



Аналити- ческая культура

от сбора данных до бизнес-результатов

[Купить книгу на сайте kniga.biz](http://kniga.biz)

Карл Андерсон

Эту книгу хорошо дополняют:

Маркетинг, основанный на данных
Марк Джейфри

Управление на основе данных
Тим Филлипс

О чем говорят цифры
Том Дэвенпорт и Ким Джин Хо

Большие данные
Виктор Майер-Шенбергер и Кеннет Кукиер

Верховный алгоритм
Педро Домингос

[Купить книгу на сайте kniga.biz.ua >>>](http://kniga.biz.ua)

Carl Anderson

Creating a Data-Driven Organization

O'Reilly
2015

[Купить книгу на сайте kniga.biz.ua >>>](http://kniga.biz.ua)

Карл Андерсон

Аналитическая культура

От сбора данных до бизнес-результатов

Перевод с английского Юлии Константиновой

Москва
«Манн, Иванов и Фербер»
2017

[Купить книгу на сайте kniga.biz.ua >>>](http://kniga.biz.ua)

УДК 658.5:005
ББК 65.291.213
A65

Научный редактор Руслан Салахиев

Издано с разрешения O'Reilly Media, Inc.

На русском языке публикуется впервые

Андерсон, Карл

A65 Аналитическая культура. От сбора данных до бизнес-результатов / Карл Андерсон ; пер. с англ. Юлии Константиновой ; [науч. ред. Руслан Салахиев]. — М. : Манн, Иванов и Фербер, 2017. — 336 с.

ISBN 978-5-00100-781-4

Это практическое пошаговое руководство по внедрению в вашей организации управления на основе данных. Карл Андерсон, директор по аналитике в компании Warby Parker, провел интервью с ведущими аналитиками и учеными и собрал кейсы, которые и легли в основу данной книги. Вы узнаете, какие процессы следует ввести на всех уровнях и как именно это сделать, с какими трудностями можно столкнуться на этом пути и как их преодолеть. Автор рассказывает об аналитической цепочке ценностей, которая поможет принимать правильные решения и достигать лучших бизнес-результатов.

Книга будет интересна CEO и владельцам бизнеса, менеджерам, аналитикам.

УДК 658.5:005
ББК 65.291.213

Все права защищены.

*Никакая часть данной книги не может
быть воспроизведена в какой бы то ни было форме
без письменного разрешения владельцев авторских прав.*

ISBN 978-5-00100-781-4

© 2017 Mann, Ivanov and Ferber
Authorized Russian translation of the English edition
of Creating a Data-Driven Organization,
ISBN 9781491916919 © 2015 Carl Anderson, published
by O'Reilly Media, Inc.
This translation is published and sold by permission of O'Reilly
Media, Inc., which owns or controls all rights to publish
and sell the same.
© Перевод на русский язык, издание на русском языке,
оформление. ООО «Манн, Иванов и Фербер», 2017

[Купить книгу на сайте kniga.biz.ua >>>](http://kniga.biz.ua)

Оглавление

Введение	11
Краткий обзор	11
Для кого эта книга?	14
Структура глав	14
Условные обозначения	16
Глава 1. Что значит «на основе данных»?	17
Сбор данных	17
Доступ к данным	18
Составление отчетности	21
Оповещения	22
От отчетов и оповещений к анализу	23
Критерии управления на основе данных	26
Зрелость аналитических данных	29
Краткий обзор	35
Глава 2. Качество данных	37
Аспекты качества данных	38
Происхождение данных	57
Качество данных как совместная ответственность	58
Глава 3. Сбор данных	62
Собирайте все что можно	62
Расстановка приоритетов при выборе источников данных	66
Установление взаимосвязи	69
Сбор данных	71
Покупка данных	74

Сколько стоит набор данных?	78
Хранение данных	81
Глава 4. Специалисты по аналитике	83
Типы специалистов по аналитике	83
Аналитика — это командный спорт	91
Навыки и качества	95
Еще один инструмент	98
Организация работы аналитиков в компании	104
Глава 5. Анализ данных	111
Что такое анализ данных?	112
Виды анализа данных	114
Глава 6. Разработка показателей	140
Разработка показателей	141
Ключевые показатели эффективности	149
Глава 7. Сторителлинг на основе данных	157
Сторителлинг	158
Первые шаги	161
Визуализация данных	165
Подача данных	176
Основные выводы	184
Глава 8. А/В-тестирование	187
Почему А/В-тестирование?	191
Практические рекомендации по А/В-тестированию	192
Другие подходы	205
Влияние на корпоративную культуру	210
Глава 9. Принятие решений	212
Как принимают решения?	214
Что осложняет процесс принятия решения?	219

Решения	231
Заключение	239
Глава 10. Корпоративная культура на основе данных	242
Открытость и доверие	243
Повышение квалификации в области работы с данными	247
Сначала цели	250
Задавайте вопросы	251
Итерации и обучение	253
Как противостоять HiPPO	255
Руководство на основе данных	256
Глава 11. Топ-менеджмент компании с управлением на основе данных	259
Chief Data Officer	261
Chief Analytics Officer	273
Заключение	279
Глава 12. Вопросы конфиденциальности, этики и риска	282
Уважайте конфиденциальность	284
Практикуйте эмпатию	290
Качество данных	294
Безопасность	296
Обеспечение исполнения	299
Заключение	300
Заключение	302
Дополнительная литература	309
Аналитика	309
Анализ данных	309
Принятие решений	309
Визуализация данных	310
A/B-тестирование	310

Приложение А. О необоснованной эффективности данных: почему больше данных лучше?	311
Проблемы типа «ближайший сосед»	312
Проблемы относительной частотности	315
Проблемы оценки одномерного распределения	316
Проблемы многофакторности	317
Приложение В. Заявление о видении	319
Ценность	321
Реализация	322
Благодарности	324
Об авторе	326
Колофон	327

Введение

Краткий обзор

Эта книга посвящена двум основным вопросам:

- 1) что означает для компании управление на основе данных?
- 2) как компания может к нему прийти?

Многие компании считают, что, если они генерируют множество отчетов или у них много дашбордов, значит, они относятся к категории компаний с управлением на основе данных. Хотя эти виды деятельности и составляют часть того, чем занимается компания, обычно они ретроспективны, то есть часто лишь представляют прошлые или настоящие факты без обеспечения достаточного контекста, без объяснения причинно-следственных связей, а также без рекомендаций, какие шаги предпринять. Иными словами, они фиксируют произошедшее, но ничего не предписывают. В этом отношении их потенциал роста ограничен.

В противовес следует рассматривать типы перспективного анализа, такие как прогнозные модели, которые способствуют оптимизации расходов на рекламу, пополнению цепочки поставок или снижению оттока покупателей. Они отвечают на вопросы «кто», «что», «когда», «почему» и «где». На основе моделей люди дают рекомендации, делают прогнозы и интерпретируют полученные данные. Часто они становятся ключевыми факторами роста в организациях с управлением на основе данных. Сформулированные на основе данных выводы и рекомендации, если их правильно использовать, оказывают огромное потенциальное влияние на эффективность деятельности компаний.

Однако для получения подобных выводов требуется, чтобы были собраны правильные, заслуживающие доверия данные, анализ был проведен качественно, выводы учитывались при принятии решений, а решения подразумевали конкретные действия, чтобы потенциал был полностью реализован. Уф! Я называю эту последовательность от сбора данных до конечного результата *аналитической цепочкой ценности*.

Последний шаг в этой цепочке чрезвычайно важен. Аналитику нельзя считать основанной на данных, если полученная информация не учитывается при принятии решений и не вызывает последующих действий. Если данные игнорируются, а большой босс делает что пожелает, сбор этих данных не имеет смысла. Управление на основе данных осуществляется в компании при наличии правильных процессов и корпоративной культуры, чтобы дорабатывать или стимулировать важные деловые решения с учетом проведенного анализа данных, который таким образом оказывает непосредственное влияние на развитие бизнеса.

Ключевую роль играет создание соответствующей корпоративной культуры. Это многосторонняя программа, включающая качество данных и обмен информацией, прием на работу и обучение аналитиков, коммуникацию, аналитическую организационную структуру, разработку показателей, А/В-тестирование¹, процессы принятия решений и многое другое. Эта книга поможет пролить свет на все эти понятия благодаря доступным объяснениям и наглядным примерам из целого ряда производственных отраслей. Кроме того, здесь приводятся практические советы и рекомендации от лидеров в области анализа и обработки данных. Надеюсь, эта книга вдохновит читателей на то, чтобы переориентировать свою деятельность и начать руководствоваться данными.

Более того, на протяжении всей книги подчеркивается важная роль, которая отводится самим разным специалистам в области обработки и анализа данных. Я убежден, что компанию с управлением на основе данных и соответствующую корпоративную культуру можно и нужно развивать не только сверху вниз — от руководства на места, — но и снизу вверх. Как отметил на форуме 2014 года Chief Data Officer Executive Forum руководитель направления по анализу и обработке данных

¹ Метод маркетингового исследования, суть которого заключается в том, что контрольная группа элементов сравнивается с набором тестовых групп, в которых один или несколько показателей были изменены, для того чтобы выяснить, какие из изменений улучшают целевой показатель. *Прим. ред.*

компании Trulia Тодд Холлоуэй, «лучшие идеи подают сотрудники, наиболее тесно работающие с данными». Они не только напрямую имеют дело с источниками данных и способны оценить их качество и повлиять на него, не только понимают, как лучше всего их дополнить, но также «часто подают хорошие идеи по поводу товаров». Кроме того, они могут помочь повысить уровень знаний других сотрудников компании в этой области. Частично это происходит благодаря тому, что они развиваются свои навыки и активно применяют их для качественного выполнения работы. Другая причина в том, что у них лучше развито предпринимательское мышление: они умеют задавать правильные вопросы и формулировать бизнес-проблемы, а затем убеждаться в своих выводах и рекомендациях тех, от кого зависит принятие решения, предлагая им веское обоснование, какое влияние на бизнес способны оказать эти выводы и рекомендации.

А влияние и выгоды могут быть весьма заметными. Согласно результатам одного из отчетов¹, в котором контролировались и другие факторы, в компаниях с управлением на основе данных производительность была на 5–6% выше, чем в тех, что не практикуют подобное управление. К тому же в компаниях первой категории были выше показатель использования ресурсов, коэффициент рентабельности капитала и рыночная стоимость. Согласно данным другого отчета², возврат на каждый вложенный в проведение аналитики 1 долл. составляет 13,01 долл. Управление на основе данных окупается!

Ориентацию на использование данных можно представить в виде непрерывного процесса: компания всегда может повысить свой уровень управления на основе данных, улучшить качество собираемых данных и аналитического процесса, провести больше тестирований. Более того, всегда можно усовершенствовать качество процесса принятия решений. В этой книге мы обсудим отличительные черты эффективных компаний с управлением на основе данных. Мы остановимся на инфраструктуре, навыках, корпоративной культуре, необходимых для создания компаний, где к данным относятся как к основному активу и используют их для принятия бизнес-решений. Кроме того, мы

¹ Brynjolfsson E., Hitt L. M. and Kim H. H. Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? Social Science Research Network (2011). URL: http://ebusiness.mit.edu/research/papers/2011.12_Brynjolfsson_Hitt_Kim_Strength%20in%20Numbers_302.pdf.

² Nucleus Research. Analytics pays back \$13.01 for every dollar spent. O204 (Boston, MA: Nucleus Research, 2014), 5. URL: <http://nucleusresearch.com/research/single/analytics-pays-back-13-01-for-every-dollar-spent/>.

рассмотрим некоторые примеры поведения, которое, наоборот, мешает бизнесу максимально эффективно использовать получаемые данные.

Таким образом, цель этой книги — вдохновить специалистов по анализу и обработке данных в компаниях эффективно выполнять свои функции, время от времени делать паузу, чтобы ответить на вопросы, максимально ли использует компания свои данные и можно ли делать это еще эффективнее. Еще одна цель — стимулировать обсуждение: для каких еще целей возможно применение этого ключевого ресурса. Никогда не рано думать об этом. Основатели компаний и руководство высшего звена должны постараться внедрить принципы управления на основе данных на самых ранних этапах развития организации. Давайте узнаем больше о том, что эти принципы собой представляют.

Для кого эта книга?

Информация, здесь изложенная, поможет разработать программу внутренней аналитики и управлять ею: принимать решения, какие данные собирать и хранить, как их получать и интерпретировать, и самое важное — как действовать на их основе.

Неважно, единственный ли вы специалист по анализу и обработке данных в стартапе (и притом вынуждены выполнять еще с десяток других функций) или руководитель отдела с кучей подчиненных в зрелой компании. Если вы работаете с данными и стремитесь действовать быстрее, рациональнее и эффективнее, эта книга поможет создать не просто аналитическую программу, а соответствующую корпоративную культуру.

Структура глав

Структура книги соответствует этапам создания цепочки аналитической ценности. Первые главы посвящены непосредственно данным, в частности выбору правильных источников, обеспечению качества и достоверности. Следующий шаг в этой цепочке — анализ данных. Для качественного выполнения анализа, результаты которого можно будет эффективно использовать в дальнейшей работе, нужны профессионалы, владеющие определенными навыками и инструментами. Для обозначения этой группы сотрудников намеренно используется общий

термин «специалисты по аналитической работе», который объединяет сотрудников, занимающихся сбором, обработкой, анализом данных. Это сделано на основании убеждения, что любой член команды — от младшего аналитика без опыта работы до суперзвезды в области анализа данных — вносит свою лепту в общее дело. Мы подробнее остановимся на том, какими компетенциями должен обладать хороший аналитик, как можно развивать профессиональные навыки в этой области, а также на организационных аспектах — как помочь специалисту по аналитической работе стать частью команды или подразделения. Следующие главы посвящены непосредственно аналитической работе: выполнению анализа, разработке показателей, А/В-тестированию и рассказыванию истории. Затем мы перейдем к следующему этапу в цепочке аналитической ценности — принятию решений на основе результатов анализа. Мы рассмотрим, что может затруднять процесс принятия решения и как с этим бороться.

На протяжении всей книги прослеживается основная мысль: суть процесса управления компанией на основе данных не сводится к данным как таковым или к обладанию самым современным набором инструментов по работе с большими данными. Самое важное в этом — корпоративная культура. 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. По этой причине в главе, посвященной корпоративной культуре, мы объединим все уроки, извлеченные на разных этапах цепочки аналитической ценности. В одной из последних глав обсудим роль двух относительно новых позиций в высшем руководстве компаний: CDO (Chief Data Officer, директор по управлению данными) или CAO (Chief Analytics Officer, директор по аналитике). Тем не менее рядовые сотрудники тоже в значительной мере влияют на формирование корпоративной культуры организации, поэтому на протяжении книги мы будем напрямую обращаться к специалистам по работе с данными, подчеркивая, что именно они способны сделать для повышения своего влияния на эффективность деятельности компании. В компании, для которой управление на основе данных не просто модная тенденция, сотрудники на всех уровнях уделяют большое внимание качеству данных и их оптимальному использованию при принятии взвешенных решений и для повышения конкурентного преимущества компании.

Условные обозначения

В книге используются следующие условные обозначения.

Выделение курсивом

Применяется для обозначения новых терминов, адресов сайтов (URL), адресов электронной почты, имен файлов и расширений файлов.

Моноширинный шрифт

Применяется для обозначения программных элементов, таких как переменные, названия функций, базы данных, типы данных, переменные окружения, утверждения и ключевые слова.

Моноширинный шрифт с полужирным выделением

Применяется для обозначения команд или другого текста, который должен внести пользователь.

Моноширинный шрифт с курсивом

Применяется для обозначения текста, который нужно заменить переменными пользователя или переменными, которые определяются контекстом.



Этот элемент обозначает совет или рекомендацию.



Этот элемент обозначает общую информацию.

ГЛАВА 1

Что значит «на основе данных»?

Без данных вы просто еще один человек с собственным мнением.

Уильям Эдвардс Деминг¹

Управление на основе данных подразумевает формирование инструментов, способностей и, что самое важное, *корпоративной культуры*, которая опирается на данные. В этой главе мы рассмотрим, что отличает компанию с управлением на основе данных. Начнем с базовых требований к их сбору и доступности. Затем остановимся подробнее на весьма важном отличии — подготовке отчетов и получении оповещений в противовес процессу анализа. Существует много различных типов перспективного анализа, отличающихся по степени сложности. Мы уделим некоторое время изучению этих типов с точки зрения их «уровня аналитики» и «аналитической зрелости», а также обсудим основные признаки «аналитически зрелой» организации. Какой она должна быть?

Начнем с ответа на первый вопрос: что означает для компании управление на основе данных?

Сбор данных

Давайте сразу озвучим несколько очевидных требований.

Требование № 1: в компании должен осуществляться сбор данных.

Несомненно, данные — ключевой компонент. При этом речь идет не о любых данных, а о *правильных*. Необходимо, чтобы набор данных

¹ Уильям Эдвардс Деминг (William Edwards Deming, 1900–1993) — американский ученый, статистик и консультант по менеджменту. Создатель теории менеджмента, основанной на предложенной им же теории глубинных знаний. *Прим. перев.*

соответствовал вопросу, который требуется решить. Помимо этого, данные должны быть своевременными, точными, чистыми, объективными, и, что важнее всего, они должны заслуживать доверия.

Это не так-то просто. Данные никогда не бывают настолько чистыми, как вам кажется. Они могут быть предвзятыми, что может повлиять на результат анализа, а очистка данных может стать трудоемким и дорогим процессом, требующим времени. Часто приходится слышать, что специалисты по работе с данными до 80% времени тратят на их сбор, очистку и подготовку и только 20% — на построение моделей, процесс анализа, визуализацию и формулировку заключений на основе этих данных¹. Как показывает опыт, это вполне вероятно.

В следующей главе мы поговорим о качестве данных подробнее.

Даже если у вас есть действительно качественные данные и даже если у вас много качественных данных, это означает только то, что вы обладаете этими данными, но не то, что в вашей компании действует управление на основе данных. Некоторые люди, особенно специалисты организаций, предоставляющих услуги по работе с большими данными, называют большие данные практически панацеей: если собирать абсолютно всё, где-то должен попасться алмаз (или крупинки золота, или искомая иголка, или любая другая метафора) и компания станет успешной. Горькая правда в том, что одних только данных недостаточно. Небольшое количество чистой, достоверной информации может быть гораздо более ценно, чем петабайты мусора.

Доступ к данным

Требование № 2: данные должны быть общедоступными.

Наличие точных и своевременных данных по теме еще не делает управление в вашей компании управлением на основе данных. Данные также должны отвечать еще ряду требований.

Данные могут быть объединены

Их формат должен при необходимости допускать объединение с другими данными компании. Варианты могут быть разные: реляционные базы данных, хранилища NoSQL или Hadoop. Используйте инструмент, который отвечает вашим конкретным требованиям. Например, в течение длительного времени финансовые

¹ См., например: <http://bit.ly/nyt-janitor> и <http://bit.ly/im-data-sci>.

аналитики в компании Warby Parker использовали Excel для вычисления основных показателей, которые они предоставляли высшему руководству. Они собирали огромное количество сырых данных из разных источников и запускали функцию ВПР (VLOOKUP — функцию в Excel для поиска перекрестных ссылок в данных), чтобы объединить весь массив данных и взглянуть на них в перспективе. Изначально это работало, но по мере того как базы данных по клиентам и продажам быстро росли и информации становилось все больше, объем файла в Excel начал приближаться к 300 МБ, загрузка оперативной памяти компьютеров была максимальной, а обработка файла с помощью функции ВПР начала занимать до десяти часов и больше, при этом программа периодически зависала, и ее приходилось запускать заново. Специалисты компании применяли этот инструмент и подход так долго, как могли, но если когда-то Excel была вполне удобным инструментом, то динамичный рост компаний изменил ситуацию. Механика получения этих данных превратилась для аналитиков в «пожиратель времени» и источник стресса: они никогда не знали, получат ли необходимые им данные или через десять часов им вновь придется перезапускать функцию ВПР. Условно говоря, из специалистов по анализу данных они превратились в специалистов Microsoft по сбору данных. Моя команда помогла перенести весь массив информации в реляционную базу данных в MySQL. Мы написали запросы для обработки данных для аналитиков, чтобы они могли сосредоточиться на анализе, выявлении трендов и презентации этих данных, что было гораздо более эффективным использованием их рабочего времени. Теперь, когда в их распоряжении более эффективные инструменты и больше времени, они способны проводить более глубокий анализ.

Данные можно использовать совместно

Внутри организации следует развивать культуру обмена данными, чтобы была возможность их сопоставлять и объединять, например связать историю поисковых запросов пользователя и историю осуществленных им покупок. Представим ситуацию: пациента доставили в отделение экстренной медицинской помощи, где ему оказали первую помощь, а затем выписали, и теперь ему необходимо обратиться за амбулаторным лечением и провести обследования. Очевидно, что качество обслуживания и, что важнее, качество лечения пострадают, если между этими медицинскими учреждениями не будет организован обмен информацией:

когда и по какой причине пациент обратился за медицинской помощью, какое лечение ему было оказано и так далее. С точки зрения представителей здравоохранения, невозможно проанализировать или улучшить процесс в отсутствие связной и четкой картины потока пациентов, процесса диагностики и полных данных наблюдения за этими пациентами за длительный срок. Таким образом, разрозненные данные всегда стараются охватить все, что возможно. Когда больший объем данных доступен для большего количества частей системы, целое всегда бывает лучше суммы частей.

Доступны по запросу

Необходимы адекватные инструменты для работы с данными и предоставления информации по запросу. В процессе анализа и составления отчетности огромный объем сырых данных необходимо отфильтровать, сгруппировать и объединить в небольшие наборы высокоуровневых показателей, чтобы обеспечить понимание того, что происходит в бизнесе. Например, мне нужно увидеть тренд или понять разницу между сегментами покупателей. У специалистов по работе с данными должны быть инструменты, позволяющие сделать это относительно просто.

(Все эти аспекты мы подробнее проанализируем в следующих главах.)

Итак, теперь у нас есть данные и доступ к ним. Достаточно ли этого? Нет, пока недостаточно. Нужны квалифицированные специалисты, которые смогут работать с этими данными. И здесь важны не только механизмы сортировки и систематизации данных, например посредством языка запросов или макросов Excel, но, главным образом, специалисты, которые будут выбирать соответствующие показатели (подробнее об этом в главе 6). К этим показателям могут относиться уровень повторной подписки (для таких сервисов, как Netflix или Wall Street Journal), долгосрочные показатели ценности или показатели роста, но в любом случае кто-то должен решать, какие именно это будут показатели, и кто-то должен создать процесс их получения.

Таким образом, человеческий фактор в управлении компанией на основе данных — важнейший: необходимы люди, способные задавать правильные вопросы, люди с необходимыми навыками для получения нужных данных и показателей, люди, использующие данные для планирования следующих шагов. Иными словами, одни лишь данные мало чем помогут компании.

Составление отчетности

Предположим, у вас есть аналитическая группа с доступом к точным данным. Эта группа получает данные по объему продаж и гордо рапортует о росте портфеля заказов компании на 5,2% с апреля по май (рис. 1.1).

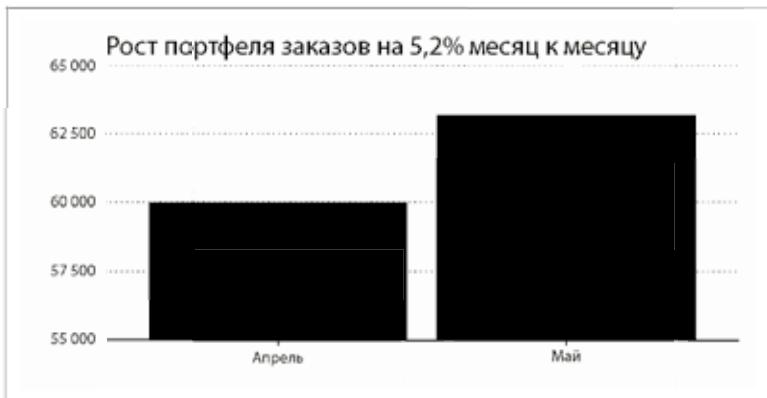


Рис. 1.1. Рост уровня продаж на 5,2% месяц к месяцу!

Кажется, что в компании осуществляется управление на основе данных. Однако этого по-прежнему недостаточно. Разумеется, хорошо, что специалисты отслеживают данные по продажам. Генерального и финансового директоров эти цифры, несомненно, заинтересуют. И тем не менее — о чём на самом деле говорит показатель 5,2%? Практически ни о чём. Возможны самые разные причины роста объема продаж компании.

- Предположим, вы продаете сезонный товар, например купальные костюмы. Может быть, рост в 5,2% — это гораздо ниже, чем обычно. Может быть, в предыдущие годы рост объема продаж в мае составлял более 7%, а в этом году он ниже обычного.
- Возможно, директор по маркетингу потратил кучу денег на национальную кампанию по повышению узнаваемости бренда. Какой процент роста из этих 5,2% обусловлен проведенной кампанией? Насколько эффективным оказалось подобное вложение средств?
- Может быть, генерального директора вашей компании пригласили поучаствовать в телешоу *Good Morning America*¹, или ваш

¹ *Good Morning America* («Доброе утро, Америка») — американское телевизионное шоу, которое транслируется по утрам на канале ABC. Выходит в эфир с 1975 г. Прим. ред.

продукт был упомянут в Techcrunch¹, или ваше видео стало «вирусным», и это послужило фактором роста продаж. То есть причина — какое-то конкретное событие, способное обеспечить временный или устойчивый рост.

- Возможно, продажи за месяц характеризуются низким объемом и широким ассортиментом. Возможно, это было лишь удачным стечением обстоятельств, а общая тенденция — *нисходящая*. (Если вы когда-нибудь пробовали играть на бирже, то понимаете, о чем речь.)
- Может быть, ошибка в самих данных. Если уровень продаж относительно стабилен и вы видите резкий скачок без каких-либо предпосылок к тому, возможно, все дело в качестве данных.

Все это возможные объяснения. Цифра в отчете представляет собой именно это — числовой показатель без контекста.

«По мере того как компании становятся все более крупными и сложноорганизованными, руководство все меньше зависит от личного опыта и все больше — от обработанных данных». — Джон Гарднер

Джон Маэда (@johnmaeda)
16 августа 2014 года²

Оповещения

Дзынь, дзынь, дзынь! Загрузка ЦП на сервере приложений № 14 за последние пять минут превысила 98%.

Оповещения фактически представляют собой отчеты о том, что происходит в настоящее время. Обычно они обеспечивают конкретные данные в рамках тщательно разработанных показателей. К сожалению, как и отчеты, они не сообщают, почему наблюдается рост загрузки ЦП, и не говорят, что следует предпринять прямо сейчас для решения проблемы, то есть они не дают важного контекста.

Нет причинно-следственного объяснения. Это момент, когда системные администраторы или инженеры по эксплуатации начинают изучать журнал регистрации событий, чтобы понять, что происходит,

¹ Techcrunch — сайт и одноименная компания, блог, описывающий продукты, стартапы и другие сайты, основанный Майклом Аррингтоном в 2005 г. *Прим. ред.*

² URL: <http://bit.ly/maeda-gardner>

почему и как это исправить: сделать откат назад, раскрутить дополнительные серверы, перенастроить выравниватель нагрузки и так далее.

На рис. 1.2 приведен пример загрузки сервера. С небольшими вариациями на протяжении дня очередь выполнения составляет 0,5 или меньше. В час ночи загрузка начинает расти и за 30 минут увеличивается до пяти и выше, в десять раз по сравнению с «нормой». Ситуация нестандартная. Что происходит? Возможно, требуется вмешательство? Но что нужно сделать?

В данном случае это всего лишь еженедельное резервное копирование данных. Оно осуществляется каждый четверг в час ночи. Это абсолютно штатная ситуация. Мы имеем четкие данные и ясно представленные показатели. Нет только контекста: что причина повышения загрузки — резервное копирование данных, что оно ожидаемо и запланированно происходит в определенное время и что сервер спокойно справляется

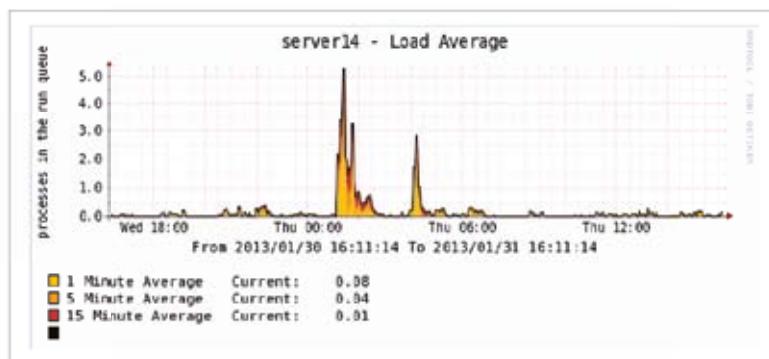


Рис. 1.2. Пример загрузки сервера

Источник: <https://blog.bigwetfish.hosting/we-got-your-back/>

От отчетов и оповещений к анализу

Составление отчетов и получение оповещений — необходимые факторы управления на основе данных, но этого недостаточно. Хотя не стоит недооценивать важность двух этих видов деятельности. Подготовка отчетов чрезвычайно важна для управления на основе данных: компания не сможет быть эффективной без этого элемента. А вот обратное не обязательно верно: существует множество организаций, сосредоточенных на отчетности, у которых может не быть качественного

анализа. Составление отчетности может быть вызвано официальными требованиями, например необходимостью исполнения закона Сарбейнза–Оксли¹ и подготовки отчетов о прибыли для акционеров, а не внутренним стремлением к повышению эффективности бизнеса.

Данные отчетов информируют, что произошло в прошлом. Кроме того, они могут быть тем фундаментом, с которого можно наблюдать за изменениями и тенденциями. Они могут представлять интерес для инвесторов и акционеров, но в целом это ретроспективный взгляд на ситуацию. Для управления на основе данных нужно двигаться дальше. Необходимо прогнозировать развитие ситуации, на основе анализа стараться понять, почему меняются показатели, и, где возможно, проводить эксперименты для сбора данных, которые могут помочь понять причины.

Давайте сравним два этих понятия. Вот варианты их возможных определений.

Отчетность — процесс организации данных в информационные сводки для отслеживания того, как функционируют разные сферы бизнеса².

Анализ — преобразование данных в выводы, на основе которых будут приниматься решения и осуществляться действия с помощью людей, процессов и технологий³.

Отчет показывает, что произошло: в четверг в 10:03 на сайте наблюдалось максимальное число посетителей — 63 000 человек. Он дает конкретные цифры.

Анализ показывает, почему это произошло: в 10:01 о компании упомянули в ТВ-шоу 60 Minutes, — и рекомендует, что компании следует делать, чтобы оставаться примерно на этом же уровне.

Отчеты ретроспективны, анализ дает рекомендации.

В табл. 1.1 суммированы отличия между этими понятиями. Теперь должно быть очевидно, почему анализ и управление на основе данных — настолько важный компонент ведения бизнеса. Это факторы,

¹ Закон от 30 июля 2002 года, названный по именам его разработчиков и инициаторов: сенатора-демократа Пола Сарбейнза и конгрессмена-республиканца Майка Оксли. В соответствии с этим законом значительно ужесточились требования к финансовой отчетности. *Прим. ред.*

² Dykes B. Reporting vs. Analysis: What's the Difference? Digital Marketing Blog, October 19, 2010. URL: <https://blogs.adobe.com/digitalmarketing/analytic/reporting-vs-analysis-whats-the-difference/>.

³ Faria M. Acting on Analytics: How to Build a Data-Driven Enterprise. BrightTALK, September 11, 2013. URL: <https://www.brighttalk.com/webcast/1829/80223>.

способные дать компании новые направления развития или вывести ее на новый уровень эффективности.

Таблица 1.1. Основные характеристики отчета и анализа

Отчет	Анализ
Описательный	Дает рекомендации
Что?	Почему?
Ретроспективный	Перспективный
Поднимает вопросы	Отвечает на вопросы
Данные → информация	Данные + информация → выводы
Отчеты, дашборды, оповещения	Наблюдения, рекомендации, прогнозы
Отсутствие контекста	Контекст + история

Источник: взято преимущественно у Б. Дайкса

Полезно для понимания аналитики ознакомиться с работой Т. Дэвенпорта и др. (см. табл. 1.2)¹.

Таблица 1.2. Гипотетические основные вопросы, на которые отвечает аналитика, по Дэвенпорту (на основе работы Дэвенпорта и др., 2010). Пункт D представляет собой ценную аналитику, пункты E и F обеспечивают управление на основе данных, если эта информация стимулирует конкретные действия (подробнее об этом ниже).

	Прошлое	Настоящее	Будущее
Информация	A) Что случилось? Отчет	B) Что происходит сейчас? Оповещение	C) Что произойдет? Экстраполяция
Выводы	D) Как и почему это произошло? Моделирование, экспериментальное планирование	E) Какой следующий оптимальный шаг? Рекомендации	F) Что самое хорошее/плохое может произойти? Прогноз, оптимизация, симуляция

В нижнем ряду таблицы отражены действия, приводящие к выводам. Как уже отмечалось ранее, составление отчетов (A) и оповещение (B) — не управление на основе данных: они отмечают, что уже произошло или что необычное или нежелательное происходит сейчас,

¹ Davenport T. H., Harris J. G. and Morison R. Competing on Analytics. Boston: Harvard Business Press, 2010.

но при этом не дают объяснений, почему это произошло или происходит, и не дают рекомендаций по улучшению ситуации. Предвестником управления на основе данных служит дальнейшее изучение причинно-следственных связей с помощью моделей или экспериментов (D). Только понимая причины произошедшего, можно сформулировать план действий или рекомендации (E). Пункты E и F обеспечивают управление на основе данных, но только если полученная информация стимулирует конкретные действия.

(Пункт C представляет собой опасную зону, поскольку слишком велик соблазн распространить существующий тренд на будущее: в Excel выберите «Диаграмма» (Chart), нажмите «Добавить линию тренда» (Add trendline) — и вот вы уже экстраполировали текущие данные на другие ячейки и делаете необоснованные прогнозы. Даже при обдуманном выборе функциональной формы модели может быть множество причин, почему этот прогноз ошибочен. Для уверенности в прогнозах следует использовать модель учета причинно-следственных связей. Подробнее об этом типе анализа — в главе 5.)

Итак, в нижнем ряду таблицы отражены перспективные виды деятельности, включающие элементы причинно-следственного объяснения. Теперь мы переходим к тому, что означает управление на основе данных.

Критерии управления на основе данных

Для компаний с управлением на основе данных характерны виды деятельности, перечисленные ниже.

- Эти компании постоянно проводят различные тестирования, например A/B-тестирование на сайте или тестирование заголовков в электронной рассылке маркетинговой кампании. Социальная сеть LinkedIn, например, проводит до 200 тестирований в день, сайт электронной коммерции Etsy одновременно может проводить до десяти тестирований. Тестирование иногда проводится непосредственно с участием конечных пользователей, чтобы компания могла получить прямую обратную связь относительно потенциальных новых характеристик или новых продуктов.
- Тестирования направлены на постоянное совершенствование деятельности компании и ее сотрудников. Это может быть постоянная оптимизация основных процессов, например сокращение производственного процесса на несколько минут или снижение

цены за конверсию, что становится возможным благодаря тщательному анализу, специально разработанным математическим или статистическим моделям и симуляции.

- Компании могут заниматься прогнозным моделированием, прогнозированием объема продаж, курса акций или выручки, но, что самое важное, они используют собственные прогнозные ошибки для улучшения своих моделей (см. главу 10).
- Практически всегда они выбирают среди будущих вариантов или действий на основе набора взвешенных показателей.

Ресурсы всегда конечны, и всегда есть аргументы за и против разных рациональных способов действий. Для принятия окончательного решения необходимо собрать данные для каждого набора показателей, которые тревожат или интересуют компанию, и определить их значимость. Например, когда компания Warby Parker собиралась открывать первый офис за пределами Нью-Йорка, то комплексно рассматривала и оценивала целый ряд переменных в отношении нового места: индекс благополучия Gallup (Well-being index), кадровый потенциал, прожиточный уровень, стоимость билетов до Нью-Йорка и так далее. Марисса Майер (CEO компании Yahoo!) делилась похожей историей: как она выбирала между разными предложениями о работе и приняла решение работать в компании Google¹.

Компания с управлением на основе данных будет делать хотя бы что-то из перечисленного, что направлено на будущее и имеет акцент на данных.

Итак, у нас в компании есть качественные данные и квалифицированные специалисты по работе с этими данными, которые занимаются деятельностью, направленной на перспективу. Теперь-то нас можно назвать компанией с управлением на основе данных?

К сожалению, не совсем. Это все равно что в лесу падает дерево, но никто этого не слышит. Если специалисты по работе с данными проводят анализ, но никто не обращает на него внимания, и если результаты этого анализа никак не отражаются на процессе принятия решений в компании, то это нельзя считать управлением на основе данных. Специалисты по работе с данными должны информировать тех, кто при-

¹ Bosker B. Google Exec Marissa Mayer Explains Why There Aren't More Girl Geeks. The Huffington Post, July 6, 2011. URL: http://www.huffingtonpost.com/2011/07/06/google-marissa-mayer-women-in-tech_n_891167.html.

нимает решения, и последние должны делать это, учитывая результаты работы аналитиков.

Дайкс предлагает термин «аналитическая цепочка ценности» (см. табл. 1.3). Данные ложатся в основу отчетов, которые будут способствовать проведению более глубокого анализа. Результаты анализа предоставляются лицам, принимающим решения, и процесс принятия решений строится на их основе. Это ключевой шаг. Данные и результаты анализа, о которых идет речь, требуются для принятия решения, способного повлиять на стратегию или тактику компании или ее развитие.

Технологии и обучение могут обеспечить первую часть плана: помочь специалистам по работе с данными с проведением анализа и представить результаты этого анализа. Однако именно от корпоративной культуры компании зависит, обратят ли на данные и результаты анализа внимание, будут ли им доверять и предпринимать на их основе конкретные действия.

Наконец мы добрались до самого важного аспекта, определяющего управление на основе данных. Для компании с управлением на основе данных именно данные — основной фактор, обуславливающий стратегию и влияющий на нее. В такой компании формируется конструктивная корпоративная культура, при которой данным доверяют, а результаты анализа бывают высокозначимыми, информативными и используются для определения следующих шагов.

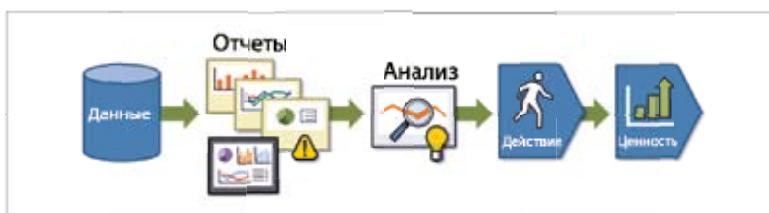


Рис. 1.3. Аналитическая цепочка ценности (по Дайксу, 2010). В компании с управлением на основе данных данные ложатся в основу отчетов, способствующих проведению более глубокого анализа. Результаты анализа влияют на процесс принятия решений, определяющий направление, в котором движется компания, и обеспечивающий ценность

Источник: <https://blogs.adobe.com/digitalmarketing/analytics/reporting-vs-analysis-whats-the-difference/>

В этом-то и заключается сложность. Если решения в компании принимаются на основе интуиции, как вывести ее на уровень управления на основе данных? Это процесс нелегкий и небыстрый, поэтому не стоит ожидать мгновенных изменений, однако все сотрудники компании могут внести свой вклад в этот процесс. Мы рассмотрим несколько способов, как стимулировать развитие в компании управления на основе данных.

Зрелость аналитических данных

В 2009 году Джим Дэвис, старший вице-президент и директор по маркетингу SAS Institute, выделил восемь уровней аналитических данных¹.

Стандартные отчеты

Что произошло? Когда произошло? Например, ежемесячные финансовые отчеты.

Ad hoc² отчеты

Как много? Как часто? Например, специальные отчеты.

Детализация по запросу

(или интерактивная аналитическая обработка, OLAP)

В чем конкретно проблема? Как найти ответы? Например, исследование данных о типах сотовых телефонов и поведении их пользователей.

Оповещения

Когда нужно действовать? Какие действия нужно предпринять немедленно? Например, загрузка ЦП, о которой говорилось ранее.

Статистический анализ

Почему это происходит? Какие возможности я упускаю? Например, почему все больше клиентов банков перекредитовываются для выплаты ипотеки.

¹ SAS, Eight Levels of Analytics (Cary, NC: SAS Institute, Inc., 2008), 4. URL: https://www.sas.com/en_us/news.htmlsascom/Analytics_levels.pdf.

² Латинская фраза, означающая «к этому, для данного случая, для этой цели». В данном контексте — специальные отчеты для исследования какой-то конкретной темы. *Прим. науч. ред.*

Прогнозирование

Что, если этот тренд продолжится? Какой объем потребуется? Когда он потребуется? Например, компании, работающие в розничной торговле, могут прогнозировать спрос на продукты в зависимости от магазина.

Прогнозное моделирование

Что произойдет дальше? Как это повлияет на бизнес? Например, казино прогнозируют, кто из VIP-посетителей будет больше заинтересован в конкретных пакетных предложениях по отдыху.

Оптимизация

Как улучшить наши процессы? Какое решение сложной проблемы будет самым эффективным? Например, каков лучший способ оптимизировать ИТ-инфраструктуру с учетом многочисленных конфликтующих ограничений с точки зрения бизнеса и ресурсов?

Представленные идеи формируют график из книги Дэвенпорта и Харриса *Competing on Analytics* (2006)^{1,2}, как показано на рис. 1.4.

(Как видите, табл. 1.2 основана на этом графике. Можно соотнести первые четыре уровня графика с верхним рядом таблицы, а вторые четыре — с нижним рядом.)

Мне нравится общая концепция и названия. Однако, исходя из того, как Дэвис (2009) и Дэвенпорт и Харрис (2007) представили свои идеи, особенно с большой восходящей стрелой, можно интерпретировать эти уровни как последовательность, своего рода иерархию, где подняться на следующий уровень можно только при условии прохождения предыдущего.

Эту псевдопрогрессию часто называют зрелостью аналитических данных. Если забьете в поисковую строку Google ключевые слова «*analytics maturity*», то поймете, что я имею в виду. Многочисленные специалисты представляют этот график как набор последовательных шагов для достижения цели, где односторонние стрелки указывают переход на новый уровень.

¹ Издана на русском языке: Дэвенпорт Т., Харрис Д. Аналитика как конкурентное преимущество. Новая наука побеждать. М. : BestBusinessBooks, 2010. Прим. ред.

² Несмотря на то что книга Дэвенпорта и Харриса появилась на два года раньше, этот источник называют «адаптация графика, сделанного компанией SAS».

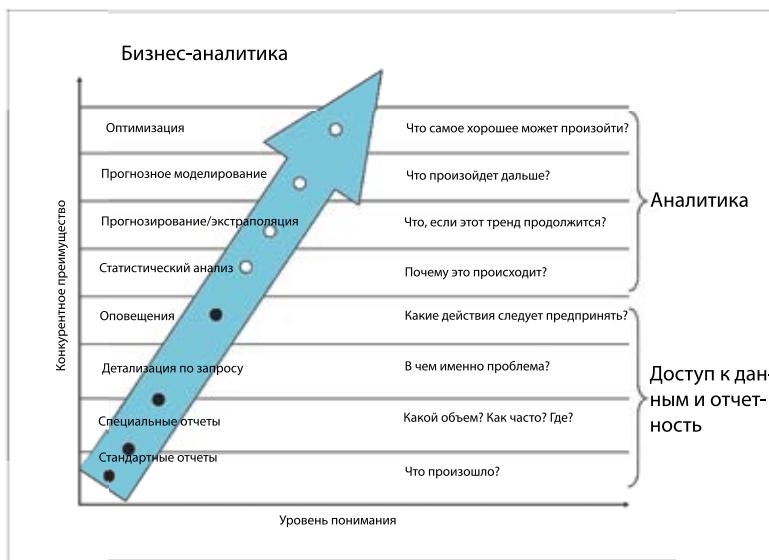


Рис. 1.4. «Бизнес-информация и аналитика» из книги Дэвенпорта и Харриса *Competing on Analytics*

Источник: HBR Press, ранее взято из уровней аналитических данных Джима Дэвиса

Аналитическая работа отличается от этого представления: в одно и тоже время разные подразделения компании могут проводить анализ разной степени сложности.

Рон Шевлин рационально отмечает¹:

С точки зрения возможностей нет причин, почему компания не может прогнозировать, например, объем продаж («уровень» 6), не зная, в чем конкретно «проблема» с продажами («уровень» 3)... Но как я, будучи руководителем, должен отвечать на вопрос «Какие действия нужно предпринять немедленно?» без понимания «Что будет, если этот тренд продолжится?» и «Что произойдет дальше?» («уровни» 6 и 7)?

Мне кажется, верный способ интерпретации — подумать о том, что **максимальный** уровень развития аналитики в компании положительно

¹ Shevlin R. The Eight Levels Of Analytics? The Financial Brand, October 27, 2009. URL: <http://thefinancialbrand.com/46761/the-eight-levels-of-analytics/>.