

Predicting Land Values of Tokyo Metropolitan Area with Regression

Kiyohiko Nishi

March 2020

*Submitted in partial fulfillment of the requirements of the Coursera course,
“Applied Data Science Capstone”*

1. Introduction

1.1. Background

Owning one's own land to live is a hard decision to make. There are a lot of things to be considered from advantage of neighborhood - convenient to commute, calm and peaceful, etc. – to disadvantage of the neighborhood – high crime rate, high risk of natural hazard, and so on. Generally speaking, however, the most important factor which affects our decision is the land price.

As for land price in Tokyo metropolitan area, where it is considered expensive area to live compared to other areas in Japan, increase tendency has still been seen despite of economic stagnation that has caused drop in land price except Tokyo area.

In this circumstance, a simple tool providing support for making decisions is strongly expected.

1.2. Problem to be solved

A model which is able to predict proper land price by analysis of the data of land price in Tokyo metropolitan area is prepared. Then, as a verification of the model, the most affordable place to live is suggested.

2. Data

2.1. Data Sources

Following sources are used in this study.

Land Attribution Data	Data Source	Description
Area name Latitude Longitude Land price Nearest station name Distance to nearest station	National Land Numerical Information: Publication of Land Price Data http://nlftp.mlit.go.jp/ksj/index.html Provided by Ministry of Land, Infrastructure, Transport and Tourism.	Various GIS data of each prefecture is provided. Shape file format, GML, and GeoJSON format is available. Dataset is able to be downloaded as zip file by each prefecture, containing all relevant files. Multiple attribution data is included in a dataset, such as regulation and land usage, other than those used in this study.
Number of stations in neighborhood	Foursquare location data https://foursquare.com/ Provided by Foursquare Labs Inc.	Data is retrieved using Foursquare API endpoint at the 'search' endpoint with query as "station" and limitation in category id which is relevant to train/metro station.
Distance to Tokyo station Distance to city hall of the capital of each prefecture	Coordinates of each landmark: OpenStreetMap data Obtained by Nominatim geocoder of Geopy module https://wiki.openstreetmap.org/wiki/Nominatim Distance between two position: calculated by Haversine Formula https://en.wikipedia.org/wiki/Haversine_formula	Obtaining latitude and longitude of each landmark and calculated the distance between each position in the dataset using Haversine Formula.

2.2. Data Description

Overview of the dataset is as follows. In total, 1,019 points of data is used to analyze.

	City	Latitude	Longitude	Stations	Station	DistToSt	ToCityHall	ToTokyoST	LandValue
0	武蔵野	35.710699	139.584039	0	吉祥寺	960	9.998406	17.018053	560000
1	調布	35.670318	139.586601	0	仙川	1000	9.735226	16.498822	351000
2	三鷹	35.678447	139.587580	0	三鷹台	2000	9.487364	16.371411	301000
3	武蔵野	35.700362	139.587689	3	吉祥寺	900	9.467431	16.504097	653000
4	武蔵野	35.704320	139.588387	0	吉祥寺	800	9.470161	16.505738	595000

3. Method

3.1. Data Understanding

3.2. Data Preparation

3.3. Modeling

3.4. Evaluation

3.5. Deployment

4.