

Marketing & Data Science Meetup

Bigdata for Marketers

2019.07.24.

안중호/대흥기획

- analytics
- infographic
- word cloud
- data science
- data analytics
- companies
- spark
- definition
- hadoop
- predictive analytics
- virality aha
- 2018
- business
- machine learning
- data analysis



Article] We Are Big Data: New Technolog...
eduardomagrani.com



Big data - Wikipedia
en.wikipedia.org



rhipe | Big Data and the Cloud - rhipe
rhipe.com



Data Science vs. Big Data vs. Data Analytics
simplilearn.com



10 Key Technologies that enable Big Data Analyti...
towardsdatascience.com



The Future of Big Data is Cloudy -- and Bumpy | Transformi...
tdwi.org



Getting Started With Big Data
digitalvidya.com



Where does Big Data go from here? - SD TI...
sdtimes.com



Big Data and the Cloud: A Perfect Combinati...
datafoq.com



5 Innovative Ways Small Companies Can Collect B...
smartdatacollective.com



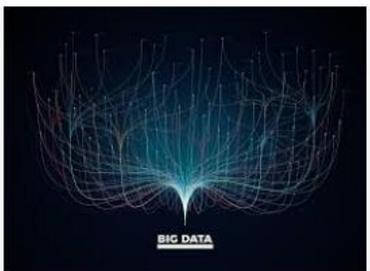
Big Data 2018: Cloud storage becomes the de facto d...
zdnet.com



Big Data in Healthcare Market Value Share o...
medgadget.com



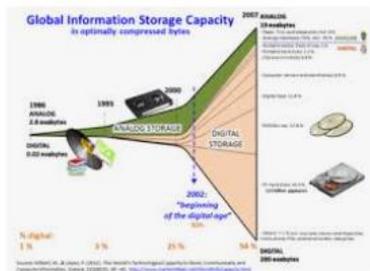
5 Things You Need to Know abo...



Mayer Brown's Tech Talks, Episode 3: The...



Big Data Analytics | Information Builders



Big data - Wikipedia

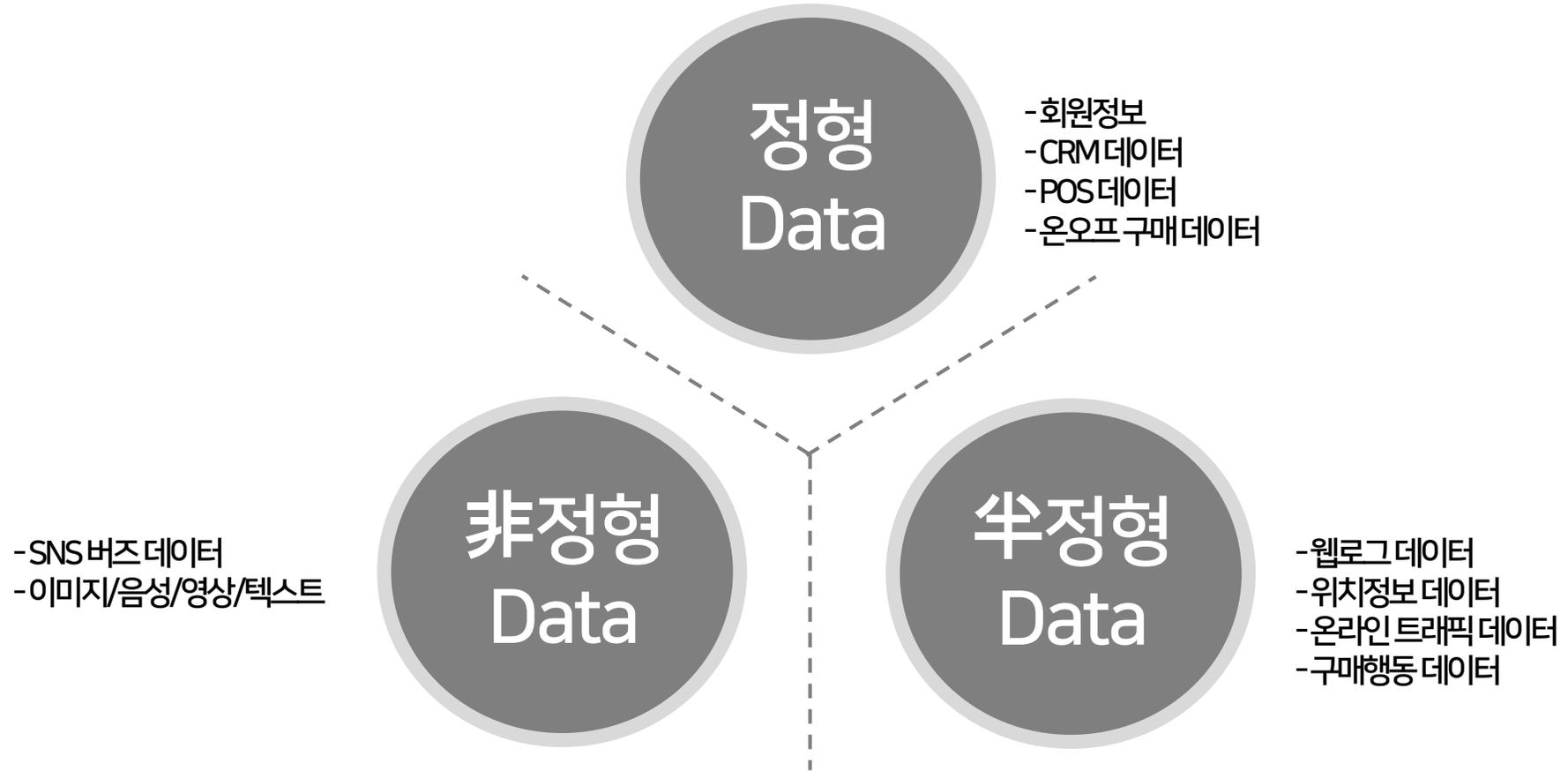


What Is Big Data? | SAS US



What is big data analytics? | IT PRO

[Big Data]



Big Data

디지털 환경에서 생성되는 데이터로
숫자뿐 아니라 문자와 이미지,
영상 데이터를 포함하는 대규모 데이터

데이터의 양 폭증과 함께 종류도 다양해져
사람들의 행동은 물론 위치 정보와 SNS를 통해

생각과 의견까지 분석 및 예측 가능하다

Dilemma

[구체적]

알아들을 수 없거나
기업기밀/ 개인정보

[개념적]

빅데이터로 하나되는
꿈과 희망의 요술나라

으어어

데이터 혁신 기업의 급부상

amazon

DONG, Welcome to Your Amazon.com™ (If you're not DONG YAL SEO, [click here.](#))

Today's Recommendations For You

Here's a daily sample of items recommended for you. Click here to [see all recommendations.](#)

Essex Fash (DVD) My Favorite Year (Paperback) A Season with Veronica (Paperback) Among the Thugs (Paperback)

Action & Adventure All Categories Applied British Chan, Jackie Coaching Comedy Comic Action Cops & Thrills Essays Europe Health, Mind & Body History & Criticism Literature & Fiction Martial Arts Nonfiction Police & Detective Films Politics Project Management Social Sciences Statistics Travel Web, Yuen World

Your Recently Viewed Items

구매 직후
연관 상품 판매 비율은
전체 판매량 중 35%

사용자 평가 DATA
상품간 유사도 DATA

NETFLIX

NETFLIX

Popular on Netflix Dark Movies Romantic

ALPHAS NEW EPISODES THE GREY TALL GIRL

개인화 추천이 전체 화면에
90% 이상 차지
전체 시간 중 75%

컨텐츠 연관도 DATA

REVOLVE

REVOLVE

도움이 필요하세요? | KO \$100 이상 3-5일 무료배송

WOMENS MENS 뷰티 검색 내 찜임

핫 리스트 신상품 의류 원피스 슈즈 악세서리 디자

LOOKS WE LOVE: Trends to Try

From luscious velvet dresses to big, statement sleeves, these trending looks are a definite must try

SHOP LOOKS WE LOVE

잘 팔리는 셔츠의 버튼수까지
분석하여 최적화
연평균성장률 2배 상승

SHOP DRESSES & SNEAKERS SHOP VELVET SHOP FALL NEUTRALS SHOP

사용자 행태 DATA

Data Science는 공학과 마케팅 사이



무게중심이 마케팅쪽으로 옮겨지고 있다.

우리는 점점 사장님과 가까워집니다.



무서운 일이죠. =_ =

사장님은 왜 무서운가?

사장님이 궁금한 것은 단 한가지

“그거 돈이 되냐?”



“Half the money I spend on advertising is wasted; the trouble is I don't know which half.”

John Wanamaker

US department store merchant (1838 - 1922)

표 9-1 업리프트 모델링의 사분면

개입 안 함	개입함	범주	대응 방안
전환하지 않음	전환하지 않음	무관심	비용이 드는 개입은 자제한다.
전환하지 않음	전환함	설득 가능	가능한 한 개입 대상에 포함시킨다.
전환함	전환하지 않음	청개구리	개입 대상에 절대 포함시키지 않는다.
전환함	전환함	잡은 물고기	비용이 드는 개입은 자제한다.

무관심 Do Not Disturbs 혹은 Sleeping Dogs 은 개입 유무와 상관없이 전환하지 않는 부류다. 예를 들어 과거 3년간 구매기록이 없는 고객은 광고 메일을 보내봤자 구매로 이어지지 않을 가능성이 높다. 그러므로 이런 고객은 휴면 고객으로 분류하여 광고 대상에서 제외하는 편이 비용을 아낄 수 있다.

✓ **설득 가능** Persuadables 은 개입이 첫 전환을 이끌어내는 부류다. 업리프트 모델링을 통해 찾아내려는 부류이기도 하다. 예를 들어 '여러 번 방문했지만 다른 곳과 가격을 비교하며 결정을 못 내린' 고객이라면 할인 쿠폰을 발급하여 구매로 이어지도록 할 수 있다.

청개구리^{Lost Causes}는 그냥 내버려 두었으면 전환했을 것이나 개입을 하면 전환하지 않게 되는 부류다. 상점에 들어섰다가도 점원이 접근하면 나가는 손님이나 대출 상환 현황을 안내하는 연락을 취했더니 조기 상환하는 대출 고객 등이 이런 부류에 해당한다.¹ 이런 유형의 고객에게 판촉

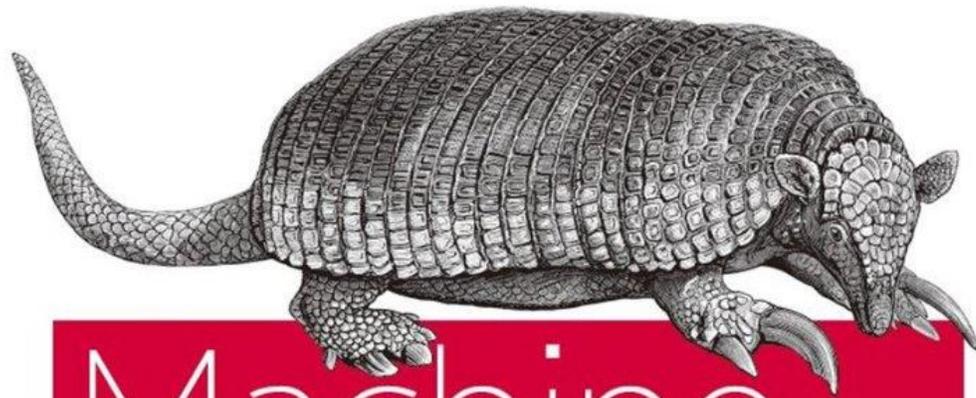
.....
¹ 은행의 주 수입원은 이자이므로 대출금을 조기 상환하면 수익이 줄어든다.

활동을 시도하면 오히려 매출이 줄어들게 된다.

잡은 물고기^{Sure Things}는 개입 유무와 상관없이 전환하는 부류다. 이미 계산대 앞에 선 고객이라면 할인 쿠폰을 제시하지 않아도 어차피 구매할 것이므로 개입이 매출 증가로 이어지지 않는다. 증가는커녕 할인분 만큼 감소하게 될 것이다. 개입하면 거의 다 반응할 것이므로 개입의 효과를 실제보다 높아 보이게 하므로 KPI 수치를 높이고자 개입 대상에 포함시키는 경우가 있으나, 개입에 비용이 드는 경우라면 자제하여야 한다.²

O'REILLY®

실전에 필요한 머신러닝 시스템 설계, 데이터 수집, 효과 검증 노하우

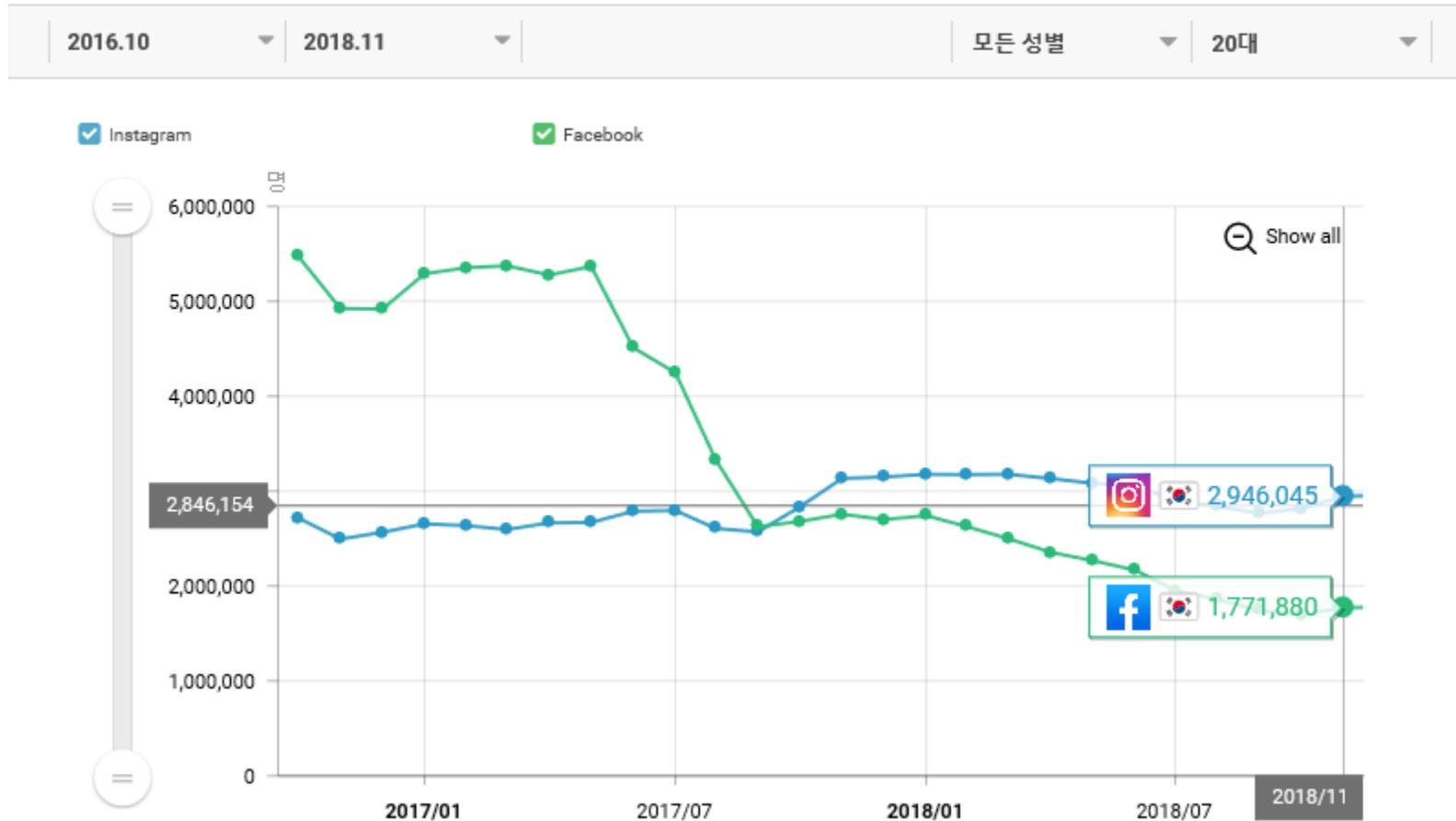


Machine Learning at Work

머신러닝 실무 프로젝트

10/20 유저의 페이스북 이용 감소 -> 페이스북 마케팅 비용 최적화 (돈새는곳 막기)

MAU(월간 활성 사용자 수) 추이



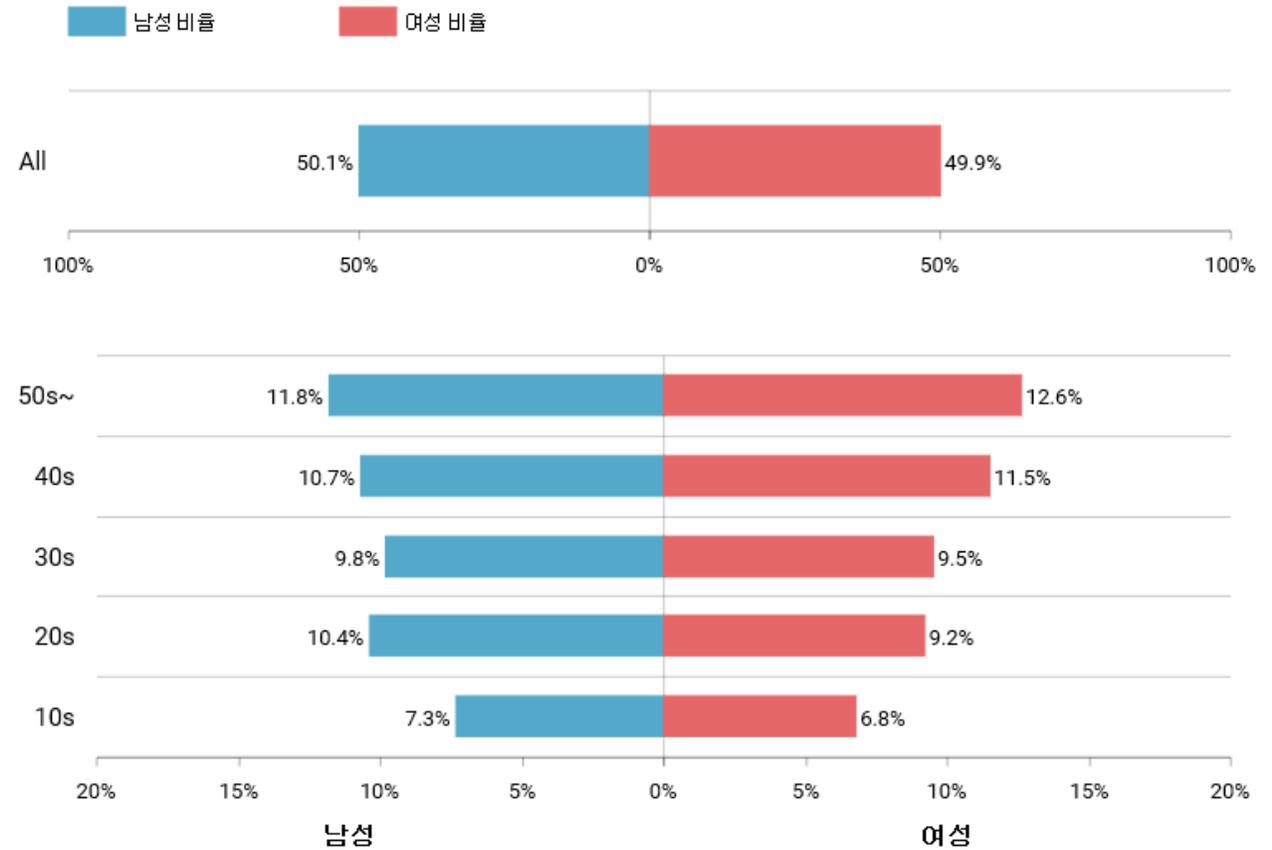
성별 · 연령대 비율

50대 이상 여성 에 인기

성별 · 연령대 비율 ?

2019.06

설치 사용자 기준



*대흥기획 앱에이프

다음앱 뉴스 탭



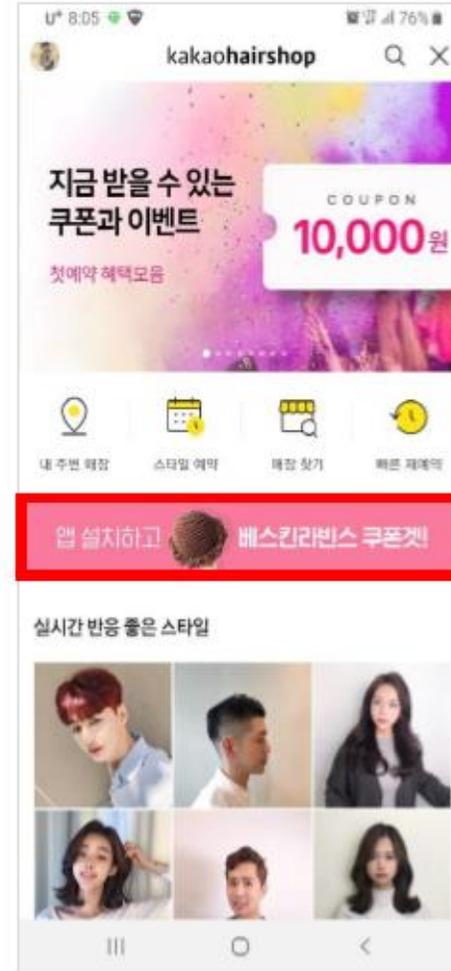
카카오톡 더보기 탭



카카오토리 피드



카카오헤어샵 피드

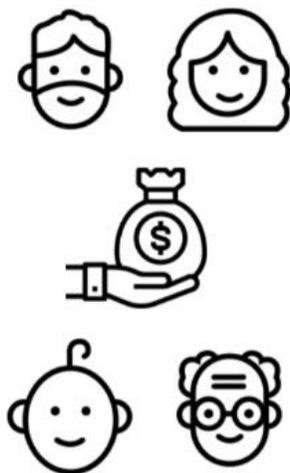


카카오톡 검색 탭



광고 타게팅 진화 양상

유저 프로필만 보고 그룹을 나눴다구요? 검색한 상품들
위주로 그룹을 나눴다구요? PABII는 유저의 평소 스마트폰
사용 내역 전체를 다 이용한 타게팅 광고를 제공합니다

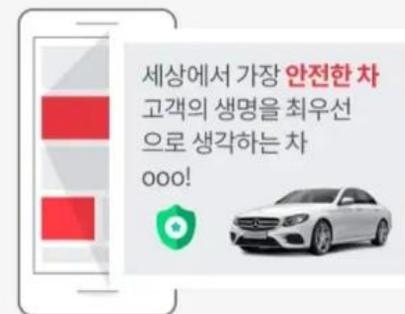


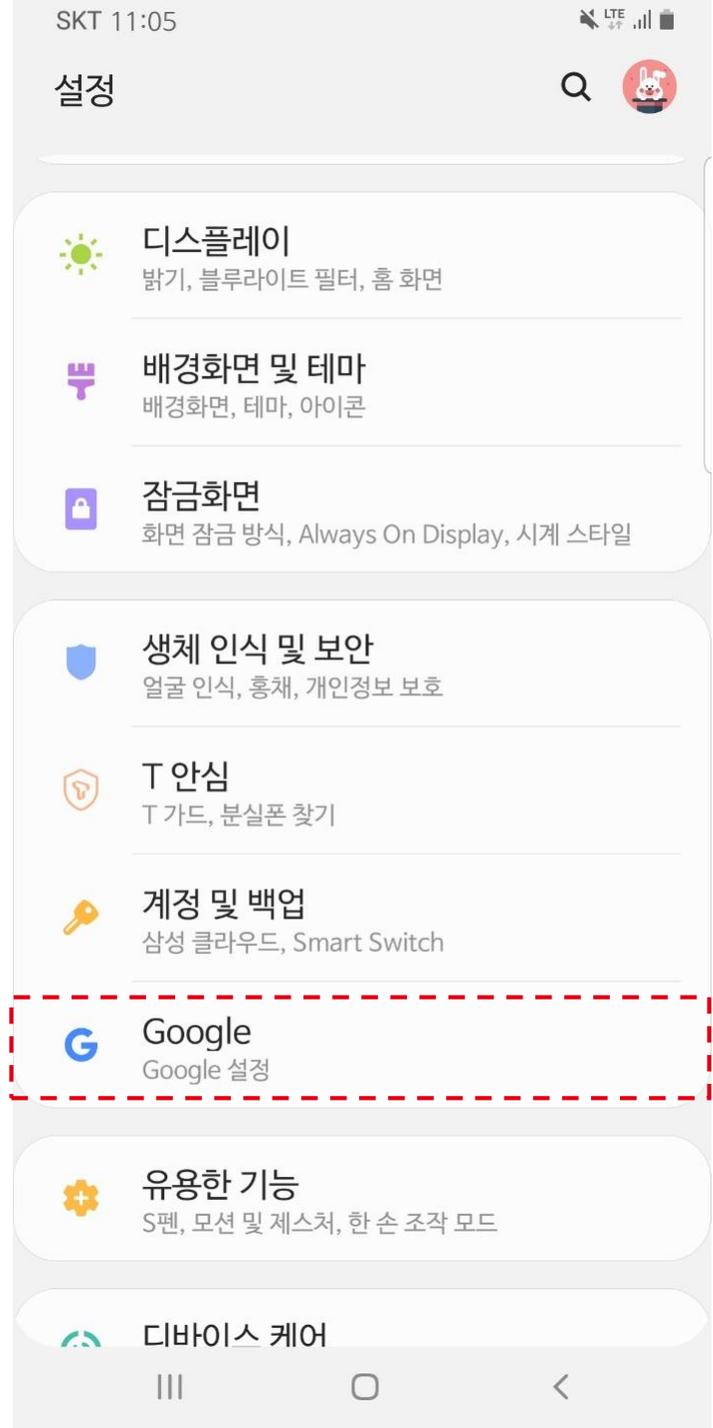
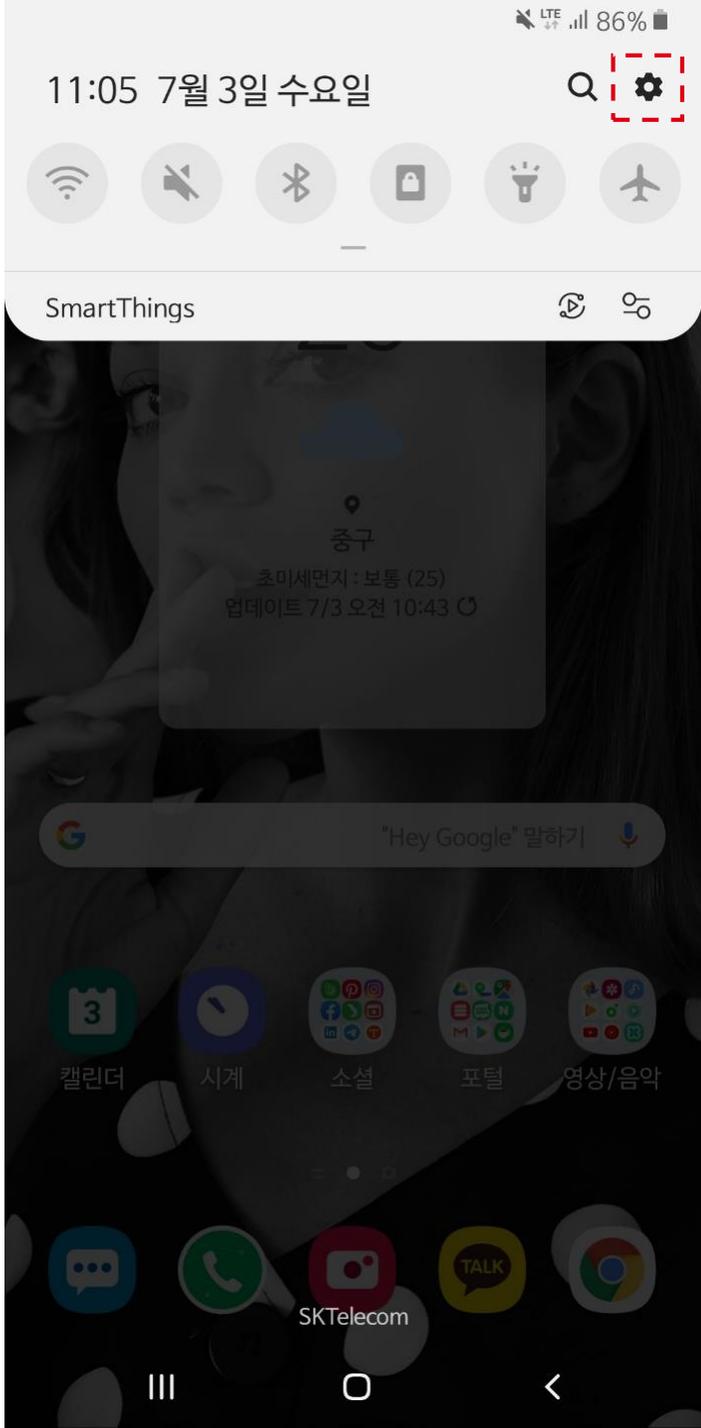
Demographic Targeting

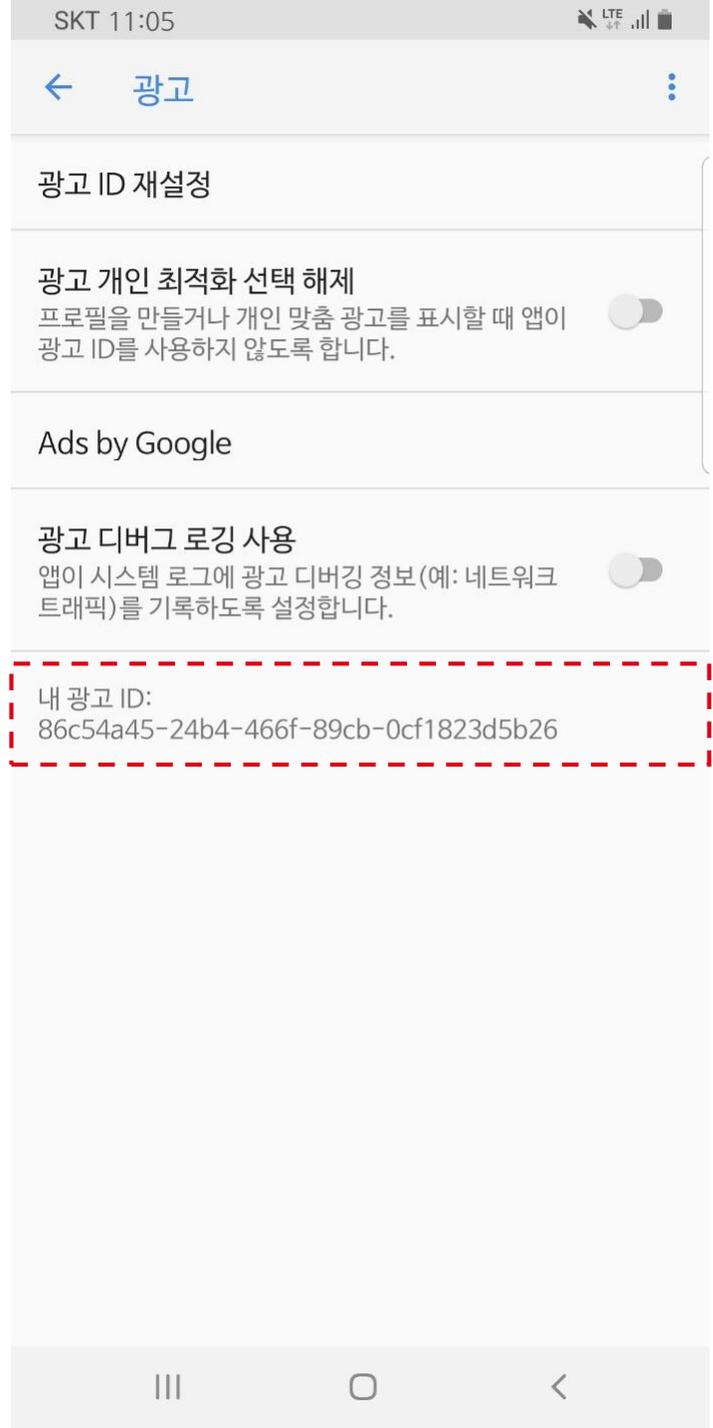
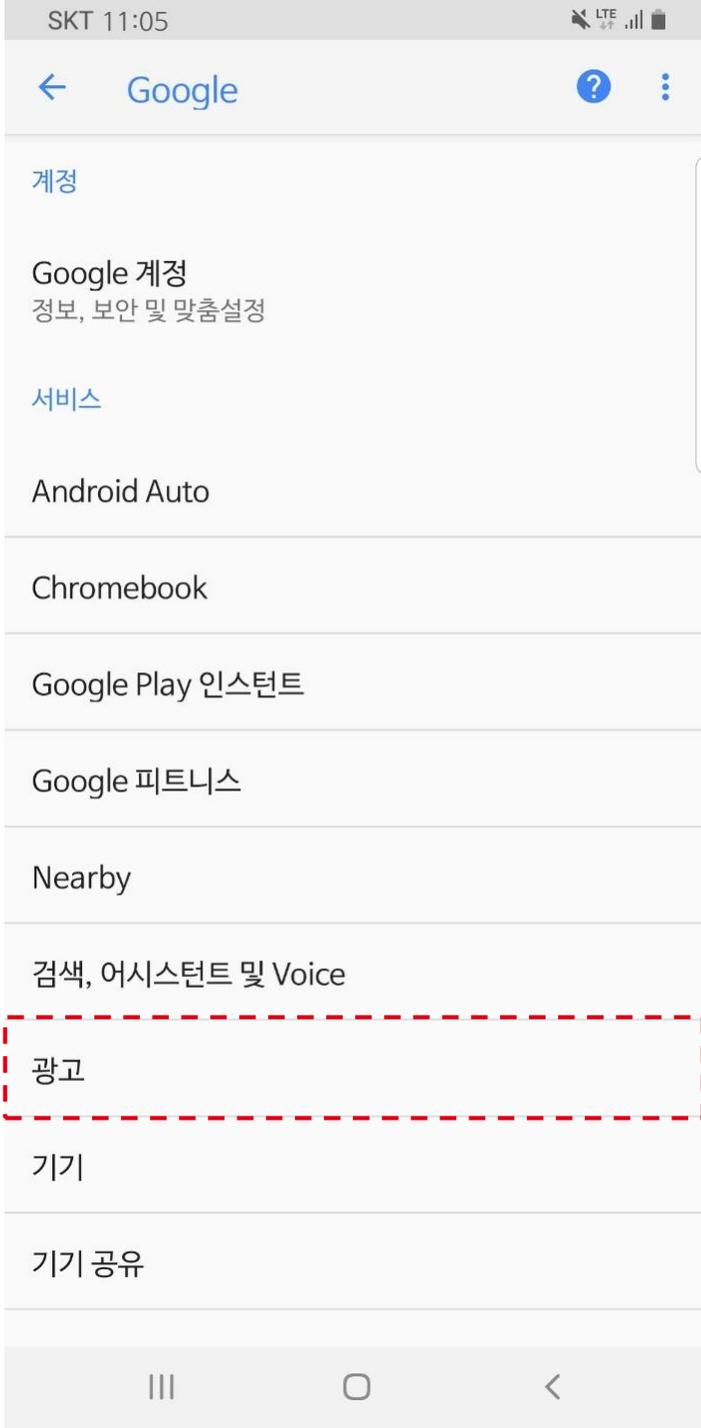
성별, 연령, 소득수준, 결혼유무...

Profile

Message









시시각각 변하는 데이터를 파악하는 자동화의 세계

빅데이터를 지탱하는 기술

시스템은 데이터를 생성하고 데이터를 통해 서로 연결된다

규모의 제약을 넘어서는 시스템 개발을 위해

데이터 처리를 시스템화하는 효율적 방법



Hadoop/시계열 데이터 수집
NoSQL/데이터 통합/실시간 집계
효율적인 전처리/Spark/분산 시스템
파이프라인/데이터 축적/워크플로 관리



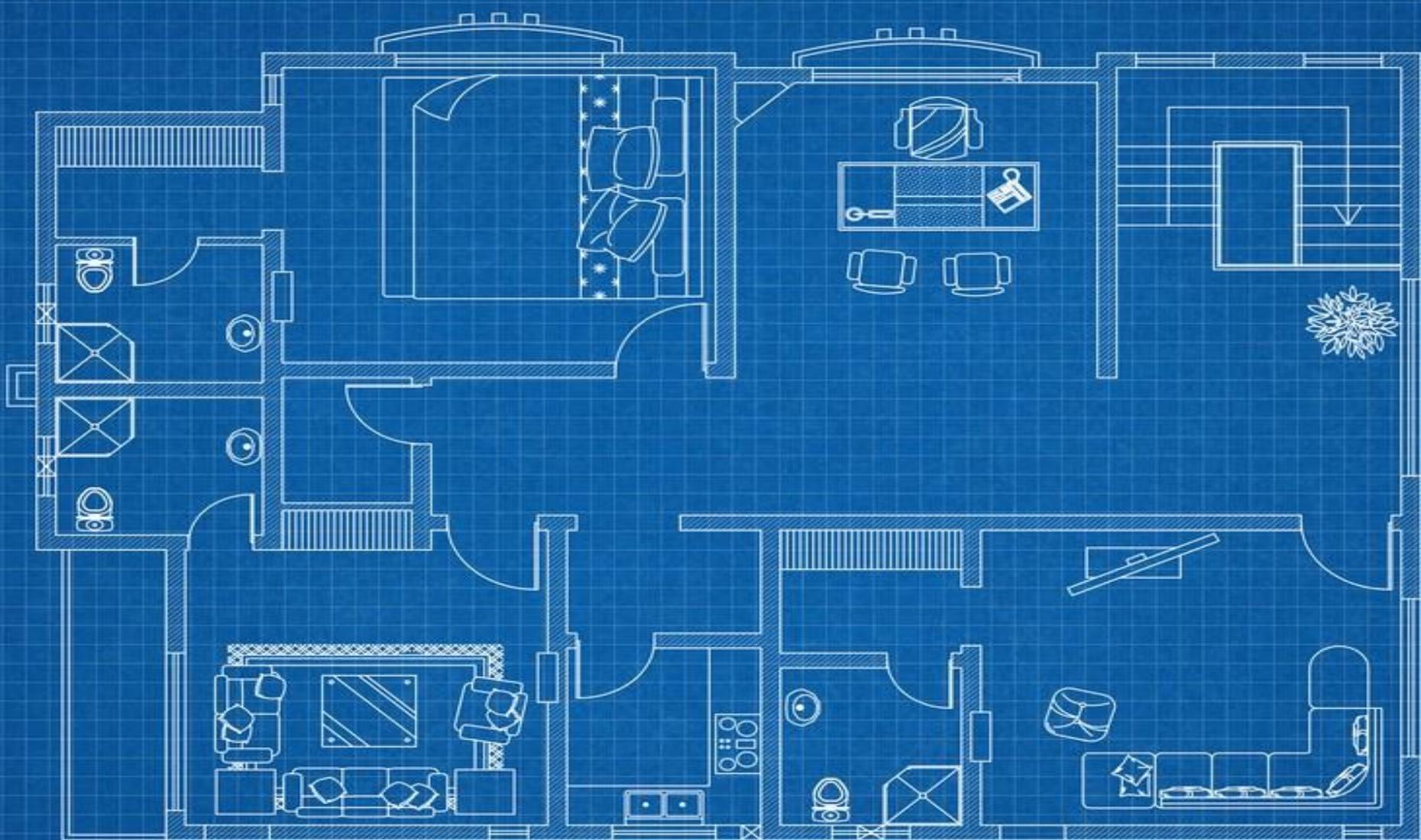
시계열 데이터
데이터

통한

의 지

11558

8640



864

2700

324

702

1242

594

648

756

540

2970

918

1027

701

1350

2160

2484

현업의 로고는 어떤 상황?

없거나

(그런거 없ㅋ영ㅋ)

쓸 수 없거나

(현재 상태 DB만 있는뎡)

왜 없나?

로그를 남기려면 고민이 필요

의식의 흐름

“뭘 남기지? (KPI)”

“어떤 형식으로 남기지?”

“어떻게 전송하지?”

“어디에 저장하지?”

몰라@.@

정하기 어려워

안남기면
해결된다.



뭘 남기지? 예시 : 고객등급 변동

비즈니스 관점에서 고객 등급 관리가 최우선
Micro-clustering을 위해 등급 변동에 대한 파생변수 정밀 제작

upcnt

고객의 등급이 올라간 횟수

down_cnt

고객의 등급이 내려간 횟수

g_mean

고객의 등급의 평균

start_grade

고객의 시작 등급

end_grade

고객의 마지막 등급

grade_sd

고객 등급 변화의 표준편차

e_cnt

연속적인 E 등급의 갯수

왜 쓸 수가 없나?

“우리도 데이터 있어요”

막상 가보면 DB에 현재의 state만 남아있음

예를들어 같은 통장 잔고 100만원도

- 1000만원 벌고 900만원 탕진한 100만원인 사람과
- 110만원 벌고 10만원 아껴써서 100만원인 사람 다름
- **이력이 필요합니다. State가 아니라 history**

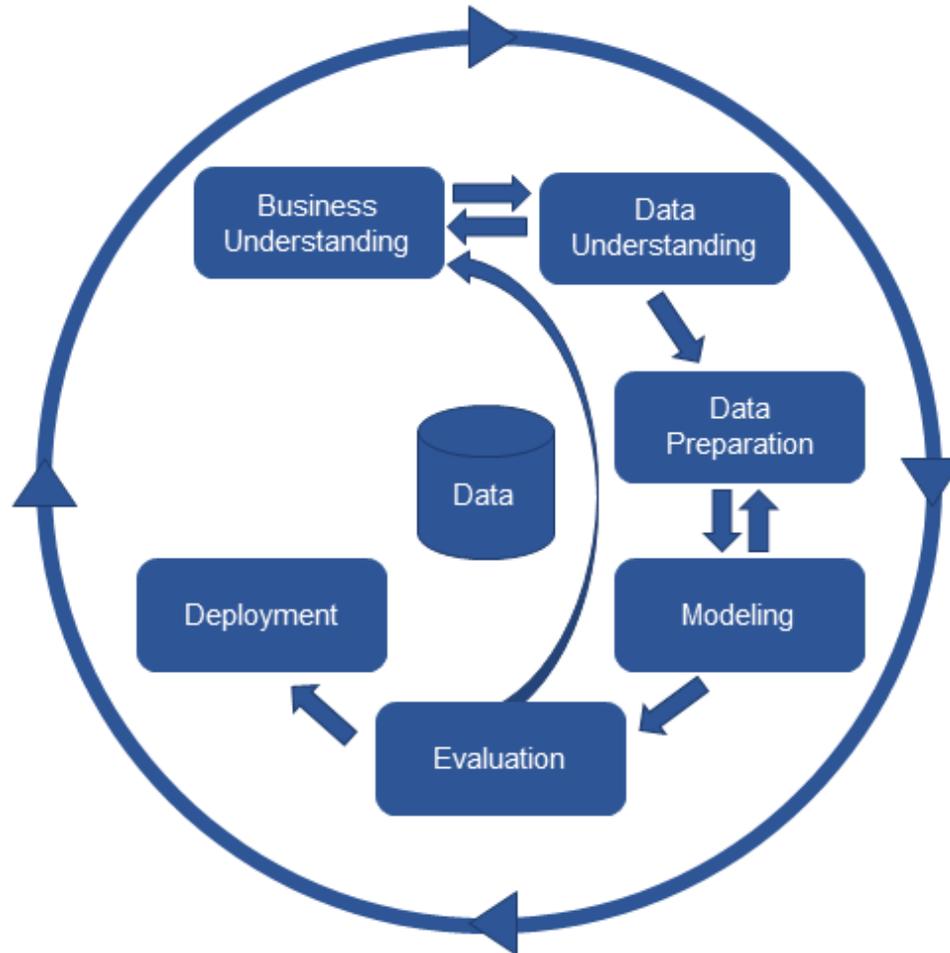
분석했더니 현업이
아는 것들이래요

정상입니다.

분석 == 액션

*CRISP-DM

Agile해야 하는 건 분석이 아니라 액션
“보고”먼저가 아니라 “실행”먼저



데이터 액션 기업의 급부상

amazon

DONG, Welcome to Your Amazon.com™ (If you're not DONG YAL SEO, [click here.](#))

Today's Recommendations For You

Here's a daily sample of items recommended for you. Click here to [see all recommendations.](#)

1 2 3 4 5

Essex Fash (DVD) My Favorite Year (Paperback) A Season with Veronica (Paperback) Among the Thugs (Paperback)

Action & Adventure All Categories Applied British Chan, Jackie Coaching Comedy Comic Action Cops & Thras Essays Europe Health, Mind & Body History & Criticism Literature & Fiction Martial Arts Nonfiction Police & Detective Films Politics Project Management Social Sciences Statistics Travel Web, Yuen World

Your Recently Viewed Items

구매 직후
연관 상품 판매 비율은
전체 판매량 중 35%

사용자 평가 DATA
상품간 유사도 DATA

NETFLIX

NETFLIX

Popular on Netflix Dark Movies Romantic

ALPHAS NEW EPISODES THE GREY TALL GIRL

NEW EPISODES

개인화 추천이 전체 화면에
90% 이상 차지
전체 시간 중 75%

컨텐츠 연관도 DATA

REVOLVE

REVOLVE

도움이 필요하세요? | KO \$100 이상 3-5일 무료배송

WOMENS MENS 뷰티 검색 내 찜임

핫 리스트 신상품 의류 원피스 슈즈 악세서리 디자

LOOKS WE LOVE: Trends to Try

From luscious velvet dresses to big, statement sleeves, these trending looks are a definite must try

SHOP LOOKS WE LOVE

잘 팔리는 셔츠의 버튼수까지
분석하여 최적화
연평균성장률 2배 상승

SHOP DRESSES & SNEAKERS SHOP VELVET SHOP FALL NEUTRALS SHOP

사용자 행태 DATA

데이터 분석 조언

비공개 · CiWY10

2018.04.27.

안녕하세요, 저는 게임업계에서 통계 분석을 하고 있습니다. 주로 rpg류 게임 내에 유저 행동패턴 및 매출 데이터를 분석하여 인사이트 도출을 하고 있습니다.

그런데 참.. 분석 기술이나 다른 것들은 많이 변하고 있는데 저는 정체되는 듯한 느낌만 받습니다.

현재 기술은 SQL과 엑셀로 기술통계에 기반하여 데이터 분석을 진행하고 있고, 이걸로도 지금까지 계속 무난하게 계속 하고 있습니다. (문제는 SQL도 아주 기본적인 수준으로도 데이터를 뽑을 수 있기 때문에... 중급 이상 실력이 안 느는 것 같습니다.)

저는 게임에 대한 열정이 있어서 해당 맡은 게임의 인사이트 도출을 위해 아주 많은 플레이 시간을 보내고 있습니다. 그리하여 어느 테이블에서 어떤 데이터를 봐야되는지가 먼저라 생각했습니다. 예를 들면 콘텐츠의 재미를 느끼는 요소, 유저들이 더 오래 머물게 할만한 요소, 이를 통한 고객 세그멘테이션 등을 연구하여 추후 이러한 기반을 삼아 게임 데이터 분석 후 전략적으로 움직일 수 있는 방안 등이 있습니다.

그렇다보니 정착 발전시켜야 될 분석기술, 통계 지식이 계속 부족해 지는것 같습니다. 왜냐하면 실상 인게임 내에 유저 행동 패턴 도출 등을 분석 가공하여 시각화하는데는 그저 기술통계학만 쓰이기 때문입니다. 그래서 요새은 제 미래를 생각해서라도 이게 잘못된 방향인가 하는 걱정이 앞섭니다.

해당 맡은 게임에 대한 깊은 지식을 통한 데이터 분석보다, 통계 데이터 기술이 더 게임업계에서는 요구하는지 궁금합니다. 해당 업계에 종사하시는 분이라면, 혹은 다른 업계 분들도 괜찮습니다. 주니어라 생각해주시고 많은 조언 부탁드립니다.

더불어 더 발전시킬 수 있는 데이터 분석 기술 관련도 알려주시면 감사하겠습니다.

다시말하면, 결정내용을 지지하지 않는 분석결과는
보고서에서 삭제당합니다.

2018.04.29. · 👍 좋아요 1

넷마블게임즈 · 구로O지털O지

대부분의 기획/분석 부서가 그런 것 같습니다. 쉽게 말해
답정너 짓 하는 이유는 경영진이 가설 검증의 도구로 통계나
데이터를 활용하는게 아니라, 경영진의 잘난 자아를
확인하는 도구로 통계를 쓰기 때문인 것 같습니다. 이걸 뭐
답이 없는 것 같아요.

강 싫으면 짜르던가 라는 마인드로 보고서를 올리고
잡마켓을 항시 주시하고 있어야 ㅎㅎ;

2018.04.30. · 👍 좋아요

비공개 · uKUA21

이걸 보면 대표님이 좋아하시겠냐고 데이터를
수정시켜대는데 어이가 없어요.

2018.04.30. · 👍 좋아요

미래에셋대우 · glory00

정말 공감합니다. 뼈빠지게 자료만들어도 의사결정과
상반된 데이터 자료는 언제든 폐기됩니다.

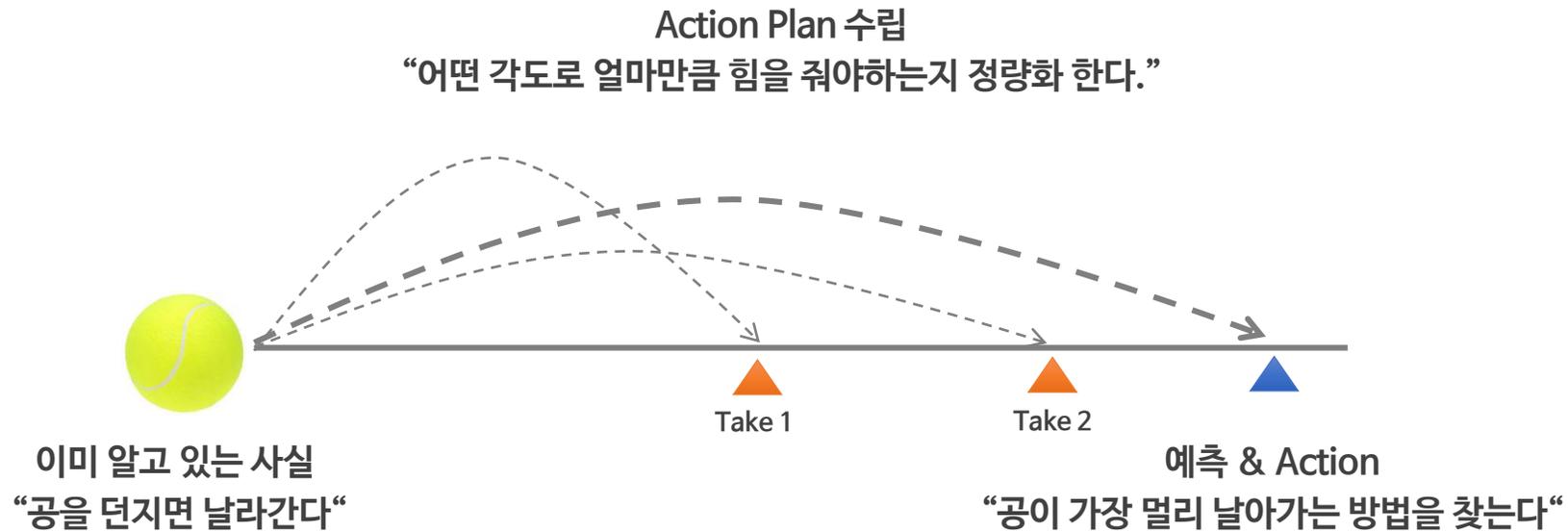
2018.05.01. · 👍 좋아요

실험 (액션) 하자

이미 알고 있는 문제를 왜 분석 해야 하는가?
우리가 찾는 건 “해결방안이지 문제제기가 아니라”고 합니다.

이것은 매우 정상적인 반응 입니다.

분석의 목적은 기적의 창조가 아니고 “행동 (Action) 하기 위함 입니다.”
대부분의 경향은 현업도 알고 있습니다. 하지만 분석의 힘은 정량화에 있습니다.



- 1 데이터는 절대 깨끗하지 않다.
- 2 당신은 분석 대부분 시간을 전처리 단계에서 보내게 될 것이다.
- 3 95% 일에는 딥러닝(Deep Learning)이 필요하지 않다.
- 4 실제 분석의 90%는 일반화 선형모형(GLM)으로 해결된다.
- 5 빅데이터는 단지 도구일 뿐이다.
- 6 당신은 베이지안 접근을 포용해야 한다.
- 7 사용자 입장에서 네가 어떤 방법을 사용했는가는 중요하지 않다.
- 8 학계와 산업계는 서로 다른 세계다.
- 9 프레젠테이션이 핵심이다. 슬라이드쇼(파워포인트) 마스터가 돼라.
- 10 모든 모형(Model)은 틀렸다, 하지만 몇몇은 유용하다.
- 11 완전 자동화된 데이터 사이언스 같은 것은 없다. 인간이 개입해야 할 부분이 있다.

분석 결과를 기반으로 **Small Action**을 실행 내부 UI/UX 점진적 개선, 최적화

예) “주소 인증 프로세스”가 1명의 이탈자를 생성하는데 -0.5를 기여하고
“앱 실행 속도”가 -0.25 만큼 기여한다면 주소 인증 프로세스 개선에 **우선순위**를 두고 대응전략 수립

• 사용자 제고 방안 분석

- KPI(재방문율, 구매빈도) 기준 회귀 분석



소비자 피드백 2: “휴대폰 가입은 인증번호 입력후 회색화면에서 넘어가지 않습니다.”



소비자 피드백 1: “결제는 되지도 않고 가입하려고 해도 페이지는 넘어가지도 않음. 이 기업 IT팀은 대체 뭐하는 건지ㅋ”



소비자 피드백 2: “이렇게 느린 앱은 처음입니다. 다시 한번 이벤트로 햄버거 하나 먹겠다고 ...”

사용율에 미치는 중요도 분석

가입 오류

영향력 평가 (-0.5 높음)

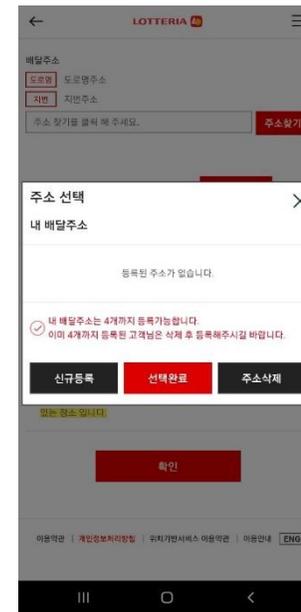
우선순위

결제 오류

영향력 평가 (-0.1 보통)

로딩 속도

영향력 평가 (-0.25 낮음)



UI 최적화

주소 인증 프로세스 개선

결제 프로세스 개선

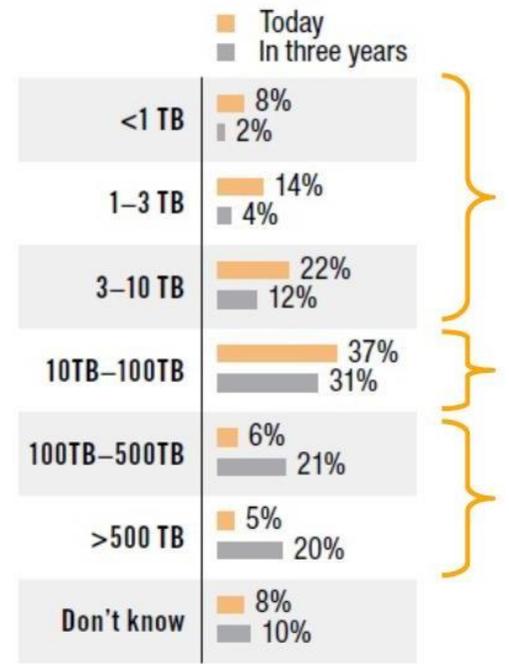
앱 실행 속도 프로세스 개선

빅데이터 분석에 대한 오해와 진실

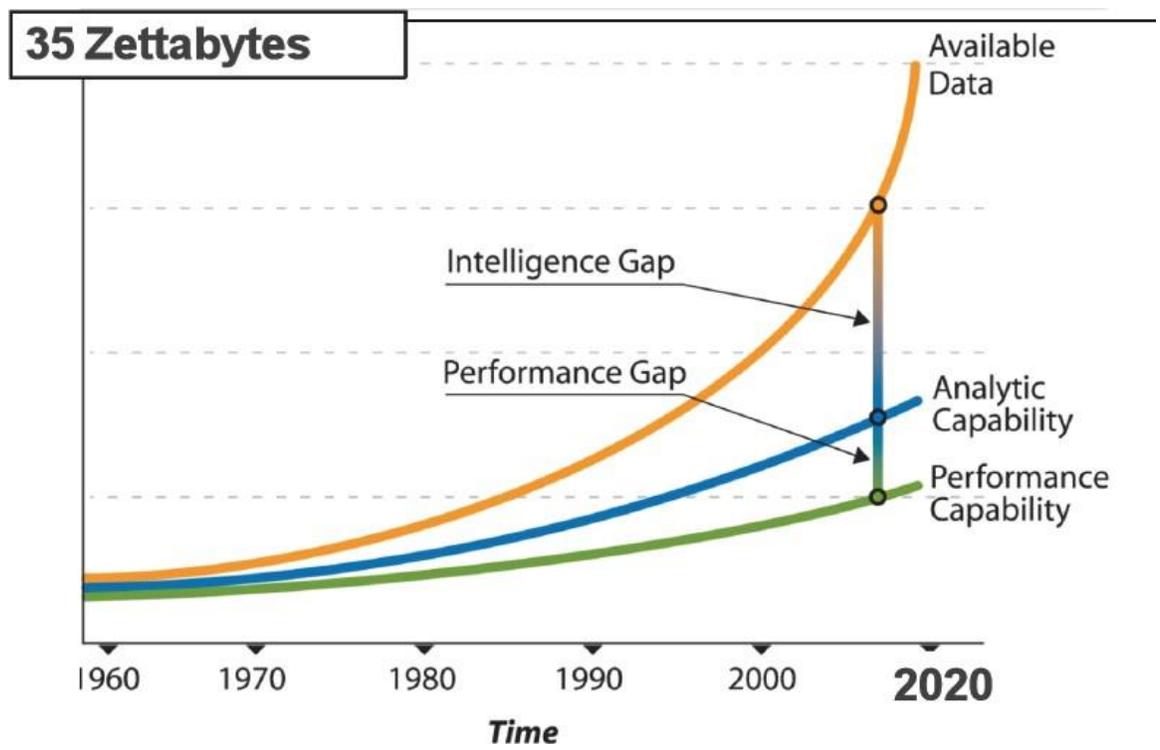


1. 빅데이터 분석에서 중요한 것은 데이터 규모이다

데이터 규모가 빠르게 증가하고 있다
 더 많은 데이터를 분석함으로써 더 큰 효과를 얻을 수 있다



Source TDWI. Based on 325 responses, June 2011



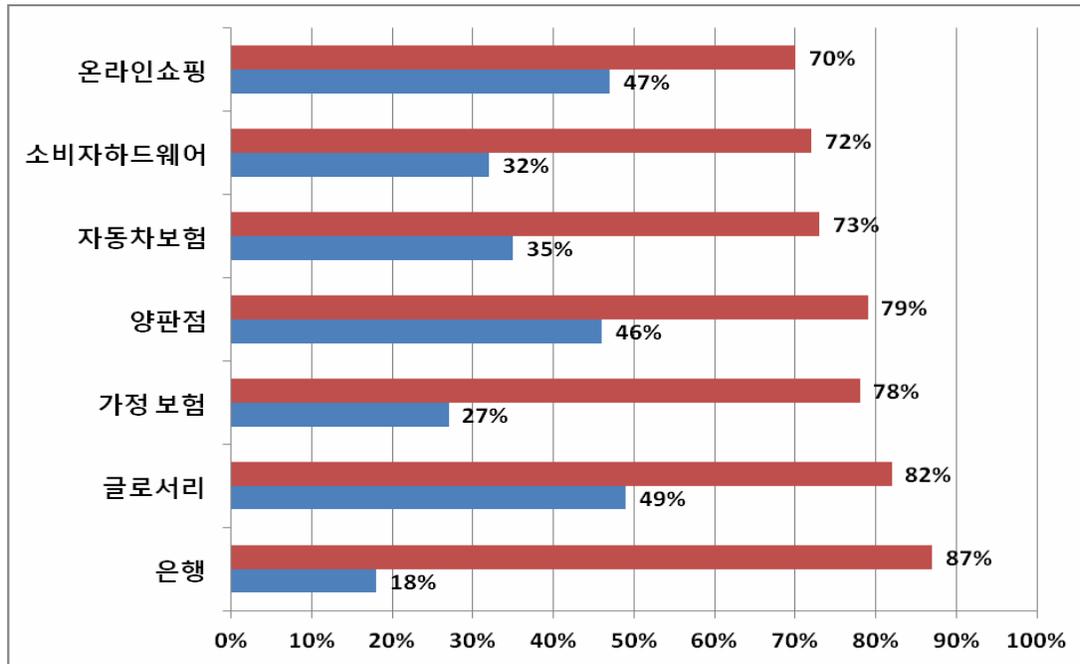
Source:Teradata

1. 데이터의 다양성과 속도가 보다 중요하다

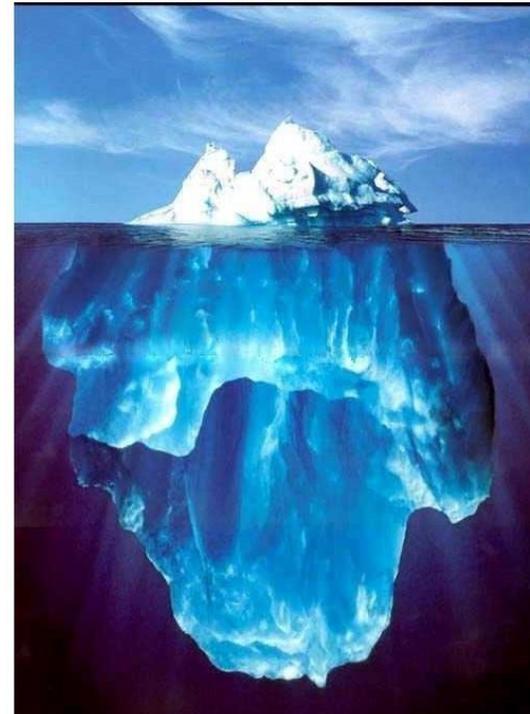


다양한 데이터는 사실을 보다 잘 알게 해준다

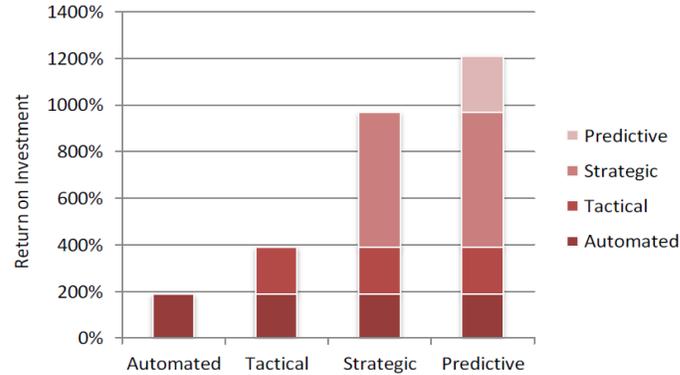
빠른 의사결정과 실행은 경쟁력을 강화한다



Source: 2011 Net Promoter Industry Benchmark



2. 분석 기법을 적용하면 획기적 성과를 거둘 수 있다



Big data can generate significant financial value across sectors

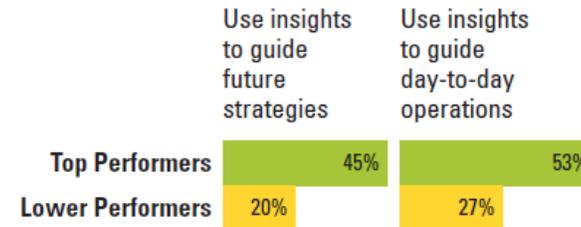


SOURCE: McKinsey Global Institute analysis

THE ANALYTICS HABITS OF TOP PERFORMERS

Top-performing organizations were twice as likely to use analytics to guide day-to-day operations and future strategies as lower performers.

Percent of respondents whose organizations do these activities well or very well.



Source: Nucleus Research m17 - The stages of an Analytic Enterprise, February 2012 Big data—the next frontier for innovation competition and productivity—McKinsey Big Data Analytics and the Path From Insights to value - MIT Sloan

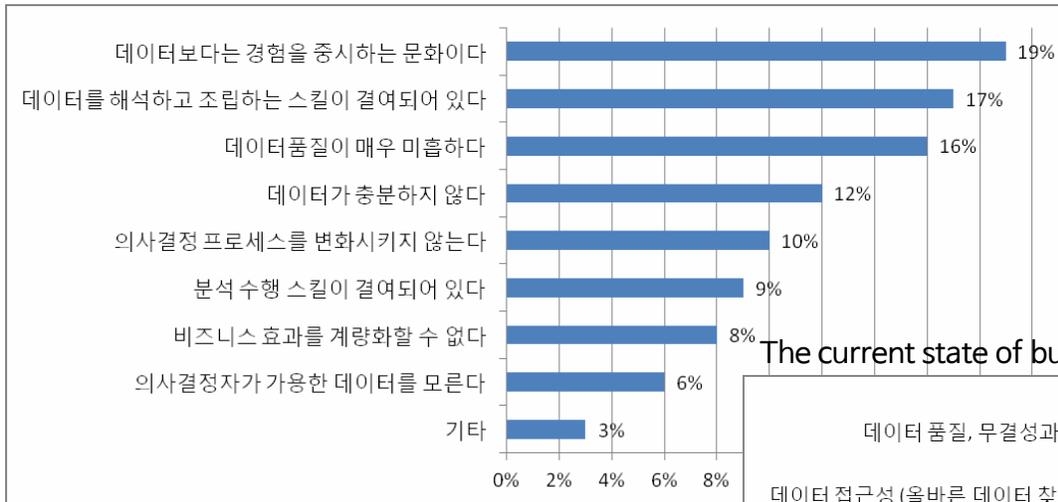
2. 분석 성과는 데이터 활용 성숙도에 비례한다



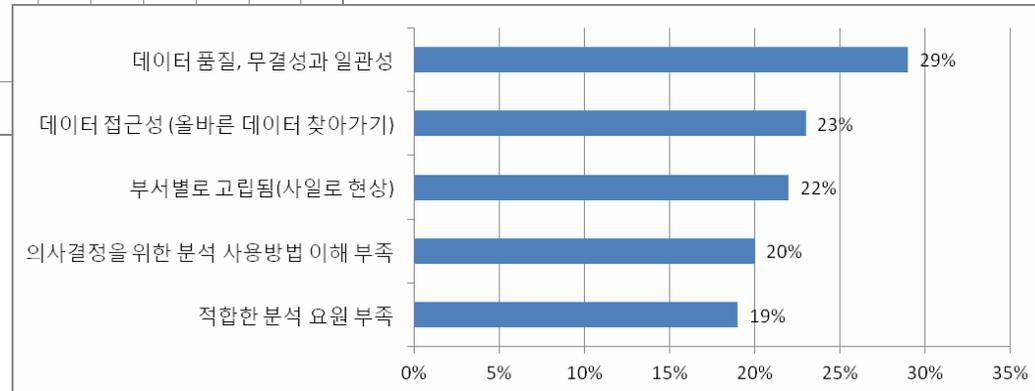
분석 자체만으로 획기적 성과를 얻는 것은 기대하기 어렵다

사실에 근거한 의사결정 문화와 분석이 내재된 프로세스에 의해 성과가 발생한다

A rising role for IT: Mckinsey Global Survey results, McKinsey



The current state of business analytics, Bloomberg Businessweek



Source : McKinsey / Bloomberg

3. 분석 프로젝트는 성공해야 한다

프로젝트는 납기, 품질, 예산이 성공지표이다

분석 프로젝트도 동일한 기준이 적용되어야 한다



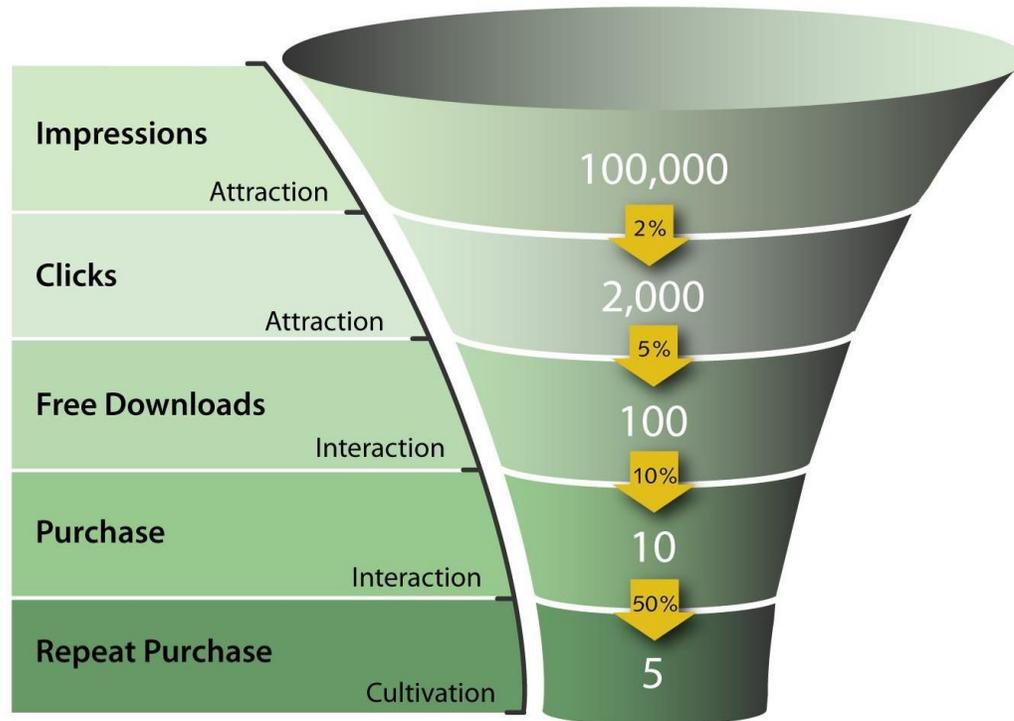
	Gulf	Chasm
주요 극복 과제	전사적 표준 도입 및 적용	BI를 비즈니스 프로세스에 내재화
아키텍처	Data Warehouse, Data Mart	Enterprise Data Warehouse, Analytics
적용범위	본부 또는 부서 단위	전사 통합, 기업 간 통합
분석기법	대쉬보드	시나리오, 내재화된 BI
사용자	지식 작업자, 관리자	경영진, 고객
초점	영감 (insight)	실행 (action)

3. 분석 프로젝트는 실패하는 것이 당연하다



85%의 분석 프로젝트는 원하는 결과를 주지 못한다
그러나, 한번 성공하면 10.5배의 성과를 얻을 수 있다

Conversion

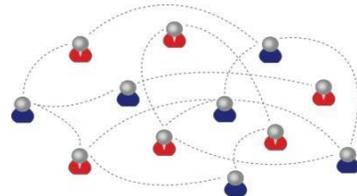


4. 빅데이터가 등장하게 된 것은 소셜미디어 때문이다

소셜미디어가 급속하게 확산됨에 따라서 소셜데이터가 폭증하고 있다
 소셜데이터는 기업 전략과 의사결정에 중요한 분석 기회를 제공한다

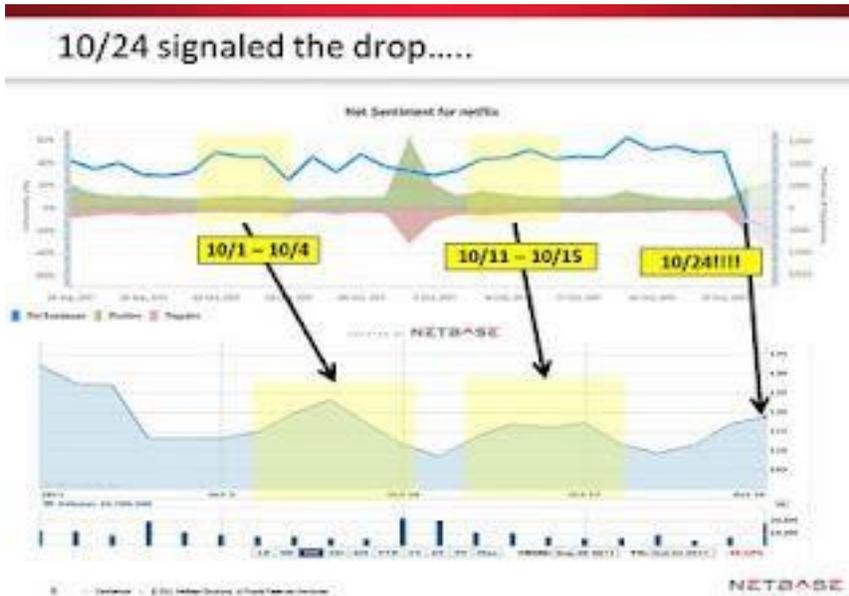
Social CRM

- 기업광고를 믿지 않고
- 기능은 덜 중요하며
- 희소성을 추구하고
- 자랑하고 싶어 한다.



Augmented Humanity

- 최신 정보를 알고 있고
- 가격을 비교할 수 있으며
- 무엇이 문제인가를 알고 있고
- 자신의 불만을 즉시 터트릴 수 있다.



4. 빅데이터 분석의 중요 대상은 기업 내부 데이터이다



디지털화 촉진, 모바일기기 확산, 센서 활용증가 등이 빅데이터 발생 원인이다
내부 데이터가 보다 중요하다, 또한 구조적 데이터는 여전히 중요하다

Which of the following data types are you collecting as big data and/or using with advanced analytics today? Select all that apply.

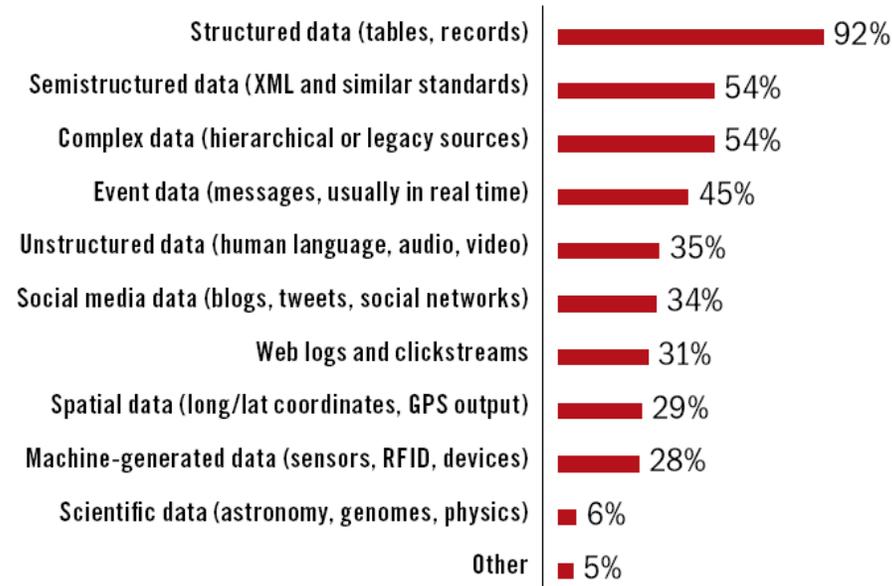


Figure 2. Source: TDWI, June 2011. Based on 450 responses from 109 respondents who report practicing big data analytics.

데이터 분석 기획서 작성시 유의사항

1. 이 분석이 왜 필요한지가 명확히 설명되어야 함
2. 이 분석이 나의 어떤 업무에 어떻게 활용될 것인지가 명확히 설명되어야 함
3. 그래서 이 분석이 나의 업무에 잘 활용되면 분석을 활용하지 않았을 때보다 업무 수행 시 어떤 성과를 기대할 수 있는지 설명이 되어야 함
4. 분석에 필요한 데이터가 충분히 고려되어야 함
5. 그 데이터가 가용한 데이터인지, 획득이 어려운 데이터인지도 명확히 파악
이 되어야 함

감사합니다
