

Анализ данных

Пожилая саламандра

March 2020

1. Реализация решения

1.1. Функции

- Функция *daily()* вычисляет дневное состояние склада и сохраняет это в файл, а также вычисляет дневные продажи и возвращает их в качестве нового dataframe. Так как при обычном обращении по столбцу и номере строки создается копия соответствующей строки, то для изменения необходимой ячейки нужно использовать *.loc[...]*. В конце функции из соответствующих датафреймов удаляются лишние столбцы, а также столбец дат делается индексным.
- функция *stolen()* вычисляет количество украденного товара в месяц как разность между «идеальным» состоянием склада и реальным состоянием после инвентаризации.
- функция *yearly()* принимает на вход dataframe с дневными продажами и вычисляет кол-во проданных и украденных товаров в каждый из годов. Для оптимальности это делается за один проход, то есть dataframe с дневными продажами — основной, и при совпадении даты в проданном и украденном прибавляется очередное значение к переменной годовых краж.
- для удобства перебора всех файлов сделан массив с именами магазинов, а также соответствующие переменные с путями к входным/выходным данным и типом файла — склад/продажи/поставки.

1.2. Примененные средства

Библиотеки:

- *numpy*;
- *pandas*;
- *datetime*.

Для доступа к Google Storage использовался пакет *gcloud* и команда *gsutil*.

2. Вклад

- Никита Денисов - *sales.py*
- Андрей Ловягин - ноутбучный файл *.ipynb*
- Михаил Иванков - *README*