Capstone Project - The Battle of the Neighborhoods

Table of contents

Introduction: Business Problem of knowledge of neighborhood for real estate agent

Real estate agents compete for the same customers and customers want to find the best home and also best neighborhood for their family. Deep knowledge of area and neighborhood brings advantage in competition. From real estate point of view there is a battle of neighborhoods.

Nobody can remember or know all venues in Helsinki Finland area and so cannot promote all venues and categories which can found through Foursquare API. We would like to provide targeted information near properties to sold for real estate agents. We would also cluster and categorize living areas to quickly tell in which category of property belongs and what are the unique characteristics of that area for example good parks and cafeterias. This could make the difference when agents have sales meeting with owners. This information can be even crucial when families deciding where they are going to move and buy new home.

Data

Based on definition of our problem, factors that will help real estate agents are:

- All venues of neighborhood
- Top venue categories in neighborhood
- Overall style for example cafes and parks

The following data sources will be needed to generate the required information:

- Wikipedia page of Helsinki neighborhood including
- All venues or neighborhood area through Foursquare API
- Geolocator to get coordinates of neighborhoods

We will use the **explore** function to get the most common venue categories in each neighborhood of Helsinki. We will also cluster neighborhoods to give similarity information to end customer.

Methodology

We are providing characteristic information about Helsinki neighborhoods combining venue and pricing information and making clusters of neighborhoods.

First phase for project was that:

- We collected all neighborhoods with sub-neighborhoods.
- Added coordinates to all neighborhoods
- Added average m2 pricing to all neighborhoods, when data was available

Second phase

- We cluster all neighborhoods and venues correlated
- Then we got that number down to the top 10 unique venues in that neighborhood
- And then down to the top 5 based on the most frequented venues

```
In [126]: #Example of how real estate agent could fetch information by neighborhood when going to sales meeting
          helsinki_merged.loc['Eira']
```

```
Out[126]: lat                                  60.1562
          lng                                  24.9384
          Price                                   8051
          Venues                                    37
          Cluster Labels                             0
          1st Most Common Venue        Chinese Restaurant
          2nd Most Common Venue                     Zoo
          3rd Most Common Venue               Gastropub
          4th Most Common Venue           Garden Center
          5th Most Common Venue                  Garden
          6th Most Common Venue     Furniture / Home Store
          7th Most Common Venue          French Restaurant
          8th Most Common Venue                Fountain
          9th Most Common Venue                  Forest
          10th Most Common Venue             Food Truck
          Name: Eira, dtype: object
```

Discussion

Helsinki City is a relatively small city so for effective data analysis we would need even more venues in Foursquare. Lots of neighborhoods have small amounts of many kind of venues to offer people who want to move there. With adjusting k-means clustering we could have even better results. There is a big difference in average m2 housing price inside Helsinki and it look's like that relates more to other things than only venues in a neighborhood.

Conclusion

This kind of data analysis can be helpful for a real estate agent and provides competitive advantage. Data should be used to build an app and Python code of this analysis should be provided as a microservice.