МИНИСТЕРСТВО ЦИФРОВОГО РАЗВИТИЯ, СВЯЗИ И

МАССОВЫХ КОММУНИКАЦИЙ РОССИЙСКОЙ ФЕДЕРАЦИИ

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

Выполнил:

студент группы БФИ2201

Жанбаев И.

Проверил:

Симонов С. Е.

Москва 2024

Оглавление

[Введение 3](#_Toc135165313)

[Техническое задание 4](#_Toc135165314)

[Ход выполнения работы 5](#_Toc135165315)

[Вывод 9](#_Toc135165316)

Введение

В данной курсовой работе мы будем использовать различные структуры данных и алгоритмы для написания программы, которая рассчитывает общую выручку магазина, находит товар, который был продан наибольшее количество раз, и находит товар, который принес наибольшую выручку.

Одной из используемых структур данных будет хеш-таблица. Хеш-таблица представляет собой структуру данных, в которой элементы хранятся в виде пар ключ-значение. Ключ является уникальным числом, используемым для индексации значений, а значение представляет данные, связанные с этим ключом.

Когда хеш-функция генерирует один и тот же индекс для нескольких ключей, возникает коллизия хеш-таблицы. Для разрешения коллизий мы будем использовать метод цепочек. Этот метод заключается в том, что если хеш-функция выделяет один и тот же индекс для нескольких элементов, то они будут храниться в одном и том же индексе с помощью двусвязного списка.

Для нахождения товара, который был продан наибольшее количество раз, мы будем использовать методы сортировки.

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 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

Ход работы

Используем оператор try-except, для обработки ввода/вывода без ошибок.

Для с начала импортируем следующие библиотеки: pandas, matplotlib.pyplot и csv. После этого мы приступим к написанию кода для работы с хэш-таблицей. В этой таблице мы сначала создадим класс для узла списка (Node), а затем класс для самой хэш-таблицы (HashTable). В классе HashTable мы определим следующие методы: инициализация (init), добавление элемента в таблицу (add), извлечение значения по ключу (get) и изменение размера таблицы (resize).

Далее мы создадим функцию для сортировки данных. Мы выбрали метод "пузырьковой сортировки", потому что он простой и понятный, хотя и не самый эффективный для больших наборов данных. Этот метод сортировки заключается в многократном просмотре списка, сравнении каждой пары соседних элементов и замене их местами, если они расположены в неправильном порядке. Этот процесс повторяется до тех пор, пока список не будет полностью отсортирован.

Мы объявляем переменную, которая будет содержать наш исходный файл csv. Затем мы вызываем нашу функцию для расчета.

Мы определим следующие переменные: общий доход и максимальное количество проданных товаров (vyr, max\_tovar); строковую переменную, в которую мы будем записывать название товара, проданного в наибольшем количестве (max\_count); переменную для хранения количества строк в csv-файле (count\_row) и двух массивы: массив для ключей / названий продуктов и массив для выручки (names\_array, vyr\_array).

Затем мы инициализируем хэш-таблицы. Мы пройдемся по всем строкам в файле, добавив product/key в names\_array. Используя функцию append, мы добавим в массив строку с индексом 2, которая содержит информацию о названии продукта. В поле vyr\_array мы добавим строку с индексом 6, которая содержит информацию об общей стоимости. Мы преобразуем это значение в тип int.

Используя прямой поиск, найдем товар, который продался больше всех, а также найдем название этого товара.

В хеш-таблицу hsht2 мы записываем значения выручки/ключа и товара. Добавляем в hsht1 товар/ключ и количество его продаж, а также товар/ключ и его выручку. Для получения значений из хеш-таблицы мы используем метод get, который определен в классе хеш-таблицы. Затем мы сортируем массив с выручками. Для возвращения общей выручки, названия товара, который был продан наибольшее количество раз, и товара, который принес наибольшую выручку, мы используем переменные vyr, max\_count и hsht2.get(str(result[-1])).

С помощью for выводим информацию об общей выручке магазина, количестве проданных единиц каждого товар и доле каждого товара в общей выручке обращаясь к hsht1.

Также с библиотекой matplotlib.pyplot визуализируем результаты в двумерной графике.

Результаты работы:

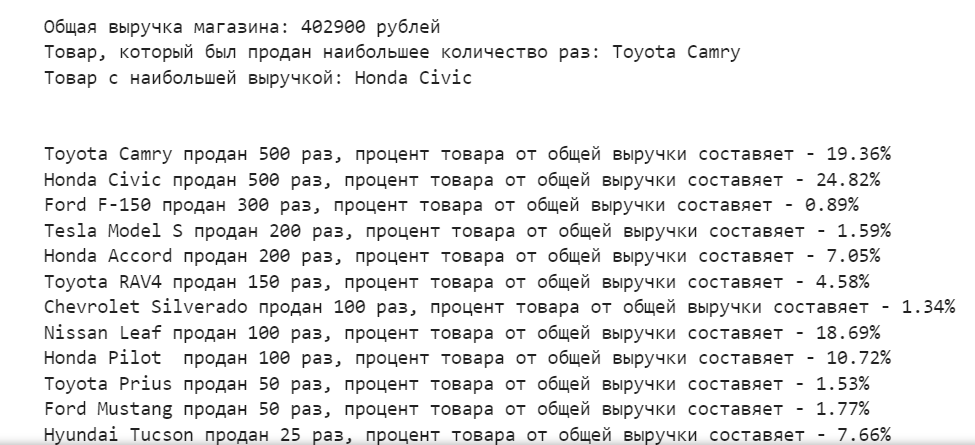


Рисунок 1 – Вывод общей информаций

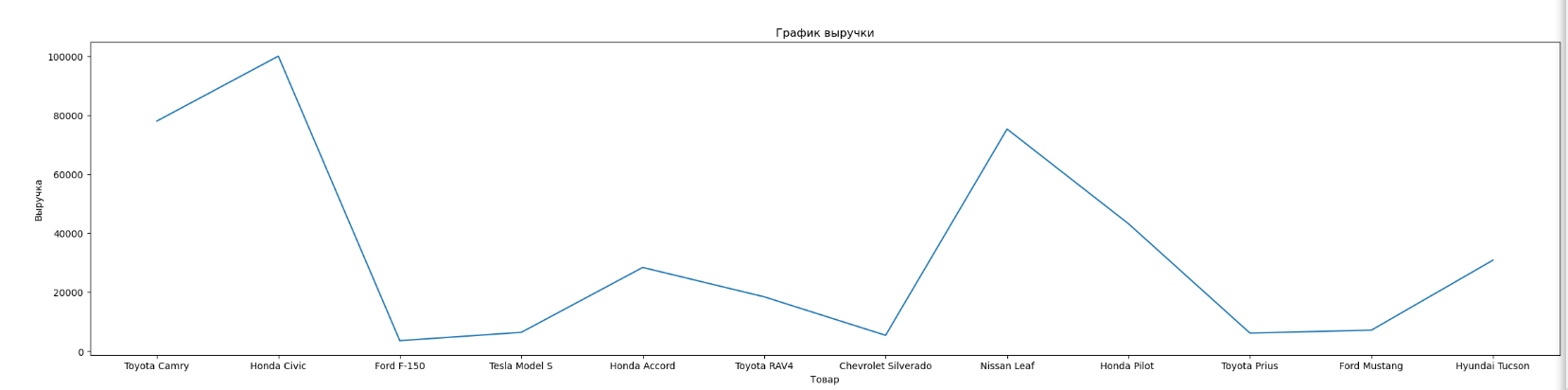


Рисунок 2 – График выручки всех товаров

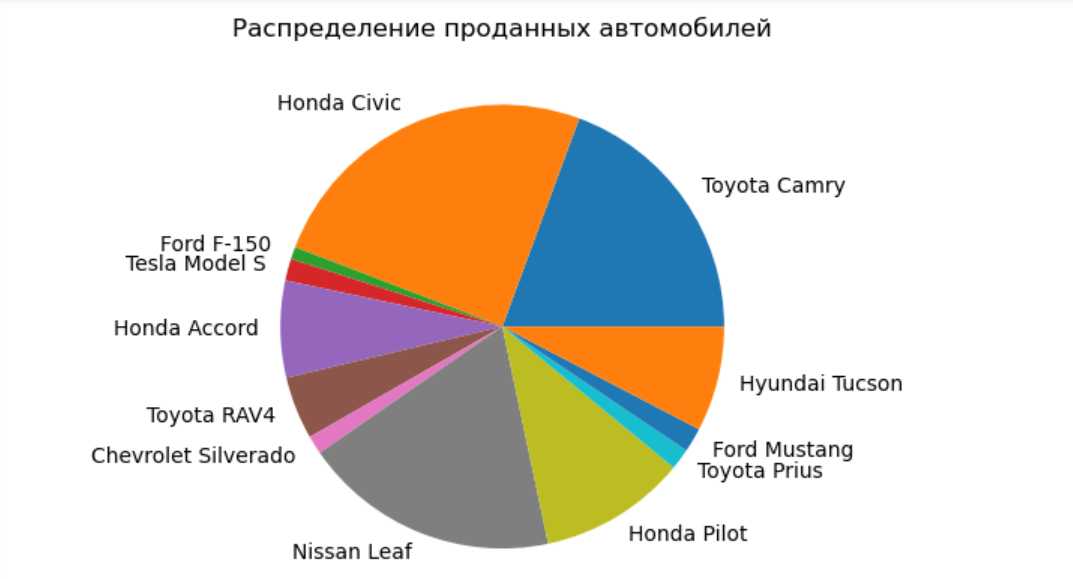


Рисунок 3 – Круговая диаграмма долей товаров от общей стоимости

Вывод

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 товар и доле каждого товара в общей выручке.