**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 БФИ2202

Охримчук П.А.

Проверил:

Симонов С. Е.

Москва, 2024

Оглавление

[1. Введение 3](#_Toc167283936)

[2. Постановка задачи 4](#_Toc167283937)

[3. Использование алгоритмов и структур данных в проекте 5](#_Toc167283938)

[4. Использование созданных методов и алгоритмов для отображения данных 7](#_Toc167283939)

[5. Заключение 9](#_Toc167283940)

# Введение

Алгоритмы и структуры данных являются основой эффективного программирования. Каждому разработчику важно не только использовать их, но и понимать, как они работают и как их правильно применять. Для этого необходимо глубокое знание их внутреннего устройства и механизмов функционирования.

В течение семестра мы изучали разнообразные алгоритмы и структуры данных. Основываясь на этих знаниях, нам было поручено разработать программу, которая демонстрирует применение этих принципов в реальной задаче. Цель данной работы заключается в создании приложения для анализа данных о продажах в магазине с использованием изученных алгоритмов и структур данных.

В первом разделе данного отчета описана поставленная задача. Во втором разделе детализированы используемые алгоритмы и структуры данных, а также принципы их работы. Итогом работы стало разработанное приложение на языке Python, которое эффективно анализирует данные о продажах, используя необходимые алгоритмы и структуры данных.

# Постановка задачи

Для выполнения работы необходимо 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 |

Также необходимо выполнить следующие пункты:

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.

# Использование алгоритмов и структур данных в проекте

В этом проекте использовались различные алгоритмы и структуры данных для эффективного решения задач по анализу данных о продажах в магазине. В частности, были использованы стек, массивы, хеш-таблица и сортировка слиянием. Также в проекте применялись несколько модулей Python для работы с данными и их визуализации.

Сортировка слиянием

В проекте используется алгоритм сортировки слиянием для упорядочивания данных о продажах. Этот алгоритм реализован с помощью рекурсии. На первом этапе массив данных разделяется на две равные части. Затем каждая часть сортируется отдельно с помощью того же алгоритма, после чего отсортированные части сливаются обратно в один массив. В реализации этого алгоритма использован стек для управления процессом разделения и объединения массивов. После сортировки каждая часть массива сохраняется в отдельном стеке, затем элементы из двух стеков сравниваются и объединяются в отсортированный массив.

Хеш-таблица

Для хранения информации о продажах и выручке каждого товара используется хеш-таблица. Реализация хеш-таблицы основана на методе цепочек для разрешения коллизий. Каждый элемент, добавляемый в хеш-таблицу, хэшируется по ключу, и значение помещается в соответствующий элемент хеш-таблицы. В случае коллизии (когда два ключа хэшируются в один и тот же индекс), элементы добавляются в связанный список, который хранится в этом индексе. Таким образом, возможность коллизий обрабатывается путем добавления элементов в цепочку.

Стек

Стек используется в проекте для реализации алгоритма сортировки слиянием. Стек представляет собой структуру данных, работающую по принципу Last In, First Out (LIFO), что означает, что последний добавленный элемент будет первым извлеченным. В контексте сортировки слиянием, стек используется для управления процессом разделения и объединения массивов данных.

Используемые модули Python

Для реализации проекта использовались следующие модули Python:

csv: Этот модуль использовался для чтения данных из CSV-файлов.

pandas: Модуль pandas применялся для обработки и анализа данных. Он обеспечил гибкие структуры данных и функции, которые упростили манипуляции с большими объемами информации.  
matplotlib: Модуль matplotlib использовался для визуализации данных. С его помощью были построены графики, отображающие долю каждого товара в общей выручке магазина, что сделало результаты анализа более наглядными.  
tabulate: Этот модуль был применен для форматирования и вывода данных в табличной форме в консоли. Он позволил представить результаты анализа в удобочитаемом виде.

# Использование созданных методов и алгоритмов для отображения данных

Для отображения данных о продажах и результатов анализа в проекте активно используются методы класса SalesDataProcessor и реализованные алгоритмы.

Описание процесса отображения данных:

Чтение данных из файла: Сначала данные о продажах считываются из CSV файла с использованием модуля csv.

Обработка данных: Каждая строка данных обрабатывается с помощью метода process\_row() класса SalesDataProcessor. Этот метод вычисляет общую выручку магазина, максимальное количество проданных единиц товара, заполняет хеш-таблицы с информацией о продажах и количестве проданных товаров.

Вычисление статистики: После обработки всех данных вызывается метод calculate\_statistics(), который вычисляет общую выручку магазина, товар, который был продан наибольшее количество раз, и товар, который принес наибольшую выручку.

Вывод статистики: Результаты анализа выводятся с помощью метода print\_statistics(), который печатает информацию о количестве проданных единиц каждого товара и их долю в общей выручке магазина.

Визуализация данных: Для наглядного представления доли каждого товара в общей выручке магазина используется модуль matplotlib, который строит гистограмму с данными, полученными из хеш-таблицы с информацией о продажах.

Используемые алгоритмы:

Сортировка слиянием: для упорядочивания данных о продажах применяется алгоритм сортировки слиянием. Он обеспечивает эффективное упорядочивание больших объемов данных, что позволяет быстро анализировать данные и находить наиболее продаваемые товары.

Описание функций:

Функция plot\_results: Эта функция создает гистограмму, отображающую долю выручки каждого товара в общем доходе магазина. Сначала она добавляет новый столбец в датасет, рассчитывая долю выручки для каждого товара как процент от общей выручки. Затем формирует гистограмму, задавая размер графика и параметры осей, и отображает её с вертикальными подписями для каждого товара.

Функция print\_table\_to\_console: Эта функция выводит содержимое датасета в консоль в виде таблицы. Она форматирует таблицу с помощью библиотеки tabulate, чтобы данные были представлены в удобочитаемом виде.

Пример использования функций:

Датасет загружается из CSV файла. Создается экземпляр класса SalesDataProcessor, который обрабатывает строки датасета и собирает статистику. Вычисляются общая выручка, самый продаваемый товар и товар с наибольшей выручкой. Эти результаты выводятся на экран. Таблица данных выводится в консоль. Строится и отображается гистограмма, показывающая долю выручки каждого товара.

‘

# Заключение

В рамках работы была разработана программа на Python, предназначенная для анализа данных о продажах и визуализации результатов. Программа использует разнообразные алгоритмы и структуры данных для эффективной обработки и представления информации.

Создана программа, осуществляющая чтение данных из CSV-файла.  
Проведены расчеты общей выручки магазина. Определены товары, обеспечившие наибольшую выручку и проданные в наибольшем количестве. Сформирован отчет, включающий данные о товарах, такие как количество продаж и их доля в общей выручке магазина.

Проведена сортировка товаров по имени с выводом списка в виде таблицы, используя библиотеку tabulate. Визуализация доли каждого товара в общей выручке с помощью гистограммы, построенной с использованием библиотеки matplotlib. Вывод подробной таблицы с данными о продажах в консоль.

Для выполнения этих задач использовались структуры данных, такие как стек и хеш-таблицы, что позволило оптимизировать процессы обработки и хранения информации.