

Московский государственный университет  
имени М. В. Ломоносова  
Факультет вычислительной математики и кибернетики

## Отчет по заданию №3

Выполнили:  
Турсуналиев Чингиз  
Серебрякова София  
Сарджаев Мередкули  
Ройтман Андрей

Преподаватель:  
Гусева Юлия

Москва

2020

## Постановка задачи

На вход подается информация о:

- Закупках (поставки яблок и карандашей два раза в месяц)
- Продажах (лог транзакций, по записи на каждую проданную позицию)
- Инвентаре (месячные данные общего количества яблок и карандашей на складе)

Данные доступны в формате CSV. Внутри файла данные отсортированы по дате. Нам необходимо получить следующие данные в CSV-файлах:

1. Состояние склада на каждый день;
2. Месячные данные о количестве сворованного товара;
3. Агрегированные данные об объемах продаж и количестве сворованной продукции по штату и году

## Ход решения

Будем работать с каждой тройкой файлов, содержащих информацию для определённого магазина штата.

Сначала создадим таблицу, где для каждой даты будет храниться количество купленных яблок и карандашей в этот день.

1. Чтобы узнать состояние склада на определённый день, надо суммировать все произведенные поставки и вычесть общее количество проданных товаров к текущему дню.
2. Чтобы получить число украденных в каждый месяц товаров, нужно сначала из теоретически высчитанного на предыдущем этапе состояния склада в конце месяца вычесть фактическое состояние склада, известное после инвентаризации. То, что мы получим, будет являться суммарным количеством товара, украденного к концу данного месяца, т. е. будет включать в себя товары, которые были украдены и во все предыдущие месяцы. Значит, чтобы получить количество сворованного товара только для текущего месяца, осталось из полученной величины вычесть суммарное количество сворованного товара, украденного к концу уже предыдущего месяца.
3. Для агрегирования данных об объёмах продаж и количестве сворованной продукции по штату и году необходимо найти сумму годовых продаж и суммарное количество украденных товаров по всем магазинам каждого штата.

# Описание программы

## Первый этап

Считываем данные при помощи функции `read`. Затем группируем по дням данные о продажах ручек и яблок, используя функцию `sell_per_day`. Кладем полученные для ручек и яблок в `df1` и `df2` соответственно.

Теперь все готово, чтобы посчитать количество товара на складе за каждый день. Делаем это учитывая следующее:

1. В первый день вычитаем из поставок продажи
2. Во второй из полученной в `p1` разности продажи во второй день и тд
3. Каждого 1 и 15 числа прибавляем новые поставки
4. Каждого 30 числа заменяем данные в нашей таблице данными из таблицы о состоянии склада на конец месяца (таблица дана нам в условии)

Проверяем наш результат, сравнивая его с данными из таблицы с ответами при помощи функции `check_daily_on_stock`

## Второй этап

Скачиваем таблицы с помощью библиотеки `requests`. Приводим данные в удобный формат, присваивая индексам `dataFrame` дату, для удобства манипулирования ими.

Создаем нужный временной интервал. Подсчитываем сворованный товар по формуле:  $(\text{сумма поставок товара за месяц}) - (\text{сумма продаж за месяц})$ . Берем разницу с реальным количеством товара на складе + товар, оставшийся на складе за прошлый месяц.

Сохраняем таблицы сворованных товаров в `csv`.

## Третий этап

Сперва определяем строки - названия столбцов будущей таблицы. Далее идёт функция вычисления продаж за год, написанная аналогично функции продаж за месяц из первого пункта. То есть из строк в логах о покупках выделяем информацию, продана ручка или яблоко. Затем считаем каждые из них по отдельности, распределяя в `Series`.

Создаётся словарь (чтобы была возможность по строке названия конкретного магазина получить доступ к соответствующей таблице статистики).

Начинается цикл, который проходит одновременно по всем заявленным именам магазинов (`markets`),

по таблицам ежемесячных краж из магазинов и по таблицам с логами продаж для каждого магазина.

Для каждого имени магазина создаётся DataFrame с заготовками лет, штата (он один — вывод из ручного просмотра файлов) и пока столбцами ( нулевыми остальными ).

Из функции `sell_per_year` получаем две Series, которые поэлементно вносим в таблицу. Суммируем по каждому году данные о кражах, которые были посчитаны во втором этапе.

## Итоги

### Результат выполнения Второго Этапа

	date	apple	pen
0	2006-01	0	1
1	2006-02	2	1
2	2006-03	4	3
3	2006-04	1	1
4	2006-05	2	2
...	...	...	...
115	2015-08	2	3
116	2015-09	5	4
117	2015-10	2	2
118	2015-11	1	0
119	2015-12	2	2

120 rows x 3 columns

## Результат выполнения Третьего Этапа

```
Out[229]: {'s1':      year state  apple_sold  apple_stolen  pen_sold  pen_stolen
          0  2006     MS      33930         31      3683         26
          1  2007     MS      33608         25      3729         13
          2  2008     MS      34071         19      3658         23
          3  2009     MS      33671         33      3657         33
          4  2010     MS      33806         31      3760         22
          5  2011     MS      33918         30      3676         27
          6  2012     MS      34046         19      3670         29
          7  2013     MS      33729         19      3521         25
          8  2014     MS      33871         24      3549         27
          9  2015     MS      34099         17      3629         24}
```

## Участники

Серебрякова София – Первый этап

Турсуналиев Чингиз – Второй этап

Ройтман Андрей – Третий этап

Сарджаев Мередкули - README