

R 프로그래밍

(13주차)

2016. 05. 28(토)

장운호

(ADP 002-0004)

목차

I. 장바구니 데이터 소개

II. reshape2 패키지

III. RFM 분석 실습



I. 장바구니 데이터 소개

1. 장바구니 (Basket)

PARIS CROISSANT

서울역 파리크라상
106-85-18197 Tel:023127740
서울 용산 동자 43-205번지
POS:020 2016-05-20 06:20:56

ITEM	QTY	D/C	AMT
타르타르햄에그PC	1	0	7,000
그릴드토마토모짜렐라	1	0	7,000
(PET)생오렌지주스	1	0	6,800
과세금액			18,907
부가세			1,893
총매출액			20,800
합계			20,800
받은 돈			20,800
[카 드]			20,800

* 케익 및 냉장제품은 교환 및 환불이 불가합니다. 양해부탁드립니다.

<< 하나기업카드 승인 >>

[카드번호]
[승인금액]
[승인번호]

[고객용]

고객님~ 만족도 설문에 참여해 주세요♡
매 월 1,000분께 해피콘 3,000원 쏘 드립니다

위치

선호 상품
카테고리

지불수단 및
금액

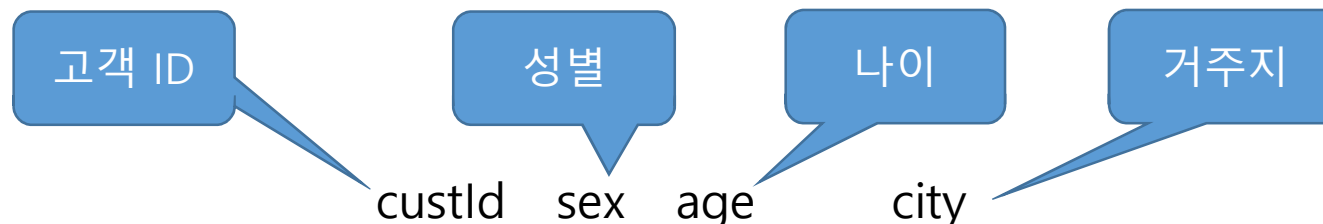
구매정보 외에
마일리지 포인트,
포인트 활용 여부
등도 활용 가능

시간

2. 구매 DB

	고객 ID	영수증번호	지점 ID	상품 대분류	상품 중분류				
	custId	receiptNum	branchId	date	time	category	division	purchase Count	amount
1	C0322	R00001	branch_02	20120405	19	beverages	hard drinks	2	4500
2	C0322	R00001	branch_02	20120405	19	beverages	hard drinks	2	4500
3	C0322	R00001	branch_02	20120405	19	beverages	hard drinks	2	4500
4	C0322	R00002	branch_02	20120522	16	fruits	banana/imported fruits	1	4490
5	C0322	R00002	branch_02	20120522	16	fruits	banana/imported fruits	1	4490
6	C0322	R00002	branch_02	20120522	16	fruits	banana/imported fruits	1	4490
		날짜	시간	구매건수	구매금액				

3. 고객 DB



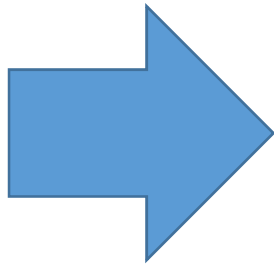
	고객 ID	성별	나이	거주지
	custId	sex	age	city
1	C0001	F	23	Kyeongki
2	C0002	F	68	Seoul
3	C0003	F	21	Kyeongki
4	C0004	F	45	Seoul
5	C0005	F	50	Seoul
6	C0006	F	61	Seoul



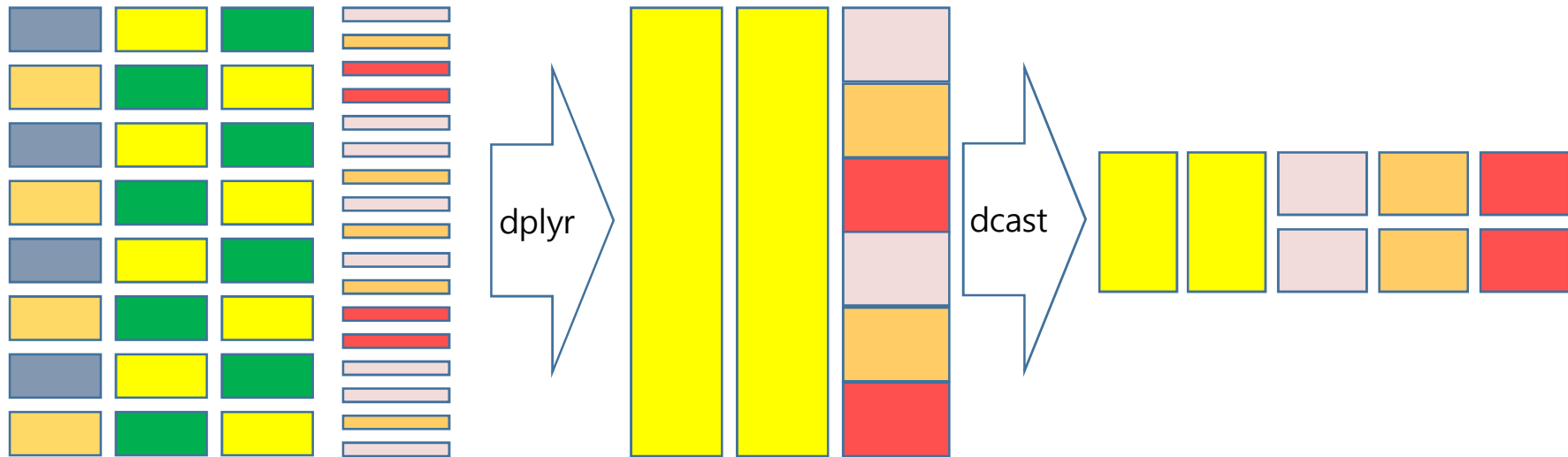
II. reshape2 패키지

1. cast 함수

데이터를 wide-form으로 재구조화 해주는 함수

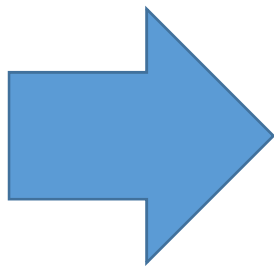


```
dcast( data이름,  
       열방향 벡돌1 + 열방향 벡돌2 ~ 행방향 벡돌,  
       value.var = ,  
       fun.aggregate= )
```

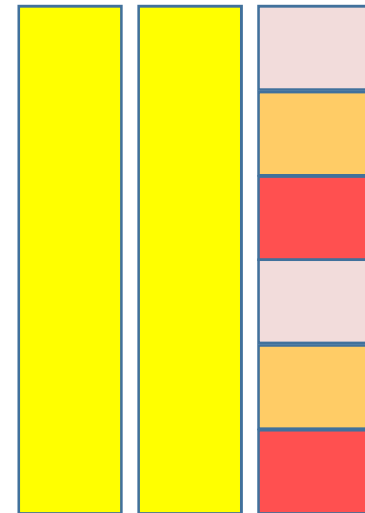
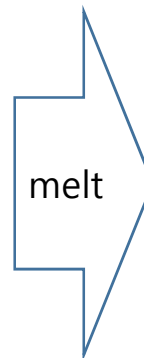
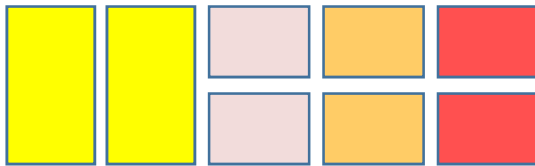


2. melt 함수

데이터를 long-form으로 재구조화 해주는 함수



```
melt ( data이름,  
       id.var =  
       )
```



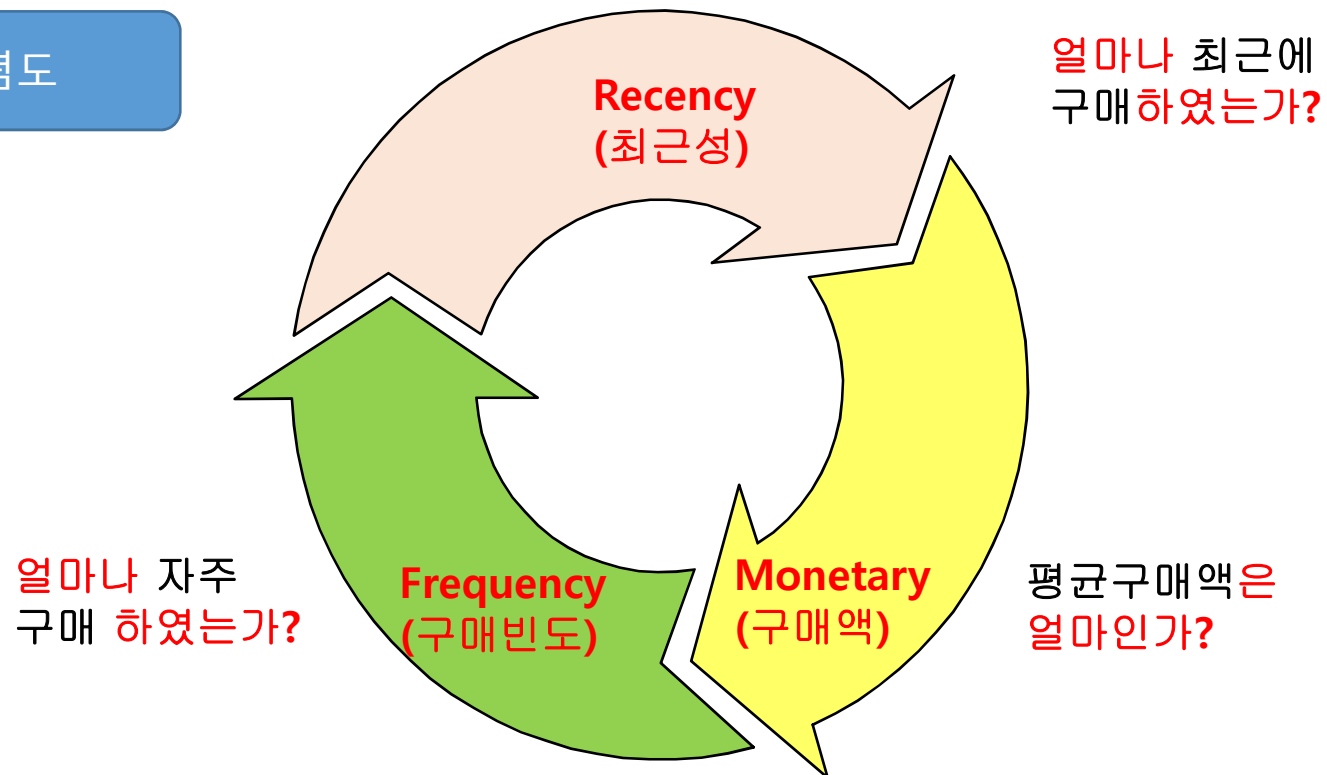


Ⅲ. RFM 분석 실습

1. RFM 개요

고객의 최근 방문 여부(**R**ecency), 구매 빈도(**F**requency), 사용 금액(**M**onetary) 등을 기준으로 지수화 함으로써, 고객을 세분화(**S**egmentation)하여 관리하는 고객 기법

RFM 개념도



2. RFM 등급 구분

계산의 단순화를 위해 RFM 각각의 Range를 5등분 하는 Quantile(백분위수)를 기준으로 등급을 구분하는 것으로 가정함.

- 백분위수(Quantile) 산출 함수 : quantile(데이터벡터, 백분위 지정 벡터)

구분	등급 구분	Point
최우수	상위 20%	5
우수	상위 20 ~ 40%	4
양호	상위 40 ~ 60%	3
보통	하위 20 ~ 40%	2
기타	하위 20%	1

3. RFM Score 산출

RFM 점수는 3개 RFM 분야별 등급 점수에 가중치를 곱한 후, 합산하여 산출

- $\text{RFM score} = A * \text{R등급} + B * \text{F등급} + C * \text{M등급}$

단, RFM 분야별 가중치 A, B, C는 매출액을 중심으로 결정하는 것으로 가정함.

각 등급별 상위 20%의 매출액을 가중치로 결정

$$A = \frac{\text{R등급 상위 20\%의 매출액}}{\text{전체 매출액 합계}}$$

$$B = \frac{\text{F등급 상위 20\%의 매출액}}{\text{전체 매출액 합계}}$$

$$C = \frac{\text{M등급 상위 20\%의 매출액}}{\text{전체 매출액 합계}}$$



A+B+C의 합계가 1이 되도록 평준화

$$A = \frac{A}{A + B + C}$$

$$B = \frac{B}{A + B + C}$$

$$C = \frac{C}{A + B + C}$$

자료) 위메프 송동근 (DK's Portfolio)

3. RFM Score 평준화

고객간 비교를 편리하게 하기 위하여 RFM 점수를 1에서 100사이의 점수로 평준화(Normalizing)함.

- 100점 환산 : $\text{RFM점수} * 100 / \text{RFM 점수 Range}$

등급	구분	평준화 점수	ABCD 등급
최우수	상위 20%	87.1	A
우수	상위 20 ~ 40%	69.8	B
양호	상위 40 ~ 60%	50.2	C
보통	하위 20 ~ 40%	37.3	D
기타	하위 20%	27.1	E

4. RFM 분석 활용 방안(descriptive)

주기적인 분석을 통해 RFM 고객 등급의 변화를 추적하는 것이 가능함.

1차	2차	변화유형	변화명수	1차고객/변화명수	총고객/변화고객수
최우수(2848명)	최우수	최우수 → 최우수	1590	<u>55.8%</u>	7.9%
	우수	최우수 → 우수	495	17.4%	2.4%
	일반	최우수 → 일반	445	15.6%	2.2%
	기타	최우수 → 기타	318	11.2%	1.6%
우수(3735명)	최우수	우수 → 최우수	1003	<u>26.9%</u>	5.0%
	우수	우수 → 우수	819	21.9%	4.0%
	일반	우수 → 일반	1019	27.3%	5.0%
	기타	우수 → 기타	894	23.9%	4.4%
일반(7062명)	최우수	일반 → 최우수	1079	15.3%	5.3%
	우수	일반 → 우수	1423	20.2%	7.0%
	일반	일반 → 일반	2293	32.5%	11.3%
	기타	일반 → 기타	2267	32.1%	11.2%
기타(6602명)	최우수	기타 → 최우수	682	10.3%	3.4%
	우수	기타 → 우수	1247	18.9%	6.2%
	일반	기타 → 일반	2244	34.0%	11.1%
	기타	기타 → 기타	2429	36.8%	12.0%
총합계			20247		100.0%

자료) 위메프 송동근 (DK's Portfolio)

4. RFM 분석 활용 방안(descriptive)

등급 하향 고객에 대해서는 이탈방지(Retention) 마케팅을 적용하고,
등급 상향 고객에 대해서는 Up-sell, Cross-sell 등의 마케팅 수단 적용을 고려할 수 있음.

[등급 하향 고객]



[등급 상향 고객]



자료) 위메프 송동근 (DK's Portfolio)

End of Document.

감사합니다.