

Loving your Life

"Be the L.BA"

제3회 L.POINT Big Data Competition

데이터 설명서



1. 대회취지

- 데이터 분석을 통해 새롭고 차별화된 가치를 창출할 수 있는 Data Analytics Specialist 발굴
- 빅데이터 분석에 관심있는 도전자에게 다양한 분석 기회 제공

2. 대회일정

- 1) 참가신청 : 2016. 11. 21(월) ~ 2016. 12. 16(금)
- 2) 데이터 다운로드 : 2016. 12. 19(월) ~ 2017. 2. 3(금) 15:00 까지
 - Competition 공식홈페이지 > 분석 Data 다운로드 (팀장만 다운로드 가능)
- 3) 결과물 제출 마감 : 2017. 2. 3(금) 15:00 까지
 - Competition 공식홈페이지 > 접수 및 확인 > 분석결과물 제출
- 4) 본선 진출자 발표 : 2017. 2. 17(금), Competition 공식홈페이지 '공지사항' 게시
- 5) 본선 PT 심사 및 시상 : 2017. 2. 24(목)

3. 결과물 제출

- 분석 보고서와 분석 Data/Code/Node 등을 Zip파일로 압축하여 홈페이지 내 '분석결과물 제출' 메뉴에서 업로드(총 50MB미만)

1) 1차 심사 제출물

- ① 분석보고서 : PPT 파일, PDF 변환 파일 제출 (분량 제한 없음)
- ② 분석에 사용한 Data, Code 및 Node 제출(SAS, SQL, R)
- ③ 모델 검증 시트 : 데이터와 함께 첨부된 모델 검증 시트를 작성하여 제출

2) 2차 심사 제출물

- ④ 본선발표 PPT : 본선 진출자에 한함. (15분 발표 분량.)
대상 및 제출 시기는 추후 별도 공지
- ⑤ 재학 및 졸업 증명서 : 참가 자격의 확인을 위하여 본선 진출자에 한하여 제출.
대상 및 제출 시기는 추후 별도 공지

※ R을 사용할 경우 R코드 전체(R package 및 R Function포함)를 제출해야 합니다.

※ 주최측이 제공한 데이터 외에도, 모델링에 필요한 '외부데이터'를 추가적으로 분석에 활용하여도 무방합니다. 단, 데이터 출처를 명시하고 적용한 외부데이터도 함께 제출해 주시기 바랍니다.

4. 결과물 제출 유의사항

- 1) 상기 명시된 제출물 중 어느 하나라도 제출하지 않은 경우 심사대상에서 제외됩니다.
- 2) 대회 공식 프로그램은 SAS, SQL, R 입니다. 이외에 다른 프로그램 사용시 심사대상에서 제외됩니다.
- 3) 데이터 시각화에 사용되는 제품은 어떠한 제품이라도 무방합니다. 단, 상업용 제품인 경우에는 개인적으로 구매한 제품에 대해서만 허용합니다. (상업용 제품의 무단사용으로 인한 문제 발생시 L.POINT에서는 어떠한 책임도 지지 않습니다.)

5. 평가 및 시상

1) 1차 서면평가

- 참가자가 제출한 분석 보고서를 하기의 기준에 의하여 평가 (배점 100점 만점)
- 분석 보고서 필수 포함 항목
 - (1) 데이터 전처리 과정
 - (2) 모델 개발 과정
 - (3) 추천 로직 및 활용 방안
- 평가 기준

▶ **평가 최소 기준** : 첨부된 모델 검증 시트 기준, 일정 정확도 이상 획득

자세한 평가 시트 검증 기준은 모델 검증 시트에 수록.

(정확도 산정 공식 : 추천된 상품 中 실제 구매한 고객 수 / 전체 고객 수 X 100 (%))

평가항목	평가항목상세	배점
모델 적합도에 대한 우수성	데이터전처리 - 데이터 탐색 - 데이터 클렌징 - 데이터 변환	20
	모델 개발 - 변수 선택 - 모델 개발 과정 - 추천 로직의 논리성	30
활용 변수의 다양성	외부 데이터 - 활용된 외부 데이터 종류 - 데이터 적합성	20

	파생 변수 - 파생 변수의 종류 - 생성 변수의 논리성 - 생성 변수의 활용성	10
시각화	분석 보고서의 시각화 자료 활용	20
합계		100

2) 2차 본선 심사 (배점 100점 만점)

평가항목	평가항목상세	배점
서면 평가 결과	1 차 서면 평가 점수 (40 점 만점 환산)	40
	심사위원 조정 - 프레젠테이션에 따른 1 차 평가 점수 조정	
프레젠테이션	추천 상품의 논리성 - 상품 추천 모델과 추천 상품의 논리적 연계성	20
	추천 로직의 활용 - 추천 로직의 실제 비즈니스 활용 정도	20
	명확한 표현 및 효과적인 의사전달 - 발표 자료의 구성 - 의사전달 능력 - 시각화 자료 활용을 통한 표현력	20
합계		100

6. 시상내역

구분	금상	은상	동상
상 금	300 만원	200 만원	100 만원
수 상 자	1 팀/인	2 팀/인	3 팀/인
수상특전	○	○	X

※ 수상특전 : 수상자 중 선발하여 인턴/공채(통계분석 관련 직군) 서류심사 면제

(기졸업자·졸업예정자 중 2017년 지원에 한함)

7. 규정 및 운영방안

- 제출 과제가 심사 기준에 부합하지 않을 경우, 수상작을 선정하지 않을 수 있습니다.
- 표절 및 저작권 관련 문제의 소지가 있는 작품은 심사에서 제외되며, 수상 후에도 수상이 취

소되고 모든 책임은 응모자에게 있습니다.

- 공모전에 제출된 응모작에 대한 권리는 응모자에게 있으며, 제출된 응모작을 바탕으로 주최측의 책임하에 개선·발전시킨 경우에는 응모자와의 합의에 의해 공동 소유 등 소유권을 결정할 수 있습니다.
- 공모된 응모작에 관한 내용은 타인에게 누설하지 아니하며, 정보보호에 대한 보안장치 등을 마련하여 비밀유지를 위한 조치를 취합니다.
- 수상작의 저작권을 공모전 주최측에 귀속시키고자 하는 경우, 주최측은 수상자와의 충분한 합의 과정을 거쳐 그에 대한 합당한 충분한 대가를 지급하도록 합니다. 이후 출품된 응모작을 사용함에 있어서도 그 가치에 대한 합리적인 대가를 지급하도록 합니다.
- 공모전 응모자는 응모작의 반환을 요구할 수 있으며, 이 때 주최측은 작품 반환 후 응모작과 관련된 자료 일체를 즉시 폐기합니다.
- 공모전에 응모된 아이디어와 관련하여 주최측과 아이디어 제안자 사이에 분쟁이 발생할 경우, 주최측과 아이디어 제안자는 우선적으로 대화와 협상을 통하여 분쟁을 해결하도록 최선을 다해야 합니다.
- 제세공과금은 수상자 부담입니다. (※ 수상금은 제세공과금을 제외 후 지급)
- 표절 및 저작권 관련한 부분은 관련 법규 및 규정을 준수합니다.

8. 데이터 설명

주제 : 개인화 상품 추천

1) 분석 개요 :

- 과거와 달리 요즘의 소비자들은 현명하고 합리적인 소비를 추구하고자 하는 경향이 강합니다. 이와 더불어 소비자들의 수준도 함께 높아짐에 따라 다양한 욕구를 지닌 소비자들이 존재하게 되고 과거와 같이 단순히 다른 소비자들이 많이 구매하는 소위 ‘베스트셀러’ 제품을 추천할 지라도 개개인의 욕구에 맞지 않는 제품이라면 소비자들은 외면하고 맙니다. 기존의 습관적으로 구매하던 상품 뿐만 아니라 종전에 보지 못했던 새로움을 추구하기도 합니다. 물건을 판매하는 기업이라면 누구나 다양한 욕구를 가진 소비자들을 만족시키기를 원하며 소비자들의 목소리를 들어 보고자 노력 할 것입니다. 과거에는 이러한 소비자들의 목소리를 듣기 위하여 단편적인 도구들만을 이용 했다면, 현재와 같은 Big Data 시대에는 상상만 했었던 혹은 상상도 하지 못했던 많은 데이터들을 활용 할 수 있게 됩니다. 본 Competition을 통하여 여러분들의 상상력과 창의성을 유감 없이 보여주시기를 기대합니다.

2) 분석 목적 : 고객이 선호 하거나 선호 할 것으로 예상되는 상품 추천

3) 데이터 구성

가. 분석 주제 : 고객의 2년간의 소비 활동 및 경쟁사 활동, 소비 패턴 데이터를 활용하여

1) 고객의 선호 상품 추천

2) 고객의 새로운 구매를 위한 상품 추천

※ 1)과 2) 조건 중 한 가지 조건만 만족하여도 무관하며, 모델 검증 시트는 1), 2) 중 어떠한 조건으로 추천된 상품이라도 무관함.

나. 데이터 상세 설명

- 분석 대상 : 2년간 롯데 그룹의 4개의 계열사에서 구매한 고객

※ 2014~2015년도의 구매 고객 중 일부 발체

- 제공 범위 : 롯데 그룹의 4개 계열사의 구매 이력 (계열사 비공개)

고객의 성향을 파악 할 수 있는 데이터 제공 (상세 : 데이터 목록 참고)

※ 본 데이터는 가공된 샘플 데이터로 실제 시장 데이터와 차이가 있습니다.

3) 데이터 목록

구분	데이터 항목	상세 설명
Demo.	고객번호	고객의 고유 식별번호
	성별	고객의 성별 (남성=M/여성=F)
	연령대	고객의 연령대
	거주지	거주지역 신우편번호 앞 3 자리 (서울 - 구 단위, 서울 외 지역 - 시/도 단위 변환)
구매내역 정보	제휴사	롯데 그룹 4 개 계열사 (A / B / C / D)
	영수증번호	구매내역의 고유 식별번호
	대분류코드	상품 대분류 카테고리 코드성 정보
	중분류코드	상품 중분류 카테고리 코드성 정보
	소분류코드	상품 소분류 카테고리 코드성 정보
	고객번호	고객의 고유 식별번호
	점포코드	구매가 발생한 점포 코드성 정보
	구매일자	구매가 발생한 일자 (ex. 2015 년 12 월 1 일은 '20151201'로 표기)
	구매시간	구매가 발생한 시간대 (ex. 17 시는 '17'로 표기)
	구매금액	구매한 상품의 금액

		* 복수의 상품 구매 시 구매금액의 총 합계를 표기
상품분류	제휴사	롯데 그룹 4 개 계열사 (A / B / C / D)
	대분류코드	상품 대분류 카테고리 코드성 정보
	중분류코드	상품 중분류 카테고리 코드성 정보
	소분류코드	상품 소분류 카테고리 코드성 정보
	중분류명	상품 중분류 카테고리 한글명
	소분류명	상품 소분류 카테고리 한글명
경쟁사이용	고객번호	고객의 고유 식별번호
	제휴사	롯데 그룹 4 개 계열사 (A / B / C / D)
	경쟁사	계열사별 경쟁사 코드성 정보 (A01/A02/B01/B02/C01/C02/C03/D01/D02) (ex. A 사의 경쟁사 A01, A02)
	이용년월	경쟁사 이용년월 (ex. 2015 년 12 월은 '201512'로 표기)
멤버십여부	고객번호	고객의 고유 식별번호
	멤버십명	추가혜택 이용을 위해 가입한 멤버십명 (더영 / 룩스 / 하이마트 / 다동이)
	가입년월	멤버십 가입 년월 (ex. 2015 년 12 월은 '201512'로 표기)
채널이용	고객번호	고객의 고유 식별번호
	제휴사	롯데 그룹 4 개 계열사 관련 모바일/app (A_MOBILE/APP / B_MOBILE/APP / C_MOBILE/APP / D_MOBILE/APP) 롯데 그룹 2 개 계열사 관련 온라인쇼핑몰 (B_ONLINEMALL / C_ONLINEMALL)
	이용횟수	최근 3 개월간의 모바일/APP 로그인 횟수 /최근 3 개월간의 온라인쇼핑몰 구매 횟수

9. 관련 문의

- Competition 공식 홈페이지 내 Q&A 게시판 및 사무국 이메일(competition@lotte.net)

-EOD-