

Assignment: Applied Data Science Capstone project

Marc Jellema

December 17, 2020

Abstract

This report is a deliverable for my capstone project *Battle of the neighbourhoods*, part of my graduation for my **IBM Data Science Professional certificate**. First leg of the capstone project is the introduction and data part of this report that you can find below. The other sections will be filled in the second and final leg of the capstone project. Feel free to leave comments or send me an email: mjellema@omnia.nl

Introduction

I tasked myself with the assignment to find additional ways to enhance automatic estimation of house prices in the Netherlands. The use of the **FourSquare** API is mandatory for this assignment, part of the IBM Data Science professional certificate [1]. As a tech savvy person myself, I'd heard of FourSquare and I do remember seeing the FourSquare logo on venues in the past, but never came across its footprint in the Netherlands. Wondering if I could use the FourSquare API in finding an answer for a challenge in the Netherlands, I did some research and as it turns out, the FourSquare database is more than comprehensive enough to use for data science challenges in the Netherlands as well. Not knowing a single person actively using FourSquare to let their friends know where they check-in, it puzzled me how FourSquare actually accumulates their data. James D. Walsh wrote a nice article [2] on how FourSquare accomplishes this. Enough about FourSquare for now, let's focus on introducing the question I want to answer with help of the FourSquare API and the data it can provide:

*Based on **FourSquare** data, can we find a correlation between the average house price, the number of restaurants and the number of inhabitants in the provinces of the Netherlands?*

Although the COVID pandemic is having an impact on the overall economy world wide, it seems house prices have not yet taken a hit due to the pandemic. Startups like **Promodomo** [3] use algorithms to calculate house prices, even when they are not on the market yet. Algorithms like Promodomo's always benefit from additional sources to enrich their calculation. Searching for a correlation based on FourSquare data is just one off the many possibilities to augment the estimation of house prices and this is why I choose to take this challenge and come up with answers in this capstone project of my IBM Data Science Professional certificate.