비즈니스 애널리틱스

엑셀을 활용한 RFM 분석 실습

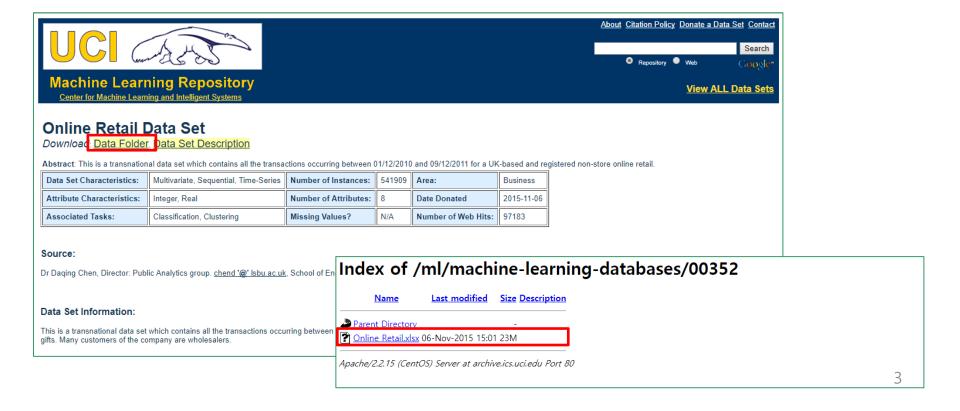
SNU Business School

RFM 실습



데이터 셋 다운로드

- · 아래 페이지에서 Online Retail Data Set을 다운받는다
 - URL: http://archive.ics.uci.edu/ml/datasets/online+retail
 - [Data Folder]를 클릭해 들어간 후 Online Retail.xlsx 엑셀 파일을 다운받는다



데이터 설명

· 2010년 12월 1일부터 2011년 12월 9일까지 1년 동안의 온라인 리테일 거래 내역 데이터

- InvoiceNo: 송장 번호(각 배송을 구분하기 위한 key)

- StockCode: 각 제품에 부여된 서로 구분되는 코드

- Description: 각 제품에 대한 설명

- Quantity: 구매 수량

- InvoiceDate: 발송 날짜

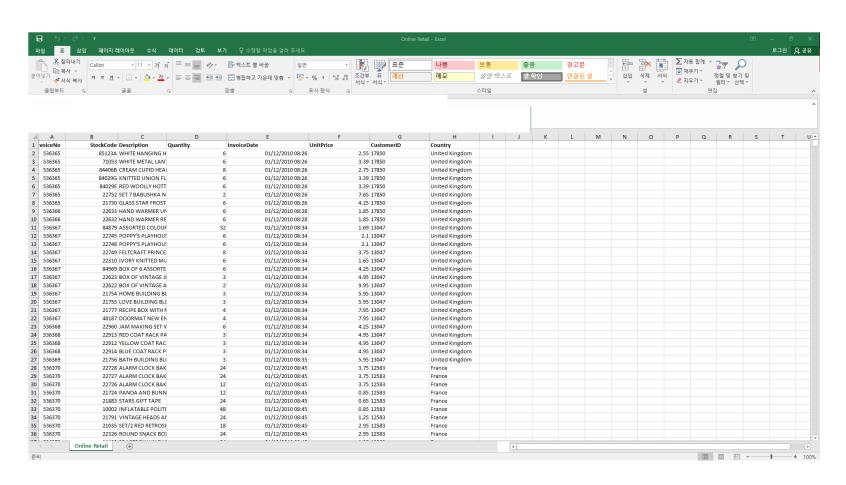
- UnitPrice: 개당 가격

- CustomerID: 각 고객에게 유일하게 부여되는 고객 ID

- Country: 배송 국가

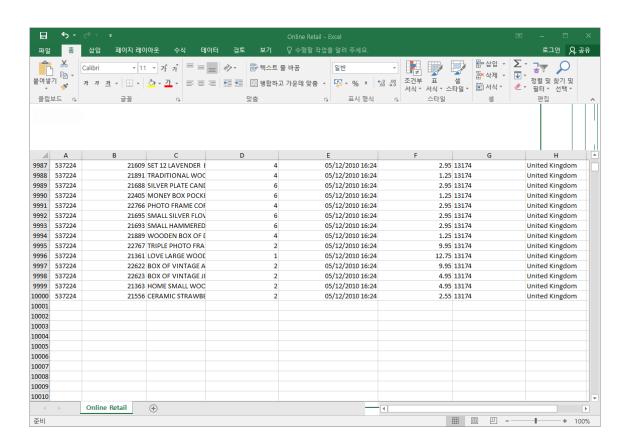
데이터 셋 다운로드

·로 데이터(Online Retail.xlsx)는 8개 행(A~H)과 541911개의 열로 이루어진 엑셀 파일이다



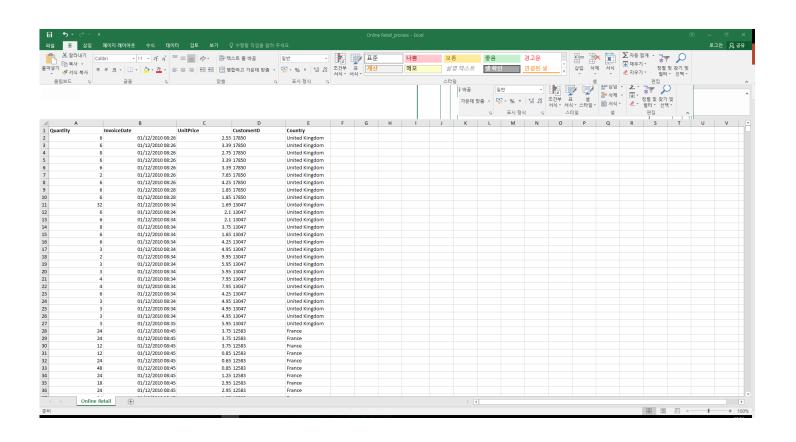
데이터 축소

- ·데이터가 너무 크므로 본 실습에서는 10000행까지의 데이터만 포함해서 분석해 본다
 - 10001행부터 마지막 행까지는 데이터를 모두 삭제한다



데이터 축소

- · voiceNo, StockCode, Description 데이터를 삭제한다
 - A-C열을 삭제한다



- •전체 데이터에 필터를 씌운다
 - 전체 데이터를 선택한 후(Ctrl + A), [데이터] 탭에서 [필터] 클릭



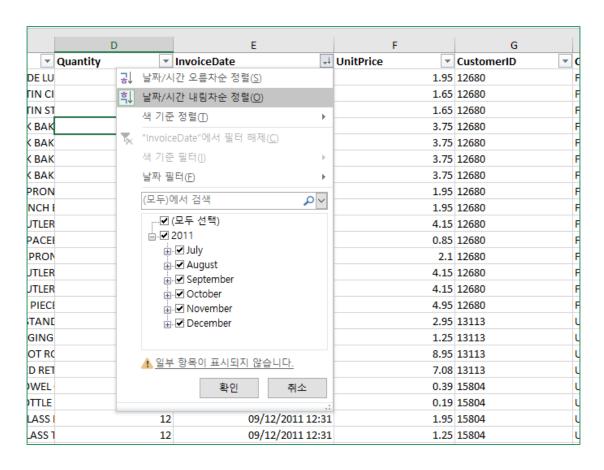
- 맨 위의 행의 각 열마다 ▼ 모양의 버튼이 생긴다

4	Α	В	С	D	Е	F	G
1	Quantity	InvoiceDate 🔻	UnitPrice <	CustomerID	Country		
2	6	01/12/2010 08:26	2.55	17850	United Kingdom		
3	6	01/12/2010 08:26	3.39	17850	United Kingdom		
4	8	01/12/2010 08:26	2.75	17850	United Kingdom		
5	6	01/12/2010 08:26	3.39	17850	United Kingdom		
6	6	01/12/2010 08:26	3.39	17850	United Kingdom		
7	2	01/12/2010 08:26	7.65	17850	United Kingdom		
8	6	01/12/2010 08:26	4.25	17850	United Kingdom		
9	6	01/12/2010 08:28	1.85	17850	United Kingdom		

- · CustomerID 열을 A열로 옮긴다
 - Quantity 열 옆에 새로운 열을 삽입한 후 CustomerID 열의 데이터를 잘라내기 후 붙인다

4	Α	В	С	D	Е	F	G
1	CustomerID	Quantity	InvoiceDate 🔻	UnitPrice 🔻	Country		
2	17850	6	01/12/2010 08:26	2.55	United Kingdom		
3	17850	6	01/12/2010 08:26	3.39	United Kingdom		
4	17850	8	01/12/2010 08:26	2.75	United Kingdom		
5	17850	6	01/12/2010 08:26	3.39	United Kingdom		
6	17850	6	01/12/2010 08:26	3.39	United Kingdom		
7	17850	2	01/12/2010 08:26	7.65	United Kingdom		
8	17850	6	01/12/2010 08:26	4.25	United Kingdom		
9	17850	6	01/12/2010 08:28	1.85	United Kingdom		
10	17850	6	01/12/2010 08:28	1.85	United Kingdom		
11	13047	32	01/12/2010 08:34	1.69	United Kingdom		
12	13047	6	01/12/2010 08:34	2.1	United Kingdom		
13	13047	6	01/12/2010 08:34	2.1	United Kingdom		
14	13047	8	01/12/2010 08:34	3.75	United Kingdom		
15	13047	6	01/12/2010 08:34	1.65	United Kingdom		
16	13047	6	01/12/2010 08:34	4.25	United Kingdom		
17	13047	3	01/12/2010 08:34	4.95	United Kingdom		
18	13047	2	01/12/2010 08:34	9.95	United Kingdom		
19	13047	3	01/12/2010 08:34	5.95	United Kingdom		
20	13047	3	01/12/2010 08:34	5.95	United Kingdom		
21	13047	4	01/12/2010 08:34	7.95	United Kingdom		
22	13047	4	01/12/2010 08:34	7.95	United Kingdom		
23	13047	6	01/12/2010 08:34	4.25	United Kingdom		
24	13047	3	01/12/2010 08:34	4.95	United Kingdom		
25	13047	3	01/12/2010 08:34	4.95	United Kingdom		
26	13047	3	01/12/2010 08:34	4.95	United Kingdom		
27	13047	3	01/12/2010 08:35	5.95	United Kingdom		

- · InvoiceDate 열을 내림차순으로 정렬한다
 - ☑ 버튼 클릭 후 [날짜/시간 내림차순 정렬] 클릭



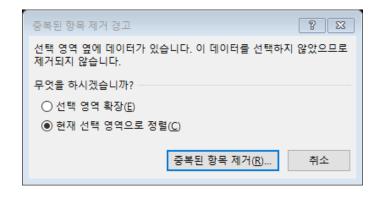
- · TotalPrice 열을 새로 만들고 그 값은 Quantity와 UnitPrice를 곱해 계산한다
 - E열에 TotalPrice를 만든다
 - E2에 "=B2*D2"를 입력하고 E10000까지 복사/붙여넣기를 한다

1	А	В	С	D	E	F
1	CustomerID	Quantity	InvoiceDate 🚭	UnitPrice <u></u>	TotalPrice 🔻	Country
2	17850	2	05/12/2010 16:24	9.95	=B2*D2	United Kingdom
3	17850	1	05/12/2010 16:24	1.65	1.65	United Kingdom
4	17850	2	05/12/2010 16:24	2.1	4.2	United Kingdom
5	17850	1	05/12/2010 16:24	1.65	1.65	United Kingdom
6	17850	1	05/12/2010 16:24	1.65	1.65	United Kingdom
7	17850	1	05/12/2010 16:24	1.65	1.65	United Kingdom
8	17850	1	05/12/2010 16:24	2.1	2.1	United Kingdom
9	17850	1	05/12/2010 16:24	2.95	2.95	United Kingdom
10	17850	1	05/12/2010 16:24	1.65	1.65	United Kingdom
11	13047	7	05/12/2010 16:24	1.25	8.75	United Kingdom
12	13047	4	05/12/2010 16:24	1.25	5	United Kingdom
13	13047	3	05/12/2010 16:24	1.25	3.75	United Kingdom
14	13047	6	05/12/2010 16:24	1.95	11.7	United Kingdom
15	13047	11	05/12/2010 16:24	0.85	9.35	United Kingdom
16	13047	4	05/12/2010 16:24	2.1	8.4	United Kingdom
17	13047	1	05/12/2010 16:24	1.65	1.65	United Kingdom
18	13047	3	05/12/2010 16:24	2.1	6.3	United Kingdom
19	13047	3	05/12/2010 16:24	2.1	6.3	United Kingdom
20	13047	5	05/12/2010 16:24	1.65	8.25	United Kingdom
21	13047	11	05/12/2010 16:24	1.65	18.15	United Kingdom
22	13047	1	05/12/2010 16:24	1.65	1.65	United Kingdom
23	13047	1	05/12/2010 16:24	1.95	1.95	United Kingdom
			! !			

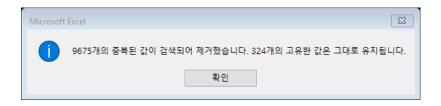
- · 기준일(date)를 J2 셀에 입력한다
 - 본 실습에서는 2012년 1월 1일(01/01/2012)로 입력

_ A	В	С	D	E	F	G	Н	1	J
1 CustomerID	Quantity	InvoiceDate 🚭	UnitPrice <	TotalPrice 💌	Country				Date
2 17850	2	05/12/2010 16:24	9.95	19.9	United Kingdom				01/01/2012
3 17850	1	05/12/2010 16:24	1.65	1.65	United Kingdom				
4 17850	2	05/12/2010 16:24	2.1	4.2	United Kingdom				
5 17850	1	05/12/2010 16:24	1.65	1.65	United Kingdom				
6 17850	1	05/12/2010 16:24	1.65	1.65	United Kingdom				
7 17850	1	05/12/2010 16:24	1.65	1.65	United Kingdom				
8 17850	1	05/12/2010 16:24	2.1	2.1	United Kingdom				
9 17850	1	05/12/2010 16:24	2.95	2.95	United Kingdom				
10 17850	1	05/12/2010 16:24	1.65	1.65	United Kingdom				
11 13047	7	05/12/2010 16:24	1.25	8.75	United Kingdom				
12 13047	4	05/12/2010 16:24	1.25	5	United Kingdom				
13 13047	3	05/12/2010 16:24	1.25	3.75	United Kingdom				
14 13047	6	05/12/2010 16:24	1.95	11.7	United Kingdom				
15 13047	11	05/12/2010 16:24	0.85	9.35	United Kingdom				
16 13047	4	05/12/2010 16:24	2.1	8.4	United Kingdom				
17 13047	1	05/12/2010 16:24	1.65	1.65	United Kingdom				
18 13047	3	05/12/2010 16:24	2.1	6.3	United Kingdom				
19 13047	3	05/12/2010 16:24	2.1	6.3	United Kingdom				
20 13047	5	05/12/2010 16:24	1.65	8.25	United Kingdom				
21 13047	11	05/12/2010 16:24	1.65	18.15	United Kingdom				
22 13047	1	05/12/2010 16:24	1.65	1.65	United Kingdom				
23 13047	1	05/12/2010 16:24	1.95	1.95	United Kingdom				

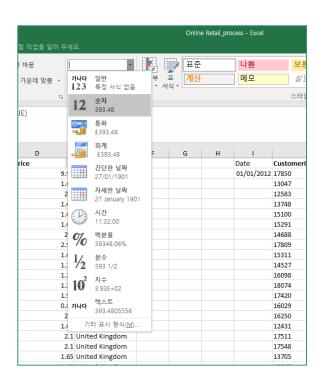
- · 고객 ID를 중복됨 없이 K1에 복사한다
 - A1:A10000을 복사해 K1에 붙여넣는다
 - [데이터] 탭에서 [중복된 항목 제거] 클릭 후 [현재 선택 영역으로 정렬(C)] 선택



- [중복된 항목 제거(R)] 클릭



- · L2:L10000에 RECENCY 지표를 계산해 넣는다
 - L2에 "=\$J\$2-VLOOKUP(K2,\$A\$2:\$F\$10000,3,TRUE)" 라고 입력 후 [표시 형식]을 숫자로 변환



- L10000까지 복사/붙여넣기를 한다

- · M2:M10000에 FREQUENCY 지표를 계산해 넣는다
 - M2에 "=COUNTIF(A2:A10000,K2)" 라고 입력한다
 - M10000까지 복사/붙여넣기를 한다

J	K	L	M	N	0
Date	CustomerID	Recency	Frequency	Monetary	
01/01/2012	17850	393.48	=COUNTIF	(A2:A1000	0,K2)
	13047	391.50	COUNTI	F(range, crite	eria)
	12583	391.50	20		
	13748	391.50	1		
	15100	391.50	1		
	15291	391.50	2		
	14688	391.50	19		
	17809	393.48	15		
	15311	391.50	69		
	14527	391.50	30		
	16098	391.50	12		
	18074	394.24	13		
	17420	393.48	7		
	16029	391.50	12		

- · N2:N10000에 MONETARY 지표를 계산해 넣는다
 - N2에 "=SUMIF(A2:A10000,K2,E2:E10000)" 라고 입력한다
 - N10000까지 복사/붙여넣기를 한다

J	K	L	M	N	0	Р	
Date	CustomerID	Recency	Frequency	Monetary			
01/01/2012	17850	393.48	297	=SUMIF(A	2:A10000,k	(2,E2:E100	000)
	13047	391.50	17	110.7			
	12583	391.50	20	95.05			
	13748	391.50	1	14.85			
	15100	391.50	1	10.2			
	15291	391.50	2	32.65			
	14688	391.50	19	205.99			
	17809	393.48	15	314.78			
	15311	391.50	69	329.98			
	14527	391.50	30	119			
	16098	391.50	12	58.3			
	18074	394.24	13	56.24			
	17420	393.48	7	79.78			
	16029	391.50	12	257.83			
	16250	391.50	14	73.75			
	12431	#N/A	14	147.23			
	17511	393.48	24	818.1			
	17548	393.48	7	219.94			
	13705	391.50	10	770.56			

- · 계산된 지표를 바탕으로 O2:Q10000까지 RFM 테이블을 완성한다
 - 기준은 각자 마음대로 정해본다
 - IF문을 활용한다
 - 예시
 - 만약 FREQUENCY가 100보다 작으면 F = 1, 100과 200 사이면 F = 2, 기타 F = 3

"=IF(M2<100,1,IF(M2<200,2,3))"

완성된 RFM 테이블

· RFM 테이블

J	K	L	M	N	0	Р	Q
Date	CustomerID	Recency	Frequency	Monetary	R	F	M
01/01/2012	17850	393.48	297	4365.18	1	3	3
	13047	391.50	17	110.7	3	1	1
	12583	391.50	20	95.05	3	1	1
	13748	391.50	1	14.85	3	1	1
	15100	391.50	1	10.2	3	1	1
	15291	391.50	2	32.65	3	1	1
	14688	391.50	19	205.99	3	1	1
	17809	393.48	15	314.78	1	1	1
	15311	391.50	69	329.98	3	1	1
	14527	391.50	30	119	3	1	1
	16098	391.50	12	58.3	3	1	1
	18074	394.24	13	56.24	1	1	1
	17420	393.48	7	79.78	1	1	1
	16029	391.50	12	257.83	3	1	1
	16250	391.50	14	73.75	3	1	1
	12431	#N/A	14	147.23	#N/A	1	1
	17511	393.48	24	818.1	1	1	1
	17548	393.48	7	219.94	1	1	1

- 결측치 등으로 인해 무효한 값(#N/A)가 있을 수 있음

References

· Fader, P. S., & Hardie, B. G. (2008). Creating an RFM summary using Excel.