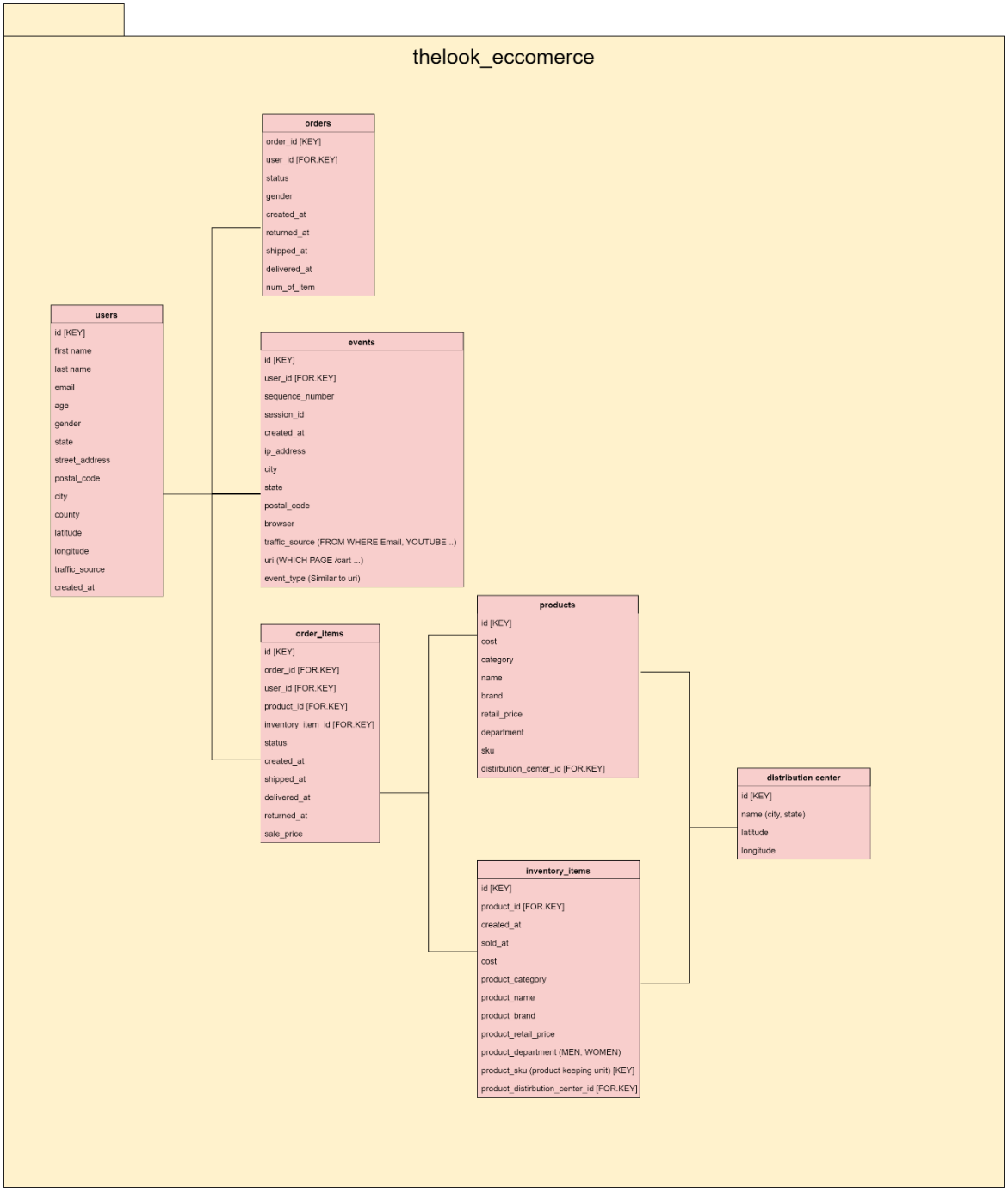

그림 1 이커머스 데이터 스키마 및 테이블 관계

프로젝트 주제 선정 이유

SQL과 익숙하지 않고 첫 SQL 프로젝트이기 SQL 활용에 초점을 두었습니다. 그래서 *그림 1*과 같이 하나의 데이터베이스에 여러 테이블이 있는 데이터를 활용하기로 했습니다. 활용된 데이터는 가상 의류 이커머스 데이터로 구글 빅 쿼리에서 제공하는 공공 데이터 셋입니다.

고객 세분화 분석은 고객들을 공통된 특징을 기준으로 그룹화 하는 분석입니다. users 테이블과 order_items 테이블을 사용해서 이메일 하나당 생성된 계정의 수로 고객을 그룹화하고 그룹당 평균 주문 건 수를 분석했습니다.

고객 세분화 분석 결과

고객 세분화 분석에서는 users 테이블과 order_items 테이블을 활용해서 고객 세분화를 진행했습니다. users 테이블을 보면 유니크한 아이디가 100,000개로 사용자가 웹사이트 이용자가 100,000명 있는 것으로 보입니다. 하지만 유니크한 이메일로 보면 웹사이트 이용자가 83,732명 인 것을 알 수 있습니다. 즉, 단일 이메일로 다수의 계정 생성이 가능했습니다.

*표 4*를 통해 유니크한 아이디 수 대비 유니크한 이메일 계정 수의 비율을 보면 중국, 미국, 브라질 순으로 단일 이메일로 여러 개의 계정을 생성한 사용자들이 많았습니다. 콜롬비아, 폴란드, 그리고 오스트리아는 단일 이메일로 단일 계정만 생성한 것을 알 수 있었습니다.

country	email_cnt	id_cnt	email_id_ratio
China	31335	33859	108.0549
United States	21208	22397	105.6064
Brasil	14166	14710	103.8402
Colombia	19	19	100
Poland	228	228	100
Austria	7	7	100

표 4 나라별 아이디 수 대비 이메일 수

*표 5*을 통해 중국 남성 이메일의 약 10%가 이러한 경향을 나타낸 것을 확인할 수 있습니다. 그 다음으로 미국 남성 이메일의 약 7%, 중국 여성 이메일의 약 6%가 하나의 이메일로 여러 계정을 만들었습니다.

country	gender	email_cnt	id_cnt	account_email_ratio
China	M	15366	16918	110.1002
United States	M	10402	11128	106.9794
China	F	15978	16941	106.027

표 5 나라별 성별 별 아이디 수 대비 이메일 수

이것을 통해 두가지 가설을 세울 수 있었습니다. 첫번째 중국, 미국 등의 거주하는 사용자들은 비밀번호를 찾는 것 보다 새로운 아이디를 만드는 것이 더 시간이 절약된다고 생각한다는 것입니다. 두번째, 해당 이커머스 웹사이트에서 고객 정보 변경을 불가하게 만들었기에 이러한 결과를 갖고 왔을 수도 있습니다.

가설 검증을 위해서 더 세부적으로 하나의 이메일로 생성된 아이디들을 나라, 자치 주, 생성된 날들로 살펴보았습니다. *표 6*를 살펴본 결과 다수의 계정을 보유한 사용자들을 보면 한 나라의 한 자치 주에서 계속 거주하는 것이 아닌 거주지를 옮길 때마다 새로운 아이디를 생성한 것을 확인할 수 있었습니다. 그러므로 아까 세웠던 가설에서 해당 웹사이트에서 고객 정보를 변경 불가하게 했기에 이러한 결과를 갖고 왔을 수 있다고 봅니다.

email	id	country	state	created_at
heatherbrown@example.org	37517	Spain	Cataluña	2023-12-05 11:01:00 UTC
	17790	United States	New Mexico	2022-03-22 11:00:00 UTC
	16224	Brasil	Acre	2021-02-19 00:51:00 UTC
	20782	Brasil	Minas Gerais	2019-06-05 07:19:00 UTC
angelarogers@example.net	21917	Brasil	Acre	2020-10-29 16:03:00 UTC
	87361	United Kingdom	Wales	2019-01-11 10:08:00 UTC
cynthiataylor@example.net	12722	France	Normandie	2021-07-16 00:11:00 UTC
	6340	China	Zhejiang	2021-05-13 13:35:00 UTC
	32781	Brasil	Acre	2021-02-09 17:50:00 UTC

표 6 이메일 별 생성 아이디, 국가, 자치 구, 그리고 생성일

분석 결과를 바탕으로 고객 세분화를 위해 이메일 별 아이디 개수로 그룹화를 했습니다. 아이디가 많은 고객이면 더 많은 구입 패턴을 나타내는지 검증을 하기 위해 아이디 개수와 주문 건수의 비율을 구하려고 했습니다. 데이터를 조회한 결과 하나의 이메일로 최다 19개의 계정을 만들었습니다. 그래서 먼저 1부터 19로 계정 수를 분류했습니다. 그리고 같은 이메일 계정으로 주문한 주문 건 수들을 살펴보았습니다. 단일 이메일로 단일 계정만 만든 사용자들이 많아서 아이디가 하나일 때 총 주문량이 많았습니다. 그래서 *표 7*와 같이 비율로 비교했습니다. 같은 개수의 아이디로 발생한 총 구매량 대비 해당 아이디 개수의 이메일 수로 구해 보았습니다. 분석 결과로 19개의 계정을 갖고 있는 이메일 하나당 평균적으로 24건의 주문이 있었습니다.

id_count	email_count	total_orders	order_email_ratio
19	2	48	24
17	2	47	23.5

16	3	58	19.33333
3	1731	6488	3.748122
2	7781	19416	2.495309
1	73023	91466	1.252564

표 7 아이디 개수 별 이메일 개수, 총 주문 건수, 이메일 별 평균 주문건수

*표 7*에서도 볼 수 있다시피 보유 아이디 개수가 많은 사용자들이 평균적으로 더 많은 주문을 했습니다. 이 트렌드의 의미는 많은 아이디를 가진 사용자들이 해당 웹사이트의 충성 고객이거나 해당 웹사이트에서 판매하는 물건이 해당 고객에게 필수적으로 필요했기에 평균 주문 건수가 많은 것으로 보여집니다.

그렇기에 이러한 고객들이 다른 웹사이트로 넘어가지 않게 이들을 타켓한 마케팅 전략이 필요하다고 보여집니다. 주소를 옮길 때 마다 불편을 감수하고 아이디를 새로 생성해서 웹사이트를 이용하는 사용자들의 경향을 기반으로 다른 곳으로 이사를 계획하고 있는 사용자들에게 특별한 세일 및 서비스를 제공해서 유지율을 높이고 웹사이트의 불편한 점을 개선하는 것이 중요해 보입니다.