IBM Data Science certificate – Capstone project

In this document I will describe the business problem that I will try to solve together with a description of the data that I will use to formulate an answer to the problem.

Description of the idea

Intro

It is not a secret that the real estate prices in New York are some of the highest and most outrageous in the world. It does however depend on different factors if the price of real estate is higher or lower than average. Prices in the Upper East Side are different from prices in Harlem for example.

Research questions

How is it that prices in different neighborhoods are so different from each other? Is it just the location of the real estate, or are nearby venues also a factor of influence on the price? In my research I will try to answer some specific questions:

-   Do surrounding venues have any effect on the price of real estate?
-   If so, what kind of venues have and effect and to what extend?
-   Can we predict the price of a real estate based on the neighborhood and surrounding venues?

Target group

The information that we derive from this research can be used by home buyers or realtors.

Description of the data

Data sources

 Most important sources of data that will be used are as follows:

-   The average price by neighborhoods in New York city according to propertyshark.com
-   The venues in each neighborhood (FourSquare API)

The prices of real estate in neighborhoods can be scraped. The coordinates of the neighborhood can be determined by use Geocoder in combination with the Foursquare API we can get the surrounding venues.

Techniques

To determine if venues have any effect on housing prices, the correlation has to be determined. IF there is any correlation I will use a regression analysis to analyze the data and determine what kind of venues have the most effect on house prices.