Contents

[Introduction 1](#_Toc5955700)

[Data 1](#_Toc5955701)

[Methodology 2](#_Toc5955702)

[Results 2](#_Toc5955703)

[Discussion 3](#_Toc5955704)

[Conclusion 3](#_Toc5955705)

# Introduction

# This data science capstone project will try to solve the following problem: Real state valuation. Real state valuation is the task of assigning a monetary value to a specific property based on its attributes. These attributes relate to different aspects of the property such as: location, distance to convenience stores, size, neighborhood, etc.

# Real state valuation is complex and dynamic problem that can be affected by a variety of factors such as: the economy, crime rates, geology, etc. Every single aspect of everyday life can influence the value of a given house. Is it located in a decent neighborhood? Is it the right size? How old is it?

# This problem is of interest to two groups: 1) Real state agencies and 2) potential buyers (clients). For real state agencies, understanding the nuances of the situation may help them achieve good profits at reasonable prices. On the other hand, for potential buyers, a deeper understanding of the scenario may help them buy decent properties at a reasonable price.

# The formal definition of the problem is as follows: Can we accurately predict the price of a given property using information such as: house age and/or location?

# Data

# The data that will be used in this project is the so called “Real Estate Valuation Data Set”. This data set was obtained from the well-known UCI Machine Learning Repository. This is a multivariate data set with 414 instances. The number of attributes for each instance is 7 and there are no missing values. This data set is usually used for Regression tasks.

# The following table describes the dataset in more detail:

|  |  |  |
| --- | --- | --- |
| **Attribute name** | **Description** | **Attribute type** |
| No | Instance number, identifier | Numeric: Integer |
| X1 transaction date | The date the transaction was made | Numeric: Real |
| X2 house age | House age in years | Numeric: Real |
| X3 distance to the nearest MRT station | Distance to nearest subway station | Numeric: Real |
| X4 number of convenience stores | Number of convenience stores | Numeric: Integer |
| X5 latitude | Latitude | Numeric: Real |
| X6 longitude | Longitude | Numeric: Real |
| Y house price of unit area | House price of unit area | Numeric: Real |

As we can see, most of the attributes are real values, with a few exceptions that are expressed in integer. Also, most of the attributes relate to location information.

In the following section this dataset will be preprocessed and analyzed further.

# Methodology

In this section I will go over the whole process of data science. This section is divided into the following subsections: Data Preprocessing, Data exploration and Modeling.

# Results

Discuss the results

# Discussion

observation

# Conclusion

conclussion