

Final Assignment IBM-Data Science-Coursera-Course: The Battle of Neighborhoods

Title: "Basis for a decision-making concerning an investment in establishing a shopping mall in Oxford, Great Britain"

1. Szenario (Introduction, Business Problem)

Suppose there's an investor who plans to establish a shopping mall in Oxford. One of his key-questions at the beginning of the development-process regards the number of competitors (existing shopping malls) in the particular regions of Oxford.

Your job is to give a recommendation on the basis of existing data, in what region the investor should open a shopping mall.

Your approach: The decision-criterion is the number of competitors in the region. A "no go-area" is on the one side an area with a high number of competitors. On the other side, regions with no shopping malls could imply, that there's absolute no demand for a shopping mall. So the most interesting area is that "in between".

2. Data

At first I build a dataframe of neighborhoods in Oxford by scraping the data from Wikipedia. Then I get the geographical coordinates of the concerning neighborhoods and obtain the corresponding venue data for the neighborhoods from my Foursquare API.

3. Model-Approach: Clustering

Finally I run a k-means-algorithm to cluster the neighborhoods in Oxford into 3 clusters. On this basis I give a recommendation to the investor.

4. Recommendation

Cluster 0 (red dots in the map) shows the area(s) with a neglectible number of shopping malls in Oxford. They are in the north/east and a little in south/west. Here the investor should be reserved to open a shopping mall. It is possible that there is not sufficient demand for it.

Cluster 1 (purple dots in the map): It's the middle of Oxford. Here seems to be the main potential for making money with a shopping mall: Not so many competitors but a apparently sufficient high demand.

Cluster 2 (green dots in the map): It's a small region in the west of Oxford. Here is a high numbers of competitors. On the other side: Here is apparently the biggest demand. If the investor is succesfull in cluster 1 and has a reputation, why at a later time also succussful in this cluster?