



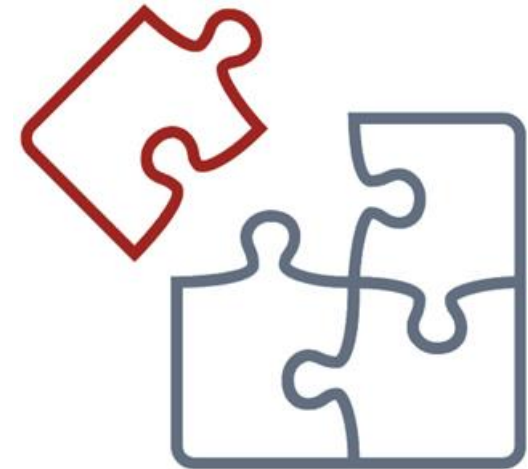
# **Библиотека Loginom Silver Kit: ПОЛЕЗНЫЕ КОМПОНЕНТЫ ДЛЯ АНАЛИТИКА**

Вебинар

6 июля 2023 года

# На вебинаре

- Познакомимся с компонентами библиотеки
- Научимся устанавливать библиотеки в Loginom
- Разберем несколько сценариев с использованием компонентов Silver Kit



# Loginom Silver Kit



Библиотека готовых компонентов Loginom.  
Облегчает рутинный труд аналитика по  
разработке сценариев обработки данных.



# Преимущества Loginom Silver Kit

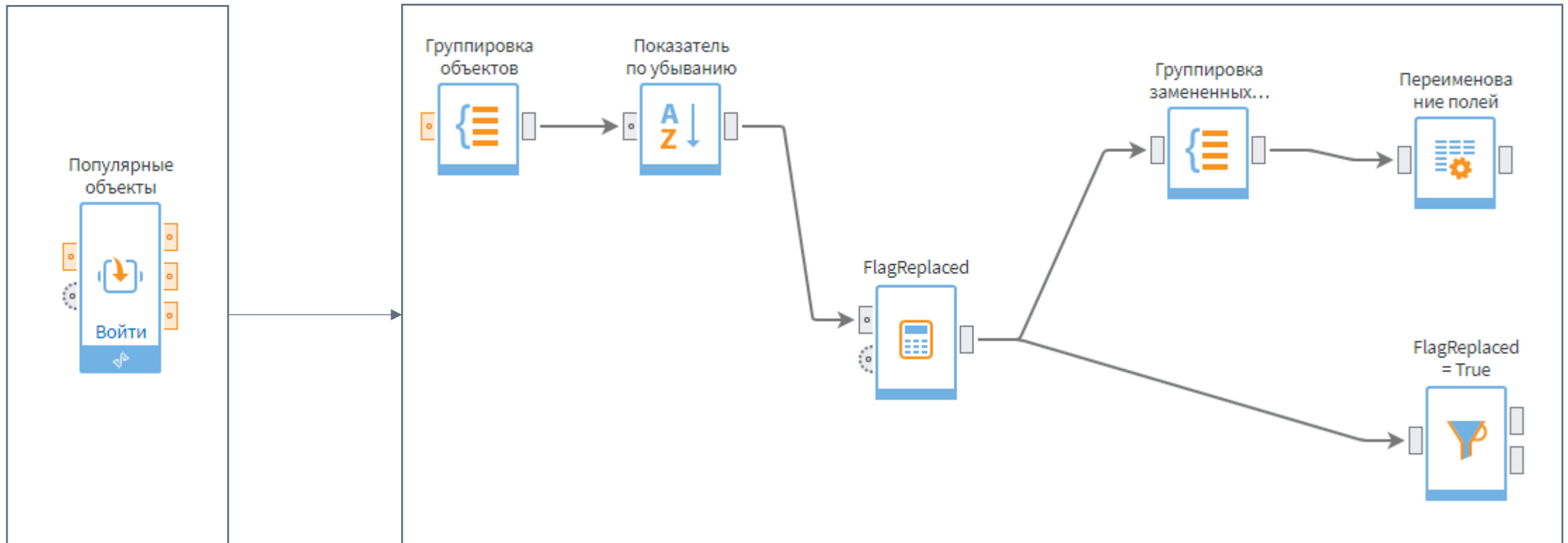


Бесплатна  
я



Открыта  
я

Каждый компонент – подмодель с реализованной логикой обработки данных





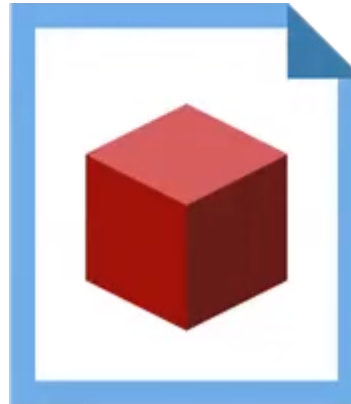
**LOGINOM**  
SKILLS

# 1. Установка библиотеки

# Что внутри?



Документаци  
я

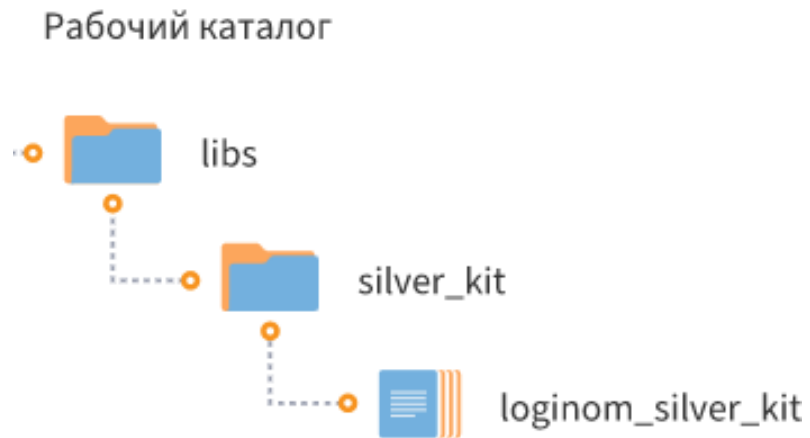


Пакет  
Loginom



Кейсы  
использования

# Расположение



Библиотеки располагаются в рабочих каталогах:

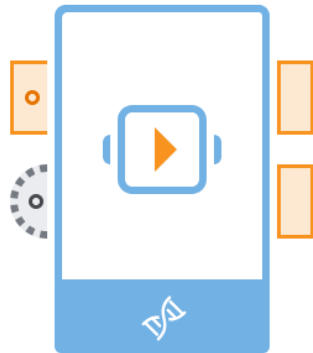
- для серверных редакций — в рабочем каталоге Loginom Server (в папке пользователя или в общей папке пользователей)
- для настольных редакций — в любой папке на локальном диске



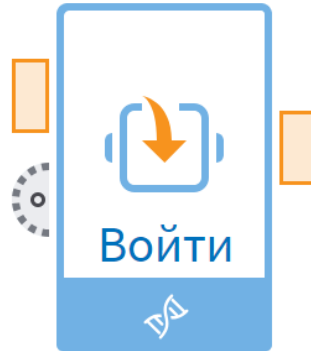
Для корректной работы необходимо установить версию Silver Kit, совместимую с вашей версией Loginom



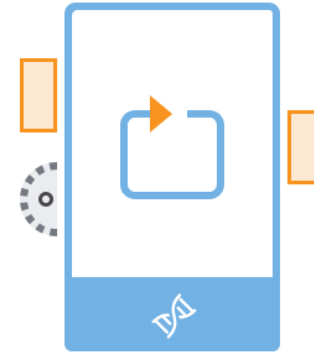
# Способы использования компонентов в сценарии



Выполнение узла



Производный узел



Цикл



**LOGINOM**  
SKILLS

## 2. Компоненты библиотеки

# Группы компонентов

## Утилиты

- рутинные операции

## ETL и подготовка данных

- предобработка данных

## Data Mining

- анализ данных

# УТИЛИТЫ

- Генератор календаря
- Генератор списка
- Генератор пустых полей
- Категория к числу
- Разность дат
- IF-объединение JS
- Список полей JS
- Фильтр столбцов JS
- Случайная категория JS

# ETL и подготовка данных

- Выделение периода
- Заполнение периодов
- Редкие значения
- Разбиение на пачки
- Разметка событий
- Разделитель строк (переменные)
- ABC-анализ
- ABC-анализ (метод касательных)
- XYZ-анализ
- Сэмплинг объектов
- Популярные объекты
- Разделитель строк

# ABC-анализ

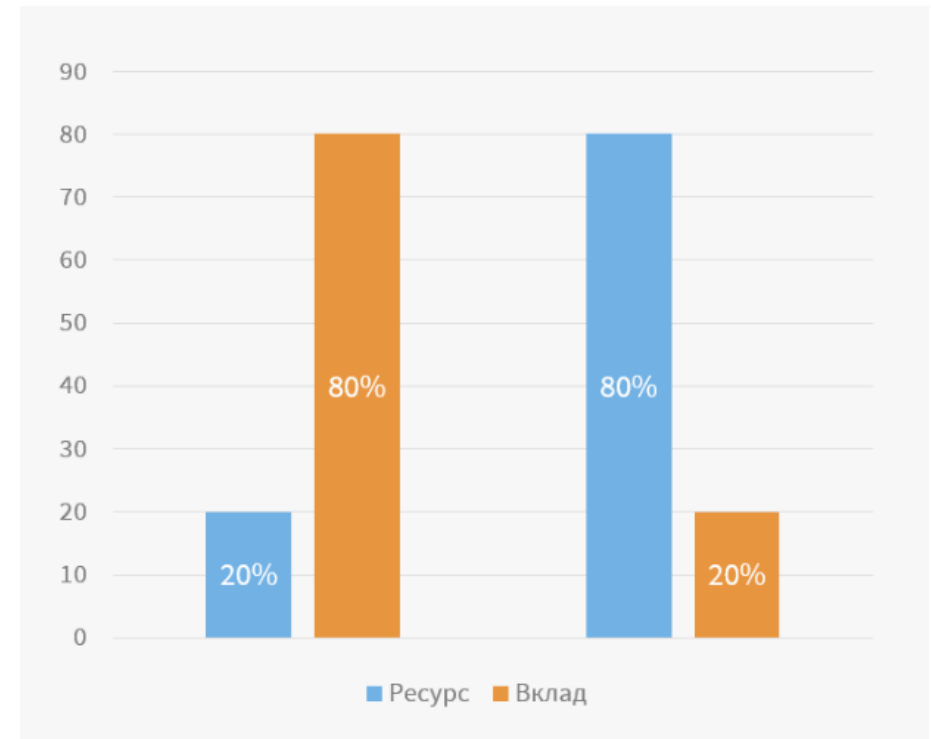
Интерпретация на примере анализа продаж

ABC подход позволяет классифицировать объекты по степени их важности. В его основе лежит принцип Парето: 20% усилий обеспечивают 80% результата.

A – наиболее ценные

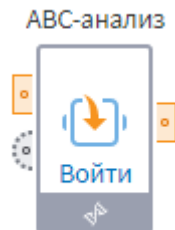
B – промежуточные

C – наименее ценные

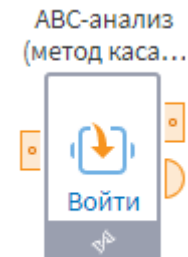


# ABC-анализ

Реализуется с помощью двух компонентов библиотеки



Границы групп определяются аналитиком самостоятельно и задаются в переменных.



Границы групп определяются автоматически методом касательных, их значения можно увидеть на выходном порте переменных.

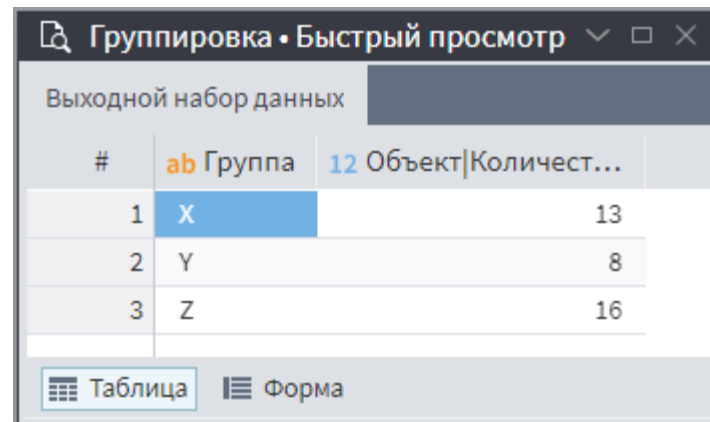
# XYZ-анализ

XYZ-анализ позволяет классифицировать ресурсы в зависимости от стабильности их потребления на основе оценки значений коэффициентов вариации.

X – постоянные

Y – нерегулярные

Z – единичные



#	ab	Группа	12 Объект Количест...
1	X		13
2	Y		8
3	Z		16

Таблица   Форма



# Интерпретация результатов

**AX**

Топовая категория. В нее попадают объекты со стабильно высокой доходностью и прогнозируемым спросом.

**AY**

Здесь также заложена хорошая прибыль, но продажи не такие предсказуемые.

**AZ**

Хороший заработок, но кривая имеет значительные отклонения. Другими словами: «то густо, то пусто».

**BX**

Средние, но стабильные доходы. Легко смоделировать дальнейшие действия.

**BY**

Нестабильное потребление при усредненном показателе маржинальности.

**BZ**

Непредсказуемый спрос в совокупности со средними значениями по выручке.

**CX**

Низкая стоимость со стабильным прогнозом. «Мал золотник, да дорог».

**CY**

Маленькая цена, и нерегулярный спрос.

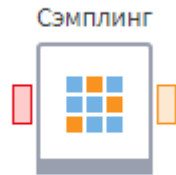
**CZ**

Ни заработка, ни постоянства. От таких объектов в 99% случаев нужно избавляться.

# СЭМПЛИНГ

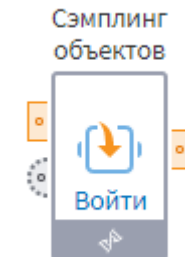
Формирование выборки путем отбора единиц наблюдения из генеральной совокупности

## Стандартный компонент Loginom



Производит отбор записей в выборку из исходного набора данных

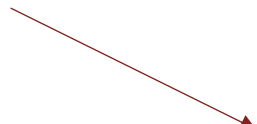
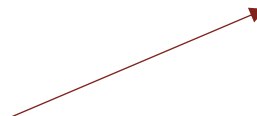
## Компонент библиотеки Silver Kit



Отбирает подмножество уникальных объектов заданного размера из множества транзакций

# Пример

ID клиента	Сумма покупки, руб
20050034	2 358
20050017	4 812
20050034	2 647
20050017	45 241
20050017	7 235
20050008	8 245
20050034	10 268
20050017	2 568
20050008	18 759
20050008	3 658
20050034	4 568
20050008	14 268



## Сэмплинг

ID клиента	Сумма покупки, руб
20050034	2 358
20050034	4 568
20050017	2 568

## Сэмплинг объектов

ID клиента
20050034
20050017
20050008

# Data Mining

- AUC
- ROC-анализ
- Динамика события
- Кластерные силуэты
- Сравнение кластеризаций
- Статистическая значимость
- Метрики классификации
- Тривиальная модель
- IV-отбор
- Сравнение метрик

# Конечные классы

Преобразует входные столбцы в последовательность интервалов, называемых конечными классами, каждому из которых присваивается определенная метка.



Применяется для

- сокращения размерности данных
- восстановления пропусков
- борьбы с выбросами и экстремальными значениями
- упрощения описания исследуемых объектов

# Конечные классы

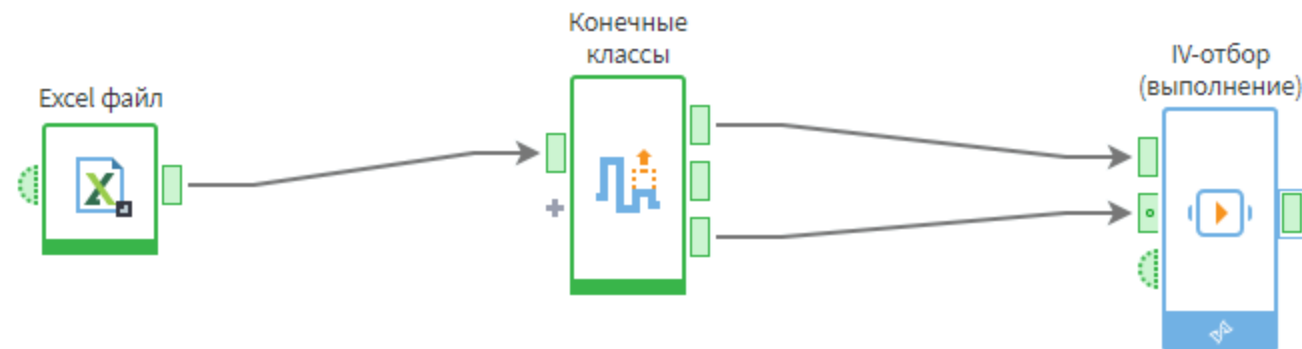
Для каждого входного столбца может быть вычислен уровень значимости для отбора переменных в модели бинарной классификации.

Значение информационного индекса	Значимость признака
$IV < 0,02$	Отсутствует
$0,02 \leq IV < 0,1$	Низкая
$0,1 \leq IV < 0,3$	Средняя
$IV \geq 0,3$	Высокая

# IV-отбор

Компонент предназначен для постобработки результатов компонента Конечные классы:

- отбор столбцов, которые содержат метку конечного класса;
- отбор значимых столбцов.



В результате каждое исходное значение признака было заменено на метку интервала квантования, а также выявлены и исключены из итоговой выборки незначимые признаки.

Excel файл • Быстрый просмотр

Набор данных

#	12 Возраст	ab Социаль...	ab Социальн...	12 Пол	12 Кол-во детей	12 Кол-во иждив...	12 Идентифика...	12 Откл...
1	49	работает	не пенсионер	1	2	1	59 910 150	0
2	32	работает	не пенсионер	1	3	3	59 910 230	0
3	52	работает	не пенсионер	1	4	0	59 910 525	0
4	39	работает	не пенсионер	1	1	1	59 910 803	0
5	30	работает	не пенсионер	0	0	0	59 911 781	0
6	29	работает	не пенсионер	0	0	0	59 911 784	0
7	35	работает	не пенсионер	0	1	1	59 911 832	0
8	41	работает	не пенсионер	1	0	0	59 912 034	0
9	53	работает	не пенсионер	0	2	0	59 912 560	0
10	43	работает	не пенсионер	1	0	0	59 912 659	0
11	54	работает	не пенсионер	1	2	0	59 912 692	0
12	26	работает	не пенсионер	0	0	0	59 913 108	1
13	62	не работает	пенсионер	0	3	0	59 913 134	1
14	45	работает	не пенсионер	1	3	1	59 913 205	0
15	38	работает	не пенсионер	1	0	1	59 913 395	0
16	41	работает	не пенсионер	1	0	0	59 913 542	1
17	65	не работает	пенсионер	1	2	0	59 913 603	0
18	29	работает	не пенсионер	1	1	1	59 913 652	0
15 223	34	работает	не пенсионер	0	0	0	59 913 736	1

Таблица | Форма

IV-отбор (выполнение) • Быстрый просмотр

Выходной набор данных

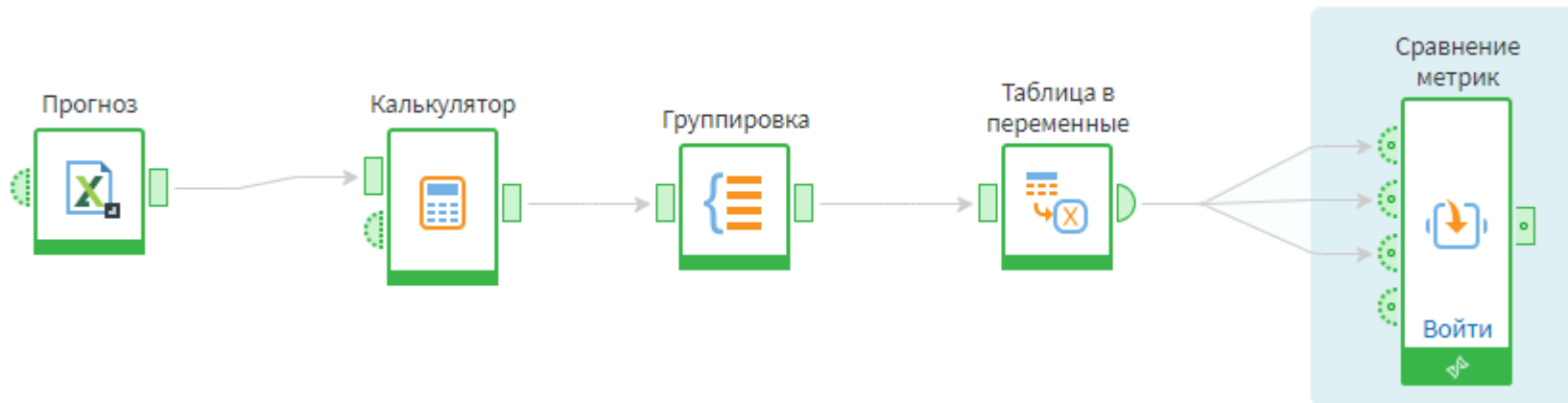
#	ab Возра...	ab Социальны...	ab Кол-во иждиве...	ab Социальн...
1	43 ≤... < 53	[не пенсионер]	1 ≤... < 2	[работает]
2	30 ≤... < 43	[не пенсионер]	от 2	[работает]
3	43 ≤... < 53	[не пенсионер]	до 1	[работает]
4	30 ≤... < 43	[не пенсионер]	1 ≤... < 2	[работает]
5	30 ≤... < 43	[не пенсионер]	до 1	[работает]
6	до 30	[не пенсионер]	до 1	[работает]
7	30 ≤... < 43	[не пенсионер]	1 ≤... < 2	[работает]
8	30 ≤... < 43	[не пенсионер]	до 1	[работает]
9	53 ≤... < 58	[не пенсионер]	до 1	[работает]
10	43 ≤... < 53	[не пенсионер]	до 1	[работает]
11	53 ≤... < 58	[не пенсионер]	до 1	[работает]
12	до 30	[не пенсионер]	до 1	[работает]
13	от 58	[пенсионер]	до 1	[не работает]
14	43 ≤... < 53	[не пенсионер]	1 ≤... < 2	[работает]
15	30 ≤... < 43	[не пенсионер]	1 ≤... < 2	[работает]
16	30 ≤... < 43	[не пенсионер]	до 1	[работает]
17	от 58	[пенсионер]	до 1	[не работает]
18	до 30	[не пенсионер]	1 ≤... < 2	[работает]
15 223	30 ≤... < 43	[не пенсионер]	до 1	[работает]

Таблица | Форма



# Сравнение метрик

Компонент позволяет сравнивать нескольких метрик. Для AUC, Gini, KS и Индекса силуэта встроены эталонные шкалы. В случае необходимости, можно самостоятельно задать значения шкалы.



# Годовой доступ в Loginom Skills

- Все курсы Loginom и статистика их прохождения
- Регулярные воркшопы в рамках Мастерской
- Все библиотеки компонентов



# Loginom Skills

Всем участникам вебинара дарим скидку **15%** на годовой доступ ко всем продуктам Loginom Skills.





Спасибо за внимание!