

Validating overbuilding of **shopping centers** in Shanghai and identifying which with investment potential

01 Data summary analysis

Chi Zhang
Candidate MUSA, 2022

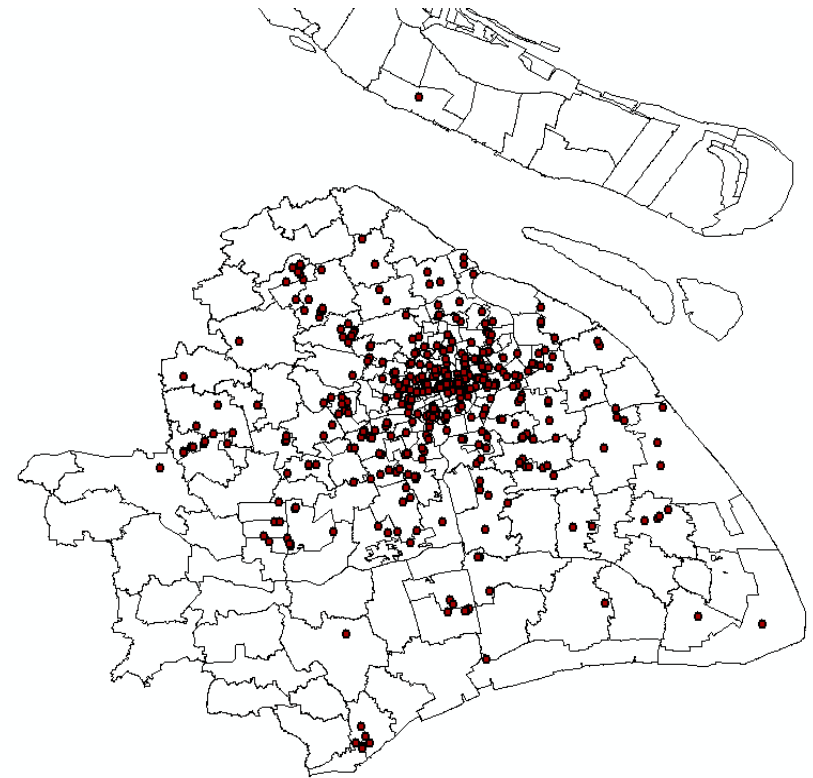
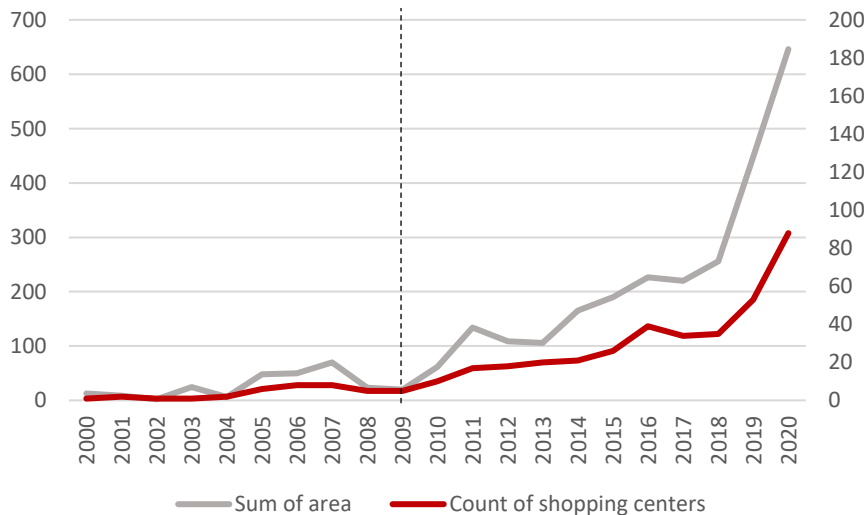


Background

Shopping centers area per capital in Shanghai is 3 times of the U.S.

This year the number of shopping centers opened or to be opened in Shanghai will be as high as 419. For 2020, The area of shopping centers per person is 0.97 sqm. Both count and area of malls are dramatically increasing since 2009. Considering the consumption capabilities, the supply area of shopping centers in Shanghai is almost 3 times of that in the United States. Overbuilt may cause severe energy waste and disturb the social economics.

Booming of shopping centers since 2009



419 malls opened or to be opened

Goals and values

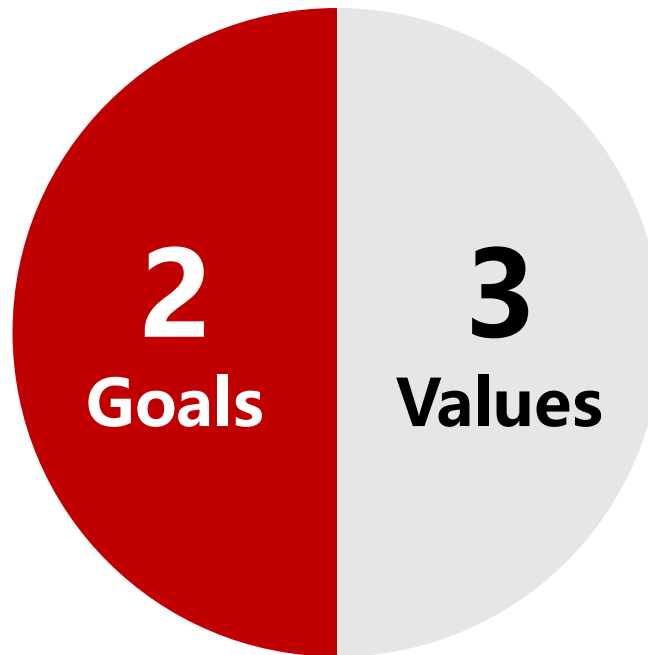
From the perspective of researchers, government and investors



- **Validate** if overbuilding exists



- **A tool** to identify surplus malls with investment potential



- **Researchers:** Fill a research gap



- **Government:** Land use planning and market interfering



- **Investors:** Acquire commercial assets with lower price

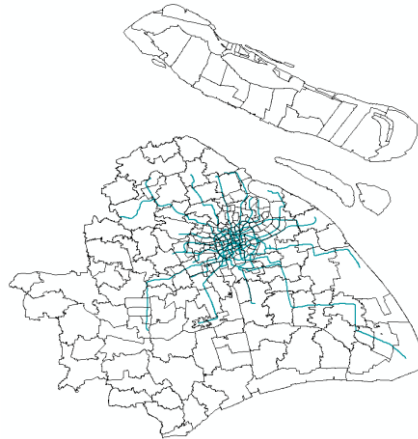
Data

Multivariate data resources are involved to build the model

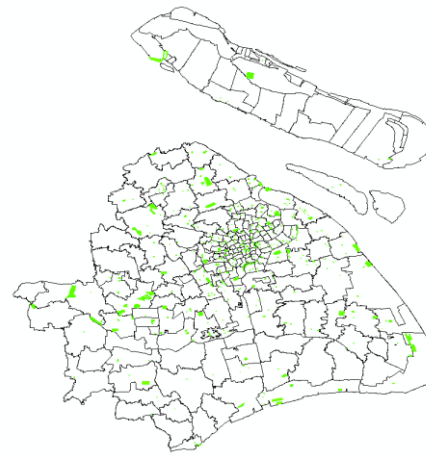
Categories	Data Sources	Columns	Type
Urban environment	Purchased from third party	road	geometry
		parks	geometry
		subway_station	geometry
		POIs	geometry
		building_outlines	geometry



Road network



Subway lines



Parks

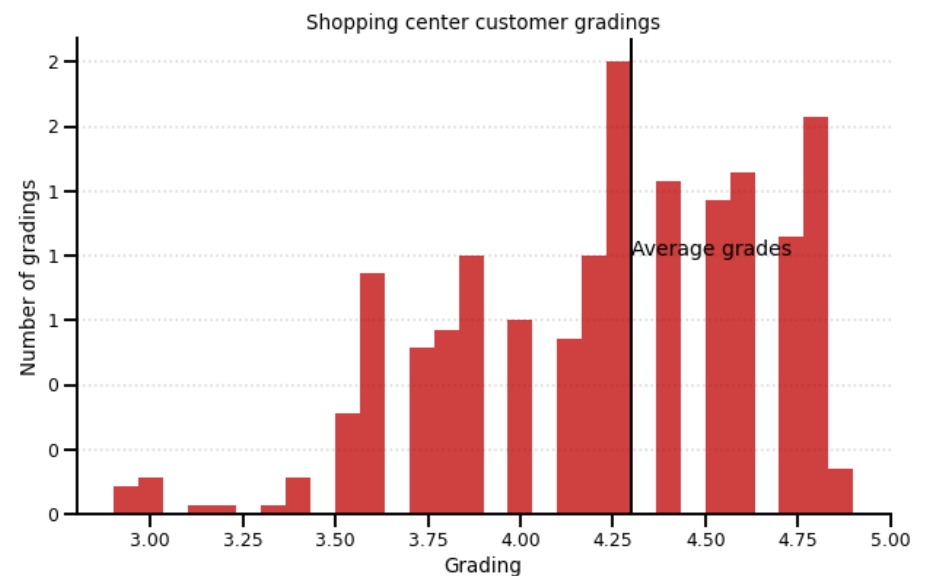
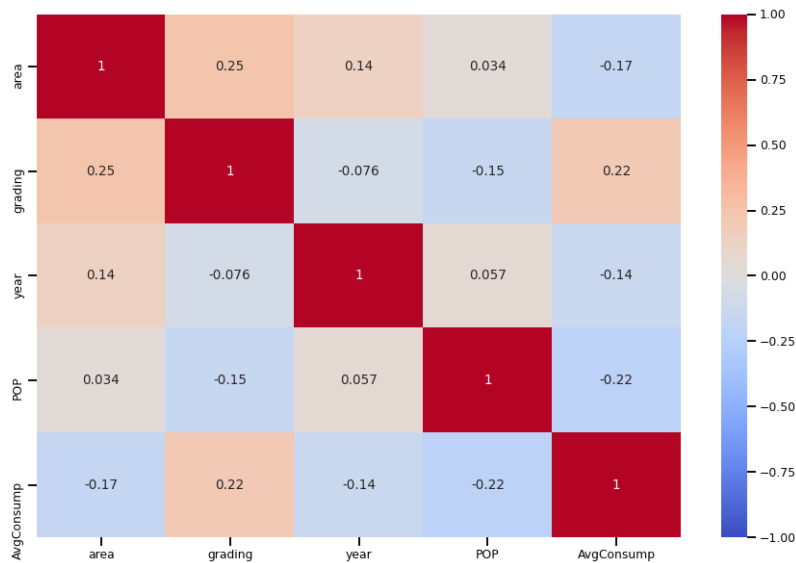


POIs

Data

Multivariate data resources are involved to build the model

Categories		Columns	Type
Shopping centers	Wrangled from Yingshang.com	building_area	float
		rentable_area	float
		built_year	integer
	Baidu Map API	coordinate	geometry
Socioeconomics	Chinese socioeconomical year book of 2020	population	integer
		disposable_income	integer
		consumption expense	integer
Social Media	Dianping.com	customer_gradings	integer



Data exploration

Use gravity model to analyze the customer coverage rates of each mall

Gravity model: a spatial model calculating the potential number of customers in each neighborhood

$$P_{ij} = \frac{W_i / D_{ij}^{\alpha}}{\sum_{i=1}^n \left(W_i / D_{ij}^{\alpha} \right)}$$

- P_{ij} = the probability of consumer j shopping at store i .
- W_i = a measure of the attractiveness of each store or site i .
- D_{ij} = the distance from consumer j to store or site i .

Gravity model results

	1	10	100	101	102	104	105	106	107	108	...	93	94	95	96	97
0	0.001741	0.000152	0.000580	0.001578	0.004740	0.003061	0.000196	0.001601	0.015114	0.006493	...	0.000623	0.005329	0.003351	0.002660	0.002303
1	0.001761	0.000157	0.000597	0.001542	0.004887	0.003114	0.000203	0.001729	0.010063	0.006131	...	0.000623	0.005482	0.003378	0.002697	0.002285
2	0.001002	0.000085	0.000298	0.002049	0.002515	0.001983	0.000095	0.000426	0.000705	0.029481	...	0.000518	0.002707	0.002420	0.001757	0.002317
3	0.001527	0.000129	0.000496	0.001464	0.004037	0.002662	0.000166	0.001272	0.026363	0.006388	...	0.000557	0.004555	0.002957	0.002327	0.002085
4	0.001128	0.000089	0.000367	0.001402	0.003102	0.002153	0.000125	0.000774	0.015585	0.007256	...	0.000483	0.003354	0.002479	0.001908	0.001869
5	0.001081	0.000088	0.000315	0.003505	0.002893	0.002159	0.000113	0.000421	0.000518	0.055122	...	0.000727	0.002863	0.003234	0.002280	0.002715
6	0.000562	0.000046	0.000192	0.000641	0.001636	0.001085	0.000067	0.000448	0.202322	0.002845	...	0.000237	0.001756	0.001244	0.000972	0.000874
7	0.000541	0.000039	0.000173	0.000755	0.001462	0.001055	0.000057	0.000314	0.001833	0.004872	...	0.000241	0.001579	0.001206	0.000913	0.000997
8	0.001935	0.000135	0.000574	0.002055	0.004850	0.003710	0.000179	0.000922	0.001823	0.022705	...	0.000773	0.005226	0.003800	0.002817	0.004014
9	0.001616	0.000146	0.000554	0.001371	0.004534	0.002858	0.000189	0.001669	0.004567	0.005319	...	0.000564	0.005094	0.003076	0.002467	0.002057

Note: Columns represents each neighborhood, row index is the ID of each shopping center

Data exploration

Use supply & demand model to find out surplus shopping malls

Supply & demand estimation model:

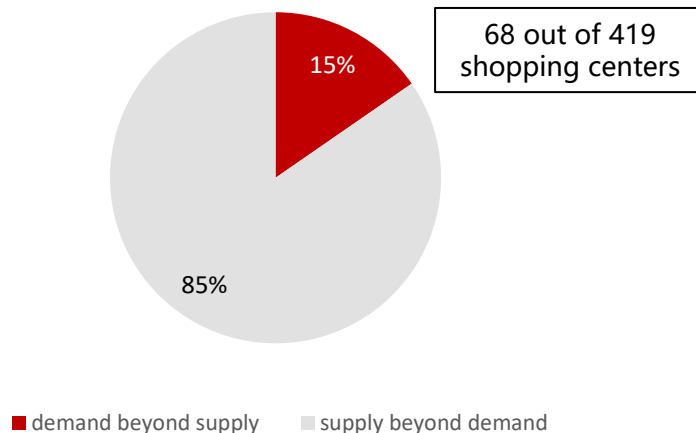
$$S = N * I * R$$

- N = count of customers in market area
- I = average consumption expense
- R = market penetration rate

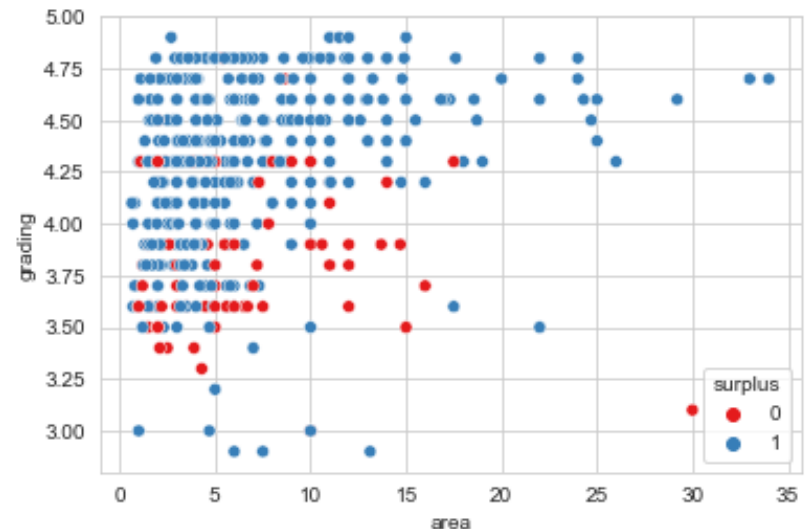
$$D_{\min} = A * S$$

- A = rentable area
- S = minimum sales income per sqm to pay off the building and operation fees within 40 years mortgage (assuming 10,000 RMB/sqm/year)

Surplus malls accounts for 15% area



Low graded and median-scale malls are identified

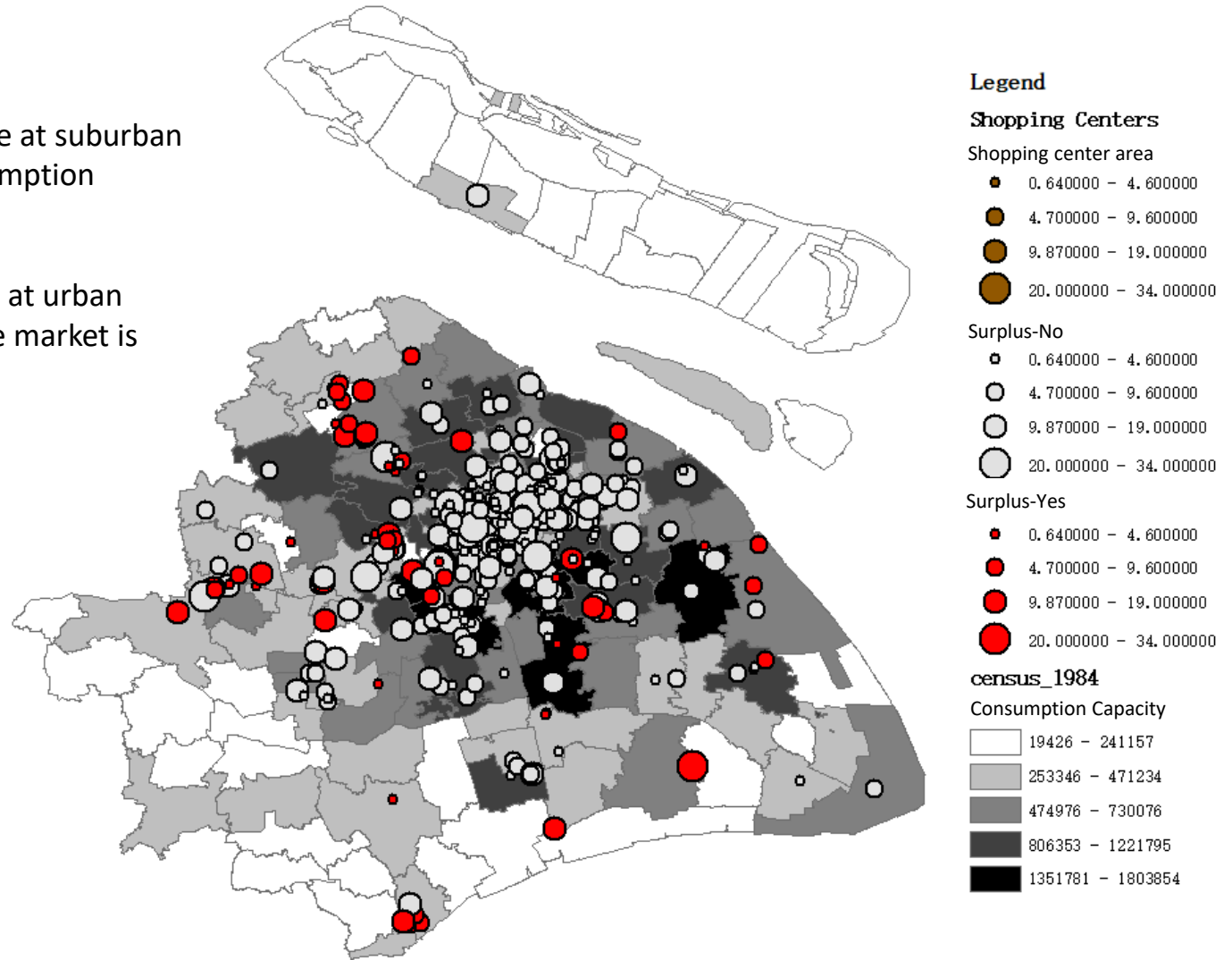


Data exploration

Surplus malls locate at suburban neighborhoods where consumption capacity is small

Most of surplus malls locate at suburban area where the total consumption capacity is small.

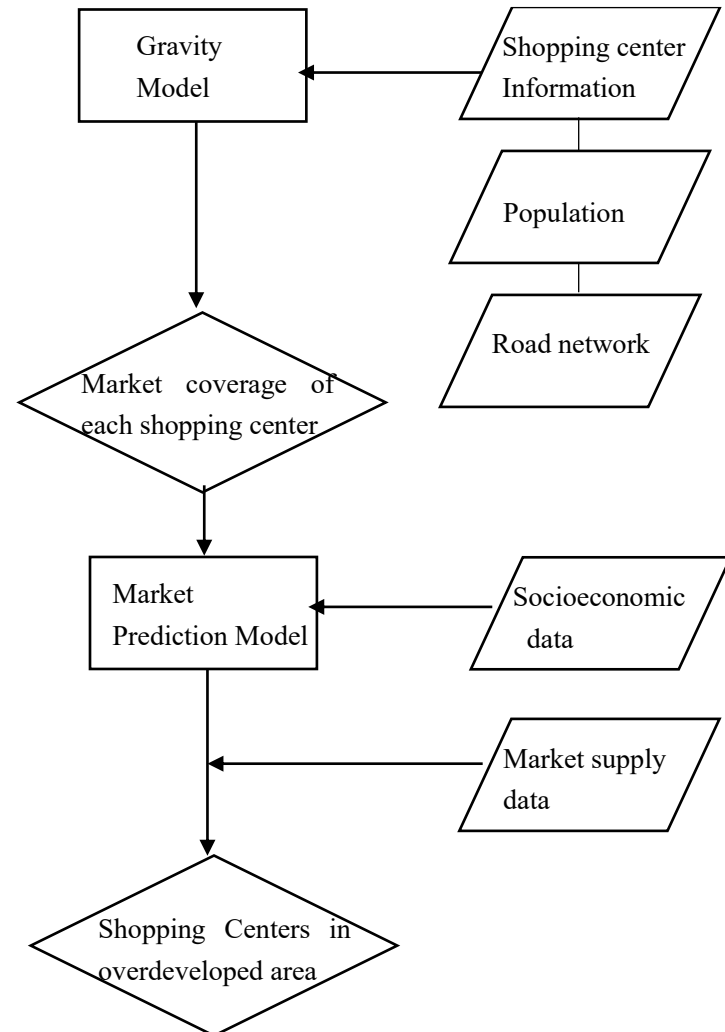
Some surplus malls located at urban center periphery where the market is competitive.



Open question

Model improvement and result validation

- **Gravity model improvement**
Straight line distance VS Driving distance
- **Model validation**
Significancy test



Next steps

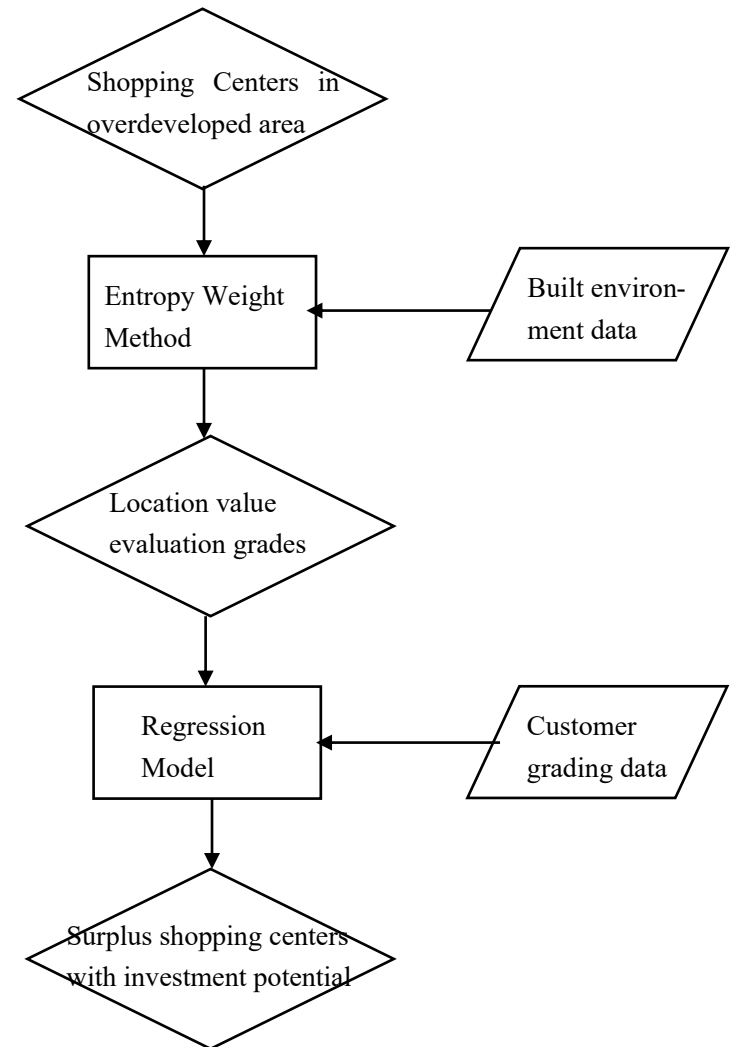
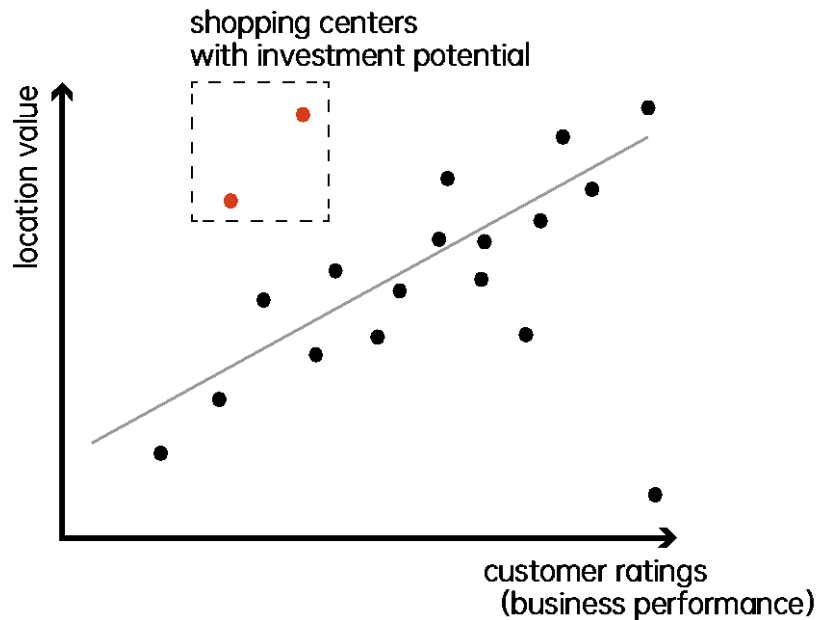
Identify shopping centers with investment value

- **Entropy weight method**

Quantify location value

- **Regression model**

Identify outliers with high location value but low customer grades



Thank you!

