

다이나믹하나

대시보드를 활용한 다이나믹 카드명세서

하나금융티아이 교육생 구본성

01 배경 및 기대효과

배경	초개인화 서비스 확대 → 정보의 주체자인 고객에 맞게 개별적인 금융서비스 제공
목표	하나카드의 앱은 거래내역 제공, 서비스 가입 등 제공 → 정보성 페이지 제공을 통해 스마트한 소비 주도, Lock-in 효과
기대효과	데이터 분석을 통한 고객 소비 생활, 소비 패턴 파악 → 각 고객에게 알맞은 카드 추천, 관리자의 유연한 대처 가능

02 주요기능 : 사용자 기능



나만의 카드

- 고객이 원하는 혜택 기반 카드 검색
- 고객의 지난달 소비 기반 예상 혜택 제공



실시간 거래 내역서

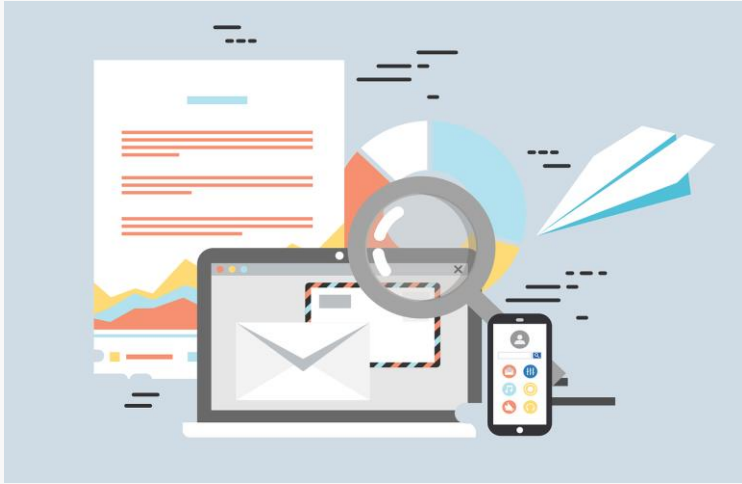
- 고객의 최근 90일 거래 내역 제공
- 최근 거래 추세 제공



월간 거래 리포트

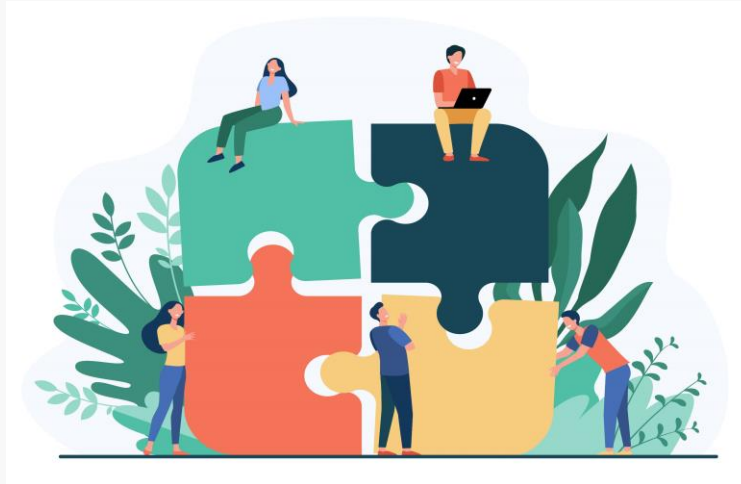
- 1달간의 고객 소비 정리, 메일링
- 고객 소비에 최적화된 카드 추천

02 주요기능 : 관리자 기능



고객 대시보드

- 고객의 시간대별, 일간, 월간의 거래량 비교
- 가장 많이 사용된 카드 확인



회원 분석

- 매월 1일 형성된 군집 정보 확인, 군집의 주요 소비처와 군집 소속 회원 확인
- 고객들의 인구통계학적 정보 제공



안내 메일 발송

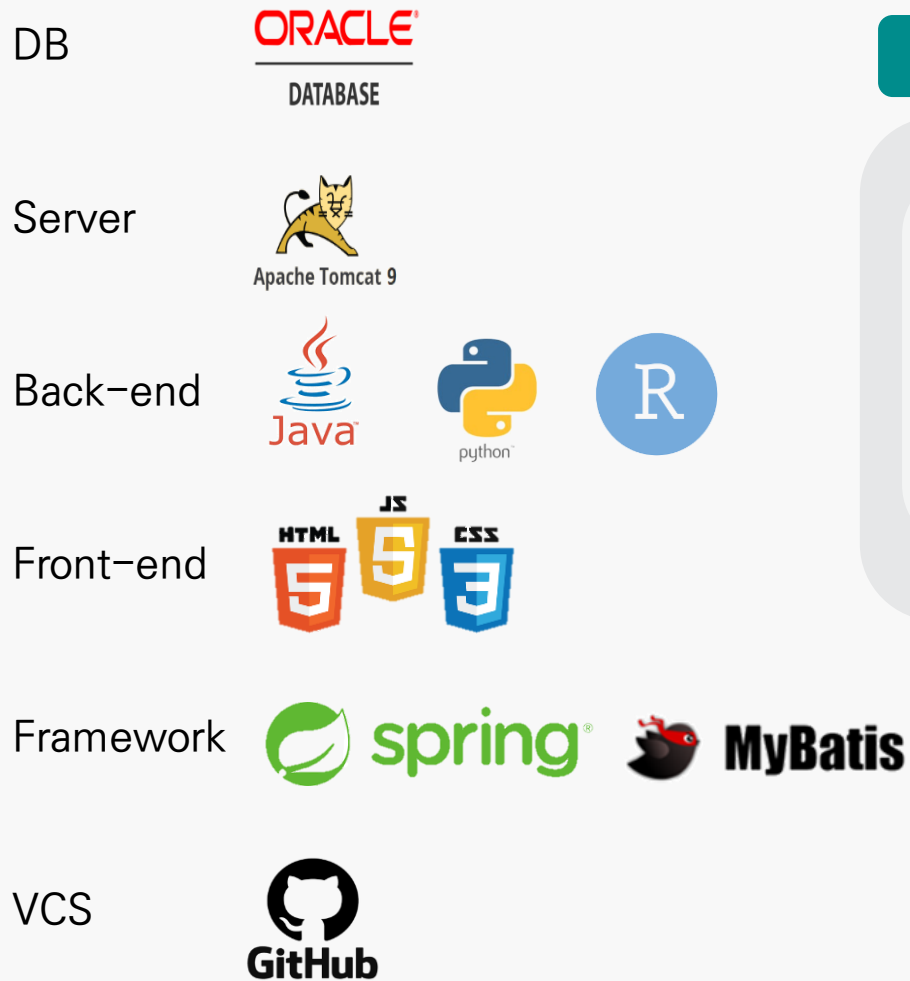
- 형성된 고객 군집에 대해 추천 카드를 기반으로 한 메일링

하나카드 다이나믹하나 2021.09.06 ~ 2021.10.05

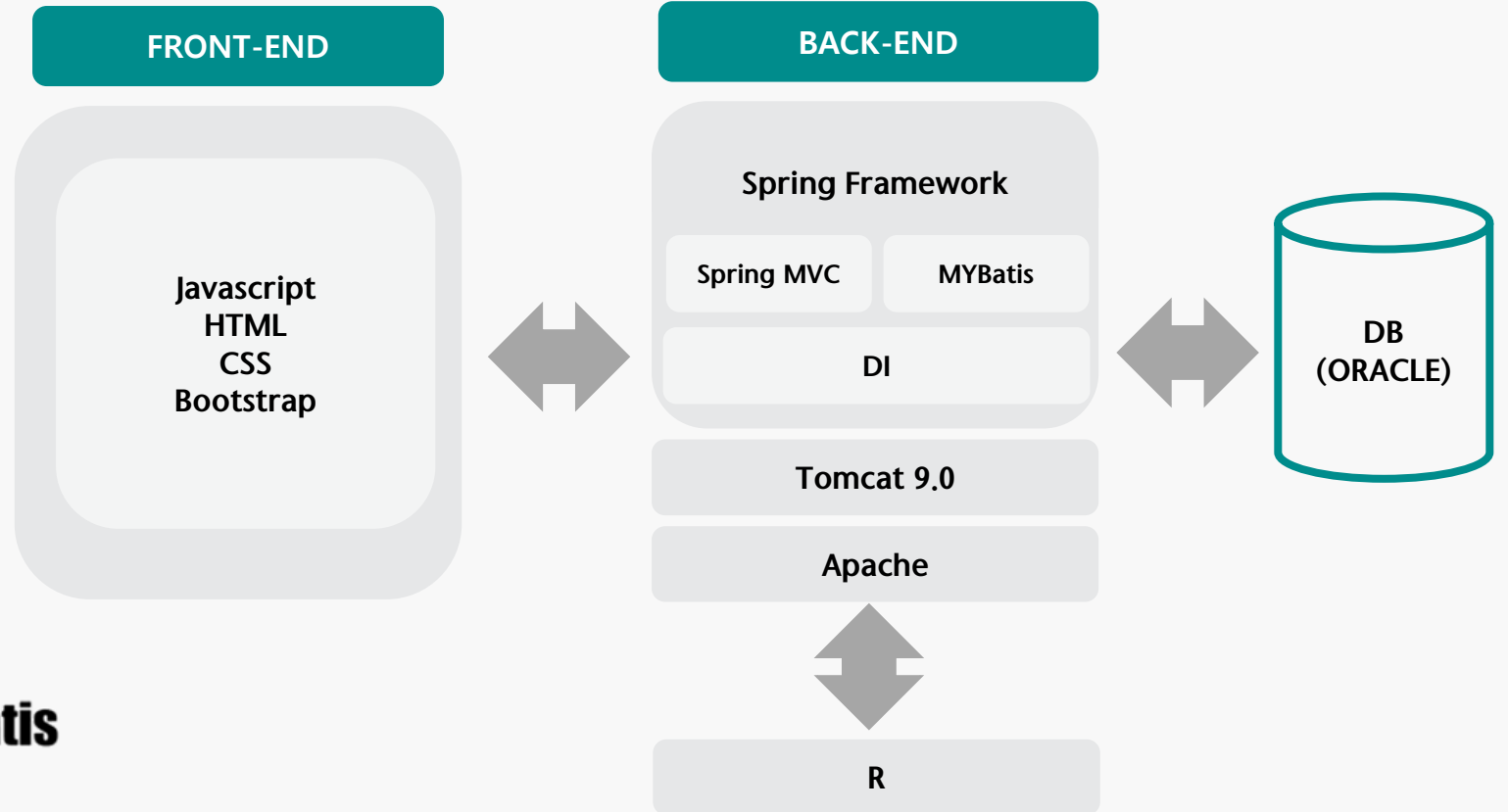
	9/6	9/11	9/16	9/21	9/26	10/1	10/6
프로젝트 주제 선정	2						
프로젝트 설계 및 데이터 생성	4						
사용자 기능 개발		10					
데이터 분석 모듈 개발			5				
관리자 기능 개발				7			
단위 테스트 및 미비된 기능 구현					4		
최종 테스트 및 기능점검						4	
프로젝트 최종 발표							1

04 개발 환경 및 아키텍처

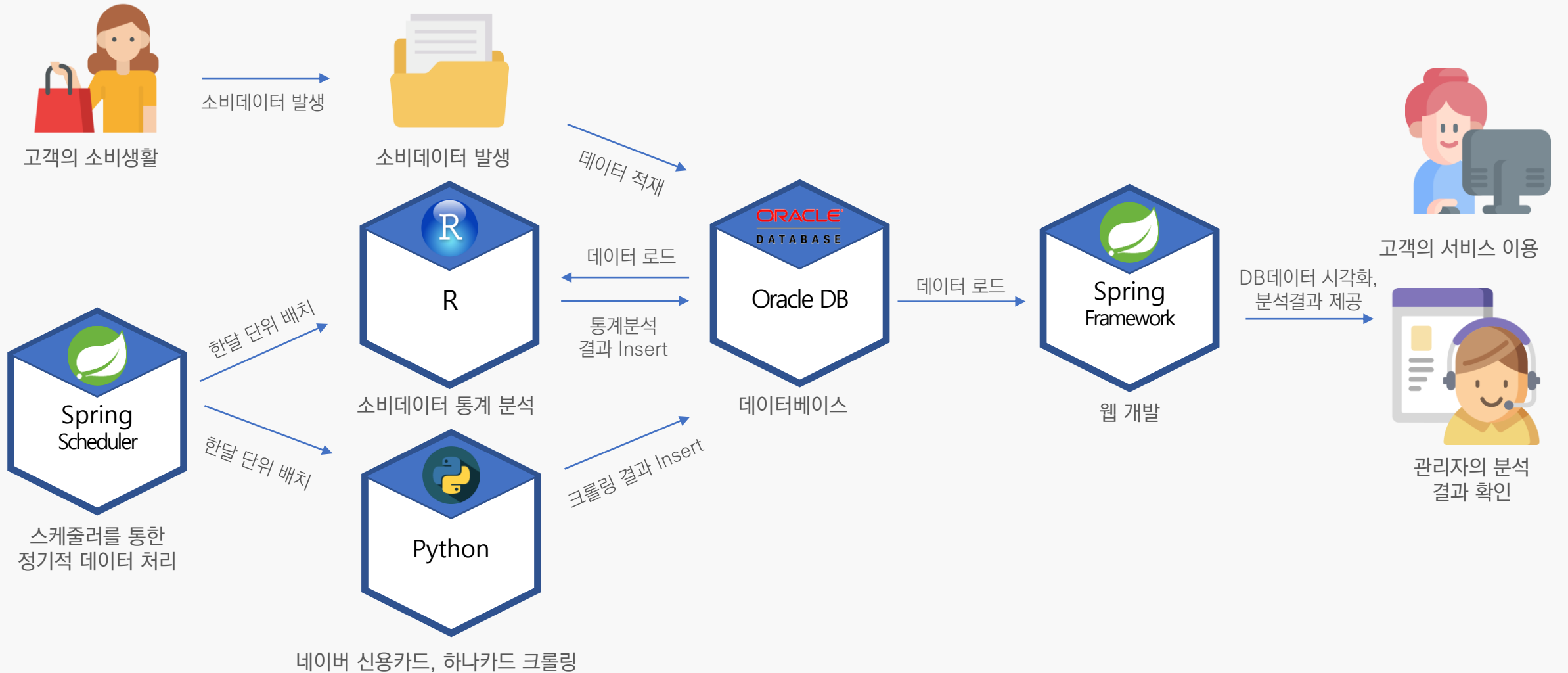
개발 환경



시스템 아키텍처

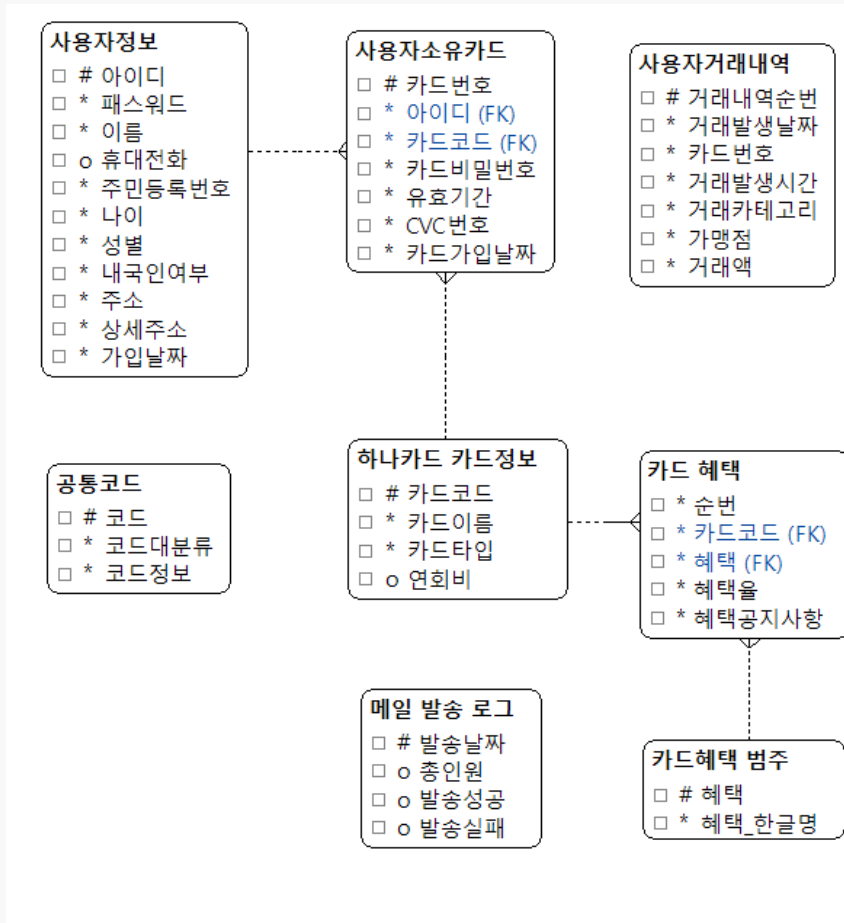


04 개발 환경 및 아키텍처

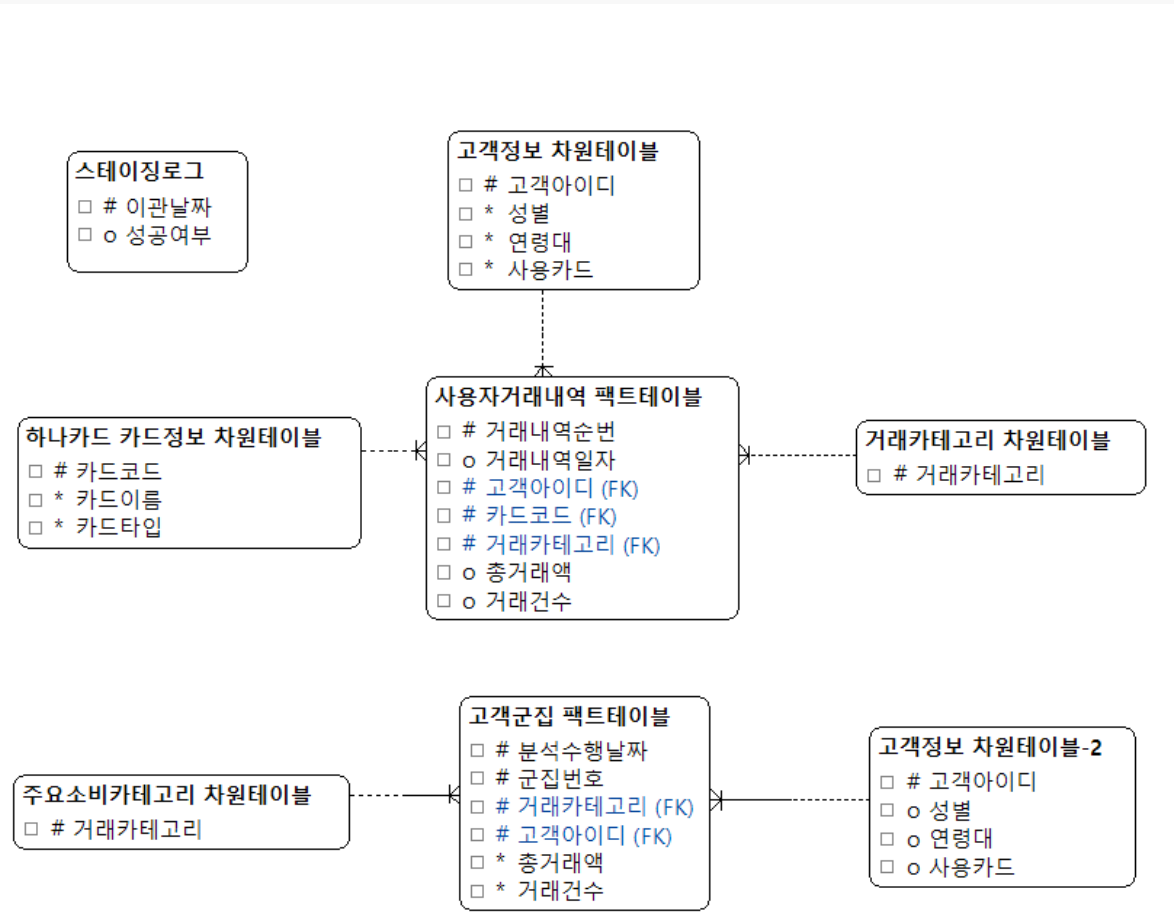


05 ERD 및 데이터 구조

<계정계>

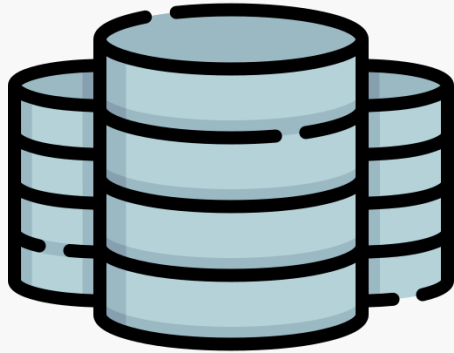


<정보계 - 스타 스키마>



데이터 생성

① 데이터 생성



서울시 빅데이터 캠퍼스에서 제공하는
신한카드 고객 소비 데이터 재구성

② 하나카드 고객 데이터



5천명의 고객 가정, 1인 1카드

③ 하나카드 고객 소비 데이터



1인당 약 200건의 결제건수
약 1백만건의 데이터 생성

06 적용 특화 기술

① 주성분 분석을 통한 고객 군집화

- ✓ 매월 1일, 한달간 축적된 고객 소비데이터를 기반
- ✓ 데이터의 차원을 축소시키지만, 공통의 변수로 원래 데이터의 의미를 유지할 수 있도록



② 대용량 데이터 Bulk Insert

- ✓ Bulk Insert를 통한 Insert 시간 절약
- ✓ 최적의 Batch Size 탐색
- ✓ 약 2천만건의 데이터 Bulk Insert
(해당 서비스는 1백만건만 사용)

③ 데이터베이스 Indexing을 통한 성능 향상

- ✓ Full Scan → Index Range Scan
- ✓ Cost 감소, 쿼리 수행 시간 감소

Bulk Insert 성능 비교

Batch Size	Data 갯수	소요 시간
0	1883,1554개	2시간 10분
1000		3분 25초

대시보드 일간 소비액 변화율 쿼리 성능 테스트

	인덱스 설정 전	인덱스 설정 후
COST	104754	448
쿼리 수행시간	4 sec	0.1 sec

④ 스케줄러를 이용한 군집 갱신

- ✓ 한달 단위의 소비데이터를 분석 한 후, 결과를 바탕으로 군집을 갱신
- ✓ 새롭게 갱신된 정보를 바탕으로 고객에게 카드를 추천

⑤ 데이터 크롤링

- ✓ 하나카드 웹사이트, 네이버 신용카드 이나 공지사항 등의 주요정보 크롤링
- ✓ 신용카드의 주요 혜택과 공지사항을 카드별로 정리

⑥ Java MailSender

- ✓ 월간 거래 레포트, 카드 안내 메일 등 정기적으로 고객에게 주요 정보 전달

07 시연영상

1

OLAP 구현 필요

→ 계정계와 정보계에 대한 DB 분리, 데이터의 효율적인 관리

2

빅데이터에 대한 기존과는 다른 접근 방법 필요

→ 분산처리, 쿼리 최적화 등 빅데이터 성능 향상

3

Logic에 기반한 개발을 하자!

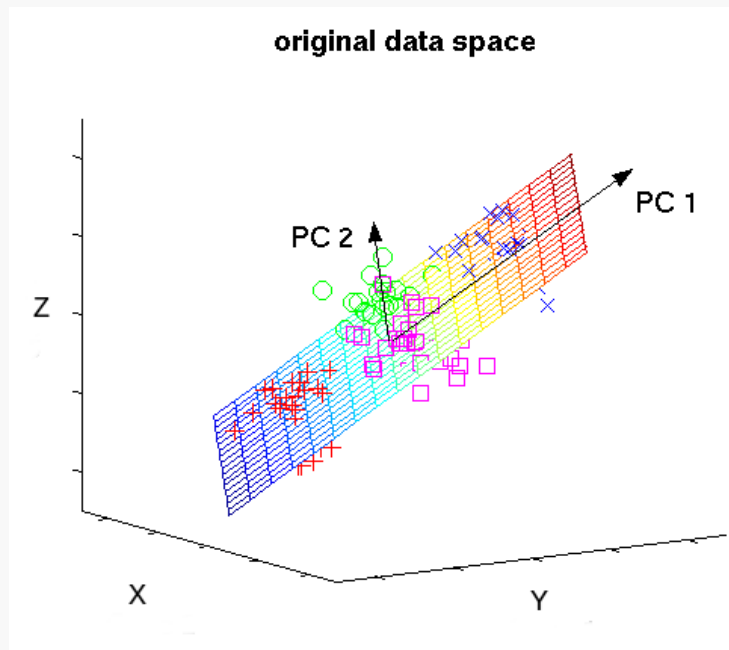
→ 논리적으로 좋은 구조를 가진 프로그램을 짜도록 노력하자

감사합니다 :)

하나금융티아이 교육생 구본성

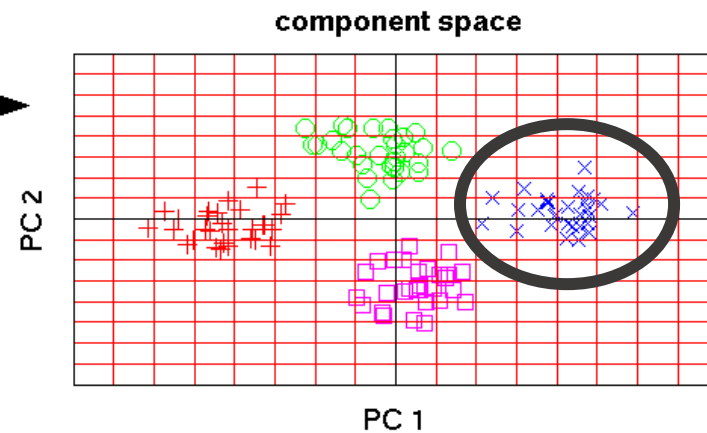
PCA & K-means 군집화 방법

① 번



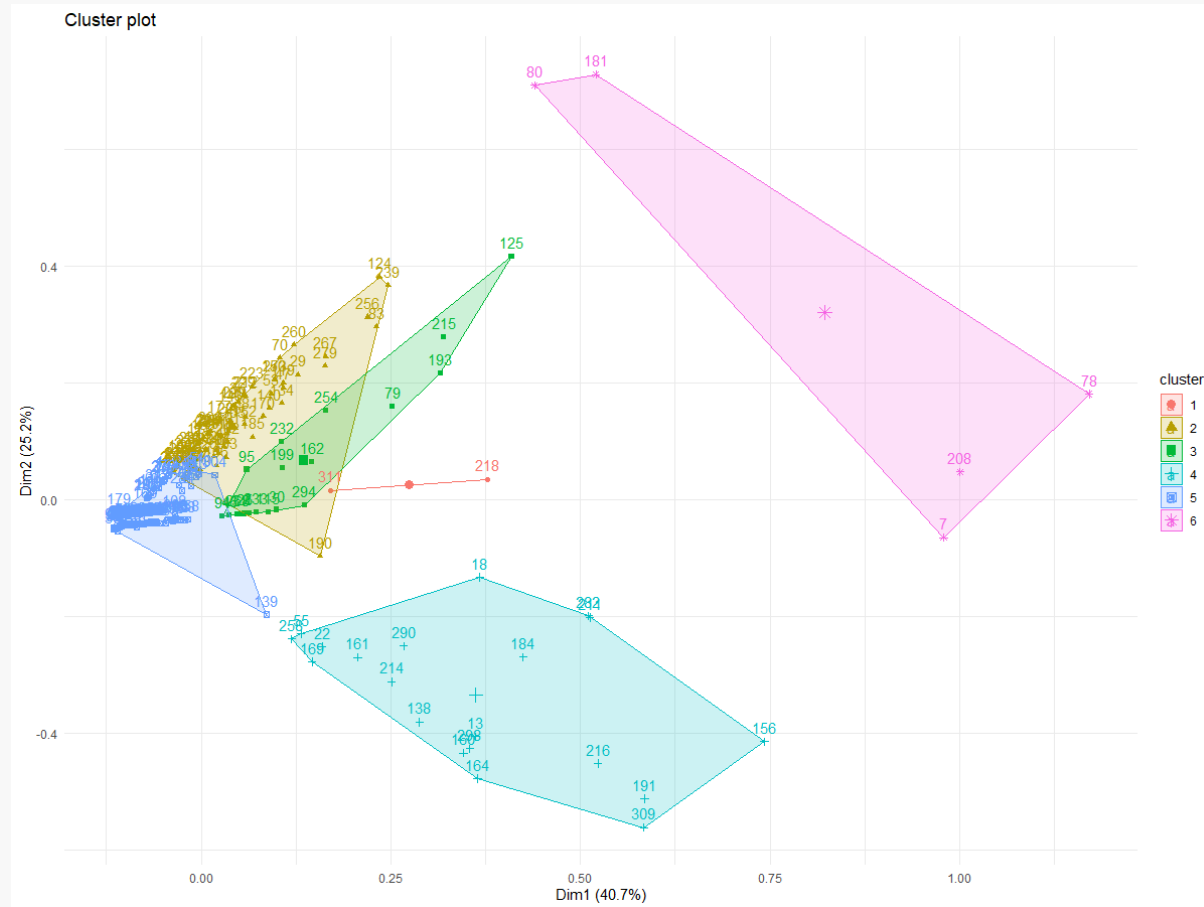
② 번

PCA



- ✓ 3차원의 데이터의 축을 두개의 주성분으로 축약한 후, 데이터를 k-means 방법을 통해 군집화
- ✓ 이때 검은색 원에 둘러싸인 데이터는 PC1에서 양의방향으로 뻗어있기 때문에 주성분 1의 성질을 가진다고 판단

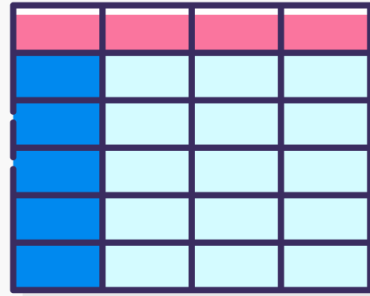
2021년 8월 군집화 결과



Ex 데이터 수집 절차



서울시 빅데이터 캠퍼스 방문
신한카드 고객 소비데이터 열람



약 **12억건의 데이터** 요약집계
(2018. 01 ~ 2018. 12)



데이터 마스킹 후 심사,
데이터 반출 후 재가공
(정책상 raw데이터 반출 불가)

Ex 데이터 수집 절차

신한카드 고객 소비데이터
요약집계 테이블

월	성별	연령대	소비카테고리	분포	평균소비금액	분산
1월	남	10대	외식	0.14	140,000	10000
1월	남	20대	카페/베이커리	0.02	8,000	800
...						
12월	여	70대 이상	의료	0.04	40,000	1000



분포(확률)에 근거한 추출

하나카드 고객
가상 소비데이터

일자	성별	연령대	이름	소비카테고리	소비금액	카드번호
2021-01-04	남	10대	안재훈	외식	130,000	1234-1234
2021-01-03	남	20대	황준호	카페/베이커리	7,500	2345-2345
...						
2021-12-11	여	70대 이상	윤소영	의료	48,000	5678-5678