



# Ведение DS- проектов

Лекция 3



**Алексей Кузьмин**

Директор  
разработки  
ДомКлик.ру

Работаю в ДомКлик.ру с 2016 года  
Руководжу направлением Data Science и  
работы с данными  
До этого работал в компании АBBYY, где  
занимался распознаванием языков со  
сложной письменностью  
Окончил мехмат МГУ

# О чем поговорим?

# Сегодня на лекции

4

**01** Что такое отчет по исследованию

**02** Рекомендации по его написанию

**03** Пример хорошего отчета

# Отчет по исследованиям

**Зачем он нужен?**

## Аналитический продукт

Результат работы аналитика:

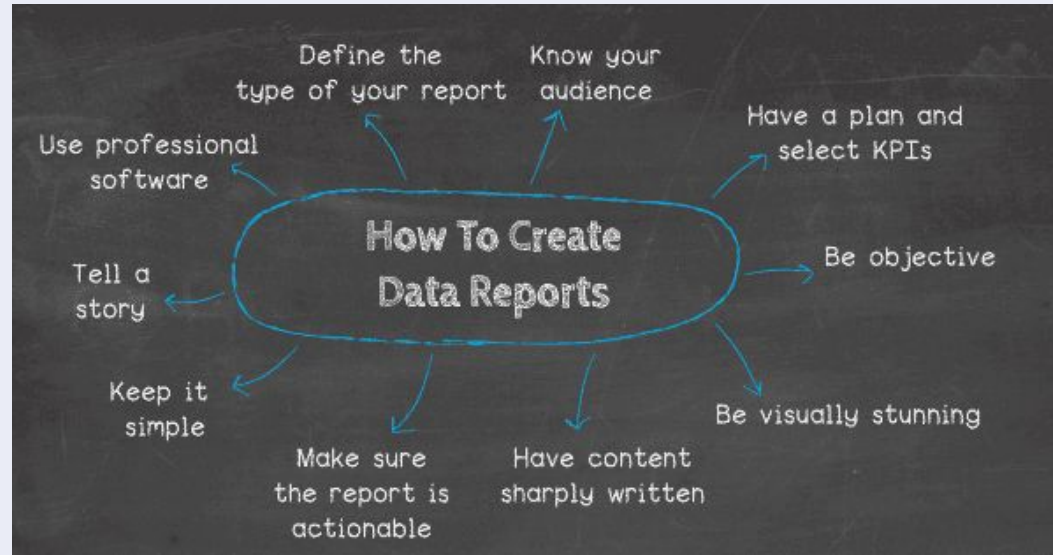
- Это **бизнес-польза**, которую можно получить сделав определенное действие
- Оно может быть оформлено в виде отчета или веб-сервиса
- **Но!** Оно **бесполезно**, пока оно не используется и не приносит пользу

## Продавайте свой продукт





## Как написать хороший отчет?



## Определите тип отчета

- Информационный
  - Просто показываем фактическую информацию - меньше в размере, структура на так важна
  
- **Аналитический**
  - Показывает рекомендации и советы, подводит к выводам

## Оцените аудиторию

- Для кого мы показываем отчет?
- Какие вопросы могут задать?
- Какой уровень детализации и оформления нужен?

## Аудитория - руководство

- Высокоуровневые показатели
- Упор на ожидаемый эффект и риски
- Что получим / что потеряем?
- Краткость - сестра таланта

# Аудитория - руководство

## Financial KPI Dashboard

### Current Working Capital

**Current Assets** **\$129,000**

Cash	\$34,000
Accounts Receivable	\$59,000
Inventory	\$31,000
Pre-Paid Expenses	\$5,000

==

**Current Liabilities** **\$68,000**

Accounts Payable	\$25,000
Credit Card Debt	\$2,000
Bank Operating Credit	\$27,000
Accrued Expenses	\$9,000
Taxes Payable	\$5,000

==

**Working Capital** **\$61,000**

**Current Ratio** **1.90**

### Cash Conversion Cycle in Days - Last 3 Years

$$59.0 \text{ DSO} + 36.0 \text{ DIO} - 48.3 \text{ DPO} = 46.8 \text{ CCC}$$



### Vendor Payment Error Rate - Last 12 Months



# Аудитория - руководство

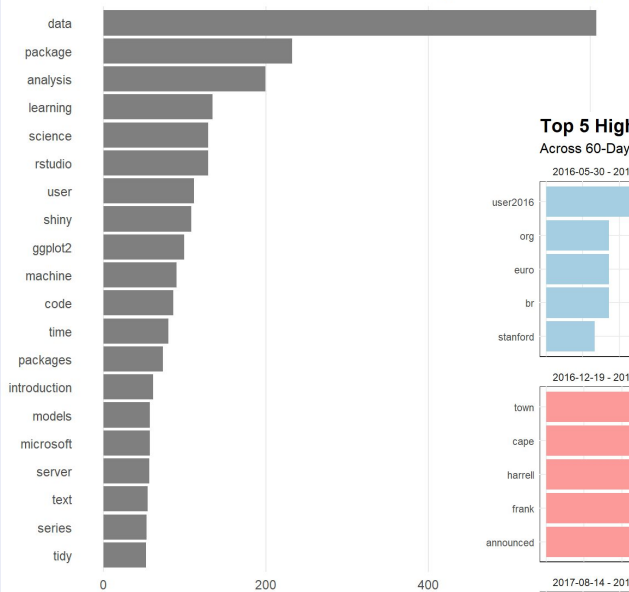


## Аудитория - коллеги

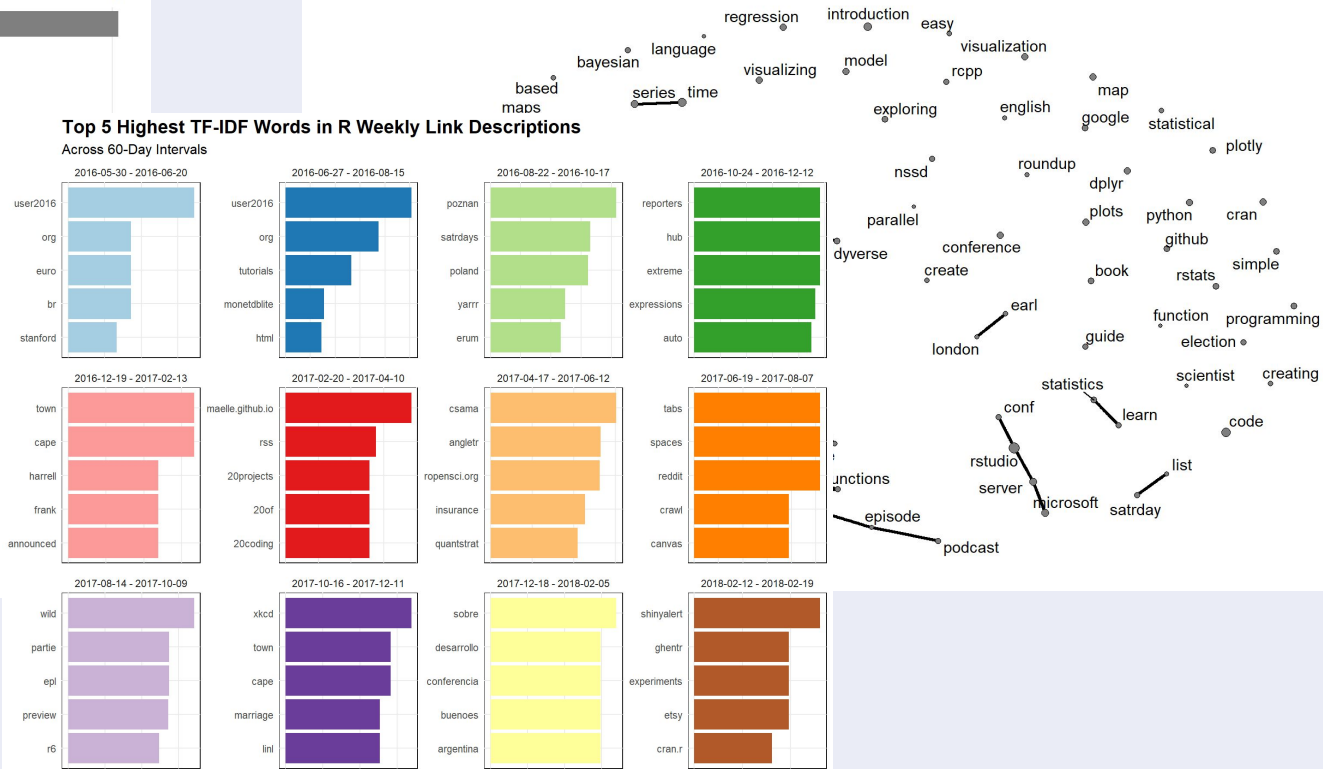
- Больше деталей
- Быть готовым отвечать на глубокие вопросы по существу

## Аудитория - коллеги

## Most Common Words in R Weekly Link Descriptions



## Network of Pairwise Correlations of Words in R Weekly Link Descriptions

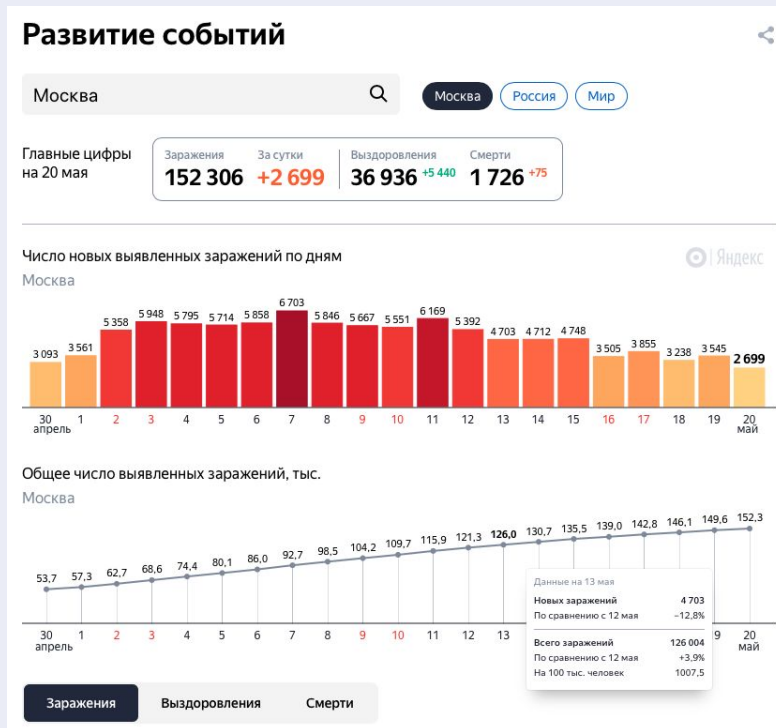




## Аудитория - широкая аудитория

- Проще текст
- Больше графиков и визуализаций
- По возможности - интерактив

# Аудитория - широкая аудитория



## Иметь детальный план

- Четко сформулируйте - что Вы хотите донести
- Выберите ключевые метрики / показатели, которые подчеркнут Ваш тезис
- Сфокусируйтесь на ключевой мысли

## Быть объективным

- Объективный и нейтральный способ повествования
- Исследователь - он смотрит сверху и изучает. Формально - он не участник исследуемого процесса

## Используйте визуализации

- Визуализация облегчает восприятие сложных концепций
- Подбирайте правильную визуализацию
- Про это не говорят - но убедитесь, что отчет будет выглядеть хорошо там, где его будут смотреть (desktop vs mobile и тп)

## Четко формулируйте мысли

- Хотя визуализации важны, многие отчеты также содержат текст
- Используйте корректный деловой язык
- Сокращайте, где возможно

## Содержит призыв к действию

- Если это аналитический отчет - то самое важное - это предложения
- Предложения - это логическое завершение отчета
- Способствует повышению data-driven в компании

## Максимальная простота

- Переусложнение - корень всех проблем
- Можно сделать отчет, который будет виден только в AR-реальности. Но зачем?
- Не используете сложные цветовые схемы, сложные диаграммы и тп там, где без этого можно обойтись

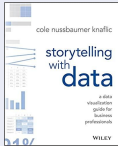


<https://www.ozon.ru/context/detail/id/137634276/>



# Расскажите историю

- В слайдах должна быть логика, которая рассказывает историю исследования и подводит к выводам



[https://www.amazon.com/Storytelling-Data-Visualization-Business-Professionals/dp/11190022](https://www.amazon.com/Storytelling-Data-Visualization-Business-Professionals/dp/1119002257)

57

- <https://netology.ru/programs/data-journalist#/audience>

# Как оформить работу?

## Общие правила

- **Быть проще**

Читатели отчета не погружены в детали Вашего исследования. Выливание на них всех имеющихся в Вашем распоряжении графиков/выводов/таблиц только помешает восприятию результатов

## Общие правила

- **Сначала общее, потом частное**

Лучше показать основную канву исследования, основные пути и тп. Затем уже переходить к детальным срезам/инсайтам по каждому из этапов.

## Общие правила

- **Задаем вопросы, потом отвечаем. Не наоборот**

Данные должны отвечать на вопросы исследования, определенные заранее (в начале исследования/отчета).

## Общие правила

- **Не надо подробно описывать процесс сбора данных**

Интерес могут представлять только источники. Детали работы и их взаимодействия никого не интересуют.

## Общие правила

- **Текст/таблица/график - одна цифра - один способ визуализации**

Не повторяйте одну и ту же информацию в более чем одном формате. Выберите лучший способ передать сообщение.

## Текст

- Лучше для выводов/интерпретаций (цифры лучше в таблицах или графиках)
- Но если цифр не очень много - можно и в тексте
  - “Группа для исследования состояли из 80% мужчин и 20% женщин”
- Слова для подчеркивания качества - можно не использовать. “Уникальные результаты”, “экстремальные значения” и тп. Точные цифры лучше скажут все за вас
  - “Уникальный результат - повышение конверсии на 1%”
  - “В результате удалось достичь повышения конверсии на 1.32%”



## Таблицы

- Таблицы полезны для выделения точных числовых значений
  - отношения и тренды лучше показывать диаграммами и графиками.
- Суммируют большое количество связанных данных и позволяют проводить сравнение между группами переменных.

## Таблицы - Заголовок

- Краткое
- Соответствует содержанию
- Резюмировать суть, а не повторять заголовки столбцов и строк
  - «Сравнение результатов полного анализа крови среди разных рас»
  - «Сравнение гемоглобина, количества тромбоцитов и общего количества лейкоцитов среди малайцев, китайцев и индийцев»

## Таблицы - колонки и строки

- Столбцы содержат сравниваемые данные
- Столбцов не должно быть очень много
- Важный результат в таблице лучше дополнительно акцентировать

## Таблицы - значения

- Единый формат данных в рамках столбца
- Округлите числа до наименьшей значимой точности
- Избегайте большого количества нулей ( $11,3 \times 10^6$  vs 11 300 000). При этом степень лучше использовать единую в рамках столбца

## Таблицы - плохо

Table 1A:

**Baseline waist circumference, HbA<sub>1c</sub>, Blood pressure and LDL-cholesterol level among Malay, Chinese, Indian and others races**

	Malay	Chinese	Indian	Others
Waist circumference	98	102	105	95
HbA <sub>1c</sub>	8.89	8.66	9.0	8.7
SBP	165.1	164.0	170.34	168
DBP	98.5	101	99.3	97.6
LDL-C	3.8	3.9	3.4	3.1

## Таблицы - хорошо

Table 1B:

**Mean (SD) baseline diabetic metabolic control among different races**

	WC <sup>*</sup> (SD)	HbA <sub>1c</sub> (SD), %	SBP <sup>†</sup> (SD)	DBP <sup>‡</sup> (SD)	LDL-C <sup>§</sup> (SD)
Malay	98 (15)	8.9 (1.5)	165 (21)	98 (13)	3.8 (0.9)
Chinese	102 (18)	8.7 (2.1)	164 (28)	101 (15)	3.9 (0.7)
Indian	105 (22)	9.0 (1.8)	170 (36)	99 (22)	3.4 (1.2)
Others	95 (28)	8.7 (2.5)	168 (40)	97 (28)	3.1 (1.0)

\*WC, waist circumference (in cm)

†SBP, systolic blood pressure (in mmHg)

‡DBP, diastolic blood pressure (in mmHg)

§LDL-cholesterol (in mmol/L)

## Графики

- Используйте, чтобы показать тренды и тенденции
- Приводит к потере деталей и точных значений

## Графики

- Гистограммы - для отображения категориальных данных
- Не используйте трехмерные графики и диаграммы
- Линейные графики - для отображения изменений значений между переменными в течении определенного периода времени
- Графики с накоплением - для отображения достижения какой-то величины / отслеживания KPI со временем (% текучки с начала года)



# Примеры

## Отчет по модели

<http://cs229.stanford.edu/proj2014/Jithin%20Thomas,%20Bhrugurajsinh%20Chudasama,%20Chinmay%20Duvedi,%20Learning%20Facial%20Expressions%20From%20an%20Image.pdf>

# Итоги

## Итоги

1. Узнали что такое отчет по исследованию.
2. Проговорили об основных принципах его написания
3. Посмотрели пример

# Домашнее задание

## Домашнее задание

Возьмите свое решение задачи с винами (ДЗ #2) и напишите небольшой отчет по этой задаче (1-1.5) страницы



**Спасибо за  
внимание**