

제6회 L.POINT

BIG DATA Competition

최대 300만원 수상과 채용 특전까지



L.pay | L.POINT

제6회 L.POINT BIG DATA Competition 설명자료

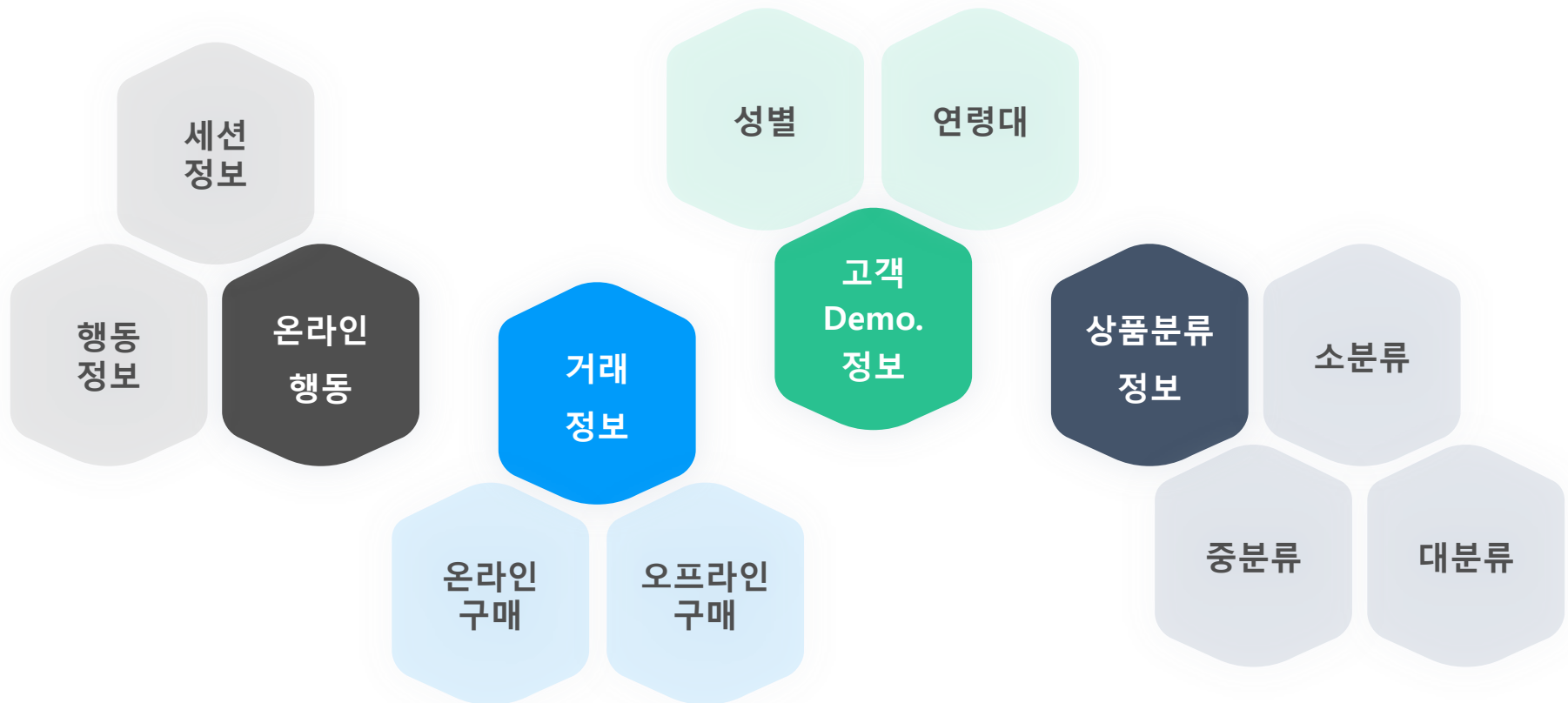
※ 본 자료는 2019년 11월 18일 진행된
제6회 L.POINT BIG DATA Competition
설명회 발표자료 중 일부와 당시 Q&A
내용을 가공하여 작성되었습니다.



분석용 배포 데이터

분석데이터는 온라인 행동 정보, 거래 정보, 고객의 Demographic 정보, 그리고 상품분류정보 총 4개 종류의 데이터로 구성되어 있습니다.

※ 분석 데이터 기간 : 총 3개월 ('19년 7월 ~ 9월)



분석용 배포 데이터

온라인 행동 정보

온라인 행동 정보란, 고객의 온라인 행동에 대한 기록으로써 유입부터 구매까지 모든 행동 과정을 분석할 수 있는 데이터입니다.



분석용 배포 데이터

온라인 행동 정보 - 세션

세션(Sessions)은 방문자가 일정한 기간 내에 사이트에 방문하여 활동하는 영역을 의미하며, 세션 시작 시 세션 ID가 발급됩니다.

세션 ID가 변경되는 경우

1) 접속 사이트가 달라지는 경우

예)



2) 세션 내 행동이 없이 30분 경과하는 경우

예) 세션 내 행동이 30분 이상 없을 경우 신규 세션 ID 발급

3) 기존 접속 경로와 다르게 접속하는 경우(30분 내 기준)

예) URL을 입력하여 사이트 접속한 뒤 검색 포털을 통해 접속하는 경우 신규 세션 ID 발급

※ URL을 직접 입력하여 사이트 접속한 뒤 동일하게 URL을 직접 입력하여 접속하는 경우 기존 세션 ID 유지

분석용 배포 데이터

온라인 행동 정보 - 조회

조회(Hits)는 방문자의 세션 내 최소 행동 단위를 의미하며, 페이지 조회, 클릭, 검색, 상품 장바구니 담기, 상품 구매 등 모든 행위를 포함합니다.

조회(Hits)



1

2



3

[데이터 예시]

고객ID	날짜	세션 ID	Hit 일련번호	행동	거래ID
00001	11/29	030	1	삼겹살 검색	-
00002	12/2	008	1	우유 검색	-
			2	장바구니	-
			3	구매확정	108484

분석용 배포 데이터

클라이언트 ID

클라이언트(Client) ID는 고객을 고유하게 식별할 수 있도록 랜덤으로 부여된 ID입니다.

※ 쿠키(cookie) : 방문자가 웹사이트를 방문할 때 저장되는 기록 정보로써 이용한 디바이스(종류, 모델) 및 브라우저(익스플로러, 크롬 등)에 따라 달라지며 삭제가 가능하여 삭제 후 해당 웹사이트를 방문할 경우 새로운 쿠키 정보가 생성됨

클라이언트 ID 부여

• 쿠키 정보와 구매 정보에 따라 동일 고객의 여부를 판단

- 세션 간 쿠키 및 접속경로 정보가 같은 경우 동일 고객으로 판단
- 세션 간 쿠키 및 접속경로 정보가 다른 경우
 - 1) 각 세션에서 구매 정보가 있는 경우 고객의 식별이 가능하기 때문에 세션 간 동일고객 여부를 판단할 수 있음
 - 2) 구매 정보가 없거나 한 세션에만 있는 경우 세션 간 동일고객 여부를 판단할 수 없음

[예시 - 동일고객의 온라인 행동 정보]

Case	쿠키 정보	날짜	세션ID	행동유형	클라이언트 ID
1	A	12/3	011	검색	00789
	A	12/3	011	검색	00789
2	A	12/4	015	제품 목록	00789
	A	12/4	015	제품 목록	00789
3	B	12/5	022	검색	03302
	B	12/5	022	제품 목록	03302
4	B	12/6	032	제품 목록	03302
	B	12/6	032	구매 완료	03302
5	C	12/7	067	제품 목록	03302
	C	12/7	067	구매 완료	03302
6	C	12/7	069	제품 목록	03302
	C	12/7	069	구매 완료	03302

[Case 1 ↔ 2 / Case 3 ↔ 4 / Case 5 ↔ 6]

세션 간 쿠키 정보가 동일하기 때문에
같은 클라이언트ID 부여

[Case 2 ↔ 3]

세션 간 쿠키 정보가 다르기 때문에
다른 클라이언트ID 부여

[Case 4 ↔ 5, 6]

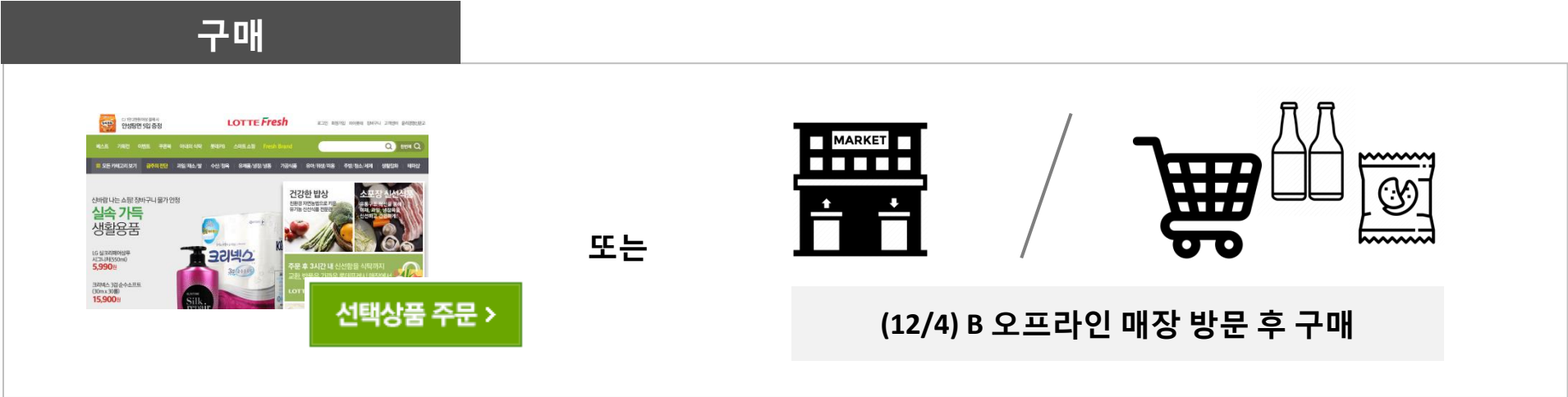
세션 간 쿠키 정보가 다르지만 구매 정보를
기반으로 같은 클라이언트ID 부여

※ 쿠키 정보는 별도로 제공되지 않습니다.

분석용 배포 데이터

거래정보 & 상품분류정보

거래 정보는 온라인 또는 오프라인에서 구매한 내역이며, 거래 일련번호는 복수의 상품이 동시에 구매된 경우를 구분합니다.



[데이터 예시]

고객ID	날짜	거래 ID	거래 일련번호	상품 (소분류)	구매수량	구매금액
A포이	11/30	ORDR33356	1	치즈	5	5,500
A포이	12/4	ORDR44005	1	수입맥주	2	8,000
			2	일반스낵	1	2,000

[데이터 예시]

상품 (소분류)	상품 (중분류)	상품 (대분류)
치즈	유가공품	유제품
수입맥주	맥주	주류
일반스낵	스낵류	과자

별첨) 데이터 레아웃 (1/3)

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

온라인 행동 정보

No.	변수명(영문)	변수명(국문)	상세설명	PK
1	CLNT_ID	클라이언트ID	• 고객을 고유하게 식별할 수 있도록 랜덤으로 부여된 ID	Y
2	SESS_ID	세션ID	• Web/App에 접속 후 세션이 시작될 때 부여된 순번 ID ※ 하나의 클라이언트ID에 여러 개의 세션ID가 발급될 수 있음	Y
3	HIT_SEQ	조회일련번호	• 조회 순서를 알 수 있도록 부여된 일련번호	Y
4	ACTION_TYPE	행동유형	• 총 8가지의 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 0으로 설정됩니다.	
9	TRANS_ID	거래ID	• 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]	

별첨) 데이터 레이어아웃 (2/3)

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

거래 정보

No.	변수명(영문)	변수명(국문)	상세설명	PK
1	CLNT_ID	클라이언트ID	• 고객을 고유하게 식별할 수 있도록 랜덤으로 부여된 ID	Y
2	TRANS_ID	거래ID	• 구매 내역을 고유하게 식별할 수 있도록 랜덤으로 부여된 ID	Y
3	TRANS_SEQ	거래일련번호	• 해당 거래ID의 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	

별첨) 데이터 레이아웃 (3/3)

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

고객 Demo. 정보

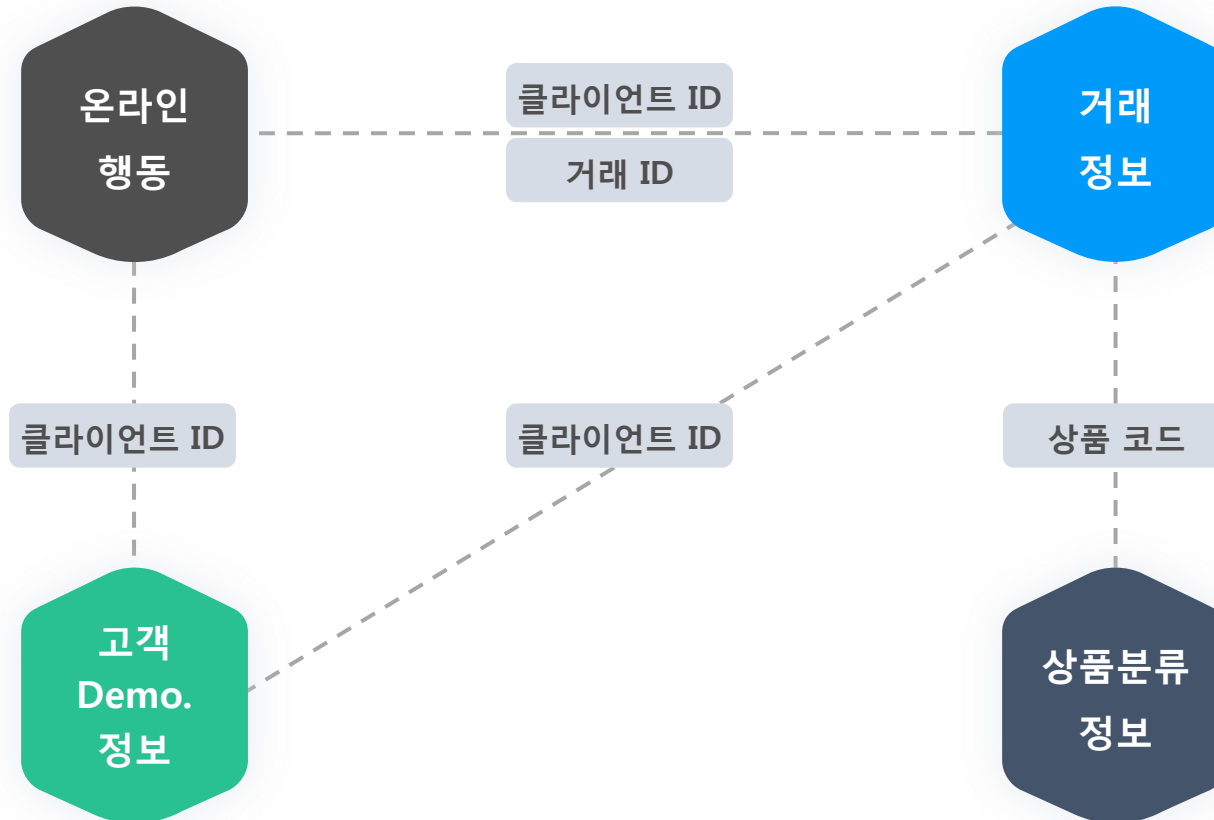
No.	변수명(영문)	변수명(국문)	상세설명	PK
1	CLNT_ID	클라이언트ID	• 고객을 고유하게 식별할 수 있도록 랜덤으로 부여된 ID	Y
2	CLNT_GENDER	성별	• 성별정보 [남자 : M / 여자 : F / 정보없음 : unknown]	
3	CLNT_AGE	연령대	• 연령대 정보 [10대이하 / 20대 / 30대 / 40대 / 50대 / 60대이상 / 정보없음 : unknown]	

상품분류정보

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	

별첨) 데이터 흐름도

※  : 테이블 조인 Key



설명회 Q&A

설명회 당시 Q&A 내용 중 주요 내용을 정리하였습니다.

Q. 시각화의 경우에도 사용 가능한 Tool 제한이 있나요?

A. 시각화 Tool의 경우 제한이 없습니다. 다만 분석 Tool은 SAS, SQL, R, Python 으로 제한되어 있습니다.

Q. 제공하는 데이터가 실제 데이터인가요?

A. 제공해 드리는 데이터는 실제 데이터를 기반으로 가공한 데이터로 실제 데이터와 동일하지는 않습니다.

Q. 알고리즘이나 인사이트 도출만 하면 되는지 아니면 비즈니스/서비스 활용까지 제안해야 하는건가요?

A. 도출한 알고리즘이나 인사이트를 활용하여 비즈니스/서비스 활용까지 제안해 주셔야 합니다.

Q. 세션 내에서 30분간 행동이 없을 경우 세션ID가 변경되는데 여기서 말하는 행동의 의미가 궁금합니다.

A. 행동이라 함은 조회, 검색, 장바구니 추가, 구매 등 온라인 행동 유형으로 규정되어 있는 것을 의미합니다.

Q. 평가/시상은 주제별로 나누어서 하나요?

A. 주제별로 나누어서 평가하거나 시상하지 않습니다. 주제와 상관없이 동일한 기준으로 평가할 예정입니다.