**Карл Андерсон. Аналитическая культура**

Это практическое руководство по внедрению управления на основе данных. Написано, скорее для менеджеров, чем для специалистов, хотя интересна книга будет обеим категориям читателей. Математики почти нет. Аналитическая культура основана на аналитической цепочке ценностей – последовательности действий от сбора данных до принятия решения. Последний шаг в этой цепочке чрезвычайно важен. Если данные игнорируются, а большой босс делает что пожелает, сбор этих данных не имеет смысла. Автор уделяет внимание «продаже» аналитических результатов руководству компании; рассматривает новые должности в современной компании: директор по данным и директор по аналитике; доступно и интересно рассказывает о А/В-тестировании. Книга имеет обширную библиографию (почти всю на английском языке).

Карл Андерсон. Аналитическая культура. От сбора данных до бизнес-результатов. – М.: Манн, Иванов и Фербер, 2017. – 336 с.



Купить цифровую книгу в [ЛитРес](https://www.litres.ru/karl-anderson-11438349/analiticheskaya-kultura-ot-sbora-dannyh-do-biznes-rezultatov/?lfrom=13042861), бумажную книгу в [Лабиринте](https://www.labirint.ru/books/591689/?p=13320)

### Глава 1. Что значит «на основе данных»?

Без данных вы просто еще один человек с собственным мнением.  
[*Уильям Эдвардс Деминг*](http://baguzin.ru/wp/?p=2138)

Управление на основе данных подразумевает формирование инструментов, способностей и, что самое важное, корпоративной культуры, которая опирается на данные. Культура организации — доминирующий фактор, который устанавливает ожидания относительно того, насколько демократичным будет процесс работы с данными, как эти данные станут использоваться внутри организации, какие ресурсы, в том числе образовательные, станут инвестироваться в использование данных как стратегического актива компании.

Следует различать отчетность и анализ (рис. 1), информацию и выводы (рис. 2).

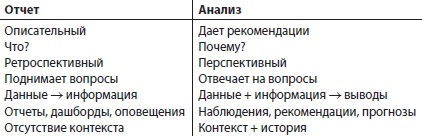


Рис. 1. Основные характеристики отчета и анализа

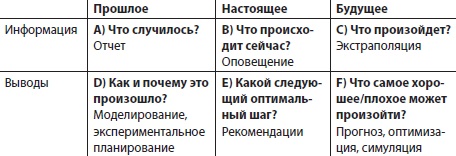


Рис. 2. Основные вопросы, на которые отвечает аналитика (на основе книги Тома Дэвенпорта [Аналитика как конкурентное преимущество](http://baguzin.ru/wp/?p=2923))

### Глава 2. Качество данных

При получении любой информации аналитику в первую очередь следует в той или иной форме провести разведочный анализ данных для оценки их качества. Простой способ – сделать сводку из данных. Для каждого показателя можно составить пятичисловую сводку: два крайних значения (максимальное и минимальное значение), нижний (25-й процентиль) и верхний (75-й процентиль) квартили и медиану. Посмотрите на крайние значения. Насколько они адекватны? Они выше или ниже значений, которые вы могли бы ожидать?

Два важных навыка, которые должны развивать в себе аналитики, — прогнозирование возможных результатов и способность предварительно оценивать данные. Одна из наиболее существенных проблем — неполные или пропущенные данные (подробнее см. Литтл Р., Рубин Д. [Статистический анализ данных с пропусками](https://www.ozon.ru/context/detail/id/19146639/?partner=baguzin). – М.: Финансы и статистика, 1990).

При обнаружении проблемы с качеством данных важно отследить источник данных. Для метаданных, хранящих информацию об источнике данных и историю их изменений, я использую термин «происхождение данных». Эти метаданные делятся на два типа: история источников (отслеживает, откуда появились данные) и история преобразований (отслеживает, какие изменения претерпевали данные).

### Глава 3. Сбор данных

Собирайте все доступные данные. Никогда не знаешь, какая информация может понадобиться, а шанс собрать данные часто выдается только один. Чем лучше компания поймет своих покупателей, их вкусы, намерения, желания, тем успешнее ей удастся улучшить пользовательский опыт своих клиентов благодаря персонализации, рекомендациям или совершенствованию сервиса, что будет способствовать возникновению так называемого [длинного хвоста](http://baguzin.ru/wp/?p=4273). Контекст — наше все.

### Глава 4. Специалисты по аналитике

Специалистов по работе с данными можно разделить на пять групп по ключевому навыку: бизнес; математика / анализ операций; машинное обучение / большие данные; программирование; статистика.

Если вы аналитик, бросьте вывоз самому себе: в течение следующего месяца или квартала освойте еще один инструмент или программу. Если вы руководите аналитиками, поставьте перед ними такую задачу. Попробуйте и увидите, какой будет результат. Вы будете удивлены.

Есть веские основания утверждать, что владение такими инструментами, как R, Python, инструментарием Hadoop, D3, а также масштабируемыми инструментами машинного обучения, свидетельствует о более высокой квалификации аналитика, позволяя ему претендовать на более высокооплачиваемую позицию, чем, когда аналитик владеет такими инструментами, как SQL, Excel и платформы RDB [реляционных баз данных]. Мы также пришли к выводу, что чем б***о***льшим числом инструментов способен пользоваться аналитик, тем лучше: если вы задумываетесь о том, чтобы научиться применять инструмент из набора Hadoop, лучше изучите сразу несколько.

Опрос 2014 года показал разницу в оплате труда почти в 15 тыс. долл. между аналитиками, умеющими работать с программным кодом, и не умеющими. Так что если это ваше слабое место, окажите себе услугу, научитесь программировать!

### Глава 5. Анализ данных

Если достаточно долго мучить данные, они признаются [в чем угодно].  
[Рональд Коуз](http://baguzin.ru/wp/?p=5116)

Анализ — преобразование данных в выводы, на основе которых будут приниматься решения и строиться действия с помощью людей, процессов и технологий. Данные представляют собой сырые, необработанные факты об окружающем мире. Информация — собранные, обработанные данные, в то время как знания — это набор ментальных моделей и убеждений об окружающем мире, который сформировался на основе информации, полученной на протяжении какого-то периода времени.

Джеффри Лик, старший преподаватель биостатистики в Университете Джонса Хопкинса, а также один из редакторов блога о статистике, выделяет шесть типов анализа данных:

* описательный (descriptive);
* разведочный (exploratory);
* индуктивный (inferential);
* прогностический (predictive);
* каузальный (причинно-следственный) (causal);
* механистический (mechanistic).

*Описательный анализ.* Ранее уже упоминали составление пятичисловой сводки. Можно также выделить меры среднего уровня («середина» данных), меры рассеивания (разброса данных) и формы распределения. Описательный анализ — важный первый шаг. При этом просто итоговых цифр может быть недостаточно. Одна из проблем заключается в том, что большое число значений сводится к нескольким итоговым цифрам. А потому не стоит удивляться, что одни и те же итоговые статистические показатели могут описывать разные выборки с разным распределением данных, формами и свойствами.

Предположим, у вас 4 набора данных с двумя переменными со следующими характеристиками:

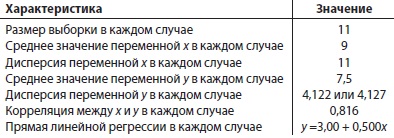


Рис. 3. Характеристики четырех наборов данных

Это система с жесткими заданными ограничениями. Значит, графики этих четырех наборов данных с идентичными статистическими характеристиками должны быть достаточно похожими, не так ли? А вот рис. 4 показывает, что это далеко не так.

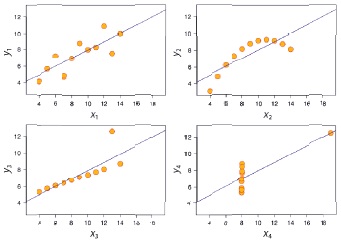


Рис. 4. Квартет Энскомба. В каждом из четырех наборов данных идентичны среднее значение *х*, среднее значение *y*, дисперсия *х*, дисперсия *y*, корреляция и прямая линейной регрессии

Применение графиков для визуализации и изучения данных получило название *разведочного анализа данных*. Наибольшую известность он приобрел благодаря продвижению американским математиком Джоном Тьюки в книге [Анализ результатов наблюдений. Разведочный анализ](http://baguzin.ru/wp/?p=15897), опубликованной в 1977 году.

*Индуктивный анализ.* Как правило, мы делаем выводы обо всей генеральной совокупности на основе взятой из нее выборки, так как полный сбор данных бывает слишком дорогим, непрактичным, а иногда и просто невозможным.

*Прогностический анализ* строится на индуктивном анализе. Цель в том, чтобы изучить взаимосвязи между переменными на основе существующего набора данных и разработать статистическую модель, способную прогнозировать значения для новых, неполных или будущих точек данных. Реальность такова, что мы не в состоянии точно предсказать какие-то неожиданные события и катастрофы, так называемых [черных лебедей](http://baguzin.ru/wp/?p=1533). При этом во многих аспектах бизнеса и других областях знаний есть достаточные сигналы, с обработкой которых прогностический анализ отлично справляется. Например, в 2008 году [Нейту Сильверу](http://baguzin.ru/wp/?p=15853) удалось предсказать результаты выборов в Сенат и победителей в 49 штатах из 50 (рис. 5).

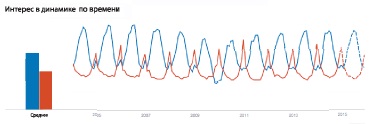


Рис. 5. Инструмент Google Trends отражает предсказуемую сезонную закономерность интереса к солнечным очкам (верхняя синяя кривая) и перчаткам (нижняя красная кривая) в период 2004-2014 годов и прогноз на год, до 2015-го

### Глава 6. Разработка показателей

В компании с управлением на основе данных должна быть четкая стратегия, а также конкретный набор ключевых показателей эффективности деятельности (KPI) — для отслеживания, в верном ли направлении и насколько успешно идет развитие бизнеса. Как и цели, KPI должны соответствовать критериям SMART и быть:

* конкретными (Specific);
* измеримыми (Measurable);
* достижимыми (Achievable);
* ориентированными на результат (Result-oriented);
* ограниченными во времени (Time-bound).

### Глава 7. Сторителлинг на основе данных

«Каждый набор, каждая база данных, каждая таблица способны рассказать целую историю», — уверен Стюарт Франкел, CEO компании Narrative Science. Работа специалиста по анализу данных заключается в том, чтобы увидеть эту историю или хотя бы историю, интересную для компании, сформулировать ее и донести до аудитории. Более того, аналитикам следует позаботиться о точности истории, которая должна быть подтверждена практикой. В противном случае люди придумают свою историю, опираясь на сомнительные данные.

История должна содержать основные выводы, особенности данных или присущие им закономерности, чтобы по возможности раскрывать причины происходящего, а также смотреть в будущее, делать прогнозы и формулировать рекомендации для компании. Взгляните на рис. 6. Вам ничего не кажется необычным или интересным?

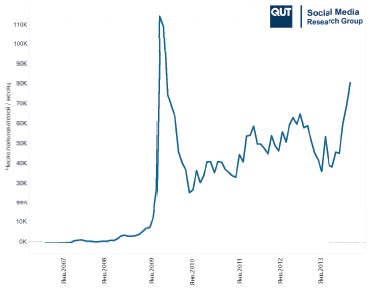


Рис. 6. Регистрация новых пользователей из Австралии в Twitter

Рис. 6, дополненный описательной частью, более полезен. График и описание дополняют друг друга. Требуется качественная визуализация, чтобы обнаружить закономерности в данных в ходе проведения анализа, а затем продемонстрировать их аудитории. И помимо этого требуется знание точной и достоверной истории для интерпретации данных и построения возможных прогнозов.

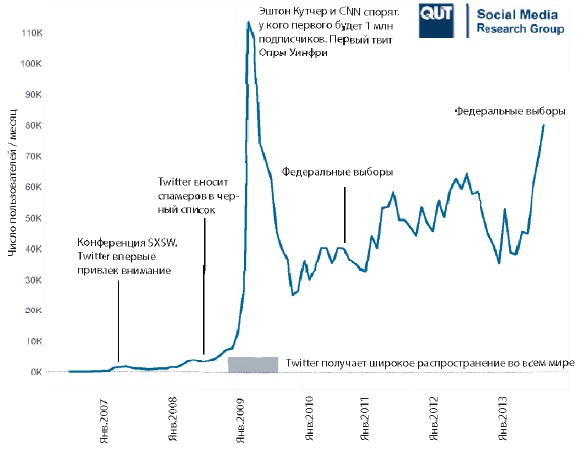


Рис. 7. Аннотированная версия предыдущего рисунка

Если вас интересует визуализации данных рекомендую начать с золотого стандарта – книг Эдварда Тафти Envisioning Information («Представление информации»), Visual Explanations («Визуальные объяснения») и The Visual Display of Quantitative Information (Graphics Press) («Визуальное отображение количественной информации»). Книги на русском языке не издавались.

Выбор типа диаграммы — основной фактор с точки зрения способности сделать презентацию данных понятной для пользователей. Сосредоточьтесь на одной из четырех причин, по которым мы вообще строим диаграмму:

* Сравнение. Например, сравнение групп или сравнение изменений во времени.
* Распределение. Необходимость показать изменчивость набора данных.
* Взаимосвязи. Необходимость отразить корреляцию между переменными.
* Состав. Необходимость показать, как распределяются данные между двумя или более категориями Рис. 8).

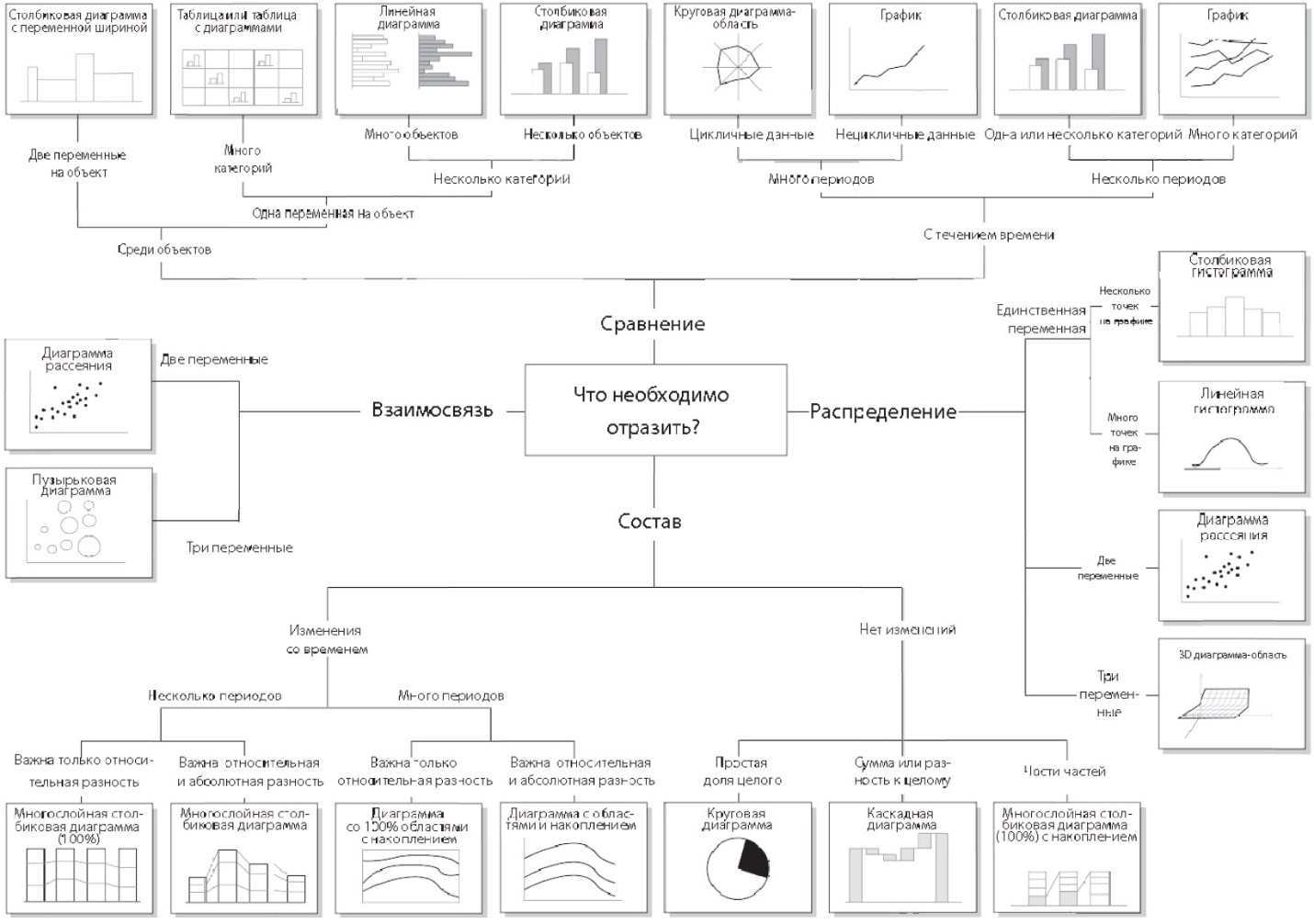


Рис. 8. Типы диаграмм, каждый из которых отвечает определенной задаче (см. также постер [Graphic Continuum](https://www.scribblelive.com/blog/2014/10/01/graphic-continuum/), [Джин Желязны. Говори на языке диаграмм](http://baguzin.ru/wp/?p=190) и [Новые диаграммы в Excel 2016](http://baguzin.ru/wp/?p=16080))

Исключите «графический мусор» и излишества и сконцентрируйтесь на данных и сообщении. Термин [графический мусор](http://baguzin.ru/wp/?p=2071) ввел в употребление Эдвард Тафти для обозначения элементов, отвлекающих внимание. «Графический мусор» — все визуальные элементы диаграмм и графиков, в которых нет необходимости для понимания представленной информации или которые отвлекают от нее. Минималистский подход Тафти отличается категоричностью.

Сравните диаграммы в [r/dataisbeautiful](https://www.reddit.com/r/dataisbeautiful/) и [r/dataisugly](https://www.reddit.com/r/dataisugly/). Почему первые такие ясные, а вторые такие бестолковые? Спросите себя, что бы вы сделали иначе.

В контексте управления на основе данных я не большой поклонник инфографики: сегодня инфографика превратилась в «веселые картинки», приправленные парой фактов, которые обычно создают дизайнеры, а не аналитики. По моему мнению, у подобной инфографики слишком низкое соотношение данных и чернил (data-to-ink ratio), как его определил Эдвард Тафти. Фактически в большинстве случаев инфографика страдает от «графического мусора» и от недостатка данных.

Многие компании ошибочно измеряют степень управления на основе данных количеством производимых ими отчетов и числом дашбордов, которыми они пользуются. Дашборды можно условно разбить на три категории: управленческие или стратегические, аналитические, операционные. Стратегические дашборды (рис. 9) обеспечивают общий обзор деятельности компании и, как правило, концентрируются на системе показателей (например, KPI и их цели).

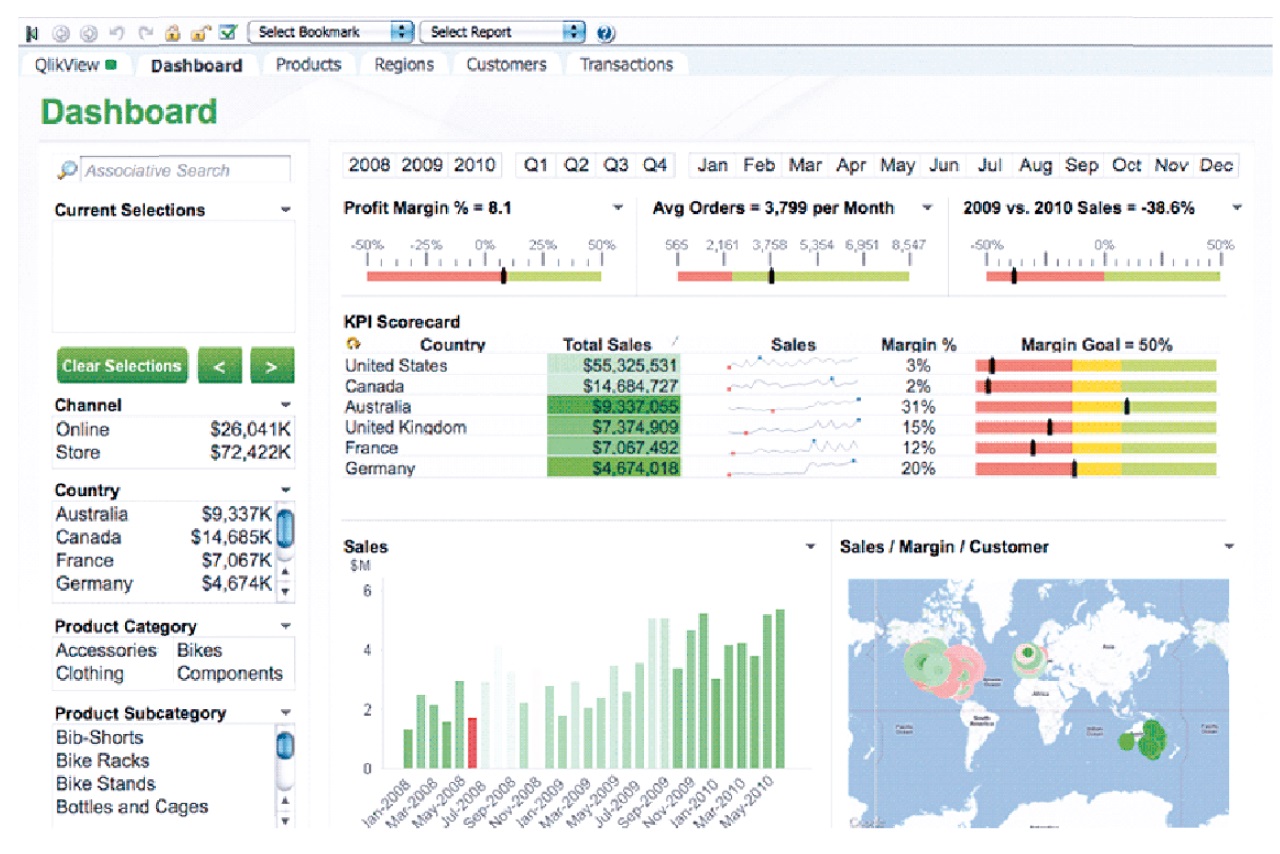


Рис. 9. Дашборд для топ-менеджмента компании

Аналитические дашборды отражают основные тенденции развития и показатели в рамках одного подразделения компании или направления деятельности, например, цепочку продаж, маркетинг или цепочку поставок. Операционные дашборды дают подробное представление об отдельных аспектах ведения бизнеса, таких как, например, объем продаж в режиме реального времени, интернет-трафик, практические случаи при работе с клиентами или время ожидания, когда вы пытаетесь дозвониться клиенту.

Довольно редко один отчет или дашборд, содержащие аналитическую информацию, служат основой для принятия важного решения. Гораздо чаще пользователи задаются вопросом: почему? Почему в северо-восточном регионе продажи упали на 30%?

*Возможно, дашборд бесполезен сам по себе, но он точно будет таковым, если его никто не использует.*

Можно выделить следующие цели презентации данных: убедить других, оформить мысль, стимулировать действие.

### Глава 8. A/B-тестирование

Во многих ситуациях, особенно новых для нас, интуиция не всегда срабатывает верно. Часто мы бываем удивлены результатом. Например, предложение о покупке в рекламном объявлении. С точки зрения количества переходов (индекс CTR), какое из них сработает лучше и насколько?

* Получите скидку 10 долл. с первой покупки. Заказывайте онлайн сейчас!
* Получите дополнительную скидку 10 долл. Заказывайте онлайн сейчас.

На практике второй вариант оказался эффективнее первого, его индекс CTR был в два раза выше. Именно поэтому качественно подготовленный эксперимент имеет такую ценность. Он переводит диалог из плоскости «Мне кажется...» в плоскость «Согласно данным...». Таким образом, это неоценимый компонент компании с управлением на основе данных.

Один из приятных аспектов A/B-тестирования в том, что вам не требуется предварительного причинно-следственного объяснения, почему что-то должно сработать. Нужно просто провести тест, изучить результаты и найти те факторы, которые обеспечивают позитивное влияние. В Amazon половина экспериментов не приносила результатов, а в Microsoft — две трети. Чтобы выиграть в долгосрочной перспективе, совсем не обязательно, чтобы срабатывал каждый эксперимент. Единственное положительное изменение способно оказать огромное влияние на итоги всей деятельности.

Как долго нужно проводить тестирование? Причина, по которой непросто дать ответ на этот вопрос, заключается в том, что существует множество факторов, которые мы пытаемся оптимизировать (рис. 10).

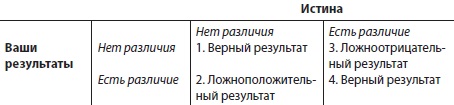


Рис. 10. Наша цель — попытаться оптимизировать вероятность верного заключения (1 или 4) и минимизировать вероятность сделать ложноположительное (2) или ложноотрицательное (3) заключение.

Используйте калькулятор размера выборки. Возьмите критерий статистической мощности = 0,8 и уровень статистической значимости = 5% (рис. 11).



Рис. 11. Калькулятор размера выборки для определения конверсии; см. [здесь](http://www.evanmiller.org/ab-testing/sample-size.html)

Чтобы избежать последствий возможных ошибок при создании нового функционала, постепенно наращивайте количество пользователей в тестовой группе. Доведите его в итоге до 50% от совокупной выборки. Когда нужно протестировать много вариантов используйте многовариантные тесты. Сервис YouTube, у которого огромная посещаемость, в 2009 году провел эксперимент, включавший тестирование 1024 комбинаций. Оптимальная комбинация привела к росту количества подписчиков сервиса на 15%.

A/B-тестирование осуществляется в рамках классического, или частотного, статистического подхода. При этом по умолчанию предполагается, что внешние условия не меняются, то есть мы в вероятностном смысле делаем выводы из распределения, но само распределение и его параметры со временем остаются неизменными.

В рамках байесовского подхода стартовой точкой служит предпосылочное убеждение. Однако, если вы ошиблись с ним, не страшно, так как со временем вы будете обновлять и изменять их по мере получения новых доказательств. Это ключевое отличие от частотного подхода: любая новая информация — просмотр, продажа или переход по ссылке — становится дополнительным доказательством, которое следует включать в базу знаний. Это итеративный подход. Более того, в его рамках не стоит вопрос «Есть ли различие между сравниваемыми вариантами?», вместо этого задают другой вопрос: «Что эффективнее: контрольный параметр или тестовый?»

Со временем система будет менять соотношение пользователей, которые получают более эффективную характеристику. Грубо говоря, тестирование может начаться с соотношения 50/50. Предположим, что тестируемая характеристика действительно очень эффективна (мы наблюдаем гораздо больше переходов), тогда система снижает пропорцию посетителей, которые пользуются контрольной характеристикой, и увеличивает пропорцию тех, кто пользуется тестируемой характеристикой. Теперь соотношение составляет 40% (контрольная группа) и 60% (тестовая).

У этого подхода два очевидных преимущества. Во-первых, нет необходимости проводить анализ, чтобы понять, какой вариант лучше, — можно просто оценить относительную пропорцию. Во-вторых, поскольку более эффективная характеристика применяется дольше, у нас есть возможность сразу же воспользоваться этим преимуществом. (В терминах статистики, нам не придется сожалеть об упущенной выгоде за период проведения эксперимента, когда у нас все еще действовала менее эффективная характеристика.)

Можно установить ограничивающий критерий: например, если эффективность тестируемой характеристики превышает 5% по сравнению с контрольной характеристикой, 100% трафика переключается на нее.

При байесовском, как и при частотном подходе положительный результат тестируемой характеристики может быть, как ее эффектом, так и делом случая. Если результат был случайным, то дальнейшее использование этой характеристики, скорее всего, приведет к снижению коэффициента CTR, и пропорция тестовой группы будет скорректирована в сторону снижения.

Байесовский подход набирает популярность, хотя и медленно. Гораздо сложнее объяснить неспециалистам принцип работы системы. В компании должна быть очень хорошо развита культура работы с данными, чтобы сотрудники могли доверять этому процессу. К сожалению, для многих компаний эта система не более чем волшебный черный ящик (см. [Скотт Хартшорн. Теорема Байеса: визуальный курс для начинающих](http://baguzin.ru/wp/?p=18517)).

### Глава 9. Принятие решений

HiPPO — аббревиатура от highest paid person’s opinion, то есть «мнение самого высокооплачиваемого сотрудника». Этот термин ввел в употребление Авинаш Кошик для обозначения концепции, полностью противоположной управлению на основе данных. Каждый из нас сталкивался с такими людьми — это эксперты с многолетним опытом (рис. 12). Им наплевать на данные, особенно когда те идут вразрез с их персональным мнением, и они всегда придерживаются своего плана, потому что знают лучше. Кроме того, «они здесь начальники».



Рис. 12. Решения должны приниматься на основе данных, а не мнения HiPPO

В целом ряде опросов, топ-менеджеры заявили, что при принятии решений следует прислушиваться к практическому опыту, а не к цифрам, часто необходимо и даже предпочтительно полагаться на интуицию и «мягкие» факторы. Почему субъективному опыту и интуиции отдают предпочтение по сравнению с объективным аналитическим подходом? Можно выделить три фактора: сами данные, корпоративная культура компании, человеческий мозг.

По моему мнению, термин «управление на основе данных» можно использовать в следующем смысле: можно стремиться к тому, чтобы данные оказывали влияние на принятие решений, то есть чтобы опытные руководители располагали адекватными данными, на которые они могут опираться в поисках ответов на правильные вопросы о том, что и насколько эффективно они делают и что, возможно, им следует делать в будущем.

Кнапп считает, что данные более полезны для практики управления, чем для принятия решений как таковых. Данные в большей степени задают вопросы и стимулируют размышления, чем указывают на конкретные варианты решения проблемы.

Серьезная проблема заключается в том, что многие топ-менеджеры не умеют работать с данными. Статистика не входит в программу МВА, и коучи ей тоже не обучают.

Есть еще один огромный барьер, препятствующий эффективному принятию решений и поддерживающий (плохую) интуицию, — наш мозг. В человеческом сознании процесс принятия решений происходит двумя основными способами: быстро, непреднамеренно, неосознанно (лауреат Нобелевской премии [Даниэль Канеман](http://baguzin.ru/wp/?p=7840) назвал это системой 1) и медленно и намеренно (система 2). Система 1 — это наше «шестое чувство», интуиция, в то время как система 2 — это наше сознание, мы пользуемся ею для тщательного обдумывания и глубокого математического анализа.

Искажения, способные негативно повлиять на наши суждения:

* «Ошибка выжившего». Мы слушаем истории успеха. Неудачники безмолвствуют.
* Предвзятость подтверждения. Человек ищет или предпочитает выбирать данные, подтверждающие то, что он уже знает.
* Эффект новизны. Мы склонны больше вспоминать недавние события и фокусироваться на них.
* Эффект «свой-чужой». Когда кто-то сообщает вам информацию, первое, что вы делаете, — оцениваете собеседника.

Я являюсь сторонником поведенческой модели Фогга. В рамках этой модели формулируется набор условий для выполнения какого-либо действия и предполагается следующее: человек должен быть достаточно мотивирован, человек должен обладать возможностью выполнить действие, на человека должен воздействовать стимул, побуждающий его выполнить действие.

Стимулируйте развитие более открытой и прозрачной корпоративной культуры, чтобы было очевидно, кто и какие решения принимает, а также к каким результатам это приводит. Повышая прозрачность самих решений и результатов этих решений с помощью презентаций, отчетов или дашбордов, вы запускаете мотивирующий фактор социального принятия.

Аналитики могут облегчить процесс принятия решений для руководителей и снизить время принятия решений, если подберут правильную форму для презентации своих выводов и рекомендаций, отразят, почему это важно, и сфокусируются на влиянии (рис. 13).



Рис. 13. Привяжите действия к результатам. Укажите действие с привязкой к конкретному результату, а ниже представьте причинно-следственное доказательство

Скотт Беркен отмечает: «Когда кто-то говорит “данные показывают”, он притворяется, что существует единственная интерпретация этих данных, но это далеко не так. Подобное ложное убеждение мешает задавать важные вопросы, например, “Можно ли на основании этих же данных выстроить альтернативную и в равной степени убедительную гипотезу, ведущую к другому заключению?”» Основное в этом процессе — начать с правильных вопросов и сконцентрироваться на вопросе и решении, а не на данных.

### Глава 10. Корпоративная культура на основе данных

Корпоративная культура определяет, кто имеет доступ к данным, какие данные можно распространять, какие вложения будут сделаны в развитие сотрудников и в инструменты. Корпоративная культура также определяет, HiPPO или данные будут влиять на решение.

Руководители должны думать о том, как поощрять сотрудников, распространяющих данные, как стимулировать сотрудников и отделы, развивающие и поддерживающие открытые, точные и доступные для использования данные и аналитику.  
*Дженнифер Кобб*

Компании с управлением на основе данных отличаются большей открытостью и прозрачностью, данные демократизированы и доступны многим сотрудникам. Компания должна быть уверена, что не произойдет утечки информации к конкурентам, что ее данные не будут использовать в различных политических интригах, а станут исключительно способствовать росту и развитию бизнеса. Многие решения, особенно касающиеся операционной деятельности, можно делегировать на места, если обеспечен доступ к нужным данным.

Специалисты по аналитической работе должны пройти обучение по планированию экспериментов, развитию навыков критического мышления, презентации данных, применению инструментов бизнес-аналитики и статистики. Однако, чтобы вся компания стала ориентированной на данные, этот набор навыков, а также подход, опирающийся на факты и доказательства, должен быть внедрен на более широком уровне. Руководители и другие лица, ответственные за принятие решений, также должны быть компетентны в области работы с данными.

В компании с управлением на основе данных должна поощряться атмосфера здоровой любознательности. Нужно стимулировать конструктивные обсуждения, в ходе которых участники запрашивают дополнительную информацию, подвергают сомнениям предположения, обсуждают результаты тестирования.

Поощряйте несогласие. Если приоритет для вас — создание компании с управлением на основе данных, то вы должны принять наличие определенного уровня несогласия.

Предпринимая определенные действия на перспективу, такие как построение прогнозных моделей, важно не забывать о петле обратной связи, в рамках которой вы проводите регулярный обзор результатов, изучаете отдельные случаи (так называемый анализ ошибок), выясняете, где вы могли бы действовать эффективнее (рис. 14).



Рис. 14. Петля обратной связи: планирование, измерение, выводы и повтор

### Глава 11. Топ-менеджмент компании с управлением на основе данных

В последнее десятилетие наблюдается определенный сдвиг, так как все больше компаний уже не ассоциируют данные только с расходами и обязательствами, но оценивают их как актив. В компании должен быть топ-менеджер, отвечающий за данные.

Chief Data Officer (CDO) преимущественно встречаются в следующих областях: банковский и финансовый сектор (40% от общего количества), государственное управление, здравоохранение. Все эти области подчиняются строгому регулированию. Основная обязанность CDO заключается в стратегическом использовании данных. CDO должен менять корпоративную культуру, оказывать влияние на других (как на топ-менеджеров, так и на простых сотрудников), чтобы изменить их отношение к использованию данных. Ему необходимо способствовать созданию в компании открытой корпоративной культуры, основанной на обмене данными, а также демократизировать данные, делая доступными их источники, что включает ликвидацию обособленных закрытых хранилищ данных.

Желательно, чтобы CDO отчитывался в своих действиях перед СЕО и занимал равное положение с другими топ-менеджерами. CDO смогут принести компании больше пользы, если будут подчиняться людям, отвечающим за коммерческую составляющую, а не за техническую.

Думаю, роль CDO, как она понимается сейчас, то есть определенного топ-менеджера, отвечающего за преобразование данных в актив для повседневного использования, должна исчезнуть в ближайшие пару десятков лет, поскольку использование данных и проведение анализа станут неотъемлемой частью ведения бизнеса.

Chief Analytics Officer (САО) сосредоточен на стратегическом использовании данных, то есть, на их анализе. Роль САО — повысить эффективность методов работы и корпоративной культуры на основе данных. Один из подходов, который оправдал себя, по крайней мере, с одной из компаний, производящих товары широкого потребления, состоял в привлечении CEO как силы воздействия. По его указанию руководитель направления по работе с данными и руководитель бизнес-подразделения, который имел весьма слабое представление о больших данных, должны были совместно разработать план для максимальной реализации потенциала аналитических данных. «В результате этого сотрудничества, объединившего эксперта в области данных и опытного руководителя в сфере работы с клиентами, аналитические цели, обозначенные в плане работы, были сконцентрированы на действительно важных и актуальных бизнес-решениях.

### Глава 12. Вопросы конфиденциальности, этики и риска

По моему мнению, больше данных, больше доступа, больше аналитики означают большую власть и больше риска. Основополагающий принцип, которого компании должны придерживаться в своей деятельности, – эмпатия.

Один из примеров использования конфиденциальной информации, когда все происходило в рамках закона, но привело к неблагоприятным последствиям, связан с компанией Target. В статье, опубликованной в New York Times и вызвавшей оживленное обсуждение среди специалистов по работе с данными, журналист [Чарльз Дахигг](http://baguzin.ru/wp/?p=10673) рассказывает, как специалисты по маркетингу компании Target попросили одного из аналитиков компании, Эндрю Пола, определить группу покупательниц, которые были беременны, чтобы знать эту информацию до того, как появятся официальные сведения о рождении ребенка. Маркетологи предположили, что, если выделить эту категорию женщин достаточно рано, есть больше шансов заинтересовать их купонами и создать базу лояльных клиентов.

Примерно через год после того, как Пол разработал свою прогнозную модель, в офис компании Target в Миннеаполисе вошел мужчина и потребовал встречи с менеджером. Как рассказал один из сотрудников компании, присутствовавший при разговоре, мужчина, сжимавший в руке пачку купонов, был в бешенстве.

«Моя дочь получила это по почте, — заявил он. — Она еще учится в старшей школе, а вы посылаете ей купоны на покупку детской одежды и кроватки. Вы что, занимаетесь пропагандой подростковой беременности?»

Менеджер понятия не имел, о чем говорит этот мужчина. Он взглянул на буклет. Никаких сомнений: буклет был адресован дочери этого мужчины, и содержал рекламу детской одежды и мебели, а еще фотографии розовощеких младенцев. Менеджер принес свои извинения, а затем позвонил через несколько дней, чтобы извиниться еще раз. Отец на другом конце провода был явно смущен. «Я поговорил с дочерью, — объяснил он. — Кажется, в моем доме происходит нечто, о чем я не имел ни малейшего представления. Она должна родить в августе. Это я должен принести вам извинения».

Социальная сеть Facebook постоянно испытывает разногласия со своими пользователями, часто выходя за рамки того, какой информацией можно делиться и с кем. Изменения в настройках конфиденциальности по умолчанию для различных аспектов на сайте — просто поразительны. Сравните два графика (рис. 15). Первый показывает настройки по умолчанию в 2005 году, а второй — те же самые настройки через пять лет в 2010 году.

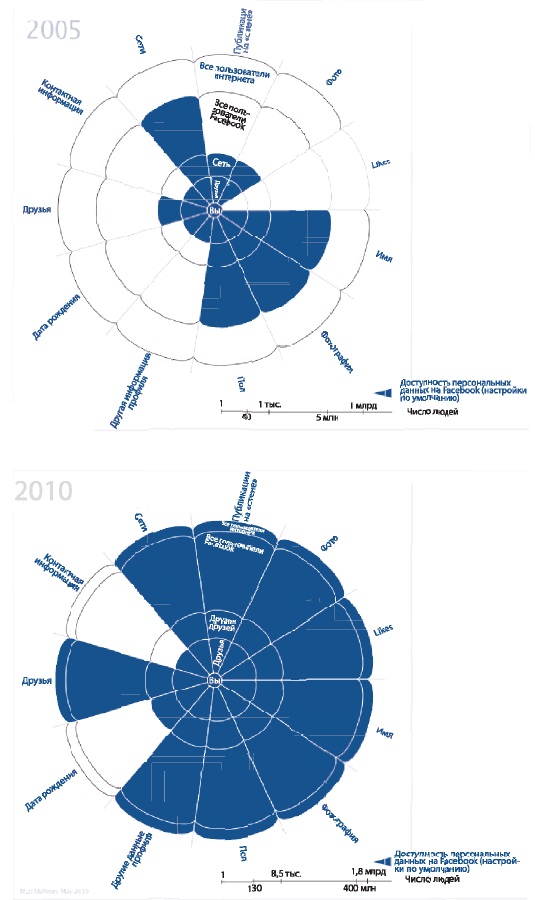


Рис. 15. Изменения в настройках конфиденциальности по умолчанию Facebook

В компаниях с активным использованием данных всегда будет наблюдаться некоторое здоровое напряжение между разными командами: так, например, аналитики всегда будут стремиться создавать самые современные продукты с использованием данных, а более консервативные юристы — минимизировать риски для компании. В то время как законодательные ограничения непреложны, существует обширная серая зона, деятельность в которой не нарушает закон, но может вызывать сомнения с морально-этической точки зрения. Компания должна уважать своих пользователей и разработать руководство, что считать приемлемым и неприемлемым использованием данных.

### Заключение

Итак, концепция управления на основе данных охватывает всю аналитическую цепочку ценности и всю структуру компании (рис. 16).



Рис. 16. Обзор компонентов, из которых складывается компания с управлением на основе данных

### Дополнительная литература

Дэвенпорт Т., Харрис Дж. [Аналитика как конкурентное преимущество](http://baguzin.ru/wp/?p=2923). Новая наука побеждать. – М.: BestBusinessBooks, 2010.

Сигель Э. Просчитать будущее. Кто кликнет, купит, соврет или умрет. – М.: Альпина Паблишер, 2014.

Сильвер Н. [Сигнал и шум](http://baguzin.ru/wp/?p=15853). Почему одни прогнозы сбываются, а другие — нет. – М.: Азбука-Аттикус: КоЛибри, 2015.

Канеман Д. [Думай медленно... Решай быстро](http://baguzin.ru/wp/?p=7840). – М.: ACT, 2016.