parking spaces available for each parking lot. This information is available in the second dataset form the city of Grenoble. This would certainly help to make more precise recommendations.

Another interesting piece of data that we could leverage on would be the location of all bus and tramway stations. This information is probably available on the web and it would enrich the model in the sense that the need of a parking lot is less relevant if a venue is accessible by bus or tramway.