

Слайд 1

Добрый день! В рамках финального проекта, мы работали над кейсом Амияма Авто, под названием "Данные из различных источников".

Слайд 2

Поставленная нам проблема заключалась в следующем. Для предоставления услуги "отчет об истории эксплуатации автомобиля по выставленному объявлению", Дром использует данные из трёх различных источников, условно источник 1, источник 2 и источник 3. Было известно, что эти источники поставляют данные по трем различным моделям подписки, информация о которых и представлена на слайде.

Слайд 3

От нас требовалось проанализировать данные о процессе оптимизации запросов, начатом командой разработчиков в начале ноября 24 года, и ответить на ряд вопросов, который вы так же можете видеть на экране.

Первым блоком вопросов занимался преимущественно мой товарищ, поэтому он про них и расскажет.

- А дальше про 1 блок
-

Слайд 6

Теперь касаемо второго блока поставленных вопросов. Ответить на них требовалось из предположения, что каждый день Дром продает по 1000 отчетов.

Для расчета себестоимости источников, было положено, что показатели получены при использовании источников независимо друг от друга.

Как можем видеть, в силу большой арендной платы и небольшого числа продаваемых отчетов, источник под 2 номером является наиболее затратным.

Слайд 7

В целом же, отмечу, что средние суточные затраты на второй источник не зависят от числа продаваемых отчетов, а значит, при достаточном их числе, он станет наименее затратным.

Слайд 8

Перейдем к вопросу о качестве отчетов. Под качеством отчетов, в рамках задачи, требовалось понимать количество проданных отчетов, в которых есть данные хотя бы из одного коммерческого источника.

Так же, известно, что источники используются по очереди и, если на каком-то из источников были получены записи, дальнейшие запросы не делаются. В силу этого, вычисление качества отчетов сводится к простой вероятностной задаче.

Как можем видеть, при отказе от 2 или 3 источников, качество отчетов падает. При этом, при отказе от 2 источника средние потери в качестве составили бы 46.5%, а при отказе от 3 - 4.2%, что, впрочем, больше поставленной планки в 2%.

Слайд 9

Последним шагом стал подбор оптимальной стратегии использования источников. Поскольку позволено менять лишь порядок, у нас есть 15 возможных комбинаций источников.

Для каждой из стратегий была рассчитана средняя себестоимость одного отчета. Снова видим, что в вакууме, наиболее дорогим является источник 2, а использование источника 1 наиболее качественно снижает себестоимость отчета.

Слайд 10

Так же, было рассчитано среднее качество отчетов. С точки зрения качества, что логично, наиболее эффективны стратегии, использующие все 3 источника.

Слайд 11

Для определения лучшей стратегии в общем, была рассчитана предполагаемая прибыль, которую получила компания за прошедший период, если бы использовала каждую из стратегий, а стоимость одного отчета была бы 175 рублей. Именно такая сумма на момент расчетов запрашивалась на дrome за услугу "Проверка авто". Таким образом лучшей стратегией будет использование всех источников в порядке 2-1-3.