Assignment1 report - Data Science

Instructor : 채동규
Student name / ID : 손승일 / 2018008504

Apriori algorithm

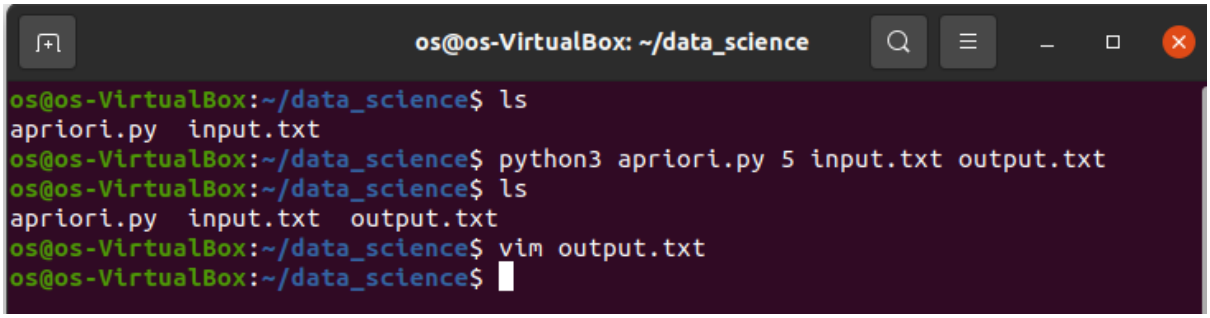
1. Environment

Ubuntu : 20.04.5 LTS

Python : 3.8.10

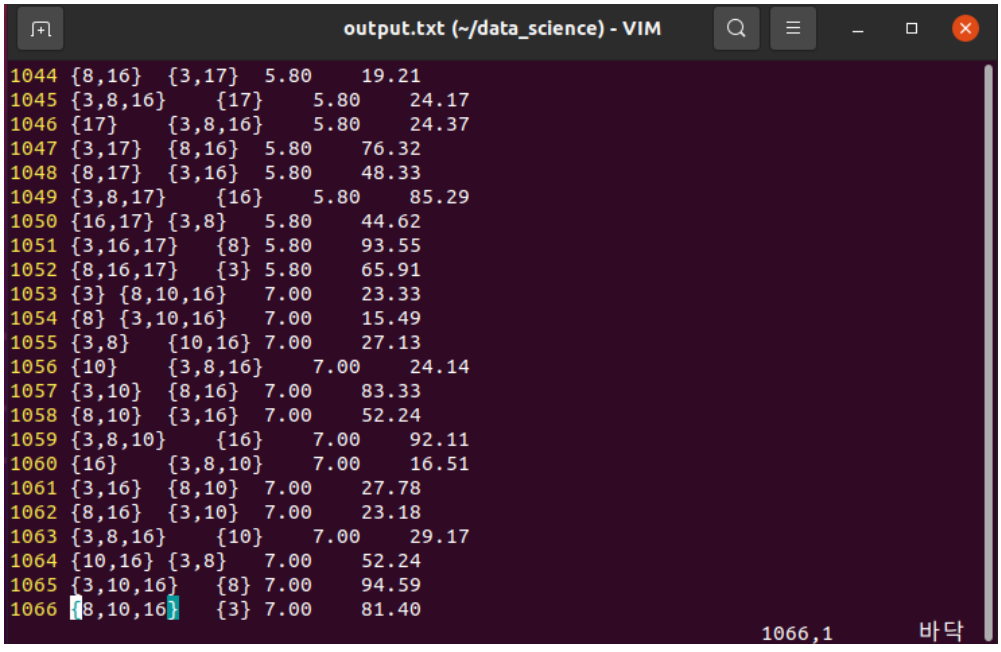
(vscode로 짜다가 터미널에서 input file을 못읽어서, Ubuntu_linux 환경으로 옮겨보니 작동됨)

2. How to run this code



Python3	apriori.py	5	input.txt	output.txt
	File name	minimum support	input file	output file

3. Result



Minimum support를 5로 잡고 올려주신 input.txt 파일을 입력받고 실행하면 1066개의 결과값이 output 파일에 저장됨.

4. Specifics of the code

1. def read_file : Read input file

2. def generate_first_candidate : Generate 1- candidate item set from the db

3. def generate_candidate : Generate K-candidate item set. Make K-candidate item set from (k-1) frequent item set

4. def prune : Check if the candidate's support matches the minimum support and if it doesn't, eliminate the candidate and if it does, insert it to k-candidate item set.

5. def get_sup_cnt : calculate the candidate's support count

6. def generate_frequent : make K-frequent item set from k-candidate item set

7. def write_file : write the frequent item sets at the output file

8. main : run till the candidate itemset and frequent item set doesn't come out. If the While true ends, function write_file will be executed and the code will end.