

«ХРАНИЛИЩЕ BIG DATA ДОЛЖНО БЫТЬ НЕ СЛОЖНЕЕ СКЛАДА ИКЕА»

Основатель агентства делового туризма «Аэро клуб»
Денис Матюхин о том, как работа с большими
данными меняет бизнес и людей.



каким объемом данных работает индустрия бизнес-туризма, проще всего объяснить на примере авиабилетов. Их оформление — основная услуга, которую оказывает «Аэро клуб», составляющая примерно 60% оборота компании. Вид авиационных билетов за последние 20 лет заметно изменился: сначала они были бумажными и выписывались от руки, потом стали электронными, но объем информации в них все это время продолжает оставаться огромным — это сотни строк. Для пользователей это не всегда очевидно, основной объем данных скрыт. По сути, в билете остается только та информация, которая требуется для выставления счета, — фамилия пассажира, название авиакомпании, данные рейса. Все остальное долгое время просто «выкидывалось». Так делали 100% компаний нашей отрасли — ничего необычного в этом не было. Понимание того, как это все может работать, пришло не сразу.

Пионерами использования Big Data в туристической индустрии были сервисы, ориентированные на массовую аудиторию. И очень долгое время мы, представители B2B-компаний, доказывали всем, что никакие B2C-технологии нам неинтересны. Казалось, это два разных полюса. Но в 2014 году ситуация изменилась. Начался кризис, в любой момент отрасль могла столкнуться с высокой конкуренцией за клиентов, и мы впервые задумались о персонализации собственных сервисов как о чем-то особенном, что можем предложить рынку. Тогда оказалось, что инструменты B2C могут работать и в сегменте B2B, прежде всего я говорю о технологиях подбора различных вариантов авиаперелетов. Ведь, по сути, разница не такая большая: контракт на перевозку всегда заключается между физическим лицом, например Ивановым, и авиакомпанией, например «Аэрофлотом». И физическому лицу нужно получить качественный сервис — неважно, предоставляет его B2C- или B2B-компания.

И когда мы наконец задумались о персонализации сервисов, выяснилось, что необходимых данных у нас нет. Хорошо еще, что сведения, которые когда-то казались лишними, мы не буквально «выбрасывали на помойку», а просто отправляли в архив. Когда мы их оттуда достали, остро встал вопрос технической организации процесса хранения и обработки данных. Сейчас многие компании заявляют о том, что у них есть Big Data. Честно говоря, верю я им далеко не всегда, а если говорить о сегменте делового туризма, то данными в полном объеме владеют буквально единичные игроки. Во всех остальных случаях это какие-то усеченные или вообще придуманные данные. Механизм тут достаточно простой: говорить о том, что все сведения собраны, — круто, а когда их приходится подтверждать — все можно придумать из головы. И доказать это очень сложно: маловероятно, что клиент залезет во внутреннюю кухню компании.

Но мы в «Аэро клубе» пошли честным путем. И могу точно сказать — этот процесс, во-первых, очень трудоемкий, а во-вторых, крайне дорогой. Перед тем как сформировать «озеро данных», так называемое Big Data Lake, мы должны их сначала собрать, а потом очистить, привести в состояние, в котором они смогут использоваться в рамках Business Intelligence (BI). Сначала нужно создать склад. Он должен быть устроен не сложнее, чем стандартный магазин ИКЕА, где, если вы выберете себе диван, сразу можете узнать, в каком складском стеллаже и на какой полке будут находиться все его комплектующие. На организацию процесса хранения и обработки данных ушло много времени: только в одном авиабилете сто с лишним полей, а в гостиничном бронировании ненамного меньше. Все это нужно соединять, правильно распределять и добавлять к этому данные, которые уже есть на стороне клиента.

Здесь остро встает проблема размера хранилища. Когда вы храните сто или тысячу транзакций — это один вопрос, но когда речь о нескольких миллионах, ситуация резко меняется. Нужен не один, не два и не три сервера, плюс они должны быть очень высокого качества, защищенные. Также нужно понимать, что поток данных у нас не кончается, каждая новая оказанная услуга, выписанный билет формируют дополнительный объем. По мере роста количества информации стоимость ее хранения постоянно растет. При этом прямой монетизации Big Data в на-

шей отрасли пока нет. Именно поэтому я редко верю небольшим компаниям, которые говорят о том, что у них есть Big Data. Как человеку, который работает на этом рынке, мне очевидно, что если чистая прибыль компании составляет 30–50 млн руб., то она вряд ли будет распределять их на направления, которые не приносят доход. При этом данные накапливают, и не вкладываются в их хранение просто невозможно. Поэтому транзакции не сохраняют, а если данные нужны для демонстрации клиенту — их придумывают.

Сейчас, по сути, мы монетизируем Big Data только косвенно, в рамках BI. Это очень интересная вещь, которая позволяет оперативно доставать и использовать информацию. Я люблю приводить один наглядный пример: имея определенные данные, мы сможем, обработав их, выявить прямое влияние восхода и захода солнца на стоимость авиабилетов. Конечно, на практике это вряд ли кого-нибудь заинтересует. Но речь о том, что любую, даже самую несуразную закономерность можно проверить, используя продукты на базе BI.

Проект для нас оказался непростым. Изначально он запускался как сервис для конкретного клиента, в котором мы сделали несколько интересных шаблонных решений. Когда мы начинали писать панели индикаторов, по сути, мы применяли BI в отчетах, где ранее традиционно использовался Excel. У последнего есть один большой минус — если объем больше чем 64 тыс. строк, сделать там ничего невозможно. У нас есть клиенты, которые подбираются к этому объему, и с ними приходилось достаточно не просто: файл тяжелый, его не так просто отправить, потом не так просто получить, а в итоге он просто у кого-нибудь не будет грузиться. Когда вместо Excel мы предложили клиентам пользоваться нашим BI-инструментом АТОМ, стало очевидно, что этот продукт сильно помогает сотрудникам самой компании-клиента. Например, приходит человек, ответственный в компании за тревел-направление, отчитываться на совет директоров и сразу с помощью планшета может показать, как структурируются расходы на командировки. Его спрашивают: а что если мы сократим эти траты на 5%? Он может ввести эти данные в шаблон и сразу смоделировать несколько возможных сценариев.