

# 고객을 세그멘테이션하자 [프로젝트] 최효근

## 11-2. 데이터 불러오기

### 데이터 살펴보기

- 테이블에 있는 10개의 행만 출력하기

```
select *  
from `artful-striker-473603-f9.modulabs_project.data`  
LIMIT 10;
```

[결과 이미지를 넣어주세요]

행	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
1	536365	851228A	WHITE HANGING HEART FLAG...	6	2010-12-01 08:26:00 UTC	2.55	17850	United Kingdom
2	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00 UTC	3.39	17850	United Kingdom
3	536365	844068	CREAM CUPID HEARTS COAT H...	8	2010-12-01 08:26:00 UTC	2.75	17850	United Kingdom
4	536365	840295	KNITTED UNION FLAG HOT WA...	6	2010-12-01 08:26:00 UTC	3.39	17850	United Kingdom
5	536365	840296	RED WOOLLY HOTTIE WHITE H...	6	2010-12-01 08:26:00 UTC	3.39	17850	United Kingdom
6	536365	22752	SET 7 BABUSHKA NESTING BO...	2	2010-12-01 08:26:00 UTC	7.65	17850	United Kingdom
7	536365	21790	GLASS STAR FROSTED FLIGHT...	6	2010-12-01 08:26:00 UTC	4.25	17850	United Kingdom
8	536366	22633	HAND WARMER UNION JACK	6	2010-12-01 08:28:00 UTC	1.85	17850	United Kingdom
9	536366	22632	HAND WARMER RED POLKA DOT	6	2010-12-01 08:28:00 UTC	1.85	17850	United Kingdom
10	536367	84879	ASSORTED COLOUR BIRD ORN...	32	2010-12-01 08:34:00 UTC	1.69	13047	United Kingdom

- 전체 데이터는 몇 행으로 구성되어 있는지 확인하기

```
SELECT COUNT(*) AS total_rows  
FROM `artful-striker-473603-f9.modulabs_project.data`;
```

[결과 이미지를 넣어주세요]

행	total_rows
1	541909

## 데이터 수 세기

- COUNT 함수를 사용해서, 각 컬럼별 데이터 포인트의 수를 세어 보기

```
SELECT  
COUNT(InvoiceNo) AS InvoiceNo_count,  
COUNT(StockCode) AS StockCode_count,  
COUNT(Description) AS Description_count,  
COUNT(Quantity) AS Quantity_count,  
COUNT(InvoiceDate) AS InvoiceDate_count,  
COUNT(UnitPrice) AS UnitPrice_count,  
COUNT(CustomerID) AS CustomerID_count,  
COUNT(Country) AS Country_count  
FROM `artful-striker-473603-f9.modulabs_project.data`;
```

[결과 이미지를 넣어주세요]

행	InvoiceNo_count	StockCode_count	Description_count	Quantity_count	InvoiceDate_count	UnitPrice_count	CustomerID_count	Country_count
1	541909	541909	540455	541909	541909	541909	406829	541909

## 11-4. 데이터 전처리 방법(1): 결측치 제거

### 컬럼 별 누락된 값의 비율 계산

- 각 컬럼 별 누락된 값의 비율을 계산
  - 각 컬럼에 대해서 누락 값을 계산한 후, 계산된 누락 값을 UNION ALL을 통해 합치기

```
SELECT 'InvoiceNo' AS InvoiceNo, COUNT(*) - COUNT(InvoiceNo) AS null_count
FROM `artful-striker-473603-f9.modulabs_project.data`

UNION ALL

SELECT 'StockCode' AS StockCode, COUNT(*) - COUNT(StockCode) AS null_count
FROM `artful-striker-473603-f9.modulabs_project.data`

UNION ALL

SELECT 'Description' AS Description, COUNT(*) - COUNT(Description) AS null_count
FROM `artful-striker-473603-f9.modulabs_project.data`

UNION ALL

SELECT 'Quantity' AS Quantity, COUNT(*) - COUNT(Quantity) AS null_count
FROM `artful-striker-473603-f9.modulabs_project.data`

UNION ALL

SELECT 'InvoiceDate' AS InvoiceDate, COUNT(*) - COUNT(InvoiceDate) AS null_count
FROM `artful-striker-473603-f9.modulabs_project.data`

UNION ALL

SELECT 'UnitPrice' AS UnitPrice, COUNT(*) - COUNT(UnitPrice) AS null_count
FROM `artful-striker-473603-f9.modulabs_project.data`

UNION ALL

SELECT 'CustomerID' AS CustomerID, COUNT(*) - COUNT(CustomerID) AS null_count
FROM `artful-striker-473603-f9.modulabs_project.data`

UNION ALL

SELECT 'Country' AS Country, COUNT(*) - COUNT(Country) AS null_count
FROM `artful-striker-473603-f9.modulabs_project.data`;
```

[결과 이미지를 넣어주세요]

행	InvoiceNo	null_count
1	Quantity	0
2	Country	0
3	CustomerID	135080
4	UnitPrice	0
5	InvoiceDate	0
6	Description	1454
7	InvoiceNo	0
8	StockCode	0

## 결측치 처리 전략

- StockCode = '85123A' 의 Description 을 추출하는 쿼리문을 작성하기

```
SELECT Description
FROM `artful-striker-473603-f9.modulabs_project.data`
WHERE StockCode = '85123A';
```

[결과 이미지를 넣어주세요]

행	Description
1	WHITE LONGSLEEVE SHIRT
2	WHITE HANDED HEART T SHIRT
3	WHITE LONGSLEEVE SHIRT
4	WHITE HANDED HEART T SHIRT
5	WHITE LONGSLEEVE SHIRT
6	WHITE HANDED HEART T SHIRT
7	WHITE LONGSLEEVE SHIRT
8	WHITE HANDED HEART T SHIRT
9	WHITE LONGSLEEVE SHIRT
10	WHITE HANDED HEART T SHIRT
11	WHITE LONGSLEEVE SHIRT
12	WHITE HANDED HEART T SHIRT
13	WHITE LONGSLEEVE SHIRT
14	WHITE HANDED HEART T SHIRT
15	WHITE LONGSLEEVE SHIRT
16	WHITE HANDED HEART T SHIRT
17	WHITE LONGSLEEVE SHIRT
18	WHITE HANDED HEART T SHIRT
19	WHITE LONGSLEEVE SHIRT
20	WHITE HANDED HEART T SHIRT
21	WHITE LONGSLEEVE SHIRT
22	WHITE HANDED HEART T SHIRT
23	WHITE LONGSLEEVE SHIRT
24	WHITE HANDED HEART T SHIRT
25	WHITE LONGSLEEVE SHIRT
26	WHITE HANDED HEART T SHIRT
27	WHITE LONGSLEEVE SHIRT
28	WHITE HANDED HEART T SHIRT
29	WHITE LONGSLEEVE SHIRT
30	WHITE HANDED HEART T SHIRT
31	WHITE LONGSLEEVE SHIRT
32	WHITE HANDED HEART T SHIRT
33	WHITE LONGSLEEVE SHIRT
34	WHITE HANDED HEART T SHIRT
35	WHITE LONGSLEEVE SHIRT
36	WHITE HANDED HEART T SHIRT
37	WHITE LONGSLEEVE SHIRT
38	WHITE HANDED HEART T SHIRT
39	WHITE LONGSLEEVE SHIRT
40	WHITE HANDED HEART T SHIRT
41	WHITE LONGSLEEVE SHIRT
42	WHITE HANDED HEART T SHIRT
43	WHITE LONGSLEEVE SHIRT
44	WHITE HANDED HEART T SHIRT
45	WHITE LONGSLEEVE SHIRT
46	WHITE HANDED HEART T SHIRT
47	WHITE LONGSLEEVE SHIRT
48	WHITE HANDED HEART T SHIRT
49	WHITE LONGSLEEVE SHIRT
50	WHITE HANDED HEART T SHIRT

## 결측치 처리

- DELETE 구문을 사용하며, WHERE 절을 통해 데이터를 제거할 조건을 제시

```
DELETE FROM `artful-striker-473603-f9.modulabs_project.data`
WHERE StockCode = '85123A';
```

[결과 이미지를 넣어주세요]

이 문으로 data의 행 2,313개가 삭제되었습니다.

## 11-5. 데이터 전처리(2): 중복값 처리

### 중복값 확인

- 중복된 행의 수를 세어보기
  - 8개의 컬럼에 그룹 함수를 적용한 후, COUNT가 1보다 큰 데이터를 세어보기

```
SELECT *
FROM project_name.modulabs_project.data
# [[YOUR QUERY]]
```

```
SELECT
  InvoiceNo,
  StockCode,
  Description,
  Quantity,
  InvoiceDate,
  UnitPrice,
  CustomerID,
  Country,
```

```

COUNT(*) AS cnt
FROM `artful-striker-473603-f9.modulabs_project.data`
GROUP BY
InvoiceNo,
StockCode,
Description,
Quantity,
InvoiceDate,
UnitPrice,
CustomerId,
Country
HAVING COUNT(*) > 1;

```

[결과 이미지를 넣어주세요]

#	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerId	Country	cnt
1	539733	22962	JAM JAR WITH A PINK LID	2	2010-12-21 15:00 UTC	6.85	nafl	United Kingdom	2
3	539856	22518	CHILD'S GARDEN BRUSH BLUE	1	2010-12-22 14:41 UTC	0.0	nafl	United Kingdom	2
8	541975	22452	MEASURING TAPE BABUSHKA...	24	2011-01-24 14:24 UTC	2.95	nafl	Hong Kong	2
4	541975	22454	MEASURING TAPE BABUSHKA...	24	2011-01-24 14:24 UTC	2.95	nafl	Hong Kong	2
5	541975	20749	ASSORTED COLOUR MINI CANS	6	2011-01-24 14:24 UTC	7.95	nafl	Hong Kong	2
6	541975	20682	RED RETROSPOT CHILDRENS U...	6	2011-01-24 14:24 UTC	9.25	nafl	Hong Kong	2
7	542867	22131	FOOD CONTAINER SET 3 LOVE...	2	2011-02-01 12:22 UTC	1.95	nafl	United Kingdom	2
8	543899	21121	SET/10 RED POLKA DOT PARTY...	24	2011-02-14 12:11 UTC	1.25	nafl	IRE	2
9	548604	22453	MEASURING TAPE BABUSHKA...	1	2011-04-01 10:58 UTC	2.95	nafl	United Kingdom	2
10	551921	22584	FRENCH BLUE METAL KITCHEN...	2	2011-05-17 14:55 UTC	0.0	nafl	United Kingdom	2
11	556813	23172	REGENCY TEA PLATE PINK	10	2011-06-14 17:38 UTC	1.65	nafl	United Kingdom	2
12	557644	23290	SPACEBOY CHILDRENS BOWL	7	2011-06-21 17:06 UTC	1.25	nafl	United Kingdom	2
13	558840	22430	ENAMEL WATERING CAN CREAM	1	2011-06-28 14:01 UTC	0.0	nafl	United Kingdom	2
14	559520	90122A	PINK CRYSTAL GLASS BRACEL...	1	2011-07-08 16:26 UTC	4.95	nafl	United Kingdom	2
15	559593	23339	EGG FRYING PAN BLUE	2	2011-07-11 16:21 UTC	2.08	nafl	United Kingdom	2
16	560245	22494	EMERGENCY FIRST AID TIN	1	2011-07-17 12:15 UTC	1.25	nafl	Switzerland	2
17	560245	23298	SPOTTY BUNTING	1	2011-07-17 12:15 UTC	4.95	nafl	Switzerland	2
18	562905	23296	LUNCH BAG APPLE DESIGN	1	2011-08-19 17:12 UTC	1.65	nafl	United Kingdom	2
19	562965	23033	SMALL PASSIONNE HEART PL...	1	2011-08-19 17:12 UTC	2.49	nafl	United Kingdom	2
20	565303	21558	SKELL LUNCH BOX WITH OUTL...	1	2011-09-02 12:17 UTC	2.95	nafl	Unspecified	2
21	565455	22429	ENAMEL MEASURING JUG CRE...	1	2011-09-04 16:09 UTC	4.35	nafl	United Kingdom	2
22	571061	23201	JUMBO BAG ALPHABET	6	2011-10-13 14:33 UTC	4.13	nafl	United Kingdom	2
23	573036	21210	SET OF 72 RETROSPOT PAPER...	1	2011-10-28 14:38 UTC	1.45	nafl	Switzerland	2
24	573286	23287	RED VINTAGE SPOT BEAKER	4	2011-10-28 14:38 UTC	6.85	nafl	Switzerland	2
25	573290	23396	LE JARDIN BOTANIQUE CUSHIO...	1	2011-10-28 14:51 UTC	3.75	nafl	Switzerland	2
26	577489	23321	RENDER HEART DECORATION...	1	2011-11-20 12:07 UTC	6.83	nafl	United Kingdom	2
27	578744	23010	CIRCUS PARADE BABY GIFT SET	1	2011-11-30 14:38 UTC	16.95	nafl	United Kingdom	2
28	OS40580	5	SAMPLES	-1	2011-02-21 14:25 UTC	9.74	nafl	United Kingdom	2
29	CS47725	84629	ASSTD FRUIT+1 LONKES FRDG...	-144	2011-03-25 10:43 UTC	6.55	nafl	United Kingdom	2
30	CS57663	21121	SET/10 RED POLKA DOT PARTY...	-24	2011-06-21 17:59 UTC	1.25	nafl	ERE	2
31	571054	23245	SET OF 5 REGENCY CAKE TINS	4	2011-10-13 12:47 UTC	4.95	12359	Cyprus	2
32	571054	23239	SET OF 4 KNOCK KNOCK TINS P...	6	2011-10-13 12:47 UTC	4.15	12359	Cyprus	2
33	577228	22144	CHRISTMAS CRAFT LITTLE PR...	1	2011-11-18 12:07 UTC	2.1	12391	Cyprus	2
34	577228	84080	MEUSE TOP WITH PINK S-BART	1	2011-11-18 12:07 UTC	3.75	12391	Cyprus	2
35	577228	23048	SET OF 10 LANTERNS FAIRY L...	1	2011-11-18 12:07 UTC	4.15	12391	Cyprus	2
36	577228	23156	SET OF 5 MINI GROCERY MAG...	1	2011-11-18 12:07 UTC	2.08	12391	Cyprus	2
37	577228	22435	SET OF 9 HEART SHAPED BALL...	1	2011-11-18 12:07 UTC	1.25	12391	Cyprus	2
38	588174	23226	ROUND SNACK BOXES SET OFF...	12	2010-12-10 09:35 UTC	2.95	12471	Germany	2
39	555162	22734	WRAP RED APPLES	25	2011-08-01 18:15 UTC	6.42	12473	Germany	2
40	576607	22940	FELTCRAFT CHRISTMAS FAIRY	8	2011-11-15 15:42 UTC	4.35	12474	Germany	2
41	576607	22086	PAPER CHAIN KIT 50'S CHRIST...	6	2011-11-15 15:42 UTC	2.95	12474	Germany	2
42	553731	23238	LUNCH BAG VINTAGE LEAF DE...	10	2011-05-19 08:26 UTC	1.65	12481	Germany	2
43	564734	72131	COLUMBIAN CANDLE RECTRA...	1	2011-08-28 13:32 UTC	1.95	12484	Spain	2
44	564734	47890	TEA TIME DES TEA COOY	1	2011-08-28 13:32 UTC	2.95	12484	Spain	2
45	574860	23043	SET OF TEA COFFEE SUGAR TL...	2	2011-11-04 15:18 UTC	4.95	12484	Spain	2
46	575692	21818	GILTIED HEART DECORATION...	1	2011-11-19 16:27 UTC	6.39	12508	France	2
47	569527	21992	VINTAGE FAIRLY STATIONERY...	1	2011-10-04 14:03 UTC	1.25	12512	Israel	2
48	555383	22648	PINK BABY BUNTING	5	2011-06-02 15:13 UTC	2.95	12517	Germany	2
49	555383	21828	EIGHT PIECE SNAKE SET	2	2011-06-02 15:13 UTC	1.25	12517	Germany	2
50	575885	22059	CERAMIC STRAWBERRY DESIG...	1	2011-11-11 13:57 UTC	1.49	12517	Germany	2

## 중복값 처리

- 중복값을 제거하는 쿼리문 작성하기
  - CREATE OR REPLACE TABLE 구문을 활용하여 모든 컬럼(\*)을 DISTINCT 한 데이터로 업데이트

```

SELECT DISTINCT *
FROM `artful-striker-473603-f9.modulabs_project.data`
LIMIT 100;

```

```

CREATE OR REPLACE TABLE `artful-striker-473603-f9.modulabs_project.data` AS
SELECT DISTINCT *
FROM `artful-striker-473603-f9.modulabs_project.data`;

```

[결과 이미지를 넣어주세요]

일	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
1	536414	22139	red	56	2010-12-01 11:52:00 UTC	0.0	red	United Kingdom
2	530544	21189	WHITE HONEYCOMB PAPER G...	4	2010-12-01 14:32:00 UTC	3.36	red	United Kingdom
3	530544	21870	I CAN ONLY PLEASE ONE PER...	1	2010-12-01 14:32:00 UTC	3.36	red	United Kingdom
4	530544	901200	GREEN MURANO TWIST BRACE...	1	2010-12-01 14:32:00 UTC	4.24	red	United Kingdom
5	530544	84923	PINK BUTTERFLY HANDBAG W...	1	2010-12-01 14:32:00 UTC	4.21	red	United Kingdom
6	530544	90194	SILVER LABIAI BLACK STONE ...	1	2010-12-01 14:32:00 UTC	4.24	red	United Kingdom
7	530544	21218	RED SPOTTY DISCOUT TIN	1	2010-12-01 14:32:00 UTC	6.47	red	United Kingdom
8	530544	21914	BLUE HANNAKING IN BOX	3	2010-12-01 14:32:00 UTC	2.51	red	United Kingdom
9	530544	170125	ORGANIC SANDELYWOOD NICEN...	1	2010-12-01 14:32:00 UTC	2.51	red	United Kingdom
10	530544	21012	ANTIQUE ALL GLASS CANDLES...	2	2010-12-01 14:32:00 UTC	2.51	red	United Kingdom
11	530544	21569	STRAWBERRY LUNCH BOX WIT...	1	2010-12-01 14:32:00 UTC	5.06	red	United Kingdom
12	530544	21169	YOU'RE CONFUSING ME META...	1	2010-12-01 14:32:00 UTC	4.21	red	United Kingdom
13	530544	22219	LOWBIRD HANGING DECORATI...	3	2010-12-01 14:32:00 UTC	1.66	red	United Kingdom
14	530544	805020	VALETTE IMAGES GIFT WRAP ...	2	2010-12-01 14:32:00 UTC	1.66	red	United Kingdom
15	530544	21888	BINGO SET	1	2010-12-01 14:32:00 UTC	7.62	red	United Kingdom
16	530544	826130	METAL SIGN CUPCAKE SINGLE...	1	2010-12-01 14:32:00 UTC	2.51	red	United Kingdom
17	530544	17129F	BLUE GLASS GEMS IN BAG	2	2010-12-01 14:32:00 UTC	1.66	red	United Kingdom
18	530544	85017A	ENVELOPE 50 ROMANTIC IMA...	1	2010-12-01 14:32:00 UTC	0.85	red	United Kingdom
19	530544	22719	GUMBALL MONOCHROME OOA...	1	2010-12-01 14:32:00 UTC	2.51	red	United Kingdom
20	530544	21911	GARDEN METAL SIGN	1	2010-12-01 14:32:00 UTC	3.36	red	United Kingdom
21	530544	21429	RED GINGHAM ROSE JEWELL E...	5	2010-12-01 14:32:00 UTC	3.36	red	United Kingdom
22	530544	21563	RED HEART SHAPE LOVE BUCK...	2	2010-12-01 14:32:00 UTC	5.91	red	United Kingdom
23	530544	90200A	PURPLE SWEETHEART BRACEL...	1	2010-12-01 14:32:00 UTC	4.24	red	United Kingdom
24	530544	22910	PAPER CHAIN KIT VINTAGE CH...	4	2010-12-01 14:32:00 UTC	5.91	red	United Kingdom
25	530544	35971	ROSE FOLKART HEART DECOR...	1	2010-12-01 14:32:00 UTC	2.51	red	United Kingdom

26	530544	21858	GLASS BEURRE DISH	1	2010-12-01 14:32:00 UTC	8.47	red	United Kingdom
27	530544	22988	GOLDEN EGG CUP	3	2010-12-01 14:32:00 UTC	2.51	red	United Kingdom
28	530544	85182	HAND OVER THE CHOCOLATE ...	1	2010-12-01 14:32:00 UTC	4.21	red	United Kingdom
29	530544	71063	WHITE METAL LANTERN	1	2010-12-01 14:32:00 UTC	8.47	red	United Kingdom
30	530544	21871	SAVE THE PLANET MUG	5	2010-12-01 14:32:00 UTC	3.36	red	United Kingdom
31	530544	21349	HVY HEART WREATH	1	2010-12-01 14:32:00 UTC	13.57	red	United Kingdom
32	530544	21677	PACK OF 60 PINK PRISPY CAK...	1	2010-12-01 14:32:00 UTC	1.28	red	United Kingdom
33	530544	47591B	SCOTTIES CHILDREN APRON	2	2010-12-01 14:32:00 UTC	1.66	red	United Kingdom
34	530544	84590	MOLINE TOW WITH PINK T-SHIRT	1	2010-12-01 14:32:00 UTC	4.21	red	United Kingdom
35	530544	22196	SMALL HEART MEASURING SP...	2	2010-12-01 14:32:00 UTC	1.66	red	United Kingdom
36	530544	85099B	JUMBO BAG RED RETROSPOT	1	2010-12-01 14:32:00 UTC	4.21	red	United Kingdom
37	530544	22130	PARTY COME CHRISTMAS DEC...	2	2010-12-01 14:32:00 UTC	8.47	red	United Kingdom
38	530544	37449	CERAMIC CAKE STAND + HAN...	1	2010-12-01 14:32:00 UTC	21.23	red	United Kingdom
39	530544	84949	SILVER HANGING T-LIGHT HOL...	1	2010-12-01 14:32:00 UTC	3.36	red	United Kingdom
40	530544	22141	CHRISTMAS CRAFT TREE TOP ...	4	2010-12-01 14:32:00 UTC	4.21	red	United Kingdom
41	530544	22574	HEART WOODEN CHRISTMAS ...	1	2010-12-01 14:32:00 UTC	1.66	red	United Kingdom
42	530544	22515	CHILD GARDEN SPADE PINK	1	2010-12-01 14:32:00 UTC	4.21	red	United Kingdom
43	530544	21679	SKULLS STICKERS	3	2010-12-01 14:32:00 UTC	1.66	red	United Kingdom
44	530544	79030D	TUMBLER BAROQUE	1	2010-12-01 14:32:00 UTC	12.72	red	United Kingdom
45	530544	22866	RECIPE BOX BLUE SKETCHBOOK...	1	2010-12-01 14:32:00 UTC	5.91	red	United Kingdom
46	530544	90129A	PINK GLASS TASSLE BAG CHA...	2	2010-12-01 14:32:00 UTC	1.66	red	United Kingdom
47	530544	22345	PARTY PIZZA DISH BLUE POLK...	1	2010-12-01 14:32:00 UTC	1.66	red	United Kingdom
48	530544	21944	KITTENS DESIGN PLANNER	1	2010-12-01 14:32:00 UTC	1.66	red	United Kingdom
49	530544	20966	SANDWICH BATH SPONGE	4	2010-12-01 14:32:00 UTC	2.51	red	United Kingdom
50	530544	22153	ANGEL DECORATION STARS G...	1	2010-12-01 14:32:00 UTC	0.85	red	United Kingdom

## 11-6. 데이터 전처리(3): 오류값 처리

### InvoiceNo 살펴보기

- 고유(unique)한 InvoiceNo 의 개수를 출력하기

```
SELECT COUNT(DISTINCT InvoiceNo) AS unique_invoice_count
FROM `artful-striker-473603-f9.modulabs_project.data`;
```

[결과 이미지를 넣어주세요]

행	unique_invoice_c...
1	25871

- 고유한 InvoiceNo 를 앞에서부터 100개를 출력하기

```
SELECT DISTINCT InvoiceNo
FROM `artful-striker-473603-f9.modulabs_project.data`
LIMIT 100;
```

[결과 이미지를 넣어주세요]

행	InvoiceNo
1	5366414
2	536544
3	536545
4	536546
5	536547
6	536549
7	536550
8	536552
9	536553
10	536554
11	536555
12	536558
13	536565
14	536589
15	536592
16	536596
17	536640
18	536755
19	536756
20	536764
21	536765
22	536780
23	536801
24	536862
25	536864

26	536865
27	536866
28	536876
29	536883
30	536885
31	536920
32	536941
33	536942
34	536982
35	536991
36	536995
37	536996
38	536997
39	536998
40	536999
41	537000
42	537001
43	537002
44	537003
45	537004
46	537005
47	537006
48	537007
49	537008
50	537009

- **InvoiceNo** 가 'C'로 시작하는 행을 필터링 할 수 있는 쿼리문을 작성하기 (100행까지만 출력)

```
SELECT *
FROM `artful-striker-473603-f9.modulabs_project.data`
WHERE InvoiceNo LIKE 'C%'
LIMIT 100;
```

[결과 이미지를 넣어주세요]

행	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
1	C537251	22429	ENAMEL MEASURING JUG CRE...	-2	2010-12-06 10:45:00 UTC	4.25	na1	United Kingdom
2	C537251	22418	10 COLOUR SPACEBOY PEN	-7	2010-12-06 10:45:00 UTC	0.85	na1	United Kingdom
3	C537251	22454	MEASURING TAPE BARUSHKA ...	-8	2010-12-06 10:45:00 UTC	2.95	na1	United Kingdom
4	C537251	21915	RED HARMONICA IN BOX	-4	2010-12-06 10:45:00 UTC	1.25	na1	United Kingdom
5	C537251	21891	TRADITIONAL WOODEN SKIPP...	-3	2010-12-06 10:45:00 UTC	1.25	na1	United Kingdom
6	C537251	22141	CHRISTMAS CRAFT TREE TOP...	-8	2010-12-06 10:45:00 UTC	2.1	na1	United Kingdom
7	C537251	22328	ROUND SNACK BOXES SET OF ...	-2	2010-12-06 10:45:00 UTC	2.95	na1	United Kingdom
8	C537251	22327	ROUND SNACK BOXES SET OF ...	-4	2010-12-06 10:45:00 UTC	2.95	na1	United Kingdom
9	C537251	22536	MAGIC DRAWING SLATE PURD...	-11	2010-12-06 10:45:00 UTC	6.42	na1	United Kingdom
10	C537251	22911	PAPER CRAN KIT LONDON	-2	2010-12-06 10:45:00 UTC	2.95	na1	United Kingdom
11	C537251	21826	EIGHT PIECE DINGOJAR SET	-4	2010-12-06 10:45:00 UTC	1.25	na1	United Kingdom
12	C537251	22520	4 TRADITIONAL SPINNING TOPS	-9	2010-12-06 10:45:00 UTC	1.25	na1	United Kingdom
13	C537251	21890	SUN WOODEN SKITTLES IN COT...	-2	2010-12-06 10:45:00 UTC	2.95	na1	United Kingdom
14	C537251	23564	ALPHABET STENCILS CRAFT	-5	2010-12-06 10:45:00 UTC	1.25	na1	United Kingdom
15	C537251	22943	CHRISTMAS LIGHTS 10 VINTA...	-6	2010-12-06 10:45:00 UTC	4.95	na1	United Kingdom
16	C537251	84347	ROTATING SILVER ANGELS TLL...	-9	2010-12-06 10:45:00 UTC	2.65	na1	United Kingdom
17	C537251	22748	PORPHY'S PLAYHOUSE KITCHEN	-4	2010-12-06 10:45:00 UTC	2.1	na1	United Kingdom
18	C537251	22466	FAIRY TALE COTTAGE NIGHTILL...	-3	2010-12-06 10:45:00 UTC	1.95	na1	United Kingdom
19	C537251	22940	FELTRAFT CHRISTMAS FAIRY	-5	2010-12-06 10:45:00 UTC	4.25	na1	United Kingdom
20	C537251	22747	PORPHY'S PLAYHOUSE BATHRO...	-6	2010-12-06 10:45:00 UTC	2.1	na1	United Kingdom
21	C537672	BANK CHARGES	Bank Charges	-1	2010-12-07 12:00:00 UTC	95.88	na1	United Kingdom
22	C537581	S	SAMPLES	-1	2010-12-07 12:03:00 UTC	52.0	na1	United Kingdom
23	C537581	S	SAMPLES	-1	2010-12-07 12:03:00 UTC	12.95	na1	United Kingdom
24	C537600	AMAZONFEE	AMAZON FEE	-1	2010-12-07 12:41:00 UTC	1.0	na1	United Kingdom
25	C537610	M	Manual	-1	2010-12-07 13:23:00 UTC	631.31	na1	United Kingdom

26	C537613	M	Manual	-1	2010-12-07 13:28:00 UTC	313.78	na1	United Kingdom
27	C537630	AMAZONFEE	AMAZON FEE	-1	2010-12-07 15:04:00 UTC	13541.33	na1	United Kingdom
28	C537644	AMAZONFEE	AMAZON FEE	-1	2010-12-07 15:34:00 UTC	13474.79	na1	United Kingdom
29	C537647	AMAZONFEE	AMAZON FEE	-1	2010-12-07 16:41:00 UTC	5079.25	na1	United Kingdom
30	C537651	AMAZONFEE	AMAZON FEE	-1	2010-12-07 16:49:00 UTC	5594.32	na1	United Kingdom
31	C537652	AMAZONFEE	AMAZON FEE	-1	2010-12-07 16:50:00 UTC	4706.71	na1	United Kingdom
32	C538109	M	Manual	-1	2010-12-13 10:35:00 UTC	133.00	na1	United Kingdom
33	C538680	BANK CHARGES	Bank Charges	-1	2010-12-13 17:10:00 UTC	964.92	na1	United Kingdom
34	C538681	M	Manual	-1	2010-12-13 17:12:00 UTC	316.3	na1	United Kingdom
35	C539692	M	Manual	-1	2010-12-13 17:14:00 UTC	1130.9	na1	United Kingdom
36	C539886	22457	GUMBALL COAT RACK	-1	2010-12-14 09:49:00 UTC	2.65	na1	United Kingdom
37	C539756	22720	SET OF 3 CAKE TINS PANTRY D...	-1	2010-12-21 16:31:00 UTC	4.95	na1	United Kingdom
38	C539948	21888	BINGO SET	-4	2010-12-23 11:48:00 UTC	3.75	na1	ERE
39	C540117	AMAZONFEE	AMAZON FEE	-1	2011-01-05 09:55:00 UTC	16888.02	na1	United Kingdom
40	C540118	AMAZONFEE	AMAZON FEE	-1	2011-01-05 09:57:00 UTC	16493.71	na1	United Kingdom
41	C540155	728028	OCEAN SCENT CANDLE IN JEW...	-64	2011-01-05 11:31:00 UTC	3.81	na1	Bahrain
42	C540266	M	Manual	-1	2011-01-06 11:05:00 UTC	459.29	na1	United Kingdom
43	C540266	POST	POSTAGE	-1	2011-01-06 11:05:00 UTC	35.09	na1	United Kingdom
44	C540559	21888	BINGO SET	-4	2011-01-10 10:07:00 UTC	3.75	na1	ERE
45	C540834	22461	SAVOY ART DEC'D CLOCK	-1	2011-01-12 09:34:00 UTC	12.75	na1	United Kingdom
46	C541492	37500	TEA TIME TEAPOT IN GIFT BOX	-1	2011-01-18 14:24:00 UTC	9.95	na1	United Kingdom
47	C541492	85174	S/4 CACTI CANDLES	-1	2011-01-18 14:24:00 UTC	4.95	na1	United Kingdom
48	C541492	848708	BLUR GRISIA GRL	-1	2011-01-18 14:24:00 UTC	3.75	na1	United Kingdom
49	C541492	85039A	SET 4 RED MINI ROSE CANDLE ...	-1	2011-01-18 14:24:00 UTC	1.65	na1	United Kingdom
50	C541492	85040A	S/4 PINK FLOWER CANDLES IN...	-1	2011-01-18 14:24:00 UTC	1.65	na1	United Kingdom

- 구매 건 상태가 **Canceled** 인 데이터의 비율(%) - 소수점 첫번째 자리까지

```
SELECT
  ROUND(
    100 * SUM(CASE WHEN InvoiceNo LIKE 'C%' THEN 1 ELSE 0 END) / COUNT(*),
    1
  ) AS canceled_ratio_percent
FROM `artful-striker-473603-f9.modulabs_project.data`;
```

[결과 이미지를 넣어주세요]

행	canceled_ratio_p...
1	1.7

## StockCode 살펴보기

- 고유한 **StockCode** 의 개수를 출력하기

```
SELECT COUNT(DISTINCT StockCode) AS unique_stockcode_count
FROM `artful-striker-473603-f9.modulabs_project.data`;
```

[결과 이미지를 넣어주세요]

행	unique_stockcod...
1	4069

- 어떤 제품이 가장 많이 판매되었는지 보기 위하여 **StockCode** 별 등장 빈도를 출력하기
  - 상위 10개의 제품들을 출력하기

```
SELECT StockCode, COUNT(*) AS sell_cnt
FROM `artful-striker-473603-f9.modulabs_project.data`
GROUP BY StockCode
```

```
ORDER BY sell_cnt DESC
LIMIT 10;
```

[결과 이미지를 넣어주세요]

행	StockCode	sell_cnt
1	22423	2192
2	85099B	2156
3	47566	1720
4	20725	1626
5	84879	1489
6	22720	1469
7	22197	1468
8	21212	1367
9	22383	1328
10	20727	1323

- **StockCode**의 컬럼에 있던 값 중에서 숫자를 제외한 문자만 남기고 문자가 몇 자리 수 인지 세고
  - 숫자가 0~1개인 값들에는 어떤 코드들이 들어가 있는지 출력하기

```
SELECT DISTINCT StockCode, number_count
FROM (
  SELECT StockCode,
    LENGTH(StockCode) - LENGTH(REGEXP_REPLACE(StockCode, r'[0-9]', '')) AS number_count
  FROM `artful-striker-473603-f9.modulabs_project.data`
)
WHERE number_count >= LENGTH(StockCode) - 1; -- 숫자가 0~1개인 경우
```

[결과 이미지를 넣어주세요]

행	StockCode	number_count
1	22139	5
2	22153	5
3	22966	5
4	22699	5
5	20733	5
6	21823	5
7	21912	5
8	20973	5
9	22064	5
10	35818P	5
11	22560	5
12	47559B	5
13	17012B	5
14	21787	5
15	22517	5
16	22082	5
17	18098C	5
18	22161	5
19	22090	5
20	22822	5
21	22276	5
22	22371	5
23	90116	5
24	21559	5
25	22910	5



26	21507	5
27	22603	5
28	22432	5
29	22847	5
30	22945	5
31	840296	5
32	21213	5
33	21169	5
34	21871	5
35	20840	5
36	22149	5
37	22582	5
38	22758	5
39	79321	5
40	21822	5
41	84598	5
42	21991	5
43	22189	5
44	22224	5
45	22043	5
46	22134	5
47	35471D	5
48	22071	5
49	22835	5
50	22962	5

- **StockCode**의 컬럼에 있던 값 중에서 숫자를 제외한 문자만 남기고 문자가 몇 자리 수 인지 세고
  - 숫자가 0~1개인 값들을 가지고 있는 데이터 수는 전체 데이터 수 대비 몇 퍼센트인지 구하기 (소수점 두 번째 자리까지)

```
SELECT
  ROUND(
    100 * SUM(CASE WHEN LENGTH(StockCode) - LENGTH(REGEXP_REPLACE(StockCode, r'[0-9]', '')) <= 1
      THEN 1 ELSE 0 END)
    / COUNT(*), 2
  ) AS zero_one_number_ratio_percent
FROM `artful-striker-473603-f9.modulabs_project.data`;
```

[결과 이미지를 넣어주세요]

행	zero_one_number...
1	0.55

- 제품과 관련되지 않은 거래 기록을 제거하기

```
DELETE FROM `artful-striker-473603-f9.modulabs_project.data`
WHERE StockCode IN (
  SELECT DISTINCT StockCode
  FROM (
    SELECT StockCode
    FROM `artful-striker-473603-f9.modulabs_project.data`
    WHERE LENGTH(REGEXP_REPLACE(StockCode, r'[0-9]', '')) = LENGTH(StockCode) -- 숫자가 전혀 없는 코드
  )
);
```

[결과 이미지를 넣어주세요]

이 문으로 data의 행 2,790개가 삭제되었습니다.

## Description 살펴보기

- 고유한 Description 별 출현 빈도를 계산하고 상위 30개를 출력하기

```
SELECT Description, COUNT(*) AS description_cnt
FROM `artful-striker-473603-f9.modulabs_project.data`
GROUP BY Description
ORDER BY description_cnt DESC
LIMIT 30;
```

[결과 이미지를 넣어주세요]

행	Description	description_cnt
1	REGENCY CAKESTAND 3 TIER	2189
2	JUMBO BAG RED RETROSPOT	2156
3	PARTY BUNTING	1720
4	LUNCH BAG RED RETROSPOT	1625
5	ASSORTED COLOUR BIRD ORN...	1488
6	SET OF 3 CAKE TINS PANTRY D...	1465
7	null	1449
8	PACK OF 72 RETROSPOT CAKE ...	1367
9	LUNCH BAG BLACK SKULL	1323
10	NATURAL SLATE HEART CHAL...	1272
11	JUMBO BAG PINK POLKADOT	1245
12	HEART OF WICKER SMALL	1230
13	JAM MAKING SET WITH JARS	1221
14	JUMBO STORAGE BAG SUKI	1211
15	PAPER CHAIN KIT 50'S CHRIST...	1194
16	JUMBO SHOPPER VINTAGE RE...	1192
17	LUNCH BAG CARS BLUE	1185
18	JAM MAKING SET PRINTED	1177
19	LUNCH BAG SPACEBOY DESIGN	1177
20	RECIPE BOX PANTRY YELLOW ...	1173
21	SPOTTY BUNTING	1168
22	ROSES REGENCY TEACUP AND ...	1128
23	WOODEN PICTURE FRAME WHI...	1124
24	LUNCH BAG PINK POLKADOT	1121
25	LUNCH BAG SUKI DESIGN	1121
26	SET OF 4 PANTRY JELLY MOUL...	1099
27	ALARM CLOCK BAKELIKE RED	1092
28	GREEN REGENCY TEACUP AND...	1066
29	VICTORIAN GLASS HANGING T...	1066
30	LUNCH BAG APPLE DESIGN	1066

- 서비스 관련 정보를 포함하는 행들을 제거하기

```
SELECT DISTINCT Description
FROM `artful-striker-473603-f9.modulabs_project.data`
WHERE UPPER(Description) LIKE '%TEST%'
OR UPPER(Description) LIKE '%ADJUST%'
OR UPPER(Description) LIKE '%SAMPLE%'
OR UPPER(Description) LIKE '%POSTAGE%';
```

[결과 이미지를 넣어주세요]

행	Description ▼
1	adjustment
2	adjust
3	samples
4	test
5	taig adjust
6	OOPS I adjustment
7	amazon adjust
8	reverse 21/5/10 adjustment
9	samples/damages
10	Adjustment
11	reverse previous adjustment
12	Show Samples
13	Damages/samples
14	taig adjust no stock
15	temp adjustment
16	Amazon Adjustment
17	re-adjustment
18	dotcom adjust

- 대소문자를 혼합하고 있는 데이터를 대문자로 표준화 하기

```
CREATE OR REPLACE TABLE `artful-striker-473603-f9.modulabs_project.data` AS
SELECT
  * EXCEPT (Description),
  UPPER(Description) AS Description
FROM `artful-striker-473603-f9.modulabs_project.data`;

SELECT Description
FROM `artful-striker-473603-f9.modulabs_project.data`
LIMIT 100;
```

[결과 이미지를 넣어주세요]

행	Description ▼
1	<i>null</i>
2	PAPER CHAIN KIT RETROSPOT
3	60 CAKE CASES VINTAGE CHRI...
4	PIZZA PLATE IN BOX
5	CHILDRENS APRON SPACEBOY...
6	BOX OF 24 COCKTAIL PARASOLS
7	RED WHITE SCARF HOT WATE...
8	WHITE SKULL HOT WATER BOT...
9	15CM CHRISTMAS GLASS BAL...
10	SET 2 TEA TOWELS I LOVE LON...
11	HOT WATER BOTTLE I AM SO P...
12	JAM MAKING SET PRINTED
13	AIRLINE BAG VINTAGE JET SET...
14	HAND WARMER UNION JACK
15	TEA COSY RED STRIPE
16	TOY TIDY SPACEBOY
17	FAMILY ALBUM WHITE PICTUR...
18	ORIGAMI VANILLA INCENSE/C...
19	WHITE WOOD GARDEN PLANT ...
20	TRADITIONAL WOODEN CATCH...
21	HERB MARKER MINT
22	CHILDRENS DOLLY GIRL MUG
23	BLUE GLASS GEMS IN BAG
24	DOORMAT WELCOME PUPPIES
25	OFFICE MUG WARMER CHOC+...

26	DOORMAT RED RETROSPOT
27	BOYS ALPHABET IRON ON PAT...
28	COFFEE MUG DOG + BALL DESI...
29	BASKET OF FLOWERS SEWING ...
30	GLITTER STAR GARLAND WITH...
31	BINGO SET
32	EMPIRE UNION JACK TV DINNE...
33	METAL SIGN,CUPCAKE SINGLE ...
34	SPOTTY PINK DUCK DOORSTOP
35	GOLD MINI TAPE MEASURE
36	DAISY GARDEN MARKER
37	CHRISTMAS CRAFT TREE TOP ...
38	ACRYLIC JEWEL ICICLE, PINK
39	ALARM CLOCK BAKELIKE IVORY
40	RIBBON REEL FLORA + FAUNA
41	SET OF 2 TINS VINTAGE BATHR...
42	PINK NEW BAROQUECANDLES...
43	BEADED CRYSTAL HEART PINK...
44	MINI JIGSAW CIRCUS PARADE
45	BLUE DRAGONFLY HELICOPTER
46	HOMEMADE JAM SCENTED CA...
47	SEWING SUSAN 21 NEEDLE SET
48	HEARTS GIFT TAPE
49	ASSORTED TUTTI FRUTTI KEYR...
50	12 PENCILS TALL TUBE WOOD...

## UnitPrice 살펴보기

- UnitPrice 의 최솟값, 최댓값, 평균을 구하기

```
SELECT
  MIN(UnitPrice) AS min_price,
  MAX(UnitPrice) AS max_price,
  ROUND(AVG(UnitPrice), 2) AS avg_price
FROM `artful-striker-473603-f9.modulabs_project.data`;
```

[결과 이미지를 넣어주세요]

행	min_price	max_price	avg_price
1	0.0	649.5	3.29

- 단가가 0원인 거래의 개수, 구매 수량(Quantity)의 최솟값, 최댓값, 평균 구하기

```
SELECT
  COUNT(*) AS cnt_quantity,
  MIN(Quantity) AS min_quantity,
  MAX(Quantity) AS max_quantity,
  ROUND(AVG(Quantity), 2) AS avg_quantity
FROM `artful-striker-473603-f9.modulabs_project.data`
WHERE UnitPrice = 0;
```

[결과 이미지를 넣어주세요]

행	cnt_quantity	min_quantity	max_quantity	avg_quantity
1	2490	-9600	12540	-57.22

- UnitPrice = 0 를 제거하고 일관된 데이터셋을 유지하기

```
CREATE OR REPLACE TABLE `artful-striker-473603-f9.modulabs_project.data` AS
SELECT *
FROM `artful-striker-473603-f9.modulabs_project.data`
WHERE UnitPrice <> 0;
```

```
SELECT UnitPrice, Quantity, StockCode, Description
FROM `artful-striker-473603-f9.modulabs_project.data`
LIMIT 100;
```

[결과 이미지를 넣어주세요]

행	UnitPrice	Quantity	StockCode	Description
1	2.98	1	22949	36 DOILIES DOLLY GIRL
2	1.49	1	84535B	FAIRY CAKES NOTEBOOK A6 SI...
3	14.43	1	20685	DOORMAT RED RETROSPOT
4	5.91	1	22470	HEART OF WICKER LARGE
5	0.85	4	22419	LIPSTICK PEN RED
6	4.21	1	22182	CAKE STAND VICTORIAN FILIG...
7	3.36	4	22198	LARGE POPCORN HOLDER
8	3.36	2	22174	PHOTO CUBE
9	2.51	3	17012A	ORIGAMI VANILLA INCENSE/C...
10	14.43	1	21523	DOORMAT FANCY FONT HOME ...
11	5.91	2	21844	RED RETROSPOT MUG
12	0.85	4	22539	MINI JIGSAW DOLLY GIRL
13	0.85	1	21703	BAG 125G SWIRLY MARBLES
14	4.21	1	22876	NUMBER TILE COTTAGE GARD...
15	2.51	2	22095	LADS ONLY TISSUE BOX
16	2.98	1	21210	SET OF 72 RETROSPOT PAPER ...
17	0.85	2	85131B	BEADED CRYSTAL HEART GREE...
18	7.62	5	21479	WHITE SKULL HOT WATER BOT...
19	1.28	1	21975	PACK OF 60 DINOSAUR CAKE C...
20	8.47	2	85132A	CHARLIE + LOLA BISCUITS TINS
21	3.36	1	21874	GIN AND TONIC MUG
22	2.51	1	90116	FRUIT SALAD BAG CHARM
23	1.66	1	85177	BASKET OF FLOWERS SEWING ...
24	2.51	1	82613B	METAL SIGN,CUPCAKE SINGLE ...
25	10.17	2	85127	SMALL SQUARE CUT GLASS CA...
26	0.85	1	22153	ANGEL DECORATION STARS O...
27	1.26	1	90071	RUBY GLASS CLUSTER FARRIN...
28	7.62	2	22550	HOLIDAY FUN LUDO
29	3.36	2	72799E	IVORY PILLAR CANDLE SILVER ...
30	8.47	3	22114	HOT WATER BOTTLE TEA AND ...
31	11.87	2	21662	VINTAGE GLASS COFFEE CADDY
32	0.85	1	90214D	LETTER 'D' BLING KEY RING
33	5.91	1	22314	OFFICE MUG WARMER CHOC+...
34	3.36	3	22077	6 RIBBONS RUSTIC CHARM
35	3.36	1	21935	SUKI SHOULDER BAG
36	5.91	4	22910	PAPER CHAIN KIT VINTAGE CH...
37	1.69	8	71459	HANGING JAM JAR T-LIGHT HO...
38	12.72	1	20902	VINTAGE KEEPSAKE BOX PARL...
39	2.51	1	21949	SET OF 6 STRAWBERRY CHOPS...
40	0.85	8	22531	MAGIC DRAWING SLATE CIRCU...
41	5.06	3	22588	CARD HOLDER GINGHAM HEART
42	1.65	6	20854	BLUE PATCH PURSE PINK HEART
43	5.91	2	22798	ANTIQUE GLASS DRESSING TA...
44	0.85	1	84692	BOX OF 24 COCKTAIL PARASOLS
45	4.21	1	22380	TOY TIDY SPACEBOY
46	8.47	1	21218	RED SPOTTY BISCUIT TIN
47	2.51	3	22988	SOLDIERS EGG CUP
48	12.72	40	22423	REGENCY CAKESTAND 3 TIER
49	1.66	2	22196	SMALL HEART MEASURING SP...
50	1.66	1	21884	CAKES AND BOWS GIFT TAPE

## 11-7. RFM 스코어

### Recency

- **InvoiceDate** 컬럼을 연월일 자료형으로 변경하기

```
SELECT
DATE(InvoiceDate) AS InvoiceDay,
*
FROM `artful-striker-473603-f9.modulabs_project.data`
LIMIT 100;
```

[결과 이미지를 넣어주세요]

일	InvoiceDay	InvoiceNo	StockCode	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	Description
1	2010-12-01	536544	22949	1	2010-12-01 14:32:00 UTC	2.98	no	United Kingdom	36 DOLLES DOLLY GIRL
2	2010-12-01	536544	845358	1	2010-12-01 14:32:00 UTC	1.49	no	United Kingdom	FAIRY CAKES NOTEBOOK AS SL...
3	2010-12-01	536544	20685	1	2010-12-01 14:32:00 UTC	14.43	no	United Kingdom	DOORMAT RED RETROSPOT
4	2010-12-01	536544	22470	1	2010-12-01 14:32:00 UTC	5.91	no	United Kingdom	HEART OF WICKER LARGE
5	2010-12-01	536544	22419	4	2010-12-01 14:32:00 UTC	0.85	no	United Kingdom	LIPSTICK PEN RED
6	2010-12-01	536544	22182	1	2010-12-01 14:32:00 UTC	4.21	no	United Kingdom	CAKE STAND VICTORIAN FILIG...
7	2010-12-01	536544	22189	4	2010-12-01 14:32:00 UTC	3.36	no	United Kingdom	LARGE POPCORN HOLDER
8	2010-12-01	536544	20174	3	2010-12-01 14:32:00 UTC	3.36	no	United Kingdom	PHOTO CLIP
9	2010-12-01	536544	17012A	3	2010-12-01 14:32:00 UTC	2.51	no	United Kingdom	ORGANIC VANILLA ICENSENCE...
10	2010-12-01	536544	21923	1	2010-12-01 14:32:00 UTC	14.43	no	United Kingdom	DOORMAT FANCY FONT HOME...
11	2010-12-01	536544	21844	2	2010-12-01 14:32:00 UTC	5.91	no	United Kingdom	RED RETROSPOT MUG
12	2010-12-01	536544	22539	4	2010-12-01 14:32:00 UTC	0.85	no	United Kingdom	MINI JESSAM DOLLY GIRL
13	2010-12-01	536544	21703	1	2010-12-01 14:32:00 UTC	0.85	no	United Kingdom	BAG 12SG SWIRLY MARBLES
14	2010-12-01	536544	22875	1	2010-12-01 14:32:00 UTC	4.21	no	United Kingdom	NUMBER TILE COTTAGE GARD...
15	2010-12-01	536544	22395	2	2010-12-01 14:32:00 UTC	2.51	no	United Kingdom	LADS ONLY TISSUE BOX
16	2010-12-01	536544	21210	1	2010-12-01 14:32:00 UTC	2.98	no	United Kingdom	SET OF 72 RETROSPOT PAPERS
17	2010-12-01	536544	851218	2	2010-12-01 14:32:00 UTC	0.85	no	United Kingdom	BEADED CRYSTAL HEART ORN...
18	2010-12-01	536544	27479	5	2010-12-01 14:32:00 UTC	7.62	no	United Kingdom	WHITE SKULL HOT WATER BOT...
19	2010-12-01	536544	21975	1	2010-12-01 14:32:00 UTC	1.28	no	United Kingdom	PACK OF 60 DINOSAUR CAKE C...
20	2010-12-01	536544	85132A	2	2010-12-01 14:32:00 UTC	8.47	no	United Kingdom	CHARLIE + LOLA BISCUITS TINS
21	2010-12-01	536544	21874	1	2010-12-01 14:32:00 UTC	3.36	no	United Kingdom	GIN AND TONIC MUG
22	2010-12-01	536544	90115	1	2010-12-01 14:32:00 UTC	2.51	no	United Kingdom	FRUIT SALAD BAG CHARM
23	2010-12-01	536544	85177	1	2010-12-01 14:32:00 UTC	1.66	no	United Kingdom	BASKET OF FLOWERS SEWING ...
24	2010-12-01	536544	826138	1	2010-12-01 14:32:00 UTC	2.51	no	United Kingdom	METAL SIGN/CUPCAKE SINGLE ...
25	2010-12-01	536544	85127	2	2010-12-01 14:32:00 UTC	10.17	no	United Kingdom	SMALL SQUARE CUT GLASS CA...

26	2010-12-01	536544	22153	1	2010-12-01 14:32:00 UTC	0.85	no	United Kingdom	ANGEL DECORATION STARS O...
27	2010-12-01	536544	90071	1	2010-12-01 14:32:00 UTC	1.26	no	United Kingdom	RUBY GLASS CLUSTER EARRIN...
28	2010-12-01	536544	22550	2	2010-12-01 14:32:00 UTC	7.62	no	United Kingdom	HOLIDAY FUN LUGG
29	2010-12-01	536544	72796	2	2010-12-01 14:32:00 UTC	3.36	no	United Kingdom	IVORY PILLAR CANDLE SILVER...
30	2010-12-01	536544	22114	3	2010-12-01 14:32:00 UTC	8.47	no	United Kingdom	HOT WATER BOTTLE TEA AND...
31	2010-12-01	536544	21842	2	2010-12-01 14:32:00 UTC	11.87	no	United Kingdom	VINTAGE BLACK COFFEE CADDY
32	2010-12-01	536544	90140	1	2010-12-01 14:32:00 UTC	0.85	no	United Kingdom	LETTER 'O' SLING KEY RING
33	2010-12-01	536544	22314	1	2010-12-01 14:32:00 UTC	5.91	no	United Kingdom	OFFICE MUG WARMER CHECK...
34	2010-12-01	536544	22077	3	2010-12-01 14:32:00 UTC	3.36	no	United Kingdom	6 REBORN RUSTIC CHARM
35	2010-12-01	536544	21885	1	2010-12-01 14:32:00 UTC	3.36	no	United Kingdom	SLAM SHOULDER BAG
36	2010-12-01	536544	22910	4	2010-12-01 14:32:00 UTC	5.91	no	United Kingdom	PAPER CHAIN KIT VINTAGE CH...
37	2010-12-01	536544	71499	8	2010-12-01 14:32:00 UTC	1.69	no	United Kingdom	HANGING JAM JAR 14.5CM HD...
38	2010-12-01	536544	20052	1	2010-12-01 14:32:00 UTC	12.72	no	United Kingdom	VINTAGE KEPSAKE BOX PAIR...
39	2010-12-01	536544	21949	1	2010-12-01 14:32:00 UTC	2.51	no	United Kingdom	SET OF 6 STRAWBERRY CAKES...
40	2010-12-01	536544	22523	8	2010-12-01 14:32:00 UTC	0.85	no	United Kingdom	MASCOT BEARING SLEET CIGAR...
41	2010-12-01	536544	22588	3	2010-12-01 14:32:00 UTC	5.06	no	United Kingdom	CARE HOLDER GINGHAM HEART
42	2010-12-01	536544	20684	6	2010-12-01 14:32:00 UTC	7.65	no	United Kingdom	BLUE PATCH PURSE PINK HEAT...
43	2010-12-01	536544	22798	2	2010-12-01 14:32:00 UTC	5.91	no	United Kingdom	ANTIQUE GLASS DRESSING TA...
44	2010-12-01	536544	84992	1	2010-12-01 14:32:00 UTC	0.85	no	United Kingdom	BOX OF 24 COCKTAIL PARASOL...
45	2010-12-01	536544	22380	1	2010-12-01 14:32:00 UTC	4.21	no	United Kingdom	TOY TREN SPACEROY
46	2010-12-01	536544	21218	1	2010-12-01 14:32:00 UTC	8.47	no	United Kingdom	RED SPOTTY BISCUIT TIN
47	2010-12-01	536544	22989	3	2010-12-01 14:32:00 UTC	2.51	no	United Kingdom	SOLIDPERS CUP
48	2010-12-01	536544	22623	40	2010-12-01 14:32:00 UTC	12.72	no	United Kingdom	REGENCY CAKESTAND 3 TIER
49	2010-12-01	536544	22176	2	2010-12-01 14:32:00 UTC	1.66	no	United Kingdom	SMALL HEART MEASURING SP...
50	2010-12-01	536544	21884	1	2010-12-01 14:32:00 UTC	7.66	no	United Kingdom	CAKES AND BOWS GIFT TAPE

- 가장 최근 구매 일자를 MAX() 함수로 찾아보기

```
SELECT
(SELECT MAX DATE(InvoiceDate)) FROM `artful-striker-473603-f9.modulabs_project.data`) AS most_recent_date,
DATE(InvoiceDate) AS InvoiceDay,
*
FROM `artful-striker-473603-f9.modulabs_project.data`
LIMIT 100;
```

[결과 이미지를 넣어주세요]

일	most_recent_date	InvoiceDay	InvoiceNo	StockCode	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	Description
1	2011-12-09	2010-12-01	536544	22949	1	2010-12-01 14:32:00 UTC	2.98	no	United Kingdom	36 DOLLES DOLLY GIRL
2	2011-12-09	2010-12-01	536544	845358	1	2010-12-01 14:32:00 UTC	1.49	no	United Kingdom	FAIRY CAKES NOTEBOOK AS SL...
3	2011-12-09	2010-12-01	536544	20685	1	2010-12-01 14:32:00 UTC	14.43	no	United Kingdom	DOORMAT RED RETROSPOT
4	2011-12-09	2010-12-01	536544	22470	1	2010-12-01 14:32:00 UTC	5.91	no	United Kingdom	HEART OF WICKER LARGE
5	2011-12-09	2010-12-01	536544	22419	4	2010-12-01 14:32:00 UTC	0.85	no	United Kingdom	LIPSTICK PEN RED
6	2011-12-09	2010-12-01	536544	22182	1	2010-12-01 14:32:00 UTC	4.21	no	United Kingdom	CAKE STAND VICTORIAN FILIG...
7	2011-12-09	2010-12-01	536544	22189	4	2010-12-01 14:32:00 UTC	3.36	no	United Kingdom	LARGE POPCORN HOLDER
8	2011-12-09	2010-12-01	536544	20174	3	2010-12-01 14:32:00 UTC	3.36	no	United Kingdom	PHOTO CLIP
9	2011-12-09	2010-12-01	536544	17012A	3	2010-12-01 14:32:00 UTC	2.51	no	United Kingdom	ORGANIC VANILLA ICENSENCE...
10	2011-12-09	2010-12-01	536544	21923	1	2010-12-01 14:32:00 UTC	14.43	no	United Kingdom	DOORMAT FANCY FONT HOME...
11	2011-12-09	2010-12-01	536544	21844	2	2010-12-01 14:32:00 UTC	5.91	no	United Kingdom	RED RETROSPOT MUG
12	2011-12-09	2010-12-01	536544	22539	4	2010-12-01 14:32:00 UTC	0.85	no	United Kingdom	MINI JESSAM DOLLY GIRL
13	2011-12-09	2010-12-01	536544	21703	1	2010-12-01 14:32:00 UTC	0.85	no	United Kingdom	BAG 12SG SWIRLY MARBLES
14	2011-12-09	2010-12-01	536544	22875	1	2010-12-01 14:32:00 UTC	4.21	no	United Kingdom	NUMBER TILE COTTAGE GARD...
15	2011-12-09	2010-12-01	536544	22395	2	2010-12-01 14:32:00 UTC	2.51	no	United Kingdom	LADS ONLY TISSUE BOX
16	2011-12-09	2010-12-01	536544	21210	1	2010-12-01 14:32:00 UTC	2.98	no	United Kingdom	SET OF 72 RETROSPOT PAPERS
17	2011-12-09	2010-12-01	536544	851218	2	2010-12-01 14:32:00 UTC	0.85	no	United Kingdom	BEADED CRYSTAL HEART ORN...
18	2011-12-09	2010-12-01	536544	27479	5	2010-12-01 14:32:00 UTC	7.62	no	United Kingdom	WHITE SKULL HOT WATER BOT...
19	2011-12-09	2010-12-01	536544	21975	1	2010-12-01 14:32:00 UTC	1.28	no	United Kingdom	PACK OF 60 DINOSAUR CAKE C...
20	2011-12-09	2010-12-01	536544	85132A	2	2010-12-01 14:32:00 UTC	8.47	no	United Kingdom	CHARLIE + LOLA BISCUITS TINS
21	2011-12-09	2010-12-01	536544	21874	1	2010-12-01 14:32:00 UTC	3.36	no	United Kingdom	GIN AND TONIC MUG
22	2011-12-09	2010-12-01	536544	90115	1	2010-12-01 14:32:00 UTC	2.51	no	United Kingdom	FRUIT SALAD BAG CHARM
23	2011-12-09	2010-12-01	536544	85177	1	2010-12-01 14:32:00 UTC	1.66	no	United Kingdom	BASKET OF FLOWERS SEWING ...
24	2011-12-09	2010-12-01	536544	826138	1	2010-12-01 14:32:00 UTC	2.51	no	United Kingdom	METAL SIGN/CUPCAKE SINGLE ...
25	2011-12-09	2010-12-01	536544	85127	2	2010-12-01 14:32:00 UTC	10.17	no	United Kingdom	SMALL SQUARE CUT GLASS CA...

26	2011-12-09	2010-12-01	536544	22193		1	2010-12-01 14:32:00 UTC	9.85	🇬🇧	United Kingdom	ANGEL DECORATION STARS G...
27	2011-12-09	2010-12-01	536544	90371		1	2010-12-01 14:32:00 UTC	1.28	🇬🇧	United Kingdom	RUBY GLASS CLUSTER EARRIN...
28	2011-12-09	2010-12-01	536544	22593		2	2010-12-01 14:32:00 UTC	7.42	🇬🇧	United Kingdom	HOLIDAY PAIL LUGO
29	2011-12-09	2010-12-01	536544	75796		2	2010-12-01 14:32:00 UTC	9.36	🇬🇧	United Kingdom	WOM PILLAR CANDLE SILVER ...
30	2011-12-09	2010-12-01	536544	22114		3	2010-12-01 14:32:00 UTC	8.47	🇬🇧	United Kingdom	HOT WATER BOTTLE TEA AND ...
31	2011-12-09	2010-12-01	536544	21462		2	2010-12-01 14:32:00 UTC	11.87	🇬🇧	United Kingdom	VINTAGE BLASS COFFEE CADDY
32	2011-12-09	2010-12-01	536544	902140		1	2010-12-01 14:32:00 UTC	9.85	🇬🇧	United Kingdom	LETTER "D" BLING KEY RING
33	2011-12-09	2010-12-01	536544	22214		1	2010-12-01 14:32:00 UTC	9.91	🇬🇧	United Kingdom	OFFICE MAG WARMER CHOCY...
34	2011-12-09	2010-12-01	536544	22577		9	2010-12-01 14:32:00 UTC	9.36	🇬🇧	United Kingdom	4-WHEELS AUTO CYCLES
35	2011-12-09	2010-12-01	536544	21885		1	2010-12-01 14:32:00 UTC	9.36	🇬🇧	United Kingdom	SUMI SHOULDER BAG
36	2011-12-09	2010-12-01	536544	22910		4	2010-12-01 14:32:00 UTC	9.91	🇬🇧	United Kingdom	PAPER CHAIN KIT VINTAGE CH...
37	2011-12-09	2010-12-01	536544	71499		8	2010-12-01 14:32:00 UTC	1.89	🇬🇧	United Kingdom	HANUKING JAM JAR LIGHT HO...
38	2011-12-09	2010-12-01	536544	20902		1	2010-12-01 14:32:00 UTC	12.72	🇬🇧	United Kingdom	VINTAGE KEEPSAKE BOX PAR...
39	2011-12-09	2010-12-01	536544	21149		1	2010-12-01 14:32:00 UTC	2.51	🇬🇧	United Kingdom	SET OF 8 SHAWARMOP CYCLES...
40	2011-12-09	2010-12-01	536544	22521		8	2010-12-01 14:32:00 UTC	9.85	🇬🇧	United Kingdom	MAGIC DRAWING SLATE CIRCUL...
41	2011-12-09	2010-12-01	536544	22588		3	2010-12-01 14:32:00 UTC	9.96	🇬🇧	United Kingdom	CARD HOLDER GINGHAM HEART
42	2011-12-09	2010-12-01	536544	30654		6	2010-12-01 14:32:00 UTC	1.68	🇬🇧	United Kingdom	BLUE PATCH PURSE PINK HEART
43	2011-12-09	2010-12-01	536544	22798		2	2010-12-01 14:32:00 UTC	9.91	🇬🇧	United Kingdom	ANTIQUE GLASS DRESSING TAL...
44	2011-12-09	2010-12-01	536544	94902		1	2010-12-01 14:32:00 UTC	9.85	🇬🇧	United Kingdom	BOX OF 24 COCKTAIL PARASOLS
45	2011-12-09	2010-12-01	536544	22389		1	2010-12-01 14:32:00 UTC	4.21	🇬🇧	United Kingdom	TOY TONY SPICEBOY
46	2011-12-09	2010-12-01	536544	21218		1	2010-12-01 14:32:00 UTC	8.47	🇬🇧	United Kingdom	RED SPOTTY BISCUIT TIN
47	2011-12-09	2010-12-01	536544	22088		3	2010-12-01 14:32:00 UTC	2.51	🇬🇧	United Kingdom	SOLDIERS EGG CUP
48	2011-12-09	2010-12-01	536544	22423		40	2010-12-01 14:32:00 UTC	12.72	🇬🇧	United Kingdom	REGENCY CANDLESTAND 3 TIER
49	2011-12-09	2010-12-01	536544	22196		2	2010-12-01 14:32:00 UTC	1.66	🇬🇧	United Kingdom	SMALL HEART MEASURING SP...
50	2011-12-09	2010-12-01	536544	21884		1	2010-12-01 14:32:00 UTC	1.66	🇬🇧	United Kingdom	CARDS AND BOWS GIFT TAPE

- 유저 별로 가장 큰 InvoiceDay를 찾아서 가장 최근 구매일로 저장하기

```
SELECT
    CustomerID,
    MAX(DATE(InvoiceDate)) AS InvoiceDay
FROM `artful-striker-473603-f9.modulabs_project.data`
GROUP BY CustomerID;
```

[결과 이미지를 넣어주세요]

행	CustomerID	InvoiceDay
1	null	2011-12-09
2	12346	2011-01-18
3	12347	2011-12-07
4	12348	2011-09-25
5	12349	2011-11-21
6	12350	2011-02-02
7	12352	2011-11-03
8	12353	2011-05-19
9	12354	2011-04-21
10	12355	2011-05-09
11	12356	2011-11-17
12	12357	2011-11-06
13	12358	2011-12-08
14	12359	2011-12-02
15	12360	2011-10-18
16	12361	2011-02-25
17	12362	2011-12-06
18	12363	2011-08-22
19	12364	2011-12-02
20	12365	2011-02-21
21	12367	2011-12-05
22	12370	2011-10-19
23	12371	2011-10-11
24	12372	2011-09-29
25	12373	2011-02-01

26	12374	2011-11-14
27	12375	2011-12-07
28	12377	2011-01-28
29	12378	2011-08-02
30	12379	2011-09-19
31	12380	2011-11-18
32	12381	2011-12-05
33	12383	2011-06-08
34	12384	2011-11-11
35	12386	2011-01-06
36	12388	2011-11-24
37	12390	2011-09-21
38	12391	2011-11-18
39	12393	2011-09-28
40	12394	2011-10-07
41	12395	2011-11-24
42	12397	2011-11-04
43	12398	2011-10-25
44	12399	2011-08-12
45	12401	2011-02-09
46	12402	2011-01-20
47	12403	2011-10-21
48	12405	2011-07-14
49	12406	2011-11-17
50	12407	2011-10-21

- 가장 최근 일자( **most\_recent\_date** )와 유저별 마지막 구매일( **InvoiceDay** )간의 차이를 계산하기

```

SELECT
  CustomerID,
  EXTRACT(DAY FROM MAX(InvoiceDay) OVER () - InvoiceDay) AS recency
FROM (
  SELECT
    CustomerID,
    MAX(DATE(InvoiceDate)) AS InvoiceDay
  FROM `artful-striker-473603-f9.modulabs_project.data`
  GROUP BY CustomerID
);

```

[결과 이미지를 넣어주세요]



행	CustomerID	recency
1	12371	59
2	12650	3
3	12704	7
4	12740	63
5	12798	10
6	13243	203
7	13385	329
8	13466	100
9	13682	187
10	13803	255
11	14293	3
12	14314	106
13	14705	198
14	14739	2
15	15319	283
16	15454	362
17	15551	36
18	15728	49
19	15730	65
20	15755	1
21	15910	0
22	16024	12
23	16150	38
24	16240	8
25	16262	360

26	16812	360
27	16913	290
28	16940	52
29	16969	39
30	17325	280
31	17708	187
32	17711	10
33	17877	170
34	17888	18
35	17892	56
36	18014	28
37	18217	29
38	12406	22
39	12488	9
40	12576	35
41	12577	35
42	12580	246
43	12683	4
44	12710	96
45	12714	9
46	12817	281
47	12845	267
48	13050	9
49	13137	10
50	13474	236

- 최종 데이터 셋에 필요한 데이터들을 각각 정제해서 이어붙이고 지금까지의 결과를 **user\_r** 이라는 이름의 테이블로 저장하기

```
CREATE OR REPLACE TABLE `artful-striker-473603-f9.modulabs_project.user_r` AS
WITH cleaned_data AS (
  SELECT *
  FROM `artful-striker-473603-f9.modulabs_project.data`
  WHERE UnitPrice <> 0
),
user_last_purchase AS (
  SELECT
    CustomerID,
    MAX(DATE(InvoiceDate)) AS InvoiceDay
  FROM cleaned_data
```

```
GROUP BY CustomerID
),
most_recent AS (
SELECT MAX(InvoiceDay) AS most_recent_date
FROM user_last_purchase
),
recency_calc AS (
SELECT
    u.CustomerID,
    u.InvoiceDay,
    EXTRACT(DAY FROM m.most_recent_date - u.InvoiceDay) AS recency
FROM user_last_purchase u
CROSS JOIN most_recent m
)
SELECT
    d.*,
    r.InvoiceDay,
    r.recency
FROM cleaned_data d
LEFT JOIN recency_calc r
ON d.CustomerID = r.CustomerID;
```

[결과 이미지를 넣어주세요]

△기타	세부정보	마리텔기	데이터 필터기	종계	계보	데이터 프로필	데이터 품질			
일	InvoiceNo	StockCode	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	Description	InvoiceDay	recency
1	530544	22077	3	2010-12-01 14:32:00 UTC	3.36	naid	United Kingdom	6 RIBBONS PLSTIC CHARM	naid	naid
2	530544	22045	1	2010-12-01 14:32:00 UTC	3.36	naid	United Kingdom	CHRISTMAS PUDDING TRINET	naid	naid
3	530544	22087	1	2010-12-01 14:32:00 UTC	9.92	naid	United Kingdom	HOT WATER BOTTLE BABUSHKA	naid	naid
4	530544	22318	1	2010-12-01 14:32:00 UTC	5.91	naid	United Kingdom	FIVE HEART HANGING DECORA...	naid	naid
5	530544	20719	2	2010-12-01 14:32:00 UTC	1.66	naid	United Kingdom	WOODLAND CHARLOTTE BAG	naid	naid
6	530544	22818	6	2010-12-01 14:32:00 UTC	0.85	naid	United Kingdom	CARD CHRISTMAS VILLAGE	naid	naid
7	530544	22100	1	2010-12-01 14:32:00 UTC	2.91	naid	United Kingdom	SKULLS SQUARE TISSUE BOX	naid	naid
8	530544	21913	2	2010-12-01 14:32:00 UTC	1.66	naid	United Kingdom	GLASS HEART FLIGHT HOLDER	naid	naid
9	530544	21957	1	2010-12-01 14:32:00 UTC	0.85	naid	United Kingdom	ELEPHANT BIRTHDAY CARD	naid	naid
10	530544	21577	2	2010-12-01 14:32:00 UTC	5.96	naid	United Kingdom	SAVE THE PLANET COTTON TO...	naid	naid
11	530544	22948	2	2010-12-01 14:32:00 UTC	6.77	naid	United Kingdom	METAL DECORATION NAUGHT...	naid	naid
12	530544	20770	4	2010-12-01 14:32:00 UTC	2.91	naid	United Kingdom	ABSTRACT CIRCLE JOURNAL	naid	naid
13	530544	22476	1	2010-12-01 14:32:00 UTC	11.02	naid	United Kingdom	EMPIRE UNION JACK TV DINNE...	naid	naid
14	530544	20975	1	2010-12-01 14:32:00 UTC	1.28	naid	United Kingdom	12 PENCILS SMALL TUBE RED ...	naid	naid
15	530544	22694	4	2010-12-01 14:32:00 UTC	4.21	naid	United Kingdom	WICKER STAR	naid	naid
16	530544	21758	1	2010-12-01 14:32:00 UTC	12.72	naid	United Kingdom	VINTAGE SHELLS PRINT	naid	naid
17	530544	22757	9	2010-12-01 14:32:00 UTC	2.91	naid	United Kingdom	LARGE RED BABUSHKA NOTIC...	naid	naid
18	530544	21564	2	2010-12-01 14:32:00 UTC	5.91	naid	United Kingdom	PINK HEART SHAPES LOVE BUC...	naid	naid
19	530544	21922	1	2010-12-01 14:32:00 UTC	4.21	naid	United Kingdom	GLITTER CHRISTMAS TREE WIT...	naid	naid
20	530544	840206	1	2010-12-01 14:32:00 UTC	7.62	naid	United Kingdom	RED WOOLLY HOTTE WHITE H...	naid	naid
21	530544	21165	1	2010-12-01 14:32:00 UTC	4.21	naid	United Kingdom	BEWARE OF THE CAT METAL SL...	naid	naid
22	530544	22135	5	2010-12-01 14:32:00 UTC	0.85	naid	United Kingdom	LADLE LOVE HEART PINK	naid	naid
23	530544	21914	3	2010-12-01 14:32:00 UTC	2.91	naid	United Kingdom	BLUE HARMONICA IN BOX	naid	naid
24	530544	22136	2	2010-12-01 14:32:00 UTC	1.66	naid	United Kingdom	SMALL HEART MEASURING SP...	naid	naid
25	530544	87669	1	2010-12-01 14:32:00 UTC	4.21	naid	United Kingdom	HOT BATHS METAL SIGN	naid	naid
26	530544	22296	1	2010-12-01 14:32:00 UTC	2.91	naid	United Kingdom	MAGNETS PACK OF 4 RETRO R...	naid	naid
27	530544	20856	6	2010-12-01 14:32:00 UTC	1.66	naid	United Kingdom	DEANNA PATCH PURSE PINK BL...	naid	naid
28	530544	83116	3	2010-12-01 14:32:00 UTC	1.66	naid	United Kingdom	BLACK CANDLEMABRA FLIGHT ...	naid	naid
29	530544	22224	1	2010-12-01 14:32:00 UTC	5.91	naid	United Kingdom	WHITE LOVERBIRD LANTERN	naid	naid
30	530544	22379	2	2010-12-01 14:32:00 UTC	4.21	naid	United Kingdom	RECYCLING BAG RETROSPOT	naid	naid
31	530544	22560	7	2010-12-01 14:32:00 UTC	6.47	naid	United Kingdom	JAM MAKING SET WITH JARS	naid	naid

## Frequency

- 고객마다 고유한 InvoiceNo의 수를 세어보기

```
SELECT
    CustomerID,
    COUNT(DISTINCT InvoiceNo) AS purchase_cnt
FROM `artful-striker-473603-f9.modulabs_project.data`
GROUP BY CustomerID;
```

[결과 이미지를 넣어주세요]

행	CustomerID	purchase_cnt
1	null	1410
2	12346	2
3	12347	7
4	12348	4
5	12349	1
6	12350	1
7	12352	8
8	12353	1
9	12354	1
10	12355	1
11	12356	3
12	12357	1
13	12358	2
14	12359	6
15	12360	3
16	12361	1
17	12362	13
18	12363	2
19	12364	4
20	12365	1
21	12367	1
22	12370	4
23	12371	1
24	12372	3
25	12373	1

26	12374	1
27	12375	3
28	12377	2
29	12378	1
30	12379	3
31	12380	5
32	12381	4
33	12383	6
34	12384	3
35	12386	2
36	12388	6
37	12390	1
38	12391	1
39	12393	4
40	12394	2
41	12395	15
42	12397	2
43	12398	1
44	12399	4
45	12401	1
46	12402	1
47	12403	1
48	12405	1
49	12406	3
50	12407	5

- 각 고객 별로 구매한 아이템의 총 수량 더하기

```
SELECT
  CustomerID,
  SUM(Quantity) AS total_quantity
FROM `artful-striker-473603-f9.modulabs_project.data`
GROUP BY CustomerID;
```

[결과 이미지를 넣어주세요]

행	CustomerID	total_quantity
1	null	416362
2	12346	0
3	12347	2458
4	12348	2332
5	12349	630
6	12350	196
7	12352	463
8	12353	20
9	12354	530
10	12355	240
11	12356	1573
12	12357	2708
13	12358	242
14	12359	1599
15	12360	1156
16	12361	90
17	12362	2180
18	12363	408
19	12364	1499
20	12365	173
21	12367	172
22	12370	2255
23	12371	582
24	12372	788
25	12373	196
26	12374	339
27	12375	175
28	12377	843

- 전체 거래 건수 계산과 구매한 아이템의 총 수량 계산의 결과를 합쳐서 `user_rf` 라는 이름의 테이블에 저장하기

```
CREATE OR REPLACE TABLE `artful-striker-473603-f9.modulabs_project.user_rf` AS
```

```
-- (1) 전체 거래 건수 계산
```

```
WITH purchase_cnt AS (
  SELECT
    CustomerID,
    COUNT(DISTINCT InvoiceNo) AS purchase_cnt
  FROM `artful-striker-473603-f9.modulabs_project.data`
  GROUP BY CustomerID
),
```

```
-- (2) 구매한 아이템 총 수량 계산
```

```
item_cnt AS (
  SELECT
    CustomerID,
    SUM(Quantity) AS item_cnt
  FROM `artful-striker-473603-f9.modulabs_project.data`
  GROUP BY CustomerID
)
```

```
-- 기존의 user_r에 (1)과 (2)를 통합
```

```
SELECT
  pc.CustomerID,
  pc.purchase_cnt,
  ic.item_cnt,
  ur.recency
FROM purchase_cnt AS pc
JOIN item_cnt AS ic
  ON pc.CustomerID = ic.CustomerID
JOIN `artful-striker-473603-f9.modulabs_project.user_r` AS ur
  ON pc.CustomerID = ur.CustomerID;
```

[결과 이미지를 넣어주세요]

user\_rf 쿼리 다음에서 열기 공유

스키마	세부정보	미리보기	테이블 탐색기	프리뷰
행	CustomerID	purchase_cnt	item_cnt	recency
1	12423	9	1312	0
2	12423	9	1312	0
3	12423	9	1312	0
4	12423	9	1312	0
5	12423	9	1312	0
6	12423	9	1312	0
7	12423	9	1312	0
8	12423	9	1312	0
9	12423	9	1312	0
10	12423	9	1312	0
11	12423	9	1312	0
12	12423	9	1312	0
13	12423	9	1312	0
14	12423	9	1312	0
15	12423	9	1312	0
16	12423	9	1312	0
17	12423	9	1312	0
18	12423	9	1312	0
19	12423	9	1312	0
20	12423	9	1312	0
21	12423	9	1312	0
22	12423	9	1312	0
23	12423	9	1312	0
24	12423	9	1312	0
25	12423	9	1312	0
26	12423	9	1312	0
27	12423	9	1312	0
28	12423	9	1312	0
29	12423	9	1312	0
30	12423	9	1312	0
31	12423	9	1312	0
32	12423	9	1312	0

## Monetary

- 고객별 총 지출액 계산 (소수점 첫째 자리에서 반올림)

```
SELECT
  CustomerID,
  ROUND(SUM(UnitPrice * Quantity), 1) AS user_total
FROM `artful-striker-473603-f9.modulabs_project.data`
GROUP BY CustomerID;
```

[결과 이미지를 넣어주세요]

행	CustomerID	user_total
1	null	1502666.8
2	12346	0.0
3	12347	4310.0
4	12348	1437.2
5	12349	1457.5
6	12350	294.4
7	12352	1265.4
8	12353	89.0
9	12354	1079.4
10	12355	459.4
11	12356	2487.4
12	12357	6207.7
13	12358	928.1
14	12359	6183.0
15	12360	2302.1
16	12361	174.9
17	12362	4665.6
18	12363	552.0
19	12364	1208.1
20	12365	320.7
21	12367	150.9
22	12370	3170.2
23	12371	1528.0
24	12372	1196.0
25	12373	324.6
26	12374	622.9
27	12375	375.4
28	12377	1548.1

- 고객별 평균 거래 금액 계산

- 고객별 평균 거래 금액을 구하기 위해 1) `data` 테이블을 `user_rf` 테이블과 조인(LEFT JOIN) 한 후, 2) `purchase_cnt`로 나누어서 3) `user_rfm` 테이블로 저장하기

```
CREATE OR REPLACE TABLE `artful-striker-473603-f9.modulabs_project.user_rfm` AS
SELECT
  rf.CustomerID AS CustomerID,
  rf.purchase_cnt,
  rf.item_cnt,
  rf.recency,
  ut.user_total,
  ROUND(ut.user_total / rf.purchase_cnt, 2) AS user_average
FROM `artful-striker-473603-f9.modulabs_project.user_rf` rf
LEFT JOIN (
  -- 고객 별 총 지출액
  SELECT
    CustomerID,
    SUM(UnitPrice * Quantity) AS user_total
  FROM `artful-striker-473603-f9.modulabs_project.data`
  GROUP BY CustomerID
) ut
ON rf.CustomerID = ut.CustomerID;
```

[결과 이미지를 넣어주세요]

user\_rfm

쿼리

다음에서 열기

공유

복사

스냅샷

스키마	세부정보	미리보기	테이블 탐색기	프리뷰	통계	계보	데이터 프로파일
행	CustomerID	purchase_cnt	item_cnt	recency	user_total	user_average	
1	12423	9	1312	0	1624.110000...	180.46	
2	12423	9	1312	0	1624.110000...	180.46	
3	12423	9	1312	0	1624.110000...	180.46	
4	12423	9	1312	0	1624.110000...	180.46	
5	12423	9	1312	0	1624.110000...	180.46	
6	12423	9	1312	0	1624.110000...	180.46	
7	12423	9	1312	0	1624.110000...	180.46	
8	12423	9	1312	0	1624.110000...	180.46	
9	12423	9	1312	0	1624.110000...	180.46	
10	12423	9	1312	0	1624.110000...	180.46	
11	12423	9	1312	0	1624.110000...	180.46	
12	12423	9	1312	0	1624.110000...	180.46	
13	12423	9	1312	0	1624.110000...	180.46	
14	12423	9	1312	0	1624.110000...	180.46	
15	12423	9	1312	0	1624.110000...	180.46	
16	12423	9	1312	0	1624.110000...	180.46	
17	12423	9	1312	0	1624.110000...	180.46	
18	12423	9	1312	0	1624.110000...	180.46	
19	12423	9	1312	0	1624.110000...	180.46	
20	12423	9	1312	0	1624.110000...	180.46	
21	12423	9	1312	0	1624.110000...	180.46	
22	12423	9	1312	0	1624.110000...	180.46	
23	12423	9	1312	0	1624.110000...	180.46	
24	12423	9	1312	0	1624.110000...	180.46	
25	12423	9	1312	0	1624.110000...	180.46	
26	12423	9	1312	0	1624.110000...	180.46	
27	12423	9	1312	0	1624.110000...	180.46	
28	12423	9	1312	0	1624.110000...	180.46	
29	12423	9	1312	0	1624.110000...	180.46	
30	12423	9	1312	0	1624.110000...	180.46	
31	12423	9	1312	0	1624.110000...	180.46	
32	12423	9	1312	0	1624.110000...	180.46	

## RFM 통합 테이블 출력하기

- 최종 user\_rfm 테이블을 출력하기

```
SELECT *
FROM `artful-striker-473603-f9.modulabs_project.user_rfm`
LIMIT 100;
```

[결과 이미지를 넣어주세요]

행	CustomerID	purchase_cnt	item_cnt	recency	user_total	user_average
1	12423	9	1312	0	1624.110000000...	180.46
2	12423	9	1312	0	1624.110000000...	180.46
3	12423	9	1312	0	1624.110000000...	180.46
4	12423	9	1312	0	1624.110000000...	180.46
5	12423	9	1312	0	1624.110000000...	180.46
6	12423	9	1312	0	1624.110000000...	180.46
7	12423	9	1312	0	1624.110000000...	180.46
8	12423	9	1312	0	1624.110000000...	180.46
9	12423	9	1312	0	1624.110000000...	180.46
10	12423	9	1312	0	1624.110000000...	180.46
11	12423	9	1312	0	1624.110000000...	180.46
12	12423	9	1312	0	1624.110000000...	180.46
13	12423	9	1312	0	1624.110000000...	180.46
14	12423	9	1312	0	1624.110000000...	180.46
15	12423	9	1312	0	1624.110000000...	180.46
16	12423	9	1312	0	1624.110000000...	180.46
17	12423	9	1312	0	1624.110000000...	180.46
18	12423	9	1312	0	1624.110000000...	180.46
19	12423	9	1312	0	1624.110000000...	180.46
20	12423	9	1312	0	1624.110000000...	180.46
21	12423	9	1312	0	1624.110000000...	180.46
22	12423	9	1312	0	1624.110000000...	180.46
23	12423	9	1312	0	1624.110000000...	180.46
24	12423	9	1312	0	1624.110000000...	180.46
25	12423	9	1312	0	1624.110000000...	180.46
26	12423	9	1312	0	1624.110000000...	180.46
27	12423	9	1312	0	1624.110000000...	180.46
28	12423	9	1312	0	1624.110000000...	180.46
29	12423	9	1312	0	1624.110000000...	180.46
30	12423	9	1312	0	1624.110000000...	180.46
31	12423	9	1312	0	1624.110000000...	180.46
32	12423	9	1312	0	1624.110000000...	180.46
33	12423	9	1312	0	1624.110000000...	180.46
34	12423	9	1312	0	1624.110000000...	180.46
35	12423	9	1312	0	1624.110000000...	180.46
36	12423	9	1312	0	1624.110000000...	180.46
37	12423	9	1312	0	1624.110000000...	180.46
38	12423	9	1312	0	1624.110000000...	180.46
39	12423	9	1312	0	1624.110000000...	180.46
40	12423	9	1312	0	1624.110000000...	180.46
41	12423	9	1312	0	1624.110000000...	180.46
42	12423	9	1312	0	1624.110000000...	180.46
43	12423	9	1312	0	1624.110000000...	180.46
44	12423	9	1312	0	1624.110000000...	180.46
45	12423	9	1312	0	1624.110000000...	180.46
46	12423	9	1312	0	1624.110000000...	180.46
47	12423	9	1312	0	1624.110000000...	180.46
48	12423	9	1312	0	1624.110000000...	180.46
49	12423	9	1312	0	1624.110000000...	180.46
50	12423	9	1312	0	1624.110000000...	180.46

## 11-8. 추가 Feature 추출

### 1. 구매하는 제품의 다양성

- 1) 고객 별로 구매한 상품들의 고유한 수를 계산하기
- 2) `user_rfm` 테이블과 결과를 합치기
- 3) `user_data` 라는 이름의 테이블에 저장하기



```

CREATE OR REPLACE TABLE project_name.modulabs_project.user_data AS
WITH unique_products AS (
  SELECT
    CustomerID,
    COUNT(DISTINCT StockCode) AS unique_products
  FROM project_name.modulabs_project.data
  GROUP BY CustomerID
)
SELECT ur.*, up.* EXCEPT (CustomerID)
FROM project_name.modulabs_project.user_rfm AS ur
JOIN unique_products AS up
ON ur.CustomerID = up.CustomerID;

```

[결과 이미지를 넣어주세요]

## 2. 평균 구매 주기

- 고객들의 쇼핑 패턴을 이해하는 것을 목표 (고객 별 재방문 주기 살펴보기)
  - 군 구매 소요 일수를 계산하고, 그 결과를 `user_data` 에 통합

```

CREATE OR REPLACE TABLE project_name.modulabs_project.user_data AS
WITH purchase_intervals AS (
  -- (2) 고객 별 구매와 구매 사이의 평균 소요 일수
  SELECT
    CustomerID,
    CASE WHEN ROUND(AVG(interval_), 2) IS NULL THEN 0 ELSE ROUND(AVG(interval_), 2) END AS average_interval
  FROM (
    -- (1) 구매와 구매 사이에 소요된 일수
    SELECT
      CustomerID,
      DATE_DIFF(InvoiceDate, LAG(InvoiceDate) OVER (PARTITION BY CustomerID ORDER BY InvoiceDate), DAY) AS interval_
    FROM
      project_name.modulabs_project.data
    WHERE CustomerID IS NOT NULL
  )
  GROUP BY CustomerID
)

SELECT u.*, pi.* EXCEPT (CustomerID)
FROM project_name.modulabs_project.user_data AS u
LEFT JOIN purchase_intervals AS pi
ON u.CustomerID = pi.CustomerID;

```

[결과 이미지를 넣어주세요]

user_data							
<div> <span>쿼리</span> <span>다음에서 열기</span> <span>공유</span> <span>복사</span> <span>스냅샷</span> <span>삭제</span> </div>							
스키마	세부정보	미리보기	테이블 탐색기	프리뷰	통계	계보	데이터 프로파일
행	CustomerID	purchase_cnt	item_cnt	recency	user_total	user_average	unique_prod...
1	16446	3	2	0	2.899999999...	0.97	3
2	16446	3	2	0	2.899999999...	0.97	3
3	16446	3	2	0	2.899999999...	0.97	3
4	16446	3	2	0	2.899999999...	0.97	3
5	15344	5	106	0	476.5800000...	95.32	21
6	15344	5	106	0	476.5800000...	95.32	21
7	15344	5	106	0	476.5800000...	95.32	21
8	15344	5	106	0	476.5800000...	95.32	21
9	15344	5	106	0	476.5800000...	95.32	21
10	15344	5	106	0	476.5800000...	95.32	21
11	15344	5	106	0	476.5800000...	95.32	21
12	15344	5	106	0	476.5800000...	95.32	21
13	15344	5	106	0	476.5800000...	95.32	21
14	15344	5	106	0	476.5800000...	95.32	21
15	15344	5	106	0	476.5800000...	95.32	21
16	15344	5	106	0	476.5800000...	95.32	21
17	15344	5	106	0	476.5800000...	95.32	21
18	15344	5	106	0	476.5800000...	95.32	21
19	15344	5	106	0	476.5800000...	95.32	21
20	15344	5	106	0	476.5800000...	95.32	21
21	15344	5	106	0	476.5800000...	95.32	21
22	15344	5	106	0	476.5800000...	95.32	21
23	15344	5	106	0	476.5800000...	95.32	21
24	15344	5	106	0	476.5800000...	95.32	21
25	15344	5	106	0	476.5800000...	95.32	21
26	15344	5	106	0	476.5800000...	95.32	21
27	15344	5	106	0	476.5800000...	95.32	21
28	15344	5	106	0	476.5800000...	95.32	21
29	15344	5	106	0	476.5800000...	95.32	21
30	15344	5	106	0	476.5800000...	95.32	21
31	15344	5	106	0	476.5800000...	95.32	21
32	15344	5	106	0	476.5800000...	95.32	21

### 3. 구매 취소 경향성

- 고객의 취소 패턴 파악하기
  - 1) 취소 빈도(cancel\_frequency) : 고객 별로 취소한 거래의 총 횟수
  - 2) 취소 비율(cancel\_rate) : 각 고객이 한 모든 거래 중에서 취소를 한 거래의 비율
    - 취소 빈도와 취소 비율을 계산하고 그 결과를 user\_data 에 통합하기  
(취소 비율은 소수점 두번째 자리)

```
CREATE OR REPLACE TABLE `artful-striker-473603-f9.modulabs_project.user_data` AS

WITH TransactionInfo AS (
  SELECT
    CustomerID,
    COUNT(*) AS total_transactions,
    SUM(CASE WHEN InvoiceNo LIKE 'C%' THEN 1 ELSE 0 END) AS cancel_frequency
  FROM `artful-striker-473603-f9.modulabs_project.data`
  GROUP BY CustomerID
)

SELECT
  u.*, t.* EXCEPT(CustomerID),
  ROUND(t.cancel_frequency / t.total_transactions, 2) AS cancel_rate
FROM `artful-striker-473603-f9.modulabs_project.user_data` AS u
LEFT JOIN TransactionInfo AS t
ON u.CustomerID = t.CustomerID;
```

[결과 이미지를 넣어주세요]

user\_data

🔍 쿼리

📄 다음에서 열기

👤 공유

📋 복사

📊 스냅샷

🗑️ 삭제

📤 내보내기

스키마	세부정보	미리보기	테이블 탐색기	프로뷰	통계	계보	데이터 프로파일	데이터 품질		
명	CustomerID	purchase_cnt	item_cnt	recency	user_total	user_average	unique_prod...	total_transac...	cancel_frequ...	cancel_rate
1	17307	1	-144	365	-152.640000...	-152.64	1	1	1	1.0
2	13829	1	-12	359	-102.0	-102.0	1	1	1	1.0
3	18141	1	-12	360	-35.4000000...	-35.4	1	1	1	1.0
4	16579	1	-12	365	30.5999999...	30.6	1	1	1	1.0
5	16061	1	-1	269	-29.95	-29.95	1	1	1	1.0
6	14119	1	-2	354	-19.9	-19.9	1	1	1	1.0
7	16138	1	-1	368	-7.95	-7.95	1	1	1	1.0
8	12943	1	-1	301	-3.75	-3.75	1	1	1	1.0
9	16428	1	-1	81	2.95	2.95	1	1	1	1.0
10	14679	1	-1	371	-2.55	-2.55	1	1	1	1.0
11	16995	1	-1	372	-1.25	-1.25	1	1	1	1.0
12	16738	1	3	297	3.75	3.75	1	1	0	0.0
13	16454	1	2	64	5.9	5.9	1	1	0	0.0
14	17956	1	1	249	12.75	12.75	1	1	0	0.0
15	17763	1	12	263	15.0	15.0	1	1	0	0.0
16	13307	1	4	120	15.0	15.0	1	1	0	0.0
17	16093	1	20	106	17.0	17.0	1	1	0	0.0
18	17986	1	10	56	20.8	20.8	1	1	0	0.0
19	16953	1	10	30	20.8	20.8	1	1	0	0.0
20	17102	1	2	261	25.5	25.5	1	1	0	0.0
21	15657	1	24	22	30.0	30.0	1	1	0	0.0
22	13120	1	12	238	30.5999999...	30.6	1	1	0	0.0
23	16765	1	4	294	34.0	34.0	1	1	0	0.0
24	14576	1	12	372	35.4000000...	35.4	1	1	0	0.0
25	15940	1	4	311	35.8	35.8	1	1	0	0.0
26	18184	1	60	15	49.8	49.8	1	1	0	0.0
27	14351	1	12	164	51.0	51.0	1	1	0	0.0
28	15313	1	25	110	52.0	52.0	1	1	0	0.0
29	13366	1	144	50	56.1600000...	56.16	1	1	0	0.0
30	13391	1	4	203	59.8	59.8	1	1	0	0.0
31	13302	1	5	155	63.75	63.75	1	1	0	0.0
32	13185	1	12	267	71.4	71.4	1	1	0	0.0

- 다양한 컬럼들을 활용하여 고객의 구매 패턴과 선호도를 보다 심층적으로 이해할 수 있도록 최종적으로 `user_data`를 출력하기

```
SELECT *
FROM `artful-striker-473603-f9.modulabs_project.user_data`
LIMIT 100;
```

[결과 이미지를 넣어주세요]

명	CustomerID	purchase_cnt	item_cnt	recency	user_total	user_average	unique_products	total_transactions	cancel_frequency	cancel_rate
1	17307	1	-144	365	-152.64000000...	-152.64	1	1	1	1.0
2	13829	1	-12	359	-102.0	-102.0	1	1	1	1.0
3	18141	1	-12	360	-35.400000000...	-35.4	1	1	1	1.0
4	16579	1	-12	365	30.599999999...	30.6	1	1	1	1.0
5	16061	1	-1	269	-29.95	-29.95	1	1	1	1.0
6	14119	1	-2	354	-19.9	-19.9	1	1	1	1.0
7	16138	1	-1	368	-7.95	-7.95	1	1	1	1.0
8	12943	1	-1	301	-3.75	-3.75	1	1	1	1.0
9	16428	1	-1	81	-2.95	-2.95	1	1	1	1.0
10	14679	1	-1	371	-2.55	-2.55	1	1	1	1.0
11	16995	1	-1	372	-1.25	-1.25	1	1	1	1.0
12	16738	1	3	297	3.75	3.75	1	1	0	0.0
13	16454	1	2	64	5.9	5.9	1	1	0	0.0
14	17956	1	1	249	12.75	12.75	1	1	0	0.0
15	17763	1	12	263	15.0	15.0	1	1	0	0.0
16	13307	1	4	120	15.0	15.0	1	1	0	0.0
17	16093	1	20	106	17.0	17.0	1	1	0	0.0
18	17986	1	10	56	20.8	20.8	1	1	0	0.0
19	16953	1	10	30	20.8	20.8	1	1	0	0.0
20	17102	1	2	261	25.5	25.5	1	1	0	0.0
21	15657	1	24	22	30.0	30.0	1	1	0	0.0
22	13120	1	12	238	30.599999999...	30.6	1	1	0	0.0
23	16765	1	4	294	34.0	34.0	1	1	0	0.0
24	14576	1	12	372	35.400000000...	35.4	1	1	0	0.0
25	15940	1	4	311	35.8	35.8	1	1	0	0.0

26	18184	1	60	15	49.8	49.8	1	1	0	0.0
27	14351	1	12	164	51.0	51.0	1	1	0	0.0
28	15313	1	25	110	52.0	52.0	1	1	0	0.0
29	13366	1	144	50	56.16000000000...	56.16	1	1	0	0.0
30	13391	1	4	203	59.8	59.8	1	1	0	0.0
31	13302	1	5	155	63.75	63.75	1	1	0	0.0
32	13185	1	12	267	71.4	71.4	1	1	0	0.0
33	18113	1	72	368	76.32000000000...	76.32	1	1	0	0.0
34	15668	1	72	217	76.32000000000...	76.32	1	1	0	0.0
35	16148	1	72	296	76.32000000000...	76.32	1	1	0	0.0
36	14090	1	72	324	76.32000000000...	76.32	1	1	0	0.0
37	15753	1	144	304	79.2	79.2	1	1	0	0.0
38	16078	1	16	283	79.2	79.2	1	1	0	0.0
39	13747	1	8	373	79.6	79.6	1	1	0	0.0
40	17732	1	192	359	80.64	80.64	1	1	0	0.0
41	13841	1	100	252	85.0	85.0	1	1	0	0.0
42	12814	1	48	101	85.92	85.92	1	1	0	0.0
43	13703	1	10	318	99.5	99.5	1	1	0	0.0
44	18068	1	6	289	101.6999999999...	101.7	1	1	0	0.0
45	18174	1	50	7	104.0	104.0	1	1	0	0.0
46	15070	1	36	372	106.2	106.2	1	1	0	0.0
47	15562	1	39	351	134.55	134.55	1	1	0	0.0
48	15316	1	100	326	165.0	165.0	1	1	0	0.0
49	16144	1	16	246	175.2	175.2	1	1	0	0.0
50	17331	1	16	123	175.2	175.2	1	1	0	0.0

고객별 최근 구매일, 거래 건수, 총 지출액을 기반으로 구매 패턴을 파악할 수 있었습니다.

구매 다양성과 평균 구매 간격을 통해 충성도와 재구매 가능성을 예측할 수 있었고,

취소 빈도와 비율을 분석하여 주문 관리가 필요한 고객을 식별할 수 있었습니다.

## 회고

[회고 내용을 작성해주세요]

**Keep :**

데이터 전처리 과정에서 결측치, 중복, 단가 0원, 제품과 관련 없는 거래 기록 등 체계적으로 제거

RFM 분석 과정에서 Recency, Frequency, Monetary 지표를 개선하고, 고객별 구매 패턴을 이해하기 위해 평균 구매 간격, 구매 다양성, 취소 빈도 및 취소 비율까지 통합

MAX() 함수 관련 오류를 경험하며 서브쿼리와 집계함수 활용에 대한 이해가 깊어짐

**Problem :**

11-7 RFM스코어 중 '가장 최근 구매 일자를 MAX() 함수로 찾아보기' 에서 MAX() 함수와 일반 컬럼을 함께 SELECT할 때 GROUP BY 가 없어서 에러 발생 → 서브쿼리로 처리해서 오류가 안나게 해결

**Try :**

집계함수와 일반 컬럼을 함께 쓸 때는 서브쿼리, 윈도우 함수, GROUP BY를 적절히 활용할 것

데이터 전처리 단계에서 이상치, 결측치, 불필요한 기록을 꼼꼼히 확인하고 제거하는 습관을 유지할 것