

Introduction

The task is to explore cities in Hertfordshire (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 Hertfordshire.

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 Hertfordshire, as also entrepreneurs going to open any eating place in Hertfordshire.

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 Hertfordshire. 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.

Methodology

- Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.