**Coursera Capstone Project Report:**

**Predicting House Prices in Zagreb, Croatia**

Introduction

The problem to be investigated is to what extent do nearby venues determine house prices. The location chosen for this project is the city of Zagreb, Croatia. The people interested in solving this sort of problem would be as follows: people looking to buy houses in Zagreb, real-estate agents, city planners, investors etc.



Figure 1. Zagreb, Croatia

The data

In this project the Foursquare location data is leveraged to explore and compare neighborhoods, and the house prices (as well their size in meters squared) is parsed from [www.gohome.hr](http://www.gohome.hr) website using BeautifulSoup4, html2text and urllib python libraries and modules. By manually creating and filling the corresponding .csv file with neighborhood’s latitudes and longitudes and assigning those latitudes and longitudes to all the houses in the same neighborhood, the first part of the dataset is obtained. As already mentioned, using the Foursquare location data to each house a vector of the number of nearby venues is joined. That’s how the whole dataset is obtained. The data will be used to find correlations between the selected features (venue categories and size of the house) and the house prices in EUR.

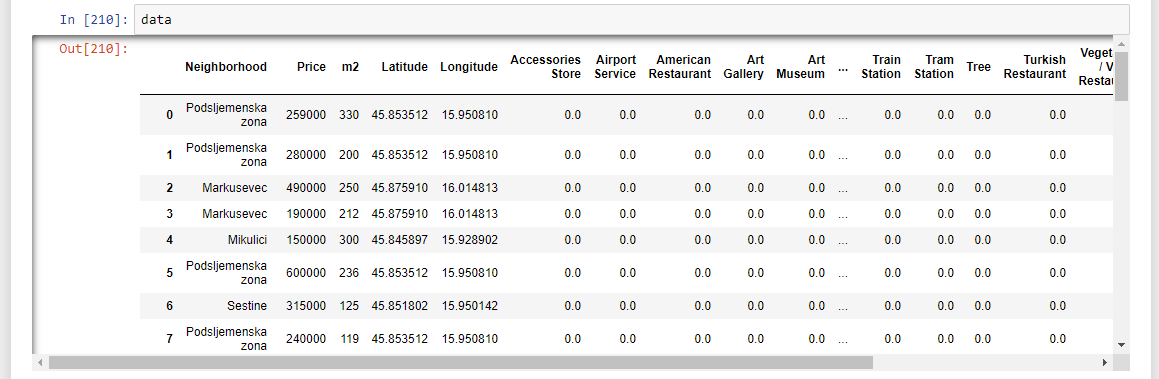


Figure 2. The dataset