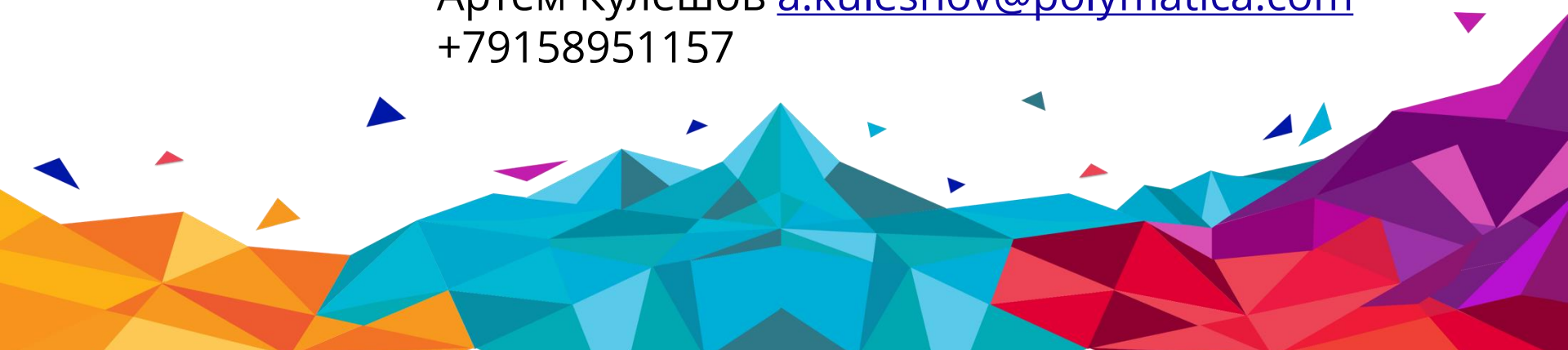


POLY//ATICA

ЛОМАЕМ СТЕРЕОТИПЫ В АНАЛИЗЕ
БОЛЬШИХ ДАННЫХ: ВЫВОДЫ В
НЕСКОЛЬКО КЛИКОВ МЫШИ

Артем Кулешов a.kuleshov@polymatica.com
+79158951157



О КОМПАНИИ ПОЛИМАТИКА

Полиматика – российский разработчик уникальной аналитической платформы Polymatica, позволяющей обрабатывать любой объём данных в режиме реального времени для поиска точек роста бизнеса

7 ЛЕТ

Компания работает на рынке России и СНГ

2010 ГОД

Началась разработка аналитической платформы Polymatica

4 ОФИСА

Собственные офисы в Москве, Лондоне, Барселоне и Цюрихе

№1 В ОТРАСЛИ

Мы являемся самым быстрорастущим поставщиком решений для анализа данных по результатам 2017 года*

5 РАЗ

Во столько раз в среднем сокращается время на подготовку аналитических срезов после внедрения нашей платформы

30+ ПАРТНЁРОВ

К партнёрской сети компании присоединяются интеграторы и дистрибьютеры по всему миру

* CNews: Аналитика 3.0 – 2018



БИЗНЕС ПОДХОД
К АНАЛИЗУ ДАННЫХ

ПОНИМАНИЕ, КАКОЙ СИЛОЙ ОБЛАДАЮТ ДАННЫЕ



- Руководителей всё больше заботит потенциал использования больших данных для улучшения бизнеса
- Они ищут оптимальные решения, способные извлечь скрытую выгоду из данных
- Они видят, как данные и цифровые технологии становятся ключевой компетенцией будущего
- Они знают, что смена парадигмы наступит очень быстро. Процесс идет – по прогнозам IDC, общий мировой объём данных достигнет 163 зеттабайт к 2025 г.



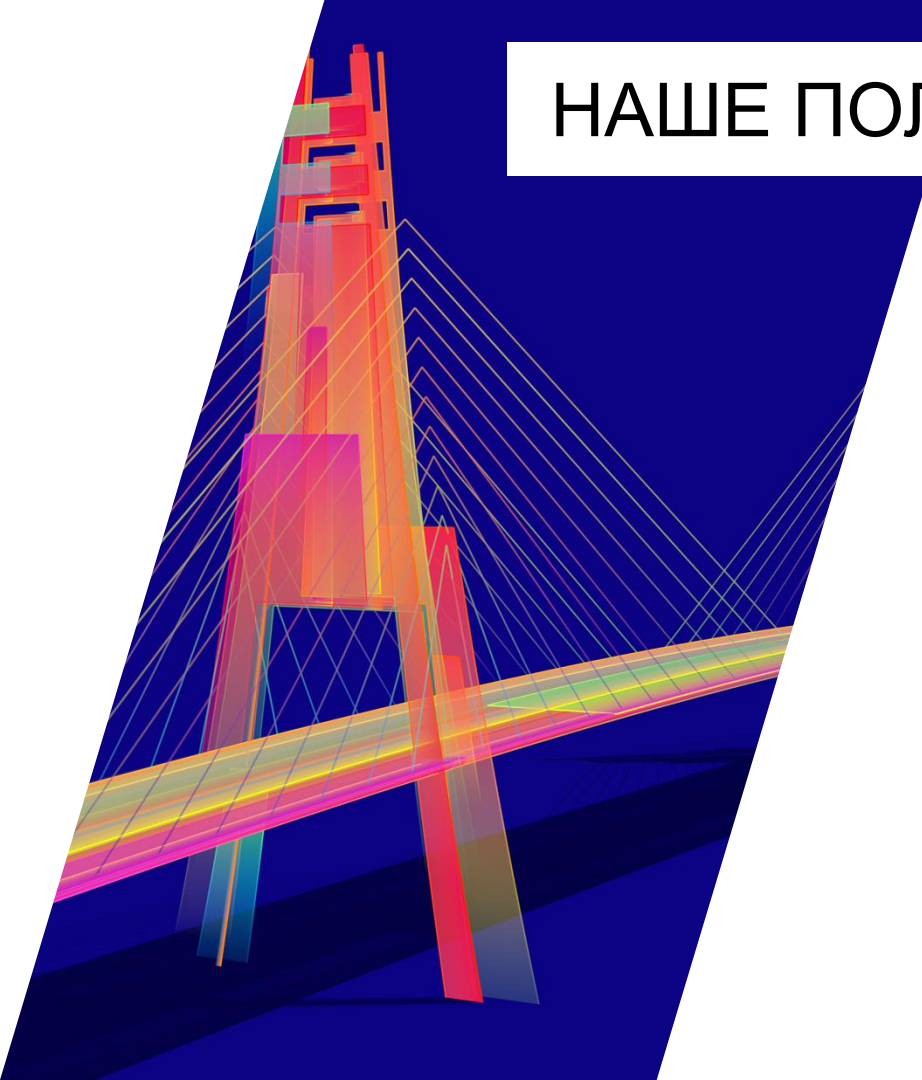


МНОГИЕ ХОТЯТ НЕ «ПРОСТО ПОНИМАТЬ», НО И ИЗВЛЕКАТЬ ВЫГОДУ

- Бизнес-пользователи недовольны тем, что не могут извлекать полезную информацию из данных быстро и легко
- Им не хватает навыков профессиональных аналитиков, поэтому они вынуждены действовать, опираясь на собственную интуицию и предположения

Согласно отчёту McKinsey, лишь **18%** компаний считают, что обладают навыками, необходимыми для извлечения и использования полезной информации, содержащейся в данных





НАШЕ ПОЛОЖЕНИЕ УНИКАЛЬНО

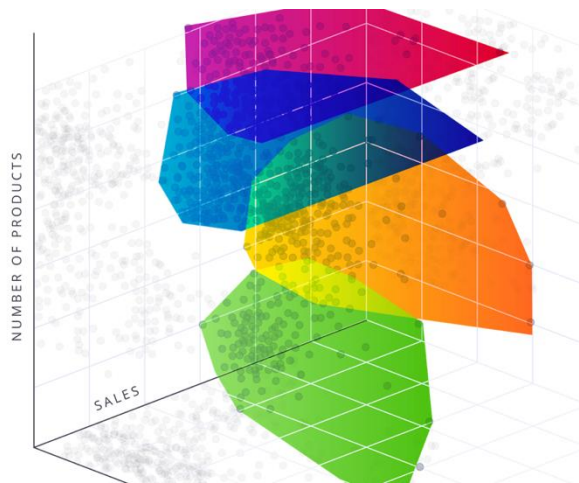
Строим мост от понимания данных к действию

- Ускорение исследования данных
 - Извлекайте пользу из данных в один клик за секунды, а не часы
- Уменьшение нагрузки на профессиональных аналитиков
 - Предоставьте аналитикам возможность сосредоточиться на будущем вместо рутинных отчетов
- Вовлечённость каждого в поиск новых идей
 - Пусть каждый попробует наметить пути совершенствования работы с данными в рамках своей компетенции
- Развитие экспертизы
 - Воспользуйтесь нашим передовым опытом в содействии росту производительности компаний, помогая поставить данные во главу угла бизнеса



БИЗНЕС ПОДХОД
К АНАЛИЗУ ДАННЫХ

ПРОРЫВНАЯ ЗАПАТЕНТОВАННАЯ ТЕХНОЛОГИЯ, ПОДДЕРЖИВАЮЩАЯ ЗАЯВЛЕННЫЙ ПОДХОД



- **Больше никаких ограничений**

Никаких ограничений по детализации, скорости и объёму данных

- **Встроенные инструменты глубинного анализа**

Специальные алгоритмы и машинное обучение для простого решения сложных задач

- **Ускорение на базе сочетания GPU + CPU**

Скорость и мощь при минимальных затратах

- **Удобное представление аналитической информации**



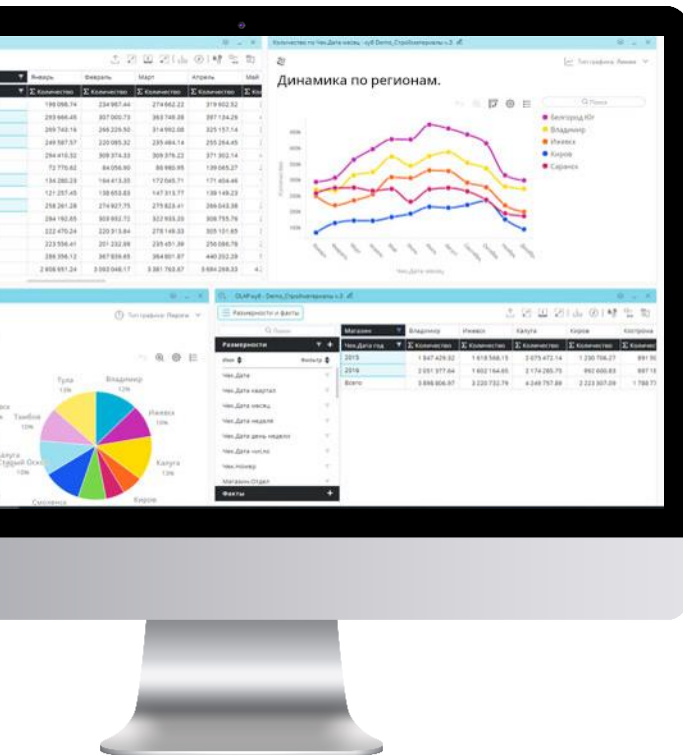
ИНТЕЛЛЕКТУАЛЬНАЯ КЛАСТЕРИЗАЦИЯ



- Распределение объектов по группам по одному или нескольким фактам.
- Для запуска модуля не требуется выбора модели, дополнительных настроек, программирования и прочих сложных операций.
- Распределение кластера выстраивается автоматически



ДЕЛАЙТЕ ПРОГНОЗЫ НА БУДУЩЕЕ



- Множество моделей прогнозирования запускаются одновременно
ARIMA, ARIMA-T, фильтр Кальмана, нелинейная регрессия, логистическая регрессия
- Независимые испытания
Каждая модель тестируется независимо на последних 10% наборов данных

НАХОДИТЕ СКРЫТЫЕ ПАТТЕРНЫ ПОВЕДЕНИЯ



Поиск ассоциативных связей между объектами основан на классическом алгоритме Apriori

В системе используется модифицированный вариант алгоритма, оптимизированный с точки зрения потребления оперативной памяти и времени работы. За счет этого операция поиска ассоциативных связей даже на миллионах объектах выполняется за минуты

	A	B
№1	1	0
№2	0	1
№3	0	1
№4	1	1

A → B
Популярность **25%**
Достоверность **50%**

B → A
Популярность **25%**
Достоверность **33%**



ВЫЯВЛЕНИЕ ЗАВИСИМОСТЕЙ КЛИКОМ



Модуль **Ассоциативные правила** позволяет проводить поиск ассоциативных связей между объектами на всем массиве данных, определять популярности и достоверности совместного возникновения событий.

- Поиск популярных связей между продуктами / услугами
- Определение поведения клиента внутри категорий
- Поиск подходящих категорий предложений для кросс-продаж
- Определение доли сопутствующих предложений при проведении акций
- Распределение категорий предложений согласно выявленному поведению и предпочтениям клиентов



КЛИЕНТЫ ГОСУДАРСТВЕННОГО СЕКТОРА



Полиматика - Стандарт
Аналитического Слоя.
Проекты: Аналитическая
Подсистема ЗАГС,
Отчетность 2.0,
Маркировки



Аналитическая
подсистема актуарных
расчетов,
аналитические
подсистемы для ФРИ,
ЕГИССО



Ситуационно-аналитический центр
Минэнерго России

Интерактивная карта с
общей аналитической
информацией по
состоявшимся 18 марта
выборам Президента РФ



ФАНО
России
ФЕДЕРАЛЬНОЕ
АГЕНТСТВО
НАУЧНЫХ
ОРГАНИЗАЦИЙ

Аналитическая платформа
для оценки
эффективности работы
научных организаций
ФАНО России



БИЗНЕС ПОДХОД
К АНАЛИЗУ ДАННЫХ

ДИТ: СКРЫТЫЕ ВЗАИМОСВЯЗИ ОСНОВНЫХ ПРОЦЕССОВ ГОРОДСКИХ ОРГАНИЗАЦИЙ

ЗДРАВООХРАНЕНИЕ



ОБРАЗОВАНИЕ



ГРАДОСТРОИТЕЛЬСТВО



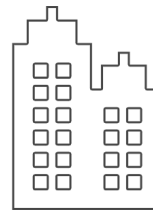
ТРАНСПОРТНАЯ
ИНФРАСТРУКТУРА



СВЯЗЬ
И КОММУНИКАЦИИ



БИЗНЕС



БИЗНЕС ПОДХОД
К АНАЛИЗУ ДАННЫХ

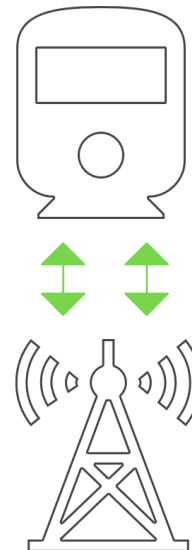
ТРАНСПОРТ И КОММУНИКАЦИИ

СЛУЖБЫ ГОРОДСКОГО УПРАВЛЕНИЯ

АНАЛИЗ ПОССАЖИРОПОТОКА НА ОСНОВАНИИ ДАННЫХ СОТОВЫХ ОПЕРАТОРОВ

Результаты проекта

- Составлен рейтинг станции по уровню нагрузки
- Определено время наивысшей загруженности транспорта
- Создана почасовая тепловая карта областей, формирующих большие пассажиропотоки
- Выполнена визуализация пассажиропотока в разрезе станций и времени
- Выявлены особенности некоторых станций: пассажиров входит больше чем выходит и наоборот.



БИЗНЕС ПОДХОД
К АНАЛИЗУ ДАННЫХ

ПРИМЕР ВЫВОДА

- На топ-10 станций приходится 13,2% пассажиропотока.
- Лидером по пассажиропотоку является станция «Текстильщики».
- Возможные причины: станцией пользуются жители близлежащих к Москве городов Подольска и Щербинки, пересаживаясь на ней с железной дороги на метро

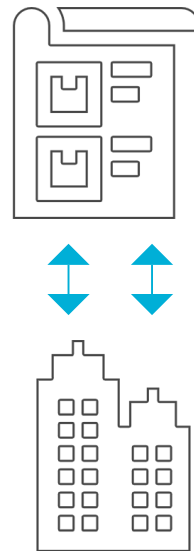


ДАННЫЕ ЖКХ И ПЛАТЕЖИ

КОММУНАЛЬНЫЕ ПЛАТЕЖИ В БЮДЖЕТ ГОРОДА

АНАЛИЗ ЗАДОЛЖЕННОСТИ ЖИТЕЛЕЙ И УПРАВЛЯЮЩИХ КОМПАНИЙ ПЕРЕД ГОРОДОМ

- Анализ задолженности в разрезе дома, района, округа.
- Данные в разрезе УК и параметров постройки
- Анализ эффективности сбора задолженности: KPI колл-центра
- Сопоставление расхода энергии (данные со счетчиков – температурный график) и данных по оплате (договор с поставщиками услуг)



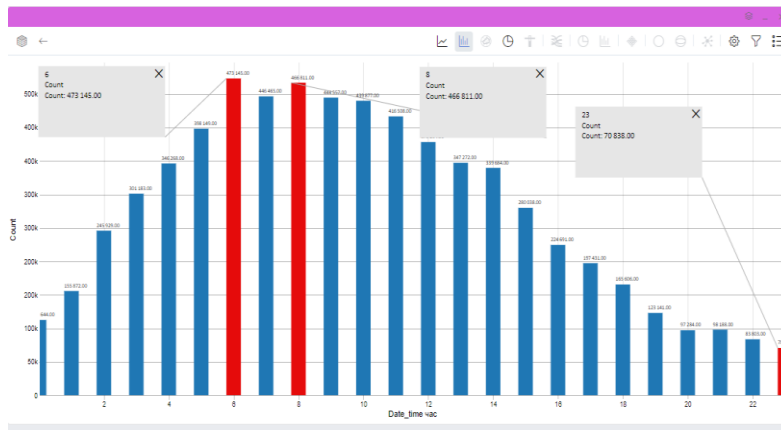
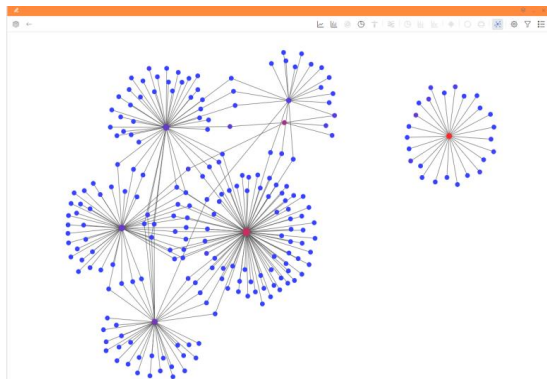
БИЗНЕС ПОДХОД
К АНАЛИЗУ ДАННЫХ

ДАННЫЕ ИСПОЛЬЗОВАНИЯ WEB-РЕСУРСОВ

АНАЛИЗ ИСПОЛЬЗОВАНИЯ СЕРВИСОВ

ИЗУЧЕНИЕ ПОВЕДЕНИЯ ПОЛЬЗОВАТЕЛЕЙ НА ИНФОРМАЦИОННЫХ РЕСУРСАХ

- Анализ переходов между сайтами по типам устройств, протоколам и др. параметрам.
- Динамика использования сайтов и сервисов.



БИЗНЕС ПОДХОД
К АНАЛИЗУ ДАННЫХ

КЕЙС: ПРИМЕНЕНИЕ АНАЛИТИЧЕСКОЙ ПЛАТФОРМЫ ПРИ ПРОГНОЗИРОВАНИИ ОТТОКА АБОНЕНТОВ ТЕЛЕКОМ-КОМПАНИЙ



ИСХОДНЫЕ ДАННЫЕ

БЫЛИ ПОЛУЧЕНЫ В EXCEL – ПО АБОНЕНТАМ И ПО БЛОКИРОВКАМ

	A	B	C
1	ID абонента	Дата начала блокировки	Дата окончания блокировки
2	1950	25-июл-2016 11:00:00	13-авг-2016 08:34:04
3	1950	24-авг-2016 03:23:38	17-ноя-2016 19:37:05
4	1950	03-дек-2016 03:31:42	19-январь-2017 18:30:06
5	1950	04-фев-2017 03:39:13	28-фев-2017 19:09:34
6	1950	16-мар-2017 03:25:40	20-мар-2017 11:52:04
7	1950	05-апр-2017 03:23:34	20-апр-2017 17:08:05
8	1950	06-май-2017 03:01:56	
9	1950	15-май-2017 02:21:04	

	A	B	C	D	E	H	I	J	K	L	M	N
11-я	PERIOD	ID абонента	ЗУЭС	РУЭС	Дата подключения	Количество	Наличие	Наличие	Наличие	Сумма пл	Количество	Начислен
2	201706	1950	Набер	Актан	19.07.16 1	1	1	1	0	91,51	2	85,37
3	201707	1950	Набер	Актан	19.07.16 1	1	1	1	0	92,01	1	83,42
4	201708	1950	Набер	Актан	19.07.16 1	1	1	1	0	0	0	14,79
5	201709	1950	Набер	Актан	19.07.16 1	1	1	1	0	0	0	0
6	201710	1950	Набер	Актан	19.07.16 1	1	1	1	0	0	0	0
7	201711	1950	Набер	Актан	19.07.16 1	1	1	1	0	0	0	0
8	201712	1950	Набер	Актан	19.07.16 1	1	1	1	0	0	0	0
9	201801	1950	Набер	Актан	19.07.16 1	1	1	1	0	0	0	0
10	201802	1950	Набер	Актан	19.07.16 1	1	1	1	0	0	0	0
11	201803	1950	Набер	Актан	19.07.16 1	1	1	1	0	0	0	0
12	201804	1950	Набер	Актан	19.07.16 1	1	1	1	0	0	0	0
13	201805	1950	Набер	Актан	19.07.16 1	1	1	1	0	0	0	0
14	201706	1962	Арски	Рыбн	19.07.16 1	1	1	1	0	0	0	0
15	201707	1962	Арски	Рыбн	19.07.16 1	1	1	1	0	100	1	0

ОБЩАЯ ИНФОРМАЦИЯ

цели

- Прогнозирование оттока
- Формирование списков для передачи в дирекцию маркетинга

подход

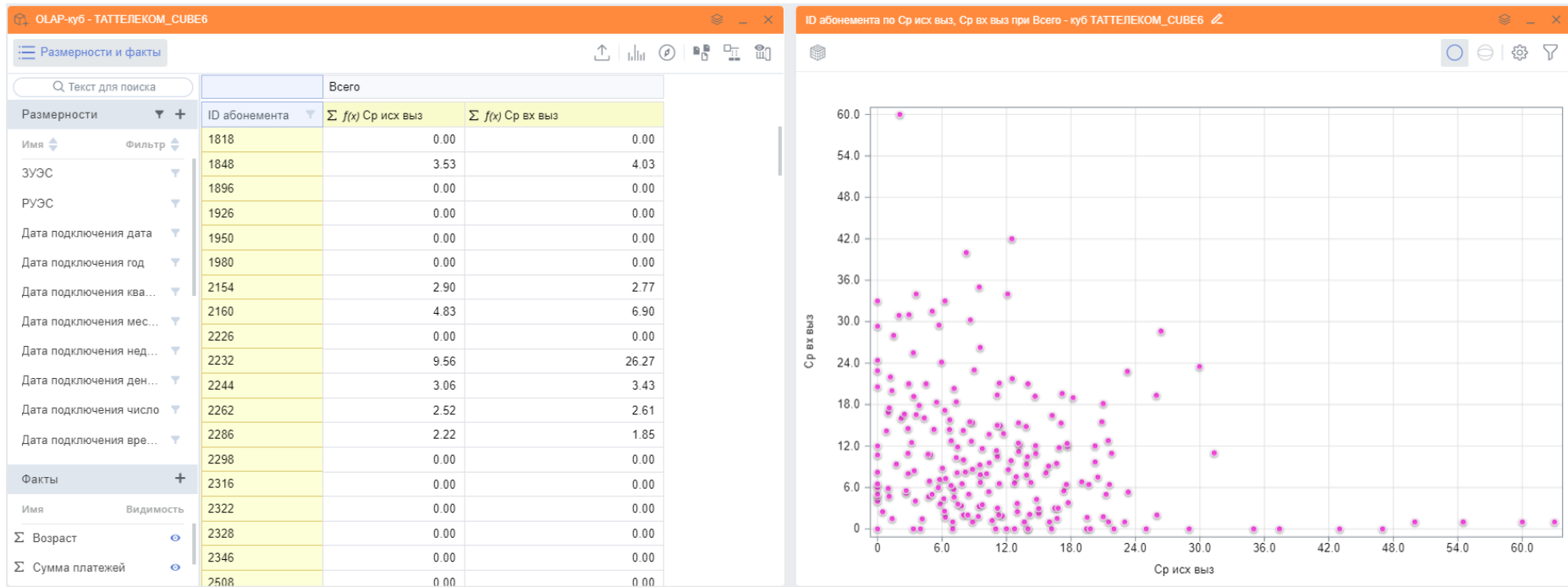
- Очистка данных
- Кластеризация и анализ данных

результат

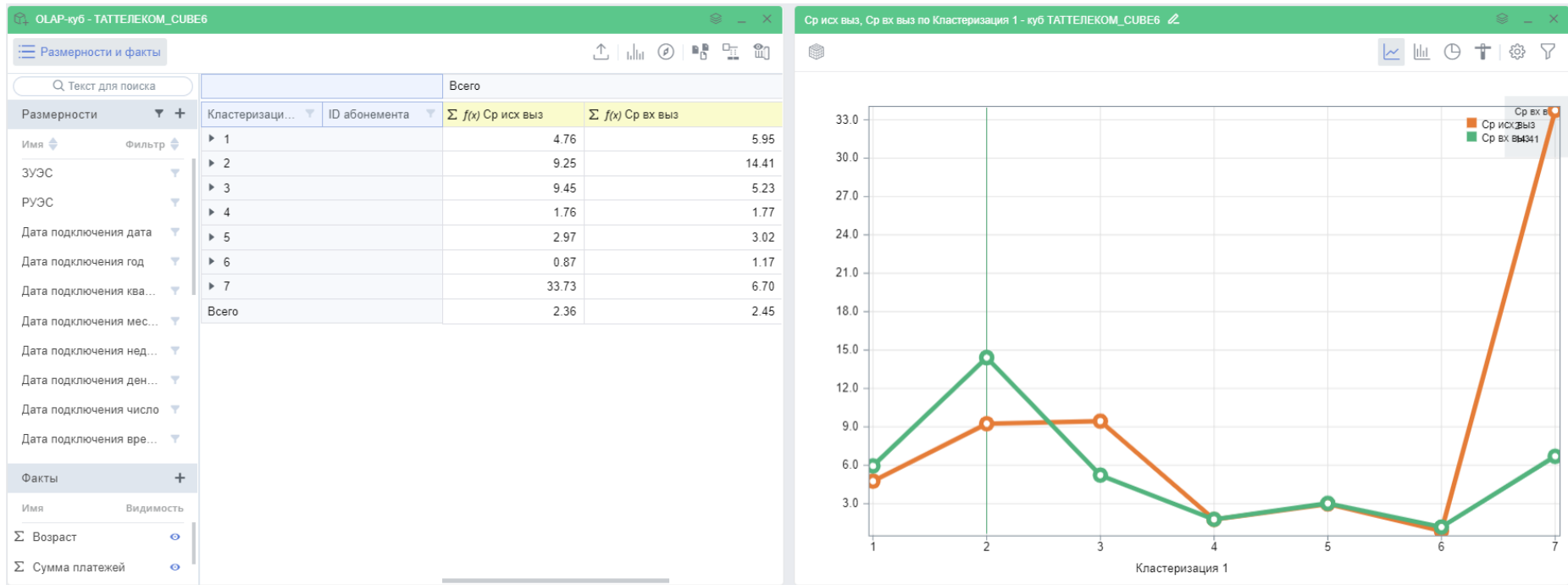
- Выделение групп абонентов в зоне риска
- Excel-файл со списком абонентов



POLYMATICA: РАСПРЕДЕЛЕНИЕ АБОНЕНТОВ ПО СРЕДНЕЙ ПРОДОЛЖИТЕЛЬНОСТИ ВЫЗОВОВ



POLYMATICA: СИСТЕМА РАЗБИЛА ДАННЫЕ НА 7 КЛАСТЕРОВ



POLYMATICA: АНАЛИЗ УЧАСТНИКОВ ПРОГРАММЫ ЛОЯЛЬНОСТИ В РАЗРЕЗЕ ЗАБЛОКИРОВАННЫХ АБОНЕНТОВ

OLAP-куб - ТАТТЕЛЕКОМ_CUBE6

Размерности и факты

Текст для поиска

Участие в ПЛ	0	1	Всего
Заблокирован	ID абонента	Кластеризация 1	// Возраст
Да			9 165.00
Нет			20 732.00
Всего			2.00

Имя Фильтр

ЗУЭС

РУЭС

Дата подключения дата

Дата подключения год

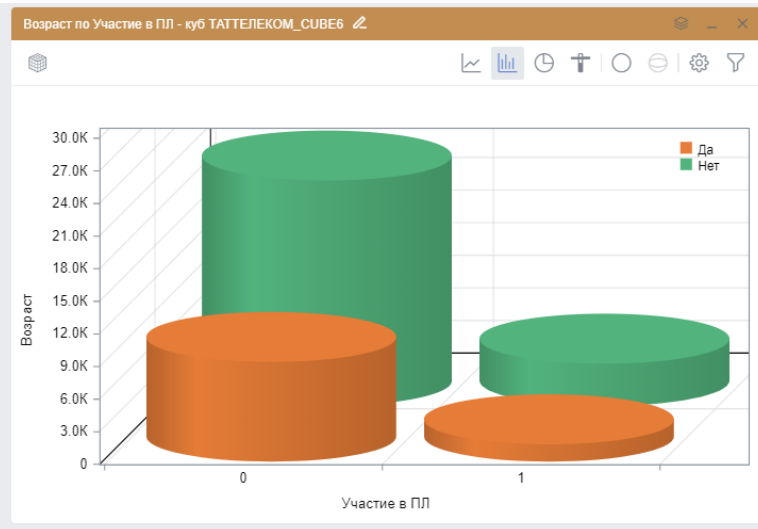
Дата подключения ква...

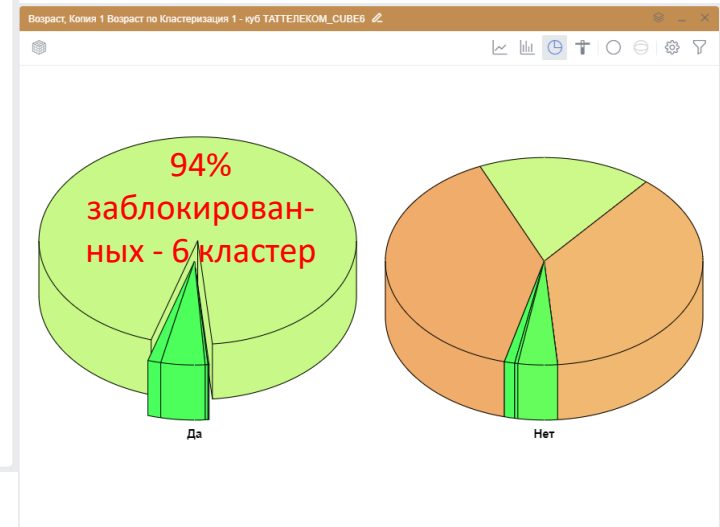
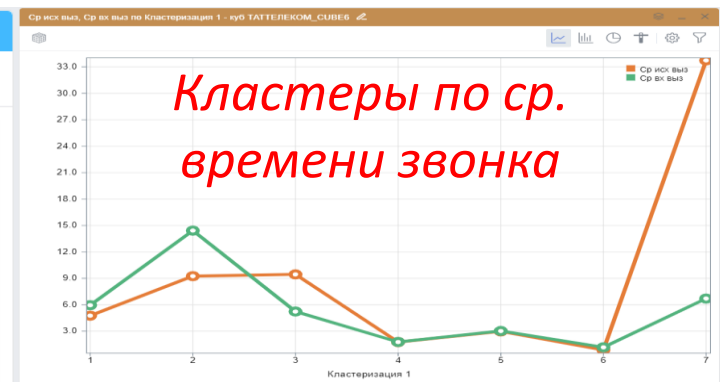
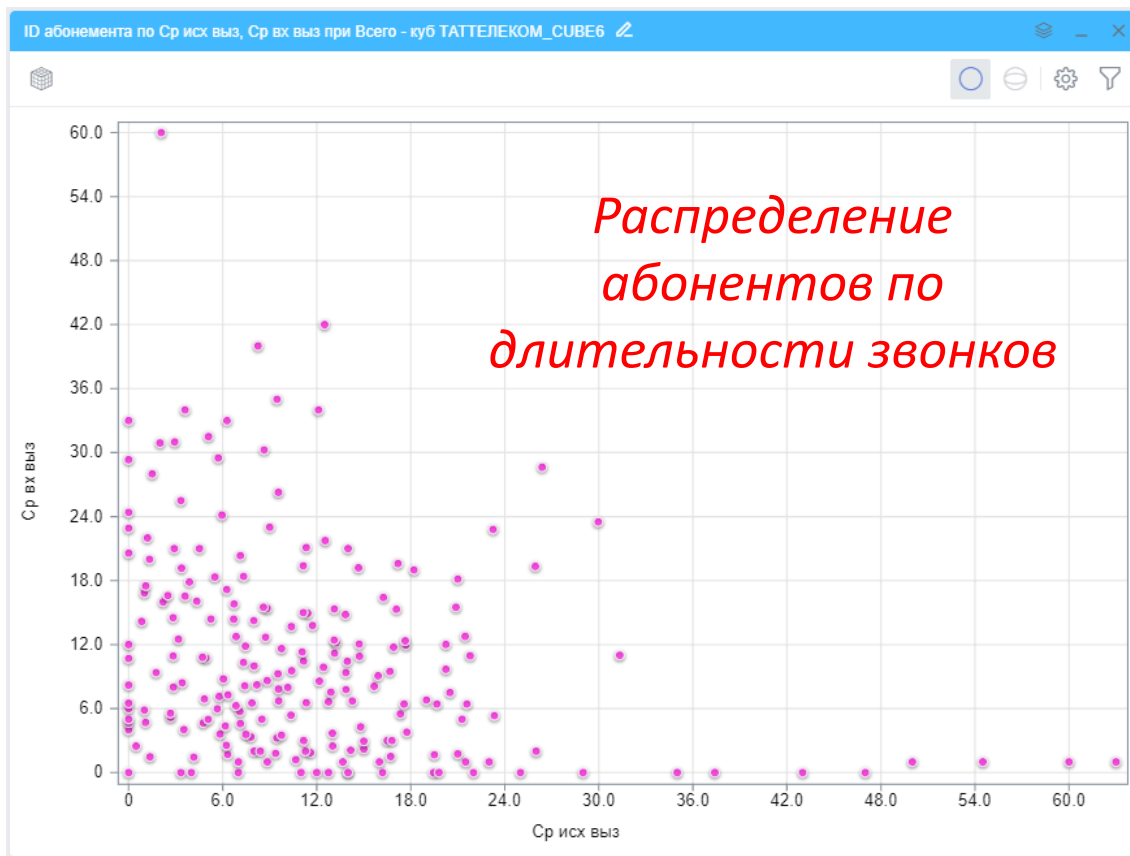
Дата подключения мес...

Дата подключения нед...

Факты

Имя Видимость





РЕЗУЛЬТАТЫ ПЕРВИЧНОГО АНАЛИЗА ДАННЫХ

- **Количество финансовых блокировок** не влияет на статус. Есть активные пользователи с 180-304 финансовыми блокировками. Есть заблокированные с 0-1 финансовых блокировок.
- **Количество смен ТП** не влияет на отток – наибольшее количество смен ТП (7, 8) у тех, кто сейчас Активен.
- **Динамика интернет-трафика** статистически не влияет на блокировки.
- Провели автоматическую **кластеризацию** по **среднему времени исходящего и входящего вызова** - результат - 7 кластеров.
- **94% заблокированных – в одном кластере**, куда вошли абоненты с самой короткой длительностью звонков.
- Статистически значимыми параметрами при прогнозировании оттока являются: **средняя продолжительность звонка, возраст абонента, участие в программах лояльности, тарифный план.**



ПРИМЕР ПРИМЕНЕНИЯ POLYMATICA: ПРОДАЖА СТРАХОВКИ «ЗАЩИТА+»

ВЫБОР ПУТИ КОММУНИКАЦИИ И ЦЕЛЕВОЙ АУДИТОРИИ ДЛЯ МАКСИМИЗАЦИИ КОНВЕРСИИ ПРЕДЛОЖЕНИЯ

Конверсия предложения

35%

(вместо 14% традиционными методами)

Окупаемость коммуникации

2 месяца

при жизни продукта 6 месяцев
– 4 месяца чистой прибыли!

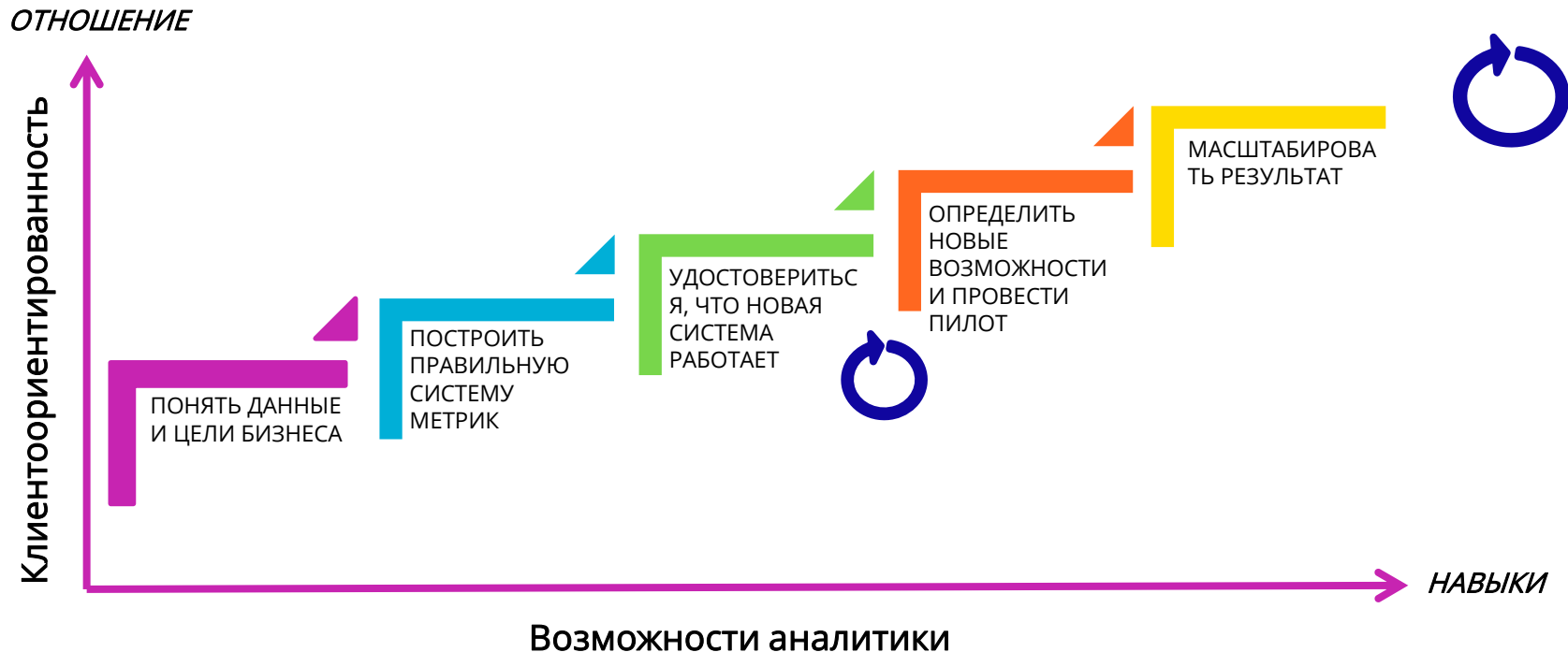
- Кластеризация по сочетаниям «портрет – тип карты – расходы – поведение»
- Апробация 3 (трех) способов коммуникации и 8 возможных сценариев для каждого кластера
- Выработка решения за 8 часов

**ПОДРАЗДЕЛЕНИЕ БАНКА DIGITAL В НЕСКОЛЬКО РАЗ
ПРИБЫЛЬНЕЕ ДРУГИХ ПОДРАЗДЕЛЕНИЙ БАНКА**



БИЗНЕС ПОДХОД
К АНАЛИЗУ ДАННЫХ

НАПРАВЛЯЕМ ВАС И ВАШ БИЗНЕС ПО ПРАВИЛЬНОМУ ПУТИ РАБОТЫ С ДАННЫМИ



БИЗНЕС ПОДХОД
К АНАЛИЗУ ДАННЫХ