Predicting house prices in King County, Washington, USA

By Panagiotis Petsas

This report is part of the course "Applied Data Science Capstone" from "IBM Data Science Professional Certificate" on Coursera.

Part 1: Introduction

1 Introduction

1.1 Background

King County is one of the most populated counties in USA, with an estimated population of more than 2.25 million people [1]. It is located in Washington, where Seattle is the county seat. It has an area of almost 6000 km^2, where 490 of them being covered by water. This county consists from more than 30 cities and a big variety of historical places [2]. Considering the importance of King County, it is of great interest to identify how house prices vary in this area. In addition, it would be important to identify the factors that drive these prices.

1.2 Problem

Houses are one of the most important types of tangible assets. They are the place where people can settle, or start their business. If someone owns multiple houses, it is possible that they will maintain one for settlement and use the remaining ones to produce economic value, either by renting them or by selling them. House prices might vary depending on many different factors. The house features (e.g. total area, amount of bedrooms or bathrooms) play a major role in pricing. Other important factors are the city/area where the house is located or its proximity from different venues, such as shops, restaurants, utilities or services. Social and economic factors might also take place, which means that a house's value might change at different time intervals.

Stakeholders might be aware of many factors that determine the house prices from their experience, but they might not be able to quantify the magnitude of each factor. For example, houses with more bathrooms and bedrooms might have a higher price, but one would not know which of the two increases the price more. It is therefore important to utilize house data in order to identify patterns between house features and house prices.

1.3 Interest

Real estate agents would be the most interested individuals in this subject. Gaining knowledge about the factors that drive the house prices would be beneficial, as they could identify spatial and temporal trends, evaluate houses in a more informed way and help their clients to sell their property, satisfying both parties. Even in the case where an individual wants to sell its house

without an agent, this could provide useful information in order to set the right selling price. While this study will be held in King County, interested stakeholders can utilize the results for similar places or use the same methods to repeat the analysis for other places that might not be similar to this one.

References

[1] Most populated USA cities

(https://en.wikipedia.org/wiki/List_of_the_most_populous_counties_in_the_United_States)

[2] King County, Washington – Wikipedia

(https://en.wikipedia.org/wiki/King_County,_Washington)