Dynamic evaluation model of urban public transport service level

Abstract

The paper solves dynamic evaluation problem of urban public transport service level by adopting big data treatment, clustering integration, computer simulation analysis, comprehensive energy efficiency evaluation and other methods.

Regarding problem one, we use three methods to initially pre-treat data, in order to make it helpful for analysis and solution of next problems. The first method is to remove interfering noise by binning and clustering and other method, remove irrelevant attribute values to achieve the effect of initial cleaning, use Apriori arithmetic to get one-dimension Boolean association rules frequent item-set, make comprehensive statistics huge GPS data amount, use the method of fast generating large item-set to treat data. The second method is to consider damage to original data in clustering method, remove some invalid items and defect items, After initially clean data, grey relational degree clustering analysis is used. It judges influence of identified abject on research object and make specific analysis of grey relational degree to treat data by comparing various relevancy. The third method is to remove some same, defect and invalid items, whose data cleaning process and methods are similar to previous two methods, considering limitation of k-average value on data type, it uses k-central value arithmetic, after data treatment, it uses relevancy report to get the influence of data pre-treatment on original data, in order to treat data. Regarding problem two, we use pre-treated data in problem one, make supplementary integration of bus GPS data and bus IC card information through proper data integration principle, get basic information of citizens' bus transport through comprehensive analysis. Based on citizens' bus transport two analysis points model and Erlung-K distribution curve, it gets single-line bus OD matrix. It divides the whole city into 647 bus communities according to corresponding principle and standard of integrating bus stop groups, adjusts and calculates corresponding contents in citizens' basic transport information table to get bus OD matrix. Based on bus OD matrix, it analyzes transport time distribution characteristics, transport space distribution and characteristics of matrix contents. Through analysis, we can obviously see peak of bus transport, the tide phenomenon of urban transport in the morning and evening is relatively obvious; meanwhile, we also find transport space distribution has close relation with urban land use layout, population density and bus system distribution. Urban central area accumulates numerous employment opportunities, various public service and entertainment equipments, and becomes main destination of citizen's transport.

Regarding problem three, we use systematic efficiency energy value to measure the abstract concept of public transport system service quality. Systematic efficiency energy value is composed of two parts, one is evaluation system of public transport service quality, which is divide into six different aspects, namely safety, convenience, economy, fastness, accuracy and comfort, it gets systematic service quality through evaluation of these six dimensions; the other is systematic reliability, which is divided into four aspects, namely service time reliability, transport time reliability, accurate point time reliability and transport amount time reliability. Finally we find East Door bus station, Meilong over-bridge bus station and Shiyan Jinbei village bus station are most improper through comprehensive evaluation. Roads are narrow and passenger flow volume is huge in East Door bus station, which causes traffic jam and influences its systematic efficiency energy value. Passenger flow volume is huge and bus operation distance is long in Meilong over-bridge bus station, which influences systematic efficiency energy value. Bus point distance is extremely large, distance to residential area is long and bus number is less , which can influence systematic efficiency energy value of Shiyan Jinbei village bus station.

In conclusion, we make initial cleaning of data through attribute and noise treatment, use clustering, K-central point, Apriori and ADC arithmetic to solve dynamic evaluation of urban public transport service level, combined with MATLAB, SQL Server, ArcGis and other software. In the end of the paper, we have expectations and exploration, which have big use value in our practical life.

Key words: data mining, OD matrix, clustering, system energy efficiency, Apriori

1. Problems restatement

The service evaluation of urban public transit is an important part of the construction of urban public transit system and public transport operation efficiency improvement. For transit enterprises, the management and planning department, the data for the planning of traditional bus stops, lines and Transfer Hubs are simply based on the statistics collected by competent authorities and manual inventory.

Increasingly developed in Auto-collecting Technology today, if we could automatically analyze the Resident Travel Demand using the Bus GPS data, Bus Card consuming data, metro card consuming data and taxi GPS data, and evaluate dynamically about the existing transit planning facilities service (including regular bus stops and metro stations), then we may improve the work efficiency and quality of traditional transit planning, design and management dramatically.

The data of urban bus, bus card consuming, metro card consuming and taxi GPS are included in the attachment. Please analyze these data carefully and answer the following questions.

1. Use at least three methods to preform preprocessing of the data, and compare the processing effect of data.

2. Analyze the OD matrix of residents travel data based on GPS and bus card consuming.

3. Please build mathematical model to make dynamic evaluation on the service of bus stops, metro stations, and reprioritize the most unreasonable three bus stops.

2. Issues analysis

2.1 Analysis of problem one

Problem one asks us to adopt at least three methods to carry on the preliminary processing to the data, and analyze the influence to the data. After analysis, we get the process of dealing with large data. First of all, cleaning and integration of important data, then transform and reduction. Finally, using Server SQL tool to generate the processing results to To analyze the impact of processing results on data.

First methods: In order to eliminate a large number of "dirty data" in the mass of data"，we first introduce the binning principle and the method of clustering，then give weight to time variable. Finally, the key effective information of urban public transport system service quality is obtained. Meanwhile，remove some invalid attribute values and streamlining the data size. Then, the Apriori algorithm is used to obtain the one dimensional Boolean association rules. Considering the diversity of the factors affecting the service quality of public transportation system, we use multiple iterations to generate multiple candidate item sets, then draw a dependency graph and select different nodes to evaluate the accuracy of the results.

Second methods：Taking into account the damage to the original data in the clustering method, we only remove some invalid entries and the defect, and then carry out regression analysis on the data. Because data are given a number of different indicators, and therefore the use of multiple linear regression approach to data clustering, classification and processing.

Third methods: The process of data cleaning is similar to that of the above mentioned method,

but some duplication, defect and invalid items are removed, then considering the k - average algorithm to the limitations of the data type, so we use the k - midpoint algorithm. After the data processing, call the dissimilarity of the report, get influence of the data preprocessing to the original data.

2.2 Analysis of problem two

Problem two asks us to analyze the OD matrix of residents travel data based on GPS and bus card consuming. After analysis, we found that the useful attribute values of the GPS information and the integration of the IC card data are limited. So we choose to remove the relevant invalid basic information, so as to form a more comprehensive, practical basis data. By getting under the station, passengers get on and off time of these three kinds of information from the database, to complete the research on the start and end points of the transit trip. We can directly extract information of passengers' get on time and the station from the database. In order to obtain the get-off-station and the get-off-time, we introduce the "bus travel analysis two site model", combining with the GIS bus electronic maps to get the full information about the end station and starting station. Two site models do not take into account most of the public transport network is the existence of transformation on site location, so the introduction of el Aaron - K curve is modified. After the above analysis, fuse the "single line bus passenger OD" information and auxiliary parameters to get the city bus OD matrix Then we according to "site corresponding principle" will be merged into the site, and according to this standard will be divided into a number of public transport area. Then adjust the public transport travel information based on the table to get the city bus OD matrix. Based on the bus OD matrix, the matrix's content and features is analyzed, and the corresponding conclusions are drawn.

2.3 Analysis of problem three

Problem three asks us to build mathematical model to make dynamic evaluation on the service of bus stops, metro stations, and reprioritize the most unreasonable three bus stops. After analysis, we think that the abstract concept of service quality should be transformed into the system energy efficiency value to measure. There are two parts in the system energy efficiency: One is the public transport service quality evaluation system, which is divided into six different aspects: safety, convenience, and economy, rapid, accurate and comfortable. Safety is the evaluation of operational safety and operational safety of the situation.Convenience is the evaluation of its service interval, service time and ease of use. Economy is the evaluation of the passenger's transportation costs. Rapid evaluation of the speed of delivery and passenger travel time. The analysis is accurate arrival time offset and punctuality rate. Comfort is the evaluation of the full load rate and the seat of the transport system. The two is the reliability of the system, which is divided into Business Hours reliability, time reliability and good reliability, reliability of the four aspects of the number of passenger. For the above ten aspects of the evaluation index of the database as well as the results of previous studies and the actual situation of Shenzhen set the evaluation criteria, and finally establish the system energy efficiency formula based on the ADC model to calculate the energy efficiency of each station. Then integrated all the energy efficiency values to evaluate the whole system. The lowest efficiency of the station is the most unreasonable station，and we can get the most reasonable arrangement according to the scoring process.

3. Model assumptions and Symbol Description

3.1 Model hypothesis

1、Suppose provide complete data can represent the actual situation in Shenzhen traffic

2、Suppose the data attribute values defects recorded as invalid information

3、Suppose urban public transport system is in a complete state of standby;

4、Assuming the punctuality rate, travel time, Business Hours accord with normal distribution, Passenger numbers are accord with the Poisson distribution.

3.2 Symbol Description

Symbolic name	Symbolic meaning
L_1	A ride distance（km）
$\overline{L_1}$	Average ride distance（km）
k_1	Coefficient，integer representation
$\overline{L}(m)$	Average station distance of bus line network
S_l	Actual service time interval of a certain period of time
α , β , γ , λ	Correction factor
S_{l0}	Theoretical service time of a certain period of time
L	Traffic line group
$\bar{t}$	OD between the actual average travel time
$\overline{t_0}$	OD to the theoretical average time
σ_i	Actual offset of vehicle i arrival time
σ_{i0}	Limit offset of i arrival time for vehicle
p_i	Actual passenger number of vehicle i
p_{i0}	Vehicle I limited number of passenger vehicles i

4、Establishment and solution of the model

4.1 Model building and solving of the Question one:

4.1.1 Apriori algorithm:

In view of the massive amounts of data given to the subject, the first thing we think of is

the data cleaning. To remove a large number of "dirty data" contained in the mass of data，we first introduce the principle of binning and clustering, giving weights to the time variable. Finally, the key effective information of urban public transport system service quality is obtained. Meanwhile, remove some invalid attribute value and reduced data size。

Then, we use the Apriori algorithm to draw one dimensional Boolean association rules frequent itemsets. Taking into account the impact of the public transport system for the quality of service quality and the reasons for the huge data capacity, we adopt the algorithm for fast generation of strong set. Specific procedures are as follows:

(1) Scan the database for each table to produce the Candidate 1 itemsets' collection C_1;

(2) According to the minimum support min_sup, via the Candidate 1 itemsets' collection C1,we produce strong itemsets' collection L_1，for in the database occurrences less than minimum support min_sup counts attribute columns for logical Tags, skip these properties in a subsequent scan;

(3) solve k item set，order k=1;

(4)By L_k produce candidate（k+1）item set's collection C_{k+1};

(5) According to minimum support min_sup, by candidate (k+1) item set's collection C_{k+1},produce (k+1) strong set's collection L_{k+1}, the method is to scan the database, when executed to the i line:

① If the length of the item set is less than (k+1), make logical markup for the line. In the next scan, we all can skip the line no longer scan;

② If the length of the line is equal to (k+1), determine the pattern of the line item set to match the pattern of the candidate set. Matching success is the support of the set of the counter +1. Other modes of candidate set, in the bank no longer scan; If the match is not successful, skipped;

③ If the length of the line is greater than (k+1), the support counter for the item set which will be matched with the candidate item set (k+1) add 1. Comparing the support of all the itemsets in the candidate set Ck+1 with that of min_sup, produce L_{k+1}.

（6）If L_{k+1} is not equal to the empty set, k=k+1, jump to step (7);

（7）According to the minimum confidence min_conf, generate association rules from a strong point set, end.

Take 20140609 metro station credit card data as an example, through the Apriori association rules analysis card records encoding, trading amount (before discount), actual transaction amount (after discount), transaction times, corporate name, potential link between get in and out of the station. As shown in the following table:

Table 4-1-1 20140609 Metro station credit card database

卡片记录编码	交易金额（打折前）	实际交易金额（打折后）	交易时间	公司名称	进出站
251386615	2	1	2014/6/9 0:00	地铁五号线	出站
280228491	3	2.85	2014/6/9 3:12	地铁五号线	进站
290392057	7	6.65	2014/6/9 5:34	地铁五号线	出站
293865243	5	4.35	2014/6/9 9:00	地铁五号线	出站
320179727	3	2.85	2014/6/9 14:01	地铁五号线	进站
323341960	3	2.85	2014/6/9 13:21	地铁五号线	出站
330039918	6	5.7	2014/6/9 2:21	地铁五号线	出站
880026438	2	1.9	2014/6/9 16:55	地铁五号线	出站
293899247	7	6.65	2014/6/9 9:02	地铁三号线	进站
330157913	4	3.75	2014/6/9 19:21	地铁三号线	出站

1. First convert the actual problem of DBS into logical values:
make the transaction amount (discount before) discretization (1:the transaction amount (discount before)<5,2: the transaction amount (discount before)>5);make the actual transaction amount (after a discount) discretization(3:the actual transaction amount (after a discount)<5,4:the actual transaction amount (after a discount)>5);make exchange hour discretization（5:the transaction times between 2014/6/9 00:00:00 and 2014/6/9 08:00:00,6:the transaction times between 2014/6/9 08:00:00and2014/6/9 16:00:00,7::the transaction times between 2014/6/9 16:00:00 and 23：59:59）；make the name of the company discretization(8:line 3,9:line 5）；dialyzed the import and export station(10:get in,11:get out of). Through the above steps, the tables 4-1-2 give a logical table corresponding to Table 4-1-1.

Table4-1-2 Logical database corresponding to database

卡片记录编码	交易金额（打折前）		实际交易金额（打折后）		交易时间			公司名称		进出站	
	1	2	3	4	5	6	7	8	9	10	11
251386615	1	0	1	0	1	0	0	1	0	0	1
280228491	1	0	1	0	1	0	0	1	0	1	0
290392057	0	1	0	1	1	0	0	1	0	0	1
293865243	0	1	1	0	0	1	0	1	0	0	1
320179727	1	0	1	0	0	1	0	1	0	1	0
323341960	1	0	1	0	0	1	0	1	0	0	1
330039918	0	1	0	1	1	0	0	1	0	0	1
880026438	1	0	1	0	0	0	1	1	0	0	1
293899247	0	1	0	1	0	1	0	0	1	1	0
330157913	1	0	1	0	0	0	1	0	1	0	1

Using association rule algorithm to find out value and potential information among attributes in the Table 4-1-2, the value of each Item in each record is obtained by the logical database. As shown in Table 4-1-3。

Table 4-1-3 Value collection of attribute items recorded in the database

卡片记录编码	Items	卡片记录编码	Items
251386615	1, 3, 5, 8, 11	323341960	1, 3, 6, 8, 11
280228491	1, 3, 5, 8, 10	330039918	2, 4, 5, 8, 11
290392057	2, 4, 5, 8, 11	880026438	1, 3, 7, 8, 11
293865243	2, 3, 6, 8, 11	293899247	2, 4, 6, 9, 10
320179727	1, 3, 6, 8, 10	330157913	1, 3, 7, 9, 11

2、Set minimum support degree min_sup=0.5，minimum confidence min_conf=0.7 ,to find association rules. By querying the database, we get k candidate, K's strong set (Ln) and association rules.

(1)solve 1 candidate and 1 strong set，As shown in table 4-1-4。

Table 4-1-4 1 candidate and 1 strong set

Item	sum	sup(I)	L_1	Item	sum	sup(I)	L_1
{1}	6	6/9	✓	{7}	2	2/9	
{2}	4	4/9		{8}	8	8/9	✓
{3}	7	7/9	✓	{9}	2	2/9	
{4}	3	3/9		{10}	3	3/9	
{5}	4	4/9		{11}	7	7/9	✓
{6}	4	4/9					

So strong set of 1 L1={{1}，{3},{8},{11}}.

（2）Get strong set of 1 by strong set of 2，As shown in table 4-1-5.

Table 4-1-5 2 candidate and 2 strong set

Items	Sum	$sup(I_m \cup I_n)$	L_2
{1, 3}	6	6/9	✓
{1, 8}	5	5/9	✓
{1, 11}	4	4/9	
{3, 8}	6	6/9	✓
{3, 11}	5	5/9	✓
{8, 11}	6	6/9	✓

So strong set of 2 L_2={{1,3}，{1,8}，{3,8}，{3,11}，{8,11}}。

（3）By spending degree of strength set 1 sup（A）,calculate the credibility of the strength set 2 conf(Im=>In)=sup(Im∪In)/sup(Im)，getting two association rules, as shown in table 4-1-6.

Table 4-1-6 the credibility of the strength set 2

Items	$sup(I_m \cup I_n)$	sup(Im)	sup(In)	conf（Im=>In)	2项关联规则
{1, 3}	6/9	6/9	7/9	1	✓
{1, 8}	5/9	6/9	8/9	5/6	
{3, 8}	6/9	7/9	8/9	6/7	
{3, 11}	5/9	7/9	7/9	5/7	
{8, 11}	6/9	8/9	7/9	6/8	

Calculate the credibility and association rules of a number of strengths, and so on.

Then convert data into a form suitable for mining。GPS monitoring data of ground bus as an example, in order to prevent a larger range of initial attribute and attribute has a smaller range than the initial weight is too large, we standardize data processing, select min-max normalization for linear transformation of raw data. Scaling attribute data by scale, make it fall into a small specific range. Set minA and maxA as the minimum and maximum values for the attribute A. Minimum-maximum normalized calculation:

$$V_1 = \frac{v - \min A}{\max A - \min A}(new\_\max A - new\_\min A) + new\_\min A$$

Mapping the value V of A to V1 ,which in the interval [new_minA,new_maxA]. The relationship between the original data and the value of the data is still maintained.

Finally, the data is clustered using k-means clustering method. The data of the database is divided into k clustering to satisfy the obtained clustering: The similarity of objects in the same cluster is relatively high, But the object similarity in different clustering is small. The concrete working process is as follows: Firstly, the K object is selected as the initial clustering center from the N data object; but for the rest of the object, according to their similarity with the cluster centers，They are assigned to their most similar clustering centers; Then calculate the clustering center of each new cluster(the mean of all the objects in the cluster); Continue to repeat this process until the standard measure function is beginning to converge. Use mean square deviation as a standard measure function，generally.

4.1.2 Association algorithm

Considering the destruction of the original data in the clustering method，we only remove some invalid and defective items during the cleaning. To the wrong data tuple, such as the direction of the bus GPS records, some of the data is abnormal, with the practical problems of the response to the data, need to analyze data, change, delete or ignore the processing. And if the missing data, analysis of the final results will also have a great impact, data capacity of

this database is large, we can be directly to the vacant value of the delete operation.
After a preliminary clean-up of the data, Then analyses the gray correlation degree of clustering. Grey relation analysis is a method to measure the degree of correlation among the factors in the system. That is, according to the degree of the similarity of the gray time series curve geometry to determine whether the connection is close. The closer the curve is, the more the correlation between the corresponding grey time series is, and the smaller the vice versa. Grey relational degree analysis method will be the research object and the factor of influencing factors as a line of points, compare with the curves drawn by the factor values of the object and the factors to be identified, and compare the degree of closeness between them, quantified, calculate the degree of correlation between the study object and the influence factors of the object to be identified. Compare the degree of the degree of the degree of relevance to determine the impact of the object to be identified. Specific analysis process is as follows. (1) Determine the reference sequence of the behavior characteristics of the system and the comparison sequence of the system behavior. Data sequence, which reflects the behavior of the system, is called a reference sequence. Data sequence, which is composed of factors that influence the system behavior, is called the sequence of numbers. (2) Non dimensional processing of reference and comparative sequence. Because the physical meaning of each factor is different in the system, the dimension of the data is not necessarily the same, it is not easy to compare, or difficult to get the right conclusions in comparison. Therefore, when analyze the gray relational, the general need to conduct non dimensional data processing. (3) Solve the grey correlation coefficient between the reference sequence and the comparison sequenceξ（Xi）. The so-called correlation degree, essentially, the difference in the geometry of the curves is the difference between the curves. So the difference between the curves can be used as a measure of the degree of correlation. For a reference sequence X0 ,there are a number of comparative numbersX1, X2,…, Xn, the correlation coefficient of the comparison sequence and the reference sequence in every moment (i.e., the points in the curve)ξ（Xi）can be calculated by the following formula.
The correlation coefficientξ（Xi）can be simplified as the following formula:

$$\xi_{0i} = \frac{\Delta(\min) + \rho\Delta(\max)}{\Delta_{0i}(k) + \rho\Delta(\max)}$$

(4)Get the relative degree. Because the correlation coefficient is the correlation degree between the comparison sequence and the reference sequence in every moment (that is, the points in the curve), so it's more than one. And the information is too scattered to carry out the overall comparison, so it is necessary to focus on the correlation coefficient of every moment (that is, every point in the curve), that is the average value, as the number of the correlation between the number of the sequence and the reference sequence, correlation formula is as follows:

$$r_i = \frac{1}{N}\sum_{k=1}^{N}\xi_i(k)$$

r_i as gray correlation degree of Xi of the reference sequence X0, or called sequence correlation, average correlation degree and linear correlation degree。ri value is closer 1， the

better the correlation。

(5) Correlation ranking

The correlation degree between the factors, which is mainly used to describe the size of the correlation degree, but not only the size of the correlation degree。The correlation degree of the m sequence to the same parent sequence is arranged in order of magnitude, credited as {x}, it reflects the "good and bad" relationship of the sub sequences of the parent sequence. If r0i>r0j, The {xi} is better than {x0} for the same parent {xj},called{xi}>{xj} ;r0i representate of the sequence of I in the first sequence.

Commonly used gray correlation degrees, such as the degree of correlation, the concrete steps are as follows:

Step one: find the initial value of the sequence (or the mean image). Order:

$X_i' = X_i / x_i(1) \Rightarrow x_i'(1), x_i'(2), \cdots, x_i'(n)$

$\Delta_i(k) = \left| x_0'(k) - x_i'(k) \right| \Rightarrow \Delta_i = \Delta_i(1), \Delta_i(2), \cdots, \Delta_i(n) \quad i = 1, 2, \cdots, m$

Step two: seeking the maximum difference between two poles, remember to:

$M = \max_i \max_k \Delta_i(k), \quad m = \min_i \min_k \Delta_i(k)$

Step three：Figure out coefficient of correlation.

$$r(x_0(k), x_i(k)) = \frac{m + \rho M}{\Delta_i(k) + \rho M} \quad (k = 1, 2, \cdots, n; i = 1, 2, \cdots, m)$$

Step four: calculate correlation degree.

$$r_{0i} = \frac{1}{n} \sum_{k=1}^{n} r_{0i}(k); \quad i = 1, 2, \cdots, m$$

Finally, the data is based on the objective function to achieve the clustering algorithm. Iterative optimization algorithm——K-means, In essence, the iterative optimization is a local search method, which is easy to fall into local extremum and is sensitive to the initial value. If n samples $x_j \in \Re^N (j = 1, 2, \cdots n)$ are divided into c categories，For i=1,2,... , C and j=1,2,... , n can be defined:

$$\mu_{ij} = \begin{cases} 1, \text{如果第} j \text{个样本属于第} i \text{类} \\ 0, \text{否则} \end{cases}$$

The matrix $\mu = (\mu_{ij})$ has the following properties::

$\mu_{ij} \in \{0, 1\}$ 且 $\sum_{i=1}^{c} \mu_{ij} = 1 \quad (j = 1, 2, \cdots, n)$

Let n_i denote the number of samples contained in the i class.

$$n_i = \sum_{j=1}^{n} \mu_{ij} \quad i = 1, 2, \cdots, c$$

Let's set $\overline{x_i} \in \Re^N$ be the center of class i.

$$\overline{x_i} = \frac{\sum_{j=1}^{n} \mu_{ij} x_j}{\sum_{j=1}^{n} \mu_{ij}} = \frac{1}{n} \sum_{j=1}^{n} \mu_{ij} x_j \qquad i = 1, 2, \cdots, c$$

The i class in the difference:

$$S^{(i)}(\mu) = \sum_{j=1}^{n} \mu_{ij} \left\| x_j - \overline{x_j} \right\|^2$$

Within the overall class difference:

$$S(\mu) = \sum_{i=1}^{c} S^{(i)}(\mu) = \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij} \left\| x_i - \overline{x_i} \right\|^2$$

k- mean iterative optimization aims to findμ_{ij}, So get the minimum $S(\mu)$, namely:

$$S(\mu^*) = \min\{S(\mu)\}$$

4.1.3 K-median algorithm

The process of data cleaning is similar to that of the two methods mentioned above, but only the removal of some duplication, defect, and invalid item. Then consider the k-averaging algorithm for data types of limitations, therefore the use of k- center algorithm, after data processing, call reports dissimilarity, obtained pretreatment on the impact of the data generated by the original data.

k- center algorithm according to the spatial distribution of the data set of samples neighborhood radius and neighborhood information to define the sample, first select the area of the most intensive in the sample as the first class of the initial cluster centers, and centralized omitted from the sample data Its all samples neighborhood; then select in the most dense areas of the samples from the remaining sample data set for the second class of the initial cluster centers, and concentrated by deleting the sample of all samples and their neighborhood from the data; This process continues until the selected K initial centers, the initial allocation of the sample to the nearest center, get results K initial cluster centers algorithm to calculate the cluster sum of squared errors; to all kinds of manipulation cluster clustering error sum of squares The minimum sample such as a new center cluster Update Center each class clusters; redistribution of sample to the nearest cluster center class, get new clustering results, calculate the cluster sum of squared errors; if no clustering error sum of squares change, then the algorithm ends; otherwise choose to make all types of cluster clustering square error and minimum sample update class cluster centers in various types of clusters; allocation of the sample to the nearest cluster center class, get new clustering results, calculate the cluster error sum of squares, squared error and determine whether the cluster changes, in order to determine the iteration to continue or end of the iteration. Here you select the mutual distance of the K sample distribution in dense regions of data objects as the initial cluster centers, so that clusters and the smallest error square sample point update class cluster center, until Clustering Squared error and changes do not occur, so that not only improve the convergence speed, but also achieve better clustering.

Set up a collection of data containing N data objects X={x1,x2,…,xn} Each data object

contains P dimensional features, and is now divided into k class Cj ， j ＝ 1 ， 2 ， …， k ， k ＜ n 。The first j I of the data object is x_{ij}.

Definition 1 Euclidean distance definition for any two data objects，as follows：

$$d(x_i, x_j) = \sqrt{\sum_{n=1}^{p} (x_{ia} - x_{ja})^2}$$

Definition 2 Data focus on the density of each data object Xi Density(x_i) be defined as：

$$Density(x_i) = \sum_{j=1}^{n} \frac{d(x_i, x_j)}{\sum_{l=1}^{n} d(x_i, x_j)}, \qquad i = 1, 2, \cdots, n$$

Definition 3 The sum of squared errors is defined as $E = \sum_{j=1}^{k} \sum_{t=1}^{n_j} \| x_{jt} - m_j \|^2$

Where x_{ij} is the j-Class t sample points, m_j is j-class cluster center, n_j j-class cluster is the number of samples.

Definition4 The average distance between data objects Mean Dist. defined as:

$$DistMean = \frac{2}{n(n-1)} \times \sum d(x_i, x_j$$

n is the number of sample data sets, the size of the dataset. Dataset is any distance between two different samples and.

Definition 5 According neighborhood radius R is defined as an object $R = \frac{DistMean}{n^{cR}}$ ， cR is the neighborhood radius adjustment coefficient， $0 < cR \le 1$； n is the size of the data set.

Definition 6 Data objects neighborhood: For any data object xi to xi as the center R is the radius of the circular area contains a collection of data objects, called data objects xi neighborhood, with δ expressed formally defined as:

$$\delta = \{ x_j \mid 0 < d(x_i, x_j) < \}$$

Algorithm description:

（1）Initialization center

Step1: according to the definition of 2 to calculate each data object Xi density Density (Xi), and on the basis of Density (Xi) value will be $X = \{x_1, x_2, \cdots, x_n\}$ in ascending order sample set; center initialization set M is empty, $M = \{\}$.

Step2: Select the value of the minimum focus, that is the data object $x_{\min}$ densest region as the first initial centers will be added to the center of focus, namely $M = M \cup \{x_{\min}\}$, and concentrated by deleting the object from the data, that $X = X - \{x_{\min}\}$ from the data. According

to the definition 5 and definition of neighborhood 6 calculates $x_{\min}$ from the data set by deleting all the data objects in its neighborhood (samples).

Step3：Repeat until k initial centers concentrate containing a center ,$|M|=k$。

Step4：Output initial centers set M.

（2）Update class cluster center

Step1：The data is centralized data assigned to its nearest center, and calculated according to the definition of cluster 3 square errors.

Step2： For each class of looking for a new center, so that the point to which it is the sum of the minimum distance in class cluster of other sample points.

Step3：Update all class cluster center.

（3）UP date all class cluster center.

Step1：The distribution data set of data objects to its closest class center。

Step2：According to the definition computing cluster 3 square error and, if the clustering error square and there is no change, then the algorithm ends; otherwise go to (2) to continue.

K center clustering algorithm described above in the initialization clustering centers, the introduction of the concept of neighborhood distribution information defines the radius of the neighborhood and its neighbors based on the sample space sample the entire data set. Select samples in the data set around the most densely distributed sample for the first initial centers, concentrated by deleting the sample of all samples and neighborhood from the data; in turn elected the same way a second, third, or even all K initial center. So you can choose the distance to each other and in a dense area of the sample distribution K samples as the best initial cluster centers. Clustering and choose to make the smallest error square class cluster sample update class cluster center. In the algorithm, the start time to calculate the distance between all the samples do not need to repeat the calculation behind, thereby reducing the computational burden, but also accelerate the convergence of the algorithm.

Finally, the data AGNES algorithm processing. AGNES algorithm is hierarchical clustering method cohesion. AGNES algorithm initially each objects as a cluster, then these clusters are combined according to certain guidelines step by step. For example, if a cluster to an object in the distance C1 and C2 in a cluster between objects belonging to different clusters are all representatives of all the objects, the similarity between two clusters by two different clusters in the nearest data point for determining similarity. Cluster merging process is repeated until all the objects eventually combined to form a cluster. In clustering, the user can define the number of clusters hope to get as an end condition. AGNES algorithm process is as follows:

Input: contains n data object database, the number of clusters k termination condition

Output: up to k clusters termination specified conditions

Step1: each object as an initial cluster;

Step2: find the nearest two clusters according to the most recent data point of the two clusters;

Step3: merge two clusters, generating a new set of clusters;

Step4: number to reach defined clusters until Step3 Step4 cycle.

4.1.4 Solving the problem one

Due to the huge amount of data, detailed results in Appendix screenshot

4.2 Model building and solving of the Problem two

Problem two asks us to analyze the OD matrix of residents travel data based on GPS and bus card consuming. We think it consists of the following steps.

Step one: Introduce the theory of data fusion principle. With the help of this principle, the purpose of which is to GPS data and IC card data be made one more effective, the formation of data, improve the maximum utilization value of two sets of data;

Step two: The introduction of bus travel analysis of two site model "and Earl give -K curve, so as to obtain a complete travel passenger information, single line passenger OD matrix drawing.

Step three: The introduction of the concept of" Public transport area", the city will be divided into a number of public transport area according to the specific principle of dividing principle, then the total number of auxiliary parameters, so as to calculate the city bus OD matrix;

Step four: According to the city bus OD matrix, the travel time distribution characteristics and travel space distribution and characteristics of the analysis, get the conclusion.

4.2.1 The introduction of data fusion theory

After obtaining buses GPS data, we can analyze the basic data of buses number, the time distribution of the bus running position and the direction of the bus running. After inserting the IC card data into the GPS database, we streamlined the data, only to retain the IC card number, the time on the buses, credit card amount and the bus number and other data. Then generated two sets of the data: the table 4-2-1 bus GPS basic data table and table 4-2-2 bus IC card based information table two sets of the data.

Table 4-2-1 Bus GPS basic data table

license plate number	Travel time	Latitude and longitude	Driving direction
粤 BN4646	2014/6/8 7:14	（114.18,22.6383）	56

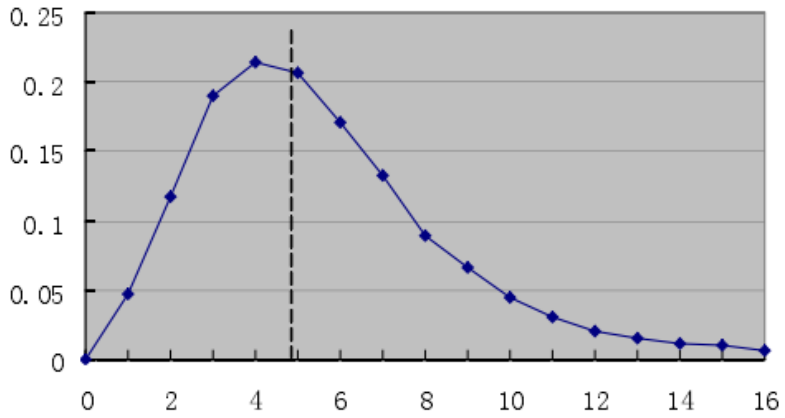
Table 4-2-2 Bus IC card based information table

IC card number	Boarding Time	Amount deducted from card（Yuan）	License plate number
292915914	2014/6/9 0:00	1.2	粤 BW2839

4.2.2 The introduction of Bus travel analysis model of the two site and Earl laenger -K curve and the Establishment of single line bus OD matrix

There are two models of urban residents to travel by bus:1.One-way model, bus passengers from the starting point to get on the bus and get off at the end to complete model of the bus travel; 2.Double-way model, bus passengers from the starting point to get on the bus and get off at the end, bus travel mode from the point of return to the starting point. As the information provided in the table is not record the sites of which passengers get off, We based on "direction criterion" and "random function" random allocation passengers to the get off station: At this time the vehicle driving direction is the premise, from the passenger on the site of the next station, through random function to get off site. At this time the vehicle driving direction is the premise, from the passenger on the site of the next station, through random

function to get off site.



Picture 4-2-1 Earl laenger -K curve

The probability density calculation formula of the Earl laenger -K curve:

$$P(L_1)=\frac{\left[k_1-\frac{1}{\overline{L_1}}\right]^{k_1}}{(k_1-1)!}\overline{L_1}^{k_1-1}e^{k\frac{L_1}{\overline{L_1}}}$$

L_1 A ride distance（km）;

$\overline{L_1}$ Average ride distance（km）;

k_1 Coefficient，integer representation.

We can use the following formula to calculate the coefficient k_1： $k_1=\left(\frac{\overline{L_1}}{\Delta L}\right)^2$

In the equation: ΔL The allowable deviation value of once travel distance value (km)

For several station that only have passengers get on, We use "Cumulative usage" as Weight，and distribution by random functions. If in a given direction criterion, did not meet the conditions of the station, the site off by Earl laenger K curve random distribution.

If in a given direction criterion, did not meet the conditions of the station, the site off by Earl laenger K curve random distribution. According to the above analysis, we can obtain the complete information on bus travel. As shown in figure 4-2-3.

Table 4-2-3 Residents travel basic information table

IC card number	boarding station	Boarding Time	Driving direction	Alighting station	Alighting Time	License plate number
292915914	00001	07:00:19	56	00012	07:26:14	粤 BN4646

According to each record in the form of data, through the corresponding processing, the final form of the bus OD matrix. Specific steps are as follows：

1、Using Shenzhen bus route 62 as a model, the number of its 42 stations 0001-0042 ,the line of the bus station as the matrix, starting point and the end column";

2、Read the records of the basic information table in the system order, and add 1 in the corresponding matrix form;

Arranging the line's card passenger information, and generating the OD matrix.

Figure 4-2-4 62 Road OD matrix partial screenshot

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	0	0	1
2	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
3	1	0	0	0	0	1	1	1	1	1	1	2	1	2	1	2	1	1	1	1
4	1	0	0	0	0	1	1	1	2	1	1	2	2	2	2	2	1	1	1	2
5	1	0	1	1	0	1	1	2	2	2	2	3	3	3	2	3	2	2	2	2
6	0	1	2	0	1	0	1	1	2	1	1	2	2	1	2	2	2	1	2	2
7	0	1	1	1	2	2	0	2	4	3	3	4	4	4	4	5	3	3	3	4
8	3	0	1	1	1	2	2	0	4	2	2	2	4	4	3	4	3	3	3	3
9	4	1	2	3	2	3	3	4	0	5	5	8	8	8	7	8	6	5	5	7
10	2	1	1	1	1	3	3	3	5	0	4	6	6	6	5	6	4	4	4	5
11	1	1	1	1	1	2	3	3	5	3	0	6	5	6	5	6	4	4	4	5
12	2	1	1	1	3	2	2	3	5	4	3	0	6	6	5	6	4	4	4	5
13	4	1	1	1	2	3	3	4	6	4	4	7	0	7	6	7	5	5	5	6
14	3	1	1	1	1	2	2	3	5	3	3	6	6	0	5	6	4	4	4	5
15	2	1	1	1	1	2	2	2	4	3	3	4	4	4	0	5	3	3	3	4
16	0	1	1	1	1	2	2	3	4	3	3	5	5	5	4	0	4	3	3	4
17	1	0	1	1	1	2	2	2	3	2	2	4	4	4	3	4	0	3	3	3
18	1	3	2	1	1	1	2	2	3	2	2	3	3	3	3	4	2	0	2	3
19	1	0	1	1	1	2	2	2	3	2	2	4	4	4	3	4	3	3	0	3
20	1	2	1	3	2	3	3	3	5	4	4	6	6	6	5	6	4	4	4	0
21	3	1	1	1	2	2	2	3	5	3	3	5	5	5	5	5	4	4	4	4

4.2.3 The introduction of the concept of “traffic zone" and generation of OD matrix of bus lines in the city

4.2.3.1 Method of traffic area merging

Because of the traditional administrative divisions to establish the method of traffic area has some defects, cannot be a good response to the basic information, so we use the bus station and City Road intersection as the center, set up a number of station set as a small area of the bus, and then apply the cluster analysis method to divide the whole city into a suitable number of bus. Specific methods are as follows:

1、Taking the city road intersection as the center, and combining surrounding bus station. The existing bus routes are based on the existing road, which can be used to construct the traffic flow.

2、Putting the center of each road intersection as the central geographic coordinates as the sub district of encoding.

3、Setting up the data dictionary of traffic sub district and public transport links. Setting up the data dictionary of traffic sub district and public transport links. The mapping relationship between the data dictionary of the data dictionary and the bus station is set up in the database.

4、Using clustering analysis method to merge the sub regions into the public transportation area.

4.2.3.2 Comparison of two kinds of clustering algorithms

In accordance with the above steps，but we have too many traffic areas, in order to get the ideal number of residential areas, we should optimize our clustering method. After analysis, we believe that the method of pedigree clustering and fast clustering can accomplish the task.

Pedigree clustering algorithm：As long as the given clustering level, there will be a corresponding classification number。When using this method, the clustering of similar areas is completely determined. However, for fast clustering algorithm, the aggregation point of the areas is given by the experience, and the final classification is related to the selection of the

station.

As for the pedigree clustering algorithm of the Public transport area，it is usually "a calculation, multiple use". As long as the bus station and the classification principle is the same, the district will not be re calculated. This shows that the public transport area of the cluster, in the time of the request is not very prominent, the advantages of fast clustering method is not very good. The hierarchical clustering method can be used to guarantee the number of the final public transportation cell to achieve a predetermined level.

4.2.3.3 Clustering method in public traffic area

We cluster merging traffic area. Firstly, Dividing the urban area into the area and the big area, and then the large areas in the application of the spectrum of the cluster analysis method for the merger of each district.

4.2.3.4 Generation of OD matrix of public transportation in Shenzhen

The generating method of the OD matrix of Shenzhen city bus is similar to that of line OD matrix. In order to avoid the occurrence of passengers in the "Two sites with line selection" model, the cardholder's travel information is recorded in a number of lines, resulting in artificially fragmented data, the IC card data of the bus line that should be directly related to the study area should be merged into a table when generating of the Shenzhen OD matrix.

According to the basic information of residents travel, by the same method as above, the bus travel matrix of the region's card passengers can be generated.

4.2.4 Bus OD matrix analysis

According to the generated OD matrix, we analyze the travel time distribution characteristics and the distribution characteristics of the travel space of the matrix, and then draw the conclusion.

4.2.5 Model solving of the Problem two

To establish a line of single bus OD matrix, the results as shown below:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
1	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	0	0	1	0	1	1	0	0	0	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	1
2	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0
3	1	0	0	0	0	1	1	1	1	1	1	2	1	2	1	2	1	1	1	1	1	1	2	1	1	1	1	1	1	2	1	1	2	2	1	1	1	1	1	0	0	0
4	1	0	0	0	0	1	1	1	2	1	1	2	2	2	2	2	1	1	1	2	1	2	2	1	1	1	2	1	2	2	2	2	3	1	1	1	1	1	1	1	0	
5	1	0	1	1	0	1	1	2	2	2	2	3	3	3	2	3	2	2	2	2	2	2	3	2	2	2	2	2	3	3	3	3	3	1	2	2	1	1	1	1	1	0
6	0	1	2	0	1	0	1	1	2	1	1	2	2	1	2	2	2	1	2	2	2	2	2	2	1	2	2	2	2	3	2	2	2	3	2	2	1	2	1	1	1	2
7	0	1	1	1	2	2	0	2	4	3	3	4	4	4	4	5	3	3	3	4	3	4	4	3	3	3	4	3	4	5	4	4	5	6	3	3	2	2	1	0	1	1
8	3	0	1	1	1	2	2	0	4	2	2	2	4	4	3	4	3	3	3	3	3	4	4	3	2	3	3	3	4	5	4	4	4	5	3	3	2	2	1	1	1	0
9	4	1	2	3	2	3	3	4	0	5	5	8	8	8	7	8	6	5	5	7	5	7	8	5	5	5	7	6	8	9	8	7	8	11	5	6	3	4	3	2	1	1
10	2	1	1	1	1	3	3	3	5	0	4	6	6	6	5	6	4	4	4	5	4	5	6	4	4	4	5	5	6	7	6	6	6	10	4	5	2	3	2	2	1	1
11	1	1	1	1	1	2	3	3	5	3	0	6	5	6	5	6	4	4	4	5	4	5	6	4	3	4	5	4	5	6	5	5	6	7	4	4	2	3	2	1	1	1
12	2	1	1	1	3	2	2	3	5	4	3	0	6	6	5	6	4	4	4	5	4	5	6	4	3	4	5	4	5	7	5	5	6	8	4	4	2	3	2	1	1	1
13	4	1	1	1	2	3	3	4	6	4	4	7	0	7	6	7	5	5	5	6	5	6	7	5	4	4	6	5	7	8	7	6	7	9	5	5	3	3	2	2	2	2
14	3	1	1	1	1	2	2	3	5	3	3	6	6	0	5	6	4	4	4	5	4	5	6	4	3	4	5	4	5	7	5	5	6	8	4	4	2	3	2	1	1	3
15	2	1	1	1	1	2	2	2	4	3	3	4	4	4	0	5	3	3	3	4	3	4	4	3	3	3	4	3	4	5	4	4	5	6	3	3	2	2	1	1	1	1
16	0	1	1	1	1	2	2	3	4	3	3	5	5	5	4	0	4	3	3	4	3	4	5	3	4	4	4	4	5	6	5	5	5	7	3	4	2	2	4	1	1	1
17	1	0	1	1	1	2	2	2	3	2	2	4	4	4	3	4	0	3	3	3	3	3	4	1	2	2	3	3	4	4	4	4	4	5	3	3	2	2	1	1	1	0
18	1	3	2	1	1	1	2	2	3	2	2	3	3	3	3	4	2	0	2	3	2	3	4	2	2	2	3	3	3	4	3	3	4	5	2	3	1	2	1	1	1	0
19	1	0	1	1	1	2	2	2	3	2	2	4	4	4	3	4	3	3	0	3	3	4	4	3	2	3	3	3	4	5	4	4	4	5	3	3	2	2	1	1	1	0
20	1	2	1	3	2	3	3	3	5	4	4	6	6	6	5	6	4	4	4	0	4	5	6	4	4	4	5	4	6	7	6	6	6	8	4	4	2	3	2	2	1	1
21	3	1	1	1	2	2	2	3	5	3	3	5	5	5	5	5	4	4	4	4	0	5	5	4	3	3	4	4	5	6	5	5	5	7	4	4	2	2	2	1	2	1
22	0	0	1	1	1	2	2	2	3	2	2	4	4	4	3	4	3	3	3	3	3	0	4	3	2	3	3	3	4	5	4	4	4	5	3	3	2	2	1	1	1	2
23	1	1	2	1	1	2	2	3	4	3	3	5	5	5	4	5	4	3	3	4	3	4	0	3	3	3	6	4	5	6	5	5	5	7	3	4	2	2	2	1	2	3
24	2	0	3	1	1	2	2	2	4	2	2	4	4	4	4	4	3	3	3	3	3	4	4	0	2	3	3	3	4	5	4	4	4	5	3	3	2	2	1	1	1	0
25	5	0	1	1	1	2	2	2	4	2	2	4	1	4	3	4	3	3	3	3	3	4	1	3	0	3	3	3	4	5	4	4	4	5	3	3	2	2	1	1	2	1
26	0	1	1	1	1	2	2	2	4	2	2	4	4	4	4	4	3	3	3	4	3	4	4	3	2	0	4	3	4	5	4	4	4	6	3	3	2	2	1	1	2	1
27	3	1	1	1	1	3	3	3	5	4	4	6	6	6	5	6	4	4	4	5	4	5	6	4	4	4	0	5	6	7	6	6	6	8	4	5	2	5	2	2	1	1
28	0	1	2	1	3	2	2	3	4	3	3	5	5	5	4	5	4	3	3	4	3	5	5	3	3	3	4	0	5	6	5	5	5	7	3	4	5	2	2	1	1	1
29	2	6	1	1	3	3	3	4	5	3	3	6	2	6	5	6	4	4	4	5	7	5	6	4	3	4	5	4	0	7	6	6	3	8	4	4	2	1	2	2	3	1
30	0	1	1	1	1	2	3	3	5	3	3	6	6	6	5	6	4	4	4	5	6	5	6	4	3	4	5	4	6	0	6	5	6	8	4	4	2	3	2	2	3	4
31	3	1	2	3	1	3	3	3	5	3	3	6	6	6	5	6	4	4	4	5	4	5	6	4	3	4	5	4	6	7	0	6	6	8	4	4	2	3	2	2	1	1
32	2	1	1	4	1	1	2	2	3	2	2	3	3	3	3	4	2	2	2	3	2	3	4	2	2	2	3	3	3	4	3	0	4	5	2	3	1	2	1	1	1	0
33	4	3	1	1	1	1	2	2	3	2	2	3	3	3	3	4	2	2	2	3	2	3	4	2	2	2	3	3	3	4	3	3	0	5	2	3	1	2	1	1	1	0
34	5	2	1	1	1	2	2	3	5	3	3	5	5	5	4	5	4	4	4	4	3	5	5	3	3	3	4	4	5	6	5	5	5	0	4	4	2	2	2	1	1	4
35	3	1	1	1	1	1	1	2	3	2	2	3	3	3	3	3	2	2	2	3	2	3	3	2	2	2	3	2	3	4	3	3	3	4	0	2	1	2	1	1	1	0
36	1	0	1	0	1	1	1	1	2	2	2	3	3	3	2	3	2	2	2	2	2	2	3	2	2	2	2	2	3	2	2	3	3	3	2	0	1	1	1	1	0	0
37	1	1	0	0	1	1	1	1	2	1	1	3	2	3	2	3	2	2	2	2	2	2	3	2	1	2	2	2	2	3	2	2	3	3	2	2	0	1	1	1	0	0
38	2	3	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	0	0	0	0	0
39	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0
40	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	0	1	0	0	0	0
41	1	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	0	0	1	0	1	1	0	0	0	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0
42	1	1	0	2	0	0	0	0	1	0	2	1	1	1	1	1	1	0	0	1	0	1	1	0	0	0	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0

All lines OD matrix solver, solving the results shown in Figure 4-2-5 (due to the number of data is too large, we only show part of the theme, detailed results see Appendix)

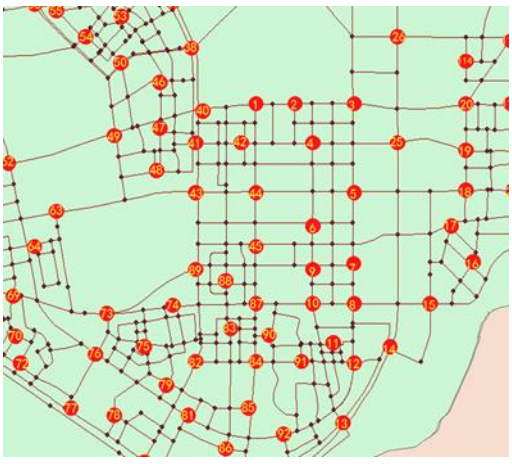
Figure 4-2-5 Bus OD matrix partial view

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0	78	127	117	156	273	283	351	565	380	380	643	633	643	555	663	458	438
2	1	0	190	175	234	409	424	526	848	570	570	965	950	965	833	994	687	658
3	1	166	0	248	331	580	601	745	1201	808	808	1367	1346	1367	1180	1408	973	932
4	1	214	348	0	429	750	777	965	1554	1045	1045	1769	1742	1769	1527	1822	1259	1206
5	1	302	491	453	0	1057	1095	1359	2190	1473	1473	2492	2454	2492	2152	2568	1775	1699
6	0	1	2	395	526	0	954	1184	1907	1283	1283	2171	2138	1	1875	2236	1546	1
7	0	477	776	716	2	1671	0	2149	3462	2328	2328	4	3879	3939	3402	4059	2805	2686
8	3	438	713	658	877	1535	1590	0	3179	2138	2138	2	3563	3618	3124	3727	2576	2467
9	4	858	1393	3	1715	3001	3108	3859	0	4180	4180	7074	6967	7074	6110	7289	5038	4823
10	2	663	1077	994	1325	2319	2402	2982	4804	0	3230	5467	5384	5467	4721	5632	3893	3727
11	1	604	982	906	1208	2115	3	2719	4380	2945	0	4984	4909	4984	4305	5135	3549	3398
12	2	614	998	921	3	2149	2	2763	4451	4	2993	0	4988	5065	4374	5218	3607	3453
13	4	750	1219	1125	1501	2626	2720	3376	5440	3658	3658	6190	0	6190	5346	6378	4408	4221
14	3	614	998	921	1228	2149	2225	2763	4451	2993	2993	5065	4988	0	4374	5218	3607	3453
15	2	477	776	716	955	1671	1731	2149	3462	2328	2328	3939	3879	3939	0	4059	2805	2686
16	0	536	871	804	1072	1876	1943	2412	3886	2613	2613	4422	4355	4422	3819	0	3149	3015
17	1	409	665	614	819	1432	1484	1842	2967	1995	1995	3376	3325	3376	2916	3479	0	2302
18	1	3	2	570	760	1330	1378	1710	2755	1853	1853	3135	3088	3135	2708	3230	2233	0
19	1	429	697	643	858	1501	1554	1929	3108	2090	2090	3537	3484	3537	3055	3644	2519	2412
20	1	2	1061	3	2	2285	2367	2938	4733	3183	3183	5386	5305	5386	4652	5549	3836	3672
21	3	575	934	862	2	2012	2084	2587	4168	2803	2803	4743	4671	4743	4096	4887	3378	3234
22	0	429	697	643	858	1501	1554	1929	3108	2090	2090	3537	3484	3537	3055	3644	2519	2412
23	1	546	2	819	1091	1910	1978	2456	3956	2660	2660	4502	4434	4502	3888	4638	3206	3069
24	2	448	3	672	896	1569	1625	2017	3250	2185	2185	3698	3642	3698	3194	3810	2633	2521
25	5	438	713	658	877	1535	1590	1973	3179	2138	2138	3618	1	3618	3124	3727	2576	2467
26	0	458	744	687	916	1603	1660	2061	3320	2233	2233	3778	3721	3778	3263	3893	2691	2576
27	3	663	1077	994	1325	2319	2402	2982	4804	3230	3230	5467	5384	5467	4721	5632	3893	3727
28	0	1	2	1	3	1944	2013	2499	4027	2708	2708	4582	4513	4582	3957	4721	3263	3124
29	2	6	1045	965	3	2251	2331	4	4663	3135	3135	5306	2	5306	4582	5467	3778	3618
30	0	633	1029	950	1267	2217	2296	2850	4592	3088	3088	5225	5146	5225	4513	5384	3721	3563
31	3	643	2	3	1286	2251	2331	2894	4663	3135	3135	5306	5225	5306	4582	5467	3778	3618
32	2	1	618	4	760	1330	1378	1710	2755	1853	1853	3135	3088	3135	2708	3230	2233	2138
33	4	3	618	570	760	1330	1378	1710	2755	1853	1853	3135	3088	3135	2708	3230	2233	2138
34	5	2	918	848	1130	1978	2049	2543	4098	2755	2755	4663	4592	4663	4027	4804	3320	3179
35	3	1	570	526	702	1228	1272	1579	2543	1710	1710	2894	2850	2894	2499	2982	2061	1973
36	1	283	459	424	565	989	1024	1272	2049	1378	1378	2331	2296	2331	2013	2402	1660	1590
37	1	1	443	409	546	955	989	1228	1978	1330	1330	2251	2217	2251	1944	2319	1603	1535

Transportation cell division

According to the clustering criteria, we will Shenzhen city is divided into 647 traffic zone, dividing the results in Appendix, Figure 4-2-6 is a partial view in Futian District in traffic cell division

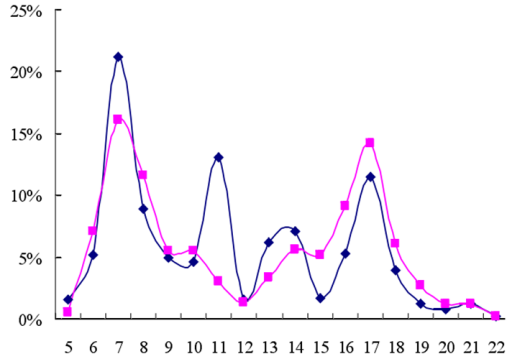
Figure 4-2-6 division of traffic zone, Futian District,



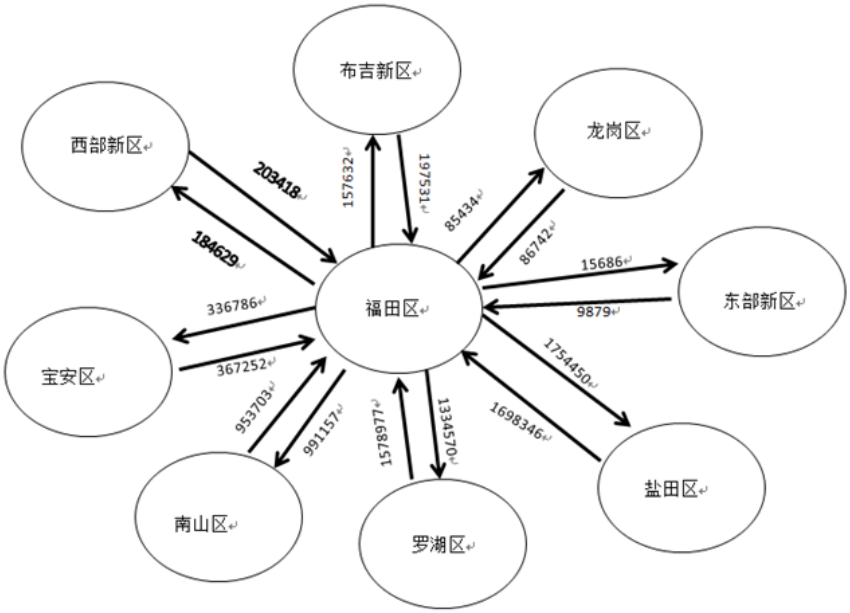
Analysis of the Trip Characteristics

Travel time distribution analysis:

We choose June 9 and 10, the data plotted travel time profile shown in Figure 4-2-7, from the chart we can clearly see Shenzhen concentrated in the morning peak travel 7-9 pm, 4-7 points, so we recommend the transport sector in these two time periods can increase the number of vehicles, slowing the peak passenger flow.



Travel spatial distribution analysis: we buses according to GPS data, draw the Shenzhen district people flow diagram, shown in Figure 4-2-8.



Based on the results, we issued between huge Luohu and Nanshan District, Futian District, traffic, we recommend that companies can increase the operating frequency of the bus and run to open the line, and for what two new districts can increase the distance between each station, reducing transportation cost.

4.3 Model building and solving of the Problem three

Problem three asks us to establish a dynamic model to evaluate the service quality of the bus station and the subway station to find out the three most unreasonable bus station to correct。
We believe that it should be composed of the following steps:
Step 1: the introduction of public transport service quality evaluation system, respectively, from the safety, convenience, economy, rapid, accurate and comfortable several different perspectives to be evaluated; Step two: define the specific method for calculating the six dimensions; Step three: the introduction of public transport service reliability evaluation system;
Step four: the introduction of public transport service energy efficiency evaluation system, the current service quality evaluation and reprioritize the most unreasonable three bus stops.

4.3.1 The introduction of public transport service quality evaluation system

We believe that the evaluation of the service quality of the bus station and the subway station is essentially an evaluation of the service efficiency of a city's public transport system. The service efficiency of the public transportation system is composed of two aspects: the service quality evaluation system and the reliability evaluation system. First we start with the quality of service evaluation system.

Service quality evaluation system consists of the following six dimensions: namely, safety, convenience, economic, rapid, accurate and comfortable. Safety is the evaluation of operational safety and operational safety of the situation. Convenience is the evaluation of its service interval, service time and ease of use. Economy is the evaluation of the passenger's transportation costs. Rapid evaluation of the speed of delivery and passenger travel time. The analysis is accurate arrival time offset and punctuality rate .Comfort is the evaluation of the full load rate and the seat of the transport system.

4.3.2 Six dimension index specific calculation method：Safety, convenience, economy, rapid, accurate and comfortable.

4.3.2.1 Safety

（1）Order A1 as vehicle safety operation interval mileage，Formula is as follows:

Vehicle safety operation interval mileage = Total run / Number of traffic accidents

（2）Bus operation safety status. Main evaluation content①Vehicle components and operating system, compliance②Qualification certificate of division by personnel③Division by bus lines and professional standards of operation④Bus peak load rate is low⑤Develop a safety quality management system⑥To formulate the regulations on traffic safety management⑦Joint development of public security departments to combat criminal acts of the bus

（3）Evaluation criterion

Table 4-3-1 Average safe operating interval mileage rating scale（10^4km/time）

Evaluation level	1	2	3	4	5
A1	≥125	[100,125）	[75,100）	[50,75）	[0,50）
Index	[90,100]	[80,90）	[70,80）	[60,70）	[0,60）

Table 4-3-2 Vehicle running safety status rating scale

Evaluation level	1	2	3	4	5
A2	Meet seven contents	Meet six contents	Meet five contents	Meet for contents	Less than four contents
Index	[90,100]	[80,90）	[70,80）	[60,70）	[0,60）

4.3.2.2 Convenience

（1）Order A3 as Service interval. Formula is as follows：

Service interval = Line turnaround time（min）/ Vehicle number

（2）Order A4 as service time. Specifically referring to the first bus line and the last bus daily operational interval.

（3）Order A4 as travel convenience degree. Main evaluation content①The site provides information about dynamic time②Bus IC card fees③Complete information station reported clear④Car line and around the site simple map, place names, building names, and so the

guidance information guiding ⑤ Transfer indication map ⑥ Other convenient service broadcasting or television media

（4）evaluation standard

Table 4-3-3 Service interval（min）and Service frequency（vel/h）hierarchical list

Evaluation level		1	2	3	4	5
A3	Service interval	<4	[5,4）	[7.5,5）	[10,7.5）	≥10
	Service frequency	>15	[12,15)	[8,12)	[6,8)	≤6
Index		[90,100]	[80,90）	[70,80)	[60,70)	[0,60)

Table 4-3-4 Daily service time Ines（h）

Evaluation level	1	2	3	4	5
A4	≥18	[16,18）	[13,16)	[11,13)	<11
Index	[90,100]	[80,90)	[70,80)	[60,70)	[0,60)

表 4-3-5 Convenience level classification table

Evaluation level	1	2	3	4	5
A5	Meet six contents	Meet five contents	Meet four contents	Meet three contents	Less than three contents
Index	[90,100]	[80,90）	[70,80）	[60,70）	[0,60）

4.3.2.3 Economy

Order A6 as Passenger traffic rate (%) . The calculation formula is as follows：

Passenger traffic rate = Average per 100km ride from the actual travel expenses paid / Transport Services District employees' average monthly wage. The evaluation form is as follows:

Table 4-3-6 Passenger transport rate rating scale（%）

Evaluation level	1	2	3	4	5
A6	[11,8)	[14,11)	[17,14)	[20,17)	>20
Index	[90,100]	[80,90)	[70,80)	[60,70)	[0,60)

4.3.2.4 Rapid

（1）Order A7 as transporting velocity（km/h），The calculation formula is as follows：

Transporting velocity = Distance（km）/Time（h）

（2）Order A8 as Passenger travel time（min）The calculation formula is as follows：Passenger travel time =2* Walking time + Waiting time + Riding time + Transfer time. The indicators identified as follows:

① Passenger walking time：The density of urban public transportation line network is δ（km/km^2）The average distance between the passengers from the starting point to the nearest distance to the bus line is 1/3, so the average passenger walking distance is:L_1=2000/3δ.Let the average station distance of the bus as $\bar{L}(m)$ ，Passengers walk along the road to the average walking distance of the site from the average station 1/4，Average distance to the station： $L_2 = \bar{L}/4$

Average passenger walking distance： $L(m) = 2000/3\delta + \bar{L}/4$.

From statistical data，Passengers walking speed is generally 4km/. The average walking time of the passengers is appropriate in 5-8 minutes;

② Waiting time：Generally no more than 3 minutes；

③ Riding time：Taking into account the size of the city of Shenzhen City, located in the 35-40 minutes is appropriate;

④ Transfer time：Average transfer time is generally not more than 5 minutes.

（3）Evaluation criterion

Table 4-3-7 Transport speed rating scale

Evaluation level	1	2	3	4	5
A7	>25	[16,25）	[13,16)	[10,13)	<10
Index	[90,100]	[80,90)	[70,80)	[60,70)	[0,60)

Table 4-3-8 Passenger travel time scale（min）

Evaluation level	1	2	3	4	5
A8	<50	[60,50）	[70,60)	[80,70)	>80
Index	[90,100]	[80,90)	[70,80)	[60,70)	[0,60)

4.3.2.5 Accurate

（1）Order A9 as Arrival offset. The calculation formula is as follows：

Arrival time offset = Actual arrival time - Expected arrival time

（2）Order A10 as The bus punctuality rate（%）

（3）Evaluation criterion

Table 4-3-9 Arrival time offset rating scale

Evaluation level	1	2	3	4	5
A9	[1,0）	[3,1）	[5,3)	[7,5)	>7
Index	[90,100]	[80,90)	[70,80)	[60,70)	[0,60)

Table 4-3-10 The arrival punctuality rate scale

Evaluation level		1	2	3	4	5
A10	start	[95,100）	[90,95）	[85,90)	[80,85)	<80
	run	[85,90）	[80,85)	[75,80)	[70,75)	<70
	peak	[80,85）	[75,80)	[70,75)	[65,70)	<65
	Index	[90,100]	[80,90)	[70,80)	[60,70)	[0,60)

4.3.2.6 Comfortable

（1）Order A11 as Full-load ratio，Full-load ratio = Actual number of passengers in the car / Full load rating;

（2）Order A12 as Seating rate（%），Seating rate = Car seat number / Actual number of passengers.

（3）Evaluation criterion

Table 4-3-11 Full load rating scale

Evaluation level		1	2	3	4	5
A11	Flat	[55,45)	[65,55)	[75,65)	[85,75)	>85

	peak					
	peak	[100.95）	[105,100)	[110,105)	[115,110)	>115
Index		[90,100]	[80,90)	[70,80)	[60,70)	[0,60)

Table 4-3-12 Rating scale（%）

Evaluation level		1	2	3	4	5
A12	Flat peak	[70,80)	[60,70)	[50,60)	[40,50)	>40
	peak	[38,40）	[36,38)	[34,36)	[32,34)	>34
Index		[90,100]	[80,90)	[70,80)	[60,70)	[0,60)

4.3.3 The introduction of public transport reliability evaluation system

We believe that the level of public transport service is also related to the reliability of the service, and the reliability of public transport can be measured from the following four dimensions:1、Service time reliability 2、Travel time reliability 3、Punctuality reliability 4 Passenger quantity reliability.

4.3.3.1 Service time reliability

Generally speaking, the operation time of the line section and the time of the station are two factors which determine the reliability of the service time of the transportation system. So the service time reliability of each line can be defined by the following:

$$SR_l = P(S_l \le \alpha \le S_{l0}), \forall l \in L$$

S_l Actual service time interval of a certain period of time

α Correction factor

S_{l0} Theoretical service time of a certain period of time

L Traffic line group S_l

GPS and card data can be extrapolated actual observations. Comprehensive Shenzhen Century traffic conditions, we believe that to achieve the three above criteria can be considered reliable service.

4.3.3.2 Travel Time Reliability

Travel time is the time between the OD of the time consuming, so travel time reliability of the public transport system is defined by the following formula: $TTR = P(\bar{t} \le \beta \cdot \bar{t_0})$

$\bar{t}$ OD between the actual average travel times β Correction factor

$\bar{t_0}$ OD to the theoretical average time

Comprehensive Shenzhen's traffic conditions, we believe that the average travel time to reach the three standards is the reliable service. GPS and card data can be extrapolated actual observations. Comprehensive Shenzhen Century traffic conditions, we believe that to achieve the three above criteria can be considered reliable service.

4.3.3.3 Punctuality reliability

Reliability refers to a reference point in the dynamic transportation network; the vehicle can

point to the ability of quasi-off station, expressed by the following formula. $OTR = P\{|\sigma_i| \le \gamma \cdot \sigma_{i0}\}$

σ_i The actual offset vehicle i arrival time

σ_{i0} Vehicle i arrival time is defined offset

γ Correction factor

Comprehensive Shenzhen actual situation, and with reference to A9 grading standards, we believe that to reach the three criteria, namely whichever is 5min, can be identified reliably prevail point

4.3.3.4 Number of passengers Reliability

Comfort level is determined by the number of passengers, so the number of passengers we define this standard of reliability.

Define the following formula： $LR = P\{p_i \le \lambda p_{i0}\}$

p_i Actual passenger number of vehicle i

p_{i0} Limited number of passenger vehicles i

λ Correction factor

Comprehensive Shenzhen traffic conditions, we believe that the number of passengers reached the three criteria is reliable.

Integrated database data and the actual traffic situation, we believe that the first three criteria to measure the amount of a normal distribution, the number of passengers in line with the Poisson distribution.-

4.3.4 System energy efficiency evaluation system

Commonly used system US industry based on energy efficiency ADC model, we have established the urban public transportation system efficiency analysis ADC model that E = ADC, wherein A refers to the system of energy efficiency, A for the transportation system the probability of starting a usable state, in this model, we the value is 1, that the urban public transport system at any time in a complete state; D reliable transport system is running; C represents service capacity transport system.

This can be listed as follows: $E = SR \cdot A3 + TTR \cdot A8 + OTR \cdot A9 + LR \cdot A11$。

4.3.5 Model solving of the Problem three

According to the model, we analyzed the operational capability of Shenzhen total of 1,880 bus stops, bus stops to get a quality of service reports.

We will bus stops named 0001-1880, respectively for each bus station to calculate the system efficiency values obtained reported values of energy efficiency, because the data is too large, we only show part of the theme, the detailed results in the appendix.

Figure 4-3-1 some bus station value of energy efficiency report

Finally we find East Door bus station, Meilong over-bridge bus station and Shiyan Jinbei

village bus station are most improper through comprehensive evaluation. Roads are narrow and passenger flow volume is huge in East Door bus station, which causes traffic jam and influences its systematic efficiency energy value. Passenger flow volume is huge and bus operation distance is long in Meilong over-bridge bus station, which influences systematic efficiency energy value. Bus point distance is extremely large, distance to residential area is long and bus number is less, which can influence systematic efficiency energy value of Shiyan Jinbei village bus station.

5、Model Evaluation and Error Analysis

For a problem, we used three different data mining techniques from different angles data preprocessing, removed a large number of invalid information, and organize effective information consolidation, as well as preliminary statistical analysis, good reach the purpose of the pre-treatment, for the analysis of the following issues save a lot of time, but since the amount of data is too huge and our limited computing power, data processing is still a bit rough, the clustering process, classification principles are not particularly clear, the classification result there is no obvious relevance.

For question two, we analyzed the GPS data and bus IC card data, the establishment of a bus OD matrix, especially the introduction of the concept of traffic zone, making the OD matrix can be greatly simplified analysis speed has greatly improved. Eventually we were analyzing residents travel in time and space for the generated OD matrix, whereby the advice in order to optimize public transport system, a good topic to complete the requirements. However, due to the acquisition of information is only a few days of data, is not necessarily representative, and traffic cell division seems to be rather rough, clustering is not thorough enough, OD matrix is still a bit large.

For question three, we first establish a service quality evaluation system and evaluation system reliability and energy efficiency of the evaluation mechanism introduced to the system, the quality of service of this abstract concept concrete, eventually obtained a good change of bus stops need to fix information. However, many of the present evaluation system evaluation criteria used are reference values, not necessarily very fit the actual situation in Shenzhen, and therefore if they can have more of the monitoring data, the accuracy of the model will be greatly enhanced.

6、Reference material

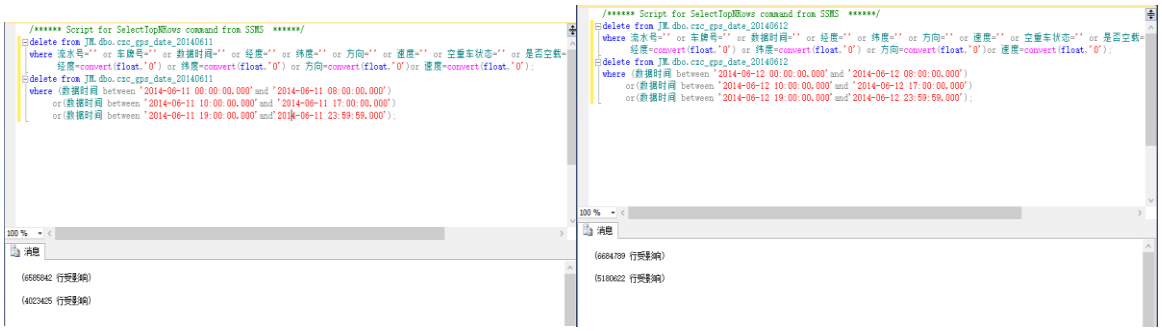
【1】 Jiang Qiyuan, mathematical model，Higher Education Press，August 2011 2nd printing.

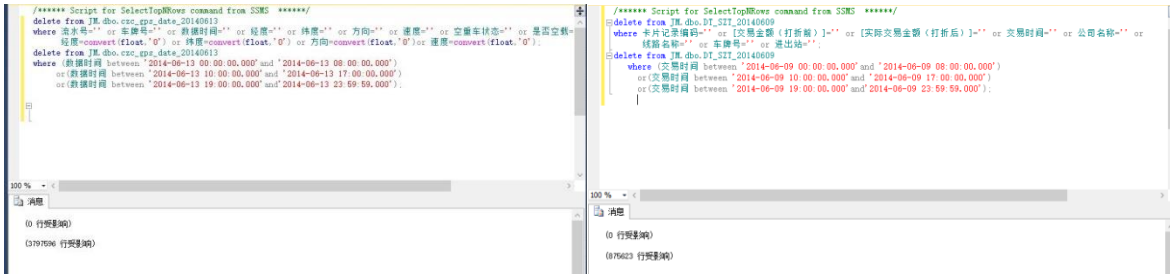
【2】 Si Shoukui，Mathematical modeling algorithm，National Defense University

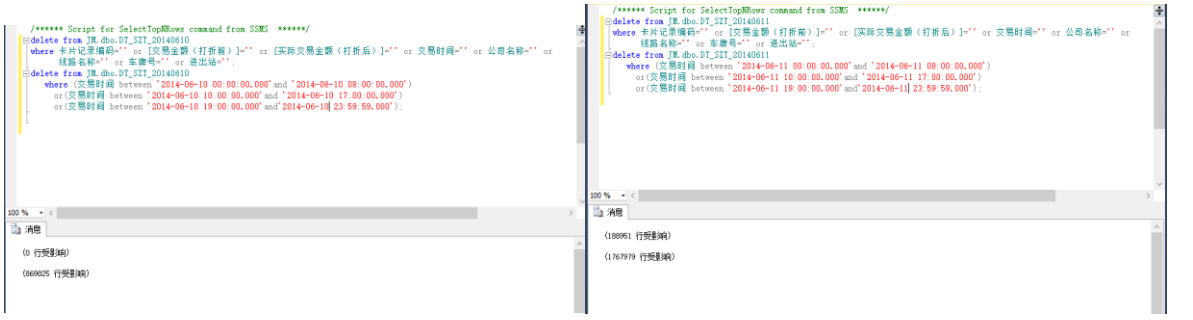
【3】 Press.Abraham Silberschatz， Database System Concepts，Machinery Industry Press，By the end of July 2014 edition

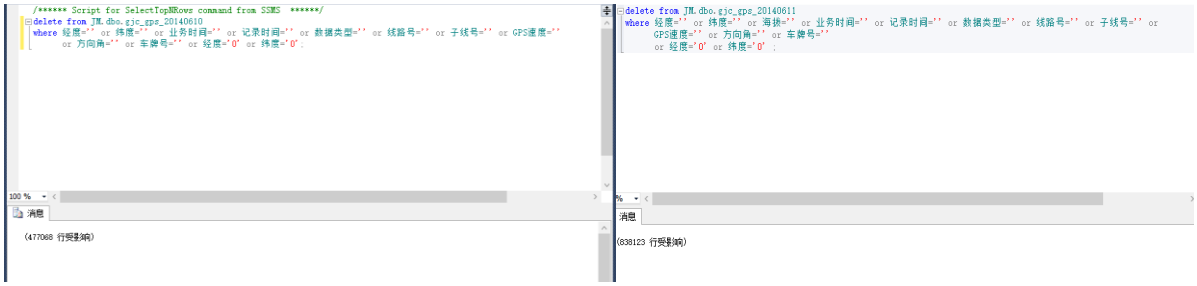
【4】 Zhang Xinghui， Data warehouse and data mining technology，Tsinghua University Press，June 2011 1st edition

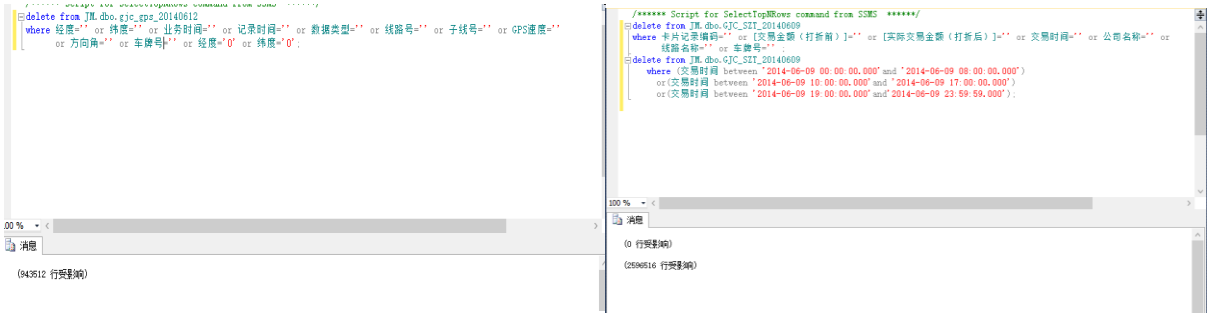
Appendix I: a data preprocessor problem and results

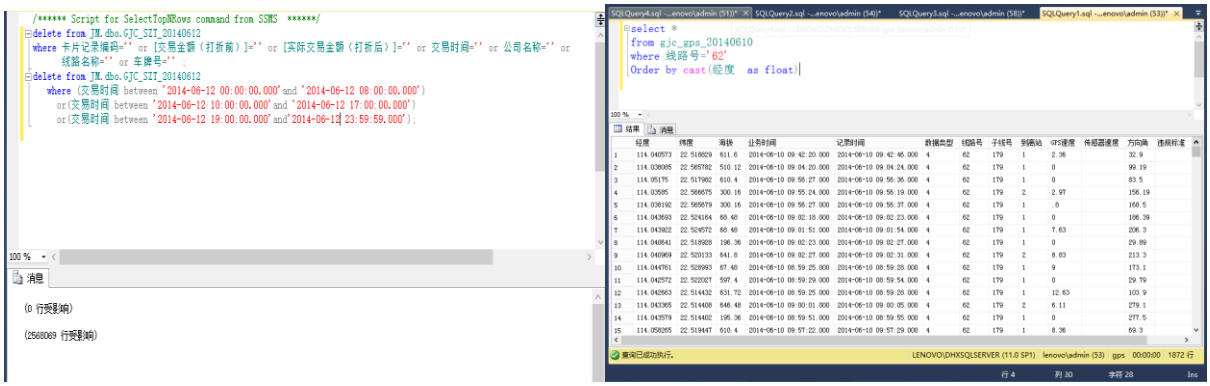


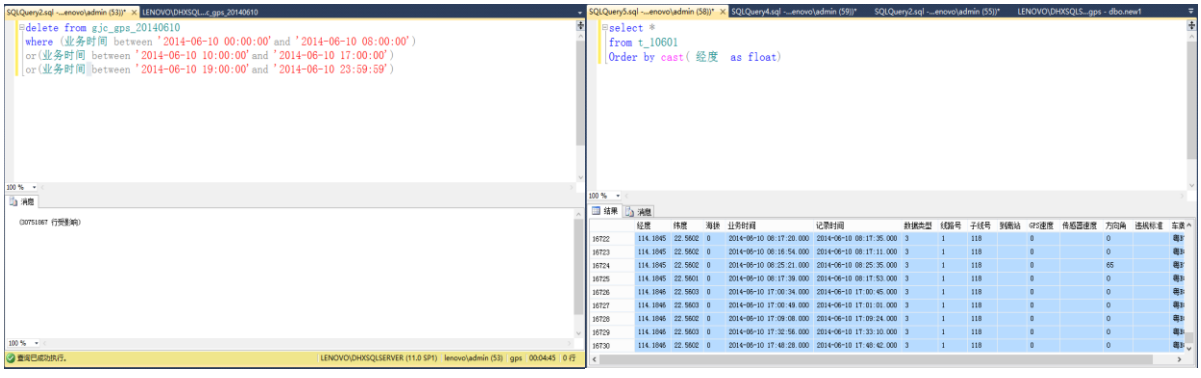








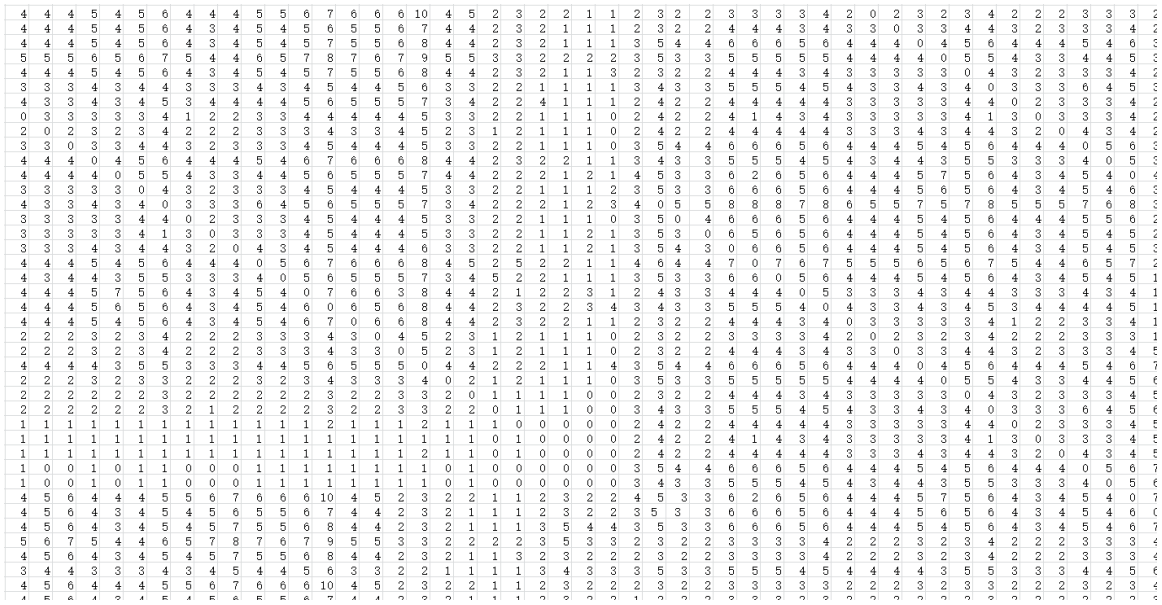


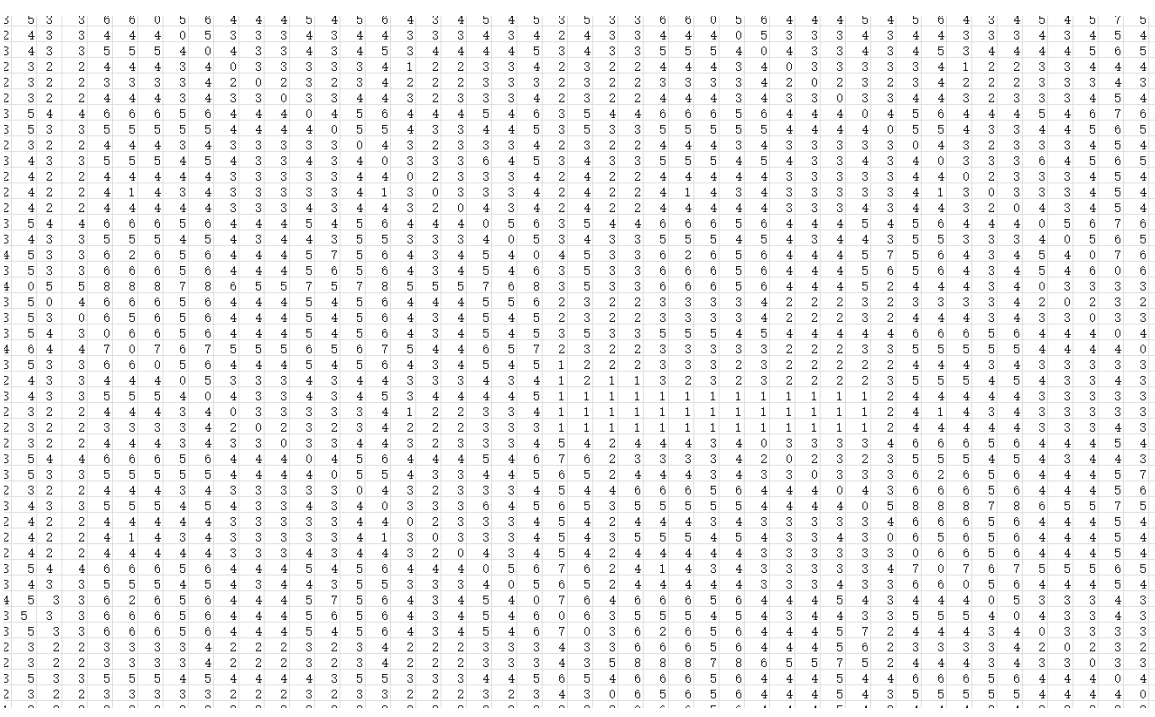


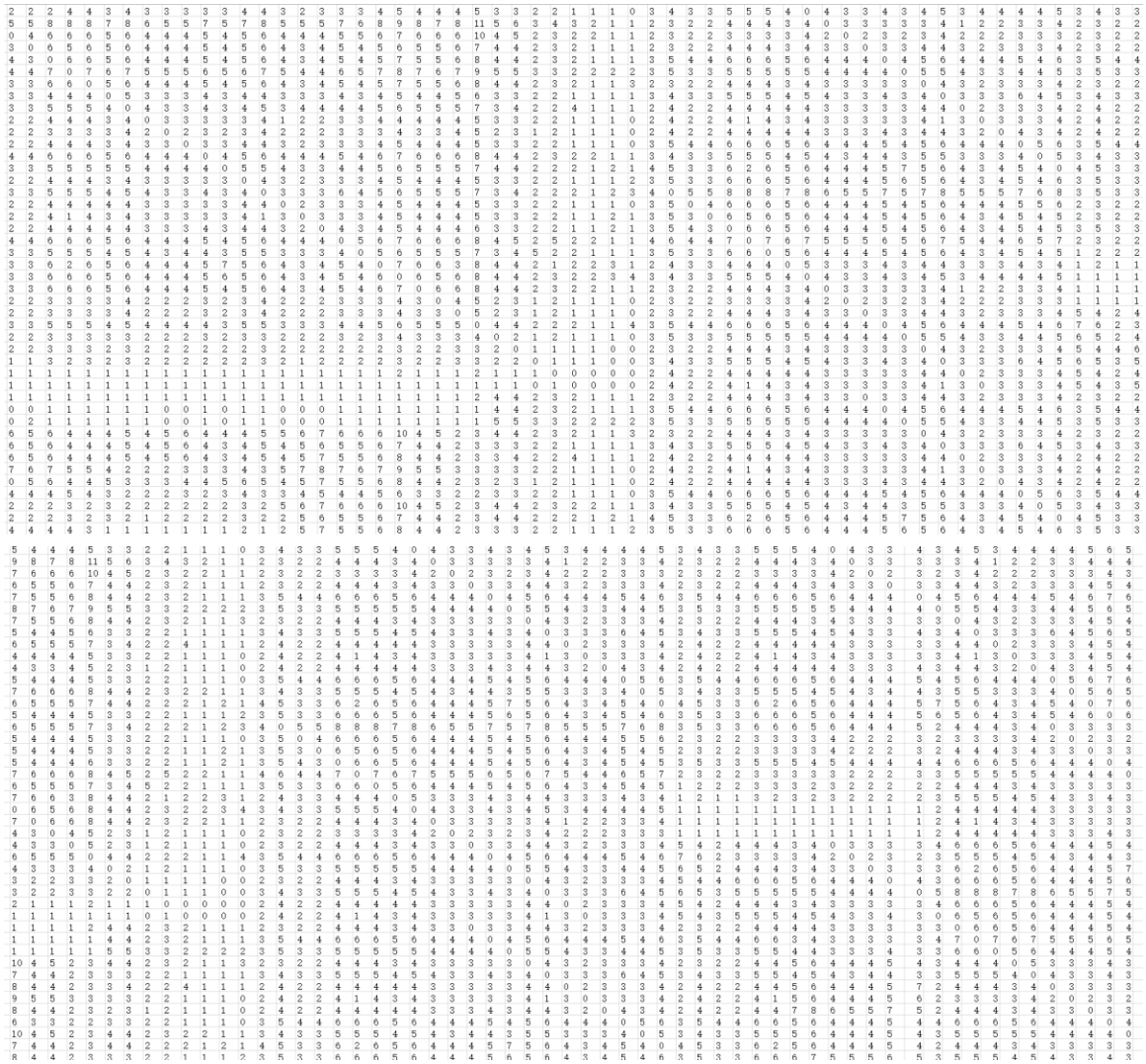
Appendix 2: Bus OD Matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62
1	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	0	0	1	0	1	1	0	0	0	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	1	4	0	5	5	8	8	8	7	8	6	5	5	7	5	7	8	5	5	5	7
2	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	3	5	0	4	6	6	6	5	6	4	4	4	5	4	5	6	4	4	4	5
3	1	0	0	0	0	1	1	1	1	1	1	2	1	2	1	2	1	1	1	1	1	1	2	1	1	1	1	1	1	2	1	1	2	2	1	1	1	1	1	0	0	0	3	5	3	0	6	5	6	5	6	4	4	4	5	4	5	6	4	3	4	5
4	1	0	0	0	0	1	1	1	2	1	1	2	2	2	2	2	1	1	1	2	1	2	2	1	1	1	2	1	2	2	2	2	2	3	1	1	1	1	1	1	1	0	3	5	4	3	0	6	6	5	6	4	4	4	5	4	5	6	4	3	4	5
5	1	0	1	1	0	1	1	2	2	2	2	3	3	3	2	3	2	2	2	2	2	2	3	2	2	2	2	2	3	3	3	3	3	1	2	2	1	1	1	1	1	0	4	6	4	4	7	0	7	6	7	5	5	5	6	5	6	7	5	4	4	6
6	0	1	2	0	1	0	1	1	2	1	1	2	2	1	2	2	2	1	2	2	2	2	2	2	1	2	2	2	2	3	2	2	2	3	2	2	1	2	1	1	1	2	3	5	3	3	6	6	0	5	6	4	4	4	5	4	5	6	4	3	4	5
7	0	1	1	1	2	2	0	2	4	3	3	4	4	4	4	5	3	3	3	4	3	4	4	3	3	3	4	3	4	5	4	4	5	6	3	3	2	2	1	0	1	1	2	4	3	3	4	4	4	0	5	3	3	3	4	3	4	4	3	3	3	4
8	3	0	1	1	1	2	2	0	4	2	2	2	4	4	3	4	3	3	3	3	3	4	4	3	2	3	3	3	4	5	4	4	4	5	3	3	2	2	1	1	1	0	3	4	3	3	5	5	5	4	0	4	3	3	4	3	4	5	3	4	4	4
9	4	1	2	3	2	3	3	4	0	5	5	8	8	8	7	8	6	5	5	7	5	7	8	5	5	5	7	6	8	9	8	7	8	11	5	6	3	4	3	2	1	1	2	3	2	2	4	4	4	3	4	0	3	3	3	3	3	4	1	2	2	3
10	2	1	1	1	1	3	3	3	5	0	4	6	6	6	5	6	4	4	4	5	4	5	6	4	4	4	5	5	6	7	6	6	6	10	4	5	2	3	2	2	1	1	2	3	2	2	3	3	3	3	4	2	0	2	3	2	3	4	2	2	2	3
11	1	1	1	1	1	2	3	3	5	3	0	6	5	6	5	6	4	4	4	5	4	5	6	4	3	4	5	4	5	6	5	5	6	7	4	4	2	3	2	1	1	1	2	3	2	2	4	4	4	3	4	3	3	0	3	3	4	4	3	2	3	3
12	2	1	1	1	3	2	2	3	5	4	3	0	6	6	5	6	4	4	4	5	4	5	6	4	3	4	5	4	5	7	5	5	6	8	4	4	2	3	2	1	1	1	3	5	4	4	6	6	6	5	6	4	4	4	0	4	5	6	4	4	4	5
13	4	1	1	1	2	3	3	4	6	4	4	7	0	7	6	7	5	5	5	6	5	6	7	5	4	4	6	5	7	8	7	6	7	9	5	5	3	3	2	2	2	2	3	5	3	3	5	5	5	5	5	4	4	4	4	0	5	5	4	3	3	4
14	3	1	1	1	1	2	2	3	5	3	3	6	6	0	5	6	4	4	4	5	4	5	6	4	3	4	5	4	5	7	5	5	6	8	4	4	2	3	2	1	1	3	2	3	2	2	4	4	4	3	4	3	3	3	3	3	0	4	3	2	3	3
15	2	1	1	1	1	2	2	2	4	3	3	4	4	4	0	5	3	3	3	4	3	4	4	3	3	3	4	3	4	5	4	4	5	6	3	3	2	2	1	1	1	1	3	4	3	3	5	5	5	4	5	4	3	3	4	3	4	0	3	3	3	6
16	0	1	1	1	1	2	2	3	4	3	3	5	5	5	4	0	4	3	3	4	3	4	5	3	4	4	4	4	5	6	5	5	5	7	3	4	2	2	4	1	1	1	2	4	2	2	4	4	4	4	4	3	3	3	3	3	4	4	0	2	3	3
17	1	0	1	1	1	2	2	2	3	2	2	4	4	4	3	4	0	3	3	3	3	3	4	1	2	2	3	3	4	4	4	4	4	5	3	3	2	2	1	1	1	0	2	4	2	2	4	1	4	3	4	3	3	3	3	3	4	1	3	0	3	3
18	1	3	2	1	1	1	2	2	3	2	2	3	3	3	3	4	2	0	2	3	2	3	4	2	2	2	3	3	3	4	3	3	4	5	2	3	1	2	1	1	1	0	2	4	2	2	4	4	4	4	4	3	3	3	4	3	4	4	3	2	0	4
19	1	0	1	1	1	2	2	2	3	2	2	4	4	4	3	4	3	3	0	3	3	4	4	3	2	3	3	3	4	5	4	4	4	5	3	3	2	2	1	1	1	0	3	5	4	4	6	6	6	5	6	4	4	4	5	4	5	6	4	4	4	0
20	1	2	1	3	2	3	3	3	5	4	4	6	6	6	5	6	4	4	4	0	4	5	6	4	4	4	5	4	6	7	6	6	6	8	4	4	2	3	2	2	1	1	3	4	3	3	5	5	5	4	5	4	3	4	4	3	5	5	3	3	3	4
21	3	1	1	1	2	2	2	3	5	3	3	5	5	5	5	5	4	4	4	4	0	5	5	4	3	3	4	4	5	6	5	5	5	7	4	4	2	2	2	1	2	1	4	5	3	3	6	2	6	5	6	4	4	4	5	7	5	6	4	3	4	5
22	0	0	1	1	1	2	2	2	3	2	2	4	4	4	3	4	3	3	3	3	3	0	4	3	2	3	3	3	4	5	4	4	4	5	3	3	2	2	1	1	1	2	3	5	3	3	6	6	6	5	6	4	4	4	5	6	5	6	4	3	4	5
23	1	1	2	1	1	2	2	3	4	3	3	5	5	5	4	5	4	3	3	4	3	4	0	3	3	3	6	4	5	6	5	5	5	7	3	4	2	2	2	1	2	3	4	0	5	5	8	8	8	7	8	6	5	5	7	5	7	8	5	5	5	7
24	2	0	3	1	1	2	2	2	4	2	2	4	4	4	4	4	3	3	3	3	3	4	4	0	2	3	3	3	4	5	4	4	4	5	3	3	2	2	1	1	1	0	3	5	0	4	6	6	6	5	6	4	4	4	5	4	5	6	4	4	4	5
25	5	0	1	1	1	2	2	2	4	2	2	4	1	4	3	4	3	3	3	3	3	4	1	3	0	3	3	3	4	5	4	4	4	5	3	3	2	2	1	1	2	1	3	5	3	0	6	5	6	5	6	4	4	4	5	4	5	6	4	3	4	5
26	0	1	1	1	1	2	2	2	4	2	2	4	4	4	4	4	3	3	3	4	3	4	4	3	2	0	4	3	4	5	4	4	4	6	3	3	2	2	1	1	2	1	3	5	4	3	0	6	6	5	6	4	4	4	5	4	5	6	4	3	4	5
27	3	1	1	1	1	3	3	3	5	4	4	6	6	6	5	6	4	4	4	5	4	5	6	4	4	4	0	5	6	7	6	6	6	8	4	5	2	5	2	2	1	1	4	6	4	4	7	0	7	6	7	5	5	5	6	5	6	7	5	4	4	6
28	0	1	2	1	3	2	2	3	4	3	3	5	5	5	4	5	4	3	4	4	3	5	5	3	3	3	4	0	5	6	5	5	5	7	3	4	5	2	2	1	1	1	3	5	3	3	6	6	0	5	6	4	4	4	5	4	5	6	4	3	4	5
29	2	6	1	1	3	3	3	4	5	3	3	6	2	6	5	6	4	4	4	5	7	5	6	4	3	4	5	4	0	7	6	6	3	8	4	4	2	1	2	2	3	1	2	4	3	3	4	4	4	0	5	3	3	3	4	3	4	4	3	3	3	4
30	0	1	1	1	1	2	3	3	5	3	3	6	6	6	5	6	4	4	4	5	6	5	6	4	3	4	5	4	6	0	6	5	6	8	4	4	2	3	2	2	3	4	3	4	3	3	5	5	5	4	0	4	3	3	4	3	4	5	3	4	4	4
31	3	1	2	3	1	3	3	3	5	3	3	6	6	6	5	6	4	4	4	5	4	5	6	4	3	4	5	4	6	7	0	6	6	8	4	4	2	3	2	2	1	1	2	3	2	2	4	4	4	3	4	0	3	3	3	3	3	4	1	2	2	3
32	2	1	1	4	1	1	2	2	3	2	2	3	3	3	3	4	2	2	2	3	2	3	4	2	2	2	3	3	3	4	3	0	4	5	2	3	1	2	1	1	1	0	2	3	2	2	3	3	3	3	4	2	0	2	3	2	3	4	2	2	2	3
33	4	3	1	1	1	1	2	2	3	2	2	3	3	3	3	4	2	2	2	3	2	3	4	2	2	2	3	3	3	4	3	3	0	5	2	3	1	2	1	1	1	0	2	3	2	2	4	4	4	3	4	3	3	0	3	3	4	4	3	2	3	3
34	5	2	1	1	1	2	2	3	5	3	3	5	5	5	4	5	4	4	4	4	3	5	5	3	3	3	4	4	5	6	5	5	5	0	4	4	2	2	2	1	1	4	3	5	4	4	6	6	6	5	6	4	4	4	0	4	5	6	4	4	4	5
35	3	1	1	1	1	1	1	2	3	2	2	3	3	3	3	3	2	2	2	3	2	3	3	2	2	2	3	2	3	4	3	3	3	4	0	2	1	2	1	1	1	0	3	5	3	3	5	5	5	5	5	4	4	4	4	0	5	5	4	3	3	4
36	1	0	1	0	1	1	1	1	2	2	2	3	3	3	2	3	2	2	2	2	2	2	3	2	2	2	2	2	2	3	2	2	3	3	2	0	1	1	1	1	0	0	2	3	2	2	4	4	4	3	4	3	3	3	3	3	0	4	3	2	3	3
37	1	1	0	0	1	1	1	1	2	1	1	3	2	3	2	3	2	2	2	2	2	2	3	2	1	2	2	2	2	3	2	2	3	3	2	2	0	1	1	1	0	0	3	4	3	3	5	5	5	4	5	4	3	3	4	3	4	0	3	3	3	6
38	2	3	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	0	0	0	0	0	2	4	2	2	4	4	4	4	4	3	3	3	3	3	4	4	0	2	3	3
39	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	2	4	2	2	4	1	4	3	4	3	3	3	3	3	4	1	3	0	3	3
40	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	0	1	0	0	0	0	2	4	2	2	4	4	4	4	4	3	3	3	4	3	4	4	3	2	0	4
41	1	0	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	0	0	1	0	1	1	0	0	0	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	3	5	4	4	6	6	6	5	6	4	4	4	5	4	5	6	4	4	4	0
42	1	1	0	2	0	0	0	0	1	0	2	1	1	1	1	1	1	0	0	1	0	1	1	0	0	0	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	3	4	3	3	5	5	5	4	5	4	3	4	4	3	5	5	3	3	3	4
43	1	3	3	3	5	0	4	6	6	6	5	6	4	4	4	5	4	5	6	4	4	4	5	5	6	7	6	6	6	10	4	5	2	3	2	2	1	1	2	3	2	2	4	5	3	3	6	2	6	5	6	4	4	4	5	7	5	6	4	3	4	5

1	1	2	1	1	2	2	1	2	2	2	1	2	2	2	2	2	2	1	2	2	2	2	3	2	2	2	3	2	2	1	2	1	1	1	2	3	5	3	3	6	6	0	5	6	4	4	4	5	4	5	6
0	2	4	3	3	4	4	4	4	5	3	3	3	4	3	4	4	3	3	3	4	3	4	5	4	4	5	6	3	3	2	2	1	0	1	1	2	4	3	3	4	4	4	0	5	3	3	3	4	3	4	4
2	0	4	2	2	2	4	4	3	4	3	3	3	3	3	4	4	3	2	3	3	3	4	5	4	4	4	5	3	3	2	2	1	1	1	0	3	4	3	3	5	5	5	4	0	4	3	3	4	3	4	5
3	4	0	5	5	8	8	8	7	8	6	5	5	7	5	7	8	5	5	5	7	6	8	9	8	7	8	11	5	6	3	4	3	2	1	1	2	3	2	2	4	4	4	3	4	0	3	3	3	3	3	4
3	3	5	0	4	6	6	6	5	6	4	4	4	5	4	5	6	4	4	4	5	5	6	7	6	6	6	10	4	5	2	3	2	2	1	1	2	3	2	2	3	3	3	3	4	2	0	2	3	2	3	4
3	3	5	3	0	6	5	6	5	6	4	4	4	5	4	5	6	4	3	4	5	4	5	6	5	5	6	7	4	4	2	3	2	1	1	1	2	3	2	2	4	4	4	3	4	3	3	0	3	3	4	4
2	3	5	4	3	0	6	6	5	6	4	4	4	5	4	5	6	4	3	4	5	4	5	7	5	5	6	8	4	4	2	3	2	1	1	1	3	5	4	4	6	6	6	5	6	4	4	4	0	4	5	6
3	4	6	4	4	7	0	7	6	7	5	5	5	6	5	6	7	5	4	4	6	5	7	8	7	6	7	9	5	5	3	3	2	2	2	2	3	5	3	3	5	5	5	5	5	4	4	4	4	0	5	5
2	3	5	3	3	6	6	0	5	6	4	4	4	5	4	5	6	4	3	4	5	4	5	7	5	5	6	8	4	4	2	3	2	1	1	3	2	3	2	2	4	4	4	3	4	3	3	3	3	3	0	4
2	2	4	3	3	4	4	4	0	5	3	3	3	4	3	4	4	3	3	3	4	3	4	5	4	4	5	6	3	3	2	2	1	1	1	1	3	4	3	3	5	5	5	4	5	4	3	3	4	3	4	0
2	3	4	3	3	5	5	5	4	0	4	3	3	4	3	4	5	3	4	4	4	4	5	6	5	5	5	7	3	4	2	2	4	1	1	1	2	4	2	2	4	4	4	4	4	3	3	3	3	3	4	4
2	2	3	2	2	4	4	4	3	4	0	3	3	3	3	3	4	1	2	2	3	3	4	4	4	4	4	5	3	3	2	2	1	1	1	0	2	4	2	2	4	1	4	3	4	3	3	3	3	3	4	1
2	2	3	2	2	3	3	3	3	4	2	0	2	3	2	3	4	2	2	2	3	3	3	4	3	3	4	5	2	3	1	2	1	1	1	0	2	4	2	2	4	4	4	4	4	3	3	3	4	3	4	4
2	2	3	2	2	4	4	4	3	4	3	3	0	3	3	4	4	3	2	3	3	3	4	5	4	4	4	5	3	3	2	2	1	1	1	0	3	5	4	4	6	6	6	5	6	4	4	4	5	4	5	6
3	3	5	4	4	6	6	6	5	6	4	4	4	0	4	5	6	4	4	4	5	4	6	7	6	6	6	8	4	4	2	3	2	2	1	1	3	4	3	3	5	5	5	4	5	4	3	4	4	3	5	5
2	3	5	3	3	5	5	5	5	5	4	4	4	4	0	5	5	4	3	3	4	4	5	6	5	5	5	7	4	4	2	2	2	1	2	1	4	5	3	3	6	2	6	5	6	4	4	4	5	7	5	6
2	2	3	2	2	4	4	4	3	4	3	3	3	3	3	0	4	3	2	3	3	3	4	5	4	4	4	5	3	3	2	2	1	1	1	2	3	5	3	3	6	6	6	5	6	4	4	4	5	6	5	6
2	3	4	3	3	5	5	5	4	5	4	3	3	4	3	4	0	3	3	3	6	4	5	6	5	5	5	7	3	4	2	2	2	1	2	3	4	0	5	5	8	8	8	7	8	6	5	5	7	5	7	8
2	2	4	2	2	4	4	4	4	4	3	3	3	3	3	4	4	0	2	3	3	3	4	5	4	4	4	5	3	3	2	2	1	1	1	0	3	5	0	4	6	6	6	5	6	4	4	4	5	4	5	6
2	2	4	2	2	4	1	4	3	4	3	3	3	3	3	4	1	3	0	3	3	3	4	5	4	4	4	5	3	3	2	2	1	1	2	1	3	5	3	0	6	5	6	5	6	4	4	4	5	4	5	6
2	2	4	2	2	4	4	4	4	4	3	3	3	4	3	4	4	3	2	0	4	3	4	5	4	4	4	6	3	3	2	2	1	1	2	1	3	5	4	3	0	6	6	5	6	4	4	4	5	4	5	6
3	3	5	4	4	6	6	6	5	6	4	4	4	5	4	5	6	4	4	4	0	5	6	7	6	6	6	8	4	5	2	5	2	2	1	1	4	6	4	4	7	0	7	6	7	5	5	5	6	5	6	7
2	3	4	3	3	5	5	5	4	5	4	3	4	4	3	5	5	3	3	3	4	0	5	6	5	5	5	7	3	4	5	2	2	1	1	1	3	5	3	3	6	6	0	5	6	4	4	4	5	4	5	6
3	4	5	3	3	6	2	6	5	6	4	4	4	5	7	5	6	4	3	4	5	4	0	7	6	6	3	8	4	4	2	1	2	2	3	1	2	4	3	3	4	4	4	0	5	3	3	3	4	3	4	4
3	3	5	3	3	6	6	6	5	6	4	4	4	5	6	5	6	4	3	4	5	4	6	0	6	5	6	8	4	4	2	3	2	2	3	4	3	4	3	3	5	5	5	4	0	4	3	3	4	3	4	5
3	3	5	3	3	6	6	6	5	6	4	4	4	5	4	5	6	4	3	4	5	4	6	7	0	6	6	8	4	4	2	3	2	2	1	1	2	3	2	2	4	4	4	3	4	0	3	3	3	3	3	4
2	2	3	2	2	3	3	3	3	4	2	2	2	3	2	3	4	2	2	2	3	3	3	4	3	0	4	5	2	3	1	2	1	1	1	0	2	3	2	2	3	3	3	3	4	2	0	2	3	2	3	4
2	2	3	2	2	3	3	3	3	4	2	2	2	3	2	3	4	2	2	2	3	3	3	4	3	3	0	5	2	3	1	2	1	1	1	0	2	3	2	2	4	4	4	3	4	3	3	0	3	3	4	4
2	3	5	3	3	5	5	5	4	5	4	4	4	4	3	5	5	3	3	3	4	4	5	6	5	5	5	0	4	4	2	2	2	1	1	4	3	5	4	4	6	6	6	5	6	4	4	4	0	4	5	6
1	2	3	2	2	3	3	3	3	3	2	2	2	3	2	3	3	2	2	2	3	2	3	4	3	3	3	4	0	2	1	2	1	1	1	0	3	5	3	3	5	5	5	5	5	4	4	4	4	0	5	5
1	1	2	2	2	3	3	3	2	3	2	2	2	2	2	2	3	2	2	2	2	2	2	3	2	2	3	3	2	0	1	1	1	1	0	0	2	3	2	2	4	4	4	3	4	3	3	3	3	3	0	4
1	1	2	1	1	3	2	3	2	3	2	2	2	2	2	2	3	2	1	2	2	2	2	3	2	2	3	3	2	2	0	1	1	1	0	0	3	4	3	3	5	5	5	4	5	4	3	3	4	3	4	0
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	0	0	0	0	0	2	4	2	2	4	4	4	4	4	3	3	3	3	3	4	4
0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	2	4	2	2	4	1	4	3	4	3	3	3	3	3	4	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	0	1	0	0	0	0	2	4	2	2	4	4	4	4	4	3	3	3	4	3	4	4
0	0	1	0	0	1	1	1	1	1	1	0	0	1	0	1	1	0	0	0	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	3	5	4	4	6	6	6	5	6	4	4	4	5	4	5	6
0	0	1	0	2	1	1	1	1	1	1	0	0	1	0	1	1	0	0	0	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	3	4	3	3	5	5	5	4	5	4	3	4	4	3	5	5
4	6	6	6	5	6	4	4	4	5	4	5	6	4	4	4	5	5	6	7	6	6	6	10	4	5	2	3	2	2	1	1	2	3	2	2	4	5	3	3	6	2	6	5	6	4	4	4	5	7	5	6
0	6	5	6	5	6	4	4	4	5	4	5	6	4	3	4	5	4	5	6	5	5	6	7	4	4	2	3	2	1	1	1	2	3	2	2	3	5	3	3	6	6	6	5	6	4	4	4	5	6	5	6
3	0	6	6	5	6	4	4	4	5	4	5	6	4	3	4	5	4	5	7	5	5	6	8	4	4	2	3	2	1	1	1	3	5	4	4	3	5	3	3	6	6	6	5	6	4	4	4	5	4	5	6
4	7	0	7	6	7	5	5	5	6	5	6	7	5	4	4	6	5	7	8	7	6	7	9	5	5	3	3	2	2	2	2	3	5	3	3	2	3	2	2	3	3	3	3	4	2	2	2	3	2	3	4
3	6	6	0	5	6	4	4	4	5	4	5	6	4	3	4	5	4	5	7	5	5	6	8	4	4	2	3	2	1	1	3	2	3	2	2	2	3	2	2	3	3	3	3	4	2	2	2	3	2	3	4
3	4	4	4	0	5	3	3	3	4	3	4	4	3	3	3	4	3	4	5	4	4	5	6	3	3	2	2	1	1	1	1	3	4	3	3	3	5	3	3	5	5	5	4	5	4	4	4	4	3	5	5
4	6	6	6	5	6	4	4	4	5	4	5	6	4	4	4	5	5	6	7	6	6	6	10	4	5	2	3	2	2	1	1	2	3	2	2	2	3	2	2	3	3	3	3	3	2	2	2	3	2	3	3







Appendix 3: Transportation cell division