1. Министерство науки и высшего образования Российской Федерации
2. Санкт-Петербургский политехнический университет Петра Великого
3. —
4. Институт прикладной математики и механики
5. Кафедра «Информационная безопасность компьютерных систем»

**Лабораторная работа №4**

**«Исследование инструментов сбора статистики клиентах сайта, проведение маркетинговых»**

1. по дисциплине «Разработка интернет-приложений»
2. Выполнили
3. студенты гр.  43609/3 Корольков А.А

<*подпись*>

1. Романюк Ю.В.

<*подпись*>

1. Проверил
2. преподаватель Печенкин А.И.

<*подпись*>

1. Санкт-Петербург
2. 2019

# Формулировка задания

Изучить функционал, предоставляемый современными инструментами аналитики для веб-сайтов.

**Задачи**

1. Подключить инструмент аналитики к реализованному сайту.
2. Сгенерировать искусственной трафик пользователей на сайт.
3. Исследовать функционал, предоставляемый выбранным средством аналитики.
4. Исследовать статистики, применяемые в product management.

# Результаты работы

В ходе выполнения 1 лабораторной работы было принято решение разрабатывать онлайн сервис коротких текстовых заметок. Для выполнения лабораторной работы 4 изначально были разработаны программы-боты:

* 1-ая- осуществляет создание аккаунтов для ботов;
* 2-ая- производит генерацию искусственного трафика.

При написании ботов использовался инструмент Selenium IDE, позволяющий автоматизировать действия пользователя на странице. Запуск ботов для генерации трафика осуществляется в 3 потока.

Алгоритм работы бота, осуществляющего регистрацию:

1. создается элемент класса Chrome: google\_chrome= webdriver.Chrome()
2. происходит перенаправление на страницу регистрации: google\_chrome.get(url)
3. для i от 0 до 100 выполняются следующие действия:
   1. поиск элементов по имени: поле ввода для полного имени пользователя, поле ввода для email, поле ввода для логина, поле ввода пароля, кнопка регистрации:

full\_name = google\_chrome.get\_elem\_by\_name('full\_name')

email = google\_chrome.get\_elem\_by\_name('email')

username = google\_chrome.get\_elem\_by\_name('username')

password = google\_chrome.get\_elem\_by\_name('password')

button = google\_chrome.get\_elem\_by\_name('register')

* 1. заполнение полей:

full\_name.send\_keys('Julia' + str(i))

email.send\_keys('julia' + str(i) + '@gmail.com')

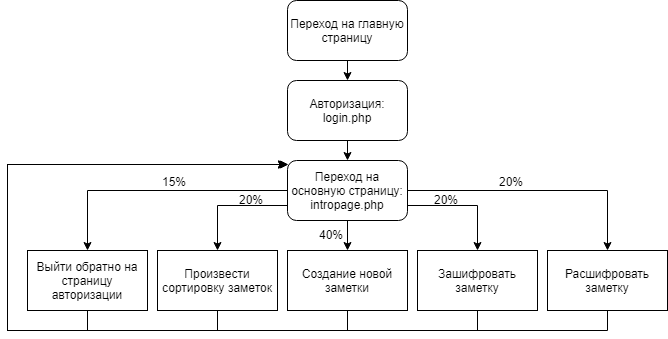
username.send\_keys('Julia' + str(i))

password.send\_keys('1234-julRom')

* 1. нажатие кнопки: button.click()

1. закрытие окна браузера: google\_chrome.quit()

Алгоритм взаимодействия бота для генерации трафика с сайтом:

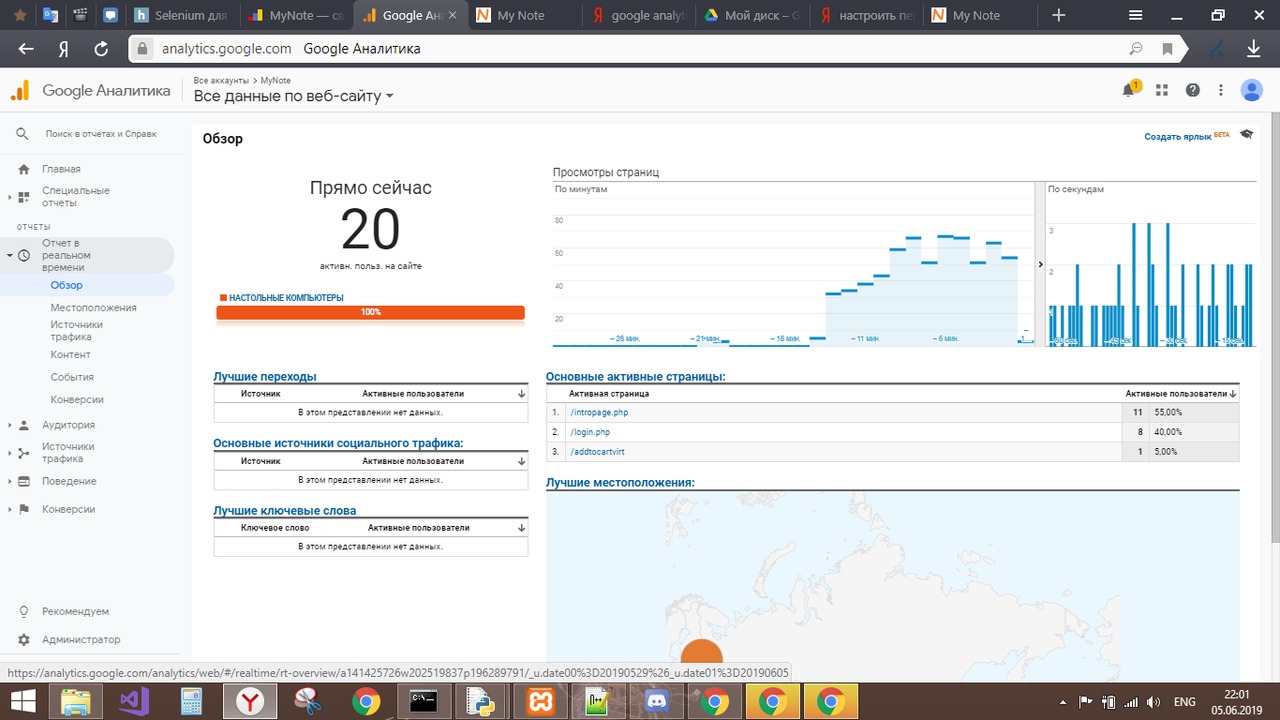


Для выполнения лабораторной работы были выбраны инструменты Яндекс Метрика и Google Analytics. Данный выбор в пользу одновременного использования обусловлен тем, что функционал инструментов несколько отличается. Оба инструмента подключаются к сайту путём добавления JavaScript-кода, реализующего логику сбора аналитики, на все подотчётные страницы сайта.

Так же в ходе сбора аналитики был несколько модифицирован код сайта: добавлены обработчики событий для нажатия кнопок. Пример:

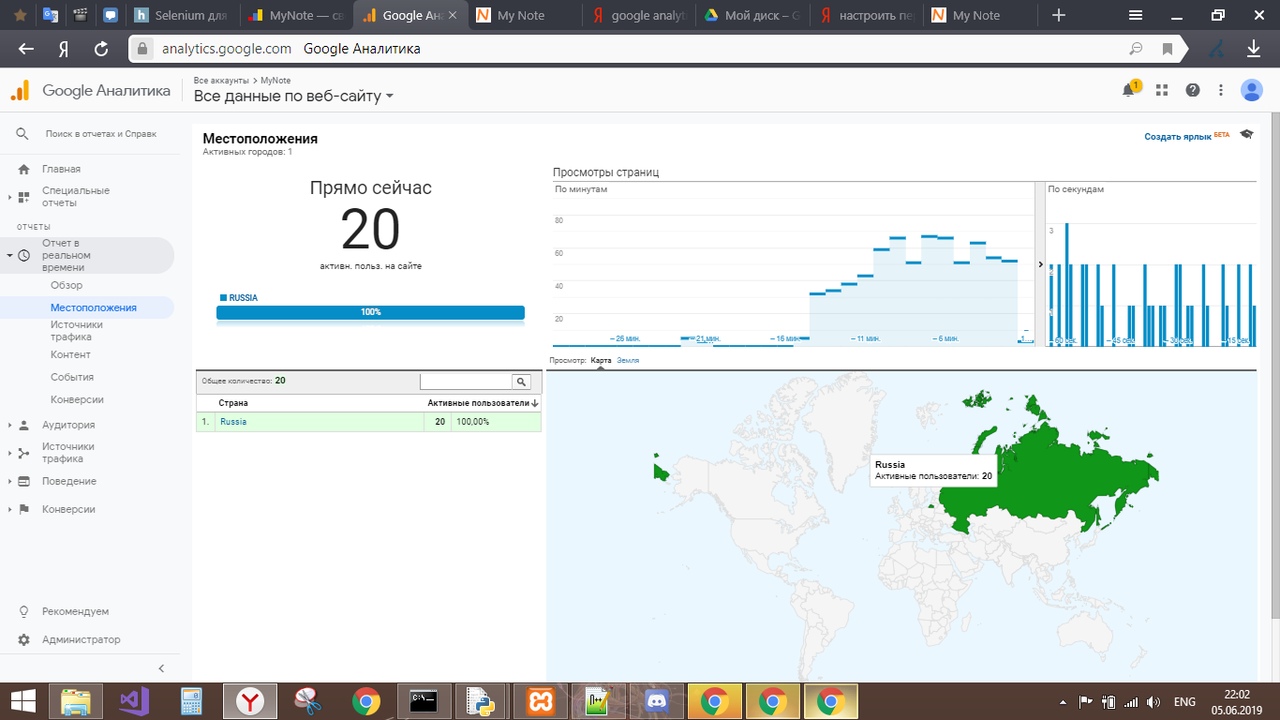
<input type="submit" name="Text" class="btn btn-primary" id="submit\_button" value="Submit" **onClick="gtag('event','page\_view',{'page\_path':'/addtocartvirt'});"** />

Для исследования функционала средств аналитики был написан скрипт запускающий бота для генерации трафика в нескольких потоках в бесконечном цикле. Пример отчета из Google Analytics в режиме реального времени:

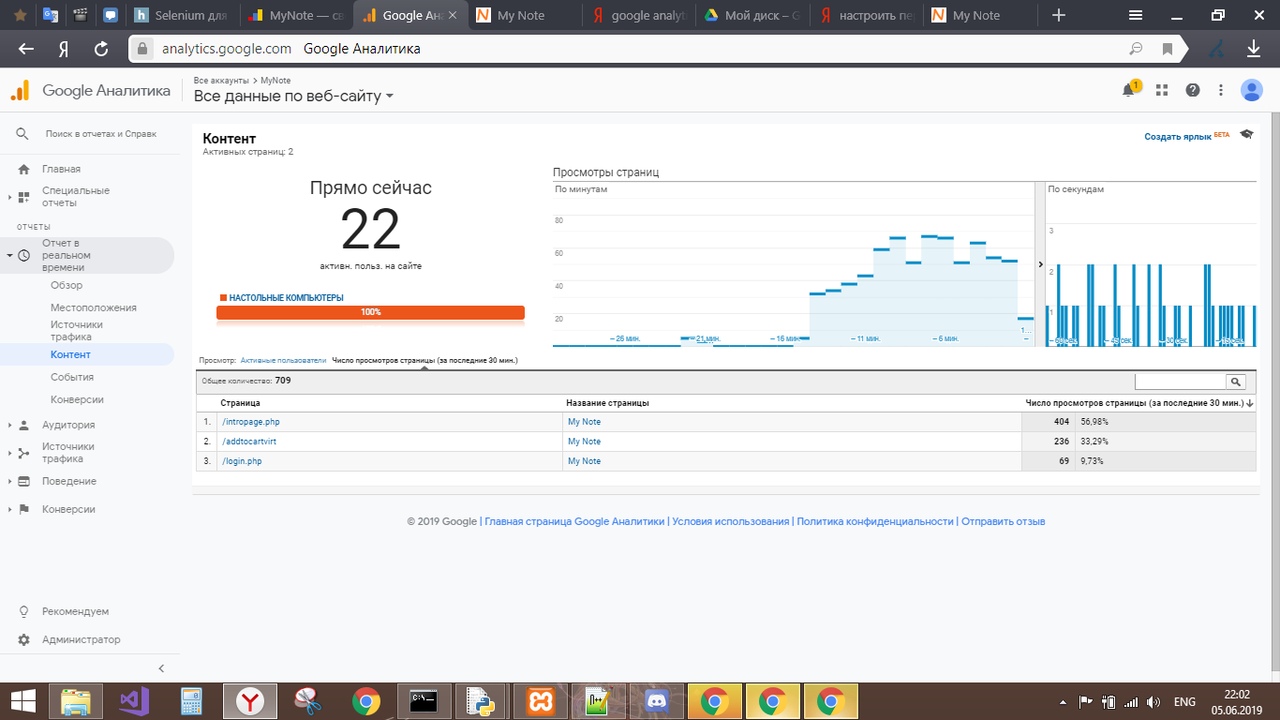


На рисунке выше можно увидеть текущее количество активных пользователей, используемая ими платформа, на какой странице находятся в данный момент.

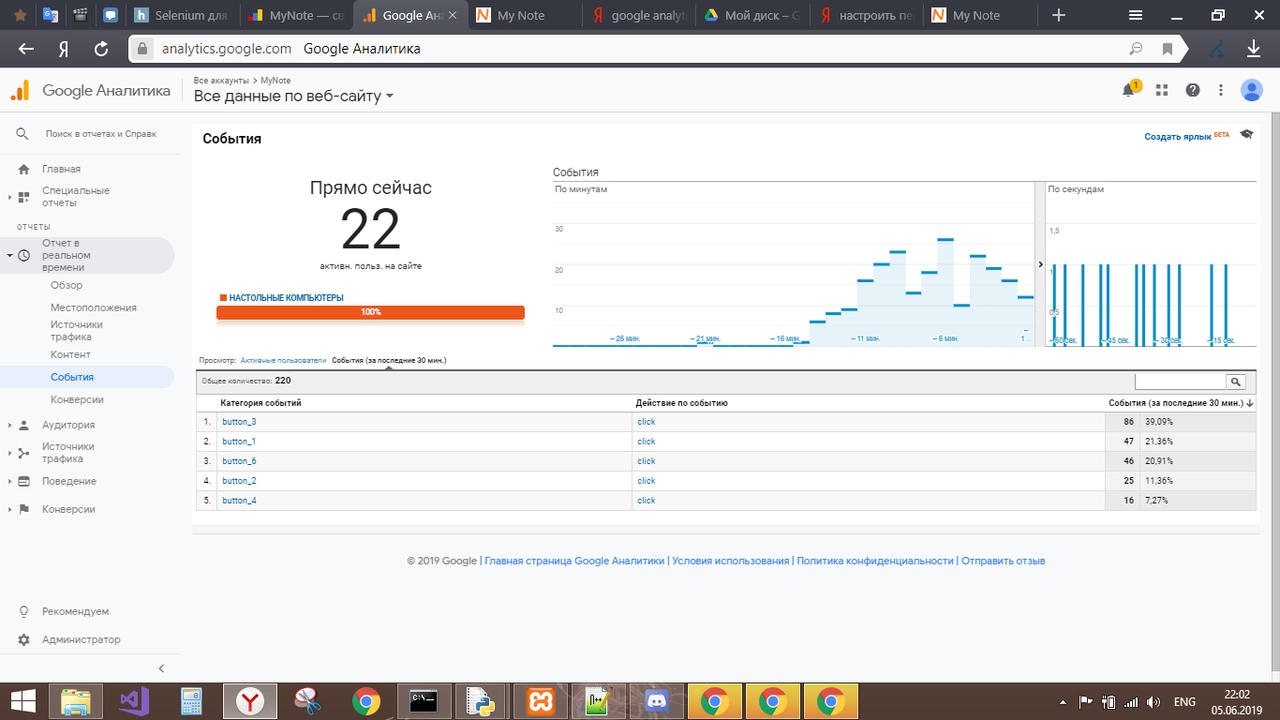
При переходе в режим отслеживания в реальном времени становиться доступно больше информации. Например, предоставляется статистика по географическому распределению клиентов:



Также предоставляется статистика по распределению посетителей по страницам сайта, в данном случае она отображает результат «случайного» обхода сайта ботов:



До запуска ботов, как было сказано выше, на кнопки были добавлены обработчики. Далее в Google Analytics были добавлены цели на выполнение событий - нажатия кнопок. Таким образом, инструмент предоставляет возможность отслеживать действия ботов:



События:

button\_1 – нажатие кнопки «Sort by deadline»

button\_2 - нажатие кнопки «Sort by creation date»

button\_3 - нажатие кнопки «Encrypt»

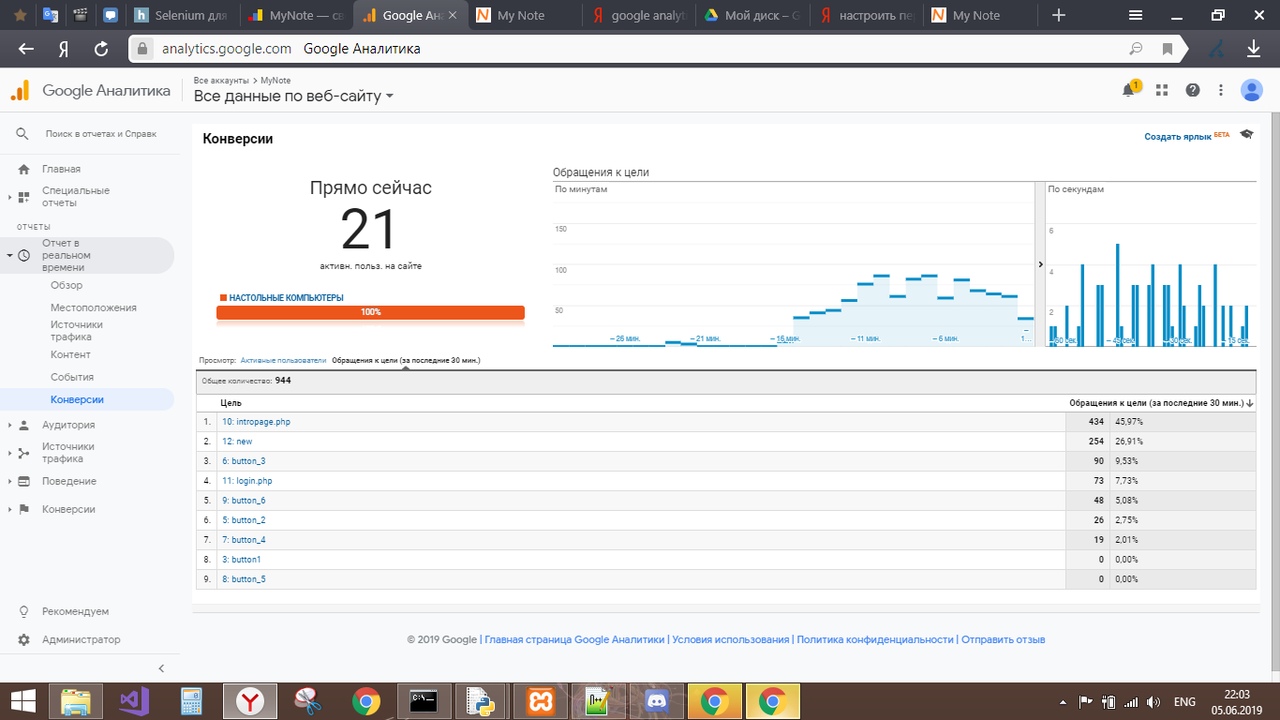
button\_4 - нажатие кнопки «Decrypt»

button\_6 – нажатие кнопки «Logout»

Также были созданы цели для перехода по страницам в определенном порядке. То есть цель intropage – переход на страницу login.php, а затем на intropage.php.

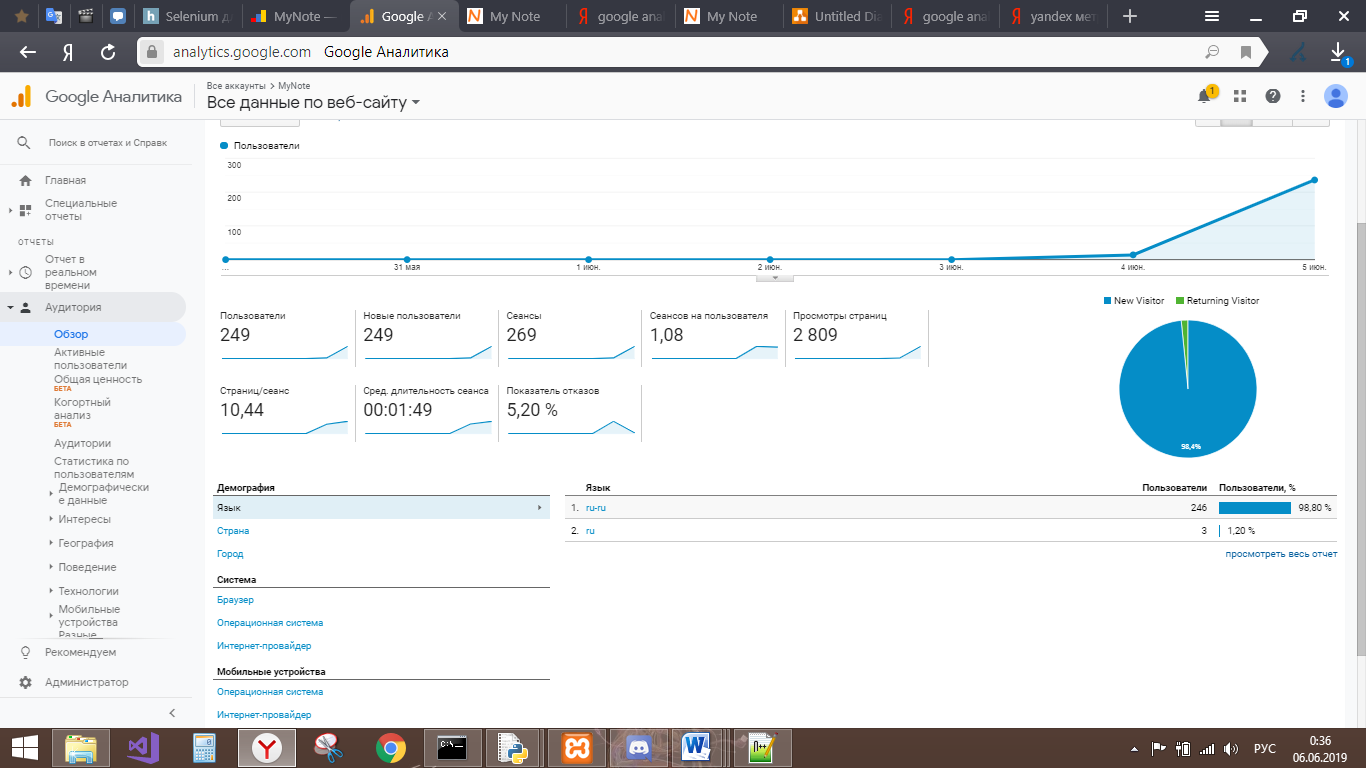
Важную роль здесь играет цель new. Она определяет следующий порядок действий: переход на login.php, переход на intropage.php, нажатие кнопки создания заметки. Чтобы отследить нажатие кнопки создания заметки код был повторно модифицирован. Обработчик нажатия кнопки выполнял переход на виртуальную страницу.

Конверсия по всем целям:

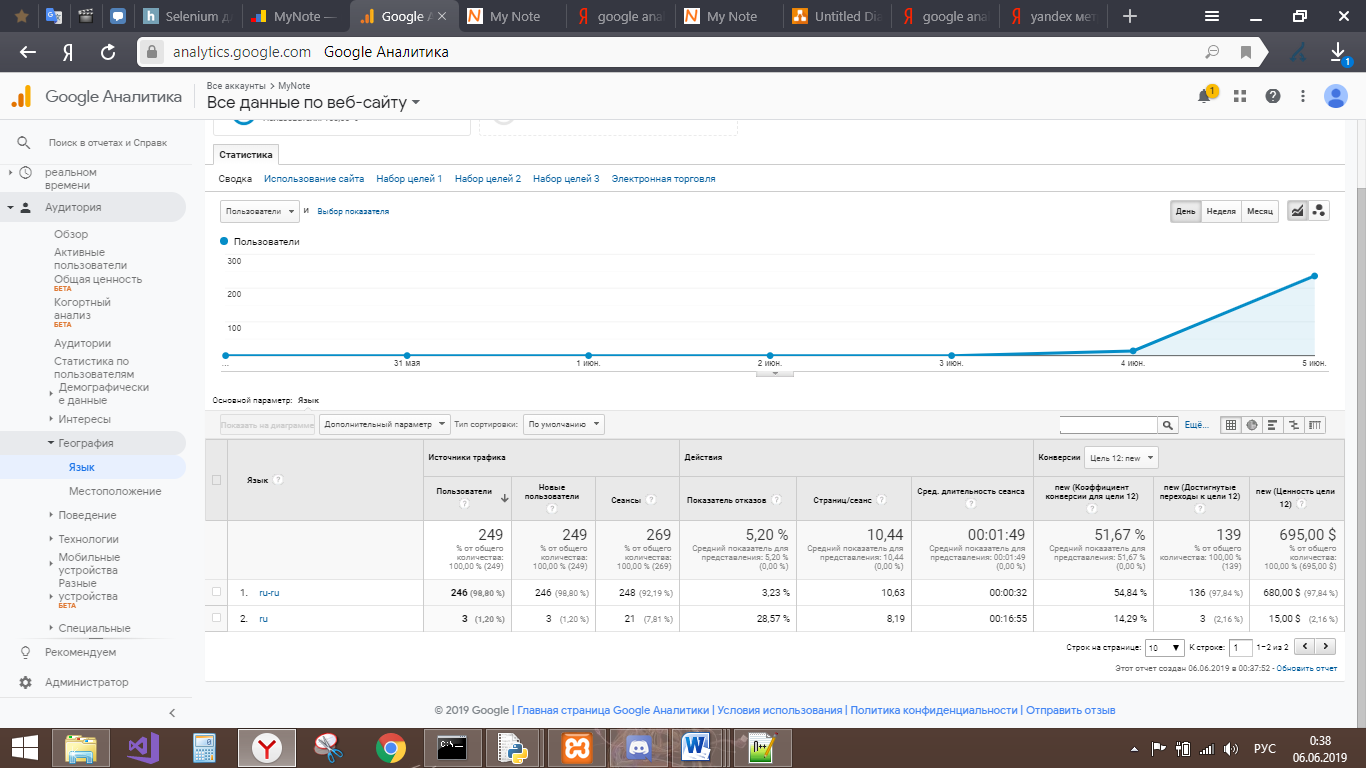


О конверсии (воронке) будет описано ниже.

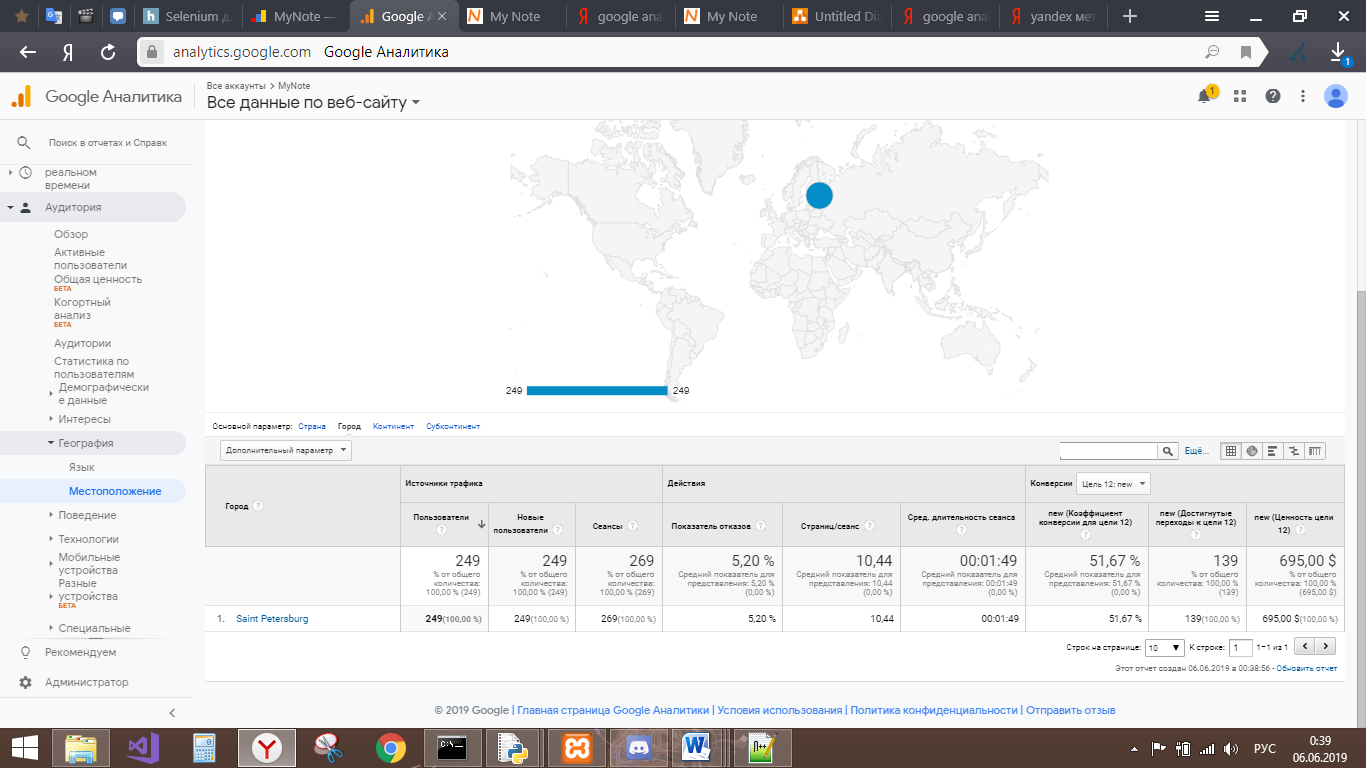
Рассмотрим общую статистику по пользователям:



Распределение по языку:



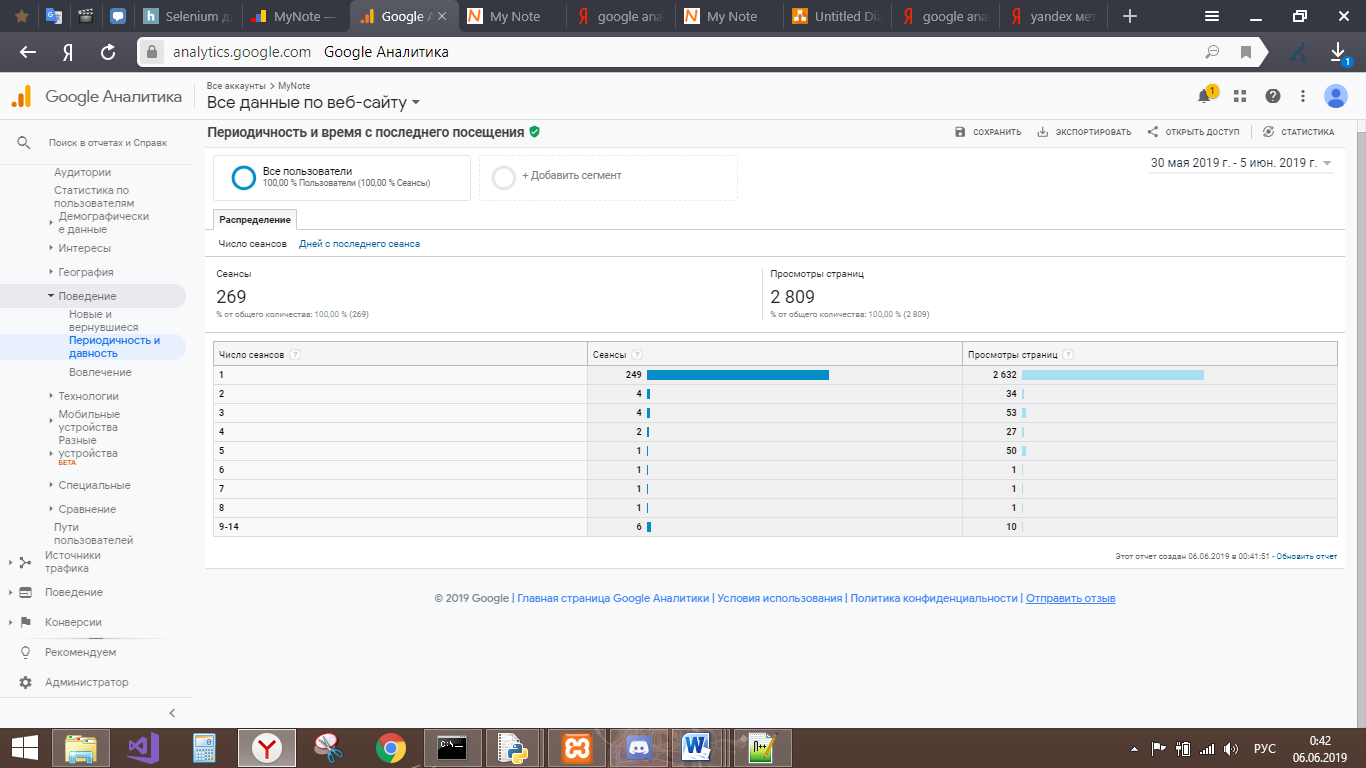
Местоположение:



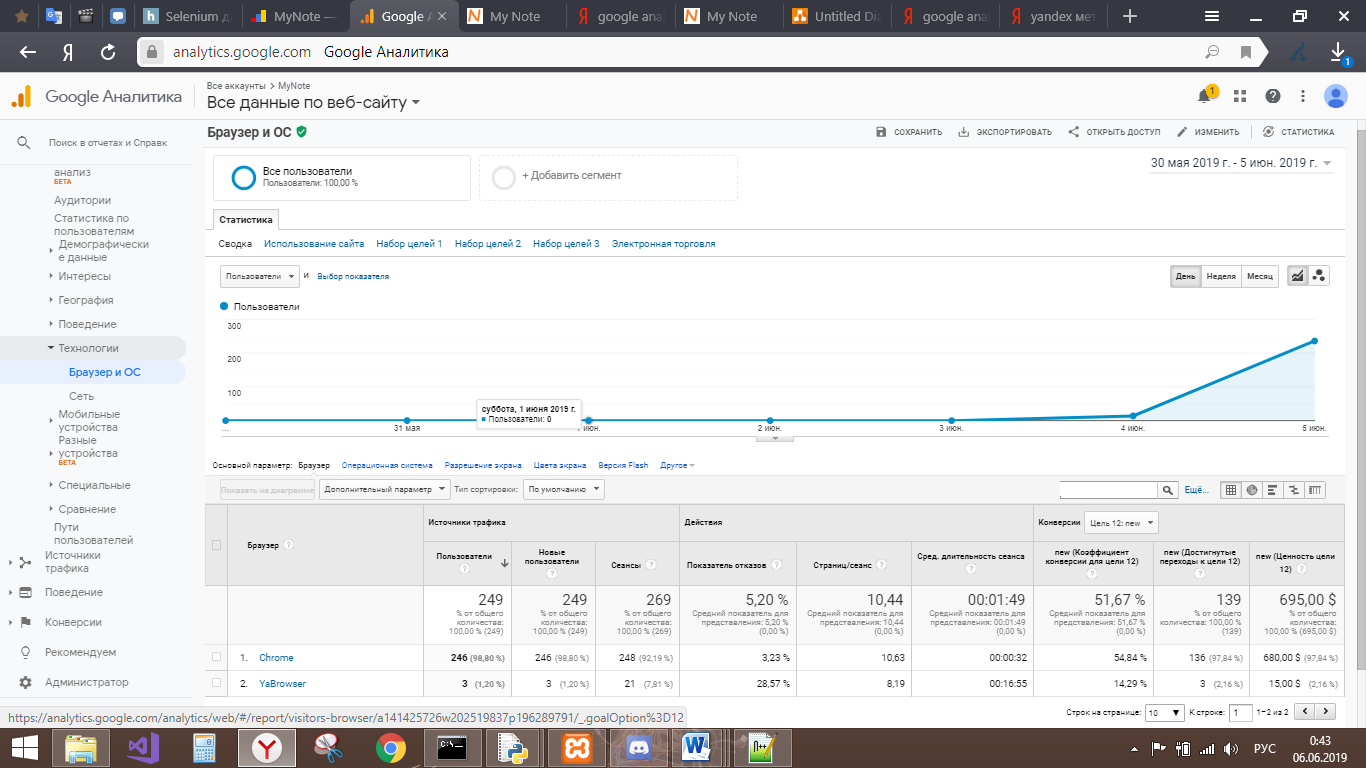
Новые и вернувшиеся пользователи:



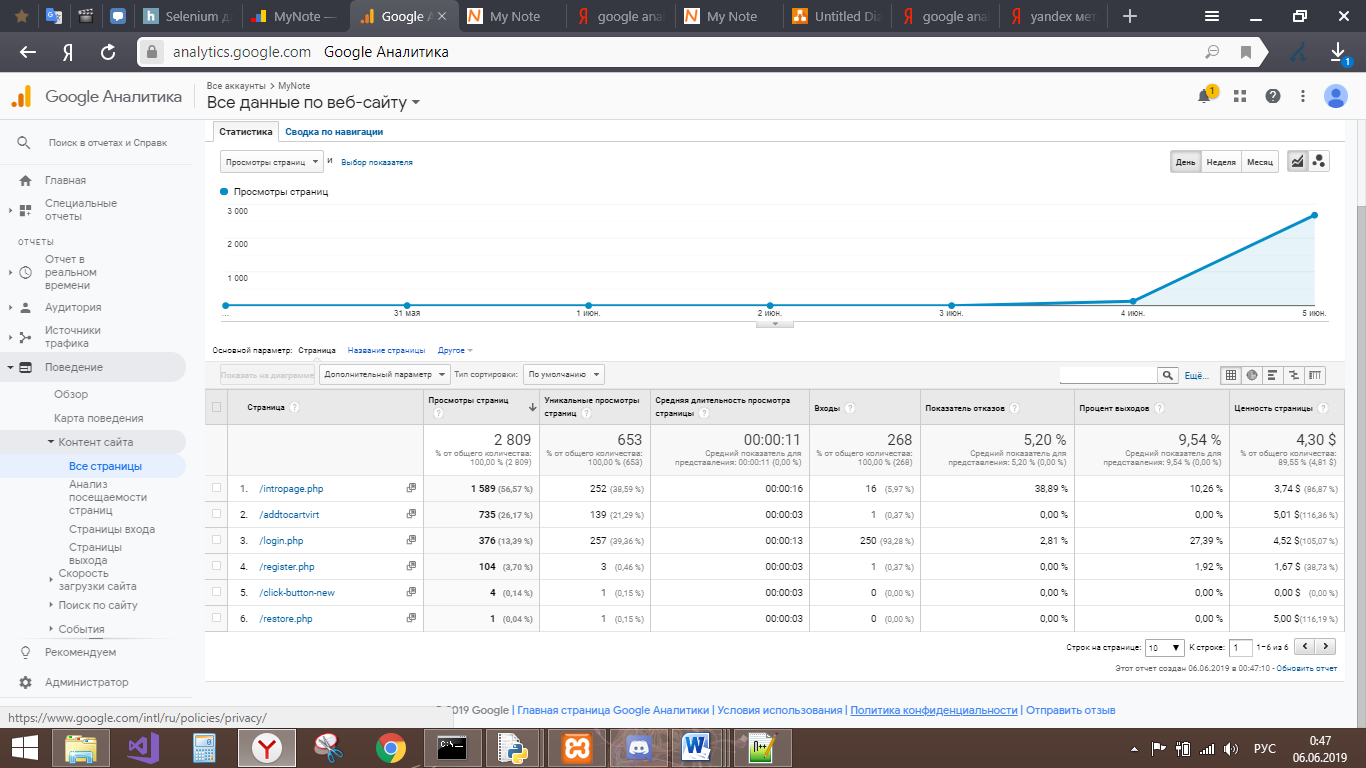
Периодичность и давность:



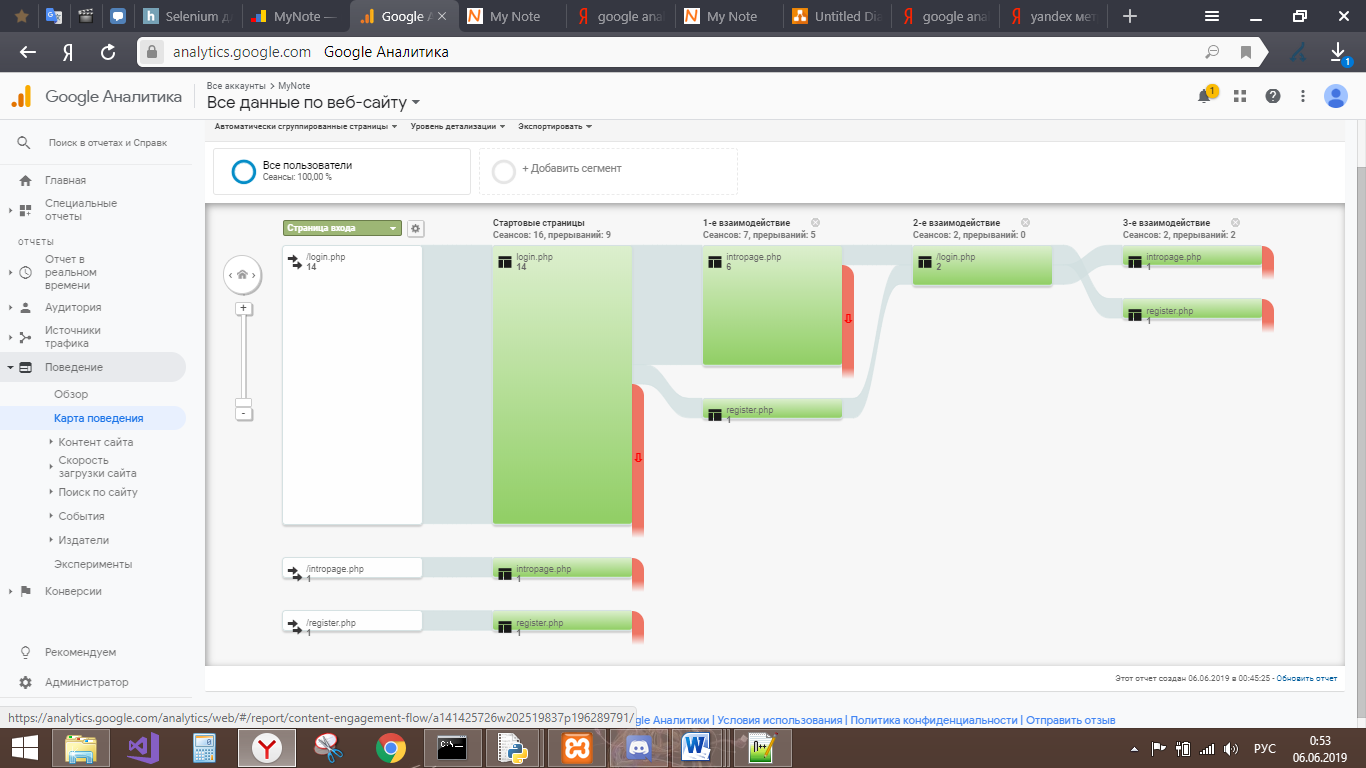
Используемые браузеры:



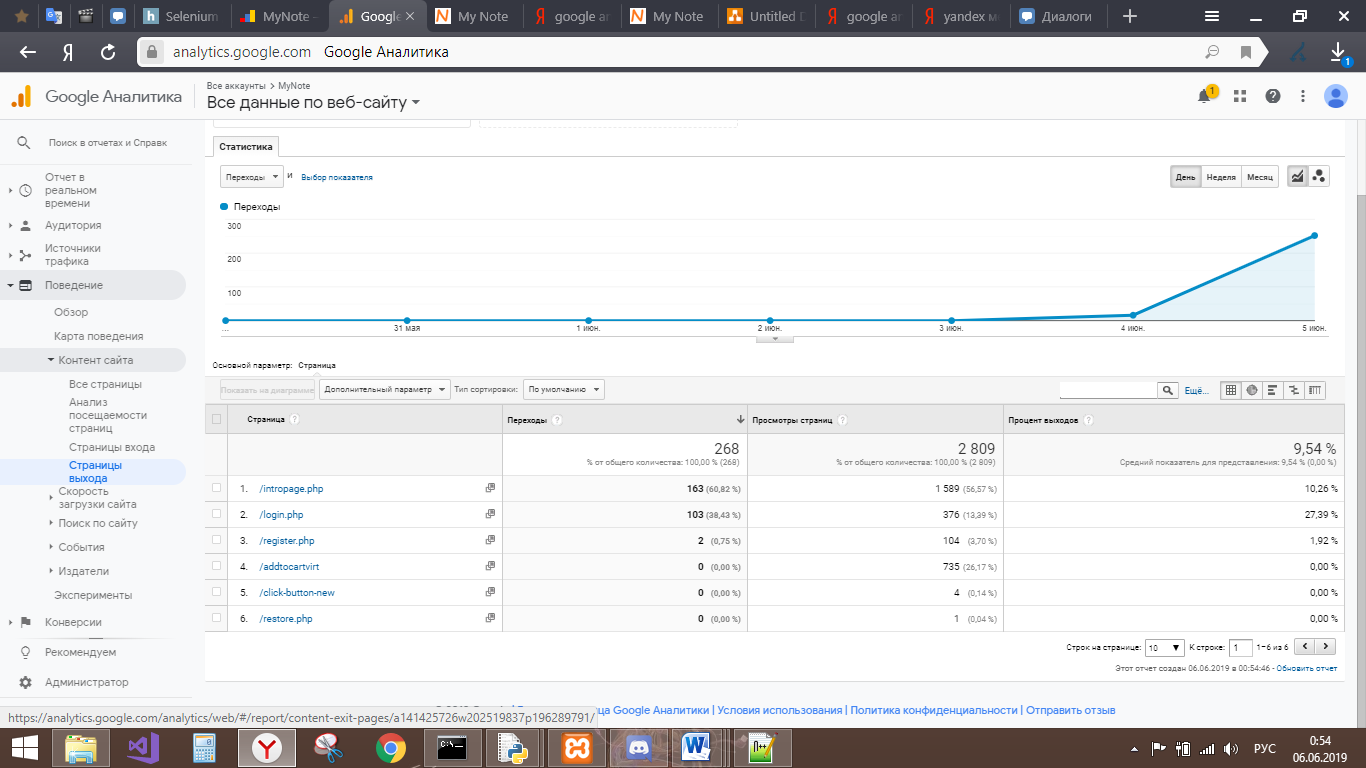
Используемый инструмент аналитики позволяет просматривать распределение трафика по страницам сайта, отсюда можно сделать выводы о самых популярных страницах. Как можно видеть ниже основная страница сайта и виртуальная страница создания новой заметки являются наиболее популярными, так как весь смысловой функционал сайта находится именно здесь.



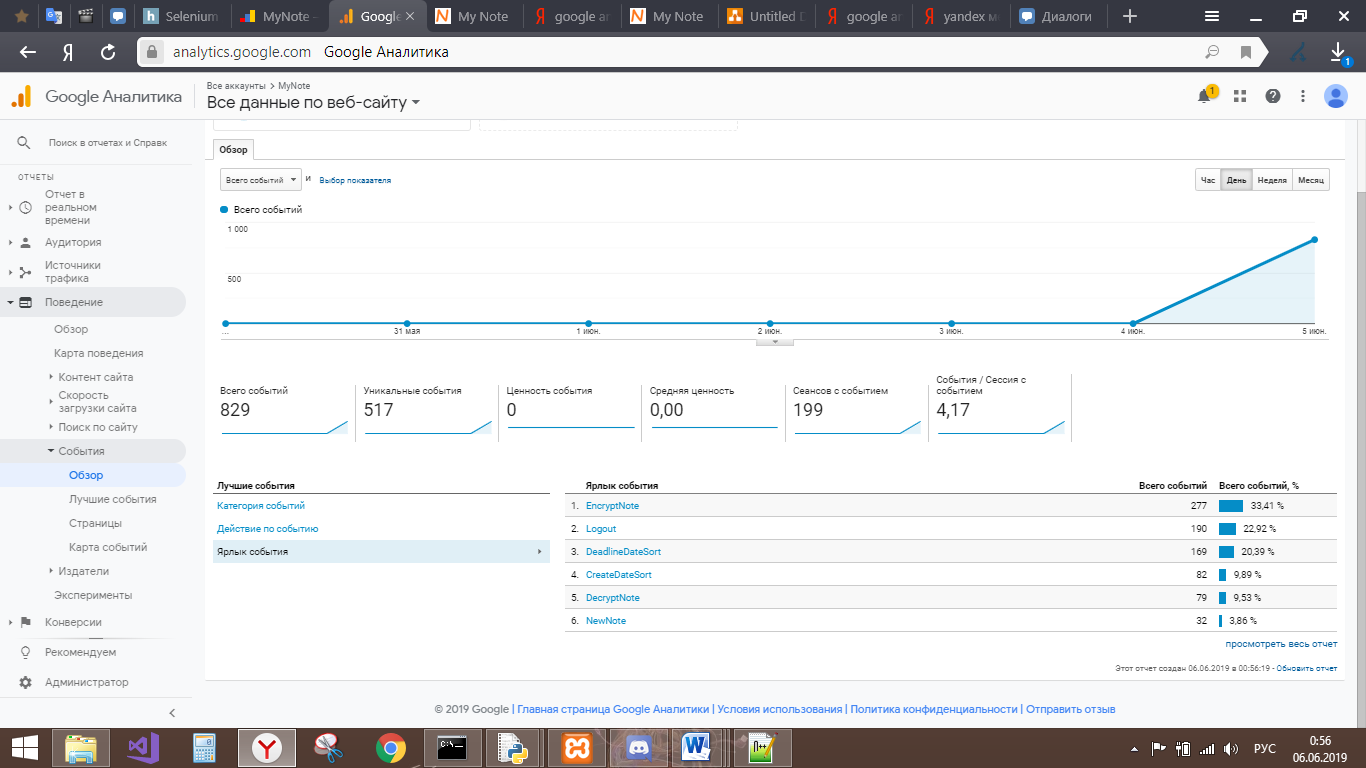
Так же имеется возможность отследить путь пользователя на сайте и понять, каким образом он переходит на те или иные страницы.



Страницы входа\выхода:



Распределение по событиям, установленным на кнопки:

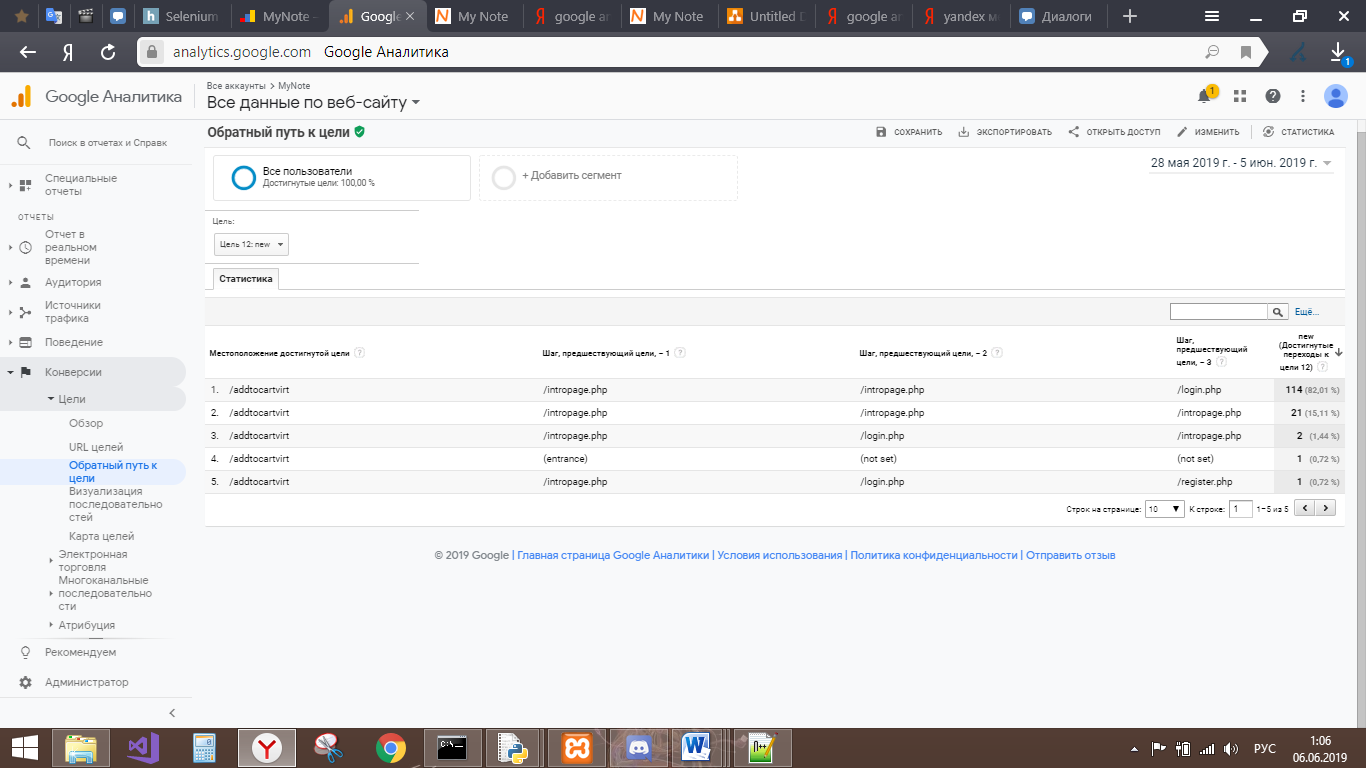


Отсюда можно сделать вывод, что пользователи предпочитают сортировку по сроку выполнения.

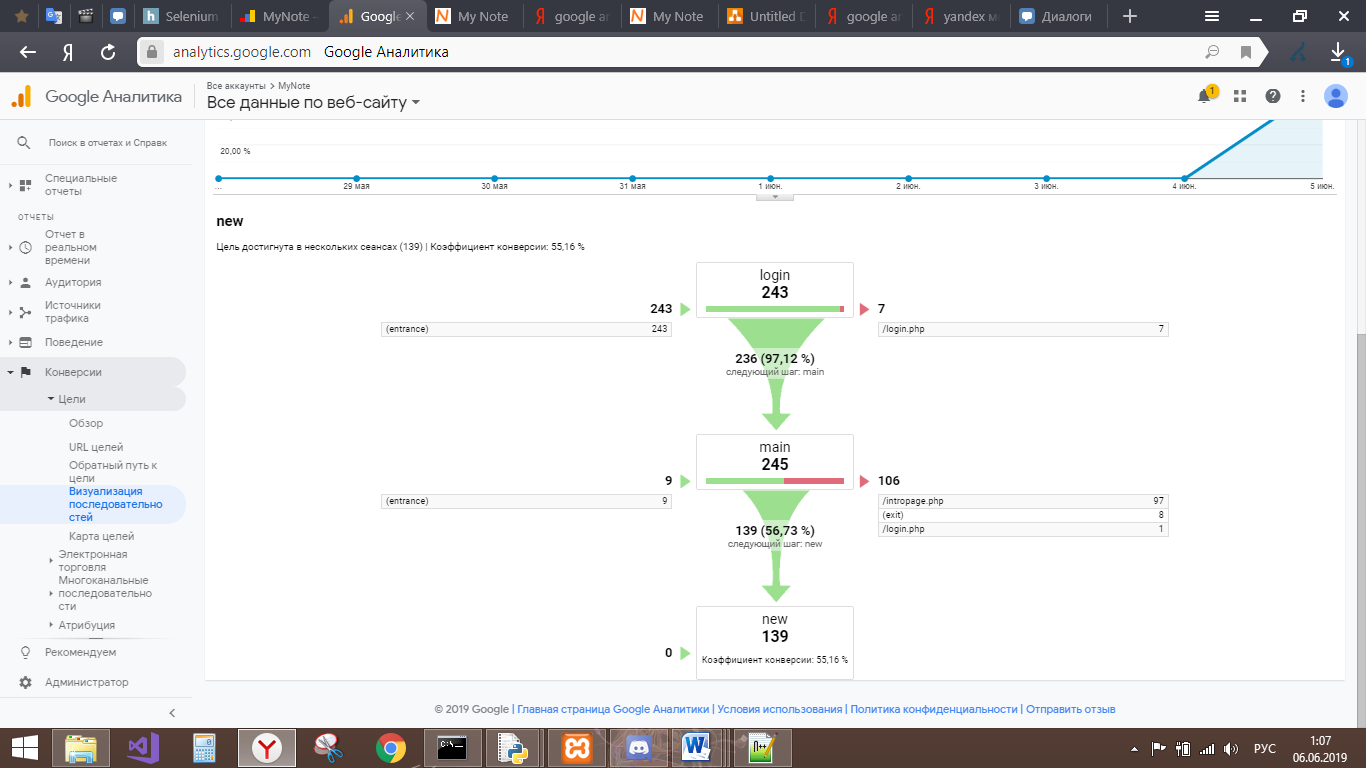
Для рассмотрения воронки была создана цель: гарантированное создание новой заметки при входе на сайт. Чтобы ее достичь пользователь должен выполнить следующие действия:

* 1. пройти авторизацию: login.php
  2. перейти на основную страницу: intropage.php
  3. создать новую заметку – перейти на виртуальную страницу: /addcartvirt

Обратный путь к поставленной цели:



Визуализация воронки:



Рассмотрим отчеты Яндекс Метрики.

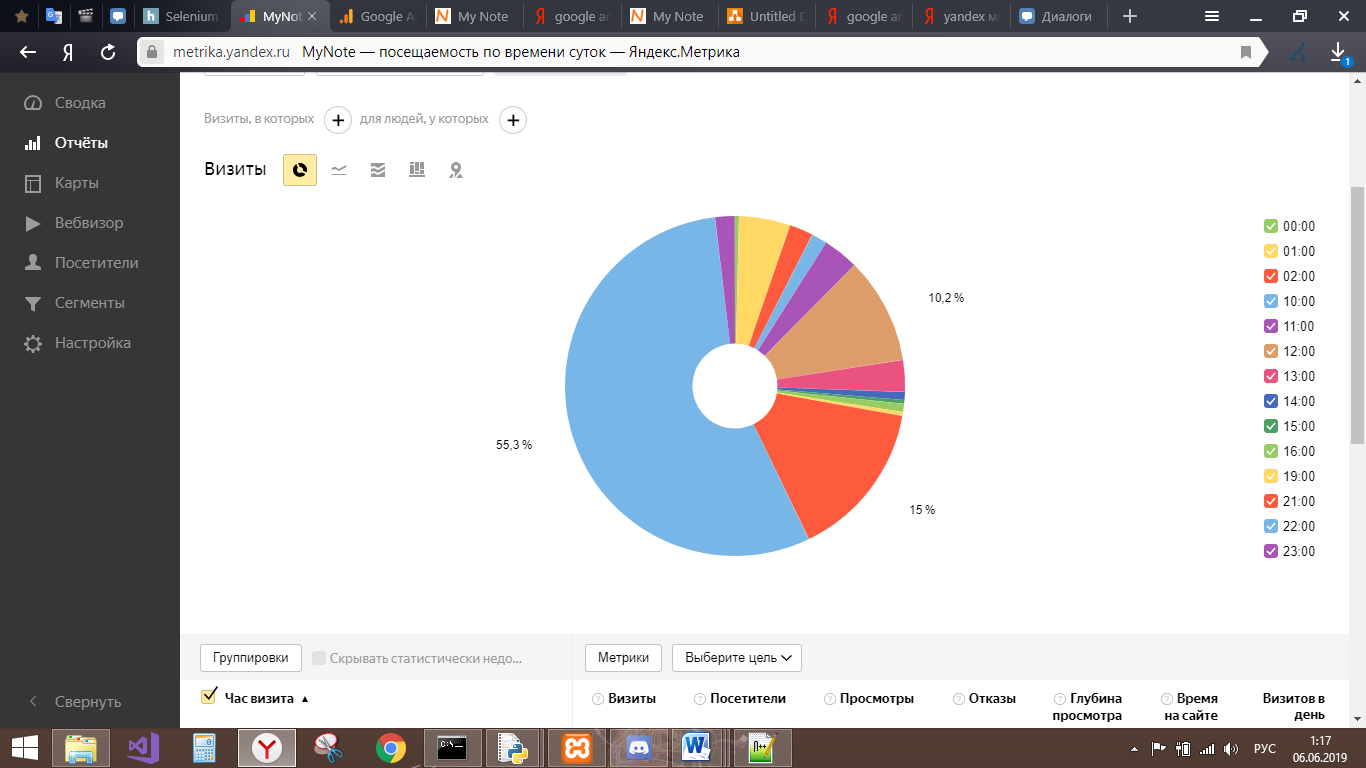
Общая информация о посещаемости:



Глубина просмотра:



Время посещения сайта:



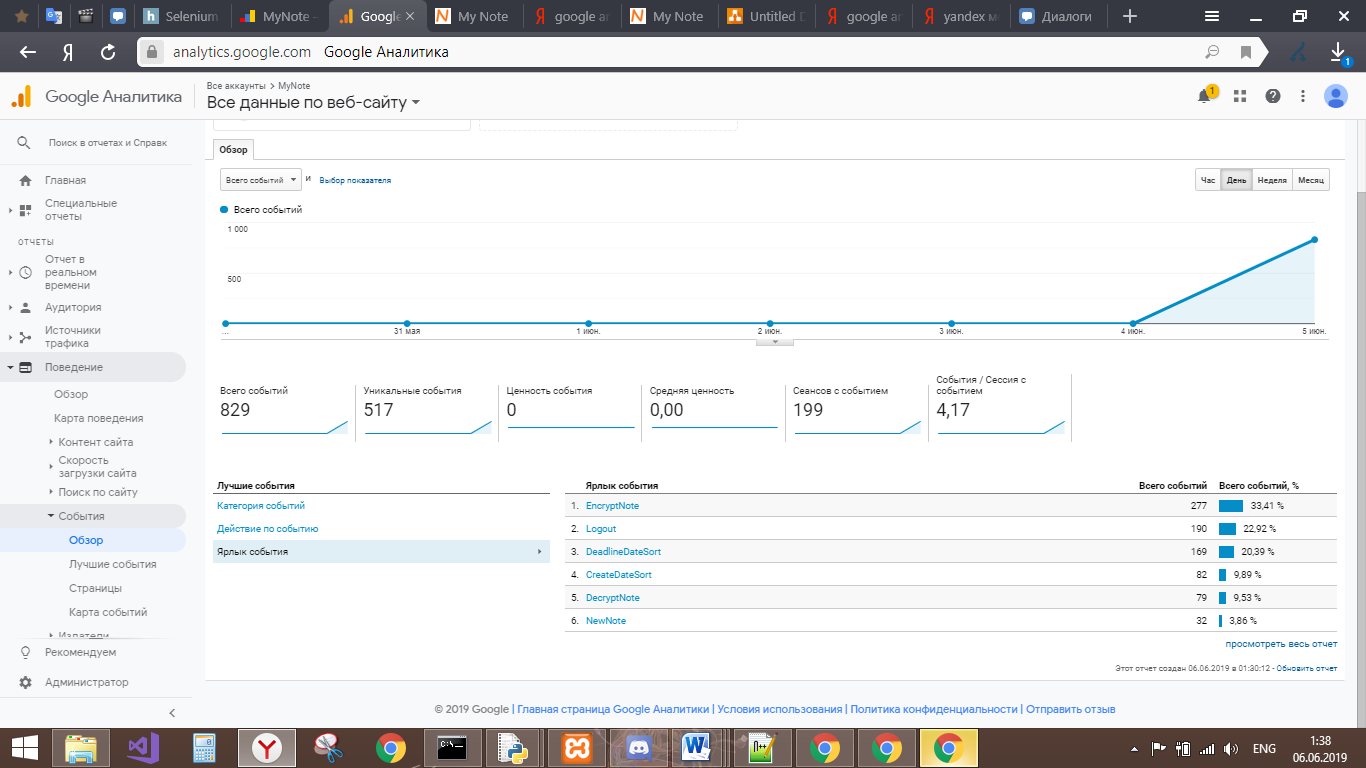
При изучении отчетов, мы пришли к выводу, что Яндекс Метрики предоставляют схожий, по сравнению с Google Analytics, функционал инструмента аналитики. Более легким в использовании оказался инструмент Яндекс Метрика: все возможные отчеты вынесены в отдельный пункт в меню, что значительно облегчает поиск необходимой информации. Что касается Google Analytics, данный инструмент позволяет отслеживать большее количество действий на сайте.

А/B тестирование – маркетинговое исследование, в рамках которого контрольной группе предоставляется исходный вариант сайта, а тестовой группе – изменённый.

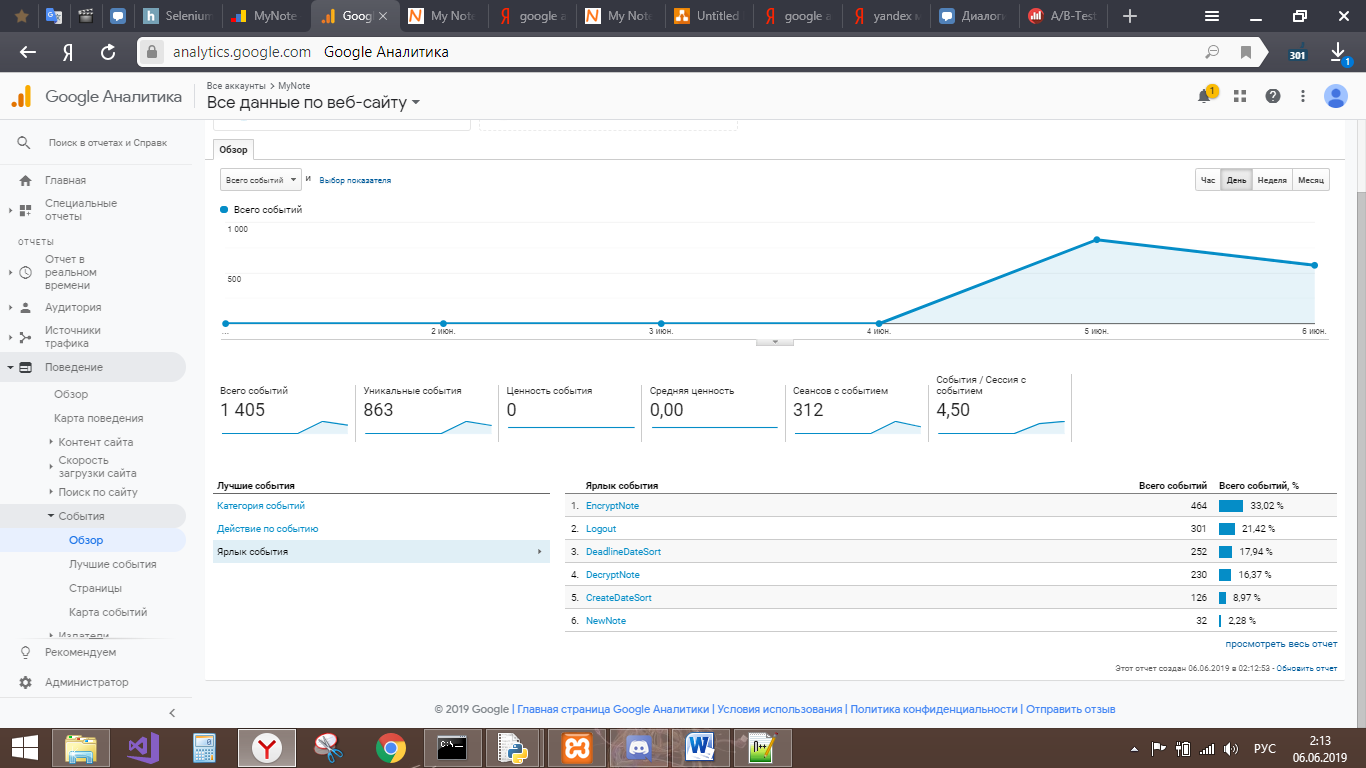
В течение определённого времени собирается интересующая исследователей статистика (число взаимодействия с изменённым функционалом). В ходе сбора установленных в Google Analytics событий было обнаружено, что боты реже всего используют функционал расшифровки заметки при ее создании. В данный функционал было внесено изменение (цвет кнопки расшифровки был изменён с серого на красный) и были произведены дополнительные замеры числа срабатывания события.

В боте генерации трафика были изменены вероятности выбора определенного действия. То есть исходя из того, что кнопка стала ярче, она будет привлекать пользователей.

Количество событий до изменения:



Количество событий после:



По результатам видно, что количество событий увеличилось. Данный факт можно объяснить тем, что, во-первых, искусственно была увеличена вероятность нажатия кнопки, во-вторых, при нажатии кнопки не обрабатывается ситуация выполнения данного действия повторно. В результате тестирования были собраны следующие статистики: в контрольной группе число нажатий кнопки «Decrypt» составило 79, а в тестовой 230-79=151 раз. Причем в контрольной группе действие совершили 120 ботов, а в тестовой 115.

Для проведения расчетов зафиксируем параметры следующим образом: количество пользователей – 200 (, количество нажатий в контрольной группе – (и количество нажатий в тестовой группе - (

Показатель конверсии:

Получили прирост conversion rate в

Оценим достоверность эксперимента. Для этого выдвигается две гипотезы и производится проверка достоверности альтернативной гипотезы с желаемым уровнем значимости:

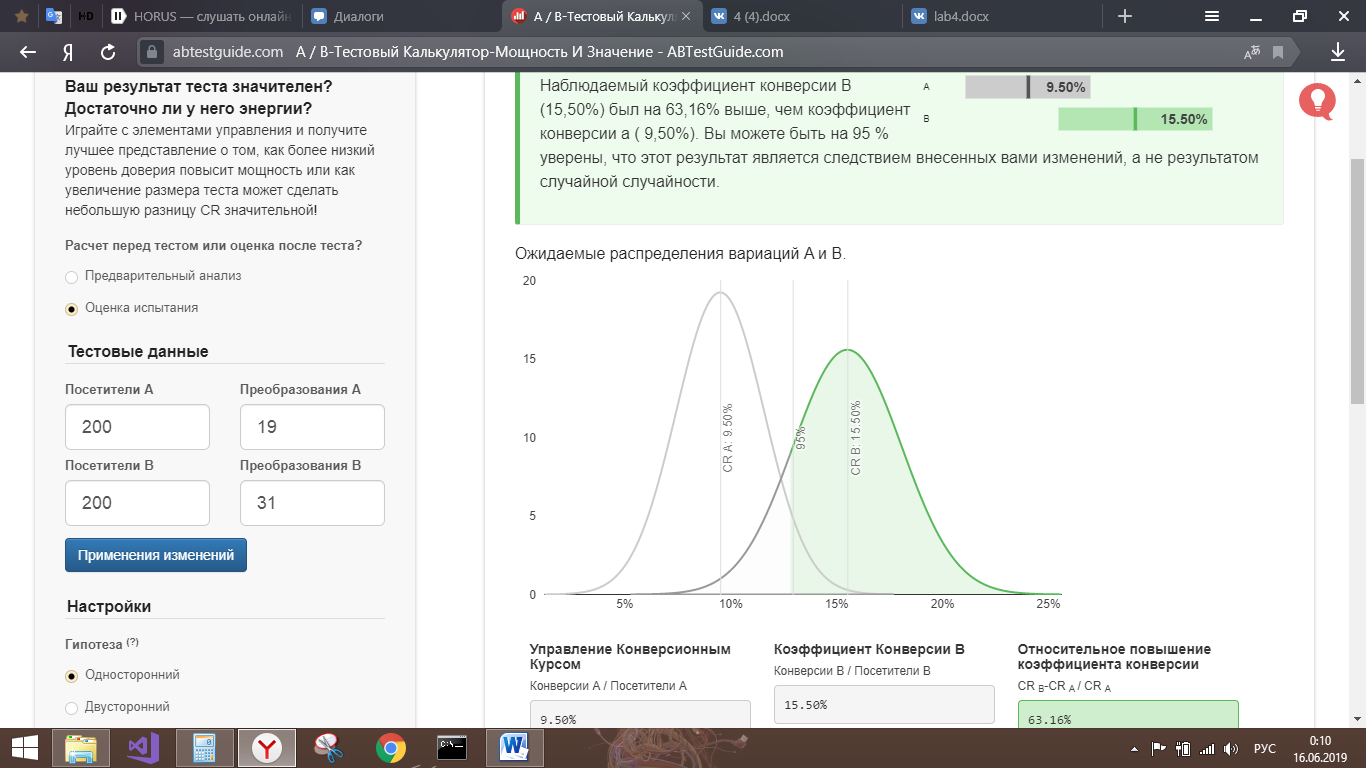
* Нулевая гипотеза: . То есть нулевая гипотеза предполагает, что изменение не привело к изменению в conversion rate.
* Альтернативная гипотеза: Альтернативная гипотеза соответствует увеличению conversion rate в результате изменений.

Для проверки гипотезы сначала рассчитываем стандартные ошибки для conversion rates по формуле:

Стандартная ошибка изменений:

Z-оценки являются стандартными отклонениями.

Ожидаемое распределение показателей:

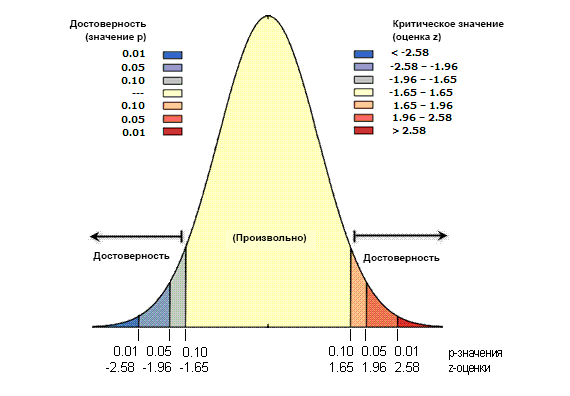


Для того чтобы считать результат достоверным, необходимо оценить площадь пересечения распределений.

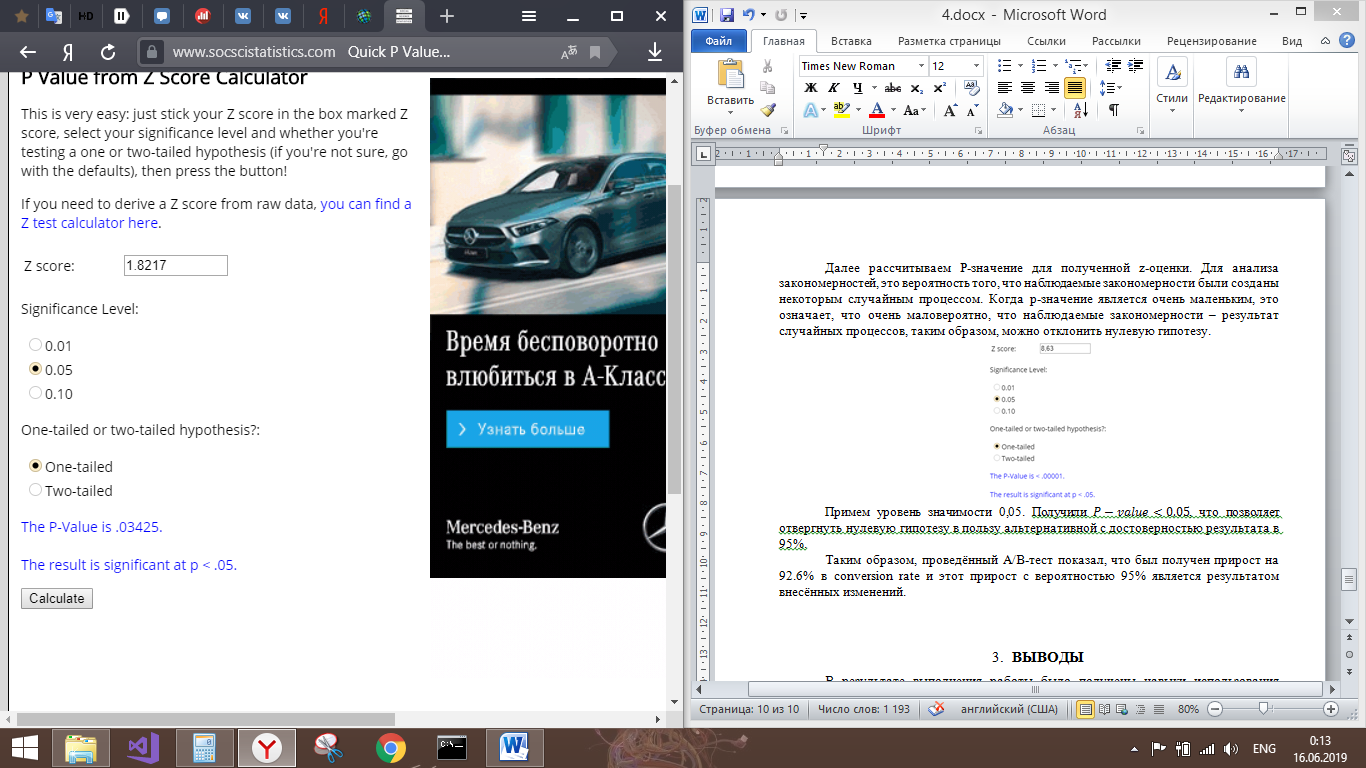
Чем меньше пересечение, тем с большей уверенностью мы можем сказать, что эффект действительно значим. Эта “уверенность” в статистике называется значимостью результата.

Для принятия положительного решения об эффективности изменений уровень значимости обычно выбирают равным 90%, 95% или 99%. Пересечение распределений при этом равно соответственно 10%, 5% или 1%. При невысоком уровне значимости существует опасность сделать ошибочные выводы об эффекте, полученном в результате изменения.

Соотношение z-меры и P-значения:



Используя онлайн калькулятор рассчитываем P-значение для полученной z-оценки. Для анализа закономерностей, это вероятность того, что наблюдаемые закономерности были созданы некоторым случайным процессом. Когда p-значение является очень маленьким, это означает, что очень маловероятно, что наблюдаемые закономерности – результат случайных процессов, таким образом, можно отклонить нулевую гипотезу.



Примем уровень значимости 0,05, что связано с понятием ошибки первого рода. Уровень значимости 0,05 означает, что существует 5% вероятность ложного срабатывания, то есть ошибки первого рода. Таким образом, если бы полученное , то мы бы приняли нулевую гипотезу.

Также выберем one-tailed гипотезу, что означает, что окончательное утверждение не может быть сделано, если коэффициент конверсии ниже, чем , то есть будет рассматриваться один «конец» распределения.

Получили , что позволяет отвергнуть нулевую гипотезу в пользу альтернативной с достоверностью результата в 95%.

Таким образом, проведённый A/B-тест показал, что был получен прирост на 63.16% в conversion rate и этот прирост с вероятностью 95% является результатом внесённых изменений.

Метрики для отслеживания

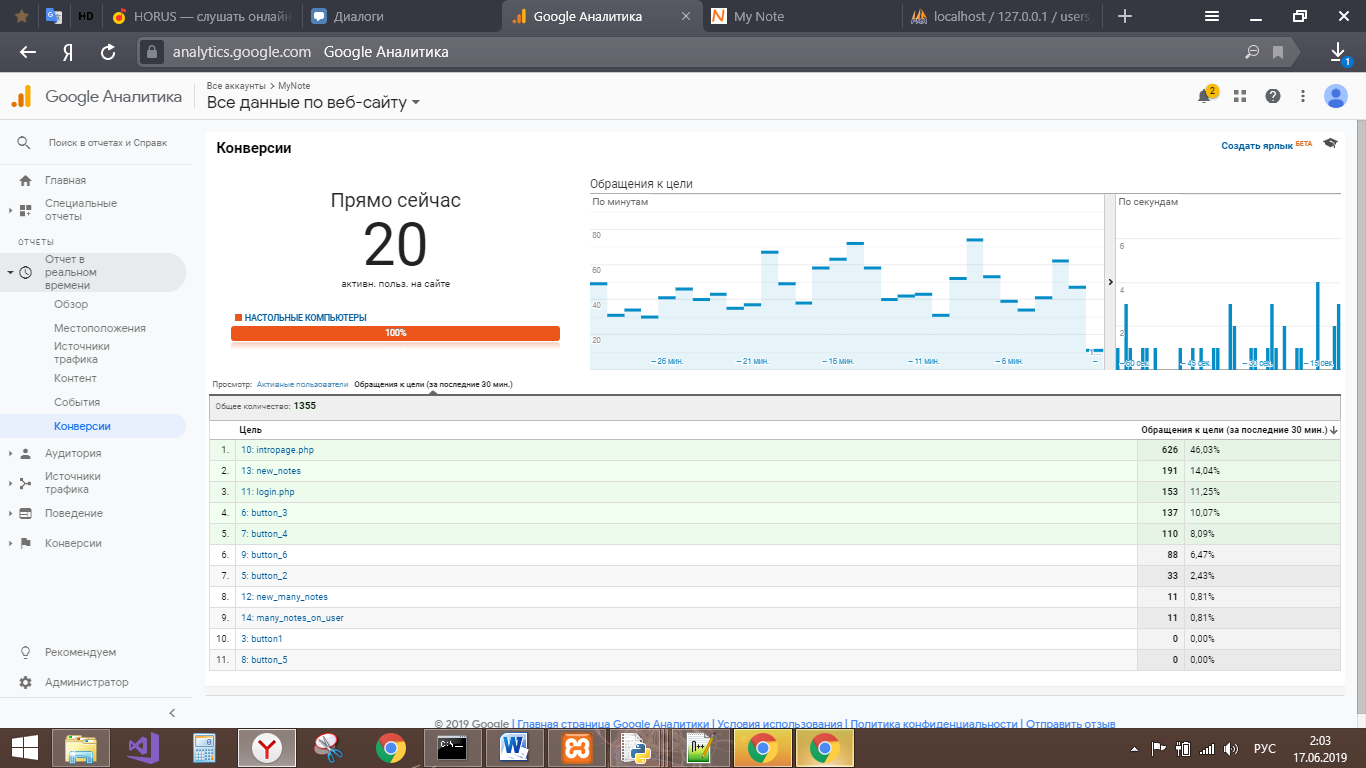
Иногда разработчикам хочется отслеживать рост своего продукта среди аудитории и сравнивать свой продукт с аналогичными. Для этого придумываются различные метрики, которые являются уникальными для данного продукта и отображают его текущую популярность. В качестве подобных метрик для онлайн-сервиса заметок было решено использовать следующие критерии, которые обновляются каждый день:

1. Количество зарегистрированных пользователей за день/неделю/месяц.

Данную информацию можно получить, посмотрев отчет в Google Analytics – число выполненных целей «new\_user» за интересующий период.

1. Количество новых заметок в день/неделю/месяц.

Так же можно узнать из отчетов Google Analytics – число выполненных целей «new\_notes» за необходимый период. Так, например, за пол часа работы 100 ботов создали 191 заметку.



1. Среднее количество новых заметок на пользователя за день/неделю/месяц.

Для высчитывания данной метрики необходимо воспользоваться предыдущими: количество новых заметок в день (неделю,месяц) / количество пользователей за период. В нашем случае за пол часа каждый бот создает в среднем 191/100 ≈ 2 заметки.

# выводы

В результате выполнения работы было получены навыки использования инструментов сбора аналитики веб-сайта, собирающих важную информацию о трафике пользователей.

На практике было проведено исследование функционала инструментов Яндекс Метрика и Google Analytics путём создания искусственного трафика и исследование результатов обработки информации об этом трафике. В результате изучения пришли к выводу, что инструмент аналитики от Яндекс представляет схожий функционал с Google, но является более удобным для людей незнакомых с системой аналитики.