Development of Effective Marketing Strategies for a Housing Equipment Company Based on Analysis of Sales Data and Sales Personnel

Kakunoshin Nishiyama, Yuya Nakata, Ushio Sumita, Takaki Hayashi, and Masahiro Okada
Graduate School of Business Administration
Keio University
Yokohama, Kanagawa, Japan
kaku@keio.jp, nakata.yuya@keio.jp, sumita@kbs.keio.ac.jp, takaki@kbs.keio.ac.jp, m-okada@kbs.keio.ac.jp

Abstract*—A housing equipment company under consideration has a variety of products such as sash windows, bath modules, toilets, and bathroom sinks. The company also provides construction and maintenance services through a network of partner building companies. It typically takes a few months to establish an order since the initial contact to a client and another few months before the construction for the order actually takes place. Because of this, the company often suffers from losing orders. One client may buy only one product or a combination of products at a time. Since sales channels for different products are not necessarily the same, multiple sales persons from different sales offices may approach the same client without realizing it. In order to eliminate such unnecessary complications, it is important to develop effective marketing strategies for coordinating sales personnel better. The purpose of this paper is to achieve this goal by analyzing sales data and sales personnel. More specifically, by identifying the characteristics of profitable clients and those of clients who are likely to cancel orders, we provide a basis to establish recommendation rules concerning how combinations of products might be sold to one client and how sales personnel should be assigned to that client.

***Keywords*—Marketing Strategy, Statistical Analysis, Sales Data, Combination of Products, Characteristics of Clients**

I. INTRODUCTION

Because of the rapid advancement of information technologies surrounding the Internet, including mobile devices from the hardware side, and cloud computing and SNS services from the software side, it is now affordable to gather massive amount of data with speed and efficiency. This phenomenon called Big Data has been drastically changing how the management should be done in every industry all over the world. As discussed in LaValle et al. (2010) [1], Chen, Chiang, and Storey (2012) [2], and Davenport, Barth, and Bean (2012) [3], however, it is a difficult challenge for many corporations to find a way to establish a path from insights to values through Big Data. In this regard, a new approach named "Business Intelligence and Analytics" has emerged, where a variety of analytical techniques would be applied to Big Data so as to extract useful managerial implications from the sea of massive data. Some of such techniques include two-dimensional association rules derived using semidefinite programming in Fujisawa et al. (1999) [4], smoothness prior approach for analyzing large time series data in Kitagawa, Higuchi, and Kondo (2003) [5], machine learning in automated text categorization in Sebastiani (2002) [6], Magnetic Agile Deep (MAD) data analysis skills discussed in Cohen et al. (2009) [7], and predictive analytics for Customer Relationship Management in Waller and Fawcett (2013) [8], to name only a few. Along this line of research, the purpose of this paper is to demonstrate the usefulness of Big Data Analytics, where the massive sales data provided by a housing equipment company would be analyzed so as to extract useful managerial implications.

The structure of this paper is as follows. In Section II, we describe the content of real data provided by the company. Furthermore, certain notation would be introduced so as to facilitate the subsequent analyses to follow. Since the amount of data that we deal with is so huge, in Section III, we narrow down the original set of data covering the entire market in Japan into a subset corresponding to the five regions of key interest for our purpose. Fundamental characteristics of the selected five regions would then be captured in some detail. Section IV is devoted to the analysis of purchase patterns of clients of the company for identifying combinations of products that produce large sales revenues, as well as those which suffer from losing orders with big potential sales revenues lost. Finally, in Section V, some concluding remarks are given.

II. DATA DESCRIPTION AND NOTATION

In this section, we describe a set of data to be used for the study. The company under consideration is a housing equipment company with the following eleven sales regions in Japan: Hokkaido; Tohoku; Kitakanto; Tokyo; Chiba; Saitama; Kanagawa; Chubu; Kansai; Chushikoku; and Kyusyu. The company sells seven different types of products, which are 1) sash windows; 2) floorings; 3) bath modules; 4) toilets; 5) kitchen sinks and counters ; 6) bathroom sinks; and 7) exterior walls and gates. We employ the actual data provided by the company over the period of December 1, 2013 through June 28, 2015, involving the sales amount and the lost sales amount for each product type from each client, as well as the sales persons assigned to the client.

Given a region, let $I = \{1, 2, \cdots, I\}$ and $J = \{1, 2, \cdots, 7\}$ be a set of clients and a set of product types, respectively. For each $i \in I$, the sales amount for product type $j \in J$ is denoted by $s(i, j)$, and $ls(i, j)$ is defined similarly for the amount of lost sales. The sales amount and the lost amount for client $i \in I$ are then given by $s(i) = \sum_{j \in J} s(i, j)$ and $ls(i) = \sum_{j \in J} ls(i, j)$, respectively. Also of interest is the number of sales persons assigned to client $i \in I$, which is denoted by $m(i)$. Consequently, the profile vector $\underline{p}(i)$ for client $i \in I$ can be constructed as given in Table I.

TABLE I. PROFILE VECTOR $\underline{p}(i)$ FOR CLIENT $i \in I$

Client	$s(i,1)$	...	$s(i,7)$	$ls(i,1)$	...	$ls(i,7)$	$s(i)$	$ls(i)$	$m(i)$
…	72	…	93	0	…	0	180	12	2
i	32	…	0	0	…	15	80	60	1
…	98	…	90	0	…	0	199	0	1

Associated with $s(i, j)$ and $m(i)$, the following flags are introduced.

$$sf(i, j) = \begin{cases} 1 & \text{if } s(i, j) > 0 \\ 0 & \text{otherwise} \end{cases} \quad , \quad mf(i, j) = \begin{cases} 1 & \text{if } m(i) > 0 \\ 0 & \text{otherwise} \end{cases} \quad .$$

The flag profile vector $\underline{fp}(i)$ for client $i \in I$ is then defined as given in Table II.

TABLE II. FLAG PROFILE VECTOR $\underline{fp}(i)$ FOR CLIENT $i \in I$

Client	$sf(i,1)$	...	$sf(i,7)$
…	1	…	0
I	1	…	1
…	0	…	0

It should be noted that the purchase pattern of client $i \in I$ can be characterized by $\underline{fp}(i)$, which belongs to a set of seven dimensional binary vectors, having components being either 0 or 1, denoted by B^7.

Let $\underline{q} \in \mathrm{B}^7$ be a purchase pattern and define $C(\underline{q}) = \{i : \underline{fp}(i) = \underline{q}\}$. Then total sales amount and the total lost sales amount associated with $\underline{q} \in \mathrm{B}^7$, denoted by $S(\underline{q})$ and $LS(\underline{q})$ respectively, can be obtained as

$$S(\underline{q}) = \sum_{i \in C(\underline{q})} s(i) \qquad \text{and} \qquad LS(\underline{q}) = \sum_{i \in C(\underline{q})} ls(i) \quad .$$

In the next section, we identify several regions of importance for our analysis by ranking them in terms of the sum of $s(i)$ and $ls(i)$ within each region and their respective averages. The selected regions are then examined so as to understand the overall perspective of the regional markets.

III. KEY REGIONS OF INTEREST

In order to understand the characteristics of the Japanese market that the company faces, we first attempt to identify key regional markets out of the eleven regions. In Tables III A) through D), the eleven regions are ranked in descending order along two axes and then grouped into ten cells with 10 % distance. In Table III A), for example, Chubu, Chushikoku, and

Kansai collectively contribute to 50 % of No. of Clients with Sales and 50 % of Total Sales. We realize that the following five regions are ranked in the 90 % or more percentile in every table: Tohoku, Kitakanto, Tokyo, Chubu, and Kansai, as indicated by grey cells in the four tables. Throughout the remainder of this paper, we focus on analysis of these five regions.

TABLE III. TWO DIMENSIONAL RANKING OF REGIONAL MARKETS

A) NO. OF CLIENTS WITH SALES VS. AMOUNT OF TOTAL SALES

		Amount of Total Sales									
		0.1	***0.2***	***0.3***	***0.4***	***0.5***	***0.6***	***0.7***	***0.8***	***0.9***	***1.0***
	0.1										
	0.2				Chushikoku						
	0.3										
	0.4		Chubu								
No. of Clients with Sales	0.5					Kansai					
	0.6							Tokyo			
	0.7										
	0.8						Kitakanto		Tohoku		
	0.9									Kanagawa	1
	1.0									1	2

B) NO. OF CLIENTS WITH LOST SALES VS. AMOUNT OF TOTAL LOST SALES

		Amount of Total Lost Sales									
		0.1	***0.2***	***0.3***	***0.4***	***0.5***	***0.6***	***0.7***	***0.8***	***0.9***	***1.0***
	0.1										
	0.2				Chushikoku						
	0.3										
	0.4		Chubu								
No. of Clients with Lost Sales	0.5					Kansai					
	0.6							Tokyo			
	0.7						Kitakanto				
	0.8								Tohoku		
	0.9									Kyushu	1
	1.0									1	2

C) NO. OF CLIENTS WITH SALES VS. AVERAGE AMOUNT OF SALES

		Average Amount of Sales									
		0.1	***0.2***	***0.3***	***0.4***	***0.5***	***0.6***	***0.7***	***0.8***	***0.9***	***1.0***
	0.1										
	0.2										1
	0.3										
	0.4					Chubu					
No. of Clients with Sales	0.5				Kansai						
	0.6								Tokyo		
	0.7										
	0.8						Tohoku	Kitakanto			
	0.9								kanagawa		1
	1.0		1	1						1	

D) NO. OF CLIENTS WITH LOST SALES VS. AVERAGE AMOUNT OF LOST SALES

		Average Amount of Lost Sales									
		0.1	***0.2***	***0.3***	***0.4***	***0.5***	***0.6***	***0.7***	***0.8***	***0.9***	***1.0***
No. of Clients with Lost Sales	0.1										
	0.2										1
	0.3										
	0.4				Chubu						
	0.5			Kansai							
	0.6								Tokyo		
	0.7		Kitakanto								
	0.8					Tohoku					
	0.9						Kyushu				1
	1.0							2		1	

Tables IV A) and B) summarize Total Sales and Total Lost Sales for these five regions, where the former describes those cases with a single sales person assigned to each client, while the latter represents those cases with multiple sales persons assigned to each client. It can be seen that the cases of the single sales person assignment supersedes those of the multiple sales person assignment in Average Total Sales in all of the five regions. This is somewhat unintuitive since sales efforts with more people assigned are likely to yield larger revenues. In contrast, sales teams with multiple persons reduce Average Total Lost Sales, in comparison with the single sales person case, in all the regions except Tokyo. Perhaps Tokyo is exceptional because of the severe competitiveness of the Tokyo market.

For the single sales person assignment case, Total Sales is exceeded by Total Lost Sales in Tohoku, Kitakanto, and Tokyo, while this phenomenon is reversed in Chubu and Kansai. When multiple sales persons are assigned to each client, Total Sales is much larger than Total Lost Sales, perhaps demonstrating the advantage of having sales teams with multiple persons so as to keep better relationships with clients.

TABLE IV. TOTAL SALES AND TOTAL LOST SALES BY REGION

A) SINGLE SALES PERSON ASSIGNED TO EACH CLIENT

	Sales			**Lost Sales**		
	Total Sales (¥1,000)	***No. of Clients with Sales***	***Average (¥1,000)***	***Total Sales (¥1,000)***	***No. of Clients with Lost Sales***	***Average (¥1,000)***
Tohoku	11,940,946	28,501	419.0	11,557,352	26,698	432.9
Kitakanto	14,627,104	33,842	432.2	16,015,102	41,716	383.9
Tokyo	12,735,191	32,553	391.2	13,785,611	38,266	360.3
Chubu	19,584,874	48,490	403.9	19,373,665	49,001	395.4
Kansai	15,852,131	33,457	473.8	13,785,611	38,266	360.3

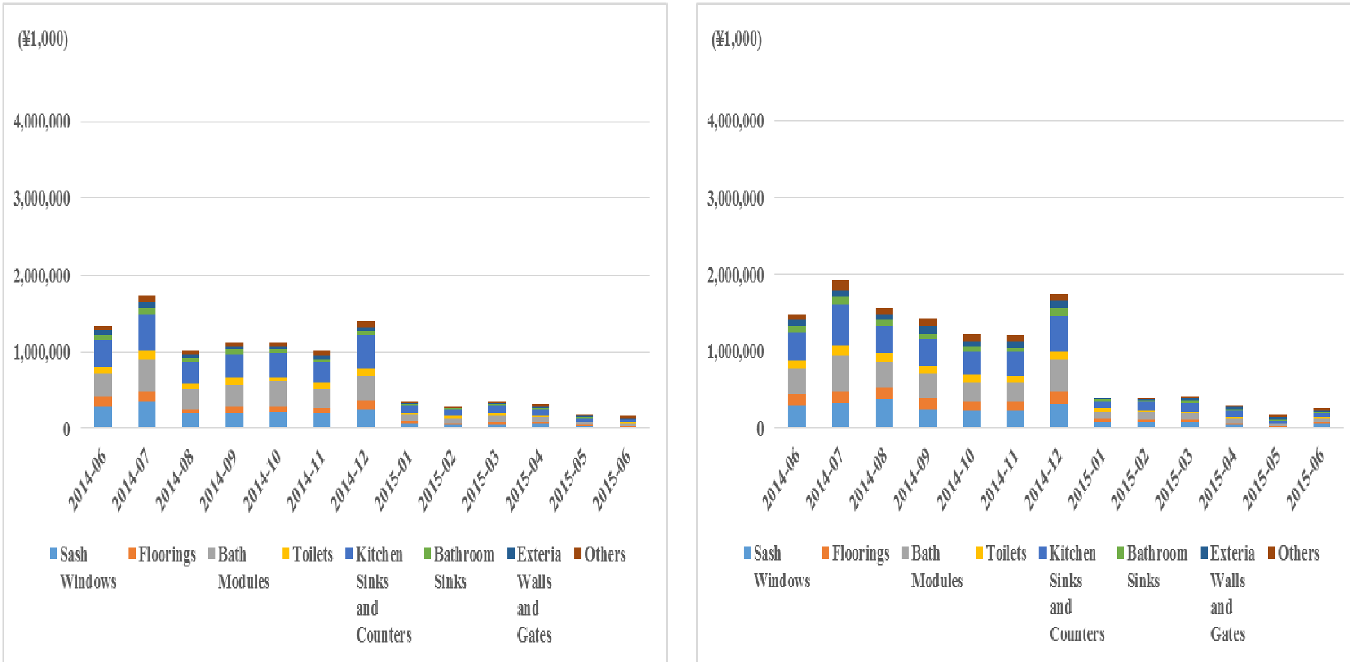
B) MULTIPLE SALES PERSONS ASSIGNED TO EACH CLIENT

	Sales			**Lost Sales**		
	Total Sales (¥1,000)	***No. of Clients with Sales***	***Average (¥1,000)***	***Total Sales (¥1,000)***	***No. of Clients with Lost Sales***	***Average (¥1,000)***
Tohoku	1,343,176	3,913	343.3	1,208,100	3,549	340.4
Kitakanto	2,181,515	5,926	368.1	1,232,979	3,648	338.0
Tokyo	605,889	1,960	309.1	386,237	1,044	370.0
Chubu	4,778,205	12,980	368.1	3,667,020	11,622	315.5
Kansai	605,889	1,960	309.1	386,237	1,044	370.0

Fig. 1 a) and b) depict a sequential transition of Total Sales by product types over the period of June 1, 2014 through June 28, 2015, and that of Total Lost Sales by product types over the same period, respectively, for the region of Tohoku, which is the most northern region. In the order of Kitakanto, Tokyo, Chubu and Kansai from the north to the south, Fig. 2 a) and b) through Fig. 5 a) and b) illustrate the same set of data in a similar manner. One realizes that both Total Sales and Total Lost Sales decrease substantially from 2014 to 2015. This reduction is more visible in Tohoku and Kitakanto in comparison with Tokyo, Chubu, and Kansai. This difference may be explained by the fact that the former regions have Total Sales and Total Lost Sales more equally balanced across different product types, while Total Sales and Total Lost Sales for the latter regions are dominated by two types of products: Bath Modules, and Kitchen Sinks and Counters.

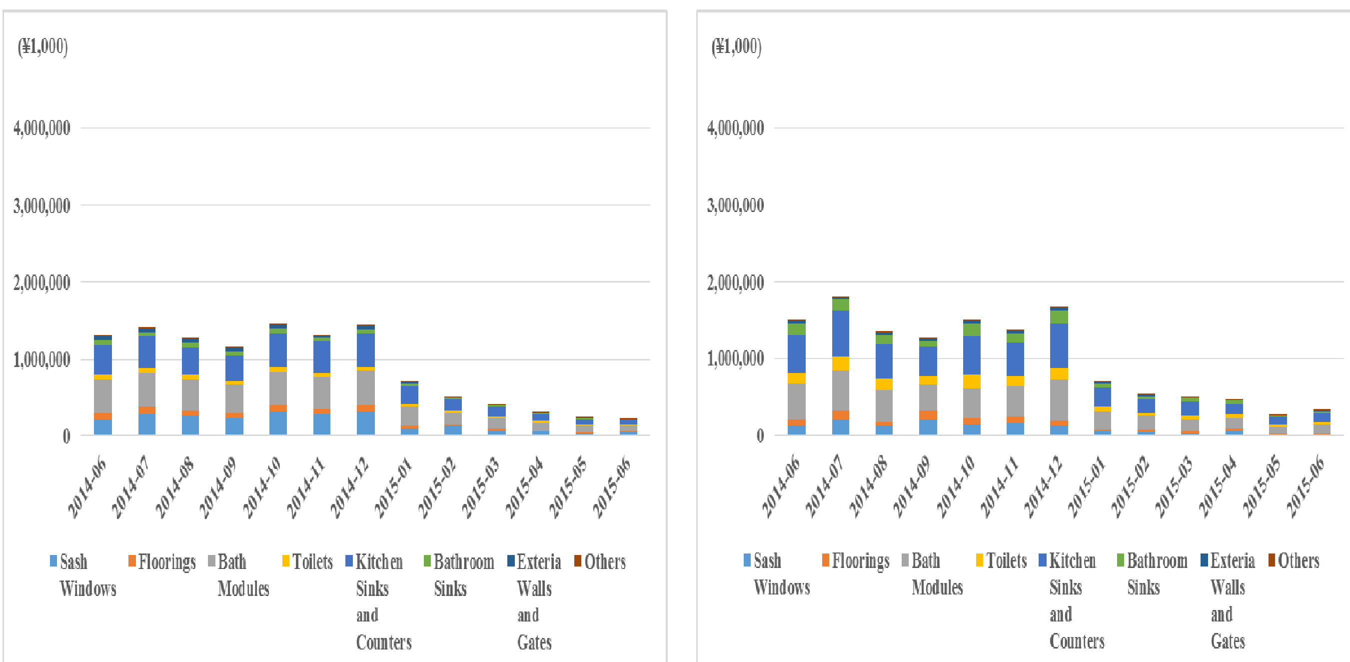
a) Total Sales b) Total Lost Sales

Fig. 1. Time Series of Total Sales and Total Lost Sales by Region : Tohoku



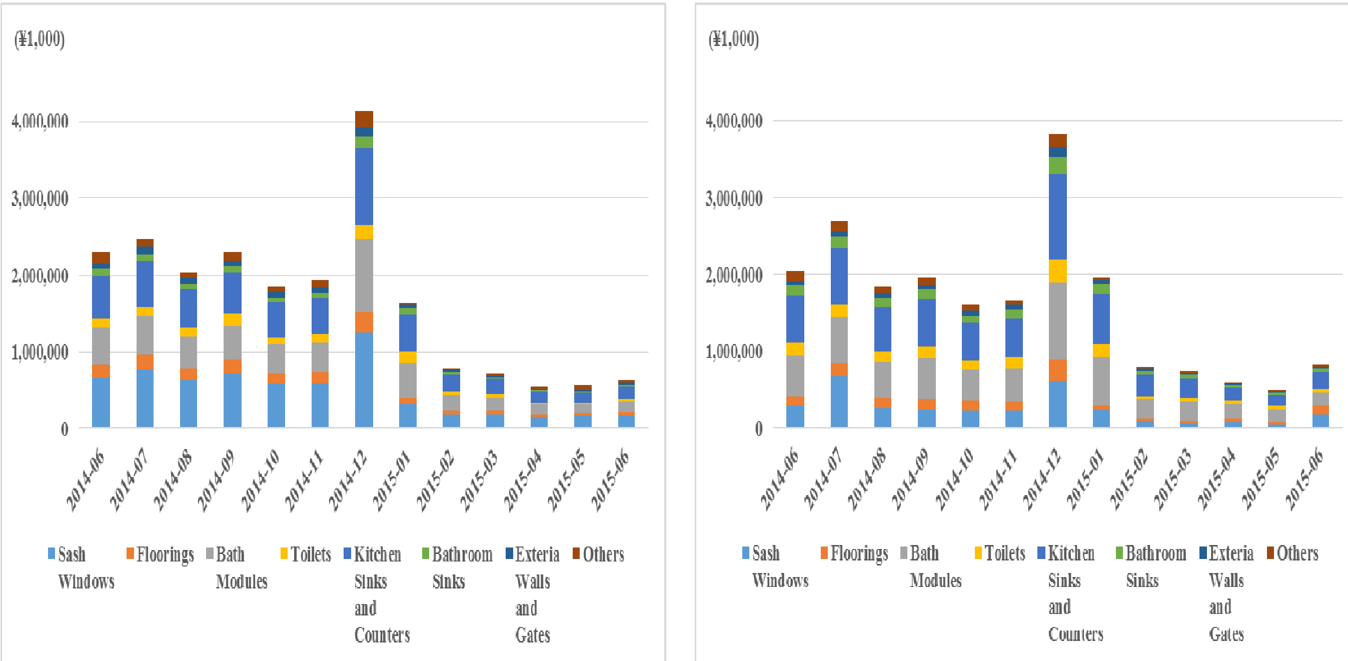
a) Total Sales b) Total Lost Sales

Fig. 2. Time Series of Total Sales and Total Lost Sales by Region : Kitakanto



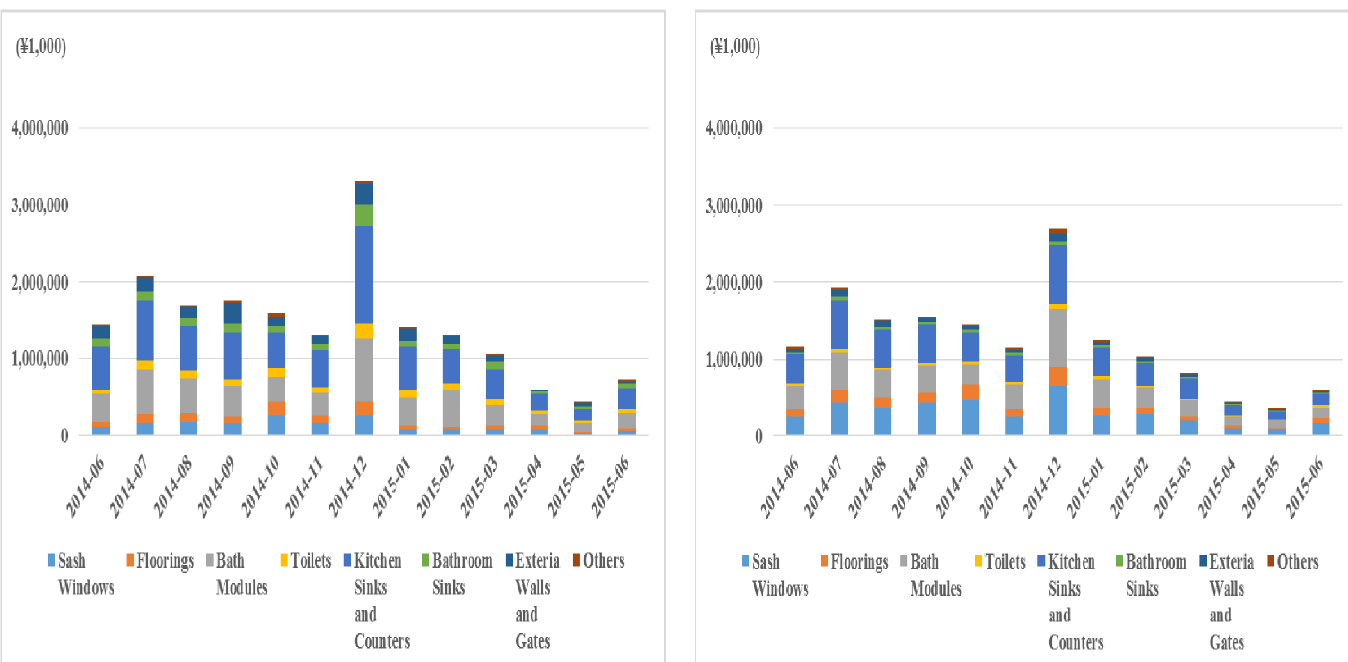
a) Total Sales b) Total Lost Sales

Fig. 3. Time Series of Total Sales and Total Lost Sales by Region : Tokyo



a) Total Sales b) Total Lost Sales

Fig. 4. Time Series of Total Sales and Total Lost Sales by Region : Chubu



a) Total Sales b) Total Lost Sales

Fig. 5. Time Series of Total Sales and Total Lost Sales by Region : Kansai

Table V exhibits Amount, No. of Clients, and Average of Total Sales as well as those of Total Lost Sales over two years 2014 and 2015 for the five regions. It should be noted that Amount of Total Lost Sales supersedes that of Total Sales in every year and in every region, with exception of Chubu in 2014. This phenomenon can be also observed for No. of Clients except Tohoku and Chubu in 2014. It may be worth noting that only Chubu has Total Sales dominating Total Lost Sales in both Amount and No. of Clients in the total of 2014 and 2015. As for the differences between 2014 and 2015, one observes that both Amount and No. of Clients decrease in every region in every year. This reduction is relatively minor in Kansai with 60.9 % reduction in Total Sales and 58.6 % in Total Lost Sales, followed by Chubu with 66.8 % and 63.6 %, in comparison with other regions having about 70 % or more of reduction in both Total Sales and Total Lost Sales. In Tohoku and Kitakanto, both Average Total Sales and Average Total Lost Sales slightly increase, while those in all other regions decrease.

TABLE V. COMPARISON OF FIVE REGIONS

		Amount (¥1,000)			*No. of Clients*			*Average (¥1,000)*		
		(1) Total Sales	*(2) Total Lost Sales*	*(1) - (2)*	*(1) Total Sales*	*(2) Total Lost Sales*	*(1) - (2)*	*(1) Total Sales*	*(2) Total Lost Sales*	*(1) - (2)*
Tohoku	2014	7,905,714	8,118,322	-212,608	21,331	19,377	1,954	370.6	419.0	-48.3
	2015	2,165,912	2,625,341	-459,429	5,572	5,995	-423	388.7	437.9	-49.2
	Total	10,071,626	10,743,663	-672,037	26,903	25,372	1,531	374.4	423.4	-49.1
Kitakanto	2014	8,741,676	10,538,752	-1,797,076	24,113	28,759	-4,646	362.5	366.5	-3.9
	2015	1,655,223	1,922,878	-267,655	4,345	5,067	-722	380.9	379.5	1.5
	Total	10,396,899	12,461,630	-2,064,731	28,458	33,826	-5,368	365.3	368.4	-3.1
Tokyo	2014	9,301,972	10,489,471	-1,187,499	25,062	28,987	-3,925	371.2	361.9	9.3
	2015	2,370,854	2,737,220	-366,366	6,725	7,700	-975	352.5	355.5	-2.9
	Total	11,672,826	13,226,691	-1,553,865	31,787	36,687	-4,900	367.2	360.5	6.7
Chubu	2014	16,988,673	15,622,575	1,366,098	41,511	40,167	1,344	409.3	388.9	20.3
	2015	4,881,288	5,407,577	-526,289	13,774	14,608	-834	354.4	370.2	-15.8
	Total	21,869,961	21,030,152	839,809	55,285	54,775	510	395.6	383.9	4.5
Kansai	2014	11,441,569	13,143,544	-1,701,975	24,076	27,212	-3,136	475.2	483.0	-7.8
	2015	4,472,338	5,498,870	-1,026,532	9,419	11,256	-1,837	474.8	488.5	-13.7
	Total	15,913,907	18,642,414	-2,728,507	33,495	38,468	-4,973	475.1	484.6	-9.5

IV. CHARACTERISTIC ANALYSIS OF PURCHASE PATTERNS

The housing equipment company under consideration offers a variety of products as discussed in Section II. Accordingly, a client may purchase one product or multiple products at a time. In this chapter, we examine whether or not certain purchase patterns over seven types of products [1) sash windows; 2) floorings; 3) bath modules; 4) toilets; 5) kitchen sinks and counters; 6) bathroom sinks; and 7) exterior walls and gates] can be characterized for generating larger Total Sales or Total Lost Sales. In other words, we are interested in categorizing purchase patterns in terms of Total Sales and Total Lost Sales.

We recall from Section II that a purchase pattern of client $i \in I$ over the seven types of products is expressed by a seven dimensional binary vector $\underline{q} \in \mathbf{B}^7$, as shown in Table II. The set of clients having the purchase pattern $\underline{q}$ is denoted by $C(\underline{q}) = \{i : \underline{fp}(i) = \underline{q}\}$. The purpose of this section is to find subsets of $\mathbf{B}^7$ so as to establish efficient marketing strategies for increasing Total Sales and decreasing Total Lost Sales. For this purpose, two key measures of interest are

$$S(\underline{q}) = \sum_{i \in C(\underline{q})} s(i) \quad \text{and} \quad LS(\underline{q}) = \sum_{i \in C(\underline{q})} ls(i) \quad .$$

The first step of our analysis is to rank $\underline{q}$ in descending order in terms of $S(\underline{q})$ and $|C(\underline{q})|$, and choose those which belong to the percentile of 90 % or more. We define the set of those purchase patterns by $Total-Sales$. Similarly, we rank them in terms of $S(\underline{q})/|C(\underline{q})|$ and $|C(\underline{q})|$, yielding a subset of B^7 denoted by $Ave-Sales$. The above procedure is repeated by replacing $S(\underline{q})$ by $LS(\underline{q})$, producing $Total-Lost-Sales$ and $Ave-Lost-Sales$. For notational convenience, we define

$$Sales = Total-Sales \cup Ave-Sales \quad \text{and}$$

$$Lost-Sales = Total-Lost-Sales \cup Ave-Lost-Sales \qquad .$$

We are now in a position to introduce a matrix consisting of nine cells as shown in Table VI, where (m, n) cells are defined in the following manner.

$(1,1): Total-Sales \setminus (Ave-Sales \cup Lost-Sales)$

$(1,2): (Total-Sales \cap Ave-Sales) \setminus Lost-Sales)$

$(1,3): Ave-Sales \setminus (Total-Sales \cup Lost-Sales)$

$(2,1): (Total-Sales \cap Total-Lost-Sales) \setminus (Ave-Sales \cup Ave-Lost-Sales)$

$(2,2): Total-Sales \cap Total-Lost-Sales \cap Ave-Sales \cap \mathrm{Ave}-Lost-Sales$

$(2,3): (Ave-Sales \cap Ave-Lost-Sales) \setminus (Total-Sales \cup Total-Lost-Sales)$

$(3,1): Total-Lost-Sales \setminus (Sales \cup Ave-Lost-Sales)$

$(3,2): (Total-Lost-Sales \cap Ave-Lost-Sales) \setminus Sales$

$(3,3): Ave-Lost-Sales \setminus (Sales \cup Total-Lost-Sales)$

TABLE VI. NINE SUBSETS OF PURCHASING PATTERNS

	Only Total	**Both of Total and Average**	**Only Average**
Only Sales	(1, 1)	(1, 2)	(1, 3)
Both of Sales and Lost Sales	(2, 1)	(2, 2)	(2, 3)
Only Lost Sales	(3, 1)	(3, 2)	(3, 3)

In Table VII, all purchasing patterns in $Sales \cup Lost - Sales$, which appear at least once in one of the five regions over the data period, are listed with indication of which cell in Table VI it belongs to. It should be noted that only four cells appear more than three times on a certain row in this table: $(1, 2), (2, 1), (2, 2), and\ (3, 1)$.

TABLE VII. PURCHASING PATTERNS BY REGIONS

	Tohoku	Kitakanto	Tokyo	Chubu	Kansai
0000001	(2, 1)		(1, 1)	(1, 1)	
0000010	(3, 1)	(3, 1)	(3, 1)		(3, 1)
0000100	(2, 1)	(2, 1)	(2, 1)	(2, 1)	(2, 1)
0000110		(3, 1)	(3, 1)		
0001000	(2, 1)	(2, 1)	(2, 1)	(2, 1)	(3, 1)
0001010	(3, 1)	(1, 1)		(1, 1)	
0001110	(1, 1)	(2, 2)	(3, 2)	(3, 3)	
0010000	(2, 1)	(2, 1)	(2, 1)		(2, 1)
0010010	(2, 1)	(1, 1)	(2, 1)		
0010100	(2, 2)	(2, 2)		(2, 2)	(2, 2)
0010110	(2, 2)	(2, 2)		(2, 2)	(2, 2)
0011010		(2, 2)			(3, 2)
0011100	(2, 2)	(3, 2)		(3, 2)	
0011110	(2, 2)	(2, 2)	(2, 2)	(2, 2)	(2, 2)
0100000	(1, 1)				(1, 1)
1000000				(2, 2)	(2, 1)
1000001	(1, 2)	(2, 2)		(1, 2)	(1, 2)
1100000	(2, 2)	(2, 2)		(2, 2)	(1, 2)
1100001	(2, 2)	(1, 2)		(1, 2)	
1111110	(2, 2)	(2, 2)		(2, 2)	(1, 2)
1111111	(2, 2)	(2, 2)	(2, 2)	(2, 2)	(2, 2)

In Table VIII, actual purchasing patterns constituting each cell are listed, along with Total Sales and Total Lost Sales in terms of Amount and No. of Clients for the five regions. It is somewhat surprising to find that every region has the heaviest weight in the cell $(2, 2)$ in Amount in both Sales and Lost Sales, while $(2, 1)$ in No. of Clients in both Sales and Lost Sales. We note that the cell $(2, 2)$ contains those purchasing patterns constituting the 90 percentile in all of Total Sales, Average Total Sales, Total Lost Sales, and Average Total Lost Sales. The cell $(2, 1)$ consists of those purchasing patterns constituting the 90 percentile in Total Sales and Total Lost Sales. Accordingly, the two cells contribute mostly to both Total Sales and Total Lost Sales. This type of contributions could happen basically in two different ways: 1) individual clients in the two cells have both Sales and Lost Sales; or 2) a group of clients largely contribute to Sales, while other clients largely cause Lost Sales.

TABLE VIII. SALES AND LOST SALES FOR FOUR PURCHASING PATTERNS

		Tohoku				**Kitakanto**			
		Total Sales		***Total Lost Sales***		***Total Sales***		***Total Lost Sales***	
		Amount (¥1,000)	***No. of Clients***	***Amount (¥1,000)***	***No. of Clients***	***Amount (¥1,000)***	***No. of Clients***	***Amount (¥1,000)***	***No. of Clients***
(1, 2)	1000001	310,565	433	39,342	47	466,405	631	260,520	334
	Sub Total	310,565	433	39,342	47	466,405	631	260,520	334
(2, 1)	0000100	844,575	2,589	1,391,104	3,196	1,614,055	5,128	1,382,079	3,577
	0001000	140,241	883	251,874	1,566	317,597	1,584	340,834	1,743
	0010000	664,557	1,317	983,110	1,379	804,360	1,903	904,193	1,639
	Sub Total	1,649,373	4,789	2,626,088	6,141	2,736,012	8,615	2,627,106	6,959
(2, 2)	0010100	703,550	771	820,345	494	627,382	749	551,364	516
	0010110	290,163	360	609,869	460	456,102	439	392,015	402
	0011110	1,578,655	1,690	2,887,538	2,414	2,633,379	2,447	2,422,133	2,168
	1100000	826,618	919	423,734	348	625,287	580	765,865	510
	1111110	657,287	359	388,807	350	507,939	284	590,315	347
	1111111	648,093	269	617,239	328	1,828,082	905	4,225,817	2,239
	Sub Total	4,704,366	4,368	5,747,532	4,394	6,678,171	5,404	8,947,509	6,182
(3, 1)	0000010	31,409	234	213,573	872	101,592	794	199,180	975
	Sub Total	31,409	234	213,573	872	101,592	794	199,180	975
Others		5,245,233	7,673	4,138,917	4,153	4,644,924	6,146	5,213,766	5,395
Total		11,940,946	17,497	12,765,452	15,607	14,627,104	21,590	17,248,081	19,845

		Tokyo				**Chubu**				**Kansai**			
		Total Sales		***Total Lost Sales***		***Total Sales***		***Total Lost Sales***		***Total Sales***		***Total Lost Sales***	
		Amount (¥1,000)	***No. of Clients***	***Amount (¥1,000)***	***No. of Clients***	***Amount (¥1,000)***	***No. of Clients***	***Amount (¥1,000)***	***No. of Clients***	***Amount (¥1,000)***	***No. of Clients***	***Amount (¥1,000)***	***No. of Clients***
(1, 2)	1000001	137,282	113	83,016	50	915,578	1,251	248,377	315	299,299	401	33,008	28
	Sub Total	137,282	113	83,016	50	915,578	1,251	248,377	315	299,299	401	33,008	28
(2, 1)	0000100	2,100,978	6,677	2,320,709	5,971	2,476,580	6,164	2,555,017	5,462	2,607,398	4,643	4,476,707	7,723
	0001000	159,327	663	527,779	1,689	330,944	1,743	441,287	2,399	59,647	444	403,121	2,093
	0010000	1,654,624	4,400	1,677,253	3,457	1,357,436	2,970	1,579,480	2,477	1,687,669	3,311	2,071,438	3,298
	Sub Total	3,914,929	11,740	4,525,741	11,117	4,164,960	10,877	4,575,784	10,338	4,354,714	8,398	6,951,266	13,114
(2, 2)	0010100	866,686	1,316	724,143	811	1,190,094	1,483	1,020,305	960	1,824,503	1,770	734,504	764
	0010110	395,186	534	571,649	554	650,113	687	823,435	749	348,988	331	841,029	526
	0011110	1,404,879	1,360	2,060,721	1,725	2,997,216	2,782	4,723,356	3,950	1,032,934	850	2,206,468	1,561
	1100000	254,876	161	171,867	88	1,301,162	976	752,893	419	1,006,770	912	275,274	184
	1111110	387,036	206	418,449	279	1,087,213	676	786,312	522	1,051,364	411	644,673	344
	1111111	1,237,097	763	1,739,432	1,173	3,415,571	1,856	2,960,672	2,062	1,430,035	891	2,869,239	1,331
	Sub Total	4,545,760	4,340	5,686,261	4,630	10,641,369	8,460	11,066,973	8,662	6,694,594	5,165	7,571,187	4,710
(3, 1)	0000010	95,805	613	368,531	1,684	110,111	782	267,926	1,510	32,996	200	497,767	2,078
	Sub Total	95,805	613	368,531	1,684	110,111	782	267,926	1,510	32,996	200	497,767	2,078
Others		4,041,415	4,712	3,508,299	4,039	8,556,588	10,843	6,891,335	7,005	5,648,977	7,958	4,835,122	3,913
Total		12,735,191	21,518	14,171,848	21,520	24,388,606	32,213	23,050,395	27,830	17,030,580	22,122	19,888,350	23,843

In order to understand how contributions to Total Sales and Total Lost Sales in the two cells $(2,1)$ and $(2,2)$ are made for each of the five regions, we decompose a set of clients $U(m,\mathrm{n})$ in the cell (m,n) for each region into three sets: Sales Type, Balanced Type, and Lost Sales Type. For the cell (m,n), the decomposition is made in the following manner. Let $Diff(i)$ be the difference between Total Sales and Total Lost Sales for client $i \in U(m,\mathrm{n})$ given by $Diff(i) = s(i) - ls(i)$, and denote its mean by $\mu(m,\mathrm{n})$ and its standard deviation by $\sigma(m,\mathrm{n})$. A client $i \in U(m,\mathrm{n})$ is said to be of

Sales Type $\Leftrightarrow$ $\mu(m,\mathrm{n}) + 0.3\times\sigma(m,\mathrm{n}) < Diff(i)$,
Balanced Type $\Leftrightarrow$ $\mu(m,\mathrm{n}) - 0.3\times\sigma(m,\mathrm{n}) \le Diff(i) \le \mu(m,\mathrm{n}) + 0.3\times\sigma(m,\mathrm{n})$,
Lost Sales Type $\Leftrightarrow$ $Diff(i) < \mu(m,\mathrm{n}) - 0.3\times\sigma(m,\mathrm{n})$.

Tables IX A) through C) exhibit how each of the above three types of clients contributes to Total Sales and Total Lost Sales in the two cells $(2,1)$ and $(2,2)$ in the five regions. In the cell $(2,1)$, Balanced Type dominates Sales Type and Lost Sales Type in terms of No. of Clients in Tohoku, Kitakanto, and Tokyo with Balanced Type constituting 50.5 %, 47.4 % and 45.8 %, respectively, while the contributions of Sales Type or Lost Sales Type remain under 30 %. Kansai has the reversed dominance with 36.5 % for Sales Type and 35.1 % for Lost Sales Type exceeding 28.4 % of Balanced Type. Chubu has more or less equal contribution from each type with 33.3 % for Sales Type, 36.5 % for Balanced Type and 30.2 % for Lost Sales Type. In terms of Amount, however, it is clear that Sales Type dominates Total Sales with contribution of 84 % or more, and Lost Sales Type governs Total Lost Sales with 80 % or more in all of the five regions.

As for the cell $(2,2)$, Sales Type and Lost Sales Type dominate Balanced Type in both Amount and No. of Clients throughout the five regions. In terms of Amount, Sales Type of Tohoku accounts for 89.6 % of Total Sales and a similar phenomenon is visible for other regions as well: Kitakanto with 91.3 %, Tokyo with 88.8 %, Chubu with 91.1 %, and Kansai with 86.1 %. This dominance structure is even stronger for Total Lost Sales, where Lost Sales Type contribute to Total Lost Sales significantly more than other two types: 89.3 % for Tohoku, 93.4 % for Kitakanto, 91.6 % for Tokyo, 90.6 % for Chubu, and 91.6 % for Kansai. Concerning No. of Clients, an almost equal contribution from each of Sales Type and Lost Sales Type exceeds that from Balanced Type. One observes that the contributions of Sales Type, Balanced Type, and Lost Sales Type are 47.7 %, 13.2 %, and 39.1 %, respectively, in Tohoku; 46.3 %, 5.9%, and 47.8 % in Kitakanto; 38.9 %, 6.0 %, and 55.0 % in Tokyo; 52.2 %, 7.7 %, and 40.1 % in Chubu; and 43.3 %, 10.2 %, and 46.5 % in Kansai.

TABLE IX. SALES AND LOST SALES BY CLIENT TYPES IN FIVE REGIONS (¥1,000)

A) TOHOKU

			Tohoku					
			(2, 1)		*(2, 2)*		*Total*	
Sales Type	Sales	Amount	1,987,217	85.5%	4,971,569	89.6%	6,958,786	88.4%
		Average	578.9		1,277.7		950.1	
	Lost Sales	Amount	9,885	0.3%	250,051	4.0%	259,936	2.8%
		Average	154.5		293.5		283.8	
	No. of Clients		3,433	26.6%	3,891	47.7%	7,324	34.7%
Balanced Type	Sales	Amount	332,728	14.3%	402,780	7.3%	735,508	9.3%
		Average	120.7		417.4		197.7	
	Lost Sales	Amount	564,247	18.4%	424,401	6.8%	988,648	10.6%
		Average	148.1		472.6		210.0	
	No. of Clients		6,524	50.5%	1076	13.2%	7,600	36.1%
Lost Sales Type	Sales	Amount	3,704	0.2%	174,586	3.1%	178,290	2.3%
		Average	336.7		412.7		410.8	
	Lost Sales	Amount	2,491,448	81.3%	5,603,961	89.3%	8,095,409	86.6%
		Average	841.4		1,754.0		1,315.0	
	No. of Clients		2,961	22.9%	3,195	39.1%	6,156	29.2%
Total	Sales	Amount	2,323,649		5,548,935		7,872,584	
		Average	374.8		1,051.1		685.8	
	Lost Sales	Amount	3,065,580		6,278,413		9,343,993	
		Average	448.6		1,269.6		793.3	
	No. of Clients		12,918		8,162		21,080	

B) KITAKANTO AND TOKYO

			Kitakanto						Tokyo					
			(2, 1)		*(2, 2)*		***Total***		*(2, 1)*		*(2, 2)*		***Total***	
Sales Type	Sales	Amount	2,297,637	84.0%	6,955,243	91.3%	9,252,880	89.3%	3,501,061	85.1%	2,346,886	88.8%	5,847,947	86.5%
		Average	520.4		1,286.1		942.0		505.6		1,517.1		690.3	
	Lost Sales	Amount	2,387	0.1%	358,383	3.7%	360,770	2.9%	14,081	0.3%	184,973	4.9%	199,054	2.3%
		Average	217.0		370.2		368.5		132.8		382.2		337.4	
	No. of Clients		4,415	28.5%	5,408	46.3%	9,823	36.1%	6,925	29.5%	1,547	38.9%	8,472	30.8%
Balanced Type	Sales	Amount	434,681	15.9%	290,328	3.8%	725,009	7.0%	611,871	14.9%	124,517	4.7%	736,388	10.9%
		Average	103.8		507.6		152.3		118.2		541.4		136.2	
	Lost Sales	Amount	316,547	12.0%	277,443	2.9%	593,990	4.8%	672,610	14.2%	135,853	3.6%	808,463	9.5%
		Average	98.6		445.3		154.9		119.7		601.1		138.3	
	No. of Clients		7,348	47.4%	687	5.9%	8,035	29.6%	10,746	45.8%	240	6.0%	10,986	40.0%
Lost Sales Type	Sales	Amount	3,694	0.1%	374,620	4.9%	378,314	3.7%	1,947	0.0%	170,573	6.5%	172,520	2.6%
		Average	284.2		521.8		517.5		177.0		493.0		483.2	
	Lost Sales	Amount	2,308,172	87.9%	9,039,884	93.4%	11,348,056	92.2%	4,057,467	85.5%	3,479,327	91.6%	7,536,794	88.2%
		Average	617.7		1,619.5		1,217.7		697.5		1,590.2		941.5	
	No. of Clients		3,737	24.1%	5,582	47.8%	9,319	34.3%	5,817	24.8%	2,188	55.0%	8,005	29.1%
Total	Sales	Amount	2,736,012		7,620,191		10,356,203		4,114,879		2,641,976		6,756,855	
		Average	317.6		1,137.7		676.3		339.7		1,244.5		474.7	
	Lost Sales	Amount	2,627,106		9,675,710		12,302,816		4,744,158		3,800,153		8,544,311	
		Average	377.5		1,348.9		870.6		411.0		1,311.3		591.6	
	No. of Clients		15,500		11,677		27,177		23,488		3,975		27,463	

C) CHUBU AND KANSAI

			Chubu						Kansai					
			(2, 1)		*(2, 2)*		***Total***		*(2, 1)*		*(2, 2)*		***Total***	
Sales Type	Sales	Amount	2,568,443	91.5%	12,184,017	91.1%	14,752,460	91.2%	5,716,661	93.3%	3,992,824	86.1%	9,709,485	90.2%
		Average	490.5		1,257.8		988.6		665.1		1,458.8		856.8	
	Lost Sales	Amount	2,396	0.1%	850,779	6.5%	853,175	5.3%	970	0.0%	190,816	2.9%	191,786	1.4%
		Average	184.3		500.2		497.8		64.7		305.8		300.1	
	No. of Clients		5,236	33.3%	9,687	52.2%	14,923	43.5%	8,595	36.5%	2,737	43.3%	11,332	37.9%
Balanced Type	Sales	Amount	235,302	8.4%	415,955	3.1%	651,257	4.0%	408,928	6.7%	351,098	7.6%	760,026	7.1%
		Average	88.4		339.6		167.5		124.7		638.4		198.4	
	Lost Sales	Amount	291,799	9.7%	379,512	2.9%	671,311	4.2%	565,128	8.1%	367,250	5.5%	932,378	6.8%
		Average	94.2		493.5		173.6		165.8		591.4		231.4	
	No. of Clients		5,730	36.5%	1434	7.7%	7,164	20.9%	6,679	28.4%	642	10.2%	7,321	24.5%
Lost Sales Type	Sales	Amount	3,779	0.1%	769,044	5.8%	772,823	4.8%	1,176	0.0%	292,538	6.3%	293,714	2.7%
		Average	419.9		642.5		640.8		588.0		527.1		527.3	
	Lost Sales	Amount	2,702,109	90.2%	11,879,979	90.6%	14,582,088	90.5%	6,452,041	91.9%	6,093,174	91.6%	12,545,215	91.8%
		Average	568.9		1,598.1		1,196.8		780.0		2,074.6		1,119.2	
	No. of Clients		4,750	30.2%	7,434	40.1%	12,184	35.6%	8,272	35.1%	2,937	46.5%	11,209	37.5%
Total	Sales	Amount	2,807,524		13,369,016		16,176,540		6,126,765		4,636,460		10,763,225	
		Average	355.1		1,104.1		808.2		515.9		1,206.8		684.7	
	Lost Sales	Amount	2,996,304		13,110,270		16,106,574		7,018,139		6,651,240		13,669,379	
		Average	381.2		1,323.7		906.6		600.0		1,590.4		860.9	
	No. of Clients		15,716		18,555		34,271		23,546		6,316		29,862	

V. CONCLUDING REMARKS

In this paper, we analyzed a set of real data of large magnitude provided by a housing equipment company. So as to deal with the massive data with speed and efficiency for extracting useful managerial implications with speed and efficiency, we first extracted the five key regions which cover approximately 80 % or more of total sales revenues as well as that of total lost sales revenues. With focus on these five regions, we next identified combinations of products sold to a client at a time, which contribute to either sales revenues or lost sales revenues. This provides a basis for developing segmentation algorithms to identify potential clients so that sales revenues could be enhanced and lost sales revenues could be decreased. Such segmentation algorithms will be developed in due course and reported elsewhere.

ACKNOWLEDGMENT

This research is supported by MEXT Grand-in-Aid for Scientific Research (B) 25282086.

REFERENCES

[1] S. LaValle, et al., "Big data, analytics and the path from insights to value," MIT sloan management review 21, vol. 52, no. 2, 2010.

[2] H. Chen, R. H.L. Chiang, and V. C. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact," MIS quarterly, vol. 36, no. 4, pp. 1165-1188, 2012.

[3] T. H. Davenport, P. Barth, and R. Bean, "How 'big data'is different," MIT Sloan Management Review, vol. 54, no. 1, 2012.

[4] K. Fujisawa, et al., "Approximation of optimal two-dimensional association rules for categorical attributes using semidefinite programming," Discovery Science. Springer Berlin Heidelberg, pp. 148-159, 1999.

[5] G. Kitagawa, T. Higuchi, and F. N. Kondo, "Smoothness prior approach to explore mean structure in large-scale time series," Theoretical computer science, vol. 292, no. 2, pp. 431-446, 2003.

[6] F. Sebastiani, "Machine learning in automated text categorization," ACM computing surveys, vol. 34, no. 1, pp. 1-47, 2002.

[7] J. Cohen, et al., "MAD skills: new analysis practices for big data," Proceedings of the VLDB Endowment, vol. 2, no. 2, pp. 1481-1492, 2009.

[8] M. A. Waller and S. E. Fawcett, "Click here for a data scientist: Big data, predictive analytics, and theory development in the era of a maker movement supply chain," Journal of Business Logistics, vol. 34, no. 4, pp. 249-252, 2013.

BIOGRAPHY

Kakunoshin Nishiyama, a graduate of Faculty of Law at Keio University majoring in Criminal Law, has his career goal to establish himself as an investment banker for enhancing the global competitiveness of Japanese companies, where he intends to provide a bridge role between the financial management and the legal aspects of corporate businesses. He is currently a second year MBA student at Graduate School of Business Administration, Keio University, and will assume a position of Investment Banking Analyst at Merrill Lynch Japan Securities Co., Ltd. next April.

Yuya Nakata graduated from Faculty of Engineering, Hokkaido University majoring in Socio-Environmental Engineering. After graduation, he worked for Tsuchiya Co., Ltd. as a second class architect for 2 years, designing houses of many clients. In April 2014, he entered Graduate School of Business Administration, Keio University. He is currently a second year MBA student at Keio. His career goal is to work as a consultant for facilitating the growth of various municipal governments of Japan. His research interest includes development of practical tools for better managing the new hometown tax system by utilizing big data analytics.

Ushio Sumita is a Research Project Professor at Graduate School of Business Administration, Keio University, Japan. Prior to joining Keio, he has worked for such universities as Graduate School of Systems and Information Engineering, University of Tsukuba, Japan, for which he is a Professor Emeritus, Graduate School of International Management, The International University of Japan, William E. Simon Graduate School of Business Administration, University of Rochester, U.S.A, and Department of Industrial Engineering and Operations Research, Syracuse University, U.S.A. With his first Ph.D. in 1981 from the University of Rochester, U.S.A. and his second Ph.D. in 1987 from the Tokyo Institute of Technology, Japan, he has a wide range of research interests in both theoretical and functional areas, including applied probability, stochastic processes, financial engineering, e-marketing, information and communication technology, production and logistics systems, organizational theory, and comparative analysis of global management approaches, among others. He has published more than 160 papers in leading archive journals in such areas. He also has extensive experiences in business consulting in both the United States and Japan.

Takaki Hayashi is Professor of Management Science at Keio Business School. Previously, he was Assistant Professor of Statistics at Columbia University in the City of New York. He was also appointed as COE Visiting Associate Professor at the Graduate School of Mathematical Sciences, the University of Tokyo in 2004. Prior to joining academia, he worked as a financial engineer at the Industrial Bank of Japan, Ltd. His research interests are in statistical science for business and management, quantitative finance and financial econometrics, in particular statistical modeling and analysis of large-scale, highfrequency data. His research has appeared in a number of academic journals, including Bernoulli, Mathematical Finance, and Stochastic Processes and their Applications. He earned his Ph.D. in statistics from the University of Chicago. He earned his master's and bachelor's degrees in engineering from the University of Tokyo. He has been a Director of the Japanese Association of Financial Econometrics and Engineering (JAFEE) since 2006.

Masahiro Okada is a Professor of Strategic Management at the Graduate School of Business Administration, Keio University, Japan. He earned his Ph.D. in Business Administration from The Ohio State University in 1999. He also has an extensive working experience both as an administrator in Honda Motor Company and as a consultant in Arthur D. Little (Japan), a strategy consulting company. His research interests lie in the strategy theories and their application to business activities in emerging markets such as South East Asian countries and Sub-Sahara African countries. His current research topic includes how to measure the value of firm strategy not only in terms of economic value but also in terms of social value, which is the value generated by firms for stakeholders other than shareholders. His recent published articles include "Can Creating Shared Value Be a Source of Competitive Advantage? The Destination of a New View of the Firm" Diamond Harvard Business Review (Japanese edition) January 2015.