**Технологии аналитической обработки информации**

**Задание 2 (Поиск шаблонов).**

1. Доработайте программу из задания 1, чтобы она также выполняла поиск ассоциативных правил. Список результирующих правил должен выдаваться в удобочитаемом виде (антецедент→консеквент) с указанием поддержки и достоверности каждого правила. Дополнительные параметры программы: порог достоверности, способ упорядочивания результирующего списка наборов (по убыванию значения поддержки или лексикографическое).
2. Проведите эксперименты на наборах из задания 1. В экспериментах Зафиксируйте значение пороговое значение поддержки (например, 10%), варьируйте пороговое значение достоверности (например, от 70% до 95% с шагом 5%).
3. Выполните визуализацию полученных результатов в виде следующих диаграмм:
   * + сравнение быстродействия поиска правил на фиксированном наборе данных при изменяемом пороге достоверности;
     + общее количество найденных правил на фиксированном наборе данных при изменяемом пороге достоверности.
4. Подготовьте список правил, в которых антецедент и консеквент суммарно включают в себя не более семи объектов (разумное количество). Проанализируйте и изложите содержательный смысл полученного результата.
5. Подготовьте отчет о выполнении задания и загрузите отчет в формате PDF в систему. Отчет должен представлять собой связный и структурированный документ со следующими разделами:
   * + формулировка задания;
     + гиперссылка на каталог репозитория с исходными текстами, наборами данных и др. сопутствующими материалами;
     + рисунки с результатами визуализации.

**Результат работы.**

На рисунке 1 представлена реализация метода для нахождения частых наборов и ассоциативных правил с двумя разными сортировками.

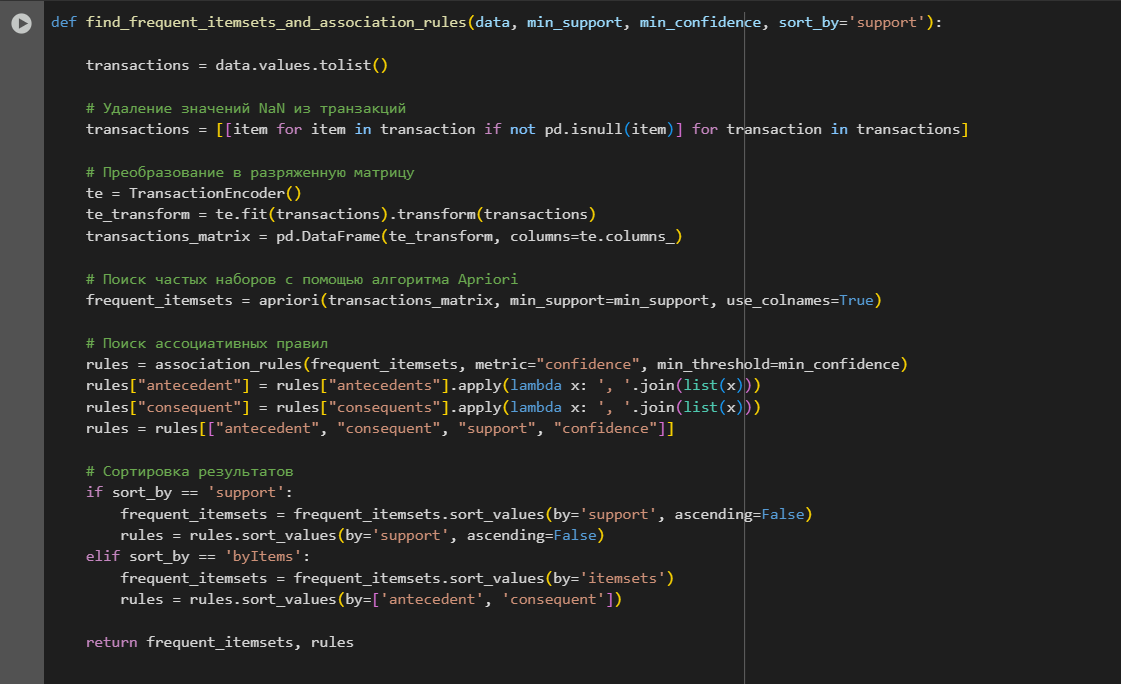
****

Рисунок 1 – Метод find\_frequent\_itemsets\_and\_association\_rules

На рисунке 2 представлен результат вызова метода.

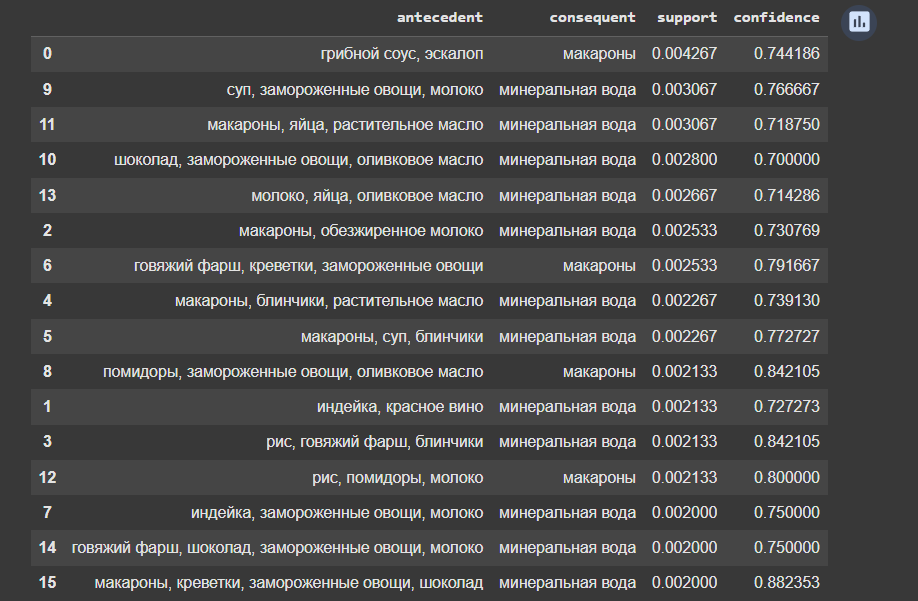
****

Рисунок 2 – Вызов метода с минимальной поддержкой 0.002 и достоверностью 0.7

На рисунке 3 представлена реализация метода для отображения диаграмм.

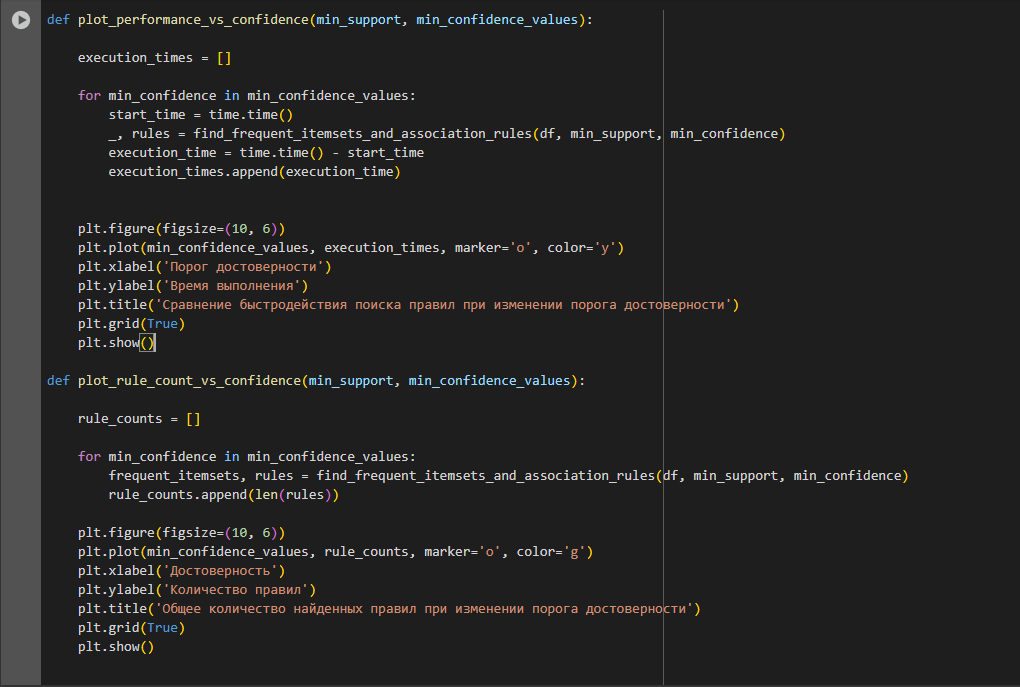
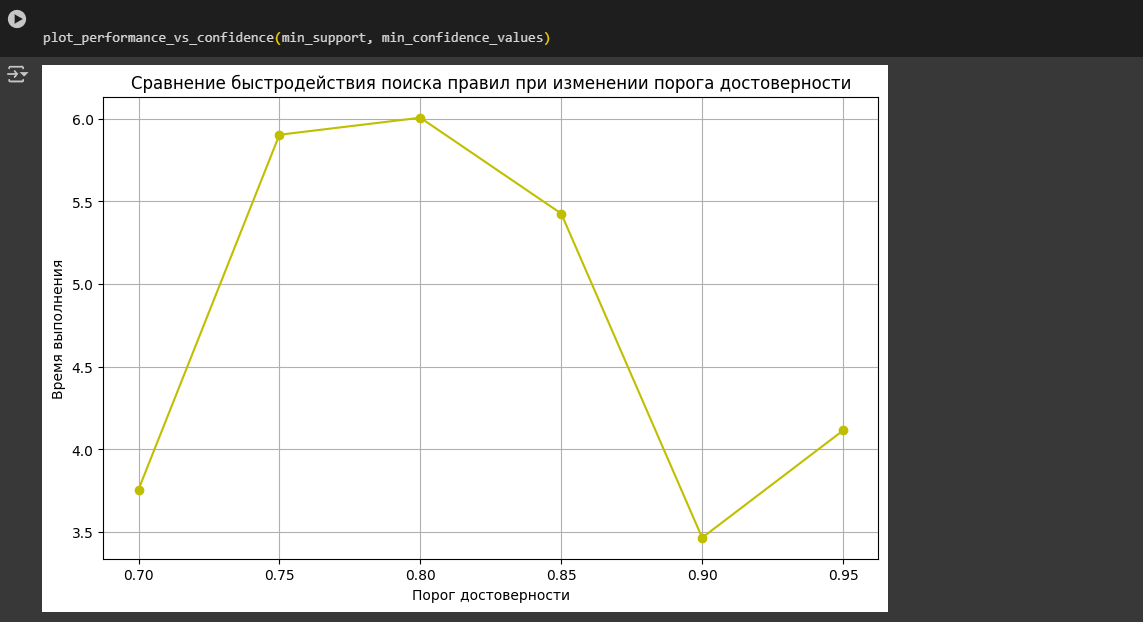


Рисунок 3 – Методы plot\_performance\_vs\_confidence и plot\_rule\_count\_vs\_confidence

Была выполнена визуализация полученных результатов в виде следующих диаграмм:

* + - сравнение быстродействия поиска правил на фиксированном наборе данных при изменяемом пороге достоверности;
    - общее количество найденных правил на фиксированном наборе данных при изменяемом пороге достоверности.

На рисунке 4 представлена диаграмма зависимости достоверности от времени.

Рисунок 4 – Диаграмма зависимости достоверности от времени

На рисунке 5 представлена диаграмма зависимости достоверности от количества правил.

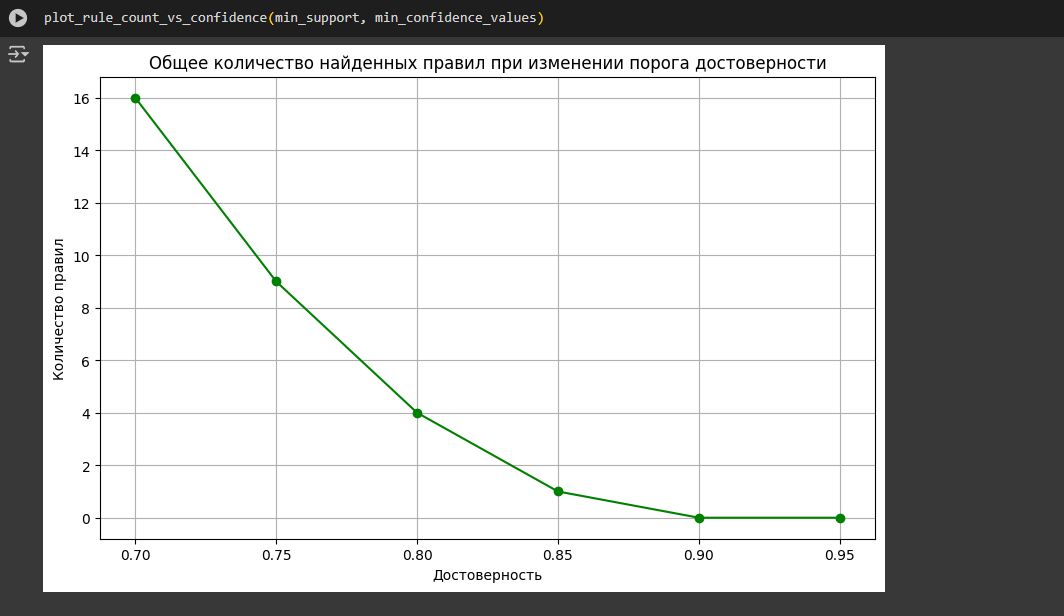


Рисунок 5 – Диаграмма зависимости достоверности от количества правил

На рисунке 6 представлен код для вывода правил, антецедент и консеквент которых включают в себя не более семи объектов.

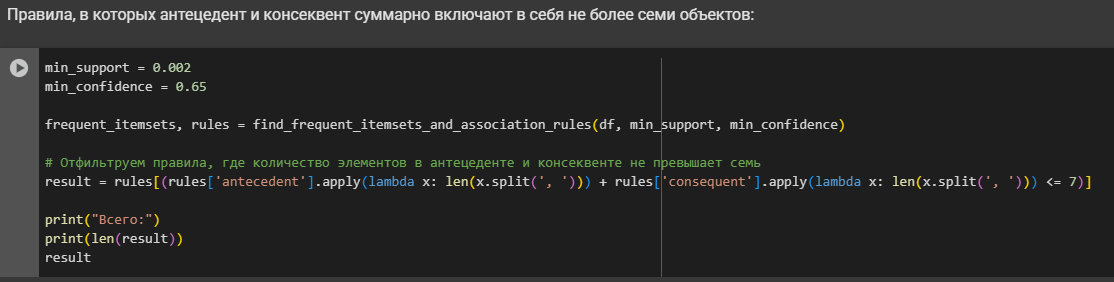


Рисунок 6 – Код для задания 4

На рисунке 7 представлен результат кода.

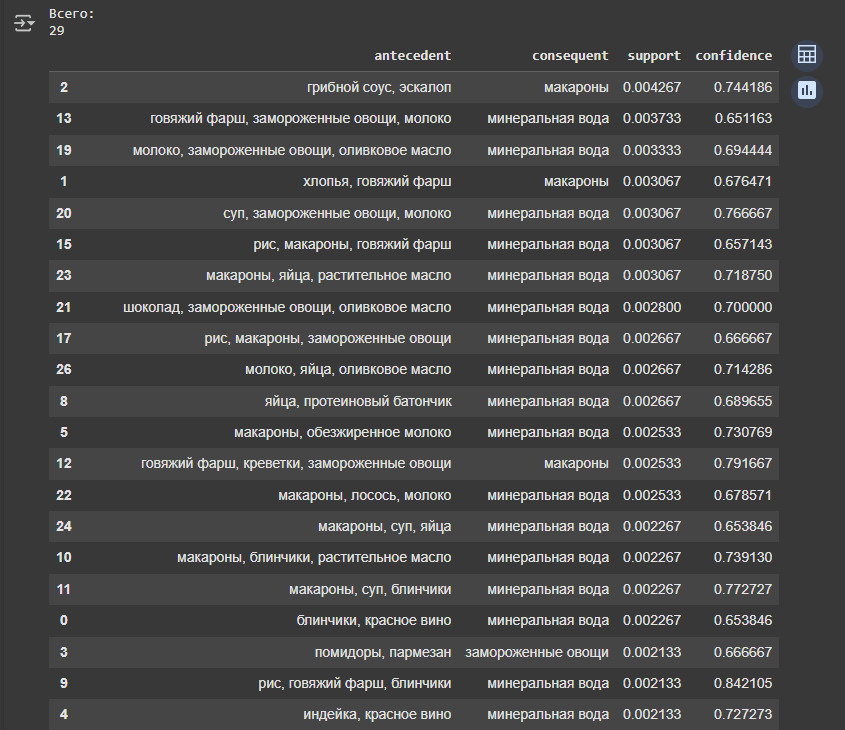


Рисунок 7 – Вывод правил, антецедент и консеквент которых включают в себя не более семи объектов.