

Loving your Life

*"Be the L.BA"*

## 제2회 L.POINT Big Data Competition

### 데이터 설명서 (주제1)



## 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 : 2015. 11. 16(월) ~ 2015. 12. 15(화)
- 2) 데이터 다운로드 : 2015. 12. 16(수)
  - Competition 공식홈페이지 > 분석 Data 다운로드 (팀장만 다운로드 가능)
- 3) 결과물 제출 마감 : 2016. 1. 29(금) 23:59 까지
  - Competition 공식홈페이지 > 접수및확인 > 분석결과물 제출
- 4) 본선 진출자 발표 : 2016. 2. 12(금) , Competition 공식홈페이지 '공지사항' 게시
- 5) 본선 PT 심사 : 2016. 2. 25(목)
- 6) 최우수작 통합 발표 및 시상 : 2015. 2월 말 (예정)

## 3. 결과물 제출

- 선택한 주제의 분석 보고서와 분석 Data/Code/Node 등을 Zip파일로 압축하여 홈페이지 내 '분석결과물 제출' 메뉴에서 업로드(총 50MB미만)
- ① 분석보고서 : 표지포함 분량 20장 내외 PPT 파일, PDF 변환 파일 제출
  - ② 분석에 사용한 Data, Code 및 Node 제출(SAS, SQL, R)
    - ※ R을 사용할 경우 R코드 전체(R package 및 R Function포함)를 제출해야 합니다.
    - ※ 주최측이 제공한 데이터 외에도, 모델링에 필요한 '공공데이터'를 추가적으로 분석에 활용하여도 무방합니다. 단, 데이터 출처를 명시하고 적용한 공공데이터도 함께 제출해 주시기 바랍니다.

## 4. 결과물 제출 유의사항

- 1) 상기 명시된 제출물 중 어느 하나라도 제출하지 않은 경우 심사대상에서 제외됩니다.
- 2) 선택한 주제의 분석결과물 미제출시 심사대상에서 제외됩니다.  
(예시) 주제1 선택시 주제1 제출, 주제2 선택시 주제2 제출, 주제1/ 주제2 모두 선택시 주제1/주제2 모두 제출하여야 합니다.  
주제1/주제2 모두 선택 후, 주제1만(또는 주제2만) 제출 시 심사대상에서 제외됩니다.

- 3) 대회 공식 프로그램은 SAS, SQL, R 입니다. 이외에 다른 프로그램 사용시 심사대상에서 제외됩니다.

## 5. 평가기준

### 1) 1차 서면평가

- 주제1. 모델 적합도에 대한 우수성 및 보고서/프리젠테이션 표현력
- 주제2. 분석 결과 활용도 및 내용 우수성, 보고서/프리젠테이션 표현력

### 2) 2차 본선 심사 (배점 100점 만점)

주제1		주제2	
모델 적합성 (45점)	분석 데이터 가공 통계적 적용에 따른 최적의 방법론	구조화 (30점)	분석 결과물에 대한 명확한 표현 및 도식화
독창성 (20점)	분석 방법의 다양성 및 창의성	도식화 (30점)	분석 설계에 따른 스토리라인
결과활용도 (25점)	분석 결과 해석의 적합성 및 활용성	통계지표 활용 (30점)	다양한 통계지수 활용
프레젠테이션 (10점)	명확한 표현 및 효과적인 의사전달	프레젠테이션 (10점)	효과적인 의사전달

## 6. 시상내역

구 분	주제1			주제2
	금상	은상	동상	특별상
상 금	300만원	200만원	100만원	100만원
수상자	1팀/인	1팀/인	1팀/인	1팀/인

※ 수상자 중 선발하여 인턴/공채 채용기회 부여(기졸업자·졸업예정자 중 2016년 지원에 한함)

## 7. 규정 및 운영방안

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

### 주제1 : 출시상품별 Next 트렌드 예측

1) 분석 목적 : 상품 출시 후 반응 진단 및 향후 매출 예측

2) 데이터 구성

가. 분석 주제 : 출시 후 26주간 구매내역 데이터를 기반으로 출시상품별 Next 4주간의 주별 매출액 예측

나. 데이터 상세 설명

- 분석 대상 : 3개의 카테고리(감자스낵·건강음료·유제품) 중 택1, 선택한 카테고리에 속한 상품 4개(상품A·상품B·상품C·상품D)가 예측 대상  
※ 2014~2015년도에 출시된 상품으로 구성
- 제공 기간 : 각 상품이 출시된 날로부터 26주(182일)간의 구매 내역
- 제공 범위 : 서울·경기·부산 3개 지역에 위치한 일부 점포의 구매 내역  
※ 본 데이터는 가공 데이터로 실제 시장 데이터와 차이가 있습니다.

## 3) 데이터 목록

구분	데이터 항목	상세 설명
Demo.	ID	고유 식별번호
	성별	ID 의 성별 (남성=1/여성=2/식별불가=0)
	연령	ID 의 연령 (식별불가인 연령은 '0'으로 표기)
구매내역 정보	ID	고유 식별번호
	카테고리	3 개의 카테고리 중 1 개 선택 (감자스낵/건강음료/유제품)
	상품 구분	1 개의 카테고리별 4 개의 상품으로 구성 (상품 A/상품 B/상품 C/상품 D)
	구매 지역	구매가 발생한 지역 (서울/경기/부산)
	구매 시점	출시일로부터 182 일간의 구매 시점 (최초 구매시점: 1/마지막 구매시점: 182)
	구매 시간	구매가 발생한 시간대 (ex. 19 시는 '19'로 표기)
	구매 요일	구매가 발생한 요일 (월=1/화=2/수=3/목=4/금=5/토=6/일=7)
	구매 수량	구매한 상품의 수량
	구매 금액	구매한 상품의 금액 * 구매 수량이 '2'인 경우 2 개에 해당하는 금액을 표기
	취소 여부	구매한 상품의 취소 여부 (취소함=1/취소하지 않음=0)
	평균 기온	구매가 발생한 시점의 시도별 평균 기온

## 4) 분석 요건

- 데이터 탐색, 이상값 탐지 및 처리 방법 기술
- 변수 생성, 선택, 변환 등을 통한 최적의 모델 구축 및 검증
- 예측 결과값 산출 및 결과 해석 (여러 모델 생성시 최종 모델 선정 후 선정 이유 기술)

## 9. 관련 문의

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

-EOD-