

20 сентября 2021

Кластеризация в ценообразовании: основные принципы

Кластеризация магазинов для теста и контроля

Любой эксперимент по оптимизации ценообразования начинается с теста. Для этого выбираем два магазина для проверки гипотез: тестовый и второй, максимально похожий на первый, будет контрольным. Сравнивая результаты между ними, мы увидим отклонения, которые и будут сигнализировать об изменении покупательской способности. Также на базе исторических данных и поведения магазина во время ценового эксперимента можно проследить стандартное покупательское поведение.

При установлении новой цены на только введенный в магазин товар или изменении существующей важно правильно спрогнозировать реакцию покупателя во второй торговой точке. При этом тестовая выборка должна была максимально похожа на контрольную.

Замер будет «чистым», если выборка тестовой и контрольной групп будет максимально репрезентативна. Для понимания сопоставимости двух магазинов нужно создать критерии их сравнения по показателям: количество покупателей, структура чека, средний спрос, выручка и доля категорий в ней. Дополнительно можно выбрать категории, которые описывают структуру спроса. Всё это станет бизнес-ограничениями для разрабатываемого алгоритма.

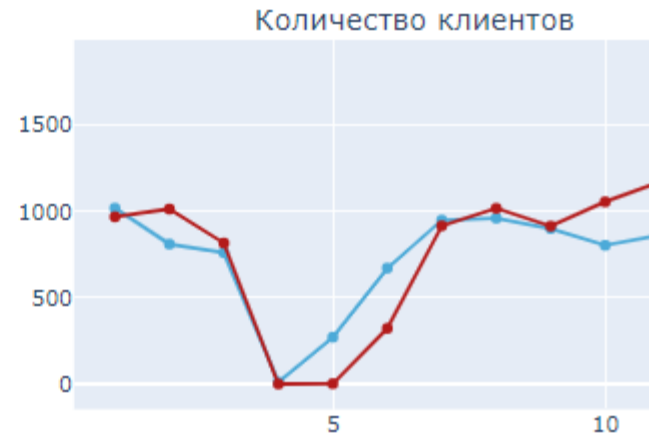
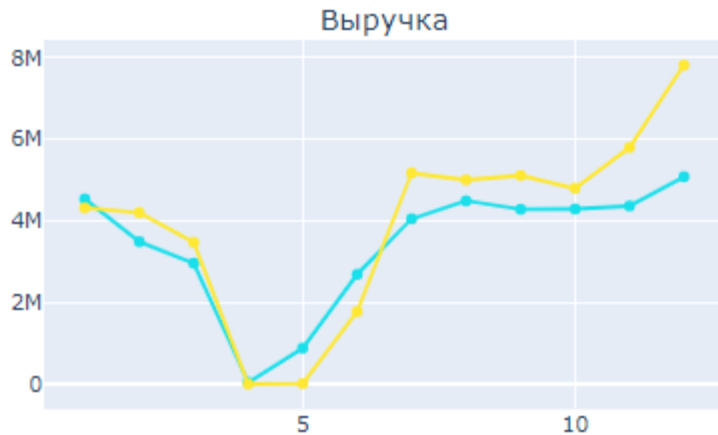
Как помогает ML

В рамках созданных критериев можно сделать предположение, что алгоритм должен сводиться к задаче кластеризации временных рядов, то есть последовательно измеренных через некоторые промежутки.

Однако даже при группировке статистических данных часто получаются сомнительные результаты, тоже самое получается с информацией, собранной из временных рядов. Такая задача создает дополнительные сложности. Например, временные ряды могут быть не информативны и иметь разное количество отсчетов, а интерпретации при работе с последовательными данными могут быть разными. Кроме того, при кластеризации рядов более сильный вклад вносит семантика данных.

Поэтому с возможными семантическими отличиями данных полезно выделять несколько типов подобия:

- 1) Подобие по особым точкам и интервалам, которые точно соответствуют друг другу во времени. При этом необязательно, чтобы значения экстремумов, (максимально и минимально допустимых значений) и наклоны кривых совпадали.
- 2) Поиск временных рядов с одинаковыми законами изменения: периодичность и её паттерны, тренд.



- Выручка_магазин_1
- Кол-во_клиентов_магазин_1
- Средний_чек_магазин_1
- Доля_выручки_категорий_магазин_1
- Выручка_магазин_2
- Кол-во_клиентов_магазин_2
- Средний_чек_магазин_2
- Доля_выручки_категорий_магазин_2

Чтобы выделить в массиве информации ясные закономерности, нужно сконструировать хорошую функцию близости. Если представить, что временной ряд – это сигнал, изменяющийся во времени, то алгоритмы для анализа и распознавания речевых сигналов хорошо справляются с задачей нахождения максимально близких друг другу магазинов. Метрик близости

сигналов очень много, их можно объединить в несколько существующих алгоритмов:

1) Динамическая трансформация временной шкалы (Dynamic Time Warping, DTW)

Одинаковые части временных рядов не только отстают друг от друга по оси времени, но и растянуты или сжаты. Не так много метрик расстояний с этим могут справиться или дать показательные результаты. DTW-алгоритм строит матрицу возможных отображений одного ряда на другой с учетом того, что отсчеты рядов могут как сдвигаться, так и изменять уровень. Путь в этой матрице с минимальным значением стоимости и есть DTW-расстояние.

2) Дискретное расстояние Фреше (сцепленное расстояние)

Представим, что у нас есть две последовательности (P и Q). Будем двигать точку A по последовательности P, а точку B – по последовательности Q, необязательно равномерно. В этом случае расстояние Фреше между P и Q будет минимально возможное расстояние между точками A и B.

3) Площадь между двумя кривыми

Площадь фигуры между двумя кривыми в прямоугольных координатах определяется интегралом. Если же кривые не пересекаются, пределы интегрирования остаются неизвестными.

Замер осуществляется каждую неделю и накопительно. Период до теста фиксируется и не меняется в течение пилота. Замер будет «чистым», если тестовая и контрольная группа максимально сопоставимы.

Теперь мы можем проверять по этим метрикам наши новые рекомендуемые цены. Ситуации, которые мы увидим, могут быть разные. Например, магазины могут значительно отличаться друг от друга по выручке, графики с выручкой

могут пересекаться, а могут вести себя нестабильно.

В тестовую группу магазинов сети могут входить более одного магазина. Тогда в контрольную группу подбираем аналогичное количество максимально близких магазинов. Эта методология удобна и тем, что позволяет к одной и другой группе прибавлять новые торговые точки и проверять то, к чему приведет оптимизация ценообразования. Даже не самые сложные модели машинного обучения могут в этом с лёгкостью помочь.