



제6회 L.POINT Big Data Competition

데이터 설명서

목차

대회취지	3
대회일정	3
결과물 제출	3
결과물 제출 유의사항	3
평가 및 심사	4
시상내역	5
규정 및 운영방안	5
데이터 설명	6
관련 문의	9

1. 대회취지

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

2. 대회일정

- 1) 참가신청 : 2019. 10. 21(월) ~ 2019. 11. 26(화)
- 2) 데이터 다운로드 : 2019. 11. 27(수) ~ 2020. 1. 8(수) 15:00 까지
 - Competition 공식홈페이지 > 분석 Data 다운로드 (팀장만 다운로드 가능)
- 3) 결과물 제출 마감 : 2020. 1. 8(수) 15:00 까지
 - Competition 공식홈페이지 > 접수 및 확인 > 분석결과물 제출
- 4) 본선 진출자 발표 : 2020. 2. 3(월), Competition 공식홈페이지 '공지사항'에 게시
- 5) 본선 PT 심사 : 2020. 2. 18(화)(예정)
- 6) 시상식 : 2020. 2. 21(금)(예정)

3. 결과물 제출

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

1) 1차 심사 제출물

- ① 분석보고서 : PPT 파일, PDF 변환 파일 2부 제출 (분량 제한 없음)
- ② 분석에 사용한 Data, Code 및 Node 제출(SAS, SQL, R, Python)

2) 2차 심사 제출물

- ① 본선발표 PPT : 본선 진출자에 한하며, 대상 및 제출 시기는 추후 별도 공지(15분 발표 예정)
- ② 재학/졸업 증명서 : 참가 자격의 확인을 위해 본선 진출자에 한하여 제출(추후 별도 공지)

※ R을 사용할 경우 R package 및 R Function 등을 함께 제출해야 합니다.

※ 주최측이 제공한 데이터 외에도, 모델링에 필요한 '외부데이터'를 추가적으로 분석에 활용하여도 무방합니다.(단, 데이터 출처 명시)

4. 결과물 제출 유의사항

- 1) 상기 명시된 제출물 중 어느 하나라도 제출하지 않은 경우 심사대상에서 제외됩니다.
- 2) 분석에 사용되는 대회 공식 프로그램은 SAS, SQL, R, Python 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 롯데멤버스에서는 어떠한 책임도 지지 않습니다.)

5. 평가 및 심사

1) 1차 서면평가

- 참가자가 제출한 분석 보고서를 하기의 기준에 의하여 평가 (배점 100점 만점)
- 분석 보고서 평가 항목 및 기준

평가항목 (총 배점)	평가항목상세	배점
고객 분석 및 알고리즘 도출 과정 (60 점)	<ul style="list-style-type: none"> · 데이터 전처리 <ul style="list-style-type: none"> - 데이터 탐색 - 데이터 클렌징 	10
	<ul style="list-style-type: none"> · 고객 분석 <ul style="list-style-type: none"> - (주제 1) 고객의 행동/소비 유형 도출 과정의 논리성 - (주제 2) 고객 성향 및 라이프스타일 파악의 논리성 - 분석기법의 적정성 - 인사이트 도출 	30
	<ul style="list-style-type: none"> · 추천 알고리즘 도출 <ul style="list-style-type: none"> - (주제 1) AI 기반 고객 맞춤 상품 추천 알고리즘 개발의 논리성 - (주제 2) AI 기반 고객 행동 콘텐츠 맞춤 추천 알고리즘 개발의 논리성 - 알고리즘의 적정성 	20
활용 변수의 다양성 (20 점)	<ul style="list-style-type: none"> · 외부 데이터 <ul style="list-style-type: none"> - 활용된 외부 데이터 종류 - 데이터 적합성 	10
	<ul style="list-style-type: none"> · 파생 변수 <ul style="list-style-type: none"> - 파생 변수의 종류 - 생성 변수의 논리성, 활용성 	10
비즈니스 전략의 창의성 및 합리성 (10 점)	<ul style="list-style-type: none"> · 비즈니스 전략 아이디어의 창의성 · 비즈니스 전략의 합리성 	10
시각화 (10 점)	<ul style="list-style-type: none"> · 분석 보고서의 시각화 자료 활용 	10
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 평가점수 조정	
프레젠테이션	· 고객 분석 및 도출된 알고리즘 - 고객 분석 및 알고리즘 도출의 논리성 - 분석 기법의 적정성	20
	· 제안 비즈니스 전략 - 제안 비즈니스 전략의 창의성 및 합리성	20
	· 명확한 표현 및 효과적인 의사전달 - 발표 자료의 구성 - 의사전달 능력 - 시각화 자료 활용을 통한 표현력	20
합계		100

6. 시상내역

구분	대 상	최우수상	우수상	장려상
상 금	300 만원	200 만원	100 만원	-
수 상 자	1 팀	2 팀	3 팀	2 팀
수상특전	○	○	○	-

※ 수상특전 : 대상, 최우수상, 우수상 수상자가 롯데멤버스 통계분석 관련 직무 지원 시 2020년 인턴 또는 신입사원 1차 서류전형 면제(기졸업자·졸업예정자 중 2020년 내 1회에 한함)

※ 본선 진출자 전원에게 소정의 포인트 지급 예정

※ 장려상 수상 시 상장 지급 예정 (단, 상금 및 수상특전 제외)

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 마련하

여 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 : Audience Targeting (부제 : 잠재고객 맞춤 콘텐츠 추천)

주제2 : AI 기반 초개인화 콘텐츠 추천 (부제 : 고객 행동/소비 컨텍스트 맞춤 콘텐츠 추천)

1) 데이터 구성

- 온라인 행동 정보
- 온/오프라인 거래 정보
- 고객 Demographic 정보
- 상품 분류 정보

2) 데이터 상세

- 분석데이터 기간 : 총 3개월('19년 7월 ~ 9월)
- 제공 범위 : 롯데그룹 온라인/오프라인 계열사의 이용 이력(계열사는 임의로 선정)
※ 본 데이터는 가공된 샘플 데이터로 실제 시장 데이터와 차이가 있습니다.
- 분석 데이터 관련 기본 개념 및 구성은 제6회 빅데이터 컴피티션 설명회 발표자료를 재구성한 “제6회 빅데이터 컴피티션 설명회 자료.pdf”를 참조해 주시기 바랍니다.

3) 데이터 목록

- 온라인 행동 정보

No.	변수명(영문)	변수명(국문)	상세설명	PK
1	CLNT_ID	클라이언트 ID	• 고객을 고유하게 식별할 수 있도록 랜덤으로 부여된 ID	Y
2	SESS_ID	세션 ID	• Web/App 에 접속 후 세션이 시작될 때 부여된 순번 ID ※ 하나의 클라이언트 ID 에 여러 개의 세션 ID 가 발급될 수 있음	Y
3	HIT_SEQ	조회일련번호	• 조회 순서를 알 수 있도록 부여된 일련번호	Y
4	ACTION_TYPE	행동유형	• 총 8 가지의 행동 유형을 구분한 코드 [0.검색 / 1.제품 목록 / 2.제품 세부정보 보기 / 3.장바구니 제품 추가 / 4.장바구니 제품 삭제 / 5.결제 시도 / 6.구매 완료 / 7.구매 환불 / 8.결제 옵션]	
5	BIZ_UNIT	업종단위	• 온라인 및 오프라인 이용처를 구분하는 단위코드	
6	SESS_DT	세션일자	• 세션일자 (YYYYMMDD 형식으로 표시)	
7	HIT_TM	조회시각	• 조회시각 (HH:MM 형식으로 표시)	
8	HIT_PSS_TM	조회경과시간	• 세션이 시작된 이후 해당 조회까지 경과한 시간(단위 : 밀리초) ※ 첫 번째 조회이면 0 으로 설정됩니다.	
9	TRANS_ID	거래 ID	• 구매 내역을 고유하게 식별할 수 있도록 랜덤으로 부여된 ID	
10	SRCH_KWD	검색 키워드	• 고객이 검색한 키워드	
11	TOT_PAG_VIEW_CT	총페이지조회건수	• 세션 내의 총 페이지(화면)뷰 수	
12	TOT_SESS_HR_V	총세션시간값	• 세션 내 총 시간(단위: 초)	
13	TRFC_SRC	유입채널	• 고객이 유입된 채널 [DIRECT / PUSH / WEBSITE / PORTAL_1 / PORTAL_2 / PORTAL_3 / unknown]	
14	DVC_CTG_NM	기기유형	• 기기 유형 [mobile_web / mobile_app / PC]	

- 거래 정보

No.	변수명(영문)	변수명(국문)	상세설명	PK
1	CLNT_ID	클라이언트 ID	• 고객을 고유하게 식별할 수 있도록 랜덤으로 부여된 ID	Y
2	TRANS_ID	거래 ID	• 구매 내역을 고유하게 식별할 수 있도록 랜덤으로 부여된 ID	Y
3	TRANS_SEQ	거래일련번호	• 해당 거래 ID의 구매 내역 내 구매 순서를 알 수 있도록 부여된 일련번호	Y
4	BIZ_UNIT	업종단위	• 온라인 및 오프라인 이용처를 구분하는 단위코드	
5	PD_C	상품소분류코드	• 상품의 소분류코드	
6	DE_DT	구매일자	• 상품을 구매한 일자 (YYYYMMDD 형식)	
7	DE_TM	구매시각	• 상품을 구매한 시각 (HHMM 형식)	
8	BUY_AM	구매금액	• 구매한 상품의 금액	
9	BUY_CT	구매수량	• 구매한 상품의 수량	

- 고객 Demographic 정보

No.	변수명(영문)	변수명(국문)	상세설명	PK
1	CLNT_ID	클라이언트 ID	• 고객을 고유하게 식별할 수 있도록 랜덤으로 부여된 ID	Y
2	CLNT_GENDER	성별	• 성별정보 [남자 : M / 여자 : F / 정보없음 : unknown]	
3	CLNT_AGE	연령대	• 연령대 정보 [10 대이하 / 20 대 / 30 대 / 40 대 / 50 대 / 60 대이상 / 정보없음 : unknown]	

- 상품 분류 정보

No.	변수명(영문)	변수명(국문)	상세설명	PK
1	PD_C	상품 소분류코드	• 상품의 소분류코드(최소단위)	Y
2	CLAC1_NM	상품 대분류명	• 상품 대분류명	
3	CLAC2_NM	상품 중분류명	• 상품 중분류명	
4	CLAC3_NM	상품 소분류명	• 상품 소분류명	

9. 관련 문의

- Competition 공식 홈페이지 내 Q&A 게시판

- EOD -