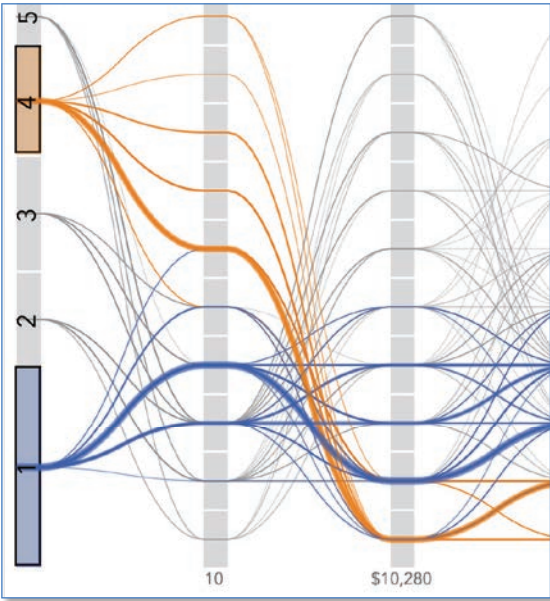


# SAS<sup>®</sup> Visual Statistics on SAS<sup>®</sup> Viya<sup>™</sup>

개방형 분산 환경에서 분석 모델을 인터랙티브 방식으로  
구축, 조정 및 평가하여 보다 신속하고 심층적인 인사이트 확보



## SAS<sup>®</sup> Visual Statistics on SAS<sup>®</sup> Viya<sup>™</sup>는 어떤 솔루션인가?

SAS<sup>®</sup> Visual Statistics on SAS<sup>®</sup> Viya<sup>™</sup>는 데이터를 인터랙티브 방식으로 탐색하고 드래그-앤-드롭 인터페이스나 프로그래밍 인터페이스를 사용해 탐색 및 예측 모델을 구축할 수 있도록 지원합니다. 데이터 사이언티스트들은 비즈니스 분석가들과 협업하여 모델을 조정함으로써 보다 심층적인 인사이트를 얻을 수 있습니다. 분산형 인메모리 프로세싱은 데이터 탐색 및 모델 개발에 소요되는 시간을 단축합니다.

## SAS<sup>®</sup> Visual Statistics on SAS<sup>®</sup> Viya<sup>™</sup>가 중요한 이유는?

시장 상황이나 고객 행동 패턴의 변화를 분석 모델에 신속하게 반영할 수 있습니다. 각 세그먼트나 그룹에 적합한 모델을 신속하게 실험하고 반복적으로 구축하며 개선하면 실행 가능하고 유용한 인사이트를 얻을 수 있습니다. 또한 분석 및 비즈니스 팀에게는 생산성 및 협업 능력을, IT 팀에게는 거버넌스 및 중앙집중식 관리를 지원할 수 있습니다.

## SAS<sup>®</sup> Visual Statistics on SAS<sup>®</sup> Viya<sup>™</sup>는 누구를 위한 솔루션인가?

SAS Visual Statistics는 심층적인 인사이트 확보를 위해 예측 모델을 구축하거나 조정 및 평가해야 하는 통계 전문가, 데이터 사이언티스트, 프로그래머, 일반 데이터 사이언티스트를 주요 대상으로 설계되었습니다.



시장 상황과 고객 선호도가 빠르게 변화하면서 대부분의 분석 소프트웨어 패키지와 아키텍처들은 그 속도를 따라잡지 못하고 있습니다. 또한 데이터 준비, 탐색 및 예측 분석에서 기술 격리 현상이 발생하면서 필요에 따라 다양한 모델을 생성하기가 어려워졌습니다. 여러 세그먼트에 대한 모델을 실행하는 데 너무 많은 시간이 소요되고 수동 모델 평가에 많은 인력이 투입되고 있다면 SAS Visual Statistics가 해결 방안이 될 수 있습니다.

## 주요 특징

- **새로운 기회에 대한 신속한 포착으로 경쟁력 강화** 데이터 사이언티스트와 통계 전문가들은 가장 적합한 분석 모델링 기법을 사용하여 해결하기 어려운 문제의 관찰 내용을 세분화해 처리할 수 있습니다. 그 결과, 신속하게 인사이트를 발견하고 수익을 창출할 수 있는 새로운 방법을 찾아낼 수 있습니다.
- **분석가의 생산성 개선** SAS Visual Statistics를 사용하면 수동 작업이 줄어들고 협업이 강화됩니다. 여러 명의 사용자가 시각적인 방식으로 데이터를 다루면서(변수 추가/변경, 특이치 제거 등) 이러한 변화가 예측력에 미치는 영향을 신속하게 파악할 수 있습니다. 모델이 비즈니스 또는 도메인 전문가의 요구를 충족하지 못할 경우, 분석가는 즉시 이를 조정할 수 있습니다.
- **모델 개발 시간 대폭 단축 및 빠른 시간 내 더 나은 모델 적용** 특정 그룹 또는 세그먼트를 목표로 손쉽게 모델을 구축 및 조정하고 다양한 시나리오를 동시에 실행할 수 있습니다. 이렇게 조정된 모델은 보다 정확한 결과를 산출하기 때문에 분석 전문가들은 가상의 질문을 던지고 적절한 답을 들을 수 있습니다. 한편 적시에 결과를 도출할 수 있도록 자동 생성된 스코어 코드를 통해 이러한 결과를 적용할 수 있습니다.
- **사용자가 원하는 프로그래밍 언어 선택** Python, Java, R, Lua 프로그래머들은 SAS에서의 프로그래밍 방법을 배우지 않고도 SAS Visual Statistics를 사용할 수 있습니다. 따라서 다른 코딩 환경을 사용하는 사용자도 테스트를 거쳐 믿을 수 있는 SAS 머신러닝 및 통계 알고리즘에 유연하게 액세스할 수 있습니다.
- **최고의 확장성** 분산형 인메모리 프로세싱 플랫폼은 데이터 요구와 사용자 수가 증가하고 보다 복잡한 문제를 다루야 하는 경우에도 손쉽게 확장할 수 있습니다. 분산 환경에 내장된 파일오버 관리 기능은 제출된 작업이 항상 완료되도록 보장하고 필요하면 언제든지 컴퓨팅 리소스를 사용할 수 있도록 보장합니다.

## 주요 기능

SAS® Visual Statistics on SAS® Viya™는 시각적 인터페이스로 데이터의 규모에 관계 없이 탐색 및 예측 모델을 신속하게 생성할 수 있도록 지원합니다. 또한 SAS나 다른 언어로 코딩하고 싶어하는 사용자에게는 인터랙티브 방식의 프로그래밍 인터페이스를 제공하는 한편, 강력한 SAS 통계 모델링 및 머신러닝 기법을 제공합니다. 이와 같은 분석 모델링 기법들을 사용하는 사용자는 결과를 예측함으로써 목표에 더욱 부합하는 효과적인 조치를 취할 수 있습니다.

SAS Viya는 개방형 인메모리 플랫폼으로, 데이터 탐색 및 분석을 위한 초고속 프로세싱을 제공합니다. SAS Viya는 데이터 및 분석 워크로드 작업을 단일 서버의 코어 또는 대규모 컴퓨팅 클러스터 노드로 자동 분산시켜 병렬 처리할 뿐 아니라 단일한 확장형 통합 환경을 제공하여 관리, 유지 및 거버넌스를 손쉽게 수행할 수 있도록 지원합니다.

## 시각적 데이터 탐색

SAS Visual Statistics를 SAS Visual Analytics에 추가하면 동일한 인터랙티브 방식의 데이터 탐색 및 예측 모델 인터페이스를 공유할 수 있습니다. 즉, 통합 프로세스로 데이터를 분석하여 인사이트로 전환할 수 있습니다. 포인트-앤-클릭 인터페이스와 프로그래밍 인터페이스는 모두 다양한 설명 변수 중에서 예측 인자를 손쉽게 식별하고 이상값과 데이터 불일치를 시각적으로 표시하고 파악할 수 있도록 지원합니다. 시각적 데이터 탐색을 통해 훨씬 손쉽게 데이터의 관계를 이해하고 새로운 변수를 도출하며 관련된 변수를 선택하여 모델 개발 프로젝트를 개선할 수 있습니다. 또한 모델 반영에 적합한 변수와 세분화 전략을 가장 잘 정의하는 변수가 무엇인지 확인할 수 있습니다.

모델 구축과 시각적 데이터 탐색이 하나로 통합되어 있기 때문에 워크플로우를 중단 없이 유지하는 동시에 가설을 신속하게 검증할 수 있어 모델링의 신뢰도, 생산성, 정확도를 높일 수 있습니다.

- 대규모 데이터 세트에서 모델링 결과에 영향을 미치는 복잡한 관계나 중요 변수를 신속하게 해석
- 관찰 결과를 필터링하여 변수가 전반적인 모델 리프트에 미치는 영향 이해
- 이상값 및 IP(Influence Point) 탐지 및 포착하고 다운스트림 분석(예: 모델)에서 제거 지원
- 막대 차트, 히스토그램, 박스 플롯, 열 지도, 버블 도표, 지형도 등을 이용하여 데이터 탐색
- 다른 모델링 또는 시각화 작업에서 직접 사용할 수 있는 예측 결과나 세분화 결과 도출 및 모델 구축과 무관한 사용자에게도 공유 가능

## 예측 모델링 기법에 대한 GUI 기반의 액세스

시각적 웹 브라우저 인터페이스를 사용하면 일반 데이터 사이언티스트도 간단한 드래그-앤-드롭 방식으로 강력한 탐색 및 예측 모델을 테스트하고 개발할 수 있습니다. 또한 모델링 프로젝트가 밀려 있거나 다양하게 세분화된 세그먼트를 대상으로 테스트해야 할 경우 프로그래머 및 데이터 사이언티스트 역시 시각적 GUI를 활용해 모델을 신속하게 개발할 수 있습니다.

시각적 인터페이스는 선형 회귀 분석, 로지스틱 회귀 분석, 일반화 선형 회귀 분석 모델, 2개의 차원 축소 기법(군집 분석 및 의사결정 트리)에 포인트-앤-클릭 방식으로 손쉽게 액세스할 수 있도록 지원합니다.

한편 복잡한 대규모 데이터를 처리할 때는 차원 축소 기법을 통해 모델의 정확도를 높일 수 있고 K-평균 군집 분석, 산점도 및 상세 요약 통계를 이용해 세그먼트를 탐색하고 평가해 추가 분석을 수행할 수 있으며 범주화 및 회귀 분석을 위해 의사결정 트리를 구축할 수도 있습니다. 의사결정 트리를 생성하고 나면 인터랙티브 방식으로 트리를 제거하고 서브-트리를 구축할 수 있습니다.

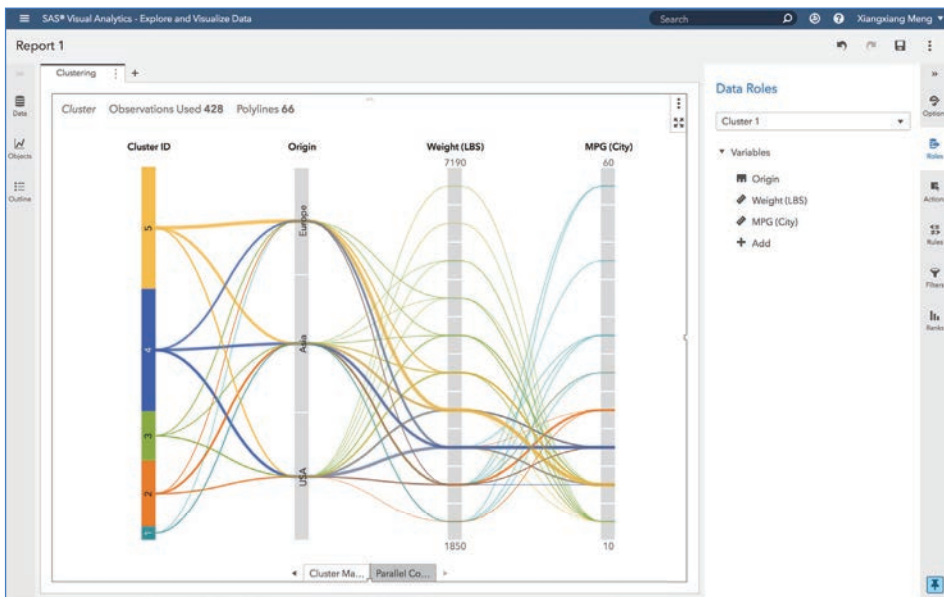


그림 1: 그림 1: 포인트-앤-클릭 방식의 시각적 인터페이스를 사용해 클러스터를 구축 및 조정

SAS Visual Statistics 사용자들은 자신의 모델을 Model Studio로 복사하여 SAS Visual Data Mining and Machine Learning에서 데이터 마이닝 또는 머신 러닝 프로세스를 계속 수행할 수 있습니다.

• 클러스터링:

- K-평균 군집화
- 군집들을 인터랙티브 방식으로 평가할 수 있는 평형 좌표 그림
- 소규모 데이터 세트를 위한 군집 프로파일이 오버레이된 입력 산점도와 대규모 데이터 세트를 위한 군집 프로파일이 오버레이된 열 지도
- 상세 요약 통계 (각 군집의 평균, 관찰 횟수 등)
- 필요 시 군집 ID를 새로운 열로 생성
- 모델 평가를 위한 홀드아웃 데이터(교육 및 검증용) 지원

• 의사결정 트리:

- 변수 중요성의 측도 계산
- 범주 및 회귀 트리 지원
- 수정된 C4.5 알고리즘이나 비용 복잡도 제거 기반
- 인터랙티브 방식의 트리 육성 및 제거 및 인터랙티브 방식의 서브트리 육성
- 트리 깊이, 최대 가지 수, 리프(leaf) 크기, 트리 제거 정도 등을 설정
- 트리 맵 디스플레이를 이용해 인터랙티브 방식으로 트리 구조 탐색
- 필요 시 리프 ID를 새로운 열로 생성
- 모델 평가를 위한 홀드아웃 데이터(교육 및 검증용) 지원

• 선형 회귀 분석:

- 정규, 이진, 감마, 포아송, 트위디(Tweedie), 역 가우시안, 음이항을 비롯한 다양한 분포 지원
- 전진, 후진, 단계 및 lasso 변수 선택 지원

- 영향력 통계
- 변수 선택
- 빈도 및 가중치 변수
- 잔차 진단
- 모든 ANOVA, 모델 차원, 적합도 통계, 모델 ANOVA, Type III 테스트, 모수 추정치가 포함되어 있는 요약 테이블
- 필요 시 예측 값과 잔차를 새로운 열로 생성
- 모델 평가를 위한 홀드아웃 데이터(교육 및 검증용) 지원
- 일반화 가산 모델:
  - 1/2차원 스플라인(spline) 효과 지원
  - Smoothing effect를 위한 GCV, GACV 및 UBRE 방법
  - 보정 변수 지원
- 빈도 및 가중치 변수
- 잔차 진단
- 모델 요약, 반복 히스토리, 적합도 통계 및 파라미터 추정이 포함된 요약 테이블
- 모델 평가를 위한 홀드아웃 데이터(교육 및 검증) 지원
- 비모수 로지스틱 회귀분석:
  - logit, probit, log-log 및 c-log-log 연결 함수가 적용된 이진 데이터 모델
  - 1/2차원 스플라인 효과 지원
  - Smoothing effect를 위한 GCV, GACV 및 UBRE 방식
  - 보정 변수 지원
- 빈도 및 가중치 변수
- 잔차 진단
- 모델 요약, 반복 히스토리, 적합도 통계 및 파라미터 추정이 포함된 요약 테이블
- 모델 평가를 위한 홀드아웃 데이터(교육 및 검증용) 지원

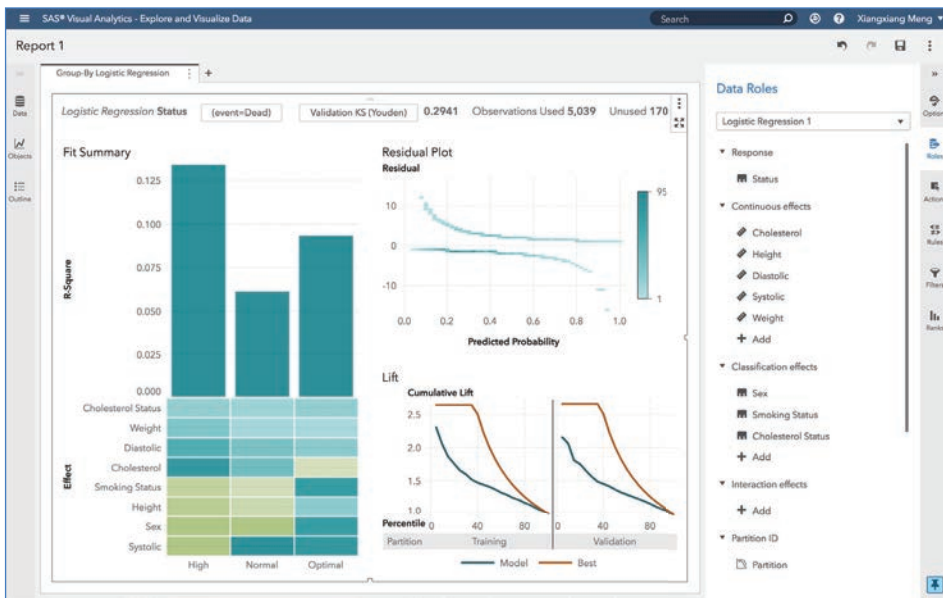


그림 2: 시각적 인터페이스를 사용하여 로지스틱 회귀 분석 모델을 신속하게 구축 및 조정할 수 있습니다. 프로그래밍 인터페이스에서는 보다 고급 기능이 제공됩니다.

- 로지스틱 회귀 분석:
  - 로짓(logit) 및 프로빗(probit) 연결 함수용 이진 데이터를 위한 모델
  - 영향력 통계
  - 변수 선택
  - 빈도 및 가중치 변수
  - 잔차 진단
  - 모델 차원, 반복 히스토리, 적합도 통계, 수렴 상태, Type III 테스트, 모수 추정, 응답 프로파일 포함 요약 테이블
  - 필요 시 예측 라벨과 예측 이벤트 확률을 새로운 열로 생성 예측 컷오프를 조정하여 관찰 내용을 이벤트나 비이벤트로 구분하여 라벨링
  - 모델 평가를 위한 홀드아웃 데이터(교육 및 검증용) 지원
- 일반화 선형 모델:
  - 베타, 정규, 이진, 지수, 감마, 기하 분포, 푸아송, 역 가우시안, 음이항을 비롯한 다양한 분포 지원
  - 보정 변수 지원
  - 빈도 및 가중치 변수
  - 잔차 진단
  - 모델 요약, 반복 히스토리, 적합도 통계, Type III 테스트 테이블, 모수 추정이 포함된 요약 테이블
  - 예측 변수에 대한 결측 값을 처리하는 결측 옵션
  - 필요 시 예측 값과 잔차를 새로운 열로 생성
  - 모델 평가를 위한 홀드아웃 데이터(교육 및 검증용) 지원

Python, R, Lua, Java 같은 다른 언어를 사용하는 프로그램에서 호출할 수 있습니다.

뿐만 아니라 공용 REST API는 코더가 기존 애플리케이션에 입증된 SAS Analytics를 추가할 수 있도록 지원합니다. 따라서 SAS를 익숙하게 사용하지 못하는 프로그래머들도 SAS를 손쉽게 사용할 수 있습니다. SAS Viya 기반의 SAS Visual Statistics는 새로운 분산형 인메모리 환경에서 실행되고, 효과적인 SAS 알고리즘에 프로그래밍 방식으로 액세스하여 여러 통계 분석 기능들을 지원합니다. 이와 같이 유연한 개방형 프로그래밍 환경을 통해 데이터 사이언티스트와 통계 전문가들은 프로그래밍 능력과 기호에 따라 데이터 조작 및 고급 분석 기능을 쉽게 사용할 수 있습니다.

- SAS PROCs 및 기타 작업을 이용해 SAS Studio에서 SAS Viya (CAS 서버)에 액세스
- PROC CAS을 이용해 CAS 작업을 수행하거나 Python, R, Lua, Java 등 프로그래밍 환경
- 공용 REST API를 사용해 자체 애플리케이션에서 SAS Viya (CAS 서버)에 액세스
- Python Pandas DataFrames에 기본 통합 기능 제공: Python 프로그래머는 CAS에 DataFrames 업로드 및 CAS에서 DataFrames 형태의 결과를 다른 Python 패키지(Pandas, matplotlib, Plotly, Bokeh 등)에 적용 가능
- SAS/STAT® 프로시저 및 SAS/GRAPH® 지원
- 주성분 분석(PCA):
  - 주성분을 계산하여 차원 축소 수행
  - 고유 값 분해, NIPALS 및 ITERGS 알고리즘 제공
  - 관찰 내용 전반에 대한 주요 구성요소 스코어 출력
  - 화면 플롯 및 패턴 프로파일 플롯 생성
- 의사결정 트리:
  - 범주 및 회귀 트리 지원
  - 범주 및 수치 기능 지원

### 개방형 코드 기반 모델 개발

SAS Visual Statistics의 시각적 GUI는 강력하고 효과적이지만, 많은 통계 전문가, 데이터 사이언티스트, 분석가들은 예측 모델을 자체적으로 코딩해서 더 많은 옵션을 활용해 이를 세부 조정하는 방법을 선호합니다. SAS Viya 기반의 SAS Visual Statistics에서 실행되고 있는 분석 기능들은 SAS Studio 프로그래밍 인터페이스에서 프로그래밍 방식으로 액세스하거나,

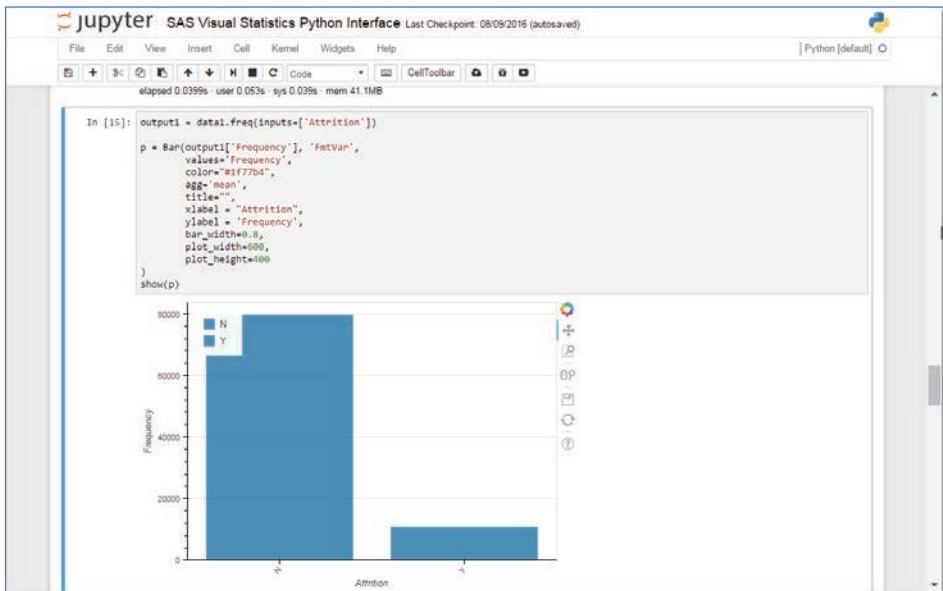


그림 3: 그림 3: Python을 비롯한 오픈 소스 언어를 사용하여 탐색 및 예측 모델 구축

- 불순물 측도 및 통계 테스트로 노드 분할을 위한 기준 제공
- 트리 제거 시 비용 복잡성 및 오류 감소 방법 제공
- 교육, 검증, 테스트 역할로 데이터 파티셔닝 지원
- 최상의 서브트리를 선택할 수 있는 검증 기능 지원
- 최종 트리 모델 평가를 위한 테스트 데이터 지원
- 대리 규칙 포함 결측 값 처리 방법 제공
- 트리 다이어그램 생성
- 모델 기반(재치환) 통계 및 모델 적합도 평가
- 변수 중요성의 측도 계산
- 관찰을 위한 리프 할당 및 예측 값 출력
- 클러스터링:
  - 연속형 변수 군집화를 위한 K-평균 알고리즘 제공
  - 명목형 변수 군집화를 위한 K-모드 알고리즘 제공
  - 유사성을 위한 다양한 거리 측도 제공
  - 군집 수 예측을 위한 정렬 상자 기준 방법 제공
  - 관찰 내용 전반에서 군집 소속 및 거리 측도 출력
- 선형 회귀 분석:
  - 연속형 및 범주형 변수에서 선형 모델 지원
  - 범주 효과를 위한 다양한 모수화 지원
  - 모든 수준의 상호 작용 및 중첩 효과 지원
  - 다항식과 스플라인 효과 지원
  - 전진, 후진, 단계, LARS(Least Angle Regression), lasso 선택 방법 지원
  - 모델 선택을 제어할 수 있도록 정보 기준 및 검증 방법 지원
  - 범주 효과의 각 수준을 선택할 수 있는 옵션 제공
  - 효과들 간의 계층 구조 보존
  - 교육, 검증, 테스트 역할로 데이터를 파티셔닝할 수 있도록 지원
  - 다양한 기술 통계 제공
  - 운영 환경에서의 스코어링을 위한 SAS 코드 생성
- 로지스틱 회귀 분석:

- 이진 및 이항 응답 지원
- 범주 효과를 위한 다양한 모수화 지원
- 모든 수준의 상호 작용 및 중첩 효과 지원
- 다항식과 스플라인 효과 지원
- 전진, 후진, 고속 후진, lasso 선택 방법 지원
- 모델 선택을 제어할 수 있도록 정보 기준 및 검증 방법 지원
- 범주 효과의 각 수준을 선택할 수 있는 옵션 제공
- 효과들 간의 계층 구조 보존
- 교육, 검증, 테스트 역할로 데이터를 파티셔닝할 수 있도록 지원
- 모델 평가를 위한 다양한 통계 제공
- 최우 추정법을 위한 다양한 최적화 방법 제공
- 일반 선형 모델:
  - 이진, 정규, 포아송, 감마 등 다양한 분포 지원
  - 분류 효과를 위해 다양한 매개 변수로 변환하는 기능 지원
  - 모든 수준의 인터랙션 및 중첩 효과 지원
  - 다항식 및 스플라인 효과 지원
  - 전진, 후진, 고속 후진, 단계, 그룹 lasso 선택 방법 지원
  - 모델 선택할 수 있도록 평가 기준 및 검증 방법 지원
  - 분류 효과를 각 개별로 선택할 수 있는 옵션 제공
  - 효과의 계층 구조 유지
  - 데이터를 교육, 검증 및 테스트 역할로 분할하도록 지원
  - 모델 평가를 위한 다양한 통계 제공
  - 최우 추정법을 위한 다양한 최적화 방법 제공
- 비선형 회귀분석 모델:
  - 표준 또는 일반 분포에 따른 비선형 회귀분석 모델 적합화
  - 더욱 정확한 모수 추정을 위해 사용자 입력 표현식의 분석 도함수 계산
  - ESTIMATE 및 PREDICT 문(프로시저에 한함)을 사용하여 사용자 입력 표현식을 평가

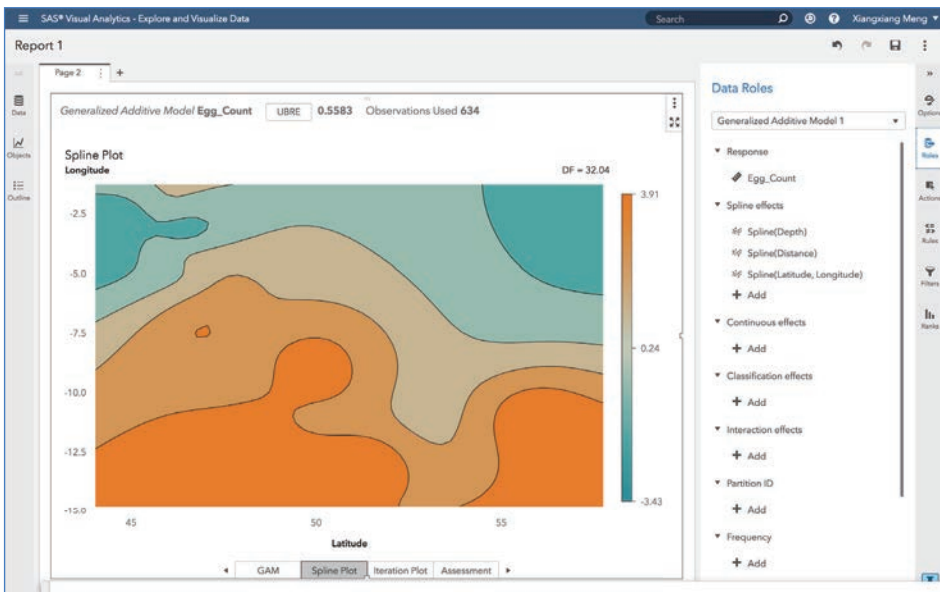


그림 4: 일반화 가산 모델이 SAS Visual Statistics에 추가되었습니다. 시각적 인터페이스를 사용하거나 코드를 통해 일반화 가산 모델을 구축하십시오.

- PROC NLMOD를 사용하지 않는 경우 CMP 항목 저장소가 포함된 데이터 테이블이 필요
- 최소 제곱법을 사용한 모수 추정
- 최우 기법을 사용한 모수 추정
- 분위수 회귀 모델:
  - 단일 또는 다중 분위 수준에 대해 분위수 회귀 분석 지원
  - 범주 효과를 위한 다양한 모수화 지원
  - 모든 수준의 상호 작용(교차 효과) 및 중첩 효과 지원
  - 효과들 간 계층적 모델 선택 전략 지원
  - 다양한 효과 선택 방법 제공
  - 다양한 선택 기준에 따른 효과 선택 지원
  - 중단 및 선택 규칙 지원
- 예측을 위한 부분적 최소 제곱법 모델:
  - 범주형 변수, 연속형 변수, 상호 작용 및 중첩에서 프로그래밍 구문 제공
  - 다항식 및 스플라인 효과를 위한 효과 구축 구문 제공
  - 교육, 검증 역할로 데이터를 파티셔닝할 수 있도록 지원
  - 추출된 요인의 수를 선택할 수 있도록 테스트 세트 검증 제공
  - 주요 구성요소 회귀 분석, 감소된 순위 회귀 분석, 부분적 최소 제곱법 회귀 분석 등의 방법 실행
- 기술 통계
  - 카디널리티(cardinality) 이해를 위한 고유 수치
  - 1개 이상의 변수에 대한 이상값을 포함해 중심성 및 확산성을 평가하기 위한 박스 플롯
  - 변수 세트에서 피어슨 상관 계수를 측정하기 위한 상관관계
  - 가중치 지원을 포함한 교차표
  - 연관성 측도를 포함한 통계 분할표
- 비닝(binning) 값, 최대 값 임계치, 특이치 등을 제어하기 위한 옵션이 있는 히스토그램

- 데이터 단일 패스 시 다차원적 요약
- 1개 이상 변수에 대한 백분위
- 관찰 회수, 결측 값 개수, 비결측 값 합계, 평균, 표준 편차, 표준 오차, 보정 및 미보정 제곱합, 최소값 및 최대값, 변동 계수 같은 요약 통계
- 정상, 트라이큐브(tri-cube) 및 2차 커널 함수를 이용한 커널 밀도 추정값
- 일반화 가산 모델:
  - 저차원 회귀분석 스플라인에 따른 일반화 가법 모델 적합화
  - Penalized likelihood estimation을 사용한 회귀분석 파라미터 추정
  - 성과 반복 기법 또는 외부 반복 기법을 사용한 평균화된 모수 추정
  - 최우 기법을 사용한 회귀분석 모수 추정
  - 왈드(Wald) 통계량에 따른 각 스플라인 항의 총 기여도 검증
  - 분류 변수, 연속 변수, 상호작용 및 중첩 효과가 포함되는 모델 개발 구문 제공
  - 다양한 변수를 사용한 스플라인 항 구성 가능
- 비례 위험 회귀분석:
  - 생존 데이터에 대한 Cox 비례 위험 회귀분석 모델 적합화 및 변수 선택
  - 모델 개발 구문에 분류 변수, 연속형 변수, 인터랙션 및 중첩 효과 제공
  - 다항식 및 스플라인 효과를 위한 효과 구성 구문 제공
  - 최우 부분 추정, 층화 분석 및 변수 선택 실행
  - 데이터를 교육, 검증 및 테스트 역할로 분할
  - 가중치 분석 및 그룹 분석 제공
- 통계적 공정 관리:
  - 슈하트(Shewhart) 관리도 분석 실행

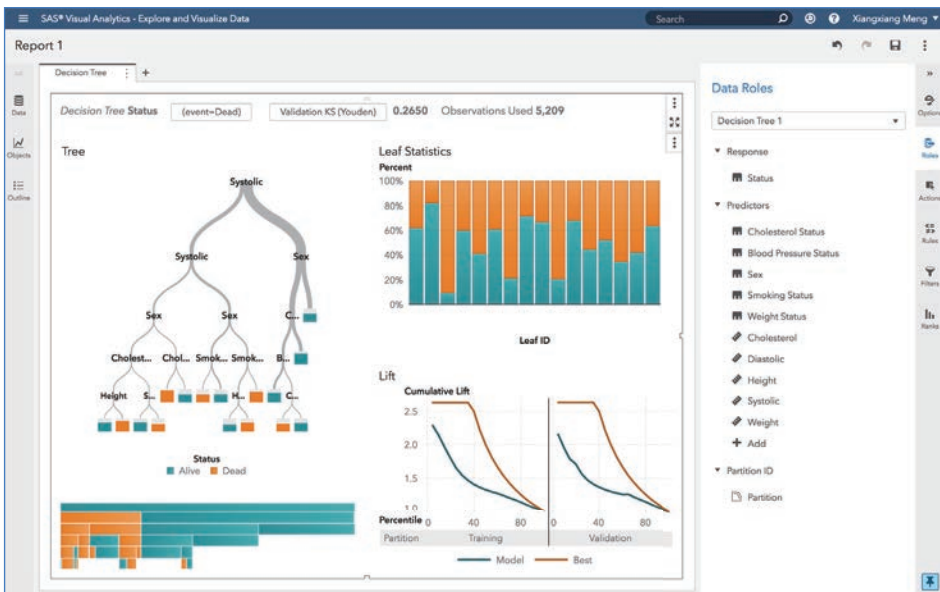


그림 5: 의사결정 트리는 가장 가능성 있는 결과를 시각적으로 보여줍니다. 시각적 인터페이스를 사용하거나 코드를 통해 의사결정 트리를 구축하십시오.

- 통계에서 벗어나는 과정을 구분하기 위한 다수의 공정 변수 분석
- 서로 크기가 다른 서브그룹 분석을 위한 관리 한계값(control limit) 조정
- 데이터에서 관리 한계값 추정, 모수의 특정 값(알려진 표준 값)에서 관리 한계값 계산 또는 입력 데이터 테이블에서 임계값 판독
- 실행 패턴에 따른 이상 원인 검정 실행(Western Electric 규칙)
- 다양한 기법을 사용한 공정 표준 편차 추정(변수 관리도에 한함)
- 관리도 통계 및 관리 임계값을 출력 데이터 테이블에 저장

## 그룹화 처리

SAS Visual Statistics를 사용하면 다수의 사용자가 다양한 모델을 동시에 구축할 수 있을 뿐 아니라 데이터를 매번 정렬 또는 인덱싱하지 않고도 각 그룹 또는 세그먼트에 맞게 결과를 처리할 수 있습니다. 그룹화 변수 나 그 속성은 작업에 따라 달라질 수 있는데, 데이터 서플링 또는 재배열을 하지 않고도 그룹을 처리할 수 있습니다.

따라서 추가적인 처리 부담 없이 더 많은 세그먼트나 그룹에 대해 더 많은 모델을 신속하게 생성할 수 있으므로 결과적으로 각 세그먼트나 그룹 고유의 요구를 충족하는 모델을 생성할 수 있습니다.

- 데이터 정렬 또는 인덱싱 없이 각 세그먼트에 대해 신속하게 모델 구축 및 결과 계산 처리
- 의사결정 트리나 클러스터링 분석을 통해 세그먼트 기반 모델을 즉시 구축(충별 모델링)

## 모델 비교 및 평가

모델이 생성되고 나면 시각적 인터페이스나 프로그래밍 인터페이스에서 리프트 차트, ROC 차트, 일치율 통계, 1개 이상의 모델에 대한 오분류 테

이블 같이 다양한 통계 정보를 사용하여 모델을 손쉽게 비교 및 평가할 수 있습니다.

한편 시각적 인터페이스에서 인터랙티브 방식의 슬라이더를 활용한 임계값 조작을 통해 서로 다른 백분위에서 리프트를 시각적으로 손쉽게 평가할 수 있을 뿐 아니라 모델 적합(fitting)과 모델 진단을 하나로 결합하여 성능에 미치는 영향을 신속하게 파악하고 이해할 수 있습니다.

모델 평가 기능을 사용하면 최상의 리프트와 ROI를 산출하는 모델을 식별할 수 있습니다.

## 모델 스코어링

최종 결정된 모델을 새로운 데이터에 손 쉽게 적용할 수 있습니다. 정교한 모델의 모든 데이터 준비 작업과 이를 알고리즘을 개발 환경에서 운영 시스템으로 이동시키는 일은 예측 모델링 및 머신러닝에서 어려운 작업입니다. 그러나 SAS Visual Statistics를 사용하면 SAS DATA 단계 코드로 모델을 내보내서 새로운 데이터에 손 쉽게 적용할 수 있습니다. 예측 모델을 운영 환경에 적용하여 더 나은 결정을 내리고 최적의 조치를 취하는데 필요한 인사이트를 확보할 수 있습니다.

## SAS® Viya™를 통한 분산형 인메모리 분석 프로세싱

SAS Visual Statistics는 인메모리 분석 프로세싱을 사용해 확장형 단일 거버넌스 환경에서 다양한 비즈니스 문제를 해결하는 혁신적인 고성능 플랫폼인 SAS Viya에서 실행됩니다. SAS Viya는 CAS (Cloud Analytic Services) 서버라는 새로운 인메모리 엔진을 채용하고 있습니다. CAS 엔진은 인메모리 프로세싱을 수행하고 처리를 클러스터의 노드 전반으로 분산할 수 있습니다. 매우 빠른 분산형 병렬 처리 기술을 활용하기 때문에 데이터 탐색 및 모델 개발 시간을 대폭 줄일 수 있습니다. 뛰어난 가용성과 확장형 프로세싱을 제공하는 분산형 인메모리 컴퓨팅 플랫폼인 SAS Viya를 사용하는 IT 부서는 컴퓨팅 용량을 수직/수평 확장하여 더 많

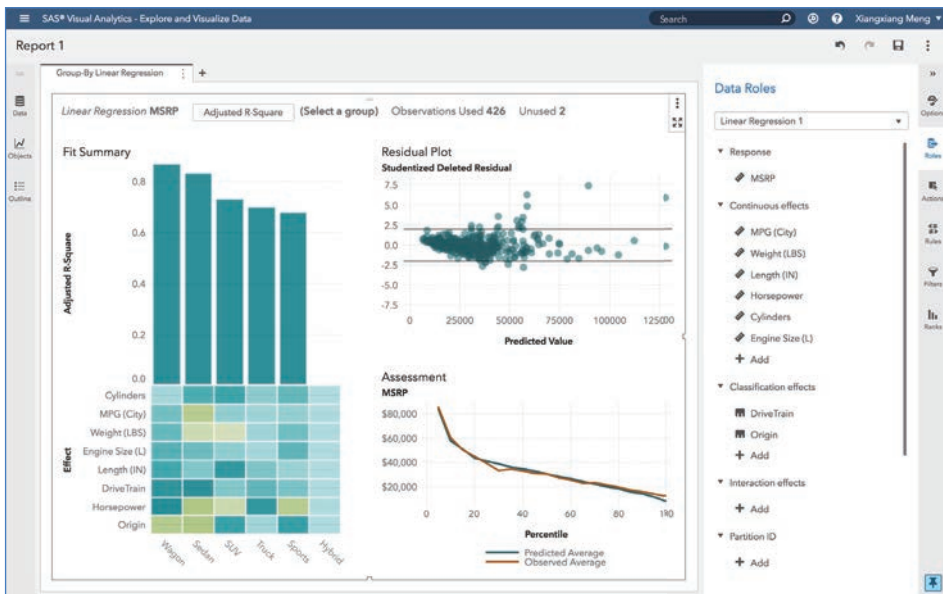


그림 6: 그림 4: 동적 그룹화 처리를 통해 데이터를 매번 정렬하거나 인덱싱하지 않고도 각 그룹, 파티션 또는 세그먼트에 대한 결과 계산

은 사용자의 요구를 충족시키고 증가하는 데이터와 점차 복잡해지는 분석 문제를 처리할 수 있습니다.

사용자를 위한 강력한 분산 프로세싱을 내제한 SAS Viya는 내결함성 설계를 통해 멀티플랫폼 프로세싱 환경에서도 서버 결함을 자동으로 감지하고 필요에 따라 프로세싱을 다시 분배합니다. 한편, SAS Viya는 프로세싱 클러스터의 데이터 사본도 관리하는데, 클러스터에 속한 노드 중 하나를 사용할 수 없게 되거나 결함이 발생하는 경우, 다른 블록에서 필요한 데이터를 가져와 프로세싱을 이어갑니다. 이와 같은 자동 복구 메커니즘 덕분에 사용자는 중단 없는 프로세싱 및 자동 복구를 통한 고가용성을 누릴 수 있습니다.

### 그룹별 프로세싱

- 데이터를 매번 정렬하거나 인덱싱할 필요 없이 각 그룹 또는 세그먼트마다 모델을 개발하여 결과 계산 및 처리
- 의사결정트리나 군집 분석을 통해 세그먼트 기반 모델을 즉시 구축(증화 모델링)

### 모델 비교, 평가 및 스코어링

- 1개 이상의 모델에 대해 리프트 차트, ROC 차트, 일치도 통계 및 오분류 테이블 등 모델 비교/요약 작성
- 평가 통계 및 분류 테이블의 자동 업데이트를 위한 대화형 방식의 예측 한계 감소
- 각 백분위별로 대화형 방식의 리프트 평가
- 모델을 SAS DATA 스텝 코드로 내보내서 다른 애플리케이션과 통합. 모델이 다른 모델에서 나온 출력 데이터(리프트 ID, 군집 ID 등)를 사용하는 경우 스코어 코드를 자동으로 연결

### SAS® Viya® 인-메모리 런타임 엔진

- 인-메모리 서버인 CAS를 통해 메모리 내에서 프로세싱을 실행하고 처리 값을 군집의 각 노드로 분산
- 분산 환경에서 처리에 필요한 변수를 사용해 사용자 요청(프로시저 언어로 표현됨)을 작업으로 변환. 사용자의 추가 작업을 위해 결과 집합 및 메시지가 프로시저로 다시 전달
- 데이터를 블록으로 관리하고 필요에 따라 메모리에 로드. 테이블의 메모리 용량이 초과하는 경우 서버가 블록을 디스크에 캐싱
- 데이터 및 중간 결과는 사용사의 사용 범위 안에서 필요한 시간 동안 메모리에 저장
- 알고리즘이 지정된 작업에 가장 적합한 노드 수를 결정
- 통신층(communication layer)이 고장 방지 기능을 지원. 실행 중인 상태에서도 서버에서 노드를 제거하거나 추가 가능. 고가용성을 위해 모든 아키텍처 구성요소의 복제 가능
- 제품을 멀티테넌트(multitenant) 모드로 배포하여 공유 소프트웨어 패키지 안전하게 격리된 테넌트(tenant)를 지원

### 메모리 내 데이터에 동시 액세스

SAS Viya 플랫폼을 사용하면 필요 시 메모리 내 데이터를 로드해서 영구적으로 유지하고 멀티패스 분석 계산을 수행할 수 있습니다. 또한 인터랙티브 방식의 시각적 탐색이나 고급 분석 프로세싱 등 관계없이 모든 데이터, 테이블, 개체를 필요한 기간동안 메모리에 유지할 수 있습니다. 따라서 많은 사용자들이 협업을 통해 동일한 데이터를 탐색하고 동시에 모델을 구축할 수 있습니다. 인메모리 프로세싱은 반복 단계 또는 요청 동안 불필요하고 비용 손실이 큰 데이터 셔플링을 제거합니다. 따라서 모델 변경(예: 새로운 변수 추가, 이상값 제거 등)으로 인한 결과를 즉시 눈으로 확인할 수 있습니다.

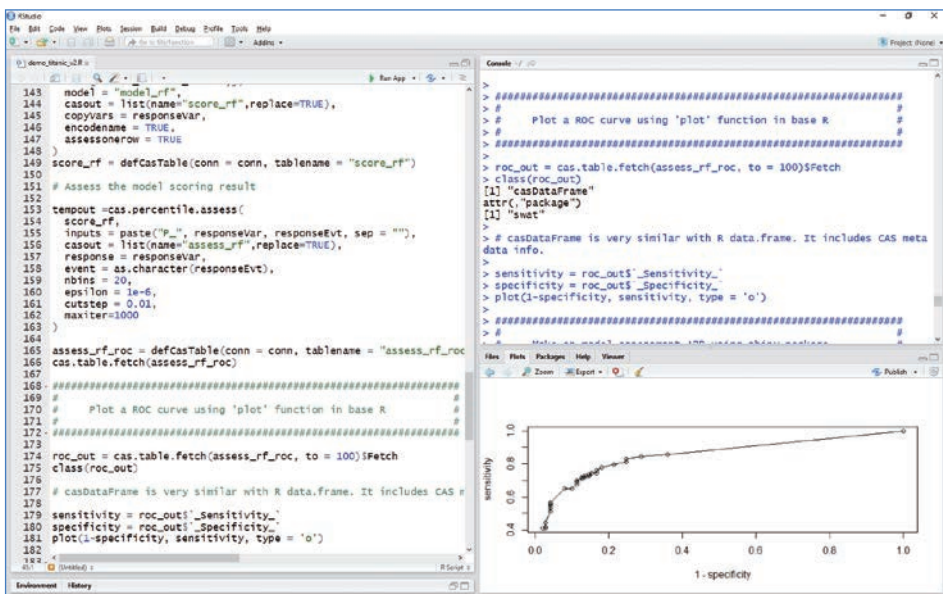


그림 7: 기타 오픈 소스 프로그래밍 언어는 물론, R을 사용해서 SAS Visual Statistics에서 분석 절차 및 기능을 호출할 수 있습니다.



## 유연한 배포 옵션

SAS Visual Statistics는 데이터 규모, 워크로드 유형, 성능 요구사항이 서로 다른 조직들을 위해 다양한 배포 옵션을 제공합니다. 부서 워크그룹이나 중소기업 위한 단일 머신 환경부터 수 백 개의 노드가 포함되어 있는 클러스터를 운영하는 대기업을 위한 대규모 분산형 시스템에 이르기까지 다양한 규모로 배포할 수 있습니다. 또한 사설 및 공용 클라우드 인프라에서 실행할 수 있도록 설계되어 있습니다.

- 온프레미스(On-premise) 배포:
  - 중소기업의 요구를 지원하는 단일 시스템 모드
  - 증가하는 데이터, 워크로드, 확장성 요구사항을 충족하는 분산형 모드
- 클라우드 배포:
  - 엔터프라이즈 호스팅, 사설/공용 클라우드 (예: Amazon의 BYOL) 인프라

더 자세한 내용은 [sas.com/korea/vs](https://sas.com/korea/vs)에서 확인하실 수 있습니다.

