REPORT

1. Data Scraping

The known data from week 3 were used:

Location data in Toronto, which includes a table of postal codes and boroughs for all of Toronto's districts, is taken from a Wikipedia page. Longitude and latitude data for all districts are included in a file that can be downloaded online. All venues and their details can be accessed via the Foursquare API.

The location data is scraped from Wikipedia is converted into a Pandas DataFrame. The DataFrame is then cleaned up by removing all lines with unassigned boroughs. All districts with an unassigned value are replaced with the name of the borough. Using the Nominatim longitude and latitude data read into a Pandas DataFrame. Following both DataFrames are merged.

A DataFrame of up to 100 top venues venues within a 500m radius for each neighborhood is created using the Foursquare's API.

From the Foursquare data neighborhoods without a bank were extracted.In addition neighborhoods of Toronto were identified according to the number of venues (counting with a limit of 100 venues). They were sorted decreasingly. These data were used to determine a ranking for the site selection (locations with many venues were preferred) by merging both DataFrames.

BLOGPOST

#1. Introduction/Business Problem: I was commissioned by the CEO to find a new location for a bank in Toronto. However, I am from New York and do not know the neighborhoods in Toronto. The location should be chosen in such a way that no other bank is located in the neighborhood and a high customer frequency can be expected. A selection of 5 possible locations should be the result.

#2. The Foursquare location data from Toronto is used. These are used to find out if there is already a bank in the neighborhood. In addition, the data enables the identification of neighborhoods with many ventures. A location should be determined by combining both information.

#3. Methodology: The known data from week 3 of the Capstone project were used. From this data neighborhoods without a bank were extracted. In a second step, the neighborhoods of Toronto were identified according to the number of venues. These data were used to determine a ranking for the site selection (locations with many venues were preferred).

#4. Results: As requested by the CEO, 5 locations have been identified as candidates for a new location, as they do not yet have a bank and customer traffic is expected to be high due to many local businesses. To enable the CEO to make an objective decision, the top 5 business types in the neighborhood were identified.

#5. Discussion: The data analysis gives a good first impression which location could be suitable. For a closer look it would be necessary to find out which types of business are frequently visited in combination with a bank. This would be possible with a similar approach.

#6. Conclusion: There are 5 possible locations that should bring the greatest benefits to our bank. In an upcoming discussion with the CEO, a decision will either be made on the basis of the top 5 businesses at the respective location or alternatively further parameters will be determined which are decisive for our choice.