

## Домашнее задание №3: Аналитика аудиторных показателей поиска

**дедлайн** - 15 ноября, 23:30

**Задание (наконец) проверено!** Критерии оценки можно посмотреть [здесь](#). Вы можете попробовать что-то исправить, прислав правки не позже 27 декабря (в этом случае вы даже сможете получить "хорошо" или "отлично"). После Нового Года тоже можно дослать исправления, но уже только на "зачёт".

Задача -- описать и нарисовать картинки о том, как менялся поисковый рынок в основных десктопных браузерах с 2014 года по данным мейла (<https://top.mail.ru/overall>), а именно:

- как менялась десктопная доля Яндекса?
- как менялась поисковая траффикогенерация в разбивке по браузерам?
- как менялась доля внутри разных браузеров?

Эту задачу будем решать поэтапно:

1. (4 балла) собрать датасет про траффик (переходы с поисковиков) по данным мейла с разбивкой по браузерам
  - <https://top.mail.ru/overall>
  - переходы из Яндекса с разбивкой по браузерам:  
<https://top.mail.ru/browsers?id=1839501&period=0&date=&aggregation=sum&side=s=mob,chrome,firefox,opera,msie,chrome/chrome-yab&ytype=hits&gtype=line>
  - берем хиты, а не пользователей, чтобы можно было суммировать по срезам
  - за 2014-01-01 - 2017-10-01
  - берем основные десктопные браузеры :  
'chrome','firefox','opera','msie','chrome/chrome-yab'
  - проверить, что абсолюты сходятся с теми, что видны в веб интерфейсе
  - собрать в csv с колонками
    - date

- se (поисковая система), принимающая значения yandex, search (все поисковики)
- browser
  - отметить, что chrome/chrome-yab - это подмножество chrome.
  - в browser=chrome положить все chrome, кроме chrome/chrome-yab
  - в browser=yabrowser положить chrome/chrome-yab
  - firefox,opera,msie оставляем без изменений
- hits
  - прислать в контекст (ссылку на контекст добавим)

## 2. подготовить данные (8 баллов)

- 2.1) пофиксить аномалии (4 балла)
  - аномалии фиксируются, чтобы дальше получать правдоподобные stl-разложение в срезах (se, browser)
  - то есть, надо пофиксить явные всплески, дырки в данных, ступеньки для срезов (se, browser)
  - какие-то аномалии могут быть во всех срезах, какие-то - нет
  - падение трафика в праздники и выходные - это наблюдаемое сезонное явление, его не надо выпиливать как аномалию
  - для каждого фикса нарисовать ряд до и после фикса
- 2.2) сделать stl-разложение рядов (4 балла)
  - надо построить stl-разложения всех (очищенных) рядов 'трафик яндекса, поисковый трафик' для каждого браузера
  - строим сначала недельное разложение (freq=7), по полученному тренду строим годовое разложение (freq=365)
  - можно воспользоваться R или прилагаемым питоновским пакетом, там же есть пример использования

- архив с оберткой над stl приложен к странице (для его работы должны быть установлены r и гру2, пример использования есть в репозитории): <file:forecast-master.zip>
  - на выходе задачи должен получиться
    - ноутбук / r-сгенеренная страничка
      - с фиксами аномалий, где для каждой найденной аномалии нарисован на одном графике ряд до и после фикса
      - разложения рядов можно делать как удобно
    - датасет вида date, se, browser, hits, hits\_fixed, hits\_week\_trend, hits\_week\_seasonal, hits\_year\_trend, hits\_year\_seasonal
3. (8 баллов) используя полученный датасет, рассказать минимумом простых картинок (чтобы было понятно гуманитарному менеджеру)
- как в общих чертах менялась доля Яндекса в исследуемых браузерах
  - как менялась доля браузеров, если учитывать только поисковую траффикогенерацию
  - как менялась доля поисков внутри браузеров
  - на выходе - ноутбук, в котором по датасету из п.4 строятся картинки, отвечающие на эти вопросы. Картинки надо сопроводить поясняющей в двух словах фразой о том, что видно на этой картинке
  -