

# 02 Data Selection

## PC & Mobile 로그 데이터

- 수집기간: 2014.04.01 ~ 2014.06.30
- 크기: 34GB
- 주요 컬럼명
  - UID: 패널 고유의 ID값
  - Time: 사이트의 클릭 날짜/시간
  - Full\_URL: 패널이 방문한 사이트의 full url
  - Domain: 패널이 방문한 사이트의 도메인 정보
  - ownership\_1: 사이트 오너십 기준 카테고리 대분류
  - ownership\_2: 사이트 오너십 기준 카테고리 중분류

## URL 클릭스트림 데이터 활용

## 구매행태 서베이 응답 데이터

- 수집시기: 2014.07.14
- 크기: 1066명의 패널 대상, 151개의 질문 응답
- 주요 서베이 질문
  - Q. 귀하께서 최근 3개월 이내에 구매한 제품을 모두 선택해주십시오.  
A1-가전제품 A2-의류/잡화 A3-화장품 A4-금융상품
  - Q. 귀하께서는 해당 제품을 몇 월에 구매하셨습니다?  
A1-4월 A2-5월 A3-6월 A4-7월
  - Q. 구매하신 시기는 해당 월 중 언제쯤 이십니까?  
A1-초순 A2-중순 A3-하순

## 구매 시기 Label 추출

데이터 핸들링과 응답 데이터 신뢰도를 고려해,  
서베이 응답 시기와 가장 가까우면서 한 달동안 완전히 수집된 달인 6월 로그 데이터 활용  
로그 데이터와 서베이 응답 데이터에 동시에 존재하는 UID(1063개)만을 고려

약 3495만 rows