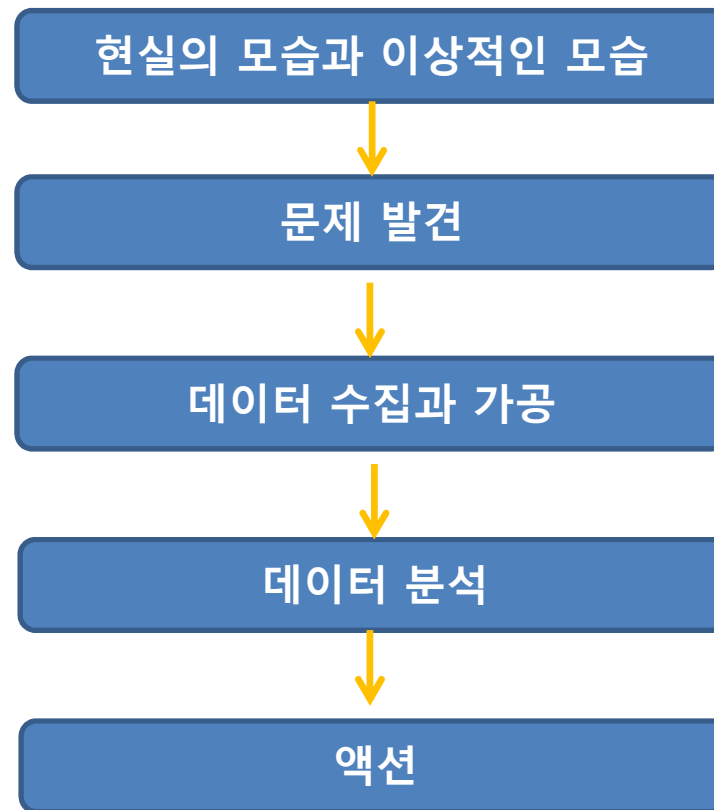




# 비즈니스 데이터 분석 모델링

- 비즈니스 데이터 분석 모델링



# 비즈니스 데이터 분석 모델링

- **비즈니스 데이터 분석 모델링**

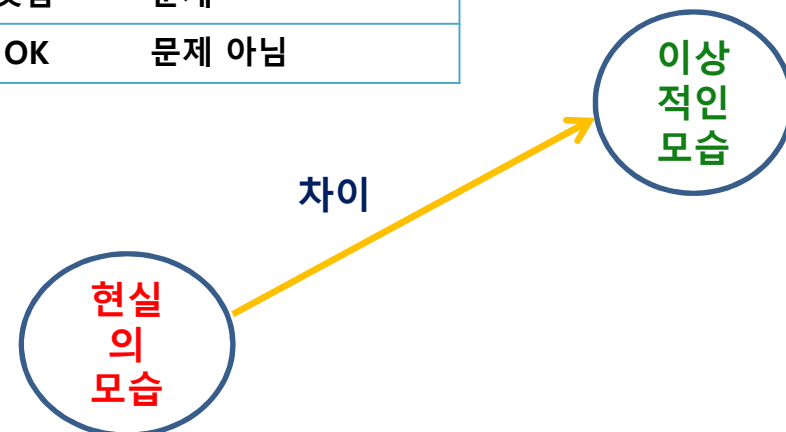
- 해결해야 할 문제에 맞게 분석 기법을 설계하고 실행하는 것이 중요
- 현실의 모습과 이상적인 모습을 정확히 파악하여 둘의 차이를 이루는 요소에 대한 문제 발견을 실시
- 문제가 되는 요소는 복수일 경우가 많으므로 각각에 대해 어떻게 데이터 수집과 가공을 하지 검토한 후에 실시
- 데이터 분석
  - 분석 대상의 구조를 분해하여 각 요소가 지닌 영향력의 크기를 파악
  - 요소 간의 비교를 통해 현실과 이상의 차이를 이루는 주요인을 찾아냄
- 액션을 제안하여 추진
  - 노동 비용이나 금전적 비용을 고려

# 비즈니스 데이터 분석 모델링

- 현실의 모습과 이상적인 모습

- '문제'는 어떤 시점에 어떤 비즈니스가 처한 환경의 '이상적인 모습'에 의해 바뀔 수 있다
- '이상적인 모습'을 당사자 간에 공유하는 것이야말로 효과적인 데이터 분석 수행을 위해 필수적
- 최대한 많은 비즈니스 당사자와 분석자 사이에 이상적인 모습이 공유되지 않으면 비즈니스에 가치가 떨어지는 분석 결과가 나오게 됩니다.

현상	전체	이상적인 모습	문제인가?
매상이 떨어졌다	매상 비율이 낮다	지금 상태로 OK	문제 아님
	매상 비율이 높다	호조기였을 때 매상	문제
매상이 올랐다	광고비율이 높다	광고비율을 낮춤	문제
	광고비율이 적절하다.	지금 상태로 OK	문제 아님



# 비즈니스 데이터 분석 모델링

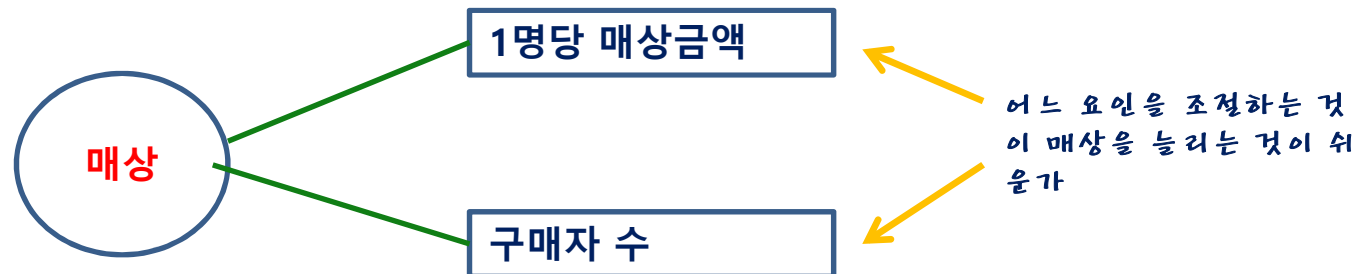
- 문제 발견을 위한 세가지 관점

- 크기 비교

- 이상적인 모습과 현실의 모습의 차이에 대한 요인의 영향도 비교
    - 영향도가 작다면, 본질적인 문제가 아니라고 판단하고, 다른 요인을 찾아야 함

- 분해해서 보기

- Mutually Exclusive : 상호 배제 (요소가 서로 중복되지 않고)
    - Collectively Exhaustive : 전체 포괄 (빠짐없이 모으도록)
    - MECE에 따라 분해한 각 요소의 시계열 추이를 그래프로 확인하면 어느 것이 떨어지고 있는지 알 수 있다.
    - 조절할수 있는 요소인지 조절할 수 없거나 힘든 요소인지 구별하는 것이 중요
    - MECE에 따라 분해한 각 요소 중에 조절할 수 있는 요소가 포함되어 있으면 문제의 원인을 쉽게 파악할 수 있다



# 비즈니스 데이터 분석 모델링

- 문제 발견을 위한 세가지 관점

- 비교해서 보기

- 문제가 있는 데이터와 문제가 없는 데이터를 비교해서 그 차이가 무엇에 의해 발생하는지 찾아내는 방법

시계열로 비교	전일과 당일 비교 전주와 금주 비교 과거에 같은 캠페인을 실시했던 시기와 이번에 캠페인을 실시한 시기 비교
다른 유사 상품이나 서비스와 비교	경쟁사와 매상 비교 자사의 서비스 간 이익 비교
유저 속성을 이용한 분석	20대와 50대는 구매의욕이 다름 (연령별) 남성과 여성 간에는 구매율이 다름(성별) 한국과 미국에서는 인기 있는 색상이나 형태가 다름 (지역별)

# 비즈니스 데이터 분석 모델링

- 데이터 수집과 가공

- 문제를 검증하기 위해 어떤 데이터가 필요한가?
- 분석자가 사요할 수 있는 곳에 필요한 데이터가 보존되어 있는가?
- 분석자가 신청하면 필요한 데이터를 사용할 수 있는가?
- 필요한 데이터가 보존되어 있지 않을 경우, 새로 데이터를 취득할 수 있는가?
- 필요한 데이터가 보존되어 있지 않고 또한 새로 취득하기에는 시간이나 비용이 많이 들 경우 그것을 대체할 수 있는 다른 데이터가 있는가?

# 비즈니스 데이터 분석 모델링

- 데이터 가공

- 데이터 분석의 목적, 데이터 보존 상태, 데이터 형태 등에 따라 달라지기 때문에 기본적으로는 개별적으로 다루어야 한다
- 데이터 결합
  - 분석하고자 하는 여러 데이터를 하나의 데이터로 모으는 처리
- 판정용 변수 작성
- 연속된 수치를 이산화 변수로 작성



# 비즈니스 데이터 분석 모델링

- 데이터 분석

- 문제의 종류에 따라 '의사결정지원'과 '자동화.최적화'
- 의사결정지원
  - 문제 해결을 위한 행동을 사람이 결정하고 실행할 수 있도록 지원하는 것이 주목적
  - 분석 결과를 사람이 이해할 수 있고 적절한 판단을 할 수 있는지가 중요
  - 단순하고 이해하기 쉬운 모델이 효과적 (단순집계, 크로스 집계, 유효한 경우가 많음)
  - 예측모델은 어떤 요인이 결과에 어떻게 영향을 끼치는지 인과관계를 명확히 할 수 있게 함
  - 회귀분석, 공분산구조분석은 어떤 요인이 1만큼 변화했을 때 전체 현상이 어떻게 변동하는지를 조사
- 자동화.최적화
  - 문제 해결을 위한 행동을 컴퓨터에 실행시키기 위한 알고리즘을 구축하는 것이 주목적
  - 알고리즘의 계산량과 정확도가 중요시됨
  - 기계학습을 이용하여 과거에 축적해놓은 로그 데이터로부터 패턴이나 규칙을 찾아서 모델을 작성

# 비즈니스 데이터 분석 모델링

- 데이터 분석

전체	의사결정지원	자동화 최적화
목적	사람의 행동결정을 지원	컴퓨터의 행동결정을 지원
목표	의사소통비용 절감	추정 정확도 향상, 계산량 삭감
주로 사용되는 기법	단순집계, 크로스 집계	기계학습, 알고리즘 구축

# 비즈니스 데이터 분석 모델링

- 액션



의사 결정 지원의 경우

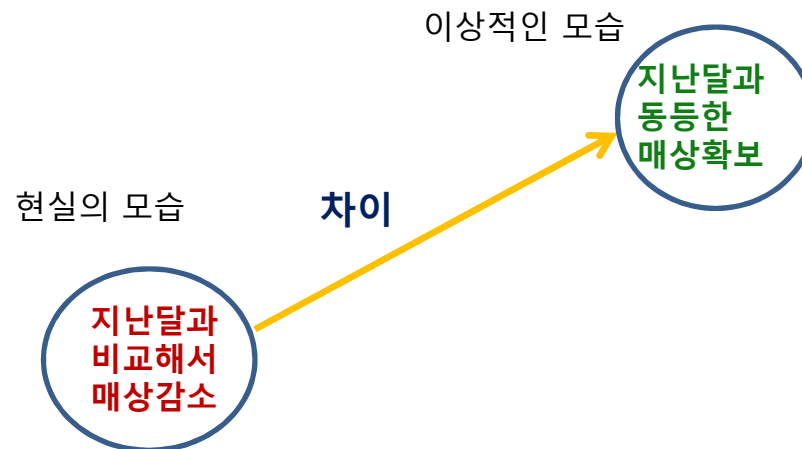


자동화.최적화의 경우

# 데이터 분석 workshop2

- 소셜게임의 매상 분석

소셜게임 회사가 운영하는 소셜게임 '퍼즐 & 검은 고양이 컬렉션'은 지금까지 순조롭게 매상이 늘었으나, 이달 들어 매상이 감소했습니다.  
(total payment가 228만에서 211만으로 감소)  
시장환경이나 게임의 상태를 고려하면 아직은 더 성장할 것으로 예상했기 때문에 사내에서 큰 문제로 인식하고 있습니다.  
원인을 밝혀서 대책을 세우기 위해 분석 담당자에게 의뢰했습니다.



# 데이터 분석 workshop2

## • 소셜게임의 매상 분석

매상 감소의 원인 가설 :

1. 광고에 문제가 생겼다
2. 매월 테마를 바꿔서 개최하던 게임의 이벤트에 문제가 생겼다.

마케팅 부서와 게임 개발 부서에 질의 응답을 한 결과 :

1. 이달에는 예산 관계로 지난달만큼 광고를 하지 못했다.
2. 이벤트 내용은 지난달과 거의 비슷하다.

문제

- 지난달에 비해 매상이 감소했다 (사실)
- 이달은 지난달에 비해 광고를 적게 했다 (사실)
- 그래서 신규 유저수가 줄어들었다. (가설)

분석 스토리

- 퍼즐컬렉션의 매상이 지난달에 비해 줄어들었다(사실)
- 매상 내역을 보았더니 신규 유저의 매상이 감소했다 (가설)
- 광고를 지난달과 동등한 수준으로 실시한다 (해결책)

# 데이터 분석 workshop2

- 소셜게임의 매상 분석
  - 데이터 수집

데이터 수집 :

- DAU(Daily Active User) 데이터 - 하루에 한번 이상 게임을 이용한 유저
- DPU(Daily Payment User) 데이터 - 하루에 1원 이상 지불한 유저
- Install 데이터 - 유저별로 게임을 이용하기 시작한 날짜 기록

DAU

데이터 내용	데이터형	R 언어에서의 표기
로그인한 날	string (문자열)	log_date
앱 이름	string (문자열)	app_name
유저 ID	int (수치)	user_id

# 데이터 분석 workshop2

- 소셜게임의 매상 분석
  - 데이터 수집

DPU

데이터 내용	데이터형	R 언어에서의 표기
과금일	string (문자열)	log_date
앱 이름	string (문자열)	app_name
유저 ID	int (수치)	user_id
과금액	int (수치)	payment

lstart

데이터 내용	데이터형	R 언어에서의 표기
이용 시작한 날	string (문자열)	install_date
앱 이름	string (문자열)	app_name
유저 ID	int (수치)	user_id

# 데이터 분석 workshop2

- 소셜게임의 매상 분석

- 분석에 필요한 데이터 가공(전처리)

- 데이터를 분석할 수 있도록 데이터를 정돈하는 가공 작업
    - 노이즈로 분류될 법한 데이터를 삭제하기 위한 가공 작업

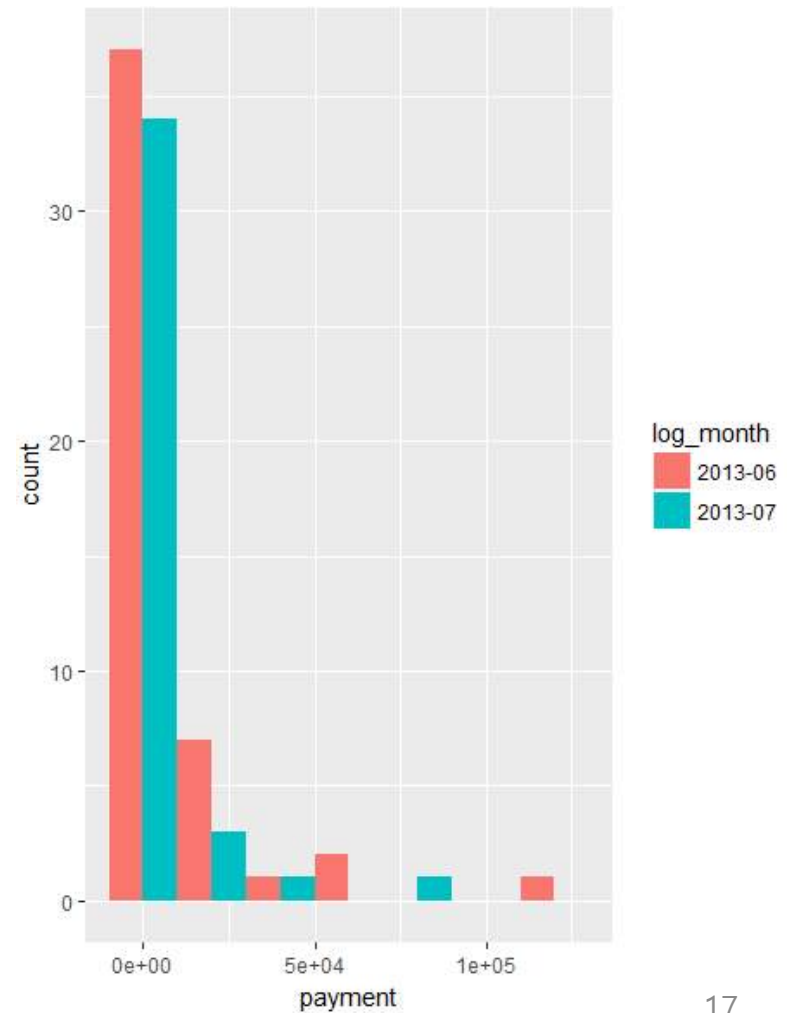
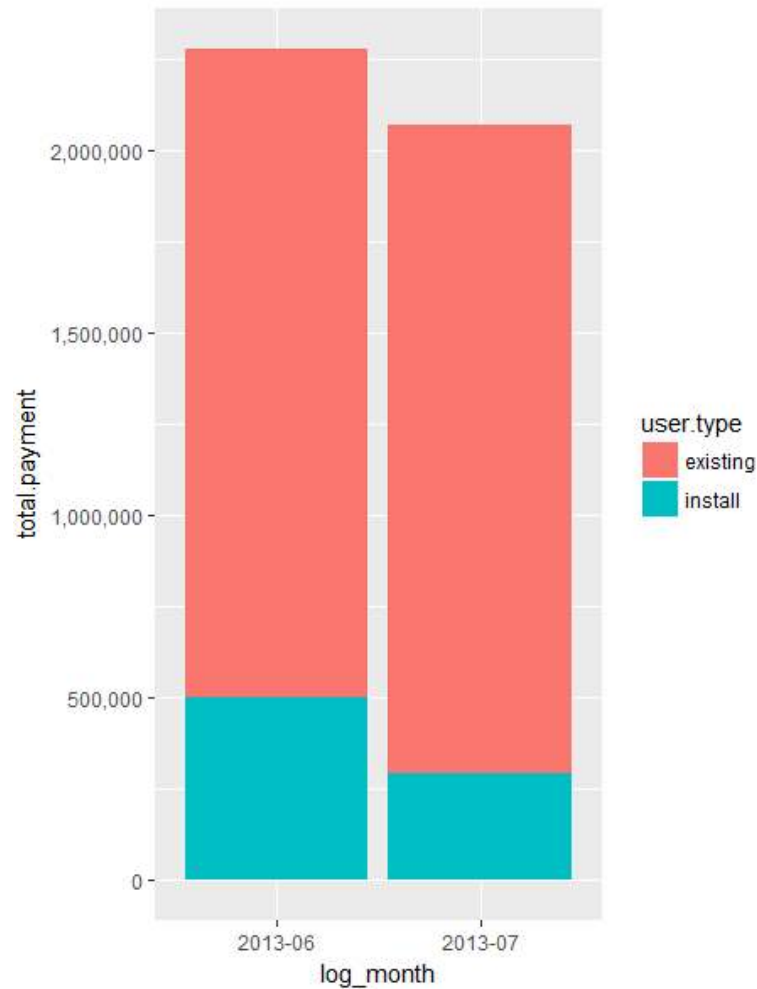
매상 감소가 신규 유저의 영향인지 아닌지 판단하기 위한 데이터 가공 순서 :

1. 유저(DAU) 데이터에 이용시작(Install) 데이터 결합하기 (유저 ID 키)
2. 1단계에서 결합된 데이터에 과금(DPU) 데이터 결합하기 (유저 ID와 과금일 키)  
결합된 레코드뿐만 아니라 결합되지 않은 레코드도 남기도록 한다
3. 비과금 유저의 과금액에 0 넣기
4. 매월 유저별로 집계한 데이터로 변환
5. 4단계의 월차 데이터에 신규 유저인지 기존 유저인지 구분하는 항목 추가하기  
2013년 6월과 2013년 7월 데이터를 기존 유저와 신규 유저로 나누어 과금액을 집계
6. 히스토그램으로 시각화



# 데이터 분석 workshop2

- 소셜게임의 매상 분석



# 데이터 분석 workshop3

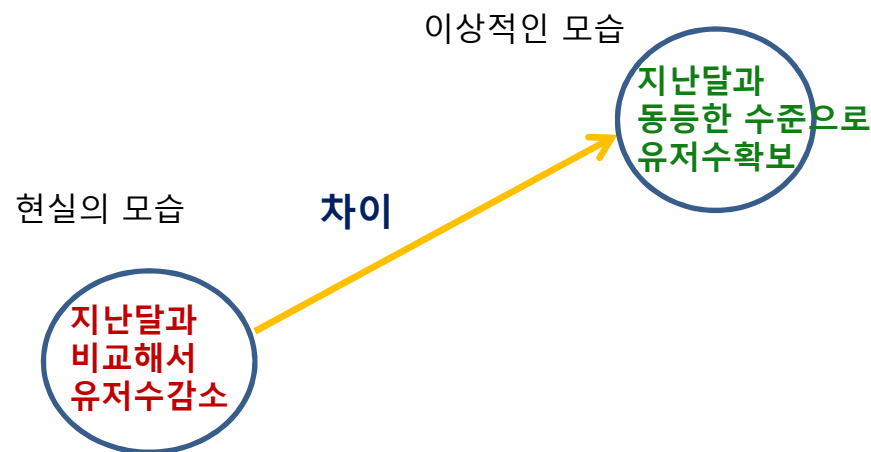
- 소셜게임의 이용자 탈퇴 이유 분석

퍼즐컬렉션 서비스를 시작한지 1년 3개월이 지났습니다. 서비스 직후에는 사전 등록자를 중심으로 유저수가 크게 증가하였고, 몇 주 후에는 심각한 버그 때문에 유저가 탈퇴하다가, 1년 후에는 광고 효과로 다시 유저수가 증가하였고, 그 후에는 그 광고를 보고 가입했던 유저드리 조금씩 떠나는 등 유저수가 단기적으로는 오르락 내리락하면서 월 단위로 반년 정도 계속해서 상승하였습니다.

그 후 8개월동안 유저수가 정점을 찍은 상태로 유지되었습니다.

이달 들어 퍼즐컬렉션 유저수가 크게 줄어들었고, 회사를 대표하는 히트 게임이었던 만큼 경영진을 중심으로 유저수의 감소를 큰 문제로 인식하고 있습니다.

원인을 밝혀서 대책을 세우기 위해 분석 담당자에게 의뢰했습니다.



# 데이터 분석 workshop3

## • 소셜게임의 이용자 탈퇴 이유 분석

가설 :

1. 광고에 문제가 있어서 신규 유저수보다 탈퇴 유저수가 더 많았다
2. 매월 테마를 바꿔서 개최하던 게임 이벤트가 식상해져서 그만둔 유저가 많았다
3. 성별 혹은 연령 등 특정 유저 층에서 탈퇴한 유저가 많았다

마케팅 부서와 게임 개발 부서에 질의 응답을 한 결과 :

1. 광고는 지난달과 비교해서 거의 같은 수준이었으며 신규 유저수도 거의 같은 수준이었다.
2. 이벤트 내용도 지난달과 거의 바뀌지 않았다.

문제

- 지난달과 비교해서 유저수가 줄어들었다 (사실)
- 어떤 특정한 유저층에서 문제가 생겼을 것이다. (가설)

분석 스토리

- 퍼즐컬렉션의 유저수가 지난달보다 줄어들었다. (사실)
- 어딘가 숫자가 줄어든 세그먼트가 있을 것이다. (가설)
- 그 세그먼트에 적합한 대책을 세워서 유저수를 지난달과 같은 수준으로 회복한다. (해결책)

# 데이터 분석 workshop3

- 소셜게임의 이용자 탈퇴 이유 분석

- 탐색형 데이터 분석

- 사전에 원인을 짐작하기 어려운 상황에서 출발하여 그 원인을 데이터 분석을 통해 탐색해나가는 방법

데이터 수집 :

- DAU(Daily Active User) 데이터 - 하루에 한번 이상 게임을 이용한 유저
  - user\_info 데이터 - 유저의 속성정보

DAU

데이터 내용	데이터형	R 언어에서의 표기
로그인한 날	string (문자열)	log_date
앱 이름	string (문자열)	app_name
유저 ID	int (수치)	user_id

# 데이터 분석 workshop3

- 소셜게임의 이용자 탈퇴 이유 분석
  - 데이터 수집

DAU

데이터 내용	데이터형	R 언어에서의 표기
이용시작일	string (문자열)	install_date
앱 이름	string (문자열)	app_name
유저 ID	int (수치)	user_id
성별(여성, 남성)	string (문자열)	gender
연령대(10, 20, 30, 40, 50)	int (수치)	generation
단말기 종류(iOS, Android)	string (문자열)	device_type

# 데이터 분석 workshop3

- 소셜게임의 사용자 탈퇴 이유 분석
  - 데이터 가공

DAU 데이터에 user\_info 데이터 연관시키기

데이터 내용	데이터형	R 언어에서의 표기
유저 ID	int (수치)	user_id
앱 이름	string (문자열)	app_name
로그인한 날짜	string (문자열)	log_date
이용시작일	string (문자열)	install_date
성별(여성, 남성)	string (문자열)	gender
연령대(10, 20, 30, 40, 50)	int (수치)	generation
단말기 종류(iOS, Android)	string (문자열)	device_type

# 데이터 분석 workshop3

- 소셜게임의 이용자 탈퇴 이유 분석
  - 데이터 분석
    - 어떤 속성을 가진 유저층이 지난달에 비해 감소하였는지 그 세그먼트(계층)을 찾아내기 위해 크로스집계를 실시
- 크로스 집계
  - 2개 변수의 인과관계를 교차해서 집계하는 분석 기법

# 데이터 분석 workshop3

- 소셜게임의 이용자 탈퇴 이유 분석

- 어떤 속성을 가진 유저층이 지난달에 비해 감소하였는지 그 세그먼트(계층)를 찾아내기 위해 크로스 집계를 실행하고, 원인으로 생각되는 속성이 발견되면 시각화한다.
- 어떤 상태의 데이터와 유저의 속성정보(원인 데이터)를 연관시켜서 특정한 속성(원인)이 어떠한 상태(결과)를 야기하는 것이 아닌진 검토하는 방식으로 분석

세그먼트별로 크로스집계하기(세그먼트 분석):

1. 성별 세그먼트 분석
2. 연령대 세그먼트 분석
3. 성별 X 연령대 세그먼트 분석
4. 단말기별 세그먼트 분석
5. 세그먼트 분석 결과 시각화



# 데이터 분석 workshop3

- 소셜게임의 이용자 탈퇴 이유 분석

- 데이터 분석

- 세그먼트별로 크로스집계하기(세그먼트 분석) : 성별

로그인한 달	성별	
	여성	남성
2013년 8월		
2013년 9월		

- 세그먼트 분석 : 연령대

로그인한 달	연령대				
	10대	20대	30대	40대	50대
2013년 8월					
2013년 9월					

# 데이터 분석 workshop3

- 소셜게임의 이용자 탈퇴 이유 분석
  - 데이터 분석
    - 세그먼트 분석 : 연령대

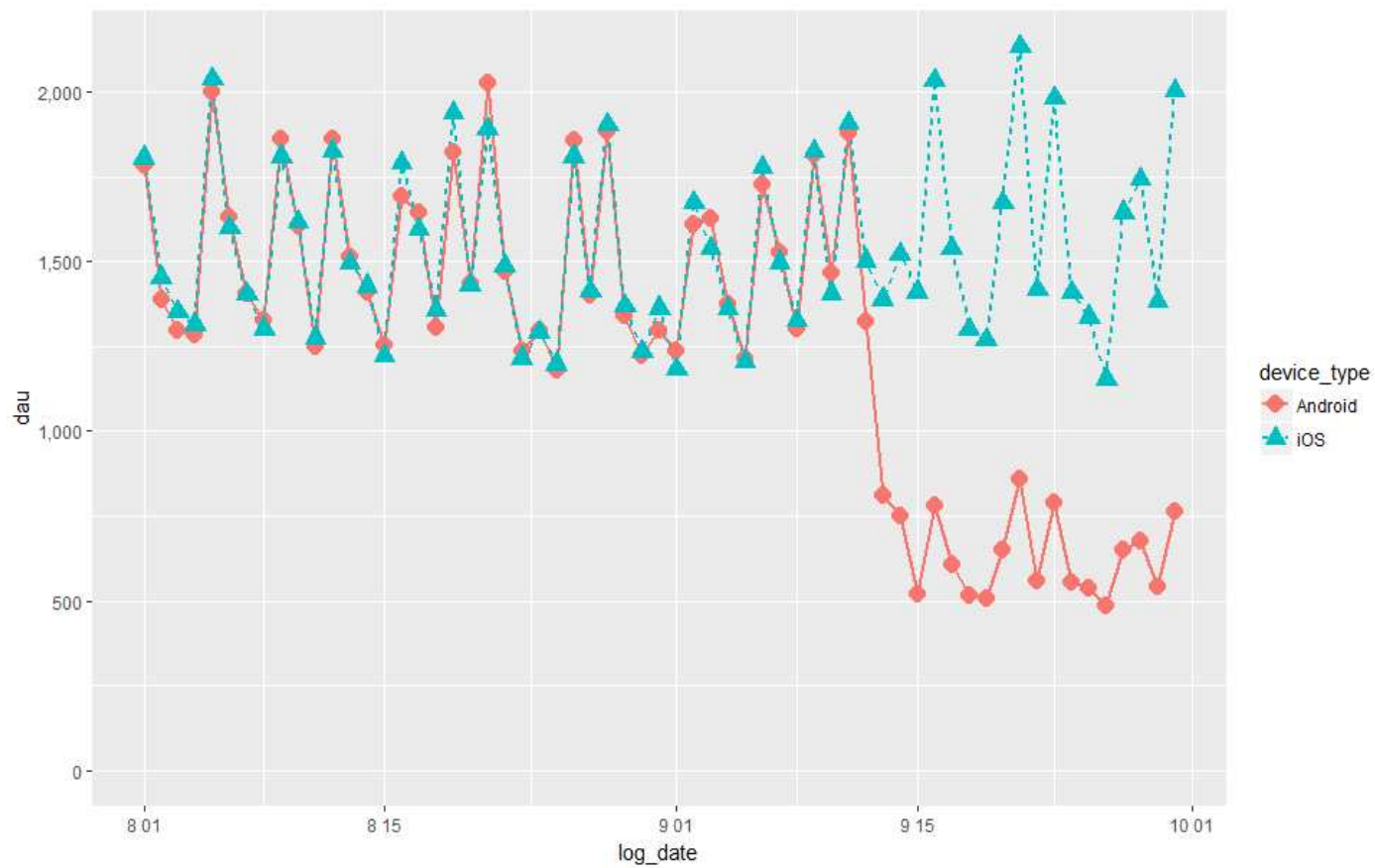
로그인한 달	여성					남성				
	10대	20대	30대	40대	50대	10대	20대	30대	40대	50대
2013년 8월										
2013년 9월										

- 세그먼트 분석 : 단말기별

로그인한 달	단말기	
	Android	iOS
2013년 8월		
2013년 9월		

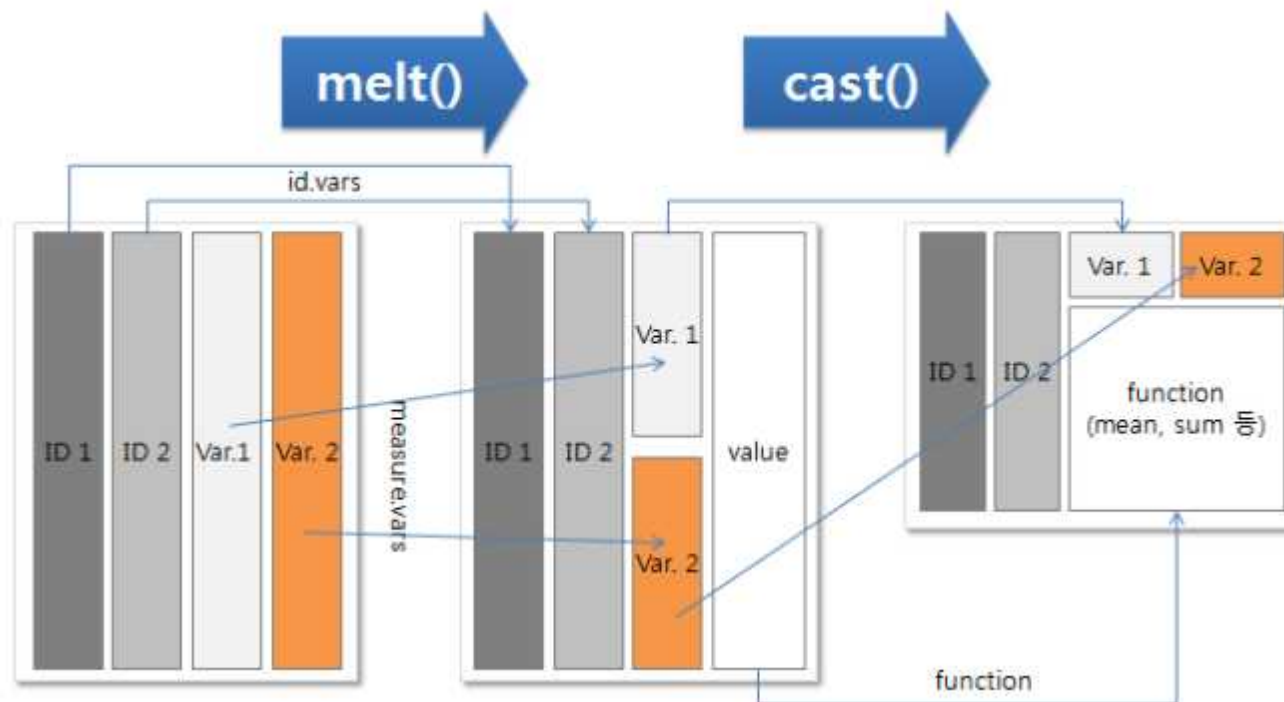
# 데이터 분석 workshop3

- 소셜게임의 이용자 탈퇴 이유 분석
  - 데이터 분석 결과 시각화하기



# 데이터 분석 workshop

- R 분석 함수



# 데이터 분석 workshop

- R 분석 함수

**tables in R**

어떤 값이나 리스트를 원소로 갖는 이차 배열

- ✓ table()의 첫 번째 인수 : 팩터나 팩터 리스트
- ✓ table()의 데이터 인수 : 데이터 프레임

“ 대부분의 수학적 용도가 아닌 행렬, 배열 연산 사용 가능 ”

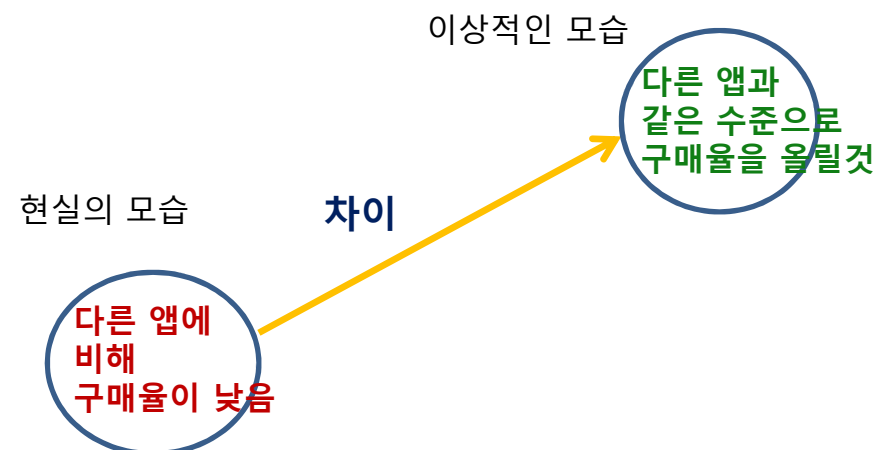
aggregate( )      cut( )

# 데이터 분석 workshop 4

- 배너광고 반응 비교 검증
  - 현실의 모습과 이상적인 모습

퍼즐컬렉션에서는 매월 게임에서 이용하는 아이템을 세일하고 있습니다. 이 세일은 매상 비율이 높은 중요한 이벤트인데, 어느 날 회사 경영진으로부터 '매상 자체는 높지만 구매율은 그리 높지 않은 것 아닌가?' 라는 지적을 받게 되었습니다. 실제로 다른 앱의 세일과 비교한 결과, 퍼즐컬렉션의 구매율이 상대적으로 낮음을 알게 되었습니다. 구매율을 개선하기 위해 구매율이 낮은 원인을 데이터 분석으로 찾아내도록 분석 담당자에게 의뢰가 왔습니다.

앱 이름	구매율
퍼즐 컬렉션	6%
앱 A	12%
앱 B	12%



# 데이터 분석 workshop 4

- 배너광고 반응 비교 검증
  - 문제 발견

가설 :

- 아이템 세일 내용에 문제 있음
  - => 지금 별로 필요하지 않은 아이템을 내놓고 있다
  - => 세일폭이 별로 크지 않아서 유저에게 크게 어필하지 못하고 있다.
- 배너광고의 표시내용에 문제가 있음

기획부로부터 아이템에 대한 응답 결과 :

- 아이템 세일에 쓸 만한 것을 내놓았다고 생각하지만 얼마나 사용하고 있는지 크게 자신은 없음
- 세일폭은 50%로 다른 앱과 같고 유저 입장에서 충분히 싸게 느낄 것임

마케팅부로부터 배너광고에 대한 응답 결과 :

- 아이템 세일의 배너광고는 해당 앱의 디자이너가 만들어 앱에 따라 품질이 제각각임
- 퍼즐컬렉션의 배너광고는 항상 클릭률이 낮음

세일 배너광고의  
클릭을 비교

앱 이름	구매율
퍼즐 컬렉션	6%
앱 A	12%
앱 B	12%

# 데이터 분석 workshop 4

- **배너광고 반응 비교 검증**
  - 검증 방법의 검토

## 문제

- 퍼즐컬렉션의 배너 광고 클릭률이 다른 앱보다 낮다(사실)
  - 퍼즐컬렉션의 배너 광고 품질에 문제가 있다. (가설)
- 
- **클릭률이 높은 배너광고를 조사해서 퍼즐컬렉션의 배너광고를 개선한다. (해결책)**
  - 어떤 배너 광고가 자주 클릭되는지 알 필요가 있습니다.
  - 퍼즐컬렉션에서는 지금껏 매월 실시해온 아이템 세일의 배너광고를 한번도 바꾸지 않아 분석을 하려 해도 데이터가 없습니다.
  - 2개의 배너광고를 새로 작성해서 어느 쪽이 더 나은지 데이터를 수집하기로 합니다



# 데이터 분석 workshop 4

- **A/B 테스트**

- 여러 선택지 중에서 어느 것이 가장 좋은 결과를 가져다줄지 알아보기 위한 검증방법
- 초기 도입 시 개발비용이 많이 들지만 비교적 낮은 비용으로 실시할 수 있으며, 수집한 데이터를 통계적으로 취급하기 쉬워서 Web 업계에서 많이 이용
- 일부 광고업이나 제조업 쪽에서도 실시
- 예) A와 B라는 2개의 제품을 만들고 동시에 판매해서 데이터를 수집하고, 타킷 고객층에 크게 어필한 것이 어느 쪽인지 검증
- 같은 시기에 같은 타킷 고객층에 대한 분명한 인과관계를 알 수 있기 때문에 많이 사용
- A와 B의 구분은 임의적이어야 함
  - 성별, 연령대, 게임이용시작일등의 조건이 모든 그룹에 균등하게 배분될 것

# 데이터 분석 workshop 4

- **KPI(Key Performance indicator)**
  - 중요업적 평가지표
  - PV(Page View) : 액세스 횟수
  - DAU(Daily Active Users) : 1일 이용 유저수
  - ARPU(Average Revenue Per User) : 유저 1명의 월간 평균 지출 금액
  - 지속율
  - 과금율

# 데이터 분석 workshop 4

- 가설검정

- 어떤 두개의 그룹에 차이가 있는지 없는지
- 가설검정은 사람수가 많아질수록 '통계적'으로 유의한 차이가 있다'는 결론을 내리기 쉬워짐
- 통계적으로 유의한 차이가 있어도 비즈니스에서 의미가 있는 차이라고 단정할 수 없다.
- '커트라인'으로 실시
- '가설 검정에서 차이가 나타났기 때문에 이걸로 끝'이 아니라 '적어도 가설검정에서는 의미가 있는 차이가 나타났으므로, 이것이 비즈니스상에서 의미가 있는 차이인지 검토해야 한다.

# 데이터 분석 workshop 4

- **배너광고 반응 비교 검증**
  - 데이터 수집
    - as\_test\_imp (배너광고의 표시횟수정보)

데이터 내용	데이터형	R 언어에서의 표기
표시된 날짜	string (문자열)	log_date
테스트 이름	string (문자열)	test_name
테스트 케이스 : A 또는 B	string (문자열)	test_case
유저 ID	int (수치)	user_id
트랜잭션 ID	int (수치)	transaction_id

# 데이터 분석 workshop 4

- **배너광고 반응 비교 검증**
  - 데이터 수집
    - as\_test\_goal (배너광고의 클릭횟수정보)

데이터 내용	데이터형	R 언어에서의 표기
클릭한 날짜	string (문자열)	log_date.g
테스트 이름	string (문자열)	test_name.g
테스트 케이스 : A 또는 B	string (문자열)	test_case.g
유저 ID	int (수치)	user_id.g
트랜잭션 ID	int (수치)	transaction_id.g

# 데이터 분석 workshop4

- **배너광고 반응 비교 검증**
  - 데이터 가공

- 배너광고의 표시횟수정보에 클릭횟수정보 결합하기
- 클릭했는지 하지 않았는지 플래그 작성하기

데이터 내용	데이터형	R 언어에서의 표기
트랜잭션 ID	int (수치)	
표시할 날짜	string (문자열)	
테스트 케이스 : A 또는 B	string (문자열)	
유저 ID	int (수치)	
클릭여부	int (수치)	ls_goal

# 데이터 분석 workshop 4

- 배너광고 반응 비교 검증

- 데이터 분석

- 클릭률 집계 plyr 패키지의 ddpby 함수 사용
  - Test\_case 항목별로 집계 (클릭한 사람의 합계/ 배너광고가 표시된 유저수)
  - $\chi^2$  검정 실행하기 `chisq.test ( )`
  - 2개의 배너광고와 '클릭했음/하지 않았음'에 대한 관계를 가설 검정
  - P값이 0에 가까우면 가까울수록 차이가 있다는 것을 나타내고, 일반적으로 0.05보다 작은 값이면 '통계적으로 차이가 있다'고 할 수 있다.

# 데이터 분석 workshop 4

## • 배너광고 반응 비교 검증

1. 배너광고의 표시횟수 정보데이터와 클릭횟수정보 데이터를 결합 (merge함수)
2. 클릭했는지 하지 않았는지 나타내는 플래그 작성 (ifelse 함수)  
배너광고 클릭회수 정보의 userid값이 NA인 경우 0, 이외의 경우 1로 기입
3. 클릭률 집계(plyr 패키지의 ddply 함수)  
test case 항목별로 집계  
클릭한 사람의 합계 / 배너광고가 표시된 유저수
4.  $\chi^2$  검정 실행하기 (`chisq.test()`)  
2개의 배너광고와 '클릭했음/하지 않았음'에 대한 관계를 가설 검정  
P값이 0에 가까우면 가까울수록 차이가 있다는 것을 나타내고,  
일반적으로 0.05보다 작은 값이면 '통계적으로 차이가 있다'고 할 수 있다
5. 날짜별, 테스트 케이스별로 클릭률 산출
6. 테스트 케이스별로 클릭률 산출
7. 테스트 케이스별 클릭률의 시계열 추이 그래프로 시각화



# 데이터 분석 workshop 4

- 배너광고 반응 비교 검증

- 시각화

- 가설검정으로 통계적으로 유의한 차이가 있는지 확인 후 비즈니스 상에서 의미 있는 차이인지 검토
  - 시계열 추이를 확인
  - 시계열 추이에서 전체적으로 배너광고 B의 클릭률이 높으면 문제가 없지만, 어떤 특정 시점에만 차이가 크게 나타난다고 하면 전혀 다른 요인일 가능성이 있음

