

행정으로 국민의 삶의 질을 개선하라!
해커톤

분석결과 보고서

코로나19 시대, 문화누리카드 활성화 방안 제시 방안

참여자: 김명희, 김연선, 손주웅,
위은진, 이영현, 이윤지, 전재현

씨에스리 컨소시엄

CSLEE 한국생산성본부

1. 분석 개요	1
가. 분석 배경 및 개요	2
1) 문화 격차 및 문화누리카드 사업 개요	2
2) 문화누리카드 활성화 방안의 필요성	3
나. 분석 목적 및 방향	5
다. 분석 결과 활용 방안	5
2. 분석 데이터	6
가. 분석 데이터 목록	6
나. 데이터 상세 설명	8
다. 데이터 정제 방안	10
3. 분석 프로세스	11
가. 분석 프로세스	11
1) 전체 분석 프로세스	11
2) 지역별 여가 선호도 및 소비 변화 분석 프로세스	12
3) 잔액 사용 가맹점 추천 알고리즘 개발	12
나. 분석 내용 및 방법	12
1) 지역별 여가 선호도 및 소비 변화 분석	12
가) 지역별 주요 여가 니즈 추출	13
나) 포스트코로나 시대 주요문화 소비 트렌드 추출	13
다) 소비 니즈 및 소비 트렌드 기반 지역별 가맹점 평가지표 개발	18
라) 지역별 문화누리카드 가맹점 분포 평가	21
2) 잔액 사용 가맹점 추천 알고리즘 개발	24
가) 가상 고객 거래 데이터 셋 구축	24
나) 가맹점 추천을 위한 선호도 점수 생성 및 테이블 재구조화	26
다) 1차 추천 수행	27
라) 고객 메타데이터를 바탕으로 필터링 후 최종 가맹점 추천	29
마) 사용 분석기법	32

4. 분석결과 34

가. 문화누리카드 가맹점 취약 지역 제시 및 결과해석 34

- 1) 취미*자기개발 오프라인 가맹점 분포 현황 35
- 2) 도서 분야 최하위 취약 지역 10곳 및 신규 가맹점 제안 36
- 3) 음악 분야 가맹점 수=0인 취약 지역 106곳 및 신규 가맹점 제안 36
- 4) 공예/문화체험 분야 최하위 취약 지역 20곳 및 신규 가맹점 제안 37
- 5) 도서, 공예/문화체험, 음악 관련 가맹점 수가 각각 0개인 지역 38

나. 잔액 사용이 가능한 가맹점 추천 결과 및 해석 39

- 1) user_1000의 시뮬레이션 39
 - 가) user_1000의 고객 EDA 분석 39
 - 나) user_1000 고객 선호 소분류 예측 40
 - 다) user_1000 고객 선호 가맹점 예측 40
 - 라) user_1000 고객 잔액, 지역, 코로나 위험 단계 필터 적용한 최종 가맹점 추천 41
- 2) user_5000 시뮬레이션 42
 - 가) user_5000의 고객 EDA 분석 42
 - 나) user_5000 고객 선호 소분류 예측 42
 - 다) user_5000 고객 선호 가맹점 예측 43
 - 라) user_5000 고객 잔액, 지역, 코로나 위험 단계 필터 적용한 최종 가맹점 추천 43

5. 활용 방안 44

가. 문제점 개선 방안 44

- 1) 가맹점 취약지역 제시 44
- 2) 잔액 사용 가맹점 추천 알고리즘 활용 및 확대 방안 46
- 3) 포스트코로나 시대 문화누리카드 활성화로서의 의의 48

나. 업무 활용 방안 48

- 1) 지자체별 가맹점 확대 활용방안 제시 48
- 2) 잔액 사용 가맹점 추천 알고리즘의 표준분석 모델화 48

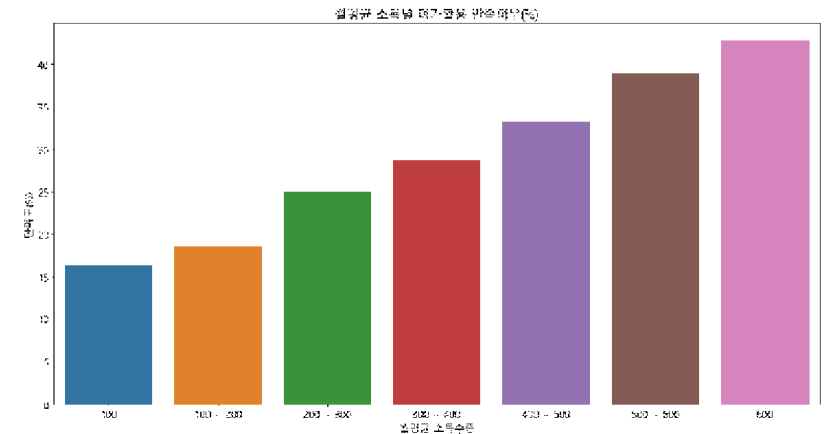
1. 분석 개요

가. 분석 배경 및 개요

1) 문화 격차 및 문화누리카드 사업 개요

- 시장 소득기준 상대적 빈곤율의 증가로 소득 양극화에 따른 문화 격차가 심화되고 있음. 더불어 SNS의 발달로 다른 이들이 문화생활 수준을 엿볼 수 있어 계층적 위화감의 심화 가능성도 존재함

(통계청의 2019년 사회조사 결과에 따르면, 19년 월 소득 600만 원 이상 가구의 여가 만족도가 42.7%인 것에 비해 월 소득 100만 원 미만 가구의 여가 만족도는 16.3%에 그치고 격차 또한 10년 전에 비해 2%p 증가)



[그림 1-1] 월 평균 소득별 여가활동 만족여부(%) (2019)

- 문화 격차는 문화 차이가 자본의 부족에서 발생하여 노력으로 극복하는데 한계가 있을 때 발생하며, 문화에 대한 상대적 박탈감이나 불평등으로 나타남. 국토교통부 정책 브리핑에 따르면 문화향유 기회나 문화향유 만족도는 일상적인 삶의 질뿐만 아니라 지역의 경제적·사회적 활력과 경쟁력에도 영향을 미치기 때문에 지역 간 문화격차는 시급히 해소되어야 함

* 문화의 상대적 박탈은 자신이 속한 사회 구성원 다수가 관습적으로 또는 널리 소비하는 문화에서 소외되는 것을 말함

- 문화누리카드 사업은 이러한 문화격차 완화와 국민 삶의 질 향상을 위해 기초생활수급자와 차상위계층을 대상으로 문화예술, 국내여행, 체육활동을

지원하는 사업이며 바우처 형식의 카드를 지급하는 형식

(추진 주제: 문화체육관광부, 한국문화예술위원회, 전국 광역·기초 지자체 및 17개 시도 지역 주관, 추진 근거: 문화예술진흥법 제15조의 3, 제15조의 4 국정과제 67. 지역과 일상에서 문화를 누리는 생활문화시대 (67-1. 국민의 기초 문화생활 보장), 지원 내용: 문화예술·여행·체육 분야에 사용 가능한 ‘문화누리카드’ 발급 지원 (1인당 연간 9만 원, 예산 범위 내 신청자 발급))

2) 문화누리카드 활성화 방안의 필요성

[표 1-1] 연도별(2015~2020년) 문화누리카드 발급 대상자 및 발급률

구분	문화누리카드 대상자(A)	예산대비 발급가능인원(명) (B)	대상자 대비 지원가능비율(%) (B/A*100)	실제 발급인원 (C)	대상자 대비 실제 발급비율(%) (C/A*100)	예산가능인원 대비 실제 발급비율(%) (C/B*100)
2015년	242만 명	1,579,712	65.3	1,378,368	57.0	87.3
2016년	251만 명	1,530,000	61.0	1,450,801	57.8	94.844
2017년	263만 명	1,610,600	61.2	1,523,506	57.9	94.6
2018년	256만 명	1,645,371	64.3	1,591,777	62.2	96.7
2019년	220만 명 (2,195,699명)	1,600,000	72.7	1,629,036	74.2	101.8
2020년	238만 명	1,610,000	67.8	-	-	-

자료원 : 한국문화예술위원회(2019.12) 「2019년도 통합문화이용권 이용자 만족도 조사 결과보고서」

- 예상 가능 인원대비 실제 발급인원(%) : 현재 예산에 비해 실제 발급 받는 비율을 의미. 19년 기준, 당시 예산 내에서 모든 사람이 발급받고 있다는 것을 알 수 있음
- 대상자 대비 실제 발급비율(%) : 문화누리카드 발급 대상자인 ‘기초생활수급자’와 ‘차상위계층’ 전체 인원 대비 실제 문화누리카드가 얼마나 발급되고 있는지의 비율. 이 비율은 매년 증가하고 있지만, 19년 기준 여전히 25% 이상의 대상자가 발급받지 못하고 있기 때문에 지속적인 예산 증액이 필요한 상황

[표 1-2] 문화누리카드 온·오프라인 가맹점 이용내역(2.1 ~ 6.5 기준)

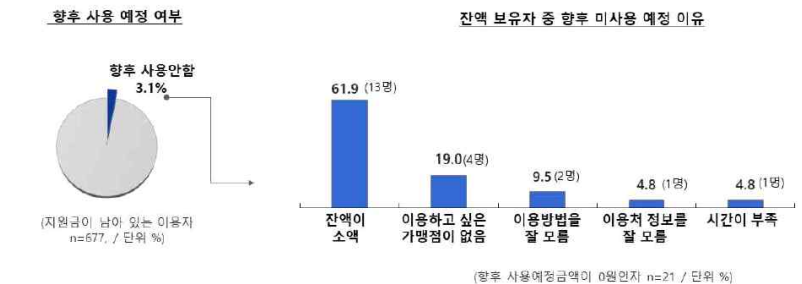
구분	2019년			2020년		
온라인	이용건수	약 19만 건	6.4%	이용건수	약 29만 건	12.6%
		약 49억 원	9.1%		약 78억 원	15.2%
오프라인	이용금액	약 491억 원	90.9%	이용금액	약 435억 원	84.8%
		약 540억 원	100%		약 513억 원	100%

자료원 : 문화체육관광부(2020.06) 「『집콕』 길어지자 온라인 문화 활동 2배 늘었다」

- 문화누리카드 발급 대상자 대비 지원 가능 비율은 19년 기준 72%로, 꾸준히 상승했으나 대상자를 전부 지원하기에는 현재 예산으로는 부족함. 그러나 최근 코로나19 장기화로 2020년 상반기 문화누리카드 이용률이 전년 대비 10%(27억 원) 감소했음. 더불어 이러한 양상이 지속되어 잔고가 소진되지 못한다면 향후 사업 평가에 반영되어 예산 감축의 위험이 존재함. 특히 코로나19 이후 제휴 가맹점의 90%를 차지하는 오프라인 가맹점 사용이 어려워져 하반기 이용률은 더 떨어질 것으로 추정됨

자료원 : 한국문화예술위원회, 「2019년도 통합문화이용권 이용자 만족도 조사 결과보고서」

Q 귀하께서는 문화누리카드에 현재 금액이 남아 있는데, 올해말까지 문화누리카드를 이용하지 않으려는 이유는 무엇입니까?

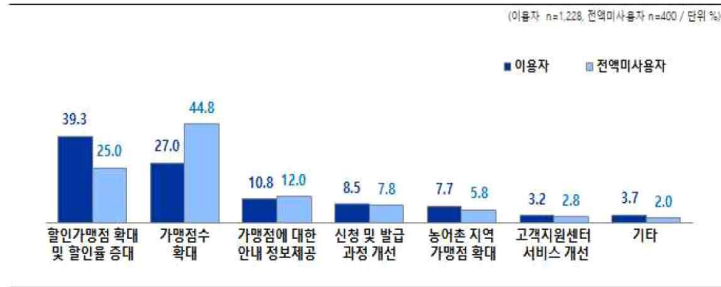


[그림 1-2] 잔액 보유자 중 향후 미사용 예정 이유

- 잔액 소진을 촉진하기 위하여 문화누리카드 측은 매년 연말에 잔액 소진 이벤트를 진행하지만, 해당 이벤트의 인지도가 낮아 사업 담당자들이 일일이 연락을 취하는 등의 리소스가 낭비되고 있음
- 또한 지역 간 사용 가능한 가맹점 인프라의 격차가 존재함. 서울이 약 3,200개의 가맹점이 존재하는 것에 비해 강원은 서울의 절반인 약 1,600개에 그쳐 지방에 거주하는 사용자들의 이용에 어려움이 있음
- 코로나19 이후 비대면 문화가 가속화되면서 이에 대한 여가생활 수요가 급증했으나, 문화누리카드의 오프라인 대비 온라인 가맹점 수는 현저히 적어 코로나19 시대의 온라인 가맹점을 다양하게 확보할 필요가 있음

(20년 9월 기준, 오프라인 대비 온라인 가맹점 3.25%)

자료원 : 한국문화예술위원회, 「2019년도 통합문화이용권 이용자 만족도 조사 결과보고서」



[그림 1-3] 문화누리카드 개선요청 사항

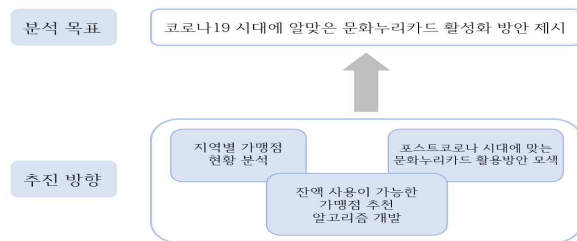
- 사용처에 대한 정보 전달 및 최신 정보 업데이트 부족으로 사용상의 불편함 존재

나. 분석 목적 및 방향

1) 목적

- 코로나19 시대에 실질적으로 활용 가능한 문화누리카드 활성화 방안 제시

2) 분석 추진방향



[그림 1-4] 분석 추진 방향

다. 분석 결과 활용 방안

1) 지역 문화 인프라를 활용한 지역 맞춤 가맹점 확대 방안을 통해 지역 간 문화격차 감소 기대

2) 잔액 사용이 가능한 가맹점 추천 알고리즘(이하 잔액 사용 가맹점 추천 알고리즘) 제시로 소규모 잔액 소진 가능. 잔액 사용 가맹점 추천 알고리즘을 2~3년 이상 사용할 시, 데이터 축적을 통해 더욱 정교화된 초개인화 '잔액 소진' 추천 시스템으로 활용 가능 기대

3) 잔액 사용 가맹점 추천 알고리즘을 현재 문화예술진흥원에서 진행하고 있는 모바일 애플리케이션 개발 사업에 적용한다면, 기존의 잔액 잔여율(17%)이 10% 이하(9.3%)로 감소할 것으로 기대

* 모바일 애플리케이션 사용이 원활한 65세 미만 청장년층의 50%가 해당 서비스를 사용했을 경우 예상 감소치

4) 데이터 기반 사용자 관점에서 문화누리카드 공식 애플리케이션 입찰 공고 개편안 아이디어 제시

5) 문화누리카드 사업 외 기존의 획일화된 사회복지형 바우처 공공 서비스 사업을 넘어, 복지사업 수혜 대상자의 인구통계적 특성과 거주 지역 특성을 고려한 맞춤형 바우처 서비스로 확장 기대

6) 코로나19 이후 문화 관련 시장의 급격한 변화에 비해 현재 문화누리카드 활용을 위한 적절한 방안은 존재하지 않음. 고객 수요에 부합하며 적시성을 고려한 현실적인 신규 가맹점 제휴 제안 및 코로나19 시대의 현명한 문화누리카드 활성화 방안 제시

2. 분석 데이터

가. 분석 데이터 목록

[표 2-1] 분석 데이터 목록

번호	분석 데이터	데이터 소스
1	대상자 좌표 데이터 (시군구별 대상자 수, 좌표)	- 기초생활보장수급자수 (한국사회보장정보원, 사회보장통계)

2	2020 문화누리카드 가맹점 데이터 (분류, 상호, 주소, 좌표)	- 문화누리카드 홈페이지 오프라인 가맹점 (22,000여개), 온라인 (700여개) 리스트 - Naver OpenAPI 사용 좌표 추출
3	소셜네트워크 문화이슈 데이터	- 18,131건 인스타그램 게시물 태그 데이터 #문화생활:5,135개의 게시물 (2020-09-15~2020-04-19) #취미:11438개의 게시물 (2020-09-17 ~ 2020-09-15) #코로나취미:1,177개의 게시물 (2020-09-15 ~ 2020-03-27) #코로나취미생활:381개의 게시물 (2020-09-17 ~ 2020-03-10)
4	문화누리카드 홈페이지 게시물 (문의사항, 이벤트, Q&A, 지역, 등록일, 제목)	- 지역주관처 알림글, 이벤트 글 - 약 1만 건의 문의사항 게시물 (문화누리카드 홈페이지)
5	기초생활보장수급자 현황 (날짜, 시도명, 시군구명, 연령, 수급)	- 국민기초생활보장수급자현황 (보건복지부, KOSIS)
6	여가활동 설문자료 (연령, 경제활동 상태, 가구소득별 하고 싶은 여가활동)	- 국민여가활동조사 (2019, 문화체육관광부, KOSIS) - 사회조사 (2019, 통계청, KOSIS)
7	문화활동 설문자료 (성별, 시도별, 문화 활동별 관람 횟수)	- 문화예술 및 스포츠 관람 현황 (사회조사 2019, 통계청)
8	가맹점 데이터 셋 (분류, 가맹점명, 지역, 물품 가격, 코로나)	- 문화누리카드 가맹점의 분류, 지역 및 각 품목별 금액 기반 데이터 셋
9	가상 고객 거래 데이터 셋 (고객 ID, 분류, 가맹점명, 구매 금액)	- 8번에서 직접 생성한 가맹점 데이터 셋 기반으로 구축
10	시군구별 인구통계학 자료	- 시군구별 남녀 성비 통계 (통계청, KOSIS)

		- 시군구별 연령별 통계 (인구총조사, 통계청, KOSIS)
11	소비자동향지수 현황	- 문화체육관광 소비자동향조사 (2020, 문화체육관광부, e-나라지표)

나. 데이터 상세 설명

1) 문화누리카드 대상자 데이터

- ☐ 출처 : 한국사회보장정보원, 사회복지통계 (행복e음), Naver OpenAPI 바탕으로 직접 생성
- ☐ 형태 : 정형
- ☐ 설명 : 20'2월 시군구별 기초생활수급자와 차상위계층의 수를 더하여 문화누리카드 발급 대상자 수를 구하고, Naver OpenAPI 활용하여 각 시군구의 좌표를 산출, 카드 발급 대상자 데이터와 합쳐 직접 생성

2) 문화누리카드 가맹점 데이터

- ☐ 출처 : 문화누리카드 홈페이지
- ☐ 형태 : 정형
- ☐ 설명 : 20'9월 홈페이지에 공개된 온·오프라인 가맹점 목록을 크롤링을 통해 수집

3) 소셜네트워크 문화이슈 데이터

- ☐ 출처 : 인스타그램
- ☐ 형태 : 비정형
- ☐ 설명 : 문화생활의 최근 트렌드를 보기위해 SNS(인스타그램)에 올라온 글 중 문화생활, 취미, 코로나취미, 코로나취미생활 등의 태그로 검색하여 나온 각각의 결과를 크롤링 하여 수집함
(#문화생활 : 5,135개, 2020-09-15~2020-04-19 / #취미 : 11,438개, 2020-09-17 ~ 2020-09-15 / #코로나취미 : 1,177개, 2020-09-15 ~ 2020-03-27 / #코로나취미생활 : 381개, 2020-09-17 ~ 2020-03-10)

4) 문화누리카드 홈페이지 게시물

- ☐ 출처 : 문화누리카드 홈페이지

- ☐ 형태 : 비정형
- ☐ 설명 : 2019년 문화누리카드 홈페이지에 올라온 지역 주관처 알림글과 이벤트, Q&A 글을 크롤링 하여 수집함
(공지사항 279개, 이벤트 196개, Q&A 10148개)
- 5) 기초생활보장수급자 현황
 - ☐ 출처 : 한국사회보장정보원, 사회복지통계 (행복e음)
 - ☐ 형태 : 정형
 - ☐ 설명 : 2020년 사회복지통계에 공개된 기초생활보장수급자의 연령별, 성별, 시도별 데이터를 CSV 파일 형태로 다운로드
- 6) 여가활동 설문자료
 - ☐ 출처 : 국민여가활동조사 (2019, 문화체육관광부), 사회조사 (2019, 통계청)
 - ☐ 형태 : 정형 설문조사 통계표)
 - ☐ 설명 : 문화체육관광부와 통계청에서 매년 발간하는 국민여가활동조사와 사회조사를 바탕으로 현재 우리나라 국민의 여가생활에 대한 전반적인 통계자료 취사 다운로드
- 7) 문화활동 설문자료
 - ☐ 출처 : 사회조사 (2019, 통계청)
 - ☐ 형태 : 정형 (설문조사 통계표)
 - ☐ 설명 : 통계청에서 매년 발간하는 사회조사를 바탕으로 현재 우리나라 국민의 문화 활동에 대한 전반적인 통계자료 취사 다운로드
- 8) 가맹점 데이터 셋
 - ☐ 출처 : 문화누리카드 홈페이지, 각 온·오프라인 가맹점 사이트 바탕으로 직접 생성
 - ☐ 형태 : 정형
 - ☐ 설명 : 문화누리카드 홈페이지에 공개된 온라인 가맹점 정보를 바탕으로 분류별 세부 가맹점을 조사하여 무작위로 상품 추출, 이를 바탕으로 가맹점별 물품의 최저값과 평균값을 구하여 물품 데이터 셋을 구축하고, 이를 바탕으로 가상 고객 거래 데이터 셋을 구축
- 9) 가상 고객 거래 데이터 셋
 - ☐ 출처 : 가맹점 데이터를 바탕으로 가상 고객 거래 데이터 셋 구축
 - ☐ 형태 : 정형
 - ☐ 설명 : 가맹점 데이터의 대분류, 소분류, 가맹점, 가격 정보를 랜덤으로

- 순차 선택하여 총 5000명의 가상고객에 대해 생성한 데이터 셋
- 10) 시군구별 인구통계학 자료
 - ☐ 출처 : 인구총조사 (2019, 통계청)
 - ☐ 형태 : 정형
 - ☐ 설명 : 통계청에서 매년 발간하는 인구총조사를 바탕으로 시군구별 성비, 연령별 통계 취사 다운로드
- 11) 소비자동향지수 현황
 - ☐ 출처 : 문화체육관광 소비자 동향조사 (2020 2분기, 문화체육관광부)
 - ☐ 형태 : 정형
 - ☐ 설명 : 문화체육관광부에서 분기마다 발간하는 문화체육관광 소비자 동향조사를 바탕으로 소비자동향지수 등의 필요 데이터 취사 다운로드

다. 데이터 정제 방안

- 1) 문화누리카드 대상자 데이터
 - ☐ 한국사회보장정보원 (사회보장통계)에서 제공하는 시군구별 기초생활보장수급자 인원수 자료와 차상위계층 인원수 자료를 합쳐 문화누리카드 전체 발급 대상자 수를 구함
 - ☐ Naver geocoding API를 사용하여 각 시군구의 위도와 경도 좌표를 구하여 카드 발급 대상자 데이터와 합쳐 시군구별 문화누리카드 대상자 인원수와 좌표의 데이터를 생성
- 2) 문화누리카드 가맹점 데이터
 - ☐ 문화누리카드 홈페이지 '사용하기' 탭의 오프라인 가맹점 검색을 통해 출력되는 가맹점 리스트의 정보를 Selenium을 사용한 크롤링을 활용하여 Dataframe 형식으로 구축
- 3) 소셜네트워크 문화이슈 데이터
 - ☐ SNS (인스타그램)에 올라온 게시물 중 문화와 관련된 글을 해시태그 ('문화생활', '취미', '코로나취미', '코로나취미생활') 검색을 통해 선별, Selenium을 사용한 크롤링을 활용하여 Dataframe 형식으로 구축
- 4) 문화누리카드 홈페이지 게시물
 - ☐ 문화누리카드 홈페이지 '알림방', '고객센터' 탭의 지역주관처 알림글, 이벤트, Q&A 게시판의 글을 Selenium을 사용한 크롤링을 활용하여 Dataframe 형식으로 구축

5) 가맹점 데이터 셋

- 문화누리카드 홈페이지에 공개된 온·오프라인 가맹점 정보를 바탕으로 분류별 세부 가맹점을 조사하여 무작위로 10개씩 상품명과 가격을 추출
- 가맹점별 물품의 최저값과 평균값을 구하여 물품 데이터 셋을 구축

6) 가상 고객 거래 데이터셋

- 만들어진 물품 데이터 셋을 기반으로 구매 총액이 9만원이 넘지 않는 5000명분의 가상 고객 거래 데이터 셋 직접 생성

7) 시군구별 인구통계학 자료

- KOSIS에 공개된 통계청 인구총조사 데이터 중에서 시군구별 성비, 연령별 인구 데이터 다운로드
- 데이터별로 다른 연연구간 통일하여 시군구별 인구 총합 산출

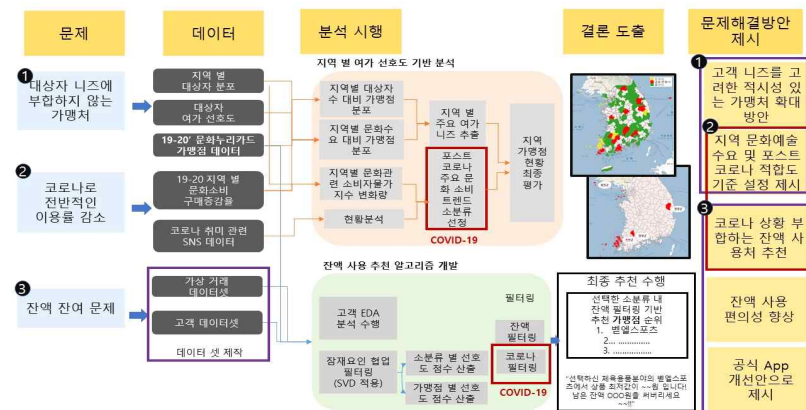
8) 소비자동향지수 현황

- 나라캐터 공개된 '소비자 동향 지수(CSI)' 데이터 중에서 2019-2020년도 2분기 데이터만 추출
- 기초생활수급자 및 차상위계층에 해당하는 평균 가구 월 소득 200만 원 미만에 해당하는 데이터만 추출하여 소득 기준 설정

3. 분석 프로세스

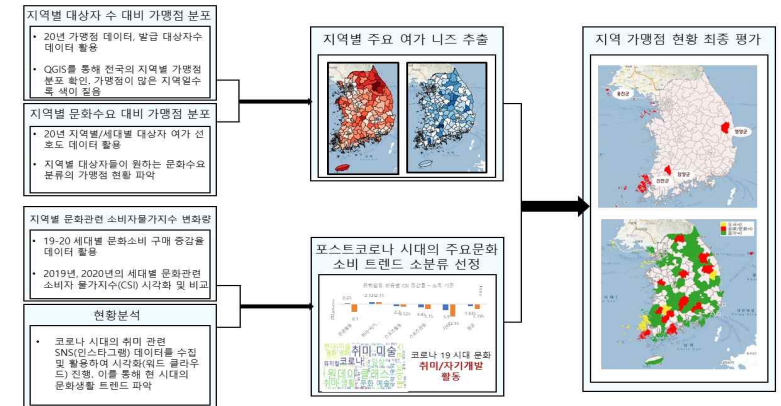
가. 분석 프로세스

1) 전체 분석 프로세스



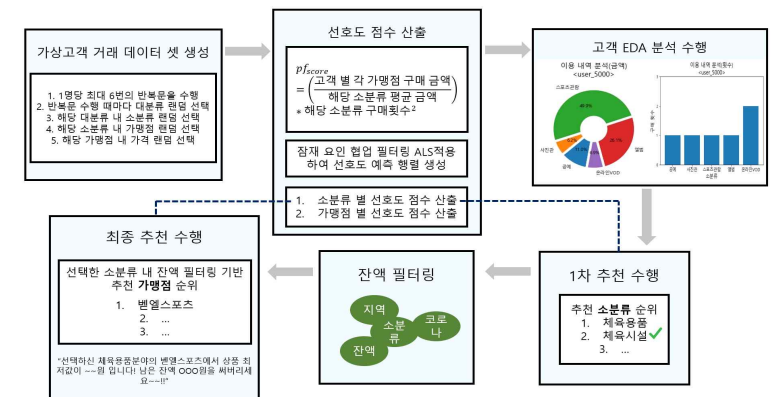
[그림 3-1] 전체 분석 프로세스

2) 지역별 여가 선호도 및 소비 변화 분석 프로세스



[그림 3-2] 지역별 여가 선호도 기반 분석

3) 잔액 사용 가맹점 추천 알고리즘 개발



[그림 3-3] 잔액 사용 가능한 가맹점 추천 알고리즘 개발 흐름도

나. 분석 내용 및 방법

1) 지역별 여가 선호도 및 소비 변화 분석 후 가맹점 현황 분석

가) 지역별 주요 여가 니즈 추출

- 시군구별 인구통계학 자료와 여가활동 설문자료 데이터를 결합하여 시군



[그림 3-10] 인스타그램 #코로나취미 크롤링 워드클라우드

□ 분석내용

: 클래스가 가장 많이 언급된 단어로 나타났으며 집콕, 셀프 등의 단어가 함께 나타난 것은 집에서 생활하며 즐길 수 있는 공예 분야가 트렌드가 되고 있음을 시사.

라) #코로나취미생활

□ 개요 : 인스타그램 게시물 중 #코로나취미생활 태그 검색 결과(381개)

□ 분석결과 및 시각화



[그림 3-11] 인스타그램 #코로나취미생활 크롤링 워드클라우드

□ 분석 내용

: 코로나취미와 비슷하게 집에서 할 수 있는 취미생활이 대부분. 자수, 만들기, 코바늘, 미술 등의 단어이며 집콕 등의 단어와 함께 나타남

다) 소비 니즈 및 소비 트렌드 기반 지역별 가맹점 평가지표 개발

□ 코로나19 전후, 문화소비에 대한 소비자들의 실질적인 소비 심리를 분석하기 위해 19-20' 2분기 소비자동향지수(CSI)를 활용하여 소비 심리를 정량적으로 파악함

* CSI(소비자동향지수) : 전 분기 대비 소비자 구매심리에 관한 지수를 종합적으로 평가한 국가공인평가지수

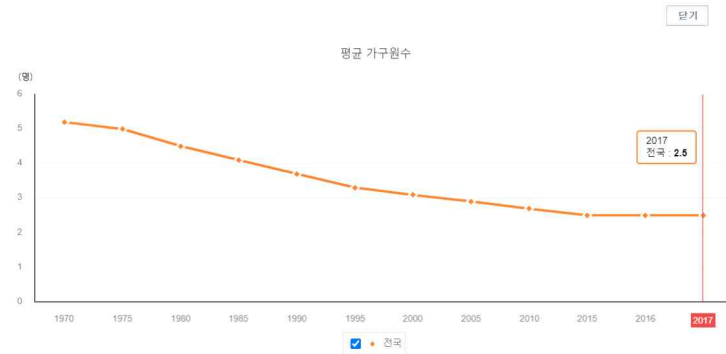
분류	항목	세부항목	2020년 02	2019년 02분기
오프라인	전체	전체	67.7	96
오프라인	가구주연령20대		56.8	97.2
오프라인	가구주연령30대		61.4	97.1
오프라인	가구주연령40대		59.9	97.4
오프라인	가구주연령50대		65.7	94.8
오프라인	가구주연령60대 이상		88.7	94.4
오프라인	가구원 수 1인 가구		71.1	97.9
오프라인	가구원 수 2-3인 가구		69.5	94.5
오프라인	가구원 수 4인 이상		59.4	96.9
오프라인	월평균 가·100만원 이하		84.7	96.9
오프라인	월평균 가·100만원-1		76	93.1
오프라인	월평균 가·200만원-2		69.5	93.3
오프라인	월평균 가·300만원-3		64.3	96.7

[그림 3-12] 19-20' 2분기 소비자 동향지수(종합 CSI), 문화센터

□ 소비자 동향지수 table 전처리. 평균 가구원수 2.5명, 1-3인 가구 중위소득 기준으로 차상위 계층에 속하는 99만원 미만, 100만원-199만원 미만 외에 나머지 월평균 가구 소득에 해당하는 200만원 이상 row 삭제

[표 3-1] 2020년 기준 1-3인 가구 중위소득

구분	1인 가구	2인 가구	3인 가구
기본 중위소득	175만 7194	299만 1980	387만 5770
차상위계층	87만 8597	149만 5990	193만 6289



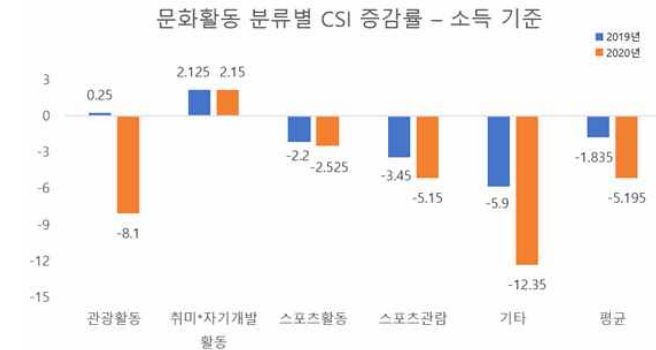
[그림 3-13] 평균 가구원수 2.5명(2017년 기준), 통계청_인구총조사

- 소비자 동향지수 table을 excel 피벗 테이블을 활용하여 년도 별 평균 CSI (소비자동향지수)를 제시된 문화 분류 기준에 따라 연령기준(전 연령), 소득 기준(월 평균 가구소득 200만원 미만) 평균 값 도출

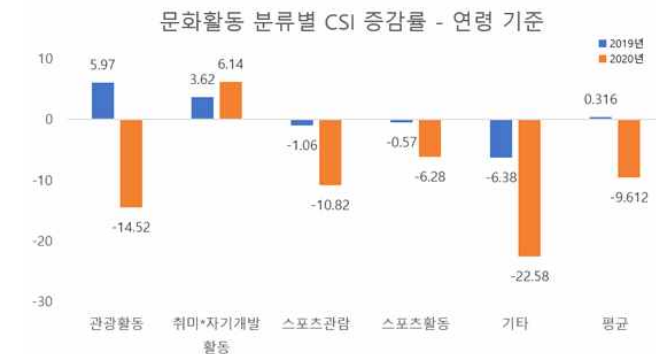
연령 기준	평균 : 2020_CSI	평균 : 2019_CSI
관광·여행비	85.6	106.78
국내 여행비	85.36	105.16
도서 구입비	107.54	104.44
미디어기기 구입비	104.74	102.8
스포츠 학원비	96.5	98.76
스포츠경기 관람료	89.18	98.94
스포츠기구 구입비	96.16	101.48
스포츠시설 이용료	86	99.54
아웃도어용품 구입비	96.22	97.94
오락시설 이용료	77.42	93.62
오락용품 구입비	104.04	100.04
오프라인 문화생활비	66.5	96.18
온라인 문화생활비	108.05	101.24
음악·미술 학원비	98.42	99.24
해외 여행비	85.7	99.7
총합계	92.28513514	100.3907

[그림 3-14] 문화 관련 CSI 19~20'

- 테이블을 막대그래프로 시각화



[그림 3-15] 문화활동 분류별 CSI 증감률 - 소득기준



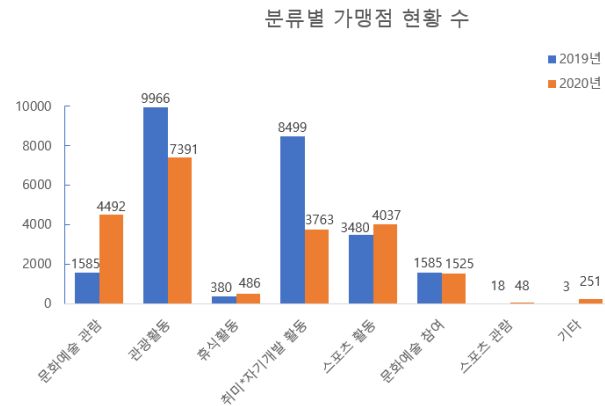
[그림 3-16] 문화활동 분류별 CSI 증감률 - 연령기준

- (1) 시각화 결과, 코로나19의 영향으로 전반적으로 전년도 대비 문화에 관련한 CSI 지수가 하락한 것을 알 수 있음. 특히 관광활동의 경우 매우 큰 폭으로 하락하여 코로나 19 시대에 적절한 문화누리카드 사용 방향이라 할 수 없음. 그에 반해, 취미, 자기개발 활동의 경우, 소득기준 및 연령기준 전체에서 수요지수 상승폭을 보임
- (2) 워드 클라우드와 문화활동 분류별 CSI 증감률을 확인한 결과, 코로나19 이후 전반적인 문화생활 소비 및 수요 하락에도 불구하고 **취미 및 자기개발 활동 영역**에 대한 관심과 문화 소비에 대한 니즈가 존재함을 파악
- (3) 문화동향 변화에 따라 문화누리카드 기존 가맹점을 평가하기 위해 소비자 동향지수 칼럼과 문화누리카드 가맹점 소분류를 대분류-소분류로 재설정

[표 3-2] 선호 문화 여가생활 분류표-(1)

소비자 동향지수 파악 대분류	문화누리카드 가맹점 소분류
오프라인 문화 생활비	오프라인 가맹점 전체
온라인 문화 생활비	온라인 가맹점 전체
스포츠 시설 이용료	체육시설
스포츠 경기 관람료	스포츠관람
아웃도어용품 구입비	체육용품
미디어 기기 구입비	음악, TV
관광여행비	숙박, 철도, 시외/고속버스, 국내항공, 여객선, 렌트카, 여행사, 관광명소, 휴양림/캠핑장, 온천, 체험관광, 테마파크
도서 구입비	도서
오락용품 구입비	X
오락시설 이용료	영화, 공연, 전시, 공예, 사진관, 문화체험, 직업체험, 동식물원, 체험관광, 테마파크,

□ 위의 분류별 19-20' 가맹점 현황 수 시각화



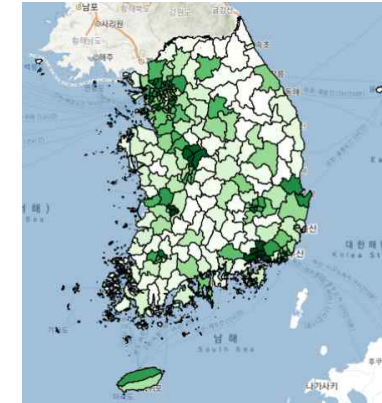
[그림 3-17] 문화활동 분류별 가맹점 현황

라) 지역별 문화누리카드 가맹점 분포 평가

전체적인 문화누리카드 발급 대상자와 가맹점 분포 현황부터, 코로나 시대 문화 트렌드를 반영한 현황 분석까지 시행

(1) 지역별 문화누리카드 발급 대상자와 오프라인 가맹점 현황 파악

- 지역별 문화누리카드 대상자 현황 파악
 - 문화누리카드 대상자 데이터를 이용해 시각화

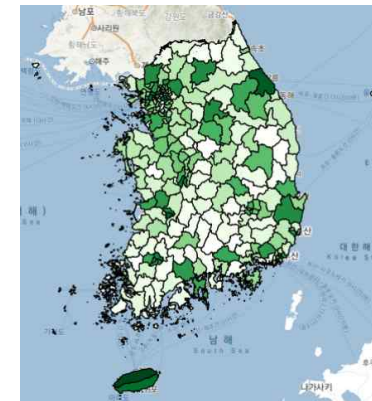


[그림 3-18] 20'2 월 기준 지역별 발급 대상자 수 분포 현황

*색이 짙을수록 더 많은 발급 대상자 분포

□ 지역별 문화누리카드 오프라인 가맹점 분포 현황

- 문화누리카드 가맹점 데이터를 이용해 시각화



[그림 3-19] 20'9 월 기준 지역별 문화누리카드 오프라인 가맹점 분포 현황

*색이 짙을수록 더 많은 가맹점 분포

□ 지역별 발급 대상자 수 대비 가맹점 수 분포 현황

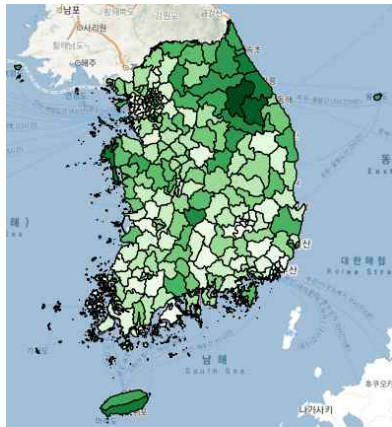
- 지역별 문화누리카드 발급 대상자 수 대비 오프라인 가맹점이 얼마나 분포되어 있는지 파악. 위의 2)번의 Q-GIS에서 위치 속성 결합으로 인해 산

출된 시군구별 오프라인 가맹점 개수를 이용하여 해당 지역의 발급 대상자 수로 나눈 후 백분율(%)로 나타내어 지역별로 시각화

- 식 : 지역별 가맹점 수 / 발급 대상자 수 * 100 (단위 : %)

	A	B	C	D	E	F
1	SIG_CD	SIG_KOR_NM	off_가맹점	발급대상자수	가맹점수/대상자수	
2	11110	종로구	316	7011	4.5072	
3	11140	중구	189	6779	2.788	
4	11170	용산구	95	9564	0.9933	
5	11200	성동구	78	13064	0.5971	

[그림 3-20] 시군구별 발급 대상자수 대비 오프라인 가맹점 수 백분율(%).xlsx



[그림 3-21] 시군구별 발급 대상자수 대비 오프라인 가맹점 수 백분율(%) 분포

*색이 짙을수록 대상자수 대비 많은 가맹점 분포

(2) 지역별 취미*자기개발 활동 분야 문화 가맹점 현황 파악

코로나 발생 이후 문화생활 관련 키워드 중심 SNS 워드 클라우드와 문화활동 분류별 CSI 증감률에 따라 유일하게 증가한 취미*자기개발 활동 분야에 대한 가맹점 현황 파악

□ 오프라인 가맹점 중 취미*자기개발 활동 분야 가맹점 현황 분석

오프라인 가맹점을 아래의 표와 같이 분류

[표 3-3] 선호 문화 여가생활 분류표-(2)

대분류	소분류
TV시청	TV, 영화
문화예술 관람	영화, 공연, 사진관, 전시
문화예술 참여	공예, 문화체험
스포츠 관람	스포츠관람
스포츠 활동	체육용품, 체육시설
관광활동	숙박, 철도, 시외/고속버스, 국내항공, 여객선, 렌트카, 여행사, 관광명소, 휴양림/캠핑장, 온천, 체험관광, 테마파크
컴퓨터 게임, 인터넷 검색 등	분류 없음
취미*자기개발 활동	도서, 음악, 공예, 문화체험
휴식활동	음악, 렌트카, 휴양림/캠핑장, 온천
사회 및 기타활동	분류 없음
기타	분류 없음

지역별로 취미*자기개발 활동에 해당하는 오프라인 가맹점 수 산출. 지역별 해당 문화 선호 대상자 수 대비 가맹점 수 비율 현황 분석

식 : 지역별 취미*자기개발 오프라인 가맹점 수 / 해당 문화 선호 대상자 수 * 100 (단위:%)

	A	B	C	D	E	F	G	H	I	J	K	L
1	SIG_CD	SIG_ENG_시도	SIG_KOR_주소	취미_자기:도서	음악	공예	문화체험	공예/문화체험				
2	11110	Jongno-gu	서울특별시 종로구	서울특별시	12.121	6.667	1.697	1.939	1.818	3.757		
3	11140	Jung-gu	서울특별시 중구	서울특별시	8.53	3.806	1.444	0.262	3.018	3.28		
4	11170	Yongsan-g	서울특별시 용산구	서울특별시	3.406	1.892	0.378	0	1.135	1.135		
5	11200	Seongdon	서울특별시 성동구	서울특별시	1.405	1.213	0	0	0.192	0.192		
6	11215	Gwangjin-	서울특별시 광진구	서울특별시	1.382	1.01	0.106	0	0.266	0.266		
7	11230	Dongdaen	서울특별시 동대문구	서울특별시	1.248	0.98	0.089	0.045	0.134	0.179		
8	11260	Jungnang-	서울특별시 중랑구	서울특별시	0.554	0.343	0.053	0	0.158	0.158		
9	11290	Seongbuk	서울특별시 성북구	서울특별시	1.577	0.986	0.039	0	0.552	0.552		
10	11305	Gangbuk-	서울특별시 강북구	서울특별시	0.927	0.447	0.032	0	0.447	0.447		

[그림 3-22] 시군구별 선호 대상자수 대비 오프라인 가맹점 수 백분율(%).xlsx

□ 최종 결과는 분석 결과에 제시

2) 잔액 사용 가맹점 추천 알고리즘 개발

가) 가상 고객 거래 데이터 셋 구축

□ 문화누리카드 온·오프라인 가맹점 정보를 토대로 분류별 세부 가맹점을 각각 조사하여 무작위로 10개씩 상품명과 가격을 추출하여 가맹점별 물품의 최저값과 평균값, 코로나 단계별 위험시설 여부와 함께 물품 데이터셋을 구축

[표 3-4] 가맹점 데이터(일부)

구분	대분류	소분류	가맹점	가맹점ID	온라인	지역	상품1	...	상품10	최저값	평균값	코로나
문화	도서	서점	세이북스	8	0	인천	3500		30000	1000	30700	1
문화	영화	영화관	CGV	12	0	전국	8500		9500	6500	7900	2
문화	공예	공예	서령필방	18	1		11500		8500	4000	9750	0
체육	체육용품	체육용품	4레인	82	1		33000		9800	38860	9800	0
체육	체육시설	체육시설	쌍쌍	93	0	서울	1500		2000	1400	1920	0

- ☐ 만든 가맹점 데이터 셋을 기반으로 구매 총액이 9만 원이 넘지 않는 5000명의 가상 고객 거래 데이터 셋 구축 (핵심 코드 부록 첨부)
1. 문화누리카드 인당 1년 평균 6~7회 사용하므로 가상고객 1명당 최대 6번의 반복문을 수행
 2. 반복문 수행 때마다 대분류 랜덤 선택
 3. 해당 대분류 내 소분류 랜덤 선택
 4. 해당 소분류 내 가맹점 랜덤 선택
 5. 해당 가맹점 내 가격 랜덤 선택. 만일 마지막으로 구매한 상품과 이전 구매한 상품과의 가격 합산이 9만원이 넘으면 해당 물품은 사지 않고 다음 반복문 수행
 6. 생성할 고객 수(5000번)만큼 수행

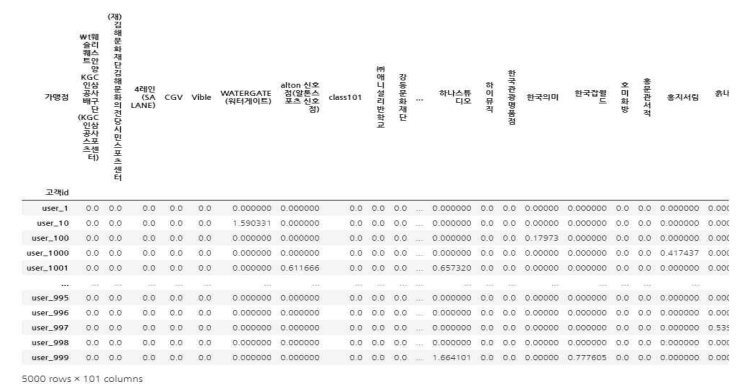
	고객id	대분류	소분류	가맹점	구매금액
0	user_1	체육용품	체육용품	운동장스포츠샵(운동장스포츠)	42000
1	user_1	도서	온라인서점	리디북스	8400
2	user_1	음악	앨범	뮤직플랜트	29800
3	user_2	체육시설	체육시설	알파카(주식회사 매스아시아)	1760
4	user_2	직업체험	직업체험	드림잡스쿨	12000
...
22250	user_5000	스포츠관람	스포츠관람	웨슬리웨스트 안양 KCG인삼공사 농구단(안양실내체육관)	34000
22251	user_5000	영화	온라인VOD	네이버버시리즈온	1925
22252	user_5000	영화	온라인VOD	네이버버시리즈온	2750
22253	user_5000	사진관	사진관	포토몬	4200
22254	user_5000	음악	앨범	뮤직플랜트	17800

22255 rows x 5 columns

[그림 3-23] 가상 고객 거래 데이터 셋

나) 가맹점 추천을 위한 선호도 점수 생성 및 테이블 재구조화

- 각 가상 고객이 구매한 소분류와 가맹점에 대해 얼마만큼의 선호도를 가지고 있는지 파악하기 위해 선호도 점수 생성
- 선호도 점수 = (고객 별 각 가맹점 구매 금액/해당 소분류 평균 금액)*해당 소분류 구매횟수²
- 고객이 소비한 구매 금액이 해당 소분류 내에서 얼마만큼의 가격우위를 가지고 있는지 알아보기 위하여 구매 금액을 해당 소분류 평균 금액으로 나눈 값을 점수에 반영. 이때 해당 소분류 평균 금액은 각 소분류의 데이터에서 이상치 제거¹⁾ 후 데이터로 평균을 계산함.
- : 고객 별 각 가맹점 구매 금액/해당 소분류 평균 금액) > 1 이면, 해당 소분류 평균금액보다 더 많은 금액을 내고 지출하여 선호가 있음을 의미
- : (고객 별 각 가맹점 구매 금액/해당 소분류 평균 금액) < 1 이면, 해당 소분류 평균금액보다 더 적은 금액을 냈음을 의미
- 해당 소분류 구매횟수를 제공한 것은 같은 소분류를 샀을 때 해당 분류에 더 높은 점수를 주고 우선순위로 나타나게 하기 위함. 즉, 구매횟수에 가중치를 크게 줌
- 최종적으로 선호도 점수가 높을수록 소분류와 가맹점에 선호가 높다는 것을 의미함
- 잠재요인 협업 필터링 수행을 위한 사용자-아이템 테이블로 재구조화



[그림 3-24] 고객-가맹점 평점 행렬

- 1) 이상치제거: $Q1 - 1.5 \cdot IQR$ 이하, $Q3 + 1.5 \cdot IQR$ 이상의 값 제거
IQR: 사분위수의 상위 75% 지점의 값과 하위 25% 지점의 차이

소분류	공연/전시	공예	네이버랩톤	사진관	서점	스포츠관람	약기	앨범	영화관	오디오북	온라인 VOD	온라인서점	직업체험	체육시설
고객id														
user_1	0.000000	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	1.461501	0.000000	0.000000	0.000000	0.493682	0.000000	0.000000
user_10	0.199005	1.133255	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.078834
user_100	1.592040	0.000000	0.0	0.000000	0.000000	0.000000	0.247177	0.799411	0.986842	0.000000	0.000000	0.000000	0.000000	0.000000
user_1000	0.000000	0.000000	0.0	1.248076	0.417437	0.508349	0.000000	0.000000	0.000000	0.000000	0.000000	0.320190	0.000000	0.000000
user_1001	1.194030	0.000000	0.0	0.657320	0.268352	0.000000	0.000000	0.000000	0.000000	0.754717	0.000000	0.000000	0.000000	0.141536
...
user_995	0.000000	0.000000	0.0	24.861672	0.000000	0.547452	0.000000	0.000000	0.000000	1.301275	0.205701	0.000000	0.000000	0.000000
user_996	0.398010	0.000000	0.0	0.000000	0.000000	4.066789	0.000000	0.838646	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
user_997	0.000000	0.000000	0.0	2.254857	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
user_998	0.000000	0.273544	0.0	0.000000	0.000000	6.882259	0.000000	0.490438	0.000000	0.532340	0.000000	0.411673	0.000000	0.000000
user_999	0.796020	0.000000	0.0	9.391688	0.000000	0.000000	0.000000	0.730750	0.000000	0.000000	0.000000	1.555210	0.000000	0.000000

5000 rows × 16 columns

[그림 3-25] 고객-소분류 평점 행렬

다) 1 차 추천 수행

- 잠재요인 협업 필터링 알고리즘을 통하여 고객 선호에 알맞은 소분류 추출
- 기존의 고객-소분류 행렬에서 고객과 소분류의 잠재 요인을 한 번씩 번갈아가며 학습하여 거대 희소행렬일 때 성능이 좋은 ALS 행렬분해 기법을 활용해 평점 0인(구매하지 않아서 선호도 계산 불가) 소분류의 선호도 예측
- ALS 의 프로세스를 따라 모델 학습 후 소분류 선호 점수 예측(사용기법에 자세히 기술, 적용 및 코드는 부록 첨부)

fran_id	1	2	3	4	5	6	7	8	9	10	11	12	13
user_id													
user_1	0.184185	0.157213	0.075507	0.220957	0.092396	0.165509	0.129651	0.977419	0.152629	0.108066	0.132260	0.911286	0.131530
user_10	0.906820	0.982221	0.104671	0.379226	0.114904	0.224799	0.152586	0.173617	0.199720	0.132684	0.196156	0.129554	0.204066
user_100	0.987448	0.259523	0.149777	0.324692	0.135675	0.250663	0.879879	0.966670	0.974640	0.150796	0.214775	0.119873	0.189167
user_1000	0.309007	0.320950	0.118649	0.993199	0.905858	0.970666	0.186102	0.209751	0.269079	0.156173	0.218723	0.177457	0.919127
user_1001	0.985922	0.321485	0.125812	0.988114	0.857046	0.325611	0.192927	0.226618	0.252099	0.951834	0.224383	0.186963	0.239405
...
user_995	0.251601	0.245234	0.096099	0.994809	0.118723	0.959068	0.146790	0.140302	0.165665	0.116904	0.977830	0.821977	0.165461
user_996	0.945486	0.221382	0.097569	0.349565	0.114813	0.994962	0.142347	0.964107	0.227252	0.137859	0.194854	0.114913	0.193462
user_997	0.161515	0.171678	0.067170	0.983979	0.071322	0.166498	0.076578	0.106654	0.106818	0.065573	0.097431	0.071881	0.097165
user_998	0.283981	0.936090	0.141456	0.370213	0.143136	0.998151	0.187809	0.944061	0.233782	0.141458	0.948441	0.120807	0.931450
user_999	0.971731	0.263638	0.121817	0.994429	0.114217	0.251739	0.133468	0.957290	0.195313	0.121610	0.182965	0.113467	0.988675

5000 rows × 16 columns

[그림 3-26] 그림번호, ALS 모델을 학습해 생성한 소분류 선호 점수

- user_1000 소분류 추천 예시

[표 3-5] 표번호, user_1000 결제 정보

고객id	대분류	소분류	가맹점	구매금액
------	-----	-----	-----	------

user_1000	직업체험	직업체험	드림잡스쿨	14000
user_1000	사진관	사진관	포토젠	15000
user_1000	스포츠관람	스포츠관람	웹슬리퀘스트 서울 SK나이트(서울특별시 학생체육관)	13000
user_1000	문화체험	클래스	도약닷컴	10000
user_1000	체육용품	체육용품	두산 베어스(㈜위팬)	19000
user_1000	도서	서점	홍지서림	7000

선호도

사진관	0.993199
스포츠관람	0.970666
체육용품	0.956629
직업체험	0.919127
서점	0.905858
클래스	0.892105
체육시설	0.351993
공예	0.320950
공연/전시	0.309007
영화관	0.269079
온라인VOD	0.218723
앨범	0.209751
약기	0.186102
온라인서점	0.177457
오디오북	0.156173
네이버랩톤	0.118649

[그림 3-27] 그림번호, user_1000 잔액 필터링 적용 전 소분류 추천 결과

- user_1000 고객에 대한 소분류 선호도를 기반으로 추천했을 시, 사진관, 스포츠관람, 체육용품 등 결제한 소분류와 비슷한 소분류를 적절히 추천해 주는 것을 볼 수 있음, 또한 구매한 소분류들 외에 체육시설과 공예에 다음 선호가 있을 것이라고 추정 가능
- user_5000 소분류 추천 예시

[표 3-6] 표번호, user_5000 결제 정보

고객id	대분류	소분류	가맹점	구매금액
------	-----	-----	-----	------

user_5000	공예	공예	한국관광명품점	7500
user_5000	스포츠관람	스포츠관람	웨슬리퀘스트 안양 농구단(안양실내체육관)	34000
user_5000	영화	온라인VOD	네이버시리즈온	1925
user_5000	영화	온라인VOD	네이버시리즈온	2750
user_5000	사진관	사진관	포토몬	4200
user_5000	음악	앨범	뮤직플랜트	17800

선호도	
온라인VOD	0.988432
스포츠관람	0.981860
사진관	0.969681
앨범	0.964766
공예	0.937121
공연/전시	0.300324
체육시설	0.288097
체육용품	0.255134
영화관	0.227066
클래스	0.200511
직업체험	0.193145
악기	0.179641
네이버웹툰	0.136028
서점	0.134944
오디오북	0.134322
온라인서점	0.115350

[그림 3-28] 그림번호, user_5000 잔액 필터링 적용 전 소분류 추천 결과

- user_5000 고객에 대한 소분류 선호도를 기반으로 추천 했을 시, 온라인 VOD, 스포츠관람 등 결제한 가맹점과 비슷한 소분류를 적절히 추천해 주는 것을 볼 수 있음, 또한 구매한 소분류들 외에 공연/전시, 체육시설에 다음 선호가 있을 것이라고 추정 가능

라) 고객 메타데이터를 바탕으로 필터링 후 최종 가맹점 추천

- 잠재요인 협업 필터링 알고리즘을 통하여 고객 선호에 알맞은 가맹점 추천
- 기존의 고객-가맹점 행렬에서 고객과 가맹점의 잠재 요인을 한번씩 번갈

아가며 학습하여 거대 희소행렬일 때 성능이 좋은 ALS 행렬분해 기법을 활용해 평점 0인 가맹점의 선호도 예측

- ALS 의 프로세스를 따라 모델 학습 후 소분류 선호 점수 예측(사용기법에 자세히 기술, 적용 및 코드는 부록 첨부)

	W(웨슬리 퀘스트안양 KGC 안양 농구 배구 단 (KGC인 삼공사 스 포츠센터))	(제)김래문 화재단 김 래문의 전담 시민 스포츠센 터	4라인(SA LANE)	CGV	Vible	WATERGATE (워터게이트)	alton 신호 점(알토스 포츠 신호 점)	class101	해운대리 한림고	강동문화 재단	...	하남시 디오	하이유직	한국관광 명품점	
0	0.058933	0.034202	0.015498	0.053568	0.000058	0.078441	0.022068	0.067166	-0.007603	0.059201	...	0.021655	-0.003115	0.051741	0
1	-0.015265	0.044757	0.069195	0.129833	0.000276	0.903149	0.123642	0.063818	0.021567	0.027048	...	0.123131	0.014979	0.062870	-0
2	0.103823	0.041522	0.069338	0.091268	0.000157	0.079106	0.058937	0.068850	0.011086	0.071485	...	0.059168	0.091739	0.092028	0
3	0.083872	0.090603	0.090096	0.035993	-0.000305	0.073126	0.076226	0.033029	0.083241	0.117816	...	0.078611	0.044565	0.131760	0
4	-0.014715	0.049691	0.022422	0.165157	0.000097	0.089756	0.891698	0.074071	-0.004035	0.062171	...	0.905666	0.031078	0.147658	0
...
4995	0.089800	0.070458	0.073403	0.117682	0.000087	-0.004886	0.044910	0.081166	0.003680	0.020177	...	0.062924	0.019531	0.114733	0
4996	0.082877	0.072620	0.096861	0.087664	-0.000036	0.010676	0.101207	0.088830	-0.024030	0.075579	...	0.054047	0.005075	0.075104	0
4997	0.042583	0.034650	0.011759	0.038570	-0.000176	0.045478	0.018276	0.013244	0.061080	0.050794	...	0.094574	0.072640	0.117867	0
4998	0.097087	0.076555	0.141611	0.056494	-0.000011	0.063774	0.126354	0.105919	-0.025157	0.068109	...	0.083698	0.028725	0.123344	0
4999	0.047936	0.054198	0.066491	0.102993	-0.000042	0.087218	0.074273	0.108158	0.003546	0.127192	...	0.917563	0.018890	0.114975	0

5000 rows × 101 columns

[그림 3-29] 그림번호, ALS 모델을 학습해 생성한 가맹점 선호 점수

- user_1000 가맹점 추천 예시

	가맹점	선호도	소분류
0	포토몬	0.956051	사진관
1	웨슬리퀘스트 서울 SK나이트(서울특별시 학생체육관)	0.898301	스포츠관람
2	두산 베어스(뽕위팬)	0.895933	체육용품
3	홍지서림	0.873129	서점
4	드림잡스	0.865254	직업체험
...
96	엘로이피아노	0.021939	악기
97	Vible	-0.000007	클래스
98	올리브그린	-0.000067	클래스
99	웨슬리퀘스트 수원 현대건설 힐스테이트(수원실내체육관)	-0.016870	스포츠관람
100	환율아트센터	-0.029991	공연/전시

[그림 3-30] 그림번호, user_1000 잔액 필터링 적용 전 가맹점 추천 결과

- 가맹점에 따른 추천을 수행한 결과, 소분류 선호와 연관된 가맹점 위주로 추천해 주는 것을 확인할 수 있음

- user_5000 가맹점 추천 예시

	가맹점	선호도	소분류
0	네이버시리즈온	0.986305	온라인VOD
1	웨슬리퀘스트 안양 KCG인삼공사 농구단(안양실내체육관)	0.954439	스포츠관람
2	뮤직플랜트	0.940143	앨범
3	포토론	0.852640	사진관
4	한국관광광명점	0.817359	공예
...
96	보앤봉	0.009667	공예
97	㈜애니설리반학교	0.003208	직업체험
98	Vible	0.000007	클래스
99	올리브그린	-0.000005	클래스
100	반디앤루니스	-0.002770	온라인서점

[그림 3-31] 그림번호, user_5000 잔액 필터링 적용 전 가맹점 추천 결과

- 가맹점에 따른 추천을 수행한 결과, 소분류 선호와 연관된 가맹점 위주로 추천해 주는 것을 확인할 수 있음

- 필터링 조건: 잔액, 코로나, 지역

- 필터링 조건 기준 및 근거 서술

* 잔액: 고객이 9만원 내에서 구매하고 남은 잔액으로 소비 가능한 가맹점을 추천하기 위함, 분석의 주목적, 더하여 잔액을 고려하지 않고 선호 가맹점을 추천해줄 수 있는 선택지 또한 제공

* 코로나: 체육시설을 좋아하는 고객에 대해 현 상황에 맞지 않은 '수영장'과 같은 곳을 추천하는 것을 방지하기 위함. 또한, 사회적 거리두기 단계 2단계가 시행될 경우, '스포츠관람'과 같은 추천을 방지하기 위해 사회적 거리두기 단계에 따른 적절한 가맹점 추천, 가맹점 별 0~3까지의 위험 단계를 할당한 후 고객의 선택에 따라 결과 필터링 후 추천

* 지역: 각 고객 별 방문할 수 있는 오프라인 가맹점을 추천하기 위함

- 필터링 함수 프로세스

1. 유저 id, 선호 소분류, 코로나 위험단계, 잔액 필터 적용 여부²⁾를 입력 받음
2. 유저의 메타데이터에서 유저의 지역 정보 불러옴

2) 잔액 필터 적용시, 각 고객 별 잔액에 한하는 상품이 있는 가맹점만을 추천, 잔액필터 적용 해제시, 해당 고객의 선호도만으로만 가맹점 추천

3. 고객별 잔액 계산
4. 고객이 선택한 소분류내의 가맹점만을 리스트 업
5. 지역 필터 적용
6. 코로나 필터 적용
7. ALS 점수로 정렬

- 최종 결과는 분석 결과에 게시

마) 사용 분석기법

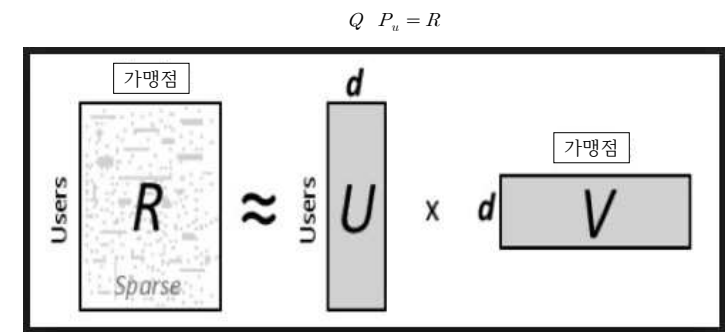
- 잠재요인 협업 기반 필터링(Latent Factor Collaborative Filtering)

: 사용자-아이템 평점 행렬에 잠재되어 있는 어떤 요인(factor)이 있다고 가정하고, 행렬 분해를 통해 그 요인들을 찾아내는 방식

- Matrix Factorization(행렬 분해)

: user-item 행렬을 2개 또는 3개 이상의 행렬로 분해하고 분해한 행렬을 변수로써 학습하는 방법, 잠재 요인 협업 필터링을 구현하는 가장 좋은 방법

: 평점 패턴으로부터 추론한 요인 벡터(d)들을 통해 사용자와 아이템의 특성을 잡아냄

[그림 3-32] MF 도식화³⁾

- * R : 우리의 선호도 점수를 반영한 사용자-소분류, 사용자-가맹점 행렬
- * Q_i : Item latent matrix로, 열은 아이템(가맹점 또는 소분류)을 의미하고 d는 잠재변수(특징값)를 의미
- * P_u : User latent matrix로, 행은 사용자(고객)를 의미하고 d는 잠재변수

3) 출처: <https://analytics4everything.tistory.com/104>

수(특징값)를 의미

* \hat{R} : 원 데이터를 추정한 행렬, 분해된 행렬을 다시 곱하여 예측 평점 행렬을 계산

: Latent Factor Matrix가 적절하게 학습이 되었다면 예측된 \hat{R} 은 R 과 유사한 결과를 낼 것이라 예상할 수 있음

: 이를 학습시키기 위한 Loss function

$$\min_{x^*, y^*} \sum_{u,i} (r_{ui} - x_u^T y_i)^2 + \lambda (\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2)$$

Loss function 계산식⁴⁾

- 원래 평점에서 예측된 평점을 빼준 값의 제곱 사용

: 해당 Loss Function을 최소화하는 사용자와 아이템의 Latent Factor Matrix를 찾아내는 것이 목표

□ ALS(Alternative Least Square: 교대 최소 제곱법)

: 사용자와 아이템의 Latent Factor를 한 번씩 번갈아가며 학습



[그림 3-33] ALS 도식화⁵⁾

: ALS에서의 Loss Function

$$\min_{x^*, y^*} \sum_{u,i} c_{ui} (p_{ui} - x_u^T y_i)^2 + \lambda (\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2)$$

ALS Loss function⁴⁾

- i (기존의 평점 행렬) 대신 c_{ui} 와 p_{ui} 가 새롭게 추가, 해당 식을 최소화하는 방향으로 학습

* $p_{ui} = \begin{cases} 1 & \text{if } r_{ui} > 0 \\ 0 & \text{if } r_{ui} < 0 \end{cases}$: 평점 예측 선호도를 나타냄, 평점 데이터가 존재하는 경우 선호함을 나타내는 1, 반대의 경우 0으로 바꿔 줌

* $c_{ui} = 1 + \alpha r_{ui}$: 해당 데이터의 신뢰도를 나타냄, 평점이 남아있지 않은 데이터에 대한 예측 값도 전체 Loss Function에 영향을 주도록 만들

* ALS 프로세스

1. 사용자 혹은 아이템의(한번 씩 번갈아가며 학습하므로) Latent Factor 행렬을 아주 작은 랜덤 값으로 초기화

2. 선호도 행렬 P 구하기

: 기존의 R 행렬에서 선호가 존재하는 부분은 1, 없는 부분은 0

3. 신뢰도 행렬 C 구하기

: 고객-소분류 선호도의 점수에 신뢰를 주기 위한 절차

: 선호가 0인 값에는 1, 0이 아닌 값은 일정 수준이 되도록 값 변환

4. Loss Function 구하기

5. Loss Function을 최적화시키는 Optimizer 작성

$$x_u = (Y^T C^u Y + \lambda I)^{-1} Y^T C^u p(u)$$

User Latent Factor Optimizer⁴⁾

$$y_i = (X^T C^i X + \lambda I)^{-1} X^T C^i p(i)$$

Item Latent Factor Optimizer⁴⁾

6. 잠재 행렬 학습

7. 학습 후 최종 선호도 행렬 예측

4. 분석결과

가. 문화누리카드 가맹점 취약 지역 제시 및 결과해석

□ 취미*자기개발 오프라인 가맹점 취약 지역 및 가맹점 제시

: 최종 취미*자기개발 오프라인 가맹점 취약 지역을 평가하여 스코어 순으로 정렬, 시각화하여 평가 점수 하위에 있는 지역에 가맹점 확대처 추천

4) 참고문헌 : Hu, Yifan & Koren, Yehuda & Volinsky, Chris. (2008). Collaborative Filtering for Implicit Feedback Datasets. Proceedings - IEEE International Conference on Data Mining, ICDM. 263-272. 10.1109/ICDM.2008.22.

5) 출처: <https://yeomko.tistory.com/4>

1) 취미*자기개발 관련 오프라인 가맹점 분포 현황

□ 지역별 취미*자기개발 관련 예상 수요자 대비 관련 가맹점 분포 현황

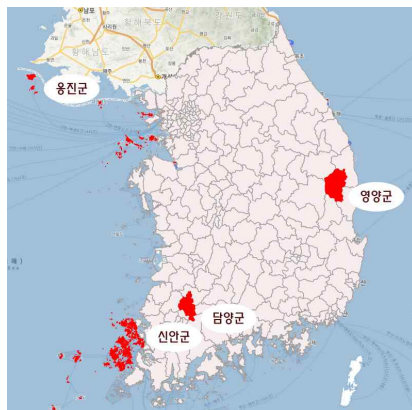
[표 4-1] 취미*자기개발 가맹점 부족 하위 10곳

하위 10곳			
no	지역	예상 수요 자	수요 대비 가맹점 수
1	인천광역시 옹진군	138	0
2	전라남도 담양군	452	0
3	전라남도 신안군	408	0
4	경상북도 영양군	196	0
5	경상남도 창원시 진해구	5548	0.18
6	전라남도 고흥군	880	0.227
7	전라남도 완도군	792	0.24
8	충청북도 청주시 청원구	4776	0.253
9	경상남도 창원시 마산회원구	5696	0.272
10	경상북도 봉화군	302	0.281

[그림 4-1] 취미*자기개발 관련 오프라인 가맹점 분포 현황

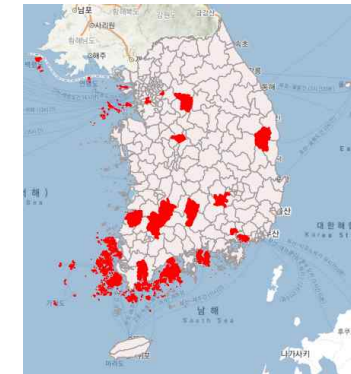
*색이 짙을수록 취미*자기개발 관련 오프라인 가맹점 부족 지역

□ 취미*자기개발 관련 가맹점 수가 0인 지역, 즉 최우선 가맹점 확대 지역으로 인천광역시 옹진군, 전라남도 신안군, 전라남도 담양군, 경상북도 영양군이 도출됨.



[그림 4-2] 취미*자기개발 관련 가맹점 수가 0인 지역

2) 도서 분야 최하위 취약 지역 10곳 및 신규 가맹점 제안



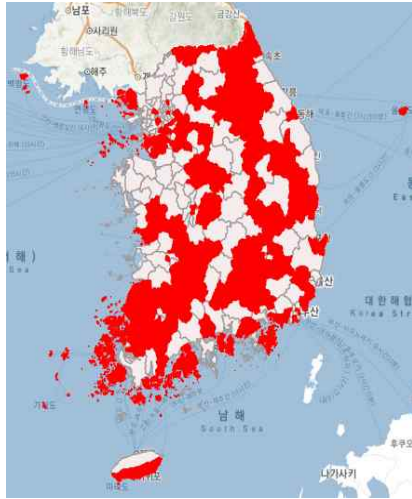
[그림 4-3] 도서 분야 최하위 취약 지역 20곳

[표 4-2] 도서 분야 취약 지역 10곳 및 신규 가맹점 제안

주소	취미*자기개발 활동 수요 예측(명)	수요 대비 가맹점수 - 도서	가맹점 제안
경상남도 창원시 진해구	5548	0.18	
전라남도 고흥군	880	0.227	문화책서림
전라남도 진도군	416	0.24	인우서점
전라남도 완도군	792	0.253	완도살롱
충청북도 청주시 청원구	4776	0.272	라임오렌지서점, 원더 박스
경상남도 창원시 마산회원구	5696	0.281	가고파서적
경상북도 봉화군	302	0.331	막장책방
경기도 성남시 중원구	4362	0.344	짱구네
부산광역시 영도구	1736	0.346	손목서가, 녹색광선, 오늘의 산책
경기도 안산시 상록구	4407	0.386	대동서적사동본점, 마 을상점생활관

*조사 결과, 파란색에 해당하는 부분은 지역에 관련 인프라가 풍부한 것에 반해, 문화누리카드 가맹점 협약이 잘 이뤄지지 않는 것으로 파악됨. 이 부분을 우선적으로 가맹점 협약을 이끌어내는 것을 추천

3) 음악 분야 가맹점 수=0인 취약 지역 106곳 및 신규 가맹점 제안

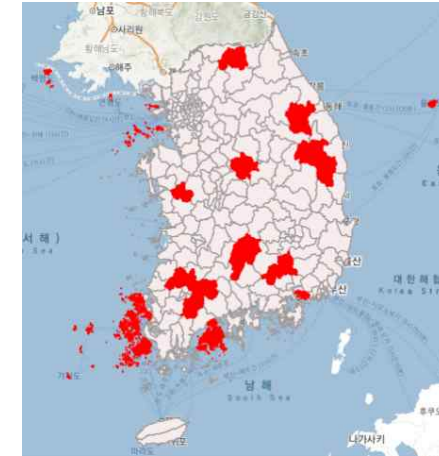


[그림 4-4] 음악 분야 가맹점 수가 0인 지역 106곳

[표 4-3] 음악 분야 가맹점 수가 0인 지역 106곳 및 신규 가맹점 제안

주소	취미*자기개발 활동 수요 예측(명)	수요 대비 가맹점수 - 음악	가맹점 제안
인천광역시 옹진군	138	0	다드림
전라남도 담양군	452	0	소리사랑 음악사
전라남도 신안군	408	0	지은소리사, 암태소리사
경상북도 영양군	196	0	지구레코드 음악사, 동서종합음악사
경상남도 창원시 진해구	5548	0	마음소리 램포레, 하트플러스 램포레
전라남도 고흥군	880	0	
전라남도 진도군	416	0	
전라남도 완도군	792	0	
충청북도 청주시 청원구	4776	0	192dollars, 시그널레코드
경상북도 봉화군	302	0	

4) 공예/문화체험 분야 최하위 취약 지역 20곳 및 신규 가맹점 제안

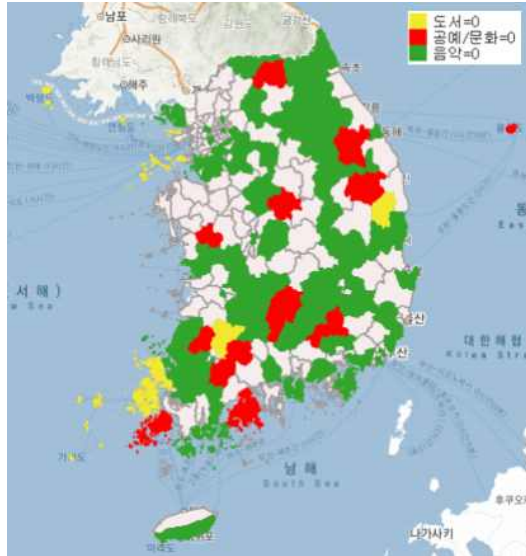


[그림 4-5] 공예/문화체험 분야 가맹점 수가 0인 지역

[표 4-4] 공예/문화체험 분야 가맹점 수가 0인 지역 및 신규 가맹점 제안

주소	취미*자기개발 활동 수요 예측(명)	수요 대비 가맹점수 - 공예, 문화체험	가맹점 제안
인천광역시 옹진군	138	0	등대박물관, 니오타니곤충사육관
전라남도 담양군	452	0	담양군 추억의골목, 한국대나무박물관, 담양곤충박물관
전라남도 신안군	408	0	소금박물관, 해양문화체험공간, 세계화석광물박물관
경상북도 영양군	196	0	목재문화체험장, 두들 문화마을
전라남도 고흥군	880	0	고흥우주천문과학관, 나로우주센터우주과학관, 힐링파크 쑥섬쑥섬
전라남도 진도군	416	0	
경상북도 울릉군	83	0	울릉예술문화체험장
경상북도 봉화군	302	0	봉화목재문화체험장, 봉화전통마을
충청남도 청양군	214	0	백제문화체험박물관, 청양목재문화체험관
전라남도 장성군	483	0	

5) 도서, 공예/문화체험, 음악 관련 가맹점 수가 각각 0개인 지역



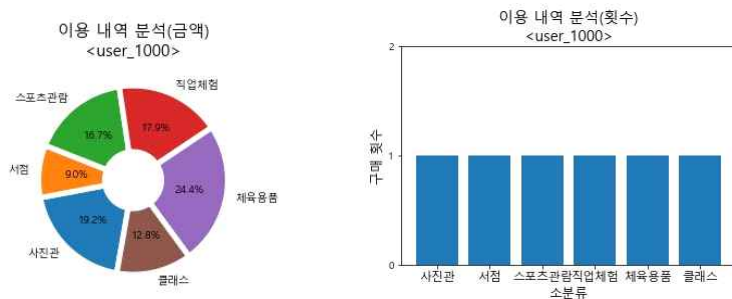
[그림 4-6] 분류 별 취약지역 최종 분석 결과 도출

* 각 분류에 해당하는 가맹점이 0곳인 곳을 추출

나. 잔액 사용 가능 가맹점 추천 결과 및 해석

1) user_1000 시뮬레이션

가) user_1000의 고객 EDA 분석

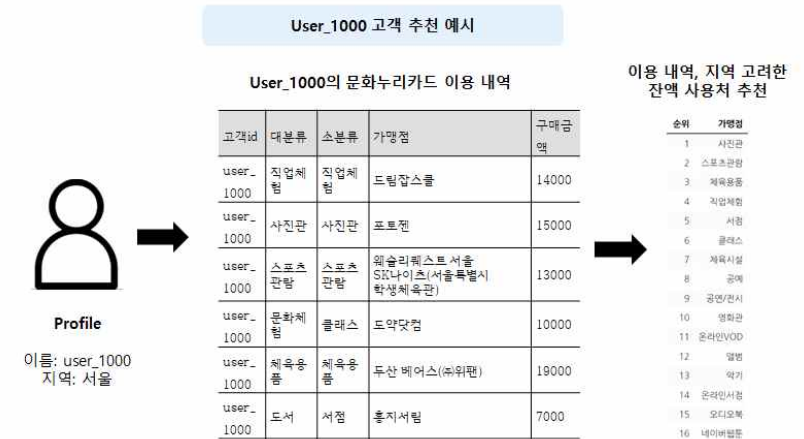


[그림 4-7] 그림번호, user_1000 고객 EDA 분석

- user_1000은 스포츠관람 외 5개의 분야에서 소비를 하였으며 파이차트를 통해 구매 금액의 비중이 가장 높은 소분류는 '체육용품'이고 그 다음 비중을 차지하는 소분류는 '사진관'임을 알 수 있음

- 막대그래프를 통해 각 소분류별 구매 횟수는 모두 1회임을 알 수 있음

나) user_1000 고객 선호 소분류 예측



[그림 4-8] 그림번호, user_5000 고객 1차 추천 결과

- 그 후 소분류에 관하여 1차 추천을 수행하였고, 그 결과 사진관, 스포츠관람, 체육용품 순으로 선호를 예측하였음

다) user_1000 고객 선호 가맹점 예측

순위	가맹점	소분류
1	포토젠	사진관
2	웹슬리퀘스트 서울 SK나이트(서울특별시 학생체육관)	스포츠관람
3	두산 베어스(취위팬)	체육용품
4	홍지서림	서점
5	드림잡스쿨	직업체험
...
96	노원문화예술회관	공연/전시
97	엘로이피아노	악기
98	Vible	클래스
99	올리브그린	클래스
100	웹슬리퀘스트 수원 현대건설 힐스테이트(수원실내체육관)	스포츠관람

[그림 4-9] 그림번호, user_1000 고객 가맹점 추천 결과

- 그 후 가맹점에 관한 추천을 수행하였고, 그 결과 사진관, 스포츠관람, 체육용품 순으로 선호를 예측하였던 소분류와 관련된 가맹점을 추천함을 볼 수 있음

라) user_1000 고객 잔액, 지역, 코로나 위험 단계 필터를 적용한 최종 가맹점 추천

```
>>> 유저 ID를 입력하세요 : user_1000
>>> 분류를 선택하세요 : 사진관
>>> 코로나 위험단계를 입력하세요 : 2
>>> 잔액필터 적용 여부를 선택하세요(0,X) : o
```

소분류	가맹점	지역
0 사진관	픽스	온라인
1 사진관	하나스튜디오	서울시
2 사진관	포토몬	온라인

[그림 4-10] 그림번호, user_1000 고객 최종 가맹점 추천(1)

- 고객이 소분류로 사진관을 선택하고, 위험단계 2로 선택하고, 고객의 잔액에 맞춘 최저값이 존재하는 가맹점을 추천하도록 선택했을 시의 결과
- 고객의 지역은 서울시로 설정되어 있었기에 다른 지역의 오프라인 가맹점은 추천대상에서 제외됨

```
>>> 유저 ID를 입력하세요 : user_1000
>>> 분류를 선택하세요 : 스포츠관람
>>> 코로나 위험단계를 입력하세요 : 2
>>> 잔액필터 적용 여부를 선택하세요(0,X) : o
```

소분류	가맹점	지역
-----	-----	----

[그림 4-11] 그림번호, user_1000 고객 최종 가맹점 추천(2)

```
>>> 유저 ID를 입력하세요 : user_1000
>>> 분류를 선택하세요 : 스포츠관람
>>> 코로나 위험단계를 입력하세요 : 1
>>> 잔액필터 적용 여부를 선택하세요(0,X) : o
```

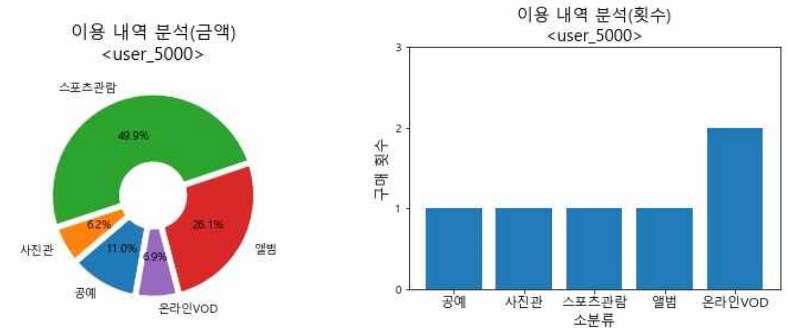
소분류	가맹점	지역
0 스포츠관람	웨슬리웨스트 서울SK나이스(서울특별시 학생체육관)	서울시
1 스포츠관람	울산현대 축구단	온라인
2 스포츠관람	인터파크 티켓 스포츠	온라인
3 스포츠관람	티켓링크 스포츠	온라인
4 스포츠관람	웨슬리웨스트 서울 우리카드 위비(서울장충체육관)	서울시
5 스포츠관람	웨슬리웨스트 서울 삼성전드스(잠실실내체육관)	서울시

[그림 4-12] 그림번호, user_1000 고객 최종 가맹점 추천(3)

- 스포츠관람 분류의 코로나 위험 단계 선택에 따른 결과 비교

2) user_5000 시뮬레이션

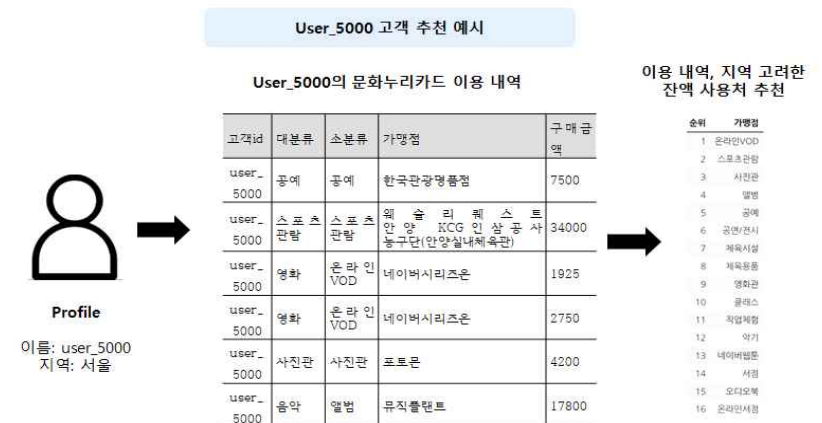
가) user_5000의 고객 EDA 분석



[그림 4-13] 그림번호, user_5000 고객 EDA 분석

- user_5000은 스포츠관람 외 4개의 분야에서 소비를 하였으며 파이차트를 통해 구매 금액의 비중이 가장 높은 소분류는 ‘스포츠관람’이고 그 다음 비중을 차지하는 소분류는 ‘앨범’임을 알 수 있음
- 막대그래프를 통해 각 소분류별 구매 횟수는 온라인VOD 2회를 제외한 나머지 모두 1회임을 알 수 있음

나) user_5000 고객 선호 소분류 예측



[그림 4-14] 그림번호, user_5000 고객 1차 추천 결과

- 그 후 소분류에 관하여 1차 추천을 수행하였고, 그 결과 온라인VOD, 스포츠관람, 사진관 순으로 선호를 예측하였음

다) user_5000 고객 선호 가맹점 예측

순위	가맹점	소분류
1	네이버시리즈온	온라인VOD
2	웨슬리퀘스트 안양 KCG인삼공사 농구단(안양실내체육관)	스포츠관람
3	뮤직플랜트	앨범
4	포토몬	사진관
5	한국관광명품점	공예
...
96	드림오피스 경관점	서점
97	보애평	공예
98	(주)애니설리반학교	직업체험
99	Vible	클래스
100	올리브그린	클래스

[그림 4-15] 그림번호, user_5000 고객 가맹점 추천 결과

- ☐ 그 후 가맹점에 관한 추천을 수행하였고, 그 결과 온라인VOD, 스포츠관람, 앨범 순으로 선호를 예측하였던 소분류와 관련된 가맹점을 추천함을 볼 수 있음

라) user_5000 고객 잔액, 지역, 코로나 위험 단계 필터를 적용한 최종 가맹점 추천

```
>>> 유저 ID를 입력하세요 : user_5000
>>> 분류를 선택하세요 : 온라인VOD
>>> 코로나 위험단계를 입력하세요 : 2
>>> 잔액필터 적용 여부를 선택하세요 (0,X) : o
```

소분류	가맹점	지역
0	온라인VOD	네이버시리즈온 온라인
1	온라인VOD	티빙 온라인

[그림 4-16] 그림번호, user_5000 고객 최종 가맹점 추천(1)

- ☐ 고객이 소분류로 온라인VOD를 선택하고, 코로나 위험단계 2로 선택하고, 고객의 잔액에 맞춘 최저값이 존재하는 가맹점을 추천하도록 선택했을 시의 결과
- ☐ 고객의 지역은 서울시로 설정되어 있었고 온라인VOD는 가맹점 데이터에서 온라인 가맹점에 속하기 때문에 지역은 모두 온라인으로 표시

```
>>> 유저 ID를 입력하세요 : user_5000
>>> 분류를 선택하세요 : 앨범
>>> 코로나 위험단계를 입력하세요 : 2
>>> 잔액필터 적용 여부를 선택하세요 (0,X) : o
```

소분류	가맹점	지역
0	앨범	뮤직플랜트 온라인
1	앨범	서울레코드 온라인

[그림 4-17] 그림번호, user_5000 고객 최종 가맹점 추천(2)

- ☐ 고객이 소분류로 앨범을 선택하고, 코로나 위험단계를 2로 선택하고, 고객의 잔액에 맞춘 최저값이 존재하는 가맹점을 추천하도록 선택했을 시의 결과
- ☐ 처음 구축했던 가맹점 데이터 셋에서 소분류 앨범에 대한 가맹점은 총 3곳으로 온라인 가맹점 2곳과 부천시의 오프라인 가맹점 1곳
- ☐ 해당 고객의 지역은 서울시이므로 자동으로 부천시 가맹점은 추천에서 제외

5. 활용 방안

가. 문제점 개선 방안

1) 가맹점 취약지역 제시

- ☐ 취미-자기개발 문화누리카드 가맹처 수가 0곳인 4지역(인천광역시 옹진군, 전라남도 담양군, 전라남도 신안군, 경상북도 영양군)에 대한 가맹처 확대 방안을 제안. 총 33건의 소외지역 신규 가맹처 확대 방안을 우선제시 및 시각화
- ☐ 옹진군 신규 가맹점 제시

no	신규 가맹처 제시
1	백령기독교역사관
2	등대박물관
3	조기역사관
4	연평도안보교육장
5	마음풍경갤러리
6	보산미술관
7	옹진군대이작도해양생태관

[그림5-1] 옹진군 신규 가맹처 제시

□ 영양군 신규 가맹점 제시

[표5-1] 영양군 신규 가맹점 제시

no	신규 가맹처 제시
1	지훈문학관
2	영양문화원(영양작은영화관)
3	영양산촌생활박물관
4	장계항문화체험교육
5	전통음식체험관

[그림5-1] 영양군 신규 가맹처 제시

□ 담양군 신규 가맹점 제시

[표5-2] 담양군 신규 가맹점 제시

no	신규 가맹처 제시
1	담양추억의골목
2	호남기후변화체험관
3	갤러리 아트14
4	남촌미술관
5	한국대나무박물관
6	담양우표박물관
7	미암박물관
8	한국가사문학관

[그림5-2] 담양군 신규 가맹처 제시

□ 신안군 신규 가맹점 제시

[표5-3] 신안군 신규 가맹점 제시

no	신규 가맹처 제시
1	소금박물관
2	신안갯벌센터슬로시티센터
3	세계조개박물관
4	에로스서각박물관
5	저녁노을미술관
6	신안군기록관
7	세계화석광물박물관
8	천사상미술관
9	신안철새전시관

[그림5-3] 신안군 신규 가맹처 제시

- 도서 분야 취약 지역 하위 18곳에 대해 추가 가맹점이 생긴다면 총 55,000여명에 대한 소비 니즈 충족 가능 예상. 이는 문화누리카드 발급 대상자의 3.43%에 해당하는 인원 수
- 음악 분야와 공예/문화체험 분야 취약 지역 하위 18곳에 대해 추가 가맹점이 생긴다면 각각 38,000여명과 7535명에 대한 소비 니즈 충족 가능 예상. 문화누리카드 발급 대상자의 2.37%와 0.4%에 해당하는 인원 수 . 특히 이 두 취약 지역은 해당 가맹점이 0개이므로 더욱 극적인 효과 기대

2) 잔액 사용 가맹점 추천 알고리즘 활용 및 확대 방안

- 현재 국가종합물자조달 시스템인 '나라장터'를 통해 한국문화예술위원회에서 문화누리카드의 모바일 앱 개발 공고를 냈고(2020.08.21.) 개찰 완료되어 사업자 선정까지 완료(2020.08.25.)
- 사업 선정자는 내년 1월 상용화를 목표로 약 4개월 동안 애플리케이션 개발
- 모바일 애플리케이션에서 잔액 사용 가맹점 추천 알고리즘의 활용 방안은 아래 예시와 같음



[그림 5-4] 모바일 앱에서의 잔액 사용 가맹점 추천 모델 활용 방안 예시

- 가) 현재 제작되는 모바일 애플리케이션에서 해당 모델이 활용되도록 함
- 나) 문화누리카드 잔액 표시 기능
- 다) 사회적 거리두기 단계별로 달라지는 잔액 사용 가맹점 확인
- 라) 본인 잔액에 맞는 가맹점 추천 여부 설정 기능
- 마) 분류별 추천 여부 설정 기능
- 바) 개인 맞춤 설정에 따라 추천 가맹점 표시
- 사) 하단에 본인의 최근 구매 내역을 표시함으로써 추천 가맹점 선택에 도움

- 한국문화예술위원회에서 공정한 통합문화이용권(문화누리카드) 모바일 애플리케이션 개발 사업의 제안요청서에 대한 제언 (기능 요구사항 (SFR) 중심)

- 가) 본인인증 기능 (SFR-003)
- : 현재 디지털 원패스, 아이핀, 공인인증서, 휴대폰 인증 등 4가지 방법으로 본인인증 가능. 이에 지문인식, 얼굴인식 등의 바이오 인식 방법 추가 제언

- 나) 카드발급 기능 (SFR-004)
- : 현행(우편수령, 지점수령) 외 모바일 상의 가상 카드 발급 기능 제언 (애플리케이션 사용)

- 다) 시각 장애인용 점자 스티커 신청 (SFR-008)
- : 청각 장애인을 위한 음성 안내 서비스 추가 제언

- 라) 수령등록/인터넷 사용등록 (SFR-009)
- : 현재 수령등록과 인터넷 사용등록이 이원화 되어 있는 것을 일원화 추진

3) 포스트코로나 시대, 문화누리카드 활성화 방안으로서의 의의

- '위드 코로나(With Corona)' 시대의 새로운 현실(New Reality)에 적응해 가는 시점에서, 문화 활동 위축에 따라 문화 소외계층의 삶의 질 하락과 사회 우울감이 심화됨
- 코로나 사태 이전(Before Corona)과는 다른 코로나 이후(After Corona)의 새로운 세계의 세팅(setting)에 맞춰 지역별 문화 선호도와 새로운 문화 트렌드를 분석하여 코로나 시대에 문화소외계층이 현실적으로 즐길 수 있는 문화 인프라를 구축하기 위한 방안을 제시하고, 잔액 사용 가맹점 추천 알고리즘을 개발하여 문화누리카드 사용자에게 친화적인 프로세스로의 전환과 상황에 알맞은 문화 활동을 추천 실현
- 니즈와 적시성, 사용 편의성을 고려하여 문화누리카드 대상자와 담당 사업부 모두가 만족할 수 있는 문화누리카드 활성화
- 결과적으로, 코로나19로 지친 문화소외계층의 마음을 위로하고 삶의 즐거움을 줄 것이라 기대

나. 업무 활용 방안

1) 지자체별 가맹점 확대 활용방안 제시

- 3 Step 가맹점 확대 컨설팅 방안 제시



[그림 5-5] 3 Step 가맹점 확대 컨설팅 방안 예시

Step 1. 취약지역으로 평가받은 지자체 컨택

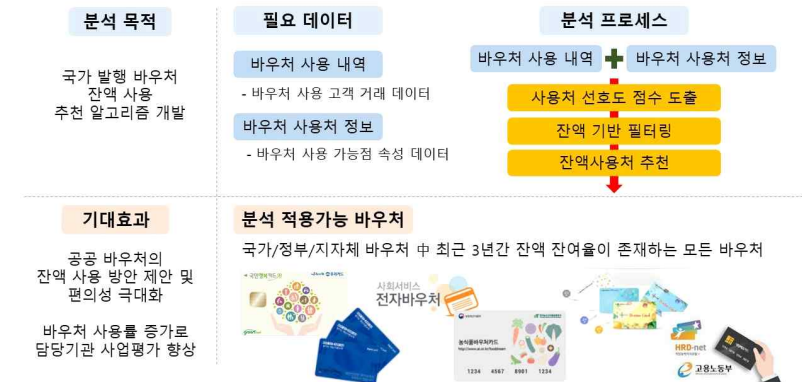
(Ex. 취미-자기개발 분야 대상자의 니즈 대비 가맹점 최하위 점수를 받은 인천 옹진군, 전라남도 신안군과 담양군, 경상북도 영양군을 최우선으로 컨택)

Step 2. 각 지자체에 필요한 신규 문화누리카드 가맹점 관련 컨설팅 진행

(Ex. 전라남도 담양군의 경우, 452명의 대상자들이 도서, 음악, 공예, 문화생활을 원하는 것으로 예측 됐으나 담양군에는 현재 관련 가맹점이 존재하지 않음.)

Step 3. 가맹점에 협약 제안

(Ex. 가맹점 후보 리스트 업에 포함된 약 10여 곳의 가맹점들과 협약 추진. 문화누리카드 대상자 452명의 수요 충족 기대)

2) 잔액 사용 가맹점 추천 알고리즘의 표준분석 모델화**□ 잔액 사용 가맹점 추천 알고리즘의 표준분석 모델화 제안****잔액 사용 추천 알고리즘 표준분석모델화 제안**

[그림 5-6] 잔액 사용 추천 알고리즘 표준분석모델화 제안

: 문화누리카드 잔액 사용 가맹점 추천 알고리즘을 타 정부 사회복지 바우처에도 적용할 수 있을 것으로 기대함. 해당 알고리즘을 국가 발행 바우처의 잔액 사용 추천 알고리즘 표준분석모델로 확장하여, 기존에 잔액이 발생하던 모든 국가·정부·지자체 바우처 애플리케이션에 잔액 사용 추천 서비스를 제공할 수 있을 것이라 예상함. 기존 바우처의 고객 이용 내역 및 바우처 사용처 속성 데이터를 기반으로 사용자 선호도 점수를 추출하여 잔액 사용처를 추천하는 표준 분석모델을 확립. 공공 바우처의 잔액사용 방안 제안으로 사용자의 편의를 극대화하고, 바우처 사용률을 높여 해당 바우처 담당기관 사업 평가 향상에도 기여할 수 있을 것이라 기대함.

* 정부주도 바우처 사업 약 30개에 적용 가능⁶⁾

(표 2-5) 바우처 사업의 소관부처별 분류

소관부처	개수	예시
고용노동부	2	실업자 내일배움카드, 재직자 내일배움카드
교육부	3	유아학비, 방과후학교자유수강권, 치료지원서비스
국가유공자	2	국가유공자 등 LTC차량 세금인상분 지원, 장기복무제대군인취업지원
문화체육관광부	2	통합문화이용권, 스포츠강화이용권
보건복지부	18	인0-5세보육료지원사업, 방과후보육지원, 시간영양보호료지원, 장애인보육료지원, 다문화보육지원, 지역사회서비스투자사업, 산모신생아건강관리지원사업, 가사간접지원사업, 노인돌봄종합서비스, 노인단기가사서비스, 치매관리와 가족돌봄지원서비스, 임신출산전면지원(건강보험), 임신출산전면지원(형소년산모), 저소득층 기저귀조세분류 지원, 장애인활동지원, 발달장애서비스, 발달장애인 부모상담지원, 언어발달지원사업
산림청	1	산림복지서비스이용권
산업통상자원부	1	에너지바우처
여성가족부	1	아이돌봄서비스
계	30	

[그림 5-7] 정부 및 지자체 바우처

6) 출처: 보건복지부 바우처사업 정책효과

[부록]

1. 사례조사 한 내용

가. 관련 기사 리스트

- "영화관 문 닫고 외출도 못하는데"...문화누리카드 '그림의 떡' (2020.08.31.)
<https://www.yna.co.kr/view/AKR20200828143100064?input=1195m>
- 문화누리카드 전액 사용을 절반에 못 미쳐 (2019.09.23.)
<http://news.heraldcorp.com/view.php?ud=20190923000090>
- '카드사용 자체가 숙제' 누리지 못할 문화누리카드 (2020.02.11.)
<https://www.kado.net/news/articleView.html?idxno=1009620>
- 카드 쓰라고 전화 오는데 쓸 수 있는 곳 없어 (2020.02.18.)
<https://www.kado.net/news/articleView.html?idxno=1010583>
- 전남 문화누리카드 상반기 이용률 52%...전국 1위 (2020.07.10)
<https://www.mk.co.kr/news/politics/view/2020/07/708372/>
- 지역문화재단, 코로나 여파에 문화누리카드 소비 촉진 이벤트 시행 (2020.09.02.)
<http://www.newspim.com/news/view/20200902000877>
- 코로나 '집콕' 길어지자 온라인 문화 활동 2배 늘었다 (2020.06.08.)
<http://www.korea.kr/news/policyNewsView.do?newsId=148873193>

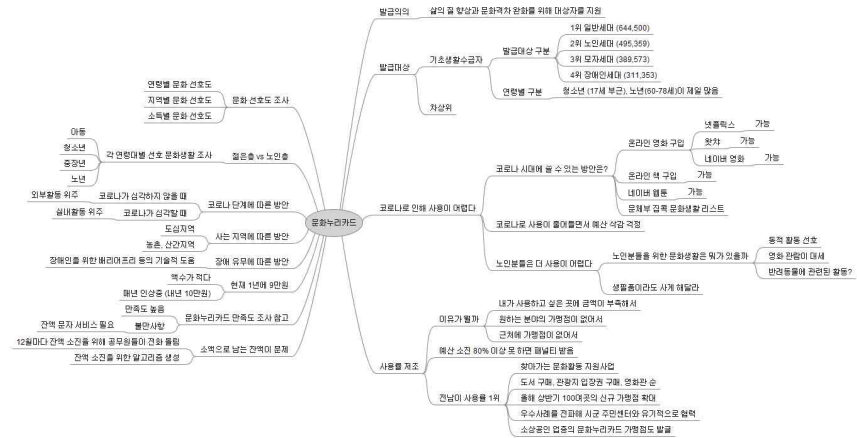
나. 참고 자료

- 문화체육관광부 2021년 예산안 정보그림(인포그래픽 pdf 합본)
- 문화체육관광부 외 - 코로나 일상 속 비대면 예술 지원방안
- 2020년도 문화체육관광부 성과관리 시행계획
- 2019년도 통합문화이용권(문화누리카드) 이용자 만족도 조사 결과보고서
- 코로나바이러스감염증-19 중앙재난안전대책본부 정례브리핑(20.09.01)
- Hu, Yifan & Koren, Yehuda & Volinsky, Chris. (2008). Collaborative Filtering for Implicit Feedback Datasets. Proceedings - IEEE International Conference on Data Mining, ICDM. 263-272.

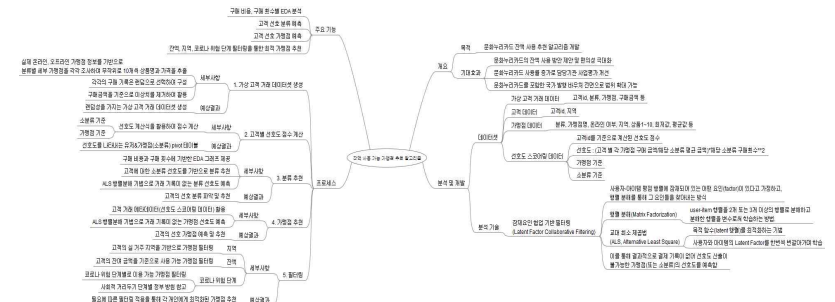
10.1109/ICDM.2008.22.

2. 주제설계를 위한 마인드맵

1) 문화누리카드 마인드맵



2) 잔액 사용이 가능한 가맹점 추천 알고리즘 개발 마인드맵



3. 분석 상세코드

1) 문화누리카드 오프라인 가맹점 데이터 크롤링

```
1 addr = 'https://www.mnuri.kr/useOfCard/offlineMerchants.do' #문화누리카드 오프라인 가맹점 페이지
2 driver.implicitly_wait(5)
3 driver.get(addr)
4 driver.find_element(By.XPATH, '//*[@id="contents"]/div[4]/a').click() #전체분류에 대해 검색

1 df = pd.DataFrame(columns=['구분', '분류', '가맹점', '주소', '전화번호', '할인&이용정보'])
2 num = 1
3 df_list = [] # 디렉터리로 저장된 각 행(가맹점)들의 list
4 max_page = 2214 #마지막 페이지
5
6 while num < max_page:
7     for j in range(3,8): #한 페이지에 나타나는 페이지 번호
8         for k in range(1,12): #한 페이지에 나타나는 가맹점 행
9             value_list = []
10            for i in [1,2,3,4,5,6]: # 각 가맹점 행에 대한 컬럼 인덱스
11                value = driver.find_element(By.XPATH, '//*[@id="contents"]/div[5]/table/tbody/tr[{}]/td[{}].format(k, i))
12                value_list.append(value.text)
13
14            # 각 가맹점 컬럼별 데이터 디렉터리 형태로 저장
15            v1 = {'구분':value_list[0],
16                  '분류':value_list[1],
17                  '가맹점':value_list[2],
18                  '주소':value_list[3],
19                  '전화번호':value_list[4],
20                  '할인&이용정보':value_list[5]}
21
22            df_list.append(v1) # 디렉터리 형태로 저장된 가맹점 정보를 list에 추가
23            driver.find_element(By.XPATH, '//*[@id="list_paging"]/a[{}].format(j)).click() #다음 페이지 클릭
24            num = int(driver.find_element(By.XPATH, '//*[@id="list_paging"]/strong').text) #화면에 보이는 페이지에서 마지막 페이지 설정

1 df = df.append(df_list, ignore_index=True) # 가맹점 정보가 포함된 list를 df에 추가
2 df = df.drop_duplicates() # 중복된 행 제거
3 df = df.reset_index()
4 df
```

2) 가상고객 거래 데이터 생성 코드

```
# 대분류 랜덤 선택 -> 소분류 랜덤 선택 -> 가격 랜덤 선택

def lcat_rc(df):
    l_list = list(df.대분류.unique()) #대분류 값들이 들어갈 리스트(도서, 영화 등)
    return random.choice(l_list)

def scat_rc(lcat, df):
    s_list = list(df[df['대분류']==lcat].소분류.unique()) #해당하는 대분류 내의 소분류들 리스트
    return random.choice(s_list)

def gamaeng_rc(scat, df):
    g_list = list(df[df['소분류']==scat].가맹점.unique()) #해당하는 소분류 내의 가맹점 리스트
    return random.choice(g_list)

def amtrc(gamaeng, df):
    s_amt = df[df['가맹점']==gamaeng].iloc[:,7:-3].values #해당하는 가맹점의 가격 10개
    lower = np.percentile(s_amt, 10) #하위 10%센트
    upper = np.percentile(s_amt, 90) #상위 10%센트
    s_amt = s_amt[s_amt>=lower] #하위금액 제거 (중호 포함한 이유: 아예 데이터가 없어서 버리면 계산오류 발생하므로)
    s_amt = s_amt[s_amt<=90000] #9만원이상 제거
    s_amt = s_amt[s_amt<=upper] #상위금액제거
    return random.choice(s_amt)
```

```
# 가상 고객 데이터 셋 생성

def virtual_user(cnt,df):
    f_df = pd.DataFrame(index=range(0), columns=['고객id', '대분류', '소분류', '가맹점', '구매금액', '가맹점 ID']) #빈데이터프레임 생성
    for i in range(1, cnt+1): #만들어줄 가상고객 데이터 수만큼
        a = 0
        while(a<6): #인당 평균 6~7회 사용
            a+=1
            total_amt = 0
            large_cate = lcat_rc(df) #대분류 선택
            small_cate = scat_rc(large_cate, df) #소분류 선택
            ga_cate = gamaeng_rc(small_cate, df) #가맹점 선택
            money = amtrc(ga_cate, df) #가격 선택
            if (f_df[f_df['고객id']==f'user_{str(i)}']['구매금액'].sum() + money) >= 90000: #만약 마지막으
                로 구매한 상품과 이전 구매한 상품과의 가격이 9만원이 넘으면 continue
            else:#아니면 해당 값을 적재
                f_df=f_df.append({'고객id': 'user_'+str(i), '대분류': large_cate, '소분류': small_cate, '가맹점':ga_cate, '구매금액':money}, ignore_index=True)
        return f_df

final_df = virtual_user(5000, data)#생성
final_df
```

3) ALS 모델링

```
# 초기화 변수를 논문에서 가장 좋은 결과를 냈다고 나온 값인 40, 200, 40으로 초기화
r_lambda = 40 #정규화에 필요한 lambda
nf = 200 # 사용자와 아이템의 Latent Factor 행렬의 차원 nf
alpha = 40 # confidenc level 조정에 필요한 alpha
```

```
R = data.values
```

```
# 사용자 5000 / 가맹점 101
print(R.shape)
```

```
nu = R.shape[0] # 사용자
ni = R.shape[1] # 가맹점
```

```
# X, Y 값을 아주 작은 랜덤 값으로 설정
X = np.random.rand(nu, nf) * 0.01
Y = np.random.rand(ni, nf) * 0.01
```

```
#선호도 행렬
P = np.copy(R)
P[P > 0] = 1
print(P)
```

```
#신뢰도 행렬
C = 1 + alpha * R
print(C)
```

#loss 측정 설정

```
def loss_function(C, P, xTy, X, Y, r_lambda):
    predict_error = np.square(P - xTy)
    confidence_error = np.sum(C * predict_error)
    regularization = r_lambda * (np.sum(np.square(X)) + np.sum(np.square(Y)))
    total_loss = confidence_error + regularization
    return np.sum(predict_error), confidence_error, regularization, total_loss
```

#사용자 optimizer 설정

```
def optimize_user(X, Y, C, P, nu, nf, r_lambda):
    yT = np.transpose(Y)
    for u in range(nu):
        Cu = np.diag(C[u])
        yT_Cu_y = np.matmul(np.matmul(yT, Cu), Y)
        lI = np.dot(r_lambda, np.identity(nf))
        yT_Cu_pu = np.matmul(np.matmul(yT, Cu), P[u])
        X[u] = np.linalg.solve(yT_Cu_y + lI, yT_Cu_pu)
```

#가맹점 optimizer 설정

```
def optimize_item(X, Y, C, P, ni, nf, r_lambda):
    xT = np.transpose(X)
    for i in range(ni):
        Ci = np.diag(C[:, i])
        xT_Ci_x = np.matmul(np.matmul(xT, Ci), X)
        lI = np.dot(r_lambda, np.identity(nf))
        xT_Ci_pi = np.matmul(np.matmul(xT, Ci), P[:, i])
        Y[i] = np.linalg.solve(xT_Ci_x + lI, xT_Ci_pi)
```

```
predict_errors = []
confidence_errors = []
regularization_list = []
total_losses = []
```

#모델 학습 진행

#학습 횟수 = 15

```
for i in range(15):
    if i!=0:
        optimize_user(X, Y, C, P, nu, nf, r_lambda)
        optimize_item(X, Y, C, P, ni, nf, r_lambda)
        predict = np.matmul(X, np.transpose(Y))
        predict_error, confidence_error, regularization, total_loss = loss_function(C, P, predict, X, Y, r_lambda)
```

```
predict_errors.append(predict_error)
confidence_errors.append(confidence_error)
regularization_list.append(regularization)
total_losses.append(total_loss)
```

#각 step별 Loss function 출력

```
print('-----step %d-----' % i)
print("predict error: %f" % predict_error)
print("confidence error: %f" % confidence_error)
print("regularization: %f" % regularization)
print("total loss: %f" % total_loss)
```

```
predict = np.matmul(X, np.transpose(Y))
print('final predict')
print([predict])
```

#결과 시각화

```
from matplotlib import pyplot as plt
%matplotlib inline
```

```
plt.subplots_adjust(wspace=100.0, hspace=20.0)
fig = plt.figure()
fig.set_figheight(10)
fig.set_figwidth(10)
predict_error_line = fig.add_subplot(2, 2, 1)
confidence_error_line = fig.add_subplot(2, 2, 2)
regularization_error_line = fig.add_subplot(2, 2, 3)
total_loss_line = fig.add_subplot(2, 2, 4)
```

```
predict_error_line.set_title("Predict Error")
predict_error_line.plot(predict_errors)
```

```
confidence_error_line.set_title("Confidence Error")
confidence_error_line.plot(confidence_errors)
```

```
regularization_error_line.set_title("Regularization")
regularization_error_line.plot(regularization_list)
```

```
total_loss_line.set_title("Total Loss")
total_loss_line.plot(total_losses)
plt.show()
```

```
predict_df = pd.DataFrame(predict)
predict_df['user_id'] = data['user_id'].tolist()
predict_df.set_index('user_id', inplace=True)
predict_df.columns.names = ["frn_id"]
als_pred = predict_df
# als_pred.to_csv('./필수데이터/최종/가맹점_ALS_pred.csv', encoding='utf-8')
```

#ALS 모델을 통해 생성된 고객 선호도 점수를 기반으로 추천 가맹점 목록 추출

```
def recommend(df_preds, frn, user_id):
    sorted_user_predictions = df_preds.loc[user_id].sort_values(ascending=False)

    recommendations = pd.DataFrame(sorted_user_predictions).reset_index()

    recommendations = recommendations.rename(columns= {user_id:'선호도'})

    recommendations['frn_id'] = pd.to_numeric(recommendations['frn_id'])

    recommendations1 = pd.merge(frn, recommendations, left_on='frn_id', right_on='frn_id',
    how='left').sort_values(by= ['선호도'], ascending=False)

    recommendations1.set_index('frn_id', inplace=True)

    return recommendations1
```


4) 고객의 잔액, 지역, 코로나 위험단계에 따른 필터링 함수 세부 코드

```
#추출된 가맹점 목록을 각 요소별로 필터링 후 최종 추천 목록 제시
def balance_filter():
    #정보 입력 받기
    user_id = input(">>> 유저 ID를 입력하세요 : ")
    choice = input(">>> 분류를 선택하세요 : ")
    corona = int(input(">>> 코로나 위험단계를 입력하세요 : "))
    c_balance = input(">>> 잔액필터 적용 여부를 선택하세요(0,X) : ")

    #유저 지역 불러오기
    find_local = user_local['user_id'] == user_id
    u_local = user_local[find_local]
    local = u_local.iloc[0,1]

    #고객별 잔액 구하기
    is_user = filter_data['고객id'] == user_id
    subset = filter_data[is_user]
    balance = int(90000 - sum(subset['구매금액']))

    #소분류 선택
    is_choice = fran['소분류'] == choice
    is_under = fran['최저값'] <= balance
    if c_balance == '0' or 'o' :
        get_content = fran[is_choice & is_under]
    elif c_balance == 'X' or 'x' :
        get_content = fran[is_choice]

    #지역 filter
    local_c = get_content['지역'] == local
    result_local_df_offline = get_content[local_c]
    online_c = get_content['온라인'] == 1
    online_df = get_content[online_c]
    result_local_df = pd.concat([result_local_df_offline,online_df])

    #코로나 filter
    if corona == 3 :
        find_corona = result_local_df['코로나'] <= 1
        result_local_corona_df = result_local_df[find_corona]
        if result_local_corona_df.empty:
            print("\n 가맹점이 없습니다. \n")
    elif corona == 2 :
        find_corona = result_local_df['코로나'] <= 2
        result_local_corona_df = result_local_df[find_corona]
    else :
        result_local_corona_df = result_local_df

    #SVD_score로 sorting
    svd = recommend(als_pred,fran,f'{user_id}')
    sort_df = pd.merge(svd, result_local_corona_df, how='right')
    sort_df.sort_values(by='선호도', ascending=False)
    sort_result = sort_df.iloc[:,2:6]
    sort_result_df = pd.DataFrame(sort_result)
    sort_result_df.fillna('온라인', inplace=True)
    sort_result_df.drop(['온라인'],axis='columns',inplace=True)
    return sort_result_df
```

5) ALS 모델 세분류 선호 예측 코드 실행 및 결과

```
# 논문에서 발췌한 파라미터 최적값들
r_lambda = 40
nf = 200
alpha = 40

import numpy as np

# 기존 고객-소분류 선호도 데이터셋을 행렬로 변환
R = data.values

print(R.shape)

(5000, 16)

nu = R.shape[0]
ni = R.shape[1]

# X(사용자 잠재요인 행렬) 랑 Y(아이템 잠재요인 행렬)값을 아주 작은 랜덤 값으로 초기화
X = np.random.rand(nu, nf) * 0.01
Y = np.random.rand(ni, nf) * 0.01

print(X)

[[0.00287093 0.00170631 0.00381943 ... 0.00632751 0.00871744 0.00666814]
 [0.00529067 0.00209407 0.00059403 ... 0.00427393 0.00880139 0.00450293]
 [0.00852484 0.00985254 0.00139841 ... 0.00049172 0.00328156 0.00827634]
 ...
 [0.00630343 0.00533351 0.0098431 ... 0.00783114 0.0010889 0.00510775]
 [0.00115 0.0077677 0.00388197 ... 0.0030143 0.00727343 0.00520802]
 [0.00645163 0.00433986 0.00163682 ... 0.00143895 0.00888088 0.00917207]]

# 기존 고객-소분류 선호도에서 선호가 있는 부분은 1로 없는 부분은 0으로 선택하여 P선호도 행렬 생성
P = np.copy(R)
P[P > 0] = 1
print(P)

[[0. 0. 0. ... 0. 1. 0.]
 [1. 1. 0. ... 1. 1. 0.]
 [1. 0. 0. ... 0. 0. 1.]
 ...
 [0. 0. 0. ... 0. 0. 1.]
 [0. 1. 0. ... 0. 0. 0.]
 [1. 0. 0. ... 0. 0. 0.]]

# 고객-소분류 선호도의 경우에 신뢰를 주기 위한 절차
# 선호가 0인 값에는 1, 0이 아닌 값은 일정 수준이 되도록 값 변환
C = 1 + alpha * R
print(C)

[[ 1.          1.          1.          ...  1.          42.10393424
  1.          ]
 [ 8.960199    46.33020712  1.          ...  4.073352    184.98903896
  1.          ]
 [ 64.68159204  1.          1.          ...  1.          1.
  8.18918748]
 ...
 [ 1.          1.          1.          ...  1.          1.
 374.83774804]
 [ 1.          11.94177412  1.          ...  1.          1.
 1.          ]
 [ 32.840796   1.          1.          ...  1.          1.
 1.          ]]
```

* 부록 3)에 첨부한 ALS 모델 loss 함수와 optimizer을 통해 잠재 요인 행렬 학습 및 결과

```

-----step 0-----
predict error: 19282.676265
confidence error: 1220897.137077
regularization: 1337.557691
total loss: 1222234.694769
-----step 1-----
predict error: 47466.679090
confidence error: 307417.068424
regularization: 76867.396515
total loss: 384284.464939
-----step 2-----
predict error: 16356.844932
confidence error: 25102.508948
regularization: 35738.839522
total loss: 60841.348470
-----step 3-----
predict error: 7597.575995
confidence error: 12072.239669

-----step 13-----
predict error: 2178.282372
confidence error: 3224.455460
regularization: 36098.591849
total loss: 39323.047310
-----step 14-----
predict error: 2154.708782
confidence error: 3179.388070
regularization: 36000.685088
total loss: 39180.073158
final predict
[array([[0.1841854 , 0.15721323, 0.07550708, ..., 0.18099048, 0.97729298,
         0.14811797],
        [0.9068202 , 0.98222126, 0.10467111, ..., 0.80525287, 1.00313882,
         0.18073085],
        [0.98744783, 0.25952262, 0.14977743, ..., 0.26837395, 0.27026143,
         0.87025164],
        ...,
        [0.16151498, 0.17167803, 0.06717045, ..., 0.15370968, 0.14206503,
         1.01032651],
        [0.28398057, 0.93608953, 0.14145556, ..., 0.30248417, 0.26652033,
         0.21859751],
        [0.97173087, 0.26363795, 0.12181718, ..., 0.28415271, 0.23116764,
         0.17252486]])]

```