**МИНОБРНАУКИ РОССИИ**

**Санкт-Петербургский государственный**

**электротехнический университет**

**«ЛЭТИ» им. В.И. Ульянова (Ленина)**

**Кафедра МО ЭВМ**

отчет

**по лабораторной работе №3**

**по дисциплине «Машинное обучение»**

Тема: **Частотный анализ**

|  |  |  |
| --- | --- | --- |
| Студент гр. 8304 |  | Сергеев А.Д. |
| Преподаватель |  | Жангиров Т. Р. |

Санкт-Петербург

2021

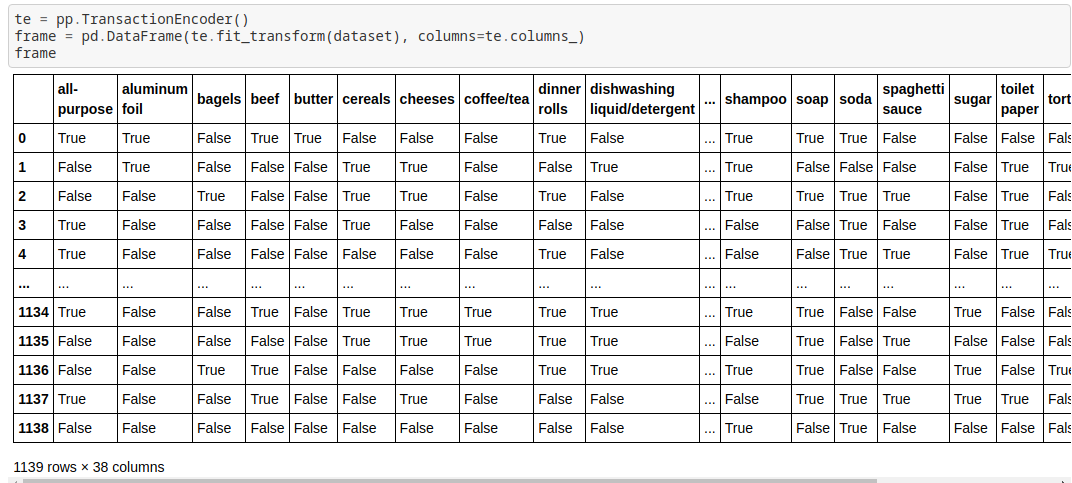
**Цель работы**

Ознакомиться с методами частотного анализа из библиотеки MLxtend.

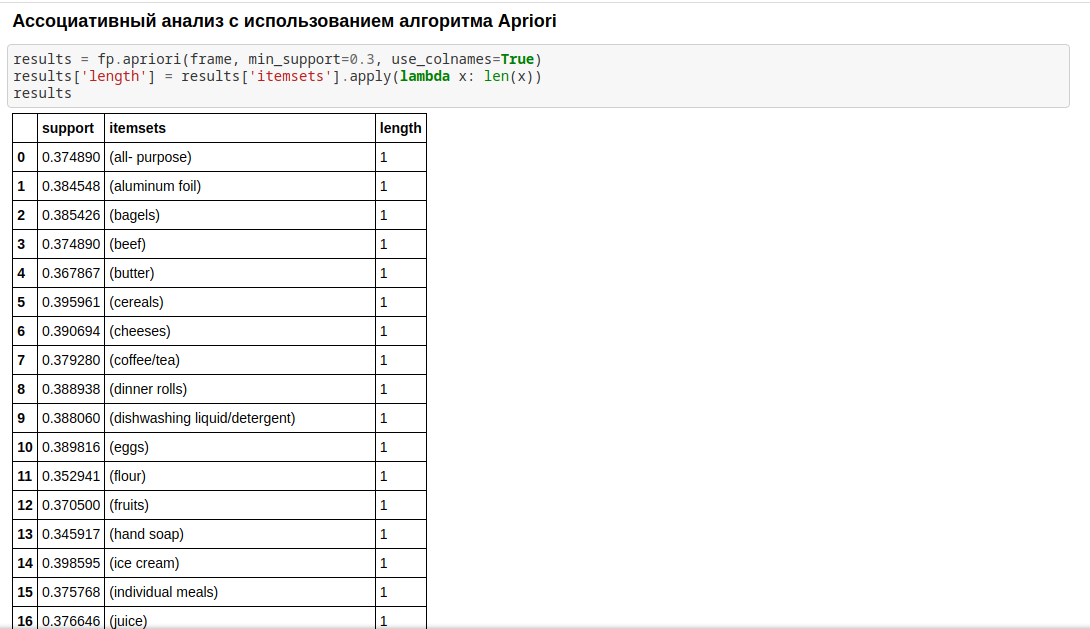
**Ход работы**

* 1. **Загрузка данных**

1. Данные загружены в датафрейм.
2. Были получены список id всех покупателей, которые есть в файле (38 различных), а также список всех товаров, которые есть в файле (1139 различных).
3. Далее был сформирован датасет, подходящий для частотного анализа. Для этого были слиты все товары одного покупателя в один список. В дальнейшем частотном анализе id покупателя будет не нужен.
4. **Подготовка данных**
5. Данные были конвертированы в матричный датасет, подходящий для использования в частотных алгоритмах при помощи TransactionEncoder.

Рисунок 1 – Полученный датасет

1. **Ассоциативный анализ с использованием алгоритма Apriori**
2. Был применен алгоритм apriori с уровнем поддержки 0.3. На основе полученных данных можно сделать вывод, что уровнем поддержки набора элементов является вероятность его встречи в транзакции. Всего получено 52 набора.

Рисунок 2 – Результаты работы алгоритма apriori с уровнем поддержки 0.3

1. Был применен алгоритм apriori с теми же параметрами, но с ограничением на максимальный размер набора, равным 1. Всего получено 38 наборов.

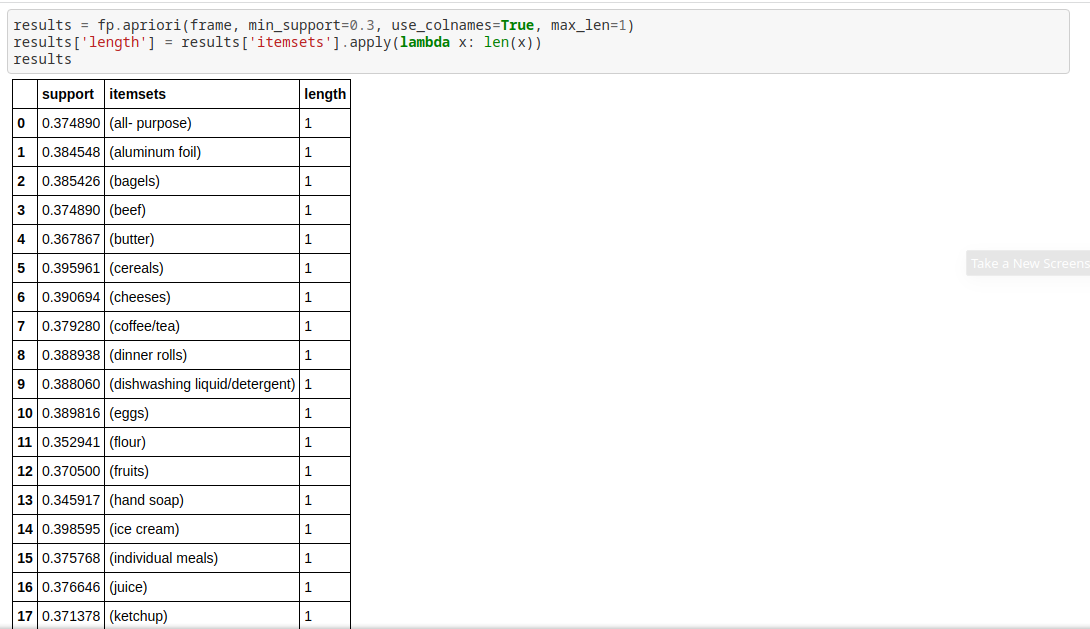


Рисунок 3 – Результаты алгоритма apriori с уровнем поддержки 0.3 для наборов размером 1

1. Был применен алгоритм apriori с теми же параметрами, но с ограничением на размер набора, равным 2, а также количество таких наборов. Всего получено 14 наборов.

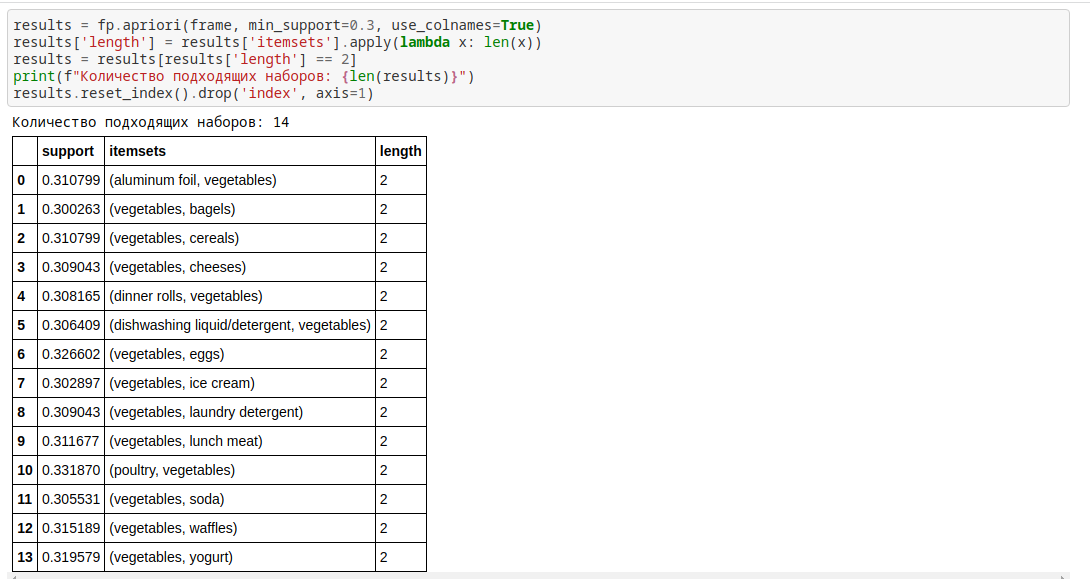


Рисунок 4 – Результаты алгоритма apriori с уровнем поддержки 0.3 для наборов размером 2

1. Было посчитано количество наборов при различных уровнях поддержки. Начальное значение поддержки 0.05, шаг 0.01. Был построен график зависимости количества наборов от уровня поддержки. Значения уровня поддержки, при котором генерируются наборы разного максимального размера юыли отмечены разными цветами.

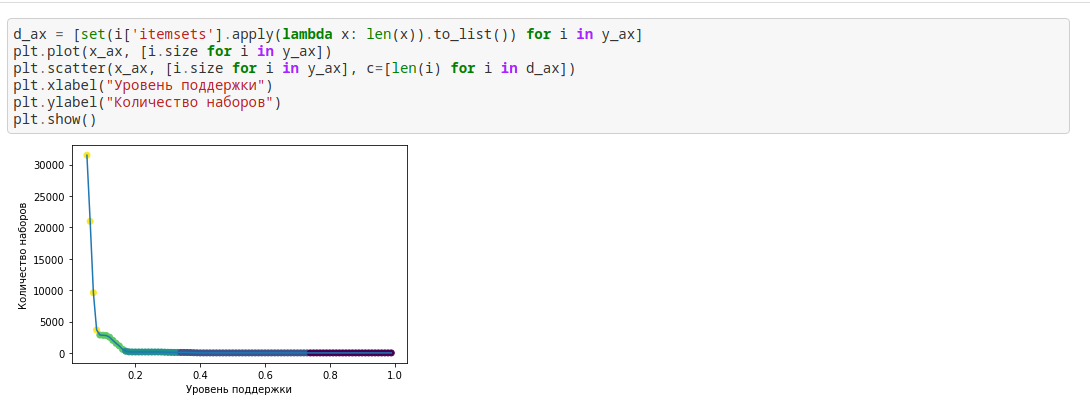
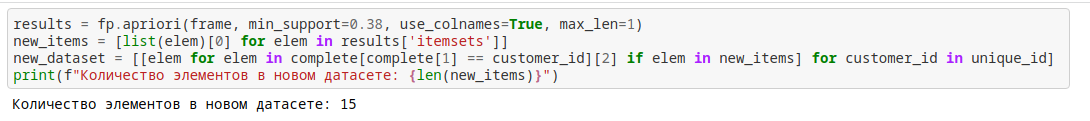


Рисунок 5 – График зависимости количества наборов от уровня поддержки

1. Был построен датасет только из тех элементов, которые при минимальном уровне поддержки 0.38 попадают в наборы размера 1. Всего в датасет попало 15 элементов.

Рисунок 6 - Полученный датасет

1. Был применен алгоритм apriori при уровне поддержки 0.3 для нового датасета. В результате представлены товары, которые аналогичны п.2, но имеют минимальный уровень поддержки 0.38. Всего получено 28 наборов.

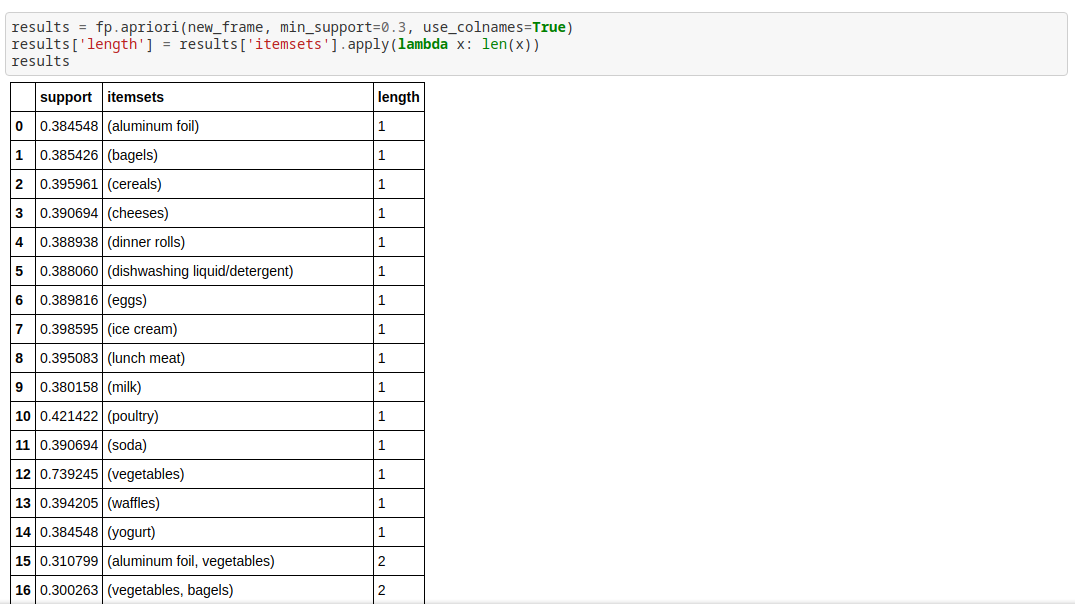
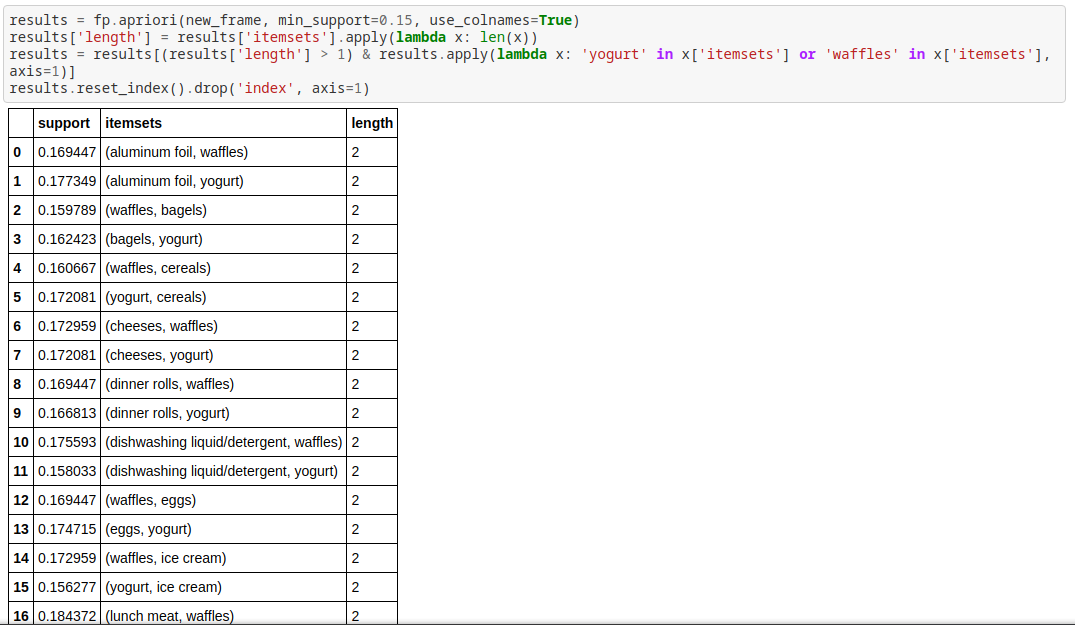
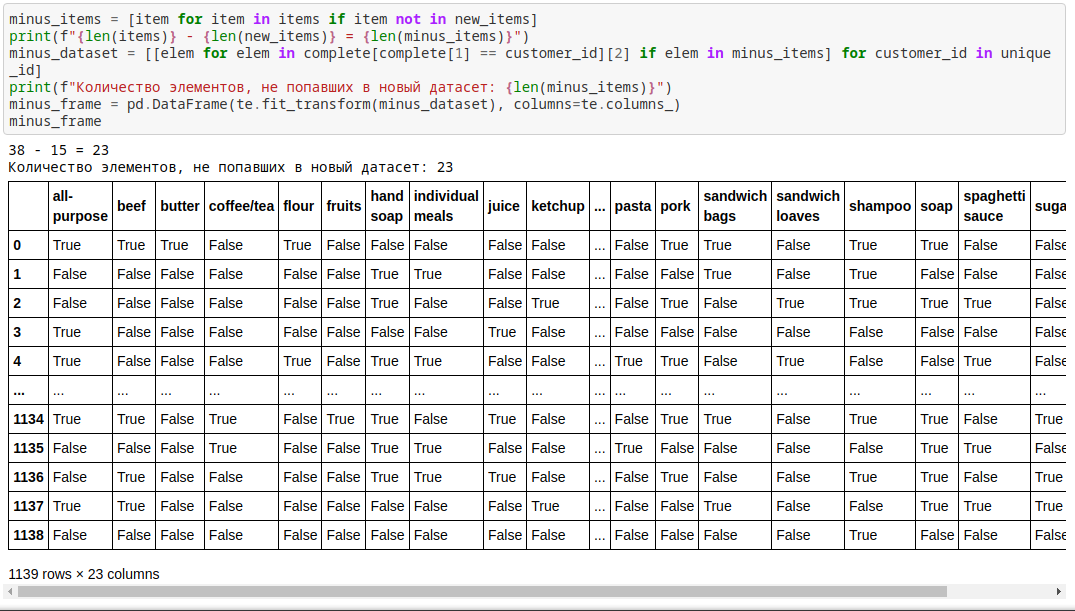


Рисунок 7 – Результаты алгоритма apriori с уровнем поддержки 0.38 для наборов размером 1

1. Был применен алгоритм apriori при уровне поддержки 0.15 для нового датасета. Выведены все наборы, размер которых больше 1 и в которых есть элементы «yogurt» или «waffles».

Рисунок 8 – Результаты алгоритма apriori с уровнем поддержки 0.15 для наборов размером больше 1, содержащих «yogurt» или «waffles»

1. Был построен датасет из тех элементов, которые не попали в датасет в п.6. Всего в датасет попало 23 элемента.

Рисунок 9 – Полученный датасет

1. Был применен алгоритм apriori с минимальным уровнем поддержки 0.3 для полученного датасета.

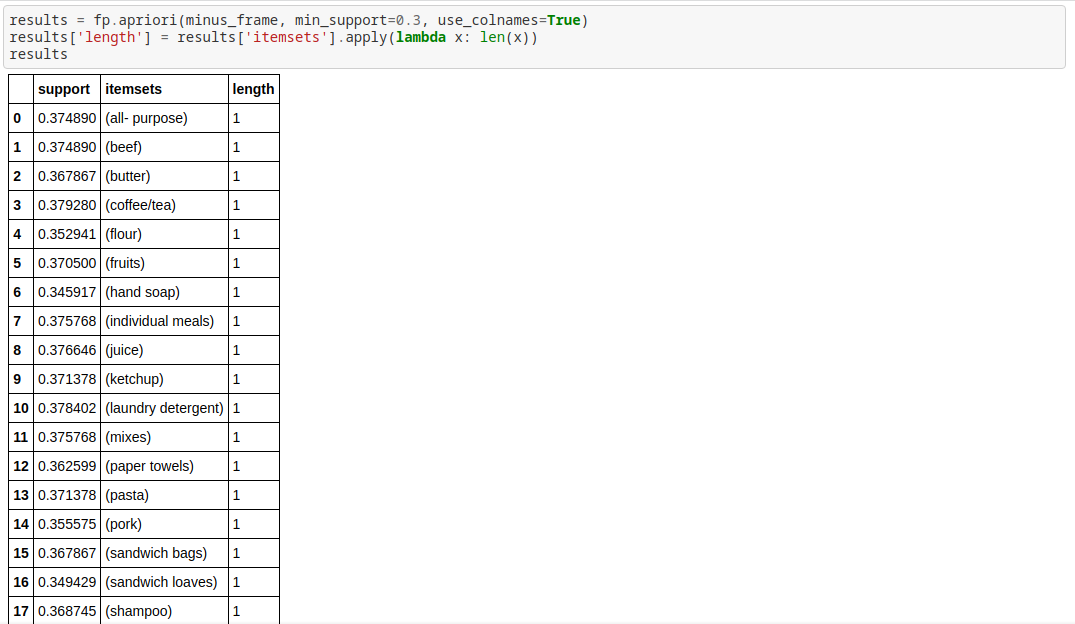


Рисунок 10 – Результаты алгоритма apriori с уровнем поддержки 0.3 для полученного датасета

1. Было написано правило для вывода всех наборов, в которых хотя бы два элемента начинаются на «s».

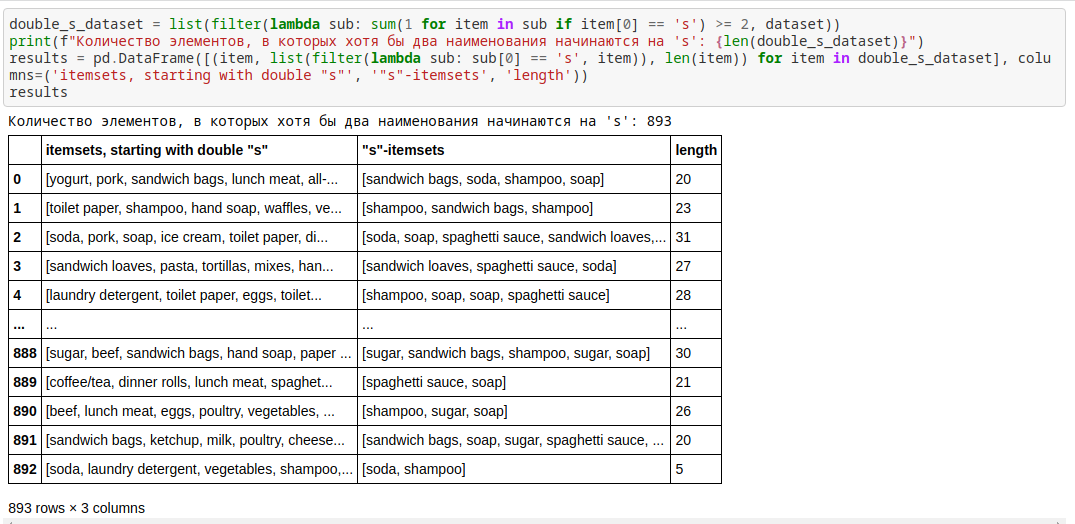


Рисунок 11 – Наборы, в которых хотя бы два элемента начинаются на «s»

1. Было написано правило для вывода всех наборов, для которых уровень поддержки изменяется от 0.1 до 0.25.

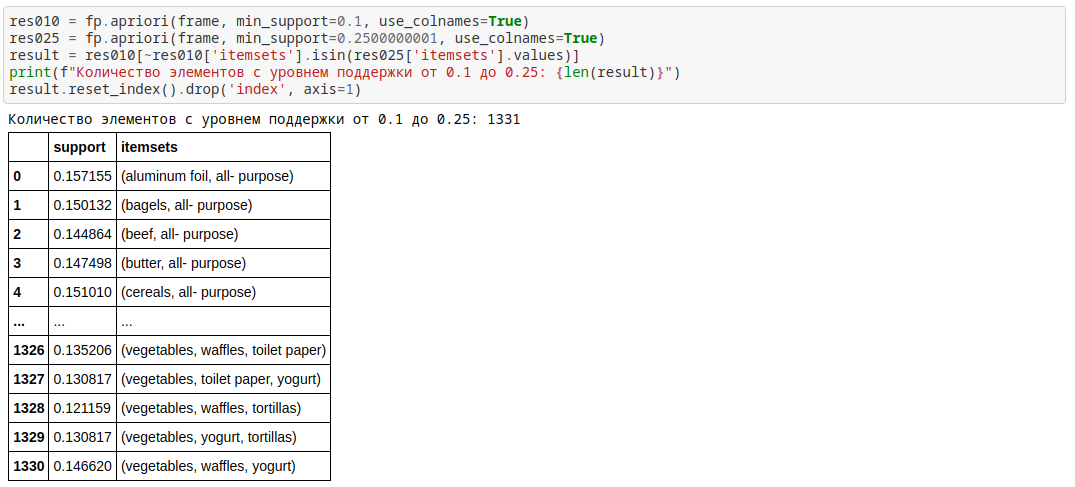


Рисунок 12 – Наборы, для которых уровень поддержки изменяется от 0.1 до 0.25

**Выводы**

В ходе выполнения данной лабораторной работы было выполнено ознакомление с методами частотного анализа из библиотеки MLxtend.

На основе данной лабораторной работы был изучен алгоритм Apriori, а также было установлено, что увеличение минимального уровня поддержки ведет к уменьшению максимальной длины подходящих наборов.