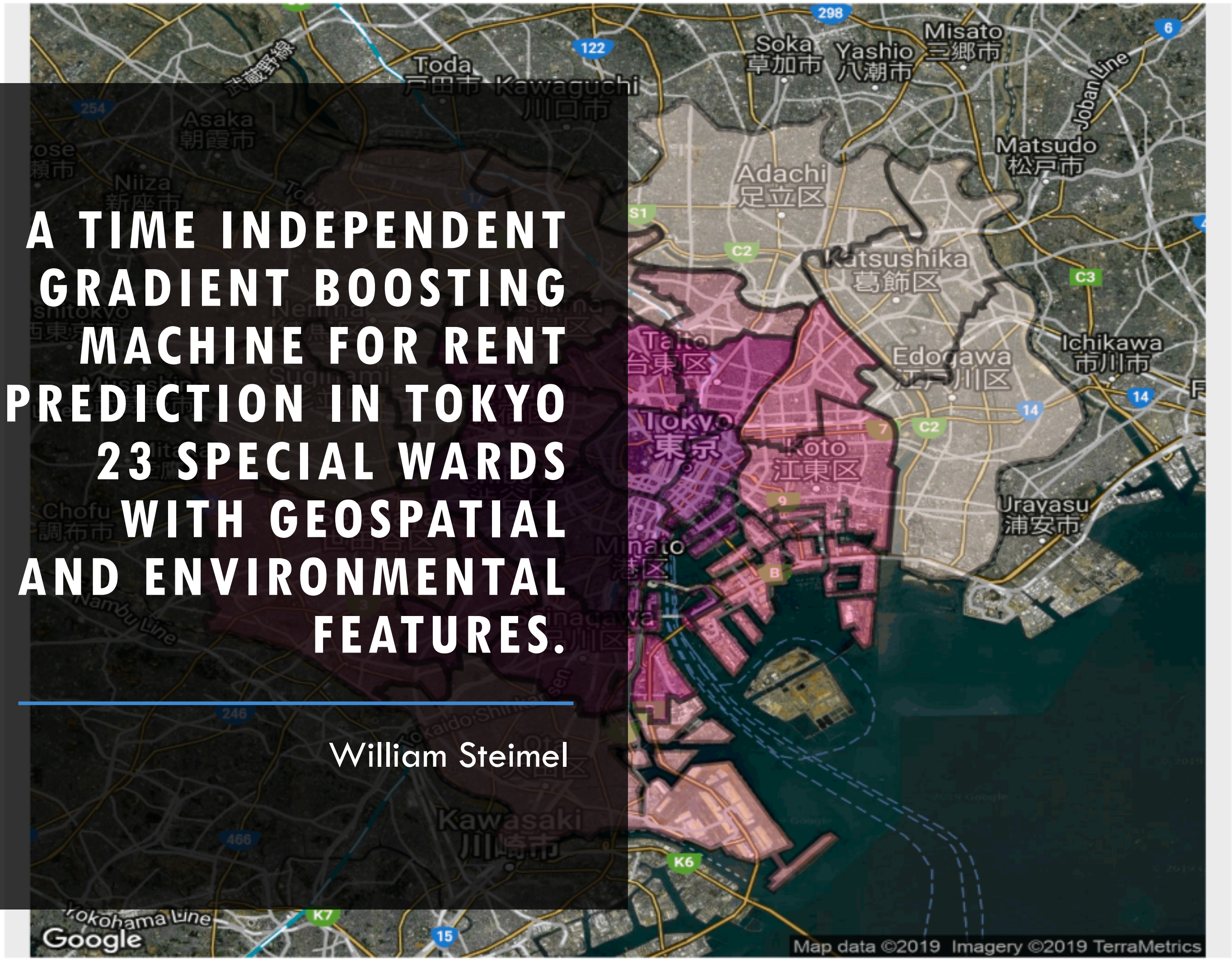


A TIME INDEPENDENT GRADIENT BOOSTING MACHINE FOR RENT PREDICTION IN TOKYO 23 SPECIAL WARDS WITH GEOSPATIAL AND ENVIRONMENTAL FEATURES.

William Steimel

SOPHIA UNIVERSITY
MASTERS OF SCIENCE IN
GREEN SCIENCE AND ENGINEERING
YUICHIRO MIYAMOTO

TABLE OF CONTENTS

TABLE OF CONTENTS

INTRODUCTION

This research seeks to build a time independent predictive model that can accurately **predict rent** in **Tokyo** through the combination of a **rental** listings descriptive features as well as **environmental** and **geospatial features** at the administrative district and Ku level.

Rent is considered one of the **biggest housing expenditures** for people and also has a **huge impact** on the **economy.**

According to the United Nations, 55 % of the world's **population** live in **urban areas** and is **expected to increase** to 68 % by 2050.

Real estate estimation algorithms are going to be increasingly important:

- public policy developers
- urban planners
- real estate companies
- banks

Real estate price prediction algorithms have recently **disrupted traditional** methods for **pricing**

Companies:

- Zillow
- GeoPhy
- PriceHubble
- Suumo

Zillow Competition: $ 1,000,000 Grand Prize Zestimate

RESEARCH QUESTIONS

Machine Learning model performance is highly influenced by the features used as inputs.

Goal 1- This research wishes to test the viability of environmental and geospatial features and their effectiveness in contributing to rent price estimation in Tokyo's rental market.

Goal 2- Feature engineering techniques will also be experimented with against real estate, geospatial, and environmental features to build the optimal rent estimation model.

- Feature Engineering is the process of generating features through the transformation of variables to augment predictive accuracy.

REAL ESTATE MARKET MODELING CHALLENGES

- **Many Variables (features)**
  - Location
  - Size
  - Rooms
  - Transportation
  - Amenities
- **Each Geographic market is Different**
  - New York vs Tokyo- different factors that influence rent.
- **Price Fluctuates**
  - Temporal Aspect- Seasonal Supply/Demand
- **Multiple Market Segments**
  - Luxury Housing vs Cheap Housing

Machine Learning formulated as a **Regression** problem can be used to address some of these challenges.

LITERATURE REVIEW

Three Perspectives for Real Estate Estimation

- Traditional Econometric Approaches
- Machine Learning Approaches
- Tokyo Real Estate Market Research

LITERATURE REVIEW – TRADITIONAL APPROACHES REAL ESTATE PRICING MODELS

Traditional approaches for Real Estate Estimation relied on Hedonic Modeling:

- Hedonic Regression
  - Proposed in Sherin Rosen’s paper “Hedonic Pricing and Implicit Markets: Product Differentiation in Pure Competition” (1974)

$$y = f(x) + \epsilon$$

  - It breaks a property or real estate property up into each of its descriptive characteristics to establish a relationship between these characteristics and price.
    - Highly interpretable
    - Multivariate Regression
  - Still used often in Econometric Research - Has been applied recently to real estate markets like Singapore and Turkey (2014, 2019)
- Hedonic Regression –**Machine Learning** vs **Econometric** Thought (2017)
  - **Machine Learning** – produce predictions y based on input features x
  - **Econometrics** – Focuses on Parameter Estimation where the goal is not prediction but to understand how x and y are related.

Traditional **Hedonic Methods** are **effective** at **modeling relationships** but state of the art **machine learning** techniques may be **more suited** for more **accurate prediction**.

LITERATURE REVIEW – MACHINE LEARNING APPROACHES

- Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data (2015)

- Identifying Real Estate Opportunities Using Machine Learning (2018)
  - Samalanca district Madrid, Spain

- Housing Price Prediction using machine learning algorithms: The case of Melbourne city, Australia (2019)

- Ensemble Learning Based Rental Apartment Price Prediction Model by Categorical Features Factoring (2019)

So far no papers focus on **LightGBM** from a Gradient Boosting perspective. Most papers **focus** on **algorithm comparison** rather than the **impact of features** on **predictive accuracy**.

LITERATURE REVIEW – TOKYO HOUSING MARKET

- Spatial Analysis of Tokyo apartment Market (Jan 2007)
- Earthquake Risk and housing rents: Evidence from the Tokyo Metropolitan Area (January 2007)
- Do Urban Amenities Drive Housing Rent? (July 2014)
- Measuring the spatial effect of multiple sites: An application to housing rent and public transportation in Tokyo, Japan (May 2018)

No studies to my knowledge **use** state-of-the-art **machine learning** techniques to **predict rent** from the context of the **Tokyo** Housing Market. Interesting feature ideas were gained from these studies.

TABLE OF CONTENTS

DATASETS

This research builds a dataset with environmental and geospatial features from multiple sources to predict rent in Tokyo.

Dataset	Source	Description	Join Key
Suumo	Suumo	Apartment and mansion listings for 23 Ku's in Tokyo	location/Ku
Google Cloud API's	Google Cloud	Geospatial coordinates, distance and time calculations to hub stations	location
Tokyo Regional Earthquake Risk Survey	Tokyo Bureau of Urban Development	Earthquake risk by administrative district in Tokyo (chome)	location
Tokyo Air Quality	Tokyo Ministry of the Environment	Air quality by sensor in Tokyo	location
Tokyo Parks	Tokyo Bureau of Construction	Park statistics by Ku	Ku
Tokyo Crime Data	Tokyo Metropolitan Police	Crime statistics by administrative district in Tokyo (chome)	location
Land Use by District	Tokyo Bureau of Urban Development	Land use classification breakdown in Tokyo by Ku	Ku
Complaints about Pollution by Kind	Tokyo Ministry of the Environment	Total complaints about pollution by Ku	Ku

SUUMO DATA

What is Suumo? –is an online platform in Japan for housing, real estate buying and selling, and rental support information.

Sample Listing from Dataset- Suumo Website

Table 1: Suumo (base) dataset

	Feature	Description
1	title	Suumo listing description
2	Ku	Tokyo Ku that the unit is located in
	location	Administrative district (Also known as chome)
3	station_1	Closest station and travel time
4	station_2	Second closest station and travel time
5	station_3	Third closest station and travel time
6/7	yrs	Age of the building in years
	heights	Height/depth of the building
8	floor	Floor the rental unit is on
9/10	rent	Monthly rent
	admin	Monthly administrative Fee
11/12	deposit	Deposit fee when starting a contract
	gratuity	Gratuity fee when starting a contract
13/14	floor_plan	Room layout of rental unit
	area	Total space in meters squared

Feature Dictionary

title	location	station1	station2	station3	yrs	heights	floor	rent	admin	deposit	gratuity	floor_plan	area	Ku
コーポ碑文谷	東京都目黒区碑文谷２	東急東横線/学芸大学駅 歩10分	東急東横線/都立大学駅 歩18分	東急目黒線/西小山駅 歩22分	築50年	4階建	4	14.4万円	6000円	14.4万円	-	1LDK	47.83m2	Meguro
ボナール	東京都練馬区田柄１	東京メトロ有楽町線/平和台駅 歩12分	都営大江戸線/練馬春日町駅 歩15分	東京メトロ有楽町線/地下鉄赤塚駅 歩20分	築28年	3階建	3	10.5万円	-	10.5万円	-	2DK	43.74m2	Nerima
東京メトロ千代田線 千駄木駅 11階建 築28年	東京都文京区千駄木２	東京メトロ千代田線/千駄木駅 歩2分	ＪＲ山手線/日暮里駅 歩13分	東京メトロ千代田線/根津駅 歩9分	築28年	11階建	5	16万円	-	32万円	16万円	2DK	44.56m2	Bunkyo
ＪＲ山手線 目白駅 14階建 築21年	東京都豊島区雑司が谷３	ＪＲ山手線/目白駅 歩8分	東京メトロ副都心線/雑司が谷駅 歩1分	ＪＲ山手線/池袋駅 歩13分	築21年	14階建	7	7.5万円	5000円	7.5万円	7.5万円	1K	20.47m2	Toshima
日暮里・舎人ライナー 熊野前駅 10階建 築30年	東京都荒川区東尾久３	日暮里・舎人ライナー/熊野前駅 歩2分	東京メトロ千代田線/町屋駅 歩16分	都電荒川線/宮ノ前駅 歩9分	築30年	10階建	3	8.5万円	5000円	8.5万円	8.5万円	2DK	41.4m2	Arakawa

Five Row Sample Before Pre-Processing

GOOGLE MAPS API - GEOCODING

"**Geocoding** is the process of converting addresses into geographic coordinates (latitude and longitude)"

Map for 35.63392, 139.7156

Example:

- 目黒駅 – Meguro Station
  - latitude 35.63392 and longitude 139.7156

Table 2: Google Maps Geocoding Data (Location)

Feature	Description
location	Administrative district (link to base data on location)
LAT	Latitude
LON	Longitude

GOOGLE MAPS API – DISTANCE MATRIX

The Distance Matrix API was used to calculate time (minutes) and distance (km) between all stations and Tokyo's major hub stations. (Via Google Maps)

- Tokyo
- Shinjuku
- Shibuya
- Ikebukuro
- Ueno
- Shinagawa

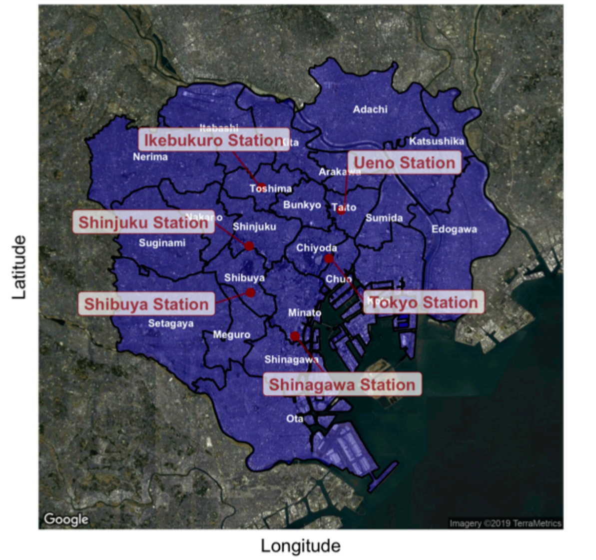
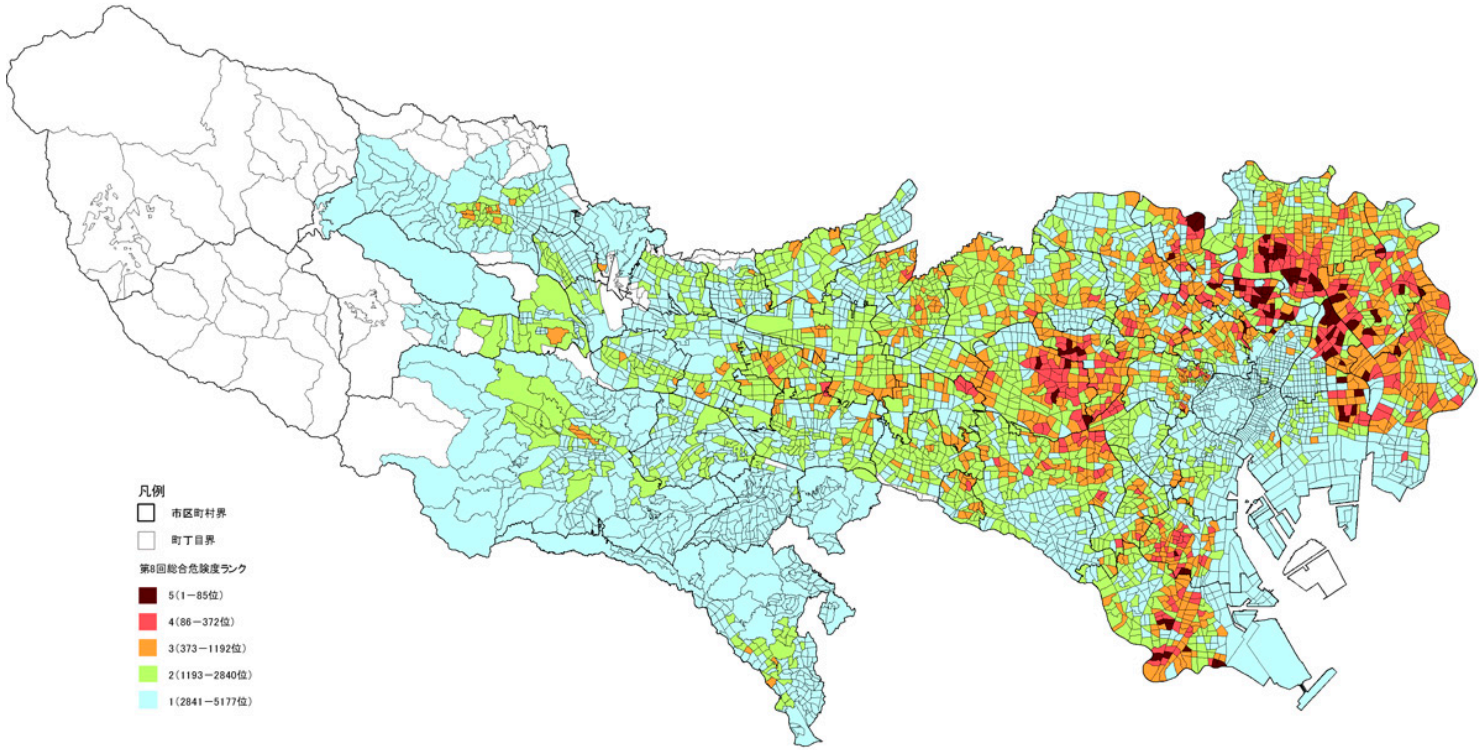


Figure 2: Tokyo Hub Stations

Table 3: Google Maps Distance Matrix Data

Feature	Description
Stations	Station name (link to base data on Station_1)
LAT	Latitude
LON	Longitude
ShinjukuMade	Time to Shinjuku
ShinjukuDist	Distance to Shinjuku (km)
TokyoMade	Time to Tokyo
TokyoDist	Distance to Tokyo (km)
ShibuyaMade	Time to Shibuya
ShibuyaDist	Distance to Shibuya (km)
IkebukuroMade	Time to Ikebukuro
IkebukuroDist	Distance to Ikebukuro (km)
UenoMade	Time to Ueno
UenoDist	Distance to Ueno (km)
ShinagawaMade	Time to Shinagawa
ShinagawaDist	Distance to Shinagawa (km)



Source: Tokyo Bureau of Urban Development

TOKYO REGIONAL EARTHQUAKE RISK SURVEY

- The Tokyo Regional Earthquake Risk Survey from the Tokyo Bureau of Urban Development details the risk posed to each administrative district in the event of an earthquake.
- Tokyo surveys earthquake risk in 5,177 administrative districts to support planning and disaster resilience.

Table 4: Tokyo Regional Earthquake Survey

Feature	Description
Land Classification	Land classification by administrative district
Building_Risk	Risk of building destruction
Building_Risk_Rank	Ranking in dataset for building destruction risk
Building_Risk_Level	Level of risk to buildings which is evaluated relative to other districts
Fire_Risk	Risk of fire
Fire_Risk_Rank	Ranking in dataset for fire risk
Fire_Risk_Level	Level of risk to fire which is evaluated relative to other districts
Life_Difficulty_Risk	Risk to livelihood in event of earthquake based on access to everyday services.
Life_Difficulty_Risk_Rank	Ranking in dataset for livelihood risk
Life_Difficulty_Risk_Level	Level of risk to livelihood which is evaluated relative to other districts
Total_Risk	Overall risk
Total_Risk_Rank	Overall risk ranking in dataset
Total_Risk_Level	Level of overall risk which is evaluated relative to other districts

TOKYO ENVIRONMENT - ATMOSPHERIC REGIONAL OBSERVATION SYSTEM: AEROS

Atmospheric Regional Observation System AEROS was used to create air quality features for every Administrative District (chome).

- February 2018 – February 2019 Data
- 33 Sensors
- Euclidean Distance
  - Linked to Chome (Administrative District) by closest sensor.
- Time Series Data (10 minute intervals) - Aggregates calculated for each measurement (min, mean, median, max, std)

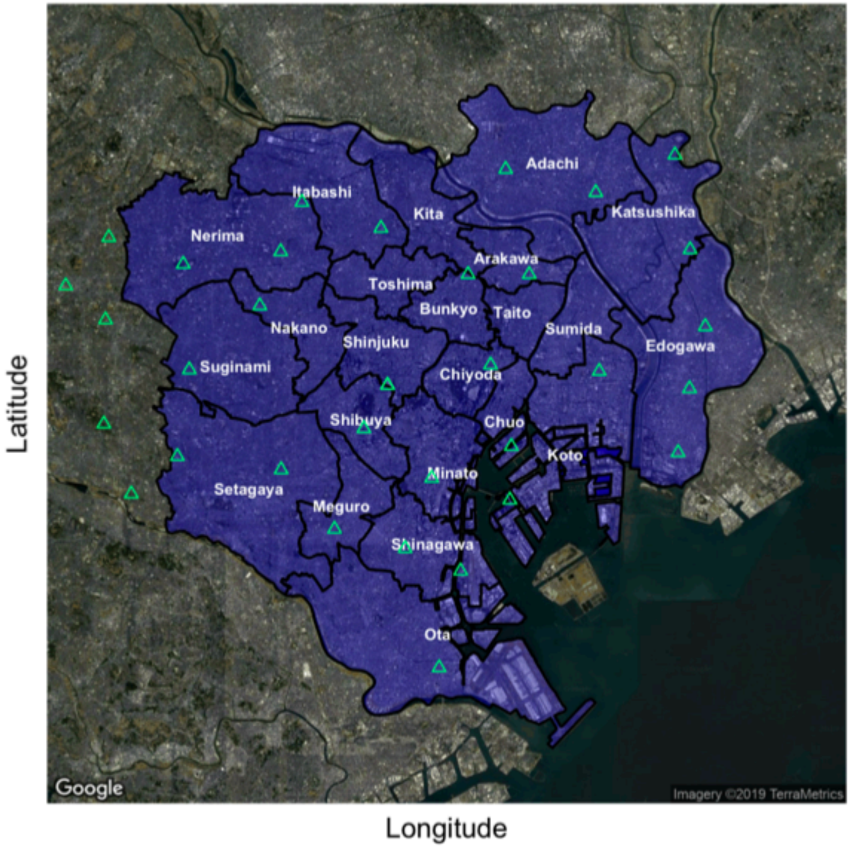


Figure 3: Tokyo Air Quality Sensors

Table 5: Tokyo Air Quality

Feature	Description	Aggregates (1yr)	Total Features
SO2 (ppm)	Sulfur Dioxide	min, median, mean, max, std	5
NO (ppm)	Nitrogen Monoxide	min, median, mean, max, std	5
NO2 (ppm)	Nitrogen Dioxide	min, median, mean, max, std	5
Ox (ppm)	Photochemical Oxidant	min, median, mean, max, std	5
NMHC (ppmC)	Non-Methane Hydrocarbons	min, median, mean, max, std	5
SPM (mg/m3)	Suspended Particulate Matter	min, median, mean, max, std	5
PM2.5 (ug/m3)	Fine Particulate Matter	min, median, mean, max, std	5
WS (m/s)	Wind Speed	min, median, mean, max, std	5
TEMP	Temperature	min, median, mean, max, std	5
HUM	Humidity	min, median, mean, max, std	5
WD	Wind Direction Counts by direction	count by 18 wind directions	18

Table 6: Tokyo Parks

Category	Feature Group	Aggregates	Total Features
General	Area	Total area for Ku (hectares)	3
	Population	Total population for Ku	
	Population Density	Total area per person	
Urban Parks	Metropolitan Parks	Totals, area (hectares)	10
	Municipal Parks	Totals, area (hectares)	
	National Government Parks	Totals, area (hectares)	
	Urban Parks (subtotal)	Totals, area (hectares)	
	Urban Parks (ratio)	Area per person, park to area ratio	
Non-Urban Parks	Seaside Parks	Totals, area (hectares)	8
	Municipal Parks	Totals, area (hectares)	
	Non-Urban Parks (subtotal)	Totals, area (hectares)	
	Non-Urban Parks (ratio)	Area per person, park to area ratio	
Municipal Parks	Public Parks (subtotal)	Totals, area (hectares)	4
	Public Parks (ratio)	Area per person, park to area ratio	
Other Parks	National Parks	Totals, area (hectares)	4
	Private Parks	Totals, area (hectares)	
Totals	Total (subtotal)	Totals, area (hectares)	4
	Total (ratio)	Area per person, park to area ratio	

TOKYO PARKS

- Parks in Tokyo are considered important for social gathering, cleaning the air, and evacuation sites during times of disaster making them vital to local communities.
- Parks data was collected from the Tokyo Bureau of Construction by Ku.
- The parks data details aggregates for the various types of parks including urban parks, non-urban parks, and natural parks as well as their subclassifications by Ku.

TOKYO CRIME

- Aggregated Crime Statistics for 2018 (Heisei 30) were collected for each administrative district in Tokyo from the Tokyo Metropolitan Police Department website.
- Features were translated from Japanese to English and contain totals and subcategory totals for:
  - Felonious crimes
  - Violent crime
  - Burglary and Larceny
  - Non-Intrusive Larceny
  - Other Crimes

Table 7: Tokyo Crime

Category	Features	Description	Total Features
Felonious Crime	Robbery	Total robbery	3
	Other	Total for other felonies	
	Total	Total all felonies	
Violent Crime	Rape	Total rapes	5
	Assault	Total assaults	
	Threat	Total threats	
	Blackmail	Total blackmail	
	Total	Total violent crime	
Burglary and Larceny	Safe_Theft	Total safe thefts (including banks)	9
	School	Total thefts at school	
	Office	Total thefts at office	
	FoodStall	Total thefts at food stall	
	House	Total thefts at house	
	burglary	Total break and enter thefts	
	Abandoned_house	Total thefts at abandoned property	
	Other	Total other burglary and larceny	
	Total	Total for burglary and larceny	
Non-Intrusive Larceny	Bike	Total bicycle theft	11
	Motorcycle	Total motorcycle theft	
	Car	Total motor vehicle theft	
	Vending_Machine	Total vending machine theft	
	Construction_Site	Total theft at construction sites	
	Pickpocket	Total pickpockets	
	purse_snatching	Total purse thefts	
	bag	Total bag thefts	
	shoplifting	Total shoplifting incidents	
	other	Total for other larceny	
	Total	Total non-intrusive larceny	
Other Crimes	Fraud	Total frauds	6
	Embezzlement	Total embezzlement	
	Intellectual Crime	Total intellectual crimes	
	Gambling	Total gambling offense	
	Other Offenses	Total other offenses	
	Total	Total other crimes	
Total	Total_Crimes	Aggregate total of crimes	1

Table 8: Tokyo Land Use by District

Feature	Description (Area in hectares)
Total	Total area
Residential	Residential Areas
OtherUse	Quarries, refuse dumping-grounds.
OtherUse(Open)	Storage space, parking lots, exhibition space, construction camps.
Parks	Parks, athletic fields, baseball grounds
Unused	Residential sites before construction, demolished sites, deserted buildings, reclaimed land
Roads	Urban roads, pavements, bicycle roads
roads2	Roads, railway tracks, monorail tracks, airports, seaports
Farm	Rice paddies, ordinary fields, orchards
Water	Rivers, canals, lakes, ponds
WoodsForest	Woodlands, bamboo groves
Fields	Grasslands and other uncultivated land

LAND USE BY DISTRICT (2016)

The Land Use By District details how land is used by Ku.

POLLUTION COMPLAINTS (2016)

Total Complaints of Pollution in each Ku

- Ministry of Environment

Table 9: Tokyo Pollution Complaints

Feature	Description
Total_Complaints	Total aggregate complaints
Air_Pollution	Air pollution related complaints
Water_Pollution	Water pollution related complaints
Soil_Pollution	Soil pollution related complaints
Noise	Noise complaints
Vibration	Vibration complaints
ground_subsidence	Ground subsidence complaints
Offensive_odors	Offensive odor complaints
other_pollution	Other complaints

DATASET CHALLENGES

Outliers

- Luxury and Cheap Rentals

Dirty Data

- Many data sources and text cleanup required.

Administration Fee Noise

- Adds noise to the target value

High Categorical Feature Cardinality

- Station, Administrative District, and Lines

Japanese/English Data

- Requires Translation and appropriate pre-processing

TABLE OF CONTENTS

DATA PRE-PROCESSING

Minimal pre-processing was done on all datasets which will be discussed in this section.

- Outlier removal
- Creation of Target value
- Train/Test/Validation split

OUTLIER REMOVAL

Upper and Low Outliers - **351 samples**

- > 800,000 yen per month
- < 23,000 yen per month

Rent per meter 5x higher or 5x lower than average for its age – **15 samples**

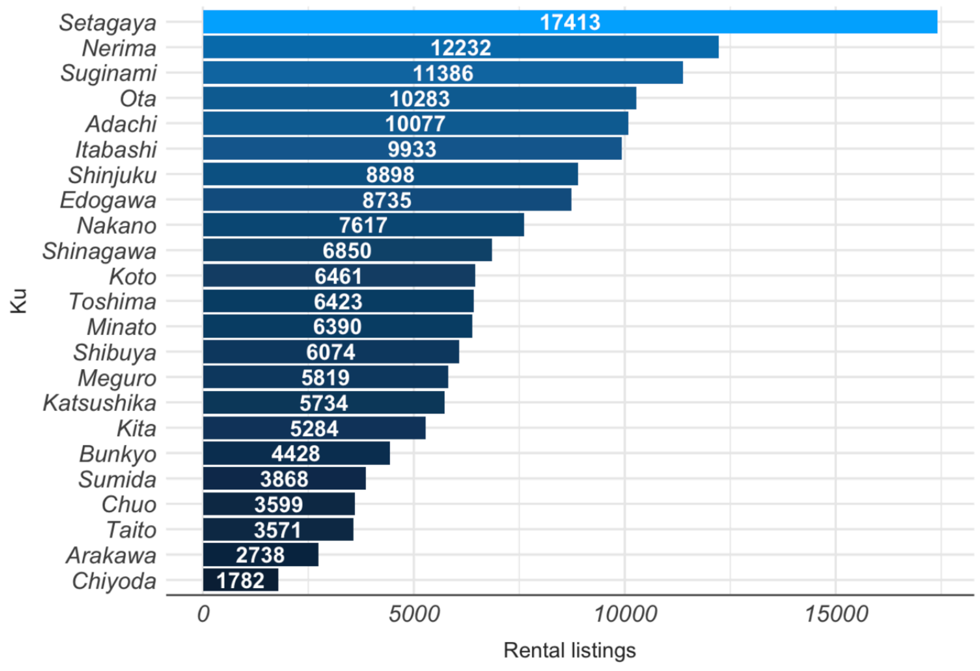
Apartments coded as 99 years – **20 samples**

Unbelievable Room Layouts - **2 samples**

- Example: 22DK

Total Removed – **388 Samples**

Remaining Samples in Data – **165,696**



TARGET VALUE

Target = Rent + Administration Fee

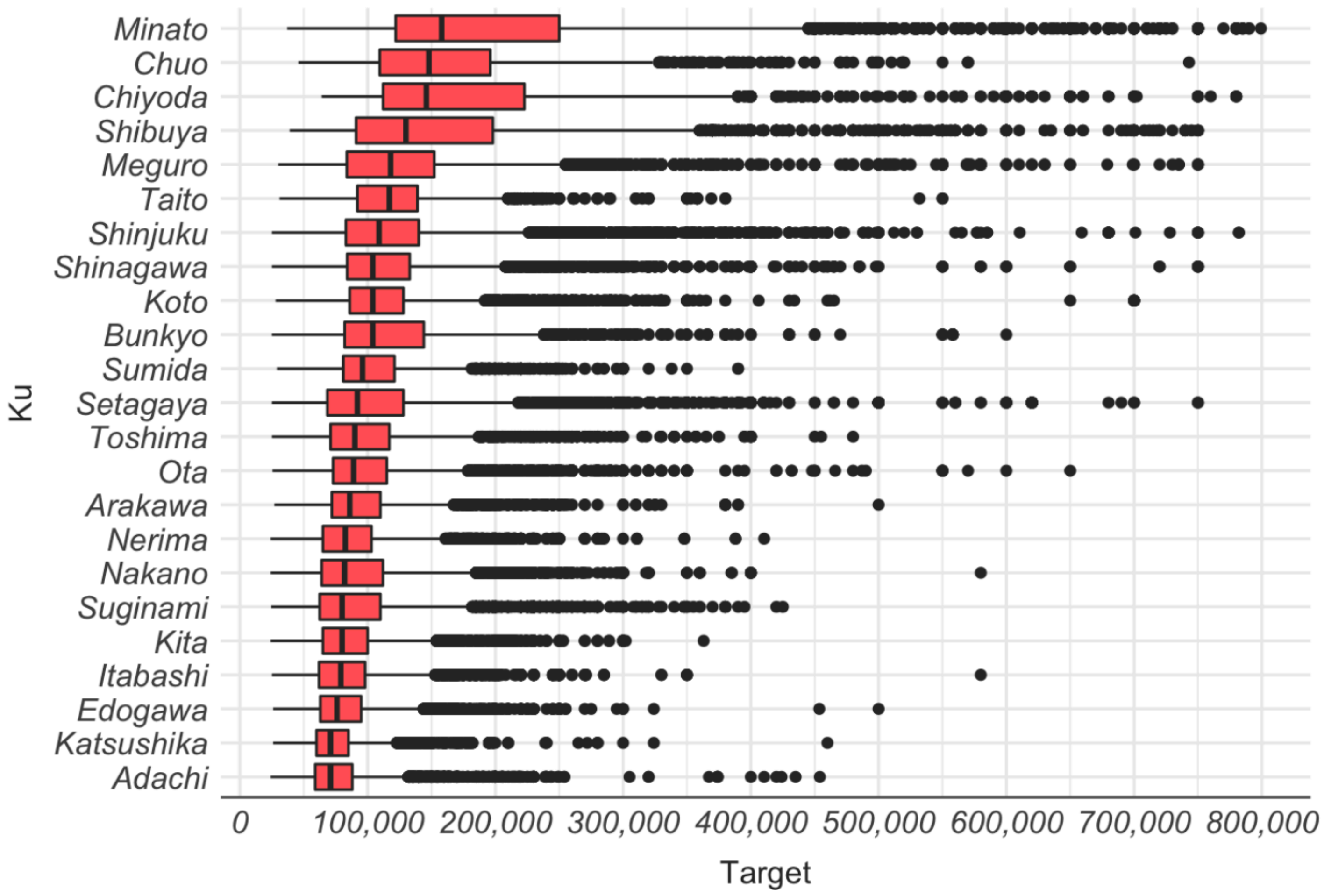


Figure 5: Price distributions of target value by Ku

TRAIN/TEST/VALIDATION SPLIT

To preserve the distribution the dataset was split into **15 price quantile bins** to perform **stratified splits.**

- Split 1- Entire Dataset split **80/20** to create **Training** & **Test Set**
- Split 2- Training Set split **80/20** to create a **Training** / **Validation Set**

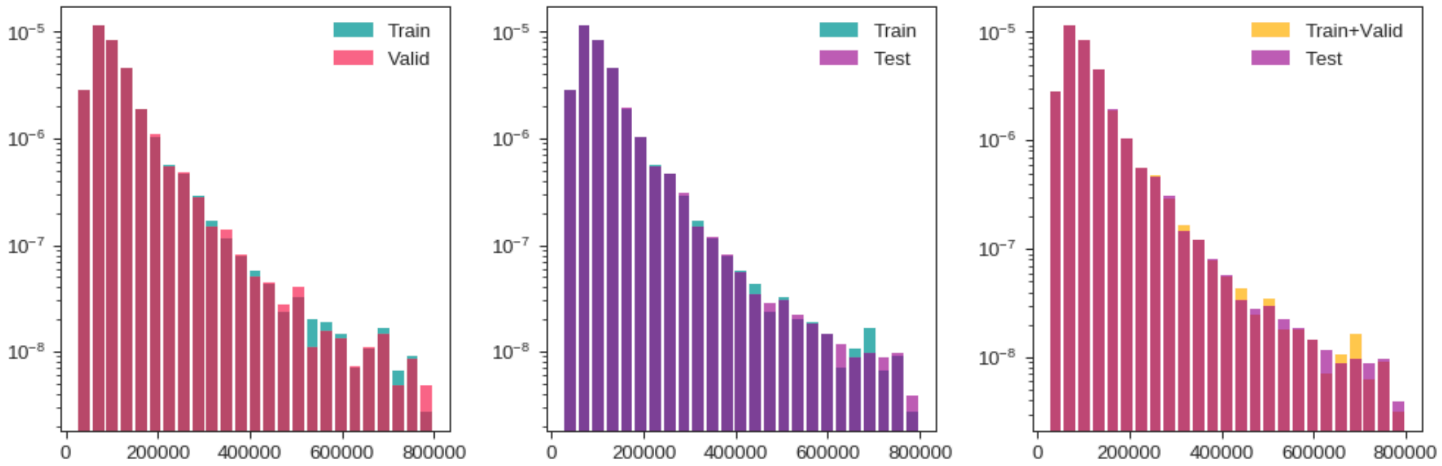


Figure 6: Distributions for Train, Validation, and Test sets (logarithm scale)

Table 10: Total Samples per Dataset and Target Descriptive Statistics

dataset	count	mean	std	min	25%	50%	75%	max
train	105,980	108,596.47	65,599.13	24,000	70,000	91,000	124,500	799,310
validation	26,496	108,593.61	65,337.07	24,500	70,000	91,000	125,000	782,000
test	33,119	108,555.13	65,454.43	24,000	70,000	91,000	125,000	790,000

TABLE OF CONTENTS

METHODOLOGY

This section will review the performance metrics, experimental setup, pre-processing steps, algorithm, and hyperparameter optimization methods used to conduct the experiments.

PERFORMANCE EVALUATION METRICS

Evaluation Metric 1:

- Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Evaluation Metric 2:

- Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{1}{n}\Sigma_{i=1}^{n}|\frac{(y_i - \hat{y}_i)}{y_i}| \times 100$$

Objective Function:

- Mean Squared Error (MSE)

$$MSE = \frac{1}{n}\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Evaluation Function: (LGBM Hyperparameter)

- Root Mean Squared Logarithmic Error (RMSLE)
- Helps control Early Stopping

$$RMSLE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}log(\hat{y}_i + 1) - log(y_i + 1)^2}$$

EXPERIMENTAL SETUP

Google Colaboratory

- Jupyter Notebook Environment
- GPU Tesla T4, 12.6 GB RAM, 320 GB of disk

- Three Experiments:
  - **Experiment 1: Baseline Model**
    - Suumo base features only
  - **Experiment 2: Geospatial and Environmental Model (Geo-Env)**
    - Suumo Base Features
    - Geospatial and Environmental Features
  - **Experiment 3: Feature Engineering Model (FE)**
    - Suumo Base Features
    - Geospatial and Environmental Features
    - Feature Engineering

MODEL PRE-PROCESSING-BASELINE MODEL

Kanji and Text Cleanup

- Kanji and text characters removed to create numeric features

Feature Transformation

- Binary Features
  - Deposit
  - Gratuity
  - Administrative Fee
- Features created by splitting other features
- Logarithm transform of area

Table 11: Baseline Model Features

Feature	Type	Pre-Processing	Definition
yrs	Numeric	Removed Kanji	Age in years of apartment
floor	Numeric	Removed Kanji, imputed highest floor value for multi-floor apartments	Floor of rental unit
deposit	Numeric	Binarized	1 if deposit fee
gratuity	Numeric	Binarized	1 if gratuity fee
area	Numeric	Removed meters squared, log transformed	Area in meters squared after log transform
admin_flag	Numeric	Binarized	1 if admin fee
new	Numeric	Binarized	1 if apartment is new
Basement_Depth	Numeric	Basement value inferred from Kanji in height feature	Building basement depth
Height	Numeric	Total height of building inferred from Kanji in height feature	Building height
entirehouse	Numeric	Binarized	1 if rental is entire house
mezanine	Numeric	Binarized	1 if rental is on mezanine floor
multifloor	Numeric	Binarized	1 if rental has multiple floors
basementdweller	Numeric	Binarized	1 if living in basement
Car_1	Numeric	Extracted from Station_1 Kanji	Car travel time to closest station
Walk_1	Numeric	Extracted from Station_1 Kanji	Walk travel time to closest station
Bus_1	Numeric	Extracted from Station_1 Kanji	Bus travel time to closest station
Car_2	Numeric	Extracted from Station_2 Kanji	Car travel time to second closest station
Walk_2	Numeric	Extracted from Station_2 Kanji	Walk travel time to second closest station
Bus_2	Numeric	Extracted from Station_2 Kanji	Bus travel time to second closest station
Car_3	Numeric	Extracted from Station_3 Kanji	Car travel time to third closest station
Walk_3	Numeric	Extracted from Station_3 Kanji	Walk travel time to third closest station
Bus_3	Numeric	Extracted from Station_3 Kanji	Bus travel time to third closest station
location	Categorical	Text cleanup	Administrative district (chome)
floor_plan	Categorical	Text cleanup	Room layout
Ku	Categorical	None	Ku
Line_1	Categorical	Split from Station_1	Closest station's line
Station_1	Categorical	Split from Station_1	Closest station
Line_2	Categorical	Split from Station_2	Second closest station's line
Station_2	Categorical	Split from Station_2	Second closest station
Line_3	Categorical	Split from Station_3	Third closest station's line
Station_3	Categorical	Split from Station_3	Third closest station

MODEL PRE-PROCESSING – GEOSPATIAL AND ENVIRONMENTAL MODEL

The Geospatial-Environmental model uses the baseline model features and adds features from each environmental and geospatial dataset.

- Features from each dataset can be referenced in the previous slides:

Table 12: Geospatial/Environmental Model Datasets

Dataset	Source	Description	Join Key
Suumo	Suumo	Apartment and mansion listings for 23 Ku's in Tokyo	location/Ku
Google Cloud API's	Google Cloud	Geospatial coordinates, distance and time calculations to hub stations	location
Tokyo Regional Earthquake Risk Survey	Tokyo Bureau of Urban Development	Earthquake risk by administrative district in Tokyo (chome)	location
Tokyo Air Quality	Tokyo Ministry of the Environment	Air quality by sensor in Tokyo	location
Tokyo Parks	Tokyo Bureau of Construction	Park statistics by Ku	Ku
Tokyo Crime Data	Tokyo Metropolitan Police	Crime statistics by administrative district in Tokyo (chome)	location
Land Use by District	Tokyo Bureau of Urban Development	Land use classification breakdown in Tokyo by Ku	Ku
Complaints about Pollution by Kind	Tokyo Ministry of the Environment	Total complaints about pollution by Ku	Ku

MODEL PRE-PROCESSING: FEATURE ENGINEERING MODEL

The Feature Engineering model uses the previous models features with the addition of new features created through intuition and experience to boost performance.

- Binary Encoding
- Expansion Encoding
- Consolidation Encoding
- Interaction Features
- Categorical Binning
- Ranking Features
- Target-Mean Encoding
- Geospatial Coordinate System Projection

BINARY, EXPANSION, AND CONSOLIDATION ENCODING

Binary Encoding

First Floor? 1 or 0

Floor		first_floor_FE
1	→	1
6	→	0

Expansion Encoding

Table 13: Expansion Encoding Features (room_layout)

New Feature	Description	Correlation to Target
total_rooms_FE	Total number of rooms	0.368
Living_FE	Living room?	0.595
Dining_FE	Dining room?	0.469
Kitchen_FE	Separate kitchen?	0.218
Service_Room_FE	Service room?	0.194
Room_Only_FE	1 room layout	-0.218

Room_layout		rooms	L	D	K
2LDK	→	2	1	1	1
1K	→	1	0	0	1

Consolidation Encoding

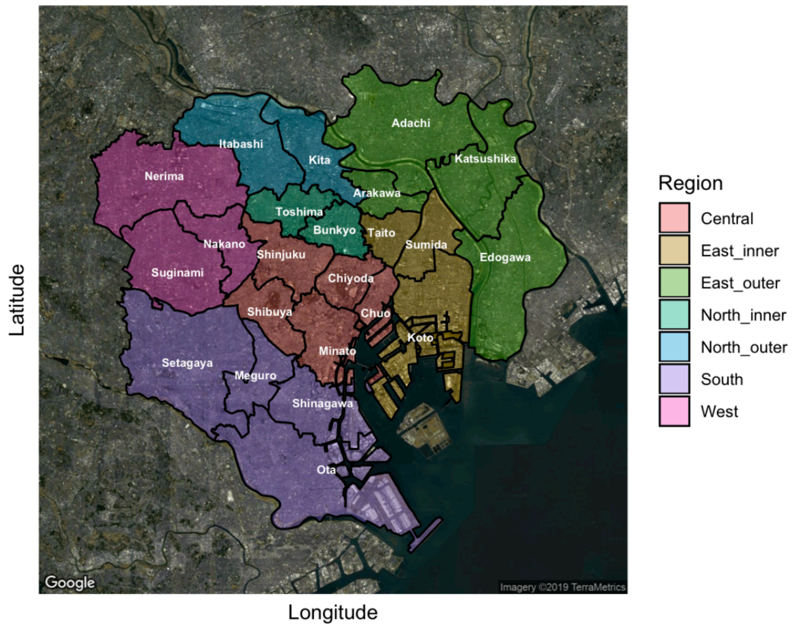


Figure 7: Consolidation Encoding Map

INTERACTION FEATURES

Table 14: Interaction Features (Building)

Feature Name	Description	Correlation to Target
area_per_room_FE	area / total_rooms	-0.056
height_ratio_FE	floor / Height	-0.086
area_height_FE	area $\times$ Height	0.549
floor_area_FE	floor $\times$ area	0.499

Table 15: Interaction Features (Risk)

Feature	Description	Correlation to Target
yrs_risk_FE	yrs $\times$ Total_Risk	-0.180
yrs_lifediff_FE	yrs $\times$ Life_Difficulty_Risk	-0.225
yrs_fire_FE	yrs $\times$ Fire_Risk	-0.153
yrs_building_FE	yrs $\times$ Building_Risk	-0.248

Table 16: Interaction Features (Transit)

Feature	Description	Correlation to Target
Hub_Aggregate_FE	Sum of transit times to major hubs (made features)	-0.313
Shinjuku_Commute_FE	Transit time to Shinjuku + Time to Station_1	-0.213
Tokyo_Commute_FE	Transit time to Tokyo + Time to Station_1	-0.301
Shinagawa_Commute_FE	Transit time to Shinagawa + Time to Station_1	-0.306
Ueno_Commute_FE	Transit time to Ueno + Time to Station_1	-0.254
Ikebukuro_Commute_FE	Transit time to Ikebukuro + Time to Station_1	-0.072
Shibuya_Commute_FE	Transit time to Shibuya + Time to Station_1	-0.356
Aggregate_Commute_FE	Sum of commutes to all major hubs	-0.297

Table 17: Interaction Features (Interaction x Interaction)

Feature	Description	Correlation to Target
agg_commute_floor_area_FE	Aggregate_Commute_FE $\times$ floor_area_FE	0.422
agg_commute_area_height_FE	Aggregate_Commute_FE $\times$ area_height_FE	0.485
agg_commute_area_per_room_FE	Aggregate_Commute_FE $\times$ area_per_room_FE	-0.290
agg_commute_yrs_building_FE	Aggregate_Commute_FE $\times$ yrs_building_FE	-0.267
agg_commute_yrs_fire_FE	Aggregate_Commute_FE $\times$ yrs_fire_FE	-0.156
agg_commute_yrs_lifediff_FE	Aggregate_Commute_FE $\times$ yrs_lifediff_FE	-0.253
agg_commute_yrs_risk_FE	Aggregate_Commute_FE $\times$ yrs_risk_FE	-0.183

Table 18: Interaction Features (Land Classification)

Feature	Description	Correlation to Target
Residential_land_ratio_FE	Residential / Total_District_Land	0.050
OtherUse_land_ratio_FE	OtherUse / Total_District_Land	-0.287
Parks_land_ratio_FE	Parks / Total_District_Land	-0.016
Fields_land_ratio_FE	Fields / Total_District_Land	-0.231
WoodForest_land_ratio_FE	WoodsForest / Total_District_Land	-0.063
Water_land_ratio_FE	Water / Total_District_Land	-0.021
Farm_land_ratio_FE	Farm / Total_District_Land	-0.191
Roads_land_ratio_FE	Roads / Total_District_Land	0.110
Unused_land_ratio_FE	Unused / Total_District_Land	0.157

Table 19: Interaction Features (Pollution Complaints)

Feature	Description	Correlation to Target
Air_Pollution_Complaints_per_person_FE	Air_Pollution / Population	-0.137
Water_Pollution_Complaints_per_person_FE	Water_Pollution / Population	0.237
Soil_Pollution_Complaints_per_person_FE	Soil_Pollution / Population	-0.067
Other_Pollution_complaints_per_person_FE	Other_Pollution / Population	0.120
Offensive_odors_Complaints_per_person_FE	Offensive_odors / Population	0.136
Vibration_Complaints_per_person_FE	Vibration / Population	0.070
Noise_Pollution_Complaints_per_person_FE	Noise / Population	0.247

Table 20: Interaction Features (Crime)

Feature	Description	Correlation to Target
Felony_Offense_Ratio_FE	Felonious_Offense_Total / Total_Crimes	0.026
Violent_Crime_Ratio_FE	Violent_Crime_Total / Total_Crimes	0.111
Burglary_Larceny_Ratio_FE	Burglary_Larceny_Total / Total_Crimes	0.009
Non_Intrusive_Larceny_Ratio_FE	Non_Intrusive_Larceny / Total_Crimes	-0.116
Other_Crime_Ratio_FE	Others_Total / Total_Crimes	0.081

RANK ENCODING

CATEGORICAL BINNING

Table 21: Rank Encoding (Crime)

New Feature	Description	Correlation to Target
Crime_Rank_FE	Ranking by location based on total overall crime	-0.029
Felony_Rank_FE	Ranking by location based on total felonies	-0.057
Violent_Crime_Rank_FE	Ranking by location based on total violent crimes	-0.082
Burglary_Larceny_Rank_FE	Ranking by location based on total burglary and larceny	0.009
Non_Intrusive_Larceny_Rank_FE	Ranking by location based on total non intrusive larceny	-0.003
Other_Crime_Rank_FE	Ranking by location based on other crimes	-0.063

Table 22: Categorical Binning

Feature	Category Mapping	Threshold
location	(Ku)_Minority (23 Categories)	< 5 samples
Line_1	Line_Minority	< 30 samples
Station_1	Station_Minority	< 30 samples
Line_2	Line_Minority	< 30 samples
Station_2	Station_Minority	< 30 samples
Line_3	Line_Minority	< 30 samples
Station_3	Station_Minority	< 30 samples

TARGET MEAN ENCODING / COORDINATE SYSTEM PROJECTION

Target Mean Encoding: Encodes a categorical feature by its expected value of target.

- This research employs a form of target-encoding that uses the prior *probability as* a regularization method.
- Line_1, Line_2, Line_3, Station_1, Station_2, Station_3, location (Administrative District),
- Can reduce categorical variable to 1-dimension

Station_1	Station_1_FE
四ツ谷駅	112124.52
四ツ谷駅	112286.66
四ツ谷駅	112366.83

Table 23: Coordinate System Projections

Features	Source LAT/LON
x_cart, y_cart, z_cart	location
x_cart_Eki1, y_cart_Eki1, z_cart_Eki1	Station_1
x_cart_Eki2, y_cart_Eki2, z_cart_Eki2	Station_2
x_cart_Eki3, y_cart_Eki3, z_cart_Eki3	Station_3

LAT	LON
35.64817	139.7505

x_cart_FE	y_cart_FE	z_cart_FE
-0.02324904	-0.4613190	-0.8869297

FINAL PRE-PROCESSING ALL MODELS

Feature Scaling (Numeric Features)

- Scikit-Learn
- Gives each feature a mean and unit variance of 1

One-Hot-Encoding (Categorical Features)

- Scikit-Learn
- All Models were encoded with One-Hot-Encoding with the exception of the Feature Engineering Model which encodes features with high cardinality through Target-Mean-Encoding

Table 24: Feature Breakdown for Each Model

Model	Categorical Feature Encoding Method	Numeric Columns	Encoded Categorical Features	Total Features
Baseline (1)	One-Hot-Encoding	22	5,277	5,299
Geospatial and Environmental (2)	One-Hot-Encoding	207	5,287	5,494
Feature Engineering (3)	One-Hot-Encoding/Target Mean Encoding	283	56	339

LIGHTGBM/GRADIENT BOOSTING

LightGBM – Gradient Boosting Decision Tree (GBDT) developed by Microsoft

Gradient Boosting Decision Tree - Ensemble that trains multiple weak learners (decision tree's) sequentially to make a stronger model.

$$y(\hat{x})^K = \sum_{i=1}^{K} f_i(x)$$

f_i – output of the ith regression tree of the Kth ensemble

$y(\hat{x})^K$ – predicted output.

$$L = \sum_{i=1}^{n} L(y_i, \hat{y}_i^K + f_{K+1}(x_i)) + \Theta(f_{K+1})$$

To build the K+1th tree a regularized objective function is minimized:

(a) Level Wise Tree Growth

(b) Leaf Wise Tree Growth

Figure 8: Level vs Leaf Wise Tree Growth [39]

Machine Learning Challenge Winning Solutions

Place	Competition	Solution	Date
1st	TalkingData AdTracking Fraud Detection Challenge	link	2018.5
1st	DonorsChoose.org Application Screening	link	2018.4
1st	Toxic Comment Classification Challenge	link	2018.3
1st	Mercari Price Suggestion Challenge	link	2018.2
1st	IEEE's Signal Processing Society, Camera Model Identification	link	2018.2
1st	Recruit Restaurant Visitor Forecasting	link	2018.2
1st	WSDM CUP 2018 - KKBox's Music Recommendation Challenge	link	2017.12
1st	Porto Seguro's Safe Driver Prediction	link	2017.11
1st	Quora Question Pairs	link	2017.6

Source: LightGBM

LIGHTGBM/GRADIENT BOOSTING

State of the Art GBDT's
- XGBoost
- LightGBM
- CatBoost

LightGBM is popular and has been successful in Machine Learning competitions:
- Recruit Visitor Forecasting
- Mercari Price Prediction

Recent research also show it as generally being able to converge to a solution that generalizes better.

Big challenge with GBDT's
- Determining the optimal tree splits

LightGBM
- Leaf Wise Growth
- Relies on histogram-based algorithms to bucket continuous features into discrete bins

XGBoost
- Level Wise Growth
- Pre-sort based algorithms

Table 25: Bayesian Optimization Hyperparameter Search Range and Final Parameters (rounded to 2 decimals)

Parameter	Search Range	Baseline	Geo_Env	FE
bagging_fraction	0.8-1	0.99	0.95	0.98
colsample_by_tree	0.5-0.7	0.59	0.66	0.68
feature_fraction	0.1-0.9	0.76	0.57	0.86
lambda_l1	0-5	4.31	2.23	0.31
lambda_l2	0-3	2.07	1.21	2.63
learning_rate	.05-0.2	0.16	0.17	0.13
max_depth	5-9	8	9	7
min_child_weight	5-50	29.69	47.42	17.67
min_split_gain	0.001-0.1	0.03	0.06	0.01
num_leaves	24-45	25	40	41
subsample	.5-.7	0.59	0.52	0.63
device_type	gpu	gpu	gpu	gpu
early_stopping_rounds	1000	1000	1000	1000

HYPERPARAMETER OPTIMIZATION

Optimal Hyperparameters were determined through the use of **Bayesian Optimization.**

- Bayesian Optimization is a method for finding optimal parameters that uses **gaussian processes** and an **acquisition function** to guide the exploration process.
  - Acquisition Function - Expected Improvement

Bayesian Optimization Methodology

- **5-fold Cross-Validation** on the training set
- Optimal parameters were selected based on predictive performance on the validation set.

TABLE OF CONTENTS

Table 26: MAPE Scores by Model

Evaluation Method	Baseline	Geo-Env	Feature Engineering
5-Fold CV (x5)	5.89	5.36 -.53 ↓	5.12 -.77 ↓
Test Set Evaluation	5.91	5.32 -.59 ↓	5.01 -.9 ↓

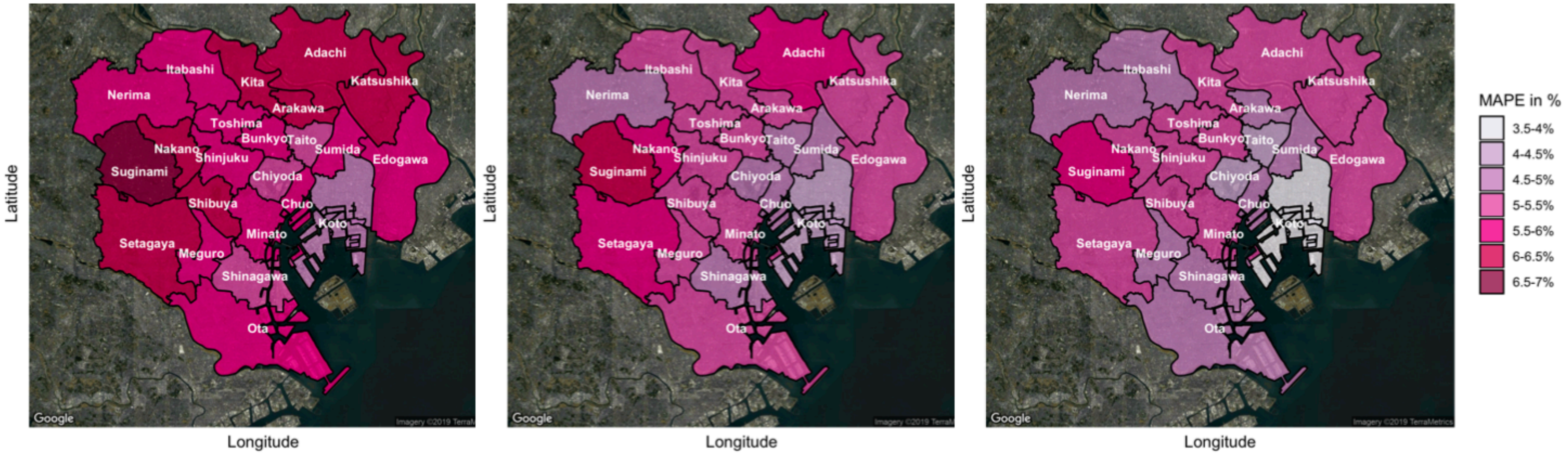
Table 27: RMSE Scores by Model

Evaluation Method	Baseline	Geo-Env	Feature Engineering
5-Fold CV (x5)	12,494.14	11,692.04 -802 ↓	11,269.30 -1,225 ↓
Test Set Evaluation	12,204.35	11,328.02 -876 ↓	11,202.46 -1,002 ↓

RESULTS

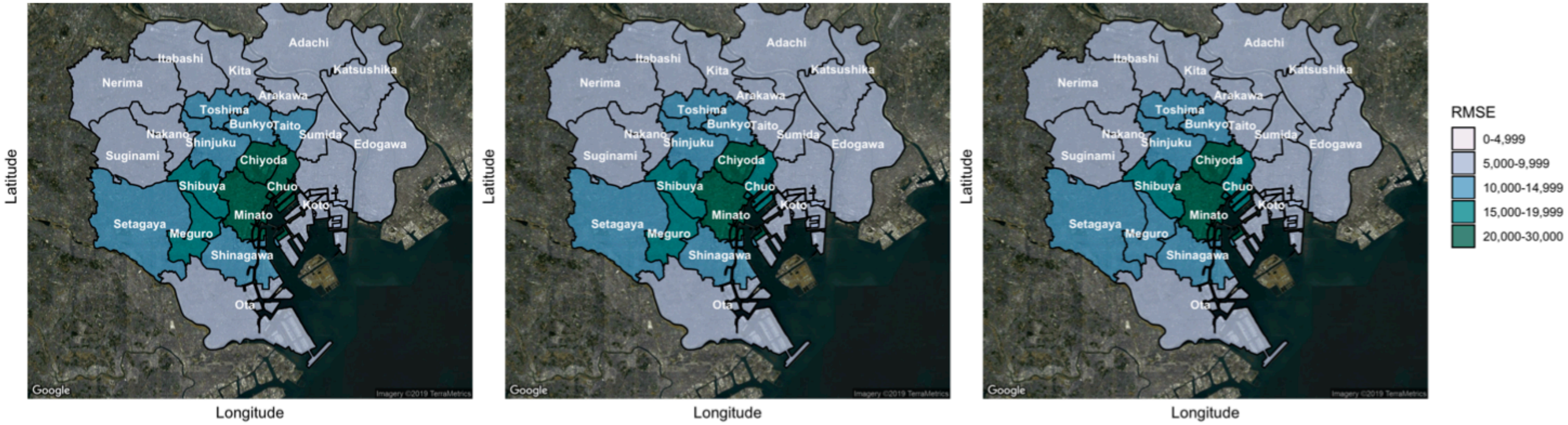
- This research **quantifies** the **effectiveness** of the use of feature engineering and environmental/geo-spatial features by **comparing** the baseline model to the two constructed models.
- The results were compared based on **5-Fold Cross Validation** conducted **5x** with different random seeds and a final test set evaluation.
- The **biggest increase** in performance came with the addition of geospatial and environmental features.
- The proposed feature engineering approach **greatly improves** all proposed **evaluation metrics** and **reduces** the total **number of features** required for model training.
  - MAPE 5.89 -> 5.12
  - RMSE 12,494.14 -> 11,269.30

RESULTS BY KU



(a) Baseline (b) Geo-Env Model (c) FE Model

Figure 9: MAPE scores for each Ku by Model



(a) Baseline (b) Geo-Env Model (c) FE Model

Figure 10: RMSE scores for each Ku by Model

Table 28: MAPE by Model and Comparison to Baseline by Ku

Ku	Baseline	Geo-Env	FE	Geo-Env Change	FE Change
Arakawa	6.28	5.38	5.03	-0.90	-1.25
Katsushika	6.22	5.33	5.12	-0.89	-1.10
Nakano	6.44	5.89	5.34	-0.55	-1.10
Meguro	5.63	5.07	4.57	-0.56	-1.06
Suginami	7.00	6.35	5.96	-0.65	-1.04
Setagaya	6.29	5.56	5.25	-0.73	-1.03
Kita	6.07	5.37	5.06	-0.70	-1.01
Nerima	5.68	4.84	4.67	-0.84	-1.00
Adachi	6.09	5.57	5.14	-0.52	-0.96
Ota	5.89	5.30	5.00	-0.59	-0.89
Chiyoda	5.09	4.52	4.22	-0.57	-0.87
Shibuya	6.13	5.49	5.27	-0.64	-0.86
Sumida	5.53	4.95	4.69	-0.58	-0.83
Koto	4.75	4.22	3.94	-0.53	-0.81
Chuo	5.54	4.91	4.73	-0.63	-0.80
Shinagawa	5.40	4.83	4.59	-0.56	-0.80
Edogawa	5.90	5.35	5.13	-0.54	-0.77
Taito	5.20	4.68	4.45	-0.51	-0.75
Itabashi	5.70	5.12	4.99	-0.59	-0.71
Toshima	5.79	5.40	5.11	-0.39	-0.68
Minato	5.69	5.21	5.06	-0.49	-0.63
Bunkyo	5.97	5.47	5.37	-0.49	-0.59
Shinjuku	5.71	5.44	5.12	-0.27	-0.59

Table 29: RMSE by Model and Comparison to Baseline by Ku

Ku	Baseline	Geo-Env	FE	Geo-Env Change	FE Change
Meguro	16,108	15,437	13,272	-671	-2,836
Chuo	22,438	19,728	19,804	-2,710	-2,634
Shibuya	19,083	15,849	16,520	-3,235	-2,563
Taito	11,922	9,915	9,512	-2,007	-2,410
Toshima	12,630	10,853	10,384	-1,777	-2,246
Setagaya	12,054	10,623	10,074	-1,431	-1,980
Nerima	9,134	7,814	7,292	-1,320	-1,842
Arakawa	9,387	8,079	7,923	-1,308	-1,464
Kita	8,961	8,300	7,654	-661	-1,307
Ota	9,440	8,483	8,199	-957	-1,240
Edogawa	8,147	7,527	6,966	-620	-1,182
Shinagawa	13,447	11,027	12,404	-2,420	-1,043
Sumida	9,146	8,531	8,157	-614	-989
Itabashi	7,745	6,847	6,908	-899	-837
Suginami	9,653	9,202	8,841	-450	-812
Katsushika	7,135	6,761	6,382	-374	-753
Koto	9,328	8,932	8,581	-395	-747
Nakano	8,744	8,697	8,017	-47	-727
Shinjuku	13,654	13,304	13,234	-350	-421
Adachi	7,448	7,983	7,087	535	-361
Bunkyo	12,114	12,193	12,863	79	749
Minato	23,681	24,291	24,456	610	775
Chiyoda	21,584	21,842	25,742	258	4,158

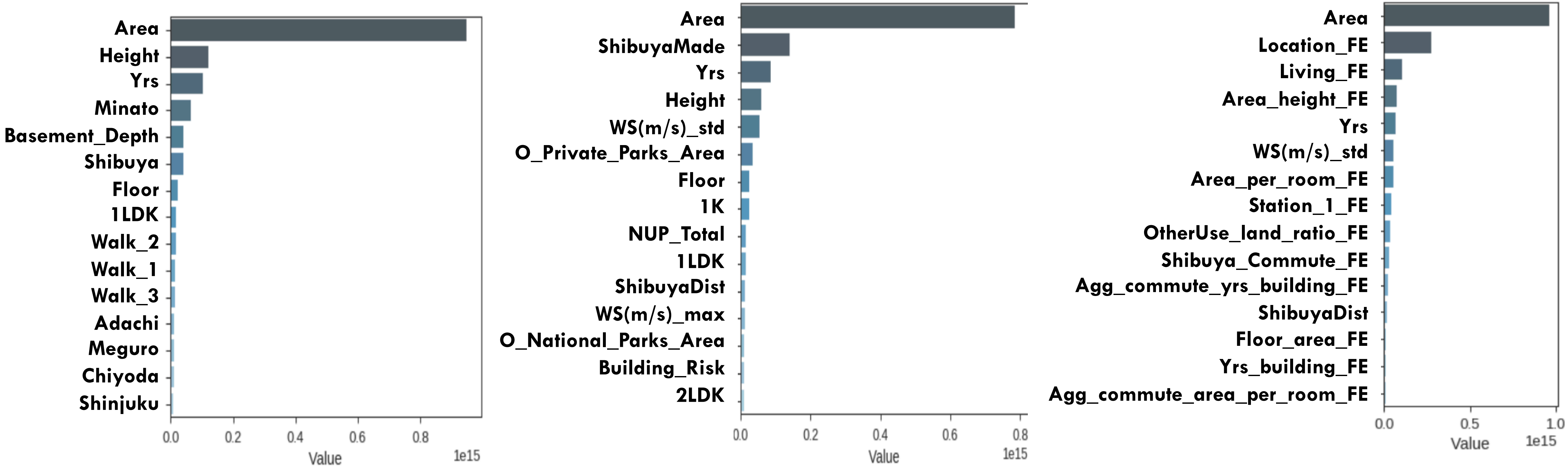
RESULTS BY PRICE BIN

Table 30: MAPE by Model and Comparison to Baseline by Price Group

Price Group	Baseline	Geo-Env	FE	Geo-Env Change	FE Change
23,999-53,000	10.13	9.2	8.94	-0.93	-1.19
53,000-60,000	6.78	6.13	5.56	-0.65	-1.22
60,000-66,000	6.27	5.67	5.12	-0.6	-1.15
66,000-72,000	5.59	5.16	4.8	-0.43	-0.79
72,000-77,000	5.81	5.29	5	-0.52	-0.81
77,000-82,500	5.67	5.18	4.69	-0.49	-0.98
82,500-88,000	5.52	5.03	4.88	-0.49	-0.64
88,000-95,000	5.72	5.12	4.88	-0.59	-0.83
95,000-102,000	5.2	4.6	4.21	-0.6	-0.99
102,000-110,000	5.09	4.61	4.29	-0.48	-0.81
110,000-121,000	4.69	4.07	3.89	-0.62	-0.8
121,000-135,000	4.85	4.33	4.06	-0.52	-0.79
135,000-155,000	5.27	4.61	4.36	-0.66	-0.91
155,000-203,000	5.7	4.97	4.74	-0.73	-0.95
203,000-790,000	6.36	5.63	5.63	-0.73	-0.73

Table 31: RMSE by Model and Comparison to Baseline by Price Group

Price Group	Baseline	Geo-Env	FE	Geo-Env Change	FE Change
23,999-53,000	6,412	6,103	5,824	-309	-588
53,000-60,000	5,369	5,067	4,670	-302	-699
60,000-66,000	5,469	5,201	4,683	-267	-785
66,000-72,000	5,606	5,533	4,998	-73	-609
72-000-77,000	6,323	5,893	5,562	-430	-761
77,000-82,500	6,430	6,267	5,575	-163	-855
82,500-88,000	6,920	6,664	6,442	-256	-478
88,000-95,000	8,090	7,852	7,252	-237	-838
95,000-102,000	7,609	7,155	6,490	-455	-1,119
102,000-110,000	7,865	7,515	7,025	-350	-840
110,000-121,000	8,376	7,524	7,288	-852	-1,088
121,000-135,000	9,972	8,987	8,416	-986	-1,557
135,000-155,000	11,939	10,912	10,649	-1,027	-1,290
155,000-203,000	16,523	15,495	14,564	-1,027	-1,959
203,000-790,000	34,862	32,090	33,051	-2,772	-1,811



(a) Baseline

(b) Geo-Env Model

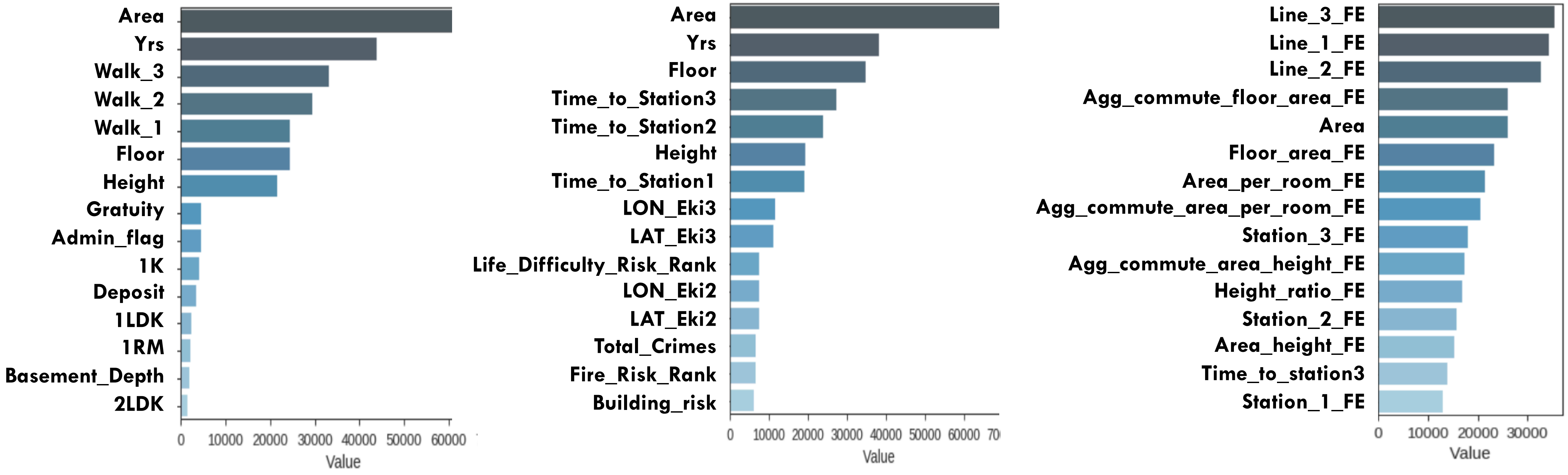
(c) FE Model

Figure 11: Feature Importance by Gain for Each Model

The total gain measures the relative contribution of a feature in comparison to others and can often be an indicator for the amount a feature contributes to a prediction.

- **All models-** area, Height, yrs, floor (Fundamental features for Tokyo rent estimation)
- **Geospatial/Environmental model –** Time/Distance to Shibuya, park related features, building risk
- **Feature Engineering model-** 11/15 of important features were created by Feature Engineering.

Surprising – Wind Speed(m/s)_std



(a) Baseline (b) Geo-Env Model (c) FE Model

Figure 12: Feature Importance by Total Splits for Each Model

The total splits measures how many times a feature was used in tree splitting criteria. It does not always translate to predictive power but we can assume that features with high gain and high splits are very important.

- **All models-** area, Height, yrs, floor, travel time to stations (Fundamental features for Tokyo)
- **Geospatial/Environmental model –** Latitude/Longitude, disaster risk, crime features
- **Feature Engineering model**- 13/15 of most important features were created by Feature Engineering.

Analysis of feature importance show the contribution the addition features had on each model's predictive performance.

TABLE OF CONTENTS

CONCLUSION

This research presented a time independent gradient boosting approach to rent prediction in Tokyo through the use of environmental and geospatial features as well as feature engineering techniques.

- This study provides insight into the viability of environmental features in rental prediction for Tokyo
- Feature Engineering is often associated with the generation of more features however it can be used as a technique to reduce features and maintain accurate representation of the data.
- Environmental Features in conjunction with creative feature engineering can be used as useful inputs to improve predictive performance of real estate estimation algorithms.

FUTURE RESEARCH DIRECTIONS

Rent Estimation in other areas of Japan

- Tokyo is a special case in Japan and the hardest market to predict

Ensemble Modeling

Feature Selection

Separate Models for Different Market Segments (Luxury vs Inexpensive)

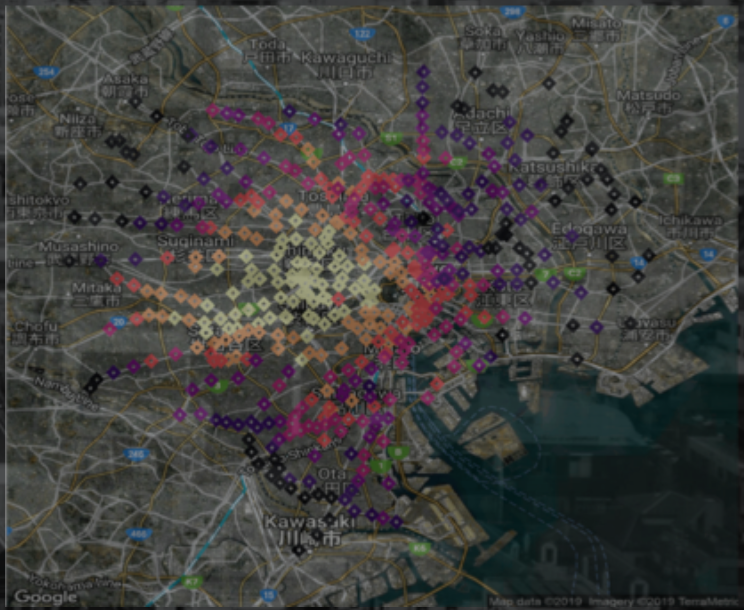
Neural Networks for Unstructured Data

- Building Names
- Room Layout Images

Categorical Entity Embedding vs Target Mean Encoding

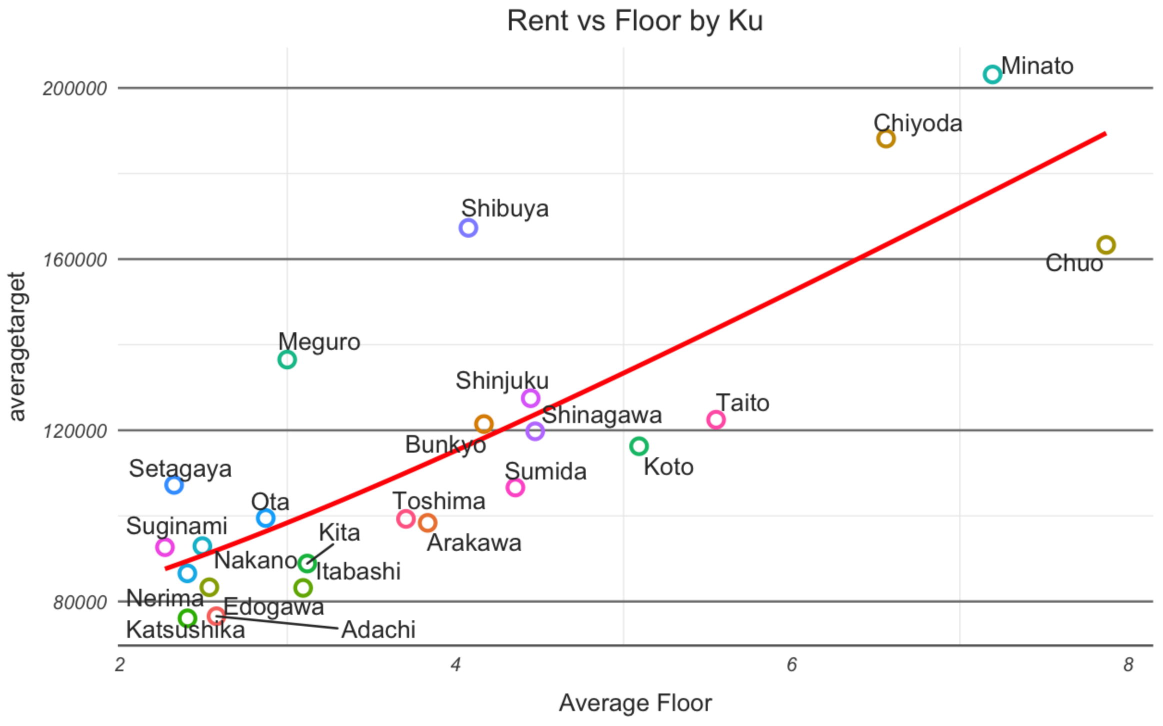
Amenity Data

APPENDIX GRAPHS
TRAINING SET

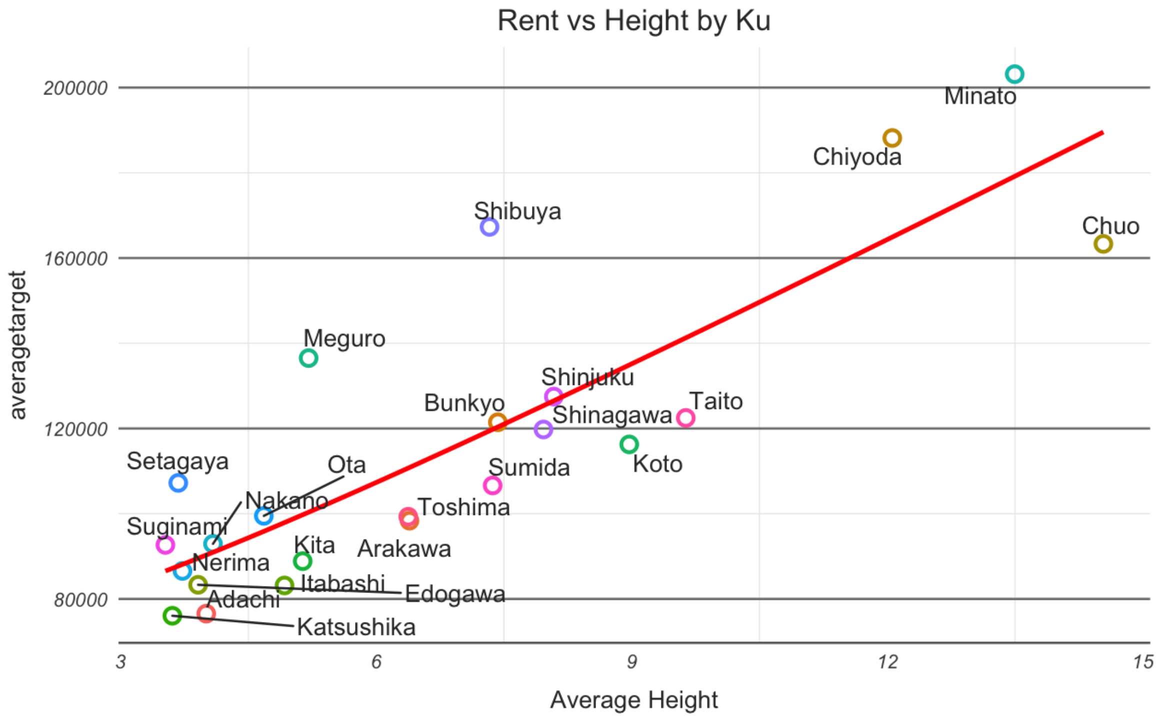


FUNDAMENTAL FEATURES FOR TOKYO RENT PREDICTION

Higher Floor– More Rent

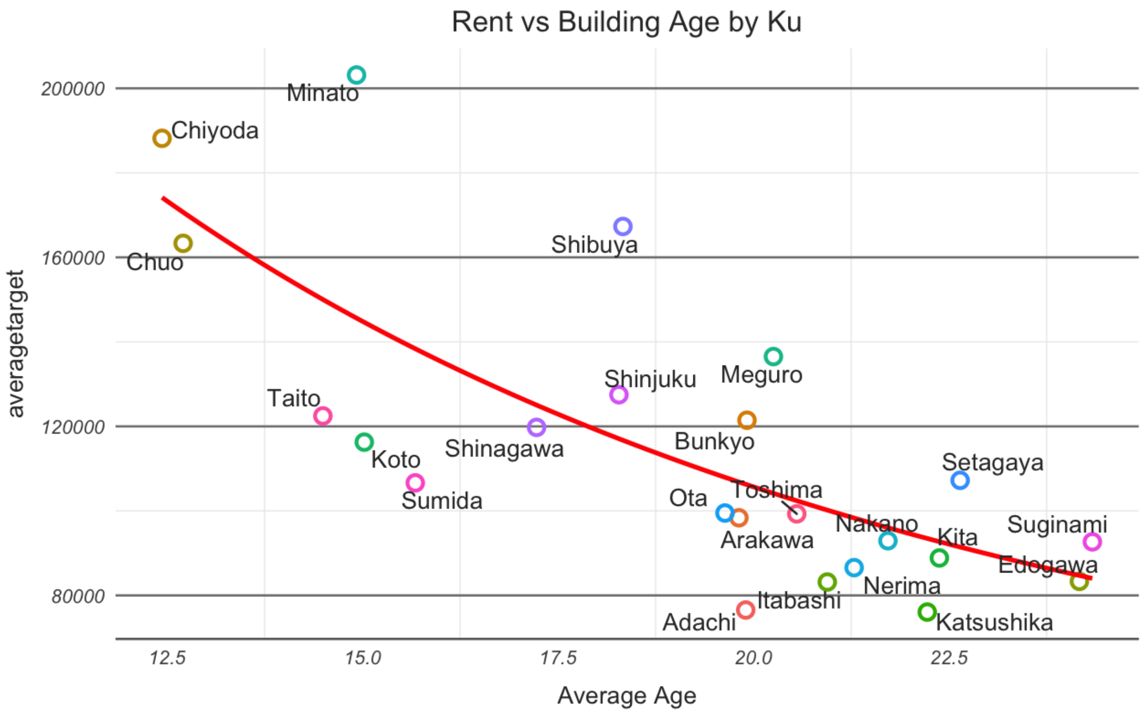


Higher Building Height – More Rent

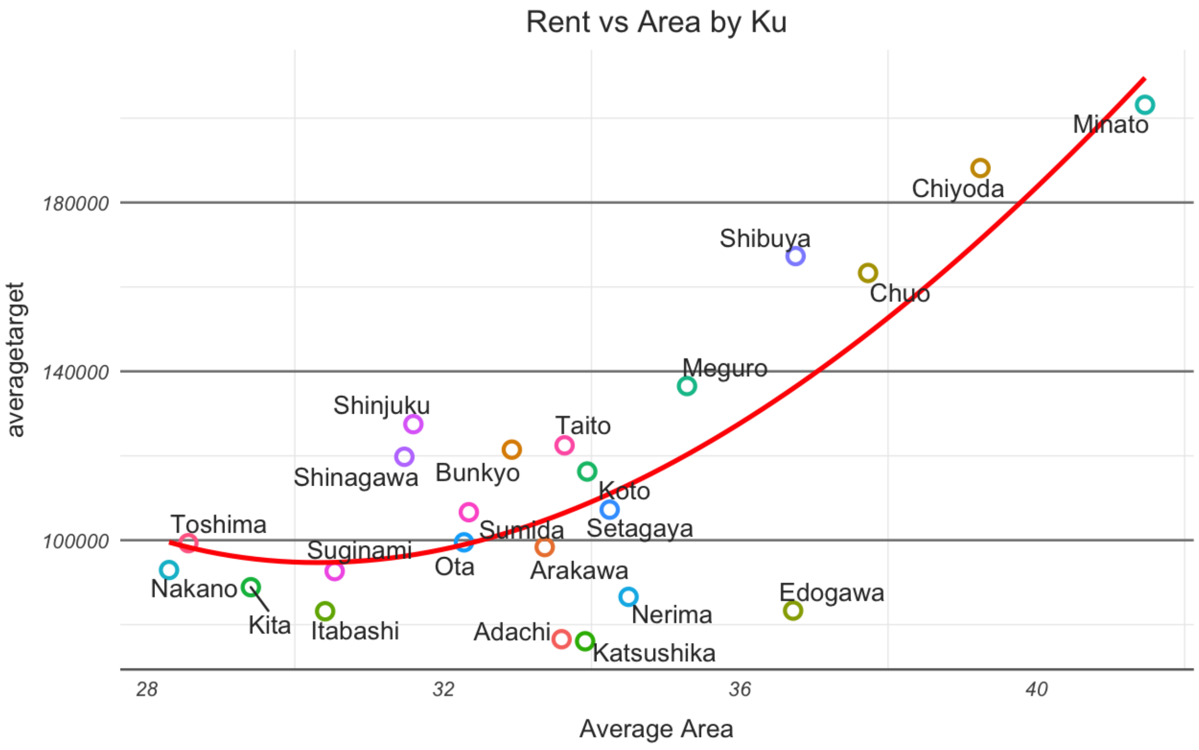


FUNDAMENTAL FEATURES FOR TOKYO RENT PREDICTION

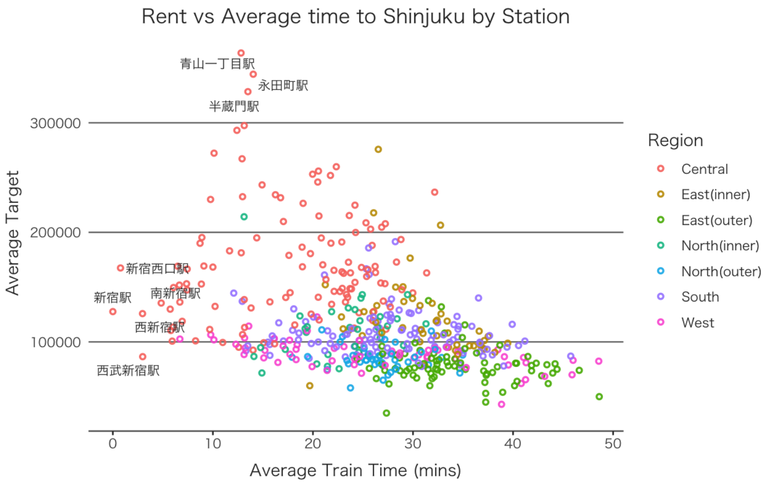
Higher Age – Less Rent

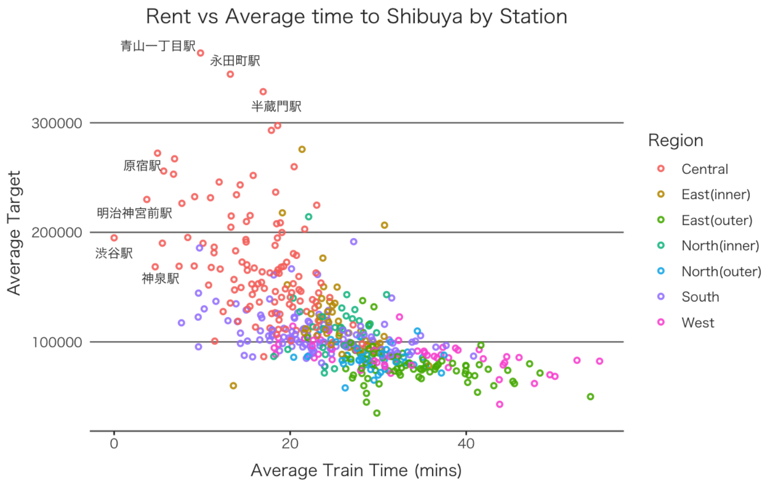


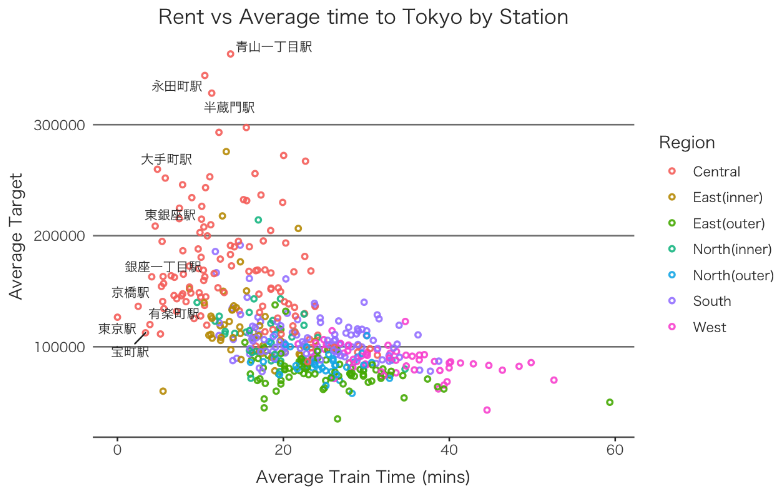
Higher Area – More Rent

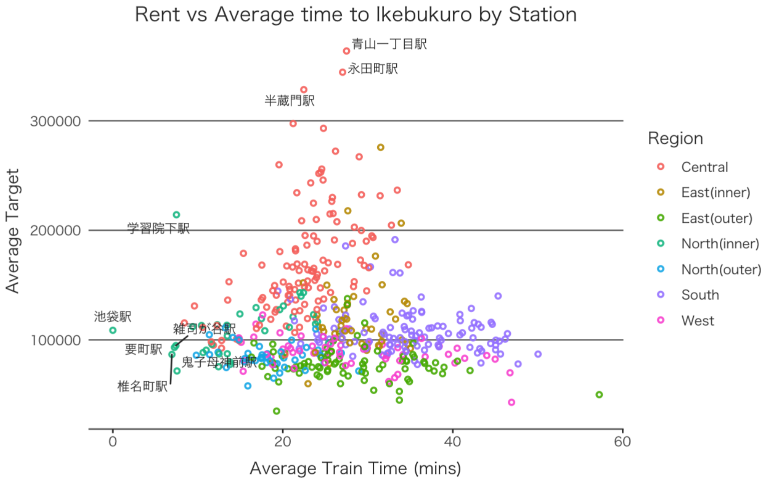


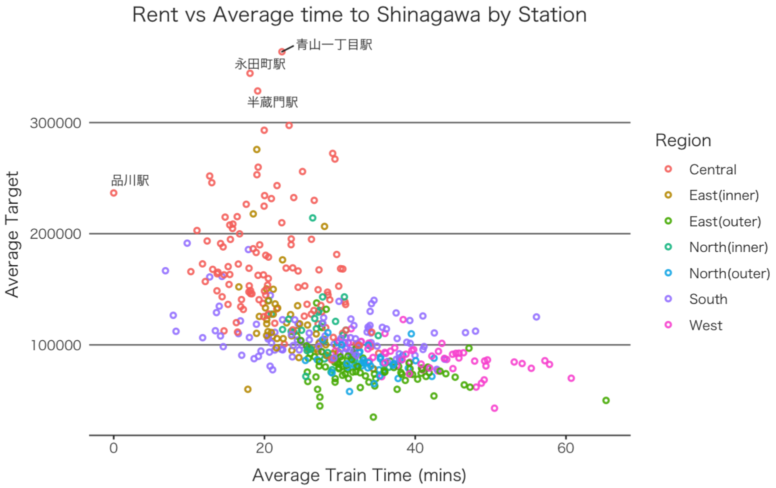
COMMUTE TIME VS AVERAGE RENT BY STATION (TRANSIT TIME)

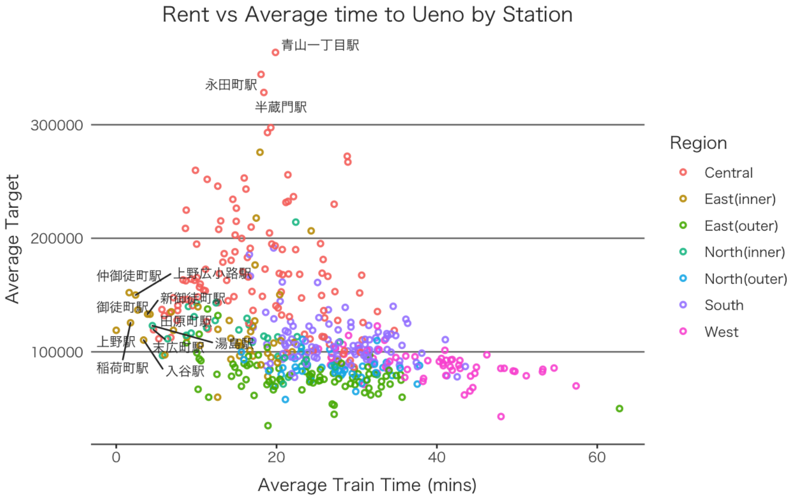






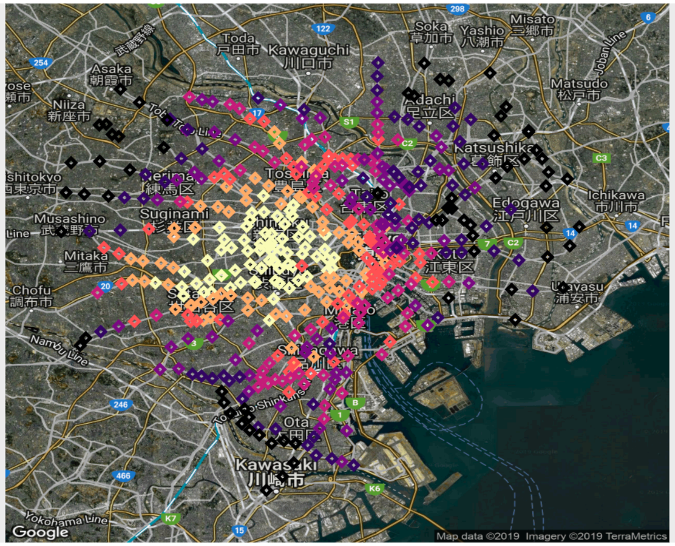




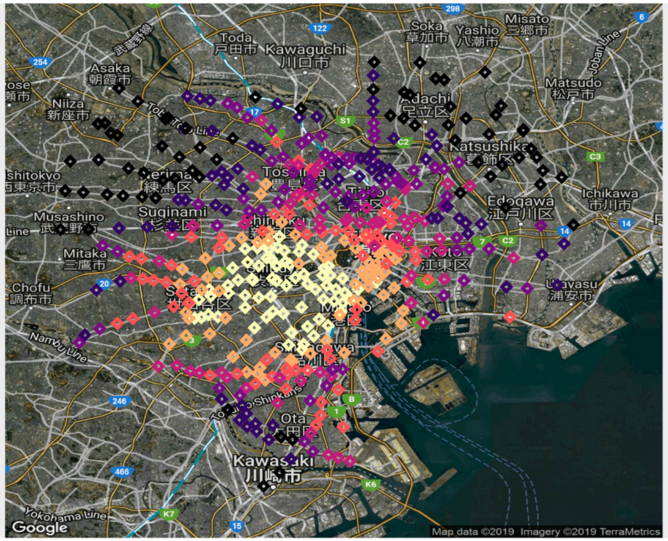


AVERAGE COMMUTE HEATMAP BY STATION

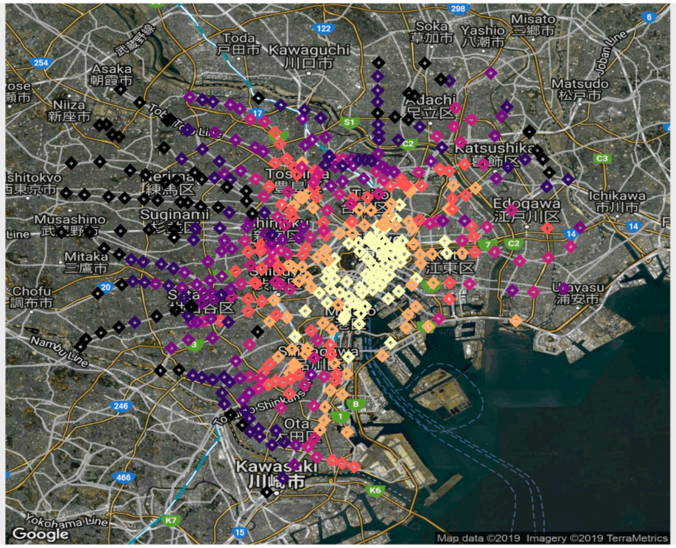
Average commute to Shinjuku



Average commute to Shibuya



Average commute to Tokyo



TOKYO REGIONAL EARTHQUAKE RISK SURVEY

Building Risk

location	Ku	averagetarget	totalrisk	buildingrisk	firerisk	lifequality	averagetargetKu
墨田区京島2	Sumida	86818.18	7	1	2	917	106632.64
墨田区京島3	Sumida	119315.79	36	2	9	1423	106632.64
台東区鳥越1	Taito	141519.23	1549	3	213	2634	122476.24
足立区柳原2	Adachi	66922.73	11	4	12	474	76546.82
足立区千住柳町	Adachi	63900.00	2	5	1	332	76546.82
荒川区南千住1	Arakawa	86323.53	23	6	25	725	98369.81
台東区日本堤1	Taito	107534.48	2970	7	11	2780	122476.24
荒川区町屋4	Arakawa	76633.33	1	8	6	42	98369.81
足立区千住4	Adachi	95200.00	46	9	57	726	76546.82
足立区千住寿町	Adachi	87143.48	30	10	19	1035	76546.82

Fire Risk

location	Ku	averagetarget	totalrisk	buildingrisk	firerisk	lifequality	averagetargetKu
足立区千住柳町	Adachi	63900.00	2	5	1	332	76546.82
墨田区京島2	Sumida	86818.18	7	1	2	917	106632.64
墨田区墨田3	Sumida	79784.09	5	11	3	400	106632.64
足立区千住大川町	Adachi	73511.43	4	20	4	279	76546.82
荒川区荒川6	Arakawa	85392.86	3	24	5	168	98369.81
荒川区町屋4	Arakawa	76633.33	1	8	6	42	98369.81
足立区千住元町	Adachi	77666.67	10	84	7	333	76546.82
墨田区押上3	Sumida	93442.86	12	34	8	556	106632.64
墨田区京島3	Sumida	119315.79	36	2	9	1423	106632.64
墨田区東向島1	Sumida	83523.81	13	12	10	641	106632.64

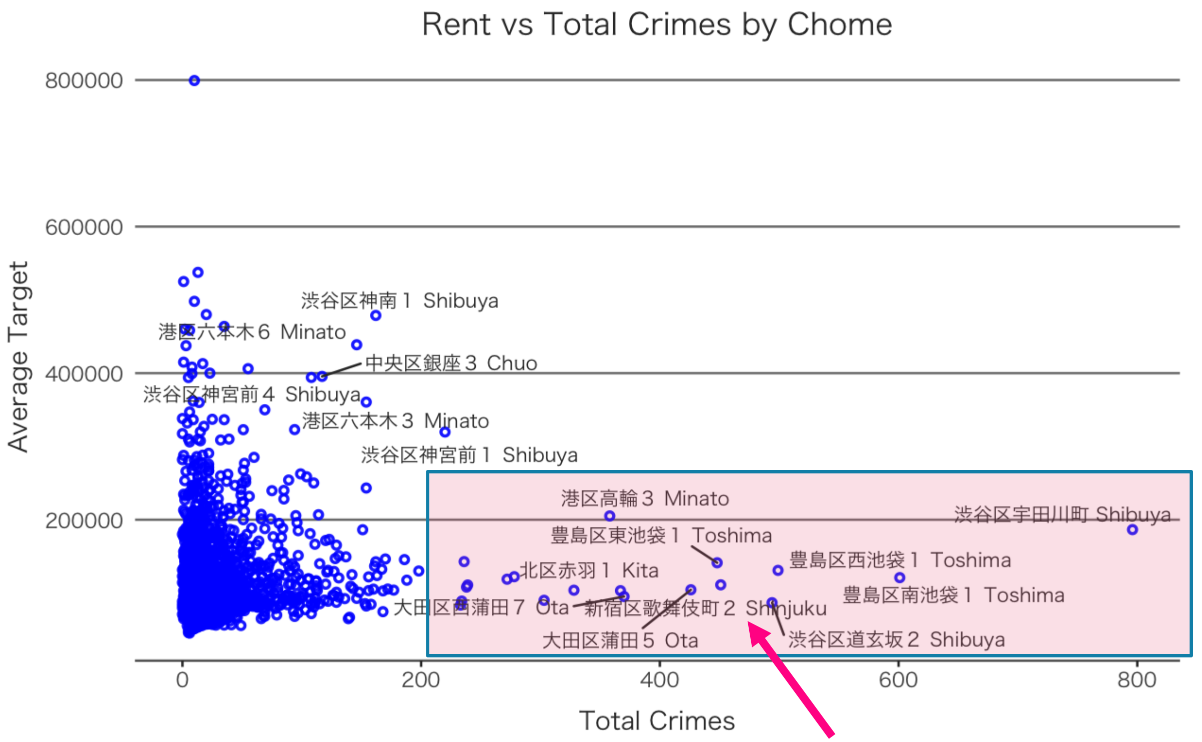
Total Risk

location	Ku	averagetarget	totalrisk	buildingrisk	firerisk	lifequality	averagetargetKu
荒川区町屋4	Arakawa	76633.33	1	8	6	42	98369.81
足立区千住柳町	Adachi	63900.00	2	5	1	332	76546.82
荒川区荒川6	Arakawa	85392.86	3	24	5	168	98369.81
足立区千住大川町	Adachi	73511.43	4	20	4	279	76546.82
墨田区墨田3	Sumida	79784.09	5	11	3	400	106632.64
北区志茂4	Kita	79534.09	6	111	14	66	88865.02
墨田区京島2	Sumida	86818.18	7	1	2	917	106632.64
江東区北砂4	Koto	81200.00	8	28	23	105	116270.66
大田区羽田6	Ota	77187.50	9	23	13	280	99495.90
足立区千住元町	Adachi	77666.67	10	84	7	333	76546.82

TOKYO CRIME

Top Areas for Crime

location	Ku	averagetarget	totalcrime	averagetargetKu
渋谷区宇田川町	Shibuya	186500.00	796	167309.31
豊島区南池袋１	Toshima	120976.19	601	99299.45
豊島区西池袋１	Toshima	130888.89	499	99299.45
渋谷区道玄坂２	Shibuya	86725.00	494	167309.31
新宿区歌舞伎町２	Shinjuku	111142.86	451	127490.59
豊島区東池袋１	Toshima	141256.36	448	99299.45
大田区蒲田５	Ota	104420.34	426	99495.90
大田区西蒲田７	Ota	95180.72	370	99495.90
北区赤羽１	Kita	103177.45	367	88865.02
港区高輪３	Minato	205166.67	358	203145.53



Super High Crime – Flat Rent

WIND SPEED AND TARGET

Lowest Rents

location	Ku	target	WS(m/s)_std
足立区神明1	Adachi	47750.000000	1.551383
江戸川区鹿骨6	Edogawa	48000.000000	1.267263
足立区平野3	Adachi	49000.000000	1.551383
足立区大谷田4	Adachi	49951.612903	1.551383
江戸川区瑞江3	Edogawa	52000.000000	1.267263
足立区梅田1	Adachi	53488.000000	1.551383
足立区西新井3	Adachi	53580.645161	1.461738
足立区入谷5	Adachi	54125.000000	1.461738
足立区花畑7	Adachi	55333.333333	1.551383
足立区谷中5	Adachi	55462.264151	1.551383
足立区入谷9	Adachi	55666.666667	1.461738
足立区佐野2	Adachi	56135.514019	1.551383
板橋区赤塚5	Itabashi	56260.000000	1.582644
江戸川区新堀2	Edogawa	56263.157895	1.267263
北区岸町2	Kita	56312.500000	1.060387
足立区古千谷本町4	Adachi	56333.333333	1.461738
足立区六月3	Adachi	56355.555556	1.461738
足立区大谷田2	Adachi	56500.000000	1.551383
北区十条仲原4	Kita	57057.142857	1.060387
足立区加平2	Adachi	57066.666667	1.551383

Highest Rents

location	Ku	target	WS(m/s)_std
港区北青山1	Minato	799310.000000	0.596205
千代田区六番町	Chiyoda	586000.000000	0.596205
千代田区九段北2	Chiyoda	549000.000000	0.436493
港区愛宕2	Minato	535000.000000	0.485434
中央区日本橋室町2	Chuo	520000.000000	0.436493
港区赤坂1	Minato	501000.000000	0.485434
渋谷区広尾4	Shibuya	461861.095890	0.485434
渋谷区神南1	Shibuya	449543.859649	0.702081
千代田区五番町	Chiyoda	444750.000000	0.596205
千代田区永田町2	Chiyoda	436800.000000	0.596205
中央区銀座4	Chuo	434000.000000	0.851103
千代田区富士見1	Chiyoda	432166.666667	0.436493
港区北青山2	Minato	424375.000000	0.596205
港区元麻布1	Minato	421857.142857	0.485434
港区愛宕1	Minato	420000.000000	0.485434
港区六本木6	Minato	414944.444444	0.485434
渋谷区神宮前4	Shibuya	408705.882353	0.702081
千代田区四番町	Chiyoda	407060.000000	0.596205
千代田区鍛冶町1	Chiyoda	400000.000000	0.436493
港区南青山5	Minato	394080.000000	0.702081
渋谷区広尾2	Shibuya	386666.666667	0.485434

Lowest WS

location	Ku	target	WS(m/s)_std
文京区本郷5	Bunkyo	99040.277778	0.436493
台東区小島1	Taito	114408.450704	0.436493
台東区蔵前3	Taito	124857.142857	0.436493
千代田区一番町	Chiyoda	376482.352941	0.436493
千代田区九段北1	Chiyoda	142107.826087	0.436493
中央区日本橋横山町	Chuo	208686.567164	0.436493
中央区日本橋茅場町2	Chuo	155345.215385	0.436493
中央区日本橋人形町2	Chuo	131280.952381	0.436493
中央区東日本橋3	Chuo	148192.307692	0.436493
中央区日本橋馬喰町1	Chuo	153809.375000	0.436493
千代田区内神田1	Chiyoda	206435.897436	0.436493
中央区日本橋蛎殻町1	Chuo	131241.645570	0.436493
千代田区三番町	Chiyoda	349766.666667	0.436493
千代田区岩本町2	Chiyoda	133706.250000	0.436493
中央区日本橋堀留町1	Chuo	102767.241379	0.436493
墨田区両国2	Sumida	107677.949153	0.436493
中央区日本橋浜町2	Chuo	134381.338983	0.436493
墨田区両国3	Sumida	137603.448276	0.436493
中央区日本橋富沢町	Chuo	167111.111111	0.436493
台東区台東3	Taito	175794.642857	0.436493
中央区日本橋堀留町2	Chuo	128592.592593	0.436493

Highest WS

location	Ku	target	WS(m/s)_std
江戸川区東葛西6	Edogawa	85485.074627	2.010427
江戸川区中葛西3	Edogawa	82963.190184	2.010427
江戸川区中葛西5	Edogawa	81886.178862	2.010427
江戸川区中葛西8	Edogawa	70175.000000	2.010427
江戸川区東葛西4	Edogawa	82870.229008	2.010427
江戸川区東葛西5	Edogawa	87086.956522	2.010427
江戸川区西葛西7	Edogawa	83330.508475	2.010427
江戸川区東葛西8	Edogawa	80166.019417	2.010427
江戸川区中葛西7	Edogawa	74559.633028	2.010427
江戸川区南葛西6	Edogawa	91639.639640	2.010427
江戸川区南葛西2	Edogawa	87494.736842	2.010427
江戸川区西葛西6	Edogawa	94106.382979	2.010427
江戸川区西葛西3	Edogawa	92386.904762	2.010427
江戸川区中葛西6	Edogawa	104914.772727	2.010427
江戸川区東葛西7	Edogawa	104466.666667	2.010427
江戸川区南葛西3	Edogawa	100137.662338	2.010427
江戸川区西葛西5	Edogawa	93714.285714	2.010427
江戸川区南葛西4	Edogawa	93123.076923	2.010427
江戸川区西葛西2	Edogawa	101620.000000	2.010427
江戸川区中葛西4	Edogawa	76272.727273	2.010427
江戸川区西葛西8	Edogawa	69513.888889	2.010427

WIND SPEED AND TARGET