Optimizing Location of Car-sharing Stations Based on Potential Travel Demand and Present Operation Characteristics: The Case of Chengdu

Yu CHENG
Xu CHEN, Xiaohua Ding, Linting Zeng
Shanghai Electric Vehicle Public Data Collecting, Monitoring and Research Center

EVData
SHANGHAI CHINA

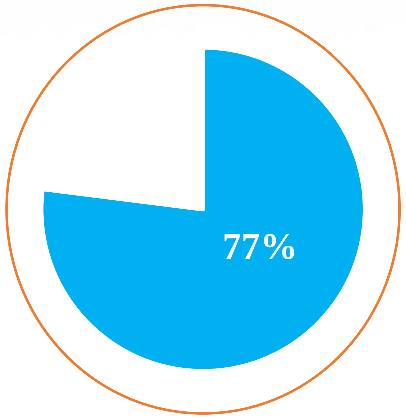
Overview of Shanghai NEV public data platform

- Register at the end of 2014
- The Data Center is a social service agency
- The first local monitoring platform for the promotion and application of new energy vehicles
- A local monitoring platform with the biggest data access of new energy vehicles up to now

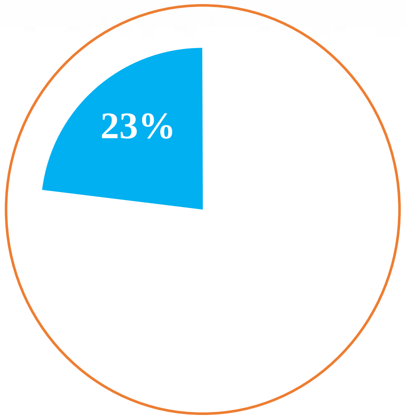
Overview of Shanghai NEV public data platform

Vehicle types and number of vehicles connected to Shanghai NEV Public Data Collecting, Monitoring and Research Center (Up to October 01, 2018)

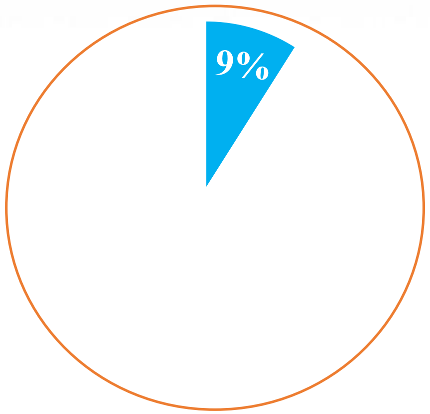
Vehicles connected to the center	Passenger car	Commercial vehicle	Total
Number of companies	58	52	110
Number of models	340	258	598
Number of vehicles	194,023	18,834	212,857



Proportion of PHEV passenger cars



Proportion of BEV passenger cars



Proportion of commercial vehicles

Overview of Shanghai NEV public data platform

Vehicle data items –static vehicle information (44 items of GB standards)

vehicle information

- SIM card number (ICCID)
- VIN number
- Vehicle model
- Drive motor layout/location
- Maximum vehicle speed
- All-electric range
- Gear ratio

Fuel part information

- Engine number
- Fuel type
- Fuel code
- Maximum power output
- Maximum torque

Information about rechargeable energy storage device and drive motor

- Rechargeable energy storage system code
- On-board energy storage device type
- Capacity of on-board energy storage device
- Cooling method for on-board energy storage device type
- Cooling method for drive motor
- Rated voltage
- Maximum current for working drive motor
- Drive motor serial number
- Drive motor model
- Peak power of drive motor
- Maximum drive motor speed
- Peak torque of drive motor
- Maximum output torque of drive motor

General alarm limit

- Temperature change alarm
- Battery high temperature alarm
- Overvoltage alarm for on-board energy storage device type
- Undervoltage alarm for on-board energy storage device type
- Low SOC alarm
- Overvoltage alarm for single cell battery
- Undervoltage alarm for single cell battery
- High SOC alarm
- SOC jump alarm
- Rechargeable energy storage system mismatch alarm
- Consistency bias alarm for single cell battery
- Insulation alarm
- DC-DC temperature alarm
- Braking system alarm
- DC-DC status alarm
- Drive motor controller temperature alarm
- HVIL alarm
- Drive motor temperature alarm
- Overcharge alarm for on-board energy storage device type

Overview of Shanghai NEV public data platform

Vehicle data items –real time vehicle information (80 items of GB standards)

vehicle data

- Vehicle status
- Charging status
- Operation mode
- Vehicle speed
- Accumulated mileage
- Total voltage
- Total current
- SOC
- DC-DC status
- Gear
- Insulation resistance (vehicle)

Alarm data

- Highest alarm level
- General alarm sign
- Total number of faults in rechargeable energy storage device N1
- List of rechargeable energy storage device fault codes
- Total number of faults in driver motor N2
- List of drive motor fault codes
- Total number of engine faults
- List of engine faults
- Total number of other faults N4
- List of other fault codes

Drive motor data

- Number of drive motors
- Information list of drive motor assembl
- Drive motor serial number
- Drive motor status
- Temperature of drive motor controller
- Drive motor speed
- Drive motor torque
- Drive motor temperature
- Input voltage of drive motor controller
- DC bus current of driver motor controll

Vehicle location data

- Location status
- Longitude
- Latitude

Fuel cell data

- Fuel cell voltage
- Fuel cell current
- Fuel consumption rate
- Number of temperature sensors for fuel cell
- Temperature value from the sensor
- Maximum temperature in the hydrogen system
- Code of sensor with maximum temperature value from the hydrogen system
- Maximum hydrogen concentration
- Maximum Hydrogen Pressure
- Code of sensor with maximum hydrogen concentration value
- Code of sensor with maximum hydrogen pressure value
- High voltage DC-DC status

Engine data

- Engine status
- Crankshaft speed
- Fuel consumption rate

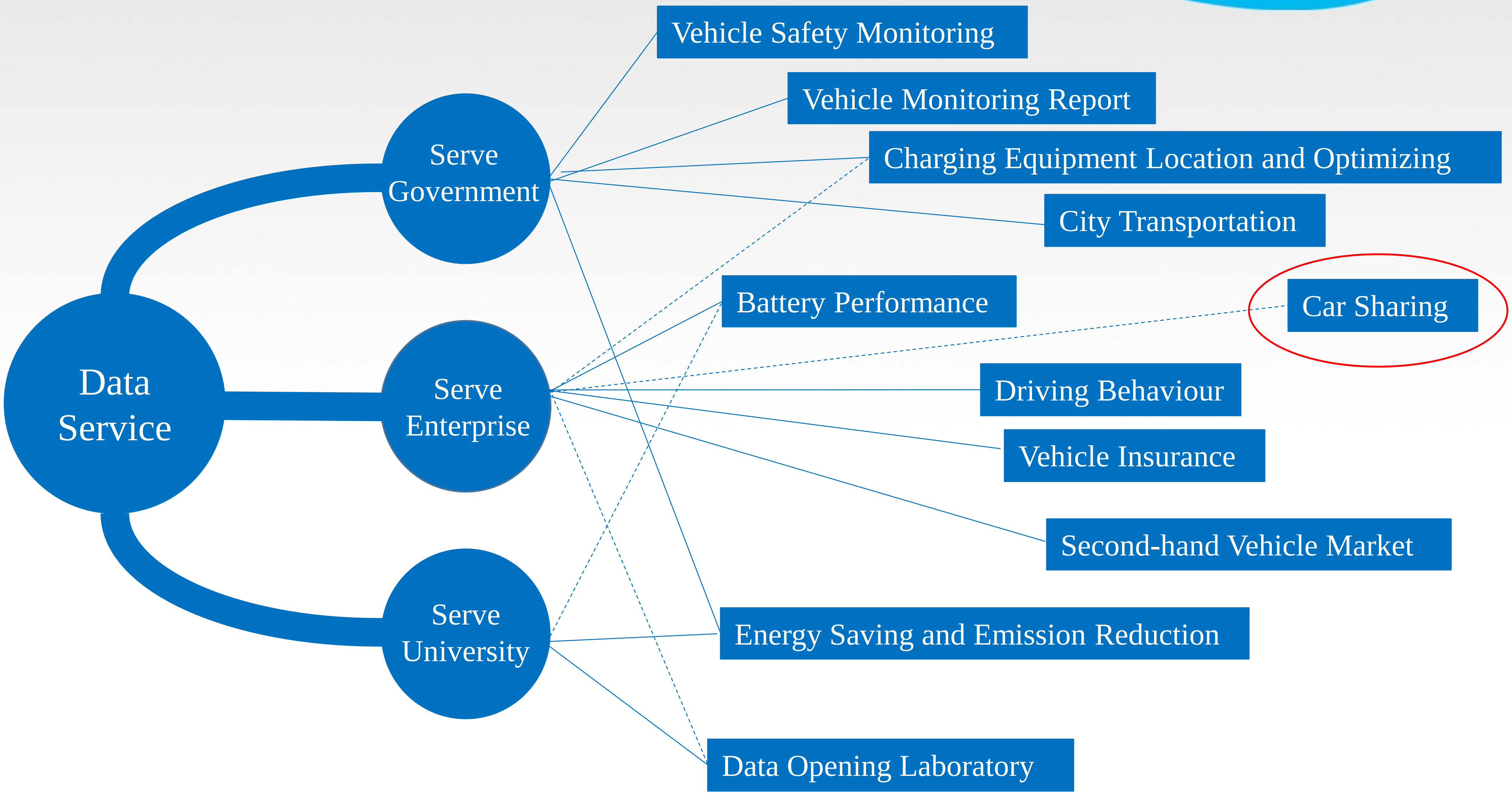
Extreme value data of power storage battery

- Number for battery subsystem with maximum voltage
- Code of single cell battery with maximum voltage
- Maximum voltage value for single cell battery
- Number for battery subsystem with mimum voltage
- Code of single cell battery with minimum voltage
- Minimum voltage value for single cell battery
- Number for battery subsystem with maximum temperature
- Serial number for sensor with maximum temperature value
- Maximum temperature value
- Number for battery subsystem with minimum temperature
- Serial number for sensor with minimum temperature value
- Minimum temperature value

General alarm sign

- Temperature change alarm
- Battery high temperature alarm
- Overvoltage alarm for on-board energy storage device type
- Undervoltage alarm for on-board energy storage device type
- Low SOC alarm
- Overvoltage alarm for single cell battery
- Undervoltage alarm for single cell battery
- High SOC alarm
- SOC jump alarm
- Rechargeable energy storage system mismatch alarm

Overview of Shanghai NEV public data platform

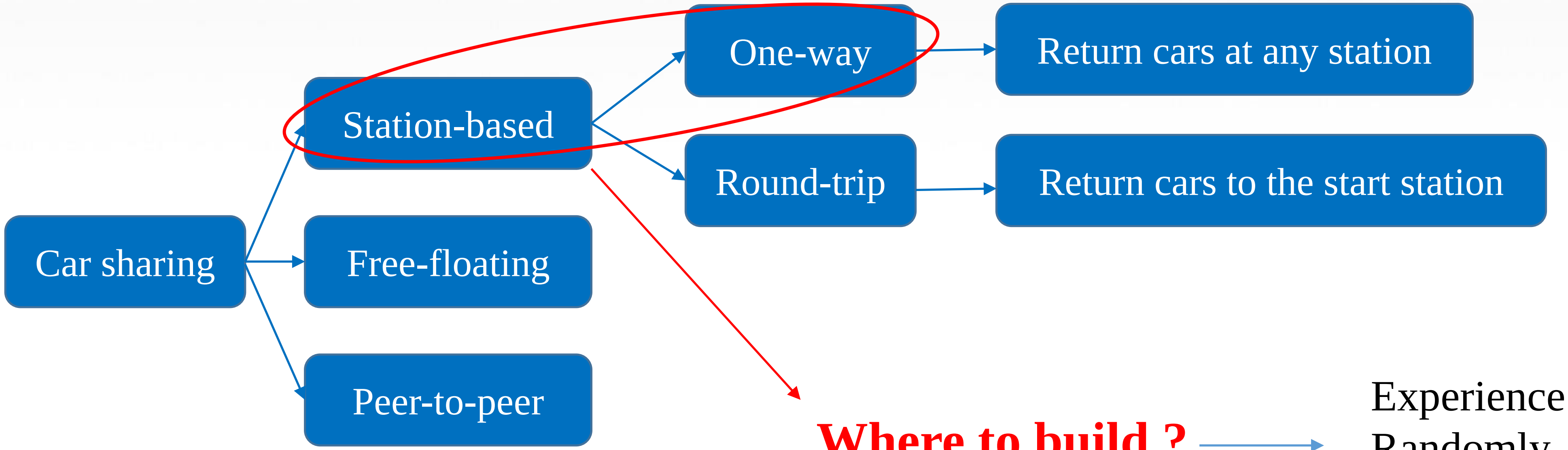


Optimizing Location of Car-sharing Stations Based on Potential Travel Demand and Present Operation Characteristics: The Case of Chengdu

- Introduction
- Data
- Methodology
- Results
- Conclusion

1. Introduction

- Access a fleet of vehicles for short-term use without ownership
- Reserve a vehicle online/by mobile app ⟶ move to parking lots ⟶ drive the car ⟶ pay the fee after travelling
- Increasingly important with the development of electric vehicles
- Reducing vehicle ownership, vehicle kilometers travelled and greenhouse gas emission



Where to build ? ⟶ Experience
Randomly
Mathematical Models

1. Introduction

Mathematical Models

- Mixed-integer Programming Model
- Multi-linear Model
- Multi-linear Model with LASSO
- Analytic Hierarchy Process (AHP)

Shanghai EVCARD

	Chengdu Market	
	Operator F	**Operator H**
Common	Station-based mode	
	Electric Vehicles	
Difference	Number of Electric Vehicles	
	Number of Stations	
	Vehicle Models	
	Charging Mode	

Potential Demand Heat
Existing Order Heat

2. Data

58724 squared grids
500m*500m
Allocate data into grids

1

Order Data:
Average Daily Order Numbers
from March 28 to April 17

2

POI Data:
AMAP
860,192 POIs
Fourteen classes

3

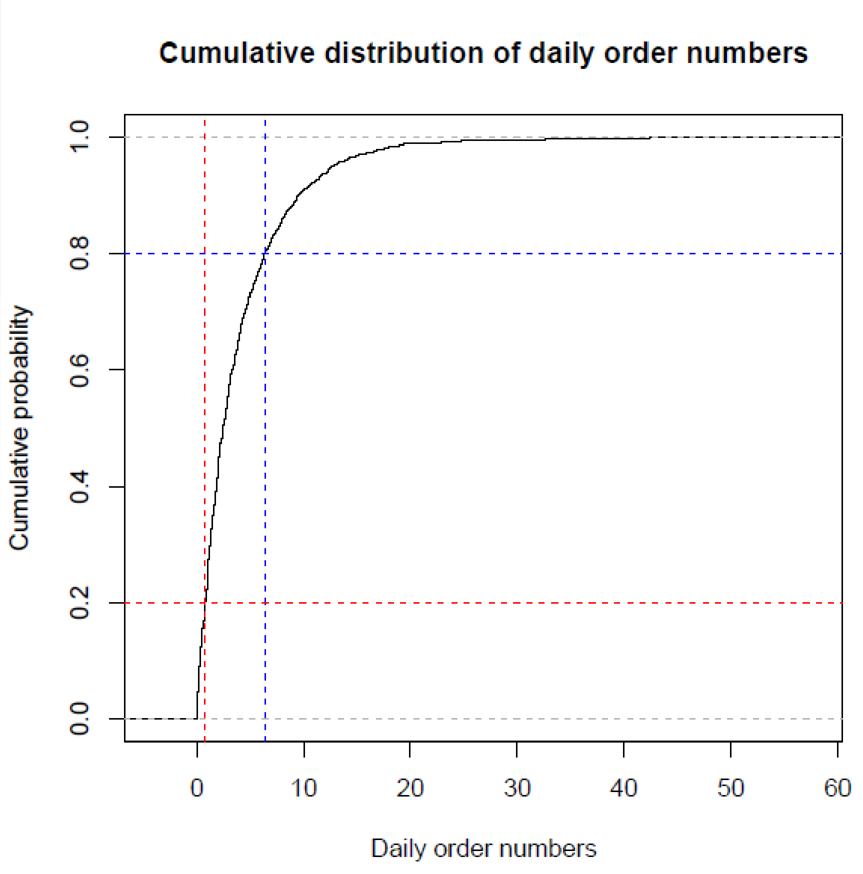
Population Data:
GPW v4 from NASA
2015 Population Estimation

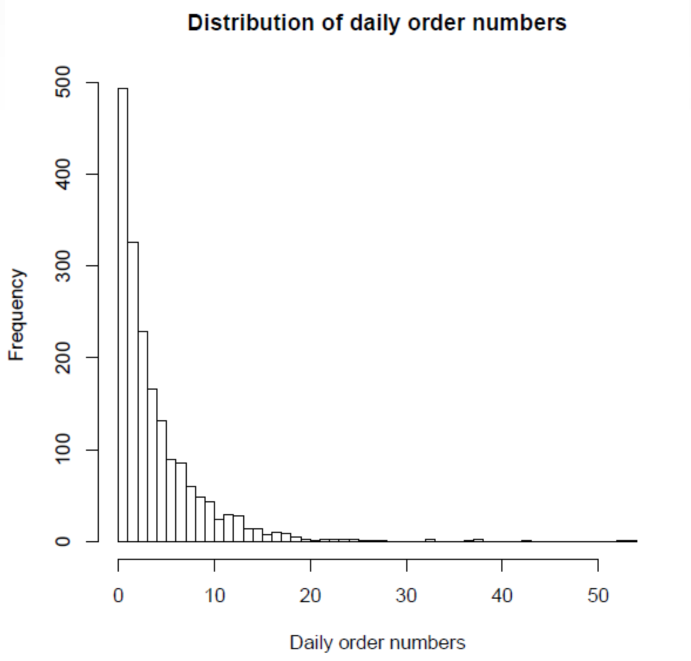
2. Data

Category		Abbreviation	Type	Mean	Std.Error	Max	Min
Auto Service	Car-rental Service	CRS	Numerical	0.0221	0.1836	7	0
Food & Beverages		FB	Numerical	2.4100	12.5445	373	0
Shopping		SH	Numerical	4.9470	26.9275	900	0
Daily Life Service		DS	Numerical	2.4800	11.7907	226	0
Sports & Recreation		SR	Numerical	0.3965	1.9526	78	0
Medical Service		MS	Numerical	0.6307	3.3228	148	0
Accommodation Service		AS	Numerical	0.3428	3.6490	251	0
Tourist Attraction		TA	Numerical	0.0609	0.4368	29	0
Commercial House		CH	Numerical	0.4638	2.4126	63	0
Governmental Organization & Social Group		GS	Numerical	0.4416	2.1701	78	0
Science/Culture & Education Service		SS	Numerical	0.5213	2.7930	117	0
Transportation Service	Bus Station	BS	Numerical	0.2033	0.6535	10	0
	Underground Station	US	Numerical	0.0053	0.0727	1	0
	Train Station	TS	Numerical	0.0013	0.0359	1	0
	Airport	AP	Numerical	0.0003	0.0184	1	0
	Parking Lots	PL	Numerical	0.6332	2.9611	81	0
Finance & Insurance Service		FS	Numerical	0.0874	0.5926	27	0
Enterprises		EN	Numerical	1.2750	7.1559	358	0

3. Methodology

- True value for the potential demand unknown
- Full sample questionnaire/present order numbers
- Change to binary question
- Whether demand of car-sharing exists or not
- Classification algorithm
- 1834 grids with at least one car-sharing station
- Lower 20% and upper 20% grids as the sample set
- Class ’demand =1’ for demand existing and
  Class ’demand =0’ for no demand





3. Methodology

- Logistic Regression

$$p(\mathbf{X}) = \Pr(Y = 1|\mathbf{X}) = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}{1+e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}$$

Log-likelihood function:

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{N}\{y_i \boldsymbol{\beta}^T \mathbf{X}_i - \log\left(1 + e^{\boldsymbol{\beta}^T \mathbf{X}_i}\right)\}$$

where $\boldsymbol{\beta} = \left(\beta_0, \beta_1, \ldots, \beta_p\right)$, $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{ip})$

- Logistic Regression with LASSO

A penalized term $\lambda \sum_{j=1}^{p} |\beta_j|$ is added

Log-likelihood function:

$$\max_{\beta}\{\sum_{i=1}^{N}\left[y_i \boldsymbol{\beta}^T \mathbf{X}_i - \log\left(1 + e^{\boldsymbol{\beta}^T \mathbf{X}_i}\right)\right] - \lambda \sum_{j=1}^{p} |\beta_j|\}$$

- Bayes' theorem:

$$\Pr(Y = k|\mathbf{X} = \mathbf{x}) = \frac{\pi_k f_k(\mathbf{x})}{\sum_{l=1}^{K} \pi_l f_l(\mathbf{x})},$$

where $f_k(\mathbf{x}) = \Pr(\mathbf{X} = \mathbf{x}|Y = k)$

- Naive Bayes

Assume each feature is independent given class k

$$\Pr(Y = k|\mathbf{X} = \mathbf{x}) = \frac{\pi_k f_k(\mathbf{x})}{\sum_{l=1}^{K} \pi_l f_l(\mathbf{x})} = \frac{\pi_k \prod_{j=1}^{p} f_{kj}(x_j)}{\sum_{l=1}^{K} \pi_l \prod_{j=1}^{p} f_{lj}(x_j)}$$

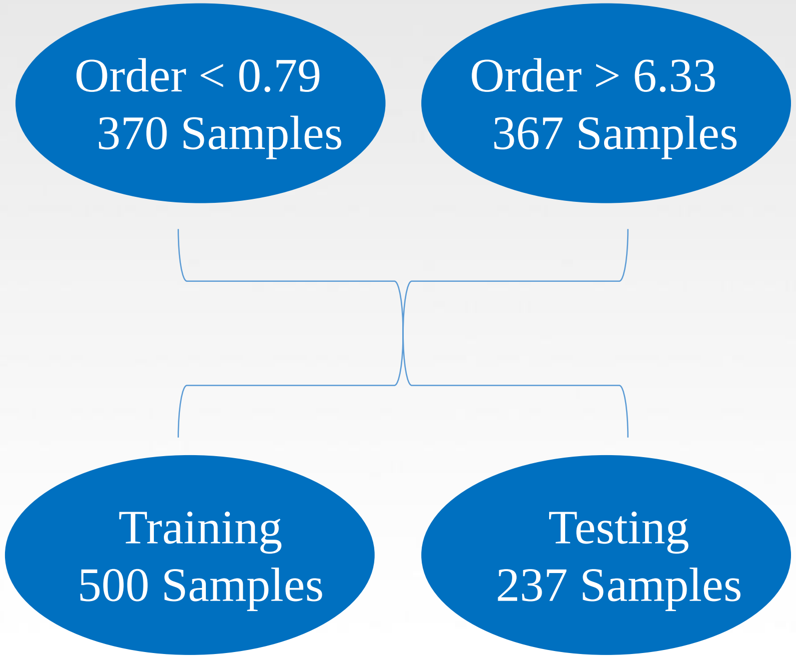
- Linear Discriminant Analysis (LDA)

$$f_k(\mathbf{x}) = \Pr(\mathbf{X} = \mathbf{x}|Y = k)$$
$$= \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(x - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1}(x - \boldsymbol{\mu}_k)\right)$$

- Quadratic Discriminant Analysis (QDA)

$$f_k(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}_k|^{1/2}} \exp\left(-\frac{1}{2}(x - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(x - \boldsymbol{\mu}_k)\right)$$

4. Results



	CRS	FB	SH	DS	SR	MS	AS	TA	CH	GS
Mean	0.3	32.35	57.49	29.31	4.93	6.32	5.83	0.37	5.12	4.14
Std.error	0.67	43.51	90.96	34.67	6.83	8.32	14.78	1.24	7.51	6.77
	SS	BS	US	TS	AP	PL	FS	EN	POP	
Mean	7.34	1.51	0.072	0.008	0.004	8.89	1.23	16.24	1882	
Std.error	11.35	1.4	0.26	0.09	0.06	10.09	2.23	33.81	5621.53	

4. Results

- The results of Logistic Regression

	Coefficient	Std.error	P value
Intercept	0.2569	0.1211	0.0338
FB (Food & Beverages)	0.9960	0.2404	<0.001
MS (Medical Service)	-0.3401	0.1581	0.0315
GS (Governmental Organization & Social Group)	-0.5480	0.1383	<0.001
BS (Bus Station)	0.2814	0.1307	0.0313
PL (Parking Lots)	1.3981	0.2182	<0.001

- The results of Logistic Regression with LASSO ($\lambda = 0.01973$)

	Intercept	FB (Food & Beverage)	SR (Sports & Recreation)	GS (Governmental Organization & Social Group)	BS (Bus Station)	TS (Train Station)	AP (Airport)	PL (Parking Lots)
Coefficient	0.1318	0.4652	0.1587	-0.2717	0.1641	0.0172	0.0336	1.0114

4. Results

- LDA and QDA: mean vectors for classes: demand=0 and demand=1

	Class 0	Class 1		Class 0	Class 1
CRS	-0.2109	0.1552	SS	-0.3451	0.372
FB	-0.4108	0.3963	BS	-0.2901	0.2864
SH	-0.285	0.3203	US	-0.0778	0.0947
DS	-0.3978	0.4089	TS	-0.0905	0.0435
SR	-0.3692	0.3401	AP	-0.0014	0.0622
MS	-0.2867	0.2754	PL	-0.4602	0.5125
AS	-0.2561	0.2615	FS	-0.2755	0.3065
TA	-0.1173	0.039	EN	-0.2301	0.3012
CH	-0.3205	0.3431	POP	-0.1	0.1076
GS	-0.0866	0.1573			

- Naive Bayes: Mean and standard deviation for the predictors grouped by class

	Class 0		Class 1	
Variable	Mean	Std.error	Mean	Std.error
CRS	-0.2109	0.7259	0.1552	1.1169
FB	-0.4108	0.6177	0.3963	1.1262
SH	-0.2850	0.6372	0.3203	1.2895
DS	-0.3978	0.6668	0.4089	1.1068
SR	-0.3692	0.5560	0.3401	1.0806
MS	-0.2867	0.8778	0.2754	1.0337
AS	-0.2561	0.3561	0.2615	1.3175
TA	-0.1173	0.5521	0.0390	0.6665
CH	-0.3205	0.7296	0.3431	1.1695
GS	-0.0866	0.9128	0.1573	1.1696
SS	-0.3451	0.4995	0.3720	1.2910
BS	-0.2901	0.8360	0.2864	1.0089
US	-0.0778	0.8589	0.0947	1.1439
TS	-0.0905	0.0000	0.0435	1.2157
AP	-0.0014	0.9907	0.0622	1.4038
PL	-0.4602	0.6131	0.5125	1.1715
FS	-0.2754	0.6450	0.3065	1.2305
EN	-0.2301	0.5698	0.3012	1.3990
POP	-0.1000	0.7764	0.1076	1.1253

- The priors are $\pi_0 = 0.502$ and $\pi_1 = 0.498$

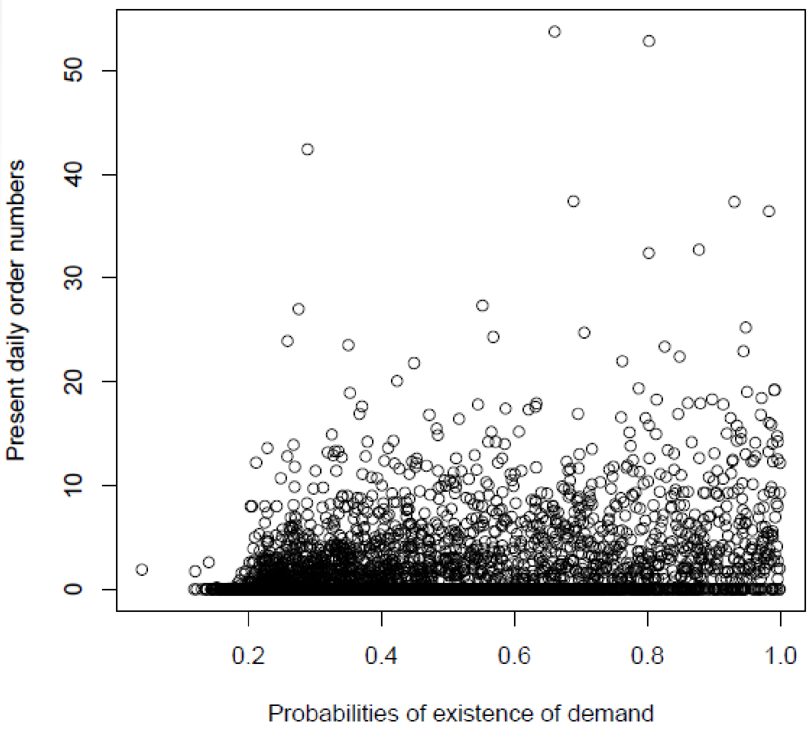
4. Results

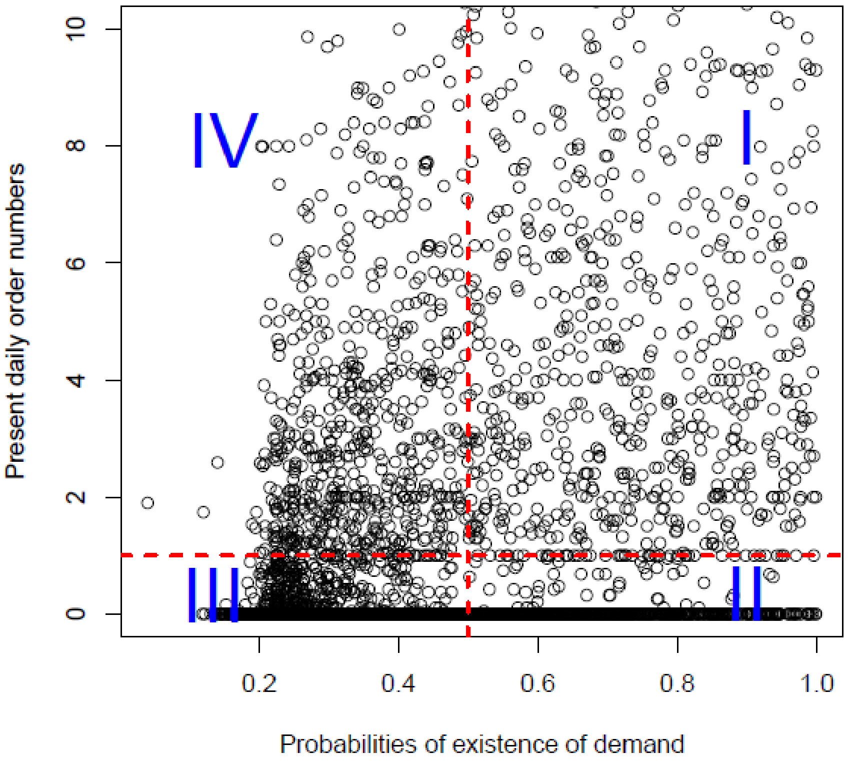
- Comparing these five models
  AUC: Area under receiver operating characteristic (ROC) curve
  Accuracy Rate: number of correct predictions over total number of predictions

	AUC value	Accuracy Rate
Logistic regression	0.8500	0.7722
Logistic regression with LASSO	0.8545	0.7637
LDA	0.8513	0.7553
QDA	0.8020	0.6835
Naive Bayes	0.8146	0.7215

4. Results

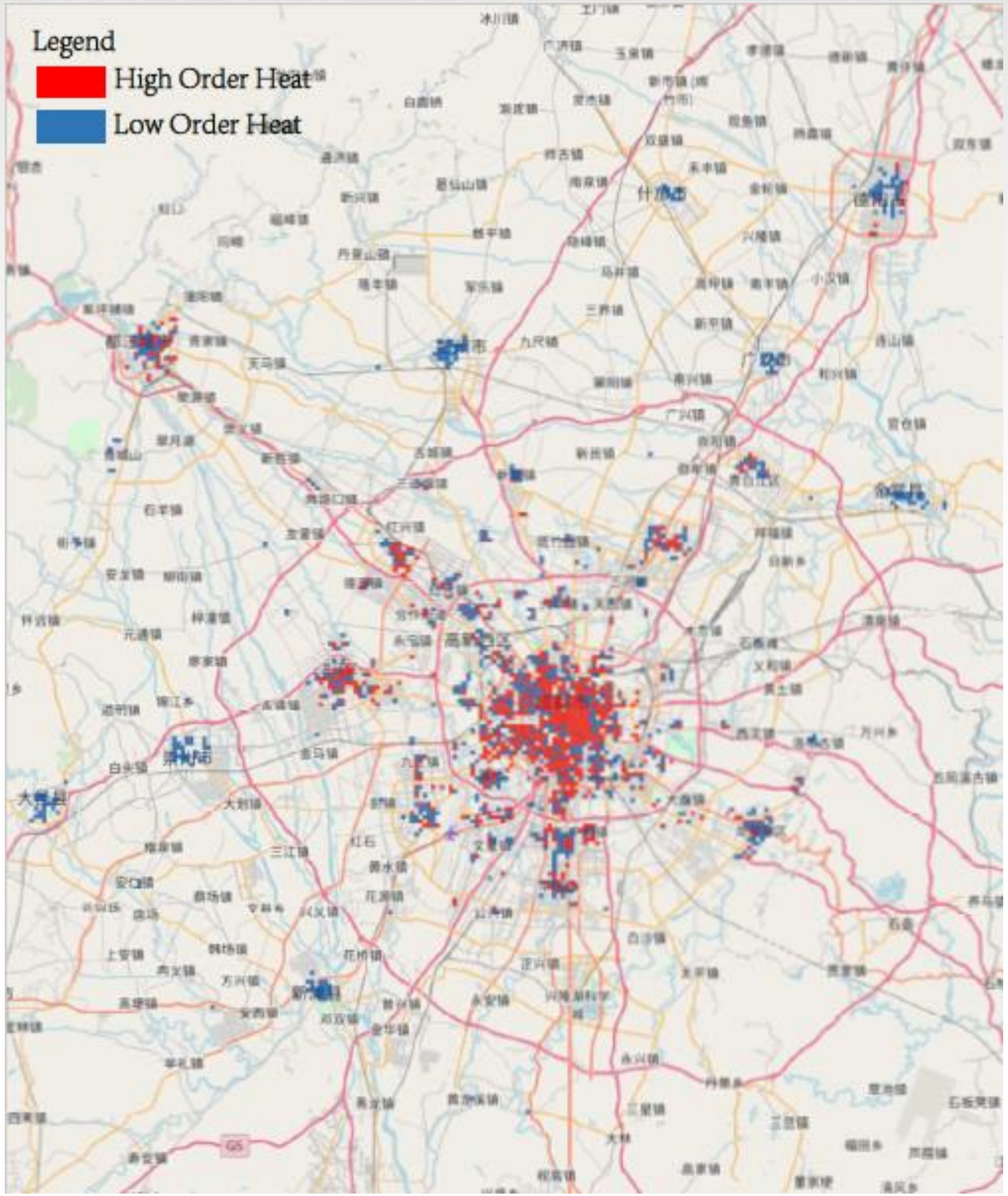
- Logistic Regression with LASSO





		Demand Heat	
		Low	High
Order Heat	Low	III	II
	High	IV	I

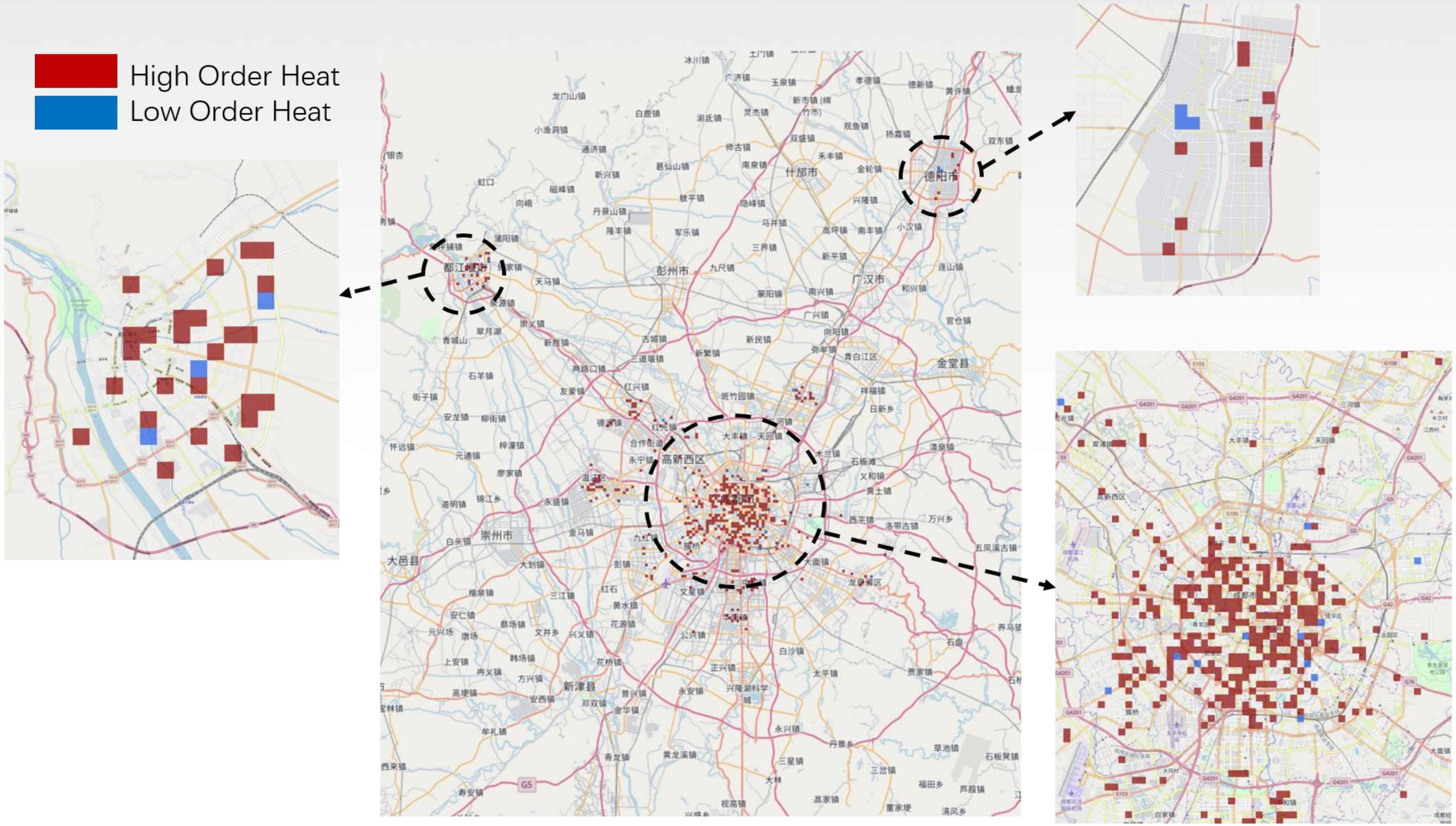
4. Results



- Case 1: Grids with no operator F stations and at least one operator H stations
- Case 2: Grids with no operator H stations and at least one operator F stations
- Case 3: Grids with no operator H stations and no operator F stations

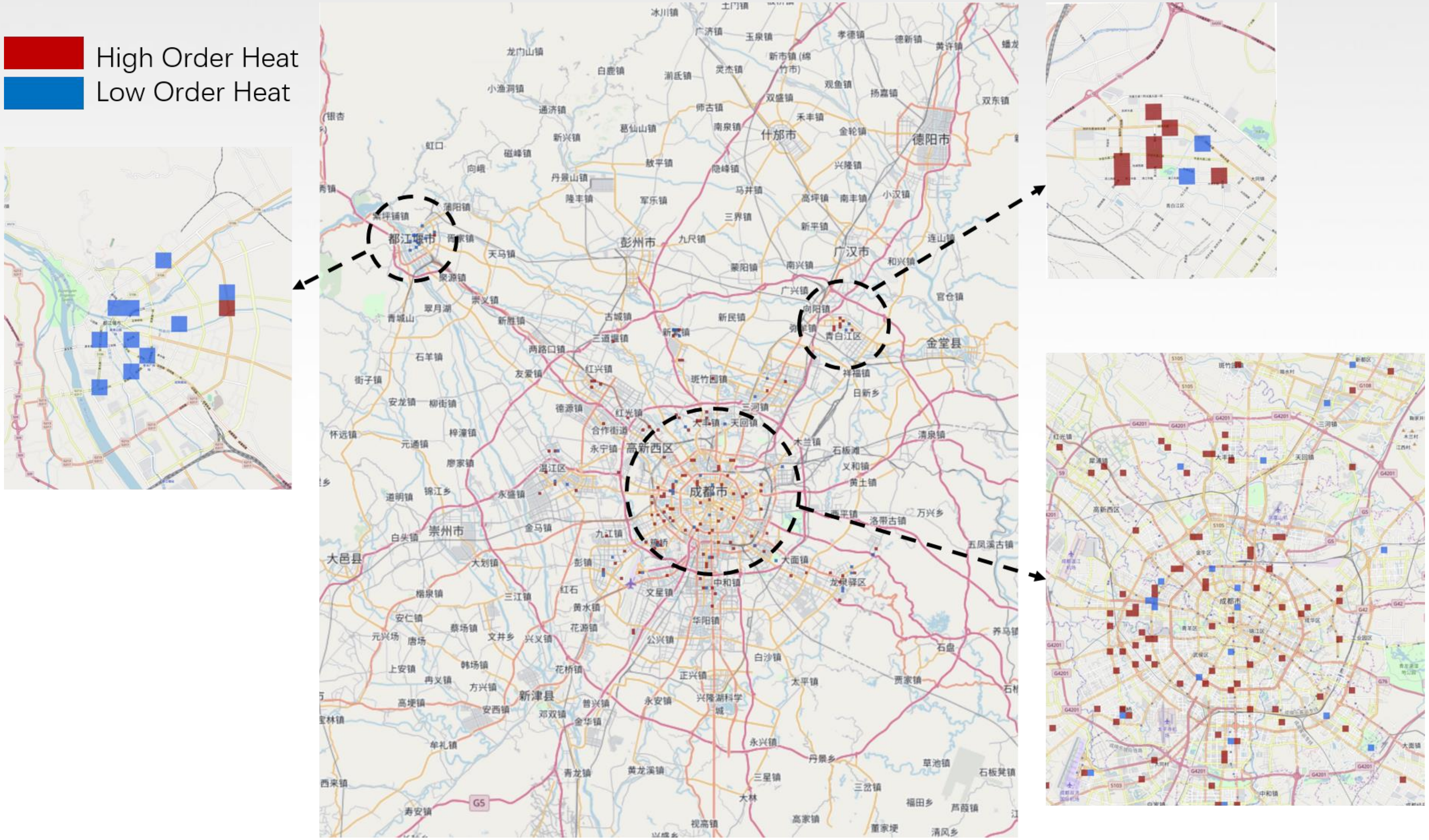
4. Results

- Case 1: Grids with no operator F stations and at least one operator H stations



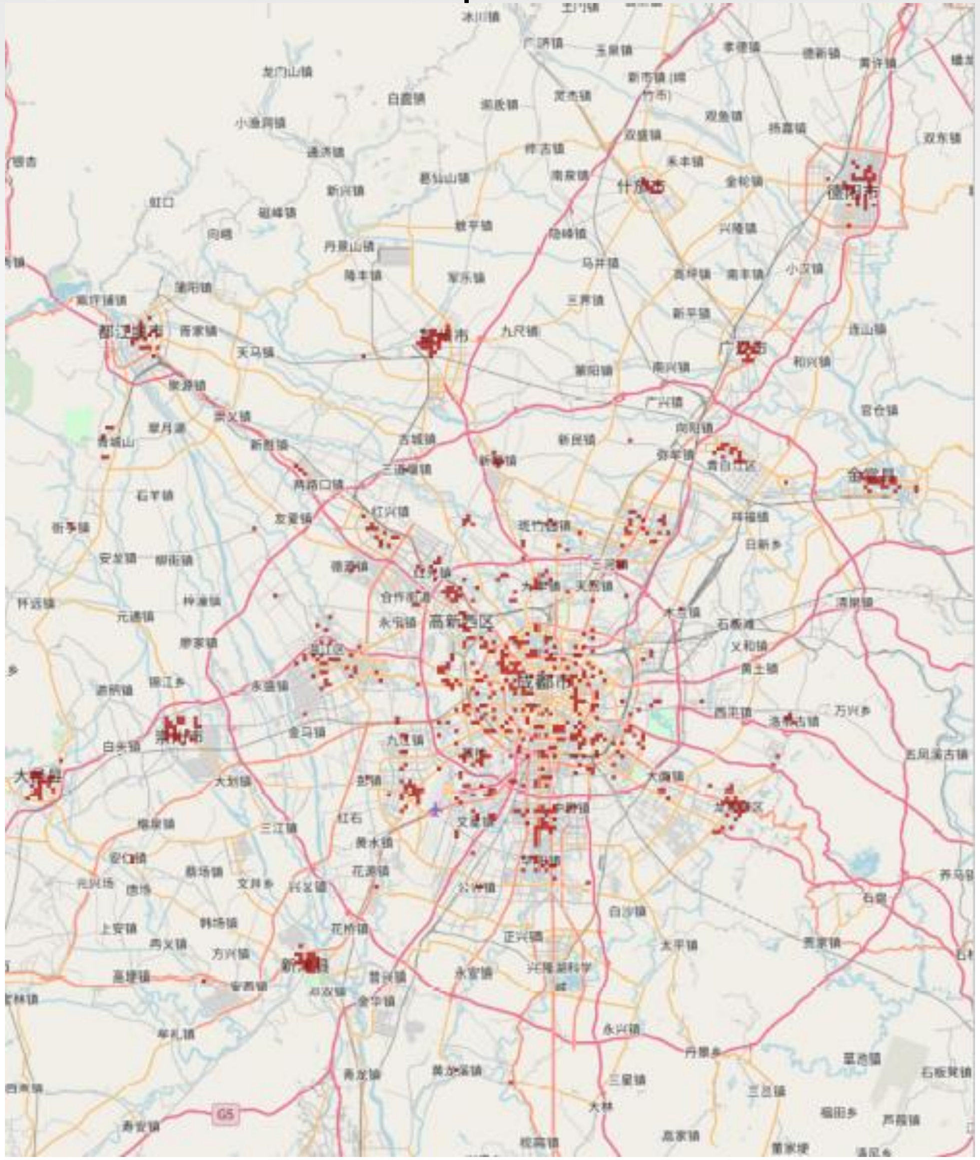
4. Results

- Case 2: Grids with no operator H stations and at least one operator F stations



4. Results

- Case 3: Grids with no operator H stations and no operator F stations



5. Conclusion

- Logistic Regression
  Positive effect: Food& Beverages, Bus Station and Parking Lots
  Negative effect: Medical Service and Governmental Organization & Social Group
- Logistic Regression with LASSO
  Positive effect: Food& Beverages, Sports& Recreation, Bus Station, Train Station, Airport and Parking Lots
  Negative effect: Governmental Organization & Social Group
- Logistic Regression and Logistic Regression with LASSO perform best,
  QDA and Naive Bayes perform worst
  Linear models work better in our case

- Grids with high demand heat are concentrated in the city center or town center
- Both operators are advised to build stations in the grid with demand heat
- It is suggested that both operators should remove stations in the grid with low order heat and demand heat
- Operator H is advised to build stations in the north-west of Chengdu

5. Conclusion

Limitations:

- Only two operators are considered
- 500m*500m grids are not suitable for low building density area
- The cost of building stations is not considered that the definition of order heat is based on
- Subjectively choose samples
- Models are based on strong assumptions
- Only consider the order that starts and omit the return behaviour
- Variable selection work can be done before train the model

T H A N K S

Oct. 2018

E mail: chengyu@shevdc.org
Official Web: www.shevdc.org