
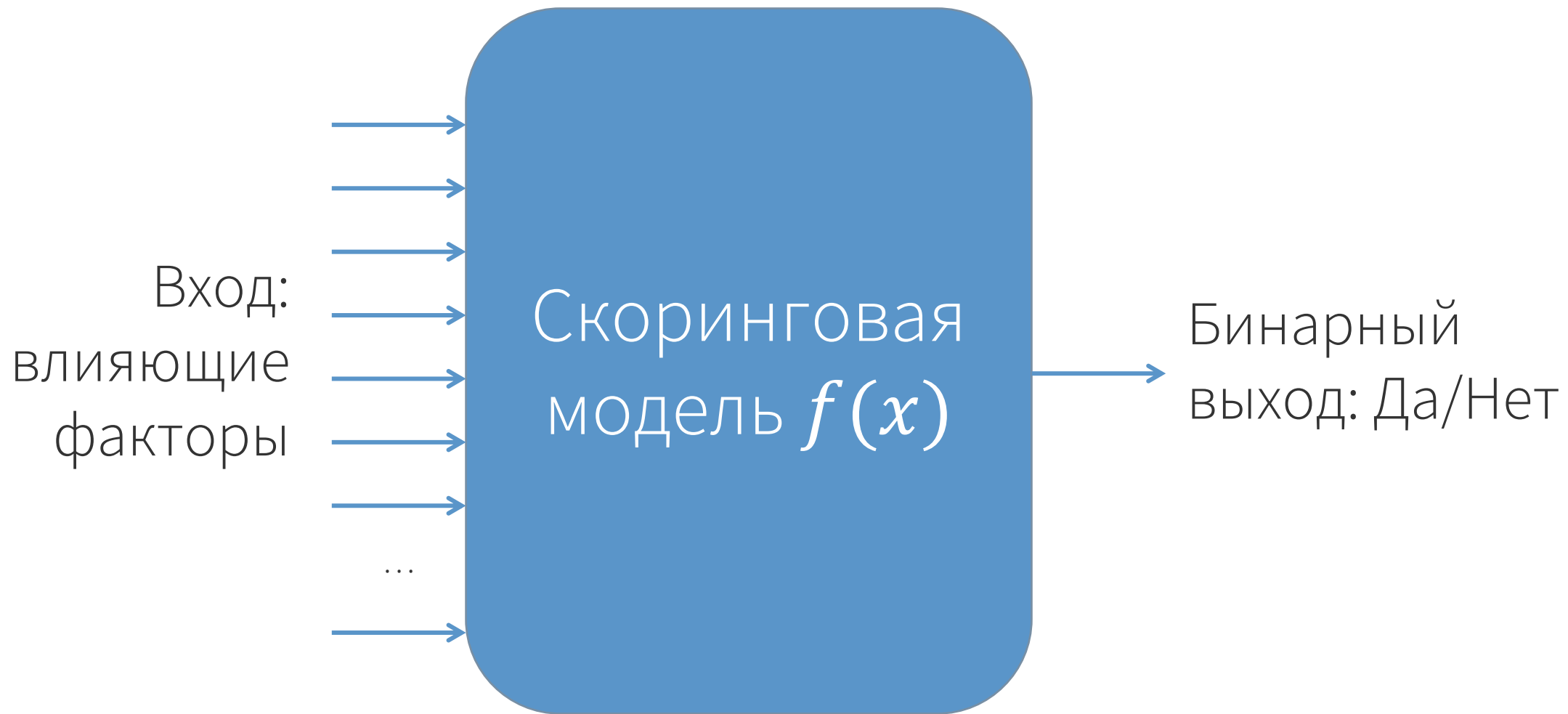


Loginom Scorecard  
Modeler – построение  
скоринговых карт

Скоринг – метод решения задачи бинарной классификации – одной из популярных задач бизнес-аналитики:

- Вернет кредит / Не вернет
  - Будет страховой случай / Не будет
  - Будет поломка / Не будет
  - Лояльный клиент / Не лояльный
- 





Скоринг —  
отраслевой стандарт  
оценки рисков в  
сфере кредитования.

# Проблемы управления без скоринга



Влияние  
человеческого  
фактора



Отсутствие  
объективной  
оценки



Учет малого  
количества  
факторов



Большая доля  
ручной работы



Сложность  
управления  
качеством  
портфеля



Построение  
скоринговых карт

```
graph TD; A[Построение скоринговых карт] --> B[Встраивание карты в бизнес-процесс]
```

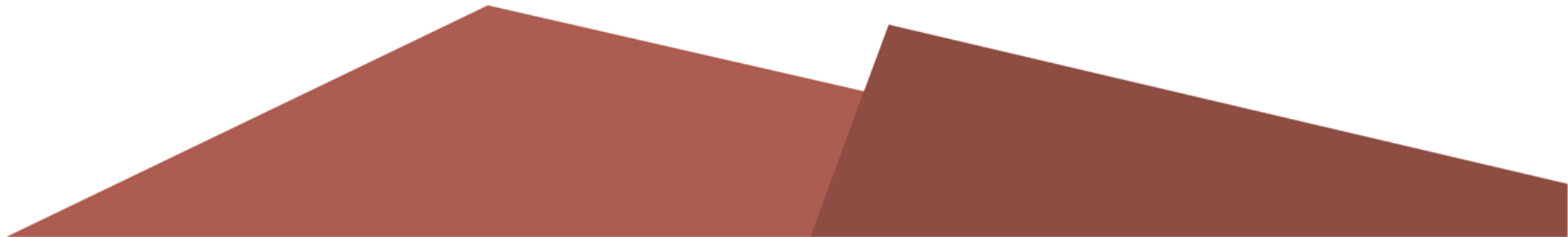
Встраивание карты  
в бизнес-процесс

Logiном Scorecard  
Modeler решает одну  
из двух задач –  
**построение карты.**

Loginom Scorecard Modeler — комплексное решение, автоматизирующее процесс построения скоринговых карт.

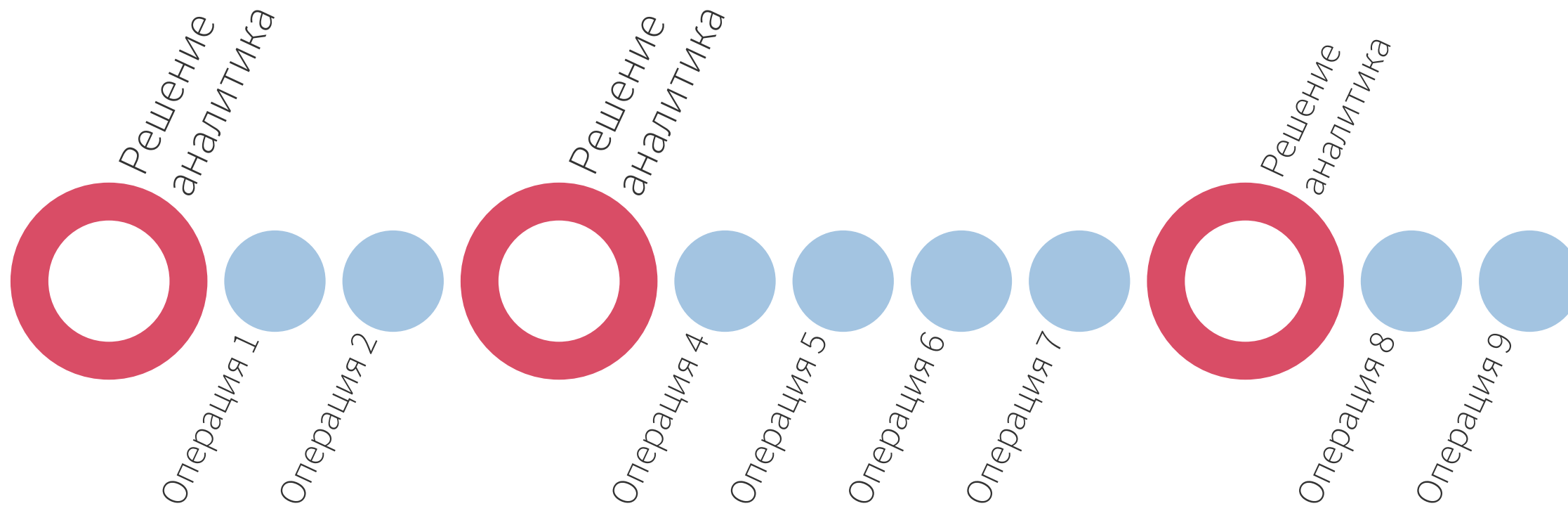
Результат работы – модель, позволяющая на основании множества характеристик заемщика оценивать риски и вероятность возврата им кредита.

# Описание решения





# Автоматизируем все рутинные операции

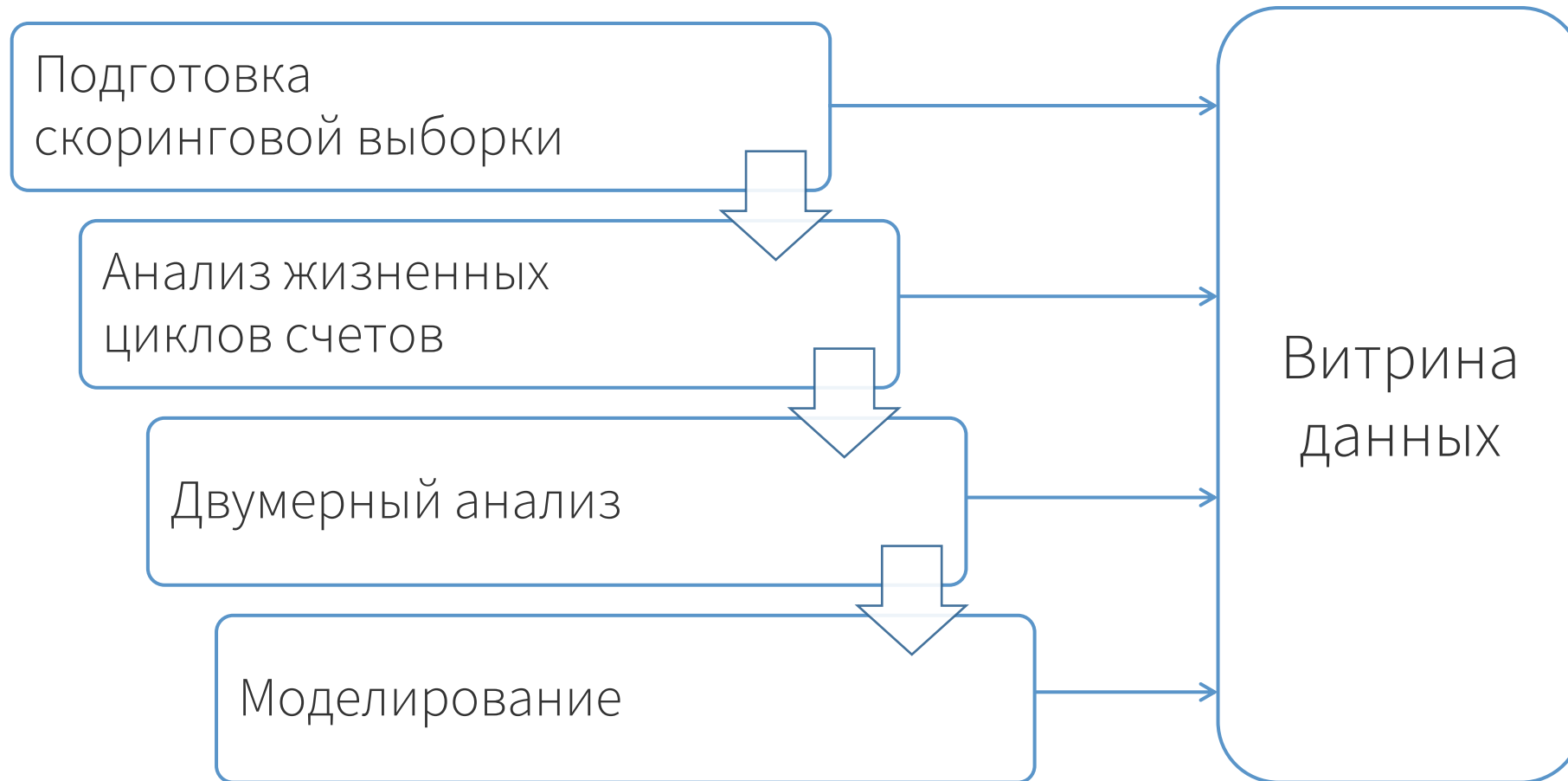


# Подготовка к работе

В начале нужно задать основные параметры:

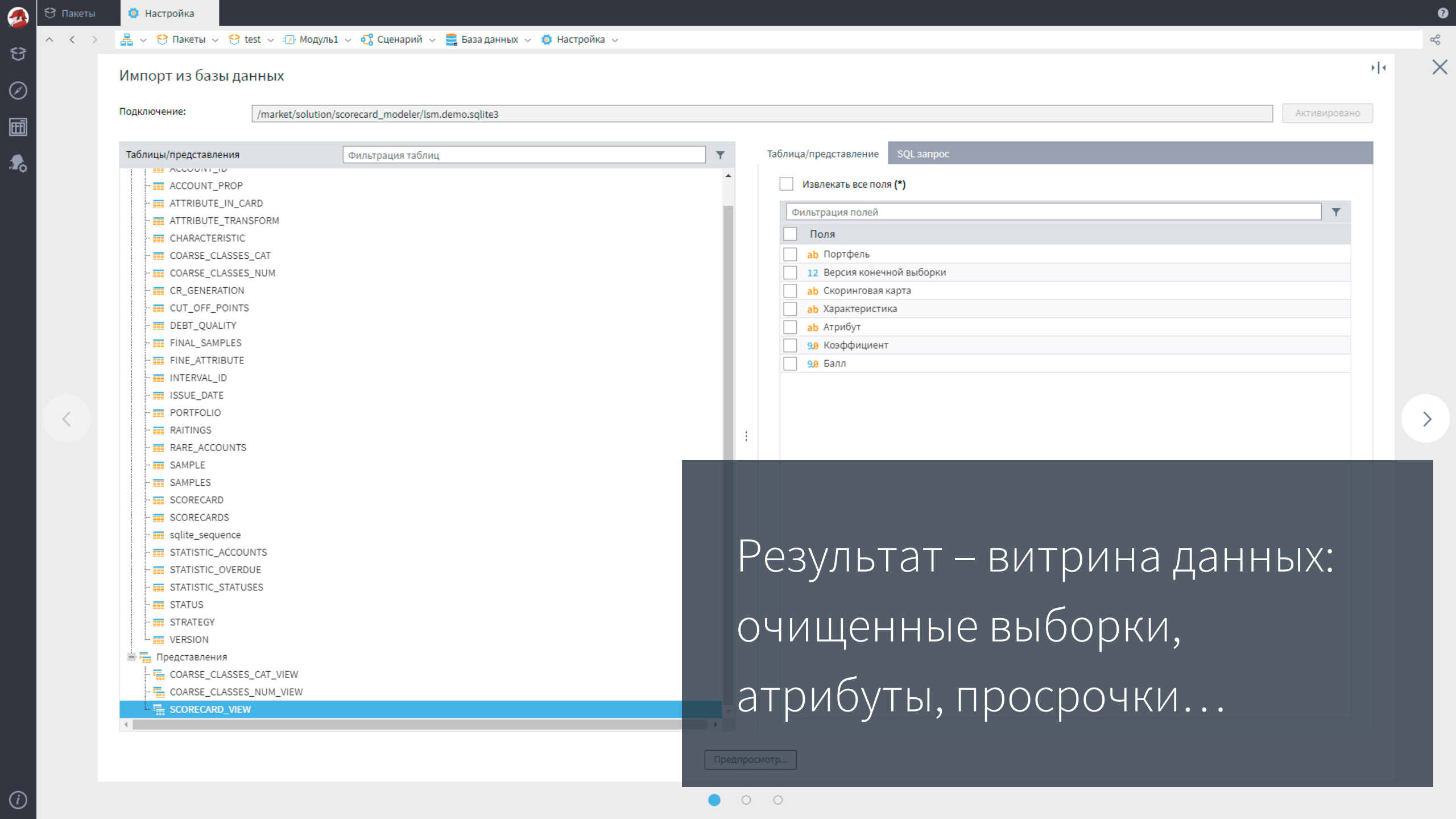
- Название портфеля
- Начальная и конечная дата учета счетов
- Минимальная частота редкого атрибута
- Разбиение на обучающую и тестовую выборку
- ...

# Разработка скоринговой карты



# Подготовка скоринговой выборки

- Загрузка счетов
- Загрузка характеристик
- Загрузка просрочек (опционально)
- Загрузка флагов хороший/плохой (опционально)
- Мониторинг корректности загрузки данных
- Формирование начальных выборок



## Импорт из базы данных

Подключение:

/market/solution/scorecard\_modeler/lsm.demo.sqlite3

Активировано

Таблицы/представления

Фильтрация таблиц

- ACCOUNT\_ID
- ACCOUNT\_PROP
- ATTRIBUTE\_IN\_CARD
- ATTRIBUTE\_TRANSFORM
- CHARACTERISTIC
- COARSE\_CLASSES\_CAT
- COARSE\_CLASSES\_NUM
- CR\_GENERATION
- CUT\_OFF\_POINTS
- DEBT\_QUALITY
- FINAL\_SAMPLES
- FINE\_ATTRIBUTE
- INTERVAL\_ID
- ISSUE\_DATE
- PORTFOLIO
- RAITINGS
- RARE\_ACCOUNTS
- SAMPLE
- SAMPLES
- SCORECARD
- SCORECARDS
- sqlite\_sequence
- STATISTIC\_ACCOUNTS
- STATISTIC\_OVERDUE
- STATISTIC\_STATUSES
- STATUS
- STRATEGY
- VERSION
- Представления
  - COARSE\_CLASSES\_CAT\_VIEW
  - COARSE\_CLASSES\_NUM\_VIEW
  - SCORECARD\_VIEW**

Таблица/представление

SQL запрос

Извлекать все поля (\*)

Фильтрация полей

- Поля
- ab Портфель
- 12 Версия конечной выборки
- ab Скоринговая карта
- ab Характеристика
- ab Атрибут
- 9.0 Коэффициент
- 9.0 Балл

Предпросмотр...

Результат – витрина данных:  
очищенные выборки,  
атрибуты, просрочки...

# Анализ жизненных циклов

- Аудит данных по просрочкам
- Винтажный анализ
- Построение матрицы миграции
- Определение статуса счета
- Балансировка классов
- Загрузка в витрину данных

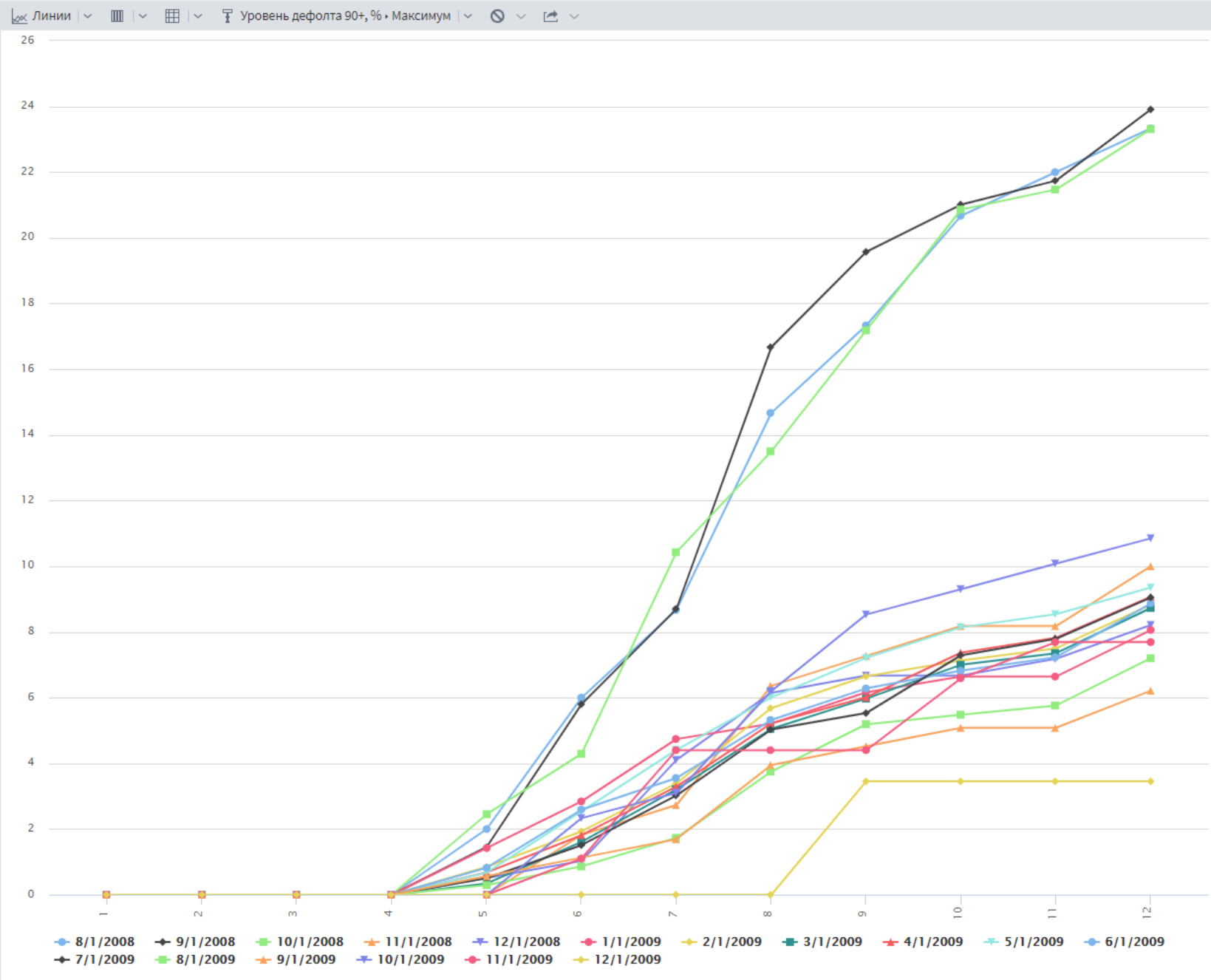
Аудит данных по просрочкам выявляет грубые ошибки и проблемные счета, которые могут внести неточность в винтажный анализ :

- Проверка на пропуски в начальной истории и в течение жизненного цикла счета;
- Проверка правильности движения просрочки по группам;
- Проверка заполнения просрочки по методам ее начисления LIFO/FIFO



		Поколение			
		Σ Факты			
Счёт.Период		8/1/2008	9/1/2008	10/1/2008	11/1/200
1		0.00	0.00	0.00	0
2		0.00	0.00	0.00	0
3		0.00	0.00	0.00	0
4		0.00	0.00	0.00	0
5		2.00	1.45	2.45	0
6		6.00	5.80	4.29	1
7		8.67	8.70	10.43	2
8		14.67	16.67	13.50	6
9		17.33	19.57	17.18	7
10		20.67	21.01	20.86	8
11		22.00	21.74	21.47	8
12		23.33	23.91	23.31	10
Итого:		23.33	23.91	23.31	10

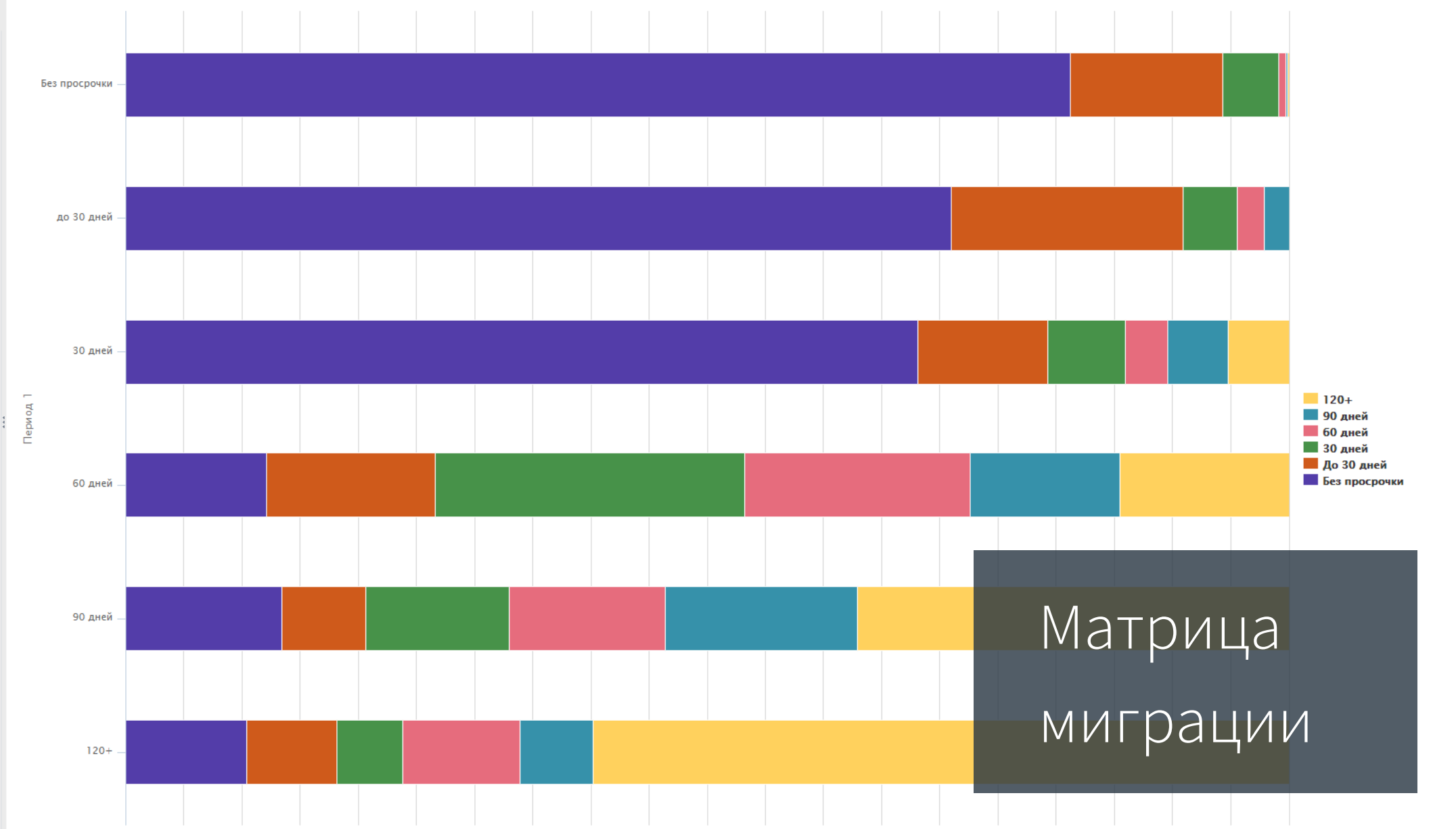
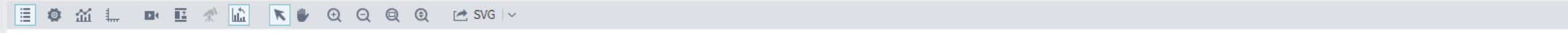
Винтажный анализ – расчет накопительной суммы потерь поколений счетов





Поля

- ab Период 1
- 9.0 Без просрочки
- 9.0 До 30 дней
- 9.0 30 дней
- 9.0 60 дней
- 9.0 90 дней
- 9.0 120+



- 120+
- 90 дней
- 60 дней
- 30 дней
- До 30 дней
- Без просрочки

Матрица миграции

# Определение статусов счетов

- Определение события хороший/плохой
- Разделение на обучающую и тестовую выборки
- Балансировка плохих и хороших счетов
- Загрузка в витрину данных

### Разбиение на множества

Состояние входа:  [Активировать](#)

Общее число записей:

Множество	Способ	% Размер в процентах	Размер в строках
Обучающее	% =	<input type="text" value="x/a"/> Обучающая выборка, %	<input type="text" value="x/a"/> Обучающая выборка, %
Тестовое	% =	<input type="text" value="x/a"/> Тестовая выборка, %	<input type="text" value="x/a"/> Тестовая выборка, %
Итого:		100,00%	100

Метод сэмпинга:

#### Параметры метода

Приоритет тестового множества


Положение приоритетного тестового множества

- Фильтрация
- Поля, определяющие страты
  - 9.0 Группа редкого счета
  - ab Статус
  - 9.0 Хороший
  - 0/1 Флаг HARD\_BAD
  - 0/1 Служебный фильтр для HARD\_BADS
  - 9.0 Плохой
  - 12 Выделять ли неопределенные счета
  - 12 Включать счета HARD-BAD
  - 12 Включать счета HARD-BAD незрелые
  - ab Портфель
  - ab Счет.Код
  - 12 Закрытая сделка
  - 9.0 1-30
  - 9.0 31-60
  - 9.0 61-90
  - 9.0 91-120

Полнота списка уникальных значений

Разделение на обучающую и тестовую выборки

# Двумерный анализ

- Квантование непрерывных данных
  - Квантование дискретных данных
  - Отбор значимых факторов
  - Загрузка в витрину данных
- 

# Двумерный анализ: конечные классы

Необходимо квантование, т.к. скоринговая карта строится на категориальных атрибутах. Конечные классы – механизм поиска компромисса между:

- Точностью
- Интерпретируемостью
- Значимостью

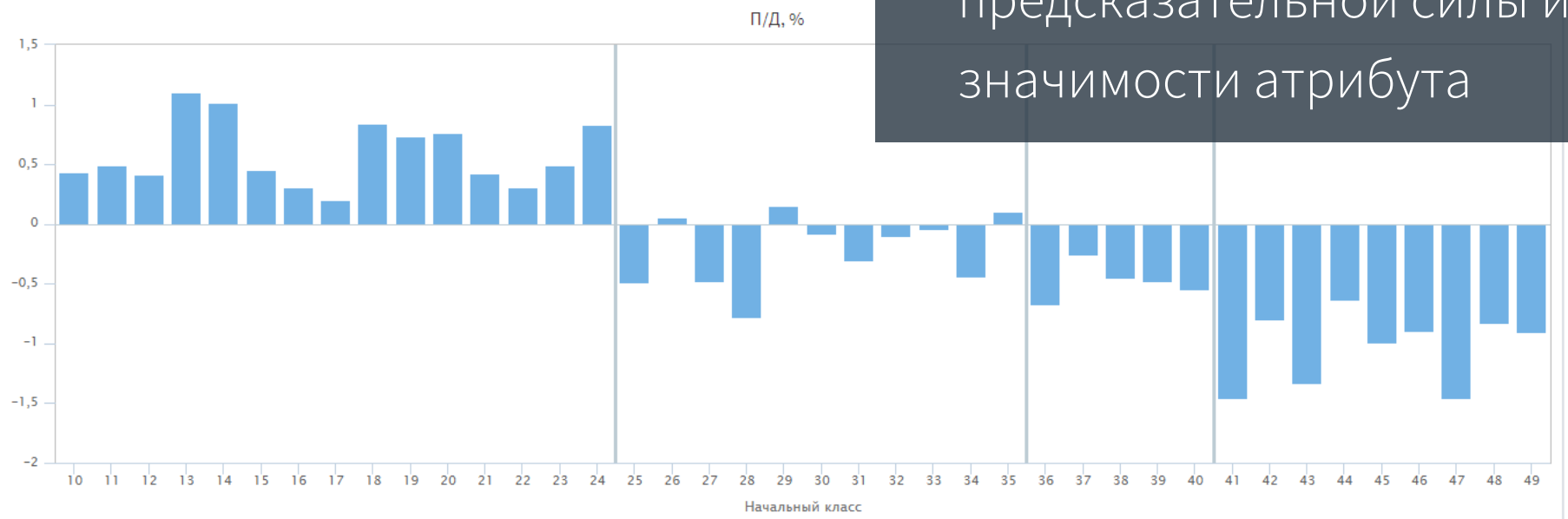


### Настройка конечных классов

Состояние входа: Вход активирован



Столбец	Индекс
9.0 Возраст	0,04
9.0 П/Д, %	0,35
9.0 Стаж в отрасли	0,11
9.0 Стаж на последнем ...	0,15



**Конечные классы**

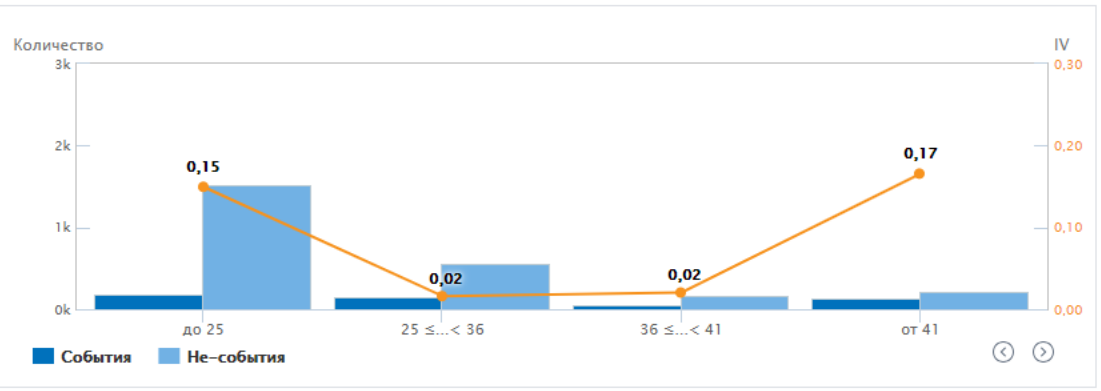
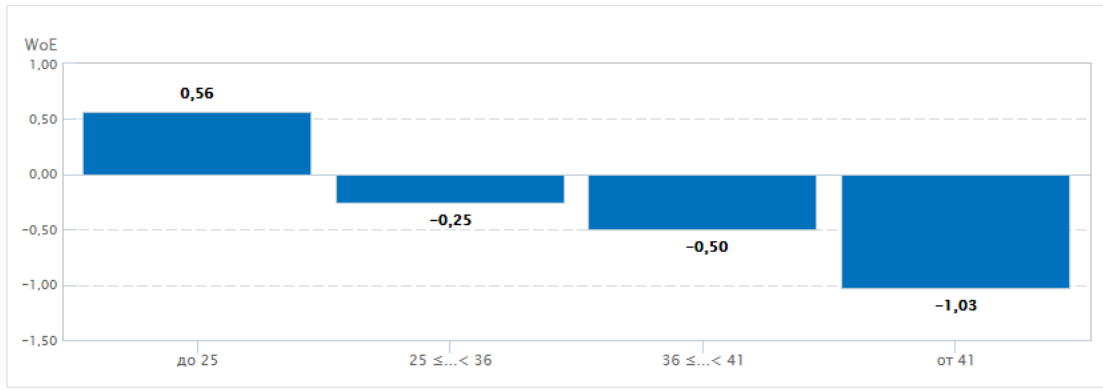
Минимальный вес, %:

Максимальное кол-во:


**Оптимизация**

Равномерность, %:

Выбор лучшего варианта квантования: оценка предсказательной силы и значимости атрибута



# Моделирование

- Построение модели
  - Оценка качества модели
  - Преобразование коэффициентов регрессии в скоринговые баллы
  - Загрузка карты в витрину данных
- 

### Выбор диаграммы

- ROC-кривая
- PR-кривая
- Базовые показатели
- Диаграмма точности
- Диаграмма равновесия
- % распознанных событий
- Диаграмма роста
- Диаграмма отклика
- Диаграмма выигрыша

Кумулятивная

10 диапазонов

### Множества

Обучающее  Тестовое

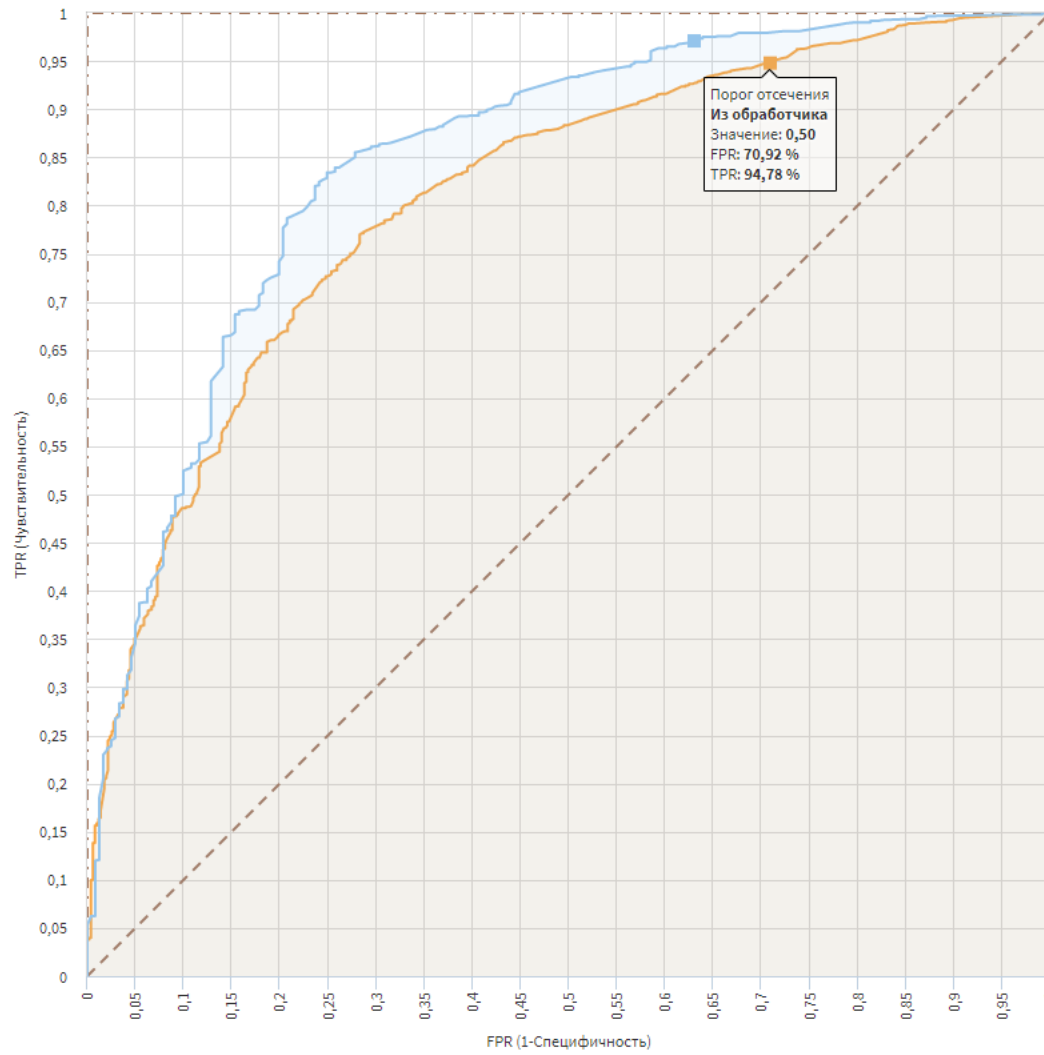
### Порог отсеечения

Из обработчика

Значение порога:

### ROC-кривая

Событие: Статус = хороший



- Обучающее множество
- Порог отсеечения
- - - Базовая линия
- - - Идеальная линия
- Тестовое множество
- Порог отсеечения

### Оценки классификации

Показатель	Множества	
	Обучающее	Тестовое
Оценки классификатора		
AUC ROC	0,8087	0,8482
AUC PR	0,9555	0,9585
Коэффициент Джини	0,6174	0,6964
KS	48,7831	58,6147
Порог отсеечения: Из обработчика		
Значение	0,5000	0,5000
TPR (Чувствительность)	0,9478	0,9709
TNR (Специфичность)	0,2908	0,3693
FPR (1-Специфичность)	0,7092	0,6307
PPV	0,8824	0,8847
F1 Score	0,9139	0,9258
MCC	0,3014	0,4527

### Матрицы ошибок

Классифицировано	Фактически		Итого
	Событие	Не-событие	
Обучающее	2 857	509	
Событие	2 708	361	3 069
Не-событие	149	148	297
Тестовое	1 201	241	
Событие	1 166	152	1 318
Не-событие	35	89	124

### Распознано

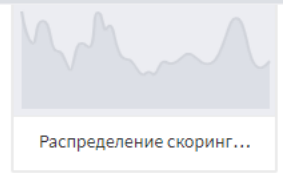
Обучающее	2 856/3 366
Тестовое	1 255/1 442

Оценка качества модели регрессии



- Визуализаторы
- Табличное представление
  - Диаграмма
  - Куб
  - Статистика
  - Таблица

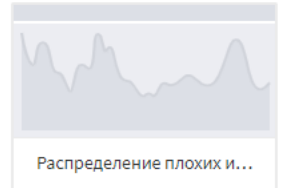
Переименовать Удалить Удалить все Клонировать Добавить в отчеты



+

Добавить визуализатор

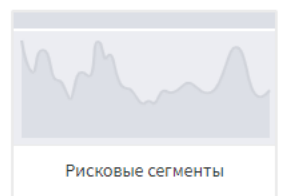
Распределение плохих и хороших



+

Добавить визуализатор

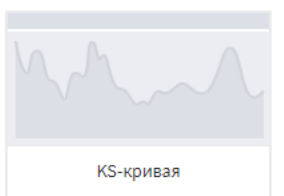
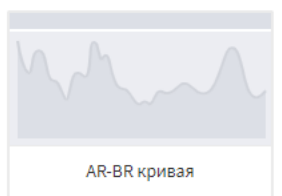
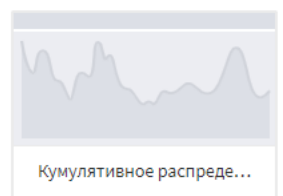
Рисковые сегменты



+

Добавить визуализатор

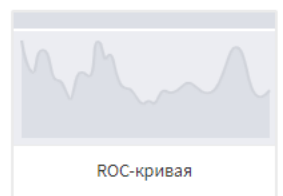
Кумулятивное распределение счетов, AR-BR кривая, KS-кривая



+

Добавить визуализатор

ROC, Lift кривая(CAP)



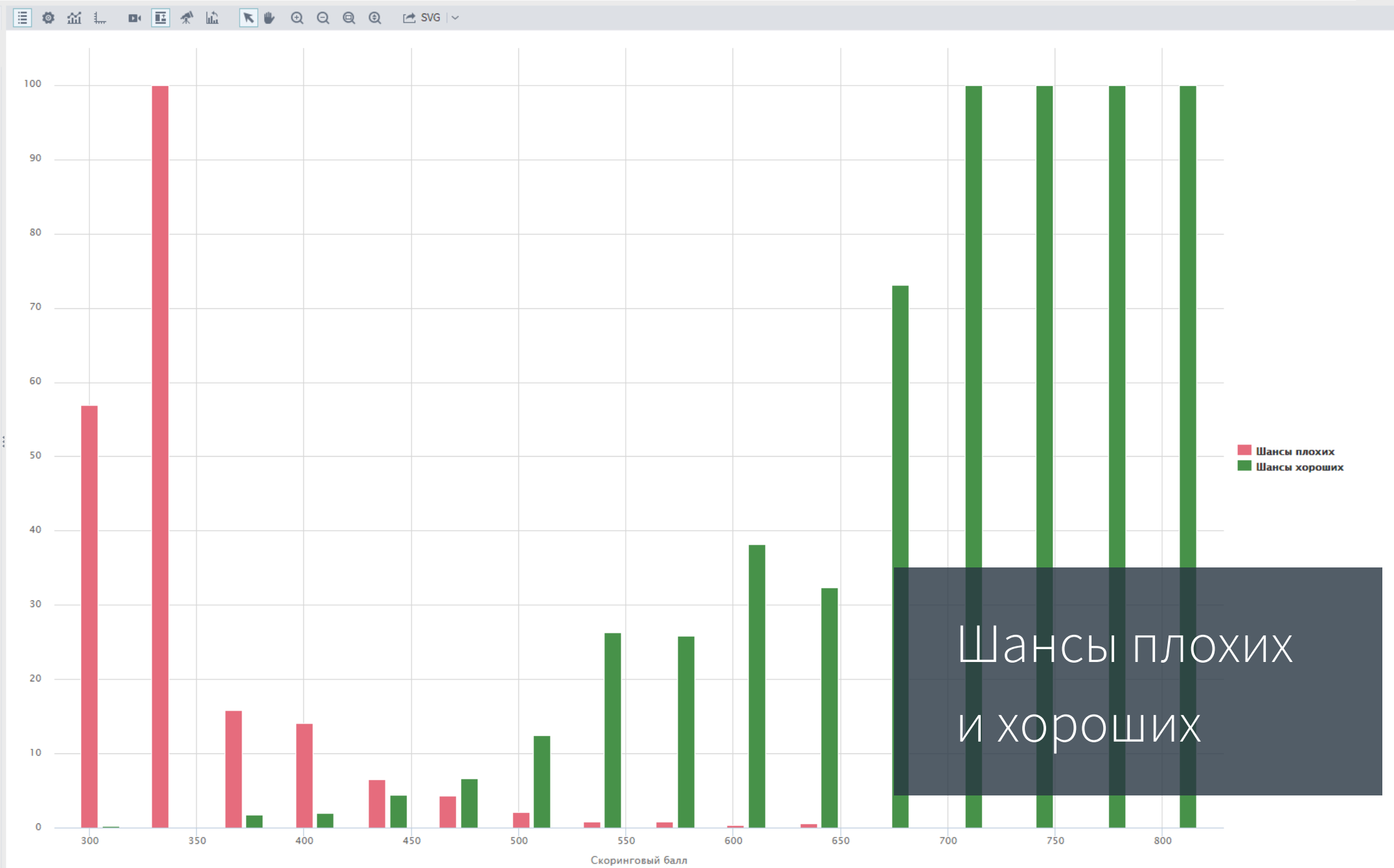
+

Добавить визуализатор

# Оценка скоринговой карты – все показатели

Поля

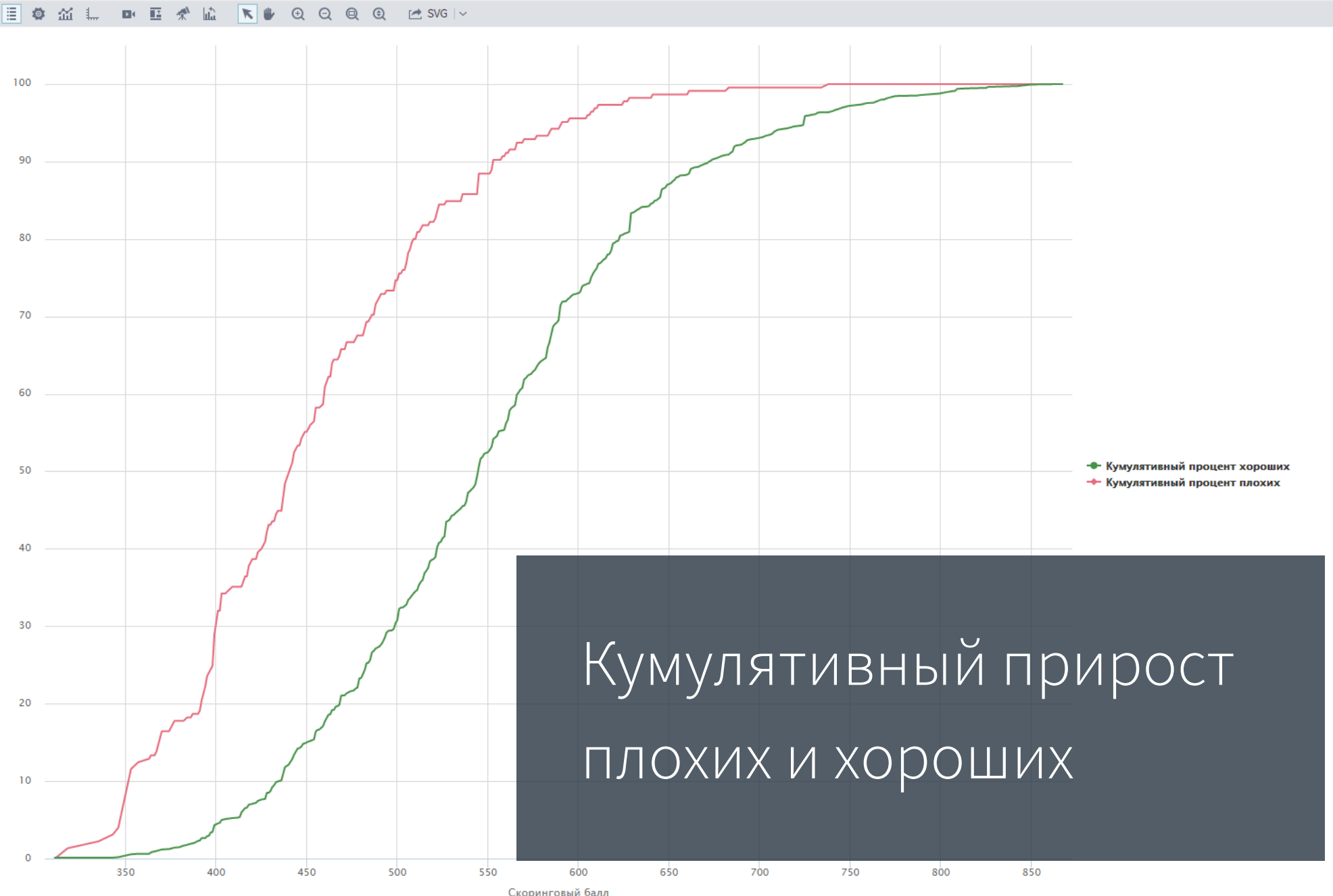
- 9.0 Шансы плохих
- 9.0 Шансы хороших
- 12 Скоринговый балл



Шансы плохих  
и хороших



- Поля
- 9.0 Уровень одобрений, %
  - 9.0 Уровень дефолтов, %
  - 9.0 Кумулятивный процент плохих
  - 9.0 Кумулятивный процент хороших
  - 9.0 KS
  - 12 Скоринговый балл



Кумулятивный прирост плохих и хороших

#	ab Скоринговая карта	ab Характеристика	ab Атрибут	9.0 Балл	9.0 Коэффициенты	9.0 Уровень дефолтов %	12 Общее количество счетов	9.0 Количество плохих счетов	9.0 Количество хороших счетов
1	Скоринговая карта demo 2	Constant_LR	Константа	301	0.4260761308				
2	Скоринговая карта demo 2	Возраст	49 ≤... < 51	73	1.095196529	14.29	98	14	84
3	Скоринговая карта demo 2	Возраст	51 ≤... < 54	0	-0.1685932437	8.97	156	14	142
4	Скоринговая карта demo 2	Возраст	до 49	10	0	13.41	1178	158	1020
5	Скоринговая карта demo 2	Возраст	от 54	27	0.3054731411	9.31	419	39	380
6	Скоринговая карта demo 2	Дети	[[пусто]]	31	0.5368556024	11.11	324	36	288
7	Скоринговая карта demo 2	Дети	[1; 2 и более]	63	1.094869612	5.44	662	36	626
8	Скоринговая карта demo 2	Дети	[отсутствуют]	0	0	17.69	865	153	712
9	Скоринговая карта demo 2	Карты	[дебетовая]	122	2.11595942	3.91	230	9	221
10	Скоринговая карта demo 2	Карты	[кредитная]	84	1.463244528	4.95	666	33	633
11	Скоринговая карта demo 2	Карты	[нет]	0	0	19.16	955	183	772
12	Скоринговая карта demo 2	Кредитная история	[отрицательная; нет; средняя]	0	0	13.42	1639	220	1419
13	Скоринговая карта demo 2	Кредитная история	[положительная]	180	3.126216946	2.36	212	5	207
14	Скоринговая карта demo 2	П/Д, %	25 ≤... < 36	46	-0.6711424564	14.29	441	63	378
15	Скоринговая карта demo 2	П/Д, %	36 ≤... < 41	48	-0.6360535118	22.61	115	26	89
16	Скоринговая карта demo 2	П/Д, %	до 25	85	0	6.47	1067	69	998
17	Скоринговая карта demo 2	П/Д, %	от 41	0	-1.475672266	29.39	228	67	161
18	Скоринговая карта demo 2	Стаж в отрасли	4 ≤... < 8	0	-0.1171854225	9.49	158	15	143
19	Скоринговая карта demo 2	Стаж в отрасли	8 ≤... < 13	33	0.443829959	7.06	170	12	158
20	Скоринговая карта demo 2	Стаж в отрасли	до 4	0	-0.1207774157	12.28	114	14	100
21	Скоринговая карта demo 2	Стаж в отрасли	Класс пропусков	7	0	14.27	1254	179	1075
22	Скоринговая карта demo 2	Стаж в отрасли	от 13	68	1.05571469	3.23	155	5	150
23	Скоринговая карта demo 2	Стаж на последнем месте	13 ≤... < 17	59	0.4086476633	10.32	339	35	304
24	Скоринговая карта demo 2	Стаж на последнем месте	3 ≤... < 13	35	0	14.08	774	109	665
25	Скоринговая карта demo 2	Стаж на последнем месте	до 3	0	-0.606216647	19.8	101	20	81
26	Скоринговая карта demo 2	Стаж на последнем месте	Класс пропусков	79	0.7580670382	8.43	534	45	489
27	Скоринговая карта demo 2	Стаж на последнем месте	от 17	39	0.07027888202	15.53	103	16	87

Итог моделирования –  
скоринговая карта

# Комплект поставки:

1. Пакеты с компонентами
2. Тематическая витрина данных
3. Демонстрационные данные и сценарии
4. Документация

## Компоненты (packages)

Файл	Назначение
подключения.lgp	Пакет для настройки подключений к базам данных в компонентах решений. По умолчанию пакет с подключением настроен на lsm.sqlite3.
загрузка счетов и характеристик.lgp	Загрузки исходных данных по счетам и их характеристикам в витрину данных
флаги плохим (статусы известны).lgp	Загрузка в витрину данных имеющихся статусов по счетам
конечные классы дискретные.lgp	Формирование конечной выборки для дискретных характеристик начальной выборки и загрузка в витрину
конечные классы непрерывные.lgp	Формирование конечной выборки для непрерывных характеристик начальной выборки и загрузка в витрину
переменные для моделирования и оценки качества.lgp	Задание всех необходимых параметров для моделирования и оценки качества
подготовка данных на регрессию.lgp	Подготовка конечной выборки для моделирования
оценка качества регрессии.lgp	Расчет основных метрик качества модели <b>логистической регрессии</b>
масштабирование скоринговой карты.lgp	Масштабирование модели логистической регрессии в <b>скоринговую карту</b> . По умолчанию используется стандарт FICO.
расчет баллов по счетам.lgp	Расчет скоринговых баллов по счетам обучающей и тестовой выборок
вероятности дефолтов.lgp	Расчет вероятности дефолтов по счетам обучающей и тестовой выборок и экспорт результатов в витрину

# Пакеты с готовыми компонентами

## Импорт из базы данных

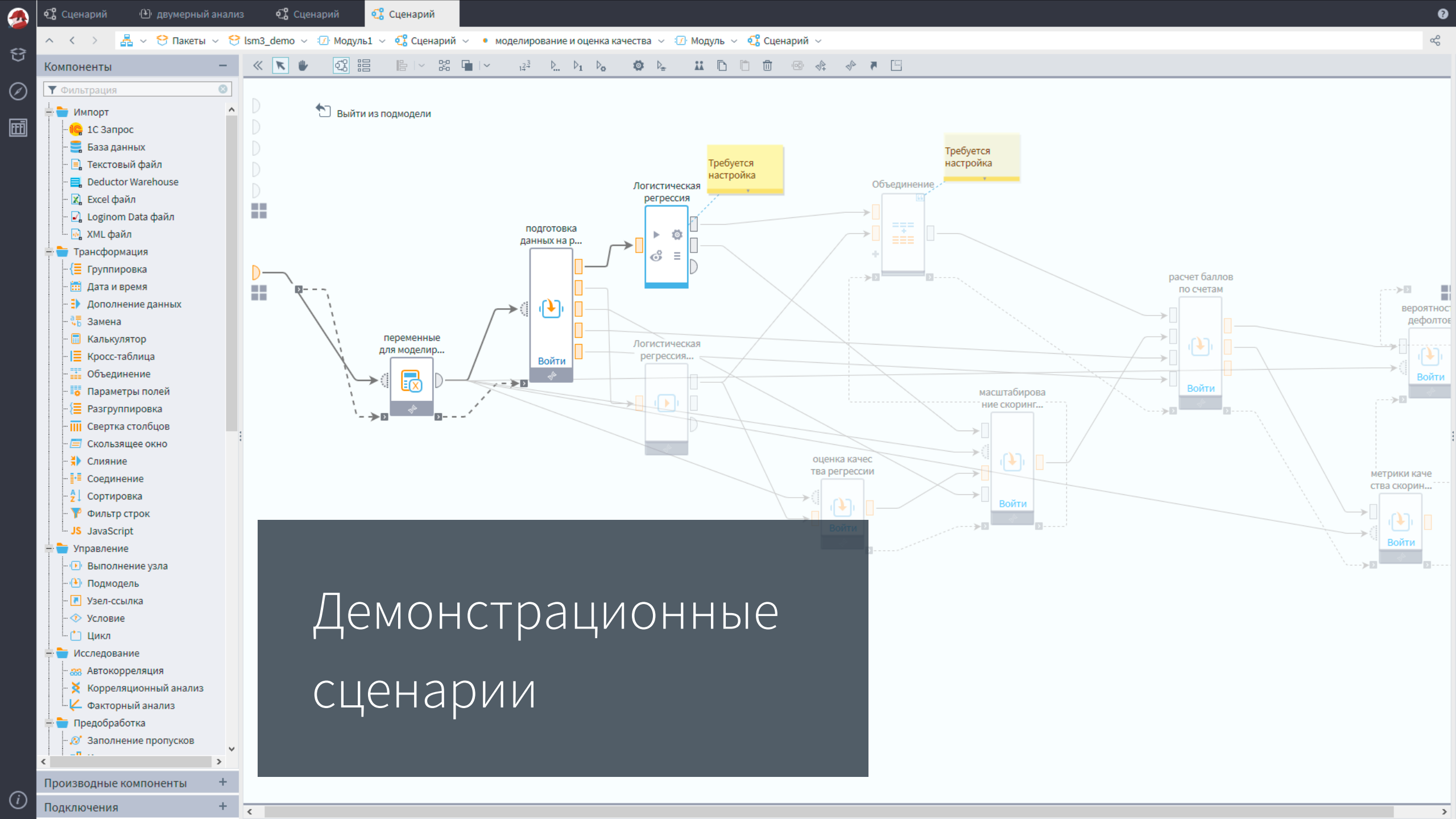
Подключение:  Активировано

- Таблицы/представления
- ACCOUNT\_ID
  - ACCOUNT\_PROP
  - ATTRIBUTE\_IN\_CARD
  - ATTRIBUTE\_TRANSFORM
  - CHARACTERISTIC
  - COARSE\_CLASSES\_CAT
  - COARSE\_CLASSES\_NUM
  - CR\_GENERATION
  - CUT\_OFF\_POINTS
  - DEBT\_QUALITY
  - FINAL\_SAMPLES
  - FINE\_ATTRIBUTE
  - INTERVAL\_ID
  - ISSUE\_DATE
  - PORTFOLIO
  - RAITINGS
  - RARE\_ACCOUNTS
  - SAMPLE
  - SAMPLES
  - SCORECARD
  - SCORECARDS
  - sqlite\_sequence
  - STATISTIC\_ACCOUNTS
  - STATISTIC\_OVERDUE
  - STATISTIC\_STATUSES
  - STATUS
  - STRATEGY
  - VERSION
  - Представления
    - COARSE\_CLASSES\_CAT\_VIEW
    - COARSE\_CLASSES\_NUM\_VIEW
    - SCORECARD\_VIEW

- Таблица/представление
- Извлекать все поля (\*)
- Фильтрация полей
- Поля
  - ab Характеристика
  - 12 Версия
  - ab Портфель
  - 12 Номер интервала
  - 9.0 Нижняя граница
  - 9.0 Верхняя граница
  - ab Конечный класс
  - 9.0 Вес
  - 9.0 Информационный индекс
  - ab Значимость

Предпросмотр...

Тематическая витрина данных



# Демонстрационные сценарии



## Scorecard Modeler

[Главная](#)
[Комплект поставки](#)
[Установка и настройка](#)
[Требования к данным](#)
[Общая схема работы](#)
[Загрузка исходных данных](#)
[Двумерный анализ](#)
[Моделирование и оценка качества](#)
[Демо-пример](#)
[Инструкция по использованию с...](#)
[Wiki](#)
[User guide](#)

## Инструкция по использованию карты

### Необходимые данные

#### Данные по заявке

Для использования **скоринговой карты** необходимо входную заявку представить в виде следующей структуры:

- **Номер заявки** – это идентификатор, номер договора, позволяющий однозначно определить значения атрибутов заявки;
- **Характеристика** – это название параметров в заявке (*пол, гражданство, возраст* и т.д.);
- **Атрибут** – значение характеристики (*мужской, Россия, 25* и т.д.).

#### Пример набора

Номер заявки	Характеристика	Атрибут
123456	Пол	мужской
123456	Гражданство	Россия
123456	Возраст	25
...	...	...

#### Конечные классы дискретные

Для использования карты необходимо использовать 3 столбца из **дискретных** конечных классов:

- **Характеристика** и **Атрибут** соответствуют понятиям, описанным в разделе **Данные по заявке**;
- **Конечный класс** – объединенные по степени влияния на выходную переменную атрибуты характеристики. Атрибуты с относительно равной силой влияния образуют один конечный класс. Это делается для понижения размерности входных данных.

#### Пример набора

Документация,  
включая инструкцию  
по встраиванию карты

# Выгоды применения:

- В 5-6 раз сократить трудозатраты аналитиков
- Обосновать корректность каждого этапа построения карты
- Изменять логику без кодирования, в том числе используя машинное обучение

[loginom.ru](http://loginom.ru)