

Практика «Инжиниринг и анализ данных» для студентов в сбытовом блоке группы «Газпром нефть»

Докладчик:

Михаил Казанцев

Руководитель Группы разработки
и сопровождения BI-системы ДРП

03.10.2018



Реализация нефтепродуктов оптом
Хранение нефтепродуктов
Управление **>50** собств. нефтебазами

36 регионов
присутствия в РФ +
4 страны СНГ



Розничная реализация нефтепродуктов
Продажи корпоративным клиентам
Управление АЗС \ АСК **>1500** объектов

Сотрудников в блоке
– порядка **23 тыс.**
человек



Реализация сопутствующих товаров,
оказание услуг моек, СТО
>800 магазинов

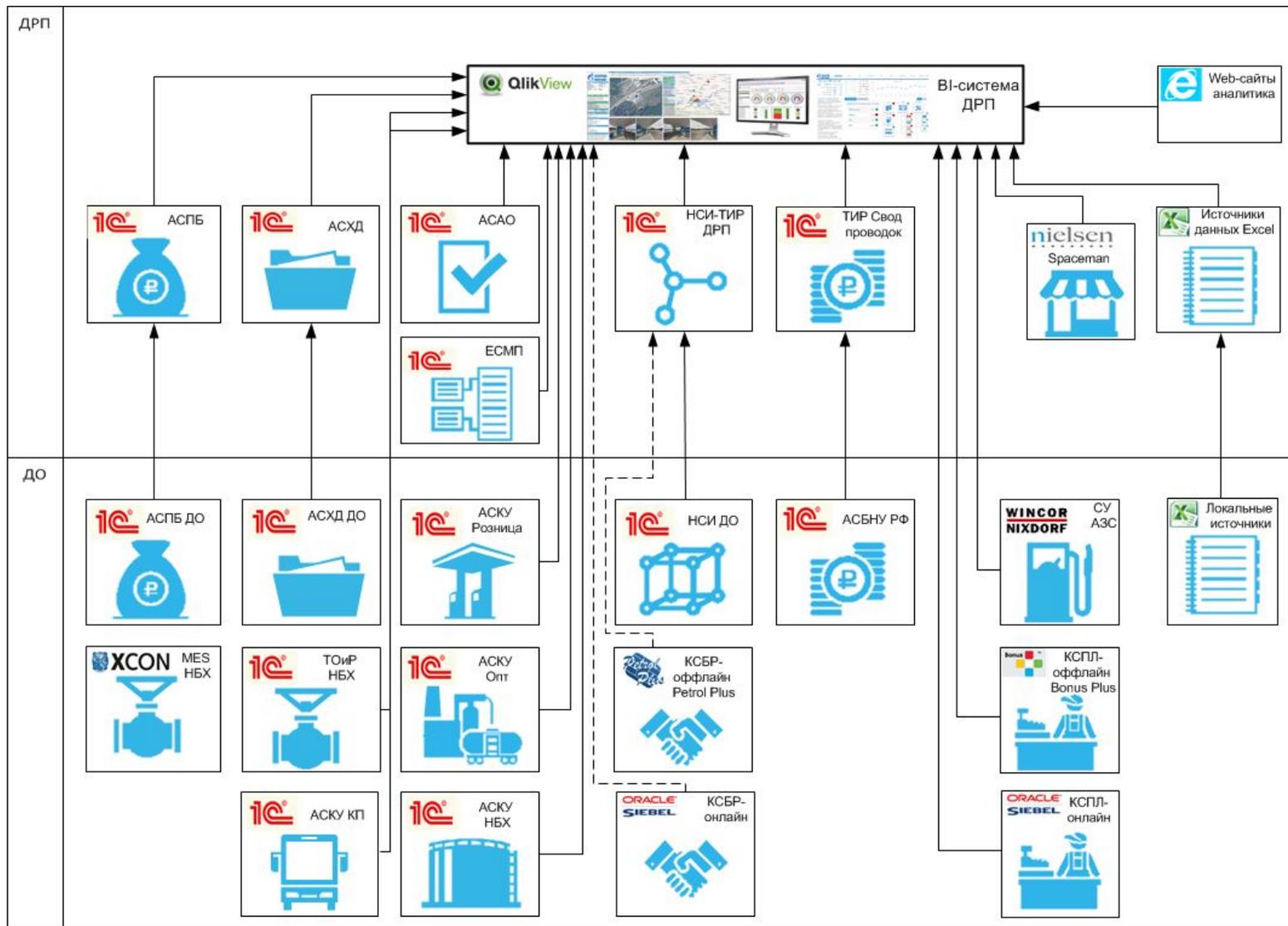
Клиентов-
участников
бонусной
программы > 9
млн. человек



Управление автотранспортом:
бензовозы/газовозы
Услуги метрологии

Кто из вас знает что такое VI?

Объединяет данные из **>140** источников и является зеркалом ключевых данных



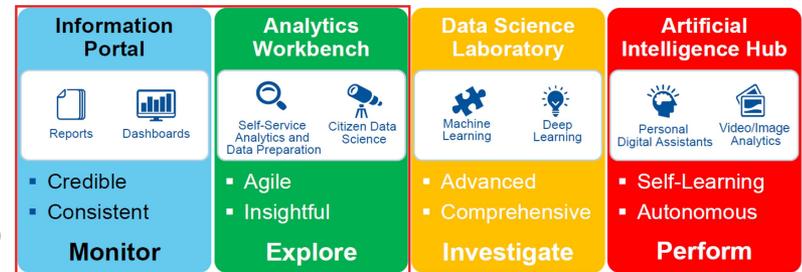
Начало 2000-х годов – персональные компьютеры и массовые инструменты аналитики

1993 год: Э.Кодд: недостаток реляционной модели БД - невозможность «объединять, просматривать и анализировать данные с точки зрения множественности измерений, т.е. самым понятным для аналитиков способом», статья «OLAP для пользователей-аналитиков: каким он должен быть» – 12 правил оперативной аналитической обработки данных

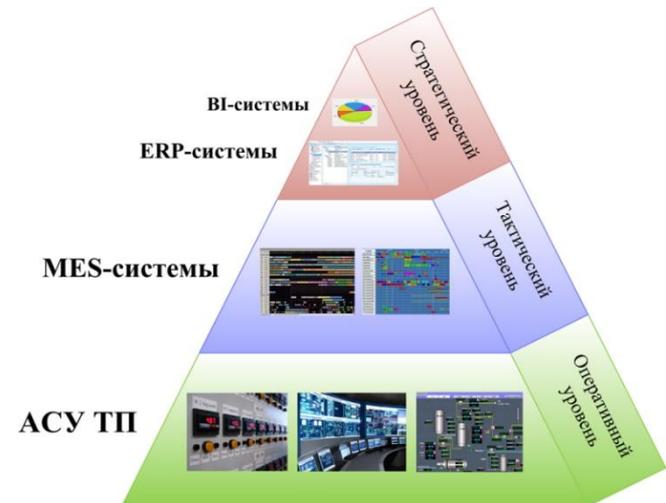
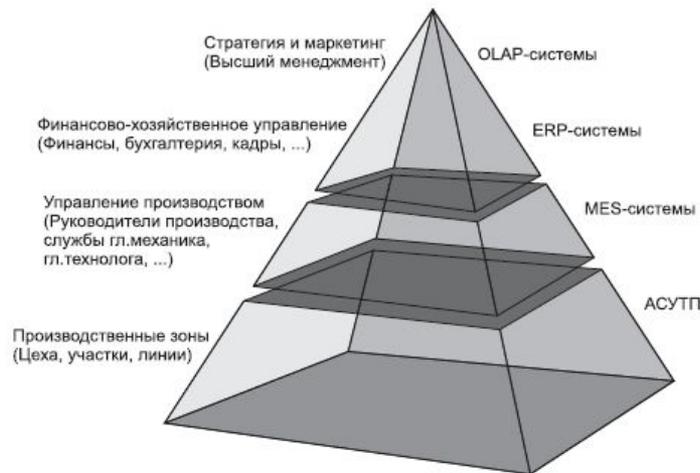
Конец 1990-х годов – появление промышленных OLAP-систем

Начало 2000-х годов – массовое появление систем Business Intelligence 2.0 (BI), включающих:

- Бизнес-отчётность
- Информационный поиск
- Аналитическую обработку в реальном времени (OLAP),
- Инструменты предупреждения об отклонениях от ожиданий
- Бизнес-аналитику (статистику, прогнозирование и оптимизацию)



BI-системы являются развитием идей OLAP, но лишены недостатков OLAP-систем



Business Intelligence (BI): многоэтапная логика принятия решений

Пример: концепция DAR (Dashboard, Analytics, Reports) компании Qlik

Детали



* Выберите ФП1 для результата только по сравнению

Финплан | Территории | Вид | Тип | Группа ОСАГО | Группа ОСАГО ЮЛ

млн. руб. | Каналы | АСП | ДВС | «Баз. группа» | «Баз. группа» ЮЛ

ДТО (стоимость) | ОПС | Рентабельность | Рентабельность

Показатель | Сортировка: Факт

Линии бизнеса



* Выберите ФП1 для результата только по сравнению

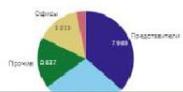
Финплан | Территории | Каналы

млн. руб. | ЛБГ | Бизнес | Бизнес1 | Бизнес2 | Бизнес3

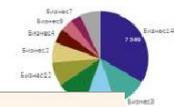
Регионы | Прочие

Факт	Доля от Итого, %
78.0%	100.0%
70.0%	94.0%
100.0%	12.0%
97.0%	10.0%
111.0%	10.0%
97.0%	8.0%
105.0%	7.0%
95.0%	8.0%
33.0%	4.0%
126.0%	4.0%
60.0%	2.0%
72.0%	2.0%
100.0%	1.0%
20.0%	1.0%
110.0%	0.0%
2.0%	0.0%

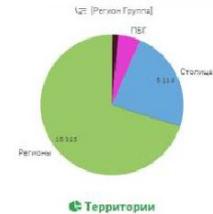
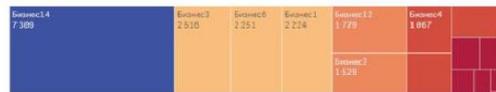
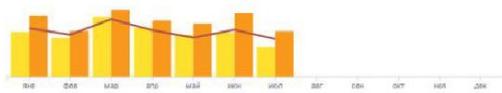
Каналы



Виды



Основные показатели





- Текущие выборы**
- Объект: АЗС 006 Кр кр
 - Управления: Горячключевской район М4
 - Наименование: Газпромнефть-Юг
 - Подразделение: Газпромнефть-Юг
 - в Отделении: Вывучка
 - Показатель: Вывучка
 - ТипКарты: Гибрид

Назад Очистить Вперед

Отобрано количество транзакций: 771 514

- Отделение**
- Газпромнефть-Юг
 - Газпромнефть-Азия
 - Газпромнефть-Казахстан
 - Газпромнефть-Таджикистан
 - Отделение Екатеринбург

- Организации**
- Юг Азия Казахстан
 - Новосибирск Северо-Запад Таджикистан

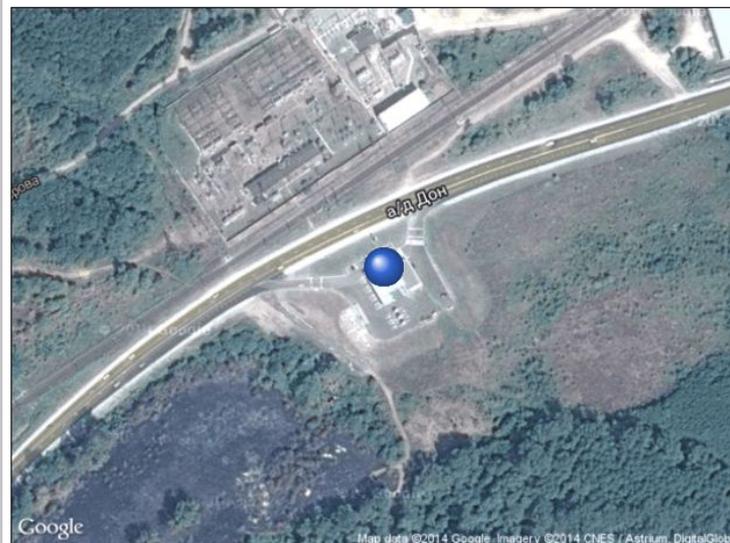
- Каналы продаж:**
- Корпоративные клиенты
 - Розница
 - Сопутств ующий Бизнес

- Номенклатурная группа**
- Автотовары
 - Автохимия и автокосметика
 - Алкогольные напитки
 - Безалкогольные напитки
 - Бензин 92

- Номенклатура ЕКТУ**
- CD/MP3
 - DVD
 - Автотоксосуфры
 - Автотовары в багажник
 - Автотовары для детей

- Номенклатура**
- Gardex Baby Детский бальзам-стик после ...
 - Gardex Baby Детский спрей от комаров 50 мл
 - Gardex Extreme Аэрозоль от клещей 150 мл
 - SF-0518 Скребок водосгон силиконовый
 - Aln GE 1056 PY21 12B 21Вт BAU15s

Год	Квартал		Месяц			День месяца											Неделя							День недели																
2013	1 кв	2 кв	январь	февраль	март	апрель	май	июнь	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	1	2	3	4	5	6	7	8	9	10	11	12	Пн	Вт	Ср	Чт
2014	3 кв	4 кв	июль	август	сентябрь	октябрь	ноябрь	декабрь	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	13	14	15	16	17	18	19	20	21	22	23	24	Пт	Сб	Вс		



Тип карты

- Гибрид
- Карта
- Спутник

Показатель на карте

- Вывучка
- Объем реализации
- Среднесуточная вывучка
- Среднесуточная реализация

Регион Группа

- Киргизская Республика
- Республика Казахстан
- Республика Таджикистан
- Российская Федерация

Регион Наименование

- Краснодарский край
- Алматинская область
- Алтайский край
- Баткенская область
- Вологодская область

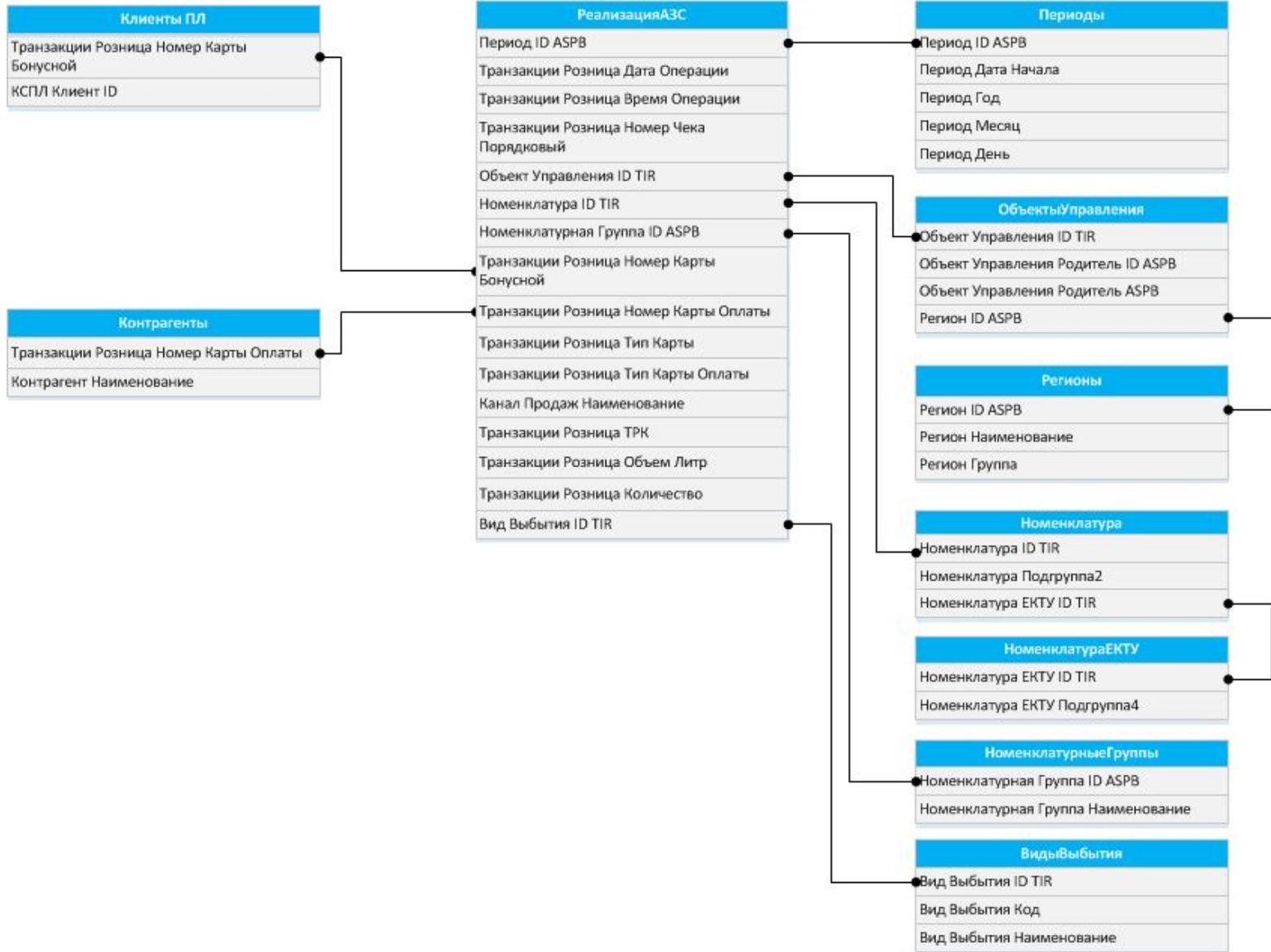
АЗС Формат

- ТРАНЗИТ-XL
- АКТИВ-A
- АКТИВ-L
- АКТИВ-M
- АКТИВ-S
- АКТИВ-XL
- НЕ ТИПОВОГО ФОРМАТА
- ТРАКСТОП-XXL
- ТРАНЗИТ-L
- ТРАНЗИТ-M

Объект Управления

- АЗС 006 Кр кр Горячключевской район М4
- АЗС 001 Кр-р Селезнева 197/2
- АЗС 002 Кр-р Уральская 194/1
- АЗС 003 Кр-р Дорожная 33
- АЗС 004 Кр-р Суворова 2/1
- АЗС 005 Кр кр Новотиторовская
- АЗС 007 Кр кр Ивановская Рыночная 1а
- АЗС 008 Кр кр Горячий Ключ Кучерявого 64
- АЗС 009 РА Майкоп Шовгенова/Строителей 313/2
- АЗС 010 Кр-р Краснодар - Кропоткин 16 км





>1100

активных
пользователей

>1900

реализованных
отчетов, диаграмм,
дашбордов

>200 BI

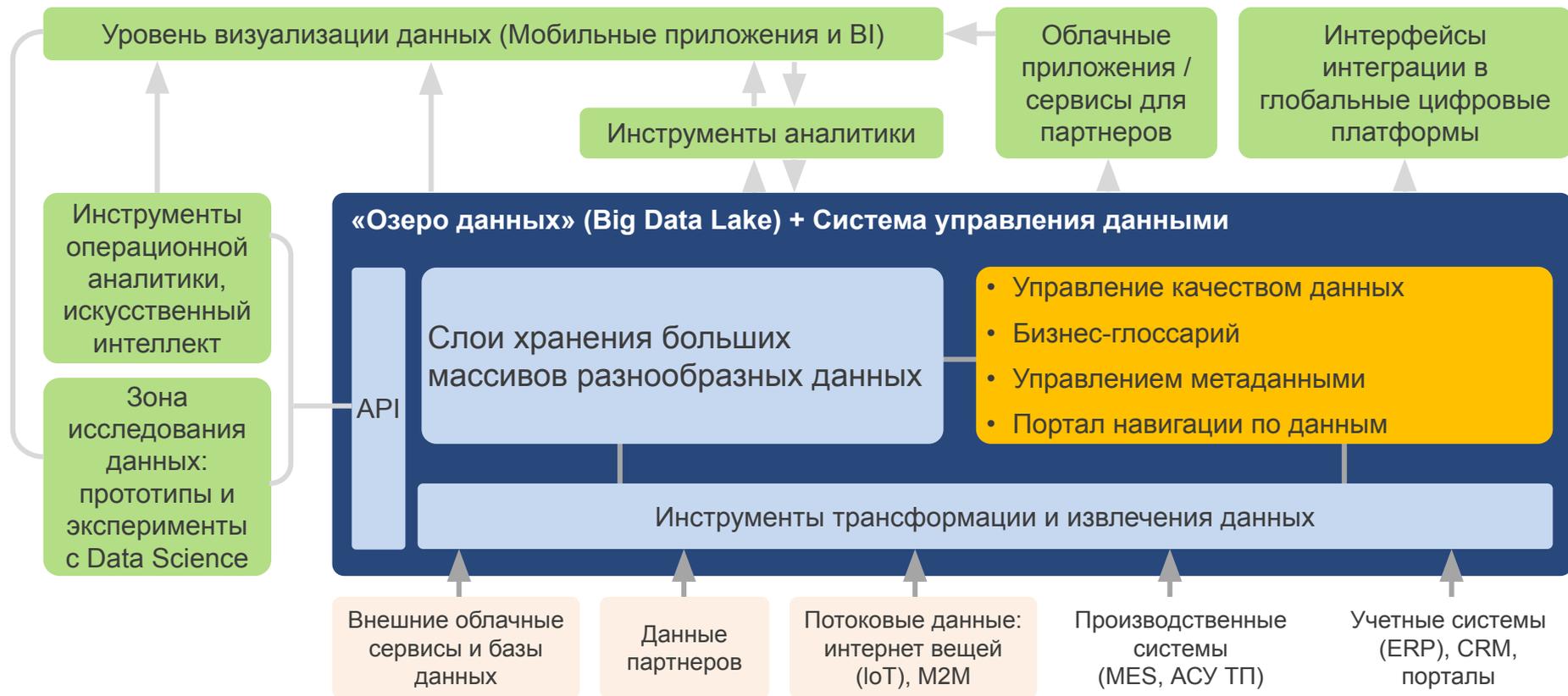
приложений в
продуктиве

>20 Тбайт данных
обрабатывается

>140 источников
данных

Состав подсистем BI показывает разнообразие решаемых аналитических задач

- Финансовый учет
- Расходы
- Экономический анализ
- Финансы
- Инвестиции
- ИТ-сервисы
- Маржинальный доход
- Материальный баланс
- Первичная логистика
- Розничные продажи
- Аналитика розничных продаж
- Сервисы для АЗС
- Сравнительный анализ АЗС
- Модель распределения нефтепродуктов
- Эффективность бизнес-процессов
- Мониторинг цен и ценообразование
- Общая отчетность и АРМы руководителей
- Управление ассортиментом АЗС
- Аналитика объектов партнеров
- Анализ численности персонала
- Аналитика нефтебазового хозяйства



- Создание тематических BI-приложений (варианты: отчетность, дизайн, сложные расчетные алгоритмы)
- Прогнозирование и исследовательские кейсы на поиск закономерностей в данных с использованием машинного обучения (SQL / Python / специализированное ПО)
- Оптимизация, в том числе нелинейная (солверы и специализированное ПО)
- Разработка Big Data (Java, SQL + компоненты Hadoop)
- Имитационное моделирование (AnyLogic)

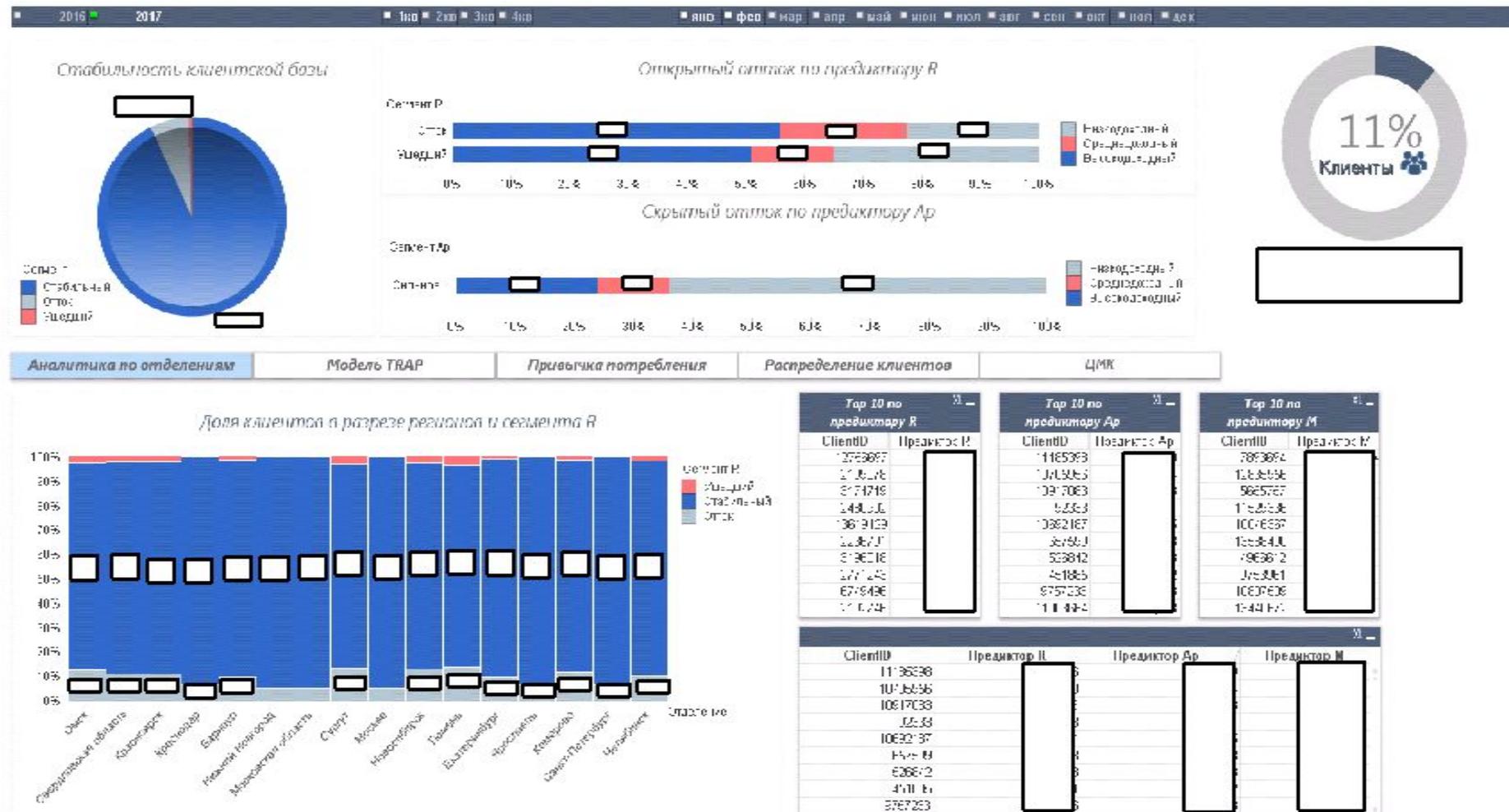
- Студенты 3-6 курсов по профильным направлениям (бизнес-аналитика, бизнес-информатика, прикладная и фундаментальная математика, системный анализ и программирование), желающие развиваться как аналитики / инженеры данных / разработчики BI и работать в бизнес-компаниях
- Приветствуются знание SQL, Python, Java, технологий ML и Hadoop, знакомство с платформой QlikView
- Приветствуется опыт личных достижений в творческой или спортивной сфере
- Форматы привлечения при наличии направления от ВУЗа:
 - Стажировка/производственная практика = 2-3 месяца
 - Стажировка + Дипломная работа = 4-6 месяцев
 - Для более крупных задач с исследовательской составляющей может быть сформирована мини-группа из 2-3 студентов для совместной работы
- Для приема на стажировку студенты проходят собеседование и решают кейсы (кейсы выполняются заочно)
- **Основной отбор: Кейс-чемпионат GPN Intelligence Cup**

- **Где:** в офисах в Санкт-Петербурге, Екатеринбурге
- **График:** период и график работы обсуждаются индивидуально, обычно на 0,5 занятости (2,5 дня в неделю по свободному графику)
- **Оплата стажировки:** для успешно прошедших отборы студентов
- **Обучения и тренинги:** профильные обучения на внутренних курсах и на продуктивных кейсах/приложениях
- **Кураторство:** за стажером закрепляются 1-2 задачи и куратор-эксперт
- **Оформление:** по итогам стажировки формируется отзыв/рекомендательное письмо от имени профильного руководителя Дирекции/руководителя практики
- **Трудоустройство:** для стажеров-выпускников возможно предложение постоянной работы по итогам стажировки или по окончании ВУЗа

2017-2018 учебный год

- «Построение комплексной системы управления данными компании»
- «Совершенствование функции управления данными в рамках цифровой трансформации бизнеса»
- «Разработка BI-приложения для управления бизнес-процессом компании»
- «Проектирование BI-приложения в среде QlikView»
- «Экономико-математические модели исследования потребителя»
- «Применение метода частичного обучения Co-training в задаче классификации клиентской базы»

Сегментация клиентов сети АЗС в R, формирование групп для отправки целевых маркетинговых предложений, визуализация с помощью BI-инструментов



BI-приложение для анализа оценок клиентов в мобильном приложении сети АЗС ГПН

Общие данные
Средняя оценка
Оценки и комментарии
Автоматические АЗС
Автоматические АЗС. Жалобы
Рейтинги

Очистить все

Вперед

Назад

Текущие выборки

Регион	Отделение	Год	Неделя (дата начала недели)	День недели	День месяца
<input type="checkbox"/> Сибирский кластер	<input type="checkbox"/> Екатеринбург	2017	26.12.2016 02.01.2017 09.01.2017	Пн Вт Ср Чт Пт Сб Вс	01 02 03 04 05 06 07 08 09 10 11 12 13 14
<input type="checkbox"/> Уральский кластер	<input type="checkbox"/> Кемерово		16.01.2017 23.01.2017		15 16 17 18 19 20 21
<input type="checkbox"/> Центральный кластер	<input type="checkbox"/> Краснодар				22 23 24 25 26 27 28
	<input type="checkbox"/> Красноярск				29
	<input type="checkbox"/> Москва				
	<input type="checkbox"/> Московская Область				

Обновить данные

Общие данные по оценкам

Оценки ниже "5"

Предустановленные ответы

Для увеличения масштаба кликните на название интересующего региона
Для возврата нажмите на стрелку в левом верхнем углу гистограммы

Доля предустановленных ответов

Вариант предустановленного ответа	Процент
Очереди	29%
Другое (пожалуйста, укажите в комментарии)	27%
Обслуживание кассира	16%
Обслуживание заправщика	16%
Чистота	7%
Продукция кафе/магазина	5%

Для увеличения масштаба в таблице кликните на ячейку интересующего региона:

Кластер	Причины недовольства клиентов	Доля по РФ
Центральный кластер		57,33%
Сибирский кластер		21,96%
	Без предустановленного ответа	19,07%
	Очереди	0,75%
	Другое (пожалуйста, укажите в комментарии)	0,77%
	Обслуживание кассира	0,52%
	Обслуживание заправщика	0,50%
	Чистота	0,21%
	Продукция кафе/магазина	0,14%
Уральский кластер		20,72%
	Без предустановленного ответа	18,20%
	Очереди	0,73%
	Другое (пожалуйста, укажите в комментарии)	0,64%
	Обслуживание кассира	0,47%
	Обслуживание заправщика	0,35%
	Чистота	0,22%
	Продукция кафе/магазина	0,12%

Количество ответов

Кластер	Очереди	Другое	Обслуживание кассира	Обслуживание заправщика	Чистота	Продукция кафе/магазина
Центральный кластер	5130	4586	2561	2686	1183	737
Сибирский кластер	1781	1813	1218	1175	493	
Уральский кластер	1728	1502	1099	820	511	

Вернуться ко всем вариантам ответа

VI-приложение для анализа эффективности закупок по тендерным процедурам

Основной
Планирование закупочной деятельности
Организация закупочной деятельности
Исполнение решений по результатам закупочной деятельности
Данные
Фильтры



Год: 2017

Квартал: Q2 Q1

Месяц: мар апр май июн июл

ДеньМесяца:

1	5	9	13	17	21	25	29
2	6	10	14	18	22	26	30
3	7	11	15	19	23	27	31

ДеньНедели:

Пн	Ср	Пт	Вс
Вт	Чт	Сб	

 Реестр объявленных конкурентных процедур

 Анализ исполнения Плана закупок

 Точность планирования закупочной деятельности

 Информация о проведенной процедуре

 Обеспечение экономической эффективности закупок

 Качество организации и проведения процедуры

 Соблюдение сроков организации и проведения конкурентных отборов

 Уровень конкуренции (с учетом закупки у единственного источника)

 Обеспечение экономической эффективности закупок

Текущие выборки

Заказ На Отбор Вид П...
 Организация Наименование

- СТ и У, МТР, Работы и услуги
- ООО "Газпромнефть-Центр" (Розница), ООО "Газпром нефть - Таджикистан", ООО "Газпромнефть-Красноярск", ООО...

Назад Очистить Вперед

Сценарий Наименование

- Бизнес-план. Год 1 (помесечно)
- Бизнес-план. Год 1 (помесечно) итерация 2
- Бизнес-план. Год 1 (помесечно) итерация 3
- Бизнес-план. Год 1 (помесечно) альтерн...
- Бизнес-план. Год 2 (годовой)
- Бизнес-план. Год 2 (годовой) итерация 2

Заказ На Отбор Вид Предмета Закупки

МТР ПИР и СМР
 СТ и У Работы и услуги

Заказ На Отбор Способ Выбора Пост ав...

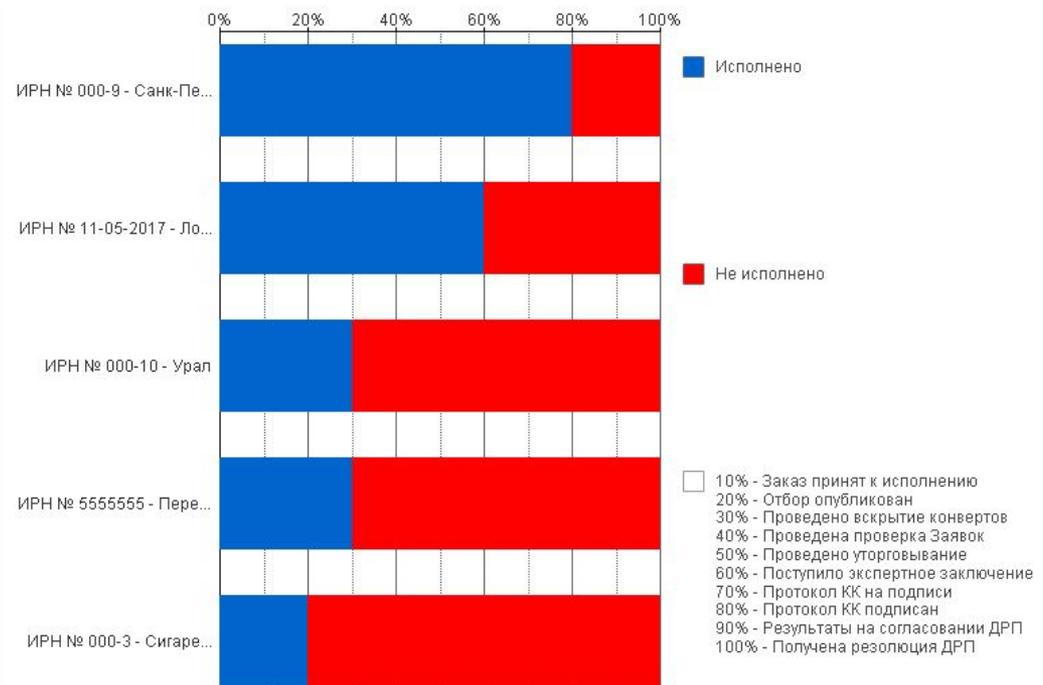
Закрытый отбор Закупка у единстве...
 Открытый отбор

Товарная Категория Наименование

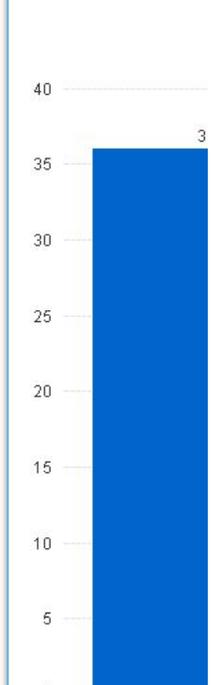
- Инвентарь пожарный
- Мебель корпусная
- Обслуживание пожарных ав томатических...
- Одежда
- Оформление документов для получения...
- Периодические медицинские осмотры
- Прочие ПИР, СМР
- Разработка и согласование проектов но...

Анализ исполнения Плана закупок

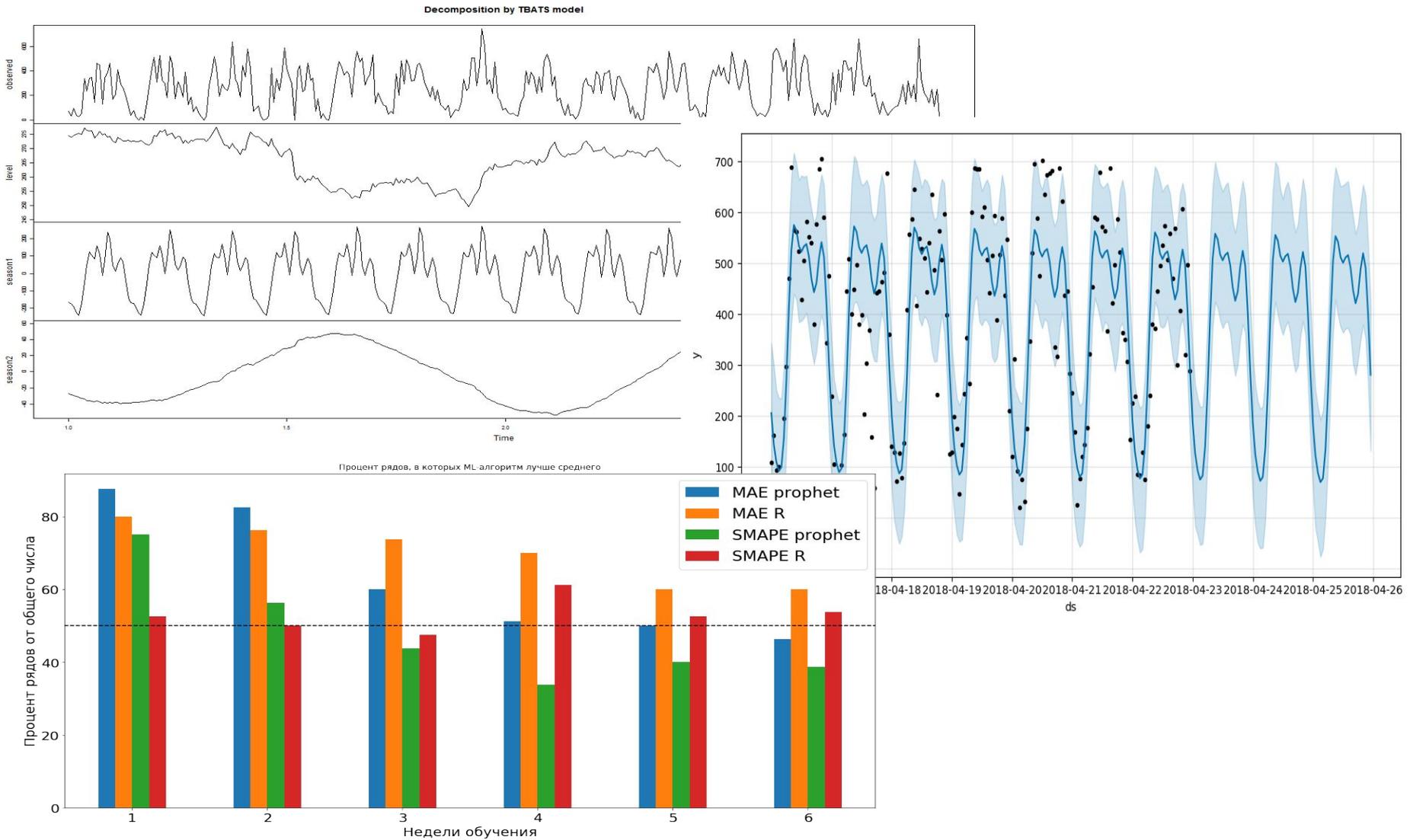
Исполнено



Точность планирования



Прогнозирование продаж АЗС с использованием R, Python и пакетов для работы с временными рядами



Совместное использование QlikView с R для задач прогнозирования, ансамбль моделей для временных рядов

Главная
Прогноз
Анализ проблемных рядов
Качество данных



Год: 2014 2015 2016

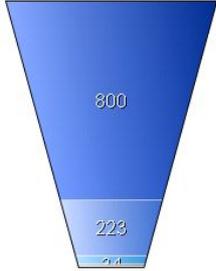
Квартал: 1 кв 2 кв 3 кв 4 кв

Месяц: янв фев мар апр май июн июл авг сен окт ноя дек

Качество ретропрогноза

Кластер

- Центр: 1,22%
- Новосибирск: 2,15%
- Урал: 2,22%



Откл. от факта, %

- 0 ≤ x < 5: 800
- 5 ≤ x < 10: 223
- 10 ≤ x < 15: 34
- 15 ≤ x < 20: 7
- 25 ≤ x < 30: 2
- Другие: 4

Норма отклонения: 3%



Период	Год	Месяц	декабрь 2015	январь 2016	февраль 2016	За 3 мес
Отделение (кол-во АЗС)						
Тюмень (82)			0,5%	1,9%	3,8%	2,0%
Барнаул (37)			0,6%	3,7%	3,6%	2,5%
Сургут (31)			1,1%	2,1%	2,9%	1,1%
Красноярск (36)			0,2%	2,5%	12,9%	4,7%
Санкт-Петербург (95)			2,0%	6,1%	1,7%	1,8%
Ярославль (66)			0,1%	3,0%	0,5%	1,2%
Московская область (103)			1,4%	4,5%	0,3%	1,0%
Эмск (85)			2,9%	1,4%	3,5%	0,4%
Новосибирск (98)			1,0%	2,4%	4,4%	2,5%
Свердловская область (78)			0,2%	1,6%	3,0%	1,5%
Лосква (90)			2,9%	7,0%	0,8%	0,9%
Кемерово (93)			1,2%	0,1%	4,6%	1,9%
Нижний Новгород (40)			3,0%	6,0%	2,3%	1,5%
Екатеринбург (60)			0,7%	4,3%	3,9%	2,3%
Челябинск (76)			0,3%	7,4%	5,7%	4,0%

Выбрано АЗС: 1 070

Модель: Best model, AutoArima, ETS, Trend+Fourier, Trend+Season, Trend+Fourier

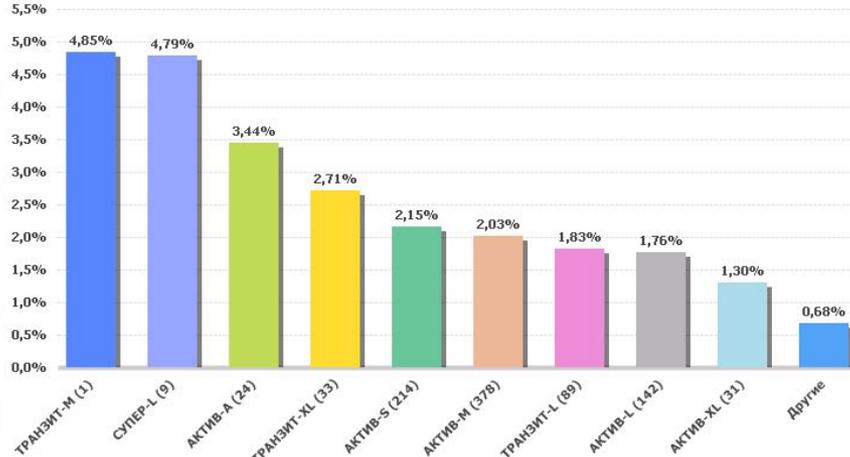
Очистить все

Анализ отклонений КЛАСТЕР - ОТДЕЛЕНИЕ - АЗС

Анализ отклонений ФОРМАТ - АЗС - ИГ

Динамика во времени

Анализ отклонений ФОРМАТ - АЗС - ИГ



Формат	Процент отклонения
ТРАНЗИТ-М (1)	4,85%
СУПЕР-L (9)	4,79%
АКТИВ-A (24)	3,44%
ТРАНЗИТ-XL (33)	2,71%
АКТИВ-S (214)	2,15%
АКТИВ-M (378)	2,03%
ТРАНЗИТ-L (89)	1,83%
АКТИВ-L (142)	1,76%
АКТИВ-XL (31)	1,30%
Другие	0,68%

Частота выбора модели

Лучшая Мо...	Рядов	%
ETS	4 372	100,0%
ETS	1 267	29,0%
Trend+Fourier	1 107	25,3%
Trend+Season	631	14,4%
AutoArima	602	13,8%
Trend*Fourier	445	10,2%
Trend*Season	320	7,3%

Подразделение Кластер

Объект Управление Подразделени...

Регион Наименование

Объект Управление Формат АЗС

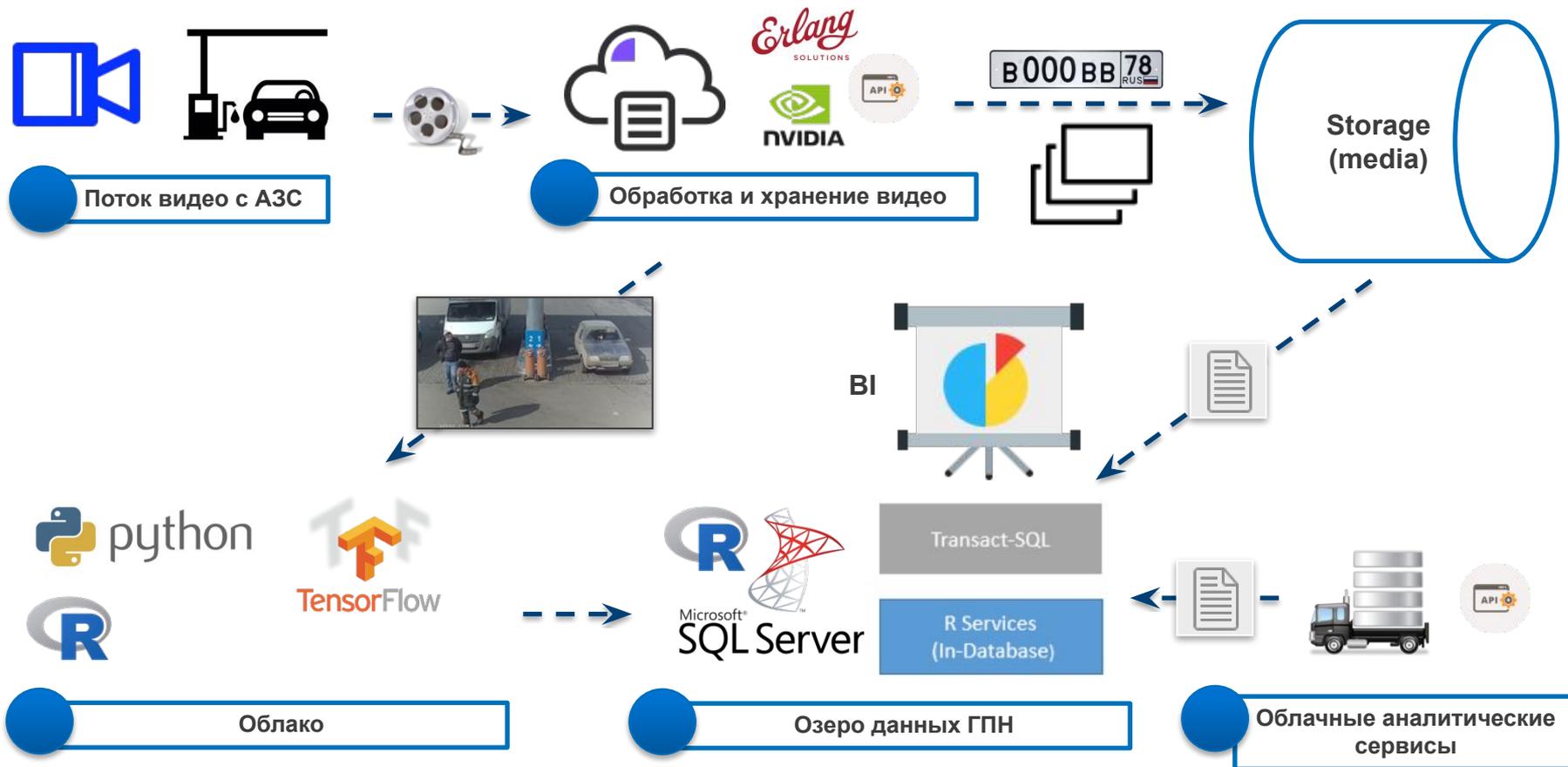
Объект Управление Родитель ASPB ...

Номенклатурная Группа Наименова...

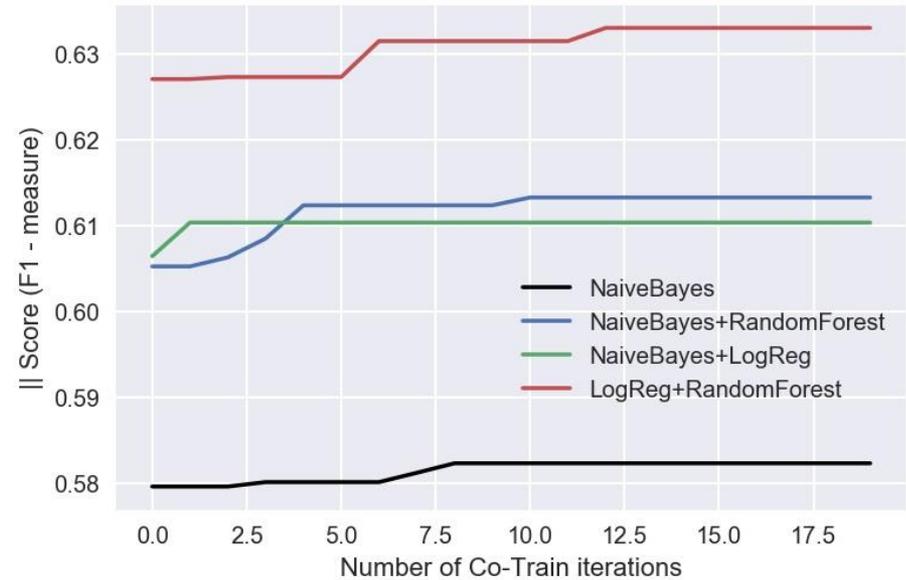
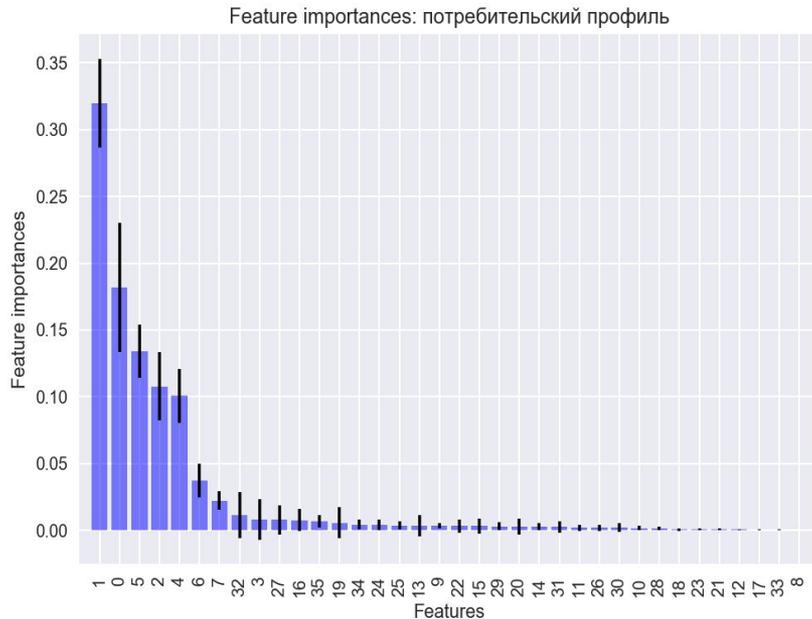
VI-приложение «Операционные показатели эффективности программы лояльности»



Система распознавания номеров, цвета и сегмента автомобилей на АЗС



Классификация клиентской базы с использованием инструментов Python



	Precision	Recall	F1-score	Total F1-score
NaiveBayes	0,53	0,34	0,41	0,59
LogReg	0,62	0,36	0,46	0,63
RandomForest (500 est)	0,61	0,32	0,42	0,61
CT:NaiveBayes	0,55	0,24	0,33	0,58
CT:LogReg	0,52	0,57	0,55	0,63
CT:RandomForest (500est)	0,53	0,28	0,37	0,59
CT:NaiveBayes+RandomForest	0,55	0,29	0,38	0,60
CT:NaiveBayes+LogReg	0,52	0,37	0,43	0,60
CT:LogReg+RandomForest	0,55	0,42	0,48	0,63

Открыта регистрация на GPN Intelligence Cup — ежегодный кейс-чемпионат по решению аналитических задач Дирекции региональных продаж «Газпром нефти» для студентов старших курсов ВУЗов по направлениям в области экономики, информатики и технологий обработки и анализа данных.

Участникам предстоит погрузиться в работу компании и решить актуальные задачи бизнеса по разработке стратегии продвижения продуктов, продвинутой аналитике и инжинирингу данных.

Победители GPN Intelligence Cup смогут попасть на оплачиваемую стажировку в компанию!

Регистрация проходит на сайте <http://gpn-cup.ru> по двум направлениям:

- «Анализ и инжиниринг данных» - до 22 октября 2018.

Регистрация по направлению "Анализ и инжиниринг данных" возможна на все или произвольное количество тем на ваш выбор, темы:

- 1). Разработка BI-приложений;
- 2). Разработка алгоритмов обработки больших массивов данных;
- 3). Продвинутая аналитика

- «Бизнес и аналитика» - до 9 ноября 2018г

- ✓ Ксения Гальперина – эксперт по качеству данных
Galperina.KS@ekb.gazprom-neft.ru
- ✓ Михаил Казанцев – Руководитель группы разработки BI
Kazantsev.MP@gazprom-neft.ru
- ✓ Иван Черницын – Руководитель Офиса Управления данными
Chernitsyn.IG@gazprom-neft.ru