SEGMENTING CUSTOMERS WITH DATA MINING TECHNIQUES

Sabri Serkan Güllüoğlu
Computer Engineering Department
Istanbul Arel University
İstanbul, Turkey
Türkoba Mah. Erguvan Sok. 26-K
Tepekent/Büyükçekmece

***Abstract*—Retail marketers are constantly looking for ways to improve the effectiveness of their campaigns. One way to do this is to target customers with the particular offers most likely to attract them back to the store and to spend more time and money on their next visit. Demographic market segmentation is an approach to segmenting markets. A company divides the larger market into groups based on several defined criteria. Age, gender, marital status, occupation, education and income are among the commonly considered demographics segmentation criteria.**

A sample case study has been done in order to explain the theory of segmentation applied on a Turkish süpermarket chain. The purpose of this case study is to determine dependency on products and shopping habits. Furthermore forecast sales determine the promotions of products and customer profiles. Association rule mining was used as a method for identifying customers buying patterns and as a result customer profiles were determined.

Besides association rules, interesting results were found about customer profiles, such as "What items do female customers buy?" or "What do consumers(married and 35-45 aged) prefer mostly?". For instance, female customers purchase feta cheese with a percentage of 60% whereas male customers purchase tomato with a percentage of 46%. Regarding to customers age, 65 and older customers purchase tea with a percentage of 58%, and customers aged between 18-25 preferred pasta with a percentage of 57%.

Keywords— association rule mining; customer segmentation; market analysis

I. INTRODUCTION

Apriori Algorithm is one of the fastest and earliest tools for Association Mining. In this study, the apriori algorithm for mining Association Rules were used for the large database.

Today, most supermarkets record sales and collect customers' shopping details via a card dedicated to customer which holds customer's personal information (e.g. age, gender, job, income). Data mining helps using this huge amount of data in an efficient way and provides statistical information, thus predicting future customer behavior.

One of the most important data mining methods is, also used in this study, association rule mining. The main purpose of this method is to determine correlations among the sales of items using a set of customer transactions on items. Association rule mining is also known market basket analysis. Market basket analysis helps to understand about the sets of items that are likely to be purchase together.

In this paper, some questions were explored such as "Which products are commonly purchase together?". Generally it is being sought the dependency between two products X and Y. This information will be gathered by processing given a transaction database which is huge in size. This information will be used for store layout, promotions, discounts, catalog design, etc. These results will be analyzed whether they are related to customers' data (age, gender, income, marital status, etc.).

II. STUDY

This part shows all steps from start to end of data mining process, including gathering raw data, normalizing, preparing raw data to be processed, as well as processing data with data mining software. After data is completely processed, results are shown in specific sub-part.

A. Data Preparation

Transaction data is needed to produce association rules, which will be used to find frequent items that are purchase together. These data should be provided from a supermarket. In this study, data is provided by a supermarket chain. Data is extracted from the data pool, filtered by July 2012 from several stores and 300 customers.

Raw data consist of 2 parts, one is customer information which has age, gender, educational status, income attributes and other one is transaction data. Below are some examples for customer information and transaction data.

TABLE I. AN EXAMPLE OF CUSTOMER TRANSACTION RAW DATA

	A	B	C	D	E	F	G	H	I
1	ITEM COD	ITEM NAME	UNIT	ITEM S	ITEM	ITEM SA	ITEM	CUSTOME	ITEM QU
2	3000	POLONEZ KANGAL SUCUK KG	Kilogram	472	16	15,17	0	0,205	1
3	3001	POLONEZ MACAR SALAM KG	Kilogram	254	9,4	8,67	0	0,117	1
4	49376	MAGNUM MINI 360 GR KLASIK-ANTEP-FINDIK	Adet	1	7,5	6,93	1,5	0,094	1
5	44117	SYOSS 750 ML SAC KRM.YOGUN PARLAKLIK	Adet	1	6,9	5,85	2	0,086	1
6	38054	NESTLE CRUNCH 45 GR FRAMBUAZLI	Adet	4	5,6	5,16	0	0,07	1
7	47721	BALONEVI SU BOMBASI BLISTER	Adet	2	5	4,24	0	0,063	1
8	18317	BREF WC TEMIZLEME KUPLERI 100 GR HIJYEN	Adet	1	5	4,23	0	0,062	1
9	16230	CIF KREM 750ML LIMON	Adet	1	4	3,38	0	0,05	1
10	16237	CIF JEL 750ML BANYO	Adet	1	4	3,38	3	0,05	1
11	2646	ERGOR 10 LU YUMURTA	Adet	1	3,2	2,95	0	0,04	1
12	385	MAYDANOZ	Adet	3	3	2,76	0	0,037	1
13	47771	IHE AYVALIK TOSTU EKMEGI 500 GR	Adet	1	2	1,98	0	0,025	1
14	186	DOMATES LUX	Kilogram	#####	2	1,81	0	0,025	1
15	831	OZKAYNAK MADEN SUYU 200 ML*6 CAM	Adet	1	1,8	1,66	0,46	0,022	1
16	16763	IHE.ORGANIK TAM BUGDAY EKMEGI 500 GR	Adet	1	1,8	1,73	0	0,022	1
17	3913	FILIZ MAKARNA KLASIK 500GR BONCUK	Adet	1	0,9	0,82	0,2	0,011	1
18	16503	ETI 37313 PUF RENKLI 18GR	Adet	3	0,6	0,54	0	0,007	1
19	16506	ETI 39312 PUF H.CEVIZLI 18 GR	Adet	3	0,6	0,53	0	0,007	1
20	47149	ETI 14317 CIN PORTAKAL JOLELI BISKUVI 48	Adet	1	0,5	0,46	0	0,006	1

TABLE II. A PORTION OF CUSTOMER INFORMATION RAW DATA

ISBN: 978-1-4799-6376-8/15/$31.00 ©2015 IEEE

	CUSTOMER ID	EXPENDITURE	INCOME	JOB	EDUCATIONAL STATUS	DATE OF BIRTH	GENDER	MARITAL STATUS
2	'5460000000000524'	239,15	400-750	WORKER	PRIMARY	01.03.1957	MALE	MARRIED
3	'5460000000001907'	162,77	400-750	HOUSEWIFE	PRIMARY	15.02.1980	FEMALE	MARRIED
4	5460000000002415'	305,06	750-1000	WORKER	HIGH	01.03.1978	MALE	MARRIED
5	'5460000000003010'	137,04	1000-1500	WORKER	PRIMARY	27.12.1972	MALE	MARRIED
6	'5460000000003240'	200,11	750-1000	TOURISM	HIGH	22.06.1962	MALE	MARRIED
7	'5460000000003823'	43,37	750-1000	TOURISM	UNIVERSITY	10.08.1983	FEMALE	SINGLE
8	'5460000000005230'	108,93	2000+	TRADE	PRIMARY	10.10.1959	MALE	SINGLE
9	'5460000000005797'	106,93	750-1000	WORKER	HIGH	18.08.1977	MALE	MARRIED
10	'5460000000006012'	103,27	400-750	WORKER	HIGH	23.03.1961	MALE	MARRIED
11	'5460000000006033'	226,14	2000+	ENGINEER	UNIVERSITY	07.01.1969	MALE	MARRIED
12	'5460000000006061'	85,42	400-750	STUDENT	HIGH	15.08.1993	MALE	SINGLE
13	'5460000000006151'	140,55	400-750	WORKER	PRIMARY	05.05.1980	MALE	MARRIED
14	'5460000000006330'	124,26	1000-1500	TEACHER	UNIVERSITY	15.04.1978	MALE	MARRIED
15	'5460000000006398'	146,56	2000+	WORKER	PRIMARY	11.08.1955	MALE	MARRIED
16	'5460000000006511'	92,53	1500-2000	WORKER	HIGH	13.09.1969	MALE	MARRIED
17	'5460000000006578'	273,32	750-1000	WORKER	PRIMARY	28.02.1960	MALE	MARRIED
18	'5460000000006827'	166,53	400-750	WORKER	PRIMARY	16.08.1978	MALE	MARRIED
19	'5460000000007878'	59,65	750-1000	RETIRED	HIGH	09.03.1956	FEMALE	MARRIED
20	'5460000000007099'	70,99	1000-1500	HEALTH PERSONNEL	UNIVERSITY	01.02.1971	MALE	MARRIED
21	'5460000000007368'	139,53	1000-1500	OFFICER	UNIVERSITY	14.06.1967	FEMALE	MARRIED
22	'5460000000009025'	120,24	400-750	HOUSEWIFE	PRIMARY	05.06.1972	FEMALE	MARRIED
23	'5460000000008387'	164,16	400-750	WORKER	PRIMARY	15.02.1973	FEMALE	MARRIED
24	'5460000000008733'	40,39	400-750	RETIRED	HIGH	31.03.1955	MALE	MARRIED
25	'5460000000011702'	489,55	2000+	TRADE	HIGH	02.02.1952	MALE	SINGLE
26	'5460000000011989'	184,99	1000-1500	TEACHER	UNIVERSITY	23.04.1968	FEMALE	MARRIED
27	'5460000000009600'	105,4	750-1000	WORKER	PRIMARY	17.10.1963	MALE	MARRIED
28	'5460000000010147'	150,71	750-1000	HOUSEWIFE	PRIMARY	12.05.1974	FEMALE	MARRIED
29	'5460000000012357'	45,24	1000-1500	TEACHER	UNIVERSITY	04.01.1959	FEMALE	SINGLE
30	'5460000000010459'	54,48	750-1000	HOUSEWIFE	PRIMARY	11.02.1958	FEMALE	MARRIED
31	'5460000000010956'	60,16	400-750	WORKER	PRIMARY	25.12.1970	FEMALE	MARRIED
32	'5460000000011041'	118,83	750-1000	ISLETMECI	PRIMARY	28.07.1958	MALE	MARRIED
33	'5460000000011048'	105,98	400-750	HOUSEWIFE	PRIMARY	14.08.1976	FEMALE	MARRIED
34	'5460000000032608'	42,25	1000-1500	RETIRED	PRIMARY	15.12.1958	MALE	MARRIED
35	'5460000000011727'	128,67	1500-2000	ENGINEER	UNIVERSITY	21.06.1964	FEMALE	MARRIED
36	'5460000000012517'	56,48	750-1000	WORKER	HIGH	12.08.1973	FEMALE	SINGLE
37	'5460000000013068'	280,33	400-750	RETIRED	PRIMARY	22.05.1967	FEMALE	MARRIED
38	'5460000000012178'	164,49	1500-2000	KASAP	PRIMARY	02.04.1974	MALE	MARRIED

TABLE III. AN EXAMPLE OF CUSTOMER TRANSACTION RAW DATA

1	ITEM CODE	ITEM NAME	UNIT	ITEM S	ITEM SA	ITEM	ITEM	CUSTC	ITEM Q
2	5819	AIRWICK FM KIT+2 YEDEK(TEM.ES)	Adet	1	22,99	19,5	0	0,062	1
3	5597	SOLO TUV.KAGIDI 32 LI	Adet	1	19,45	16,5	0	0,052	1
4	45343	BINGO AUTOMAT 5 KG ERGUVAN	Adet	1	17,49	14,8	0	0,047	1
5	47057	S.BRITE MIKROFIBER TEMIZLIK BEZI 3 AL 2	Adet	1	12,49	11,6	0	0,033	1
6	14931	PERMATIK BANYO 10LU	Adet	1	12,49	10,6	0	0,033	1
7	4272	YUDUM 2 LT PET AYCICEK YAGI	Adet	1	10,99	10,2	1	0,029	1
8	6103	ORKID ULT.KANAT.NRM. 4 LU ANNE KIZ	Adet	1	9,99	8,47	0	0,027	1
9	44609	DURU DUS JELI 500+250 ML SHINE YAKUT	Adet	1	9,99	8,47	0	0,027	1
10	36510	BLENDAX 700ML SAMP.+180ML KEPEK	Adet	1	8,9	7,54	0	0,024	1
11	5573	SELPAK HAVLU 6 LI %25 IND.	Adet	1	8,9	7,54	0	0,024	1
12	5813	AIRWICK FRESH MATIC MINI KIT LAVENDER	Adet	1	8,79	7,45	0	0,024	1
13	882	RED BULL 250ML KUTU	Adet	2	7,78	7,2	0	0,021	1
14	318	KAVUN	Kilogram	#####	7,75	7,18	0	0,021	1
15	13012	S.BRITE TEMIZLIK BEZI 5 LI	Adet	1	7,25	6,72	0	0,019	1
16	16198	DOMESTOS 2160GR KAR BEYAZI	Adet	1	6,99	5,92	0	0,019	1
17	2393	SUTAS BEYAZ PEYNIR 500 GR	Adet	1	6,99	6,47	0	0,019	1
18	18684	CAYKUR RIZE TURIST CAY 500 GR	Adet	1	6,99	6,47	0	0,019	1
19	46696	CALVE 700 GR KETCAP+600 GR MAYONEZ SET	Adet	1	6,99	6,48	2,6	0,019	1
20	11534	KOROPLAST COP TOR.BUZ.ORTA BOY 55*60	Adet	1	6,49	5,5	0	0,017	1
21	314	KARPUZ	Kilogram	#####	5,99	5,55	0	0,016	1
22	6492	BINGO SOFT 2 LT SENSITIVE	Adet	1	5,99	5,08	0	0,016	1
23	1856	PINAR SUT TAM YAGLI UHT 1000 ML	Adet	3	5,97	5,52	0	0,016	1
24	13007	S.BRITE SUNGER 5+1 YESIL OLUKLU	Adet	1	5,8	4,92	0	0,016	1
25	4009	TAT KONS.MISIR 220GR 3AL 2ODE	Adet	1	5,39	5	0	0,014	1
26	18689	CAYKUR TOMURCUK CAY 125 GR	Adet	2	5,3	4,9	0	0,014	1
27	6264	MR.MUSCLE AQUA MAVI	Adet	1	4,99	4,23	0	0,013	1
28	37085	P.BAHCE 54166 KULLUK OVAL DISK 2 LI	Adet	1	4,59	3,89	0	0,012	1
29	190	DOMATES SALKIM KG	Kilogram	#####	4,52	4,19	0	0,012	1
30	5913	ACTIVEX SIVI SABUN 300ML DOGAL KORUMA	Adet	1	4,49	3,8	0	0,012	1
31	15187	FE CIMBIZ RENKLI YAN UCLU	Adet	1	4,29	3,64	0	0,011	1
32	678	LIPTON EARL GREY CAY 100 GR	Adet	1	4,29	3,97	0	0,011	1
33	43401	TAMEK 830 GR SALCA	Adet	1	3,99	3,7	0	0,011	1
34	6587	MOLPED GUNLUK 40 LI ULTRA LIGHT DEO FLOR	Adet	1	3,99	3,38	0	0,011	1
35	47047	HOBBY 300 ML SOFT KREM	Adet	1	3,99	3,38	0	0,011	1
36	43408	TAMEK 370 CC CAM BIBER SALCASI TATLI	Adet	1	3,99	3,69	0	0,011	1
37	4086	TUKAS TATLI KORNISON TURSU 660 CC	Adet	1	3,69	3,42	0	0,01	1
38	4089	TUKAS BIBERIYE 370 CC	Adet	1	3,59	3,32	0	0,01	1
39	38756	TAMEK 720 CC CAM KARISIK TURSUSU	Adet	1	3,59	3,33	0	0,01	1
40	4615	BIZIM MARGARIN 250 GR PAKET	Adet	2	3,5	3,24	0	0,009	1
41	5924	DURU GOURMET SIVI SABUN 300 ML VISNELI T	Adet	1	3,29	2,79	0	0,009	1
42	37355	BREF POWER AKTIF 51 GR OKYANUS	Adet	1	3,25	2,75	0,7	0,009	1
43	43456	TAMEK 370 CC CAM DILIMLI JALAPENO	Adet	1	3,25	3,01	0	0,009	1
44	1823	NEVZAT KAYSERI MANTI 500GR	Adet	1	3,25	3,01	0	0,009	1
45	4206	KNORR FESLEGENLI VE KEKIKLI SALATA SOSU	Adet	1	3,25	3,01	0	0,009	1
46	424	PATATES TAZE KG	Kilogram	#####	3,22	2,98	0	0,009	1
47	3392	INCI PUL BIBER 70 GR	Adet	1	3,19	2,96	0	0,009	1
48	44323	MAGGI ET BULYON 240 GR YENI	Adet	1	2,95	2,73	0	0,008	1
49	759	MEHMET EFENDI KAHVE FOLYO 100 GR	Adet	1	2,95	2,73	0	0,008	1
50	3457	HUNKAR PILAVLIK PIRINC 1000GR	Adet	1	2,75	2,55	0	0,007	1
51	466	SALATALIK	Kilogram	#####	2,37	2,19	0	0,006	1
52	43770	PAREX STREC FILM 33 MT	Adet	1	2,35	1,99	0	0,006	1
53	11406	IWO COK AMACLI ELDIVEN	Adet	1	2,25	2,08	0	0,006	1
54	3467	HUNKAR KOFTELIK BULGUR 1000GR	Adet	1	2,25	2,08	0	0,006	1

B. *Preperation of Raw Data*

Data mining softwares cannot process raw data, as they need tabular forms in order to mine. It is needed to convert these raw data to such a form so that softwares can process. Below are the examples of converted raw data, ready to process.

Table IV. shows customer informations in group form, in order to make processing the data easier. For example, if a customer has a value "1" for gender, it means this customer is a male, whereas if it has a value "4" for job, he/she is an engineer. Value "3" for expenditure, means customer spent 150-250 TRY.

TABLE IV. CUSTOMER INFORMATION GROUPS

	ATTRIBUTE	OPTIONS	VALUES
1	Gender	Male	1
		Female	2
2	Income	0-750	1
		750-1000	2
		1000-1500	3
		1500-2000	4
		2000+	5
3	Educational Status	Primary School	1
		High School	2
		University	3
4	Marital Status	Married	1
		Single	2
6	Job	Unemployed	1
		Worker	2
		Student	3
		Engineer	4
		Teacher	5
		Tourism-Trade	6
		Retired	7
		Medical Personnel	8
		Officer	9
7	Expenditure	0-50	1
		50-150	2
		150-250	3
		250-350	4
		350+	5

TABLE VI. CONVERTED RAW DATA FOR TRANSACTIONS-1

1	Customer ID	Job	Gender	Marital Status	Age	Educational Status	Income	Expenditure	Pear	Ayran	Spices	Honey, Jam	Diaper
2	'5460000000000524'	2	1	1	5	1	1	3	F	F	F	T	F
3	'5460000000001907'	1	2	1	2	1	1	3	F	F	F	F	F
4	5460000000002415'	2	1	1	3	2	2	4	F	F	F	F	F
5	'5460000000003010'	2	1	1	3	1	3	2	F	F	F	F	F
6	'5460000000003240'	2	1	1	4	2	2	3	F	F	F	T	F
7	'5460000000003823'	6	2	1	2	3	2	1	F	F	F	F	F
8	'5460000000005230'	6	1	1	4	1	5	2	F	F	F	F	F
9	'5460000000006012'	2	1	1	4	2	1	2	F	F	F	F	F
10	'5460000000006033'	2	1	1	3	1	1	3	F	F	F	F	F
11	'5460000000006330'	2	1	1	3	2	3	2	F	F	F	F	F
12	'5460000000006578'	6	1	1	4	1	4	4	F	F	F	F	T
13	'5460000000006827'	3	1	1	2	1	1	3	F	F	F	F	T
14	'5460000000006984'	1	2	1	3	1	1	3	F	F	F	F	T
15	'5460000000007878'	7	2	1	5	2	2	2	F	T	F	F	F
16	'5460000000008733'	7	1	1	5	2	1	1	F	F	F	F	F
17	'5460000000010147'	1	2	1	3	1	2	3	F	F	F	F	F
18	'5460000000010385'	1	2	1	4	2	1	2	F	F	F	F	F
19	'5460000000010459'	1	2	1	5	1	2	2	F	F	F	F	F
20	'5460000000011702'	6	1	2	5	2	5	5	F	F	F	F	F

TABLE VI. CONVERTED RAW DATA FOR TRANSACTIONS-II

	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	Feta Cheese	Pepper	Biscuits	Insecticide	Dish Soap	Cracked Wheat	Chips	Laundry Soap	Bleacher	Tea	Nuts	Chocolat e	Trash Bag
2	T	T	T	F	F	F	F	F	T	F	T	T	F
3	F	T	T	F	T	F	F	F	T	T	T	T	F
4	F	F	F	F	F	F	F	F	F	F	F	F	F
5	F	T	F	F	F	F	F	T	F	F	F	F	F
6	T	F	T	F	T	F	F	T	F	F	T	T	F
7	F	F	F	F	F	F	F	T	F	F	F	T	F
8	F	T	F	F	T	F	F	F	F	F	F	F	F
9	F	F	F	F	T	F	F	F	F	F	F	T	F
10	T	F	T	F	F	F	F	T	F	F	F	T	F
11	F	T	T	F	F	F	F	F	F	F	F	T	F
12	F	F	F	F	F	F	F	T	T	T	T	T	F
13	T	F	T	F	F	F	T	F	F	F	T	T	F
14	T	T	F	F	T	F	F	T	F	F	F	F	F
15	T	T	T	F	F	F	F	F	F	F	F	T	F
16	F	F	F	F	T	F	F	F	F	F	F	F	F
17	T	T	T	F	F	F	F	F	F	F	F	F	F
18	T	F	F	F	F	F	F	F	F	F	F	T	F
19	F	T	F	F	F	F	F	F	F	F	F	T	F
20	F	T	F	F	F	F	F	F	F	F	F	F	F

Table VI. shows converted transaction data. First part of table shows customer information which is explained in Table V. Second part shows customers' transaction information which consists of 95 items, and each of these items has True and False values to indicate whether the customer purchase the item or not. If a customer has "T" for "tomato" it means customer purchase tomato, whereas if there is an "F" value in column "milk" customer did not buy milk.

III. APPLICATION

In the study, SPSS Clementine v12 was used to find the association rules from the dataset.

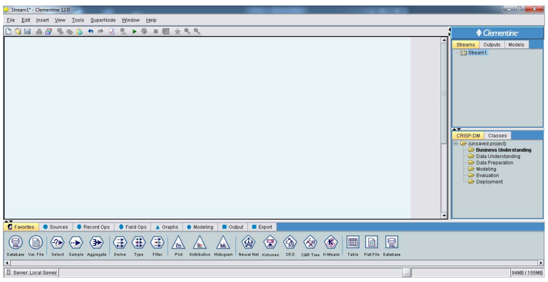

FIGURE I. SPSS CLEMENTINE V12 INTERFACE

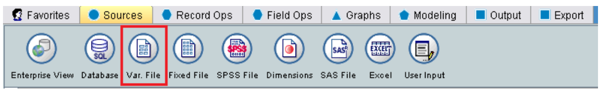

FIGURE II. ADDING SOURCE TO STREAM IN CLEMENTINE

As shown in Figure II, Var. File needs to be added from Sources tab to stream, in order to import transaction data. To specify source file, it can be double-clicked.

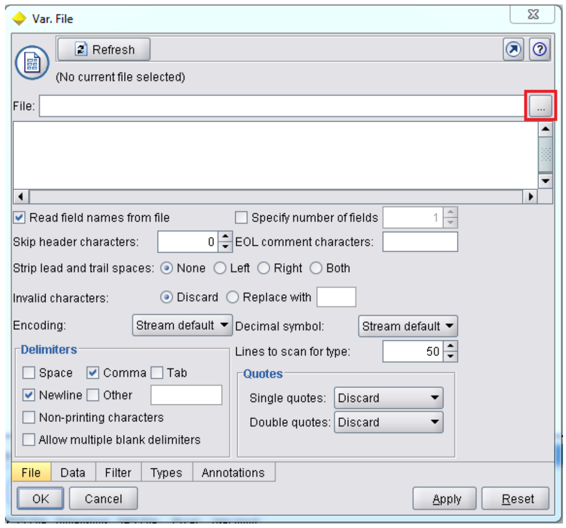

FIGURE III. IMPORTING DATA FROM SOURCE FILE

By clicking highlighted browse button, the file to be imported to software can be chosen. The structure of the file, which is used in this study, is shown in Figure IV.

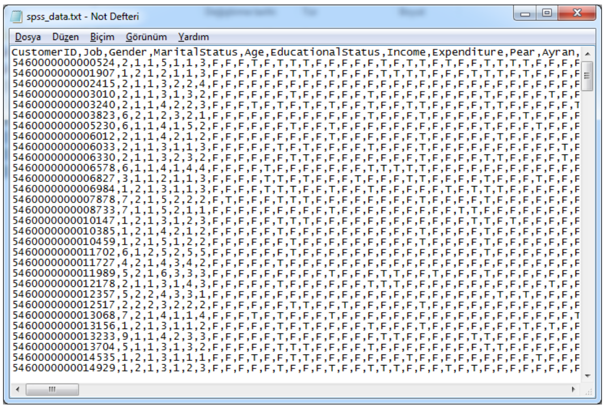

FIGURE IV. STRUCTURE OF THE FILE IMPORTED TO CLEMENTINE

After importing data, a "type" node needs to be added into the stream. This node allows which fields in the data will be used in apriori algorithm to find association rules, and which ones will be ignored.

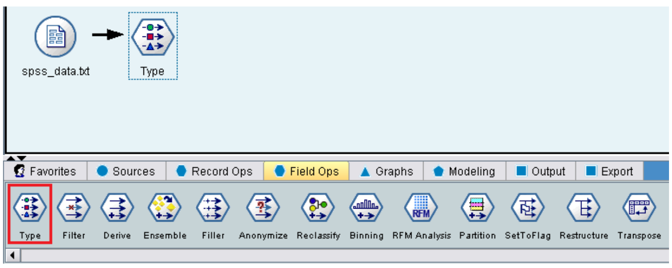

FIGURE V. ADDING TYPE NODE INTO STREAM IN CLEMENTINE

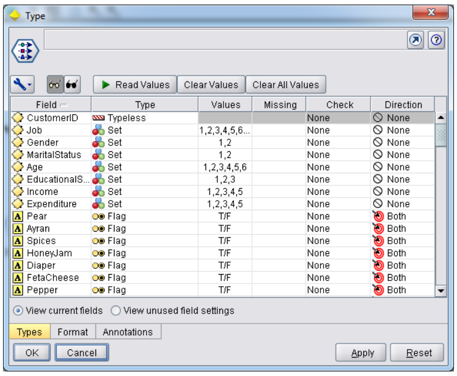

FIGURE VI. VIEWING THE TYPES OF FIELDS IN CLEMENTINE

After adding type node, a "table" node needs to be added into the stream as well. Table node enables to view imported data in tabular form, after executing stream.

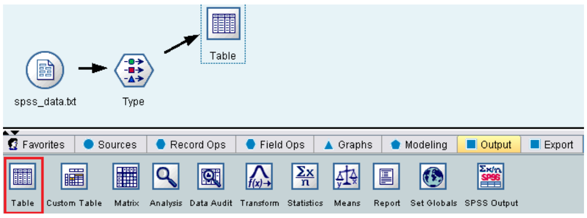

FIGURE VII. ADDING TABLE NODE INTO THE STREAM

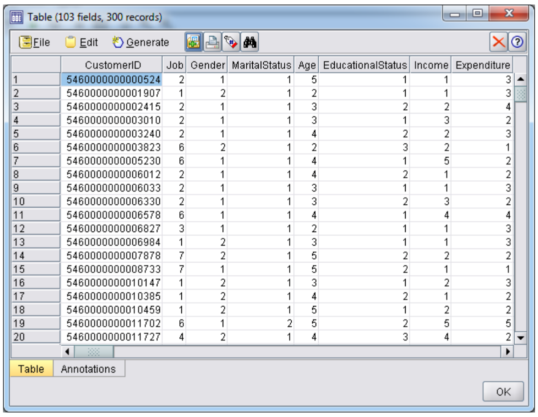

FIGURE VIII. VIEWING THE DATA IN CLEMENTINE

The last step is to add "Apriori" node into the stream. Apriori node allows to produce association rules from the given dataset.

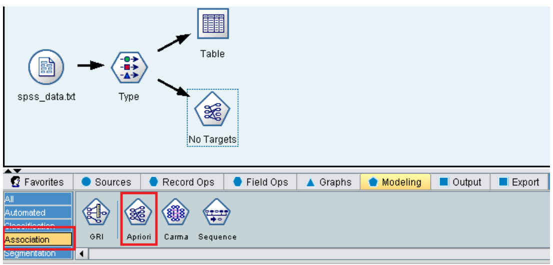

FIGURE IX. ADDING APRIORI NODE INTO THE STREAM IN CLEMENTINE

IV. RESULTS

There are total of 95 items in the dataset. Number of buying amounts for each item is shown in TableVI. as True and False values.

In the study, best known association rule mining algorithm Apriori is used.

To have the most optimized results the parameters are set as follows:

TABLE VII. APRIORI SETTINGS: 1 ANTECEDENT

Build Settings	Value
Maximum number of antecedents	1
Minimum antecedent support (%)	5
Minimum rule confidence (%)	50

The results achieved from using these parameters are shown below:

TABLE VIII. ANALYSIS RESULTS: 1 ANTECEDENT

Analysis	Value
Number of Rules	426
Number of Valid Transactions	300
Minimum Support(%)	5,333
Maximum Support(%)	53
Minimum Confidence(%)	50
Maximum Confidence(%)	91,429

A. Market Segmentation

Geographic, demographic, psychographic and behavioral segmentation can be used in market segmentation. In this research demographic segmentation was explored as sample.

Geographic segmentation separates the market into different geographical units such as nations, regions, states, counties, cities, or even neighborhoods. Psychographic segmentation divides buyers into different groups based on social class, lifestyle, or personality characteristics. Behavioral segmentation divides buyers into groups based on their knowledge, attitudes, uses, or responses to a product.

Demographic segmentation divides the market into groups based on variables such as age, gender, family size, family life cycle, income, occupation, education, religion, race, generation, and nationality.

B. Customer Profile

In the study, besides association rules for each item the customer profile could be extracted from raw data, as it includes customer information like age, gender, income. Customer profile have information like which customer group prefers what kind of items. Below are the figures that demonstrate all customers grouped in different attributes.

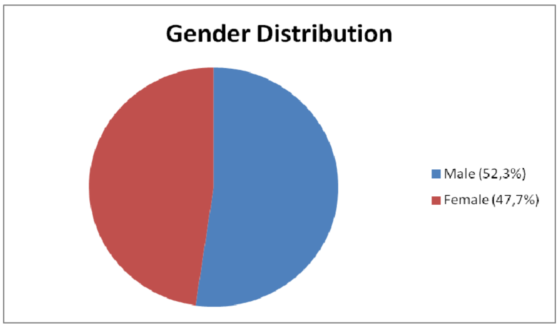

FIGURE IX. GENDER DISTRIBUTION

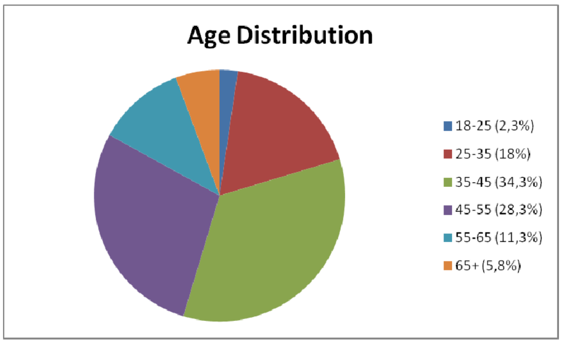

FIGURE IX. AGE DISTRIBUTION

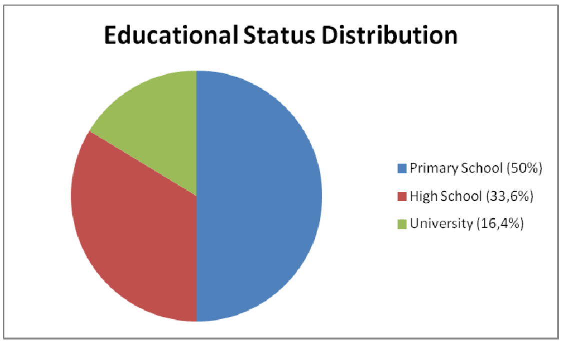

FIGURE IX. EDUCATIONAL STATUS DISTRIBUTION

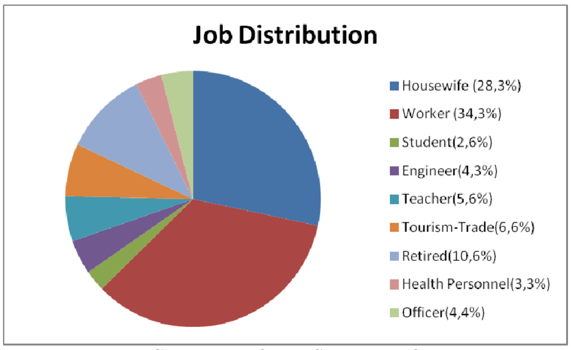

FIGURE IX. JOB DISTRIBUTION

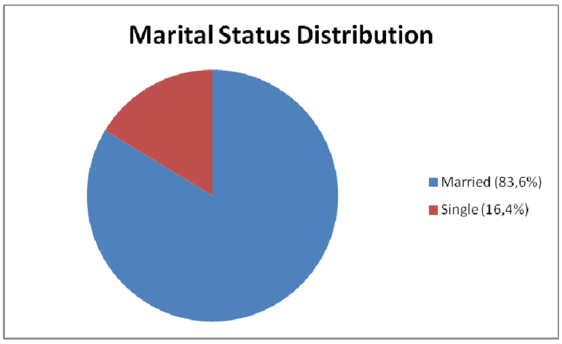

FIGURE IX. MARITAL STATUS DISTRIBUTION

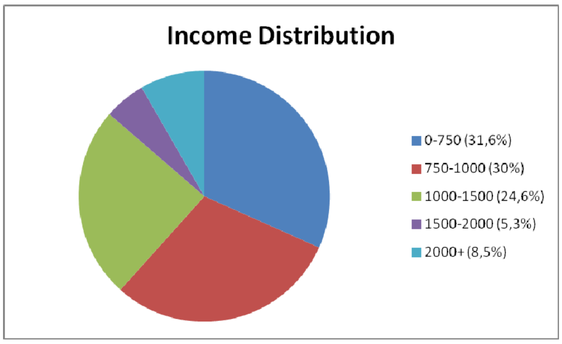

FIGURE IX. INCOME DISTRIBUTION

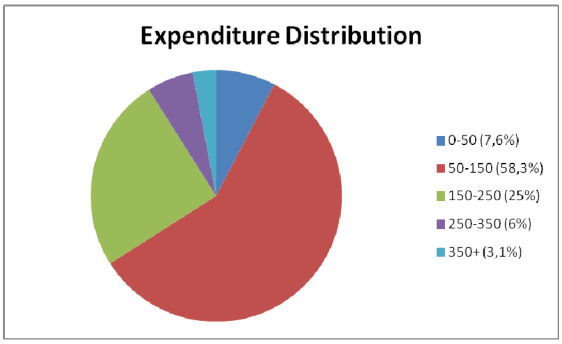

FIGURE IX. EXPENDITURE DISTRIBUTION

V. RESULTS

With the advance of technology, databases are becoming more and more important for current information technology. Databases stores huge amount of data, and data mining allows extracting valuable information from this datasets. There are many techniques to mine these data and association rule mining is one of the most important among these.

Apriori algorithm is one of the most important tools for association rule mining. In this study, apriori algorithm is applied for mining association rules in database of Turkish supermarket chain. This database included customer informations like age, income and gender but without confidential information like name, address and phone. A unique customer ID is used to distinct customers. This database was extracted in July 2012 from several Turkish supermarket chain stores located in Istanbul and following results had been found.

ISBN: 978-1-4799-6376-8/15/$31.00 ©2015 IEEE

The item which is sold the most was feta cheese, which was in 53 percent of all transactions. The second most sold item was tomato and it was present in 49 percent of transactions. These items were followed by, milk and pasta, respectively. For example, customers who purchase cucumber also purchase tomato with a confidence of 88%, having support 36%. It should be noted that, this rule is not same as saying customers who purchase tomato also purchase cucumber with a confidence of 88%. Because customers who purchase tomato also purchase cucumber with a confidence of 64%. It is also should be noted that support percentage doesn't change between two rules.

In this study, not only 1 antecedent association rules are generated, but also 2 and 3 antecedents are generated as well. For example, for 2 antecedents, customers who purchase cucumber and pepper also purchase tomato with a confidence of 94%. This may be interpreted as, the items which are used in salad making are purchase together. Customers who purchase egg and yoghurt also purchase feta cheese with a confidence of 78%. This is not a surprising association rule because these items can be assumed as staple food. Another example is that customers who purchase coke and biscuits also purchase chocolate with a confidence of 73%.

Besides association rules, interesting results are found about customer profiles, like "What are the items that are purchase most by female customers?" or "What do married and 35-45 aged customers prefer mostly?". For example, female customers purchase feta cheese most, with a percentage of 60% whereas male customers purchase tomato with a percentage of 46%. Regarding to customers age, 65 and older customers purchase tea most, with a percentage of 58%, and customers aged between 18-25 preferred pasta with a percentage of 57%. Feta cheese is the most sold item regardless to marital status of customers with a percentage of 57% and 52% for single and married customers respectively. This is also not a surprising result, as feta cheese is one of the staple foods. Pasta is one of the most preferred items, considering marital status and gender together, when customer is single regardless of gender.

With the help of these results, much useful information can be found. For instance, most sold items can be in promotion together to increase sales by attracting customers. Least sold items can also be combined with most sold items to increase sales of it. Moreover, store layouts can be optimized to increase sales, shopping time of customers.

REFERENCES

I. L. B. Cristopher, "Mining Rules In Single-Table and Multiple-Table Databases", 2002.

II. R .Agrawal, R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases", Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), Santiago, Chile, September 1994.

III. G. Shaw, "Discovery & Effective Use Of Quality Association Rules In Multi-Level Datasets", 2010.

IV. H. Mannila, "Methods and Problems in Data Mining", Proceedings of the 6th International Conference on Database Theory, 1997.

V. M. Berry, G. Linoff, "Data Mining Techniques for Marketing, Sales, and Customer Relationship Management", Second Edition, John Wiley & Sons, 2004.

VI. D. E. Sharp, "Customer Relationship Management Systems Handbook", CRC Press, 2003.

VII. G. Bukhbinder, M. Krumenaker, A. Phillips, "Insurance Industry Decision Support: Data Marts, OLAP and Predictive Analytics", 2005.

VIII. G. Galfond, "Data Mining Can Unearth A Competitive Edge", October 6, 1997.

IX. R. Wirth, J. Hipp, "CRISP-DM: Towards a Standard Process Model for Data Mining", Proceedings of the Fourth International Conference on the Practical Applications of Knowledge Discovery and Data Mining, Manchester, UK, 2000.

X. D. T. Larose, "Discovering Knowledge in Data: An Introduction to Data Mining", John Wiley & Sons, 2005.

XI. B. Moxon, "Defining Data Mining", DBMS Data Warehouse Supplement, August, 1996.

XII. R. Agrawal, A. Bollinger, T.Mehta, M. Shafer, J. and Srikant, "The Quest Data Mining System", Proc. of the 2nd lnt'l Conference on Knowledge Discovery in Databases and Data Mining, Portland, Oregon, August 1996.

XIII. R. Agrawal, T. Imielinski, and A.Swami, "Database Mining: A Performance Perspective", IEEE Transactions on Knowledge and Data Engineering, Special issue on Learning and Discovery in Knowledge-Based Databases, 5(6): 914-925, December 1993.

XIV. G. Lefait, T. Kechadi, "Customer Segmentation Architecture Based on Clustering Techniques" Digital Society, ICDS'10, Fourth International Conference, 10-02-2010.

XV. P. Kotler, G. Armstrong, "Principles of Marketing", Prentice Hall, 11th Edition, 2005.