Predicting **RESIDENTIAL REAL ESTATE** In Taipei,Taiwan

Workflow

01
Problem Statement

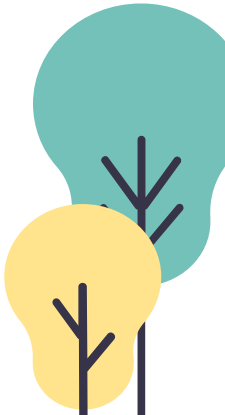
02
Data Acquisition & Preprocessing

03
Cluster

04
Predictive Model: Linear Model

05
Socioeconomic Analysis

06
Conclusion



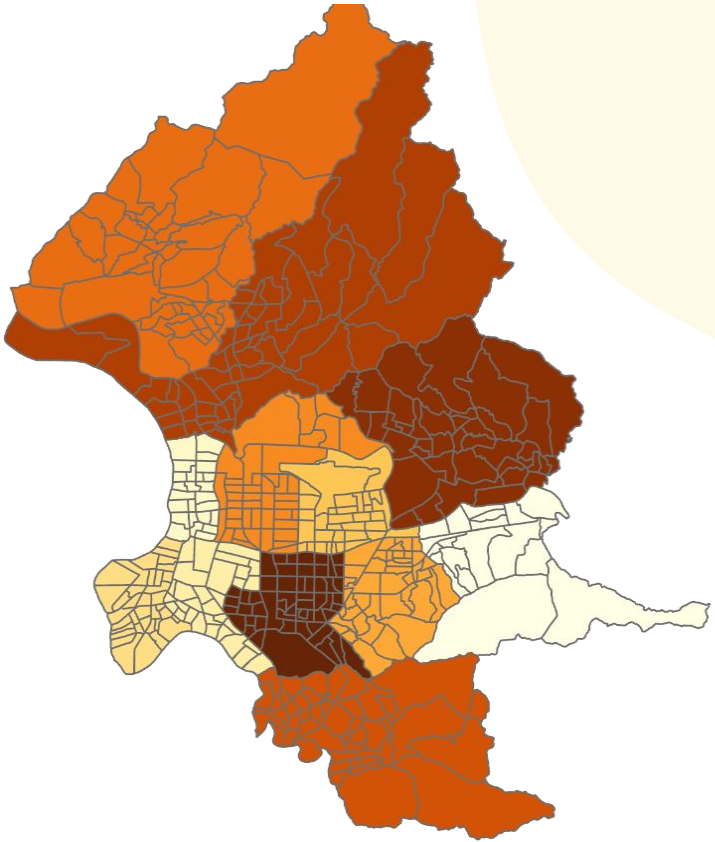
Problem Statement

- Housing prices around the world have experienced **tremendous increases,** and this has created business opportunities while **worsening inequality.**

- The ability to develop a city that is **balanced in living conditions,** prices, and ability to invest accurately in real estate properties has become essential skills for city governments and its citizens.

Through this project, our team would attempt to evaluate the impact of certain **points of interests (POIs)**, or **“residential attractors"**, on the housing market within the city.

Taipei in Numbers



Total Area

271.80 km²

Districts

12

Density

9,071.42 / km

Population

2,465,610

Avg. House Price

US$ 924,249

Data Acquisition

2020

Housing Sales Price
Taiwan Ministry of Interior

40000+ Entries
Price & Properties

2022

Point of Interest (POI)
Google Maps API

4118 Entries
12 POI Types
3 Distances
(500m, 1km, 3km)

2021

Socioeconomic Factors
Taiwan Ministry of Interior,
Ministry of Finance,
Taipei City Government

Education, Income per Sub-district,
Population per district

2021

Human Activity Flow
Taiwan Ministry of Interior
Social Economical GIS (SEGIS)

Phone Signal Data
Workday vs. Weekend
Day vs. night
(District Level)

Data Pre-Processing

Translation

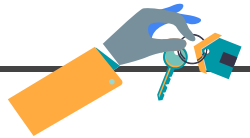
Data CSV file contained columns record in Traditional Chinese

Data Cleaning

Certain houses have missing information or unique info for specific columns

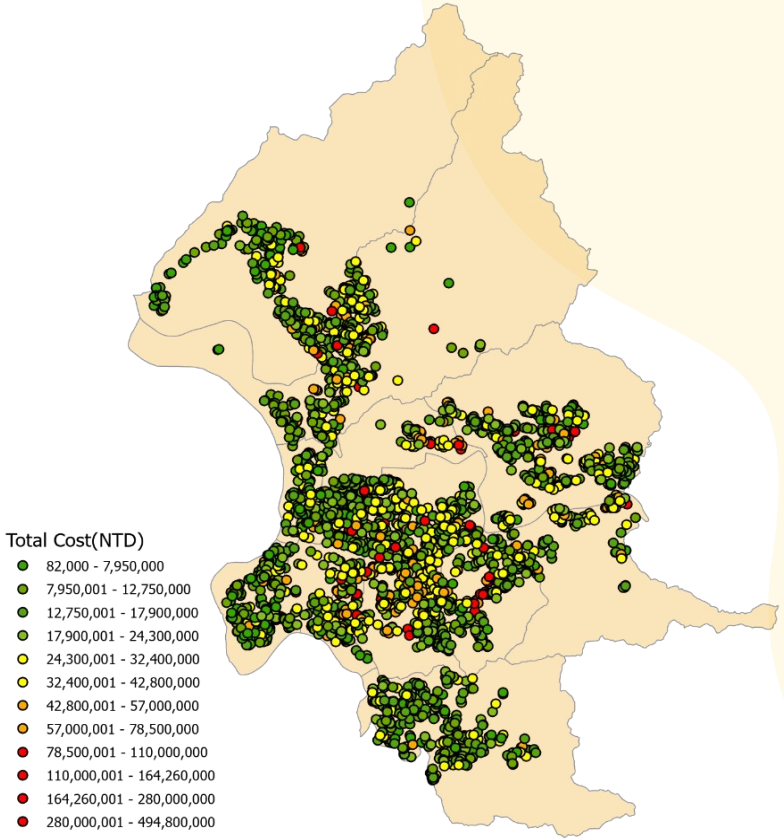
Data Transformation

Normalization of values, One-hot encode



Ready for Analysis

Data Visualization: Total Price



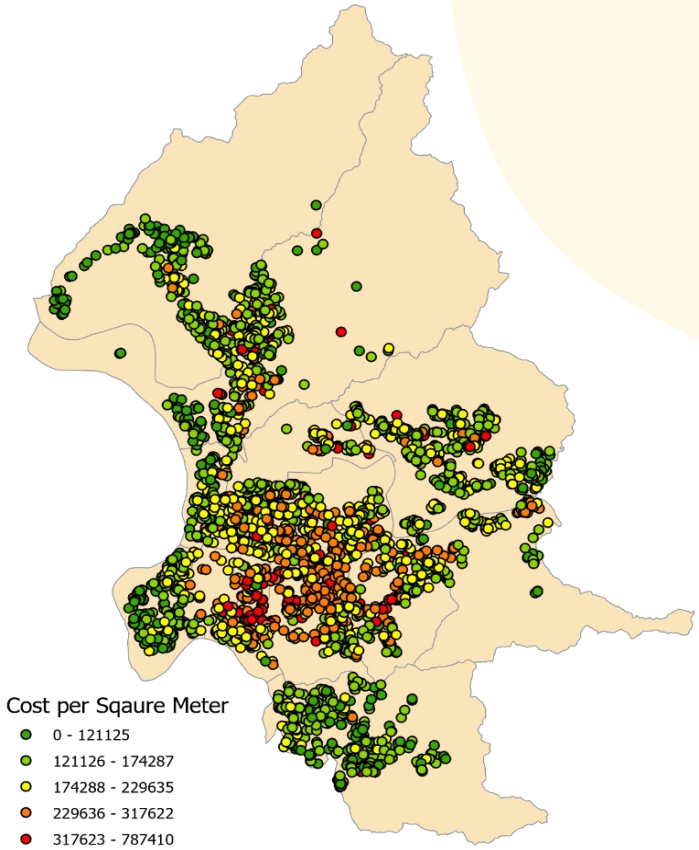
- ≤ NTD $25,000,000 (USD $817,750)
- ≥ NTD $80,000,000 (USD $2,616,800)

There are a few red spots in most of the districts.

Cluster of medium to expensive houses around city center and sub-centers.

Areas further away from the city center have lower house prices on average.

Data Visualization: Price per Area



- ≤ NTD $200,000 (USD $6,542)
- ≥ NTD $400,000 (USD $13,084)

This clears up the distribution of prices. We can see natural clustering in many districts.

POIs






Police Station
Supermarkets
Hospitals

Bus station
SubwayStation

University
Primary School
Library

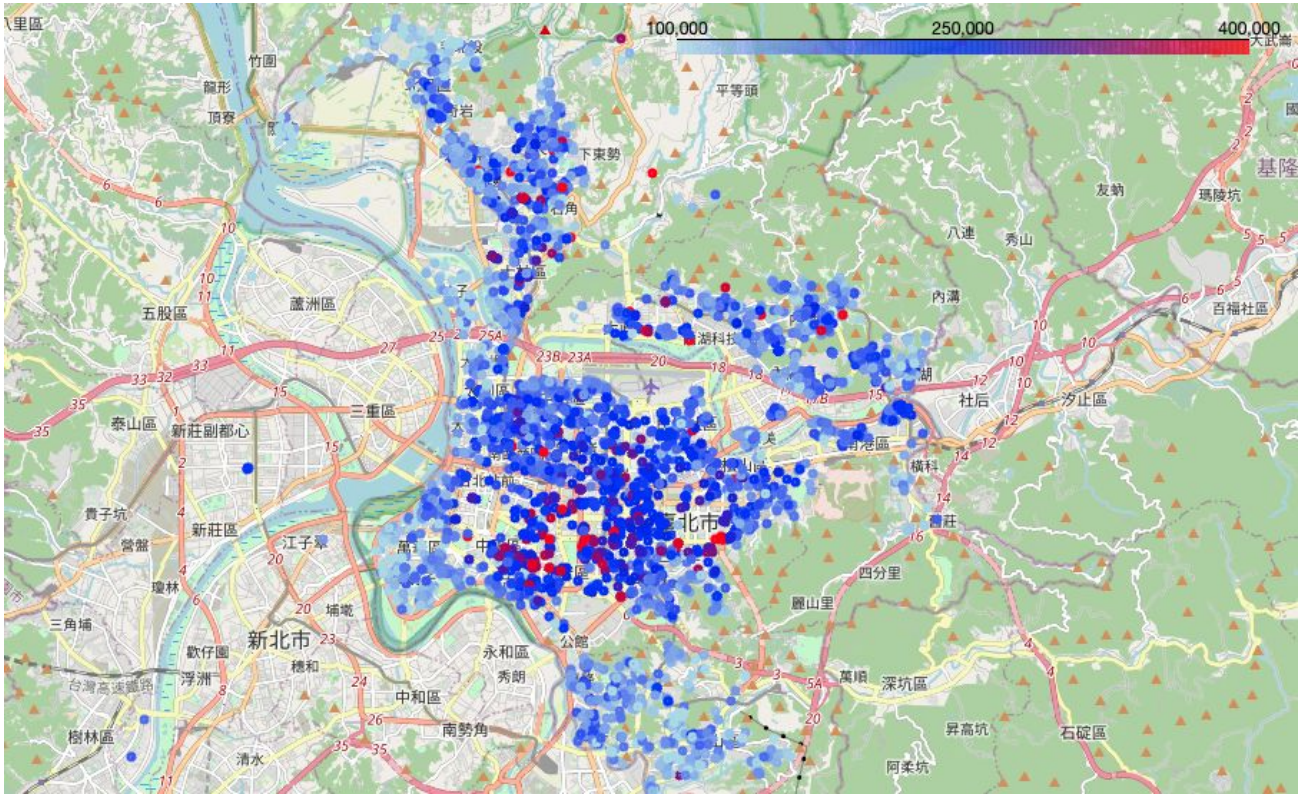
Church
Night Club
Shopping Mall
Park

500 meters
Walking Distance

1 kilometer
Public Transport/
Scooter Distance

3 kilometers
Car Distance

Data Visualisation : Price per Square Meter (m^2)



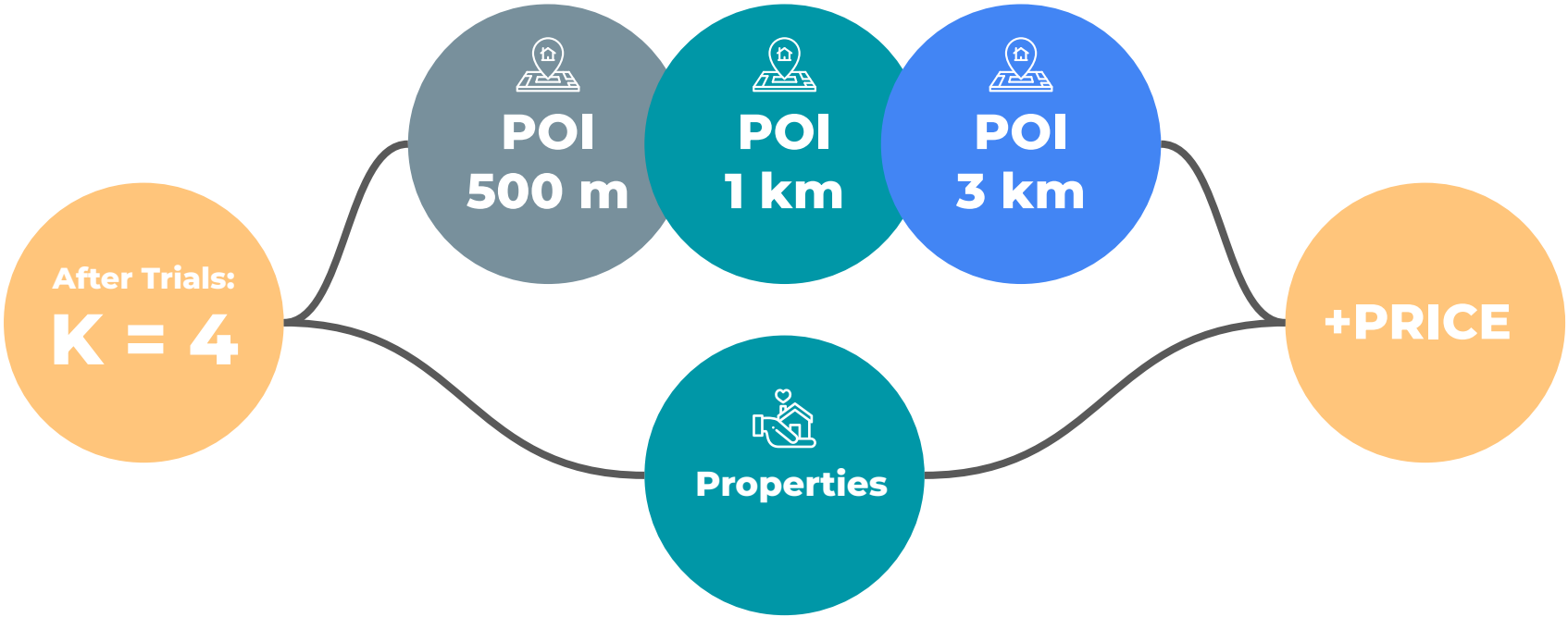
Index:

- ≤ NTD $100,000 (USD $3,231)
- ≥ NTD $400,000 (USD $12,925)

This clears up the distribution of prices even more clearly. Colors for each districts and areas are less different.

The red spots have also decreased as this takes the size of houses into consideration.

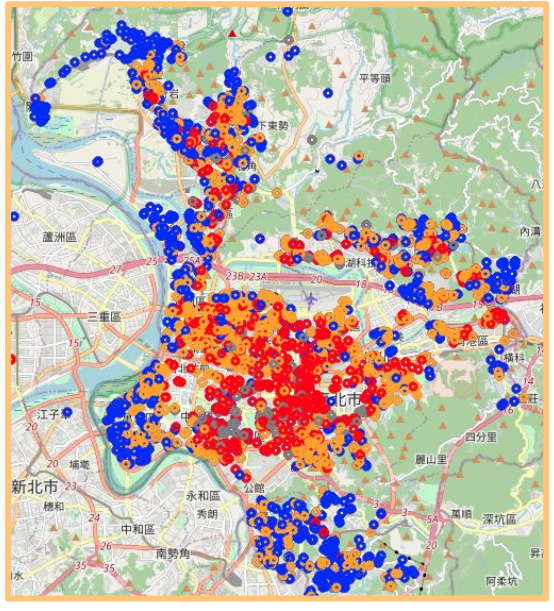
Clustering: K-Means



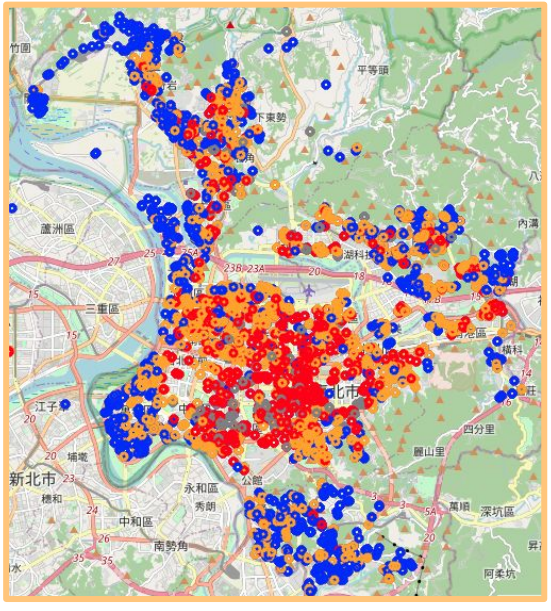
POI vs Properties

Cluster Legend:
0: Red, 1: Blue,
2: Gray, 3: Orange

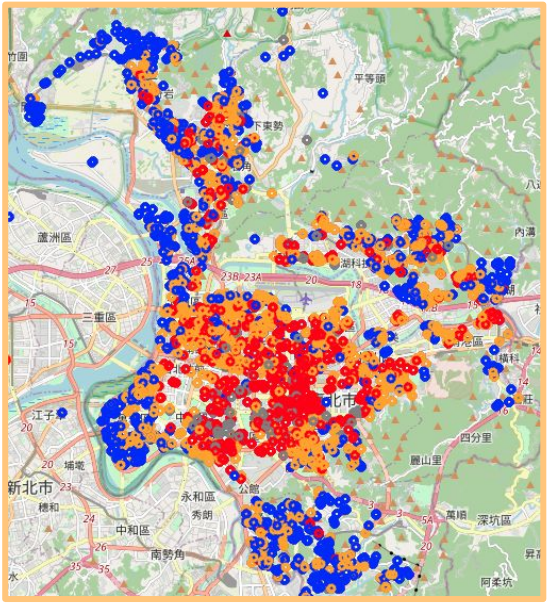
K-Means: POIs + Price



500m



1 km

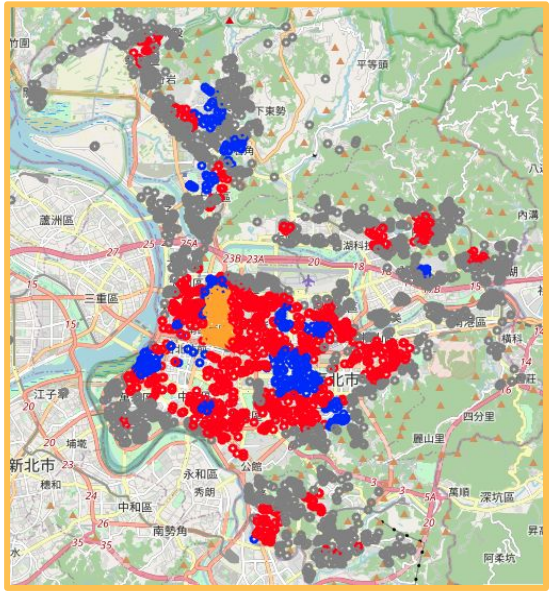


3 km

When price of the house is taken into consideration, it skews the results significantly. Therefore, it should not be included in the clustering model.

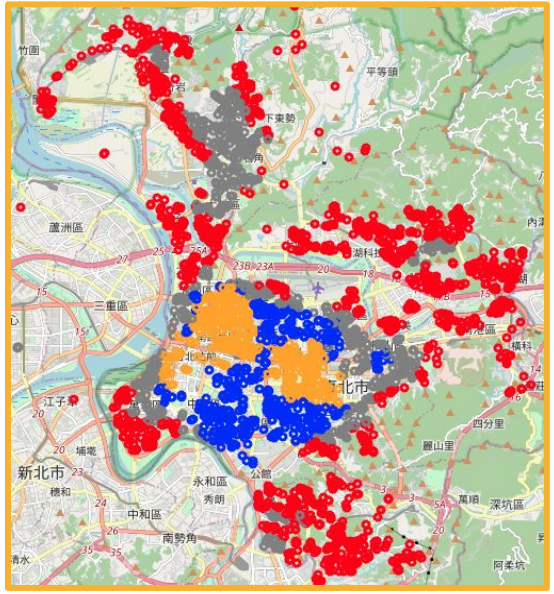
Cluster Legend:
0: Red, 1: Blue,
2: Gray, 3: Orange

K-Means: Only POIs



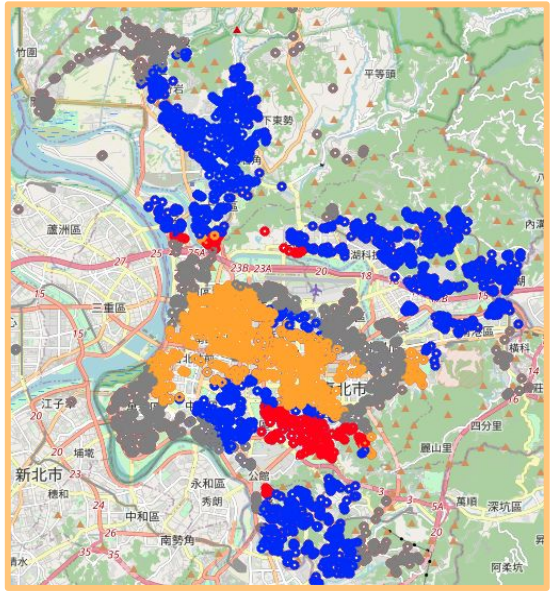
500m

There are small clusters everywhere in the city.



1 km

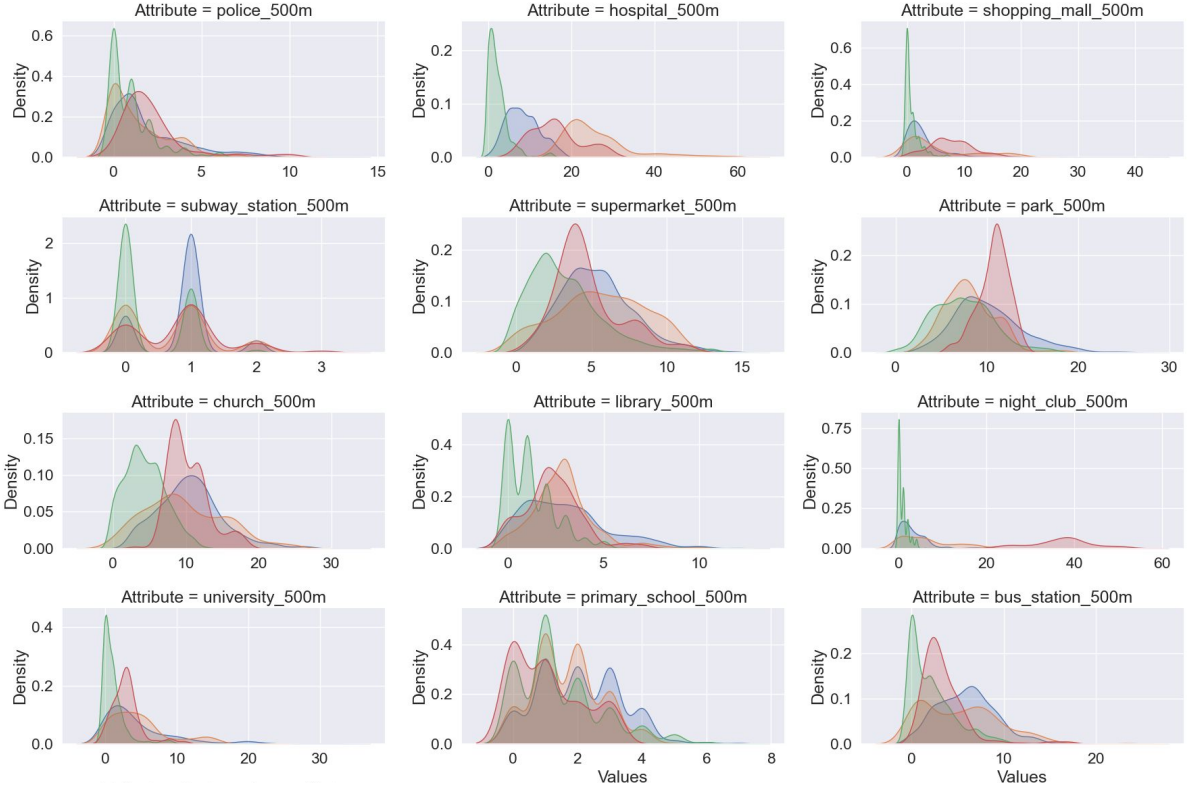
Two main cluster appear in the city, one is old town and CBD.

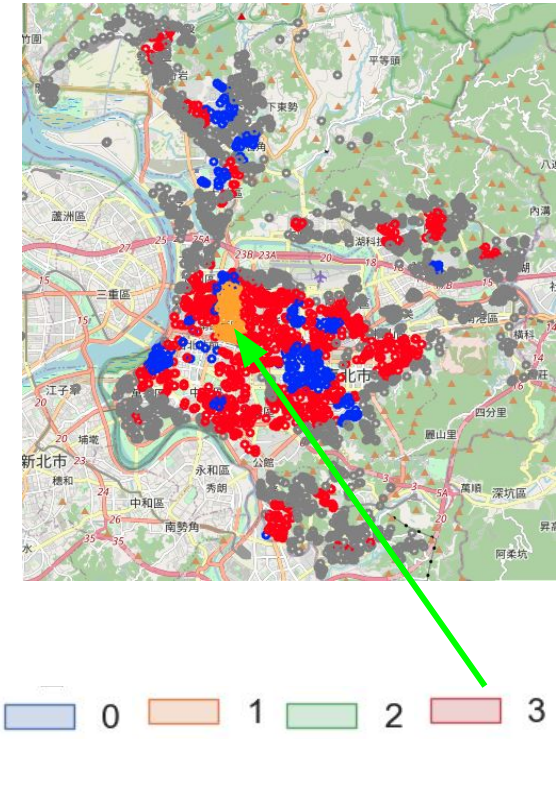


3 km

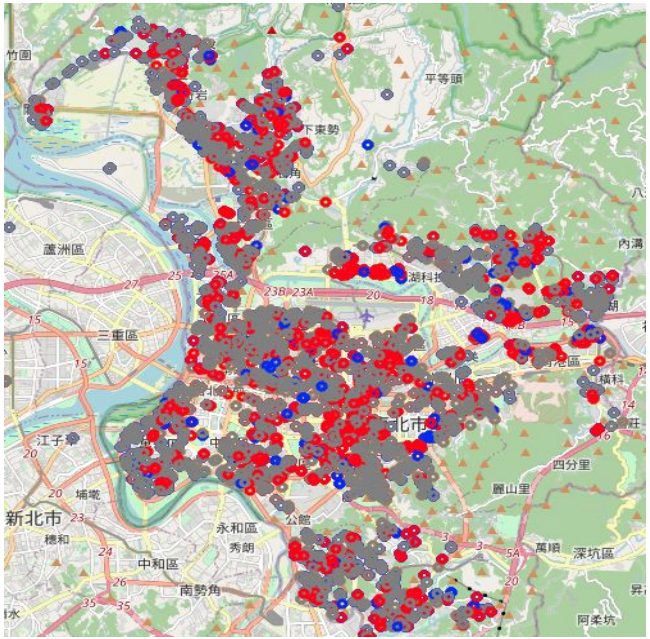
The city is separated into the inner cluster, middle, and outer.

POI distributions within 500m in different clusters

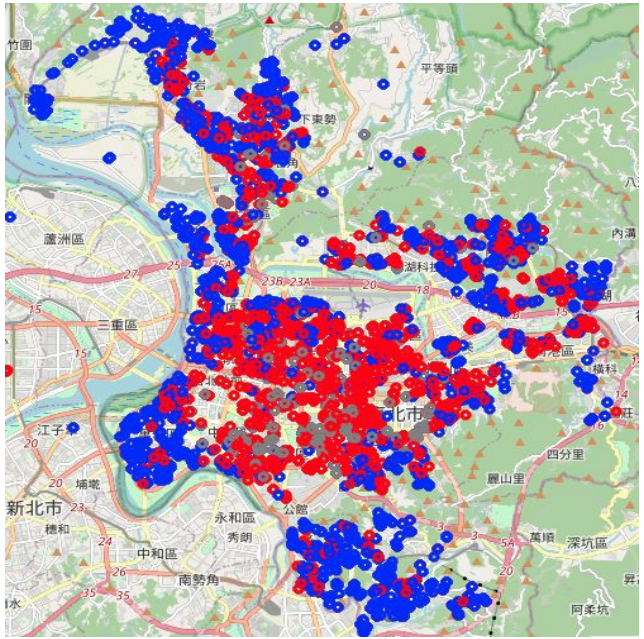




K-Means: Properties / Properties + Price



Properties



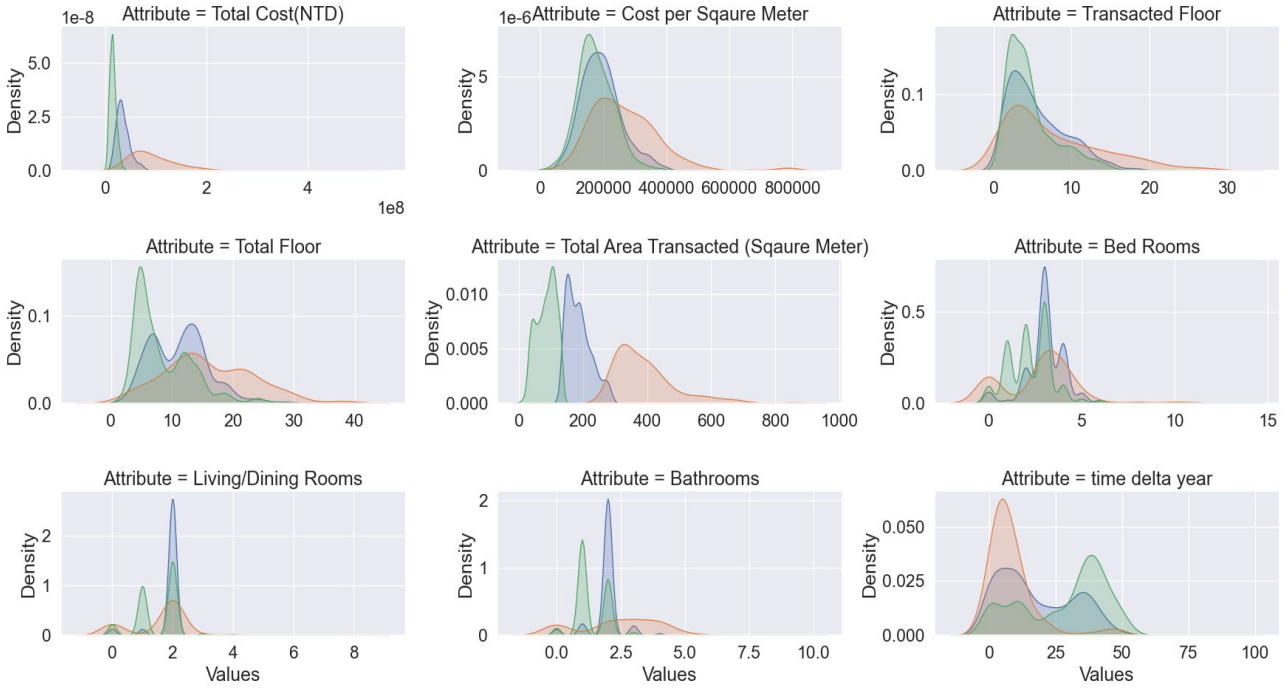
Properties + Price

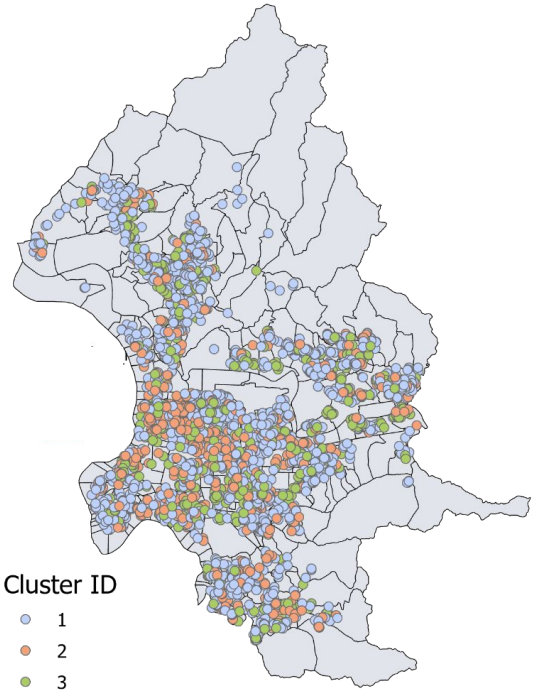
Findings:
When price is considered, it identifies the city centers and areas outside of city centers.

Therefore, a closer look at the clustering without price is needed.

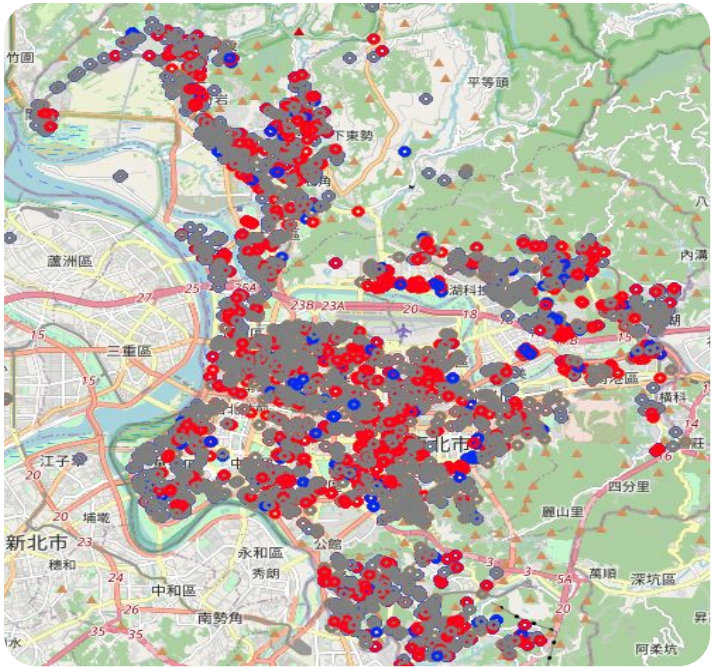
Cluster Legend:
0: Red, 1: Blue, 2: Gray

Distributions of properties in each cluster





Cluster Legend:
0: Red, 1: Blue,
2: Gray,



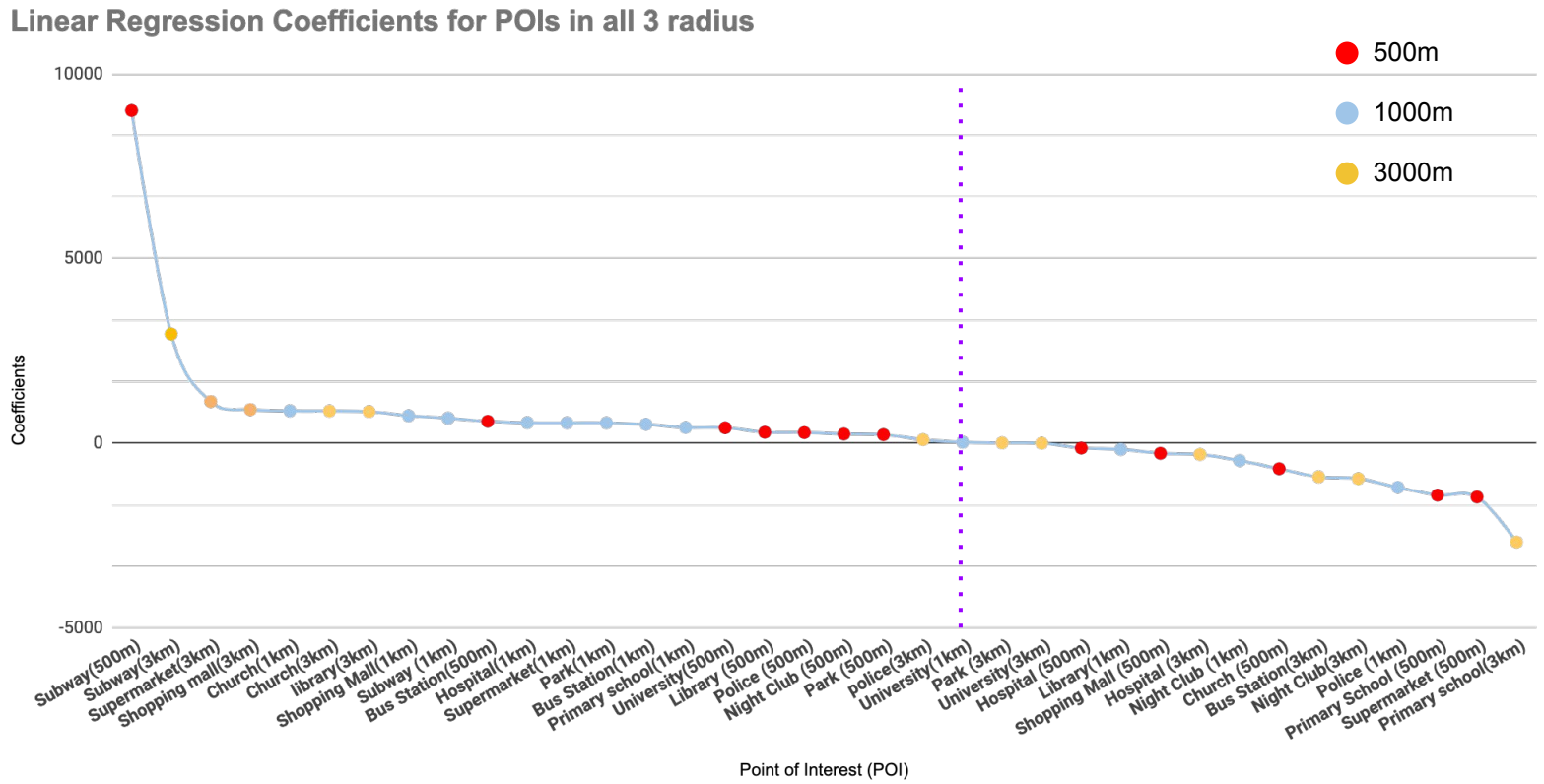
Properties

	Cluster 0	Cluster 1	Cluster 2
Total Cost	34458925	99205981	14354527
Cost m2	193140	269623	175929
Floor	5.6845	7.74	4.80
Total floor	11.318	16.051	8.4
Total Area	186.191	391.259846	83.94
Bed rooms	3.079381	2.77	2.24
Liv/Dine Rm	1.9020	1.64	1.49
Bathrooms	1.948454	2.46	1.39
Age	17	8.32	28

Linear Regression Model

$$Unit\ Price = \sum_{i} \alpha_i * F_i + bias$$

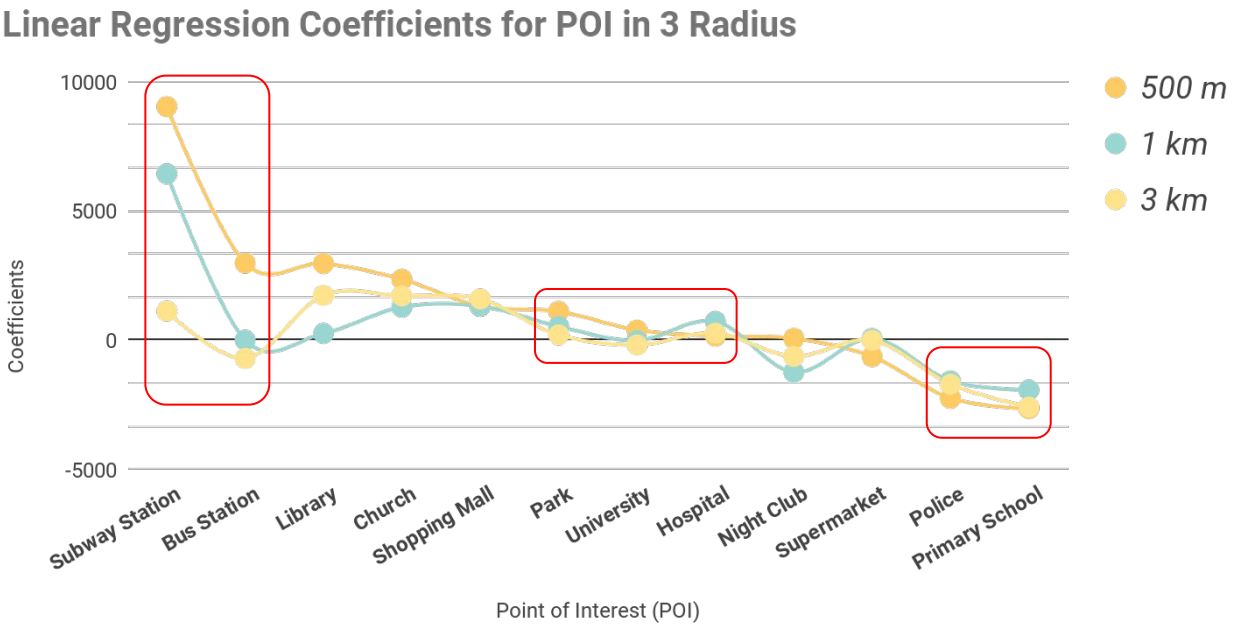
Linear Regression Model — All POI in 3 Radius



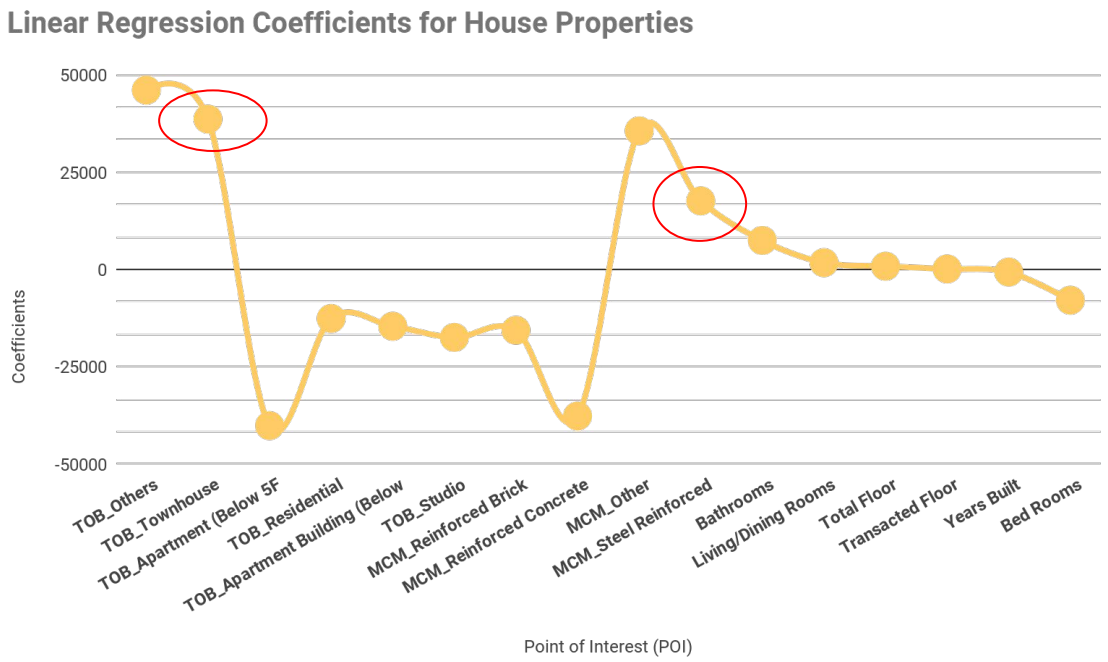
Linear Regression Model — POI

Findings:

- **Subway stations** and **Bus stations** are important in shorter distances, and becomes less important further away.
- At 3 km, **library, church, and shopping malls** has the highest importance since these POIs may only be occasional visits.
- **Park** and **University**'s importance would decrease as it becomes further away from the house.
- **Hospitals** generally have a limited positive influence.
- **Nightclubs** influence is still unclear, as the influence does not correlate with the distance.
- **Supermarkets, police, and primary schools** have negative influence, since these POIs may be prevalent in most density areas.



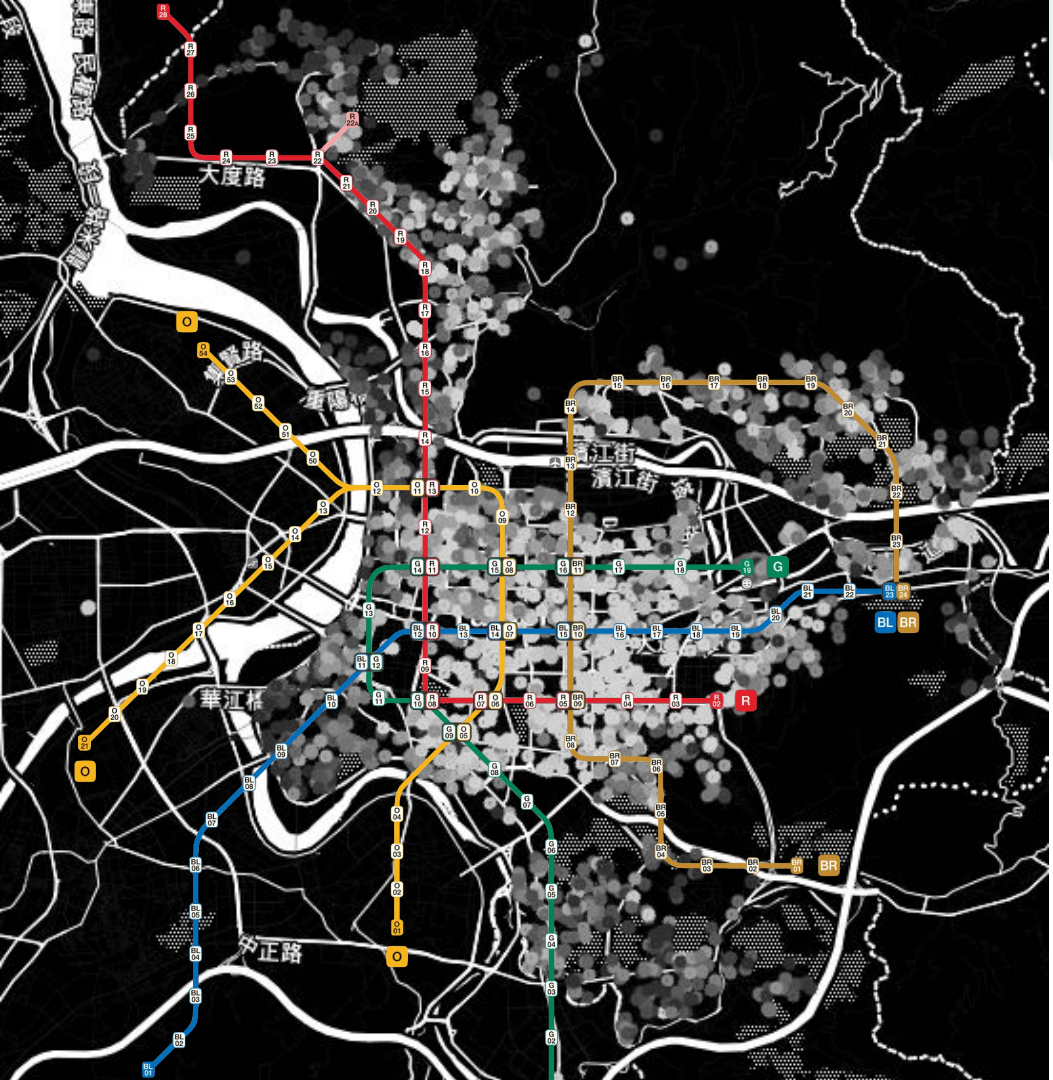
Linear Regression Model — House Properties



TOB ⟶ Type of Buildings
MCM ⟶ Main Construction Method

Findings:

- **TOB_Others** include office buildings, factories, and storefront that are registered as residentials, so it could be considered as an outlier.
- **Townhouses** are rare in Asian cities and higher priced as seen in the model. Other types of buildings does not have large influence as townhouse does.
- **Steel Reinforced Concrete(SRC)** has a large influence compared to other factors, and as most houses built are **Reinforced Concrete(RC)** it has less influence.

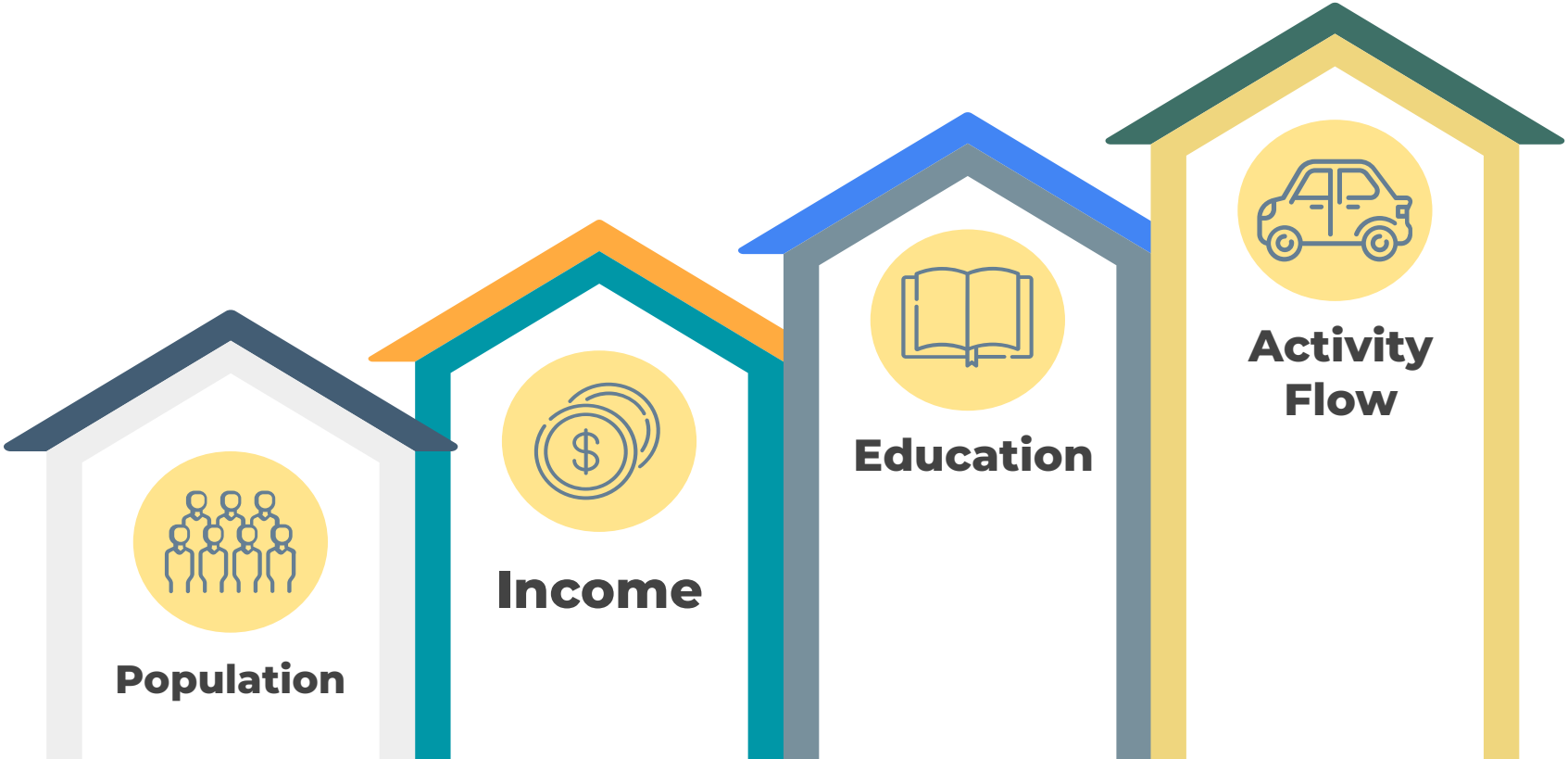


In- Situ Validation:

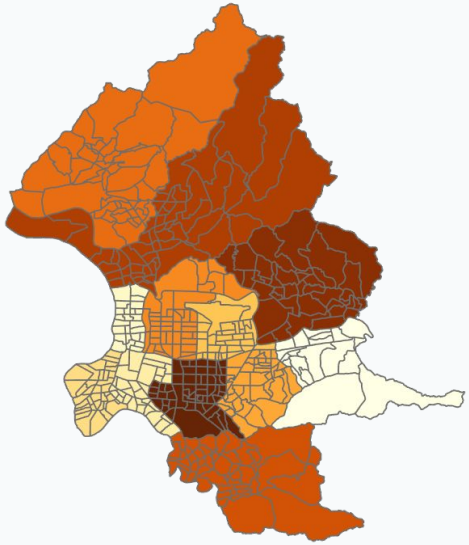
Subway Station's High importance seen on real map

Lighter areas on this map is most expensive.

Socioeconomic Factors

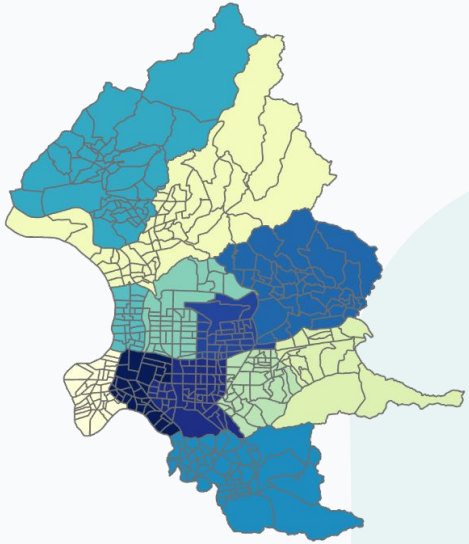


Socioeconomic Factors Overview



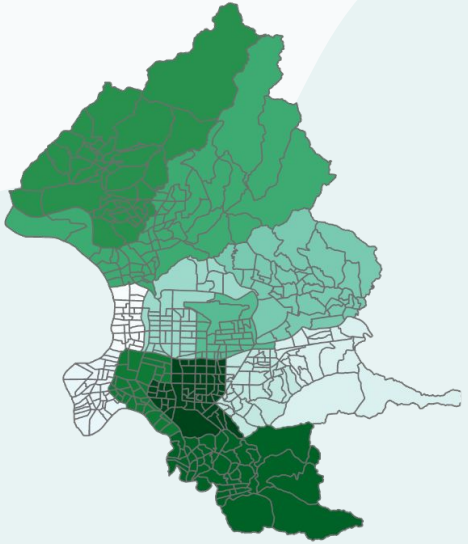
Population

113612	202369 - 210776
113613 - 118354	210777 - 239780
118355 - 146970	239781 - 256418
146971 - 172819	256419 - 263360
172820 - 187993	263361 - 271749
187994 - 202368	271750 - 281411



Avg Income

574	999 - 1010
575 - 855	1011 - 1041
856 - 914	1042 - 1053
915 - 919	1054 - 1254
920 - 944	1255 - 1420
945 - 998	1421 - 1849

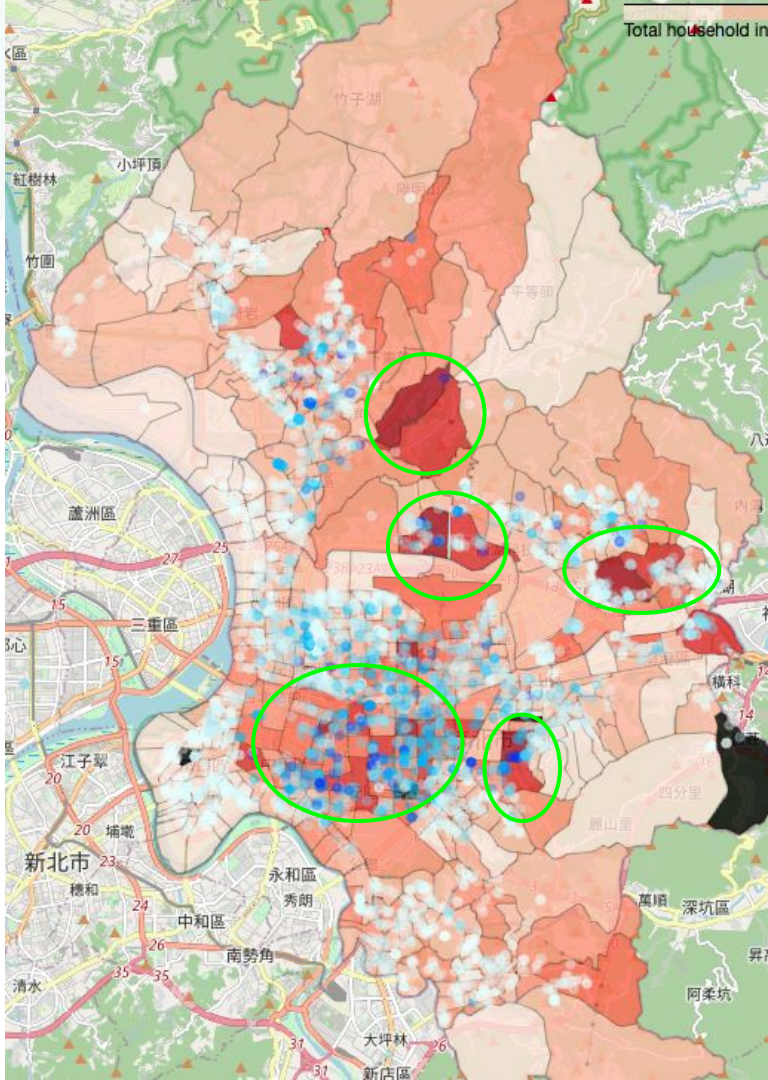


Doctorate degree

694	1956 - 2029
695 - 890	2030 - 2131
891 - 1018	2132 - 2142
1019 - 1706	2143 - 2422
1707 - 1954	2423 - 3347
1955	3348 - 5749

Average Income vs. Total House Price

Houses with higher cost per square meter tends to land in places with higher average income.



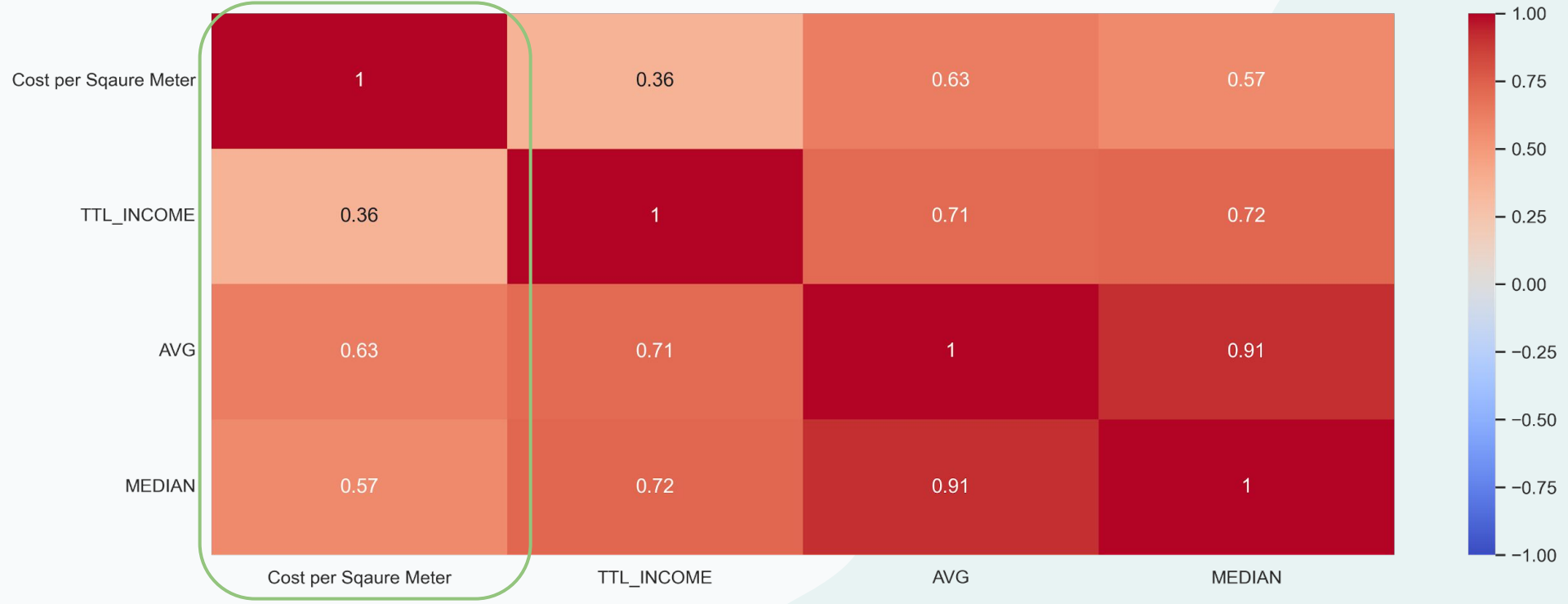
Average Income in NTD in Thousands

516 791 1,066 1,341 1,615 1,890 2,165

Cost per Square Meter

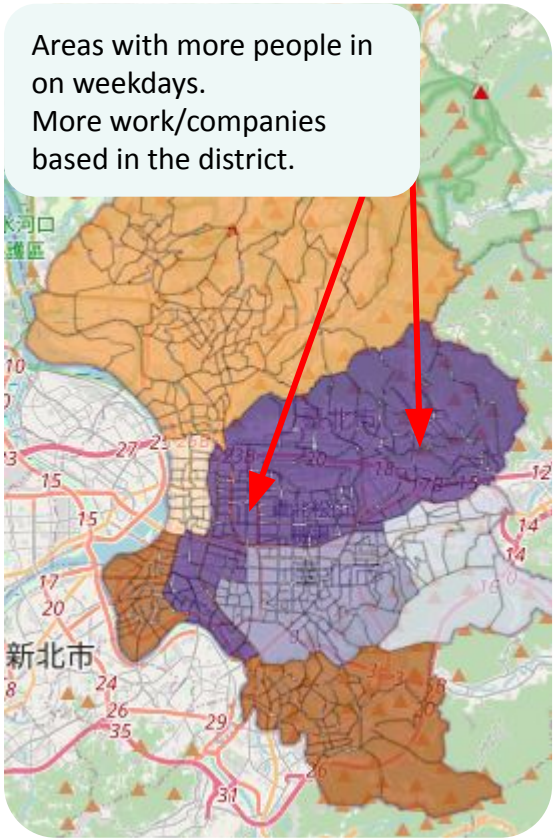
100,000 160,000 220,000 280,000 340,000 400,000

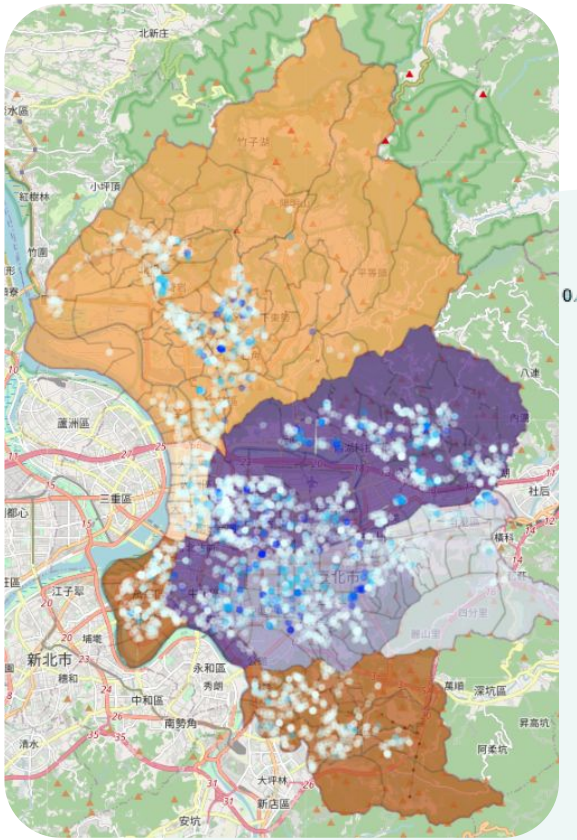
House Price v.s. Income Correlation

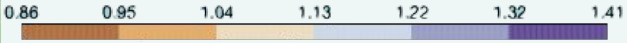


Activity Flow Day Time

Areas with more people in on weekdays.
More work/companies based in the district.







Weekday Daytime Count / Weekend Daytime Count

If ratio:

- > 1: Higher Weekday Count
- = 1: Small or No Differences
- < 1: Lower Weekday Count

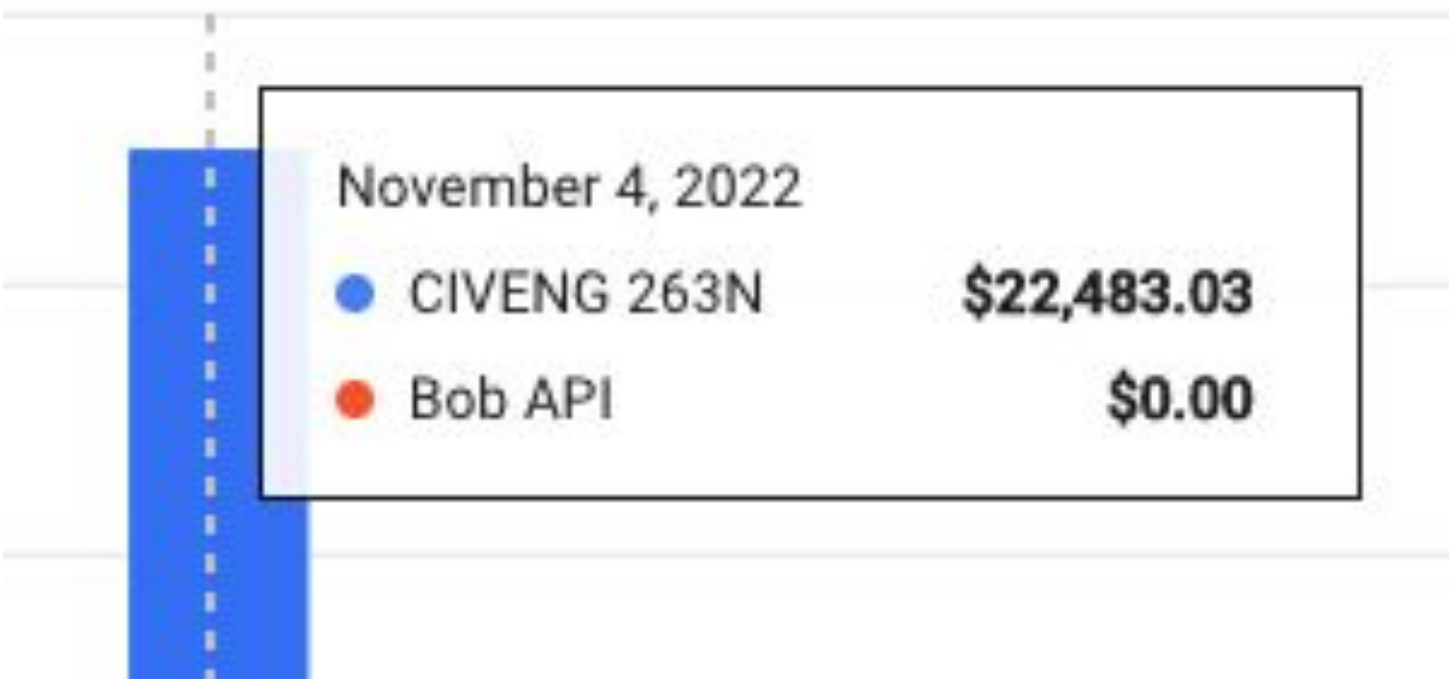
Therefore:

- >1 would represent more people working, and <1 would be more people living there or leisure zones.

Insights

1. **Transit-oriented Development (TOD):**
    - Transportation POIs, subway stations, and bus stations has high importance
        - TOD development methodology help create a more sustainable city.
        - **Walk more to save your wallet!**
    - POIs for leisure (shopping mall) are important if there are at a certain distance (3km).
    - Good city structure makes common POIs (hospital, park, universities) less important.
2. **House Properties:**
    - Important but not as POIs do.
3. **Socioeconomics:**
    - **Earn more, liver better**
    - Working areas have the most expensive houses
        - **Spend some time commuting to save money!**

Data is not free!



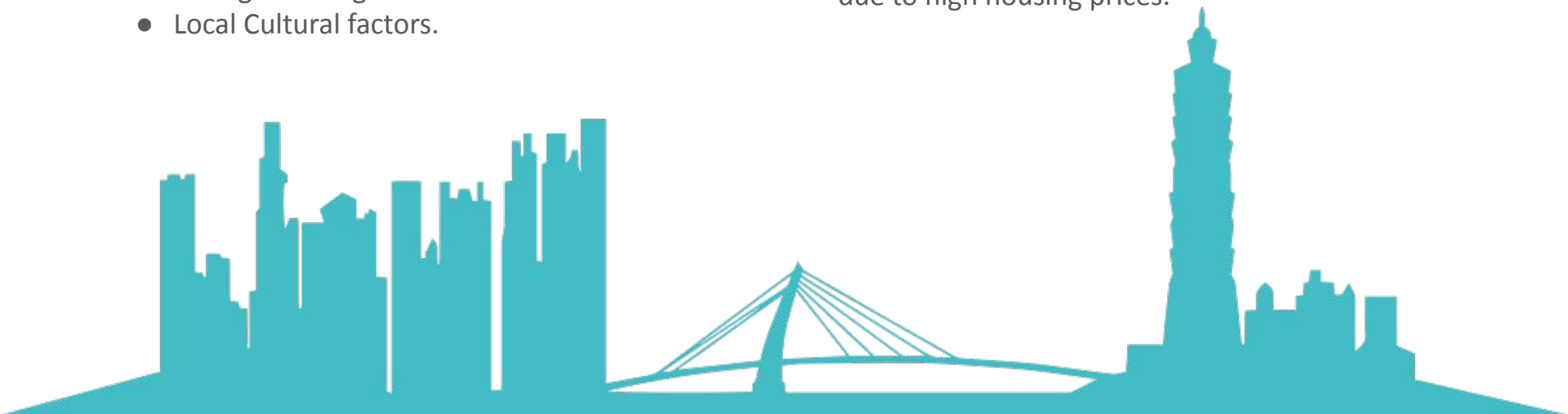
Challenges, Caveats, and Opportunities

Challenges & Caveats

- Cost of Google Maps POI Query
  - USD $22,000 Bill
- POIs change over time.
- Data are not from the same year.
- Certain data have been dropped due to large missing entries.
- Local Cultural factors.

Opportunities

- Sustainable city development.
- Improve equality, housing access and purchase power.
- Reduce further urban sprawl due to high housing prices.



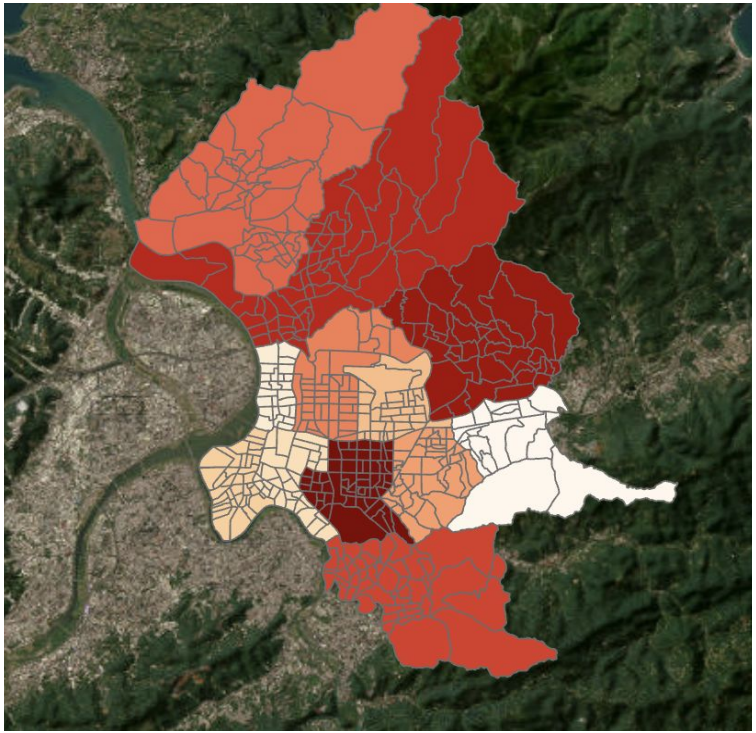
Thank you!

Questions?

Appendix Slides

Questions?

Taipei City Background Information



Administrative Districts:	12
Total Area:	271.80 km²
Compared to San Francisco	121.0 km²
Household Registered Population	2,465,610
Metropolitan Area Population:	6,894,643
Population Density:	9,071.42 / km
GDP Per capita	$35,513
Purchasing power parities (PPP)	US$ 760,00 in 2019
Average House Price	US$ 924,249 in 2020

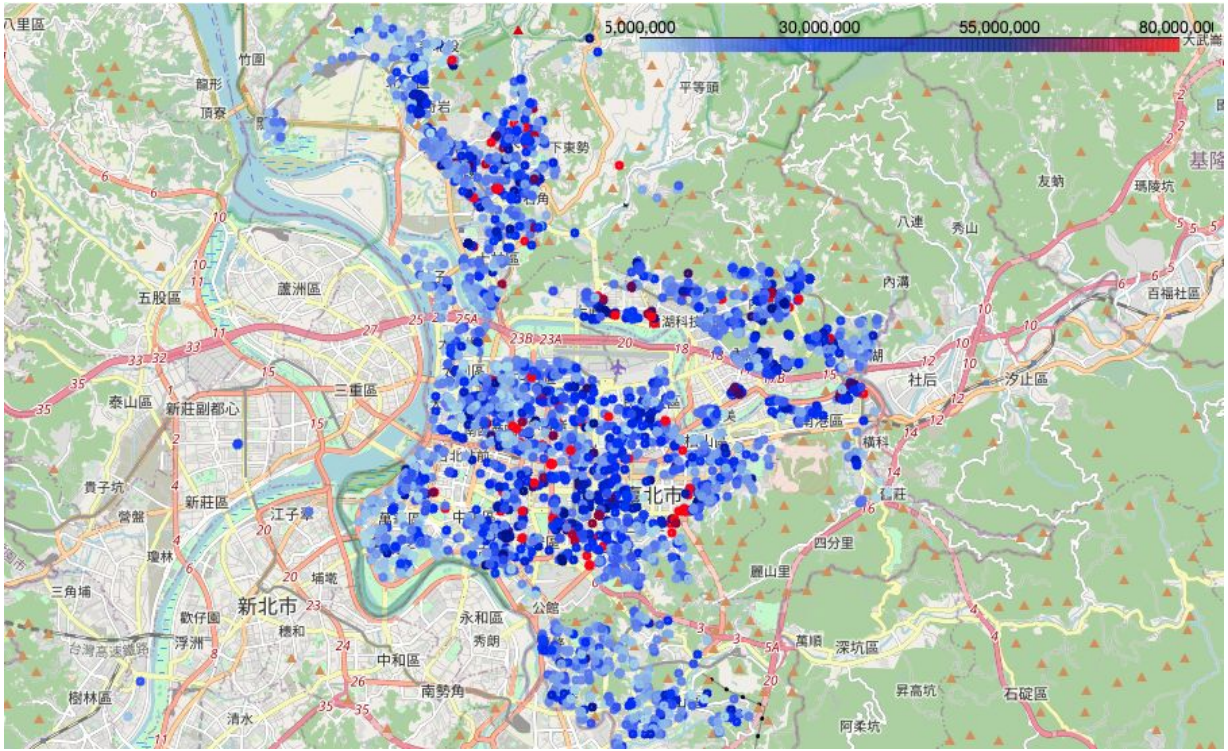
Point of Interest List & Background Information






POI	Total Number	Remarks
Police Station	104	
Hospital	36 Hospital 3,541 Clinics	478 in Taiwan
Supermarket	211	
Bus Station	280 Lines, 2437 stops	
Subway Station	131	6 Lines
Library	76	
University	28	
Primary School	151	142 Public, 9 Private
Church	420	Buddhist/Tao Temple: 1857
Nightclub	N/A	Definition Uncom
Shopping Mall	62	1 per 88k population
Park	616	

Data Preview: Total Price



Index:

▣ ≤ NTD $5,000,000 (USD $161,574)

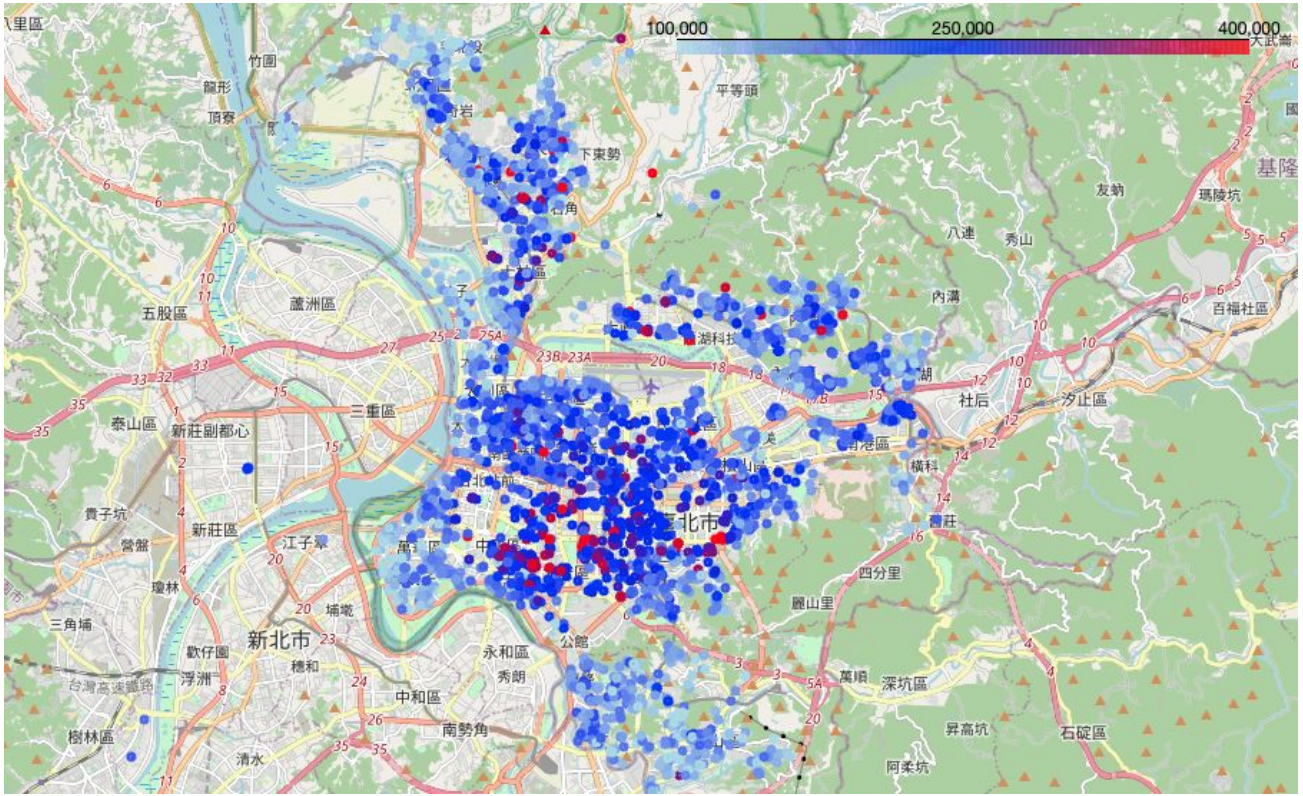
▣ ≥ NTD $80,000,000 (USD $2,585,184)

There are a few red spots in most of the districts.

Cluster of medium to expensive houses around city center and sub-centers.

Areas further away from the city center have lower house prices on average.

Data Preview: Price per Square Meter (m²)



Index:

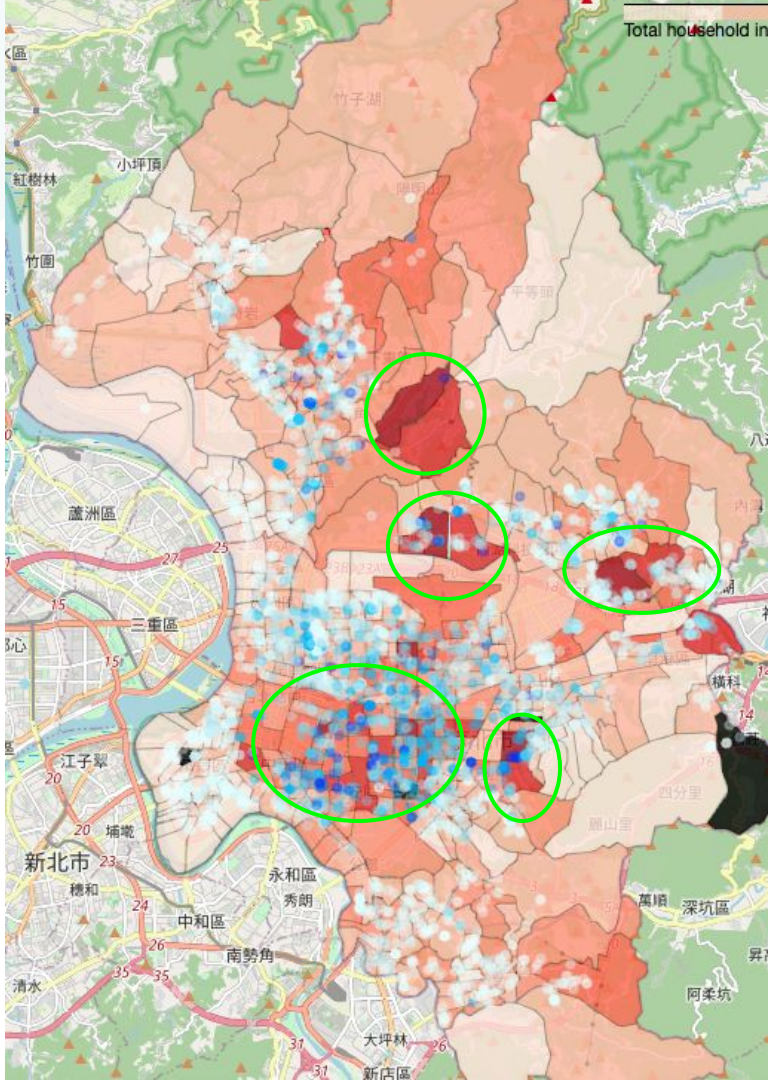
- ≤ NTD $100,000 (USD $3,231)
- ≥ NTD $400,000 (USD $12,925)

This clears up the distribution of prices even more clearly. Colors for each districts and areas are less different.

The red spots have also decreased as this takes the size of houses into consideration.

Average Income vs. Total House Price

Houses with higher cost per square meter tends to land in places with higher average income.



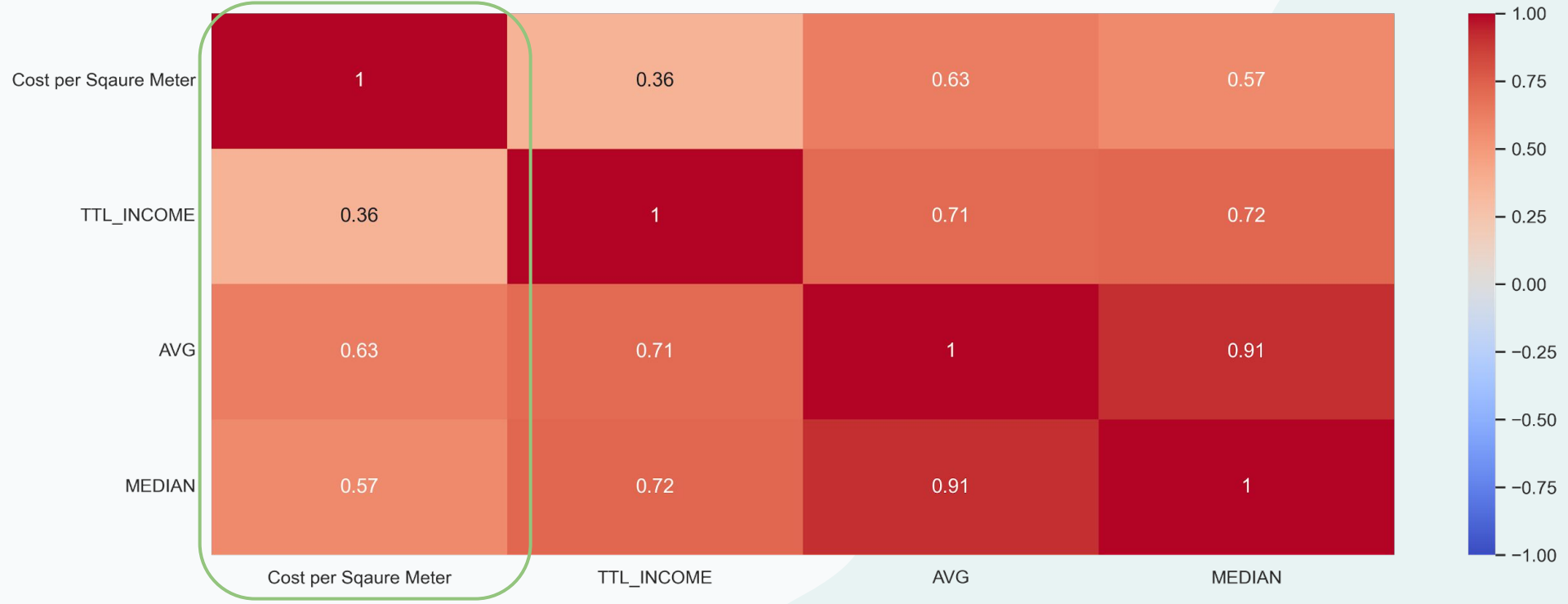
Average Income in NTD in Thousands

516 791 1,066 1,341 1,615 1,890 2,165

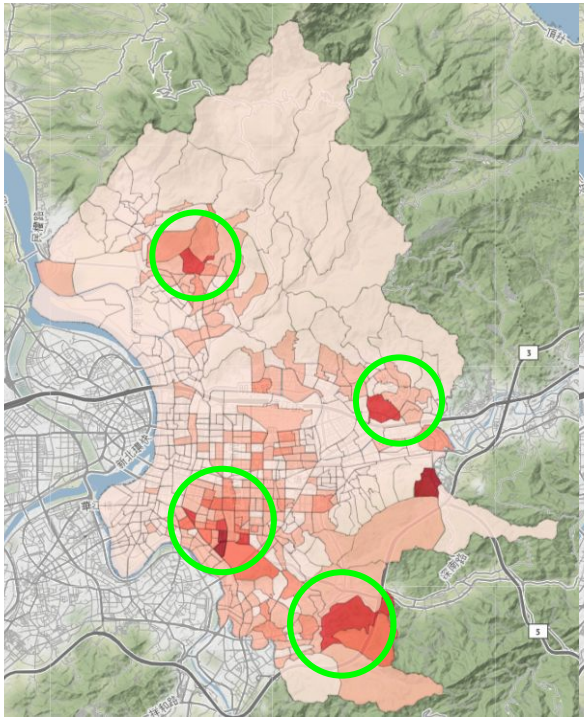
Cost per Square Meter

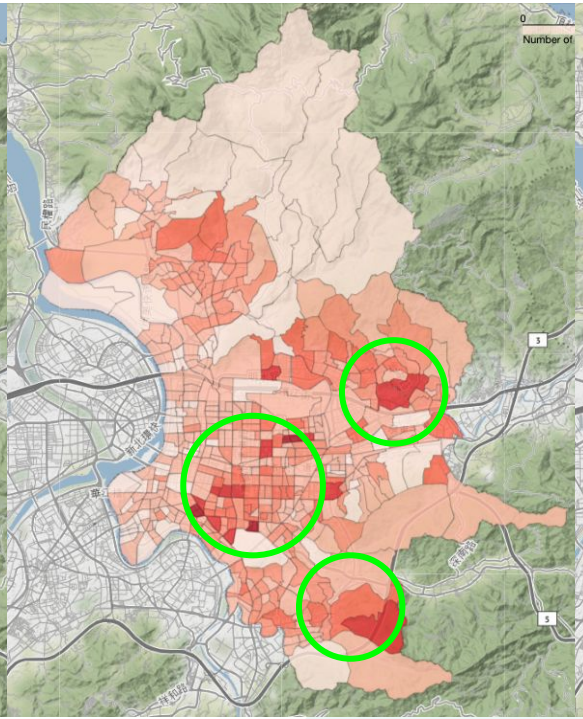
100,000 160,000 220,000 280,000 340,000 400,000

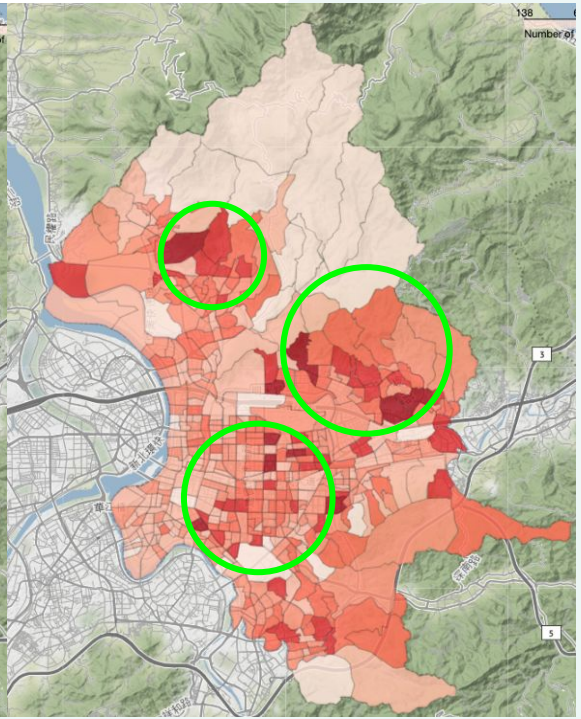
House Price v.s. Income Correlation



Education

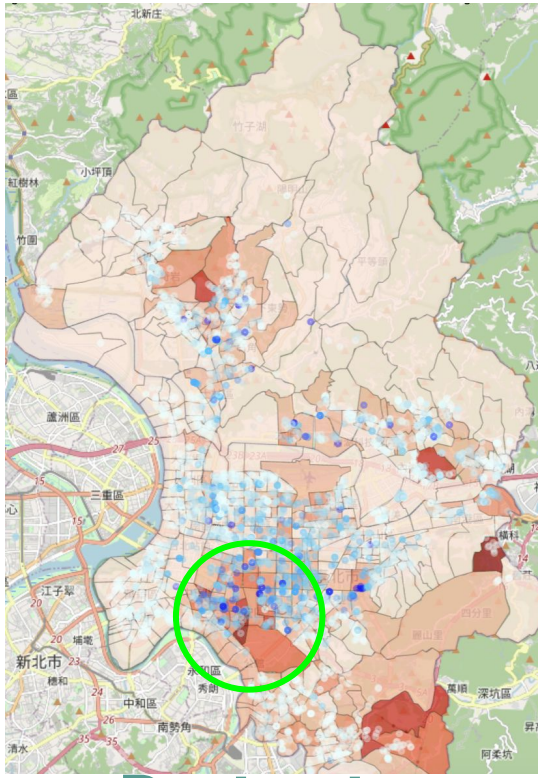




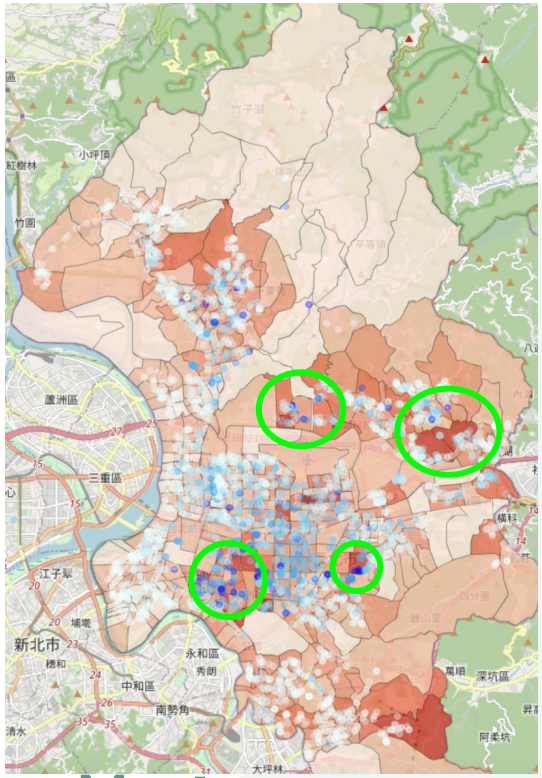


- **Doctorate**
- **Masters**
- **Undergrad**

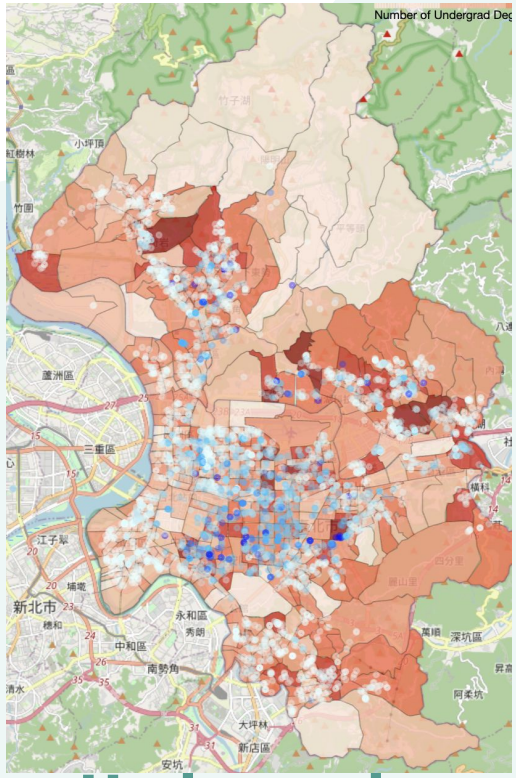
Education v.s. House Price



- Doctorate

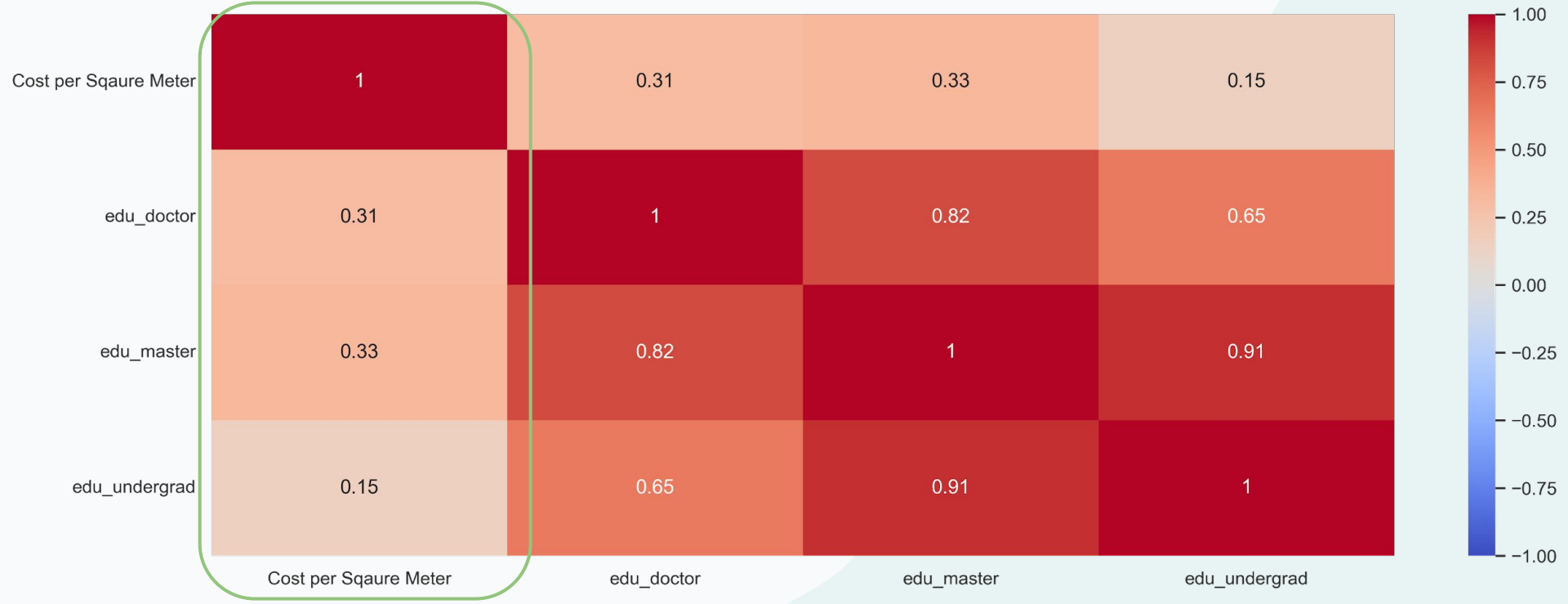


- Masters



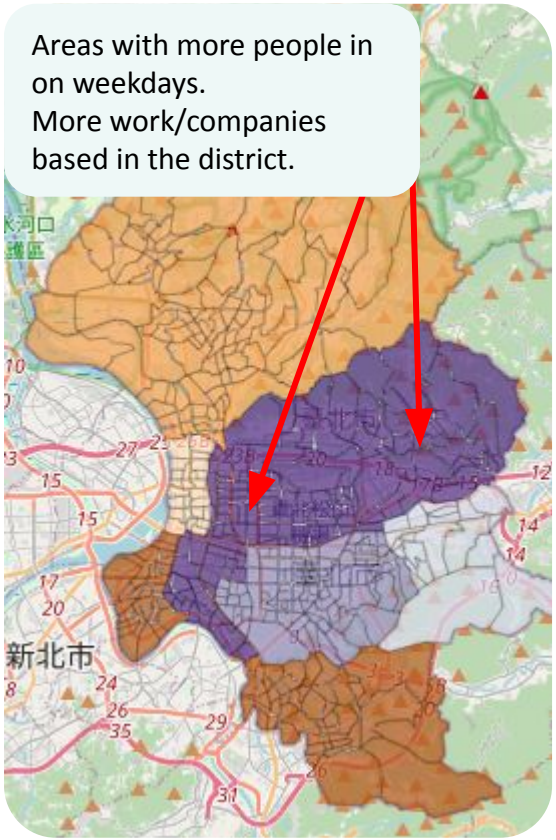
- Undergrad

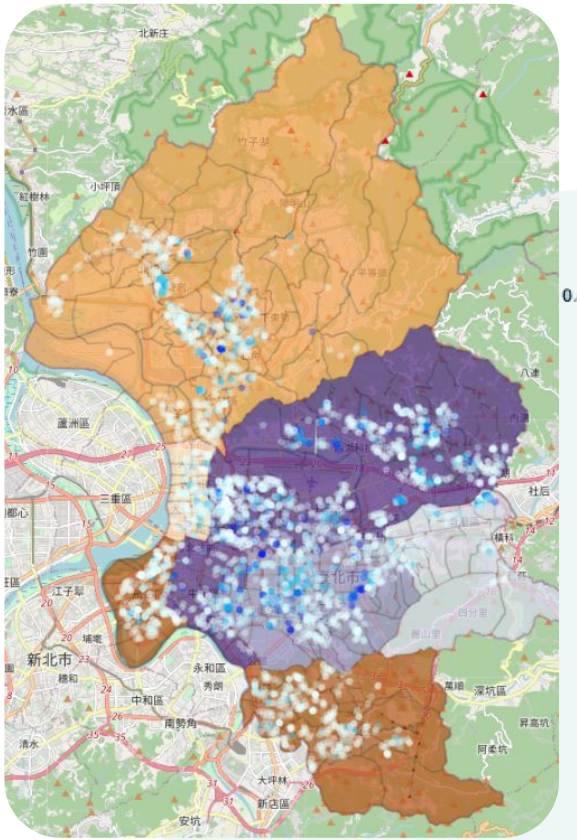
House Price v.s. Education Correlation

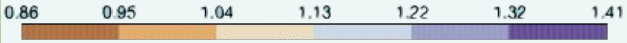


Activity Flow Day Time

Areas with more people in on weekdays.
More work/companies based in the district.







Weekday Daytime Count / Weekend Daytime Count

If ratio:

- > 1: Higher Weekday Count
- = 1: Small or No Differences
- < 1: Lower Weekday Count

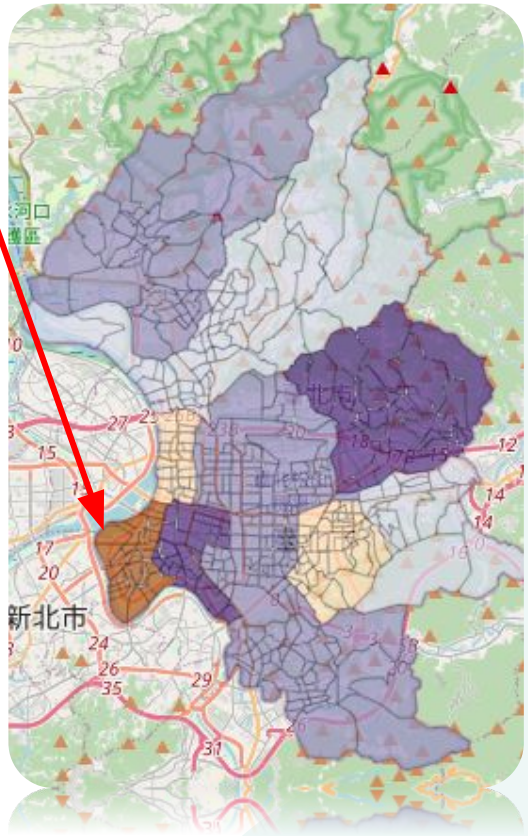
Therefore:

- >1 would represent more people working, and <1 would be more people living there or leisure zones.

Activity Flow Night Time

Differences are smaller, but areas with more weekend visits are identified.

Overall, the night time activity count has a smaller difference compared to day time.



0.987 0.994 1.001 1.008 1.016 1.023 1.030

Weekday Night time Count

Weekend Night time Count

If ratio:

- > 1: Higher Weekday Count
- = 1: Small or No Differences
- < 1: Lower Weekday Count

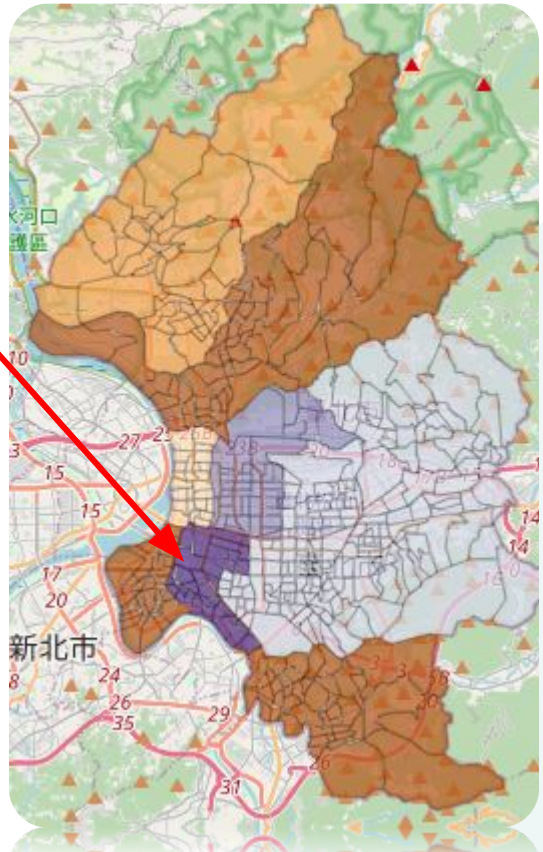
Therefore:

- >1 would represent more people working, and <1 would be more people living there or leisure zones.

Activity Flow Workday Day vs. Night

Overall, this shows the the city faces a large inflow of workers in the day time.

The more purple the district is would indicate more people work there and less people live there.



Weekday Day time Count

Weekday Night time Count

If ratio:

- > 1: Higher Weekday Count
- = 1: Small or No Differences
- < 1: Lower Weekday Count

Therefore:

- >1 would represent more people working, and <1 would be more people living there or leisure zones.