

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ М.В.ЛОМОНОСОВА

ФАКУЛЬТЕТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И
КИБЕРНЕТИКИ

ОТЧЕТ ПО ЗАДАНИЮ 3 "APPLEPEN"

Работу выполнили:
Гладков Денис
Кислюк Лев
Назаренко Тарас

Москва, март 2020 г.

Содержание

1. Постановка задачи.	1
2. Метод решения	2
2.1. Состояние склада на каждый день	2
2.2. Число украденных товаров	2
2.3. Агрегирование данных	3
3. Программная реализация на языке Python	3
4. Список участников и их вклад в проект	4

1. Постановка задачи.

Аррлереп - большая торговая сеть, которая занимается продажей всего двух продуктов: яблок и карандашей. Её магазины расположены в различных уголках Соединённых Штатов и более 10 лет обслуживает покупателей. В CSV-файлах дана информация, отсортированная по дате:

- 1) о закупках (поставки яблок и карандашей два раза в месяц),
- 2) о продажах (лог транзакций, по записи на каждую проданную позицию),
- 3) об инвентаре (месячные данные общего количества яблок и карандашей на складе).

Необходимо получить следующие данные в CSV-файлах:

- 1) состояние склада на каждый день,
- 2) месячные данные о количестве сворованного товара,
- 3) агрегированные данные об объёмах продаж и количестве сворованной продукции по штату и году.

Наборы входных и выходных данных для тестирования решения скачаны по ссылке: <https://console.cloud.google.com/storage/browser/artem-pyanykh-smc-prac-task3-seed17> В скачанной папке с исходными данными содержится 3 типа файлов для каждого магазина в штате:

- 1) **(имя штата)-(название магазина)-supply.csv** - информация по поставкам (1го и 15го числа месяца),
- 2) **(имя штата)-(название магазина)-sell.csv** - транзакции на каждый проданный продукт,
- 3) **(имя штата)-(название магазина)-inventory.csv** - месячные данные общего количества яблок и карандашей на складе.

2. Метод решения

В предложенном задании, необходимо работать с тройками файлов, каждая из которых содержит определенную информацию для каждого из магазинов штата. В самом начале, мы создаем таблицу, где для каждой даты хранится количество купленных яблок и карандашей в эту дату

2.1. Состояние склада на каждый день

В 1 и 15 числа каждого месяца в магазин происходит поставка товара. Так что исходя из этого, будем считать что со склада никто не может ничего украсть, то есть то что привезли на склад оттуда может пропасть, только если соответствующий товар продали. Количество инвентаря на складе, посчитанное в начале каждого месяца как раз и говорит о том, сколько товара пропало со склада, то есть с учетом продажи и воровства. Для того чтобы узнать состояние склада на определенный день, нам необходимо суммировать произведенные поставки и вычесть из них количество проданных товаров до этого дня включительно.

2.2. Число украденных товаров

Для получения числа украденных за каждый месяц товаров, нужно из числа товаров, которое получилось во время инвентаризации за предыдущий месяц, вычесть количество проданных за текущий месяц товаров, а так же фактическое состояние склада

2.3. Агрегирование данных

Для агрегирования данных об объемах продаж и количестве утраченной в результате воровства продукции, необходимо найти сумму годовых продаж и суммарное количество украденных товаров по всем магазинам конкретного штата.

3. Программная реализация на языке Python

Язык программирования - Python.

Реализация и запуск кода осуществлялись через Python3 с помощью команды `./python3 main.py`

Используемые библиотеки:

- pandas - для регулярных выражений
- re - для работы с табличными данными, их хранения и обработки

4. Список участников и их вклад в проект

- Гладков Денис - выполнение пунктов 1, 2
- Кислюк Лев - написание Readme
- Назаренко Тарас - выполнение пункта 3