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»

Курсовая работа

по дисциплине «Структуры и алгоритмы обработки данных»

Выполнила:

студентка группы БФИ2203

Мячина Лада Андреевна

Проверил:

Симонов Сергей Евгеньевич

Москва, 2024 г.

Оглавление

[Введение 3](#_Toc167312158)

[Ход выполнения работы 5](#_Toc167312159)

[Заключение 9](#_Toc167312160)

Введение

Структуры данных и алгоритмы обработки данных играют важную роль в современном мире, где информация является одним из самых ценных ресурсов. Знание структур данных и алгоритмов позволяет эффективно организовывать и хранить данные. Структуры данных, такие как списки, массивы, деревья и графы, позволяют представить данные в удобной форме и обеспечивают быстрый доступ к ним. Алгоритмы обработки данных, такие как сортировка и поиск, позволяют эффективно оперировать большими объемами данных и выполнять операции быстро и точно. структуры данных и алгоритмы обработки данных помогают улучшить производительность и оптимизировать использование ресурсов. Эффективные алгоритмы позволяют обрабатывать большие объемы данных за короткое время, что важно для современных приложений и систем. Это также помогает снизить нагрузку на процессор и память, что позволяет оптимизировать использование ресурсов и повысить производительность системы в целом.

Разработка программы, способной анализировать данные о продажах товаров, позволит бизнесу эффективно управлять своей деятельностью. Расчет общей выручки и определение наиболее популярных и прибыльных товаров поможет магазину принимать обоснованные решения по улучшению своей работы и увеличению прибыли.

Цель курсовой работы: разработка сервиса парсинга и валидации данных.

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 выполнения работы

Программа для курсовой работы была реализована на языке программирования Python.

Часть содержимого CSV файла представлено на рисунке 1.

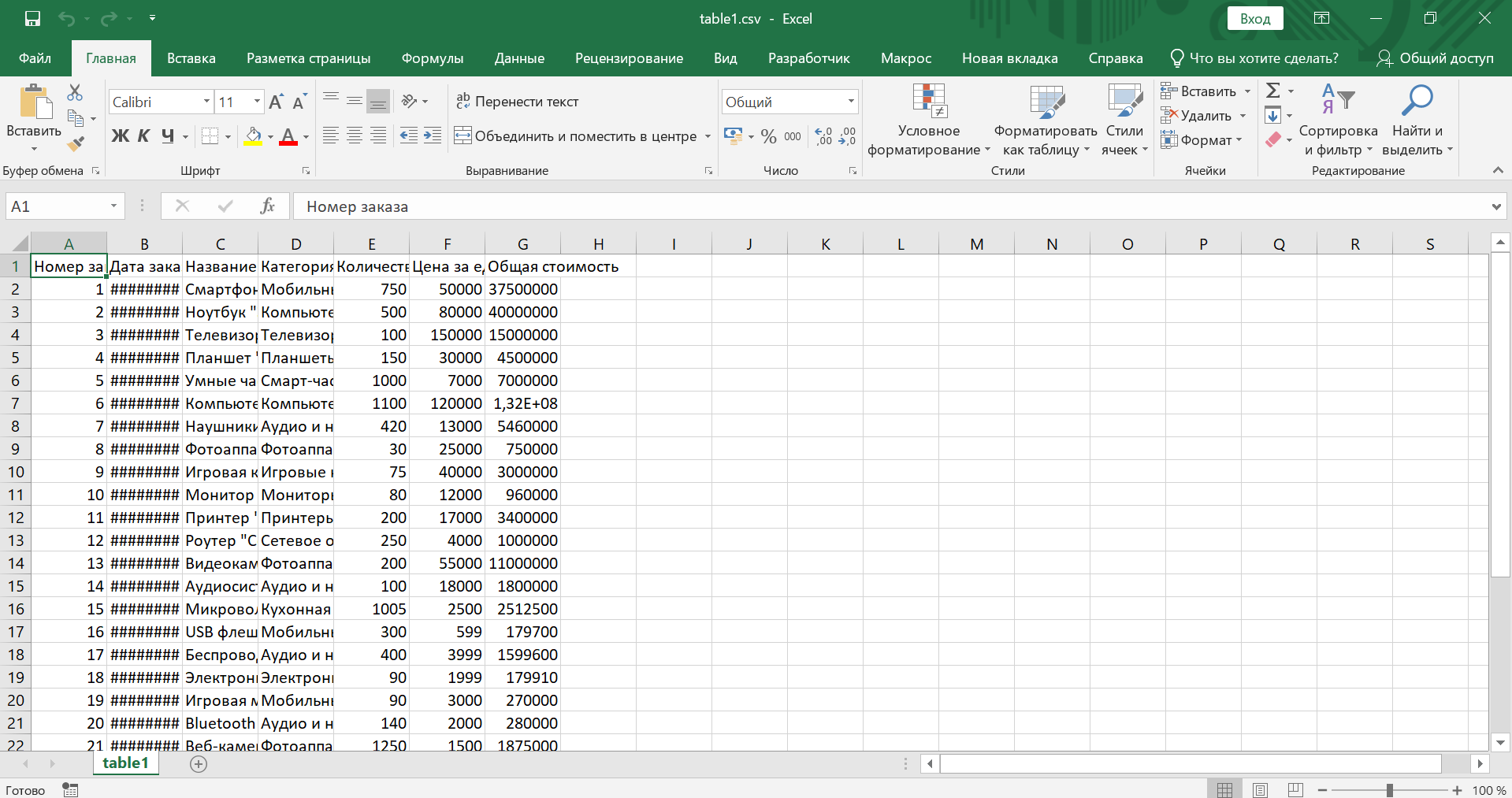


Рисунок 1 – Содержимое CSV файла

Для выполнения курсовой работы импортированы библиотеки: для работы с CSV файлами, для работы с таблицами данных (DataFrame) и для построения наглядной гистограммы.

После этого реализована функция открытия CSV файла, которая считывает все данные в список, так чтобы, каждая строка файла стала отдельным элементом списка.

Следующий шаг - создание хеш-таблицы, где ключ название товара, а значение - количество продаж и общая выручка для этого товара. Создана функция, которая проходит по каждой строке в списке, начиная со второго элемента, извлекает необходимые данные о продажах, обновляет словарь с информацией о количестве продаж и общей выручке для каждого товара, при этом если товар уже есть в словаре продаж по товару, увеличивается количество продаж и общая выручка, если нет, то создается новая запись. Эта проверка необходима для выявления ошибки ввода-вывода.

Следующим шагом проходимся по всей таблице в поисках наибольших значений по количеству продаж, а также общей выручке. Параллельно рассчитывается вся выручка магазина. Здесь реализована сортировка вставками. Общая суть сортировки: 1) Перебираются элементы в неотсортированной части массива. 2) Каждый элемент вставляется в отсортированную часть массива на то место, где он должен находиться. Общую выручку магазина вычисляем, просуммировав общую стоимость из каждой строки CSV файла.

Для того, чтобы создать отчет, содержащий информацию об общей выручке магазина, количестве проданных единиц каждого товара и доле каждого товара в общей выручке, необходимо подготовить данные, это реализовано с помощью библиотеки Pandas. Сначала созданы пустые списки для хранения информации о продуктах, количестве проданных единиц, выручке и доле в общей выручке. Далее внутри цикла происходит следующее:

1. Количество проданных единиц и общая выручка извлекаются из данных о продажах.
2. Значения добавляются в соответствующие списки.
3. Вычисляется доля в общей выручке путем деления выручки за товар на общую выручку и умножения на 100.

После этого создан DataFrame из списков (они стали столбцами) и установлен 'Товар' как индекс для удобства чтения таблицы. Также установлена опция display.width, чтобы таблица отображалась полностью. В результате выполнения кода получается DataFrame, который содержит информацию о продуктах, количестве проданных единиц, выручке и доле в общей выручке для каждого продукта.

Результат работы программы представлен на рисунках 2 - 5

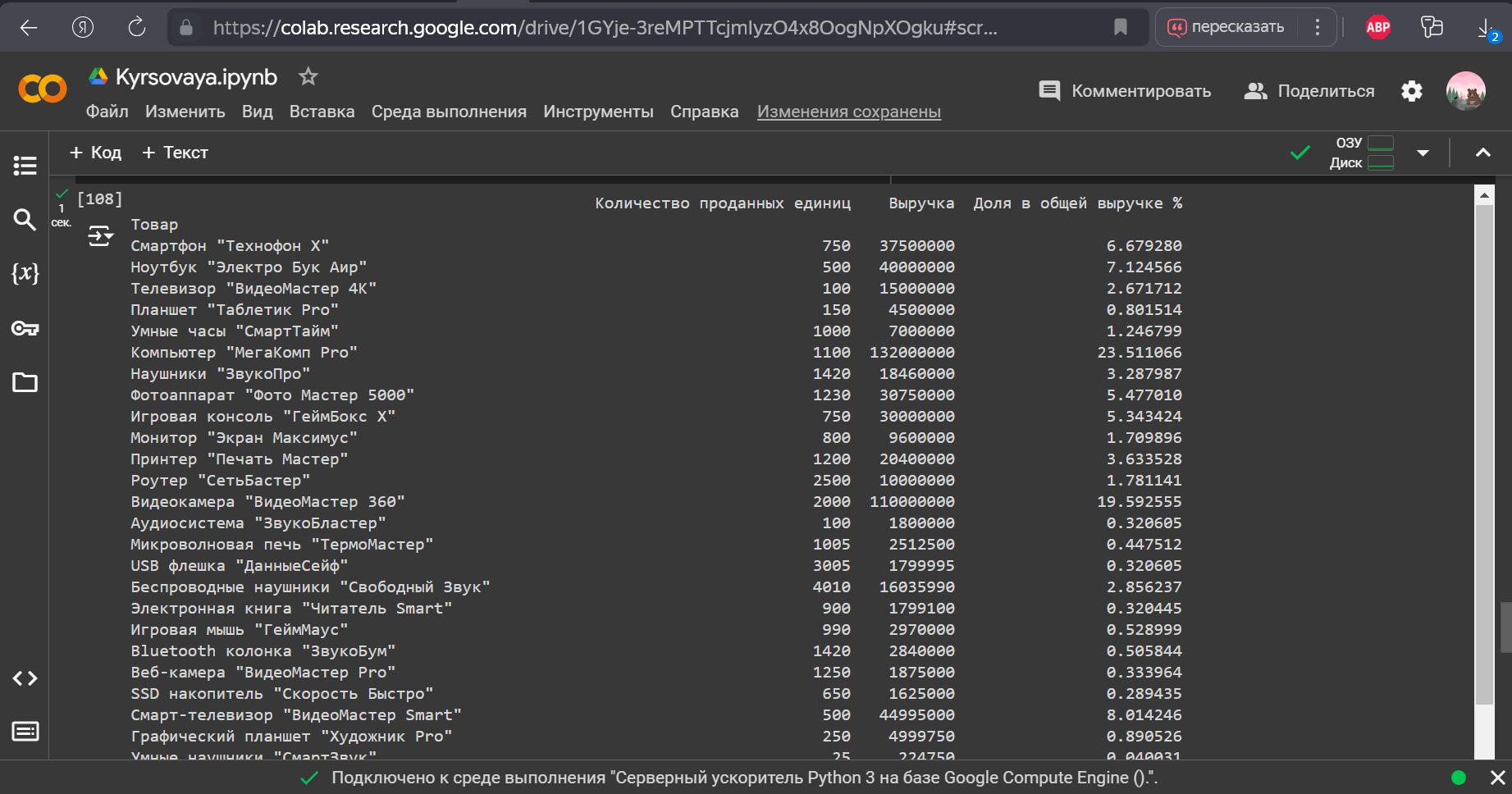


Рисунок 2 – Результат работы программы (общая информация о всех товарах)

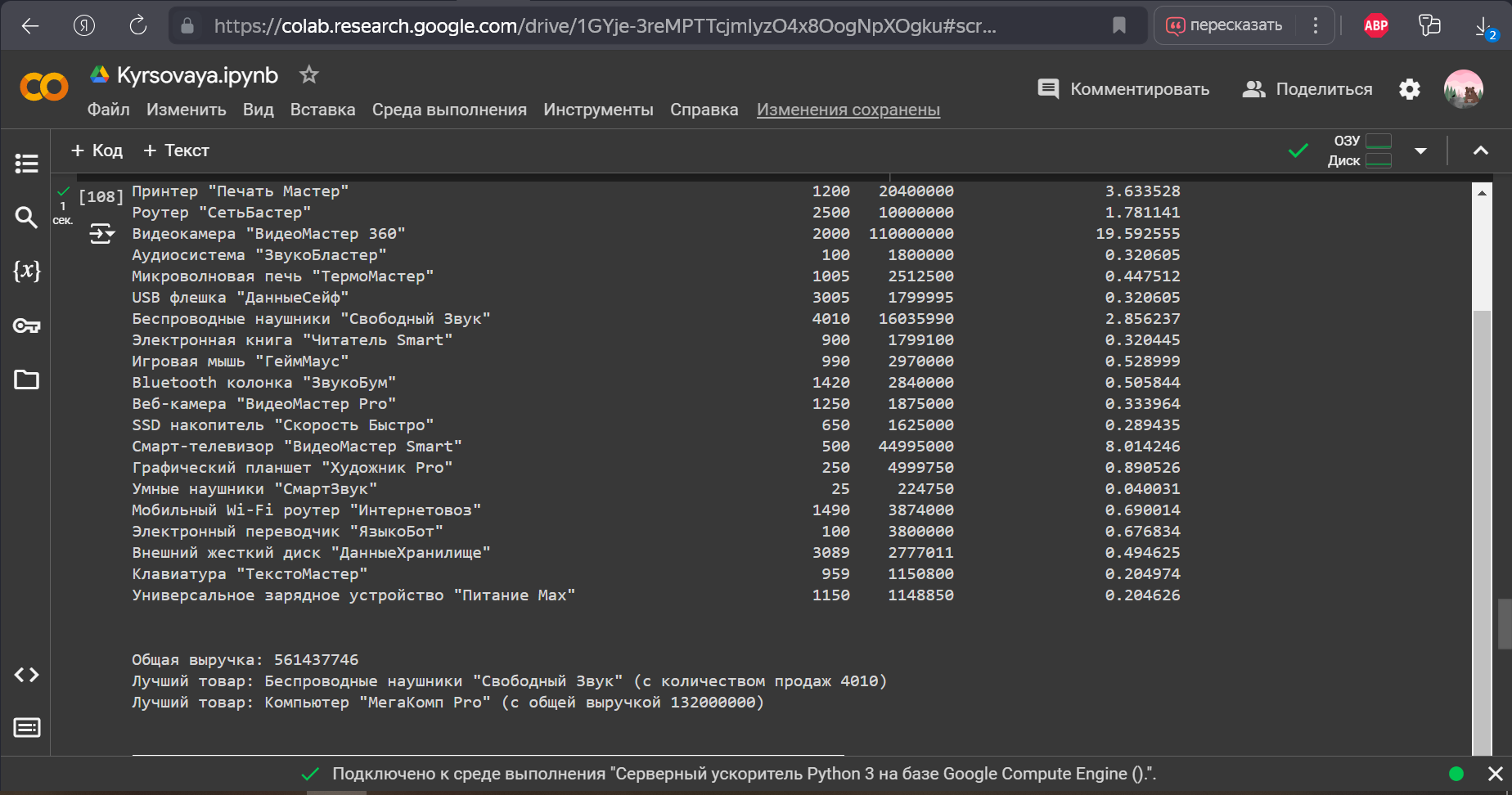


Рисунок 3 – Результат работы программы (лучшие товары и общая выручка)

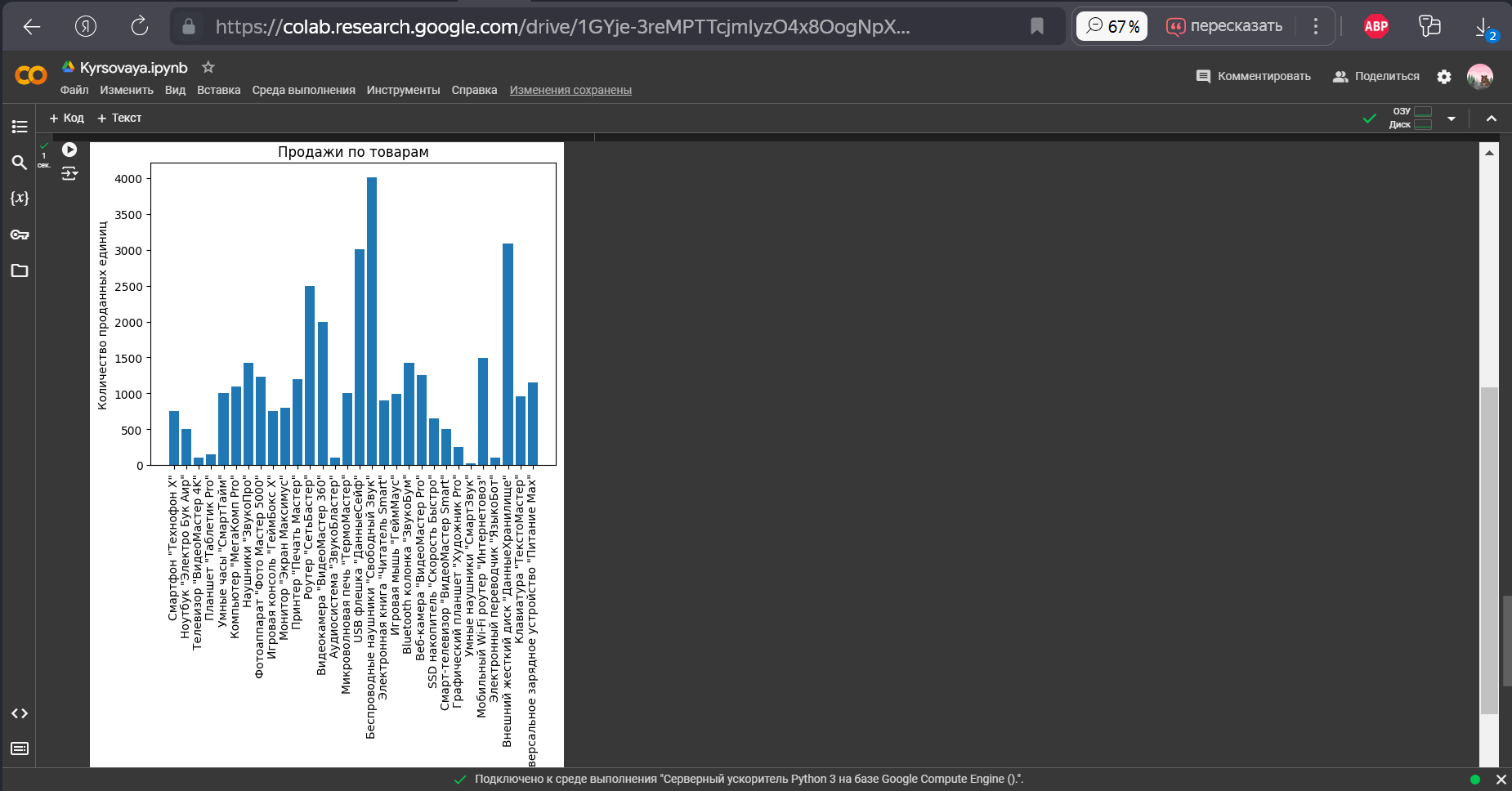


Рисунок 4 – Результат работы программы (гистограмма, отображающая количество проданных единиц по каждому товару)

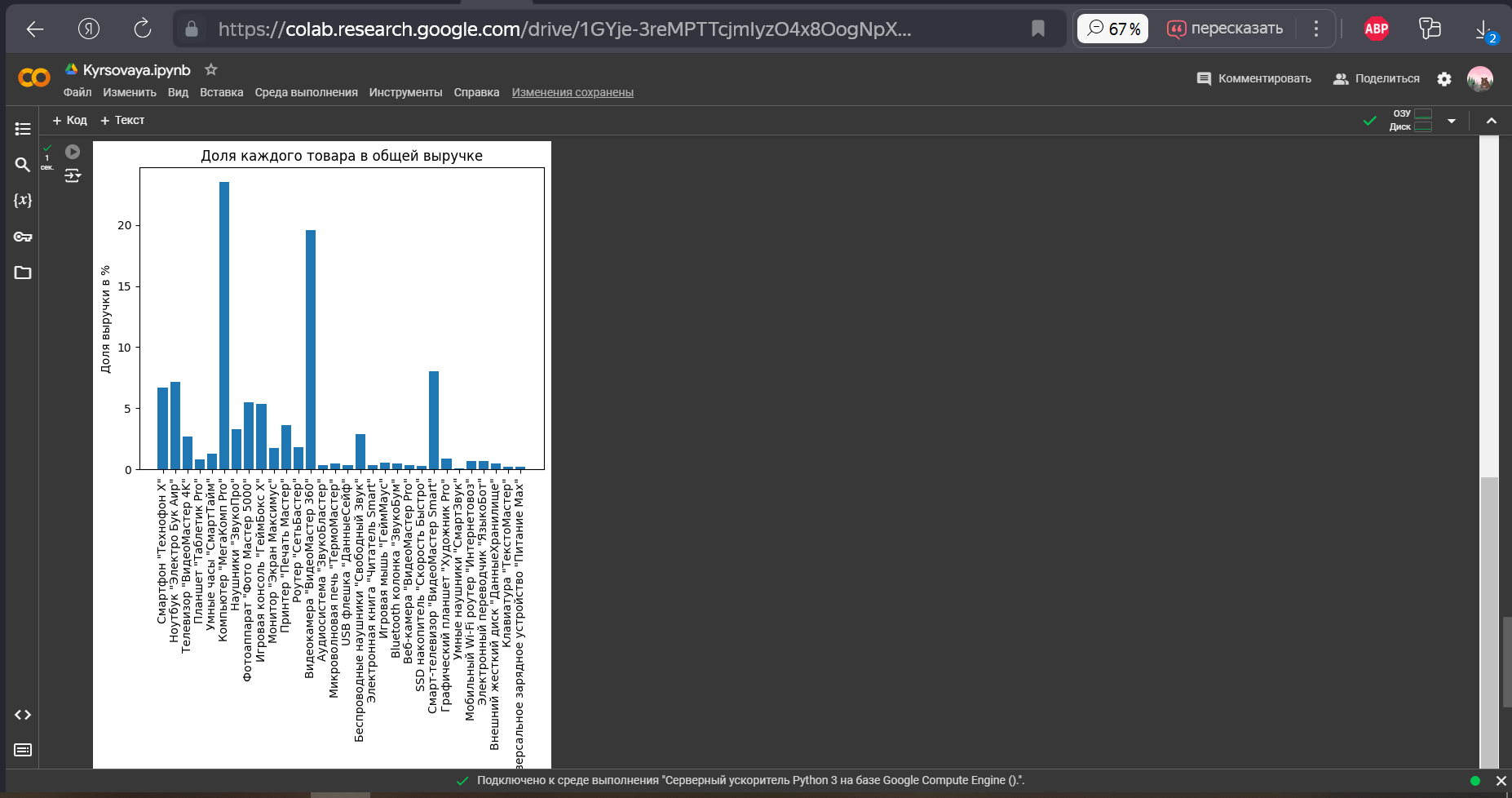


Рисунок 5 – Результат работы программы (гистограмма, отображающая долю в общей выручке)

Заключение

В процессе выполнения данной курсовой работе, был разработан сервис парсинга csv-файла, валидации данных, а также добавлена возможность вывода конечного результата. Выполнены следующие задачи:

1. Рассчитана общая выручка магазина.

2. Найден товар, который был продан наибольшее количество раз.

3. Найден товар, который принес наибольшую выручку.

4. Составлен отчет, содержащий информацию об общей выручке магазина, количестве проданных единиц каждого товар и доле каждого товара в общей выручке.

При этом использовались структуры данных, такие как массивы и хеш-таблицы, и алгоритм обработки данных - сортировка вставками. Также необходимо учтены возможные ошибки ввода-вывода и обработаны в соответствии с требованиями.