**МИНИСТЕРСТВО ЦИФРОВОГО РАЗВИТИЯ СВЯЗИ И МАССОВЫХ КОММУНИКАЦИЙ**

**Ордена Трудового Красного Знамени**

**Федеральное государственное бюджетное образовательное учреждение высшего образования**

**«Московский технический университет связи и информатики»**

Кафедра «Математическая кибернетика и информационные технологии»

**Курсовая работа**

по дисциплине «Структуры и алгоритмы обработки данных»

Выполнил: студент группы БСТ2203

Иванов Д.

Проверил:

Симонов С.Е.

Москва, 2024

**Содержание**

[Введение 3](#_Toc167207791)

[1 Задание 4](#_Toc167207792)

[2 Выполнение работы 5](#_Toc167207793)

[Заключение 11](#_Toc167207794)

# Введение

Данная курсовая работа посвящена написанию программы на языке python, которая получает на вход csv-файл, содержащий информацию о товарах, их продажах, датах продаж, выручке. Программа должна обработать данный файл, выводя отчет по продажам и его графическое представление.

Для работы с csv-файлом в программе будет использована библиотека csv.

В данной программе будет использована структура хеш-таблица – набор пар ключ-значение.

В программе будут реализованы функции сортировки для графического отображения, поиска необходимых по заданию товаров, рехеширования на случай коллизий и форматированного вывода.

Для вывода в виде таблицы будет использована библиотека pandas.

Для графического вывода в виде гистограмм и круговых диаграмм будет использована библиотека matplotlib.

# 1 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 товар 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.

# 2 Выполнение работы

Сначала подключаются необходимые библиотеки: csv, pandas и matplotlib.

Далее объявляется класс хеш-таблица (HashTable), в котором указываются методы инициализации (\_\_init\_\_) и вставки (insert). Вставка будет использоваться при чтении файла.

После объявляется функция специальной сортировки выбором SelectionSortSpec. Под словом «специальная» подразумевается, что она создана только для работы с этой программой. На вход подаются три массива: revenues (выручка), quantity (количество продаж) и names (названия товаров). Сортировка происходит по массиву выручек, а остальные массивы меняют элементы местами в соответствии с массивом revenues. Это нужно для сохранения соответствия «товар – количество продаж – выручка».

Далее объявляется функция поиска названия товара по числу продаж SearchQuan. На вход поступают хеш-таблица (sales) и количество продаж (quantity). В функции осуществляется построчный поиск количества продаж. Если значение в столбце таблицы «Количество продаж» совпадает со значением входной переменной quantity, то функция вернет значение столбца «Название товара».

Далее идет аналогичная функция поиска названия товара по выручке от него SearchRev. Её алгоритм такой же, как и у SearchQuan, с одним отличием в том, что на вход вместо quantity поступает переменная revenue (выручка), и поиск осуществляется по ней и по значению столбца «Общая стоимость».

Следующая функция – простое рехеширование Rehash. На входе переменная ключа key и массив существующих ключей keys. Функция увеличивает значение ключа на единицу и возвращает его.

Последняя функция – функция форматированного вывода FormatPrint. У неё на входе хеш-таблица sales, массив количества продаж quantity и массив выручек товаров revenues. Сначала определяется общая выручка как сумма всех элементов массива revenues, значение записывается в переменную total\_revenue. Далее определяется товар, проданный наибольшее число раз. Сначала определятся само число продаж – максимальный элемент массива quantity. Это число записывается в переменную max\_sold\_num. После при помощи функции SearchQuan осуществляется поиск названия этого товара, которое записывается в переменную max\_sold. Аналогичным образом находятся названия самого прибыльного товара, наименее продаваемого и наименее прибыльного. Далее функция начинает текстовый вывод отчета. Осуществляется построчный вывод хеш-таблицы. Для каждой строки определяется выручка для того, чтобы определить её долю от общей, а после осуществляется вывод формата «Товар X был продан N раз(-а), выручка – Q руб., доля от общей выручки – K%». После этого функция выводит значение общей выручки, названия самого продаваемого и самого прибыльного, наименее продаваемого и наименее прибыльного товаров.

После этого инициализируем хеш-таблицу Sales, объявляем массивы количества продаж (Quantity), выручек (Revenues), названий товаров (Names) и ключей (Keys). Далее в защищенном блоке try осуществляется чтение файла формата csv. Из каждой строки добавляем соответствующие значения в массивы Quantity, Revenues и Names, в качестве ключа берем столбец «Номер заказа» и проверяем его на коллизию через цикл while: пока данный ключ key есть в массиве Keys, мы вызываем функцию Rehash. После цикла добавляем ключ в массив Keys. Значением для ключа является переменная-словарь data, в которую записываются остальные столбцы данной строки. После чтения и добавления всех строк вызывается функция FormatPrint с таблицей Sales и массивами Quantity и Revenues на входе. В случае ошибок с вводом предусмотрен блок except, в котором устанавливается флаг error = True.

Далее, используя библиотеку pandas осуществляем табличный вывод данных. Для начала вызывается функция сортировки SelectionSortSpec для массивов Revenues, Quantity и Names. После создается переменная общей выручки total\_revenue и массив долей от общей выручки Shares, значения в который вводятся через цикл по длине массива Revenues. Сами доли определяются как выручка товара X делить на общую выручку умножить на сто процентов, итоговое значение округляется до 1 знака после запятой. Далее через DataFrame создается и выводится таблица со столбцами: «Название товара», «Количество продаж», «Выручка, руб.», «Доля от общей выручки, %». Вывод происходит только в том случае, если флаг error не равен True, также и для круговых диаграмм.

В конце, используя методы библиотеки matplotlib, производится графический вывод выручек по товарам и количеств продаж по товарам. При помощи метода bar создаются гистограммы, причем названия товаров повернуты на 66 градусов, чтобы они друг друга не перекрывали. При помощи метода pie создаются круговые диаграммы с отображением процентных долей.

Итоговый вывод показан на рисунках 1-6.

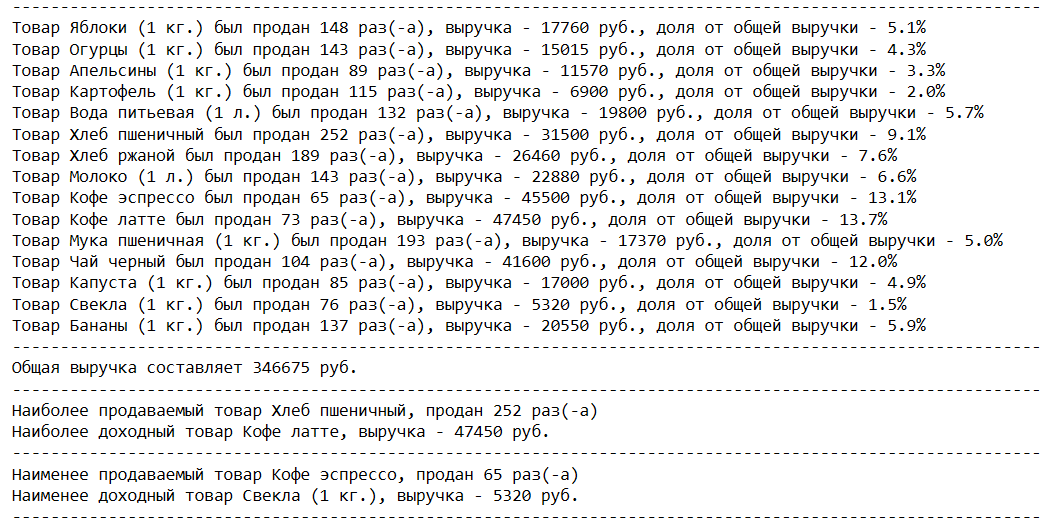


Рисунок 1 – Текстовый вывод отчета



Рисунок 2 – Табличный вывод отчета

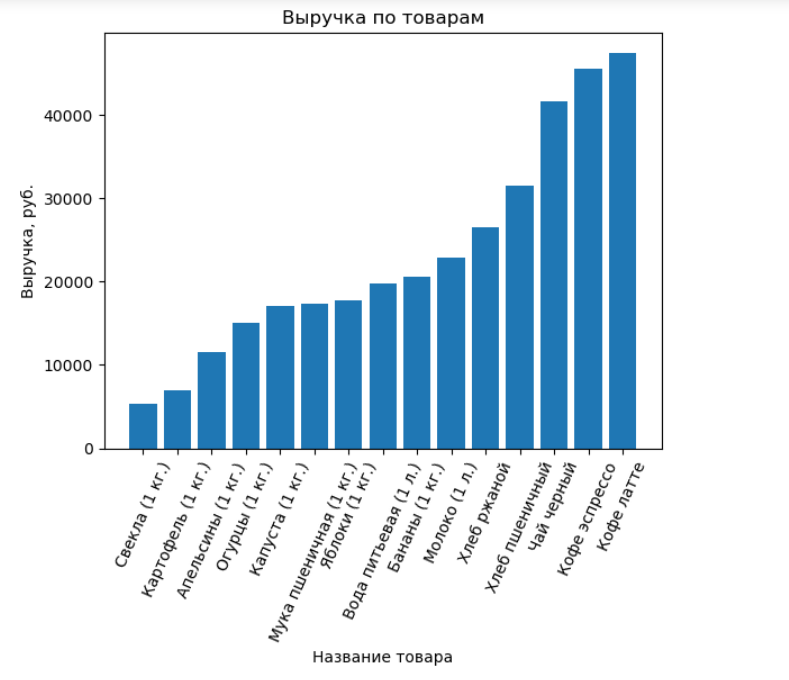


Рисунок 3 – Гистограмма выручек по товарам

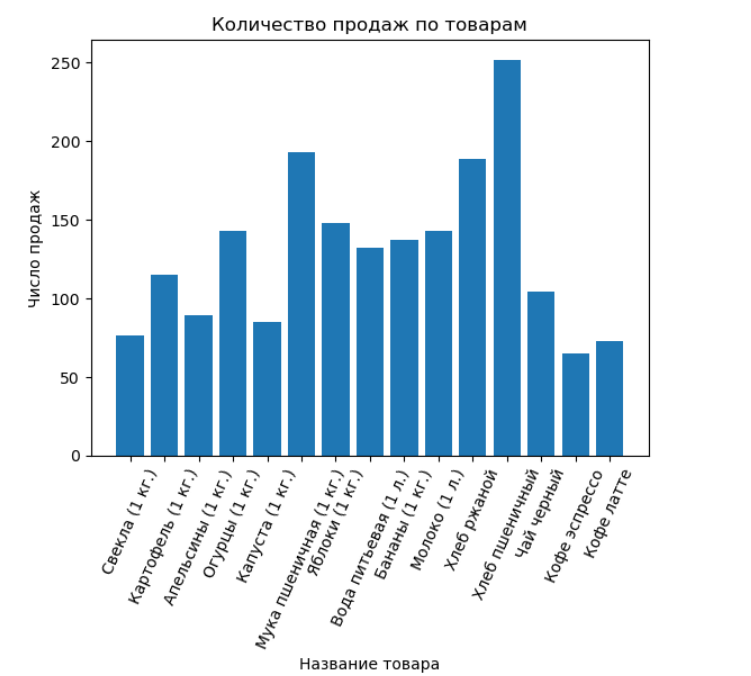


Рисунок 4 – Гистограмма числа продаж по товарам

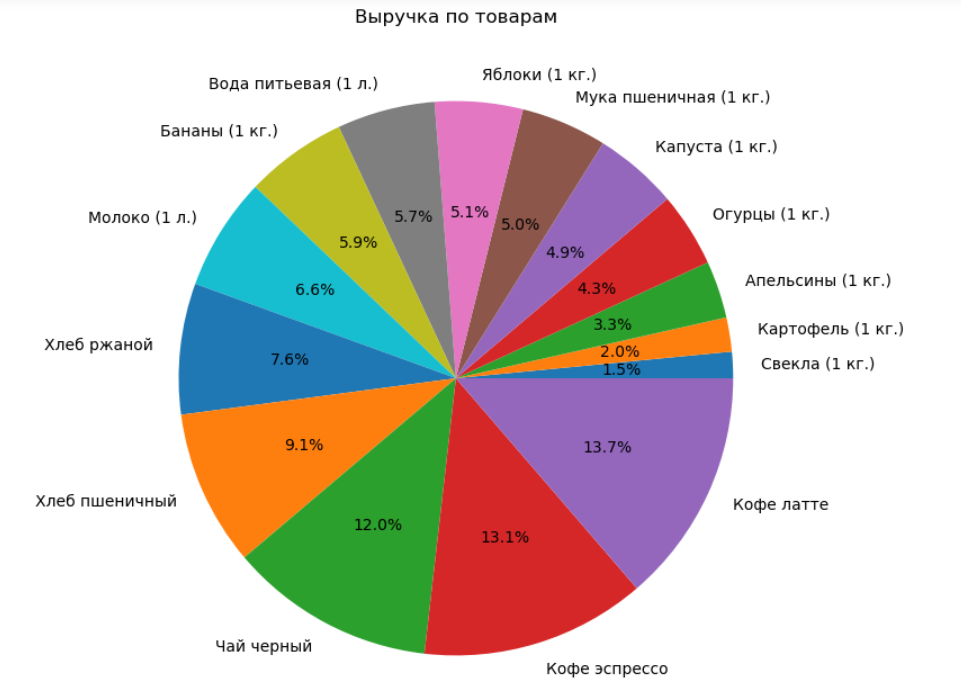


Рисунок 5 – Круговая диаграмма выручек по товарам

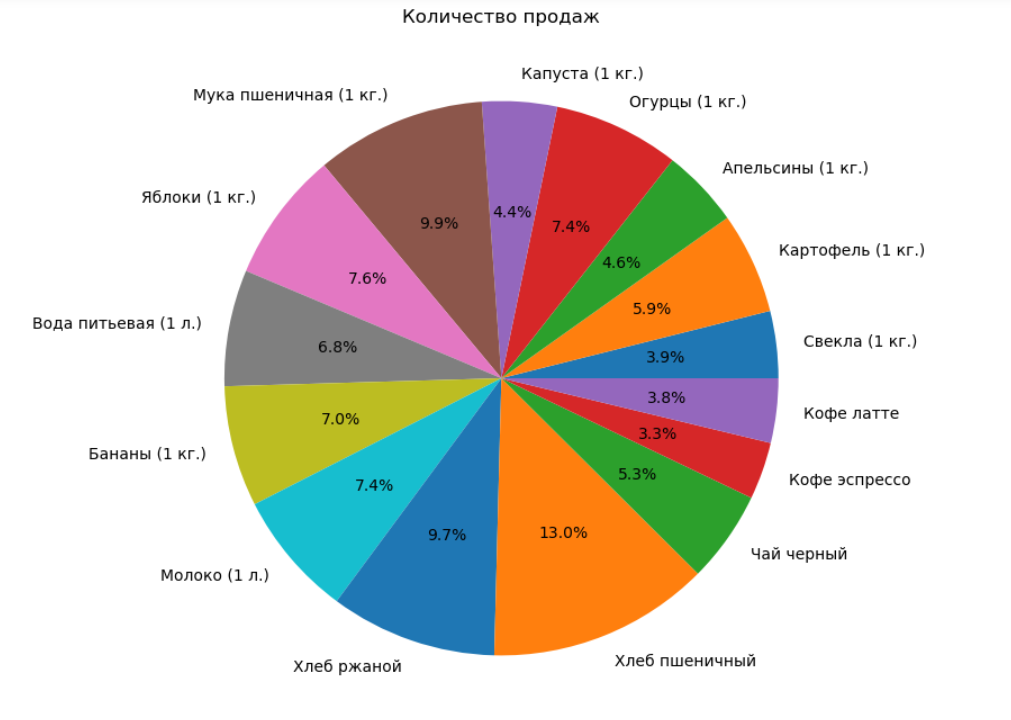


Рисунок 6 – Круговая диаграмма числа продаж по товарам

# Заключение

В ходе данной курсовой работы была создана программа, которая получает на вход csv-файл, данные из которого, содержащие информацию о числе продаж и выручке от разных товаров, записываются в хеш-таблицу. После этого программа выводит отчет по файлу: сколько раз какой товар был продан, какова была выручка и её доля от общей выручки, которая также выводится. Определяются и выводятся самые продаваемые и прибыльные, минимально проданные и минимально прибыльные товары. Вывод осуществляется текстом и в виде таблицы. Для наглядности программа также создает гистограммы и круговые диаграммы для графического отображения товаров по их выручке и продажам.