

Coursera course, “Applied Data Science Capstone”

Predicting Land Values of Tokyo Metropolitan Area with Regression

Kiyohiko Nishi

March 2020

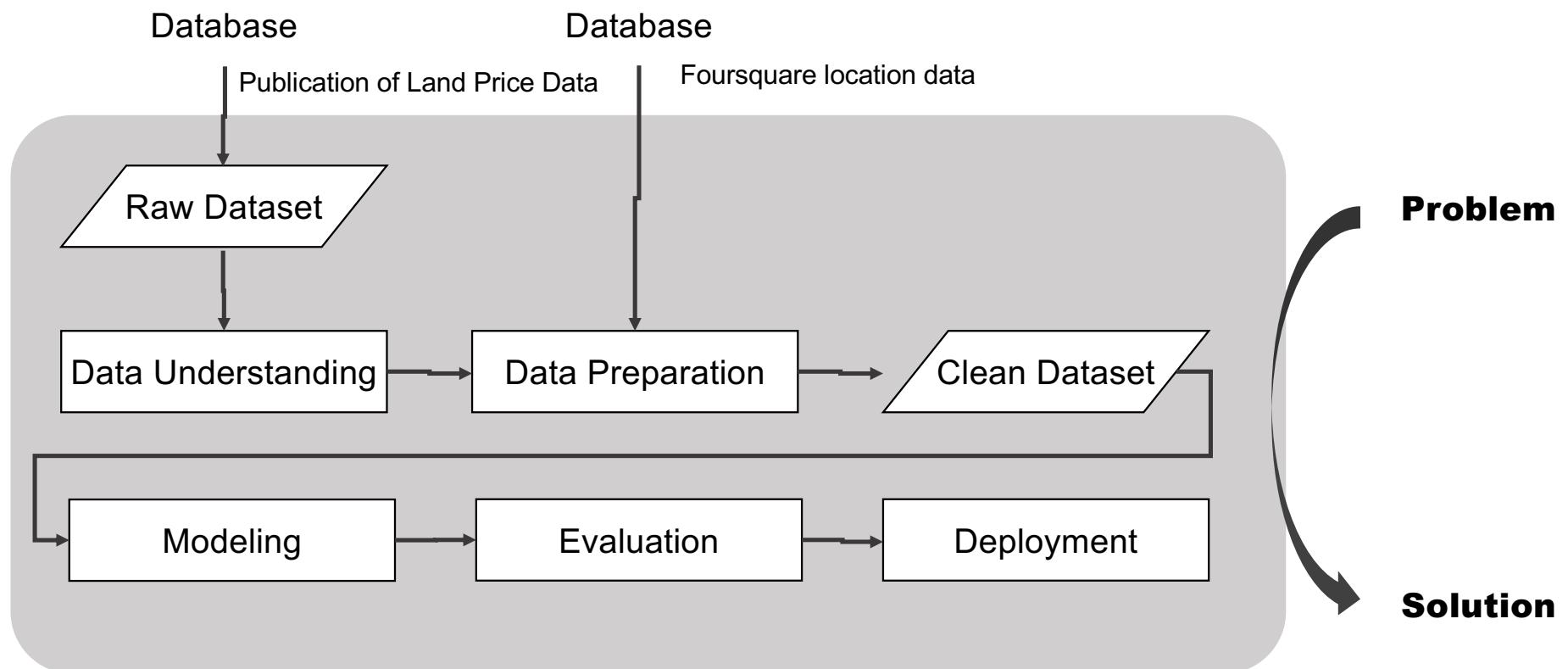
Problem to be Solved

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 own land to live is suggested.

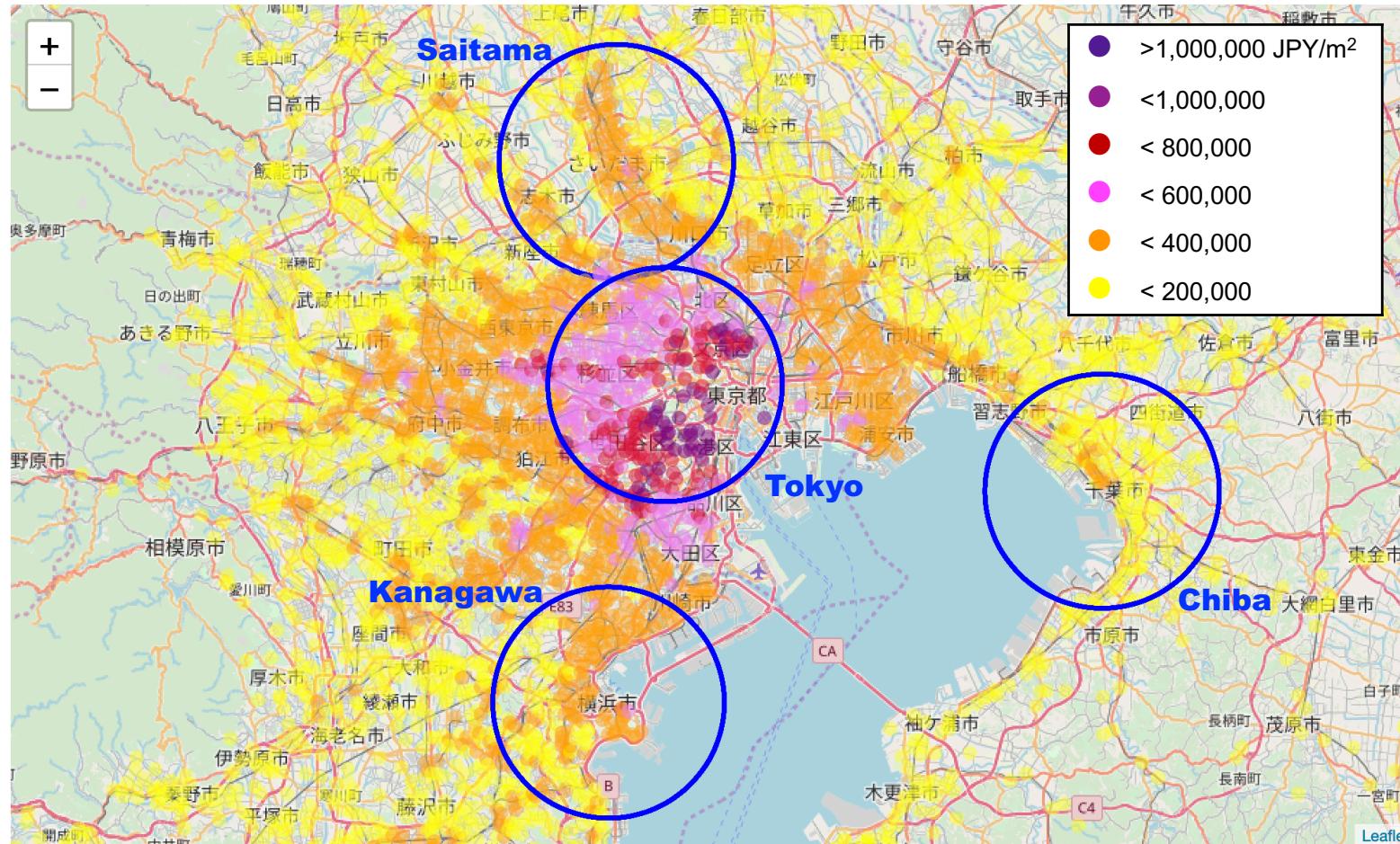
Data Sources

Land Attribution Data	Data Source	Description
Area name	National Land Numerical Information: Publication of Land Price Data http://nlftp.mlit.go.jp/ksj/index.html	Various GIS data of each prefecture is provided. Shape file format, GML, and GeoJSON format is available.
Latitude		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.
Longitude	Provided by Ministry of Land, Infrastructure, Transport and Tourism.	
Land price (JPY/m ²)		
Nearest station name		
Distance to nearest station (m)		
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 (Km)	Coordinates of each landmark: OpenStreetMap data	Obtaining latitude and longitude of each landmark and calculated the distance between each position in the dataset using Haversine Formula.
Distance to city hall of the capital of each prefecture (Km)	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	

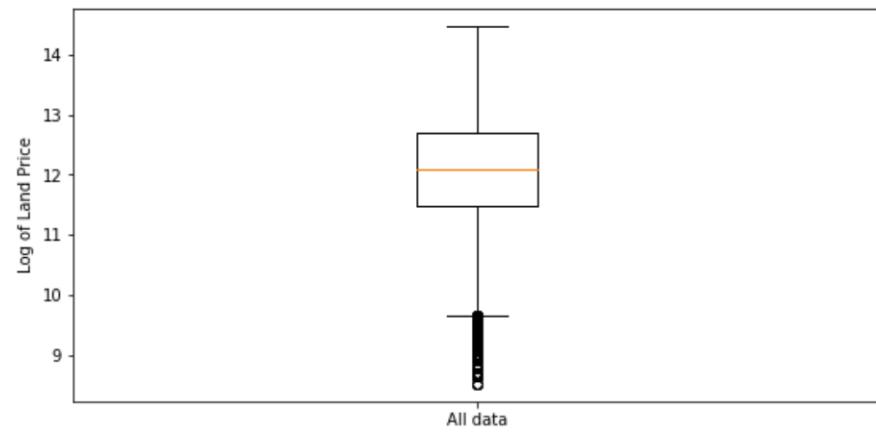
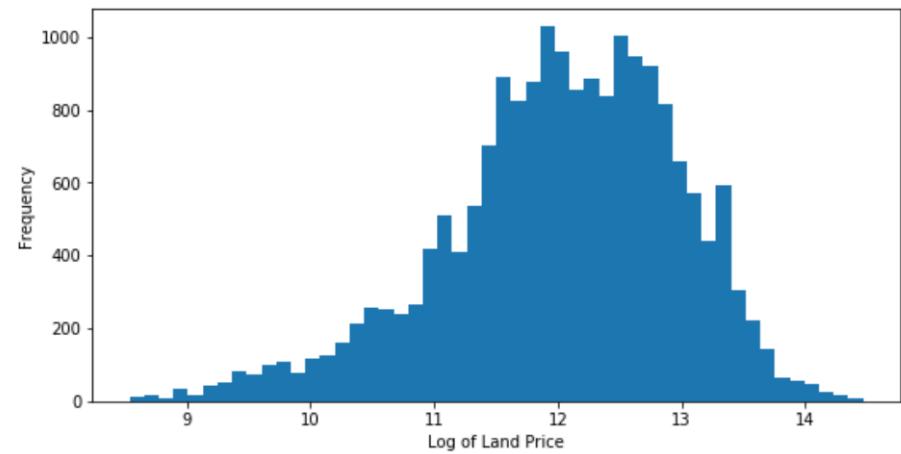
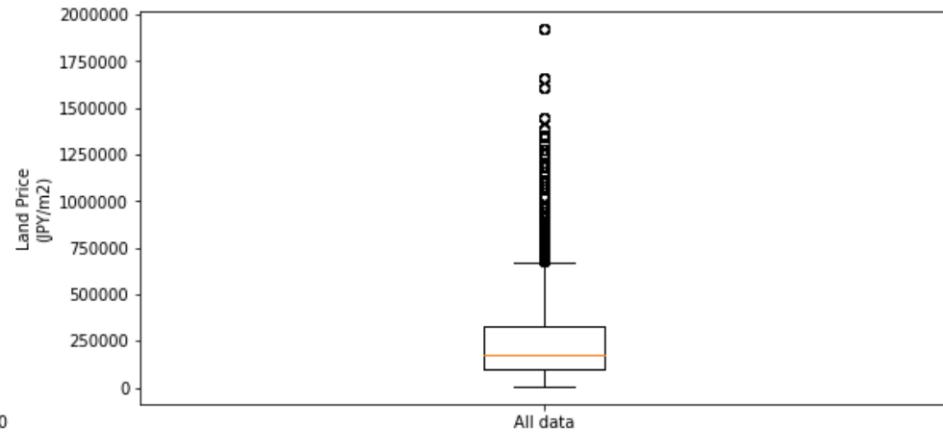
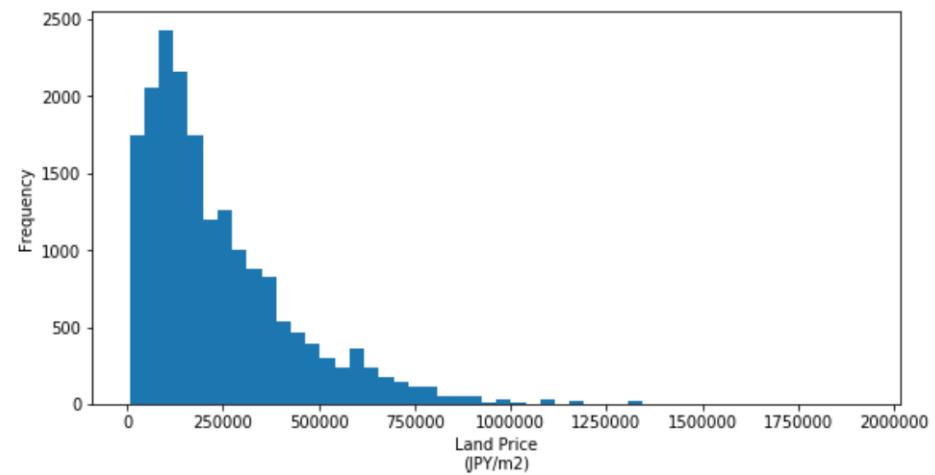
Summary of Workflow



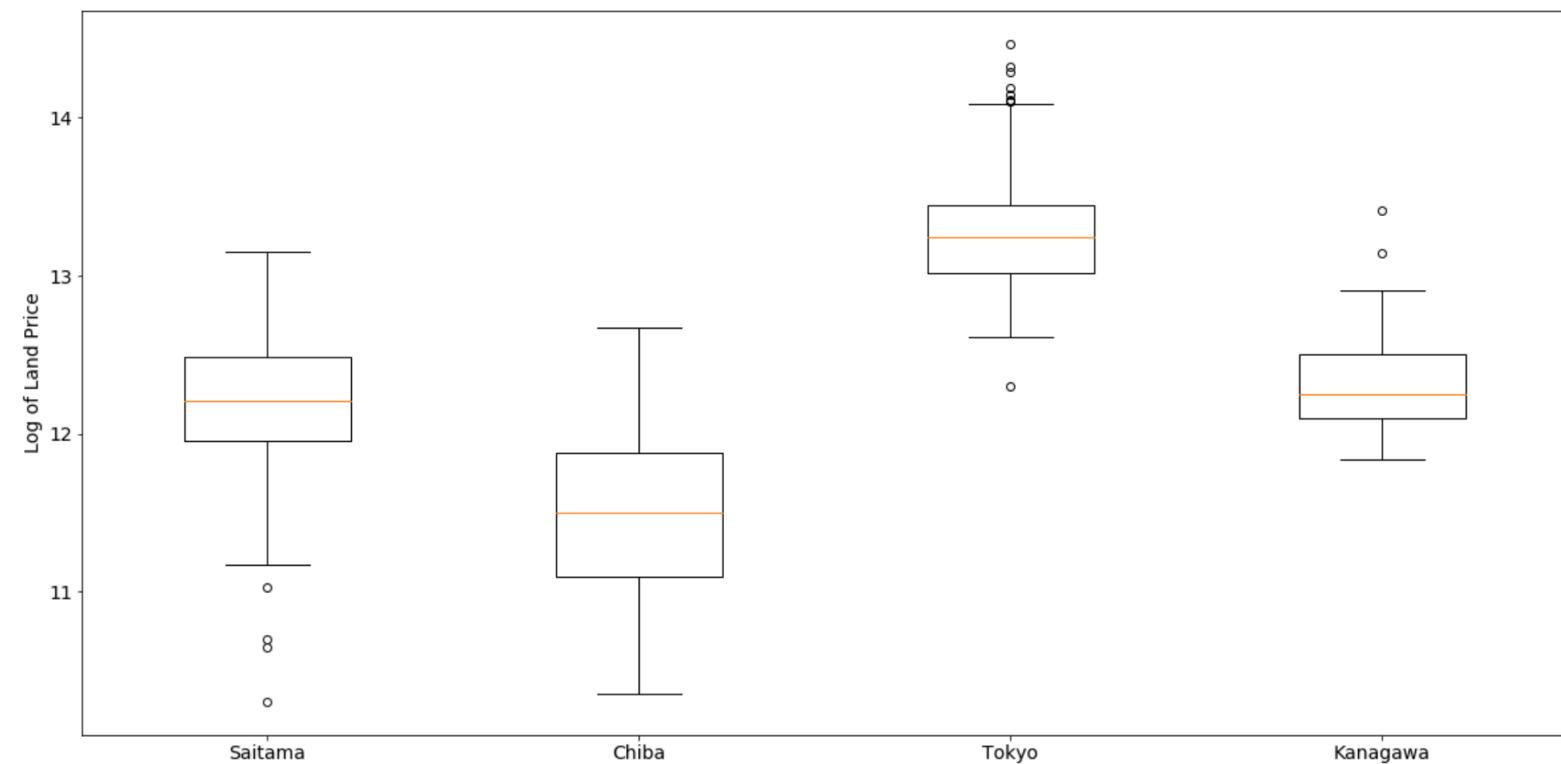
Data Understanding: Mapping Data



Data Understanding: Data Distribution

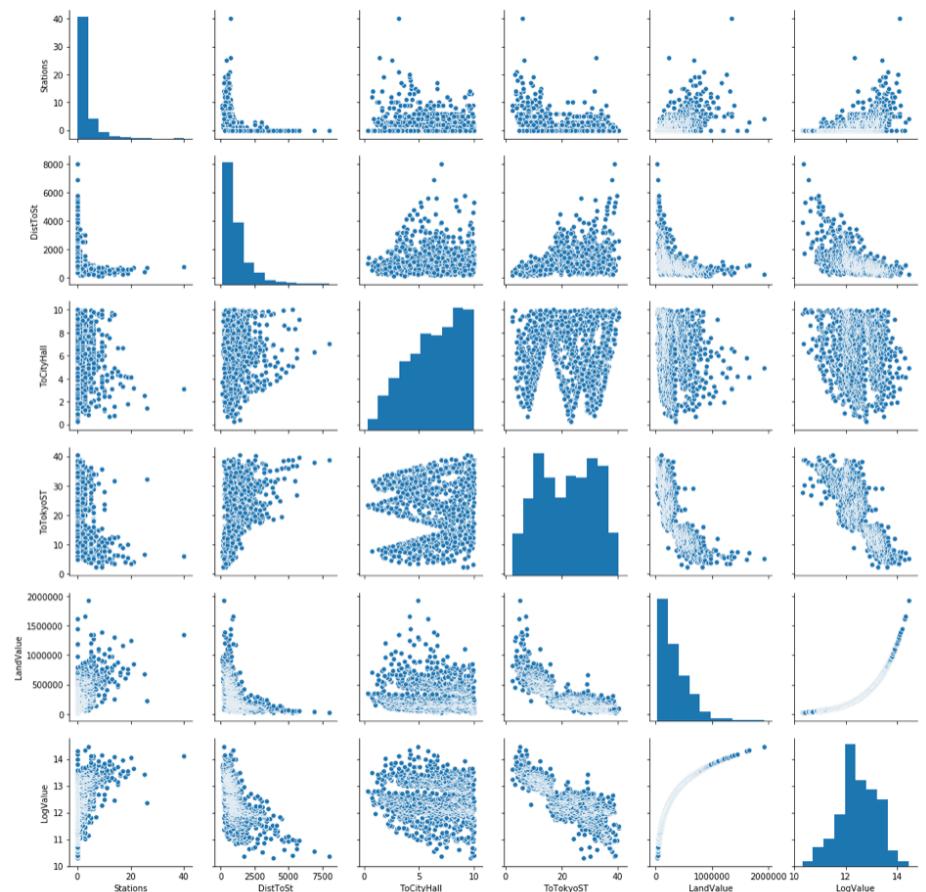


Data Preparation: Description of Clean Data

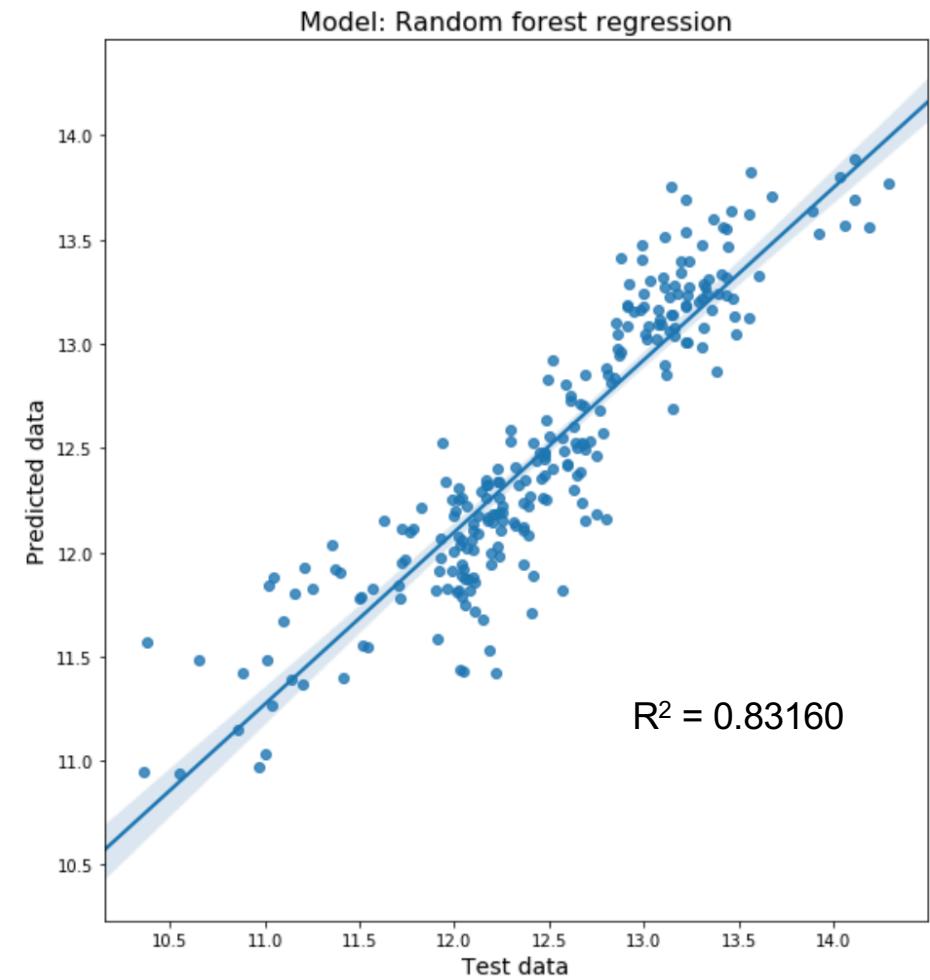
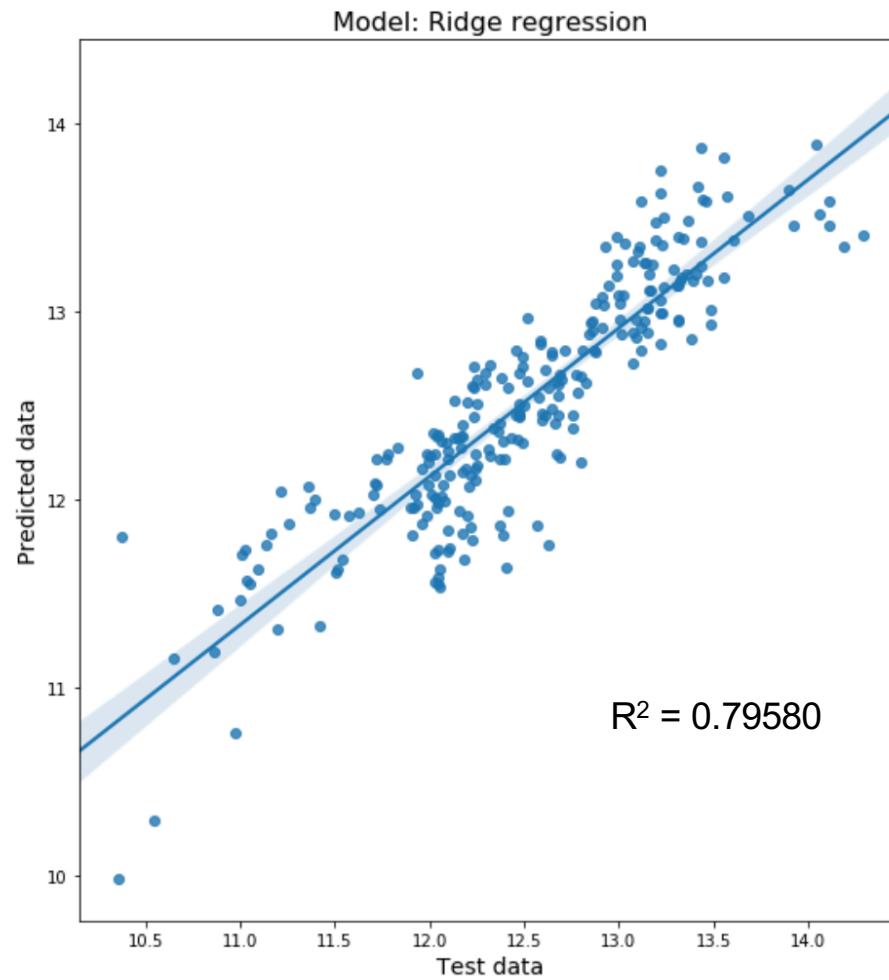


Modeling: Correlation of Each Variables

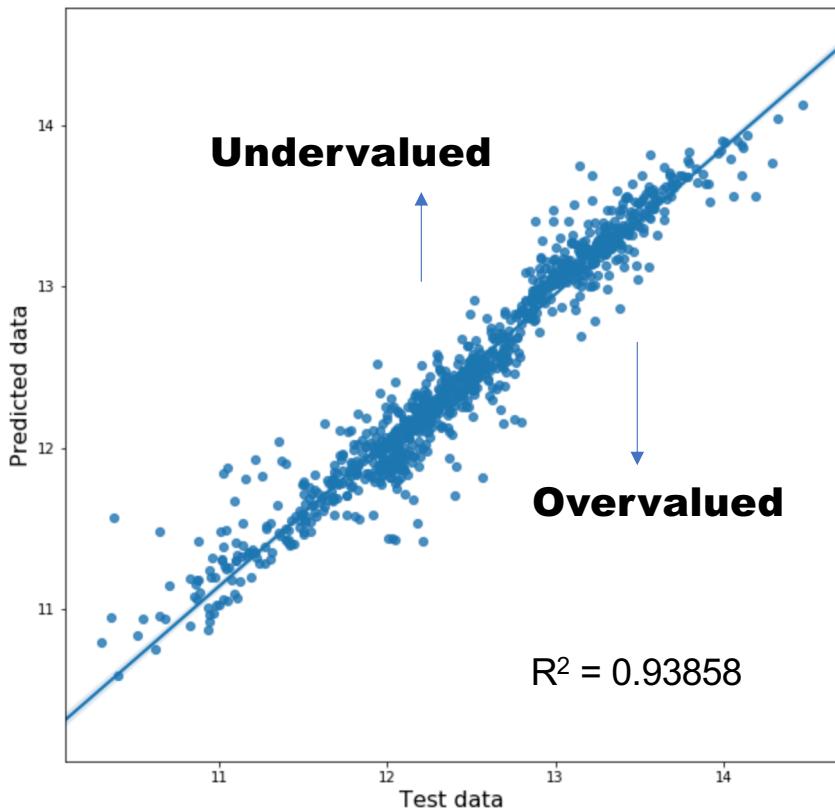
	City	Latitude	Longitude	Stations	Station	DistToSt	ToCityHall	ToTokyoST	LandValue	
0	ふじみ野	35.872763	139.537278	0	ふじみ野	2500	9.823200	29.884129	134000	
1	富士見	35.855106	139.537822	0	鶴瀬	1100	9.723353	28.478432	165000	
2	富士見	35.859945	139.540321	0	鶴瀬	1900	9.472695	28.684975	142000	
3	富士見	35.840844	139.542992	0	鶴瀬	1000	9.515581	27.070457	181000	
4	三芳	35.826179	139.544807	0	みずほ台	1600	9.887371	25.899016	74000	
...	
220	横浜鶴見	35.505880	139.686688	0	京急鶴見	900	8.181897	20.806649	287000	
221	横浜鶴見	35.512986	139.692186	0	鶴見市場	900	9.115972	19.891581	284000	
222	川崎	35.514899	139.694512	2	川崎	2500	9.408876	19.619226	279000	
223	川崎	35.504901	139.694512	2	川崎	2500	9.408876	19.619226	279000	
224	川崎	35.509975	139.694512	2	川崎	3900.0	9.679512	27.572543	29800.0	10.302264
1019 rows × 9 columns										
1				0	8000.0	7.039010	38.625657	31500.0	10.357743	
2				0	1700.0	9.821689	30.219071	32000.0	10.373491	
3				0	5300.0	9.997029	38.890954	32800.0	10.398184	
4				0	3100.0	9.840500	37.759704	36800.0	10.513253	
...				
1014				9	310.0	6.370946	5.811658	1390000.0	14.144814	
1015				0	720.0	4.176302	7.417678	1450000.0	14.187074	
1016				0	840.0	5.832020	4.778504	1610000.0	14.291745	
1017				3	900.0	4.124020	7.274164	1660000.0	14.322328	
1018				4	250.0	4.901122	5.065084	1920000.0	14.467836	
1019 rows × 6 columns										



Modeling: Evaluation of Models



Deployment: Prediction of Land Price

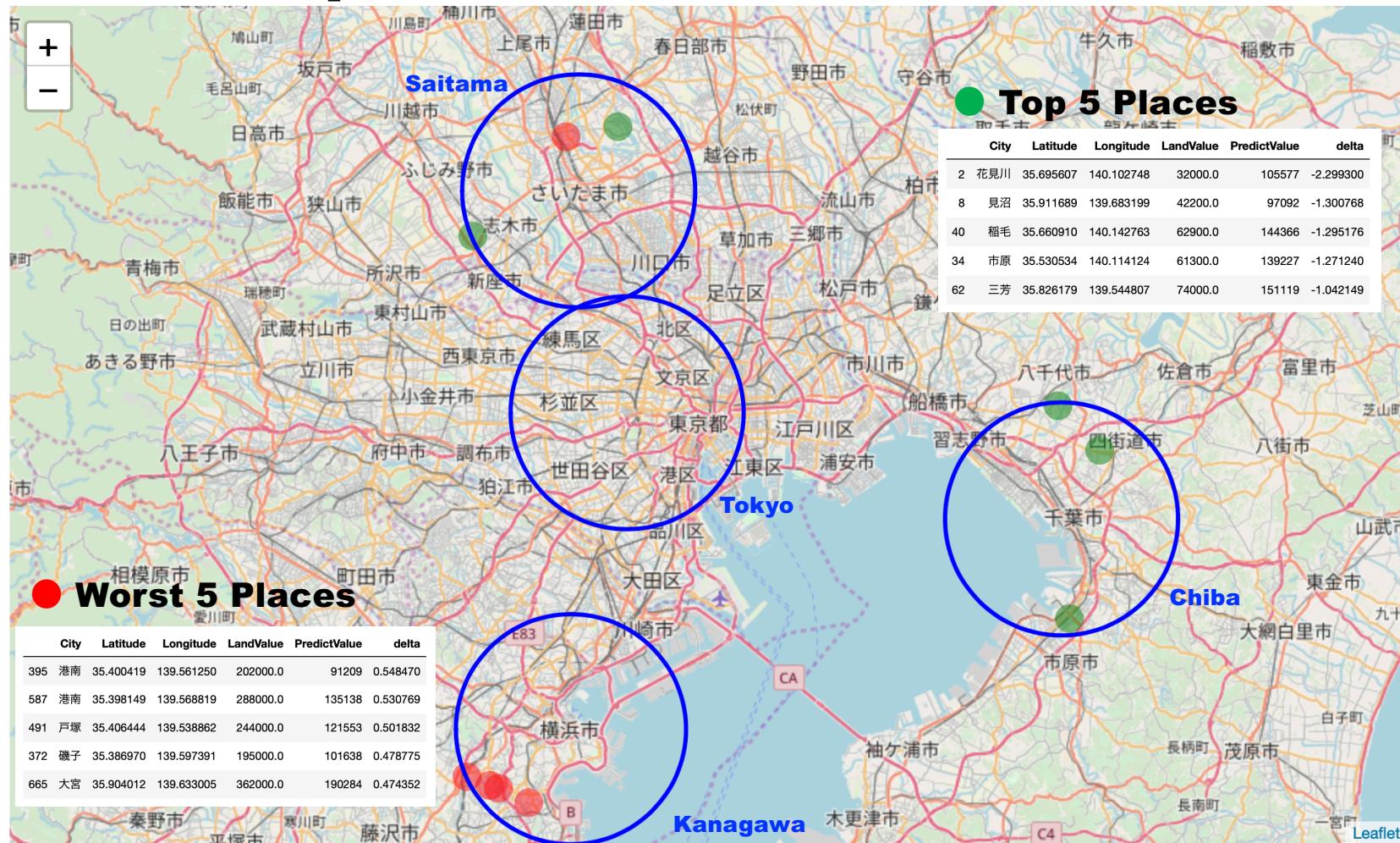


City	Latitude	Longitude	Stations	Station	DistToSt	ToCityHall	ToTokyoST	LandValue	PredictValue	delta
2 花見川	35.695607	140.102748	0	八千代台	1700.0	9.821689	30.219071	32000.0	105577	-2.299300
8 見沼	35.911689	139.683199	0	大宮	5600.0	6.530860	26.834771	42200.0	97092	-1.300768
40 稲毛	35.660910	140.142763	0	四街道	2300.0	6.796555	33.866578	62900.0	144366	-1.295176
34 市原	35.530534	140.114124	0	八幡宿	1000.0	8.572809	35.415164	61300.0	139227	-1.271240
62 三芳	35.826179	139.544807	0	みずほ台	1600.0	9.887371	25.899016	74000.0	151119	-1.042149
...
665 大宮	35.904012	139.633005	0	大宮	1100.0	4.847476	27.701503	362000.0	190284	0.474352
372 磯子	35.386970	139.597391	0	洋光台	1200.0	7.326629	36.159718	195000.0	101638	0.478775
491 戸塚	35.406444	139.538862	0	戸塚	920.0	9.906969	36.919170	244000.0	121553	0.501832
587 港南	35.398149	139.568819	5	上永谷	600.0	8.078220	36.258996	288000.0	135138	0.530769
395 港南	35.400419	139.561250	3	下永谷	450.0	8.475124	36.389084	202000.0	91209	0.548470

Undervalued

Overvalued

Results: Top 5 and Worst 5



Conclusion

Publicly disclosed land price data in Tokyo Metropolitan Area is examined in order to suggest the most affordable place to own land to live. Comparison to predicted land price data obtained by machine learning model with random forest regression shows Top 5 under evaluated places, and it is found out that Chiba prefecture is generally under evaluated and affordable.