

Работа выполнена студентами 312 группы:

- Арбузов П.А.
- Кюнченкова Д.Д
- Семенов А.В.

## Постановка задачи

Необходимо по некоторым известным для каждого магазина данным посчитать новые данные для этих магазинов или для всей торговой сети. Данные находятся в Google Storage и даны в формате CSV. Результат так же должен быть в формате CSV.

## Известные данные

1. Информация о закупках  
Поставки яблок и карандашей два раза в месяц.
2. Информация о продажах  
Лог транзакций, по записи на каждую проданную позицию.
3. Информация об инвентаре  
Месячные данные общего количества яблок и карандашей на складе.

## Необходимый результат

1. Состояние склада на каждый день

date	apple	pen
2006-01-01	14105	770
...	...	...

2. Месячные данные о количестве сворованного товара

date	apple	pen
2006-01-31	4	4
2006-02-28	5	0
...	...	...

### 3. Агрегированные данные об объёмах продаж и количестве сворованной продукции по штату и году

year	state	apple_sold	apple_stolen	pen_sold	pen_stolen
2006	AK	4	2	2	1
2007	AL	5	0	6	2
...	...	...	...	...	...

## Решение

### Загрузка данных

Для быстрой загрузки файлов с Google Storage воспользуемся командой `gsutil` из пакета консольных команд `gcloud`. Для корректной работы программы все файлы с данными должны находиться в той же директории, где и сама программа.

CSV — текстовый формат, предназначенный для представления табличных данных. Каждая строка файла такого формата — это одна строка таблицы. Разделителем значений колонок является символ запятой. Для чтения данных из файлов в формате CSV будем использовать модуль `csv`. Для каждого магазина получим три списка с информацией о закупках, продажах и инвентаре. Для этого воспользуемся функцией `csv.reader(file)`.

### Создание таблицы №1

Информацию о том, когда и какой именно товар был продан будем получать из столбца с транзакциями таблицы продаж: 'ар' — яблоки, 'ре' — ручки. Количество яблок и ручек на каждый день, кроме последнего дня месяца будем считать по следующей формуле:

$$\text{количество\_сегодня} = \text{количество\_вчера} - \text{продажи} + \text{закупки}.$$

Причём так как закупки производятся два раза в месяц, то и учитываться они будут 2 раза. Для последнего дня месяца количество яблок и ручек будем брать из известных данных об инвентаре.

### Создание таблицы №2

Для того, чтобы посчитать количество сворованного товара по месяцам необходимо разницу между известным количеством товара на складе на конец месяца и количеством товара, которое должно было оказаться на складе после всех закупок и продаж.

Точное количество товара на складе мы будем брать из данных об инвентаре, а ожидаемый остаток товара будем считать аналогично первой таблице: каждый день вычитая продажи и добавляя закупки, если они были.

## Создание таблицы №3

Чтобы получить агрегированные данные по годам и штатам об объёмах продаж и сворованной продукции сначала посчитаем эту информацию по годам для каждого магазина отдельно.

Далее создадим общую таблицу, в которую добавим таблицы по каждому магазину. Отсортируем её с помощью функции `sort()` сначала по штату, потом по году. Для получения финальной таблицы посчитаем суммарные объёмы продаж и сворованного товара для каждого года и штата.

## Сохранение результата

Все полученные таблицы с помощью `csv.writer()` сохраним в исходной директории программы. Предварительно, полученные результаты сверим с ответами.

## Используемые средства

Модули:

- `csv`
- `copy`

Функции:

- `csv.reader(file)` — возвращает объект для чтения, который будет итерироваться по строкам
- `csv.writer(file, delimiter)` — возвращает объект для записи данных по строкам, используя разделитель
- `writerow(row)` — записывает в файл следующую строку
- `copy.deepcopy()` — создаёт полную копию
- `sort(key)` — сортирует список, используя параметр `key`, как функцию, возвращающую элемент для сравнения