New Italian restaurant in Hertforshire

Introduction

The task is to explore cities in Hertforshire (England) and their popular places (venues).

To understand in which city center is better to open a new medium class Italian restaurant, considering that we are going to open one in Hertforshire.

As we see, usually centers of all cities located relatively close to each other contains very similar food/entertainment and other popular places, therefore it will be interesting to explore if the similar amount of Italian places close to the center of a city can define a similarity in other popular places close to the center.

The audience is an entrepreneur going to open an Italian restaurant in Hertforshire, as also entrepreneurs going to open any eating place in Hertforshire.

Data + source of data

Data source for the list of the cities and their population a table in Wikipedia page https://en.wikipedia.org/wiki/List_of_settlements_in_Hertfordshire_by_population

For the cities' coordinates we can use geolocator latitude and longitude. As there exist cities with same names located in different countries, better to add "Hertfordshire, England" to the city names for searching.

For the information about popular places (venues) we can get data from foursquare database using coordinates from geolocator.

As soon as we got all the data, we could explore more popular places in radius 1 km from city center in each city of Hertforshire. Can take a look in what cities there are Italian restaurants in top 10 places; also what popular places from the other cities each city do not have. Can compare relative amount of Italian places (amount of Italian places divided by population) for each city; and relative amount of all popular eating places. With clustering could see what cities are similar by their popular places, and if there is or not any correlation with similarity in relative amount of Italian places.

Data preparation

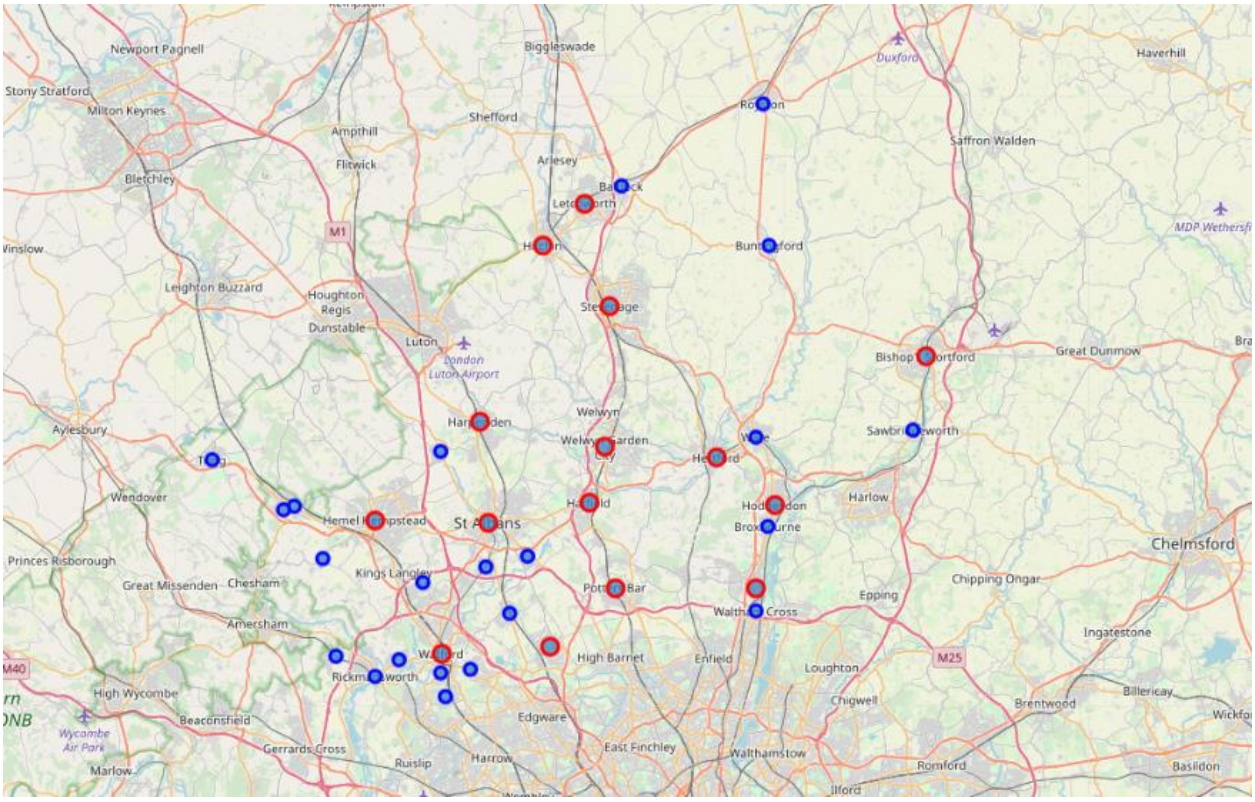
As soon as we got from Wikipedia a table into data frame, we can delete all irrelevant columns, and add a new column containing longer city's name with addition "Hertfordshire, East of England, England" to ordinary city name. We need to add this because there are other cities around the world with the same names as some cities in Hertfordshire, and it will be impossible to extract proper coordinated for them using only short city's name.

It appeared that from Wikipedia data apart from all other single city's names we got one collaborative place name "Eastbury/Moor Park/Northwood", we can keep in mind that it is not just a one place, but need to change its name to just "Eastbury", because other was coordinates will not be found.

After adding of coordinates from geolocator we are getting a data frame with 5 main columns (city, population, city long name , latitude and longitude) and 37 rows each representing one city in Hertfordshire.

Methodology

From the biginning we check how all our cities look like on the map



Trying to test how many venue we could find from Foursquare for the biggest city in Hertforshire – Watford using radius 500m from the city center.

We have found only 42 venues for the city with biggest population, so decided to increase radius to 1000m to find more venues and get more representative data.

Now for Watford we have 95 popular venues, for smaller cities less

City	
Watford	95
St Albans	92
Hemel Hempstead	48
Borehamwood	44
Berkhamsted	42
Hitchin	42
Bishop's Stortford	40
Hatfield	39

Getting data about venues from 1000m from centers of all cities, we got venues from 146 different categories.

We can take a look to 10 most popular places for each city, for example, for Watford it looks like

```
----Watford----
                  venue  freq
0           Coffee Shop  10.0
1              Platform   6.0
2     Italian Restaurant   4.0
3  Furniture / Home Store   4.0
4       Electronics Store   3.0
5        Stationery Store   3.0
6          Clothing Store   3.0
7                     Pub   3.0
8                 Theater   3.0
9           Burger Joint   3.0
```

After we have the information about these 10 popular places for each city inside data frame, we can take a look what top 3 popular other cities places each particular city does not have in 1000m radius around the center inside its 10 most popular. It is rather interesting to see what places are more popular elsewhere and less popular in particular city, for example for Watford the list of these is

```
['Indian Restaurant', 'Café', 'Hotel', 'Market', 'Park', 'Grocery Store',
'Gas Station', 'Restaurant', 'Deli / Bodega', 'Convenience Store', 'Fast F
ood Restaurant', 'Train Station', 'Supermarket', 'Fish & Chips Shop', 'Piz
za Place', 'River', 'Light Rail Station', 'Pool']
```

Starting to explore the data about Italian restaurant (as we are interested in medium class restaurant we add Pizza place to this as well), from the beginning we can try to use the same 10 popular places data frame and find all the cities that do not have Italian restaurant/Pizza place inside their 10 most popular places. These cities are:

```
Abbots Langley
Baldock
Bovingdon
Broxbourne
Buntingford
Croxley Green
Eastbury
How Wood
Redbourn
Royston
Sawbridgeworth
South Oxhey
Waltham Cross
```

It seems that all these cities without Italian/Pizza place by population are quite small cities , so let's better concentrate on bigger places there are less Italian/Pizza places than elsewhere. Now let's get from dataset only full amount of Italian Restaurants and Pizza Places (still in 1000m from the center radius) normalized by the size of the population and try to visualize it, the bar chart will show us clearly what cities have less Italian places per person than another. As we are interested in opening a new restaurant, we can concentrate more attention on relatively big cities (population >20000), because they are usually more visited by people living in other places for weekend entertainment or for work, so in reality there could be even more people that city's population, other way small cities usually (if they

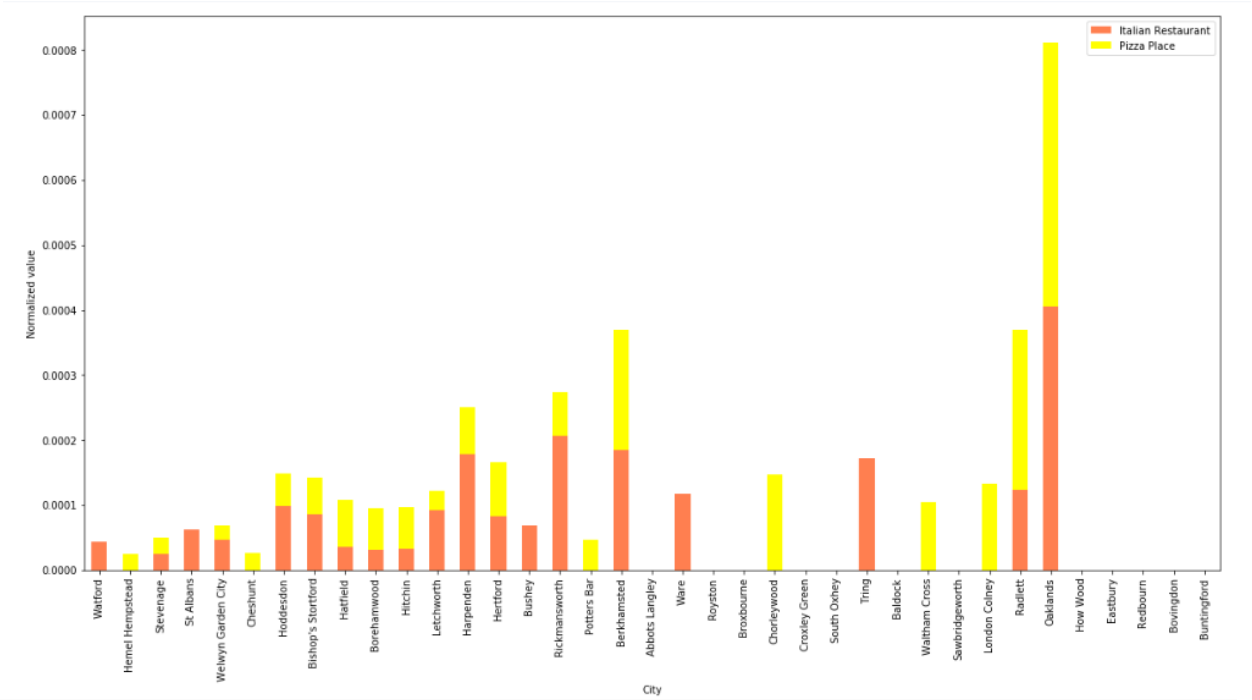
do not contain something extra special like tourist place or other unique place) usually are not crowed with people living in other places, even their own population often is going to work and rest to bigger crowded places.

The same way we can visualize amount of other popular food places in comparison with Italian/Pizza places. As we have very many different venue categories, among them there are many different food places categories, let's group all food places into 5 categories: Italian (because we have a particular interest in them), National (Asian, Indian, English, and so on), Café (Café, Snak places), Pubs, and others.

After we will get clear about where to open a new Italian restaurant we can use clustering to explore if the similar amount of Italian places close to the center of a city can define a similarity in other popular places close to the center.

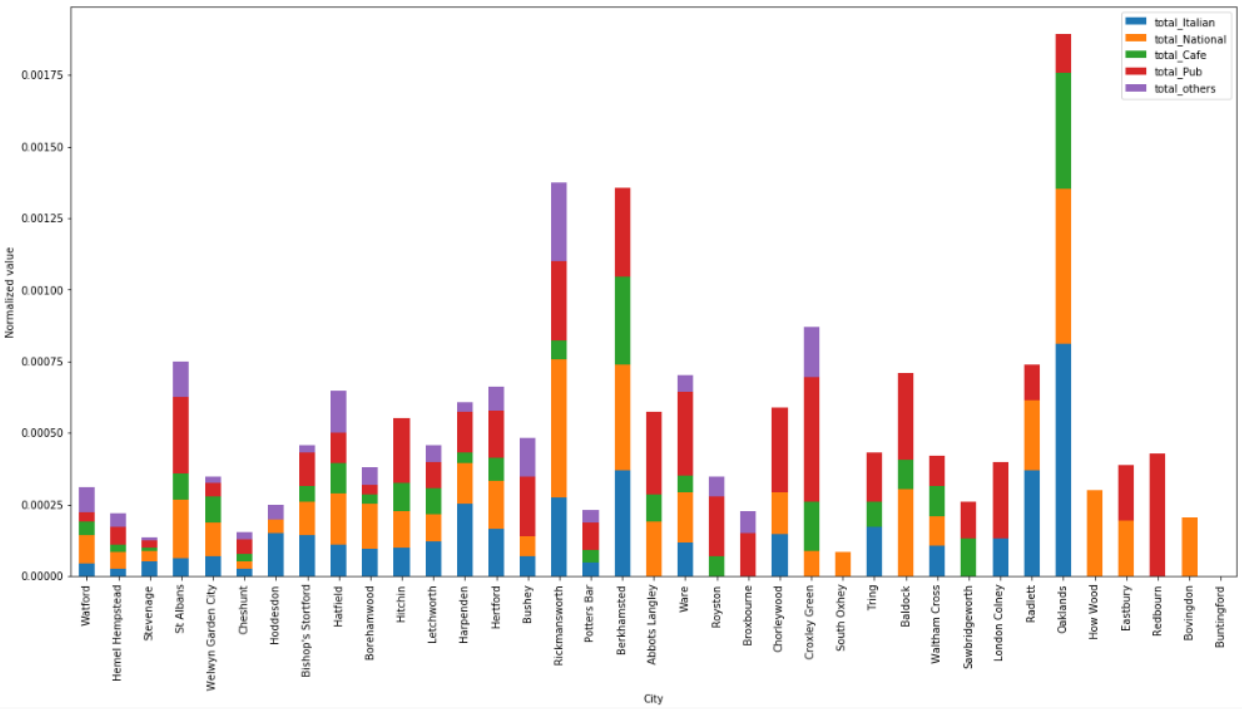
Results

From the bar chart below we see that from the relatively big cities (population>20000) the smallest coverage of popular Italian/Pizza places is in Hemel Hemstead and Cheshunt, so these 2 are the candidates for opening a new Italian please.



It is logical to ask, may be Hemel Hemstead and Cheshunt are full of other popular food places, therefore Italian places are relatively not so popular.

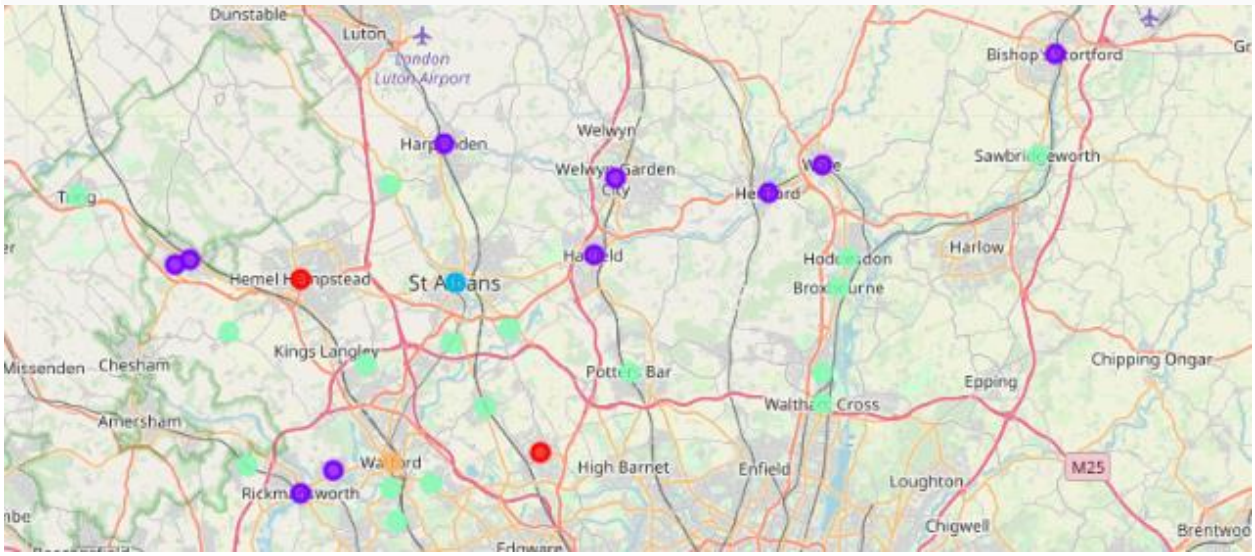
The next chart helps to understand the amount of other popular food places in comparison with Italian/Pizza places (all food places are grouped into 5 categories: Italian (because we have a particular interest in them), National (Asian, Indian, English, and so on), Café (Café, Snak places), Pubs, and others)



We still see that in both Hemel Hempstead and Cheshnut there are not so many other popular food places, so the idea to open a new Italian Restaurant in one of them looks fine.

Discusson

After we get clear about where to open a new Italian restaurant we can use clustering to explore if the similar amount of Italian places close to the center of a city can define a similarity in other popular places close to the center.



From the clustering we see that Hemel Hempstead and Cheshunt, what both has similarity in the luck of popular Italian/Pizza places, are in different clusters, so do not have a lot of other similarities; the same about Hoddesdon and Bishop's Stortford - they also have similar Italian/Pizza popular places amount and for both of them Italian restaurant is in top 3, but they are in different clusters, so there are not a lot of more similarities. From the other side, we see also an opposite example - Berkhamsted and Rickmansworth are in one cluster and also have similar amount of popular Italian/Pizza places. So actually there are no big correlation between general similarity in popular places and similarity in having close relative amount of Italian/Pizza places

Conclusion

We had successfully explored cities in Hertforshire (England) and their popular places (venues).

For the opening a new medium class Italian restaurant in Hertforshire we had offered either of 2 relatively big cities (population >20000) Hemel Hempstead and Cheshnut .

The analyze showed that there is no big dependence and correlation between amount of Italian places close to the center of a city and similarity of other popular places close to the center.