

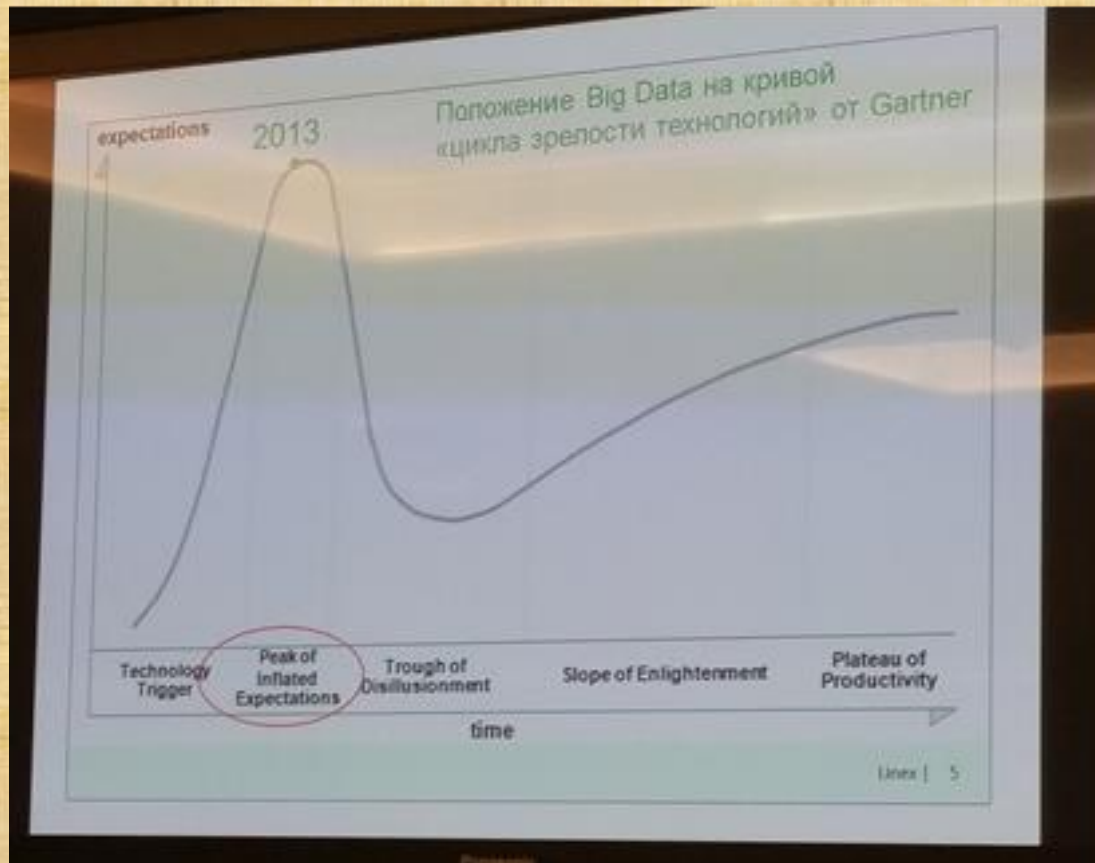
# **Большие данные в банковской деятельности**

**Ф.Т. Алескеров  
(НИУ ВШЭ и ИПУ РАН)**

- Круглый стол**
- «Форсайт финансовых профессий»**

**14 марта 2017 г.**

# Спрос на Big Data



# Мифы о данных (Gartner, 2015)

- На Западе только 13% компаний используют Большие данные. Миф — все уже используют
- Большие данные ценны только, если аналитики найдут нужную информацию. Миф — Большие данные всегда ценны
- Данные всегда предвзятые. Механика сбора ведет к предвзятости. Миф — считается, что сам объем данных исключает предвзятость. Нужен план, как избежать предвзятости.
- Все данные — структурированные на самом деле (структура может быть неочевидной). Миф — 80% данных неструктурированы
- Даже самые большие данные не могут гарантировать допустимую погрешность предсказаний. Миф — чем больше данных, тем точнее
- Потребность в специалистах удовлетворена в 2015 г. на треть, в 2016, 2017 будет хуже. Миф — развитие инструментов снижает потребность в специалистах.

# Наши разработки

- Эффективность отделений банка
- Размещение новых отделений
- Значимость коммерческих клиентов
- Распределение персонала по отделениям банка
- Сетевые модели анализа финансовой деятельности

## Цели моделирования

Как компания или банк может повысить лояльность клиентов?

Как сделать маркетинговые действия более целенаправленными?

В каких новых предложениях покупатели будут заинтересованы?

Что сможет привлечь новых клиентов от конкурентов?

Как получить большую отдачу от существующих клиентов?

...

### Следствия:

- интерес к изучению фактического поведения и предпочтений клиентов
- новые технологии управления, основанные на прогнозировании поведения клиентов



# Математические модели для маркетинговых задач

- **Методология выявления потенциала роста продаж**
  - Сегментация клиентов по типу потребления и составу корзины
  - Оценка доли компании в структуре расходов клиента (share-of-wallet)
  - Выявление потенциала продаж данному клиенту
  - Подготовка адресных маркетинговых кампаний и предложений
  - Прогнозирование реакции клиентов на новые предложения
- **Методология анализа жизненного цикла клиентов**
  - Сегментация клиентов по стадиям их жизненного цикла
  - Выявление групп риска (уходящих клиентов)
  - Оценка степени лояльности клиентов
  - Подготовка адресных маркетинговых кампаний и предложений

# Выявление потенциала продаж: **Логика модели**

## Если мы знаем:

- тип деятельности клиента
- его «типичную» структуру потребления
- относительный масштаб бизнеса

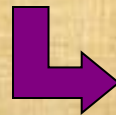
## То мы можем предсказать:

- Недопотребление
- Потенциал продаж каждому клиенту

**ВХОД:** Данные об операциях целевой группы клиентов



1. Стартовая подготовка данных
2. Выбор репрезентативного набора продуктов
3. Сегментация по типу потребления (кластеризация)
4. Построение референтных корзин потребления
5. Оценка «недопотребления» для каждого клиента
6. Структурирование потенциала продаж



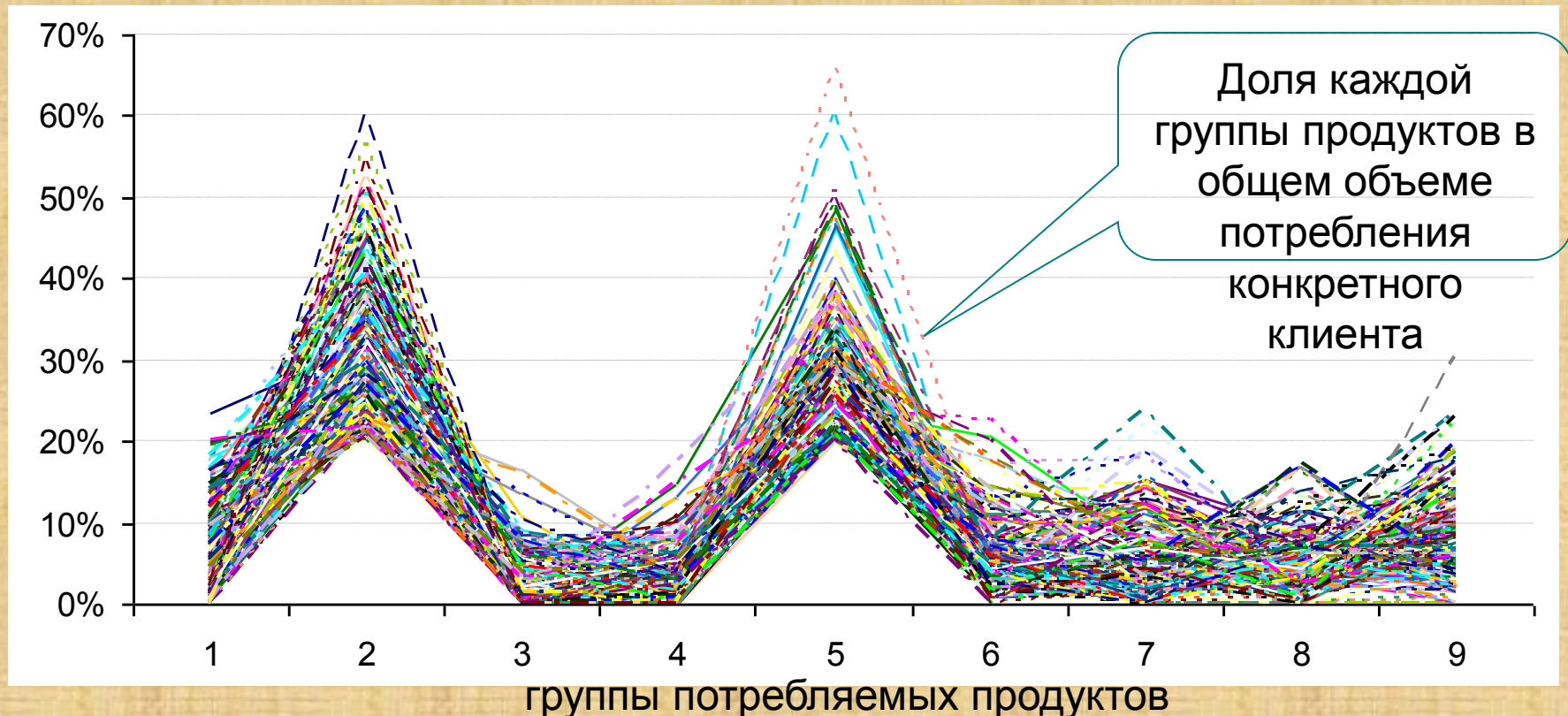
**ВЫХОД:** Результаты оценок:

- Какие продукты недопотребляются?
- Какими клиентами?
- В каком объеме?

# Метод туннельной кластеризации

Позволяет выявить в большом множестве данных об операциях подгруппы клиентов, имеющих аналогичной структурой потребления. То есть, позволяет сегментировать клиентскую базу по типу потребления продуктов компании.

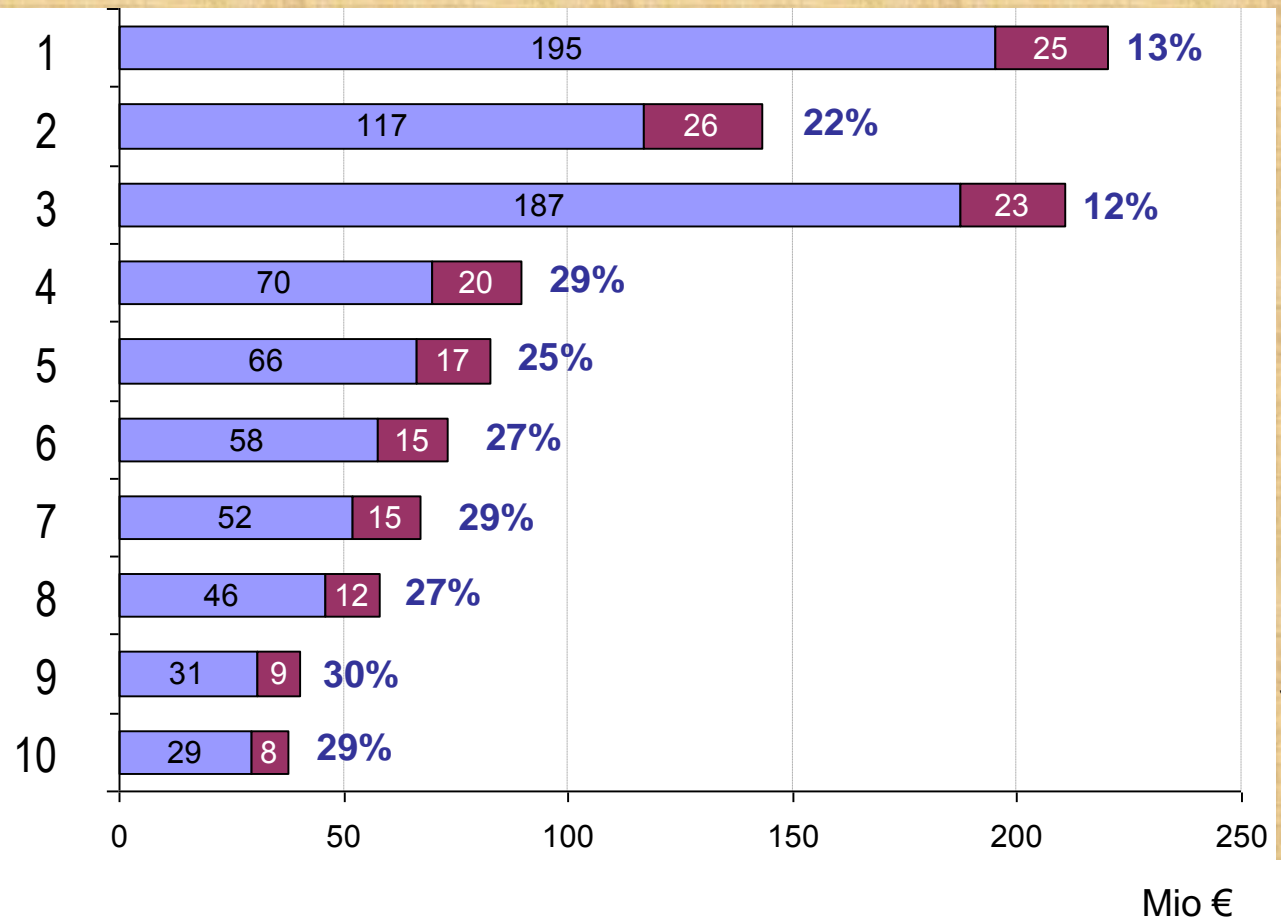
Пример кластера с однородной структурой потребления:





# Пример одного из возможных результатов

Целевая группа в 6000 клиентов может обеспечить 20% рост продаж по 10 наиболее покупаемым типам продуктов



# Некоторые характеристики модели

## Специально разработанные методы

- Алгоритм стартовой подготовки данных
- Метод определения потребности клиента
- Метод туннельной кластеризации
- Модель построения референтных корзин
- Алгоритм оценки потенциала продаж

## Настраиваемые параметры модели

- Уровень детализации для сегментации клиентов
- Уровень детализации анализа потребления (типы продуктов – продукты – их модификации)
- Критерии однородности потребления
- Допустимая погрешность кластеризации
- Набор используемых атрибутов

## Характеристики применения

- Обработка больших объемов данных
- Оценка на уровне каждого клиента
- Высокая точность прогнозирования
- Легко используемые результаты
- Возможность местной настройки
- Надежности и устойчивость

## Использованные стат. методы

- Стохастический анализ
- Многовекторная кластеризация
- Многокритериальное агрегирование
- Корреляционный и факторный анализ

## Фундаментальные теории

- Общая теория выбора (Гипотеза рационального поведения)
- Неклассические модели теории полезности

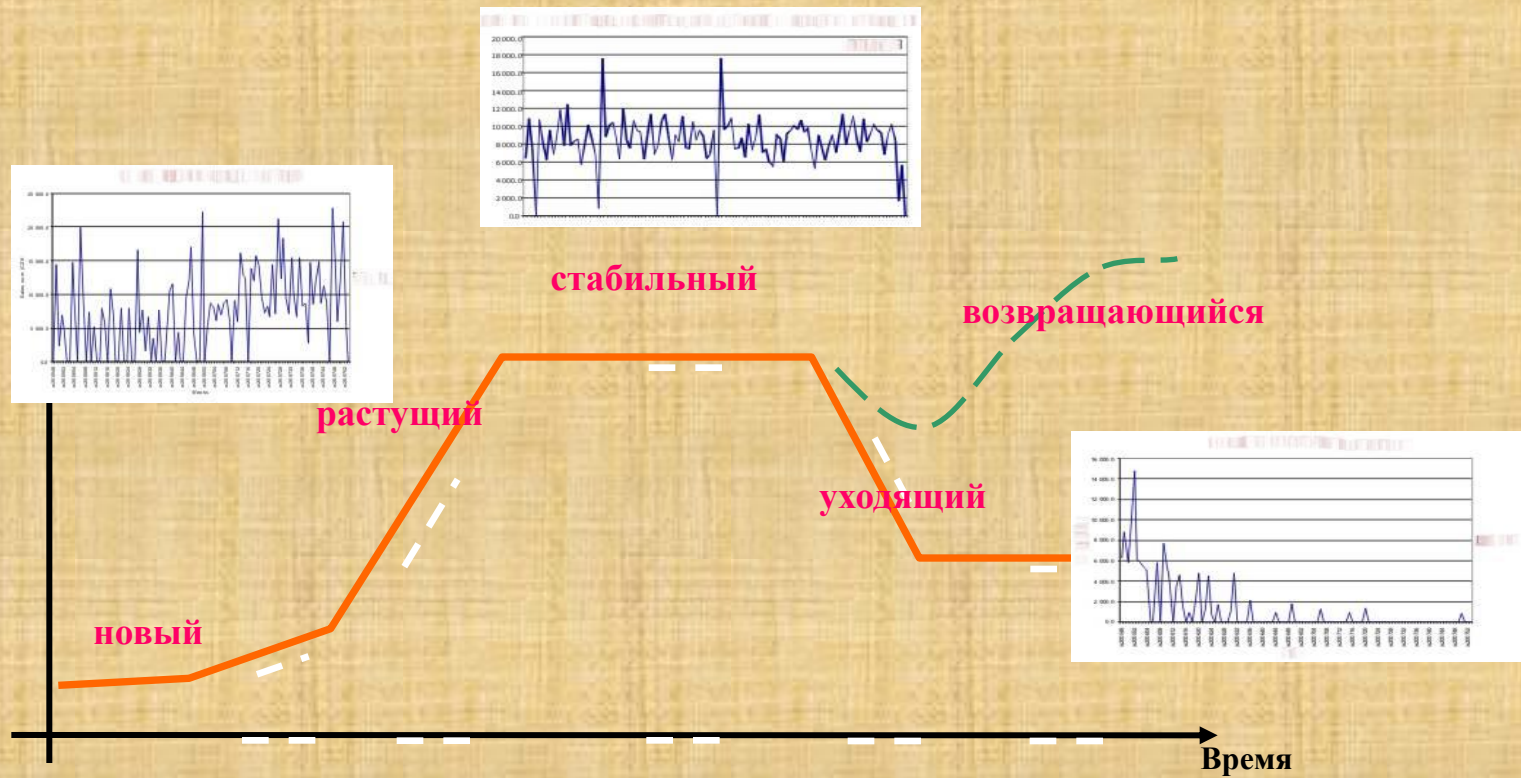
# Методология анализа жизненного цикла клиентов



**Проблема:** Как выявить устойчивые линии поведения клиента при высоких колебаниях значений?

# Анализ жизненного цикла клиентов

График жизненного цикла клиента





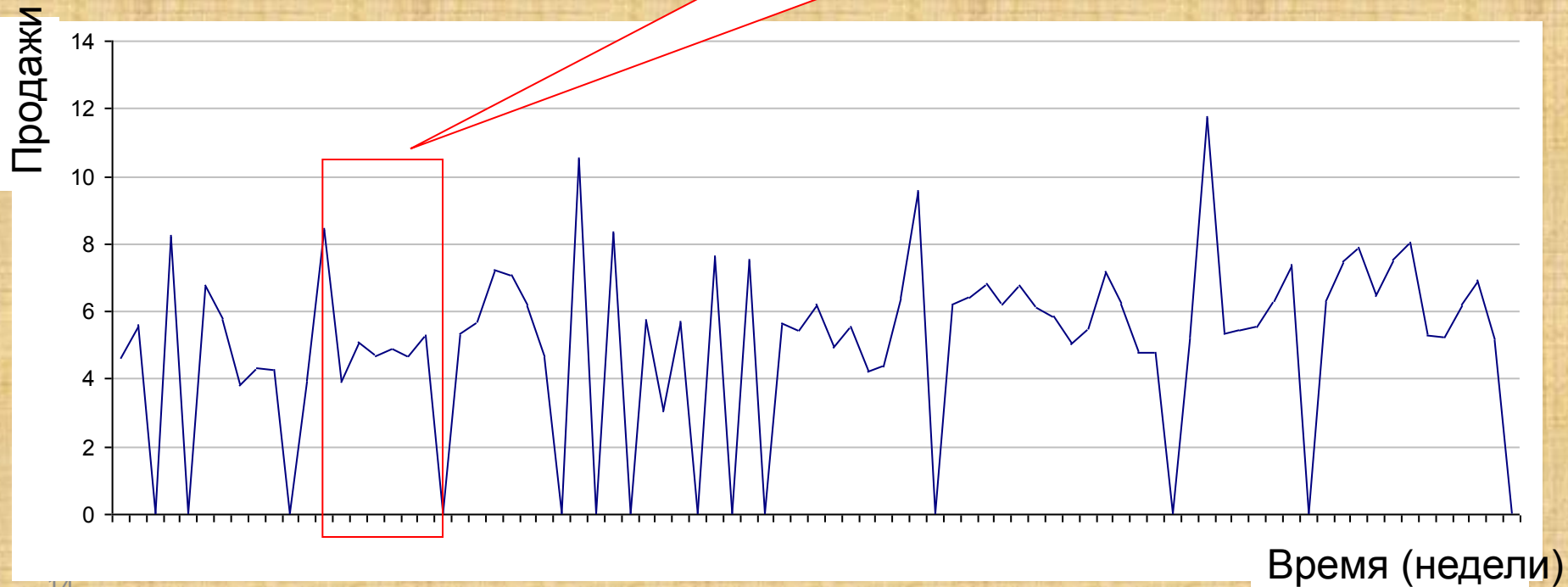
## Пример графика операций клиента (временной ряд)

**Это – уходящий клиент?**



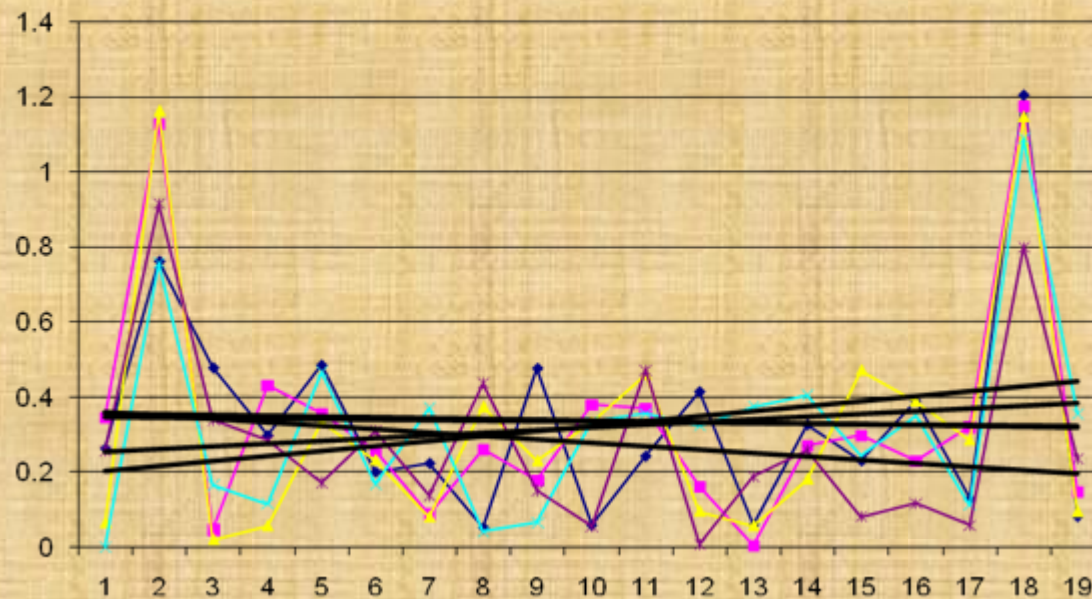
## Пример графика операций клиента (продолжение)

**Нет! Это стабильный клиент.**



# Линейный тренд

Графики потребления с двумя пиками, в начале и в конце периода.  
Тренды различны



Как видно из рисунка, после пика потребления идет период низкого потребления, заканчивающийся ростом потребления. Однако, тренды этих клиентов (см. черные прямые) показывают как падение, так и рост для клиентов с одинаковой моделью потребления.

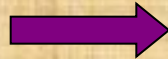
# Сегментация клиентов по их жизненному циклу

## Вход:

Данные об  
операциях целевой  
группы клиентов



Библиотека  
паттернов



1. Выявление типа поведения для каждого клиента
2. Привязка к образцу из Библиотеки
3. Выявление текущей стадии жизненного цикла
4. Присвоение атрибута ЖЦ каждому клиенту
5. Прогноз ближайших изменений для клиента
6. Обновление Библиотеки образцов
7. Формирование результатов работы модели



## Выход:

1. Значения атрибута жизненного цикла для каждого клиента
2. Перечень клиентов, требующих срочного внимания
3. Прогноз ближайших изменений атрибута ЖЦ для каждого клиента



# Некоторые характеристики модели

## Специально разработанные методы

- Алгоритм стартовой подготовки данных
- Алгоритм вычисления Индекса Неравномерности
- Метод ретроспективного анализа нестационарных временных рядов
- Алгоритм формирования Библиотеки паттернов

## Настраиваемые параметры модели

- Критерии группировки клиентов на входе
- Критерии выделения стадий жизненного цикла
- Допустимая ошибка сегментации
- Набор используемых атрибутов
- Формат представления выходных данных

## Характеристики применения

- Сегментация для 100% клиентов
- Самообучаемость (Библиотека)
- Обработка больших объемов данных
- Оценки для каждого клиента
- Высокая точность прогнозирования
- Легко применимые результаты
- Гибкость в настройке по месту
- Надежность и устойчивость

## Используемые стат. методы

- Динамический анализ поведения
- Анализ нестационарных временных рядов
- Вероятностные методы сегментирования

## Фундаментальные теории

- Общая теория выбора (Гипотеза рационального поведения)
- Неклассические модели теории полезности

# Аналитика помогает получать преимущества

Модели предоставляют надежные сведения о поведении клиентов компании.

Анализируя их и задавая более детальные вопросы ...

- С чем связаны изменения в поведении данного клиента?
- Что может быть причиной для снижения продаж данному клиенту?
- Почему именно клиент не покупает «недопотребленные» продукты?
- Позволит ли более комплексное предложение стимулировать клиента на покупки?
- Что следует включить в персональное предложение для клиента?
- ...

... компании повышают лояльность клиентов, улучшают результаты своих маркетинговых инициатив и получают преимущества перед своими конкурентами.

**Спасибо за внимание !**