

Отчет по заданию №1

Содержание

1. Постановка задачи	2
2. Ход решения	2
3. Используемые программные средства	3
4. Исполнение	3
5. Участники	4

Постановка задачи

Applepen — это большая торговая сеть, которая занимается продажей всего двух продуктов: яблок и карандашей. Ее магазины расположены в различных уголках Соединенных Штатов и более 10 лет обслуживают покупателей.

Недавно топ-менеджмент компании решил более активно использовать имеющиеся у них данные в принятии решений. Каждый магазин собирает информацию о:

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

Данные доступны в формате CSV. Внутри файла данные отсортированы по дате. К сожалению, данные никогда не консолидировались и не проверялись. Нам необходимо получить следующие данные в CSV-файлах.

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

Ход решения

При выполнении поставленной задачи будем работать с тройками файлов `inventory`, `sell` и `supply`, в которых содержится вся необходимая информация для каждого предприятия штата.

Для хранения и обработки данных о состоянии складов на каждый день, поставках и общих продажах, определим их в отдельные “таблицы”, с которыми будем работать в дальнейшем.

1. Чтобы получить данные о состоянии склада на каждый день, нам необходимо найти количество поставок (если таковые имеются), произведенных в этот день, и вычесть количество произведенных продаж.
2. Чтобы получить месячные данные о количестве украденного товара, нам необходимо воспользоваться данными о состоянии склада, полученными нами теоретически в предыдущем пункте задания, и вычесть данные о фактическом состоянии склада. Однако стоит заметить, что результатом данных вычислений будет суммарное количество украденного товара на конец текущего месяца. Это означает, что он будет учитывать кражи и всех предыдущих месяцев, но

поскольку для выполнения данного пункта нам необходимо получить информацию о кражах на каждый месяц, то из уже полученной величины нужно вычесть общее количество товара, украденного к концу предыдущего месяца.

3. Чтобы получить агрегированные данные об объемах продаж и количестве украденного товара по штату и году, нам необходимо найти общее количество продаж за год, а также количество украденного товара по всем предприятиям штата.

Используемые программные средства

Python — интерпретируемый язык программирования, используемый для реализации задания.

Jupyter Notebook — основанный на web программный интерфейс, используемый в работе для написания кода на Python.

В реализации применялись следующие библиотеки:

1. pandas — библиотека, предоставляющая инструменты для обработки и анализа данных (в реализации используется для работы с временными рядами);
2. datetime — модуль, предоставляющий классы для обработки времени и даты разными способами;
3. calendar — модуль, позволяющий производить различные операции с датами;
4. glob — модуль, который находит все пути, совпадающие с заданным шаблоном в соответствии с правилами, используемыми оболочкой Unix;
5. os — модуль, предоставляющий множество функций для работы с операционной системой.

Исполнение

1. На вход подаются файлы, содержащие информацию о продажах, закупках и инвентаре, -- sell, supply и inventory соответственно. Считываем данные из этих файлов с помощью функции read_csv и организуем их в отдельные списки для дальнейшей работы с ними.
2. Заводим некоторые переменные, необходимые нам для дальнейших вычислений.
3. Создаем цикл, внутри которого будем “пробегать” по данным о продажах, поставках и инвентаре для каждого предприятия. Здесь будем производить

подсчет величин, необходимых нам для выполнения каждого из трех пунктов задания, а именно:

- a. `apple_stolen` -- содержит информацию о количестве украденных яблок;
- b. `pen_stolen` -- содержит информацию о количестве украденных карандашей;
- c. `apple_sold` -- содержит информацию о количестве проданных яблок;
- d. `pen_sold` -- содержит информацию о количестве проданных карандашей;
- e. `apple_day` -- содержит информацию о количестве яблок в инвентаре предприятия на конец каждого дня;
- f. `pen_day` -- содержит информацию о количестве карандашей в инвентаре предприятия на конец каждого дня.

- 4. Производим все необходимые вычисления, описанные в нами в разделе “Ход решения”;
- 5. Для выполнения третьего пункта задания заводим словарь `by_year_dict`, где будут обрабатываться данные обо всех предприятиях штата за год;
- 6. Для вывода результатов, полученных нами при выполнении всех пунктов задания, организуем их в словари и, используя функцию `file.to_csv`, записываем в соответствующие файлы;
- 7. На выводе, учитывая все необходимые требования, получаем файлы csv формата, где содержатся результаты проведенной работы.

Участники

Азимжанова Инаара (312), Узакбай Каби (312), Югай Светлана (312).
Все этапы работы были выполнены в команде.