Coursera Capstone

IBM Applied Data Science Capstone

Real Estate

By: Deepak Raichandani

(December 2019)

Table of contents

Introduction

Business Problem of knowledge of neighbourhood for Real estate agents compete for the same customers and customers want to find the best home and also best neighbourhood for their family. Deep knowledge of area and neighbourhood brings advantage in competition. From real estate point of view there is a battle of neighbourhoods. 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 categorise 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 neighbourhood
- Top venue categories in neighbourhood
- Overall style for example cafes and parks

The following data sources will be needed to generate the required information:
- Wikipedia page of Helsinki neighbourhood including
- All venues or neighbourhood area through Foursquare API
- Geo-locator to get coordinates of neighbourhoods

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

Methodology

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

First phase for project was that:

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

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