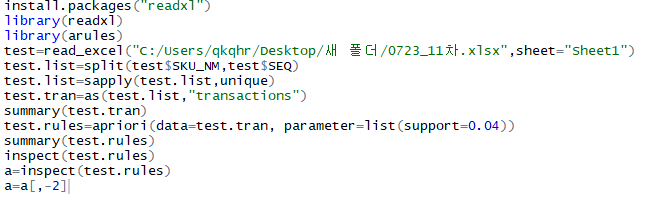
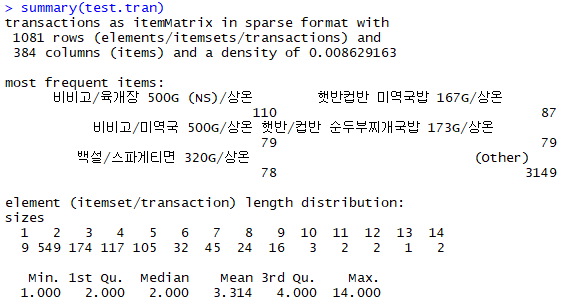
주문 상품에 대한 유사도를 분석하기 위해 R의 arules 라이브러리를 이용하여 장바구니 분석을 실시하였다. 알고리즘을 테스트하기 위해 우선 각 월별 주문 상품 수 상위 5개를 sampling 하였다. 그 결과, 아래와 같이 6월과 7월 각각 상위 5개의 샘플이 추출되었다.

|  |  |  |
| --- | --- | --- |
| 일자 및 차수 | 주문 상품 수 | SKU 수 |
| 6월 23일 6차 | 4296 | 352 |
| 6월 13일 6차 | 3594 | 384 |
| 6월 28일 5차 | 3498 | 317 |
| 6월 15일 5차 | 3396 | 350 |
| 6월 29일 9차 | 3124 | 336 |
| 7월 20일 9차 | 5028 | 390 |
| 7월 6일 5차 | 4282 | 262 |
| 7월 23일 11차 | 4267 | 368 |
| 7월 19일 7차 | 4175 | 392 |
| 7월 26일 7차 | 3325 | 343 |

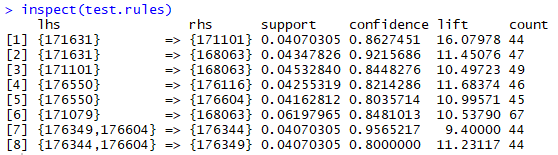
장바구니 분석을 위한 코드는 아래와 같다.



<6월 13일 – 6차>



6월 13일 6차수 데이터에 대한 분석을 보면 가장 많이 팔린 상품 TOP 5개는 ‘비비고/육개장 500G (NS)/상온’, ‘햇반컵반 미역국밥 167G/상온’, ‘햇반/컵반 순두부찌개국밥 173G/상온’, ‘비비고/미역국 500G/상온’, ‘백설/스파게티면 320G/상온’ 이다. 또한, 한 상품에 대해 한번에 주문한 개수는 최대 14개이다.

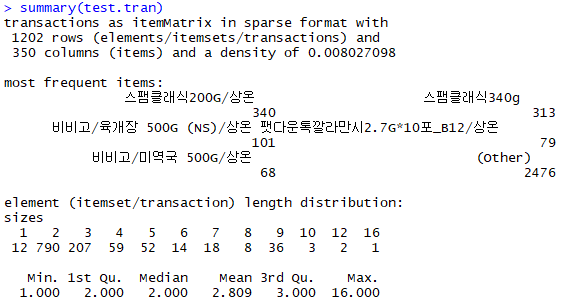


여러 개의 조합 중에서 지지도가 0.04 이상인 조합들을 추출하였는데 총 8개가 추출되었다. 지지도는 좋은 규칙을 찾거나, 불필요한 연산을 줄일 때의 기준으로 사용한다.



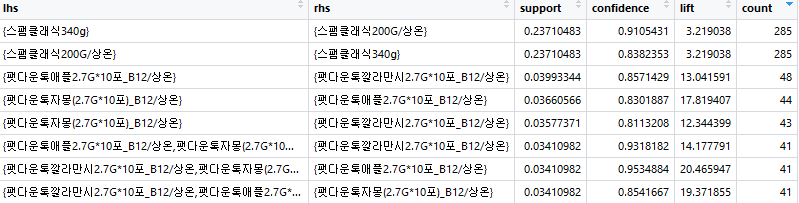
가장 많은 횟수의 조합은 ‘햇반/컵반 순두부찌개국밥 173G/상온’과 ‘햇반컵반 미역국밥 167G/상온’의 조합으로 총 67번이다.

<6월 15일 – 5차>



6월 15일 5차에 대한 데이터를 분석해 보았을 때 가장 많이 팔린 상위 5개 제품은 스팸클래식200G/상온, 스팸클래식340g, 스팸클래식340g, 팻다운톡깔라만시2.7G\*10포\_B12/상온, 비비고/미역국 500G/상온이다.

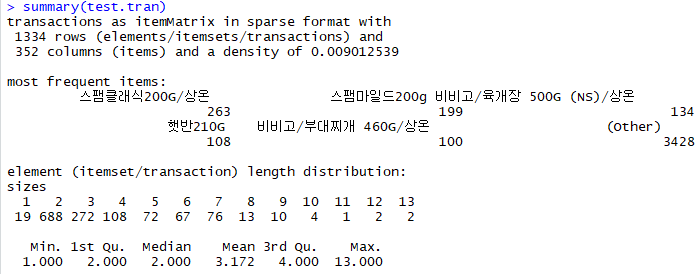
이 차수의 데이터에서 조합 분석을 실시할 때 지지도를 0.04로 기준을 잡으니 압도적으로 횟수가 많은 두 개의 조합이 추출되었다. 개수를 늘리기 위해 지지도를 0.03으로 낮추니 8개의 조합이 추출되었다.



가장 많은 상품의 조합은 ‘스팸클래식340g’, ‘스팸클래식200G/상온’으로 총 285개의 주문이 있었다.

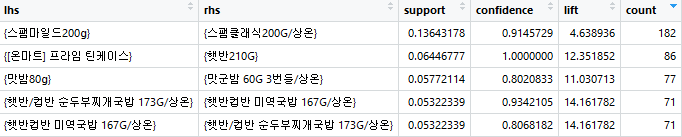
이번 차수에서는 ‘스팸클래식340g’, ‘스팸클래식200G/상온’, ‘팻다운톡애플2.7G\*10포\_B12/상온’, ‘팻다운톡깔라만시2.7G\*10포\_B12/상온’, ‘팻다운톡자몽(2.7G\*10포)\_B12/상온’ 이 5개의 제품에 대한 조합이 대부분이었다.

<6월 23일 - 6차>



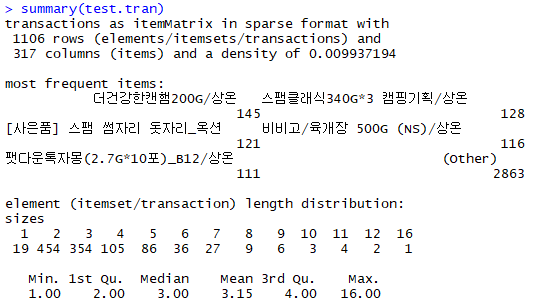
6월 23일 6차에 대한 데이터를 분석해 보았을 때 가장 많이 팔린 상위 5개 제품은 스팸클래식200G/상온, 스팸마일드200g, 비비고/육개장 500G (NS)/상온, 햇반210G, 비비고/부대찌개 460G/상온이다.

이번 차수에서는 지지도를 낮게 설정했더니 너무 많은 조합이 추출되어 지지도를 0.05로 높여 5개의 조합을 추출하였다

. 

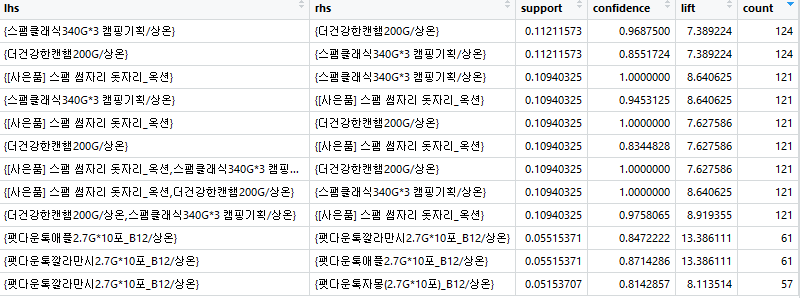
이 차수에서 가장 많은 조합은 ‘스팸마일드200g’와 ‘스팸클래식200G/상온’로 총 182번의 조합이 있었다.

<6월 28일 – 5차>



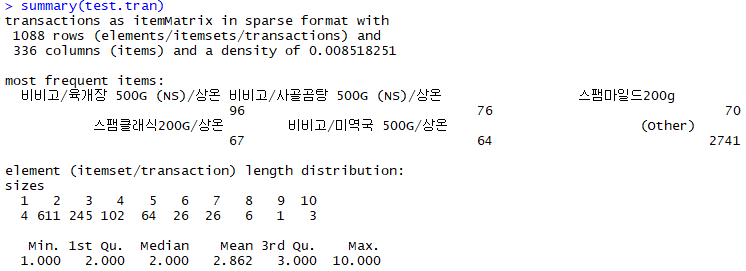
6월 28일 5차수에 가장 많이 팔린 상위 5개는 더건강한캔햄200G/상온, 스팸클래식340G\*3 캠핑기획/상온, [사은품] 스팸 썸자리 돗자리\_옥션, 비비고/육개장 500G (NS)/상온, 팻다운톡자몽(2.7G\*10포)\_B12/상온 이다.

이번 차수에서는 지지도를 0.05로 설정했는데 동일 상품에 대한 여러 조합이 대부분이었다.



이 차수의 상품 조합에서는 ‘스팸클래식340G\*3 캠핑기획/상온’, ‘더건강한캔햄200G/상온’, ‘[사은품] 스팸 썸자리 돗자리\_옥션’, 이 세가지 상품의 조합이 가장 많았다.

<6월 29일 - 9차>



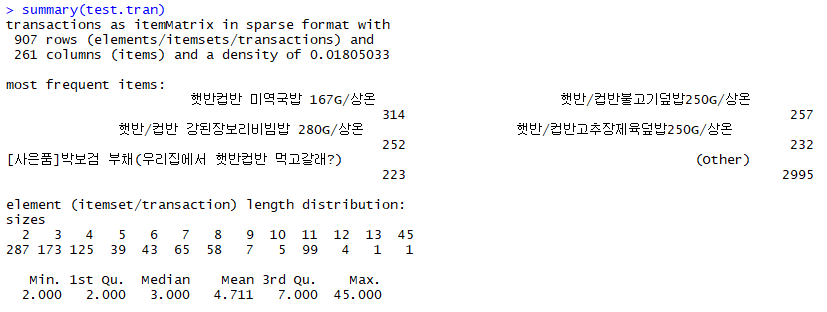
6월 29일 9차수에 가장 많이 팔린 상위 5개는 비비고/육개장 500G (NS)/상온, 비비고/사골곰탕 500G (NS)/상온, 스팸마일드200g, 스팸클래식200G/상온, 비비고/미역국 500G/상온이다.

이 차수에서는 상품 조합의 지지도 기준을 0.023으로 설정했고 8개의 조합이 추출되었다.



가장 많은 횟수의 조합은 ‘햇반/컵반 사골곰탕국밥 166G/상온‘, ‘햇반컵반 미역국밥 167G/상온‘이고, 이번 차수에서는 8개의 조합이 비슷한 횟수로 추출되었다.

<7월 6일 – 5차>



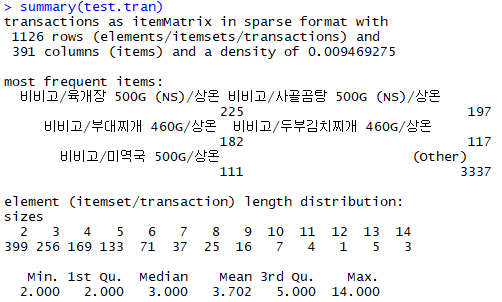
7월 6일 5차에서 가장 많이 팔린 상위 5개 상품은 ‘햇반컵반 미역국밥 167G/상온’, ‘햇반/컵반불고기덮밥 250G/상온’, ‘햇반/컵반 강된장보리비빔밥 280G/상온’, ‘햇반/컵반고추장제육덮밥250G/상온’, ‘[사은품]박보검 부채(우리집에서 햇반컵반 먹고갈래?)’이다.

이번 차수에서는 지지도를 0.25 이상으로 기준을 잡았다. 총 7개의 조합이 추출되었다.



가장 많은 조합은 ‘햇반/컵반 강된장보리비빔밥 280G/상온’, ‘햇반컵반 미역국밥 167G/상온’으로 총 241번의 조합이 있었다.

<7월 19일 - 7차>



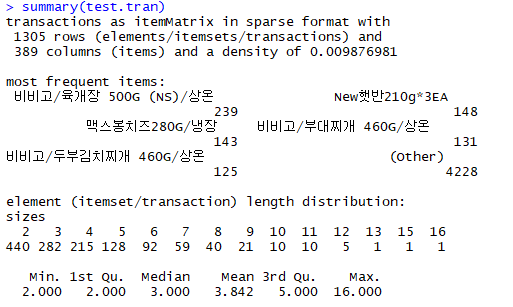
가장 많이 팔린 상위 다섯개의 상품은 ‘비비고/육개장 500G (NS)/상온’, ‘비비고/사골곰탕 500G (NS)/상온’, ‘비비고/부대찌개 460G/상온’, ‘비비고/두부김치찌개 460G/상온’, ‘비비고/미역국 500G/상온’이다.

이번 차수에서는 지지도 기준을 0.035로 세웠고, 총 14개의 조합이 추출되었다.



이 차수에서 가장 많은 조합은 ‘크림토마토파스타소스455G/상온’, ‘백설/스파게티면 320G/상온’으로 총 47번의 조합이 있었다.

<7월 20일 – 9차>



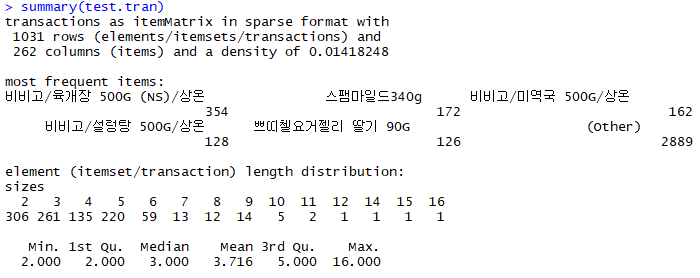
7월 20일 9차 주문에서 가장 많이 팔린 상위 5개 상품은 ‘비비고/육개장 500G (NS)/상온’, ‘New햇반210g\*3EA’, ‘맥스봉치즈280G/냉장’, ‘비비고/부대찌개 460G/상온’, ‘비비고/두부김치찌개 460G/상온’이다.

이번 차수의 지지도는 0.04를 기준으로 잡았고, 총 15개의 조합이 추출되었다.



가장 많은 조합은 ‘쁘띠첼워터젤리사과130ML’, ‘쁘띠첼워터젤리복숭아 130ML’로 총 74번의 조합이 있었다.

<7월 23일 – 11차>



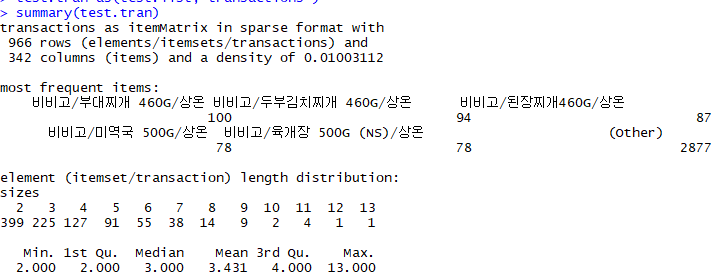
7월 23일 11차 주문에서 가장 많이 판매된 상위 5개 제품은 ‘비비고/육개장 500G (NS)/상온’, ‘스팸마일드340g’, ‘비비고/미역국 500G/상온’, ‘비비고/설렁탕 500G/상온’, ‘쁘띠첼요거젤리 딸기 90G’이다.

이번 차수에서는 지지도 기준을 0.095로 지정했고, 16개의 조합이 추출되었다.



가장 많은 횟수의 조합은 ‘쁘띠첼요거젤리 밀감 90G’, ‘쁘띠첼요거젤리 복숭아 90G’으로 총 118번의 조합이 있었다.

<7월 26일 – 7차>



7월 26일 7차 주문 데이터에서 가장 많이 판매된 상품 상위 5개는 ‘비비고/부대찌개 460G/상온’, ‘비비고/두부김치찌개 460G/상온’, ‘비비고/미역국 500G/상온’, ‘비비고/된장찌개460G/상온’, ‘비비고/육개장 500G (NS)/상온’이다.

이번 차수에서는 지지도를 0.043으로 기준을 정했고, 총 12개의 조합이 추출되었다.



가장 많은 조합은 ‘비비고/된장찌개460G/상온’, ‘비비고/두부김치찌개 460G/상온’로 총 72번의 조합이 있었다.

* 각 차수 별 데이터에서 적당한 개수의 조합을 얻기 위한 지지도는 각각 달랐다. 각 조합들의 횟수가 많으면 지지도를 높게 설정해도 적정량의 조합이 추출되지만 조합들의 횟수가 적으면 지지도를 낮게 설정해야 조합이 추출되어서 이를 위한 기준이 모호하다. 또한, 코드를 실행해서 조합을 추출했을 때, 똑같은 조합이 순서만 다르게 똑같이 추출되는 경우가 대부분이다. 이에 대한 원인을 찾고 제대로 조합을 추출할 필요가 있다. 코드를 수정하고 다시 조합을 찾는다면 좀 더 정확한 조합을 더 많이 추출할 수 있을 것으로 보인다.