

# **Renting a House in Seoul**

Pham Ngoc Son

2021-06-01

## **1. Introduction**

### **1.1 Background**

Renting a house in any place is never an easy task. From renting price to service and utility, everybody want the best apartment with the least pay. Currently, the apartment price in Seoul is readily available in many website such as Naver land or KB land. However, price is only one part of the problem. Customer also cares about many other facility around the place such as transport, park, restaurant, supermarket, cinema to name a few. The problem is this data is not available on any website so customer will have to take a lot of time to investigate each area. This is a quite frustrating task and often they will give up and choose any place with reasonable price.

### **1.2 Problem**

Location data from each apartment can be used together with Foursquare database to get venues around it within a predetermined radius. In this project the venue data from Foursquare will be used to classify each apartment and give user more useful information on their choice of living place.

### **1.3 Interest**

This analysis is not only good for the house finder in Seoul but also for the housing data provider such as Naver & KB. The government also can use this analysis to improve living condition in Seoul.

## **2. Data acquisition and cleaning**

### **2.1 Data sources**

Data for all apartments, buildings and houses in Seoul can be found in the below link:

<https://www.juso.go.kr/addrlink/addressBuildDevNew.do?menu=match>

The size of this data is very large so take care before downloading it.

The data included all building name, code, detail address included city, district, ward, road number ... The X, Y location of each apartment and building is also provided but it is in Korean standard.

### **2.2 Data cleaning**

Due to the size of the data, after downloaded, ~ 1000 rows were randomly extracted to be used

for this analysis.

The X, Y coordinates in Korean standard must be transformed into longitude and latitude to be useful in the analysis. This proves to be a quite difficult task since the transformation from GRS80 UTM-K to latitude and longitude is not readily available. Fortunately, a free software provided by Korean government named NGI pro can be used to do this transformation.

There are some missing values in the building name feature and all of them were removed since this is an important feature.

The feature names were also translated from Korean to English for non-Korean reader to be able to understand the data.

The price was added randomly to each apartment based on the actual renting price in Seoul. It will be better if exact data was used but because the price info was not free so for the scope of this report, we will only use assumed value.

## 2.3 Feature selection

After data cleaning, some data rows were removed so that exact 1000 rows remained in the database. The data consisted of many not-useful features such as city code, building code, road code ...

After carefully checking, the selected features are as below:

City – District – Ward – Road – BuildingName – PostalCode – RentPrice – Longitude – Latitude.

	City	District	Ward	Road	BuildingName	PostalCode	RentPrice	Longitude	Latitude
673	서울특별시	강서구	등촌동	공항대로58나길	힐하우스	7666	970	126.862080	37.549676
378	서울특별시	중랑구	상봉동	신내로7다길	상봉빌라	2079	2295	127.091766	37.603401
118	서울특별시	중구	신당동	동호로8나길	신당힐빌리지	4594	840	127.014236	37.552235
747	서울특별시	강서구	공항동	공항대로	무성빌딩	7623	1335	126.812512	37.558736
8	서울특별시	종로구	누상동	필운대로5가길	창원예가	3037	2865	126.966781	37.580645

The Longitude and Latitude of each building will be used to get venue information using Foursquare API. This venue data will then be used to classify the building into different categories and allow customer to easily select building of their choice.