

## Урок 7

# Поиск зависимостей в данных

### 7.1. Взаимное влияние в продажах товаров

Третья неделя курса посвящена методам выявления взаимосвязей и поиска закономерностей. Это вещи, которые статистика умеет делать лучше всего.

Одна из наиболее ярких областей, в которых постоянно требуется решать задачи с помощью статистики, — это ритейл, или розничная продажа. Такие задачи сопровождают продажи практически на всех этапах, начиная от планирования и аренды складских помещений, заканчивая разработкой гибких скидочных политик для постоянных клиентов. Например, при выборе места для склада важно, чтобы, с одной стороны, оно стоило дешево, а с другой — находилось где-то в центре и до него было удобно добираться. Помимо локации еще нужно выбрать площадь склада, а оценить ее тоже довольно сложно. Если площадь склада будет слишком маленькой, то какую-то часть времени товары может быть негде хранить, и придется арендовать дополнительную площадь, что влечёт за собой дополнительные расходы. С другой стороны, если сразу арендовать большие складские помещения, то в течение какого-то времени они будут пустовать. Это тоже дополнительные расходы, но уже из-за простоя.

Выше приведён пример небольшой задачи, которая в ритейле, фактически, решается еще до старта бизнеса. Такие задачи сопровождают бизнес на протяжении всего жизненного цикла и их количество не позволяет решать их все в голове. Получается, что для того, чтобы успешно вести бизнес в области ритейла, необходимо грамотное планирование. Заблаговременно требуется получить оценку того, какой будет спрос на различные товары, сколько места требуется на их хранение, а также сколько места будет свободно с учетом уже хранящихся товаров. Это означает, что нужно уметь строить модель, способную делать прогнозы такого вида. Однако для этого недостаточно одних моделей, способных связать площадь склада и объем продаж. Помимо этого необходимо опираться на существующую статистику работы предприятия, например, на статистику по продажам. Получается, для грамотного планирования необходимо строить модели, сочетающие методы прогнозирования и методы статистики.

Также в этих задачах так много параметров, что очень легко построить модель, которая не будет интерпретируемой. Интерпретируемость — это очень важный фактор, потому что люди, принимающие решения в бизнесе, не готовы доверять моделям-черным ящикам. Поэтому необходимо, чтобы в моделях было как можно больше здравого смысла. Требуется уметь не только вносить здравый смысл в модели, но и извлекать его из данных. Именно таким методам посвящена третья неделя курса.

### 7.2. Внешние факторы, влияющие на продажи

Анализировать данные на предмет выявления различных закономерностей не только очень интересно, но и позволяет принести практическую пользу бизнесу. Так, проанализировав продажи магазина, можно сделать вывод о том, какие товары покупают наиболее часто, а эта информация предоставляет целый ряд практических преимуществ. Её можно использовать для планирования продаж, закупок, рекламных кампаний, а также для оптимизации размещения товаров в торговом зале или, в случае онлайн-магазинов, для правильного построения блока «популярные товары».

Построение промо-блоков — это очень актуальная задача. Так, проанализировав чеки оффлайн-магазина или содержимое корзины онлайн-магазина, можно сделать вывод о том, какие товары чаще всего покупают вместе. Это, в свою очередь, позволяет строить так называемые товарные рекомендации, рекомендации вида

«С товаром А наиболее часто покупают товар В». По данным различных онлайн-магазинов, такие рекомендации сильно способствуют продажам. Они позволяют принести 15–30% дополнительных продаж, а это очень много.

Имея достаточное количество информации о товарах, можно делать очень много вещей: анализировать взаимосвязи между ними (например, выявлять аксессуары, аналоги или комплементарные товары), анализировать успешность товаров или брендов у той или иной целевой аудитории, оценивать долю продукта на рынке, смотреть, как она меняется во времени, или, например, оценивать эффект каннибализации (каннибализация — это известный маркетинговый эффект, заключающийся в том, что один из товаров отбирает целевую аудиторию у похожего товара или бренда). Такие экономические явления, как, например, каннибализация, существенно сказываются на продажах. Однако важно понимать, что нужно учитывать и такие, казалось бы, незначительные вещи, как расположение товара в магазине, его близость к кассе, высота полки, на которой он находится, хорошо ли товар видно, могут ли до него дотянуться дети. Всё это тоже очень сильно сказывается на продажах. Более того, важно учитывать контекст, в котором товар существует, например, его ближайшее окружение: какие товары и бренды находятся на полке рядом с ним.

Структура продаж магазина определяется в первую очередь людьми, которые в него ходят. Часто оказывается, что в городах и в деревнях покупают совершенно разные вещи, причем иногда эти зависимости совершенно не такие, как можно было ожидать. Например, в одном из проектов были проанализированы продажи в сети доступных супермаркетов, и выяснилось, что в деревнях люди чаще покупают дорогой алкоголь.

Факторы, связанные со временем, — это, пожалуй, самые важные факторы для прогнозирования продаж. Многие товары по-разному продаются в разное время. Например, шампанское очень хорошо продается в Новый год, мороженое лучше продается летом и т.д. Кроме того, есть всякие длинные, медленно меняющиеся во времени тренды, связанные, например, с тем, что какой-то товар постепенно с рынка выходит, а какой-то товар только на нём появляется.

Также важно учитывать экономическую ситуацию регионов продаж. Так, колебания в курсах валют, изменение налогового или таможенного законодательства способны существенно повлиять на продажи. Например, в моменты нестабильной экономической ситуации товары и бренды из более низкого ценового сегмента начинают отъедать долю рынка у товаров из более высоких ценовых сегментов. Однако и обратное тоже верно: в период стабильности экономики и роста доходов населения товары из премиум и высоких классов начинают возвращать свою долю рынка обратно.

Выше перечислено большое количество факторов, которые могут быть полезны при прогнозировании продаж. На самом деле, таких факторов может быть еще больше, и методы, которые будут обсуждаться на этой неделе, способны эти факторы из данных выявлять.

Задачи ритейла очень интересные и большую часть времени речь будет идти именно о них. Однако выявление закономерностей из данных активно применяется и в других областях. Методы выявления закономерностей традиционно широко используются в социальных науках. Например для выявления гендерной дискриминации при приеме на работу или для оценки эффективности государственных социальных программ. Очень часто они применяются в биологии и медицине для выбора оптимального лечения или для поиска генов, которые в организме действуют совместно, или для поиска подгруппы пациентов, у которых побочные эффекты наблюдаются в наиболее тяжелой форме.