

Тестовое задание Margarito

Апрель 2020 г.

Задача

Есть **гипотеза**, что Margarito теряет лояльность пользователей из-за неэффективности текущей системы поддержки:

Либо ответы даются слишком долго, либо проблема пользователя не решена.

Задача:

— Подтвердить или опровергнуть гипотезу.

Если гипотеза подтвердится, то оценить примерный урон для сайта Авито: в недополученных кол-ве объявлений и пользователях.

Вывод

Гипотеза **подтверждается**.

Существенная часть пользователей теряет свою лояльность к сайту Margarito из-за неэффективной системы тех.поддержки.

Несмотря на преобладание положительных оценок, наблюдается выраженная часть неудовлетворительных оценок — **20.5%**.

Нелояльные пользователи оставляют негативные оценки тех.поддержке из-за нерешённости проблемы, либо превышения терпимого времени ожидания ответа. Время ожидания для нелояльных пользователей составляет **7 часов, а иногда и 4 дня**. Бывают случаи, когда ответы даются через **20 и 30 дней**.

Мы проверяли нашу гипотезу о разнице во времени ожидания ответа для лояльных и нелояльных пользователей на уровень статистической значимости. Среднее время ожидания ответа в группе нелояльных пользователей значительно **дольше**, в сравнении с другой группой. Разница статистически значима.

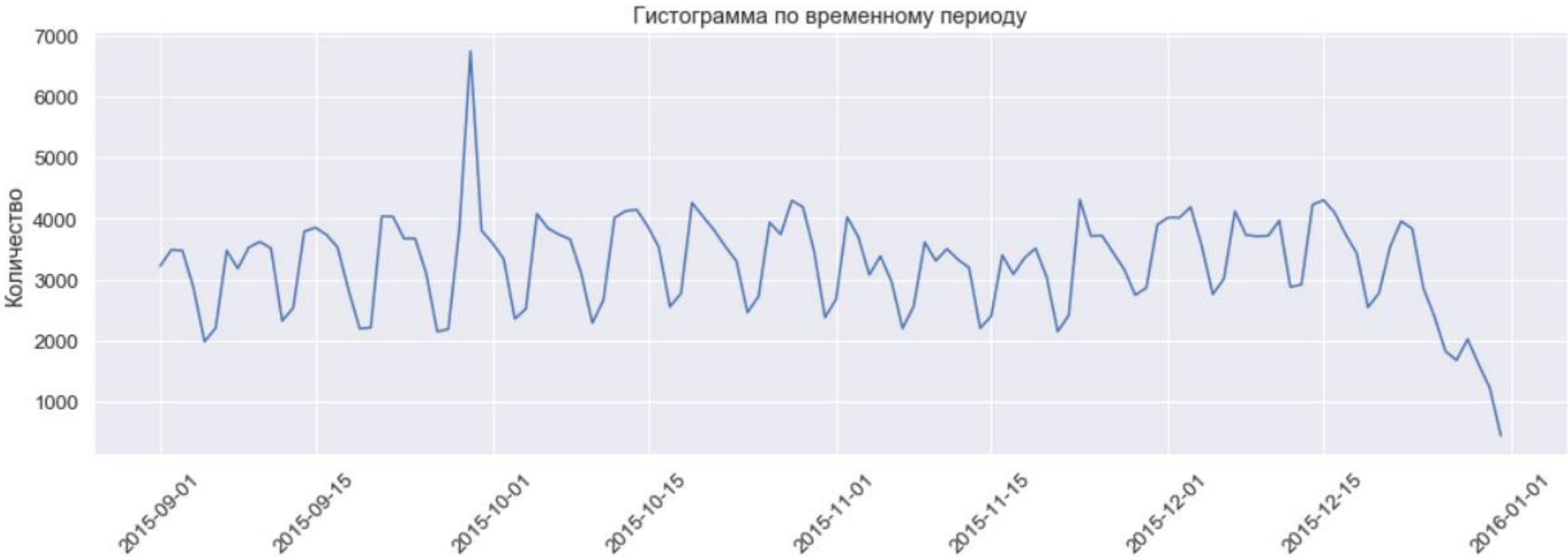
В целом, нелояльных пользователей **меньше**, в сравнении с лояльными. Но они склонны публиковать **больше объявлений**. Гипотезу о разнице в среднем количестве объявлений между группой лояльных и нелояльных пользователей мы проверили на уровень статистической значимости.

Мы оценили примерный урон для сайта Margarito из-за нелояльных пользователей:

- недополученное количество объявлений — **441 090 — 11.1%** от общего количества за представленный период;
- недополученное количество пользователей — **16 965 — 6.8%** от общего количества;

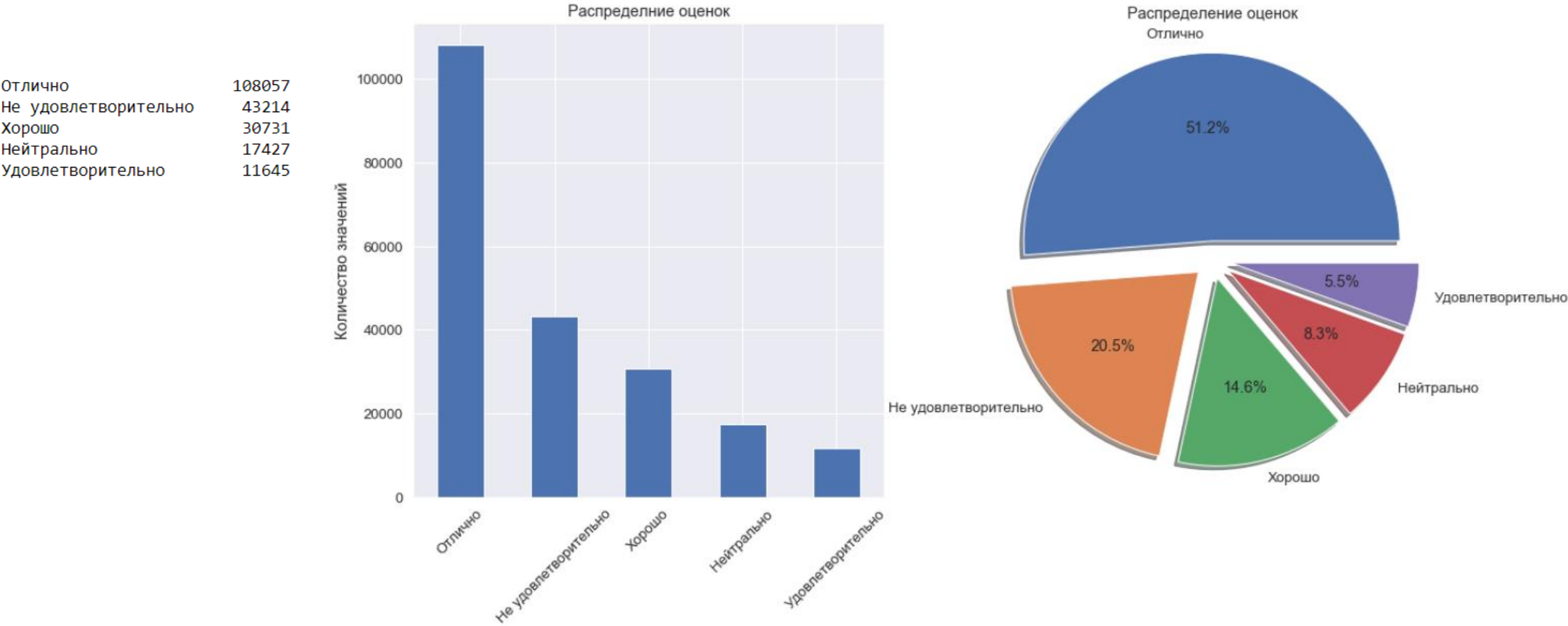
Шаг 1. Провели предобработку данных и провели EDA

- Удалили пропуски и дубликаты;
- Привели дату к нужному формату;
- Изучили временной период обращений;



Полные данные представлены с 1го сентября 2015 по конец декабря 15го.

Изучили оценки содержащиеся в ответах пользователей и их количество



В значительной степени преобладают положительные оценки от пользователей — 51.2% — «отлично» и 14.6% — «хорошо». Однако, наблюдается выраженная часть неудовлетворительных оценок — 20.5%.

Таким образом, предположительно, негативная оценка от пользователей может отрицательно повлиять на лояльность.

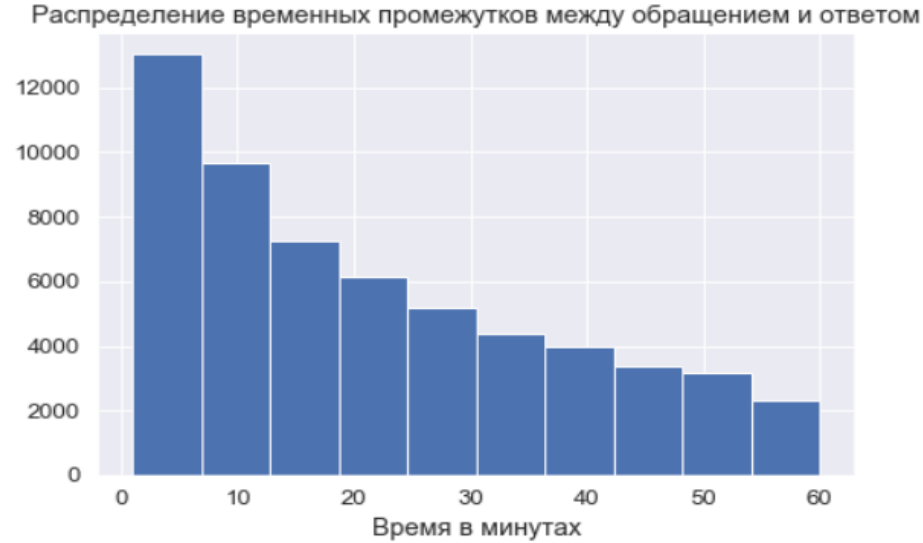
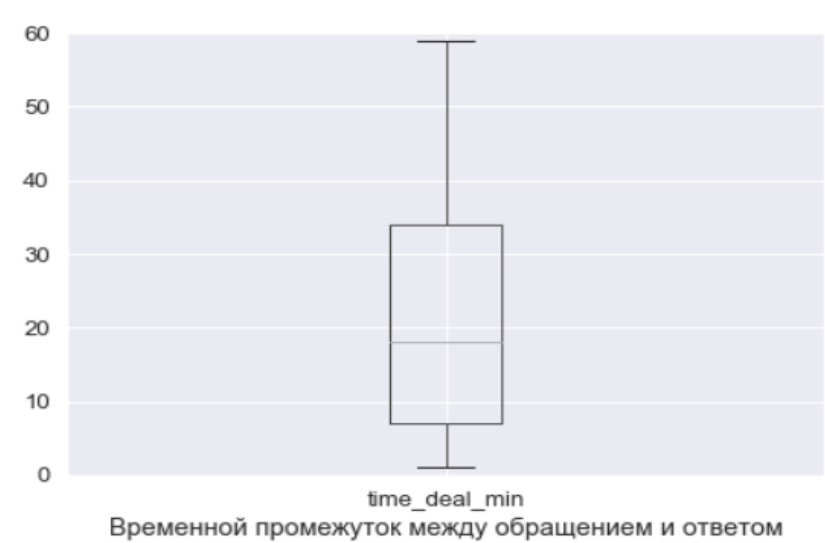
Детально изучили среднее время ожидания ответа от тех.поддержки по минутам, часам, дням. Также, изучили долю ожидания ответов

В основном, время ожидания ответа от тех.поддержки для всех пользователей ранжируется на несколько групп (по минутам, часам и дням):

- Есть быстрые ответы через 4-8 минут. Но среднее время ожидания ответа составляет 18 минут — доля в 19%
- Есть быстрые ответы через 1-5 часа. Но среднее время ожидания ответа составляет 6 часов — доля в 80%
- Есть быстрые ответы через 1-3 дня. Но среднее время ожидания ответа составляет 3 дня — доля в 42%

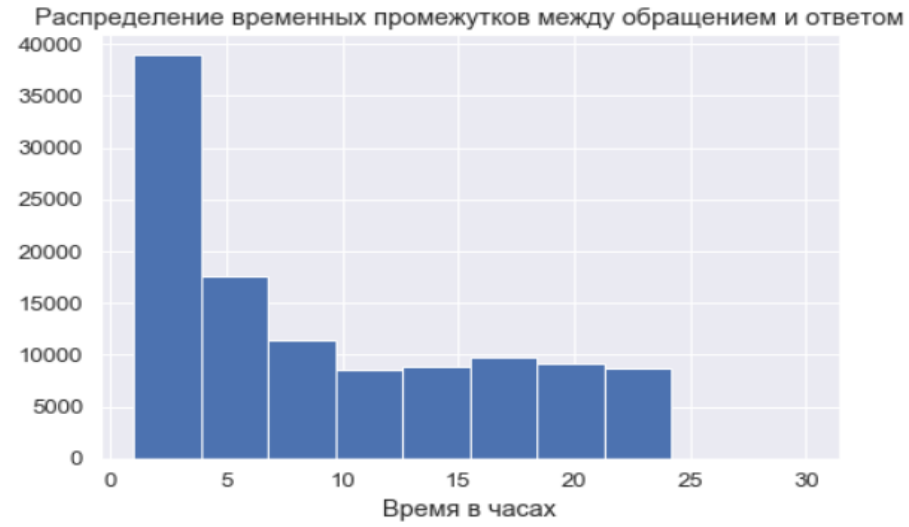
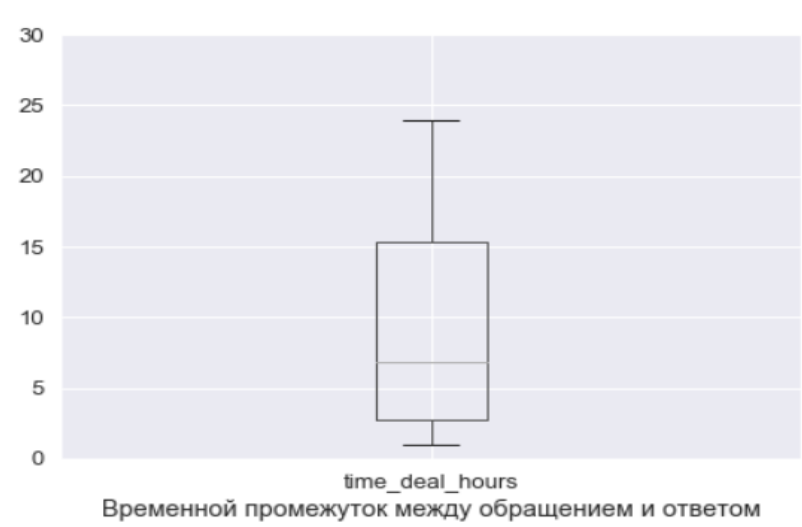
Таким образом, большая часть ответов приходится на длительное время ожидания от 6 часов до 3х дней. На данный момент такая работа тех.поддержки не эффективна и негативно сказывается на лояльности пользователей.

Среднее время ожидания ответа от тех.поддержки по минутам, часам, дням

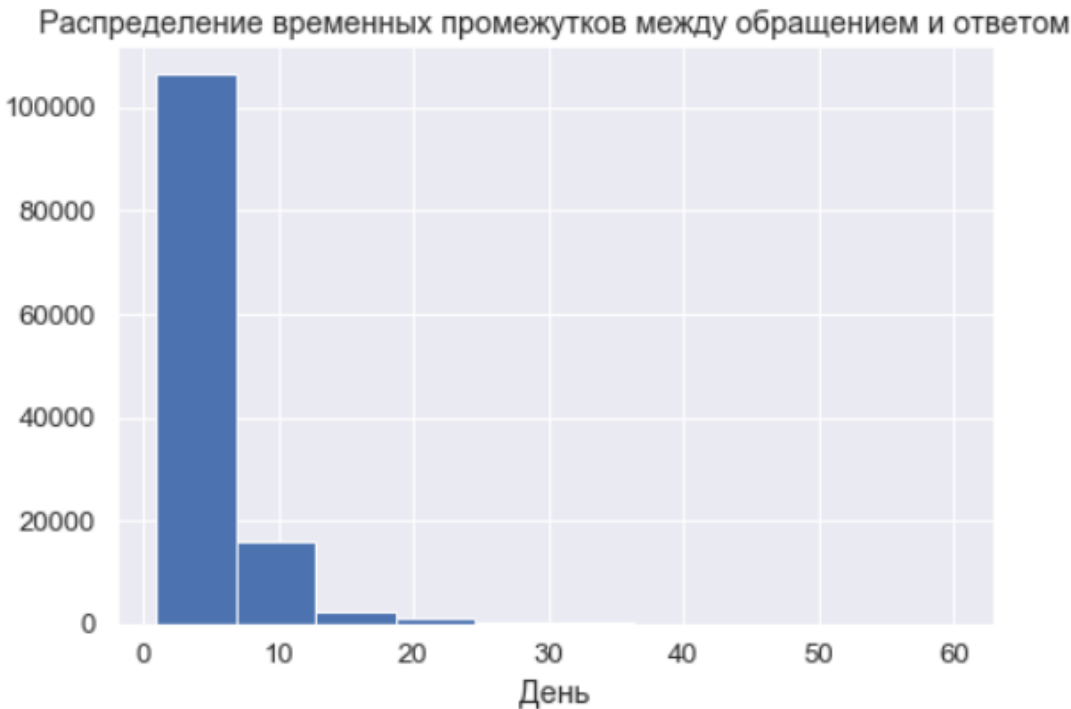
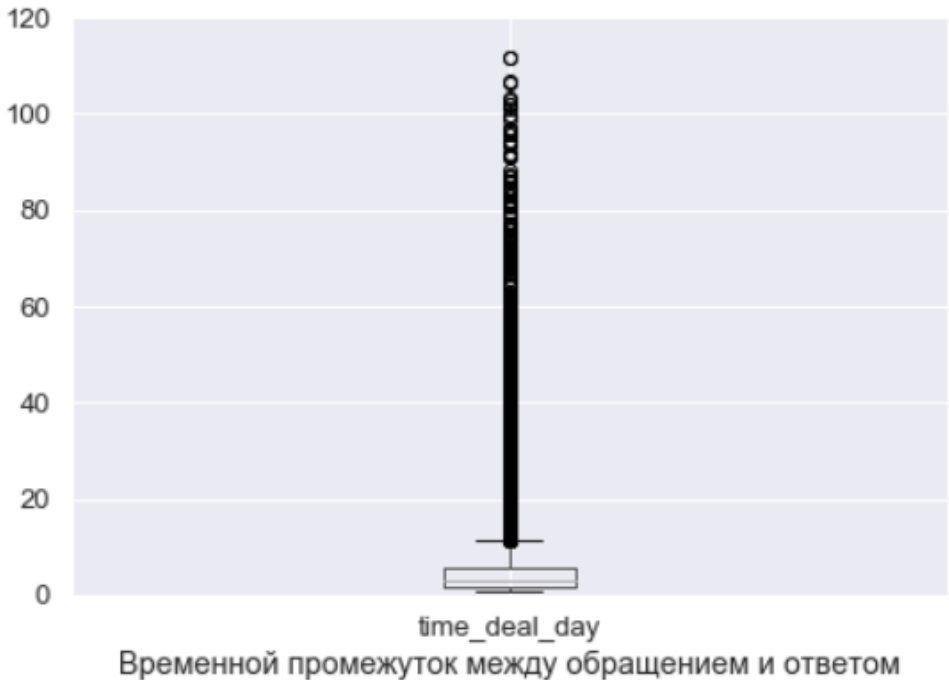


большая часть пользователей получают ответ через 4-8 минут. В среднем, после обращения пользователи получают ответ через 18 минут.

Основная часть пользователей получает ответ через 1-4 часа. В среднем, через 6 часов.



Распределение времени обращения в часах



В целом, пользователи получают ответ через 1-3 дня. В среднем, через 3 дня. Однако, есть значения через 10 и 20 дней.

Шаг 2. Изучили влияние негативной оценки и времени

Предположим, что:

- случаи положительной оценки и опр. временной показатель соответствуют решению проблемы;
- случаи отрицательной оценки и длительное время ожидания соответствуют недовольству в отношении превышения терпимого времени;
- случаи отрицательной оценки и короткого времени ожидания ответа соответствуют нерешённой проблеме;

Вывод

В целом в группе пользователей с негативной оценкой преобладает длительное время ожидания ответа от тех. поддержки в сравнении с другой группой.

Среднее время ожидания ответов в минутах, часах, днях:

Группа с положительной оценкой / группа с отрицательной оценкой

18 минут / 20 минут

5 часов / 7 часов

2 дня / 4 дня

Таким образом, в группе с негативной оценкой либо не решена проблема пользователя, либо превышено терпимое время ожидания ответа.

Проверили нашу гипотезу методами статистики

Анализируем различие в среднем времени ответа по двум группам.

Для проверки нормальности распределения использовали тест Андерсона-Дарлинга.

Для проверки гипотезы
использовали t- тест Стьюдента.
Значение альфа-Кронбаха 5%

р-значение: 5189.0

Не получилось отвергнуть нулевую гипотезу, всё нормально

р-значение: 4.8297188846595226e-260

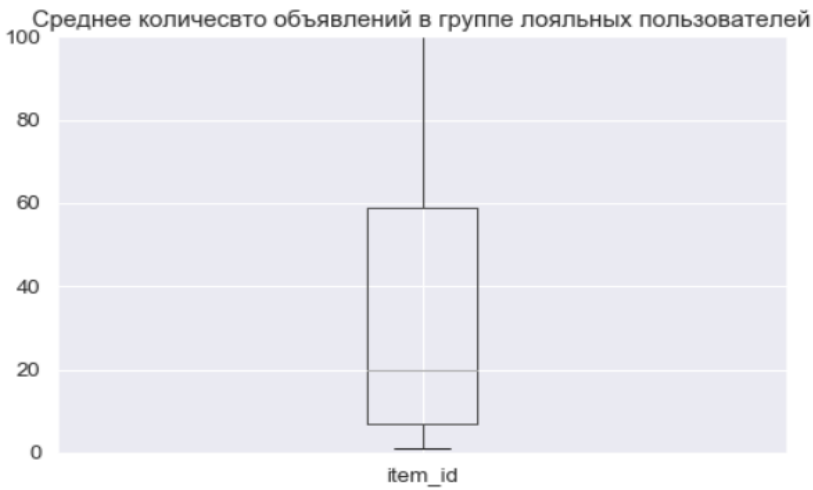
Отвергаем нулевую гипотезу: разница статистически значима

Шаг 3. Изучили негативное влияние на сайт Margarito от группы «нелояльных» пользователей

- Выделили группу нелояльных пользователей с негативной оценкой, по которым будем сравнивать урон для сайта
- Изучили распределение среднего количества публикуемых объявлений для каждого пользователя.

Примерно 16 объявлений публикует каждый пользователь

- Объединили два фрейма и выделили две группы: с неудовлетворительной оценкой и все остальные.
- Для двух групп сделали группировку по уникальным пользователям и рассчитали количество объявлений для каждого пользователя.
- Нашли среднее количество объявлений для пользователей в каждой из групп и сравнили



Группа нелояльных пользователей в среднем размещает больше объявлений, чем в другая. 26 против 20.

Проверили нашу гипотезу методами статистики. Для проверки использовали t- тест Стьюдента

Анализируем различие в среднем количестве объявлений по двум группам

р-значение: 0.002672597377015311

Отвергаем нулевую гипотезу: разница статистически значима

- Рассчитали количество пользователей в каждой из групп
- Умножили среднее количество объявлений на количество пользователей, в каждой из групп. Провели сравнение по примерному недополученному количеству объявлений.
- Нашли доли от общего количества

Количество нелояльных пользователей: 16965

Доля нелояльных пользователей от общей выборки: 6.8 %

Количество объявлений от нелояльных пользователей: 441090

Доля объявлений от нелояльных пользователей от общей выборки: 11.1 %

Вывод

В целом, из-за снижения лояльности пользователей, сайт Margarito может потерять примерно 16 965 пользователей — 6.8% от общего количества за представленный период. И примерно 441 090 объявлений — 11.1% от общего количества.

Несмотря на полученные результаты, мы должны понимать, что данный анализ урона является примерным, так как нам неизвестны средняя выручка по объявлениям, покупка рекламы и продвижения, платное размещение и т.д. Также, мы не учитывали группу нейтральных оценок и тех, кто не оставлял оценок. Бывает так, что такие безучастные пользователи быстро «отваливаются» от сайта.

Мы рекомендуем разработать уникальный классификатор запросов для тех. поддержки. Он позволит ранжировать обращения, выделяя срочные и наоборот. Возможно, это поможет в эффективности работы.