

Министерство цифрового развития, связи и массовых коммуникаций РФ  
ордена Трудового Красного Знамени федеральное государственное  
бюджетное образовательное учреждение высшего образования  
«МОСКОВСКИЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ СВЯЗИ  
И ИНФОРМАТИКИ»

Кафедра «Математическая кибернетика и информационные технологии»  
Дисциплина «Структуры и алгоритмы обработки данных»

Курсовая работа

по дисциплине «Структуры и алгоритмы обработки данных»  
на тему: «Считывать данные из CSV файла»

Выполнил:

студент группы БФИ2201

Лопеш С.С.

Проверил:

Павликов А. Е.

Москва, 2024 г.

## Оглавление

Введение.....	3
Техническое задание .....	4
Ход выполнения работы .....	5
Заключения .....	7

## Введение

В данной курсовой работе мы должны, используя разные структуры данных и алгоритмы, написать программу, которая рассчитает общую выручку магазина, найдёт товар, который был продан наибольшее количество раз и найдёт товар, который принес наибольшую выручку.

В курсовой работе будем использовать такую структуру данных, как хеш-таблица. Хеш-таблица — это структура данных, в которой все элементы хранятся в виде пары ключ-значение, где:

- **ключ** — уникальное число, которое используется для индексации значений;
- **значение** — данные, которые с этим ключом связаны.

Когда хеш-функция генерирует один индекс для нескольких ключей, возникает конфликт: неизвестно, какое значение нужно сохранить в этом индексе. Это называется **коллизией хеш-таблицы**. Для разрешения коллизий будем использовать метод цепочек.

Суть этого метода проста: если хеш-функция выделяет один индекс сразу двум элементам, то храниться они будут в одном и том же индексе, но уже с помощью двусвязного списка.

В курсовой работе будем использовать методы сортировки, для получения товара, который был продан наибольшее количество раз.

### **Техническое задание**

Написать программу, которая будет считывать данные из CSV файла, содержащего информацию о продажах товаров в магазине. Данные в файле содержатся в следующем формате: | Номер заказа | Дата заказа | Название товара | Категория товара | Количество продаж | Цена за единицу | Общая стоимость | Необходимо:

1. Рассчитать общую выручку магазина.
2. Найти товар, который был продан наибольшее количество раз.
3. Найти товар, который принес наибольшую выручку.
4. Составить отчет, содержащий информацию об общей выручке магазина, количестве проданных единиц каждого товара и доле каждого товара в общей выручке. Для решения задач необходимо использовать структуры данных, такие как массивы и хеш-таблицы, а также различные алгоритмы обработки данных, например, сортировку и поиск. Также необходимо учитывать возможные ошибки ввода-вывода и обрабатывать их в соответствии с требованиями

### Ход выполнения работы

Подключим необходимые библиотеки для работы: `pandas`, `matplotlib.pyplot`, `csv`. После подключения модулей, напишем код для работы с хеш-таблицей. В хеш-таблице мы создаем класс узла списка (`Node`), после чего создаем класс хэш-таблицы (`HashTable`). В классе хеш-таблицы пропишем методы для работы: инициализация (`__init__`), добавление элемента в таблицу (`add`), получение значения по ключу (`get`), изменение емкости таблицы (`resize`).

Создадим функцию для сортировки. Использовать мы будет `quick_sort` (“быструю сортировку”), так как она быстрее всех выполняет фильтрацию элементов от наименьшего к наибольшему. Чтобы отсортировать элементы, мы должны создать переменную, в которой будет храниться опорный элемент. С этим опорным элементом будут сравниваться остальные элементы. Для реализации быстрой сортировки создадим 3 массива, в которые будут отправляться элементы, которые меньше опорного, равны опорному, больше опорного. Далее напишем функцию для выполнения расчетов.

Объявим переменные, которые будут хранить: общую выручку и максимального количества проданного товара (`revenue`, `max_tovar`); переменную типа `string`, в которую мы запишем названия товара, проданный максимальное количество раз (`max_sales`); переменную для хранения количество строк в `csv` файле (`count_row`) и два массива: массив для ключей / название товаров и массив для выручки (`names_array`, `reven_array`).

Вызовем хеш-таблицу, записав ее в переменную (`h1`, `h2`). Будем проходить по всем строкам в файле. В `names_array` добавляем товар/ключ. Используя функцию `append`, добавим в массив строку с индексом 2. Строка с индексом 2 содержит данные о названии товара. В `revenue` добавим строку с индексом 6, которая отвечает за общую стоимость. Добавлять мы будем, преобразовав значение в тип `int`.

Используя оператор if, найдем товар, который принес наибольший доход, а также найдем название этого товара.

Если max\_tovar меньше, чем сравниваемый элемент из строки, то обновляем значения в max\_tovar и max\_sales, которое содержится в строках под индексом 2. В массив для выручки добавим значение выручки из i-ой строки в индекс, который отвечает за общую стоимость.

В h2 записываем значение выручки/ключ и товар. Добавляем в h1 товар/ключ и количество его продаж и добавляем товар/ключ и его выручку. Выводим наши значения из хеш-таблиц, используя метод get, который прописали в классе хеш-таблицы. Сортируем массив с выручками. А возвращать мы будем общую выручку, названия товара, который был продан наибольшее количество раз и товар, который принес наибольшую выручку (revenue, max\_sales, ht2.get(str(result[-1]))). Последнее мы будем выводим из хеш-таблицы методом get последний элемент, преобразованный в тип string.

Используем оператор try-except, который будет отвечать за ввод/вывод без ошибок. Объявим переменную, в которой будет наш исходный файл. Вызовем нашу функцию для расчета.

Итоговый вывод оказался таким:

```
Grundig 50 NANO GH 8100 был продан 200 раз, доля товара от выручки составляет 12.89%
Bose QuietComfort был продан 150 раз, доля товара от выручки составляет 1.34%
Apple MacBook Air 13 был продан 100 раз, доля товара от выручки составляет 7.16%
HONOR Pad 8 был продан 50 раз, доля товара от выручки составляет 1.25%
Apple iPhone 13 был продан 250 раз, доля товара от выручки составляет 6.8%
Beko WSPE7612W был продан 75 раз, доля товара от выручки составляет 5.67%
Electrolux EACS-07HAR_A/N3_21Y был продан 30 раз, доля товара от выручки составляет 1.72%
HUAWEI Pura70 был продан 300 раз, доля товара от выручки составляет 9.67%
Thunderobot 911S Core был продан 150 раз, доля товара от выручки составляет 5.32%
Candy CCRN 6180W был продан 60 раз, доля товара от выручки составляет 1.4%
Nest Learning Thermostat был продан 250 раз, доля товара от выручки составляет 1.7%
Lenovo V15 G4 AMN был продан 350 раз, доля товара от выручки составляет 10.9%
Hisense WFQP6012VM был продан 100 раз, доля товара от выручки составляет 1.61%
Apple iPhone 15 Pro был продан 550 раз, доля товара от выручки составляет 17.72%
Hi VHX-40F152MSY был продан 400 раз, доля товара от выручки составляет 14.32%
Rapid RAM-07HJ/N1_23Y был продан 25 раз, доля товара от выручки составляет 0.54%

Общая выручка магазина: 279304000 рублей
Товар, который был продан наибольшее количество раз: Apple iPhone 15 Pro
Товар, который принес наибольшую выручку: Apple iPhone 15 Pro
```

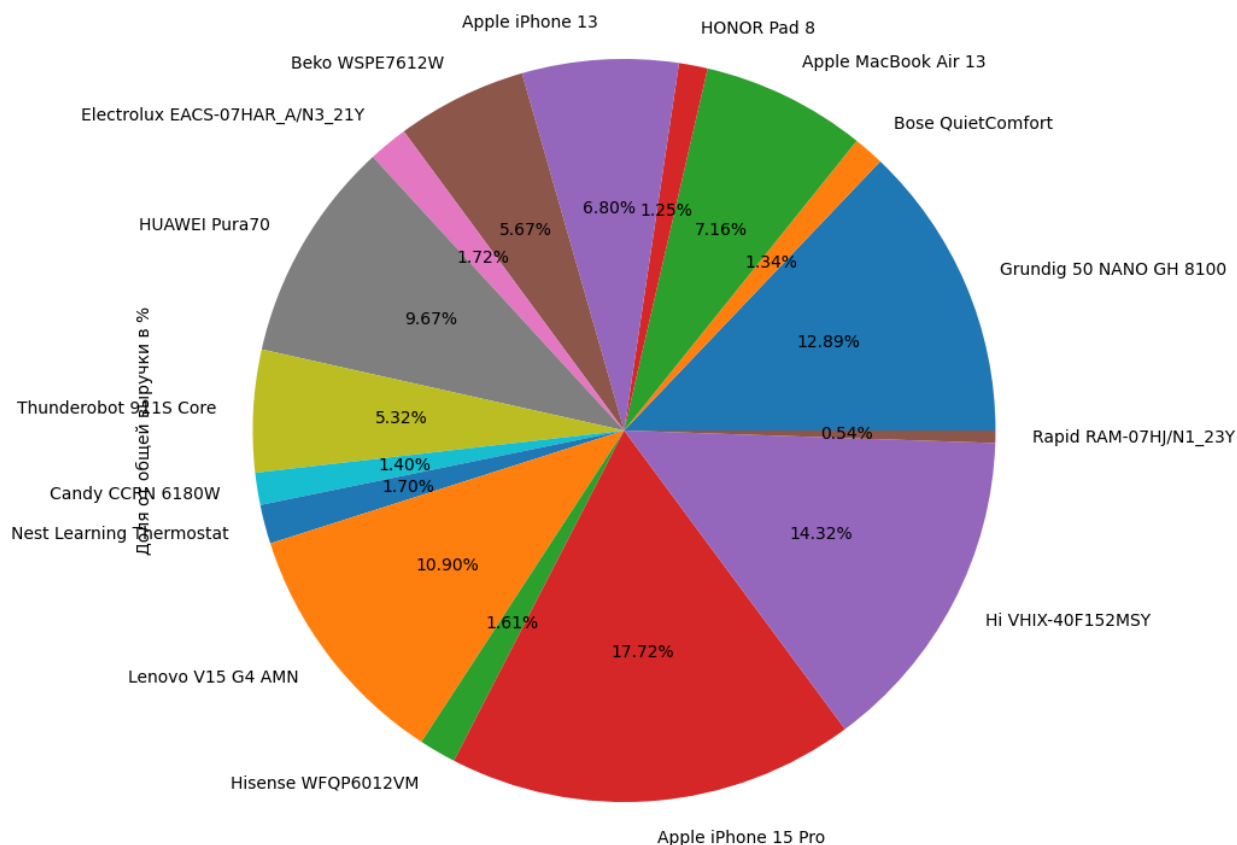


Рисунок 1 – Круговая диаграмма долей товаров от общей стоимости

### Заключения

Написал программу, которая будет считывать данные из CSV файла, содержащего информацию о продажах товаров в магазине. Рассчитал общую выручку магазина. Нашёл товар, который был продан наибольшее количество раз. Нашёл товар, который принес наибольшую выручку. Составил отчет, содержащий информацию об общей выручке магазина, количестве проданных единиц каждого товара и доле каждого товара в общей выручке.