

# 고객을 세그먼테이션하자 [프로젝트] 승희

## 11-2. 데이터 불러오기

### 데이터 살펴보기

- 테이블에 있는 10개의 행만 출력하기

```
SELECT *  
FROM `trim-field-479202-b8.aiffel_dataset.data`  
LIMIT 10;
```

[결과 이미지를 넣어주세요]

행	InvoiceNo	StockC...	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
1	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	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	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00 UTC	2.75	17850	United Kingdom
4	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00 UTC	3.39	17850	United Kingdom
5	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00 UTC	3.39	17850	United Kingdom
6	536365	22752	SET 7 BABUSHKA NESTING BOXES	2	2010-12-01 08:26:00 UTC	7.65	17850	United Kingdom
7	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	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 ORNAMENT	32	2010-12-01 08:34:00 UTC	1.69	13047	United Kingdom

- 전체 데이터는 몇 행으로 구성되어 있는지 확인하기

```
SELECT COUNT(*) AS row_count  
FROM `trim-field-479202-b8.aiffel_dataset.data`;
```

[결과 이미지를 넣어주세요]

행	row_count
1	541909

## 데이터 수 세기

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

```
SELECT
  COUNT(InvoiceNo) AS cnt_InvoiceNo,
  COUNT(StockCode) AS cnt_StockCode,
  COUNT(Description) AS cnt_Description,
  COUNT(Quantity) AS cnt_Quantity,
  COUNT(InvoiceDate) AS cnt_InvoiceDate,
  COUNT(UnitPrice) AS cnt_UnitPrice,
  COUNT(CustomerID) AS cnt_CustomerID,
  COUNT(Country) AS cnt_Country
FROM `trim-field-479202-b8.aiffel_dataset.data`;
```

[결과 이미지를 넣어주세요]

행	cnt_InvoiceNo	cnt_StockCode	cnt_Description	cnt_Quantity	cnt_InvoiceDate	cnt_UnitPrice	cnt_CustomerID	cnt_Country
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 col, COUNTIF(InvoiceNo IS NULL) / COUNT(*)
AS null_ratio
FROM `trim-field-479202-b8.aiffel_dataset.data`
UNION ALL
SELECT 'StockCode', COUNTIF(StockCode IS NULL) / COUNT(*) AS nu
```

```

    ll_ratio
FROM `trim-field-479202-b8.aiffel_dataset.data`
UNION ALL
SELECT 'Description', COUNTIF(Description IS NULL) / COUNT(*) AS null_ratio
FROM `trim-field-479202-b8.aiffel_dataset.data`
UNION ALL
SELECT 'Quantity', COUNTIF(Quantity IS NULL) / COUNT(*) AS null_ratio
FROM `trim-field-479202-b8.aiffel_dataset.data`
UNION ALL
SELECT 'InvoiceDate', COUNTIF(InvoiceDate IS NULL) / COUNT(*) AS null_ratio
FROM `trim-field-479202-b8.aiffel_dataset.data`
UNION ALL
SELECT 'UnitPrice', COUNTIF(UnitPrice IS NULL) / COUNT(*) AS null_ratio
FROM `trim-field-479202-b8.aiffel_dataset.data`
UNION ALL
SELECT 'CustomerID', COUNTIF(CustomerID IS NULL) / COUNT(*) AS null_ratio
FROM `trim-field-479202-b8.aiffel_dataset.data`
UNION ALL
SELECT 'Country', COUNTIF(Country IS NULL) / COUNT(*) AS null_ratio
FROM `trim-field-479202-b8.aiffel_dataset.data`;

```

[결과 이미지를 넣어주세요]

행	col	null_ratio
1	InvoiceNo	0.0
2	StockCode	0.0
3	Description	0.002683107311...
4	Quantity	0.0
5	InvoiceDate	0.0
6	UnitPrice	0.0
7	CustomerID	0.249266943342...
8	Country	0.0

## 결측치 처리 전략

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

```
SELECT DISTINCT Description
FROM `trim-field-479202-b8.aiffel_dataset.data`
WHERE StockCode = '85123A';
```

[결과 이미지를 넣어주세요]

행	Description
1	WHITE HANGING HEART T-LIGHT HOLDER
2	?
3	wrongly marked carton 22804
4	CREAM HANGING HEART T-LIGHT HOLDER

## 결측치 처리

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

```
DELETE FROM `trim-field-479202-b8.modulabs_project.data`  
WHERE CustomerID IS NULL  
OR Description IS NULL;
```

[결과 이미지를 넣어주세요]

작업 정보

결과

실행 세부정보

실행 그래프



이 문으로 data의 행 135,080개가 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  
  COUNT(*) AS duplicated_row_count  
FROM (  
  SELECT  
    InvoiceNo,  
    StockCode,  
    Description,  
    Quantity,  
    InvoiceDate,  
    UnitPrice,  
    CustomerID,
```

```
Country,
COUNT(*) AS cnt
FROM `trim-field-479202-b8.aiffel_dataset.data`
GROUP BY
InvoiceNo, StockCode, Description, Quantity,
InvoiceDate, UnitPrice, CustomerID, Country
HAVING cnt > 1);
```

[결과 이미지를 넣어주세요]

행	duplicated_row_count ▾
1	4837

## 중복값 처리

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

```
CREATE OR REPLACE TABLE `trim-field-479202-b8.aiffel_dataset.data`
AS
SELECT DISTINCT *
FROM `trim-field-479202-b8.aiffel_dataset.data`;
```

[결과 이미지를 넣어주세요]

작업 정보

**결과**

실행 세부정보

실행 그래프



이 문으로 이름이 data인 테이블이 교체되었습니다.

## 11-6. 데이터 전처리(3): 오류값 처리

### InvoiceNo 살펴보기

- 고유(unique)한 InvoiceNo 의 개수를 출력하기

```
SELECT COUNT(DISTINCT InvoiceNo) AS unique_invoice_no_cnt  
FROM `trim-field-479202-b8.aiffel_dataset.data`;
```

[결과 이미지를 넣어주세요]

행	unique_invoice_no_cnt
1	22190

- 고유한 InvoiceNo 를 앞에서부터 100개를 출력하기

```
SELECT DISTINCT InvoiceNo  
FROM `trim-field-479202-b8.aiffel_dataset.data`  
ORDER BY InvoiceNo  
LIMIT 100;
```

[결과 이미지를 넣어주세요]

행	InvoiceNo
1	536365
2	536366
3	536367
4	536368
5	536369
6	536370
7	536371
8	536372
9	536373
10	536374
11	536375
12	536376
13	536377
14	536378
15	536380
16	536381
17	536382
18	536384
19	536385
20	536386
21	536387
22	536388
23	536389
24	536390
25	536392
26	536393
27	536394
28	536395
29	536396
30	536397
31	536398
32	536399
33	536400
34	536401
35	536402
36	536403
37	536404
38	536405
39	536406
40	536407
41	536408
42	536409
43	536412
44	536415
45	536416
46	536420
47	536423
48	536425
49	536437
50	536446

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

```
SELECT *
FROM `trim-field-479202-b8.aiffel_dataset.data`
WHERE InvoiceNo LIKE 'C%'
LIMIT 100;
```

[결과 이미지를 넣어주세요]

행	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
1	C541433	23166	MEDIUM CERAMIC TOP STORA...	-74215	2011-01-18 10:17:00 UTC	1.04	12346	United Kingdom
2	C545329	M	Manual	-1	2011-03-01 15:47:00 UTC	183.75	12352	Norway
3	C545329	M	Manual	-1	2011-03-01 15:47:00 UTC	280.05	12352	Norway
4	C545330	M	Manual	-1	2011-03-01 15:49:00 UTC	376.5	12352	Norway
5	C547388	22413	METAL SIGN TAKE IT OR LEAVE...	-6	2011-03-22 16:07:00 UTC	2.95	12352	Norway
6	C547388	21914	BLUE HARMONICA IN BOX	-12	2011-03-22 16:07:00 UTC	1.25	12352	Norway
7	C547388	22645	CERAMIC HEART FAIRY CAKE ...	-12	2011-03-22 16:07:00 UTC	1.45	12352	Norway
8	C547388	22701	PINK DOG BOWL	-6	2011-03-22 16:07:00 UTC	2.95	12352	Norway
9	C547388	37448	CERAMIC CAKE DESIGN SPOTT...	-12	2011-03-22 16:07:00 UTC	1.49	12352	Norway
10	C547388	84050	PINK HEART SHAPE EGG FRYN...	-12	2011-03-22 16:07:00 UTC	1.65	12352	Norway
11	C547388	22784	LANTERN CREAM GAZEBO	-3	2011-03-22 16:07:00 UTC	4.95	12352	Norway
12	C549955	22666	RECIPE BOX PANTRY YELLOW ...	-2	2011-04-13 13:38:00 UTC	2.95	12359	Cyprus
13	C549955	22839	3 TIER CAKE TIN GREEN AND C...	-2	2011-04-13 13:38:00 UTC	14.95	12359	Cyprus
14	C580165	22797	CHEST OF DRAWERS GINGHA...	-2	2011-12-02 11:21:00 UTC	16.95	12359	Cyprus
15	C580165	22826	LOVE SEAT ANTIQUE WHITE M...	-1	2011-12-02 11:21:00 UTC	42.5	12359	Cyprus
16	C580165	23245	SET OF 3 REGENCY CAKE TINS	-2	2011-12-02 11:21:00 UTC	4.95	12359	Cyprus
17	C580165	22720	SET OF 3 CAKE TINS PANTRY D...	-1	2011-12-02 11:21:00 UTC	4.95	12359	Cyprus
18	C544902	22629	SPACEBOY LUNCH BOX	-1	2011-02-24 13:05:00 UTC	1.95	12362	Belgium
19	C544902	22273	FELTCRAFT DOLL MOLLY	-1	2011-02-24 13:05:00 UTC	2.95	12362	Belgium
20	C563752	22659	LUNCH BOX I LOVE LONDON	-6	2011-08-19 10:38:00 UTC	1.95	12362	Belgium
21	C563752	22891	TEA FOR ONE POLKADOT	-1	2011-08-19 10:38:00 UTC	4.25	12362	Belgium
22	C563752	22630	DOLLY GIRL LUNCH BOX	-1	2011-08-19 10:38:00 UTC	1.95	12362	Belgium
23	C579178	22941	CHRISTMAS LIGHTS 10 REINDE...	-2	2011-11-28 14:55:00 UTC	8.5	12362	Belgium
24	C579178	22942	CHRISTMAS LIGHTS 10 SANTAS	-2	2011-11-28 14:55:00 UTC	8.5	12362	Belgium
25	C579178	22943	CHRISTMAS LIGHTS 10 VINTA...	-3	2011-11-28 14:55:00 UTC	4.95	12362	Belgium
26	C544577	M	Manual	-1	2011-02-21 14:02:00 UTC	320.69	12365	Cyprus
27	C581071	85099C	JUMBO BAG BAROQUE BLACK ...	-1	2011-12-07 11:27:00 UTC	2.08	12375	Finland
28	C559193	22659	LUNCH BOX I LOVE LONDON	-1	2011-07-07 10:26:00 UTC	1.95	12379	Belgium
29	C572532	22960	JAM MAKING SET WITH JARS	-1	2011-10-24 16:03:00 UTC	4.25	12380	Belgium
30	C565050	23243	SET OF TEA COFFEE SUGAR TI...	-2	2011-08-31 17:10:00 UTC	4.95	12381	Norway
31	C565050	21791	VINTAGE HEADS AND TAILS C...	-12	2011-08-31 17:10:00 UTC	1.25	12381	Norway
32	C565050	85159B	WHITE TEA,COFFEE,SUGAR JARS	-1	2011-08-31 17:10:00 UTC	1.95	12381	Norway
33	C565050	23198	PANTRY MAGNETIC SHOPPIN...	-10	2011-08-31 17:10:00 UTC	1.45	12381	Norway
34	C541586	21218	RED SPOTTY BISCUIT TIN	-3	2011-01-19 14:36:00 UTC	3.75	12383	Belgium
35	C567540	23173	REGENCY TEAPOT ROSES	-1	2011-09-21 10:00:00 UTC	9.95	12384	Switzerland
36	C567540	84279P	CHERRY BLOSSOM DECORATIV...	-2	2011-09-21 10:00:00 UTC	3.75	12384	Switzerland
37	C567540	23295	SET OF 12 MINI LOAF BAKING ...	-2	2011-09-21 10:00:00 UTC	0.83	12384	Switzerland
38	C540152	22666	RECIPE BOX PANTRY YELLOW ...	-1	2011-01-05 11:21:00 UTC	2.95	12395	Belgium
39	C544419	85232A	SET/3 POLKADOT STACKING TI...	-3	2011-02-18 15:16:00 UTC	4.95	12395	Belgium
40	C578432	21559	STRAWBERRY LUNCH BOX WIT...	-1	2011-11-24 12:20:00 UTC	2.55	12395	Belgium
41	C572187	POST	POSTAGE	-1	2011-10-21 11:00:00 UTC	18.0	12403	Denmark
42	C562728	22846	BREAD BIN DINER STYLE RED	-4	2011-08-09 09:41:00 UTC	14.95	12406	Denmark
43	C562728	22849	BREAD BIN DINER STYLE MINT	-1	2011-08-09 09:41:00 UTC	14.95	12406	Denmark
44	C562728	22848	BREAD BIN DINER STYLE PINK	-4	2011-08-09 09:41:00 UTC	14.95	12406	Denmark
45	C549253	22328	ROUND SNACK BOXES SET OF ...	-1	2011-04-07 12:20:00 UTC	2.95	12408	Belgium
46	C549253	22725	ALARM CLOCK BAKELIKE CHO...	-1	2011-04-07 12:20:00 UTC	3.75	12408	Belgium
47	C549253	22727	ALARM CLOCK BAKELIKE RED	-1	2011-04-07 12:20:00 UTC	3.75	12408	Belgium
48	C549253	20712	JUMBO BAG WOODLAND ANIM...	-1	2011-04-07 12:20:00 UTC	2.08	12408	Belgium
49	C549253	22466	FAIRY TALE COTTAGE NIGHTLI...	-1	2011-04-07 12:20:00 UTC	1.95	12408	Belgium
50	C549533	22628	PICNIC BOXES SET OF 3 RETRO...	-3	2011-04-08 17:08:00 UTC	4.95	12408	Belgium

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

```
SELECT ROUND(SUM(CASE WHEN InvoiceNo LIKE 'C%' THEN 1 ELSE 0 END) / COUNT(*) * 100, 1) AS canceled_rate_pct
FROM `trim-field-479202-b8.aiffel_dataset.data`;
```

[결과 이미지를 넣어주세요]

행	canceled_rate_pct
1	2.2

## StockCode 살펴보기

- 고유한 StockCode 의 개수를 출력하기

```
SELECT COUNT(DISTINCT StockCode) AS unique_stockcode_cnt
FROM `trim-field-479202-b8.aiffel_dataset.data`;
```

[결과 이미지를 넣어주세요]

행	unique_stockcod...
1	3684

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

```
SELECT StockCode, COUNT(*) AS sell_cnt
FROM `trim-field-479202-b8.aiffel_dataset.data`
GROUP BY StockCode
ORDER BY sell_cnt DESC
LIMIT 10;
```

[결과 이미지를 넣어주세요]

행	StockCode	sell_cnt
1	85123A	2065
2	22423	1894
3	85099B	1659
4	47566	1409
5	84879	1405
6	20725	1346
7	22720	1224
8	POST	1196
9	22197	1110
10	23203	1108

- **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 `trim-field-479202-b8.aiffel_dataset.data`
)
WHERE number_count <= 1;
```

[결과 이미지를 넣어주세요]

행	StockCode	number_count
1	POST	0
2	M	0
3	C2	1
4	D	0
5	BANK CHARGES	0
6	PADS	0
7	DOT	0
8	CRUK	0

- **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 퍼센트 인지 구하기 (소수점 두 번째 자리까지)

```
WITH stock_number_cnt AS (SELECT StockCode, LENGTH(StockCode) - LENGTH(REGEXP_REPLACE(StockCode, r'[0-9]', '')) AS number_count
FROM `trim-field-479202-b8.aiffel_dataset.data`)
SELECT ROUND(100.0 * COUNTIF(number_count <= 1) / COUNT(*), 2) AS ratio_pct
FROM stock_number_cnt;
```

[결과 이미지를 넣어주세요]

행	ratio_pct
1	0.48

- 제품과 관련되지 않은 거래 기록을 제거하기

```
DELETE FROM `trim-field-479202-b8.aiffel_dataset.data`  
WHERE StockCode IN (SELECT DISTINCT StockCode  
  FROM (SELECT StockCode, LENGTH(StockCode) - LENGTH(REGEXP_  
    REPLACE(StockCode, r'[0-9]', '')) AS number_count  
  FROM `trim-field-479202-b8.aiffel_dataset.data`)  
  WHERE number_count <= 1);
```

[결과 이미지를 넣어주세요]



이 문으로 data의 행 1,915개가 삭제되었습니다.

## Description 살펴보기

- 고유한 Description 별 출현 빈도를 계산하고 상위 30개를 출력하기

```
SELECT  
  Description,  
  COUNT(*) AS description_cnt  
FROM `trim-field-479202-b8.aiffel_dataset.data`  
GROUP BY Description  
ORDER BY description_cnt DESC  
LIMIT 30;
```


[결과 이미지를 넣어주세요]

행	Description ▼	description_cnt ▼
1	WHITE HANGING HEART T-LIG...	2058
2	REGENCY CAKESTAND 3 TIER	1894
3	JUMBO BAG RED RETROSPOT	1659
4	PARTY BUNTING	1409
5	ASSORTED COLOUR BIRD ORN...	1405
6	LUNCH BAG RED RETROSPOT	1345
7	SET OF 3 CAKE TINS PANTRY D...	1224
8	LUNCH BAG BLACK SKULL.	1099
9	PACK OF 72 RETROSPOT CAKE ...	1062
10	SPOTTY BUNTING	1026
11	PAPER CHAIN KIT 50'S CHRIST...	1013
12	LUNCH BAG SPACEBOY DESIGN	1006
13	LUNCH BAG CARS BLUE	1000
14	HEART OF WICKER SMALL	990
15	NATURAL SLATE HEART CHAL...	989
16	JAM MAKING SET WITH JARS	966
17	LUNCH BAG PINK POLKADOT	961
18	LUNCH BAG SUKI DESIGN	932
19	ALARM CLOCK BAKELIKE RED	917
20	REX CASH+CARRY JUMBO SHO...	900
21	WOODEN PICTURE FRAME WHI...	900
22	JUMBO BAG PINK POLKADOT	897
23	LUNCH BAG APPLE DESIGN	890
24	SET OF 4 PANTRY JELLY MOUL...	890
25	BAKING SET 9 PIECE RETROSP...	885
26	RECIPE BOX PANTRY YELLOW ...	883
27	JAM MAKING SET PRINTED	883
28	LUNCH BAG WOODLAND	850
29	ROSES REGENCY TEACUP AND ...	844
30	VICTORIAN GLASS HANGING T...	843

- 서비스 관련 정보를 포함하는 행들을 제거하기

```
DELETE
FROM `trim-field-479202-b8.aiffel_dataset.data`
WHERE Description LIKE '%POSTAGE%'
      OR Description LIKE '%MANUAL%'
      OR Description LIKE '%SAMPLES%';
```


[결과 이미지를 넣어주세요]

 이 문으로 data의 행 0개가 삭제되었습니다.

- 대소문자를 혼합하고 있는 데이터를 대문자로 표준화 하기

```
CREATE OR REPLACE TABLE `trim-field-479202-b8.aiffel_dataset.data`
AS
SELECT
  * EXCEPT (Description),
  UPPER(Description) AS Description
FROM `trim-field-479202-b8.aiffel_dataset.data`;
```

[결과 이미지를 넣어주세요]

 이 문으로 이름이 data인 테이블이 교체되었습니다.

**UnitPrice** 살펴보기

- **UnitPrice** 의 최솟값, 최댓값, 평균을 구하기

```
SELECT
  MIN(UnitPrice) AS min_price,
  MAX(UnitPrice) AS max_price,
  AVG(UnitPrice) AS avg_price
FROM `trim-field-479202-b8.modulabs_project.data`;
```

[결과 이미지를 넣어주세요]

행	min_price	max_price	avg_price
1	0.0	649.5	2.907457172951...

- 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,
  AVG(Quantity) AS avg_quantity
FROM `trim-field-479202-b8.aiffel_dataset.data`
WHERE UnitPrice = 0;
```

[결과 이미지를 넣어주세요]


행	cnt_quantity	min_quantity	max_quantity	avg_quantity
1	33	1	12540	420.51515151...

- **UnitPrice = 0** 를 제거하고 일관된 데이터셋을 유지하기

```
CREATE OR REPLACE TABLE `trim-field-479202-b8.aiffel_dataset.data`
AS
```

```
SELECT *  
FROM `trim-field-479202-b8.aiffel_dataset.data`  
WHERE UnitPrice > 0;
```

[결과 이미지를 넣어주세요]

 이 문으로 이름이 data인 테이블이 교체되었습니다.

## 11-7. RFM 스코어

### Recency

- **InvoiceDate** 컬럼을 연월일 자료형으로 변경하기

```
SELECT  
  DATE(InvoiceDate) AS InvoiceDay,  
  *  
FROM `trim-field-479202-b8.aiffel_dataset.data`;
```

[결과 이미지를 넣어주세요]

행	InvoiceDay	InvoiceNo	StockCode	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	Description
1	2011-01-18	541431	23166	74215	2011-01-18 10:01:00 UTC	1.04	12346	United Kingdom	MEDIUM CERAMIC TOP STORA...
2	2011-01-18	C541433	23166	-74215	2011-01-18 10:17:00 UTC	1.04	12346	United Kingdom	MEDIUM CERAMIC TOP STORA...
3	2010-12-07	537626	22729	4	2010-12-07 14:57:00 UTC	3.75	12347	Iceland	ALARM CLOCK BAKELIKE ORA...
4	2010-12-07	537626	22375	4	2010-12-07 14:57:00 UTC	4.25	12347	Iceland	AIRLINE BAG VINTAGE JET SET...
5	2010-12-07	537626	85116	12	2010-12-07 14:57:00 UTC	2.1	12347	Iceland	BLACK CANDELABRA T-LIGHT ...
6	2010-12-07	537626	22728	4	2010-12-07 14:57:00 UTC	3.75	12347	Iceland	ALARM CLOCK BAKELIKE PINK
7	2010-12-07	537626	22771	12	2010-12-07 14:57:00 UTC	1.25	12347	Iceland	CLEAR DRAWER KNOB ACRYLI...
8	2010-12-07	537626	84558A	24	2010-12-07 14:57:00 UTC	2.95	12347	Iceland	3D DOG PICTURE PLAYING CAR...
9	2010-12-07	537626	21064	6	2010-12-07 14:57:00 UTC	5.95	12347	Iceland	BOOM BOX SPEAKER BOYS
10	2010-12-07	537626	22497	4	2010-12-07 14:57:00 UTC	4.25	12347	Iceland	SET OF 2 TINS VINTAGE BATHR...
11	2010-12-07	537626	22774	12	2010-12-07 14:57:00 UTC	1.25	12347	Iceland	RED DRAWER KNOB ACRYLIC E...
12	2010-12-07	537626	85232D	3	2010-12-07 14:57:00 UTC	4.95	12347	Iceland	SET/3 DECOUPAGE STACKING ...
13	2010-12-07	537626	85167B	30	2010-12-07 14:57:00 UTC	1.25	12347	Iceland	BLACK GRAND BAROQUE PHOT...
14	2010-12-07	537626	21731	12	2010-12-07 14:57:00 UTC	1.65	12347	Iceland	RED TOADSTOOL LED NIGHT LI...
15	2010-12-07	537626	22725	4	2010-12-07 14:57:00 UTC	3.75	12347	Iceland	ALARM CLOCK BAKELIKE CHO...
16	2010-12-07	537626	84969	6	2010-12-07 14:57:00 UTC	4.25	12347	Iceland	BOX OF 6 ASSORTED COLOUR T...
17	2010-12-07	537626	71477	12	2010-12-07 14:57:00 UTC	3.25	12347	Iceland	COLOUR GLASS. STAR T-LIGHT ...
18	2010-12-07	537626	84997C	6	2010-12-07 14:57:00 UTC	3.75	12347	Iceland	BLUE 3 PIECE POLKADOT CUTL...
19	2010-12-07	537626	22772	12	2010-12-07 14:57:00 UTC	1.25	12347	Iceland	PINK DRAWER KNOB ACRYLIC ...
20	2010-12-07	537626	22726	4	2010-12-07 14:57:00 UTC	3.75	12347	Iceland	ALARM CLOCK BAKELIKE GREEN
21	2010-12-07	537626	22773	12	2010-12-07 14:57:00 UTC	1.25	12347	Iceland	GREEN DRAWER KNOB ACRYLI...
22	2010-12-07	537626	22212	6	2010-12-07 14:57:00 UTC	2.1	12347	Iceland	FOUR HOOK WHITE LOVEBIRDS
23	2010-12-07	537626	22195	12	2010-12-07 14:57:00 UTC	1.65	12347	Iceland	LARGE HEART MEASURING SP...
24	2010-12-07	537626	20782	6	2010-12-07 14:57:00 UTC	5.49	12347	Iceland	CAMOUFLAGE EAR MUFF HEA...
25	2010-12-07	537626	84997B	6	2010-12-07 14:57:00 UTC	3.75	12347	Iceland	RED 3 PIECE RETROSPOT CUTL...
26	2010-12-07	537626	22494	12	2010-12-07 14:57:00 UTC	1.25	12347	Iceland	EMERGENCY FIRST AID TIN
27	2010-12-07	537626	22805	12	2010-12-07 14:57:00 UTC	1.25	12347	Iceland	BLUE DRAWER KNOB ACRYLIC ...
28	2010-12-07	537626	20780	12	2010-12-07 14:57:00 UTC	4.65	12347	Iceland	BLACK EAR MUFF HEADPHONES
29	2010-12-07	537626	22775	12	2010-12-07 14:57:00 UTC	1.25	12347	Iceland	PURPLE DRAWERKNOB ACRYLI...
30	2010-12-07	537626	84997D	6	2010-12-07 14:57:00 UTC	3.75	12347	Iceland	PINK 3 PIECE POLKADOT CUTL...
31	2010-12-07	537626	22492	36	2010-12-07 14:57:00 UTC	0.65	12347	Iceland	MINI PAINT SET VINTAGE
32	2010-12-07	537626	21171	12	2010-12-07 14:57:00 UTC	1.45	12347	Iceland	BATHROOM METAL SIGN
33	2010-12-07	537626	22727	4	2010-12-07 14:57:00 UTC	3.75	12347	Iceland	ALARM CLOCK BAKELIKE RED
34	2011-01-26	542237	22729	4	2011-01-26 14:30:00 UTC	3.75	12347	Iceland	ALARM CLOCK BAKELIKE ORA...
35	2011-01-26	542237	22423	3	2011-01-26 14:30:00 UTC	12.75	12347	Iceland	REGENCY CAKESTAND 3 TIER
36	2011-01-26	542237	22725	4	2011-01-26 14:30:00 UTC	3.75	12347	Iceland	ALARM CLOCK BAKELIKE CHO...
37	2011-01-26	542237	21154	10	2011-01-26 14:30:00 UTC	1.25	12347	Iceland	RED RETROSPOT OVEN GLOVE
38	2011-01-26	542237	22417	24	2011-01-26 14:30:00 UTC	0.55	12347	Iceland	PACK OF 60 SPACEBOY CAKE C...
39	2011-01-26	542237	84558A	12	2011-01-26 14:30:00 UTC	2.95	12347	Iceland	3D DOG PICTURE PLAYING CAR...
40	2011-01-26	542237	84991	24	2011-01-26 14:30:00 UTC	0.55	12347	Iceland	60 TEATIME FAIRY CAKE CASES
41	2011-01-26	542237	21976	24	2011-01-26 14:30:00 UTC	0.55	12347	Iceland	PACK OF 60 MUSHROOM CAKE...
42	2011-01-26	542237	21041	6	2011-01-26 14:30:00 UTC	2.95	12347	Iceland	RED RETROSPOT OVEN GLOVE ...
43	2011-01-26	542237	20719	10	2011-01-26 14:30:00 UTC	0.85	12347	Iceland	WOODLAND CHARLOTTE BAG
44	2011-01-26	542237	21731	12	2011-01-26 14:30:00 UTC	1.65	12347	Iceland	RED TOADSTOOL LED NIGHT LI...
45	2011-01-26	542237	84625A	24	2011-01-26 14:30:00 UTC	0.85	12347	Iceland	PINK NEW BAROQUECANDLES...
46	2011-01-26	542237	84969	6	2011-01-26 14:30:00 UTC	4.25	12347	Iceland	BOX OF 6 ASSORTED COLOUR T...
47	2011-01-26	542237	22376	4	2011-01-26 14:30:00 UTC	4.25	12347	Iceland	AIRLINE BAG VINTAGE JET SET...
48	2011-01-26	542237	22134	12	2011-01-26 14:30:00 UTC	0.42	12347	Iceland	MINI LADLE LOVE HEART RED
49	2011-01-26	542237	22728	4	2011-01-26 14:30:00 UTC	3.75	12347	Iceland	ALARM CLOCK BAKELIKE PINK
50	2011-01-26	542237	84992	24	2011-01-26 14:30:00 UTC	0.55	12347	Iceland	72 SWEETHEART FAIRY CAKE C...

## 가장 최근 구매 일자를 MAX() 함수로 찾아보기

```
SELECT
  MAX(InvoiceDate) OVER () AS most_recent_date,
  DATE(InvoiceDate) AS InvoiceDay,
```

\*

FROM `trim-field-479202-b8.aiffel\_dataset.data`;

[결과 이미지를 넣어주세요]

행	most_recent_date	InvoiceDay	InvoiceNo	StockCode	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	Description
1	2011-12-09	2011-01-18	541431	23166	74215	2011-01-18 10:01:00 UTC	1.04	12346	United Kingdom	MEDIUM CERAMIC TOP STORA...
2	2011-12-09	2011-01-18	C541433	23166	-74215	2011-01-18 10:17:00 UTC	1.04	12346	United Kingdom	MEDIUM CERAMIC TOP STORA...
3	2011-12-09	2010-12-07	537626	22729	4	2010-12-07 14:57:00 UTC	3.75	12347	Iceland	ALARM CLOCK BAKELIKE ORA...
4	2011-12-09	2010-12-07	537626	22375	4	2010-12-07 14:57:00 UTC	4.25	12347	Iceland	AIRLINE BAG VINTAGE JET SET...
5	2011-12-09	2010-12-07	537626	85116	12	2010-12-07 14:57:00 UTC	2.1	12347	Iceland	BLACK CANDELABRA T-LIGHT ...
6	2011-12-09	2010-12-07	537626	22728	4	2010-12-07 14:57:00 UTC	3.75	12347	Iceland	ALARM CLOCK BAKELIKE PINK
7	2011-12-09	2010-12-07	537626	22771	12	2010-12-07 14:57:00 UTC	1.25	12347	Iceland	CLEAR DRAWER KNOB ACRYLI...
8	2011-12-09	2010-12-07	537626	84558A	24	2010-12-07 14:57:00 UTC	2.95	12347	Iceland	3D DOG PICTURE PLAYING CAR...
9	2011-12-09	2010-12-07	537626	21064	6	2010-12-07 14:57:00 UTC	5.95	12347	Iceland	BOOM BOX SPEAKER BOYS
10	2011-12-09	2010-12-07	537626	22497	4	2010-12-07 14:57:00 UTC	4.25	12347	Iceland	SET OF 2 TINS VINTAGE BATHR...
11	2011-12-09	2010-12-07	537626	22774	12	2010-12-07 14:57:00 UTC	1.25	12347	Iceland	RED DRAWER KNOB ACRYLIC E...
12	2011-12-09	2010-12-07	537626	85232D	3	2010-12-07 14:57:00 UTC	4.95	12347	Iceland	SET/3 DECOUPAGE STACKING ...
13	2011-12-09	2010-12-07	537626	85167B	30	2010-12-07 14:57:00 UTC	1.25	12347	Iceland	BLACK GRAND BAROQUE PHOT...
14	2011-12-09	2010-12-07	537626	21731	12	2010-12-07 14:57:00 UTC	1.65	12347	Iceland	RED TOADSTOOL LED NIGHT LI...
15	2011-12-09	2010-12-07	537626	22725	4	2010-12-07 14:57:00 UTC	3.75	12347	Iceland	ALARM CLOCK BAKELIKE CHO...
16	2011-12-09	2010-12-07	537626	84969	6	2010-12-07 14:57:00 UTC	4.25	12347	Iceland	BOX OF 6 ASSORTED COLOUR T...
17	2011-12-09	2010-12-07	537626	71477	12	2010-12-07 14:57:00 UTC	3.25	12347	Iceland	COLOUR GLASS. STAR T-LIGHT ...
18	2011-12-09	2010-12-07	537626	84997C	6	2010-12-07 14:57:00 UTC	3.75	12347	Iceland	BLUE 3 PIECE POLKADOT CUTL...
19	2011-12-09	2010-12-07	537626	22772	12	2010-12-07 14:57:00 UTC	1.25	12347	Iceland	PINK DRAWER KNOB ACRYLIC ...
20	2011-12-09	2010-12-07	537626	22726	4	2010-12-07 14:57:00 UTC	3.75	12347	Iceland	ALARM CLOCK BAKELIKE GREEN
21	2011-12-09	2010-12-07	537626	22773	12	2010-12-07 14:57:00 UTC	1.25	12347	Iceland	GREEN DRAWER KNOB ACRYLI...
22	2011-12-09	2010-12-07	537626	22212	6	2010-12-07 14:57:00 UTC	2.1	12347	Iceland	FOUR HOOK WHITE LOVEBIRDS
23	2011-12-09	2010-12-07	537626	22195	12	2010-12-07 14:57:00 UTC	1.65	12347	Iceland	LARGE HEART MEASURING SP...
24	2011-12-09	2010-12-07	537626	20782	6	2010-12-07 14:57:00 UTC	5.49	12347	Iceland	CAMOUFLAGE EAR MUFF HEA...
25	2011-12-09	2010-12-07	537626	84997B	6	2010-12-07 14:57:00 UTC	3.75	12347	Iceland	RED 3 PIECE RETROSPOT CUTL...
26	2011-12-09	2010-12-07	537626	22494	12	2010-12-07 14:57:00 UTC	1.25	12347	Iceland	EMERGENCY FIRST AID TIN
27	2011-12-09	2010-12-07	537626	22805	12	2010-12-07 14:57:00 UTC	1.25	12347	Iceland	BLUE DRAWER KNOB ACRYLIC ...
28	2011-12-09	2010-12-07	537626	20780	12	2010-12-07 14:57:00 UTC	4.65	12347	Iceland	BLACK EAR MUFF HEADPHONES
29	2011-12-09	2010-12-07	537626	22775	12	2010-12-07 14:57:00 UTC	1.25	12347	Iceland	PURPLE DRAWERKNOB ACRYLI...
30	2011-12-09	2010-12-07	537626	84997D	6	2010-12-07 14:57:00 UTC	3.75	12347	Iceland	PINK 3 PIECE POLKADOT CUTL...
31	2011-12-09	2010-12-07	537626	22492	36	2010-12-07 14:57:00 UTC	0.65	12347	Iceland	MINI PAINT SET VINTAGE
32	2011-12-09	2010-12-07	537626	21171	12	2010-12-07 14:57:00 UTC	1.45	12347	Iceland	BATHROOM METAL SIGN
33	2011-12-09	2010-12-07	537626	22727	4	2010-12-07 14:57:00 UTC	3.75	12347	Iceland	ALARM CLOCK BAKELIKE RED
34	2011-12-09	2011-01-26	542237	22729	4	2011-01-26 14:30:00 UTC	3.75	12347	Iceland	ALARM CLOCK BAKELIKE ORA...
35	2011-12-09	2011-01-26	542237	22423	3	2011-01-26 14:30:00 UTC	12.75	12347	Iceland	REGENCY CAKESTAND 3 TIER
36	2011-12-09	2011-01-26	542237	22725	4	2011-01-26 14:30:00 UTC	3.75	12347	Iceland	ALARM CLOCK BAKELIKE CHO...
37	2011-12-09	2011-01-26	542237	21154	10	2011-01-26 14:30:00 UTC	1.25	12347	Iceland	RED RETROSPOT OVEN GLOVE
38	2011-12-09	2011-01-26	542237	22417	24	2011-01-26 14:30:00 UTC	0.55	12347	Iceland	PACK OF 60 SPACEBOY CAKE C...
39	2011-12-09	2011-01-26	542237	84558A	12	2011-01-26 14:30:00 UTC	2.95	12347	Iceland	3D DOG PICTURE PLAYING CAR...
40	2011-12-09	2011-01-26	542237	84991	24	2011-01-26 14:30:00 UTC	0.55	12347	Iceland	60 TEATIME FAIRY CAKE CASES
41	2011-12-09	2011-01-26	542237	21976	24	2011-01-26 14:30:00 UTC	0.55	12347	Iceland	PACK OF 60 MUSHROOM CAKE...
42	2011-12-09	2011-01-26	542237	21041	6	2011-01-26 14:30:00 UTC	2.95	12347	Iceland	RED RETROSPOT OVEN GLOVE ...
43	2011-12-09	2011-01-26	542237	20719	10	2011-01-26 14:30:00 UTC	0.85	12347	Iceland	WOODLAND CHARLOTTE BAG
44	2011-12-09	2011-01-26	542237	21731	12	2011-01-26 14:30:00 UTC	1.65	12347	Iceland	RED TOADSTOOL LED NIGHT LI...
45	2011-12-09	2011-01-26	542237	84625A	24	2011-01-26 14:30:00 UTC	0.85	12347	Iceland	PINK NEW BAROQUECANDLES...
46	2011-12-09	2011-01-26	542237	84969	6	2011-01-26 14:30:00 UTC	4.25	12347	Iceland	BOX OF 6 ASSORTED COLOUR T...
47	2011-12-09	2011-01-26	542237	22376	4	2011-01-26 14:30:00 UTC	4.25	12347	Iceland	AIRLINE BAG VINTAGE JET SET...
48	2011-12-09	2011-01-26	542237	22134	12	2011-01-26 14:30:00 UTC	0.42	12347	Iceland	MINI LADLE LOVE HEART RED
49	2011-12-09	2011-01-26	542237	22728	4	2011-01-26 14:30:00 UTC	3.75	12347	Iceland	ALARM CLOCK BAKELIKE PINK
50	2011-12-09	2011-01-26	542237	84992	24	2011-01-26 14:30:00 UTC	0.55	12347	Iceland	72 SWEETHEART FAIRY CAKE C...

- 유저 별로 가장 큰 InvoiceDay를 찾아서 가장 최근 구매일로 저장하기

```
SELECT
CustomerID,
```

```
MAX(InvoiceDate) AS InvoiceDay  
FROM `trim-field-479202-b8.aiffel_dataset.data`  
WHERE CustomerID IS NOT NULL  
GROUP BY CustomerID;
```

[결과 이미지를 넣어주세요]

행	CustomerID	InvoiceDay
1	12346	2011-01-18
2	12347	2011-12-07
3	12348	2011-09-25
4	12349	2011-11-21
5	12350	2011-02-02
6	12352	2011-11-03
7	12353	2011-05-19
8	12354	2011-04-21
9	12355	2011-05-09
10	12356	2011-11-17
11	12357	2011-11-06
12	12358	2011-12-08
13	12359	2011-12-02
14	12360	2011-10-18
15	12361	2011-02-25
16	12362	2011-12-06
17	12363	2011-08-22
18	12364	2011-12-02
19	12365	2011-02-21
20	12367	2011-12-05
21	12370	2011-10-19
22	12371	2011-10-11
23	12372	2011-09-29
24	12373	2011-02-01
25	12374	2011-11-14
26	12375	2011-12-07
27	12377	2011-01-28
28	12378	2011-08-02
29	12379	2011-09-19
30	12380	2011-11-18
31	12381	2011-12-05
32	12383	2011-06-08
33	12384	2011-11-11
34	12386	2011-01-06
35	12388	2011-11-24
36	12390	2011-09-21
37	12391	2011-11-18
38	12393	2011-09-28
39	12394	2011-10-07
40	12395	2011-11-24
41	12397	2011-11-04
42	12398	2011-10-25
43	12399	2011-08-12
44	12401	2011-02-09
45	12402	2011-01-20
46	12403	2011-10-21
47	12405	2011-07-14
48	12406	2011-11-17
49	12407	2011-10-21
50	12408	2011-11-07

- 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 `trim-field-479202-b8.aiffel_dataset.data`
  WHERE CustomerID IS NOT NULL
  GROUP BY CustomerID
);
```


[결과 이미지를 넣어주세요]

행	CustomerID ▼	recency ▼
1	12375	2
2	12415	24
3	12778	19
4	12808	36
5	12870	366
6	12908	176
7	12923	64
8	13090	8
9	13092	70
10	13094	21
11	13210	93
12	13226	273
13	13313	22
14	13449	23
15	13466	100
16	13473	60
17	13647	9
18	13709	3
19	13735	24
20	13853	8
21	13854	8
22	14155	266
23	14273	51
24	14439	319
25	14898	23
26	15048	65
27	15262	323
28	15289	29
29	15478	40
30	15568	2
31	15649	336
32	15923	372
33	15937	64
34	16086	59
35	16113	183
36	16164	17
37	16202	360
38	16318	35
39	16351	332
40	16657	9
41	17205	53
42	17218	4
43	17306	10
44	17429	120
45	17542	269
46	17674	74
47	17701	252
48	18005	3
49	18219	2
50	12609	78

- 최종 데이터 셋에 필요한 데이터들을 각각 정제해서 이어붙이고 지금까지의 결과를 `user_r` 이라는 이름의 테이블로 저장하기

```
CREATE OR REPLACE TABLE `trim-field-479202-b8.aiffel_dataset.user_r` AS
SELECT
  CustomerID,
  EXTRACT(DAY FROM MAX(InvoiceDay) OVER () - InvoiceDay) AS recency
FROM (
  SELECT
    CustomerID,
    MAX(DATE(InvoiceDate)) AS InvoiceDay
  FROM `trim-field-479202-b8.aiffel_dataset.data`
  WHERE CustomerID IS NOT NULL
  GROUP BY CustomerID
);
```

[결과 이미지를 넣어주세요]

 이 문으로 이름이 user\_r인 새 테이블이 생성되었습니다.

## Frequency

- 고객마다 고유한 InvoiceNo의 수를 세어보기

```
SELECT
  CustomerID,
  COUNT(DISTINCT InvoiceNo) AS purchase_cnt
FROM `trim-field-479202-b8.aiffel_dataset.data`
WHERE CustomerID IS NOT NULL
GROUP BY CustomerID;
```

[결과 이미지를 넣어주세요]

행	CustomerID ▼	purchase_cnt ▼
1	12346	2
2	12347	7
3	12348	4
4	12349	1
5	12350	1
6	12352	8
7	12353	1
8	12354	1
9	12355	1
10	12356	3
11	12357	1
12	12358	2
13	12359	6
14	12360	3
15	12361	1
16	12362	13
17	12363	2
18	12364	4
19	12365	1
20	12367	1
21	12370	4
22	12371	1
23	12372	3
24	12373	1
25	12374	1
26	12375	3
27	12377	2
28	12378	1
29	12379	3
30	12380	5
31	12381	4
32	12383	6
33	12384	3
34	12386	2
35	12388	6
36	12390	1
37	12391	1
38	12393	4
39	12394	2
40	12395	15
41	12397	2
42	12398	1
43	12399	4
44	12401	1
45	12402	1
46	12403	1
47	12405	1
48	12406	3
49	12407	5
50	12408	8

- 각 고객 별로 구매한 아이템의 총 수량 더하기

```
SELECT
  CustomerID,
  SUM(Quantity) AS item_cnt
FROM `trim-field-479202-b8.aiffel_dataset.data`
WHERE CustomerID IS NOT NULL
GROUP BY CustomerID;
```

[결과 이미지를 넣어주세요]

행	CustomerID ▾	item_cnt ▾
1	12346	0
2	12347	2458
3	12348	2332
4	12349	630
5	12350	196
6	12352	463
7	12353	20
8	12354	530
9	12355	240
10	12356	1573
11	12357	2708
12	12358	242
13	12359	1599
14	12360	1156
15	12361	90
16	12362	2180
17	12363	408
18	12364	1499
19	12365	173
20	12367	172
21	12370	2349
22	12371	582
23	12372	788
24	12373	196
25	12374	339
26	12375	175
27	12377	942
28	12378	2529
29	12379	401
30	12380	1109
31	12381	772
32	12383	1515
33	12384	117
34	12386	354
35	12388	1462
36	12390	350
37	12391	293
38	12393	816
39	12394	808
40	12395	2092
41	12397	1250
42	12398	1389
43	12399	1173
44	12401	10
45	12402	92
46	12403	94
47	12405	849
48	12406	1809
49	12407	1403
50	12408	1382

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

```
CREATE OR REPLACE TABLE `trim-field-479202-b8.aiffel_dataset.user_
rf` AS

-- (1) 전체 거래 건수
WITH purchase_cnt AS (
  SELECT
    CustomerID,
    COUNT(DISTINCT InvoiceNo) AS purchase_cnt
  FROM `trim-field-479202-b8.aiffel_dataset.data`
  WHERE CustomerID IS NOT NULL
  GROUP BY CustomerID
),

-- (2) 구매한 아이템 총 수량
item_cnt AS (
  SELECT
    CustomerID,
    SUM(Quantity) AS item_cnt
  FROM `trim-field-479202-b8.aiffel_dataset.data`
  WHERE CustomerID IS NOT NULL
  GROUP BY CustomerID
)

-- (3) user_r와 합치기
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 `trim-field-479202-b8.aiffel_dataset.user_r` AS ur  
ON pc.CustomerID = ur.CustomerID;
```

[결과 이미지를 넣어주세요]



이 문으로 이름이 user\_rf인 새 테이블이 생성되었습니다.

## Monetary

- 고객별 총 지출액 계산 (소수점 첫째 자리에서 반올림)

```
SELECT  
  CustomerID,  
  ROUND(SUM(Quantity * UnitPrice), 1) AS user_total  
FROM `trim-field-479202-b8.aiffel_dataset.data`  
WHERE CustomerID IS NOT NULL  
GROUP BY CustomerID;
```

[결과 이미지를 넣어주세요]

행	CustomerID	user_total
1	12346	0.0
2	12347	4310.0
3	12348	1437.2
4	12349	1457.5
5	12350	294.4
6	12352	1265.4
7	12353	89.0
8	12354	1079.4
9	12355	459.4
10	12356	2487.4
11	12357	6207.7
12	12358	928.1
13	12359	6183.0
14	12360	2302.1
15	12361	174.9
16	12362	4665.6
17	12363	552.0
18	12364	1208.1
19	12365	320.7
20	12367	150.9
21	12370	3421.9
22	12371	1528.0
23	12372	1196.0
24	12373	324.6
25	12374	622.9
26	12375	375.4
27	12377	1548.1
28	12378	4008.6
29	12379	775.3
30	12380	2423.6
31	12381	1433.5
32	12383	1617.3
33	12384	406.2
34	12386	401.9
35	12388	2780.7
36	12390	504.8
37	12391	439.7
38	12393	1582.6
39	12394	1080.5
40	12395	2662.3
41	12397	2214.9
42	12398	1435.7
43	12399	1033.6
44	12401	69.3
45	12402	195.6
46	12403	391.7
47	12405	1390.4
48	12406	3280.8
49	12407	1507.1
50	12408	2587.6

- 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 `trim-field-479202-b8.aiffel_dataset.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 `trim-field-479202-b8.aiffel_dataset.user_rf` AS rf
LEFT JOIN (
  SELECT
    CustomerID,
    ROUND(SUM(Quantity * UnitPrice), 1) AS user_total
  FROM `trim-field-479202-b8.aiffel_dataset.data`
  WHERE CustomerID IS NOT NULL
  GROUP BY CustomerID
) AS ut
ON rf.CustomerID = ut.CustomerID;
```

[결과 이미지를 넣어주세요]



이 문으로 이름이 user\_rfm인 새 테이블이 생성되었습니다.

## RFM 통합 테이블 출력하기

- 최종 `user_rfm` 테이블을 출력하기

```
SELECT *  
FROM `trim-field-479202-b8.aiffel_dataset.user_rfm`;
```

[결과 이미지를 넣어주세요]

행	CustomerID	purchase_cnt	item_cnt	recency	user_total	user_average
1	12713	1	505	0	794.5	794.5
2	15520	1	314	1	343.5	343.5
3	13436	1	76	1	196.9	196.9
4	13298	1	96	1	360.0	360.0
5	14569	1	79	1	227.4	227.4
6	15195	1	1404	2	3861.0	3861.0
7	15471	1	256	2	454.5	454.5
8	14204	1	72	2	150.6	150.6
9	15318	1	642	3	312.6	312.6
10	12650	1	250	3	242.4	242.4
11	15992	1	17	3	42.0	42.0
12	16569	1	93	3	124.2	124.2
13	12442	1	181	3	144.1	144.1
14	16528	1	171	3	244.4	244.4
15	17914	1	457	3	329.4	329.4
16	12478	1	233	3	546.0	546.0
17	14578	1	240	3	168.6	168.6
18	14219	1	78	4	89.9	89.9
19	18015	1	157	4	120.0	120.0
20	17383	1	148	4	193.4	193.4
21	13790	1	748	4	348.8	348.8
22	16597	1	184	4	90.0	90.0
23	15097	1	170	4	248.1	248.1
24	12367	1	172	4	150.9	150.9
25	15773	1	311	5	635.7	635.7
26	16988	1	43	5	126.1	126.1
27	16535	1	356	5	593.8	593.8
28	13153	1	418	5	486.2	486.2
29	17936	1	257	5	380.1	380.1
30	18174	1	50	7	104.0	104.0
31	15539	1	410	7	538.5	538.5
32	13017	1	48	7	204.0	204.0
33	12587	1	50	7	104.0	104.0
34	13560	1	63	7	171.4	171.4
35	17942	1	38	7	77.5	77.5
36	14804	1	121	8	353.3	353.3
37	15904	1	84	8	164.7	164.7
38	16789	1	32	8	80.8	80.8
39	12953	1	84	9	329.9	329.9
40	17911	1	223	9	366.2	366.2
41	12966	1	75	9	160.2	160.2
42	18058	1	88	9	170.2	170.2
43	15790	1	113	10	218.8	218.8
44	14349	1	86	10	133.5	133.5
45	15148	1	187	10	301.3	301.3
46	13428	1	151	10	201.8	201.8
47	13349	1	224	10	197.3	197.3
48	14601	1	168	10	214.0	214.0
49	15619	1	136	10	336.4	336.4
50	15783	1	212	10	246.3	246.3


## 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 `trim-field-479202-b8.aiffel_dataset.user_data` AS
WITH unique_products AS (
  SELECT
    CustomerID,
    COUNT(DISTINCT StockCode) AS unique_products
  FROM `trim-field-479202-b8.aiffel_dataset.data`
  WHERE CustomerID IS NOT NULL
  GROUP BY CustomerID
)
SELECT ur.*, up.* EXCEPT (CustomerID)
FROM `trim-field-479202-b8.aiffel_dataset.user_rfm` AS ur
JOIN unique_products AS up
ON ur.CustomerID = up.CustomerID;
```

[결과 이미지를 넣어주세요]

 이 문으로 이름이 user\_data인 새 테이블이 생성되었습니다.

### 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 `trim-field-479202-b8.aiffel_dataset.user_data` AS
WITH purchase_intervals AS (
  -- (2) 고객 별 평균 구매 간격
  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 사이의 일수
    SELECT
      CustomerID,
      DATE_DIFF(
        InvoiceDate,
        LAG(InvoiceDate) OVER (PARTITION BY CustomerID ORDER BY InvoiceDate),
        DAY
      ) AS interval_
    FROM `trim-field-479202-b8.aiffel_dataset.data`
    WHERE CustomerID IS NOT NULL
  )
  GROUP BY CustomerID
)
SELECT u.*, pi.* EXCEPT (CustomerID)
FROM `trim-field-479202-b8.aiffel_dataset.user_data` AS u
LEFT JOIN purchase_intervals AS pi
ON u.CustomerID = pi.CustomerID;

```

[결과 이미지를 넣어주세요]



이 문으로 이름이 user\_data인 테이블이 교체되었습니다.

### 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 `trim-field-479202-b8.aiffel_dataset.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 `trim-field-479202-b8.aiffel_dataset.data`
  WHERE CustomerID IS NOT NULL
  GROUP BY CustomerID
)

SELECT
  u.* EXCEPT(total_transactions, cancel_frequency, cancel_rate),
  t.total_transactions,
  t.cancel_frequency,
  ROUND(
    SAFE_DIVIDE(t.cancel_frequency, t.total_transactions),
    2
  ) AS cancel_rate
FROM `trim-field-479202-b8.aiffel_dataset.user_data` AS u
LEFT JOIN TransactionInfo AS t
ON u.CustomerID = t.CustomerID;
```

[결과 이미지를 넣어주세요]



이 문으로 이름이 user\_data인 테이블이 교체되었습니다.

- 다양한 컬럼들을 활용하여 고객의 구매 패턴과 선호도를 보다 심층적으로 이해할 수 있도록 최종적으로 **user\_data** 를 출력하기

```
SELECT *
FROM `trim-field-479202-b8.aiffel_dataset.user_data`;
```

[결과 이미지를 넣어주세요]

행	CustomerID	purchase_cnt	item_cnt	recency	user_total	user_average	unique_products	average_interval	total_transactions	cancel_frequency	cancel_rate
1	17307	1	-144	365	-152.6	-152.6	1	0.0	1	1	1.0
2	18184	1	60	15	49.8	49.8	1	0.0	1	0	0.0
3	18133	1	1350	212	931.5	931.5	1	0.0	1	0	0.0
4	15389	1	400	172	500.0	500.0	1	0.0	1	0	0.0
5	12791	1	96	373	177.6	177.6	1	0.0	1	0	0.0
6	16344	1	18	158	101.1	101.1	1	0.0	2	0	0.0
7	17925	1	72	372	244.1	244.1	1	0.0	1	0	0.0
8	16257	1	1	176	21.9	21.9	1	0.0	1	0	0.0
9	17291	1	72	308	550.8	550.8	1	0.0	1	0	0.0
10	14119	1	-2	354	-19.9	-19.9	1	0.0	1	1	1.0
11	15118	1	1440	134	244.8	244.8	1	0.0	1	0	0.0
12	17752	1	192	359	80.6	80.6	1	0.0	1	0	0.0
13	17715	1	384	200	326.4	326.4	1	0.0	1	0	0.0
14	14576	1	12	372	35.4	35.4	1	0.0	1	0	0.0
15	15562	1	39	351	134.6	134.6	1	0.0	1	0	0.0
16	18174	1	50	7	104.0	104.0	1	0.0	1	0	0.0
17	13185	1	12	267	71.4	71.4	1	0.0	1	0	0.0
18	15657	1	24	22	30.0	30.0	1	0.0	1	0	0.0
19	13270	1	200	366	590.0	590.0	1	0.0	1	0	0.0
20	17923	1	50	282	207.5	207.5	1	0.0	1	0	0.0
21	13391	1	4	203	59.8	59.8	1	0.0	1	0	0.0
22	18113	1	72	368	76.3	76.3	1	0.0	1	0	0.0
23	14705	1	100	198	179.0	179.0	1	0.0	1	0	0.0
24	13747	1	8	373	79.6	79.6	1	0.0	1	0	0.0
25	13302	1	5	155	63.8	63.8	1	0.0	1	0	0.0
26	16138	1	-1	368	-8.0	-8.0	1	0.0	1	1	1.0
27	16093	1	20	106	17.0	17.0	1	0.0	1	0	0.0
28	16323	1	50	196	207.5	207.5	1	0.0	1	0	0.0
29	15524	1	4	24	440.0	440.0	1	0.0	1	0	0.0
30	17347	1	216	86	229.0	229.0	1	0.0	1	0	0.0
31	13366	1	144	50	56.2	56.2	1	0.0	1	0	0.0
32	16454	1	2	64	5.9	5.9	1	0.0	1	0	0.0
33	16148	1	72	296	76.3	76.3	1	0.0	1	0	0.0
34	16428	1	-1	81	-3.0	-3.0	1	0.0	1	1	1.0
35	14351	1	12	164	51.0	51.0	1	0.0	1	0	0.0
36	16061	1	-1	269	-29.9	-29.9	1	0.0	1	1	1.0
37	13120	1	12	238	30.6	30.6	1	0.0	1	0	0.0
38	12814	1	48	101	85.9	85.9	1	0.0	1	0	0.0
39	15313	1	25	110	52.0	52.0	1	0.0	1	0	0.0
40	15195	1	1404	2	3861.0	3861.0	1	0.0	1	0	0.0
41	15316	1	100	326	165.0	165.0	1	0.0	1	0	0.0
42	12943	1	-1	301	-3.8	-3.8	1	0.0	1	1	1.0
43	12603	1	56	21	613.2	613.2	1	0.0	1	0	0.0
44	18141	1	-12	360	-35.4	-35.4	1	0.0	1	1	1.0
45	17102	1	2	261	25.5	25.5	1	0.0	1	0	0.0
46	15510	1	2	330	250.0	250.0	1	0.0	1	0	0.0
47	15668	1	72	217	76.3	76.3	1	0.0	1	0	0.0
48	16765	1	4	294	34.0	34.0	1	0.0	1	0	0.0
49	15753	1	144	304	79.2	79.2	1	0.0	1	0	0.0
50	16579	1	-12	365	-30.6	-30.6	1	0.0	1	1	1.0

## 회고

[회고 내용을 작성해주세요]

Keep :

- 모르는 부분은 바로 AI에게 질문하며 해결한 점.
- 단순히 정답을 붙여넣는 것이 아니라, 왜 그런 결과가 나오는지 계속 이해하려고 노력했습니다.
- 오류가 나면 그대로 넘기지 않고 원인을 찾으며 수정.
- SQL 문제를 순서대로 따라가며 점진적으로 개념을 익힌 방식이 도움이 되었습니다.

Problem :

- SQL 구조 자체가 처음이라, JOIN·GROUP BY·집계 함수 등 기본 개념이 직관적으로 다가오지 않았아 힘들었습니다.
- 문제들이 단계별로 연결되어 있는데, 그 흐름을 제대로 파악하기 어려웠습니다.
- BigQuery에서 CREATE OR REPLACE로 테이블을 교체하는 구조 때문에 “이게 통합 인지, 교체인지” 헷갈리는 부분 등 어려움이 많았습니다.
- 설명 없이 쿼리만 보면 의미를 이해하기 어려웠고, 결국 AI의 설명을 기반으로 다시 이해하면서 진행하였고 물어볼때는 이해하고 넘어갔지만 다시 보면 헷갈립니다.

Try :

- SQL 문법을 한글 문장으로 직접 번역해보면서 “내 언어로 이해하는 연습”을 많이 해야 할것 같습니다.
- 앞으로는 학습 중 작성하는 모든 SQL 쿼리에 주석을 붙여 내가 다시 봐도 이해 가능한 형태로 정리할 계획.
- 노선에 문제·쿼리·설명·결과·내 해석을 정리하는 개인 템플릿을 만들어야 겠다 생각했습니다.
- GitHub 제출뿐 아니라, 앞으로의 학습 기록도 저장해서 변화 과정을 추적해보려고 합니다.