BUSINESS INTELLIGENCE MODEL FOR CUSTOMERS ANALYSIS FOR A SALES COMPANY

LJILJANA KAŠĆELAN, Faculty of Economics, Podgorica
BILJANA RAJKOVIĆ, T-Mobile, Montenegro

***Abstract**: This paper defines a theoretical framework for developing business intelligence support systems for customers analysis for a sales company. It will contribute significantly to the quality of analysis of certain segments of the market and to efforts to make best decisions in this area. This paper represents a theoretical basis for a practical implementation of all system components. A concrete company example shows that the suggested business intelligence model is producing expected results.*
***Key words:** business intelligence systems, data warehouse, data mart, star schema, customer analysis, custommer segmentation, etl, olap, oracle tools.*

***Sadržaj:** U ovom radu je definisan konceptualni okvir za razvoj sistema poslovne inteligencije za analizu kupaca trgovinskog preduzeća, koji će bitno doprijeniti kvalitetu analize pojedinih segmenata tržišta i donošenju pravih poslovnih odluka u ovoj oblasti. Rad predstavlja teorijsku osnovu za praktičnu realizaciju svih komponenti sistema. Na primjeru konkretnog trgovinskog preduzeća je pokazano da predloženi model sistema poslovne inteligencije daje očekivane rezultate.*
***Ključne riječi**: sistemi poslovne inteligencije, skladište podataka, data mart, zvjezdasta šema, analiza kupaca, segmentacija kupaca, etl, olap, oracle alati.*

(JEL clasification: M21)
Original scientific paper; Recived: Mart 12, 2008

1. Introduction

In the modern market economy everything is changing and the frequency and rate of diffusion of changes is increasing. The main characteristics of the modern economy are business globalization and an increase of the level of competition. In the increasingly complex management environment quality, innovation and flexibility become business imperatives.

Business intelligence systems provide support to organizations that are analysing changes of market trends, changes in customers' behaviour and their buying power, changes of customer needs, internal strength and weaknesses as well as external opportunities and threats.

In a competitive market atmosphere it is necessity for organizations to have timely and correct information about customers needs, so that they could provide a rapid response to the altered demands. Business intelligence systems provide that kind of information based on which organizations will introduce new products and services on the market - in that way anticipating the needs of customers.

The objective of this paper is to model a business intelligence system for customers analysis for a sales company, which includes:

- Basic decision making processes for customer analyses of sales company are identified.
- Information sources for identified decision making processes for sales analyses are defined.
- Data mart's design which supports identified decision making processes is drafted.
- The concept and methodology of Data mart's realization using business intelligence tools is defined.
- OLAP models which support the identified decision making processes are realized.

The basic goal of this paper is a conceptual framework for designing and developing business intelligence system for customer analysis, as well as confirmation of theoretical results on a concrete example of sales company.

Applying chosen methods should give better insight in the potential of their implementation in the area of customer analysis for a sales company.

2. Identification of Decision Models

The customer holds the central place in the modern economy. The customer is not only a consumer of our products and services, he also represents data. The data on who is the buyer of the product is more important than the act of buying. A segmentation of the market refers to efforts to divide the market into segments with common characteristics.

As far as the purpose of definition of target markets is concerned, we can use different strategies of segmentation:

- Geographic segmentation-deals with allocations of customers according to different geographic variables like city area, municipality, town, region, state, postal code, climate, neighbourhood etc.
- Demographic segmentation-deals with allocations of customers according to demographic variables like gender, age, education, occupation, religion, family size, family lifecycle, income, race, nationality etc.
- Psychographic segmentation-deals with allocations of customers according to their social status, lifestyle or personal characteristics. Psychographic variables include activities, interests, attitudes, values etc.
- Behavioural segmentation-deals with allocations of customers according to their knowledge, attitude, usage rate or reaction to products or services.
- Segmentation based on customer status-deals with allocations of customers into three categories: regular (permanent), potential and ex-customers.
- Geodemographic segmentation- is based on the assumption that people share similar demographic characteristics with their near neighbourhood, like taste, system of values, shopping habits etc. The results of this segmentation are demographic information summarized according to geographic variables, like city areas, towns, municipalities, regions etc.
- Segmentation based on benefits deals with allocations of customers according requested benefits like additional services, special services, special prices, delivery term etc.
- B2B-Business To Business segmentation could be based on the same variables as in the customers segmentation (B2C).

There is no unique way of market segmentation, but the best results can be achieved by employing good combinations of different segmentation strategies.

A customers analysis of a sales company could be made in many different ways by combining segmentation strategies. In this paper, the accent is on customer analysis according to geographic and demographic variables. The business intelligence system for customer analysis of a sales company in this paper should afford answers to the following questions: Who are our customers? Where do they live? Where are they buying? What are their occupations and gender? What is their buying power? What kind of age structure of our customers is etc. All these analyses could be done according to time and geographic variables of the shopping process.

By recognizing specific needs of each segment we allow organizations to develop highly specific marketing programmes, improve their current products, determine the best price of their product etc. This segmentation becomes a starting point the organizations use to allocate resources during development of product, marketing and services.

3. Data Mart Implementation

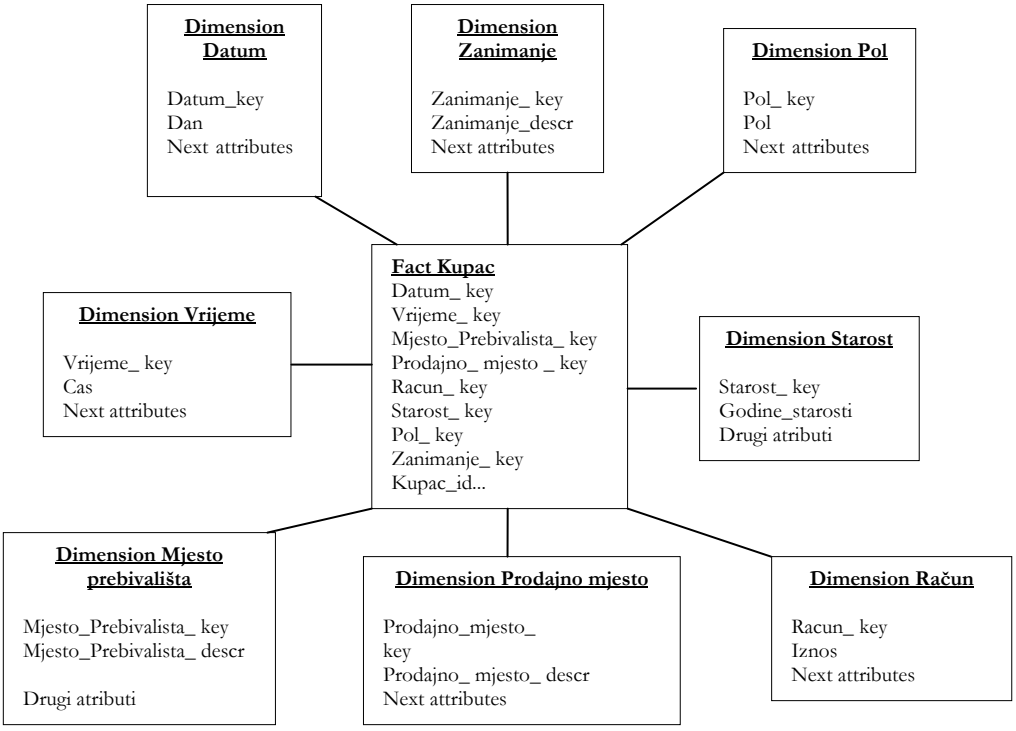
The main goals of Data mart projects are to allow the management of the company to make better business decision within the processes of customer's analysis. It will contribute significantly to the quality of analysis of certain segments of the market and to efforts to make best decisions in this area. One of the goals of Data mart projects is to provide all necessary conditions for realization of OLAP models. These decision models will be providing:

- Customer analysis according to demographic variables: gender, age, occupation,
- Customer analysis according to geographic variables: street, city area, town, municipality, region, postal code etc.
- Customer analysis according to geodemographic variables which assumes summarizing demographic values by geographic units.
- Customers analysis in predefined period and comparative time analysis.
- Customers' habit analysis according to retail shop, time of shopping, buying power (Where are they buying, when and how much they spend?).

Data mart will be implemented in a way that will provide its full integration with the integrated Data warehouse, and its integration with the standard OLAP tools. The drafted Data mart will provide access to endusers via internet or intra net, with full security of data.

Data Mart is based on data. The basis of this dimensional model is a star schema. The star schema consists of a fact table which references dimensional tables. This dimensional model will enable analytical ROLAP processing.

The central part of the star schema is the '*Kupac*' fact table, which is connected to dimensional tables through relations. The star schema is a type of modelling which is used for designing Data mart at the logic level.



Picture 3.1 Star schema 'Kupac'

By creating a definition of dimensions, we create a definition for dimensional objects and dimensionnal tables. Dimensional object consists of levels and hierarchies defined between those levels. A level represents a level of aggregations of data and a hierarchy represents relations between those levels (parent –child). When designing the fact table for each dimension we choose the level of granularity, which is in accordance with the level of chosen fact granularity. After that for each dimension we will identify the production key and create a warehouse key. The warehouse key is an artificial key that allows us to preserve the dimension history.

Than we have to determine levels of granulation of entities' attributes which are related to production keys. Than we create hierarchies of attributes which enable data aggregation according to defined levels of aggregation.

The fact table is related to dimensions exclusively via warehouse keys, which are defined for each dimension. This way the performance of executing queries will be improved, since warehouse keys are normally shorter than unique keys.

The fact table consists of columns that represent measures of facts and foreign key columns that represent relations between fact table and dimensions. Attributes of this table will represent quantitative indicators of business in customer analyses for a sales company. The primary key of the fact table will be defined over the group of foreign key columns.

The data in tables (fact and dimensions) are denormalized, which enables analytical queries to process an enormous quantity of data.

In this paper, for the purpose of realization of the designated Data mart, we will define eight dimensions: '*Datum*', '*Vrijeme*', '*Pol*', '*Zanimanje*', '*Prodajno mjesto*', '*Mjesto prebivališta*', '*Starosna struktura*' and '*Račun*'.

The '*Prodajno mjesto*' dimension is determined by the structure of a sales company. The hierarchy of this dimension makes an aggregate analysis according to organizational units, city area, town and region possible.

The '*Račun*' dimension allows for an analysis of the customers' buying power and sales analysis according to time and geographic dimensions. The buying power of customers influences the realization of personal spending. The buying power represents income which customer could set aside for buying products or services. The '*Grupaiznos*' hierarchy makes it possible for business people to monitor the aggregation of data on buying power according to different categories which are of interest as far as the analysis is concerned.

The 'Mjesto prebivalista' dimension enables customer analysis according to geographic variables. The hierarchy enables data aggregation according to city area, municipality, and the region of customer's residential address.

The '*Starosna struktura*' dimension makes a customer analysis across according to demographic variable age possible. The '*Grupagodište*' hierarchy enables data aggregation according to age groups.

The realization of Data mart includes a selection and description of the Warehouse builder tools, definition phase (definition of dimensional models, aggregations, data sources and ETL procedures), generation phase (configuration, validation, script generation and objects generation), loading phase (defining a sequence of execution of ETL procedure scripts).

The suggested Data mart model is realized through application of Oracle tools: Oracle9i RDBMS, Oracle9i Warehouse Builder. We will use Oracle Discoverer for realization of identified decision models.

4. Case Study

A concrete company example shows that the suggested business intelligence model is producing expected results. For purposes of analysis within this Case Study we used data from financial bookkeeping of a sales company from Podgorica. This company offers in its assortment a variety of types of products such as cosmetics products for adults and children, makeup products, napkins for children and adults, products for health care etc. This company has a chain of stores, which consists of seven stores in Podgorica, Budva, Nikšić, Berane, Bijelo Polje, Bar and Igalo, with plans to extend to another town in Montenegro. During 2006 and 2007 the company has organized a prize game for customers for the purpose of marketing presentation of the company and in order to collect customers' data. These reports were entered in all sales stores in the '*Sales-&retalis*' applicative system. For the purpose of Data mart realization we used over 40,000 records from this period. In order to protect the identity of customers, we have not used their personal names. As far as the hardware and software used for this case study are concerned we met the following requirements:

- We used notebook Fujitsu Siemens Lifebook S series, model: S7020D, with processor Intel Pentium, 1.86 GHz, 1 GB RAM and 74.5 GB hard disc
- Operative system- Microsoft Windows XP Professional Service Pack 2.
- Oracle9i RDBMS 9.2.0.7.0, as well as tools from its environment:
  - Oracle Warehouse Builder 9.0.2
  - Oracle9i Discoverer ver. 9.0.2
  - Oracle9i Designer ver. 9.0.2
  - SQLPlus 9.2.0.7.0
  - SQLoader.

Initial fulfilling of Data warehouse consists of several steps:

Importing Case study's data in Oracle relational database-company data from export flat files (from Oracle database of the sales company), which are imported in the local Oracle relational database via the SQLoader (about 40,000 records from 2 fiscal years).

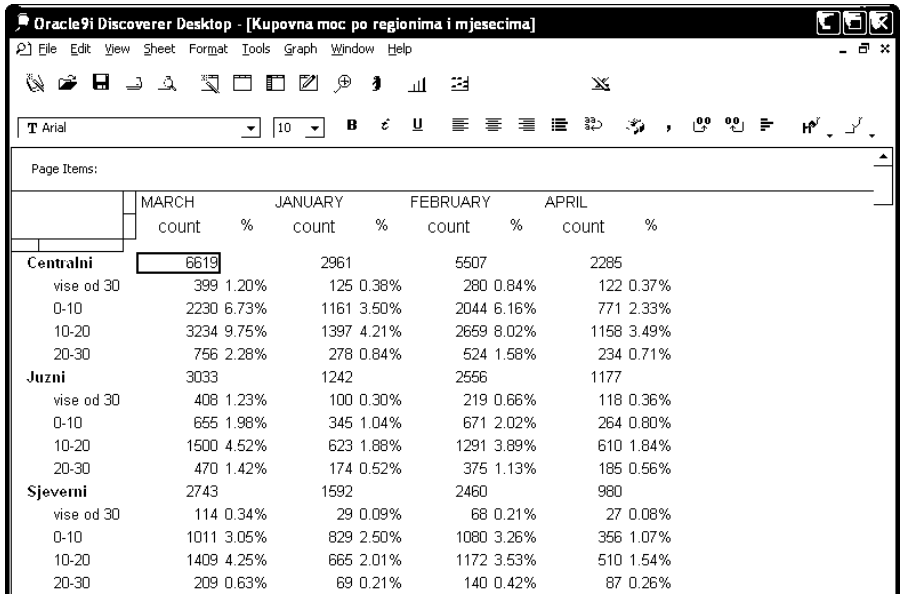
Executing the ETL procedures - while observing the sequence of loading data into the Warehouse target module '*Kupci_TP* ' we started PL/SQL scripts for extracting, transforming and loading data into defined dimensions and fact. All mapping procedures are successfully finished, thereby we initially fill the target warehouse schema. The duration the initial loading of the Warehouse structures i.e. mapping procedures was excellent and was less than 3 sec.

Verifying - the success of the process of initial loading is verified via SQLPlus.

The implemented Data mart was a basis base for generation of models in Oracle Discoverer. The models in the '*Kupci*' fact table, which has 33,158 records, are generated on the average in less than 3 sec. All operations of drill-down and pivot, drill-up are executed immediately.

The *Picture 4.1* shows a multi-dimensional model for analysis of customer's buying power

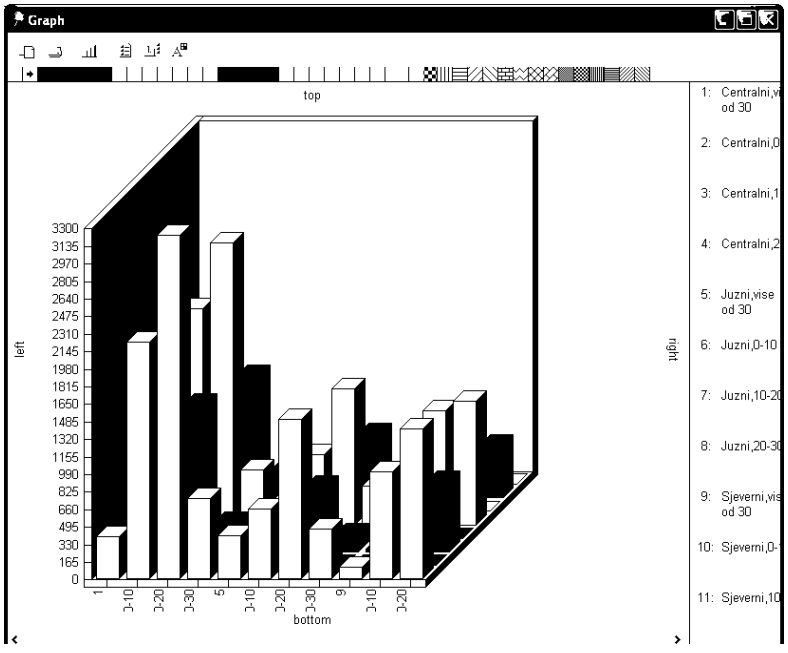
according to retail shops on the monthly level. From this model we can see that the biggest number of customers with the biggest buying power (amount>30 euros) has been buying in the store in Budva in March. This could indicate a significant increase of sales in March, probably because of the 8 March, keeping in mind the fact that this company is mostly selling cosmetics products. Likewise, we can see that the store in Budva has more customers than the store in Podgorica, what could indicate harsher competition in Podgorica or insufficient coverage of the Podgorica municipality in the geographic sense (Podgorica is significantly bigger than Budva). In addition, we could deduce that the high buying power of customers in Budva is related to sales and the increase prices of real estate along the coast.



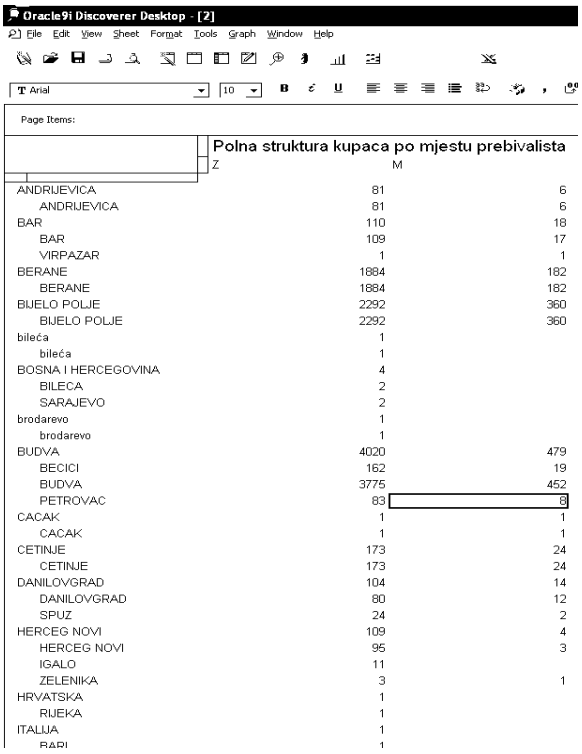
	MARCH		JANUARY		FEBRUARY		APRIL	
	count	%	count	%	count	%	count	%
Centralni	6619		2961		5507		2285	
vise od 30	399	1.20%	125	0.38%	280	0.84%	122	0.37%
0-10	2230	6.73%	1161	3.50%	2044	6.16%	771	2.33%
10-20	3234	9.75%	1397	4.21%	2659	8.02%	1158	3.49%
20-30	756	2.28%	278	0.84%	524	1.58%	234	0.71%
Juzni	3033		1242		2556		1177	
vise od 30	408	1.23%	100	0.30%	219	0.66%	118	0.36%
0-10	655	1.98%	345	1.04%	671	2.02%	264	0.80%
10-20	1500	4.52%	623	1.88%	1291	3.89%	610	1.84%
20-30	470	1.42%	174	0.52%	375	1.13%	185	0.56%
Sjeverni	2743		1592		2460		980	
vise od 30	114	0.34%	29	0.09%	68	0.21%	27	0.08%
0-10	1011	3.05%	829	2.50%	1080	3.26%	356	1.07%
10-20	1409	4.25%	665	2.01%	1172	3.53%	510	1.54%
20-30	209	0.63%	69	0.21%	140	0.42%	87	0.26%

Picture 4.1 Multi-dimensional model for analysis of customer's buying power

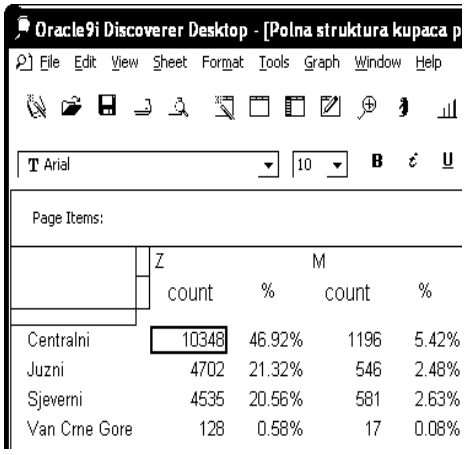
The *Picture 4.2* shows a graphical presentation of the multidimensional model used for analysis of the buying power of customers.



Picture 4.2 Graphical presentation of the model for analysis of customer's buying according to regions

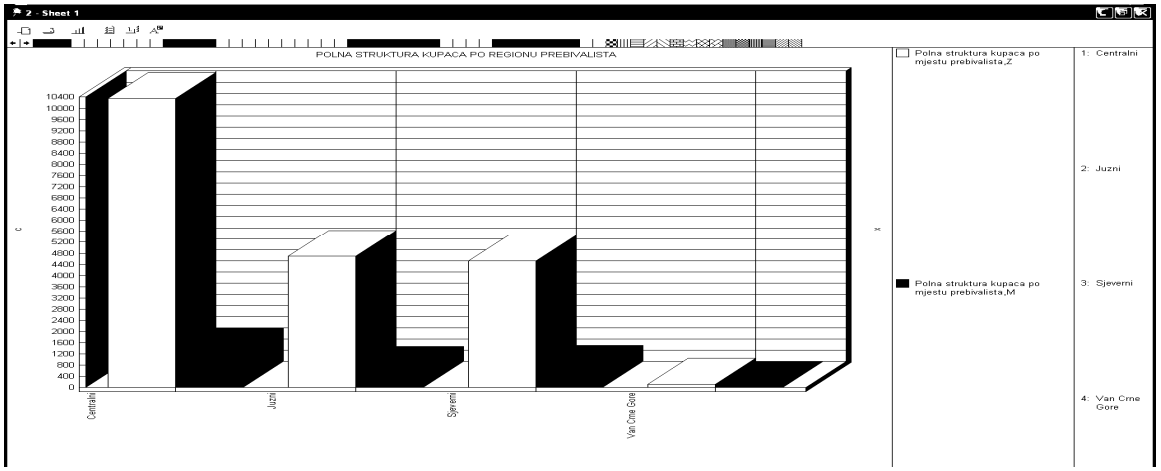


The *Picture 4.3* shows a multi-dimensional model for analysis of customers' gender structure for the sales company according to customer's residential address. From the model we can see that the number of female customers exceeds the number of male customers, as far as the customers with residential addresses within Montenegro are concerned. When we analyze customers that are not from Montenegro we can see that the number of male and female customers is almost the same.



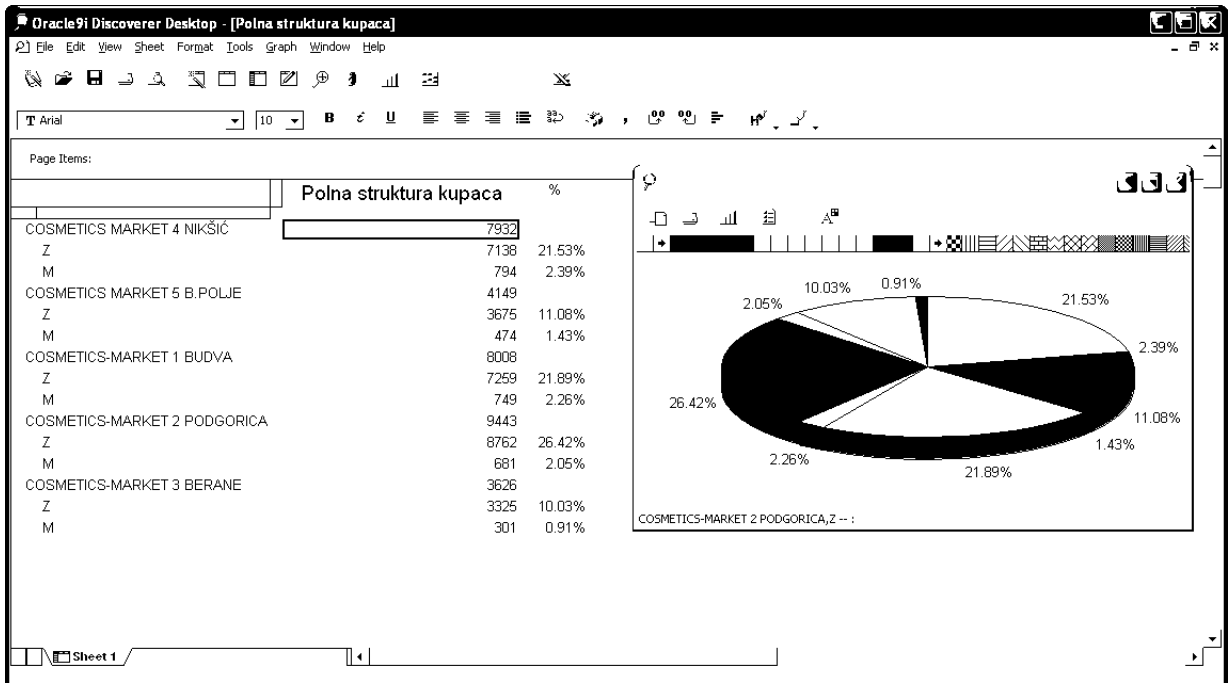
Picture 4.3 Model for customer's gender structure analysis according to customer's residential address (region)

The *Picture 4.4* shows a graphical presentation of the model for customer's gender structure analysis.



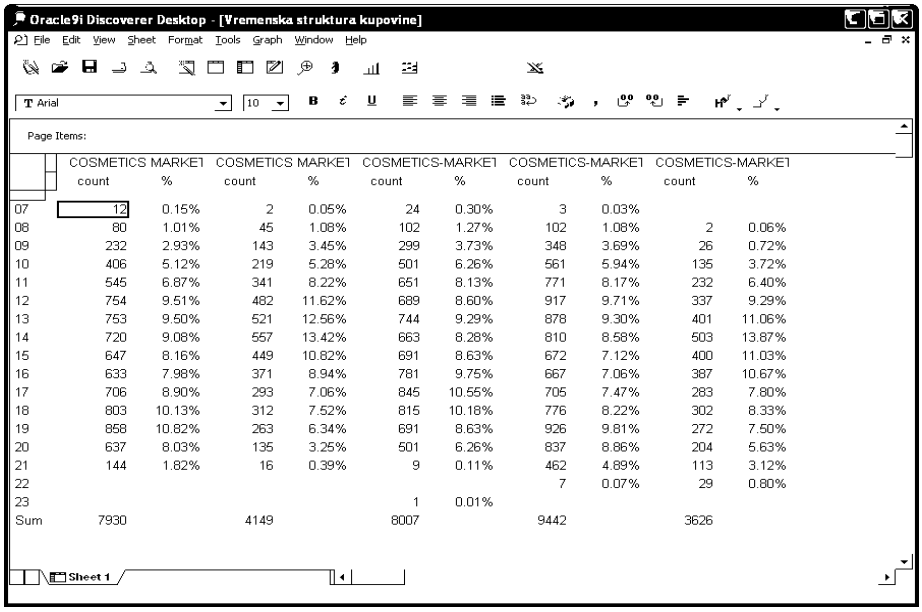
Picture 4.4 Graphical presentation of model for customer's gender structure analysis according to region

The *Picture 4.5* shows a multidimensional model for customer's gender structure analysis of a sales company according to shop locations.

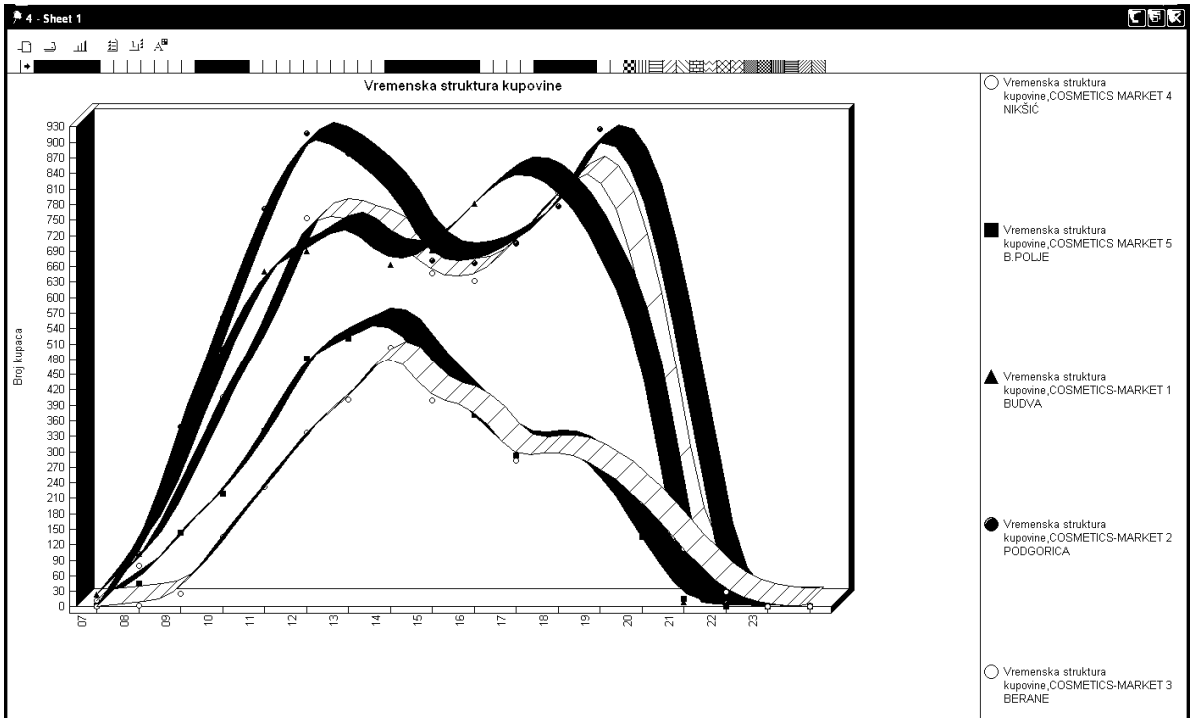


Picture 4.5 Model for customer's gender structure analysis according to retail shop and graphical presentation

The *Picture 4.6* shows a multi-dimensional model for analysis of time structure of sales by shops. From this model we can conclude that the customers in Niksic are mostly buying at 6 pm and 7 pm, in Bijelo Polje at 1 pm and 2 pm, in Budva at 5 pm and 6 pm, in Podgorica at 12 am and 8 pm, in Berane at 2 pm... These conclusions can be used for optimization of human resources, marketing promotions of new products, etc. Based on these conclusions the management can plan the working hours in one shift or in two. This model allows comparative time analyses and improves the chances of making best business decision in the area of tactical management. The model allows us to organize the working hours, plan vacations etc.

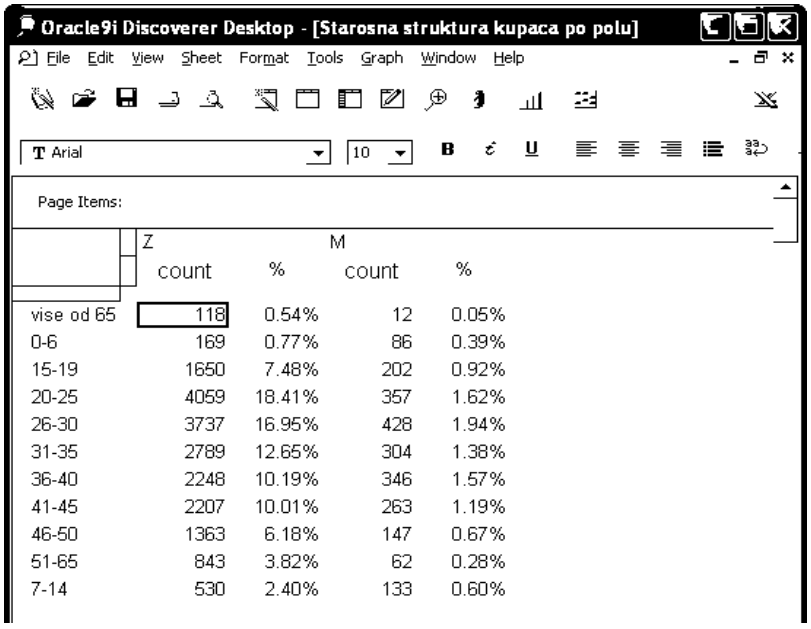


Picture 4.6. Model for analysis of time structure of sales by shops



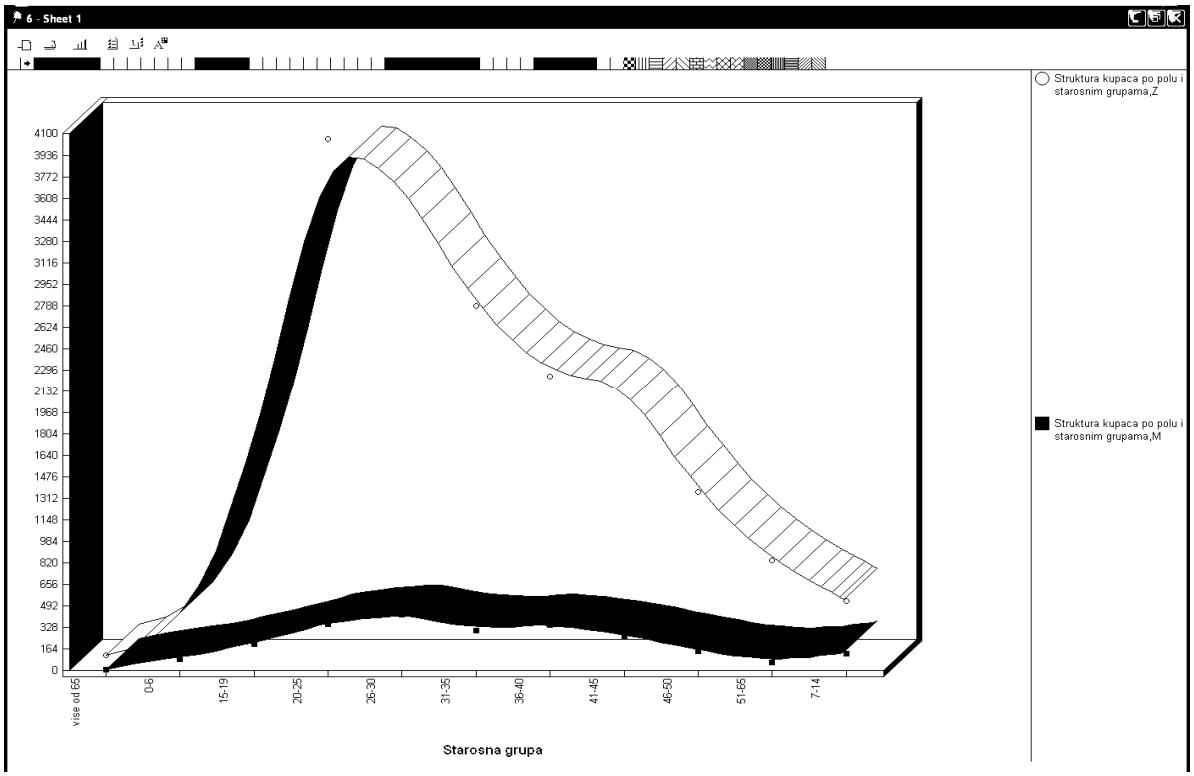
Picture 4.7. Graphical presentation of model for analysis of time structure of sales by shops

The *Picture 4.8* shows a multi-dimensional model for demographic analysis of customers according to gender and age structure. From this model we can conclude that the biggest customer's segment consists of female customers, age between 20 and 25. Based on this conclusion we can made decisions on opening new shops closely to some university units etc. In addition, based on this conclusion we could make a decision to store and purchase more products used by the largest segment of customers. Furthermore, we could make a decision about concrete marketing activities that will be focused on a particular segment in order to recruit these customers.



	Z		M	
	count	%	count	%
vise od 65	118	0.54%	12	0.05%
0-6	169	0.77%	86	0.39%
15-19	1650	7.48%	202	0.92%
20-25	4059	18.41%	357	1.62%
26-30	3737	16.95%	428	1.94%
31-35	2789	12.65%	304	1.38%
36-40	2248	10.19%	346	1.57%
41-45	2207	10.01%	263	1.19%
46-50	1363	6.18%	147	0.67%
51-65	843	3.82%	62	0.28%
7-14	530	2.40%	133	0.60%

Picture 4.8 Model for demographic analysis of customers according to gender and age structure



Picture 4.9 Graphical presentation of the model for demographic analysis of customers according to gender and age structure

The *Picture 4.10* shows a multi-dimensional model for demographic analysis of customers according to their occupation and geographic region. From this model we could conclude for instance that students make up the largest segment in all the regions. Based on that we could make a decision to open new shops close to some university units.

The *Picture 4.11* shows a multi-dimensional model for customers analysis according to residential address and shop locations. From this model we could conclude where customers live and where they are buying things. For instance, if we have information on locations in particular city areas this model will allow us to make decisions on opening new shops in the areas where customers live but where there are no shops yet.

5. Conclusion

The business intelligence model for customer analysis for a sales company based on concepts such as Data warehouse and tools such as Warehouse builder, OLAP and business intelligence tools has produced expected results. The implemented model could be basis for realization of business intelligence models for other companies. This paper is giving insight in the potential of the Oracle tools for realization of all components of a business intelligence system.

References:

[1] Balaban N., Ristić Ž.: '*Poslovna inteligencija*', Ekonomski fakultet, Subotica, 2006

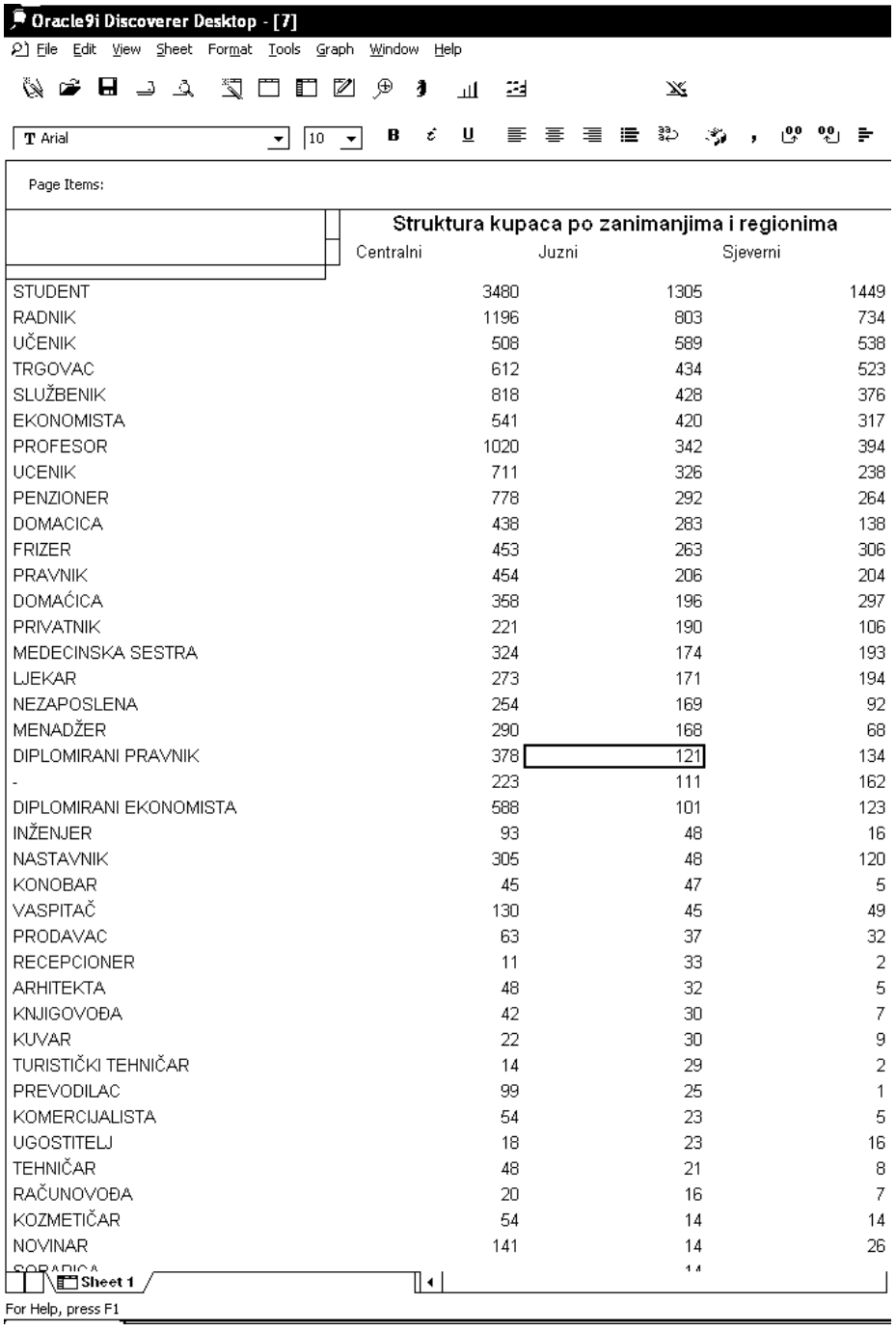
[2] Chenoweth T., Luis R., Schuff D., '*A method for developing Data Marts*', Communications of the ACM, Dec. 2003, Vol46. No.12

[3] Kašćelan, Lj., '*Kreiranje OLAP modela u Oracle Discoverer-u* ', INFO-M, Br. 13/2005, str. 18-22

[4] Pendse, N., '*What is OLAP?* ', OLAP Report, 1998, http://www.olapreport.html/FASMI.HTM

[5] *Oracle Warehouse Builder User's Guide 10*g *Release 1 (10.1)*, PDF version, http://download.oracle.com/docs/pdf/B13916_04.pdf , Oct 2005

[6] 'Rječnik marketinga', 1993, Masmedia, Zagreb.


Oracle9i Discoverer Desktop - [7]

File Edit View Sheet Format Tools Graph Window Help

T Arial 10 B I U

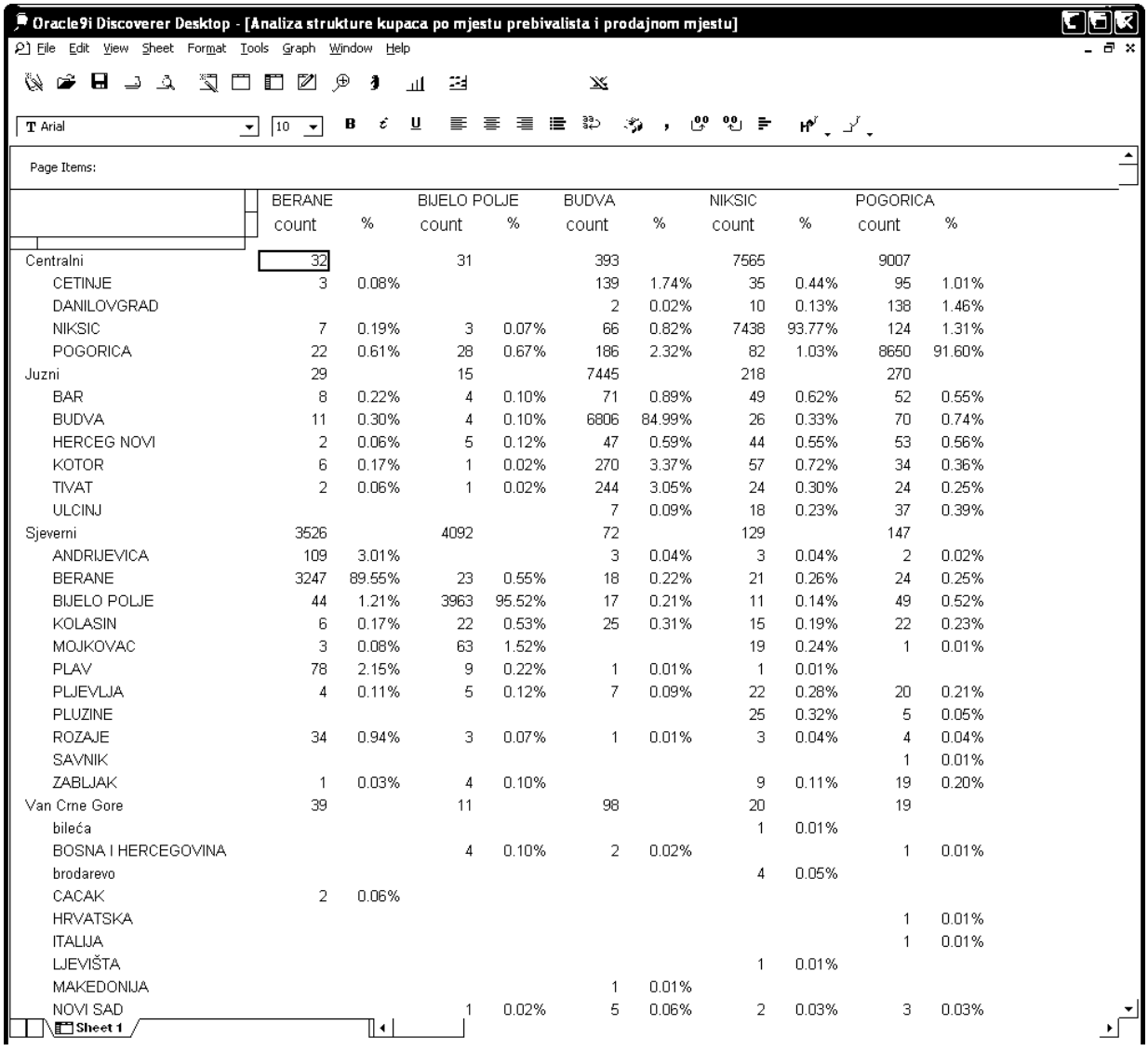
Page Items:

	Struktura kupaca po zanimanjima i regionima		
	Centralni	Juzni	Sjeverni
STUDENT	3480	1305	1449
RADNIK	1196	803	734
UČENIK	508	589	538
TRGOVAC	612	434	523
SLUŽBENIK	818	428	376
EKONOMISTA	541	420	317
PROFESOR	1020	342	394
UCENIK	711	326	238
PENZIONER	778	292	264
DOMACICA	438	283	138
FRIZER	453	263	306
PRAVNIK	454	206	204
DOMAĆICA	358	196	297
PRIVATNIK	221	190	106
MEDECINSKA SESTRA	324	174	193
LJEKAR	273	171	194
NEZAPOSLENA	254	169	92
MENADŽER	290	168	68
DIPLOMIRANI PRAVNIK	378	121	134
-	223	111	162
DIPLOMIRANI EKONOMISTA	588	101	123
INŽENJER	93	48	16
NASTAVNIK	305	48	120
KONOBAR	45	47	5
VASPITAČ	130	45	49
PRODAVAC	63	37	32
RECEPCIONER	11	33	2
ARHITEKTA	48	32	5
KNJIGOVOĐA	42	30	7
KUVAR	22	30	9
TURISTIČKI TEHNIČAR	14	29	2
PREVODILAC	99	25	1
KOMERCIJALISTA	54	23	5
UGOSTITELJ	18	23	16
TEHNIČAR	48	21	8
RAČUNOVOĐA	20	16	7
KOZMETIČAR	54	14	14
NOVINAR	141	14	26
SORADICA		14	

Sheet 1

For Help, press F1

Picture 4.10. Model for demographic analysis of customers according to their occupation and region



Oracle9i Discoverer Desktop - [Analiza strukture kupaca po mjestu prebivalista i prodajnom mjestu]

File Edit View Sheet Format Tools Graph Window Help

Arial 10

Page Items:

	BERANE		BIJELO POLJE		BUDVA		NIKSIC		POGORICA	
	count	%	count	%	count	%	count	%	count	%
Centralni	32		31		393		7565		9007	
CETINJE	3	0.08%			139	1.74%	35	0.44%	95	1.01%
DANILOVGRAD					2	0.02%	10	0.13%	138	1.46%
NIKSIC	7	0.19%	3	0.07%	66	0.82%	7438	93.77%	124	1.31%
POGORICA	22	0.61%	28	0.67%	186	2.32%	82	1.03%	8650	91.60%
Juzni	29		15		7445		218		270	
BAR	8	0.22%	4	0.10%	71	0.89%	49	0.62%	52	0.55%
BUDVA	11	0.30%	4	0.10%	6806	84.99%	26	0.33%	70	0.74%
HERCEG NOVI	2	0.06%	5	0.12%	47	0.59%	44	0.55%	53	0.56%
KOTOR	6	0.17%	1	0.02%	270	3.37%	57	0.72%	34	0.36%
TIVAT	2	0.06%	1	0.02%	244	3.05%	24	0.30%	24	0.25%
ULCINJ					7	0.09%	18	0.23%	37	0.39%
Sjeverni	3526		4092		72		129		147	
ANDRIJEVICA	109	3.01%			3	0.04%	3	0.04%	2	0.02%
BERANE	3247	89.55%	23	0.55%	18	0.22%	21	0.26%	24	0.25%
BIJELO POLJE	44	1.21%	3963	95.52%	17	0.21%	11	0.14%	49	0.52%
KOLASIN	6	0.17%	22	0.53%	25	0.31%	15	0.19%	22	0.23%
MOJKOVAC	3	0.08%	63	1.52%			19	0.24%	1	0.01%
PLAV	78	2.15%	9	0.22%	1	0.01%	1	0.01%		
PLJEVLJA	4	0.11%	5	0.12%	7	0.09%	22	0.28%	20	0.21%
PLUZINE							25	0.32%	5	0.05%
ROZAJE	34	0.94%	3	0.07%	1	0.01%	3	0.04%	4	0.04%
SAVNIK									1	0.01%
ZABLJAK	1	0.03%	4	0.10%			9	0.11%	19	0.20%
Van Crne Gore	39		11		98		20		19	
bileća							1	0.01%		
BOSNA I HERCEGOVINA			4	0.10%	2	0.02%			1	0.01%
brodarevo							4	0.05%		
CACAK	2	0.06%								
HRVATSKA									1	0.01%
ITALIJA									1	0.01%
LJEVIŠTA							1	0.01%		
MAKEDONIJA					1	0.01%				
NOVI SAD			1	0.02%	5	0.06%	2	0.03%	3	0.03%

Sheet 1

Picture 4.11. Model for customers analysis according to residential address and shop location

MODEL SISTEMA POSLOVNE INTELIGENCIJE ZA ANALIZU KUPACA TRGOVINSKOG PREDUZEĆA

Zaključak: *Model sistema poslovne inteligencije za analizu kupaca trgovinskog preduzeća, zasnovanog na konceptima kao što su Data warehouse i alatima kao što su Warehouse builder, OLAP i business intelligence alati, daje očekivane rezultate. realizovani model može biti osnosa za realizaciju modela poslovne inteligencije drugih trgovinskih preduzeća. Rad je dao uvid u potencijal Oracle alata za realizaciju komponenti sistema poslovne inteligencije.*