

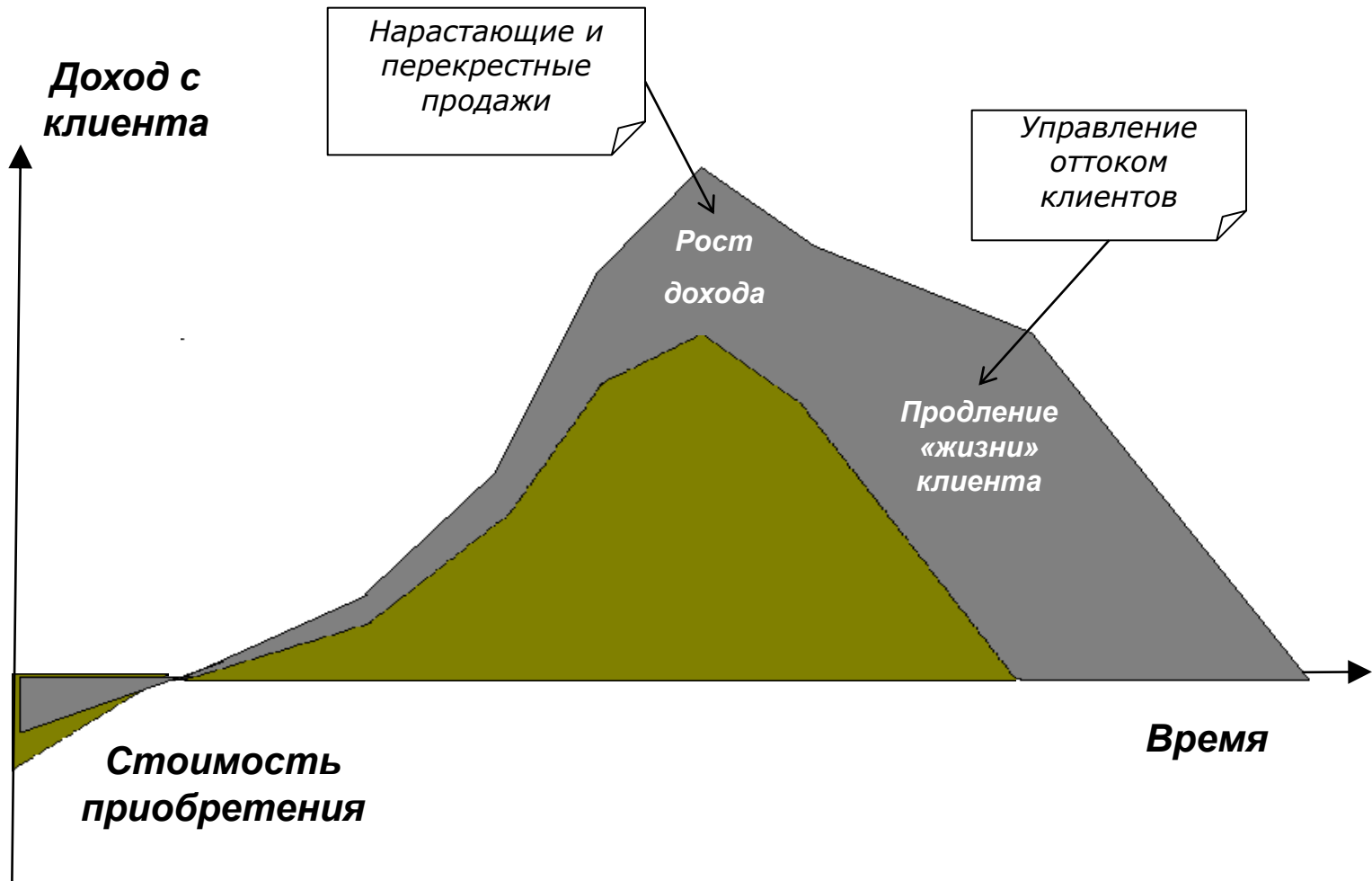
***Повышение рентабельности инвестиций  
маркетинга в телекоммуникациях с помощью  
математического аппарата.***

***Владимир Маслик***

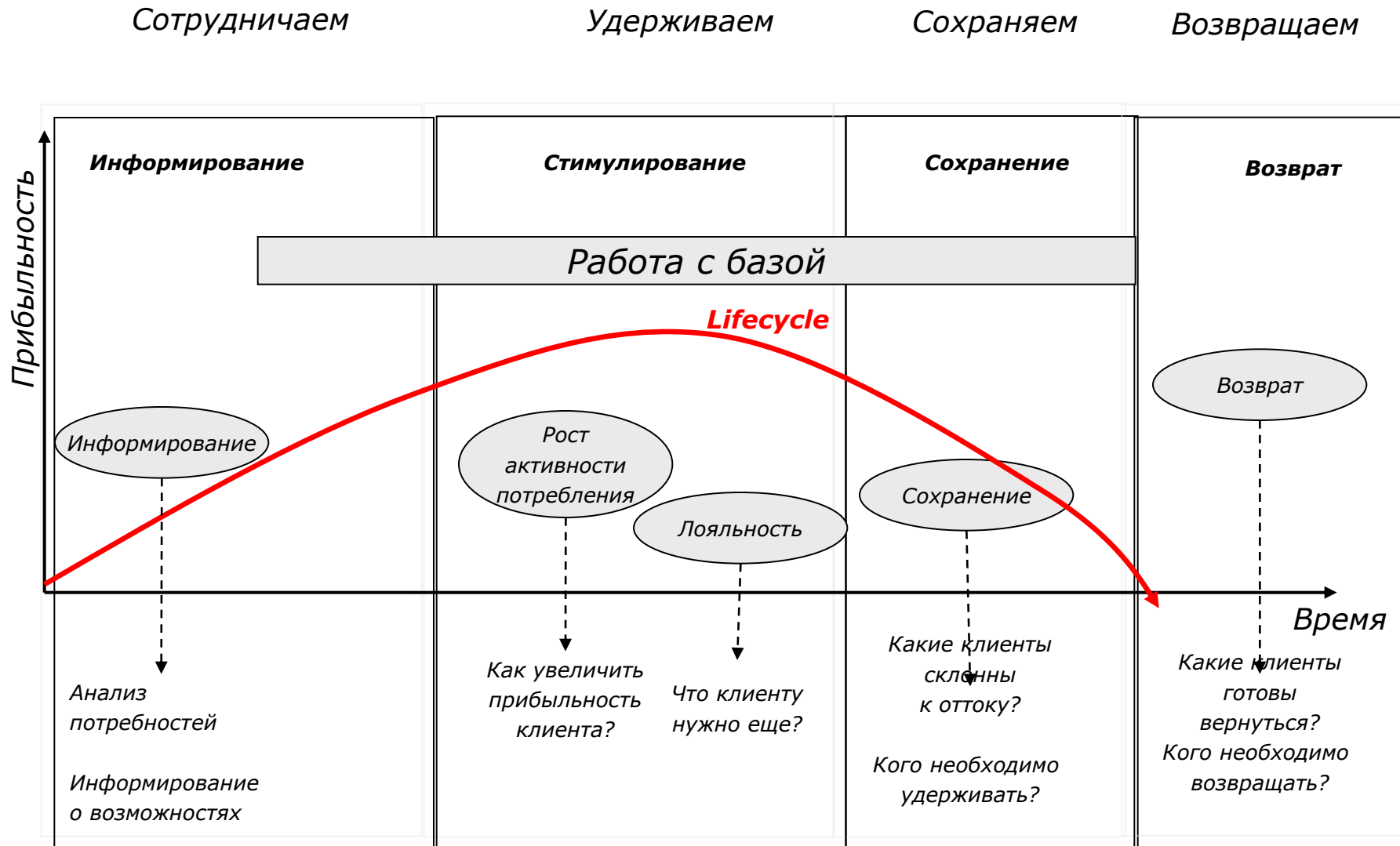
# Содержание

- *Общая стратегия работы с абонентской базой*
- *Стратегия информирования*
- *Стратегия стимулирования*
- *Стратегия сохранения*
- *Организация доступа к информации*
- *Применяемые математические методы*
- *Практическая часть*

# Жизненный цикл клиента



# Активности на всем жизненном цикле клиента

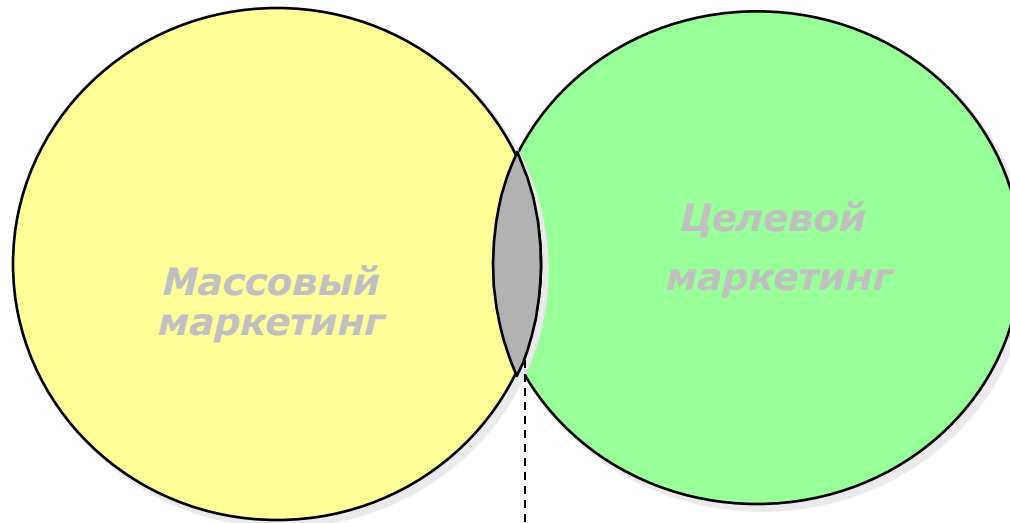


[2] – Arthur Middleton Hughes Strategic Database Marketing; McGraw- Hill, 2005 г.

# Стимулирование абонента



# МЕСТО ЦЕЛЕВОГО МАРКЕТИНГА



- Действенный способ привлечения новых клиентов
- Отлично подходит для продажи простых продуктов
- Единое открытое предложение для всех абонентов
- Активное внешнее продвижение
- Долгосрочное предложение

**Максимальный охват**

- Работает на конкретную аудиторию. Коммуникация Person2Person
- Хороший способ укрепить лояльность
- Если нужны разъяснения при продаже сложных VAS
- Создает дополнительную ценность для клиента в отношениях с оператором
- Позволяет управлять стоимостью предложения для разных клиентов
  - Дизайн и сроки предложений диктуются целями. Может быть неограниченное количество целевых предложений

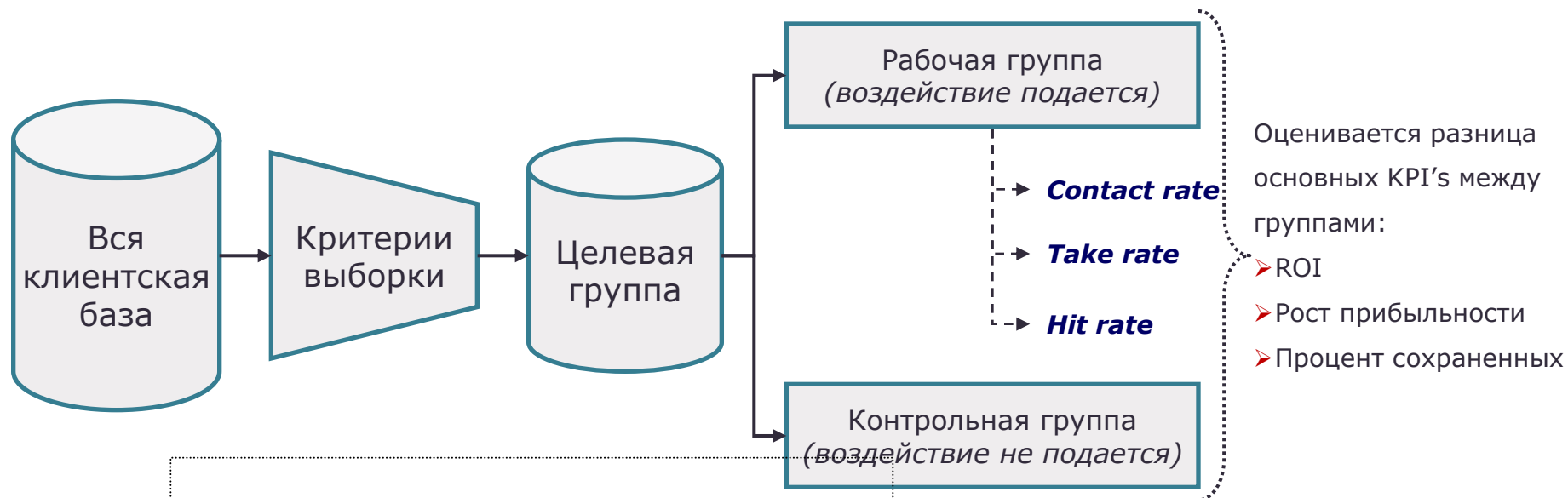
**Гибкость на операционном уровне**

# Бизнес-процесс целевой маркетинговой акции

*Принцип построения бизнес-процесса целевой маркетинговой акции*



# Анализ целевых компаний.[3]



Оценивается разница основных KPI's между группами:

- ROI
- Рост прибыльности
- Процент сохраненных

ROI = 
$$\frac{\text{Lift ARPU} - \text{Затраты на коммуникацию}}{\text{Retention effect} - \text{Стоимость бонуса/скидки}}$$

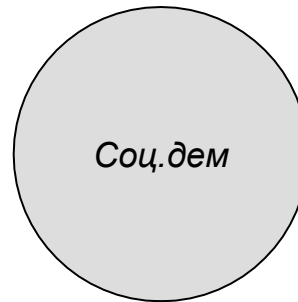
- Contact rate** – сколько клиентов из совокупной выборки удалось проинформировать об условиях акции (данный показатель влияет на затраты компании на коммуникацию предложения).
- Take rate** – сколько клиентов согласились принять участие в акции (данный показатель демонстрирует заинтересованность клиентов предложенными условиями).
- Hit rate (real take rate)** – сколько клиентов выполнили условия акции (на основе данного показателя рассчитываются затраты на вознаграждение клиентов или недополученная выручка).

[3] – Sara Madeira, João M. Sousa, "Comparison of Target Selection Methods in Direct Marketing". In Proc. of European Symposium on Intelligent Technologies, Hybrid Systems and their Implementation on Smart Adaptive Systems, Eunate'02, Albufeira, Portugal, September 2002.

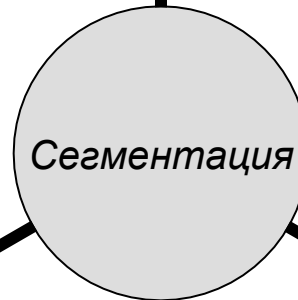


# Сегментация. Основные положения

## **Соц-дем сегментация на основе потребностей клиентов**



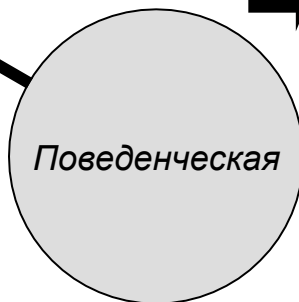
- Студенты моложе 25 лет
- Молодые работающие мужчины
- Молодые работающие женщины
- Работающие мужчины с маленькими детьми
- Женщины с маленькими детьми



## Кластеризация абонентов

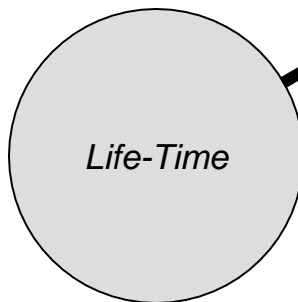


- ARPU
- Состав услуг
- Зона сети
- Платежи
- Блокировки
- Обращения абонентов

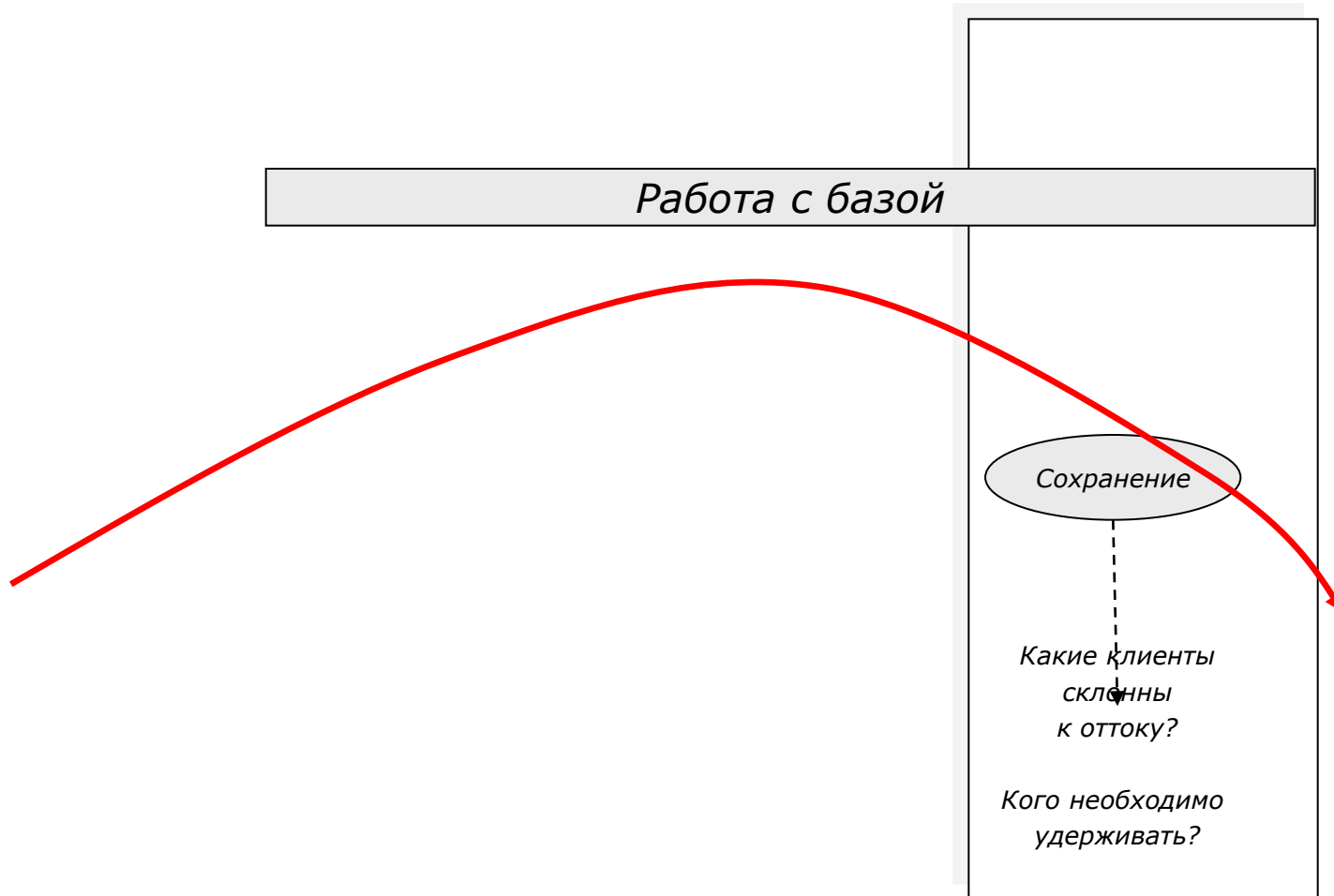


## **Сегментация по жизненному циклу**

- 0-6 месяцев
- 6-12 месяцев
- 1-2 года
- 2-3 года
- Более 3-х лет



# Сохранение и повышение ЛОЯЛЬНОСТИ



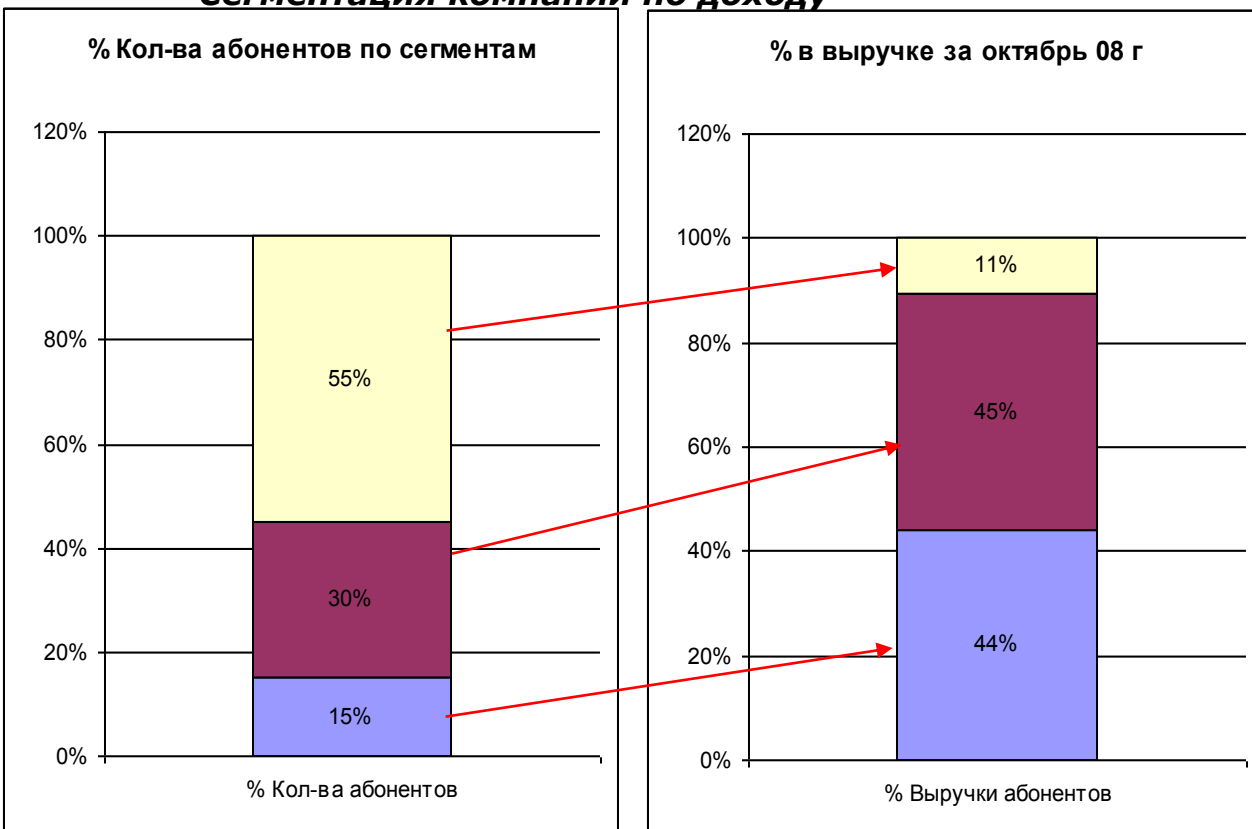
# Необходимо внедрить и применять сегментированный подход!

*Абоненты неравноценны с т.зр. поступающей от них выручки и распределены в базе неравномерно. Необходимо удерживать высокодоходных абонентов!*

*Разбиваем базу на сегменты по ср.месячному доходу с абонента*

*Выделенные сегменты по доходности:*

## Сегментация компании по доходу



**Top 15% (Consumer High) –**  
15% абонентской базы с максимальным ARPU  
Приносят 44% всей выручки

**Next 30% (Consumer Middle) –**  
30% абонентской базы со средним ARPU  
Приносят 45% выручки

**Bottom 55% (Consumer Low) –**  
остальная абонентская база  
Приносят 11% выручки

Применяем про-активного взаимодействия с  
потенциальными churn-ми.

Выявляем кого удерживаем(сегмент)!

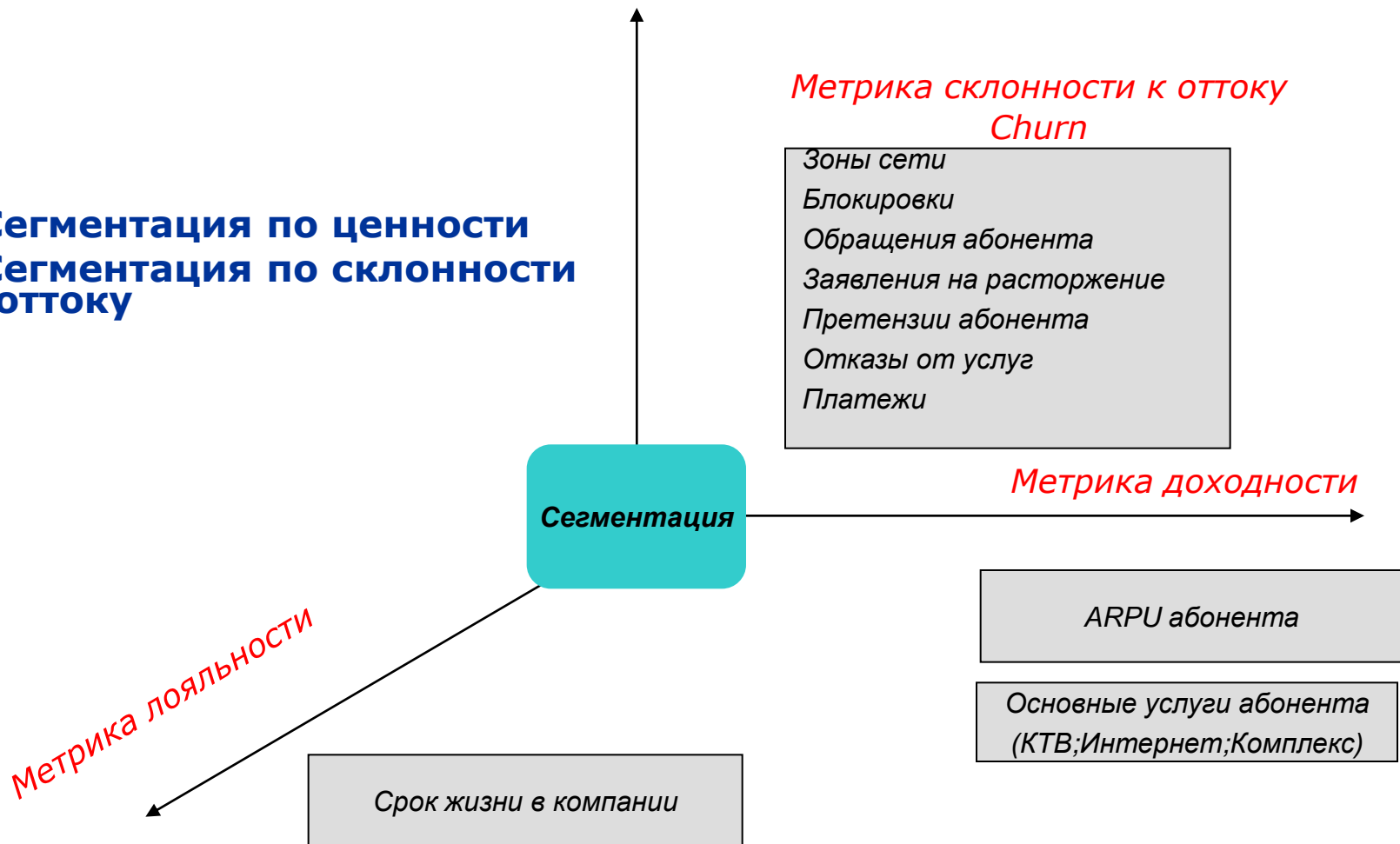
Выявляем каким инструментом удерживаем!

Определяем момент удержания!

Определяем канал коммуникации!

# Принципы сегментации абонентской базы

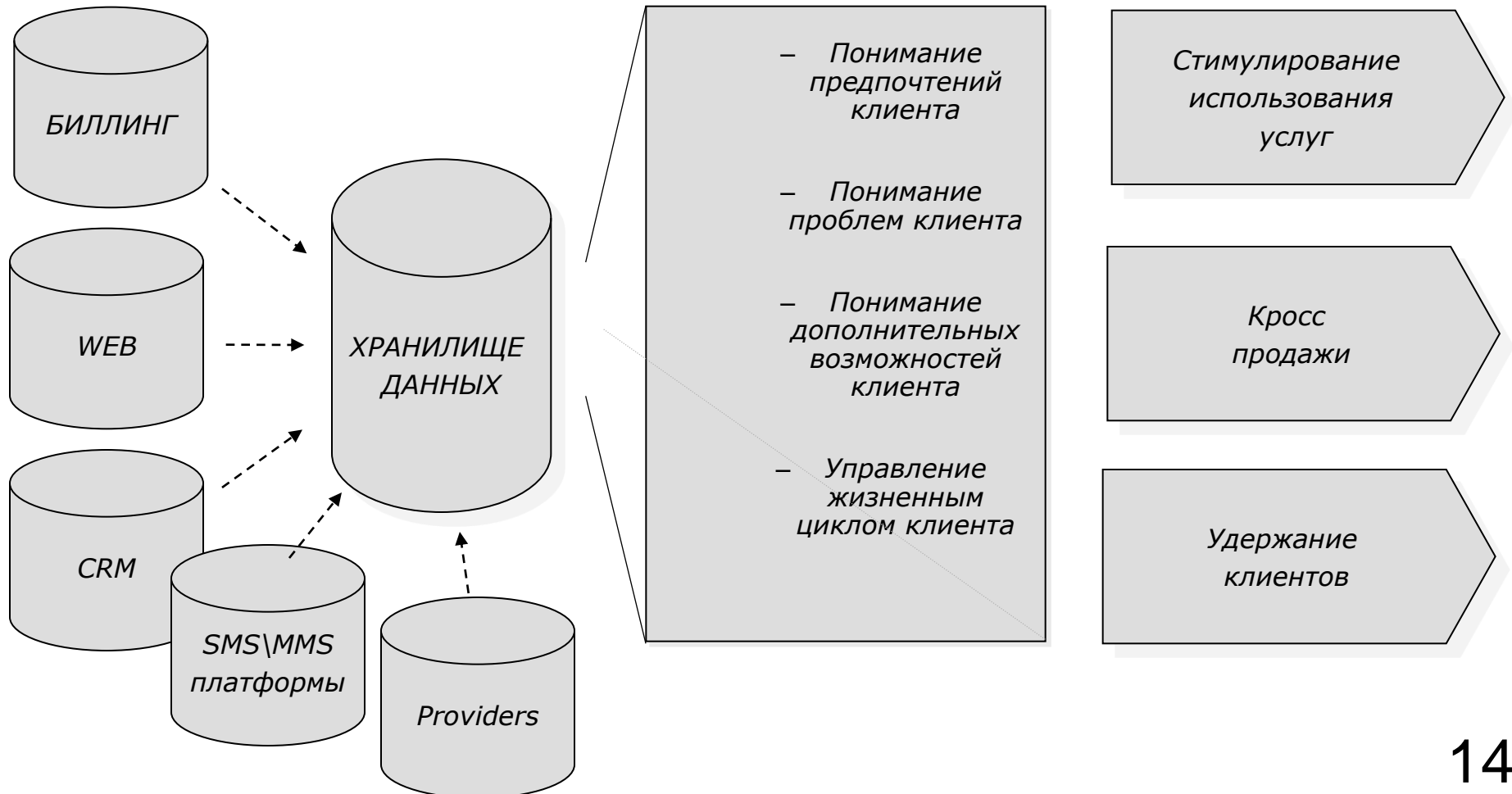
- Сегментация по ценности
- Сегментация по склонности к оттоку



# Организация доступа к информации

# Структура данных.

*Интеграция данных из всех источников в единое хранилище позволяет компании проводить детальную клиентскую аналитику, строить различные поведенческие модели, что предоставляет много возможностей для создания целевых предложений*



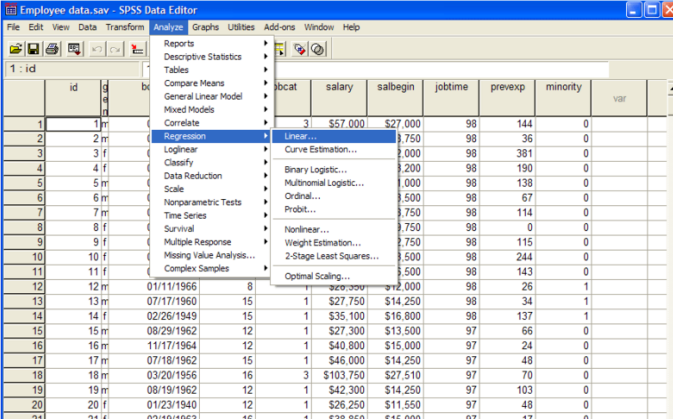
# Используемое программное обеспечение.

## Хранилище данных.

*Microsoft SQL Server — реляционная СУБД, разработанная корпорацией Microsoft. Основным используемый язык запросов — Transact-SQL, создан совместно Microsoft и Sybase. Используется для небольших и средних по размеру баз данных, и в последние 5 лет — для крупных баз данных масштаба предприятия, конкурирует с другими СУБД в этом сегменте рынка.[5]*

## Инструменты анализа данных.

*SPSS (аббревиатура англ. «Statistical Package for the Social Sciences», «статистический пакет для социальных наук») — компьютерная программа для статистической обработки данных, один из лидеров рынка в области коммерческих статистических продуктов, предназначенных для проведения прикладных исследований в социальных науках.[6]*



	id	sex	age	jobcat	salary	salbegin	jobtime	prevexp	minority	var
1	1	m								
2	2	m								
3	3	f			\$57,000	\$27,000	98	144	0	
4	4	f								
5	5	m								
6	6	m								
7	7	m								
8	8	f								
9	9	f								
10	10	f								
11	11	f								
12	12	m	01/11/1966	8						
13	13	m	07/17/1960	15	1	\$27,750	\$14,250	98	34	1
14	14	f	02/26/1949	15	1	\$35,100	\$16,800	98	137	1
15	15	m	08/29/1962	12	1	\$27,300	\$13,500	97	66	0
16	16	m	11/17/1964	12	1	\$40,800	\$15,000	97	24	0
17	17	m	07/18/1962	15	1	\$46,000	\$14,250	97	48	0
18	18	m	03/20/1956	16	3	\$103,750	\$27,510	97	70	0
19	19	m	08/19/1962	12	1	\$42,300	\$14,250	97	103	0
20	20	f	01/23/1940	12	1	\$26,250	\$11,550	97	48	0
21	21	f	02/19/1963	16	1	\$38,850	\$15,000	97	17	0

[5] – Сайт Microsoft.com

[6] – spss.ru



Математические методы.

# Этапы анализа данных.

- 1. Определение целей анализа.**
- 2. Отбор данных.**
- 3. Проверка данных.** – нормальность распределения; эксцесс; асимметрию
- 4. Предварительная обработка входных данных .**– нормализация; замена пропущенных значений
- 5. Выбор метода анализа.**
- 6. Анализ данных .**– построение моделей; выявление закономерностей и т.д.
- 7. Проверка результатов анализа и корректировка входных переменных; методов анализа.**
- 8. Экстраполяция результатов анализа.**
- 9. Применение результатов в бизнесе .**

# Математические алгоритмы анализа и прогнозирования.

**Бинарная логистическая регрессия** – это разновидность множественной регрессии, общее назначение которой состоит в анализе связи между несколькими независимыми переменными (называемыми также регрессорами или предикторами) и зависимой переменной.

Бинарная логистическая регрессия, как следует из названия, применяется в случае, когда зависимая переменная является бинарной (т.е. может принимать только два значения).

Иными словами, с помощью логистической регрессии можно оценивать вероятность того, что событие наступит для конкретного испытуемого (больной/здоровый, возврат кредита/дефолт и т.д.).

**Кластерный анализ**— задача разбиения заданной выборки объектов (ситуаций) на непересекающиеся подмножества, называемые кластерами, так, чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров существенно отличались.

## **Иерархический кластерный анализ**

В иерархических методах каждое наблюдение образует сначала свой отдельный кластер. На первом шаге два соседних кластера объединяются в один; этот процесс может продолжаться до тех пор, пока не останутся только два кластера. В методе, который в SPSS установлен по умолчанию (*Between-groups linkage* (Связь между группами)), расстояние между кластерами является средним значением всех расстояний между всеми возможными парами точек из обоих кластеров.

## **Деревья принятия решений .**

Обычно используются для решения задач классификации данных или, иначе говоря, для задачи аппроксимации заданной булевой функции. Ситуация, в которой стоит применять деревья принятия решений, обычно выглядит так: есть много случаев, каждый из которых описывается некоторым конечным набором дискретных атрибутов, и в каждом из случаев дано значение некоторой (неизвестной) булевой функции, зависящей от этих атрибутов

**Искусственные нейронные сети (ИНС)**— математические модели, построенные по принципу организации и функционирования биологических нейронных сетей — сетей нервных клеток живого организма. Это понятие возникло при изучении процессов, протекающих в мозге при мышлении, и при попытке смоделировать эти процессы. Первой такой моделью мозга был перцептрон. Впоследствии эти модели стали использовать в практических целях, как правило в задачах прогнозирования.

# Математические алгоритмы анализа и прогнозирования.

## Факторный анализ

Факторный анализ это процедура, с помощью которой большое число переменных, относящихся к имеющимся наблюдениям сводит к меньшему количеству независимых влияющих величин, называемых факторами. При этом в один фактор объединяются переменные, сильно коррелирующие между собой. Переменные из разных факторов слабо коррелируют между собой. Таким образом, целью факторного анализа является нахождение таких комплексных факторов, которые как можно более полно объясняют наблюдаемые связи между переменными, имеющимися в наличии.

## Дискриминантный анализ

Такая постановка задачи, в особенности в случае двух заранее заданных групп, очень сильно напоминает постановку задачи для метода логистической регрессии (см. гл. 16.4). Ядром дискриминантного анализа является построение так называемой дискриминантной функции

$$d = b_1x_1 + b_2x_2 + \dots + b_nx_n + a,$$

где  $x_1$  и  $x_n$  — значения переменных, соответствующих рассматриваемым случаям, константы  $b_1$ - $b_n$  и  $a$  — коэффициенты, которые и предстоит оценить с помощью дискриминантного анализа. Целью является определение таких коэффициентов, чтобы по значениям дискриминантной функции можно было с максимальной четкостью провести разделение по группам.

# Практическая часть

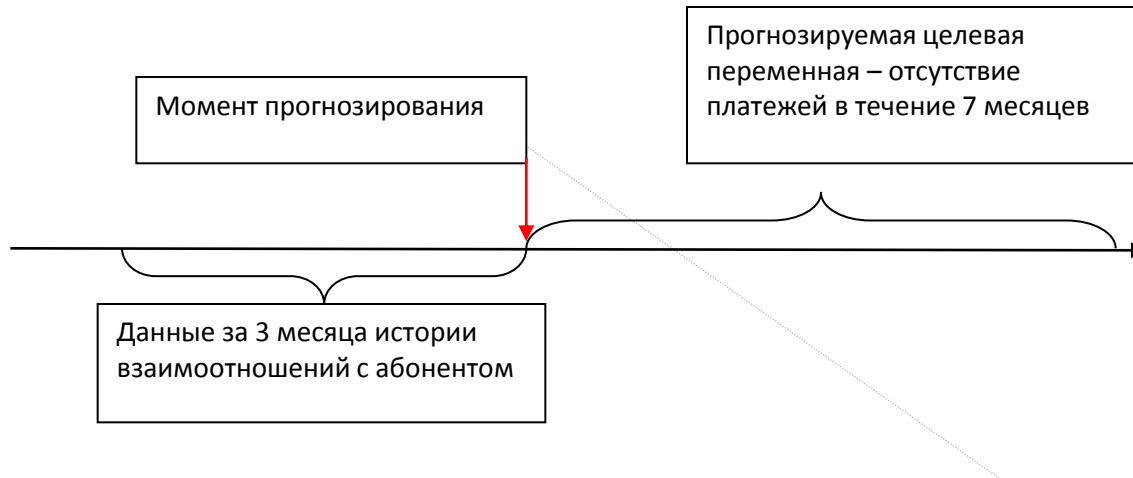
# Пример определения склонности абонента к оттоку.

## **Цель работы:**

*По имеющимся данным необходимо составить предикативную модель оттока абонентов.*

*Оттоком абонентов считается отсутствие платежей в течение 7 месяцев по истории абонента за 3 месяца.*

*3 месяца было выбрано для оптимальной скорости реагирования на изменения в поведении абонента.*



# Анализ данных.

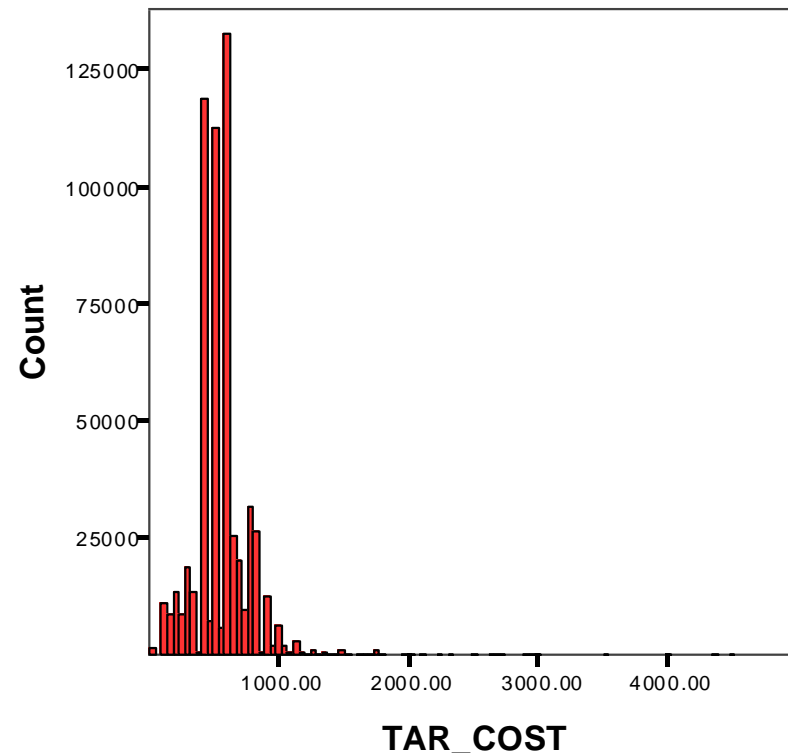
## АНАЛИЗ ДАННЫХ:

### Подготовительный этап – обработка и оценка данных

До анализа данных необходимо произвести предварительную обработку и чистку данных (В случае ошибок ввода)

Для начала необходимо заменить пропущенные значения (Для платежей и начислений), проверить на дублируемость строк. Для заполнения пробелов выполняю процедуру перекодировки переменных в самих себя, при этом выбираю функцию перекодировки пропущенных переменных в 0

Вторым этапом необходимо оценить (визуально разброс данных) – оценим разброс данных стоимости тарифа на интернет с помощью гистограммы



# Отбор репрезентативной выборки.

Для того, чтобы к примеру с вероятностью 0,95 доля абонентов с значением  $AIM=1$  (неплательщики) в выборке отличалась от ген.совокупности не более, чем на 0,004 (по абсолютной величине), необходимо произвести следующие расчеты:

1) Найдем состоятельную оценку доли неплательщиков в ген. совокупности.

По данным частотного анализа получаю:

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	637333	80.3	80.3	80.3
	1.00	156435	19.7	19.7	100.0
Total		793768	100.0	100.0	

Следовательно, доля = 0,197

Учитывая, что требуемая вероятность = 0.95, то по таблицам значений функции Лапласа найду, что значению параметра  $t$ , соответствует значение 1.96 и следовательно объем выборки определяется по формуле для бесповторной выборки

$$= N * t^2 * \text{доля} * (1 - \text{доля}) / (t * \text{Доля} * (1 - \text{Доля}) + \text{Размер ген.совокупности} * \text{процент максимально допустимого отклонения}^2) \\ = 36\ 000$$



# Выбор метода прогнозирования.

*Дерево принятия решений — это дерево, на ребрах которого записаны атрибуты, от которых зависит целевая функция, в листьях записаны значения целевой функции, а в остальных узлах — атрибуты, по которым различаются случаи. Чтобы классифицировать новый случай, надо спуститься по дереву до листа и выдать соответствующее значение. Для начала необходимо в данном методе выбрать переменную, которую необходимо предугадать с большей точностью - в нашем случае — это неплательщик (при построении дерева мы стремимся угадать всех неплательщиков, даже с потерей точности классификации плательщиков)*

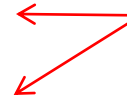
*На следующем шаге необходимо выбрать, то что мы будем выводить - это основные статистики (которые стоят по умолчанию) и саму модель в формате скриптов SPSS. Так же в настройках дерева решений выбираем валидацию данных - то есть отберем 50 % выборки, для тестирования на ней полученных результатов. Для дерева выберем максимальную глубину — 3 уровня. Так же отберем минимальное число случаев, которые попадут в родительские ветки — 2000 случаев и конечные ветки — 1000.*

*Так же выберу метод включения переменных в анализ - в данном случае наиболее эффективен метод включения переменных в порядке максимальной корреляции. Так же необходимо выбрать метод отличия кластеров - выберу кси- квадрат.*

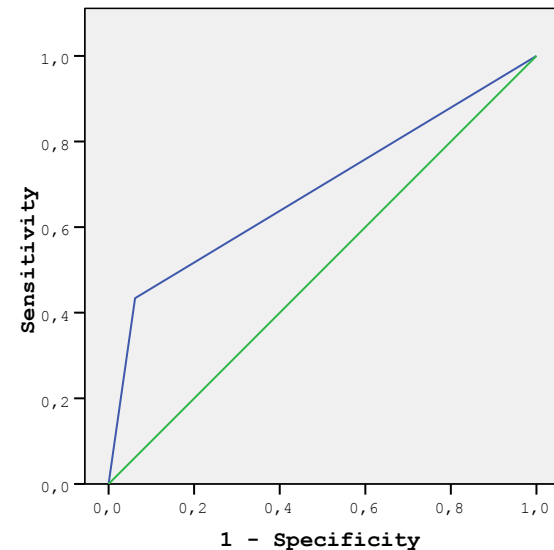
# Итоговая модель.

Sample	Observed	Predicted		
		Плательщик	Неплательщик	Percent Correct
Training	Плательщик	29869	1955	93.9%
	Неплательщик	4318	3448	44.4%
	Overall Percentage	86.4%	13.6%	84.2%
Test	Плательщик	29642	1977	93.7%
	Неплательщик	4628	3407	42.4%
	Overall Percentage	86.4%	13.6%	83.3%

Как видно по результатам в данном случае модель правильно классифицировала 44.4% неплательщиков по тренировочной выборке и 42.4 по тестовой



ROC Curve

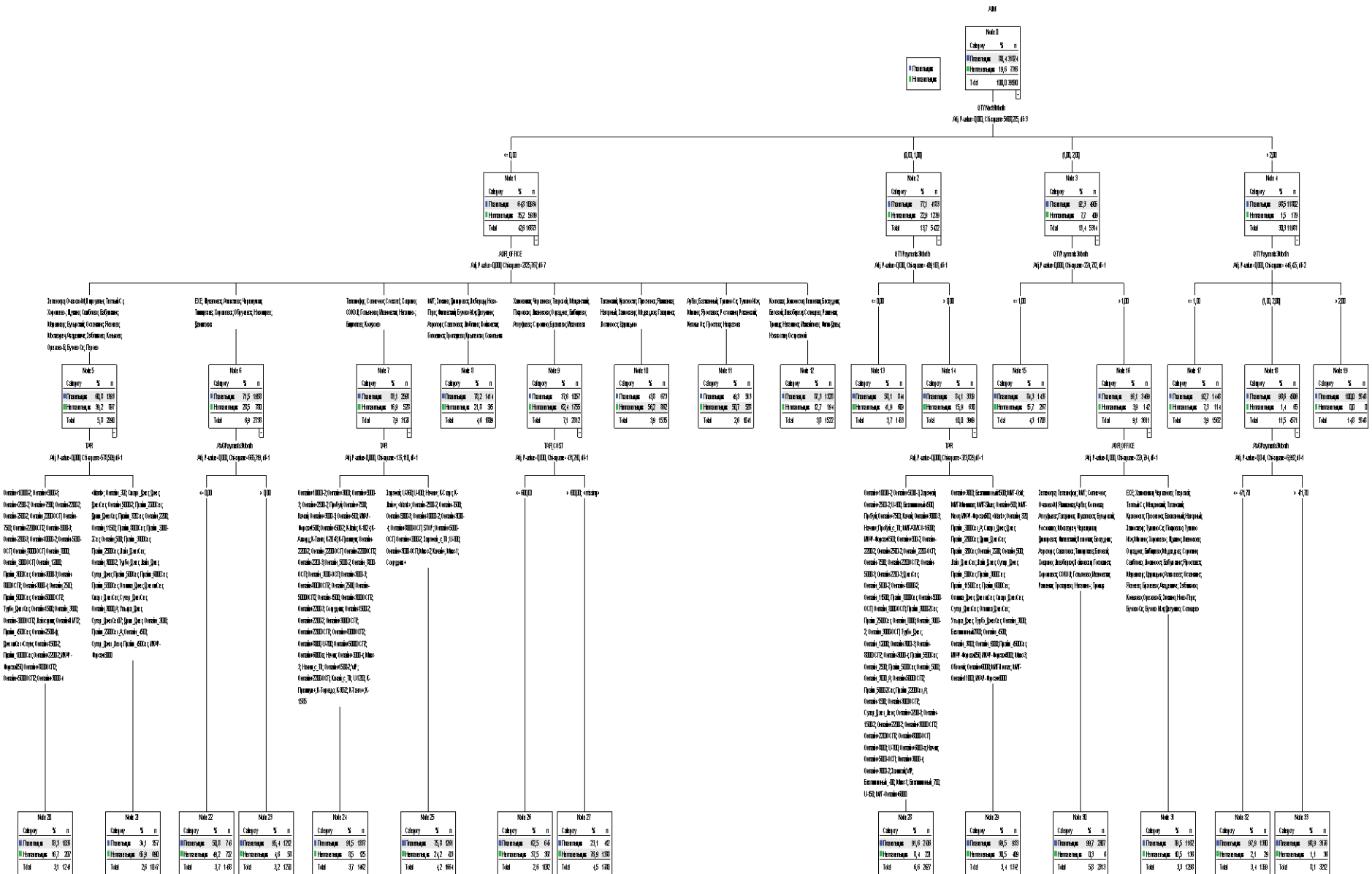


Diagonal segments are produced by ties.

Площадь под кривой = 0.67

По графику видно, что точность прогнозирования не максимальна, так как кривая проходит далеко от верхнего левого угла и площадь под кривой = 0.67 (средняя точность прогнозирования).

# Итоговая модель.



# Итоговая модель.

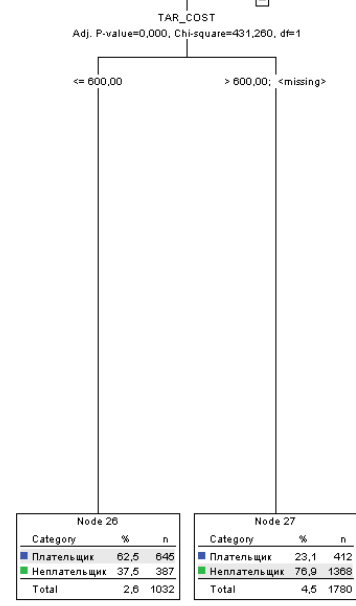
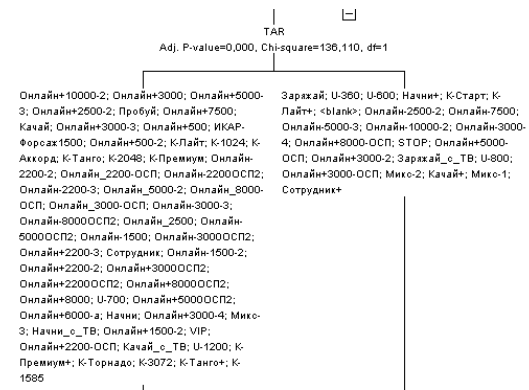
Таганский; Красносел; Преоненск; Якиманка; Нагорный; Замосквор; Медведков; Гагаринск; Лосиноост; Царицыно

Котловка; Ломоносов; Вешняки; Бескудник; Беговой; Левобереж; Солнцево; Раменки; Троицк; Нагатинск; Измайлово; Фили-Давы; Новокосин; Островной

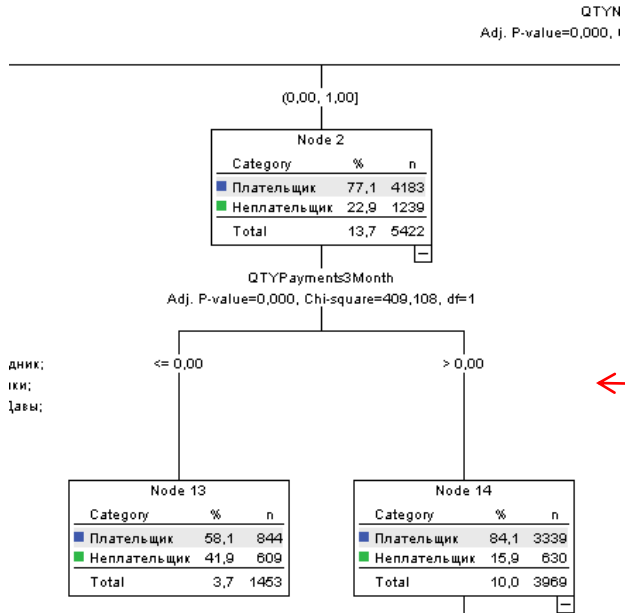
← Модель выявила проблемные районы!

Node 10		
Category	%	n
Плательщик	43,8	673
Неплательщик	56,2	862
Total	3,9	1535

Node 12		
Category	%	n
Плательщик	87,3	1328
Неплательщик	12,7	194
Total	3,8	1522



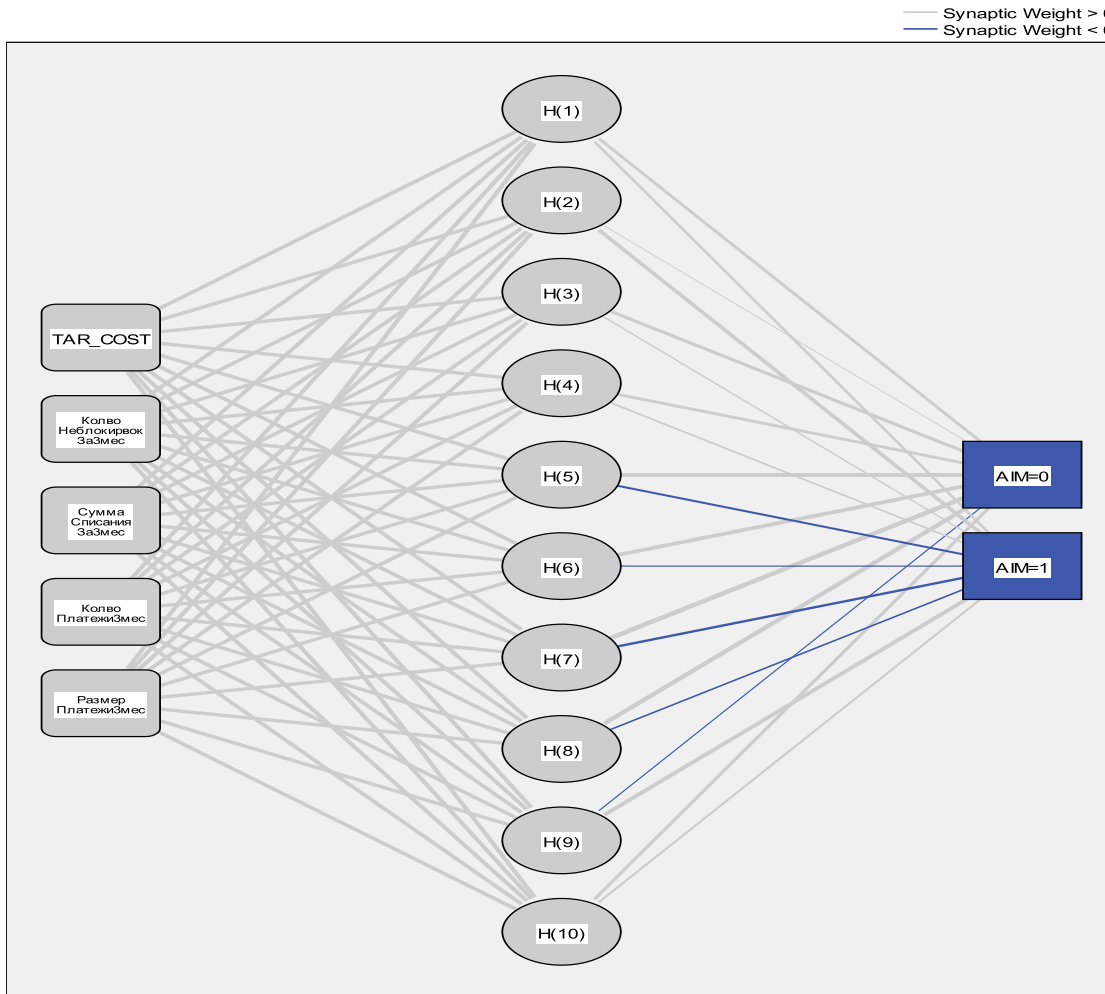
Модель выявила тарифы в зоне "риска" →



← Поведение абонента влияет на его склонность к оттоку!

# Выбор метода прогнозирования.

*ИНС(Нейронные сети)- Математические модели, построенные по принципу организации и функционирования биологических нейронных сетей — сетей нервных клеток живого организма*



*56 % всех абонентов, которые ушли в отток были классифицированы верно.*

Hidden layer activation function: Softmax

Output layer activation function: Identity

# Выбор метода прогнозирования.

Бинарная логистическая регрессия вычисляет следующее уравнение,

$$P=1/(1+e^{(-Z)})$$

где  $z = b_1 * X_1 + b_2 * X_2 + \dots + b_n * X_n + a$ ,

$X_1$  — значения независимых переменных,  $b_1$  — коэффициенты, расчёт которых является задачей бинарной логистической регрессии,  $a$  — некоторая константа.

Classification Table<sup>a</sup>

Observed		Predicted		Percentage Correct
		AIM		
		not churn	churn	
Step 1	AIM not churn	167526	33815	83,2
	churn	42140	53180	55,8
Overall Percentage				74,4

a. The cut value is .500

# Выбор метода прогнозирования.

Дискриминантный анализ.

С помощью дискриминантного анализа на основании некоторых признаков (независимых переменных) индивидум может быть причислен к одной из двух (или к одной из нескольких) заданных заранее групп.

		Predicted Group for Analysis 1		
		not churn	churn	
AIM	not churn	246594	158329	39%
	churn	39205	150618	79%

# Заключение.

## *Ключевые результаты достижения целей:*

- 1. В работе проанализирован комплекс маркетинга в телекоммуникационной отрасли.*
- 2. Определены места возможного увеличения рентабельности инвестиций*
- 3. Изучены мат. Методы прогнозирования и анализа данных*
- 4. Разработан и внедрен системный подход к анализу данных.*
- 5. Разработана и внедрена информационная составляющую Data mining.*
- 6. Разработана и внедрена система прогнозирования оттока абонентов.*
- 7. Внедрена проактивная работа по сохранению абон. базы*

## *Социальный и экономический эффект:*

- 1. Улучшено понимание структуры БД*
- 2. Снижен отток (на 1 %, что позволило сэкономить до 7 000 000\$ выручки ежегодно.*
- 3. Увеличена лояльность абонентов к компании*
- 4. Повышен уровень NPS*



**Спасибо за внимание!**