Advances in Social Science, Education and Humanities Research, volume 72

International Conference on Management, Education and Social Science (ICMESS 2017)

A Descriptive Statistic Analysis Based on the Operation of Public Bicycle in Shenzhen

Yuxin Yang, Xinyi Zeng, Ling Lv, Yali Peng

Software College of Jiangxi Normal University

Nanchang, China

***Abstract*—Currently, traveling by bike has been widely accepted by the public. However, the public bicycle system is in need of an optimization scheme owing to its poor operating efficiency. In this paper, real operating data of Shenzhen city public bicycle system during July 2016 are used as the data bases, all of which were analyzed by EXCEL, SPSS and RStudio. After the analysis, the following operating rules are found: there are morning and evening peaks in bike borrowing and return; change of bike lending and return flow during same period or at same outlet type of valid date is slight; the public bike flow at varied outlet types differ in workdays and holiday; large number of users lend bike once a week in workdays, while some users lend bike three or more times a week. To improve the demand of public bicycle, frequent users should be delved deeper. The research results may provide a theoretical basis and ideas for the operation optimization of public bicycle system.**

Keywords—Statistics; Public bicycle; Regression analysis; Running law

I. INTRODUCTION

At present, the problem of traffic congestion is becoming increasingly severe, [1] and environmental problems, like air quality and PM2.5, [2] are also affecting people's lives. Traveling by bike has the advantage of low-carbon, environmental, convenient and efficient. Therefore, it is more and more accepted by people. At the same time, local government departments established and improved the public bicycle system for encouraging people to travel by bike. [3]But in the usage of public bicycles, the citizens are often troubled by problems --- lack of bikes and piles.

In this thesis, data of borrowing and return of Shenzhen public bicycle will be analyzed by means of the descriptive statistic method. [4] This paper attempts to attain the basic law of the borrowing and return of bike, thereby providing a theoretical basis and ideas for the optimization of public bicycle system.

II. THE SOURCES AND ORGANIZATION OF DATA

From the real Shenzhen public bicycle operators did the data of this paper come. The data include the date and place of lending, and the stock and capacity of bicycles outlet. All outlets consist of four categories: residential area, common service area, and transportation hub and business center. In which, the common service area includes public services area, government agency and industrial park; commercial center include Business Center, leisure and entertainment complex. Therefore, the four categories can be divided into seven categories.

To understand the operation status of Shenzhen public bicycle system from a macro perspective, we must find the best indicator that can reflect the overall level. We leave the bicycle outlet type first, and add up the times of borrowing and return of bikes respectively. In workday and holiday, users' travel rule diverges greatly, so it's necessary to analyze them respectively. Here, we will analyze the data of workday in detail mainly, and the same analysis method will be applied to analyze holiday's data. The table 1 contains data of Shenzhen public bicycle in five consecutive working days. They will be used in analysis of the time distribution of public bicycle traffic in Shenzhen:

TABLE I USE TIME DISTRIBUTION OF PUBLIC BICYCLE IN SHENZHEN

Borrowing situation	rent	return	rent	return	rent	return	rent	return	rent	return	rent	return
0:00—1:00	26	34	20	27	21	31	25	32	23	34	115	158
1:00—2:00	8	9	7	9	8	9	1	6	7	13	31	46
2:00—3:00	2	3	1	2	2	3	4	2	4	6	13	16
3:00—4:00	2	2	5	4	2	3	1	1	2	1	12	11
4:00—5:00	4	1	8	4	8	3	3	5	8	4	31	17
5:00—6:00	40	18	43	16	32	16	54	19	49	19	218	88
6:00—7:00	202	127	150	106	122	71	170	120	179	121	823	545
7:00—8:00	461	409	469	368	327	270	491	395	462	406	2210	1848

TABLE 1, cont.												
8:00—9:00	521	560	547	601	475	484	505	565	525	531	2573	2741
9:00—10:00	257	276	279	297	260	237	286	295	281	284	1363	1389
10:00—11:00	205	223	204	234	162	216	200	225	189	243	960	1141
11:00—12:00	175	183	131	148	139	161	157	149	139	163	741	804
12:00—13:00	154	153	136	140	94	99	134	148	150	145	668	685
13:00—14:00	117	129	131	124	111	103	138	122	151	131	648	609
14:00—15:00	149	141	170	156	175	154	191	195	178	185	863	831
15:00—16:00	194	163	192	159	181	174	153	150	164	170	884	816
16:00—17:00	279	221	238	228	226	191	225	212	208	150	1176	1002
17:00—18:00	338	345	298	283	324	296	333	284	331	313	1624	1521
18:00—19:00	418	421	485	469	441	458	515	489	475	489	2334	2326
19:00—20:00	388	411	339	385	327	362	356	404	337	366	1747	1928
20:00—21:00	233	268	240	255	217	250	287	311	246	250	1223	1334
21:00—22:00	190	220	187	231	189	204	172	230	166	211	904	1096
22:00—23:00	140	170	113	138	105	143	122	143	150	167	630	761
23:00—24:00	40	53	60	80	41	53	42	50	42	73	225	309
Total	4543	4540	4453	4464	3989	3991	4565	4552	4466	4475	22016	22022

In order to understand the space distribution of public bicycle flow in Shenzhen further, we need to analyze the flow direction of the data. We analyze the basic flow through the statistics of public bike outlet types. The table 2 below is the borrowing and return data of Shenzhen public bicycle in five consecutive workdays, which will be used to analyze the space distribution of public bicycle traffic in Shenzhen city:

TABLE II SPACE DISTRIBUTION OF PUBLIC BICYCLE IN SHENZHEN

Date	**2016-7-4**		**2016-7-5**		**2016-7-6**		**2016-7-7**		**2016-7-8**		**Total**	
Borrowing situation	rent	borrow	rent	borrow	rent	borrow	rent	borrow	rent	borrow	rent	borrow
industrial park	543	581	555	561	561	536	544	496	533	535	2736	2709
public service	633	596	639	592	459	489	629	589	588	646	2948	2912
Traffic hub	835	808	807	840	730	769	809	843	834	831	4015	4091
Commercial Centre	353	336	359	350	273	281	374	378	337	331	1696	1676
Living quarters	1001	1101	946	925	867	867	1053	1104	1046	982	4913	4979
Recreation & Entertainment	167	152	156	161	169	145	162	174	181	167	835	799
Government office	863	839	834	870	803	782	821	796	794	828	4115	4115
Unknown	148	127	157	165	127	122	173	172	153	155	758	741
total	4543	4540	4453	4464	3989	3991	4565	4552	4466	4475	22016	22022

III. ANALYSIS OF SHENZHEN PUBLIC BICYCLE FLOW

From the data sheet above, we can know the conditions of Shenzhen public bicycle system in different periods at outlet of varied types. To achieve the goal of analyzing and understanding the change of bicycle traffic at outlet of varied types during different period, we will organize the data with EXCEL, SPSS and RStudio, and then use descriptive analysis and regression analysis [5-6] to analyze the bicycle traffic distribution rules of Shenzhen city public bicycle system. Also, we will analyze and describe operation feature of Shenzhen bicycle system from the view of travel time of user and types

of outlet. These spatial analysis methods are the most basic and commonly used statistical analysis methods.

A. Descriptive analysis on the distribution law of the total flow of bicycles

The data in table 1 show that the borrowing data are basically balance in varied period of different workdays, which can reflect the time distribution feature of public bicycle system under current circumstances. To show the characteristics more intuitively, we make two line charts, using the lending and return data of workdays in July, as shown in figure 1 and figure 2 respectively:

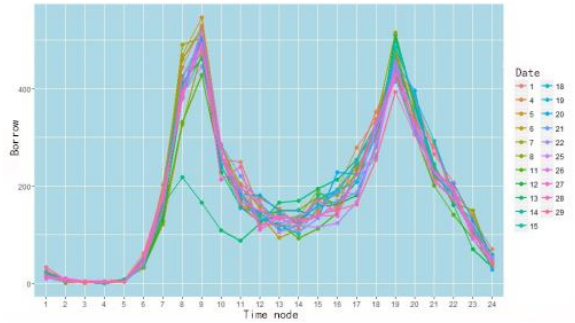

Fig. 1. Statistics of the number of bikes lending in July

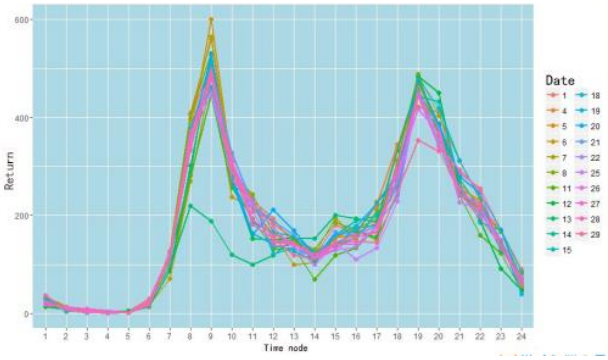

Fig. 2. Statistics of the number of bikes return in July

From the above line chart, we can clearly see that the borrowing and return trend of public bicycle in workdays of July is basically the same. There is a peak around 9:00, and another peak around 19:00.

In the meantime, using the data in table 1, we can calculate the average value, standard deviation and coefficient of variation of bike lending and return in Shenzhen city as follow table 3:

TABLE III PARAMETERS OF BIKE LENDING FLOW IN EACH PERIOD OF SHENZHEN BICYCLE SYSTEM

Date	**0:00 — 1:00**	**1:00— 2:00**	**2:00— 3:00**	**3:00— 4:00**	**4:00— 5:00**	**5:00— 6:00**	**6:00— 7:00**	**7:00— 8:00**	**8:00— 9:00**	**9:00— 10:00**	**10:00 — 11:00**	**11:00 — 12:00**	**TOTAL**
mean value	23	6.2	2.6	2.4	6.2	43.6	164.6	442	514.6	272.6	192	148.2	
standard deviation	2.550	2.950	1.342	1.517	2.490	8.444	30.262	65.414	26.736	13.164	17.930	17.754	
Variation coefficient	0.111	0.476	0.516	0.632	0.402	0.194	0.184	0.148	0.052	0.048	0.093	0.120	
Date	12:00 — 13:00	13:00 — 14:00	14:00 — 15:00	15:00 — 16:00	16:00 — 17:00	17:00 — 18:00	18:00 — 19:00	19:00 — 20:00	20:00 — 21:00	21:00 — 22:00	22:00 — 23:00	23:00 — 24:00	
mean value	133.6	129.6	172.6	176.8	235.2	324.8	466.8	349.4	244.6	180.8	126	45	4403.2
standard deviation	23.766	16.087	15.307	17.852	26.715	15.802	37.976	23.965	26.063	11.032	18.695	8.426	236.489
Variation coefficient	0.178	0.124	0.089	0.101	0.114	0.049	0.081	0.069	0.107	0.061	0.148	0.187	0.054

On the one hand, according to the average value, we can know that the total amount of bicycle lending is about 4403 times a day, and during 8:00 - 9:00, the bicycle lending was significantly more frequent than other periods. On the other hand, we can come to the conclusion that the coefficient of variation in different period of a day is 0.054, which means that the change of bike lending in varied period is slight. Also, we find that the traffic flow parameters own the same feature. Thus, we can draw the conclusion that user's bike time and behavior habits are daily cycled, and the total amount of each hour is basically fixed [7].

B. Analysis on flow direction between bicycle outlets

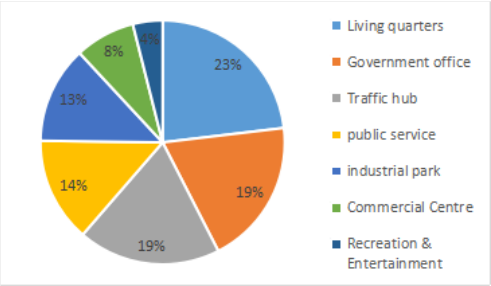

Fig. 3. The statistics of the bike lending outlets

We make the following pie chart, as is shown in figure 3, so as to understand users' flow between outlets of different types more intuitively.

From the above chart, we can easily know that the proportion of lending from residential area is the largest, while the proportion of lending from other sites are relatively small. This phenomenon can provide us with the site planning basis for the initial layout of the bike outlets.

Besides, using the data of table 2, we can attain the mean value, standard deviation and coefficient of variation of bike lending and return in varied outlets as follows table 4 and table 5:

TABLE IV BICYCLE LENDING FLOW PARAMETERS OF SHENZHEN PUBLIC BICYCLE SYSTEM

Dot type	industrial park	public service	Traffic hub	Commercial Centre	Living quarters	Recreation & Entertainment	Government office	Unknown	total
mean value	547.2	589.6	803	339.2	982.6	167	823	151.6	4403.2
standard deviation	10.964	75.722	42.913	39.309	77.487	9.301	27.230	16.637	236.489
Variation coefficient	0.020	0.128	0.053	0.116	0.079	0.056	0.033	0.110	0.054

TABLE V FLOW PARAMETERS OF SHENZHEN PUBLIC BICYCLE SYSTEM

Dot type	industrial park	public service	Traffic hub	Commercial Centre	Living quarters	Recreation & Entertainment	Government office	Unknown	total
mean value	541.8	582.4	818.2	335.2	995.8	159.8	823	148.2	4404.4
standard deviation	31.948	57.204	30.736	35.379	105.554	11.563	35.000	22.532	234.312
Variation coefficient	0.059	0.098	0.038	0.106	0.106	0.072	0.043	0.152	0.053

On the one hand, according to the average value, we can see that the total lending amount is about 4403 times per day, and the returning amount is about 4404 bikes per day, besides, the lending and return amount in residential area outlets are the highest. On the other hand, the variation coefficient of lending and return of bicycle is around 0.1 --- the minimum value is 0.02, and the maximum value is 0.152 --- which means the lending and return amount at various outlets all day is basically the same every day.

In order to analyze the user flow between different outlets in detail, we make a pie chart based on the data of user flow between outlets of varied types in five consecutive days of the first week of July 2016. The pie chart is as follows figure 4:

Obviously, the largest flow emerges between government agencies in working days, it's about 10.99%, followed by residential area - residential area (about 8.29%), transportation hub - transportation hub (about 5.62%), industrial park - industrial park (about 5.11%) and traffic hub - residential area (about 5.01%), while the flow amount between other outlets are relatively small.

Similarly, the distribution feature of flow between different outlets in holidays can be attained, basing on the data of all holidays in July. The characteristic is different from that of workdays: during holidays, the largest flow appears in residential area - residential area, which accounts for 10.25%. Followed by government agency - government agency (about 8.64%), transportation hub - transportation hub (about 6.64%), industrial park - industrial park (about 6.40%). The flow amount of users between other outlets are relatively small.

Other traffic between users is relatively small.

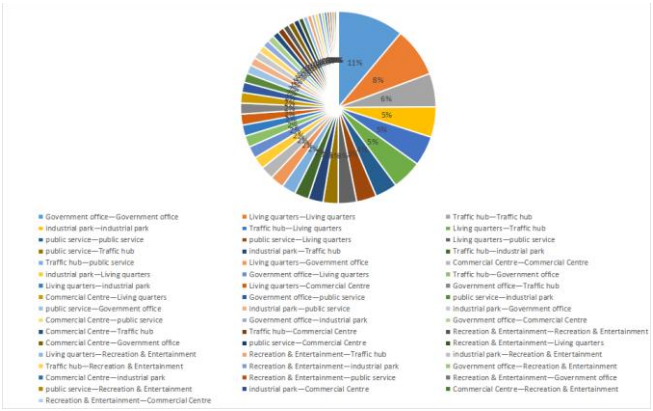

Fig. 4. Flow distribution of public bicycles among varied types of outlet in Shenzhen

C. The characteristics of public bicycle users in Shenzhen

In order to improve the operating efficiency of public bicycles, we not only need to understand the current operation situation of public bicycle, but also need to further identify the user characteristics and mine potential benefits [8].

We know appeared 15531 users by statistical data of four weeks consecutive working days in July. We analyze by the frequency of the use of bicycles, classify the data around the statistics of the 5 days a week, the number of days the user, the following table 6:

TABLE VI NUMBER OF DAYS PER WEEK FOR USERS

Number of days	0	1	2	3	4	5	5 days of user accounts	3 days and more than 3 days the proportion of users
First week	1726	6686	866	686	495	376	0.035	0.144
Second weeks	1803	6881	826	607	457	261	0.024	0.122
Third weeks	1685	6826	892	568	499	365	0.034	0.132
Fourth week	3242	3992	1942	847	495	317	0.029	0.153

From the table, we can find a large number of users use once a week, we regard as the law of users who’ s frequency of use is greater than or equal to 3for a week, and about 3% of users within a week every day, and 3 days and more users accounted for more than 12%, the difference between our two that is the mining of user benefits.

For further statistical analysis of the above data, the following table 7:

TABLE VII THE NUMBER OF WEEKS MORE THAN 3 DAYS PER WEEK

The number of weeks	0	1	2	3	4
user	7840	1428	616	491	460
User proportion	0.724	0.132	0.057	0.045	0.042

Obviously, the number of users per week accounted for 4.2%, but there are still a large number of regular users within four weeks of 2 - 3 weeks appear more than 3 days, obviously, there are a lot of potential user benefits.

IV. CONCLUSION

In this paper, based on basic data of a month in Shenzhen city public bicycle system, we make the descriptive analysis by EXCEL, SPSS and Studio software. Taking July 2016 data as an example to illustrate the flow law of Shenzhen city public bicycle, and then borrow the car dot type analysis, get the characteristics of user travel.

First, we found the bike running all day long appeared the first peaking the 8::00 - 9:00 morning , evening peak appeared in 18:00 19:00; secondly, each time by bike fluctuation are smaller; from the living area to lend the largest amount of work on the government to the government departments, the largest flow of holidays in the living quarters living area to the user among the largest flow; and various outlets daily fluctuations, showed a steady state; finally, we analyze the frequency for the user, found that 5 consecutive working days about 13.78% users appear not less than 3 days, there are about 4.2% users each week for four consecutive weeks of not less than 3 days; at the same time, we found the statistics a large number of potential users. According to the different classification rules can also be excavated to different potential users, such as: according to the user often appear network information classification. I believe that through continuous data mining, bicycle system will achieve the best operating efficiency.

.

ACKNOWLEDGEMENT

The corresponding author of this paper is Yali Peng. This work was supported by National Natural Science Foundation of China No. 71661015, The Major Soft Science Research Program of Jiangxi Province No. 20161BBA10015.

REFERENCES

[1] Yu Sheng, Sheng’ao Yang, Wenling Tang, Xiukun Yang. Analysis of urban traffic congestion and scientific management [J]. Green technology, 2016, (22):123-124.

[2] Mei Zheng,Yanjun Zhang,Caiqing Yan,Xianlei Zhu ,James J.Schauer,Yuanhang Zhang. A survey of PM2.5 source apportionment methods in China [J]. Journal of Peking University (Natural Science Edition), 2014, (06):1141-1154.

[3] Tianjun Feng. Analysis of the effect of public bicycle transportation system [D].Jilin University, 2016.

[4] Hongmeng Song, Tiantian Ma, Jianxun Jiang. Descriptive statistics analysis of South Zhongguancun streetcar based on traffic. [J]. Journal of Minzu University of China (Natural Science Edition), 2013, (S1):140-144.

[5] Runxing Liu. Principal component regression analysis using SPSS [J]. Chinese public health, 2001, (08):74-76.

[6] Shixin Ruan. The implementation, improvement and application of some regression analysis methods [D].Jilin University, 2005.

[7] Peng Yali, Lv Ling, Xie Zhongfu, Zhang Zhiming, Xu Hong. Research on Layered-Partition Communication Model on Intelligent Urban VANET [J] Journal of Jiangxi Normal University, 2016, 6 (40): 627-634.

[8] Lu Bai. Potential Group Mining Based on user annotation behavior and interest points [J]. Technology entrepreneurship monthly, 2015, (22):103-10.