

Business Analytics의 완성과 미래

-산업별 예측분석(Predictive Analytics) 사례

SPSS Korea (주)데이터솔루션
허준 이사 (hoh@spss.co.kr)



Business Analytics와 Predictive Analytics

Business Analytics.

Business Analytics란?

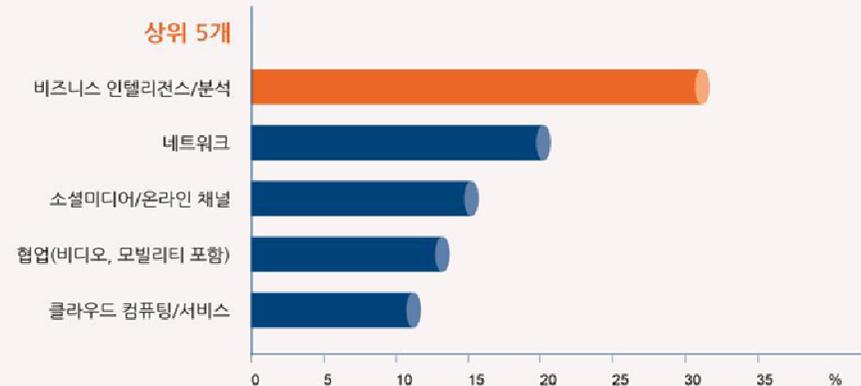
단순한 시각화를 넘어선 **과학적인 예측 기법** 을 이용하여 문제에 대한 **원인을 파악** 하고 향후를 **예측함으로써, 비즈니스 의사결정** 을 위해 필요한 **핵심 정보를 도출** 해주는 **솔루션을 의미**.

Business Analytics 시장

- 세계적인 IT분야 시장 예측기관인 Gartner Group을 비롯한 여러 연구 기관에서 가장 성장 가능성이 높은 분야로 예측.
- 기업 경쟁력에 가장 필요한 기술 영역 1위.
- 이를 바탕으로, BA 시장의 2014년 Asia/Pacific Revenue는 약 20억 달러(2224억)으로 예측.
- 각종 IT 투자 1순위가 BI/BA 분야임

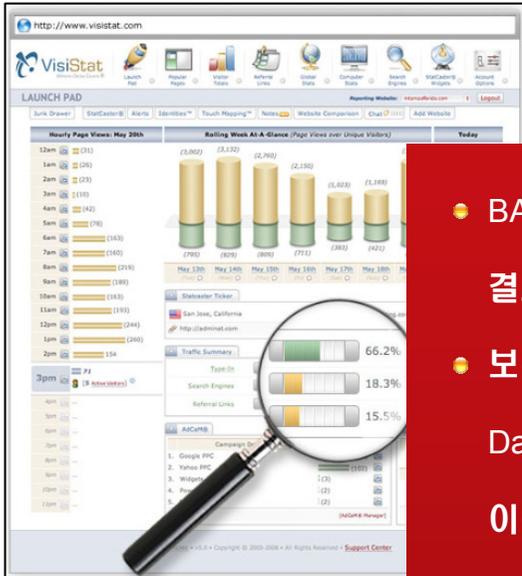
그림 1: 비즈니스 분석 부문에 대한 높은 선호도

질문: 귀하(CIO/CTO)는 경쟁력 제고를 위해 귀사의 ICT 부문을 강화한다고 언급하였는데, 아래 기술 및 솔루션 중에서 어떤 것을 최우선 항목으로 선택 하시겠습니까?

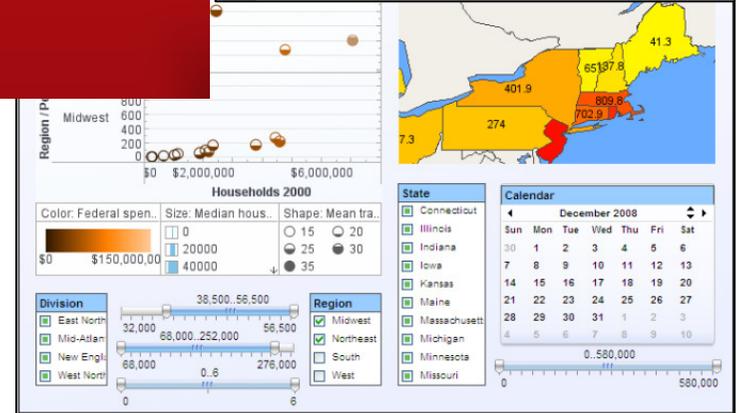


Source: IDC, 2011

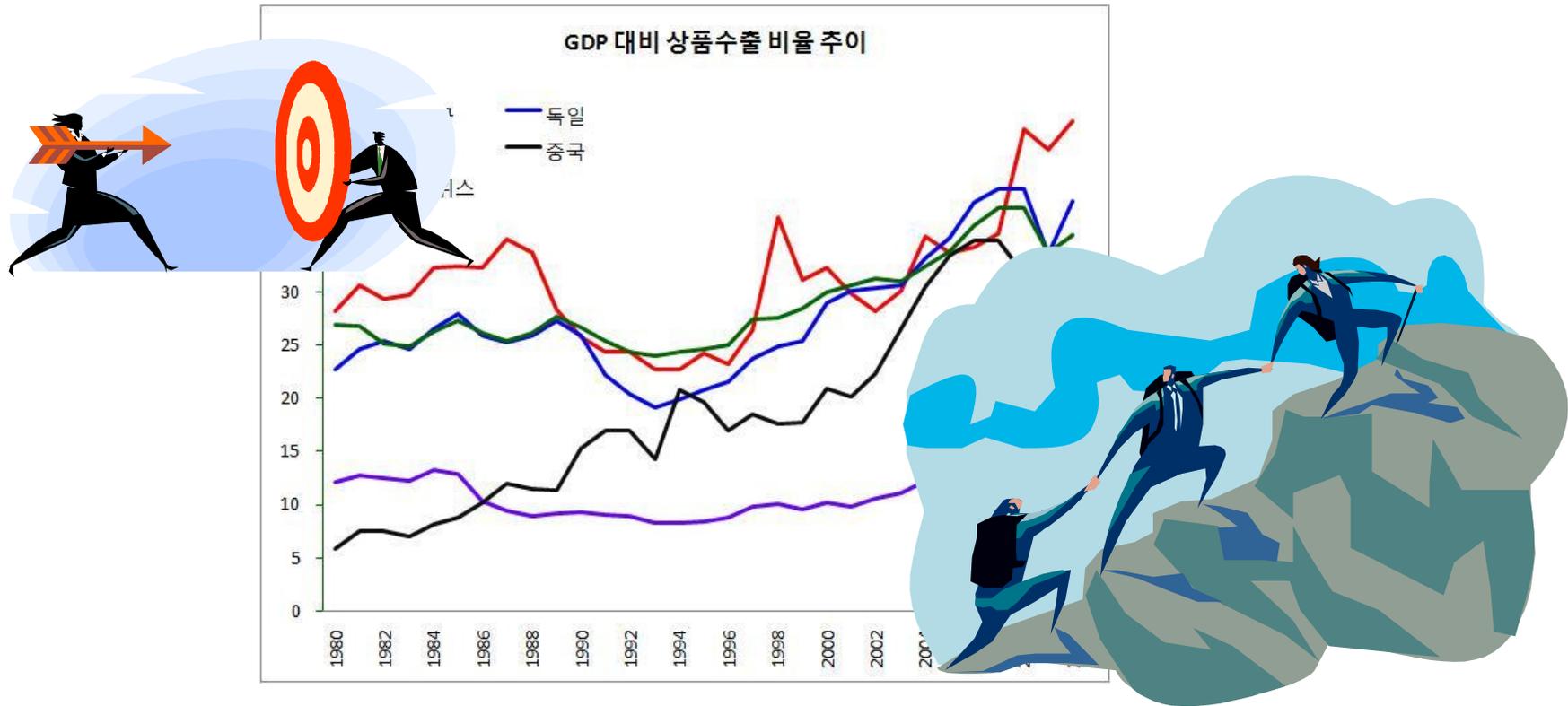
현재 Business Analytics를 위한 솔루션의 의미



- BA는 기업에서 의사결정 및 각종 내부 분석을 수행, 조회, 결과 점검을 지원하는 모든 IT적 솔루션을 의미합니다.
- 보통 OLAP(On-line Analytical Processing)이나 Web Dash-board Report, Visual Query 등 기업 내 데이터를 이용한 각종 정보 조회 및 분석을 지원하는 각종 SW 및 IT 컨설팅 기술을 BA 솔루션이라고 한다.
- 고객 분석 BA 솔루션인 CRM, 경영계획, 마케팅 조사/관리 시스템, 각종 생산/자산/판매 관리 예측, 부정적발, 준법 감시까지 BA 솔루션은 다양.

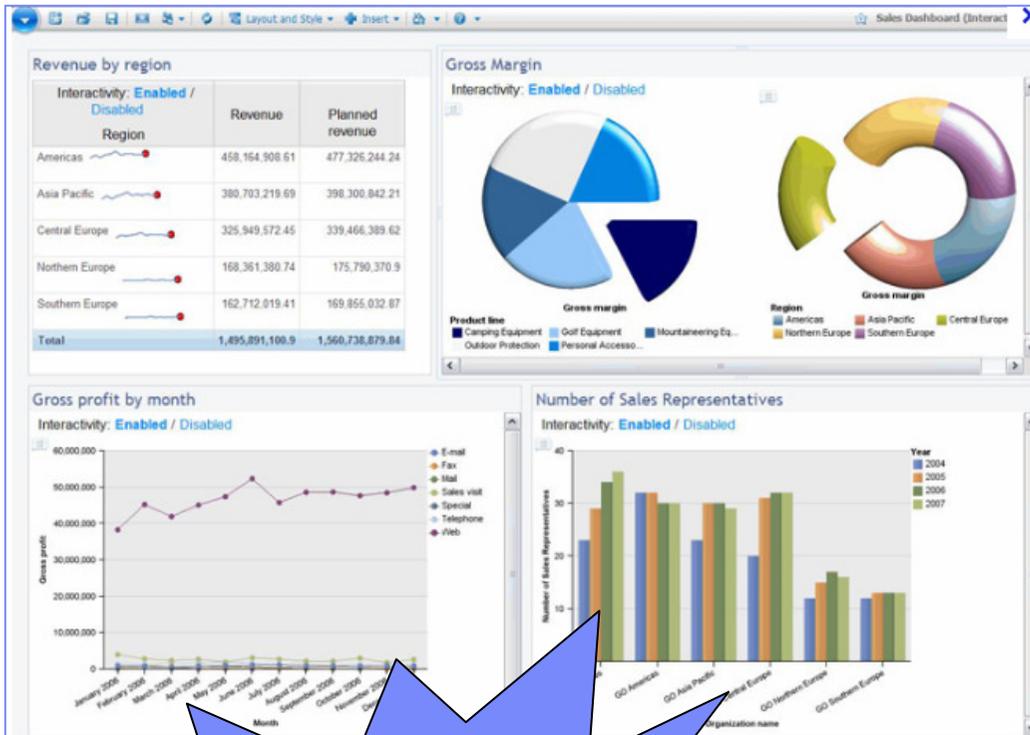


기업과 우리들은 BA를 통해 무엇을 얻고자 할 것인가?



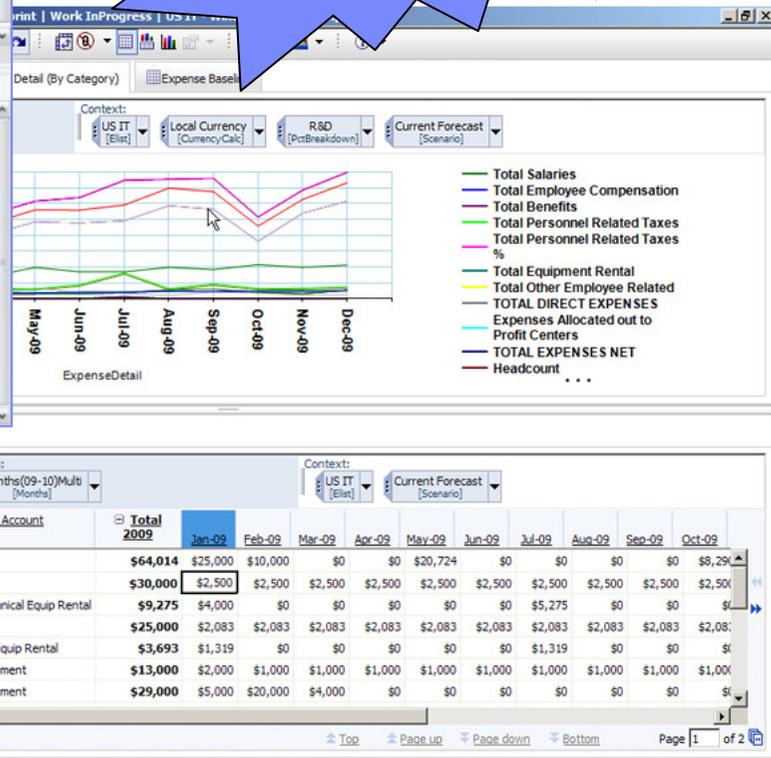
- 기업이 Business 분석을 하고, 이를 위해 BA 솔루션을 도입하는 가장 큰 이유는 과거의 정보를 자유롭게 분석하여, 미래를 **예측/대비**하기 위해서이다.

그렇다면...현재 BA는 무엇을 중심으로 분석하고 있을까?(1)



현재의 비즈니스 분석이라는 것은...

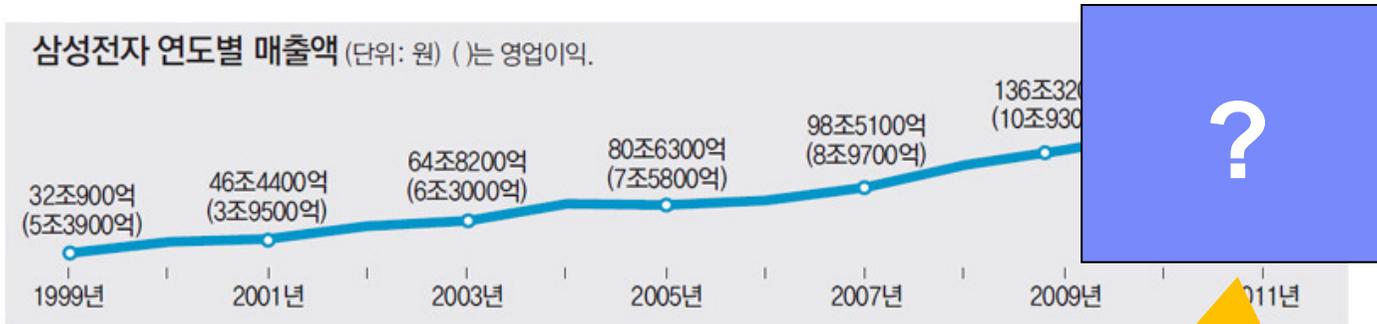
어찌 보면... 과거의 데이터를 그냥 보고 있을 뿐이다!!



Manual Entry

Description	Expense Account	Total 2009	Jan-09	Feb-09	Mar-09	Apr-09	May-09	Jun-09	Jul-09	Aug-09	Sep-09	Oct-09
1 ReportNet Research Overtime		\$64,014	\$25,000	\$10,000	\$0	\$0	\$20,724	\$0	\$0	\$0	\$0	\$8,290
2 Printer Incentive Comp		\$30,000	\$2,500	\$2,500	\$2,500	\$2,500	\$2,500	\$2,500	\$2,500	\$2,500	\$2,500	\$2,500
3 servers for Trade Show Computers & Technical Equip Rental		\$9,275	\$4,000	\$0	\$0	\$0	\$0	\$5,275	\$0	\$0	\$0	\$0
4 Bring down que program Incentive Comp		\$25,000	\$2,083	\$2,083	\$2,083	\$2,083	\$2,083	\$2,083	\$2,083	\$2,083	\$2,083	\$2,083
5 servers for Trade Show Office AV Studio Equip Rental		\$3,693	\$1,319	\$0	\$0	\$0	\$0	\$1,319	\$0	\$0	\$0	\$0
6 Corporate rollover Travel & Entertainment		\$13,000	\$2,000	\$1,000	\$1,000	\$1,000	\$1,000	\$1,000	\$1,000	\$1,000	\$1,000	\$1,000
Team Dinner Travel & Entertainment		\$29,000	\$5,000	\$20,000	\$4,000	\$0	\$0	\$0	\$0	\$0	\$0	\$0

그렇다면...현재 BA는 무엇을 중심으로 분석하고 있을까?(2)

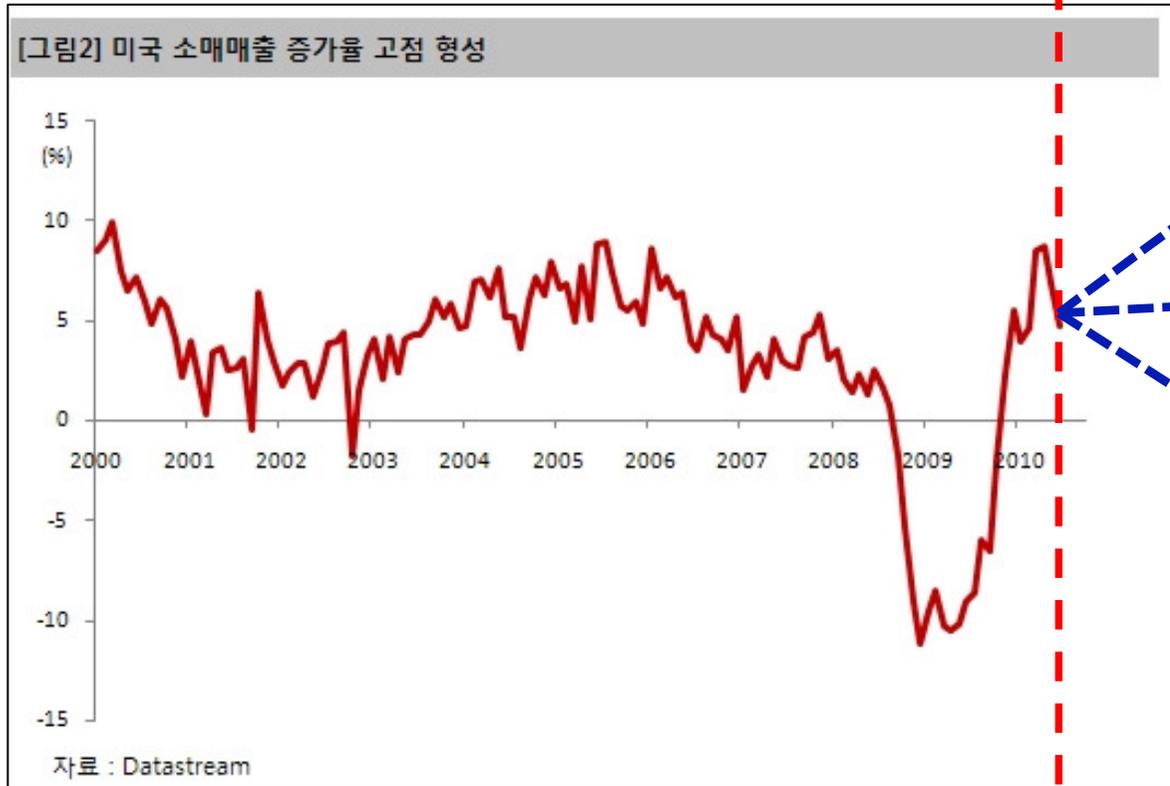


이게 만약 BA를 통해서 나온 정보라면, 그 다음 해의 결과는 어떻게 될까요?

→ 현재 우리가 주로 사용하는 BA는 다음을 사람이 직접 추정하게 합니다.



그렇다면...현재 BA는 무엇을 중심으로 분석하고 있을까?(3)

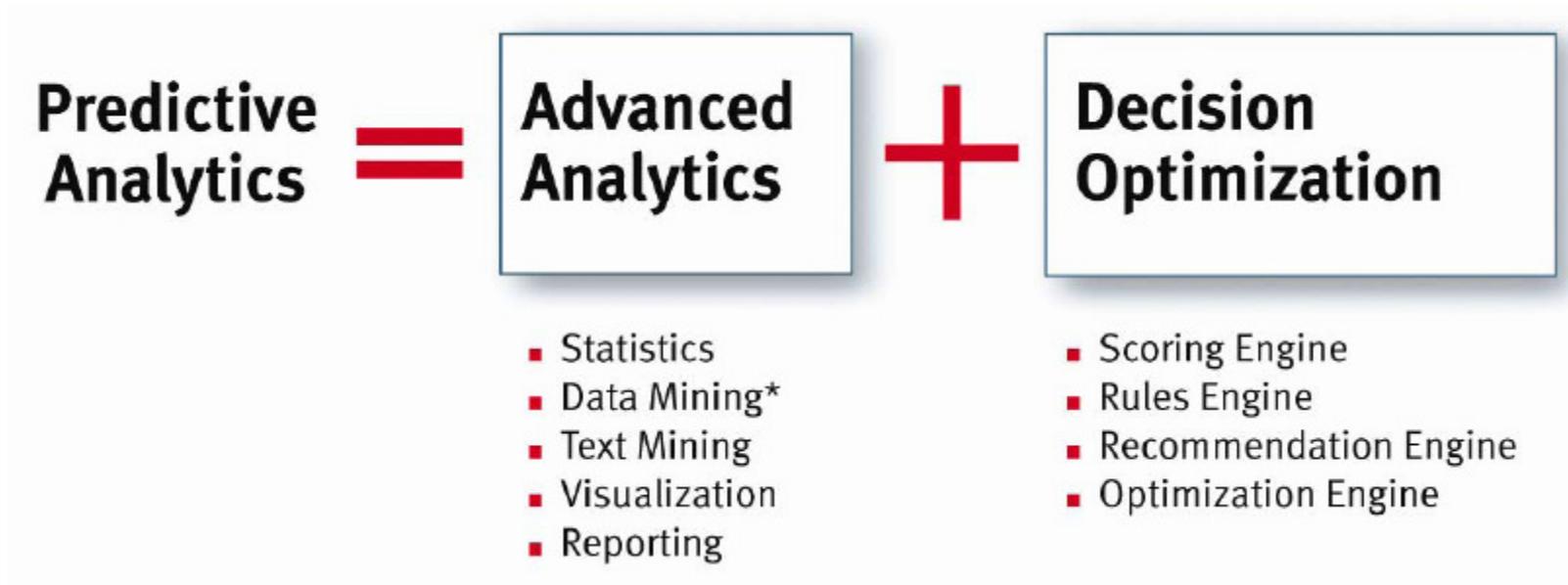


으아~ 이렇게 복잡하면, 다음은 어찌될 것인가?



Predictive Analytics (1)

- 이제 PA(Predictive Analytics)를 빼고, BA를 논하는 것은 아무 의미가 없다.
- PA란 무엇인가?

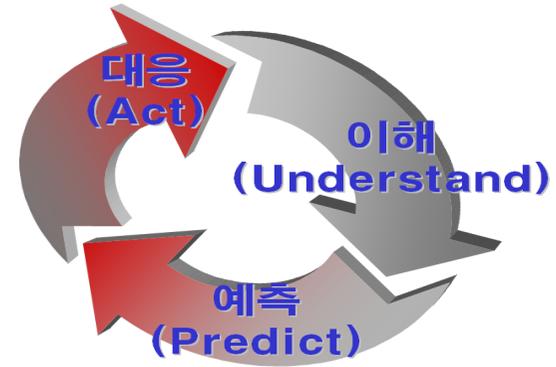


Predictive Analytics (2)

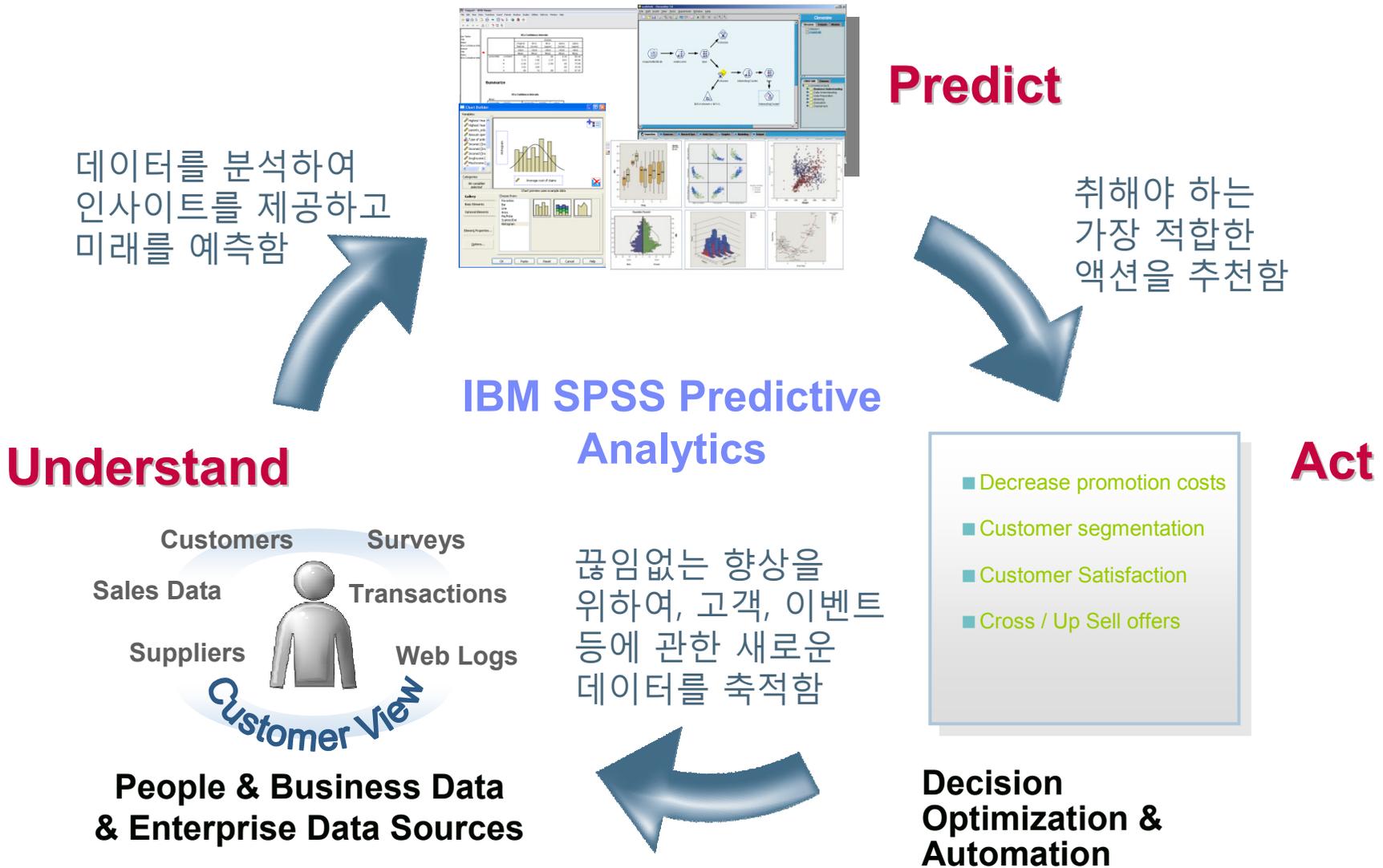


What?에 대한 해답을 주는 것
과거와 현재의 비즈니스를 요약 하는 것

Why?에 대한 해답을 주는 것.
과거와 현재의 비즈니스를 이해하여
미래의 비즈니스를 예측 하고 최적화하는 것



Predictive Analytics (3)



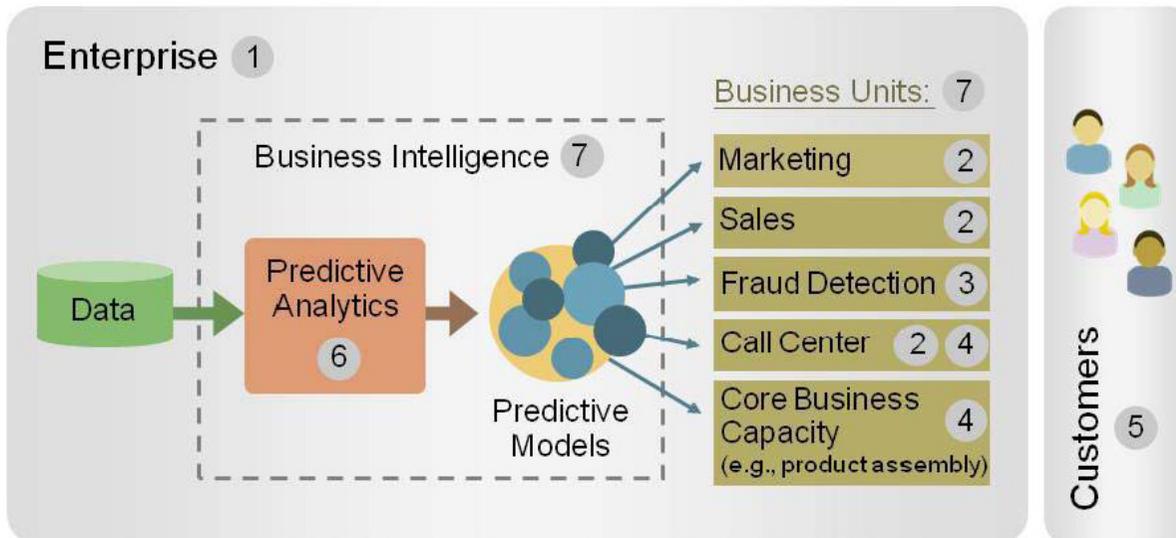
Predictive Analytics를 해야 하는 7가지 이유

Seven Reasons You Need Predictive Analytics — Key Strategic Objectives Attained:

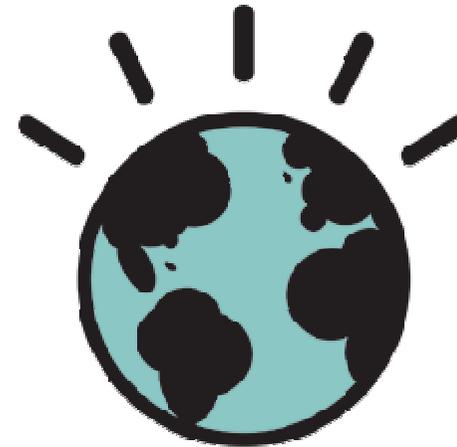
These strategic objectives can be attained to their full potential only by employing predictive analytics.

1. **Compete** – Secure the Most Powerful and Unique Competitive Stronghold
2. **Grow** – Increase Sales and Retain Customers Competitively
3. **Enforce** – Maintain Business Integrity by Managing Fraud
4. **Improve** – Advance Your Core Business Capacity Competitively
5. **Satisfy** – Meet Today's Escalating Consumer Expectations
6. **Learn** – Employ Today's Most Advanced Analytics
7. **Act** – Render Business Intelligence and Analytics Truly Actionable

이제 BA는 PA를 중심으로 이루어진다!!!



IBM Smarter Planet



IBM이 PA를 중심으로 한 Smarter Planet의 사례 – i-BEE(1)



Smarter Energy

통합모니터링, 알람

- 각종 Meter 및 센서 정보 통합
- 알람을 위한 임계값 관리

에너지 효율 시뮬레이션 및 최적화

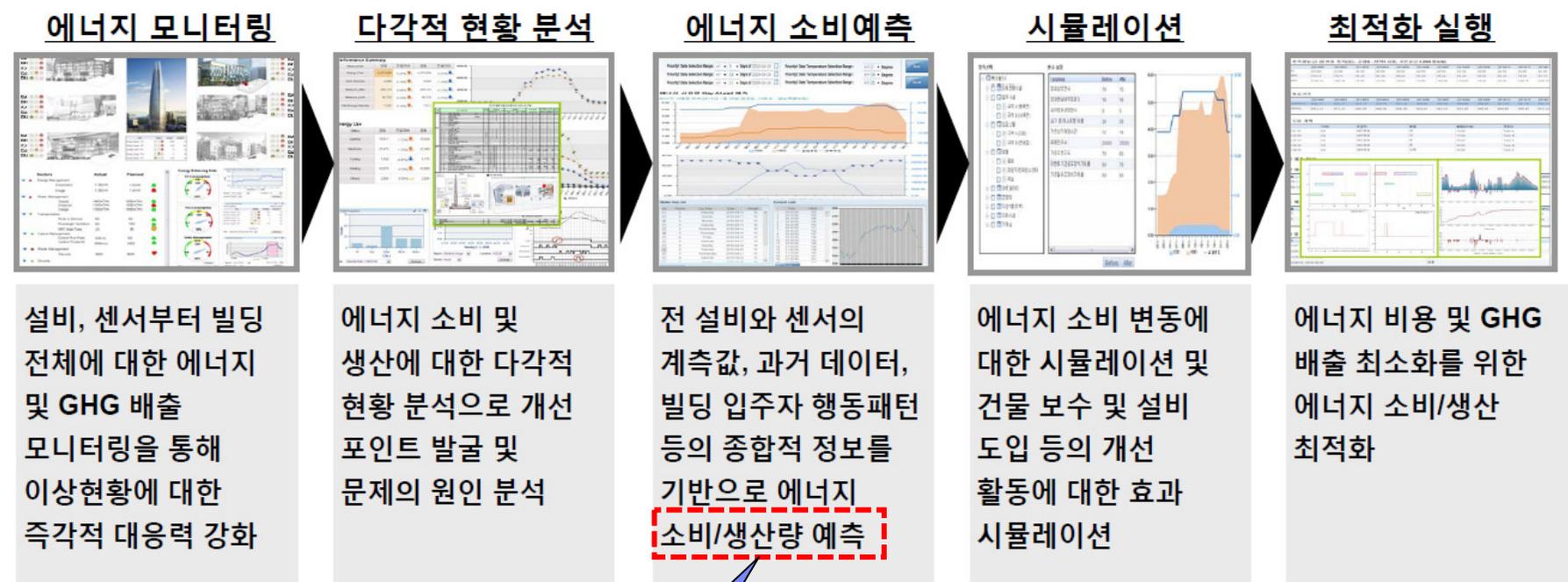
- 에너지 수요/공급 예측 및 최적화
- Emission 배출량 산출 및 최소배출을 위한 에너지 사용 최적화

최적 모델 적용을 위한 Automation

- 에너지 수요/공급 최적 스케줄 적용
- 에너지 Peak 관리를 위한 자동제어

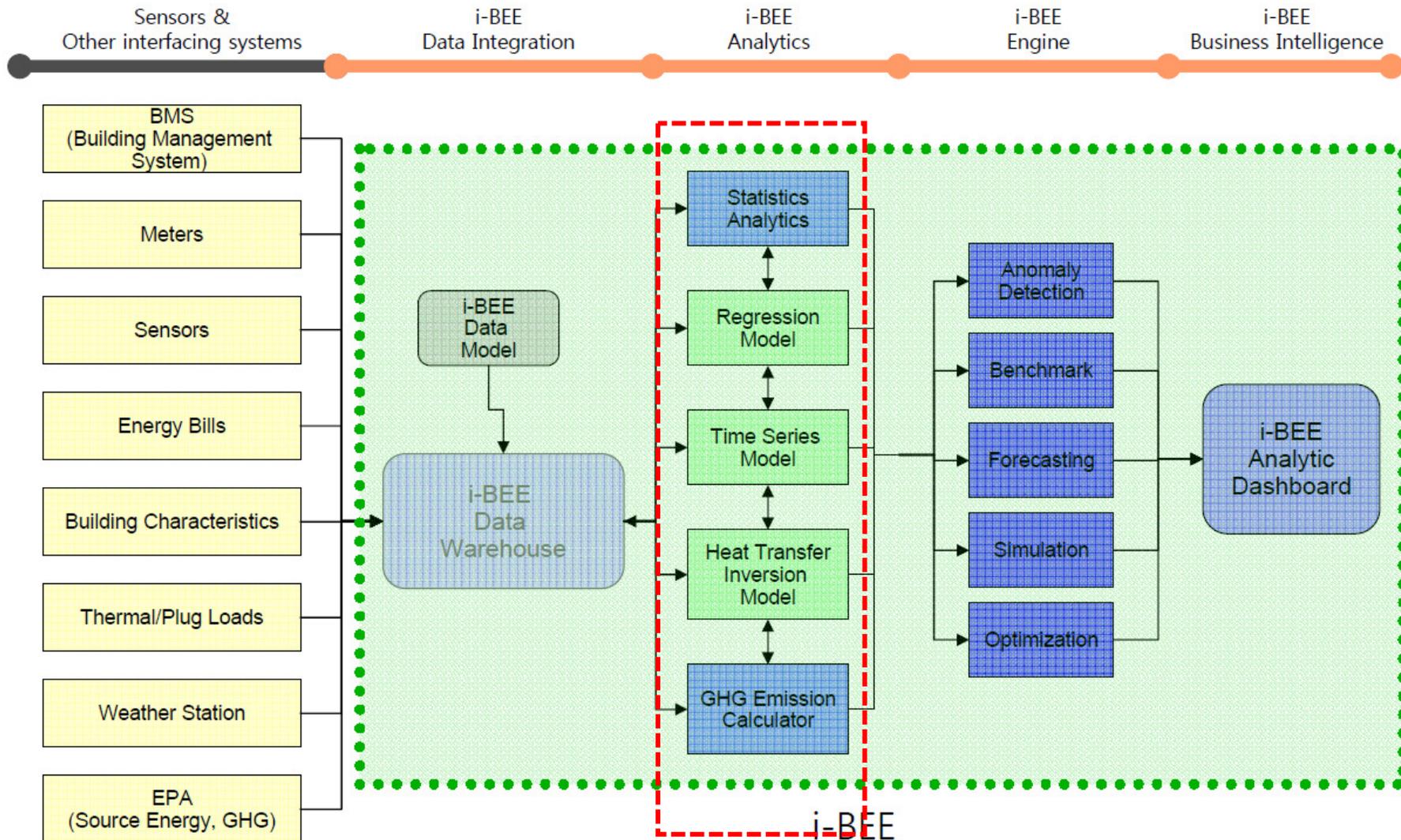
IBM이 PA를 중심으로 한 Smarter Planet의 사례 - i-BEE(2)

에너지 최적화 관리



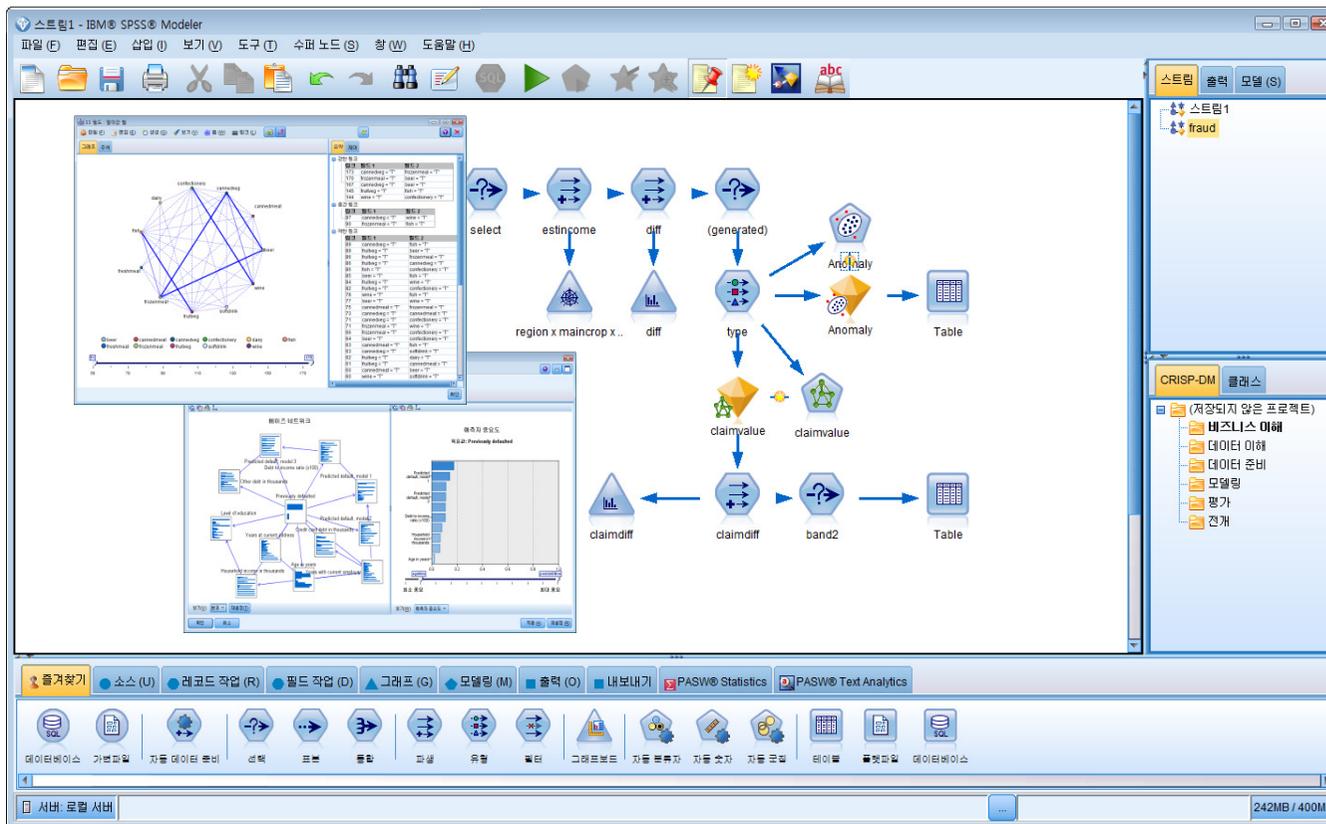
차별화의 중심
예측!!!

IBM이 PA를 중심으로 한 Smarter Planet의 사례 – i-BEE(3)



IBM의 PA 솔루션 IBM SPSS Modeler 14.2

IBM SPSS Modeler은 데이터를 로딩, 변환, 정제, 모델링, 그래프, 결과의 출력까지 하나의 소프트웨어 내에서 가능하며, 이것은 모두 대화상자와 아이콘, 메뉴를 이용하여 완벽한 GUI를 구현한 소프트웨어입니다.



- Visual한 작업방식을 통한 손쉬운 접근
- 작업의 유연성을 보장하는 인터페이스
- 최단 시간 내에 가장 다양한 모델 생성
- 개방형 구조 (Open Architecture)

PA를 위한 IBM SPSS Modeler의 다양한 기능들



- ### Sources Node

 - ✓ 데이터 연결 노드
 - ✓ 데이터베이스 연결 또는 가변형식, 고정형식 파일의 데이터, IBM SPSS Statistics, SAS 파일 등의 다양한 파일들을 데이터로 읽음
- ### Operations Node

 - ✓ 데이터 변환 작업 노드
 - ✓ 샘플링, 레코드 또는 필드 단위의 데이터 병합 및 필터, 변수파생, 모형평가를 위한 파티션 작업 포함
- ### Graphs Node

 - ✓ 데이터 도식화 노드
 - ✓ 크게 데이터 탐색으로 이용되는 히스토그램, 2차원 및 3차원 도표와 ROI Chart 등과 같은 평가도표로 이용
- ### Modeling Node

 - ✓ 데이터 모형화 노드
 - ✓ Decision Tree, Regression, Neural Network, Clustering, Association 등 다양한 종류 이용가능
- ### Output Node

 - ✓ 마이닝 결과 출력 노드
 - ✓ 최종 결과를 테이블, 외부파일로 출력하는 기능, 생성된 모델들 간의 예측력을 평가하는 기능 포함

*Data Mining
솔루션을
넘어 PA
솔루션으로...*

PA를 위한 IBM SPSS Modeler의 다양한 모델링 기능들

의사결정나무 분석

C&RT 탐색 CHAID C5.0

- 최신의 알고리즘 C5.0, Quest 탑재
- 대화식 모형으로 Tree 사용자 정의 가능

연관성 분석

Apriori 카르마 순차규칙

- 다양한 연관성 분석 가능
- 동시 후항을 지원하는 CARMA
- 시간의 개연성을 고려한 순차규칙

통계 분석

선형회귀 PCA/요인 로지스틱 GenLin Cox 회귀

- 연속형의 선형회귀모형과 범주형의 로지스틱 회귀모형, 요인분석 등 다양한 회귀모형 가능

군집화 / 세그먼트

K-평균 코포넨 TwoStep

- 대용량 자료의 효율적인 Two-Step Clustering 알고리즘
- 편리한 군집 Rule 찾기 기능

Screening 모델

필드 선택 Anomaly

- 모델링을 더욱 강력하고, 편리하게 하기 위한 Cleaning 작업
- 군집분석을 이용한 특이치, 이상치 제거 기능

AI기법 / 기타

신경망 Bayes Net Text Mining

- 전문가를 위한 신경망의 다양한 Option 조정 가능
- 불완전한 데이터에서 패턴을 찾아내는 Text Mining 기능 (별도)

기업 현장에서 범용성을 가지고, 적용할 수 있는 가장 많은 예측적 모델링 분석 기능을 보유한 "IBM SPSS Modeler"

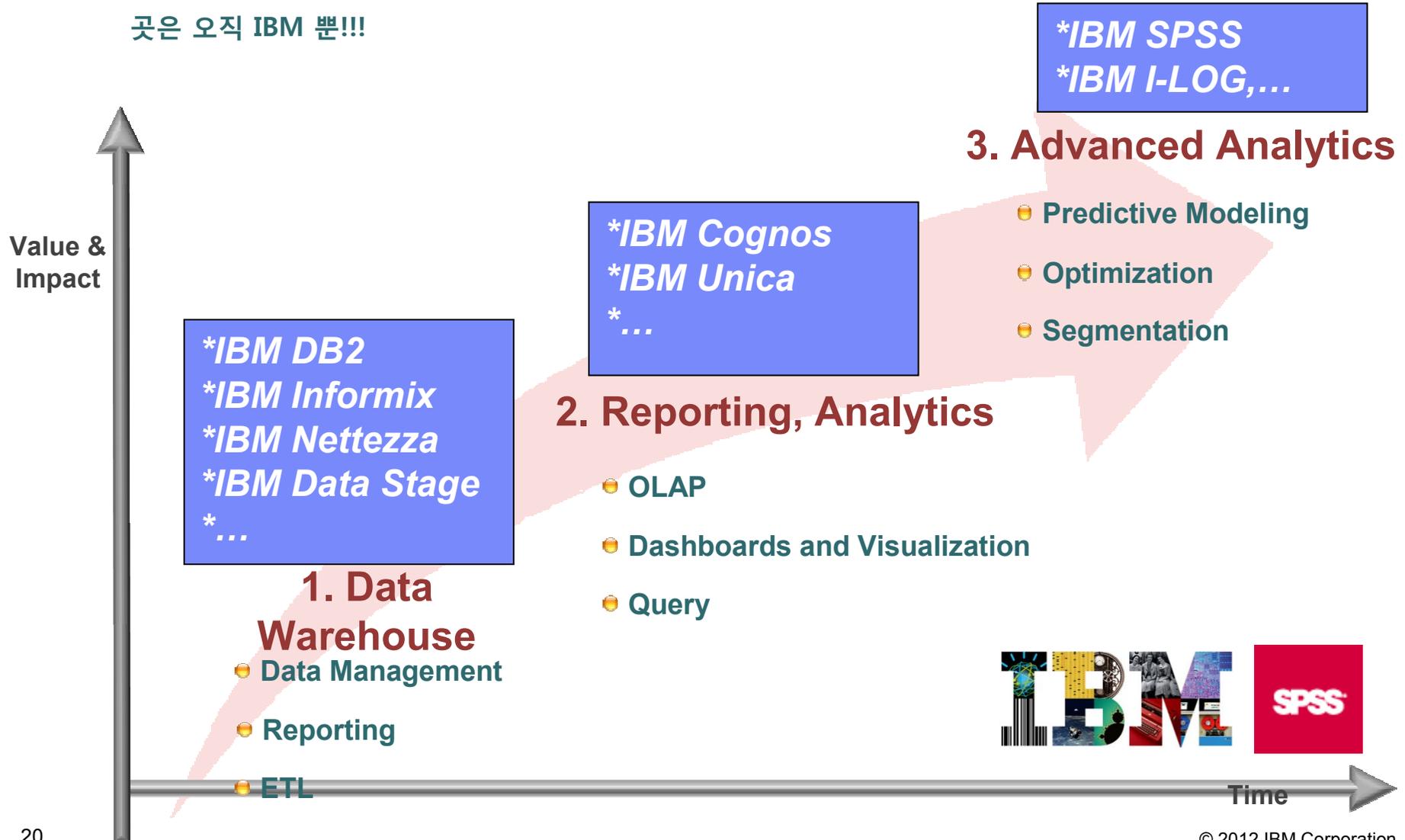
통계 분석

시계열 분석 선형 선형회귀 PCA/요인 판별분석 로지스틱 GenLin Cox 회귀 SVM Bayes Net SLRM KNN

- 시계열 분석 뿐만 아니라 연속형의 선형회귀모형과 범주형의 로지스틱 회귀모형, 요인분석 등 다양한 회귀모형 가능
- SVM, Bayesian Network, KNN, 등 그 외 다양한 고급 통계 분석 기능 제공

IBM의 PA를 포함한 고도화 BA 단계별 Line-up

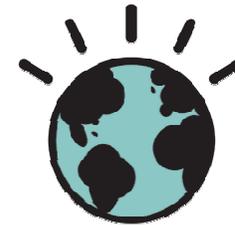
- 현재 BA 전 영역에 대하여, 실용적이고, 실제 기업 기반 하에서 검증된 솔루션을 보유/제공할 수 있는 곳은 오직 IBM 뿐!!!



PA의 그림자...그래도 PA는 향후 미래의 방향이다!!!

Predictive Analytics가 옳고,
방향은 맞지만, 너무 전문성을
요하는 것은 같아...

Data Mining기술, 어려운
통계적 기법, 인공지능 기법,
고도의 계량경제 바탕의
다양한 수학적 기법들...



**IBM의 Smarter
Planet이 바로 어려운
PA의 실제 적용을
위한 방향임.**

IBM SPSS Modeler를 이용한 Predictive Analytics의 사례들

공공-고속도로(1)

사례개요

- *특별 교통기간(추석, 설, 휴가기간)에 전국 고속도로의 주요 거점 간 소요시간 예측 모델의 개발
- *소요시간 예측에 따른 대국민 정보 제공→ 궁극적인 고속도로 정체 해소/ 소통원활

How? What?

- *통계분석 기법의 시계열(Time Series) 분석 기능과 회귀분석을 이용하여, 개발
- *시공도 및 교통공학의 예측 기법 접목하여 모델 개발

결과 및 시사점

- *2008년도 추석 연휴 예측율 30분 이내에서 정확
- *추석연휴 귀성/귀경길 소통원활(2008년)
- *세계 최초의 통계분석 / 데이터마이닝을 이용한 소요시간 예측 시스템화

공공-상업용 차량 운전자 운전정밀검사(1)

사례개요

*사업용 차량을 운전하는 운수종사자의 경우 반드시 운전정밀 검사를 수행함

*사전 운전정밀 검사와 사후 교통사고 DB를 결합하여, 향후 사업용 차량 운수종사자의 사고유발 경향성을 예측하고 이를 안전교육에 활용함.

How? What?

*의사결정나무 분석/ 군집분석 등을 이용하여, 운수종사자를 세분화하고, 각 정밀 검사 시험 성적 기반에 따라서 안전도 및 사고 유발 확률 계산

결과 및 시사점

*운수 종사자들 대상 교육 및 안전 주의용으로 사용

*교통사고 방지에 효과적이었다는 내부 평가

유통-면세점의 경영계획

사례개요

- *호텔 산하 면세점의 경영계획 수립을 위해 가장 중요한 정보는 면세품의 수요예측
- *ERP를 통한 실질적인 활용도 높은 계획 수립을 위한 호텔 취급 면세품의 수요 예측 모델 개발(1년 이상 장기)

How? What?

- *기존 면세품의 판매 현황 및 각종 경제 지표(유가, 환율)를 고려한 최적 회귀 모형을 개발
- *약 2만 여종의 면세품 시장 수요/개별 수요 예측 모델 개발/자동화

결과 및 시사점

- *경영계획의 모든 시작점은 수요예측 결과 기반으로 시작
- *각 기업이 도입하고 있는 ERP 시스템의 가장 핵심이자 시작점이 바로 수요예측 모델
- *SAP ERP-BPC 시스템 과 연동

유통-인터넷 쇼핑몰 추천상품

사례개요

- *인터넷 유통 기업의 추천상품 모델 개발
- *고객이 구매하고자 하는 물품과 가장 연관성(과거 유사판매 정도)이 높은 물품을 동시에 추천하는 모델을 통한 추가 매출 달성

How? What?

- *데이터 마이닝의 연관성 분석 알고리즘 이용
- *최적의 연관 다빈도 조합을 생성하여, 고객에게 상품을 추천

결과 및 시사점

- *간단한 정보 제공으로 매출 상승 기여
- *무엇보다도, 투자대비 효과가 높은 점이 매우 중요
- *가끔 전략적인 추천 상품을 제시하는 등의 융통성있는 모델 개발
- *유통업의 기본적인 모델이 되었음(URC)

제조-제철회사의 불량인자 탐색

사례개요

- *제철회사에서 최종 제품이 나왔을 때, 다양한 불량요인이 발생할 수 있음
- *그러나, 여러 단계의 공정 중 어느 단계/어떤 요인으로 불량이 발생했는지 파악 어려움
→ 마이닝 모델 필요

How? What?

- *마이닝의 의사결정나무 분석 기법을 이용하여, 일정 공정이 지나면 실제 불량이 발생할지 예측하는 모델 개발
- *동시에 QC 기법도 도입하여, 품질 제어

결과 및 시사점

- *최종 제품 전에 불량 발생 공정 제거/수정
- *최종 품질 향상
- *생산관리의 QC 관리의 새로운 방법을 제시(6시그마의 과제)

금융-저축은행의 연체 예측

사례개요

*저축은행의 가장 큰 문제 중 하나는 소액 대출자들의 대출 연체 → 1금융권 보다 2배 이상 연체율이 높음
*기존 신용등급 모델로는 한계점 봉착 → 신규 연체 예측 모델 개발 필요

How? What?

*마이닝의 신경망 분석을 이용하여, 실제 연체가 될지 안 될지에 대한 예측 모델 개발
*대출 상품별(대학생, 직장인,... 등)로 세부적인 맞춤형 예측 모델 개발

결과 및 시사점

*기존 신용평가 사의 모델보다 2~3배 높은 정확도 발휘

기타-방송 시청률 예측

사례개요

- *광고 기획사에서 핵심 포인트는 TV의 광고 시청률
- *광고 시청률에 따라, 광고 시간대를 구매하고, 구매하지 않는 것이 결정
- *시청률 예측 모델을 개발하는 것이 필요

How? / What?

- *통계적 시계열 모델, Bass의 확산 모델, SOV-TOM 등의 기법을 이용하여, 다양한 수요예측
- *광고 시간대, 시청 연령대, TV 프로그램 장르, 경쟁사 TV 방송 프로그램 등의 정보를 이용하여 모델 개발

결과 및 시사점

- *5% 이하의 예측율
- *광고 기획사의 마케팅 정보 시스템의 출발점
- *각종 경영 지표/ Simulation 자료로 적극 활용
- *예측의 중요성

그 외 Predictive Analytics 적용사례

공공분야

- *세금 부정 납세자 파악
- *교통안전공단 운행기록계 이용 사고 예측 모형 개발
- *신용회복위원회 데이터 마이닝
- *FIU 자금세탁 방지 모형 개발

금융분야

- *생명보험사 고객세분화 / 각종 고객 추가 매출 가입 모형 개발
- *화재보험사 언더라이팅 심사 예측 지원 모형 개발
- *제1,2 금융권 은행권 바젤 2 모형(리스크 관리 모형 개발)
- *증권사 우수고객/ 카드사 고객 이탈 방지 모형 개발

통신/제조/유통분야

- *화장품 회사의 URC-Selling 모형 개발
- *홈쇼핑 고객 데이터 마이닝 모형 개발
- *통신회사 트래픽 예측 모형 개발
- *제조사 Quality Control 시스템/관련 모형 개발

기타 분야

- *대형 종합 병원의 의료 CRM
- *종합 유통사/프랜차이즈 점 상권분석
- *제약회사 생동성 실험 모듈 개발
- *기타 다수...