

Capstone Project Blog by AAFunian

Capstone Project - The Battle of the Neighborhoods

Table of contents

- [Introduction: Business Problem](#)
- [Data](#)
- [Methodology](#)
- [Analysis](#)
- [Results and Discussion](#)
- [Conclusion](#)

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

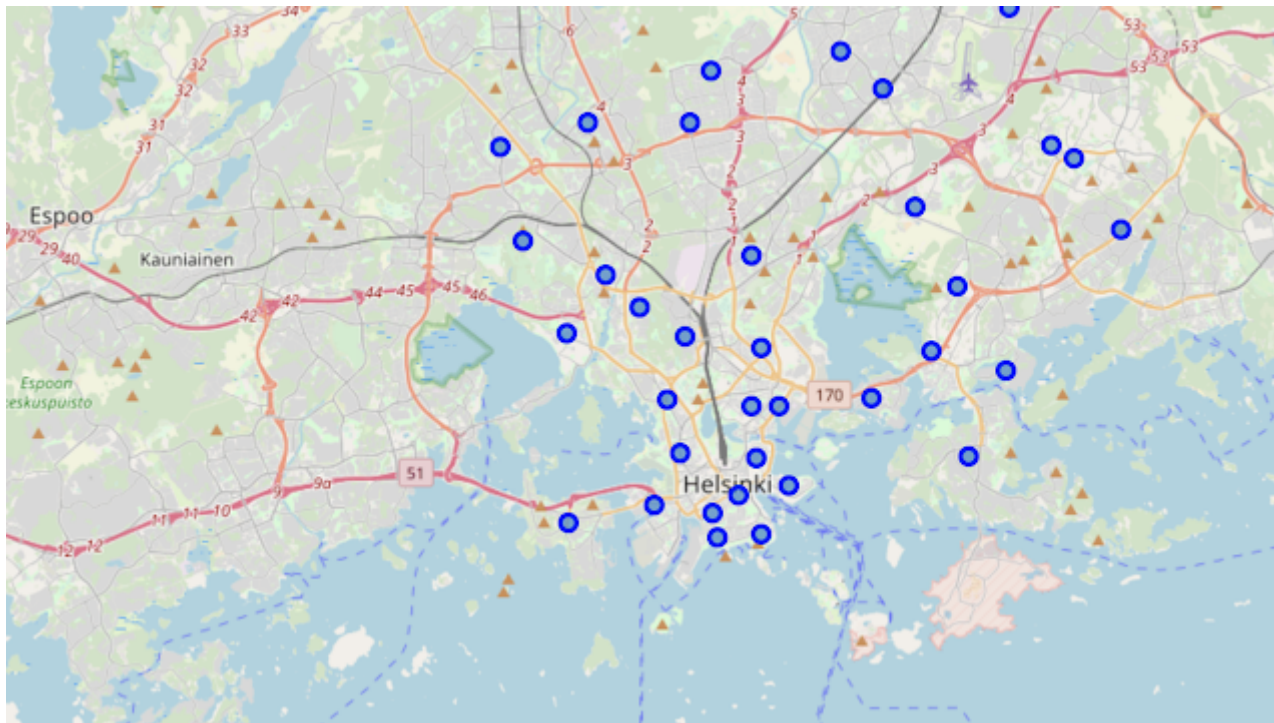
Based on definition of our problem, factors that will help real estate agents are:

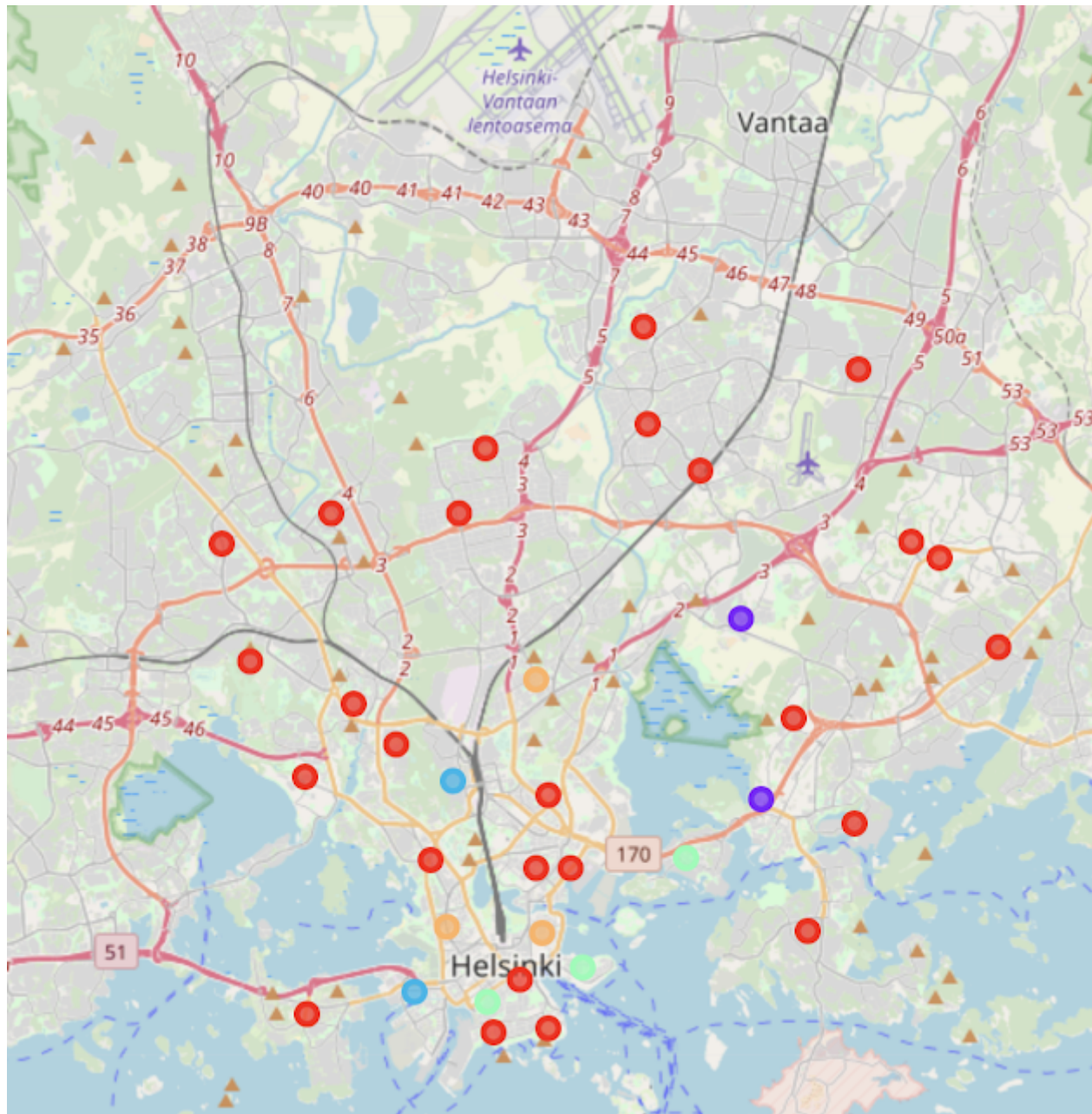
- All venues of neighborhood
- Top venue categories in neighborhood
- Overall style for example cafes and parks

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

