

# 빅데이터 이노베이션

조성준

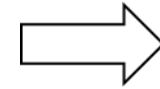
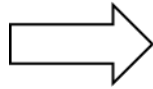
서울대학교 데이터마이닝센터  
공공데이터전략위원회

1. 데이터 > 인사이드 > 밸류

# 빅데이터

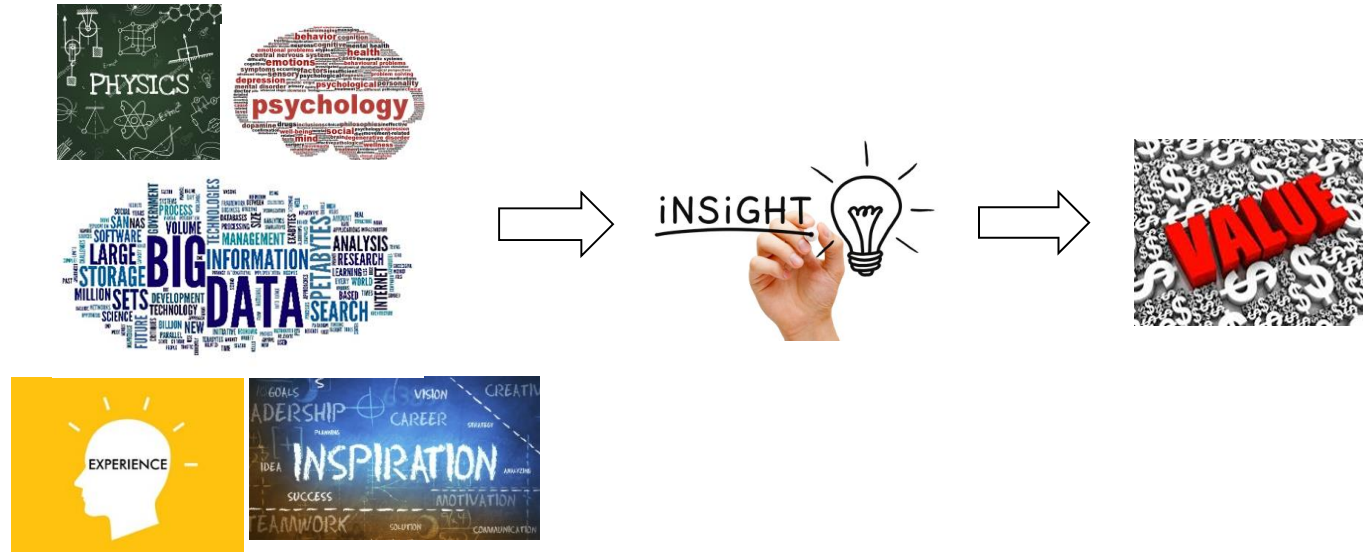
- **Volume** 대용량
- **Velocity** 생성 속도 high
- **Variety** 숫자, 문서, 이미지, 동영상, 스피치
  
- 누구의 관점?

# 재료 요리 만족





# Data Insight Value



# 1.1 신 사업 개발

# Data Insight Value



항공 엔진 운항 데이터 → 고장 가능 확률 → 최적 보전 서비스



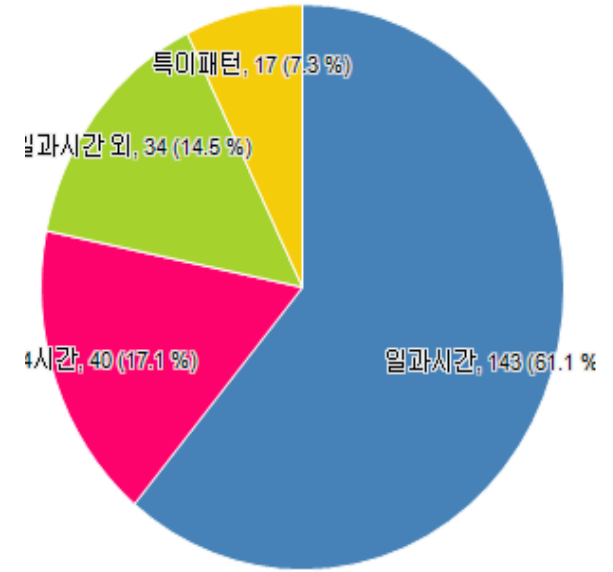
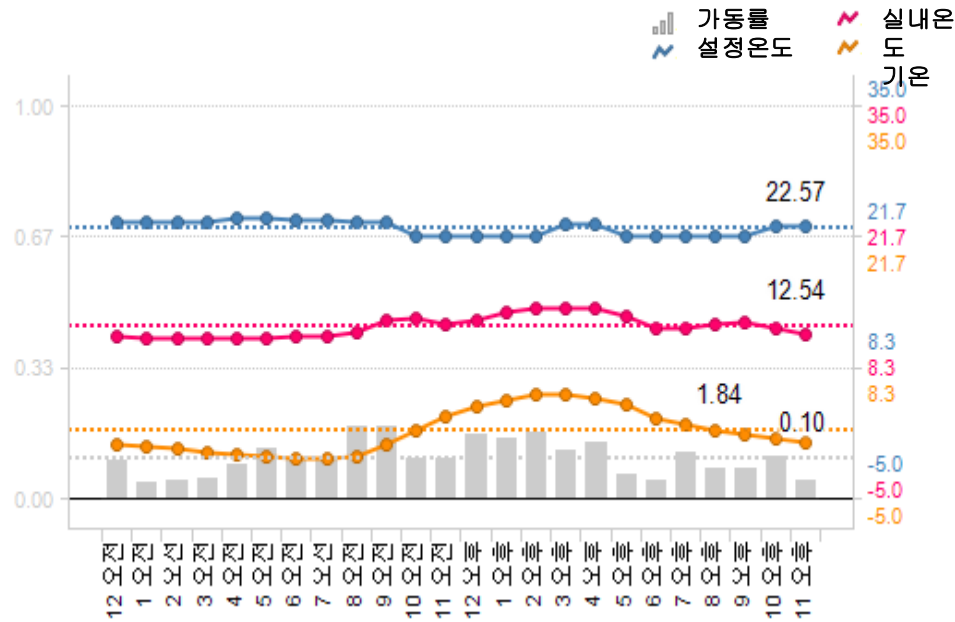
# Data Insight Value



에어컨 실내기 작동 → 에너지 소비 패턴 → 서비스 판매

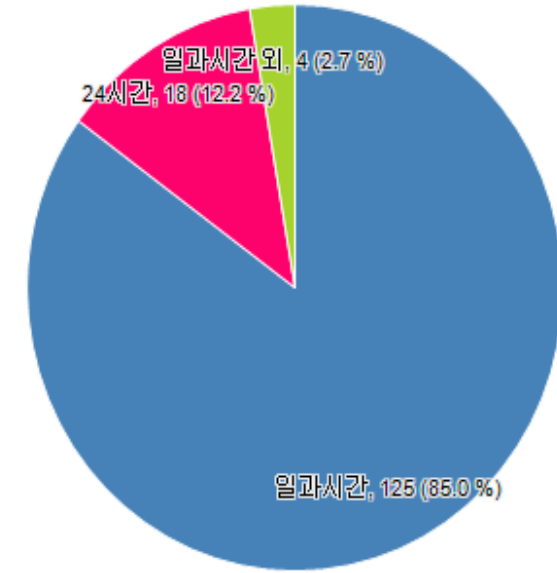
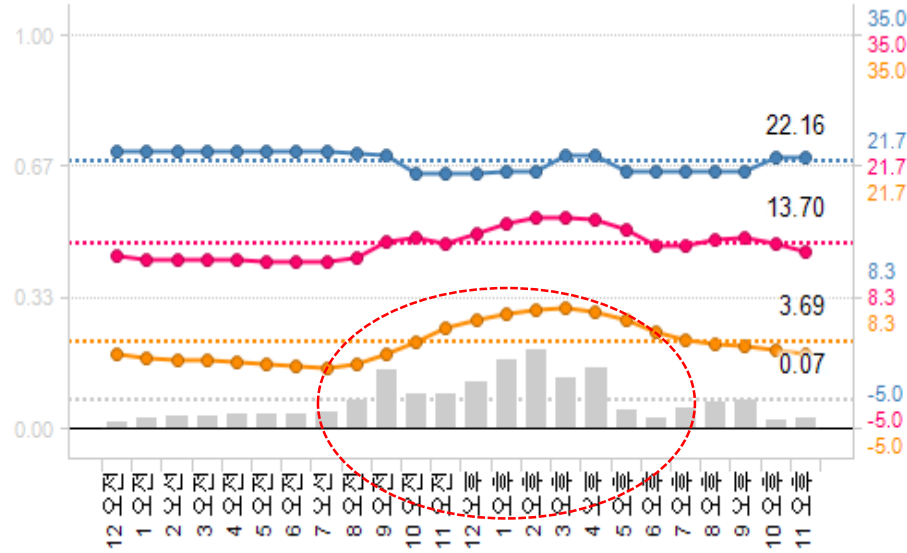
(12월~2월)

University (Avg)



부경대 대연캠퍼스의 시간 별 사용

Pukyung U





## 1.2 NPD 소비자 원하는 기능?



... 엄마가 오븐활용을 못하는거 같은데 별로라고 팔아버린대...



...쿠키 진짜 오븐에 굽는데 수분조절실패에 크기 조절 실패해서 오븐 판에 12개가 ...

... 레시피따라서 했더니... 주방이....난장판에 올리브오일냄새에... 오븐은 타버리고..

... 집에 오븐이 있는데...ㅋㅋ...할 줄 아는 게 없으니까 그냥 뭐 데워먹을 때밖에 쓸일이 없어..!(한심

### SNS를 통한 니즈 분석:

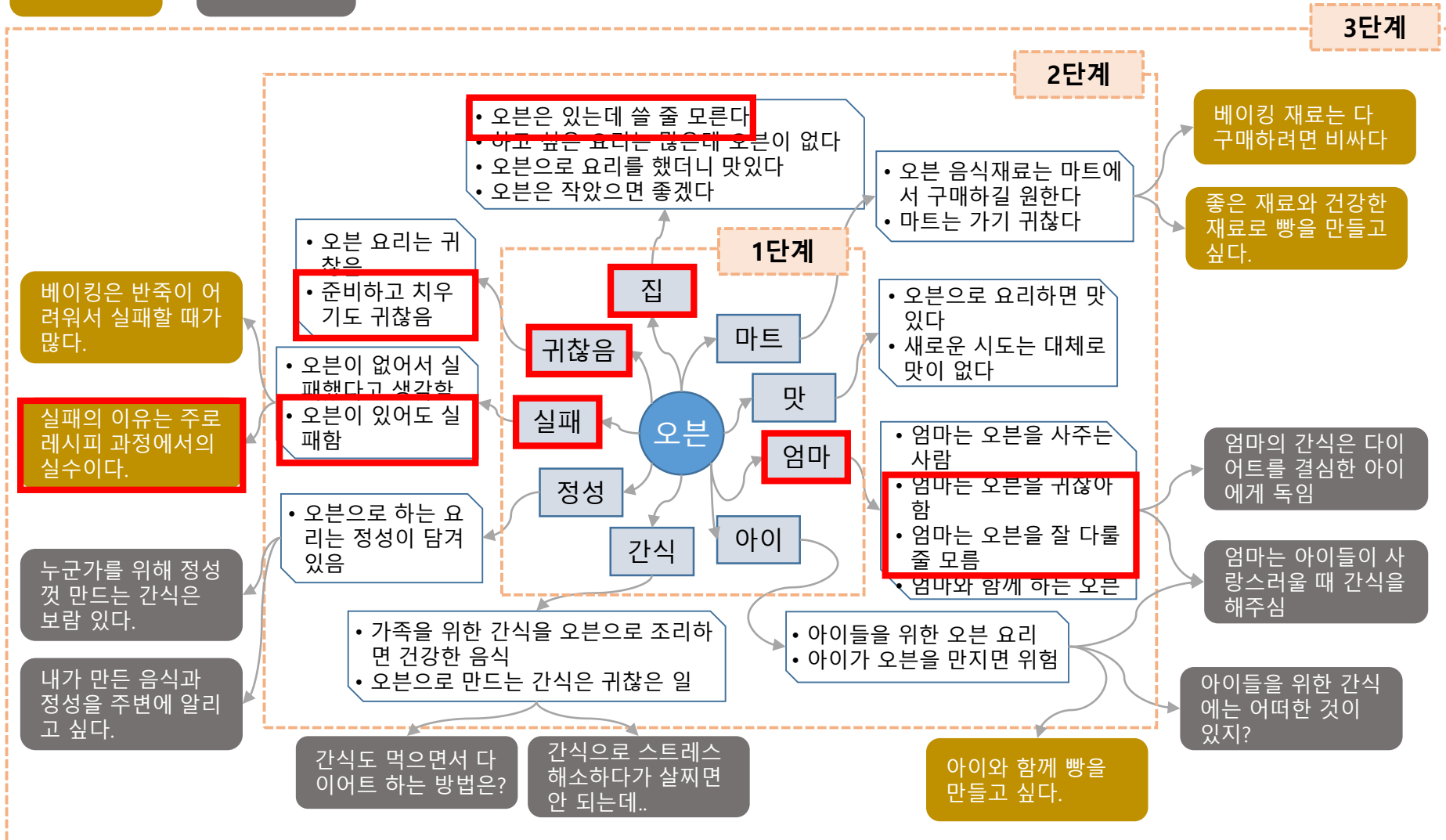
- 오븐을 이용하여 레시피에 따라 성공적인 조리를 하고 싶음.
- 내가 가지고 있는 오븐을 레시피에서 지시하는 대로 작동해도 요리가 제대로 되는지 궁금함

# 3 단계: 키워드 + 키워드 조합

베이킹

간식

“기계 맞춤형 레시피”



## 1.3 수요 예측

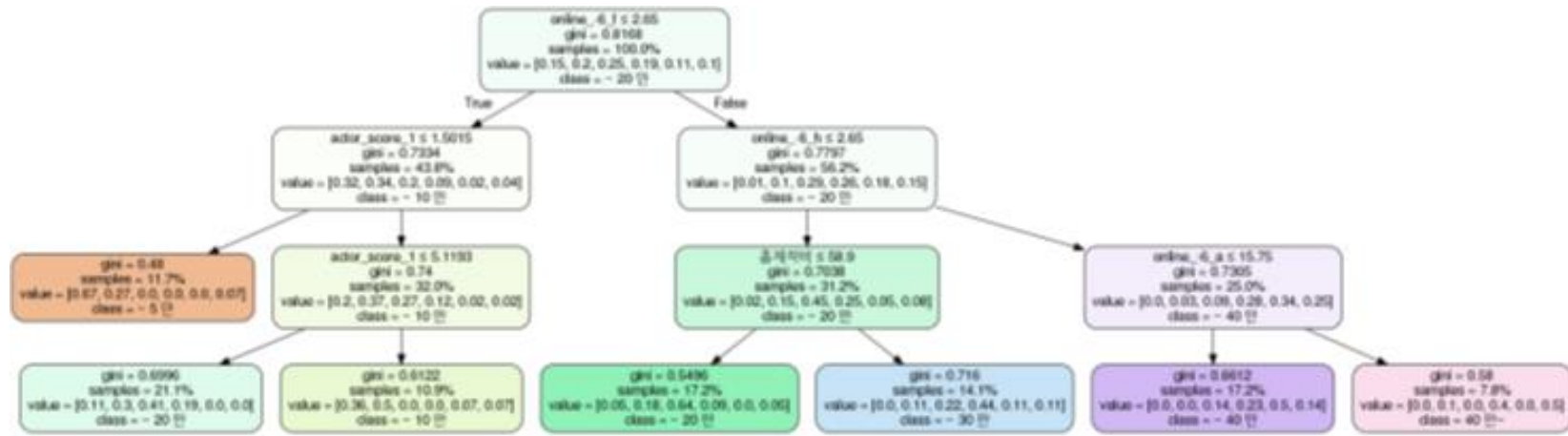


# Data Insight Value



과거 흥행 데이터 → 첫 주말 관객 수 → BEP 관객 수 달성

# 영화 관객 수 예측



	DecisionTreeClassifier	RandomForestClassifier	GradientBoostingClassifier
제작비_감독배우파워	0.7297297297	0.7567567568	0.6756756757
제작비	0.7297297297	0.6756756757	0.6756756757
설문조사_6	0.7297297297	0.6486486486	0.7027027027
설문조사_5	0.5945945946	0.6216216216	0.6756756757
설문조사_4	0.6486486486	0.5675675676	0.6216216216
설문조사_3	0.6216216216	0.6216216216	0.6486486486
설문조사_2	0.5675675676	0.6486486486	0.7027027027
설문조사_1	0.6216216216	0.6756756757	0.7027027027

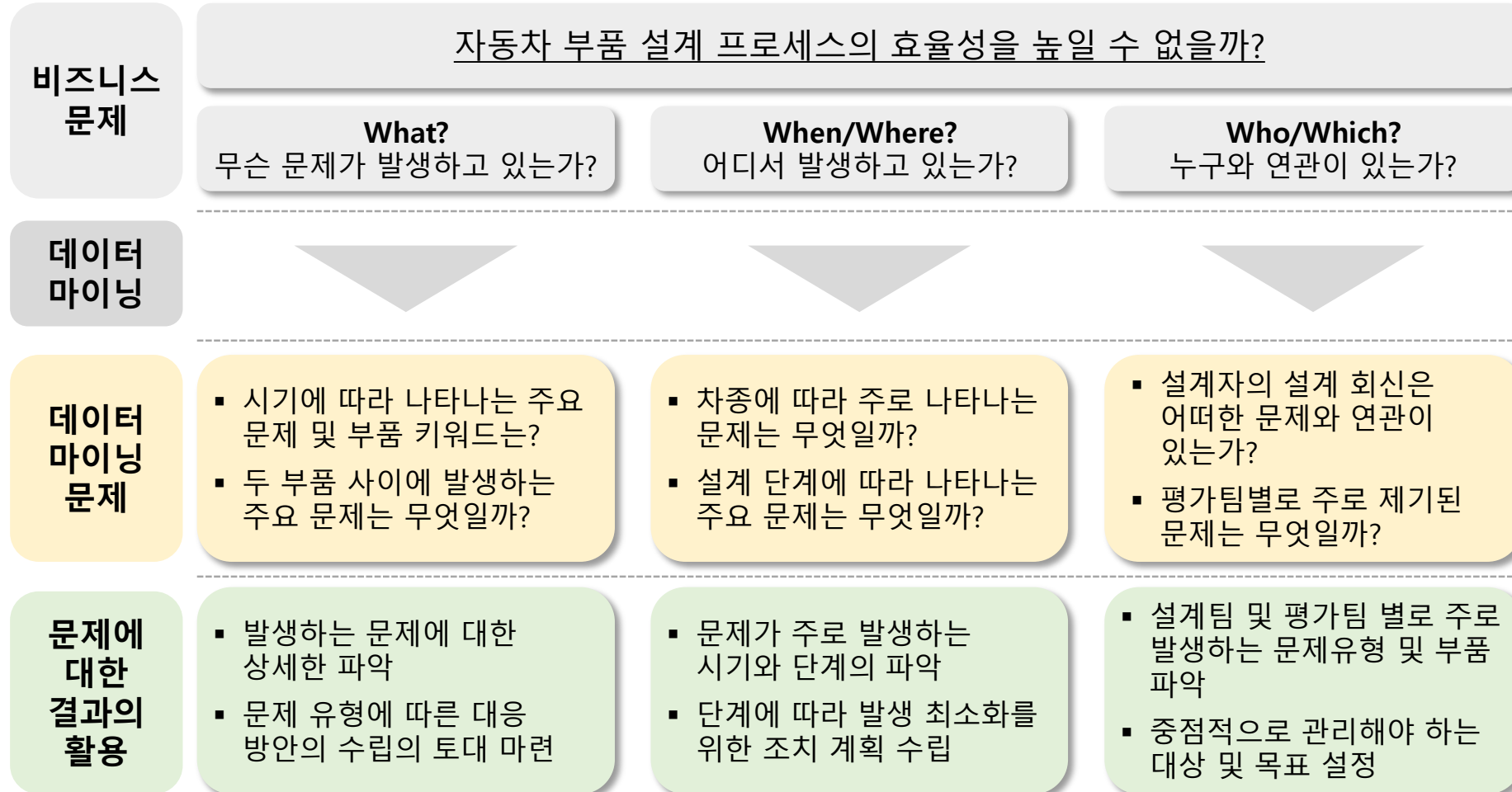
# 1.4 Operational Innovation



# 설계검증시스템



# Questions



# Insight

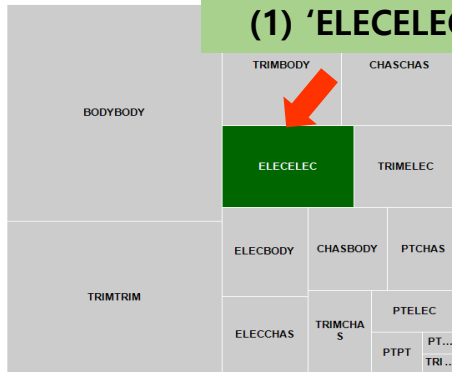
탐색

전처리

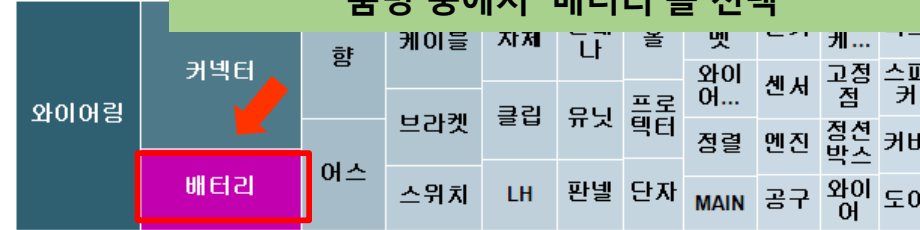
TM

## Approach

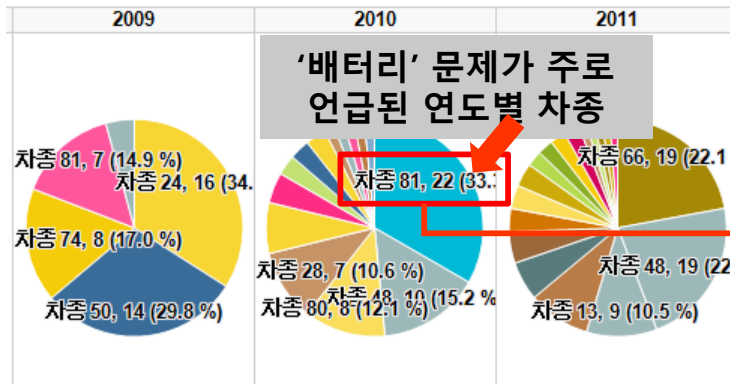
### (1) 'ELECELEC' 을 선택



### (2) 전자 부품계열의 시스템에서 많이 나타난 부품명 중에서 '배터리'를 선택



## Results



### 2010년 '차종81'에서 '배터리'와 관련된 문제점/원인 22건

문제점/원인	평가자
배터리와 와이어 프로텍터간 간섭	한동균
배터리 트레이와 배터리 패드간 간섭으로 체결 안됨	구자근
배터리와 와이어 프로텍터간 간섭(기재기)	구자근
배터리 와이어와 트레이 마운팅부 간섭	구자근
배터리 와이어링 커넥터 하단부에 판넬에 고정되는 클립 없음	한동균
프론트 와이어링 J/BOX항 분기와 배터리간 공간 협소해 간섭 및 와이어 손상 발생 / J/...	구자근

# Data Insight Value



Transporter log → 선박 블록 물류 이해 → 생산 기간 단축



# 선박 조립

- 200여 개의 블록 생산하여 도크에서 조립함
- 6,000 건의 블록 이동 발생

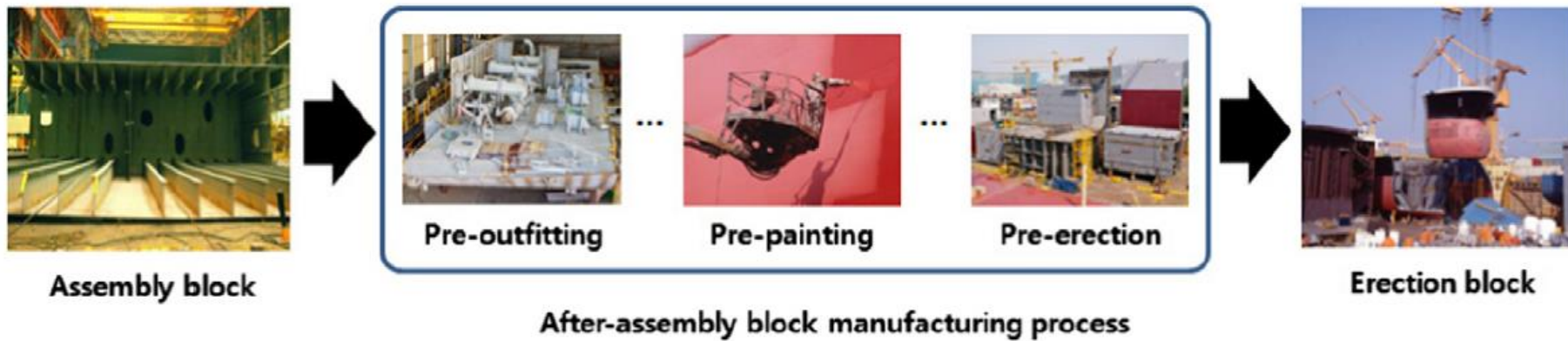


Fig. 1. After-assembly block manufacturing process.

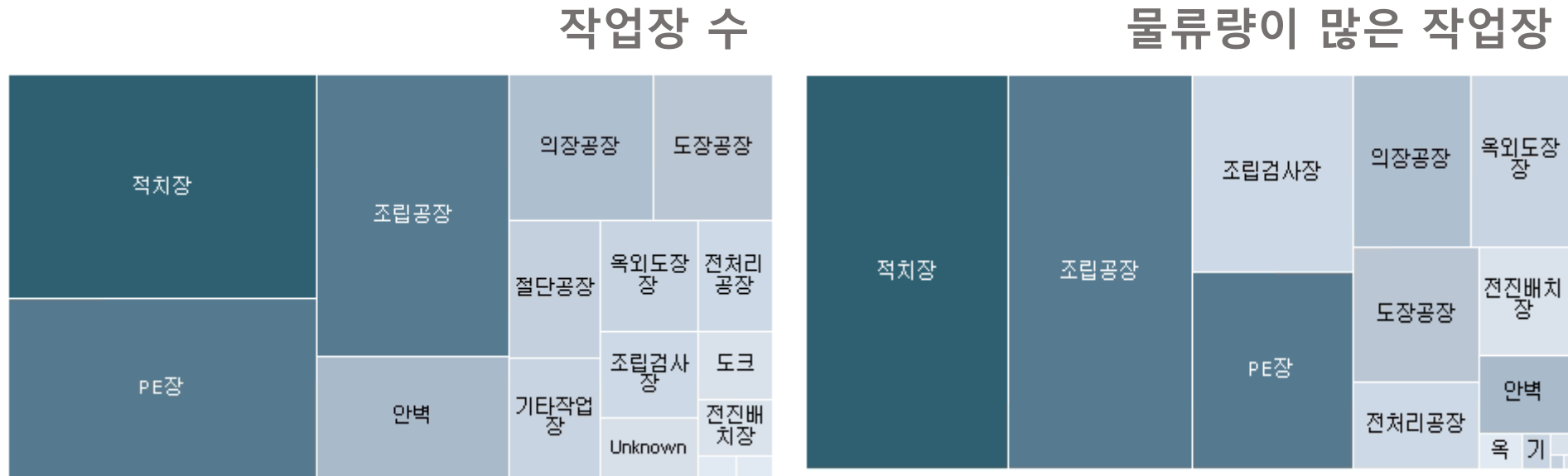
# 블록 이동 패턴 이해

- 블록 이동은 언제 어떻게?
- 생산 납기에 영향을 주는 병목 프로세스는?





# 블록이 어떻게 움직이고 있나? - 어디로 많이 움직이는가

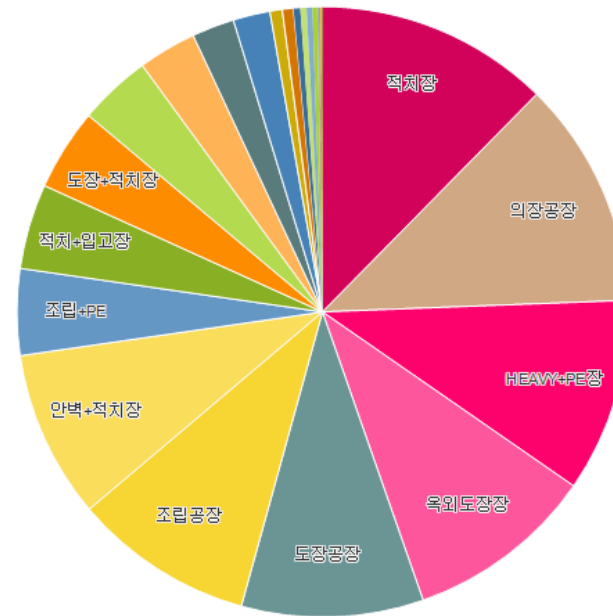
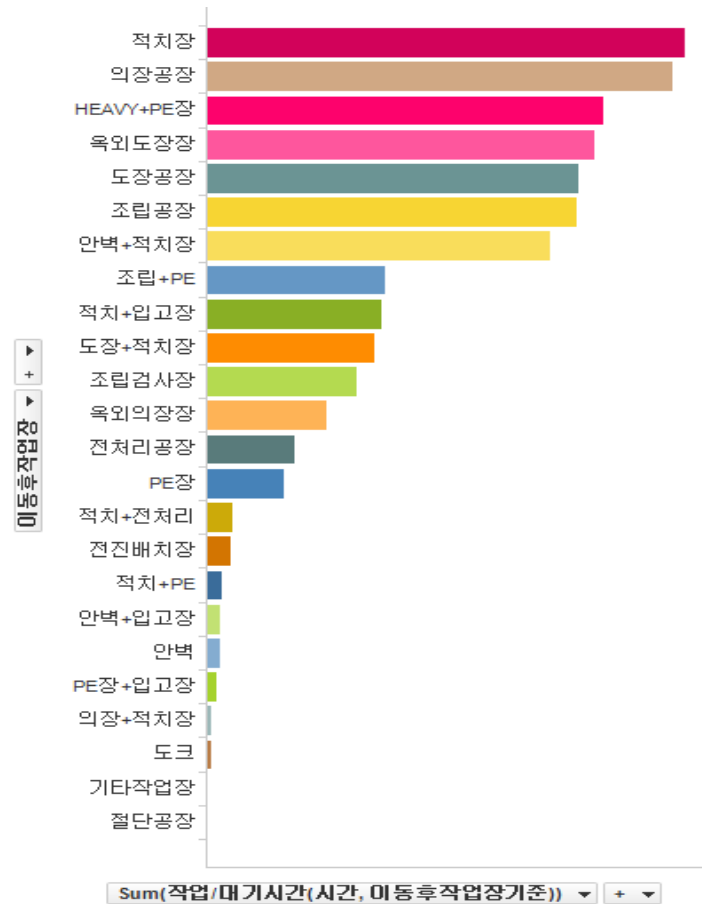


- PE장은 많이 배치되어 있지만 물류량은 적음
- 조립검사장은 적게 배치되어 있지만 물류량은 많음
- 조립공장, 적치장은 많이 배치되어 있고, 물류량도 많음

**적치장 ! → 얼마나 오래 머무는가?**

## 분석 결과: 최상위 계층에서의 표준 생산 프로세스 모델

- 표준 생산 프로세스 모델을 시각적으로 확인할 수 있을 뿐만 아니라, 작업/대기 시간 등의 performance 지표를 검토해 볼 수 있음



블록 생산 프로세스는 의장공정에서 가장 많은 시간이 소요되며, 블록이 적치되어 있는 시간도 상당함

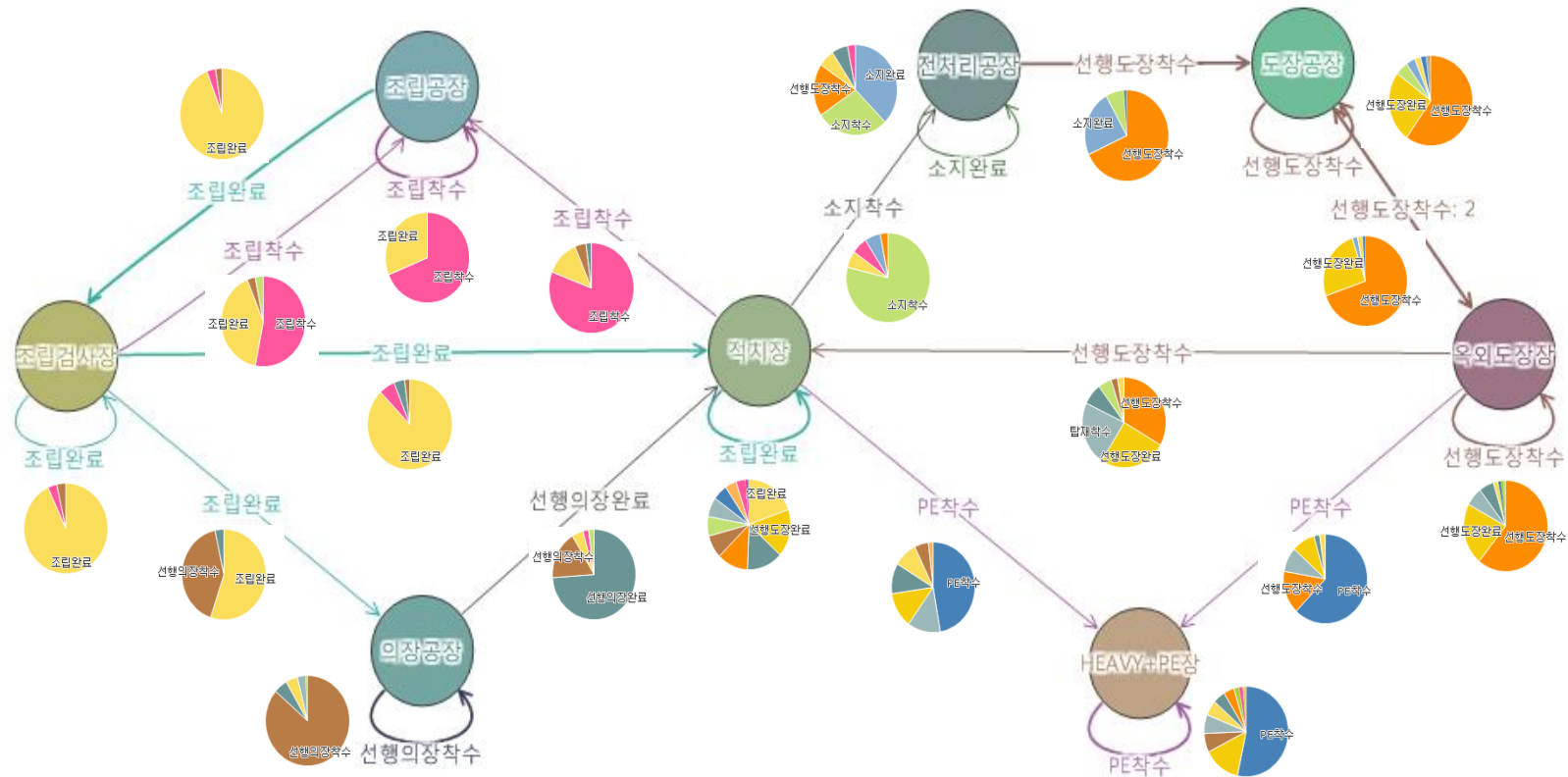


어느 작업에서 병목 현상이 발생하는가?

## 분석 결과: 최상위 계층에서의 표준 프로세스 모델

- 최소 선행, 후행 공정의 Support 값 = 0.5, 최소 선/후행 공정 간 Confidence 값 = 0.2
- 블록의 상태 값에는 입력 과정의 노이즈나, 입력 정책에 따른 혼선이 있음

예) 조립검사장 → 의장공장: 선행의장착수일 수도 있고, 조립완료 일수도 있음



# Data Insight Value



생산 환경 데이터 → 미래 품질 예측 → 생산성 향상 / 투자 기회





ORTS  
ATOUR  
OMER DE CHATEAU LATOUR  
2010  
ILLAC

CHATEAU MARGAUX  
2006  
GRAND VIN  
PREMIER GRAND CRU CLASSE

CHATEAU MARGAUX  
2009  
GRAND VIN  
PREMIER GRAND CRU CLASSE

MIS EN B  
PAVIL  
CHATEA  
APPELLA  
E.C.A. CHATEAU MARGAUX

WINERY DIRECT  
Forts De Latour  
2010  
**\$349.97**  
750ml  
Wine Spectator  
Margaux, Bordeaux, France- "A solid,  
berry, tart, Paulliac, with a  
to the kirsch, blackberry,  
and steeped fig notes,  
ed with anise and tar. Shows  
y through the finish, with a  
note echoing Best from  
igh 2015."

WINERY DIRECT  
Ch Margaux Margaux 2006  
**\$499.99** | **\$449.99\***  
Single Bottle 750ml | MIX 6  
95 pts  
Wine Spectator  
Margaux, Bordeaux, France- "This has a  
wonderful nose of perfume, sandalwood,  
ripe plum and berries. Incredible. Full-  
bodied, with seamless tannins and a long  
finish. Gorgeous."  
Concentrated, Cassia, Violet, Full-  
Bod

WINERY DIRECT  
Ch Margaux Margaux  
2009  
**\$1499.97**  
98 pts 750ml  
Wine Enthusiast  
Margaux, Bordeaux, France- "A massive  
Margaux, packed with tannins  
and ripe fruit... giving intense black  
currant flavors with enticing acidity  
balanced by the sweetness of the fruit.  
Ripe swathes of this opulent fruit are  
also elegant and structured."

# 보르도 와인



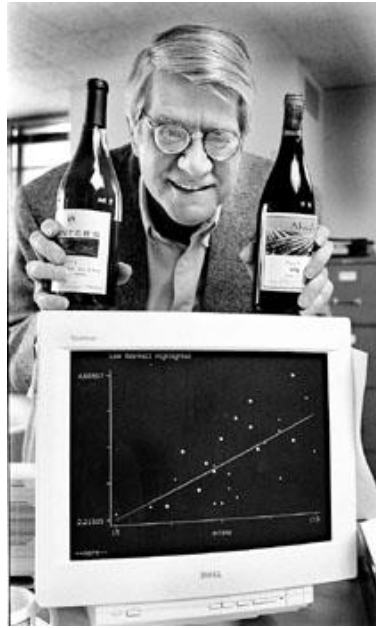
6개월~18개월



6개월



# 와인의 품질은?



온도, 햇빛, 강수량...

# 와인의 품질 예측 가능?

X: 1952~1980 사이의 30년간의 보르도 지방 기후  
데이터 (월별온도, 햇빛, 강수량)

VS

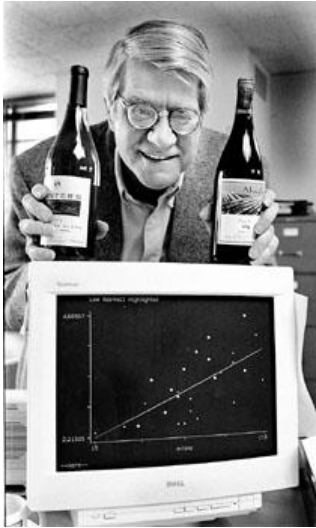
Y: 당해 보르도 와인 평균 가격 (품질)



혹시,  $Y = f(X)$  인  $f$  를 찾을 수 있을까?

찾을 수 있다면,  $x$  로  $y$  예측 가능!!!

# 와인의 품질 예측 가능!!!



$$\begin{aligned} \text{품질} &= 12.145 \\ &+ 0.00117 * \text{전년도 겨울 강수량} \\ &+ 0.06140 * \text{당해년도 평균기온} \\ &- 0.00386 * \text{수확기 강수량} \end{aligned}$$

# 1986 vintage???



# Data Insight Value

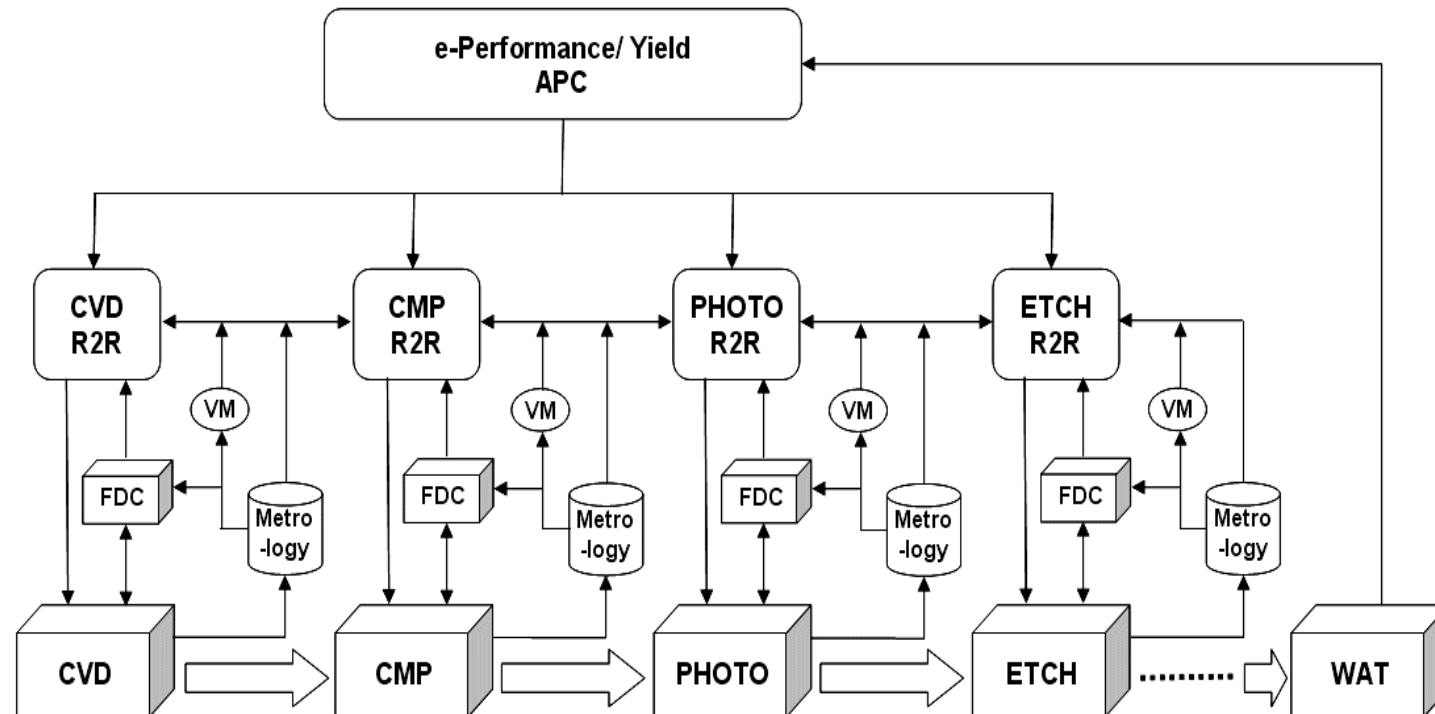


반도체 공정 센서 → 품질 추정치 → 품질 및 생산성 향상



# 반도체 제조 공정

- 수 많은 개별 공정 스텝으로 구성
- 설비의 상태는 웨이퍼의 품질에 큰 영향을 미치며 이를 관리하기 위하여 계측을 수행
- 샘플링 기반의 계측 시스템은 많은 한계점을 내포



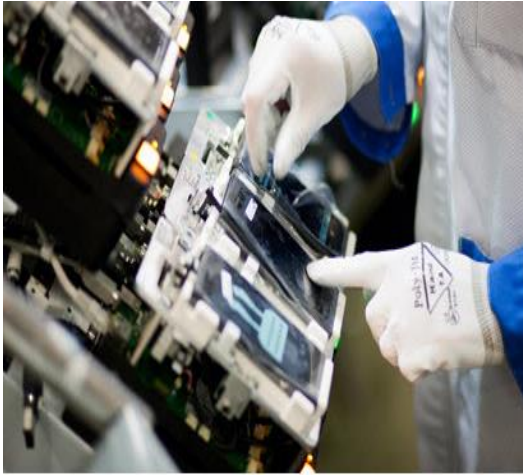


# Data Insight Value



검사 + AS → 제조 불량과 검사 결과 관계 → 품질 및 생산성 향상

## 제조 (제조 테스트 데이터)



### P/N으로 관리

검사 항목	수행시간	검사	검사 항목	측정치	상한	하한	통과여부
2013-02-09 오전	1:20:31	FINAL	RF FINAL START				P
2013-02-09 오전	1:20:31	FINAL	Read Headinfo				P
2013-02-09 오전	1:20:37	FINAL	OB3 MID FER	0	5	0	P
2013-02-09 오전	1:20:40	FINAL	OB3 MID RSRP1	-77	-70	-85	P
2013-02-09 오전	1:20:41	FINAL	OB3 MID CINR1	29	100	20	P
2013-02-09 오전	1:20:44	FINAL	OB3 MID RSRP2	-130	-70	-85	F
2013-02-09 오전	1:20:50	FINAL	OB3 MID FER	0	5	0	P
...	...	...	...	...	...	...	...
2013-02-09 오전	1:21:10	FINAL	4G Verification				F

## 품질 (AS 데이터)



### S/N으로 관리

증상코드	증상명	부위그룹	부위코드	부위명	처리코드	처리명
TGR	Wi-Fi무선연불량	적용,응용 프로그램	SW41	통신망	3A	사용설명
TC2	데이터백업요청	적용,응용 프로그램	SW52	고객요청	ZD	데이터 백업(PC-Link)
TFY	기타 기본S/W 오류	적용,응용 프로그램	SW41	통신망	Z4	외부환경문제+UPGrade
TC2	데이터백업요청	적용,응용 프로그램	SW52	고객요청	ZD	데이터 백업(PC-Link)
TM7	데이터통신불량	회로 ASSY류	AT12	메인보드(H/W성)	3A	사용설명
TMD	FOTA오류	적용,응용 프로그램	SW21	기본탑재S/W(App)	Z1	F/W Bug(Up,Grade)
TBE	백리현상(영)	적용,응용 프로그램	SW16	고객설치 S/W(App)	3A	사용설명
TSF	수화음억통	회로 ASSY류	AT12	메인보드(H/W성)	RP	착하불량(제조불량)

MES  
(Manufacturing Execution System)

Mapping

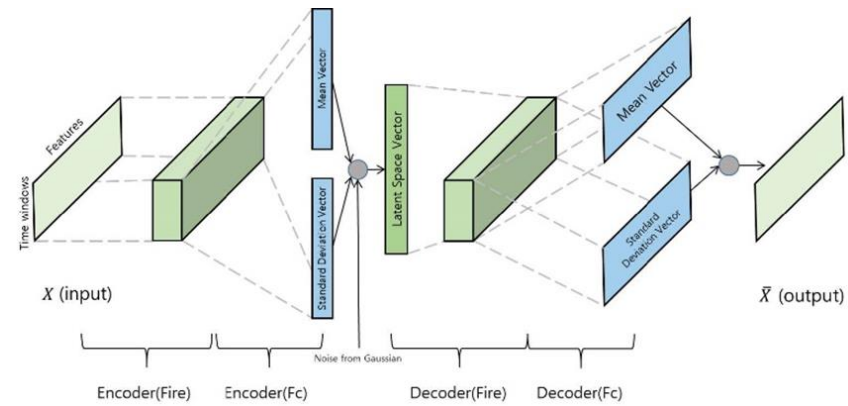


# Data Insight Value



CNC 머신 센서 → anomaly score → 사전 점검을 통한 down 방지

# 센서를 통한 딥러닝



Squeezed Convolutional Variational Autoencoder

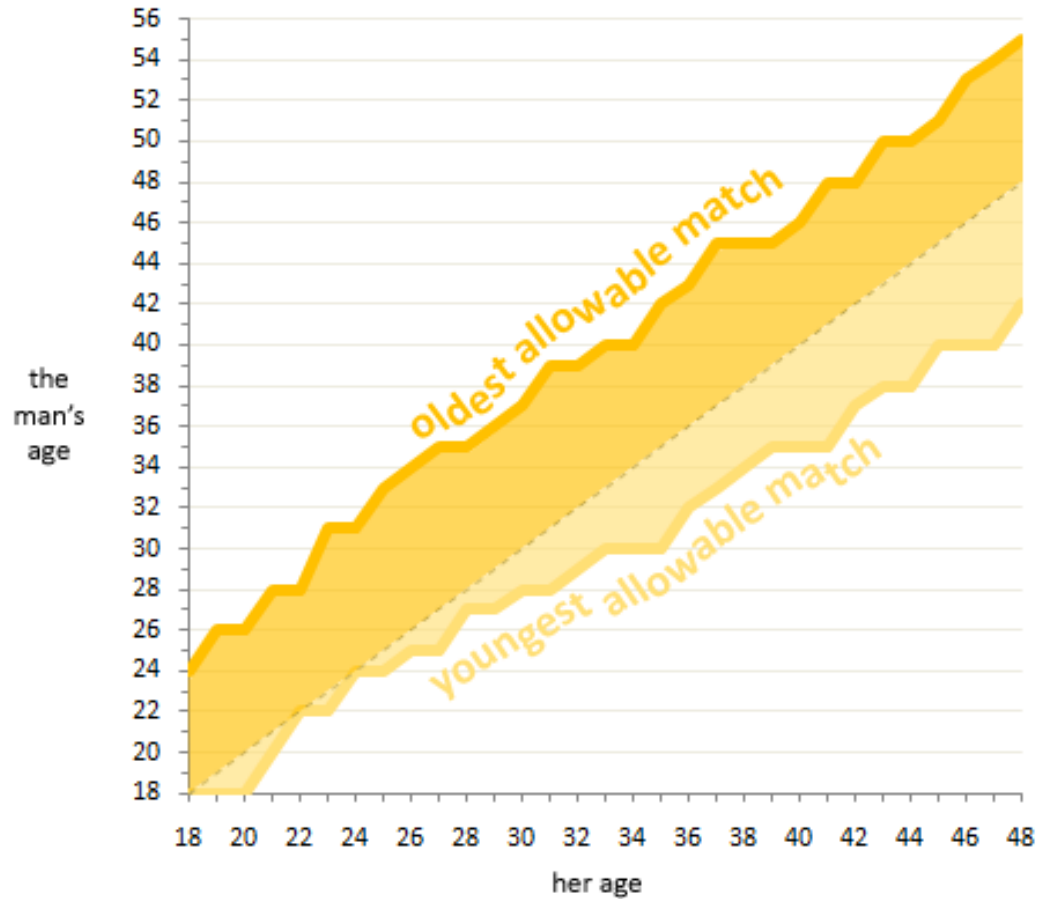
# Data Insight Value



고객 behavior → 고객 생각 <> 행동 → 정확한 추천

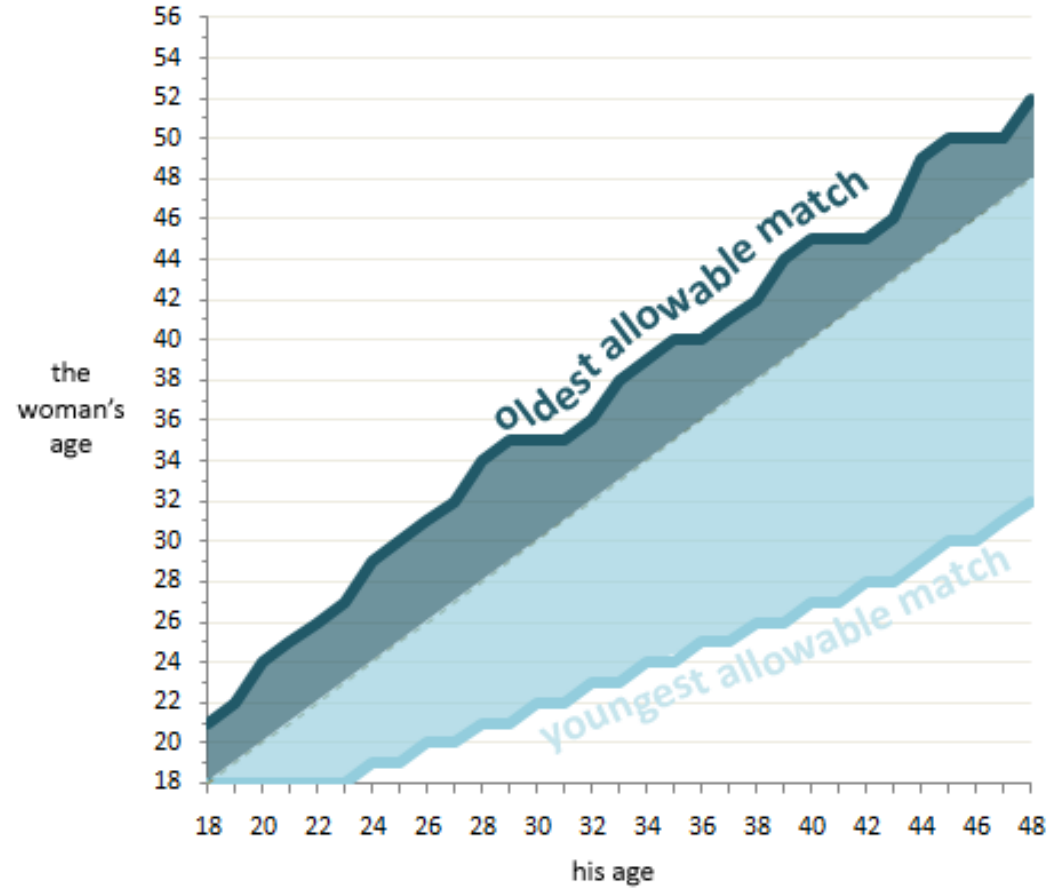
### How Female Match Preferences Change With Age

median min. & max. age values in our female users'  
"allowable match" settings



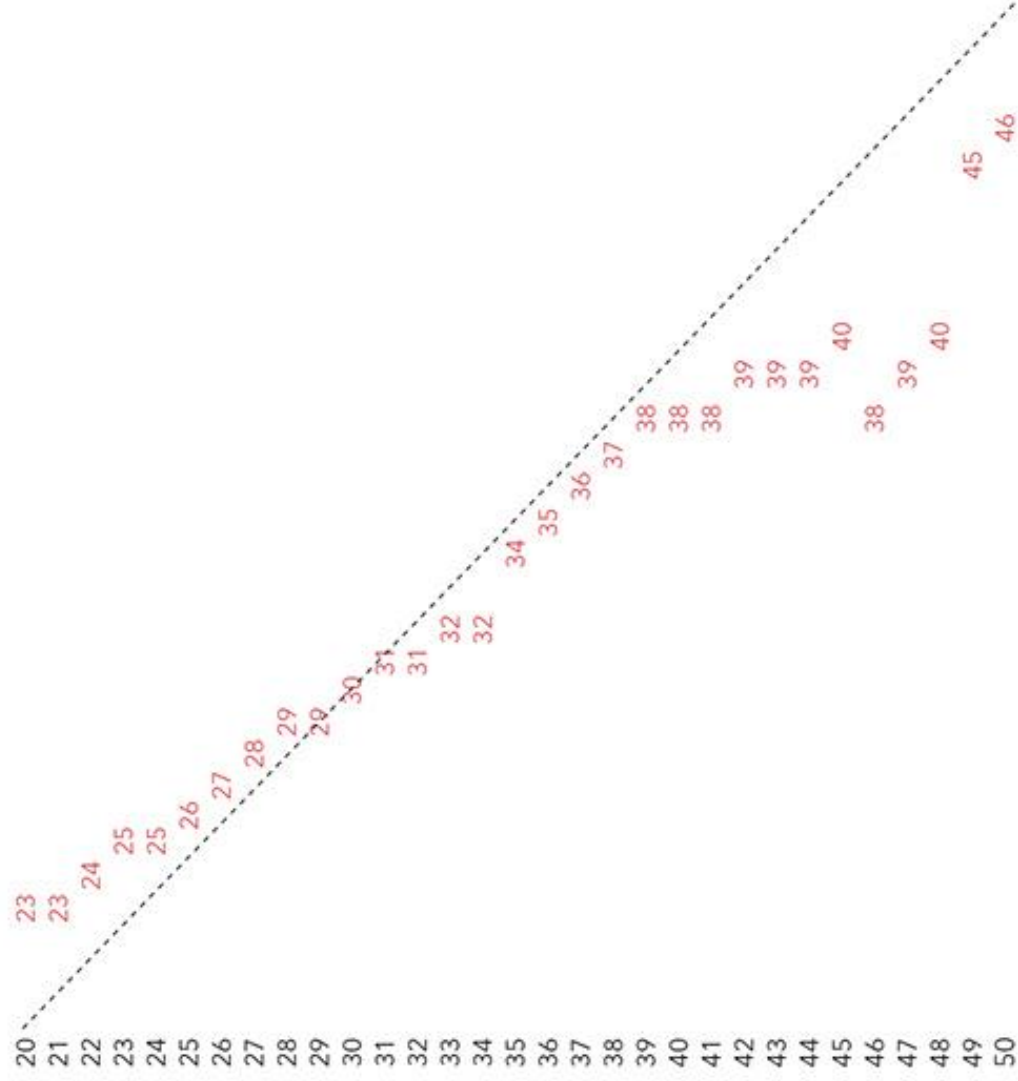
### How Male Match Preferences Change With Age

median min. & max. age values in our male users'  
"allowable match" settings

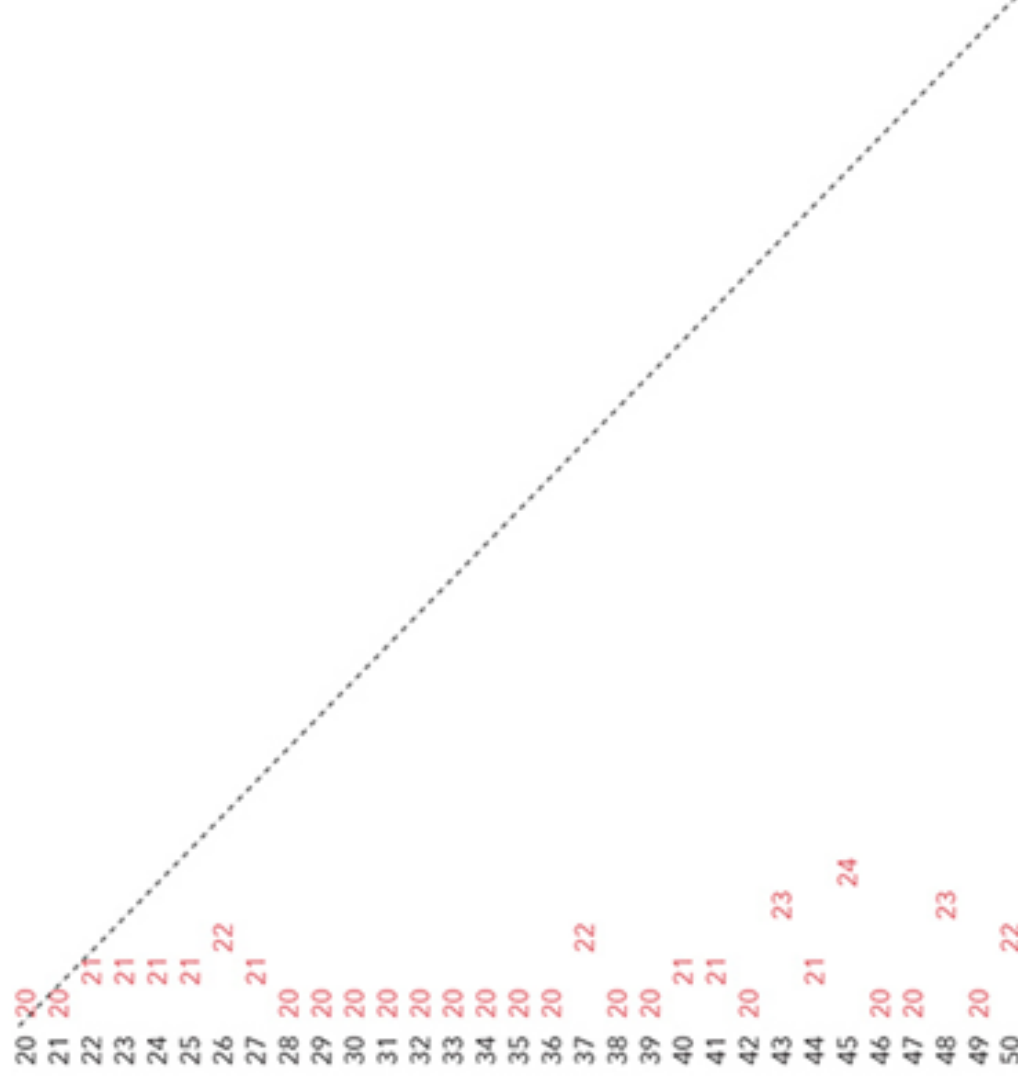


Source: <https://theblog.okcupid.com>

*a woman's age vs. the age of the men who look best to her*



*a man's age vs. the age of the women who look best to him*

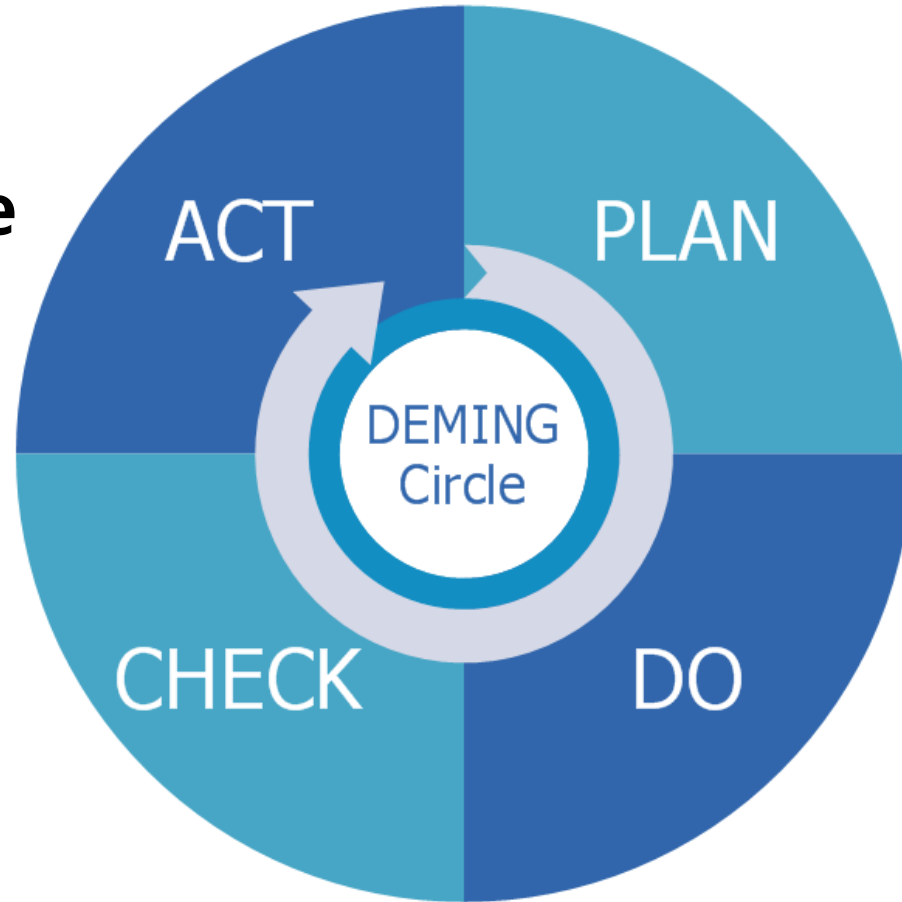


## 2. 기술에서 혁신으로



# Who?

4. Insight → Value



1. 빅데이터 분석 기획

3. Insight

2. Data → Insight

Event vs Process



**기획** data <= insight <= value





# PLAYERS



# 분석가



- Who?
  - 데이터사이언티스트, 데이터마이너, 분석가
- To Do
  - 데이터에서 비즈니스 인사이트/포사이트 도출
- 조건
  - **AI 지식**: 머신러닝, 컴퓨터, 통계학, 산업공학
  - **적용 지식**: 해당 비즈니스
  - **스킬**: 커뮤니케이션, fast learning
- 어디서 구하나?
  - 내부 교육, 고용, 임대

# 의사결정자



- Who? 마케터, 상품기획자, 공정 엔지니어
- To Do
  - AI 인사이트를 의사결정에 적용
  - AI "기획"
- 조건
  - AI 이해: AI 종류, 예상 결과, 한계
  - How? 교육!





# 리더



- Who? CEO, 본부장, 담당 임원,
- To Do
  - 인적 **투자**: 분석가 영입 + 의사결정자 교육
  - AI/빅데이터-인사이트 기반 의사결정 **지지**
- 조건:
  - AI/빅데이터 value 이해

