

*“Be the L.BA”*

**제5회 L.POINT Big Data Competition**

**데이터 설명서**



**목차**

**……………………………………………….………… 3**

**………………...…………………………….………… 3**

**…………..….……………………………….………… 3**

**………………..…….……………………….………… 4**

**………………………………………..…….………… 4**

**……………………………………..……….………… 5**

**………………………………..…………….………… 5**

**………………………………..…………….………… 6**

**………………………………..…………….………… 9**

**대회취지**

**대회일정**

**결과물 제출**

**결과물 제출 유의사항**

**평가 및 심사**

**시상내역**

**규정 및 운영방안**

**데이터 설명**

**관련 문의**

**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 : 2018. 10. 22(월) ~ 2018. 11. 30(금)  
2) 데이터 다운로드 : 2018. 12. 3(월) ~ 2019. 1. 16(수) 15:00 까지  
 - Competition 공식홈페이지 > 분석 Data 다운로드 (팀장만 다운로드 가능)   
3) 결과물 제출 마감 : 2019. 1. 16(수) 15:00 까지   
 - Competition 공식홈페이지 > 접수 및 확인 > 분석결과물 제출   
4) 본선 진출자 발표 : 2019. 2. 1(금), Competition 공식홈페이지 ‘공지사항’에 게시   
5) 본선 PT 심사 및 시상 : 2019. 2. 19(화)(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 한함 (15분 발표 분량)

대상 및 제출 시기는 추후 별도 공지

② 재학 및 졸업 증명서 : 참가 자격의 확인을 위해 본선 진출자에 한하여 제출. 제출대상 및 시기는 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 필수 포함 항목

(1) 상품군별 온라인 선호지수 및 수요 트렌드 예측에 대한 우수성

(2) 활용 변수의 다양성

(3) 제안 서비스의 창의성 및 시각화

- 평가 기준

|  |  |  |
| --- | --- | --- |
| **평가항목** | **평가항목상세** | **배점** |
| 상품군별  온라인 선호지수 및  수요 트렌드 예측에 대한 우수성 | 데이터전처리  - 데이터 탐색  - 데이터 클렌징 | 10 |
| 온라인 선호지수 개발  - 지수 개발 과정  - 개발지수의 논리성  - 분석기법의 적정성 | 20 |
| 수요 트렌드 예측  - 예측모형 개발 과정  - 분석기법의 적정성, 다양성  - 트렌드 예측의 정확성 | 20 |
| 인사이트 도출 | 10 |
| 활용 변수의 다양성 | 외부 데이터  - 활용된 외부 데이터 종류  - 데이터 적합성 | 10 |
| 파생 변수  - 파생 변수의 종류  - 생성 변수의 논리성, 활용성 | 5 |
| 제안 서비스의 창의성 | 서비스 제안 아이디어 창의성 | 15 |
| 시각화 | 분석 보고서의 시각화 자료 활용 | 10 |
| 합계 | | 100 |

**2) 2차 본선 심사** (배점 100점 만점)

|  |  |  |
| --- | --- | --- |
| **평가항목** | **평가항목상세** | **배점** |
| 서면 평가 결과 | 1차 서면 평가 점수   (40점 만점 환산) | 40 |
| 심사위원 조정  - 프레젠테이션에 따른   1차 평가 점수 조정 |
| 프레젠테이션 | 개발된 지수/트렌드 예측의 논리성  - 지수/트렌드 등 제안 모델과   콘텐츠 간의 논리적 연계성 | 20 |
| 서비스 제안 구현 로직의 활용  - 서비스 제안 구현 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** | **대 상** | **최우수상** | **우수상** |
| 상 금 | 300만원 | 200만원 | 100만원 |
| 수 상 자 | 1팀 | 2팀 | 3팀 |
| 수상특전 | O | O | O |

※ 수상특전 : 롯데멤버스 통계분석 관련 직무 지원 시 2019년 인턴 또는 신입사원 1차 서류전형 면제  
(기졸업자·졸업예정자 중 2019년 내 1회에 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**

**주제 : Digital Trend Analyzer**

1) 분석 목적 : 온라인 행동 기반 트렌드 예측

2) 데이터 구성   
가. 분석 주제 : 온라인행동 데이터 및 상품분류 데이터를 활용하여   
 1) 주요 상품군별 온라인 선호지수 생성  
 2) 상품군별 수요 트렌드 예측 및 인사이트 도출  
 3) 1), 2)를 활용한 신규 서비스 제안

나. 데이터 상세 설명  
 - 분석데이터 기간 : 총 6개월(‘18년 4월 ~ 9월)  
 - 제공 범위 : 롯데그룹 온라인 계열사의 구매 및 이용 이력(온라인 계열사는 임의로 선정)  
 ※ 본 데이터는 가공된 샘플 데이터로 실제 시장 데이터와 차이가 있습니다.  
 - 온라인 행동데이터 기본 개념 및 분석 데이터 구성은 **“자료.데이터설명.PDF” 참조**

3) 데이터 목록

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **구분** | **영문명** | **한글명** | **상세 설명** | **pk** |
| 상품구매  (Product) | CLNT\_ID | 클라이언트ID | ■ 방문자(Visitors)의 쿠키1)에 랜덤으로 부여된 고유 ID  ※ 기기, 디바이스, 브라우저 체제에 따라 다른 방문자로 인식되어 동일고객이라도 여러 개의 클라이언트ID를 보유할 수 있음 | PK |
| SESS\_ID | 세션ID | ■ Web/App에 접속 후 세션이 시작될 때 부여된 고유 ID ※ 하나의 클라이언트ID에 여러 개의 세션ID가 발급될 수 있음 | PK |
| HITS\_SEQ | 히트일련번호 | ■ Web/App에서 페이지 또는 화면 클릭, 이벤트 참여, 검색 등 방문자의 행위에 대해 순서대로 배열된 일련번호 ※ 세션 내에서 발생되며, 첫번째 행위에 대해서는 1로 설정됨 | PK |
| PD\_C | 상품코드 | ■ 구매한 상품의 코드(최소단위) | PK |
| PD\_ADD\_NM | 상품추가정보 | ■ 구매한 상품의 추가 정보 | PK |
| PD\_BRA\_NM | 상품브랜드 | ■ 구매한 상품의 브랜드 |  |
| PD\_BUY\_AM | 단일상품금액 | ■ 구매한 상품 1개의 금액 |  |
| PD\_BUY\_CT | 구매건수 | ■ 구매한 상품의 수량 |  |
| 검색어1  (Search1) | CLNT\_ID | 클라이언트ID | ■ 방문자(Visitors)의 쿠키1)에 랜덤으로 부여된 고유 ID  ※ 기기, 디바이스, 브라우저 체제에 따라 다른 방문자로 인식되어 동일고객이라도 여러 개의 클라이언트ID를 보유할 수 있음 | PK |
| SESS\_ID | 세션ID | ■ Web/App에 접속 후 세션이 시작될 때 부여된 고유 ID ※ 하나의 클라이언트ID에 여러 개의 세션ID가 발급될 수 있음 | PK |
| KWD\_NM | 검색키워드명 | ■ 검색창에 입력한 검색 키워드 | PK |
| SEARCH\_CNT | 검색건수 | ■ 세션 내 해당 검색어 검색량 |  |
|  |  |  |  |  |
| 검색어2  (Search2) | SESS\_DT | 세션일자 | ■ 세션일자 (YYYYMMDD 형식으로 표시) | PK |
| KWD\_NM | 검색키워드명 | ■ 검색창에 입력한 검색 키워드 | PK |
| SEARCH\_CNT | 검색건수 | ■ 세션 내 해당 검색어 검색량 |  |
| 회원  (Custom) | CLNT\_ID | 클라이언트ID | ■ 방문자(Visitors)의 쿠키1)에 랜덤으로 부여된 고유 ID  ※ 기기, 디바이스, 브라우저 체제에 따라 다른 방문자로 인식되어 동일고객이라도 여러 개의 클라이언트ID를 보유할 수 있음 | PK |
| CLNT\_GENDER | 성별 | ■ 성별정보 [남자 : M / 여자 : F] |  |
| CLNT\_AGE | 연령대 | ■ 연령대 정보 [10대 : 10 / 20대 : 20 / 30대 : 30 ...] |  |
| 세션  (Session) | CLNT\_ID | 클라이언트ID | ■ 방문자(Visitors)의 쿠키1)에 랜덤으로 부여된 고유 ID  ※ 기기, 디바이스, 브라우저 체제에 따라 다른 방문자로 인식되어 동일고객이라도 여러 개의 클라이언트ID를 보유할 수 있음 | PK |
| SESS\_ID | 세션ID | ■ Web/App에 접속 후 세션이 시작될 때 부여된 고유 ID ※ 하나의 클라이언트ID에 여러 개의 세션ID가 발급될 수 있음 | PK |
| SESS\_SEQ | 세션일련번호 | ■ Web/App에 접속 후 세션이 시작될 때 부여된 일련번호 ※ 첫 번째 세션이면 1로 설정됩니다. 세션ID의 순서를 알 수 있는 번호 |  |
| SESS\_DT | 세션일자 | ■ 세션일자 (YYYYMMDD 형식으로 표시) |  |
| TOT\_PAG\_VIEW\_CT | 총페이지조회건수 | ■ 세션 내의 총 페이지(화면) 뷰 수 |  |
| TOT\_SESS\_HR\_V | 총세션시간값 | ■ 세션 내 총 시간(단위: 초) |  |
| DVC\_CTG\_NM | 기기유형 | ■ 기기 유형(1: desktop , 2: mobile, 3.tablet ) |  |
| ZON\_NM | 지역대분류 | ■ 세션기준 발생 대분류 지역명 (대분류IP 주소 또는 지역 ID 기준) |  |
| CITY\_NM | 지역중분류 | ■ 세션기준 발생 중분류 지역명 (중분류IP 주소 또는 지역 ID 기준) |  |
| CLNT\_ID | 클라이언트ID | ■ 방문자(Visitors)의 쿠키1)에 랜덤으로 부여된 고유 ID  ※ 기기, 디바이스, 브라우저 체제에 따라 다른 방문자로 인식되어 동일고객이라도 여러 개의 클라이언트ID를 보유할 수 있음 |  |
|  |  |  |  |  |
| 상품분류 | PD\_C | 상품코드 | ■ 상품의 코드(최소단위) | PK |
| PD\_NM | 상품명 | ■ 상품명 |  |
| CLAC1\_NM | 상품 대분류명 | ■ 상품 대분류명 |  |
| CLAC2\_NM | 상품 중분류명 | ■ 상품 중분류명 |  |
| CLAC3\_NM | 상품 소분류명 | ■ 상품 소분류명 |  |

※ PK(Primary Key) : 데이터 테이블 설계 시 기준으로 설정되는 기본Key

**9. 관련 문의**

- Competition 공식 홈페이지 내 Q&A 게시판

-EOD-