1. **Introduction**

Many factors affect in the price of a house. The most commons are, normally, the size, amount of rooms, amount of rooms and others. Many papers have been written specifying the correlations between such variables, emphasizing the overall gain in having one more room as a bathroom or as a bedroom or the rise in evaluation from an expansion of one square meter.

Even with that knowledge, there are more impacts to be considered. Therefore, this paper tries to expand the vast literature on the topic by considering the impact of the amount of venues in a 500m radius of a sold house in the overall price of the same through the data available for King Count in the USA. By gaining such knowledge, investors in the real estate market will be able to make even more predictions based on the overall growth of certain area in the price of homes in that area or, if there is no impact, consider acquiring houses in areas with a smaller perspective of growth.

1. **Data Acquisition and Cleaning**

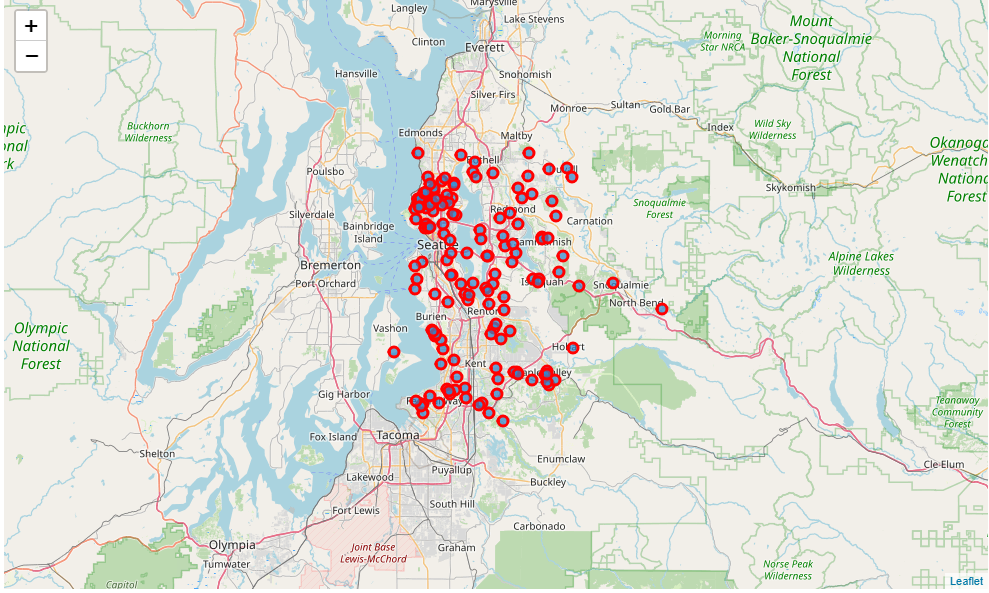
For this paper, the data utilized was gathered by two main sources, IBM and Foursquare. From IBM, it was possible to obtain price and information about the houses sold in King County, USA between May of 2014 and May of 2015. This data also included latitudinal and longitudinal coordinates of these houses, enabling us to access the Foursquare API in order to pinpoint those positions. To get the amount of venues, Foursquare API retrieve a list of venues in a 500m radius from the geographical coordinates in 2020, that will be used as approximation for venues in that time period.

After having these data, some changes were required. For the house sales, most of it was already organised in a file with 21 thousand rows with the amount of bedrooms, of bathrooms, of floors, of square foot in the lot, of square foot in the actual residence, latitudinal and longitudinal coordinates, if it is in front of water, an overall grade, the condition of the house, the year built, the year renovated, the zip code and the date of the transaction. The data was already organised, so only the column with a defective index was dropped.

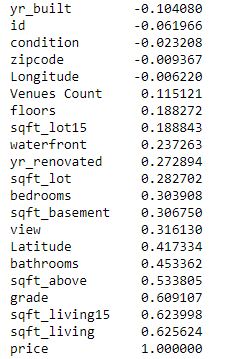
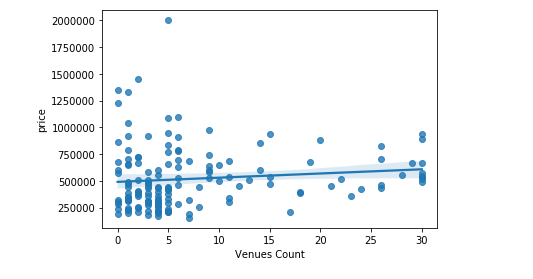
Nevertheless, the Foursquare data required some refinement. After retrieving the coordinates for some rows and applying to the API, a new data frame was formed with each of the venues connected to a certain house ID. Therefore, it was needed to create a new data frame with the amount of times that each ID appeared, that is, the amount of venues designed to each ID and merge it to the IBM data frame, adding a column named Venues Count to a smaller set of data.

1. **Exploratory Data Analysis**

Initializing 150 data samples of the record, we obtained the display of houses as follows using the Folium package:



After that, the analysis focused on obtaining some kind of correlation between the shown variables. Firstly, it was tried the Pearson Correlation and then a linear regression model, which appeared as following:



Besides that, the obtained for the equation was positive, but it represented less than 0.05 of the total reinforcing the results of the Pearson correlation.

1. **Conclusions and Next Steps**

The results obtained for King County, USA shown close to no statistical significant correlation between the amount of venues close to a house and its overall sales price. It is possible that this could be caused by the difference in time period of the venues and the sales, by a possibility that house away from venues might be bigger or because there is not enough impact among variables.

Going forward with the research, it would be needed more data and usage of Foursquare API, which limited the search size. Getting more recent data for US house sales in different states might get a certain result of the overall increase by the proximity of the venues.

Even though, for the moment, there is a strong indicator that there is no real correlation between the amount of venues and the price of the house. Because of this, it is possible to expand real estate market actions into areas that do not have a perspective of economic growth in stores and still receive profit for the activity.