|  |
| --- |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования **«МИРЭА − Российский технологический университет»**  **РТУ МИРЭА** |

**Институт информационных технологий (ИИТ)**

**Кафедра прикладной математики (ПМ)**

**ОТЧЕТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ**

по дисциплине «Технологии и инструментарий анализа больших данных»

**Практическая работа № 8**

|  |  |  |  |
| --- | --- | --- | --- |
| Выполнил студент группы ИНБО-01-17 | ИВБО-06-20 Чурилов А.К. | (подпись) | |
| Принял | Парамонов А.А. | (подпись) | |
| Отчет представлен | «7» декабря 2023 г. | |  | |

Москва 2023 г.

**Задание**

1. Загрузить данные Market\_Basket\_Optimisation.csv.

2. Визуализировать данные (отразить на гистограммах относительную и

фактическую частоту встречаемости для 20 наиболее популярных

товаров).

3. Применить алгоритм Apriori, используя 3 разные библиотеки

(apriori\_python, apyori, efficient\_apriori). Подобрать гиперпараметры для

алгоритмов так, чтобы выводилось порядка 10 наилучших правил.

4. Применить алгоритм FP-Growth из библиотеки fpgrowth\_py. Подобрать

гиперпараметры для алгоритма так, чтобы выводилось порядка 10

наилучших правил.

5. Сравнить время выполнения всех алгоритмов и построить гистограмму.

6. Загрузить данные data.csv.

7. Визуализировать данные (отразить на гистограммах относительную и

фактическую частоту встречаемости для 20 наиболее популярных

товаров).

8. Применить алгоритм Apriori, используя 3 разные библиотеки

(apriori\_python, apyori, efficient\_apriori). Подобрать гиперпараметры для

алгоритмов так, чтобы выводилось порядка 10 наилучших правил.

9. Применить алгоритм FP-Growth из библиотеки fpgrowth\_py. Подобрать

гиперпараметры для алгоритма так, чтобы выводилось порядка 10

наилучших правил.

10.Сравнить время выполнения всех алгоритмов и построить гистограмму.

11.Сформулировать выводы и сделать отчет.

**Ход работы**

**Задание 1 и 6**

1. Загрузить данные Market\_Basket\_Optimisation.csv.

6. Загрузить данные data.csv.

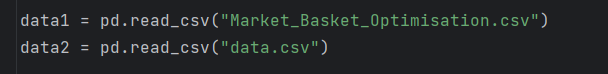


Рис. 1 – задание 1, 6

**Задание 2 и 7**

2. Визуализировать данные (отразить на гистограммах относительную и

фактическую частоту встречаемости для 20 наиболее популярных

товаров).

7. Визуализировать данные (отразить на гистограммах относительную и

фактическую частоту встречаемости для 20 наиболее популярных

товаров).

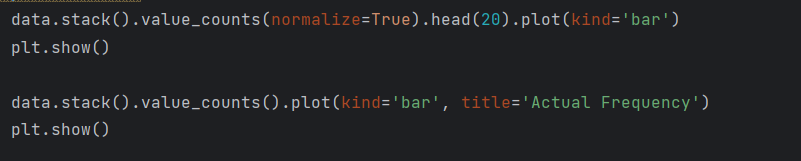


Рис. 2 – задание 2, 7

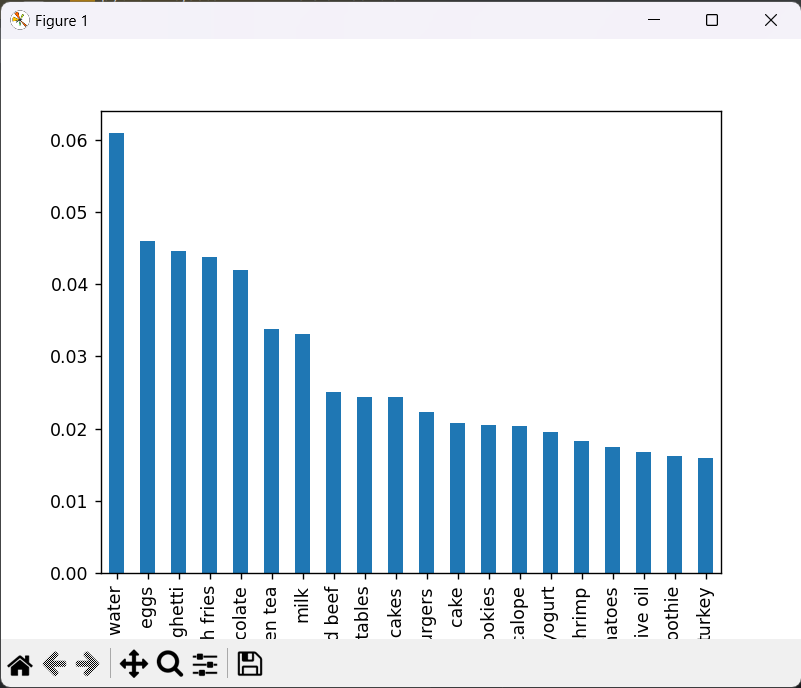


Рис. 3 – 20 наиболее популярных товаров

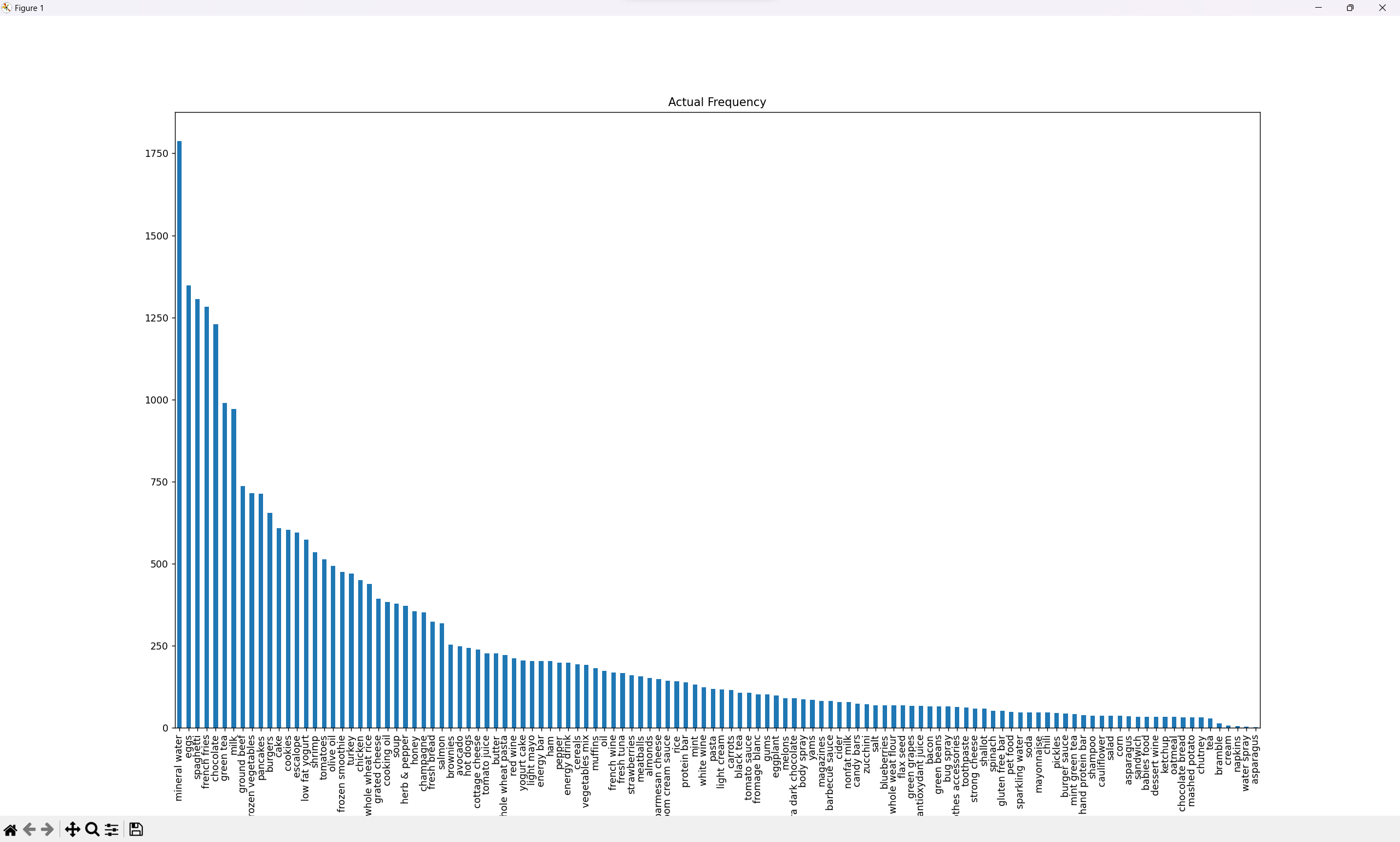


Рис. 4 – популярность товаров

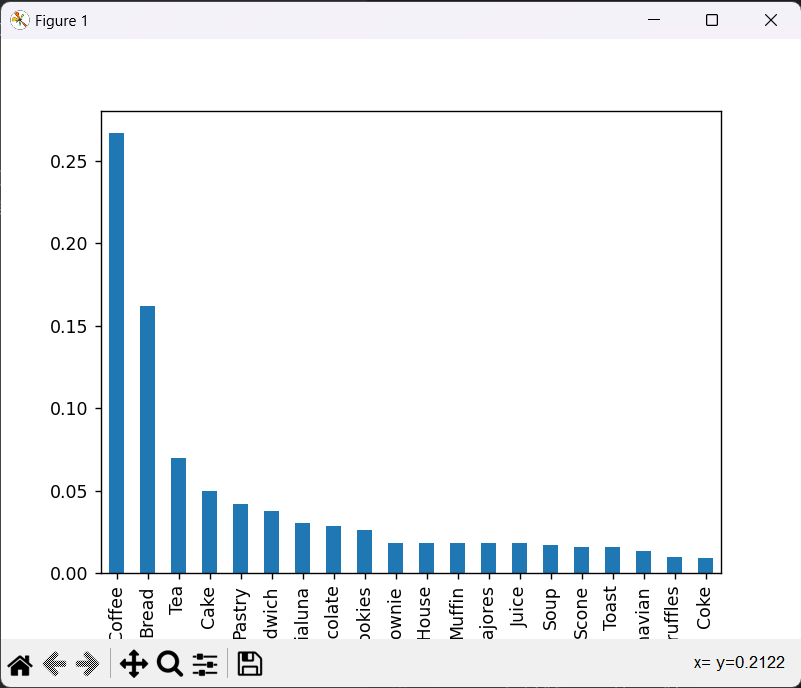


Рис. 5 – 20 наиболее популярных товаров

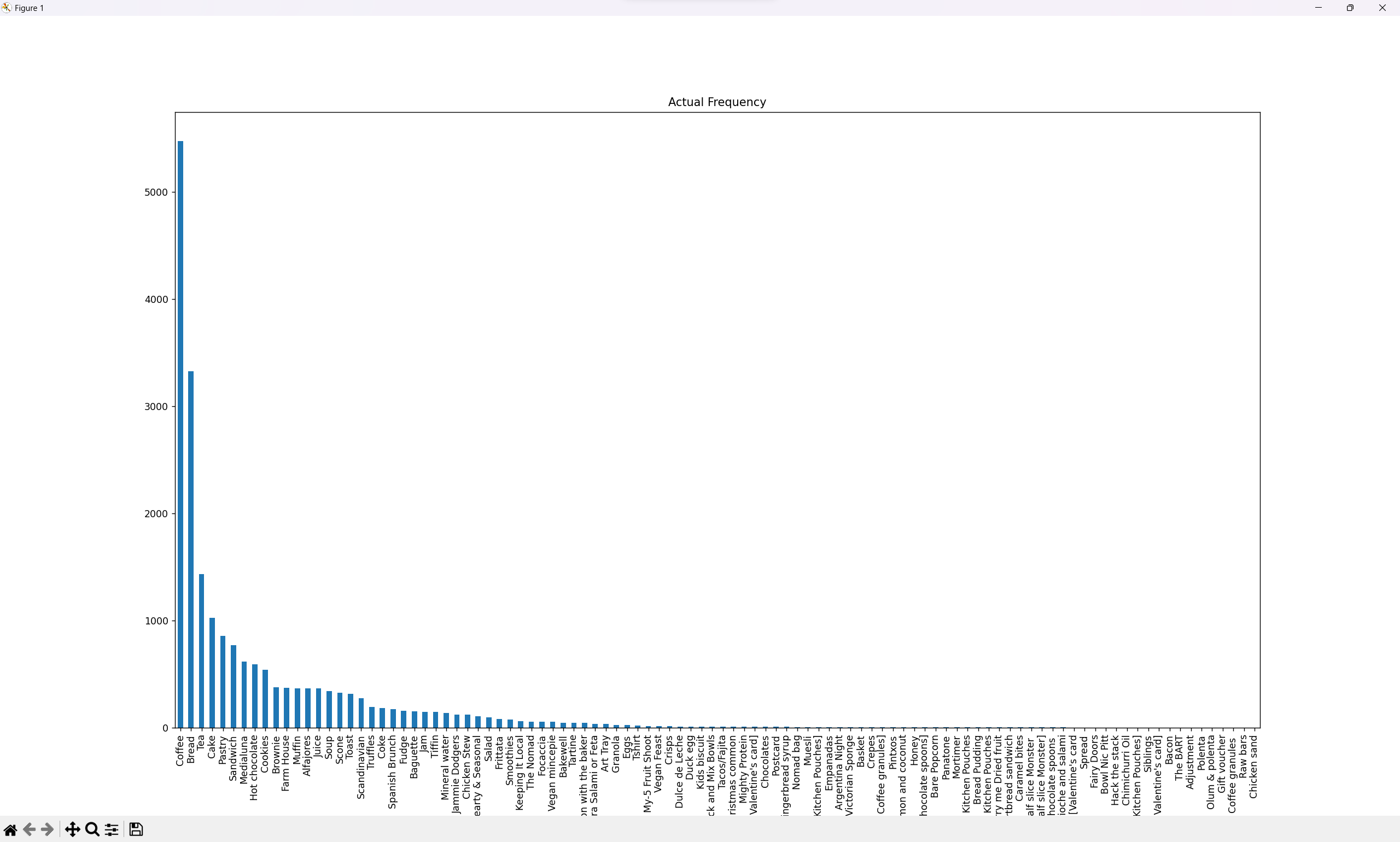


Рис. 6 – популярность товаров

# Задание 3 и 8

3. Применить алгоритм Apriori, используя 3 разные библиотеки

(apriori\_python, apyori, efficient\_apriori). Подобрать гиперпараметры для

алгоритмов так, чтобы выводилось порядка 10 наилучших правил.

8. Применить алгоритм Apriori, используя 3 разные библиотеки

(apriori\_python, apyori, efficient\_apriori). Подобрать гиперпараметры для

алгоритмов так, чтобы выводилось порядка 10 наилучших правил.

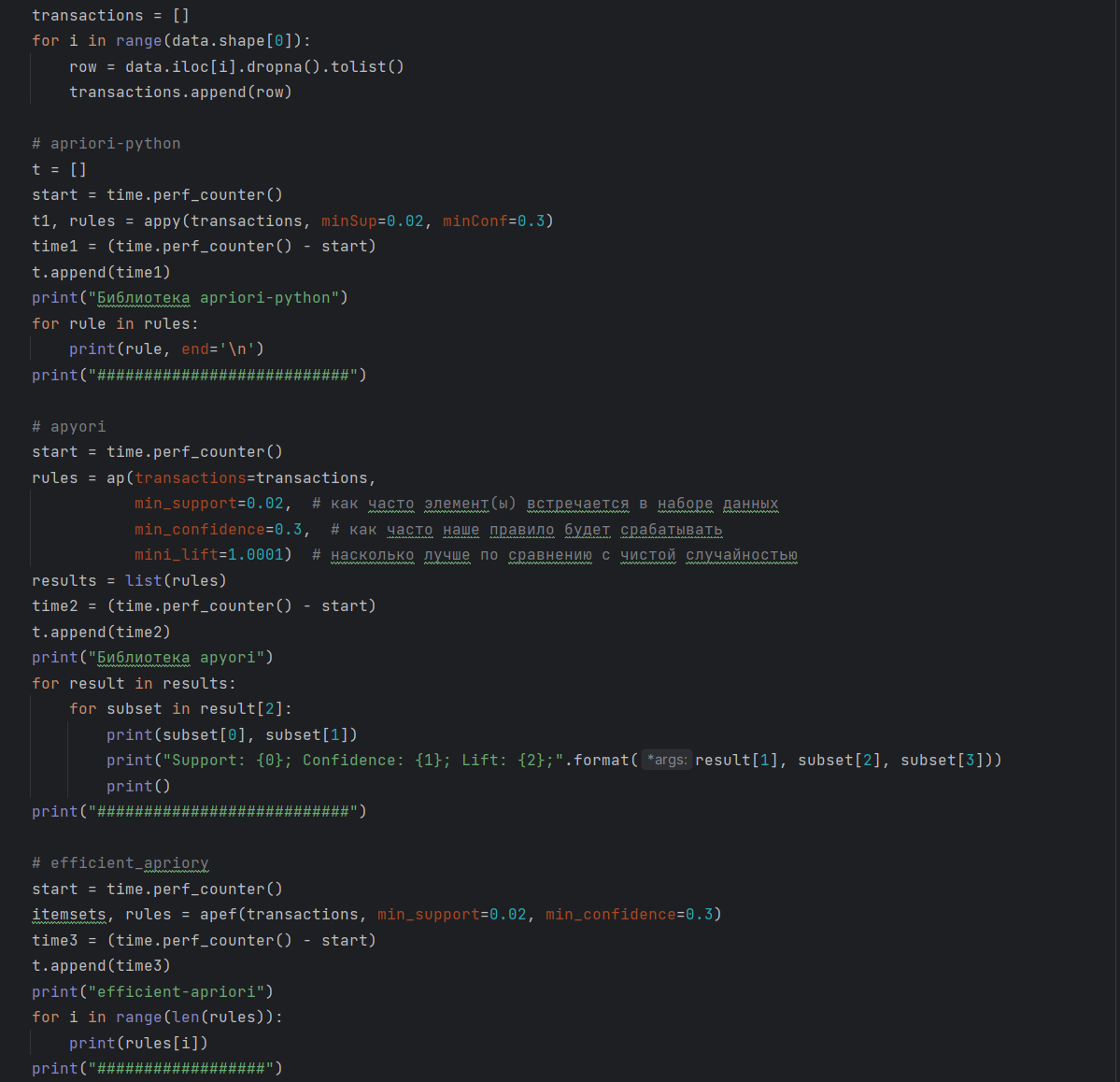


Рис. 7 – задание 3 и 8

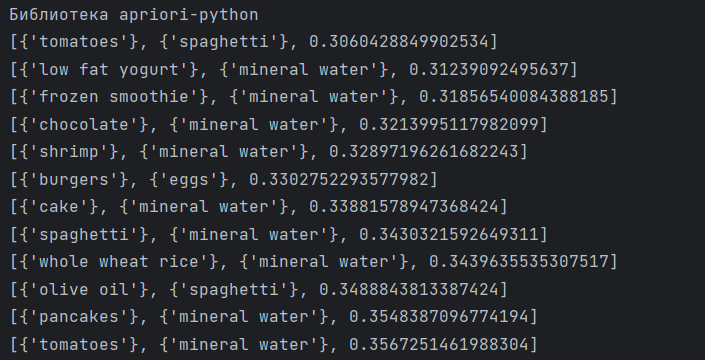


Рис. 8 - apriori\_python для Market\_Basket\_Optimisation.csv

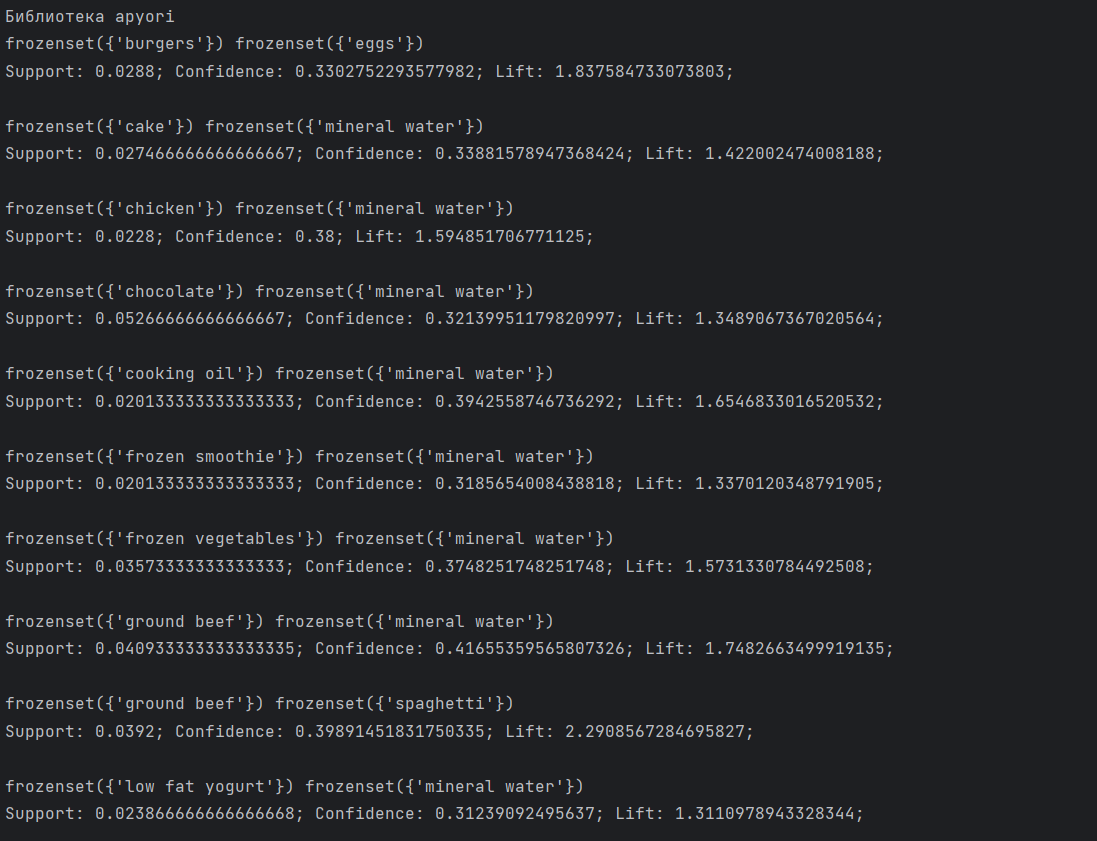


Рис. 9 – apyori для Market\_Basket\_Optimisation.csv

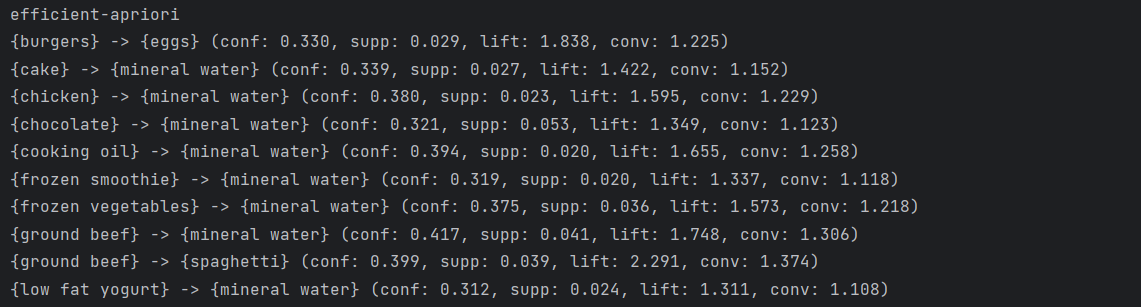


Рис. 10 - efficient\_apriori для Market\_Basket\_Optimisation.csv

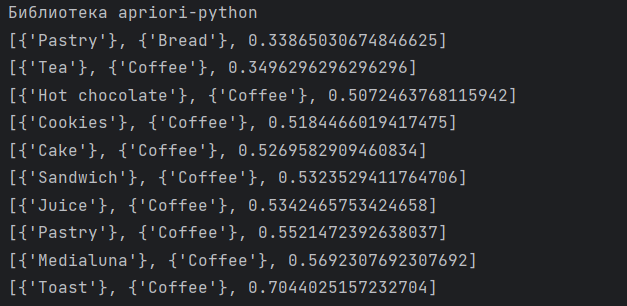


Рис. 11 - apriori\_python для data.csv

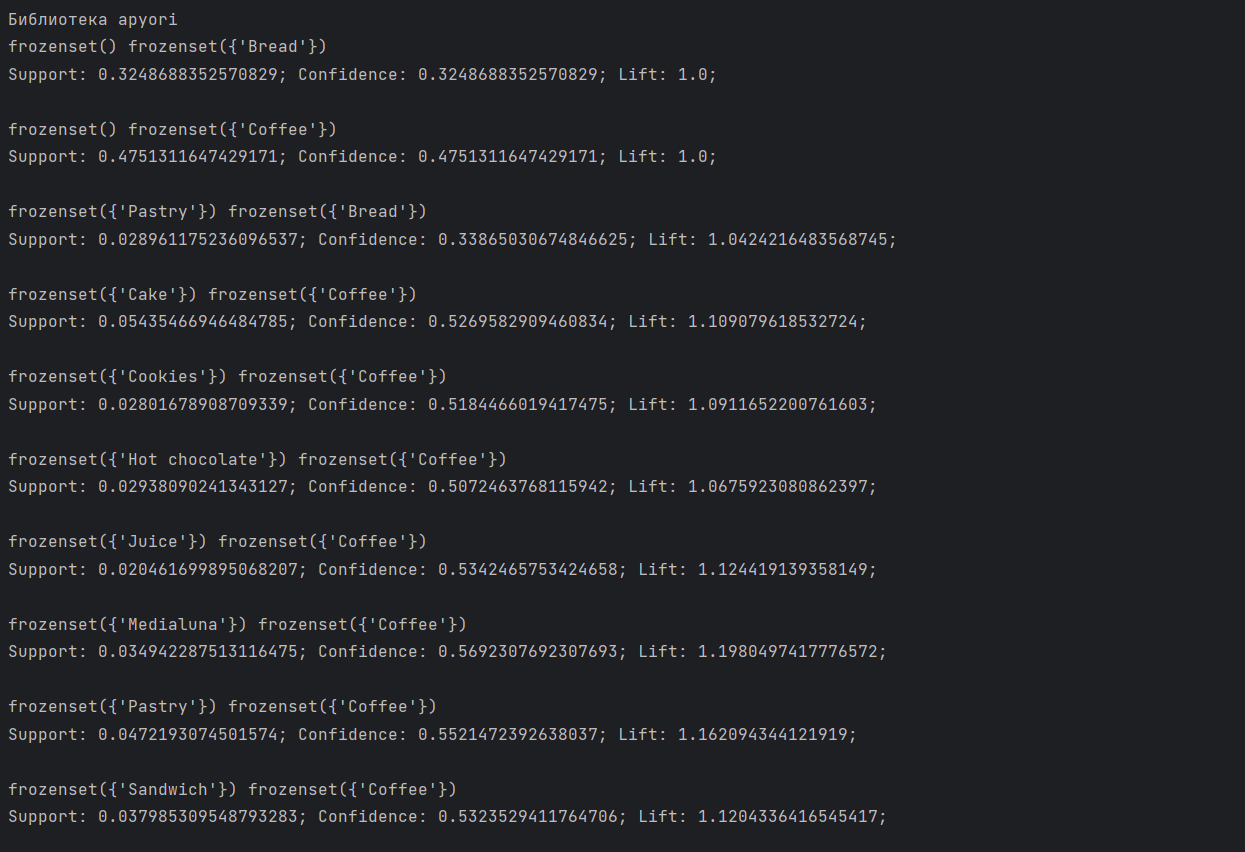


Рис. 12 - apyori для data.csv

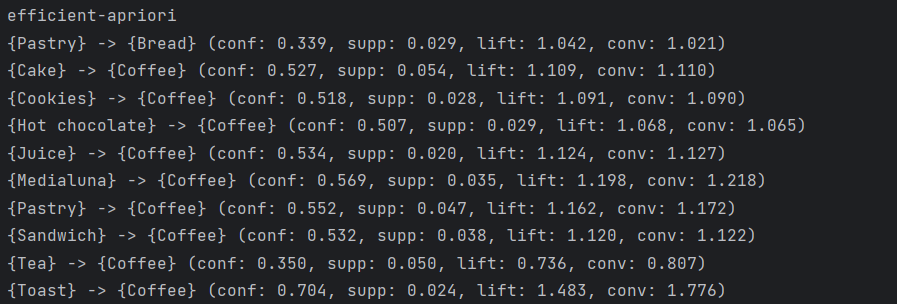


Рис. 13 - efficient\_apriori для data.csv

# Задание 4 и 9

4. Применить алгоритм FP-Growth из библиотеки fpgrowth\_py. Подобрать

гиперпараметры для алгоритма так, чтобы выводилось порядка 10

наилучших правил.

9. Применить алгоритм FP-Growth из библиотеки fpgrowth\_py. Подобрать

гиперпараметры для алгоритма так, чтобы выводилось порядка 10

наилучших правил.



Рис. 14 – Задания 4 - 9

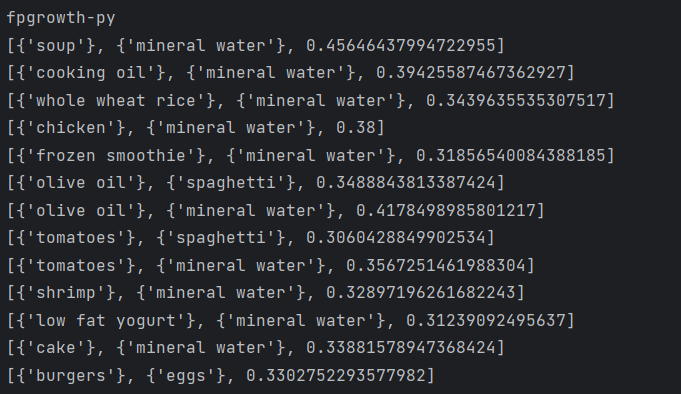


Рис. 15 - fpgrowth\_py для Market\_Basket\_Optimisation.csv

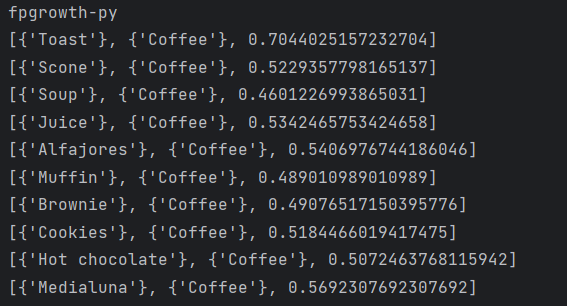


Рис. 16 - fpgrowth\_py для data.csv

# Задание 5 и 10

5. Сравнить время выполнения всех алгоритмов и построить гистограмму.

10.Сравнить время выполнения всех алгоритмов и построить гистограмму.

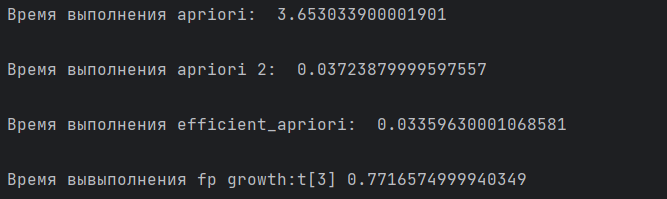


Рис. 17 – Время выполнения для Market\_Basket\_Optimisation.csv

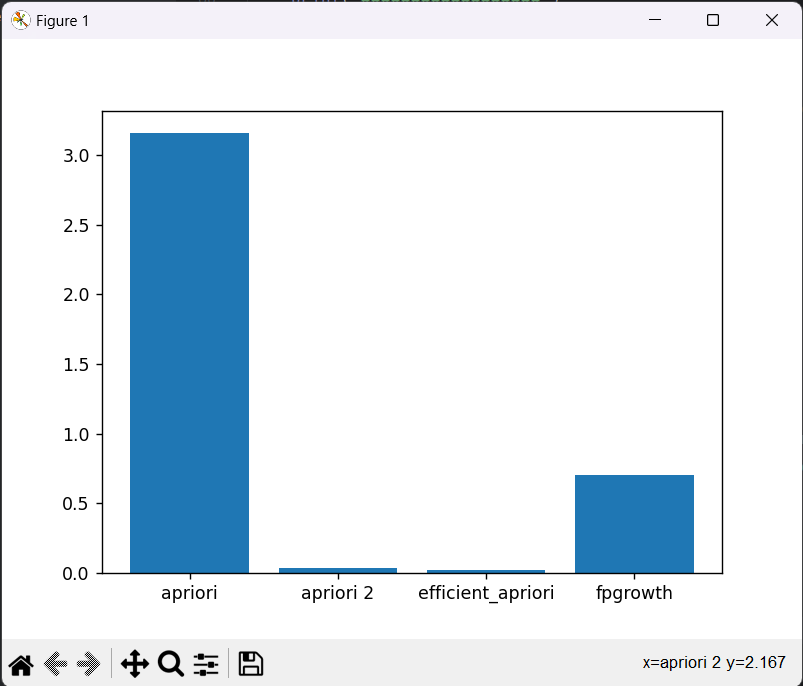


Рис. 18 – Гистограмма времени выполнения для Market\_Basket\_Optimisation.csv

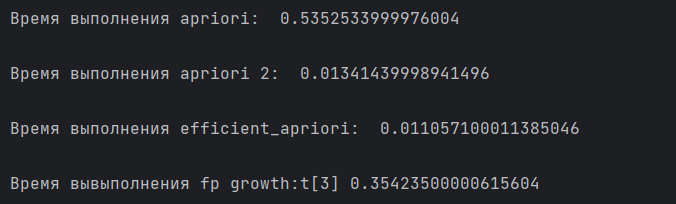


Рис. 19 – Время выполнения для data.csv

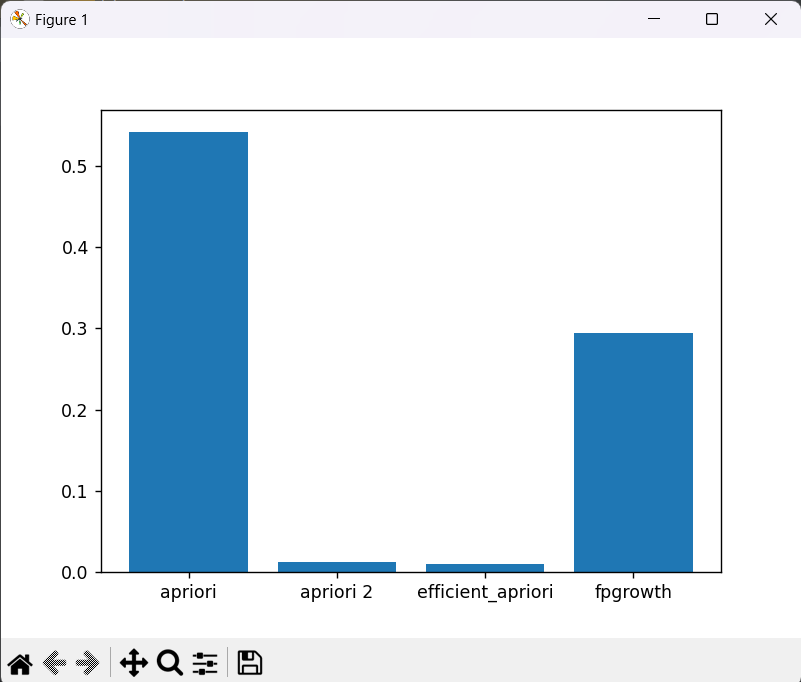


Рис. 20 – Гистограмма впемени выполнения для data.csv

**Вывод**

В ходе выполнения данной практической работы мы познакомились с обучением на ассоциативных правилах.