

АНАЛИЗ МНОГОМЕРНЫХ БАЗ ДАННЫХ НА ПРИМЕРЕ БАНКОВСКОЙ СИСТЕМЫ

Попов Александр Митрофанович,

к.ф.-м.н, доцент, Российский Университет Дружбы Народов (РУДН)
thenickys@gmail.com

Куценко Ирина Львовна,

к.ф.-м.н, доцент, Российский Университет Дружбы Народов (РУДН)
nicky2004@list.ru

Аннотация. В данной статье рассматривается многомерная система построения OLAP-кубов на примере банковской системы. Основными задачами работы является выявление наиболее часто потребляемых видов кредитов и проверка анализа рисков выдачи кредита клиентам. Для достижения поставленных задач был проделан анализ данных, на основании которого были получены результаты оценки деятельности баков, а так же выявлены наиболее популярные кредитные программы и рассчитан прогноз дефолта клиента.

Ключевые слова: Многомерные базы данных, OLAP — куб, разбиение с поворотом, анализ рисков, дефолт клиента, прогнозирование в банковской система.

MULTIDIMENSIONAL DATABASES ANALYSIS ON THE EXAMPLE OF BANKING SYSTEM

Popov Aleksandr Mitrofanovich,

Ph.D. of physico-mathematical sciences, Russian Peoples' Friendship University

Kutsenko Irina Livovna,

Ph.D. of physico-mathematical sciences, Russian Peoples' Friendship University

Abstract. This article describes a multidimensional system of building OLAP-cubes on the example of the banking system. The main problems of the paper are to identify the most frequently used types of credit risk analysis and verification of loan customers. To achieve these problems solution were necessary to make the analysis of data. Which based on the evaluation of the bank system, as well as identify the most popular loan programs and reveal the default clients.

Keywords: Multidimensional database, OLAP – cube, slising ang dising, risk analysis, default client, prognosis in banking system.

1. Представление многомерных данных

Таблица 1

Реляционная модель представления данных

Вид кредита	Месяц	Объем сделок
Потребительский	сентябрь	5
Авто	октябрь	3
Ипотека	ноябрь	2
Авто	сентябрь	2
Авто	октябрь	2
Потребительский	ноябрь	3
Потребительский	сентябрь	4
Ипотека	октябрь	5

Для представления более легких запросов потребуется простая таблица. Например: рассчитать объемы продаж в сентябре, октябре и ноябре месяце.

В реляционном виде таблица будет выглядеть следующим образом:

Вид кредита	Месяц	Объем сделок
Потребительский	ноябрь	4
Авто	сентябрь	3
Потребительский	октябрь	4
Ипотека	ноябрь	1
Авто	сентябрь	3
Потребительский	октябрь	4
Потребительский	ноябрь	5
Ипотека	сентябрь	6
Потребительский	октябрь	3
Ипотека	ноябрь	3
Авто	сентябрь	6
Потребительский	октябрь	3
Ипотека	ноябрь	3
Потребительский	сентябрь	3
Авто	октябрь	3
Авто	ноябрь	2
Ипотека	сентябрь	5
Авто	октябрь	2
Потребительский	ноябрь	3

Таблица 2

Многомерная модель данных

Вид кредита	Сентябрь	Октябрь	Ноябрь
Потребительский	12	14	15
Авто	14	10	12
Ипотека	19	12	9

2. Анализ и построение OLAP-кубов.

Главным достоинством OLAP является наличие у них способности предоставлять пользователям своевременную информацию, необходимую для принятия эффективных решений по выбору стратегических направлений развития организации. [1]

Одной из основных аналитических операций сервера многомерных баз данных является разбиение с поворотом. Эта операция позволяет получить представление данных с разных точек зрения. Например один срез данных о доходах может отображать все сведения о доходах от продаж объектов недвижимости указанного типа по каждому городу. Другой срез может представлять все данные о доходах отделений компании в каждом из городов.

3. Построение Базы Данных для банка

В рассмотренной базе данных представлены следующие объекты соответствующие реальному миру:

«Клиент», «Вид_кредита», «Сотрудники», «Договор», «Отделения», «Параметры_кредитования», «Предмет_договора», «База_авто», «База_квартир», «Подтверждение_доходов».

В итоге получим общую Базу Данных, удовлетворяющую ссылочной целостности. (Рис. 1).

Для анализа выявления кредита, который пользовался спросом за последний год нам понадобится многомерная модель представления данных.

В этой таблице измерениями являются «Месяцы» и «Виды кредита», а в качестве измерений будет «Объем сделок». Данные собраны поквартально.

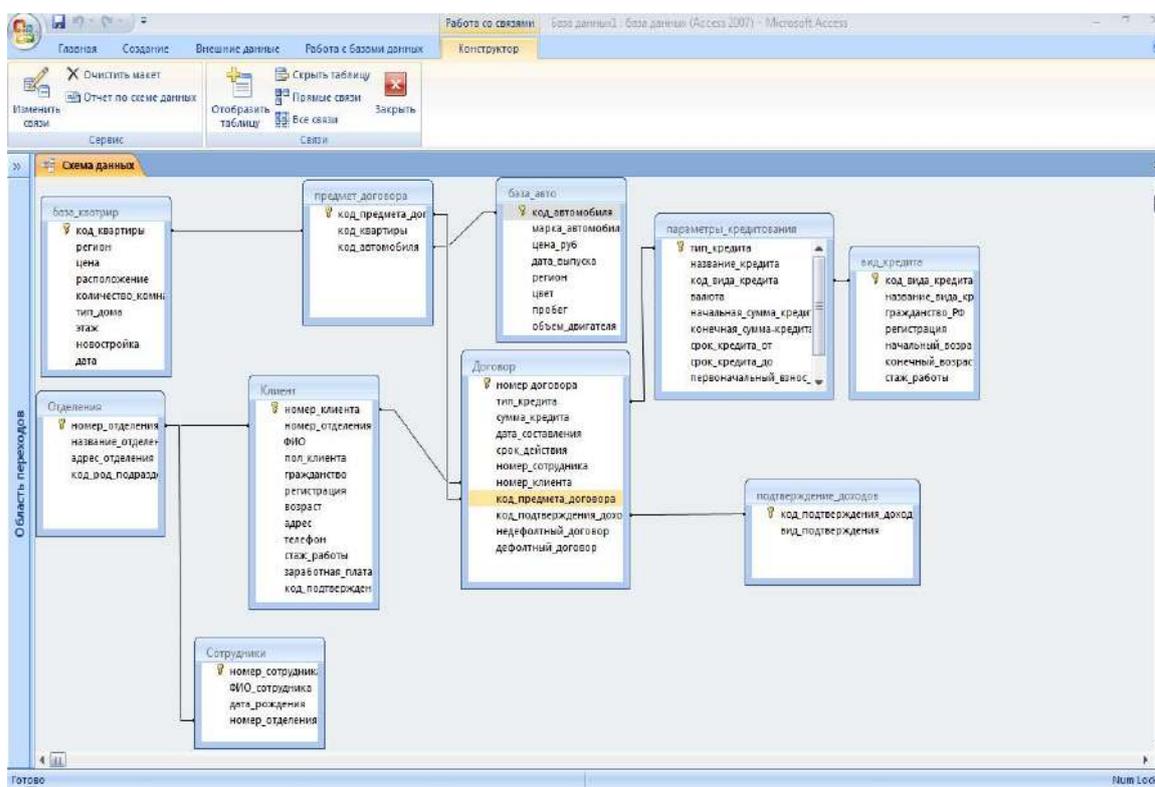


Рис. 1. ER — модель данных

Таблица 3

Многомерная таблица данных (по кварталам)

	1 квартал	2 квартал	3 квартал	4 квартал
Кредиты физическим лицам	147	120	132	102
Потребительский	53	41	42	39
Автокредит	37	38	40	31
Ипотечный	57	41	50	32
Кредиты юридическим лицам	76	76	96	88
Банковский	22	16	23	30
Тендерный	15	24	28	23
Лизинг	20	24	29	21
Факторинг	19	12	16	14

Из таблицы 3 видим, что низкая выдача кредитов по физическим лицам была в апреле и составила 7 шт., а уже в мае стала самой высокой по всему году и составила 19 шт. А по юридическим лицам максимальный объем продаж был в сентябре и составил 9 шт.

Таблица 4

Многомерная таблица данных (по кварталам)

	1 квартал	2 квартал	3 квартал	4 квартал
Кредиты физическим лицам	147	120	132	102
Потребительский	54	41	42	39
Автокредит	38	39	40	31
Ипотечный	57	40	50	32
СРЗНАЧ	50	40	44	34
Кредиты юридическим лицам	76	76	96	88
Банковский	22	15	24	30
Тендерный	14	24	28	23
Лизинг	22	25	29	21
Факторинг	18	11	16	12
СРЗНАЧ	19	19	24	22

Из таблицы 4 можно сделать вывод, что лучший объем продаж по физическим лицам был в 1 квартале, т.е. с января по март и составил 50 шт, а по юридическим лицам в 3 квартале и составил 24 шт.

Глядя на таблицы, приведенные выше, можно сделать следующий вывод:

Среди выдачи кредитов по физическим лицам пользуется спросом Ипотечный.

4. Прогноз дефолта клиента

На сегодняшний день выдача кредитов очень востребована. Все больше и больше людей ими пользуются. Примерно% населения. Но порой люди не могут их до конца погасить, в этом случае необходимо рассчитывать возможность погашения кредита. С этим нужно быть предельно осторожным и использовать все имеющиеся возможности для анализа платежеспособности заемщиков.[2], [3]

В случае полного и своевременного погашения кредита вероятность возврата кредита $P=1$, а вероятность дефолта клиента $Q=0$

$$Q=n/N \quad (4.1)$$

где N-общее число выданных кредитов
n-количество невозвратов кредитов

Таблица 5

Таблица расчета кредитов пользующихся за весь год

	1 квартал	2 квартал	3 квартал	4 квартал	
Кредиты физическим лицам	147	120	132	102	Какой кредит пользовался спросом за весь
Потребительский	53	41	42	39	44
Автокредит	37	38	40	31	37
Ипотечный	57	41	50	32	45
Кредиты юридическим лицам	76	76	96	88	Какой кредит пользовался спросом за весь
Банковский	22	16	23	30	23
Тендерный	15	24	28	23	23
Лизинг	20	24	29	21	24
Факторинг	19	12	16	14	15

	A	B	C	D	E	F	G	H	I	J	K
1											интервальная оценка вероятности дефолта
2		всего кредитов		n	N	Q1	Q2	P	DQ	SQ	
3	Сидоров И.С.	1		0	1	0	0,5	0,5	0,083333	0,2886751	(0.04; 0.54)
4	Иванов В.П.	2		0	2	0	0,333333	0,666667	0,055556	0,2357023	(0.03; 0.45)
5	Петров Н.К.	3		0	3		0,25	0,75	0,0375	0,1936492	(0.02; 0.36)
6	Тимофеев И.В	3		0	3	0	0,25	0,75	0,0375	0,1936492	(0.02; 0.36)
7	Тапкин С.С.	4		0	4	0	0,2	0,8	0,026667	0,1632993	(0.02; 0.3)
8	Володин П.В	5		0	5	0	0,166667	0,833333	0,019841	0,140859	(0.02; 0.26)
9	Цветков А.Г.	1		0	1		0,5	0,5	0,083333	0,2886751	(0.04; 0.54)
10	Бондалетов Г.П.	2		0	2	0	0,333333	0,666667	0,055556	0,2357023	(0.03; 0.45)
11	Степанов Н.М.	1		0	1	0	0,5	0,5	0,083333	0,2886751	(0.04; 0.54)
12											
13											
14											

Рис. 2. Данные о недефолтных клиентах.

Это выражение вероятности предназначено только для $n > 0$

Из теории вероятности и математической статистики вероятность дефолта будем оценивать по следующей формуле:

$$Q = (n+1)/(N+1) \quad (4.2)$$

А вероятность возврата кредита по выражению:

$$P = 1 - Q = 1 - (n+1)/(N+1) \quad (4.3)$$

Дисперсия вероятности дефолта клиента оценивается по формуле:

$$D(Q) = PQ/(N+2) \quad (4.4)$$

А среднее квадратичное отклонение указанной вероятности по формуле:

$$S(Q) = PQ/(N+2) \quad (4.5)$$

Обращаясь к кубу и используя вышеперечисленные формулы, вычислим клиентов, которые вызывают наибольшее доверие со стороны банка. (Рис. 2).

По приведенным аналитическим вычислениям можно заметить, что клиент Володин вызывает наибольшее доверие со стороны банка, вероятность возврата кредита которого минимальна.

Список литературы

1. И.Ф. Готовчиков. Практика использования математических методов при управлении кредитными рисками в розничном кредитовании // Банковское кредитование. 2006 №5.
2. Т. Коннолли, К. Бегг. Базы данных. Проектирование, реализация и сопровождение. Теория и практика. 3-е издание.: Пер. с англ. – ООО «И.Д. Вильямс», 2003. – 1228 с. Под ред. К.А. Птицина.
3. Билл Джелен, Майкл Александер. Сводные таблицы в Microsoft Excel 2007.; Пер. с англ. -М.; ООО «И.Д. Вильямс», 2008-336с