

## 1. 전처리

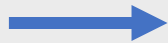
1. 결측값 (none)을 Null값(공백)으로 변경
2. 데이터 타입 변경(모든 컬럼 -> text)

## 2. 데이터 Import

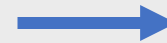
1. purchase\_data.csv -> purchase\_data 테이블
2. visit\_data .csv -> visit\_data 테이블
3. user\_action\_data .csv -> user\_action\_data 테이블

3. **데이터 이해:** 각 고객이 다양한 출처를 통해 여러 차례의 visit 후 purchase하는 여정 데이터. (구매가 없을 수도 있다)

browser\_id



visit



purchase

source

direct

daum

naver

instagram

youtube

google

kakao

source

direct

naver

instagram

youtube

google

kakao

# 사용자별 첫 번째 유입출처(source)와 마지막 유입출처(source)를 구한 결과

## Window 함수 사용한 쿼리

```
197 SELECT distinct browser_id,  
198 FIRST_VALUE(source) OVER (PARTITION BY browser_id ORDER BY datetime ROWS UNBOUNDED PRECEDING) as first_source,  
199 LAST_VALUE(source) OVER (PARTITION BY browser_id ORDER BY datetime ROWS BETWEEN UNBOUNDED PRECEDING AND UNBOUNDED FOLLOWING) as last_source  
200 FROM user_action_data  
201 order by browser_id;
```

{{ }}

≡

⚡

☒ LIMIT 1000

Save\*

Execute

Table + Add Visualization

browser_id	first_source	last_source
00008afde9465ff034eb48546bb486ba0519735fc75a95d06b9d20796243d0d6	direct	direct
00009aabd3617485078ed5cf382f7ab028586e2b3de34475695a3ec27d562f41	naver	direct
0001252c0bd99a6687406b4ff6ab9b9e00939f3424980ed8b7475d1d17b6acf7	direct	direct
0003d7f5ccf66276b0c4f6ac5786fa9f9d0b94d0fd6d4704e9949d840b5753ca	direct	direct
000a7350e8ac8544c5bb2bb89f00a457cd02a5f65d27899eadbffac2c1e81657	direct	direct
000c4580c2906689d15f1a3b8b34c8762b7a9b5c55d314abab4f2b17e976ae5b	direct	direct
000d915eb5f87cbbfdc21c4b25b3d398c67935a70ea29688de3146d6b215c9df	direct	direct
000e622b37decd71d63e622a455d019497cfea464afeeb09e9575a3328fbf948	direct	direct
0011fa8e668612ca5157467661b8b534bec406ba3db3ab0685e700cfb82e9e9d	naver	naver

## 사용자별 첫 번째 유입출처(source)와 마지막 유입출처(source)를 구한 결과

### Window 함수 사용하지 않은 쿼리

```
22 select distinct full_first.browser_id, full_first.first_source, full_last.last_source
23 from
24 (select full.browser_id, full.source as 'first_source'
25 from user_action_data full JOIN (
26     select min(datetime) min_date, browser_id from user_action_data group by browser_id) as first
27     on full.browser_id=first.browser_id
28     where full.datetime=first.min_date) full_first
29 join
30 (select full.browser_id, full.source as 'last_source'
31 from user_action_data full
32     JOIN (select max(datetime) max_date, browser_id from user_action_data group by browser_id) as last
33     on full.browser_id=last.browser_id
34     where full.datetime=last.max_date) full_last
35 on full_first.browser_id=full_last.browser_id;
```

{{ }}



LIMIT 1000

Save\*

Execute

Table

+ Add Visualization

browser_id	first_source	last_source
eca50c7a5cde0ab2ce52141405613c1577384e1f4ff7338e8792d5b0a4044dcd	direct	direct
9a4d435ab83d6a019a967ff0350997508dfbbb920786bc55d896868668c4e771	direct	direct
2a403e6bca25b766e8b4677791963cfc9c92ac42cdffbee1b641ea295d3348fd	direct	direct
818e3a7eec2ab0889de1619a56289b85c8bd31cab256718a71ef5d6b67dacca3	direct	direct
fd9e91b971e1d709e76718625ef4c8928b999e40cd8fc0473a4001e674e63ef3	direct	direct

# First Click 기여도 모델을 기반으로 구한 각 유입 출처(source)별 전환율 분석

## First Click 기여도 모델 기반

First Click 기여도 모델은 고객이 유입한 첫 번째 출처가 전환 가치에 100% 기여했다고 간주한다.

다음 쿼리 결과에 따르면 957번의 방문 중 79번의 구매가 발생하여 구매전환율이 8.26인 youtube의 구매 전환율이 가장 높다. 이는 youtube를 통해 해당 상품/브랜드를 처음 인지하여 구매까지 이어지는 비율이 높음을 의미한다. 따라서 youtube를 통한 상품/브랜드 광고에 예산을 투자하여 구매 증가를 기대해 볼 수 있다.

하지만 First Click 기여도 모델의 상품/브랜드 인지 후 구매 이전까지의 보조채널의 기여효과를 무시한다는 단점과 함께 해당 데이터에서 상품/브랜드를 직접 검색하여 방문, 구매한 절대적인 수치가 높으므로 인지도를 알리는 단계는 아닌 것으로 생각되기 때문에 다른 기여도 모델에 따른 분석이 필요한 것으로 보인다.

```
143 SELECT first_source as '유입 출처', max(visit) as '방문수', max(purchase) as '구매수',  
144        (max(purchase) / max(visit) * 100) as '구매 전환율'  
145 FROM (  
146 SELECT first.first_source,  
147        IF(full.event_type='purchase', count(full.browser_id), NULL) as purchase,  
148        IF(full.event_type='visit', count(full.browser_id), NULL) as visit  
149 FROM user_action_data full JOIN (  
150 SELECT distinct browser_id,  
151        FIRST_VALUE(source) OVER (PARTITION BY browser_id ORDER BY datetime ROWS UNBOUNDED PRECEDING) as first_source  
152 FROM user_action_data) first  
153 ON full.browser_id = first.browser_id  
154 GROUP BY first.first_source, full.event_type  
155 ORDER BY full.event_type, first.first_source) A  
156 GROUP BY A.first_source ORDER BY 구매 전환율;
```

{{ }}

≡

↺

☒ LIMIT 1000

Save\*

Execute

Table

+ Add Visualization

유입 출처	방문수	구매수	구매전환율
daum	40		
kakao	141	1	0.71
instagram	2,562	23	0.90
google	978	21	2.15
direct	85,406	5,102	5.97
naver	12,301	796	6.47
youtube	957	79	8.26

# Last Click 기여도 모델을 기반으로 구한 각 유입 출처(source)별 전환율 분석

## Last Click 기여도 모델 기반

Last Click 기여도 모델은 구매 전환 이전에 마지막으로 유입한 출처가 전환 가치에 100% 기여했다고 간주한다.

다음 쿼리 결과에 따르면 1227번의 방문 중 90번의 구매가 발생한 youtube의 구매 전환율이 7.34로 가장 높다. 이는 youtube에서 마지막으로 구매를 결심하는 고객이 많음을 의미한다. 따라서 youtube를 통한 구매 절차를 간소화하거나 할인율 등의 구매 촉진 요소를 첨가해 구매수 증가를 기대해 볼 수 있다.

또한 방문수, 구매수의 절대적인 수치가 가장 높은 direct 유입 고객들의 구매전환율을 높이는 방법을 고려해볼 수도 있다. direct의 방문수에 비해 구매전환율이 낮은 것은 구입 직전에 고민하는 경우가 많음을 의미하므로 할인 기간이 임박함을 표시하는 등 구매에 확신이 들도록 하는 방법을 찾아볼 수 있다.

```
159 SELECT last_source as '유입 출처', max(visit) as '방문수', max(purchase) as '구매수',  
160        (max(purchase) / max(visit) * 100) as '구매 전환율'  
161 FROM (  
162 SELECT last.last_source,  
163        IF(full.event_type='purchase', count(full.browser_id), NULL) as purchase,  
164        IF(full.event_type='visit', count(full.browser_id), NULL) as visit  
165 FROM user_action_data full JOIN (  
166 SELECT distinct browser_id,  
167        LAST_VALUE(source) OVER (PARTITION BY browser_id ORDER BY datetime ROWS UNBOUNDED PRECEDING) as last_source  
168 FROM user_action_data) last  
169 ON full.browser_id = last.browser_id  
170 GROUP BY last.last_source, full.event_type  
171 ORDER BY full.event_type, last.last_source) A  
172 GROUP BY A.last_source ORDER BY 구매 전환율;
```

{{ }} 쿼리 LIMIT 1000

Save\*

Execute

Table + Add Visualization

유입 출처	방문수	구매수	구매전환율
daum	47		
kakao	149	1	0.67
instagram	2,857	28	0.98
google	1,432	46	3.21
direct	89,591	5,511	6.15
naver	14,481	907	6.26
youtube	1,227	90	7.34

## 각 유입 출처(source)별 지원 전환수, 직접 전환수 분석

```
161 SELECT direct.source as '유입 출처', direct.direct as '직접 전환수', support.support as '지원 전환수'
162 FROM
163 (SELECT source, count(*) as 'direct'
164 FROM purchase_data
165 GROUP BY source) direct,
166 (SELECT source, count(*) as 'support'
167 FROM visit_data
168 GROUP BY source) support
169 WHERE support.source = direct.source
170 ORDER BY direct.direct DESC
```

{{ }} LIMIT 1000

Save\*

Execute

Table

+ Add Visualization

유입 출처	직접 전환수	지원 전환수
direct	5,317	86,925
naver	596	10,926
youtube	79	882
instagram	23	2,568
google	6	907
kakao	1	137

지원 전환수를 '구매' 전환이 일어나기 이전의 채널로 설정하여 event\_type이 'visit'인 모든 행을 count한 결과인데, '구매' 전환이 발생하지 않은 (곧, visit만 있고 purchase로 이어지지 않은) 값은 제외되지 않은 결과이다. 우선은, 모든 id에서 구매가 이루어진 것을 전제로 하여 쿼리 결과를 분석했다.

유입 출처가 direct인 직접/지원 전환수가 압도적으로 높음을 알 수 있다. '구매' 전환을 위한 직접 유입이 많은 것은 해당 상품/브랜드에 대한 인지도가 이미 꽤 높아 다른 플랫폼 광고를 통하지 않고 바로 구매하거나, 고민하게 만든다.

따라서 상품/브랜드에 대한 좋은 이미지를 확고하게 굳히기 위해 direct 페이지의 사용성을 높이고 좋은 이미지 형성에 투자해 볼 수 있다. 또한 다른 플랫폼을 통한 유입이 아닌, TV, 야외 광고 등을 통한 유입인지 확인해 볼 필요가 있다. TV, 라디오, 옥외 광고 등의 효과인 것으로 분석될 경우 해당 광고에 더 투자해볼 수 있다.

또한 direct 다음으로 직접/지원 전환수가 높은 naver에 광고예산을 더 많이 투자하여 구매 전환을 높일 수 있을 것으로 생각된다.