

МОНИТОРИНГ СРЕДНИХ ЦЕН НА ТИПОВЫЕ ПРОДУКТЫ В РАМКАХ КОНТРАКТОВ ГОСУДАРСТВЕННОГО ЗАКАЗА

Алина Пишняк, Наталья Халина, Илья Черкасов

Краткое описание проекта

Проект представляет собой онлайн-мониторинг средних цен на 13 типовых продуктов из состава потребительской корзины Российской Федерации на основе данных контрактов государственного заказа, представленных на портале <http://zakupki.gov.ru>. Проект позволяет получить представление о динамике цен на продукты в виде сводных графиков в период с 2016 г. по время публикации последних контрактов, а также сравнивать динамику средних цен по регионам.

Сайт проекта <http://prices.contragents.ru>

Цель проекта

Основная цель проекта – создание мониторинга закупочных цен на базовые продукты потребительской корзины россиян, как дополнительного инструмента описания динамики и сопоставления цен на российском и региональных рынках.

Предоставляемые данные могут быть использованы:

- для анализа проводимых закупок, в том числе с целью повышения их открытости;
- в качестве информационного инструмента для специалистов в области государственных закупок в случае необходимости расчета начальной цены контракта;
- журналистами и другими представителями медиа, работающими с открытыми данными;
- экспертами в области потребления и уровня жизни россиян в качестве альтернативной и/или дополнительной системы оценки имеющихся статистических данных в области цен, предоставляемых Росстатом.

Методология

На этапе прототипа для мониторинга отобраны товары, которые отражают основные категории товаров, входящих в потребительскую корзину РФ. Отбор товаров происходит по коду ОКПД2. Список товаров с кодами представлен ниже.

1. Хлеб (10.71.11.111)
2. Макароны (10.73.11.110)
3. Крупа гречневая (10.61.32.113)
4. Картофель (01.13.51.120)
5. Томаты (помидоры) (01.13.34.000)
6. Яблоки (01.24.10.000)
7. Сахар белый свекловичный (10.81.12.110)
8. Мясо кур, замороженное (10.12.20.110)
9. Рыба морская мороженая (10.20.13.122)

10. Молоко питьевое коровье пастеризованное (10.51.11.111)
11. Яйца куриные в скорлупе (01.47.21.000)
12. Масло подсолнечное рафинированное (10.41.54.000)
13. Чай черный (10.83.13.120)

Исходные данные, представленные в контрактах, не унифицированы, содержат ошибки, полученные массивы информации требуют корректировки, поэтому для расчета средней цены проводится следующая обработка массива.

На первом этапе для каждой из категорий отбираются контракты с соответствующим указанием на стандартную единицу измерения продукта: килограммы, литры, штуки и т.д. Информация о закупках в нерелевантных/некалькулируемых единицах измерения (например, условные единицы) отсеивается.

На втором этапе для расчета средней цены и исключения статистических выбросов производится корректировка с учетом медианной стоимости единицы продукции по региону, поскольку согласно данным Росстата средние потребительские цены для разных регионов могут существенно различаться.

В соответствии с экспертными оценками для каждого региона допускается отклонение медианной стоимости единицы измерения продукции на $[0,25*Me;3*Me]$, где Me – значение медианы для конкретного региона. Значения, выходящие за пределы данных величин, рассматриваются как статистические выбросы и удаляются из массива.

Технические характеристики проекта

1. Сложность сбора исходных данных.

Данные представлены в виде множества файлов, распределенных по директориям FTP-сервера (ftp://free:free@ftp.zakupki.gov.ru/fcs_regions/). Каждый каталог соответствует региону Российской Федерации. Присутствует разбивка каталогов по годам и месяцам. Каталоги каждый день в непрерывном режиме пополняются новыми файлами с новыми опубликованными контрактами (исправлениями к контрактам).

Для учета уже скачанных и обработанных файлов на нашем сервере баз данных ведется таблица всех HASHей от имен файлов, включая полный путь к ним на FTP-сервере. Если HASH от имени найденного на FTP файла присутствует в данной таблице, то это означает, что файл уже обработан - скачивать его не нужно.

2. Стек технологий.

Язык программирования – PHP.

СУБД - MySQL (общее количество контрактов - примерно 4 400 000, общее количество цен на поставленные товары - примерно 26 000 000).

Redis - средство кэширования рассчитанных результатов.

HTTP-сервер – NGINX.

Операционная система Debian 6.

Хостинг - выделенный сервер 64Гб RAM, 2x480Гб SSD (База данных) + 2x2Тб SATA (backup)

Место на сервере, а также поддомен prices.contragents.ru предоставлены владельцем сайта и домена contragents.ru - компанией ООО "ИНФОКОМ" (тел. +74956410077). Авторы проекта выражают благодарность руководству ООО "ИНФОКОМ" за предоставленные ресурсы для некоммерческого проекта "Мониторинг средних цен".

3. Визуализация

Для отрисовки графиков по рассчитанным параметрам использован продукт HIGHCHARTS на основе Javascript (www.highcharts.com). Этот пакет бесплатен для некоммерческого использования.

4. Руководство пользователя сайта <http://prices.contragents.ru>

4.1 На главной странице - <http://prices.contragents.ru> открывается три графика (один под другим) показывающие изменение средней цены на основные продукты в среднем по России.

4.2 Пользователь с помощью формы вверху страницы может изменять следующие параметры:

- анализируемый продукт (13 основных товаров);
- регион/ы (для множественного выбора используется клавиша ctrl);
- год;
- месяц.

В случае если пользователь выбирает регион/ы, оставляя пустым поле выбора продукта, в качестве результата будут представлены графики средних цен на основные продукты (гречневая крупа, молоко, картофель) для каждого региона. В случае выбора конкретного продукта данные по регионам выводятся на один график.

4.3 Нажатие на кнопку поиска отправляет данные формы на обработку, и через 1-2 секунды страница обновляется и выводит результаты, соответствующие новому запросу.

4.4 Диаграммы на движке HIGHCHARTS позволяют сохранять графики в файл в различных графических форматах. Для этого используется кнопка в правом верхнем углу диаграммы (три горизонтальные черточки)

4.5 Для каждого запроса представлена возможность выгрузки информации по контрактам в формате csv. Файл содержит информацию о выбранном продукте (код ОКПД2), стоимости за единицу продукции, единице измерения, количестве, регионе, годе и месяце заключения/публикации контракта, а также ссылку на соответствующий контракт на портале <http://zakupki.gov.ru>.