

# Capstone project IBM Data Science

## 1 Introduction

House prices have huge margins and prices are often very volatile. Predicting house prices using data can lead to a more efficient housing market. The goal of this research will be to quantify subjective house prices. Real estate agents often give certain properties to houses, e.g. number of bedrooms, bathrooms, square meters and location. With data becoming more publicly available, distances to certain venues and number of venues in the area can be incorporated in the pricing models. The goal of this research is to find out which type of venues or whether venue data increase the price of a house in King County, USA.

## 2 Data

The data that are used is the King County data set<sup>1</sup>. The dataset contains house prices, the date the house was sold, number of bedrooms, bathrooms, square foot of the house, living room and lot, if the house has a view to a waterfront, if the house has been viewed, general conditions, grades based on King County grading system, year the house was built, year the house was renovated, and finally some location data. We will also be using data from Foursquare<sup>2</sup> to find venues in the area.

---

<sup>1</sup>Data is retrieved from <https://data.kingcounty.gov>

<sup>2</sup>Data retrieved from <https://foursquare.com/>