



Анализ данных о вызовах колл-центров с целью определения возможностей по выявляемости неэффективных операторов колл-центра на основе собираемых данных

ИССЛЕДОВАНИЕ ДЛЯ ОПРЕДЕЛЕНИЯ НАПРАВЛЕНИЙ ДОРАБОТКИ SW

Для продуктового маркетинга «Нупозвони»

Москва, ноябрь 2022

Исполнители:

курс по аналитике данных Яндекс Практикум

студент Севастьянов Александр / DA-49

Резюме основных результатов проекта

УДАЛОСЬ

Построен эффективный персональный скоринг для определения неэффективных операторов колл-центра – смешанный сразу по двум услугам: горячей линии и целевого обзвона.

Доказано, что мотивация операторов по предложенным продуктовым маркетингам KPI для каждой из услуг статистически достоверно приводит к росту утилизации операторов – тем самым и к достижению бизнес-целей по снижению расходов компании.

Обнаружено, что основной вклад в потери звонков абонентов на горячие линии вносят не операторы, а проблемы с IVR, принимающим и проключающим вызовы на операторов.

Опробован критерий определения нехватки операторов в колл-центре - определены проблемные колл-центры с нехваткой ресурса.

НЕ УДАЛОСЬ

Разделить колл-центры по типу услуг на две группы. Точнее, вторая группа получилась принципиально смешанная по типу услуг.

Выяснить причину таких больших потерь на IVR – возможно, что это и не потери, а ситуация, когда абонент получил автоматический ответ в голосовом меню IVR, хотя среднее время проведенное на IVR не соответствует данному предположению.

Объяснить присутствие загадочных “стахановцев” с суточной выработкой, превышающей продолжительность смены – возможно, это относится к случаям, когда один эккаунт в системе колл-центра переиспользуется разными физическими операторами, хотя наличие оператора с чистым временем разговоров с абонентами за сутки выше 24 часов не подходит и под данное предположение.

План презентации

1. Формализация постановки задачи: цели проекта и модель изучаемого объекта
2. Замечания об исходных данных: состав и качество
3. Характерные значения статистик датасета – потенциально интересных для заказчика
4. Целевое исследование 1: разделение колл-центров по типу услуги на два класса
5. Целевое исследование 2: построение скоринга неэффективности операторов
6. Целевое исследование 3: исследование статистики вызовов потерянных на IVR
7. Проверка гипотез и заключительные выводы:
 1. Предложенные продуктовым маркетингом мотивационные KPI эффективности операторов также приводят к повышению утилизации операторов
 2. У колл-центров, обслуживающих звонки на горячую линию по остаточному принципу, (т.е. у колл-центров по типу целевого дозвона) время ожидания на IVR больше, чем у колл-центров по типу горячей линии.

Цели проекта

В проекте два уровня целевого потребителя его результатов:

- бизнес-заказчик проекта - продуктовый маркетинг компании “Нупозвони”
 - клиент бизнес-заказчика – поставщик услуг колл-центра (закупает SW и SLA) у “Нупозвони”
-

Цели бизнес-заказчика:

На данный момент у «Нупозвони» готова только измерительная часть SW и модуль пайплайна сбора данных (медиатор).

Также по результатам Customer Development от продуктового маркетинга определено мнение клиентов в отношении мотивационных KPI операторов.

От аналитика требуется на основе реальных данных, собранных готовой частью SW, создать метод определения неэффективных операторов колл-центра – для реализации презентационной части SW. Плюс, провести EDA на предмет проблем с данными.

Цели клиента (колл-центра):

С помощью правильной установки мотивационных KPI операторам адресовать бизнес-задачу компании по минимизации расходов за счет повышения утилизации рабочего времени операторов - то есть минимизировать количество операторов, необходимое для обслуживания актуального объема вызовов своего колл-центра.

Узнать, какие еще проблемы есть у услуг колл-центра с точки зрения конечного абонента.

Инфомодель объекта исследования

При анализе данных обнаружено:

- хорошее соответствие данных инфомодели – кроме идеи о специализации колл-центров
- необходимость задания мотивационных KPI до начала сбора данных – нельзя менять KPI

Упрощенная модель колл-центра:

Есть две услуги колл-центра: целевой обзвон и горячая линия для абонентов.

Целевой обзвон инициируется только операторами колл-центра (нет автообзвона).

Вызовы на горячую линию сперва приземляются на IVR и после успешного прохождения голосового меню проключаются далее на операторов.

Все операторы одинаковы (нет специализации) и проключение идет на менее занятого (минимальная очередь).

Операторы также могут звонить друг-другу.

Взаимосвязь KPI и бизнес-целей:

Операторам назначаются KPI на языке их сферы ответственности.

Задача технолога бизнес-процессов - изобрести такие KPI для всех участников процесса, чтобы они, в ходе их выполнения, оптимизировали достижение целей компании.

Поэтому эффективность операторов проверяется на параметрах обслуживания внешних вызовов, а бизнес-целью компании является снижение расходов - то есть при некоторой нагрузке на колл-центр минимизировать число операторов в нем - за счет повышения их утилизации.

Роль данного исследования в первую очередь проверить на данных статистики - приводят ли указанные KPI в селекции операторов в пользу выполнения бизнес-цели.

Замечания о качестве исходных данных

Основной датасет содержит 53.9 тыс. записей с посуточной статистикой вызовов – с тремя группами полей:

- Дата суточной агрегации статистик и ID оператора и колл-центра
- Атрибуты типа вызова (8 типов по трем бинарным атрибутам) – хорошо стратифицируются по KPI
- Агрегируемые характеристики вызовов: количество, продолжительности разговора и ожидания на линии

Дубликаты и отсутствующие данные:

Данные содержат 9.1% полных дубликатов (удалены).

Имеются отсутствующие данные в полях:.

Атрибут вызова (внутренний/внешний) – пропусков 0.2% и почти полностью могут быть заполнены по логике датасета (заменены)

ID оператора – пропусков 15.2%, из них только 0.5% реально отсутствующих данных (удалены), остальные 14.7% означают ситуацию, когда вызов не был проключен на оператора. То есть абонент дал отбой еще в меню IVR, либо произошел сбой в маршрутизации вызова внутри колл-центра.

Семантическая противоречивость данных:

У ряда пропущенных вызовов обнаружена ненулевая продолжительность чистого времени разговора оператора с абонентом (удалены 0.6%).

У внутренних вызовов (1.5-2.0%) количество входящих и исходящих ног не совпадает на порядок величины – а должно быть одинаково (впрочем, это не мешает анализу).

Обнаружены лже-«стахановцы» с абсолютно нереальной посуточной продолжительностью чистого времени общения с абонентами – есть даже более 24 часов!

- здесь предполагается либо баг SW, либо кривая его установка в колл-центре.

Ключевые значения статистик данных

Данные выводы могут быть потенциально интересны для бизнес-заказчика исследования

1. В собранных данных присутствует менее половины колл-центров подписавшихся на услуги, при этом:
 - Тарифом А пользуются наиболее крупные колл-центры - в среднем по 10+ операторов и большей нагрузкой по числу вызовов,
 - Тарифом В пользуются небольшие колл-центры - в среднем по 4-5 операторов и более длинным временем общения с абонентом,
 - Тарифом С пользуются самые многочисленные маленькие колл-центры всего на 3-4 оператора и с самыми короткими разговорами.
2. Новые клиенты появляются с регулярной скоростью с течение первых 3 месяцев рассматриваемого периода - исключительно по рабочим дням
3. Количество вышедших на работу операторов линейно растет в течение всех 4 месяцев, по которым есть данные в основном датасете - причем в рабочие дни их в 4 раза больше, чем в выходные,
4. Доля внутренних разговоров операторов составляет 2.0 % по числу разговоров и 1.4 % по их чистой продолжительности,
5. Доля пропущенных колл-центрами внешних вызовов составляет 52.7 % от всех внешних вызовов, из них доля пропущенных операторами внешних вызовов составляет 0.6 % от вызовов, успешно проключенных на операторов,
6. Для абонентов эффективное время общения с операторами составляет 79.3 % от всего времени звонка в колл-центр,
7. Для операторов эффективное время общения с абонентом составляет 88.6 % от всего времени обслуживания проключенного к оператору звонка (с ожиданием),
8. Доля успешных дозвонков абонентам при целевых кампаниях составляет 61.2 % от всех вызовов,
9. Для целевого обзвона медианная продолжительность разговора составляет 85 секунд.

Объем собранных данных линейно растет со временем

Что частично объясняется постоянными темпами подписки новых колл-центров на услуги «Нупозвони»

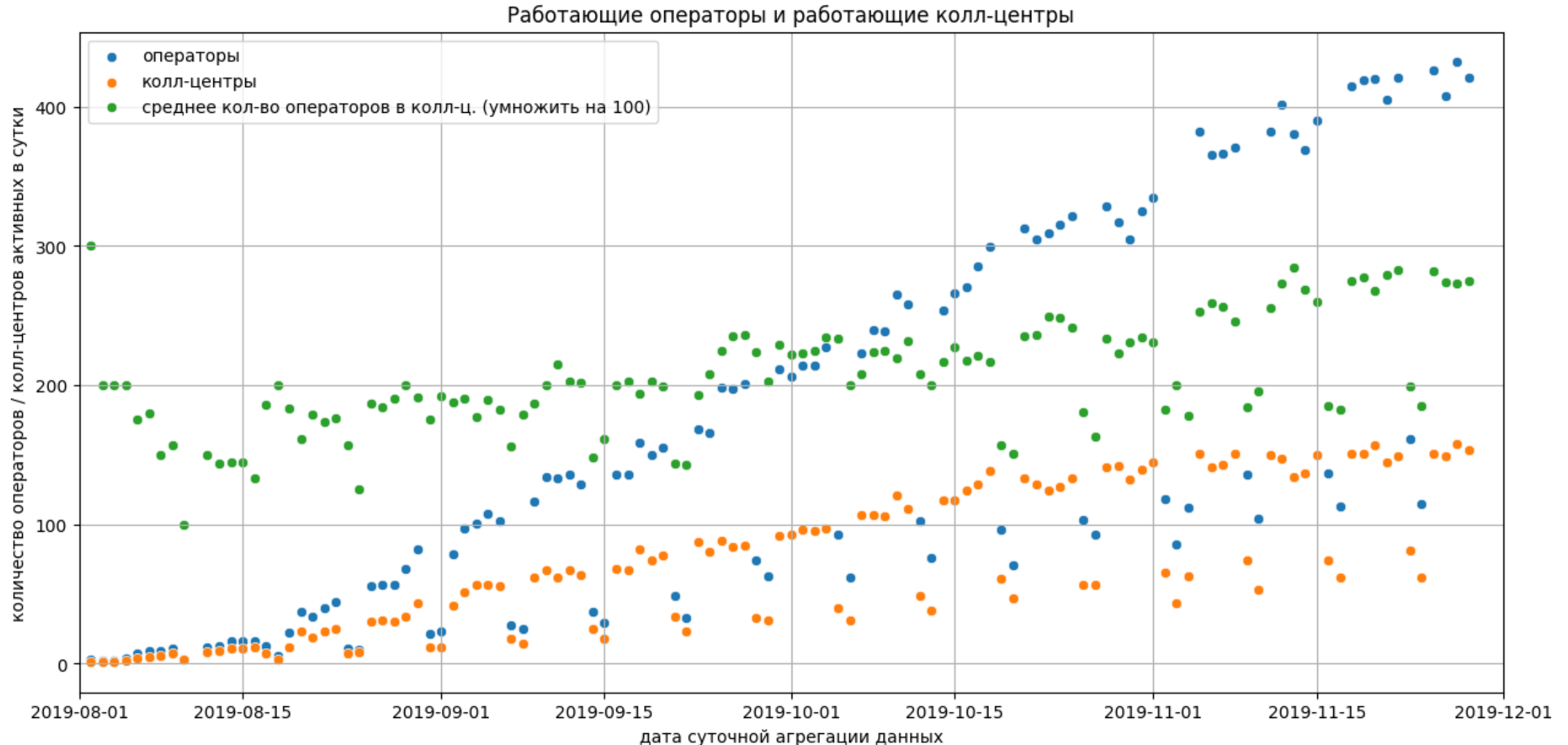


Особенности:

- Данные по вызовам присутствуют за 4 месяца август-ноябрь 2019.
- Четко виден недельный ритм нагрузки на колл-центры (5 рабочих - 2 выходных), при этом активность колл-центров по выходным в 4 раза ниже, чем в рабочие дни.
- Постоянные темпы подписки новых колл-центров на услуги «Нупозвони» - в среднем 8 клиентов ежедневно, кроме выходных.
- В последний месяц новых подписок нет, но линейный рост нагрузки продолжается с той же скоростью.

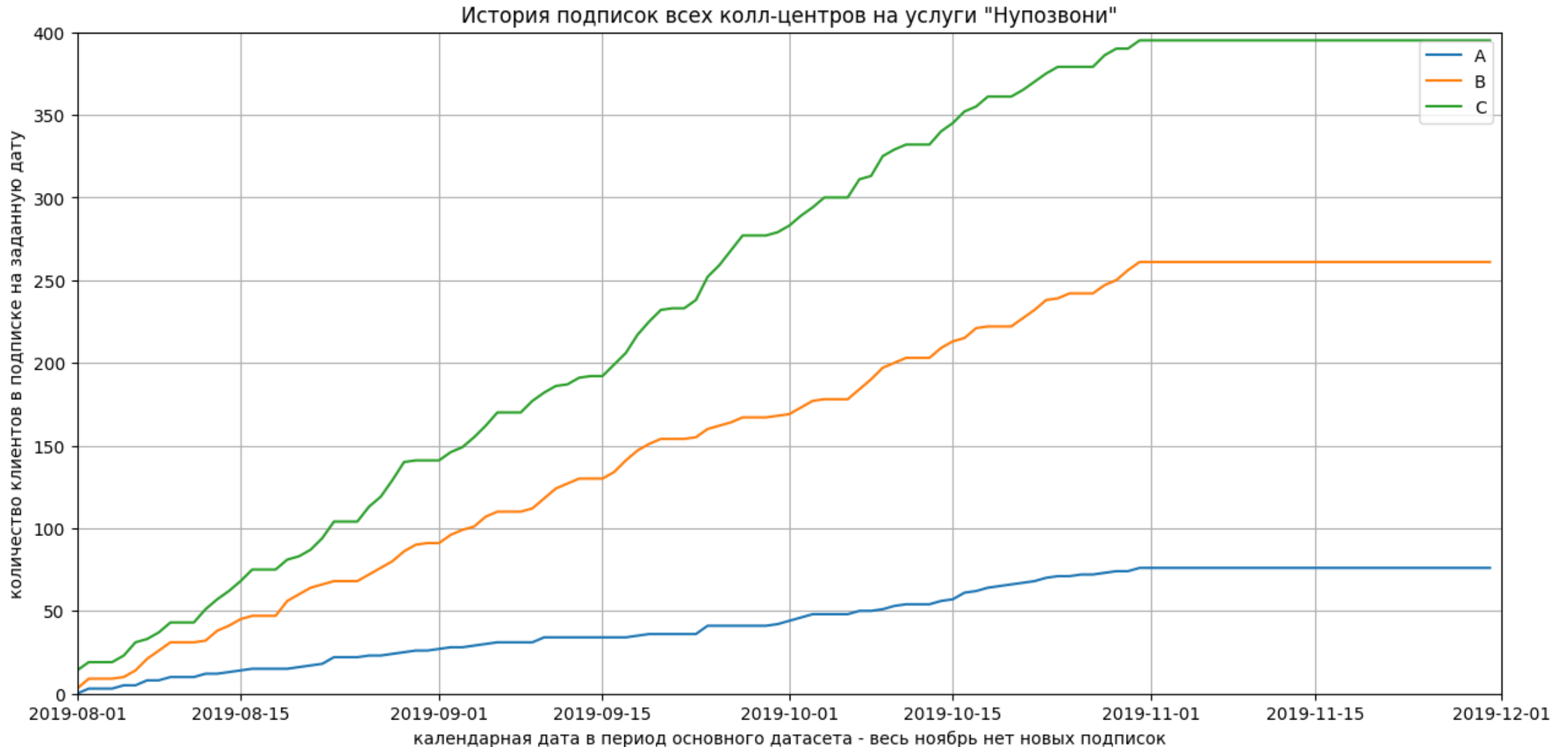
Количество операторов в смену линейно растёт со временем

При этом количество операторов в одну смену невелико – в пределах 1-3 в среднем



История подписок клиентов на разные тарифные планы

Тариф А – крупные клиенты, тариф В – средние и тариф С - мелкие



Целевое исследование разделения колл-центров на два типа: горячая линия и целевой обзвон

ПОСТАНОВКА ЗАДАЧИ

- Мотивационные KPI для операторов разбиваются на два типа услуг:
 - Горячая линия,
 - Целевой обзвон.
- Вопрос как замешивать их при оценке одного оператора неоднозначен, поэтому самый простой способ – поделить колл-центры по типу услуги и применять только один критерий с весом 100%.
- Так, возникает задача по разделению колл-центров на две группы.

ПРИМЕНЕННЫЙ МЕТОД

- Поскольку типы вызовов хорошо стратифицируются по типу услуги, то каждой записи датасета можно присвоить флаг:
 - 1 – горячая линия,
 - -1 – целевой обзвон,
 - 0 – любой.
- Далее можно просуммировать с весом кол-ва вызовов по всем записям для данного колл-центра и по знаку результата выбрать тип.
- Можно также построить скоринг, нормируя результат на суммарное кол-во вызовов.

РЕЗУЛЬТАТ(Ы)

- Только 1/4 колл-центров точно относится к горячим линиям – четко видна мода.
- Остальные колл-центры ровно размазаны по скорингу – нет яркой моды целевого обзвона.
- Знак скоринга не дает уверенного разделения – оцениваемая точность 2/3.
- Видимо, колл-центры диверсифицируют бизнес и предлагают обе услуги.
- Чистая мода у горячих линий может быть просто линия ТП производителей SW.
- Вывод: подход не годится для скоринга эффект. операторов.

Четко видна только одна мода – колл-центров горячей линии

$\frac{3}{4}$ колл-центров диверсифицируют свой бизнес и предоставляют все виды услуг

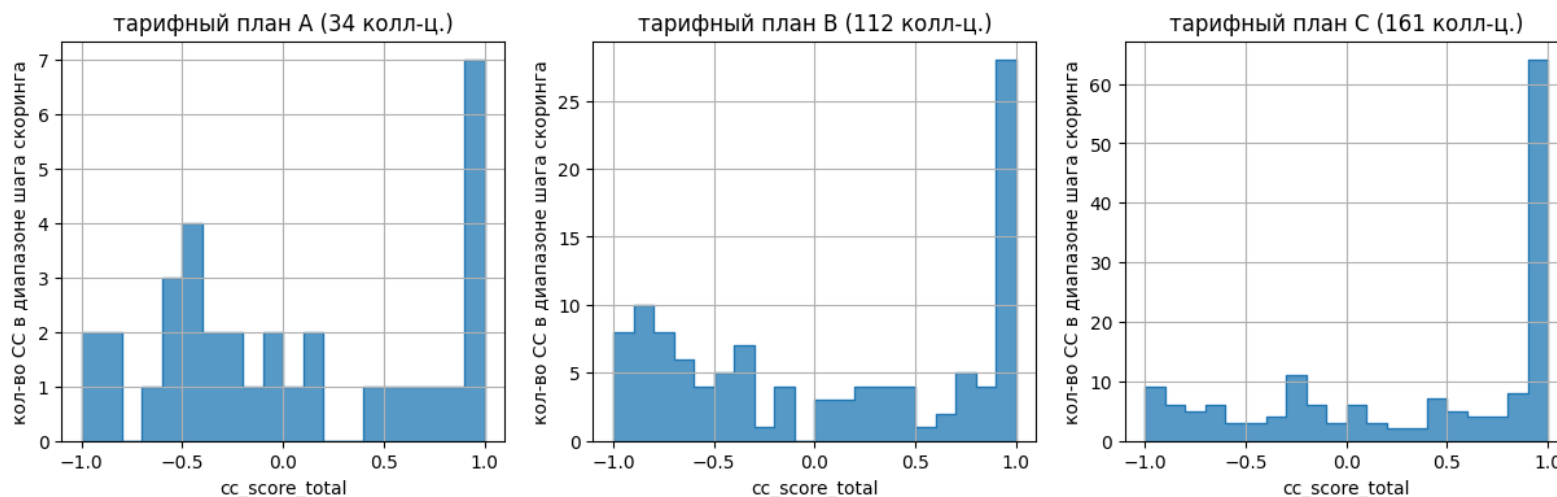


Особенности:

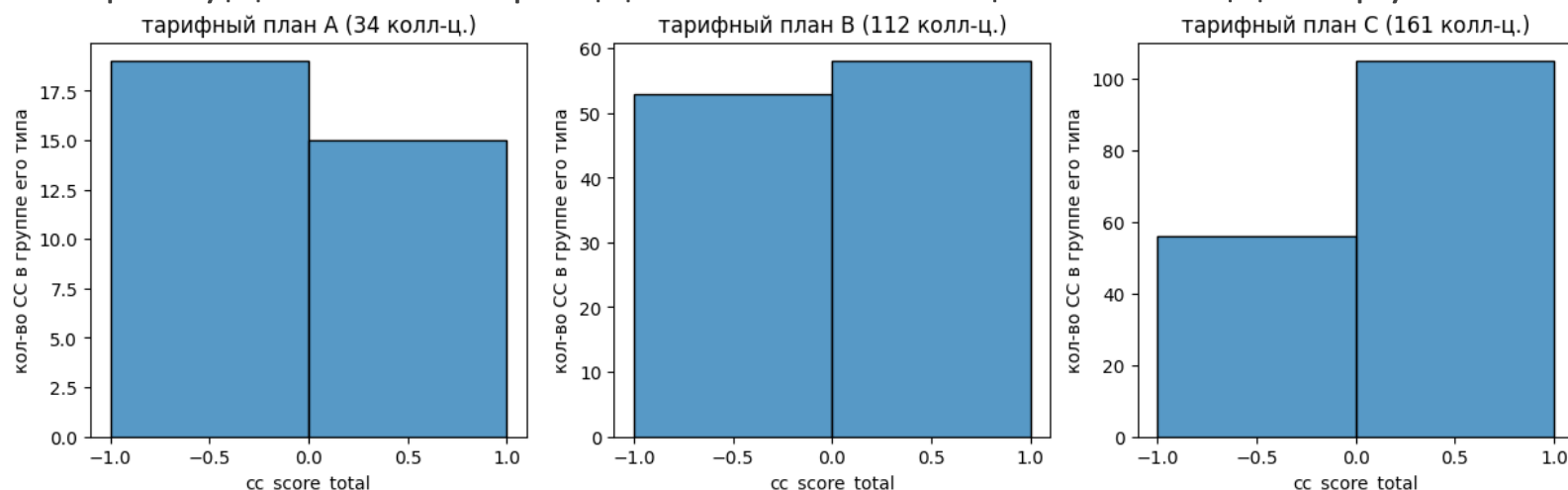
- За каждый вызов по типу услуги Г.Л. колл-центр получает плюс 1 балл,
- За каждый вызов по типу услуги Ц.О. колл-центр получает плюс -1 балл,
- Потом нормируем на число вызовов, получая скоринг колл-центра в пределах то -1 до 1.
- Чувствительность скоринга повышена срезом вызовов с баллом 0.
- Мода горячей линии, скорее всего, – это ТП от производителя, таких колл-центров в наших данных $\frac{1}{4}$.
- Остальные колл-центры занимаются аутсорсом и предлагают смешанные услуги – моды Ц.О. нет совсем.

Соотношение плеч скоринга показывает соотношение двух типов услуг – как видим, кроме моды 25% горячих линий остальные предлагают в принципе обе услуги

гистограммы распределения колл-центров по скорингу



принудительное разделение колл-центров на две группы



Особенности:

- В диапазоне $|\text{score}| < 0.5$ находится примерно 1/3 всех колл-центров, поэтому оценка точности метода порядка 2 успеха из 3-х – что независимо подтвердилось при проверке гипотезы №2.
- Наиболее ярко выражены предпочтения типа у тарифного плана С – видимо туда попали in-house SLA ТП производителей.
- У тарифов А и В почти поровну скоринга разных знаков.
- Вывод: для 3/4 колл-центров расчет эффективности операторов только по одному из KPI был бы неверен!

Целевое исследование построения смешанного скоринга эффективности операторов по всем KPI

ПОСТАНОВКА ЗАДАЧИ

- Мотивационные KPI для операторов разбиваются на два типа услуг:
 - Горячая линия,
 - Целевой обзвон.
- Так как идея с разделением колл-центров забракована, то нужно предложить логику замешивания разных KPI в общий скоринг эффективности оператора.
- Результирующий скоринг нужно проверить на эффективность разделения операторов и прогноза утилизации рабочего времени.

ПРИМЕНЕННЫЙ МЕТОД

- Поскольку типы вызовов хорошо стратифицируются по типу услуги, то каждой записи датасета можно поставить в соответствие только один KPI.
- Пороги по KPI определяются из анализа всего датасета как:
 - Г.Л. – 90% перцентиль распределения ожидания,
 - Ц.О. – 25% от средней выработки за сутки.
- Далее определяется дневной скоринг оператора по нарушению порога с весами количеств вызовов данного типа.
- Дневной скоринг усредняется по рабочим дням оператора.

РЕЗУЛЬТАТ(Ы)

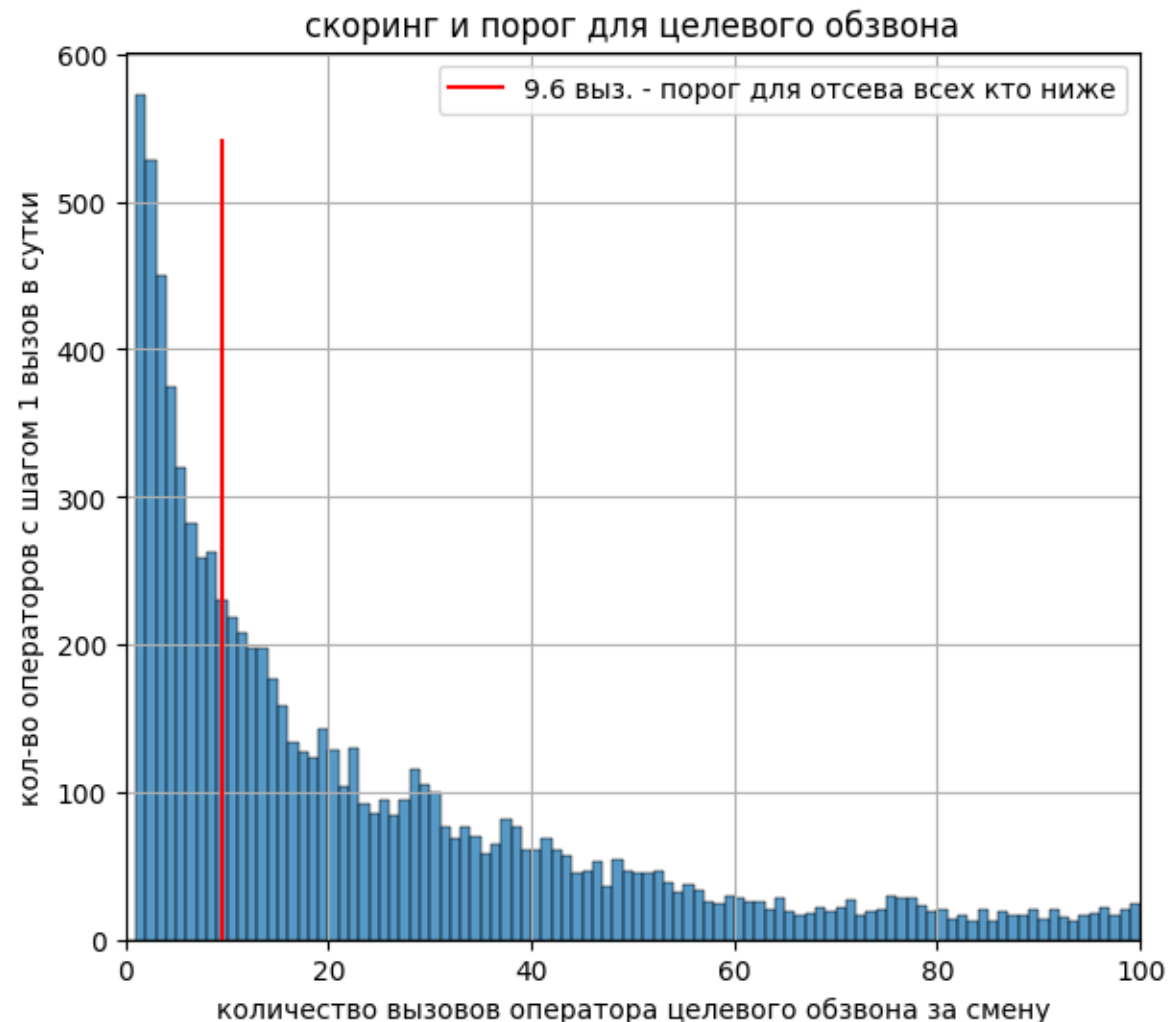
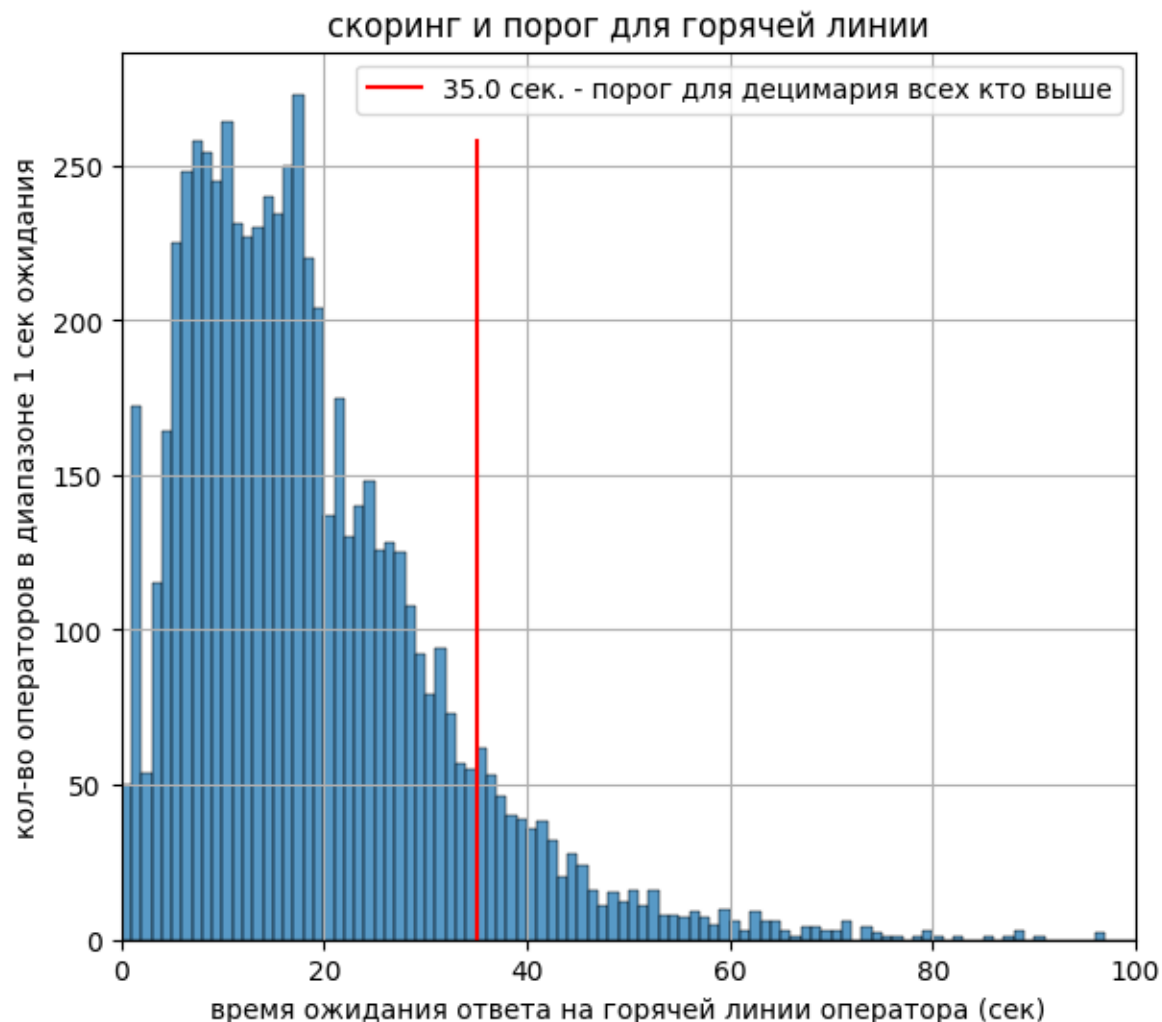
- Скоринг принимает значения от 0 до 1: 1 означает полную неэффективность.
- Скоринг оказался чрезвычайно эффективным:
- 15.5% операторов полностью неэффективны (черный список)
- 23.0% операторов полностью эффективны (список премий)
- 61.5% операторов частично эффективны – по одному из видов услуг
- Утилизация рабочего времени операторов черного списка на порядок ниже таковой у остальных – статистическая достоверность более 10 сигм!

Определение порогов по статистике основного датасета

Идея – посмотреть некоторые типичные значения и предложить разумный порог:

Для горячей линии порог определяет децимарий 10% самых плохих операторов

Для целевого обзвона порог определяет 25% от нормы выработки (среднего)



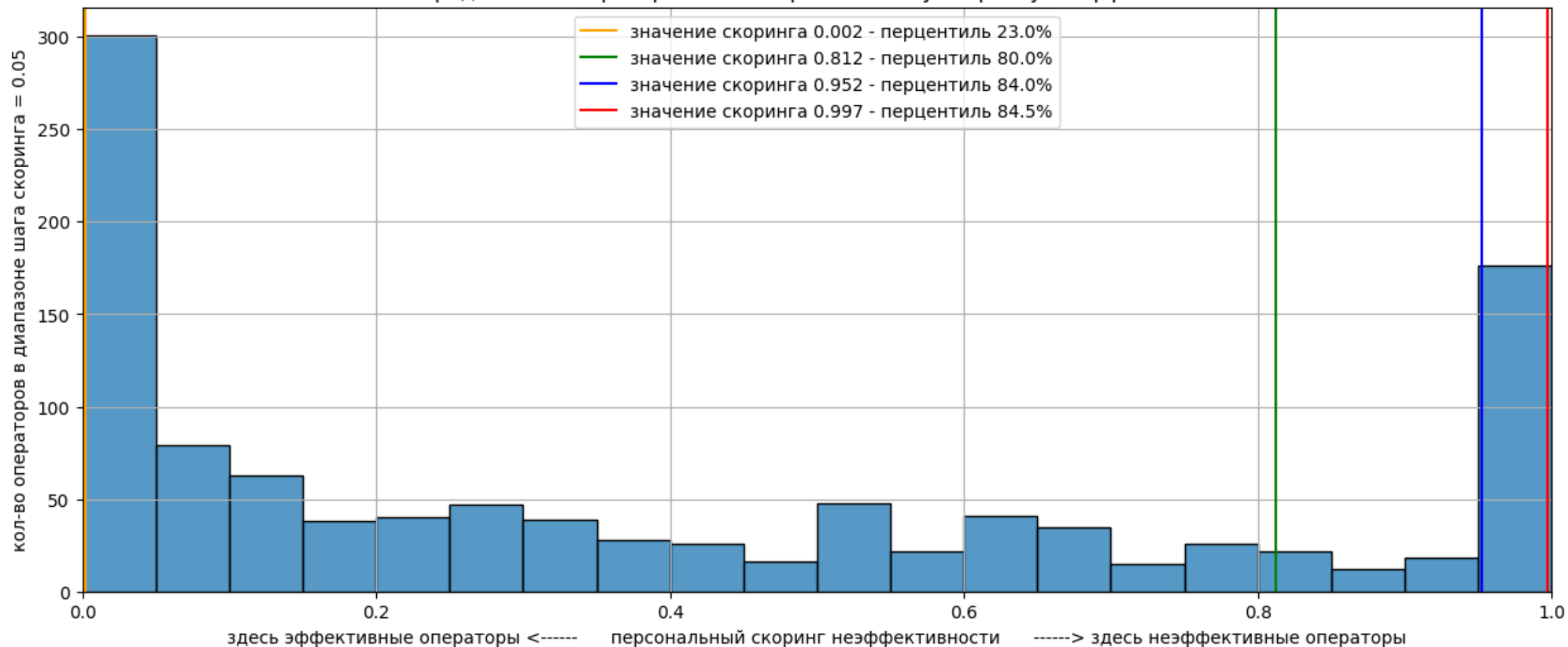
Расчет смешанного скоринга неэффективности операторов

15.5% - абсолютно неэффективных операторов (черный список кандидатов на увольнение)

23.0% - полностью эффективных операторов (белый список на премию)

61.5% - операторов эффективных по одному из KPI и неэффективных по другому

Распределение операторов по их персональному скорингу неэффективности



Целевое исследование статистики не проключенных на операторов вызовов

ПОСТАНОВКА ЗАДАЧИ

- Более половины потерянных входящих вызовов от абонентов приходится не маршрутизируются на операторов.
- Таким образом основной вклад в потерю звонков на горячие линии вносят не операторы, а проблемы на IVR.
- Требуется более детально исследовать статистику распределения проблем с IVR по колл-центрам.

ПРИМЕНЕННЫЙ МЕТОД

- Никакой специальной методологии не используется – проведена обычная EDA по срезу записей основного датасета с пропущенным ID оператора.
- На основе агрегированных данных из среза выше также рассчитаны по каждому колл-центру:
 - среднее время ожидания,
 - количество потерянных вызовов на оператора,
 - соотношение вызовов, обслуженных операторами к потерянным на IVR
- Графический анализ не нужен

РЕЗУЛЬТАТ(Ы)

- IVR вносит основной вклад в потерю вызовов на Г.Л.
- Потерянные до проключения на операторов вызовы имеет 304 колл-центров из 306 всего.
- 3/4 всех колл-центров теряет более 10% вызовов абонентов горячей линии на IVR.
- 1/4 всех колл-центров теряет более 50% вызовов абонентов горячей линии на IVR.
- 5% всех колл-центров просто не хватает операторов – время ожидания в очереди IVR превышает порог децимария для времени ожидания в очереди к оператору (35 сек).

Проверка гипотез

Гипотеза №1: Предложенные продуктовым маркетингом мотивационные KPI эффективности операторов также приводят к повышению утилизации операторов

Результат: ПОДТВЕРЖДЕНА

Выводы:

Мы можем подтвердить с любой разумной статистической достоверностью, что утилизация рабочего времени 15.5% неэффективных операторов, отобранных нашим методом, на порядок меньше утилизации остальных операторов, не попавших в черный список.

Данный успех в подтверждении гипотезы является доказательством эффективности как нашего метода скоринга эффективности операторов, так и выбора KPI для мотивации операторов - в плане его соответствия бизнес-цели клиента по снижению расходов на зарплаты операторам (за счет увеличения утилизации их рабочего времени).

Гипотеза №2 У колл-центров, обслуживающих звонки на горячую линию по остаточному принципу, (т.е. у колл-центров по типу целевого дозвона) время ожидания на IVR больше, чем у колл-центров по типу горячей линии

Результат: НЕПОДТВЕРЖДЕНА

Выводы:

Нам не хватило статистической мощности теста, чтобы подтвердить статистическую значимость обнаруженного превышения среднего времени ожидания на IVR для колл-центров, предоставляющих услугу горячей линии по остаточному принципу - по сравнению с колл-центрами, для которых это основная услуга.

Данный результат еще раз подтвердил правильность нашего заключения о нецелесообразности подхода к скорингу на основе присваивания всему колл-центру и его операторам только основного типа услуги.

Заключение по основным результатам проекта

УДАЛОСЬ

Построен эффективный персональный скоринг для определения неэффективных операторов колл-центра – смешанный сразу по двум услугам: горячей линии и целевого обзвона.

Доказано, что мотивация операторов по предложенным продуктовым маркетингам KPI для каждой из услуг статистически достоверно приводит к росту утилизации операторов – тем самым и к достижению бизнес-целей по снижению расходов компании.

Обнаружено, что основной вклад в потери звонков абонентов на горячие линии вносят не операторы, а проблемы с IVR, принимающим и проключающим вызовы на операторов.

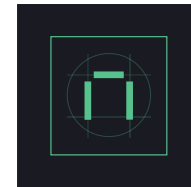
Опробован критерий определения нехватки операторов в колл-центре - определены проблемные колл-центры с нехваткой ресурса.

НЕ УДАЛОСЬ

Разделить колл-центры по типу услуг на две группы. Точнее, вторая группа получилась принципиально смешанная по типу услуг.

Выяснить причину таких больших потерь на IVR – возможно, что это и не потери, а ситуация, когда абонент получил автоматический ответ в голосовом меню IVR, хотя среднее время проведенное на IVR не соответствует данному предположению.

Объяснить присутствие загадочных “стахановцев” с суточной выработкой, превышающей продолжительность смены – возможно, это относится к случаям, когда один эккаунт в системе колл-центра переиспользуется разными физическими операторами, хотя наличие оператора с чистым временем разговоров с абонентами за сутки выше 24 часов не подходит и под данное предположение.



СПАСИБО ЗА ВНИМАНИЕ !