

왜 할까, 그리고 왜 중요할까
데이터 프로젝트 접근법

대흥기획 디지털사업부문
안중호 cem

bakky14@gmail.com

CONTENTS

1. 왜 해야 하는가

2. 어떻게 해야 하는가

3. 프로젝트 유형별 접근

4. 되게 만드는 ppt

5. 은근히 알려지지 않은 몇가지

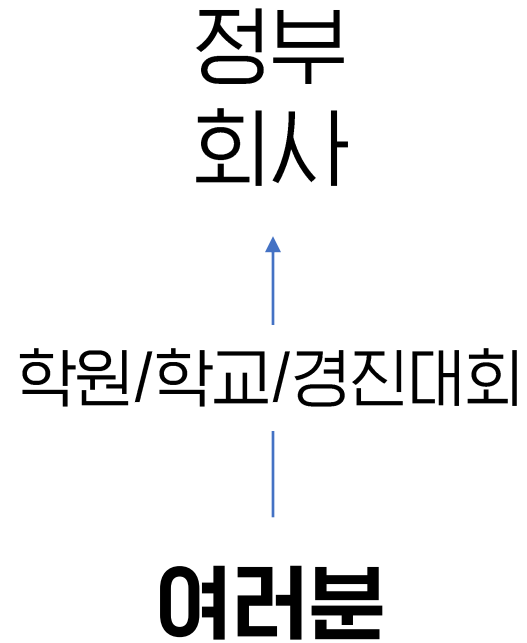


1. 왜 해야 하는가

- 분석의 목적부터 정의하기

- 1. 왜 해야 하는가
 - 분석의 목적부터 정의하기

왜 하나요?



- 1. 왜 해야 하는가
 - 분석의 목적부터 정의하기

돈이 되니까!

정부
회사
↑
생각해보면
돈이 안되면 아무도 여기에 없었을 것이다
즉, 데이터 분석은 보고서가 아니라
돈이 되어야 한다
↓
여러분

1. 왜 해야 하는가

- 분석의 목적부터 정의하기

근데, 빅데이터는 너무 흑하고 들어왔다
“목적”을 생각하지 않고 무작정 시작했던 빅데이터 1세대
여기에서 살짝 비극이 시작됐다



1. 왜 해야 하는가

- 분석의 목적부터 정의하기

목적이 없으면 망

알파고의 목적

바둑이라는 규칙에서 승리

19 x 19
검은돌 -> 흰돌
에워싼 공간이 많으면 승리

-> 목적 달성^_^

≠

사람들의 목적

우월어 빅데이터

일단 수집
일단 저장
일단 조직

-> ??????=_=;;;

1. 왜 해야 하는가

- 분석의 목적부터 정의하기

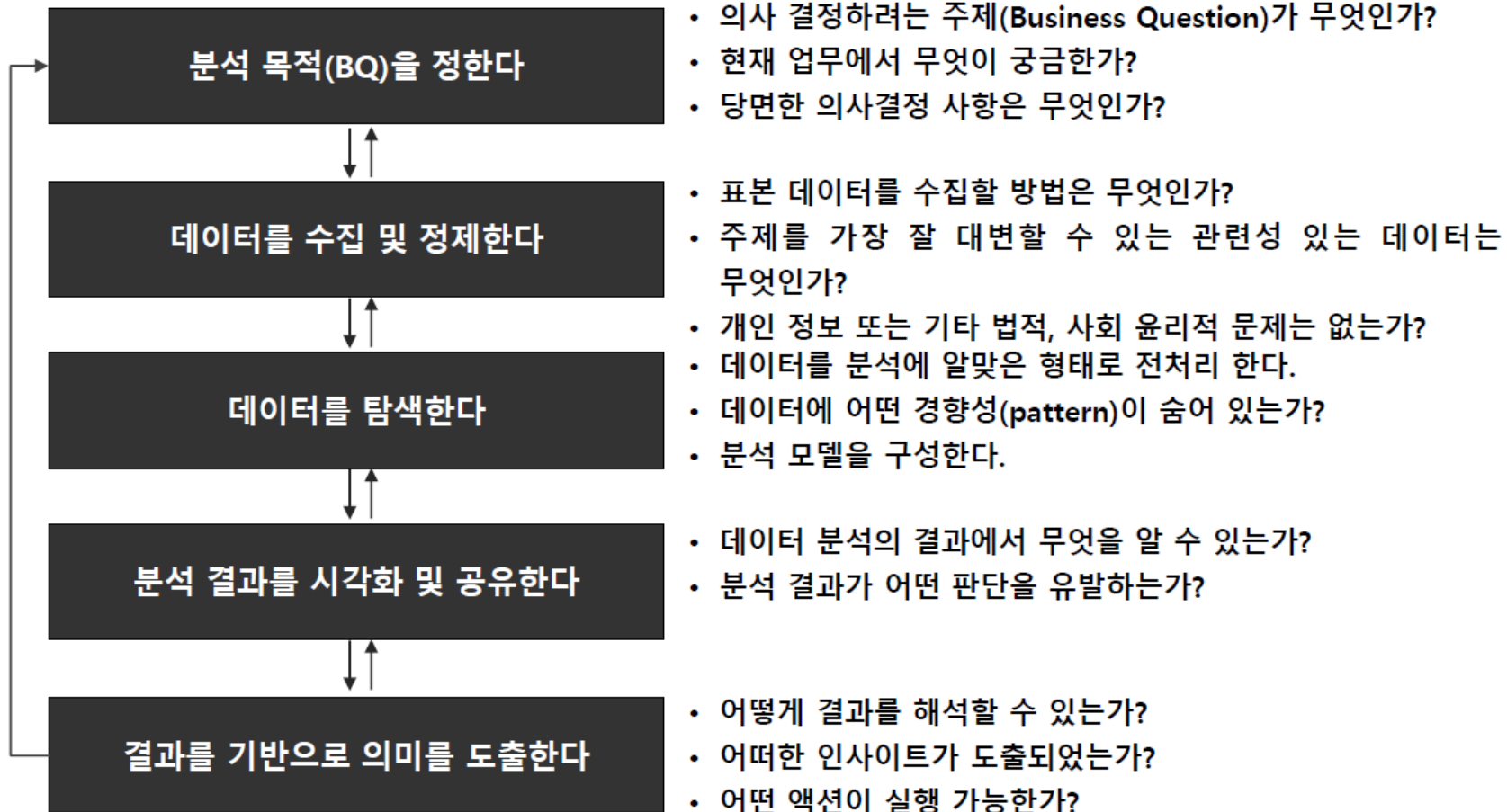
그때 목적 : 데이터를 모아서 쌓는 것
쌓기만 해도 잘한다고 칭찬받던 시절이 있었다
지금 목적 : 어찌지? 이제 잘 쌓아야겠다



1. 왜 해야 하는가

- 분석의 목적부터 정의하기

“뭘 알고 싶은지를 결정하고, 정확한 데이터 찾기”에서 시작
“방대한 빅데이터에서 다양한 인사이트를” 이런거 금지



1. 왜 해야 하는가

-분석의 목적부터 정의하기

좋은 비즈니스 질문(BQ)이 있어야 한다

알고 싶은게 뭔지 모른다? 그럼 “질문”부터 분석가가 만들어야지
좋은 질문은 현업과 분석가가 티키타카 함께 만들어간다 예상적

매출

- 매출에 영향을 미친 고객 지표는 무엇인가?
- 추가 이익을 얻기 위해 매장과 상품을 어떻게 개선해야 하는가?

Range

- 카테고리 매대 진열을 어떻게 변경할 것인가?
- 어떤 상품을 pb로 개발할 것인가

프로모션

- 지난 프로모션 성과는 어떠하고, 이번 프로모션은 어떤 상품으로 구성해야 하나?
- 어떤 상품을 EDLP(every day low price)로 진행해야 하나?

매장

- 온라인 전환 속도는 어떠하며, 가장 효과적인 활동은 무엇인가?
- 매장 리뉴얼을 통해 신규 유입된 고객은 몇 명이고, 기존 고객의 장바구니는 커졌는가?

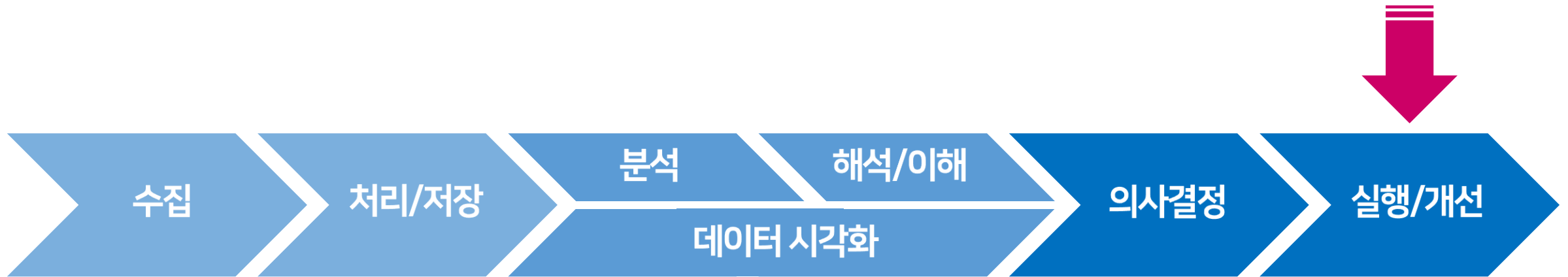
고객

- 충성 고객이 계속 구매하고 있는가?
- 온라인 채널로 유입된 고객의 구매 전환율은 어떠한가?

1. 왜 해야 하는가

- 분석의 목적부터 정의하기

비즈니스 질문에 답해오다 보니 “의사결정/실행” 단까지
...이제 사장님의 질문을 받는 단계 -_-



Engineering

**Marketing
(CEO)**

1. 왜 해야 하는가
_분석의 목적부터 정의하기

사장님의 질문은 심플해요

그거 돈이 돼냐?

내 월급은 회사가 준다

프로젝트도 돈이 든다

“그냥 한번 해보는” 건 의미가 없다

반드시! 분석하는 투자 대비 이익(ROI)


듣다보면 돈이 나오겠지 하고 참아주시던 좋은 시절이 아니다

1. 왜 해야 하는가
_분석의 목적부터 정의하기

그렇다면 분석의 아웃풋은

뭘 하면 뭐가(얼마가) 나오는지

사장님한테 납득시켜
실행하게 하는 것이 목적
목적에 집중하지 않으면 수요 없는 공급을 만들 수 있다
그리고 수요 없는 공급도 받아주던 시절은 끝났다

A top-down view of a business meeting around a table. The table is covered with various documents, including colorful bar and pie charts, and several white papers. There are two white coffee cups with dark coffee on the table. A person's hand is visible pointing at a document. The background shows the legs and hands of several people in business attire sitting around the table.

2. 어떻게 해야 하는가

- 인사이트는 액션이 되어야 한다

- 2. 어떻게 해야 하는가
 - _인사이트는 액션이 되어야 한다

분석 결과에 대한 흔한 피드백

**재밌네요 or 좋네요
그래서 뭘 하자는 건지??**

그리고 아무것도 일어나지 않는다

- Ⅰ 2. 어떻게 해야 하는가
 - _인사이트는 액션이 되어야 한다

재밋네요
Insight는 있어도

그래서 뭘 하자는 건지?
Actionable한 Plan이 없다면

“분석”을 아무리 많이 해도
현업의 “액션”으로 연결되지 않는다

2. 어떻게 해야 하는가

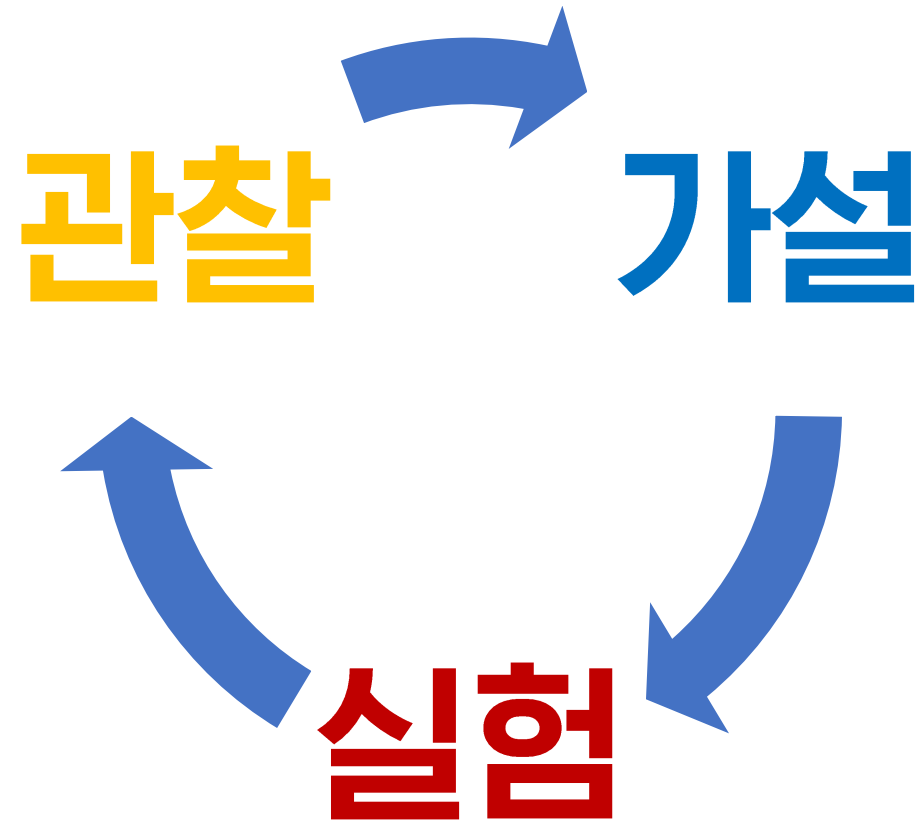
-인사이트는 액션이 되어야 한다

애초에 왜 Data **Science**일까? 를 생각해보면



- 2. 어떻게 해야 하는가
 - 인사이트는 액션이 되어야 한다

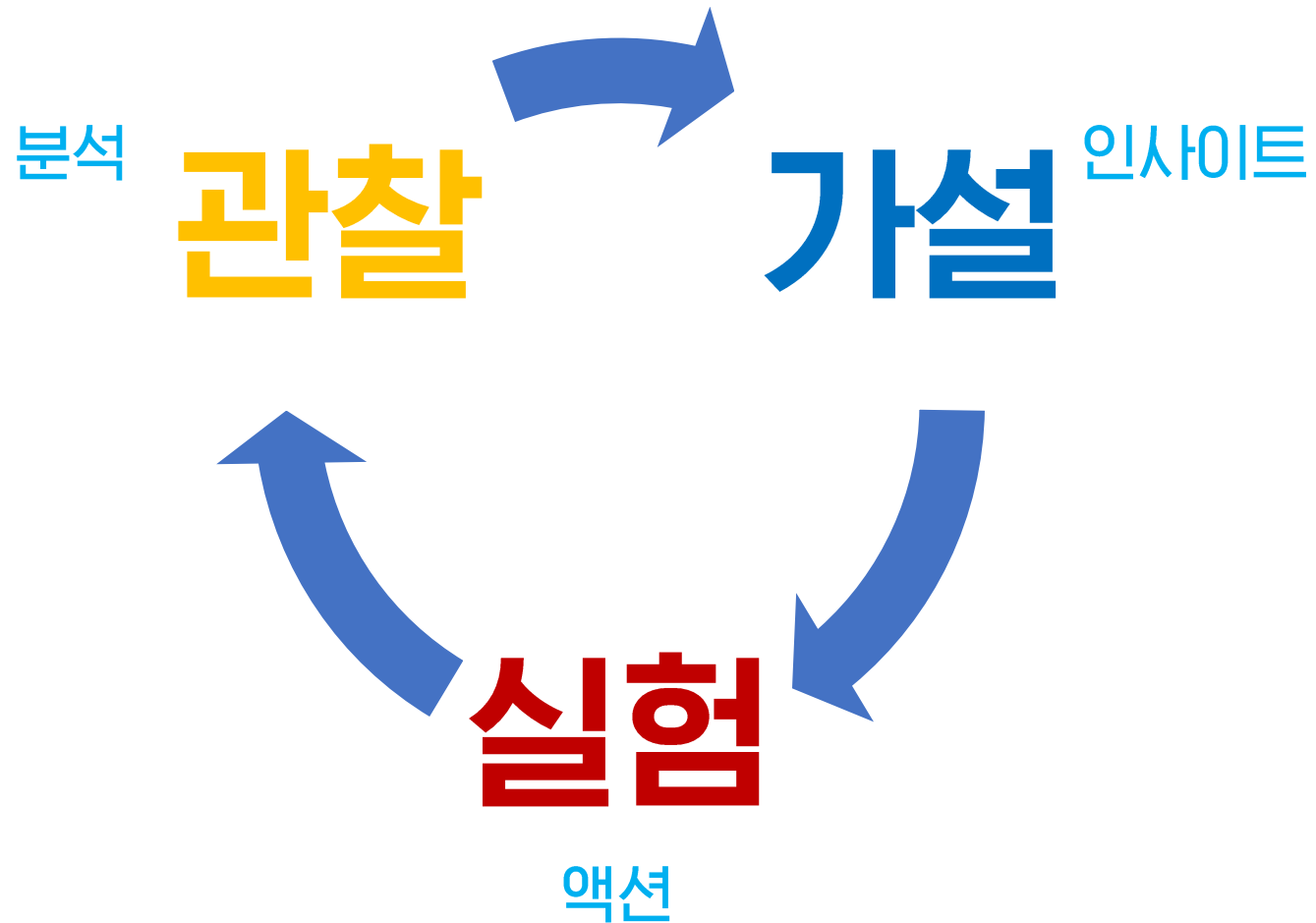
과학의 방법론대로 일하기 때문



2. 어떻게 해야 하는가

-인사이드는 액션이 되어야 한다

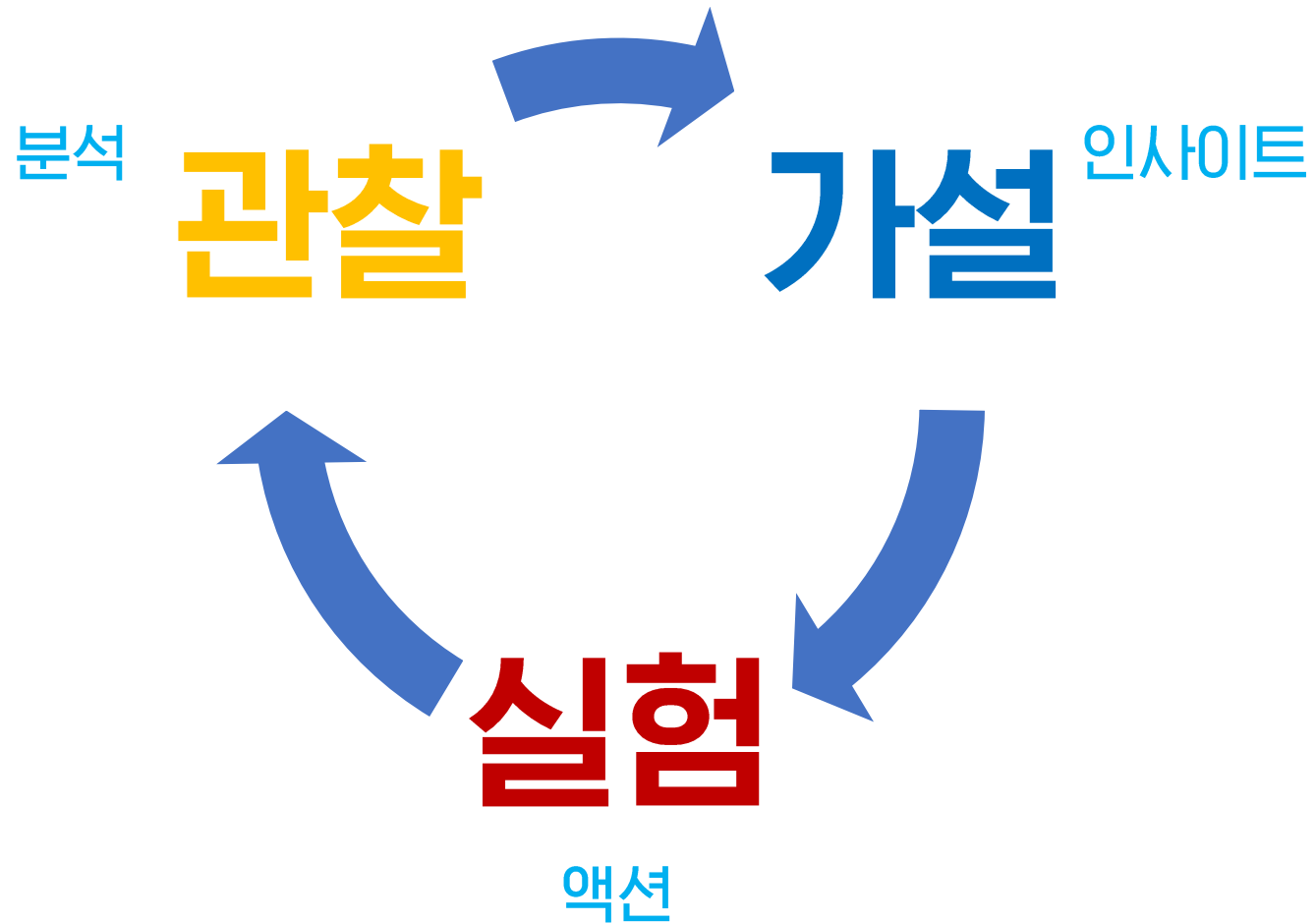
그중 “**실험**”부분이 바로 액션



Ⅱ 2. 어떻게 해야 하는가

-인사이트는 액션이 되어야 한다

작은 가설의 실험을 **빠르게 반복**해 개선하는 프로세스



2. 어떻게 해야 하는가

-인사이트는 액션이 되어야 한다

그런데 “분석”은 많이 하지만 액션”은 잘 하지 않는다



2. 어떻게 해야 하는가

-인사이드는 액션이 되어야 한다

“액션”에는 돈이 든다 돈 돈
실험? 여기가 회사지 연구소냐?
실험으로 나아가는 부담을 줄여줘야 한다
즉, 결과가 어느 정도 예상가능해야



넘을 수 없는 4차원의 벽

Ⅱ 2. 어떻게 해야 하는가

-인사이트는 액션이 되어야 한다

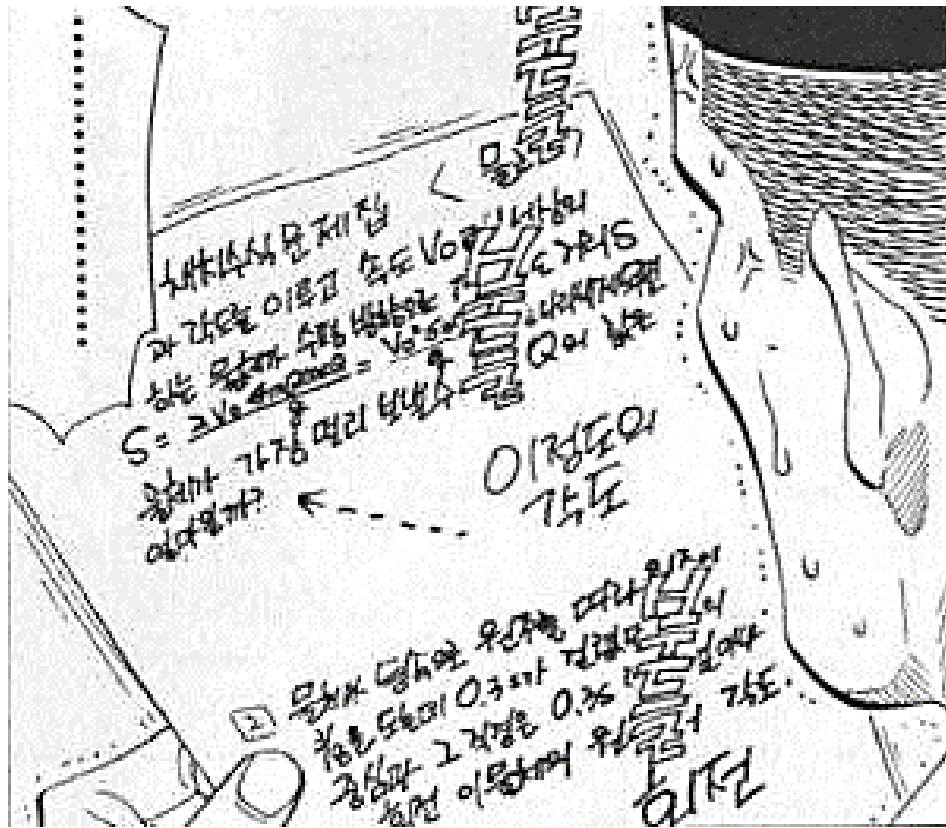
분석한 결과를 보면
“당연한건데?” “꼭 분석해야 알아?”
경향성만 파악했다면 그럴 수 있다
코로나로 색조화장이 감소했어요
이 틈트는 10대 여성이 주로 구매하네요

여기에서 한 걸음 더 나가는게 중요하다
분석의 힘은 정량화 : 양과 크기를 아는 것
“양을 정확히 안다” -> “예상과 액션이 가능하다”

2. 어떻게 해야 하는가

-인사이드는 액션이 되어야 한다

공을 던지면 날아간다는 원손은 거들뿐 누구나 알지만
힘, 각도, 날아간 거리를 재면 인공위성도 띄운다





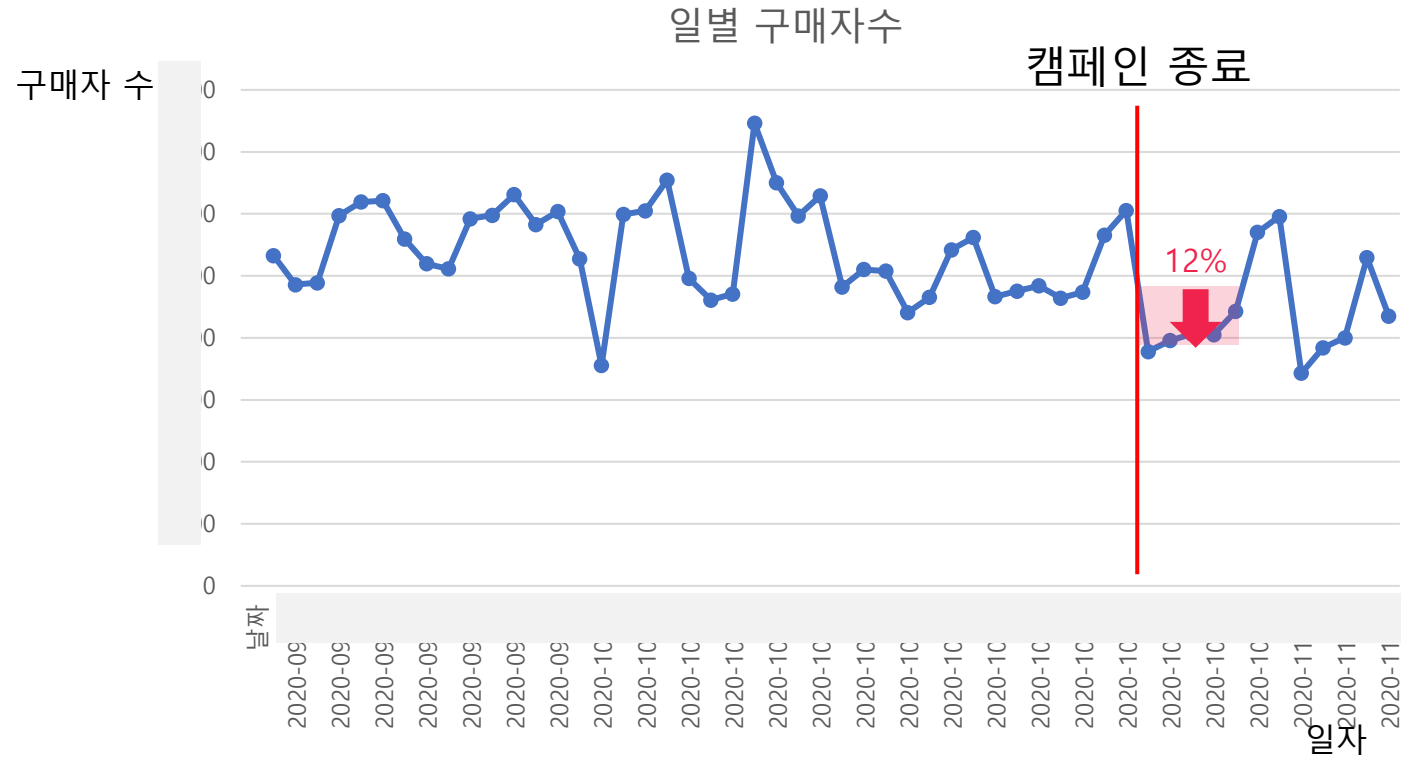
**내 광고비의 절반이
낭비되고 있다는 것을 알지만
문제는 그 절반이
어느 쪽인지 모르겠다는 점이다.**

- 존 워너메이커 -

2. 어떻게 해야 하는가

-인사이트는 액션이 되어야 한다

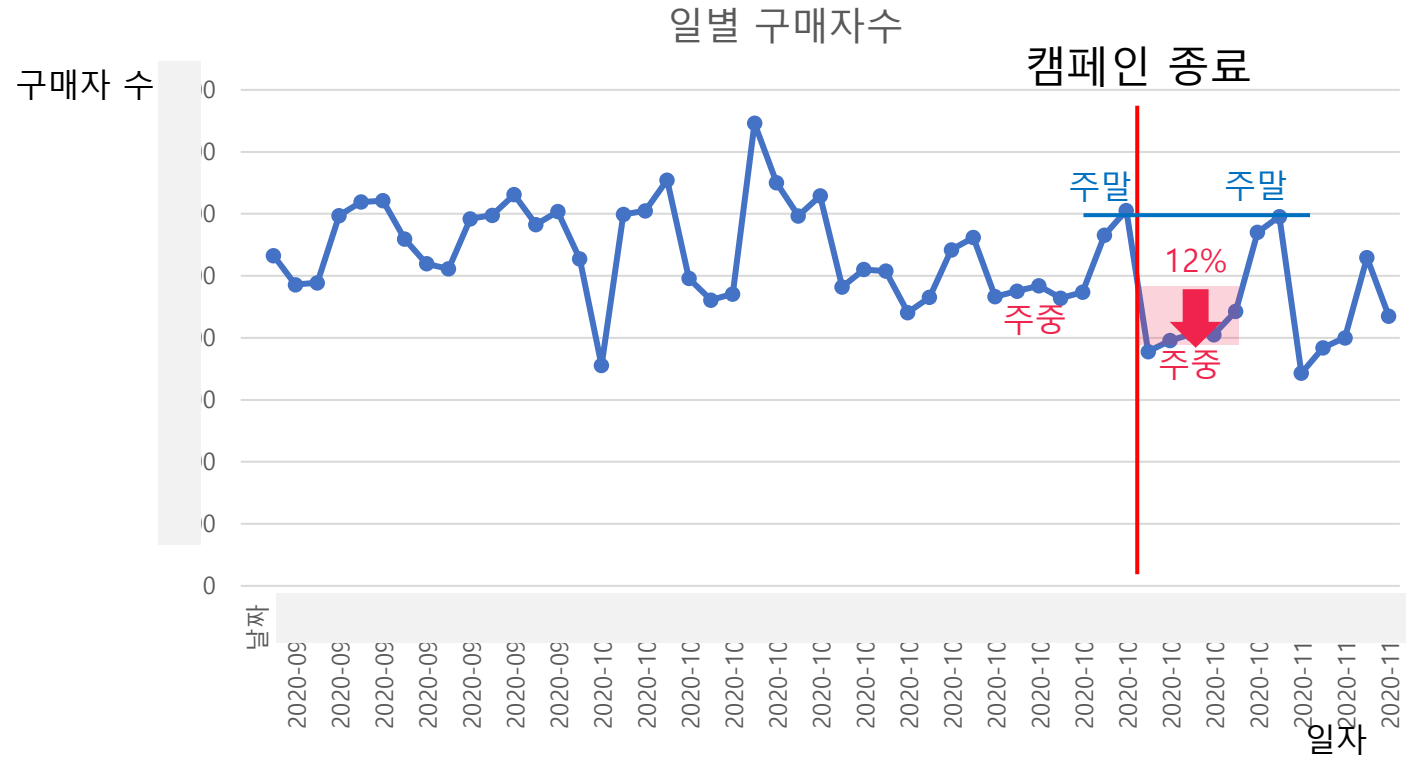
관찰 : 광고를 끊었더니 고객은 12%감소



2. 어떻게 해야 하는가

-인사이트는 액션이 되어야 한다

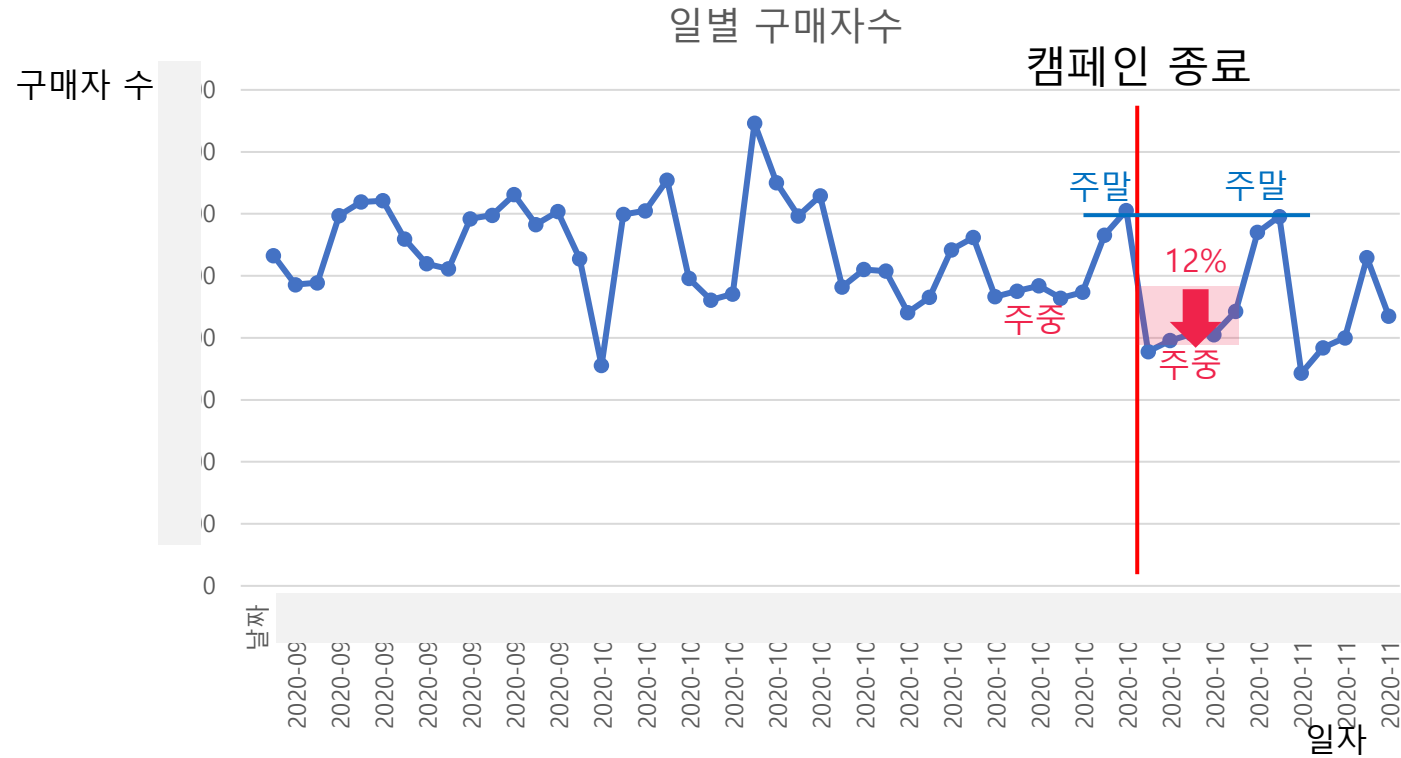
좀더 관찰 : 광고를 끊었더니 **주중 고객만 12%감소**,
주말 고객은 그대로



2. 어떻게 해야 하는가

-인사이트는 액션이 되어야 한다

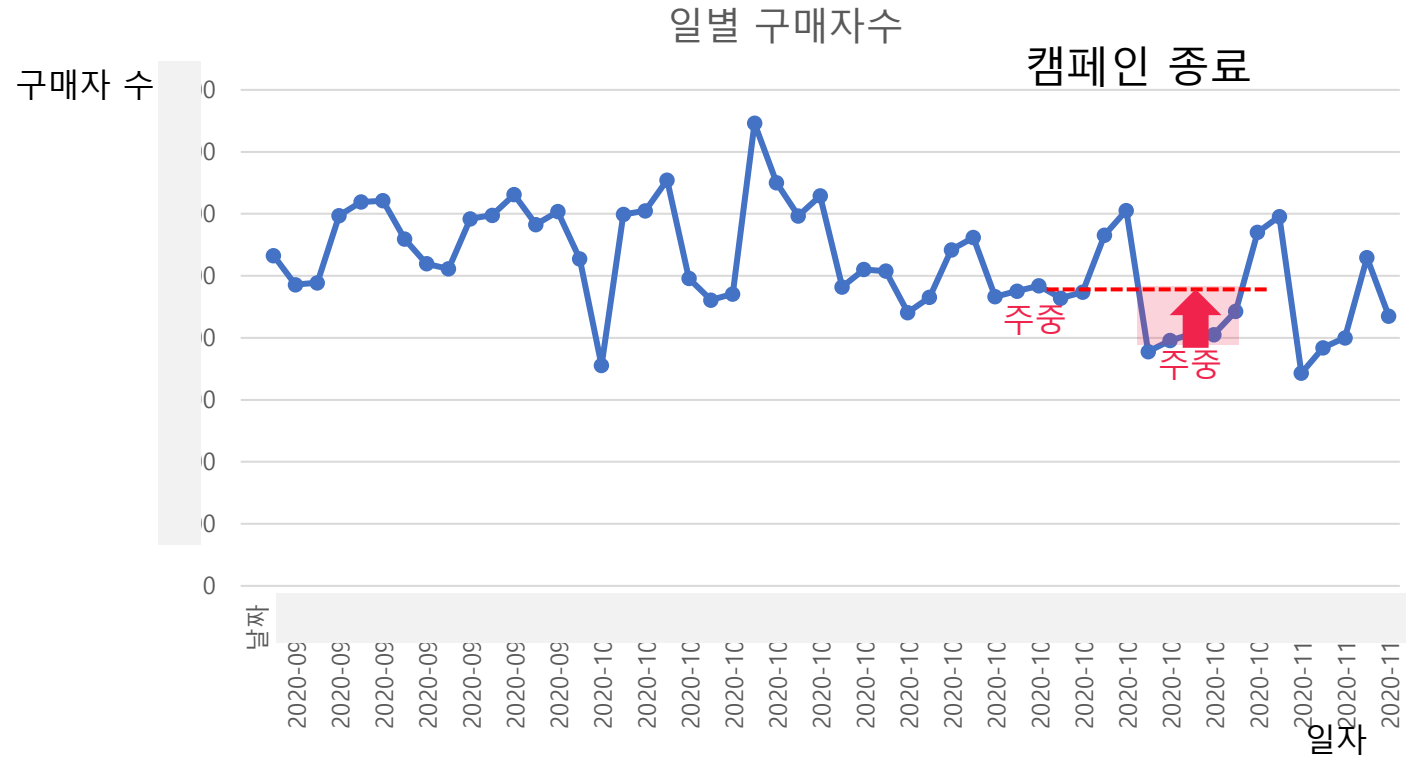
가설 : 주말 고객은 광고가 없어도 왔을 사람들!
-> 주말에 광고하는건 낭비



2. 어떻게 해야 하는가

-인사이트는 액션이 되어야 한다

실험 : 주말에 쓸 광고비를 주중으로 돌리자!



Ⅱ 2. 어떻게 해야 하는가

-인사이드는 액션이 되어야 한다

정량적으로 아는 데이터 :

일별 구매자수, 구매액, 주중/주말 광고비
주중에 감소한 고객의 숫자와 매출액

“결과가 어느 정도 예측 가능하다”

가설 -> 액션 :

주말의 (불필요한) 광고비를 주중으로 돌리면

12% 감소한 주중고객이 복귀할 것이다

실험 :

주중/주말 광고비 최적 배분을 알고싶다!
얼마 넣으면 얼마 돌아오는지 찾아내자!

“될거같다, 해보자!”

2. 어떻게 해야 하는가 _인사이드는 액션이 되어야 한다

“수치화”의 가치 : 현재 상태 진단, 향후 개선의 우선순위 판단

환자 상태를 알아야
뭐라도 해보지

돈과 시간은 언제나 부족하다
선택과 집중이다

서울교통공사, 지하철 혼잡도 산출 서비스 '철도 10대 기술상' 수상

황순호 | 승인 2021.11.16 10:06 | 댓글 0

혼잡도 4단계 색으로 안내, 시간대별 표준편차 약 30% 감소, 승객 분산 효과 검증
'또타지하철'과 'TMAP대중교통' 앱 실시간 표출 시범 운영...향후 전 노선 확대 추진



2021년 철도 10대 기술상 시상식에서 서울교통공사와 SK텔레콤 관계자들이 기념촬영을 하고 있다. 사진=서울

조사원이 직접 탑승인원을 확인하는 방식, 교통카드를 이용한 경로추적 방식 등 다양한 방법으로 측정·서비스한데 이어 한층 과학적인 빅데이터 융합 방식으로 그동안의 한계를 넘어 혼잡도 측정의 정확성을 높였다고 설명했다.

▷ 목측(目測)법 : 교통량 조사 시 주로 사용되는 가장 기초적인 방법으로 조사원이 직접 지하철에 탑승해 인원을 확인·계산하는 방식이다. 대략적인 혼잡도는 파악할 수 있지만 신뢰도가 높지 않고 조사 시기 값만 반영된다 는 문제가 있다.

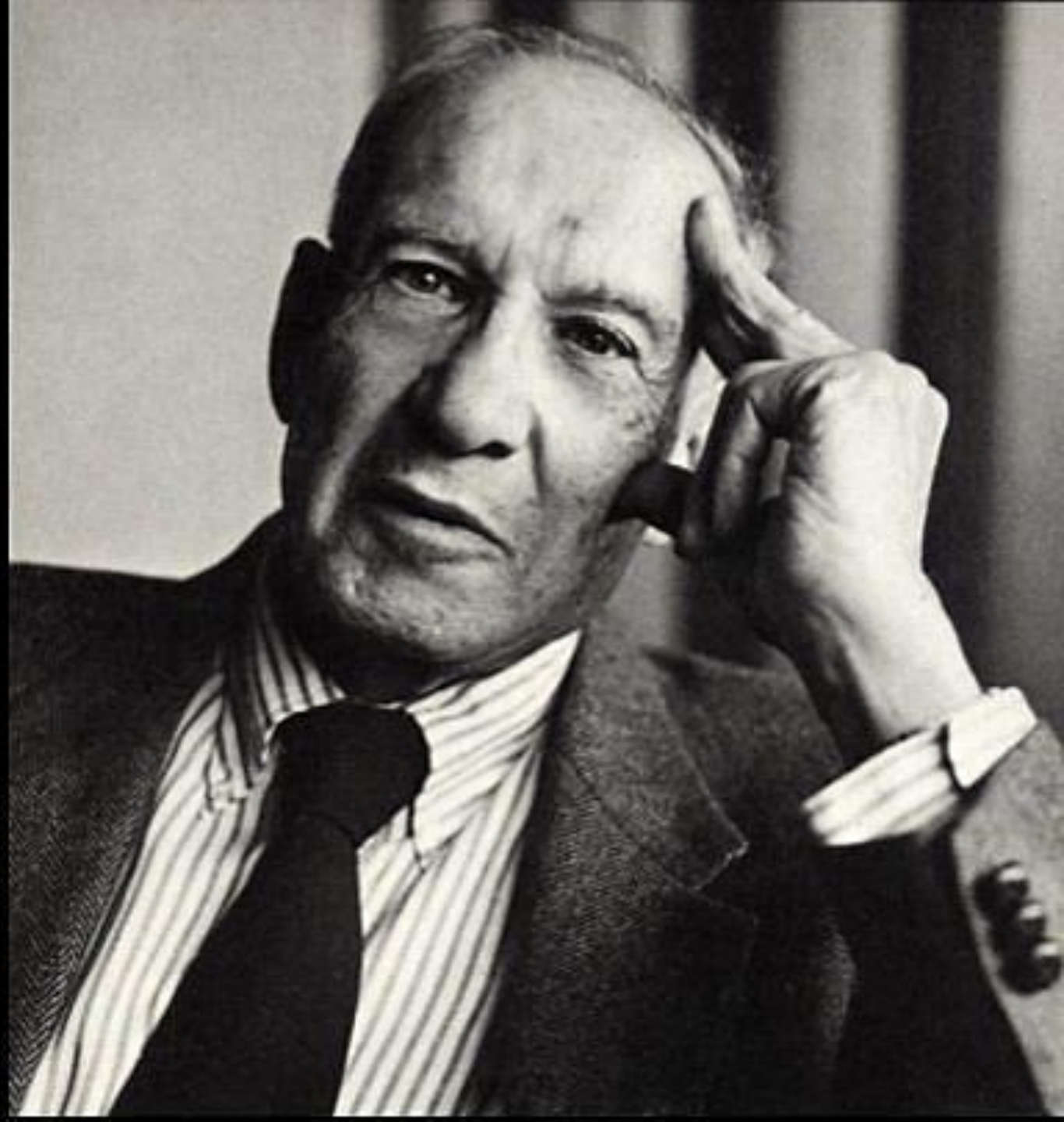
▷ 교통카드 자료 기반 경로추적 : 승객이 지하철을 타고 내릴 때 단말기에 입력되는 교통카드 정보를 기반으로 이동경로를 추정, 지하철 탑승인원을 산출하는 방식이다. 공사는 이 방식을 활용한 '지하철 혼잡도 예보제' SNS 알림 서비스를 '20년 중순부터 실시하고 있다.

▷ 탑승인원 무게 센서 방식 : 2018년부터 운행 중인 2호선 신조 차량의 경우, 탑승인원의 무게를 자동 센서로 감지, 실시간으로 측정해 혼잡도를 안내하는 장치가 승강장 내에 설치돼 있다.

“

측정할 수 없으면
관리할 수 없고
관리할 수 없으면
개선시킬 수도 없다

- 피터 드러커 -





3. 프로젝트 유형별 접근

- 과거를 분석, 현재를 개선, 미래를 예측

3. 프로젝트 유형별 접근

_과거를 분석, 현재를 개선, 미래를 예측

분석의 관점 : 회사의 “어느 시점”에 관심있는가?



프로젝트가 어떤 성격에 해당하는지 파악하고
목적에 부합하는 접근방식 채택

3. 프로젝트 유형별 접근

-과거를 분석, 현재를 개선, 미래를 예측

“과거를 분석한다는 것”



회사의 과거를 제대로 확인 하는 것

그리고

과거를 분석할 수 있도록, 데이터가 목적에 맞게 쌓이게 하는 것

3. 프로젝트 유형별 접근

-과거를 분석, 현재를 개선, 미래를 예측

“과거를 분석한다는 것”

지난 매출에 대한 분석
현재 상품에 대한 분석
현재 고객에 대한 분석

BQ 예시 (커머스) :

최근 6개월 동안 판매된 상품 중 제품 카테고리별 구매 주기가 가장 긴 상품은 무엇인가?

대표 브랜드/제품의 구매 주기는?

최근 6개월간 카테고리별/브랜드별 구매 집중 시기는 언제인가?

특정 제품을 구매하는 고객의 특성은 무엇인가?

3. 프로젝트 유형별 접근

_과거를 분석, 현재를 개선, 미래를 예측

“과거를 분석한다는 것”

경험에만 의한 추측과 가설들을
오랜 동안 덮어뒀던 경우가 많다

우리가 제대로 해 왔는지
맞다고 생각했던 가설이 정말 맞는지
여태 그냥 해왔던 것들을
숫자로 제대로 파악하고 싶다

그리고 다음 목적을 세우게 하는 것

3. 프로젝트 유형별 접근

과거를 분석, 현재를 개선, 미래를 예측

“과거를 분석한다는 것”

순번	데이터 사이트	데이터 명칭	데이터 설명 (요약)	
1	서울시 빅데이터캠퍼스	서울시민의 업종별 카드소비 패턴 데이터	국내 전 지역 가맹점을 대상으로 서울시민의 개인카드 (신용카드와 체크카드 포함, 법인카드 제외) 사용한 패턴을 집계 (서울시민 거주지 기준으로 가맹점소재지별/일별/업종대분류별, 업종대분류별/일별/시간대별, 업종소분류별/월별/상월/연령대별로 카드이용금액과 이용건수, 업종소분류별/월별/시간대별로 카드이용건수와 소액결제건수를 집계)	https://bigdata.seoul.go.kr/data/selectSampleD
2	서울시 빅데이터캠퍼스	서울시민 KCB 생활금융 통계	경제 활동 중인 시민의 금융 정보를 기로로 가공된 소득, 부채, 지출 등 데이터. 개인 금융정보를 거주지 기반 가구화(가구/행정동 기준)를 통해 총 소득, 부채 등 추정하여 행정동/성/연령대별 급여소득자/자영업자/기타소득자의 생활수입 소득 등을 집계	https://bigdata.seoul.go.kr/data/selectSampleD
3	서울시 빅데이터캠퍼스	서울시 업종별 내외국인 신한카드 매출데이터	서울시 내에 위치한 가맹점의 카드이용 건을 대상으로 내국인은 개인카드 (신용카드/체크카드) 및 법인카드 이용 내역을 포함하고, 외국인은 외국인의 카드이용 내역을 포함	https://bigdata.seoul.go.kr/data/selectSampleD
4	서울시 빅데이터캠퍼스	서울시 지하철 30분단위 이용 통계	*서울시 대중교통 및 지하철 1회권 승하차 데이터* 중 거래내역 데이터를 출발 시간을 기준으로 30분 단위씩 지하철역별 승하차 인원을 집계	https://bigdata.seoul.go.kr/data/selectSampleData.d
5	서울시 빅데이터캠퍼스	서울시 버스 30분단위 이용 통계	*서울시 대중교통 및 지하철 1회권 승하차 데이터* 중 거래내역 데이터를 출발 시간을 기준으로 30분 단위씩 버스정류장별 승하차 인원을 집계	https://bigdata.seoul.go.kr/data/selectSampleD
6	서울시 빅데이터캠퍼스	서울시 00~14년 집계구별 사업체 통계	집계구 단위 산업분류별 사업체수와 종사자수 통계정보	https://bigdata.seoul.go.kr/data/selectSampleD
7	서울시 빅데이터캠퍼스	서울시 19년도 소비행태분석을 위한 신한은행 금융정보	신한은행 거래고객의 연간 금융정보를 상권/그리드 단위, 기준 년/월별 소득(급여, 가맹점 매출, 연금), 소비 (신용/체크카드 현금인출), 금융자산 (총 수신평균잔액) 등의 건수와 총금액에 대한 금융정보를 포함한 ※ 상권은 골목상권, 발달상권, 진통시장, 관광특구로 분류	https://bigdata.seoul.go.kr/data/selectSampleD
8	서울시 빅데이터캠퍼스	서울시 19년도 소비행태분석결과 샘플	신한은행의 소득, 소비, 금융자산 정보와 LG+U+IPTV 채널 종류별 시청 시간 정보, CJ올리브네트웍스의 온라인 구매 상품(중분류)별 주문량, 서울시에서 제공하는 골목상권 배후지의 상주인구 정보 CJ대한통운 택배 데이터를 지치구/블록 단위로 출지와 착지 데이터를 각각 추출하여 유형별로 월 송장건수를 집계	https://bigdata.seoul.go.kr/data/selectSampleData.d
9	서울시 빅데이터캠퍼스	서울시 CJ대한통운 택배 유형별 월 데이터	서울시 주요 공공 빅데이터, 교통연구원의 교통플리커와 KT의 휴대폰 LTE+5G시그널 데이터 이용하여 생활이동 데이터를 개발 (특정 시점, 특정 지역간에 이동하는 서울의 모든 인구로, 서울 안에서 이동하거나 서울 외부에서 서울로 오고 간 모든 사람들의 이동을 측정)	https://bigdata.seoul.go.kr/data/selectSampleData.d
10	서울시 빅데이터캠퍼스	서울시 내국인 KT 생활이동 데이터	매년 토지소재지 지치구청장이 결정공시하는 번지별 1㎡당 토지단가(원/㎡) 정보	https://bigdata.seoul.go.kr/data/selectSampleD
11	서울시 빅데이터캠퍼스	서울시 번지별 공시지가 데이터	내국인의 매출금액과 매출건수는 집계구 단위로 성/연령 별, 요일/시간대/요일지 별 25개 업종에 대하여, 외국인의 매출금액과 매출건수는 국가기초구역 단위로 요일/시간대/국가 별 22개 업종에 대하여 집계	https://bigdata.seoul.go.kr/data/selectSampleD
12	서울시 빅데이터캠퍼스	서울시 15~17년 업종별 내외국인 신한카드 매출데이터	서울시에서 매년 개최되는 3개 관광문화 축제 (김장문화축제, 빛초롱축제, 하이서울페스티벌)를 대상으로 도 5분거리의 영역에 대한 업종별, 성별, 연령대별, 시간대별로 매출금액과 건수가 집계된 신한카드 매출 데이터	https://bigdata.seoul.go.kr/data/selectSampleD
13	서울시 빅데이터캠퍼스	서울시 3개 관광문화 축제 신한카드매출데이터	블록별 43개 업종별 혹은 3개 대분류 업종별로 성별, 연령대별, 요일별, 시간대별로 매출금액과 매출건수를 집계한 KB국민카드 매출데이터	https://bigdata.seoul.go.kr/data/selectSampleD
14	서울시 빅데이터캠퍼스	서울시 블록별 43개 업종별 KB국민카드 매출데이터	서울특별시 공공자전거 시간대별 이용정보 (대여일시, 대여시간, 대여소번호, 대여소명, 정기권유무, 성별, 연령대, 한살림, 이동거리, 이동시간 정보를 제공)	http://data.seoul.go.kr/dataList/OA-15245/F/1/
15	서울열린데이터광장	서울특별시 공공자전거 이용정보 (시간대별)	서울시 기준의 코로나19 백신 접종률 현황	http://data.seoul.go.kr/dataList/OA-20914/S/1/
16	서울열린데이터광장	서울특별시 코로나19 백신 예방접종 현황	서울시 골목상권 영역 내 경포의 추정매출 정보를 제공	http://data.seoul.go.kr/dataList/OA-15572/S/1/
17	서울열린데이터광장	서울시 우의마을가게 상권분석서비스 (상권-추정매출)	서울시 골목상권 영역 내 직장인구 정보를 제공	http://data.seoul.go.kr/dataList/OA-15569/S/1/
18	서울열린데이터광장	서울시 우의마을가게 상권분석서비스 (상권-직장인구)	5개 수질자동측정소에서 측정된 서울시 한강 및 주요지천 수질 측정 자료	http://data.seoul.go.kr/dataList/OA-15488/S/1/
19	서울열린데이터광장	서울특별시 한강 및 주요지천 수질 측정 자료	서울시농수산식품공사의 식품 판매 결과 정보	http://data.seoul.go.kr/dataList/OA-20950/S/1/
20	서울열린데이터광장	서울시농수산식품공사 식품 판매 결과	서울글로벌센터 월별 상담실적률 접수유형별/분류별로 제공	http://data.seoul.go.kr/dataList/OA-15739/S/1/
21	서울열린데이터광장	서울글로벌센터 월별 접수유형별 분류별 상담실적	서울글로벌센터 분기별 상담 실적률 제공	http://data.seoul.go.kr/dataList/OA-15737/S/1/
22	서울열린데이터광장	서울글로벌센터 분기별 상담 실적률	서울왕리리포털에서 제공하는 누담일자리 사업정보	http://data.seoul.go.kr/dataList/OA-15355/S/1/
23	서울열린데이터광장	서울특별시 누담일자리 사업정보	서울시 물가미터가 주2회 지치구별 진통시장과 대형마트에 나가 농수축산물 16개 품목을 조사하고 그 가격을 공개하는 정보	http://data.seoul.go.kr/dataList/OA-1170/S/1/dataset
24	서울열린데이터광장	서울시 생활물 농수축산물 가격 정보	개인서울요금 관리대상업소의 효율적 관리 및 소비자에게 업소별 가격정보 제공을 위한 자료	http://data.seoul.go.kr/dataList/OA-1169/S/1/c
25	서울열린데이터광장	서울시 개인서비스 요금 정보	가격이 저렴하고 서비스가 좋은 가격인정보업소 (착한가게) 정보에 대한 서비스	http://data.seoul.go.kr/dataList/OA-1173/S/1/c
26	서울열린데이터광장	서울시 착한가격업소 현황	가격이 저렴하고 서비스가 좋은 가격인정보업소 (착한가게)들의 상품목록 정보에 대한 OpenAPI 서비스	http://data.seoul.go.kr/dataList/OA-1174/S/1/c
27	서울열린데이터광장	서울시 가격인정 모범업소 상품목록 현황	서대문구의 소비자를 직접 방문하여 화장품 등 재화 또는 용역을 판매하는 업소정보	http://data.seoul.go.kr/dataList/OA-18764/S/1/
28	서울열린데이터광장	서울특별시 서대문구 방문판매업 인허가 정보	교육기부활동조사규정에 따라 초등학생 현황을 제공하는 지정 · 보고통계	https://data.seoul.go.kr/dataList/201/S/2/datas
29	서울열린데이터광장	서울시 초등학교 통계	대기 환경지수, 미세먼지, 오존, 이산화질소, 일산화탄소, 이황산화가스 등의 평균 대기오염도 년별 정보를 제공	https://data.seoul.go.kr/dataList/OA-2216/S/1/
30	서울열린데이터광장	서울시 년도별 평균 대기오염도 정보	중관구로의 날씨를 파악하기 위하여 정해진 시기에 모든 관측소에서 같은 시기에 실시하는 지상관측, 기온, 강수, 바람, 기압, 습도, 일사, 일조, 눈, 구름, 시정, 지면상태, 지면 · 초상온도, 일기상상, 증발량, 현상번호 제공	https://data.kma.go.kr/data/grnd/selectAsosRtr
31	기상자료개발포털	중관구기상관측 (ASOS) - 파일셋	동해예보 구역(5km 해상도 격자)에 대한 대표 AWS 관측값. 2010년 6월부터 조회일 기준 전월 자료까지 제공	https://data.kma.go.kr/data/rmt/rmtList.do?cx
32	기상자료개발포털	초단기실황 - 파일셋		

특히, 목적을 결정하기 전에 데이터부터 수집되어 쌓여 있는 경우 대부분 회사가 그렇다 공공데이터도 그렇다

쌓인 데이터의 현재 상태를 정확히 짚고 용도와 활용성을 발견하고

목적을 찾아주는 그 자체가 의미가 있는 과정이다

“과거를 제대로 확인해야 다음 단계로 나아갈 수 있다”

3. 프로젝트 유형별 접근

-과거를 분석, 현재를 개선, 미래를 예측

“현재를 개선한다는 것”



과거에 대한 궁금증이 해결되고 나면
회사는 **빠르게 현재를 개선**하고 싶어한다
돈을 많이 쓰고 있는 쪽에, 그리고 **성장**에 매우 관심이 많다

3. 프로젝트 유형별 접근

-과거를 분석, 현재를 개선, 미래를 예측

“**현재를 개선한다는 것**”

공통된 문제

모두가 처음에는 고객이 없다
모두가 겪는 일이라 전문 분야로 특화

“그로스 해킹”

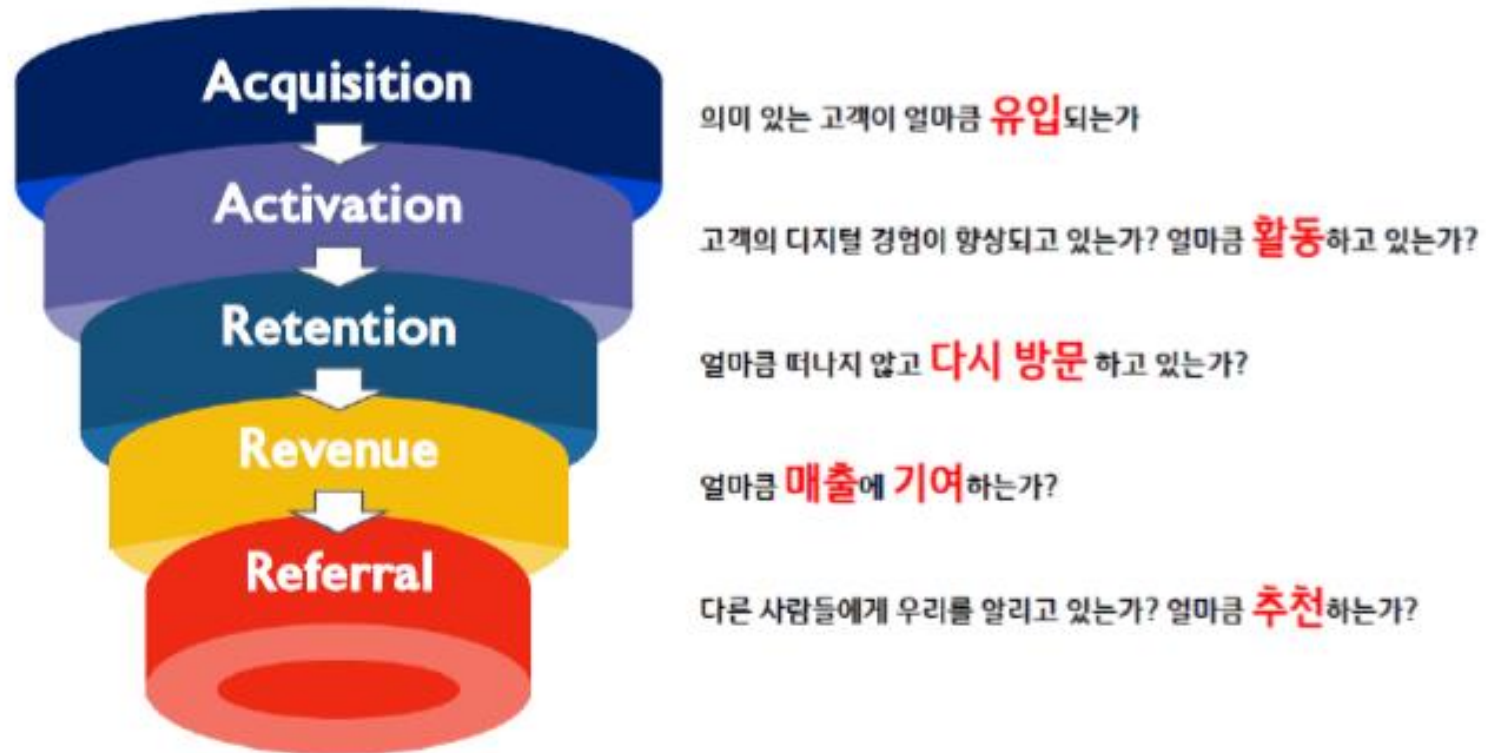
유저획득과 이탈방지, 성장에 집중한
디지털 마케팅 방법론의 등장

3. 프로젝트 유형별 접근

-과거를 분석, 현재를 개선, 미래를 예측

그로스 해킹은 **핵심 성과 지표(KPI) AARRR**을 통해
서비스의 현주소와 문제점을 파악하고 즉시 개선

AARRR Metrics Funnel Diagram



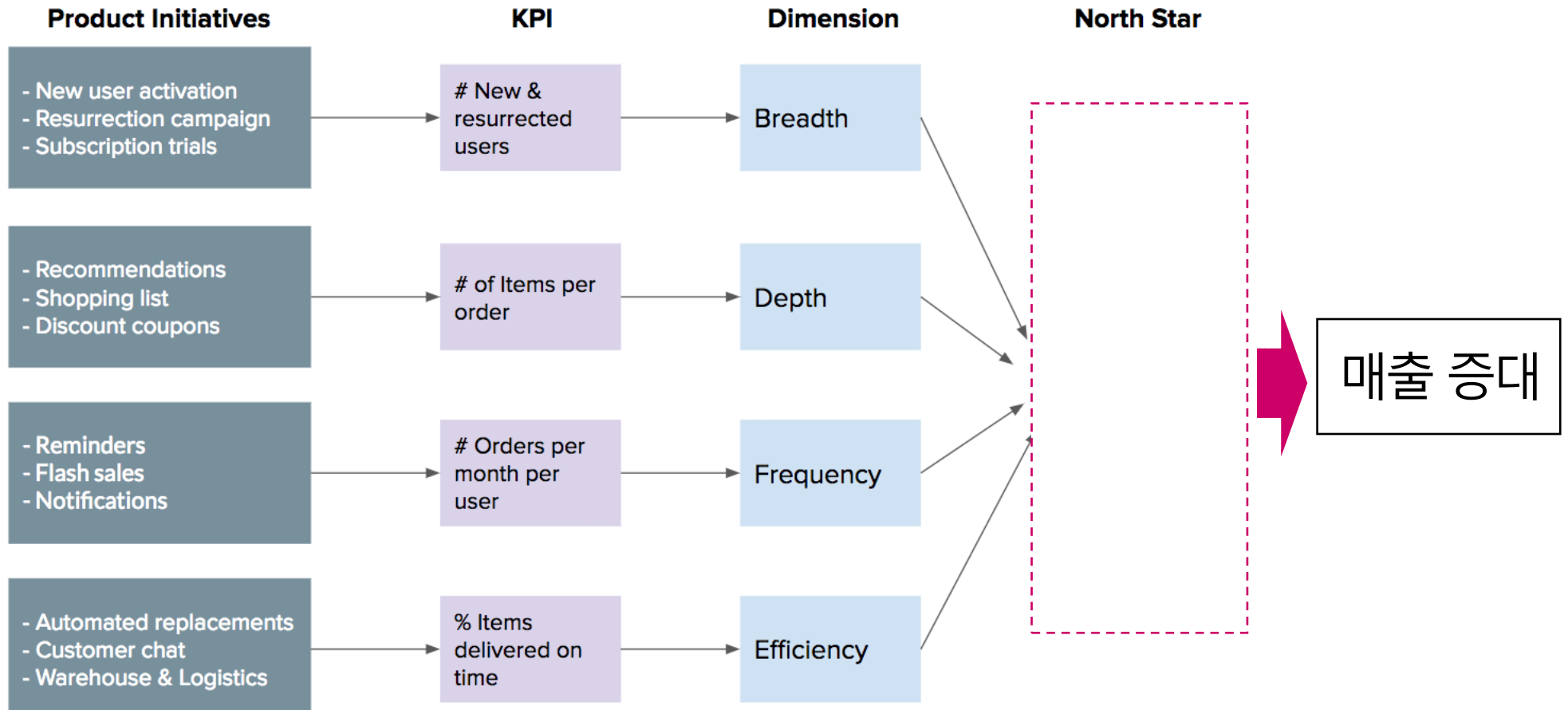
3. 프로젝트 유형별 접근

-과거를 분석, 현재를 개선, 미래를 예측

후행지표에 연관된 하위 선행지표 발굴

무엇의 결과로 일어난 것
개선할 방법이 바로 안떠오름

무엇의 원인이 되는 것
열심히만 해도 될 거 같음

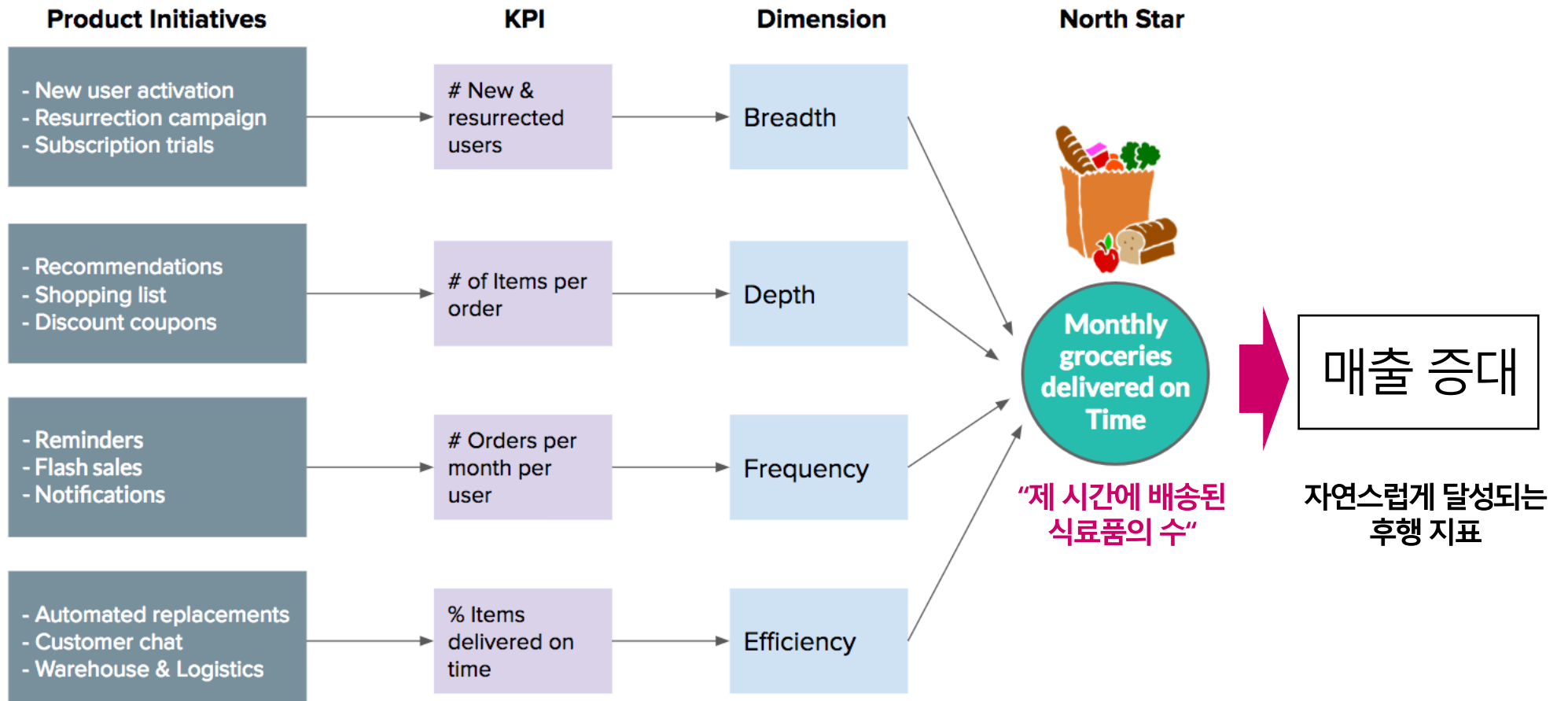


3. 프로젝트 유형별 접근

-과거를 분석, 현재를 개선, 미래를 예측

후행지표에 연관된 하위 선행지표 발굴

모든 하위 KPI를 복극성 지표의 상승을 위해 Align시켜 진정한 성장[Growth] 창출



3. 프로젝트 유형별 접근

-과거를 분석, 현재를 개선, 미래를 예측

마크 주커버그는 본인을 해커라고 정의하고
철저하게 그로스 해킹 방식을 비즈니스에 적용



10 REASONS WHY COMPANIES FAIL AT GROWTH

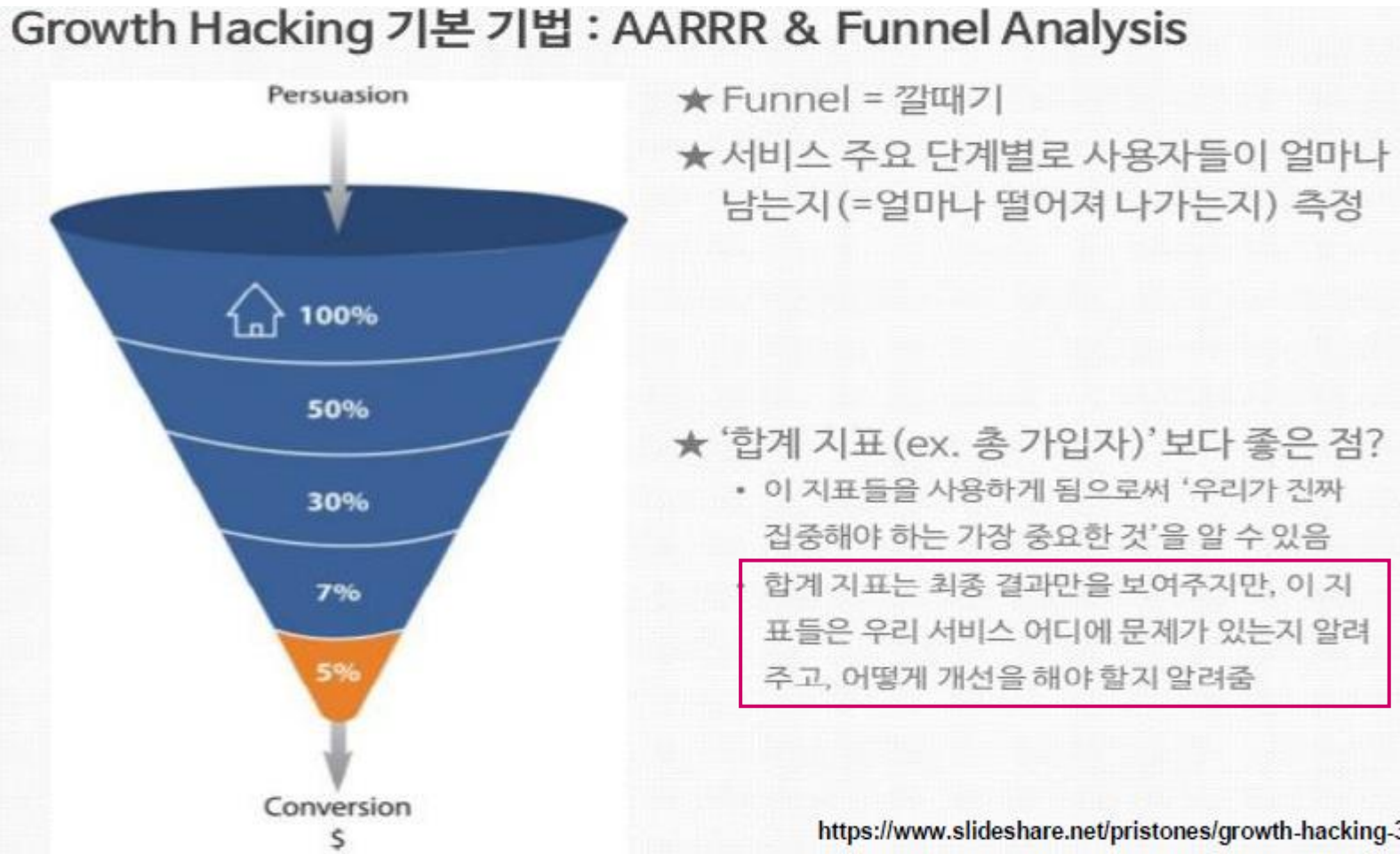
- | | |
|---|---|
| They don't focus on retention first. 1 | They don't run experiments. 6 |
| They subscribe to "product is everything" mantra. 2 | They don't dig in and learn. 7 |
| They search for a silver bullet. 3 | They don't double down. 8 |
| They don't focus. 4 | They don't dedicate the right resources towards growth. 9 |
| They don't invest enough in data and analytics. 5 | They don't change and adapt. 10 |

Brian Ballfour :: @bballfour :: <http://www.coclevate.com>

3. 프로젝트 유형별 접근

-과거를 분석, 현재를 개선, 미래를 예측

그로스 해킹의 목표는 고객이 각 단계별로 줄지 않도록
최선을 다해 숫자를 맞추는 것

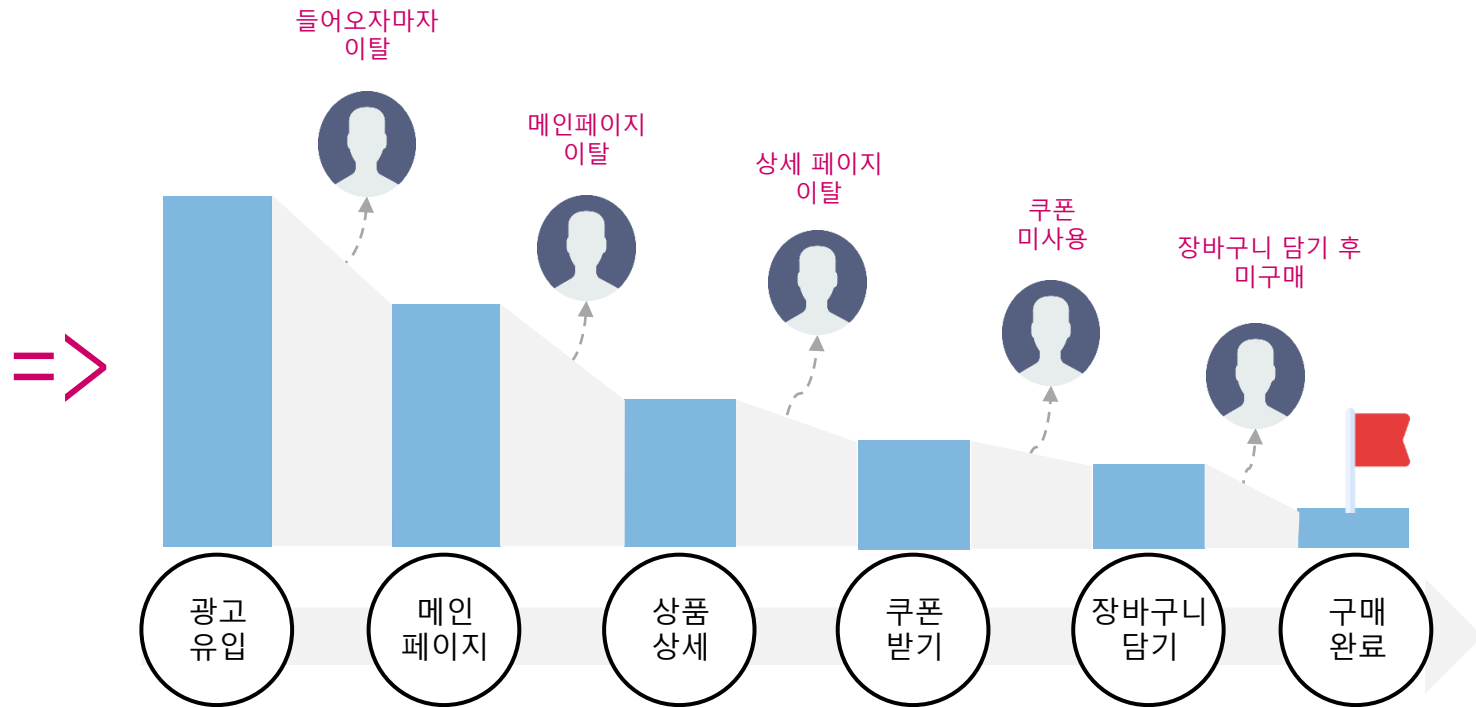


3. 프로젝트 유형별 접근

-과거를 분석, 현재를 개선, 미래를 예측

모두 다 하고 있는 그로스 해킹 실무

앱 유입 이후 단계별 이탈율 감소를 통한
최종 구매 전환율, 매출액 증대의 반복



3. 프로젝트 유형별 접근

-과거를 분석, 현재를 개선, 미래를 예측

“미래를 예측한다는 것”



앞의 두 단계가 매우 잘 진행되어야만 가능하다
엄청난 양의 잘 정리된 데이터, 엄청난 서버, 엄청난 엔지니어가 필요하다
우리 비즈니스에 과연 “예측”이 맞는지 ROI 검토가 핵심

3. 프로젝트 유형별 접근

-과거를 분석, 현재를 개선, 미래를 예측

딸기가 왔어요 **하루만 지나도 못 파는 딸기가**



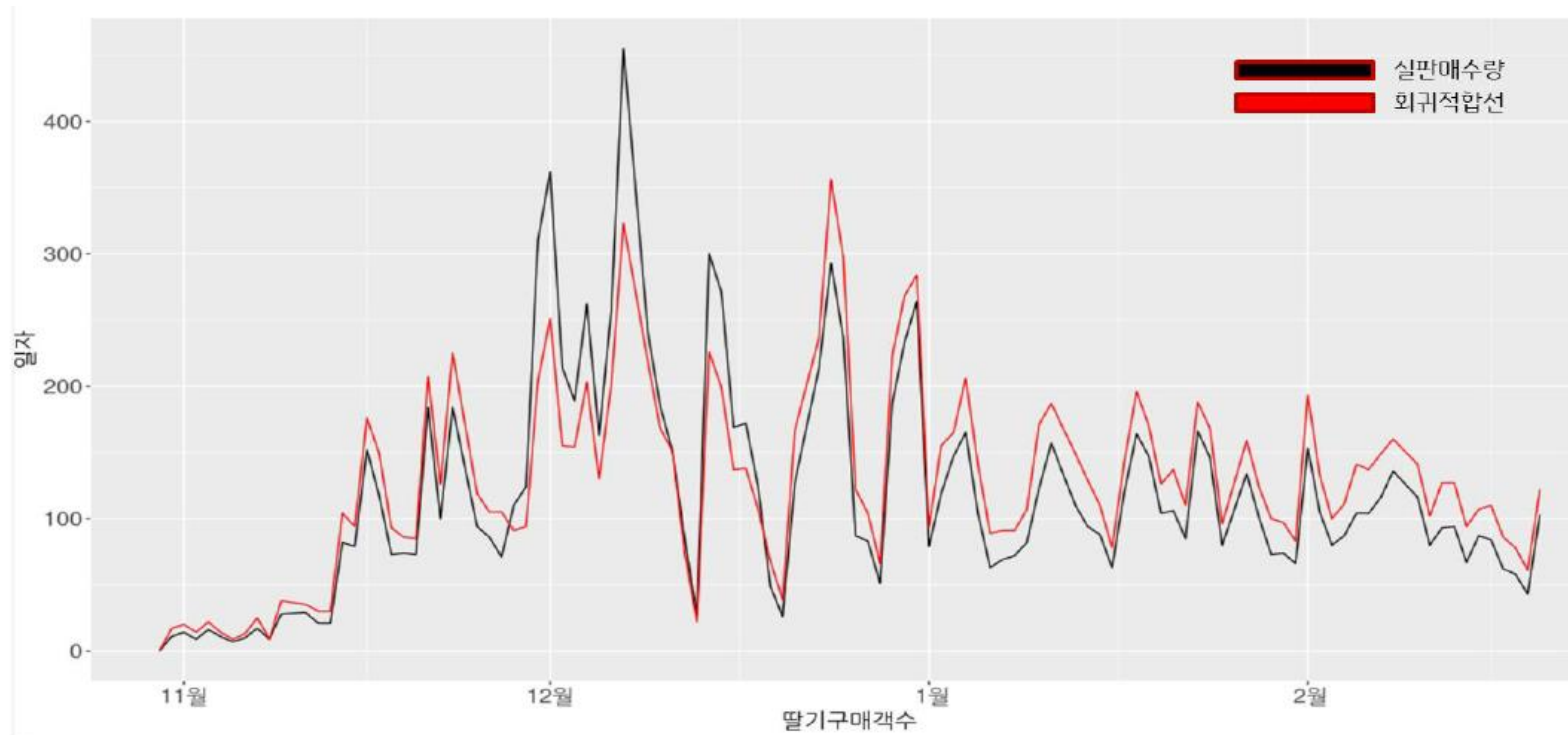
3. 프로젝트 유형별 접근

-과거를 분석, 현재를 개선, 미래를 예측

딸기 미래수요 예측 모델

- ✓ 데이터 : 00년간 일별 오프라인 / 0년간 실시간 온라인 주문내역 3년치라도 연단위집계면 데이터포인트 세개
- ✓ 엔지니어/서버 : 초기 개발비용 000억, 유지보수 연 00억
- ✓ 목적 : 딸기 발주/생산과 재고 관리 최적화로 비용감소/수익증대

딸기 구매객수 이용하여 83% 정확도로 적합한 시계열 모형



3. 프로젝트 유형별 접근

-과거를 분석, 현재를 개선, 미래를 예측

비즈니스의 본질을 고민해야 한다

과거

현황과 문제 파악
분석의 목적 도출
데이터의 용도 발견


현재

방문자 증대
재방문/재구매 증대
매출 개선

미래

개인화 추천
구매 예측
물류 예측

분석의 목적과 데이터가
비즈니스에 맞아야 의미가 있다
우리 비즈니스가 어떻게 구성되어 있는지를 생각하자

A top-down view of a business meeting around a table. Several people are seated around the table, looking at documents and laptops. The table is covered with various items including papers with charts and graphs, a coffee cup, and a laptop. The background is slightly blurred, focusing attention on the meeting activity.

4. 되게 만드는 ppt

- EDA/시각화/발표자료 작성 단계별 주요 체크

4. 되게 만드는 ppt

_EDA/시각화/발표자료 작성 단계별 주요 체크

분석가에게는 코드 한줄 그림 한장이 너무나도 소중한
다 보여 주고 싶겠지만, 듣는 사람의 입장에서

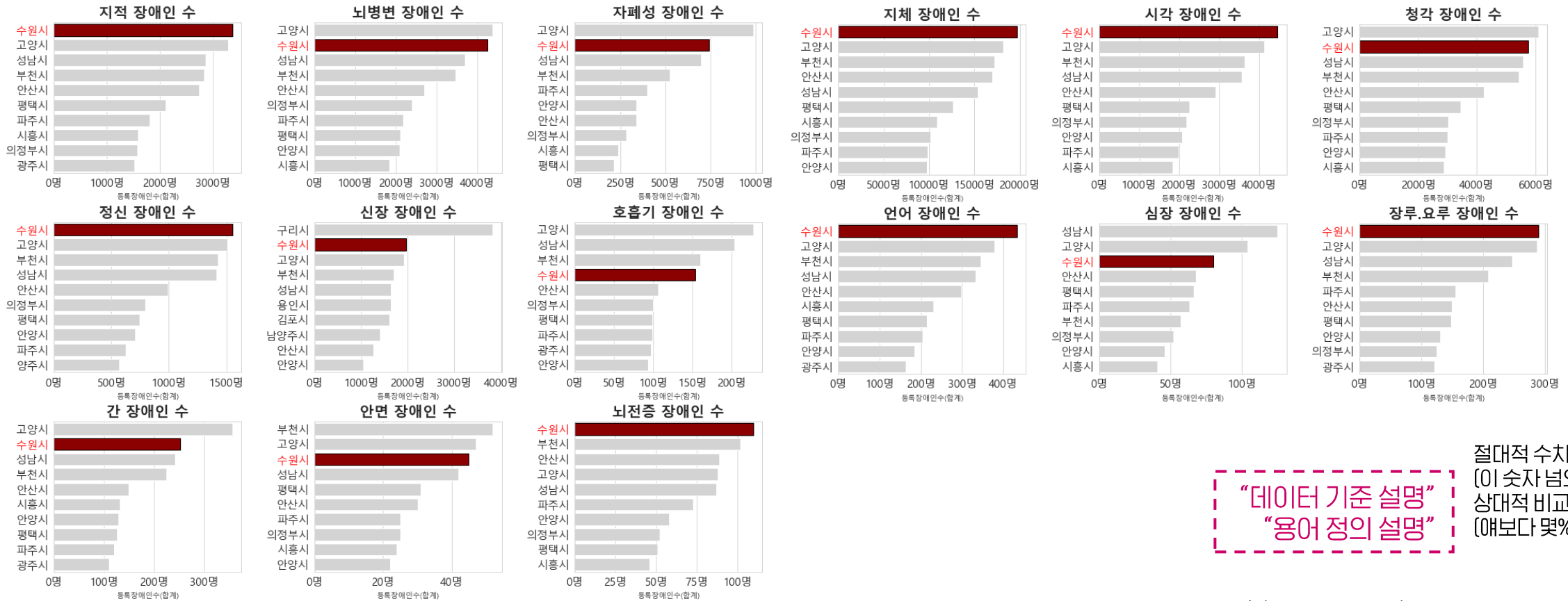


4. 되게 만드는 ppt

_EDA/시각화/발표자료 작성 단계별 주요 체크

전체 EDA는 모아서 한장으로 EDA 단을 하나하나씩 모두 설명하면 망

“Implication” - EDA로 이끌어낸 전체 데이터의 함축적 의미 (두줄정도)



“데이터 기준 설명”
“용어 정의 설명”

절대적 수치
(이 숫자 넘으면 좋다)
상대적 비교
(애보다 몇% 좋다)

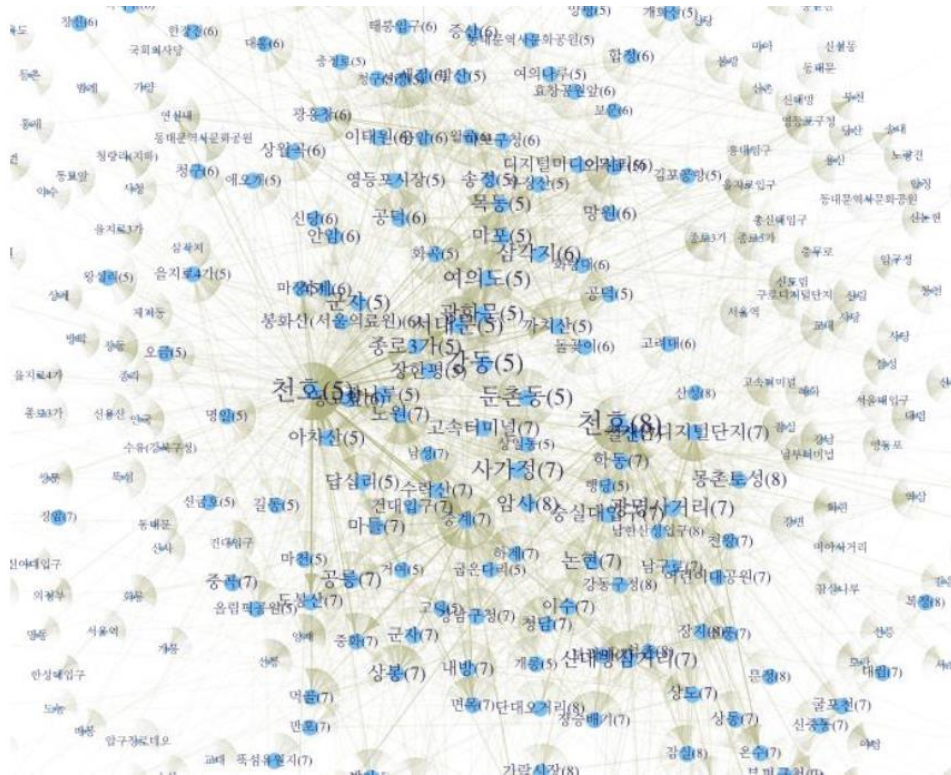
*출처 : 2021 데멘토 최우수팀 <에이블> 발표자료

4. 되게 만드는 ppt

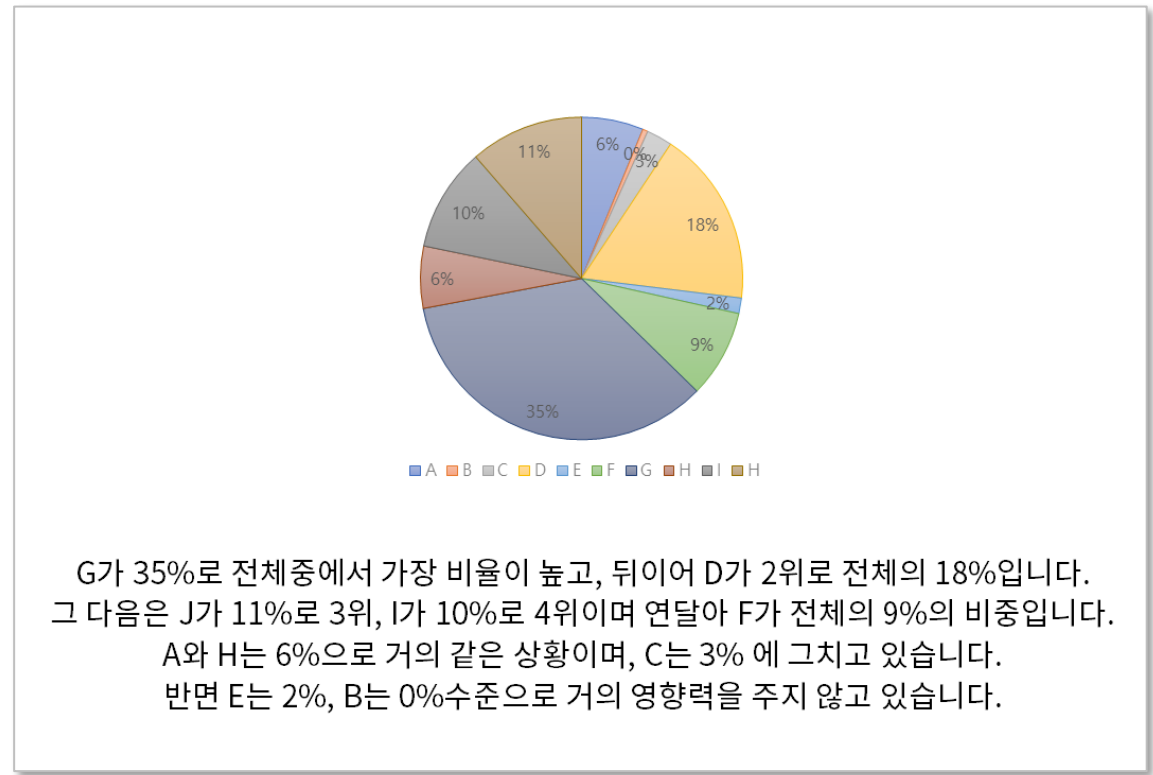
_ EDA/시각화/발표자료 작성 단계별 주요 체크

그래프가 스스로 말하게 하자

시각화의 목적은 상대방의 빠른 이해
화려한 비주얼라이제이션은 그닥 유용하지 않다 예를 들어 네트워크



특히 파이차트는 쓰지 마세요
이건 비교가 잘 안돼요. 그래프를 다시 설명하게 됩니다



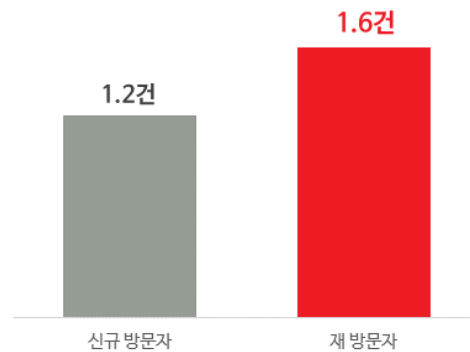
4. 되게 만드는 ppt

_EDA/시각화/발표자료 작성 단계별 주요 체크

데이터는 주장을 해야 한다

읽는 순서에 따라, 글로 된 Insight(주장)를 위에 쓰고
그 근거를 하단의 차트가 도와주는 순서가 좋다 ㅎㅎㅎ

인당구매건수는 신규방문자 1.2건, 재방문자 1.6건으로 재방문고객의 LTV가 훨씬 큼
매출 제고를 위해 기존 고객의 재방문 유도 위한 Retention 전략이 효과적

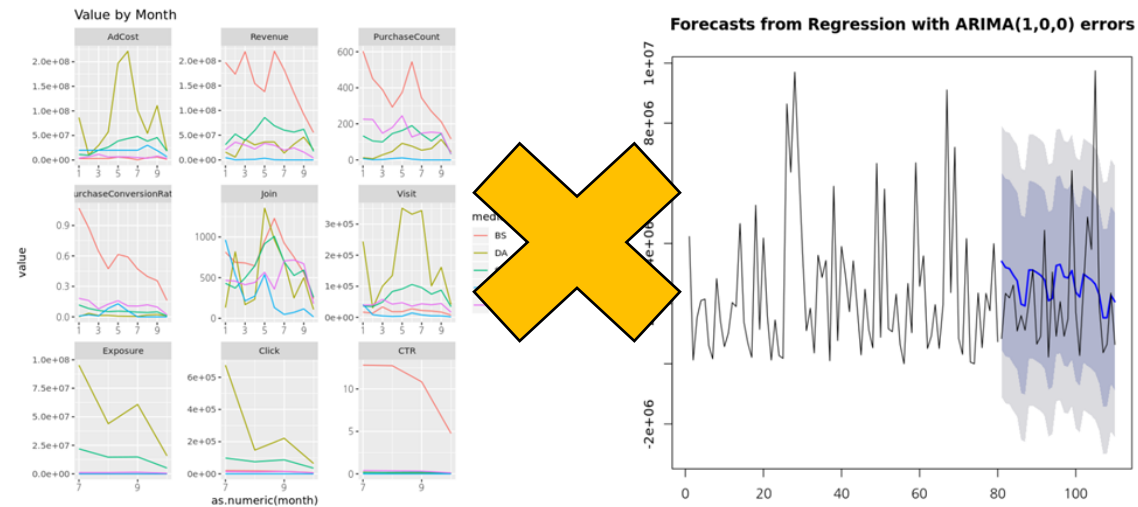


4. 되게 만드는 ppt

_EDA/시각화/발표자료 작성 단계별 주요 체크

분석가가 발표자가 될때의 함정

분석을 수행한 순서 그대로 발표
분석의 전문용어 그대로 사용



총 개월간 다양한 채널에서 수집된 A 제품의
미디어별 마케팅 비용 데이터와 구매 데이터를 융합 분석했습니다
ARIMA 시계열 분석, 회귀분석을 통해 여러 채널에서 집행된
캠페인들 간의 상호 영향력을 시계열적, 정량적으로 파악했습니다

4. 되게 만드는 ppt

_EDA/시각화/발표자료 작성 단계별 주요 체크

발표는 재정리해서 두괄식으로

결론부터! 발표 순서는 분석의 역방향

상대방의 쉬운 언어 사용 요즘은 중학생도 똑똑하다 초등 5학년 눈높이

결론

SUMMARY

-> 들을 준비 만들기

“이런 문제가 있었는데 잘 풀어서 이렇게 좋아합니다”

이유

데이터와 사례로
결론의 정당성 지원

“이런 조사/분석을 수행했고 이런 이유로 채택했습니다”

효과

정량적/정성적 이득
단기적/장기적 효과

“이런이런 이득이 있고 언제까지/얼마만큼 가능합니다”

재정리

다른 표현으로 반복/강조
확실한 전달과 설득의 마침표

“결론적으로, 이렇게 되었으니 이렇게 해 주셨으면 합니다”

4. 되게 만드는 ppt

_EDA/시각화/발표자료 작성 단계별 주요 체크

논리적 설득법 PREP

Point (결론) :

“의사결정”을 위해서는 명확한 결론을 앞에

Reason (이유) :

“왜”에 대한 대답. 전체 설명에서 세부적 이유로

전국 -> 시/도 -> 지역
전체 -> 인당/개당/면적당

Example (사례) :

이유를 확실하게 해 주는 구체적 백업

Point (결론) :

처음과 끝의 공통성 있는 내용 재강조로 방점

4. 되게 만드는 ppt

_EDA/시각화/발표자료 작성 단계별 주요 체크

채워지지 않는 부분은
-> 컴퓨터 밖에서 질문하기

한정적인 데이터에서
모든 것을 얻으려 하기 쉽지만 답은 T/F

어떤 것이 문제의 원인인가?
어떤 것이 있었으면 좋겠는가?
이 결과는 납득이 가는가?
'열린' 질문에 대한 답은 고객이 쥐고 있다

채널을 가리지 말자
컴퓨터 밖도 좋은 세상이다




5. 은근히 알려지지 않은 몇가지

- 비밀이라서가 아니라 당연해서

5. 은근히 알려지지 않은 몇가지 _비밀이라서가 아니라 당연해서


“코딩은 독학도 좋습니다”



오늘코드todaycode

구독자 2.27만명 · 동영상 321개

공공데이터 분석 데이터 시각화 캐글을 통한 머신러닝/딥러닝 튜토리얼 Pandas, Numpy, Scipy, scikit-learn, TensorFlow, Keras, ...



나도코딩

구독자 28.6만명 · 동영상 258개

코딩, 쉽고 재미있게 배우세요 nadocoding@gmail.com.

파이썬 공공데이터 분석[1/6] - 신규 민간 아파트 분양가격 데이터셋 소개

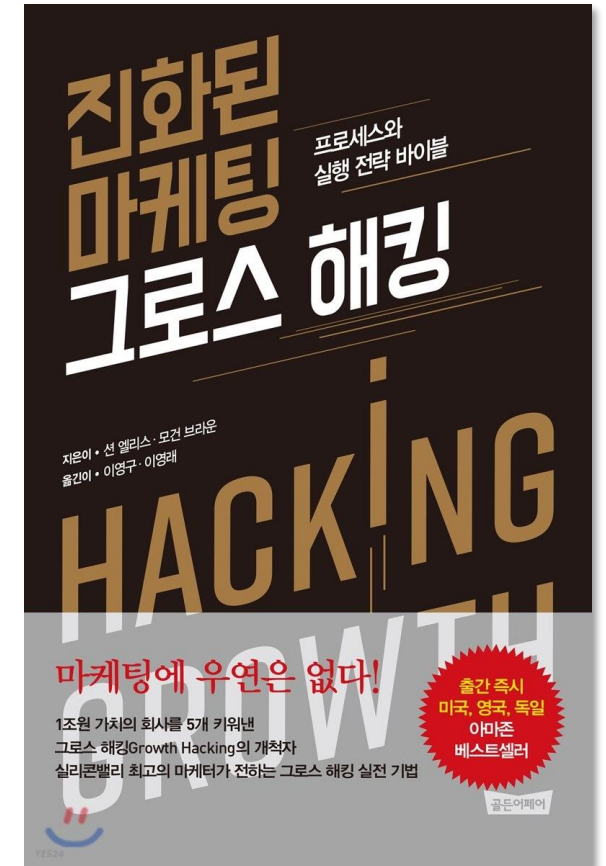
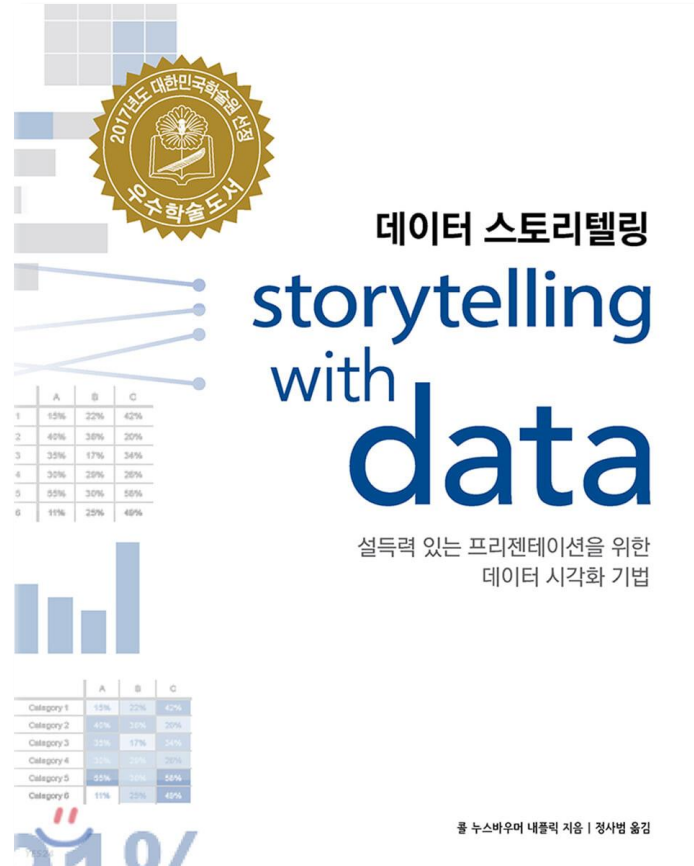
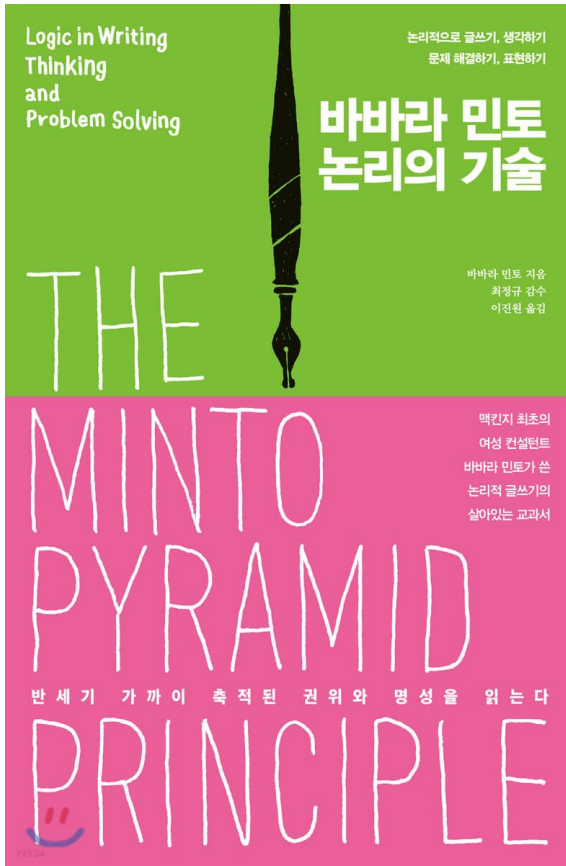
<https://www.youtube.com/c/todaycode>

#파이썬 #인공지능 #머신러닝
파이썬 코딩 무료 강의 (활용편7) - 머신러닝, 영화 추천 시스템 만들기

<https://www.youtube.com/channel/UC7iAOLiALt2rtMVAWWl4pnw>

5. 은근히 알려지지 않은 몇가지
_비밀이라서가 아니라 당연해서

“분석보다 관점이 중요합니다”



5. 은근히 알려지지 않은 몇가지 _비밀이라서가 아니라 당연해서

“코딩이 필수는 아닙니다”

<https://orangedatamining.com/>

Python
No-Code tool



Screenshots

Data Mining
Fruitful and Fun

Open source machine learning and data visualization.

Build data analysis workflows visually, with a large, diverse toolbox.

Download Orange

https://www.youtube.com/results?search_query=%EC%98%A4%EB%A0%8C%EC%A7%80+%EB%8D%B0%EC%9D%B4%ED%84%B0%EB%B6%84%EC%84%9D

<https://www.jamovi.org/>

R
No-Code tool



	len	supp	dose	Residuals
1	4.2	VC	500	-3.78
2	11.5	VC	500	3.52
3	7.3	VC	500	-0.68
4	5.8	VC	500	-2.18
5	6.4	VC	500	-1.58
6	10.0	VC	500	2.02
7	11.2	VC	500	3.22
8	11.2	VC	500	3.22
9	5.2	VC	500	-2.78

https://www.youtube.com/results?search_query=%EC%9E%90%EB%AA%A8%EB%B9%84+%EB%8D%B0%EC%9D%B4%ED%84%B0%EB%B6%84%EC%84%9D

<https://www.listly.io/>

Crawling
No-Code tool

LISTLY | How



All it takes is one click — web scraping

https://www.youtube.com/results?search_query=%EB%A6%AC%EC%8A%A4%ED%8B%80%EB%A6%AC+%ED%81%AC%EB%A1%A4%EB%A7%81

5. 은근히 알려지지 않은 몇가지 _비밀이라서가 아니라 당연해서

“데이터는 찾아야 하고, 목적도 만들어야 합니다”

한국지능정보사회진흥원 K-ICT 빅데이터 센터

<https://kbig.kr/>

한국지능정보사회진흥원 AI Hub > AI데이터

https://aihub.or.kr/ai_data

공공데이터 포털

<https://www.data.go.kr>

* 참고 - [공공데이터 개방 표준](#)

서울 열린데이터 광장

<http://data.seoul.go.kr>

서울특별시 빅데이터캠퍼스

https://bigdata.seoul.go.kr/data/selectPageListDataSet.do?r_id=P211

부산 공공데이터 포털

<https://data.busan.go.kr/>

인천 데이터 포털

<https://www.incheon.go.kr/data/DATA010201>

경기데이터드림

<https://data.gg.go.kr/>

문화·관광 빅데이터 분석대회

<http://www.tourbigdata.kr/>

경남 공공데이터 포털

<http://open.gyeongnam.go.kr/index.gyeong>

강원 공공데이터

<https://data.gwd.go.kr/>

한국데이터산업진흥원 데이터스토어

<https://www.datastore.or.kr/>

건강보험심사평가원 보건의료빅데이터개방시스템

<https://opendata.hira.or.kr/>

e-나라지표 - 국정모니터링지표

<http://www.index.go.kr/>

통계청

<http://kostat.go.kr/>

KOSIS - 국가통계포털

<http://kosis.kr/>

기상청 기상자료개방포털

<https://data.kma.go.kr/>

날씨 빅데이터 콘테스트(기상청)

<https://bd.kma.go.kr/contest/>

환경부 실시간수질정보시스템

http://www.koreawqi.go.kr/index_web.jsp

교육부 나이스 교육정보 개방 포털

<https://open.neis.go.kr/portal/data/dataset/searchDatasetPage.do>

한국전자통신연구원 공공데이터개방

https://www.etri.re.kr/korcon/sub1/sub1_07.etri

농업관측통계정보시스템

<https://oasis.krei.re.kr/>

고속도로 공공데이터 포털

<http://data.ex.co.kr/>

네이버 데이터랩

<https://datalab.naver.com/>

5. 은근히 알려지지 않은 몇가지 _비밀이라서가 아니라 당연해서

“데이터와 목적이 주어지면 성능을 소수점 세째자리까지”

서버/계산량/시간이 문제가 되지 않는 경우
한정적인 비즈니스 상황이라면
1달 걸려 5% 올리는건 의미가 없는 경우도 있다

국내

WEVIFY (공모전 공고 모음) <https://wevity.com/?c=find&s=1&gub=1&cidx=21>

데이콘 <https://dacon.io/>

네이버 AI RUSH <https://campaign.naver.com/airush/>

빅데이터콘테스트 <https://www.bigcontest.or.kr/>

삼성 AI 챌린지 <https://www.sait.samsung.co.kr/saithome/main/main.do>

카카오 아레나 <https://arena.kakao.com/>

디지털 헬스 해커톤(삼성융합의과학원)
<https://www.digitalhealthhack.org/>

해외

Kaggle

<https://www.kaggle.com/>

KDD Cup

<https://www.kdd.org/kdd-cup>

일본 SIGNATE 경진대회

<https://signate.jp/competitions>

Analytics Vidhya Data Science Hackathon

<https://datahack.analyticsvidhya.com/>

Open Data Hackathon

<http://opendatacy.com/>

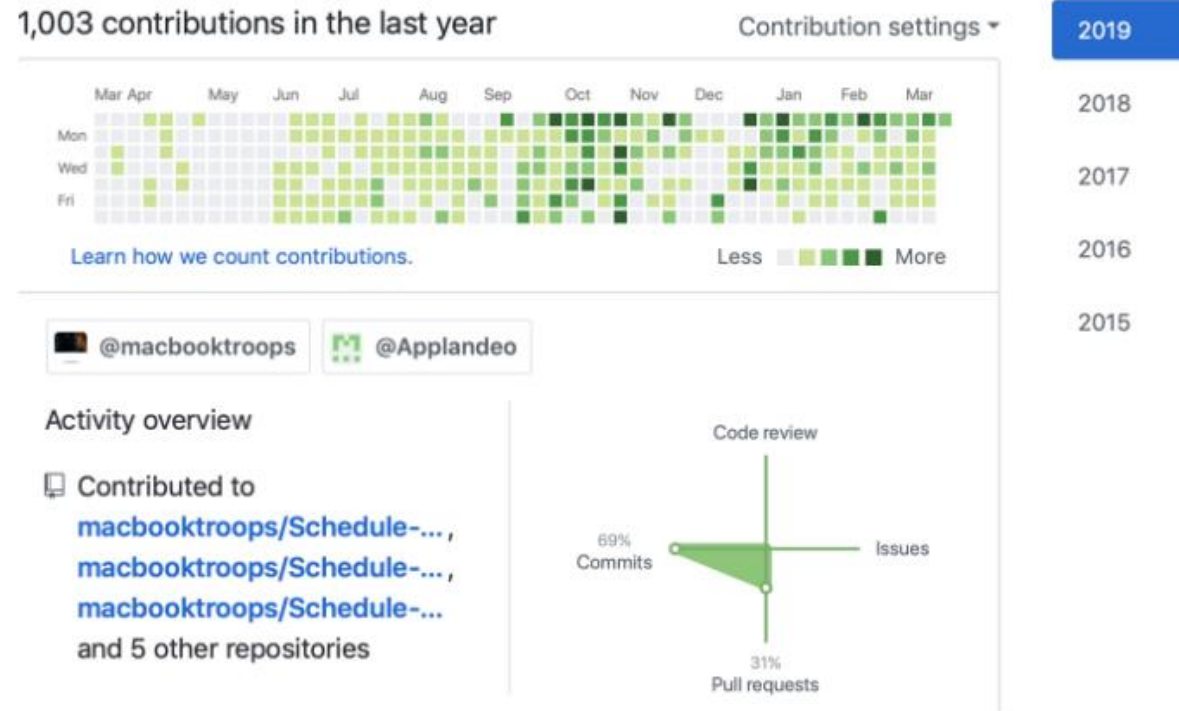
Asia Open Data Hackathon

<http://odhack.asia/>

5. 은근히 알려지지 않은 몇가지 _비밀이라서가 아니라 당연해서

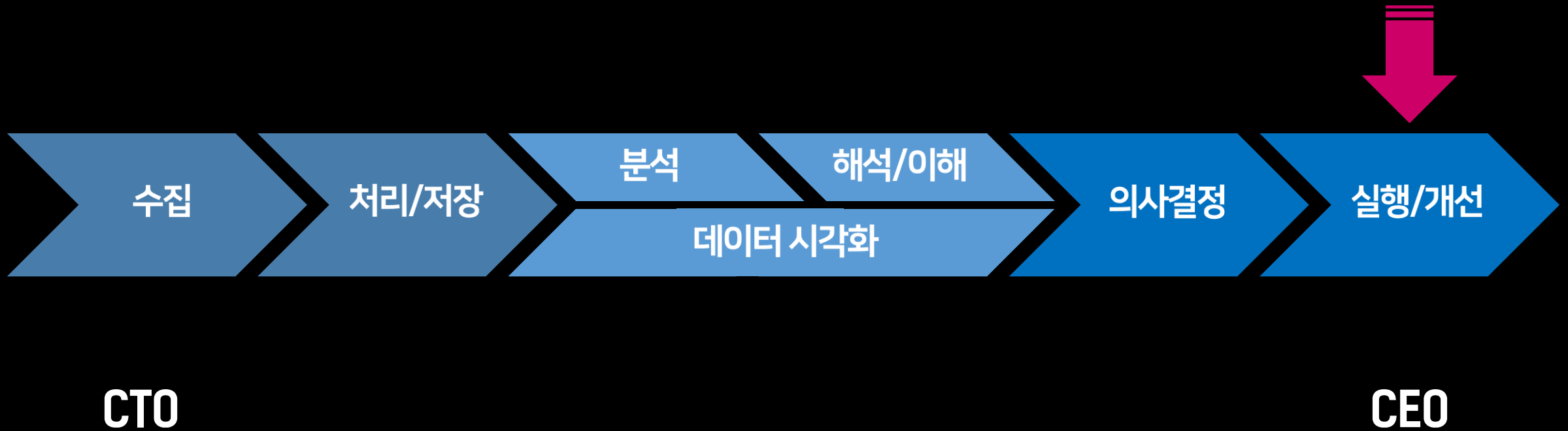
“기록을 남겨야 포트폴리오가 됩니다”

Github 1,000 commit 회고록



마무리하며

데이터 사이언스가 **마케팅**으로 왔다는 의미는
분석의 관점 R&D에서 **회사의 이익 Business**이라는 관점으로



■ 마무리하며

데이터를 통해 무언가를 분석한다는 것은
단순히 수치를 만들고 보는 것이 아니다.

데이터로 무언가를 분석한다는 것은
그래서 지금 우리 회사 혹은 서비스가 어떤 구조속에서
어떻게 작동하고 있었고, 있는 것이며,
앞으로 어떻게 작동해야 한다는 것을 이해한다는 것이다.

즉, 마지막의 마지막에 가서 보면 데이터 분석가는
**임원 레벨의 큰 청사진을 보고, 이해하고, 설계하고,
또 그 청사진을 성취하기 위해 투쟁할 수 있는 사람이어야만 한다.**



Thank You.

Ahn, Joongho

bakky14@gmail.com